Lesson #8

Aug. 3, 2011

APPLICATION OF GENETIC ENGINEERING TO IMMUNOCHEMISTRY (Tai Te Wu)

Introduction

Biomedical engineers play a vital role of applying engineering methods to biomedical problems. During the past several decades, biology and medicine have turned from descriptive sciences to more quantitative analysis by the development of biochemistry, molecular biology, cell biology, bio-informatics, analytical biology, etc. We can now manipulate biological materials for the advancement of medical treatments, using genetic engineering, protein engineering, biotechnology, etc.

Obviously we need to understand what has happened to biology and medicine during recent years. It is no longer developing hard-wares for use in medical therapy or monitoring. The basic advancement started with the discovery of restriction enzymes which can cut DNA molecules at specific sites. It then became possible to make recombinant DNA molecules consisting of human genes and bacterial phage or plasmid molecules. One of the first useful proteins manufactured by this process was human insulin. It has drastically reduced the side effects using porcine insulin in the treatment of diabetes mellitus.

Over the years, this rapidly developing field of genetic engineering and protein engineering havs provided numerous new discoveries in the biomedical area. One of such areas is immunology.

Indeed, immunology has gradually moved from immune-biology to immune-chemistry. The latter discipline involves more chemistry as well as more molecular and cell biology. Not only genetic engineering and protein engineering have provided detailed understanding of the basic principles of immune-chemistry, but also they are crucial in improving the immune system. New genes can be inserted into existing immune cells. New molecules of the immune system can be synthesized. However, before undertaking this task, we need to have a thorough understanding of the macro-molecules and cells in the immune system.

Some basic assumptions have to be made for the students in this course. All students must have had biology in high school. They must have learned the central dogma of DNA to RNA to protein. They must know the four bases of DNA and those of RNA, as well as the twenty common amino acid residues of protein. If any of the student has not be exposed to such basic knowledge, he or she needs to read up on this from a simple textbook on molecular biology. Otherwise this course will be total waste of the student's time.

July 12, 2011

Macromolecules and cells in immunochemistry

Antigen

Antigens can be any foreign material to the human system and commonly used as vaccines. In rare situations such as autoimmunity, some of the antigens can also be human macro-molecules. In most cases, it is a protein or carbohydrate. In the case of carbohydrates, they may have an overall contour similar to some protein molecules. Sometimes, they can be small molecules which may be associated with some macro-molecules. Even some small proteins will be more antigenic if they are cross-linked to form large aggregates.

In the case of invading micro-organisms, such as viruses, bacteria, parasites, etc., their major antigens are mostly proteins with some occasional carbohydrates. For example, in the case of human immune-deficiency virus (HIV), the most important antigen is the trans-membrane protein (TM) or glycoprotein 120 (gp120). Of course humans infected with HIV also consider other viral proteins as antigens.

In the case of pneumococcus, the major antigens are their surface carbohydrate molecules. Most of the studies involving carbohydrate antigens seem to suggest that they are similar to protein antigens so far subsequent biological processes, at least the major histocompatibility complex (MHC) class II molecule binding and interaction with the T cell receptor for antigen, are concerned (Cobb BA, Kasper DL (2008) Characterization of carbohydrate antigen binding to the presentation protein HLA-DR. Glycobiology, 18, 707-718) .

Although DNA itself should not be antigenic, DNA molecules can direct the synthesis of proteins so that DNA can sometimes be used as antigens for vaccination. Therefore, most of the detailed analyses of antigens are primarily concentrated in the studies of protein molecules. They can be internalized into different human cells with various processes. In the case of B lymphocytes, some of them may have a membrane bound antibody molecule with reasonably high binding affinity with the foreign antigen. The antigen-antibody complex is then internalized. For monocytes, macrophages and dendritic cells, they simply take up the antigen molecules by phagocytosis. Once inside these different human cells, the protein antigen will be degraded into fragments of various lengths. Some of these fragments constitute the processed peptides. In the case of carbohydrate molecules, they are also degraded into different fragments mimicking the properties of the protein fragments.

Thus we need to study the various cells which internalize different foreign antigens. They include B lymphocytes, monocytes, macrophages, dendritic cells, and many other cells with special designations.

Glycobiology vol. 18 no. 9 pp. 707–718, 2008
doi:10.1093/glycob/cwn050
Advance Access publication on June 4, 2008

Characteristics of carbohydrate antigen binding to the presentation protein HLA-DR

Brian A Cobb[1,2] and Dennis L Kasper[3]

[2]Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH, USA; and [3]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

Received on March 25, 2008; revised on May 27, 2008; accepted on May 27, 2008

Zwitterionic polysaccharide antigens (ZPSs) were recently shown to activate T cells in a class II major histocompatibility complex (MHCII)-dependent fashion requiring antigen presenting cell (APC)-mediated oxidative processing although little is known about the mechanism or affinity of carbohydrate presentation (Cobb BA, Wang Q, Tzianabos AO, Kasper DL. 2004. Polysaccharide processing and presentation by the MHCII pathway. *Cell*. 117:677–687). A recent study showed that the helical conformation of ZPSs (Wang Y, Kalka-Moll WM, Roehrl MH, Kasper DL. 2000. Structural basis of the abscess-modulating polysaccharide A2 from *Bacteroides fragilis*. *Proc Natl Acad Sci USA*. 97:13478–13483; Choi YH, Roehrl MH, Kasper DL, Wang JY. 2002. A unique structural pattern shared by T-cell-activating and abscess-regulating zwitterionic polysaccharides. *Biochemistry*. 41:15144–15151) is closely linked with immunogenic activity (Tzianabos AO, Onderdonk AB, Rosner B, Cisneros RL, Kasper DL. 1993. Structural features of polysaccharides that induce intra-abdominal abscesses. *Science*. 262:416–419) and is stabilized by a zwitterionic charge motif (Kreisman LS, Friedman JH, Neaga A, Cobb BA. 2007. Structure and function relations with a T-cell-activating polysaccharide antigen using circular dichroism. *Glycobiology*. 17:46–55), suggesting a strong carbohydrate structure–function relationship. In this study, we show that PSA, the ZPS from *Bacteroides fragilis*, associates with MHCII at high affinity and 1:1 stoichiometry through a mechanism mirroring peptide presentation. Interestingly, PSA binding was mutually exclusive with common MHCII antigens and showed significant allelic differences in binding affinity. The antigen exchange factor HLA-DM that catalyzes peptide antigen association with MHCII also increased the rate of ZPS association and was required for APC presentation and ZPS-mediated T cell activation. Finally, the zwitterionic nature of these antigens was required only for MHCII binding, and not endocytosis, processing, or vesicular trafficking to MHCII-containing vesicles. This report is the first quantitative analysis of the binding mechanism of carbohydrate antigens with MHCII and leads to a novel model for non-traditional MHCII antigen presentation during bacterial infections.

[1]To whom correspondence should be addressed: Tel: +1-216-368-1263; Fax +1-216-368-0494; e-mail: brian.cobb@case.edu

Keywords: antigen binding/antigen processing/carbohydrate/MHCII presentation

Introduction

The cornerstone of adaptive immunity is the specific molecular recognition of foreign molecules by immune system receptors. For conventional exogenous protein antigens, this is achieved through proteolytic processing to short peptides, class II major histocompatibility complex (MHCII) binding (Jardetzky et al. 1990; Smith et al. 1998), and cell surface presentation (Abbas et al. 2000). These peptides associate with MHCII in a groove that accommodates a single antigen and is formed by two parallel α-helices sitting on top of a β-sheet. Peptides are linear in structure while bound and are characterized by relatively low affinities, usually approximately 1 μM (Abbas et al. 2000) with very slow k_{on} and k_{off} rates, facilitating the long complex half life required for MHCII-peptide complexes presented at the cell surface. For in vitro binding studies, affinity can be measured with extended incubation times (>24 h) to overcome the slow rates; however, the intracellular loading of MHCII is dependent upon HLA-DM (DM) which increases the rate of association through promoting the exit of self-peptides that associate with MHCII during synthesis in the ER (Denzin and Cresswell 1995). DM functions at a pH optimum between 4.5 and 6.0, correlating well with the acidic environment of the MHCII compartment where peptide loading occurs. Furthermore, hydrogen bonds between the peptide backbone and hydrophobic interactions with peptide side chains dominate peptide–MHCII interactions. In contrast, superantigens are MHCII-dependent molecules that are presented without processing as an intact protein bound outside of the normal peptide binding domain, usually to MHCII molecules already loaded with a peptide (Jardetzky et al. 1994; Hogan et al. 2001). While both of these mechanisms are well understood from biochemical and structural interrogation, essentially nothing is known about the binding mechanism for carbohydrate antigens known to activate immune responses via MHCII-dependent presentation (Cobb et al. 2004).

Until recently, MHCII was only known to present peptide or superantigens; thus, the discovery that the MHCII pathway is also responsible for carbohydrate antigen presentation and ultimately T cell recognition alters the fundamental model of MHCII-antigen binding. Zwitterionic polysaccharide antigen (ZPS) molecules require processing to a low molecular weight form, much like conventional peptide antigens, only the mechanism for cleavage is nitric oxide dependent (Cobb et al. 2004). The presented form ranges from 3 to 15 kDa (Cobb et al. 2004), which is larger than most peptides that are usually between 13 and 18 amino acids in length (i.e., <2 kDa). In T cell assays, antibodies against the MHCII allele family HLA-DR blocked ZPS-mediated T cell activation while neither HLA-DP nor HLA-DQ

Downloaded from glycob.oxfordjournals.org at Northwestern University Library, Serials Department on July 12, 2011

© The Author 2008. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org

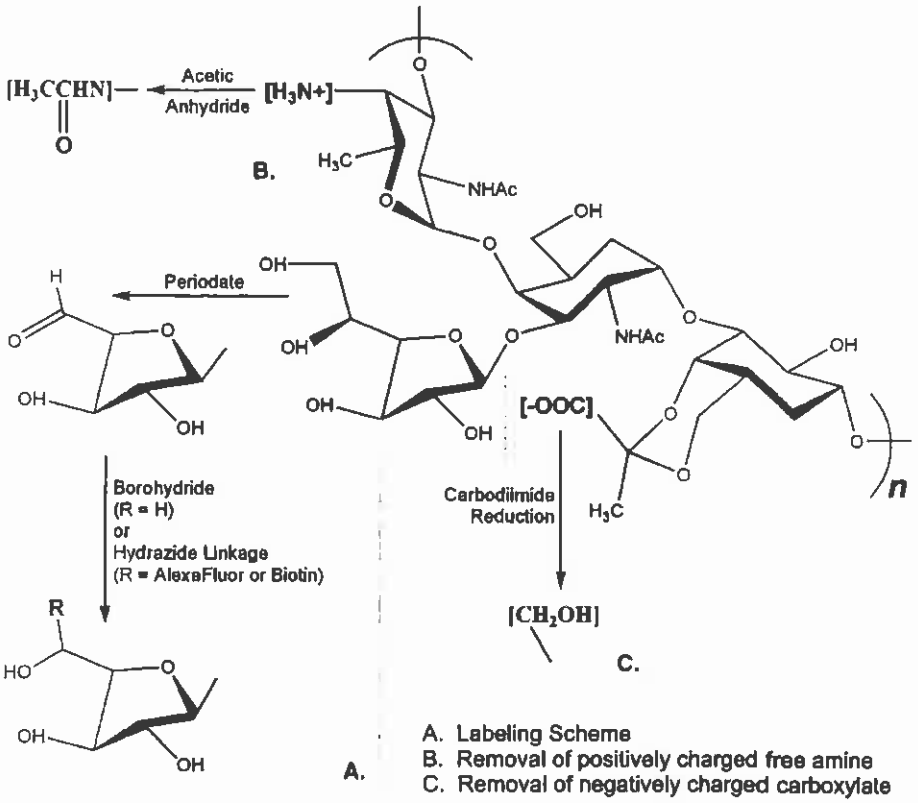


Fig. 1. Chemical structure of PSA from *B. fragilis*. (**A**) Labeling reactions with hydrazide-linked biotin and AlexaFluor594 were targeted to the periodate-activated side chain galactofuranose. Radiolabeling was accomplished by tritiated borohydride reduction of periodate-created aldehydes. (**B**) Positive free amines were neutralized through acetylation by acetic anhydride. (C) Negative carboxylates were neutralized by carbodiimide and borohydride reduction.

antibodies had an effect (Kalka-Moll et al. 2002). These data, although limited by the number of specific alleles tested, suggest that ZPS binding and presentation is restricted to the HLA-DR family. In vitro binding assays and co-immunoprecipitation (co-IP) experiments have confirmed a direct MHCII–ZPS interaction that seems to be affected by the presence of DM (Cobb et al. 2004) although these assays were limited to the affirmation of binding rather than elucidating the biochemical binding mechanism or the functional role of DM.

All known ZPS molecules have in common a zwitterionic charge motif within the repeating unit. These include the capsular polysaccharides Sp1 from type I *Streptococcus pneumoniae* (Tzianabos et al. 1993; Stingele et al. 2004), CP from types 5 and 8 *Staphylococcus aureus* (Tzianabos et al. 2001), and PSA from *Bacteroides fragilis* (Tzianabos et al. 1993; Brubaker et al. 1999). The primary structure of PSA consists of a tetrasaccharide repeating unit (Figure 1) containing 4,6-pyruvate attached to a D-galactopyranose, 2,4-dideoxy-4-amino-D-FucNAc, D-*N*-acetylgalactosamine, and D-galactofuranose (Baumann et al. 1992). Within this repeating unit, one positively charge free amine and one negatively charged carboxylate are present. Interestingly, NMR solution structures of PSA and Sp1 show a stable, extended right-handed helix with three repeating units per helical turn and a pitch of 20 Å (Wang et al. 2000; Choi et al. 2002). The large negatively charged grooves within the helix are 10 Å wide, 10 Å long, and 5 Å deep, with the outer surfaces strongly positive. In silico modeling further suggests that these large negatively charged carbohydrate grooves could envelop a protein α-helix in a manner analogous to α-helix binding to the major groove of DNA, providing one plausible structure-based model for carbohydrate–protein interactions.

More recently, circular dichroism has been used to monitor the helical content of PSA to better understand the role of conformation in antigenicity (Kreisman et al. 2007). The PSA helix is remarkably stable to temperature and yet is sensitive to nonpolar solvents and polymer length. It was shown that the inclusion of 20% acetonitrile increases the molar ellipticity to the same extent seen in the α-helix-rich protein myoglobin. Since three repeating units (~3 kDa) are required to make a full helical turn, it is also not surprising that PSA molecules below 3 kDa are in a random coil conformation that lacks the ability to associate with HLA-DR1. Interestingly, PSA fragments between 3 and 10 kDa that maintain a helical conformation seem to bind better than larger ones (Kreisman et al. 2007) although the affinity has not yet been determined, and this agrees with an earlier study that demonstrated ZPSs smaller than 5 kDa fail to activate T cells. Collectively, these data indicate that carbohydrate processing is required to generate fragments between 3 and 10 kDa that maintain helical conformation to facilitate MHCII binding and presentation.

Finally, the zwitterionic motif is also known to be critical for antigenic function of ZPSs. It was demonstrated nearly 15 years ago that the removal of either a positive or negative group results in an immunologically inert carbohydrate (Tzianabos et al. 1993). In fact, the introduction of a zwitterionic motif into a normally nonzwitterionic polysaccharide converts

Downloaded from glycob.oxfordjournals.org at Northwestern University Library, Serials Department on July 12, 2011

it to a T cell-activating antigen. To extend those findings, CD has been used to show that this charge motif is required for the stability of the helical conformation of PSA and leads to less binding to MHCII (Kreisman et al. 2007), pointing to key electrostatic contributions to the structure and function of ZPS antigens. Together, these observations point to a relationship between ZPS helical content, activity, electrostatic interactions and MHCII-mediated presentation.

Although direct interactions between HLA-DR2 (DR2) (Cobb et al. 2004), HLA-DR1 (DR1) (Kreisman et al. 2007), and PSA have been demonstrated, the binding affinity, stoichiometry, types of bonds that dominate the interaction, and allelic specificity remain completely undetermined. DM has been shown to influence the binding of PSA to DR2 (Cobb et al. 2004), but we have no knowledge of the carbohydrate-specific mechanism of DM action, if it influences MHCII presentation in live cells, or if DM is required for T cell activation. Finally, no information is available on how ZPS binding to MHCII relates to peptide and superantigen association. In this report, we demonstrate high affinity interactions between the DR family and preprocessed PSA. Our results indicate that the interaction is one to one and shows allelic selectivity between DR1, DR2, and DR4 as measured by dissociation constants (K_d). Binding is sensitive to ionic strength and pH, indicating that both electrostatic bonds and ionizable groups play a central role in PSA presentation. We also show competitive binding between PSA, peptides, and superantigens, suggesting that the binding domains for these antigens overlap. DM is shown to be critically important for cellular surface presentation of PSA and in vivo T cell activation due to its ability to increase the rate of binding. Finally, we show that the loss of the alternating charge motif results in the ablation of in vitro and in vivo MHCII binding, whereas this motif does not significantly affect uptake, vesicular trafficking, or processing. These results are the first biochemical characterization of carbohydrate antigen presentation by MHCII proteins and demonstrate the requirement for acidic pH, self-peptide release, and catalysis by DM to achieve surface presentation and T cell recognition.

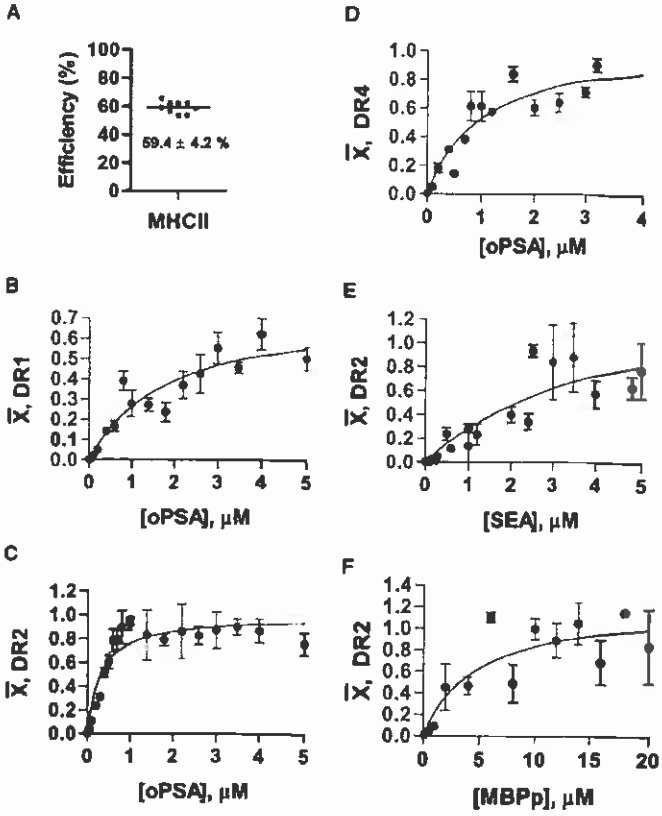


Fig. 2. Analysis of MHCII antigen binding. **(A)** The plate capture efficiency for immobilization of DR protein was measured spectroscopically ($n = 9$) to facilitate the calculation of binding stoichiometries. 59.4% of the DR incubated in the Immulon 2 HB wells became immobilized, with 40.6% of the protein remaining in the supernatant as measured by absorbance at 280 nm. In subsequent binding assays, 0.1 μg DR construct was coated into wells (0.0594 μg actually immobilized) at 37°C and pH 5.0, except SEA (at pH 7.3). The degree of binding ($\bar{X}$) is the moles ligand bound per mole of MHCII; thus, the B_{max} represents the stoichiometry. All data points represent the average of at least four replicates. PSA saturably binds to DR1 (panel **B**; $K_d = 1.9 \pm 0.4$ μM; $B_{max} = 0.8 \pm 0.08$), DR2 (panel **C**; $K_d = 0.31 \pm 0.05$ μM; $B_{max} = 1.0 \pm 0.04$), and DR4 (panel **D**; $K_d = 1.0 \pm 0.3$ μM; $B_{max} = 1.0 \pm 0.1$). Control binding with SEA (**E**) and MBPp (**F**) with DR2 show K_d values of 4.7 ± 2.0 μM and 4.3 ± 2.6 μM with B_{max} values of 1.5 ± 0.5 and 1.2 ± 0.2, respectively.

Results

MHCII–PSA binding affinity and stoichiometry

The MHCII affinity, stoichiometry, and specificity for the model zwitterionic polysaccharide antigen PSA was carried out using in vitro preprocessed (average MW = 15 kDa) and biotinylated (Figure 1A) PSA in equilibrium binding experiments with previously described recombinant MHCII alleles (Table I) (Gauthier et al. 1998; Stratikos et al. 2002; Day et al. 2003). The DP and DQ proteins were previously shown not to mediate T cell activation with ZPSs (Kalka-Moll et al. 2002); thus, we focused this study on the DR family. First, the amount of DR protein immobilized on the plates following conjugation was determined spectroscopically to facilitate the calculation of the degree of binding ($\bar{X}$); the moles of bound carbohydrate per mole of immobilized MHCII at each data point. We found that 59.4 ± 4.2% of the input protein conjugated to the well (Figure 2A).

For the three HLA-DR alleles, $\bar{X}$ was measured with varying amounts of ligand, and then plotted as a function of ligand concentration to calculate the maximal degree of binding (i.e., stoichiometry; B_{max}) and affinity (K_d) of interaction. We found that PSA has a K_d of 1.9 ± 0.4 μM and a B_{max} of 0.8 ± 0.08 with DR1 (Figure 2B). With DR2 ($K_d = 0.31 \pm 0.05$ μM; Figure 2C), the affinity is 6-fold tighter than DR1 and 3-fold tighter than DR4 ($K_d = 1.0 \pm 0.3$ μM, Figure 2D), demonstrating allelic selectivity. Like DR1, both DR2 and DR4 binding

Table I.

HLA-DR constructs		
Construct name (reference)	Allele haplotype	Encoded peptide
HLA-DR1 (20)	DRA*0101, DRB1*0101	None
HLA-DR2 (19)	DRA*0101, DRB1*1501	MBPp (residues 85–99)
HLA-DR4 (22)	DRA*0101, DRB1*0401	CLIP

Downloaded from glycob.oxfordjournals.org at Northwestern University Library, Serials Department on July 12, 2011

show 1:1 stoichiometry (B_{max} = 1.0 ± 0.04 and 1.0 ± 0.1, respectively). As controls, the superantigen staphylococcal enterotoxin A (SEA; Figure 2E) and myelin basic protein peptide (MBPp; residues 85–99; Figure 2F) were used in DR2 binding curves, both showing μM affinity (K_d = 4.7 ± 2.0 μM and 4.3 ± 2.6 μM, respectively). MBPp bound with the expected 1:1 stoichiometry (B_{max} = 1.2 ± 0.2) while SEA shows a stoichiometry between 1 and 2 to 1 (B_{max} = 1.5 ± 0.5), likely reflecting the ability of SEA to simultaneously bind at two sites on DR. Since it is well known that MBPp binds at 1:1 mole ratio with all MHCII proteins, the unity stoichiometry measured here for MBPp binding confirms the immobilization efficiency of DR and demonstrates that essentially all DR molecules are accessible to antigen binding in these assays (Nag et al. 1994; Kalandadze et al. 1996; Ryan et al. 2004).

Salt and pH effects on binding

One of the hallmarks of carbohydrates that activate T cells is the highly charged zwitterionic motif. It is therefore reasonable to hypothesize that electrostatic interactions may play an important role in MHCII presentation. Using varying concentrations of salt to inhibit these potential electrostatic interactions between MHCII and antigens, we found that neither MBP peptide nor superantigen binding was significantly affected by the addition of NaCl (Figure 3A). In contrast, the addition of various NaCl concentrations to block electrostatic bonds inhibited PSA binding by approximately 60% (Figure 3B), showing that critical electrostatic bonds are formed between MHCII and PSA during binding and presentation. To confirm that high salt conditions do not lead to alterations in PSA itself, sizing analysis on a Superose 12 column was performed. No changes in size or aggregation state of PSA were seen as a result of low or high ionic strength conditions (Figure 3C).

Within antigen presenting cells (APCs), the acidification of endo/lysosomes is a characteristic of the MHCII endocytic pathway. In fact, carbohydrate-mediated T cell activation can be inhibited by the addition of an acidification blocker, such as bafilomycin A1 (Stephen et al. 2005), and sensitivity to pH in binding has already been observed previously (Cobb et al. 2004). In this study, we performed DR2 binding assays at pH 7.3 and 5.0 specifically to compare the pH dependency of peptide, superantigen, and carbohydrate presentation by MHCII. We found that PSA (Figure 4B) shows a strong preference for binding in acidic conditions (5-fold change) with far greater sensitivity to pH than either MBPp (2-fold; Figure 4A) or SEA (less than 2-fold; Figure 4A). These results for MBPp are similar to those previously reported (Mukku et al. 1995), showing a 2-fold preference for binding at neutral pH. Finally, to confirm that near neutral and acidic pH conditions do not lead to alterations in PSA independently from MHCII binding, sizing analysis on a Superose 12 column was performed. No significant difference in size or aggregation state of PSA was seen at pH 7.2 and pH 5.0 (Figure 4C). These data point to a novel mode of MHCII-dependent antigen presentation in which cellular entry into acidic endosomal compartments and the presence of electrostatic interactions play dominant roles.

Binding competition

MHCII proteins are associated with the self-peptide CLIP within the vesicular compartments of APCs. For antigenic peptides to achieve presentation, DM catalyzes the competitive binding and exchange of self and foreign peptides into the canonical peptide-binding groove of MHCII. As such, only a single peptide may be bound to MHCII at any one time. In contrast, superantigens tend to bind to peptide-loaded MHCII proteins at the cell surface, forming critical contacts outside of the normal peptide-binding groove (Hogan et al. 2001). In order to demonstrate both the exclusivity of PSA binding and to provide a framework of potential contacts made within and outside the peptide groove of MHCII, we performed competitive binding experiments. First, titrations of nonlabeled tetanus toxin peptide (TTp; Figure 5A) (Boitel et al. 1992) and 10 kDa dextran (not shown) competitors were collected with a static concentration of preprocessed biotin-PSA (1 μM) and DR2 (0.1 μg) at pH 5.0. We found that nonzwitterionic carbohydrates like dextran failed to inhibit PSA interactions, yet TTp efficiently and completely competed for MHCII binding (Figure 5A). Similar results were obtained with a TTp competition of PSA for DR1 binding (data not shown).

In other experiments, biotinylated MBPp and SEA superantigen were monitored for binding to DR2 at pH 7.3 using nonla-

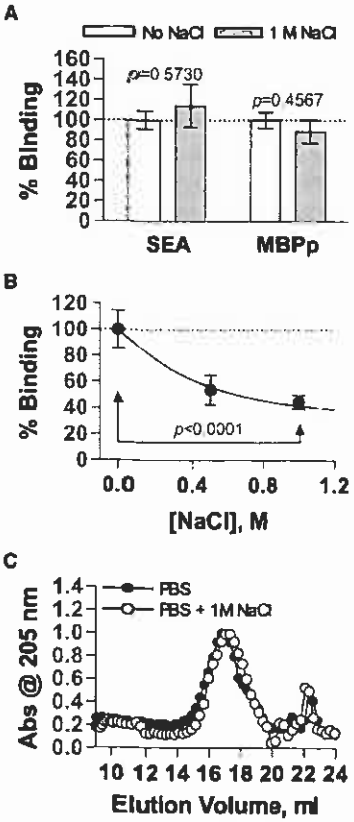


Fig. 3. Salt effects on in vitro MHCII binding. 0.1 μg of DR was used in all of the assays. **(A)** 4 μM MBPp and 4 μM SEA binding to DR2 with zero (open) or 1 M NaCl (shaded) added to the reaction buffer, normalized to the no salt added data with each data point representing the average of three replicates. Only very small effects are observed. **(B)** 1 μM PSA binding to DR2 with 0, 0.5, or 1.0 M NaCl added with each data point representing the average of eight replicates, showing significant loss of binding as a result of high ionic strength conditions. **(C)** Size exclusion elution profiles of PSA ligand used in these assays in high and low ionic strength buffer using a Superose 12 column, confirming that changes in binding are not due to changes in size or aggregation state under the specified ionic conditions.

Downloaded from glycob.oxfordjournals.org at Northwestern University Library, Serials Department on July 12, 2011

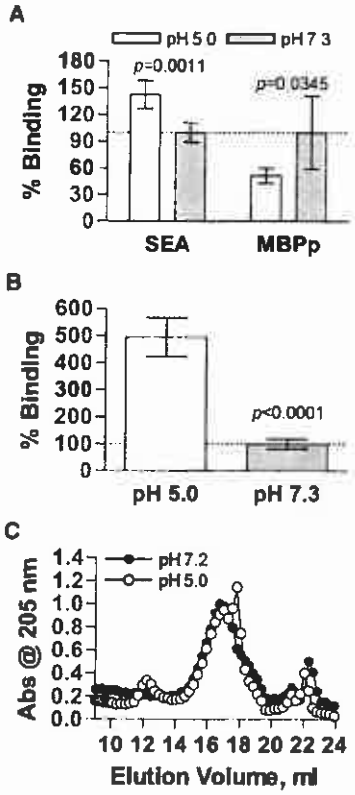


Fig. 4. pH effects on in vitro MHCII binding. (A) Peptide and superantigen binding to DR2 at pH 5.0 (open) compared to pH 7.3 (shaded), normalized to pH 7.3 data, demonstrating modest alterations in binding (2-fold with peptide). (B) PSA binding to DR2 at pH 5.0 and 7.3, showing a profound increase in binding at acidic pH (5-fold), showing that binding of PSA to MHCII is strongly dependent upon an acid environment. (C) Size exclusion elution profiles of PSA ligand used in these assays at neutral and acidic pH using a Superose 12 column, confirming that changes in binding are not due to changes in size or aggregation state at the specified pH value.

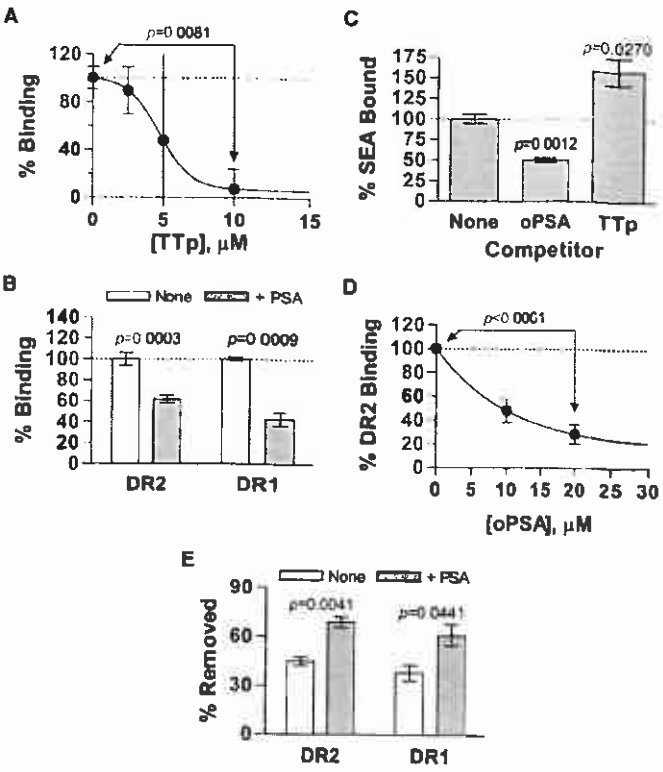


Fig. 5. MHCII competitive binding assays (all normalized to no competitor), with each data point representing at least three replicates. (A) Titration of 1 μM PSA binding competition with a TTp competitor in DR2 binding at pH 5.0, showing complete inhibition of PSA binding at high peptide concentrations. (B) 4 μM MBPp binding (open) is competed with 20 μM PSA (shaded) using DR2 and DR1 at pH 7.3. (C) 4 μM SEA binding to DR1 competes with 20 μM PSA but is enhanced by 20 μM peptide at pH 7.3. (D) 4 μM SEA binding to DR2 with varying concentrations of PSA competitor, showing dose-dependence. (E) MBPp from 0.2 μg of MBPp-saturated DR2 and DR1 was displaced with the addition of a buffer alone (open) but more efficiently with 50 μM PSA (shaded).

beled preprocessed PSA as a competitive inhibitor. At 10-fold molar excess over MBPp, PSA competed with peptide binding to a similar extent as previous peptide-to-peptide competition studies (Nag et al. 1996) although to a lesser degree as seen in Figure 5A. This is likely due to the suboptimal binding of PSA to DR molecules at pH 7.3 (see Figure 4B). Interestingly, PSA also inhibited SEA binding to DR1 by 2-fold at 10-fold molar excess, while TTp enhanced SEA binding (Figure 5C). This agrees with previous reports showing enhanced superantigen binding with varying peptides bound to MHCII grooves. In order to extend these findings, two concentrations of PSA were used in competitive binding assays with DR2 and SEA. We found that at least 80% of SEA binding to MHCII can be blocked by PSA at high concentrations (Figure 5D). Finally, peptide displacement assays were performed by incubating biotinylated MBPp-saturated DR molecules with and without nonlabeled PSA (Figure 5E). We found that with buffer alone, approximately 50% of the bound peptide is lost from the protein after 24 h incubation through the re-establishment of equilibrium; however, the addition of PSA increases the amount of peptide lost from the MHCII protein to approximately 75%. These observations are likely the result of PSA forming contacts inside and outside the peptide-binding groove, rendering PSA binding mutually exclusive (i.e., PSA cannot bind while MHCII is bound to either superantigen or peptide) with all other antigens through the steric hindrance of the necessary contacts.

The role of HLA-DM

We showed previously that the exchange factor DM, which facilitates competitive peptide binding on MHCII, plays an uncharacterized role in PSA presentation (Cobb et al. 2004). Since PSA competitively binds to MHCII, we analyzed the mechanistic role for DM in carbohydrate presentation. PSA time course binding experiments were conducted in the presence and absence of equimolar quantities of DM to DR2 (Figure 6A). Without DM, the time to reach half completion ($T_{1/2}$) is 107.6 ± 16.4 min, whereas the $T_{1/2}$ is 33.3 ± 5.0 min with DM, demonstrating a 3-fold increase in the binding exchange rate in the presence of DM, which is consistent with previous findings using peptides (Busch et al. 2002). This observation was extended using wild-type (wt) and DM-ablated mice ($DM^{-/-}$). In ex vivo presentation assays, MHCII complexes were immunoprecipitated from wt or $DM^{-/-}$ primary splenocytes incubated with radiolabeled PSA. We found that the amount of radioactive carbohydrate coprecipitating with MHCII was reduced significantly in the

Downloaded from glycob.oxfordjournals.org at Northwestern University Library, Serials Department on July 12, 2011

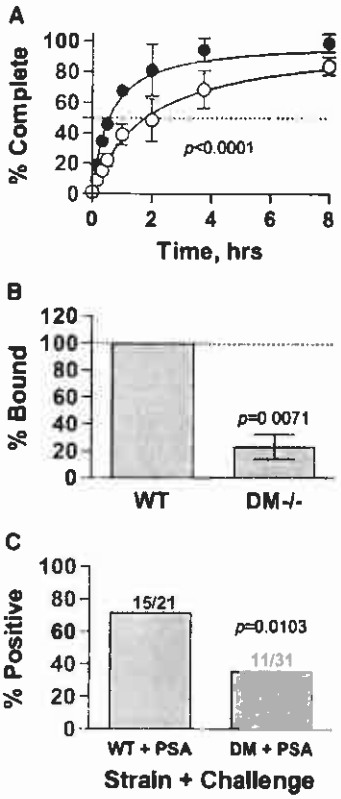


Fig. 6. DM effects on binding. **(A)** 1.0 μM PSA binding time course with 0.1 μg DR2 in the presence (filled) or absence (open) of equimolar DM. Time to reach half completion (dotted line, $T_{\frac{1}{2}}$) in the absence of DM is 107.6 ± 16.4 min, but the rate of binding is increased 3-fold to $T_{\frac{1}{2}} = 33.3 \pm 5.0$ min with DM. Each data point represents the average of three replicates. **(B)** PSA co-IP with an anti-MHCII antibody from wild-type and $DM^{-/-}$ splenocytes. $DM^{-/-}$ cells fail to present significant quantities of PSA in MHCII. Each data point represents the average of two independent experiments with 12 mice in each group per experiment. **(C)** PSA-induced abscess formation (a measure of in vivo T cell activation by PSA) in either wild type or $DM^{-/-}$ animals, showing a 2-fold decrease in abscess formation in the animals lacking DM. The numbers above each bar represent the number of mice with abscesses compared to total.

$DM^{-/-}$ animals (Figure 6B). In addition, in vivo T cell activation studies in these animals, as measured by abscess induction, were impaired by approximately 50% in $DM^{-/-}$ animals compared to wt (Figure 6C). These data collectively demonstrate that DM functions to catalyze PSA binding and that this is required for the efficient cellular presentation of PSA to T cells.

The role of charges in PSA presentation

In previous studies, it was demonstrated that the loss of the zwitterionic charge motif on these antigens by deamination, *N*-acetylation, or carbodiimide reduction results in the ablation of T cell stimulatory activity (Tzianabos et al. 1993); however, it was unclear whether these charged groups were necessary because of a failure in endocytosis, processing, or MHCII binding. To understand the mechanistic role of the zwitterionic motif on the ability of PSA to be presented by MHCII, charge-modified preprocessed PSA molecules (mod-PSA; Figure 1B and C) were created by *N*-acetylation to remove the positive charges (NAc-PSA), carbodiimide reduction to remove the negative charges (Carbo-PSA) or both to generate a neutral antigen (NAc-Carbo-PSA).

First, confocal microscopy studies were performed with fluorescently conjugated native and mod-PSA antigens in APCs to determine if PSA endocytosis and vesicular localization rely upon charge-dependent interactions with host molecules. Native PSA colocalized with MHCII inside and on the surface of human Raji B cells (Figure 7A), whereas all of the mod-PSA antigens colocalized with MHCII only within internal vesicles (Figure 7B–D). These data show that APC internalization and trafficking of PSA are independent of the charge motif, yet the charges are critically important for surface localization, potentially due to a failure to be processed, to bind to MHCII, or both.

In order to determine if mod-PSA antigens are efficiently processed once inside the APCs, radiolabeled native and mod-PSA antigens were incubated with Raji B cells overnight and the endosomal fraction of the cells was isolated and analyzed by chromatography as previously described (Cobb et al. 2004). Native PSA, NAc-PSA, and an anionic polysaccharide control (type III group B streptococcal capsule) were present inside the endosomal fraction in a reduced molecular weight form (open circles; Figure 8A–C, arrows) compared to the input polysaccharide (filled circles). An overlay of the elution profiles shown in panels A and B for the input PSA with and without charge modification is shown in Figure 8D to illustrate that charge modification does not alter the size or state of aggregation of the polysaccharide. These data show that the zwitterionic charge motif neither facilitates nor interferes with antigen processing within the endo/lysosomal pathway.

Since uptake, processing, and vesicular localization were normal for mod-PSA antigens, we then used these antigens in binding assays to determine if the lack of presentation and T cell activation seen with these modifications were the result of a reduction in ZPS affinity for MHCII. We found that both singly charged and neutral mod-PSA completely fail to bind the MHCII protein (Figure 9A). For cellular confirmation, radiolabeled mod-PSA antigens were used in co-IP experiments with anti-MHCII antibodies as described for the $DM^{-/-}$ cells above (Figure 6B). In support of the in vitro binding data, mod-PSA molecules were not presented in the context of MHCII on the cell surface (Figure 9B). Based on these findings, the zwitterionic motif promotes the specific interaction of PSA and MHCII, which also explains why nonzwitterionic carbohydrates are not able to bind MHCII or specifically activate T cells.

Discussion

The recent demonstration of carbohydrate antigen processing and presentation by a MHCII-dependent mechanism in professional APCs (Cobb et al. 2004) shifted the long-standing paradigm limiting MHC presentation to protein antigens (Watts 1997; Watts and Powis 1999). These carbohydrate molecules enter APCs through endocytosis, where they are processed in a NO-dependent fashion to low molecular weight antigens that are loaded onto MHCII proteins for surface presentation and recognition by αβ T cell receptors (Cobb et al. 2004). In the present report, the carbohydrate PSA from the capsule of *B. fragilis* has been used in a series of studies aimed at understanding

Downloaded from glycob.oxfordjournals.org at Northwestern University Library, Serials Department on July 12, 2011

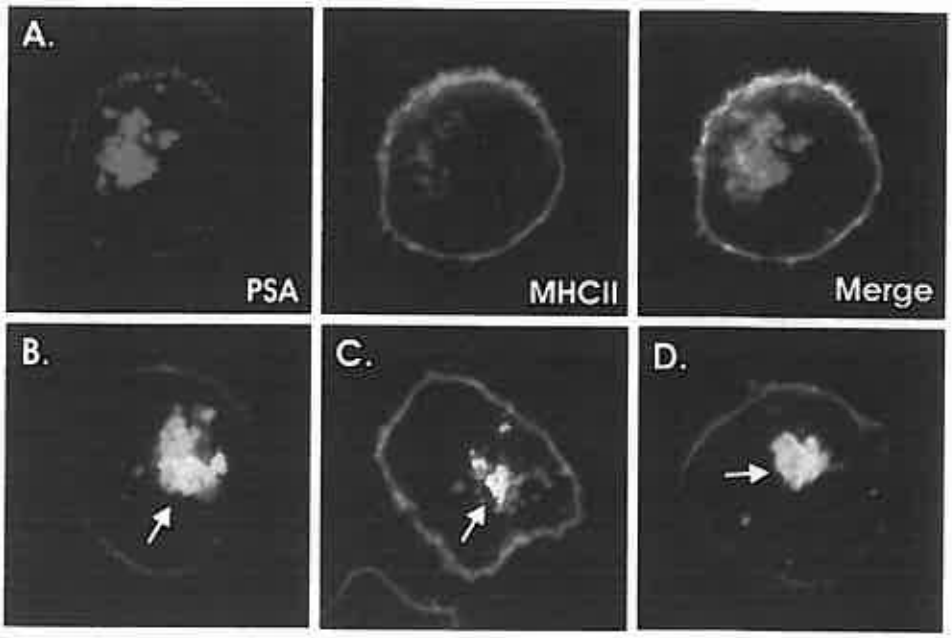


Fig. 7. Charge effects on cellular localization of PSA antigens (red) and MHCII (green) in murine MΦs and Raji cells by confocal microscopy. Surface colocalized PSA and MHCII (yellow) are indicated by arrow heads and internal colocalization (yellow) by arrows. (**A**) The native PSA images, shown as separate channels to illustrate the weakly stained MHCII positive vesicles obscured in the overlay (right) by the strong red signal of the polysaccharide. (**B**) NAc-PSA. (**C**) Carbo-PSA. (**D**) NAc-Carbo-PSA. In agreement with the in vitro binding experiments, none of the mod-PSA samples are presented by APCs at the cell surface, though all can enter vesicular compartments that contain MHCII protein (arrows).

the biochemical mechanism of MHCII-dependent carbohydrate presentation. Our data show for the first time that PSA binds to the HLA-DR family at a 1:1 stoichiometry with high affinity and allelic selectivity, suggesting that specificity and restriction could play an important role in carbohydrate-driven T cell responses. Moreover, PSA binding to MHCII is mutually exclusive with peptide and superantigens, leading to a binding model where PSA could either share or mask contacts both inside and outside of the canonical peptide-binding groove, or induces a conformational shift in MHCII that precludes peptide and superantigen interactions. We previously reported that DM increases the amount of PSA bound to DR2 in vitro (Cobb et al. 2004), but here we demonstrate for the first time that DM increases the binding rate of antigens other than peptides and is required for cellular presentation by MHCII and in vivo T cell activation. Finally, we found that the zwitterionic motif is not required for APC entry, processing, or vesicular colocalization with MHCII, but is critical for MHCII loading and presentation. Indeed, all carbohydrates tested were processed to low molecular weight fragments, regardless of their charge nature and lack of ability to bind MHCII proteins. These data differentiate key attributes between T cell-dependent and -independent carbohydrate antigens while providing a step-by-step picture of the presentation mechanism within host immune cells (Figure 10).

In vitro equilibrium binding assays using three DR alleles show a direct and saturable interaction between preprocessed PSA and MHCII. Interestingly, our data using DR1, DR2, and DR4 suggest that specificity and perhaps even restriction may play an important role in T cell responses against ZPS antigens. Peptides generally average 1 μM affinity with MHCII proteins, which is significantly weaker than many antibody–antigen interactions, such as HyHEL-5 binding to lysozyme at a K_d of 400 pM (Davies and Cohen 1996). It is therefore interesting to note that at 315 nM, the binding affinity of PSA to DR2 is relatively high in comparison to superantigens (Lee and Watts 1990) and peptide antigens (Babbitt et al. 1985).

Many superantigens bind to MHCII molecules in a promiscuous fashion, often with both a high and low affinity site (Papageorgiou and Acharya 1997), as reflected in the 1.5:1 average apparent stoichiometry of SEA binding to DR2 in our studies. The common lack of a one-to-one interaction in such circumstances correlates with the overall lack of specificity. In contrast, peptide antigens are strictly one to one in their association with MHCII molecules and this correlates with a high level of specificity both in presentation as well as recognition (Abbas et al. 2000). Given the relatively high affinity and the observation that all ZPS binding data show B_{max} values at or near unity, our data lend strong support to the conclusion that ZPS antigens are presented by specific MHCII proteins as a one-to-one binary complex at the cell surface.

Furthermore, most peptides and superantigens associate with MHCII proteins primarily through hydrophobic and hydrogen bonding (Stern et al. 1994; Petersson et al. 2002). The in vitro binding experiments at high ionic strength suggest that while hydrogen bonding and hydrophobic interactions likely contribute to PSA binding, electrostatic interactions play a central role in anchoring the ZPS antigen to MHCII. This observation is extended by the results using charge-modified PSA antigens where the alternating charge motif on PSA was found to be specifically required for binding and presentation by MHCII. Although we cannot completely rule out the possibility that the introduction of space-filling acetyl groups is responsible for the reduced activity rather than the simple loss of the positive charge in the NAc-PSA sample, it has been previously demonstrated that a lack

Downloaded from glycob.oxfordjournals.org at Northwestern University Library, Serials Department on July 12, 2011

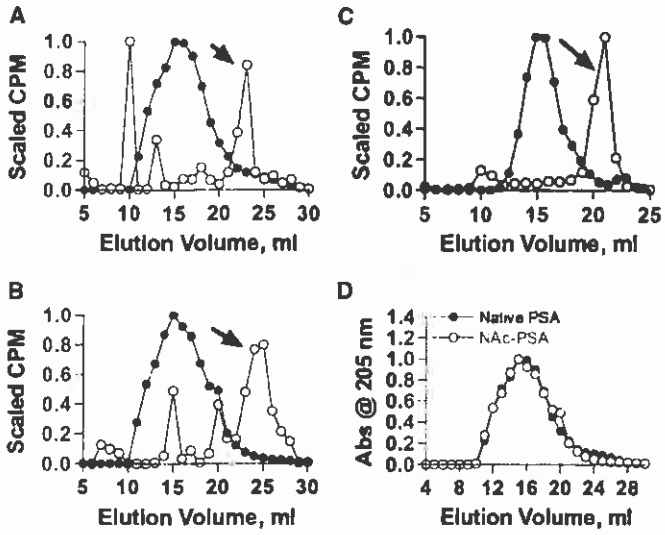


Fig. 8. Charge effects on carbohydrate processing. Superose 12 analysis of (A) native PSA, (B) NAc-PSA, and (C) group B streptococcal (GBS) polysaccharide from APC endosomal compartments (open) compared to untreated controls (filled). PSA is reduced to a low molecular weight product (arrow) that is known to be presented by MHCII. Likewise mod-PSA and nonzwitterionic GBS polysaccharide also enter APCs and are processed (arrow). (D) Overlay plot of the control PSA elution profiles from panels A and B, showing no measurable difference in size or aggregation state of the native polysaccharide (filled) compared to charge-modified PSA (open) prior to endocytosis and processing.

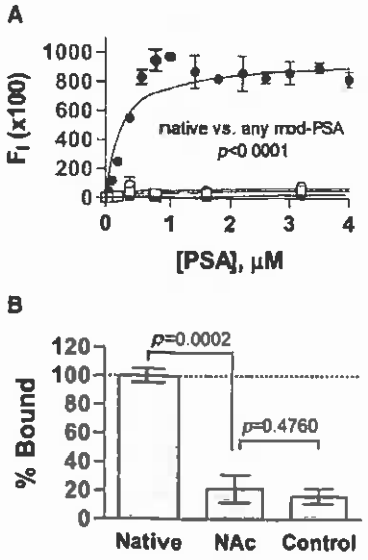


Fig. 9. Charge effects on carbohydrate binding. (A) Native PSA (filled circles), NAc-PSA (open circles), Carbo-PSA (filled squares), and NAc-Carbo-PSA (open squares) binding to DR2 at 37°C and pH 5.0. All mod-PSA antigens failed to bind MHCII. (B) Native and NAc-PSA co-IP with anti-MHCII antibody from wt splenocytes, showing that mod-PSA antigens fail to be presented by APCs over an isotype background control.

of T cell stimulatory activity is also seen in PSA that has been simply deaminated (Tzianabos et al. 1993) and that loss of the zwitterionic motif results in significant conformational changes in PSA (Kreisman et al. 2007). As a result, it is unlikely that the additional acetyl group is playing a role beyond the elimination of the positive charge and change in the structure.

Our data also directly demonstrate that cellular entry and processing are not dependent upon the zwitterionic motif, implying that neither of these mechanisms are mediated by zwitterion-specific receptors or enzymes. Indeed, our observations confirm that mod-PSA antigens arrive in the correct vesicle for MHCII loading and presentation following processing but fail to be presented on the cell surface. Together with the binding data at various ionic strengths, we report that the alternating charge distribution on PSA is the key factor for MHCII presentation due in part to a requirement of electrostatic interactions during presentation.

Previous findings demonstrated a requirement of cellular entry for ZPS presentation to occur (Cobb et al. 2004). In comparing in vitro MHCII binding experiments, we found that PSA binding was remarkably more sensitive to pH (i.e., preferring acidic pH) than other antigen types despite the fact that no significant change in net carbohydrate charge would be expected within that range. Previously published observations even show a lack of T cell activation in bafilomycin A1-treated APCs (Stephen et al. 2005). These data show that unlike many peptides that bind quite well at pH 7.3 and are thus able to bind MHCII proteins on the surface of fixed APCs (Mukku et al. 1995), cell entry is necessary for ZPS molecules to achieve presentation because the association requires an acidic environment.

Published results have also implicated the exchange factor DM in the mechanism of ZPS presentation in that greater in vitro binding was seen with DM present (Cobb et al. 2004). In the present study, we show for the first time that DM catalyzes the rate of PSA binding to MHCII and is required for surface presentation on live APCs. In vivo abscess induction studies in $DM^{-/-}$ mice confirm the importance of DM in the efficiency of PSA-mediated T cell responses, strongly suggesting that peptides and PSA compete for MHCII-mediated presentation that is facilitated by DM. In agreement with this interpretation, time course experiments show an increase in the rate of PSA binding that correlates well with rate enhancements in other recombinant systems. In one landmark report, it was shown that the k_{off} for peptides was enhanced as a linear function of DM concentration (Weber et al. 1996). In the same article, it was shown that without DM, the half-time ($t_{1/2}$) of CLIP dissociation from DR1 was 5.7 h, whereas the presence of DM at 1.5 nM decreased the $t_{1/2}$ to 2.4 h. Increasing the DM concentration to 3.4 nM further decreased $t_{1/2}$ to 1.2 h. This correlates to enhancements of 2.4- and 4.8-fold over assays without DM, which is very similar to the observed 3-fold rate increase for PSA. In addition, we used a 1:1 molar ratio of DR and DM, which is much less than was used in another report that demonstrated an 82-fold increase in k_{on} with DR1 and MBPp (Sloan et al. 1995). In that study, a 200-fold molar excess of DM was used over DR1. Given the linear relationship with DM concentration and activity (Weber et al. 1996), our 3-fold increase extrapolates to an estimated 600-fold increase at 1:200. Finally, with the involvement of DM in exchanging antigens on MHCII molecules, it is not surprising that our competitive binding experiments demonstrate that peptides and superantigens effectively compete for PSA binding.

Given the importance of pH on the binding of ZPS antigens to MHCII, the exclusivity of the interaction, and the requirement

Downloaded from glycob.oxfordjournals.org at Northwestern University Library, Serials Department on July 12, 2011

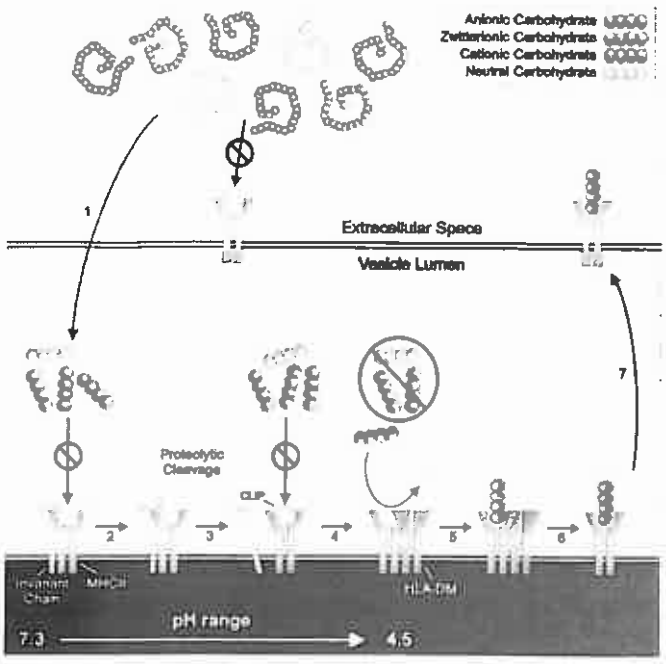


Fig. 10. Schematic of ZPS Antigen Presentation by MHCII. In step 1, carbohydrate antigens enter the vesicular traffic of APCs and are quickly processed to low molecular weight forms. These same molecules cannot bind to surface localized MHCII molecules due to the requirement of low pH and HLA-DM, nor can the processed carbohydrates bind to vesicular Ii-bound MHCII since peptides must first be removed from the binding groove in order for binding to occur. In step 2, the pH drops and cathepsin S cleaves Ii from CLIP. In step 3, carbohydrate antigens cannot bind MHC-CLIP efficiently enough within the cell because HLA-DM is not present. By step 4, the pH has reached 4.5, enabling HLA-DM to catalyze the exchange of CLIP for ZPS molecules—but not singly charged or neutral carbohydrates. In step 5, high affinity electrostatic interactions are formed between PSA and MHCII in an allelic selective manner. In step 6, HLA-DM dissociates from the complex, enabling step 7, where the MHCII-ZPS complex is shuttled to the cell surface for T cell recognition.

for acidic vesicular compartments for APC presentation (Cobb et al. 2004) and T cell activation (Kalka-Moll et al. 2002), a number of general MHCII pathway characteristics become important considerations in understanding the ZPS mechanism of presentation. For example, the pH optimum for DM activity is acidic (Denzin and Cresswell 1995) and the invariant chain (Ii) is cleaved away from MHCII through the action of the acid-activated protease cathepsin S, leaving behind CLIP (Riese et al. 1998). Ii cleavage makes the peptide-binding groove of MHCII available for antigenic peptides to exchange with CLIP because two peptides cannot bind to a single MHCII protein (Abbas et al. 2000). In addition, most superantigens bind to MHCII molecules with or without peptides because these two antigen types form nonexclusive and distinct molecular contacts on MHCII (Hogan et al. 2001).

Integrating these known MHCII pathway characteristics with the observations in this report, a number of important conclusions can be reached. First, the pH must be reduced to allow cathepsin S activation and Ii cleavage to allow for peptide removal prior to PSA binding because PSA and peptide cannot be bound at the same time. Second, a low pH is required to facilitate PSA binding and enable DM activity, which is critical for enabling cellular presentation and in vivo T cell activation through increasing the rate of exchange between peptide and PSA. Third, the competitive nature of the binding suggests that ZPS molecules either share and/or block some portion of the peptide and superantigen binding surfaces, or that binding results in a conformational change in MHCII that significantly reduces the recognition of peptides and superantigens. Finally, our data suggest that any successful ZPS-based vaccine would require entry into the acidic vesicular traffic of APCs to achieve MHCII presentation. Indeed, our observations generate a binding model in which a mutually exclusive and specific interaction of processed PSA and MHCII occurs through the formation of an unusual and uncharacterized antigen/MHCII complex that is recognized by T cells.

This report represents the first mechanistic characterization of carbohydrate antigen binding by MHCII. Interestingly, these data explain why many polysaccharides fail to activate T cells. In fact, nonzwitterionic carbohydrates enter the MHCII pathway and are even processed, but they cannot bind to MHCII even with the assistance from DM. This also provides a rationale for the efficacy of glycoconjugate vaccines in that the carrier protein is clearly not required for APC entry and trafficking, but for polysaccharide presentation by MHCII proteins—likely in the form of glycopeptides. Since the absence of a zwitterionic motif relegates carbohydrates to T cell independence, it is intriguing to point out that the modification of carbohydrates to create a zwitterionic motif introduces the ability to activate T cells (Tzianabos et al. 1993). As a result, this binding study represents the foundation of a potential new vaccine paradigm that allows for the development of carbohydrate-based vaccines against the surface glycans on bacterial pathogens through the introduction of a zwitterionic motif to enable carbohydrates to induce T cell-dependent immune responses via MHCII presentation.

Materials and methods

Cell culture

Raji human lymphoma B cells (ATCC, Cat. CCL-86) were cultured in RPMI 1640 media with 10% fetal bovine serum (FBS). Murine raw MΦs (ATCC, Cat. TIB-71) were cultured in Dulbecco's Modified Eagle Media also with 10% FBS.

Mice

Mouse strains were wt, C57BL/6J, stock 000664, and DM$^{-/-}$, B6.129S4-H2-DMa^{tm1Dim}/J, stock 004513 (Jackson Labs, Bar Harbor, ME) (Miyazaki et al. 1996).

Carbohydrate purification and labeling

Culturing, purification, and analysis by NMR spectroscopy, SDS–PAGE, and wavelength scans were performed as previously described (Tzianabos et al. 1992). PSA was judged to be >99% pure. Radioisotope, AlexaFluor® 594 (Molecular Probes, Eugene, OR), and biotin conjugation were performed as previously described (Cobb et al. 2004) and illustrated in Figure 1.

Peptides, protein expression, and purification

MBPp (residues 85–99) and TTp (residues 830–844) were obtained from New England Peptide (Gardner, MA). DR2 with

Downloaded from glycob.oxfordjournals.org at Northwestern University Library, Serials Department on July 12, 2011

MBPp (Gauthier et al. 1998) was purified from CHO cells as previously described using an antibody affinity column (Cobb et al. 2004). Pure DR1 (Stratikos et al. 2002) and DM (Mosyak et al. 1998) were gifts from Dr. Stephen DeWall, Harvard Medical School. Pure DR4 (Day et al. 2003) was a gift from Dr. Kai Wucherpfennig, Dana Farber Cancer Center.

In vitro binding assays and analysis

For all binding reactions, PSA was preprocessed with ozone for 1 h (MW 15 kDa) and then biotinylated as previously described (Cobb et al. 2004) and illustrated in Figure 1. ELISA wells (Immulon 2 HB, Fisher Scientific, Pittsburgh, PA) were coated with 0.1 μg DR or blocking agent. Unless noted, all PSA binding assays were performed in an acetate/phosphate saline buffer (A/PBS) at 37°C and pH 5.0 for 24 h with mixing. All peptide and superantigen binding assays were similarly performed with the same A/PBS buffer at 37°C for 24 h, only at pH 7.3, unless noted, which is similar to published peptide-binding conditions (Kalandadze et al. 1996; Day et al. 2003). Time course experiments ± 0.1 μg DM (1:1 mole ratio with DR) were performed as described for equilibrium binding assays with variable times. Bound antigens were detected using europium-conjugated streptavidin (Eu-SA) and time-resolved fluorometry (TRF) as previously established (Cobb et al. 2004).

The specific activity of the biotinylated PSA ligand (fluorescence units/mole) was determined using TRF prior to the completion of the binding studies. Briefly, PSA-specific rabbit sera was coated into the wells of Immulon 2 HB plates. 2 μM of radiolabeled PSA with a known cpm/mole specific activity was incubated in the wells for 1 h at room temperature. The unbound PSA was removed and the wells washed twice with phosphate-buffered saline, pooled with the original supernatant, and quantified by scintillation counting. The capture efficiency of the sera was calculated using the unbound and total PSA [efficiency = 1 − (unbound cpm/total cpm)]. Using this value, the experiment was repeated with biotinylated PSA. The amount of biotinylated PSA bound to the plate was calculated using the measured capture efficiency above and divided by the TRF signal upon detection with Eu-SA to yield the specific activity in fluorescence units/mole for conversion of fluorescence units into moles ligand bound in all binding assays reported.

Plating efficiency of the MHCII molecules was determined by coating a known quantity of purified DR into Immulon 2 HB wells overnight at 4°C. The supernatant from those wells was then analyzed for absorbance at 280 nm and compared to control samples that contained equal amounts of DR used for coating. The percentage of protein concentration loss as a result of coating was then used as the immobilization efficiency (Figure 2A).

For analysis of binding data, the degree of binding (moles ligand bound/mole MHCII) was plotted against the ligand concentration and fit by nonlinear curve regression using Prism v. 3.03 (GraphPad Software, San Diego, CA). Affinity (K_d) and maximal binding (B_{max}) were derived from a single-site, single-ligand binding model. For salt and pH effects, the normal saline buffer and pH 7.3 were used as the references (Figures 3 and 4). In general, normalization was performed by dividing the reference value into the corresponding experimental sample values and expressed as a percentage of the reference, which was set to 100%.

Competition

The competition assays were performed as described above only with nonbiotinylated competing antigens added concurrently with the ligand. All assays used 1 μM biotinylated ligand and varying amounts of competitor, as indicated in *Results*. In assays where PSA binding was monitored in the presence of competitors, the pH was 5.0. In assays where biotinylated peptide or SEA binding was monitored in the presence of competitor, the pH was 7.3. For all assays, the reference data were set to the no-competitor control for normalization.

Immunoprecipitations

Primary splenocytes were isolated from 12 wt or $DM^{-/-}$ mice using a Ficoll-Hypaque gradient and cultured in RPMI 1640 media with 10% FBS as previously described (Tzianabos et al. 2000). Cells were incubated for 24 h at 37°C with 1 mg of radiolabeled PSA antigens. Surface MHCII was solublized and immunoprecipitated as previously described (Cobb et al. 2004). Radioactive immunoprecipitates were quantified by scintillation counting. For the knock-out study, normalization was performed with the wt cells as the reference and then calculating the percentage of those counts found in the $DM^{-/-}$ cell precipitates. For mod-PSA co-IP experiments, native PSA was used as the reference.

Abscess induction

Induction of abscesses in mice was performed as previously reported with 50 μg of PSA in sterilized cecal contents (Tzianabos et al. 1999). Animals were examined for the presence of at least one abscess 6 days after challenge.

Psa charge modifications

Positive free amine groups and negative carboxylate groups were eliminated from the PSA repeating unit as previously described (Tzianabos et al. 1993). Briefly, amines were acetylated using acetic anhydride in dH_2O, whereas carboxylates were neutralized using carbodiimide and borohydride reduction.

Processing assays/chromatography

Raji antigen processing assays were performed as previously described (Cobb et al. 2004). Vesicular lysates were analyzed Superose 12 chromatography on a FPLC system (BioLogic HR, BioRad). Fractions were assayed for radioactivity to determine the elution profile.

The effects of 1 M NaCl, acidic pH (pH 5.0), and charge modifications (*N*-acetylation, analyzed in PBS) on the aggregation state of PSA samples were also analyzed by a Superose 12 elution profile on a FPLC system. In all cases, the PSA samples were allowed to incubate for 30 min in the respective conditions at room temperature prior to loading onto the column that was also equilibrated in the same treatment conditions (saline, saline + 1 M NaCl, pH 5.0, or pH 7.2). Elution profiles were collected with absorbance at 205 nm.

Confocal microscopy

Human Raji B cells (~10^6) were incubated with 50 μg of AlexaFluor®594-conjugated native or mod-PSA for 3 h at 37°C.

Downloaded from glycob.oxfordjournals.org at Northwestern University Library, Serials Department on July 12, 2011

Cells were fixed with 4% paraformaldehyde and stained with AlexaFluor® 488-conjugated anti-DR (Raji, clone L243) antibodies for viewing with a Zeiss Pascal confocal microscope (63× objective lens; Carl Zeiss, Thornwood, NY).

Funding

National Institutes of Health, National Institute of Allergy and Infectious Disease (AI039576 to D.L.K., AI062707 to B.A.C.); the Ford Foundation and the National Academy of Sciences (to B.A.C.).

Acknowledgements

We thank Dr. Kai Wucherpfennig for many helpful suggestions in addition to providing the cell line producing HLA-DR2 and for providing the purified HLA-DR4 protein. Also, we thank Dr. Stephen DeWall for helpful advice and providing purified HLA-DR1 and HLA-DM. Finally, we thank Dr. Qun Wang for ^{1}H-NMR analysis of the mod-PSA samples.

Conflict of interest statement

None declared.

Abbreviations

A/PBS, acetate-phosphate saline buffer; APC, antigen presenting cell; carbo-PSA, carbodiimide reduced PSA; co-IP, co-immunoprecipitation; DM, HLA-DM; DR1, HLA-DR1; DR2, HLA-DR2; Eu-SA, europium-conjugated streptavidin; FBS, fetal bovine serum; MBPp, myelin basic protein peptide; MHCII, class II MHC; mod-PSA, charge-modified PSA; NAc-Carbo-PSA, neutral PSA; NAc-PSA, *N*-acetylated PSA; PSA, polysaccharide A; SEA, staphylococcal enterotoxin A; TRF, time-resolved fluorescence; TTp, tetanus toxin peptide; ZPS, zwitterionic polysaccharide.

References

Abbas AK, Lichtman AH, Pober JS. 2000. *Cellular and Molecular Immunology*. 4th ed. New York: W.B. Saunders.

Babbitt BP, Allen PM, Matsueda G, Haber E, Unanue ER. 1985. Binding of immunogenic peptides to Ia histocompatibility molecules. *Nature* 317:359–361.

Baumann H, Tzianabos AO, Brisson JR, Kasper DL, Jennings HJ. 1992. Structural elucidation of two capsular polysaccharides from one strain of *Bacteroides fragilis* using high-resolution NMR spectroscopy. *Biochemistry*. 31:4081–4089.

Boitel B, Ermonval M, Panina-Bordignon P, Mariuzza RA, Lanzavecchia A, Acuto O. 1992. Preferential V beta gene usage and lack of junctional sequence conservation among human T cell receptors specific for a tetanus toxin-derived peptide: Evidence for a dominant role of a germline-encoded V region in antigen/major histocompatibility complex recognition. *J Exp Med*. 175:765–777.

Brubaker JO, Li Q, Tzianabos AO, Kasper DL, Finberg RW. 1999. Mitogenic activity of purified capsular polysaccharide A from *Bacteroides fragilis*: Differential stimulatory effect on mouse and rat lymphocytes in vitro. *J Immunol*. 162:2235–2242.

Busch R, Pashine A, Garcia KC, Mellins ED. 2002. Stabilization of soluble, low-affinity HLA-DM/HLA-DR1 complexes by leucine zippers. *J Immunol Methods*. 263:111–121.

Choi YH, Roehrl MH, Kasper DL, Wang JY. 2002. A unique structural pattern shared by T-cell-activating and abscess-regulating zwitterionic polysaccharides. *Biochemistry*. 41:15144–15151.

Cobb BA, Wang Q, Tzianabos AO, Kasper DL. 2004. Polysaccharide processing and presentation by the MHCII pathway. *Cell*. 117:677–687.

Davies DR, Cohen GH. 1996. Interactions of protein antigens with antibodies. *Proc Natl Acad Sci USA*. 93:7–12.

Day CL, Seth NP, Lucas M, Appel H, Gauthier L, Lauer GM, Robbins GK, Szczepiorkowski ZM, Casson DR, Chung RT, et al. 2003. Ex vivo analysis of human memory CD4 T cells specific for hepatitis C virus using MHC class II tetramers. *J Clin Invest*. 112:831–842.

Denzin LK, Cresswell P. 1995. HLA-DM induces CLIP dissociation from MHC class II alpha beta dimers and facilitates peptide loading. *Cell*. 82:155–165.

Gauthier L, Smith KJ, Pyrdol J, Kalandadze A, Strominger JL, Wiley DC, Wucherpfennig KW. 1998. Expression and crystallization of the complex of HLA-DR2 (DRA, DRB1*1501) and an immunodominant peptide of human myelin basic protein. *Proc Natl Acad Sci USA*. 95:11828–11833.

Hogan RJ, VanBeek J, Broussard DR, Surman SL, Woodland DL. 2001. Identification of MHC class II-associated peptides that promote the presentation of toxic shock syndrome toxin-1 to T cells. *J Immunol*. 166:6514–6522.

Jardetzky TS, Brown JH, Gorga JC, Stern LJ, Urban RG, Chi YI, Stauffacher C, Strominger JL, Wiley DC. 1994. Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen. *Nature*. 368:711–718.

Jardetzky TS, Gorga JC, Busch R, Rothbard J, Strominger JL, Wiley DC. 1990. Peptide binding to HLA-DR1: A peptide with most residues substituted to alanine retains MHC binding. *EMBO J*. 9:1797–1803.

Kalandadze A, Galleno M, Foncerrada L, Strominger JL, Wucherpfennig KW. 1996. Expression of recombinant HLA-DR2 molecules. Replacement of the hydrophobic transmembrane region by a leucine zipper dimerization motif allows the assembly and secretion of soluble DR alpha beta heterodimers. *J Biol Chem*. 271:20156–20162.

Kalka-Moll WM, Tzianabos AO, Bryant PW, Niemeyer M, Ploegh HL, Kasper DL. 2002. Zwitterionic polysaccharides stimulate T cells by MHC class II-dependent interactions. *J Immunol*. 169:6149–6153.

Kreisman LS, Friedman JH, Neaga A, Cobb BA. 2007. Structure and function relations with a T-cell-activating polysaccharide antigen using circular dichroism. *Glycobiology*. 17:46–55.

Lee JM, Watts TH. 1990. Binding of staphylococcal enterotoxin A to purified murine MHC class II molecules in supported lipid bilayers. *J Immunol*. 145:3360–3366.

Miyazaki T, Wolf P, Tourne S, Waltzinger C, Dierich A, Barois N, Ploegh H, Benoist C, Mathis D. 1996. Mice lacking H2-M complexes, enigmatic elements of the MHC class II peptide-loading pathway. *Cell*. 84:531–541.

Mosyak L, Zaller DM, Wiley DC. 1998. The structure of HLA-DM, the peptide exchange catalyst that loads antigen onto class II MHC molecules during antigen presentation. *Immunity*. 9:377–383.

Mukku PV, Passmore D, Phan D, Nag B. 1995. pH dependent binding of high and low affinity myelin basic protein peptides to purified HLA-DR2. *Mol Immunol*. 32:555–564.

Nag B, Arimilli S, Mukku PV, Astafieva I. 1996. Functionally active recombinant alpha and beta chain-peptide complexes of human major histocompatibility class II molecules. *J Biol Chem*. 271:10413–10418.

Nag B, Mukku PV, Arimilli S, Phan D, Deshpande SV, Winkelhake JL. 1994. Antigenic peptide binding to MHC class II molecules at increased peptide concentrations. *Mol Immunol*. 31:1161–1168.

Papageorgiou AC, Acharya KR. 1997. Superantigens as immunomodulators: Recent structural insights. *Structure*. 5:991–996.

Petersson K, Thunnissen M, Forsberg G, Walse B. 2002. Crystal structure of a SEA variant in complex with MHC class II reveals the ability of SEA to crosslink MHC molecules. *Structure (Camb)*. 10:1619–1626.

Riese RJ, Mitchell RN, Villadangos JA, Shi GP, Palmer JT, Karp ER, De Sanctis GT, Ploegh HL, Chapman HA. 1998. Cathepsin S activity regulates antigen presentation and immunity. *J Clin Invest*. 101:2351–2363.

Ryan KR, McNeil LK, Dao C, Jensen PE, Evavold BD. 2004. Modification of peptide interaction with MHC creates TCR partial agonists. *Cell Immunol*. 227:70–78.

Sloan VS, Cameron P, Porter G, Gammon M, Amaya M, Mellins E, Zaller DM. 1995. Mediation by HLA-DM of dissociation of peptides from HLA-DR. *Nature*. 375:802–806.

Smith KJ, Pyrdol J, Gauthier L, Wiley DC, Wucherpfennig KW. 1998. Crystal structure of HLA-DR2 (DRA*0101, DRB1*1501) complexed with a peptide from human myelin basic protein. *J Exp Med*. 188:1511–1520.

Downloaded from glycob.oxfordjournals.org at Northwestern University Library, Serials Department on July 12, 2011

Stephen TL, Niemeyer M, Tzianabos AO, Kroenke M, Kasper DL, Kalka-Moll WM. 2005. Effect of B7-2 and CD40 signals from activated antigen-presenting cells on the ability of zwitterionic polysaccharides to induce T-cell stimulation. *Infect Immun.* 73:2184–2189.

Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC. 1994. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature.* 368:215–221.

Stingele F, Corthesy B, Kusy N, Porcelli SA, Kasper DL, Tzianabos AO. 2004. Zwitterionic polysaccharides stimulate T cells with no preferential Vbeta usage and promote anergy, resulting in protection against experimental abscess formation. *J Immunol.* 172:1483–1490.

Stratikos E, Mosyak L, Zaller DM, Wiley DC. 2002. Identification of the lateral interaction surfaces of human histocompatibility leukocyte antigen (HLA)-DM with HLA-DR1 by formation of tethered complexes that present enhanced HLA-DM catalysis. *J Exp Med.* 196:173–183.

Tzianabos AO, Finberg RW, Wang Y, Chan M, Onderdonk AB, Jennings HJ, Kasper DL. 2000. T cells activated by zwitterionic molecules prevent abscesses induced by pathogenic bacteria. *J Biol Chem.* 275:6733–6740.

Tzianabos AO, Onderdonk AB, Rosner B, Cisneros RL, Kasper DL. 1993. Structural features of polysaccharides that induce intra-abdominal abscesses. *Science.* 262:416–419.

Tzianabos AO, Pantosti A, Baumann H, Brisson JR, Jennings HJ, Kasper DL. 1992. The capsular polysaccharide of *Bacteroides fragilis* comprises two ionically linked polysaccharides. *J Biol Chem.* 267:18230–18235.

Tzianabos AO, Russell PR, Onderdonk AB, Gibson FC III, Cywes C, Chan M, Finberg RW, Kasper DL. 1999. IL-2 mediates protection against abscess formation in an experimental model of sepsis. *J Immunol.* 163:893–897.

Tzianabos AO, Wang JY, Lee JC. 2001. Structural rationale for the modulation of abscess formation by *Staphylococcus aureus* capsular polysaccharides. *Proc Natl Acad Sci USA.* 98:9365–9370.

Wang Y, Kalka-Moll WM, Roehrl MH, Kasper DL. 2000. Structural basis of the abscess-modulating polysaccharide A2 from *Bacteroides fragilis*. *Proc Natl Acad Sci USA.* 97:13478–13483.

Watts C. 1997. Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu Rev Immunol.* 15:821–850.

Watts C, Powis S. 1999. Pathways of antigen processing and presentation. *Rev Immunogenet.* 1:60–74.

Weber DA, Evavold BD, Jensen PE. 1996. Enhanced dissociation of HLA-DR-bound peptides in the presence of HLA-DM. *Science.* 274:618–620.

Downloaded from glycob.oxfordjournals.org at Northwestern University Library, Serials Department on July 12, 2011

B lymphocytes

B lymphocytes are probably the most well defined cells. They have numerous developmental stages specified by various surface molecules, designated as early or late, pre-lymphocytes (lymphoblasts) or lymphocytes, etc. (Santos PM, Borghesi L (2011) Molecular resolution of the B cell landscape. Cur Opin Immunol., 23, 163-170). The developmental stages are closely related to the step-wise synthesis of various classes of antibodies to be discussed in detail later. Roughly, they are:

Pre-pro-B cells

Early pro-B cells

Late Pro-B cells

Large Pre-B cells

Small Pre-B cells

Immature B cells

Mature B cells

At the stage of immature B cells, membrane bond antibody molecules will be synthesized and appear on the cell surface of these cells. The terminal stage of B lymphocytes is known as plasma cells with the major function of secreting a specific antibody. In addition, there are several other types:

Memory B cells

B1 cells

B2 cells

Marginal zone B cells

Follicular B cells

Humans as well as animals require numerous different antibodies, probably of the order of over one million, for protection against foreign invaders. Thus when a specific foreign antigen is introduced into the system, one of these many different B lymphocytes with surface antibodies may be able to bind the foreign antigen reasonably tightly. This antigen-antibody complex will be internalized by the B lymphocyte. In a special compartment of the cell cytoplasm, the protein antigen will be degraded into various peptides, and carbohydrate antigen also into different fragments mimicking the peptides.

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Immunology

Molecular resolution of the B cell landscape

Patricia M Santos and Lisa Borghesi

The progression of hematopoietic stem cells (HSCs) to the B lymphocyte lineage requires that uncommitted progenitors successfully negotiate the transition from multipotency to unipotency, including the loss of self-renewal potential. Previous work identified essential transcription factors that mediate B lineage development. Major advances build on this knowledge and reveal coordinated changes in gene expression occurring within single cells at sequential stages in the B cell differentiation pathway. Recent studies on epigenetic mechanisms also provide a framework within which transcription factor activity, chromatin modifications, and gene expression patterns can be viewed at hierarchical levels to link genotype and phenotype.

Address
Department of Immunology, University of Pittsburgh School of Medicine, 200 Lothrop Street, Pittsburgh, PA 15261, United States

Corresponding author: Borghesi, Lisa (borghesi@pitt.edu)

Current Opinion in Immunology 2011, 23:163–170

This review comes from a themed issue on
Lymphocyte development
Edited by Kris Hogquist and Eugene Oltz

Available online 12th January 2011

0952-7915/$ – see front matter
© 2010 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.coi.2010.11.014

Introduction

Hematopoiesis is a key model system for understanding the mechanisms that control tissue regeneration, developmental plasticity, and lineage fate decisions. This is a scientifically exciting area because the basic processes that occur during hematopoiesis are also fundamental to embryogenesis and tissue repair/regeneration. As such knowledge about the molecular mechanisms that control hematopoiesis have broad biological implications.

Differentiation of HSCs to B lymphocytes involves progression through multipotent progenitor (MPP), lymphoid/myeloid-primed multipotent progenitor (LMPP) and common lymphoid progenitor (CLP) stages of development, accompanied by the sequential loss of megakaryocyte/erythroid (MEP) and myeloid potentials [1,2] (Figure 1). One powerful characteristic of the hematopoietic model system is the ability to isolate, based on surface phenotype, intermediates that represent developmental stage-specific progenitors. Using defined surface markers, studies over the past ten years have progressed from isolation of phenotypic subsets to isolation of single cells. A core transcriptional network that orchestrates B cell fate specification within sequential developmental contexts is established. The transcription factors Ikaros, PU.1 and E2A regulate lymphoid versus myeloid fate choice while EBF and Pax5 control B cell specification and commitment (Figure 2). Recent progress connects genetic and epigenetic mechanisms during specific developmental transitions between lymphoid progenitors and B cell precursors.

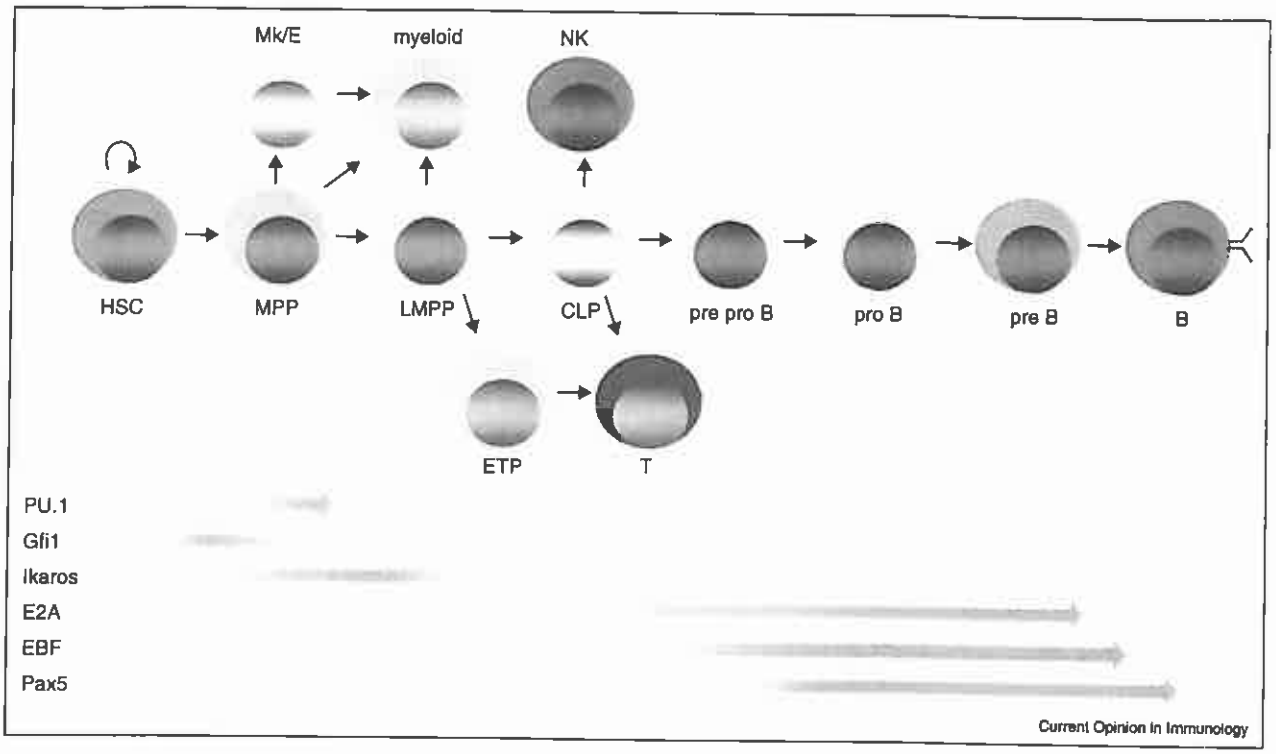
Here, we review current knowledge of transcription factors and their cis-targets crucial for B lineage development. We integrate this transcriptional network within a broader context of recent genome-wide analyses that reveal the patterns of transcription factor binding at sequential developmental stages. Emerging from these studies is a clearer understanding of how the concerted binding of multiple transcription factors (trans-acting factors) at defined regulatory regions (cis-acting elements) regulates gene expression in a developmental stage-specific manner (phenotype). In other words, the gap between genotype and phenotype just got a little smaller.

Transcriptional regulation of B cell development

Progression of HSCs to the B lineage is marked by the loss of pluripotent self-renewal activity and activation of a B cell specific program. During this transition, characteristic low-level gene expression associated with multi-lineage priming gives way to increased expression of genes associated with the B cell fate [3,4]. Genes expressed during multilineage priming commonly exhibit bivalent chromatin marks in which both activating and inhibitory modifications to histones and DNA are present [5,6]. Within $CD150^+$ LSK (lineage marker-negative, $Sca\text{-}1^+$, $c\text{-}kit^+$) stem cells, for example, ~4% of *ebf* and *pax5* promoters bear coincident activating H3K4me3 and inhibitory H3K27me3 modifications [7••]. These bivalent histone marks are subsequently resolved concomitant with changes in the activity of transcription factors that upregulate lymphoid lineage genes and repress genes associated with alternative lineage fates.

The Ikaros transcription factor is a major regulator of HSC progression to the lymphoid lineages. Encoded by the *Ikzf1* gene [8], Ikaros suppresses stem cell associated genes including the receptor tyrosine kinases *tie1*, *tie2*, and *mpl*, and induces lymphoid-specific genes including the *Dntt* nucleotide transferase [9••]. In a powerful experiment, the lymphoid induction potential of Ikaros was shown within single HSCs, the functional level at which lineage fate

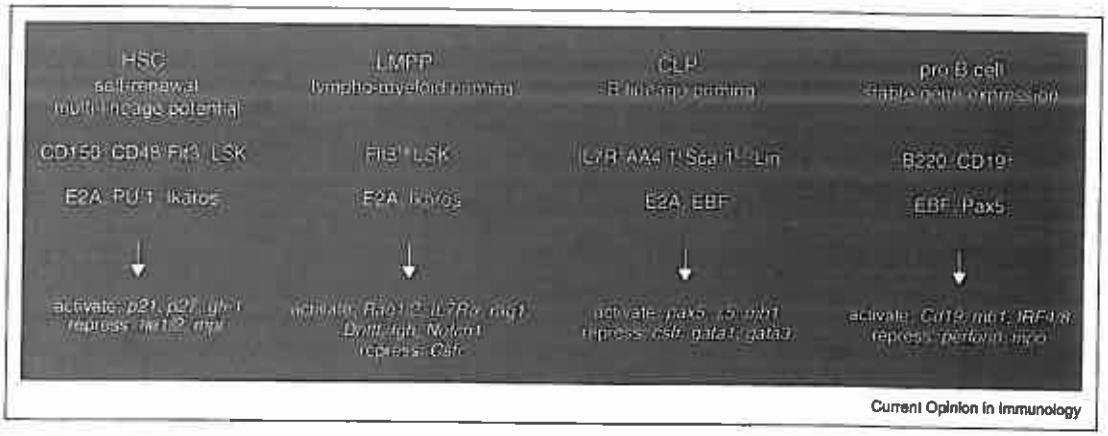
Figure 1



The development of B cells from HSCs involves progression to multipotent progenitor (MPP), lymphoid/myeloid-primed multipotent progenitor, common lymphoid progenitor, (CLP) and B cell progenitor intermediates before B cell commitment. In one current model of developmental progression, establishment of the B lineage fate is accompanied by sequential loss of megakaryocyte/erythrocyte (Mk/E) potential, myeloid potential and T/NK cell potentials.

decisions are made [9••]. Ikaros activity is required in LMPPs and early B cell precursors where it regulates expression of the cytokine receptor *flt3*, the λ5 pre-B cell receptor chain [10] and *rag1/2* genes. Ikaros also regulates immunoglobulin chromatin remodeling [11•]. One crucial function of Ikaros is to antagonize the transcription factor PU.1 (purine box factor-1) to direct the lymphoid versus myeloid fate [12].

Figure 2



Hallmark developmental transitions during HSC progression to the B cell fate. Indicated below each major phenotypic stage is the defining functional criteria, surface markers, key transcription factors, and associated target genes. Emphasis is placed on E2A to reflect recent scientific advances.

17

PU.1, an Ets family transcription factor, regulates gene targets important to both the lymphoid and myeloid lineages including receptors for the cytokines IL-7 and the granulocyte/macrophage (G/M) colony stimulating factors, respectively. PU.1 expression in MPPs restricts MEP fate [13] after which the coordinated interaction of PU.1 and the Ikaros-induced Gfi-1 (growth factor independent-1) transcription factor establishes B versus myeloid fate choice by stabilizing PU.1 levels [12]. PU.1 is maintained at low levels in B cells and at high levels in myeloid cells [14,15]. Engagement of PU.1 motifs within the *Sfp1* promoter itself can promote autoamplification in a feed forward loop [16,17]. During lymphopoiesis however, Ikaros induces Gfi-1, which, in turn, blocks PU.1 autoamplification through physical displacement of PU.1 at the *Sfp1* promoter [11•]. The biological effect is to reduce PU.1 expression to lymphoid-appropriate levels. In addition to limiting PU.1, Gfi-1 reinforces lymphoid progression by enhancing E2A activity through direct suppression of the inhibitor of DNA binding factor-2 (Id2), an E2A antagonist [18]. Finally, the ability of PU.1 to direct major downstream factors E2A, EBF, Oct-2 and NFκB to defined target genes in a lymphoid lineage-specific manner is one likely mechanism that favors lymphopoiesis versus myelopoiesis [19].

E2A, a basic helix–loop–helix transcription factor, is emerging as an important regulator of HSC self-renewal and of LMPP – but not HSC – lineage restriction. E2A deficient HSCs exhibit diminished self-renewal capability following adoptive transfer, hyperproliferation associated with loss of the *p21* cell cycle inhibitor, and a failure to prime lymphoid lineage-associated genes [20••,21,22••,23]. Loss of HSC activity is also observed in mice lacking either the E2A inhibitor Id1 [24,25] or the E2A interaction partner Stem Cell Leukemia (SCL) [26], suggesting an interacting network of E proteins at this stage. E2A is dispensable for the restriction of MEP and myeloid lineage potentials as assessed using single HSCs [20••,22••]. E2A may regulate later dynamics as erythroid and megakaryocyte intermediates are diminished [21]. MPPs and LMPPs are numerically reduced in the absence of E2A [20••,23], and downstream CLPs are virtually ablated [27]. *In vivo* administration of antioxidant restores MPPs but not the LMPP or CLP subsets in E2A deficient mice [22••], thereby defining new activities of E2A independent of lymphoid lineage differentiation.

E2A is an essential regulator of lymphoid specification and restriction of myeloid potential at the LMPP stage [20••]. Gain-of-function and loss-of-function studies show that the balance between E proteins and Id proteins regulates lymphoid versus myeloid fate decisions of LMPPs [28]. In addition to being reduced in number, remnant LMPPs from E2A-deficient mice fail to fully upregulate the Flt3 cytokine receptor [20••,22••] that is essential for development of $Flt3^{hi}$VCAM-1^{-}LMPPs [29]. E2A does not regulate *flt3* transcription *per se* [20••]. Rather, Hoxa9 and PU.1 regulate *flt3* expression, although cooperative interactions between E2A, hoxa9, Ikaros, and PU.1 in LMPPs remain to be established [30]. E2A-deficient LMPPs additionally lack V(D)J recombinase activity [27]. At the molecular level, E2A-deficient LMPPs have decreased expression of the lymphoid associated genes *IL7Rα*, *rag1*, *Dntt*, *Igh-6*, *notch1*, and *ccr9* [20••], and fail to generate pro-B cells [31]. The E2A inhibitor, HEBalt, an alternative splice product of the HEB locus, emerges as a context-dependent regulator of cell fate choice. HEBalt inhibits B cell outgrowth to the benefit of myeloid cells, but in the presence of Notch ligands, favors T cells over myeloid cells [32].

In summary, three major roles for E2A are emerging. First, E2A appears to be required for efficient HSC self-renewal and persistence, and for lymphoid-lineage priming at this stage. Second, E2A promotes the production and/or maintenance of MPPs and LMPPs, and is required for myeloid restriction in LMPPs. Third, as detailed immediately below, E2A is required for promotion of B lineage progression through a cascade involving the EBF, Pax5, and Foxo1 regulatory factors.

EBF and Pax5 sequentially specify and commit lymphoid precursors to the B cell fate. Forced expression of EBF in MPPs activates the lineage appropriate genes *pax5*, *lambda5*, *VpreB*, and *Cd79b* and represses genes associated with alternative fates including *cJEBPα* [33] (Figure 2). Binding of the interferon-regulatory factor IRF8 to the *ebf* promoter results in transcriptional activation of EBF while binding to the *Sfp1* promoter represses PU.1 [34]. EBF, in turn, reinforces E2A activity by repressing the Id inhibitors of E2A [35]. Unlike E2A-deficient mice that lack CLPs, EBF-deficient mice have a numerically replete CLP compartment that lacks B lineage potential. At the single cell level, EBF KO CLPs fail to upregulate *pax5*, *Pou2af1*, and *mb-1* transcripts [36]. Single cell tracing reveals that high levels of *rag1* accompany loss of myeloid and NK potential [37••,38•]. Almost all rag^{hi} CLPs express EBF, with ~31% of these cells also expressing *pax5* and retaining both B and T cell potential *in vitro*. T cell potential is lost concomitant with full transcriptional activity of EBF. *Rag* expression in CLPs correlates with Ly6D expression suggesting the value of this surface marker for distinguishing CLPs with global lymphoid potential versus CLPs that are B lineage biased [37••,39•]. Compound haploinsufficiency of EBF and RUNX1 diminishes B lineage specific gene expression resulting in a block in B cell progression [40].

Pax5 maintains the B cell fate [41••,42]. Conditional ablation of *pax5* in $CD19^{+}$ splenocytes leads to a loss of the B cell program and a striking conversion to the T

cell fate. In support of these observations, Notch1 signaling in response to delta like ligand 4 leads to a dose dependent inhibition of Pax5 and the B lineage program in the thymus [43]. In humans, ablation of the E2A-EBF-Pax5 pathway is associated with Hodgkin lymphoma in which B cells exhibit a loss of identity uncannily reminiscent of *pax5* conditional deletion [42].

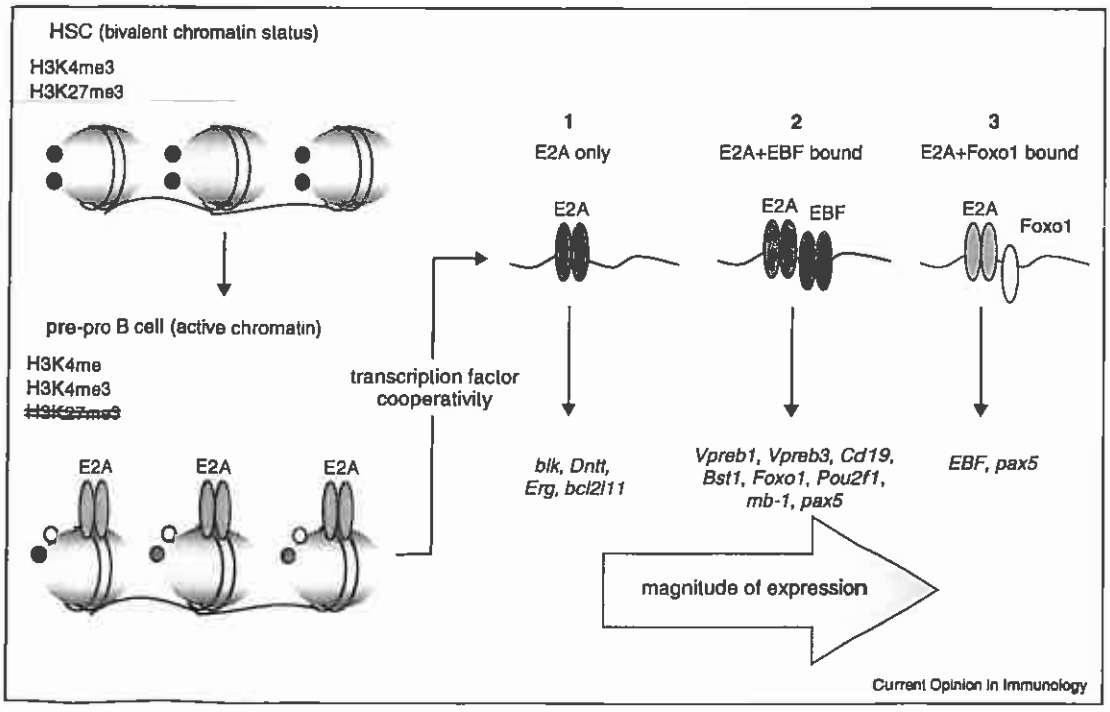
Together, these studies provide new insights into activities of major transcription factors at the single cell level. Ikaros induces lympho-myeloid priming as early as HSCs while E2A mediates myeloid restriction in LMPPs. EBF induces B lineage specific gene expression, and Pax 5 promotes B lineage restriction. These findings raise questions about the specific epigenetic mechanisms that act in concert with transcription factor activity to promote B cell development.

Epigenetic regulation of the B cell development

Cell fate decisions are dictated by the activation state of core transcription factors and the access of these factors to select genomic regions. The epigenetic mechanisms that regulate B lineage determination are being revealed at the sequential levels of genome wide analysis (E2A, EBF), a promoter/enhancer complex (*pax5*), and B cell-specific genes (*mb-1*, *Cd19*).

Murre *et al.* detail the association of E2A binding to cis elements, epigenetic modification, coordinated occupancy by cooperating factors, and transcript abundance at sequential stages of B cell development [44••]. There are two thematic advances in this study. First, the data provide a molecular genetic snapshot of B cells at a developmental transition where they still possess a genotype identical with most other cells in the body yet have a distinct phenotype. Second, this study quantifies the positive effects on gene expression of cooperative transcription factor binding. E2A occupancy near transcription start sites is associated with activating H3K4 monomethylation marks at enhancers and H3K4 trimethylation marks at promoters (Figure 3). About half of the E2A-occupied sites in pre-pro B cells are also occupied in pro-B cells, but pro-B cells acquire ~10,000 new E2A-bound sites. E2A-occupied sites co-localize with binding motifs of factors essential to the B lineage fate including EBF and PU.1 [44••], the Foxo1 regulator of *rag* expression [45,46], the CTCF regulator of immunoglobulin locus contraction [47] and Bcl11a [48]. Co-occupancy by E2A and EBF correlates with H3K4

Figure 3



Epigenetic modification and transcription. Bivalent chromatin modifications mark B lineage-specific genes in HSCs. Such genes are considered poised but silent. E2A binding to cis elements is associated with activating H3K4 monomethylation at enhancers and H3K4 trimethylation at promoters. E2A activates EBF and Pax5, initiating a cascade of expression of B cell-specific genes. Transcription factor cooperation elevates the magnitude of gene expression.

methylation status and an abundance of the hallmark B lineage transcripts *VpreB*, *Cd19*, *foxo1*, *Pou2f1*, *Cd79a*, and *pax5*. These data define a molecular genetic basis for the functional synergy observed between E2A and EBF [49]. Loci enriched for co-occupancy of E2A and Foxo1 include *Dntt*, the *Erg* transcription factor, and the pro-apoptotic gene *bcl2l11* that encodes Bim. Mice doubly heterozygous for E2A and Foxo1 ($e2A^{+/-}foxo1^{+/-}$) have reduced B cell progenitors compared to either single heterozygote, thereby directly linking E2A, EBF, and Foxo1 in a common pathway [44••]. In another example of lineage-specific cooperativity, the ability of PU.1 to bind to defined targets is influenced by the presence of cooperating factors as enforced expression of E2A increased PU.1 binding sites by 4-fold [19].

Genome-wide analyses reveal EBF-activated and EBF-repressed targets across defined stages of B cell development [50••]. EBF-activated targets (*Cd79a*, *Gfra2*, and *pax5*) gain the activation marks H3K9me and H3ac while EBF-repressed genes acquire repressive H3K27me3 marks. A subset of developmentally regulated genes (e.g. *Cd40*) exhibits the H3K4me2 poised mark at the pro B cell stage followed by the activating H3K4me3 and H3ac at the mature stage, suggesting that EBF primes chromatin before transcriptional activation. Enforced expression of EBF in a $CD4^+CD8^+$ T cell line or in NIH3T3 cells reveals EBF capabilities for chromatin poising. In the T cell line, EBF binding and H3K4me2 modification is present at *Cd79a*, *pax5*, and *Cd40* loci, with low-level transcription detectable from the *Cd40* locus. In NIH3T3 cells, neither EBF binding nor activating histone marks are observed. Thus, EBF recognizes binding sites independent of transcriptional activation but not in closed chromatin [50••].

Busslinger *et al.* define the stepwise activation of the Pax5 locus. Pax5 expression is dictated by an enhancer in intron 5 of the *pax5* locus and a promoter element [51••]. CpG motifs in the enhancer are demethylated during the HSC to MPP transition, and acquire the active chromatin mark $H3K9ac^+H3K27me3^-$ by the pro B cell stage. The enhancer is positively activated by PU.1, IRF-4 and IRF-8, and NFκB. By contrast, the promoter is demethylated even in HSCs. However, in the absence of E2A or EBF, the promoter fails to acquire the H3K9ac activating histone marks. These findings highlight a crucial role for EBF in activation of the *pax5* promoter.

Pax5 collaborates with other factors to remodel B lineage-specific targets *mb-1* and *Cd19*. The *mb-1* promoter undergoes demethylation during the HSC to pro B transition [52]. Demethylation requires the activity of both EBF and Pax5 as enforced expression of EBF in $ebf^{-/-}pax5^{-/-}$ progenitors is not sufficient to mediate demethylation. Efficient transcription of *mb-1* requires SWI/SNF activity and is suppressed by Mi-2NuRD chromatin remodeling complexes [52]. The CD19 locus undergoes stepwise activation of the enhancer and promoter [53]. E2A binding sites within the *Cd19* enhancer are already demethylated as early as LSKs, suggesting that CD19 undergoes lineage priming well before the stage of lineage-specification. Progressive demethylation correlates with the successive binding of E2A, EBF, and Pax5 in a developmental stage specific manner. CD19 transcriptional activation is achieved only after Pax5 binding to the promoter, an interaction accompanied by H3K4 trimethylation. Both studies detail the mechanistic importance of transcription factor cooperativity in lineage priming and the stable expression of B cell-specific genes, echoing a major theme of this Review.

Summary and perspective

The ability to link coordinated changes in gene expression to developmental potential is essential for a clear vision of the mechanistic factors that drive lineage fate choice. Research progress over the past ten years has defined a core network of transcription factors that activate the B cell fate and repress alternative fates. We now add to this foundation an ability to discern the epigenomic marks that influence gene expression. The studies highlighted here describe epigenetic changes associated with HSC progression to the B cell fate at the hierarchical levels of genome-wide analysis, B lineage-specific transcription factor activation, and B cell locus-specific gene expression. On the horizon is the opportunity to evaluate chromatin poising and gene expression within single cells, the biological level at which lineage fate decisions are made. The most significant advances are likely to come from progenitors poised at crucial developmental transitions. Observations that lineage priming within HSCs is associated with bivalent histone states raises major questions. On a per cell basis, do individual HSCs bearing identical bivalent marks have comparable potential for lineage progression? Or, are some HSCs refractory to lineage progression due to other limiting co-modifications? HSCs with heterogeneous development potential appear to co-exist in the bone marrow, with some stem cells expressing lymphoid bias, others expressing myeloid bias, and still others possessing a latent potential that manifests in robust repopulation activity only after secondary transplantation [54]. Future studies will reveal how chromatin status and gene expression patterns of individual HSCs – or cells representing later development stages – track with lineage reconstitution potential and progression to the B cell fate.

Acknowledgements

Supported by NIH R01 AI079047 with seed perspective developed under an NSF/Alfred P. Sloan Foundation Fellowship in Molecular Evolution (LB). We deeply appreciate direct input from K Murre, M Sigvardsson, and K Medina.

References and recommended reading

Papers of particular interest published within the period of review have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Adolfsson J, Mansson R, Buza-Vidas N, Hultquist A, Liuba K, Jensen CT, Bryder D, Yang L, Borge OJ, Thoren LA *et al.*: **Identification of Flt3 + lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment.** *Cell* 2005, **121**:295-306.

2. Forsberg EC, Serwold T, Kogan S, Weissman IL, Passegue E: **New evidence supporting megakaryocyte-erythrocyte potential of flk2/flt3+ multipotent hematopoietic progenitors.** *Cell* 2006, **126**:415-426.

3. Hu M, Krause D, Greaves M, Sharkis S, Dexter M, Heyworth C, Enver T: **Multilineage gene expression precedes commitment in the hemopoietic system.** *Genes Dev* 1997, **11**:774-785.

4. Mansson R, Hultquist A, Luc S, Yang L, Anderson K, Kharazi S, Al-Hashmi S, Liuba K, Thoren L, Adolfsson J *et al.*: **Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors.** *Immunity* 2007, **26**:407-419.

5. Spivakov M, Fisher AG: **Epigenetic signatures of stem-cell identity.** *Nat Rev Genet* 2007, **8**:263-271.

6. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K *et al.*: **A bivalent chromatin structure marks key developmental genes in embryonic stem cells.** *Cell* 2006, **125**:315-326.

7. •• Welshaupt H, Sigvardsson M, Attema JL: **Epigenetic chromatin states uniquely define the developmental plasticity of murine hematopoietic stem cells.** *Blood* 2010, **115**:247-256.

Mini-ChIP analysis of HSCs ($CD150^+$ LSKs) reveals bivalent histone modification status of lineage-restricted promoters. ebf and pax5 have bivalent marks while Ikaros and pu.1 have only activating marks. Chromatin marks correlate with gene expression levels. Some promoters in splenic T cells bear bivalent marks raising the possibility of re-expression under select conditions (e.g. ccr9, normally expressed only in thymocytes). These data suggest one mechanism underlying developmental plasticity in 'committed' cells (see reference [41]).

8. Georgopoulos K, Bigby M, Wang JH, Molnar A, Wu P, Winandy S, Sharpe A: **The Ikaros gene is required for the development of all lymphoid lineages.** *Cell* 1994, **79**:143-156.

9. •• Ng SY, Yoshida T, Zhang J, Georgopoulos K: **Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells.** *Immunity* 2009, **30**:493-507.

Demonstrates requirement of Ikaros for repressing stem cell associated genes and inducing lympho-myeloid priming in HSCs ($Flt3^-$ LSKs). In contrast to previous models in which lymphoid priming was thought to be a relatively late event, HSC transcriptional signatures reveal lymphoid priming. Importantly, transcriptome analysis validated at the single cell level shows comparable frequency of lymphoid priming and erythroid priming, with a small proportion of HSCs exhibiting lymphoid/erythroid/myeloid co-priming.

10. Thompson EC, Cobb BS, Sabbattini P, Meixlsperger S, Parelho V, Liberg D, Taylor B, Dillon N, Georgopoulos K, Jumaa H *et al.*: **Ikaros DNA-binding proteins as integral components of B cell developmental-stage-specific regulatory circuits.** *Immunity* 2007, **26**:335-344.

11. • Reynaud D, Demarco IA, Reddy KL, Schjerven H, Bertolino E, Chen Z, Smale ST, Winandy S, Singh H: **Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros.** *Nat Immunol* 2008, **9**:927-936.

Defines B lineage specific roles of Ikaros. While forced expression of EBF in Ikaros-deficient MPPs produces $CD19^+$, IL-7 responsive cells, these 'pro-B' cells fail to undergo V(D)J recombination and still retain myeloid potential. Forced expression of Ikaros itself is required to induce rag expression, V_H region accessibility, and IgH chromatin compaction.

12. Spooner CJ, Cheng JX, Pujadas E, Laslo P, Singh H: **A recurrent network involving the transcription factors PU.1 and Gfi1 orchestrates innate and adaptive immune cell fates.** *Immunity* 2009, **31**:576-586.

13. Arinobu Y, Mizuno S, Chong Y, Shigematsu H, Iino T, Iwasaki H, Graf T, Mayfield R, Chan S, Kastner P *et al.*: **Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages.** *Cell Stem Cell* 2007, **1**:416-427.

14. DeKoter RP, Singh H: **Regulation of B lymphocyte and macrophage development by graded expression of PU.1.** *Science* 2000, **288**:1439-1441.

15. Nutt SL, Metcalf D, D'Amico A, Polli M, Wu L: **Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors.** *J Exp Med* 2005, **201**:221-231.

16. DeKoter RP, Lee HJ, Singh H: **PU.1 regulates expression of the interleukin-7 receptor in lymphoid progenitors.** *Immunity* 2002, **16**:297-309.

17. DeKoter RP, Walsh JC, Singh H: **PU.1 regulates both cytokine-dependent proliferation and differentiation of granulocyte/macrophage progenitors.** *EMBO J* 1998, **17**:4456-4468.

18. Li H, Ji M, Klarmann KD, Keller JR: **Repression of Id2 expression by Gfi-1 is required for B-cell and myeloid development.** *Blood* 2010, **116**:1060-1069.

19. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK: **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.** *Mol Cell* 2010, **38**:576-589.

20. •• Dias S, Mansson R, Gurbuxani S, Sigvardsson M, Kee BL: **E2A proteins promote development of lymphoid-primed multipotent progenitors.** *Immunity* 2008, **29**:217-227.

E2A is required for lymphoid lineage priming. E2A deficient mice have a paucity of $Flt3^{hi}$ LMPPs that fail to express lymphoid appropriate genes. This defect extends to HSCs, indicating a role for E2A in lymphoid priming even in stem cells ($Flt3^-$ LSKs). E2A is dispensable for megakaryocyte restriction but antagonizes LMPP differentiation to the myeloid fate.

21. Semerad CL, Mercer EM, Inlay MA, Weissman IL, Murre C: **E2A proteins maintain the hematopoietic stem cell pool and promote the maturation of myelolymphoid and myeloerythroid progenitors.** *Proc Natl Acad Sci USA* 2009, **106**:1930-1935.

E2A is required for HSC self-renewal. HSCs ($CD150^+Flt3^-$ LSKs) lacking E2A fail to sustain hematopoiesis following serial adoptive transfer. E2A is mechanistically linked to the known regulators of HSC competence p21, p27, and mpl.

22. •• Yang Q, Esplin BL, Borghesi L: **E47 regulates hematopoietic stem cell proliferation and energetics but not myeloid lineage restriction.** *Blood*, 10/27/10, in press.

E47 is dispensable for short-term myeloid production from HSCs. E47-deficient HSCs ($CD150^+CD48^-$ LSKs or $Flt3^-$ LSKs) within wild type hosts exhibit hyperproliferation and loss of p21. *In vivo* quantitative studies demonstrate that while long-term self-renewal and lymphoid repopulation activities are compromised, myeloid restriction remains intact. *In vivo* treatment with antioxidant restores numbers of MPPs in E47 knockout mice.

23. Yang Q, Kardava L, St Leger A, Martinelc K, Varnum-Finney B, Bernstein ID, Milcarek C, Borghesi L: **E47 controls the developmental integrity and cell cycle quiescence of multipotential hematopoietic progenitors.** *J Immunol* 2008, **181**:5885-5894.

E47 is required for HSC proliferative integrity. HSCs (defined as $CD150^+CD48^-$ LSKs, $CD27^-$ LSKs or $Flt3^-$ LSKs) lacking E47 exhibit hyperproliferation, and fail to produce a robust MPP compartment. p21 is identified as a direct E47 target gene.

24. Perry SS, Zhao Y, Nie L, Cochrane SW, Huang Z, Sun XH: **Id1, but not Id3, directs long-term repopulating hematopoietic stem-cell maintenance.** *Blood* 2007, **110**:2351-2360.

25. Suh HC, Ji M, Gooya J, Lee M, Klarmann KD, Keller JR: **Cell-nonautonomous function of Id1 in the hematopoietic progenitor cell niche.** *Blood* 2009, **114**:1186-1195.

26. Lacombe J, Herblot S, Rojas-Sutterlin S, Haman A, Barakat S, Iscove NN, Sauvageau G, Hoang T: **Scl regulates the quiescence and the long-term competence of hematopoietic stem cells.** *Blood* 2010, **115**:792-803.

21

27. Borghesi L, Aites J, Nelson S, Lefterov P, James P, Gerstein R: **E47 is required for V(D)J recombinase activity in common lymphoid progenitors.** *J Exp Med* 2005, 202:1669-1677.

28. Cochrane SW, Zhao Y, Welner RS, Sun XH: **Balance between Id and E proteins regulates myeloid-versus-lymphoid lineage decisions.** *Blood* 2009, 113:1016-1026.
Targeted manipulation of E protein or Id protein activity directs lineage fate choice in LMPPs. Id1 is expressed in HSCs, is turned off during MEP and CMP restriction, and is re-expressed in myeloid cells. Forced expression of Id1 in LSKs favors myeloid over lymphoid development. Expression of a cre inducible dominant negative inhibitor of Id activity has the reverse effect with LSKs and LMPPs showing robust lymphoid over myeloid production.

29. Lal AY, Watanabe A, O'Brien T, Kondo M: **Pertussis toxin-sensitive G proteins regulate lymphoid lineage specification in multipotent hematopoietic progenitors.** *Blood* 2009, 113:5757-5764.

30. Gwin K, Frank E, Bossou A, Medina KL: **Hoxa9 regulates Flt3 in lymphohematopoietic progenitors.** *J Immunol* 2010, 185:6572-6583.

31. Zhuang Y, Soriano P, Weintraub H: **The helix-loop-helix gene E2A is required for B cell formation.** *Cell* 1994, 79:875-884.

32. Wang D, Claus CL, Rajkumar P, Braunstein M, Moore AJ, Sigvardsson M, Anderson MK: **Context-dependent regulation of hematopoietic lineage choice by HEBAlt.** *J Immunol* 2010, 185:4109-4117.
Defines biological role for Alt (alternative) splice form of HEB. Forced expression of HEBalt in uncommitted fetal liver progenitors blocks E2A activity thereby favoring myeloid over B cell outgrowth. In combination with Notch signals, HEBalt favors T over myeloid outgrowth. HEBcan (canonical) does not suppress myeloid potential. These data demonstrate a unique function of the Alt domain as a context-dependent regulator of lineage fate choice.

33. Pongubala JM, Northrup DL, Lancki DW, Medina KL, Treiber T, Bertolino E, Thomas M, Grosschedl R, Allman D, Singh H: **Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5.** *Nat Immunol* 2008, 9:203-215.

34. Wang H, Lee CH, Qi C, Tailor P, Feng J, Abbasi S, Atsumi T, Morse HC 3rd: **IRF8 regulates B-cell lineage specification, commitment, and differentiation.** *Blood* 2008, 112:4028-4038.

35. Thal MA, Carvalho TL, He T, Kim HG, Gao H, Hagman J, Klug CA: **Ebf1-mediated down-regulation of Id2 and Id3 is essential for specification of the B cell lineage.** *Proc Natl Acad Sci USA* 2009, 106:552-557.

36. Zandi S, Mansson R, Tsapogas P, Zetterblad J, Bryder D, Sigvardsson M: **EBF1 is essential for B-lineage priming and establishment of a transcription factor network in common lymphoid progenitors.** *J Immunol* 2008, 181:3364-3372.

37. •• Mansson R, Zandi S, Welinder E, Tsapogas P, Sakaguchi N, Bryder D, Sigvardsson M: **Single-cell analysis of the common lymphoid progenitor compartment reveals functional and molecular heterogeneity.** *Blood* 2010, 115:2601-2609.
The CLP compartment is populated by progenitor subsets with constrained lineage potential as opposed to a uniform population of multipotent progenitors. This study tracks lineage fate potential of single CLPs using the dual molecular tracers rag1 (indicator of lymphoid lineage progression) expression and the λ5 light chain gene (reflects Pax5 transcriptional activity). Rag^{lo} cells retain B, T, and NK potential while rag^{hi} cells have B and T potential. Within rag^{hi} CLPs, only the λ5$^+$ subset is B lineage restricted. These data suggest that inhibition of NK potential depends on EBF but not Pax5.

38. • Welner RS, Esplin BL, Garrett KP, Pelayo R, Luche H, Fehling HJ, Kincade PW: **Asynchronous RAG-1 expression during B lymphopoiesis.** *J Immunol* 2009, 183:7768-7777.
Fate mapping study of Rag1$^+$ progenitors using a permanent red fluorescence tagging. Rag1/Red$^+$ CLPs progress more quickly to the CD19$^+$ stage as compared to their Rag1/Red$^-$ counterparts. Following exposure to TLR ligands, Rag1/Red$^+$ CLPs are lineage stable and efficiently give rise to CD19$^+$ cells while Rag1/Red$^-$ CLPs generate DC and NK cells with a near complete inhibition of B cell potential. Thus, bone marrow progenitors become increasingly committed to the B lineage concordant with Rag expression.

39. • Inlay MA, Bhattacharya D, Sahoo D, Serwold T, Seita J, Karsunky H, Plevritis SK, Dill DL, Weissman IL: **Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development.** *Genes Dev* 2009, 23:2376-2381.
Uses the algorithm MidReg to identify Ly6D as a surface marker that phenotypically resolves CLPs with lymphoid potential versus CLPs that are B lymphoid progenitors. Ly6D$^-$ CLPs have low IL7Rα expression and retain T, NK, and DC potential that is lost upon Ly6D expression.

40. Lukin K, Fields S, Lopez D, Cherrier M, Ternyak K, Ramirez J, Feeney AJ, Hagman J: **Compound haploinsufficiencies of Ebf1 and Runx1 genes impede B cell lineage progression.** *Proc Natl Acad Sci USA* 2010, 107:7869-7874.

41. •• Cobaleda C, Jochum W, Busslinger M: **Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors.** *Nature* 2007, 449:473-477.
Landmark publication demonstrating conversion of splenic B cells into functional T cells by conditional deletion of a single factor. Targeted ablation of pax5 in splenic B cells leads to loss of surface immunoglobulin and re-expression of immature cell surface markers. Following adoptive transfer into T cell deficient mice, these former B cells, permanently marked at the genetic level by historic immunoglobulin gene rearrangements, produce TCR$^+$ T cells that provide competent T cell help following antigen challenge.

42. Mathas S, Janz M, Hummel F, Hummel M, Wollert-Wulf B, Lusatis S, Anagnostopoulos I, Lietz A, Sigvardsson M, Jundt F *et al.*: **Intrinsic inhibition of transcription factor E2A by HLH proteins ABF-1 and Id2 mediates reprogramming of neoplastic B cells in Hodgkin lymphoma.** *Nat Immunol* 2006, 7:207-215.

43. Mohtashami M, Shah DK, Nakase H, Kianizad K, Petrie HT, Zuniga-Pflucker JC: **Direct comparison of Dll1- and Dll4-mediated Notch activation levels shows differential lymphomyeloid lineage commitment outcomes.** *J Immunol* 2010, 185:867-876.

44. •• Lin YC, Jhunjhunwala S, Benner C, Heinz S, Welinder E, Mansson R, Sigvardsson M, Hagman J, Espinoza CA, Dutkowski J *et al.*: **A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate.** *Nat Immunol* 2010, 11:635-643.
Integrates transcriptional regulators, epigenetic modifications, and the expression of lineage specific genes at a defined developmental transition in B cells. E2A occupancy at cis regulatory regions is associated with activating H3K4 modifications. The pre-pro-B to pro-B transition is marked by differential patterns of E2A binding and H3K4 methylation including the acquisition of appreciable new binding sites in pro-B cells. E2A sites are enriched for cis regulatory sequences associated with candidate cooperative factors, and coordinate occupancy of cis elements by E2A, EBF and Foxo1 correlates with quantitative increases in expression of B lymphoid genes.

45. Dengler HS, Baracho GV, Omori SA, Bruckner S, Arden KC, Castrillon DH, DePinho RA, Rickert RC: **Distinct functions for the transcription factor Foxo1 at various stages of B cell differentiation.** *Nat Immunol* 2008, 9:1388-1398.

46. Amin RH, Schlissel MS: **Foxo1 directly regulates the transcription of recombination-activating genes during B cell development.** *Nat Immunol* 2008, 9:613-622.

47. Degner SC, Wong TP, Jankevicius G, Feeney AJ: **Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development.** *J Immunol* 2009, 182:44-48.

48. Liu P, Keller JR, Ortiz M, Tessarollo L, Rachel RA, Nakamura T, Jenkins NA, Copeland NG: **Bcl11a is essential for normal lymphoid development.** *Nat Immunol* 2003, 4:525-532.

49. O'Riordan M, Grosschedl R: **Coordinate regulation of B cell differentiation by the transcription factors EBF and E2A.** *Immunity* 1999, 11:21-31.

50. •• Treiber T, Mandel EM, Pott S, Gyory I, Firner S, Liu ET, Grosschedl R: **Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription-independent poising of chromatin.** *Immunity* 2010, 32:714-725.
EBF remodels chromatin independent of transcription activation. Genome wide ChIP analysis defines EBF binding sites, many of which overlap with E2A and RUNX. EBF binding correlates with transcript expression,

particularly of genes associated with B cell receptor signaling. Forced expression of EBF in a T cell line leads to EBF binding to genomic regions, H3K4 dimethylation of B cell specific genes, and low level transcriptional activation of CD40. Parallel experiments on NIH3T3 cells demonstrated no substantial binding of EBF. Thus, EBF activates lineage-specific genes in the context of hematopoietic chromatin but not closed chromatin.

51. •• Decker T, Pasca di Magliano M, McManus S, Sun Q, Bonifer C, Tagoh H, Busslinger M: **Stepwise activation of enhancer and promoter regions of the B cell commitment gene Pax5 in early lymphopoiesis.** *Immunity* 2009, **30**:508-520.

Deletion mapping and epigenetic status of pax5 regulatory elements reveals stepwise epigenetic activation. The pax5 enhancer and promoter are differentially controlled by CpG methylation and chromatin marks. The enhancer is subject to developmental stage specific demethylation while the promoter remains unmethylated even in HSCs. Conversely, the enhancer bears activation marks (H3K9ac) in the absence of E2A and EBF while the promoter is repressed ($H3K9ac^{-}H3K27me3^{+}$) under these same circumstances. Functional binding sites for PU.1, IRF4, IRF8, and NFκB further regulate enhancer activity.

52. Gao H, Lukin K, Ramirez J, Fields S, Lopez D, Hagman J: **Opposing effects of SWI/SNF and Mi-2/NuRD chromatin remodeling complexes on epigenetic reprogramming by EBF and Pax5.** *Proc Natl Acad Sci USA* 2009, **106**:11258-11263.

53. Walter K, Bonifer C, Tagoh H: **Stem cell-specific epigenetic priming and B cell-specific transcriptional activation at the mouse Cd19 locus.** *Blood* 2008, **112**:1673-1682.

54. Hock H: **Some hematopoietic stem cells are more equal than others.** *J Exp Med* 2010, **207**:1127-1130.

Monocytes, macrophages, and dendritic cells

Foreign antigens can also be internalized into monocytes, macrophages, and dendritic cells by phagocytosis. This may be a more primitive method of defending against foreign invaders. In the case of monocytes and macrophages, they are critical effectors and regulators of inflammation and the innate immune response. On the other hand, dendritic cells initiate and regulate the pathogen specific adaptive immune response. However, developmentally, these cells are closely related (Geissmann F, Manz MG, Jung S, Sieweke MH, Merad M, Key K (2010) Development of monocytes, macrophages, and dendritic cells. Science, 327, 656-661; and erratum). According to these authors, there is a common monocyte, macrophage, dendritic cell precursor cell which differentiates into common dendritic cell precursor and monocyte precursor both with two subsets: plasmacytoid dendritic cells and pre-classical dendritic cells; and Ly-6C$^+$ and Ly-6C$^-$ monocytes respectively. These cells exist in bone marrow as well as in the blood stream.

In the tissue, the situation becomes very complicated. They specified three different tissues:

Lymphoid tissues:	Plasmacytoid dendritic cells
	Classic dendritic cells from pre-classical dendritic cells
	Macrophages probably from monocyte precursor
Intestinal lamina propria:	CD103$^+$ lpDC from pre-classical dendritic cells
	CX_3CR1^+ lpDC from Ly-6C$^+$ monocytes
	Alveolar macrophages from Ly-6C$^-$ monocytes
Inflamed tissues:	TipDC from Ly-6C$^+$ monocytes
	Inflammatory macrophage from Ly-6C$^+$ monocytes
	Fraction of MDSC from Ly-6C$^+$ monocytes
	Mcrophages from Ly-6C$^-$ monocytes

Lamina propria dendritic cells are abbreviated as lpDC, tumor necrosis factor (TNF) and inducible nitric oxide synthase (iNOS) producing dendritic cells as TipDC, and myeloid derived suppressor cells as MDSC. In addition, Ly-6C$^+$ monocytes may give rise to Langerhans cells and microglia, both of which can regenerate on their own.

As in the case of B lymphocytes, the internalized foreign protein antigen in these monocytes, macrophages, dendritic cells can then be degraded into peptides. There is no need of the existence of some cell receptors to bind the foreign antigen.

CORRECTED 3 DECEMBER 2010; SEE LAST PAGE

Development of Monocytes, Macrophages, and Dendritic Cells

Frederic Geissmann,[1]* Markus G. Manz,[2] Steffen Jung,[3] Michael H. Sieweke,[4] Miriam Merad,[5] Klaus Ley[6]*

Monocytes and macrophages are critical effectors and regulators of inflammation and the innate immune response, the immediate arm of the immune system. Dendritic cells initiate and regulate the highly pathogen-specific adaptive immune responses and are central to the development of immunologic memory and tolerance. Recent in vivo experimental approaches in the mouse have unveiled new aspects of the developmental and lineage relationships among these cell populations. Despite this, the origin and differentiation cues for many tissue macrophages, monocytes, and dendritic cell subsets in mice, and the corresponding cell populations in humans, remain to be elucidated.

White blood cells, or leukocytes, are a diverse group of cell types that mediate the body's immune response. They circulate through the blood and lymphatic system and are recruited to sites of tissue damage and infection. Leukocyte subsets are distinguished by functional and physical characteristics. They have a common origin in hematopoietic stem cells and develop along distinct differentiation pathways in response to internal and external cues. The mononuclear phagocyte system represents a subgroup of leukocytes originally described as a population of bone marrow–derived myeloid cells that circulate in the blood as monocytes and populate tissues as macrophages in the steady state and during inflammation (*1*). In different tissues, they can show substantial heterogeneity with respect to phenotype, homeostatic turnover, and function. The discovery of dendritic cells (DCs) as a distinct lineage of mononuclear phagocytes, specialized in antigen presentation to T cells and the initiation and control of immunity (*2*), revealed additional roles of these cells in shaping the immune response to pathogens, vaccines, and tumors, as well as additional heterogeneity. Whereas a detailed map of the relationship between monocytes and DCs and their progenitors begins to emerge, other areas like the origin and renewal of tissue macrophage subsets remain less defined.

Monocytes

Monocytes (Fig. 1A) circulate in the blood, bone marrow, and spleen and do not proliferate in a steady state (*3*, *4*). Monocytes represent immune effector cells, equipped with chemokine receptors and adhesion receptors that mediate migration from blood to tissues during infection. They produce inflammatory cytokines and take up cells and toxic molecules. They can also differentiate into inflammatory DCs or macrophages during inflammation and possibly, less efficiently, in the steady state. Migration to tissues and differentiation to inflammatory DCs and macrophages are likely determined by the inflammatory milieu and pathogen-associated pattern-recognition receptors (*5*).

Macrophages

Macrophages (Fig. 1A) are resident phagocytic cells in lymphoid and nonlymphoid tissues and are believed to be involved in steady-state tissue homeostasis, via the clearance of apoptotic cells, and the production of growth factors. Macrophages are equipped with a broad range of pathogen-recognition receptors that make them efficient at phagocytosis and induce production of inflammatory cytokines (*6*). The developmental origin and the function of tissue macrophage subsets, such as microglia (macrophages in the central nervous system), dermal macrophages, and splenic marginal zone and metallophilic macrophages remain insufficiently understood.

Classical DCs

Classical DCs (cDCs) are specialized antigen-processing and -presenting cells, equipped with high phagocytic activity as immature cells and high cytokine-producing capacity as mature cells (*7*, *8*). Although present in human circulation, cDCs are rare in mouse blood. cDCs are highly migratory cells that can move from tissues to the T cell and B cell zones of lymphoid organs via afferent lymphatics and high endothelial venules. cDCs regulate T cell responses

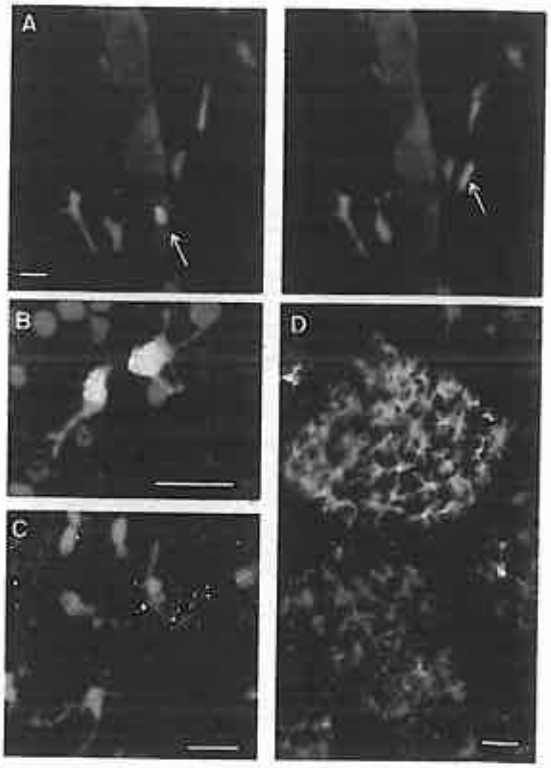


Fig. 1. (**A**) Still frames from time-lapse intravital confocal microscopy of a crawling monocyte (arrow) and perivascular macrophages in the dermis. Scale bar indicates 20 μm, initial magnification 20×. [Credit: F. Geissmann; for details, see (*52*)] (**B**) Dividing LCs in the epidermis. Scale bar, 20 μm; initial magnification 20×. [Credit: I. Chorro and F. Geissmann; for details, see (*35*)] (**C**) Confocal micrograph of aortic whole mount from a $Cx3cr1^{GFP/GFP}$ $Apoe^{-/-}$ mouse viewed from endothelial side. Scale bar, 20 μm; initial magnification 20×. [Credit: K. Ley] (**D**) Intravital two photon microscopic image of intestinal villi of $CD11c^{+}$ cell–depleted mice reconstituted by grafts of bone marrow monocytes expressing red and green fluorescent reporters and yielding red and green fluorescent lamina propria cells, respectively. Scale bar, 20 μm; initial magnification 40×. [Credit: S. Jung and G. Shakhar; for details, see (*26*)]

[1]Centre for Molecular and Cellular Biology of Inflammation, Division of Immunology, Infection, and Inflammatory Diseases, King's College London, Great Maze Pond, London SE1 1UL, UK. [2]Division of Hematology, Department of Internal Medicine, Zürich University Hospital, CH-8091 Zürich, Switzerland. [3]Department of Immunology, The Weizmann Institute of Science, Rehovot 76100, Israel. [4]Centre d'Immunologie de Marseille-Luminy (CIML), Université Aix Marseille, Campus de Luminy, Case 906, 13288 Marseille Cedex 09, France; Institut National de la Santé et de la Recherche Médicale (INSERM), U631 Marseille, France; Centre National de la Recherche Scientifique (CNRS), UMR6102, Marseille, France. [5]Department of Gene and Cell Medicine and Institute for Immunology, Icahn Medical Institute, Mount Sinai School of Medicine, New York, NY 10029, USA. [6]Division of Inflammation Biology, La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037, USA.

*To whom correspondence should be addressed. E-mail: frederic.geissmann@kcl.ac.uk (F.G.); klaus@liai.org (K.L.)

Downloaded from www.sciencemag.org on July 15, 2011

25

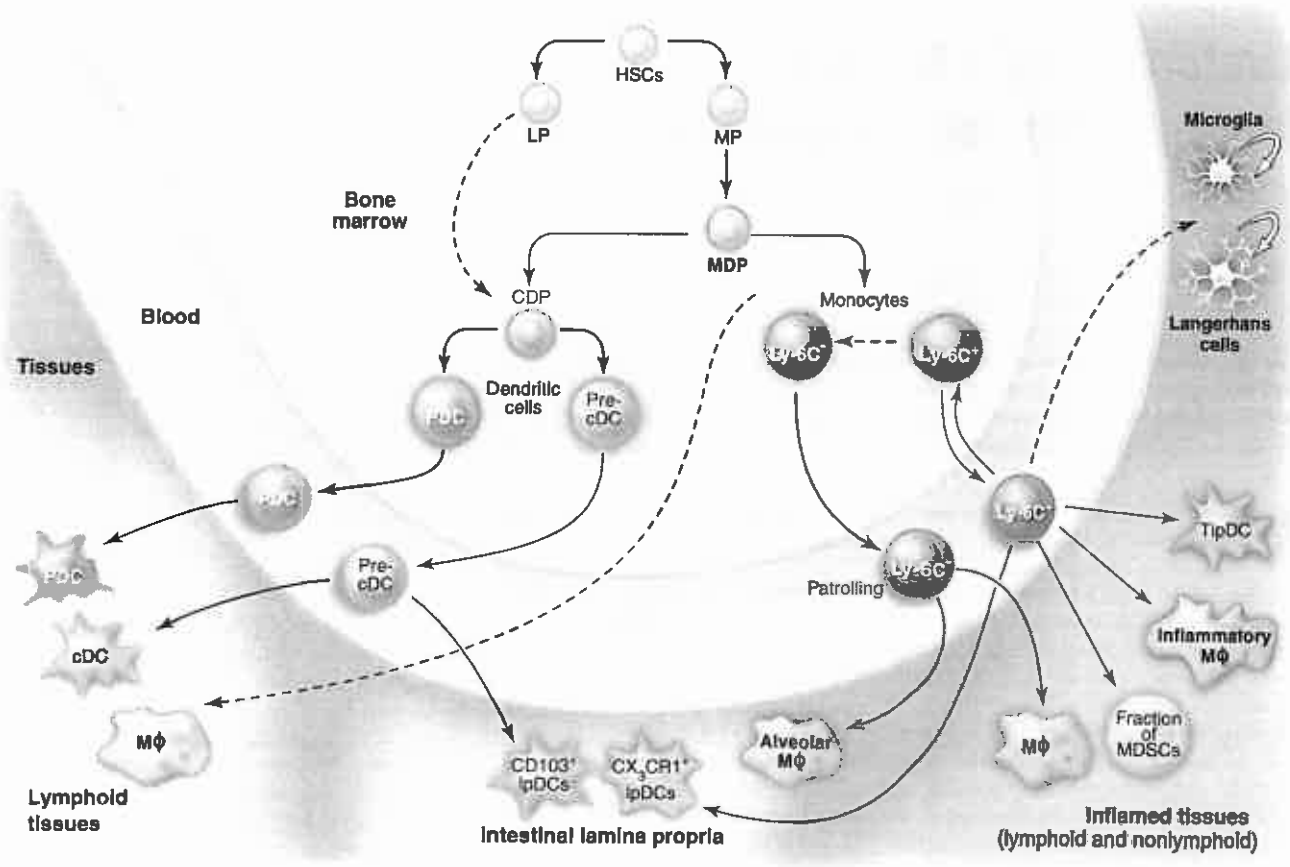


Fig. 2. Differentiation of DCs and macrophages in mice. In the bone marrow, hematopoietic stem cells (HSCs) produce myeloid (MP) and lymphoid (LP) committed precursors. MPs give rise to monocyte, macrophage, and DC precursors (MDPs). MDPs give rise to monocytes, some populations of macrophages, and common DC precursors (CDPs). Two monocyte subsets, Ly-6C$^+$ and Ly-6C$^-$, leave the bone marrow to enter the blood. CDPs give rise to preclassical dendritic cells (pre-cDCs) and plasmacytoid dendritic cells (PDCs). Pre-cDCs circulate in blood and enter lymphoid tissue, where they give rise to CD8α^+ and CD8α^- cDCs, and nonlymphoid tissues, where they may give rise to CD103$^+$ lamina propria DCs (lpDCs). Under homeostatic conditions, Ly-6C$^-$ monocytes may contribute to alveolar macrophages (MΦ), and Ly-6C$^+$ monocytes can become CX$_3$CR1$^+$ lpDCs in nonlymphoid tissues. During inflammation, Ly-6C$^+$ monocytes give rise to monocyte-derived DCs, for example, tumor necrosis factor (TNF) and inducible nitric oxide synthase (iNOS)–producing dendritic cells (TipDCs), inflammatory macrophages, and may contribute to myeloid-derived suppressor cells (MDSCs) associated with tumors. They are also suspected (dashed arrow) to contribute to microglia and Langerhans cells in selected experimental conditions. Microglia and Langerhans cells can renew independently from the bone marrow (curved arrow). HSCs can also leave their bone marrow niche (dashed arrow) and enter peripheral tissues, where they differentiate to myeloid cells during inflammation. It is unclear at this time whether LPs contribute substantially to PDCs and cDCs (dashed arrow).

both in the steady state and during infection. They are generally short-lived and replaced by blood-borne precursors (*9*, *10*). Of note, they are distinct from Langerhans cells (Fig. 1B) (DCs found in the epidermis), which are not replaced by blood-borne cells at steady state (*11*). Individual myeloid cell populations may share features of DCs and macrophages and can be difficult to ascribe to one or the other cell type (Fig. 1, C and D).

Plasmacytoid DCs

Plasmacytoid DCs (PDCs) differ from cDCs in that they are relatively long-lived and a proportion of them carry characteristic immunoglobulin rearrangements (*12*). They are present in the bone marrow and all peripheral organs. PDCs are specialized to respond to viral infection with a massive production of type I interferons (IFNs); however, they also can act as antigen-presenting cells and control T cell responses (*13*).

The development of the mononuclear phagocyte system is controlled by cytokines: small secreted proteins that promote cell-cell communication and can act as growth and differentiation factors. The generation of monocytes and macrophages, and to some extent DCs, is dependent on the cytokine and hematopoietic growth factor receptor Csf1r (also known as c-fms, M-CSFR, and CD115), expressed in monocytes, macrophages, and mononuclear phagocyte precursors (*14*–*17*). Characterization of *op/op* mice, a spontaneous mutant lacking a functional *Csf1* gene, has revealed both the role of Csf1 in the development of mononuclear phagocytes and also their broad functions (*17*). The known ligands of Csf1r, Csf1/M-CSF (*18*) and interleukin (IL)-34 (*19*), are likely both important for the development of the mononuclear phagocyte lineage, because M-CSF–deficient mice have a milder phenotype than Csf1r-deficient mice. Another cytokine receptor closely related to Csf1r, fms-related tyrosine kinase 3 (FLT3, also known as Flk2) receptor, is critical for the development of cDCs and PDCs (*10*, *20*–*22*).

Following on the early work on Csf1r, a large amount of work has been devoted in the past 20 years to investigating the origin and differentiation pathways of monocytes, macrophages, and DCs. This recently culminated in a series of studies that unveiled what is likely to be the main pathway for the develop-

Downloaded from www.sciencemag.org on July 15, 2011

ment of DCs, monocytes, and macrophages from bone marrow progenitors both in the steady state and during inflammation.

What Are the Precursors of Monocytes, Macrophages, and DCs?

The use of multicolor fluorescence-activated cell sorting has allowed identification of progenitors and differentiated cell populations on the basis of the expression of multiple cell surface proteins (table S1). These cells can then be isolated from mouse bone marrow, blood, or spleen and grafted into recipient mice to determine lineage relationships. This methodology was recently used to delineate the DC developmental program in vivo (*3*, *9*, *23*–*26*). Current models propose that blood monocytes, many macrophage subsets, and most DCs originate in vivo from hematopoietic stem cell–derived progenitors with myeloid-restricted differentiation potential (Fig. 2). Successive commitment steps in the bone marrow include common myeloid progenitors (CMPs), granulocyte-macrophage precursors (GMPs), and macrophage/DC progenitors (MDPs). MDPs are a subset of proliferating cells in the bone marrow that share phenotypic characteristics with myeloid precursor populations and give rise to many macrophages and DC subsets (*24*) (table S1) but largely cannot differentiate into granulocytes, another cell type of the myeloid lineage. Within the bone marrow, MDPs differentiate to monocytes and to the common DC precursors (CDPs).

The CDPs are proliferating cells that differentiate into PDCs and the precursors for cDCs (pre-cDCs) in the bone marrow but have lost the potential to give rise to monocytes (*9*, *23*, *27*). At steady state, pre-cDCs are found in the bone marrow, blood, and spleen (*9*, *28*). They enter lymph nodes from the blood through high endothelial venules and integrate into the DC network, where they acquire a mature cDC surface phenotype and morphology. cDC development also involves cell division in lymphoid organs, which is controlled in part by regulatory T cells (*9*), Flt3 (*10*), and the lymphotoxin-β receptor (*29*). Pre-cDCs can also differentiate into $CD103^+$ mucosal DCs in the lamina propria (*26*, *30*).

Monocytes exit to the blood and can enter tissues under inflammatory conditions. They give rise to subsets of macrophages and to inflammatory DCs that share many of the phenotypic features and functions of DCs, such as the ability to process and present antigen to T cells (*3*, *5*, *26*, *31*, *32*); however, it is now widely accepted that monocytes do not give rise to cDCs and PDCs (*9*, *24*).

Are There Exceptions to the Classical Developmental Pathways?

Although many macrophages and DC subsets are renewed from bone marrow progenitors, there are notable exceptions. For example, neither microglia nor Langerhans cells (LCs) are dependent on the bone marrow for their renewal in the steady state and possibly during inflammation (*33*, *34*). LCs were recently shown to develop from an embryonic precursor that colonizes the epidermis before birth, differentiates in situ, and then proliferates during the first week of life to establish the LC network (*35*). Adult LCs self-renew in situ and can massively proliferate during inflammation (*35*). Thus, epidermal LCs appear to be their own progenitors. In contrast, hematopoietic stem cells and precursors recirculate between the periphery and the bone marrow in adults (*36*) and may contribute directly to the differentiation of certain myeloid cells in inflamed tissues (*37*). Lymphoid-restricted progenitors contribute little to peripheral lymphoid organ or tissue DCs, with the possible exception of the thymus. There, developing T cells and potentially a proportion of thymic macrophages, DCs, PDCs, and B cells might be derived directly from blood-borne early thymocyte progenitors (*38*, *39*).

What Are the Origins of Nonlymphoid Tissue Resident Dendritic Cells?

Recent studies have established the existence of two distinct DC populations in the lung and dermis [reviewed in (*34*)], which are characterized as $CD103^+$ and CX_3CR1^+ DCs. Likewise, recent studies have revealed phenotypically definable $CD11c^+$ $CD103^+$ and CX_3CR1^+ subpopulations in the intestinal lamina propria (*26*, *30*, *40*) (table S1), which are derived from pre-cDCs and $Ly6C^+$ monocytes, respectively (*26*). Mice reconstituted exclusively with CX_3CR1^+ lamina propria cells developed severe intestinal inflammation when challenged in an innate colitis model (*26*). Thus, the balance of lamina propria DC subsets is likely important for gut homeostasis. Of note, only $CD103^+$ cells emigrate from the lamina propria to the mesenteric lymph nodes (*30*, *41*). CX_3CR1^+ cells are poor T cell stimulators (*41*) and thus may represent lamina propria macrophages rather than DCs.

What Is the Developmental Relationship Between Mature Monocyte Subsets?

Differentiation of MDPs into monocytes and macrophages is less well understood than their differentiation into DCs. Among differentiated "mature" monocytes in the bone marrow, blood, and spleen, at least two phenotypic and functional subsets can be identified [reviewed in (*3*)], whose developmental relationship is still unclear. Adoptive transfer experiments demonstrate that bone marrow $Gr1^+$/Ly-$6C^{high}$ monocytes can shuttle between the blood and the bone marrow and lose Ly-6C expression (*25*, *42*), suggesting that they give rise to $Gr1^-$/Ly-$6C^{low}$ monocytes. However, neither a genetic defect in nor antibody-mediated depletion of $Gr1^+$/Ly-$6C^{high}$ monocytes affected the generation of $Gr1^-$/Ly-$6C^{low}$ monocytes (*43*–*46*). When bone marrow is used as a source of monocytes, current adoptive transfer protocols may not always distinguish mature $Gr1^+$/Ly-$6C^{high}$ monocytes from their proliferating precursors because pre-monocytes are ill-defined. Therefore, a better characterization of monocyte precursors and new lineage tracking studies will be needed to establish the developmental relationship between monocyte subsets.

Are Monocytes Effector Cells or Circulating Precursors?

Monocytes have long been considered as a developmental intermediate between bone marrow precursors and tissue macrophages. It is now clear, however, that many DCs and tissue macrophages do not originate from monocytes in a steady state. Conversely, monocytes carry out specific effector functions during inflammation (*3*). "Inflammatory" $Gr1^+$/Ly-$6C^{high}$ monocytes are referred to as such because they give rise to macrophages and DCs in a variety of infectious models. During infection with *Listeria monocytogenes*, $Gr1^+$/Ly-$6C^{high}$ monocytes differentiate into DCs that produce inflammatory mediators such as tumor necrosis factor α, nitric oxide, and reactive oxygen species (known as TipDCs) (*5*, *47*). $Gr1^+$/Ly-$6C^{high}$ monocytes are required for resistance to the pathogen *Toxoplasma gondii* but in this case differentiate into mucosal macrophages, which differ in surface phenotype and inflammatory mediator production from TipDCs (*48*, *49*). In some tumors, up to half of the myeloid-derived suppressor cells [cells that can protect tumors from immune attack (*50*)] are $Gr1^+$/Ly-$6C^{high}$. Upon spinal cord injury, recruited $Gr1^+$/Ly-$6C^{high}$ monocytes differentiate into macrophages that critically contribute to the recovery (*51*). Thus, $Gr1^+$/Ly-$6C^{high}$ monocytes can differentiate into a variety of macrophages and DCs subtypes that can both activate and inhibit the immune response, depending on local or systemic cues received and the nature of encountered pathogen.

Are $Gr1^-$/Ly-$6C^{low}$ Monocytes a New Cell Subset with a Role in Tissue Repair?

$Gr1^-$/Ly-$6C^{low}$ monocytes represent a functionally distinct monocyte subset (*52*). They are characterized by a smaller size, lack of expression of CCR2 and L-selectin, but higher expression of the chemokine receptor CX_3CR1 and the integrin LFA-1 (*31*) (table S1). $Gr1^-$/Ly-$6C^{low}$ monocytes were initially termed "resident" in mice because of their longer half-life in vivo and their localization to both resting and inflamed tissues after adoptive transfer. Intravital microscopy revealed that $Gr1^-$/Ly-$6C^{low}$ monocytes exhibit long-range crawling on the luminal side of the vascular endothelium and are ideally located to survey endothelial cells and surrounding tissues for damage or infection (*52*). Moreover, $Gr1^-$/Ly-$6C^{low}$ monocytes are recruited at a late phase to the ischemic myocardium, where they express vascular-endothelial growth factor (VEGF), and are proposed to promote healing via myofibroblast accumulation, angiogenesis, and deposition of collagen (*53*). Transcriptional profiling of extravasated monocytes retrieved from the peritoneum

Downloaded from www.sciencemag.org on July 15, 2011

27

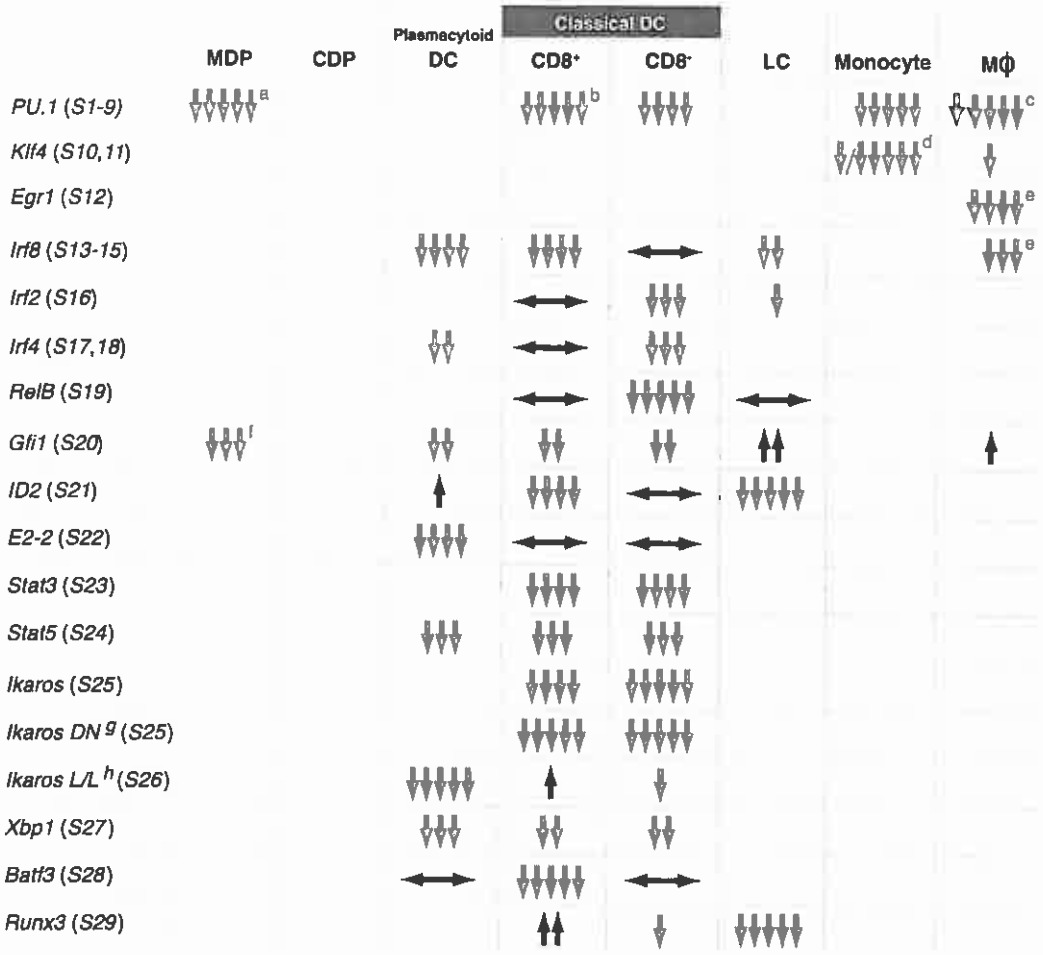


Fig. 3. Phenotype of transcription factor knockout (KO) mice in different populations of the MPs. The ↓ indicates a reduction in KO; ↑ is an increase in KO; and ↔ means unchanged numbers of the populations, suggesting a positive (green), negative (red), or no (black) effect on the development of the respective population. Number of arrows indicates the relative strength of the effect. Specific notes are as follows: (a) MDPs have not been specifically analyzed but PU.1-deficient mice lack all myeloid progenitors, of which MDPs are a subpopulation; (b) in one study (*S7*), $CD8^+$ $CD11c^+$ cells were detected in E16.5 embryos; (c) some macrophage subpopulations are present in the embryo; (d) $Ly6C^-/Ly6C^+$, respectively; (e) M-CSF–dependent differentiation in culture; (f) MDPs were not specifically studied, but the population of lin^- $ckit^+$ $Flt3^+$ progenitors, which includes MDPs, was reduced; (g) dominant negative allele; and (h) low-level expression. References *S1* to *S29* are found in the supporting online material (SOM).

after *L. monocytogenes* infection (*3*, *52*) indicates that $Gr1^-/Ly\text{-}6C^{low}$ monocytes initiate a macrophage differentiation program that resembles the one described for "alternatively activated macrophages," cells that are proposed to be involved in tissue repair (*54*) (also known as M2 macrophages). Conversely $Gr1^+/Ly\text{-}6C^{high}$ monocytes that entered the peritoneum at the same time initiate a more inflammatory "classical" transcriptional program that resembles TipDCs or M1 macrophage differentiation. Direct experimental evidence that monocytes are involved in tissue repair is missing, however, and the differential contributions of monocyte subsets to neoangiogenesis (*55*, *56*) need further study.

What Determines Monocyte Differentiation Potential?

The emerging evidence that monocyte subsets exert distinct functions and have distinct fates, such as the differentiation into cells with features of M1 and M2 macrophages, seemingly contradicts the notion of a general plasticity of monocytes and macrophages, which holds that monocytes respond to their environment by differentiating into a variety of macrophages and DC-like cells [reviewed in (*57*)]. This concept of plasticity is largely based on the effects of cytokines on monocytes in vitro. Exposure to granulocyte-macrophage colony-stimulating factor (GM-CSF) and IL-4 induces differentiation of human and mouse monocytes into DCs, irrespective of their subset (*58*, *59*). Moreover, addition of transforming growth factor-β1 confers the phenotype of LCs (*60*), whereas exposure to M-CSF induces monocytes to differentiate into macrophages. Addition of IFNγ (or lipopolysaccharide) to M-CSF induces the differentiation of M1-like macrophages, whereas addition of IL-4 induces the differentiation of M2-like macrophages (*54*, *61*, *62*). M2 macrophages have been further subsetted depending on the culture conditions (*62*). The chemokine CXCL4 induces another macrophage phenotype with possible relevance to vascular diseases (*63*).These results suggest that monocytes are phenotypically polarized by the microenvironment to mount specific functional programs. M1 macrophages, whose prototypical activating stimuli are IFNγ and LPS, exhibit potent microbicidal properties and promote strong IL-12–mediated T helper 1 responses. In contrast, M2 macrophages support T helper 2–associated effector functions and may play a role in resolution of inflammation through endocytic clearance and trophic factor synthesis (*54*). In vitro studies may not, however, fully recapitulate in vivo differentiation, and thus heterogeneity of monocytes may mediate in part their plasticity in vivo, with $Gr1^-/Ly\text{-}6C^{low}$ and $Gr1^+/Ly\text{-}6C^{high}$ monocytes being committed to differentiate more readily into M2-like cells or inflammatory DCs and M1-like macrophages, respectively.

Transcriptional Control of Lineage Commitment in the Mononuclear Phagocyte System

The successively more restricted lineage potentials of MDPs, CDPs, pre-DCs, monocytes, macrophages, and DCs involve the selection of specific gene expression programs. For some pathways, the transcription factors responsible for these cell fate choices have recently begun to be elucidated. Although several transcription factor knockout mice show deficiencies in the mononuclear phagocyte system, they often display broad effects in multiple cell types (*3*, *34*, *64*) (Fig. 3). A few exceptions have been identified, such as the loss of PDCs in mice deficient for the basic helix-loop-helix transcription factor E2-2 (*65*) or the loss of $CD8\alpha^+$ cDCs in mice deficient for the basic leucine zipper transcription factor Batf3 (*66*). Similarly selective transcription factor requirements may be responsible for monocyte (*44*) and macrophage subtypes and will require further investigation.

However, broad effects on many lineages do not necessarily exclude important functions in specific commitment events. An instructive example is the myeloid transcription factor

Downloaded from www.sciencemag.org on July 15, 2011

PU.1. Although PU.1 is required for the earliest steps of myeloid lineage commitment in hematopoietic stem cells (*67*) and its absence results in general myeloid lineage deficiencies (*3*, *34*), PU.1 has additional key selector gene functions at several branch points of myeloid lineage diversification, particularly during the macrophage-versus-DC choice of monocytes (*68*). The function of PU.1 at particular progenitor stages depends on its balance with antagonistic factors driving alternative fates. Intermediate expression of PU.1 in GMPs can overcome the neutrophil fate-inducing effects of the basic leucine zipper transcription factor C/EBPα and activate the macrophage-specifying zinc finger transcription factors Egr-1 and Egr-2 (*3*, *69*). In contrast, high expression of PU.1 is required to induce a DC fate in monocytes and to antagonize the macrophage-inducing transcription factors c-Maf and MafB (*3*, *68*). Of interest, the transcription factor balance specifying M2 or DC fate of monocytes in culture correlates with cell fate choice in vivo, as outlined above for the programs initiated by $Ly6C^{low}$ and $Ly6C^{high}$ monocytes upon *L. monocytogenes* infection of the peritoneum (*52*, *68*).

In general, gain-of-function experiments have been informative in defining transcription factor potential in driving particular cell fate choices. For example, ectopic expression of the transcription factors MafB, c-Maf, Egr1, ICSBP/IRF8, KLF4, and PU.1 in early progenitors can drive monocyte/macrophage fates. PU.1 and RelB induce DC differentiation via monocytic intermediates and E-box proteins (likely mimicking endogenous E2-2) or SpiB can induce PDC fates (*3*). In most of these experiments, however, undefined or early progenitor-enriched populations were used. Given that the function of any factor will depend on other cooperating and antagonistic transcription factors expressed at each specific progenitor stage (*70*), this makes it difficult to precisely identify the affected commitment event. Such differentiation stage-specific manipulation of transcription factors will require models that faithfully reproduce defined differentiation stages in culture, the development of new genetic tools of inducible gene regulation in vivo, or adoptive transfer of genetically marked, defined cell populations.

Another aspect of emerging importance is the coupling of cytokine receptor signaling with specific transcription factors in lineage commitment. For example, signal transducer and activator of transcription 3 (STAT3) is required for Flt3L- but not GM-CSF–mediated DC differentiation (*21*, *34*, *64*). Similarly, Id2 and Runx3 mediate TGFβ signals in LC differentiation (*34*, *64*). Furthermore, MafB limits M-CSF signals and inhibits PU.1 activation in HSCs (*67*), whereas MafB and c-Maf together inhibit M-CSF–dependent proliferative signals in mature monocytes and macrophages and assure their withdrawal from cell cycle (*71*).

Besides transcription factors, epigenetic modification, with large-scale histone and DNA modifications like acetylation and methylation, and micro-RNAs have recently been reported to regulate important functions of mononuclear phagocytes (*72*, *73*) and are also emerging as important determinants of lineage choice but will not be discussed here.

Perspectives and Future Challenges

Although the studies discussed above have shed light on the in vivo development and homeostasis of the mononuclear phagocyte system, mechanisms that control replacement of many macrophage subsets and resident cells remain mysterious. Most organs contain resident tissue macrophages, many of which have long half-lives. After irradiation and other forms of tissue injury, bone marrow–derived cells frequently repopulate the injured organs. Whether bone marrow–derived cells under inflammatory conditions assume the same phenotype and function as resident cells that populate tissues under steady state conditions remains largely unclear. Fate mapping experiments investigating steady state developmental potential will therefore have to avoid experimental inflammation. Toward this end, Cre recombinase expression under the control of promoters for key markers (such as lysozyme) have been used to activate conditional reporter genes so that all downstream daughter cells express a stable marker (*9*). It has to be cautioned, however, that "key" markers are seldom completely specific; progenitor cells can express low levels of the marker and marker-positive cells may not uniformly turn on the reporter, all of which may complicate the interpretation of such lineage-marking experiments (*74*, *75*). Parabiotic mice and adoptive transfer into mice in which cell populations have been depleted by genetic means may provide complementary approaches.

The transcription factors controlling monocyte and macrophage subtype-specific programs in vivo also need to be identified. Models that faithfully reproduce defined differentiation stages in culture and the development of new genetic tools of inducible gene regulation that enable precise differentiation by stage-specific manipulation of transcription factors in vivo could accelerate discovery in this area. The use of new model systems for the study of phagocytes, such as *Drosophila melanogaster*, which lack lymphocytes but possess a powerful innate immune system and are perfectly suited for genetic studies and in vivo imaging, has the potential to accelerate discovery of genetic pathways that control the development and functions of phagocytes in vivo (*76*–*78*).

Another challenge will be to translate knowledge obtained from mouse and other models into a better understanding of the mononuclear phagocyte system and its role in inflammation and associated diseases in humans. A large number of studies have attempted to recapitulate in vitro some of the heterogeneity of DCs and macrophages. On the basis of these studies, functional subsets of macrophages and DCs (e.g., tolerogenic DCs) inside tissues have been proposed to mediate important functions in vivo, such as tissue repair, angiogenesis, or tolerance to self. Laser capture dissection techniques are being developed to obtain gene expression data from macrophages isolated from tissues (*79*). Flow cytometry–based cell sorting approaches also hold promise for isolating pure macrophage and DC populations from various tissues and organs. These techniques are especially important in order to translate some of the findings made in mice into the human system. Global expression profiling of mouse DC, macrophages, and monocyte subsets might assist in the identification of corresponding human populations. Human hemato-lymphoid-system mouse models that will provide in vivo DC and macrophage readouts from human hematopoietic stem cells and progenitor cells will also be important (*80*). Lastly, studies of patients with deficiencies for genes involved in myeloid cell differentiation and functions will be critical to assess the relevance of models derived from studies in the mouse and in vitro experiments in human cells (*81*).

References and Notes

1. R. van Furth, Z. A. Cohn, *J. Exp. Med.* **128**, 415 (1968).
2. R. M. Steinman, D. S. Lustig, Z. A. Cohn, *J. Exp. Med.* **139**, 1431 (1974).
3. C. Auffray, M. H. Sieweke, F. Geissmann, *Annu. Rev. Immunol.* **27**, 669 (2009).
4. F. K. Swirski *et al.*, *Science* **325**, 612 (2009).
5. N. V. Serbina, T. Jia, T. M. Hohl, E. G. Pamer, *Annu. Rev. Immunol.* **26**, 421 (2008).
6. S. Gordon, *Cell* **111**, 927 (2002).
7. J. Banchereau, R. M. Steinman, *Nature* **392**, 245 (1998).
8. I. Mellman, R. M. Steinman, *Cell* **106**, 255 (2001).
9. K. Liu *et al.*, *Science* **324**, 392 (2009); published online 12 March 2009 (10.1126/science.1170540).
10. C. Waskow *et al.*, *Nat. Immunol.* **9**, 676 (2008).
11. M. Merad *et al.*, *Nat. Immunol.* **3**, 1135 (2002).
12. L. Corcoran *et al.*, *J. Immunol.* **170**, 4926 (2003).
13. M. Colonna, G. Trinchieri, Y. J. Liu, *Nat. Immunol.* **5**, 1219 (2004).
14. R. T. Sasmono *et al.*, *Blood* **101**, 1155 (2003).
15. X. M. Dai *et al.*, *Blood* **99**, 111 (2002).
16. M. G. Cecchini *et al.*, *Development* **120**, 1357 (1994).
17. W. Wiktor-Jedrzejczak, S. Gordon, *Physiol. Rev.* **76**, 927 (1996).
18. E. S. Kawasaki *et al.*, *Science* **230**, 291 (1985).
19. H. Lin *et al.*, *Science* **320**, 807 (2008).
20. H. J. McKenna *et al.*, *Blood* **95**, 3489 (2000).
21. N. Onai, A. Obata-Onai, R. Tussiwand, A. Lanzavecchia, M. G. Manz, *J. Exp. Med.* **203**, 227 (2006).
22. A. D'Amico, L. Wu, *J. Exp. Med.* **198**, 293 (2003).
23. N. Onai *et al.*, *Nat. Immunol.* **8**, 1207 (2007).
24. D. K. Fogg *et al.*, *Science* **311**, 83 (2006); published online 1 December 2005 (10.1126/science.1117729).
25. C. Varol *et al.*, *J. Exp. Med.* **204**, 171 (2007).
26. C. Varol *et al.*, *Immunity* **31**, 502 (2009).
27. S. H. Naik *et al.*, *Nat. Immunol.* **8**, 1217 (2007).
28. S. H. Naik *et al.*, *Nat. Immunol.* **7**, 663 (2006).
29. K. Kabashima *et al.*, *Immunity* **22**, 439 (2005).
30. M. Bogunovic *et al.*, *Immunity* **31**, 513 (2009).
31. F. Geissmann, S. Jung, D. R. Littman, *Immunity* **19**, 71 (2003).
32. C. Auffray *et al.*, *J. Exp. Med.* **206**, 595 (2009).
33. B. Ajami, J. L. Bennett, C. Krieger, W. Tetzlaff, F. M. Rossi, *Nat. Neurosci.* **10**, 1538 (2007).
34. M. Merad, M. G. Manz, *Blood* **113**, 3418 (2009).

Downloaded from www.sciencemag.org on July 15, 2011

29

35. L. Chorro *et al.*, *J. Exp. Med.* **206**, 3089 (2009).
36. D. E. Wright, A. J. Wagers, A. P. Gulati, F. L. Johnson, I. L. Weissman, *Science* **294**, 1933 (2001).
37. S. Massberg *et al.*, *Cell* **131**, 994 (2007).
38. L. Wu, K. Shortman, *Semin. Immunol.* **17**, 304 (2005).
39. J. Li, J. Park, D. Foss, I. Goldschneider, *J. Exp. Med.* **206**, 607 (2009).
40. E. Jaensson *et al.*, *J. Exp. Med.* **205**, 2139 (2008).
41. O. Schulz *et al.*, *J. Exp. Med.* **206**, 3101 (2009).
42. U. Yrlid, C. D. Jenkins, G. G. MacPherson, *J. Immunol.* **176**, 4155 (2006).
43. M. W. Feinberg *et al.*, *EMBO J.* **26**, 4138 (2007).
44. J. K. Alder *et al.*, *J. Immunol.* **180**, 5645 (2008).
45. J. C. Scatizzi *et al.*, *Am. J. Pathol.* **168**, 1531 (2006).
46. A. Mildner *et al.*, *Nat. Neurosci.* **10**, 1544 (2007).
47. E. Narni-Mancinelli *et al.*, *J. Exp. Med.* **204**, 2075 (2007).
48. I. R. Dunay *et al.*, *Immunity* **29**, 306 (2008).
49. P. M. Robben, M. LaRegina, W. A. Kuziel, L. D. Sibley, *J. Exp. Med.* **201**, 1761 (2005).
50. K. Movahedi *et al.*, *Blood* **111**, 4233 (2008).
51. R. Shechter *et al.*, *PLoS Med.* **6**, e1000113 (2009).
52. C. Auffray *et al.*, *Science* **317**, 666 (2007).
53. M. Nahrendorf *et al.*, *J. Exp. Med.* **204**, 3037 (2007).
54. F. O. Martinez, L. Helming, S. Gordon, *Annu. Rev. Immunol.* **27**, 451 (2009).
55. M. Grunewald *et al.*, *Cell* **124**, 175 (2006).
56. M. De Palma *et al.*, *Cancer Cell* **8**, 211 (2005).
57. P. R. Taylor, S. Gordon, *Immunity* **19**, 2 (2003).
58. F. Sallusto, A. Lanzavecchia, *J. Exp. Med.* **179**, 1109 (1994).
59. F. Geissmann *et al.*, *Immunol. Cell Biol.* **86**, 398 (2008).
60. F. Geissmann *et al.*, *J. Exp. Med.* **187**, 961 (1998).
61. M. Stein, S. Keshav, N. Harris, S. Gordon, *J. Exp. Med.* **176**, 287 (1992).
62. F. O. Martinez, A. Sica, A. Mantovani, M. Locati, *Front. Biosci.* **13**, 453 (2008).
63. C. A. Gleissner *et al.*, *Circ. Res.* **106**, 203 (2010).
64. M. Zenke, T. Hieronymus, *Trends Immunol.* **27**, 140 (2006).
65. B. Cisse *et al.*, *Cell* **135**, 37 (2008).
66. K. Hildner *et al.*, *Science* **322**, 1097 (2008).
67. S. Sarrazin *et al.*, *Cell* **138**, 300 (2009).
68. Y. Bakri *et al.*, *Blood* **105**, 2707 (2005).
69. P. Laslo *et al.*, *Cell* **126**, 755 (2006).
70. M. H. Sieweke, T. Graf, *Curr. Opin. Genet. Dev.* **8**, 545 (1998).
71. A. Aziz *et al.*, *Mol. Cell. Biol.* **26**, 6808 (2006).
72. R. M. O'Connell *et al.*, *J. Exp. Med.* **205**, 585 (2008).
73. R. M. O'Connell, K. D. Taganov, M. P. Boldin, G. Cheng, D. Baltimore, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1604 (2007).
74. T. Miyamoto *et al.*, *Dev. Cell* **3**, 137 (2002).
75. H. Shigematsu *et al.*, *Immunity* **21**, 43 (2004).
76. J. A. Martinez-Agosto, H. K. Mikkola, V. Hartenstein, U. Banerjee, *Genes Dev.* **21**, 3044 (2007).
77. D. T. Babcock *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10017 (2008).
78. L. Cuttell *et al.*, *Cell* **135**, 524 (2008).
79. E. Trogan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2234 (2002).
80. M. G. Manz, *Immunity* **26**, 537 (2007).
81. J. L. Casanova, L. Abel, *Annu. Rev. Immunol.* **20**, 581 (2002).

82. The diagram depicted in Fig. 2 is the result of a collective discussion at a recent workshop: "Monocytes, Macrophages and Dendritic Cell Heterogeneity," 2 to 7 March 2009, Paris, France, organized by the Fondation des Treilles (www.les-treilles.com/newsite/index.html). We thank V. Bronte, I. Charo, J. Van Ginderachter, G. MacPherson, A. Mantovani, E. Pamer, R. Steinman, and F. Swirski for their input into Fig. 1 and for their ideas that we have incorporated into this review. Any misrepresentation is the responsibility of the authors. Work by F.G. and K.L. on the development of monocytes, macrophages, and dendritic cells is funded by Medical Research Council UK grant G0900867 to F.G. and grant NIH R01 HL058108 to K.L.

Supporting Online Material

www.sciencemag.org/cgi/content/full/327/5966/656/DC1
Table S1
References

10.1126/science.1178331

Downloaded from www.sciencemag.org on July 15, 2011

ERRATUM

Post date 3 December 2010

Review: "Development of monocytes, macrophages, and dendritic cells" by F. Geissmann *et al.* (5 February, p. 656). Reference 71 was incorrect. It should be A. Aziz *et al.*, *Science* **326**, 867 (2009).

Downloaded from www.sciencemag.org on July 15, 2011

Processed peptides

The internalized foreign protein antigens are degraded into various peptide fragments. This is inferred by the fact that there are peptides associated with the class I and II major histocompatibility complex (MHC) molecules sitting on the cell surfaces of the cells mentioned before, namely B lymphocytes, moncytoes, macrophages and dendritic cells, with internalized antigen molecules. The detail properties of the MHC class I and II molecules will be discussed later. Two classical papers discussed the eluted peptides from MHC class I (Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, Cox AL, Appella E, Engelhard VH (1992) Characterization of peptides bound to the class I MHC molecule HLA0A2.1 by mass spectrometry. Science, 255, 1261-1263), and from MHC class II (Chicz RM, Urban RG, Lane WS, Gorga JC, Stern LJ, Vignali DAA, Strominger JL (1992) Predominant naturally processed peptides bound to HLA-DE1 are derived from MHC-related molecules and are heterogeneous in size. Nature, 358,764-768). They are mostly fragments from human proteins. Presumably, foreign proteins will be processed in a similar fashion.

When the foreign antigen protein is internalized in these specific cells, they are broken down into fragments. Some specific fragments are then loaded onto membrane bound MHC class I or II molecules. The peptides associated with MHC class I molecules are usually 9 amino acid residues in length, with possible variation of 8 to 10. Those associated with MHC class II molecules are roughly 15 amino acid residues in length, varying from about 13 to 25.

Of course, one may wonder whether normal human proteins can also be degraded outside these cells in the circulation or tissue space. Their fragments can then be loaded directly onto the cell surface bound MHC class I or II molecules outside the cell.

In the study by Hunt et al. of MHC class I, four processed peptides were isolated:

1. LLDVPTAAV Position 27 to 35 of gamma-interferon induced protein precurosr
2. SLLPAIVEL Position 415 to 423 of protein phosphatase 2A beta
3. YLLPAIVHI Position 225 to 233 of DEAD box polypeptide 17
4. TLWVDPYEV Position 103 to 111 of B-cell translocation gene 1

They are all 9 amino acid residues in length, and are located at positions listed above in human proteins as identified by the BLAST program at http://www.ncbi.nlm.nih.gov .

In the study of Chicz et al. of MHC class II, peptides of varying lengths were isolated. They are located in the same regions of human proteins. As they have summarized, in most cases, each human protein only contributes one peptide segment but with some length variations. In few cases, two to four peptides are derived from the same protein.

Characterization of Peptides Bound to the Class I MHC Molecule HLA-A2.1 by Mass Spectrometry

DONALD F. HUNT,* ROBERT A. HENDERSON,*
JEFFREY SHABANOWITZ,* KAZUYASU SAKAGUCHI,*
HANSPETER MICHEL, NOELLE SEVILIR, ANDREA L. COX,
ETTORE APPELLA, VICTOR H. ENGELHARD

Antigens recognized by T cells are expressed as peptides bound to major histocompatibility complex (MHC) molecules. Microcapillary high-performance liquid chromatography–electrospray ionization–tandem mass spectrometry was used to fractionate and sequence subpicomolar amounts of peptides isolated from the MHC molecule HLA-A2.1. Of 200 different species quantitated, eight were sequenced and four were found in cellular proteins. All were nine residues long and shared a distinct structural motif. The sensitivity and speed of this approach should enhance the analysis of peptides from small quantities of virally infected and transformed cells as well as those associated with autoimmune disease states.

CYTOTOXIC T LYMPHOCYTES (CTLs) are a part of the immune system concerned with recognition of host cells that express new antigens as a result of viral infection or transformation. CTLs do not recognize new antigens directly, but only as short peptides bound to a deep cleft in class I molecules of the MHC (*1–3*). Newly synthesized viral and cellular proteins are degraded into peptides in the cytoplasm, transported to the endoplasmic reticulum where they bind to class I molecules, and then expressed on the cell surface (*4–7*). Each of the allelic forms of the class I MHC molecule binds to a complex mixture of structurally distinct peptides (*8, 9*). Information on the nature of these peptides has been obtained from studies with synthetic peptides (*10–12*) and from Edman degradation applied to unfractionated mixtures of peptides extracted from five different class I MHC molecules (*8*). Sequences of 11 peptides extracted from HLA-B27 were identified after high-performance liquid chromatography (HPLC) fractionation and Edman degradation (*9*). Because HPLC was unable to completely resolve the complex mixture, this analysis could only be applied to the few fractions that contained one or two dominant peptides. Declining PTH (phenylthiohydantoin)–amino acid yields made it difficult to determine the exact number of residues in several peptides.

We have applied microcapillary HPLC–electrospray ionization–tandem mass spectrometry to circumvent the above problems. In a matter of hours, this technique determines the molecular mass and therefore maximum length of each peptide component, and the approximate number and quantity of individual peptides. Sequence information can be also obtained on subpicomolar amounts of peptides. We analyzed the naturally processed peptides bound to HLA-A2.1, one of the most widely distributed class I molecules within the human population. The three-dimensional structure of this molecule allows modeling of the complex (*2*).

HLA-A2.1 molecules were purified by immunoprecipitation from the human B lymphoblastoid cell line C1R-A2.1. The associated peptides were released by acid extraction and separated from HLA-A2.1 and antibody by filtration. A "mock" extract was prepared from an equal number of Hmy2.C1R (C1R) cells, which do not express any HLA-A or -B products (*13*). Since C1R-A2.1 is a transfectant of C1R, these cell lines should be identical except for the expression of HLA-A2.1. Extracts were fractionated by microcapillary reversed-phase HPLC and eluted directly into the electrospray ion source of a triple quadrupole mass spectrometer (*14*). A broad band of material was observed in the extract from C1R-A2.1 (Fig. 1A), but not in the extract from C1R (Fig. 1B). If all HLA-A2.1 molecules contain bound peptide, the total peptide yield from 10^8 C1R-A2.1 cells should be roughly 20 pmol; we estimate that our total yield of peptide was about 70 to 80% of the amount expected. This provides evidence that the material isolated for C1R-A2.1 cells by this method is specifically associated with the HLA-A2.1 molecule.

Material with a nominal mass-to-charge ratio (m/z) of 1121 was present in the extract from C1R-A2.1 cells but not that from C1R cells (Fig. 1 insets). By a similar analysis, approximately 200 distinct ions having a signal-to-noise ratio (S/N) of at least 2 were observed in the material eluted from C1R-A2.1 cells. Of these, 47% were singly charged, 45% were doubly charged, and 8% were triply charged. Calculated molecular masses ranged between 737 and 1370. Known quantities of synthetic peptides added to cell extracts allowed quantitation of peptides, and defined the lower limit of detection (S/N = 2) as 30 fmol. Of the 200 peptides detected, 10% were present at 150 to 600 fmol per 10^8 cells,

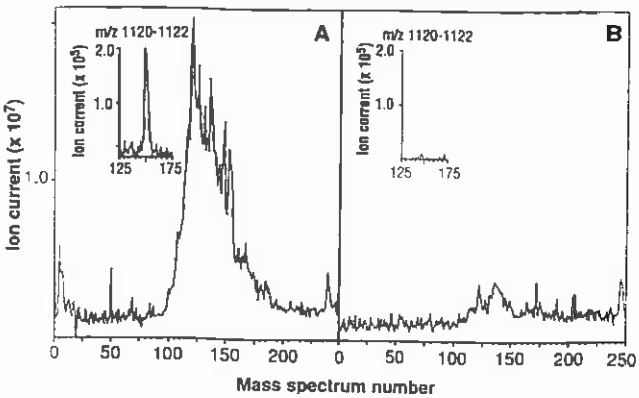


Fig. 1. Total ion chromatograms of HLA-A2.1–associated peptides. C1R-A2.1 (**A**) or C1R (**B**) cells (2×10^9) were solubilized in 20 mM tris, pH 8.0, 150 mM NaCl, 1% NP-40, 2 mM PMSF, 100 μM iodoacetamide, Aprotinin (5 μg/ml), Leupeptin (10 μg/ml), Pepstatin A (10 μg/ml), EDTA (3 ng/ml), and 0.2% sodium azide. After sample centrifugation at 100,000*g* for 1 hour, supernatants were immunoprecipitated with the HLA-A2.1–specific monoclonal antibody BB7.2 (*30*) and protein A–Sepharose (Sigma). Precipitates were extracted with 0.2 N acetic acid. The supernatants were brought to pH 2.1 with glacial acetic acid and boiled for 5 min, centrifuged through an Ultrafree-Cl unit (Millipore) with a nominal limit of 5 kD, and concentrated by vacuum centrifugation. Mass spectra were recorded on a Finnigan-MAT TSQ-70 (San Jose, California) triple quadrupole mass spectrometer equipped with an electrospray ion source and a C-18 microcapillary-HPLC column (75 μm by 10 cm) (*14*). The column was eluted with a gradient of 0 to 80% acetonitrile in 0.5% acetic acid over 10 min at 1 to 2 μl/min. The electrospray needle was operated with a voltage differential of 4 to 5 keV and a sheath flow of 2 to 4 μl/min of a 3:1 mixture of methanol:0.5% acetic acid. Total ion signal from samples equivalent to 10^8 cells was obtained by scanning the range of masses corresponding to m/z values between 300 and 1500 every 2 s and then summing each of the spectra. Insets are plots of ion current for the peptide $(M+H)^+$ ions at m/z 1121 (nominal mass).

D. F. Hunt, J. Shabanowitz, H. Michel, A. L. Cox, Department of Chemistry, University of Virginia, Charlottesville, VA 22903.
R. A. Henderson, N. Sevilir, V. H. Engelhard, Department of Microbiology and Beirne Carter Center for Immunological Research, University of Virginia School of Medicine, Charlottesville, VA 22908.
K. Sakaguchi and E. Appella, Laboratory of Cell Biology, National Cancer Institute, NIH, Bethesda, MD 20892.

*Contributions of these authors were equivalent and their order should be considered arbitrary

33

whereas 90% were present at 30 to 150 fmol. By this measure, individual peptides in the mixture represent between 0.08 and 1.2% of the total peptide bound. Depending on the cell, this would correspond to as few as 100 complexes per cell or as many as 3000. These numbers are in accord with estimates of the number of peptide-MHC complexes required to stimulate T cells (*12, 15, 16*). However, 67 of the 200 peptides detected carried two-thirds of the ion current observed for this group. Thus, HLA-A2.1 seems to be preferentially occupied by a relatively small set of peptides. On the other hand, the number of peptides detected accounts for only 50% of the total ion current. If the remaining undetected peptides are all present at amounts near that of the background, the total number of different peptides presented by HLA-A2.1 could easily exceed 1000.

Table 1 shows a partial list of data obtained on peptides chosen at random from the more than 200 species observed. Fourteen of the peptides were in extracts of immunoprecipitates from both C1R-A2.1 and another HLA-A2.1–expressing cell line, JY, but not in extracts from C1R cells. Five others afforded relatively weak signals in extracts from the JY cells but were undetected in extracts from either C1R-A2.1 or C1R cells. To obtain sequence information, $(M+H)^+$ and $(M+2H)^{2+}$ ions were subjected to collision-activated dissociation (CAD) on the triple quadrupole mass spectrometer (*17*). Representative results for the $(M+H)^+$ ion at *m/z* 1121 (peptide 19) are shown in Fig. 2. Complete sequences for 8 peptides in Table 1 have been obtained; partial sequences have been obtained on 11 others. Each of the completely sequenced peptides is a nonamer and contains either Leu or Ile at position 2. Since Ile and Leu are of identical mass, they cannot be differentiated on the triple quadrupole instrument. Leu was considered a "dominant residue" at position 2 in unfractionated peptide mixtures isolated from HLA-A2.1 (*8*). For peptide 19, a definitive assignment of Leu at position 2 was established by adding synthetic peptides containing Leu or Ile at this position to the cell extract, and demonstrating that only the Leu-containing peptide coeluted from a microcapillary HPLC column with the naturally processed species. For the 19 peptides for which COOH-terminal information is available, all end in residues with aliphatic hydrocarbon side chains: nine terminate in Val, eight in Leu/Ile, and two in Ala [Table 1 and (*18*)]. Only Val was categorized as "dominant" based on the pool sequencing technique (*8*). No other residues were present at any position in more than 50% of the peptides. None of the sequences in Table 1 contains Met at position 2, Lys at position 4, Val at position 6, or Lys at position 8, assignments that were considered "strong" in analysis of pooled peptides (*8*).

The length of the peptides bound to HLA-A2.1 is identical to that of peptides bound to HLA-B27 (*9*) and inferred for other class I molecules (*12*). This could reflect the mechanism by which these peptides were generated, or their ability to bind with particularly high affinity (*3, 19*). However, the motif for peptides bound to HLA-A2.1 is different from that of peptides bound to HLA-B27. Position 2 was also the most conserved in the peptides isolated from HLA-B27, but was an Arg (*9*). This residue is suggested to interact with negatively charged Glu^{45} in the B pocket of the HLA-B27 antigen binding site (*3*). In HLA-A2.1, this pocket is hydrophobic and is capable of accommodating a Leu or an Ile side chain

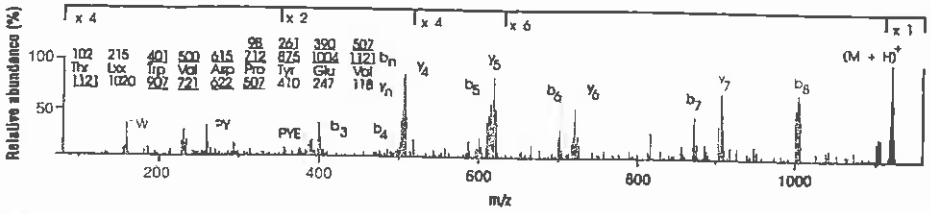


Fig. 2. CAD mass spectrum of peptide $(M+H)^+$ ions at *m/z* 1121 using 100 to 300 fmol of material derived from 2×10^8 cells. Predicted masses for fragment ions of types b and y (*17*) are shown, respectively, above and below the deduced amino acid sequence. Ions observed in the spectrum are underlined. Subtraction of *m/z* values for any two fragments that differ by a single amino acid generates a value that specifies the mass and thus the identity of the extra residue in the larger fragment. Residues 5 to 9 were deduced from fragments of type b, whereas residues 3 to 5 were defined by fragments of type y. CAD spectra recorded on additional aliquots of the sample mixture that had either been subjected to a single cycle of Edman degradation (*17*) or treated with methanolic-HCl to convert peptides to their corresponding methyl esters (*17*) confirmed the order of the first two amino acids as Thr and Leu/Ile, and the presence and location of the two acidic residues in the peptide. Total sample consumed in these analyses corresponded to the extract from 2×10^9 cells. Since Ile and Leu are of identical mass, they cannot be differentiated on the triple quadrupole instrument and are specified here as Lxx.

Table 1. Partial listing of data for peptides extracted from HLA-A2.1. HLA-A2.1 was purified and acid extracts prepared from the cell lines JY, C1R-A2.1, and C1R as outlined in Fig. 1. Aliquots corresponding to 10^8 cells were analyzed by microcapillary HPLC-electrospray ionization–mass spectrometry. Signals represent measured ion current. Synthetic peptides corresponding to sequences 8 and 14, doped at 1 pmol into samples, gave signals of 60×10^5 to 70×10^5, whereas background noise gave a signal of 1×10^5. Sequences were deduced as in Fig. 2. Single letter abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

Peptide	$(M+H)^+$ (*m/z*)	JY		C1R-A2.1		C1R	Sequence*	Database match
		Signal ($\times 10^5$)	Yield (fmol)	Signal ($\times 10^5$)	Yield (fmol)	Signal ($\times 10^5$)		
1	786	9.1	137	2.0	30	-†	SXPSGGXGV	
2	868	2.7	41	7.9	119	-		
3	871	2.9	44	6.1	98	-		
4	898	11.9	179	15.8	237	-	XXDVPTAAV	LLDVPTAAV
5	903	4.7	71	11.8	177	-		
6	930	4.9	74	-	-	-	GXVPFXVSV	
7	945	11.8	177	3.1	47	-		
8	954	11.9	179	2.5	38	-	SXXPAXVEX	SLLPAIVEL
9	971	8.5	128	4.4	66	-		
10	994	6.2	93	-	-	-		
11	999‡	3.5	53	-	-	-	SXXVRAXEV	
12	1018‡	4.3	65	2.5	38	-		
13	1037‡	10.0	150	7.3	110	-	KXNEPVXXX	
14	1038‡	39.4	591	8.1	122	-	YXXPAXVHX	YLLPAIVHI
15	1064	8.1	122	3.5	53	-		
16	1095	5.2	78	-	-	-		
17	1098	2.3	35	-	-	-		
18	1115	11.4	171	2.9	44	-		
19	1121	6.4	320§	2.0	100§	-	TLWVDPYEV	TLWVDPYEV

*Leu and Ile are not distinguishable by tandem mass spectrometry, and are designated X. †Signal was at or below background. ‡Observed as a double charged, $(M+H)^{2+}$ ion at half of this *m/z* value. §Signal obtained from 1 pmol of this synthetic standard added into sample was 2×10^6.

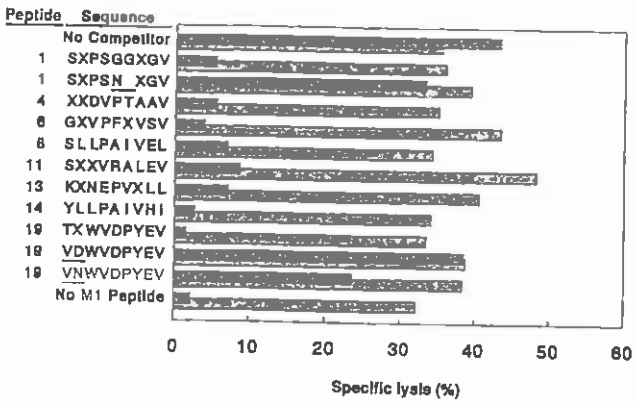


Fig. 3. Synthetic peptides corresponding to sequenced peptides bind to HLA-A2.1. Synthetic peptides are referenced by the numbers in Table 1. JY cells were incubated with the $M1_{57-68}$ peptide of influenza A (8 μg/ml), in the presence or absence of one of the synthetic peptides (200 μg/ml) as a competitor, in a total volume of 250 μl of RPMI 1640 containing 5% newborn calf serum and 50 μCi of ^{51}Cr for 3 hours. Washed target cells (2×10^3) were incubated in 200 μl with either an M1 peptide–specific, HLA-A2.1–restricted CTL clone 5.2-11 (*31*) (filled bars) or an HLA-A2.1–specific human alloreactive CTL line (stippled bars) at an effector to target ratio of 1.25:1. Data are the means of duplicate samples with an SD of less than 5% and are representative of at least three independent experiments.

(*2*). In addition, positively charged residues predominate at position 9 in peptides associated with HLA-B27, and this residue was suggested to interact with Asp^{116} in the F pocket. In HLA-A2.1, position 116 is a Tyr, rendering the F pocket smaller and more hydrophobic. Based on these considerations and the similarity of the electron density corresponding to peptide in these two molecules (*2, 3*), we suggest that interactions of Leu or Ile at position 2 with the B pocket, an aliphatic side chain at position 9 with the F pocket, and perhaps a nine-residue length are the primary determinants of peptide binding to HLA-A2.1.

The eight completely sequenced peptides in Table 1 were synthesized in order to demonstrate that the peptides identified actually bound to HLA-A2.1. When no complete match to a sequence in the protein or gene databases was found (below), a mixture of Leu and Ile was incorporated at positions specified by X. Cells of the HLA-A2.1–expressing line JY were incubated with the $M1_{57-68}$ peptide of influenza A virus in the absence or presence of a 16- to 20-fold molar excess of one of the competitor peptides and assayed for their ability to be lysed by an M1 peptide–specific, HLA-A2.1–restricted CTL clone. All eight peptides inhibited recognition under these circumstances but had no effect on recognition by a human alloreactive CTL line that is specific for HLA-A2.1 (Fig. 3). Thus, these synthetic peptides bind to HLA-A2.1 in vitro, which confirms their association with HLA-A2.1 in intact cells. Additional peptides were synthesized in which two Gly residues at positions 5 and 6 in peptide 1 were replaced by a single Asn to produce an octamer, or the Thr and Leu in positions 1 and 2 of peptide 19 were substituted with Val and Asn or Val and Asp. None of these peptides blocked recognition, indicating that a nine-residue length, the presence of Leu or Ile at position 2, and perhaps a polar residue at position 1 are important features for peptide binding.

To identify possible precursor proteins for these peptides, gene and protein sequence databases were searched in a manner that allowed X residues to be assigned as either Leu or Ile. Only four peptides gave 100% matches with proteins in the library (Table 1). All of the sequences contain Leu at position 2, although Ile is often found at other positions. A match for peptide 8 occurs in the 65-kD α and β regulatory subunits of human protein phosphatase 2A and in a highly homologous human transformation-associated protein, p61. Both are cytoplasmic proteins and p61 represents only 0.01 to 0.02% of the total cellular protein (*20*). A match for peptide 14 appears in the human nuclear protein p68, which exhibits RNA-dependent adenosine triphosphatase activity and RNA unwinding activity (*21*). A different sequence from this protein is found in a peptide bound to HLA-B27 (*9*). A match for peptide 19 is found in TIS21, a protein of unknown function that is a member of the primary response group of genes induced by growth factors (*22*). Finally, a match for peptide 4 is found in the human protein IP-30, the expression of which is induced by gamma interferon (*23*). This precursor sequence is unique among all naturally processed peptides described to date in that it is located in the signal peptide domain of the protein. This peptide is also bound to HLA-A2.1 molecules expressed on CEMx721.174(T2) (*24*), a mutant cell line defective in the normal antigen processing pathway (*6, 7*). This result suggests a second pathway through which peptides can enter the endoplasmic reticulum and associate with MHC class I molecules.

Peptides derived from the processing of normal cellular proteins located in the cytoplasm have been implicated in the correct folding and intracellular transport of class I molecules (*6, 7*) and in the epitopes recognized by alloreactive T cells (*25–28*). The use of tandem mass spectrometry for the direct analysis of these peptides has enhanced our knowledge of peptide-MHC interactions and the cellular processes that regulate formation of these complexes and should allow the direct identification of peptides that are alloreactive T cell epitopes. Peptides associated with class II MHC molecules can also be characterized by the above approach (*29*). This same methodology should also facilitate identification of peptide antigens associated with viral infection, cellular transformation, and autoimmunity.

REFERENCES AND NOTES

1. A. Townsend and H. Bodmer, *Annu. Rev. Immunol.* **7**, 601 (1989).
2. M. A. Saper, P. J. Bjorkman, D. C. Wiley, *J. Mol. Biol.* **219**, 277 (1991).
3. D. R. Madden, J. C. Gorga, J. L. Strominger, D. C. Wiley, *Nature* **353**, 321 (1991).
4. L. A. Morrison *et al.*, *J. Exp. Med.* **163**, 903 (1986).
5. M. W. Moore *et al.*, *Cell* **54**, 777 (1988).
6. V. Cerundolo *et al.*, *Nature* **345**, 449 (1990).
7. T. Spies *et al.*, *ibid.* **348**, 744 (1990).
8. K. Falk, O. Rotzschke, S. Stevanovic, G. Jung, H.-G. Rammensee, *ibid.* **351**, 290 (1991).
9. T. S. Jardetsky, W. S. Lane, R. A. Robinson, D. R. Madden, D. C. Wiley, *ibid.* **353**, 326 (1991).
10. G. M. van Bleek and S. G. Nathenson, *ibid.* **348**, 213 (1990).
11. O. Rotzschke *et al.*, *ibid.*, p. 252.
12. K. Falk *et al.*, *J. Exp. Med.* **174**, 425 (1991).
13. W. J. Storkus *et al.*, *J. Immunol.* **138**, 1657 (1987).
14. D. F. Hunt *et al.*, in *Techniques in Protein Chemistry II*, J. J. Villafranca, Ed. (Academic Press, New York, 1991), pp. 441–454.
15. S. Demotz, H. M. Grey, A. Sette, *Science* **249**, 1028 (1990).
16. E. R. Christnick, M. A. Luscher, B. H. Barber, D. B. Williams, *Nature* **352**, 67 (1991).
17. D. F. Hunt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6233 (1986).
18. D. F. Hunt, H. Michel, J. Shabanowitz, A. Cox, unpublished data.
19. V. Cerundolo *et al.*, *Eur. J. Immunol.* **21**, 2069 (1991).
20. G. Walter, F. Ferre, O. Espiritu, A. Carbone-Wiley, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8669 (1989).
21. H. Hirling, M. Scheffner, T. Restle, H. Stahl, *Nature* **339**, 562 (1989).
22. B. S. Fletcher *et al.*, *J. Biol. Chem.* **266**, 14511 (1991).
23. A. D. Luster, R. L. Weinshank, R. Feinman, J. V. Ravetch, *ibid.* **263**, 12036 (1988).
24. R. A. Henderson *et al.*, *Science* 255, 1264 (1992).
25. W. R. Heath, L. A. Sherman, *Eur. J. Immunol.* **21**, 153 (1991)
26. W. R. Heath *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 5101 (1991).
27. S. Man, R. D. Salter, V. H. Engelhard, *Int. Immunol.*, in press.
28. D. B. Crumpacker, J. Alexander, P. Cresswell, V. H. Engelhard, in preparation.
29. D. F. Hunt *et al.*, in preparation.
30. P. Parham and F. M. Brodsky, *Hum. Immunol.* **3**, 277 (1981).
31. V. H. Engelhard, E. Lacy, J. P. Ridge, *J. Immunol.* **146**, 1226 (1991).

32. Supported by USPHS grants **GM37537** (to D.F.H.) and AI20963 (to V.H.E.) and by National Science Foundation predoctoral fellowship RCD 9154717 (to N.S.).

5 November 1991; accepted 22 January 1992

35

To confirm that the current results from expression of CFTRΔF508 and to determine whether the ΔF508 mutation alters the properties of CFTR, we measured single-channel currents using excised, inside-out patches of membrane. Figure 4*d* shows that, just as with wild-type CFTR[3,4,7], there is little basal channel activity in excised patches. Addition of ATP to the cytosolic surface caused no activation, but a combination of ATP plus the catalytic subunit of cAMP-dependent protein kinase (protein kinase A; PKA) activated multiple channels. The current–voltage relationship was linear and the single-channel current at −100 mV was −1.1 ± 0.0 pA ($n = 6$), compared to −1.1 ± 0.0 pA ($n = 7$) for wild-type channels under similar conditions. When ATP and PKA were removed, activity returned to basal values. Despite these apparently normal biophysical properties and the normal regulation by cAMP (in the whole cell) and PKA (in excised patches), the channels had a reduced probability of being in the open state (P_o) at 1 mM ATP: P_o was 0.34 ± 0.02 ($n = 14$) for wild-type CFTR and 0.13 ± 0.01 ($n = 4$) for CFTRΔF508. The reduced P_o of CFTRΔF508 is similar to the value previously reported[19].

Our findings reconcile previously incongruent observations. Earlier, we had failed to observe cAMP-regulated Cl^- channels when we expressed CFTRΔF508 in airway epithelia and in heterologous cells[16,17]. Moreover, several reports indicated that the apical membrane of cystic fibrosis-affected epithelia lacks a cAMP-regulated Cl^- conductance (reviewed in ref. 12). Yet CFTRΔF508 could generate cAMP-regulated Cl^- channels when expressed in *Xenopus* oocytes[18], Vero cells[19], and Sf9 insect cells[20]. Our data suggest that the reduced temperature used to culture *Xenopus* oocytes and Sf9 insect cells may be at least partly responsible for the detection of the functional protein in the plasma membrane. Although functional channels were detected in Vero cells grown at 37 °C, CFTRΔF508 was produced by a very high-level expression system[19]. The data indicated that there might be a defect in protein transport[19], so under those expression conditions the channel function could have represented a small fraction of the mutant protein that escaped the cellular quality control mechanisms to reach the plasma membrane, where it could generate a significant cAMP-stimulated Cl^- current.

Our results suggest that CFTRΔF508 is a temperature-sensitive mutant. Temperature sensitivity of protein processing has been observed in other systems. For example, temperature-sensitive mutants of the vesicular stomatitis virus G protein have shown that oligomerization and movement in the cell can be promoted by reduced temperature[21].

What are the implications for therapy? It is difficult to imagine any clinically acceptable therapeutic approach based on significant, repeated reductions in whole-body temperature. Nevertheless, the major cause of morbidity and mortality in cystic fibrosis is lung disease, and inhalation of cold air during hyperventilation can reduce airway temperature to values approaching 30 °C (refs 24, 25). Thus, it is possible that the processing of mutant CFTR might be altered *in vivo* by a reduction of temperature. Irrespective of the feasibility of such a manoeuvre, our results suggest that therapy designed to deliver mutant protein to the plasma membrane may be possible. □

Received 23 April; accepted 13 July 1992.

1. Anderson, M. P. *et al. Science* **253**, 202–205 (1991).
2. Bear, C. E. *et al. Cell* **68**, 809–818 (1992).
3. Tabcharani, J. A., Chang, X.-B., Riordan, J. R. & Hanrahan, J. W. *Nature* **352**, 628–631 (1991).
4. Berger, H. A. *et al. J. clin. Invest.* **88**, 1422–1431 (1991).
5. Rich, D. P. *et al. Science* **253**, 205–207 (1991).
6. Cheng, S. H. *et al. Cell* **66**, 1027–1036 (1991).
7. Anderson, M. P. *et al. Cell* **67**, 775–784 (1991).
8. Rommens, J. M. *et al. Science* **245**, 1059–1065 (1989).
9. Riordan, J. R. *et al. Science* **245**, 1066–1073 (1989).
10. Kerem, B. *et al. Science* **245**, 1073–1080 (1989).
11. Boat, T. F., Welsh, M. J. & Beaudet, A. L. in *The Metabolic Basis of Inherited Disease* Vol. 6 (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 2649–2680 (McGraw-Hill, New York, 1989).
12. Quinton, P. M. *FASEB J.* **4**, 2709–2717 (1990).
13. Cheng, S. H. *et al. Cell* **63**, 827–834 (1990).
14. Gregory, R. J. *et al. Molec. cell. Biol.* **11**, 3886–3893 (1991).
15. Denning, G. M., Ostedgaard, L. S. & Welsh, M. J. *J. Cell Biol.* **118**, 551–559 (1992).
16. Rich, D. P. *et al. Nature* **347**, 358–363 (1990).
17. Anderson, M. P., Rich, D. R., Gregory, R. J., Smith, A. E. & Welsh, M. J. *Science* **251**, 679–682 (1991).
18. Drumm, M. L. *et al. Science* **254**, 1797–1799 (1991).
19. Dalemans, W. *et al. Nature* **354**, 524–528 (1991).
20. Bear, C. E., Jensen, T. J. & Riordan, J. R. *Biophys. J.* **61**, A127 (1992).
21. Machamer, C. E. & Rose, J. K. *J. biol. Chem.* **263**, 5955–5960 (1988).
22. Ljunggren, H.-G. *et al. Nature* **346**, 476–480 (1990).
23. Kartner, N. *et al. Cell* **64**, 681–691 (1991).
24. Gilbert, I. A., Fouke, J. A. & McFadden, E. R. Jr *J. appl. Physiol.* **63**, 1681–1691 (1987).
25. Deal, E. C. Jr, McFadden, E. R. Jr, Ingram, R. H. Jr & Jaeger, J. J. *J. appl. Physiol.* **46**, 484–490 (1979).

ACKNOWLEDGEMENTS. G.M.D. and M.P.A. contributed equally to this work. We thank D. Ries for helping with the immunoprecipitation and protein phosphorylation; L. DeBerg and P. Karp for culturing cells; T. Mayhew for typing the manuscript; and R. Gregory, Seng Cheng and our colleagues for discussion. This work was supported in part by the Howard Hughes Medical Institute and by grants from the National Heart, Lung and Blood Institute and the Cystic Fibrosis Foundation.

Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size

Roman M. Chicz, Robert G. Urban, William S. Lane*, Joan C. Gorga†, Lawrence J. Stern, Dario A. A. Vignali & Jack L. Strominger

Department of Biochemistry and Molecular Biology, and *Microchemistry Facility, Harvard University, Cambridge, Massachusetts 02138, USA

PEPTIDES bound to class I molecules are 8–10 amino acids long, and possess a binding motif representative of peptides that bind to a given class I allele[1–4]. In the only published study of naturally processed peptides bound to class II molecules (mouse I-A^b and I-E^b), these peptides were longer (13–17 amino acids) and had heterogenous carboxy terminals but precise amino-terminal truncations[5]. Here we report the characterization of acid-eluted peptides bound to HLA-DR1 by high-performance liquid chromatography, mass spectrometry and microsequencing analyses. The relative molecular masses of the peptides varied between 1,602 and 2,996 (13–25 residues), the most abundant individual M_r values being between 1,700 and 1,800, corresponding to an average peptide length of 15 residues. Complete sequence data were obtained for twenty peptides derived from five epitopes, of which all but one were from self proteins. These peptides represented sets nested at both the N- and C-terminal ends. Binding experiments confirmed that all of the isolated peptides had high affinity for the groove of DR1. Alignment of the peptides bound to HLA-DR1 and the sequences of 35 known HLA-DR1-binding peptides revealed a putative motif. Although peptides bound to class II molecules may have some related features (due to the nonpolymorphic HLA-DR α-chain), accounting for degenerate binding to different alleles[6], particular amino acids in the HLA-DR β-chains presumably define allelic specificity of peptide binding.

The HLA-DR1 used in this study was solubilized by papain treatment to enable the material to be used both for crystallographic[7] and bound-peptide analyses. The peptides bound to DR1 were acid-extracted and fractionated using microbore reversed-phase high-pressure liquid chromatography (RP-HPLC) (Fig. 1*a*). The absence of any detectable peptide material after a second extraction/RP-HPLC separation verified quantitative peptide extraction. Amino-acid analysis (ABI 420A/130A derivatizer/HPLC) of extracted peptide pools indicated that the yield was 70–80%, assuming total occupancy of purified DR1

† Present address: Division of Immunogenetics, Department of Pediatrics, University of Pittsburgh School of Medicine, Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA

© 1992 Nature Publishing Group

TABLE 1 Naturally processed HLA-DR1-bound peptides

Protein source	Peptide sequence	Length	Residue	Peak	Calculated M_r	Observed M_r	Yield
HLA-A2	VGSDWRFLRGYHQYAYDG	18	103–120	5	2,190.4	2,190.4	39.5
	VGSDWRFLRGYHQYA	15	103–117	4	1,855.0	1,854.4	907.5
	VGSDWRFLRGYHQY	14	103–116	4	1,784.0	1,783.6	53.3
	GSDWRFLRGYHQYA	14	104–117	3	1,755.3	1,755.2	96.5
	SDWRFLRGYHQYA	13	105–117	3	1,698.2	1,698.8	48.8
							1,145.6
Invariant chain (Ii)	LRKPPKPVSKMRMATPLLMQALPMG	25	96–120	9	2,733.5	2,734.5	40.5
	LPKPPKPVSKMRMATPLLMQALPM	24	96–119	9	2,676.4	2,675.9	80.8
	PKPPKPVSKMRMATPLLMQALPMG	24	97–120	7	2,620.2	2,619.7	91.5
	LPKPPKPVSKMRMATPLLMQALP	23	96–118	7	2,545.2	2,544.5	112.2
	PKPPKPVSKMRMATPLLMQALPM	23	97–119	8	2,563.2	2,562.3	145.0
	KPPKPVSKMRMATPLLMQALPM	22	98–119	8	2,466.1	2,465.8	101.5
	PKPPKPVSKMRMATPLLMQALP	22	97–118	6	2,432.0	2,431.7	72.5
	KPPKPVSKMRMATPLLMQALP	21	98–118	6	2,334.9	2,334.2	31.6
	PPKPVSKMRMATPLLMQALP	20	99–118	7	2,206.7	2,207.4	89.8
	KMRMATPLLMQALPM	15	105–119	9	1,732.2	1,731.9	178.5
	KMRMATPLLMQALP	14	105–118	7	1,601.0	1,600.2	162.0
							1,105.9
$(Na^+ + K^+)$ATPase	IPADLRIISANGCKVDNS	18	199–216	3	1,886.6	1,885.5	48.8
							48.8
Transferrin receptor	RVEYHFLSPYVSPKESP	17	680–696	4	2,035.3	2,036.8	30.3
							30.3
Bovine fetuin	YKHTLNQIDSVKVWPRRPT	19	56–74	1	2,338.7	2,338.5	32.5
	YKHTLNQIDSVKVWPRRP	18	56–73	2	2,237.6	2,236.5	69.0
							101.5

Amino-acid sequences (single-letter code) and mass determinations are shown for peptides isolated from HLA-DR1. Peptides were isolated as described for Fig. 1 and sequenced using an ABI 477A protein sequencer[8]. Collected fractions were as labelled in Fig. 1*a*. The peptide M_rs were calculated from the amino-acid sequences. Mass spectra were collected on a Finnigan-MAT (San Jose, California) TSQ 700 triple quadrupole mass spectrometer equipped with an electrospray ion source. The electrospray needle was operated with a voltage differential of 3 to 4 keV using a 50% methanol/0.5% acetic acid sample mixture delivered at 1 μl min^{-1} by infusion injection. Observed M_rs were determined from the resulting spectra by computer-aided deconvolution of multiply charged m/z peaks[37]. Yield refers to initial sequence levels of individual peptides measured in pmol, with the total yield of peptides from a single protein source summed and listed below the individual peptide yields.

with a molar equivalent of bound peptides corresponding to the size distribution determined by mass spectrometry (Fig. 1*b*). The RP-HPLC profiles obtained from DR1 extractions of multiple independent preparations were reproducible and profiles from either detergent-soluble or papain-solubilized DR1 were equivalent. To confirm that the peptides were in fact identical in detergent-soluble and papain-digested DR1, mass spectrometry and Edman sequencing analyses were done and revealed identical masses and sequences for analogous fractions from the two preparations.

Matrix-assisted laser desorption mass spectrometry (MALD-MS) was used to identify 111 species of unique mass in the eluted peptide pool of DR1 with an average size of 18 and a mode of 15 residues (Fig. 1*b*). Over 500 additional mass species in the M_r range of 13–25 residues were detected; however, the signal was too weak to assign individual masses with confidence. Multiple species of varying mass were detected in fractions corresponding to single RP-HPLC peaks indicating coelution of peptides. To characterize these peptides further, samples were analysed in parallel on a triple quadrupole mass spectrometer equipped with an electrospray ion source and by automated Edman degradation microsequencing[8]. Combining these two techniques allows crucial verification of both the N- and C-terminal amino acids of peptides contained in single fractions. The sequence and mass data acquired for 20 peptides isolated from DR1 are listed in Table 1. All the identified peptides aligned with complete identity to regions of proteins stored in the SWISS-PROT database[9].

Of the 20 sequenced DR1-bound peptides, 16 were derived from the self proteins HLA-A2 and class II-associated invariant chain (Ii), which represented at least 26% of the total extracted peptide mass. These isolated peptides varied in length and were truncated at both the N and C termini, suggesting that: (1) antigen processing occurs from both ends after binding to DR1; or (2) class II molecules bind antigen from a pool of randomly generated peptides. The yields from the peptide microsequencing indicated that HLA-A2 (Fig. 1*a*, peaks 3–5) and Ii (peaks 6–9) each represent at least 13% of the total DR1-bound peptides. Allorecognition of peptides derived from the class I molecule HLA-B7 presented by the class II molecule HLA-DR11.1 indicates that presentation of class I-derived peptides by class II may be a common phenomenon[10].

The other four peptides identified were derived from two self proteins, transferrin receptor and the $(Na^+ + K^+)$ATPase, and one exogenous protein, bovine serum fetuin (Table 1). Each of these peptides occupied only 0.3–0.6% of the total DR1 population, significantly less than either the HLA-A2 or Ii peptides. Several groups have demonstrated that class II molecules *en route* to the cell surface intersect the pathway of incoming endocytic vesicles[11–13]. Both recycling membrane proteins and endocytosed exogenous protein travel this common pathway. Hence the HLA-A2, transferrin receptor, $(Na^+ + K^+)$ATPase and bovine fetuin-derived peptides would all encounter DR1 in a similar manner. Ii associates with nascent class II molecules in the endoplasmic reticulum[14], preventing antigen binding until the class II/Ii complex arrives at an endocytic compartment[15], where Ii undergoes proteolysis[16,17], thus allowing peptide binding to proceed. Presumably the Ii peptides bound to DR1 were generated at this step.

Synthetic peptides corresponding to five of the peptides shown in Table 1 were made and characterized for their relative binding affinity to DR1. The influenza A haemagglutinin peptide (HA) 307–319 has been previously described as a high-affinity, HLA-DR1-restricted peptide[18,19] and was thus chosen as the control

37

© 1992 Nature Publishing Group

peptide. 'Empty' DR1 purified from recombinant insect cells was used in the binding experiments because of its higher binding capacity and 10-fold faster association kinetics than DR1 isolated from human cells[20]. All the synthetic peptides were found to compete well (inhibition constant, $K_i < 100$ nM) against the HA peptide (Table 2). At first approximation, the Ii 106–119 peptide had the highest affinity of all the competitor peptides measured, equivalent to that determined for the control HA peptide. In addition to the K_i determinations, these peptides were able to confer resistance to SDS-induced chain dissociation of empty DR1 when analysed by SDS–polyacrylamide gel electrophoresis (PAGE), indicative of stable peptide binding[21–23]. Neither of the two control peptides, β_2-microglobulin 52–64 nor Ii 96–110, was able either to confer resistance to SDS-induced chain dissociation of DR1 or compete with HA 307–319 for binding to DR1; both of these peptides lack the putative binding motif reported here.

A putative DR1-binding motif based on the sequence alignments of the core epitopes (the minimum length) of naturally processed peptides is shown in Table 3. The key residues proposed in the motif are: a positively charged group is located at the first position referred to here as the index position for orientation (I); a hydrogen-bond donor at I+5; and a hydrophobic residue at I+9. A hydrophobic residue is also often found at I+1 and/or I-1. Every naturally processed peptide sequenced from DR1 conforms to this motif (with the exception of the HLA-A2 peptide 103–116, which lacks residue I+9). Because the putative motif is not placed in a defined position with respect to the first amino acid and because of the irregular length of bound peptides, it is impossible to deduce a motif from sequencing of peptide pools, as was done for class I

TABLE 2 Peptide binding to HLA-DR1

Peptide	Length (amino-acid residues)	K_i against HA$^-$ 307–319 (nM)	SDS resistance
HLA-A2 103–117	15	49 ± 3	+
Ii 105–119	15	<10	+
Ii 96–119	24	33 ± 5	+
$(Na^+ + K^+)$ATPase 199–216	18	68 ± 9	+
Transferrin receptor 680–696	17	<10	+
Bovine fetuin 56–74	19	66 ± 18	+
HA 307–319	14	<10	+
Ii 96–110	15	$>10^4$	–
β_2m 52–64	13	$>10^4$	–

The first six entries correspond to peptides found associated with HLA-DR1 and the sequences are shown in Table 1. Two control peptides were also tested: β_2m 52–64 (SDLSFSKDWSFYL) is from human β_2-microglobulin, and Ii 96–110 (LPKPPKPVSKMRMAT) is a truncated version of the longest invariant chain-derived peptide isolated from HLA-DR1. Peptides were synthesized using solid-phase Fmoc chemistry, deprotected and cleaved using standard methods, then purified by RP-HPLC. Purified peptides were analysed by mass spectrometry and concentrations were determined by quantitative ninhydrin analysis. Inhibition constants (K_i) were measured as the concentration of test peptide that inhibited 50% of the ^{125}I-labelled HA 307–319 binding to empty HLA-DR1 produced in Sf9 insect cells[20]. HA 307–319 was labelled using Na^{125}I and chloramine-T and isolated by gel filtration. Specific activity, determined by BCA assay (Pierce) and gamma counting, was 26,000 c.p.m. pmol^{-1}. Labelled peptide (10 nM) and 10 nM purified HLA-DR1 were mixed with 10 different concentrations (10 nM to 10 μM) of synthetic cold competitor peptide in phosphate-buffered saline, pH 7.2, containing 1 mM EDTA, 1 mM PMSF, 0.1 mM iodoacetamide, and 3 mM NaN_3, and incubated at 37 °C for 85 h. Free and bound peptide were separated by native gel electrophoresis[33] and bound radioactivity was quantitated using a Fujix Imaging plate analyser (BAS 2000) after 4-h exposures on the phosphoimaging plates. Per cent inhibition was calculated as the ratio of background-corrected radioactivity in the sample to background-corrected radioactivity in a parallel sample containing no competitor peptide. Under these conditions, K_i measurements < 10 nM could not be accurately determined. The ability of the synthetic peptides to confer resistance to SDS-induced chain dissociation of HLA-DR1 produced in insect cells was determined as described[20]. Briefly, 20 μM HLA-DR1 was incubated with fivefold excess of synthetic peptide at 37 °C for 85 h, in PBS (pH 7.2) with the protease inhibitor mixture described above. After incubation, the samples were analysed by SDS–PAGE with and without boiling before loading. Peptides that prevented SDS-induced chain dissociation are indicated by plus signs and those that did not by minuses.

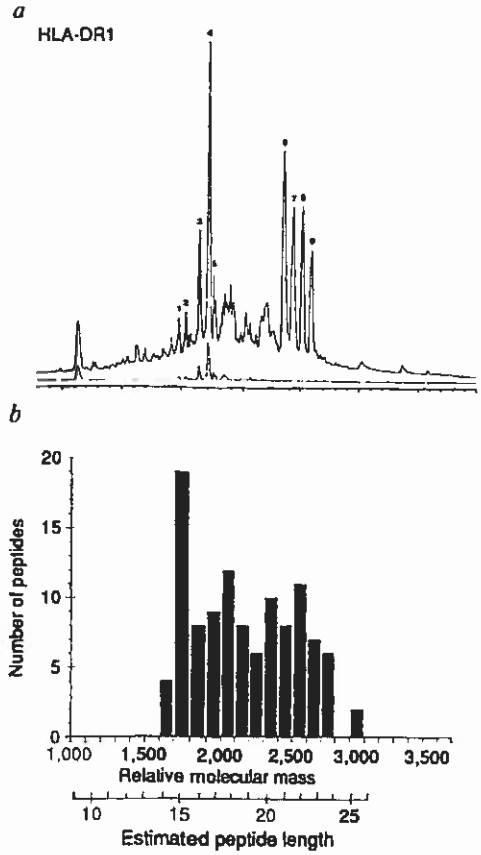


FIG. 1 Characterization of the peptide pools extracted from papain-digested HLA-DR1. *a*, The chromatogram represents the peptide repertoire as detected by ultraviolet absorbance. The ultraviolet absorbance for both 210 nm and 277 nm is shown at a full-scale absorbance of 0.500 with a retention window between 16 min and 90 min (each mark represents 2 min). *b*, Size distribution of isolated HLA-DR1-bound peptides determined by MALD-MS. The determined peptide masses in groups of 100 mass units were plotted against the number of isolated peptides identified by mass spectrometry. Peptide length was calculated by dividing the experimental mass by an average amino-acid M_r of 118. Although the weighted average for the length of bound peptide is ~18, a preference for DR1 binding of 15-amino-acid peptides is observed.

METHODS. DR1 was purified from the Epstein–Barr virus-transformed human B lymphoblastoid line LG-2. Cell growth, collection conditions and protein purification were as previously described[7,20]. Protein concentration was determined by BCA assay (Pierce). All protein samples were next purified by size-exclusion HPLC on a TSK-250 column (300 × 7.8 mm) in 25 mM MES, 150 mM NaCl, pH 6.5, at a flow rate of 1 ml min^{-1}. DR1 (1 mg) was spin-concentrated to 50 μl using a Centricon 10 microconcentrator. Acetic acid (1 ml, 10%) was added and the sample heated at 70 °C for 15 min then recentrifuged. The flow-through provided the acid-eluted peptide pool. An additional volume of acetic acid (1 ml, 10%) was added to the retentate and spin-concentrated to give a final acid-eluted volume of 2 ml. The peptide pool was concentrated to 50 μl in a Savant Speed-Vac before HPLC separation. Peptides were separated on a Vydac microbore C-18 column as previously described[3]. Chromatographic analysis was monitored at multiple ultraviolet wavelengths to permit spectrophotometric evaluation of purified species before mass and sequence analyses. Collected fractions from the RP-HPLC separation of the HLA-DR1 peptide pool were analysed on a Finnigan-MAT LASERMAT MALD-MS using a sinapinic acid matrix, an accelerating voltage of 20 keV and a nitrogen laser (337 nm). Sample calibration was achieved by static fast-atom bombardment mass spectrometry of strong signal response samples.

© 1992 Nature Publishing Group

TABLE 3 Putative HLA-DR1 peptide-binding motif

	Protein source	Peptide sequence	Length	Position	Ref.
(a)	HLA-A2	SDWRFLRGYHQYA	13	105–117	This study
	Invariant chain	KMRMATPLLMQALP	14	105–118	
	$(Na^+ + K^+)$ ATPase	IPADLRIISANGCKVDNS	18	199–216	
	Transferrin receptor	RVEYHFLSPYVSPKESP	17	680–696	
	Bovine fetuin	YGHTLNQIDSVKVWPRRP	18	56–73	
(b)	HEL	KVFGRCELAAAMKRHGLD	18	1–18	6
		RNRCKGTDVQAWIRGCRL	18	112–129	6
	β_2m	HFPHIEIQMLKNGKKI	16	31–46	6
	PLA_2	NELGRFKHTDACCRTH	16	19–34	6
		SKPKVYQWFDLRKY	14	115–128	6
	NASE	ATSTKKLHKEPATLIKAIDG	20	1–20	6
		PATLIKAIDGDTVKLMYKGQ	20	11–30	6
		DRVKLMYKGQPMTFRLLVD	20	21–40	6
		VAYVYKPNNTHEQHLSKSEA	20	111–130	6
	HIV p13	QKQEPIDKELYPLTSL	16	97–112	6
	HIV p17	GARASVLSGGELDKWE	16	1–16	6
	Influenza HA	PTLYQNVGTYVSVGTSTLNK	20	187–206	6
	Influenza HA	PKYVKQNTLKLAT	13	307–319	17
	P. falciparum p190	LKKLVFGYRKPLDNI	15	249–263	25
	P. falciparum CS	KHIEQYLKKIKNS	13	329–341	27
	Chicken OVA	DVFKELKVHHANENIF	16	15–30	6
	DR1 β-chain	GDTRPRFLWQLKFECHFFNG	20	1–20	28
		TERVRLLERCIYNQEESVRFDS	22	21–42	28
		DLLEQRRAAVDTYCRHNYGVGESFT	25	66–90	28
	p Cyt *c*	KAERADLIAYLKQATAK	17	88–104	6
	Myelin basic protein	GRTQDENPVVHFFKNIVTPRTPPP	24	75–98	6
(c)	Influenza matrix	FLKAEIAQRLEDV	13	19–31	6
	HIV p17	EQILGQLQPSLQTGSE	16	57–72	6
	β_2m	IQVYSRHPPENGKPNI	16	7–22	6
	PLA_2	INTKCYKLEHPVTGCG	16	85–100	6
	P. falciparum p190	YKLNFYFDLLRAKL	14	211–224	25
		IDTLKKNENIKEL	13	338–350	25
	DR1 β-chain	DVGEYRAVTELGRPDAEYWN	20	43–62	28
	HIV p17	ERFAVNPGLLETSEGC	16	41–56	6
	HEL	DNYRGYSLGNWVCAAKFESNFTQ	23	20–42	6
	NASE	EALVRQGLAKVAYVYKPNNT	20	101–120	6
	HIV p25	PIVQNLQGQMVHQAIS	16	1–16	6
		SALSEGATPQDLNTML	16	41–56	6
	β_2m	SFYILAHTEFTPTETD	16	61–76	6
	PLA_2	KMYFNLINTKCYKLEH	16	79–94	6

The putative HLA-DR1 peptide-binding motif includes three key amino acids; a positive charge at I, a hydrogen-bond donor at I+5, and a hydrophobic residue at I+9. The peptides are listed such that I, I+5, and I+9 are vertically aligned and underlined. *a*, The peptide core epitopes identified in Table 1; *b*, those peptides known to bind HLA-DR1 which totally conform to the described motif, and *c*, lists those peptides known to bind HLA-DR1 with either one key motif residue shifted or substituted at I by a large hydrophobic residue. These residues in *c* are in bold type.

molecules[24]. The Ii 96–110 peptide has the I and I+5 motif residues in its sequence, but is missing eight additional amino acids found in Ii 105–118 (Table 3*c*).

A sequence comparison of 35 previously described DR1-binding synthetic peptides[6,25–28] also supports this motif with 21 of the peptides (60%) having the precise motif, nine of the peptides (30%) containing a single shift at either I or I+9 and the remaining five (10%) having a single substitution at I (Table 3). In the latter, a positive charge at I is always replaced by a large hydrophobic residue (Table 3*c*); a pocket has been described in class I molecules that can accommodate this precise substitution[29]. Contributions by the other eight amino acids in the motif or the length of the peptide have not been fully evaluated and may compensate for shifted/missing residues in those peptides that bind. Evaluation of the remaining 117 peptides that do not bind DR1 cited in those studies indicates that 99 (85%) of these peptides do not contain the proposed DR1 motif. Of the remaining 18 peptides (15%) containing the motif that do not bind to DR1, 6 (5%) bind to other DR alleles; the remaining 12 peptides may have unfavourable interactions at other positions that interfere with binding. Because all HLA-DR alleles contain the same nonpolymorphic DR α-chain, the residues pointing inward from the α1 helix and half of the floor of the groove are invariant[30]. Thus some features of the motif for binding to DR1 might be expected to be shared by all DR alleles. Using T-cell recognition and/or binding of a limited number of analogue peptides, several studies have reported specific residues, primarily hydrophobic residues, at precise locations that are critical for peptide binding to DR1[19,26,31–34]. The peptides described in those studies conform in part to the putative motif described herein. But some of the naturally processed peptides identified here do not fit these described motifs. Confirmation or rejection of this motif will await the refinement of the crystallographic structure of DR1 with bound peptide or studies with analogue peptides.

In contrast to the precise N-terminal cleavages observed in the previous study of six peptides bound to mouse I-A^b and five bound to I-E^b (ref. 5), the peptides bound to DR1 are heterogeneous at both the N and C termini. What might these data tell us about antigen processing in the class II pathway?

© 1992 Nature Publishing Group

In contrast to peptides bound to class I molecules, which are predominantly nonamers[1-4], class II peptides are larger and display a high degree of heterogeneity both in length and the site of terminal truncation, implying that the mechanisms of processing for class I and class II peptides are substantially different. Furthermore, it may suggest that class II processing is a stochastic event and that a DR allele may bind peptides of different lengths from a complex random mixture. The heterogeneity observed may only be due to protection of bound peptides from further degradation. Thus, class II molecules would play an active role in antigen processing (as previously proposed[35]) by protecting the bound peptides from complete degradation. Alternatively, the predominance of 15-amino-acid peptides bound to DR1 (as detected by both the MALD-MS and the yields of sequenced peptides) could be the result of trimming of bound peptides. In any event, the absence of peptides shorter than 13 and longer than 25 residues suggests that there are length constraints intrinsic to either the mechanism of peptide binding or antigen processing. In fact, a 15-residue peptide in an extended conformation can fit in an open cleft, but a longer peptide would have to extend beyond the cleft in such a conformation.

The predominance of peptides bound to DR1 that are derived from endogenously synthesized proteins, and particularly MHC-related proteins, may result from the evolution of a mechanism for presentation of self peptides in connection with the generation of self tolerance. The plethora of self peptides also has potentially important implications for the generation of autoimmune diseases linked to particular alleles of class II molecules, resulting from a breakdown of self tolerance.

Note added in proof: Another report describing naturally processed peptides extracted from murine class II molecules has been published[38]. □

Received 21 April; accepted 9 July 1992.

1. Van Bleek, G. M. & Nathenson, S. G. *Nature* **348**, 213-216 (1990).
2. Rotzschke, O. *et al. Nature* **348**, 252-254 (1990).
3. Jardetzky, T. S., Lane, W. S., Robinson, R. A., Madden, D. R. & Wiley, D. C. *Nature* **353**, 326-329 (1991).
4. Hunt, D. F. *et al. Science* **255**, 1261-1263 (1992).
5. Rudensky, A. Y., Preston-Hurlburt, P., Hong, S.-C., Barlow, A. & Janeway, C. A. *Nature* **353**, 622-627 (1991).
6. O'Sullivan, D. *et al. J. Immun.* **145**, 1799-1808 (1990).
7. Gorga, J. C., Brown, J. H., Jardetzky, T., Wiley, D. C. & Strominger, J. L. *Res. Immun.* **142**, 401-407 (1991).
8. Lane, W. S., Galat, A., Harding, M. W. & Schreiber, S. L. *J. Prot. Chem.* **10**, 151-160 (1991).
9. Pearson, W. R. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2444-2448 (1988).
10. Chen, B. P., Madrigal, A. & Parham, P. *J. exp. Med.* **172**, 779-783 (1990).
11. Cresswell, P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8188-8192 (1985).
12. Neefjes, J. J., Stollorz, V., Peters, P. J., Geuze, H. J. & Ploegh, H. L. *Cell* **61**, 171-183 (1990).
13. Guagliardi, L. E. *et al. Nature* **343**, 133-139 (1990).
14. Jones, P. P., Murphy, D. B., Hewgill, D. & McDevitt, H. O. *Molec. Immun.* **16**, 51-60 (1978).
15. Roche, P. A. & Cresswell, P. *Nature* **345**, 615-618 (1990).
16. Thomas, L. J., Nguyen, Q. V., Elliott, W. L. & Humphreys, R. E. *J. Immun.* **140**, 2670-2675 (1988).
17. Roche, P. A. & Cresswell, P. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3150-3154 (1991).
18. Roche, P. A. & Cresswell, P. *J. Immun.* **144**, 1849-1856 (1990).
19. Rothbard, J. B. *et al. Cell* **52**, 515-523 (1988).
20. Stern, L. J. & Wiley, D. C. *Cell* **68**, 465-477 (1992).
21. Sadegh-Nasseri, S. & Germain, R. N. *Nature* **353**, 167-170 (1991).
22. Dornmair, K., Rothenhausler, B. & McConnell, H. M. *Cold Spring Harb. Symp. quant. Biol.* **54**, 409-415 (1989).
23. Springer, T. A., Kaufman, J. F., Siddoway, L. A., Mann, D. L. & Strominger, J. L. *J. biol. Chem.* **252**, 6201-6207 (1977).
24. Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. & Rammensee, H. *Nature* **351**, 290-296 (1991).
25. Guttinger, M. *et al. Int. Immun.* **3**, 899-906 (1991).
26. Hill, C. M., Heybell, J. D., Allison, A. A. & Rothband, J. B. *J. Immun.* **147**, 189-197 (1991).
27. Guttinger, M. *et al. EMBO J.* **7**, 2555-2558 (1988).
28. Harris, P. E., Liu, Z. & Suciu-Foca, N. *J. Immun.* **148**, 2169-2174 (1992).
29. Latron, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 11325-11329 (1991).
30. Brown, J. H. *et al. Nature* **332**, 845-850 (1988).
31. Rothbard, J. B. & Taylor, W. R. *EMBO J.* **7**, 93-100 (1988).
32. Kropshofer, H., Bohlinger, I., Max, H. & Kalbacher, H. *Biochemistry* **30**, 9177-9187 (1991).
33. Jardetzky, T. S. *et al. EMBO J.* **9**, 1797-1803 (1990).
34. De Magistris, M. T. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 2990-2994 (1992).
35. Donermeyer, D. L. & Allen, P. M. *J. Immun.* **142**, 1063-1068 (1989).
36. Gorga, J. C., Horejsi, V., Johnson, D. R., Raghupathy, R. & Strominger, J. L. *J. biol. Chem.* **262**, 16087-16094 (1987).
37. Hall, M. *et al.* in *Biological Mass Spectrometry* (eds Burlingame, A. L. & McCloskey, J. A.) 101-115 (Elsevier, Amsterdam, 1990).
38. Hunt, D. F. *et al. Science* **256**, 1817-1820 (1992).

ACKNOWLEDGEMENTS. We thank M. Gordy for sequence analysis of peptides. This work was supported by the Irvington Institute for Medical Research (R.M.C. and R.G.U.), the Medical Foundation/Charles A. King Trust Fund (J.C.G.), the Damon Runyon-Walter Winchell Cancer Fund (L.J.S.), the International Human Frontier Science Program Organization (D.A.A.V.) and the NIH (J.L.S.).

The structure of ribosomal protein S5 reveals sites of interaction with 16S rRNA

V. Ramakrishnan*† & Stephen W. White‡

* Biology Department, Brookhaven National Laboratory, Upton, New York 11973, USA
† MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
‡ Department of Microbiology, Duke University Medical Center, Durham, North Carolina 27710, USA

UNDERSTANDING the process whereby the ribosome translates the genetic code into protein molecules will ultimately require high-resolution structural information, and we report here the first crystal structure of a protein from the small ribosomal subunit. This protein, S5, has a molecular mass of 17,500 and is highly conserved in all lifeforms[1-4]. The molecule contains two distinct α/β domains that have structural similarities to several other proteins that are components of ribonucleoprotein complexes. Mutations in S5 result in several phenotypes which suggest that S5 may have a role in translational fidelity and translocation. These include ribosome ambiguity or *ram*[5], reversion from streptomycin dependence[6] and resistance to spectinomycin[6]. Also, a cold-sensitive, spectinomycin-resistant mutant of S5 has been identified which is defective in initiation[7]. Here we show that these mutations map to two distinct regions of the molecule which seem to be sites of interaction with ribosomal RNA. A structure/function analysis of the molecule reveals discrepancies with current models[8,9] of the 30S subunit.

The ribosomal protein S5 can be crystallized from the thermophilic organism *Bacillus stearothermophilus*[10]. The crystals are in the trigonal space group $P3_221$ with cell dimensions $a = b = 59.3$ Å, $c = 109.8$ Å. We have cloned and overexpressed the gene for *B. stearothermophilus* S5 to provide sufficient material for structural analysis[11]. The *B. stearothermophilus* protein is 55% homologous to its *Escherichia coli* counterpart[1,2] and its function is almost certainly identical in the two organisms. We were able to determine the S5 structure with a single gold ($KAu(CN)_2$) derivative[12] by exploiting the variable wavelength properties of synchrotron radiation (see Fig. 1; details will be presented elsewhere).

S5 has an α/β structure, and the polypeptide chain is folded into two distinct domains to create a somewhat elongated molecule (Fig. 2). There appears to be no flexibility between the two halves of the molecule, which are tightly associated by means of an extensive interdomain hydrophobic core. The N-terminal half consists of three contiguous β-strands arranged as an anti-parallel β-pleated sheet, and a 4-turn α-helix that lies on the sheet surface. The C-terminal half is similar in that it also contains a 3-strand antiparallel sheet, but the strands are not contiguous and there are two α-helices on one surface. The N and C termini are both disordered. The notation for the secondary structure elements and connecting loops, and their locations within the primary structure are shown in Fig. 3. A feature of the fold is that the two halves associate such that the empty surface of the β-sheet in domain 1 packs against the two α-helices of domain 2. This produces two unusual features. First, one surface of the β-sheet in domain 2 is totally exposed, and second, helices $\alpha 2$ and $\alpha 3$ are packed between two β-sheets.

S5 sequences are now known from 13 prokaryotic, eukaryotic and archaebacterial organisms. The conservation of the S5 tertiary structure throughout evolution is evident from Fig. 3. The only major differences occur at the N and C termini, and in loop 3 which connects the two domains. The structurally important residues of S5 are particularly well conserved. These include hydrophobic amino acids at the core, and glycines and alanines

© 1992 Nature Publishing Group

Antibodies

Antibodies have been extensively studied since the use of immune serum to treat bacterial infections. However, due to the large number of different antibodies each human produces, detailed biochemical and molecular studies have been delayed for many decades. For detailed discussion, see ENGINEERING DESIGN OF THERAPEUTIC ANTIBODIES. Briefly, it was the identification of Bence Jones proteins in the urine samples of multiple myeloma patients as the light chain of myeloma proteins or immunoglobulins that provided a handle to analyze antibodies molecules in detail.

The major develops in such studies are:

1. Antibody molecules could be broken into two Fab fragments and one Fc fragment by enzymatic digestion.

2. Antibody molecules could be dissociated into one portion of light chains and one portion of heavy chains by reducing its disulfide bonds.

3. Bence Jones proteins were the light chains of antibodies

4. Amino acid sequences of Bence Jones proteins showed a variable region at the N-terminal half of the molecule, and a constant region at the C-terminal half of the molecule

5. The variable regions of Bence Jones proteins contained short segments near the C-terminal region which were randomly assorted with the rest of the variable regions. That short segment was designated as the switch peptide.

6. Among different variable regions of Bence Jones proteins or light chains of antibodies, large numbers of amino acid substitutions were localized in three short segments.

7. Similarly, there were also three short hypervariable segments in the variable region of heavy chains of antibodies. Together with those from the light chains, it was hypothesized that these six short segments form the antibody combining site. They were designated as complementarity determining regions (CDR's)

8. Hybridoma technology was developed by the fusion of myeloma cells to antibody producing cells to produce monoclonal antibodies of desired specificity.

9. On co-crystallizing lysozyme and an Fab fragment of anti-lysozyme antibody, the detailed three dimensional structure unequivocally verified that the six CDR's, three from the light chain and three from the heavy chain, formed the antibody combining site.

10. On sequencing DNA from adult and fetal mice, it was found that the variable region of the light chain contained two gene segments, V_L and J_L, joined together when the B lymphocyte was committed to produce antibodies. The joining process was known as DNA re-arrangement. The J_L minigene coded the switch peptide discovered before.

11. The variable regions of antibody heavy chains contained three gene segments, V_H, D_H and J_H. The joining of these three genes could give rise to numerous different amino acid sequences in the third CDR of heavy chains. This CDRH3 thus defined the fine specificity of a given antibody.

Stage	**Heavy chain**	**Light chain**	**Ig**	**IL-7 receptor?**	**CD19?**
Progenitor (or pre-pro) B cells	germline	germline	-	Yes	No
Early Pro (or pre-pre)-B cells	undergoes D-J rearrangement	germline	-	Yes	No
Late Pro (or pre-pre)-B cells	undergoes V-DJ rearrangement	germline	-	Yes	Yes
Large Pre-B cells	is VDJ rearranged	germline	IgM in cytoplasm	Yes	Yes
Small Pre-B cells	is VDJ rearranged	undergoes V-J rearrangement	IgM in cytoplasm	Yes	Yes
Immature B cells	is VDJ rearranged	VJ rearranged	IgM on surface	No	Yes
Mature B cells	is VDJ rearranged	VJ rearranged	IgM and IgD on surface	No	Yes

The generation of CDRH3 coding segment by the joining of V_H, D_H and J_H plays a central of defining a given antibody. This involves numerous DNA re-arrangement processes as outlined below:

1. At the sites of joining, there are specific nucleotide sequences known as recombination signal sequence (RSS) adjacent to the coding segments. They contain a 7-mer, a spacer, and a 9-mer. The spacer can be a stretch of around 12 nucleotides or 23 nucleotides. To connect two coding regions, one of the RSS's should have a spacer of 12 nucleotides and the other 23 nucleotides. A group of enzymes, including Rag-1 and Rag-2, will make double strand cuts of the DNA between the coding region and the RSS.

2. The 12 nucleotide spacer RSS is then ligated to the 23 nucleotide spacer RSS precisely, removing the intervening non-coding DNA sequence between the coding regions.

3. The open ends of the coding regions can have extra nucleotide added enzymatically before joining to each other. If the extra nucleotides are G,C-rich, the added region is designated as an N-segment. If they are A,T-rich, that region is known as a P-segment. In general, it is called an N/P-segment.

4. To maintain the reading frame of J_H minigene, the total number of nucleotides from the extra segment between V_H and D_H, the coding region of D_H, and the extra segment between D_H and J_H must be divisible by three. However, this requirement can provide a mechanism to shift the reading frame of the coding region of D_H minigene.

5. The RSS on both sides of the D_H coding region are the same, i.e. 7-mer, a space of 12 nucleotides, and a 9-mer. Therefore, during D_H and J_H joining, the D_H minigene can be inverted. As a result, the complementary strand of D_H minigene may also code for the CDRH3 segment of the variable region of the heavy chain of antibodies.

6. Some longer D_H minigenes may have a 7-mer within their coding regions. Thus the 3'-end RSS may be converted into a 7-mer, a spacer of 23 nucleotides, and a 9-mer. Part of this D_H minigene can then join onto another D_H minigene, known as D-D joining. This mechanism can give rise to a longer CDRH3 segment.

Any antibody with a given specificity usually has a unique CDRH3 segment. The other five CDR's, three from the light chain and two from the heavy chain, are generally used to improve binding affinity but do not change the fine specificity of the given antibody. As illustrated before, V-D-J re-arrangement in heavy chain variable region takes place in late pro-B cells. Only after its completion, V-J re-arrangement in light chain variable region would start.

The human antibody heavy chain constant region genes are located consecutively on chromosome 14q32 in the following order: IgM, IgD, IgG3, IgG1, IgA1, IgG2, IgG4, IgE and IgA2. On the 5' side of each, except IgD, there is a special DNA segment known the switch region. Initially, the antibody produced by a B lymphocyte is an IgM, a pentamer. Subsequently, the type may be changed to one of the others, e.g. IgG1, by a process known as class switch. This involves the recombination of the switch region sequences in front of IgM and IgG1. The intervening DNA segment is removed. IgG molecules are monomers, while IgA and IgE dimers.

At the 3'-end of each of these constant regions, there are two short exons. If these exons are not used, the resulting antibody will be secreted. On the other hand, splicing may make use of these short exons, giving rise to a hydrophobic segment of amino acid sequence at the C-terminal end of the heavy chain with Lys-Val-Lys at the very end. This segment will then anchor the antibody molecule into the cell membrane with the antibody combining site on the outside of the cell. Such membrane bound antibodies can then capture foreign antigens if the bnding affinity is sufficient high. Roughly, the amounts of secreted and membrane bound antibodies manufactured by each B cell is about equal.

Each domain of an antibody molecule is denoted by an abbreviation. For example, V_L stands for the variable region of the light chain. This notation, however, is ambiguous. V_L also represents the V-gene of the light chain before joining onto the J_L minigene. Since one indicates amino acid sequence and the other nucleotide sequence, this nomenclature remains in use. Therefore, at the amino acid sequence level, we have:

Light chain	V_L and C_L
Heavy chain	V_H, C_H1, (Hinge), C_H2, C_H3, and (C_H4)

For heavy chains, depending on the class, they have either a hinge region or a C_H4 segment.

The stability of antibody molecules is due to their three dimensional folding. Each of the domains mentioned above, except the hinge region, has a very stable structure known as the immunoglobulin fold. It consists two anti-parallel beta sheets stacked together like a sandwich with hydrophobic side chains between them. Then the V_L and V_H bind to each other with four anti-parallel beta sheets stacked like a club sandwich known as Fv. Similarly, C_L binds to C_H1, C_H2 to another C_H2, and C_H3 to another C_H3.

The detailed structure of the immunoglobulin fold is illustrated on the next page showing a stereo-pair of the three dimensional structure of lysozyme binding to the Fab fragment of an anti-lysozyme antibody.

Stereo-pair of an anti-lysozyme antibody Fab fragment binding lysozyme

MHC class I molecules

Major histocompatibility complex was originally studied in consideration of tissue rejection. Each individual has a large number of these molecules on their cells surfaces and they are distinct from individual to individual. Initially, human white blood cells were used to identify their differences, thus known as human leukocyte antigens (HLA's). Their biological functions were totally obscure at that time. Since they were membrane bound, purification of one of them was very difficult. Eventually, after growing out large quantity of specific human cancer cells in highly contained environments, extra-cellular fragments were isolated and their amino acid sequences determined (Orr HT, Lopez de Castro JA, Parham P, Ploegh HL, Strominger JL (1979) Comparison of amino acid sequences of two human histocompatibility antigens, HLA-A2 abd HLA-B7: Location of putative alloantigenic sites. Proc Natl Acad Sci USA, 76, 4395-4399). On reverse transcribing short segments of these amino acid sequences back to their nucleotide sequences, although not unique, several of the human MHC class I genes were isolated and their nucleotide sequences determined. The complexity of these molecules, both MHC class I and class II, was summarized in a review article (Schreuder GM et al. (2005) The HLA dictionary 2004: a summary of HLA-A, -B, -C, -DRB1/3/4/5 and DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR and –DQ antigens. Tissue Antigens, 65, 1-55).

MHC class I molecules consist of a light chain known as β_2-microglobulin and a heavy chain divided into five regions: α_1, α_2, α_3, membrane and cytoplasmic. The three dimensional structures of β_2-microglobulin and α_3 region of the heavy chain assume the standard immunoglobulin fold configuration, binding to each other and sitting on top of the cell membrane. On the other hand, the three dimensional structures of the α_1 and α_2 regions of the heavy chain are totally different (Bjorkman PJ, Saper MA, Samraoul B, Bennett WS, Strominger JL, Wiley DC (1987) Structure of the human class I histocompatibility antigen HLA-A2. Nature, 329, 506-512; The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature, 329, 512-518).

This new structure consists of a platform made of eight anti-parallel beta-strands, four each from α_1 and α_2 regions. Each region also contributes a bent alpha-helix sitting on top of the platform. Between the two bent alpha helices exists a groove where the processed peptide binds. Due to close proximity of these two bent alpha helices, the ends of the groove are closed. Thus the size of the processed peptide is usually limited to nine amino acid residues as discussed before. When the processed peptide is derived from a foreign protein antigen, this complex of processed peptide and MHC class I molecule can thus provide a special signal on the surface of cells which have internalized the foreign antigen.

Two different complexes consisting of MHC class I and processed peptide therefore exist, one with foreign peptides and the other with self-peptides. They should trigger different biological functions subsequently.

Proc. Natl. Acad. Sci. USA
Vol. 76, No. 9, pp. 4395–4399, September 1979
Biochemistry

Comparison of amino acid sequences of two human histocompatibility antigens, HLA-A2 and HLA-B7: Location of putative alloantigenic sites

(protein structure/genetic polymorphism/alloantigenic determinants)

HARRY T. ORR, JOSE A. LOPEZ DE CASTRO, PETER PARHAM, HIDDE L. PLOEGH, AND JACK L. STROMINGER

The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138

Contributed by Jack Leonard Strominger, June 27, 1979

ABSTRACT The complete amino acid sequence for papain-solubilized HLA-B7 heavy chain is compared with the partial sequences of HLA-A2 and H-2K^b heavy chains. Although these molecules are highly conserved (i.e., 80% homology in comparing HLA-B7 with HLA-A2; 72% and 74% homology in comparing H-2K^b with HLA-A2 and HLA-B7, respectively), two stretches of greater variability are observed. These clusters of variability are discussed in terms of their possible involvement in the alloantigenic determinant(s) characteristic of these highly polymorphic membrane antigens.

The major histocompatibility complex of humans and mice encodes several membrane antigens that play a central role in the immune response (1). These include the protein products of the *HLA-A* and *HLA-B* loci in humans and the *H-2K* and *H-2D* loci in mice. The most striking feature of this system is its genetic polymorphism, which gives rise to over 20 alleles at each of the *HLA-A* and *HLA-B* loci and to much larger numbers at the *H-2K* and *H-2D* loci. This polymorphism is believed to be related in some manner to the principal functions of these molecules, including roles in cell–cell interactions and as participants in cell-bound defense mechanisms. Presumably the evolutionary pressure for the polymorphism was provided by a variety of environmental stresses. These might include the need to control the proliferation of modified syngeneic cells (e.g., virus-infected or tumor cells) and of semi-allogeneic cells arising from the highly invasive trophoblast in placental mammals.

Our understanding of the biochemical nature of these molecules has progressed substantially in the last few years (2). They are glycoproteins consisting of two polypeptide subunits (3). A large 44,000-dalton glycosylated polypeptide is associated with a nonglycosylated 12,000-dalton polypeptide. The smaller subunit is invariant, is not encoded by the major histocompatibility complex, and is identical to β_2-microglobulin (4–7). On the other hand, the large subunit is encoded by the major histocompatibility complex and contains the alloantigenic determinants characteristic of these antigens. Previous experiments have shown that, in both mice (8) and humans (9), these determinants are not located in the carbohydrate portion of the heavy chain and thus must reside in the polypeptide chain. One of the most interesting features of the sturcture of these molecules is the location of these alloantigenic site(s) in the polypeptide, because they might be the functional regions of the molecules that have varied during evolution. They may have some recognition or receptor function (just as the hypervariable regions of Ig are the sites of antigen binding).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Papain enzymatically releases from the membrane a soluble complex, HLA_{pap}, in which the heavy chain is 34,000 daltons and the small subunit, β_2-microglobulin, is uncleaved and in which the alloantigenic activity is unaltered (10, 11). This method of solubilization combined with the use of the human lymphoblastoid cell line JY has provided material for the elucidation of the complete amino acid sequence of the HLA-B7$_{pap}$ heavy chain (271 residues). A second antigen, HLA-A2, is also easily obtained from this cell line. A virtually complete amino acid sequence of the amino-terminal 167 residues of the HLA-A2 heavy chain has also been obtained. In order to determine the possible location of the variable alloantigenic determinant(s), these sequences are compared to each other and to data presently available for the H-2K^b antigen (12).

MATERIALS AND METHODS

Cell Line. The human lymphoblastoid cell line JY, homozygous at both the *HLA-A* and the *HLA-B* loci (A2,2;B7,7), was used.

Preparation of Papain-Solubilized HLA Antigens (HLA_{pap}). HLA antigens were solubilized from JY membranes and purified as described (9). The two subunits of HLA were separated by using a Sephadex G-75 column in 1 M acetic acid.

Cyanogen Bromide Cleavage. The HLA-A2$_{pap}$ heavy chain was cleaved with CNBr in 70% (vol/vol) formic acid for 16 hr at 25°C either before or after reduction and alkylation of cysteines (9). When cleavage preceded reduction and alkylation, the CNBr fragments were applied directly to a Sephadex G-100 column and eluted with 1 M propionic acid. In the case in which cleavage followed reduction and alkylation, the CNBr fragments were lyophilized, redissolved in 200 mM NH_4HCO_3, and spun at 1000 × *g*, and the supernatant was applied to a Sephadex G-75 superfine column.

Enzymatic Generation of Peptides. Cyanogen bromide fragments were digested with trypsin (enzyme:substrate = 1:33 wt/wt, 6 hr, 37°C) and chymotrypsin (enzyme:substrate = 1:200 wt/wt, 30 min, 25°C). These peptides were isolated by using a microbore ion-exchange column eluted with a pyridine/acetic acid gradient (13). Tryptic peptides were also obtained from the CNBr-3 fragment after blockage of arginine residues with cyclohexanedione (14). These peptides were separated on a column of Sephadex G-50 in 1 M acetic acid. The use of fluorescamine in the detection of peptides was as described (15).

Amino Acid Analysis. The compositions of fragments and peptides were determined on a Beckman 121M automatic amino acid analyzer after hydrolysis for 24 hr at 110°C with 5.7 M HCl and 0.1% 2-mercaptoethanol.

Abbreviation: HLA_{pap}, papain-solubilized HLA antigen.

47

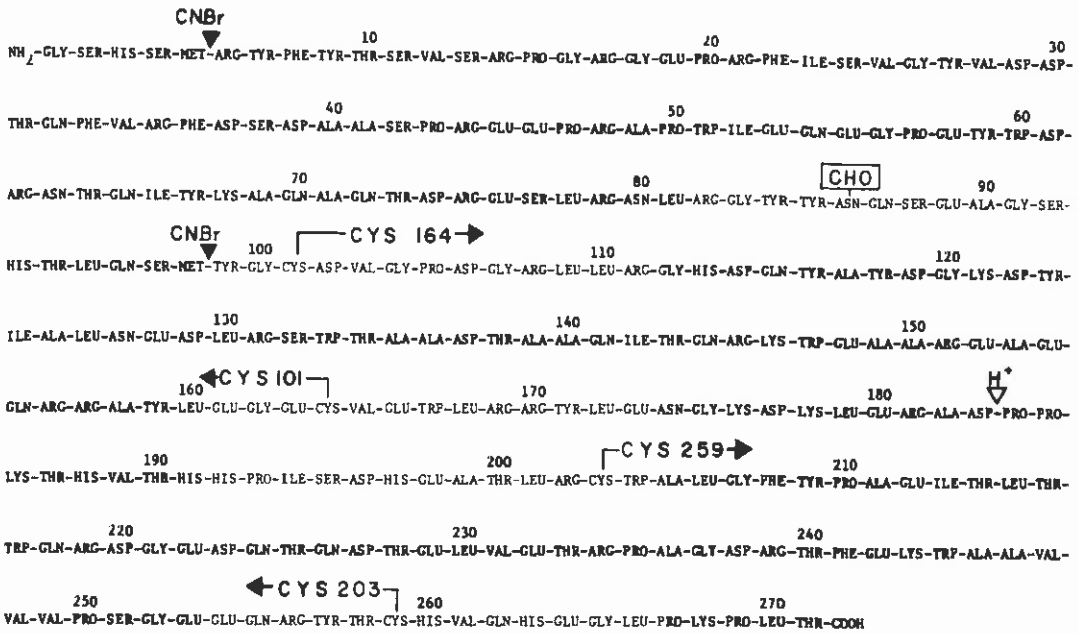


FIG. 1. The complete amino acid sequence of HLA-B7$_{pap}$. Cyanogen bromide and acid (H^+) cleavage points (26) are indicated. Also shown are the positions of the carbohydrate moiety (CHO) and the arrangement of the disulfide loops.

Sequence Determination. Automatic degradations of large fragments were performed by using an updated Beckman 890B sequenator. The 0.1 M Quadrol program (16) was utilized and degradations were performed after the addition of 3 mg of Polybrene to the cup (17, 18). Two methods were used to perform manual degradations of peptides: (*i*) the three-step manual Edman method as modified by Sauer *et al.* (19) and (*ii*) the dansyl-monitored Edman method (20). In the case of automatic degradations, phenylthiohydantoin residues were identified by using thin-layer chromatography (21, 22), gas/liquid chromatography (23, 24), and back hydrolysis followed by amino acid analysis (25).

RESULTS AND DISCUSSION

Complete Amino Acid Sequence of the Papain-Solubilized Heavy Chain of HLA-B7. The complete sequence of this molecule is shown in Fig. 1. The sequence was obtained from tryptic and chymotryptic peptides of the three large fragments derived from the heavy chain of HLA-B7 by acid and CNBr cleavages (26). A full account of this sequence determination and its analysis will be presented elsewhere.*

Cyanogen **Bromide Cleavage of HLA-A2$_{pap}$ Heavy Chain.** Unlike HLA-B7 and several other HLA antigens, the heavy chain of HLA-A2 does not contain an acid-labile Asp-Pro bond (26). Therefore, CNBr was chosen for the generation of large fragments for sequencing HLA-A2. Amino acid analysis of HLA-A2 heavy chain indicated the presence of four methionines (27), which should result in five CNBr fragments. Fig. 2 presents the elution profile of CNBr fragments obtained from HLA-A2 heavy chain without prior reduction and alkylation. Pool II was subsequently found to contain two CNBr fragments linked by a disulfide bridge. This was established by the presence of two NH_2-terminal residues, Tyr and Ala, in this pool. In addition, CNBr cleavage after reduction and alkylation yielded fragments CNBr-4 and CNBr-5, which were separated by gel filtration. Analysis of the remaining pools (Fig. 2) showed that each contained a single CNBr fragment, with pool III containing CNBr-3, pool IV CNBr-2, and the included peak, pool V, containing CNBr-1.

Alignment of CNBr fragments 1 and 2 was accomplished from the amino-terminal sequence of HLA-A2 (27). The isolation and sequence of the tryptic peptide Met-Glx-Pro-Arg from HLA-A2 heavy chain provided the overlap between CNBr-2 and CNBr-3. The remaining CNBr fragments, 4 and 5, were aligned on the basis of comparison with the HLA-B7 sequence.

Sequence Analysis of the Amino-Terminal 167 Residues of the HLA-A2 Heavy Chain. The sequence of the amino-terminal 167 residues of HLA-A2$_{pap}$ heavy chain is shown in Fig. 3. Edman degradation of HLA-A2 heavy chain combined

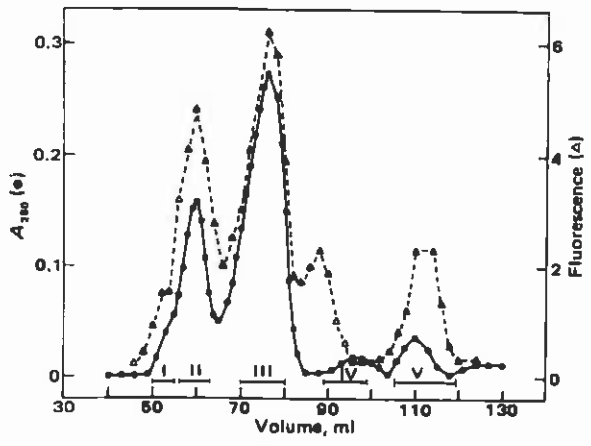


FIG. 2. Purification of HLA-A2$_{pap}$ CNBr fragments. CNBr cleavage and fragment isolation were performed without reduction and alkylation. Chromatography was on a 1 × 200 cm column of Sephadex G-100 with 1 M propionic acid. Fragments were located by A_{280} and by fluorescence (scale is arbitrary) after reaction with fluorescamine. Horizontal bars indicate pooled fractions.

* H. T. Orr, J. A. Lopez de Castro, D. Lancet, and J. L. Strominger, unpublished.

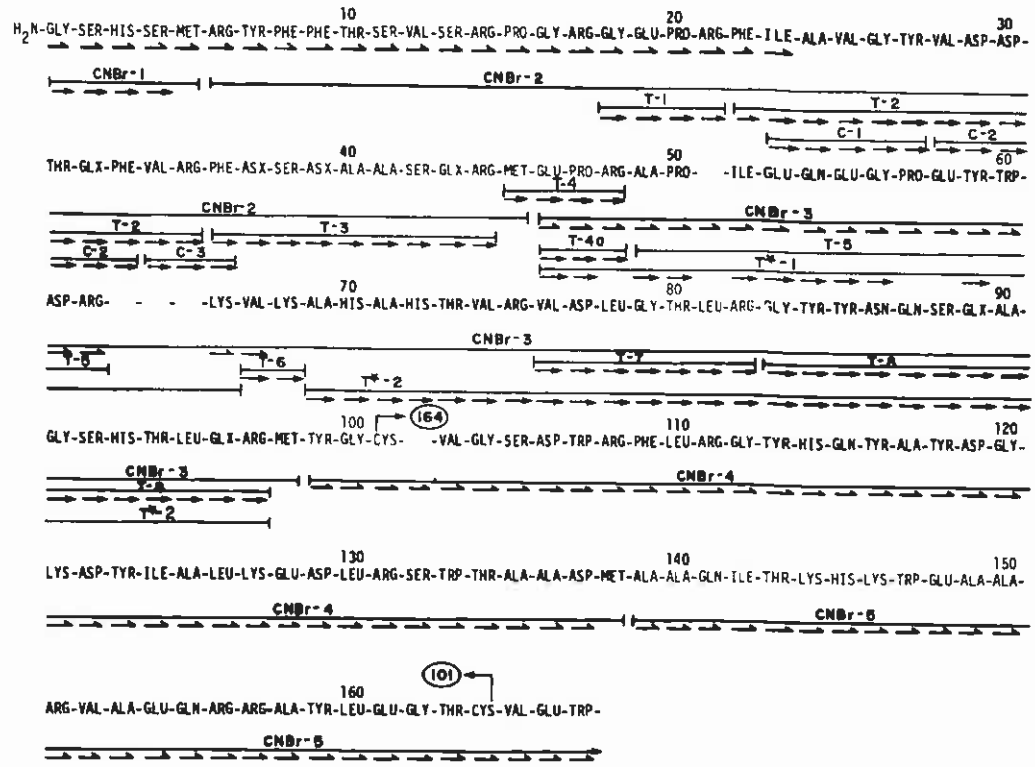


FIG. 3. Sequence of HLA-$A2_{pap}$ to residue 167. T denotes tryptic peptides; T*, tryptic peptides generated after Arg blockage; and C, chymotryptic peptides. Positions sequenced automatically are indicated by (—) and manually by (→).

with overlapping tryptic, chymotryptic, and CNBr peptides provided sequence data to residue 98. The sequence from 99 to 138 was determined from the automatic degradation of CNBr-4 to its carboxy terminus. Residues 139–167 were determined by automatic Edman degradation of CNBr-5, which by amino acid composition and comparison with HLA-B7, is thought to extend to residue 271 (unpublished data). The absence of homoserine and homoserine lactone after acid hydrolysis of CNBr-5 also indicated that this fragment extends to the carboxy terminus of HLA-$A2_{pap}$ heavy chain. The sequence of HLA-B7 heavy chain was used to align these two sequences, 99–138 and 139–167. At five positions no assignment of amino acid residues has been made yet. Therefore, the data provide the sequence of the amino-terminal two-thirds of the HLA-$A2_{pap}$ heavy chain with the exception of these residues.

The amino-terminal region of the HLA-A2 heavy chain, residues 1–98, contains the single Asn-linked carbohydrate of HLA-A2 (9). Like HLA-B7, this carbohydrate is attached to Asn-86. The tripeptide sequence Asn-Gln-Ser at residues 86–88 of HLA-A2 is compatible with a sequence of Asn-X-Ser/Thr noted for other glycoproteins at the site of glycosylation (28). Residues 99–167 contain the first disulfide loop of the HLA-A2 heavy chain. The linear arrangement of the disulfide loops in HLA-A2 is based on the dissociation of CNBr-4 and CNBr-5, each containing a single Cys residue, after reduction and alkylation. Thus, as in HLA-B7, the first disulfide loop of HLA-A2 heavy chain extends from Cys-101 to Cys-164, forming a loop of 63 residues.

The finding of significant sequence homology between the second disulfide loop of HLA-B7 heavy chain and immunoglobulin (Ig) constant domains (13) prompted a preliminary search for such homology with the HLA-A2 sequence. As with the HLA-B7 sequence, no homology between either Ig constant or variable domains and the amino-terminal 90 residues of HLA-A2 heavy chain was detected. Similarly, the first disulfide loop of HLA-B7 also failed to show Ig homologies.*

Comparison of HLA-$A2_{pap}$ Heavy Chain Data with Those from HLA-$B7_{pap}$ and H-$2K^b$ Antigens. As discussed earlier (8, 9), it is clear that the polypeptide chain of human and mouse histocompatibility antigens contain the alloantigenic determinants. The extensive sequence data available for two HLA antigens and an H-2 antigen afford the opportunity to assess the nature of primary structural differences that may contribute to the alloantigenic determinants.

Fig. 4 compares the complete sequence of HLA-$B7_{pap}$ heavy chain with the available HLA-A2 and H-$2K^b$ (12) sequences. In comparing the first 167 residues in HLA-B7 with those in HLA-A2, an overall sequence homology of 80% is obtained. Comparison of the available residues in H-$2K^b$ with those of HLA-B7 and HLA-A2 results in sequence homologies of 74% and 72%, respectively. Thus, the level of sequence homology between H-2 and HLA is only slightly less than that seen between HLA-A and HLA-B. A closer examination shows that H-2K is no more like HLA-A than it is like HLA-B—i.e., there are six positions at which H-$2K^b$ is identical to HLA-A2 and not HLA-B7 (residues 66, 76, 80, 138, 144, 163) and seven other positions at which H-$2K^b$ is like HLA-B7 and not HLA-A2 (residues 43, 70, 72, 77, 79, 145, 152). Because these residues are intermixed, there is neither an HLA-A region nor an HLA-B region in the H-$2K^b$ sequence. The high level of sequence homology between H-2 and HLA antigens, assessed from comparisons using sequence data from both disulfide loops and the amino-terminal region, clearly show that the human and mouse histocompatibility antigens must be functionally homologous. Such sequence homology also strongly supports the conclusion that an alloantigenic site(s) as it exists in H-2 is likely to be very similar to that in HLA.

The distribution of sequence differences found in comparing

FIG. 4. Comparison of the HLA-B7$_{pap}$ sequence with data on HLA-A2$_{pap}$ and H-2K^b (12). Identical residues are indicated by the solid bars. Residues are in the single-letter code: A, Ala; B, Asx; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr; and Z, Glx.

the two HLA specificities presents an interesting pattern. While differences are scattered throughout the polypeptide chains, there is a clear pattern of highly homologous regions spaced between regions of greater variability (Fig. 4). The amino-terminal 60 residues are a highly conserved stretch of the polypeptide chain. There are only 4/60 positions at which HLA-B7 differs from HLA-A2 in this region, yeilding a sequence homology of 93%. This region is followed by a cluster of differences located between residues 60 and 80. In this stretch, there are 9 positions at which HLA-B7 and HLA-A2 differ. These differences are all located within a 15-residue stretch, resulting in a drop in homology to 40%. Only one difference exists between HLA-B7 and HLA-A2 after residue 80 until residue 105, where a second cluster of differences is found. Between residues 105 and 114, 5 out of 10 positions differ. The sequence homology for this region is 50%. In the final stretch of available sequence data, residues 115 to 167, the homology is again at a high level, 88%. Of the 24 residues that differ between HLA-B7$_{pap}$ and the portion of HLA-A2$_{pap}$ whose sequence has been determined, 14 residues (58%) are found in the two clusters. It is also interesting to note that these two clusters are of similar size and have comparable levels of sequence homology.

A similar cluster of variability is found in comparing the HLA-B7 and HLA-A2 sequences with H-2K^b (12). Between residues 60 and 80 of HLA-B7 and H-2K^b heavy chains there are 11 differences. In comparing HLA-A2 with H-2K^b, this same stretch contains 10 differences. Therefore, it is highly possible that the alloantigenic site(s) consists of clusters of residues linearly located in the primary structure. Comparison of HLA with H-2 antigens indicates that one such cluster might be between residues 60 and 80. The HLA-A2 and HLA-B7 data indicate that more than one cluster might contribute to an alloantigenic specificity.

It is likely that the second disulfide loop region of these molecules is more highly conserved than the other regions. This region of HLA-B7 heavy chain shows highly significant sequence homology to Ig constant domains (13), and preliminary data for HLA-A28 and -B40 indicate a high degree of conservation in the second-disulfide loop region. Thus, it is unlikely that any additional large differences will be found between HLA-A2 and HLA-B7 in this region. Thus, the amino-terminal region of HLA antigens and the first disulfide loop region—i.e., residues 1–180—are the likely locations of the alloantigenic residues. The two clusters of variability shown in Fig. 4 are the most striking features seen in comparing HLA-A2 and HLA-B7.

A word of caution must be added regarding the variability of these molecules due to the fact that gene products of different loci are being compared. Therefore, isotypic differences may be superimposed on allotypic differences. In addition, several properties of HLA-A2 make it unique—i.e., lack of an acid cleavage site (26) and its behavior in ion-exchange chromatography (11). Such properties would be reflected in its amino acid sequence. It is clear that a full definition of the molecular basis of the alloantigenic site(s), the extent to which the clusters of variability participate in these determinants, and the differentiation between allotypic and isotypic markers of histocompatibility antigens must await additional evidence. Determining the sequences of the specificities HLA-A28 and HLA-B40 should be particularly useful in this respect, because they are crossreactive with HLA-A2 and HLA-B7, respectively (29). Thus, the number of sequence differences between these crossreactive specificities should be more restricted, perhaps enabling a clearer localization of the alloantigenic residues. In addition, the recent development of HLA-A2 (30) and HLA-B7 (31) specific monoclonal antibodies will provide access to extremely useful technology for the study of alloantigenic determinants.

We thank Mr. Thomas Kostyk for the isolation of the HLA antigens. This work was supported by National Institutes of Health Grant AI 10736, U.S. Public Health Service Research Fellowship AI 05483 to H.T.O., and Public Health Service International Research Fellowship TW 02582 to J.A.L.C.

1. Paul, W. E. & Benacerraf, B. (1977) *Science* **195**, 1293–1300.
2. Strominger, J. L., Engelhard, V. H., Fuks, A., Guild, B. C., Hyafil, F., Kaufman, J. F., Korman, A. J., Kostyk, T. G., Krangel, M. S., Lancet, D., Lopez de Castro, J. A., Mann, D. L., Orr, H. T., Parham, P., Parker, K. C., Ploegh, H. L., Pober, J. S., Robb, R. J. & Shackelford, D. A. (1979) in *The Role of the Major Histocompatibility Complex in Immunobiology*, eds. Benacerraf, B. & Dorf, M. E. (Garland, New York), in press.
3. Cresswell, P., Turner, M. J. & Strominger, J. L. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 1603–1607.
4. Grey, H. M., Kubo, R. T., Cohn, S. M., Poulik, M. D., Cresswell, P., Springer, T. A., Turner, M. & Strominger, J. L. (1973) *J. Exp. Med.* **131**, 1608–1612.

5. Nakamuro, K., Tanigaki, M. & Pressman, D. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 2863–2865.
6. Peterson, P. A., Rask, L. & Lindblom, J. B. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 35–39.
7. Silver, J. & Hood, L. (1974) *Nature (London)* **249**, 764–765.
8. Nathenson, S. G. & Muramatsu, T. (1971) in *Glycoproteins of Blood Cells and Plasma*, eds. Jamiesen, G A. & Greenwalt, T. J. (Lippincott Co., Philadelphia), pp. 245–262.
9. Parham, P., Alpert, B. N., Orr, H. T. & Strominger, J. L. (1977) *J. Biol. Chem.* **252**, 7555–7567.
10. Shimada, A. & Nathenson, S. G. (1969) *Biochemistry* **10**, 4048–4062.
11. Turner, M. J., Cresswell, P., Parham, P., Strominger, J. L., Mann, D. L. & Sanderson, A. R. (1975) *J. Biol. Chem.* **250**, 4512–4519.
12. Coligan, J. E., Kindt, T. J., Ewenstein, B. M., Uehara, H., Martinko, J. M. & Nathenson, S. G. (1979) *Mol. Immunol.* **16**, 3–8.
13. Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro, J. A. & Strominger, J. L. (1979) *Nature (London)*, in press.
14. Patthy, L. & Smith, E. L. (1975) *J. Biol. Chem.* **250**, 557–564.
15. Udenfriend, S., Stein, S., Bohlen, P., Dairman, W., Leimgruber, W. & Weigele, M. (1972) *Science* **178**, 871–872.
16. Brauer, A. W., Margolis, M. N. & Haber, E. (1975) *Biochemistry* **14**, 3029–3035.
17. Tarr, G. E., Beecher, J. F., Bell, M. & McKean, D. J. (1978) *Anal. Biochem.* **84**, 622–627.
18. Klapper, D. G., Wilde, C. E., III & Capra, J. D. (1978) *Anal. Biochem.* **85**, 126–131.
19. Sauer, R. T., Niall, H. D., Hogan, M. L., Keutmann, H. T., O'Riordan, J. L. H. & Potts, J. T., Jr. (1974) *Biochemistry* **13**, 1994–1999.
20. Gray, W. R. (1967) *Methods Enzymol.* **11**, 333–334.
21. Kulbe, K. D. (1974) *Anal. Biochem.* **59**, 564–573.
22. Summers, M. R., Smythers, G. W. & Orosylan, S. (1973) *Anal. Biochem.* **93**, 624–628.
23. Pisano, J. J. & Bronzert, T. J. (1969) *J. Biol. Chem.* **244**, 5597–5607.
24. Niall, H. D. (1973) *Methods Enzymol.* **27**, 942–1010.
25. Smithies, O., Gibson, D. M., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G. & Ballantyne, D. L. (1971) *Biochemistry* **10**, 4912–4921.
26. Terhorst, C., Robb, R., Jones, C. & Strominger, J. L. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4002–4006.
27. Terhorst, C., Parham, P., Mann, D. L. & Strominger, J. L. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 910–914.
28. Kinquist, K. K. & Lennarz, W. J. (1978) *J. Supramol. Struct.* **8**, 51–65.
29. Joysey, V. C. & Wolf, E. (1978) *Br. Med. Bull.* **34**, 217–222.
30. Parham, P. & Bodmer, W. F. (1978) *Nature (London)* **276**, 397–399.
31. Brodsky, F. M., Parham, P., Barnstable, C. J., Crumpton, M. J. & Bodmer, W. F. (1979) *Immunol. Rev.*, in press.

51

PubMed

U.S. National Library of Medicine
National Institutes of Health

Display Settings: Abstract

Full Text Online WILEY ONLINE LIBRARY

Tissue Antigens. 2005 Jan;65(1):1-55.

The HLA Dictionary 2004: a summary of HLA-A, -B, -C, -DRB1/3/4/5 and -DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR and -DQ antigens.

Schreuder GM, Hurley CK, Marsh SG, Lau M, Fernandez-Vina M, Noreen HJ, Setterholm M, Maiers M; World Marrow Donor Associations Quality Assurance and IT Working Groups; WHO Nomenclature Committee for Factors of the HLA system; 13th International Histocompatibility Workshop Serology Component; International Cell Exchange, UCLA; US National Marrow Donor Program.

Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands. g.m.l.schreuder@lumc.nl

Abstract

This report presents serologic equivalents of human leucocyte **antigen** (HLA)-A, **-B**, **-C**, -DRB1, -DRB3, -DRB4, -DRB5 and -DQB1 alleles. The dictionary is an update of the one published in 2001. The data summarize equivalents obtained by the World Health Organization Nomenclature Committee for factors of the HLA System, the International Cell Exchange, the National Marrow Donor Program, recent publications and individual laboratories. This latest update of the dictionary is enhanced by the inclusion of results from studies performed during the 13th International Histocompatibility Workshop and from neural network analyses. A summary of the data as recommended serologic equivalents is presented as expert assigned types. The tables include remarks for alleles, which are or may be expressed as **antigens** with serologic reaction patterns that differ from the well-established HLA specificities. The equivalents provided will be useful in guiding searches for unrelated hematopoietic stem cell donors in which patients and/or potential donors are typed by either serology or DNA-based methods. The serological DNA equivalent dictionary will also aid in typing and matching procedures for organ transplant programs whose waiting lists of potential donors and recipients comprise of mixtures of serologic and DNA-based typings. The tables with HLA equivalents and a questionnaire for submission of serologic reaction patterns for poorly identified allelic products will be made available through the WMDA web page: www.worldmarrow.org. and in the near future also in a searchable form on the IMGT/HLA database.

PMID: 15663741 [PubMed - indexed for MEDLINE]

Publication Types, MeSH Terms, Substances

LinkOut - more resources

Structure of the human class I histocompatibility antigen, HLA-A2

P. J. Bjorkman*†‡, M. A. Saper*, B. Samraoui*, W. S. Bennett*‡, J. L. Strominger* & D. C. Wiley*§

* Department of Biochemistry and Molecular Biology, Harvard University, Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA
† Department of Medical Microbiology, Stanford University, Stanford, California 94305, USA
‡ Present addresses: Department of Medical Microbiology, Stanford University, Stanford, California 94305, USA (P.J.B.) and Max-Planck-Institute für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin 33, FDR (W.S.B.)
§ To whom correspondence should be addressed

The class I histocompatibility antigen from human cell membranes has two structural motifs: the membrane-proximal end of the glycoprotein contains two domains with immunoglobulin-folds that are paired in a novel manner, and the region distal from the membrane is a platform of eight antiparallel β-strands topped by α-helices. A large groove between the α-helices provides a binding site for processed foreign antigens. An unknown 'antigen' is found in this site in crystals of purified HLA-A2.

HLA (human leukocyte antigen) molecules are polymorphic membrane glycoproteins found on the surface of nearly all cells. Multiple genetic loci within the major histocompatibility complex (MHC) encode these proteins, and one individual simultaneously expresses several polymorphic forms from a large pool of alleles in the population. HLA molecules (also known as class I histocompatibility antigens) are the targets of antibodies and cytotoxic T lymphocytes (CTL) during rejection of foreign transplants[1,2]. They are also recognized by T cells together with viral antigens on infected cell surfaces, a phenomenon known as MHC restricted recognition[3]. In contrast to antibodies, that can bind to free virus or soluble antigens, T-cell receptors only recognize foreign antigens that are associated with a particular HLA molecule. Because one individual expresses a limited set of different HLA molecules, a central question has been how these few HLA molecules can interact with so many foreign antigens. Limitations in the ability of a particular HLA molecule to associate with all antigens may explain the linkage of histocompatibility antigens to variations in susceptibility to human diseases[4], and the immune system's responsiveness to particular antigens[5].

Recent work has shown that virus-specific CTL will lyse an uninfected target cell of appropriate class I specificity to which peptide fragments of a viral protein have been added[6]. T-helper cells had previously been shown to recognize fragments of protein antigens in association with class II histocompatibility

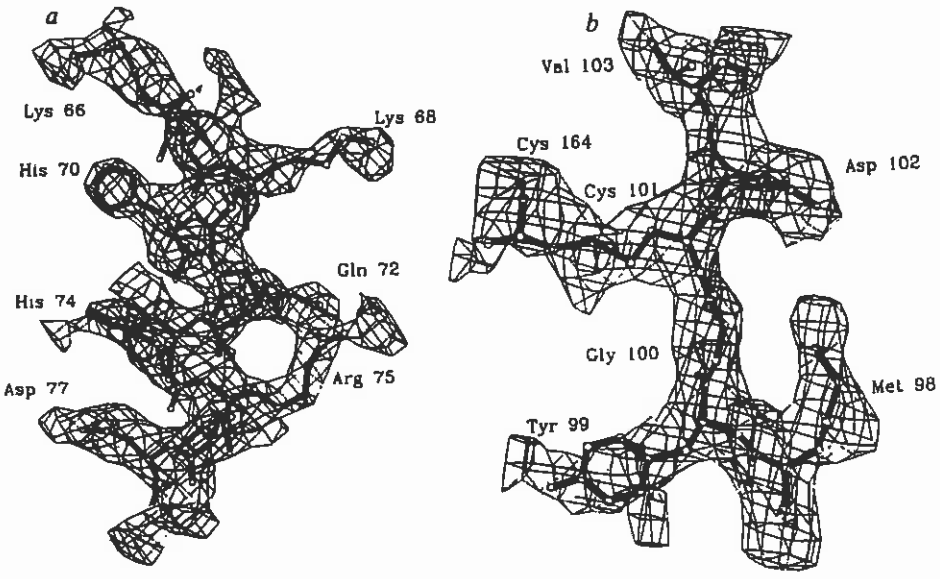


Fig. 1 Representative portions of the HLA-A2 model superimposed on the 3.5 Å averaged electron density map. Electron density is shown for residues 66–77 in the α_1 long helix (*a*) and for residues 98–103 in β-strand 1 of α_2 (*b*). Contours are drawn at 0.9 σ. Rotation and translation parameters were determined that relate the monoclinic and orthorhombic maps, and these maps were iteratively averaged[34]. The input $P2_12_12_1$ map was calculated with phases from a single derivative, and the input $P2_1$ map was calculated with phases from four derivatives combined with phases from a partial model of the structure (model did not include side chains in α_1 and α_2; see text for details). As a control against bias from the partial model included in the $P2_1$ map, the real space averaging was repeated starting with the model-free, uninterpretable monoclinic and orthorhombic maps, resulting in a similar and interpretable averaged map. It is possible that some side chains located on the surface of the molecule or near crystal contacts have different conformations in the two crystal forms. The final positioning of side chains awaits extension to higher resolution and refinement. (Current *R* factor between $P2_1$ calculated and observed amplitudes is 26.2% with good geometry after 24 cycles of 6–3.2 Å refinement using the program CORELS[13].)

53

© 1987 Nature Publishing Group

Table 1 Data collection, phasing and map averaging statistics

Mean figure of merit: Monoclinic – 0.60 for 5,759 reflections
(3.5 Å resolution) Orthorhombic – 0.40 for 4,549 reflections

Derivative	d_{lim} (Å)	Number of films (source)[‖]	Number of measurements	Number of reflections, (% complete)	R†	r.m.s. f_h/E‡ (15–3.5 Å)
$P2_1$ ($a = 60.4$ Å, $b = 80.4$ Å, $c = 56.5$ Å, $\beta = 120.4°$)						
Native	2.7	120 (DESY/CHESS)	80,891	13,513 (96%)	0.101	—
K_2PtCl_4	2.7	44 (DESY)	42,299	12,808 (89%)	0.095	1.78
HgI_2	2.7	40 (DESY)	37,101	12,696 (81%)	0.105	0.92
K_2OsCl_6	2.7	35 (DESY)	30,888	12,110 (78%)	0.091	0.52
PCMP*	2.8	54 (SSRL/CHESS)	34,295	10,247 (83%)	0.114	0.97
$P2_12_12_1$ ($a = 60.2$ Å, $b = 80.4$ Å, $c = 112.2$ Å)						
Native	2.8	24 (CHESS/XEN)	30,590	10,638 (70%)	0.087	—
K_2PtCl_4	3.5	– (XEN)	15,106	5,028 (70%)	0.094	1.45

Map averaging statistics at 3.5 Å resolution

	R§		r.m.s. phase difference¶ (degrees)	
	$P2_1$	$P2_12_12_1$	$P2_1$	$P2_12_12_1$
Cycle 1	0.342	0.343	47	38
Cycle 5	0.247	0.255	68	78
Cycle 10	0.166	0.185	73	83

Film data were collected on 5° oscillation photographs ($P2_1$) or 2–3° oscillation photographs ($P2_12_12_1$) using synchrotron X-ray sources, and processed and scaled as described[28]. The Xentronics area detector data were recorded as 5′ oscillation frames, then merged into batches of 20 frames and scaled to the native orthorhombic film data. The monoclinic derivatives used in phasing were chosen after screening over 100 heavy atom compounds, and recording complete film data sets from 13 different heavy atom derivatives, 9 of which did not prove useful in phasing. The most useful monoclinic derivative (K_2PtCl_4) was the only dataset with an overall phasing power (r.m.s. f_H/E) greater than 1.0 to 3.5 Å resolution. The other derivative datasets had phasing powers greater than 1.0 only at lower resolution (~6.0 Å). The K_2PtCl_4 orthorhombic data were truncated at 3.5 Å resolution due to poor quality at higher resolution, and these data determine the resolution limit of the averaged electron density maps (Fig. 1*a*,*b*). Statistics are presented for 10 cycles of real-space averaging of the monoclinic and orthorhombic electron density maps.

* *p*-Chloromercuriphenol; † $R_I = \sum|I - \langle I\rangle|/\sum\langle I\rangle$; ‡ f_h, heavy atom structure factor; E is the residual lack of closure; ‖ DESY, Deutches Elektronen Synchrotron; CHESS, Cornell High Energy Synchrotron Source; SSRL, Stanford Synchrotron Radiation Laboratory; XEN, Xentronics Area Detector.

§ R, crystallographic R-factor between observed structure factors and calculated structure factors from the averaged map.

¶ r.m.s. phase difference between calculated phases and phases used in the starting map.

molecules[7]. Short synthetic peptides (10–20 residues long) have also been shown to bind to purified class II proteins[8,9] at what appears to be a single binding site[10]. By analogy, it is likely that the homologous class I molecules bind antigenic peptides, and that the HLA–peptide complex is recognized by T-cell receptors on CTL[11].

Class I molecules (HLA-A, B, C in humans, H-2K, D, L in mice) are composed of two polypeptide chains[12]; a heavy chain (relative molecular mass 44,000; M_r 44K) which spans the membrane bilayer, and the non-covalently attached light chain, β_2-microglobulin (β_2m—12K)[13,14]. The extracellular portion of the heavy chain is divided into three domains, α_1, α_2 and α_3 (ref. 15), each ~90 amino acids long and encoded on separate exons[16]. The α_3 domain and β_2m are relatively conserved and show amino-acid sequence homology to immunoglobulin constant domains[15,17–20]. The polymorphic α_1 and α_2 domains show no significant sequence homology to immunoglobulin constant or variable domains, but have been reported to show weak sequence homology to each other[21]. We have crystallized a soluble fragment of HLA-A2 (ref. 22), composed of α_1, α_2, α_3 and β_2m, after removing the transmembrane anchor of the heavy chain by papain digestion[23,24]. (HLA-A2, formerly named Mac, was the first polymorphic human leukocyte antigen identified[25].)

The structure determination of HLA-A2 was undertaken to gain an understanding of where the polymorphic residues are located on the structure, how and where antibodies and T-cell receptors recognize the molecule, and how the molecule interacts with foreign antigens. Here we report a description of HLA-A2 from a 3.5-Å X-ray crystallographic structure determination. The membrane-proximal α_3 and β_2m domains have tertiary structures resembling antibody domains, but are paired by a novel interaction not previously seen in immunoglobulin structures. The α_1 and α_2 domains are nearly identical to each other in structure, and are not similar to immunoglobulin constant or variable domains. The domains α_1 and α_2 form a platform composed of a single β-pleated sheet topped by α-helices with a long groove between the helices. Electron density which is not a part of the HLA molecule is observed in this site between the helices, presumably representing an unknown bound antigen. The location of polymorphic residues, antibody binding sites, and mutations affecting recognition by CTL are analysed in the accompanying article, which defines the binding site for processed foreign antigens[26].

Structure determination

Soluble HLA-A2 was purified after papain digestion of plasma membranes from the homozygous human lymphoblastoid cell line JY[23,24]. Papain cleaves the HLA heavy chain at residue 271, thirteen residues from the transmembrane region[27], yielding a molecule composed of α_1, α_2, α_3 and β_2m. When purified, 3–4 mg protein is obtained from 200 litres of human tissue culture cells.

The crystallization and a proposed packing model for the two crystal forms of HLA-A2 and HLA-A28 have been described previously[22,28]. HLA-A2 crystallizes in two space groups: monoclinic $P2_1$ and orthorhombic $P2_12_12_1$. Both crystal forms grow as very thin (20 μm) plates ~0.5 mm square. Occasional 100 μm thick orthorhombic crystals have been grown. Data from the 20 μm thick monoclinic crystals were collected using synchrotron X-ray sources. Part of the data from the orthorhombic crystals were collected using a Xentronics area detector[29,30]. Data collection and phasing statistics are presented in Table 1.

A monoclinic electron density map was calculated to 3.5 Å resolution using isomorphous differences and anomalous scat-

© 1987 Nature Publishing Group

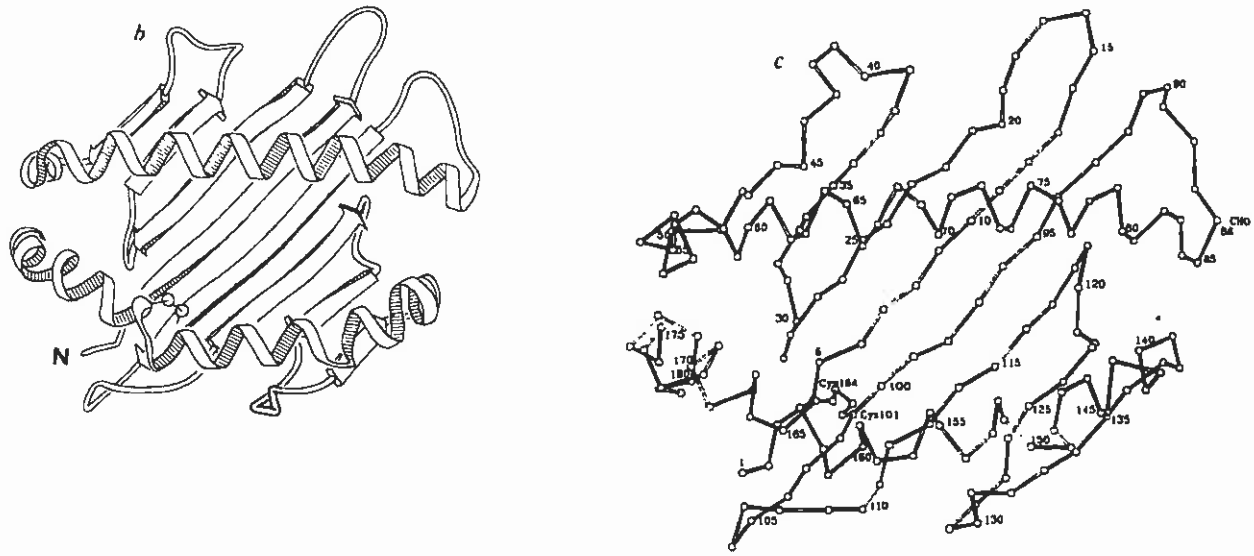


Fig. 2 Schematic representation of the structure of HLA-A2. The β-strands are shown as thick arrows in the amino to carboxy direction, α-helices are represented as helical ribbons. Connecting loops are depicted as thin lines. Disulphide bonds are indicated as two connected spheres. *a* (Facing page), schematic representation of the four domains of HLA-A2. The molecule is shown with the membrane proximal immunoglobulin-like domains (α_3, β_2m) at the bottom, and the polymorphic α_1 and α_2 domains at the top. The indicated C-terminus of α_3 is the papain cleavage site. In the membrane-bound molecule, another 13 amino acids extend past the cleavage site toward the membrane, which we assume to be horizontal at the bottom of this picture. This orientation places the helices of the α_1 and α_2 domains, and the antigen recognition site located between them, on the top surface of the molecule. The domains α_1 and α_2 form a platform with a single eight-stranded β-pleated sheet (seen edge on), covered by α-helices. *b*, Schematic representation of the top surface of HLA-A2. The α_1 and α_2 domains are shown as viewed from the top of the molecule, showing the surface that is presumably contacted by a T-cell receptor (90° from Fig. 2*a*). The N terminus of α_1 is indicated. Each domain consists of four antiparallel β-strands followed by a long helical region, and the domains pair to form a single eight-stranded β-sheet topped by α-helices. The disulphide bond connects residue 164 in the α_2 long helix to residue 101 in the first β-strand of α_2. The residue analogous to Cys 101 in α_1 is Ser 11; it is 8 Å from Ser 71 (C_α to C_α) on the α_2 long helix. *c*, α-carbon backbone of the α_1 and α_2 domains. The domains are shown in the same orientation as in *b*. Every fifth residue and the oligosaccharide attachment site (86) are labelled.

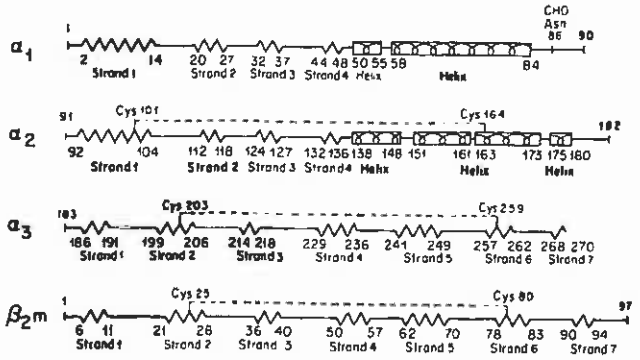


Fig. 3 (Right) Schematic representation of the secondary structure of HLA-A2. The four β-strands and helical regions in α_1 occupy equivalent locations as the β-strands and α-helices in α_2, and the two domains are folded into homologous tertiary structures (Fig. 4). In α_3 and β_2m, the seven β-strands fold into structures resembling immunoglobulin constant domains. The positions of cysteine residues involved in disulphide bonds, and the location of the glycosylation site at residue 86 are indicated.

Fig. 4 (Facing page) Comparison of the structures of the α_1 and α_2 domains. *a*, The α-carbon backbone of α_1. N and C termini are labelled. *b*, The α-carbon backbone of α_2. N and C termini are labelled. *c*, The α_1 (blue) and α_2 (red) domains as they are paired in the HLA-A2 structure. The N-terminal β-strand in each domain is depicted with main-chain atoms to emphasize the hydrogen bonds (yellow lines) between these strands that form a single, eight-stranded β-pleated sheet from the four-stranded sheets in each domain. The rest of each domain is shown as the C_α backbone. Yellow stars highlight the C_α atom of any residue with a main or side chain atom within 4 Å of an atom located in the partner domain (see Table 2). The N terminus of α_1 is labelled. *d*, The α-carbon backbones of α_1 (blue) and α_2 (red) superimposed to show structural similarity; α_1 and α_2 were aligned after a least-squares fit of the α-carbon coordinates of the β-strand and α-helical residues in each domain. The average r.m.s. positional difference is 3.1 Å.

Fig. 5 (Facing page) Location of the pseudo-symmetry axes in the HLA-A2 molecule. Domains α_1 (red) and α_2 (yellow) are related by an approximate dyad axis of symmetry (orange arrow). (The relationship is a rotation of −177° followed by a 0.7 Å translation.) Domains α_3 (blue) and β_2m (purple) are related by a 146° rotation (blue arrow) followed by a 13-Å translation. The two axes do not intersect, and deviate from colinearity by 25° (in projection).

Fig. 6 (Facing page) Van der Waals surface representation[34] of the top of the HLA-A2 molecule (*a*) showing the deep groove identified as the antigen recognition site[26] and the electron density (*b*) found in this site in crystals of HLA-A2. The surface was generated with the program GRIDCTR written by Mark Handschumacher and Fred Richards. The molecule is shown from the top with the Cα backbone (pink) oriented as in Fig. 2*b*,*c*. *a*, Van der Waals surface of the α_1 and α_2 domains (blue) showing a deep groove running between the helical regions in the α_1 and α_2 domains. This groove has been identified as the recognition site for processed antigens (see text and ref. 26 for details). *b*, The extra electron density (red contours at 0.8σ) found in both crystal forms of HLA-A2 is shown superimposed on the Van der Waals surface of α_1 and α_2 (blue). The extra density represents a bound molecule or mixture of molecules that co-purified and co-crystallized with HLA-A2. Some regions of the extra density may be wide enough to accommodate a peptide in an α-helical conformation, but in the absence of the knowledge of its composition, or whether it is one species or a mixture, it is not possible to interpret its structure unambiguously at the current resolution of the electron density maps (3.5 Å).

55

© 1987 Nature Publishing Group

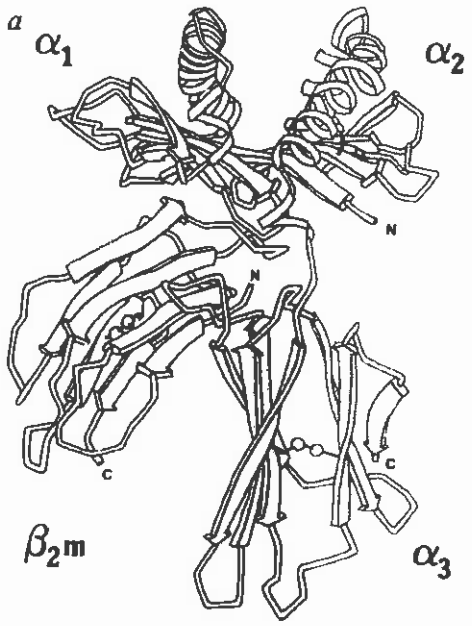


Fig. 2a

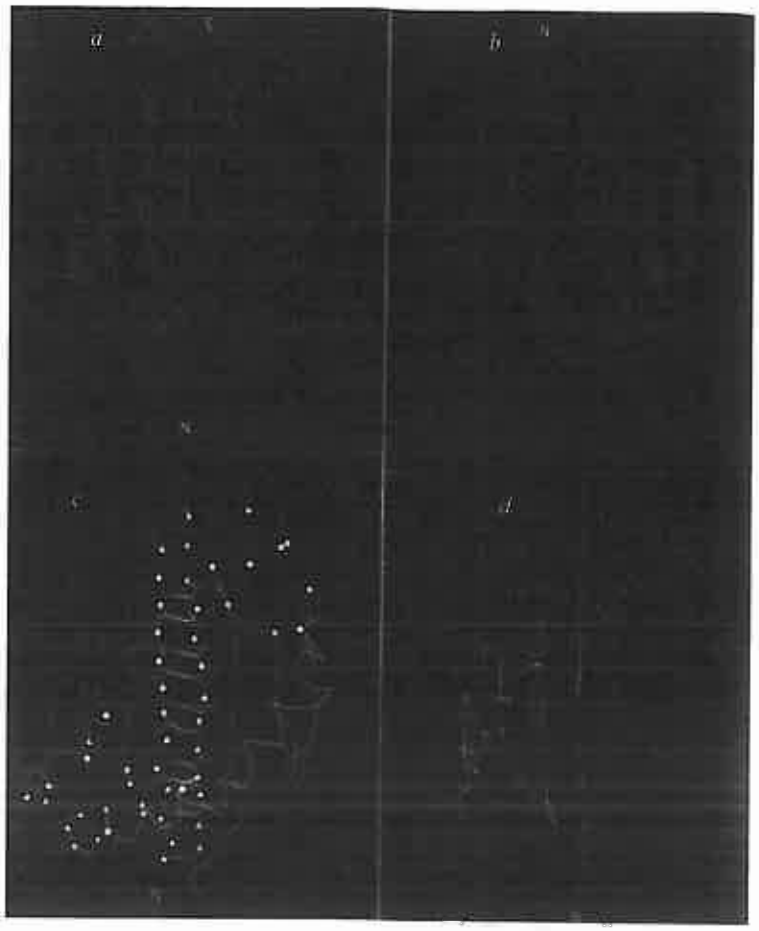


Fig. 4

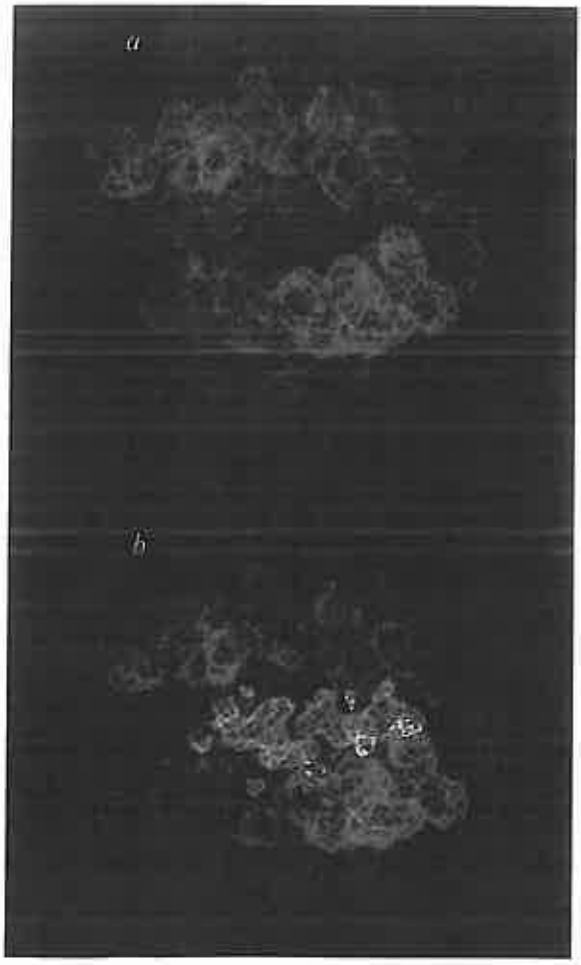


Fig. 6

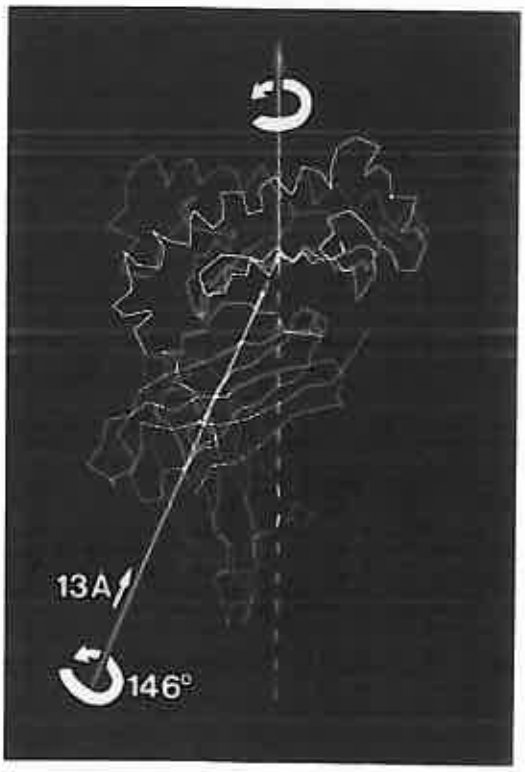


Fig. 5

© 1987 Nature Publishing Group

tering from four heavy atom derivatives and modified by iterative solvent flattening[31]. Due to the poor quality of the heavy atom derivatives, the map was not fully interpretable, although the molecular outline and the two immunoglobulin-like domains of α_3 and β_2m were evident. Subsequent electron density maps derived from many cycles of model building, phase combination[32] and CORELS refinement[33] permitted the complete structures of α_3 and β_2m to be determined, and ~80% of the main chain of α_1 and α_2 to be fit as segments of polyalanine. To complete tracing the polypeptide chain in α_1 and α_2, a single isomorphous replacement electron density map was calculated from 3.5 Å data from orthorhombic crystals. The relationship between the two crystal forms was established by a six-dimensional real-space search using the monoclinic partial model, and confirmed by the location of a common heavy atom derivative. The monoclinic and orthorhombic maps were averaged inside their molecular envelopes using the Bricogne algorithms[34]. Solvent region densities were replaced by their average value, and new phases were calculated by Fourier transformation of the maps. Ten cycles of averaging in both crystal forms produced significant phase changes and *R*-factor improvements (Table 1), allowing the complete polypeptide chain to be traced unambiguously with the aid of the amino-acid sequence[35] (details will be published elsewhere; Saper *et al.*, manuscript in preparation). Despite the limited resolution (3.5 Å), 80% of the side chains are fitted in well-defined density (Fig. 1). In the future, least squares refinement using 2.7 Å native data in either space group will allow a more precise positioning of side chains.

Description of the structure

HLA-A2 consists of two pairs of structurally similar domains: α_1 has the same tertiary fold as α_2, while α_3 has the same tertiary fold as β_2m. As oriented in Fig. 2*a*, the α_3 and β_2m domains are closest to the bottom of the figure. These domains are proximal to the cell membrane, which would probably be horizontal in this orientation (see Fig. 2 legend). Domains α_1 and α_2 form the top of the molecule with their helical sides facing away from the cell. A view of the 'top' of HLA-A2 (Fig. 2*b*) shows the surface that is presumably recognized by T cells. The cross-section of the top of the molecule is about 50 Å × 40 Å, and the molecule is about 70 Å in length.

Structures of α_3 and β_2m

The α_3 and β_2m domains are both β-sandwich structures composed of two antiparallel β-pleated sheets, one with four β-strands and one with three β-strands, connected by a disulphide bond. This tertiary structure has been described for constant regions of immunoglobulin molecules (for review, see refs 36, 37) and was expected from the significant sequence homology between α_3, β_2m and constant regions[15,17-20]. The structure of human β_2m as bound to the HLA-A2 heavy chain also appears similar to the structure of the free monomer of bovine β_2m[38]. A least-squares comparison between the α-carbon coordinates of the β-strand regions of α_3 with β_2m indicates an average r.m.s. positional difference of 1.4 Å. A similar comparison of α_3 and β_2m to an immunoglobulin constant region (C_{H3} from human Fc) yields an r.m.s. difference of 1.6 Å in each case. Similar comparisons among antibody constant domains give r.m.s. differences of 0.6–1.1 Å[38].

Structures of α_1 and α_2

The α_1 and α_2 domains each consist of an antiparallel β-pleated sheet spanned by a long α-helical region that is C-terminal to the four β-strands in the sheet (Figs 3, 4*a*,*b*). In the helical region of α_1, a short α-helix (50–55) precedes a longer curved α-helix (58–84). These helices are almost at right angles (~110°). In α_2, a short helix (138–148) precedes a longer helix (151–173) at an angle of ~130°. The long helix in α_2 is kinked at residue 162. An additional short helix in α_2 (177–181) connects α_2 to α_3. A disulphide bond in α_2 connects residue 101 in the N-terminal β-strand to the long helix at residue 164. In Fig. 4*d*, α_1 is superimposed on α_2 to show their structural similarity (overall r.m.s. difference 3.1 Å). Published secondary structure predictions[39,40] can be compared with Fig. 3.

Interdomain contacts

The α_3 and β_2m domains contact each other in an interaction not found between pairs of constant domains in the known antibody structures (for example, C_{H1} and C_L in an Fab, the C_{H3} dimer in Fc). Antibody constant domains are related by a nearly exact dyad (180°) symmetry axis, with their four-stranded β-sheets forming the contact interface[36,37]. In HLA-A2, although the immunoglobulin-like domains contact each other with their four-stranded sheets (Table 2), they are related by a 146° rotation followed by a 13 Å translation (Fig. 5).

The structurally similar α_1 and α_2 domains (Fig. 4*a*,*b*) are paired in the HLA molecule by an approximate dyad axis of symmetry, such that the four β-strands from each domain form a single antiparallel β-sheet with eight strands (Fig. 4*c*). This β-sheet is topped by the helical regions from each domain, which are separated by about 18 Å (centre to centre). The interface between α_1 and α_2 includes hydrogen bonding between the N-terminal β-strands of each domain at the centre of the eight-stranded sheet, and contacts between the C-terminal ends of the helical regions and β-sheet residues in the other domain (Fig. 4*c*, Table 2). The presence of approximate dyad symmetry has previously been observed between homologous domains in a single polypeptide chain, and is termed 'intramolecular dimerization'[41]. The particular dimeric interaction seen between α_1 and α_2, involving the creation of a single β-sheet from two domains, has also been observed in a number of intermolecular dimers (concanavalin A[42], aspartate carbamoyltransferase R chains[43], alcohol dehydrogenase[55]), indicating that the α_1–α_2 domain interface could be preserved in an intermolecular dimer such as that in class II histocompatibility antigens (see below).

Pseudo-symmetry axes, domain interfaces

The screw axis relating the two immunoglobulin-like domains of HLA-A2 (146° rotation and 13 Å translation) and the approximate dyad axis relating α_1 and α_2 are neither co-linear nor intersecting. In projection, the two axes deviate by 25° from co-linearity (Fig. 5). The presence of a local dyad axis had been anticipated by earlier crystallographic work[22,28], but the relationship between the immunoglobulin-like domains in HLA was unexpected. As this arrangement of domains is found in both crystal forms of HLA-A2, it is not likely to result from crystal-packing forces. Table 2 lists the residues in each domain that are within 4 Å of another domain. The residues in contact across the α_3–β_2m interface are in the four-stranded β-sheets, but the contact pairs are not symmetrical as would be expected if the domains were related by a dyad. The β_2m domain also interacts with the central β-strands and loops of the α_1–α_2 β-sheet. The interaction of β_2m with all three domains of the heavy chain of HLA was suggested by immunological[44] and circular dichroism data[45], and results in a molecule that is shaped quite differently from an Fab[36,37].

Carbohydrate

Both human and mouse class I histocompatibility antigens have an *N*-linked, complex oligosaccharide attached at Asn 86 (ref. 46). This residue is in a loop connecting α_1 to α_2. Carbohydrate is visible in the HLA electron-density map extending away from the protein structure, but analysis of its structure awaits high resolution refinement. Several studies have demonstrated that glycosylation at Asn 86 is not necessary for the association of β_2m with the HLA heavy chain or cell-surface expression[47,48]. In murine class I antigens there is a second glycosylation site at Asn 176, and H-2L^d, H-2D^b and H-2K^d molecules have a third at Asn 256 (ref. 49). Residue 176 is located in α_2 at a

57

© 1987 Nature Publishing Group

Table 2 Inter-domain contacts

α_1	α_2
1	104
2	104
3	103, 168, 172, 180
4	101, 102, 164, 168
5	99, 100, 101, 164, 168, 171
6	98, 99
7	98, 99
8	97
9	95, 96, 97, 99
10	95, 96
11	94, 95
12	93
13	92
28	171, 179
29	179
32	171
51	174, 175
54	174
55	170
59	167, 171
63	171
70	99
74	95, 97
78	95
81	118, 123
84	123, 139, 142, 143
85	118
87	118

β_2	α_3
8	231, 244
10	234, 242
11	242
12	238
13	188, 206
22	237
26	233, 234
65	237
97	204

β_2	α_1
33	12
53	25, 33, 35
55	27
56	8, 9, 10
62	10
63	6, 27

β_2	α_2
1	119, 120
31	94, 96, 119
56	96, 97
60	96, 115, 116, 117, 121
62	96

α_3	α_1
209	29, 30
211	30
234	27, 33

α_3	α_2
181	176, 177, 179
182	180
209	179

All residues (main chain or side chain) of one domain within 4 Å of the second domain are listed. This table is included as a guide for interpreting mutagenesis and exon shuffling experiments that exchange domains between different histocompatibility molecules. The contact list may require minor updating after crystallographic refinement. The residue numbers can be assigned to their secondary structural location on the HLA-A2 structure using Fig. 3. The α_1-α_2 contacts are shown on the structure of these domains in Fig. 4c.

position structurally homologous to Asn 86 in α_1. The presence of carbohydrate at position 176 would increase the approximate symmetric nature of the α_1-α_2 'intramolecular dimer'. Residue 256 is located on the loop between strands 5 and 6 in a_3 (Fig. 3) and is accessible to solvent in HLA-A2.

Antigen recognition site

A deep groove ~25 Å long and 10 Å wide runs between the two long α-helices of the α_1 and α_2 domains (Fig. 6*a*). The sides of the groove are formed by side chains from these helices, and the bottom is formed by side chains from the central β-strands of the α_1-α_2 β-sheet. The groove is located on the top surface of the molecule (as oriented in Fig. 2*a*), and is therefore a likely candidate for the binding site for the foreign antigen that is recognized together with HLA by a T-cell receptor. The dimensions of the site are consistent with the expectation that class I molecules bind a processed antigen, probably a peptide. The site is lined with both polar and non-polar side chains. Some of these residues have been identified to be critical for T-cell recognition, as described more fully in an accompanying article[26].

In this site, we observe a large continuous region of electron density that is not accounted for by the polypeptide chain of the HLA molecule (Fig. 6*b*). This extra density is found in electron density maps calculated in both space groups. The density level is comparable to the density level of the protein, suggesting that most or all of the crystalline HLA molecules have a molecule(s) bound in the proposed antigen binding site. It seems likely that the extra density is the image of a peptide or mixture of peptides that co-purified and co-crystallized with HLA-A2. Potential sources of such peptides are: (1) protein fragments produced during papain cleavage: (2) peptides from the fetal calf serum used in the growth of JY cells: (3) processed antigenic peptides from Epstein Barr virus used in the transformation of JY cells[50]; and/or (4) endogenous 'self' peptides created during normal recycling and degradation of cellular proteins. In the absence of the knowledge of the composition of the bound molecule(s) and whether it is one species or a mixture, it is not possible to interpret its structure unambiguously at the current resolution (3.5 Å) of the electron density maps.

Other histocompatibility antigens

The structure of HLA-A2 can be used as a starting point for modelling other class I and class II histocompatibility antigens. Using the extensive sequence homology between murine and human class I molecules, it is possible to locate residues critical for murine CTL and serological epitopes on the HLA-A2 structure (see accompanying article)[26].

Class II molecules have a domain structure similar to class I, but the four domains of the class II chains are arranged on two polypeptide chains of roughly equal size that span the membrane bilayer (reviewed in ref. 51). The membrane-proximal domains of each chain are immunoglobulin-like with homology to α_3 and β_2m. The N-terminal domains of the class II chains presumably contain the binding site for antigenic peptides, and one of these domains has weak sequence homology to class I α_1 and α_2 (ref. 52). Two features of the HLA-A2 molecule suggest that class II molecules could adopt a similar structure: (1) The 'intramolecular dimerization' between the α_1 and α_2 domains resulting in an edge-to-edge β-sheet contact is similar to intermolecular dimer contacts observed in multisubunit molecules. It is therefore possible that class II N-terminal domains can form an intermolecular dimer with a structure similar to the α_1 and α_2 region of HLA-A2, thereby creating a binding site for antigens similar to that in HLA-A2. Residues in the first half of the membrane-distal domains of class II molecules have recently been shown to be critical for chain association[53], a result consistent with the proximity of the N-terminal β-strands of the class I α_1 and α_2 domains shown in Fig. 4*c*. (2) In the HLA-A2 structure, the N-terminus of β_2m is 11 Å from the C-terminus of the α_1 domain. By replacing the loop that connects α_1 and α_2 (residues 85–93) with an extended chain connecting the end of α_1 to the N-terminus of β_2m, a four-domain structure can be constructed with two domains per polypeptide chain that is similar to HLA-A2 (J. Brown *et al.*, manuscript in preparation).

These observations indicate that the HLA-A2 structure may be a good first-order model for interpreting the locations of polymorphic and functional residues on many class I and class II histocompatibility antigens.

We thank Anastasia Haykov for excellent technical assistance; Hans Bartunik and Klaus Bartels, Keith Moffat and Wilfried Schildkamp and Paul Phizackerley for help with synchrotron data collection, Mike Silver and Tom Garrett for help with the figures and our colleagues in the structural molecular biology group. We also thank Drs Dean Mann (NIH), D. Michael Strong and James Woody (U.S. Naval Research Unit) and Don Giard (MIT Cell Culture Facility) for provision of cells which made this work possible. B.S. thanks the Algerian Ministeré de l'Enseignment et de la Recherche Scientifique for a postdoctoral leave. P.J.B. held an American Cancer Society postdoctoral fellowship during part of the work. The research was supported by the NIH and the Howard Hughes Medical Institute.

Note added in proof: Alpha-carbon coordinates are being deposited in the Brookhaven Data Bank.

© 1987 Nature Publishing Group

Received 17 August; accepted 26 August, 1987.

1. Klein, J. in *Biology of the Mouse Histocompatibility Complex* (Springer, Berlin, 1975).
2. Bach, F. H. & van Rood, J. J. *New Engl. J. Med.* 295, 806-813 (1976).
3. Zinkernagel, R. M. & Doherty, P. C. *Nature* 248, 701-702 (1974).
4. Zinkernagel, R. M. *A. Rev. Microbiol.* 33, 201-213 (1979).
5. McDevitt, H. O. & Tyan, M. L. *J. exp. Med.* 128, 1-11 (1968).
6. Townsend, A. R. M. *et al. Cell* 44, 959-968 (1986).
7. Shimonkevitz, R., Kappler, J., Marrack, P. & Grey, H. *J. exp. Med.* 158, 303-316 (1983).
8. Babbit, B., Allen, G., Matsueda, E., Haber, E. & Unanue, E. *Nature* 317, 359-361 (1985).
9. Buus, S. *et al. Proc. natn. Acad. Sci. U.S.A.* 83, 3968-2971 (1986).
10. Guillet, J. G., Lai-m-Z., Briner, T. J., Smith, J. A. & Gefter, M. L. *Nature* 324, 260-263 (1986).
11. Forman, J. *Adv. Immun.* (in the press).
12. Cresswell, P., Turner, M. J. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* 70, 1503-1507 (1973).
13. Grey, A. M. *et al. J. exp. Med.* 138, 1608-1612 (1973).
14. Peterson, P. A., Rask, L. & Lindblom, J. B. *Proc. natn. Acad. Sci. U.S.A.* 71, 35-39 (1974).
15. Orr, H. T., Lopez de Castro, J. A., Lancet, D. & Strominger, J. *Proc. natn. Acad. Sci. U.S.A.* 76, 5839-5842 (1979).
16. Malissen, M., Malissen, B. & Jordan, B. R. *Proc. natn. Acad. Sci. U.S.A.* 79, 893-897 (1982).
17. Tragardh, L., Rask, L., Wiman, K., Fohlman, J. & Peterson, P. A. *Proc. natn. Acad. Sci. U.S.A.* 76, 5839-5842 (1979).
18. Peterson, P. A., Cunningham, B. A., Berggard, I. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* 69, 1697-1701 (1972).
19. Smithies, O. & Poulik, M. D. *Science* 175, 187-189 (1972).
20. Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro, J. A. & Strominger, J. L. *Nature* 282, 266-270 (1979).
21. Orr, H. T., Lopez de Castro, J. A., Lancet, D. & Strominger, J. L. *Biochemistry* 18, 5711-5719 (1979).
22. Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. *J. molec. Biol.* 186, 205-210 (1985).
23. Turner, M. J. *et al. J. biol. Chem.* 250, 4512-4519 (1975).
24. Parham, P., Alpert, B. N., Orr, H. T. & Strominger, J. L. *J. biol. Chem.* 252, 7555-7567 (1977).
25. Dausset, J. *Acta haematol.* 20, 156 (1958).
26. Bjorkman, P. J. *et al. Nature* 329, 512-518 (1987).
27. Springer, T. A. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* 73, 2481-2485 (1976).
28. Bjorkman, P. J. thesis, Harvard University (1984).
29. Durbin, R. *et al. Science* 232, 1127-1132 (1986).
30. Blum, M., Metcalf, P., Harrison, S. C. & Wiley, D. C. *J. appl. Cryst.* 20, 235-242 (1987).
31. Wang, B. C. *Meth. Enzym.* 115, 90-111 (1985).
32. Rice, D. W. *Acta crystallogr.* A37, 491-500 (1981).
33. Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S.-H. *Acta crystallogr.* 33, 800-804 (1977).
34. Bricogne, G. *Acta crystallogr.* A32, 832-847 (1976).
35. Koller, B. H. & Orr, H. T. *J. Immun.* 134, 2727-2733 (1985).
36. Marquart, M. & Deisenhofer, J. *Immun. Today* 3, 160 (1982).
37. Amzel, L. M. & Poljak, R. J. *A. Rev. Biochem.* 48, 961-997 (1979).
38. Becker, J. W. & Reeke, G. N. *Proc. natn. Acad. Sci. U.S.A.* 82, 4225-4229 (1985).
39. Vega, M. A., Bragado, R., Ezquerra, A. & Lopez de Castro, J. A. *Biochemistry* 23, 823-831 (1984).
40. Novotny, J. & Auffray, C. *Nucleic Acid Res.* 12, 243-255 (1984).
41. McLachlan, A. D. in *Protein Folding* (ed. Jaenicke, R.) 79-99 (Elsevier, North Holland, 1980).
42. Reeke, G. N., Becker, J. W. & Edelman, G. M. *J. biol. Chem.* 250, 1525-1545 (1975).
43. Krause, K. L., Volz, K. W. & Lipscomb, W. N. *J. molec. Biol.* 193, 527-553 (1987).
44. Krangel, M. S., Biddison, W. E. & Strominger, J. L. *J. Immun.* 130, 1856-1862 (1979).
45. Lancet, D., Parham, P. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* 76, 3844-3848 (1979).
46. Nathenson, S. G. & Cullen, S. E. *Biochim. biophys. Acta* 344, 1-25 (1974).
47. Ploegh, H. L., Orr, H. T. & Strominger, J. L. *J. Immun.* 126, 270-275 (1981).
48. Santos-Aguado, J., Biro, P. A., Fuhrmann, U., Strominger, J. L. & Barbosa, J. A. *Molec. cell. Biol.* 7, 982-990 (1987).
49. Kimball, E. S. & Coligan, J. E. *Contemp. Top. molec. Immun.* 9, 1-63 (1983).
50. Terhorst, C., Parham, P., Mann, D. L. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* 73, 910-944 (1976).
51. Kaufman, J. F., Auffray, C., Korman, A. J., Shackelford, D. A. & Strominger, J. L. *Cell* 36, 1-13 (1984).
52. Larhammar, D. *et al. Proc. natn. Acad. Sci. U.S.A.* 79, 3687-3691 (1982).
53. Sant, A. J., Braunstein, N. S. & Germain, R. N. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
54. Lee, B. & Richards, F. M. *J. molec. Biol.* 55, 379-400 (1971).
55. Eklund, H. *et al. J. Molec. Biol.* 102, 27-59 (1976).

The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens

P. J. Bjorkman*†‡, M. A. Saper*, B. Samraoui*, W. S. Bennett*‡, J. L. Strominger* & D. C. Wiley*§

* Department of Biochemistry and Molecular Biology, Harvard University and Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA
† Department of Medical Microbiology, Stanford University, Stanford, California 94305, USA
‡ Present addresses: Department of Medical Microbiology, Stanford University, Stanford, California 94305, USA. (P.J.B.) and Max-Planck-Institute Für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin 33, FRG (W.S.B.)
§ To whom correspondence should be addressed

Most of the polymorphic amino acids of the class I histocompatibility antigen, HLA-A2, are clustered on top of the molecule in a large groove identified as the recognition site for processed foreign antigens. Many residues critical for T-cell recognition of HLA are located in this site, in positions allowing them to serve as ligands to processed antigens. These findings have implications for how the products of the major histocompatibility complex (MHC) recognize foreign antigens.

THE three-dimensional structure of HLA-A2, a human class I histocompatibility antigen, was determined to advance our understanding of how these polymorphic membrane glycoproteins interact with foreign antigens and with receptors on T cells. To trigger lysis of a virus-infected cell, a receptor on a cytotoxic T lymphocyte (CTL) recognizes a complex between a specific viral antigen and a class I histocompatibility molecule[1]. A high percentage of CTL are also capable of lysing uninfected cells bearing a foreign (or allo) HLA molecule, a phenomenon known as alloreactivity[2,3]. The dual recognition of a foreign antigen complexed to HLA (MHC restriction) has remained puzzling, because it seems unlikely that the limited number of HLA molecules in a given individual could be structurally complementary to the huge number of foreign antigens with which they must complex. Recent evidence, however, has suggested that foreign antigens are not recognized as native three-dimensional structures by T cells, but are processed intracellularly into short peptides[4] that are presented by histocompatibility molecules. Peptides from foreign antigens have been shown to bind to purified class II histocompatibility molecules[5,6] at a single site[7] with micromolar dissociation constants[5,8]. Because peptides from a viral protein added to uninfected cells will stimulate virus-specific, class I-restricted cell lysis[9], class I molecules are also expected to bind fragments of viral proteins

59

© 1987 Nature Publishing Group

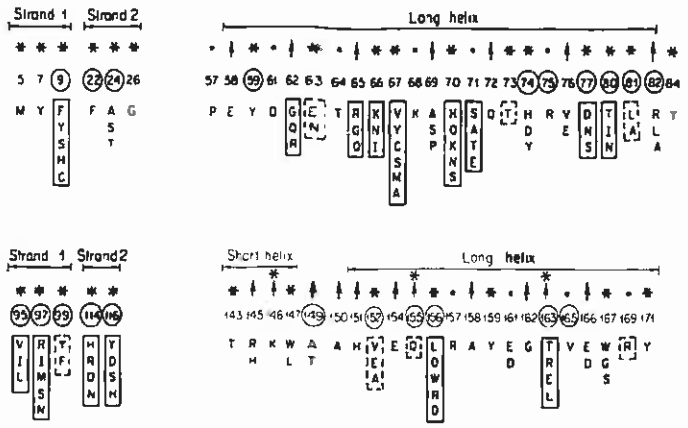


Fig. 1 Residues in the vicinity of the recognition site for processed foreign antigen. Listed are accessible residues on the two α-helices and the four central β-strands of the α_1–α_2 β-pleated sheet that could potentially interact with a bound processed antigen or a T-cell receptor bound to the antigen recognition site (see text and preceding article[12] for details). Amino acids are listed for the HLA-A2 sequence in single-letter code (top line). Amino acids found in other human class I molecules are listed below the HLA-A2 sequence. The locations of these residues on the HLA-A2 structure are noted above the residue numbers, and residues in these positions are shown on the structure of the α_1 and α_2 domains in Fig. 2. *, A residue pointing towards the site; ↑, residue on an α-helix pointing up (away from the membrane proximal immunoglobulin-like domains); ●, residue on a α-helix pointing away from the site. Residues enclosed in a rectangular box indicate a position defined as polymorphic[14,15] in human class I sequences. Residues enclosed in a dashed box indicate a position defined as polymorphic[16] in murine class I sequences. Circled residues are changed in CTL recognition mutants (see Table 1). Residues pointing towards the site are potential ligands for a bound, processed foreign antigen. Some of the ligand residues located on an α-helix might also be accessible to a T-cell receptor. The ligand residues on β-strands would probably be inaccessible to a T-cell receptor when the site is occupied. Residues pointing up on the α-helixes can potentially interact with a T-cell receptor bound to the site. Some residues are designated as pointing both into the site and up, indicating that the side chain is in an intermediate position. Residues pointing away from the site could interact with a T-cell receptor. The classification of these residues is based upon their position in the HLA-A2 structure. Amino-acid substitutions in other specificities of HLA could result in a different classification of some of these residues.

to form a complex that is recognized by CTL[10]. Some questions which remain are: how does a histocompatibility molecule bind many different peptides? Does a histocompatibility molecule impose a common conformation on peptides to facilitate complex formation and recognition by T-cell receptors? How does the T-cell receptor simultaneously recognize the peptide and the HLA molecule? Why are some peptides presented better than others, leading to a correlation between immunological response to an antigen and certain specificities of histocompatibility molecules[11]? Does alloreactive recognition of HLA involve peptide binding?

In the preceding article[12], evidence was presented that a prominent groove on the top of HLA-A2 is likely to be the foreign antigen recognition site. Its location, size and shape are consistent with the expectation that it binds a processed (for example, peptide) form of the antigen. Furthermore, in crystals of HLA-A2, the site contains a bound molecule(s) of unknown origin that may be a peptide antigen. In this report we describe the site and the location on the α_1 and α_2 domains of HLA-A2 of the polymorphic amino acids and residues critical for antibody and T-cell recognition of class I histocompatibility molecules. The observation that most polymorphic and T-cell epitope residues are positioned in this site where they could serve as ligands to a processed foreign antigen is further evidence that it serves as the foreign antigen binding site, and it will be referred to as such throughout this article. Some of the other functional residues are located on the edges of the site where they could be recognized directly by a T-cell receptor binding to the antigen recognition site on an HLA molecule.

Antigen recognition site

Figure 1 lists all of the accessible residues on the two helical regions of α_1 and α_2 and on the β-strands located between them that are in the vicinity of the binding site, and therefore likely to be important for T-cell recognition of class I histocompatibility molecules. In Fig. 2, residues of the central β-strands of the α_1–α_2 β-sheet that form the bottom of the site, and residues on the α-helices facing into the site forming its two sides are coloured red. These residues are potential ligands for antigenic peptides and are designated with an asterisk in Fig. 1. The residues on the top surfaces of the α-helices, pointing up and projecting into solution, are yellow in Fig. 2. These residues are candidates for binding directly to a T-cell receptor (Fig. 1, arrows). Amino-acid substitutions at these locations in other HLA specificities might result in side chains reaching into the site and contacting an antigenic peptide. Other residues on the helices that are solvent-accessible, but point away from the site, probably do not interact with antigenic peptides but may interact with a T-cell receptor. These are designated only by numbers in Fig. 2 and a dot in Fig. 1.

Polymorphic and conserved residues

Polymorphism at some amino-acid positions of HLA is responsible for the specificity of recognition by T cells, and for the variation in responsiveness to particular foreign antigens. A position is defined as polymorphic if its variability index (number of substitutions at a position divided by a frequency of the most common amino acid at the position) is high[13]. On the HLA-A2 structure, 15 of the 17 positions with a variability index greater than 6 (in 22 human class I sequences[14,15]) are located in the binding site region for processed antigen (residues 9, 62, 65, 66, 67, 70, 71, 77, 80 in α_1; 95, 97, 114, 116, 156, 163 in α_2). Residues 9, 95, 97, 114 and 116 are on the central β-strands of the α_1–α_2 β-sheet (Fig. 3*b*; red in Fig. 3*a*). They point up between the two helical regions, and would be buried if the binding site were occupied, suggesting that the polymorphism of these residues affects interactions with bound antigenic peptides, and not direct interactions with T-cell receptors. Residues 66, 67, 70, 77, 80 and 156 are located on the sides of the helices facing into the site (Fig. 3*c*; red in Fig. 3*a*) in positions to be ligands to foreign antigens. Residues 62, 65 and 163 are on the top face of the helices facing solvent (Fig. 3*c*, yellow in Fig. 3*a*). They are candidates for making direct contacts with the T-cell receptor. Residue 71 (blue in Fig. 3*a*) on the α_1 helix does not appear to be in a position to affect the binding of antigenic peptides. The polymorphic residues of the antigenic recognition site are enclosed in rectangles in Fig. 1 (dashed rectangles indicate positions that are polymorphic in murine, but not in human class I molecules). Two polymorphic residues (43, 45; blue in Fig. 3*a*) are not located on the helices or β-strands that form the binding site, but are found on the outside β-strand of α_1 in an area accessible to solvent. Residue 45 is immediately behind the α_1 α-helix, pointing towards it, and may be able to influence the site. The location of so many polymorphic residues close to or in the site is consistent with the genetic mapping of immune response effects (low response to specific antigens) to histocompatibility molecules[11], and with the observation that the binding of peptides from an antigen correlates with the immune response phenotype of the class II molecule for that antigen[5,6,8]

© 1987 Nature Publishing Group

Table 1 T-cell epitope residues of human and murine class I molecules

α_1 domain residue	Location on structure	Substitution	Variant (residue changed)	Classification of residue	Ref.
9	β_1**: into site	F→Y	CLA (HLA-A2.4) (9) HLA-A2.2Y (9, 43, 95, 156)	ligand	20, 30, 43
22	β_2: into site	Y→F	K^{bm8} (22–24, 30)	ligand	18
23	β_2: away from site towards β_2m	M→I	K^{bm8} (22–24, 30)	silent; see refs 22, 24	18
24	β_2: into site	E→S	K^{bm8} (22–24, 30)	ligand	18
30	Loop between β_2 and β_3	Y→N	K^{bm8} (22–24, 30)	silent; see refs 22, 24	*
40	Loop between β_3 and β_4	A→V	K^k-TNP restricted	TCR indirect effect? no affect on allo- or viral-restricted CTL	31
43	Loop between β_3 and β_4	E→R	M7 (HLA-A2) (43, 95, 156)	silent?; see refs 95, 156	20, 32
59	α-helix: into site	Y→H	HLA-B27d (59)	ligand	33
74	α-helix: into site	D→Y	HLA-B27F (74, 77, 81)	ligand	34
75	α-helix: points away from site	R→H	K^{bm8} (75, 77)	TCR	18
77	α-helix: into site	D→S	K^{bm23} (75, 77); K^{bm11} (77, 80) K^{bm3} (77, 89)	ligand	18, 34–37
		D→N	HLA-B27.2 (77, 80, 81),		
		D→S	HLA-B27.3 (77, 152)		
		D→N	HLA-B27F (74, 77, 81)		
		D→S	HLA-27.4 (77, 114, 116, 152)		
80	α-helix: into site	T→N	K^{bm11} (77, 80);	ligand	18, 35
		T→I	HLA-B27.2 (77, 80, 81)		
81	α-helix: into site	L→A	HLA-B27.2 (70, 80, 81) HLA-B27F (74, 77, 81)	ligand	34, 35
82	α-helix: points up	L→F	R8.313 (H-$2K^b$) (82)	TCR	38
89	glycosylated loop	K→A	K^{bm3} (77, 89)	silent?; see ref. 77	18
α_2 domain residues					
95	β_1: into site	V→L	M7(HLA-A2) (43, 95, 156)	ligand	32, 39
		I→V	SDM of Aw68 (95, 97)		
97	β_1: into site	M→R	SDM of Aw68 (95, 97)	ligand	39
99	β_1: into site	Y→C	KNE (HLA-A2.4) (99)	ligand	40
107	Loop between β_1 and β_2	G→W	SDM of HLA-A3, HLA-Aw68 (107)	TCR, indirect effect?	39, 41
114	β_2: into site	H→D	HLA-B27.4 (77, 114, 116, 152)	ligand	37
116	β_2: into site	Y→F	$K^{bm5.16}$ (116) $K^{bm6,7,8}$ (116, 121) HLA-B7 (CF)	ligand	18, 37, 42
		D→Y	HLA-B27.4 (77, 114, 116, 152)		
121	Loop between β_2 and β_3	C→R	$K^{bm6,7,9}$ (116, 121)	silent?; see 116	18
149	short α-helix: points up	A→T	DK1(HLA-A2) (149, 152, 156)	TCR	20, 32, 43, 44
152	α-helix: into site	E→A	K^{bm1} (152, 155, 156)	ligand/TCR	18, 20, 32, 43
		V→E	DK1 (HLA-A2) (149, 152, 156)		36, 37, 44, 45
		V→E	HLA-B27.3 (77, 152)		
		E→V	E1 (HLA-A3) (152, 156)		
		V→E	HLA-B27.4 (77, 114, 116, 152)		
155	α-helix: up and into site	R→Y	K^{bm1} (152, 155, 156)	TCR/ligand	18
156	α-helix: into site	L→Y	K^{bm1} (152, 155, 156)	ligand	18, 20, 32, 43
		L→W	DK1 (HLA-A2) (149, 152, 156)		44, 45
		L→W	M7(HLA-A2) (43, 95, 156)		
		L→Q	E1 (HLA-A3) (152, 156)		
163	α-helix: up and into site	T→A	K^{bm10} (163, 165, 173, 174)	TCR/ligand	18
165	α-helix: points away	V→M	K^{bm10} (163, 165, 173, 174)	TCR; silent? see 163	18
173	α-helix: points down	K→E	K^{bm10}(163, 165, 173, 174)	silent; see 174	18
174	α-helix: points up, not in site	N→L	K^{bm10} (163, 165, 173, 174)	TCR indirect effect?	18

Positions on α_1 and α_2 domains are listed that affect recognition by CTL in a number of Class I molecules. For each position, the location on the HLA-A2 structure (Fig. 6), the amino-acid substitution, the variant name, and the residues changed in that variant are listed. Some of the variant molecules were generated by site-directed mutagenesis (SDM). In the final column, the effect of the residue change is interpreted based on its location on the HLA structure. Residues pointing to the antigen recognition site are classified as 'ligand', and are suggested to be ligands for processed foreign antigens (see residues coloured red in Fig. 6). Residues pointing up on the central β-standards of the α_1–α_2 β-sheet would be inaccessible if the site were occupied, and are likely to affect T-cell recognition only by the way they alter (or eliminate) peptide binding. Residues on the α-helices pointing in toward the site could be ligands to bound peptides, and in some cases, could be contacted by a T-cell receptor bound to the site. Residues on the α-helices that point up (yellow in Fig. 6), or away from the site (not coloured in Fig. 6) are in positions to be contacted by a T-cell receptor and are classified as 'TCR'. In variants that contain multiple residue changes, we have interpreted certain changes as not contributing to T-cell recognition ('silent', blue in Fig. 6) and directed attention to other residue changes in the same variant. Residues outside the antigen recognition site (blue in Fig. 6) could indirectly affect T-cell receptor binding. The classification of these residues is based upon their position in the HLA-A2 structure. Amino-acid substitutions in other specificities of HLA could result in a different classification of some of these residues.

* Residue 30 substitution (G. Pfaffenbach, *et al.* unpublished data). ** β_N refers to β-sheet strand *N*.

61

© 1987 Nature Publishing Group

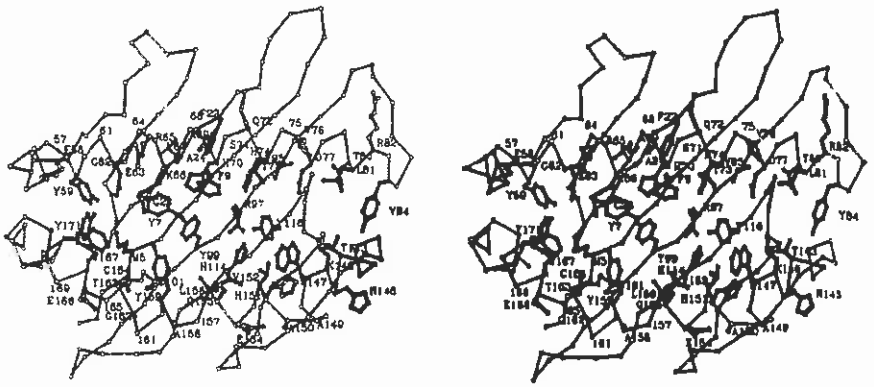


Fig. 2 Stereo view of the residues in HLA-A2 located in the vicinity of the antigen recognition site. The α-carbon backbone of the α_1 and α_2 domains is shown with side chains of the residues that are close to the site. Side chains pointing in toward the site (red) are potential ligands for bound processed antigens. Side chains on the α-helices pointing up (yellow) or away from the site (not coloured) could interact with a T-cell receptor. Positions are marked with the residue number and amino-acid type in single-letter code (see text and Fig. 1 for details).

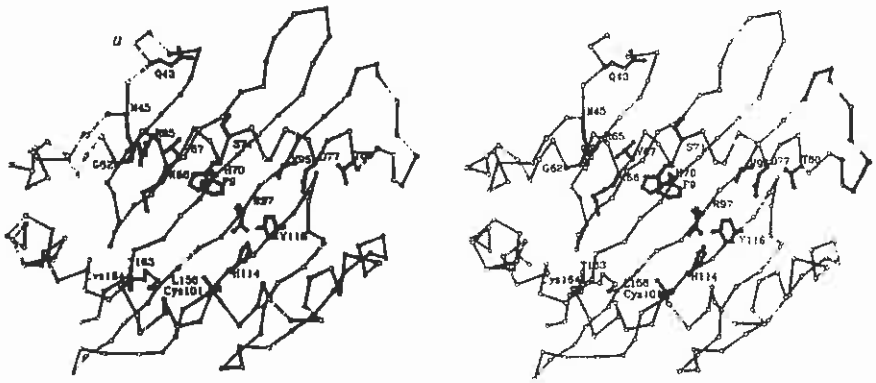


Fig. 3 Location of polymorphic amino acids in the α_1 and α_2 domains of human class I histocompatibility molecules. Polymorphic amino acids are shown highlighted on the α-carbon backbone of HLA-A2 (*a*), or schematically (*b*, *c*). Polymorphic positions are defined as those with a variability index[13] greater than 6 (using 22 human class I sequences[14,15], or 13 murine class I sequences[16]). Side chains shown are the amino acids at these positions in HLA-A2. In addition to these polymorphic amino acids, other non-conserved residues (see Fig. 4) could contribute to MHC-restricted T-cell recognition of HLA. *a*, Stereo view of polymorphic amino acids in human class I molecules. Side chains coloured are polymorphic residues in the vicinity of the antigen recognition site (Fig. 2) in positions as potential ligands to antigenic peptides (red; pointing toward site), or to T-cell receptors (yellow; pointing up from site). Residue 163, in an intermediate position, is yellow in this figure (see *c*). Side chains coloured blue are polymorphic residues either outside the antigen recognition site, or not in a position to affect the binding of antigenic peptides (see text and Figs 1, 2 for details). *b*, Schematic view of the polymorphic amino acids located in the α_1-α_2 β-pleated sheet at the bottom of the antigen recognition site. These residues point up between the α-helices and would be inaccessible if the site were occupied. Their polymorphism presumably affects interactions with bound processed foreign antigens. The β-strands are represented by an arrow in the amino to carboxy direction. Polymorphic positions are indicated by a solid circle (human) or dashed circle (murine). Polymorphic residues 114 and 116 are separated by residue 115, a conserved Gln, that points down to interact with Trp 60 in the membrane-proximal, β_2-microglobulin domain. *c*, Schematic view of the polymorphic amino acids located in the antigen recognition site on the α_1 and α_2 domain α-helixes. The helices are represented as rectangles with arrows indicating the amino to carboxy direction. Polymorphic residues pointing in, towards the antigen recognition site, are represented as circles on the right (α_1) or left (α_2) side of the rectangles. Polymorphic residues which point up into solution are depicted as circles in the centre of the rectangle. Residues 155 and 163 are depicted as pointing both up and towards the site. Residue 71 (square) points down towards the α_1-α_2 β-sheet and appears inaccessible. Polymorphic positions are indicated by a solid circle (human) or dashed circle (murine).

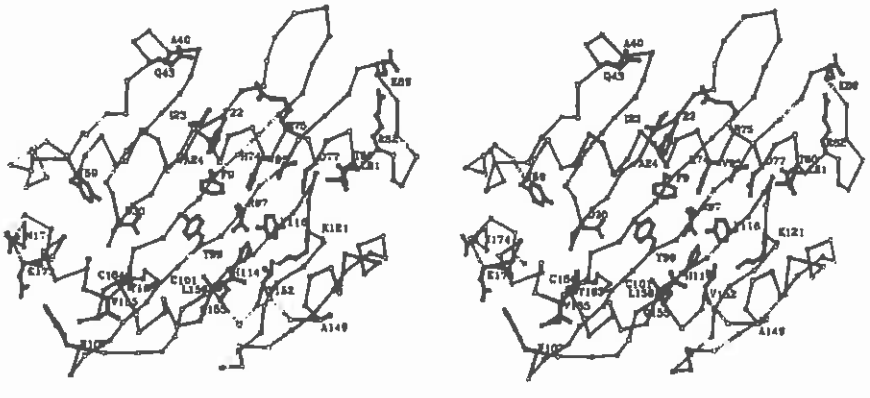


Fig. 6 Stereo view of the T-cell epitopes of human and murine class I molecules. The α-carbon backbone of the α_1 and α_2 domains is shown with side chains of those residue demonstrated to be critical for recognition by alloreactive and MHC-restricted T cells. Side chains shown are the amino acids in these positions in HLA-A2. Residues pointing toward the antigen-recognition site (red) are potential ligands for processed foreign antigens. Residues on the α-helices that point up into solution (yellow) or away from the site (not coloured) could interact with a T-cell receptor bound to the site. Other residues implicated in T cell recognition are coloured blue. Residues 155 and 163 are in intermediate positions and could have been coloured either red or yellow. (See text and Table 1 for an interpretation of the effect of residue changes on CTL recognition of class I molecules.)

© 1987 Nature Publishing Group

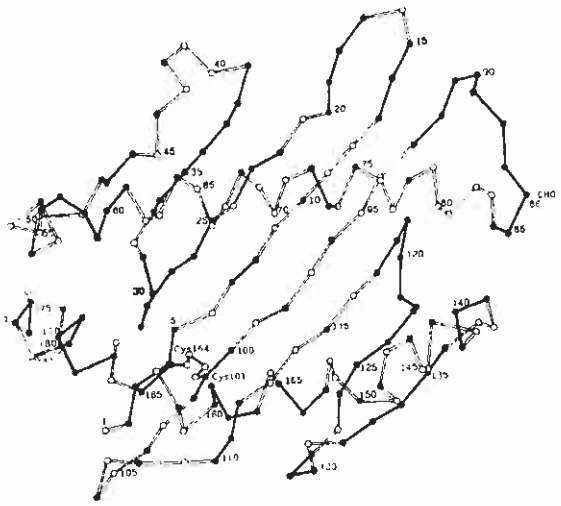


Fig. 4 Location of conserved amino acids in the α_1 and α_2 domains of human class I histocompatibility molecules (sequences from refs 14, 15). Invariant residues are shown as solid circles on the α-carbon backbone of HLA-A2. Open circles indicate positions of non-conserved (but not necessarily polymorphic, see Fig. 3) amino acids. The conserved tetrapeptide RFDS (residues 35–38) in the α_1 domain has been suggested to interact with CD8, a T-cell accessory molecule[46] based on the similarity between this sequence and the tetrapeptide involved in the cell attachment site of fibronectin[47]. The RFDS sequence is located on β-strand 3 of α_1. R35 is involved in a salt bridge to D53 of the membrane-proximal β_2-microglobulin domain. F36 is buried underneath the α_1 helix. This sequence does not appear accessible for interactions with other molecules, although conformational changes upon binding T cells cannot be ruled out. An SDGR sequence also suggested to be important for molecular interactions[46] is located on an exposed loop at 105–108, but is not conserved (SDWR on HLA-A2).

Conserved residues

Ten residues pointing into the antigenic recognition site are completely conserved in 22 human sequences[14,15] (M5, Y7, F22, G26, Y59, Y84, T143, K146, Y159, Y171). In 13 murine sequences, they are also conserved or have one occurrence of another amino acid[16]. These residues, which include five tyrosines, may recognize a constant feature of processed antigens. A pattern of every third or fourth residue is conserved on the long helix in the α_1 domain of HLA (57, 58, 61, 64, 72, 73, 75, 78, 84). These residues (excluding 78 and 84) form a completely conserved face on this helix, pointing up into solution or away from the antigen-recognition site. A similar patch (146, 150, 154, 155, 158, 162, 165, 169) is conserved on the long helix of α_2. Whether these conserved amino acids are recognized directly by constant features on T cells or are conserved to maintain the HLA structure cannot yet be determined.

The locations of invariant residues on the α_1 and α_2 domains of HLA-A2 are shown in Fig. 4, where it is evident that several loops on the molecule are highly conserved. Most of the non-conserved residues are located in and around the binding site, suggesting that the location of most of the variability in HLA molecules has been selected to generate an ability to present many different peptides.

Antibody binding sites on HLA

The residues comprising some serological epitopes were determined by comparing the pattern of substitutions in HLA subtype sequences with the distribution of a particular epitope (reviewed in ref. 17). Residues that were identified as critical for the binding of monoclonal antibodies to human or murine class I molecules are shown in Fig. 5 (see legend). All of these residues are located on the α-helices in the α_1 and α_2 domains, with the exception of E89 and W107, which are found on exposed loops. In each epitope at least one residue is solvent-exposed (62, 75, 80, 82, 89, 107, 149, 152, 161, 177) and appears accessible to direct recognition by an antibody. But one cannot exclude the possibility that some substitutions are not recognized directly, and instead cause conformational changes in other residues which affect antibody binding. None of the polymorphic residues located at the bottom of the antigen recognition site have been demonstrated to be critical for antibody recognition of class I molecules. But some of these residues, which would be buried if the site were occupied, are involved in T-cell recognition of HLA (see below). Figure 5*b* shows residues identified to be involved in antibody binding to β_2-microglobulin.

CTL mutants and variants

Specific residues involved in the recognition of murine class I molecules by cytotoxic T lymphocytes (CTL) have been identified by characterizing spontaneous mutations in the *H-2K*^{b} (class I) gene from mutants selected for skin graft incompatibility with individuals of their parental haplotype (reviewed in ref. 18). In human class I proteins, residues have been identified by isolating subtype variants from serologically-defined HLA specificities on the basis of differential reactivity with CTL (ref. 19; see ref. 20 for review). Most of the human and murine variant molecules have two or three amino-acid substitutions, and in some cases, these changes have resulted from a gene conversion event involving another class I gene[18]. Table 1 lists the residues in murine and human class I molecules that are changed in CTL variants and mutants, their location on the HLA-A2 structure, and an interpretation of the probable consequences of the substitution.

The locations of residue changes in CTL variants fall into three categories: (1) residues in the antigenic recognition site defined as potential ligands to an antigenic peptide (red in Fig. 6: 9, 22, 24, 59, 74, 77, 80, 95, 97, 99, 114, 116, 152, 156). Within this category, those residues located at the bottom of the antigenic recognition site (9, 22, 24, 95, 97, 99, 116) would be expected to be covered by a bound antigen, and are therefore likely to be distinguished by CTL only by the way they alter (or eliminate) peptide binding. (2) Residues in the antigenic recognition site that point up or away from the site (75, 82, 149, 155, 163, 165; yellow in Fig. 6). Substitutions at these positions could affect interactions with a T-cell receptor bound to the antigen recognition site. (3) Residues outside the antigen recognition site (blue in Fig. 6: 40, 43, 89, 107, 121, 173, 174). Residues 107, 173 and 174 are near one end of the site where they may affect T-cell receptor binding. The other residues are found on loops. With the exception of residue 40, these mutations (43, 121, 89) are found in variants that also have at least one other change in a residue at a potential ligand position in the antigen recognition site. Therefore they may not be residues critical for T-cell recognition (see Table 1).

Alloreactive and MHC-restricted recognition

The murine CTL variants were selected on the basis of alterations in recognition by alloreactive CTL. Some have also been tested in virus-specific, MHC-restricted killing assays[18]. In almost all cases, at least one MHC-restricted T-cell line was found whose reactivity was affected by the mutation. The residue changes reported to have the greatest effect on class I-restricted CTL are located at positions 152, 155 and 156, and 77–80 (refs 18, 20). These regions include residues that are potential ligands in the binding site.

The recognition of a foreign HLA by alloreactive T cells is thought to be a cross-reaction occurring because the structure of a self MHC molecule plus antigen (self+x) resembles a non-self (allo) MHC molecule[21]. A T-cell receptor could directly contact the polymorphic residues on an allo HLA molecule to distinguish it from self, or the receptor could recognize the

63

© 1987 Nature Publishing Group

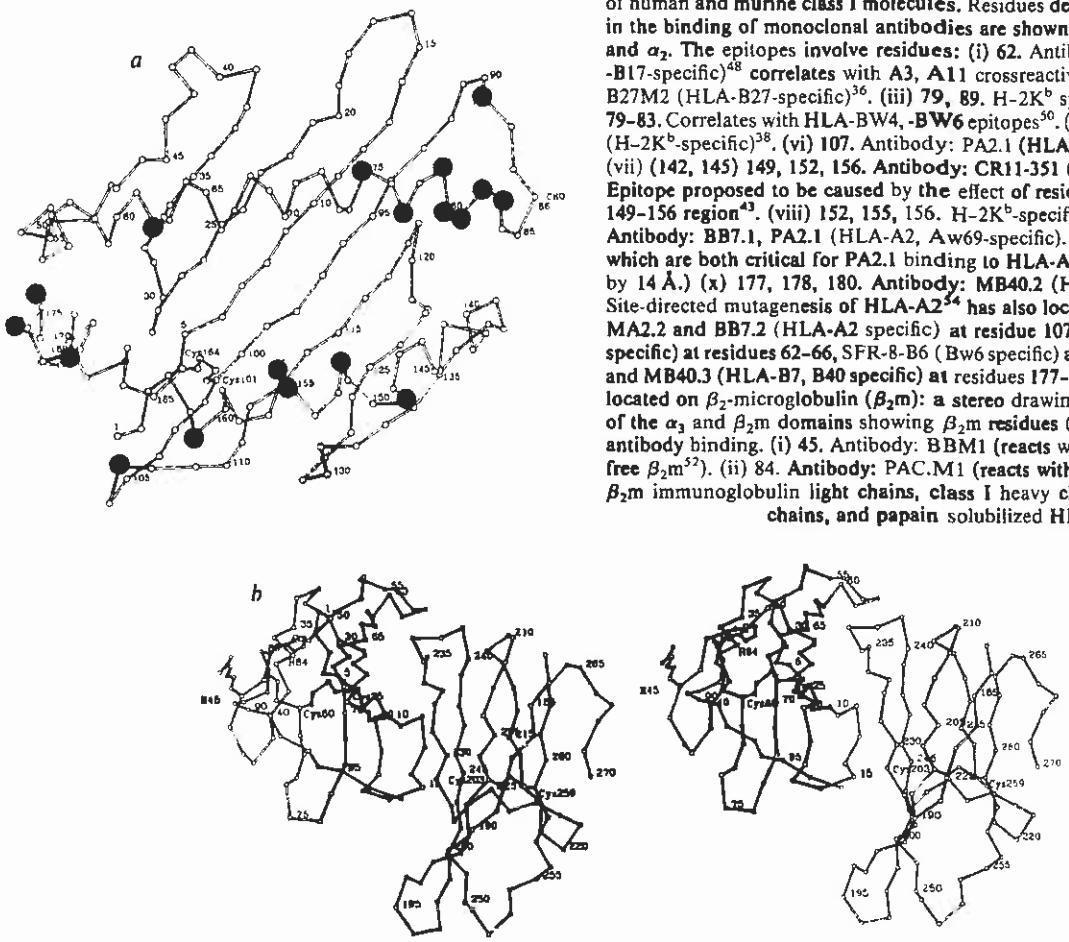


Fig. 5 Serological epitopes. *a*, Serological epitopes on the α_1 and α_2 domains of human and murine class I molecules. Residues demonstrated to be critical in the binding of monoclonal antibodies are shown with solid circles on α_1 and α_2. The epitopes involve residues: (i) 62. Antibody: MA2.1 (HLA-A2, -B17-specific)[48] correlates with A3, A11 crossreactivity[14]. (ii) 77. Antibody: B27M2 (HLA-B27-specific)[36]. (iii) 79, 89. H-2K^b specific antibodies[49]. (iv) 79–83. Correlates with HLA-BW4, -BW6 epitopes[50]. (v) 82. Antibody: EH-144 (H-2K^b-specific)[38]. (vi) 107. Antibody: PA2.1 (HLA-A2, Aw69-specific)[39,41]. (vii) (142, 145) 149, 152, 156. Antibody: CR11-351 (HLA-A2, A28-specific). Epitope proposed to be caused by the effect of residues 142 and 145 on the 149–156 region[43]. (viii) 152, 155, 156. H-2K^b-specific antibodies[49]. (ix) 161. Antibody: BB7.1, PA2.1 (HLA-A2, Aw69-specific). (Residues 107 and 161, which are both critical for PA2.1 binding to HLA-A2 (ref. 42) are separated by 14 Å.) (x) 177, 178, 180. Antibody: MB40.2 (HLA-B40, B7-specific)[51]. Site-directed mutagenesis of HLA-A2[54] has also located epitopes for PA2.1, MA2.2 and BB7.2 (HLA-A2 specific) at residue 107, MA2.1 (HLA-A2, B17 specific) at residues 62–66, SFR-8-B6 (Bw6 specific) at residue 79 and MB40.2 and MB40.3 (HLA-B7, B40 specific) at residues 177–180 (ref. 55). *b*, Epitopes located on β_2-microglobulin (β_2m): a stereo drawing of the α-carbon chain of the α_3 and β_2m domains showing β_2m residues (solid bonds) critical for antibody binding. (i) 45. Antibody: BBM1 (reacts with HLA-associated and free β_2m[52]). (ii) 84. Antibody: PAC.M1 (reacts with an epitope common to β_2m immunoglobulin light chains, class I heavy chains, class II α and β chains, and papain solubilized HLA[53]).

molecule complexed with a peptide. Although either mechanism is possible, the observation that the majority of alloreactive CTL variant residues are in the peptide-binding site, rather than on surfaces facing towards solution, suggests that alloreactivity may involve the recognition of a complex between the allo HLA and some peptide. Such a peptide could be derived from the foreign HLA itself or another molecule. Although it is conceivable that substituted residues on an allo HLA molecule are recognized on a peptide derived from itself, none of these residues (in the studies cited above) are buried in the structure of HLA-A2. Thus, it is not necessary to fragment the allo molecule to expose these residues for CTL recognition.

Recognition of foreign antigens

Several observations suggest that the prominent groove between the α_1 and α_2 domain α-helices is a foreign antigen recognition site. The site is located on a surface distal from the membrane end of the HLA molecule, in a position for recognition by receptors from the surface of another cell. The site, ~25 Å long by 10 Å wide by 11 Å deep, has a size and shape consistent with the expectation that, by analogy with class II molecules, class I molecules bind a processed antigen in the form of a peptide. Electron density representing an unknown molecule, possibly a bound peptide antigen, is found in the site in two crystal forms of HLA-A2 (see preceding article[12]). Many of the highly polymorphic amino acids that are responsible for recognition by T cells and haplotype-specific associations with antigens are located in the site. A majority of the amino-acid substitutions in murine and human class I molecules that alter recognition by T cells face into the site.

Peptide binding

Synthetic peptides bind to purified murine[5–8] and human (T. Jardetzky *et al.*, unpublished data) class II molecules, presumably mimicking processed foreign antigens. Peptide binding to class I molecules has not been reported, but synthetic peptides serve as antigens in virus-specific, class I-restricted CTL recognition[9]. Because class I and class II molecules have homologous structures[22] and T cells specific for either class of molecule use the same receptors[23,24], the type of interaction described between peptides and class II molecules is assumed to apply to peptides and class I molecules[10]. In an individual, either class of histocompatibility molecule must bind a very large number of foreign peptides, possibly in a common conformation. An α-helical conformation has been suggested for bound peptides based on the effect of amino acid substitutions on CTL recognition[25,26]. An amphipathic α-helix was suggested by analysing the sequences of CTL epitopes[27,28]. In both cases, one face of a peptide α-helix was envisioned to contact the class II molecule, leaving the other face to be contacted by a T-cell receptor.

The shape of the antigen recognition site on HLA-A2 suggests that a bound peptide would make contacts on three of its surfaces, to the bottom and two sides of the groove, rather than to one surface. The site contains a number of tyrosine and many polar or charged residues, as well as non-polar amino acids (Figs 1, 2), obviating a need for strict amphipathicity of a bound peptide. The site would accommodate an α-helical peptide of ~20 residues, or an extended polypeptide chain of ~eight residues. If the actual conformation of the bound peptide were bent, or only partly helical with unfolded ends, it could have an intermediate number of residues. One end of the groove

© 1987 Nature Publishing Group

(near amino acid 77) appears open and may allow longer antigens to extend out of the site in that direction. The other end contains four conserved residues (M5, Y5, Y59, Y159) with three tyrosine hydroxyl groups directed into the site. These conserved potential ligand residues could recognize a common feature of processed antigens, such as the main chain atoms of a peptide, a common class of terminal residue, or a chemical adduct associated with antigen processing[29].

In a recent report, residues that contact the T cell and residues that contact a class II molecule were defined on the lysozyme 52–61 peptide[26]. When the peptide was modelled as an α-helix and energy-minimized, a conformation was found in which residues important for T-cell contact oriented in one direction, and residues important for contact to the class II oriented in the opposite direction. This segregation of residues suggested that the lysozyme 52–61 peptide may bind to a class II molecule as an α-helix with the T-cell contact residues facing solvent[26]. We suggest that as two of the class II contact residues on the peptide (residues 52 and 61) are 10 residues (2.5 turns) apart, they would be on opposite sides of a standard α-helix. If class II molecules have an antigen binding groove like that on HLA-A2 (see preceding article[12]), then lysozyme residues 52 and 61 might contact opposite sides of that groove; residue 52 to the α-helix on one domain and residues 61 to the α-helix on the other domain (J. Brown *et al.*, manuscript in preparation).

The molecule or mixture of molecules bound to crystalline HLA-A2 might be processed peptides, but their composition and conformation cannot be unambiguously determined[12]. However, finding molecules bound to the antigen recognition site in crystals of HLA-A2 implies that they remained bound to HLA during the entire purification and crystallization procedure; an observation consistent with the slow kinetic off-rates (of the order of days) reported for peptides bound to class II molecules[2,6] and may indicate that a conformational change is required to release bound peptide. This may suggest that a peptide is always bound to histocompatibility antigens, and that its displacement is an essential feature of cellular immunity.

We thank Tom Garrett for help with the figures and our colleagues at Harvard and Stanford for helpful discussions. P.J.B. thanks Mark Davis for support and laboratory space during the year 1986–1987. B.S. thanks the Algerian Ministeré de l'Enseignement et de la Recherche Scientifique for a postdoctoral leave. P.J.B. held an American Cancer Society postdoctoral fellowship during part of the work. The research was supported by the NIH and the Howard Hughes Medical Institute.

Received 17 August; accepted 27 August, 1987.

1. Zinkernagel, R. M. & Doherty, P. C. *Nature* **248**, 701–702 (1974).
2. Klein, J. in *Biology of the Mouse Histocompatibility Complex* (Springer, Berlin, 1975).
3. Bach, F. H. & van Rood, J. J. *New Engl. J. Med.* 295, 806–813 (1976).
4. Shimonkevitz, R., Kappler, J., Marack, P. & Grey, H. *J. exp. Med.* **158**, 303–316 (1983).
5. Babbitt, B., Allen, G., Matsueda, E., Haber, E. & Unanue, E. *Nature* **317**, 359–361 (1985).
6. Buus, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3968–3971 (1986).
7. Guillet, J. G., Lai, M.-Z., Briner, T. J., Smith, J. A. & Gefter, M. L. *Nature* **324**, 260–263 (1986).
8. Buus, S., Sette, A., Colon, S. M., Jenis, D. M. & Grey, H. M. *Cell* **47**, 1071–1077 (1986).
9. Townsend, A. R. M. *et al. Cell* **44**, 959–968 (1986).
10. Forman, T. *Adv. Immun.* (in the press)
11. McDevitt, H. O. & Tyan, M. L. *J. exp. Med.* **128**, 1–11 (1968).
12. Bjorkman, P. J. *et al. Nature*, **329**, 506–512 (1987)
13. Wu, T. T. & Kabat, E. *J. exp. Med.* **132**, 211–250 (1970).
14. Cowan, E. P., Jelachich, M. L., Biddison, W. E. & Coligan, J. E. *Immunogenetics* **25**, 241–250 (1987).
15. Gussow, D. *et al. Immunogenetics* **25**, 313–322 (1987).
16. *Immun. Today*, Centre page diagram, 13 (Elsevier, Cambridge, U.K.)
17. Parham, P. *Immun. Rev.* **6**, 153–178 (1987).
18. Nathenson, S. G., Geliebter, J., Pfaffenbach, G. M. & Zeff, R. A. *A. Rev. Immun.* **4**, 471–502 (1986).
19. Biddison, W. E., Ward, F. E., Shearer, G. M. & Shaw, S. *J. Immun.* **124**, 548–552 (1980).
20. Lopez de Castro, J., Barbosa, J. A., Krangel, M. S., Biro, A. & Strominger, J. L. *Immunol. Rev.* **85**, 149–168 (1985).
21. Bevan, M. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2094–2098 (1977).
22. Kaufman, J. F., Auffray, C., Korman, A. J., Shackelford, D. A. & Strominger, J. L. *Cell* **36**, 1–13 (1984).
23. Rupp, R., Acha-Orbea, H., Hengartner, H., Zinkernagel, R. & Joho, R. *Nature* **315**, 425–427 (1985).
24. Marrack, P. & Kappler, J. *Adv. Immun.* **38**, 1–30 (1986).
25. Watts, T. H., Gariepy, J., Schoolnik, G. K. & McConnell, H. M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5480–5484 (1985).
26. Allen, P. M. *et al. Nature* **327**, 713–715 (1987).
27. Delisi, C. & Berzofsky, J. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7048–7052 (1985).
28. Berkower, I., Buckenmeyer, G. K. & Berzofsky, J. A. *J. Immun.* **136**, 2498–2503 (1986).
29. Nairn, R., Spengler, M. L., Hoffman, M. D., Solvay, M. J. & Thomas, D. W. *J. Immun.* **133**, 3225–3234 (1984).
30. Ezquerra, A. *et al. J. Immun.* **137**, 1642–1649 (1986).
31. Karmann, G. *et al. Immunogenetics* **22**, 35–38 (1985).
32. Mattson, D. H. *et al. Immunogenetics* (in the press).
33. Rojo, S., Aparicio, P., Hansen, J., Choos, S. Y. & Lopez de Castro, J. A. *J. Immun.* (In the press).
34. Rojo, S., Aparicio, P., Choo, S. Y., Hansen, J. A. & Lopez de Castro, J. A. *J. Immun.* **139**, 831–836 (1987).
35. Vega, M. A. *et al. Proc. natn. Acad Sci. U.S.A.* **82**, 7394–7398 (1985).
36. Vega, M. A. *et al. J. Immun.* **135**, 3323–3332 (1985).
37. Vega, M. A., Bragado, R., Ivanyi, P., Palsez, J. L. & Lopez de Castro, J. A. *J. Immun.* **137**, 3557–3565 (1986).
38. Nakagawa, M., Zeff, R. A., Geier, S. S., Bluestone, J. A. & Nathenson, S. G. *Immunogenetics* **24**, 381–385 (1986).
39. Salter, R. D., Clayberger, C., Lomen, C. E., Krensky, A. M. & Parham, P. *J. exp. Med.* **166**, 283–288 (1987).
40. Domenech, N., Ezquerra, A., Castano, R. & Lopez de Castro, J. A. *J. Immun.* (submitted).
41. Layet, C. *et al. J. Immun.* **138**, 2197–2201 (1987).
42. Taketani, S., Krangel, M. S., Pious, D. & Strominger, J. L. *J. Immun.* **131**, 2935–2938 (1983).
43. Holmes, N., Ennis, P., Wan, A. M., Denney, D. W. & Parham, P. *J. Immun.* **139**, 936–941 (1987).
44. Krangel, M. S., Biddison, W. E. & Strominger, J. L. *J. Immun.* **130**, 1856–1862 (1983).
45. Cowan, E. P., Jordan, B. R. & Coligan, J. E. *J. Immun.* **135**, 2835–2841 (1985).
46. Auffray, C. & Novotny, J. *hum. Immun.* **15**, 381–390 (1986).
47. Yamada, K. M. & Kennedy, D. W. *J. cell. Biochem.* **28**, 99–104 (1985).
48. Ways, J. P. & Parham, P. *Biochem. J.* **216**, 423–432 (1983).
49. Hammerling, G. J., Rusch, E., Tada, N., Kimura, S. & Hammerling, U. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4737–4741 (1982).
50. Wan, A. M., Ennis, P., Parham, P. & Holmes, N. *J. Immun* **137**, 3671–3674 (1986).
51. Ways, J. P., Lawlor, D. A., Wan, A. M. & Parham, P. *Immunogenetics* **25**, 323–328 (1987).
52. Parham, P., Androlewicz, M. J., Holmes, N. J. & Rothenberg, B. E. *J. biol. Chem.* **258**, 6179–6186 (1983).
53. Parham, P., Lutz, P. M. & Creaswell, P. *Immunogenetics* **25**, (in the press).
54. Santos-Aguado, J., Biro, P. A., Fuhrmann, U., Strominger, J. L. & Barbosa, J. A. *Molec. Cell Biol.* 7, 982–990 (1987).
55. Santos-Aguado, J., Barbosa, J. A., Mentzer, S., Burakoff, S., Biro, P. A. & Strominger, J. L. *Abstracts, Fifth HLA and H-2 Cloning Workshop*, Geneva, Switzerland (1986) and manuscript in preparation.
56. Maryanski, J. L., Pala, P., Corradin, G., Jordon, B. R. & Cerottini, J.-C. *Nature* **324**, 578–579 (1986).

65

© 1987 Nature Publishing Group

MHC class II molecules

In most tissue transplantations, MHC class I molecules are first matched, leaving MHC class II molecules in a less important position. Instead of one light chain and one heavy chain in the case of MHC class I molecules, MHC class II molecules have two nearly similar size chains both are membrane bound. As a result, it was more difficult to isolate and cyrstallize the extra-cellular portions of the MHC class II molecule.

To avoid confusion with the MHC class I molecule nomenclature, one of the systems of identifying the domains of MHC class II molecules is as follows:

A chain:	A1	A2	A-membrane	A-cytoplasmic
B chain:	B1	B2	B-membrane	B-cytoplasmic

Both A2 and B2 domains have the standard immunoglobulin fold binding to each other, similar to α_3 and β_2 domains of the MHC class I molecule. Also, the A1 and B1 domains fold similarly as the α_1 and α_2 domains of the MHC class I molecule. However, since A1 and B1 are not covalently linked to each other, the two bent alpha-helices sitting on the beta-sheet formed by eight anti-parallel beta-strands are somewhat more flexible. Therefore, the ends of the groove between these alpha-helices are not closed. Longer processed peptides can be bound into the groove with peptide ends hanging out at both ends of the groove. Usually nine amino acid residues sit in the middle of the groove similar to the processed peptide in MHC class I molecules. On the other hand, the total length of processed peptides binding to MHC class II molecules is around 15 amino acid residues, varying from 13 to 25.

The processing of the antigen peptides by these two different MHC class I and II molecules is summarized on the next page. In that description, the MHC class II molecule chains are designated as α and β instead of A and B as mentioned above. The subsequent biological processes are complicated and will be discussed in detail later.

The detailed three dimensional structure of the extra-cellular portion of the MHC class II molecule has also been determined (Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature, 364, 33-39).

Table 2. Characteristics of the antigen processing pathways

Characteristic	MHC-I pathway	MHC-II pathway
Composition of the stable peptide-MHC complex	Polymorphic chain α and β2 microglobulin, peptide bound to α chain	Polymorphic chains α and β, peptide binds to both
Types of antigen presenting cells (APC)	All nucleated cells	Dendritic cells, mononuclear phagocytes, B lymphocytes, some endothelial cells, epithelium of thymus
T lymphocytes able to respond	Cytotoxic T lymphocytes (CD8+)	Helper T lymphocytes (CD4+)
Origin of antigenic proteins	cytosolic proteins (mostly synthetized by the cell; may also enter from the extracellular medium via phagosomes)	Proteins present in endosomes or lysosomes (mostly internalized from extracellular medium)
Enzymes responsible for peptide generation	Cytosolic proteasome	Proteases from endosomes and lysosomes (for instance, cathepsin)
Location of loading the peptide on the MHC molecule	Endoplasmic reticulum	Specialized vesicular compartment
Molecules implicated in transporting the peptides and loading them on the MHC molecules	TAP (transporter associated with antigen processing)	DM, invariant chain

80. Lewis, R. S. & Anders, E. *Lunar planet. Sci.* **XIX**, 679–680 (1988).
81. Nichols, R. H. Jr *et al. Geochim. cosmochim. Acta* **55**, 2921–2936 (1991).
82. Verchovsky, A. B., Ott, U., Russell S. S., Fisenko, A. V. & Shukolyukov, Yu. A. *Lunar planet. Sci.* **XXIII**, 1467–1468 (1992).
83. Huss, G. R. & Lewis, R. S. *Meteoritics* **28** (in the press).
84. Alexander, C. M. O'D., Arden, J. W., Ash, R. D. & Pillinger, C. T. *Earth planet. Sci. Lett.* **99**, 220–228 (1990).
85. Huss, G. R. *Nature* **347**, 159–162 (1990).
86. Ott, U., Löhr, H. P. & Begemann, F. *Meteoritics* **26**, 382 (1991).
87. Alexander, C. M. O'D., Swan P. & Walker, R. M. *Nature* **348**, 715–717 (1990).
88. Tang, M. & Anders, E. *Astrophys. J.* **335**, L31–L34 (1988).
89. Beer, H. *Astrophys. J.* **375**, 823–832 (1991).
90. Frenklach, M., Carmer, C. S. & Feigelson E. D. *Nature* **339**, 196–198 (1989).
91. Lodders, K. & Fegley, B. Jr *Meteoritics* **27**, 250–251 (1992).
92. Bernatowicz, T. J., Amari S., Zinner E. K. & Lewis, R. S. *Astrophys. J.* **373**, L73–L76 (1991).
93. Bernatowicz, T. J., Amari, S. & Lewis, R. S. *Lunar planet. Sci.* **XXIII**, 91–92 (1992).
94. Alexander, C. M. O'D., Prombo, C. A., Swan, P. D. & Walker, R. M. *Lunar planet. Sci.* **XXIII**, 5–6 (1991).
95. Lee, M. R., Russell, S. S., Arden, J. W. & Pillinger, C. T. *Meteoritics* **27**, 248–249 (1992).
96. Zinner, E. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 956–983 (Univ. of Arizona Press, Tucson, 1988)
97. Prombo, C. A. & Clayton, R. N. *Science* **230**, 935–937 (1985).
98. Sugiura, N. & Hashizume, K. *Earth planet. Sci. Lett.* **111**, 441–454 (1992).
99. Rotaru, M., Birck, J. L. & Allègre C. J. *Nature* **358**, 465–470 (1992).
100. Huss, G. R., Hutcheon, I. D., Wasserburg, G. J. & Stone, J. *Lunar planet. Sci.* **XXIII**, 563–564 (1992).
101. Nittler, L. R., Walker, R. M. & Zinner, E. *Lunar planet. Sci.* **XXIV**, 1087–1088 (1993).
102. McCulloch, M. T., & Wasserburg, G. J. *Astrophys. J.* **220**, L15–L19 (1978).

ACKNOWLEDGEMENTS. I thank V. V. Smith for providing Fig. 2, S. Amari and E. K. Zinner for providing Figs 4–6, M. Arnould, F. Begemann, T. J. Bernatowicz, R. Gallino, P. Hoppe, A. Jorissen and E. K. Zinner for discussions and R. S. Lewis and I. P. Wright for constructive reviews and the MacDonnell Center for the Space Sciences for its hospitality during revision of the manuscript.

Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1

Jerry H. Brown*, Theodore S. Jardetzky, Joan C. Gorga†, Lawrence J. Stern*, Robert G. Urban, Jack L. Strominger & Don C. Wiley*†

Department of Biochemistry and Molecular Biology, * Howard Hughes Medical Institute, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA
† Department of Pediatrics, Children's Hospital of Pittsburgh, 6130 Rangos Research Building, 3705 Fifth Avenue, Pittsburgh, Pennsylvania 15213, USA

The three-dimensional structure of the class II histocompatibility glycoprotein HLA-DR1 from human B-cell membranes has been determined by X-ray crystallography and is similar to that of class I HLA. Peptides are bound in an extended conformation that projects from both ends of an 'open-ended' antigen-binding groove. A prominent non-polar pocket into which an 'anchoring' peptide side chain fits is near one end of the binding groove. A dimer of the class II $\alpha\beta$ heterodimers is seen in the crystal forms of HLA-DR1, suggesting class II HLA dimerization as a mechanism for initiating the cytoplasmic signalling events in T-cell activation.

CLASS II human leukocyte antigens (HLA) are cell surface $\alpha\beta$ heterodimers ($M_r \sim 60{,}000$) expressed primarily by specialized antigen-presenting cells such as macrophages, dendritic cells and B lymphocytes. Class II HLA presents peptides derived from extracellular antigens (refs 1, 2 but see refs 3, 4) to helper T cells as part of the mechanism for identifying foreign antigens and producing an immune response.

Class II HLA is polymorphic in the human population[5] and different types selectively bind different sets of peptides[1–3]. The HLA polymorphism appears to be responsible for variations in the immune response of different individuals to different antigens, and may contribute to the susceptibility to disease and autoimmune disorders[6,7].

Crystals have been grown from a number of class II HLA-DR (ref. 8) allelic variants after purification by immunoaffinity chromatography in detergent complexes and removal of the transmembrane region and cytoplasmic domain by papain digestion[9] and from a soluble construct of HLA-DR1 expressed in insect cells[10].

The three-dimensional structure of the extracellular portion of HLA-DR1 (DRA/DRB1*0101) has now been determined using three different crystal forms. In each crystal form, DR1 crystallizes as a dimer of the DR1 $\alpha\beta$ heterodimer. The class II structure is similar to that of the class I HLA molecule[11] and to a hypothetical model of the class II binding site based on the class I structure[12]. Structural differences from class I HLA include an altered position of the immunoglobulin-like β_2 domain of HLA-DR1 relative to that of the α_3 domain of class I HLA, and detailed changes in the peptide binding site so that its ends are more open. The peptide-binding site observed in the crystal structure contains electron density from the collection of peptides copurified with DR1 from human lymphoblastoid cells[4]. These peptides are bound as straight, extended chains that project out of both ends of the site. At one position of the peptide a side chain extends into a distinct pocket in the binding site. The main chain of the bound peptide contacts conserved, hydrogen-bonding residues along the HLA helices.

The presence of a parallel dimer of the $\alpha\beta$ heterodimers in crystals suggests the capacity of class II HLA molecules to dimerize before, or more probably during, recognition by helper T cells. Such a dimer of dimers could lead to an increased affinity for the CD4 coreceptor and to the crosslinking of T cell receptors to initiate cytoplasmic signalling pathways.

Structure determination

The three-dimensional structure of the extracellular portion of DR1 was determined to 3.3 Å resolution by combining X-ray diffraction data from three different types of crystals containing DR1: (1) Papain cleaved DR1 from human B lymphoblastoid cells carrying a mixture of endogenous peptides[8,9]; (2) the same DR1-peptide complex plus the 28K superantigen *Staphylococcus aureus* enterotoxin B (T.S.J. *et al.*, unpublished results); and (3) a complex of soluble DR1 expressed in insect cells with an antigenic peptide from the influenza virus haemagglutinin, HA 306-318 (ref. 10) (Table 1 legend).

An initial model of the polypeptide chain was determined from

†To whom correspondence should be addressed.

© 1993 Nature Publishing Group

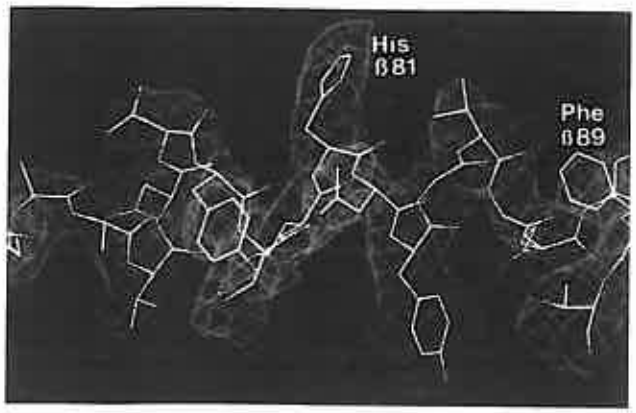


FIG. 1 Sixfold iteratively averaged 3.3 Å 'omit' electron density of the $\beta 1$ domain helix. About 83% of the polypeptide side chains show electron density of similar quality. Residues α 38, 46, 47, 50, 57, 126 and β 59, 64, 189 show no side-chain electron density, whereas residues α 1, 51, 52, 53, 132, 143, 181, β 42, 66, 68, 69, 70, 71, 72, 94, 108, 144 presently have ambiguous electron density.

an electron density map at 4.2 Å resolution, using phases from a single isomorphous heavy-atom derivative (0.12 mM K_2HgI_4, 0.3 mM KI) and iterative, fourfold, real-space averaging[13] of the first two crystal forms. Fourfold averaging was possible because each of the two crystals contained two crystallographically independent copies of the α-β heterodimer related by a non-crystallographic 2-fold symmetry axis (the symmetry axis in the dimer of heterodimers, Fig. 5*a*). Phases were improved by model building and iterative sixfold real-space averaging about the non-crystallographic 2-fold symmetry axes of all three crystal forms. Real-space averaged 'omit'[14] maps, now at 3.3 Å resolution, show excellent density for 83% of the polypeptide side chains (Fig. 1). Nine side chains show no electron density. Seventeen residues have ambiguous electron density, including β-chain residues 68–72 which have been provisionally modelled as helical. Though electron density is observed to emanate from each of the three oligosaccharide attachment sites (most extensively at α 118), it has not yet been interpreted as a molecular model. Refinement of DR1 in the three crystal forms is in progress.

Structural comparison with class I HLA

Overall the structure of DR1 is similar to that of class I HLA (Fig. 2*a*). The two α-chain domains, α_1 and α_2, of DR1 superimpose closely on the corresponding α_1 domain and the β_2-microglobulin subunit, respectively, of class I HLA. The two β-chain domains, β_1 and β_2, of DR1 superimpose on the α_2 and less closely on the α_3 domains, respectively, of class I HLA. The β_2 domain of DR1 is offset relative to the HLA-A2, class I α_3 domain by about 15° (Fig. 2*a*), but the packing between the class II immunoglobulin domains still resembles the packing in class I more than the packing in an antibody.

The peptide-binding groove of both class I and class II HLA is composed of eight strands of antiparallel β-sheet as a floor and two antiparallel helical regions as the sides. The similarity between the two structures extends even to the location of β-bulges in the β-strands and the locations of kinks in the helices (Fig. 2*b*). With the exception of the sequence alignment in strand 4 (residues α 37 to α 43) of the α_1 domain, a hypothetical model of class II (ref. 12) based on the class I HLA-A2 structure is correct. (A number of segments deemed unpredicable based on sequences and the class I structure have a different structure in class II antigens.)

Differences in the helical regions of class II HLA relative to class I may determine why class I HLA binds primarily short, nonameric peptides[15], whereas class II HLA binds longer, 15–

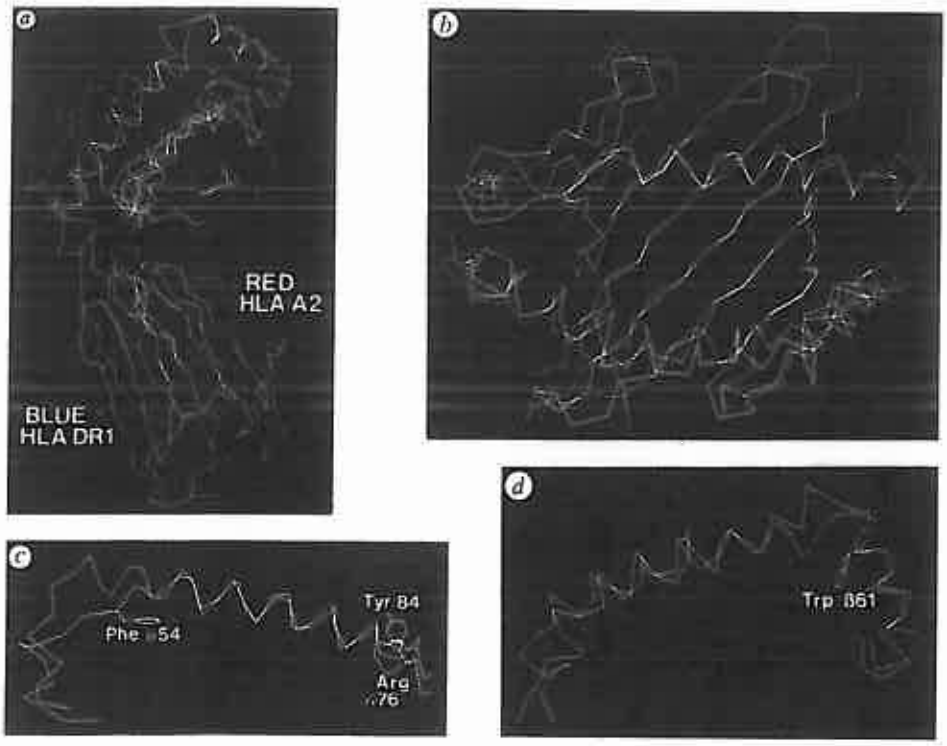


FIG. 2 Class II HLA-DR1 (blue) superimposed on class I HLA-A2 (red). *a*, One DR1 $\alpha\beta$ heterodimer and one HLA-A2 molecule[34]. The β_2 DR1 domain is offset from its homologue the α_3 HLA-A2 domain by about 15°. *b*, 'Top' view shows the overall similarity of the peptide binding sites. HLA-A2 α_1 and DR1 α_1 domains' helices are at the top, those of HLA-A2 α_2 and DR1 β_1 are at the bottom. *c*, 'Side' view of α_1 domain helical regions showing differences at both ends. On the left (also seen in *b*) DR1 has an extended chain where the HLA-A2 has helical residues. On the right, DR1 is about 2.0 Å lower (at Arg α76) than HLA-A2 Tyr 84. *d*, 'Side' view of the helical regions of the DR1 β_1 and HLA-A2 α_2 domains. DR1 has a more prominent kink and Trp β 61 superimposes almost exactly on HLA-A2 Trp 147. α-Carbon coordinates were superimposed by the program FITATOM (W. Kabsch) based on the central six β strands and two long helices of the peptide-binding site.

© 1993 Nature Publishing Group

24 residue, peptides[1-4]. Two turns near the amino terminus of the class I α_1 domain helical region are replaced by a stretch of extended chain in class II (Fig. 2*b*, *c*). The carboxy-terminal end (right, Fig. 2*c*) of the class II α_1 helical region bends more towards the floor of the site such that arginine, α 76, is 2.0 Å 'lower' in the site than the homologous class I residue tyrosine 84. These alterations help 'open' the peptide-binding site of class II HLA, so that long peptides can be accommodated. There is also a slightly more protuberant kink in the β_1 helical region of DR1 (right, Fig. 2*d*), as predicted[16], and this helical region ends (left, Fig. 2*d*) in an extended chain in DR1 rather than a kink and short helix as in class I HLA. The regions that differ in secondary structure between class I and class II also differ in their patterns of polymorphism, being relatively conserved in class I and relatively polymorphic in class II[12].

In addition to these differences in secondary structure between the class I and class II binding groove, a number of specific amino-acid side chains that 'close' the ends of the class I site are not present in the class II site. The conserved tyrosines (7, 59, 171) that bind the peptide N terminus and the conserved residues (Y84, K146, T143) that bind the peptide C terminus in the class I structure[17-19] are not found in class II sequences. Furthermore, a nearly conserved tryptophan (167) and salt bridge (Glu 55 to Arg 170) which 'close' the 'left' (Fig. 2) end of the class I groove[20] are absent in class II sequences. The significantly smaller side chain (glycine or valine) at β 86 compared to the class I homologue tyrosine 171 helps create a deep pocket in the DR1 groove surrounded by predominately non-polar residues of both the α_1 and β_1 domains (left in Fig. 3). This pocket, which provides the binding site for a peptide 'anchor' side chain in DR1, is composed of class II residues homologous to class I residues beyond the 'left' end of the class I binding groove.

Residues in the binding site

The distribution of conserved and polymorphic positions in the class II peptide-binding site suggests how conserved class II residues contact main chain atoms of bound peptides and where polymorphic pockets might bind the side chains of peptides. Figure 3 shows that three conserved asparagine residues, 62 and 69 in the α_1 helical region and 82 in the β_1 helical region are positioned such that their γ amide groups might hydrogen bond with the main-chain peptide atoms of bound peptides. Mutation of β 82 to serine affects intracellular transport and cell surface expression of murine class II molecules, consistent with this observation[21]. Conserved tryptophan β 61 is placed so that its indole nitrogen could donate a hydrogen bond to a main chain carbonyl group of a bound peptide in exact homology to the location and function of tryptophan 147 of class I HLA[17]. The orientations of the murine I-A^k equivalents of α_1 helical residues 57-75 were previously deduced from alanine scanning mutagenesis[22].

Several clusters of polymorphic residues, spaced across the binding site, appear to provide the mechanism by which different types of class II HLA selectively bind peptides of different sequences[23-26]. The first cluster (α 31, 52; β 86, 89, 90) appears to provide variability to a peptide side chain pocket composed

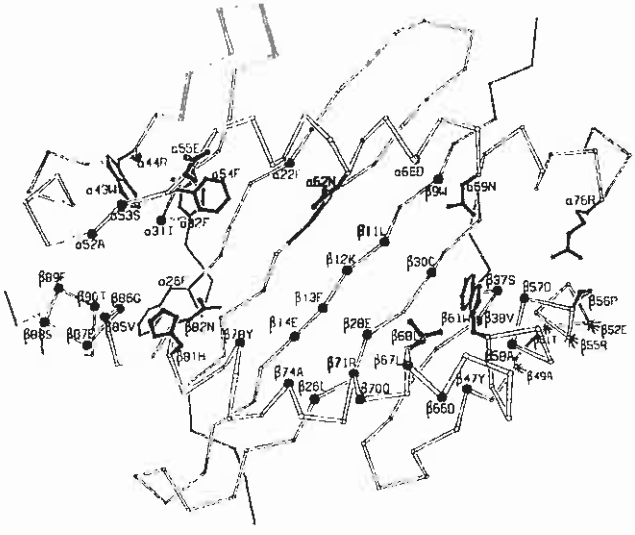


FIG. 3 The class II α_1 and β_1 domains, showing all polymorphic positions (dark circles, POL > 2[12]) and those conserved positions facing the antigen-binding groove (all atoms of side chain shown). Amino-acid residues, in one letter code, and positions are given for DR1 (DRA/DRB1*0101). The antigen-binding groove of DR1 includes the side chains of residues 7, 9, 11, 22, 24, 26, 31, 32, 43, 51, 53, 54, 55, 58, 62, 65, 68, 69, 72, and 76 of the α chain, and residues 9, 11, 13, 28, 30, 32, 37, 38, 47, 56, 60, 61, 65, 68, 70, 71, 74, 78, 81, 82, 85, 86, 88, and 89 of the β chain. Residues whose side chains are in the dimer interface (see text and Fig. 5) are indicated by 'starbursts'.

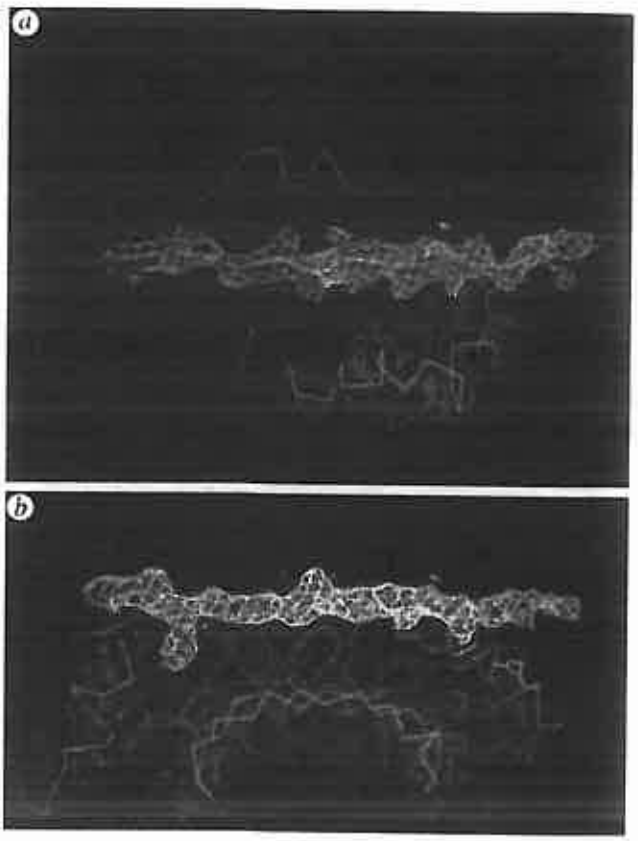


FIG. 4 Peptides bound to crystalline DR1 from human cells. Red, Electron density of the collection of endogenous bound peptides; Blue, van der Waals surface of DR1. *a*, 'Top' view, probably as seen by T cells; α_1 domain helix on top, β_1 domain helix on bottom; peptide N terminus at left. An extended, 15 amino-acid long, α-carbon, peptide backbone is placed in the electron density for scale. The side chains shown at the right-hand end of the site are arginine α76 and aspartic acid β 57 which form a salt bridge under the exiting peptide. *b*, 'Side' view, β_1 domain helix removed for clarity, β-sheet shown below and α_1 domain helix shown behind peptide electron density. The prominent downward facing peptide side chain electron density near the left end of the peptide is located about one peptide residue to the left of where the N terminus of peptides bound to class I HLA are positioned. Electron density map is fourfold iteratively averaged, 3.3 Å resolution, using DR1-LG2 and DR1-SEB crystal forms (see Table 1 legend) and $2F_o$–F_c coefficients. Density within 1.5 Å of provisional peptide atoms is included.

© 1993 Nature Publishing Group

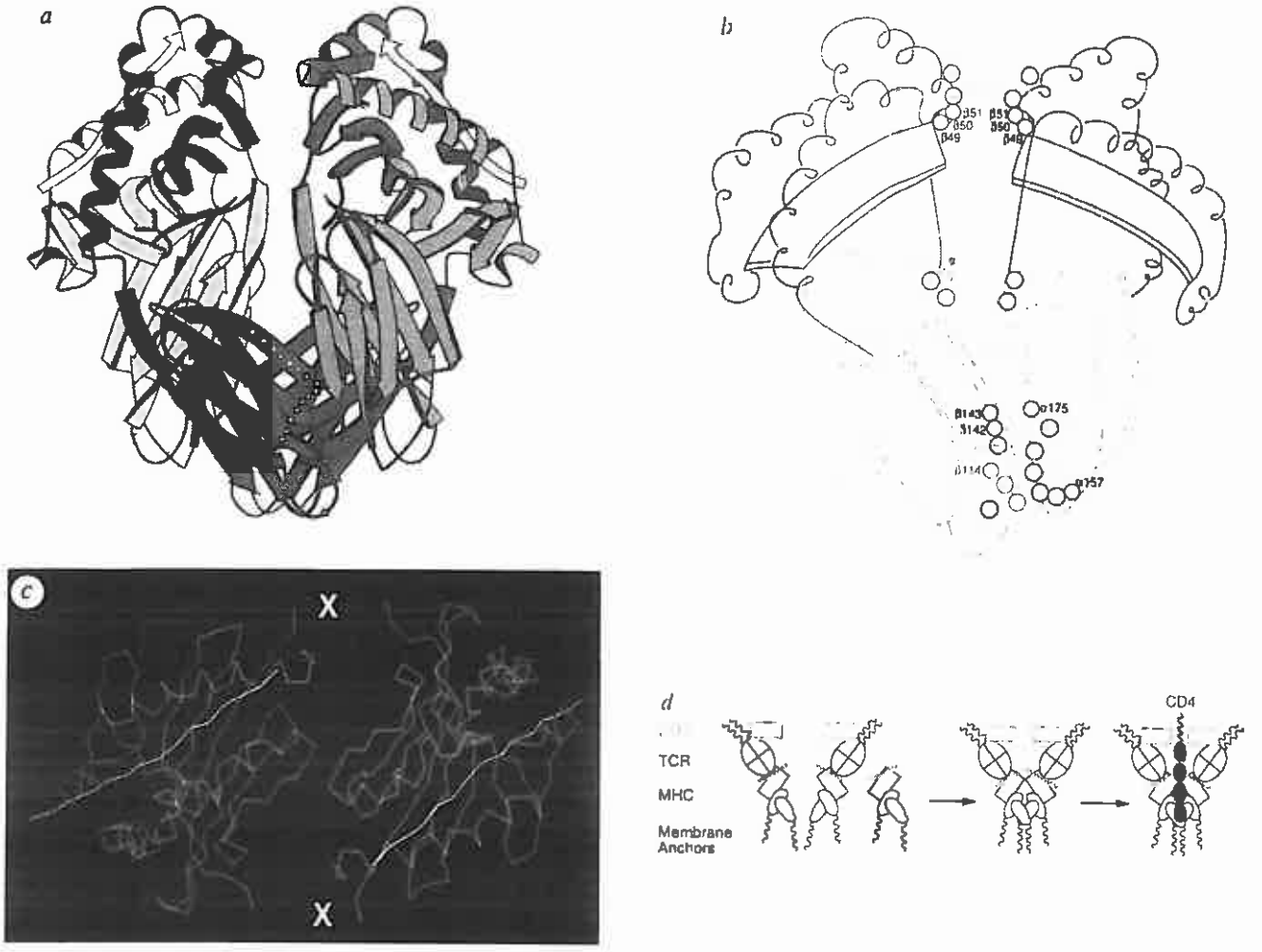


FIG. 5 DR1 crystallizes as a dimer of the $\alpha\beta$ heterodimer. *a*, One DR1 molecule is shaded in red, the other in blue, peptides are yellow. The α-chains are the lighter shades, the β-chains are darker. The width of the dimer shown is 76 Å; the height is 73 Å. β_2 residues implicated in CD4 binding by mutagenesis[40] (□), by peptide binding[41] (○). α_2 domain residues near the β_2 domain residues that have been implicated in CD4 binding are α 116, 118, 120, 122, 125, 127, 129, 164, 166, 168, 171, 172 and 173. *b*, Residues which make contact (<3.5 Å) in current, unrefined model across the DR1 dimer interfaces are indicated. Red, Asp or Glu; Blue, Arg, Lys, His. The second dimer-related interface at the bottom is not shown. A total of 1,293 Å^2 of one DR1 heterodimers solvent accessible surface is buried by the other, larger than the interface between antibody and protein antigen[68]. Surface areas in DR1 were obtained using ACCESS with a probe radius of 1.4 Å (written by M. D. Handschumacher and F. M. Richards), and may slightly change upon refinement. *c*, The dimer of antigen-binding domains as seen from the top. The closest approach of the bound portions of the two peptides (in yellow) is ~30 Å, the furthest distance being about 80 Å. 'X' marks the probable locations of the CD4 coreceptors. *d*, Hypothetical assembly pathway for an antigen-specific, cell–cell interaction complex of class II HLA, T-cell receptor, and the coreceptor CD4 (see text). Further aggregation may occur if CD4 dimerizes[69] (J. Wang and S. C. Harrison, personal communication). (An alternative model where CD4 binds before TCR has similar properties but is less consistent with CD4-independent T-cell recognition.) The box marked CD3 is intended to symbolize the CD3, ζ, and associated components of the T-cell receptor.

of mostly conserved (α 26, 32, 43, 54) non-polar residues in class II HLA (Fig. 3). Three other polymorphic clusters are found at three, five and about seven or eight residues further along the bound peptide, composed of β 13, 26, 28, 71, 74, 78; α 66, β 9, 11, 30 and β 37, 38, 58. The clustering of the murine I-E and I-A equivalents of β 71 and β 74 with β 28 has previously been suggested from mutual compensatory effects of mutations at those residues on peptide binding and T cell response[27]. The structure is also consistent with earlier observations of the effect of allelic specific residues on α–β chain pairing[28,29].

The polymorphic residue aspartic acid β 57 in DR1 appears to form a salt bridge with a conserved arginine α 76 (Fig. 3) under the exiting peptide (see below). Aspartic acid β 57 in the HLA-DQ isotype has been found to correlate with protection against insulin-dependent diabetes mellitus (IDDM) in humans, as HLA-DQ allelic products with side chains of β 57 incapable of forming a salt bridge are linked to susceptibility to IDDM[7,30]. In the non-obese diabetic mouse alterations at β 57 in I-A^k have been implicated in insulitis[31–33], but other genes also appear to be involved in generating disease[31]. In addition to a possible effect on peptide binding, polymorphism at β 57 may affect the stability or structure of the dimer of $\alpha\beta$ heterodimers because nearby residues β 52 and β 55 form part of the interface of the dimer (see below).

Structure of bound endogenous peptides

The electron density in the binding groove (Fig. 4), which represents the collection of endogenous peptides from human lymphoblastoid cells, is straight and thin as expected for a peptide in an extended conformation. This density indicates that about

© 1993 Nature Publishing Group

TABLE 1 Data collection statistics

X-ray Temp.	Protein (source)	Space group (Z*)	a, b, c (Å)	Deriv.	Resolution R_{merge} comp(red)†	Resolution R_{merge} comp(red)†	ΔF	r.m.s. f_h/E†
GX-13 4 °C	DR1(LG-2)	$C222_1$ (2)	96.7 112.6 209.2	native	100–4 Å 0.117 74% (1.5)	4.5–4.2 Å 0.222 60% (1.4)		
GX-13 4 °C	DR1(LG-2)	$C222_1$ (2)	96.6 112.6 209.2	K_2HgI_4	15–4.2 Å 0.099 73% (1.2)	4.4–4.2 Å 0.198 63% (1.1)	0.20 (15–4.2 Å)§	1.49
GX-13 −165 °C	DR1(LG-2) SEB‖	$P2_12_12_1$ (2)	95.7 115.9 150.4	native	30–3.4 Å 0.101 70% (1.44)	3.6–3.4 Å 0.24 51% (1.47)		
GX-13 −165 °C	DR1(LG-2) SEB‖	$P2_12_12_1$ (2)	95.9 117.1 150.8	K_2HgI_4	30–3.2 Å 0.077 83% (4.0)	3.4–3.2 Å 0.229 76% (2.23)	0.185 (30–3.5 Å)§	1.3
CHESS −170 °C	DR1(LG-2)	$C222_1$ (2)	96.7 114.4 216.8	native	15–2.7 Å 0.082 90% (2.4)	3.4–3.2 Å 0.168 95% (2.7)		
CHESS −170 °C	DR1(LG-2) SEB‖	$P2_12_12_1$ (2)	95.0 114.7 149.8	native	20–2.8 Å 0.057 88% (3.3)	3.0–2.8Å 0.224 89% (3.5)		
CHESS −170 °C	DR1(SF9) HA306–318	$P4_32_12$ (2)	95.4 95 4 245.7	native	20–2.7 Å 0.068 93% (3.2)	3.0–2.9 Å 0.221 94% (3.2)		

Averaging statistics

	4.2 Angstrom 4-fold DR1(LG-2)	4.2 Angstrom 4-fold DR1/SEB	3.3 Angstrom 6-fold DR1(LG-2)	3.3 Angstrom 6-fold DR1/SEB	3.3 Angstrom 6-fold DR1(SF9)
$R^{\S}$	0.23	0.23	0.31	0.30	0.30
Mean-phase Difference					
Overall	75.3°	68.7°	56.8°	67.5	51.1°
Last cycle	2.8°	3.3	3.2°	4.8	3.1°

DR1 and a number of other subtypes and isotypes of human class II histocompatibility antigens have been crystallized[8] after purification by immunoaffinity chromatography in detergent complex and removal of the transmembrane region and cytoplasmic domain by papain digestion[9]. Crystals of DR1 isolated from the human B lymphoblastoid cell line LG-2 (DR1–LG2) suitable for high-resolution study were initially grown at pH 4.0 and later at pH 5.5–6.0 from 27% PEG 8000, 0.1 M glycine, 0.01% NaN_3, 0.05 M MES. Crystals of the superantigen SEB in 1:1 complex with DR1–LG2 were grown using the same human-cell derived DR1 mixed with SEB (Sigma) further purified by gel filtration and will be described further elsewhere (T.S.J. *et al.*, manuscript in preparation). DR1 expressed as a soluble heterodimer in SF9 insect cells (DR1–SF9), purified as an empty molecule, and subsequently complexed with the influenza virus 13-mer peptide HA 306–318 crystallized from 15% PEG 8000, 0.02% NaN_3, 100 mM sodium acetate, pH 4.6[10]. Crystals of DR1–LG2 and the DR1–SEB complex were soaked overnight in 0.12 mM K_2HgI_4, 0.3 mM KI. Native and derivative X-ray data (first four sets of rows) were collected with a Xentronics area detector using an Elliot GX-13 X-ray source focused with Franks mirrors[62], then integrated and scaled using the BUDDHA[63] and CCP4 packages. Two heavy-atom positions per asymmetric unit were found to be 27.5 Å apart by analysis of difference Patterson maps of both crystals. Electron density maps were calculated at 6 Å resolution using single isomorphous replacement (SIR) phases for DR1–LG2 and SIR with anomalous scattering for DR1–SEB. A relationship between the electron density in the two crystal forms was found by a one-dimensional, real-space correlation search along the Hg to Hg vector in the two crystals using the program CYLINDER (written by P. Metcalf). The maps from the two crystals were then averaged. An envelope could be constructed around the DR1 portion of the electron density from the average map, and the remaining electron density in the DR1–SEB crystals was assigned to an SEB envelope. Thirteen cycles of iterative, real-space averaging within the DR1 envelopes and solvent flattening outside of DR1 and DR1–SEB in the two crystal forms were done using CCP4 and Joy of skewing[13]. Input for the first cycle of averaging include observed structure factors (F_{obs}) and single isomorphous replacement phases (ϕSIR) for DR1–LG2 and F_{obs} and ϕ SIR/anomalous for DR1–SEB. Input for cycles 2→4 include F_{obs} and combined phases (ϕcomb)[13]. Input for cycles 5→8 included combined structure factors (F_{comb})[64] and ϕcomb. Input for the final 5 cycles included Sim-weighted $2F_{obs}-F_{calculated}$ coefficients and calculated phases. This procedure yielded an electron density map from which the domain structure was evident for two, $\alpha\beta$ heterodimers of DR1 per asymmetric unit in each crystal. Using the 2-fold noncrystallographic rotational symmetry within each crystal's asymmetric unit, the above averaging procedure was repeated, but now real-space averaged 4-fold within the DR1 envelopes, 2-fold within the SEB envelopes, and solvent flattened outside the envelopes at 4.2 Å resolution (left side of averaging statistics). The DR1 polypeptide pathway and many large side chains were evident at this point. The Hg derivative was seen to label the only free cysteine (β 30). Higher resolution, more complete X-ray data were then collected on phosphor image plates at CHESS (and integrated using DENZO (Zbyszek Otwinowski, personal communication)) from crystals that had been soaked in cryoprotectant (row sets 5–7) and mounted in nichrome (DR1–SEB) or glass (DR1–LG2 and DR1–SF9) loops[65] and flash frozen to −165 °C in a stream of cold nitrogen. A partially refined model of DR1 (against synchrotron-cryo DR1–LG2) was fit and refined as rigid domains using X-PLOR version 2.1 (written by A. Brünger) into the 3 or 3.5 Å resolution cryo-X-ray data of the other two crystals. For each of the three crystal forms, eight 'omit models' were constructed[14], each missing about one-eighth of the DR1 model. The DR1–LG2 'omit' models were subjected to a simulated annealing refinement procedure to reduce bias[66]. 'Omit maps' were calculated from each 'omit' model. Eight cycles of sixfold, iterative, real-space averaging of the electron density within the DR1 envelopes of both monomers in the 'omit' maps of all three crystals (right side of averaging statistics) yielded the current 3.3 Å (Fig. 1) map. A model has been built using FRODO[67].

* Z, Number of DR1 heterodimers per asymmetric unit.

† Comp, % completeness = 100 × (number of independent reflections)/(maximum number theoretical), red, redundancy = (number of measurements)/(number of independent reflections). CHESS = Cornell High Energy Synchrotron Source, GX-13 = Elliot rotating anode.

‡ f_h, Heavy-atom structure factor; E the residual lack of closure.

§ Resolution range used for heavy-atom derivative statistics, ΔF and r.m.s. f_h/E.

$$R_{Merge} = \frac{\sum_{hkl}\sum_{obs}|I^{hkl}_{obs} - \langle I^{hkl}\rangle|}{\sum_{hkl}\sum_{obs} I^{hkl}_{obs}}, \qquad \Delta F = \frac{\sum_{hkl}|F_{native} - F_{derivative}|}{\sum_{hkl} F_{native}}, \qquad R = \frac{\sum_{hkl}|F_{observed} - F_{calculated\ from\ average\ map}|}{\sum_{hkl} F_{observed}}$$

‖ SEB source: SEB used in these frozen crystals was a gift from M. Sax, VA Medical Center, Pittsburgh, PA. The same crystals were also obtained from SEB purchased from Toxin Technology Inc., Sigma, or purified from an SEB producing strain of *S. aureus*.

© 1993 Nature Publishing Group

15 peptide residues contact HLA-DR, consistent with the average length of the most abundant peptides bound to DR1 from LG2 cells[4]. The electron density is not weak in the middle as observed in class I HLA-A2[34] and HLA-Aw68[35] (but not HLA-B27[17]) where peptides of different length appear to be accommodated by bulging out of the site near the middle[17,19,35]. Instead it extends out of both ends of the site, consistent with the observation that the endogenous peptides of class II HLA vary in length from 12–24 amino acids and have 'ragged' ends[1,4].

The current DR1 model is based on averaging the 3.3 Å resolution data of different crystal types and is not crystallographically refined, so no detailed model for the bound peptides has been constructed. (A more complete analysis of peptide contacts is in preparation at higher resolution for the complex between DR1 and influenza virus haemagglutinin peptide, HA 306–318 (L.J.S. *et al.*, unpublished data). But the electron density from LG2 cell-derived DR1 (Fig. 4*b*, left end) indicates the position of one prominent peptide side chain that appears bound in a hydrophobic pocket near residues α 26, 31 and β 86. This seems to be the position of the dominant 'anchoring' peptide residue predicted at position 308 of the influenza HA (306–318) peptide both by binding studies of substituted peptide sequences and by the high-affinity binding of a simplified peptide where all but two residues were made alanine[36]. The size of this peptide residue side chain seems to parallel the size of the non-polar pocket as modulated by the Gly–Val dimorphism at β 86[37–39]. These observations and the homologous role observed for tryptophan β 61 in class II and tryptophan 147 in class I (see above) suggest that peptides are bound with their N termini to the left (in Fig. 3), an orientation homologous to that in class I sites. This has been confirmed for one peptide HA306–318 complexed with HLA-DR1, which is currently being refined to 2.7 Å resolution (L.J.S. *et al.*, unpublished data). The clarity of the peptide electron density may in part be due to the predominance of a few peptides in DR1 from LG2 cells[4].

A dimer of $\alpha\beta$ heterodimers

Because a dimer (Fig. 5*a*) is observed in DR1 crystals, it may represent an intrinsic property of DR1 (see next section). There are four interfaces between DR1 molecules in the dimer (Fig. 5*b*). Two small interfaces are near the 2-fold rotational symmetry axis that relates the molecules (top two in Fig. 5*b*). β_1 49–55 and α_2 88, 111. In the most membrane-distal interface, five β_1 domain residues of one molecule contact the same residues in the β_1 domain of the second molecule, burying 305 Å^2 of solvent accessible surface on each monomer (see also Fig. 3). Four of these residues are conserved in all DR alleles. Two salt bridges from glutamic acid β 52 to arginine β 55 on the other molecule are conserved in DR, but are replaced by proline and leucine residues in some DP and DQ variants. The second interface is simply two salt bridges between glutamic acid 88 and lysine 111 of the second domain of the conserved DR1 α chain. This interface is small, burying 75 Å^2 of surface per monomer.

The bottom two interfaces are identical and are located about 12 Å from the dimer rotation axis at the membrane proximal end of the dimer (Fig. 5*b*). Each is composed of 7 residues on the β_2 domain of one molecule in contact with 7 residues on the α_2 domain of the other molecule and buries surfaces of about 900 Å^2. Most of the residues are conserved in class II HLA sequences including those participating in three salt bridges between α_2 domain residues glutamic acid 158, histidine 177, glutamic acid 179 and β_2 domain residues histidine 111, glutamic acid 162 and histidine 112.

CD4 binding site

The T-cell coreceptor molecule, CD4, is thought to bind to the β_2 domain of DR1 based on mutagenesis that indicated CD4 affinity changes when β_2 domain residues 110, 137, 140, 141 and 142 were altered[40] and based on the observation that peptides with the sequence of β_2 domain residues 134–148 (and less so, 138–152) could bind to CD4[41] (white symbols in Fig. 5*a*). If the DR1 dimer is present during T-cell recognition, the X-ray structure suggests that a region on the α_2 domain of the second DR1 molecule in the dimer (light blue domain near white symbols, Fig. 5*a*) might also contact CD4, potentially increasing the affinity of the DR1–CD4 interaction over that possible between CD4 and a single DR1 molecule. (α_2 domain positions near the β_2 CD4 site are listed in Fig. 5*a* legend.) Some of the β_2 domain positions identified by mutagenesis as important for CD4 binding are in (β 141, 142) or near (β 110 for which the affinity effect is only partial[40]) the dimer interface raising the possibility that some of the observed CD4 affiity changes may have been due to destabilization of a class II HLA dimer. Viewed from the 'top' (Fig. 5*c*), the DR1 dimer presents two peptide binding site surfaces presumably for recognition by T-cell receptors and two slots between them (X in Fig. 5*c*). The presumptive CD4-binding sites described above are located behind the Xs in Fig. 5*c*.

Conclusion

HLA class II versus class I peptide binding. The class II DR1 structure shows that the peptide-binding site is open at both ends (Fig. 4). Long peptides, at least 15 residues, are seen to be bound in an extended conformation, projecting out of both ends of the site. This contrasts with peptides bound to class I HLA molecules where mostly 9-mers bind with extended, but kinked, conformations, and the N and C termini of the peptides are sequestered in conserved, peptide-terminal binding sites[17–19,35]. In the mechanism by which MHC molecules bind many peptides tightly, the role of clusters of conserved residues in class I HLA binding of peptide termini[17,35,42] appears to be replaced in class II HLA by conserved residues (β82N, α62N, α69N, β61W, α76R) spaced so as to bind at intervals (Fig. 3) to the extended main chain of bound peptide, providing a peptide side chain independent component to the affinity.

In class I MHC molecules polymorphic residues in the binding site form pockets which accommodate peptide side chains and which form the basis for allele-specific peptide binding[17,19,34,35,43]. In class II DR1 one prominent peptide side chain at about the third position of a 15-mer identifies a critical binding pocket, and other pockets are inferred from the locations of polymorphic residues and the shape of the electron density representing multiple peptides complexed with DR1 (Figs 3 and 4). The critical binding pocket of DR1 lies near one end (left in Figs. 3 and 4) of the class II binding groove, before the location where the class I groove begins. The location of this critical pocket, the orientation of the peptide, and the extended conformation of the peptide are consistent with some of the previous studies of peptide interactions with class II proteins[16,44,45].

Possible implications of a class II HLA dimer of dimers. Two observations suggest that the dimer of DR1 $\alpha\beta$ heterodimers discovered in the three crystals studied is suitable for a physiological role. (1) In the dimer two DR1 molecules are parallel to each other so that all four C termini face one direction (down in Fig. 5*a*) and both peptide-binding sites face the opposite direction (up in Fig. 5*a*) as might be expected for glycoproteins on a membrane surface. (2) The two peptide-binding regions are oriented so that two molecules the size of the $\alpha\beta$ T-cell receptor could contact these binding sites simultaneously.

The possibility that the dimer is a crystallization artefact is not ruled out. It is also possible that a dimer stable only at high protein concentrations in solution has been stabilized by crystallization. Whether stabilization conditions might exist physiologically to take advantage of this capacity to form dimers is discussed below. Whether the dimer exists *in vivo* can be tested by site-directed mutagenesis. If the observed dimer of DR1 molecules is physiologically important, its structure suggests that mutations in the α_2 domain may affect CD4 binding (see Fig. 5*a* legend) and that mutations in the dimer interfaces (Fig. 5*b*) may affect both CD4 binding and T-cell activation.

© 1993 Nature Publishing Group

The capacity of DR1 molecules to form a parallel dimer suggests the possibility that the observed dimerization might provide part of the mechanism for initiating intracellular signalling pathways both in T cells by crosslinking two T-cell receptors[46,47] and, in antigen presenting cells, to induce the expression of costimulating molecules[48-51]. Dimerization of cell-surface receptors with the concomitant dimerization of their cytoplasmic domains is an established mechanism for signal initiation in some hormone receptor interactions, for example that of human growth hormone[52]. Indeed, T-cell receptors can be activated by crosslinking with a divalent antibody but are not activated by the cognate monovalent Fab fragment[53]. But hydrodynamic and other measurements provide no evidence that DR1 dimerizes in solution[8,10,54] (although detergent-soluble DR1 migrates anomalously on gel filtration[55]), even at the acidic conditions (pH 4.0–6.0) used in crystallization (L.J.S., unpublished data).

Low affinity in the dimer interaction may be an important feature for T-cell activation. Unless a mechanism exists for loading both peptide-binding sites of a single dimer with the same peptide, dimers preexisting on antigen-presenting cells would be expected to contain two different peptides, only one being specific for a given T cell. Weakly interacting DR1 molecules would exchange more readily, allowing T-cell receptor recognition a role in the assembly of an activation complex.

If the T-cell receptor binds to monomeric DR1 molecules and affects dimerization, many mechanisms for initiating signalling can be generated. For one example (Fig. 5*d*), assume that the $\alpha\beta$ T-cell receptor (and its associated CD3 and ζ subunits) can also form weak dimers like those of DR1, then before DR1–TCR binding neither DR1 nor TCR dimers would be stable (Fig. 5*d*, left). Once DR1 molecules with a specific peptide are bound by a TCR, each such complex would be able to form a stable dimer by using both the weak dimerization sites of DR1 and of the TCR molecules simultaneously, achieving stability by a chelating effect (Fig. 5*d*, middle)[56]. The DR1 dimer structure further suggests that the affinity for the coreceptor CD4 could be greater after dimerization due to contacts with both monomers, suggesting both that adhesion contributed by CD4 would be increased after TCR-dependent dimer assembly, and that signalling through CD4[57] could be triggered by its association with this complex (Fig. 5*d*, right). Concentrations sufficient to drive the assembly process (Fig. 5*d*) may be found in the synapse-like contacts between T cells and their target cell.

Some peptides that bind to class II MHC but engage the T-cell receptor weakly or differently do not initiate full T-cell signalling and therefore act as pure antagonists or mixed agonist/antagonists[58-60]. Such MHC–TCR complexes might be expected to assemble into the interaction complex proposed above (Fig. 5*d*, right), but could destabilize it below the threshold required to stimulate complete intracellular signalling.

A synergistic, multisubunit assembly process, where each interaction increases the affinity of subsequent interactions, would have the potential to amplify an initially small, but specific, interaction between the $\alpha\beta$ chains of TCR and HLA into a stable, antigen-specific complex that bridges both sides of the T-cell membrane and the antigen-presenting cell membrane (Fig. 5*d*, right). This mechanism substitutes direct, stable molecular contacts to initiate full intracellular signalling for previous models that invoked statistical increases in the local density of TCR–HLA complexes to lead to T-cell activation[47,61]. □

1. Rudensky, A. Y., Preston Hurlburt, P., Hong, S. C., Barlow, A. & Janeway, C. A. Jr *Nature* **353,** 622–626 (1991).
2. Hunt, D. F. *et al. Science* **256,** 1817–1820 (1992).
3. Chicz, R. M. *et al. J. exp. Med.* **178,** 27–47 (1993).
4. Chicz, R. M. *et al. Nature* **358,** 764–768 (1992).
5. Kappes, D. & Strominger, J. L. *A. Rev. Biochem.* **57,** 991–1028 (1988).
6. Tiwari, J. L. & Terasaki, P. I. *HLA and Disease Associations* (Springer, New York, 1985).
7. Todd, J. A. *et al. Science* **240,** 1003–1009 (1988).
8. Gorga, J. C., Brown, J. H., Jardetzky, T., Wiley, D. C. & Strominger, J. L. *Res. Immun.* **142,** 401–407 (1991).
9. Gorga, J. C., Horejsí, V., Johnson, D. R., Raghupathy, R. & Strominger, J. L. *J biol. Chem.* **262,** 16087–16094 (1987).
10. Stern, L. J. & Wiley, D. C. *Cell* **68,** 465–477 (1992).
11. Bjorkman, P. J. *et al. Nature* **329,** 506–512 (1987).
12. Brown, J. H. *et al. Nature* **332,** 845–850 (1988)
13. Brisogne, G. *Acta crystallogr.* **A32,**832–847 (1976).
14. Sussman, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M. & Kim, S.-H. *J. molec. Biol.* **123,** 607–630 (1978).
15. Falk, K., Rötzschke, O., Stevanovic, S., Jung, G. & Rammensee, H.-G. *Nature* **351,** 290–296 (1991).
16. Gorga, J. C. *Crit. Rev. Immun.* **11,** 305–335 (1992).
17. Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. *Cell* **70,** 1035–1048 (1992).
18. Silver, M. L., Guo, H.-C., Strominger, J. L. & Wiley, D. C. *Nature* **350,** 367–369 (1992)
19. Fremont, D. H., Matsumura, M., Stura, E. A., Peterson, P. A. & Wilson, I. A. *Science* **257,** 919–927 (1992).
20. Bjorkman, P. J. *et al. Nature* **329,** 512–518 (1987).
21. Griffith, I. J. *et al. J. exp. Med.* **167,** 541–555 (1988).
22. Peccoud, J., Dellabona, P., Allen, P., Benoist, C. & Mathis, D. *EMBO J.* **9,** 4215–4223 (1990)
23. Sette, A. *et al. J. Immun.* **143,** 3289–3294 (1989).
24. Sidney, J. *et al. J. Immun.* **149,** 2634–2640 (1992).
25. Hammer, J., Takacs, B. & Sinigaglia, F. *J. exp. Med.* **176,** 1007–1013 (1992)
26. Geluk, A. *et al. J. Immun.* **149,** 2864–2871 (1992).
27. Racioppi, L., Ronchese, F., Schwartz, R. H. & Germain, R. N. *J. Immun.* **147,** 3718–3727 (1991)
28. Braunstein, N. S. & Germain, R. N. *Proc. natn. Acad. Sci. U.S.A.* **84,** 2921–2925 (1987).
29. Braunstein, N. S., Germain, R. N., Loney, K. & Berkowitz, N. *J. Immun.* **145,** 1635–1645 (1990).
30. Horn, G. T., Bugawan, T. L., Long, C. M. & Erlich, H. A. *Proc. natn. Acad. Sci. U.S.A.* **85,** 6012–6016 (1988).
31. Miyazaki, T. *et al. Nature* **345,** 722–724 (1990).
32. Slattery, R. M. *et al. Nature* **345,** 724–726 (1990).
33. Lund, T. *et al., Nature* **345,** 727–729 (1990).
34. Saper, M. A., Bjorkman, P. J. & Wiley, D. C. *J. molec. Biol.* **219,** 277–319 (1991).
35. Guo, H.-C. *et al. Nature* **360,** 364–388 (1992).
36. Jardetzky, T. S. *et al. EMBO J.* **9,** 1797–1803 (1990).
37. Busch, R., Hill, C. M., Hayball, J. D., Lamb, J. R. & Rothbard, J. B. *J. Immun.* **147,** 1292–1298 (1991).
38. DeMotz S., Barbey, C., Corradin, G., Amoroso, A. & Lanzavecchia, A. *Eur. J. Immun.* **23,** 425–432 (1993).
39. Newton-Nash, D. K. & Eckels, D. D. *J. Immun.* **150,** 1813–1821 (1993).
40. Konig, R., Huang, L.-Y. & Germain, R. N. *Nature* **356,** 796–798 (1992).
41. Cammarota, G. *et al. Nature* **356,** 799–801 (1992).
42. Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. *Nature* **353,** 321–325 (1991).
43. Garrett, T. P. J., Saper, M. A., Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. *Nature* **342,** 692–696 (1989).
44. Krieger, J. I. *et al. J. Immun.* **146,** 2331–2340 (1991).
45. Kropshofer, H., Bohlinger, I., Max, H. & Kalbacher, H. *Biochemistry* **30,** 9177–9187 (1991).
46. Klinman, N. R. *J. exp. Med.* **136,** 241–260 (1972).
47. Janeway, C. A. Jr *et al. Cold Spring Harb. Symp. quant. Biol.* **54,** 657–666 (1989).
48. Koulova, L., Clark, E. A., Shu, G. & DuPont, B. *J. exp. Med.* **173,** 759–762 (1991).
49. Nabavi, N. *et al. Nature* **360,** 266–268 (1992).
50. Schwartz, R. H. *Cell* **71,** 1065–1068 (1992).
51. Watts, T. H., Alaverdi, N., Wade, W. F. & Linsley, P. S. *J. Immun.* **150,** 2192–2202 (1993).
52. DeVos, A. M., Ultsch, M. & Kossiakoff, A. A. *Science* **255,** 306–312 (1992).
53. Kaye, J. & Janeway, C. A. Jr *J. exp. Med.* **159,** 1397–1412 (1984).
54. Roche, P. A., Marks, M. S. & Cresswell, P. *Nature* **354,** 392–394 (1991).
55. Springer, T. A., Kaufman, J. F., Terhorst, C. & Strominger, J. L. *Nature* **268,** 213–218 (1977).
56. Jencks, W. P. *Proc. natn. Acad. Sci. U.S.A.* **78,** 4046–4050 (1981).
57. Glaichenhaus, N., Shastri, N., Littman, D. R. & Turner, J. M. *Cell* **64,** 511–520 (1991).
58. DeMagistris, M. T. *et al. Cell* **68,** 625–634 (1992).
59. Alexander, J. *et al. J. Immun.* **150,** 1-7 (1993).
60. Racioppi, L., Ronchese, F., Matis, L. A. & Germain, R. *J. exp. Med.* **177,** 1047–1060 (1993).
61. Matsui, K., *et al., Science* **254,** 1788–1791 (1991).
62. Hamson, S. C. *J. Appl. Crystallogr.* **1,** 84–89 (1968)
63. Blum, M. Metcalf, P., Harrison, S. C. & Wiley, D. C. *J. appl. Crystallogr.* **20,** 235–242 (1987)
64. Stuart, D. & Artymiuk, P. *Acta crystallogr.* **A40,** 713–716 (1984).
65. Teng, T. Y. *J. Appl. Crystallogr.* **23,** 387-391 (1990).
66. Hodel, A., Kim, S.-H. & Brünger, A. T. *Acta crystallogr.* **A48,** 851–858 (1992).
67. Jones, T. A. *J. Appl. Crystallogr.* **11,** 268–272 (1978).
68. Colman, P. M. *Adv. Immun.* **43,** 99–132 (1988).
69. Kwong, P. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **87,** 6423–6427 (1990).

ACKNOWLEDGEMENTS. We thank P. Klimovitsky, A. Haykov and M. Frayser for technical assistance, K. Svenson, assisted by M. Pietras, for the large-scale production of tissue culture cells, and other members of our group and the CHESS synchrotron staff for assistance with X-ray data collection. We acknowledge conversations with R. Germain who reached similar conclusions about how a dimer might not preexist, but only be stabilized by TCR binding. J.H.B. is supported by the Howard Hughes Medical Institute, T.S.J. by the Cancer Research Institute; J.C.G. by the Juvenile Diabetes Foundation International, L.J.S. by a Damon Runyon–Walter Winchell Cancer Fellowship and the Howard Hughes Medical Institute; R.G.U. by the Irvington Institute; J.L.S. by the NIH. D.C.W. is an Investigator with the Howard Hughes Medical Institute. Preliminary, unrefined coordinates of DR1 are available from the authors.

75

© 1993 Nature Publishing Group

Genes of MHC class I and II molecules

In the case of antibody molecules, many different amino acid sequences are produced by the V-J joining of light chain genes and V-D-J joining of the heavy chain genes as discussed before. In human, there are two light chains loci: kappa and lambda. The kappa genes, both variable and constant, are located on chromosome 2 at 2p11, and lambda genes on chromosome 22 at 22q11. The heavy chain variable region genes as well as several constant region genes are all located on chromosome 14 at 14q32. Every human can thus produce a large numer of different antibodies, just like every other human.

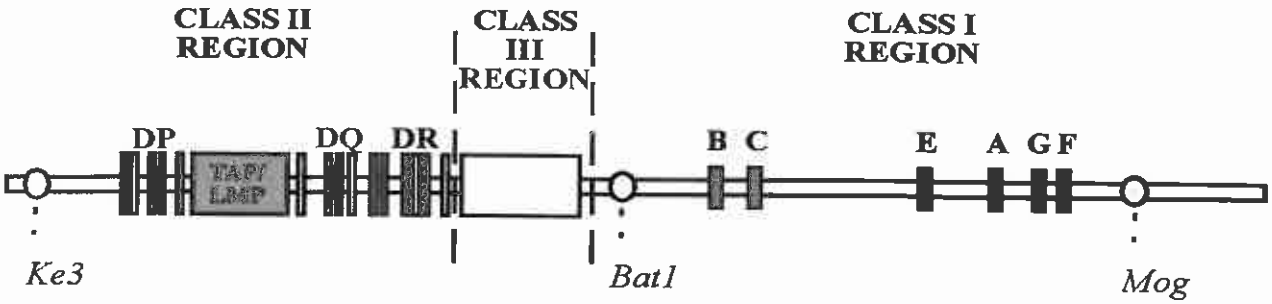
On the other hand, every human produces a distinct set of MHC class I and II molecules, except for identical twins. In other words, there are numerous alleles. Their genes are located on chromosome 6 at 6p21.3 as shown on the next two pages, one more detailed than the other. In between the cluster of MHC calss II genes and that of MHC class I genes locate the MHC class III genes which code for some of the molecules of the complement system, not directly related to tissue transplantation rejections.

On the following pages, G. Mayer and J. Nyland summarized the number of alleles in the MHC class I and II system:

Class I loci	Number of alleles
HLA-A	218
HLA-B	439
HLA-C	96
HLA-E, -F, -G	few
Class II loci	Number of alleles
HLA-DPA	12
HLA-DPB	88
HLA-DQA	17
HLA-DQB	42
HLA-DRA	2
HLA-DRB1	269
HLA-DRB3	30
HLA-DRB4	7
HLA-DRB5	12
HLA-DM, -DO	few

Genetics of MHC

The Major Histocompatibility Complex (**MHC**) is a gene rich region encoding highly polymorphic histocompatibility genes. These were first discovered and studied in graft rejection experiments. Work in this area attempts to match organ donors and recipients for the antigens important for successful transplantations. The human MHC, also known as **HLA**, consists of approximately 4,000 kilobases of DNA covering about 2.5 map units located on the short arm of **chromosome 6**. Many proteins encoded in this region are involved is diverse aspects of the immunological response mechanism.



The MHC class I locus contains smaller loci for the classical genes, named A, B and C and non-classical genes E, F and G. All code for class I molecules, but each is distinct in its structure and binding capacity. The class I gene codes only for the alpha protein of the Class I molecule. The beta-2 microglobulin gene is very constant and is located elsewhere in the genome.

A similar situation exists for MHC II, where the locus is split into three smaller loci named DP, DQ and DR. Here each gene codes for a variant of both the alpha and beta protein. Since it is possible for an alpha unit from one gene to associate with a beta of another, there are a maximum of twelve different MHC II molecules. The prevalence of different HLA types vary widely in different populations. It is within this region that the genes for LMP proteins and TAP transporters have been found.

The region between the D and B loci codes for a number of MHC class III proteins, including the complement components C2, C4 and Factor B.
(The diagram above indicates the positions of the markers Ke3, Bat1 and Mog.)

Suggested further reading

Roberto S. Accolla
MHC and the immune response.

Omim Entry **[142880]**
MAJOR HISTOCOMPATIBILITY COMPLEX

77

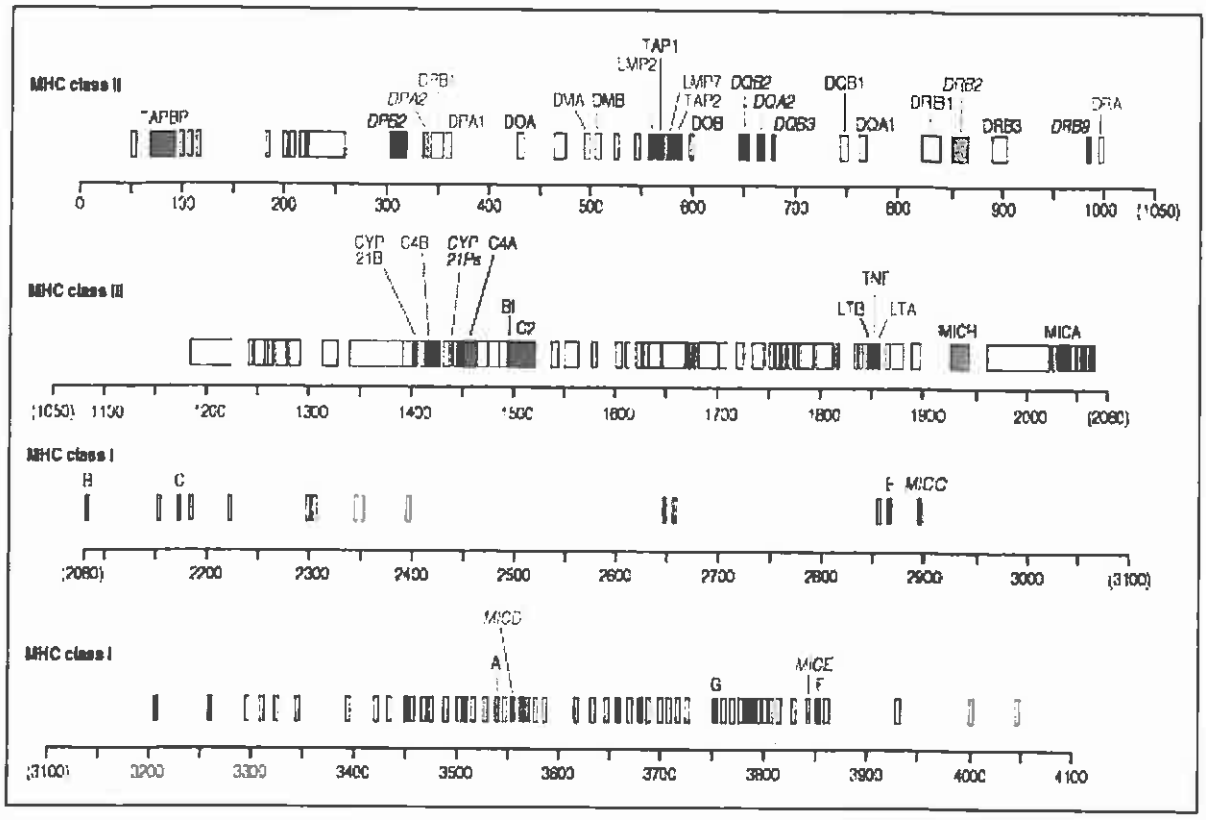


Figure 5.11 Detailed map of the human MHC

The organization of the class I, class II, and class III regions of the human MHC are shown, with approximate genetic distances given in thousands of base pairs (kb). Most of the genes in the class I and class II regions are mentioned in the text. The additional genes indicated in the class I region (for example, E, F, and G) are class I-like genes, encoding class IB molecules; the additional class II genes are pseudogenes. The genes shown in the class III region encode the complement proteins C4 (two genes, shown as C4A and C4B), C2 and factor B (shown as Bf) as well as genes that encode the cytokines tumor necrosis factor-α (TNF) and lymphotoxin (LTA, LTB). Closely linked to the C4 genes is the gene encoding 21-hydroxylase (shown as CYP 21B), an enzyme involved in steroid synthesis. Genes shown in gray and named in italic are pseudogenes. The genes are colour coded, with the MHC class I genes being shown in red, except for the MIC genes, which are shown in blue; these are distinct from the other class I-like genes and are under different transcriptional control. The MHC class II genes are shown in yellow. Genes in the MHC region which have immune functions but are not related to the MHC class I and class II genes are shown in purple.

From: The major histocompatibility complex and its functions

Immunobiology: The Immune System in Health and Disease. 5th edition.
Janeway CA Jr, Travers P, Walport M, et al.
New York: Garland Science; 2001.

Copyright © 2001, Garland Science.

NCBI Bookshelf. A service of the National Library of Medicine, National Institutes of Health.

Let us know what you think
SEARCH
Logo image © Jeffrey Nelson, Rush University, Chicago, Illinois and

Last major update July 2010

READING
7th edition chapters 5

IMMUNOLOGY - CHAPTER TEN

MAJOR HISTOCOMPATIBILITY COMPLEX (MHC) AND T-CELL RECEPTORS - ROLE IN IMMUNE RESPONSES

Dr Gene Mayer and Dr Jennifer Nyland

TEACHING OBJECTIVES
To give an overview of the role of the major histocompatibility complex in immune responses.
To describe the structure and function of class I and class II MHC molecules.
To discuss the nature of polymorphisims in class I and class II MHC molecules.
To describe the structure of the T cell receptor for antigen
To discuss the genetic basis for the generation of diversity in the TCR.
To discuss the role of the CD3 complex and co-stimulatory molecules.
To describe the nature of the immunological synapse.
To discuss the requirements for T cell activation.

HISTORICAL OVERVIEW

Cell-cell interactions of the adaptive immune response are critically important in protection from pathogens. These interactions are orchestrated by the immunological synapse whose primary components are the T cell antigen receptor (TCR) and the Major histocompatibility complex (MHC) molecule. The major function of the TCR is to recognize antigen in the correct context of MHC and to transmit an excitatory signal to the interior of the cell. Since binding of peptide within the MHC is not covalent, there are many factors while help stabilize the immunological synapse.

Gene products encoded in the MHC were first identified as being important in rejection of transplanted tissues. Furthermore, genes in the MHC were found to be highly polymorphic (i.e. in the population there were many different allelic forms of the genes). Studies with inbred strains of mice showed that genes in the MHC were also involved in controlling both humoral and cell-mediated immune responses. For example, some strains of mice could respond to a particular antigen but other strains could not and these strains differed only in one or more of the genes in the MHC. Subsequent studies showed that there were two kinds of molecules encoded by the MHC – Class I molecules and class II molecules which are recognized by different classes of T cells. Class I molecules were found on all nucleated cells (not red blood cells) whereas class II molecules were found only on antigen presenting cells, (APCs) which included dendritic cells, macrophages, B cells and a few other types (Figure 1).

It was not until the discovery of how the TCR recognizes antigen that the role of MHC genes in immune responses was understood. The TCR was shown to recognize antigenic peptides in association with MHC molecules. T cells recognize portions of protein antigens that are bound non-covalently to MHC gene products. Cytotoxic T cells (Tc) recognize peptides bound to class I MHC molecules and helper T cells (Th) recognize peptides bound to class II MHC molecules. The three dimensional structures of MHC molecules and the TCR have been determined by X-ray crystallography so that a clear picture of how the TCR, MHC gene products and antigen interact has emerged.

Figure 1
Distribution of class I and class II MHC molecules on human cells

STRUCTURE OF CLASS I MHC MOLECULES

The molecule

Class I MHC molecules are composed of two polypeptide chains, a long α chain and a short β chain called β2-microglobulin (figure 2). The α chain has four regions.

- A cytoplasmic region, containing sites for phosphoylation and binding to cytoskeletal elements.

79

- A transmembrane region containing hydrophic amino acids by which the molecule is anchored in the cell membrane.
- A highly conserved α3 immunoglubilin-like domain to which CD8 binds.
- A highly polymorphic peptide binding region formed from the α1 and α2 domains. The β2-microglobulin associates with the α chain and helps maintain the proper conformation of the molecule.

Figure 2
The MHC class I molecule has three globular domains alpha 1 (yellow), alpha 2 (green) and alpha 3 (blue). The alpha 3 domain is closely associated with the non-MHC -encoded beta 2 microglobulin (pink). The latter is stabilized by a disulfide bridge (red) and is similar to an immunoglobulin domain in three-dimensional structure. The alloantigenic sites which carry determinants specific to each individual are found in the alpha 1 and 2 domains. The latter also has a carbohydrate chain (blue, CHO) There is a phosphate in the cytoplasmic domain. Papain cleaves near the outer surface of the plasma membrane

The antigen-binding groove

An analysis of which part of the class I MHC molecules is most variable demonstrates that variability is most pronounced in the α1 and α2 domains, which comprise the peptide binding region (Figure 3). The structure of the peptide binding groove, revealed by X-ray crystallography, shows that the groove is composed of two α helices forming a wall on each side and eight β-pleated sheets forming a floor. The peptide is bound in the groove and the residues that line the groove make contact with the peptide (Figure 4). These are the residues that are the most polymorphic. The groove will accommodate peptides of approximately 8-10 amino acids long. Whether a particular peptide will bind to the groove will depend on the amino acids that line the groove. Because class I molecules are polymorphic, different class I molecules will bind different peptides. Each class I molecule will bind only certain peptides and will have a set of criteria that a peptide must have in order to bind to the groove. For example, Figure 5 shows that one class I molecule will bind peptides that have a leucine (L) as the carboxy-terminal amino acid and either tyrosine (Y) or phenylalanine (F) as the 4th amino acid from the carboxy-terminal end. As long as these two conditions are met a peptide will bind, regardless of what the other amino acids are. Similarly a different class I molecule will bind any peptide that has a tyrosine (Y) as the second amino acid from the amino-terminal end and either a valine (V), isoleucine (I) or leucine (L) at the carboxy-terminal end (Figure 5). Thus, for every class I molecule, there are certain amino acids that must be a particular location in the peptide before it will bind to the MHC molecule. These sites in the peptide are referred to as the "anchor sites". The ends of the peptide are buried within the closed ends of the class I binding groove while the center bulges out for presentation to the TCR.

Within the MHC there are 6 genes that encode class I molecules HLA-A, HLA –B, HLA-C, HLA-E, HLA-F and HLA-G. Among these HLA-A, HLA –B, and HLA-C are the most important and are most polymorphic. Table 1 shows the degree of polymorphism at each of these loci.

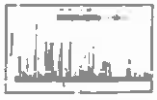


Figure 3
Most variability in amino acids at different positions along the alpha chain of class I MHC molecules occurs in the alpha 1 and alpha 2 regions. The greatest polymorphism is found for amino acids that line the wall and floor of the groove that binds the peptides

Figure 4
a. Peptide binding groove of class I MHC molecules.
b. Groove with highlighted highly variable residues. The variable residues are clustered around the peptide-binding pocket

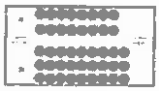


Figure 5
Anchor sites in peptides that bind to class I MHC molecules (adapted from Janeway et al. Immunobiology 6th Edition

CHIME

Chime presentation showing the regions of variability of MHC I molecules and the interaction of the alpha chain with other subunits of the MHC I complex and the bound peptide (requires Chime plug-in

Table 1. Polymorphism of class I MHC genes

Locus	Number of alleles (allotypes)
HLA-A	218
HLA-B	439
HLA-C	96
HLA-E, HLA-F and HLA-G	Relatively few alleles

Get Chime)

STRUCTURE OF CLASS II MHC MOLECULES

Figure 6
MHC class II molecules comprise two non-identical peptides (alpha and beta) which are non-covalently associated and traverse the plasma membrane with the N terminus to the outside of the cell. The domains closest to the membrane in each chain are structurally related to immunoglobulins. With the exception of the alpha 1 domain, all domains are stabilized by disulfide bridges (red). Both the alpha and beta chains are glycosylated. The beta chain is shorter than the alpha chain (beta mol. wt = 28,000) and contains the alloantigenic sites. There is some polymorphism in the alpha chain of some MHC II molecules

The molecule

Class II MHC molecules are composed of two polypeptide chains an α and a β chain of approximately equal length (Figure 6). Both chains have four regions:

- A cytoplasmic region containing sites for phosphoylation and binding to cytoskeletal elements
- A transmembrane region containing hydrophic amino acids by which the molecule is anchored in the cell membrane
- A highly conserved α2 domain and a highly conserved β2 domain to which CD4 binds
- A highly polymorphic peptide binding region formed from the α1 and β1 domains

The antigen-binding groove

As with Class I MHC molecules, an analysis of which part of the class II MHC molecule is most variable demonstrates that variability is most pronounced in the α1 and β1 domains, which comprise the peptide binding region (Figure 7). The structure of the peptide binding groove, revealed by X-ray crystallography, shows that, like class I MHC molecules, the groove is composed of two α helices forming a wall on each side and eight β-pleated sheets forming a floor. Both the α1 and β1 chain contribute to the peptide binding groove. The peptide is bound in the groove and the residues that line the groove make contact with the peptide. These are the residues that are the most polymorphic. The groove of Class II molecules is open at one end so that the groove can accommodate longer peptides of approximately 13-25 amino acids long with some of the amino acids located outside of the groove. Whether a particular peptide will bind to the groove will depend on the amino acids that line the groove. Because class II molecules are polymorphic, different class II molecules will bind different peptides. Like class I molecules, each class II molecule will bind only certain peptides and will have a set of criteria that a peptide must have in order to bind to the groove (i.e. "anchor sites").

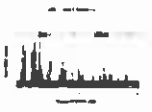


Figure 7
The greatest polymorphism for the beta chain of class II MHC molecules is found for those amino acids in the beta 1 region that line the wall and floor of the groove that binds the peptide

Within the MHC there are 5 loci that encode class II molecules, each of which contains a gene for an α chain and at least one gene for a β chain. The loci are designated as HLA-DP, HLA –DQ, HLA-DR, HLA-DM, and HLA-DO. Among these, HLA-DP, HLA –DQ, and HLA-DR are the most important and are most polymorphic. Table 2 shows the degree of polymorphism at each of these loci.

IMPORTANT ASPECTS OF MHC

- Although there is a high degree of polymorphism for a species, an individual has maximum of six different class I MHC products and only slightly more class II MHC products (considering only the major loci).
- Each MHC molecule has only one binding site. The different peptides a given MHC molecule can bind all bind to the same site, but only one at a time.
- Because each MHC molecule can bind many different peptides, binding is termed degenerate.
- MHC polymorphism is determined only in the germline. There are no recombinational mechanisms for generating diversity.
- MHC molecules are membrane-bound; recognition by T cells requires cell-cell contact.
- Alleles for MHC genes are co-dominant. Each MHC gene product is expressed on the cell surface of an individual nucleated cell.
- A peptide must associate with a given MHC of that individual, otherwise no immune response can occur. That is one level of control.
- Mature T cells must have a T cell receptor that recognizes the peptide associated with MHC. This is the second level of control.
- Cytokines (especially interferon-γ) increase level of expression of MHC.
- Peptides from the cytosol associate with class I MHC and are recognized by Tc cells. Peptides from within vesicles associate with class II MHC and are recognized by Th cells.
- Polymorphism in MHC is important for survival of the species.

81

Table 2. Polymorphism of class II MHC genes

Locus	Number of alleles (allotypes)
HLA-DPA	12
HLA-DPB	88
HLA-DQA	17
HLA-DQB	42
HLA-DRA	2
HLA-DRB1	269
HLA-DRB3	30
HLA-DRB4	7
HLA-DRB5	12
HLA-DM and HLA-DO	Relatively few alleles

HOW DO PEPTIDES GET INTO THE MHC GROOVE?

Peptides from the cytosol associate with class I MHC and are recognized by CTL cells. The peptides enter the endoplasmic reticulum and bind in the MHC class I groove. This complex is then exported to the cell surface through the Golgi. MHC class II molecules are formed with an invariant (Ii) chain as a place holder while in the ER and Golgi. The Ii chain is cleaved and removed once the complex is in a vesicle. Peptides from within the vesicle associate with class II MHC and are then exported to the cell surface where they are recognized by Th cells.

THE ROLE OF TCR IN THE IMMUNE RESPONSE

The TCR is a surface molecule found on T cells that recognizes antigen presented in the correct MHC context. The TCR is similar to immunoglobulin and is part of the immunoglobulin superfamily. There are two types of TCRs, the predominant αβ which is commonly found in lymphoid tissues, and the γδ which is found at mucosal surfaces.

Figure 8
The T cell receptor heterodimer comprises two transmembrane glycoproteins, the alpha and beta chains. There are two domains in the external part of each chain and these resemble immunoglobulin variable and constant regions. There are sugar chains on each domain. There is a short sequence similar to the immunoglobulin hinge region that connects the immunoglobulin-like domains to the transmembrane sequence. This contains cysteines that form a disulfide bridge. The hydrophobic transmembrane helical structures are unusual in that they contain positively charged amino acids (basic amino acids). The alpha chain has two positively charged residues while the beta chain has one.

STRUCTURE OF THE T CELL RECEPTOR (TCR)

The TCR is a heterodimer composed of one α and one β chain of approximately equal length (Figure 8). Each chain has a short cytoplasmic tail but it is to small to be able to transduce an activation signal to the cell. Both chains have a transmembrane region comprised of hydrophobic amino acids by which the molecule is anchored in the cell membrane. Both chains have a constant region and a variable region similar to the immunoglobulin chains. The variable region of both chains contains hypervariable regions that determine the specificity for antigen. Each T cell bears a TCR of only one specificity (*i.e.* there is allelic exclusion).

THE GENETIC BASIS FOR RECEPTOR GENERATION

The genetic basis for the generation of the vast array of antigen receptors on B cells has been discussed previously (see lecture on Ig genetics). The generation of a vast array of TCRs is accomplished by similar mechanism. The germline genes for the TCR β genes are composed of V, D and J gene segments that rearrange during T cell development to produce many different TCR β chains (Figure 9). The germline genes for the TCR α genes are composed of V and J gene segments which rearrange to produce α chains. The specificity of the TCR is determined by the combination of α and β chains.

There is a small population of T cells that express TCRs that have γ and δ chains instead of α and β chains. These gamma/delta T cells predominate in the mucosal epithelium and have a repertoire biased toward certain bacterial and viral antigens. The genes for the δ chains have V, D and J gene segments whereas the genes for the γ chains have only V and J gene segments but the repertoire is considerably smaller that than that of the alpha/beta T cells. The gamma/delta T cells recognize antigen in an MHC-independent manner unlike the alpha/beta T cells.

IMPORTANT ASPECTS OF THE TCR

- Each T cell bears a TCR of only one specificity (i.e. there is allelic exclusion).

Structure of A6-T cell receptor bound to MHC class I molecule complexed with an altered Htlv-1 Tax Peptide Y8a. The HIV peptide is shown in gray. MHC class I molecule is in dark blue, the associated beta 2 microglobulin in light blue. T cell receptor is in green and yellow. Y H Ding, B M Baker, D N Garboczi, W E Biddison & D C Wiley MMDB Id 11766 PDB Id 1QSF Image prepared using RasMol

Figure 9
Rearrangement of the TCR beta chain genes

- The αβ TCR recognizes antigen only in the context of cell-cell interaction and in the correct MHC.
- The γδ TCR recognizes antigen in an MHC-independent manner in response to certain viral and bacterial antigen.

TABLE 3
COMPARISON OF THE MAJOR PROPERTIES OF IMMUNOGLOBULIN (Ig) AND T-CELL RECEPTOR (TCR) GENES AND PROTEINS

Properties	Ig	TCR
GENES		
Many VDJs, Few C's	Yes	Yes
VDJ Rearrangement	Yes	Yes
V pairs form antigen-recognition site	Yes	Yes
Somatic hypermutation	Yes	No
PROTEINS		
Transmembrane forms	Yes	Yes
Secreted forms	Yes	No
Isotypes with distinct functions	Yes	No
Valency	2	1

Adapted from Janeway and Travers, Immunobiology

Structure of a crystal structure of a complex of a human T cell receptor, influenza Ha Antigen Peptide and an MHC Class II Molecule. The alpha and beta chains of the MHC II molecules are in dark and light blue. The T cell receptor is in yellow and green. The influenza peptide is in gray. Hennecke, J., Carfi, A., Wiley, D. C. MMDB Id 14648 PDB Id 1FYT Image prepared using RasMol

CHIME

Click on the image above to view rotatable structure and identify protein chains of MHC I and TCR interacting with HTLV tax peptide (requires Chime plug-in Get Chime ...)

CHIME

Click on the image above to view a rotatable structure and identify protein chains of MHC II and TCR interacting with an influenza HA peptide (requires Chime plug-in Get Chime ...)

Figure 10
The receptor for antigens on the T cell surface comprises eight proteins
(a) Two disulfide-bonded chains

TCR AND CD3 COMPLEX

The TCR is closely associated with a group of 5 proteins collectively called the CD3 complex (Figure 10). The CD3 complex is composed of one γ, one δ, two ε and 2 ξ chains. All of the proteins of the CD3 complex are invariant and they do not contribute to the specificity in any way. The CD 3 complex is necessary for cell surface expression of the TCR during T cell development. In addition, the CD3 complex transduces activation signals to the cell following

83

of the T cell receptor which form a heterodimer. These recognize peptides associated with MHC molecules.
(b) Four chains, collectively termed CD3, that associate with the T cell receptor dimer and participate in its transport to the surface of the cell The CD3 complex together with the zeta chains, which form a homodimer, transduce the signal after antigen has bound

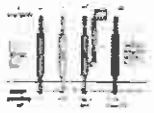
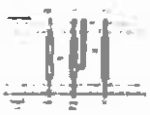


Figure 11
A. Molecules involved in the interaction between T cells and antigen-presenting cells. Some cytokines produced by each cell type are shown



B. Ligands involved in the interaction of cytotoxic T cells and their target cells

antigen interaction with the TCR.

THE "IMMUNOLOGICAL SYNAPSE"

The interaction between the TCR and MHC molecules are not very strong. Accessory molecules are necessary to help stabilize the interaction (Figure 11a,b). These include:

- CD4 binding to Class II MCH, which ensures that Th cells only interact with APCs
- CD8 binding to class I MHC, which ensures that Tc cells can interact with target cells
- CD2binding to LFA-3
- LFA-1 binding to ICAM-1

The accessory molecules are invariant and do not contribute to the specificity of the interaction, which is solely determined by the TCR. The expression of accessory molecules can be increased in response to cytokine, which is one way that cytokines can modulate immune responses.

In addition to accessory molecules which help stabilize the interaction between the TCR and antigen in association with MHC molecules, other molecules are also needed for T cell activation. Two signals are required for T cell activation – one is the engagement of the TCR with Ag/MHC and the other signal comes from the engagement of co-stimulatory molecules with their ligands. One of the most important (but not the only) co-stimulatory molecule is CD28 on T cells which must interact with B7-1 (CD80) or B7-2 (CD81) on APCs . Like accessory molecules the co-stimulatory molecules are invariant and do not contribute to the specificity of the interaction. The multiple interactions of TCR with Ag/MHC and the accessory and co-stimulatory molecules with their ligands have been termed the "immunological synapse."

Figure 12a
Activation of T cells only occurs when both TCR and co-stimulatory molecules are engaged with their respective ligands

Figure 12b
Down regulation occurs if CTLA-4 interacts with B7: CTLA-4 sends an inhibitory signal

Not only is co-stimulation necessary for T cell activation, a lack of co-stimulation may result in anergy (i.e., inability to respond to antigen) or down-regulation of the response. Figure 12 shows the possible outcomes of a T cell receiving one or both of the signals necessary for activation. Engagement of the TCR with Ag/MHC but no co-simulation results in anergy. Engagement of only the co-stimulatory molecule has no effect. Engagement of TCR with Ag/MHC and co-stimulatory molecules with their ligand results in activation. Engagement of the TCR with Ag/MHC and engagement of B7 ligand with CTLA-4, molecules similar to CD28, results in down-regulation of the response. CTLA-4/B7 interaction sends an inhibitory signal to the T cell rather than an activating signal. This is one of the ways that immune responses are regulated. CTLA-4 is expressed on T cells later in an immune response and this helps to turn off the response.

KEY STEPS IN T CELL ACTIVATION

- APC must process and present peptides to T cells
- T cells must receive a co-stimulatory signal - usually from CD28/B7
- Accessory adhesion molecules must help to stabilize the binding of T cells and APC. (CD4/class II MHC, CD8/classs I MHC, LFA-1/ICAM-1 and CD2/LFA-3)
- Signals from cell surface must be transmitted to the nucleus via second messengers
- Cytokines, including IL-2, must help drive cell division

Figure 12c
Engagement of TCR and antigen/MHC in the absence of co-stimulation may lead to anergy

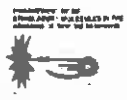


Figure 12d

TABLE 4
IMPORTANT ACCESSORY MOLECULES

T cell molecule	Ligand on second cell

Engagement of co-stimulatory molcules in the absence of TCR engagement results in no response	CD4 on helper T cells	class II MHC molecules
	CD8 on cytotoxic T cells	class I MHC molecules
	LFA-2 (CD2)	LFA-3
	LFA-1	ICAM-1, ICAM-2

LFA = Leukocyte Function-associated Antigen

ICAM = Intercellular Adhesion Molecule

Return to the Immunology Section of Microbiology and Immunology On-line

This page copyright 2010. The Board of Trustees of the University of South Carolina
This page last changed on Monday, July 12, 2010
Page maintained by

85

T cell receptor for antigen (TCR)

The complex formed by MHC class I molecule and the nine amino acid residue processed peptide or by MHC class II molecule and the roughly 15 amino acid residue processed peptide will interact with molecules on the surface of T lymphocytes known as T cell receptor for antigens (TCR's). These molecules, on one hand, resemble MHC class II molecules with two nearly equal size membrane bound chains designated as α / β. In some cases, the two chains are designated as γ / δ. These two pairs have different properties to be discussed in detail later. On the other hand, TCR's resemble the Fab fragment of antibodies. The variable regions of α and γ are assembled from V genes and J minigenes, while those of β and δ from V genes, D minigenes and J minigenes.

The search for the TCR genes took place in two separate laboratories of T. K. Mak and M. M. Davis, both using the now famous subtraction method by comparing the differences between T and B lymphocytes. Three papers were published in the 1984 issue of Nature:

Yanagi Y, Yoshikai Y, Leggett K, Clark SP, Aleksander I, Mak TW (1984) A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. Nature, 308, 145-149.

Hedrick SM, Cohen DI, Nielsen EA, Davis MM (1984) Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature, 308, 149-154.

Hedrick SM, Cohen DI, Nielsen EA, Davis MM (1984) Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. Nature, 308, 154-158.

By this method, the TCR β chain genes were isolated initially. Since morphologically T lymphocytes and B lymphocytes are very similar, most of their mRNA molecules and thus cDNA should also be similar. The major difference should thus reside in the different genes for TCR's and for immunoglobulins. Subsequently, the TCR α chain genes were also isolated. When these isolated genes were translated to amino acid sequences, they have the following domains:

TCR α chain:	V_α	C_α	α-membrane	α-cytoplasmic
TCR β chain:	V_β	C_β	β-membrane	β-cytoplasmic

Their cytoplasmic segments are relatively short, especially that of the α chain.

Similar to antibodies, the variable and constant regions have the immunoglobulin fold structure. Their variable regions also have three CDR's from both the V_α and the V_β domains, forming the TCR binding site to interact with the MHC-processed peptide complex. TCR's do not seem to differentiate MHC class I and MHC class II molecules.

A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains

Yusuke Yanagi, Yasunobu Yoshikai, Kathleen Leggett, Stephen P. Clark, Ingrid Aleksander & Tak W. Mak

The Ontario Cancer Institute, and Department of Medical Biophysics, University of Toronto, 500 Sherbourne Street, Toronto, Ontario, Canada M4X 1K9

We have cloned and sequenced a human mRNA specific for mammalian T-lymphoid cells. The message was found to be expressed in human and murine T lymphoblasts, thymocytes and phytohaemagglutinin-stimulated T lymphocytes. The protein deduced from the cDNA sequence has a molecular weight of 34,938 and shows extensive similarity to the entire length of the variable, joining and constant regions of mammalian immunoglobulin light chains. In addition, the relative positions of the cysteine residues are similar to those of the light chains of murine and human immunoglobulin molecules. These properties suggest that the cDNA clone may correspond to a message that specifies part of the human T-cell receptor.

KNOWLEDGE of the structure of immunoglobulin molecules[1,2] and the genes[3-6] encoding them has furthered our understanding of the function of B cells and the generation of diversity of immunoglobulin molecules. The findings have led to the conclusion that immunoglobulin molecules are generated by a series of gene rearrangements and RNA splicing events[3-6] that result in polypeptide chains consisting of variable and constant regions[1-6]. These regions can be subdivided into domains held together by interchain and intrachain disulphide bridges situated at the same relative positions[1-7].

In contrast, very little is known about the generation of T-cell specificity and diversity. Some investigators believe that there is a close relationship between immunoglobulin and the protein that is responsible for specificity of T cells (the T-cell receptor)[8,9], while others speculate that entirely different mechanisms or sets of genes are used by T cells[10]. During the past year, several laboratories have produced monoclonal antibodies that recognize antigens specific for individual clones of human and murine T cells[11-13]. These idiotypic proteins may be part of the human or murine T-cell receptor. Although the identification of these tentative T-cell receptor proteins has aided in the definition of their sizes and complexities[14-16], their structures, relationships to immunoglobulins, and biological functions are not clear. Progress in this area has been impeded by the lack of highly purified preparations of these proteins and the absence of molecular cones encoding T cell-specific surface antigens or receptors.

Cloning of T cell-specific mRNA

In an attempt to clone mRNAs specific for T cells, we used the human leukaemic T-cell line, MOLT-3 (ref. 17), which, like other human T-cell lines, is known to express T cell-specific antigens[18,19]. mRNA was isolated from MOLT-3 cells by oligo(dT)-cellulose column chromatography. Double-stranded cDNAs were generated using the approach described by Land *et al.*[20] and inserted into the *Bgl*II site of the vector pFP502EB5 (ref. 21) after selection of 0.5–2-kilobase (kb) fragments by preparative agarose gel electrophoresis. The cDNA was inserted by transfection into *Escherichia coli* strain HB101, and 10,000 independent clones were obtained. A survey of 25 of these chosen at random indicated that the cDNA inserts were of the expected size.

Isolation of T cell-specific clones

To screen for cDNA clones of mRNA expressed either exclusively or preferentially in T cells, the clones were grouped on the basis of their relative levels of expression in MOLT-3 cells and in the human B-cell line, HSC-58 (see Fig. 1 legend). Four of these clones, YT30 (insert size 1.3 kb), YT35 (1.3 kb), YT53 (1.3 kb) and YT76 (0.6 kb), had similar restriction enzyme maps and were all found to hybridize to mRNA from MOLT-3 cells but not to that from HSC-58 cells by Northern gel analysis. A single band of 1.3 kb was detected in MOLT-3 cells (Fig. 1*a*).

To test whether this message was specific for MOLT-3 or T cells in general we probed RNA from several haematopoietic and non-haematopoietic cell lines and cells. In addition to MOLT-3, two other human T-cell lines (Jurkat and CEM) showed a single 1.3-kb band by Northern gel analysis (Fig. 1*a*). In contrast, no hybridization was evident with RNA from the following: two other human B-cell lines (RPMI 1788 and RPMI 3638), lymphocytes from a patient with B-cell chronic lymphocytic leukaemia (B-CLL), a human myeloid cell line (K562), a cell line derived from an acute myeloblastic leukaemia (AML) patient (provided by H. Messner), normal human bone marrow cells or a human bladder tumour line (MGHU-1). Additional T-cell populations were also examined, including fresh human thymocytes (provided by E. Gelfand), phytohaemagglutinin (PHA)-stimulated human peripheral blood T cells, and the murine T-cell line RBL-5—RNA from these sources all showed the typical single band in Northern gels (Fig. 1*b*). The level of hybridization to these RNAs, however, was varied. The thymus cells showed the greatest hybridization, followed by PHA-stimulated T lymphoblasts, and finally the murine T-cell line RBL-5. In contrast, the control RNA from the human B-cell line, HSC-58, and the RNA from a murine myeloma cell line, X63, showed no hybridization.

Taken together, these data show that the cDNA clones YT30, YT35, YT53 and YT76 are similar to RNA from T cells of both human and murine origin, but not to RNA from other haematopoietic or non-haematopoietic sources.

Protein identity of clone YT35

To determine the size of the protein encoded by the mRNA corresponding to the T cell-specific clones, we used the hybrid selection method of Parnes *et al.*[22]. Total mRNA from MOLT-3

87

was hybridized to clone YT35 or YT76. After the unattached mRNA was washed off, the annealed mRNA was released and translated *in vitro* using a rabbit reticulocyte lysate system. Figure 2 shows the protein that was synthesized. Two polypeptide bands were detected, one corresponding to a molecular weight (M_r) of ~45,000 (45K) and the other ~30K. The larger polypeptide was intrinsic to the *in vitro* translation system as it was present in the control without added RNA (data not shown), whereas the 30K polypeptide appeared to be encoded by the mRNA selected by these T cell-specific cDNAs.

Fig. 1 Northern blot examination of T cell and non-T-cell RNA for hybridization to clone YT35. T cell-specific clones were obtained by a differential hybridization described previously[37]. RNA was extracted from cells in the presence of guanidine thiocyanate[38], treated with glyoxal, then 10 μg were electrophoresed through 1% agarose in 10 mM sodium phosphate buffer, *p*H 7.0. Transfer of the RNA to nitrocellulose membranes was carried out as described[39]. Hybridization was to ^{32}P-dCTP-labelled, nick-translated clone YT35 (*a*) or a 0.4-kb *Bgl*II–*Xba*I fragment of YT35 which is internal to the cDNA insert (*b*). *a*, RNA is from: lane 1, the human bladder cell line MGHU-1 (provided by Dr I. Tannock); (2) normal human bone marrow; (3) a human erythroleukaemia-like cell line from an AML patient; (4) a human myeloid cell line, K562 (ref. 33); (5) human B-CLL cells; lanes 6–8, human B-cell lines[33], and lanes 9–11, human T-cell lines[18]. Cell lines in lanes 2, 3 and 5 were provided by Dr H. Messner. *b*, RNA is from: lane 1, human thymocytes (provided by Dr E. Gelfand); (2) PHA-stimulated human T-lymphocytes; (3) a human B-cell line[33]; (4) a murine T-cell line, RBL5 (provided by Dr R. Miller); and (5) a murine myeloma cell line X63 (provided by Dr N. Hozumi).

Nucleotide and protein sequences of YT35

The nucleotide sequence of clone YT35 was determined using the dideoxy chain terminating inhibitor method of Sanger *et al.*[23]: a sequence of 1,151 nucleotides is shown in Fig. 3. Examination of this sequence reveals a long open reading frame with a TGA termination triplet at position 974. A protein that initiates at the first ATG methionine codon in this reading frame (position 38) has a deduced molecular weight of 34,938, which is in general agreement with the results of hybrid selection and *in vitro* translation (Fig. 2). There are stretches of nonpolar and uncharged amino acids adjacent to the N and C terminal, and two possible sites of *N*-glycosylation (Fig. 3).

To establish whether the deduced amino acid sequence of the T cell-specific clone is similar to any of the immunoglobulin- or T cell-related protein sequences, we surveyed these proteins by diagonal dot-matrix analysis[24]. The protein sequences deduced from clone YT35 showed a long region of similarity to a murine λ light chain[25] (Fig. 4*a*) and human λ[26] and κ[27] light chains. This homology is more easily recognized when the background is reduced by looking for ≥35% similarity over 21 contiguous amino acids. The resulting plots (Fig. 4*b–d*) show that there are two stretches having high similarity: one near the C-terminal side of the variable region and the other near the *N*-terminal side of the constant region. Figure 5 shows a direct comparison of these sequences. The highest degree of similarity that was detected in Fig. 4 corresponds to the region surrounding the cysteine residues at positions 111 and 166 in the YT35 sequence. The sequences are quite similar near the cysteine residues at positions 42 and 231 also. Figure 5 also emphasizes the similarity of the amino acid sequences that was detected along the entire light-chain protein in Fig. 4*a*. Less pronounced homology was also found to the heavy chain of human[28] and murine[29] immunoglobulin (not shown), but there was no significant similarity to murine H-2 (ref. 30), Thy-1 (ref. 31) or human histocompatibility locus antigens[32]. The stretch of 249 amino acids that overlaps with the light chains is 33% identical to the murine λ light chain and 38% identical to the constant region. The similarity increases to 59% and 58%, respectively, if one takes into consideration conservative changes. The homology with the constant region of human λ light chain is comparable, with 36% identical amino acids and 61% if conservative changes are taken into account. Note that the protein encoded by clone YT35 is equally similar to the human and

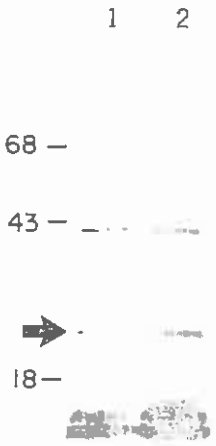


Fig. 2 SDS-polyacrylamide gel electrophoresis of the *in vitro* translation products of the mRNA that hybridizes to clones YT35 and YT76. *In vitro* translation products are from MOLT-3 RNA hybrid selected with clone YT76 (lane 1) or YT35 (lane 2). The positions of molecular weight standards run in neighbouring lanes are marked (in kilodaltons), and the 30K specific protein is indicated with an arrow. Hybrid selection was done according to the procedure of Parnes *et al.*[22]. Briefly, poly(A)$^+$ RNA (isolated by oligo(dT)-cellulose column chromatography) from MOLT-3 cells was annealed to 10 μg of plasmid YT35 or YT76 that had been bound to nitrocellulose filters. These were extensively washed to remove unbound RNA. Annealed RNA was released by boiling and translated in the presence of ^{35}S-methionine (Amersham) using rabbit reticulocyte lysate *in vitro* translation kits (BRL). The products were electrophoresed on a 10% polyacrylamide–SDS gel[40] which was treated with En^{3}Hance (NEN) dried and exposed to Kodak XAR-5 film at −70 °C.

```
                                                    MetAspSerTrpThrPheCysCysValSerLeuCysIleLeuValAlaLysHis  18
CTGGTCTAGAATATTCCACATCTGCTCTCACTCTGCCATGGACTCCTGGACCTTCTGCTGTGTGTCCCTTTGCATCCTGGTAGCGAAGCA
        10        20        30        40        50        60        70        80        90

 ThrAspAlaGlyValIleGlnSerProArgHisGluValThrGluMetGlyGlnGluValThrLeuArgCysLysProIleSerGlyHis  48
TACAGATGCTGCAGTTATCCAGTCACCCCGCCATGAGGTGACAGAGATGGGACAAGAAGTGACTCTGAGATGTAAACCAATTTCAGGCCA
       100       110       120       130       140       150       160       170       180

 AsnSerLeuPheTrpTyrArgGlnThrMetMetArgGlyLeuGluLeuLeuIleTyrPheAsnAsnAsnValProIleAspAspSerGly  78
CAACTCCCTTTTCTGGTACAGACAGACCATGATGCGGGGACTGGAGTTGCTCATTTACTTTAACAACAACGTTCCGATAGATGATTCAGG
       190       200       210       220       230       240       250       260       270

 MetProGluAspArgPheSerAlaLysMetProAsnAlaSerPheSerThrLeuLysIleGlnProSerGluProArgAspSerAlaVal 108
ATGCCCGAGGATCGATTCTCAGCTAAGATGCCTAATGCATCATTCTCCACTCTGAAGATCCAGCCCTCAGAACCCAGGGACTCAGCTGT
       280       290       300       310       320       330       340       350       360

 TyrPheCysAlaSerSerPheSerThrCysSerAlaAsnTyrGlyTyrThrPheGlySerGlyThrArgLeuThrValValGluAspLeu 138
TATTTCTGTGCCAGCAGTTTCTCGACCTGTTCGGCTAACTATGGCTACACCTTCGGTTCGGGGACCAGGTTAACCGTTGTAGAGGACCT
       370       380       390       400       410       420       430       440       450

 AsnLysValPheProProGluValAlaValPheGluProSerGluAlaGluIleSerHisThrGlnLysAlaThrLeuValCysLeuAla 168
GAACAAGGTGTTCCCACCGAGGTCGCTGTGTTTGAGCCATCAGAAGCAGAGATCTCCCACACCCAAAAGGCCACACTGGTGTGCCTGGC
       460       470       480       490       500       510       520       530       540

 ThrGlyPhePhoProAspHisValGluLeuSerTrpTrpValAsnGlyLysGluValHisSerGlyValSerThrAspProGlnProLeu 198
CACAGGCTTCTTTCCCGACCACGTGGAGCTGAGCTGGTGGGTGAATGGGAAGGAGGTGCACAGTGGGGTCAGCACAGACCCGCAGCCCCT
       550       560       570       580       590       600       610       620       630

 LysGluGlnProAlaLeuAsnAspSerArgTyrCysLeuSerSerArgLeuArgValSerAlaThrPheTrpGlnAsnProArgAsnHis 228
CAAGGAGCAGCCCGCCCTCAATGACTCCAGATACTGCCTGAGCAGCCGCCTGAGGGTCTCGGCCACCTTCTGGCAGAACCCCGCAACCA
       640       650       660       670       680       690       700       710       720

 PheArgCysGlnValGlnPheTyrGlyLeuSerGluAsnAspGluTrpThrGlnAspArgAlaLysProValThrGlnIleValSerAla 258
CTTCCGCTGTCAAGTCCAGTTCTACGGGCTCTCGGAGAATGACGAGTGGACCCAGGATAGGGCCAAACCCGTCACCCAGATCGTCAGCGC
       730       740       750       760       770       780       790       800       810

 GluAlaTrpGlyArgAlaAspCysGlyPheThrSerValSerTyrGlnGlnGlyValLeuSerAlaThrIleLeuTyrGluIleLeuLeu 288
CGAGGCCTGGGGTAGAGCAGACTGTGCCTTTACCTCGGTGTCCTACCAGCAAGGGGTCCTGTCTGCCACCATCCTCTATGAGATCCTGCT
       820       830       840       850       860       870       880       890       900

 GlyLysAlaThrLeuTyrAlaValLeuValSerAlaLeuValLeuMetAlaMetValLysArgLysAspPhe***                  312
AGGGAAGGCCACCCTGTATGCTGTGCTGGTCAGCGCCCTTGTGTTGATGGCCATGGTCAAGAGAAAGGATTTCTGAAGGCAGCCCTGGAA
       910       920       930       940       950       960       970       980       990

CTGGAGTTAGGAGCTTCTAACCCGTCATGGTTCAATACACATTCTTCTTTTGCCAGCGCTTCTGAAGAGCTGCTCTCACCTCTCTGCATC
      1000      1010      1020      1030      1040      1050      1060      1070      1080

CCAATACATCCCCCTATGTGCATGCACACCTGCACACTCACGGCTGAAATCTCCCTAACCCAGGGGGAC
      1090      1100      1110      1120      1130      1140      1150
```

Fig. 3 Nucleotide a... sequences of the T cell-sp... The deduced amino acid se... largest open reading frame is sho... the nucleotide sequence, with ove... indicating nonpolar amino acid stretches ne... the termini, and underlining indicating poss-ible sites of *N*-glycosylation. The map below shows the positions of the methionine codons (bars below line) and termination codons (bars above line) of the three coding frames. Presumptive initiation and termination codons of the protein encoded by YT35 are arrowed.

Methods: For DNA sequencing, clone YT35 was digested with restriction enzymes *Pvu*I and *Apa*I and the 1.3-kb insert, flanked by 0.1 kb of vector sequences at each end, was separated from the rest of the vector by electrophoresis through 1% low melting point agarose in 100 mM Tris-borate, 1 mM EDTA *p*H 8.3. The fragment was purified from the agarose by phenol and chloroform extraction followed by ethanol precipitation, and further digested with *Alu*I, *Fnu*4HI, *Hinf*I, *Rsa*I or *Sau*961. We rendered the termini of the resulting fragments blunt, where necessary, by treatment with the large fragment of *E. coli* DNA polymerase I (Boehringer Mannheim) and deoxynucleotide triphosphates, fragments were ligated into the *Sma*I site of the replicative form of the phage vector M13mp9, and sequenced as before[24].

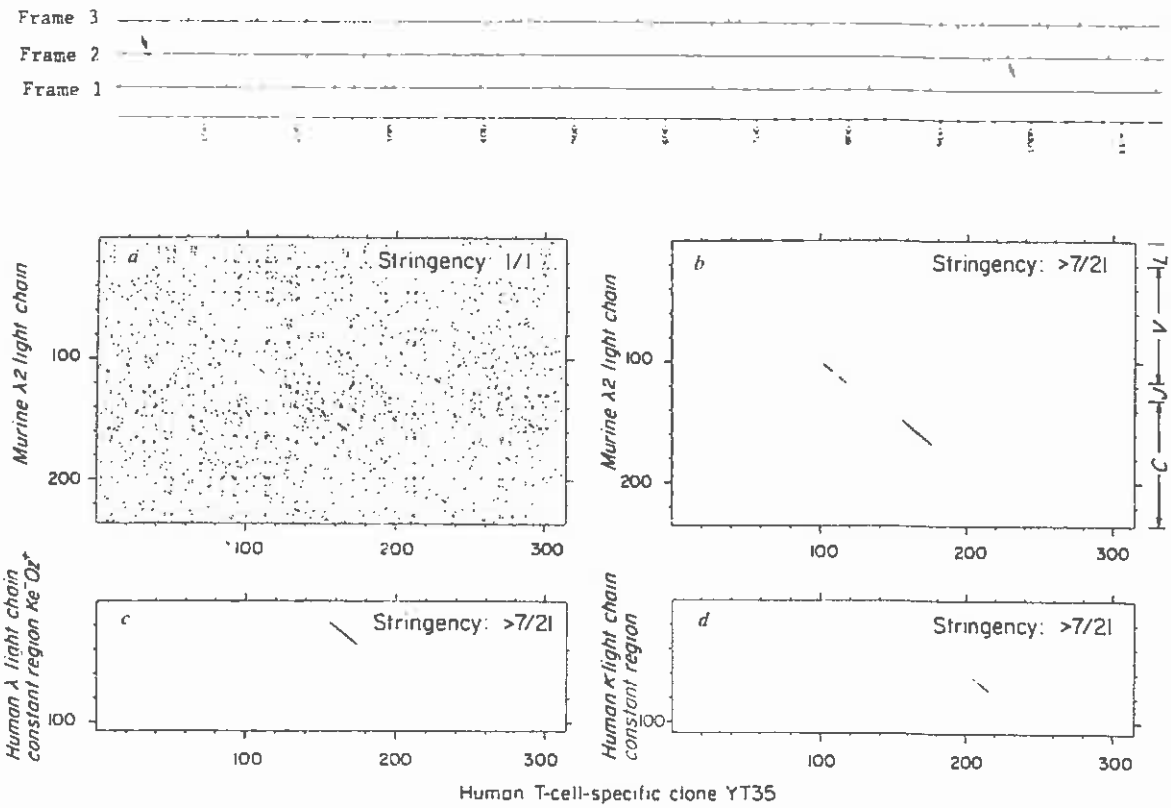


Fig. 4 Diagonal dot-matrix comparison of the deduced amino acid sequence of clone YT35 and the amino acid sequences of murine and human immunoglobulin light chains. The abscissa represents the YT35 sequence while the ordinate represents a murine λ2 light chain[25] (in *a* and *b*), and a human λ^{26} (*c*) or κ^{27} (*d*) light chain. In *a*, wherever the amino acids were the same at the corresponding *x* and *y* positions a dot was plotted, that is, the stringency was set to 1 out of 1. In the remaining panels the background was reduced by plotting a dot only when > 7 within a span of 21 consecutive positions were identical, that is, at ≥35% homology. L, leader peptide, *V*, variable region; *J*, joining segment; *C*, constant region.

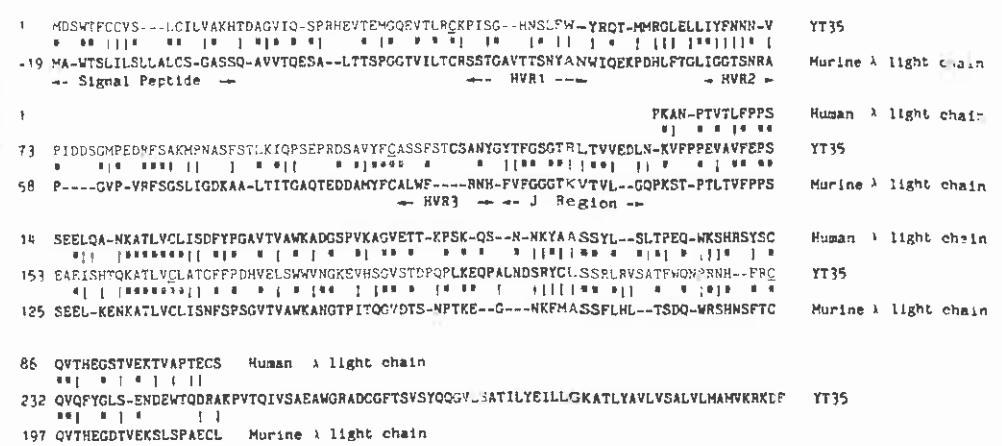


Fig. 5 Direct comparison of the deduced amino acid sequence of clone YT35 and those of murine[25] and human[26] λ light chains. The cysteine residues at positions 42, 111, 166 and 231 of YT35 are underlined. Identity between the sequences is indicated by an asterisk (*) and vertical lines (|) denote conservative changes. Gaps (–) were introduced to increase the match-up. To identify conservative changes, the amino acids were grouped as follows. Acidic: aspartic acid (D) and glutamic acid (E); basic: histidine (H), lysine (K) and arginine (R); uncharged polar: asparagine (N), glutamine (Q), serine (S) and threonine (T); nonpolar: alanine (A), cysteine (C), phenylalanine (F), glycine (G), isoleucine (I), leucine (L), methionine (M), proline (P), valine (V), tryptophan (W) and tyrosine (Y). Regions of interest in the murine λ light chain are marked. HVR, hypervariable region.

murine light chains, and that the similarity is highest where the human and murine light chains are most conserved between themselves. Also, the highest similarity to the λ and κ chains occurs at different positions (Fig. 4).

Implications

Here we have reported two important findings: first, we have isolated a human T cell-specific cDNA clone, and second, we have determined that the protein which it encodes, as deduced from the nucleotide sequence, resembles human and murine immunoglobulin light-chain molecules. The protein sequences are similar in the relative locations of the cysteine residues, but more importantly, they show extensive homology over the entire length of the variable, joining and constant regions. The stretch of hydrophobic amino acids at the N-terminus resembles a signal peptide, and the hydrophobic region near the C-terminus resembles a membrane anchor, implying that the protein may be transmembrane, with the major part exterior to the cell. These structural features suggest that the protein encoded by YT35 is part of the receptor that mediates specialized T lymphocyte function.

The finding of immunoglobulin-related sequences in a T cell-specific mRNA led us to consider the possibility that the sequence arose as an artefact during cDNA synthesis. We consider this unlikely for the following reasons. The lengths of the inserts of clones YT30, YT35 and YT53 are all 1.3 kb, which is the same size as the mRNA from the different T cells examined to which these clones hybridize. Also, the MOLT-3 cells from which we generated the clones do not produce immunoglobulin[33], thus it is unlikely that the immunoglobulin-related sequences arose during reverse transcription. Furthermore, the amino acid sequences of clone YT35 and immunoglobulin light chains, while being significantly similar, are not identical. Finally, cDNA clone YT35 does not hybridize to immunoglobulin mRNAs present in bone marrow cells and B-cell lines which produce the heavy and light chains (λ and κ) of immunoglobulin molecules[33] (Fig. 1). In fact, the homology between immunoglobulin light chains and clone YT35 is much more difficult to detect at the nucleotide level than at the amino acid level. Thus, we conclude that the light chain-like sequences are an inherent part of the T cell-specific mRNA.

This conclusion raises some interesting questions concerning the role of the YT35-encoded protein in the specialized functions of T cells. The nature and function of this protein are unknown; it may be similar to the known T cell-specific antigens[34], or to the α- or β-subunits of the recently identified putative T-cell receptors[11–16]. The physical properties of the protein encoded by clone YT35 cannot easily be compared with those of the other T cell-specific proteins which are post-translationally modified and have glycosyl side chains[11–16,34], even though there are potential *N*-glycosylation sites at positions 90 and 205 in the YT35 amino acid sequence. Nonetheless, the similarity in size, the relative locations of the cysteine residues and the homology to immunoglobulin light chains suggests that the protein encoded by YT35 and the immunoglobulin light chains may have evolved from the same ancestral protein and may have related functions. That the YT35-encoded protein is not more closely related to the human than the murine light chain suggests that these proteins may have diverged before the species diverged, while the ability of the human clone to hybridize to a murine T-cell mRNA indicates that the gene is highly conserved between the species. The conservation of sequence around the cysteine residues suggests the possibility that the disulphide linkages are maintained and that the overall structure of these molecules is similar. When the immunoglobulin light chains are aligned with the YT35-encoded protein to show maximum homology, the C-terminus of the YT35 protein, which has a stretch of 17 uncharged or nonpolar amino acids followed by four charged amino acids, extends past that of the light chains (Fig. 5)—this is reminiscent of membrane-bound immunoglobulin molecules on the surface of B cells, where the heavy chains have an extra peptide at the C-terminus with a similar structure that anchors the molecule to the membrane[35,36].

Alternatively, it is possible that the similarity of the structure and sequence of the YT35-encoded protein to the immunoglobulin molecules is fortuitous and that this protein is involved in other T-cell functions. Results in our laboratory do indeed show rearrangement of DNA fragments homologous to YT35 in T-cell lines (B. Toyonaga *et al.*, manuscript in preparation). These results, taken together with the striking similarity of the deduced amino acid sequence of YT35 to the immunoglobin chain, suggest that YT35 may indeed correspond to a message that specifies the human T-cell receptor.

We thank Drs E Gelfand, N. Hozumi, I. Tannock, H. Messner and R. G. Miller for providing cells and cell lines, and Drs E. A. McCulloch, R. G. Miller, R. A. Phillips and Cox Terhorst for discussions. This work was supported by the NCI of Canada. Y.Ya., Y.Yo. and S.P.C. are recipients of awards from the MRC of Canada.

Received 15 December 1983, accepted 24 January 1984

1 Porter, R R *Science* **180**, 713–716 (1973)
2 Edelman, G M *Science* **180**, 830–840 (1973)
3 Hozumi, N & Tonegawa, S *Proc natn Acad Sci USA* **73**, 3628–3632 (1976)
4 Honjo, T *Ann Rev Immun* **1**, 499–528 (1983)
5 Rabbits, T H, Bentley, D L & Milstein, C P *Immun Rev* **59**, 69–91 (1981)
6 Tonegawa, S *Nature* 302, 575–581 (1983)
7 Capra, J D & Kehoe, J M *Adv Immun* **20**, 1–40 (1975)
8 Eichmann, K *Adv Immun* **26**, 195–254 (1978)
9 Janeway, C A & Jason, J M *CRC Crit Rev Immun* **1**, 133–164 (1980)
10 Miller, J F A P, Marohan, G & Walker, I D *Immun Today* **4**, 141–143 (1983)
11 Allison, J P, McIntyre, B W & Bloch, D *J Immun* **129**, 2293–2300 (1982)
12 Haskins, K *et al J exp Med* **157**, 1149–1169 (1983)
13 Meuer, S C *et al J exp Med* **157**, 705–719 (1983)
14 Acuto, O *et al Cell* **34**, 717–726 (1983)
15 Kappler, J *et al Cell* **34**, 727–737 (1983)
16 McIntyre, B W & Allison, J P *Cell* **34**, 739–746 (1983)

17 Minawada, J., Okamura, T. & Moore, G. E. *J. natn. Cancer Inst.* 49, 891–895 (1972)
18 Nagasawa, K. & Mak, T. W. *Cell. Immun.* 71, 396–403 (1982)
19 Delia, D., Greaves, M. F., Newman, R. A. & Sutherland, D. R. *Int. J. Cancer* 29, 23–31 (1982)
20 Land, H. *et al. Nucleic Acids Res.* 9, 2251–2266 (1981)
21 Clark, S. P. & Mak, T. W. *Nucleic Acids Res.* 10, 3315–3330 (1982)
22 Parnes, J. R. *et al. Proc. natn. Acad. Sci. U.S.A.* 78, 2253–2257 (1981)
23 Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* 74, 5463–5467 (1977)
24 Clark, S. P. & Mak, T. W. *Proc. natn. Acad. Sci. U.S.A.* 80, 5037–5041 (1983)
25 Bo[illegible] A. L. M. *et al. Nature* 298, 380–382 (1982)
26 Meuer, P. A., Hodgkin, G. F., Korsmeyer, S. J., Waldmann, T. A. & Leder, P. *Nature* 294, [illegible]–540 (1981)
27 Heiter, P. A., Max, E. E., Seidman, J. G., Maizel, J. V. Jr & Leder, P. *Cell* 22, 197–207 (1980)
28 Matthyssens, G. & Rabbitts, T. H. *Proc. natn. Acad. Sci. U.S.A.* 77, 6561–6565 (1980)
29 Bernardo, O. & Gough, N. M. *Proc. natn. Acad. Sci. U.S.A.* 77, 3634–3636 (1980)
30 Bregegere, A. J. P. *et al. Nature* 292, 78–81 (1981)
31 Moriuchi, T., Chang, H. C., Denome, R. & Silver, J. *Nature* 301, 80–82 (1983)
32 Malissen, M., Malissen, B. & Jordan, B. R. *Proc. natn. Acad. Sci. U.S.A.* 79, 893–897 (1982)
33 Chechik, B. S. & Minowada, J. *J. natn. Cancer Inst.* 63, 609–614 (1979)
34 Reinherz, E. L. & Schlossman, S. F. *Cell* 19, 821–827 (1980)
35 Alt, F. *et al. Cell* 20, 293–301 (1980)
36 Rogers *et al. Cell* 20, 303–312 (1980)
37 Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* 113, 237–251 (1977)
38 Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* 18, 5294–5299 (1979)
39 Thomas, P. S. *Proc. natn. Acad. Sci. USA* 77, 5201–5205 (1983)
40 Laemmli, U. K. *Nature* 277, 680–685 (1970)

Isolation of cDNA clones encoding T cell-specific membrane-associated proteins

Stephen M. Hedrick*‡, David I. Cohen*†, Ellen A. Nielsen* & Mark M. Davis*†‡

* Laboratory of Immunology, NIAID, National Institutes of Health, Bethesda, Maryland 20205, USA
† Department of Medical Microbiology, Stanford University School of Medicine, Stanford, California 94305, USA

Of 10 distinct cloned DNA copies of mRNAs expressed in T lymphocytes but not in B lymphocytes and associated with membrane-bound polysomes, one hybridizes to a region of the genome that has rearranged in a T-cell lymphoma and several T-cell hybridomas. These characteristics suggest that it encodes one chain of the elusive antigen receptor on the surface of T lymphocytes.

T LYMPHOCYTES, like B lymphocytes, are capable of recognizing a wide range of different antigens[1–3]. As with B cells, the ability to recognize a given antigen is fixed in any particular clonal line of T cells. However, unlike B cells, T cells appear to recognize antigens in combination with self major histocompatibility (MHC) determinants[4–6]. In view of the similarities to B cells, early ideas as to how T cells recognize antigens centred on the use of either entire antibody molecules[7] or at least some of the separately encoded (in the germ-line genome) segments that make up the antigen-binding sites of immunoglobulin heavy and light chains[8–12]. But despite early reports that antibodies against immunoglobulin antigen-binding sites can react with T cells[13–15] and recognize a target closely linked to the immunoglobulin heavy-chain locus[16,17], attempts to demonstrate an involvement of immunoglobulins in T-cell antigen recognition have proved consistently negative[8–12].

More recent investigations have taken a route largely independent of the antibody models and have succeeded in raising at first antisera[18] and then monoclonal antibodies[19–22] that specifically recognize particular T-cell lines or hybridomas. Some of these antibodies have been shown either to inhibit[20–22] or stimulate[18] the response of a cell in a clone-specific fashion. Several groups have used these antibodies to immunoprecipitate disulphide-linked heterodimers, composed of two distinct glycoproteins of molecular weights 37,000–50,000 (refs 19–22), both of which appear to have variable and constant regions[23,24].

We have taken a different approach to this problem and have attempted to use the techniques of molecular genetics to isolate from a series of antigen-specific, MHC-restricted T helper hybridomas[25,26] the genes that encode their antigen receptors. Our approach was based on the following assumptions about the nature of the T-cell receptor genes: (1) That they should be expressed in T cells but not in B cells. (2) That the mRNAs for the T-cell receptor proteins should be found on membrane-bound polysomes, as one would expect the nascent receptor polypeptides to attach to the endoplasmic reticulum by a leader peptide, or signal sequence[27]. (3) That like immunoglobulin genes those that encode the T-cell receptor, proteins should be rearranged in T cells as a mechanism of generating diversity and consequently increasing the antigen-recognition repertoire. (4) That like immunoglobulin genes they should have constant regions (as they presumably share at least some functions) and variable regions, would confer the antigen-binding specificity.

An experimental strategy could be developed on the basis of these assumptions, as B and T cells differ in only a small fraction of their gene expression (~2%, or 200–300 different sequences[28]; M.M.D. and W. E. Paul, manuscript in preparation) and only a small proportion of lymphocyte mRNAs appear to be in the membrane-bound polysomal fraction (~3%)[29]. Thus, by synthesizing ^{32}P-labelled DNA copies (cDNAs) of the membrane-bound polysomal RNA of antigen-specific T cells and removing by RNA hybridization those sequences also expressed in B cells, one should be left with a specific probe representing a very small fraction of total T-cell gene expression and likely to include copies of the mRNAs that encode the T-cell receptor. This probe can then be used to screen a library of cloned cDNAs[30], and the clones thus identified could be used to look for somatic gene rearrangements in T cells. Further restriction mapping and sequence analysis of the cDNA clones derived from different T cells might reveal variable and constant regions. We have carried through this strategy, and have managed to obtain a cDNA clone (TM86) representing mRNA that is expressed specifically in T cells, found in the membrane-bound polysomal fraction and encoded by genomic DNA that has rearranged in T cells.

Analysis of T-cell mRNA

The number of different species in a given mRNA population can be estimated from an analysis of the kinetics of cDNA–mRNA hybridization reactions[31]. In this case, labelled cDNA was synthesized from cytoplasmic poly(A)$^+$ RNA of the T helper hybridoma M12 (ref. 26), the RNA template was removed and the cDNA hybridized to the original RNA. The resulting C_0t

91

‡ Present addresses: Department of Biology, University of California at San Diego, La Jolla, California 92093, USA (S.M.H.); Stanford University School of Medicine (M.M.D.).

Table 1 Analysis of C_0t curves

Labelled cDNAs	Component	Fraction	$C_0t_{1/2}$	$C_0t_{1/2}$ Pure	No of species	mRNAs per cell
Cytoplasmic	1	48.6%	0.52	0.18	130	1,050
	2	51.4%	47.4	17.7	12,900	12
Free	1	56.8%	0.12	0.048	—	—
	2	43.2%	25.6	7.77	—	—
Membrane-bound	1	67.6%	0.12	0.058	—	—
	2	32.4%	15.7	3.8	—	—
MB $T^*_H - B$	1	41.7%	1.25	0.25	1	—
	2	58.3%	92.8	26.0	70	—

MB T^*_H subtraction

Reaction	Double-stranded	Single-stranded	
1	91.3%	8.7%	$(T^*_{M12} - B_{L10A})$
2	32.5%	67.7%	$(T^*_{M12} - B_{L10A})$
3	44.4%	55.6%	$(T^*_{M12} + T_{M12})$

Least-squares fit analysis[32] of the reactions shown in Fig. 1. The curves were fitted to two components with root mean square (r.m.s.) errors of 0.03–0.05. The fractions were normalized to 100% and both uncorrected and 'pure' $C_0t_{1/2}$ values are shown. The number of species was determined using the corrected $C_0t_{1/2}$ values and an average mRNA size of 2,000 nucleotides[29]. Purified ovalbumin mRNA (given by G. Stanley McKnight) was used as a kinetic standard. The calculation of number of mRNAs per cell is based on the average of 300,000 cytoplasmic poly(A)$^+$ RNAs we observe in most T hybridomas in log phase growth. The MB $T^*_H - B$ analysis draws on the data in Fig. 1*b*. As we reproducibly find that the remaining B cell-reactive material consists largely of small, very slowly reacting cDNA which introduces a physiologically irrelevant slow component into the C_0t analysis (M. M. D. and W. E. Paul, manuscript in preparation), we subtracted the background curve directly from the homologous curve and computer fitted the difference. In this case, the L10A reaction with the MB $T^*_H - B$ is subtracted from the M12 reaction and fitted as shown in this table. The second part of the table summarizes the cDNA subtraction, with two successive subtractive hybridizations with B-cell (L10A) RNA follow by a back reaction with homologous (M12) RNA to remove unreactable sequences. The repetitive yield is 50–65% for each round of reaction and it is not selective for single-stranded or double-stranded nucleic acid (M. M. D. and W. E. Paul, manuscript in preparation). The single-stranded fraction is carried through to the next step from both reactions 1 and 2, and the double-stranded fraction of reaction 3 is the material analysed in Fig. 1*b* and which represents 2.6% of input membrane-bound polysomal cDNA, normalizing for yield losses. As the analytical reactions (Fig. 1*b*) indicate that 61% of this material represents T cell-specific sequences, 1.6% (2.6×0.61) of the initial cDNA is unique to T cells. As the membrane-bound RNA represents only 15% of the total, this value is further reduced by this fraction (1.6%× 0.15) to yield 0.24% of the mass of the cytoplasmic mRNA. As perhaps an equal amount of these same sequences are found in the free polysomal fraction (see text), the value of 0.24% must be revised upwards to 0.5%. This number is then divided into abundant and rare fractions as indicated in the table and the number of sequences of average size derived (~1 for component 1 and 70 for component 2).

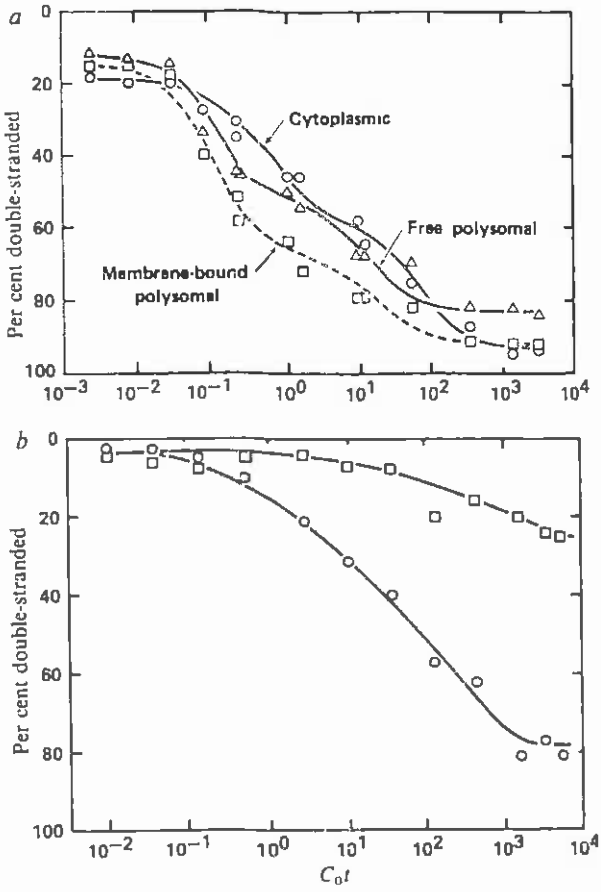


Fig. 1 Kinetic hybridization analysis of T-helper hybridoma mRNAs. *a*, Hybridization to an excess of cytoplasmic poly(A)$^+$ RNA of: ○, total cytoplasmic cDNA; △, free polysomal cDNA; and □, membrane-bound cDNA. *b*, Hybridization of the MB $T^*_{M12} - B_{L10A}$ probe to M12 cytoplasmic poly(A)$^+$ RNA (○) and hybridization of the same probe to L10A poly(A)$^+$ cytoplasmic RNA (□).

Methods. Cytoplasmic poly(A)$^+$ RNA was prepared as described[46] from an *in vitro* cultured hybridoma line (M12). Membrane-bound and free polysomal poly(A)$^+$ RNAs were prepared from these cells by the procedure of Melcher and Rabbitts[29]. ^{32}P-labelled cDNA was synthesized from these RNAs using oligo(dT) and reverse transcriptase[30]. cDNAs were depleted of template RNA by base hydrolysis and reacted in 0.5 M phosphate buffer[47], 0.1% SDS, 5 mM EDTA in sealed glass capillaries with an excess of cytoplasmic poly(A)$^+$ RNA. Reactions were assayed by hydroxyapatite chromatography and analysed as described in Table 1. C_0t values as given are normalized for the effect of the high salt[47].

(or R_0t) curve was computer-fitted by the least-squares program of Pearson, Davidson and Britten[32] and is shown in Fig. 1*a*. The analysis of this curve, presented in Table 1, indicates that there are about 13,000 different mRNA species in this cell line, most of which are in the rare abundance class (component 2 in Table 1) at about 12 molecules per cell. These findings are typical of a variety of *in vitro* cultured cell lines[33] although somewhat higher than the 8,000 different species seen in a murine plasma cell line[29]. Also shown in Fig. 1*a* are the reactions of cytoplasmic poly(A)$^+$ RNA with labelled cDNAs prepared from membrane-bound and free polysomal fractions of RNA, using the procedure of Mechler and Rabbits[29]. Although a striking feature of that report was the detection of virtually no rare mRNA species in the membrane-bound fraction of a plasmacytoma, the analysis of the data presented here (Table 1) clearly indicates the existence of a rare abundance class component of the equivalent fraction in M12 cells.

When cDNA from membrane-bound polysomal RNA is repeatedly hybridized to B-cell mRNA and fractionated on hydroxyapatite (Table 1), the non-hybridizing DNA consists of 2.6% of the input cDNA, and the reactions shown in Fig. 1*b* (and analysed in Table 1) indicate that 61% of this fraction is T cell-specific and divided as shown between rare and moderately abundant classes (components 2 and 1 respectively, the $C_0t_{1/2}$ 'pures' of which are almost indistinguishable from those of cytoplasmic cDNA). Although only 15% of the polyribosomal RNA was membrane-bound in these cells, the T cell-specific genes were found to be as prevalent in the free polysomal as in the membrane-bound fractions of RNA. We therefore estimate that the pool of membrane-bound RNA derives from ~30%

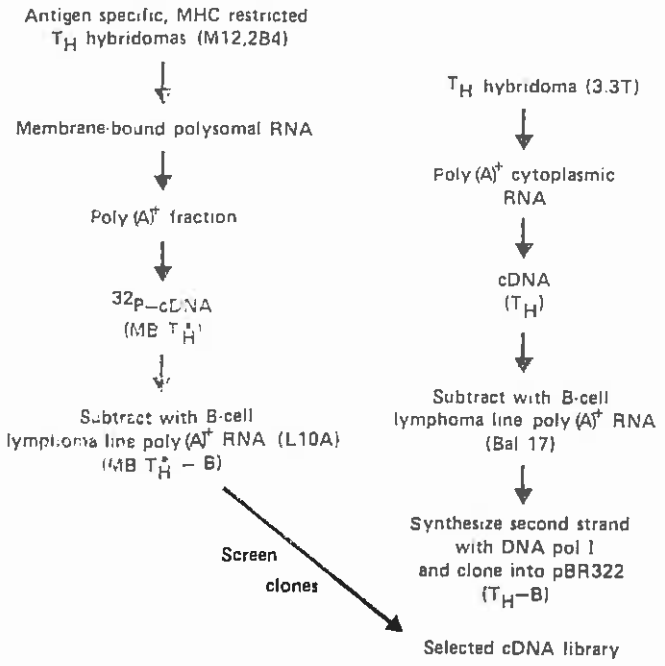


Fig. 2 Strategy of clone isolation. ^{32}P-labelled cDNA was synthesized from membrane-bound polysomal RNA of T_H hybridomas and subtracted with B-cell mRNA (L10A). These probes were then used to screen a selected cDNA library (T_H – B) constructed as described[30] from another T_H hybridoma/B cell combination.

of the total polysomal RNA. Thus, the subtracted cDNA ($MBT^*_{M12} - B_{L10A}$) represents ~0.5% of the input cDNA and ~70 different species (see Table 1 legend). This latter number could actually be an overestimate for the following reasons: (1) Rare mRNAs have, in several cases[34,35], been shown to be larger than the average mRNA. (2) A certain fraction (0.2–0.4%) of the cDNA differences between any two mRNA populations are not reproducibly expressed in other cells of the same type, indicating a certain background level of random gene expression or loss of expression (M.M.D. and S.M.H., unpublished observation). (3) The use of total cytoplasmic RNA rather than polysomal may introduce a significant number of non-translated sequences into the estimates.

cDNA clone isolation

The strategy used to isolate cDNA clones representing membrane-bound mRNAs expressed in T but not in B cells is shown in Fig. 2. Membrane-bound T helper (T_H) cell cDNA probes were subtracted with B-cell mRNA and used to screen a cDNA library that was itself the product of a $T_H - B$ cell reaction, and which was constructed as described previously[30,36] for a B cell-specific library[30]. The $T_H - B$ library was 20-fold enriched for T cell-specific sequences as judged by the fact that 95% of the mass of the cDNA was removed in the subtraction (at the hydroxyapatite stage). As indicated in Fig. 2 the library of 5,000 selected clones was screened and rescreened by standard procedures[37,38] using the $MBT^*_{2B4} - B_{L10A}$ probe. Thirty-five clones were screened and rescreened positive (TM clones). In order to determine which clones were derived from the same mRNA species and which were different, as well as to remove any false

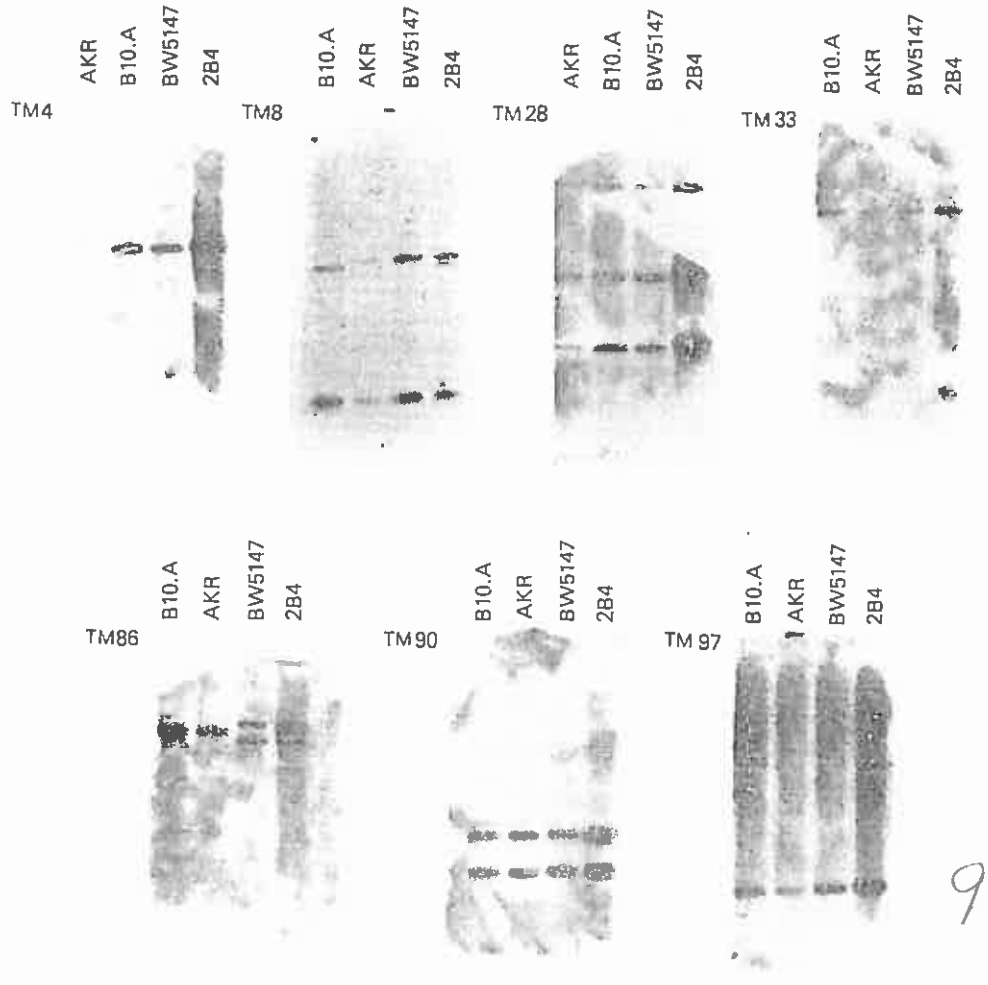


Fig. 3 Southern blot analysis of TM clones. ^{32}P-labelled inserts from the seven clones indicated were hybridized to Southern blots containing DNA from liver cells of strains B10.A and AKR, and DNA from BW5147 and the T_H hybridoma 2B4. The DNAs were prepared by standard methods[48], digested with the restriction enzyme *Pvu*II, electrophoresed through 0.9% agarose and blotted onto nitrocellulose. Hybridization was in 50% formamide, 5×SSPE[48] at 42 °C. Washing was done in 2×SSPE at 55 °C.

93

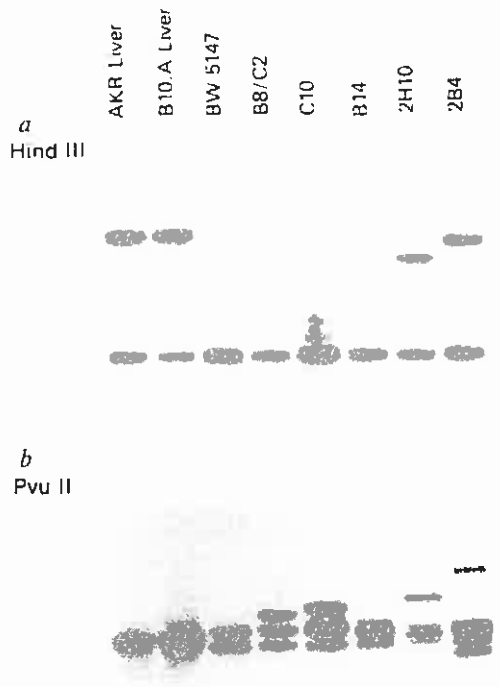


Fig. 4 Analysis of other T_H hybrids for rearrangement of TM86. In this case, four additional antigen-specific hybridoma lines were examined for rearrangement using the TM86 probe and the procedures outlined in Fig. 3 legend. B8/C2 and B14 are T_H hybrids specific for insulin and C10 is a T_H hybrid for lysozyme[49].

positives, we ^{32}P-labelled by nick-translation each positive clone to make labelled probes which were then hybridized to representative Northern blots. Five of the positive clones were reactive with B-cell mRNA, and the remaining 30 fell into one of 10 distinct patterns of mRNA size and expression. One of these 10 different clones cross-hybridized strongly with and gave a Northern blot pattern superimposable on a rat Thy-1 cDNA clone isolated by Silver and colleagues[39]. As Thy-1 is a classical T-cell membrane antigen not expressed on B cells, the fact that it is one of the genes isolated in this small set is an important internal control.

Our procedure for obtaining cDNAs of membrane-bound, T cell-specific mRNAs should have enriched for particular sequences to such an extent that even very low abundance mRNAs are represented at levels comparable to those of abundant mRNAs in the original population. This should overcome the difficulties earlier workers have found in detecting low abundance mRNAs[40,41]. In fact, hybridization of the inserts of the less abundant TM clones to the 3.3T-selected cDNA library in at least two cases indicates that we have been able to isolate DNA copies of mRNAs represented by only 1 in 100,000 clones in the unselected library.

Somatic gene rearrangements

To determine whether any of the seven TM clones which, on Northern blotting, were shown to represent mRNAs large enough to encode T-cell receptor proteins[19–22], had been copied from mRNAs encoded by genes that, like immunoglobulin genes, rearrange somatically, we prepared labelled probes from them which we hybridized to Southern blots of genomic DNA from the thymoma BW5147 (from the mouse strain AKR), AKR liver cells, the antigen-specific T-cell line 2B4 (a fusion of T cells from B10.A mice with BW5147) and B10.A liver cells. The autoradiograms from Southern blots are shown in Fig. 3. Except for the restriction polymorphism between AKR and B10.A seen with TM8 (Thy-1), the patterns of hybridization with each clone were identical for all the sources of DNA except in the case of TM86 (Fig. 3). Clearly, there was a strikingly different pattern of *Pvu*II fragments that hybridized to the clone from either BW5147 or 2B4 compared with liver DNA from either of the parental strains. The clones surveyed here were also hybridized to *Eco*RI and *Hin*dIII digests of genomic DNA, and in each case only TM86 showed a significant difference between the T-cell DNAs and liver DNAs.

It would be expected that genomic rearrangements of a receptor gene should be different for T cells of different antigen specificities. To test this possibility, genomic blots consisting of DNA from five antigen-specific T-cell hybridomas were hybridized with a nick-translated insert from clone TM86. The results, shown in Fig. 4, demonstrate that DNA from each of the antigen-specific T cells gives a different pattern. Three different B-cell lymphoma DNAs give patterns identical to that of the liver (data not shown), indicating that this type of rearrangement may only take place in T cells. It is also significant that no restriction fragment length polymorphisms can be detected with TM86 between the two strains for which data are shown here or in the three other strains that we have examined, either with the *Pvu*II enzyme or several others. This diminishes the possibility that the rearrangements could be due to the presence of a viral sequence since one would expect a highly unstable integrated viral sequence to be changing at least as dramatically in evolutionary time.

Discussion

As previously demonstrated[43,44] the techniques used in this report provide a general means of isolating sets of genes that are particularly important to the differentiated state of a cell. This set should include at least a portion of the genes required for the specialized functions of the cell, as well as genes involved in regulation and cellular differentiation. As the set of DNA clones obtained is very small, elaborate screening procedures can be used to further define them. In cases where the target population is very small (~0.5% for the set of genes examined here) the enrichment of the relevant probe (~×200) is sufficient to allow the isolation of very rare genes. For the studies reported here, we were able to screen a small set of cDNA clones for possible somatic rearrangement of the genes encoding the mRNAs from which they were derived. Such a screening procedure would be much too laborious if used to screen a large number of cDNA clones. An extensive characterization of TM86 and related cDNA clones is presented in the accompanying report[42].

The striking characteristic of TM86 is that the genes undergo somatic rearrangement in T cells but not B cells. The pattern of *Pvu*II fragments from liver DNA which hybridize to TM8 consists of three closely spaced bands of 5.2–6.5 kilobases (kb). The T lymphoma BW5147 has retained none of the germ-line fragments, and instead shows a pattern of two bands different from any in the liver DNA. We conclude that BW5147 has rearranged one or both of its gene copies; if it has rearranged only one, then it must have lost the other. Consistent with the notion that TM86 represents one chain of the T-cell receptor, each of the different antigen-specific T-cell hybridomas has one or more new homologous restriction fragments. Interestingly, each of the antigen-specific T-cell hybridomas has the same two fragments as BW5147, but has lost various of the germ-line fragments. Considering that immunoglobulin genes are the target of oncogene translocations in B lymphomas and plasmacytomas[45], it is attractive to speculate that the maintenance of the rearranged chromosomes of BW5147 reflects the translocation of an oncogene to TM86.

We thank E. Heber-Katz, D. Hansburg, L. Samuelson and R. Schwartz for providing the T-helper hybridoma lines used here; J. Kim and R. Asofsky for the B-cell lymphoma lines; D. Margulies and R. Germain for some of the cell line and liver DNAs; J. Silver for the rat Thy1 cDNA clone; and G. Stanley McKnight for the generous gift of purified ovalbumin mRNA. We thank R. Schaller, J. Rothbard and N. Gascoigne for critical reading of the manuscript. We also thank W. E. Paul for continual support and encouragement.

Received 25 January; accepted 31 January 1984.

1. Katz, D. H. *Lymphocyte Differentiation, Recognition and Regulation*, 152–187 (Academic, New York, 1977).
2. Mellief, C. J. M., deWaal, L. P., van der Meulen, M. Y., Melvold, R. W. & Kohn, H. I. *J. exp. Med.* **152**, 993–1013 (1980).
3. von Boehmer, H. & Haas, W. *Immun. Rev.* **54**, 27–56 (1981).
4. Zinkernagel, R. M. & Doherty, P. C. *Nature* **248**, 701–702 (1974).
5. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51–177 (1979).
6. Fink, P. J. & Bevan, M. *J. exp. Med.* **148**, 776–785 (1978).
7. Warr, G. W., DeLuca, D. & Marchalonis, J. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2476–2480 (1976).
8. Kronenberg, M., Davis, M. M., Early, P. W., Hood, L. E. & Watson, J. D. *J. exp. Med.* **152**, 1745–1761 (1980).
9. Nakanishi, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 6984–6988 (1982).
10. Kurosawa, Y. H. *et al. Nature* **290**, 565–570 (1981).
11. Kronenberg, M. *et al. J. exp. Med.* **158**, 210–227 (1983).
12. Kraig, E. *et al. J. exp. Med.* **158**, 192–209 (1983).
13. Eichmann, K. & Rajewsky, K. *Eur. J. Immun.* **5**, 661–666 (1975).
14. Binz, H. & Wigzell, H. *J. exp. Med.* **142**, 197–211 (1975).
15. Augustin, A. A., Julius, M. H., Cosenza, H. & Matsunaga, T. in *Regulatory T Lymphocytes* (eds Pernis, B. & Vogel, H. J.) 171–184 (Academic, New York, 1980).
16. Berek, C., Taylor, B. A. & Eichmann, K. *J. exp. Med.* **144**, 1164–1175 (1976).
17. Strassman, G., Lifshitz, R. & Mozes, E. *J. exp. Med.* **152**, 1448–1452 (1980).
18. Infante, A. J., Infante, P. D., Gilles, S. & Fathman, C. G. *J. exp. Med.* **155**, 1100–1107 (1982).
19. Allison, J. P., MacIntyre, B. W. & Block, D. *J. Immun.* **129**, 2293–2300 (1982).
20. Haskins, K. *et al. J. exp. med.* **157**, 1149–1169 (1983).
21. Meuer, S. C. *et al. Nature* **303**, 808–810 (1983).
22. Samuelson, L. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 6972–6976 (1983).
23. McIntyre, B. W. & Allison, J. P. *Cell* **34**, 739–746 (1983).
24. Acuto, O. *et al. Cell* **34**, 717–726 (1983).
25. Heber-Katz, E. *et al. J. exp. Med.* **155**, 1086–1099 (1982).
26. Hedrick, S. M. *et al. Cell* **30**, 141–152 (1982).
27. Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 835–851 (1975).
28. Davis, M. M., Cohen, D. I., Nielsen, E. A., DeFranco, A. D. & Paul, W. E. *UCLA Symp.* **24**, 215–220 (1982).
29. Mechler, B. & Rabbitts, T. H. *J. Cell Biol.* **88**, 29–36 (1981).
30. Davis, M. M. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
31. Bishop, J. O., Morton, J. G., Rosbash, M. & Richardson, M. *Nature* **250**, 199–204 (1974).
32. Pearson, W. R., Davidson, E. H. & Britten, R. J. *Nucleic Acids Res.* **4**, 1727–1735 (1977).
33. Lewin, B. in *Gene Expression* Vol. 2, 2nd edn (Wiley, New York, 1980).
34. Meyuhas, O. & Perry, R. P. *Cell* **16**, 139–148 (1979).
35. Milner, R. J. & Sutcliffe, J. G. *Nucleic Acids Res.* **11**, 5497–5520 (1983).
36. Sargent, T. & Dawid, I. *Science* **222**, 135–139 (1983).
37. Hanahan, D. & Meselson, M. *Gene* **10**, 63–67 (1980).
38. Ish-Horowicz, D. & Burke, J. F. *Nucleic Acids Res.* **9**, 2989–2998 (1981).
39. Morsuchi, T., Chang, H. C., Denome, R. & Silver, J. *Nature* **301**, 80–82 (1983).
40. Dworkin, M. B. & David, I. B. *Devl Biol.* **76**, 435–448 (1980).
41. Crampton, J., Humphries, S., Woods, D. & Williamson, R. *Nucleic Acids Res.* **8**, 6007–6017 (1980).
42. Hedrick, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I. & Davis, M. M. *Nature* **308**, 153–158 (1984).
43. Mather, E. L. *et al. Cell* **23**, 369–378 (1981).
44. Zimmerman, C. R., Orr, W. C., Leclerc, R. F., Barnard, E. C. & Timberlake, W. E. *Cell* **23**, 709–715 (1980).
45. Shen-Ong, G. L. C., Keath, E. J., Piccoli, S. P. & Cole, M. D. *Cell* **31**, 443–452 (1982).
46. Mushinski, J. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7405–7409 (1980).
47. Britten, R. J., Graham, D. E. & Neufeld, B. R. *Meth. Enzym.* **29**, 363–418 (1974).
48. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1982).
49. Glimcher, L. H. & Sherach, E. M. *J. exp. Med.* **156**, 640–645 (1982).

Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins

Stephen M. Hedrick*‡, Ellen A. Nielsen*, Joshua Kavaler†, David I. Cohen*† & Mark M. Davis*†‡

* Laboratory of Immunology, NIAID, National Institutes of Health, Bethesda, Maryland 20205, USA
† Department of Medical Microbiology, Stanford University School of Medicine, Stanford, California 94305, USA

Comparison of the sequence of a cloned T cell-specific cDNA with those of cross-reacting cloned cDNAs isolated from a thymocyte library indicates the presence of variable, constant and joining regions remarkably similar in size and sequence to those encoding immunoglobulin proteins. Together with the evidence for somatic gene rearrangements reported in the accompanying paper, this strongly suggests that the TM86 cDNA clone encodes one chain of the T-cell receptor for antigen.

IN THE preceding paper[1] we reported the isolation of a cDNA clone, TM86, which represents a species of mRNA that is expressed in T cells but not in B cells, encodes a membrane-associated protein, and is rearranged in T cells. We report here our further analysis by nucleotide sequencing of clone TM86, and also of several cross-reacting clones isolated from a thymocyte cDNA library. Our data reveal a remarkable resemblance between the amino acid sequences of the proteins encoded by the mRNAs represented by the TM86 and related cDNA clones and the immunoglobulin proteins of B cells.

Immunoglobulin proteins have a characteristic primary and secondary structure made up of a series of intrachain disulphide-bonded domains (as reviewed in ref. 2). Both heavy and light immunoglobulin chains begin with a leader peptide of 17–29 residues, followed by a variable region of 94–97 residues then a joining region of 13–17 residues (in heavy chains the variable and joining regions are separated by short 'diversity' elements of 1–14 residues). These elements are encoded separately in the germ-line genome by gene segments that are brought together during B-cell ontogeny[3–6] to generate an enormous range of different antibody genes. The domains of the constant regions of immunoglobulins (of variable number depending on polypeptide chain type and class) are encoded by separate exons[7], but they do not appear to rearrange during development. The sequences of our T cell-specific cDNA clones has revealed a similar arrangement of constant and variable regions, strongly suggesting that the mRNAs that they represent encode polypeptide components of the T-cell antigen receptor.

Characterization of cDNA clones

In view of the evidence presented in the preceding paper that the region of the genome represented by cDNA clone TM86 is rearranged in T cells, we sought to isolate cDNAs representative of independent rearrangement events by screening a thymocyte cDNA library (provided by C. Benoist) with TM86 using standard conditions[8]. Positive clones were subcloned into the *Eco*RI site of plasmid pUC9 (ref. 9). Partial restriction maps and the sequencing strategies[10] are shown in Fig. 1 for TM86 and the three thymus-derived clones 86T1, 86T3 and 86T5. The 3′ termini of the 86T series of clones are all identical, as each clone has an internal *Eco*RI site in that region, and the variation in position of the 5′ termini is presumably due to random chain termination during library construction.

As the largest mRNA detected on Northern blots with TM86 has a size of ~1,300 bases[1], by subtracting a poly(A) tail of

‡ Present address: Department of Biology, University of California at San Diego, La Jolla, California 92093, USA.

95

©1984 Nature Publishing Group

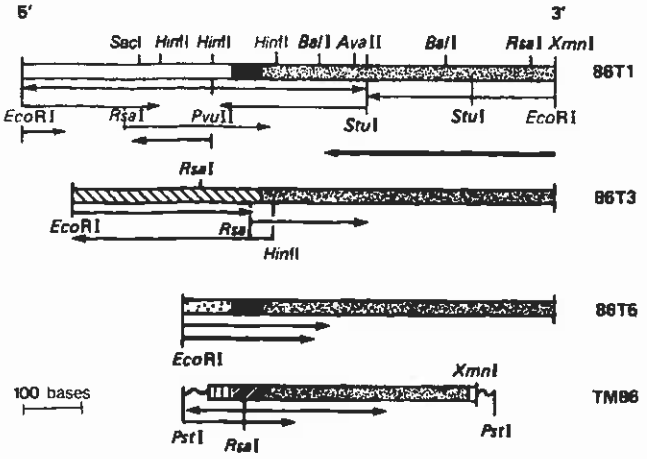


Fig. 1 cDNA clone sequencing strategy and regions of similarity. The TM86 cDNA clone was obtained as described in the accompanying article[1], and the insert excised by *PstI* from the plasmid vector. The insert was ^{32}P-labelled by nick-translation[8] and used to screen a λgt-10 cDNA library constructed from total thymocyte poly(A)$^+$ RNA from young, BALB/c mice (T. Huynh and R. W. Davis, unpublished protocol). Shaded areas indicate major similarities between the different cDNA clones. The 5' → 3' orientation was deduced from the large open reading frame of 86T1. Restriction sites indicated for 86T1 are shared within regions having the same shading on the cDNA clones. Sequencing was by the procedure of Maxam and Gilbert[10], with thick arrows representing 3'-end labelling and thin arrows 5'-end labelling.

150–250 bases we would expect a cDNA insert size of around 1,100 bases. Thus, the 938-base insert of clone 86T1 should contain most of the sequence of the mRNA it represents. The insert of 86T1 was completely sequenced and compared with partial sequences of the other clones (see Fig. 1). As implied by Fig. 1, the 3' end of TM86 appears to be different from those of the 86T series, which may have a bearing on thymic maturation and will be the subject of a separate report.

The nucleotide sequences and predicted amino acid sequences of the four different cDNA clones are given in Fig. 2. It can be seen that an open reading frame exists throughout the entire 86T1 insert, beginning with a methionine codon 22 nucleotides from the 5' end and extending for 306 amino acid residues through the 3' *Eco*RI site. The positions of cysteine residues and five potential carbohydrate attachment sites are indicated. Several interesting points are evident from these sequences: (1) Although all of the clones have identical 3' end sequences (except where noted above for TM86) the 5' end sequences are all very different. We have recently sequenced the entire 'constant' region of 86T3 (D. Becker and M.M.D., unpublished results) and it differs from the sequence of 86T1 by only a single nucleotide. This arrangement of constant and variable regions is similar to that observed with immunoglobulin cDNA clones. (2) Immediately following the methionine initiation codon there is a stretch of hydrophobic residues corresponding to the expected leader peptide (the sequence Leu-Leu-Leu is common in immunoglobulin κ light-chain leader peptides). (3) 86T1 and 86T5 share a 16-residue sequence between the constant and variable regions that is not found in the sequences of 86T3 and TM86, suggesting the existence of an independently assorting *J*-like region. (4) The positions of cysteine, and other, residues is similar to that of immunoglobulin and related molecules. (5) Clone 86T3 appears to be derived from a non-functional mRNA as it has stop codons in every reading frame.

Similarity to immunoglobulins

To identify possible evolutionary relationships between the sequence of cDNA clone 86T1 and other known sequences, we searched the Dayhoff protein sequence data bank using the rapid comparison programs of Wilbur and Lipman[11]. From the Dayhoff bank of ~2,300 sequences, 25 were found to have a similarity to the 86T1 sequence greater than, or equal to, five standard deviations from the mean. Of these 25 sequences, 24 were immunoglobulin sequences and one was of a class II human major histocompatibility molecule (HLA-DC1α). Furthermore, it was found that the apparent variable region of 86T1 matched with variable regions of immunoglobulins and the constant region with immunoglobulin constant regions (and in one case with the second domain of HLA-DC1α, which has previously been shown to have a sequence related to those of immunoglobulins[12]).

Figure 3 gives examples of sequences from the Dayhoff bank which match with parts of the 86T1 sequence. The closest match was found with a sequence just N-terminal of the third hypervariable region of the heavy chain of the monoclonal antibody 93G7 (ref. 13). This is the most highly conserved sequence of immunoglobulin variable regions, and surrounds the second cysteine at residue 92 which forms part of the characteristic intrachain disulphide bond of the variable regions domain[2].

Figure 4 shows a more detailed comparison between the 86T1 sequence and those of murine immunoglobulin variable and constant regions, and demonstrates the clear immunoglobulin-like character of most of the 86T1 sequence. The similarity of the variable region of 86T1 to three different immunoglobulin variable region sequences[13–15] is shown in Fig. 4*a*. The degree of similarity between 86T1 and the immunoglobulin variable regions is comparable to that between the variable regions of immunoglobulin heavy and light chains. Of the 18 invariant residues of murine κ variable regions and the 10 invariant residues of the heavy-chain variable regions, 13 and 6, respectively, are found in the 86T1 sequence. The outermost two cysteines of the 86T1 variable region are separated by a similar number of residues (68) to those separating the cysteines of immunoglobulin variable regions that form the intrachain disulphide bonds[2] (65 amino acids for light chains and 70 amino acids for heavy chains). The alignment in Fig. 4 also predicts that the putative leader peptide of 86T1 will be cleaved just before the asparagine (N) residue at position 20 (Fig. 2).

This similarity to immunoglobulins is also shown by the constant region of 86T1, particularly around the cysteine residue at position 164 (Fig. 4*b*). It is interesting to note that the region N-terminal of this cysteine is more similar to light chains and the C-terminal region is more similar to heavy chains. The similarity of the 86T1 sequence around the last cysteine residue (position 260) to both λ and κ light chains may be important as in light chains this is the cysteine that forms an interchain disulphide bond with the heavy chain. The similarity in sequence between 86T1 and immunoglobulins is so pronounced that it might at least in part explain the cross-reactivity of B-cell anti-idiotypic antisera with T cells[16–18]. In contrast to this amino acid sequence similarity, the nucleotide sequence of 86T1 has diverged sufficiently from those of immunoglobulin genes to explain the negative results of many investigators who have used immunoglobulin gene probes to search for the genes encoding the T-cell receptor[19–22].

Joining region-like elements

As indicated in Fig. 2, 16 amino acids are shared between 86T1 and 86T5 but not with the other two clones that we have sequenced, 86T3 and TM86. As this falls precisely into the region occupied by joining (*J*) region elements in immunoglobulins (between the variable and constant regions) we compared this region in 86T1 with the equivalent regions of 86T3 and TM86, as well as with the consensus sequences from each of the three different immunoglobulin *J* region elements[2] (Fig. 5). As shown in Fig. 5, the putative *J* regions of 86T1 and TM86 both have very similar sequences to those of all the immunoglobulin *J* regions, with no need for gaps to be introduced. As with the variable region similarities, the T-cell *J*

©1984 Nature Publishing Group

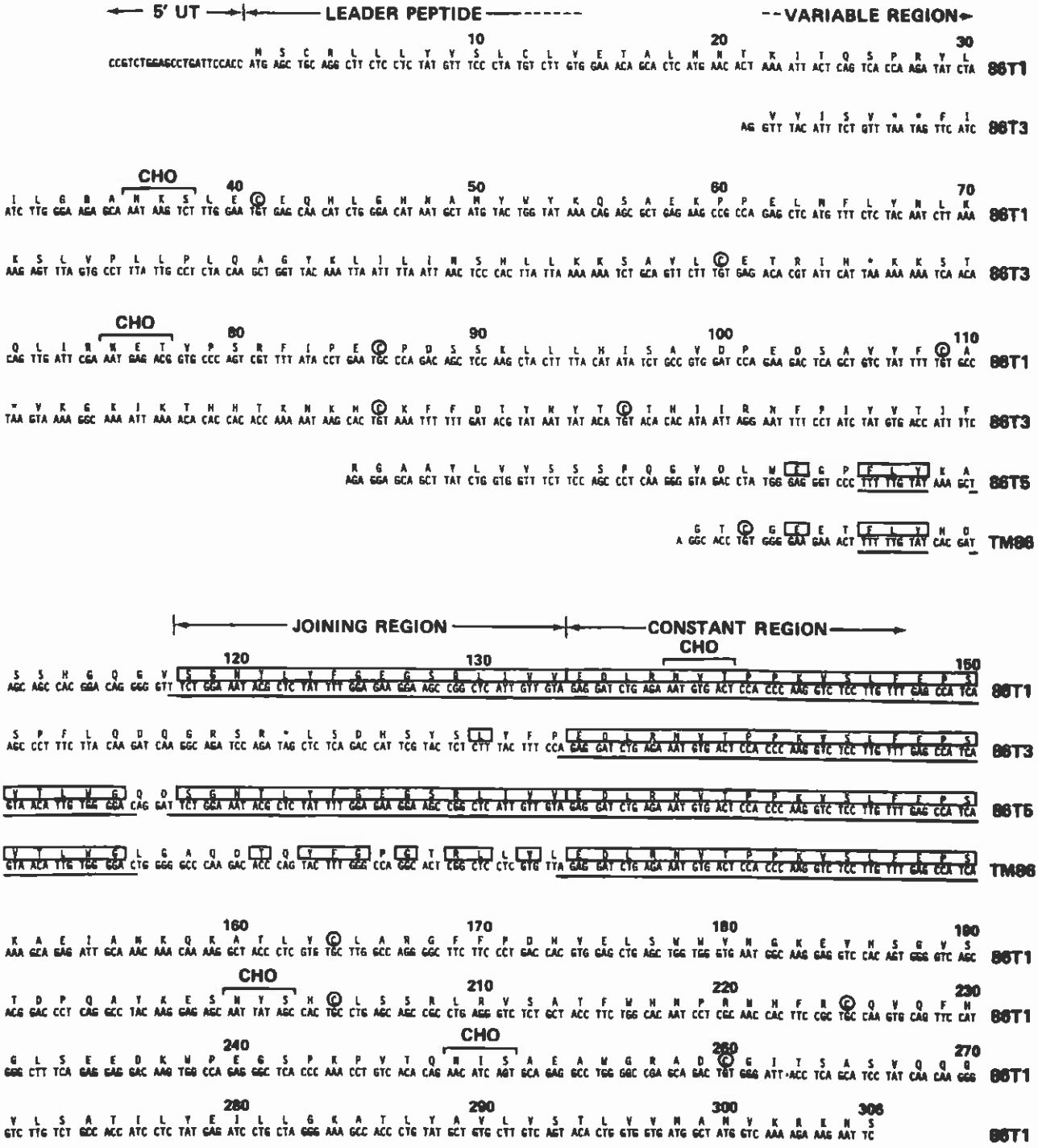


Fig. 2 Sequence of TM86 and related cDNA clones. This figure shows the complete nucleotide sequence of 86T1 and partial sequences of the other cDNA clones, indicating the 5′ untranslated region (UT), the putative leader polypeptide, and the variable, joining and constant regions. Amino acid similarities are boxed, and nucleotide sequence similarities are underlined. The numbering follows the predicted amino acid sequence of 86T1. Cysteine residues are circled and possible carbohydrate (CHO) attachment sites are indicated (following the formula: N–X–S or N–X–T)[31].

regions show an approximately equal degree of similarity to each of the different immunoglobulin *J* region types, similar to the relationship between the heavy-chain and light-chain *J* regions, but not as pronounced as that between light-chain *J* regions. In terms of size (16 amino acids) they are more closely related to heavy-chain *J* regions (17 amino acids) than light-chain *J* regions (13 amino acids) but since the putative receptor chains are approximately equal in size[23–25], the same constraints may not apply.

In addition to the *J* elements, the adjacent 5′ regions of 86T5 and TM86 between residues 103 and 115 (Fig. 2) are very similar. In particular, the 17-nucleotide and 9-nucleotide identities between these two cDNA clones suggest the existence of other possible 'mini-gene'[26] elements possibly analogous to the

©1984 Nature Publishing Group

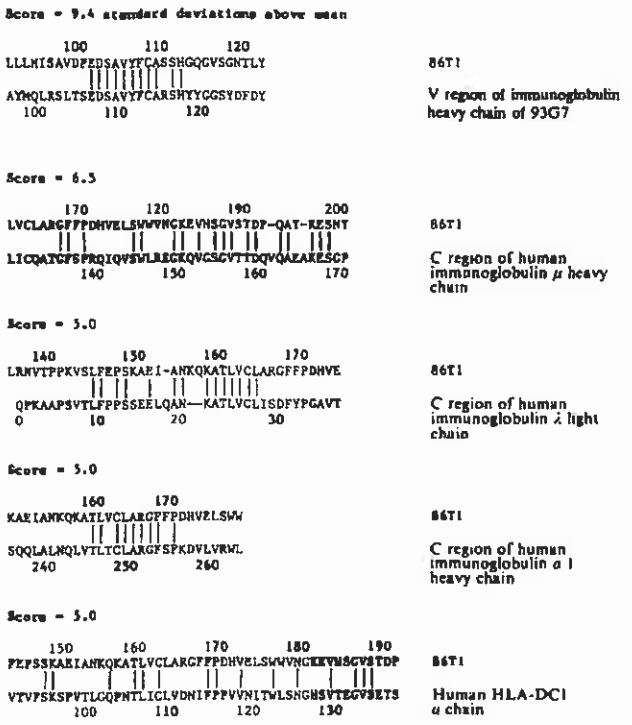


Fig. 3 Search of the Dayhoff sequence data bank for sequences closely related to that of 86T1 using the program of Wilbur and Lippman[11]. Representative similarities are shown as they appear from the search with the matches indicated (the computer program focuses on short stretches of homology and thus all of the possible matches are not shown). Gaps are represented as dashes (–).

D region of heavy-chain immunoglobulins[6,27,28]. An alternative explanation would be that these similarities represent highly conserved sequences of related variable-region genes.

Structural predictions

A number of groups have devised formulae to predict from amino acid sequences which portions of a molecule are embedded in membranes and which are exposed. These predictions are based on weighting formulae derived partly from measurements of solubility in organic and aqueous solvents and partly from empirical considerations. One such formula is that of Kyte and Doolittle[29] which is particularly suited for cell-surface molecules as most of the now considerable literature on the structure of such molecules has been taken into account. This 'hydropathicity' plot is shown in Fig. 6 and indicates a number of important features: first, that the 86T1 protein has the alternating hydrophobic–hydrophilic stretches characteristic of globular proteins[29]; second, that the predicted leader polypeptide occurs in a very adequate hydrophobic environment; and most importantly, it predicts a possible transmembrane region spanning the end of the 86T1 sequence, followed by a string of positive charges (Lys-Arg-Lys) which are characteristic of the cytoplasmic portion of a number of lymphocyte cell-surface markers[2]. The major difficulty with this prediction is the lysine residue (K) at position 284, in the middle of the putative transmembrane region, which is not characteristic, but may be masked by a salt bridge to another protein (possibly the other chain of the heterodimer).

Discussion

For the following reasons the data presented in this and the accompanying paper[1] indicate the considerable likelihood that the mRNAs related to the TM86 cDNA clone encode one chain of the T-cell antigen receptor.

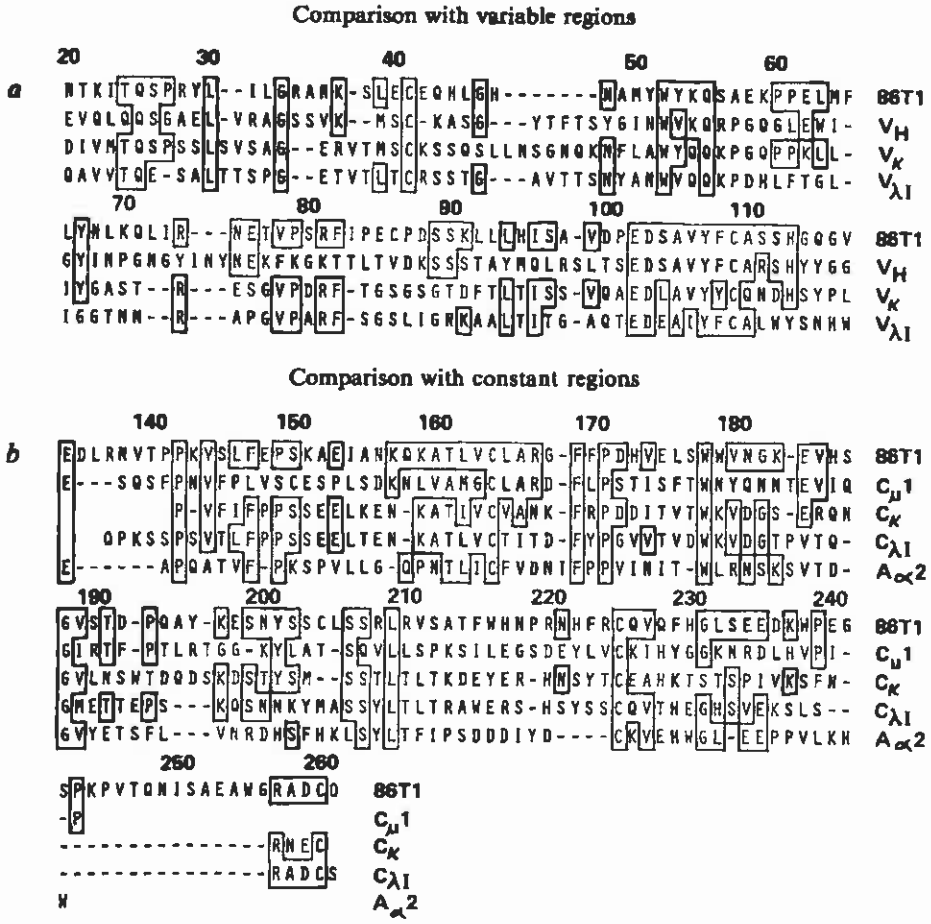


Fig. 4 Similarities of the 86T1 amino acid sequence to variable (*a*) and constant (*b*) regions of murine immunoglobulins and to the second domain of the murine Ia–A^d_{α} (ref. 32) (the murine equivalent of the human HLA chain DC_{α}). Similarities are boxed with gaps represented as dashes (–). *a* Shows a comparison of the variable-region immunoglobulin sequences with the 5′ region of 86T1; *b* shows a comparison of the 3′ region of 86T1 with immunoglobulin constant-region sequences and a region of the sequences of A^d_{α}. The 9367′CL sequence also includes part of the *D* (diversity) region. The V_{κ} is from MOPC 603[14] and the V_{λ} sequence from MOPC 104e[15].

©1984 Nature Publishing Group

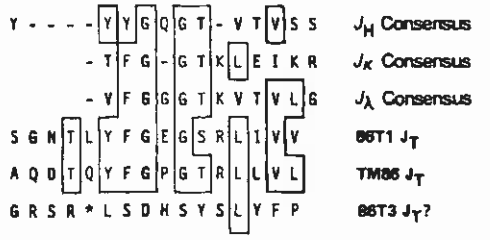


Fig. 5 Joining region sequence similarities between immunoglobulin consensus *J* sequences[2] and the apparent T-cell *J* regions. Similarities are boxed and dashes (–) indicate a non-conserved residue between the *J* region elements of a particular immunoglobulin.

The first reason is the consistent rearrangements seen in Southern blot analyses of T-cell hybridomas and lymphomas, and the fact that these appear to vary in T cells of different specificities[1]. Second, the analysis presented here shows the very close similarity of the amino acid sequences predicted by our cDNA sequences to immunoglobulin variable, constant and joining regions (Figs 3–5). This similarity is much closer than that of the postulated 'super-family' of immunoglobulin-like molecules which includes a wide variety of very distantly related cell-surface glycoproteins[30], as judged by the outcome of the computer search (Fig. 3) and the overall similarities presented in Figs 4 and 5. Third, the finding of sequence differences and similarities between the set of cDNA clones illustrated in Fig. 2 and that these can be divided into 5' variable, 3' constant and intermediate joining regions of approximately the same size as those of immunoglobulins.

Given these considerations, we conclude that the locus described here represents a type of immunoglobulin gene specifically rearranged and expressed in at least some subsets of T lymphocytes, and that it must almost certainly play a part in antigen recognition by T cells. Supporting this later contention is recent work which indicates that antisera raised against synthetic peptide fragments of 86T1 can significantly inhibit the antigen-dependent release of interleukin 2 by T helper hybridomas (J. Rothbard, S.M.H., R.H. Schwartz and M.M.D., manuscript in preparation.

On the basis of the data that we have presented we suggest that the structure of the 86T1 gene is that shown in Fig. 7, with a 19-amino acid leader polypeptide, a 98-amino acid variable region, a 16-amino acid *J* region and a single globular constant-region domain followed by transmembrane and cytoplasmic portions. By analogy to immunoglobulin, the two outermost cysteines in each globular domain would be linked (as shown in the figure) and the last cysteine at position 260 would be bound to the other chain of the receptor heterodimer. It is not clear why there is an extra cysteine within each domain, but it should be noted that rabbit light-chain immunoglobulins also have an additional extra cysteine within each domain (and at the same position in the constant region); this may confer additional stability on the native molecule, perhaps through inter-chain bonding. The protein predicted from the 86T1 nucleotide sequence has a molecular weight of 34,000 (34K), and the processed molecule would be 32K, plus whatever additional coding sequence remains in the full-sized mRNA. Given the many glycosylation sites this is thus far compatible in size with either of the two 40–48K putative T-cell receptor polypeptide chains seen in the mouse[24,25].

Another interesting finding is the relative lack of immunoglobulin variable region homology found in the 86T5 and TM86 fragments shown in Fig. 2. 86T5, in particular, has no cysteine

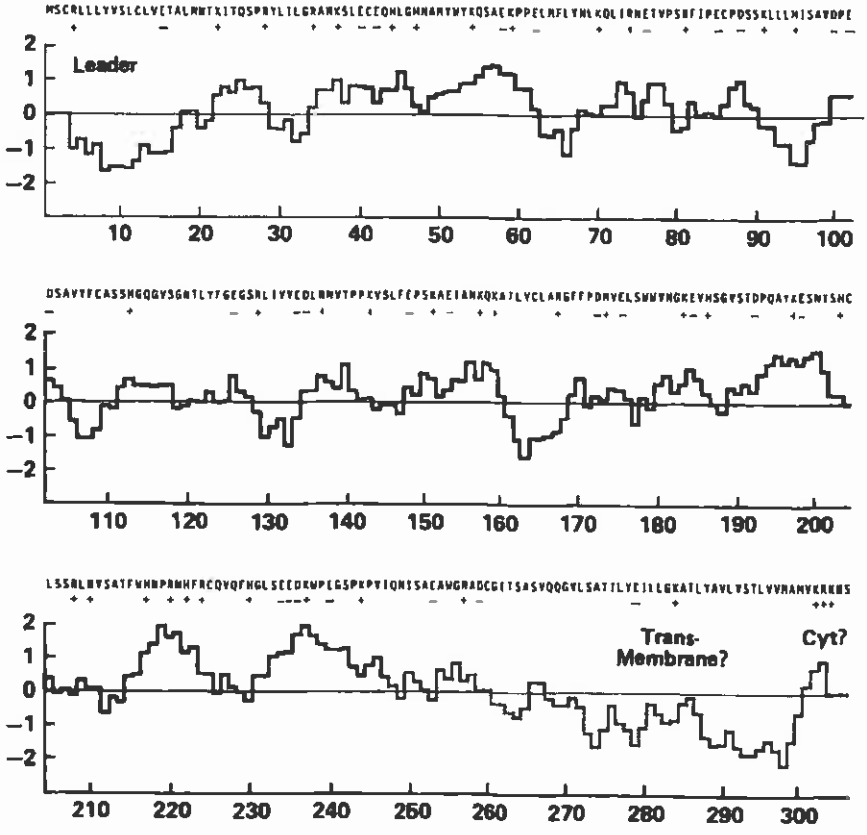


Fig. 6 Kyte–Doolittle 'hydrophathicity' plot. The weighting system of Kyte and Doolittle[29] was used to predict membrane-embedded and hydrophilic stretches in the sequence of 86T1: negative values indicate the former and positive values the latter (see text). The numbers 1 and 2 in the figure represent the numbers 10 and 20 used in ref. 29.

©1984 Nature Publishing Group

99

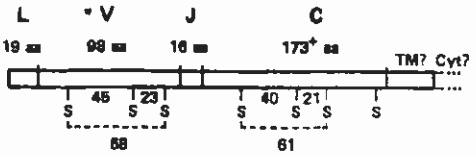


Fig. 7 Suggested structure of the 86T1 gene with leader (L), variable (V), joining (J) and constant (C) regions indicated together with the length in amino acids (aa). Also illustrated are the positions of sulphydryl side chains of the cysteine residues (S) and possible intrachain bonding, based on the immunoglobulin analogy. The transmembrane (TM) and cytoplasmic (Cyt) portions of the molecules, based on the analysis in Fig. 6, are also shown.

residue even remotely near the one at position 109 of 86T1, which is so highly conserved in all immmunoglobulin variable regions. TM86 has a cysteine nearby at position 101, but appears to lack some of the surrounding conserved sequences. This may be the result of either a large number of non-functional transcripts (particularly in the thymus) or a wide variation in the structural requirements for a binding site.

We thank W. E. Paul for encouragement and support; C. Benoist for providing the thymocyte cDNA library; and D. Lipman for assistance in searching the Dayhoff data bank. J. Rothbards' assistance in the protein analysis was invaluable. We also thank R. Schellar, J. Rothbard and H. McDevitt for critical reading of the manuscript. D.C. is a fellow of the Leukemia Society of America. This work was financed, in part, by grants to M.M.D. from the NIH (1R01 AI20320, 1P0L AI19512).

Note added in proof: Recent sequence analysis (Y. Chien, J.K. and M.M.D., work in progress) of a cDNA clone which extends further at the 3′ end than 86T1 indicates the presence of a stop codon immediately after the equivalent of residue 306 in Fig. 2. Therefore, the predicted amino acid sequence of 86T1 is most likely to be of the complete molecule.

Received 25 January; accepted 31 January 1984.

1. Hedrick, S. M., Cohen, D. I., Nielson, E. A. & Davis, M. M. *Nature* **308**, 149–153 (1984).
2. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. *Sequences of Immunological Interest* (US Dept of Health and Human Services, Washington, DC, 1983).
3. Dreyer, W. T. & Bennett, J. C. *Proc. natn. Acad. Sci. U.S.A.* **54**, 864–868 (1965).
4. Brack, C., Hirama, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* **15**, 1–14 (1978).
5. Seidman, J. G. & Leder, P. *Nature* **276**, 790–795 (1978).
6. Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. *Cell* **19**, 981–992 (1980).
7. Sakano, H. *et al. Nature* **277**, 627–635 (1979).
8. Maniatis, T., Fritsch, E. & Sambrook, J. in *Molecular cloning* (Cold Spring Harbor Laboratory, New York, 1982).
9. Viera, J. & Messing, J. *Gene* **19**, 259–268 (1982).
10. Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
11. Wilbur, N. J. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 726–730 (1983).
12. Auffray, C., Korman, A. J., Roux-Desseto, M., Bono, R. & Strominger, J. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6337–6341 (1982).
13. Estess, P., Nisonoff, A. & Capra, J. D. *Molec. Immun.* **16**, 1111–1118 (1979).
14. Hood, L., McKean, D., Farnsworth, V. & Potter, M. *Biochemistry* **12**, 741–749 (1973).
15. Appella, E. *Proc. natn. Acad. Sci. U.S.A.* **68**, 590–594 (1971).
16. Eichmann, K. & Rajewsky, K. *Eur. J. Immun.* **5**, 661–666 (1975).
17. Binz, H. & Wigzell, H. *J. exp Med.* **142**, 197–211 (1975).
18. Augustin, A. A., Julius, M. H., Cosenza, H. & Matsunaga, T. in *Regulatory T Lymphocytes* (eds Pernis, B. & Vogel, H. J.) (Academic, New York, 1980).
19. Kronenberg, M., Davis, M. M., Early, P. W., Hood, L. E. & Watson, J. D. *J. exp. Med.* **152**, 1745–1761 (1980).
20. Nakanishi, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 6984–6988 (1982).
21. Kurosawa, Y. H. *et al. Nature* **290**, 565–570 (1981).
22. Kronenberg, M. *et al. J. exp. Med.* **158**, 210–227 (1983).
23. Acuto, O. *et al. Cell* **34**, 717–726 (1983).
24. McIntyre, B. W. & Allison, J. P. *Cell* **34**, 739–746 (1983).
25. Kappler, J., Kubo, R., Haskins, K., White, J. & Marrack, P. *Cell* **34**, 727–737 (1983).
26. Kabat, E. A., Wu, T. T. & Bilofsky, H. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2429–2433 (1978).
27. Sakano, H., Kurosawa, Y., Weigert, M. & Tonegawa, S. *Nature* **290**, 562–579 (1981).
28. Kurosawa, Y. & Tonegawa, S. *J. exp. Med.* **155**, 201–218 (1982).
29. Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105–132 (1982).
30. Larhammar, D. *et al. Cell* **30**, 153–161 (1982).
31. Wagh, P. V. & Bahl, O. P. *Crit. Rev. Biochem.* **11**, 307–377 (1981).
32. Benoist, C. O. *et al. Cell* **34**, 169–177 (1983).

LETTERS TO NATURE

Pioneer 10 search for gravitational waves—no evidence for coherent radiation from Geminga

J. D. Anderson, J. W. Armstrong, F. B. Estabrook, R. W. Hellings, E. K. Lau & H. D. Wahlquist

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

It has been reported[1] that an observed 160-min solar oscillation may be caused by gravitational radiation from the intense γ-ray source Geminga, a source whose period according to the report is exactly one cycle per year different from the solar oscillation period. This remarkable coincidence has motivated us to look for a sinusoidal component having a period near 160 min in Doppler data from the 1981 gravitational wave search with Pioneer 10. We find no such component in our most quiet data. Thus, the amplitude of a possible sinusoidal variation is no more than 3.3×10^{-14} in fractional Doppler frequency, which for the direction of Geminga translates into an upper bound of 2×10^{-14} for the polarization component of gravitational spatial strain potentially observable. This places rather stringent limits on the amount of gravitational radiation available to excite the 160-min normal mode of the Sun.

We have examined Doppler data taken by the NASA/JPL Deep Space Network with Pioneer 10 over a 3-week period from 14 November to 8 December 1981 during solar opposition; at such times low-frequency phase fluctuation caused by scattering by free electrons in the interplanetary plasma is at a minimum. The data represent part of the record from gravitational wave searches with both Pioneer 10 and Pioneer 11 during their extended Heliospheric Mission. Data are also available from the 1982 opposition of Pioneer 10, as well as from two oppositions of Pioneer 11 for that matter, but these data have not been reduced to a point where they are useful for a search for the specific gravitational wave signal of interest here. The standard deviation of the Doppler noise from the best 1981 Pioneer 10 data is $\sim2\times10^{-13}$ in fractional frequency $(\Delta\nu/\nu)$ over a bandwidth of 4×10^{-5} to $\sim2\times10^{-3}$ Hz. The low-frequency cutoff is set by the round-trip light-time (or 'storage' time) of the microwave link while the system noise limits the high-frequency region[2–4].

The long light time l to Pioneer 10, now beyond the orbit of Neptune, is a definite advantage in searching for long-period coherent gravitational waves. During the 1981 opposition $l\sim12{,}490$ s and the ratio $4\pi l/P\sim16.4$ rad, where P is the period of the hypothesized gravitational wave. This ratio enters in a fairly complicated way with the cosine of the separation angle on the sky between Pioneer 10 and Geminga, $\mu=\cos\theta$, in determining the relation between the strain amplitude h of the gravitational wave and the amplitude of the fractional Doppler response $\Delta\nu/\nu$ at the same frequency. The right ascension and declination of Pioneer 10 on 28 November 1981 was 4 h 4.05 min and 24.097° respectively, while Geminga has a position of 6 h 31 min and 17°48′, both positions given with respect to the mean equator and equinox of 1950.0. This yields $\mu=0.82$. From the 'three-pulse' response of the Doppler system[5], we conclude by summing the three phasors of the coherent Geminga signal that the observable polarization component of strain amplitude of a gravitational wave at Earth is amplified by a factor of 1.63 in the Doppler data.

Two approaches were used to search for a periodic signal from Geminga in the reduced Doppler record. In the first approach, three selected long passes of Doppler data of ~12 h each were transformed into the frequency domain by a fast Fourier transform operating on 512 data, sampled at 100-s intervals. The three spectra were averaged with equal weight given to each; the resulting spectrum is shown in Fig. 1. No

©1984 Nature Publishing Group

T cell receptor for antigen (TCR)

The complex formed by MHC class I molecule and the nine amino acid residue processed peptide or by MHC class II molecule and the roughly 15 amino acid residue processed peptide will interact with molecules on the surface of T lymphocytes known as T cell receptor for antigens (TCR's). These molecules, on one hand, resemble MHC class II molecules with two nearly equal size membrane bound chains designated as α / β. In some cases, the two chains are designated as γ / δ. These two pairs have different properties to be discussed in detail later. On the other hand, TCR's resemble the Fab fragment of antibodies. The variable regions of α and γ are assembled from V genes and J minigenes similar to antibody light chains, while those of β and δ from V genes, D minigenes and J minigenes similar to antibody heavy chains.

The search for the TCR genes took place in two separate laboratories of T. K. Mak and M. M. Davis, both using the now famous subtraction method by comparing the differences between T and B lymphocytes. Three papers were published in the 1984 issue of Nature:

Yanagi Y, Yoshikai Y, Leggett K, Clark SP, Aleksander I, Mak TW (1984) A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. Nature, 308, 145-149.

Hedrick SM, Cohen DI, Nielsen EA, Davis MM (1984) Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature, 308, 149-154.

Hedrick SM, Cohen DI, Nielsen EA, Davis MM (1984) Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. Nature, 308, 154-158.

By this method, the TCR β chain genes were isolated initially. Since morphologically T lymphocytes and B lymphocytes are very similar, most of their mRNA molecules and thus cDNA should also be similar. The major differences should thus reside in the different genes for TCR's and for immunoglobulins. Subsequently, the TCR α chain genes were also isolated. When these isolated genes were translated to amino acid sequences, they have the following domains:

TCR α chain:	V_α	C_α	α-membrane	α-cytoplasmic
TCR β chain:	V_β	C_β	β-membrane	β-cytoplasmic

Their cytoplasmic segments are relatively short, especially that of the α chain.

Similar to antibodies, the variable and constant regions have the immunoglobulin fold structure. Their variable regions also have three CDR's from both the V_α and the V_β domains, forming the TCR binding site to interact with the MHC-processed peptide complex. TCR's do not seem to differentiate MHC class I and MHC class II molecules.

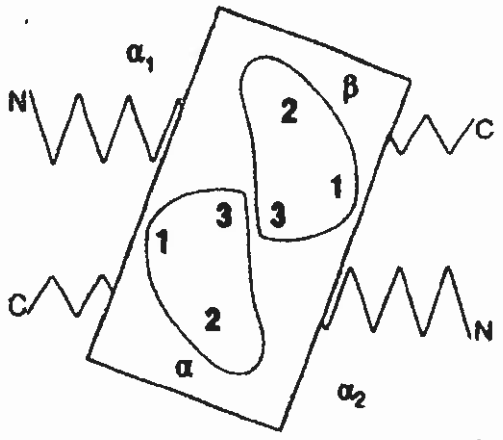
While antibodies bind to antigens with highly varied structures and contours, TCR's bind to MHC-processed peptide complexes with very restricted contours. Therefore, antibody CDRH3 segments of antibodies can have different lengths as well as amino acid compositions. To bind relatively fixed contours, the combined lengths of CDRα3 and CDRβ3 should have a fairly narrow distribution. This is achieved by random assortment of CDRα3 and CDRβ3 with even narrower length distributions (Johnson G, Wu TT (1999) Random length assortment of human and mouse T cell receptor for antigen α and β CDR3. Immunol Cell Biol, 77, 391-394).

The physical arrangement of the CDR's of α / β in the three dimensional structure of TCR is essentially identical to the CDR's of the antibody. Two laboratories managed to co-crystallize the extra-cellular portion of the MHC class I molecule, the processed peptide and the extra-cellular portion of the TCR and determined the detailed three dimensional structure of that complex:

Garcia KC, Denago M, Stanfield RL, Brunmark A, Jackson MR, Peterson PA, Teyton L, Wilson IA (1996) An αβ T cell receptor structure at 2.5 A and the orientation in the TCR-MHC complex. Science, 274, 209-219.

Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC (1996) Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. Nature 384, 134-141.

A schematic illustration of the area of contact between these molecules is illustrated below:



Immunology and Cell Biology (1999) 77, 391–394

Research Article

Random length assortment of human and mouse T cell receptor for antigen α and β chain CDR3

GEORGE JOHNSON and TAI T WU

Department of Biochemistry, Molecular Biology and Cell Biology and Department of Biomedical Engineering, Northwestern University, Evanston, Illinois, USA

Summary In view of the recently determined three-dimensional structures of complexes formed by the T cell receptor for antigen (TCR), the processed peptide and the MHC class I molecule, it is expected that the combined configuration formed by the third complementarity determining regions (CDR3) of TCR α and β chains will be very restricted in size and shape due to the limited length variations of the processed peptides. Thus, the combined TCR α and β chain CDR3 lengths should have a fairly narrow distribution. This feature can be due to the selective association of long α chain CDR3 with short β chain CDR3 and vice versa or due to random assortment of α and β chain CDR3 of even narrower length distribution. Based on existing translated amino acid sequence data, it has been found that the latter mechanism is responsible.

Key words: α chain, β chain, random assortment, T cell receptor for antigen, third complementarity determining region.

Introduction

Although TCR α and β chains exhibit many features similar to those of antibody light and heavy chains, there are a number of subtle differences. We have previously noted that TCR chains are under completely different and very restrictive evolutionary pressures and their functional V genes can be divided into two groups: similar and quite diverse.[1] Furthermore, based on recently determined three-dimensional structures formed by TCR, processed peptide and MHC class I molecule,[2,3] TCR combining sites are very restricted in size and shape due to the very limited length variations of the processed peptides inside the grooves of MHC molecules. Physically, TCR α and β chain CDR3 are in contact with each other. The CDR3 of α chain primarily binds residues 4 and 5 of the processed peptide, while that of the β chain binds residues 5–8.[3] They are located in the centre of the TCR-binding site, tightly bracketed by the other CDR, similar to those of antibodies.

In order to understand how the combination of CDR3 of TCR α and β chains can give rise to such limited configurations, we have aligned all available TCR amino acid sequences and plotted their CDR3 length distributions both separately and in combination with each other when both CDR3 have been determined from the same TCR molecule.

There are two possible mechanisms of generating narrow length distributions of combined CDR3 of TCR α and β chains. One involves selective association of long TCR α chain CDR3 with short TCR β chain CDR3 and vice versa. The other requires even narrower length distributions of TCR α and β chain CDR3 separately and they can be randomly associated with each other. In the present study, our result unequivocally supports the latter mechanism.

Materials and Methods

Aligned amino acid sequences of human and mouse TCR α and β chains can be found in the Kabat database Web server[4] at URL: <http://immuno.bme.nwu.edu>, which has been maintained in our laboratory. In this database, CDR3 are defined as segments starting from two amino acid residues away from the second cysteine and ending at the residue one position away from the glycine-X- glycine, the same as for antibody heavy chains.[5]

Complete and distinct CDR3 sequences are listed and counted. For identical sequences, only one is included. If both TCR α and β chains have been determined for the same TCR molecule, the sums of their CDR3 are calculated.

The numbers of different CDR3 are then plotted against their lengths for TCR α and β chains, separately and combined for both CDR3 sequences if known. To calculate the theoretical length distribution of combined α and β CDR3 lengths based on random assortment, frequencies of each separate combination of lengths are multiplied and summed. This sum is then normalized with the total number of cases where both TCR α and β chain CDR3 are known. Let a_i be the number of TCR α chain CDR3 of length i and b_i that for the β chain. The total numbers of TCR α and β chain CDR3 in the length range of interest in the present study, that is, 0–32 amino acid residues, are denoted by A and B:

$$A = \sum_{i=0}^{32} a_i;\ B = \sum_{i=0}^{32} b_i \qquad [1]$$

Let s_i be the number of α and β chain CDR3 from the same TCR molecule with a sum of length of i, and

$$S = \sum_{i=0}^{32} s_i \qquad [2]$$

Correspondence: Tai T Wu, Room E267, Technological Institute, 2145 Sheridan Road, Evanston, IL 60208, USA.
Email: <tt@immuno.bme.nwu.edu>

Received 31 March 1999; accepted 18 May 1999.

Then, t_i, the theoretical number of a sum of length of i, is calculated from:

$$t_i = \sum_{j=0}^{32} (a_j/a) \times (b_{i-j}/B) \quad [3]$$

for non-negative subscripts.

Results

The number versus CDR3 length plots for human and mouse are shown in Figs 1 and 2, respectively. In each figure, (a) gives the plot for TCR α chains determined experimentally, (b) the plot for β chains, (c) the sum of experimentally determined α and β chain CDR3 lengths of the same TCR molecules and (d) the theoretical sum derived from plots (a) and (b), assuming random assortment and compared with plot (c).

For human TCR α chain CDR3, the most common length is approximately eight or nine amino acid residues. Most are in the range of 6–13 residues, as illustrated in Fig. 1a. For β chain CDR3, the peak occurs at approximately 9–10 residues, while most are between eight and 15 residues (Fig. 1b). Experimental data for both α and β chain CDR3 are much lower (Fig. 1c). The sums of their lengths have a peak at approximately 19–20 residues, with a range of 14–26 residues and one example at 30 residues. The theoretical length distribution of the sum of α and β chain CDR3, based on random assortment, is superimposable on the experimental distribution (Fig. 1d).

The length distributions for mouse TCR α and β chain CDR3, both separated and combined (Fig. 2), are similar to those of humans. The peak length for mouse TCR α chain CDR3 is approximately 9–10 amino acid residues (Fig. 2a) and is the same for β chain CDR3 (Fig. 2b). The length of their sums peaks at 19 residues (Fig. 2c). Their ranges are approximately 7–14, 7–14 and 15–23 residues, respectively. Again, the theoretical distribution for their sums, based on random assortment, is similar to that found experimentally (Fig. 2d), with identical peak location but slightly broader range.

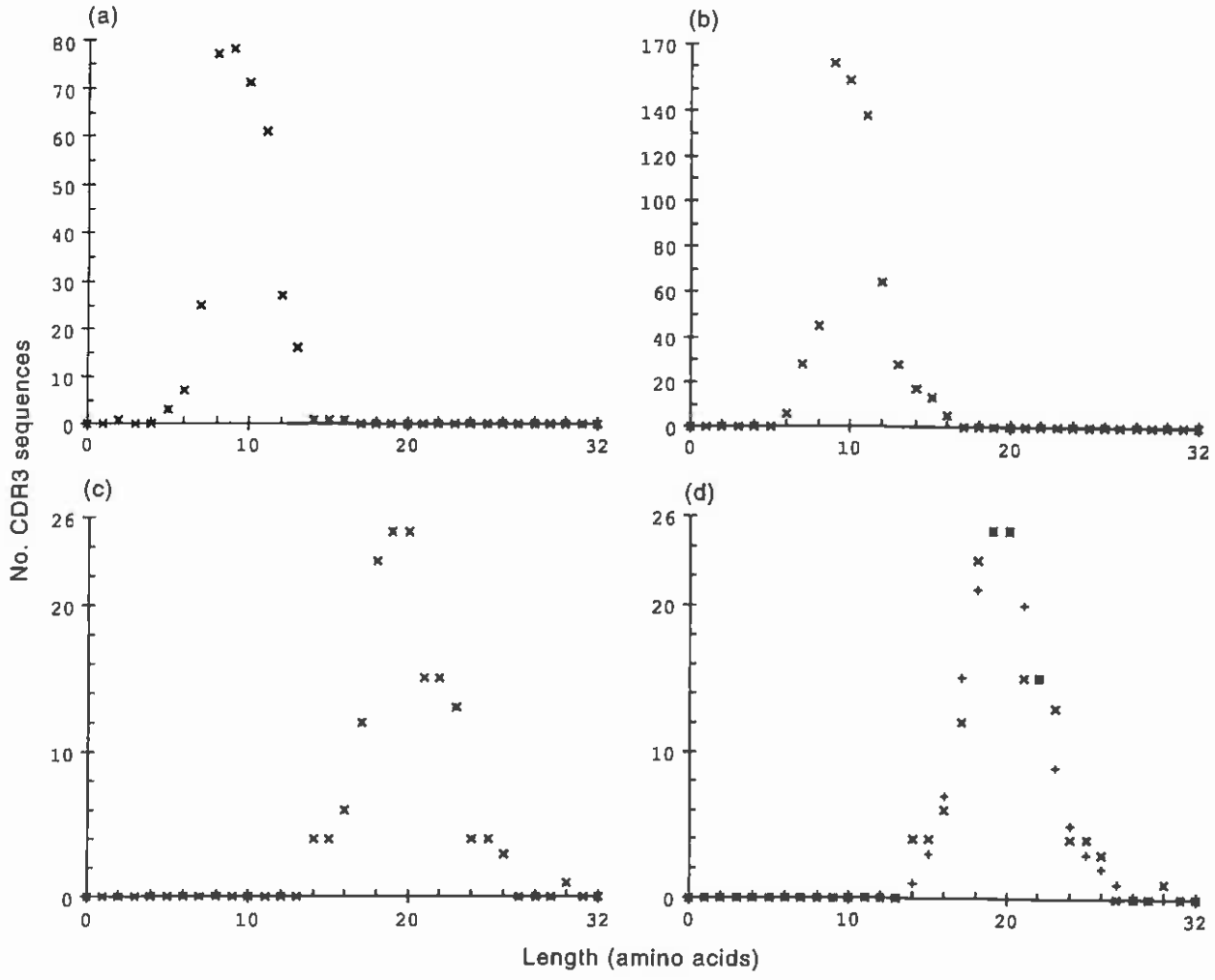


Figure 1 Length versus CDR3 length plots for human TCR (a) α chains, (b) β chains, (c) the sum of experimentally determined α and β chain CDR3 lengths of the same TCR molecules and (d) the theoretical sum derived from (a) and (b), assuming random assortment and compared with plot (c). (×), Experimental; (+), theoretical.

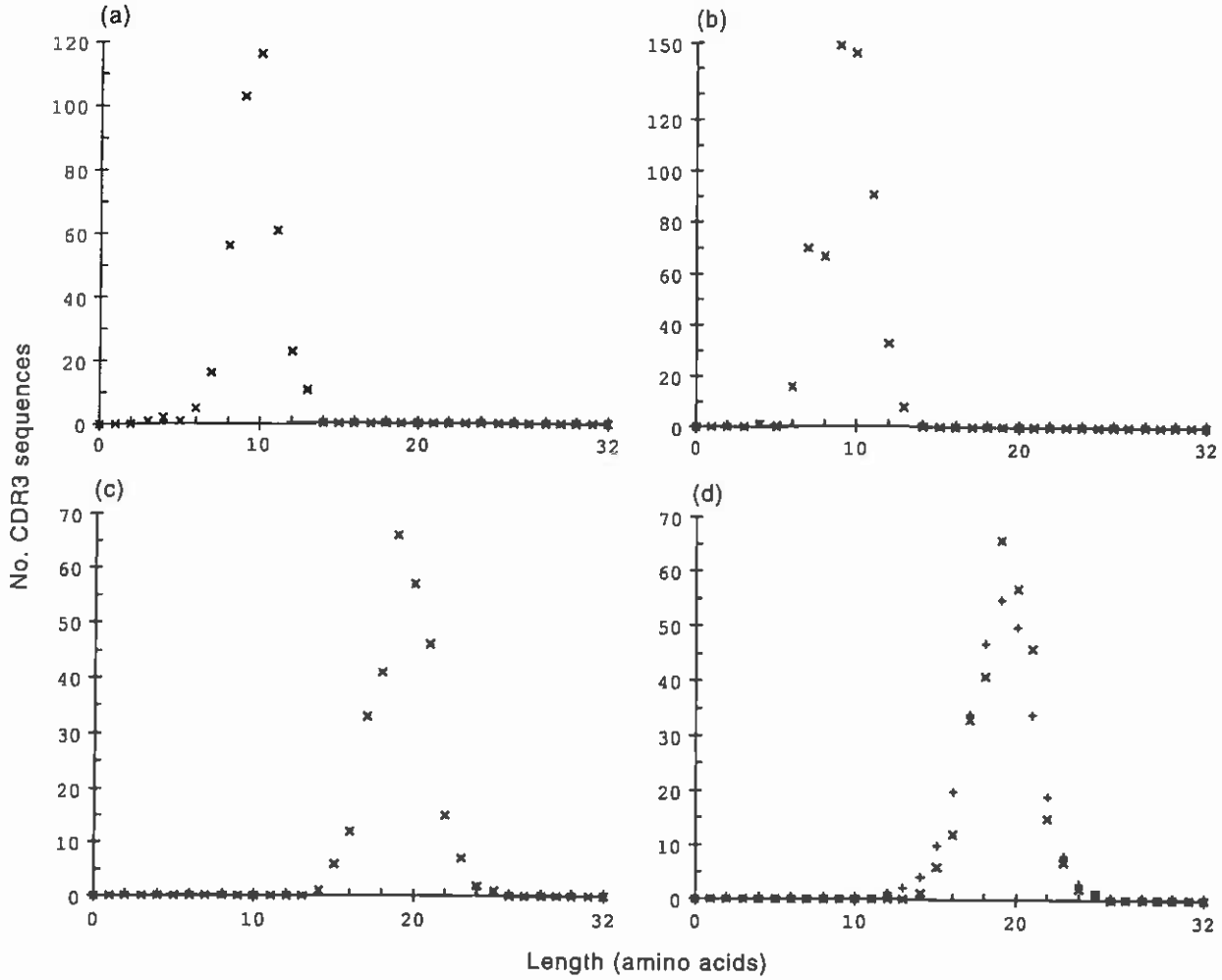


Figure 2 Length versus CDR3 length plots for mouse TCR (a) α chains, (b) β chains, (c) the sum of experimentally determined α and β chain CDR3 lengths of the same TCR molecules and (d) the theoretical sum derived from (a) and (b), assuming random assortment and compared with plot (c). (×), Experimental; (+), theoretical.

Discussion

Since the early days of sequencing TCR α and β chains,[6,7] it has been noticed that some portions of these sequences are similar to antibody light chains and other portions to heavy chains. During the past 30 years of collecting TCR and antibody sequences and storing them in a computer database,[4] we have noted that insertions and deletions in the various CDR3 regions are more extensive for antibody heavy chains[8] and TCR δ chains.[9] With the recent determination of three-dimensional structures of complexes of TCR, processed peptide and MHC class I molecules,[2,3] it has become clear that the sums of TCR α and β chain CDR3 lengths must also be very restricted, because the portions of the processed peptides in the MHC molecule grooves have little length variation.

In order to understand how this sum of lengths can have a narrow distribution, we proposed two possible alternative mechanisms. One requires the biological selective process of matching long TCR α chain CDR3 with short TCR β chain CDR3 and vice versa. In other words, if the combined length is beyond a certain range, either too short or too long, the combination would be non-functional biologically. Of course, this would also result in many wasted combinations of TCR α and β chain pairs. Another possibility is that individual TCR α and β chain CDR3 have even narrower length distributions.

In our present study, the near perfect superimposable length distributions for sums of TCR α and β chain CDR3 between experimentally determined pairs and random assortment of individual α and β chains for both human (Fig. 1) and mouse (Fig. 2) sequences suggests that the length distribution of CDR3 of each chain is narrower than the distribution of combined lengths. This type of length distribution is similar to that of antibody light chain CDR3.[4] Quantitatively, the means and standard deviations of the distributions shown in Figs 1,2 are summarized in Table 1. It is clearly shown that the theoretically predicted means for the sums of human and mouse TCR α and β chain CDR3 lengths, based on random assortment, are very close to the experimentally observed means. The predicted standard deviation is slightly smaller for human chains, but slightly larger for mouse chains.

105

Table 1 Means and standard deviations for human and mouse TCR α and β chain CDR3 length distributions in numbers of amino acid residues

	Mean	Standard deviation
Human α	9.51	1.78
Human β	10.23	1.78
Expt. sum of human α and β	19.82	2.71
Theoret. sum of human α and β	19.80	2.44
Mouse α	9.58	1.49
Mouse β	9.34	1.54
Expt. sum of mouse α and β	19.23	1.83
Theoret. sum of mouse α and β	18.91	2.15

CDR3, third complementarity determining region; Expt., experimental; Theoret., theoretical.

In summary, not only are the V genes of human and mouse TCR α and β chains under very restrictive evolutionary pressures,[1] but also their CDR3 are selected to have very limited lengths. For α chains, V-J joining can give rise to such distributions, as in the case of antibody light chains. However, for β chains, unlike antibody heavy chains or TCR δ chains, V-D-J rearrangements still result in a very restricted distribution, presumably due to the relatively uniform lengths of the few D minigenes and the scarcity of D-D joinings. Whether our present findings are also true for other species requires future sequence data for TCR α and β chains from these species.

Currently, the three-dimensional structure of TCR, processed peptide and MHC class II molecule has not yet been determined; however, it is expected to be similar to that with MHC class I molecules,[2,3] possibly with some minor variations.[4] Thus, what we have found in the present study about the very narrow length distributions of TCR α and β chain CDR3 and their random association to form TCR combining sites will most likely also be correct for TCR interacting with MHC class II molecules.

Acknowledgement

The present study was supported, in part, by NIH Grant 5 R24 AI25616-10.

References

1 Johnson G, Wu TT. A method of estimating the numbers of human and mouse T cell receptors for antigen α and β chain V-genes. *Immunol. Cell Biol.* 1997; **75**: 580–83.

2 Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* 1996; **384**: 134–41.

3 Garcia KC, Degano M, Stanfield RL *et al.* An αβ T cell receptor structure at 2.8 A and its orientation in the TCR-MHC complex. *Science* 1996; **274**: 209–19.

4 Johnson G, Kabat EA, Wu TT. Kabat database of proteins of immunological interest. In: Herzenberg LA, Weir WM, Herzenberg LA, Blackwell C (eds). *Weir's Handbook of Experimental Immunology. I. Immunochemistry and Molecular Immunology.* Cambridge: Blackwell Science, 1996; 6.1–21.

5 Johnson G, Wu TT. Preferred CDRH3 lengths for antibodies with defined specificities. *Int. Immunol.* 1998; **10**: 1801–5.

6 Yanagi Y, Yoshida Y, Legett K, Clark SP, Aleksander I, Mak TW. A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* 1984; **308**: 145–9.

7 Hedrick SM, Nielsen EA, Kavaler J, Cohen DI, Davis MM. Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. *Nature* 1984; **308**: 153–8.

8 Ikematsu H, Harindranath N, Ueki Y, Notkins AL, Casali P. Clonal analysis of a human antibody response. II. Sequences of the V_H genes of human IgM, IgG, and IgA to rabies virus reveals preferential utilization of V_HIII segments and somatic hypermutation. *J. Immunol.* 1993; **150**: 1325–37.

9 Ishiguro N, Aida Y, Shinagawa T, Shinagawa M. Molecular structure of cattle T-cell receptor γ and δ chains predominantly expressed on peripheral blood lymphocytes. *Immunogenetics* 1993; **38**: 437–43.

From *Science*, Vol. 274 (11 October 1996): 209–219 DOI: 10.1126/science.274.5285.209 Reprinted with permission from AAAS.

An αβ T Cell Receptor Structure at 2.5 Å and Its Orientation in the TCR-MHC Complex

K. Christopher Garcia, Massimo Degano, Robyn L. Stanfield, Anders Brunmark, Michael R. Jackson, Per A. Peterson, Luc Teyton, Ian A. Wilson*

The central event in the cellular immune response to invading microorganisms is the specific recognition of foreign peptides bound to major histocompatibility complex (MHC) molecules by the αβ T cell receptor (TCR). The x-ray structure of the complete extracellular fragment of a glycosylated αβ TCR was determined at 2.5 angstroms, and its orientation bound to a class I MHC–peptide (pMHC) complex was elucidated from crystals of the TCR-pMHC complex. The TCR resembles an antibody in the variable Vα and Vβ domains but deviates in the constant Cα domain and in the interdomain pairing of Cα with Cβ. Four of seven possible asparagine-linked glycosylation sites have ordered carbohydrate moieties, one of which lies in the Cα–Cβ interface. The TCR combining site is relatively flat except for a deep hydrophobic cavity between the hypervariable CDR3s (complementarity-determining regions) of the α and β chains. The 2C TCR covers the class I MHC H-2K^{b} binding groove so that the Vα CDRs 1 and 2 are positioned over the amino-terminal region of the bound dEV8 peptide, the Vβ chain CDRs 1 and 2 are over the carboxyl-terminal region of the peptide, and the Vα and Vβ CDR3s straddle the peptide between the helices around the central position of the peptide.

T lymphocytes respond to a wide variety of foreign antigens that are presented as peptides in the context of major histocompatibility molecules (MHC) (*1*). Specific recognition of peptide-MHC (pMHC) complexes is accomplished by a membrane-bound, multicomponent, cell surface glycoprotein termed the T cell receptor (TCR). The TCR complex consists of highly diverse, clonotypic αβ or γδ heterodimers and the γ, δ, ε, and ζ chains of the invariant accessory protein CD3 (*2*). The α and β chains participate in the interaction with the pMHC complex, whereas the CD3 chains participate in signal transduction. The genes encoding the TCR resemble immunoglobulin (Ig) genes not only in sequence, but in their assembly by somatic rearrangement of linked variable (V), diversity (D), joining (J), and constant (C) gene segments during lymphocyte development (*3*, *4*). The formation of functional α chain polypeptide requires the in-frame rearrangement of a V-region gene segment to a J-region gene segment, whereas functional β chain polypeptide is formed by two successive rearrangements of V-, D-, and J-region gene elements (*5*). The rearranged V-J and V-D-J regions are then attached to their respective C regions to assemble the mature α and β chain gene products. A vast number of potential protein sequences can result from these recombinations (*4*), as with Igs. Although not as striking as in Igs, four regions of hypervariable amino acid sequence are found on both the α and β chains, three of which are analogous to the antibody complementarity-determining regions (CDRs) (*6*), which serve as the primary contact points between antibody and antigen (*7*). Both CDRs 1 and 2 are encoded within the V genes; CDR3 occurs at the V-J junction in the α chain and at the V-D-J junction in the β chain (*4*).

The high degree of sequence identity between various V and C elements of TCRs and Igs (30 to 70 percent) suggested that TCR domains are folded into β-sheet sandwich structures (*8*, *9*), resembling Ig domains (*10*), that would pair in a manner similar to the heavy (H) and light (L) chains of antibodies. However, the recognition requirements of a TCR are more restricted than that of an antibody. Antibodies can bind ligands of extensive chemical and structural diversity, as reflected in the different shapes of antibody combining sites, from flat surfaces to deep grooves (*7*). The function of the TCR is to discriminate among different peptide antigens embedded in the largely flat, undulating surfaces of MHC molecules, whose dimensions and shape are relatively constant. Therefore, a more conserved binding site topology among different TCRs could be expected. Limited sequence diversity of the CDRs 1 and 2 suggested that most of the peptide specificity of the TCR would reside in CDR3, which is the most variable because of the junctional diversity of the V(D)J recombination (*4*). Site-directed mutagenesis studies confirmed that substitutions in CDR3 can either alter antigen specificity or abolish the response (*11*). Various models have been proposed for TCR recognition of pMHC in which the CDR3s of Vα and Vβ contact the peptide, whereas CDRs 1 and 2 interact primarily with the MHC α helices (*4*, *12*, *13*) or the ends of the peptide (*14*).

The x-ray structures of an individual TCR β chain from a T cell clone termed 14.3.d (*15*) and an isolated Vα fragment from a T cell clone termed 1934.4 (*16*) have confirmed that the TCR does indeed contain Ig-like domains. The Vα and Vβ domains resemble a v-type Ig fold (*17*), whereas Cβ more distantly resembles a c-type Ig fold (*17*). The monomeric β chain was proposed to be rather rigid because of extensive contacts between Vβ and Cβ, and a large protruding loop from Cβ that might limit "elbow" motion between Vβ and Cβ domains (*15*). Both CDRs 1 and 2 appear to be conformationally restricted by main-chain interactions with framework residues but, in the absence of the buttressing effect of their respective α and β chains, the CDR3s fold away from the domain surface toward the solvent.

The absence of a TCR αβ heterodimer crystal structure until now has been due to difficulties in producing large quantities of

K. C. Garcia, M. Degano, R. L. Stanfield, and I. A. Wilson are at the Department of Molecular Biology and the Skaggs Institute of Chemical Biology, Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA. A. Brunmark, M. R. Jackson, P. A. Peterson, and L. Teyton are at the R. W. Johnson Pharmaceutical Research Institute–La Jolla, 3535 General Atomic Court, San Diego, CA 92121, USA.

*To whom all correspondence should be addressed. E-mail: wilson@scripps.edu

107

From *Science*, Vol. 274 (11 October 1996): 209–219 DOI: 10.1126/science.274.5285.209 Reprinted with permission from AAAS.

soluble, homogeneous α and β chains that are correctly paired and folded. Recombinant techniques, although effective for the expression of antibodies, have not worked as well for TCRs. Prokaryotic and eukaryotic systems have been used to express TCRs (*18*), TCR-variable domain constructs (TCR-Fv) (*18*), and chimeras with Ig C domains (*19*), membrane-bound lipid linkages (*20*), the CD3 ζ chain (*21*), and leucine zippers (*22*), but with low yields. Even for correctly paired αβ TCRs, sufficient purity for obtaining diffraction quality crystals has not been achieved, even with deletion of asparagine-linked (N-linked) glycosylation sites (*23*). The conclusion is that chain pairing and secretion is less efficient (*24*), and refolding from *Escherichia coli* does not reliably yield functional TCRs.

TCR 2C ($V_\alpha 3\ J_\alpha 58\ C_\alpha$; $V_\beta 8.2\ D_\beta 2\ J_\beta 2.4\ C_\beta 2$), which was one of the first TCRs cloned (*3*, *25*), is a murine receptor of a cytotoxic T cell with specificity for MHC class I H-2K^b molecules (syngeneic), but cross-reactive (allogeneic) with H-2L^d in association with a self-peptide (p2Ca) derived from a mitochondrial protein (*25*). 2C has a relatively high affinity for the H-2L^d-p2Ca complex ($K_d \sim 0.5 \times 10^{-6}$M) (*26*), compared to H-2K^b-dEV8 ($K_d \sim 0.5 \times 10^{-5}$M) (*26*).

Structural analysis of the TCR 2C has been made possible by crystallization of a fully glycosylated form of 2C expressed in *Drosophila melanogaster* cells. The three-dimensional structure determination at 2.5 Å provides insights into the architecture, assembly, and folding of a TCR. In addition, crystallization of a complex of 2C and H-2K^b–dEV8 has allowed us to determine by molecular replacement the orientations of both the TCR and the pMHC in the TCR-pMHC complex.

TCR expression, crystallization, and structure determination. Expression of 2C from *D. melanogaster* cells (*27*) yielded a heterogeneous mixture of disulfide-bonded heterodimers and free monomers that, after purification, could be crystallized in bundles of needles that diffracted to low resolution. The COOH-terminal 10 to 15 residues were then digested away by successive treatment with carboxypeptidases A and B so that 2C could be purified on a hydrophobic interaction chromatography (HIC) column (*28*). The correctly paired disulfide-linked heterodimers (*28*) could now be crystallized and macro-seeded to a size sufficient for x-ray data collection. Interestingly, 2C expressed without N-linked glycosylation, by including the glycosylation inhibitor tunicamycin in the growth media, is soluble and binds to pMHC and all conformation-specific mAbs available (*28*), but it did not produce crystals (*29*).

The crystal structure of 2C was determined by means of a combination of molecular replacement (MR) and multiple isomorphous replacement (MIR). An initial room temperature data set to 2.9 Å was collected from a monoclinic crystal (*30*), space group $P2_1$ (Table 1). Solutions (MR) were obtained for the β chain and Vα (Table 1) but could not be determined for Cα even with a wide range of Ig-like do-

Table 1. Data collection, heavy atom, and refinement statistics. Monoclinic 2C crystals were grown in sitting drops from 2C (15 mg/ml), 1.8 M ammonium sulfate, 50 mM sodium acetate, and 50 mM sodium citrate (pH 6.4) at 22°C. The crystals were $P2_1$, with unit cell parameters a = 136.9 Å, b = 76.9 Å, c = 57.0 Å, and β = 101.1°, with two αβ TCR heterodimers in the asymmetric unit [V_m = 2.95 Å^3 per dalton (*31*); solvent content = 58 percent]. The crystals have pronounced pseudo–face centering, with $h + k = 2n + 1$ reflections being systematically weak but not absent. The frozen 2C crystal had slightly altered unit cell parameters a = 138.0 Å, b = 72.9 Å, c = 57.3 Å, and β = 101.2° that resulted in an exact *C2* centering through a slight adjustment of one molecule (*30*). X-ray data were collected on a MAR imaging Plate mounted on a Siemens rotating anode operating at 40 kV and 50 mA with Supper double-focusing long mirrors. Data were integrated, scaled, and reduced with DENZO and SCALEPACK (*31*). An MR solution was obtained with the free β chain (1bec) as a search model (*15*) and the program AMoRe (31). The Vα domain position was determined in X-PLOR with an REI light chain monomer (68) as a search model, after positioning the β chain. A set of 1156 reflections (5 percent) was excluded from refinement for cross-validation (R_{free}) (*31*). Isomorphous heavy atom derivatives were obtained by soaking 2C crystals in 20 μM ethyl mercury thiosalicilate (EMTS) or 500 μM trimethyl lead acetate (TMLA) in 2.2 M ammonium sulfate, 50 mM sodium acetate (pH 7.0). Heavy atom sites were identified and refined with PHASES (*31*) and PHASIT (*31*). Phases to 3.2 Å were improved by solvent flattening and molecular averaging with DM (*31*) and by combining MIR and partial model phases by means of SIGMAA (*31*). Noncrystallographic symmetry restraints were imposed in simulated annealing refinement in X-PLOR (*31*). After one cycle of refinement and manual model building with O (*31*), SIGMAA-weighted F_o-F_c difference maps showed density for four N-linked glycans. After several cycles, the R_{cryst} and R_{free} were 0.20 and 0.37 for data with $F > 2\sigma_f$. The model was then placed in the new *C2* unit cell with the MR procedure in AMoRe (*31*) and refined with an R_{free} data set of 1536 reflections (5.4 percent). The high quality of this electron density map allowed unambiguous tracing of Cα. The R_{cryst} and R_{free} values are 0.208 and 0.285, respectively ($F > 2\sigma_f$) with individual B values. The 2C model contains 437 amino acids [residues α(1–213) and β(3–247)] (*34*), ten carbohydrate moieties, 30 ordered solvent molecules, and three sulfate ions. Of the residues in the 2C model, 85 percent fall in the most favorable regions of a Ramachandran plot [generated from PROCHECK (*31*)] with no residues in a disallowed region. Analysis by 3D-PROFILE (*31*) and ERRAT (*31*) shows the 2C model is consistent with highly refined protein structures.

Data set	Resolution limit (Å)	R_{sym}* (last shell)	Completeness (last shell) (%)	$\langle I/\sigma(I)\rangle$ (last shell)	Sites (n)	R_{iso}†	R_{Cullis}‡	R_{Kraut}§	Phasing power‖
Native-*C2*	2.5 (2.6–2.5)	4.8 (23.7)	99.7 (99.3)	18.3 (2.9)					
Native-$P2_1$	2.9 (3.0–2.9)	11.5 (54.0)	88.3 (80.2)	10.5 (1.1)					
EMTS-$P2_1$	3.2 (3.5–3.2)	9.8 (37.0)	73.0 (71.7)	10.0 (2.0)	2	8.2	57.6	11.6	1.54
TMLA-$P2_1$	3.2 (3.5–3.2)	11.1 (45.2)	80.0 (73.0)	9.8 (1.3)	2	9.8	56.1	13.1	1.73

Mean figure of merit = 0.340 (low because the EMTS and TMLA sites are almost identical)

Refinement statistics

Resolution (Å)	R value	R_{free}	rms from ideal values			Average B values (Å^2)			
			Bond lengths	Bond angles	Impropers	Vα	Cα	Vβ	Cβ
∞–2.5 ($F > 2\sigma$)	0.224 (0.208)	0.304 (0.285)	0.011 Å	1.95°	1.45°	36.9	50.3	29.8	27.1

$R_{sym} = 100^*\Sigma(|I_i(hkl) - \langle I(hkl)\rangle|)/\Sigma I_i(hkl)$, where $I_i(hkl)$ are the intensities of multiple measurements and $\langle I(hkl)\rangle$ is the average of the measured intensities of the ith reflection, with the summation extended over all measured reflections. †$R_{iso} = 100^*\Sigma(|F_{PH} - F_P|)/\Sigma F_P$, where F_{PH} is the derivative structure factor and F_P is the native structure factor and the sum is extended over all reflections common between the two data sets. ‡$R_{Cullis} = 100^*\Sigma||F_{PH}(hkl)| \pm |F_P(hkl)|| / \Sigma|F_{PH}(hkl)| \pm |F_P(hkl)||$, where the sum is extended over all centric reflections. §$R_{Kraut} = 100^*\Sigma(|F_{PH}(hkl)| - |F_H(hkl)|)/\Sigma|F_P(hkl)|$, where the sum is extended over all acentric reflections. ‖Phasing power = $\langle F_H\rangle/\varepsilon$, where ε is the lack-of-closure error.

From *Science*, Vol. 274 (11 October 1996): 209–219 DOI: 10.1126/science.274.5285.209 Reprinted with permission from AAAS.

mains. The Vα, Vβ, and Cβ domains were refined by X-PLOR simulated annealing (*31*) to an *R* value of 0.25. The Cα region showed some continuous density only for three β strands of the back sheet of the classical Ig fold (*17*), but there was no interpretable density for a top sheet. Although Cα had very few crystal contacts and there was precedent for constant domain disorder in some Igs (*32*), extensive disorder in a single Ig domain was puzzling. An MIR structure determination was undertaken to eliminate the possibility of MR model bias (Table 1); MIR and partial structure factor phases from the refined coordinates of the other three domains were combined and used to retrace and refine the entire molecule. Four of the β strands of Cα could be traced in an unambiguous register, but the outer strands remained fragmented. Successive rounds of model building, refinement, and phase combination resulted in an *R* value of 0.21, but with an R_{free} (*31*) of 0.37 at 2.9 Å. In addition, the presence of the pseudo-centering caused half of all the observed reflections to be vanishingly weak. Finally, a crystal of 2C was successfully cryocooled to 100 K, and a complete data set with very high redundancy was collected to 2.5 Å. A small change in the crystal lattice on freezing (*33*) resulted in exact crystallographic centering and a C2 space group (*30*). Refinement improved the electron density map (Fig. 1), especially in the Cα domain, where iterative cycles of model building and refinement allowed a complete trace, up to and including the COOH-terminal disulfide-linked cysteine residues ($Cys^{213\alpha}$-$Cys^{247\beta}$) (*34*), with good geometry and a substantial drop in R_{free} to 0.285 (Table 1). The carbohydrate density also improved, and at present ten sugar moieties have been located (Figs. 1 and 2A; Table 1).

Overall structure. The 2C α and β chains [α(1–213), β(3–247)] fold into a quaternary structure that resembles the antigen binding region (Fab) of antibodies (Fig. 2). Each chain contains two domains, three of which (Vα, Vβ, and Cβ) are related to Ig domains, and one of which (Cα) substantially deviates from the standard Ig fold (*17*). The overall dimensions of 2C are 61 by 56 by 33 Å, compared to about 65 by 46 by 30 Å for a representative Fab (*35*), whose elbow angle and CDR lengths are similar to the 2C TCR. The 2C appears slightly compressed in its length and wider across its middle elbow region than does an Fab. A classical measurement made in Fabs is the angle between the vectors that represent the pseudo-dyad axes relating V_L to V_H

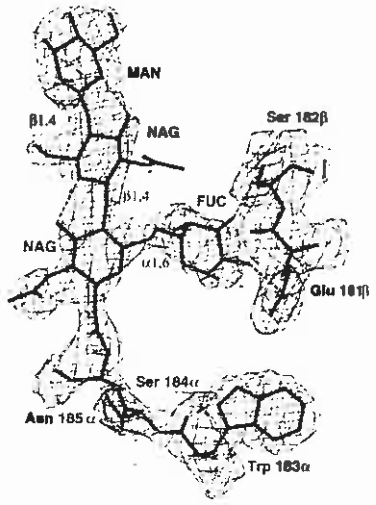


Fig. 1. Electron density in the 2C Cα-Cβ interface around $Asn^{185\alpha}$, showing an occupied N-linked glycosylation site (Asn^{185}-GlcNAc1 $\beta1,4$ GlcNAc2 $\beta1,4$ Man, $\alpha1,6$ Fuc1), as well as an invariant Trp residue ($Trp^{183\alpha}$). Close interaction with Cβ is shown by hydrogen bonds (dotted lines) of the fucose residue to side chains of Cβ. The electron density map was generated from SIGMAA-weighted $2F_o$-F_c coefficients (*31*) and contoured at 1.0 σ with the program TOM (*69*). NAG, GlcNAc.

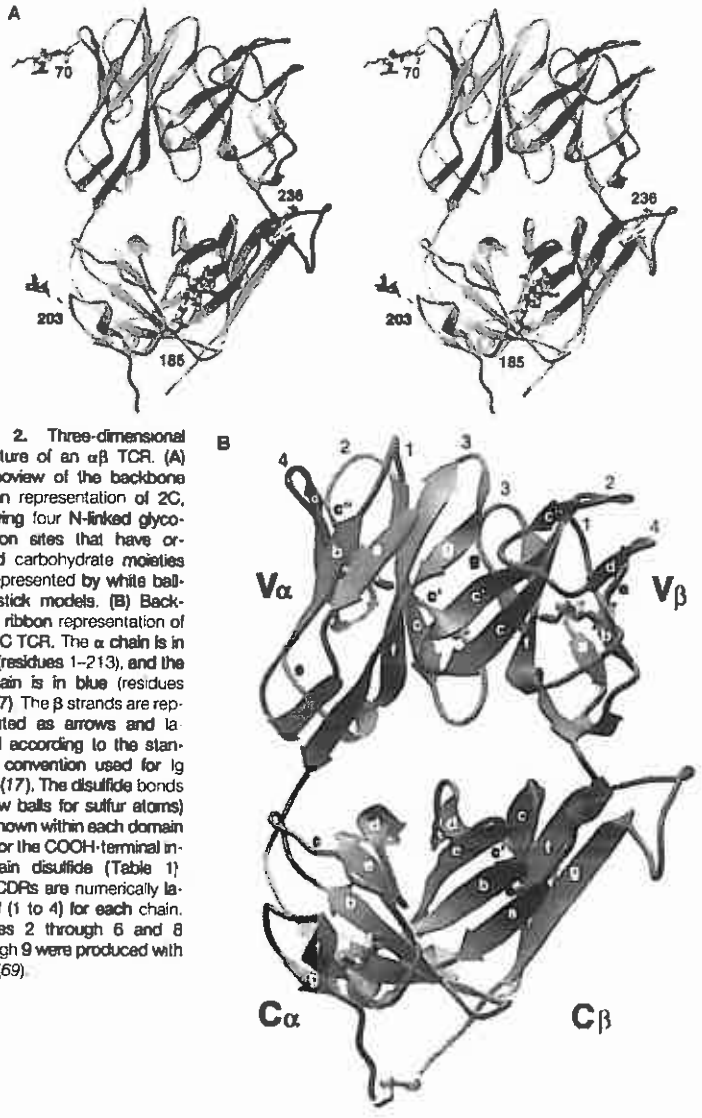


Fig. 2. Three-dimensional structure of an αβ TCR. (**A**) Stereoview of the backbone ribbon representation of 2C, showing four N-linked glycosylation sites that have ordered carbohydrate moieties as represented by white ball-and-stick models. (**B**) Backbone ribbon representation of the 2C TCR. The α chain is in pink (residues 1–213), and the β chain is in blue (residues 3–247). The β strands are represented as arrows and labeled according to the standard convention used for Ig folds (*17*). The disulfide bonds (yellow balls for sulfur atoms) are shown within each domain and for the COOH-terminal interchain disulfide (Table 1). The CDRs are numerically labeled (1 to 4) for each chain. Figures 2 through 6 and 8 through 9 were produced with AVS (*69*).

109

From *Science*, Vol. 274 (11 October 1996): 209–219 DOI: 10.1126/science.274.5285.209 Reprinted with permission from AAAS.

and C_L to C_H1, otherwise known as the elbow angle, which at present ranges from 133° to 220° (*7*). For 2C, the elbow angle is 148.6° (*36*), within the normal range for an Fab, but about 6° as small as the 14.3.d β chain structure, indicating some segmental flexibility between the V and C domains. The pseudo-dyad relating Cα to Cβ is shifted toward the α chain by about 4.6 Å relative to the Vα-Vβ pseudo-dyad, giving the TCR a somewhat more asymmetric appearance.

In the Vα crystal structure (*16*), the Vα forms homodimers that themselves dimerize, leading to a proposal (*16*) that this arrangement is spatially compatible with the dimer seen in the class II MHC crystal structure (*37*) and could therefore be part of a dimeric TCR-MHC signaling assembly.

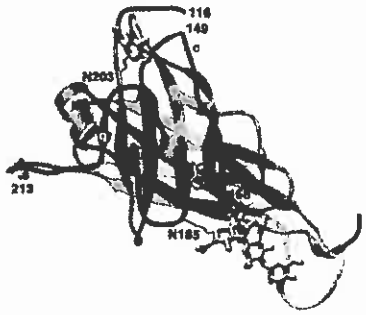


Fig. 3. Backbone ribbon representation of Cα. The Cα domain (pink) begins at residue 116α and ends at the COOH-terminal residue 213α, and is shown lying on top of the Cβ domain (blue). The β strands of the back β sheet in Cα, in lighter pink, are depicted as arrows, with the unusual Cα "top strands" in darker pink. The strands are labeled according to the designations used for standard Ig folds (*17*). The back β sheet actually comprises strands g, a, b, e, and d with strand g H-bonding in a parallel arrangement with strand a. Strand f contains a short helical segment that is disulfide bonded (partially hidden yellow balls represent the sulfurs) to the back b strand. The other side of the back β sheet of Cα (light pink) forms the interface with the back β sheet of Cβ. The carbohydrate moieties of the two N-linked glycosylation sites (Asn$^{185\alpha}$ and Asn$^{203\alpha}$) in Cα are also shown. The Trp$^{183\alpha}$ in the Cα-Cβ interface is shown as a white ball-and-stick representation. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

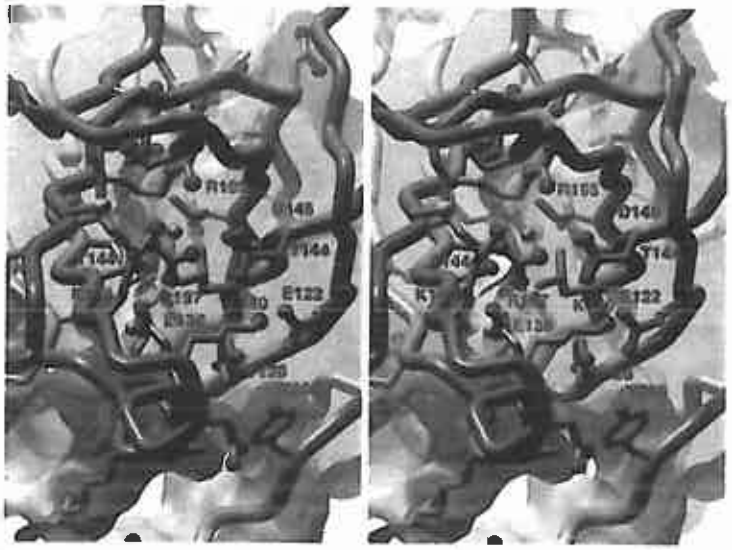


Fig. 4. The Cα-Cβ interface is highly polar. A stereoview of the interacting β sheets of Cα (pink) and Cβ (blue) is shown along with side chains making either salt bridges, H bonds, or van der Waals contacts across the interface [H bonds calculated with HBPLUS (*52*) and van der Waals and salt bridges calculated with CONTACSYM (*52*)]. The complementarity of the Cα-Cβ interface is emphasized by the superposition of a cross-section of the interacting surfaces as a translucent ghost. A predominance of basic side chains lies on the Cβ side of the interface, and acidic side chains on the Cα side, along with many other polar amino acids. Asp$^{145\alpha}$ makes salt bridges with Arg$^{197\beta}$ and Lys$^{140\beta}$. Glu$^{122\alpha}$ and Tyr$^{126\alpha}$ H-bond to Lys$^{140\beta}$ and Thr$^{144\alpha}$ to Arg$^{195\beta}$. Met$^{170\alpha}$ is in van der Waals contact with Lys$^{142\beta}$. The buried water molecules within the interface are shown. The molecular surfaces were calculated by means of AVS (*69*) with a 1.4 Å probe and the POMS program (*69*).

The 2C TCR, however, is monomeric both in solution and in the crystals (*38*).

The amino acid sequences (*39*) of mouse and human TCR C regions suggested structural homology with Ig C domains (*4*) although the sequence similarity is substantially higher for Cβ than for Cα. About 30 percent sequence identity is seen for up to 100 Cβ residues in antibody C domains C_L and C_H1, with similar spacings between the intrachain disulfide cysteine residues, and preservation of key structural residues such as Trp^{161} in the c strand. The Cβ contains an unusual insertion between residues 219β and 232β (Fig. 2B) that otherwise is found only in TCR Cγ sequences (39, 40). The Cβ2 structure of 2C is indeed similar (*41*) to the C domains of both heavy and light chains of Igs (2.0 to 2.2 Å root mean square (rms) deviation for 89 matched residues for C_L and C_H1), but it is more similar to the C_H3 domain of the human Fc fragment of an Ig (1.6 Å rms deviation for 89 residues). The 2C Cβ2 gene sequence is almost identical to the 14.3.d Cβ1, and their domain structures overlap with small differences (1.6 Å rms deviation for all residues) except in two major areas. Residues 217β to 236β, which comprise the large, solvent-exposed insertion in Cβ, has a different loop conformation from that of 14.3.d (2.0 Å rms deviation for 20 α carbons), indicating some conformational flexibility in this region (*42*). The other difference occurs for the a-b loop (residues 133β to 144β) that folds into a short α helix in 2C in the interface with Cα. In Ig C_L-C_H1 and C_H3-C_H3 pairings, the analogous residues also form a short α helix that is close to the a, b, e, d interface β sheet. In 14.3.d, these residues drift away from the β sheet in the absence of an opposing Cα chain.

Four of the seven possible N-linked glycosylation consensus sequences (*43*) have ordered carbohydrates (Figs. 1, 2A, and 3). In Vα, three carbohydrate moieties from Asn$^{70\alpha}$ (GlcNAc1 $\beta1,4$ GlcNAc2, $\alpha1,6$ Fuc1) (*43*) extend out into a large solvent channel in the crystal, with the amide nitrogen of the first *N*-acetylglucosamine moiety (GlcNAc1) H-bonded back to the side-chain carbonyl of Asn$^{70\alpha}$ (Fig. 2A). Asn$^{236\beta}$ has a visible disaccharide (GlcNAc1 $\beta1,4$ GlcNAc2) in which the O6 of GlcNAc1 is in a water-mediated H bond to Arg$^{211\beta}$ of Cβ. In Cα, density for one GlcNac moiety extends from Asn$^{203\alpha}$ toward the solvent. The largest ordered carbohydrate stretch extends from Asn$^{185\alpha}$ (GlcNAc1 $\beta1,4$ GlcNAc2 $\beta1,4$ Man, $\alpha1,6$ Fuc1) (Figs. 1 and 3) and is involved both in crystal contacts and Cα-Cβ interactions. The fucose moiety (*43*) is H-bonded to Glu$^{181\beta}$ and Ser$^{182\beta}$ (Fig. 1). A water molecule bridges the fucose exocyclic oxygen and

From *Science*, Vol. 274 (11 October 1996): 209–219 DOI: 10.1126/science.274.5285.209 Reprinted with permission from AAAS.

the side chain of $Arg^{150\beta}$. The GlcNAc2 moiety also H-bonds to a symmetry-related molecule. This N-linked sugar possibly strengthens association of Cα with Cβ and may also influence the crystal packing of 2C, which perhaps explains why non-glycosylated 2C TCR does not crystallize on its own or why it cannot be seeded with glycosylated crystals. The other carbohydrate moieties in 2C are presumably disordered or heterogeneous, or the N-linked sites are not occupied (*43*) although all the Asn residues are on exposed strands or β turns.

The Cα structure. The TCR Cα and Cδ domains are the most divergent from known Ig sequences (*4*), showing only 12 to 18 percent sequence identities with Ig C domains. Cα was predicted to be an Ig fold because of the presence of the two characteristic cysteine residues that normally link the top (g, f, c) and bottom (a, b, e, d) β sheets. Many other structurally conserved residues in Igs are missing, such as the Trp residue that normally occurs in the c strand 14 or 15 residues after the first cysteine of the intrachain disulfide. The Cα has a relatively short span (50 amino acids) between its two cysteine residues compared to Ig C domains (60 to 65 amino acids), indicating a deletion of about 12 to 15 residues whose locations are not obvious from sequence alignments.

The Cα structure shows that the a, b, e, and d strands comprising the back β sheet are highly similar to the corresponding back sheet of Ig C domains (Fig. 3). An overlap of this β sheet with the structurally equivalent residues of Ig C_L domains gives an rms deviation of ~1.3 Å for 29 Cα residues (*41*). However, large differences emerge for the c, f, and g strands, which would normally make up the top β sheet in a standard Ig-fold (Fig. 3). Instead, these strands are loosely packed against the bottom sheet (*44*) and are too far apart to form main-chain hydrogen bonds with each other. Gaps between these top elements result in exposure of some of the hydrophobic interior of the bottom sheet to solvent. The c strand connects directly from the b to the d strand and eliminates about six residues of the c′ strand found in the classical c-type Ig set (*17*). Only 14 residues connect the c and e strands, compared to 21 in c-type folds and would place the Cα in the h- or s-type of Ig folds (*17*). However, classification is complicated by the absence of an outer β sheet that represents a major structural deviation from all other Ig-type folds. The c strand is loosely tethered to the bottom sheet, as reflected in the higher thermal parameters of this strand (~69 $Å^2$) (*44*), by interactions between small hydrophobic residues (Pro, Val, Ile), as compared to a standard Ig-fold in which a conserved Trp residue (Trp^{148L}, Trp^{157H}) anchors this strand to the bottom sheet.

Another divergence from c-type folds comes in the loop connecting the e and f strands. Normally, this loop is large (about ten residues) and often adopts a minihelix containing a second conserved Trp residue that occupies the internal core of the loop. In Cα, the connection consists of a three-residue turn (Fig. 3). $Trp^{183\alpha}$ does occur at the carboxyl end of the e strand (Figs. 1 and 3), just before the turn, but it is on the opposite side of the strand, rather than in its usual interior β-sheet position and is buried in a hydrophobic pocket at the Cα-Cβ interface. The e-f turn contains a conserved N-linked glycosylation site ($Asn^{185\alpha}$) for which clear electron density exists for four carbohydrate moieties adjacent to the Cα-Cβ interface (Fig. 1). The deletion of about six residues between the c and d strands, and five in the turn connecting the f and g strands, accounts for the 10 to 15 fewer amino acids between the Cα cysteines (50 residues) than for antibody C or V domains (60 to 65 residues).

After the e-f turn, the Cα again deviates from a canonical Ig fold (Fig. 3). The f strand is also not H-bonded to a neighboring strand, although the second cysteine ($Cys^{191\alpha}$) remains disulfide linked to $Cys^{141\alpha}$ in an ideal disulfide geometry, as in Igs (*45*). Surprisingly, $Cys^{191\alpha}$ is in a short helical turn (residues 191 to 195) that places the side chain of $Phe^{195\alpha}$ into a hydrophobic pocket created by $Phe^{143\alpha}$ and $Phe^{145\alpha}$ of strand b of the back β sheet. This short α-helical stretch ($Gln^{192\alpha}$-$Asp^{193\alpha}$-$Ile^{194\alpha}$-$Phe^{195\alpha}$) does not appear to be chemically compatible with the normal hydrophobic-hydrophilic alternation of residues in a surface-exposed β strand. After the minihelix, the chain adopts a short turn from which another carbohydrate moiety extends from $Asn^{203\alpha}$ at the tip of the loop (Fig. 3). Three residues of the final g strand ($Ala^{204\alpha}$ to $Tyr^{206\alpha}$) do, in fact, form a β strand, but, surprisingly, H-bond with the a strand of the bottom sheet, a variation not seen in Ig C domains or other Ig folds (*46*). $Pro^{207\alpha}$ then kinks the β strand toward the β chain so that its C-terminal $Cys^{213\alpha}$ can make the interchain disulfide with Cβ $Cys^{247\beta}$ (*28*).

Studies on TCR biosynthesis and signaling allow us to consider the functional relevance of the unusual top strand topology and possible flexibility. The α chain is unusually unstable and susceptible to degradation in the endoplasmic reticulum (*47*), that limits high-level expression in transfected eukaryotic cells (*21*). The instability of the α chain may be partly related to a less-compact fold of the Cα domain. Several site-directed mutagenesis studies have implicated residues in the top elements of the Cα domain as being critical for interaction with the CD3 ζ chain (*48*). Some flexibility in these outer elements of Cα may provide a mechanism for adaptability in binding CD3.

The Cα-Cβ interface. The Cα-Cβ association is more like that of C_H3-C_H3 than it is C_H1-C_L-like with respect to the geometry of the domain interaction, the buried surface area, and a predominance of charged interactions in the Cα-Cβ interface (Fig. 4). The total buried surface area (*49*) between Cα and Cβ is 2049 $Å^2$, as compared to an average of 1700 $Å^2$ for C_H1-C_L (the range is 1500 to 2100 $Å^2$). The equivalent

Fig. 5. The ball-and-socket joint of antibodies is conserved in the TCR. The elbow region between variable and constant domains for (top) 2C (α, pink; β, blue) and (bottom) Fab NC41 (L chain, pink; H chain, blue) is shown for these two structures with similar elbow angles (148°) (*35*). The amino acid residues for the ball-and-socket joint are colored yellow for 2C residues $Val^{12\beta}$, $Ser^{115\beta}$, $Leu^{116A\beta}$, $Phe^{163\beta}$, and $Pro^{154\beta}$ and for Fab-NC41 (*35*) residues Leu^{11H}, Thr^{110H}, Ser^{112H}, Phe^{148H}, and Pro^{149H}. Other interdomain (Vα-Cα, Vβ-Cβ, V_L-C_L, V_H-C_H1) contact residues, as determined by CONTACSYM (*52*), are shown to highlight the comparative number of residues in the V-C interface of 2C TCR and an Fab.

From *Science*, Vol. 274 (11 October 1996): 209–219 DOI: 10.1126/science.274.5285.209 Reprinted with permission from AAAS.

C_H3 dimerization in Fc buries about 2200 Å^2. The crossing angle between opposing β sheets is −41°, as opposed to around −53° for most C_H1-C_L associations, and −45° for C_H3-C_H3 (*50*). The more parallel Cα-Cβ intersection angle gives the TCR structure its squat appearance (Fig. 2). Although the Cα and Cβ are related by a 180° dyad (*36*), similar to C_H1-C_L, a pairwise superposition also reveals a higher degree of similarity to C_H3 domain interactions (*51*). When a C_H1 domain is superimposed onto the Cβ, a 15° rotation is required to bring C_L to optimal superposition with Cα. However, when a domain of C_H3 is superimposed onto Cβ, only a 3° rotation is required to overlap the other C_H3 domain with Cα.

The Cα-Cβ interface is highly polar, with a skewed distribution of acidic residues on Cα and basic residues on Cβ (Fig. 4). A highly conserved salt bridge, found in all C_H1-C_L (Glu^{123L}-Lys^{213H}) and C_H3-C_H3 (Glu^{357H}-Lys^{370H}) interfaces, is retained in a similar location as a constellation of charged residues at one end of the Cα-Cβ interface (Fig. 4). This cluster of charges includes acidic Cα residues $Asp^{145\alpha}$ and $Glu^{122\alpha}$ that form multiple salt bridges and H-bond interactions (*52*) with Cβ basic residues $Lys^{140\beta}$, $Arg^{195\beta}$, and $Arg^{197\beta}$. In addition, $Glu^{136\beta}$ and two buried water molecules provide polar contacts along with a number of other side-chain H bonds. The other end of the interface is largely hydrophobic involving $Trp^{183\alpha}$, $Phe^{161\alpha}$, $Tyr^{179\beta}$, and $Leu^{148\beta}$. Interaction of Cβ with a carbohydrate extending from Cα $Asn^{185\alpha}$ is discussed above.

Fig. 6. Stereoview of the Vα-Vβ interface. **(Top) Backbone** representation including interacting conserved interdomain TCR side chains (α, pink; β, blue). The labeled side chains are conserved across all mouse and human TCRs and are also conserved in antibody V_H-V_L interfaces. Water molecules (small blue spheres) are present along the edges of the interface. **(Bottom)** The Vα-Vβ complementarity is shown by this molecular surface cross-section. The cluster of aromatic side chains between Vα and Vβ accounts for most of the buried surface area in the interface.

Elbow region. An extensive interaction occurs between the V and C domains (Fig. 5), with about 700 Å^2 of buried surface between Vβ and Cβ, and 650 Å^2 between Vα and Cα (*49*). The Vβ-Cβ interaction area is highly polar, with a conserved salt bridge ($Glu^{158\beta}$-$Arg^{113\beta}$) and eight interdomain H bonds, most of which are between conserved β chain residues. The Vα-Cα interface, on the other hand, is composed primarily of van der Waals interactions, with only one H bond between the $Ser^{148\alpha}$ hydroxyl and the main-chain carbonyl oxygen of $Pro^{116\alpha}$. The Vβ-Cβ interface is substantially larger than a corresponding V_H-C_H1 interface (390 Å^2) in an Fab with a similar elbow angle (*35*). The Vα-Cα interface is also larger but differs less from corresponding V_L-C_L interfaces (~500 Å^2). The conserved "ball-and-socket joint" in the V_H-C_H1 interface of antibodies is present (Fig. 5) in the 2C β chain ($Val^{12\beta}$, $Ser^{115\beta}$, $Leu^{116A\beta}$, $Phe^{153\beta}$, and $Pro^{154\beta}$), as predicted (*9*). A superposition of 2C and 14.3.d shows that 2C Vβ-Cβ has rotated around this ball-and-socket joint relative to 14.3.d so that the elbow angle in 2C is 148.6° (*36*), compared to about 155° in 14.3.d. The conservation of the ball-and-socket joint suggests some possible flexibility in the TCR elbow angle, contrary to some hypotheses of TCR signal transduction that invoke rigidity of the β chain (*40*).

The variable regions. As predicted from sequence alignments (*4*), the 2C variable domains resemble antibody variable domains (Fig. 2B). Both Vα and Vβ have residues that are conserved across all mouse and human TCRs and many of these are similarly conserved in Fab sequences. Their locations in Fab x-ray structures correspond to structurally important locations, such as the V_L-V_H interface or the β-sheet framework. The TCR has its own subset of conserved residues, which are not found in Fab sequences and, hence, indicates residues or regions that relate directly to their own

From *Science*, Vol. 274 (11 October 1996): 209–219 DOI: 10.1126/science.274.5285.209 Reprinted with permission from AAAS.

specific function.

In 2C, Vα and Vβ associate closely in a pseudo-dyad arrangement of 174°, well within the normal range for V_H-V_L associations (*7*, *53*, *54*). Within the Vα-Vβ interface (Fig. 6), 1160 Å^2 (610 Å^2 for Vα; 550 Å^2 for Vβ) of surface area (*49*) is buried, as opposed to an average value of 1420 Å^2(*7*, *54*) found in Fabs (ranges from 1068 to 1700 Å^2). About 33 percent is contributed by CDR residues (mostly CDR3s), which is again smaller than the 40 to 50 percent average in Fabs (*54*). The β-sheet crossing angle (*50*) between the inside sheets of Vα and Vβ is −62°, which is larger than for any Fab (ranges from −38° to −60°), and 10° larger than for the Vα homodimer 1934.4 (*16*).

General conclusions about trends in the Vα-Vβ association must await other TCR structures. Notwithstanding, the 2C Vα-Vβ domain buried surface area value is at the extreme end of the range seen in antibodies. The only Fab determined thus far whose V_L-V_H buried surface area is as small as 2C Vα-Vβ is Fab 50.1 (1063 Å^2), of which ~35 percent is contributed by the CDR loops (*54*). Fab 50.1 experiences a very large V_L-V_H rotational displacement of 16.3° on binding its peptide antigen, which may be the largest interdomain rotation observed so far between free and bound Fab structures (*54*). Smaller V_L-V_H contact surface areas appear to correlate with an ability to use changes in domain association for better complementarity of fit with antigen (*54*, *55*).

The interdomain contact residues between Vα and Vβ are conserved in TCRs (Fig. 6) and in the corresponding V_L-V_H interface in antibodies (*9*). These residues form a twofold symmetric core with an invariant pair of side-chain to side-chain H bonds between $Gln^{37\alpha}$ and $Gln^{37\beta}$ at the base of the interface. Both $Tyr^{35\alpha}$ and $Tyr^{35\beta}$ H-bond to the opposing backbone of CDR3 residues $Leu^{106\beta}$ and $Leu^{104\alpha}$, respectively. Three interdomain H bonds form along the edge of the inner β sheets, which are water mediated ($Leu^{45\alpha}$-water-$Leu^{106\beta}$, $Phe^{106\alpha}$-water-$Leu^{43\beta}$, $Ser^{108\alpha}$-OH-water-$Gly^{40\beta}$). The other conserved residues ($Leu^{43\alpha}$, $Leu^{45\alpha}$, $Phe^{89\alpha}$, $Ala^{91\alpha}$, $Leu^{43\beta}$, $Leu^{45\beta}$, $Phe^{91\beta}$, and $Ala^{93\beta}$) form a symmetrical hydrophobic core between the Vα-Vβ inner β sheets, as predicted (*9*).

Within each Vα and Vβ, about 24 residues form the core of the β-sheet sandwich and are conserved in mouse and human TCRs, as well as in Igs (*8*, *9*). When Vβ is superimposed onto six representative Fab V_H's (*56*) by overlapping 36 structurally equivalent framework residues, an average 1.6 Å rms deviation is obtained for up to all 70 β-sheet framework residues (*56*). A similar overlap of Vα onto six V_L's yields an average rms deviation of 1.2 Å for 76 structurally equivalent Vα residues, indicating that Vα is more similar to V_L than Vβ is to V_H (*56*). Superimposition of Vβ (*57*) is equally good with all V_H's and V_L's in the Protein Data Bank (~2.0 Å rms deviation), but Vα appears to superimpose better with V_L than with V_H (1.6 Å versus 2.0 Å rms deviation).

In the 2C Vα domain, as in the 1934.4 Vα structure (*16*), the c" strand is H-bonded to the d strand of the outer β sheet as opposed to the c' strand, as in v-type Ig domains (*17*). Framework residue $Lys^{17\alpha}$, which is highly conserved in Vα subgroup I but not in Ig frameworks, is at the carboxyl end of the e strand and H-bonds to the main chain of the d strand, possibly keeping the solvent-exposed strand-switched c"-d loop pinned next to the core of the Vα domain. The Vβ domain folds into a canonical v-type Ig set (*17*). An invariant residue in the framework of subgroup II β chains, but not in Igs, is $His^{47\beta}$, which H-bonds to the d strand and helps tether the c"-d loop, known as HV4, to the main body of Vβ. HV4 is an additional region of hypervariability observed in TCRs thought to be involved in superantigen (Vβ) or accessory molecule (Vα) interaction (*58*). Since $His^{47\alpha}$ is not conserved in any other mouse or human subgroup, the position of Vβ HV4 may differ in other TCRs.

The relative positions of the 2C CDRs, with the exception of Vα CDR2, are close to, but not exactly coincident with, corresponding CDRs in antibodies (Fig. 7A). The Vα CDR1 is eight residues long and is stabilized by hydrophobic interactions between highly conserved residues $Tyr^{74\alpha}$, $Leu^{32\alpha}$, and $Phe^{66\alpha}$ (Fig. 7B). An H bond between the $Tyr^{26\alpha}$ and $Ser^{102\alpha}$ hydroxyls serves to anchor both CDR1 and CDR3 (Fig. 7B). As the TCR CDR1 does not vary in length by more than three residues, its conformation is probably rather similar in most TCRs. However, CDR2, is more vari-

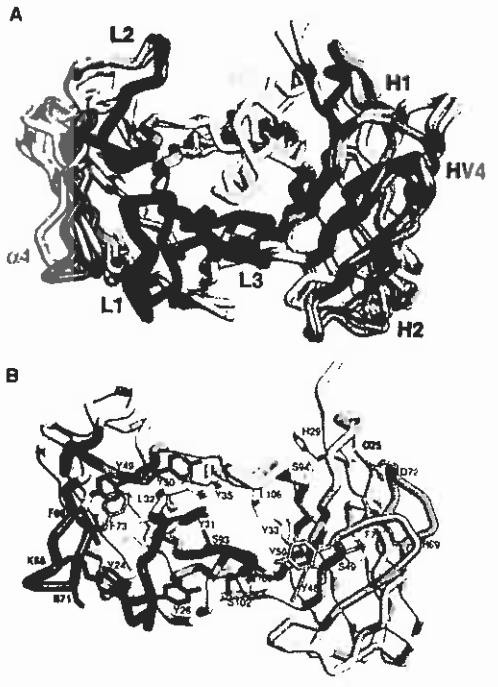


Fig. 7. The antigen binding site of a TCR. (A) Comparison of TCR and antibody combining sites. An overlap of 2C Vα and Vβ is shown with three representative Fab V_L and V_H domains (*56*). The relative dispositions of the CDRs can then be compared. Vα was optimally superimposed with three Fab V_L's, and Vβ was separately superimposed with three Fab V_H's with the program OVRLAP (*36*). The 2C CDRs are colored in red (except HV4, which is orange, and its equivalent in α chains, labeled α4). The Fab CDRs are labeled (L1, L2, L3, H1, H2, and H3) and are in different colors. The α chain is to the left and CDRs α1, α2, α3, and α4 correspond to the Fab CDRs L1, L2, L3, and L4 (only exists in TCRs, so is not colored in the Fabs), respectively. The β chain is to the right and CDRs β1, β2, β3, and HV4 correspond to the Fab CDRs H1, H2, and H3, with the fourth Vβ CDR again with no corresponding hypervariability in Fabs. (B) CDR stabilizing residues in the 2C TCR. The view into the combining site of 2C shows side chains that are important for CDR main-chain conformations and positions. The α chain is to the left, β chain to the right. Backbone representations of the 2C CDRs are as follows: α1, navy; α2, magenta; α3, green; α4, orange; β1, cyan; β2, pink; β3, yellow; and HV4, orange. The assignment of structurally important residues is based on contacts [van der Waals, H bond, or salt bridge as determined by CONTACSYM and HBPLUS (*52*)] of framework (colored white) or CDR (colored by CDR code) residues with main-chain or side-chain residues of a CDR. These residues may affect the conformation of the CDR itself or regions outside of the CDR that are critical to the positioning of the CDR within the combining site. [Figure produced with MidasPlus version 2.0 (*60*)]

113

From *Science*, Vol. 274 (11 October 1996): 209–219 DOI: 10.1126/science.274.5285.209 Reprinted with permission from AAAS.

able, as reflected by its different conformation in 2C versus 1934.4 (*16*). The top of the loop bulges away from the center of the combining site, with the phenolic ring of $Tyr^{49\alpha}$ taking the place of the main-chain path in CDR2 of 1934.4. $Tyr^{49\alpha}$ provides most of the specific interactions for this loop (Fig. 7B), through an H bond from its hydroxyl to the main-chain carbonyl of residue 52α. Hydrophobic interactions of the $Tyr^{49\alpha}$ aromatic ring with $Phe^{66\alpha}$ and $Phe^{73\alpha}$ pack the CDR2 (c′–c″ strands) against the Vα core (Fig. 7B). The main chain of Vα CDR2 is almost perpendicular to that of CDR2 in antibodies (Fig. 7A) because of the c″ strand switch. Vα CDR3 is short (eight residues long) and has a type-II β turn conformation, with Phe^{100}, protruding from the surface of the binding site (Fig. 8B). Two H bonds from the strand leading to the CDR1 turn ($Ser^{93\alpha}$-OH-$Tyr^{31\alpha}$-OH and $Ser^{102\alpha}$-OH-$Tyr^{26\alpha}$-OH) connect to residues in CDR3 (Fig. 7B). $Ala^{103\alpha}$ in CDR3 is within the Vα-Vβ interface and buried in a hydrophobic pocket composed entirely of Vβ aromatic residues (Fig. 7B).

The conformation of Vβ CDR1 is similar to 14.3.d CDR1 (*15*) and is strongly constrained by main-chain H bonds with the side chain of $Gln^{25\beta}$ (Fig. 7B). CDR2 also has the identical hairpin conformation as 14.3.d and maintains a similar set of H bonds between $Ser^{49\beta}$ and $Arg^{69\beta}$. HV4, which is entirely solvent accessible, has its conformation maintained by intraloop H bonds from the $Arg^{69\beta}$ guanidinium group to the main-chain atoms of residues 71 and 72. The tip of HV4 is held toward the center of the binding site by H bonds between $Gln^{72\beta}$ and side-chain and main-chain atoms of residues 28β and 29β at the tip of CDR1 (Fig. 7B). CDR3 has a classic gamma-type turn at its tip (*59*) with an unusual repeat of four consecutive glycine residues (Fig. 8A). The Vβ CDR3 interacts with Vα CDR1 through an H bond from $Tyr^{31\alpha}$ to the main-chain carbonyl of Vβ $Leu^{106\beta}$ (Fig. 7B). Additional stabilization of the loop position may be derived from a water-mediated H bond from the main chain of $Gly^{98\beta}$ to the $Ser^{98\alpha}$-OH in CDR3. The Vβ CDR3 moieties of many TCRs are glycine-rich, which is the consequence of TCR D regions coding for a high proportion of glycines in all three reading frames (*60*).

The structural elucidation of the CDRs of 2C enables us to address the issue of CDR boundaries. The CDR boundaries in TCRs have been chosen by analogy to the Kabat hypervariable regions in antibodies but are not as clearly delineated (*6*, *39*). A structural definition of a CDR as those residues within the hypervariable loops that are actually accessible to antigen reveals that the structural CDRs can be shorter than the sequence-defined CDRs (*61*). Using this accessibility criterion, the CDR ranges for 2C would be as follows: CDR1α, 24–31; CDR2α, 48–55; CDR3α, 93–104; CDR1β, 26–31; CDR2β, 48–55; and CDR3β, 95–107.

Five of the six CDRs in antibodies fall into a small number of main-chain structural subclasses (*61*) that can now be predicted on the basis of loop size, composition, and the presence of certain conserved residues. These commonly occuring conformations of the CDRs, called canonical structures (*61*), are universal for antibodies and are not specific to individual Kabat subgroups (*39*). Whether canonical CDR structures exist for TCRs has not been clearly identifiable from modeling of TCR sequences on antibodies (*40*). The 2C TCR structure shows that

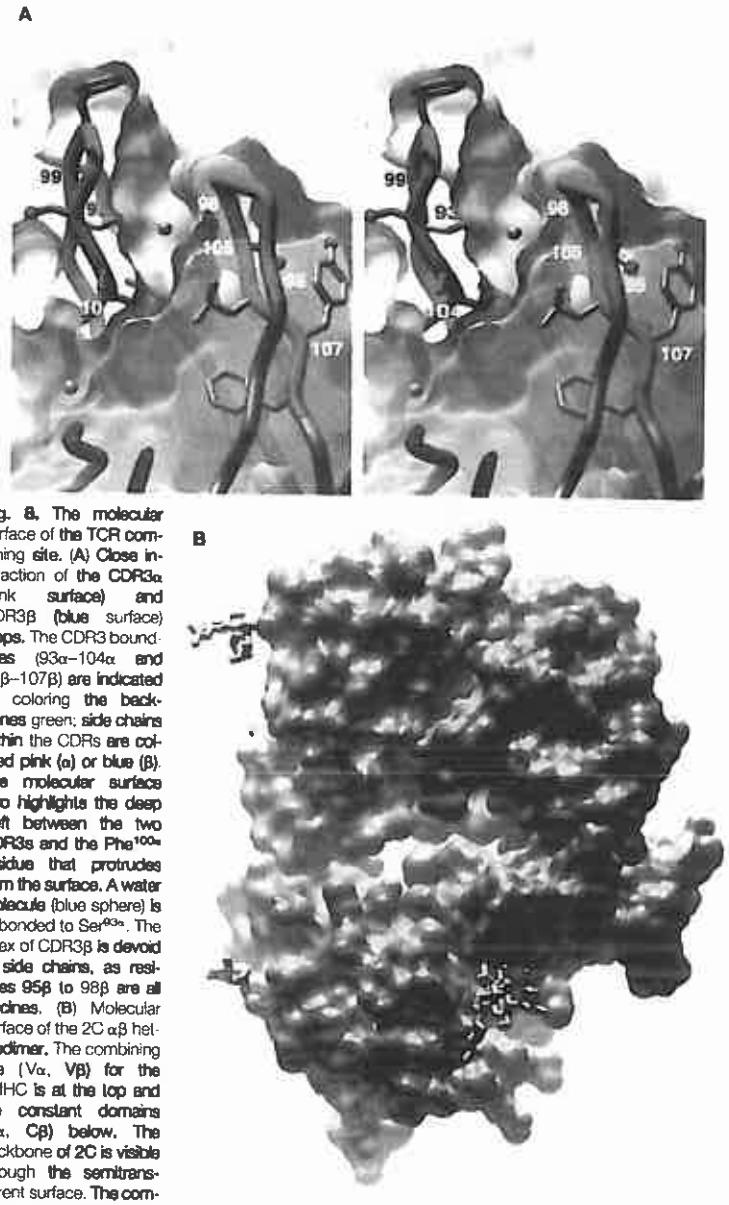


Fig. 8. The molecular surface of the TCR combining site. (A) Close interaction of the CDR3α (pink surface) and CDR3β (blue surface) loops. The CDR3 boundaries (93α–104α and 95β–107β) are indicated by coloring the backbones green; side chains within the CDRs are colored pink (α) or blue (β). The molecular surface also highlights the deep cleft between the two CDR3s and the $Phe^{100\alpha}$ residue that protrudes from the surface. A water molecule (blue sphere) is H-bonded to $Ser^{93\alpha}$. The apex of CDR3β is devoid of side chains, as residues 95β to 98β are all glycines. (B) Molecular surface of the 2C αβ heterodimer. The combining site (Vα, Vβ) for the pMHC is at the top and the constant domains (Cα, Cβ) below. The backbone of 2C is visible through the semitransparent surface. The combining site is mainly flat and undulating.

From *Science*, Vol. 274 (11 October 1996): 209–219 DOI: 10.1126/science.274.5285.209 Reprinted with permission from AAAS.

canonical CDR conformations would most likely exist only within Kabat subgroups, but not across all TCRs. Many of the CDR-stabilizing residues in 2C are conserved only within β subgroup II and α subgroup I. For instance, $Gln^{25\beta}$ appears to play a major role in stabilizing CDR1β but is invariant only in subgroup II of the mouse and human, whereas in subgroup I it is Pro. $Ser^{49\beta}$ and $Arg^{69\beta}$, H-bonded to CDR2β (Fig. 7B), are invariant in β subgroup II, but in other subgroups can be Phe (human subgroup I) or Phe(His) (mouse subgroup I). The CDR2α loop conformation and position appear to require a hydrophobic interaction between $Tyr^{49\alpha}$ and $Phe^{66\alpha}$, which are conserved in α subgroup Ia, but are small hydrophobic residues in other subgroups. Subgroup distinctions in TCRs then appear to be important determinants of antigen recognition and individual TCR functions (62).

The overall binding surface of 2C is relatively flat (Figs. 2B and 8B). With the exception of a hydrophobic pocket between CDR3α and CDR3β (Fig. 8A), the absence of large protruding loops or deep clefts is characteristic of the undulating surfaces of antibodies to proteins (7). The CDR surface is neither exceptionally hydrophobic or polar, but rather presents an even distribution of residues of both chemical types. One feature of the binding surface is the paucity of large side chains protruding from the tips of the CDRs. Many glycine and other small residues are at the apices of the CDRs ($Ala^{28\alpha}$, $Gly^{92\alpha}$, $Gly^{99\alpha}$, $Ala^{101\alpha}$, $Ala^{103\alpha}$, $Gly^{51\beta}$, $Gly^{53\beta}$, $Gly^{95\beta}$, $Gly^{96\beta}$, $Gly^{97\beta}$, and $Gly^{98\beta}$) in positions that would seem likely to be within TCR-pMHC interface. The absence of protruding side chains could allow the TCR to approach the pMHC surface more closely so that a smaller subset of side chains could supply the main binding energy for specific TCR-pMHC interactions.

Several models have been proposed to describe the relative orientation of a TCR with respect to a pMHC (*4*, *12*–*14*). The CDR3s of the α and β chains are proposed to interact predominantly with the bound peptide, whereas the locations of the less diverse CDRs 1 and 2 are thought to lie over the α helices of the MHC (*4*, *12*, *13*) or partially interact with both helices and peptide (*14*). The approximate overall dimension (63) of the 2C binding surface is 36 Å by 20 Å (Fig. 8B), as compared to the H-$2K^b$ binding surface of 32 Å by 22 Å. The tips of CDR3α (101α) and CDR3β (97β) are 7.3 Å apart (Fig. 8A). A hydrophobic pocket between the two CDR3s (~10 Å deep) appears ideally situated to accommodate one large side chain (*64*) that would presumably come from the peptide component of the pMHC complex (Fig. 8A). The exposed peptide surface in H-$2K^b$ is ~12 Å wide; that could allow both CDR3α and CDR3β to fit between the MHC α helices and hence interact with the peptide side chains. The α helices of the pMHC are about 18 Å apart from each other, nearly the same as the separation between CDRs 1 and 2 of the α chain and CDRs 1 and 2 of the β chain.

Orientation of the TCR-pMHC Complex. Crystals that diffract to beyond 3 Å have been grown of the 2C TCR in complex with the class I MHC molecule H-$2K^b$ bound to the self-peptide dEV8 (*26*, *65*). A complete x-ray data set has been collected to 3.4 Å, and the positions of the TCR and pMHC molecules have been determined by molecular replacement (*65*) with the use of the refined crystal structures of the 2C TCR reported here and the 2.5 Å H-$2K^b$-OVA class I MHC crystal structure reported previously (66).

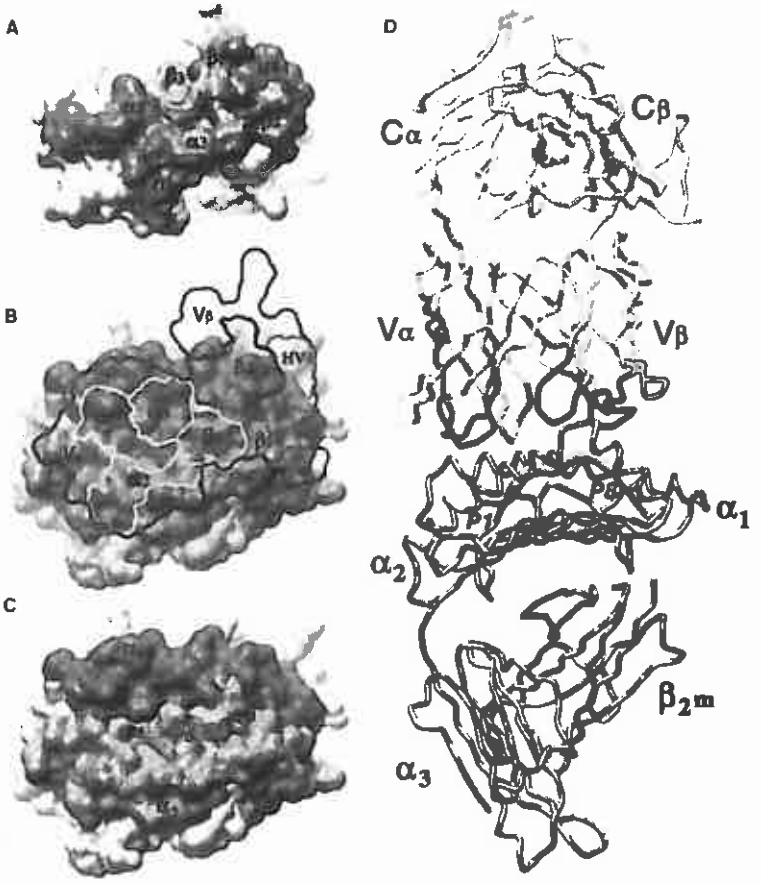


Fig. 9. The TCR-pMHC interaction. (A) Molecular surface of the 2C binding site viewed from above the CDRs. The surface and loop trace of the Vα and Vβ CDRs α1 and α2 are magenta; CDRs β1 and β2, blue; Vα and Vβ CDR3s, yellow; and Vβ HV4, orange. (B) Footprint of the 2C binding site molecular surface on the H-$2K^b$-peptide molecular surface. The outline and footprint of the TCR molecule was obtained by slicing through the TCR surface in the vicinity of the combining site and projecting that surface onto the pMHC molecule. The view is looking directly onto the H-$2K^b$-peptide surfaces as in (C) through the 2C TCR in the complex. The TCR molecular surface is outlined in black and the borders delineating CDR subregions (α1-3, β1-3, and HV4) are indicated. The pink lines separate the molecular surfaces of the CDRs α1 and α2; blue, the CDRs β1 and β2; and yellow, the CDR3s. (C) Molecular surface of the H-$2K^b$-OVA model used to determine the position of the H-$2K^b$-dEV8 in the crystals of the 2C-H-$2K^b$-dEV8 complex. The peptide in the figure still retains the OVA sequence of the search model (*66*). (D) Backbone tube representation of the oriented 2C-H-$2K^b$-peptide complex from the molecular replacement solution. The pMHC is below with the octamer peptide (P1-P8) shown as a large tube in yellow. The 2C TCR is above with the α1 and α2 CDRs colored pink; αHV4, white; CDRs β1 and β2, blue; βHV4, orange; and the CDR3s, yellow.

115

From *Science*, Vol. 274 (11 October 1996): 209–219 DOI: 10.1126/science.274.5285.209 Reprinted with permission from AAAS.

In the 2C H-2K^b-dEV8 complex (Fig. 9), the TCR combining site interacts with the α helices from α_1 and α_2 and the exposed residues of the bound peptide of the pMHC. A footprint of the TCR interaction surface on the pMHC molecule (Fig. 9, A to C) indicates that the long dimensions of the TCR and pMHC binding surfaces are not quite parallel to each other, but tilted about 20° to 30° toward the diagonal. The footprint (Fig. 9B) reveals that Vα CDRs α1 and α2 form a patch covering the peptide NH_2-terminal residues and parts of both pMHC α_1 and α_2 helices. The CDRα1 lies over both the NH_2-terminal region of the pMHC α_1 helix and the three NH_2-terminal residues of the peptide, whereas CDRα2 lies mostly over the COOH-terminal end of the pMHC α_2 helix. Conversely, Vβ CDRβ2 mostly covers the COOH-terminal end of the pMHC α_1 helix with some slight coverage of the COOH-terminal end of the peptide. CDRβ1 is mostly over the NH_2-terminal end of the α_2 helix, but also appears to have some interaction with the COOH-terminal residues of the peptide. The Vα and Vβ HV4s, even though they lie over the pMHC in the footprint (Fig. 9B), do not appear close enough to the pMHC surface (Fig. 9D) to take part in significant interatomic contacts. The Vα and Vβ CDR3s appear to play the prominent role in peptide interaction and straddle the central region of the octameric (P1 to P8) peptide within the pMHC α_1 and α_2 helices, with the P4 position of the peptide lying between the two CDR3s. The footprint of CDRβ3 extends from the P4 position of the peptide all the way to P7. Refinement of the structure should yield further, detailed information (*67*).

The orientation of 2C to H-2K^b–dEV8 resembles the model (*14*) for a TCR-MHC class II complex (D10-IA^K), in which the CDRs α1 and α2 lie over the NH_2-terminal region of the peptide, CDR3α interacts with the central P5 position of the peptide, and CDR3β has interactions extending from P5 to P8. These restrictions would place the CDRs β1 and β2 over the β_1 and α_1 helices, respectively. The 2C-H-2K^b-dEV8 orientation differs from the model (*4*, *12*), in which the TCR is rotated about 90° relative to the pMHC surface. In that model, the CDRs 1 and 2 of both the α and β chain interact predominantly with the α helices of the α and β chains of the pMHC. However, both models (*4*, *12–14*) were derived for class II TCR-MHC systems, and the absence of similar structure-function data for class I MHC interactions with TCR has precluded proposals for a more detailed model until now. In our TCR-pMHC orientation, the TCR and pMHC surfaces have substantial overlap that provides a structural framework for the central role of the Vα and Vβ CDR3s in peptide recognition; CDRs 1 and 2 contact both peptide and the α-helical segments of the MHC molecule.

REFERENCES AND NOTES

1. D. H. Katz, T. Hamaoka, B. Benacerraf, *J. Exp. Med* **137**, 1405 (1973); E. M. Shevach and A. S. Rosenthal, *ibid.* **138**, 1213 (1973); P. C. Doherty, R. V. Blanden, R. M. Zinkernagel, *Transplant Rev* **29**, 89 (1976).
2. H. Clevers, B. Alarcon, T. Wileman, C. A. Terhorst, *Annu. Rev. Immunol.* **6**, 629 (1988).
3. H. Saito *et al.*, *Nature* **312**, 36 (1984) [cloning of 2C α chain]; S. M. Hedrick, D. I. Cohen, E. A. Nielsen, M. M. Davis, *ibid.* **308**, 149 (1984); S. M. Hedrick, E. A. Nielsen, J. Kavaler, D. I. Cohen, M. M. Davis, *ibid.*, p. 153; Y. Yanagi *et al.*, *ibid.*, p. 145; Y. Chen *et al.*, *ibid.* **312**, 31 (1984).
4. M. M. Davis and P. J. Bjorkman, *ibid.* **334**, 395 (1988).
5. S. Tonegawa, *ibid.* **302**, 575 (1981); M. Kronenberg *et al.*, *Annu. Rev. Immunol.* **4**, 529 (1986).
6. P. Patten *et al.*, *Nature* **312**, 40 (1984).
7. D. R. Davies, E. A. Padlan, S. Sheriff, *Annu. Rev. Biochem.* **59**, 439 (1990); I. A. Wilson and R. L. Stanfield, *Curr. Opin. Struct. Biol.* **4**, 857 (1994); I. A. Wilson *et al.*, *Methods Enzymol.* **203**, 153 (1991); I. A. Wilson *et al.*, *Ciba Found. Symp.* **5**, 139, (1991); E. A. Padlan, *Mol. Immunol.* **31**, 169 (1994).
8. J. Novotny, S. Tonegawa, H. Saito, D. M. Kranz, H. N. Eisen, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 742 (1986).
9. C. Chothia, D. R. Bosswell, A. M. Lesk, *EMBO J.* **7**, 3745 (1988).
10. L. M. Amzel and R. J. Poljak, *Annu. Rev. Biochem.* **48**, 961 (1979).
11. I. Engel and S. M. Hedrick, *Cell* **54**, 473 (1988); C. D. Kalayama *et al.*, *EMBO J.* **14**, 927 (1995).
12. J. L. Jorgensen *et al.*, *Nature* **355**, 224 (1992).
13. J.-M. Claverie, A. Prochnika-Chalufour, L. Bougueleret, *Immunol. Today* **10**, 10 (1989).
14. D. B. Sant'Angelo *et al.*, *Immunity* **4**, 367 (1996).
15. G. A. Bentley, G. Boulot, K. Karjalainen, R. A. Mariuzza, *Science* **267**, 1984 (1995).
16. B. A. Fields *et al.*, *ibid.* **270**, 1821 (1995). Coordinates for 1934.4 Vα were provided by R. A. Mariuzza. The Vα domain has an Ig-fold topology, but with an unexpected strand switch in which the c″ strand is H-bonded to the d strand of the outer β sheet, in contrast to the inner c′ strand, as in Ig variable domains.
17. Four distinct Ig structural subtypes have been defined on the basis of the number of strands and the spacing between c′ and e strands: (i) c type, seven-stranded, as in C domains, with ~21 amino acids spacing, (ii) v type, nine-stranded, as in V domains, with ~25 amino acids spacing, (iii) s type, seven-stranded, strand-switched with ~9 amino acids spacing; and (iv) h type, a hybrid between c and s types as described by P. Bork, L. Holm, and C. Sander [*J. Mol. Biol.* **242**, 309 (1994)].
18. I. Engel, T. H. Ottenhoff, R. D. Klausner, *Science* **256**, 1316 (1992); K. L. Hilyard *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9057 (1994); Y. Rotor *et al.*, *Immunity* **2**, 281 (1995); W. F. Soo Hoo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4759 (1993), J. Novotny *et al.*, *ibid.* **88**, 8646 (1991).
19. S. Weber *et al.*, *Nature* **356**, 793 (1992); C. Gregoire *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8077 (1991).
20. A. Y. Lin *et al.*, *Science* **249**, 677 (1990).
21. S. Chung *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12654 (1994).
22. H. C. Chang *et al.*, *ibid.*, p. 11408.
23. R. K. Strong *et al.*, *J. Immunol.* **153**, 4111 (1994).
24. E. S. Ward *et al.*, *Scand. J. Immunol.* **34**, 215 (1991).
25. W. C. Sha *et al.*, *Nature* **335**, 271 (1988).
26. For H-2K^b-dEV8: M. D. Tallquist, T. J. Yan, L. R. Pease, *J. Exp. Med.*, **184**, 1017 (1996); L. Teyton *et al.*, in preparation. For H-2L^d-p2Ca: Y. Sykulev *et al.*, *Immunity* **1**, 15 (1994); M. Corr *et al.*, *Science* **265**, 946 (1994).
27. M. R. Jackson, E. S. Song, Y. Yang, P. A. Peterson, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 12117 (1992); M. Matsumura *et al.*, *J. Biol. Chem.* **267**, 23589 (1992).
28. Carboxypeptidases A and B (Boehringer Mannheim) were used at a ratio of enzyme to TCR of 1:200 (w/w) at 23°C (O/N) in phosphate buffered saline (PBS) (pH 7.2) to remove any free, disordered COOH-termini that might inhibit crystallization. The interchain disulfide (Cys$^{2.13\alpha}$-Cys$^{247\beta}$) provided an absolute stop point for the enzyme. No more than two to four residues remained after the disulfide, as estimated from mass spectrometric molecular mass determination, but no electron density was visible beyond the disulfide. To confirm that the TCR in the crystals is the disulfide-bonded heterodimer, we washed the crystals exhaustively in protein-free mother liquor and analyzed them. In the absence of reducing agent, the dissolved crystals showed on SDS-polyacrylamide gel electrophoresis a homogeneous molecular mass of 58 kD, which on reduction gave two bands of 30 kD (β chain) and 28 kD (α chain). Monomeric α and β chains, or nondisulfide-linked dimers, could not be detected in the crystal. Two-dimensional gel electrophoresis revealed a 1:1 ratio of both the acidic (α) and basic (β) chains. Sequencing of the residues NH_2-terminal confirmed the 1:1 ratio of both chains in the nonreduced 58-kD band. In addition, in BIAcore binding experiments, the expressed 2C material specifically bound both its pMHCs [H-2L^d–p2Ca (*26*) and H-2K^b-dEV8 (*26*)], and the conformation-specific mAbs–clonotypic 1B2 [D. M. Kranz, S. Tonegawa, H. N. Eisen, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7922 (1984)] and Vβ8-specific F23.1 [U. D. Staerz *et al.*, *J. Immunol.* **134**, 3994 (1985)].
29. For other TCRs expressed and purified in our laboratory, elimination or reduction of carbohydrate content has not so far been a successful strategy for improving crystal quality.
30. Although this crystal was initially assigned to the *C*2 space group (with one molecule in the asymmetric unit), further examination of the data integrated as *P*2 (with two molecules in the asymmetric unit) revealed very weak but above background intensities for $h + k = 2n + 1$ reflections, which gradually increased at higher resolutions, an indication of strong pseudo-centering of the cell.
31. X-ray crystallographic programs and methods: (i) X-PLOR: A. T. Brünger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987); A. T. Brünger, *X-PLOR, Version 3.1: A System for X-ray and NMR* (Yale Univ. Press, New Haven, CT, 1992); (ii) *R*-free: A. T. Brünger, *Nature* **355**, 472 (1992); (iii) PHASES: W. J. Furey, A program package and analysis of diffraction data from macromolecules (American Crystallographic Association Fortieth Anniversary Meeting, New Orleans, 1990); (iv) CCP4: Collaborative Computational Project No. 4, Daresbury, UK (1994); *Acta Crystallogr. D* **50**, 760 (1994); (v) O: T. A. Jones, J.-Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Crystallogr. A* **47**, 110 (1991); (vi) SIGMAA: R. J. Read, *ibid.* **42**, 140 (1986); (vii) AMoRe: J. Navaza, *Acta Crystallogr. A* **50**, 157 (1994); (viii) V_m: B. W. Matthews, *J. Mol. Biol.* **33**, 491 (1968); (ix) DENZO and SCALEPACK: Z. Otwinowski, in *Data Collection and Processing*, L. Sawyer, N. Isaacs, S. Bailey, Eds. (SERC Daresbury Laboratory, Warrington, UK, 1993), p. 56; (x) PROCHECK: R. A. Laskowski, M. W. MacArthur, S. D. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993); (xi) 3D-PROFILE: R. Luthy *et al.*, *Nature* **356**, 83 (1992); and (xii) ERRAT: C. Colovos and T. O. Yeates, *Protein Sci.* **2**, 1511 (1993).
32. In the crystal structure of an Fc fragment of an Ig [J. Deisenhofer, *Biochemistry* **20**, 2361 (1981)], regions with disorder and high *B* values are present in both the C_H2 and C_H3 domains. The entire constant region of Fab NC10 was disordered in the crystal structure of the NC10-neuraminadase complex [R. L. Malby *et al.*, *Structure* **2**, 733 (1994)].
33. Cryocooling is not easily reproducible with the 2C crystals. After screening many cryoprotectants, we obtained moderate success by the gradual introduction of glycerol at a rate of 5 percent per hour to a final concentration of 25 percent. Still, this procedure most often resulted in corrupted diffraction patterns (split reflections, loss of resolution, smearing, or high mosaicity). Eventually, after more than 50 attempts with identical soaking and cryoconditions, one crystal was successfully frozen, which diffracted to higher resolution (2.45 Å versus 2.9 Å) with lower mosa-

From *Science*, Vol. 274 (11 October 1996): 209--219 DOI: 10.1126/science.274.5285.209 Reprinted with permission from AAAS.

icity (0.4°) than did room temperature crystals (0.6°).

34. Electron density was traceable as far as the COOH terminal cysteines in both chains; the C_α-C_α distance between the cysteine residues was appropriate for a disulfide bond, but its electron density was weak.
35. The NC41 Fab from an antibody to neuraminidase has a similar elbow angle of 148° [P. M. Colman *et al.*, *Nature* **326**, 358 (1987)].
36. The elbow angle of the TCR, defined as the angle between the pseudo-twofold rotation axes relating the Vα and Vβ domains to the Cα and Cβ domains, was determined with the programs OVRLAP [M. G. Rossman and P. Argos, *J. Biol. Chem.* **250**, 7525 (1975)] and ELBOW (D. H. Fremont and T. O. Yeates, personal communication). To determine the positions of the pseudo-twofold axes, we superimposed the α chain domains with the use of a pairwise method (*51*) onto the corresponding β chain domains with the same conserved C_α atoms that were used for overlaps of the individual TCR domains with antibodies (*56*).
37. J. H. Brown *et al.*, *Nature* **364**, 33 (1994).
38. K. C. Garcia *et al.*, in preparation.
39. All TCR and Fab numbering in this article follow E. A. Kabat *et al.*, Eds., *Sequences of Proteins of Immunological Interest* (National Institutes of Health, Bethesda, MD, ed. 5, 1991).
40. G. A. Bentley and R. A. Mariuzza, *Annu. Rev. Immunol.* **14**, 563 (1996).
41. With DALI, version 2.0, L. Holm and C. J. Sander, *J. Mol. Biol.* **233**, 123 (1993).
42. Cβ residues 218 to 220 are involved in extensive van der Waals contacts with a symmetry-related molecule in the 2C crystal lattice.
43. *Drosophila melanogaster* cells are insect cells that correctly recognize N-linked glycosylation signals, Asn-X-Thr (Ser) [M. J. Fraser, *In Vitro Cell. Dev. Biol.* **25**, 225 (1989)]. However, the N-linked oligosaccharide structures produced are usually smaller than those produced in mammalian cells [K. Kuroda *et al.*, *Virology* **174**, 418 (1990)]. The fucose moiety has an α1,6 linkage to the first GlcNAc that is directly connected to the side-chain amide nitrogen of the Asn-X-Ser (Thr) sequence. The carbohydrate content was estimated by mass spectrometry of glycosylated 2C (non-carboxypeptidase digested) and unglycosylated 2C (non-carboxypeptidase digested) prepared by inducing expression in the presence of tunicamycin (1 μg/ml). The glycosylated 2C was 58,829 daltons and the unglycosylated 2C was 53,965 daltons. On the assumption that each N-linked glycosylation site has 1 to 1.5 kD of carbohydrate, then only four or five possible N-linked sites are occupied. Alternatively, N-linked glycosylation may not have been completely inhibited.
44. The average *B* values indicate that the Cα domain has higher thermal parameters (Table 1). The average *B* value for the Cα top strands (c, f, and g) is 69 Å^2, and for the bottom strands (a, b, e, and d) is 41 Å^2. However, the density is clear and unambiguous (Fig. 1). The Cα domain has no crystal lattice packing interactions that probably contribute to the higher thermal parameters.
45. The immunoglobulin disulfide has a χ_2 of ~180° [J. S. Richardson, *Adv. Prot. Chem.* **34**, 167 (1981)].
46. The program DALI (*41*) was used to search for structurally similar proteins or domains among 642 structures in the Protein Data Bank. The highest *Z* scores (4.0 to 5.0, which are very low values) were obtained with Ig fold–containing molecules based on alignments with the back β sheet (and part of strand c) of the Cα (2.8 to 3.2 Å rms deviation over 68 to 71 residues). No reasonable alignments were found in which the top strands were included in the superposition.
47. K. P. Kearse *et al.*, *EMBO J.* **13**, 4504 (1994).
48. C. Gerler *et al.*, *J. Immunol.* **148**, 3469 (1992); S. Caspar-Bauguil *et al.*, *Scand. J. Immunol.* **40**, 323 (1994).
49. All molecular surface areas buried by interaction were calculated with the program MS [M. L. Connolly, *J. Appl. Crystallogr.* **16**, 439 (1983)] with a 1.7 Å probe sphere and standard atomic radii, as described in (*7*).
50. To define the β sheet crossing angle, vectors were fit to the Cα atoms in structurally equivalent, but opposing, β strands in the interface of each of the interacting domains. The crossing angle was then the angle between these two vectors. The structurally equivalent residues chosen in opposing β strands were as follows: (i) Vα(30 to 38, 103 to 108)–Vβ(32 to 39, 105 to 112); (ii) Cα(138 to 145, 175 to 185)–Cβ(143 to 151, 190 to 198); (iii) C_H3 (363 to 370, 403 to 411)–C_H3(363 to 370, 403 to 411); (iv) C_L (132 to 138, 173 to 180)–C_H1 (140 to 145, 186 to 192); (v) V_L (30 to 38, 103 to 108)–V_H(32 to 39, 105 to 112); and (vi) Vα (31 to 37, 103 to 108)–Vα (31 to 37, 103 to 108).
51. For pairwise comparisons, an α chain TCR domain was first superimposed on the corresponding Fab L chain domain (or C_L3 domain). The same transformation was then applied to the β chain domains to superimpose on the H chain (or C_H3) domain. The additional rotation and translation required to optimize the β chain–H chain superposition describes the domain shift. This procedure is similar to that described in (*35*) with superpositions specified in (*56*).
52. Hydrogen bond interactions were based on both distance (3.5 Å cutoff) and geometrical considerations with HBPLUS [I. K. McDonald and J. M. Thornton, *J. Mol. Biol.* **238**, 777 (1994)]. The van der Waals and salt-bridge interactions were calculated with the program CONTACSYM [S. Sheriff, W. A. Hendrickson, J. L. Smith, *ibid.* **197**, 273 (1987)].
53. A pairwise superposition of all V_H-V_L heterodimers from available Fab x-ray structures in the Protein Data Bank [F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977)] with the Vα-Vβ of 2C shows a range from an essentially identical domain-domain rotational relationship with many antibodies to up to 17° differences with others.
54. R. L. Stanfield *et al.*, *Structure* **1**, 83 (1993).
55. P. M. Colman, *Adv. Immunol.* **43**, 99 (1988).
56. The six Fabs used for comparison represent a range of isotypes. The Fabs were KOL [M. Marquart *et al.*, *J. Mol. Biol.* **141**, 369 (1980)]; 3D6 [X. M. He *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7154 (1992)]; Se155-4 [M. Cygler *et al.*, *Science* **253**, 442 (1991)]; BV04 [J. N. Heron *et al.*, *Prot. Struct. Func. Genet.* **11**, 159 (1991)]; J539 [T. N. Bhat *et al.*, *ibid.* **1**, 74 (1986)]; 17/9 [J. M. Rini *et al.*, *Science* **255**, 959 (1992)]. For calculating rms deviations for the TCR and Fabs, the models were superimposed by means of OVRLAP (*36*). Conserved, structurally equivalent β-sheet framework residues in both TCR and Fabs were chosen for the superposition. For 2C Vα, residues 3 to 5, 19 to 24, 32 to 37, 44 to 47, 63 to 65, 72 to 76, 88 to 91, and 110 to 112 were superimposed on Fab V_L residues 4 to 6, 20 to 25, 33 to 38, 45 to 48, 63 to 65, 70 to 74, 86 to 89, and 102 to 104, and an rms deviation was calculated for the resulting 77 structurally equivalent framework residues. For 2C Vβ, residues 4 to 6, 20 to 25, 30 to 31, 34 to 37, 44 to 47, 66 to 68, 74 to 76, 90 to 93, and 112 to 114 were superimposed on V_H residues 4 to 6, 19 to 24, 34 to 35, 36 to 39, 46 to 49, 68 to 70, 77 to 81, 90 to 93, and 107 to 109, and the rms deviation was calculated for 70 structurally equivalent framework residues. For 2C Cβ, framework residues 126 to 129, 144 to 149, 158, 160 to 161, 172 to 174, 192 to 197, and 212 to 215 were superimposed on C_L1 residues 120 to 123, 139 to 144, 153, 156 to 157, 171 to 173, 187 to 192, and 208 to 211. For 2C Cα, 29 framework residues in the back β sheet were chosen, 123 to 127, 129, 139 to 143, 145, 147, 159 to 162, 164, 179 to 185, 190, 204, 206 to 207 for overlap with C_L residues 113 to 117, 119, 132 to 136, 138, 140, 158 to 161, 163, 174 to 180, 193, 205, and 207 to 208. The Fabs in Fig. 7A were superimposed with 2C as described above and were F9.13.7 [J. Lescar, H. Souchon, P. Alzari, *Protein Sci.* **3**, 788 (1994)]; D11.15 [V. Chitarra *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7711 (1993)]; and CNJ206 [R. Zemel *et al.*, *Mol. Immunol.* **31**, 127 (1994)].
57. Each domain of 2C was submitted to the DALI (*41*) site. The average rms deviation for overlapping about 100 residues of each TCR V domain with other antibody V domains (L and H) of Fabs was ~2.0 Å for Vβ (*Z* values of 14 and 15) and 1.6 Å for Vα (*Z* values of 14 and 15), with a slight skewing toward lower rms deviations of Vα with V_L domains.
58. Y. Choi *et al.*, *Nature* **346**, 471 (1990) [Vβ interaction with superantigen]; A. V. Chernovsky *et al.*, The 9th International Congress of Immunology, Abstract 2955 [Vα interaction with the accessory CD4].
59. Despite the high degree of flexibility expected for a tetraglycine stretch of amino acids in a solvent exposed loop, the electron density for CDR3 is excellent. The φ and ψ angles of CDR3β fall within accepted ranges for a classic gamma-type turn [J. S. Richardson, *Adv. Protein Chem.* **34**, 167 (1981)].
60. C. Abergal and J. M. Claverie, *Eur. J. Immunol.* **21**, 3021 (1991).
61. C. Chothia and A. M. Lesk, *J. Mol. Biol.* **196**, 901 (1987); C. Chothia *et al.*, *Nature* **342**, 877 (1989).
62. M. Schiffer, E. A. Kabat, T. T. Wu, *Immunogenetics* **35**, 224 (1992).
63. Length was measured from the tip of CDR1α to CDR2β and width was measured from the furthest ends of CDR1α to CDR2α or CDR1β to CDR2β.
64. A proposal for a TCR pocket that specifically binds a particular side chain is made for the class II I-E^k system in [G. J. Kersh and P. M. Allen, *J. Exp. Med.*, in press].
65. The 2C-H-2K^b-dEV8 complex was prepared by mixing a 1:1 stoichiometric ratio of carboxypeptidase-treated and HIC-purified 2C with *D. melanogaster*–expressed H-2K^b in the presence of a 1.5-fold excess of the dEV8 peptide (EQYKFYSV) (*26*). Crystals were grown by vapor diffusion in the presence of 10 percent PEG-6000, 0.2M tris-acetate (pH 7.2). Repeated macroseeding resulted in large crystals that could be cooled to 100K by the slow addition of ethylene glycol to a concentration of 22 percent. Although diffraction from these crystals now extends to beyond 3 Å, an initial data set was collected to 3.4 Å on a MAR imaging plate at the UCSD X-ray facility. The data were indexed, integrated, and scaled with DENZO and SCALEPACK (*31*). The space group was determined to be $P2_12_12_1$ with cell dimensions a = 296.7 Å, b = 89.73 Å, and c = 84.5 Å. The overall R_{sym} is 10.5 percent on I's with a completeness of 85 percent to 3.4 Å (80 percent completeness and I/σ of 4.5 in the outer 3.4 to 3.5 Å shell). The 2C MR solution was determined with normalized structure factors (E's) in AMoRe (*31*). The 2C coordinates were then placed in the cell and the rotation and translation functions were determined with H-2K^b–OVA (*66*) as a model. One round of rigid body refinement that was followed by positional refinement in X-PLOR (*31*) lowered the *R* value to 34.5 percent. Inspection of electron density maps at this point indicated clear electron density for both 2C and H-2K^b-dEV8.
66. D. H. Fremont, E. A. Stura, M. Matsumura, P. A. Peterson, I. A. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2479 (1995).
67. K. C. Garcia, M. Degano, P. A. Peterson, L. Teyton, I. A. Wilson, unpublished data.
68. O. Epp *et al.*, *Eur. J. Biochem.* **45**, 513 (1974).
69. Graphics Programs: (i) The program TOM was compiled by Christian Cambillau (1987); (ii) AVS, C. Upson *et al.*, *IEEE Computer Graphics and Application* **9**, 30 (1989); (iii) MidasPlus 2.0, T. E. Ferrin, C. C. Huang, L. E. Jarvis, R. J. Langridge, *J. Mol. Graph.* **6**, 13 (1988); and (iv) PQMS, M. L. Connolly, *ibid.* **11**, 139 (1993).

70. We thank E. Stura for advice and help with data collection and heavy atom screening; T. Yeates, L. Pease, D. Kranz, H. Eisen, C. Scott, and R. Stellanko for discussions, materials, and assistance; N. H. Xuong for use of the UCSD X-ray facility; R. Lerner for constant encouragement and support, and M. Pique for production of figures. Supported by NIH RO1 CA58896 (I.A.W.) and NIH postdoctoral training grant T32 A107244 (K.C.G.). This is publication 10295-M from The Scripps Research Institute. The coordinates for 2C have been deposited in the Protein Data Bank (Brookhaven National Laboratory, Upton, NY) with accession code 1TCR.

23 August 1996; accepted 15 September 1996

117

Reprinted by permission from Macmillan Publishers Ltd: *Nature* 384: 134–141, copyright 1996.

Structure of the complex between human T-cell receptor, viral peptide and HLA-A2

David N. Garboczi*, Partho Ghosh*, Ursula Utz†, Qing R. Fan*, William E. Biddison‡ & Don C. Wiley*§

* Department of Molecular and Cellular Biology, § Howard Hughes Medical Institute, Harvard University, 7 Divinity Avenue, Cambridge, Massachussetts 02138, USA
† Laboratoire d'Immunologie, Institut de Recherches Cliniques de Montréal, 110 Avenue des Pins Ouest, Montréal H2W 1R7, Canada
‡ Molecular Immunology Section, Neuroimmunology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA

Recognition by a T-cell antigen receptor (TCR) of peptide complexed with a major histocompatibility complex (MHC) molecule occurs through variable loops in the TCR structure which bury almost all the available peptide and a much larger area of the MHC molecule. The TCR fits diagonally across the MHC peptide-binding site in a surface feature common to all class I and class II MHC molecules, providing evidence that the nature of binding is general. A broadly applicable binding mode has implications for the mechanism of repertoire selection and the magnitude of alloreactions.

THE specificity of the T-cell response is determined by recognition of an antigenic peptide bound to a class I or class II MHC glycoprotein by the antigen-specific T-cell receptor (TCR). The TCR interaction initiates signals in T cells that, together with non-antigen-specific signals, trigger development of the T-cell repertoire, regulation of the immune response and activation of cytolytic T cells (CTL) (reviewed in ref. 1). X-ray crystal structures of several complexes of peptides with both class I and class II MHC molecules revealed how peptides are bound in a groove between two α-helices on the membrane-distal surface of MHC molecules (reviewed in refs 2, 3). MHC/peptide complexes present a compound surface composed of atoms of the peptide and the MHC molecule, providing a molecular mechanism for MHC-restricted antigen recognition by TCR[4–8].

Like antibody molecules, αβTCRs are encoded by a family of gene segments that rearrange to generate the diverse TCR repertoire, estimated to contain 10^{14} different sequences[9]. The first two variable loops (CDR1 and CDR2) of the human α- and β-chain are encoded by the 42 Vα[10] and 46 Vβ genes[11]; the third variable loops (CDR3) are encoded following the joining of *V* genes to one of a number of *J*-region and, in the case of Vβ, *D* and *J*-region gene segments. The joining reactions themselves generate further diversity both by removing nucleotides and by introducing non-germline bases ('N') at each junction.

Although the diversity-generating mechanism is very similar for antibodies and TCR, there are many more TCR *J* segments and N additions[9], and apparently less variation at CDR1 and CDR2[12]. These genetic differences suggest that TCRs differ structurally from antibody fragments (Fabs) by having less difference in overall shape as a result of reduced diversity at variable loops 1 and 2, but more variation, due to increased loop-3 diversity, in the shape of the centre of the combining site, where in Fabs the third loops converge to form a pocket. Models of the interaction of TCR with MHC molecules have been proposed that orient the diverse CDR3 loops over the (diverse) peptides and the less diverse CDR1 and CDR2 loops over the correspondingly less diverse MHC molecules[9,13–16].

We have crystallized a ternary complex containing a soluble human αβTCR (A6)[17] specific for the Tax peptide of the human T-cell lymphotropic virus HTLV-1, with that peptide bound to the human MHC molecule HLA-A2 (ref. 18). The TCR and HLA-A2 were both expressed as individual polypeptide chains in bacteria, refolded, and shown to retain the binding specificity of the molecules *in vivo*[18,19]. The X-ray structure of the HLA-A2/Tax peptide has been determined previously and predicts peptide positions that may be recognized directly by TCR[20]. HTLV-1 is associated with an adult T-cell leukaemia and a slowly progressive neurological disease known as HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP)[21]. A 40-fold higher frequency of HTLV-1-specific CTL, most of which are specific for the Tax peptide (LLFGYPVYV), is found in HLA-A2$^+$ HAM/TSP patients relative to asymptomatic HTLV-1-infected individuals[22].

The 2.6 Å structure of the TCR/Tax/HLA-A2 complex shows that the TCR is oriented diagonally over the MHC/peptide complex, with CDR1 and CDR3 from both Vα and Vβ contacting the peptide. This orientation appears similar to that in the preliminary

www.jimmunol.org

Reprinted by permission from Macmillan Publishers Ltd: *Nature* 384: 134–141, copyright 1996.

TABLE 1 Structural analysis

Data sets	Resolution limit (Å)	Mosaicity	R_{merge}†	Comp‡ (red)	I/σ_I	R_{derv}† (sites)	R_{cullis}†	F_H/e§	F.o.m.‖ (res.)
*C*2 (*Z* = 1)*									0.41 (3.5 Å)
Native	2.90	2.5°	13.9	82 (3.7)	7.3				
L1C + EMP	3.10	1.0°	10.1	98 (2.5)	8.2	22.4 (1)	75	1.13	
K91C + EMP	3.20	2.4°	8.6	81 (2.7)	9.2	21.6 (1)	89	0.79	
Y67C/K91C + EMP	2.60	1.1°	10.9	95 (3.1)	7.0	27.2 (2)	85	0.94	
*C*2 (*Z* = 1)*									0.35 (3.5 Å)
Native	2.80	1.5°	11.8	79 (1.8)	6.4				
L1C + EMP	3.40	0.6°	10.7	75 (1.8)	6.3	23.5 (2)	76	0.94	
K_2PtCl_4	3.60	1.4°	11.8	93 (2.5)	7.0	34.5 (5)	82	0.92	
$P2_1$ (*Z* = 2)*									0.30 (3.7 Å)
L1C/11C native	3.75	3.0°	14.3	86 (2.6)	6.6				
L1C/11C + EMP	3.40	2.5°	13.9	85 (2.6)	6.5	35.4 (2)	80	1.12	

Molecular replacements¶	HLA-A2/Tax9	TCR Vα	TCR Vβ	TCR Cβ
Rotation (10–3.8 Å)	7.3σ	2.3σ	4.3σ	3.4σ
Translation (15–3.8 Å)	10.3σ	8.5σ	10.3σ	5.7σ

Refinement resolution	Reflections (free)	Atoms	R_{free}†	R_{cryst}†	Bonds#	Angles#	*B*-factors**
6.0–2.6 Å	26,273 (3,006)	5,716	32.2	24.0	0.013 Å	1.81°	4.8 Å²

Domains	$\alpha_1\alpha_2$	α_3	β_2m	Tax9	Vα	Vβ	Cβ
Mean *B*-factors	27 ± 15 Å²	81 ± 32 Å²	29 ± 16 Å²	19 ± 7 Å²	36 ± 15 Å²	36 ± 15 Å²	65 ± 18 Å²
Real-space fit††	0.87 ± 0.03	0.74 ± 0.15	0.86 ± 0.09	0.89 ± 0.02	0.87 ± 0.05	0.88 ± 0.03	0.80 ± 0.12

* Space groups of crystal forms and derivatives are described in Methods. *Z*, number of MHC/Tax/TCR complexes per asymmetric unit.

† $R_{merge} = 100 \times \frac{\sum_h \sum_i |I_{h,i} - I_h|}{\sum_h \sum_i I_{h,i}}$, where I_h is the mean intensity of symmetry-related reflections, $I_{h,i}$.

$R_{derv} = 100 \times \frac{\sum_{hkl} |F_p - F_{ph}|}{\sum_{hkl} F_p}$, where F_p is the native amplitude and F_{ph} is the derivative amplitude. Sites, number of heavy atom sites per asymmetric unit.

$R_{cullis} = 100 \times \frac{\sum_{hkl} ||F_p \pm F_{ph}| - F_{hcalc}|}{\sum_{hkl} |F_p \pm F_{ph}|}$, where F_p is the native amplitude, F_{ph} is the derivative amplitude for centric reflections, and F_{hcalc} is the the calculated heavy-atom structure factor.

$R_{free} = 100 \times \frac{\sum_{hkl} |F_{obs} - F_{calc}|}{\sum_{hkl} F_{obs}}$, where R_{free} is calculated for a randomly chosen 10% of reflections ($F > 0$) omitted from refinement, and R_{cryst} is calculated for the remaining 90% of reflections ($F > 0$) included in refinement.

‡ Comp = 100 × (number of observed unique reflections)/(number of possible unique reflections); red = (number of observed reflections)/(number of observed unique reflections).

§ F_H/e is the phasing power, with F_H being the mean amplitude of the heavy-atom structure factor and e the r.m.s. lack-of-closure error.

‖ F.o.m., figure of merit; res., resolution to which figure of merit is determined.

¶ Rotation and translation solution peak heights above the mean in rotation and translation function maps.

r.m.s. deviations from ideal values.

** r.m.s. deviation between *B*-factors of bonded atoms.

†† Real-space fit correlation coefficient[48] calculated from SIGMAA-weighted[44] $2F_{obs} - F_{calc}$ electron density and the refined model of MHC/Tax/TCR.

description of a mouse TCR (2C)/peptide/H-2K^b complex[11]. All three variable loops of Vα and CDR3 of Vβ contact conserved and polymorphic positions on the α-helices of the MHC molecule. Evidence is presented that the observed binding mode is general for human and mouse class I and class II molecules. The implications of the size and shape of the TCR/peptide/MHC interaction for the specificity of TCRs and mechanisms of immunological tolerance are discussed.

Overview of the intercellular complex

The crystal structure was determined using isomorphous replacement, molecular replacement, iterative real-space averaging between crystal forms and around a non-crystallographic symmetry axis, and density modification (see Methods). All of the crystals exhibited large mosaic spreads, affecting the quality of data that could be collected (Table 1). The interface between TCR and MHC/peptide is in unambiguous electron density, as are the α1, α2, and β2-microglobulin (β2 m) domains of the MHC and the Vα and Vβ domains of the TCR. Regions distal to the interface between the MHC and TCR appear to be mobile and disordered. (Note the increase in the magnitude of temperature factors with increasing distance from the interface; Fig. 1 and Table 1.) Parts of the HLA-A2 α3 domain and some loops of TCR Cβ are unclear (see Methods). Although several strands and side chains of TCR Cα near Cβ are visible, most of Cα is disordered. The Cα domain is not included in the final model, with little effect on the crystallographic residual (Methods and Table 1).

Figure 1 shows the TCR/Tax/HLA-A2 complex and how the TCR (top) and MHC molecule (bottom) would span the space between a CTL and virally infected target cell.

Orientation of TCR variable loops

When the peptide-binding site of HLA-A2 is viewed from above, the TCR is oriented diagonally across the site (Fig. 2*a*, *b*). The TCR third variable loops of both Vα and Vβ meet at the centre of the binding site over the centre peptide residue Y5 (Fig. 2*a*). The

119

Reprinted by permission from Macmillan Publishers Ltd: *Nature* 384: 134–141, copyright 1996.

ARTICLES

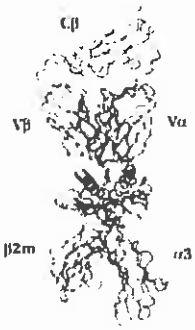
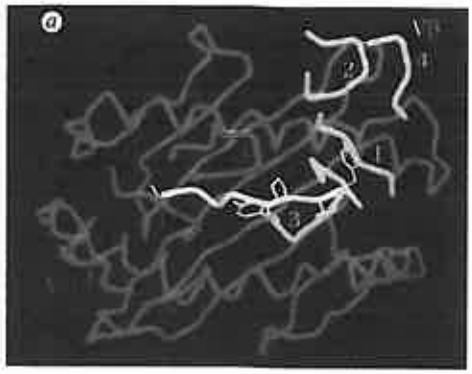
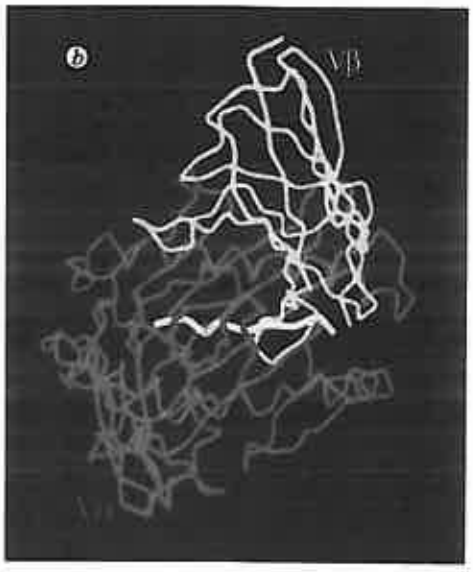


FIG. 1 Intercellular complex of A6 αβTCR (top) and the HTLV-1 Tax peptide bound to HLA-A2 (bottom) shown as an α-carbon trace, coloured by temperature factor[46]: blue is low-temperature factor, and red is high-temperature factor (>60 Å^2) and shows less ordered regions. The TCR Cα domain is not shown (see Methods).





CDR1 loops are also positioned over the peptide. CDR1α extends from the peptide N terminus to Y5, and CDR1β is over the C-terminal end of the peptide near Y8 (Fig. 2*a*). The second variable loops are over the MHC molecule, CDR2α over the α2 domain α-helix and CDR2β over the α1 domain α-helix.

Although the location of the CDR3 loops over the peptide with CDR3β toward the C terminus (Y8) and CDR3α near the centre of the peptide (Y5) is consistent with mutation data[15,16,23,25], the diagonal orientation of the TCR over the binding site is not as originally predicted[9,13,14,24]. Nathenson and colleagues[25], testing mutant mouse class I molecules against a panel of CTL, provided evidence for a diagonal orientation for TCR binding to H2-K^b like that observed here (although some of their mutations are too distant from the Tax TCR to be contacted and may have exerted their effects through distortions to the peptide). Subsequent mutation data on the class II molecule I-A^k (ref. 23), as well as the preliminary description of the mouse TCR (2C)/peptide/H-2K^b structure[31], support that orientation.

Almost the entire Tax peptide is buried in the TCR/MHC interface (Fig. 2*c*). The TCR makes contacts with peptide residues L1, L2 and from G4 to Y8. Y5 is bound in a deep pocket at the centre of the TCR where the CDR3 loops converge (Fig. 2*d*, *f*), reminiscent of the structures of peptides complexed with anti-peptide Fabs (reviewed in ref. 26). The peptide is bound much more deeply in the MHC molecule (Fig. 2*e*) than it is in the TCR (Fig. 2*d*).

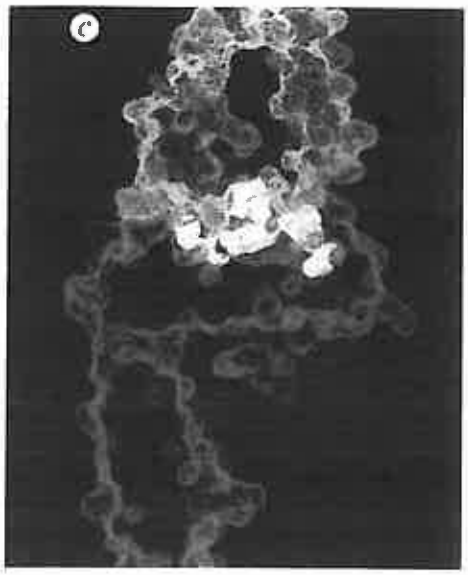
The diagonal orientation of the TCR allows the flat surface of the TCR (Fig. 2*f*) to interact with the peptide by fitting down between the two highest points on the MHC molecule, toward the N-terminal of the α1 domain α-helix (top left, Fig. 2*g*) and towards the N terminal of the α2 domain α-helix (bottom right, Fig. 2*g*), a feature of the diagonal orientation also recognized by Nathenson and colleagues[25]. The high points of the α2 domain α-helix (right, front, Fig. 2*e*) and the α1 domain α-helix (left, back, Fig. 2*e*) are a topographical feature common to all class I and class II molecules.

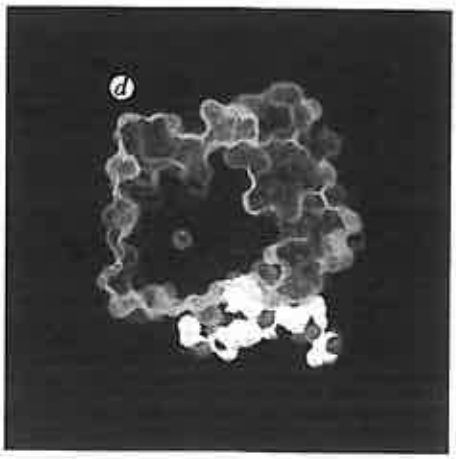
Interaction with Tax and HLA-A2

Variable loops 1 and 3 of both Vα and Vβ contact the Tax peptide, whereas loops 1, 2 and 3 of Vα and loop 3 of Vβ contact the HLA-A2 α-helices (Fig. 3*a*, Table 2). The loop with the highest potential diversity, CDRβ3, extends across the site contacting the α1 domain α-helix at residue 72, peptide residues Y5, P6, V7 and Y8, and ends by contacting the α2 domain α-helix at residues 149, 150, 151 and 155 (Fig. 3*a*). Residues 95 and 101–103 of this loop form part of the central pocket in the TCR that surrounds Y5 and the central residues of the peptide (Fig. 3*b*). CDR3α forms most of the other side of this central pocket, burying residues G4 and Y5 and the α1 domain α-helix at residues 65, 66, 68 and 69 (Fig. 3*a*). Serine 31 from CDR1α completes the central pocket, forming a hydrogen bond to the hydroxyl of peptide tyrosine, Y5. CDR1α is positioned over the N terminus of the peptide and contacts the α-helices of both the α1 and α2 domains (Fig. 3*a*), making a footprint comparable to that made by CDR3α (Fig. 3*c*). One residue of CDR1β, E30, contacts the peptide, forming a hydrogen bond to the Y8 hydroxyl group at the opposite end of the site. CDR2β, although positioned over the α1 domain, does not make any contacts; CDR2α over the α2 domain contacts helix residues 155, 158 and 166 (Fig. 3*a*).

A total of 998 Å^2 of solvent-accessible surface area of HLA-A2/Tax is buried in the interface with TCR, including parts of five polymorphic positions on HLA-A2 (open circles in Fig. 3*a*) that could provide allelic specificity, and eleven positions conserved in human class I molecules (solid circles in Fig. 3*a*) that could provide a common binding site for TCR. Of 20 hydrogen bonds between TCR and MHC/peptide, eight are to conserved and three to polymorphic class I MHC residues, and nine are to Tax peptide. About 326 Å^2 of peptide surface is buried by the TCR, two-thirds of the area buried in the average anti-peptide antibody interface (420–620 Å^2)[28]. The surface of HLA-A2 buried by the TCR is 671 Å^2, at the low end of the range for protein antigen/antibody interfaces (620–899 Å^2)[27]. Overall, the buried surface of the peptide/MHC (998 Å^2) and TCR (1,011 Å^2) is only slightly larger than typical antigen/antibody interfaces[27]. Over half of the

Reprinted by permission from Macmillan Publishers Ltd: *Nature* 384: 134–141, copyright 1996.





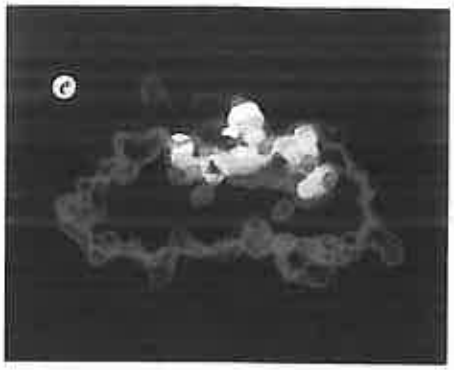
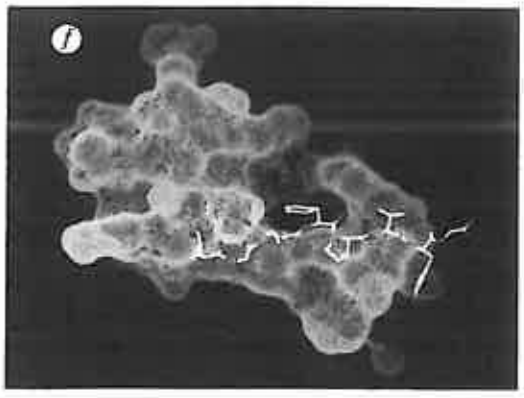
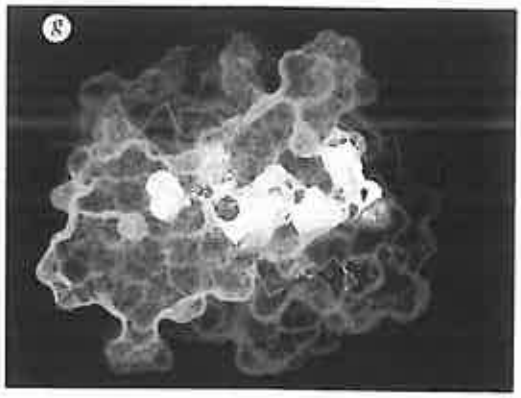


FIG. 2 Orientation of TCR on HLA-A2/Tax. *a*, Variable loops of TCR Vα (red) and Vβ (yellow), Tax peptide (white), and HLA-A2 (blue) shown as an α-carbon trace. Also shown are side chains of the peptide contacted by TCR. *b*, TCR Vα (red) and Vβ (yellow) domains bound to Tax/HLA-A2. *c*, Tax peptide is buried between the TCR and HLA-A2. The solvent-accessible surface (dot) of TCR Vα and Vβ (yellow), HLA-A2 (blue) and Tax peptide (CPK model) is shown. Surfaces were calculated with a 1.4 Å probe radius[50]. *d*, Interactions of bound Tax peptide with TCR Vα and Vβ. *e*, Interactions of bound Tax peptide with HLA-A2. *f*, Flat TCR surface with central pocket. Bound Tax peptide is shown with Y5 in the pocket and Y8 at the edge of the flat surface. *g*, TCR surface (yellow) in contact, diagonally, with HLA-A2 (blue) and Tax peptide. Figures were made using the program MIDAS[50].





121

Reprinted by permission from Macmillan Publishers Ltd: *Nature* 384: 134–141, copyright 1996.

ARTICLES

peptide surface buried by the TCR in the interface is covered by the most diverse loop, CDR3β. CDR1α and CDR3α each contribute about 25%. Over half of the MHC surface buried by TCR is also covered by CDR3 loops and the remainder is covered about equally by CDR1α and CDR2α (Fig. 3*c*).

Only one residue of loop 4, (HV4) of the TCR, α68K, contacts the MHC molecule, forming hydrogen bonds to T163 and E166, an acidic side chain conserved in mouse and human class I (corresponding to a conserved histidine in class II) (Fig. 3*a*).

Changes in interface conformation

A number of small adjustments of MHC and peptide main-chain and side-chain positions are found by comparing the Tax/HLA-A2 structure[20] to the TCR/Tax/HLA-A2 structure. The peptide appears to be pushed down further into the MHC binding site by the TCR; proline at position 6 (P6) is pressed down, and the side chain of valine 7 has moved up (Fig. 3*d*). If P6 had not moved it would have overlapped the positions of TCR residue 100 of CDR3α and residue 98 of CDR3β. MHC side chains appear to have moved to achieve close fits in the interface, and a small but significant shift in the MHC main chain is evident (Fig. 3*d*).

The structure of the uncomplexed A6 TCR is not known, but comparison of the TCR in this complex with the coordinates of the free TCRβ chain[25] and the TCRVα dimer[26] indicates that the overall conformations of the CDR1 and CDR2 loops are essentially the same in these bound and unbound TCRs. The more

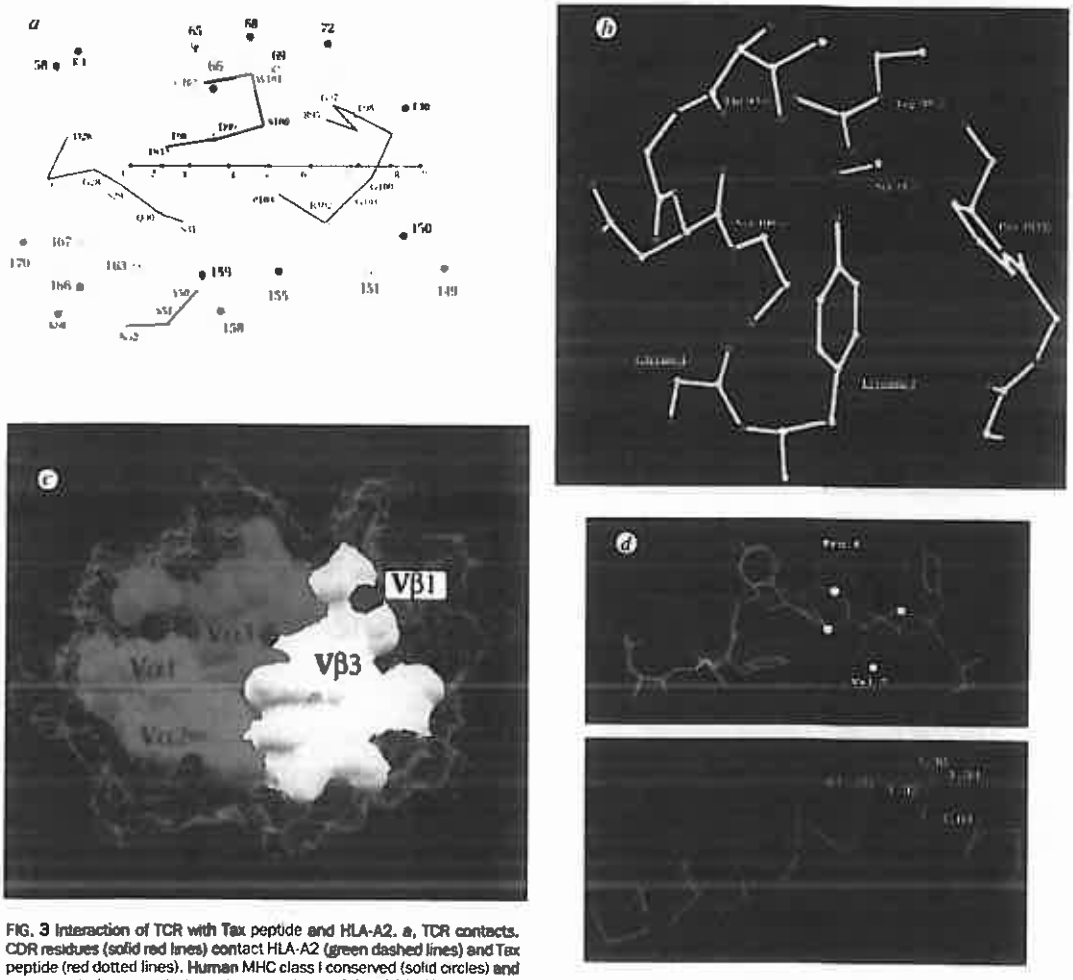


FIG. 3 Interaction of TCR with Tax peptide and HLA-A2. *a*, TCR contacts. CDR residues (solid red lines) contact HLA-A2 (green dashed lines) and Tax peptide (red dotted lines). Human MHC class I conserved (solid circles) and polymorphic (open circles) residues are shown. D26–S31, CDR1α; Y50–N52, CDR2α; T93–G102, CDR3α; E30, CDR1β; R95–P103, CDR3β. TCR (Vα2.3[AV2S1A2]–Jα24–Cα, Vβ12.3[BV13S1]–Dβ2.1–Jβ2.7–Cβ2)[15] (numbered according to ref. 51). CDR3 residues contacting MHC and peptide are encoded by N addition (α93) and J segment for Vα and N addition (β95, β102, β103) and D segment for Vβ. For clarity, the peptide is depicted as a straight line and contact between Vα30Q and HLA-A2 66 is not shown. *b*, TCR pocket around Gly 4 and Tyr 5. Figure was made using RIBBONS software[52]. *c*, MHC/peptide solvent-accessible surface buried by the CDRs of the TCR, coloured according to contact region. *d*, Conformational differences of peptide (top) and MHC α2 domain helix (bottom) in the TCR/Tax/MHC structure (yellow) and the Tax/MHC structure (blue). The Pro 6 Cα (dot) has moved 2.7 Å and the Val 7 Cγ (square) 4.6 Å. Helix displacements as measured between α-carbons (α2 148–151) are shown in ångströms (*a* and *c* were produced with MIDAS[50], D with O[46]).

Reprinted by permission from Macmillan Publishers Ltd: *Nature* 384: 134–141, copyright 1996.

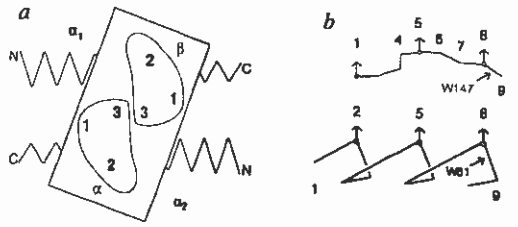


FIG. 4 Diagonal TCR binding would cover peptides bound to class I and class II MHC molecules. *a*, TCR αβ and CDR loops (numbered) shown diagonally between the N-terminal peaks of the helical regions of MHC molecules. *b*, Side-chain locations of bound peptides (side view). Peptides bind to class I molecules as extended chains with a kink (top) and to class II MHC molecules as an extended polyproline type-II conformation, with peptide side chains projecting every 120° along the chain (triangles, bottom). In both classes of MHC molecule, the carbonyl oxygen of the peptide residue 8 is hydrogen-bonded to a tryptophan (W) residue conserved in all class I (top) and class II (bottom) molecules at a homologous location of the lower helix in *a*[40,53].

variable CDR3 loops have different conformations, as predicted from Fab structures.

TCR structure compared with Fab

The overall structure of the TCR in the TCR/Tax/HLA-A2 complex is similar to the structure of a Fab. The Vα–Vβ interface buries about 1,575 Å^2 (~780 Å^2 from each domain), forming a contact surface that is among the largest observed in Fabs[30]. This suggests that the particular VαVβ combination in the A6 TCR makes a strong V–V contact that is unlikely to have rearranged upon binding antigen, like the induced-fit seen in a Fab with a small V–V interface (1,063 Å^2)[30], and discussed with respect to the small V–V interface in 2C TCR[31]. The disorder of the Cα domain in crystals of the complex precludes detailed assessment of the Cα–Cβ interface, although it is clear from the electron density that it comprises a large area of residues (see Methods). The Cα domain determined in a mouse TCR (2C) has an unconventional fold[31]. The absence of the TCR carbohydrate that is found at the Cα–Cβ interface and of the interchain disulphide bond[31] may contribute to the disorder observed.

The 'elbow' angle' between Vβ and Cβ is approximately the same in the TCR/Tax/HLA-A2 complex as in the free β-chain structure[28], the 2C TCR[31], and within the range found in Fabs[32]. The Vβ–Cβ interface in the complex buries about 800 Å^2, similar to the free β-chain[28] and 2C TCR[31] structures, and about twice the area of this interface in Fab structures[32].

TCR signalling

No gross conformational change of the TCR was observed in the complex relative to the structures of isolated TCR domains or Fabs which might indicate that it could send information to the cytoplasm that it had bound antigen. Indeed, no overall conformational change has ever been observed in Fabs that correlates with antigen binding, even when the V domains rotate up to 16° as a result of induced fit in the combining site[30]. The high-temperature factors in Cα of an uncomplexed TCR[31] suggest that flexibility is a feature of this area of isolated, soluble αβTCR, consistent with the TCR/Tax/HLA-A2 structure. It may be necessary to include CD3 and ζ-chains in crystalline complexes to be able to observe relevant conformational changes, if they exist.

The current (and preliminary) absence of structural evidence for a TCR conformational change in TCR/peptide/MHC that might initiate signalling, favours signalling mechanisms based on the assembly of complexes following TCR engagement[33], or on segregation of engaged TCRs in cell–cell attachment patches[34,35].

Specificity and tolerance

The small size of the contact between the TCR and the Tax peptide, 326 Å^2, and the observation that although substantial contacts are made to peptide residues Y5 and Y8, only a few atoms of peptide residues 1, 2, 4, 6 and 7 are in contact, places physical limits on TCR specificity for peptide. For example, when Y5, which is bound in the central pocket of the TCR, is substituted to alanine, the peptide still forms A6 TCR/peptide/HLA-A2 complexes that crystallize and is a partial agonist[17]. On the basis of the TCR/Tax/HLA-A2 structure, it seems likely that other peptide sequences will be able to satisfy the twin requirements of binding to HLA-A2 and to the A6 TCR. How many of those sequences might be expressed and presented in humans? (see ref. 36 for example). Limited TCR specificity would favour tolerance mechanisms that depend on the context of antigen presentation, including other receptors and adhesion molecules, to restrict the potential for cross-reaction[37,38].

Generality of the binding mode?

The diagonal binding observed in the TCR/Tax/HLA-A2 complex appears to result from properties shared by all classical class I and class II MHC molecules and probably all αβTCRs. The A6 and the 2C[31] αβTCRs, like Fabs that bind protein antigens[37], have a relatively flat contact surface formed by the CDR loops with a central pocket where the CDR3 loops converge (Fig. 2*f*). The top surface of class I and class II MHC molecules is not flat, but has two 'peaks' near the N termini of the α-helical regions (top left and bottom right in Fig. 4*a*); peptides bind to MHC molecules low in the groove between these α-helices (Fig. 2*e*). For the fairly flat surface of the TCR to fit at a low enough point on the MHC surface to contact the length of exposed peptide, a diagonal orientation is required. Either significant translations along the peptide binding site or rotations from the observed orientation would break this overall shape complementarity. Contacts between TCR and the α-helical 'peaks' are observed (Fig. 3*a*), but result from residues at the edges and sides of the flat TCR surface.

The peptide residues most exposed on class II MHC complexes, positions 2, 5 and 8, align closely with positions 1, 5 and 8 which the TCR covers in the Tax peptide bound to HLA-A2, despite the differences in the conformations of peptides bound to the two types of MHC molecules (see Fig. 4 of ref. 3) (Fig. 4*b*). The TCR binding mode observed here and with mouse TCR 2C[31] covers the length of the peptide with the central TCR pocket positioned over the middle of the peptide. It would also cover a peptide on a class II molecule from near pocket 1 to 9[40].

The suggestion that a diagonal orientation allows a general 'interlocking' binding mode between TCR and MHC was first proposed on the basis of mutations on mouse class I molecules[25]. Mutation data from class II mouse molecules support the same mode of binding[23], as does the preliminary description of a mouse TCR/peptide/H-2K^b complex[31]. In addition to the structural arguments presented here, functional arguments have been made in favour of a general binding mode (refs 23, 41 for example). For example, it would provide a molecular mechanism for an inherent bias of TCR for self-MHC[42] and, coupled with the physical limit seen on TCR specificity for peptide, helps to explain the ability of one peptide positively to select a nearly normal TCR repertoire[41,43]. A general TCR binding mode would simplify the processes of selection during repertoire development and provide

123

Reprinted by permission from Macmillan Publishers Ltd: *Nature* 384: 134–141, copyright 1996.

ARTICLES

TABLE 2 Contacts between TCR and MHC peptide

	α1K	N	*	HC58E	OE2
CDR1α	α26D	OD2	*	HC58E	OE1
	α27R	CG		HC167W	CZ2
		NE		HC170R	NH1
	α30Q	NE2		HC66K	NZ
		NE2		Tax1L	CD2
		NE2	*	Tax2L	O
		OE1	*	Tax4G	N
		OE1		HC159Y	CD1
		CG		HC163T	OG1
	α31S	OG	*	Tax5Y	OH
CDR2α	α50Y	CG		HC155Q	OE1
		O		HC158A	CB
	α51S	OG		HC158A	CB
	α52N	ND2	*	HC166E	OE2
HV4α	α68K	NZ	*	HC166E	OE1
		NZ	*	HC163T	OG1
CDR3α	α93T	OG1	*	Tax5Y	OH
	α98T	OG1	*	HC65R	NH2
	α99D	OD2	*	HC65R	NE
		CB		HC66K	CG
		CA		Tax4G	O
		O		Tax5Y	OH
	α100S	N	*	Tax4G	O
		OG	*	Tax4G	O
		CA		Tax5Y	CE1
	α101W	CZ3		HC69A	CA
		CZ3		HC68K	C
		O		HC65R	CD
	α102G	CA		HC65R	NE
CDR1β	β30E	OE2	*	Tax8Y	OH
CDR3β	β95R	NH2	*	Tax5Y	OH
	β98L	CD1		Tax6P	O
		CB		Tax8Y	CZ
		O	*	Tax8Y	N
		O		Tax7V	CA
		CD2		HC72Q	CG
	β100G	C		Tax7V	CG2
	β101G	N	*	HC150A	O
		O	*	HC155Q	NE2
		N		Tax7V	CG2
	β102R	NH1	*	HC149A	O
		NE	*	HC150A	O
		CD		HC151H	CD2
	β103P	CB		Tax5Y	OH
		CD		HC155Q	NE2

Asterisks denote hydrogen bonds (2.5–3.4 Å acceptor–donor distance). Unstarred residues, van der Waals contacts (<4.0 Å); only closest two atoms tabulated. α and β, TCR; HC, conserved class I MHC heavy-chain residues; underlined HC, polymorphic heavy-chain residues.

an explanation for the magnitude of the alloreactive response. One way that the danger inherent in hypervariable molecules may be controlled is by restricting their opportunity for cross-reactivity to a single binding mode on MHC molecules, which could be part of the basis of mechanisms that maintain immunological tolerance.

Note added in proof: K. C. Garcia and I. A. Wilson kindly provided the coordinates of the free, uncomplexed 2C TCR after this manuscript was submitted, permitting the disorder of the A6 Cα domain to be confirmed. Including in refinement the 2C Cα domain fitted to the A6 TCR/Tax/HLA-A2 electron density gave no improvement in the R_{free}, even at low resolution. □

Methods

Data collection and processing. Crystals with space groups C2 (a = 229.3 Å, b = 49.5 Å, c = 96.0 Å, β = 89.6°)[18] containing one complex per asymmetric unit and $P2_1$ (a = 229.9 Å, b = 49.5 Å, c = 96.0 Å, β = 88.9°) with two complexes per asymmetric unit, were soaked in 14% PEG 8000, 100 mM $Mg(CH_3COO)_2$, 50 mM Tris, pH 8.5, containing 20% glycerol for 2 min, mounted in a fibre loop, and flash-cooled in liquid N_2. X-ray data were collected on an ADSC 1K CCD detector (1°–2° oscillations, CHESS A1 beamline, λ = 0.9 Å, 100 K). All crystals exhibited large mosaic spread (~1.5°) and anisotropic diffraction (Table 1). Data were processed with DENZO, SCALEPACK (Z. Otwinowski, personal communication), and TRUNCATE[44]. Two forms of the C2 crystal (A and B) exist with identical cell constants but slightly different molecular orientations (R_{anom} = 40%; average 2° and 3 Å difference in some domains).

Molecular replacement. The position of HLA-A2/Tax peptide[20] was located with AMORE[44] in crystal A. TCR 1934.4 Vα[29] and TCR 14.3.d β-chains[28], with side chains differing from A6 TCR made alanine and Vα CDRs and the Cβ 131–142 loop omitted, were also located (R_{cryst} = 43%, 6–3.5 Å). With HLA-A2/Tax peptide fixed, the Vβ, Cβ and Vα positions were determined successively, each using the preceding domains. After density modification by solvent flattening or flipping and histogram matching[44] using envelopes calculated with ENVAT (D. Madden, personal communication), side chains unique to A6 TCR were identifiable but as the interface between HLA-A2/Tax and TCR was ambiguous, heavy-atom derivatives were used.

Multiple isomorphous replacement. Crystals with cysteine substitutions in the Tax peptide, L1C and L1C/11C (containing L1C and the C-terminal additions G10 and C11), and in β_2m (K91C and double mutant Y67C/K91C) were soaked in ethyl mercury phosphate (EMP) and yielded derivatives. L1C/11C crystallized in space group $P2_1$, breaking the centring of the C2 group, resulting in two complexes per asymmetric unit. K_2PtCl_4 was used previously for solving HLA-A2 (ref. 4). Crystals were soaked for 4–24 h in 1 mM compound. For mercurials, crystals were presoaked in buffer containing 1–2 mM dithiothreitol for 4 h. The mercury positions in L1C in form A and form B crystals were found by difference-Patterson syntheses. Other heavy atoms were located by difference Fourier syntheses using molecular replacement (MR) phases (Table 1). Heavy-atom parameters were refined with VECREF and MLPHARE[44] and re-refined with protein phases after density modification.

Iterative real-space averaging. Averaging between form A and form B crystals and fourfold averaging among form A, form B, and the two complexes in the $P2_1$ crystal were carried out to 3.5 Å with MAVE[45] or DMMULTI[44]. Seven-domain envelopes (for α1α2/Tax peptide, α3, β_2m, Vα, Cα, Vβ and Cβ) were constructed with ENVAT and transformations for each were derived from MR or from electron density correlations. Electron density from MIR phased maps, MIR/MR phase-combined maps, MR phased or MIR/model phase-combined maps were iteratively averaged. The electron density after averaging allowed an unambiguous definition of the CDR loops and TCR to MHC/peptide contacts.

Model building and refinement. The mercury derivative of Y67C/K91C, which diffracted best, was used for building[46] and refinement (Table 1) essentially as described[47] with XPLOR[48] and REFMAC[44], using a maximum-likelihood target in the latter case. An anisotropic B-factor tensor (B_{11} = −6 Å^2, and $B_{22} = B_{33}$ = 6 Å^2) determined from Wilson plots and a bulk solvent correction were applied[44,48]; data from 12 to 2.6 Å were included. Simulated annealing omit maps[48] were used extensively. The α3 domain of HLA-A2, which forms no crystal contacts, is disordered. Refinement at low resolution (12–6 Å) with and without α3, indicated that α3 makes little coherent contribution to the diffraction. As the α3 domain has previously been refined to high resolution[49], the model was maintained in refinement, although some regions (190–202, 217–227, 246–259, 267–268) have discontinuous density and high temperature factors. The Cα domain of TCR also appears to contribute little to coherent scattering even at low resolution. It has not been modelled, although several strands and side chains proximal to Cβ are visible. The model contains: residues 1–274 of heavy chain and 0–99 of β2m of HLA-A2; 1–9 of Tax peptide; 1–122 of TCRα; 3–246 of TCRβ, except for three loops (131–144, 180–189, 220–229); two ethyl mercury groups and 37 waters. One water coordinates TCR residues at the interface (Vα 101, Vβ 30, 50, 98), and three waters contact Tax peptide. Breaks in the main chain are in the Vβ–Cβ connection (117–118) and in Cβ (199–201, 207–208). Side chains of 47 residues have been omitted from later stages of refinement (mainly in α3) because they appear to be disordered. Only one, TCR α1 (lysine) is near the interface (Fig. 3a). No residues have disallowed ϕ, ψ angles; 86% of residues have favoured angles.

Received 16 October; accepted 28 October 1996.

1. Hedrick, S. M. & Eisdelman, F. J. in *Fundamental Immunology* (ed. Paul, W. E.) 363–420 (Raven, New York, 1993).
2. Madden, D. R. *Annu. Rev. Immunol.* **13**, 587–622 (1995).
3. Stern, L. J. & Wiley, D. C. *Structure* **2**, 245–251 (1994).
4. Bjorkman, P. J. *et al. Nature* **329**, 506–512 (1987).
5. Bjorkman, P. J. *et al. Nature* **329**, 512–518 (1987).
6. Shevach, E. M. & Rosenthal, A. S. *J. Exp. Med.* **138**, 1213–1229 (1973).
7. Katz, D. H., Hamaoka, T., Dorf, M. E. & Benacerraf, B. *Proc. Natl Acad. Sci. USA* **70**, 2624–2628 (1973).
8. Zinkernagel, R. M. & Doherty, P. C. *Nature* **248**, 701–702 (1974).
9. Davis, M. M. & Bjorkman, P. J. *Nature* **334**, 395–402 (1988).
10. Arden, B., Clark, S. P., Kabelitz, D. & Mak, T. *Immunogenetics* **42**, 455–500 (1995).
11. Rowen, L., Koop, B. F. & Hood, L. *Science* **272**, 1755–1762 (1996).
12. Bentley, H. A. & Manuzza, R. A. *Annu. Rev. Immunol.* **14**, 563–590 (1996).
13. Chothia, C., Boswell, D. R. & Lesk, A. M. *EMBO J.* **7**, 3745–3755 (1988).

Reprinted by permission from Macmillan Publishers Ltd: *Nature* 384: 134–141, copyright 1996.

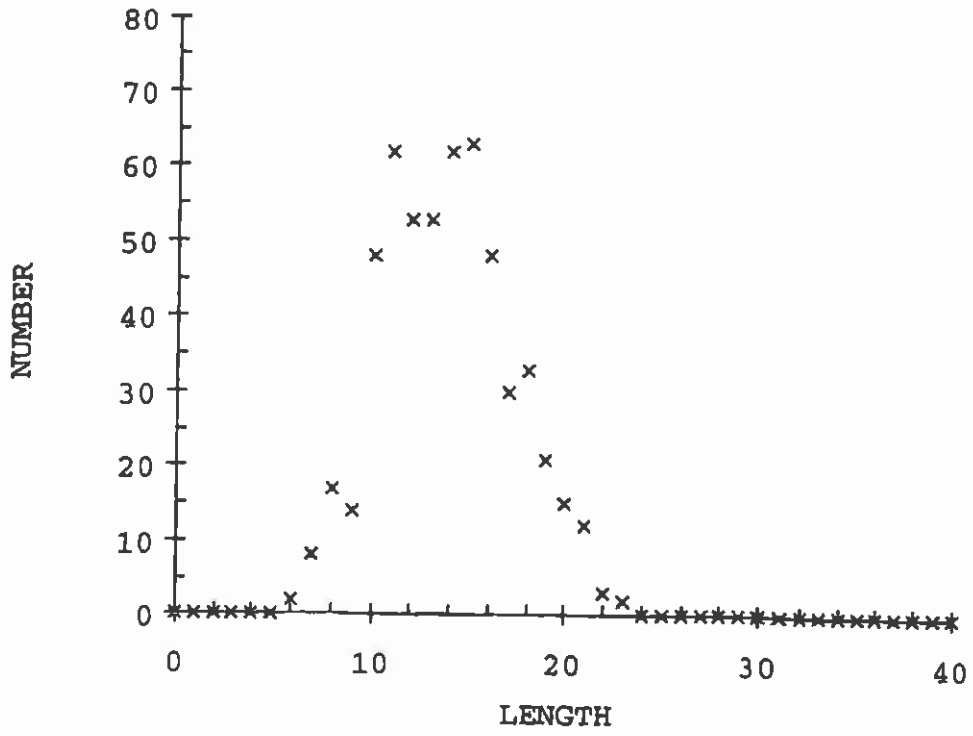
14. Clavene, J. M., Prochnicka Chalufour, A. & Bougueleret, L. *Immunol. Today* **10,** 10–14 (1989).
15. Jorgensen, J. L., Reay, P. A., Ehrich, E. W. & Davis, M. M. *Annu. Rev. Immunol.* **10,** 835–873 (1992).
16. Jorgensen, J. L., Esser, U., Fazekas de St Groth, B., Reay, P. A. & Davis, M. M. *Nature* **355,** 224–230 (1992).
17. Utz, U., Banks, D., Jacobson, S. & Biddison, W. E. *J. Virol.* **70,** 843–851 (1996).
18. Garboczi, D. N. *et al. J. Immunol.* (in the press).
19. Garboczi, D. N., Hung, D. T. & Wiley, D. C. *Proc. Natl Acad. Sci. USA* **89,** 3429–3433 (1992).
20. Madden, D. R., Garboczi, D. N. & Wiley, D. C. *Cell* **75,** 693–708 (1993).
21. Roman, G. C. & Osame, M. *Lancet* **(i),** 651 (1988).
22. **Elovaara, I.** *et al. J. Exp. Med.* **177,** 1567–1573 (1993).
23. Sent'Angelo, D. B. *et al. Immunity* **4,** 367–376 (1996).
24. Vasmatzis, G., Cornett, J., Sezerman, U. & DeLisi, C. *J. Mol. Biol* **261,** 72–89 (1996).
25. Sun, R. *et al. Immunity* **3,** 573–582 (1995).
26. Wilson, I. A., **Ghara, J. B. & Stanfield, R. L.** *A Structural View of Immune Recognition by Antibodies.* 55th Forum in Immunology, 73–78 (1996).
27. **Wilson,** I. A. & Stanfield, R. L. *Structure* **2,** 113–118 (1993).
28. Bentley, G. A., Boulot, G., Karjalainen, K. & Mariuzza, R. A. *Science* **267,** 1984–1987 (1995).
29. Fields, B. A. *et al. Science* **270,** 1821–1824 (1995).
30. Stanfield, R. L., **Takimoto-Kamimura,** M., Rini, J. M., Profy, A. T. & **Wilson, I. A.** *Structure* **1,** 83–93 (1993).
31. **Garcia, K.** C. *et al. Science* **274,** 209–219 (1996).
32. Padlan, E. A. *Mol. Immunol.* **31,** 169–217 (1994).
33. **Brown,** J. H. *et al. Nature* **364,** 33–39 (1993).
34. Janeway, C. A. Jr *et al. Cold Spring Harbor Symp. Quant. Biol* **LIV,** 657–666 (1989).
35. Matsui, K. *et al. Science* **254,** 1788–1791 (1991).
36. Wucherpfennig, K. W. & Strominger, J. L. *Cell* **80,** 695–705 (1995).
37. Janeway, C. A. Jr *Immunol. Today* **13,** 11–16 (1992).
38. Matzinger, P. *Annu. Rev. Immunol* **12,** 991–1045 (1994).
39. Webster, D. M., **Henry, A. H. & Reese,** A. R. *Curr. Opin. Struct. Biol* **4,** 123–129 (1994).
40. Stern, L. J. *et al. Nature* **368,** 215–221 (1994).
41. Ignatowicz, L., Kappler, J. & Marrack, P. *Cell* **84, 521–529 (1996).**
42. Jerne, N. *Eur. J. Immunol.* **1,** 1–9 (1971).
43. Miyazaki, T. *et al. Cell* **84,** 531–541 (1996).
44. Collaborative Computational Project No. 4. *Acta Crystallogr. D* **50, 760–763** (1994).
45. Kleywegt, G. J. & Jones, T. A. *From First Map to Final Model* (CCP4), 59–66 (1994).
46. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. *Acta Crystallogr. A* **47,** 110–119 (1991).
47. Ghosh, P., Amaya, M., Mellins, E. & Wiley, D. C. *Nature* **378,** 457–462 (1995).
48. Brünger, A. T. *Nature* **355,** 472–475 (1992).
49. Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. *Cell* **70,** 1035–1048 (1992).
50. Huang, C. C., Pettersen, E. F., Klein, T. E., Ferrin, T. E. & Langridge, R. *J. Mol. Graph.* **9,** 230–236 (1991).
51. Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S. & Foeller, C. *Sequences of Proteins of Immunological Interest* (US Department of Health and Human Services, PHS, NIH, Bethesda, 1991).
52. Carson, M. *J. Appl. Crystallogr.* **24,** 958–961 (1991).
53. Brown, J. H. *et al. Nature* **332,** 845–850 (1988).

ACKNOWLEDGEMENTS. D.N.G. and P.G. contributed equally to this work. We thank A. Haykov, N. Smitskaya and R. Crouse for technical assistance, S. C. Harrison for discussion, R. A. Mariuzza for Va 1934.4 coordinates, K. C. Garcia and I. A. Wilson for making their manuscript available before publication, members of our laboratory and the synchrotron staff for help at CHESS, and D. Takemoto for help with a figure. D.N.G. is supported by a fellowship from the Department of the Army; no official endorsement by the government should be inferred. P.G. was supported by the Irvington Institute for Medical Research and the HHMI. D.C.W. is an investigator of the HHMI, which supported this research.

CORRESPONDENCE and requests for materials should be addressed to D.C.W. Coordinates will be deposited in the Protein Data Bank (Brookhaven National Laboratory, Upton, New York) and are available before their release by e-mail (ghosh@crystal.harvard.edu).

125

There is another heterodimer of TCR other than α / β, known as γ / δ, with the genes of δ chains overlapping with those of the α chains. The latter pair can function in the same fashion as the former pair. In addition, the γ / δ TCR can also bind antigens directly. This process is designated as MHC un-restricted, and mimics the binding of antigens by antibodies. The consequence of this activity renders the CDRδ3 to become similar to CDRH3. The length distribution of CDRδ3 becomes much broader showing a bimodal pattern as shown below:



Length distribution of human TCR δ chain CDR3's.

Genes of antibodies and TCR's

The gene arrangements of antibodies and TCR's are quite complicated. However, each of them can be identified by searching the gene bank database at http://www.ncbi.nlm.nih.gov . Their constant region genes are located 3' to the variable region genes.

For the light chain, i.e. kappa and lambda, of antibodies and α and γ chains of TCR's, the variable region genes, i.e. the V gene and J minigene, can have two patterns:

1. $(V)_n$ $(J)_m$
2. $(V\ J)_n$

For the heavy chain of antibodies and β and δ chains of TCR's, the variable region genes, i.e. the V gene, D minigene and J minigene, can have several patterns:

1. $(V)_n$ $(D)_m$ $(J)_p$
2. $(V)_n$ $(D\ J)_m$
3. $(V\ D\ J)_n$

In humans, they are located on different chromosomes:

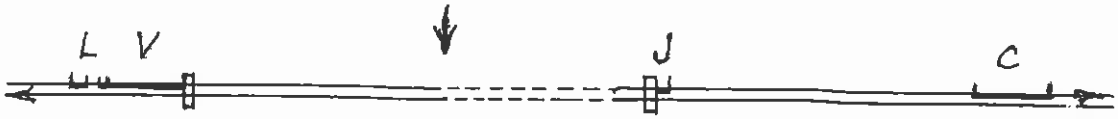
Antibody kappa light chains	2p11
Antibody lambda light chain	22q11
Antibody heavy chains	14q32
TCR α chains	14q11.2
TCR β chains	7q34
TCR γ chains	7p14
TCR δ chains	14q11.2

As mentioned before, the gene segments for the TCR α and δ chains are not only on the same chromosome but also overlap with each other. They are on the same chromosome as antibody heavy chains but at different locations. TCR β and γ chains are on the same chromosome but on two different arms.

Production of human lambda light chain

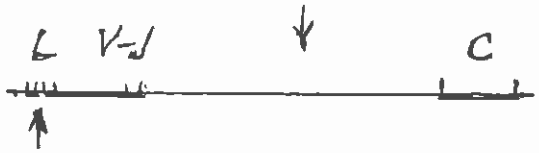
The central dogma states DNA to RNA to protein. In the case of antibodies and TCR's, the scheme is slightly more complicated. The situation of human lambda light chains provides a simple example. There are six stages as illustrated below:

1. Genomic DNA

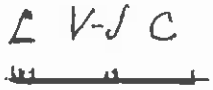


2. Re-arranged DNA

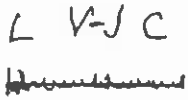
3. Pre-mRNA



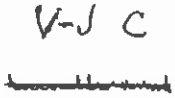
4. Mature mRNA



5. Precursor protein



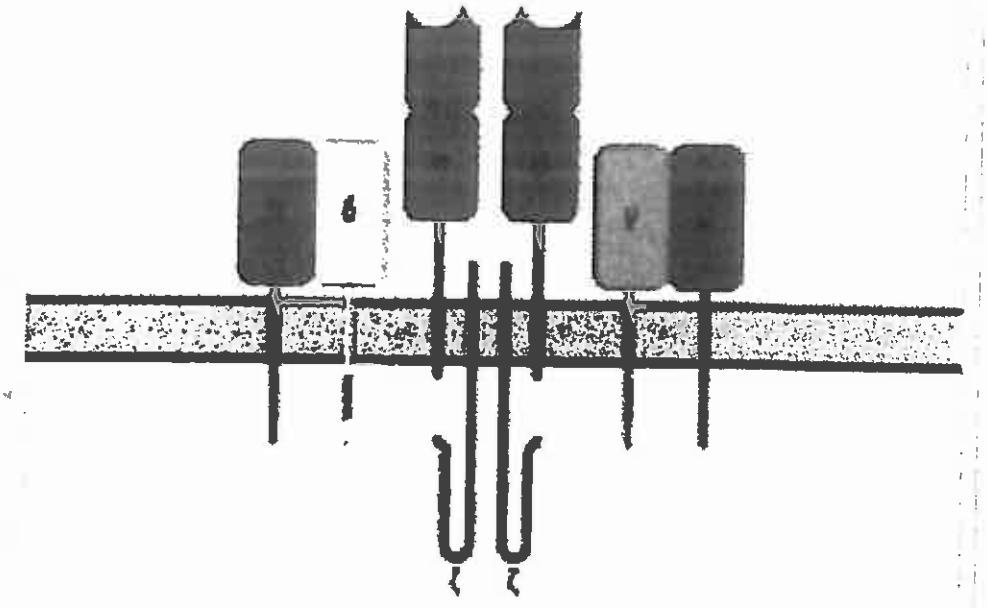
6. Mature protein



Signaling between cells by CD3

While the binding of antibodies to antigens involving specific contacting surfaces between two macro-molecules, the binding of TCR's to the MHC-processed peptide complex require more complicated interactions to bring two different cells together for biological signal transduction. As mentioned before, most of the processed peptides isolated from cell surface MHC class I or II molecules are derived from self-proteins. One of the first problems is how the tiny amount of foreign processed peptide on antigen presenting cell surfaces in the MHC grooves can be recognized. Presumably some TCR's will have very high binding affinities for the foreign peptide complex. Why? Various speculations have been proposed (van der Merwe PA, Dushek O (2011) Mechanisms for T cell receptor triggering. Nature Rev, 11, 47-55). They suggested that three different mechanisms, aggregation, conformational change, and segregation, might be involved.

In addition, TCR's are not isolated membrane bound molecules. They are closely associated with the CD3 molecules, forming the T cell receptor for antigen complex. Indeed, the TCR α and β chain cytoplasmic regions are very short and most likely cannot serve the function of signal transduction. Schematically, that complex has the following structure:



The *T-cell receptor complex* with TCR-α and TCR-β chains (top), ζ-chain accessory molecules (bottom) and *CD3* (represented by CD3γ, CD3δ and two CD3ε).

Mechanisms for T cell receptor triggering

P. Anton van der Merwe and Omer Dushek

Abstract | There is considerable controversy about the mechanism of T cell receptor (TCR) triggering, the process by which the TCR tranduces signals across the plasma membrane after binding to its ligand (an agonist peptide complexed with an MHC molecule). Three main types of mechanism have been proposed, which involve aggregation, conformational change and segregation. Here, we review recently published evidence for each type of mechanism and conclude that all three may be involved. This complexity may reflect the uniquely demanding nature of TCR-mediated antigen recognition, which requires the detection of a very weak 'signal' (very rare foreign peptide–MHC ligands) in the presence of considerable 'noise' (abundant self peptide–MHC molecules).

Immunoreceptor tyrosine-based activation motif
(ITAM) A sequence that is present in the cytoplasmic domains of the invariant chains of various cell-surface immune receptors, such as the T cell receptor–CD3 complex. Following phosphorylation of their tyrosine residue, ITAMs function as docking sites for Src homology 2 (SH2) domain-containing tyrosine kinases and adaptor molecules, thereby facilitating intracellular signalling cascades.

Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, UK
Correspondence to P.A.v.d.M.
e-mail: anton.vandermerwe@path.ox.ac.uk
doi:10.1038/nri2887
Published online
3 December 2010

The T cell receptor (TCR) consists of a variable TCRαβ heterodimer that binds to ligands. The TCR forms a multisubunit receptor complex with the non-variable signal transduction CD3 complex, which contains CD3γ, CD3δ, CD3ε and TCRζ subunits. All CD3 complex subunits contain immunoreceptor tyrosine-based activation motifs (ITAMs) in their cytoplasmic domains. TCRαβ binds complexes of peptide and MHC molecules on the surface of antigen-presenting cells (APCs) or target cells, which results in biochemical changes in the cytoplasmic portions of the CD3 complex. One biochemical change that is known to be essential for TCR signalling is phosphorylation of the ITAMs in the CD3 complex by the Src family tyrosine kinases LCK and FYN (reviewed in REFS 1,2). In addition, various conformational changes have been reported, some of which are independent of tyrosine phosphorylation[3–5], and some of which are postulated to precede, and be required for, ITAM phosphorylation[6,7]. The process by which TCR binding to peptide–MHC molecules leads to biochemical changes in the cytoplasmic regions of the CD3 complex is referred to as TCR triggering and is the main focus of this Review. Although this area has been reviewed previously, there have been considerable advances made in the past 2 years, and this Review focuses on this new research.

The challenges

While evaluating possible models of TCR triggering, it is important to bear in mind some of the unusual features of TCR antigen recognition, which distinguish it from other cell-surface receptor recognition events. These features clarify the challenges faced by any triggering mechanism and place constraints on the types of model that are mechanistically plausible.

Sensitivity. The first notable feature is the very low abundance of agonist peptide–MHC ligands on APCs or target cells. The TCR on a given T cell is 'restricted' to a subset of the MHC molecules on an APC, and the high-affinity agonist (typically 'foreign') peptide will only be present on a tiny fraction of these molecules. Hence TCR triggering needs to be very sensitive, and indeed TCRs can be triggered when only a single antigenic peptide–MHC ligand is within the contact area[8–10].

Discrimination. A second feature is the presence on APCs and target cells of abundant self peptide–MHC molecules to which the TCR can also bind. To develop into mature peripheral T cells, developing T cells need to undergo positive selection whereby only T cells expressing TCRs that bind self peptide–MHC molecules can survive. Therefore, all peripheral T cells express TCRs that bind self peptide–MHC molecules. Indeed, there is evidence that continued recognition of self peptide–MHC molecules is required for survival of peripheral T cells[11–13]. Although negative selection of developing T cells ensures that the TCRs do not have a high affinity for self peptide–MHC molecules, the affinity threshold for negative selection is sharp and close to the affinity threshold for foreign peptide–MHC ligand recognition[14]. As a result, any given TCR is likely to bind a fraction of self peptide–MHC molecules with affinities that are not much lower than the affinities for foreign peptide–MHC ligands. Furthermore, the self

© 2011 Macmillan Publishers Limited. All rights reserved

Serial-triggering model
A model that was proposed to account for the observation that small numbers of agonist peptide–MHC complexes seemed to trigger large numbers of T cell receptors (TCRs), and that postulates that a given peptide–MHC complex can serially bind to and trigger multiple TCRs. As it is the number of productive TCR engagements that determines peptide–MHC efficacy, high-affinity peptide–MHC complexes with long half-lives may be less effective. According to this model there is an optimal affinity or half-life for a TCR–peptide–MHC complex

peptide–MHC molecules that a given TCR can bind are collectively likely to be far more abundant than foreign agonist peptide–MHC ligands, only a few of which will be recognized by a specific TCR. Because T cells are sensitive to only a small number of recognition events, the stochastic nature of the molecular events involved creates a serious problem of discrimination (BOX 1): how do T cells detect very low numbers of foreign peptide–MHC ligands (signal) in the presence of high levels of self peptide–MHC molecules (noise)?

Molecular mechanisms that have been proposed to explain this remarkable ability to discriminate signal from noise either extend the time available for recognition or invoke cooperative effects, whereby individual TCRs somehow communicate with each other following ligand recognition. Given the remarkable speed of antigen recognition by T cells[15], only cooperative effects can satisfactorily account for discrimination[16]. This cooperation could either involve a direct physical interaction, which would require the TCR–CD3 complexes to form aggregates[17,18], or communication via signalling pathways, a process termed signal spreading[16,19,20].

Versatility. A third requirement of TCR recognition is that it is necessary for the same TCR to recognize multiple ligands with a range of affinities and produce different responses depending on the affinity. For example, in the thymus, developing T cells need to recognize both low-affinity and high-affinity self peptide–MHC molecules and respond differently to each[21]. Similarly, mature T cells require recognition of low-affinity self peptide–MHC molecules to promote survival, whereas recognition of high-affinity foreign peptide–MHC ligands leads to activation and proliferation. Any triggering mechanism or mechanisms must allow recognition of a range of different ligands and produce distinct signals depending on the strength of binding.

Structural diversity. A fourth unusual feature of TCR recognition is the diverse nature of the binding interface between the TCR and a peptide–MHC molecule, and the fact that TCRs need to recognize ligands (foreign peptide) that they have never previously encountered. Although there are clearly common general features in TCR–peptide–MHC complex interactions, the variability of peptide sequences and TCR complementarity-determining region 3 (CDR3) loops results in great diversity in the fine structure of the interface. There are conserved contacts between subsets of TCRs (with similar variable (V) segments) and subsets of MHC molecules[22,23], but no contacts or conformational changes at the binding interface have been identified that are conserved in all TCR–peptide–MHC complex structures[22–24]. Rossjohn *et al.*[25] observed that a triad of MHC class I residues was involved in all published structures of TCR–peptide–MHC class I complexes, but their TCR contacts were variable, as was their contribution to binding energy[26]. Thus, despite having conserved features, such as a broadly diagonal orientation, the TCR–peptide–MHC binding interface is diverse at the atomic level, with no contacts or conformational changes that are common to all TCR–peptide–MHC complexes. Any triggering mechanism needs to accommodate this structural diversity.

The key issue in TCR triggering is explaining how TCR binding to a peptide–MHC complex results in biochemical changes in the cytoplasmic domains of the TCR–CD3 complex. Models of TCR triggering invoke one or more of three basic mechanisms: aggregation, conformational change and segregation or redistribution of the TCR–CD3 complex[27]. Although it is likely that triggering will involve a combination of these mechanisms, it is useful to consider each mechanism separately. Several TCR signalling models (such as the serial-triggering model[28] and the kinetic proof-reading model[29,30]) have been proposed to account for certain quantitative features of T cell activation such as antigen discrimination. As they do not imply any particular molecular mechanisms of TCR triggering, they are not discussed further here.

Aggregation

It is not difficult to envisage how aggregation of TCR–CD3 complexes following TCR engagement could lead to enhanced phosphorylation. This aggregation could, for example, increase the proximity of associated LCK molecules, resulting in the activation of the second receptor in the aggregate by *trans*-autophosphorylation[31]. Forced aggregation of TCRs using either soluble antibodies or soluble multimeric forms of peptide–MHC complexes is sufficient to initiate TCR triggering. Although this clearly shows that artificial aggregation is sufficient for

Box 1 | Distinguishing signal from noise

T cells can detect tiny amounts of high-affinity foreign peptide–MHC ligand (signal) presented in the context of large amounts of low-affinity self peptide–MHC complex (noise). To appreciate this feat, it is helpful to consider a specific example. Consider a T cell receptor (TCR) that dissociates from a self peptide–MHC complex with an average half-life of 1 second ($k_{off} = 0.69\ s^{-1}$) and from a foreign peptide–MHC complex with a half-life of 5 seconds ($k_{off} = 0.14\ s^{-1}$), but has the same on rate for both ($k_{on} = 0.001\ \mu m^2\ s^{-1}$). Further assume that the concentration of the self peptide–MHC ligand is tenfold higher than that of the foreign peptide–MHC ligand (10 μm^{-2} and 1 μm^{-2}, respectively). How can the T cell only respond when foreign peptide–MHC complexes are present?

The simplest discrimination mechanism is the number of occupied TCRs. With some assumptions (such as the conservation of TCR and peptide–MHC complexes) we can compute the number of engaged TCR complexes (*C*) using the following equation:

$$C = \frac{A}{2}\left[P_T + T_T + K_d - \sqrt{(P_T + T_T + K_d)^2 - 4P_T T_T}\right]$$

where *A* ($\pi 5^2\ \mu m^2$) is the area of the contact interface, P_T and T_T are the total concentration of the peptide–MHC and TCR complexes, respectively, and K_d is the dissociation constant. Substituting in the values above, we find that TCR occupancy produces a poor signal-to-noise ratio of 0.33 (0.42 μm^{-2} / 1.2 μm^{-2}). Therefore, on average, more receptors will be occupied by self peptide–MHC ligand than foreign peptide–MHC ligand.

An alternative output that the T cell can potentially respond to is the rate of TCR binding events, as in the serial-triggering model[28]. The rate of binding events is simply $k_{off}C$, which is 68 s^{-1} and 4.5 s^{-1} for self and foreign peptide–MHC ligands, respectively. This produces an even lower signal to noise ratio of 0.07. This ratio can be improved if we stipulate that each TCR will only signal if bound for a threshold time (τ), as in the kinetic proof-reading model[29,30]. In this case, the rate of productive binding events is $k_{off}C\exp(-k_{off}\tau)$. Taking $\tau = 5$ seconds, we find that the revised binding rate is 2.1 s^{-1} and 2.2 s^{-1} for self and foreign peptide–MHC ligands, respectively, which corresponds to a signal-to-noise ratio of 1.1. Although this is a substantial improvement, it is still inadequate because both self and foreign peptide–MHC ligands will produce a similar number of productive binding events, making them indistinguishable.

© 2011 Macmillan Publishers Limited. All rights reserved

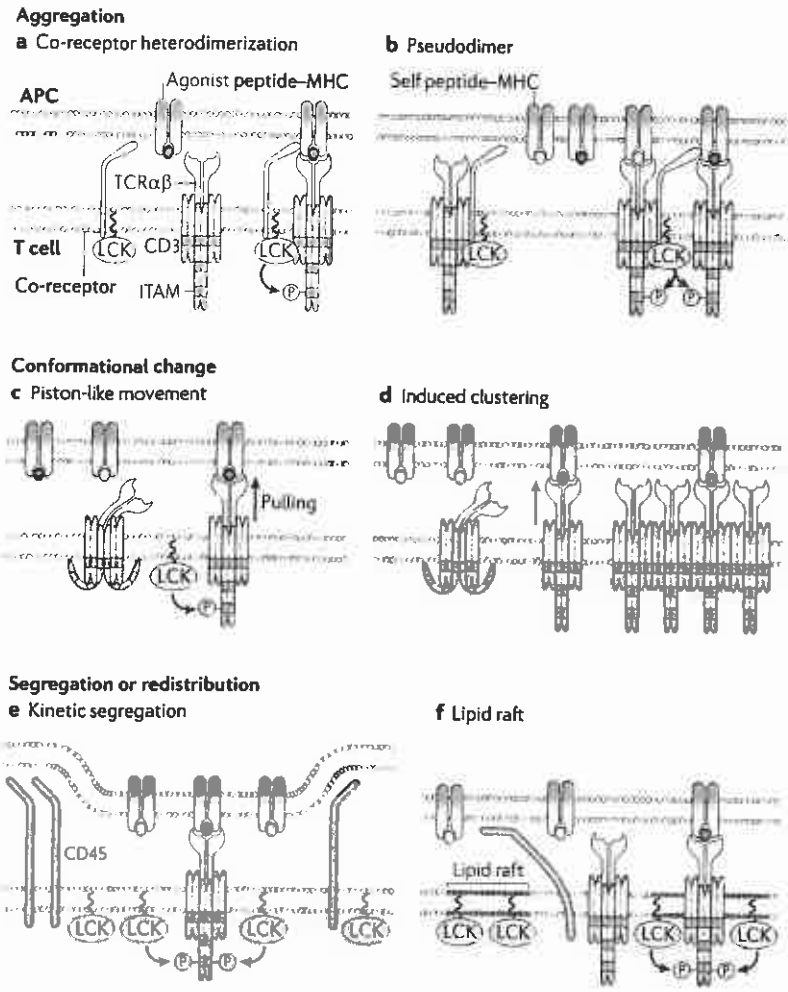


Figure 1 | **Mechanisms of TCR triggering.** Models are grouped according to whether the primary mechanism is aggregation, conformational change or segregation. **a** | In the co-receptor heterodimerization model, co-receptor binding to the same peptide–MHC complex as the T cell receptor (TCR) brings co-receptor-associated LCK into proximity with TCR–CD3 immunoreceptor tyrosine-based activation motifs (ITAMs). **b** | The pseudodimer model postulates that two TCRs are brought together by binding low-affinity self (green) or high-affinity agonist (blue) peptide–MHC ligand and that the co-receptor associated with one TCR engages the agonist peptide–MHC complex, thereby forming a dimer. **c** | A piston-like displacement of the TCR–CD3 complex is induced by the mechanical effects (primarily pulling; black arrow) of peptide–MHC binding to the TCR. This leads to a change in the conformation of the CD3 cytoplasmic domains, allowing ITAM phosphorylation. **d** | Clustering is induced by a conformational change in TCR–CD3 complex, possibly enhancing kinase activity. The conformational change in the cytoplasmic domains is depicted as dissociation from the membrane, as proposed in the 'safety-catch' model[7]. **e** | The kinetic-segregation model proposes that TCR binding to peptide–MHC ligand traps the TCR–CD3 complex in close-contact zones, thereby segregating it from the inhibitory tyrosine phosphatase CD45, leading to stable phosphorylation of TCR–CD3 ITAMs by LCK. **f** | Lipid raft models postulate that peptide–MHC engagement somehow results in partitioning of the TCR–CD3 complex into regions of membrane enriched in LCK and deficient in CD45. APC, antigen-presenting cell.

triggering, the challenge has been to explain how binding of agonist peptide–MHC ligand alone can induce aggregation when the ligand is present at such low surface densities. Indeed, TCR triggering can be observed with a single agonist peptide–MHC ligand[8–10]. Several models have been proposed to account for this.

The co-receptor heterodimerization model (reviewed in REF. 32) postulates that CD4 or CD8 co-receptors bind to the same agonist peptide–MHC complex as the TCR, thereby recruiting co-receptor-associated LCK into close proximity with CD3 complex ITAMs to mediate their phosphorylation (FIG. 1a). However, TCR triggering can occur in the complete absence of co-receptors[33,34], indicating that co-receptor heterodimerization is not essential for TCR triggering. Furthermore, in most studies, soluble agonist peptide–MHC monomers cannot induce TCR triggering, indicating that co-receptor heterodimerization is usually not sufficient for TCR triggering.

The pseudodimer model postulates a role for self peptide–MHC molecules in TCR triggering[9,35]. According to this model, one TCR binds an agonist peptide–MHC molecule and a second TCR binds a self peptide–MHC molecule. Dimerization is enhanced because the co-receptor associated with the TCR that is complexed with the self peptide–MHC molecule binds to the agonist peptide–MHC complex. A pseudodimer is hence formed by the dual interaction of a second TCR with self peptide–MHC and its associated CD4 or CD8 co-receptor with the agonist peptide–MHC complex (FIG. 1b). As self peptide–MHC molecules are present at a much higher surface density, this helps to address the problem of low surface density of agonist peptide–MHC molecules. A key prediction of this model is that self peptide–MHC molecules would enhance TCR triggering, especially at low densities of agonist peptide–MHC ligands. Evidence for this is clearest in the case of $CD4^+$ T cells, whereas there are conflicting data for $CD8^+$ T cells (reviewed in REF. 36). Interestingly, a high proportion of self peptide–MHC molecules (50–100% of those tested) seem to enhance agonist peptide–MHC ligand recognition[36,37]. It is therefore plausible that aggregates of agonist and self peptide–MHC molecules could induce clustering of TCRs and co-receptors simply by binding to the TCR.

Finally, some models propose that engagement of peptide–MHC molecules induces conformational changes in the TCR–CD3 complex that predispose it to dimerization and aggregation[38,39]. These models are considered in the next section.

There is some controversy as to the natural state of the TCR at the cell surface. Some studies report that at least a proportion of TCRs are in aggregates (or clusters)[40–43] that can form what some have termed 'protein islands'[41], whereas other studies suggest that the TCRs are primarily monomeric[44,45]. The observed TCR clusters were 30–300 nm in diameter and contained 5–20 TCRs[41–43]. It is noteworthy that in the cases in which clustering was reported, the T cells were in contact with artificial surfaces and the clusters were observed in these contact areas. These results can be reconciled if it is postulated that TCRs are primarily monomeric but are predisposed to clustering following initial triggering, and that contact with surfaces can induce clustering by inducing weak TCR triggering[46].

© 2011 Macmillan Publishers Limited. All rights reserved

Whatever the natural state of the TCR before engagement of peptide–MHC molecules, it is clear that this engagement results in increased aggregation of TCRs into what have been termed 'microclusters', containing 10–100 TCRs[43,47]. The formation of these microclusters was first visualized by total internal reflection fluorescence (TIRF) microscopy of T cells in contact with planar bilayers[47,48]. Interestingly, their size seems to vary little with the surface density of peptide–MHC molecules[43].

This raises the question as to the mechanism of microcluster formation. Is it simply a physical consequence of the binding of agonist and self peptide–MHC molecules by TCRs and their co-receptors? Or are other processes involved? In support of the latter is the observation that microcluster formation is blocked by the inhibition of actin polymerization, suggesting a role for the actin cytoskeleton in microcluster formation[47,48]. It is possible that TCR triggering itself leads to clustering, through, for example, signalling molecules that associate with and crosslink the TCR–CD3 cytoplasmic domains. Imaging studies to date have been unable to resolve whether TCR triggering precedes microcluster formation or vice versa. However, some formation of microclusters is still observed in the presence of Src tyrosine kinase inhibitors[48,49]. This shows that ITAM phosphorylation is not required for microcluster formation, but does not rule out a role for other signalling pathways.

In conclusion, recent studies on the ability of self peptide–MHC molecules to engage TCRs suggest that aggregation is a plausible mechanism of TCR triggering despite the low density of agonist peptide–MHC ligands.

Conformational change

Several models have been proposed that invoke binding-induced conformational change as a mechanism of TCR triggering. An attraction of these models has been that they can account for triggering at very low densities of agonist peptide–MHC molecules. A conformational change model needs to explain in molecular detail how TCR binding to peptide–MHC molecules can lead to changes in the CD3 cytoplasmic domains. At present there are no models that can do this satisfactorily. This is partially because we still do not know the structure of the intact TCR–CD3 complex. However, structural and other studies of different portions of the TCR–CD3 complex have provided intriguing clues.

TCR ectodomains. Although the binding of peptide–MHC molecules often induces conformational changes in the TCR[24], these are primarily in the contact area and are not widely conserved. However, crystallographic studies of a TCR in the bound and unbound state suggested a possible subtle conformational change in the membrane-proximal AB loop of the TCRα constant (Cα) domain[50]. A recent follow-up study[51] confirmed this using a different technique and reported a similar change in a different TCR. In addition, this study showed that mutation of residues in the AB loop abrogates TCR antigen recognition[51]. However, the AB loop of the TCR Cα domain is often poorly resolved in the crystal structures of TCR–peptide–MHC complexes, so it is not clear how widespread the conformational change is following TCR binding[51]. Further structural studies in which the relevant TCR regions are well resolved are needed to address this.

How could such a conserved conformational change in the TCR Cα domain be produced given the diversity at the binding interface of the TCR–peptide–MHC complex? One possible mechanism, discussed below, is through a mechanical force generated by the peptide–MHC complex pulling on the TCR. Another question is how changes in this region of the TCR could be transduced to the cell interior. Based on their observation that mutations of the AB loop affect TCR dimerization, Kuhns *et al.*[40] have proposed that the conformational change in the AB loop of the TCR Cα domain regulates TCR dimerization. Thus, binding-induced conformational changes may signal by inducing clustering. Similar models have been previously proposed based on reports that TCR binding to peptide–MHC complexes leads to clustering[38,39].

CD3 ectodomains. The ectodomains of CD3δ and CD3γ form stable heterodimers with the CD3ε ectodomain, and the structures of the CD3δε[52,53] and CD3γε[54,55] heterodimers have been determined. Although it is not yet understood how the TCRαβ ectodomain associates with the CD3 ectodomains, compelling evidence of direct association between the transmembrane regions[56] suggests that these ectodomains will be in close physical contact, at least in the portions that are close to the membrane. In the absence of structural data, indirect methods have been used to infer the arrangement of the TCR and CD3 ectodomains, with contradictory results[57]. However, recent studies support a model in which the CD3 ectodomains are arranged on one side of the TCRαβ ectodomain, with CD3δ and CD3γ contacting the TCR Cα and Cβ domains, respectively, and with the two CD3ε subunits in close proximity to each other[40,58,59]. In this model, the AB loop of the TCR Cα domain is not in contact with the CD3 heterodimers and is potentially available to mediate TCR dimerization[40].

Reinherz and colleagues[54] have observed that the CD3γε heterodimer seems to be relatively rigid and, on this basis, argued in favour of a role for mechanical piston-like forces (see below). Structural studies of the CD3γε and CD3δε heterodimers have not provided any direct evidence for conformational change, but molecular dynamic simulations have identified possible conformational changes in the CD3ε stalk[60]. Mutation of residues in this region, including in a conserved CXXC motif, abrogated TCR signalling *in vitro* and T cell development *in vivo*[60,61]. Surprisingly, these mutant forms of CD3ε inhibited TCR function even in the presence of excess wild-type CD3ε[60], supporting a model in which TCR–CD3 complexes form cooperative microclusters[18].

CD3 cytoplasmic domains. Several studies have suggested that the cytoplasmic portions of the CD3 complex undergo conformational change and that this could be implicated in TCR triggering. Alarcon and colleagues[3–5] provided evidence for conformational changes when

Kinetic proof-reading model
A model that was proposed to account for the ability of T cells to discriminate between peptide–MHC ligands that have small differences in their affinity or half-life. It postulates that T cell receptor (TCR) triggering requires multiple sequential steps that can only proceed while the TCR is engaged with a peptide–MHC complex and that are completely reversed as soon as the TCR dissociates from the complex

© 2011 Macmillan Publishers Limited. All rights reserved

the TCR is engaged by antibody or by peptide–MHC molecules, and implicated a proline-rich motif in CD3ε that binds the adaptor protein non-catalytic region of tyrosine kinase (NCK). However, this motif seems to be involved in regulating the expression of TCR and CD3 subunits, and not in TCR triggering[62,63].

Stern and colleagues[6] used circular dichroism to investigate the protein structures and showed that the cytoplasmic portion of TCRζ undergoes a conformational change upon binding to acidic lipid vesicles. This change in conformation is accompanied by changes in tyrosine fluorescence and a decreased susceptibility to phosphorylation. More recently it was shown that a basic residue rich sequence (BRS) in CD3ε mediates binding to negatively charged phospholipids in the cell membrane[7,64]. Wucherpfennig and colleagues[7] also provided nuclear magnetic resonance evidence that the tyrosine residues in the CD3ε ITAMs may be buried in the lipid bilayers. Based on these findings, and the observation that the TCRζ cytoplasmic domain also contains BRSs, a 'safety-catch' model was proposed, postulating that some TCR–CD3 ITAMs are protected from phosphorylation in the resting state (safety on) and that TCR engagement by ligand results in the dissociation of these ITAMs from the phospholipids in the cell membrane (safety off), exposing them to phosphorylation[6,65]. However, the finding that exposure of T cells to the tyrosine phosphatase inhibitor pervanadate induces dramatic TCR–CD3 ITAM phosphorylation[66–69] shows that these ITAMs are accessible to kinases in the absence of TCR engagement. Also arguing against this model is the finding that mutation of the CD3ε BRS abrogates phosphorylation[64,69]. Wucherpfennig and colleagues[70] have countered that the effects of pervanadate are secondary to lipid modification and that BRS mutants could disrupt the CD3ε conformation. Further experiments are needed to resolve these issues.

Mechanical effects. Recently, there has been increased interest in conformational change models that invoke a role for mechanical forces such as pulling or shearing[27,54,71–73]. These models have been inspired by several lines of evidence. First, the recognition that imposition of a mechanical pulling force on the TCR–CD3 complex is an inevitable consequence of binding to the peptide–MHC ligand[74], which has recently been supported by measurement of the TCR–peptide–MHC complex

Box 2 | **Two-dimensional binding properties**

Given that the T cell receptor (TCR) and peptide–MHC complex are normally associated with cell surfaces, it has long been a goal to measure the binding properties of this interaction *in situ*. As membrane-associated molecules are constricted to diffuse in two dimensions, their binding properties are stated in two-dimensional (2D) units, in contrast with solution binding properties, which are three-dimensional (3D). Two recent landmark studies have measured the 2D binding properties using very different approaches, producing intriguing results and valuable insights[105,106]. Huppa *et al.*[106] used fluorescent protein tags, total internal reflection fluorescence (TIRF) microscopy and fluorescent resonance energy transfer (FRET) to measure the surface densities of TCR, peptide–MHC and TCR–peptide–MHC complexes at the interface between T cells and planar bilayers onto which peptide–MHC complexes and other ligands had been attached. By assuming that these interactions were at equilibrium, they used these data to calculate the 2D dissociation constant (K_d) in the T cell–bilayer interface at the resolution of TIRF microscopy. By measuring the duration of FRET for a large number of TCR–peptide–MHC complexes they were also able to estimate the 2D off rate (k_{off}).

Huang *et al.*[105] used mechanical assays to measure the 2D binding parameters. In these assays, peptide–MHC molecules that were attached directly or indirectly to erythrocytes were brought into contact with T cells for varying periods, and binding was detected by deformation of the erythrocytes when the cells were pulled apart. Both studies found that the 2D k_{off} was significantly faster than the solution k_{off}, and that this difference was reduced by inhibitors of the actin cytoskeleton[105,106]. This is evidence that the TCR–peptide–MHC interaction is subjected to mechanical forces, as previously predicted[74]. A second key finding was that the 2D K_d was highly variable, and that this variability was largely the result of changes in the 2D on rate (k_{on}). Huppa *et al.* found considerable variation (~200-fold) for a given TCR–peptide–MHC combination even within a single contact interface[106], whereas Huang *et al.* found even greater variation (~2,000-fold) between a set of peptide–MHC analogues, despite the fact that their solution or 3D k_{on} did not vary much[105]. Interestingly, the dramatic differences in 2D k_{on} correlated well with activation potency[105]. Paradoxically, Huang *et al.* also found that the 2D k_{off} correlated with activation potency and inversely with 3D k_{off}. Furthermore, the 2D k_{on} was reduced by treatments that would be expected to inhibit TCR triggering and TCR clustering, and the same treatments decreased the 2D k_{off}. A plausible explanation for these intriguing results is that TCR triggering, which is induced by binding to agonist peptide–MHC ligand, rapidly leads to changes in the local surface density and/or orientation of the TCR–CD3 complex, and these changes are accompanied by an increase in the mechanical force that the TCR–peptide–MHC interaction is subjected to. Triggering-induced TCR clustering could increase the surface density of TCR–CD3 complexes over 100-fold in the immediate vicinity of the engaged TCR, accounting for most of the apparent increase in 2D k_{on}. Triggering-induced clustering could also enhance the apparent 2D k_{on} by optimizing the TCR–CD3 orientation (for example, by making it more upright). A previous study from the same group showing an increased probability of TCR–peptide–MHC binding upon subsequent encounters also supports triggering-induced enhancement of TCR–CD3 binding by clustering and/or reorientation[107].

These studies reveal that measuring 2D binding properties in the context of functional T cells is complicated by triggering-induced changes in TCR surface density and/or orientation and in the mechanical forces that act on the TCR–peptide–MHC interaction. Measurements in systems in which these effects are controlled will be necessary to fully elucidate the relationship between 2D binding properties and TCR triggering.

© 2011 Macmillan Publishers Limited. All rights reserved

off rate in intact cells (BOX 2). Second, the structural evidence that the CD3γε heterodimer may be rigid[54] and that the protruding FG loop of the TCR Cβ domain is in contact with the membrane-distal top end of the CD3γε heterodimer[58,75]. Thus any force applied to the TCRαβ heterodimer could be transmitted through the CD3γε heterodimer to the membrane. Third, the long-standing observation that T cell activation by artificial ligands is optimal when the ligands are anchored to a surface[76,77]. Fourth, an appreciation that elongation of the peptide–MHC ligand could abrogate TCR triggering by reducing any mechanical force on the TCR[71,78,79]. And, finally, reports that the application of mechanical force on the TCR enhances TCR triggering[73,80].

The mechanical forces could be generated by several processes[74]. The small size of the TCR–peptide–MHC complex would generate a force as larger molecules that are excluded from or compressed within close-contact areas try to diffuse in or straighten out. This is consistent with the observation that elongation of the peptide–MHC ligands abrogates triggering[78,79]. Active sources of force include the movement of cell processes and membranes, driven by cytoskeletal and endocytic mechanisms. This is supported by the long-standing observation that reagents that disrupt the actin cytoskeleton abrogate TCR triggering[81].

A major attraction of these mechanical force models is that they can readily explain how binding at a structurally variable peptide–MHC binding site in the TCR could induce the same conformational change in all TCR–CD3 complexes, because the mechanical force is a direct consequence of binding. Discrimination between different ligands comes about because the duration of the applied mechanical force and the resulting conformational change will be determined by the duration of binding.

It has also been proposed that peptide–MHC binding could push and/or twist the TCR[57,82]. However, an advantage of pulling or shearing mechanisms is that they are inherently specific because only specific interactions between the TCR and the peptide–MHC ligand can resist the pulling or shearing forces. By contrast, pushing does not necessarily require binding and could be nonspecific.

How might the effect of mechanical pulling on the TCR–CD3 complex be transduced into the cell interior? One possibility is that binding leads to a piston-like movement of the CD3 cytoplasmic tails, relative to the plasma membrane, that could alter their conformation (FIG 1c). Evidence cited in support of this includes the apparent rigidity of the CD3 ectodomains[54] and the observations that mutations in the conserved stalk region of CD3ε abrogate T cell activation[60,61]. In addition, the transmembrane regions of the TCR–CD3 complex specifically interact with each other, suggesting that they could form a relatively rigid structure[83,84]. However, convincing direct evidence of rigidity of the entire CD3 complex is lacking, particularly for the segments between the membrane and the ITAM region, making it unclear how piston-like movement could be transduced to the ITAMs.

A second possibility is that pulling induces a conformational change in the structure of the TCR–CD3 ectodomains and/or transmembrane domains that leads to clustering of the engaged TCR–CD3 complex with other TCR–CD3 complexes (FIG 1d). Such a mechanism has been proposed for B cell receptor (BCR) triggering[85]. Pierce and colleagues[86] have shown that binding of monomeric, surface-immobilized antigen to the IgM BCR induces clustering through its membrane-proximal Cμ4 domain.

In conclusion, evidence that a conformational change is involved in TCR triggering is mounting, but the nature of the conformational change and the mechanism by which peptide–MHC binding induces the conformational change remain unclear.

Segregation and redistribution

A third type of mechanism that has been proposed for TCR triggering is binding-induced segregation or redistribution of the TCR–CD3 complex with respect to other cell membrane-associated proteins. The TCR–CD3 complex is embedded in the plasma membrane along with molecules that favour signalling (such as the tyrosine kinase LCK) and molecules that inhibit signalling (such as the receptor tyrosine phosphatases CD45 and CD148). Treatment of resting T cells with inhibitors of tyrosine phosphatases leads to increased phosphorylation of TCR–CD3 ITAMs, resulting in TCR signalling and T cell activation even in the complete absence of TCR ligands[66–68]. Recently, Acuto and colleagues[87] reported that a substantial amount of LCK is constitutively active in resting T cells and that this did not increase detectably upon TCR engagement. These results suggest that phosphorylation by constitutively active LCK is held in check by constitutively active phosphatases. It follows that anything that disturbs this balance could induce an increase in TCR–CD3 ITAM phosphorylation and TCR signalling. One way of disturbing this balance is through the redistribution of the TCR–CD3 complex, LCK and CD45 with respect to each other.

Two types of redistribution or segregation models have been proposed. The kinetic-segregation model proposes that segregation of an engaged TCR–CD3 complex from CD45 is driven by differences in ectodomain size[88,89], whereas the lipid raft model proposes that redistribution is driven by changes in the lipid environment of the TCR[90].

As the kinetic-segregation model has recently been reviewed[91], it is discussed only briefly here. The model postulates that multiple zones of close contact (~15 nm apart) form at the cell–cell interface, from which molecules with large ectodomains, such as the inhibitory tyrosine phosphatases CD45 and CD148, are excluded (FIG. 1e). As a result, ITAM phosphorylation is strongly favoured. Binding to an agonist peptide–MHC ligand serves to trap the TCR–CD3 complex within a close-contact zone where it is protected from dephosphorylation by CD45 and CD148. This results in long-lived phosphorylation of TCR–CD3 ITAMs, leading to recruitment and activation of ζ-chain-associated protein kinase of 70 kDa (ZAP70) and subsequent phosphorylation by ZAP70 of the adaptor molecules SH2 domain-containing leukocyte protein of 76 kDa (SLP76; also known as LCP2) and linker for activation of T cells (LAT).

Lipid raft
An area of the plasma membrane that is rich in cholesterol, glycosphingolipids, glycosylphosphatidylinositol-anchored proteins and several signalling proteins — such as Src family kinases, LAT (linker for activation of T cells) and PAG (protein associated with glycolipid-enriched microdomains). These domains are also known as glycolipid-enriched microdomains (GEMs) and detergent-insoluble glycolipid-enriched membranes (DIGs).

135

© 2011 Macmillan Publishers Limited. All rights reserved

Several lines of evidence support this model: first, CD45 and CD148 are excluded from areas of TCR triggering[43,92]; second, truncation of the large CD45 and CD148 ectodomains inhibits TCR triggering[92,93]; third, elongation of the peptide–MHC complex inhibits TCR triggering[78,79]; fourth, surface-associated TCR ligands induce TCR triggering more effectively than their soluble counterparts[76,77]; and, finally, recognition by engineered TCRs is optimal when the epitope is positioned close to the plasma membrane of the target cell[94,95]. Although these results provide compelling evidence in support of the kinetic-segregation model, they do not exclude other mechanisms. Indeed, the last three lines of evidence cited above are also consistent with a mechanical pulling mechanism for TCR triggering.

Lipid raft models propose that TCR binding to peptide–MHC complexes leads to an association of the TCR–CD3 complex with lipid rafts. This enhances phosphorylation of TCR–CD3 ITAMs because lipid rafts are enriched in some proteins (such as LCK) and deficient in others (such as CD45) (FIG. 1f). The binding of peptide–MHC to TCR–CD3 may alter its lipid environment by immobilizing and/or clustering TCR–CD3 complexes. The role of lipid rafts has been controversial because the early techniques that were used to implicate them were later found to be unreliable[96]. Recent studies using new techniques have confirmed the existence of these structures and showed that they are smaller and more dynamic than previously appreciated[97]. Although there is substantial evidence supporting a role for lipid rafts in TCR signal transduction[90,98–100], some of this evidence has been challenged[101–104]. One study was unable to show co-localization of lipid raft markers with TCR microclusters[102]. Although this argues against a role for lipid rafts in microcluster formation it does not rule out a role for lipid rafts at earlier time points and/or on a smaller scale. Further studies using higher resolution imaging techniques are needed to elucidate the role of lipid rafts in the initial steps of TCR triggering. Although these findings suggest that segregation or redistribution of the TCR are likely to have an important role in TCR triggering, it is unclear whether they alone are sufficient to induce triggering.

Conclusion

Is it possible to reconcile all of the diverse experimental findings that support the different triggering mechanisms discussed above? We suggest the following scenario (FIG. 2). In resting T cells there is a delicate balance between tyrosine phosphorylation of the TCR–CD3 complex by constitutively active LCK and dephosphorylation by constitutively active phosphatases, with the phosphatases dominating (FIG. 2a). In principle, TCR triggering can be induced by any mechanism that tilts this balance in favour of phosphorylation. Engagement of the peptide–MHC complex leads simultaneously to aggregation of TCR–CD3 complexes and their segregation from other membrane proteins, which decreases the local phosphatase activity and increases the local kinase activity of LCK (FIG. 2b). Engagement also leads to conformational change in the cytoplasmic domain of the

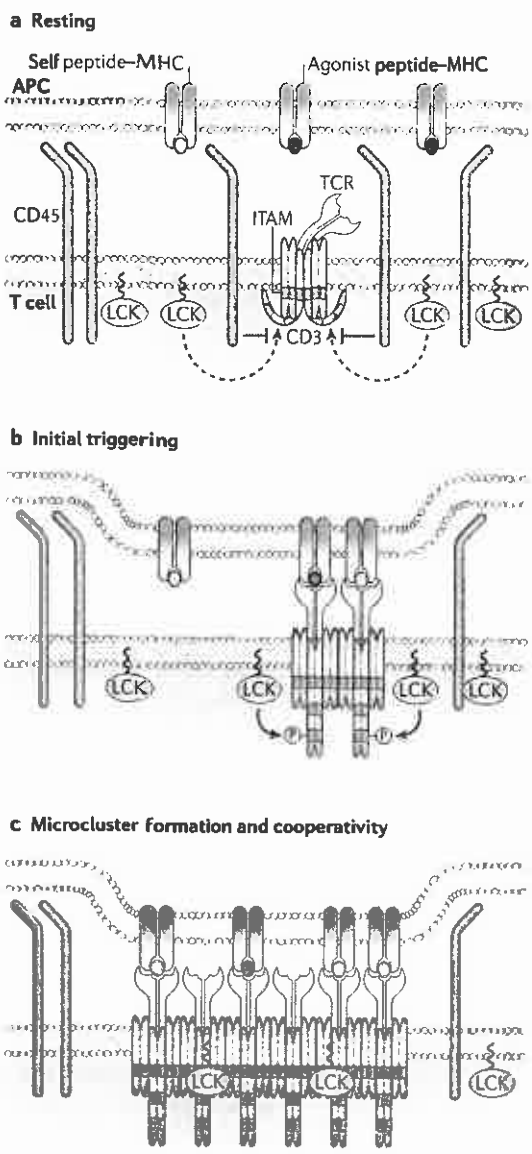


Figure 2 | **Integrated TCR triggering model. a** | In the resting state, the T cell receptor (TCR)–CD3 complex is primarily monomeric or forms very transient small aggregates. Phosphatase activity dominates and the level of TCR–CD3 immunoreceptor tyrosine-based activation motif (ITAM) phosphorylation is low. **b** | TCR engagement with peptide–MHC complex leads to segregation of the TCR–CD3 complex from phosphatases such as CD45, as well as aggregation and conformational change in the TCR–CD3 cytoplasmic domains. It is possible that aggregation alters the lipid environment of the TCR–CD3 complex (not shown). The overall result is a substantial increase in kinase activity and decrease in phosphatase activity, leading to increased ITAM phosphorylation. **c** | The TCR–CD3 complex aggregates enlarge to form microclusters as signalling proceeds. We suggest that these microclusters are the site of cooperative interactions that are required for ligand discrimination; for example, TCRs engaging agonist peptide–MHC ligand enhance signalling by other TCRs in the microcluster through positive feedback. APC, antigen-presenting cell.

(c) 2011 Macmillan Publishers Limited. All rights reserved

TCR–CD3 complex, which may alter susceptibility to phosphorylation. With continued triggering, TCR–CD3 complex aggregates enlarge to form microclusters containing 10–100 TCRs, some of which are bound to foreign peptide–MHC complexes (FIG. 2c).

The crucial requirement for TCRs to discriminate between abundant self and rare foreign peptide–MHC complexes is achieved at two levels. Individual TCRs use mechanisms such as kinetic proof-reading to discriminate binding events of different duration. However, as this is insufficient to cope with the stochastic nature of individual TCR–peptide–MHC interactions, we suggest that TCR microclusters are sites where multiple TCR–CD3 complexes cooperate to enable discrimination of rare foreign peptide–MHC ligands from more abundant self peptide–MHC molecules.

Further progress in our understanding of TCR triggering will require better information about the structure of the full TCR–CD3 complex and how this structure changes on peptide–MHC binding. Determining the structure of transmembrane proteins, particularly multisubunit proteins, is a major research bottleneck and progress in this area is likely to be slow and require considerable ingenuity. Also important is high-resolution information about the arrangement of the TCR–CD3 complex and other cell-surface and signalling molecules and how this arrangement changes following peptide–MHC engagement. Recent developments in super-resolution fluorescence microscopy suggest that advances in this area are probable over the next decade as these new techniques become more widely available.

1 Weiss, A. & Littman, D. R. Signal transduction by lymphocyte antigen receptors. *Cell* **76**, 263–274 (1994).
2 Weiss, A. & Samelson, L. E. in *Fundamental Immunology* (ed. Paul, W. E.) 321–364 (Lippincott Williams & Wilkins, Philadelphia, 2003).
3 Gil, D., Schamel, W. W., Montoya, M., Sanchez-Madrid, F. & Alarcon, B. Recruitment of Nck by CD3ε reveals a ligand-induced conformational change essential for T cell receptor signaling and synapse formation. *Cell* **109**, 901–912 (2002).
4 Gil, D., Schrum, A. G., Alarcon, B. & Palmer, E. T cell receptor engagement by peptide–MHC ligands induces a conformational change in the CD3 complex of thymocytes. *J. Exp. Med.* **201**, 517–522 (2005).
5 Risueno, R. M., Schamel, W. W. & Alarcon, B. T cell receptor engagement triggers its CD3ε and CD3ζ subunits to adopt a compact, locked conformation. *PLoS ONE* **3**, e1747 (2008).
6 Aivazian, D. & Stern, L. J. Phosphorylation of T cell receptor ζ is regulated by a lipid dependent folding transition. *Nature Struct. Biol.* **7**, 1023–1026 (2000).
7 Xu, C. *et al.* Regulation of T cell receptor activation by dynamic membrane binding of the CD3ε cytoplasmic tyrosine-based motif. *Cell* **135**, 702–713 (2008).
This paper provides evidence that the cytoplasmic domain of CD3ε associates with the plasma membrane of a cell and that the ITAM tyrosine residues are buried within the membrane, where presumably they are inaccessible to kinases.
8 Sykulev, Y., Joo, M., Vturina, I., Tsomides, T. J. & Eisen, H. N. Evidence that a single peptide–MHC complex on a target cell can elicit a cytolytic T cell response. *Immunity* **4**, 565–571 (1996).
9 Irvine, D. J., Purbhoo, M. A., Krogsgaard, M. & Davis, M. M. Direct observation of ligand recognition by T cells. *Nature* **419**, 845–849 (2002).
10 Purbhoo, M. A., Irvine, D. J., Huppa, J. B. & Davis, M. M. T cell killing does not require the formation of a stable mature immunological synapse. *Nature Immunol.* **5**, 524–530 (2004).
11 Ernst, B., Lee, D., Chang, J. M., Sprent, J. & Surh, C. D. The peptide ligands mediating positive selection in the thymus control T cell survival and homeostatic proliferation in the periphery. *Immunity* **11**, 173–181 (1999).
12 Goldrath, A. W. & Bevan, M. J. Low-affinity ligands for the TCR drive proliferation of mature CD8$^+$ T cells in lymphopenic hosts. *Immunity* **11**, 183–190 (1999).
13 Viret, C., Wong, F. S. & Janeway, C. A. Designing and maintaining the mature TCR repertoire: the continuum of self-peptide:self-MHC complex recognition. *Immunity* **10**, 559–568 (1999).
14 Palmer, E. & Naeher, D. Affinity threshold for thymic selection through a T-cell receptor–co-receptor zipper. *Nature Rev. Immunol.* **9**, 207–213 (2009).
15 Huse, M. *et al.* Spatial and temporal dynamics of T cell receptor signaling with a photoactivatable agonist. *Immunity* **27**, 76–88 (2007).
16 Altan-Bonnet, G. & Germain, R. N. Modeling T cell antigen discrimination based on feedback control of digital ERK responses. *PLoS Biol.* **3**, e356 (2005).
17 Chan, C., George, A. J. & Stark, J. Cooperative enhancement of specificity in a lattice of T cell receptors. *Proc. Natl Acad. Sci. USA* **98**, 5758–5763 (2001).
18 Schamel, W. W., Risueno, R. M., Minguet, S., Ortiz, A. R. & Alarcon, B. A conformation- and avidity-based proofreading mechanism for the TCR–CD3 complex. *Trends Immunol.* **27**, 176–182 (2006).
19 Germain, R. N. & Stefanova, I. The dynamics of T cell receptor signaling: complex orchestration and the key roles of tempo and cooperation. *Annu. Rev. Immunol.* **17**, 467–522 (1999).
20 Stefanova, I. *et al.* TCR ligand discrimination is enforced by competing ERK positive and SHP-1 negative feedback pathways. *Nature Immunol.* **4**, 248–254 (2003).
21 Starr, T. K., Jameson, S. C. & Hogquist, K. A. Positive and negative selection of T cells. *Annu. Rev. Immunol.* **21**, 139–176 (2003).
22 Garcia, K. C., Adams, J. J., Feng, D. & Ely, L. K. The molecular basis of TCR germline bias for MHC is surprisingly simple. *Nature Immunol.* **10**, 143–147 (2009).
23 Marrack, P., Scott-Browne, J. P., Dai, S., Gapin, L. & Kappler, J. W. Evolutionarily conserved amino acids that control TCR–MHC interaction. *Annu. Rev. Immunol.* **26**, 171–203 (2008).
24 Rudolph, M. G., Stanfield, R. L. & Wilson, I. A. How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.* **24**, 419–466 (2006).
25 Tynan, F. E. *et al.* T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide. *Nature Immunol.* **6**, 1114–1122 (2005).
26 Burrows, S. R. *et al.* Hard wiring of T cell receptor specificity for the major histocompatibility complex is underpinned by TCR adaptability. *Proc. Natl Acad. Sci. USA* **107**, 10608–10613 (2010).
27 Choudhuri, K. & van der Merwe, P. A. Molecular mechanisms involved in T cell receptor triggering. *Semin. Immunol.* **19**, 255–261 (2007).
28 Valitutti, S., Müller, S., Cella, M., Padovan, E. & Lanzavecchia, A. Serial triggering of many T-cell receptors by a few peptide–MHC complexes. *Nature* **375**, 148–151 (1995).
29 McKeithan, T. W. Kinetic proofreading in T-cell receptor signal transduction. *Proc. Natl Acad. Sci. USA* **92**, 5042–5046 (1995).
30 Rabinowitz, J. D., Beeson, C., Lyons, D. S., Davis, M. M. & McConnell, H. M. Kinetic discrimination in T-cell activation. *Proc. Natl Acad. Sci. USA* **93**, 1401–1405 (1996).
31 Cooper, J. A. & Qian, H. A mechanism for SRC kinase-dependent signaling by noncatalytic receptors. *Biochemistry* **47**, 5681–5688 (2008).
32 Trautmann, A. & Randriamampita, C. Initiation of TCR signalling revisited. *Trends Immunol.* **24**, 425–428 (2003).
33 Locksley, R. M., Reiner, S. L., Hatam, F., Littman, D. R. & Killeen, N. Helper T cells without CD4: control of leishmaniasis in CD4-deficient mice. *Science* **261**, 1448–1451 (1993).
34 Schilham, M. W. *et al.* Alloreactive cytotoxic T cells can develop and function in mice lacking both CD4 and CD8. *Eur. J. Immunol.* **23**, 1299–1304 (1993).
35 Krogsgaard, M. *et al.* Agonist/endogenous peptide–MHC heterodimers drive T cell activation and sensitivity. *Nature* **434**, 238–243 (2005).
36 Krogsgaard, M., Juang, J. & Davis, M. M. A role for "self" in T-cell activation. *Semin. Immunol.* **19**, 236–244 (2007).
37 Juang, J. *et al.* Peptide–MHC heterodimers show that thymic positive selection requires a more restricted set of self-peptides than negative selection. *J. Exp. Med.* **207**, 1223–1234 (2010).
This study showed that most self peptide–MHC molecules can contribute to negative selection when coupled to an agonist peptide–MHC molecule in peptide–MHC dimers, providing support for the pseudodimer model of TCR triggering, which postulates a critical role for self peptide–MHC molecules.
38 Alam, S. M. *et al.* Qualitative and quantitative differences in T cell receptor binding of agonist and antagonist ligands. *Immunity* **10**, 227–237 (1999).
39 Reich, Z. *et al.* Ligand-specific oligomerization of T-cell receptor molecules. *Nature* **387**, 617–620 (1997).
40 Kuhns, M. S. *et al.* Evidence for a functional sidedness to the αβTCR. *Proc. Natl Acad. Sci. USA* **107**, 5094–5099 (2010).
41 Lillemeier, B. F. *et al.* TCR and Lat are expressed on separate protein islands on T cell membranes and concatenate during activation. *Nature Immunol.* **11**, 90–96 (2010).
A high-resolution imaging study of the distribution of TCR and LAT on the surface of T cells. It suggests that both the TCR and LAT are clustered into small but separate islands that enlarge upon TCR triggering.
42 Schamel, W. W. *et al.* Coexistence of multivalent and monovalent TCRs explains high sensitivity and wide range of response. *J. Exp. Med.* **202**, 493–503 (2005).
43 Varma, R., Campi, G., Yokosuka, T., Saito, T. & Dustin, M. L. T cell receptor-proximal signals are sustained in peripheral microclusters and terminated in the central supramolecular activation cluster. *Immunity* **25**, 117–127 (2006).
44 Dunne, P. D. *et al.* DySCo: quantitating associations of membrane proteins using two-color single-molecule tracking. *Biophys. J.* **97**, L5–L7 (2009).
45 James, J. R. *et al.* Single-molecule level analysis of the subunit composition of the T cell receptor on live T cells. *Proc. Natl Acad. Sci. USA* **104**, 17662–17667 (2007).
46 van der Merwe, P. A., Dunne, P. D., Klenerman, D. & Davis, S. J. Taking T cells beyond the diffraction limit. *Nature Immunol.* **11**, 51–52 (2010).
47 Yokosuka, T. *et al.* Newly generated T cell receptor microclusters initiate and sustain T cell activation by recruitment of Zap70 and SLP-76. *Nature Immunol.* **6**, 1253–1262 (2005).
48 Campi, G., Varma, R. & Dustin, M. L. Actin and agonist MHC–peptide complex-dependent T cell receptor microclusters as scaffolds for signaling. *J. Exp. Med.* **202**, 1031–1036 (2005).

137

© 2011 Macmillan Publishers Limited. All rights reserved

49 Bunnell, S. C. *et al.* T cell receptor ligation induces the formation of dynamically regulated signaling assemblies. *J. Cell Biol.* **158**, 1263–1275 (2002).
50 Kjer-Nielsen, L. *et al.* A structural basis for the selection of dominant αβ T cell receptors in antiviral immunity. *Immunity* **18**, 53–64 (2003).
51 Beddoe, T. *et al.* Antigen ligation triggers a conformational change within the constant domain of the αβ T cell receptor. *Immunity* **30**, 777–788 (2009). This paper provides evidence that peptide–MHC binding induces a small but functionally significant conformational change in the distal AB loop of the TCR Cα domain.
52 Sun, Z.-Y. J. *et al.* Solution structure of the CD3εδ ectodomain and comparison with CD3εγ as a basis for modeling T cell receptor topology and signaling. *Proc. Natl Acad. Sci. USA* **101**, 16867–16872 (2004).
53 Arnett, K. L., Harrison, S. C. & Wiley, D. C. Crystal structure of a human CD3-ε/δ dimer in complex with a UCHT1 single-chain antibody fragment. *Proc. Natl Acad. Sci. USA* **101**, 16268–16273 (2004).
54 Sun, Z.-Y. J., Kim, K. S., Wagner, G. & Reinherz, E. L. Mechanisms contributing to T cell receptor signaling and assembly revealed by the solution structure of an ectodomain fragment of the CD3εγ heterodimer. *Cell* **105**, 913–923 (2001).
55 Kjer-Nielsen, L. *et al.* Crystal structure of the human T cell receptor CD3εγ heterodimer complexed to the therapeutic mAb OKT3. *Proc. Natl Acad. Sci. USA* **101**, 7675–7680 (2004).
56 Call, M. E. & Wucherpfennig, K. W. The T cell receptor: critical role of the membrane environment in receptor assembly and function. *Annu. Rev. Immunol.* **23**, 101–125 (2005).
57 Kuhns, M. S., Davis, M. M. & Garcia, K. C. Deconstructing the form and function of the TCR/CD3 complex. *Immunity* **24**, 133–139 (2006).
58 Kim, S. T. *et al.* Distinctive CD3 heterodimeric ectodomain topologies maximize antigen-triggered activation of αβ T cell receptors. *J. Immunol.* 185, 2951–2959 (2010).
59 Kuhns, M. S. & Davis, M. M. Disruption of extracellular interactions impairs T cell receptor–CD3 complex stability and signaling. *Immunity* **26**, 357–369 (2007).
60 Martínez-Martín, N. *et al.* Cooperativity between T cell receptor complexes revealed by conformational mutants of CD3ε. *Sci. Signal.* **2**, ra43 (2009).
61 Wang, Y. *et al.* A conserved CXXC motif in CD3ε is critical for T cell development and TCR signaling. *PLoS Biol.* **7**, e1000253 (2009).
62 Mingueneau, M. *et al.* The proline-rich sequence of CD3ε controls T cell antigen receptor expression on and signaling potency in preselection CD4$^+$CD8$^+$ thymocytes. *Nature Immunol.* **9**, 522–532 (2008).
63 Wang, H. *et al.* Tonic ubiquitylation controls T-cell receptor:CD3 complex expression during T-cell development. *EMBO J.* **29**, 1285–1298 (2010).
64 Deford-Watts, L. M. *et al.* The cytoplasmic tail of the T cell receptor CD3ε subunit contains a phospholipid-binding motif that regulates T cell functions. *J. Immunol.* **183**, 1055–1064 (2009).
65 Kuhns, M. S. & Davis, M. M. The safety on the TCR trigger. *Cell* **135**, 594–596 (2008).
66 Imbert, V. *et al.* Induction of tyrosine phosphorylation and T-cell activation by vanadate peroxide, an inhibitor of protein tyrosine phosphatases. *Biochem. J.* **297**, 163–173 (1994).
67 O'Shea, J. J., McVicar, D. W., Bailey, T. L., Burns, C. & Smyth, M. J. Activation of human peripheral blood T lymphocytes by pharmacological induction of protein-tyrosine phosphorylation. *Proc. Natl Acad. Sci. USA* 89, 10306–10310 (1992).
68 Secrist, J. P., Burns, L. A., Karnitz, L., Koretzky, G. A. & Abraham, R. T. Stimulatory effects of the protein tyrosine phosphatase inhibitor, pervanadate, on T-cell activation events. *J. Biol. Chem.* 268, 5886–5893 (1993).
69 Fernandes, R. A. *et al.* What controls T cell receptor phosphorylation? *Cell* **142**, 668–669 (2010).
70 Gagnon, E. *et al.* Response multilayered control of T cell receptor phosphorylation. *Cell* **142**, 669–671 (2010).
71 Ma, Z., Janmey, P. A. & Finkel, T. H. The receptor deformation model of TCR triggering. *FASEB J.* **22**, 1002–1008 (2008).
72 Choudhuri, A., Kearney, A., Bakker, T. R. & van der Merwe, P. A. Immunology: how do T cells recognize antigen? *Curr. Biol.* **15**, R382–R385 (2005).
73 Li, Y. C. *et al.* Cutting edge: mechanical forces acting on T cells immobilized via the TCR complex can trigger TCR signaling. *J. Immunol.* **184**, 5959–5963 (2010). This study provides evidence that the application of mechanical force on the TCR may induce TCR triggering.
74 van der Merwe, P. The TCR triggering puzzle. *Immunity* **14**, 665–668 (2001).
75 Ghendler, Y., Smolyar, A., Chang, H. C. & Reinherz, E. L. One of the CD3ε subunits within a T cell receptor complex lies in close proximity to the Cβ FG loop. *J. Exp. Med.* **187**, 1529–1536 (1998).
76 Geppert, T. D. & Lipsky, P. E. Accessory cell independent proliferation of human T4 cells stimulated by immobilized monoclonal antibodies to CD3. *J. Immunol.* **138**, 1660–1666 (1987).
77 Ma, Z., Sharp, K. A., Janmey, P. A. & Finkel, T. H. Surface-anchored monomeric agonist pMHCs alone trigger TCR with high sensitivity. *PLoS Biol.* **6**, e43 (2008).
78 Choudhuri, K. *et al.* Peptide–MHC dimensions control proximal kinase–phosphatase balance during T cell activation. *J. Biol. Chem.* **284**, 26096–26105 (2009).
79 Choudhuri, K., Wiseman, D., Brown, M. H., Gould, K. & van der Merwe, P. A. T cell receptor triggering is critically dependent on the dimensions of its peptide–MHC ligand. *Nature* **436**, 578–582 (2005).
80 Kim, S. T. *et al.* The αβ T cell receptor is an anisotropic mechanosensor. *J. Biol. Chem.* 284, 31028–31037 (2009).
81 Valitutti, S., Dessing, M., Aktories, K., Gallati, H. & Lanzavecchia, A. Sustained signalling leading to T cell activation results from prolonged T cell receptor occupancy. Role of the T cell actin cytoskeleton. *J. Exp. Med.* **181**, 577–584 (1995).
82 Davis, M. M. A new trigger for T cells. *Cell* **110**, 285–287 (2002).
83 Call, M. E., Pyrdol, J., Wiedmann, M. & Wucherpfennig, K. W. The organizing principle in the formation of the T cell receptor–CD3 complex. *Cell* **111**, 967–979 (2002).
84 Call, M. E. *et al.* The structure of the ζζ transmembrane dimer reveals features essential for its assembly with the T cell receptor. *Cell* **127**, 355–368 (2006).
85 Tolar, P. & Pierce, S. K. A conformation-induced oligomerization model for B cell receptor microclustering and signaling. *Curr. Top. Microbiol. Immunol.* 340, 155–169 (2010).
86 Tolar, P., Hanna, J., Krueger, P. D. & Pierce, S. K. The constant region of the membrane immunoglobulin mediates B cell-receptor clustering and signaling in response to membrane antigens. *Immunity* **30**, 44–55 (2009).
87 Nika, K. *et al.* Constitutively active Lck kinase in T cells drives antigen receptor signal transduction. *Immunity* 32, 766–777 (2010). This study showed that a large proportion of LCK in resting T cells is already in the active form and that this does not seem to change significantly upon TCR triggering.
88 Davis, S. J. & van der Merwe, P. A. The structure and ligand interactions of CD2: implications for T-cell function. *Immunol. Today* **17**, 177–187 (1996).
89 van der Merwe, P. A., Davis, S. J., Shaw, A. S. & Dustin, M. L. Cytoskeletal polarization and redistribution of cell surface molecules during T cell antigen recognition. *Semin. Immunol.* **12**, 5–21 (2000).
90 Horejsi, V. Lipid rafts and their roles in T-cell activation. *Microbes Infect.* **7**, 310–316 (2005).
91 Davis, S. J. & van der Merwe, P. A. The kinetic-segregation model: TCR triggering and beyond. *Nature Immunol.* **7**, 803–809 (2006).
92 Lin, J. & Weiss, A. The tyrosine phosphatase CD148 is excluded from the immunologic synapse and down-regulates prolonged T cell signaling. *J. Cell Biol.* **162**, 673–682 (2003).
93 Irles, C. *et al.* CD45 ectodomain controls interaction with GEMs and Lck activity for optimal TCR signaling. *Nature Immunol.* **4**, 189–197 (2003).
94 Bluemel, C. *et al.* Epitope distance to the target cell membrane and antigen size determine the potency of T cell-mediated lysis by BiTE antibodies specific for a large melanoma surface antigen. *Cancer Immunol. Immunother.* **59**, 1197–1209 (2010).
95 James, S. E. *et al.* Antigen sensitivity of CD22-specific chimeric TCR is modulated by target epitope distance from the cell membrane. *J. Immunol.* 180, 7028–7038 (2008).
96 Munro, S. Lipid rafts: elusive or illusive? *Cell* **115**, 377–388 (2003).
97 Lingwood, D. & Simons, K. Lipid rafts as a membrane-organizing principle. *Science* **327**, 46–50 (2010).
98 Gaus, K., Chklovskaia, E., Fazekas de St. Groth, B., Jessup, W. & Harder, T. Condensation of the plasma membrane at the site of T lymphocyte activation. *J. Cell Biol.* **171**, 121–131 (2005).
99 Zech, T. *et al.* Accumulation of raft lipids in T-cell plasma membrane domains engaged in TCR signalling. *EMBO J.* **28**, 466–476 (2009).
100 Harder, T. & Sangani, D. Plasma membrane rafts engaged in T cell signalling: new developments in an old concept. *Cell Commun. Signal.* **7**, 21 (2009).
101 Glebov, O. O. & Nichols, B. J. Lipid raft proteins have a random distribution during localized activation of the T-cell receptor. *Nature Cell Biol.* **6**, 238–243 (2004).
102 Hashimoto-Tane, A. *et al.* T-cell receptor microclusters critical for T-cell activation are formed independently of lipid raft clustering. *Mol. Cell. Biol.* **30**, 3421–3429 (2010).
103 Pizzo, P. *et al.* Lipid rafts and T cell receptor signaling: a critical re-evaluation. *Eur. J. Immunol.* 32, 3082–3091 (2002).
104 Zhu, M., Shen, S., Liu, Y., Granillo, O. & Zhang, W. Cutting edge: localization of linker for activation of T cells to lipid rafts is not essential in T cell activation and development. *J. Immunol.* **174**, 31–35 (2005).
105 Huang, J. *et al.* The kinetics of two-dimensional TCR and pMHC interactions determine T-cell responsiveness. *Nature* **464**, 932–936 (2010).
106 Huppa, J. B. *et al.* TCR–peptide–MHC interactions *in situ* show accelerated kinetics and increased affinity. *Nature* **463**, 963–967 (2010). References 105 and 106 measured the two-dimensional affinity and kinetics of TCR–peptide–MHC interactions for the first time.
107 Zarnitsyna, V. I. *et al.* Memory in receptor–ligand-mediated cell adhesion. *Proc. Natl Acad. Sci. USA* **104**, 18037–18042 (2007).

Acknowledgements

We thank our many scientific colleagues and collaborators, including S. Cordoba, H. Zhang, S. Davis, T. Harder, M. Brown, N. Barclay and O. Acuto, for many valuable discussions and insights, and for communicating results before publication. We also thank the referees for their comments and suggestions.

Competing interests statement

The authors declare no competing financial interests.

FURTHER INFORMATION

P. Anton van der Merwe's homepage:
http://www.path.ox.ac.uk/groups/vandermerwe

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2011 Macmillan Publishers Limited. All rights reserved

T cell receptor

From Wikipedia, the free encyclopedia

The **T cell receptor** or **TCR** is a molecule found on the surface of T lymphocytes (or T cells)[1] that is, in general, responsible for recognizing antigens bound to major histocompatibility complex (MHC) molecules. The binding between TCR and antigen is of relatively low affinity and is degenerate: that is, many TCR recognize the same antigen and many antigen are recognized by the same TCR.

The TCR is composed of two different protein chains (that is, it is a heterodimer). In 95% of T cells, this consists of an alpha (α) and beta (β) chain, whereas in 5% of T cells this consists of gamma and delta (γ/δ) chains.

When the TCR engages with antigen and MHC, the T lymphocyte is activated through a series of biochemical events mediated by associated enzymes, co-receptors, specialized accessory molecules, and activated or released transcription factors.

T cell receptor alpha locus

Identifiers

Symbol	TRA@
Alt. symbols	TCRA
Entrez	6955 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=retrieve&dopt=default&list_uids=6955&rn=1)
HUGO	12027 (http://www.genenames.org/data/hgnc_data.php?hgnc_id=12027)
OMIM	186880 (http://www.ncbi.nlm.nih.gov/omim/186880)

Other data

Locus	Chr. 14 *q11.2 (http://www.ncbi.nlm.nih.gov/Omim/getmap.cgi?chromosome=14q11.2)*

T cell receptor beta locus

Identifiers

Symbol	TRB@
Alt. symbols	TCRB
Entrez	6957 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=retrieve&dopt=default&list_uids=6957&rn=1)
HUGO	12155 (http://www.genenames.org/data/hgnc_data.php?hgnc_id=12155)
OMIM	186930 (http://www.ncbi.nlm.nih.gov/omim/186930)

Other data

Locus	Chr. 7 *q34 (http://www.ncbi.nlm.nih.gov/Omim/getmap.cgi?chromosome=7q34)*

Contents

- 1 Structural characteristics of the TCR
- 2 Generation of the TCR
- 3 The TCR complex
- 4 TCR co-receptors
- 5 Associated molecules of the TCR complex involved in T-cell activation
- 6 References
- 7 External links
- 8 See also

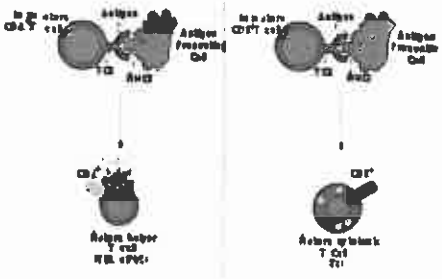
Structural characteristics of the TCR

TCR, which is anchored in the cell membrane, consists of two halves, which form a pair (or dimer) of protein chains. The halves are called the alpha (α) and beta (β) fragments (in γ/δ T cells, the halves are gamma (γ) and delta (δ) fragments). Each fragment is divided in turn into a constant (C) and variable (V) region. The constant region has an end that is anchored in the cell membrane. The variable region faces

139

outward and binds to the HLA molecule and the antigen it presents. On the α chain, the variable region is called Vα and the constant region is called Cα; on the β chain, they are called Vβ and Cβ, respectively.

The structure of TCR is very similar to immunoglobulin Fab fragments, which are regions defined as the combined light and heavy chain of an antibody arm. Each chain of the TCR is a member of the immunoglobulin superfamily and possesses one N-terminal immunoglobulin (Ig)-variable (V) domain, one Ig-constant (C) domain, a transmembrane/cell membrane-spanning region, and a short cytoplasmic tail at the C-terminal end.



Antigen presentation stimulates T cells to become either "cytotoxic" CD8+ cells or "helper" CD4+ cells.

The variable domain of both the TCR α-chain and β-chain have three hypervariable or complementarity determining regions (CDRs), whereas the variable region of the β-chain has an additional area of hypervariability (HV4) that does not normally contact antigen and, therefore, is not considered a CDR.

The residues are located in two regions of the TCR, at the interface of the α- and β-chains and in the β-chain framework region that is thought to be in proximity to the CD3 signal-transduction complex.[2] CDR3 is the main CDR responsible for recognizing processed antigen, although CDR1 of the alpha chain has also been shown to interact with the N-terminal part of the antigenic peptide, whereas CDR1 of the β-chain interacts with the C-terminal part of the peptide.

CDR2 is thought to recognize the MHC. CDR4 of the β-chain is not thought to participate in antigen recognition, but has been shown to interact with superantigens.

The constant domain of the TCR domain consists of short connecting sequences in which a cysteine residue forms disulfide bonds, which forms a link between the two chains.

Generation of the TCR

Processes for TCR formation are similar to those described for B cell antigen receptors, otherwise known as immunoglobulins.

- The TCR *alpha chain* is generated by VJ recombination, whereas the *beta chain* is generated by V(D)J recombination (both involve a somewhat random joining of gene segments to generate the complete TCR chain).
- Likewise, generation of the TCR *gamma chain* involves VJ recombination, whereas generation of the TCR *delta chain* occurs by V(D)J recombination.

The intersection of these specific regions (V and J for the alpha or gamma chain; V, D, and J for the beta or delta chain) corresponds to the CDR3 region that is important for antigen-MHC recognition (see above).

It is the unique combination of the segments at this region, along with palindromic and random N- and P-nucleotide additions, which accounts for the great diversity in specificity of the T cell receptor for processed antigen.

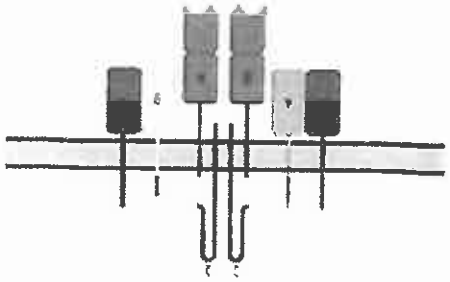
The TCR complex

The transmembrane region of the TCR is composed of positively charged amino acids.

It is thought that such structure allows the TCR to associate with other molecules like CD3, which possess three distinct chains (γ, δ, and ε) in mammals and either a ζ_2 complex or a ζ/η complex.

These **accessory** molecules have negatively charged transmembrane regions and are vital to propagating the signal from the TCR into the cell; the cytoplasmic tail of the TCR is extremely short, making it unlikely to participate in signaling.

The CD3- and ζ-chains, together with the TCR, form what is known as the *T cell receptor complex.*



The **T-cell receptor complex** with TCR-α and TCR-β chains, CD3 and ζ-chain accessory molecules.

TCR co-receptors

The signal from the T cell complex is enhanced by simultaneous binding of the MHC molecules by a specific co-receptor.

- On helper T cells, this co-receptor is CD4 that is specific for class II MHC.
- On cytotoxic T cells, this co-receptor is CD8 that is specific for class I MHC.

The co-receptor not only ensures the specificity of the TCR for an antigen but also allows prolonged engagement between the antigen-presenting cell and the T cell and recruits essential molecules (e.g., LCK) inside the cell involved in the signaling of the activated T lymphocyte.

Associated molecules of the TCR complex involved in T-cell activation

The essential function of the TCR complex is to identify specific bound antigen and elicit a distinct and critical response. The mechanism by which a T-cell elicits this response upon contact with its unique antigen is termed T-cell activation. There are myriad molecules involved in the complex biochemical process by which this occurs, which, in a wider context, is, in general, termed trans-membrane signalling.

Crystal structure of human CD8 molecule Only a fragment of extracellular portion of human CD8a is shown. Co-receptor CD 8 could bind class I MHC specifically

The most common mechanism for activation and regulation of molecules beneath the lipid bilayer is via phosphorylation/dephosphorylation by protein kinases. T-cells utilise the SRC family of kinases in transmembrane signalling largely to phosphorylate tyrosines that are part of immunoreceptor tyrosine-based activation motifs (ITAM).[3]

141

Early signally steps implicate the following kinases in TCR associated reactions.

- Lck - Associated with the transmembrane tail of CD4
- Fyn - Associated with ITAMs of the IgAlpha and Igbeta regions of the TCR complex
- CD45 - The transmembrane tail of which functions as a Tyrosine phosphatase)
- Zap70 - Binds to ITAM sequences upon phosphorylation by Lck and Fyn

References

1. ^ Thomas J. Kindt; Richard A. Goldsby; Barbara Anne Osborne; Janis Kuby (2007). *Kuby immunology* (http://books.google.com/books?id=oOsFf2WfE5wC&pg=PA223) . Macmillan. pp. 223–. ISBN 9781429202114. http://books.google.com/books?id=oOsFf2WfE5wC&pg=PA223. Retrieved 28 November 2010.
2. ^ Kieke, Michele C.; Shusta, Eric V.; Teyton, Luc; Wittrup, K. Dane; Kranz, David M. (1999). "Selection of functional T cell receptor mutants from a yeast surface-display library". *Proceedings of the National Academy of Science of the United States of America* **96** (10): 5651–5656. doi:10.1073/pnas.96.10.5651 (http://dx.doi.org/10.1073%2Fpnas.96.10.5651)
3. ^ Abram, Clare L.; Lowell, Clifford A. (2007-03-13). "The Expanding Role for ITAM-Based Signaling Pathways in Immune Cells". *Science Signalling* **2007** (377): re2.

External links

- T-cell Group - Cardiff University (http://www.tcells.org/scientific/abTCR/)
- UMich Orientation of Proteins in Membranes *protein/pdbid-2hac (http://opm.phar.umich.edu/protein.php?pdbid=2hac)* - Zeta-zeta dimer of T cell receptor
- MeSH *T-Cell+Receptor (http://www.nlm.nih.gov/cgi/mesh/2011/MB_cgi?mode=&term=T-Cell+Receptor)*

See also

- T cell
- Co-stimulation

Retrieved from "http://en.wikipedia.org/wiki/T_cell_receptor"
Categories: Genes on chromosome 14 | Genes on chromosome 7 | Cell signaling | T cells | Integral membrane proteins

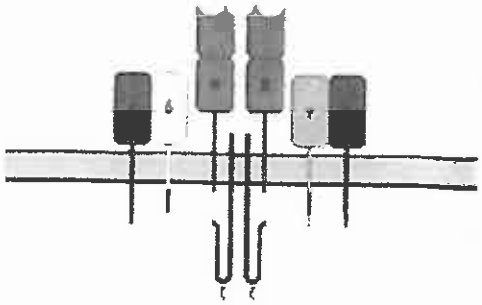
---

- This page was last modified on 9 June 2011 at 17:58.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. See Terms of use for details.
  Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

CD3 (immunology)

From Wikipedia, the free encyclopedia

In immunology, the **CD3** T-Cell Co-Receptor (CD stands for cluster of differentiation) is a protein complex and is composed of four distinct chains. In mammals, the complex contains a CD3γ chain, a CD3δ chain, and two CD3ε chains. These chains associate with a molecule known as the T cell receptor (TCR) and the ζ-chain to generate an activation signal in T lymphocytes. The TCR, ζ-chain and CD3 molecules together comprise the TCR complex.



The *T-cell receptor complex* with TCR-α and TCR-β chains (top), ζ-chain accessory molecules (bottom) and *CD3* (represented by CD3γ, CD3δ and two CD3ε).

Contents

- 1 Structure
- 2 Regulation
- 3 As a drug target
- 4 Immunohistochemistry
- 5 References
- 6 Further reading
- 7 External links

Structure

The CD3γ, CD3δ, and CD3ε chains are highly related cell surface proteins of the immunoglobulin superfamily containing a single extracellular immunoglobulin domain.

The transmembrane region of the CD3 chains is negatively charged (it contains aspartate residues in the transmembrane region), a characteristic that allows these chains to associate with the positively charged TCR chains.[1]

The intracellular tails of the CD3 molecules contain a single conserved motif known as an *immunoreceptor tyrosine-based activation motif* or ITAM for short, which is essential for the signaling capacity of the TCR.

CD3d molecule, delta

Identifiers

Symbol	CD3D
Alt. symbols	T3D
Entrez	915 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=retrieve&dopt=default&list_uids=915&rn=1)
HUGO	1673 (http://www.genenames.org/data/hgnc_data.php?hgnc_id=1673)
OMIM	186790 (http://www.ncbi.nlm.nih.gov/omim/186790)
PDB	1XIW (http://www.pdb.org/pdb/cgi/explore.cgi?pdbId=1XIW)
RefSeq	NM_000732 (http://genome.ucsc.edu/cgi-bin/hgTracks?Submit=Submit&position=NM_000732&rn=1)
UniProt	P04234 (http://www.uniprot.org/uniprot/P04234)

Other data

Locus	Chr. 11 *q23 (http://www.ncbi.nlm.nih.gov/Omim/getmap.cgi?chromosome=11q23)*

Regulation

Phosphorylation of the ITAM on CD3 renders the CD3 chain capable of binding an enzyme called ZAP70

143

(zeta associated protein), a kinase that is important in the signaling cascade of the T cell.

As a drug target

Because CD3 is required for T cell activation, drugs (often monoclonal antibodies) that target it are being investigated as immunosuppressant therapies (e.g., otelixizumab) for type 1 diabetes and other autoimmune diseases.

Immunohistochemistry

CD3 is is initially expressed in the cytoplasm of pro-thymocytes, the stem cells from which T-cells arise in the thymus. The pro-thymocytes differentiate into common thymocytes, and then into medullary thymocytes, and it is at this latter stage that CD3 antigen begins to migrate to the cell membrane. The antigen is found bound to the membranes of all mature T cells and NK cells, and in virtually no other cell type, although it does appear to be present in small amounts in Purkinje cells.

This high specificity, combined with the presence of CD3 at all stages of T cell development, makes it a useful immunohistochemical marker for T cells in tissue sections. The antigen remains present in almost all T-cell lymphomas and leukaemias, and can therefore be used to distinguish them from superficially similar B-cell and myeloid neoplasms.[2]

References

1. ^ Kuby, Janis; Kindt, Thomas J.; Goldsby, Richard A.; Osborne, Barbara A. (2007). *Kuby immunology*. San Francisco: W.H. Freeman. ISBN 1-4292-0211-4.
2. ^ Leong, Anthony S-Y; Cooper, Kumarason; Leong, F Joel W-M (2003). *Manual of Diagnostic Cytology* (2 ed.). Greenwich Medical Media, Ltd.. pp. 63-64. ISBN 1-84110-100-1.

Further reading

- Shiv Pillai MD; Abul K. Abbas MBBS; Andrew Wilson PhD (2011). *Cellular and Molecular Immunology: with STUDENT CONSULT Online Access*. Philadelphia: Saunders. ISBN 1-4377-1528-1.

CD3e molecule, epsilon

Identifiers

Symbol CD3E

Entrez 916 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=retrieve&dopt=default&list_uids=916&rn=1)

HUGO 1674 (http://www.genenames.org/data/hgnc_data.php?hgnc_id=1674)

OMIM 186830 (http://www.ncbi.nlm.nih.gov/omim/186830)

RefSeq NM_000733 (http://genome.ucsc.edu/cgi-bin/hgTracks?Submit=Submit&position=NM_000733&rn=1)

UniProt P07766 (http://www.uniprot.org/uniprot/P07766)

Other data

Locus Chr. 11 *q23 (http://www.ncbi.nlm.nih.gov/Omim/getmap.cgi?chromosome=11q23)*

CD3g molecule, gamma

Identifiers

Symbol CD3G

Entrez 917 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=retrieve&dopt=default&list_uids=917&rn=1)

HUGO 1675 (http://www.genenames.org/data/hgnc_data.php?hgnc_id=1675)

OMIM 186740 (http://www.ncbi.nlm.nih.gov/omim/186740)

RefSeq NM_000073 (http://genome.ucsc.edu/cgi-bin/hgTracks?Submit=Submit&position=NM_000073&rn=1)

UniProt P09693 (http://www.uniprot.org/uniprot/P09693)

Other data

Locus Chr. 11 *q23 (http://www.ncbi.nlm.nih.gov/Omim/getmap.cgi?chromosome=11q23)*

External links

- MeSH *CD3+Antigens (http://www.nlm.nih.gov/cgi/mesh/2011/MB_cgi?mode=&term=CD3+Antigens)*
- Mouse CD Antigen Chart (http://www.ebioscience.com/ebioscience/whatsnew/mousecdchart.htm)
- Human CD Antigen Chart (http://www.ebioscience.com/ebioscience/whatsnew/humancdchart.htm)

Retrieved from "http://en.wikipedia.org/wiki/CD3_(immunology)"
Categories: Genes on chromosome 11 | Clusters of differentiation | T cells | Chromosome 11 gene stubs

- This page was last modified on 20 July 2011 at 16:29.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. See Terms of use for details.
  Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

145

The amino acid and nucleotide sequences of the different chains of CD3 can be found in the GenBank database at http://www.ncbi.nlm.nih.gov . Their amino acid sequence regions are listed below:

CD3	Acc. No.	Signal	Extra-cell.	Membrane	Cytoplasmic
Zeta	P20963	1 – 21	22 – 30	31 – 51	52 -164
Epsilon	P07766	1 – 22	23 – 126	127 – 152	153 – 207
Gamma	P09693	1 – 22	23 – 116	117 – 137	138 – 182
Delta	P04234	1 – 21	22 – 105	106 – 136	137 – 171

Epsilon-delta (ε / δ) and epsilon-gamma (ε / γ) form dimers with their extra-cellular portion with the standard immunoglobulin fold. Their cytoplasmic regions are much longer than those of TCR α and β chains. Zeta chains form homo-dimer (ζ / ζ) with long cytoplasmic regions.

Unfortunately, the detail three dimensional structure of this T cell receptor for antigen complex has not been deciphered. Their trans-membrane regions are supposed to interact with each other so that these eight chains (TCR α and β chains, one γ, one δ, two ε, and two ζ chains of CD3) are brought together.

Since the CD3 molecules are conserved in humans, they have been the target of therapeutic antibodies to eliminate the activities of T lymphocytes.

Interaction of MHC class I with CD8

The molecular binding of TCR with the MHC class I-processed peptide complex is specific but cannot be strong enough to hold the antigen presenting cell and the T lymphocyte together. There are two large cells with a tiny bridge formed by molecular interaction. In order to strengthen this tiny bridge, another membrane bound molecule from the T lymphocyte us thus recruited. This molecule known as CD8 consists of two chains also designated as α and β. Their amino acid and nucleotide sequences are also available from the GenBank database, with the following regions for their amino acid sequences:

	Signal	Immunol.	Extra-cell.	Membrane	Cytoplasmic
CD8 α	1 – 21	31 – 133	22 – 182	183 – 203	204 – 235
CD8 β	1 – 21	29 – 134	22 – 170	171 – 191	192 – 210

The region designated by Immunol. has a three dimensional structure of the immunoglobulin fold. Both regions from CD8 α and β chains, i.e. the hetero-dimer α . β, will bind the MHC class I molecule (Devine L, Sun J, Barr MR, Kavathas PB (1999) Orientation of the Ig domain of CD8$\alpha\beta$ relative to MHC class I. J Immunol, 162, 846-851). In addition, homo-dimer α / α can also bind MHC class I molecule (Gao G, Jacobsen BK (2000) Molecular interactions of the coreceptor CD8 and MHC class I: the molecular basis for functional coordination with the T-cell receptor. Immuno Today, 21, 630-636).

Primarily, the immunoglobulin folded regions of CD8 α and β are in close contact with the α_3 region of the MHC class I heavy chain and the β_2 microglobulin, both of which also have immunoglobulin folds. The segments from position 134 to 182 of the CD8 α chain, a total of 39 amino acid residues, and from position 135 to 170 of the CD8 β chain, a total of 36 amino acid residues, are probably stretched out to allow for the spacing occupied by MHC class I α_1 and α_2 regions and TCR (Rudolph MG, Stanfield RL, Wilson IA (2006) How TCRs bind MHCs, peptides, and coreceptors. Annu Rev Immunol, 24, 419-466).

This larger complex formed by MHC class I molecule, β_2 microglobulin, processed peptide, TCR and CD8 may still not be able to hold the antigen presenting cell and the $CD8^+$ T lymphocyte together. However, that complex can now be "polymerized" with a collection of hundreds of these molecules.

Orientation of the Ig Domains of CD8$\alpha\beta$ Relative to MHC Class I[1]

Lesley Devine,* Jiaren Sun,* Mark R. Barr,* and Paula B. Kavathas[2]*†

The cell surface glycoprotein CD8 functions as a coreceptor with the TCR for interaction with MHC class I. The cocrystal structure of the CD8$\alpha\alpha$-MHC complex showed that one CD8 Ig domain provided the majority of the contact with MHC class I and that residue R4 of that domain contacted the α2 domain of MHC class I. We previously showed by mutational analysis that this residue was critical for binding to MHC class I. To determine which of the Ig domains for the CD8$\alpha\beta$ heterodimer would make the most contact with class I MHC, we expressed single-chain or dimeric forms of CD8 on COS-7 cells and measured the adhesion of MHC class I positive cells. We found that when one of the R4 residues was mutated in a CD8$\alpha\alpha$ homodimer binding comparable to that of wild type was observed, whereas a double R4 mutant severely impaired binding. However, when mutant CD8α (R4K) was coexpressed with wild-type CD8β, binding was not observed. These results support the model in which it is CD8α, not CD8β, that is making the most of the contact with MHC class I, including the α2 domain. In addition, they demonstrate that a single-chain form of CD8$\alpha\alpha$ can bind to MHC class I. *The Journal of Immunology*, **1999, 162: 846–851.**

The T cell coreceptor CD8 interacts with MHC class I and can be expressed on the cell surface in two molecular forms. The $\alpha\beta$ heterodimer is the major form present on thymocytes and mature T cells, while the $\alpha\alpha$ homodimer is exclusively present on subsets of NK cells and intraepithelial lymphocyte (IEL) cells of the intestine (1, 2).

The interaction of CD8$\alpha\alpha$ with MHC class I has been studied extensively. Salter et al. (3) identified a binding site for CD8 on a negatively charged loop of the α3 domain of MHC class I, and we identified critical residues located on the Ig domain of CD8 that were important for this interaction (4, 5). As with other Ig molecules, residues located on the surface containing the complementary determining region (CDR)[3]-like loops of CD8α were involved in recognition of MHC class I (5). However, unlike Ig molecules, residues located on the A and B β strands on the side of the dimer were also found to be critical for this interaction (4). These results allowed us to propose that while the residues located on the CDR-like loops made contact with the negatively charged loop on the α3 domain of class I, amino acids on the side of the molecule would interact with the MHC class I α2 domain. To support this, mutational studies of the α2 domain were conducted and residues found that were critical for CD8-MHC class I interaction (6). This model was further supported by functional studies with murine/human chimeric MHC class I (7).

The cocrystal structure of CD8/MHC class I confirmed that all but one of the residues that we had identified to be critical for the interaction with MHC class I by mutational analysis did, in fact, make contact (8). It also demonstrated that CD8$\alpha\alpha$ interacted with both the α2 and α3 domains of class I as well as with β_2-microglobulin and, as we predicted, that CD8$\alpha\alpha$ made contact using residues from the A and B strands. Gao et al. (8) also found that the contribution of the CD8α subunits to the binding was asymmetric, with one domain contributing 70% of the solvent-accessible area. While both subunits made contact with the α3 domain of class I (through their CDR-like loops), the subunit designated α-1 (Fig. 1) made additional contacts with the α2 domain of MHC class I and β_2-microglobulin. Within the CD8α-1 subunit, two residues, R4 and L25, located on the A and B strands, respectively, were shown to make contact with the α2 domain of MHC class I and β_2-microglobulin, respectively. We had previously found that both of these residues were critical for the interaction with MHC class I as mutation of either severely reduced binding (4).

All of these studies involved the CD8$\alpha\alpha$ homodimer. However, since the form predominantly expressed on T cells is the $\alpha\beta$ heterodimer, we were interested in studying this interaction with MHC class I. Due to the asymmetry in the binding of CD8$\alpha\alpha$ homodimers to MHC class I described in the cocrystal (8), the contribution CD8β makes to the interaction with MHC class I depends on whether it adopts the α-1 or α-2 subunit role. Based on modeling, Gao et al. (8) predicted that CD8β would function as the α-2 subunit and, therefore, would play a lesser role in the interaction.

To determine the orientation of CD8$\alpha\beta$ to MHC class I, we used a cell-cell adhesion assay to study how mutated forms of the heterodimers would interact with MHC class I. If CD8β functioned as the α-2 subunit, then coexpression with CD8α containing a mutation in a residue involved in contacting the α2 domain of class I or β_2-microglobulin (R4 or L25, respectively) would have a dramatic effect on binding. However, if CD8β functioned as the α-1 subunit, then the R4 or L25 residues of CD8α would not contact MHC class I, and coexpression with R4 or L25 mutant CD8α should have no effect on binding. Results from this study demonstrate that there was little or no binding of MHC class I positive cells to COS-7 cells expressing mutant CD8α (R4 or L25) and wild-type β, which was similar to that observed with binding to COS-7 cells expressing the R4K or L25A CD8$\alpha\alpha$ homodimers

*Department of Laboratory Medicine and Section of Immunobiology, and †Department of Genetics, Yale University School of Medicine, New Haven, CT 06520

Received for publication July 6, 1998. Accepted for publication September 22, 1998.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was funded by Grant AI35417 from the National Institutes of Health (to P.K.). L.D. was a recipient of the Brown-Coxe Fellowship.

[2] Address correspondence and reprint requests to Dr. Paula B. Kavathas, Department of Laboratory Medicine, Yale University School of Medicine, 333 Cedar Street, P.O. Box 208035, New Haven, CT 06520-8035. E-mail address: kavathas@beaker.med.yale.edu

[3] Abbreviations used in this paper: CDR, complementary determining region; WT, wild type; MT, mutant; IEL, intraepithelial lymphocyte.

Copyright © 1999 by The American Association of Immunologists

0022-1767/99/$02.00

Downloaded from www.jimmunol.org on July 27, 2011

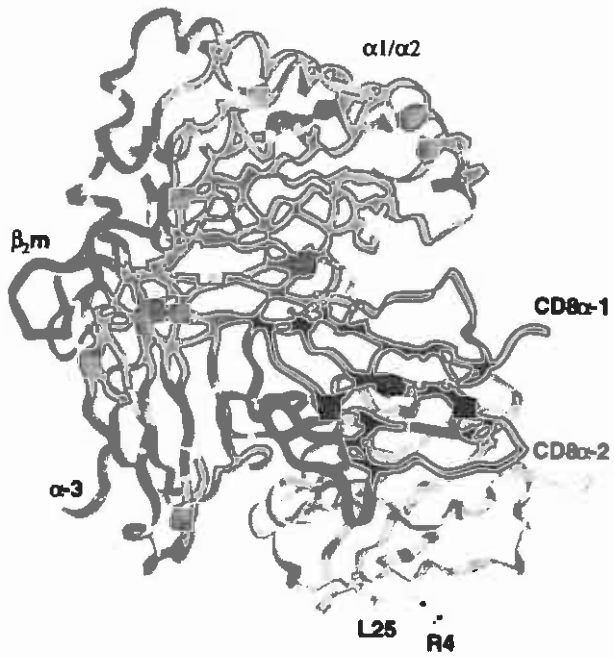


FIGURE 1. Ribbon diagram of CD8$\alpha\alpha$ MHC class I cocrystal. Diagram was taken from the Brookhaven Protein Data Bank (PDB; file lakj). CD8 Ig domains are designated CD8α-1 and CD8α-2. R4 and L25 residues of both CD8 Ig domains are shown as a ball and stick. β_2-Microglobulin interacts with residue L25 on CD8α-1 subunit, and R4, also on the CD8α-1 subunit, contacts the α2 domain of MHC class I. CDR-like loops of both CD8α domains interact with the α3 domain of class I.

alone. However, cells expressing a dimer of wild-type CD8α and mutant CD8α could bind MHC class I positive cells at a level comparable to wild-type CD8$\alpha\alpha$. Therefore, one CD8 monomer that can make contact with α2 and β_2-microglobulin is sufficient for binding. Thus, the inability of wild-type CD8β to compensate for the R4 or L25 mutation in CD8α supports the model that in the CD8$\alpha\beta$ heterodimer CD8β is analogous to the α-2 subunit.

Materials and Methods

Construction of single-chain CD8αα

Two CD8α Ig domains were linked via the carboxyl terminal of one to the amino terminal of the other by means of a peptide spacer. The second CD8α domain was linked to domains 3 and 4, together with the transmembrane and cytoplasmic domains of CD4 that allowed the chimeric protein to be expressed as a monomer. This approach has been used previously in the expression of a chimeric CD48 (13). A peptide spacer of 20 amino acids of 4 repeating units of GGGGS was determined to be sufficient for the 2 Ig-like domains to adopt the correct conformation using a computer graphic model of the CD8$\alpha\alpha$ homodimer crystal structure. Fig. 2*A* outlines the steps involved in synthesis of the single-chain construct. Briefly, the spacer was synthesized in two stages, and the PCR products of the individual α-Ig domains were subcloned into the TA cloning vector (Invitrogen, Carlsbad, CA). One *Stu*I site was deleted from the CD4 cDNA (position 934) using Kunkel mutagenesis, which allowed us to use the other site that occurred at the junction of domains 2 and 3 to subclone one CD8α domain. In addition, this mutagenesis introduced *Xho*I and *Eco*RI sites after the stop codon to facilitate future subcloning.

The second α subunit was linked to the 5′ end of the linker via the introduced *Pst*I site (a second *Pst*I site is located at the 5′ end of CD8α). All of the subcloning was conducted in pBluescript II, then the entire construct was subcloned into the expression vector pCDL-SRα296 (14).

The various mutant constructs (Fig. 2*B*) were synthesized by Kunkel mutagenesis (15) of the wild-type single-chain construct in pBluescript II. Different primers were used due to the 5′ end of both α subunits being different (subunit 1 contains the leader sequence, whereas subunit 2 has the linker peptide sequence 5′), which allowed us to introduce mutations in either subunit independently. The mutation introduced a *Dra*I site that allowed us to confirm the number of mutant sites within the entire construct. All constructs were sequenced.

Expression of different forms of CD8 on COS-7 cells

COS-7 fibroblasts were transfected using a modification of a previously described method (4). Briefly, an expression vector containing a total of 2 μg of CD8 cDNA (single-chain CD8α WT or mutants either alone or with CD8β WT) was mixed with 8 μl of lipofectamine per 100 μl of the serum-free medium Optimem (Life Technologies, Grand Island, NY) for 30 min at room temperature. The lipofectamine mixture was then added to several 35-mm dishes of nearly confluent COS-7 cells. Transfection was stopped after 18 h by replacing the lipofection mix with 2 ml of fresh medium containing 10% FCS. Cells were fed again after 24 h, and, after an additional 24 h, the dishes were either analyzed for cell surface expression of CD8 or used in the adhesion assay.

FACS analysis of cell surface molecules

For the cell surface expression of CD8α (including single-chain constructs) and CD8$\alpha\beta$, COS-7 transfectants were stained with primary mAbs OKT8 (Coulter, Westbrook, ME), which recognizes an epitope on CD8α, and 2ST85H7 (Immunotech, Westbrook, ME), which recognizes a conformational epitope formed by CD8$\alpha\beta$.

Adhesion assay of class I positive cells to CD8 transfected COS-7

The assay used was a modification of the method previously described in detail (4). CD8 transfectants were tested for their ability to bind to an MHC class I positive B cell line, UC (16). The HLA expressed by the UC cells were A1, A2, B5, B57, Cw4, DR7, DQ2, and DQ3. These cells constitutively expressed the firefly luciferase gene under the control of the CD8α promoter (17), a property that was utilized to measure binding. Transfected COS-7 cells were washed once with PBS, and 10^7 UC cells were added to each 35-mm dish. The cells were incubated for 1 h at 37°C, and the UC cells were aspirated off. After several washes, the number of bound cells was determined by measuring the amount of luciferase activity in the cell extract.

Results

The purpose of this study was to determine, for the CD8$\alpha\beta$ heterodimer, whether the CD8α or CD8β Ig domain was making the majority of contact with MHC class I. We first tested whether a CD8$\alpha\alpha$ homodimer with only one functional R4 residue could bind to MHC class I, as predicted, from the cocrystal structure of CD8$\alpha\alpha$/MHC class I. In this structure, only one CD8α subunit made contact with the α2 domain and β_2-microglobulin of MHC class I (e.g., through residues R4 or L25, respectively). To test this, we made a single-chain CD8$\alpha\alpha$ in which we could precisely control the composition of the Ig domains in the dimer. If we had simply cotransfected wild-type and mutant forms of CD8α, a mixture of dimers on the cell surface (α^{WT}/α^{WT}, α^{MT}/α^{WT}, α^{WT}/α^{MT}, α^{MT}/α^{MT}) would have resulted. Four different single-chain constructs were made (Fig. 2*B*), consisting of either two wild-type CD8α Ig domains, two mixed CD8α homodimers consisting of one wild-type Ig domain and one with the R4K mutation (in either domain), or a double mutant in which both CD8α Ig domains contained the R4K mutation. The spacer used to link the two Ig domains of CD8 consisted of four repeating units of GGGGS, as these residues have previously been shown to provide a linker with good flexibility (18–20). The constructs were expressed as a chimeric protein linked to CD4 without domains 1 and 2.

Single-chain CD8αα can support binding of MHC class I positive cells

Expression of a single-chain CD8$\alpha\alpha$ was tested by transfection into COS-7 cells. The cells were stained with a panel of four Abs against CD8α: OKT8, G10.1, Leu-2a, and 66.2. All Abs tested were found to bind in a manner similar to the wild-type

Downloaded from www.jimmunol.org on July 27, 2011

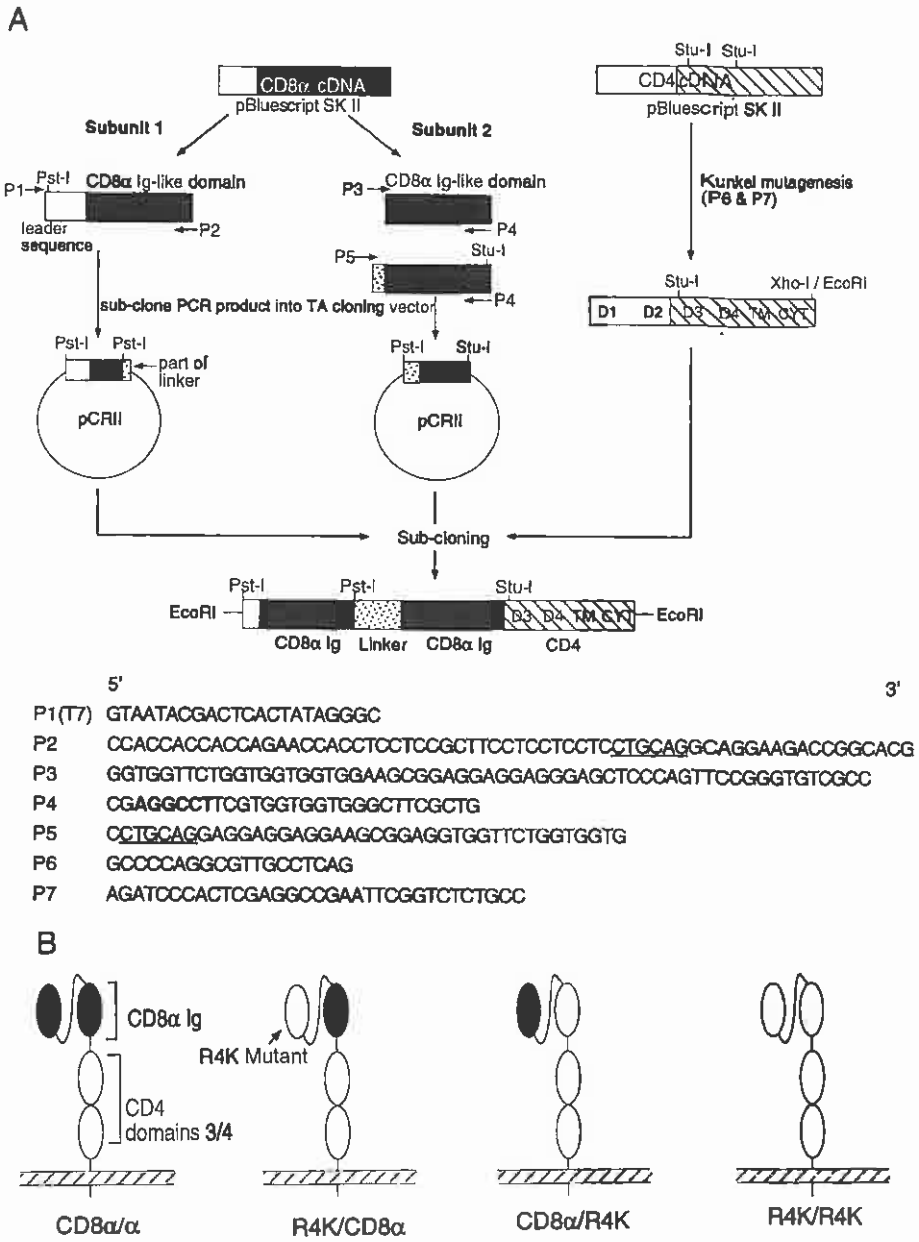


FIGURE 2. Single-chain CD8αα synthesis. *A*, Outline of the steps involved in synthesis of single-chain construct CD8αα (see *Materials and Methods* for details). P1-P7 indicates the seven primers used in either PCR or mutagenesis reactions, all of which are shown below. *B*, Schematic representation of the different single-chain constructs made.

homodimer (results not shown), indicating that no major structural changes had occurred in the Ig domains. Mutant forms of the single chain were used in a cell-cell binding assay in which we compared the binding of class I positive cells to COS-7 cells expressing either wild-type single chain, single chain with one wild-type domain and one mutant domain in which R4 was mutated to a lysine (R4K), or a double mutant as outlined in Fig. 2*B*. Expression of all forms of the single chain were similar (Fig. 3*A*); therefore, any differences in binding could not be attributed to differences in expression levels. Results shown in Fig. 3*B* demonstrate that the only construct that did not bind to class I positive cells was the double mutant form of single-chain CD8. On the other hand, a single mutant CD8α in either position in the single chain had no major effect on binding. Therefore, one functional R4 residue is sufficient for the binding of class I positive cells in this assay.

To ensure that this effect was not unique to the single-chain CD8αα, we cotransfected cells with equal amounts of wild-type CD8α and mutant (R4K) CD8α and compared the binding to cells expressing either wild-type CD8αα or double mutant CD8αα alone. The expression of the different CD8αα dimers was similar, as shown in Fig. 4*A*. Cotransfection of wild-type CD8α and mutant CD8α results in a number of different dimers on the cell surface: WT/WT, R4K/WT, WT/R4K and R4K/R4K. From the single-chain results we would predict that only one of these forms (R4K/R4K) would not be able to bind to class I. Therefore, we expected the binding of the mutant/wild-type mix in this case to be ~75% of wild-type homodimer (Fig. 4*B*). The binding of class I positive

Downloaded from www.jimmunol.org on July 27, 2011

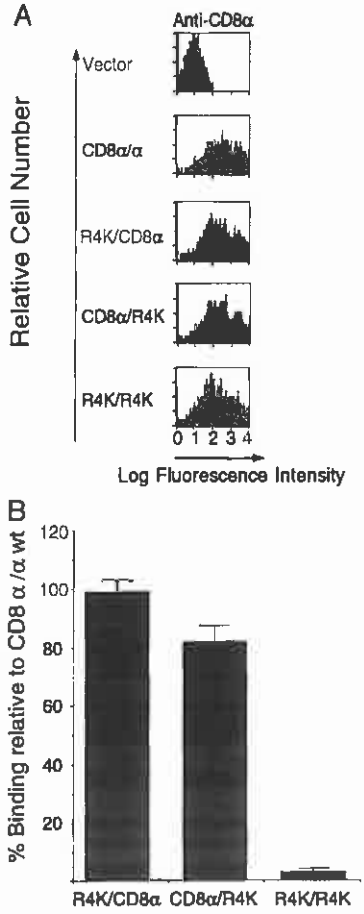


FIGURE 3. Effect of single R4 mutation in single-chain CD8$\alpha\alpha$ on MHC class I-mediated binding. *A*, Transient expression of single-chain CD8$\alpha\alpha$ constructs on COS-7 cells. Shown are: a representative flow cytometric profile from one experiment demonstrating that the level of expression of all the single-chain constructs was similar, and a comparison of these to cells transfected with vector alone as a control. *B*, Cell-cell adhesion assay determining binding of MHC class I positive UC cells to COS-7 cells expressing the various forms of single-chain CD8$\alpha\alpha$. Results from three experiments (each of which had triplicate samples) were averaged and expressed relative to the CD8$\alpha\alpha$ single-chain binding group, which on average gave a binding value of 20 times above background. Error bars represent the SE from each group.

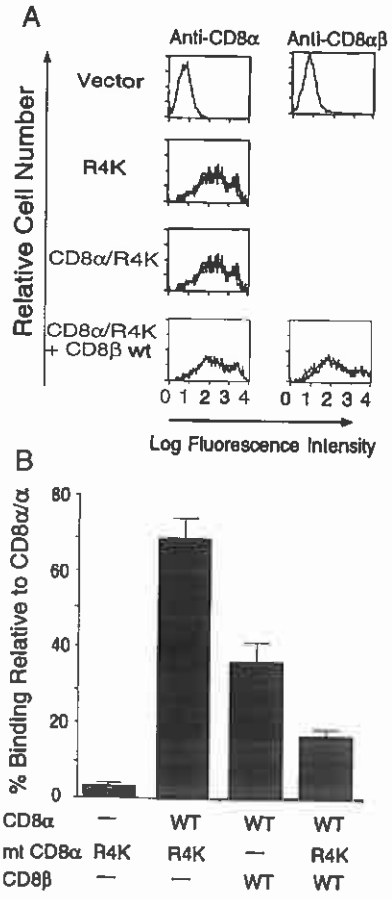


FIGURE 4. Effect of R4K mutant in presence of wild-type CD8$\alpha\alpha$ and CD8$\alpha\beta$. *A*, Flow cytometric analysis of CD8 expression. *Top panels* are negative controls in which only vector DNA has been transfected. In all other profiles, the fine lines represent cells transfected with wild-type CD8α and CD8β as a positive control. The thicker lines are the expression of mutant forms of CD8α either alone or cotransfected with wild-type CD8β. *B*, Cell-cell adhesion assay to determine whether R4K mutant inhibits binding in the presence of wild-type CD8α. COS-7 cells were transfected with equal amounts of mutant (R4K) and wild-type CD8α (total of 1 μg of CD8α) either alone or together with 1 μg of wild-type CD8β. Results from three experiments (each of which had triplicate samples) were averaged and expressed relative to the CD8$\alpha\alpha$ WT binding group, which on average gave a binding value of 35 times above background. Error bars represent the SE from each group.

cells to COS-7 cells expressing R4K/WT CD8α was found to be 69 ± 12.9% of that of wild-type CD8$\alpha\alpha$ homodimers.

Majority of contacts of CD8 with MHC class I is through the CD8α, and not the CD8β, domain

To determine the contribution that CD8β makes to the interaction, we tested the possibility that CD8β would make similar contacts as the CD8α-2 subunit. To do this, we cotransfected wild-type CD8β with either R4K or L25A CD8α. Cell surface expression of CD8 was tested with two Abs, one against CD8α and the other against an epitope formed by CD8$\alpha\beta$. While CD8β cannot come to the surface on it own (21), it is possible that CD8$\alpha\alpha$ homodimers and CD8$\alpha\beta$ heterodimers could be expressed on the cell surface. Therefore, staining of both CD8α and CD8$\alpha\beta$ was compared to ensure that the predominant form of CD8 expressed on the cell surface was that of the CD8$\alpha\beta$ heterodimer, a finding that has previously been quantified by Scatchard analysis (11). Fig. 5*A* demonstrates that this is the case, as the level of CD8α expression is similar to that of CD8$\alpha\beta$. In addition, the level of expression for wild-type CD8$\alpha\beta$ is similar to that of R4K/CD8β and L25A/CD8β, indicating that differences in binding are not due to differences in expression level. In the cell-cell adhesion assay, our prediction was that the binding would only be reduced if CD8β was not contacting the α2 domain of class I. As shown in Fig. 5*B*, binding of class I positive cells to COS-7 cells expressing R4K CD8α/WT CD8β or L25A CD8α/WT CD8β was the same as to cells expressing only double mutant CD8$\alpha\alpha$ homodimers (i.e., little or no binding). Therefore, binding was not rescued in the presence of wild-type CD8β.

To rule out the possibility that the presence of mutant CD8$\alpha\alpha$ homodimers had a dominant effect on binding, wild-type CD8α was added to the mix of wild-type CD8β and mutant CD8α DNA that was transfected into COS-7 cells. In this case, we were able to detect binding, even though some mutant CD8$\alpha\alpha$ homodimers would still be present. Fig. 4*A* shows the expression levels of the various transfectants, which indicate that the levels of expression

Downloaded from www.jimmunol.org on July 27, 2011

151

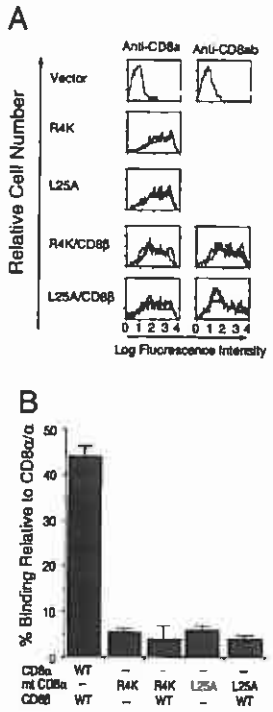


FIGURE 5. Effect of single R4 or L25 mutation in CD8$\alpha\beta$ heterodimer on MHC class I-mediated binding. *A*, Flow cytometric representation of expression of wild-type CD8β together with mutant forms of CD8α (either R4K or L25A). *Top panels* are negative controls in which only vector DNA has been transfected. In all other profiles the fine lines represent cells transfected with wild-type CD8α and CD8β as a positive control. The thicker lines are the expression of mutant forms of CD8α either alone or cotransfected with wild-type CD8β. *B*, Cell-cell adhesion assay to determine whether CD8β can rescue binding in the presence of mutant CD8α by substituting for the CD8α-1 subunit and making contact with β_2-microglobulin or α2 domain of MHC class I. Results from three experiments (each of which had triplicate samples) were averaged and expressed relative to the CD8$\alpha\alpha$ WT binding group. Error bars represent the SE from each group.

were similar and, therefore, any differences in binding could not be attributed to variations in expression level. Binding of CD8α WT/R4K/CD8β was 46 ± 5% of wild-type CD8$\alpha\beta$, as shown in Fig. 4*B*. This is what we would predict since only half of CD8$\alpha\beta$ expressed on the cell surface would be capable of binding to MHC class I if CD8β could not contact the α2 domain of MHC class I.

Discussion

Analysis of CD8 protein interactions has, in general, been performed using the CD8$\alpha\alpha$ form of CD8. However, there is a great deal of interest in studying the interactions of CD8$\alpha\beta$ with other proteins since this is the major form expressed on class I-restricted, mature T cells. The interaction of CD8 with MHC class I occurs when CD8 functions either as a coreceptor with the TCR or independent of the TCR as in the veto effect (22, 23). The cocrystal structure of CD8$\alpha\alpha$-MHC class I showed that the interaction with CD8 was asymmetric with one Ig domain contributing ~70% of the surface area buried by CD8$\alpha\alpha$ on complex formation. This domain makes contact with residues located in the α2 and α3 domains of MHC class I, as well as with β_2-microglobulin. Assuming that CD8$\alpha\beta$ would bind in a similar manner, we used a mutational approach to determine whether CD8α or CD8β made most of the contact with MHC class I. Based on our previous mutational studies and the cocrystal structure, we mutated residues on the CD8α Ig domain (R4 and L25) that could interact with the α2 domain of MHC class I and β_2-microglobulin, respectively, and we determined their effect when coexpressed with wild-type CD8β. The consequences of these mutations would be a severe reduction of CD8$\alpha\beta$/MHC class I binding or no effect, depending on the orientation of the CD8α Ig domain relative to the CD8β Ig domain. We found the former to be the case, and, therefore, our work supports the model in which the CD8α Ig domain is contacting both the α2 and α3 domains of MHC class I, whereas the CD8β Ig domain is only contacting the α3 domain.

This conclusion is consistent with respect to the previously published studies on CD8-MHC class I interaction. Biacore studies by Garcia et al. (12) found no major difference in affinity of CD8$\alpha\alpha$ or CD8$\alpha\beta$ for MHC class I, and Sun et al. (11) demonstrated that CD8β did not enhance the association of CD8 for MHC class I as had been speculated. One might expect the affinities of CD8$\alpha\alpha$ vs CD8$\alpha\beta$ to be similar if it were the CD8α Ig domain for both homodimers and heterodimers making the most contact. It is known that CD8$\alpha\beta$ functions as a better coreceptor (9, 10), a finding that in part was attributed to stronger affinity for MHC class I; however, this is unlikely to be the case.

We found that the binding of R4K/CD8α single chain was slightly higher than that of CD8α/R4K. This may be due to the different structural constraints of the two Ig domains of the single chain. The Ig domain at the N terminus of the protein that is connected to the rest of the protein by a peptide linker may be more flexible and could more readily adopt the α-2 role and bind at a position further from the membrane. The other Ig domain linked to domain 3 of CD4 appears to be the one more likely to make contact with both the α2 and α3 domains of MHC class I because it is linked to CD4, and, to adopt the α-2 subunit role, it would have to "reach" further from the cell membrane. An alternative, although not mutually exclusive, hypothesis may be the effect of the different position of the linker peptide relative to MHC class I when the different subunits make contact with the α2 domain of MHC class I (i.e., the linker peptide may be more likely to interfere in one position vs another).

There are multiple approaches for studying the orientation of two molecules relative to each other. For instance, to study TCR-MHC interaction, cocrystals were generated (24, 25), and mutations of the TCR (26) and peptide (27) have helped to elucidate the orientation of the TCR-MHC/peptide interaction. In addition, complementary mutations in CD2 and CD48 were made that restored binding or function in comparison to the affect of each mutation alone (28). These different approaches are complementary. Even a cocrystal represents the most efficient way that two molecules pack to form a crystal. This may not completely reflect what is happening on the cell surface, particularly if there are other molecules involved in the interaction that are not part of the cocrystal.

The orientation of CD8$\alpha\beta$ to MHC class I supported by the data we have generated has implications for potential binding between CD8 and the TCR. Evidence for association between coreceptor and the TCR was initially described for CD4 based on cocapping studies (29, 30). More recent cocapping studies with CD8 and the TCR showed that anti-CD8β Abs were significantly more efficient than anti-CD8α Abs at inducing cocapping of the TCR (31). This suggested to Lim et al. (31) that Abs to the CD8β polypeptide may preferentially promote a conformation of CD8 that stabilizes an association with the TCR, independent of their binding to MHC (31). In designing experiments to test this hypothesis, knowing the orientation of CD8$\alpha\beta$ to the TCR is helpful. Our results provide strong support for the orientation of CD8$\alpha\beta$ relative to MHC class I, with the CD8α Ig domain providing the majority of the contact. The CD8β stalk is shorter and yet it must reach further from the

Downloaded from www.jimmunol.org on July 27, 2011

membrane to contact MHC class I. As a result, this may constrain the orientation of CD8 and the TCR facilitating interaction between the CD8 associated tyrosine kinase $p56^{lck}$ and the TCR ζ-chain.

Acknowledgments

We thank Dr. L. Kieffer for critical reading of the manuscript, Johnny Kwon for his help with Fig. 1, and Victoria Aitken for help with all other figures.

References

1. Moebius, U., G. Kober, A. L. Griscelli, T. Hercend, and S. C. Meuer. 1991. Expression of different CD8 isoforms on distinct human lymphocyte populations. *Eur. J. Immunol. 21:1793.*
2. Terry, L. A., J. P. DiSanto, T. N. Small, and N. Flomenberg. 1990. Differential expression and regulation of the human CD8α and CD8β chains. *Tissue Antigens 35:82.*
3. Salter, R. D., R. J. Benjamin, P. K. Wesley, S. E. Buxton, T. P. J. Garrett, C. Clayberger, A. M. Krensky, A. M. Norment, D. R. Littman, and P. Parham. 1990. A binding site for the T-cell co-receptor CD8 on the α3 domain of HLA-A2. *Nature 345:41.*
4. Giblin, P. A., D. J. Leahy, J. Mennone, and P. B. Kavathas. 1994. The role of charge and multiple faces of the CD8α/α homodimer in binding to major histocompatibility complex class I molecules: support for a bivalent model. *Proc. Natl. Acad. Sci. USA 91:1716.*
5. Sanders, S. K., R. O. Fox, and P. Kavathas. 1991. Mutations in CD8 that affect interactions with HLA class I and monoclonal anti-CD8 antibodies. *J. Exp. Med. 174:371.*
6. Sun, J., D. J. Leahy, and P. B. Kavathas. 1995. Interaction between CD8 and MHC class I mediated by multiple contact surfaces that include the α2 and α3 domains of MHC class I. *J. Exp. Med 182:1275.*
7. La Face, D. M., M. Vestberg, Y. Yang, R. Srivastava, J. DiSanto, N. Flomenberg, S. Brown, L. A. Sherman, and P. A. Peterson. 1995. Human CD8 transgene regulation of HLA recognition by murine T cells. *J. Exp. Med 182:1315.*
8. Gao, G. F., J. Tormo, U. C. Gerth, J. R. Wyer, A. J. McMichael, D. I. Stuart, J. I. Bell, E. Y. Jones, and B. K. Jakobsen. 1997. Crystal structure of the complex between human CD8$\alpha\alpha$ and HLA-A2. *Nature 387:630.*
9. Wheeler, C. J., P. von Hoegen, and J. R. Parnes. 1992. An immunological role for the CD8 β chain. *Nature 357:247.*
10. Karaki, S., M. Tanabe, H. Nakauchi, and M. Takiguchi. 1992. β-chain broadens range of CD8 recognition for MHC class I molecule. *J. Immunol. 149:1613.*
11. Sun, J., and P. B. Kavathas. 1997. Comparison of the roles of CD8$\alpha\alpha$ and CD8$\alpha\beta$ in interaction with MHC class I. *J. Immunol. 159:6077.*
12. Garcia, K. C., C. A. Scott, A. Brunmark, F. R. Carbone, P. A. Peterson, I. A. Wilson, and L. Teyton. 1996. CD8 enhances formation of stable T-cell receptor/MHC class I molecule complexes. *Nature 384:577.*
13. Brown, M. H., and A. N. Barclay. 1994. Expression of immunoglobulin and scavenger receptor superfamily domains as chimeric proteins with domains 3 and 4 of CD4 for ligand analysis. *Protein Eng. 7:515.*
14. Takebe, Y., M. Seiki, J. Fujisawa, P. Hoy, K. Yokota, K. Arai, M. Yoshida, and N. Arai. 1988. SRα promoter: an efficient and versatile mammalian cDNA expression system composed of the simian virus 40 early promoter and the R-U5 segment of human T-cell leukemia virus type 1 long terminal repeat. *Mol. Cell. Biol. 8:466.*
15. Kunkel, T. A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA 82:488.*
16. Glassy, M. C., H. H. Handley, H. Hagiwara, and I. Royston. 1983. UC 729-6, a human lymphoblastoid B-cell line useful for generating antibody-secreting human-human hybridomas. *Proc. Natl. Acad. Sci. USA 80:6327.*
17. Gao, M.-H., and P. B. Kavathas. 1993. Functional importance of the cyclic AMP response element-like decamer motif in the CD8α promoter. *J. Immunol. 150:4376.*
18. Huston, J. S., D. Levinson, M. Mudgett-Hunter, M. S. Tai, J. Novotny, M. N. Margolies, R. J. Ridge, R. E. Bruccoleri, E. Haber, R. Crea, and H. Oppermann. 1988. Protein and engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA 85:5879.*
19. Mage, M. G., L. Lee, R. K. Ribaudo, M. Corr, S. Kozlowski, L. McHugh, and D. H. Margulies. 1992. A recombinant, soluble, single-chain class I major histocompatability complex molecule with biological activity. *Proc. Natl. Acad. Sci. USA 89:10658.*
20. Mottez, E., C. Jaulin, F. Godeau, J. Choppin, J.-P. Levy, and P. Kourilsky. 1991. A single-chain murine class I major transplantation antigen. *Eur. J. Immunol. 21:467.*
21. DiSanto, J. P., R. W. Knowles, and N. Flomenberg. 1988. The human Lyt-3 molecule requires CD8 for cell surface expression. *EMBO J. 7:3465.*
22. Hambor, J. E., M. C. Weber, M. L. Tykocinski, and D. R. Kaplan. 1990. Regulation of allogeneic responses by expression of CD8 α chain on stimulator cells. *Int. Immunol. 2:879.*
23. Miller, R. G. 1980. An immunological suppressor cell inactivating cytotoxic T lymphocyte precursor cells recognizing it. *Nature 287:544.*
24. Garcia, K. C., D. Massimo, R. L. Stanfield, A. Brunmark, M. R. Jackson, P. A. Peterson, L. Teyton, and I. A. Wilson. 1996. An $\alpha\beta$ T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. *Science 274:209.*
25. Garboczi, D. N., P. Ghosh, U. Utz, Q. R. Fan, W. E. Biddison, and D. C. Wiley. 1996. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature 384:134.*
26. Hong, S.-C., D. B. Sant'Angelo, B. N. Dittel, R. Medzhitov, S. T. Yoon, P. G. Waterbury, and C. A. Janeway, Jr. 1997. The orientation of a T cell receptor to its MHC class II:peptide ligands. *J. Immunol. 159:4395.*
27. Sant'Angelo, D. B., G. Waterbury, P. Preston-Hurburt, S. T. Yoon, R. Medzhitov, S.-C. Hong, and C. A. Janeway. 1996. The specificity and orientation of a TCR to its peptide-MHC class II ligands. *Immunity 4:367.*
28. van der Merwe, P. A., P. N. McNamee, E. A. Davies, A. N. Barclay, and S. J. Davis. 1995. Topology of the CD2-CD48 cell-adhesion molecule complex: implications for antigen recognition by T cells. *Curr. Biol. 5:74.*
29. Dianzani, U., A. Shaw, B. K. Al-Ramadi, R. T. Kubo, and C. J. Janeway. 1992. Physical association of CD4 with the T cell receptor. *J. Immunol. 148:678.*
30. Rojo, J. M., K. Saizawa, and C. A. J. Janeway. 1989. Physical association of CD4 and the T cell receptor can be induced by anti-T-cell receptor antibodies. *Proc. Natl. Acad. Sci. USA 86:3311.*
31. Lim, G.-E. K., L. McNeill, K. Whitely, D. L. Becker, and R. Zamoyska. 1998. Cocapping studies reveal CD8/TCR interactions after capping CD8β polypeptides and intracellular associations of CD8 with p56lck. *Eur. J. Immunol. 28:745.*

Downloaded from www.jimmunol.org on July 27, 2011

51 Casali, P. and Notkins, A.L. (1989) Probing the human B cell repertoire with EBV: polyreactive antibodies and CD5$^+$ B lymphocytes. *Annu. Rev. Immunol.* 7, 513–525
52 Kipps, T. and Carson, D.A. (1993) Autoantibodies in chronic lymphocytic leukemia and related systemic autoimmune diseases. *Blood* 81, 2475–2487
53 Diehl, L.F. and Ketchum, L.H. (1998) Autoimmune disease and chronic lymphocytic leukemia: autoimmune hemolytic anemia, pure red cell aplasia, and autoimmune thrombocytopenia. *Semin. Oncol.* 25, 80–97
54 Fehr, T. *et al.* (1996) T-independent activation of B cells by vesicular stomatitis virus: no evidence for the need of a second signal. *Cell. Immunol.* 168, 184–192

Molecular interactions of coreceptor CD8 and MHC class I: the molecular basis for functional coordination with the T-cell receptor

George F. Gao and Bent K. Jakobsen

In recent years, substantial progress has been made towards understanding the molecular basis for CD8 binding to class I MHC and the coreceptor's role in cytotoxic T-cell activation. Here, we review the structural, mechanistic and functional studies that point to a model of coordination of T-cell receptor and CD8 signaling that might provide the key to cytotoxic T-cell activation.

At the molecular level, T-cell recognition of peptide–MHC (pMHC) on antigen presenting cells (APCs) is orchestrated by a number of receptor–ligand or receptor–counterreceptor interactions[1–3], all contributing to formation of the so-called immunological synapse[4,5]. At least three classes of protein–protein interaction are involved in this process: (1) unique T-cell receptors (TCRs; TCR αβ heterodimers and CD3 complex) interact specifically with their polymorphic ligands, class I or class II pMHC; (2) T-cell coreceptors CD8 (αα or αβ dimer) or CD4 bind their ligand pMHC (class I and class II, respectively); (3) costimulatory receptors (for example, CD28 and CD152) and adhesion molecules (such as CD2) interact with their ligands or counterreceptors (for example, CD80, CD86 for CD28, CD152 respectively, and CD58/48 for CD2). Further molecules playing a part in this process are still being identified, for example, the recently discovered DC-SIGN (dendritic cell-specific ICAM-3 grabbing nonintegrin), an adhesion molecule expressed specifically on the surface of dendritic cells ('professional' APCs), which is a ligand for ICAM-3 on T-cells[6] and is complicated in HIV envelope binding and infection[7].

CD4 and CD8 augment the TCR–pMHC interaction *in vivo*[8,9]. The effect of the coreceptor is qualitative in T-cell activation but does not affect the specificity of TCR–pMHC interaction. Fewer studies have focused on the interactions between coreceptors (CD4 and CD8) and pMHC than on TCR–pMHC binding[2] and significant discrepancies exist between the *in vitro* data generated in different studies using soluble molecules[10–12].

Despite their apparently similar functions, the CD8 and CD4 proteins exhibit little structural similarity in terms of interactions with their ligands. CD8 is expressed on the T-cell surface as either an αα homodimer or an αβ heterodimer, whereas CD4 is expressed as a monomeric polypeptide. However, it should be noted that a CD4 dimer was observed in a recent crystal structure[13] and an oligomerization model was also proposed[14]. The *in vivo* significance of these observations has yet to be established.

The present data support the notion that T-cell activation requires that CD8/CD4 and TCR bind to the same pMHC molecule[1], leading to the classification of CD8 and CD4 as coreceptors. However, it is still speculated that the peptide antigen-indiscriminate binding of CD8 to pMHC molecules on the APC surface could also be involved in augmenting T-cell responses as overexpression of human CD8αα and HLA class I on the surface of the transfected cells led to their strong interaction (adhesion)[15].

Structural basis of the CD8–pMHC interaction in humans and mice

CD8, either α chain or β chain, consists of four discrete functional domains that can be related to the primary sequence as follows: the

0167-5699/01/$ – see front matter © 2000 Elsevier Science Ltd. All rights reserved.
PII: S0167-5699(00)01750-3

Ig-like ectodomain, the membrane proximal stalk (hinge) region, the transmembrane domain and the cytoplasmic domain (Fig. 1). The Ig-like domain is involved in the binding to MHC. Recently, this interaction has been characterized in detail by the complex crystal structures of both human and mouse CD8αα Ig-like domain with the MHC class I molecules HLA-A2 and H-2K^b, respectively[16,17]. Surprisingly, and refuting a previously proposed model[18], one CD8αα homodimer binds one MHC molecule in a manner similar to antibody–antigen binding. The crystal structures rule out the stoichiometric possibility that one CD8αα homodimer cross-links two MHC molecules (Fig. 2), which had been considered might play a role in T-cell activation. CD8 binding involves contacts to the α2, α3 and β_2m domains of the pMHC molecule. A striking feature of the CD8αα–pMHC complex is that the two subunits in the CD8αα bind pMHC asymmetrically, one CD8α subunit contributing approximately 75% of the contacts involved.

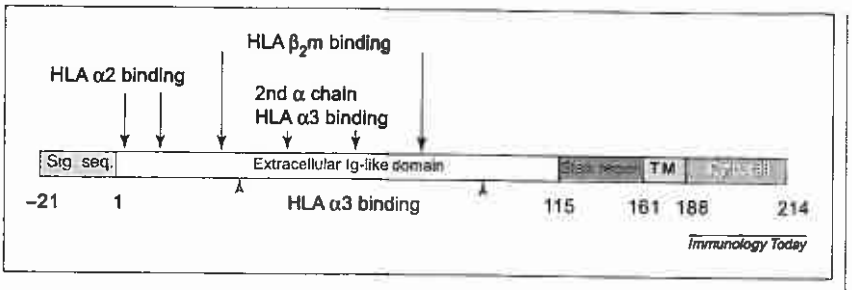


Fig. 1. *Schematic functional structure of the human CD8α. The CD8α primary structure consists of signal sequence (sig. seq.), extracellular Ig-like domain, stalk region, transmembrane domain (TM) and cytoplasmic domain (cyto. tail). The extracellular Ig-like domain is involved in the binding to MHC. 'HLA α2 binding' indicates the hydrogen bonds between amino acid R4, K21 of CD8α and the α2 domain of HLA-A2 heavy chain. 'HLA β_2m binding' indicates the hydrogen bonds between amino acid V24, D75 of CD8α and β_2-microglobulin. '2nd α chain HLA α3 binding' indicates the hydrogen bonds between amino acid S34, Y51 of the distal CD8α subunit (colored blue in Fig. 2) and the α3 domain of HLA-A2 heavy chain. 'HLA α3 binding' (red solid arrows) indicates the hydrogen bonds between amino acids cluster 1 (S27, N28, T30, Q54), cluster 2 (N99, S100) of CD8α and α3 domain of HLA-A2 heavy chain, the major CD8α-binding domain of HLA-A2. The stalk region, also called hinge region, is flexible and highly glycosylated (O-linked), which is believed important to extend the region to reach the MHC. It is also postulated to interact with TCR. The cytoplasmic domain of CD8α consists of a p56^{lck} binding motif and important for signal transduction.*

In both the human and mouse CD8αα–pMHC complexes, the MHC heavy chain α3 domains are shifted compared with the position they assume in the unliganded MHC (Fig. 2). This conformational shift appears to be necessary to facilitate CD8αα binding to the MHC heavy chain α3 domain. A similar α3 domain movement is also seen in the TCR–pMHC complex structures[19,20]. Intriguingly, in the CD8αα–pMHC crystal structures, the human and murine α3 domains shift in opposite directions, with the human HLA α3 domain being 'pulled' in the direction towards the T-cell, whereas the murine domain is being 'pushed' away. The significance, if any, of this discrepancy remains to be determined. The crystal structures also indicate that the binding of CD8αα to pMHC does not transduce any significant changes to the TCR-binding platform of the pMHC. Instead, the CD8αα–pMHC contact may help re-orientate the whole TCR–pMHC complex on the cell surface.

Detailed comparison of the human and mouse CD8αα–pMHC complex structures reveals that, although they share the common binding mode as discussed above, some species-specific features do exist. There is more contact (over 20% more buried areas) in the mouse complex than in the human counterpart. No hydrogen bonds are formed between the T-cell proximal CD8αα subunit (colored red in Fig. 2) and the MHC α2 domain in the mouse complex[17], whereas three hydrogen bonds are found in the human complex structure[16]. The CD8 contacts to the human MHC α2 domain have been shown in cell–cell adhesion assays to be critical for the functional role of the interaction in the human complex[21]. In the mouse CD8αα–pMHC complex, hydrogen bonds are formed between the heavy chain α3 and β_2m domains and the additional N-terminal amino acids in the mouse CD8α chain[17].

The molecular architecture of the coreceptor binding conformation ensures that the specificity for pMHC is determined and dominated by the antigen-specific TCR; CD8 acting as a universal co-receptor to bind, potentially, any polymorphic pMHCs. Binding of the CD8 ectodomain appears to have no consequences for the TCR binding platform. Therefore, the crystal structures reported so far have provided no clues to help distinguish the two prevalent models for CD8 function: first, that CD8 primarily enhances T-cell recognition by binding the same MHC molecules as the TCR; or second, that the much larger numbers of CD8 contacts that could be formed with any pMHC complexes on the APC surface are involved in T-cell activation. However, it is noteworthy that early work[22–24] provided strong biological evidence for co-recognition of MHC class I by specific TCR and CD8, which supports the first model.

Kinetics and affinities of CD8–pMHC interactions

The binding of CD8αα to pMHC was initially demonstrated by using cell–cell adhesion assays[15,25,26]. In recent years, technical progress in recombinant protein technology and biophysical instrumentation has allowed direct measurements of the kinetics and affinities of both human and mouse molecules by using the surface plasmon resonance (SPR) technique[10–12,27]. SPR (measured, for example, with the commonly used instrument BIAcore®) is an optical measuring technique that measures the mass concentration of proteins (refractive index) in close proximity to a specially-prepared surface. It requires immobilization of one of the interacting molecules onto a sensor surface allowing the other binder molecules to flow through it; in this way, the real-time binding is detected. It has proved to be a reliable method for measuring the affinity and kinetics

155

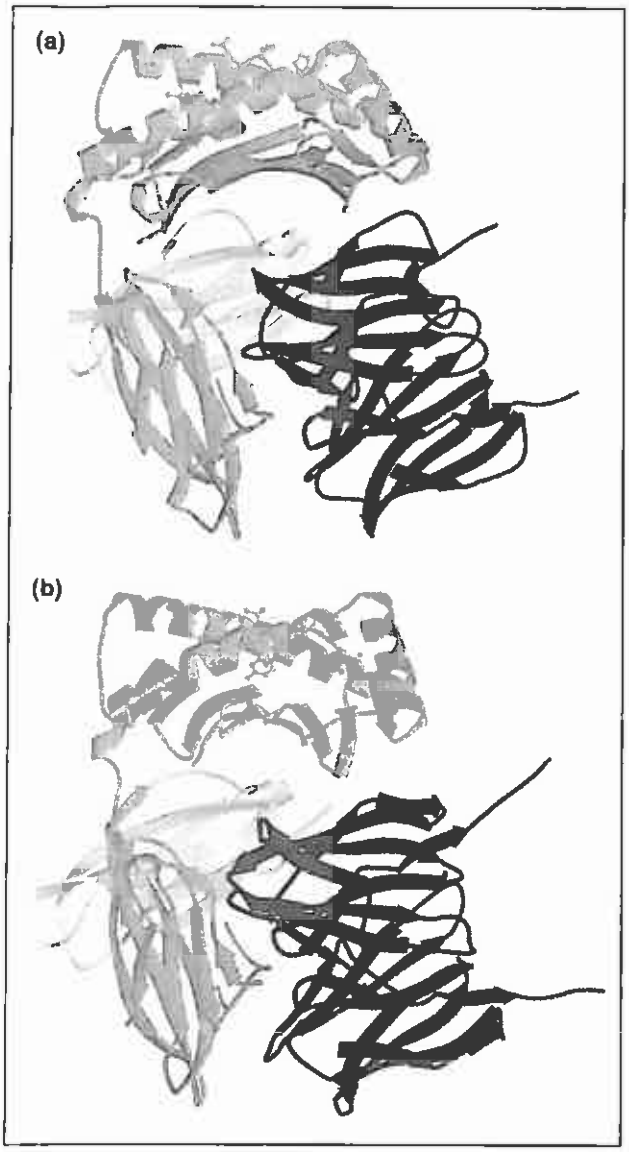


Fig. 2. *Complex structures determined by X-ray crystallography of the human CD8αα/HLA-A2/peptide (a) and the mouse CD8αα/H2-K^b/peptide (b), indicating the α3 domain shift of the MHC/HLA heavy chain upon CD8αα binding. In both human and mouse complex structures, peptides are colored in orange, heavy chain in green, β_2m in yellow, proximal CD8αα subunit in red and distal CD8αα subunit in blue. Partial structures in purple represent the relative position of heavy chain amino acids 218–230 of the unliganded HLA-A2 (a) or H-2K^b (b), in which the α1 and α2 domains of the heavy chain (unliganded structures are not shown) were superimposed very well and the α3 domain was either 'pulled' towards the CD8αα (a) or 'pushed' afar (b) upon binding in CD8αα complexed MHC. The HLA-A2 binding nonameric peptide is derived from HIV reverse transcriptase residues 309–317 (ILKEPVHGV) and the H2-K^b binding octameric peptide is derived from VSV (vesicular stomatitis virus) nucleoprotein residues 52–59 (RGYVYQGL). The peptides for complexed and unliganded HLA-A2 or H2-K^b for the structural comparison are the same, respectively. The PDB files used are as follows: unliganded HLA-A2/peptide, 1HHJ; CD8αα/HLA-A2/peptide, 1AKJ; unliganded H2-K^b/peptide, 2VAA; CD8αα/H2-K^b/peptide, 1BQH. These images were prepared using the programs MOLSCRIPT (generated by P.J. Kraulis) and Raster3D (generated by E.A. Merritt and M.E.P. Murphy).*

of very weak protein–protein interactions such as CD8–pMHC and TCR–pMHC (Ref. 28). The results for CD8–pMHC binding from these studies are summarized in Table 1. The binding of CD8 to pMHC is extremely weak [Kd = 90–220 μM (Ref. 11); 11–40 μM (Ref. 10) and 22–86 μM (Ref. 12)], with significantly lower affinities than those characterizing TCR–pMHC interactions (1–50 μM at 25°C)[2] or those characterizing typical adhesion molecule interactions (for example, CD2 and CD48, 30–50 μM at 25°C)[28]. However, the most striking feature of the CD8–pMHC interaction is its extremely fast kinetics (dissociation rate constant, $k_{off} \geq 18\ s^{-1}$; Ref. 11) compared with TCR–pMHC interactions (k_{off}, ranging from 0.01–0.1 s^{-1} for agonist peptide pMHC; Ref. 2). It is so fast that the direct measurement of the association rate constant (k_{on}) is impossible. These extremely fast kinetics have been observed in both the human and mouse CD8–pMHC complex interactions (Ref. 11 and G.F. Gao *et al.*, unpublished), although early work[10] did not indicate this. Recently, a separate study has supported the hypothesis that the mouse CD8–pMHC interaction is characterized by kinetics similar to those observed with the human molecules[12]. The transient nature of the CD8–pMHC interaction supports the notion that T-cell activation involves binding of CD8 and TCR to the same pMHC molecules (with the TCR dominating the interaction)[1]. However, this feature does not rule out CD8–pMHC binding that does not involve the TCR playing a role as well.

Thermodynamic analyses of the interactions have demonstrated that, as is the situation with natural killer (NK) cell receptor–pMHC interactions[29], the CD8–pMHC interaction can be described as a so-called 'rigid-body association'. In other words, the interface of CD8 and MHC interaction in the complex compared with the unliganded form does not show significant conformational changes: they are both in ready-binding forms. By contrast, TCR–pMHC binding falls into the 'induced-fit' category[30,31]. It is notable that there is no significant difference in affinity between CD8αα homodimer and CD8αβ heterodimer to pMHC (Refs 10, 12), the experimental values being only slightly higher for the latter (only the murine data are available). Indeed, this difference could be even explained by the range of measurement error in SPR methods. This is in contrast to other studies when CD8αα or αβ expressed on the cell surface as discussed below. The carbohydrates and the stalk region on CD8 do not affect the overall affinity and kinetics in the SPR measurement[12].

Soluble CD8 and its peptide derivatives are potent CTL inhibitors

Early work showed that peptides corresponding to the putative CD8/CD4-binding region of MHC/HLA could inhibit T cell–APC adhesion[32,33]. Recently, our group and others have shown that both human and mouse recombinant soluble CD8αα Ig-like domain

Table 1. Summary of CD8–pMHC affinities measured to date at 25°C

MHC alleles	Kd (μM) CD8αα	CD8αβ	Ref.	Soluble CD8 used
Human HLA				
HLA-A2	130 ± 30		11	Ig-like domain 1–120, expressed in *Escherichia coli*
HLA-B35	130 ± 10		39	Same as above
HLA-Cw7	220 ± 20		39	
HLA-A68	≥ 1000		39	
HLA-B48	≥ 1000		39	
HLA-G1	160 ± 3		39	
HLA-E	≥ 1000		39	
Mouse H2				
H2-L^d		11.2 ± 0.8	10	Ig-like domain plus some of stalk region, 1–156 for α chain, 1–146 for β chain; expressed in insect cells
H2-D^d		14.1 ± 0.4	10	Same as above
H2-K^b	30–40	12–14	10	Same as above
	35–86	22–30	12	Ig-like domain 1–122 for α chain, 1–115 for β chain, Ig-like domain plus stalk region, 1–165 (α), 1–149 (β), Expressed in mammalian CHO and Lec3.2.8.1 cells

homodimers, or a synthetic cyclic peptide consisting of murine CD8α residues 54–59, inhibit lysis by cytotoxic T lymphocytes (CTLs)[12,34,35]. Strikingly, the inhibition observed with soluble CD8αα is more potent than that observed with an anti-CD8α monoclonal antibody[35]. The inhibitory effect of soluble CD8 on T cells is remarkable because the binding affinity of CD8αα–pMHC is very low and it is estimated that, at the concentrations employed, only a few percent of the APC pMHC complexes are blocked[35]. It has also been reported that soluble recombinant mouse CD8αβ exerts a similar inhibitory effect on CTL activation to human CD8αα (Ref. 12). Examination of early tyrosine phosphorylation events in the presence of soluble human CD8αα showed a negating interference with the signaling competence of most TCRs (Ref. 35). Similar negating signaling phenomena have been observed with naturally occurring viral altered peptide ligand (APL) antagonists[36]. It is possible therefore that the mechanism of CTL inhibition by soluble CD8αα and antagonism by APLs may be similar. These observations provide clues to the links between TCR and CD8 in CTL activation, indicating that a close coordination of the signals transduced by the antigen receptor and the coreceptor is essential for full activation.

MHC/HLA polymorphism and CD8 interactions

The genes encoding human and mouse MHC molecules (heavy chains for class I) are the most polymorphic described to date, whereas CD8 is a non-polymorphic molecule displaying a high degree of cross-species conservation. The crystal structures of the human and mouse CD8αα–pMHC complexes show that the contact amino acid residues are all conserved in different alleles in these two species, respectively.

Despite the invariant nature of CD8αα–pMHC interactions observed in the complex structure in terms of the direct contact amino acids, cell–cell adhesion assays have indicated that HLA-A68 and B48 are unable to function as ligands for CD8αα (Refs 25, 37), though the majority of the alleles tested (including HLA-A, -B and -C) bound to CD8αα (Refs 15, 25, 26, 38). Similarly, the immunological function and/or cellular distribution of the 'non-classical' class I molecules indicate that some of these may not be involved in CD8 interactions either. Structural comparisons and SPR binding analyses of human CD8αα with different classical and non-classical class I molecules[39] showed that most HLA class I molecules, including the non-classical HLA-G, do bind to CD8αα with a similar affinity (Kd = 90–220 μM), this interaction (HLA-G and CD8αα) was clearly observed in cell–cell adhesion experiments earlier[38]. However, HLA-A68, HLA-B48 and the non-classical HLA-E bind CD8αα extremely weakly (Kd ≥ 1 mM), which biologically might be defined as non-binders. The weaker affinities of CD8 for the latter group of HLA molecules are the result of naturally occurring polymorphisms in the α3 domain in the gene products of these alleles. Each of the weak-binding HLA molecules contains polymorphisms that distort the protruding loop of the α3 domain, although the polymorphic residues are not directly involved in contact with CD8 (Ref. 16). The critical residues are a single amino acid polymorphism in A68/B48 and three amino-acid polymorphisms in HLA-E (Ref. 39).

Binding affinities similar to those observed with the human proteins have been obtained with different murine alleles tested (Table 1). However, it is speculated that murine H-2K^d and non-classical Qa-2 might be CD8αα non-binders, equivalent to HLA-A68/B48, the lower affinity caused by a single polymorphism at amino acid position 196 in the α3 domain of heavy chain[17]. Non-classical CD1 binding to CD8 has not yet been observed although the overall structure of CD1 is very close to that of classical MHC molecules[40].

CD8 enhancement of T-cell recognition

CD8 enhances T-cell recognition of pMHC on the surface of APC (Ref. 1), an effect for which the molecular and mechanistic basis is

not yet fully established. Experiments involving soluble mouse CD8, TCR and MHC molecules have suggested that CD8 enhances binding of TCR to the pMHC complex[10]. However, this effect could not be reproduced in similar experiments involving human proteins[11] and is difficult to harmonize with the structural data that show no signs of conformational changes being transmitted to the TCR binding surface by CD8–pMHC contacts[16,17]. How then does the coreceptor enhance CTL activation? First, binding of CD8 and TCR to the same pMHC would increase the overall avidity between the surfaces of APCs and T cells. Second, CD8 binding to pMHC recruits the $p56^{kk}$ tyrosine kinase through its cytoplasmic domain into the T-cell signaling complex and thus enhances signal transduction. Third, CD8 binding to pMHC possibly reduces the overall flexibility of pMHC on the cell surface, positioning the pMHC more favourably for TCR binding. It has also been speculated that the stalk region of CD8 might directly interact with TCR (Refs 2, 41, 42), although we are not aware of any experimental evidence in support of this.

We[11,16] favor the model in which the soluble CD8–pMHC and the TCR–pMHC ectodomain protein–protein interactions are independent, enhancement of CTL activation occurring as a cell-surface effect involving the whole membrane-anchored molecules as observed in the experiments involving living cells with murine CD8 (Refs 9, 43). Based on SPR binding data, it was proposed[11] that the low affinity and fast kinetics of the CD8–pMHC interaction makes it unlikely that this initiates T cell–APC recognition. It is more likely that CD8 is recruited to the pMHC–TCR complex by intracellular binding of the α chain-associated $p56^{kk}$ to TCR associated ZAP-70 (Ref. 44). Once recruited, CD8 could enhance pMHC ligation by adding to the much stronger TCR–pMHC interaction. This mode of coreceptor binding to pMHC as a consequence of stable TCR–MHC binding is widely believed in coreceptor CD4 (Refs 8, 45, 46). However, some observations showed that tailless coreceptors, for both CD4 and CD8, are able to augment the T-cell responses, though to a lesser extent than the full-length molecules[47–51]. This was speculated because of the direct interaction of the ectodomains of CD8/CD4 with TCR, for example, the domain 3 of CD4 (Refs 52, 53).

CD8αα versus CD8αβ

CD8 is expressed on the cell surface in two forms: a CD8αα homodimer and a CD8αβ heterodimer, the latter being the prevalent form on the surfaces of the αβ T-cell population. The CD8αα homodimer is also present on αβ T cells, and is exclusively expressed on subsets of γδ T cells and NK cells[41]. It is a widely held belief that the CD8αβ heterodimer enhances CTL activation better than the αα homodimer[41,54], and it has been demonstrated that the β chain can broaden the range of antigen recognition of T cells[41,55–57]. CD8β knockout mice showed dramatically reduced numbers of mature peripheral $CD8^+$ cells, also indicating the importance of CD8β in thymic differentiation and maturation[58–60]. However, experiments aimed at unequivocally designating a functional role to the CD8β chain are complicated by the fact that expression only occurs in the αβ form and that the heterodimer is coexpressed with the αα homodimer. A recent study has shown that human CD8β chain (but not the murine counterpart) can be expressed as a ββ dimer on the cell surface of transfected COS-7 cells and murine lymphocytes[61], but the physiological relevance of this is unclear. There is no $p56^{kk}$ recognition sequence in the β chain cytoplasmic domain, so it is not immediately obvious why the CD8αβ form would be better suited for mediating signal transduction than the αα form[47,62]. Nevertheless, β chain enhancement of $p56^{kk}$ signaling has been reported in a murine system and the lack of the β chain cytoplasmic domain expression resulted in a several-fold reduction of CD8α-associated $p56^{kk}$ kinase activity[63]. SPR binding data involving soluble forms of the receptors show a similar pMHC binding affinity of the CD8αα homodimer and αβ heterodimer[10,12], and so it remains a puzzling question why the CD8αβ form appears to enhance CTL activation better than the αα form.

Recently, Bosselut and colleagues[64] have reconstituted variant murine β chain chimeric constructs into CD8β-deficient mice, and demonstrated that the extracellular domain of the β chain increases the avidity of CD8 binding to pMHC. Furthermore, the cytoplasmic domain enhanced association of $p56^{kk}$ and LAT (linker for activation of T cells), and hence augmented TCR signal transduction. These observations, and an earlier work using murine CD8 (Ref. 65), challenge the previous model for the role of the β chain according to which its main function would be to promote the intracellular association of $p56^{kk}$ with the α chain or TCR (reviewed in Ref. 41). Moreover, this is in contrast to the SPR BIAcore® binding data[10,12], in which the binding of CD8αα and CD8αβ to MHC class I does not show significant difference and the importance of β chain in avidity increment is not so obvious. This does raise some questions in general for the discrepancy of SPR and cell-based binding experiments as SPR is measuring the binding of one of the partners in three dimensions (soluble protein in solution), whereas the cell-based experiment is measuring protein–protein interaction in two dimensions. There are some constraints for movement of membrane-anchored proteins on the cell surface. Therefore, further work is needed to harmonize these observations.

If the ectodomain of the β chain does increase the αβ binding avidity, what could be the structural basis for a tighter association in the CD8αβ–pMHC complex compared with the CD8αα–pMHC? Two different models have been proposed for the position of the β chain in the complex. Based on electrostatic statistics[16] and experimental mutagenesis work[66], the β chain would probably substitute the T-cell distal subunit (colored blue in Fig. 2) of the CD8αα dimer that contributes only approximately 25% of the binding contacts in the asymmetric CD8αα–pMHC interaction. An alternative model is based on the amino-acid length of the stalk regions of the α and β chains[17], the α chain possessing the longer stalk region (more amino acids) potentially reaching further towards the APC surface. According to this theoretical model (there is no experimental data as yet) it is therefore more likely that the β chain would occupy the T-cell proximal position in the complex, substituting the α subunit (colored red in Fig. 2), which contributes about 75% of the binding in the CD8αα–pMHC complex. These binding models await verification by further experimental data.

Conclusions and future perspectives

The molecular architecture and binding parameters characterizing the interaction between CD8αα and pMHC have been unraveled through recent structural and biophysical studies. These data provide a basis for mechanistic models of the coordination between coreceptor and TCR in the recognition of pMHC. The TCR dominates the T-cell interaction with pMHC, ensuring antigen specificity in the interaction. In accordance with their relative functions, the nature of the TCR–pMHC and CD8–pMHC interactions are quite different, both in architecture, as seen in the complex structures of TCR–pMHC and CD8αα–pMHC, and in binding characteristics as demonstrated by the 'induced-fit' mode of the TCR–pMHC interaction. Even so, important data are still missing in order to generate a complete picture of the protein–protein interactions involved in initiating T-cell activation. The resolution of a crystal structure of CD8αβ or a CD8αβ–pMHC complex would provide new clues to the basis for the importance of the β chain in CTL activation. A triple complex structure of TCR–pMHC–CD8 would help to finally settle the issue of whether conformational changes could be involved in cooperative binding of CD8 and TCR although the structural work reported so far has provided no basis for supporting this. Most importantly perhaps, a full understanding of T-cell activation awaits data that clarify the roles and interactions of the membrane-proximal, transmembrane and cytoplasmic domains of CD8 and TCR during the initial T-cell activation events.

We thank Professor John Bell for his support.

George F. Gao (ggao@crystal.harvard.edu) is a Wellcome Trust (UK) International Travel Fellow at the Dept. of Molecular and Cellular Biology, 7 Divinity Avenue, Cambridge, Harvard University, MA 02138, USA and the Laboratory of Molecular Medicine, Children's Hospital, 320 Longwood Avenue, Boston, Massachusetts, MA 02115, USA; Bent K. Jakobsen is at the Nuffield Dept. of Clinical Medicine, Institute of Molecular Medicine, Oxford University, Headington, Oxford, UK OX3 9DS.

References

1 Janeway, C.A., Jr. (1992) The T cell receptor as a multicomponent signalling machine: CD4/CD8 coreceptors and CD45 in T cell activation. *Annu. Rev. Immunol.* 10, 645–674
2 Davis, M.M. *et al.* (1998) Ligand recognition by αβ T cell receptors. *Annu. Rev. Immunol.* 16, 523–544
3 Garcia, K.C. *et al.* (1999) Structural basis of T cell recognition. *Annu. Rev. Immunol.* 17, 369–397
4 Paul, W.E. and Seder, R.A. (1994) Lymphocyte responses and cytokines. *Cell* 76, 241–251
5 Grakoui, A. *et al.* (1999) The immunological synapse: a molecular machine controlling T cell activation. *Science* 285, 221–227
6 Geijtenbeek, T.B. *et al.* (2000) Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell* 100, 575–585
7 Geijtenbeek, T.B. *et al.* (2000) DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* 100, 587–597
8 Hampl, J. *et al.* (1997) CD4 augments the response of a T cell to agonist but not to antagonist ligands. *Immunity*, 379–385
9 Luescher, I.F. *et al.* (1995) CD8 modulation of T-cell antigen receptor-ligand interactions on living cytotoxic T lymphocytes. *Nature* 373, 353–356
10 Garcia, K.C. *et al.* (1996) CD8 enhances formation of stable T-cell receptor/MHC class I molecule complexes. *Nature* 384, 577–581 [published erratum appears in *Nature* (1997) 387, 634]
11 Wyer, J.R. *et al.* (1999) T cell receptor and coreceptor CD8 αα bind peptide-MHC independently and with distinct kinetics. *Immunity* 10, 219–225
12 Kern, P. *et al.* (1999) Expression, purification, and functional analysis of murine ectodomain fragments of CD8αα and CD8αβ dimers. *J. Biol. Chem.* 274, 27237–27243
13 Wu, H. *et al.* (1997) Dimeric association and segmental variability in the structure of human CD4. *Nature* 387, 527–530
14 Sakihama, T. *et al.* (1995) Oligomerization of CD4 is required for stable binding to class II major histocompatibility complex proteins but not for interaction with human immunodeficiency virus gp120. *Proc. Natl. Acad. Sci. U. S. A.* 92, 6444–6448
15 Norment, A.M. *et al.* (1988) Cell–cell adhesion mediated by CD8 and MHC class I molecules. *Nature* 336, 79–81
16 Gao, G.F. *et al.* (1997) Crystal structure of the complex between human CD8αα and HLA-A2. *Nature* 387, 630–634
17 Kern, P.S. *et al.* (1998) Structural basis of CD8 coreceptor function revealed by crystallographic analysis of a murine CD8αα ectodomain fragment in complex with H-2K^b. *Immunity* 9, 519–530
18 Giblin, P.A. *et al.* (1994) The role of charge and multiple faces of the CD8αα homodimer in binding to major histocompatibility complex class I molecules: support for a bivalent model. *Proc. Natl. Acad. Sci. U. S. A.* 91, 1716–1720
19 Garboczi, D.N. *et al.* (1996) Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* 384, 134–141
20 Garcia, K.C. *et al.* (1998) Structural basis of plasticity in T cell receptor recognition of a self peptide–MHC antigen. *Science* 279, 1166–1172
21 Sun, J. *et al.* (1995) Interaction between CD8 and major histocompatibility complex (MHC) class I mediated by multiple contact surfaces that include the α2 and α3 domains of MHC class I. *J. Exp. Med.* 182, 1275–1280
22 Aldrich, C.J. *et al.* (1991) Negative and positive selection of antigen-specific cytotoxic T lymphocytes affected by the α3 domain of MHC I molecules. *Nature* 352, 718–721
23 Ingold, A.L. *et al.* (1991) Co-engagement of CD8 with the T cell receptor is required for negative selection. *Nature* 352, 721–723
24 Killeen, N. *et al.* (1992) Requirement for CD8–major histocompatibility complex class I interaction in positive and negative selection of developing T cells. *J. Exp. Med.* 176, 89–97
25 Salter, R.D. *et al.* (1989) Polymorphism in the α3 domain of HLA-A molecules affects binding to CD8. *Nature* 338, 345–347
26 Salter, R.D. *et al.* (1990) A binding site for the T-cell co-receptor CD8 on the α3 domain of HLA-A2. *Nature* 345, 41–46
27 Gao, G.F. *et al.* (1998) Assembly and crystallization of the complex between the human T cell coreceptor CD8α homodimer and HLA-A2. *Protein Sci.* 7, 1245–1249
28 Davis, S.J. *et al.* (1998) CD2 and the nature of protein interactions mediating cell–cell recognition. *Immunol. Rev.* 163, 217–236
29 Maenaka, K. *et al.* (1999) Killer cell immunoglobulin receptors and T cell receptors bind peptide–major histocompatibility complex class I with distinct thermodynamic and kinetic properties. *J. Biol. Chem.* 274, 28329–28334
30 Willcox, B.E. *et al.* (1999) TCR binding to peptide–MHC stabilizes a flexible recognition interface. *Immunity* 10, 357–365
31 Boniface, J.J. *et al.* (1999) Thermodynamics of T cell receptor binding to peptide–MHC: evidence for a general mechanism of molecular scanning. *Proc. Natl. Acad. Sci. U. S. A.* 96, 11446–11451

32 Krensky, A.M. *et al.* (1993) Peptides corresponding to the CD8 binding region of HLA class I block the differentiation of cytotoxic T lymphocyte precursors. *Transplant. Proc.* 25, 483–484
33 Clayberger, C. *et al.* (1994) Peptides corresponding to the CD8 and CD4 binding domains of HLA molecules block T lymphocyte immune responses *in vitro*. *J. Immunol.* 153, 946–951
34 Choksi, S. *et al.* (1998) A structure-based approach to designing synthetic CD8α peptides that can inhibit cytotoxic T-lymphocyte responses. *Nat. Med.* 4, 309–314
35 Sewell, A.K. *et al.* (1999) Antagonism of cytotoxic T-lymphocyte activation by soluble CD8. *Nat. Med.* 5, 399–404
36 Purbhoo, M.A. *et al.* (1998) Copresentation of natural HIV-1 agonist and antagonist ligands fails to induce the T cell receptor signaling cascade. *Proc. Natl. Acad. Sci. U. S. A.* 95, 4527–4532
37 Martinez-Naves, E. *et al.* (1997) Interactions of HLA-B*4801 with peptide and CD8. *Tissue Antigens* 50, 258–264
38 Sanders, S.K. *et al.* (1991) Cell-cell adhesion mediated by CD8 and human histocompatibility leukocyte antigen G, a nonclassical major histocompatibility complex class 1 molecule on cytotrophoblasts. *J. Exp. Med* 174, 737–740
39 Gao, G.F. *et al.* (2000) Classical and nonclassical class I major histocompatibility complex molecules exhibit subtle conformational differences that affect binding to CD8αα. *J. Biol. Chem.* 275, 15232–15238
40 Zeng, Z. *et al.* (1997) Crystal structure of mouse CD1: An MHC-like fold with a large hydrophobic binding groove. *Science* 277, 339–345
41 Zamoyska, R. (1994) The CD8 coreceptor revisited: one chain good, two chains better. *Immunity* 1, 243–246
42 Kwan Lim, G.E. *et al.* (1998) Co-capping studies reveal CD8/TCR interactions after capping CD8β polypeptides and intracellular associations of CD8 with p56(lck). *Eur. J. Immunol.* 28, 745–754
43 Daniels, M.A. and Jameson, S.C. (2000) Critical role for CD8 in T cell receptor binding and activation by peptide/major histocompatibility complex multimers. *J. Exp. Med.* 191, 335–346
44 Thome, M. *et al.* (1995) Syk and ZAP-70 mediate recruitment of $p56^{lck}$/CD4 to the activated T cell receptor/CD3/zeta complex. *J. Exp. Med.* 181, 1997–2006
45 Xu, H. and Littman, D.R. (1993) A kinase-independent function of Lck in potentiating antigen-specific T cell activation. *Cell* 74, 633–643
46 Madrenas, J. *et al.* (1997) The efficiency of CD4 recruitment to ligand-engaged TCR controls the agonist/partial agonist properties of peptide-MHC molecule ligands. *J. Exp. Med.* 185, 219–229
47 Letourneur, F. *et al.* (1990) A signaling role for the cytoplasmic segment of the CD8α chain detected under limiting stimulatory conditions. *Proc. Natl. Acad. Sci. U. S. A.* 87, 2339–2343
48 O'Rourke, A.M. *et al.* (1990) Activated CD8 binding to class I protein mediated by the T-cell receptor results in signalling. *Nature* 346, 187–189
49 Fung-Leung, W.P. *et al.* (1993) The lack of CD8α cytoplasmic domain resulted in a dramatic decrease in efficiency in thymic maturation but only a moderate reduction in cytotoxic function of $CD8^+$ T lymphocytes. *Eur. J. Immunol.* 23, 2834–2840
50 Chan, I.T. *et al.* (1993) Thymic selection of cytotoxic T cells independent of CD8α–Lck association. *Science* 261, 1581–1584
51 Itano, A. *et al.* (1996) The cytoplasmic domain of CD4 promotes the development of CD4 lineage T cells. *J. Exp. Med.* 183, 731–741
52 Vignali, D.A. *et al.* (1996) The two membrane proximal domains of CD4 interact with the T cell receptor. *J. Exp. Med.* 183, 2097–2107
53 Vignali, D.A. and Vignali, K.M. (1999) Profound enhancement of T cell activation mediated by the interaction between the TCR and the D3 domain of CD4. *J. Immunol.* 162, 1431–1439
54 Zamoyska, R. (1998) CD4 and CD8: modulators of T-cell receptor recognition of antigen and of immune responses? *Curr. Opin. Immunol.* 10, 82–87
55 Wheeler, C.J. *et al.* (1992) An immunological role for the CD8β-chain. *Nature* 357, 247–249
56 Karaki, S. *et al.* (1992) β-chain broadens range of CD8 recognition for MHC class I molecule. *J. Immunol.* 149, 1613–1618
57 Renard, V. *et al.* (1996) CD8β increases CD8 coreceptor function and participation in TCR-ligand binding. *J. Exp. Med.* 184, 2439–2444
58 Crooks, M.E. and Littman, D.R. (1994) Disruption of T lymphocyte positive and negative selection in mice lacking the CD8β chain. *Immunity* 1, 277–285
59 Fung-Leung, W.P. *et al.* (1994) Reduced thymic maturation but normal effector function of $CD8^+$ T cells in CD8β gene-targeted mice. *J. Exp. Med.* 180, 959–967
60 Nakayama, K. *et al.* (1994) Requirement for CD8β chain in positive selection of CD8-lineage T cells. *Science* 263, 1131–1133
61 Devine, L. *et al.* (2000) Human CD8β, but not mouse CD8β, can be expressed in the absence of CD8α as a ββ homodimer. *J. Immunol.* 164, 833–838
62 Veillette, A. *et al.* (1988) The CD4 and CD8 T cell surface antigens are associated with the internal membrane tyrosine–protein kinase $p56^{lck}$. *Cell* 55, 301–308
63 Irie, H.Y. *et al.* (1998) The cytoplasmic domain of CD8β regulates Lck kinase activation and CD8 T cell development. *J. Immunol.* 161, 183–191
64 Bosselut, R. *et al.* (2000) Role of CD8β domains in CD8 coreceptor function: importance for MHC I binding, signaling, and positive selection of $CD8^+$ T cells in the thymus. *Immunity* 12, 409–418
65 Witte, T. *et al.* (1999) The CD8β ectodomain contributes to the augmented coreceptor function of CD8αβ heterodimers relative to CD8αα homodimers. *Cell. Immunol.* 191, 90–96
66 Devine, L. *et al.* (1999) Orientation of the Ig domains of CD8αβ relative to MHC class I. *J. Immunol.* 162, 846–851

COMMENTARIES

Did you know that you can now access timely summaries of outstanding recently published papers on a daily basis through the Commentaries section of our recently enhanced BioMedNet online service?

Written by distinguished scientists, Commentaries provide a thought-provoking synopsis of scientific endeavour across the life sciences and biotechnology.

Access to the Commentaries is FREE: visit http://journals.bmn.com and click on the Commentary button.

Annu. Rev. Immunol. 2006.24:419-466. Downloaded from www.annualreviews.org
by Northwestern University - Evanston Campus on 07/27/11. For personal use only.

How TCRs Bind MHCs, Peptides, and Coreceptors

Markus G. Rudolph,[1] Robyn L. Stanfield,[2] and Ian A. Wilson[2,3]

[1]Department of Molecular Structural Biology, University of Göttingen, 37077 Göttingen, Germany; email: mrudolp2@gwdg.de

[2]Department of Molecular Biology, The Scripps Research Institute, and [3]The Skaggs Institute for Chemical Biology, La Jolla, California 92037; email: robyn@scripps.edu, wilson@scripps.edu

Annu. Rev. Immunol.
2006. 24:419–466

First published online as a Review in Advance on January 16, 2006

The *Annual Review of Immunology* is online at immunol.annualreviews.org

This article's doi:
10.1146/annurev.immunol.23.021704.115658

Copyright © 2006 by Annual Reviews. All rights reserved

0732-0582/06/0423-0419$20.00

Key Words

T cell receptor, major histocompatibility complex, protein-protein interaction, crystal structure, immunological synapse

Abstract

Since the first crystal structure determinations of αβ T cell receptors (TCRs) bound to class I MHC-peptide (pMHC) antigens in 1996, a sizable database of 24 class I and class II TCR/pMHC complexes has been accumulated that now defines a substantial degree of structural variability in TCR/pMHC recognition. Recent determination of free and bound γδ TCR structures has enabled comparisons of the modes of antigen recognition by αβ and γδ T cells and antibodies. Crystal structures of TCR accessory (CD4, CD8) and coreceptor molecules (CD3εδ, CD3εγ) have further advanced our structural understanding of most of the components that constitute the TCR signaling complex. Despite all these efforts, the structural basis for MHC restriction and signaling remains elusive as no structural features that define a common binding mode or signaling mechanism have yet been gleaned from the current set of TCR/pMHC complexes. Notwithstanding, the impressive array of self, foreign (microbial), and autoimmune TCR complexes have uncovered the diverse ways in which antigens can be specifically recognized by TCRs.

161

Annu. Rev. Immunol. 2006.24:419-466. Downloaded from www.annualreviews.org
by Northwestern University - Evanston Campus on 07/27/11. For personal use only.

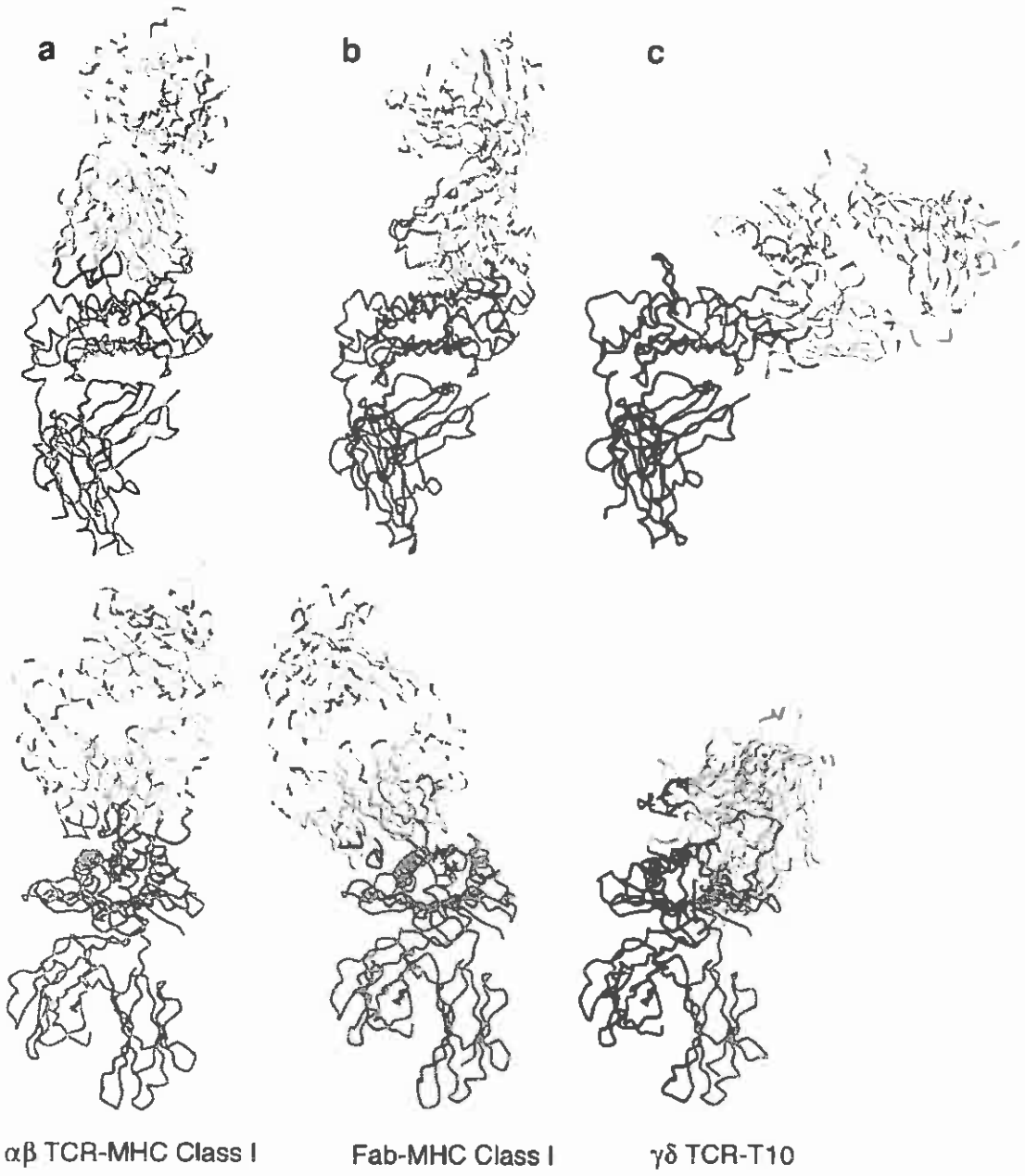


Figure 4

Overall comparison of the anatomy of complexes formed between MHC(-like) proteins and αβ TCR (*a*), Fab (*b*), and γδ TCR (*c*). The bottom panel is rotated 90° around the horizontal axis. Only one representative structure is shown for each type. The Cα trace of the TCR or Fab is on top in light gray with colored CDR loops and the MHC in dark gray below. The peptides in the αβ and Fab complexes are drawn as red ball-and-stick representations, while the CDR loops are colored as follows: CDR1 $\alpha_{(24-31)}$: dark blue, CDR2 $\alpha_{(48-55)}$: magenta, CDR3 $\alpha_{(93-104)}$: green, CDR1 $\beta_{(26-31)}$: cyan, CDR2 $\beta_{(48-55)}$: pink, CDR3 $\beta_{(95-107)}$: yellow, and HV4 $_{(69-74)}$: orange. This color scheme is continued through Figures 5 and 7.

Annu. Rev. Immunol. 2006.24:419-466. Downloaded from www.annualreviews.org
by Northwestern University - Evanston Campus on 07/27/11. For personal use only.

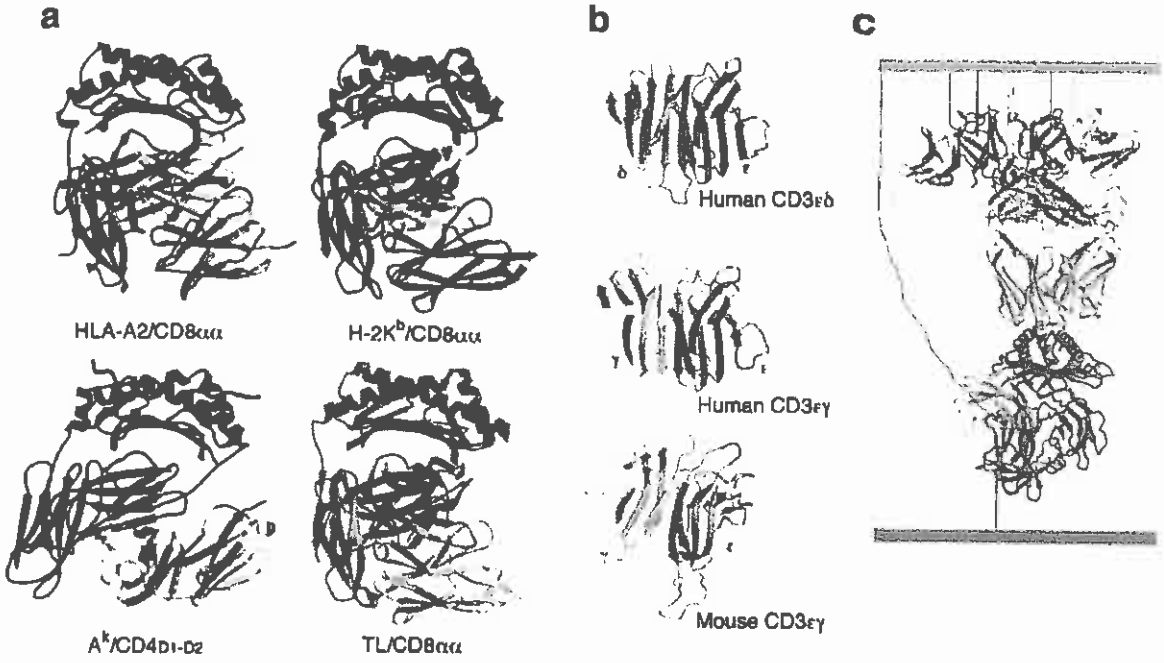


Figure 8

CD8 and CD4 coreceptor binding to class I and class II MHC. (*a*) The MHC is colored blue, CD8αα in yellow and orange. It is not yet known which of the domains of CD8αα homodimer correspond to the CD8αβ heterodimer (*upper* and *lower right panel*). The MHC is colored blue, and the CD4 (two N-terminal domains) is colored in yellow (*lower left panel*). (*b*) Human CD3εδ (*top*), human CD3εγ (*middle*), and mouse CD3εγ (*bottom*). In all three panels, the common ε-chain is colored in blue. (*c*) Hypothetical TCR/pMHC/CD3εδ/CD3εγ/CD8 complex. The TCR/pMHC-CD8αα and putative CD8αβ interaction is modeled by superimposing two structures, the HLA-A2/CD8αα complex (1akj) and the TCR A6/HLA-A2/TaxP6A complex (1qrn) on their MHC residues α1–180, with TCR (*green*), MHC (*dark blue*), peptide (*red*) and CD8 (*yellow* and *orange*). The CD3εδ (1xiw, *pink* and *blue*) and CD3εγ (1sy6, *gold* and *blue*) are shown "docked" at the top of the figure, with the common ε-chains colored in blue. This "docking" merely represents placing of the CD3 structures in the vicinity of where they are thought to bind, roughly following the cartoon diagram in Reference 125. Lines are drawn in to depict tethers connecting the different subunits to the TCR cell membrane (*top*, *green*) or the antigen-presenting cell membrane (*brown*, *bottom*).

as CD8αβ is the true αβ TCR coreceptor, whereas the main function of CD8αα on intraepithelial lymphocytes is to aid in adaptation and survival in the gut (108). Clearly, a CD8αβ structure in complex with pMHC is needed to derive the structural basis of why CD8αα cannot functionally replace or complement CD8αβ.

The low-resolution (4.3 Å) crystal structure of CD4 (domains 1 and 2) bound to I-A^k shows how the analogous coreceptor interaction is achieved in the MHC class II system (109). Whereas both domains of CD8 cooperate to bind class I MHCs, only one domain (the N-terminal variable-like region) of CD4 makes contact with the MHC with the second tandem CD4 domain being distal to the interface. Comparison of the CD4/pMHC and CD8/pMHC structures exposes a surprising structural dichotomy of the MHC class I and class II architectures, implying profoundly different modes of organization in their respective immunological synapses (**Figure 8**). As the complete, four-domain crystal

Interaction of MHC class II with CD4

Similarly, to stabilize the interaction between MHC class II-processed peptide complex with TCR requires another membrane bound molecules known as CD4. Its amino acid and nucleotide sequences are also available in the GanBank database. Its amino acid sequence consists of a signal peptide from position 1 to 25, an extra-cellular region from position 26 to 396, a trans-membrane region from position 397 to 418, and a cytoplasmic region from position 419 to 458. The extra-cellular region contains four domains with immunoglobulin folds:

Domain	Positions
D1	26 – 122
D2	123 – 204
D3	209 – 314
D4	318 – 374

Various studies have pointed to the interaction of the CD4 D1 and D2 domains with A2 and B2 regions of MHC class II molecules (Huang B, Yachou A, Fleury S, Hendrickson WA, Sekaly R-P (1997) Analysis of the contact sites on the CD4 molecule with class II MHC molecules. J Immunol, 158, 216-225). These authors concluded that a single CD4 interacted with two MHC class II molecules.

Subsequently, crystal structure were determined for a complex containing a human CD4 D1-D2 fragment and mouse MHC class II molecule (Wang I-h, Meijers R, Xiong Y, Liu J-h, Sakihama T, Zhang R, Joachimiak A, Reinherz EL (2001) Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. Proc Natl Acad Sci USA. 98, 10799-10804). The major interaction involves the CD4 D1 domian.

Therefore, similar to MHC class I molecule, a large complex is formed by MHC class II molecule, processed peptide, TCR and CD4. This complex can also be "polymerized" to make sure that the antigen presenting cell is firmly anchored to a $CD4^+$ T lymphocyte.

Analysis of the Contact Sites on the CD4 Molecule with Class II MHC Molecule

Co-Ligand Versus Co-Receptor Function[1]

Bei Huang,*† Abdelkader Yachou,*§ Sylvain Fleury,[2]*‡ Wayne A. Hendrickson,¶ and Rafick-Pierre Sekaly[3]*†‡

The CD4 molecule interacts with the α_2 and β_2 domains of the MHC class II molecules. However, the class II contact sites on CD4 are less clear. Involvement of different regions throughout D1, D2, and D3 domains have been suggested. To further delineate the class II MHC contact sites on CD4, a crystal structure-informed mutagenesis was performed. Alanine scan mutants were generated for exposed residues located throughout D1 and the FG loop of D2, and in the "hinge-like" region, a short and flexible region between D2 and D3. Mutants were tested in a co-ligand (D^d stimulation) and a co-receptor (staphylococcal enterotoxin B stimulation) assay. In the co-ligand assay, TCR and CD4 interact with two distinct ligands (D^d or HLA-DR), while in the co-receptor assay both molecules interact with the same ligand, namely HLA-DR. Results show that residues from both lateral faces of D1 and the FG loop of D2 are implicated in interaction with class II, although a bigger surface of CD4 was involved in co-receptor compared with co-ligand function. The potential involvement of residues on both the top and two side faces of CD4 is consistent with a bivalent model, which involves the interaction between a single CD4 and two class II molecules. Alternatively, our results can be interpreted with a model of a specifically organized CD4 and/or class II oligomerization event. Finally, results from the hinge region mutants revealed a very important role in maintaining the overall structural integrity of CD4, its topology, and function. *The Journal of Immunology,* **1997, 158: 216–225.**

The CD4 molecule is a cell surface glycoprotein with a molecular mass of 55 kDa. It is a member of the Ig gene superfamily. It is a monomer with four extracellular Ig-like domains, a transmembrane portion, and a cytoplasmic tail (1). The cytoplasmic tail of CD4 is noncovalently associated with the tyrosine kinase $p56^{lck}$ (2). CD4 is expressed on a subset of T lymphocytes and thymocytes. The TCRs of these $CD4^+$ cells recognize Ags presented by class II MHC molecules. The CD4 molecule binds to its ligand-class II MHC molecules to enhance the T cell response, either as an adhesion molecule (co-ligand) or in a ternary complex with the TCR as a part of the Ag recognition process, which is directly involved in the resulting signal transduction (co-receptor) (3). The CD4 class II interaction is critical for thymic selection during T cell development, and it is also important for the activation of mature T cells. The signaling functions of CD4 are mediated by a Src family tyrosine kinase $p56^{lck}$ that is associated noncovalently with the cytoplasmic tail of CD4 (2). Association of CD4 with a class II/TCR complex appears to bring this kinase into the TCR signaling pathway. CD4 has also been characterized as the primary cellular receptor for the HIV glycoprotein gp120 (4–6).

The CD4 binding site on class II molecules has been shown to be located in the α_2 and β_2 domain of class II (residue 134–140) (7–9). However, the class II contact sites on CD4 are still controversial. Results from our laboratory using a co-ligand assay (D^d) in which CD4 and TCR interact with two different MHC molecules have demonstrated that residues in the complementarity-determining region 1 (CDR1)[4] and CDR3 of D1, and the FG loop of D2, all located on the same face of CD4 that is opposite the CDR2, are implicated in binding to class II (10). Interestingly, the CDR2 deletion mutant (deletion of residues 43–49) can still interact with different murine and human class II alleles (11). Studies have shown that the CDR2 is quite divergent among species (12); therefore, it is less likely for the CDR2 to bind to a highly conserved nonpolymorphic region in class II. On the other hand, results from adhesion assays suggested that in addition to the CDR1, CDR3 of D1, and the FG loop of D2, an extended CDR2 is involved in binding to class II (13–15). Thus, the major question remains: should the extended CDR2 be implicated either in interaction with

*Laboratory of Immunology, Clinical Research Institute of Montreal, Montreal, Canada; †Department of Experimental Medicine, McGill University, Montreal, Canada; ‡Department of Microbiology and Immunology, University of Montreal, Montreal, Canada; §University of Hassan II-Mohammedia, Faculty of Sciences Ben M'sik, Casablanca, Morocco; and ¶Department of Biochemistry and Molecular Biophysics, Howard Hughes Medical Institute, Columbia University, New York, USA 10032

Received for publication July 3, 1996. Accepted for publication October 3, 1996.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work is supported by grants to R.-P.S. from the Medical Research Council (MRC) of Canada, National Cancer Institute, and the National Health Research and Development Program (NHRDP). B.H. and S.F. are fellows of NHRDP. The flow cytometry service receives a maintenance grant from the MRC and a donation from the Glaxo Foundation.

[2] Current address: Lymphocyte Biology Section, Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892-1892.

[3] Address correspondence and reprint requests to Dr. Rafick-Pierre Sekaly, Laboratoire d'Immunologie, Institut de Recherches Cliniques de Montreal, 110 avenue des Pins Ouest, Montreal, Quebec, Canada H2W 1R7.

[4] Abbreviations used in this paper: CDR, complementarity-determining region; D, domain; DEL, deletion; EC_{50}, 50% dose required to obtain 50% of the maximal response for a cell line; SEB, staphylococcal enterotoxin B; mt, mutant; wt, wild type.

Copyright © 1997 by The American Association of Immunologists

0022-1767/97/$02.00

class II or in another function of CD4 itself, such as CD4 dimerization/oligomerization or association with the TCR/CD3 complex.

Crystal structures obtained to date are the D1D2 fragment of the human CD4 (16, 17) and the D3D4 fragment of the rat CD4 molecule (18). The link between the two fragments has been suggested as a short and flexible "hinge-like" region (18, 19). This has been supported by crystallographic and proteolytic studies (20–23). When residue 180 or residues 180–181 were deleted, the CD4 molecule turned into a straightened rod-like structure instead of a bent structure, as observed under the electronic micrographs (Dr. Leo Brady, personal communication). Recently, the CD4 molecule containing the four extracellular domains has been also obtained, which suggests that a hinge-like region exists between D2 and D3 (Drs. Hao Wu and Wayne Hendrickson, unpublished observations). Furthermore, it has been proposed that the hinge region plays an important role in maintaining the overall conformation of CD4, in ligand binding-induced conformational modification or formation of dimer/oligomer, in normal signal transduction, and in membrane fusion upon HIV binding (19, 23–26). Experimental evidence for the functional importance of this hinge region has yet to be obtained.

To further delineate the class II binding site on CD4 and to understand the role of the CDR2 in the class II interaction, a large panel of CD4 mutants encompassing residues located throughout both lateral faces of D1 and the FG loop of D2 domain of CD4 was generated. Moreover, to study the role of the hinge region, alanine scan and deletion mutants of the junction region between the D2 and D3 domains were generated. All the mutated CD4 molecules were expressed in the 3DT52.5.8 murine T cell hybridoma that is $CD4^-CD8^-$, and has TCR (Vβ8) specific for the cognate alloantigen H-2D^d. This cell line is dependent on CD4 for an efficient alloresponse. Mutants were studied in the co-ligand (D^d) assay described previously (10), and in the co-receptor assay that involves the CD4/TCR interaction with class II HLA-DR molecules in the presence of the bacterial superantigen staphylococcal enterotoxin B (SEB). Results presented in this report reveal the functional importance of these different domains of CD4 in the two assays.

Materials and Methods

Plasmids

Full-length human mutant CD4 cDNAs were obtained from different sources (27–29) or generated in our laboratory by PCR overlap extension method (30). For D1 and D2 mutants, the mutated CD4 cDNAs were prepared as blunt-end *Bam*HI or *Hin*dIII-*Not*I fragments and subcloned into the pMNC-stuffer retroviral vector (10). The amphotrophic helper cell line DAMP (31) was transfected with the different pMNC-mutant CD4 constructs (pMNC-mCD4). G418 (neomycin)-resistant DAMP cells were selected and used as the producer cell lines of the recombinant retrovirus to transfect the 3DT52.5.8 cell line. The proviral pMNC-mCD4 DNA confers resistance to the selection agent G418 and allows stable long-term expression of the mutated CD4 molecules in infected cells. For mutants in the hinge region, the mutated CD4 cDNAs were generated by PCR overlap extension.

Mutagenesis

The hinge region mutants were generated by the PCR overlap extension method. Briefly, two complementary oligonucleotides bearing specific mutations were used to generate two fragments with overlapping ends. Subsequent PCR reaction was conducted on the mixture of these fragments in which the 3′ overlap of each strand serves as primer for the 3′ extension of the complementary strand. The amplified fragments were then used to replace their homologues in the CD4 wt. Mutagenesis was confirmed by DNA sequencing. The full-length cDNAs coding for the mutant CD4 molecules were prepared as fragments and subcloned at the *Xba*I site of the eukaryotic expression vector SRα (32). Then they were used for electroporation.

Transfection

D1 and D2 mutants were generated as described (10). Briefly, $CD4^+$ DAMP cells were obtained by the calcium phosphate co-precipitation technique (33, 34). The DAMP cells transfected with the various pMNC-mCD4 constructs were enriched by FACS sorting with OKT4 Ab. These DAMP $CD4^+$ cells were used to infect the 3DT52.5.8 hybridomas as described (10). G418-resistant 3DT52.5.8 cells were then analyzed for the expression of the mutated CD4 molecule by flow cytometry. Populations of cells expressing homologous levels of TCR and CD4 were generated by aseptic cell sorting.

For the hinge region mutants, 3DT52.5.8 cells were electroporated with the SRα neo-mutated CD4 constructs. For each electroporation, 5×10^6 3DT cells were resuspended in 500 μl of complete RPMI supplemented with 10% FCS. To this cell suspension, 10 μg of DNA was added, and electroporation was conducted at a voltage of 270 mV. Forty-eight hours later, cells were subjected to G418 selection. Forty thousand cells were resuspended in 2 ml of complete RPMI with 10% FCS. G418 at a concentration of 1.5 mg/ml was added. The medium was changed every 3 days. G418-resistant 3DT52.5.8 cells were analyzed for the expression of the mutated CD4 molecule by flow cytometry.

Cytofluorometric analysis of cells

Cells (5×10^5) were incubated for 60 min at 4°C with 100 μl of the appropriate mAb (10 μg/ml) diluted in complete medium. Following washes in PBS, cells were incubated with fluorescein-coupled goat-anti-mouse Ig (Becton Dickinson, San Jose, CA) for 30 min at 4°C. Cells were then washed, resuspended in PBS containing 2% FCS, and analyzed for fluorescence using Lysis II (Becton Dickinson) flow cytometry. Dead cells were excluded by propidium iodide (0.5 μg/ml) gating, and 10,000 cells were acquired for the analysis on a four-decade logarithmic scale. As a control, cells were stained only with the fluorescein-coupled goat-anti-mouse Ig.

Stimulation of effector cells and IL-2 assay

Co-ligand assay. This assay has been described previously (10, 11). Briefly, 7.5×10^4 3DT52.5.8 cells expressing the mutated CD4 molecules were cocultured with 7.5×10^4 DAP-3 cells expressing H-2D^d or H-2D^d DR4. After the coculture, supernatants were harvested for IL-2 production assay. The IL-2 ratio for a given CD4 mutant is calculated by dividing the value of the IL-2 production, upon stimulation with H-2D^d + DR4, by the IL-2 value obtained with H-2D^d alone. The responses of this co-ligand assay are displayed as such (see Fig. 2): on the x-axis, the positions of the mutated residues are given; while the IL-2 ratios are displayed on the y-axis. Furthermore, when we compare the IL-2 ratio of the mutants with that of the wt $CD4^+$ cells, we obtain the IL-2 ratio in the co-ligand assay of the mutants as compared with the wild type: IL-2 ratio of (D^d mt)/(D^d wt) (as shown in Table IV).

Co-receptor assay. 7.5×10^4 3DT52.5.8 cells expressing the mutated CD4 molecules were cocultured with 2×10^4 DAP-3 cells expressing human class II molecule DR1 in 200 μl complete medium for 20 h at 37°C in 96-well flat-bottom culture plates, in the presence of twofold serial dilutions of SEB ranging between 10^5 and 10^{-2} pg/ml (a kind gift from Drs. Kappler and Marrack, National Jewish Hospital, Denver, CO). Titration of SEB was conducted both with 3DT52.5.8 and 3DT52.5.8 transfected with the human CD4 wt. After an overnight coculture, supernatants were harvested for IL-2 measurement. The responses of the T cell hybridomas expressing different CD4 mutants are displayed as such: on the x-axis, the log concentrations of the bacterial superantigen SEB are given; on the y-axis, the IL-2 units produced by each mutant upon SEB stimulation are displayed (as shown in Fig. 3).

Anti-TCR stimulation. The TCR-specific Abs KJ12 (recognizes the TCR association of the α-chain specific for D^d alloantigen and the Vβ8.1) and F23.1 (recognizes all the Vβ8s) were used to stimulate the cells. A wide range of concentrations (3, 1, 0.3, 0.1, 0.03, 0.01, 0.003 μg/ml) of the purified mAb were coated at 4°C in 96-well flat-bottom plates overnight. To these wells, 7.5×10^4 3DT52.5.8 cells expressing different mutated CD4 molecules were added in 200 μl of complete RPMI medium and incubated for 20 h at 37°C. The supernatants were harvested for the IL-2 production assay later.

EC_{50} calculation

EC_{50} is defined as the concentration of a given stimulus (antigen, superantigen or anti-TCR Ab) required to obtain a half-maximal response for a cell line. For all the stimulation assays such as the SEB and the anti-TCR assays, the results were plotted as log concentrations of the stimulus on the x-axis against the IL-2 values on the y-axis. Curve-fitting was done with

Table I. *Summary of Abs and epitopes*

mAb	Epitope: Binding domain of CD4	Epitope: Amino acid[a]
OKT4	D3	
OKT4C	D1, D3, D4	20/27, 57, 89, (277–280), (328–330), (349–356), (363–369)
OKT4D	D1, D4	20, 47, 55, 57, 89, (349–356)
OKT4E	D4	(363–369)
OKT4F	D1, D4	48, 55, 57, (349–356)
B66.6.1	D1, D2	20/27, 23/24/27, 38, 40, 45, 47, 46/47, 42–49, 57, (99–105)
L34	D1	26–83
L110	D1	26–83
L71	D1	83–92
L83	D1, D2	26–83, 120–202
L116	D1, D2	1–25, 120–202
MT425	D1	
MT407	D1, D2, D3	19, 20, 46/47, 47, 55, 57, 60, 165, (39–43), (48–52), (99–105), (121–123), (277–280)
Q425	D3	D strand of D3, 230–235
MT151	D1, D2, D4	38, 94, 165, (328–330), (349–356), (363–369)
IF3	D1	gp120 binding site
Leu3a	D1	Del 42–49, 57

[a] The epitopes identified from previous mutagenesis studies are displayed as the position of amino acid residues that when mutated will affect the binding of the Ab. Numbers represent the positions of the residues on the human CD4 molecule, and the numbers in parentheses represent residues that affect the epitope when substituted with the murine counterparts. The number following the slash (e.g. 20/27) stands for the double or triple mutations of residues at the given position.

Table II. *Serologic analysis of D1 and D2 mutants*

Cells[a]	Ab Binding[b]: OKT4	L34	L110	L71	L83	L116	MT425
CD4 wt	+	+	+	+	+	+	+
D1 mutants							
K1A	+	+	+	+	+	+	+
K2A	+	+	+	+	+	+	+
K7A	+	+	+	+	+/−	+	+
T15A	+	+	+	+	+	+	+
T17A	+	+/−	+	+	+	+/−	+
S19A	+	+	+	+	+	+	+
Q20A	+	+	+	+/−	+	+	+
K21A	+	+	+	+	+	+	+
S23A	+	+	+	+	+	+	+
K46A	+	+	+	+	+	+	+
E77A	+	+	+	+	+	+	+
K88A	+	+	+	+	+	+	+
Q89A	+	+	+	+	+	+	+
K90A	+	+	+	+	+	+	+
D2 mutants							
Q163A	+	+	+	+	+	+	+
N164A	+	+	+	+	+	+	+
Q165A	+	+	+	+	+	+	+
K166A	+	+	+	+	+	+	+
K167A	+	+	+	+	+	+	+

[a] Mutants are designed by the name of the residue (one-letter amino acid code) and its position on the sequence, followed by the name of the substituting residue.

[b] Murine T cell hybridoma cell line 3DT52.5.8 expressing human CD4 or mutant CD4, was subjected to indirect immunofluorescent staining with all of the Abs listed here at concentration saturating for the CD4 wt. Analyses were carried out with FACScan flow cytometer. Results were calculated as follows: ([mt FL (mAb) − mt FL (neg)] / [mt FL (OKT4) − mt FL (neg)]) / ([wt FL (mAb) − wt FL (neg)] / [wt FL (OKT4) − wt FL (neg)]). FL indicates mean fluorescence value; mt, a given CD4 mutant; wt, the wild-type CD4; and neg, negative control. Symbols for binding are: +, >50%; +/−, 30 to 50%; −, <30%.

Sigma Plot Program (Jandel Scientific, San Rafael, CA). The EC_{50} value was determined from the curve for each CD4 mutant by reading out the *x*-axis value such that the *y*-axis value equals half of the peak IL-2 value on the curve.

The EC_{50} values of all the CD4 mutants were compared with that of the cells expressing wt CD4, which gives the EC_{50} ratio (as shown in Table IV). In addition, for the SEB assay, an IL-2 ratio for each mutant was calculated as follows: at the EC_{50} of SEB determined from the wt $CD4^+$ cells, the IL-2 production of the mutants tested at the same conditions were calculated from the curve; these IL-2 values were compared with that of the wt $CD4^+$ cells, giving the IL-2 ratio of [(SEB mt)/(SEB wt)] (as displayed in Table IV).

Molecular modeling

The CD4 images were generated using Quanta software (Polygen Corp., Waltham, MA) on a Silicon Graphics work station (Mountain View, CA).

Results

Description and expression of CD4 mutants

D1 and D2 mutants. To further define the molecular interactions between MHC class II and CD4, a detailed mutational analysis of CD4 was undertaken using the coordinates of the available crystal structure. We examined three major issues concerning the CD4/class II interaction: 1) the potential contribution of other residues from the CDR1, CDR3 of D1, and the FG loop of D2; 2) the involvement of other regions on the same face but outside the CDR-like regions in the D1; 3) the role of the CDR2 in class II binding. For mutagenesis, we chose amino acids that are exposed on the surface and whose side chains are highly solvent accessible, reasoning that these residues are most likely to make contact with MHC class II molecules, and that changes at these positions were least likely to result in global structural alterations. We made 16 substitutions in D1 and 5 in D2. Single amino acid substitution to alanine were introduced into CD4 by site-directed mutagenesis. Mutants are described by a single-letter code for the amino acid involved and the position of the mutated residue.

Hinge region mutants. To investigate the possible role of the hinge region in T cell activation, mutants were generated, including alanine scan for V175, V176, L177, Q180, QK180–181, and deletion (DEL) mutants DEL 175–177 and DEL 180–181. Among these residues, V176 and L177 are conserved among species including human, mouse, rat, rabbit, and cat; and residue K181 is conserved among human, rabbit, and cat.

Serologic analysis of the CD4 mutants

A detailed serologic analysis was undertaken to determine the structural impact of the mutations on the conformation of the CD4 molecule. All the Abs and their epitopes are summarized in Table I. The results of the serologic studies for D1 and D2 mutants are summarized in Table II, while results for the hinge region mutants are summarized in Table III.

All 17 mutants from the D1 and D2 domains were stained with a panel of seven different mAbs that map to different regions in D1, or D1 and D2. mAb OKT4 recognizes a region on the membrane-proximal portion of CD4, and all the CD4 mutants express the OKT4 epitope. None of the D1 or D2 mutations resulted in the loss of epitopes tested. However, some mutants had limited effects such as reduced binding for one or two epitopes. Results illustrated in Table II show that mutants K7A and Q20A had reduced binding for only one mAb; mutant T17A had reduced binding for two mAbs. These results clearly showed that there are no overall conformational changes induced by the mutations; the other regions should still be intact and able to interact with other surface molecules such as class II MHC molecules and the TCR.

167

Table III. *Serologic analysis of the hinge region mutants*

Cell	Ab Binding[a]															
	OKT4	OKT4C	OKT4D	OKT4F	OKT4F	B66 6 1	L34	L110	L71	L83	L116	MT4 25	MT407	Q425	MT151	IF3	LEU3A
CD4 wt	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
V175A	+	+/−	+/−	+	+/−	+	+	+	+	+/−	+	+	+	−	+	+	+
V176A	+	+	+	+	+	+	+	+/−	+	+	+/−	+	+	+	+	+/−	+
L177A	+	+/−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DEL 175–177	+	+/−	+/−	+	+/−	+/−	+/−	+/−	+/−	+/−	−	+/−	−	−	+	−	+/−
Q180A	+	+	+	+	+/−	+	+	+	+	+	+	+	+	+	+	+	+
QK180–181A	+	+	+/−	+	+	+	+	+/−	+	+	+/−	+	+	+/−	+	+	+
DEL 180–181	+	+	+/−	+	+/−	+	+	+	+	+	+/−	+	−	+	+	+	+

[a] Fluorescence analysis. Calculations and symbols are the same as described in Table II.

For the hinge region mutants, a panel of 17 mAbs that map from the D1 through the D4 domain of CD4 were used. All of the CD4 mutants retained the epitope for OKT4. However, the epitopes for the other anti-CD4 mAbs were differentially affected by these mutations. The deletion mutant DEL 175–177 had the most drastic effects on the conformation of the CD4 molecule, as shown by the loss of 4 of 17 epitopes and decreased binding for another 10 Abs. Only 2 mAbs of 17 still bind to this mutant at levels comparable to that of CD4 wt. Besides, the V175A mutant and the deletion mutant DEL 180–181 have lost one epitope and have several affected epitopes. The other mutants have several affected epitopes, but lost no epitopes (Table III). From these results it appears that the epitopes affected by the various hinge region mutations occur throughout D1 to D4, suggesting that this region is very important for the conformation of CD4. Among all the mutants, the DEL 175–177 mutant had the most drastic effects, indicating that the conserved residues V176 and L177 are especially critical in maintaining the overall structure of CD4. Deletions may shorten the hinge region, resulting in the close association of D2 and D3 that leads to the formation of a rod structure. This mutation may also change the orientation of the membrane-distal domain of CD4.

Functional analysis of the CD4/class II MHC interactionss

All of the D1, D2, and hinge region mutants expressed in the 3DT52.5.8 hybridoma cells were examined for their ability to interact with the class II MHC molecule in two functional assay systems: the co-ligand and the co-receptor assay (the two assay systems are illustrated in Fig. 1). Introduction of CD4 in both assays results in an enhanced production of IL-2. A minimum of three hybridoma populations expressing each CD4 mutant and comparable levels of CD4 and the TCR were used in the two assays. The responses of each one of the mutants to the TCR-specific mAbs (KJ12 and F23.1) coated on plates were comparable. Finally, each mutant was tested in at least three independent experiments.

Co-ligand stimulation

D1 and D2 mutants. T cell populations that were selected for functional assays showed comparable EC_{50} values in responses to the TCR-specific Ab KJ12, as shown in Table IV (EC_{50} ratios of the KJ12 stimulation range from 0.8 to 1.4). Although all the D1 mutants produce comparable levels of IL-2 upon KJ12 stimulation, their responses in the co-ligand assay were variable. Results presented in Table IV show that mutations K1A, K2A, K7A, T15A, S19Y, Q20A, K46A, E77A, E87A, D88A, Q89L, and Q89A affect the binding to class II, while other mutations such as T17A, S19A, K21A, S23A, DEL 43–49, and K90A did not show any effect. Interestingly, S19A can still interact with class II, although an abrogation of the interaction by substitution of this residue to tyrosine was previously reported (10).

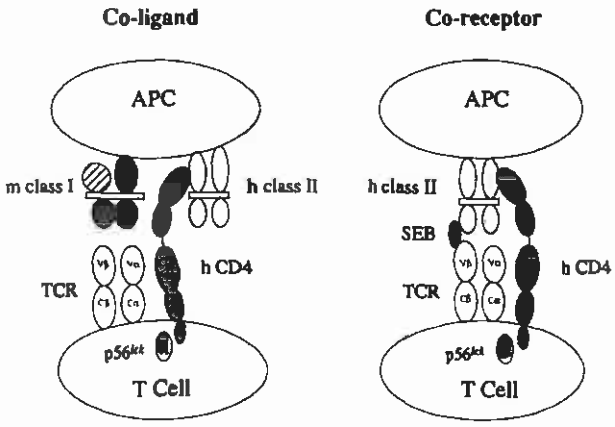


FIGURE 1. Schematic representation of the two functions of CD4 in the 3DT cell system, co-ligand vs co-receptor. In the co-ligand system, CD4 and TCR bind to different classes of MHC molecules. The TCR recognizes the class I cognate alloantigen H-2D^d, while CD4 interacts with the class II MHC molecules DR4. In the co-receptor system, the TCR and CD4 can both bind to the class II molecule DR1 in the presence of the bacterial superantigen SEB. The interaction of CD4 with the class II molecules can enhance the T cell response in both cases.

While the CDR2 deletion mutant DEL 43–49 can still interact with different human and mouse class II alleles (11), single mutations within the CDR2 (i.e., K46A) or outside the CDR2 but still on the same face as the CDR2 (i.e., E77A), can no longer interact efficiently with human class II. Though this is different from our prediction that the CDR2 is not involved in binding to class II, it is consistent with our previous study in which substitutions in this region, such as T45P, K46N/G47V, and G47R, did have an effect on the interaction with class II (10). It is possible that the deletion mutant adopts a different conformation that somehow restores its binding to class II, maybe through a different contact interface. Alternatively, it is possible, as recently suggested, that this domain of CD4 is not directly involved in CD4/class II interaction but rather involved in CD4 oligomerization (35).

All of the alanine scan mutants for residues from the FG loop of D2 were studied. Their responses in the co-ligand assay were variable. Mutants N164A and Q167A can still enhance the D^d response, showing a phenotype comparable to that of the CD4$^+$ cells. In contrast, mutants Q163A, Q165A, and K166A behave as the CD4$^-$ cells and the previously shown mutant 165[SR] (10).

Table IV. *Summary of the CD4 mutants*

Cells	Expression Level (mt/wt)[a]			Stimulation (EC_{50} Ratio: mt/wt)[b]			IL-2 Ratio-SEB mt/wt[c]	IL-2 Ratio of D^d mt/D^dwt[c]
	OLT4	KJ12	F23.1	KJ12	F23.1	SEB		
wt	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
$CD4^-$		0.8	0.9	0.6	1.1	4.6	0.1	0.2
D1								
K1A	0.6	1.9	1.1	0.9	1.4	3.0	0.1	0.3
K2A	0.9	0.6	1.0	1.2	1.3	3.0	0.1	0.2
K7A	0.5	0.7	0.7	1.1	1.3	6.7	0.1	0.2
T15A	1.0	0.9	0.9	ND	1.6	4.9	0.1	0.2
T17A	0.9	1.0	0.9	1.2	0.6	>10	0.1	1.2
S19A	1.0	1.3	0.8	1.3	1.5	5.1	0.2	1.5
S19Y	0.5	2.0	1.5	ND	2.0	6.0	0.1	0.3
Q20A	0.8	0.8	0.6	1.2	2.5	6.3	0.1	0.3
K21A	0.7	0.8	0.8	ND	ND	4.4	0.1	0.9
S23A	0.7	0.8	0.8	ND	ND	5.1	0.2	1.0
D43–49	1.1	1.4	1.6	1.1	1.2	5.0	0.1	1.1
K46A	1.1	0.8	0.8	1.3	1.1	5.0	0.1	0.1
E77A	1.5	2.6	2.5	0.8	0.9	5.0	0.1	0.1
E87A	0.5	0.8	0.6	1.4	1.4	6.0	0.1	0.2
D88A	1.0	1.0	1.0	1.1	1.4	8.5	0.1	0.2
Q89A	0.9	0.8	1.0	0.8	ND	3.0	0.1	0.2
Q89L	0.8	0.9	0.9	1.1	1	7.0	0.04	0.2
K90A	1.1	2.0	1.1	1.4	1.1	1.3	0.8	1.5
D2								
Q163A	0.7	0.7	0.7	1.2	1.0	4.3	0.1	0.1
N164A	0.8	1.3	1.0	1.4	1.2	1.2	0.8	1.4
Q165A	0.5	0.9	0.6	1.2	1.3	3.6	0.2	0.2
165[SR]	0.5	0.9	1.1	1.2	0.8	3.0	0.2	0.1
K166A	1.0	1.7	1.7	1.0	0.9	4.5	0.4	0.3
K167A	1.0	1.6	0.9	0.7	1.4	5.8	0.2	1.1
Hinge								
V175A	0.4	0.7	0.7	0.9	0.9	3.3	0.1	0.1
V176A	2.2	1.3	1.3	ND	1.3	4.3	0.1	0.1
L177A	1.2	0.7	0.7	1.6	1.3	6.7	0.1	0.1
VVL175–7A	0.7	0.9	1.0	ND	1.6	4.9	0.1	0.2
D175–7	0.5	0.8	0.9	1.2	0.6	>10	0.01	0.1
Q180A	0.9	1.5	1.6	ND	0.6	3.5	0.2	2.6
QK180–1A	0.8	1.6	1.7	ND	1.5	4.0	0.1	0.9
D180–1	0.7	1.9	1.7	ND	1.6	2.8	0.1	0.2

[a] Expression levels for CD4 (OKT4) and TCR (KJ12 and F23.1) are defined as the mean fluorescence value of a given mutant divided by the mean fluorescence value of the wt^+ cells.

[b] Stimulation by anti-TCR mAbs (KJ12 and F23.1) and SEB was defined as the EC_{50} ratios. The calculation of EC_{50} value and EC_{50} ratios is described in *Materials and Methods*.

[c] The IL-2 ratio for SEB mt/wt and IL-2 ratio for D^d mt/wt are also defined in *Materials and Methods*.

They do not produce increased levels of IL-2 in the presence of class II^+ cells (Fig. 2, Table IV).

Hinge region mutants. The hinge region mutants were also expressed in the 3DT52.5.8 cells and studied in the co-ligand and KJ12 stimulation assays. Although all the CD4 mutants gave similar responses after anti-TCR stimulation when compared with the wt $CD4^+$ cells, only mutants Q180A and QK180–181A showed a comparable response to D^d as reflected by IL-2 production ratios. The other mutants, including those that did not lose any epitopes, responded to D^d at a magnitude similar to the $CD4^-$ cells (Fig. 2). The latter result clearly suggests that this region of CD4 could play an important role in the structural integrity of CD4 and in its interaction with class II.

Co-receptor stimulation

In the present study, we confirm that the presence of the CD4 molecule in the 3DT52.5.8 T cell hybridomas can enhance the T cell response to the sub-optimal concentrations of SEB (11). Cells expressing CD4 require about 10-fold less SEB (700 pg/ml) for a half-maximal stimulation when compared with the $CD4^-$ cells (7000 pg/ml). Moreover, the IL-2 production of the $CD4^-$ cells (4000 U/ml) represents only 10% of that of the wt $CD4^+$ cells (48,000 U/ml) (Table IV and data not shown). This enhancement has been demonstrated in 11B3 (human wt $CD4^+$) (11), and two other cell lines, B57 and E10, independently transfected with the human wt CD4 (Fig. 3*A*).

D1 mutants. Mutants S19Y and Q89L that were already shown in the co-ligand assay as no longer being able to enhance the 3DT response to D^d (10) were also studied in the SEB assay. These mutants gave a significantly reduced response to SEB, since they consistently required 3 to 7 times more SEB for a half-maximal response ($n = 3$); moreover, at the EC_{50} of the $CD4^+$ cells (700 pg/ml), IL-2 production by the CD4 mutants reach about 4 to 20% of the wt $CD4^+$ cells (Fig. 3, *C* and *F*; Table IV). Thus, these CD4 mutants display a phenotype similar to that of the $CD4^-$ cells. These results indicate that at least for these residues, mutations that affect the CD4/class II interaction in a co-ligand assay also affect it in the co-receptor assay, further suggesting that the same contact residues were shared in the two functional assays. Surprisingly, when we tested the CDR2 deletion mutant (lacking residues 43–49 of D1), we found that it could no longer enhance the T cell response to SEB. Indeed, cells expressing this mutant needed at least 7 times more SEB for a half-maximal response. Moreover, at the EC_{50} (700 pg/ml), their IL-2 production was only 7% of the wt

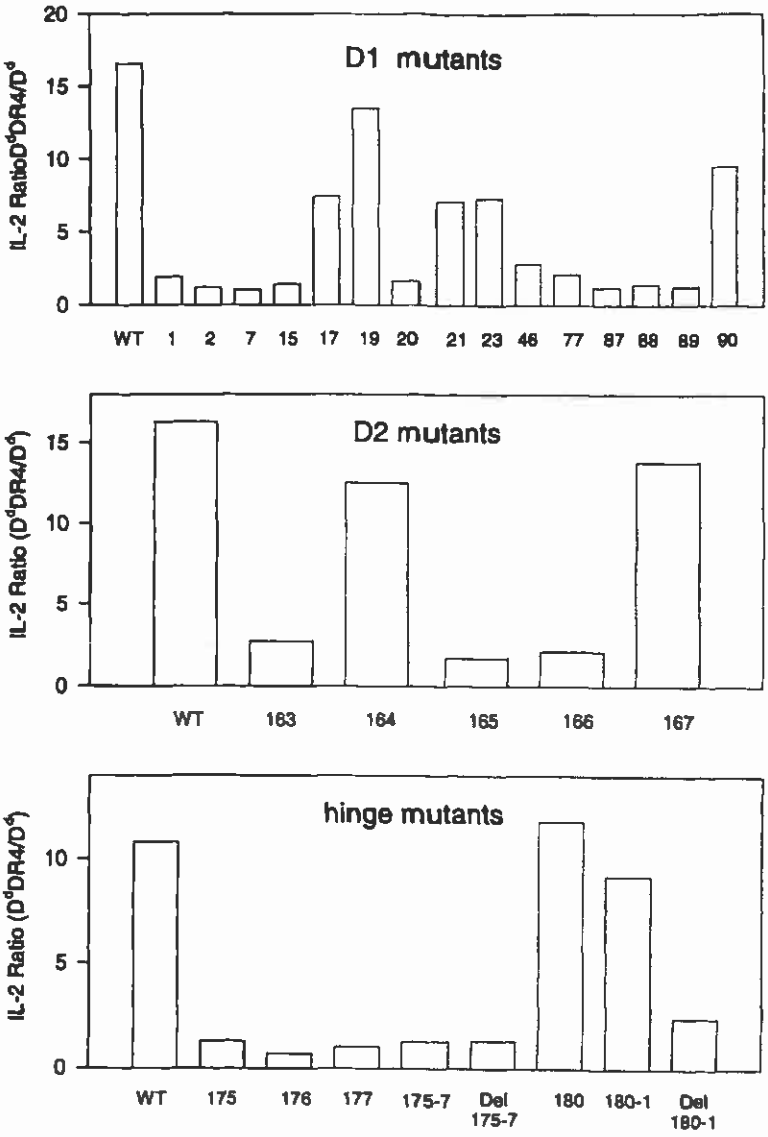


FIGURE 2. Quantitative analysis of the interaction between the CD4 mutants and the class II MHC molecules in the co-ligand D^d assay. The 3DT52.5.8 T cell hybridomas expressing the wt CD4 or different CD4 mutants in D1, D2, and hinge region were analyzed for their ability to produce IL-2 when co-cultured with DAP-3 cells expressing either D^d and D^d DR4. Results were presented as such: the *y*-axis is expressed as the IL-2 ratio (D^d DR4/D^d), as described in *Materials and Methods*. On the *x*-axis, positions of the mutated residues are indicated. Unless specifically indicated, mutants shown here are alanine substitution mutants.

$CD4^+$ cells (Fig. 3*D* and Table IV). This is different from the co-ligand assay results that showed that the CDR2 deletion mutant could still interact with different human and mouse class II alleles (10, 11). This result indicates that the contact sites on CD4 for class II in a co-ligand binding may not always be consistent with those involved in a co-receptor binding.

All the D1 alanine scan mutants except one (K90A) studied in the co-receptor assay responded to SEB at a magnitude similar to that of the $CD4^-$ cells. These CD4 mutants require 3- to 10-fold more SEB for a half-maximal response, while the IL-2 ratios are only 4 to 20% of the wt $CD4^+$ cells (Table IV). This suggests that a relatively large surface on the CD4 molecule is involved in the co-receptor contact with class II. This surface includes residues from both lateral faces of CD4. On one face it consists of residues K1, K2, K7, T15, T17, S19, Q20, K21, S23, D88, and Q89, which are from the A strand, the B strand, the CDR1 and CDR3 of D1. On the other face, it is composed of residues from the CDR2 or outside of the CDR2 (i.e., residue K46 and E77). Comparison of the CD4 contact residues mapped from the co-receptor assay with those from the co-ligand assay reveals some discrepancies. Residues T17A, S19A, K21A, and S23A are implicated only in the co-receptor assay but not in the co-ligand assay. The only mutant that did not show any effect in both assays was K90A (Fig. 3*F*), suggesting that this residue is not a direct contact residue for class II.

D2 mutants. Previous results have shown that insertion mutant 165[SR] from the FG loop of D2 affected the CD4/class II interaction in the co-ligand assay (10). When this mutant was studied in the co-receptor assay, CD4/class II interaction is still perturbed (Table IV). Moreover, all of the alanine scan mutants for residues located in the FG loop of D2 were studied in the co-receptor assay (Table IV and Fig. 3, *G–J*). The Q165A mutation affects the CD4/class II interaction in a similar way to 165[SR], further confirming the involvement of residue 165 in the interaction with class II. The other mutations that affect the class II binding are Q163A, K166A, and K167A. These mutants require four- to sixfold more SEB for the half-maximal response, and the IL-2 ratios are 10 to 40% of the wt $CD4^+$ cell ratios. However, mutant N164A can still enhance the SEB response to the level comparable to CD4 wt (Table IV). Furthermore, residue K167 appears to be differentially required in the two assays. Mutant K167A enhances the D^d stimulation as well as wt $CD4^-$ cells, as shown by an IL-2 ratio of 1.09 between the

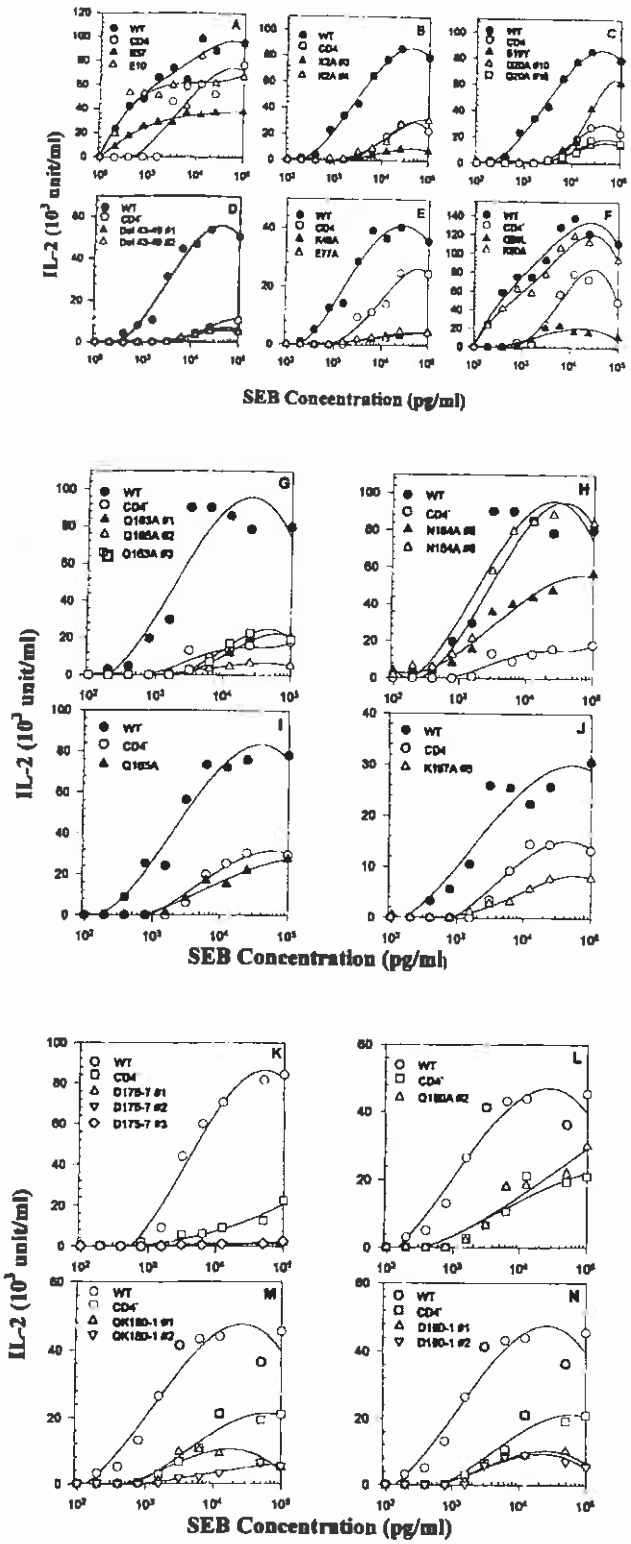


FIGURE 3. Quantitative analysis of the interaction between the CD4 mutants and the class II MHC molecules in the co-receptor SEB assay. D1, D2, and hinge region mutants were stimulated with the bacterial superantigen SEB presented by human class II molecule DR1 expressed on the DAP-3 cells. The SEB concentrations displayed on *x*-axis are: 10^5, 2.5×10^4, 1.25×10^4, 6.25×10^3, 3.13×10^3, 1.56×10^3, 7.8×10^2, 3.9×10^2, 1.95×10^2, and 0.98×10^2 pg/ml. On the *y*-axis, the IL-2 units are presented. *Top, A–F,* D1 mutants; *middle, G–J,* D2 mutants; *bottom, K–N,* the hinge region mutants.

mutant and the wild-type. However, this mutant does not enhance the SEB response. It requires about 5.8-fold more SEB to reach the half-maximal response and the IL-2 ratio is only 20% of that of the wt $CD4^+$ cells (Table IV).

Hinge region mutants. All of the mutations we generated in the hinge region showed an effect on the SEB response. Mutants V175A, V176A, L177A, VVL175–177A, Q180A, QK180–181A, and DEL 180–181 showed a reduced response to SEB (Fig. 3, *K–N,* and Table IV). At least 3 to 10 times more SEB were required to obtain the half-maximal response for cells expressing these mutant CD4 molecules. Moreover, the IL-2 ratios of the mutants are between 1 and 20% of the CD4 wt ratio. However, the implication of mutants Q180A and QK180–181A in the co-receptor assay differs from the co-ligand assay. They affect the CD4/class II interaction in the co-receptor assay, although they have no effect in the co-ligand assay. Moreover, the deletion mutant DEL 175–177 responded poorly to SEB when compared with the CD4 wt (Fig. 3*K*). In fact, this mutation has affected most of the Ab epitopes (Table III). This weak response might be due to a poor interaction between CD4 and class II molecules; alternatively, it could be due to the interference of CD4 interaction with other cell surface molecules on T cells, such as TCR or CD45.

Discussion

Based on the assumption that CD4 might use different structural components to interact with class II molecules depending on its functional role as co-ligand vs co-receptor, it is worthwhile to investigate this issue by using the 3DT52.5.8 system. The advantage of the 3DT system is that the TCR has a dual specificity. Indeed, this system allows us to study the two functional roles of CD4, co-ligand and co-receptor. It is of great interest to investigate and compare the molecular nature of the potentially different modes of CD4/class II interaction using the same cellular system.

Results from the two assays have involved residues from both the lateral faces of D1 and from the top part of D2 in interaction with class II molecules. Interestingly, though most residues implicated in both functional assay systems are quite concordant, differences do exist. The discordant residues (residues T17, S19, K21, and S23) lie mainly in the CDR1 of the D1 and are implicated in the co-receptor assay, but not in the co-ligand assay. The same pattern was noticed for residues K167 in the FG loop of D2. Computer models of the CD4 crystal structure containing D1 and D2 were generated (Fig. 4). These models reveal that residues from the D1 and D2 domains, i.e., K1, K2, K7, T15, E87, D88, Q89 from D1, and residues Q163, Q165 and K166 from D2, that when mutated have an effect in both the co-ligand and co-receptor assay, are located on the same face of the CD4 molecule (Fig. 4*A*). Furthermore, a side view of the crystal (Fig. 4*B*) shows that most of those residues cluster together, thus their implication in forming a binding site for class II is likely.

The finding that more residues are implicated in the co-receptor assay than in the co-ligand assay, suggests that a wider surface of CD4 is implicated in the co-receptor than in the co-ligand function. The exact explanation for this is not known, although it may be attributed to the differences between the two functions of CD4. In the co-receptor function of CD4, TCR may bind to the same class II molecule as CD4 binds to. Co-clustering of CD4 and TCR occurring on the T cell surface has been shown during Ag recognition (36, 37). This interaction is essential for T cell activation. Thus, it is possible that some of the mutations in CD4 may induce conformational changes that will affect the docking of CD4/TCR, rather than affect the CD4/class II interaction directly.

171

Various approaches have been used to investigate the role of the CDR2 of CD4 in its interaction with class II MHC molecules. However, different systems provided controversial results (10, 11, 13, 14, 38). In the present study, some very interesting results were revealed regarding the CDR2 of CD4 and its implication in binding to class II. First, the deletion mutant DEL 43–49, which still interacts with different class II alleles from both human and mouse in a co-ligand assay (11), can no longer enhance the IL-2 response in a co-receptor assay, suggesting that the CD4 molecule may use distinct structural components to bind to class II in either function. Moreover, single mutants K46A and E77A affect the binding to class II in both co-ligand and co-receptor function of CD4. This result is consistent with our data from rosette formation and co-ligand assay for mutants T45P, K46N/G47V, and G47R, which also seem to affect the CD4/class II interaction (10). It is likely that point mutation mutants from the CDR2 have an effect in class II interaction due to conformational effects in CDR1 and CDR3 (10). Indeed, these point mutation mutants clearly affected the serologic reactivities of mAbs that map to CDR1 and CDR3. It is also possible that these residues are involved in docking of CD4 to class II but not in the direct interaction with class II. By removing these residues, the docking of CD4 with class II is not affected. However, substitution of these residues could lead to a poor docking of these two molecules. It is also likely that these CDR2 point mutants could affect the docking of the CD4 molecule with the TCR. From this study, it seems that the CDR2 is mainly implicated in the co-receptor function, rather than its co-ligand function. A recent study using synthetic aromatically modified exocyclic analogues of CD4 supports our finding (39). These CD4 analogues were tested in a competition binding assay for their ability to inhibit the binding of recombinant soluble ^{125}I-labeled human class II DR4 to immobilized recombinant soluble CD4. They were also tested for their ability to inhibit T cell activation. Results show that a CDR3 analogue is able to bind to the β_2 domain residues 134–148 of class II. However, CDR2 analogues only show slight effect in inhibiting the CD4/class II binding and T cell activation. Thus, a direct involvement of the CDR2 in class II contact, at least for the known β_2 region sequence of class II, is not supported.

In the co-ligand assay, CD4 and the TCR bind to their respective MHC ligands, whereas in the co-receptor assay both the TCR and CD4 bind to the same class II MHC molecule. The difference between these two functions of CD4 thus can be attributed to the physical proximity of CD4 and the TCR in one function (co-receptor) but not the other (co-ligand). Therefore, the CDR2 could be a docking site for the TCR during the co-receptor function of CD4. This could explain why the CDR2 deletion mutant affects only the co-receptor function but not the co-ligand function. However, results showing the implication single mutations in both functions of CD4, and results from adhesion assay in which the binding of CD4/class II is TCR-independent do not favor such a model (Fig. 5*A*).

The recently solved crystal structure of the human class II DR1 molecule showed it as a dimer of the $\alpha\beta$ heterodimer (40). If this dimer is formed on the cell surface upon Ag stimulation of T cells, it would be logical to assume that CD4 could interact with a class II dimer using both faces. As suggested in a recent paper (9), the class II molecule DR1 could use both the α_2 and β_2 domains to interact with CD4, however, both the α_2 and β_2 domains could not bind to a single CD4 because they are separated by a 180° angle. Thus, one possibility is that one face of CD4 is involved in binding to class II α_2 domain, another face is involved in binding to class II β_2 domain. This assumption can be tested in complementation studies using both CD4 mutants and class II mutants (Fig. 5*B*).

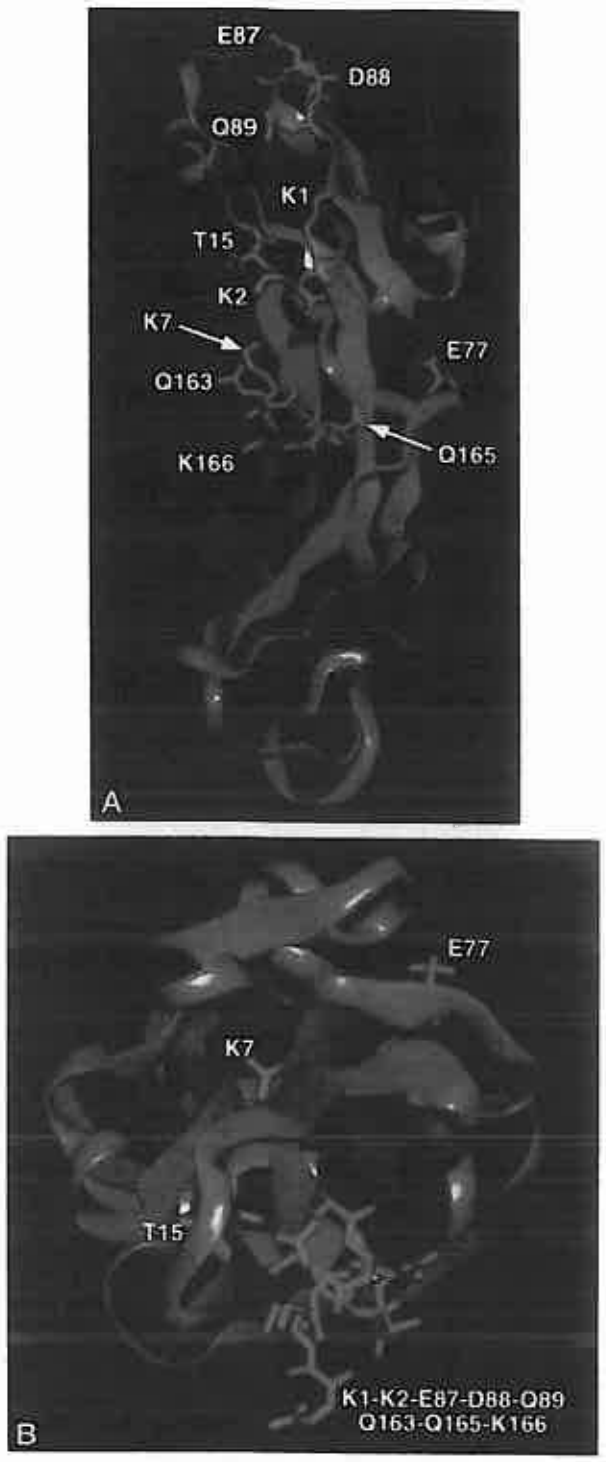


FIGURE 4. Computer model. *A*, On the crystal of the D1D2 fragment, the residues that are implicated in both the co-ligand and co-receptor assays are displayed. On one face it shows residues K1, K2, K7, T15, E87, D88, and Q89 from D1, and residues Q163, Q165, and K166 from D2; on the opposite face, E77 is presented. *B*, The side view of the crystal of the D1D2 fragment. Most of the residues implicated in class II binding from both assays are shown here to form a cluster on the surface.

Alternatively, it has been suggested that CD4 itself could dimerize (20, 35). The implication of two faces of CD4 in the above described functional assays can thus be interpreted as one face being involved in the actual contact site with class II, while the opposite face is involved in dimerization of CD4 itself. Based on

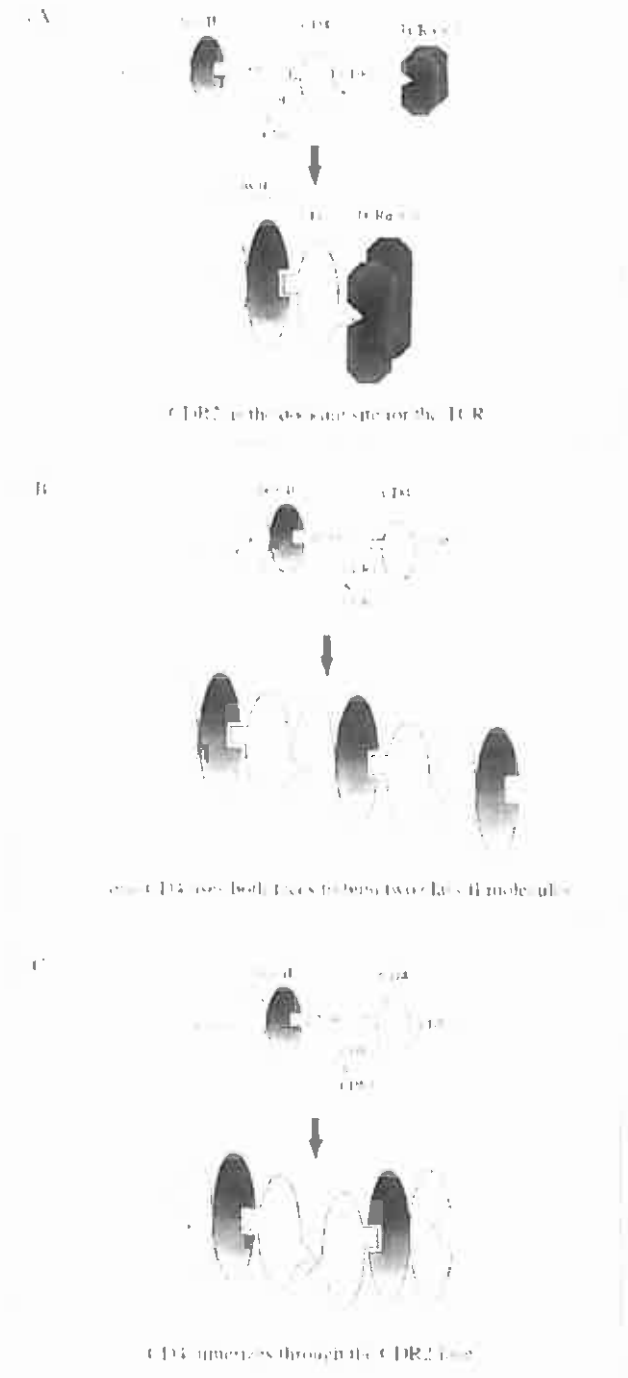


FIGURE 5. Proposed models of CD4/class II interaction. *A*, The CD4 molecule uses its residues from the face composed of the CDR1 and CDR3 to interact with the β_2 region of the class II MHC molecule, while the CDR2 is involved in the interaction with the TCR. *B*, CD4 uses both faces to interact with a class II superdimer that contains two class II $\alpha\beta$ dimers. On one face, it contains residues from CDR1 and CDR3; on the opposite face it contains the CDR2. *C*, CD4 dimerizes through the face containing the CDR2, while the opposite face containing CDR1 and CDR3 is involved in binding to the β_2 domain of class II.

the structural and functional distinction between the two assays, we can propose the following: in the co-ligand assay, dimerization or oligomerization of CD4 might not be required, since CD4 and TCR binds to different MHC molecules, there might not be any specifically organized CD4 and/or class II dimerization or oligomerization event; however, in the co-receptor assay, such a specifically organized CD4 and/or class II dimerization or oligomerization event could be induced, since CD4 and TCR bind to the same class II molecule. Whether this assumption is the reflection of the physiologic situation needs further studies (Fig. 5*C*).

It is clear that the contact sites for class II involve multiple faces of CD4, and that a bigger region is implicated in the co-receptor than in the co-ligand function of CD4. At least three molecular interactions are involved in the former function: CD4, TCR, and class II, which could explain the fact that they are more sensitive to any structural changes. It would not be surprising to find out that some of the residues mapped are the contact residues, while others are just docking residues. However, this can only be solved by the cocrystallization of CD4 and class II. It could also be possible for a specifically organized oligomerization of the three molecules to occur in the co-receptor assay, but, not in the co-ligand assay. Furthermore, and consistent with our results, studies on CD8$\alpha\alpha$ mutants revealed a multiple faces involved in interaction with class I MHC molecules, suggesting that CD4 and CD8 use several faces of the same structural unit to interact with their respective ligands (41).

To our knowledge, the hinge region of CD4 and its functional importance in the light of class II interaction was studied here for the first time. Due to lack of crystallographic data of the hinge region, the elucidation of the structure relies mostly on the Ab epitope mapping. Serologic studies on hinge region mutants have shown that many epitopes throughout D1 to D4 domains are affected by mutagenesis in this region. Furthermore, this study shows that mutations in the hinge region also affect the CD4/class II interaction drastically, regardless of the co-ligand or co-receptor function of CD4. The abrogation of class II binding following by mutations, especially single-amino acid substitutions, indicates that these residues are closely involved in maintaining a structure essential for class II interaction. These hinge residues may not be the direct contact residues for class II, as overall conformational changes occur as the result of mutations. Rather, it could suggest that loss of CD4 function may be a consequence of interfering the interaction of CD4 with other surface molecules on the T cells, such as the TCR, or CD45; alternatively, these perturbations could interfere with the possible dimerization/oligomerization of CD4. At this stage, we do not know the exact mechanism for such a drastic functional effect by the mutations. Nevertheless, this result indicates a very important role for the hinge region in maintaining the overall structure of CD4, and that such a conformation is essential for CD4 to fulfill its function.

Acknowledgments

We thank Claude Cantin for his technical assistance in flow cytometry. We thank A. Peterson, B. Seed, T. Fuerst, and D. R. Littman for plasmids. We thank A. Ashkenazi, D. Capon, A. Peterson, B. Seed, T. Mizukami, T. Fuerst, and D. R. Littman for plasmids. We thank David Buck, S. Canel, Q. Sattentau, P. Rao, A. Truneh, and Skip Meino for Abs.

References

1. Maddon, P. J., D. R. Littman, M. Godfrey, D. E. Maddon, L. Chess, and R. Axel. 1985. The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein T4: a new member of the immunoglobulin gene family. *Cell 42:93*
2. Veillette, A., M. A. Bookman, E. M. Horak, and J. B. Bolen. 1988. The CD4 and CD8 T cell surface antigens are associated with the internal membrane tyrosine-protein kinase p56^{lck}. *Cell 55:301.*
3. Janeway, C. A. 1992. The T cell receptor as a multicomponent signalling machine: CD4/CD8 coreceptors and CD45 in T cell activation. *Annu. Rev. Immunol. 10:645*

173

4. Dalgleish, A. G., P. C. L. Beverley, P. R. Clapham, D. H. Crawford, M. F. Greaves, and R. A. Weiss. 1984. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature 312:763.*
5. Klatzmann, D., E. Champagne, S. Chamaret, J. Gruest, D. Guetard, T. Hercend, J. C. Gluckman, and L. Montagnier. 1984. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature 312:767.*
6. Maddon P. J., A. G. Dalgleish, J. S. McDougal, P. R. Clapham, R. A. Weiss, and R. Axel. 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell 47.333.*
7. Konig, R., L. Y. Huang, and R. N. Germain. 1992. MHC class II interaction with CD4 mediated by a region analogous to the MHC class I binding site for CD8. *Nature 356:796.*
8. Cammarota, G., A. Scheirle, B. Takacs, D. M. Doran, R. Knorr, W. Bannwarth, J. Guardiola, and F. Sinigaglia. 1992. Identification of a CD4 binding site on the β_2 domain of HLA-DR molecules. *Nature 356:799.*
9. Konig, R., X. Shen, and R. N. Germain. 1995. Involvement of both major histocompatibility complex class II α and β chains in CD4 function indicates a role for ordered oligomerization in T cell activation. *J. Exp. Med. 182:779.*
10. Fleury, S., D. Lamarre, S. Meloche, S.-E. Ryu, C. Cantin, W. A. Hendrickson, and R.-P. Sekaly. 1991. Mutational analysis of the interaction between CD4 and class II MHC: class II antigens contact CD4 on a surface opposite to the gp120-binding site. *Cell 66:1037.*
11. Fleury, S., B. Huang, A. Zerbib, G. Croteau, E. O. Long, and R. P. Sekaly. 1996. Mutations in human CD4 impair the functional interaction with different human and mouse class II isotypes and alleles. *J. Immunol. 156:1848.*
12. Gustafson, K., S. Germana, T. M. Sundt III, D. H. Sachs, and C. LeGuern. 1993. Extensive allelic polymorphism in the CDR2-like region of the miniature swine CD4 molecule. *J. Immunol. 151:1365.*
13. Platier-Tonneau, D., L. N. Gastinel, G. Moussy, B. Benichou, F. Amblard, P. Vaigot, and C. Auffray. 1991. Mutation in the D strand of the human CD4 V1 domain affect CD4 interaction with the human immunodeficiency virus envelop glycoprotein gp120 and HLA class II antigens similarly. *Proc. Natl. Acad. Sci. USA 88:6858.*
14. Moebius, U., L. L. Clayton, S. Abraham, A. Diener, J. J. Yunis, S. C. Harrison, and E. L. Reinherz. 1992. Human immunodeficiency virus gp120 binding C'C'' ridge of CD4 domain 1 is also involved in interaction with class II major histocompatibility complex molecules. *Proc. Natl. Acad. Sci. USA 89:12008.*
15. Moebius, U., P. Pallai, S. C. Harrison, and E. L. Reinherz. 1993. Delineation of an extended surface contact area on human CD4 involved in class II major complex binding. *Proc. Natl. Acad. Sci. USA 90 8259.*
16. Wang, J., Y. Yan, T. P. J. Garrett, J. Liu, D. W. Rodgers, R. L. Garlick, G. E. Tarr, Y. Husain, E. L. Reinherz, and S. C. Harrison. 1990. Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature 348:411.*
17. Ryu, S., P. D. Kwong, A. Truneh, T. G. Porter, J. Arthos, M. Rosenberg, X. Dai, N. Xuong, R. Axel, R. W. Sweet, and W. A. Hendrickson. 1990. Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature 348:419.*
18. Brady, R. L., E. J. Dodson, G. G. Dodson, G. Lange, S. J. Davis, A. F. Williams, and A. N. Barclay. 1993. Crystal structure of domains 3 and 4 of rat CD4: relation to the NH_2-terminal domains. *Science 260:979.*
19. Hendrickson, W. A., P. D. Kwong, D. J. Leahy, S. E. Ryu, H. Yamaguchi, S. Fleury, and R. P. Sekaly. 1992. Structural aspects of CD4 and CD8 involvement in the cellular immune response. *Cold Spring Harbor Symp. Quant. Biol. 57:549.*
20. Kwong, P. D., S. E. Ryu, W. A. Hendrickson, R. Axel, R. M. Sweet, G. Folena-Wasserman, P. Hensley, and R. W. Sweet. 1990. Molecular characteristics of recombinant human CD4 as deduced from polymorphic crystals. *Proc. Natl. Acad. Sci. USA 87 6423.*
21. Richardson, N. E., N. R. Brown, R. E. Hussey, A. Vaid, T. J. Matthews, D. P. Bolognesi, and E. L. Reinherz. 1988. Binding site for human immunodeficiency virus coat protein gp120 is located in the NH_2 terminal region of T4 (CD4) and requires the intact variable region-like domains. *Proc. Natl. Acad. Sci. USA 85:6102.*
22. Ibegbu, C. C., M. S. Kennedy, P. J. Maddon, K. G. Deen, S. D. Hicks, R. W. Sweet, and J. S. McDougal. 1989. Structural features of CD4 required for binding to HIV. *J. Immunol 142:2250.*
23. Healey, D., L. Dianda, J. P. Moore, J. S. McDougal, M. J. Moore, P. Estess, D. Buck, P. D. Kwong, P. C. L. Beverley, and Q. J. Sattentau. 1990. Novel anti-CD4 monoclonal antibodies separate human immunodeficiency virus infection and fusion of CD4$^+$ cells from virus binding. *J. Exp. Med. 172:1233.*
24. Hart, T. K., R. Kirsh, H. Ellens, R. W. Sweet, D. M. Lambert, S. R. Petteway, Jr., J. Leary, and P. J. Bugelski. 1991. Binding of soluble CD4 proteins to human immunodeficiency virus type 1 and infected cells induces release of envelop glycoprotein gp120. *Proc. Natl. Acad. Sci. USA 88.2189.*
25. Sattentau, Q. J., and J. P. Moore. 1991. Conformational changes induced in the human immunodeficiency virus envelop glycoprotein by soluble CD4 binding. *J. Exp. Med. 174:407.*
26. Burkly, L. C., D. Olson, R. Shapiro, G. Winkler, J. J. Rosa, D. M. Thomas, C. Williams, and P. Chisholm. 1992. Inhibition of HIV infection by a usual CD4 domain 2-specific monoclonal antibody. *J. Immunol. 149:1779.*
27. Ashkenazi, A., L. G. Presta, S. A. Marsters, T. R. Camerato, K. A. Rosenthal, B. M. Fendly, and D. J. Capon. 1990. Mapping the CD4 binding site for human immunodeficiency virus by alanine scanning mutagenesis. *Proc. Natl. Acad. Sci. USA 87.7150.*
28. Peterson, A., and B. Seed. 1988. Genetic analysis of monoclonal antibody and HIV binding sites on the human lymphocyte antigen CD4. *Cell 54:65.*
29. Mizukami, T., T. R. Fuerst, E. A. Berger, and B. Moss. 1988. Binding region for human immunodeficiency virus (HIV) and epitopes for HIV-blocking monoclonal antibodies of the CD4 molecule defined by site-directed mutagenesis. *Proc. Natl. Acad. Sci. USA 85:9273.*
30. Ho, S. N., H. D. Hunt, R. M. Horton, J. K. Pullen, and L. R. Pease. 1989. Site-directed mutagenesis by overlap expression using the polymerase chain reaction. *Gene 77:51.*
31. Sleckman, B. P., A. Peterson, J. A. Foran, J. C. Gorga, C. J. Kara, J. L. Strominger, S. J. Burakoff, and J. L. Greenstein. 1988. Functional analysis of a cytoplasmic domain-related mutant of the CD4 molecule. *J. Immunol. 141:49.*
32. Takebe, Y. 1988. Sra promoter: an efficient and versatile mammalian cDNA expression system composed of the simian virus 40 early promoter and the R-U5 segment of human T-cell leukemia virus type 1 long terminal repeat. *Mol. Cell. Biol. 8:466.*
33. Graham, F. L., and A. J. van der Eb. 1973. Transformation of rat cells by DNA of human adenovirus 5. *Virology 52:456.*
34. Lamarre, D., A. Ashkenazi, S. Fleury, D. H. Smith, R. P. Sekaly, and D. J. Capon. 1989. The MHC-binding and gp120-binding functions of CD4 are separable. *Science 245:743.*
35. Sakihama, T., A. Smolyar, and E. L. Reinherz. 1995. Oligomerization of CD4 is required for stable binding to class II major histocompatibility complex proteins but not for interaction with human immunodeficiency virus gp120. *Proc. Natl. Acad. Sci. USA 92:6444.*
36. Kupfer, A., S. J. Singer, C. A. J. Janeway, and S. L. Swain. 1987. Coclustering of CD4 (L3T4) molecule with the T-cell receptor is induced by specific direct interaction of helper T cells and antigen-presenting cells. *Proc. Natl. Acad. Sci. USA 84:5888.*
37. Vignali, D. A. A., R. T. Carson, B. Chang, R. S. Mittler, and J. L. Strominger. 1996. The two membrane proximal domains of CD4 interact with the T cell receptor. *J. Exp. Med. 183:2097.*
38. Bowman, M. R., K. D. MacFerrin, S. L. Schreiber, and S. J. Burakoff. 1990. Identification and structural analysis of residues in the V1 region of CD4 involved in interaction with human immunodeficiency virus envelope glycoprotein gp120 and class II major histocompatibility complex molecules. *Proc. Natl. Acad. Sci. USA 87:9052.*
39. Zhang X., D. Platier-Tonneau, C. Auffray, R. Muroli, A. Mahapatra, F. Zhang, C. C. Maier, H. Saragovi, and M. I. Greene. 1996. Synthetic CD4 exocyclic peptides antagonize CD4 holoreceptor binding and T cell activation. *Nat. Biotechnol. 14:472.*
40. Brown, J. H., T. S. Jardetzky, J. C. Gorga, L. J. Stern, R. G. Urban, J. L. Strominger, and D. C. Wiley. 1993. Three-dimensional structures of the human class II antigen HLA-DR1. *Nature 364:33.*
41. Giblin, P. A., D. J. Leahy, J. Mennone, and P. B. Kavathas. 1994. The role of charge and multiple faces of the CD8 alpha/alpha homodimers in binding to major histocompatibility complex class I molecules: support for a bivalent model. *Proc. Natl. Acad. Sci. USA 91 1716.*

Downloaded from rstb.royalsocietypublishing.org on July 27, 2011

Table 3. *Summary of crystallizations*

protein	crystals	resolution of diffraction
rat sCD4		
1 sCD4	no	—
2 sCD4.C1	no	—
3 sCD4.C2	no	—
4 sCD4.C0	no	—
5 sCD4.C0 domains 3,4 papain fragment	no	
6 sCD4D1D2	no	—
7 sCD4D3D4	yes	2.8 Å
rat sCD4–Fab complexes		
1 sCD4.W3/25 IgG1 Fab	yes	
Type 1		3.5–5.0 Å[a]
Type 2		8 Å
2 sCD4:W3/25 IgG2A Fab	yes	
Type 1		3.5–5.0 Å[a]
3 sCD4.C1:W3/25 Fab	no	—
4 sCD4.C2:W3/25 Fab	yes	
Type 2		8 Å
5 sCD4.C0:W3/25 Fab	no	—
6 sCD4:OX35 Fab	no	
7 sCD4:OX38 Fab	no	—
8 sCD4:OX67 Fab	no	—
9 sCD4:OX68 Fab	no	—
10 sCD4D1D2:W3/25 Fab	no	
human sCD4		
1 native	yes	> 7 Å
2 desialated	yes	4.5 Å

[a] Crystal disorder restricts resolution along 1 axis.

restricted responses, in HIV binding and in regard to the validity of structural predictions based on its inclusion in the IgSF. The production of crystals suitable for X-ray crystallography is not a straightforward process and satisfactory crystals of the complete extracellular part of CD4 have so far not been obtained (Davis *et al.* 1990*a*; Kwong *et al.* 1990). We have expressed soluble forms of rat CD4 (sCD4) in eucaryotic cells and in order to maximize the chances of getting good crystals, mutants were made that lacked one or other or both of the glycosylation sites. The sCD4 proteins were produced by Chinese hamster ovary cell lines using the glutamine synthetase expression system developed at Celltech Ltd., Slough by C. Bebbington and colleagues. This expression system gives the rapid production of stable cell lines expressing very high levels of recombinant protein (Bebbington & Hentschel 1987; Cockett *et al.* 1990; Davis *et al.* 1990*b*). Several different constructs were expressed at around 100 mg l^{-1} when the cell cultures were grown to exhaustion (Davis *et al.* 1990*b*). The various different constructs expressed are summarized in figure 1.

Crystallization was attempted on the variants of CD4 and to maximize chances of obtaining suitable crystals for X-ray diffraction, complexes of sCD4 with Fab from various monoclonal antibodies were made and crystallization attempted. The various conditions are summarized in table 3 which shows that most protein forms did not give suitable crystals for X-ray diffraction. It had been hoped that the removal of carbohydrate sites would help crystallization but in this case it actually prevented crystallization. One complex of sCD4 with Fab gave moderate diffraction but good diffraction was obtained with the two domain form sCD4 D3D4. A structure for this protein

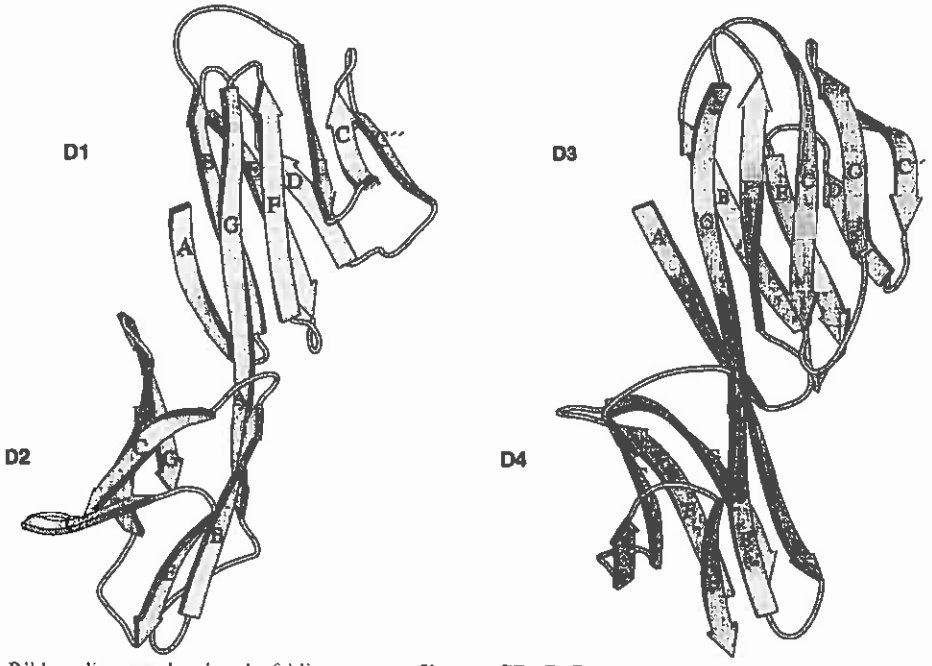


Figure 4. Ribbon diagram showing the folding pattern of human CD4 D1D2 compared to rat CD4 D3D4. Data from Ryu *et al.* (1990); Wang *et al.* (1990); Brady *et al.* (1993). The beta strands are labelled as in figures 2 and 3.

Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule

Jia-huai Wang[†‡§¶], Rob Meijers[†‖], Yi Xiong[†‖], Jin-huan Liu[†‖], Toshiko Sakihama[†‖**], Rongguang Zhang[††], Andrzej Joachimiak[††], and Ellis L. Reinherz[†‖‖]

[†]Laboratory of Immunobiology and Department of Cancer Immunology and AIDS, Dana–Farber Cancer Institute, and Departments of [‡]Pediatrics, [§]Biological Chemistry and Molecular Pharmacology, and [‖]Medicine, Harvard Medical School, Boston, MA 02115; and [††]Biosciences Division, Argonne National Laboratory, Argonne, IL 60439

Edited by Philippa Marrack, National Jewish Medical and Research Center, Denver, CO, and approved July 20, 2001 (received for review March 13, 2001)

The structural basis of the interaction between the CD4 coreceptor and a class II major histocompatibility complex (MHC) is described. The crystal structure of a complex containing the human CD4 N-terminal two-domain fragment and the murine I-A^k class II MHC molecule with associated peptide (pMHCII) shows that only the "top corner" of the CD4 molecule directly contacts pMHCII. The CD4 Phe-43 side chain extends into a hydrophobic concavity formed by MHC residues from both α2 and β2 domains. A ternary model of the CD4-pMHCII-T-cell receptor (TCR) reveals that the complex appears V-shaped with the membrane-proximal pMHCII at the apex. This configuration excludes a direct TCR–CD4 interaction and suggests how TCR and CD4 signaling is coordinated around the antigenic pMHCII complex. Human CD4 binds to HIV gp120 in a manner strikingly similar to the way in which CD4 interacts with pMHCII. Additional contacts between gp120 and CD4 give the CD4–gp120 complex a greater affinity. Thus, ligation of the viral envelope glycoprotein to CD4 occludes the pMHCII-binding site on CD4, contributing to immunodeficiency.

The adaptive immune response depends on the specific recognition by a T-cell receptor (TCR) of an antigenic peptide bound to a major histocompatibility complex (MHC) molecule (pMHC), as well as interaction of this same pMHC with a CD8 or CD4 coreceptor (1–4). The structural basis of the TCR-pMHCI, TCR-pMHCII, and CD8$\alpha\alpha$–pMHCI interactions has been crystallographically defined (1, 5–8), whereas the structural nature of CD4-pMHCII is not yet known. The extracellular rod-like CD4 segment consists of four concatamerized Ig-like domains (D1–D4) (9–11). Mutagenesis studies have suggested that residues in the membrane distal D1-D2 module bind to predominantly nonpolymorphic residues of MHC class II molecules (12–14). During pMHCII recognition, CD4 and TCR colocalize to interact with the same pMHCII (2). Hence, CD4 brings the p56^{lck} that associates with the short CD4 cytoplasmic tail to the site of immune recognition (15). In fact, such tyrosine kinase recruitment, rather than any major contribution to binding energy, appears to be the major function of CD4 (16).

Structural information about the CD4 receptor–ligand interaction would be of particular importance, given the expression of CD4 on the surface of all helper T cells and the essential role of this protein in their activation. Once triggered, helper T cells stimulate B-cell antibody production and cytolytic T-cell generation, thereby regulating the induction of the immune response. Development of methods to block CD4–pMHCII interaction might lead to an important new class of immunosuppressive compounds, depending on structural data. With this goal in mind, we have determined the crystal structure of a CD4–pMHCII complex. As reported below, it is now possible to compare the nature of coreceptor binding to pMHC for both CD4 and CD8 T-cell subsets as well as to examine how CD4 is ligated by its physiologic pMHCII ligand versus the pathologic HIV1 gp120 ligand.

Materials and Methods

Crystallization and Data Collection. The human CD4 N-terminal two-domain fragment (amino acid residues 1–183) and the murine class II molecule I-A^k with a covalently linked 13-residue hen egg conalbumin (CA) peptide (residues 134–146) were prepared as previously reported (10, 17). Crystals were grown by using the hanging-drop vapor-diffusion method. CD4 and CA/I-A^k were stoichiometrically mixed to a concentration of 20 mg/ml. The complex solution was further mixed with reservoir crystallization buffer in equal volume to 2 μl. The reservoir contains 1 ml of 17% polyethylene glycol 4,000/0.2 M Li_2SO_4/0.1 M Tris at pH 8.5. The rod-like crystals of 0.25 × 0.05 × 0.05 mm grew readily at room temperature in a few days. They belong to space group $P4_322$ with unit cell dimension: $a = b = 145.2$ Å and $c = 103.8$ Å. There is one complex in the asymmetric unit with a solvent content of ~68%. These crystals diffract very anisotropically (3.5 Å along the c-axis, the rod direction, and only around 4.5–5 Å in the perpendicular direction) and are extremely sensitive to any changes in harvest condition.

Data were collected at the SBC-CAT 19ID beamline of the Advanced Photon Source at Argonne National Laboratories. The best data set was obtained from a single prefrozen crystal that was directly picked up from the crystallization droplet with a loop, "soaked" in mineral oil for half a second, and then dipped into liquid nitrogen for freezing. Data with an oscillation range covering 160° were collected to ensure a high degree of redundancy. The data were processed by using the HKL2000 suite (18). For the overall data within 30–4.3 Å, the $I/\sigma = 12.0$, $R_{merge} = 15.2\%$; completeness = 83% (100% completeness to 5 Å); redundancy = 9.6; whereas at the outer resolution bin (4.45–4.3 Å), the corresponding figures are 1.9, 61%, 56%, and 7.8, respectively.

Structure Determination. The structure was determined with the molecular replacement method. The search for I-A^k (without a peptide) was first carried out within a data range of 10–5 Å with the program AMoRe (19) by using the pdb file 1IAK (20) as a search model. The solution was straightforward, giving an out-

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: MHC, major histocompatibility complex; pMHCII, peptide bound to a MHC class II molecule; pMHCII, peptide bound to a MHC class I molecule; TCR, T-cell receptor; CA, conalbumin.

Data deposition: The atomic coordinates reported in this paper have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID code 1JL4).

¶To whom reprint requests may be addressed. E-mail: jwang@red.dfci.harvard.edu or ellis_reinherz@dfci.harvard.edu.

**Present address: Department of Experimental Pathology, Institute for Frontier Medical Sciences, Kyoto University, 53 Shogoin Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

IMMUNOLOGY

standing rotation function peak with a correlation coefficient of 6.3 versus 3.3 for the next peak. After the fitting step, the correlation coefficient reached a value of 35.1, and the R factor was 49.8% at 5 Å. The CD4 molecule was located by using an evolutionary six-dimensional search algorithm, implemented in the program EPMR (21), also at 5 Å resolution. The properly positioned MHC molecule was added as a constraint such that the centers of gravity of the CD4 molecule and the MHC molecule are 12 Å apart. The best solution has a correlation coefficient of 60.3 versus 51.2 for the second highest. In addition, FFFEAR (22) was run to perform a reciprocal space search with data to 4.3 Å to position the CD4 molecule by using phases from the properly placed MHC molecule. The rms deviation between the models derived from FFFEAR and EPMR was 1.5 Å for all Cα atoms. Further rigid body refinement, including all measured reflections in the range 20–4.3 Å with REFMAC (23), gave an R factor of 42% (48% from 4.5 to 4.3 Å) and an R_{free} of 45% (48% from 4.5 to 4.3 Å) by using bulk solvent correction and anisotropic scaling. The rigid bodies consisted of each of the four domains of the class II MHC molecule, the peptide, and each of the two domains of the CD4 D1-D2 molecule. No major domain movements were observed on rigid body refinement. TLS refinement was performed on the same rigid bodies to get an estimate for the atomic displacements of the domains. We obtained relatively low B values for the class II MHC molecule (57 Å^2 for α1, 52 Å^2 for α2, 63 Å^2 for β1, and 47 Å^2 for β2) and slightly higher values for the peptide (82 Å^2) and the respective CD4 domains (73 Å^2 for D1 and 100 Å^2 for D2). Molecular dynamics were performed on the side chains of Phe-43 and Arg-59 with CNS (24) to ascertain their optimal position, whereas the remainder of the model was not adjusted further. The structure coordinates arrived at after rigid body refinement, as well as the structure factor amplitudes, have been deposited in the Protein Data Base under code 1JL4.

Mutagenesis, Transfection, and Binding Assay. Mutagenesis of human CD4 by using single-stranded DNA and oligonucleotides (18–22 bases) was carried out as described (12). Mutant CD4 was subcloned into the *Xba*I site of the CDM8 vector. COS7 cells were plated into each well of Falcon six-well dishes and were transiently transfected with 5 μg of wild-type or mutant CD4 DNA by the calcium phosphate/chloroquine method. Two days after transfection, binding of MHC class II$^+$ human B lymphoblastoid Raji cells to CD4-transfected COS7 cells was assayed as described (12). Briefly, 1×10^7 Raji cells were added onto the COS7 cell monolayer and incubated at 37°C for 1 h. Unbound Raji cells were removed by aspiration and further washing, and binding of CD4 to class II MHC was enumerated by counting the number of COS7-cell-B-cell rosettes in 10 random optical fields under a microscope. Cells transfected with CDM8 vector alone served as negative controls for rosette formation. Cell-surface CD4 expression was determined by using mAbs and indirect immunofluorescence, as described (13, 14).

Results and Discussions

CD4–pMHCII Complex. Prior studies established that human CD4 interacts with murine as well as human pMHCII (25–27). To examine the CD4–pMHCII interaction structurally, the human CD4 two domain fragment (amino acids 1–183) (10) and the murine class II molecule I-A^k (17) were stoichiometrically mixed and cocrystallized. The tetragonal crystals diffract anisotropically, to beyond 3.5 Å in the best direction and to within 4.5–5 Å in the worst direction. The structure was determined by using the molecular replacement method and rigid-body refined to 4.3 Å with an R factor reaching 42%. Although the complex structure has a limited resolution, the result can be interpreted with confidence in light of known high-resolution structures of each of the two molecular components. This information includes the general binding topology and interacting structural elements involved. However, at this relatively low resolution, any description of detailed interaction is tentative.

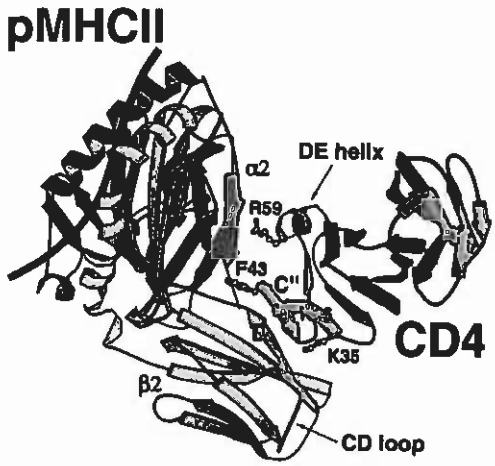


Fig. 1. Ribbon diagram of the CD4–pMHCII complex. The murine I-A^k MHC class II molecule with a CA peptide (green) bound to the antigen-presenting platform interacts with hCD4 (cyan) through both the α2 (red) and the β2 (yellow) domains of pMHCII and domain 1 (D1) of hCD4. Residues Lys-35, Phe-43, Lys-46, and Arg-59 on CD4 D1 essential for binding are highlighted. The CD loop (delineated with an arrow) on the β2 domain of the I-A^k molecule is shown to have no direct interactions with CD4. Note also how CD4 D2 (unlabeled, cyan) makes no contact to pMHCII. All the figures were prepared with MOLSCRIPT (46) and RASTER 3D (47).

The complex structure shows that the CD4 N-terminal immunoglobulin variable region-like domain is directed toward and reaching into the two membrane-proximal domains of the MHC class II molecule. The C terminus of the CD4 molecule points away from pMHCII (Fig. 1). Consequently, domain 2 of the CD4 fragment makes no contact with the pMHCII molecule. In other words, the long axes of the two molecules are neither antiparallel, as expected, nor perpendicular, as in the case of the CD8–pMHCI interaction (5, 8). There are three major binding elements on the N-terminal domain of CD4: Phe-43, the C" strand, and the short α-helical segment between β strands D and E. Collectively, these structural elements are located at the "top corner" of the domain, encompassing what has been previously referred to as the C′C" ridge (10).

The most striking feature of the interactions is that the aromatic ring of Phe-43 at the beginning of the C" strand is inserted into a site surrounded by a conserved group of hydrophobic pMHCII residues, appearing to include Val-91, Phe-92, and Trp-178 from the α2 domain as well as Ile-148 and Leu-158 from the β2 domain. Fig. 2*A* is a local view of this area and depicts how α2 and β2 domains of the pMHCII molecule together form the concavity for CD4 binding. Fig. 2*B* shows the position of these hydrophobic MHC class II residues relative to the CD4 C" strand. Superimposed is an omit map showing a local region where the C" strand lies. The rather striking conservation of residues in this region between human and mouse alleles and across loci is evident from the Fig. 2*C* sequence alignment.

The second major CD4–MHC class II interaction is between the middle portions of the CD4 C" strand and the pMHCII β2 domain D strand. The backbones of CD4 Lys-46–Leu-44 and pMHCII Ser-144–Gln-146 segments meet to form a much twisted, antiparallel mini-β-sheet. This mini-β-sheet helps bring the Phe-43 at the beginning of the C" strand into the contact site to perform its key binding function.

The third interaction site involves the helical region between

177

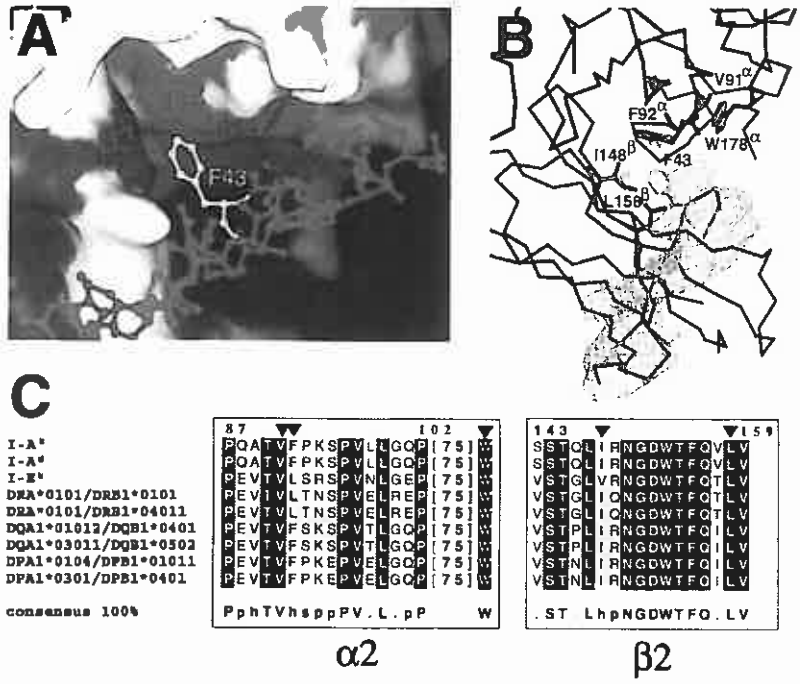


Fig. 2. The conserved hydrophobic pocket formed by class II MHC α2 and β2 domain residues into which CD4 Phe-43 inserts. (*A*) Class II MHC is depicted in surface representation form and adjacent CD4 residues in ball-and-stick format. Hydrophobic residues from domain α2 (light green) and domain β2 (dark green) of the I-A^k molecule are shown. The positions of the side chains of Phe-43 (yellow) and Arg-59 (blue) have been optimized by molecular dynamics by using CNS (24). The surface representation was prepared with SPOCK. (*B*) The same region is presented with residues that contribute to the hydrophobic pocket shown, including Val-91, Phe-92, and Trp-178 from the α2 domain and Ile-148 and Leu-158 from the β2 domain of the MHC class II molecule in green. The Cα trace of the MHC class II molecule in this region is depicted in dark red. The Cα trace of the C″ strand of the CD4 molecule is depicted in cyan, and the Phe-43 residue is shown in yellow. A portion of an omit map at a contour level of 1σ is shown in gold for the C″ strand of CD4, which was obtained by rigid body refinement with this segment excluded from the calculation. (*C*) Sequence alignment of the highly conserved contact regions in MHC class II for different murine (I-A^k, I-A^d, and I-E^k) and human (DRA*0101/DRB1*0101, DRA*0101/DRB1*04011, DQA1*01012/DQB1*0401, DQA1*03011/DQB1*0502, DPA1*0104/DPB1*01011, DPA1*0301/DPB1*0401) alleles by using CLUSTALX (48). Residues Val-91, Phe-92, and Trp-178 of the α2 domain and Ile-148 and Leu-158 of the β2 domain are indicated by a triangle. The consensus is defined so that entirely conserved residues are depicted by their amino acid code, whereas *p* indicates conservation of a polar residue, *h* conservation of a hydrophobic residue, and *s* conservation of a small side-chain-containing residue.

β strands D and E of the CD4 N-terminal domain and approximated residues of the pMHCII α2 domain. There is a group of charged and other hydrophilic residues in the contact area for both interacting partners that could potentially form a series of salt bridges and hydrogen bonds between the two molecules. These hydrophilic contacts may include Ser-60 and Asp-63 from CD4 and Gln-88α and Lys-176α of pMHCII, respectively. In addition, other hydrophilic interactions might involve CD4 residues Lys-35, Lys-46, and Arg-59 that are within reach of Glu-162β and Gln-146β as well as Gln-88α and Asp-110α of pMHCII. As suggested by analysis of other surface proteins involved in cell–cell interactions and their "hot spots," a hydrophobic residue such as Phe-43 may provide the major binding energy, whereas surrounding charged residues facilitate specificity (1). A role for both β2 and α2 pMHCII domains in the CD4 interaction was anticipated by domain swap and mutagenesis studies, although the functional disruption from those pMHCII point mutations is secondary to internal structural rearrangements (28–30).

Although the CD8αα homodimer also uses the N-terminal portions of its paired variable-like domains to contact a pMHCI molecule, it does so differently, in several respects, from the way in which CD4 binds to a pMHCII (5, 8). The major CD8-binding site on pMHCI is the CD loop of the α3 domain, the counterpart of the β2 domain in pMHCII. By contrast, the major CD4-binding site on pMHCII involves the junction of α2 and β2 domains (the D and E strands of β2 and the A and G strands of α2). There are no direct contacts between CD4 and the CD loop of the pMHCII β2 domain. CD8αα clamps onto this CD loop of the pMHCI α3 domain in an antibody-like manner with all six CDRs (three from each V-like domain) across the top of the two CD8 subunits. On the other hand, CD4 interacts with pMHCII by using only the "top corner." Finally, CD8 binding causes the α3 domain to swing relative to the rest of the pMHCI molecule (5, 8), whereas CD4 binding does not seem to induce significant pMHCII domain movements.

The structure agrees well with CD4 mutational data if one considers those mutations that do not alter CD4 expression (≥90% of wild-type CD4 levels) yet significantly impair class II MHC binding (≤30% rosetting), as determined by adhesion assays between hCD4 transfected COS cell and class II MHC expressing human B lymphoblastoid cells and, in the case of Phe-43→Ile, analysis of transgene mice (13, 14, 31). On the other hand, mutations that more modestly impair cell binding and/or are associated with reduced CD4 cell surface expression in the functional adhesion assay do not correspond to CD4-pMHCII contact sites observed crystallographically. The precise geometry of the C′C″ ridge relative to the rest of CD4 D1 is undoubtedly important because internal alterations that shift or eliminate hydrogen bonds, for example, mutations of Trp-62→Tyr or Ser-49→Val, strongly disrupt pMHCII binding. Only CD4 Lys-72 and Thr-81 residues fail to make direct contacts to pMHCII in the crystallographic structure yet virtually eliminate pMHCII binding when mutated to alanine. We did not find crystal contacts between symmetry-related molecules in the unit cell involving Lys-72 or Thr-81 to suggest that they represent points of CD4-CD4 oligomerization. Thus, these two residues are neither sites of pMHC ligand contacts nor oligomer contacts. However, effects of those two residues may be secondary such that, for example, the Thr-81→Ala mutation impacts on a hydrogen bonding network involving Gln-94 near the CD4 D1–D2 junction.

Ternary Model of pMHCII, TCR, and CD4. To date, there has been no structural report on a ternary complex that contains a pMHCII, a TCR, and an intact CD4 molecule. Because a binary complex structure of the TCR–CA/I-A^k complex (17) and a structure of the entire extracellular four-domain fragment of hCD4 (CD4 D1-D4) (11) are available, a ternary complex model can be

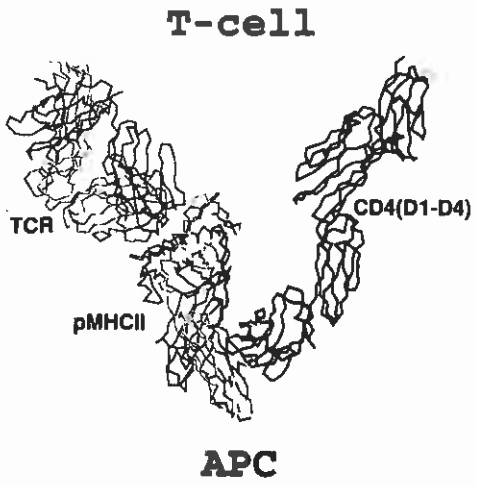


Fig. 3. Model of the ternary CD4–pMHCII–TCR complex. The assembly of several crystal structures reveals a ternary complex composed of the hCD4 D1-D4, pMHCII, and TCR having a V-shaped configuration. The D10 TCR–CA/I-A^k complex consists of the α (bright red) and β (yellow) subunits of the I-A^k molecule, the CA peptide (green), and the two TCR subunits α (dark red) and β (beige). The hCD4 D1-D4 is shown in blue, with the approximate position of the D4 glycan given as an orange ball. The assembly is oriented with the T-cell membrane on top and the antigen-presenting cell membrane on the bottom.

constructed in conjunction with the current binary complex structure of the hCD4 D1D2-CA/I-A^k. We first superimposed the I-A^k components of the CD4 D1D2-CA/I-A^k structure and our recently published structure of the D10 TCR–CA/I-A^k complex (17). The D10-CA/I-A^k structure contains only the V-module of the TCR. Because all TCRs share a common C module that is structurally similar in human and mouse, and the V modules of D10 and a class I-restricted human TCR B7 (32) can be superimposed onto each other extremely well (17), the C module of D10 TCR can thus be modeled from the B7 structure. We then applied the transform derived from the superposition of the CD4 D1D2 components of the current hCD4 D1D2-CA/I-A^k structure and the hCD4 D1-D4 structure (11) to incorporate the C-terminal D3D4 module of hCD4. The final ternary complex model (Fig. 3) gives a view of how an MHC class II-restricted TCR and an intact CD4 molecule derived from a T-helper cell might simultaneously bind to the same pMHCII on an opposing antigen-presenting cell.

From the perspective of the T-cell surface membrane (Fig. 3 *Upper*), both TCR and CD4 are tilted rather than oriented vertically. Assuming the insertion point of the CD4 ectodomain into lipid rafts where CD4 is a resident protein (33) and that of the non-raft resident TCR proteins are at equivalent vertical positions in the cell membrane, this "tilt" is likely because of the intrinsic features of the TCR complex components. The D1–D2 junction and D3–D4 junction of the CD4 molecule are quite rigid, with the D2–D3 junction itself having only limited flexibility (11). Thus, the total length of the concatenated four-domain CD4 rod is ≈120 Å, with virtually no capacity to extend further. The length of the binary TCR–pMHCII complex is around 130–140 Å, slightly longer than CD4. On the basis of prior analysis of TCR–pMHCII interactions (17, 34), it is likely that the variation of TCR docking topology onto pMHCII is limited, making the D10–CA/I-A^k complex a reasonable surrogate for all TCR–pMHCII interactions. The stalk regions connecting the membrane proximal domain and transmembrane-spanning segment of CD4, TCR Cα, and TCR Cβ are, respectively, 9, 19, and 13 amino acids in length. If the TCR–pMHCII binary complex oriented vertically to the cell membrane, then the CD4 molecule would not be of sufficient length to reach out from the T-cell membrane to maintain the V-shaped binding mode. Movement at the CD4 D2–D3 junction or alteration of the relative TCR and CD4 insertion points and intervening membrane may modify the picture to some degree. However, such variation is unlikely to change the overall binding topology of the components.

Several interesting features about the ternary complex model shown in Fig. 3 can be noted. There are no direct contacts between CD4 and the TCR αβ heterodimer, consistent with recent Biacore studies (16). In fact, the membrane-proximal ends of these two molecules are as far as 100 Å apart. On the basis of mAb epitope mapping and chain association studies (35, 36), CD3ε and CD3δ subunits may sit within or near this interval,

Table 1. CD4 D4 mutations are without effect on CD4–MHC class II interaction

Mutation	Residue location	CD4 level	Rosette formation
Arg-293→Ala	A	I	++
Lys-318→Ala[a]	C	II	+
Asn-321→Ala	CC′	I	++
Lys-322→Ala	CC′	I	++
Glu-323→Ala	CC′	I	++
Glu-320→Lys/Asn-321→Ala/Glu-323→Ala	CC′	I–II	+
Lys-328→Ala/Arg-329→Ala/Glu-330→Ala	C′,C′E	I	++
Lys-328→Glu/Arg-329→Glu/Glu-330→Gln[b]	C′,C′E	I–II	±
Lys-328→Glu/Arg-329→Glu[b]	C′,C′E	III	−
Lys-328→Glu/Glu-330→Gln[b]	C′,C′E	II	++
Arg-329→Glu/Glu-330→Gln[b]	C′E	I	++
Gln-344→Ala[a]	F	I	++
Gln-344→Ala/Glu-356→Lys[a]	F,G	I	++
Glu-356→Ala[a]	G	I	++

Mutations are designated by the name of the residue (three-letter amino acid code) and the position in the sequence followed by the name of the substituting residue. a, mutation of the residues observed in the study of Wu *et al.* (11) to be involved in CD4 dimer contacts in different crystals of hCD4 D1–D4. b, mutation of human CD4 residues to those of murine CD4. Relative levels of CD4 expression were defined by groups I–III, as follows: I, comparable to wild-type CD4; II, reduced to 31–60% of wild-type CD4; III, reduced to 10–30% of wild-type CD4. Rosette formation between CD4-expressing COS7 cells and MHC class II-expressing Raji B cells were measured and categorized as follows: ++, comparable to the rosette formation of wild-type CD4; +, reduced (31–60%) of wild-type CD4; ±, poor (10–30%) of wild-type CD4; −, <10% rosette formation. Results are representative of three independent experiments.

179

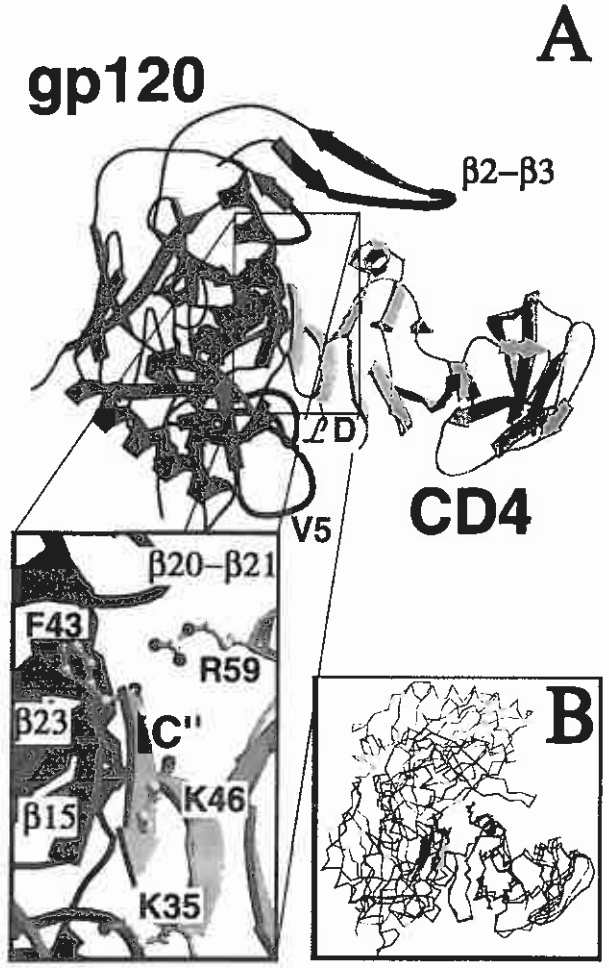


Fig. 4. Molecular mimicry of gp120 binding to CD4 relative to MHC class II. (*A*) The CD4 D1D2–gp120 complex (ref. 42, with Fab not included for clarity) is shown in similar orientation to the complex of CD4 D1D2/I-A^k in Fig. 1. The interactions that are in common are displayed in the box, which shows the Phe-43 inserted in a hydrophobic pocket involving gp120 β15 and β23, the formation of an antiparallel mini-β-sheet between the C" strand of CD4 and β15 from gp120 as well as interactions between CD4 Arg-59 and the loop connecting β20 to β21 in gp120. Besides these core interactions, gp120 has additional contacts through the V1/V2 loop connecting the β2 and β3 strands, the FD loop, and the V5 loop. (*B*) Overlay of CD4–CA/I-A^k and CD4–gp120 (42) complexes based on hCD4 D1 (blue and pink, respectively) with I-A^k in yellow and gp120 in red. Phe-43 and Arg-59 side chains are shown.

but the ample space makes direct CD4-CD3 contact unlikely. The antigen-binding groove of pMHCII is no longer parallel to the cell surface but makes about a 45° angle with the membrane. Despite the overall predicted V shape for the ternary complex, the membrane-proximal domains of each of the components are all roughly vertical, including domain 4 of CD4, the α2 and β2 domains of pMHCII, as well as the Cβ domain of TCR. The only exception is the unique Cα domain, which hangs almost parallel above the membrane with its relatively lengthy stalk region bridging the space. Human CD4 has one glycan attached to Asn-300, a residue located on the AB loop of domain 4 and pointing toward the membrane and away from TCR-pMHCII. This glycan may help preconfigure CD4 to assume a favorably tilted orientation for ligand binding in a role analogous to that observed in the structure of intercellular adhesion molecule-2 (37). Overall, this unexpected V-shaped binding mode raises intriguing questions as to how the cytoplasmic portions of the TCR subunits, including their immune receptor tyrosine activation motifs and the CD4-associated $p56^{lck}$, coordinate in signaling, the precise disposition of the CD3 components within the TCR machinery, and whether this ectodomain model represents a pre- and/or post-activation complex.

In vitro data have suggested that functional CD4-pMHCII binding may depend on CD4 oligomerization (29, 38). Dimerization of the CD4 D1-D4 via D4 in several crystal lattices has been interpreted as consistent with this view (11). Assuming this is the case, then the V shape of the ternary complex modeled herein could become a "W" shape by crosslinking of two adjacent V-shaped CD4–pMHC–TCR complexes. However, this is unlikely to occur, because the two protomers within the CD4 D1-D4 dimer have a very large opening angle. Consequently, the vertical distance from the apex of the CD4 D1-D4 dimer interface to the other end of each CD4 molecule is only ~80 Å. In comparison, excluding their respective stalks, the ternary complex spans 120 Å between the two opposing cell membranes. Thus, the different geometries make it difficult to imagine how CD4 dimers could create an oligomer array between the T cell and the antigen-presenting cell. In our current crystal structure, two CD4 D1D2 molecules are related by a crystal dyad in a parallel fashion. Both D1 and D2 are so closely juxtaposed to their symmetry mates that CD4 D3D4 cannot be modeled onto these CD4 D1D2 fragments without serious molecular collision. We have also examined molecular packing modes in published crystal structures that contain CD4 fragments (CD4 D1D2, CD4 D3D4, CD4 D1-D4), and pMHCII and relevant TCR molecules (9–11, 17, 39, 40). In fact, the CD4-binding site on the class II MHC molecule is nearly coincident with the "dimer of class II MHC heterodimers" interface previously proposed (40). From this analysis, we conclude that dimerization of the ternary complex is not likely to occur via the ectodomains of components of the complex.

Mutagenesis of CD4-D4. Nevertheless, to directly assess the role of potential CD4-D4 oligomerization in pMHCII binding, a series of CD4-D4 mutations was made and examined for their capacity to alter CD4-pMHCII-dependent cell conjugate formation. On the basis of the dimer interaction in the CD4 D1-D4 crystals, Glu-356, Gln-344, and Lys-318 residues of one CD4 molecule interact with Lys-318, Gln-344, and Glu-356 residues, respectively, of the pseudodyad related molecules (11). As shown in Table 1, mutations at these positions separately or collectively are without significant effect on CD4-pMHCII binding, excluding this CD4 dimerization mode as essential for physiologic pMHCII interaction. Likewise, a potential tetramer model involving the D4 CC' loop of one CD4 molecule and the C' strand of a neighboring protomer as necessary for pMHCII binding is implausible. Note that earlier human–mouse species shuffle mutations at CD4 residues 328–330, which demonstrated significant reduction in pMHCII binding (41), altered surface charge; alanine substitutions of these same residues are without functional effect. Although not shown, mutations on the BC and FG loops of CD4 D4 also provide no evidence for tetramerization via these D4 elements. From these results, the ternary model in Fig. 3, molecular packing within our CD4 D1D2-CA/I-A^k crystal and related structural data (9–11, 17, 39), physiologic CD4 oligomerization seems unlikely despite the indirect functional suggestions to the contrary. Perhaps the dominant negative effect of the Phe-43→Ile CD4 variant in blocking wild-type CD4–pMHCII interaction (38) is a consequence of local clustering of variant and wild-type CD4, which are both compartmentalized in lipid rafts, rather than because of a molecular oligomerization.

Comparison of CD4 Binding to pMHCII and HIV1 gp120. CD4 is the primary receptor for the HIV virus. The mature HIV gp160

envelope spike protein consists of the CD4-binding gp120 portion and the noncovalently associated gp41. CD4–gp120 interaction constitutes the first step of virus attachment, followed by cellular chemokine receptor binding (ref. 42 and refs. therein). The virus envelope gp160 spike protein then undergoes dramatic conformational changes, exposing its fusogenic gp41 portion, which coalesces viral and host cell membranes for viral genome entry. The binding of CD4-gp120 is orders of magnitude stronger than that of CD4-pMHCII (16, 43). Although some studies map the HIV- and pMHCII-binding sites to the same region of CD4, others have mapped the two ligands to opposite sides (12, 13, 44, 45). Having obtained the current CD4-CA/I-A^k structure, the direct three-dimensional–structural comparison with the previously published CD4-gp120-Fab structure (42) becomes possible.

Fig. 4*A* is a picture of the HIV gp120–CD4 complex with a view such that CD4 is in a similar orientation to CD4 in the Fig. 1 CD4–pMHCII complex. Strikingly, CD4 uses the same "top corner" for both pMHCII and gp120 binding. The key CD4 residue Phe-43 inserts into the junction of gp120's outer and inner domains as well as the bridging sheet (42). The other two CD4 D1-binding regions associated with pMHCII interaction, the mini-antiparallel β sheet involving CD4 segment Lys-46–Leu-44 on the C″ strand and the small DE helix, also play a very similar role in gp120 binding. For example, gp120 uses its β15 strand Gly-366–Asp-368 segment to pair with the CD4 C″ segment Lys-46–Leu-44 in an identical manner as the MHC class II molecule does to help position the CD4 Phe-43 residue (Fig. 4*A* and *Inset*). The Phe-43-binding environment seems more favorable in the CD4–pMHCII complex than in the CD4–gp120 complex, because the gp120-binding pocket contains a negatively charged glutamic acid and includes a large cavity. The gp120 also makes less favorable interactions with the CD4 DE helix. However, gp120 binding to CD4 is accompanied by significant conformational changes that may be energetically favorable (43) and involves a broader area than CD4-pMHCII binding. Details of the way in which the six segments of gp120 bind to CD4 have already been described (42). The buried surface area for CD4-pMHCII and CD4-gp120 is 850 Å^2 and 1,880 Å^2, respectively (using probe radius = 1.7 Å and density = 4). From Fig. 4, it is evident that the gp120 loop between β2 and β3 strands (V1V2), the £D loop, and the V5 loop may all contribute to greater binding. Because the current CD4-gp120-Fab structure comprises the core part of gp120, the intact molecule may make additional contacts with CD4 and/or further kinetically facilitate its binding. Superposition of CD4 D1D2-CA/I-A^k and CD4 D1D2-gp120 complexes on the basis of hCD4 D1 further emphasizes the common CD4-binding mode (Fig. 4*B*).

In conclusion, the crystal structure of the CD4-pMHCII presented here argues for a V-shaped CD4–pMHCII–TCR ternary complex as the basis of helper T-cell activation. It also demonstrates that gp120 binds to the identical structural elements of CD4 used by pMHCII, as well as to additional sites. Hence the virus has evolved to mimic the normal pMHCII interaction with CD4 but, in so doing, has augmented its interaction to bolster binding affinity. In this way, the viral protein usurps the normal CD4-pMHCII binding necessary for helper T-cell function (12), hence causing the human immunodeficiency. Whether some other viral envelope proteins morph themselves to mimic the normal ligands of cellular receptors as well remains to be seen.

1. Wang, J.-H. & Reinherz, E. L. (2000) *Curr. Opin. Struct. Biol.* **10**, 656–661.
2. Janeway, C. A., Jr. (1992) *Annu. Rev. Immunol.* **10**, 645–674.
3. Kruisbeek, A. M., Mond, J. J., Fowlkes, B. J., Carmen, J. A., Bridges, S. & Longo, D. L. (1985) *J. Exp. Med.* **161**, 1029–1047.
4. Marrack, P. & Kappler, J. (1986) *Adv. Immunol.* **38**, 1–30.
5. Gao, G. F., Tormo, J., Gerth, U. C., Wyer, J. R., McMichael, A. J., Stuart, D. I., Bell, J. I., Jones, E. Y. & Jakobsen, B. K. (1997) *Nature (London)* **387**, 630–634.
6. Garcia, K. C., Teyton, L. & Wilson, I. A. (1999) *Annu. Rev. Immunol.* **17**, 369–397.
7. Hennecke, J. & Wiley, D. C. (2001) *Cell* **104**, 1–4.
8. Kern, P., Teng, M.-K., Smolyar, A., Liu, J.-H., Liu, J., Hussey, R. E., Chang, H.-C., Reinherz, E. L. & Wang, J.-H. (1998) *Immunity* **9**, 519–530.
9. Ryu, S. E., Kwong, P. D., Truneh, A., Porter, T. G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N., Axel, R., Sweet, R. W. & Hendrickson, W. A. (1990) *Nature (London)* **348**, 419–426.
10. Wang, J.-H., Yan, Y., Garrett, T. P. J., Liu, J., Rodgers, D. W., Garlick, R. L., Tarr, G. E., Husain, Y., Reinherz, E. L. & Harrison, S. C. (1990) *Nature (London)* **348**, 411–418.
11. Wu, H., Kwong, P. D. & Hendrickson, W. A. (1997) *Nature (London)* **387**, 527–530.
12. Clayton, L. K., Sieh, M., Pious, D. A. & Reinherz, E. L. (1989) *Nature (London)* **339**, 548–551.
13. Moebius, U., Clayton, L. K., Abraham, S., Diener, A., Yunis, J. J., Harrison, S. C. & Reinherz, E. L. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 12008–12012.
14. Moebius, U., Pallai, P., Harrison, S. C. & Reinherz, E. L. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 8259–8263.
15. Xu, H. & Littman, D. R. (1993) *Cell* **74**, 633–643.
16. Xiong, Y., Kern, P., Chang, H.-C. & Reinherz, E. L. (2001) *J. Biol. Chem.* **276**, 5659–5667.
17. Reinherz, E. L., Tan, K., Tang, L., Kern, P., Liu, J.-H., Xiong, Y., Hussey, R. E., Smolyar, A., Hare, B., Zhang, R., *et al.* (1999) *Science* **286**, 1913–1921.
18. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276**, 307–326.
19. Navaza, J. (1994) *Acta Crystallogr. A* **50**, 157–163.
20. Fremont, D. H., Monnaie, D., Nelson, C. A., Hendrickson, W. A. & Unanue, E. R. (1998) *Immunity* **8**, 305–317.
21. Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999) *Acta Crystallogr. D* **55**, 484–491.
22. Cowtan, K. (1998) *Acta Crystallogr. D* **54**, 750–756.
23. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999) *Acta Crystallogr. D* **55**, 247–255.
24. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54**, 905–921.
25. Killeen, N., Sawada, S. & Littman, D. R. (1993) *EMBO J.* **12**, 1547–1553.
26. Law, Y. M., Yeung, R. S. M., Mamalaki, C., Kioussis, D., Mak, T. W. & Havell, R. A. (1994) *J. Exp. Med.* **179**, 1233–1242.
27. von Hoegen, P., Miceli, M. C., Tourvieille, B., Schilham, M. & Parnes, J. R. (1989) *J. Exp. Med.* **170**, 1879–1886.
28. König, R., Huang, L.-Y. & Germain, R. N. (1992) *Nature (London)* **356**, 796–798.
29. König, R., Shen, X. & Germain, R. N. (1995) *J. Exp. Med.* **182**, 779–787.
30. Vignali, D. A. A., Moreno, J., Schiller, D. & Hämmerling, G. J. (1992) *J. Exp. Med.* **175**, 925–932.
31. Sakihama, T., Hunsicker, M. E., Hussey, R. E. & Reinherz, E. L. (2000) *Eur. J. Immunol.* **30**, 279–290.
32. Ding, Y. H., Baker, B. M., Garboczi, D. N., Biddison, W. E. & Wiley, D. C. (1999) *Immunity* **11**, 45–56.
33. Xavier, R., Brennan, T., Li, Q., McCormack, C. & Seed, B. (1998) *Immunity* **8**, 723–732.
34. Hennecke, J., Carfi, A. & Wiley, D. C. (2000) *EMBO J.* **19**, 5611–5624.
35. Ghendler, Y., Smolyar, A., Chang, H.-C. & Reinherz, E. L. (1998) *J. Exp. Med.* **187**, 1529–1536.
36. Manolios, N., Kemp, O. & Li, Z. G. (1994) *Eur. J. Immunol.* **24**, 84–92.
37. Casasnovas, J. M., Springer, T. A., Liu, J.-H., Harrison, S. C. & Wang, J.-H. (1997) *Nature (London)* **387**, 312–315.
38. Sakihama, T., Smolyar, A. & Reinherz, E. L. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 6444–6448.
39. Brady, R. L., Dodson, E. J., Dodson, G. G., Lange, G., Davis, S. J., Williams, A. F. & Barclay, A. N. (1993) *Science* **260**, 979–983.
40. Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L. & Wiley, D. C. (1993) *Nature (London)* **364**, 33–39.
41. Boulgatte, R., Scarmato, P., el Marhomy, S., Martin, M., Ostankovitch, M., Lafosse, S., Vervisch, A., Auffray, C. & Platier-Tonneau, D. (1994) *J. Immunol.* **152**, 4475–4488.
42. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1998) *Nature (London)* **393**, 648–659.
43. Myszka, D. G., Sweet, R. W., Hensley, P., Brigham-Burke, M., Kwong, P. D., Hendrickson, W. A., Wyatt, R., Sodroski, J. & Doyle, M. L. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 9026–9031.
44. Fleury, S., Lamarre, D., Meloche, S., Ryu, S. E., Cantin, C., Hendrickson, W. A. & Sekaly, P. (1991) *Cell* **66**, 1037–1049.
45. Platier-Tonneau, D., Gastinel, L.-N., Moussy, G., Benichou, B., Amblard, F., Vaigot, P. & Auffray, C. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6858–6862.
46. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950.
47. Merritt, E. A. & Bacon, D. J. (1997) *Methods Enzymol.* **277**, 505–524.
48. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res.* **24**, 4876–4882.

181

$CD8^+$ T lymphocytes

As discussed above, $CD8^+$ T lymphocytes will selectively bind to antigen presenting cells (APC's) with surface MHC class I-processed peptide complexes. After this specific interaction, CD28 on surfaces of T cells will then interact with CD80 or CD86 on APC's to further stabilize the cell-cell association. These T lymphocytes are also known as cytotoxic T cells. After activation, they secrete perforin, granzymes, and granulysin and start to express CD95L on their surfaces. Both perforin and granulysin can form pores on the cell attached to the cytotoxic T cell, and allow granzymes, a group of serine proteases, to enter the target cell to induce the activities of cysteine proteases. These proteases can digest all contents in the target cell cytoplasm and thus induce apoptosis. In addition, the surface expressed CD95L will bind CD95 also to induce apoptosis.

In addition to cytotoxic T cells, there are two other types of cells which can induce apoptosis in target cells. They are:

Natural killer cells, or NK cells, are cells with no membrane bound TCR molecules, CD3 molecules, nor immunoglobulins. They are $CD16^+$ and $CD56^+$, and also $CD8^+$ in 80 % of their population. Like cytotoxic T cells, they secrete perforin and granzymes to kill target cells. They can be activated by several cytokines such as IL-12, IL-15, IL-18, IL-2, CCL5, etc. They also have Fc receptors on their cell surfaces, similar to macrophages, and can thus be activated by antibodies binding to cell surface molecules on tumor or foreign cells. They play an important role in the innate immune system.

Natural killer T cells, or NKT cells, have properties of both cytotoxic T cells and natural killer cells. They recognize CD1d. Such cells are very rare in the human system.

Therefore $CD8^+$ T lymphocytes and related cells play essential roles in the elimination of foreign invaders into the human system, as well as destruction of tumor cells. Induction of these cells by vaccination is one the most important aspects of immunotherapy for cancer treatment.

$CD4^+$ T lymphocytes

$CD4^+$ T lymphocytes are also known as helper T cells. After the binding of CD4 with MHC class II-processed peptide complex on antigen presenting cells (APC's), additional interaction between these two types of cells involve the binding of CD28 on helper T cells to CD80 or CD86 of APC's, similar to the situation with cytotoxic T cells binding to APC. Helper T cells also secrete many different macro-molecules with different biological functions. Primarily, these cytokines will stimulate the target cells to divide, instead of killing them. They also stimulate themselves to divide.

Unlike the cytotoxic T cells, helper T cells have many different types. They include T_h0, T_h1, T_h2, T_h3, T_h17, etc. as summarized by Wikipedia on the next several pages. Basically, they secrete different cytokines and stimulate different cells to divide. Many more new types are being identified recently, such as T_h9. T_{reg} is $CD4^+$, $CD25^+$, $Foxp3^+$, so that may also be considered as helper T cells. Regulatory T cells are also known as suppressor T cells some of which are $CD8^+$, $CD26^-$. Recently, four major types of regulatory T cells have been identified.

Roughly, T_h1 cells are involved in the cellular immune system while T_h2 cells in the humoral immune system. Other types are more complicated. The cytokines produced by these two different types of helper T cells are:

T_h1: interferon-γ

Tumor necrosis factor-β

IL-2 (also by all helper T cells during early stages of their activation)

IL-10 (induced in activated cells)

T_h2: IL-4

IL-5

IL-6

IL-10

IL-13

T helper cell

From Wikipedia, the free encyclopedia

T helper cells (also known as **T_h cells**) are a sub-group of lymphocytes, a type of white blood cell, that play an important role in establishing and maximizing the capabilities of the immune system. These cells are unusual in that they have no cytotoxic or phagocytic activity; they cannot kill infected host cells or pathogens. Rather, these cells provide signals for orchestrating the immune response. T_h cells are involved in activating and directing other immune cells and are particularly important in the immune system. They are essential in determining B cell antibody class switching, in the activation and growth of cytotoxic T cells, and in maximizing bactericidal activity of phagocytes such as macrophages. It is this diversity in function and their role in influencing other cells that gives T helper cells their name.

Mature T_h cells are believed to always express the surface protein CD4 and are referred to as **$CD4^+$ T cells**. $CD4^+$ T cells are generally treated as having a pre-defined role as helper T cells within the immune system, although there are known rare exceptions. For example, there are sub-groups of regulatory T cells, natural killer T cells, and cytotoxic T cells that are known to express CD4 (although cytotoxic examples have been observed in extremely low numbers in specific disease states, they are usually considered non-existent). All of the latter $CD4^+$ T cell groups are not considered T helper cells.

The importance of helper T cells can be seen from HIV, a virus that infects cells that are $CD4^+$ (including helper T cells). Towards the end of an HIV infection the number of functional $CD4^+$ T cells falls, which leads to the symptomatic stage of infection known as the acquired immunodeficiency syndrome (AIDS). There are also some rare disorders that result in the absence or dysfunction of $CD4^+$ T cells. These disorders produce similar symptoms, and many of these are fatal.

Contents

- 1 Activation of naïve helper T cells
  - 1.1 Recognition (Signal 1)
  - 1.2 Verification (Signal 2)
  - 1.3 Proliferation
  - 1.4 Maturation
- 2 Determination of the effector T cell response
  - 2.1 T_h1/T_h2 Model for helper T cells
  - 2.2 Limitations to the T_h1/T_h2 model
- 3 Role of helper T cells in disease
  - 3.1 Helper T cells and hypersensitivity
  - 3.2 HIV infection
- 4 See also
- 5 References
- 6 External links

Activation of naïve helper T cells

Following T cell development, matured, naïve (meaning they have never been exposed to the antigen to which they can respond) T cells leave the thymus and begin to spread throughout the body, including the lymph nodes. Like all T cells, they express the T cell receptor-CD3 complex. The T cell receptor (TcR) consists of both constant and variable regions, the latter of which determines what antigen the T cell can respond to. $CD4^+$ T cells have TcRs with an affinity for Class II MHC, and it is believed that CD4 is involved in determining MHC affinity during maturation in the thymus. Class II MHC proteins are generally only found on the surface of specialised antigen-presenting cells (APCs). Specialised antigen presenting cells are primarily dendritic cells, macrophages and B cells, although B cells are the only cell group that expresses MHC Class II constitutively (at all times). Some APCs also bind native (or unprocessed) antigens to their surface, such as follicular dendritic cells, but unprocessed antigens do not interact with T cells and are not involved in their activation. The antigens that bind to MHC proteins are always short peptides, 8-10 amino acids long for MHC Class I, and up to 25 or so for MHC Class II.

Recognition (Signal 1)

During an immune response, professional antigen-presenting cells (APCs) endocytose (absorb) foreign material (typically bacteria or viruses), which undergoes processing, then travels from the infection site to the lymph nodes. Once at the lymph nodes, the APC begins to present antigen peptides that are bound to Class II MHC, allowing $CD4^+$ T cells that express the specific TCRs against the peptide/MHC complex to activate.

When a T_h cell encounters and recognises the antigen, on an APC, the TCR-CD3 complex binds strongly to the peptide-MHC complex present on the surface of professional APC's. CD4, a co-receptor of the TCR complex, also binds to a different section of the MHC molecule. These interactions bring these proteins closer together, allowing the intracellular kinases present on the TCR, CD3 and CD4 proteins to activate each other via phosphorylation. With the assistance of a phosphatase present on the intracellular section of CD45 (common leukocyte antigen), these molecules activate the major biochemical pathways in the cytosol of the T_h cell. These active pathways are known as Signal 1 of T cell activation, as it is the first and primary pro-activation signal in a T_h cell. Upon subsequent encounters with a given antigen, memory T cells are re-activated using the same TCR pathways.

The binding of the antigen-MHC to the TCR complex and CD4 may also help the APC and the T_h cell adhere during T_h cell activation, but the integrin protein LFA-1 on the T cell and ICAM on the APC are the primary molecules of adhesion in this cell interaction.

It is unknown what role the relatively bulky extracellular region of CD45 plays during cell interactions, but CD45 has various isoforms that change in size depending on the T_h cell's activation and maturation status. For example, CD45 shortens in length following T_h activation ($CD45RA^+$ to $CD45RO^+$), but whether this change in length influences activation is unknown. It has been proposed that the larger CD45RA+ may decrease the accessibility of the T cell receptor for the antigen-MHC molecule, thereby necessitating an increase in the affinity (and specificity) of the T cell for activation. Once the activation has occurred however, CD45 shortens, allowing easier interactions and activation as an effector T helper cell.[*citation needed*]

185

Verification (Signal 2)

Having received the first TcR/CD3 signal, the naïve T cell must activate a second independent biochemical pathway, known as Signal 2. This verification step is a protective measure to ensure that a T cell is responding to a foreign antigen. If this second signal is not present during initial antigen exposure, the T cell presumes that it is auto-reactive. This results in the cell becoming anergic (anergy is generated from the unprotected biochemical changes of Signal 1). Anergic cells will not respond to any antigen in the future, even if both signals are present later on. These cells are generally believed to circulate throughout the body with no value until they apoptose at the end of their lifespan.

The second signal involves an interaction between CD28 on the $CD4^+$ T cell and the proteins CD80 (B7.1) or CD86 (B7.2) on the professional APCs. Both CD80 and CD86 activate the CD28 receptor. These proteins are also known as co-stimulatory molecules.

Although the verification stage is necessary for the activation of naïve helper T cells, the importance of this stage is best demonstrated during the similar activation mechanism of $CD8^+$ cytotoxic T cells. As naïve $CD8^+$ T cells have no true bias towards foreign sources, these T cells must rely on the activation of CD28 for confirmation that they recognise a foreign antigen (as CD80/CD86 is only expressed by active APC's). CD28 plays an important role in decreasing the risk of T cell auto-immunity against host antigens.

Once the naïve T cell has both pathways activated, the biochemical changes induced by Signal 1 are altered, allowing the cell to activate instead of anergise. The second signal is then obsolete; only the first signal is necessary for future activation. This is also true for memory T cells, which is one example of learned immunity. Faster responses occur upon reinfection because memory T cells have already undergone confirmation and can produce effector cells much sooner.

Proliferation

Once the two-signal activation is complete the T helper cell (T_h) then allows itself to proliferate. It achieves this by releasing a potent T cell growth factor called interleukin 2 (IL-2) which acts upon itself in an autocrine fashion. Activated T cells also produce the alpha sub-unit of the IL-2 receptor (CD25 or IL-2R), enabling a fully functional receptor that can bind with IL-2, which in turn activates the T cell's proliferation pathways.

The autocrine or paracrine secretion of IL-2 can bind that same T_h cell or neighboring T_h's via the IL-2R thus driving proliferation and clonal expansion. The T_h cells receiving both signals of activation will then become T_h0 cells (T helper 0) cell that secrete IL-2, IL-4 and interferon gamma (IFN-γ). The T_h0 cells will then differentiate into T_h1 or T_h2 cells depending on cytokine environment. IFN-γ drives T_h1 cell production while IL-10 and IL-4 inhibit T_h1 cell production. Conversely, IL-4 drives T_h2 cell production and IFN-γ inhibits T_h2 cells. It should be noted that these cytokines are pleiotropic and carry out many other functions of the immune response.

Maturation

After many cell generations, the T_h cell's progenitors differentiate into **effector T_h** cells, **memory T_h** cells,

and **regulatory T_h cells**.

- Effector T_h cells secrete cytokines, proteins or peptides that stimulate or interact with other leukocytes, including T_h cells.
- Memory T_h cells retain the antigen affinity of the originally activated T cell, and are used to act as later effector cells during a second immune response (e.g. if there is re-infection of the host at a later stage).
- Regulatory T cells do not promote immune function, but act to decrease it instead. Despite their low numbers during an infection, these cells are believed to play an important role in the self-limitation of the immune system; they have been shown to prevent the development of various auto-immune diseases.

The production of IL-2 by helper T cells is also necessary for the proliferation of activated $CD8^+$ T cells. Without helper T cell interactions, $CD8^+$ T cells do not develop effector function and eventually undergo apoptosis. This cross-reliance on helper T cells is another way the immune system works to prevent T cell-mediated auto-immune disease.

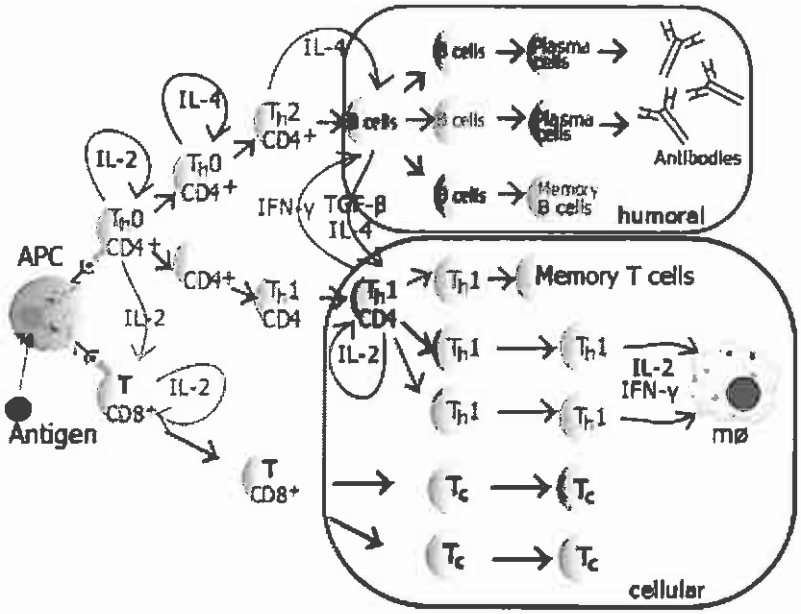
Determination of the effector T cell response

Helper T cells are capable of influencing a variety of immune cells, and the T cell response generated (including the extracellular signals such as cytokines) can be essential for a successful outcome from infection. In order to be effective, helper T cells must determine which cytokines will allow the immune system to be most useful or beneficial for the host. Understanding exactly how helper T cells respond to immune challenges is currently of major interest in immunology, because such knowledge may be very useful in the treatment of disease and in increasing the effectiveness of vaccination.

T_h1/T_h2 Model for helper T cells

Proliferating helper T cells that develop into effector T cells differentiate into two major subtypes of cells known as T_h1 and T_h2 cells (also known as Type 1 and Type 2 helper T cells, respectively).

187



T_h1/T_h2 Model for helper T cells. An antigen is ingested and processed by an APC. It presents fragments from it to T cells. The upper, T_h0, is a T helper cell. The fragment is presented to it by MHC2.[1] IFN-γ, interferon γ; TGF-β, transforming growth factor β; mø, macrophage; IL-2, interleukin 2; IL-4, interleukin 4

	Type 1/ T_h1	**Type 2/ T_h2**
Main partner cell type	Macrophage	B-cell
Cytokines produced	interferon-γ and tumor necrosis factor-beta. (Interleukin-2 was classically associated with T_h1 cells, but this association may be misleading; IL-2 is produced by all helper T cells early in their activation.) interleukin-10 production has been shown to be induced in activated Th1 cell[2]	interleukin-4, interleukin-5, interleukin-6, interleukin-10, interleukin-13
Immune stimulation promoted	Cellular immune system. Maximizes the killing efficacy of the macrophages and the proliferation of cytotoxic $CD8^+$ T cells. Also produces opsonizing antibodies	Humoral immune system. Stimulates B-cells into proliferation, to induce B-cell antibody class switching, and to increase neutralizing antibody production.
Other functions	The Type 1 cytokine IFN-γ increases the production of interleukin-12 by	The Type 2 response promotes its own profile using two different cytokines. Interleukin-4 acts

dendritic cells and macrophages, and via positive feedback, IL-12 stimulates the production of IFN-γ in helper T cells, thereby promoting the T_h1 profile. IFN-gamma also inhibits the production of cytokines such as interleukin-4, an important cytokine associated with the Type 2 response, and thus it also acts to preserve its own response.

on helper T cells to promote the production of T_h2 cytokines (including itself; it is auto-regulatory), while interleukin-10 (IL-10) inhibits a variety of cytokines including interleukin-2 and IFN-γ in helper T cells and IL-12 in dendritic cells and macrophages. The combined action of these two cytokines suggests that once the T cell has decided to produce these cytokines, that decision is preserved (and also encourages other T cells to do the same).

While we know about the types of cytokine patterns helper T cells tend to produce, we understand less about how the patterns themselves are decided. Various evidence suggests that the type of APC presenting the antigen to the T cell has a major influence on its profile. Other evidence suggests that the concentration of antigen presented to the T cell during primary activation influences its choice. The presence of some cytokines (such as the ones mentioned above) will also influence the response that will eventually be generated, but our understanding is nowhere near complete.

Limitations to the T_h1/T_h2 model

The interactions between cytokines from the T_h1/T_h2 model can be more complicated in some animals. For example, the T_h2 cytokine IL-10 inhibits cytokine production of both T_h subsets in humans. Human IL-10 (hIL-10) suppresses the proliferation and cytokine production of all T cells and the activity of macrophages, but continues to stimulate plasma cells, ensuring that antibody production still occurs. As such, hIL-10 is not believed to truly promote the T_h2 response in humans, but acts to prevent over-stimulation of helper T cells while still maximising the production of antibodies.

There are also other types of T cells that can influence the expression and activation of helper T cells, such as natural regulatory T cells, along with less common cytokine profiles such as the **T_h3 subset** of helper T cells. Terms such as "regulatory" and "suppression" have become ambiguous after the discovery that helper $CD4^+$ T cells are also capable of regulating (and suppressing) their own responses outside of dedicated regulatory T cells.

One major difference between regulatory T cells and effector T cells is that regulatory T cells typically serve to modulate and deactivate the immune response, while effector T cell groups usually begin with immune-promoting cytokines and then switch to inhibitory cytokines later in their life cycle. The latter is a feature of T_h3 cells, which transform into a regulatory subset after its initial activation and cytokine production.

Both regulatory T cells and T_h3 cells produce the cytokine transforming growth factor-beta (TGF-β) and IL-10. Both cytokines are inhibitory to helper T cells; TGF-β suppresses the activity of most of the immune system. There is evidence to suggest that TGF-β may not suppress activated Th2 cells as effectively as it might suppress naive cells, but it is not typically considered a Th2 cytokine.

189

The characterisation of another novel T helper subtype, T helper 17 cells (T_h17)[3] has cast further doubt on the basic T_h1/T_h2 model. These IL-17 producing cells were initially described as a pathogenic population implicated in autoimmunity but are now thought to have their own distinct effector and regulatory functions.

Many of the cytokines in this article are also expressed by other immune cells (see individual cytokines for details), and it is becoming clear that while the original T_h1/T_h2 model is enlightening and gives insight into the functions of helper T cells, it is far too simple to define its entire role or actions. Some immunologists question the model completely, as some *in vivo* studies suggest that individual helper T cells usually do not match the specific cytokine profiles of the T_h model, and many cells express cytokines from both profiles.[*citation needed*] That said, the T_h model has still played an important part in developing our understanding of the roles and behaviour of helper T cells and the cytokines they produce during an immune response.

Recent scientific studies by Stockinger et al. revealed that another T helper subset may exist. Th9 cells are claimed to be a IL9 (interleukin 9) producing T cell subset focused on defending helminth infections.[4]

Role of helper T cells in disease

Considering the diverse and important role helper T cells play in the immune system, it is not surprising that these cells often influence the immune response against disease. They also appear to make occasional mistakes, or generate responses that would be politely considered non-beneficial. In the worst case scenario, the helper T cell response could lead to a disaster and the fatality of the host. Fortunately this is a very rare occurrence.

Helper T cells and hypersensitivity

The immune system must achieve a balance of sensitivity in order to respond to foreign antigens without responding to the antigens of the host itself. When the immune system responds to very low levels of antigen that it usually shouldn't respond to, a hypersensitivity response occurs. Hypersensitivity is believed to be the cause of allergy and some auto-immune disease.

Hypersensitivity reactions can be divided into four types:

- Type 1 hypersensitivity includes common immune disorders such as asthma, allergic rhinitis (hay fever), eczema, urticaria (hives) and anaphylaxis. These reactions all involve IgE antibodies, which require a T_h2 response during helper T cell development. Preventative treatments, such as corticosteroids and montelukast, focus on suppressing mast cells or other allergic cells; T cells do not play a primary role during the actual inflammatory response. It's important to note that the numeral allocation of hypersensitivity "types" does not correlate (and is completely unrelated) to the "response" in the T_h model.

- Type 2 and Type 3 hypersensitivity both involve complications from auto-immune or low affinity antibodies. In both of these reactions, T cells may play an accomplice role in generating these auto-specific antibodies, although some of these reactions under Type 2 hypersensitivity would

be considered normal in a healthy immune system (for example, Rhesus factor reactions during child-birth is a normal immune response against child antigens). The understanding of the role of helper T cells in these responses is limited but it is generally thought that T_h2 cytokines would promote such disorders. For example, studies have suggested that lupus (SLE) and other auto-immune diseases of similar nature can be linked to the production of T_h2 cytokines.

- Type 4 hypersensitivity, also known as delayed type hypersensitivity, are caused via the over-stimulation of immune cells, commonly lymphocytes and macrophages, resulting in chronic inflammation and cytokine release. Antibodies do not play a direct role in this allergy type. T cells play an important role in this hypersensitivity, as they activate against the stimulus itself and promote the activation of other cells; particularly macrophages via T_h1 cytokines.

Other cellular hypersensitivities include cytotoxic T cell mediated auto-immune disease, and a similar phenomenon; transplant rejection. Helper T cells are required to fuel the development of these diseases. In order to create sufficient auto-reactive killer T cells, interleukin-2 must be produced, and this is supplied by $CD4^+$ T cells. $CD4^+$ T cells can also stimulate cells such as natural killer cells and macrophages via cytokines such as interferon-gamma, encouraging these cytotoxic cells to kill host cells in certain circumstances.

The mechanism that killer T cells use during auto-immunity is almost identical to their response against viruses, and some viruses have been accused of causing auto-immune diseases such as Type 1 Diabetes mellitus. Cellular auto-immune disease occurs because the host antigen recognition systems fail, and the immune system believes, by mistake, that a host antigen is foreign. As a result, the $CD8^+$ T cells treat the host cell presenting that antigen as infected, and go on to destroy all host cells (or in the case of transplant rejection, transplant organ) that express that antigen.

Some of this section is a simplification. Many auto-immune diseases are more complex. A well known example is rheumatoid arthritis, where both antibodies and immune cells are known to play a role in the pathology. Generally the immunology of most auto-immune diseases is not well understood.

HIV infection

Perhaps the best example of the importance of $CD4^+$ T cells is demonstrated with human immunodeficiency virus (HIV) infection. HIV targets cells that express CD4, and can infect macrophages, dendritic cells (both groups express CD4 at low levels) and $CD4^+$ T cells.

It has been proposed that during the non-symptomatic phase of HIV infection, the virus has a relatively low affinity towards T cells (and has a higher affinity for macrophages), resulting in a slow kill rate of $CD4^+$ T cells by the immune system. This is initially compensated for via the production of new helper T cells from the thymus (originally from the bone marrow). Once the virus becomes lymphotropic (or T-tropic) however, it begins to infect $CD4^+$ T cells far more efficiently (likely due to a change in the co-receptors it binds to during infection), and the immune system is overwhelmed.

At this point, functional $CD4^+$ T cell levels begin to decrease, eventually to a point where the $CD4^+$ T cell population is too small to recognize the full range of antigens that could potentially be detected. The lack of full antigen cover results in the core symptoms of acquired immune deficiency syndrome (AIDS). CD4 T cell depletion during AIDS allows various pathogens to escape T cell recognition, thus allowing

191

opportunistic infections that would normally elicit a helper T cell response to bypass the immune system. While these complete bypass situations only occur when the helper T cell response is absolutely necessary for infection clearance, most infections increase in severity and/or duration because the immune system's helper T cells provide a weaker contribution to a less efficient immune response.

Two components of the immune system are particularly affected in AIDS, due to its $CD4^+$ T cell dependency:

1. $CD8^+$ T cells are not stimulated as effectively during the AIDS stage of HIV infection, making AIDS patients very susceptible to most viruses, including HIV itself. This decline in killing of $CD4^+$ T cells results in the virus being produced for a longer period (the infected $CD4^+$ T cells are not killed as quickly), increasing the proliferation of the virus, and accelerating the development of the disease.
2. Antibody class switching declines significantly once helper T cell function fails. The immune system loses its ability to improve the affinity of their antibodies, and are unable to generate B cells that can produce antibody groups such as IgG and IgA. These effects are primarily due to the loss of any helper T cell that can interact with the B lymphocyte correctly. Another symptom of AIDS is the reduction in antibody levels due to a decrease in T_h2 cytokines (and less interactions by helper T cells). All of these complications result in an increased susceptibility to aggressive bacterial infections, especially in areas of the body not accessible by IgM antibodies.

If the patient does not respond to (or does not receive) HIV treatment they will succumb usually to either cancers or infections; the immune system finally reaches a point where it is no longer coordinated or stimulated enough to deal with the disease.

Said et al. showed that inhibition of CD4 T-cell expansion during HIV infection is also due to microbial translocation in an IL-10-dependent way. Indeed, triggering PD-1, expressed on monocytes and up-regulated upon monocytes activation, by its ligand PD-L1 induces IL-10 production which inhibits CD4 T-cell function.[5]

See also

- Cd4+ T cells and antitumor Immunity

References

1. ^ Rang, H. P. (2003). *Pharmacology*. Edinburgh: Churchill Livingstone. ISBN 0-443-07145-4. Page 223
2. ^ Interleukin-10 Production by Th1 Cells Requires Interleukin-12-Induced STAT4 Transcription Factor and ERK MAP Kinase Activation by High Antigen Dose
3. ^ Harrington, LE; Hatton, RD; Mangan, PR; others (2005), "Interleukin 17-producing CD4+ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages", *Nature Immunology* **6** (11): 1023–32, PMID 16200070
4. ^ (Veldhoen and Stockinger, Dardalhon and Kuchroo both papers from Nature Immunology 2008)
5. ^ Elias A. Said et al. 2009, PD-1 Induced IL10 Production by Monocytes Impairs T-cell Activation in a Reversible Fashion. Nature Medicine. 2010; 452-9.

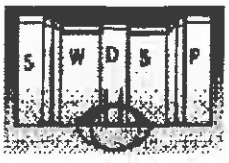
External links

- T-cell Group - Cardiff University (http://www.tcells.org/scientific/helper/)

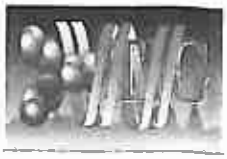
Retrieved from "http://en.wikipedia.org/wiki/T_helper_cell"
Categories: T cells | Human cells

- This page was last modified on 5 July 2011 at 18:05.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. See Terms of use for details.
  Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

193



Different Types of Regulatory T Cells

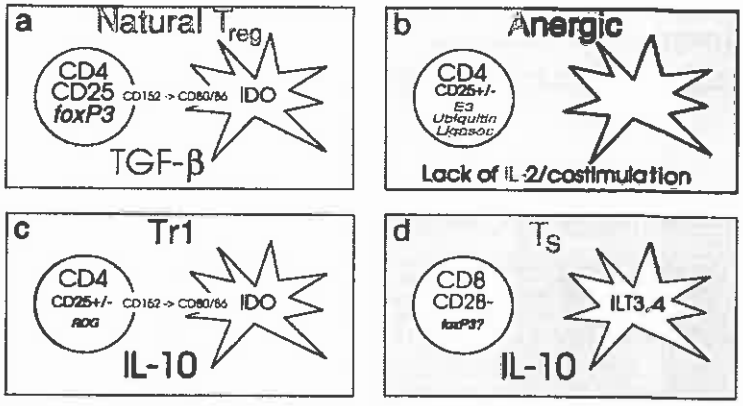


Search within Therapeutic Immunology Group Web Site:

Find: Enter text

Google powered by Google Search

- TIG/TAC Index
- Tolerance Lab
- Cytokine Lab
- Dendritic Lab
- FACS Machine Info
- Other TIG Links
- Therapeutic Antibody Centre
- Dunn School of Pathology



Four of the major regulatory T cell subsets that have been described are depicted in the left of each panel (a-d) together with their appropriate antigen presenting cells on the right. (a) The "natural" regulatory T cells originally recognised by their constitutive expression of CD4 and CD25 are further defined by expression of the transcription factor foxP3 and surface CD152. Theirgeneration and some of their suppressive activity is dependent on TGF-beta, and it has been shown that they can induced IDO in appropriate DCs by CD152 mediated ligation of CD80/86. (b) Anergic CD4+ T cells generated by antigen stimulation in the absence of costimulation seem to be characterised by an intrinsic raising of their threshold for antigen stimulation, that may be maintained by expression of E3 ubiquitin ligases such as GRAIL, c-cbl and Itch. Anergic cells can act as regulatory T cells by competing at the sites of antigen presentation and adsorbing out stimulatory cytokines such as IL-2. (c) Tr1 cells represent an induced subset of CD4 helper T cells that are dependent on IL-10 for their differentiation and for some of their regulatory properties. They do not express foxP3 but may express markers associated with Th2 cells and repressor of GATA (ROG). Like natural Tregs, they express high levels of surface CD152 and can induce IDO and trypophan catabolism in appropriate DCs. (d) CD8+CD28- suppressor T (Ts) cells were first characterised in human, but have recently also been demonstrated in rodents. Like Tr1 cells, they are induced in the presence of IL-10, and IL-10 may be involved in the downregulation of dendritic cell costimulation and the upregulation of ILT-3 and ILT-4 (in human DC) that seem to play an important role in presenting antigen to tolerise further cohorts of T cells.

IDO — indoleamine 2,3-dioxygenase

ROG — repressor of GATA

"Developing short-term treatments for long-term outcomes"

Related topics:

- Antibodies in therapy
- Making mAbs in large (gram) amounts
- TIG Monoclonal Antibodies
- A Guide to Tolerance

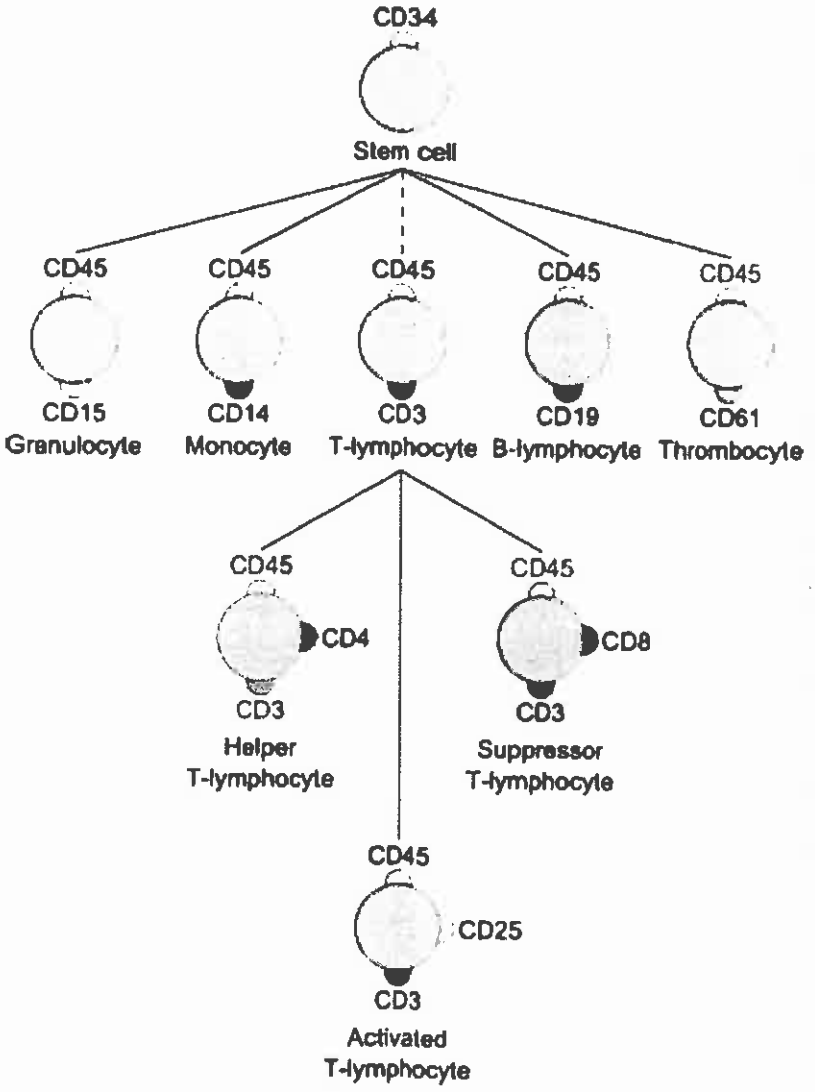
Last Updated 28th May 2004 by Steve Cobbold

Other CD molecules

CD is the abbreviation for cluster of differentiation or cluster of designation. They include a large set of more than three hundred different cell surface molecules. Many are being discovered continuously. For complete listing, please go to the website listed on the next page.

We have mentioned about CD3 associated with TCR, CD4 of helper T cells, CD8 of cytotoxic T cells, etc. Their expressions are not plus or minus, but can be of various degrees denoted by high or low. Most of them are shared by different cell types. Some are relatively specific. For example, most cells of the B lymphocyte lineage express CD20. Therapeutic antibodies are thus designed against CD20 to treat B cell lymphomas. An example is illustrated below:

Type of cell	CD markers
stem cells	CD34+, CD31-
all leukocyte groups	CD45+
Granulocyte	CD45+, CD15+
Monocyte	CD45+, CD14+
T lymphocyte	CD45+, CD3+
T helper cell	CD45+,CD3+, CD4+
Cytotoxic T cell	CD45+,CD3+, CD8+
B lymphocyte	CD45+, CD19+ or CD45+, CD20+
Thrombocyte	CD45+, CD61+
Natural killer cell	CD16+, CD56+,CD3-



CD differentiation

Click to make ImmunologyLink your homepage

The Immunology Link

Search

HOME

CD Antigens Assignment

CD Antigens
Graduate Programs
Search Jobs
Literature Searches
FDA Approved Antibodies
Antibodies in Clinical Trials
Online Journals
Fellowships/Grants
Find Antibodies
Search Books
Find Products
Associations
Cancer Info
Gene Databases
Protein Databases
KO/Transgenic Mice
Biotechnology Info
Find a Scientist
Clinical Trials
Search Patents
Meeting/Conferences
Immunology Nobel Prizes
Immunology Videos
The Human Protein Atlas

Human CD Antigens:

- Table for CD1-CD100
- Table for CD101-CD200
- Table for CD201-CD300
- Table for CD301-CD339

Quick Links
Biotech News
Pharma News
Search Wikipedia
Immune Tolerance Network
Stock Quotes
Immunology News
Web MD
Immunology Books
Madame Curie Biosci Database

Search Catalogs
Jackson Labs
Linscott's Directory
ATCC
Kinase Inhibitors
Randox Life Sciences
R&D Systems
Antibody Directory
Abnova
eBioScience
LifeSpan BioSciences
BD Biosciences
Imgenex
Abcam
BioLegend
Invitrogen

Copyright © 1997-2010 The Immunology Link. All rights reserved. Revised: January 14, 2011 | CONTACT US

Interleukins

Interleukins were originally used to designate cytokines secreted by leukocytes. More recent studies suggest that they are secreted by other cells as well. They are the major players in the immune systems regulating cell-cell interactions. Several of these interleukins are secreted by T_h2 cells with some of their functions summarized below:

IL-4: One of its major functions is to stimulate the proliferation of T_h2 cells thus providing a positive feedback loop for these cells to increase in number. It also stimulates the differentiation of naïve helper T cells, T_h0 cells, into T_h2 cells. IL-4 is also secreted by mast cells, basophils and eosinophils. Basophils may be responsible for the initial differentiation of T_h0 cells into T_h2 cells. IL-4 has many other biological activities.

IL-5: This interleukin is a homo-dimer. It stimulates B lymphocytes to grow and increases the secretion of antibodies. However, its major function is to stimulate eosinophils to proliferate, due to the large amounts of IL-5 receptors on these cells. It was initially isolated as an eosinophil differentiation factor.

IL-6: It is also known as B-stimulatory factor-2, playing an essential role in the terminal differentiation of B lymphocytes into antibody secreting cells. IL-6 is required for hybridoma growth. In addition, it has many other functions on various cells. It can cross the blood-brain barrier and induce fever. IL-6 can also be secreted by macrophages, smooth muscle cells, adipocytes, etc.

IL-10: It stimulates the maturation of B lymphocytes and antibody production. On the other hand, it down-regulates the cytokines secreted by T_h1 cells, such as interferon-γ, IL-2, etc. It also inhibits the activity of macrophages. IL-10 is also expressed in monocytes, mast cells, T_{reg} cells, activated B lymphocytes, etc.

IL-13: This interleukin is similar to IL-4. The major difference is that IL-13 provides an environment hostile to parasites. It also inhibits inflammatory cytokine production.

An overview of the functions of other interleukins can be found in Wikipedia as enclosed. Of particular interest is the recent discovery of the activity of IL-17 in relation to autoimmunity (Hemdan NY, Birkenmeier G, Wichmann G, Abu El-Saad AM, Krieger T, Conrad K, Sack U (2010) Interleukin-17-producing T helper cells in auto immunity. Autoimmunity Rev, 9, 785-792). For this reason, a special helper T lymphocyte, T_h17, has been identified.

Interleukin

From Wikipedia, the free encyclopedia

Interleukins are a group of cytokines (secreted proteins/signaling molecules) that were first seen to be expressed by white blood cells (leukocytes). The term *interleukin* derives from (*inter-*) "as a means of communication", and (*-leukin*) "deriving from the fact that many of these proteins are produced by leukocytes and act on leukocytes". The name is something of a relic though (the term was coined by Dr. Paetkau, University of Victoria); it has since been found that interleukins are produced by a wide variety of body cells. The function of the immune system depends in a large part on *interleukins*, and rare deficiencies of a number of them have been described, all featuring autoimmune diseases or immune deficiency. The majority of interleukins are synthesized by helper CD4+ T lymphocytes, as well as through monocytes, macrophages, and endothelial cells. They promote the development and differentiation of T, B, and hematopoietic cells.

Contents

- 1 Common families of interleukins
  - 1.1 Interleukin-1
  - 1.2 Interleukin-2
  - 1.3 Interleukin-3
  - 1.4 Interleukin-5
  - 1.5 Interleukin-6
  - 1.6 Interleukins 7 and 9
  - 1.7 Interleukin-10
  - 1.8 Interleukin-11
  - 1.9 Interleukin 12
  - 1.10 Interleukin-13
  - 1.11 Interleukin-15
  - 1.12 Interleukin-17
- 2 List of human interleukins
- 3 References
- 4 External links

Common families of interleukins

Interleukin-1

Interleukin

Interleukin-1 alpha and interleukin-1 beta (IL-1 alpha and IL-1 beta) are cytokines that participate in the regulation of immune responses, inflammatory reactions, and hematopoiesis.[1] Two types of IL-1 receptor, each with three extracellular immunoglobulin (Ig)-like domains, limited sequence similarity (28%) and different

pharmacological characteristics have been cloned from mouse and human cell lines: these have been termed type I and type II receptors[2] The receptors both exist in transmembrane (TM) and soluble forms: the soluble IL-1 receptor is thought to be post-translationally derived from cleavage of the extracellular portion of the membrane receptors.

Both IL-1 receptors (CD121a/IL1R1, CD121b/IL1R2) appear to be well conserved in evolution, and map to the same chromosomal location.[3] The receptors can both bind all three forms of IL-1 (IL-1 alpha, IL-1 beta and IL-1RA).

The crystal structures of IL1A and IL1B[4] have been solved, showing them to share the same 12-stranded beta-sheet structure as both the heparin binding growth factors and the Kunitz-type soybean trypsin inhibitors.[5] The beta-sheets are arranged in 3 similar lobes around a central axis, 6 strands forming an anti-parallel beta-barrel. Several regions, especially the loop between strands 4 and 5, have been implicated in receptor binding.

Molecular cloning of the Interleukin 1 Beta converting enzyme is generated by the proteolytic cleavage of an inactive precursor molecule. A complementary DNA encoding a protease that carries out this cleavage has been cloned. Recombinant expression enabled cells to process precursor Interleukin 1 Beta to the mature form of the enzyme.

Interleukin-2

T-Lymphocytes regulate the growth and differentiation of certain lymphopoietic and haemopoietic cells through the release of various secreted protein factors.[6] These factors, which include interleukin-2 (IL2), are secreted by lectin- or antigen-stimulated T-cells, and have various physiological effects. IL2 is a lymphokine that induces the proliferation of responsive T-cells. In addition, it acts on some B-cells,

Identifiers

Symbol ?

Other data

Interleukin-1

Identifiers

Symbol	Interleukin_1
Pfam	PF00340 (http://pfam.sanger.ac.uk /family?acc=PF00340)
InterPro	IPR000975 (http://www.ebi.ac.uk/interpro /DisplayIproEntry?ac=IPR000975)
SMART	SM00125 (http://smart.embl-heidelberg.de/smart /do_annotation.pl?DOMAIN=SM00125)
PROSITE	PDOC00226 (http://www.expasy.org/cgi-bin/prosite-search-ac?PDOC00226)

Available protein structures:

Pfam	structures (http://pfam.sanger.ac.uk/family /PF00340#tabview=tab9)
PDB	RCSB PDB (http://www.rcsb.org/pdb/search /smartSubquery.do?smartSearchSubtype=PfamIdQuery& pfamID=PF00340) ; PDBe (http://www.ebi.ac.uk/pdbe-srv/view/search?PFAM_domain=PF00340& id_type=PFAM_domain&id_value=PF00340& search_type=advanced)
PDBsum	structure summary (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum /GetPfamStr.pl?pfam_id=PF00340)

Interleukin-2

Identifiers

Symbol	Interleukin-2
Pfam	PF00715 (http://pfam.sanger.ac.uk /family?acc=PF00715)

199

via receptor-specific binding,[7] as a growth factor and antibody production stimulant.[8] The protein is secreted as a single glycosylated polypeptide, and cleavage of a signal sequence is required for its activity.[7] Solution NMR suggests that the structure of IL2 comprises a bundle of 4 helices (termed A-D), flanked by 2 shorter helices and several poorly-defined loops. Residues in helix A, and in the loop region between helices A and B, are important for receptor binding. Secondary structure analysis has suggested similarity to IL4 and granulocyte-macrophage colony stimulating factor (GMCSF).[8]

Interleukin-3

Interleukin-3 (IL3) is a cytokine that regulates blood-cell production by controlling the production, differentiation and function of granulocytes and macrophages.[9][10] The protein, which exists in vivo as a monomer, is produced in activated T-cells and mast cells,[9][10] and is activated by the cleavage of an N-terminal signal sequence.[10]

IL3 is produced by T-lymphocytes and T-lymphomas only after stimulation with antigens, mitogens, or chemical activators such as phorbol esters. However, IL3 is constitutively expressed in the myelomonocytic leukaemia cell line WEHI-3B.[10] It is thought that the genetic change of the cell line to constitutive production of IL3 is the key event in development of this leukaemia.[10]

Interleukin-5

Interleukin-5 (IL5), also known as eosinophil differentiation factor (EDF), is a lineage-specific cytokine for eosinophilpoiesis.[11][12] It regulates

InterPro	IPR000779 (http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR000779)
SMART	SM00189 (http://smart.embl-heidelberg.de/smart/do_annotation.pl?DOMAIN=SM00189)
PROSITE	PDOC00349 (http://www.expasy.org/cgi-bin/prosite-search-ac?PDOC00349)

Available protein structures:

Pfam	structures (http://pfam.sanger.ac.uk/family/PF00715#tabview=tab9)
PDB	RCSB PDB (http://www.rcsb.org/pdb/search/smartSubquery.do?smartSearchSubtype=PfamIdQuery&pfamID=PF00715) ; PDBe (http://www.ebi.ac.uk/pdbe-srv/view/search?PFAM_domain=PF00715&id_type=PFAM_domain&id_value=PF00715&search_type=advanced)
PDBsum	structure summary (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPfamStr.pl?pfam_id=PF00715)

Interleukin-3

Identifiers

Symbol	Interleukin_3
Pfam	PF02059 (http://pfam.sanger.ac.uk/family?acc=PF02059)
InterPro	IPR002183 (http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR002183)

Available protein structures:

Pfam	structures (http://pfam.sanger.ac.uk/family/PF02059#tabview=tab9)
PDB	RCSB PDB (http://www.rcsb.org/pdb/search/smartSubquery.do?smartSearchSubtype=PfamIdQuery&pfamID=PF02059) ; PDBe (http://www.ebi.ac.uk/pdbe-srv/view/search?PFAM_domain=PF02059&id_type=PFAM_domain&id_value=PF02059&search_type=advanced)
PDBsum	structure summary (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPfamStr.pl?pfam_id=PF02059)

Interleukin-5

eosinophil growth and activation,[11] and thus plays an important role in diseases associated with increased levels of eosinophils, including asthma.[12] IL5 has a similar overall fold to other cytokines (e.g., IL2, IL4 and GCSF),[12] but while these exist as monomeric structures, IL5 is a homodimer. The fold contains an anti-parallel 4-alpha-helix bundle with a left handed twist, connected by a 2-stranded anti-parallel beta-sheet.[12][13] The monomers are held together by 2 interchain disulphide bonds.[13]

Interleukin-6

Interleukin-6 (IL6), also referred to as B-cell stimulatory factor-2 (BSF-2) and interferon beta-2, is a cytokine involved in a wide variety of biological functions.[14] It plays an essential role in the final differentiation of B-cells into IG-secreting cells, as well as inducing myeloma/plasmacytoma growth, nerve cell differentiation, and, in hepatocytes, acute-phase reactants.[14][15]

A number of other cytokines may be grouped with IL6 on the basis of sequence similarity.[14][15][16] These include granulocyte colony-stimulating factor (GCSF) and myelomonocytic growth factor (MGF). GCSF acts in hematopoiesis by affecting the production, differentiation, and function of 2 related white cell groups in the blood.[16] MGF also acts in hematopoiesis, stimulating proliferation and colony formation of normal and transformed avian cells of the myeloid lineage.

Cytokines of the IL6/GCSF/MGF family are glycoproteins of about 170 to 180 amino acid residues that contains four conserved cysteine residues involved in two disulphide bonds:.[16] They have a compact, globular fold (similar to other

Identifiers

Symbol	Interleukin_5
Pfam	PF02025 (http://pfam.sanger.ac.uk/family?acc=PF02025)
InterPro	IPR000186 (http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR000186)

Available protein structures:

Pfam	structures (http://pfam.sanger.ac.uk/family/PF02025#tabview=tab9)
PDB	RCSB PDB (http://www.rcsb.org/pdb/search/smartSubquery.do?smartSearchSubtype=PfamIdQuery&pfamID=PF02025) ; PDBe (http://www.ebi.ac.uk/pdbe-srv/view/search?PFAM_domain=PF02025&id_type=PFAM_domain&id_value=PF02025&search_type=advanced)
PDBsum	structure summary (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPfamStr.pl?pfam_id=PF02025)

Interleukin-6/G-CSF/MGF

Identifiers

Symbol	IL6_MGF_GCSF
Pfam	PF00489 (http://pfam.sanger.ac.uk/family?acc=PF00489)
InterPro	IPR003573 (http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR003573)
SMART	SM00126 (http://smart.embl-heidelberg.de/smart/do_annotation.pl?DOMAIN=SM00126)
PROSITE	PDOC00227 (http://www.expasy.org/cgi-bin/prosite-search-ac?PDOC00227)

Available protein structures:

Pfam	structures (http://pfam.sanger.ac.uk/family/PF00489#tabview=tab9)
PDB	RCSB PDB (http://www.rcsb.org/pdb/search/smartSubquery.do?smartSearchSubtype=PfamIdQuery&pfamID=PF00489) ; PDBe (http://www.ebi.ac.uk/pdbe-srv/view/search?PFAM_domain=PF00489&id_type=PFAM_domain&id_value=PF00489&search_type=advanced)

20/

interleukins), stabilised by the 2 disulphide bonds. One half of the structure is dominated by a 4-alpha-helix bundle with a left-handed twist;[17] the helices are anti-parallel, with 2 overhand connections, which fall into a 2-stranded anti-parallel beta-sheet. The fourth alpha-helix is important to the biological activity of the molecule.[15]

PDBsum structure summary (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPfamStr.pl?pfam_id=PF00489)

Interleukins 7 and 9

Interleukin-7 (IL-7)[18] is a cytokine that serves as a growth factor for early lymphoid cells of both B- and T-cell lineages. Interleukin-9 (IL-9)[19] is a cytokine that supports IL-2 independent and IL-4 independent growth of helper T-cells. Interleukin-7 and -9 seems to be evolutionary related.[20]

Interleukin-7/interleukin-9

Identifiers

Symbol Interleukin_7_9

Pfam PF01415 (http://pfam.sanger.ac.uk/family?acc=PF01415)

InterPro IPR000226 (http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR000226)

PROSITE PDOC00228 (http://www.expasy.org/cgi-bin/prosite-search-ac?PDOC00228)

Available protein structures:

Pfam structures (http://pfam.sanger.ac.uk/family/PF01415#tabview=tab9)

PDB RCSB PDB (http://www.rcsb.org/pdb/search/smartSubquery.do?smartSearchSubtype=PfamIdQuery&pfamID=PF01415) ; PDBe (http://www.ebi.ac.uk/pdbe-srv/view/search?PFAM_domain=PF01415&id_type=PFAM_domain&id_value=PF01415&search_type=advanced)

PDBsum structure summary (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPfamStr.pl?pfam_id=PF01415)

Interleukin-10

Interleukin-10 (IL-10) is a protein that inhibits the synthesis of a number of cytokines, including IFN-gamma, IL-2, IL-3, TNF, and GM-CSF produced by activated macrophages and by helper T cells. In structure, IL-10 is a protein of about 160 amino acids that contains four conserved cysteines involved in disulphide bonds.[21] IL-10 is highly similar to the *Human herpesvirus 4* (Epstein-Barr virus) BCRF1 protein, which inhibits the synthesis of gamma-interferon and to *Equid herpesvirus 2* (Equine herpesvirus 2) protein E7. It is also similar, but to a lesser degree, with human protein mda-7.[22] a protein that has antiproliferative properties in human melanoma cells. Mda-7 contains only two of the four cysteines of IL-10.

Interleukin-10

Identifiers

Symbol Interleukin_10

Pfam PF00726 (http://pfam.sanger.ac.uk/family?acc=PF00726)

InterPro IPR000098 (http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR000098)

Interleukin-11

Interleukin-11 (IL-11) is a secreted protein that stimulates megakaryocytopoiesis, resulting in increased production of

platelets, as well as activating osteoclasts, inhibiting epithelial cell proliferation and apoptosis, and inhibiting macrophage mediator production. These functions may be particularly important in mediating the hematopoietic, osseous and mucosal protective effects of interleukin 11.[23] Family members seem to be restricted to mammals.

Interleukin 12

Interleukin-12 (IL-12) is a disulphide-bonded heterodimer consisting of a 35kDa alpha subunit and a 40kDa beta subunit. It is involved in the stimulation and maintenance of Th1 cellular immune responses, including the normal host defence against various intracellular pathogens, such as Leishmania, Toxoplasma, *Measles virus*, and *Human immunodeficiency virus 1* (HIV). IL-12 also has an important role in pathological Th1 responses, such as in inflammatory bowel disease and multiple sclerosis. Suppression of IL-12 activity in such diseases may have therapeutic benefit. On the other hand, administration of recombinant IL-12 may have therapeutic benefit in conditions associated with pathological Th2 responses.[24][25]

Interleukin-13

Interleukin-13 (IL-13) is a pleiotropic cytokine that may be important in the regulation of the inflammatory and immune responses.[26] It inhibits inflammatory cytokine production and synergises with IL-2 in regulating interferon-gamma synthesis. The sequences of IL-4 and IL-13 are distantly related.

Interleukin-15

Interleukin-15 (IL-15) is a cytokine that

SMART SM00188 (http://smart.embl-heidelberg.de/smart/do_annotation.pl?DOMAIN=SM00188)

PROSITE PDOC00450 (http://www.expasy.org/cgi-bin/prosite-search-ac?PDOC00450)

Available protein structures:

Pfam structures (http://pfam.sanger.ac.uk/family/PF00726#tabview=tab9)

PDB RCSB PDB (http://www.rcsb.org/pdb/search/smartSubquery.do?smartSearchSubtype=PfamIdQuery&pfamID=PF00726) ; PDBe (http://www.ebi.ac.uk/pdbe-srv/view/search?PFAM_domain=PF00726&id_type=PFAM_domain&id_value=PF00726&search_type=advanced)

PDBsum structure summary (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPfamStr.pl?pfam_id=PF00726)

Interleukin 11

Identifiers

Symbol IL11

Pfam PF07400 (http://pfam.sanger.ac.uk/family?acc=PF07400)

InterPro IPR010873 (http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR010873)

Available protein structures:

Pfam structures (http://pfam.sanger.ac.uk/family/PF07400#tabview=tab9)

PDB RCSB PDB (http://www.rcsb.org/pdb/search/smartSubquery.do?smartSearchSubtype=PfamIdQuery&pfamID=PF07400) ; PDBe (http://www.ebi.ac.uk/pdbe-srv/view/search?PFAM_domain=PF07400&id_type=PFAM_domain&id_value=PF07400&search_type=advanced)

PDBsum structure summary (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPfamStr.pl?pfam_id=PF07400)

Interleukin-12 alpha subunit

Identifiers

203

possesses a variety of biological functions, including stimulation and maintenance of cellular immune responses.[27] IL-15 stimulates the proliferation of T-lymphocytes, which requires interaction of IL-15 with components of IL-2R, including IL-2R beta and probably IL-2R gamma, but not IL-2R alpha.

Interleukin-17

Interleukin-17 (IL-17) is a potent proinflammatory cytokine produced by activated memory T cells.[28] The IL-17 family is thought to represent a distinct signalling system that appears to have been highly conserved across vertebrate evolution.[28]

List of human interleukins

A list of *interleukins*:

Symbol IL12

Pfam PF03039 (http://pfam.sanger.ac.uk/family?acc=PF03039)

InterPro IPR004281 (http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR004281)

Available protein structures:

Pfam structures (http://pfam.sanger.ac.uk/family/PF03039#tabview=tab9)

PDB RCSB PDB (http://www.rcsb.org/pdb/search/smartSubquery.do?smartSearchSubtype=PfamIdQuery&pfamID=PF03039) ; PDBe (http://www.ebi.ac.uk/pdbe-srv/view/search?PFAM_domain=PF03039&id_type=PFAM_domain&id_value=PF03039&search_type=advanced)

PDBsum structure summary (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPfamStr.pl?pfam_id=PF03039)

Interleukin-13

Identifiers

Symbol Interleukin_13

Pfam PF03487 (http://pfam.sanger.ac.uk/family?acc=PF03487)

InterPro IPR003634 (http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR003634)

Available protein structures:

Pfam structures (http://pfam.sanger.ac.uk/family/PF03487#tabview=tab9)

PDB RCSB PDB (http://www.rcsb.org/pdb/search/smartSubquery.do?smartSearchSubtype=PfamIdQuery&pfamID=PF03487) ; PDBe (http://www.ebi.ac.uk/pdbe-srv/view/search?PFAM_domain=PF03487&id_type=PFAM_domain&id_value=PF03487&search_type=advanced)

PDBsum structure summary (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPfamStr.pl?pfam_id=PF03487)

Interleukin 15

Identifiers

Symbol Interleukin_15

Pfam PF02372 (http://pfam.sanger.ac.uk/family?acc=PF02372)

InterPro IPR003443 (http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR003443)

Available protein structures:

Pfam structures (http://pfam.sanger.ac.uk/family/PF02372#tabview=tab9)

PDB RCSB PDB (http://www.rcsb.org/pdb/search/smartSubquery.do?smartSearchSubtype=PfamIdQuery&pfamID=PF02372) ; PDBe (http://www.ebi.ac.uk/pdbe-srv/view/search?PFAM_domain=PF02372&id_type=PFAM_domain&id_value=PF02372&search_type=advanced)

PDBsum structure summary (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPfamStr.pl?pfam_id=PF02372)

Interleukin-17

Identifiers

Symbol IL17

Pfam PF06083 (http://pfam.sanger.ac.uk/family?acc=PF06083)

InterPro IPR010345 (http://www.ebi.ac.uk/interpro/DisplayIproEntry?ac=IPR010345)

Available protein structures:

Pfam structures (http://pfam.sanger.ac.uk/family/PF06083#tabview=tab9)

PDB RCSB PDB (http://www.rcsb.org/pdb/search/smartSubquery.do?smartSearchSubtype=PfamIdQuery&pfamID=PF06083) ; PDBe (http://www.ebi.ac.uk/pdbe-srv/view/search?PFAM_domain=PF06083&id_type=PFAM_domain&id_value=PF06083&search_type=advanced)

PDBsum structure summary (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPfamStr.pl?pfam_id=PF06083)

Name	Source [29]	Target receptors[29][30]	Target cells[29]	Function[29]

205

IL-1	macrophages, B cells, monocytes,[31] dendritic cells [31]	CD121a/IL1R1, CD121b/IL1R2	T helper cells	co-stimulation [31]
			B cells	maturation & proliferation [31]
			NK cells	activation[31]
			macrophages, endothelium, other	inflammation,[31] small amounts induce acute phase reaction, large amounts induce fever
IL-2	Th1-cells	CD25/IL2RA, CD122/IL2RB, CD132/IL2RG	activated[31] T cells and B cells, NK cells, macrophages, oligodendrocytes	stimulates growth and differentiation of T cell response. Can be used in immunotherapy to treat cancer or suppressed for transplant patients. Has also been used in clinical trials (ESPIRIT. Stalwart) to raise CD4 counts in HIV positive patients.
IL-3	activated T helper cells,[31] mast cells, NK cells, endothelium, eosinophils	CD123/IL3RA, CD131/IL3RB	hematopoietic stem cells	differentiation and proliferation of myeloid progenitor cells [31] to e.g. erythrocytes, granulocytes
			mast cells	growth and histamine release[31]
IL-4	Th2 cells, just activated naive CD4+ cell, memory CD4+ cells, mast cells, macrophages	CD124/IL4R, CD132/IL2RG	activated B cells	proliferation and differentiation, IgG1 and IgE synthesis.[31] Important role in allergic response (IgE)
			T cells	proliferation[31]
			endothelium	
IL-5	Th2 cells, mast cells, eosinophils	CD125/IL5RA, CD131/IL3RB	eosinophils	production
			B cells	differentiation, IgA production
			activated B cells	differentiation into plasma cells
			plasma cells	antibody secretion
IL-6	macrophages, Th2 cells, B cells, astrocytes, endothelium	CD126/IL6RA, CD130/IR6RB	hematopoietic stem cells	differentiation
			T cells, others	induces acute phase reaction, hematopoiesis, differentiation, inflammation

IL-7	Bone marrow stromal cells and thymus stromal cells	CD127/IL7RA, CD132/IL2RG	pre/pro-B cell, pre/pro-T cell, NK cells	differentiation and proliferation of lymphoid progenitor cells, involved in B, T, and NK cell survival, development, and homeostasis, ↑proinflammatory cytokines
IL-8	macrophages, lymphocytes, epithelial cells, endothelial cells	CXCR1/IL8RA, CXCR2/IL8RB /CD128	neutrophils, basophils, lymphocytes	Neutrophil chemotaxis
IL-9	Th2 cells, specifically by CD4+ helper cells	CD129/IL9R	T cells, B cells	Potentiates IgM, IgG, IgE, stimulates mast cells
IL-10	monocytes, Th2 cells, CD8+ T cells, mast cells, macrophages, B cell subset	CD210/IL10RA, CDW210B/IL10RB	macrophages	cytokine production[31]
			B cells	activation [31]
			mast cells	
			Th1 cells	inhibits Th1 cytokine production (IFN-γ, TNF-β, IL-2)
			Th2 cells	Stimulation
IL-11	bone marrow stroma	IL11RA	bone marrow stroma	acute phase protein production, osteoclast formation
IL-12	dendritic cells, B cells, T cells, macrophages	CD212/IL12RB1, IR12RB2	activated [31] T cells,	differentiation into Cytotoxic T cells with IL-2,[31] ↑ IFN-γ, TNF-α, ↓ IL-10
			NK cells	↑ IFN-γ, TNF-α
IL-13	activated Th2 cells, mast cells, NK cells	IL13R	TH2-cells, B cells, macrophages	Stimulates growth and differentiation of B cells (IgE), inhibits TH1-cells and the production of macrophage inflammatory cytokines (e.g. IL-1, IL-6), ↓ IL-8, IL-10, IL-12
IL-14	T cells and certain malignant B cells		activated B cells	controls the growth and proliferation of B cells, inhibits Ig secretion
IL-15	mononuclear phagocytes (and some other cells),	IL15RA	T cells, activated B cells	Induces production of Natural killer cells

207

	especially macrophages following infection by virus(es)			
IL-16	lymphocytes, epithelial cells, eosinophils, CD8+ T cells	CD4	CD4+ T cells (Th-cells)	CD4+ chemoattractant
IL-17	T helper 17 cells (Th17)	CDw217/IL17RA, IL17RB	epithelium, endothelium, other	osteoclastogenesis, angiogenesis, ↑ inflammatory cytokines
IL-18	macrophages	CDw218a/IL18R1	Th1 cells, NK cells	Induces production of IFNγ, ↑ NK cell activity
IL-19	-	IL20R		-
IL-20	-	IL20R		regulates proliferation and differentiation of keratinocytes
IL-21	activated T helper cells, NKT cells	IL21R	All lymphocytes, dendritic cells	costimulates activation and proliferation of CD8+ T cells, augment NK cytotoxicity, augments CD40-driven B cell proliferation, differentiation and isotype switching, promotes differentiation of Th17 cells
IL-22	-	IL22R		Activates STAT1 and STAT3 and increases production of acute phase proteins such as serum amyloid A, Alpha 1-antichymotrypsin and haptoglobin in hepatoma cell lines
IL-23	-	IL23R		Increases angiogenesis but reduces CD8 T-cell infiltration
IL-24	-	IL20R		Plays important roles in tumor suppression, wound healing and psoriasis by influencing cell survival.
IL-25	-	LY6E		Induces the production IL-4, IL-5 and IL-13, which stimulate eosinophil expansion

IL-26 -	IL20R1	Enhances secretion of IL-10 and IL-8 and cell surface expression of CD54 on epithelial cells
IL-27 -	IL27RA	Regulates the activity of B lymphocyte and T lymphocytes
IL-28 -	IL28R	Plays a role in immune defense against viruses
IL-29 -		Plays a role in host defenses against microbes
IL-30 -		Forms one chain of IL-27
IL-31 -	IL31RA	May play a role in inflammation of the skin
IL-32 -		Induces monocytes and macrophages to secrete TNF-α, IL-8 and CXCL2
IL-33 -		Induces helper T cells to produce type 2 cytokine
IL-35 regulatory T cells		Suppression of T helper cell activation

References

1. ^ MacDonald HR, Sims JE, March CJ, Cosman D, Widmer MB, Grubin CE, Wignall JM, Jackson JL, Call SM, McMahan CJ (1988). "cDNA expression cloning of the IL-1 receptor, a member of the immunoglobulin superfamily". *Science* **241** (4865): 585–589. doi:10.1126/science.2969618 (http://dx.doi.org/10.1126%2Fscience.2969618) . PMID 2969618 (http://www.ncbi.nlm.nih.gov/pubmed/2969618) .
2. ^ Maki RA, Clevenger W, De Souza EB, Liu C, Hart RP, Liu XJ (1996). "Cloning and characterization of an alternatively processed human type II interleukin-1 receptor mRNA". *J. Biol. Chem.* **271** (34): 20965–20972. doi:10.1074/jbc.271.34.20965 (http://dx.doi.org/10.1074%2Fjbc.271.34.20965) . PMID 8702856 (http://www.ncbi.nlm.nih.gov/pubmed/8702856) .
3. ^ Jenkins NA, Mosley B, Cosman D, Slack JL, Grubin CE, Wignall JM, Lupton SD, Brunton LL, Brannan CI, McMahan CJ (1991). "A novel IL-1 receptor, cloned from B cells by mammalian expression, is expressed in many cell types". *EMBO J.* **10** (10): 2821–2832. PMC 452992 (http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=452992) . PMID 1833184 (http://www.ncbi.nlm.nih.gov/pubmed/1833184) .
4. ^ Priestle JP, Grutter MG, Schar HP (1989). "Crystallographic refinement of interleukin 1 beta at 2.0 A resolution". *Proc. Natl. Acad. Sci. U.S.A.* **86** (24): 9667–9671. doi:10.1073/pnas.86.24.9667 (http://dx.doi.org/10.1073%2Fpnas.86.24.9667) . PMC 298562 (http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=298562) . PMID 2602367 (http://www.ncbi.nlm.nih.gov/pubmed/2602367) .
5. ^ Murzin AG, Chothia C, Lesk AM (1992). "beta-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 beta and 1 alpha and fibroblast growth factors". *J. Mol. Biol.* **223** (2): 531–543. doi:10.1016/0022-2836(92)90668-A (http://dx.doi.org/10.1016%2F0022-2836%2892%2990668-A) .

209

PMID 1738162 (http://www.ncbi.nlm.nih.gov/pubmed/1738162) .

6. ^ Arai K, Yokota T, Arai N, Lee F, Rennick D, Mosmann T (1985). "Use of a cDNA expression vector for isolation of mouse interleukin 2 cDNA clones: expression of T-cell growth-factor activity after transfection of monkey cells". *Proc. Natl. Acad. Sci. U.S.A.* **82** (1): 68–72. doi:10.1073/pnas.82.1.68 (http://dx.doi.org/10.1073%2Fpnas.82.1.68) . PMC 396972 (http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=396972) . PMID 3918306 (http://www.ncbi.nlm.nih.gov/pubmed/3918306) .
7. ^ *a b* Gillis S, Cerretti DP, McKereghan K, Larsen A, Cantrell MA, Anderson D, Cosman D, Baker PE (1986). "Cloning, sequence, and expression of bovine interleukin 2". *Proc. Natl. Acad. Sci. U.S.A.* **83** (10): 3223–3227. doi:10.1073/pnas.83.10.3223 (http://dx.doi.org/10.1073%2Fpnas.83.10.3223) . PMC 323485 (http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=323485) . PMID 3517854 (http://www.ncbi.nlm.nih.gov/pubmed/3517854) .
8. ^ *a b* Driscoll PC, Campbell ID, Mott HR, Boyd J, Cooke RM, Weir MP (1992). "Secondary structure of human interleukin 2 from 3D heteronuclear NMR experiments". *Biochemistry* **31** (33): 7741–7744. doi:10.1021/bi00148a040 (http://dx.doi.org/10.1021%2Fbi00148a040) . PMID 1510960 (http://www.ncbi.nlm.nih.gov/pubmed/1510960) .
9. ^ *a b* Dorssers L, Burger H, Bot F, Delwel R, Lowenberg B, Wagemaker G, Geurts van Kessel AH (1987). "Characterization of a human multilineage-colony-stimulating factor cDNA clone identified by a conserved noncoding sequence in mouse interleukin-3". *Gene* **55** (1): 115–124. doi:10.1016/0378-1119(87)90254-X (http://dx.doi.org/10.1016%2F0378-1119%2887%2990254-X) . PMID 3497843 (http://www.ncbi.nlm.nih.gov/pubmed/3497843) .
10. ^ *a b c d e* Ymer S, Tucker WQ, Sanderson CJ, Campbell HD, Young IG, Hapel AJ (1985). "Constitutive synthesis of interleukin-3 by leukaemia cell line WEHI-3B is due to retroviral insertion near the gene". *Nature* **317** (6034): 255–258. doi:10.1038/317255a0 (http://dx.doi.org/10.1038%2F317255a0) . PMID 2413359 (http://www.ncbi.nlm.nih.gov/pubmed/2413359) .
11. ^ *a b* Tucker WQ, Sanderson CJ, Campbell HD, Young IG, Hort Y, Martinson ME, Mayo G, Clutterbuck EJ (1987). "Molecular cloning, nucleotide sequence, and expression of the gene encoding human eosinophil differentiation factor (interleukin 5)". *Proc. Natl. Acad. Sci. U.S.A.* **84** (19): 6629–6633. doi:10.1073/pnas.84.19.6629 (http://dx.doi.org/10.1073%2Fpnas.84.19.6629) . PMC 299136 (http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=299136) . PMID 3498940 (http://www.ncbi.nlm.nih.gov/pubmed/3498940) .
12. ^ *a b c d* Proudfoot AE, Wells TN, Milburn MV, Hassell AM, Lambert MH, Jordan SR, Graber P (1993). "A novel dimer configuration revealed by the crystal structure at 2.4 A resolution of human interleukin-5". *Nature* **363** (6425): 172–176. doi:10.1038/363172a0 (http://dx.doi.org/10.1038%2F363172a0) . PMID 8483502 (http://www.ncbi.nlm.nih.gov/pubmed/8483502) .
13. ^ *a b* Proudfoot AE, Turcatti G, Davies JG, Wingfield PT (1991). "Human interleukin-5 expressed in Escherichia coli: assignment of the disulfide bridges of the purified unglycosylated protein". *FEBS Lett.* **283** (1): 61–64. doi:10.1016/0014-5793(91)80553-F (http://dx.doi.org/10.1016%2F0014-5793%2891%2980553-F) . PMID 2037074 (http://www.ncbi.nlm.nih.gov/pubmed/2037074) .
14. ^ *a b c* Hirano T, Harada H, Nakajima K, Iwamatsu A, Yasukawa K, Taga T, Watanabe Y, Matsuda T, Kashiwamura S, Koyama K (1986). "Complementary DNA for a novel human interleukin (BSF-2) that induces B lymphocytes to produce immunoglobulin". *Nature* **324** (6092): 73–76. doi:10.1038/324073a0 (http://dx.doi.org/10.1038%2F324073a0) . PMID 3491322 (http://www.ncbi.nlm.nih.gov/pubmed/3491322) .
15. ^ *a b c* Heinrich PC, Rose-John S, Lutticken C, Kruttgen A, Moller C (1991). "Evidence for the importance of a positive charge and an alpha-helical structure of the C-terminus for biological activity of human IL-6". *FEBS Lett.* **282** (2): 265–267. doi:10.1016/0014-5793(91)80491-K (http://dx.doi.org/10.1016%2F0014-5793%2891%2980491-K) . PMID 2037043 (http://www.ncbi.nlm.nih.gov/pubmed/2037043) .
16. ^ *a b c* Clogston CL, Boone TC, Crandall BC, Mendiaz EA, Lu HS (1989). "Disulfide structures of human interleukin-6 are similar to those of human granulocyte colony stimulating factor". *Arch. Biochem. Biophys.* **272** (1): 144–151. doi:10.1016/0003-9861(89)90205-1 (http://dx.doi.org/10.1016%2F0003-9861%2889%2990205-1) . PMID 2472117 (http://www.ncbi.nlm.nih.gov/pubmed/2472117) .

17. ^ Cook WJ, Bugg CE, Walter MR, Ealick SE, Zhao BG, Reichert P, Nagabhushan TL, Trotta PP, Cameron Jr RP, Walter Jr RL (1992). "Crystal structure of recombinant human interleukin-4". *J. Biol. Chem.* **267** (28): 20371–20376. PMID 1400355 (http://www.ncbi.nlm.nih.gov/pubmed/1400355) .
18. ^ Henney CS (1989). "Interleukin 7: effects on early events in lymphopoiesis". *Immunol. Today* **10** (5): 170–173. doi:10.1016/0167-5699(89)90175-8 (http://dx.doi.org/10.1016%2F0167-5699%2889%2990175-8) . PMID 2663018 (http://www.ncbi.nlm.nih.gov/pubmed/2663018) .
19. ^ Renauld JC, Goethals A, Houssiau F, Merz H, Van Roost E, Van Snick J (1990). "Human P40/IL-9. Expression in activated CD4+ T cells, genomic organization, and comparison with the mouse gene". *J. Immunol.* **144** (11): 4235–4241. PMID 1971295 (http://www.ncbi.nlm.nih.gov/pubmed/1971295) .
20. ^ Paul WE, Boulay JL (1993). "Hematopoietin sub-family classification based on size, gene organization and sequence homology". *Curr. Biol.* **3** (9): 573–581. doi:10.1016/0960-9822(93)90002-6 (http://dx.doi.org/10.1016%2F0960-9822%2893%2990002-6) . PMID 15335670 (http://www.ncbi.nlm.nih.gov/pubmed/15335670) .
21. ^ Wlodawer A, Zdanov A, Schalk-Hihi C, Gustchina A, Tsang M, Weatherbee J (1995). "Crystal structure of interleukin-10 reveals the functional dimer with an unexpected topological similarity to interferon gamma". *Structure* **3** (6): 591–601. doi:10.1016/S0969-2126(01)00193-9 (http://dx.doi.org/10.1016%2FS0969-2126%2801%2900193-9) . PMID 8590020 (http://www.ncbi.nlm.nih.gov/pubmed/8590020) .
22. ^ Jiang H, Lin JJ, Goldstein NI, Fisher PB, Su ZZ (1995). "Subtraction hybridization identifies a novel melanoma differentiation associated gene, mda-7, modulated during human melanoma differentiation, growth and progression". *Oncogene* **11** (12): 2477–2486. PMID 8545104 (http://www.ncbi.nlm.nih.gov/pubmed/8545104) .
23. ^ Leng SX, Elias JA (1997). "Interleukin-11". *Int. J. Biochem. Cell Biol.* **29** (8-9): 1059–1062. doi:10.1016/S1357-2725(97)00017-4 (http://dx.doi.org/10.1016%2FS1357-2725%2897%2900017-4) . PMID 9416001 (http://www.ncbi.nlm.nih.gov/pubmed/9416001) .
24. ^ Park AY, Scott P (2001). "Il-12: keeping cell-mediated immunity alive". *Scand. J. Immunol.* **53** (6): 529–532. doi:10.1046/j.1365-3083.2001.00917.x (http://dx.doi.org/10.1046%2Fj.1365-3083.2001.00917.x) . PMID 11422900 (http://www.ncbi.nlm.nih.gov/pubmed/11422900) .
25. ^ Presky DH, Gately MK, Gubler U, Renzetti LM, Magram J, Stern AS, Adorini L (1998). "The interleukin-12/interleukin-12-receptor system: role in normal and pathologic immune responses". *Annu. Rev. Immunol.* **16**: 495–521. doi:10.1146/annurev.immunol.16.1.495 (http://dx.doi.org/10.1146%2Fannurev.immunol.16.1.495) . PMID 9597139 (http://www.ncbi.nlm.nih.gov/pubmed/9597139) .
26. ^ Minty A, Chalon P, Dumont X, Kaghad M, Labit C, Leplatois P, Liauzun P, Miloux B, Derocq JM, Guillemot JC (1993). "Interleukin-13 is a new human lymphokine regulating inflammatory and immune responses". *Nature* **362** (6417): 248–250. doi:10.1038/362248a0 (http://dx.doi.org/10.1038%2F362248a0) . PMID 8096327 (http://www.ncbi.nlm.nih.gov/pubmed/8096327) .
27. ^ Arena A, Merendino RA, Bonina L, Iannello D, Stassi G, Mastroeni P (2000). "Role of IL-15 on monocytic resistance to human herpesvirus 6 infection". *New Microbiol.* **23** (2): 105–112. PMID 10872679 (http://www.ncbi.nlm.nih.gov/pubmed/10872679) .
28. ^ *a b* Gurney AL, Aggarwal S (2002). "IL-17: prototype member of an emerging cytokine family". *J. Leukoc. Biol.* **71** (1): 1–8. PMID 11781375 (http://www.ncbi.nlm.nih.gov/pubmed/11781375) .
29. ^ *a b c d* Unless else specified in boxes, then ref is: Lippincott's Illustrated Reviews: Immunology. Paperback: 384 pages. Publisher: Lippincott Williams & Wilkins; (July 1, 2007). Language: English. ISBN 0781795435. ISBN 978-0781795432. Page 68
30. ^ Noosheen Alaverdi & David Sehy (2007-05-01). "Cytokines - Master Regulators of the Immune System" (http://www.ebioscience.com/ebioscience/whatsnew/pdf/Cytokines.pdf) . eBioscience. http://www.ebioscience.com/ebioscience/whatsnew/pdf/Cytokines.pdf. Retrieved 2008-02-28.
31. ^ *a b c d e f g h i j k l m n o p* Cytokine tutorial, The University of Arizona (http://microvet.arizona.edu/Courses/MIC419/Tutorials/cytokines.html)

External links

211

- Cytokines & Cells Online Pathfinder Encyclopedia (http://www.copewithcytokines.de/)
- HGNC Gene Family Nomenclature: Interleukin and Interleukin Receptor Gene Symbols (http://www.gene.ucl.ac.uk/nomenclature/genefamily/il.php)
- Interleukin Antibody Review (http://www.exactantigen.com/review/Interleukin.html)

This article includes text from the public domain Pfam and InterPro IPR000779 (http://www.ebi.ac.uk /interpro/IEntry?ac=IPR000779)

Retrieved from "http://en.wikipedia.org/wiki/Interleukin"
Categories: Cytokines

- This page was last modified on 27 July 2011 at 19:07.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. See Terms of use for details.
  Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

Autoimmunity Reviews 9 (2010) 785–792

Contents lists available at ScienceDirect

Autoimmunity Reviews

journal homepage: www.elsevier.com/locate/autrev

Review

Interleukin-17-producing T helper cells in autoimmunity

Nasr YA. Hemdan [a,b,*], Gerd Birkenmeier [c], Gunnar Wichmann [b], Ahmed M. Abu El-Saad [a], Thorsten Krieger [d], Karsten Conrad [e], Ulrich Sack [f]

[a] Department of Zoology, Faculty of Science, University of Alexandria, Egypt
[b] ENT-Research Lab, Department of Otolaryngology, Head and Neck Surgery, Faculty of Medicine, University of Leipzig, Germany
[c] Institute of Biochemistry, Faculty of Medicine, University of Leipzig, Germany
[d] Asculabor Hamburg, Germany
[e] Institute of Immunology, Faculty of Medicine, University of Dresden, Germany
[f] Institute of Clinical Immunology, Faculty of Medicine, University of Leipzig, Germany

ARTICLE INFO

Article history:
Received 2 July 2010
Accepted 15 July 2010
Available online 18 July 2010

Keywords:
Autoimmune diseases
Cytokines
IL-17
Interleukin 17
Regulatory T cells
T helper 17 (T_H17) cells

ABSTRACT

With all the incredible progress in scientific research over the past two decades, the trigger of the majority of autoimmune disorders remains largely elusive. Research on the biology of T helper type 17 (T_H17) cells over the last decade not only clarified previous observations of immune regulations and disease manifestations, but also provided considerable information on the signaling pathways mediating the effects of this lineage and its seemingly dual role in fighting the invading pathogens on one hand, and in frightening the host by inducing chronic inflammation and autoimmunity on the other hand. In this context, recent reports have implicated T_H17 cells in mediating host defense as well as a growing list of autoimmune diseases in genetically-susceptible individuals. Herein, we summarize the current knowledge on T_H17 in autoimmunity with emphasis on its differentiation factors and some mechanisms involved in initiating pathological events of autoimmunity.

© 2010 Elsevier B.V. All rights reserved.

Contents

1. Introduction . . . 785
2. Key factors in the differentiation program of T_H17 cells . . . 786
   2.1. Interleukin-6 . . . 786
   2.2. Transforming growth factor-β . . . 787
   2.3. Interleukin-21 . . . 787
   2.4. Interleukin-23 . . . 788
   2.5. Interleukin-1β and tumor necrosis factor-α . . . 788
   2.6. Signal transducer and activator of transcription 3 . . . 788
3. Differential regulation of T_H17 cytokines and some underlying mechanisms of autoimmunity . . . 788
4. Concluding remarks . . . 789
Take-home messages . . . 790
Acknowledgements . . . 790
References . . . 790

1. Introduction

Although, the last years have witnessed an extensive research in the field of autoimmunity that included both experimental as well as epidemiological cohort studies, most of the autoimmunity triggers remain largely hidden and many questions persist. One most important question concerns the contribution of various T_H cell subsets. The T_H1/T_H2 paradigm [1] was attained that upon activation, $CD4^+$ T cells differentiate into T_H1 cells (indispensable for resolution of intracellular pathogens) or T_H2 cells (targeting intracellular pathogens and inducing B cell differentiation into antibody-producing cells) in presence of IL-12/IFN-γ or IL-4, respectively. Both subsets attain their maturation through self-promoting and reciprocal regulatory mechanisms mediated by

* Corresponding author. ENT-Research Lab, Department of Otolaryngology, Head and Neck Surgery, Faculty of Medicine, University of Leipzig, Germany. Tel.: +49 17620781360; fax: +49 3419721709.
E-mail address: nhemdan@yahoo.de (N.Y.A. Hemdan).

1568-9972/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.autrev.2010.07.003

intrinsic signals, reviewed in Refs. [2–4]. Activation of signal transducer and activator of transcription (STAT)1 by IFN-γ induces T-bet expression, which, in turn, upregulates IL-12 receptor β2-subunit (IL-12Rβ2). IL-12, acting through STAT4, prolongs T_H1 survival and IFN-γ expression. Commitment of T_H2 is established through activation of STAT6 by IL-4 and subsequently induction of the transcription factor GATA3 that, in turn, upregulates IL-4 expression. Indeed, this T_H1/T_H2 paradigm seems paradoxical due to the fact that anti-IFN-γ antibody (Ab)-treated, IFN-$\gamma^{-/-}$ or IFN-$\gamma R^{-/-}$ mice develop collagen-induced arthritis (CIA) or experimental autoimmune encephalomyelitis (EAE), a model of human multiple sclerosis [5], previously known as T_H1-associated diseases.

Recent progress has been made in our understanding of T cell differentiation following the first detection of IL-17 as a transcript from a rodent T cell hybridoma by Rouvier and colleagues [6] originally named cytotoxic T lymphocyte associated antigen (CTLA)-8, and the later discovery of human IL-17 [7] and IL-17 receptor (IL-17R) [8], which together with IL-22R are ubiquitously expressed in various tissues [9]. Following identifying other related effecter cytokines, the T_H1/T_H2 paradigm has been expanded to include T_H17 cells, distinct from both T_H1 and T_H2 cells, and since then a new era has begun. This included implication of T_H17 in induction and/or aggravation of autoimmunity in experimental models and in genetically-predisposed individuals (Table 1), that may assign T_H17 cell lineage as a hallmark for autoimmunity.

IL-17 is a pronociceptive cytokine [10], and together with IL-17F both are key factors for activation and recruitment of neutrophils, which are of pivotal roles in inflammatory and autoimmune diseases including rheumatoid arthritis (RA) [10,11], amongst others. Both IL-17 and IL-22 are mainly produced by T_H17 cells and promote production of antimicrobial peptides [12,13] constituting thereby a link between innate and adaptive responses [14,15]. Other sources of IL-17 include NK cells, which may be the major innate source of IL-17 [16], γδT cells [15,17,18] and $CD8^+$ T cells [19]. IL-$17^{-/-}$ mice reveal a reduced IFN-γ production by $CD4^+$ T cells, impaired granuloma formation and expression of the chemokines CXCL9, CXCL10 and CXCL11 [20,21], which all signal via CXCR3 mainly expressed on T_H1 cells [22], involved in wound healing [23], in autoimmune diseases [24,25] and allograft rejection [26]. Therefore, treatments, such as Natura-alpha, both prevented and treated CIA mediated by inhibiting IL-17 production and anti-type II collagen IgG development, amongst other pathogenic factors [27].

IL-22, a member of the IL-10 cytokine family, is also produced by NK cells (NK22), and lymphoid tissue inducer cells following IL-23 stimulation [28–30], and is also produced by a distinct lineage, "T_H22" cells [31,32]. IL-22 is linked to induction of tissue repair and protection against experimental autoimmune myocarditis [33], IBD [34] and liver injury [35,36].

2. Key factors in the differentiation program of T_H17 cells

Studies of T_H17 defined important factors essential for establishing a sustained T_H17 cell lineage. These, presumably similar in both mice and human, include differentiation factors TGF-β, IL-6 and IL-21, the stabilization factor IL-23, and the transcription factors STAT3, and the retinoic acid-related orphan nuclear receptor (ROR)γt and RORα.

2.1. Interleukin-6

IL-6 is a pleiotropic and proinflammatory cytokine secreted by macrophages, DCs, epithelial and T cells, and exerts manifold effects in T cell differentiation. This includes induction of pathogenic T_H1 response [37]. The same cytokine, however, activates transcription mediated by the nuclear factor (NF)-AT leading to production of IL-4

Table 1
Protective and detrimental effecter molecules of T_H17 cells in autoimmune diseases.

Disease	Role of T_H17 cells	Observation(s)/conclusion	Reference
Multiple sclerosis (MS) and EAE	Pathogenic	Elevated IL-17 response in blood and CSF cells of MS patients during clinical exacerbations	[19,106,107]
		Inhibition of chemokine expression in brain by anti-IL-17 antibody treatment	[108]
		Reduction of incidence in IL-$1RI^{-/-}$ mice correlates with failure to induce autoantigen-specific T_H17 cells	[85]
		Resistance of IL-$6^{-/-}$ and anti-IL-6 Ab-treated mice due to compromised T_H17 response	[41,42]
		Reduction of severity and recruitment of T_H17 cells to inflammatory tissues on deletion of CCR6 or of IL-17	[109,110]
		Induction of T_H17 response promotes EAE	[111]
Psoriasis-like dermatitis	Pathogenic	Dependence of disease development on an intact IL-23/IL-17 axis	[112]
Psoriasis	Pathogenic	Presence of T_H17 in psoriasis skin lesions. Correlation of IL-17 and IL-22 mRNA levels with disease activity	[113]
		Anti-IL-12/23 p40 Ab is effective in the treatment of psoriasis	[114]
		Presence of T_H17 in psoriasis skin and contribute to IL-6-mediated resistance to T_{reg} suppression	[39]
		Increase of IL-17F $CD4^+$ T cells in psoriatic skin lesions	[115]
		IL-22- and IL-17-mediated induction of human β-defensin 2, S100A9, S100A7 and S100A8 in psoriatic skin	[12]
Rheumatoid arthritis and Ag-induced Arthritis	Pathogenic	IL-1β, TNF-α, IL-17, and IL-10 mRNA levels in synovial membrane, pronounced with shorter disease duration, correlate with joint damage	[116]
		Induction of T_H17 response in arthritis patients (IL-17, TNF-α and GM-CSF).	[117,118]
		IL-17-dependent enhancement of K/BxN serum-transferred; enrichment of IL-17-producing KRN T cells in inflamed joints by arthritogenic AAs	[119]
		Inhibition of hypernociception in $TNFR^{-/-}$ mice, by pre-treament with anti-IL-17, anti TNF-α, or anti IL-1R	[10]
		Disease attenuation and reduction of systemic IL-6 and synovial IL-1- and RANKL-positive cells on absence of IL-17	[120,121]
		Induction of inflammation by IFN-γ and IL-17 in proteoglycan-induced arthritis	[122]
Autoimmune diabetes	Pathogenic	Reduction of peri-islet T-cell infiltrates and GAD65 AA levels by anti-IL-17 and IL-25	[123]
Anti-neutrophil cytoplasm antibody (ANCA)-associated vasculitis	Pathogenic	Induction of serum IL-17, IL-23 and autoantigen-specific IL-17-, but not IFN-γ-producing cells during disease convalescence	[124]
Sjogren's syndrome	Pathogenic	Domination of IL-17-positive cells in the inflammatory lesions, and abundant expression of TGF-β, IL-6 and IL-23; Induction of T_H17-associated mRNA transcripts in cornea and conjunctiva following disruption of IL-2 signalling in $CD25^{-/-}$ mice results in age-dependent SS-like autoimmune lacrimal-keratoconjunctivitis	[125,126]
Allergen-induced airway and contact Hypersensitivity	Pathogenic	Detection of infiltrating neutrophils and cells expressing IL-17, IL-22, CCR6, and IL-22R in inflamed skin of nickel-challenged allergic individuals. T_H17 and T_H1 cells in memory PB cells	[127]
		Airway neutrophilia following allergic sensitization mediated by natural and primed T_H17 cells	[50,128]
		Induction of IL-17 levels in asthmatics, that correlates with severity. Reduction of contact, delayed-type, and airway hypersensitivity responses, and T-dependent Ab production in IL-$17^{-/-}$ mice	[129]
Autoimmune thyroid diseases	Pathogenic	Induction of T_H17 response in patients (high IL-17 and IL-22, IL-23 and IL-23R)	[130]
Thrombocytopenia	Pathogenic	Elevation of IL-21 plasma levels and IL-21^+, IL-17^+ and IFN-γ^+ T cells in patients	[131]

by naïve Th0 cells and their differentiation into effecter T_H2 cells [38]. The latter authors found that IL-6 inhibits T_H1 differentiation via upregulating expression of the transcription factor suppressor of cytokine signaling (SOCS)-1 that interferes with IFN-γ signaling and the development of T_H1 cells. This may constitute a negative feedback loop to culminate T_H1 induction, and moreover can support a more tolerant state. Furthermore, IL-6 signaling through STAT3 enables T cells to escape from T_{reg} suppression [39].

Applying neutralizing anti-IL-6 Ab, which completely abolished differentiation of T_H17, highlights the importance of IL-6 in the direct induction of T_H17 [40], and inhibits T_H17-mediated development of EAE [41]. T_{reg} cells can be substituted by TGF-β1, which, together with IL-6, subverts T_H1 and T_H2 differentiation for the generation of T_H17 cells [40], a process that is amplified by IL-1β and TNF-α. All of these findings endorse earlier studies, pre-dating the discovery of T_H17 cells, which implicate IL-6 in development of EAE [42], SLE [43] or Ag-induced arthritis [44] probably due to defective T_H17 repertoires. Taken together, and because of the perception that IL-6 release is a key cytokine of the acute phase response during infection [45], may be similar to other downstream proinflammatory cytokines, a subversion of T_{reg} cells into IL-17-producers is most likely to occur.

Recent results driven from investigations on IL-$6R^{-/-}$ mice reveal that the local control of IL-6 trans-signaling regulates the effector characteristics of T cell infiltrates and promotes the maintenance of T_H17 cells [46]. Another interesting recent study using $Lyn^{-/-}$ mice, which develop Ab-mediated autoimmune disease resembling SLE, demonstrated that $Lyn^{-/-}$ leukocytes, notably B cells, overproduce IL-6 and thereby facilitate hyperactivation of B and T cells, enhance myelopoiesis, splenomegaly and ultimately the generation of pathogenic autoreactive antibodies (AA) [47]. Moreover, deleting IL-6 on a $Lyn^{-/-}$ background does not alter B cell development, plasma cell accumulation, or DC hypermaturation, but hyperactivation of B and T cells, extramedullary hematopoiesis, expansion of the myeloid lineage and autoimmune disease are all ameliorated [47]. The authors elegantly formulated that although $Lyn^{-/-}$ B cells may be autoreactive, it is the IL-6-dependent inflammatory environment they engender that dictates their disease-causing potential. However, a possible interference with IL-17 response has not been directly issued. We propose that a possible IL-17-mediated mechanism might orchestrate such an autoimmune disorder. Altogether, it may then not be surprising that several groups have demonstrated the feasibility of IL-6-targeting therapies of some autoimmune diseases such as RA [48], or of Graft-versus-host disease mediated by T_H17 cells [49]. A novel set of memory-like natural T_H17 (termed nT_H17), however, can develop in the absence of the IL-6/STAT3 signaling and express substantial amounts of IL-17, and IFN-γ but not IL-17F, under the control of RORγt, and play a crucial role in induction of neutrophilia [50].

2.2. Transforming growth factor-β

The transforming growth factor (TGF)-β, together with IL-10, was mostly assigned as an anti-inflammatory cytokine, produced mostly by $CD4^+$ $CD25^+$ T_{reg} cells, which attained a distinct immunosuppressive property on various arms of immune response and maintain self-tolerance [51]. This view does not, however, fit into the recent perception that TGF-β is indispensable for the differentiation of murine and human T_H17 cells. Because of its key role in the differentiation program of T_H17 cells and in the differentiation of T_{reg} cells in the periphery, TGF-β is at the crossroads of both T cell subsets, which are mutually exclusive. Both lineages are descendants of a common $Foxp3^+$ $ROR\gamma t^+$ precursor that develops on convergence of IL-2 and TGF-β signaling, as reviewed in Ref. [4]. Subsequent TGF-β or IL-6 signaling upregulates one lineage and inhibits the other via activating Foxp3 or RORγt, respectively. A recent finding that also challenges the old view of TGF-β is that under certain conditions, it cooperates with IL-4 to support a distinct population of IL-9-producing "T_H9" T cells [52].

Whereas, local administration of neutralizing anti-TGF-β Ab in target sites of chronic inflammation ameliorated the disease through limiting T_H17 cell differentiation, similar interventions at mucosal sites exacerbated disease in experimental models of colitis [53], implying distinct functions of T_H17 cells as well as their protective role in mucosal immunity, see below. In a model of EAE, the adoptive transfer of induced T_{reg} cells suppresses EAE when administrated before induction or during ongoing disease due to the direct inhibition of cell proliferation [54]. Because the inhibitory effects are influenced by Foxp3 [55], suppressing IL-2, IL-4 and IFN-γ transcription through interactions with NF-κB and NF-AT [56], it is then the balance between TGF-β and Foxp3 that determines the outcome of the T_{reg}–T_H17 balance in favor of one subset at the expense of the other, an assumption that accords well with the association of impaired T_{reg} cell activities with autoimmune disorders including SLE [57]. Indeed, the reciprocal interrelationship between T_{reg} and T_H17 cells has been proved in several models, however, both lineages were increased in psoriasis patients both in the peripheral circulation and skin tissue lesions and were positively correlated with disease severity [58].

In the absence of TGF-β signaling, mice did not develop EAE due to the absence of T_H17 cells [59]. Furthermore, TGF-$\beta1^{-/-}$ mice developed lethal immunopathology in multiple organs, which was associated with differentiated T_H1 and T_H2 cells. In a transfer model, TGF-β1 produced by $Foxp3^+$ T_{reg} cells was required to inhibit differentiation of T_H1 cells as well as IBD [59], implying that T_{reg} cells might be the source of TGF-β in this model. This finding is consistent with the induced differentiation of T_H17 on the adaptive transfer of T_{reg} cells, although it prevented the disease when administered early in the course by suppressing expansion of T cells [54]. TGF-β1 is for sustained expression of IL-17F and IL-17 [60] and is also necessary for IL-17-competent γδ T thymocytes, an innate source of IL-17 that retained in thymus particularly during the postnatal life [61].

Incubating $CD4^+$ T cells from DO11.10/$RAG2^{-/-}$ mice with $C5aR^{-/-}$ spleen-derived DC results in increased production of TGF-β, the *de novo* differentiation of T_{reg} cells and induced high levels of IL-6 and IL-23 but minor amounts of IL-12 lead to differentiation of T_H17 cells [62]. This implies an important role of C5aR in DC directing the differentiation of T cells [62]. Induction of T_H17 cells is achieved through inhibition of MEK-ERK signaling that otherwise promotes production of IL-2 and activation STAT5 [63], and thereby T_{reg} cells. A recent identified group of $CCR6^+$ T_{reg} cells in human PB and lymphoid tissue is generated in the periphery, coexpresses FOXP3 and RORγt, and produces IL-17 upon activation [64]. Furthermore, $CD4^+$ $FOXP3^+$ $CCR6^-$ T_{reg} cells differentiate into IL-17 producers upon TCR stimulation in the presence of IL-1β, IL-2, IL-21, IL-23, and human serum (probably a source of TFG-β).

2.3. Interleukin-21

IL-12 is an IL-2 cytokine family member produced by activated T and NKT cells but not by APC [65], with an autocrine activity that regulates its own production through potently suppressing Foxp3 expression, enhancing T_H17 proliferation and expression of RORc [66,67]. In the absence of T_{reg} cells, IL-$6^{-/-}$ mice developed a T_H17-mediated EAE, probably induced by IL-21, as IL-$21R^{-/-}$ T cells are compromised in T_H17 [68]. The contribution of IL-21, in the absence of IL-6, seems to suppress TGF-β-induction of Foxp3, and together with TGF-β to induce T_H17. The combination of both cytokines promotes the differentiation of human naïve $CD4^+$ T cells into T_H17 cells expressing RORc2 [69] and RORγt [70,71]. A recent report indicates that IL-21 expression correlates with the appearance of T_H17 cells in synovial fluid and peripheral blood in RA patients [67], indicating that targeting IL-21 may be tractable to dampen autoimmune induction

[72] that results in downregulation of IL17 production [73], B cell activity and production of Abs. [74].

2.4. Interleukin-23

Similar to IL-12 for T_H1 and IL-4 for T_H2 [75,76], IL-23, an IL-12 cytokine family member that shares with IL-12 a common subunit, p40, is a key factor for sustained commitment of T_H17 cells. Therefore, following their initial induction via TGF-β signaling, T_H17 cells upregulate IL-23R to confer their responsiveness to IL-23 [77–79]. In a model of CIA [80], IL-12p35$^{-/-}$ mice developed more IL-17-producing T cells, accompanied by elevated mRNA expression of TNF-α, IL-1β and IL-6 in affected tissues of diseased than in wild-type mice. In the absence of TGF-β, however, both of IL-23 and IL-12 suppress IL-17 and enhance IFN-γ production in a STAT4- and T-bet-dependent manner [81]. This implies that IL-23 alone cannot initiate T_H17 differentiation.

Giving the crucial importance of IL-23, it is then not surprising that p19$^{-/-}$ mice, deficient in functional IL-23, are resistant to EAE [82]. Interestingly, McGreachy et al. [83] found that stimulation of myelin-reactive T cells with TGF-β plus IL-6 completely abrogated their pathogenic function despite upregulation of IL-17 production. These cells failed to induce CNS inflammation and produced IL-10 and thereby acted as "regulatory"-cells. Stimulation with IL-23, however, promoted expression of IL-17 and proinflammatory chemokines but not IL-10, and therefore the authors concluded that TGF-β and IL-6 drive the initial commitment of T_H17 cells but simultaneously restrain their pathogenic potential, implying that IL-23 is indispensable for full acquisition of effecter T_H17 cells.

2.5. Interleukin-1β and tumor necrosis factor-α

Both of IL-1β and TNF-α are important factors of pathogenesis in various diseases, e.g. MS [84]. The importance of IL-1β in induction of T_H17 response is evident from IL-1β/IL-1βR$^{-/-}$ mice studies. For example, upon immunization with myelin peptide, the incidence of EAE is significantly lower in IL-1RI$^{-/-}$ compared with wild-type mice, and this correlates with a failure to induce autoantigen-specific T_H17 cells [85]. Similarly, IL-1β induces the differentiation of human T_H17 cells, most of them also produced IFN-γ [86], and the combination of IL-1β and IL-6 boosts IL-17 production but not IFN-γ. Furthermore, Lee and colleagues found that IL-1β induces IL-17 production from human CD4$^+$ T cells and this can be blocked by IL-1R antagonist [87]. That IL-7, IL-15 and TGF-β upregulate IL-1R type I in naïve umbilical cord-driven CD4$^+$ T cells that lack this receptor indicates that this takes place in the periphery. Because of the finding that IL-1RI$^+$ memory CD4$^+$ T cells increase gene expression of IL-17, RORc and IRF4 even before TCR triggering, the authors suggest that the effect of IL-1β is programmed in these cells via IL-1RI.

2.6. Signal transducer and activator of transcription 3

The role of STAT3 in the commitment of T_H17 cells is presented in a recent excellent review [4,88]. IL-6 upregulates IL-23R on the differentiated T_H17 cells, and together with IL-23 or IL-21 promotes T_H17 differentiation via activation of STAT3 [88]. Mice with targeted-deletion of STAT3 in CD4$^+$ T cells do not develop experimental autoimmune uveoretinitis (EAU) or EAE [89], revealing a defective T_H17 response and exaggerated increases in Foxp3-, IL-10-, IL-4-, and IFN-γ-expressing T cells. Mice with EAU reveal a high percentage of IL-17-expressing T cells in their peripheral lymphoid organs, characterized also by secretion of IFN-γ, and hence propose the involvement of T_H17 and IL-17-expressing T_H1 cells. In presence of TGF-β, STAT3 induces RORα and RORγt [88,90,91], and hence is indispensable for the susceptibility to EAE [90]. Therefore, differentiation of T_H17 was greatly impaired in STAT3 deficient T cells, marked by impaired RORγt expression and instead elevated expression of T-bet and Foxp3 [88].

Through Smad signal transduction pathway, TGF-β inhibits IL-6- and IL-21-induced SOCS3 promoter activity, i.e. SOCS3 expression, and thereby enhances and prolongs STAT3 activation in naïve T cells [92] (Fig. 1). In mice with impaired TGF-β signaling, IL-6-induced expression of SOCS3 in T cells is higher, whereas STAT3 activation is lower compared with wild-type cells [92]. Interestingly, similar to the need of T_{reg} cells for IRF4 to suppress T_H2 [93] or for T-bet to suppress T_H1 [94], Chaudhry and colleagues found that it is the same requisite transcription factor STAT3 that is also required by T_{reg} cells to suppress T_H17 response [95], ensuring, thereby, as the authors proposed, a class-specific control of immune-mediated inflammation.

3. Differential regulation of T_H17 cytokines and some underlying mechanisms of autoimmunity

As mentioned above, T_H17 cells produce IL-6, IL-10, TNF-α, IFN-γ, IL-26 and mainly IL-17, IL-21 and IL-22. The production of these cytokines is differentially regulated. In their interesting work, Volpe et al. [96] have used a multi-parametric analysis of T_H17-associated cytokines and reported that the regulation pattern of IL-22 was most closely related to IFN-γ and not to IL-17. Moreover, whereas IL-12 and IL-23 induce high levels of IL-22 but not IL-17, TGF-β, inhibits IL-22 production but promoted IL-17 [96], and more extremely, its absence induces a shift from a T_H17 profile to a T_H1-like profile [97]. Interestingly, Sofi et al. [98] found that thymus (T)-CD4$^+$ T cells express very low levels of TGF-βR and IL-21R, which are essential to induce IL-17 expression. Under T_H17-skewing conditions, these cells revealed a compromised induction of RORγt and hence reduced IL-17 expression, which could be restored via ectopic expression of RORγt. Therefore, the authors delivered an excellent conclusion that the defect of IL-17 and RORγt expression in these cells is cell intrinsic, a characteristic that helps control the generation of the T_H17 lineage.

IL-27, a member of the IL-12 cytokine family, which is build up of two non-covalently bound subunits, Epstein Barr virus-induced gene-3 and p28, and can be also secreted as a homodimeric IL-27p28 [99], is an inhibitor of T_H17 cells [100]. Altogether, and recalling that T_{reg} and T_H17 cells are reciprocally regulated and that low levels of TGF-β plus IL-6 or IL-21 is enough for initiation of effecter T_H17 but not for production of IL-22-producing cells, unless in presence of IL-23, an immune response is programmed to function in a way that: 1) following infection, pathogen-specific induction of IL-12/IFN-γ axis favors proliferation of T_H1 response, that together with IL-27 keep T_H17 in their steady-state; 2) the increase in IL-6 production inhibits T_H1 and T_{reg} cells and together with low TGF-β levels promotes differentiation of T_H17 cells with a regulatory function that still needs support of IL-23 to attain their full effecter potency producing IL-22, CXC chemokines, antimicrobial peptides and IL-21; 3) long-lasting apoptosis-resistant T_H17 cells [101] activate B cells and their immunoglobulin production mediated by IL-21 [74]. T_H17 cells may lead to the induction of latent inflammatory T_H17 response and hyper-activation of B cells and production of auto-reactive Abs (AA).

Distinct from its proinflammatory effects, IL-17 promotes autoimmune disease by enhancing formation of spontaneous germinal centers (GCs), as shown by autoimmune BXD2 mice that express more IL-17 than wild-type counterparts and show spontaneous development of GCs before they increase production of pathogenic AA [102]. IL-17R$^{-/-}$ mice have reduced GC B cell development and humoral responses. Production of IL-17 correlates with induction of the genes Rgs13 and Rgs16, which encode regulators of G-protein signaling, and results in suppression of the B cell chemotactic response to the chemokine CXCL12. The latter chemokine is constitutively expressed by the CNS [103], its expression on the

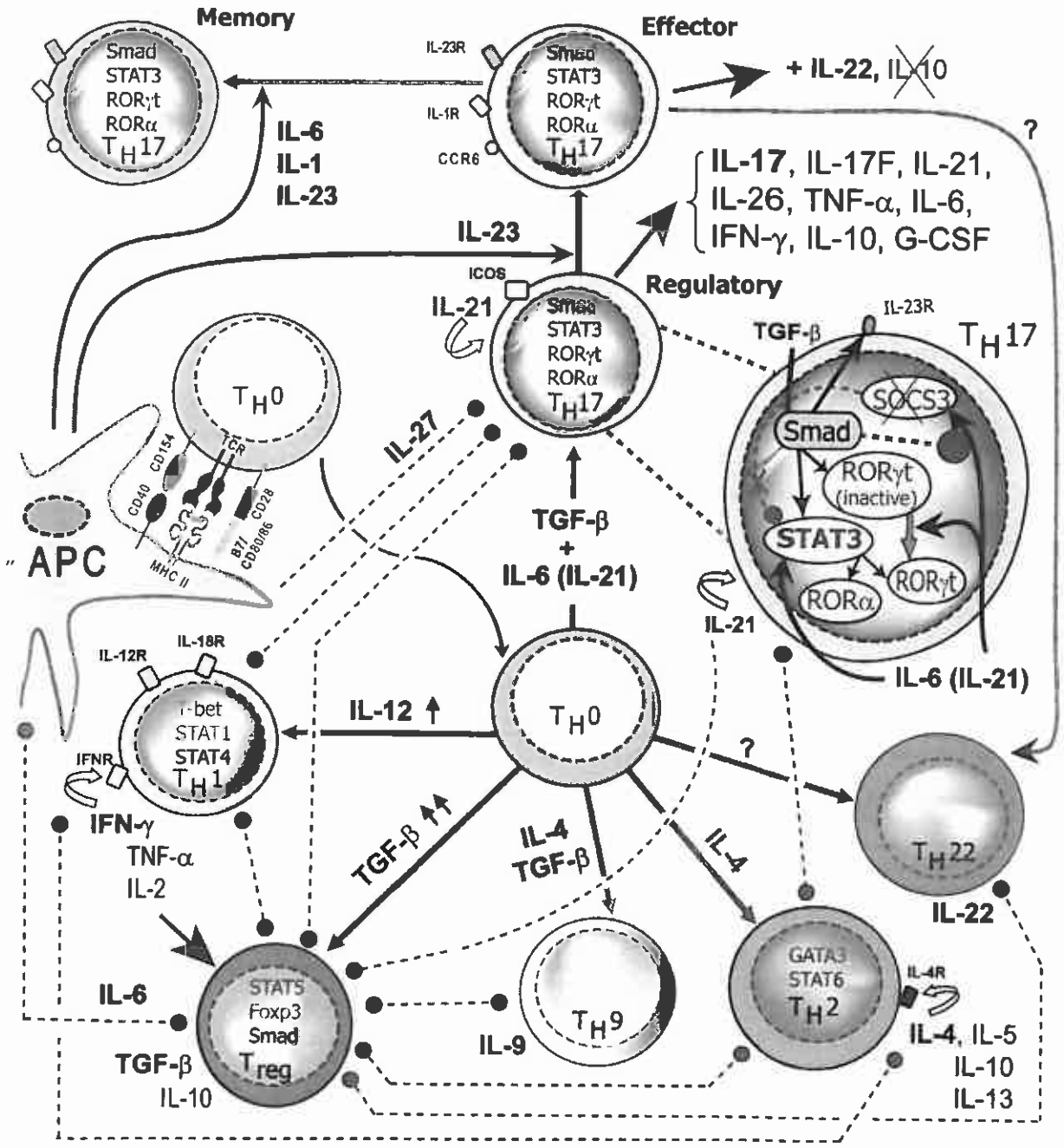


Figure 1. Model of differentiation of IL-17-producing T helper cells (T_H17). Differentiation of T_H17 from uncommitted naïve (T_H0) cells showing cytokines and transcriptional factors at the nexus of various T_H cell lineages is depicted. Upon stimulation with pathogens or allergens, antigen presenting cells (APC) induce differentiation of T_H0 cells into distinct effecter lineages T_H1, T_H2, T_H9, T_H17 or T_H22 as well as the forkhead box P3 (Foxp3)-expressing T regulatory (T_{reg}) cells according to local cytokine milieu produced by various innate immune cells. Whereas activation of STAT1 induces expression of T-bet (the key differentiation factor of T_H1 cells), STAT6 signaling up-regulates the expression of the master regulator of T_H2 cell differentiation GATA3. T_H1 and T_H2 cells reciprocally regulate each other and negatively regulate the generation of T_H17 cells through their hallmark effecter cytokines, interferon (IFN)-γ and interleukin (IL)-4, respectively. Differentiation of T_H17 requires co-stimulatory signals of CD28 and ICOS and the absence of T_H1 and T_H2 cytokines (IL-12/IFN- and IL-4, respectively) and their key transcriptional factors (T-bet and GATA3). Induction of T_{reg} and T_H17 is mutually exclusive. In presence of low levels of TGF-β, that signals through the Smad pathway, IL-6 (as well as other proinflammatory cytokines IL-1β, TNF-, IL-18 or IL-21) favor differentiation of TH17 cells. TGF-β alone induces Foxp3 expression and hence T_{reg} differentiation, also is indispensable for induction of expression of RORt (RORc in human) which is yet inactive and can be only activated in presence of proinflammatory cytokines to start the full transcript machinery of T_H17. TGF-β maintains the cell responsiveness to IL-6 signaling that, in turn, activates STAT3. STAT3 inhibits SOCS3 and thereby induces the transcription factors RORt and ROR and subsequently IL-23R. IL-23 is crucial for survival and effecter functions of T_H17 cells and production of IL-22 (TH22?).

parenchymal surface of the CNS endothelium normally functions to localize infiltrating leukocytes to peri-vascular spaces, preventing CNS infiltration [104]. In patients with CNS autoimmune diseases, this polarized pattern is destructed leading to inappropriate leukocyte trafficking and disease progression [104].

4. Concluding remarks

Amongst $CD4^+$ T helper cell subsets are the most prominent "T_H17" cells, constituting a distinct lineage of T cells bridging the innate and adaptive immunity. These T_H17 cells are characterized by expression of

the transcription factors RORγt and RORα, the surface markers CCR4, CCR6 and IL-23R, the production of the potent proinflammatory molecules IL-17, IL-17F, IL-21, IL-22, IL-26 and G-CSF as well as the chemokine CCL20. Although intensive research approaches have been made, the mechanisms underlying the generation of these cells *in vivo* remain incomplete. In addition to the naturally occurring nT_H17 cells, differentiation of T_H17 from naïve $CD4^+$ requires co-stimulatory signals of CD28 and ICOS and the absence of T_H1 and T_H2 cytokines (IL-12 plus IFN-γ and IL-4, respectively) as well as the transcriptional factors T-bet and GATA3. At low levels of TGF-β, signaling through Smad pathway, IL-6 (as well as other proinflammatory cytokines IL-1β, TNF-α, or IL-21) favors differentiation of T_H17 cells. TGF-β (that induces Foxp3 expression and hence T_{reg} cell differentiation) is indispensable for expression of RORγt (RORc in human), which can be only activated in presence of proinflammatory cytokines IL-6 or IL-21 to start the full transcript machinery of T_H17. TGF-β, maintaining the cell's responsiveness to IL-6 signaling, activates STAT3, which, in turn, inhibits SOCS3 and thereby induces RORγt and RORα and subsequently IL-23R. IL-23 is required for survival and development of effecter T_H17 cells capable of producing IL-22. Despite its participation at host defense, the functional behavior of T_H17 comprises a detrimental role of this seemingly innocent lineage in induction of chronic inflammation triggered by latent probably "uncontrolled" infections, resulting in fostering allergic and autoimmune diseases. Despite the corroboration that both T_H17 and T_H1 are critical mediators of various autoimmune diseases, clinical trials targeting both subsets revealed mixed results [105], and call for more investigations to delineate the underlying mechanisms.

Take-home messages

- **T_H17 cells are potent inflammatory mediators that are, due to the ubiquitous expression of their effecter cytokine receptors, with manifold contributions to a growing list of disorders including asthma, allergy, psoriasis, RA, lupus and IBD.**
- **The differentiation factors of T_H17 include: (i) Cytokines: TGF-β plus IL-6/IL-21; (ii) Transcription factors: STAT3, IRF4, RORα and RORγt, and (iii) The survival/maturation factor IL-23.**
- **T_{reg} and T_H17 cells are reciprocally regulated via the induction of either the transcription factors Foxp3 or RORγt, respectively, together with low or high levels of IL-6.**
- **Effecter T_H17 cytokines include: (i) IL-17 (key factor for activation and recruitment of neutrophils); (ii) IL-21 (autologous activation of T_H17 and B cells and Ig production); and (iii) IL-22 (involved in skin inflammation and mucosal immunity, provokes wound healing and protects against hepatic injury).**
- **Therapeutic strategies aiming at attenuating but not abrogating IL-17 response or other associated factors may be promising to ameliorate inflammatory and autoimmune diseases.**

Acknowledgements

Authors acknowledge Engy Mohamed (ENT-Research Lab, Department of Otolaryngology, Head and Neck Surgery, Faculty of Medicine, University of Leipzig) for reading through the manuscript.

References

[1] Mosmann TR, Coffman RL. TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. Annu Rev Immunol 1989;7:145–73.
[2] Murphy KM, Reiner SL. The lineage decisions of helper T cells. Nat Rev Immunol 2002;2(12):933–44.
[3] Szabo SJ, Sullivan BM, Peng SL, Glimcher LH. Molecular mechanisms regulating Th1 immune responses. Annu Rev Immunol 2003;21:713–58.
[4] Egwuagu CE. STAT3 in CD4+ T helper cell differentiation and inflammatory diseases. Cytokine 2009;47(3):149–56.
[5] Furuzawa-Carballeda J, Vargas-Rojas MI, Cabral AR. Autoimmune inflammation from the Th17 perspective. Autoimmun Rev 2007;6(3):169–75.
[6] Rouvier E, Luciani MF, Mattei MG, Denizot F, Golstein P. CTLA-8, cloned from an activated T cell, bearing AU-rich messenger RNA instability sequences, and homologous to a herpesvirus saimiri gene. J Immunol 1993;150(12):5445–56.
[7] Yao Z, Painter SL, Fanslow WC, et al. Human IL-17: a novel cytokine derived from T cells. J Immunol 1995;155(12):5483–6.
[8] Shi Y, Ullrich SJ, Zhang J, et al. A novel cytokine receptor–ligand pair. J Biol Chem 2000;275(25):19167–76.
[9] Ouyang W, Kolls JK, Zheng Y. The biological functions of T helper 17 cell effector cytokines in inflammation. Immunity 2008;28(4):454–67.
[10] Pinto LG, Cunha TM, Vieira SM, et al. IL-17 mediates articular hypernociception in antigen-induced arthritis in mice. Pain 2010;148(2):247–56.
[11] Cascao R, Rosario HS, Souto-Carneiro MM, Fonseca JE. Neutrophils in rheumatoid arthritis: more than simple final effectors. Autoimmun Rev 2010;9(8):513–35.
[12] Liang SC, Tan XY, Luxenberg DP, et al. Interleukin (IL)-22 and IL-17 are coexpressed by Th17 cells and cooperatively enhance expression of antimicrobial peptides. J Exp Med 2006;203(10):2271–9.
[13] Zhang Z, Clarke TB, Weiser JN. Cellular effectors mediating Th17-dependent clearance of pneumococcal colonization in mice. J Clin Invest 2009;119(7):1899–909.
[14] Stockinger B, Veldhoen M, Martin B. Th17 T cells: linking innate and adaptive immunity. Semin Immunol 2007;19(6):353–61.
[15] Schulz SM, Kohler G, Holscher C, Iwakura Y, Alber G. IL-17A is produced by Th17, gammadelta T cells and other CD4-lymphocytes during infection with *Salmonella enterica* serovar Enteritidis and has a mild effect in bacterial clearance. Int Immunol 2008;20(9):1129–38.
[16] Passos ST, Silver JS, O'Hara AC, Sehy D, Stumhofer JS, Hunter CA. IL-6 promotes NK CELL Production of IL-17 during toxoplasmosis. J Immunol 2010;184(4):1776–83.
[17] Sellge G, Magalhaes JG, Konradt C, et al. Th17 cells are the dominant T cell subtype primed by *Shigella flexneri* mediating protective immunity. J Immunol 2010;184(4):2076–85.
[18] Riol-Blanco L, Lazarevic V, Awasthi A, et al. IL-23 receptor regulates unconventional IL-17-producing T cells that control bacterial infections. J Immunol 2010;184(4):1710–20.
[19] Tzartos JS, Friese MA, Craner MJ, et al. Interleukin-17 production in central nervous system-infiltrating T cells and glial cells is associated with active disease in multiple sclerosis. Am J Pathol 2008;172(1):146–55.
[20] Umemura M, Yahagi A, Hamada S, et al. IL-17-mediated regulation of innate and acquired immune response against pulmonary *Mycobacterium bovis* bacille Calmette-Guerin infection. J Immunol 2007;178(6):3786–96.
[21] Khader SA, Bell GK, Pearl JE, et al. IL-23 and IL-17 in the establishment of protective pulmonary CD4+ T cell responses after vaccination and during *Mycobacterium tuberculosis* challenge. Nat Immunol 2007;8(4):369–77.
[22] Qin S, Rottman JB, Myers P, et al. The chemokine receptors CXCR3 and CCR5 mark subsets of T cells associated with certain inflammatory reactions. J Clin Invest 1998;101(4):746–54.
[23] Yates CC, Whaley D, Kulasekeran P, et al. Delayed and deficient dermal maturation in mice lacking the CXCR3 ELR-negative CXC chemokine receptor. Am J Pathol 2007;171(2):484–95.
[24] Jiang D, Liang J, Hodge J, et al. Regulation of pulmonary fibrosis by chemokine receptor CXCR3. J Clin Invest 2004;114(2):291–9.
[25] Panzer U, Steinmetz OM, Paust HJ, et al. Chemokine receptor CXCR3 mediates T cell recruitment and tissue injury in nephrotoxic nephritis in mice. J Am Soc Nephrol 2007;18(7):2071–84.
[26] Hancock WW, Lu B, Gao W, et al. Requirement of the chemokine receptor CXCR3 for acute allograft rejection. J Exp Med 2000;192(10):1515–20.
[27] Glatigny S, Blaton MA, Mencher SK, et al. Treatment of collagen-induced arthritis by Natura-alpha via regulation of Th-1/Th-17 responses. Eur J Immunol 2010;40(2):460–9.
[28] Hruz P, Dann SM, Eckmann L. STAT3 and its activators in intestinal defense and mucosal homeostasis. Curr Opin Gastroenterol 2009.
[29] Munoz M, Heimesaat MM, Danker K, et al. Interleukin (IL)-23 mediates *Toxoplasma gondii*-induced immunopathology in the gut via matrixmetalloproteinase-2 and IL-22 but independent of IL-17. J Exp Med 2009;206(13):3047–59.
[30] Cella M, Fuchs A, Vermi W, et al. A human natural killer cell subset provides an innate source of IL-22 for mucosal immunity. Nature 2009;457(7230):722–5.
[31] Eyerich S, Eyerich K, Pennino D, et al. Th22 cells represent a distinct human T cell subset involved in epidermal immunity and remodeling. J Clin Invest 2009;119(12):3573–85.
[32] Duhen T, Geiger R, Jarrossay D, Lanzavecchia A, Sallusto F. Production of interleukin 22 but not interleukin 17 by a subset of human skin-homing memory T cells. Nat Immunol 2009;10(8):857–63.
[33] Chang H, Hanawa H, Liu H, et al. Hydrodynamic-based delivery of an interleukin-22-Ig fusion gene ameliorates experimental autoimmune myocarditis in rats. J Immunol 2006;177(6):3635–43.
[34] Zenewicz LA, Yancopoulos GD, Valenzuela DM, Murphy AJ, Stevens S, Flavell RA. Innate and adaptive interleukin-22 protects mice from inflammatory bowel disease. Immunity 2008;29(6):947–57.
[35] Zenewicz LA, Yancopoulos GD, Valenzuela DM, Murphy AJ, Karow M, Flavell RA. Interleukin-22 but not interleukin-17 provides protection to hepatocytes during acute liver inflammation. Immunity 2007;27(4):647–59.
[36] Schulz SM, Kohler G, Schutze N, et al. Protective immunity to systemic infection with attenuated *Salmonella enterica* serovar enteritidis in the absence of IL-12 is associated with IL-23-dependent IL-22, but not IL-17. J Immunol 2008;181(11):7891–901.

[37] Yamamoto M, Yoshizaki K, Kishimoto T, Ito H. IL-6 is required for the development of Th1 cell-mediated murine colitis. J Immunol 2000;164(9):4878–82.
[38] Diehl S, Rincon M. The two faces of IL-6 on Th1/Th2 differentiation. Mol Immunol 2002;39(9):531–6.
[39] Goodman WA, Levine AD, Massari JV, Sugiyama H, McCormick TS, Cooper KD. IL-6 signaling in psoriasis prevents immune suppression by regulatory T cells. J Immunol 2009;183(5):3170–6.
[40] Veldhoen M, Hocking RJ, Atkins CJ, Locksley RM, Stockinger B. TGFbeta in the context of an inflammatory cytokine milieu supports de novo differentiation of IL-17-producing T cells. Immunity 2006;24(2):179–89.
[41] Serada S, Fujimoto M, Mihara M, et al. IL-6 blockade inhibits the induction of myelin antigen-specific Th17 cells and Th1 cells in experimental autoimmune encephalomyelitis. Proc Natl Acad Sci USA 2008;105(26):9041–6.
[42] Samoilova EB, Horton JL, Hilliard B, Liu TS, Chen Y. IL-6-deficient mice are resistant to experimental autoimmune encephalomyelitis: roles of IL-6 in the activation and differentiation of autoreactive T cells. J Immunol 1998;161(12):6480–6.
[43] Cross JT, Benton HP. The roles of interleukin-6 and interleukin-10 in B cell hyperactivity in systemic lupus erythematosus. Inflamm Res 1999;48(5):255–61.
[44] Boe A, Baiocchi M, Carbonatto M, Papoian R, Serlupi-Crescenzi O. Interleukin 6 knock-out mice are resistant to antigen-induced experimental arthritis. Cytokine 1999;11(12):1057–64.
[45] Hirano T. Interleukin 6 and its receptor: ten years later. Int Rev Immunol 1998;16(3–4):249–84.
[46] Jones GW, McLoughlin RM, Hammond VJ, et al. Loss of CD4+ T cell IL-6R expression during inflammation underlines a role for IL-6 trans signaling in the local maintenance of Th17 cells. J Immunol 2010;184(4):2130–9.
[47] Tsantikos E, Oracki SA, Quilici C, Anderson GP, Tarlinton DM, Hibbs ML. Autoimmune disease in Lyn-deficient mice is dependent on an inflammatory environment established by IL-6. J Immunol 2010;184(3):1348–60.
[48] Woo P, Wilkinson N, Prieur AM, et al. Open label phase II trial of single, ascending doses of MRA in Caucasian children with severe systemic juvenile idiopathic arthritis: proof of principle of the efficacy of IL-6 receptor blockade in this type of arthritis and demonstration of prolonged clinical improvement. Arthritis Res Ther 2005;7(6):R1281–8.
[49] Chen X, Das R, Komorowski R, et al. Blockade of interleukin-6 signaling augments regulatory T-cell reconstitution and attenuates the severity of graft-versus-host disease. Blood 2009;114(4):891–900.
[50] Tanaka S, Yoshimoto T, Naka T, et al. Natural occurring IL-17 producing T cells regulate the initial phase of neutrophil mediated airway responses. J Immunol 2009;183(11):7523–30.
[51] Corthay A. How do regulatory T cells work? Scand J Immunol 2009;70(4):326–36.
[52] Soroosh P, Doherty TA. Th9 and allergic disease. Immunology 2009;127(4):450–8.
[53] Veldhoen M, Stockinger B. TGFbeta1, a "Jack of all trades": the link with pro-inflammatory IL-17-producing T cells. Trends Immunol 2006;27(8):358–61.
[54] Carrier Y, Yuan J, Kuchroo VK, Weiner HL. Th3 cells in peripheral tolerance. I. Induction of Foxp3-positive regulatory T cells by Th3 cells derived from TGF-beta T cell-transgenic mice. J Immunol 2007;178(1):179–85.
[55] Rouse BT. Regulatory T cells in health and disease. J Int Med 2007;262(1):78–95.
[56] Bettelli E, Dastrange M, Oukka M. Foxp3 interacts with nuclear factor of activated T cells and NF-{kappa}B to repress cytokine gene expression and effector functions of T helper cells. PNAS 2005;102(14):5138–43.
[57] Gerli R, Nocentini G, Alunno A, et al. Identification of regulatory T cells in systemic lupus erythematosus. Autoimmun Rev 2009;8(5):426–30.
[58] Zhang L, Yang XQ, Cheng J, Hui RS, Gao TW. Increased Th17 cells are accompanied by FoxP3(+) Treg cell accumulation and correlated with psoriasis disease severity. Clin Immunol 2009;135(1):108–17.
[59] Li MO, Wan YY, Flavell RA. T cell-produced transforming growth factor-beta1 controls T cell tolerance and regulates Th1- and Th17-cell differentiation. Immunity 2007;26(5):579–91.
[60] Lee YK, Turner H, Maynard CL, et al. Late developmental plasticity in the T helper 17 lineage. Immunity 2009;30(1):92–107.
[61] Do JS, Fink PJ, Li L, et al. Cutting edge: spontaneous development of IL-17-producing gamma}{delta T cells in the thymus occurs via a TGF-{beta}1-dependent mechanism. J Immunol 2010;184(4):1675–9.
[62] Weaver Jr DJ, Reis ES, Pandey MK, et al. C5a receptor-deficient dendritic cells promote induction of Treg and Th17. Eur J Immunol 2009;40(3):710–21.
[63] Tan AH, Lam KP. Pharmacologic inhibition of MEK-ERK signaling enhances Th17 differentiation. J Immunol 2010;184(4):1849–57.
[64] Voo KS, Wang YH, Santori FR, et al. Identification of IL-17-producing FOXP3+ regulatory T cells in humans. Proc Natl Acad Sci USA 2009;106(12):4793–8.
[65] Parrish-Novak J, Dillon SR, Nelson A, et al. Interleukin 21 and its receptor are involved in NK cell expansion and regulation of lymphocyte function. Nature 2000;408(6808):57–63.
[66] Liu Z, Yang L, Cui Y, et al. Il-21 enhances NK cell activation and cytolytic activity and induces Th17 cell differentiation in inflammatory bowel disease. Inflamm Bowel Dis 2009;15(8):1133–44.
[67] Niu X, He D, Zhang X, et al. IL-21 regulates Th17 cells in rheumatoid arthritis. Hum Immunol 2010;71(4):331–41.
[68] Korn T, Bettelli E, Gao W, et al. IL-21 initiates an alternative pathway to induce proinflammatory TH17 cells. Nature 2007;448(7152):484–7.
[69] Yang L, Anderson DE, Baecher-Allan C, et al. IL-21 and TGF-[bgr] are required for differentiation of human TH17 cells. Nature 2008;454(7202):350–2.
[70] Fantini MC, Rizzo A, Fina D, et al. IL-21 regulates experimental colitis by modulating the balance between Treg and Th17 cells. Eur J Immunol 2007;37(11):3155–63.
[71] Manel N, Unutmaz D, Littman DR. The differentiation of human T(H)-17 cells requires transforming growth factor-beta and induction of the nuclear receptor RORgammat. Nat Immunol 2008;9(6):641–9.
[72] Li J, Shen W, Kong K, Liu Z. Interleukin-21 induces T-cell activation and proinflammatory cytokine secretion in rheumatoid arthritis. Scand J Immunol 2006;64(5):515–22.
[73] Rovedatti L, Kudo T, Biancheri P, et al. Differential regulation of interleukin 17 and interferon gamma production in inflammatory bowel disease. Gut 2009;58(12):1629–36.
[74] Borte S, Pan-Hammarstrom Q, Liu C, et al. Interleukin-21 restores immunoglobulin production ex vivo in patients with common variable immunodeficiency and selective IgA deficiency. Blood 2009;114(19):4089–98.
[75] Hemdan NY, Emmrich F, Faber S, Sack U. Alterations of Th1/Th2 reactivity by heavy metals. Possible consequences include induction of autoimmune diseases. Ann NY Acad Sci 2007;1109(1):129–37.
[76] Hemdan NY. The role of interleukin-12 in the heavy metal-elicited immunomodulation: relevance of various evaluation methods. J Occup Med Toxicol 2008;3:25.
[77] Mangan PR, Harrington LE, O'Quinn DB, et al. Transforming growth factor-{beta} induces development of the TH17 lineage. Nature 2006;441(7090):231–4.
[78] McGeachy MJ, Chen Y, Tato CM, et al. The interleukin 23 receptor is essential for the terminal differentiation of interleukin 17-producing effector T helper cells in vivo. Nat Immunol 2009;10(3):314–24.
[79] Godinez I, Raffatellu M, Chu H, et al. Interleukin-23 orchestrates mucosal responses to *Salmonella enterica* serotype Typhimurium in the intestine. Infect Immun 2009;77(1):387–98.
[80] Murphy CA, Langrish CL, Chen Y, et al. Divergent pro- and antiinflammatory roles for IL-23 and IL-12 in joint autoimmune inflammation. J Exp Med 2003;198(12):1951–7.
[81] Lee YK, Turner H, Maynard CL, et al. Late developmental plasticity in the T helper 17 lineage. Immunity 2009;30(1):92–107.
[82] Cua DJ, Sherlock J, Chen Y, et al. Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain. Nature 2003;421(6924):744–8.
[83] McGeachy MJ, Bak-Jensen KS, Chen Y, et al. TGF-[beta] and IL-6 drive the production of IL-17 and IL-10 by T cells and restrain TH-17 cell-mediated pathology. Nat Immunol 2007;8(12):1390–7.
[84] Gosselin D, Rivest S. Role of IL-1 and TNF in the brain: twenty years of progress on a Dr. Jekyll/Mr. Hyde duality of the innate immune system. Brain Behav Immun 2007;21(3):281–9.
[85] Sutton C, Brereton C, Keogh B, Mills KHG, Lavelle EC. A crucial role for interleukin (IL)-1 in the induction of IL-17-producing T cells that mediate autoimmune encephalomyelitis. J Exp Med 2006;203(7):1685–91.
[86] Acosta-Rodriguez EV, Napolitani G, Lanzavecchia A, Sallusto F. Interleukins 1 [beta] and 6 but not transforming growth factor-[beta] are essential for the differentiation of interleukin 17-producing human T helper cells. Nat Immunol 2007;8(9):942–9.
[87] Lee WW, Kang SW, Choi J, et al. Regulating human Th17 cells via differential expression of IL-1 receptor. Blood 2010;115:530–40.
[88] Yang XO, Panopoulos AD, Nurieva R, Chang SH, Wang D. STAT3 regulates cytokine-mediated generation of inflammatory helper T cells. J Biol Chem 2007;282(13):9358–63.
[89] Liu X, Lee YS, Yu CR, Egwuagu CE. Loss of STAT3 in CD4+ T cells prevents development of experimental autoimmune diseases. J Immunol 2008;180(9):6070–6.
[90] Harris TJ, Grosso JF, Yen HR, Xin H, Kortylewski M, Cutting edge: an in vivo requirement for STAT3 signaling in Th17 development and Th17-dependent autoimmunity. J Immunol 2007;179:4313.
[91] Lafdil F, Wang H, Park O, et al. Myeloid STAT3 inhibits T cell-mediated hepatitis by regulating T helper 1 cytokine and interleukin-17 production. Gastroenterology 2009;137(6):2125–35.
[92] Qin H, Wang L, Feng T, et al. TGF-{beta} promotes Th17 cell development through inhibition of SOCS3. J Immunol 2009;183(1):97–105.
[93] Zheng Y, Chaudhry A, Kas A, et al. Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control T(H)2 responses. Nature 2009;458(7236):351–6.
[94] Koch MA, Tucker-Heard G, Perdue NR, Killebrew JR, Urdahl KB, Campbell DJ. The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation. Nat Immunol 2009;10(6):595–602.
[95] Chaudhry A, Rudra D, Treuting P, et al. CD4+ regulatory T cells control TH17 responses in a Stat3-dependent manner. Science 2009;326(5955):986–91.
[96] Volpe E, Touzot M, Servant N, et al. Multiparametric analysis of cytokine-driven human Th17 differentiation reveals a differential regulation of IL-17 and IL-22 production. Blood 2009;114(17):3610–4.
[97] Volpe E, Servant N, Zollinger R, et al. A critical function for transforming growth factor-beta, interleukin 23 and proinflammatory cytokines in driving and modulating human T(H)-17 responses. Nat Immunol 2008;9(6):650–7.
[98] Sofi MH, Liu Z, Zhu L, Yu Q, Kaplan MH, Chang CH. Regulation of IL-17 expression by the developmental pathway of CD4 T cells in the thymus. Mol Immunol 2010;47(6):1262–8.
[99] Pflanz S, Timans JC, Cheung J, et al. IL-27, a heterodimeric cytokine composed of EBI3 and p28 protein, induces proliferation of naive CD4(+) T cells. Immunity 2002;16(6):779–90.
[100] Stumhofer JS, Laurence A, Wilson EH, et al. Interleukin 27 negatively regulates the development of interleukin 17-producing T helper cells during chronic inflammation of the central nervous system. Nat Immunol 2006;7(9):937–45.
[101] Shi G, Ramaswamy M, Vistica BP, et al. Unlike Th1, Th17 cells mediate sustained autoimmune inflammation and are highly resistant to restimulation-induced cell death. J Immunol 2009;183(11):7547–56.

219

[102] Hsu HC, Yang P, Wang J, et al. Interleukin 17-producing T helper cells and interleukin 17 orchestrate autoreactive germinal center development in autoimmune BXD2 mice. Nat Immunol 2008;9(2):166–75.
[103] Stumm RK, Rummel J, Junker V, et al. A dual role for the SDF-1/CXCR4 chemokine receptor system in adult brain: isoform-selective regulation of SDF-1 expression modulates CXCR4-dependent neuronal plasticity and cerebral leukocyte recruitment after focal ischemia. J Neurosci 2002;22(14):5865–78.
[104] McCandless EE, Piccio L, Woerner BM, et al. Pathological expression of CXCL12 at the blood–brain barrier correlates with severity of multiple sclerosis. Am J Pathol 2008;172(3):799–808.
[105] Steinman L. Mixed results with modulation of TH-17 cells in human autoimmune diseases. Nat Immunol 2010;11(1):41–4.
[106] Brucklacher-Waldert V, Stuerner K, Kolster M, Wolthausen J, Tolosa E. Phenotypical and functional characterization of T helper 17 cells in multiple sclerosis. Brain 2009;132(Pt 12):3329–41.
[107] Edwards LJ, Robins RA, Constantinescu CS. Th17/Th1 phenotype in demyelinating disease. Cytokine 2010;50(1):19–23.
[108] Park H, Li Z, Yang XO, et al. A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. Nat Immunol 2005;6(11):1133–41.
[109] Yamazaki T, Yang XO, Chung Y, et al. CCR6 regulates the migration of inflammatory and regulatory T cells. J Immunol 2008;181(12):8391–401.
[110] Komiyama Y, Nakae S, Matsuki T, et al. IL-17 plays an important role in the development of experimental autoimmune encephalomyelitis. J Immunol 2006;177(1):566–73.
[111] Qian G, Qin X, Zang YQ, et al. High doses of alpha-galactosylceramide potentiate experimental autoimmune encephalomyelitis by directly enhancing Th17 response. Cell Res 2010.
[112] van der FL, Mourits S, Voerman JS, et al. Imiquimod-induced psoriasis-like skin inflammation in mice is mediated via the IL-23/IL-17 axis. J Immunol 2009;182(9):5836–45.
[113] Lowes MA, Kikuchi T, Fuentes-Duculan J, et al. Psoriasis vulgaris lesions contain discrete populations of Th1 and Th17 T cells. J Invest Dermatol 2008;128(5):1207–11.
[114] Krueger GG, Langley RG, Leonardi C, et al. A human interleukin-12/23 monoclonal antibody for the treatment of psoriasis. N Engl J Med 2007;356(6):580–92.
[115] Fujishima S, Watanabe H, Kawaguchi M, et al. Involvement of IL-17F via the induction of IL-6 in psoriasis. Arch Dermatol Res 2010;302(7):499–505.
[116] Kirkham BW, Lassere MN, Edmonds JP, et al. Synovial membrane cytokine expression is predictive of joint damage progression in rheumatoid arthritis: a two-year prospective study (the DAMAGE study cohort). Arthritis Rheum 2006;54(4):1122–31.
[117] Infante-Duarte C, Horton HF, Byrne MC, Kamradt T. Microbial lipopeptides induce the production of IL-17 in Th cells. J Immunol 2000;165(11):6107–15.
[118] von Delwig A, Locke J, Robinson JH, Ng WF. Response of Th17 cells to a citrullinated arthritogenic aggrecan peptide in patients with rheumatoid arthritis. Arthritis Rheum 2010;62(1):143–9.
[119] Jacobs JP, Wu HJ, Benoist C, Mathis D. IL-17-producing T cells can augment autoantibody-induced arthritis. Proc Natl Acad Sci USA 2009;106(51):21789–94.
[120] Nakae S, Nambu A, Sudo K, Iwakura Y. Suppression of immune induction of collagen-induced arthritis in IL-17-deficient mice. J Immunol 2003;171(11):6173–7.
[121] Lubberts E, Koenders MI, Oppers-Walgreen B, et al. Treatment with a neutralizing anti-murine interleukin-17 antibody after the onset of collagen-induced arthritis reduces joint inflammation, cartilage destruction, and bone erosion. Arthritis Rheum 2004;50(2):650–9.
[122] Doodes PD, Cao Y, Hamel KM, et al. IFN-{gamma} regulates the requirement for IL-17 in proteoglycan-induced arthritis. J Immunol 2010;184(3):1552–9.
[123] Emamaullee JA, Davis J, Merani S, et al. Inhibition of Th17 cells regulates autoimmune diabetes in NOD mice. Diabetes 2009;58(6):1302–11.
[124] Nogueira E, Hamour S, Sawant D, et al. Serum IL-17 and IL-23 levels and autoantigen-specific Th17 cells are elevated in patients with ANCA-associated vasculitis. Nephrol Dial Transplant 2010;25(7):2209–17.
[125] Katsifis GE, Rekka S, Moutsopoulos NM, Pillemer S, Wahl SM. Systemic and local interleukin-17 and linked cytokines associated with Sjogren's syndrome immunopathogenesis. Am J Pathol 2009;175(3):1167–77.
[126] De Paiva CS, Hwang CS, Pitcher III JD, et al. Age-related T-cell cytokine profile parallels corneal disease severity in Sjogren's syndrome-like keratoconjunctivitis sicca in CD25KO mice. Rheumatology 2010;49(2):246–58.
[127] Larsen JM, Bonefeld CM, Poulsen SS, Geisler C, Skov L. IL-23 and T(H)17-mediated inflammation in human allergic contact dermatitis. J Allergy Clin Immunol 2009;123(2):486–92.
[128] Wilson RH, Whitehead GS, Nakano H, Free ME, Kolls JK, Cook DN. Allergic sensitization through the airway primes Th17-dependent neutrophilia and airway hyperresponsiveness. Am J Respir Crit Care Med 2009;180(8):720–30.
[129] Oboki K, Ohno T, Saito H, Nakae S. Th17 and allergy. Allergol Int 2008;57(2):121–34.
[130] Figueroa-Vega N, Alfonso-Perez M, Benedicto I, Sanchez-Madrid F, Gonzalez-Amaro R, Marazuela M. Increased circulating pro-inflammatory cytokines and Th17 lymphocytes in Hashimoto's thyroiditis. J Clin Endocrinol Metab 2010;95(2):953–62.
[131] Zhu X, Ma D, Zhang J, et al. Elevated interleukin-21 correlated to Th17 and Th1 cells in patients with immune thrombocytopenia. J Clin Immunol 2009;30(2):253–9.

The interferon-continuum: a common ring in autoimmune diseases

One of the challenges of the most recent researches is to find common pathways between autoimmune disorders. Within the "mosaic" of autoimmunity many dowels are shared by different diseases and at least some of these plugs may represent key targets for therapy. Interferons (IFNs) are proteins made and released by lymphocytes in response to the presence of pathogens that besides this protective functions they may participate in the development of autoimmune diseases. Type-I IFN has been associated with systemic lupus erythematosus, where plasmocytoid dendritic cells (PDCs) represent the main source of this molecule leading to the production of autoantibodies and immune complexes. The evidence that there is a spectrum IFN-mediated diseases is looming up and also systemic sclerosis (SSc) can be part of this IFN-mediated continuum. Assassi and coworkers raised such interesting hypothesis (Assassi et al. Systemic sclerosis and lupus: points in an interferon-mediated continuum. Arthritis Rheum. 2010;62:589-98). In this recently published paper SSc has demonstrated to have several similarities with SLE at the level of transcript profile of common genes. A modular analysis framework was used to achieve these results, that allowed to define that the "IFN-signature" is a quantitative trait that can be expressed as an IFN score. Furthermore, SSc patients could be sub-grouped on the basis the strength of this IFN signature, with patients with anti-U1 RNP or anti-topoisomerase antibodies more likely to have this feature. All this evidences can be merged together with the recent results of a clinical trial of subcutaneous IFNa in early SSc, that showed a worsening of lung function and a trend towards skin deterioration, suggesting that SSc belongs to the family of IFN-mediated disease as well as SLE, with subgroups of patients showing "lupus-like" high IFN-inducible gene expression pattern correlating with the presence of specific autoantibodies. More targeted therapeutic interventions in SSc are now awaited.

Other macro-molecules in immunochemistry

There are numerous other macro-molecules involved in the immune system. For example, one of the commercial companies, PROSPEC (http://www.prospecbio.com), lists twenty one classes of cytokines, including the commonly known ones such as interferons, interleukins, tumor necrosis factors, etc. In the case of interferons, there are many different types designated as α, β, γ, τ, ω, etc.

Another groups of macro-molecules closely associated with antigen-antibody interactions are the members of the complement system, sometimes denoted as MHC class III molecules. They consist of the following proteins:

B

C1: C1q, C1r, C1s

C2; C2a, C2b

C3: C3a, C3b

C4: C4a, C4b

C5: C5a, C5b

C6

C7

C8

C9

D

These components form complexes through a cascade of enzymatic reactions. The final complex is C5b6789 also known as the membrane attack complex (MAC), These complexes are then inserted into cell membrane to punch holes.

There are also membrane bound macro-molecules such as Fc-receptors, various molecules associated with membrane bound antibodies, etc. Some of these molecules have been renamed as CD molecules. More of them are being discovered constantly.

PROSPEC
Protein Specialists

Search by name or Catalog #:

Home About How To Order Shipping Products Services Customer Service Contact 0 items | 0 $

VERIFIED & SECURED

Protein Catalog

- Cytokines
- Growth Factors
- Chemokines
- CD Antigens
- Neurotrophins
- Hormones
- Enzymes
- Viral Antigens
- Recombinant Proteins
- Natural Proteins
- Monoclonal Antibodies
- Polyclonal Antibodies

Cytokines

Adiponectin (12 products) View Products	**Angiopoietin** (3 products) View Products	**Apolipoprotein** (12 products) View Products
B type Natriuretic Peptide (3 products) View Products	**B-Cell Activating Factor** (3 products) View Products	**Beta Defensin** (4 products) View Products
Bone Morphogenetic Protein (17 products) View Products	**EBI3** (3 products) View Products	**Endoglin** (3 products) View Products
Flt3 Ligand (4 products) View Products	**Hedgehog Protein** (5 products) View Products	**Interferon** (21 products) View Products
Interleukin (119 products) View Products	**Leukemia Inhibitory Factor** (3 products) View Products	**Other** (17 products) View Products
Resistin (10 products) View Products	**Serum Amyloid A** (3 products) View Products	**Thrombopoietin** (3 products) View Products
Trefoil Factor (6 products) View Products	**Tumor Necrosis Factor** (16 products) View Products	**Visfatin** (3 products) View Products

About Cytokines:

Cytokines/immunocytokines, were initially used to separate the immunomodulatory proteins, also called immunotransmitters, from other growth factors that modulate proliferation and bioactivities of non-immune cells. Some cytokines are produced by a rather limited number of different cell types while others are produced by almost the entire spectrum of known cell types.

Recombinant cytokines are in clinical use, and continues attempts are made to develop hybrid molecules from cytokines. 1 must be aware of the fact that current knowledge is still limited. Cytokines are powerful two-edged weapons that can trigger a cascade of reactions, and may show activities that often go beyond the single highly specific property that it is hoped they possess. New factors are being discovered constantly and they extend our knowledge about the Cytokine network.

The understanding of the biological mechanisms governing cytokine actions are an important contribution to medical knowledge. The biochemistry and molecular biology of cytokine actions explain some well-known and sometimes also some of the more obscure clinical aspects of diseases. Knowledge that cytokines create regulatory hierarchies and provide independent and/or interrelated regulatory mechanisms that can confer distinct and interactive developmental functions lays a solid, albeit rather complicated foundation, for current and future clinical experiences.

The concept of "1 producer cell - 1 cytokine - 1 target cell" has been falsified for practically every cytokine investigated more closely. To classify these factors based on their producer or target cells is therefore also problematic.

Classifications based upon identical or shared biological activities of cytokines especially with broad definitions is also

problematic for example: BCDF (B-cell differentiation factors), BCGF (B-cell growth factors), Motogenic cytokines, Chemotactic cytokines (Chemokines), CSF (colony stimulating factors), angiogenesis factors, or TRF (T-cell replacing factors). Designations such as HBGF group (heparin binding growth factors) take into account biochemical shared properties by a variety of cytokines which also problematic.

The term cytokine is used today as a generic name for a diverse group which includes proteins and peptides that act in nano-picomolar concentrations as humoral regulators and modulate the functional activities of individual cells & tissues. Cytokines also mediate interactions between cells & regulate processes taking place in the extracellular environment. Many growth factors and cytokines act as cellular survival factors by preventing programmed cell death.

Cytokines resemble hormones in their biological activity & systemic level, for example, inflammation, systemic inflammatory response syndrome, acute phase reaction, wound healing, and the neuroimmune network.

Cytokines act on a wider spectrum of target cells than hormones. The major feature distinguishing cytokines from hormones is the fact that cytokines are not produced by specialized cells organized in specialized glands. Cytokines are secreted proteins which means that their expression sites does not predict where they exert their biological function.

Several cytokines primary structure was found to be identical with enzymes. Cytokines do not possess enzymatic activities although there is a growing list of exceptions. Other cytokines require proteolytic activation.

Cytokines include Interleukins, Lymphokines, Monokines, Interferons (IFN), colony stimulating factors (CSF), Chemokines and a variety of other proteins.

Type-1 cytokines are cytokines produced by Th1 T-helper cells while Type-2 cytokines are those produced by Th2 T-helper cells. Type-1 cytokines include IL-2 (IL2), IFN-gamma (IFN-G), IL-12 (IL12) & TNF-beta (TNF-b), while Type 2 cytokines include IL-4 (IL4), IL-5 (IL5), IL-6 (IL6), IL-10(IL10), and IL-13 (IL13).

Comparison of cytokine sequences shows that primate cytokines (non-human) are closely related. An example: IL-1 alpha (IL1a), IL-1 beta (IL1b), IL-2 (IL2), IL-4 (IL4), IL-5 (IL5), IL-6 (IL6) , IL8 (IL-8), IL-10 (IL10), IL-12 (IL12), IL-15 (IL15), IFN-alpha (IFNA), IFN-gamma (IFN-G), and TNF-alpha (TNFA) which share 93% to 99 % homology at the protein & nucleic acid level with the human sequences.

Cytokines can also be classified into family groups according to the types of secondary and tertiary structure. An example: IL-6 (IL6), IL-11 (IL11), CNTF (C-NTF), LIF, OSM (Oncostatin-M), EPO (Erythropoietin), G-CSF (GCSF), GH (Growth Hormone), PRL (Prolactin), IL-10 (IL10), IFN-alpha (IFN-A), IFN-beta (IFN-B) form long chain 4 helix bundles. IL-2 (IL2), IL-4 (IL4), IL-7 (IL7), IL-9 (IL9), IL-13 (IL13), IL-3 (IL3), IL-5 (IL5), GM-CSF (GMCSF), M-CSF (MCSF), SCF, IFN-gamma (IFNG) form short chain 4 helix bundles. Beta-trefoil structures are formed by IL1-alpha (IL1A), IL1-beta (IL1B), aFGF (FGF-acidic), bFGF (FGF-basic), INT-2 (INT2), KGF (FGF7). EGF, TGF-alpha (TGF-A), Betacelluin (BTC), SCDGF, Amphireguin, HB-EGF, form EGF-like antiparallel beta-sheets.

Many cytokines are secreted by cells using secretory pathways and therefore are considered glycoproteins. Most genes encoding cytokines give rise to variety of cytokines by means of alternative splicing, yielding molecules with slightly different but biologically significant bioactivities. Usually the expression patterns of different forms of cytokines or of members of a cytokine family are overlapping only partially, suggesting a specific role for each factor.

Membrane-bound cytokines have been are associated with the extracellular matrix. The switching between soluble and membrane forms of cytokines is an important regulatory event. In some cases membrane forms of a cytokine have been found to be indispensable for normal development, with soluble forms being unable to entirely substitute for them.

Numerous cytokines are not stored inside cells though TGF-beta (TGF-b) and PDGF (P-DGF) are stored in platelets or TNF-alpha (TNF-A) and IL-8 (IL8) are found in human skin mast cells. Most of the cytokine's expression is regulated tightly at practically all levels. The factors are usually produced only by cells after cell activation in response to an induction signal. The production and secretion of cytokines and growth factors frequently is context dependent, i.e., their expression is influenced by individual signals received but also by the balance of signals received through one or more receptors (which themselves may be subject to inducible/repressible expression).

Cytokine's expression is regulated at the transcription level, translation level, and protein synthesis. The expression of cytokines also seems to be regulated differentially, depending on cell type and developmental age. Secretion or release from the producer cells is a regulated process. Once released, their behavior in the circulation may be regulated by soluble receptors and specific or unspecific binding proteins. Regulation also is at work at the receptor level on target cells and at the level of signaling pathways governing alterations in the behavior of responder cells.

Numerous cytokines are pleiotropic effectors showing multiple biological activities. Multiple cytokines have overlapping activities therefore a single cell frequently interacts with multiple cytokines with seemingly identical responses (cross-talk). A possible consequence of this functional overlap is the observation that 1 factor may frequently functionally replace another factor altogether or at least partially compensate for the lack of another factor. Since most cytokines have ubiquitous biological activities, their physiologic significance as normal regulators of physiology is often difficult to assess.

Gene function studies in experimental transgenic knock-out animals in which a cytokine gene has been functionally inactivated by gene targeting are very important in cytokine-research since, unlike in vitro studies, they provide information about the true in vivo functions of a given cytokine by highlighting the effects of their absence. In many instances these studies have shown that null mutations of particular cytokine genes do not have the effects in vivo expected from their activities in vitro.

Cytokines show stimulating or inhibitory activities and synergize or antagonize the actions of other factors. 1 sole Cytokine elicits reactions under certain circumstances that are the reverse of those shown under other circumstances. The type, the duration, and also the extent of cellular activities induced by a particular cytokine can be influenced considerably by the

223

micro-environment of a cell, depending, for example, on the growth state of the cells (sparse or confluent), the type of neighboring cells, cytokine concentrations, the combination of other cytokines present at the same time, and even on the temporal sequence of several cytokines acting on the same cell. Under such circumstances combinatorial effects thus allow a single cytokine to transmit diverse signals to different subsets of cells.

Although some cytokines are known to share at least some biological effects, the observations that single cells usually show different patterns of gene expression in response to different cytokines can be taken as evidence for the existence of cytokine receptor-specific signal transduction pathways. Shared and different transcriptional activators that transduce a signal from a cytokine receptor to a transcription regulatory element of DNA are involved in these processes such as STAT proteins.

Basic FGF (bFGF) is a strong mitogen for fibroblasts at low concentrations and a chemoattractant at high concentrations. FGFb (FGF-b) has been shown also to be a biphasic regulator of human hepatoblastoma-derived HepG2 cells, depending upon the concentration. Interferon-gamma (IFN-gamma) can stimulate the proliferation of B-cells prestimulated with Anti-IgM, and inhibits the activities of the same cells induced by IL-4 (IL4). On the other hand, IL-4 (Interleukin-4) activates B-cells and promotes their proliferation while inhibiting the effects induced by IL2 in the same cells. The activity of at least two cytokines such as IL1-A (IL1A) & IL1-B (IL1B) is regulated by an endogenous receptor antagonist, the IL1 receptor anagonist (IL1TA). Cytokines, such as TNFA (TNF-A), IFN-gamma (IFN-G), IL-2 (IL2) & IL-4 (IL4), are inhibited by soluble receptors. Cytokines including IL-10 (IL10) and TGF-beta (TGF-B), inhibit other cytokines.

Early Cytokines preactivate cells so that they then can respond to late-acting cytokines. Cytokines induce the synthesis of novel gene products once they have bound to their corresponding. Several of the novel products are themselves cytokines. In addition, there are a variety of biological response modifiers that function as Anti-cytokines.

Cytokine mediators swiftly remote areas of a multicellular organism & multiple target cells can be degraded quickly, One can assume that cytokines play a pivotal role in all sorts of cell-to-cell communication processes although many of the mechanisms of their actions have not yet been elucidated in full detail.

Thorough examination of the physiological effects of the expression of cytokines in complex organisms has shown that these mediators are involved in all systemic reactions of an organism, including the important processes as regulation of immune responses, for example: BCDF(B-cell growth and differentiation factors), BCGF (B-cell growth factors) TRF (T-cell replacing factors), isotype switching, inflammatory processes, hematopoiesis, and wound healing.

Embryogenesis and organ development inivolves important mediators called Cytokines. Their activities in these processes may differ from those observed postnatally. Cytokines play a key role in neuroimmunological, neuroendocrinological, and neuroregulatory processes. Cytokines also regulate cell cycle, differentiation, migration, cell survival & cell death, and cell transformation. Viral infectious agents exploit the cytokine repertoire of organisms to evade immune responses of the host. Virus-encoded factors affect the activities of cytokines in at least four different ways: by inhibiting the synthesis and release of cytokines from infected cells; by interfering with the interaction between cytokines and their receptors; by inhibiting signal transmission pathways of cytokines; and by synthesizing virus-encoded cytokines that antagonize the effects of host cytokines mediating antiviral processes. Bacteria and micro-organisms also appear to produce substances with activities resembling those of cytokines and which they utilize to subvert host responses.

Cytokines are rarely related among eachother in their primary sequences. Some appear to have common 3 dimensional features and some of them can be grouped into families. An example is the TNF ligand superfamily members (with the exception of LT-alpha) are type 2 membrane glycoproteins (N-terminus inside) with homology to TNF in the extracellular domain (overall homologies, 20 %). The HBNF family includes members of the group of fibroblast growth factors. The chemokine group which contain diverse factors also have conserved sequence features. Analysis of crystal structures of several cytokines with very little sequence homology has revealed a common overall topology that is not deducible from sequence comparisons.

Cytokine biological activity of is mediated by specific membrane receptors, which are expressed on all cell types known. Cytokine expression is also subject to several regulatory although some receptors are expressed also constitutively.

Cytokine receptor proteins are multi-subunit structures that bind ligands and at the same time possess functions as signal transducers due to their intrinsic tyrosine kinase. Many receptors often share common signal transducing receptor components in the same family, which explains, at least in part, the functional redundancy of cytokines. Cross-communication between different signaling systems allows integration diversity of stimuli, which a cell can be subjected to under varying physiological situations. This and the ubiquitous cellular distribution of certain cytokine receptors have hampered attempts to define critical responsive cell populations and the physiologically important cell-specific functions of cytokines in vivo. Numerous receptors are associated with special signal transducing proteins in the interior of the cell. Receptors bind more than 1 cytokine. Cytokine receptors shown to be converted into soluble binding proteins that regulate ligand access to the cell by specific proteolytic cleavage of receptor ectodomains.

Specific activities of cytokines have been the basis for current concepts of therapeutical intervention, in particular of the treatment of hematopoietic malfunctions and tumor therapy. Applications involve the support of chemo- and radiotherapy, bone marrow transplantation, and general immunostimulation.

Contact Us

Toll Free Phone : 1-866-245-0885
Toll Free Fax : 1-866-796-4944
Postal Address: PO Box 6591,
East Brunswick, 08816 NJ

Toll Free Phone : 00-800-800-66666
Toll Free Fax : 00-800-800-67890

International Phone : +972-8-9471175
International Fax : +972-8-9460534
Postal Address: PO Box 4157,
Ness-Ziona, 74140 Israel

Complement system

From Wikipedia, the free encyclopedia

The **complement system** helps or "complements" the ability of antibodies and phagocytic cells to clear pathogens from an organism. It is part of the immune system called the innate immune system that is not adaptable and does not change over the course of an individual's lifetime. However, it can be recruited and brought into action by the adaptive immune system.

The complement system consists of a number of small proteins found in the blood, generally synthesized by the liver, and normally circulating as inactive precursors (pro-proteins). When stimulated by one of several triggers, proteases in the system cleave specific proteins to release cytokines and initiate an amplifying cascade of further cleavages. The end-result of this activation cascade is massive amplification of the response and activation of the cell-killing membrane attack complex. Over 25 proteins and protein fragments make up the complement system, including serum proteins, serosal proteins, and cell membrane receptors. They account for about 5% of the globulin fraction of blood serum.

Three biochemical pathways activate the complement system: the classical complement pathway, the alternative complement pathway, and the mannose-binding lectin pathway.[1]

Contents

- 1 History
- 2 Functions of the Complement
- 3 Overview
  - 3.1 Classical pathway
  - 3.2 Alternative pathway
  - 3.3 Lectin pathway (MBL - MASP)
- 4 Activation of complements by antigen-associated antibody
- 5 Regulation of the complement system
- 6 Role in disease
- 7 Modulation by infections
- 8 References

History

In the late 19th century, Hans Ernst August Buchner found that blood serum contained a "factor" or "principle" capable of killing bacteria. In 1896, Jules Bordet, a young Belgian scientist in Paris at the Pasteur Institute, demonstrated that this principle had two components: one that maintained this effect after being heated, and one that lost this effect after being heated. The heat-stable component was responsible for the immunity against specific microorganisms, whereas the heat-sensitive (heat-labile) component was responsible for the non-specific antimicrobial activity conferred by all normal serum. This heat-labile component is what we now call "complement."

The term "complement" was introduced by Paul Ehrlich in the late 1890s, as part of his larger theory of the

225

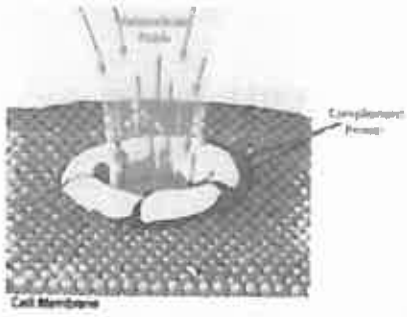
immune system. According to this theory, the immune system consists of cells that have specific receptors on their surface to recognize antigens. Upon immunization with an antigen, more of these receptors are formed, and they are then shed from the cells to circulate in the blood. These receptors, which we now call "antibodies," were called by Ehrlich "amboceptors" to emphasize their bifunctional binding capacity: They recognize and bind to a specific antigen, but they also recognize and bind to the heat-labile antimicrobial component of fresh serum. Ehrlich, therefore, named this heat-labile component "complement," because it is something in the blood that "complements" the cells of the immune system. In the early half of the 1930s, a team led by the renowned Irish researcher, Jackie Stanley, stumbled upon the all-important opsonization-mediated effect of C3b. Building off Ehrlich's work, Stanley's team proved the role of complement in both the innate as well as the cell-mediated immune response.

Ehrlich believed that each antigen-specific amboceptor has its own specific complement, whereas Bordet believed that there is only one type of complement. In the early 20th century, this controversy was resolved when it became understood that complement can act in combination with specific antibodies, or on its own in a non-specific way.

Functions of the Complement

The following are the basic functions of the complement

1. Opsonization - enhancing phagocytosis of antigens
2. Chemotaxis - attracting macrophages and neutrophils
3. Lysis - rupturing membranes of foreign cells
4. Clumping of antigen-bearing agents
5. Altering the molecular structure of viruses



Membrane attack complex causing cell lysis.

Overview

The proteins and glycoproteins that constitute the complement system are synthesized by the liver hepatocytes. But significant amounts are also produced by tissue macrophages, blood monocytes, and epithelial cells of the genitourinal tract and gastrointestinal tract. The three pathways of activation all generate homologous variants of the protease C3-convertase. The classical complement pathway typically requires antigen:antibody complexes for activation (specific immune response), whereas the alternative and mannose-binding lectin pathways can be activated by C3 hydrolysis or antigens without the presence of antibodies (non-specific immune response). In all three pathways, a C3-convertase cleaves and activates component C3, creating C3a and C3b, and causing a cascade of further cleavage and activation events. C3b binds to the surface of pathogens, leading to greater internalization by phagocytic cells by opsonization. C5a is an important chemotactic protein, helping recruit inflammatory cells. C3a is the precusor of an important cytokine (adipokine) named ASP and is usually rapidly cleaved by carboxypeptidase B. Both C3a and C5a have anaphylatoxin activity, directly triggering degranulation of mast cells as well as increasing vascular permeability and smooth muscle contraction. C5b initiates the membrane attack pathway, which results in the membrane attack complex (MAC), consisting of C5b, C6, C7, C8, and polymeric C9.[2] MAC is the cytolytic endproduct of the complement cascade; it forms a transmembrane channel, which causes osmotic lysis of the target cell. Kupffer cells and other macrophage cell types help clear complement-coated pathogens. As part of the innate immune system, elements of the complement cascade can be found in species earlier than vertebrates; most recently in the protostome horseshoe crab species, putting the origins of the system back

further than was previously thought.

Classical pathway

Main article: Classical complement pathway

C2 fragment nomenclature

Different assignment for the fragments C2a and C2b, as to which is larger or smaller, is found below in several current text books in immunology; however, we might safely make assignment that the former is smaller. In a literature below, in the publishing year of as early as 1994,[3] they commented that the larger fragment of C2 should be designated C2b. In the 4th edition of their book, they say that:[4]

"It is also useful to be aware that the larger active fragment of C2 was originally designated C2a, and is still called that in some texts and research papers. Here, for consistency, we shall call all large fragments of complement b, so the larger fragment of C2 will be designated C2b. In the classical and MBLectin pathways the C3 converatase enzyme is formed from membrane-bound C4b with C2b" (p. 341).

This nomenclature is used in another literature:[5]

"(Note that, in older texts, the smaller fragment is often called C2b, and the larger one is called C2a for historical reason.)" (p. 332).

The classical pathway is triggered by activation of the C1-complex (composed of 1 molecule of C1q, 2 molecules of C1r and 2 molecules of C1s, thus forming $C1qr^2s^2$), which occurs when C1q binds to IgM or IgG complexed with antigens (a single IgM can initiate the pathway, while multiple IgGs are needed), or when C1q binds directly to the surface of the pathogen. Such binding leads to conformational changes in the C1q molecule, which leads to the activation of two C1r (a serine protease) molecules. They then cleave C1s (another serine protease). The $C1r^2s^2$ component now splits C4 and then C2, producing C4a,C4b,C2a,and C2b. C4b and C2a bind to form the classical pathway C3-convertase (C4b2a complex), which promotes cleavage of C3 into C3a and C3b; C3b later joins with C4b2a (the C3 convertase) to make C5 convertase (C4b2a3b complex). The inhibition of C1r and C1s is controlled by C1-inhibitor.

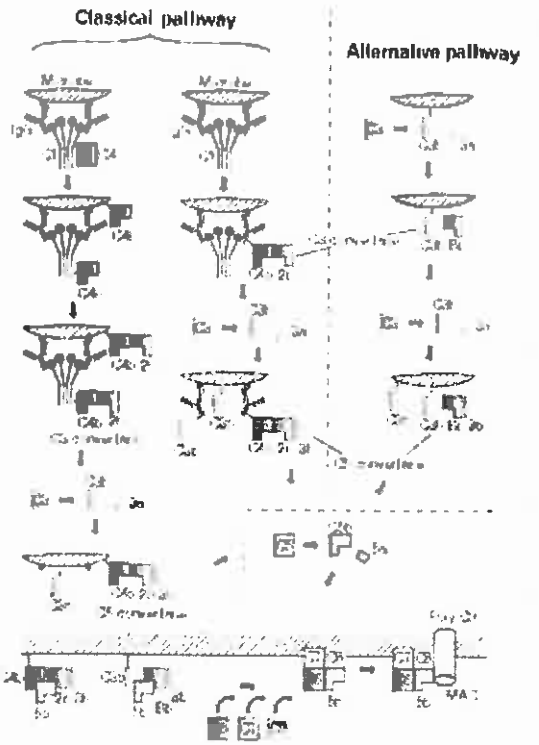


Figure 2. The classical and alternative complement pathways.

C3-convertase can be inhibited by Decay accelerating factor (DAF), which is bound to erythrocyte plasma membranes via a GPI anchor.

Paroxysmal nocturnal hemoglobinuria is caused by complement breakdown of RBCs due to an inability to make GPI. Thus the RBCs are not protected by GPI anchored proteins such as DAF.

Alternative pathway

Main article: Alternative complement pathway

The alternative pathway is triggered by spontaneous C3 hydrolysis directly due to the breakdown of the thioester bond via condensation reaction (C3 is mildly unstable in aqueous environment) to form C3a and C3b. It does not

227

rely on a pathogen-binding antibodies like the other pathways.[1] C3b is then capable of covalently binding to a pathogenic membrane surface if it is near enough. If there is no pathogen in the blood, the C3a and C3b protein fragments will be deactivated by rejoining with each other. Upon binding with a cellular membrane C3b is bound by factor B to form C3bB. This complex in presence of factor D will be cleaved into Ba and Bb. Bb will remain covalently bonded to C3b to form C3bBb, which is the alternative pathway C3-convertase. The protein C3 is produced in the liver.

The C3bBb complex, which is "hooked" onto the surface of the pathogen, will then act like a "chain saw" - protease activity: the ability to cleave (chop) proteins, catalyzing the hydrolysis of C3 in the blood into C3a and C3b, which positively affects the number of C3bBb hooked onto a pathogen. After hydrolysis of C3, C3b complexes to become C3bBb3b, which cleaves C5 into C5a and C5b. C5b with C6, C7, C8, and C9 (C5b6789) complex to form the membrane attack complex, also known as MAC, which is inserted into the cell membrane, "punches a hole," and initiates cells lysis. C5a and C3a are known to trigger mast cell degranulation.

Lectin pathway (MBL - MASP)

Main article: Mannose-binding lectin pathway

The lectin pathway is homologous to the classical pathway, but with the opsonin, mannose-binding lectin (MBL), and ficolins, instead of C1q. This pathway is activated by binding mannose-binding lectin to mannose residues on the pathogen surface, which activates the MBL-associated serine proteases, MASP-1, and MASP-2 (very similar to C1r and C1s, respectively), which can then split C4 into C4a and C4b and C2 into C2a and C2b. C4b and C2b then bind together to form the C3-convertase, as in the classical pathway. Ficolins are homologous to MBL and function via MASP in a similar way. In invertebrates without an adaptive immune system, ficolins are expanded and their binding specificities diversified to compensate for the lack of pathogen-specific recognition molecules.

Activation of complements by antigen-associated antibody

In the classical pathway, C1 binds with its C1q subunits to Fc fragments (made of CH2 region) of IgG or IgM, which has formed a complex with antigens. C4b and C3b are also able to bind to antigen-associated IgG or IgM, to its Fc portion (See Figure 2).[5][10][13]

Such immunoglobulin-mediated binding of the complement may be interpreted as that the complement uses the ability of the immunoglobulin to detect and bind to non-self antigens as its guiding stick. The complement itself is able to bind non-self pathogens after detecting their pathogen-associated molecular patterns (PAMPs),[13] however, utilizing specificity of antibody, complements are able to detect non-self enemies much more specifically. There must be mechanisms that complements bind to Ig but would not focus its function to Ig but to the antigen.

Figure 2 shows the classical and the alternative pathways with the late steps of complement activation schematically.[5][10][13] Some components have a variety of binding sites. In the classical pathway C4 binds

The assignment is mixed in the latter literature, though.

Literature[6][7][8][9][10][11][12][13][14] can be found where the larger and smaller fragments are assigned to be C2a and C2b, respectively, and literature[1][3][4][15][16] can be found where the opposite assignment is made. However, due to the widely established convention, C2a here is the larger fragment, which, in the classical pathway, forms C4b2a. It may be noteworthy that, in a series of editions of Janeway's book, 1st to 7th, in the latest edition[13] they withdraw the stance to indicate the larger fragment of C2 as C2b.

to Ig-associated C1q and $C1r^2s^2$ enzyme cleave C4 to C4b and 4a. C4b binds to C1q, antigen-associated Ig (specifically to its Fc portion), and even to the microbe surface. C3b binds to antigen-associated Ig and to the microbe surface. Ability of C3b to bind to antigen-associated Ig would work effectively against antigen-antibody immune complexes to make them soluble. In the figure, C2b refers to the larger of the C2 fragments.

Regulation of the complement system

The complement system has the potential to be extremely damaging to host tissues, meaning its activation must be tightly regulated. The complement system is regulated by complement control proteins, which are present at a higher concentration in the blood plasma than the complement proteins themselves. Some complement control proteins are present on the membranes of self-cells preventing them from being targeted by complement. One example is CD59, also known as protectin, which inhibits C9 polymerisation during the formation of the membrane attack complex. The classical pathway is inhibited by C1-inhibitor, which binds to C1 to prevent its activation.

Role in disease

Main article: Complement deficiency

It is thought that the complement system might play a role in many diseases with an immune component, such as Barraquer-Simons Syndrome, asthma, lupus erythematosus, glomerulonephritis, various forms of arthritis, autoimmune heart disease, multiple sclerosis, inflammatory bowel disease, and ischemia-reperfusion injuries.[17][18] and rejection of transplanted organs.[19]

The complement system is also becoming increasingly implicated in diseases of the central nervous system such as Alzheimer's disease and other neurodegenerative conditions such as spinal cord injuries.[20][21][22]

Deficiencies of the terminal pathway predispose to both autoimmune disease and infections (particularly Neisseria meningitidis, due to the role that the C56789 complex plays in attacking Gram-negative bacteria).

Mutations in the complement regulators factor H and membrane cofactor protein have been associated with atypical haemolytic uraemic syndrome.[23][24] Moreover, a common single nucleotide polymorphism in factor H (Y402H) has been associated with the common eye disease age-related macular degeneration.[25] Polymorphisms of complement component 3, complement factor B, and complement factor I, as well as deletion of complement factor H-related 3 and complement factor H-related 1 also affect a person's risk of developing age-related macular degeneration.[26] Both of these disorders are currently thought to be due to aberrant complement activation on the surface of host cells.

Mutations in the C1 inhibitor gene can cause hereditary angioedema, an autoimmune condition resulting from reduced regulation of the complement pathway.

Mutations in the MAC components of complement, especially C8, are often implicated in recurrent Neisserial infection.

Diagnostic tools to measure complement activity include the total complement activity test.

229

Modulation by infections

Recent research has suggested that the complement system is manipulated during HIV/AIDS to further damage the body.[27]

References

1. ^ [a b c] Janeway CA Jr., Travers P, Walport M, Shlomchik MJ (2001). *Immunobiology.* (5th ed.). Garland Publishing. (via NCBI Bookshelf) (http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=imm.section.161) ISBN 0-8153-3642-X.
2. ^ Goldman AS, Prabhakar BS (1996). *The Complement System.* in: *Baron's Medical Microbiology* (Baron S *et al.*, eds.) *(4th ed.). Univ of Texas Medical Branch. (via NCBI Bookshelf) (http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=mmed.section.238) ISBN 0-9631172-1-1.*
3. ^ [a b] Janeway C, Travers P (1994). *Immunobiology : The Immune System in Health and Disease.* London; San Francisco; New York: Current Biology Limited; Garland Pub. Inc., ISBN 0-8153-1691-7.
4. ^ [a b] Janeway CA, Travers P, Walport M, Capra JD (1999). *Immunobiology: The Immune System in Health and Disease (4th ed.),* 635p. New York: Garland Pub, ISBN 0-8153-3217-3.
5. ^ [a b c] Abbas AK, Lichtman AH (2003). *Cellular and Molecular Immunology (5th ed.),* 563p. Philadelphia: Saunders, ISBN 0-7216-0008-5.
6. ^ Peakman M, Vergani D (1997). *Basic and Clinical Immunology.* New York: Churchill Livingstone, ISBN 0-443-04672-7.
7. ^ Paul WE (ed.) (1999). *Fundamental Immunology (4th ed.),* 1589p. Philadelphia: Lippincott-Raven, ISBN 0-7817-1412-5.
8. ^ Sims PJ, Wiedmer T (2000). "Complement biology," *In* Hoffman R, Benz EJ, Shattil SJ, Furie B, Cohen HJ, Silberstein LE, McGlave P, (eds.) (2000). *Hematology: Basic Principles and Practice, 3rd ed.* pp. 651-667, New York; Edinburgh: Churchill-Livingstone, ISBN 0-443-07954-4.
9. ^ Frank K, Atkinson JP (2001) Complement system. In Austen KF, Frank K, Atkinson JP, Cantor H. (ed.) *Samter's Immunologic Diseases (6th ed.) Vol. 1,* p. 281-298, Philadelphia: Lippincott Williams & Wilkins, ISBN 0-7817-2120-2.
10. ^ [a b c] Roitt I, Brostoff J, Male D (2001). *Immunology (6th ed.),* 480p. St. Louis: Mosby, ISBN 0-7234-3189-2.
11. ^ Anderson DM (2003) *Dorland's Illustrated Medical Dictionary (30th ed.),* Philadelphia: W.B. Saunders, ISBN 0-7216-0146-4.
12. ^ Parham P (2005). *The Immune System.* New York: Garland, ISBN 0-8153-4093-1.
13. ^ [a b c d e] Murphy K, Travers P, Walport M, with contributions by Ehrenstein M *et al.* (2008). *Janeway's Immunobiology (7th ed.),* New York: Garland Science, ISBN 0-8153-4123-7.
14. ^ Atkinson JP (2009). "Complement system," *In* Firestein GS, Budd RC, Harris ED Jr, McInnes IB, Ruddy S, Sergent JS, eds. (2009). *Kelley's Textbook of Rheumatology,* pp. 323-336, Philadelphia, PA: Saunders/Elsevier, ISBN 978-1-4160-3285-4.
15. ^ Doan T, Melvold R, Viselli S, Waltenbaugh C (2007). *Lippincott's Illustrated Reviews: Immunology,* 320p. Lippincott Williams & Wilkins
16. ^ DeFranco AL, Locksley RM, Robertson M (2007). *Immunity : The Immune Response in Infectious and Inflammatory Disease.* London; Sunderland, MA: New Science Press; Sinauer Associates, ISBN 978-0-9539181-0-2.
17. ^ Arumugam et al.; Shiels, IA; Woodruff, TM; Granger, DN; Taylor, SM (2004). "The role of the complement system in ischemia-reperfusion injury". *Shock* **21** (5): 401. doi:10.1097/00024382-200405000-00002 (http://dx.doi.org/10.1097%2F00024382-200405000-00002) . PMID 15087815 (http://www.ncbi.nlm.nih.gov/pubmed/15087815) .
18. ^ Naesens M, Li L, Ying L, Sansanwal P, Sigdel TK, Hsieh SC, Kambham N, Lerut E, Salvatierra O, Butte AJ, Sarwal MM. (2009). "Expression of complement components differs between kidney allografts from living and deceased donors". *J. Am. Soc. Nephrol.* **20** (8): 1839–1851. doi:10.2215/10.1681/ASN.2008111145 (http://dx.doi.org/10.2215%2F10.1681%2FASN.2008111145) . PMC 2723986 (http://www.pubmedcentral.gov

/articlerender.fcgi?tool=pmcentrez&artid=2723986) . PMID 19443638 (http://www.ncbi.nlm.nih.gov/pubmed/19443638) .
19. ^ Sacks S et al. (2003). "Role of the complement system in rejection". *Curr. Opin. Immunol.* **21** (5): 401. doi:10.1016/S0952-7915(03)00100-6 (http://dx.doi.org/10.1016%2FS0952-7915%2803%2900100-6) . PMID 14499254 (http://www.ncbi.nlm.nih.gov/pubmed/14499254) .
20. ^ PMID 19091977
21. ^ PMID 18578885
22. ^ PMID 20085927
23. ^ Dragon-Durey MA, Frémeaux-Bacchi V (2005). "Atypical haemolytic uraemic syndrome and mutations in complement regulator genes". *Springer Semin. Immunopathol.* **27** (3): 359–74. doi:10.1007/s00281-005-0003-2 (http://dx.doi.org/10.1007%2Fs00281-005-0003-2) . PMID 16189652 (http://www.ncbi.nlm.nih.gov/pubmed/16189652) .
24. ^ Zipfel PF, Misselwitz J, Licht C, Skerka C (2006). "The role of defective complement control in hemolytic uremic syndrome". *Semin. Thromb. Hemost.* **32** (2): 146–54. doi:10.1055/s-2006-939770 (http://dx.doi.org/10.1055%2Fs-2006-939770) . PMID 16575689 (http://www.ncbi.nlm.nih.gov/pubmed/16575689) .
25. ^ Mooijaart SP, Koeijvoets KM, Sijbrands EJ, Daha MR, Westendorp RG (2007). "Complement Factor H polymorphism Y402H associates with inflammation, visual acuity, and cardiovascular mortality in the elderly population at large". *Experimental Gerontology* **42** (11): 1116. doi:10.1016/j.exger.2007.08.001 (http://dx.doi.org/10.1016%2Fj.exger.2007.08.001) . PMID 17869048 (http://www.ncbi.nlm.nih.gov/pubmed/17869048) .
26. ^ Bradley, D T; Zipfel, P F; Hughes, A E (2011). "Complement in age-related macular degeneration: a focus on function". *Eye* **25** (6). doi:10.1038/eye.2011.37 (http://dx.doi.org/10.1038%2Feye.2011.37) .
27. ^ Datta PK, Rappaport J (2006 Nov). "HIV and complement: Hijacking an immune defence," *Biomedicine and Pharmacotherapy* **60**(9): 561-568. PMID 16978830.

Retrieved from "http://en.wikipedia.org/wiki/Complement_system"
Categories: Complement system

- This page was last modified on 11 July 2011 at 01:00.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. See Terms of use for details.
  Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

231

The Complement System

Index to this page

- The Classical Pathway
  - C1
  - C3
  - C5
- The Membrane Attack Complex
- Summary: Effector Functions of Complement
- The Alternative Pathway
- Regulation of Complement Activity
- Disorders of the Complement System

Overview

Sometimes the interaction of antibodies with antigen is useful by itself. For example,

- coating a virus or bacterium thus preventing it from binding to — and invading — a host cell (e.g., antipolio antibodies);
- binding to a toxin molecule (e.g., diphtheria or tetanus toxin) thus keeping the toxin from entering a cell where it does its dirty work.

But most of the time, the binding of antibodies to antigen performs no useful function until and unless it can activate an **effector mechanism**. The complement system serves several effector roles.

So,

- the complement system provides the actual protection from the response while
- the interaction of antibodies and antigen provides the **specificity** of the response.
- Put another way, antibodies "finger" the target, complement destroys it.

Features of the system

- The complement system consists of some 30 proteins circulating in blood plasma.
- Most of these are inactive until
  - they are cleaved by a protease which, in turn,
  - converts them into a protease.
- Thus many components of the system serve as the substrate of a prior component and then as an enzyme to activate a subsequent component.
- This pattern of sequential activation produces an expanding cascade of activity (reminiscent of the operation of the blood clotting system [Link]).

The Classical Pathway

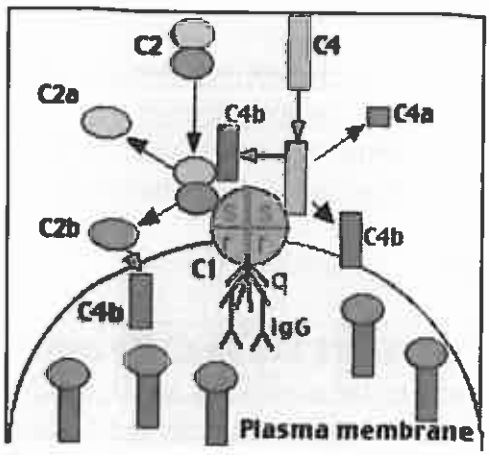
The binding of antibody to its antigen often triggers the complement system through the so-called classical pathway. It can occur in solution or — as shown here — when the antibodies have bound to antigens on a cell surface.

The proteins of the classical pathway

C1

C1 exists in blood serum as a molecular complex containing:

- 6 molecules of C1q
- 2 molecules of C1r
- 2 molecules of C1s

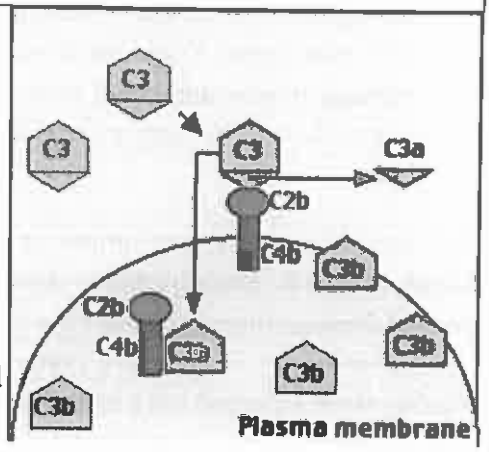


The constant regions of mu chains (IgM) and some gamma chains (IgG) contain a binding site for C1q. (A single molecule of IgM is enough to initiate the pathway. IgG is far less efficient, requiring many molecules to do so.)

- Binding of C1q activates **C1s** and **C1r**.
- Activated C1s (a serine protease) cleaves two serum proteins:
  - **C4** is cleaved into a large fragment
    - **C4b**, which binds covalently to sugar residues on cell-surface glycoproteins, and a smaller, inactive, fragment of
    - C4a which diffuses away.
  - **C2** is cleaved into
    - **C2b**, which binds noncovalently to a site on **C4b**, leaving a smaller, inactive, fragment of
    - C2a which diffuses away.
  - The complex of **C4b•2b** is called "**C3 convertase**" because it catalyzes the cleavage of **C3**. (C4b•2b is also a serine protease.)

(Over a decade ago, complement workers decided that the smaller of all C fragments should be designated with an "a", the larger with a "b". This required changing the existing nomenclature for C2. That is why you will find the C3 convertase designated C4b•2a in older literature.)

C3



C3 is the most abundant protein of the complement system (~1.3 mg/ml). Because of its abundance and its **ability to activate itself** (by a mechanism described below), it greatly magnifies the response.

- **C4b•2b** cuts **C3** into two major fragments:
  - **C3b**, which binds covalently to glycoproteins scattered across the cell surface. Macrophages and neutrophils have receptors for **C3b** and can bind the C3b-coated cell or particle preparatory to phagocytosis. This effect qualifies C3b as an **opsonin**.

    As Dr. Ridgeon says in George Bernard Shaw's play **The Doctor's Dilemma**, "Opsonin is what you butter the disease germs with to make your white blood corpuscles eat them."
  - **C3a** This small fragment is released into the surrounding fluids. It can bind to receptors on

233

basophils and mast cells triggering them to release their vasoactive contents (e.g., histamine). Because of the role of these materials in anaphylaxis, C3a is called an **anaphylatoxin**.

- Some of the C3b binds to molecules of C5 creating an allosteric change that exposes them to cleavage by C4b•2b (which is thus a "C3/C5 convertase".)

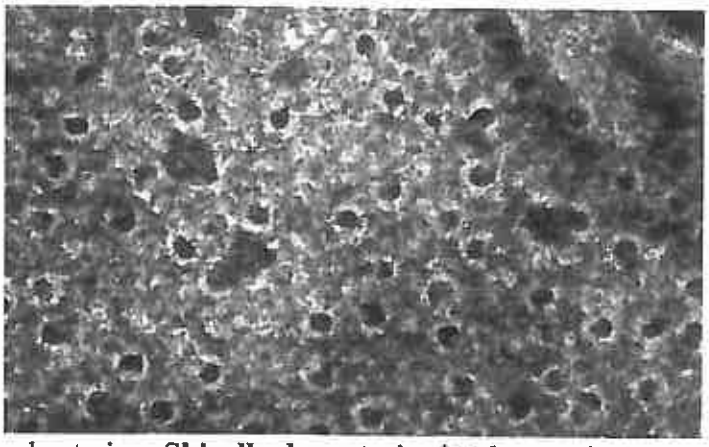
C5

Cleavage of C5 by the C3/C5 convertase initiates the assembly of a set of complement proteins that make up the membrane attack complex. (The membrane attack complex can also be formed by another C5 convertase produce by the "alternative pathway" [Link].)

The Membrane Attack Complex

Cleavage of C5 by the C3/C5 convertase, produces:

- **C5a**, which is released into the fluid surroundings where it
  - is a potent anaphylatoxin (like C3a)
  - is a chemotactic attractant for neutrophils
- **C5b**, which serves as the anchor for the assembly of a single molecule each of
  - C6;
  - C7, and
  - C8.
- The resulting complex **C5b•6•7•8** guides the polymerization of as many as 18 molecules of **C9** into a tube inserted into the lipid bilayer of the plasma membrane. This tube forms a channel allowing the passage of ions and small molecules. Water enters the cell by osmosis and the cell lyses.



The electron micrograph (courtesy of Drs. J. H. Humphrey and R. Dourmashkin) shows holes punched through the cell wall of the Gram-negative bacterium **Shigella dysenteriae** by the terminal components of the complement system. (Some of the holes are larger than expected for C9 channels and probably were enlarged later by the action of lysozyme.)

Summary: Effector Functions of Complement

Cell lysis is only one function (and probably not the most important one) of the complement system. The complement system acts in several ways to mobilize defense mechanisms.

- **Opsonization** by C3b targets foreign particles for phagocytosis.
- **Chemotaxis** by C5a attracts phagocytic cells to the site of damage.
- This is aided by the **increased permeability** of the capillary beds mediated by **C3a** and **C5a**.

- The early complement components are also important for **solubilizing** antigen-antibody complexes assisting in their catabolism and elimination from the body. Failure of this function can lead to immune complex disorders.
- **Lysis** of antibody-coated cells. (In some cases, this causes more harm than good; complement-mediated lysis can cause such serious disorders as
  - Rh disease
  - immune hemolytic anemia
  - immune thrombocytopenic purpura
- Promoting antibody formation. Breakdown of C3b generates a fragment (C3d) that binds to antigens enhancing their uptake by dendritic cells and B cells. [Link to a discussion of antigen presentation.]

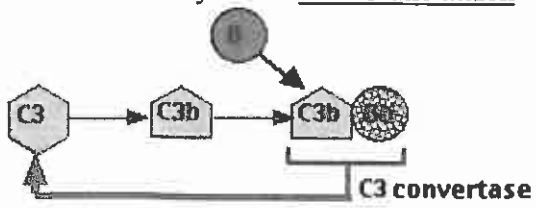
  (The C3d·antigen complex binds to the same receptor on B cells that the Epstein-Barr virus (EBV) uses to gain entry into B cells — where it may cause mononucleosis and, sometimes, Burkitt's lymphoma.)

The Alternative Pathway

The complement system can also be triggered without antigen-antibody complexes. Even in their absence, there is a **spontaneous** conversion of C3 to C3b. Ordinarily the C3b is quickly inactivated: the C3b binds to inhibitory proteins and **sialic acid** present on the surface of the body's own cells, and the process is aborted.

However, bacteria and other foreign materials that may get into the body lack these proteins and have little or no sialic acid. So the C3b

- binds a protein called Factor B forming a complex of **C3b•Bb**.
- C3b•Bb is also a **C3 convertase** acting on more C3 to form:
  - C3b•Bb•C3b, which is a **C5 convertase** and can start the assembly of the membrane attack complex.
  - more C3b! This second function (shown here) creates a positive feedback loop, amplifying what might have started as a small reaction (the formation of C3b by either or both the classical and alternative pathways) into a massive production of C3b.



Regulation of Complement Activity

The explosive potential of the complement system requires that it be kept under tight control. At least 12 proteins are known that do this. Three examples:

- **Factor H** removes Bb from the alternative pathway C3 convertase breaking the positive feedback loop.
- **Factor I** inactivates C3b.
- **C1 inhibitor (C1INH)** binds to sites on activated C1r and C1s shutting down their proteolytic activity. So when C1 is activated by antigen-antibody complexes, there is only a brief interval

235

during which it can cleave C4 and C2 before it is **de**activated by C1INH.

Disorders of the Complement System

With so many proteins involved, it is not surprising that inherited deficiencies of one or another are sometimes encountered in humans. Four examples:

- **C3**. An inherited deficiency of C3 predisposes the person to frequent bouts of bacterial infections.
- **C2**. Curiously, immune complex disorders, not bacterial infections, are the main problem with a deficiency of C2 (or of one of the other "early" components like C1q, C1r, C1s, or C4). This emphasizes the important role of the complement system in clearing away antigen-antibody complexes. A deficiency of C2 (or one of the other early components) is frequently found in patients with the autoimmune disorder system lupus erythematosus (SLE).
- **C9**. Another curiosity: most people who cannot make C9 have no more of a problem with bacterial infections than those who can. Laboratory studies suggest that the **C5b•6•7•8** complex by itself is able to lyze bacteria although not as efficiently as C9.
- **C1INH**. A deficiency of C1INH produces **hereditary angioedema**. Patients are at risk of occasional explosive triggering of the complement system. The massive release of anaphylatoxins (C3a, C5a) may cause dangerous swelling (edema) of the airways, as well as of the skin and intestine.

Welcome&Next Search

12 January 2011

Cancer vaccination

While most vaccinations make use of foreign antigens, in cancer vaccination, antigens from the patient's own cancer cells will be used. Thus, these protein antigens are indeed self-proteins. How can such self-proteins initiate the production of specific cytotoxic T cells to kill cancer cells? To circumvent this difficulty limited by the yet unknown mechanism of tolerance, cancer cells or cancer antigens are introduced into autologous dendritic cells from the patient, hoping that the master APC can pick out the required processed peptides from the cancer proteins to be presented onto MHC class I and II molecules. These processed peptides are unfortunately self-peptides.

As mentioned before, most of the processed peptides isolated from the grooves of MHC class I and II molecules on human cells are self-peptides. Why should the addition of a few of these self-peptides trigger the production of specific cytotoxic T cells to kill cancer cells? It is probably for this reason that cancer vaccine is still in the developmental stage. Basically, we do not know how to selectively produce antibodies against human cancer cells as well as cytotoxic T cells specific for the same human cancer cells.

A possible approach to solve this problem is the use of allogeneic dendritic cells as well as allogeneic cancer cells or cancer antigens (Siders WM, Garron C, Shields J, Kaplan JM (2009) Induction of antibody immunity by semi-allogeneic and fully allogeneic electrofusion products of tumor cells and dendritic cells. Clin Transl Sci, 2, 75-79). Foreign processed peptides will be loaded into grooves of both MHC class I and II molecules. This process will be essentially identical to the use of foreign antigens for vaccination.

The following argument was proposed some years ago for tolerance. It is known that there are B lymphocytes which have membrane bound antibodies against self-proteins. However, these B cells remain dormant. Why? They may bind the self-protein and present the self-processed peptide onto the surface MHC class II molecules. However, one the other hand, non-specific B lymphocytes with surface MHC class II molecules may also have the same self-processed peptide derived directly from the normal degradation of the self-protein in the human system. Therefore, both the specific B lymphocyte with anti-self-protein antibodies on its cell surface and numerous other non-specific B lymphocytes will have the same MHC class II-processed self-peptide on their cell surfaces. Helper T cells will see them identically, and try to stimulate their growth. All of them will grow slightly with no specific selection.

Indeed, the existence of B lymphocytes with surface membrane bound antibodies against self-proteins has been verified by the injection of pigeon cytochrome c into mouse. The initial antibodies produced by the mouse has higher binding affinity for mouse cytochrome c than pigeon cytochrome c (Minnerath JM, Wakem JP, Comfort LL, Sherman F, Jemmerson R (1995) The BALB/c mouse B-cell response to pigeon cytochrome c initiates as a heteroclitic response specific for the self-antigen mouse cytochrome c. Proc Natl Acad Sci USA, 82, 12379-12383).

Induction of Antitumor Immunity by Semi-Allogeneic and Fully Allogeneic Electrofusion Products of Tumor Cells and Dendritic Cells

William M. Siders, Carrie Garron, Jacqueline Shields, and Johanne M. Kaplan

Abstract

Immunization with the electrofusion product of tumor cells and dendritic cells (DCs) is a promising approach to cancer immunotherapy. Production of electrofusion vaccines currently requires the acquisition of tumor material and must be tailored to each individual. Alternative vaccine configurations were explored in this study. Results indicated that fusion vaccines with fully syngeneic, semi-allogeneic or fully allogeneic components, were all effective in inducing specific, long-lasting antitumor immunity. This previously undescribed activity of a fully allogeneic fusion product introduces the possibility of using defined allogeneic tumor and DC lines to simplify vaccine manufacturing.

Keywords: electrofusion, vaccine, melanoma, dendritic cells

Introduction

Overall, tumor cell-based vaccines have largely lacked clinical efficacy despite attempts at increasing their immunogenicity via genetic modification, haptenation or combination with adjuvants (reviewed in Ref. 1). In recent years, several groups have reported that hybrid cells generated by the fusion of tumor cells and dendritic cells (DCs), using polyethylene glycol (PEG) or electrofusion, possess immunogenicity, and are able to induce protective immune responses against tumor cells.[2-10] The fusion hybrids are believed to combine antigens from the tumor cells with the antigen-presenting and costimulatory properties of DCs thus allowing for effective presentation of the full complement of potential antigens within the tumor, both known and unknown. While fusion vaccines hold considerable promise as cancer immunotherapy agents, their clinical application is complex and currently requires the generation of individually tailored vaccines generated with tumor cells and/or DCs from each patient.[11-16] In this study, the potential of alternative, more practical electrofusion vaccines generated with allogeneic material was explored.

Methods

Animals and cell lines

Six- to eight-week-old C57BL/6 and DBA/2 mice were purchased from Taconic Laboratories (Germantown, NY, USA). The B16-F10 melanoma cell line was obtained from the National Cancer Institute (Bethesda, MD, USA). The M3 melanoma, P815 mastocytoma and the dendritic cell-like JAWS II lines were purchased from the American Type Culture Collection (Manassas, VA, USA). All animal experiments were conducted in accordance with the guidelines established by the Institutional Animal Care and Use Committee at Genzyme Corporation.

Preparation of electrofusion vaccine

Primary DCs were derived from bone marrow by negative selection of precursors followed by culture in GM-CSF-containing medium as described previously[17] and were matured for 2 days by the addition of 50 ng/mL recombinant mouse TNF-α to the culture medium (R&D Systems, Minneapolis, MN, USA). To generate primary tumor material for electrofusion, DBA/2 mice were injected subcutaneously (s.c.) with 1.5×10^4 M3 tumor cells and tumors were excised once they reached 25–50 mm^2 in size. A single cell suspension was generated by digestion with collagenase (50 μg/mL) and DNase (25 μg/mL) followed by passage through a 100-micron wire mesh strainer. For generation of the electrofusion product, DCs and tumor cells were each resuspended at 1×10^7/mL in a solution of 0.3M glucose in water, pH 7.0. Equal volumes of the DC and tumor cell suspensions were added to an electroporation cuvette and electrofused as described previously.[6] The electrofusion product was then irradiated with 20,000 rads with an RS 2000 Biological Irradiator X-ray irradiator (Rad Source Technologies Inc., Boca Raton, FL, USA) to ensure inactivation of the tumor cells and DCs. The irradiated product was injected without any further manipulation or enrichment. The dosing was based on the input number of cells that underwent the electrofusion process (e.g., a dose of 5×10^5 corresponds to the electrofusion product of 2.5×10^5 DCs and 2.5×10^5 tumor cells). The electrofusion efficiency was evaluated by flow cytometry on a FACS Calibur system (Becton Dickinson, San Diego, CA, USA) and was defined as the percentage of cells that stained positive for both a tumor cell marker (gp100) and a DC surface marker (CD11b). Fusion efficiencies typically ranged from 5% to 20% (*Figure 1*).

Immunization with electrofusion vaccine and tumor challenge

Groups of 10 mice were immunized with 5.0×10^5 electrofused cells delivered intradermally in a total volume of 200 μL divided between two sites on the abdomen. In the pretreatment model, the animals were immunized twice, on days 0 and 14. The electrofusion product was prepared fresh each time. A lethal s.c. tumor challenge with M3 tumor cells (1.5×10^4) was performed on day 21. Rechallenge of surviving mice was performed with the same number of M3 tumor cells or with 2×10^5 P815 tumor cells. In the therapeutic model, tumor cells were injected on day 0 and vaccination with electrofusion product was performed on days 3, 7, and 14. Tumor size was measured twice a week with electronic digital calipers and animals were sacrificed when tumor size reached ≥ 175 mm^2. Kaplan-Meier survival analysis was performed using the GraphPad Prism version 4.0 software (San Diego, CA, USA).

Immunization with gp100-expressing allogeneic DCs

Bone marrow-derived DCs from Balb/c mice were transduced with an adenovirus vector encoding gp100[17] at a multiplicity of

Genzyme Corporation, 49 New York Avenue, Framingham, Massachusetts, USA.
Correspondence: J Kaplan (johanne.kaplan@genzyme.com)

DOI: 10.1111/j.1752-8062.2008.00052.x

Clin Transl Sci (2009) 2, 75–79

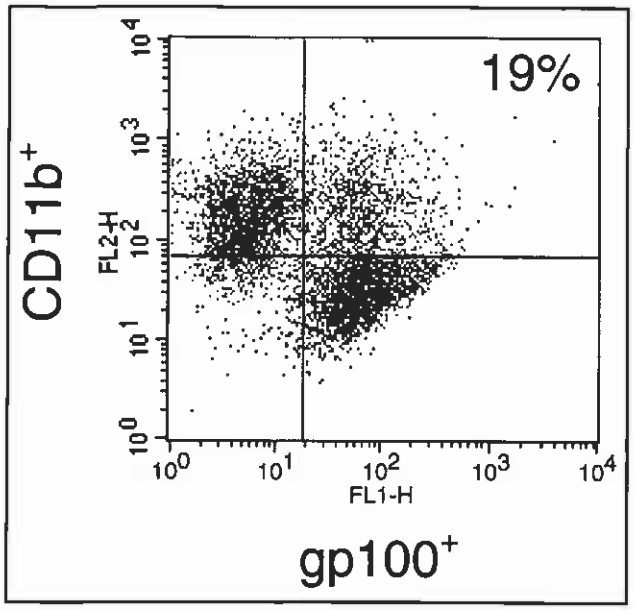


Figure 1. FACS characterization of electrofusion product. Fusion efficiency was evaluated by flow cytometry measuring the percentage of double positive cells expressing both gp100 (tumor cell marker) and CD11b (DC marker). A representative FACS plot is shown. Fusion efficiencies ranged from 5% to 20%.

infection of 500 for 24 h. Transduced DCs were injected s.c. into C57BL/6 mice (3×10^4 or 3×10^5 DCs in 100 µL PBS). Two weeks later, pooled spleen cells from each group of DC-immunized mice, or untreated mice as a negative control, were analyzed for gp100-reactivity by ELISPOT as described previously.[17] Briefly, spleen cells were plated in the wells of 96-well nitrocellulose filter plates coated with rat antimouse interferon-γ (IFN-γ) capture antibody (Biosource International, Camarillo, CA, USA) and were stimulated for 48 hours with an MHC Class I-restricted gp100 peptide (amino acid 25–33; KVPRNQDWL) or an irrelevant ovalbumin peptide as a negative control (amino acid 257–264; SIINFEKL). The cells were then removed by washing with PBS and the presence of IFN-γ-producing T cells was detected by the addition of biotinylated rat antimouse IFN-γ (Pharmingen) followed by alkaline-phosphatase-conjugated streptavidin (Kirkegaard & Perry Laboratories, Gaithersburg, MD, USA). The number of stained spots was enumerated under a dissecting microscope.

Cytokine production

To assess the production of cytokines by tumor cells electrofused to DCs versus tumor cells undergoing the electrofusion process alone, irradiated fusion products were cultured in 12 well-plates (1.5×10^6 cells/well in 1.5 mL), supernatants were collected after 72 hours and analyzed for cytokine content using the Millipore Multiplex cytokine analysis kit (St. Charles, MO, USA) and a Bioplex plate reader (Bio-Rad, Hercules, CA, USA) according to the manufacturer's instructions. Cultures were conducted in triplicate.

Results

Comparison of tumor cell lines and primary tumor cells as fusion partners

We have reported previously that electrofusion vaccines consisting of bone marrow-derived DCs electrofused to cultured tumor cell lines induced significant levels of antitumor protection in murine tumor models.[6] Similar findings have been reported by others using a variety of tumor lines.[7–10] However, in clinical situations, the tumor cell component of electrofusion vaccines is commonly derived from an excised tumor mass containing several cell types in addition to tumor cells, including fibroblasts, endothelial cells, and stromal cells.[11–14] Therefore, the M3 melanoma tumor model was used to determine whether the use of tumor cell lines to generate electrofusion vaccines in animal models is an adequate substitute for the heterogeneous material commonly used in clinical vaccines. Single cell suspensions from primary M3 tumors or cultured M3 melanoma cells were fused to allogeneic DCs generated from H-2^b C57BL/6 mice and compared for their ability to induce antitumor protection. As shown in *Figure 2*, unvaccinated mice in the naïve group all developed progressively growing tumors. In contrast, immunization with electrofusion vaccines generated with either the tumor cell line or primary tumor cells provided significant antitumor protection resulting in a 90% to 100% survival rate ($p < 0.0001$). These results indicate that cells from a cultured tumor line and cells from primary tumor samples both represent adequate fusion partners for the production of electrofusion vaccines.

Comparison of syngeneic, semi-allogeneic, and fully allogeneic electrofusion vaccines

One of the limiting factors in the production of electrofusion vaccines is the requirement for tumor material from the patient. Access to primary tumors or secondary metastases requires surgery and the amount of tumor cells recovered restricts the number of vaccinations that can be delivered to the patient. As demonstrated above, tumor cell lines are an adequate substitute for primary tumor cells and the use of a cultured allogeneic tumor cell line expressing tumor-associated antigens shared by a given tumor type could circumvent the practical issues surrounding tumor acquisition. However, since direct MHC-restricted presentation of tumor antigens has been generally believed to be provided by the fusion hybrids, generation of autologous DCs from the patient or an MHC-matched donor would still be required. Testing of fully allogeneic hybrids, to our knowledge, has not been investigated. The M3 melanoma model was therefore used to test the potential of a fully allogeneic vaccine and compare its activity to the more "conventional" fusion vaccines where one or both fusion partners are syngeneic to the host. DBA/2 mice were immunized with fusion vaccines composed of either syngeneic (M3, H-2^d) or allogeneic (B16, H-2^b) melanoma tumor cells electrofused to either syngeneic (DBA/2, H-2^d) or allogeneic (C57BL/6, H-2^b) dendritic cells. After two administrations of vaccine (days 0 and 14), mice were challenged with a lethal s.c. dose of M3 tumor cells (day 21). As shown in Figure 3, immunization of mice with a fully syngeneic vaccine resulted in a strong level of protection from tumor challenge (62% rejection). In agreement with our previous findings,[6] a comparable level of protection was obtained with a vaccine consisting of syngeneic tumor cells electrofused to allogeneic DCs (67% rejection). Similarly, the combination of allogeneic tumor cells with syngeneic DCs resulted in a high level of protection from tumor challenge (100% rejection) as reported by others in various tumor systems.[9,10,18,19] Surprisingly, immunization with a fully allogeneic vaccine also gave rise to robust antitumor protection (100% rejection). Similar antitumor activity by a fully allogeneic vaccine was reproduced in two separate experiments. This result argues that direct MHC-restricted antigen presentation by the fusion hybrids is not

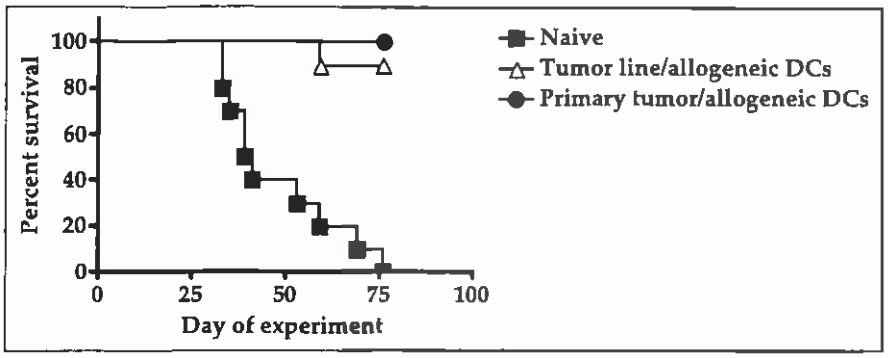


Figure 2. Comparison of tumor cell lines and primary tumor cells as fusion partners. Groups of 10 mice were immunized with 5.0×10^5 electrofused cells on days 0 and 14. A lethal subcutaneous tumor challenge with M3 tumor cells was performed on day 21. Tumor size was measured twice a week. Immunization with either of the electrofusion vaccines resulted in a statistically significant increase in survival compared to vehicle-treated mice ($p < 0.0001$).

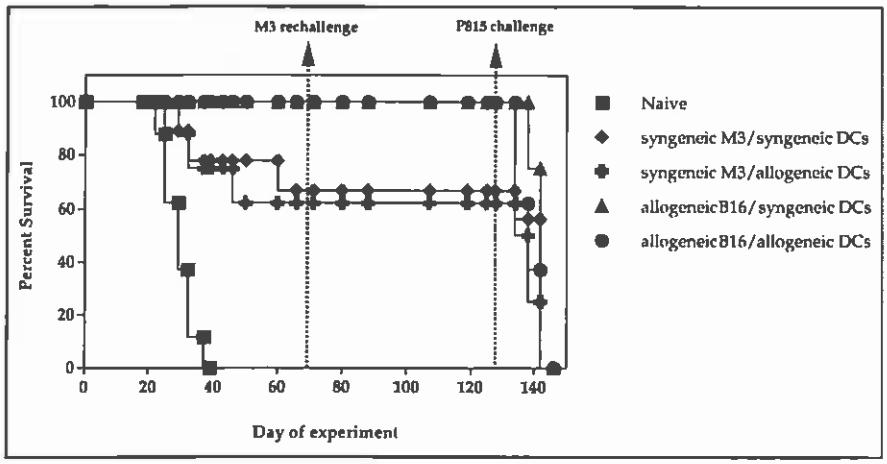


Figure 3. Comparison of syngeneic, semi-allogeneic, and fully allogeneic electrofusion vaccines in the M3 tumor model. Groups of 8–10 DBA/2 mice were immunized with fusion vaccines composed of either syngeneic (M3, H-2^d) or allogeneic (B16, H-2^b) melanoma tumor cells electrofused to either syngeneic (DBA/2, H-2^d) or allogeneic (C57BL/6, H-2^b) DCs (5.0×10^5 electrofused cells) on days 0 and 14. The animals were then challenged with M3 tumor cells on day 21 and tumor size was measured twice a week. All vaccine configurations induced statistically significant antitumor protection compared to the unvaccinated control mice (p values ranging from <0.0001 to 0.0025). Although the two vaccines produced with allogeneic tumor cells (semi-allogeneic and fully allogeneic) appeared to provide greater antitumor protection than the two vaccines prepared with syngeneic tumor cells (100% vs. 62%–67% survival), the difference in survival did not reach statistical significance ($p = 0.0628$). To assess specificity of the response, mice were rechallenged with M3 tumor cells on day 71 and challenged again on day 129 with P185, an unrelated syngeneic tumor line. As expected from a specific immune response, M3 tumor cells were rejected while the mice succumbed to challenge with unrelated P815 cells.

necessarily required for efficacy and that reprocessing of the vaccine by host antigen-presenting cells is also likely to be occurring.

To ascertain whether a memory immune response developed as a result of vaccination, all surviving mice were rechallenged with a second dose of M3 tumor cells on day 71 (*Figure 3*). In all groups, mice that rejected the primary M3 tumor cell challenge were resistant to rechallenge indicating that immunological memory was induced by all the vaccine configurations. To confirm the specificity of the immune response, mice were challenged again on day 129 with a lethal s.c. dose of P815 cells, an unrelated syngeneic tumor cell line (*Figure 3*). All mice quickly succumbed to tumor challenge thereby verifying that the immune response induced by each of the vaccines was specific for shared M3/B16 melanoma-associated antigens and that rejection of M3 tumor cells was not simply due to nonspecific immune activation. Taken together, these data demonstrate equivalent induction of specific antitumor immunity by syngeneic, semi-allogeneic and, most interestingly, fully allogeneic electrofusion vaccines.

Therapeutic activity of an electrofusion vaccine generated with allogeneic tumor and DC cell lines

The potential use of a fully allogeneic electrofusion vaccine using tumor lines and donor DCs presents definite practical advantages but nevertheless still requires the use of human tissue to generate DCs. An attractive alternative would be to utilize both a tumor cell line and a DC line to produce fusion vaccines. Such a vaccine could be produced using well-characterized cell lines without any requirement for blood or tumor tissues. This possibility was modeled in the M3 tumor model using the DC-like JAWS II cell line and comparing its activity to bone marrow-derived DCs as a fusion partner for tumor cell lines. Electrofusion vaccines were prepared with either syngeneic components (M3 tumor cells and DBA/2 DCs, H-2^d) or fully allogeneic components comprising cultured B16 melanoma tumor cells (H-2^b) electrofused to primary C57BL/6 DCs (H-2^b) or cells from the JAWS II line (H-2^b). In these studies, the vaccines were tested under a therapeutic as opposed to a prophylactic regimen to better reflect the clinical situation. DBA/2 mice were first challenged with a lethal dose of M3 tumor cells (day 0) and were subsequently treated with the electrofusion vaccines (days 3, 7, and 14). As expected, the degree of tumor growth inhibition of established tumor cells was not as great as that observed in a prophylactic setting but significant levels of tumor rejection and enhanced survival were obtained with all three vaccines compared to control ($p < 0.0011$, *Figure 4*). Interestingly, the vaccine generated with the two allogeneic lines produced the greatest degree of tumor growth inhibition although the difference in survival between the different vaccines did not reach statistical significance. These results suggest that fusion vaccines generated between well-characterized tumor lines and a DC cell line may represent a viable, practical alternative to the current custom-made, patient-specific vaccines.

Exploration of mechanism of action of allogeneic electrofusion vaccines

The ability of a fully allogeneic electrofusion vaccine to induce antitumor immunity suggests that reprocessing of the vaccine by host antigen-presenting cells is taking place. This concept is supported by results from additional studies demonstrating the existence of cross-presentation from allogeneic antigen-containing DCs. As shown in *Figure 5*, C57BL/6 mice (H-2^b)

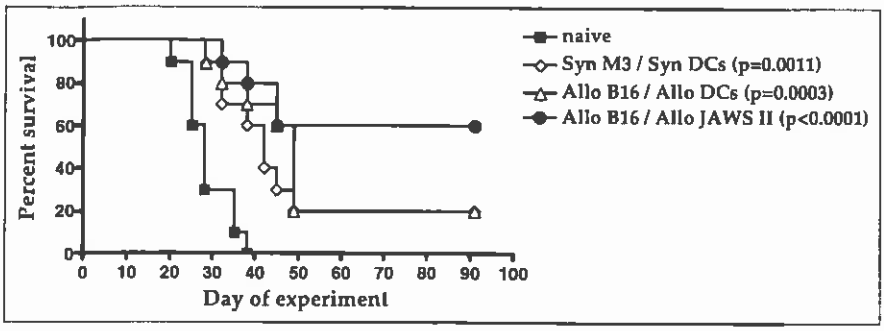


Figure 4. Therapeutic activity of an electrofusion vaccine generated with allogeneic tumor and DC cell lines. Electrofusion vaccines were prepared with either syngeneic components (M3 tumor cells and DBA/2 DCs, H-2^d) or fully allogeneic components comprising cultured B16 melanoma tumor cells (H-2^b) electrofused to primary C57BL/6 DCs (H-2^b) or cells from the JAWS II line (H-2^b; ATCC). Groups of 10 DBA/2 mice were first challenged with a lethal s.c. dose of M3 tumor cells and were then treated with 5.0×10^5 electrofused cells on days 3, 7, and 14. Tumor size was measured twice a week. All vaccine configurations induced statistically significant antitumor protection compared to the unvaccinated control mice (p values ranging from <0.0001 to 0.0011). The allogeneic cell line vaccine provided the greatest degree of protection (60% vs. 20% survival) but the difference with the other two vaccines was not statistically significant ($p = 0.17$).

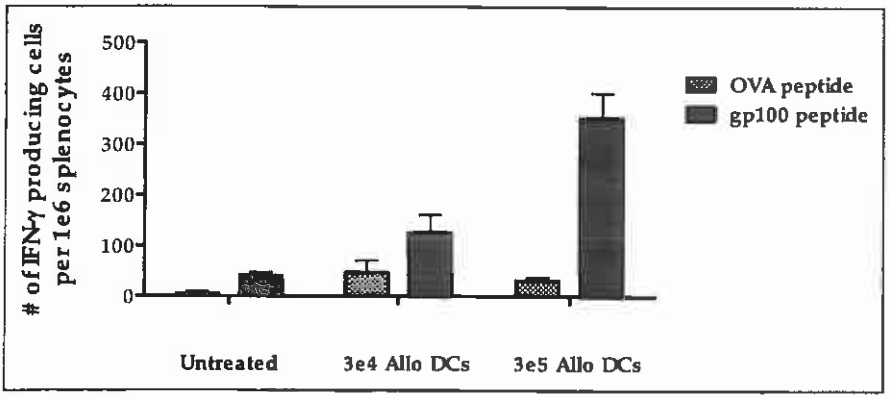


Figure 5. Antigen cross-presentation by allogeneic DCs. Bone marrow-derived DCs from Balb/c mice transduced with an adenovirus vector encoding gp100 were injected s.c. into C57BL/6 mice. The induction of gp100-reactivity in the spleen was assessed 2 weeks later by ELISPOT. Spleen cells were stimulated with an MHC Class I-restricted gp100 peptide or an irrelevant ovalbumin peptide as a negative control and the number of cells induced to produce IFN-γ was enumerated. Results shown are the mean ± SEM of triplicate wells.

immunized with fully allogeneic Balb/c (H-2^d) DCs transduced with an adenovirus vector expressing the gp100 melanoma antigen were induced to develop a specific CD8$^+$ T cell response against gp100 as determined by ELISPOT. The observed stimulation of antigen-specific T cells by allogeneic DCs incapable of direct MHC-restricted presentation implicates reprocessing and cross-presentation of antigen as operative mechanisms in the host.

As reported by ourselves and others, the DC component of fusion vaccines is required for efficacy and the antitumor response induced is not simply due to immunization with inactivated tumor cells.[3,4,6] The actual role of the DC component in an allogeneic fusion vaccine that is reprocessed by the host remains to be defined. As described above, the allogeneic DCs can act as a source of antigen for cross-presentation. It is also possible that the DCs (fused and unfused) may act by producing cytokines that deliver a "danger signal" and promote the development of cell-mediated immune responses. To test this possibility, culture supernatants from B16 tumor cells electrofused with DCs, compared to B16 tumor cells undergoing the electrofusion process alone, were collected and analyzed for cytokine content. The results indicate that DCs electrofused with tumor cells do in fact produce significant amounts of pro-inflammatory cytokines and chemokines with little or no cytokine release by tumors cells that underwent the electrofusion process alone (*Table 1*).

Discussion

Fusion vaccines generated from tumors cells and DCs represent a promising approach to cancer immunotherapy. However, their wide scale application involves significant practical and technical challenges. Production of electrofusion vaccines typically requires the acquisition and processing of patient and/or donor material and must be tailored to each individual. To address these issues, alternative vaccine configurations utilizing allogeneic tissue or cell lines were explored in this study. First, it was determined that cells from a cultured tumor line and cells from primary tumor samples both represent adequate fusion partners for DCs as the resulting electrofusion vaccines induced equivalent levels of protection from tumor challenge in the M3 melanoma tumor model. Cultured allogeneic tumor cell lines expressing tumor-associated antigens shared by a given tumor type could therefore be used to circumvent the practical issues surrounding surgical acquisition of tumor material from the patient. Indeed, the fusion product of allogeneic tumor cell lines and autologous or MHC-matched DCs has been reported to effectively induce human CD8$^+$ T cell responses *in vitro*[18-20] and induce tumor rejection in animals *in vivo*.[9] While this approach does not require the acquisition of tumor from the patient, it still requires the generation of autologous DCs from the patient or an MHC-matched donor. Because direct MHC-restricted antigen presentation by fusion hybrids has been postulated to account for the induction of antitumor immunity, testing of fully allogeneic hybrids, to our knowledge, has not

	B16 cells alone	B16/DBA DC	B16/C57 DC
IL-1	18 ± 5	1234 ± 64*	6509 ± 242*
IL-6	132 ± 42	100 ± 19	490 ± 49*
KC	376 ± 184	1631 ± 261*	8836 ± 87*
MCP-1	22 ± 4	128 ± 29	3916 ± 209*
MIP-1β	32 ± 7	81 ± 19*	1848 ± 21*
TNFα	29 ± 4	345 ± 80*	1647 ± 61*
IFN-γ	94 ± 6	103 ± 25	175 ± 23*
IL-12	157 ± 18	138 ± 14	192 ± 34

*$p < 0.05$ versus B16 cells alone.

Table 1. Cytokine release (pg/mL ± SD) from electrofused cells.

241

been reported. Therefore, the M3 melanoma model was used to test the potential of a fully allogeneic vaccine compared to fusion vaccines where one or both fusion partners are syngeneic to the host. Interestingly, our results clearly showed that fully allogeneic vaccines were as efficacious as syngeneic or semi-allogeneic vaccines and reproducibly provided equivalent levels of antitumor protection (*Figure 3*). Moreover, a DC-like cell line (JAWS II) was found to represent an adequate substitute for primary DCs in the fusion process (*Figure 4*).

The latter result suggests that "off-the shelf" fusion vaccines generated between well-characterized tumor and DC cell lines may represent a viable, practical alternative to the current custom-made, patient-specific vaccines. It has been generally assumed that the tumor/DC hybrids themselves are responsible for directly presenting antigen to host T cells in an MHC-restricted manner. Our findings clearly show that direct antigen presentation by the hybrids, although it may occur, is not an absolute requirement for induction of specific antitumor responses as evidenced by the activity of fully allogeneic vaccines. The most obvious explanation for this observation is the likely reprocessing of the vaccine by host antigen-presenting cells. The existence of such a mechanism is supported by our results showing induction of a gp100-specific response following immunization with fully allogeneic DCs expressing the gp100 antigen (*Figure 5*). In a recent review, Shu et al.[21] have similarly suggested that, given the low percentage of documented hybrids in most fusion vaccines, antigen presentation by host cells is likely to be involved in the process.

It must be noted that ourselves and others have observed that the DC component of the fusion vaccine is required for efficacy and that the antitumor response induced is not simply due to immunization with inactivated tumor cells.[3,4,6] The role of the DC component in a fusion vaccine that is reprocessed by the host remains to be defined. As suggested by the results in Table 1, it is possible that, rather than acting primarily as antigen-presenting cells, the DCs (fused and unfused) may also act by producing cytokines and chemokines that deliver a "danger signal" and promote the development of cell-mediated immune responses. The exact mechanism of action requires further investigation but the possibility of producing well-characterized fusion vaccines from allogeneic cell lines represents a significant advance in simplifying the manufacturing of electrofusion vaccines and expanding their applicability.

Conclusion

Our results indicate that fully allogeneic electrofusion vaccines can induce antitumor immunity equivalent to that of syngeneic or semi-allogeneic fusion products. This finding introduces the possibility of using defined allogeneic tumor and DC lines to simplify vaccine manufacturing.

Acknowledgments

The authors wish to thank the Genzyme Department of Comparative Medicine for the animal care provided.

References

1. Kaplan JM. New cancer vaccine approaches. *Drugs Today* 2004; 40: 913–929.

2. Gong J, Chen D, Kashiwaba M, Kufe D. Induction of antitumor activity by immunization with fusions of dendritic and carcinoma cells. *Nat Med* 1997; 3: 558–561.

3. Wang J, Saffold S, Cao X, Krauss J, Chen W. Eliciting T-cell immunity against poorly immunogenic tumors by immunization with dendritic cell-tumor fusion vaccines. *J Immunol* 1998; 161: 5516–5524.

4. Lespagnard L, Mettens P, Verheyden A-M, Tasiaux N, Thielemans K, van Meirvenne S, Geldhof A, De Baetselier P, Urbain J, Oberdan L, Moser M. Dendritic cells fused with mastocytoma cells elicit therapeutic antitumor immunity. *Int J Cancer* 1998; 76: 250–258.

5. Akasaki Y, Kikuchi T, Homma S, Abe T, Kufe D, Ohno T. Antitumor effect of immunizations with fusions of dendritic and glioma cells in a mouse brain tumor model. *J Immunother* 2001; 24: 106–113.

6. Siders WM, Vergilis KL, Johnson C, Shields J, Kaplan JM. Induction of specific antitumor immunity in the mouse with the electrofusion product of tumor cells and dendritic cells. *Mol Ther* 2003; 7: 498–505.

7. Orentas RJ, Schauer D, Bin Q, Johnson BD. Electrofusion of a weakly immunogenic neuroblastoma with dendritic cells produces a tumor vaccine. *Cell Immunol* 2001; 212: 4–13.

8. Hayashi T, Tanaka H, Tanaka J, Wang R, Averbook BJ, Cohen PA, Shu S. Immunogenicity and therapeutic efficacy of dendritic-tumor hybrid cells generated by electrofusion. *Clin Immunol* 2002; 104: 14–20.

9. Lee WT, Shimazu K, Kuriyama H, Tanaka H, Kjaergaard J, Shu S. Tumor-dendritic cell fusion as a basis for cancer immunotherapy. *Otolaryngol Head Neck Surg* 2005; 132: 755–764.

10. Yasuda T, Kamigaki T, Kawasaki K, Nakamura T, Yamamoto M, Kanemitsu K, Takase S, Kuroda D, Kim Y, Ajiki T, Kuroda Y. Superior anti-tumor protection and therapeutic efficacy of vaccination with allogeneic and semiallogeneic dendritic cell/tumor cell fusion hybrids for murine colon adenocarcinoma. *Cancer Immunol Immunother* 2007; 56: 1025–1036.

11. Kugler A, Stuhler G, Walden P, Zoller G, Zobywalski A, Brossart P, Trefzer U, Ullrich S, Muller CA, Becker V, Gross AJ, Hemmerlein B, Kanz L, Muller GA, Ringert R-H. Regression of human metastatic renal cell carcinoma after vaccination with tumor cell-dendritic cell hybrids. *Nature Med* 2000; 6: 332–336.

12. Haenssle HA, Krause SW, Emmert S, Zutt M, Kretschmer L, Schmidberger H, Andreesen R, Sorun A. Hybrid cell vaccination in metastatic melanoma. Clinical and immunologic results of a Phase I/II study. *J Immunother* 2004; 27: 147–155.

13. Avigan DE, Vasir B, George DJ, Oh WK, Atkins MB, McDermott DF, Kantoff PW, Figlin RA, Vasconcelles MJ, Xu Y, Kufe D, Bukowski RM. Phase I/II study of vaccination with electrofused allogeneic dendritic cells/autologous tumor-derived cells in patients with stage IV renal cell carcinoma. *J Immunother* 2007; 30: 749–761.

14. Trefzer U, Herberth G, Wohlan K, Milling A, Thiemann M, Sharav T, Sparbier K, Sterry W, Walden P. Tumour-dendritic hybrid cell vaccination for the treatment of patients with malignant melanoma: immunological effects and clinical results. *Vaccine* 2005; 23: 2367–2373.

15. Marten A, Renoth S, Heinicke T, Albers P, Pauli A, Mey U, Caspari R, Flieger D, Hanfland P, Von Ruecker A, Eis-Hubinger AM, Muller S, Schwaner I, Lohmann U, Heylmann G, Sauerbruch T, Schmidt-Wolf IGH. Allogeneic dendritic cells fused with tumor cells: preclinical results and outcome of a clinical phase I/II trial in patients with metastatic renal cell carcinoma. *Human Gene Ther* 2003; 14: 483–494.

16. Avigan D, Vasir B, Gong J, Borges V, Wu Z, Uhl L, Atkins M, Mier J, McDermott D, Smith T, Giallambardo N, Stone C, Schadt K, Dolgoff J, Tetreault J-C, Villaroel M, Kufe D. Fusion cell vaccination of patients with metastatic breast and renal cancer induces immunological and clinical responses. *Clin Cancer Res* 2004; 10: 4699–4708.

17. Kaplan JM, Yu Q, Piraino ST, Pennington SE, Shankara S, Woodworth LA, Roberts BL. Induction of antitumor immunity with dendritic cells transduced with adenovirus vector encoding endogenous tumor-associated antigens. *J Immunol* 1999; 163: 699–707.

18. Parkhurst MR, DePan C, Riley JP, Rosenberg SA, Shu, S. Hybrids of dendritic cells and tumor cells generated by electrofusion simultaneously present immunodominant epitopes from multiple human tumor-associated antigens in the context of MHC Class I and Class II molecules. *J Immunol* 2003; 170: 5317–5325.

19. Matsumoto S, Saito H, Tsujitani S, Ikeguchi M. Allogeneic gastric cancer cell-dendritic cell hybrids induce tumor antigen (carcinoembryonic antigen) specific CD8+ T cells. *Cancer Immunol Immunother* 2006; 55: 131–139.

20. Zhang Y, Ma B, Zhou Y, Zhang M, Qiu X, Sun Y, Zhang X, Ma B, Fan Q. Dendritic cells fused with allogeneic breast cancer cell line induce tumor antigen-specific CTL responses against autologous breast cancer cells. *Breast Cancer Res Treat* 2007; 105: 277–286.

21. Shu S, Zheng R, Lee WT, Cohen PA. Immunogenicity of dendritic-tumor fusion hybrids and their utility in cancer immunotherapy. *Crit Rev Immunol* 2007; 27: 463–483.

Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 12379–12383, December 1995
Immunology

The BALB/c mouse B-cell response to pigeon cytochrome *c* initiates as a heteroclitic response specific for the self antigen mouse cytochrome *c*

JEANNE M. MINNERATH*, L. PAUL WAKEM†, LINDA L. COMFORT†, FRED SHERMAN†, AND RONALD JEMMERSON*‡

*Department of Microbiology, University of Minnesota Medical School, Minneapolis, MN 55455; and †Department of Biochemistry, University of Rochester Medical School, Rochester, NY 14642

Contributed by Fred Sherman, August 21, 1995

ABSTRACT **Direct evidence is presented in support of the longstanding but unproven hypothesis that B lymphocytes specific for self antigens (Ags) can be used in the immune response to foreign Ags. We show that the B cells in BALB/c mice responding early to pigeon cytochrome *c* (CYT) produce antibodies that recognize and bind the major antigenic site on mouse CYT with greater affinity than they bind pigeon CYT—i.e., they are heteroclitic for the self Ag. Furthermore, these B cells express the same combination of immunoglobulin variable region (*V*) genes that are known to be used in B-cell recognition of mouse CYT. Over time, the response to pigeon CYT becomes more specific for the foreign Ag through the recruitment of B cells expressing different combinations of *V* genes and, possibly, somatic mutation of the mouse CYT-specific B cells from early in the response. Cross-recognition of pigeon CYT by mouse CYT-specific B cells results from the sharing of critical amino acid residues by the two Ags. Although B-cell recognition of the self Ag, mouse CYT, is very specific, which limits the extent to which foreign Ags can cross-activate the autoreactive B cells, it is possible that polyreactive B cells to other self Ags may be used more frequently in response to foreign Ags.**

Self antigen (Ag) interaction with receptors on B lymphocytes is known to lead to immunological tolerance (1), whereby the self Ag-reactive cells are either eliminated [i.e., deleted (2, 3)] or functionally inactivated [i.e., rendered anergic (4, 5)]. However, if the self-Ag concentration is low, autoreactive B cells are not necessarily tolerized (6, 7). In fact, there is a large body of data showing that functional autoreactive B cells persist in healthy individuals (8, 9). Although these B cells may secrete natural autoantibodies that populate normal blood (10–12), their role in the immune system is not fully understood. It has been proposed that autoreactive B cells, through cross-recognition and subsequent somatic mutation, may be used in the immune response to foreign Ags (13, 14).

There is some evidence supporting this hypothesis. For example, antibodies produced early in the immune response to the hapten *p*-azophenylarsonate (Ars) bind the self Ag, DNA, and Ars, while later in the response the antibodies bind only Ars (15). However, it is not clear that the primary B cells are actually specific for DNA or that the B cells in the primary response are also recruited into the secondary response. Neither DNA-binding antibodies from the preimmune repertoire (15) nor those observed in autoimmune-prone mice (16) derive from the same combination of immunoglobulin variable region genes (*V*) as the Arg-specific monoclonal antibody (mAb). Whether B cells specific for a self Ag can, in fact, be used in response to foreign Ags remains to be demonstrated.

Cytochromes *c* (CYTs) are useful Ags to address this question. Recently, we reported that B cells of mice do not appear to be tolerant to mouse CYT (17). Indeed, early (4 days) in the B-cell response of BALB/c mice to CYT coupled to ovalbumin (OVA; CYT-OVA), the frequency of B cells responding to the self Ag is significantly higher than the frequency of B cells responding to foreign CYTs (17). The only foreign CYTs that elicit a strong B-cell response early are those that cross-react with mAbs specific for mouse CYT and, thus, have the potential to activate the autoreactive B cells.

In the present study we asked whether the B cells responding early to the closely related pigeon CYT are, in fact, specific for the self Ag. This was examined by characterizing the mAbs elicited by pigeon CYT at different times during the response in terms of affinities for pigeon versus mouse CYT, fine specificities, and immunoglobulin gene usage. We found that B cells specific for mouse CYT respond early to pigeon CYT and that as the response progresses, the responding B cells become more specific for pigeon CYT. Evidence is presented suggesting that this progression is due to the recruitment of B cells expressing different heavy (H) chain V_H and light (L) chain V_κ gene combinations and, possibly, somatic mutation of the V_H and V_κ genes used in response to mouse CYT. The extent to which CYT-specific autoreactive B cells may be used in response to less closely related Ags is also examined.

MATERIALS AND METHODS

Pigeon, rat, rabbit, and chicken CYTs were purchased from Sigma. Since rat and mouse CYTs are identical (18), rat CYT will be referred to herein as the self Ag, mouse CYT. All of these CYTs were further purified on carboxymethyl-Sephadex G-50 (17). Guanaco CYT was provided by E. Margoliash (The University of Illinois, Chicago), and yeast iso-2 CYT was provided by B. Nall (University of Texas Health Science Center, San Antonio).

Pigeon CYT was also prepared from the strain of *Saccharomyces cerevisiae* B-7840, which contains the synthetic gene *CYC1-PIGEON* (19). Plasmid pAB855 and the oligonucleotides listed in Table 1 were used to construct, by oligonucleotide-directed mutagenesis (20), the genes encoding the altered forms of pigeon CYT having the following replacements: Glu-44 → Pro(E44P), Gly-60 → Lys(G60K), and Ala-89 → Asp(A89D). The alterations were confirmed by DNA sequencing of the appropriate regions. The normal and variant pigeon CYTs were extracted from yeast and purified as described (21). Acetylated and unacetylated forms of CYT were separated in the final purification step by ion-exchange HPLC (21).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Ag, antigen; Ars, *p*-azophenylarsonate; CDR, complementarity-determining region; CYT, cytochrome *c*; H chain, immunoglobulin heavy chain; L chain, immunoglobulin light chain; mAb, monoclonal antibody; OVA, ovalbumin; V_H, heavy chain variable region; V_κ, kappa light chain variable region.

‡To whom reprint requests should be addressed.

§The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L42265–L42276).

243

Table 1. Variant oligonucleotides used in site-directed mutagenesis of the pigeon CYT gene

Oligo-nucleotide	Sequence	Site-directed changes: Nucleotides	Site-directed changes: Amino acids
OL93.174	GGTCAAGCTCCAGGGTTCTC	GAA → CCA	E44P
OL91.250	TTACTTGGAAGGAAGACAC	GGT → AAG	G60K
OL91.251	AGAAGAAGGACGAAAGAGC	GCT → GAC	A89D

B lymphocytes secreting mAbs reactive with pigeon CYT were activated in splenic fragment cultures (17, 22). Briefly, BALB/c mice were immunized with 50 μg of pigeon CYT-OVA, and at various times splenic B lymphocytes were transferred in limiting dilution to hemocyanin-primed BALB/c mice. The recipients' spleens were chopped 18 hr later and cultured for 3 days with pigeon CYT coupled to hemocyanin. Culture fluid was collected every 3 days. B-cell hybridomas were prepared from Ag-primed mice by fusing splenic lymphocytes with the myeloma cells P3X63Ag8.653 by using PEG (23). To obtain mAbs prior to somatic mutation so that the V_H and V_κ genes could be defined, mice were primed with OVA and challenged for 4 days with pigeon CYT-OVA (17).

The specificities of the mAbs were determined by ELISA (17). Two different assays were performed. For indirect ELISA, mAb binding to microtiter plates coated with the variants of CYT was detected by using goat anti-mouse IgG conjugated to horseradish peroxidase (Sigma). In competitive ELISA, different soluble CYTs were examined for their ability to inhibit the binding of the mAbs to the plate-adsorbed Ag.

Nucleotide sequences of mAb H and L chain cDNA were determined as described elsewhere (17, 24).

RESULTS

mAbs Elicited Early in the B-Cell Response to Pigeon CYT Are Heteroclitic for Mouse CYT. Essentially all of the B cells from BALB/c mice that respond early (day 4) to the self Ag, mouse CYT, recognize a single site on the Ag around residue 62 (17). A single amino acid change at that position, Asp-62 → Glu, inhibits the binding of mAbs produced by these cells. Most of these mAbs (70%) bind pigeon CYT, which has an aspartic acid residue at position 62, although with lower affinity because of amino acid differences from mouse CYT at other positions in the antigenic site.

Surprisingly, when BALB/c mice were immunized with pigeon CYT, mAbs from the vast majority of the B cells responding early bound mouse CYT with higher affinity than they bound pigeon CYT—i.e., these mAbs were heteroclitic for the self Ag. Results of a competitive ELISA for a typical mAb elicited early by pigeon CYT (day 8 after immunization) are shown in Fig. 1. Since the mAb bound identically to mouse and rabbit CYT and bound weakly if at all to guanaco CYT, it recognized the region around residue 62. This is the only residue position at which rabbit and mouse CYTs are identical, and both differ from guanaco CYT (see Table 2). The lower affinity binding to pigeon CYT indicates that one or more of residues 3, 100, 103, and 104, where mouse and rabbit CYT are identical and differ from pigeon CYT, are involved in mAb recognition. The mAbs elicited early in the response to pigeon CYT have the same specificity as the mAbs elicited against mouse CYT (17). These results show that the B cells responding early to pigeon CYT are specific for mouse CYT and suggest that these are the same B cells that respond to the self Ag.

The Same *V* Genes Are Used by B Cells Responding Early to Pigeon CYT as Are Used by B Cells in the Response to the Self Ag, Mouse CYT. The dominant *V* genes used to encode mAbs elicited against the major antigenic site on mouse CYT

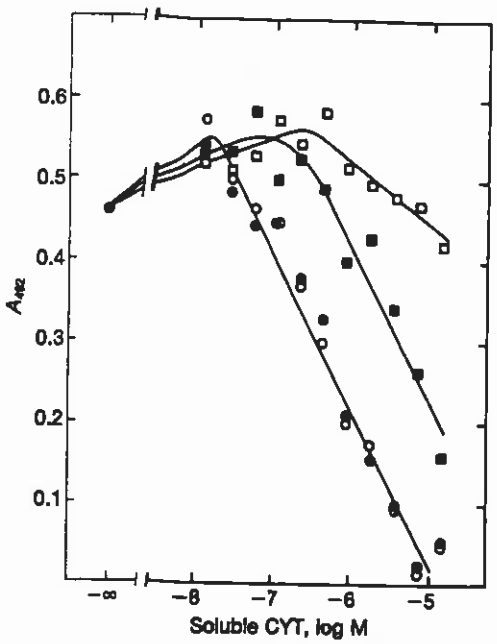


FIG. 1. Competitive ELISA of a mAb obtained 8 days after injecting BALB/c mice with pigeon CYT-OVA. mAbs, such as the one tested here, were generated in splenic fragment cultures. For the competitive ELISA, pigeon CYT was adsorbed to microtiter plates, and the ability of several CYTs to inhibit mAb binding to the plate-adsorbed Ag was examined. ELISA readings were averaged from duplicate trials and background readings were subtracted. The same results were observed when mouse CYT was adsorbed to the microtiter plates. The soluble CYTs tested for inhibition included mouse (●), rabbit (○), pigeon (■), and guanaco (□).

(17) are the germ-line V_H gene *19.1.2* (26) and the germ-line V_κ gene *R9* (27). The cDNA sequences for the H and L chains of the two mAbs, 5D9 and 6A5, from B-cell hybridomas generated early in the response to pigeon CYT show that these same *V* genes were used (Fig. 2). The mAb 1F6 derives from the same V_H gene and a related V_κ gene (*H3*; ref. 27). Although this V_κ segment has not been observed among anti-mouse CYT-specific mAbs, it could be used infrequently. In fact, several different V_κ genes are known to pair with the V_H gene *19.1.2* to encode specificity for mouse CYT (17). The variability observed in complementarity-determining region (CDR) 3 among the three H chain cDNA sequences (Fig. 2*A*) is consistent with that previously seen among cDNA sequences of 25 different mAbs elicited against mouse CYT that derive from the V_H gene *19.1.2* (ref. 17; data not shown). None of these mAbs are identical in H chain CDR3. Variability in CDR3 of the L chain was also observed among the mouse CYT-specific mAbs, although the L chain of mAb 5D9 (Fig. 2*B*) is identical to the L chain of the mouse CYT-specific mAb 1G2 (17). The identity in *V* gene usage in B cells responding early to pigeon and mouse CYT (for at least two of three anti-pigeon CYT-responding B cells) and the similar affinities of their mAbs for mouse CYT (data not shown) strongly suggest that they are the same population of B cells (see *Discussion*).

mAbs Elicited Against Pigeon CYT Become More Specific for Pigeon CYT and Less Heteroclitic for Mouse CYT as the B-Cell Response Progresses. Previously we had found that as the B-cell response to mouse CYT progressed, mAbs binding other antigenic sites began to be observed in addition to those binding the major site. To simplify analysis of any changes in anti-pigeon CYT mAb specificity over time, we focused on the major site. Variants of pigeon CYT obtained from yeast transformed with the wild-type or mutated pigeon CYT genes were used to identify mAbs that bound this region. The mAbs

Table 2. Amino acid sequences of CYTs employed in this study

CYT	Amino acid residue position					
	−9	1 10	20	30	40	50
Mouse		GDVEKGKKIF	VQKCAQCHTV	EKGGKHKTGP	NLHGLFGRKT	GQAAGFSYTD
Rabbit		----------	----------	----------	----------	---V------
Guanaco		----------	----------	----------	----------	---V------
Pigeon		--I-------	----S-----	----------	----------	---E------
Chicken		--I-------	----S-----	----------	----------	---E------
Yeast iso-2	AKESTGFKP	-SAK--ATL-	KTR-Q----I	-E--PN-V--	----I---HS	--VK-Y----
	60	70	80	90	100	104
Mouse	ANKNKGITWG	EDTLMEYLEN	PKKYIPGTKM	IFAGIKKKGE	RADLIAYLKK	ATNE
Rabbit	----------	----------	----------	--------D-	----------	----
Guanaco	----------	-E--------	----------	----------	----------	----
Pigeon	----------	----------	----------	--------A-	---------Q	--AK
Chicken	----------	----------	----------	--------S-	-V-------D	--SK
Yeast iso-2	--I--NVK-D	--SMS---T-	----------	A---L--EKD	-N---T-MT-	-AKδ*

The amino acid sequences are as previously listed (25).
*δ, Deletion of residue 104.

elicited early in the response to pigeon CYT bound poorly in ELISA to the variant with a Gly-60 → Lys change and to unacetylated pigeon CYT compared with wild-type pigeon CYT and other variants of pigeon CYT that were tested, including a variant with an Ala-89 → Asp change (results not shown). This is also the phenotype of mAbs elicited against the major antigenic site on mouse CYT (ref. 29; unpublished results). This same phenotype is dominant among mAbs obtained later in the response to pigeon CYT. However, other sites were also observed later in the immune response. For example, a small number of mAbs late in the response bound poorly a variant of pigeon CYT with a Glu-44 → Pro change compared with their binding to pigeon CYT and were not affected by amino acid substitutions elsewhere on the Ag (results not shown).

A

mAb	94	CDR3										
	R	H	Y	Y	G	S	S	S	F	A	Y	
19.1.2	:AGG	CAT	TAC	TAC	GGT	AGT	AGC	TCG	TTT	GCT	TAC	J_H3
		P										
5D9	1°:--A	-C.	...	...	...	...	...	..-	---	---	---	J_H3
	T	L	W	D		D	Y	Y		D		
6A5	1°:--- ACC	-TC	-GG	G--	--G	GAC	TA-	-AC	---	-AC	---	J_H2
		G		D			A	W				
1F6	1°:---	GGG	---	G--	...	...	GC-	-G-	---	---	---	J_H3

B

mAb	30	40	71	CDR3								
	I	P	Y	H	Q	R	S	S	Y	P	L	T
R9	:ATA	CCA	TAT	CAT	CAG	CGG	AGT	AGT	TAC	CCA	CTC	ACG	J_k5
5D9	1°:---	---	---	---	---	---	---	---	---	--.	..-	---	J_k5
											C		
6A5	1°:---	---	---	---	---	---	---	---	---	---	TG-	---	J_k2
	V	S		Q		W			N		P		
1F6	1°:G--	T--	--C	--G	---	T--	---	---	A--	---	-C-	---	J_k5

FIG. 2. cDNA and deduced amino acid sequences of anti-pigeon CYT mAb H chains (*A*) and L chains (*B*) from B-cell hybridomas prepared 4 days after challenging OVA-primed BALB/c mice with pigeon CYT-OVA. The H chains are compared to the H chain of an α(1 → 6) dextran-specific mAb derived from the germ-line V_H gene *19.1.2* (26), and the L chains are compared to the L chain of a lymphoma-produced mAb derived from the germ-line $V_κ$ gene *R9* (28). The mAbs are identical in the V_H segment and differ only in CDR3. Dashes indicate identity with the sequences used for comparison, and dots indicate nucleotide deletions. The GenBank accession numbers for the H and L chain sequences, respectively, are as follows: for mAb 5D9, L42271 and L42272; for mAb 6A5, L42275 and L42276; and for mAb 1F6, L42265 and L42266.

Employing the variants of pigeon CYT, we were readily able to identify mAbs early and late in the immune response that recognized the major antigenic site. These mAbs then were examined for their binding to mouse CYT relative to pigeon CYT by using indirect ELISA. The mAbs were grouped into three categories: those that bound mouse CYT with higher affinity than pigeon CYT [mouse CYT > pigeon CYT (M > P)], those that bound the two Ags similarly [mouse CYT = pigeon CYT (M = P)], and those that bound pigeon CYT with higher affinity than the self Ag [mouse CYT < pigeon CYT (M < P)]. The fraction of the total mAb with the heteroclitic phenotype (M > P) decreased as the B-cell response to pigeon CYT progressed, while the mAb fraction with the opposite phenotype (M < P) increased (Fig. 3). Those mAbs at days 8 and 10 that bound pigeon and mouse CYTs identically in indirect ELISA (M = P) were also found by competitive ELISA to be heteroclitic for mouse CYT (on average one-third as much mouse CYT as pigeon CYT was required to inhibit binding of the mAbs to plate-adsorbed mouse or pigeon CYT). In contrast, at days 45 and 90, this category of mAbs was found by competitive ELISA to have a higher affinity for pigeon CYT than for mouse CYT (on average 1/10th as much pigeon CYT as mouse CYT was required to inhibit binding of the mAbs to plate-adsorbed mouse or pigeon CYT). These results indicate that the mAbs were more specific for pigeon CYT later in the response than they were early on.

Varied *V* Gene Usage by B Cells Late in the Response to Pigeon CYT. The shift in specificity from self to foreign CYT during the B-cell response to pigeon CYT could be due to somatic mutation of the *V* genes encoding mouse CYT-specific mAbs, recruitment of B cells using different *V* genes, or both. To begin to address this issue, cDNA sequences were determined for the H and L chains of three mAbs obtained three months into the response to pigeon CYT. All of these mAbs bound the major antigenic site and had a higher affinity for pigeon CYT than for mouse CYT (data not shown).

Two of these mAbs, 1G7 and 5A3, derive from different V_H and $V_κ$ genes than are known to be used in the response to mouse CYT (data not shown; see GenBank accession numbers L42267–L42270). Not only do they differ at a large number of nucleotide positions, but the lengths of the H-chain CDR1 and CDR2 and L-chain CDR1 are different from the lengths of the corresponding CDRs of mouse CYT-specific mAbs. Therefore, these mAbs derive from different B cells than are used either early in the immune response to pigeon CYT or in the response to mouse CYT.

The V_H gene of the third mAb, 5H11, could derive from the V_H gene *19.1.2* (see Fig. 4*A*). They differ at 10 nucleotides in the V_H segment (which does not include CDR3 or the nucle-

245

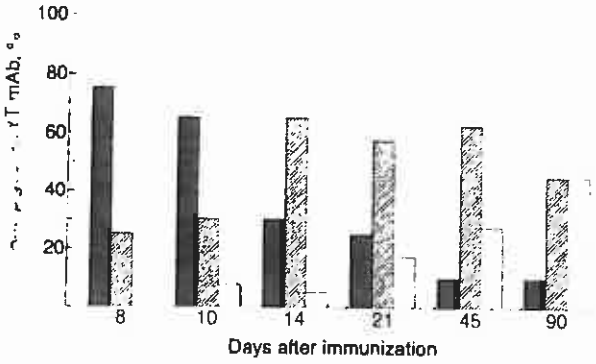


FIG. 3. Relative binding of anti-pigeon CYT mAbs to pigeon and mouse CYTs at various times during the B-lymphocyte response to pigeon CYT-OVA. ■, Binding to mouse CYT > pigeon CYT; ■, binding to mouse CYT = pigeon CYT; □, binding to mouse CYT < pigeon CYT. Binding was assessed by indirect ELISA. Results are shown for mAbs binding the major antigenic site around residue 62. These mAbs represented the total population of anti-pigeon CYT mAbs at day 8 and by day 90 represented approximately half of the total population. The mAbs were obtained from splenic fragment cultures. B cells were harvested from immunized mice at the indicated times after immunization, and functionally cloned cells were cultured for 2–3 weeks *in vitro*. The frequency of B lymphocytes responding to pigeon CYT in the splenic fragment assay varied from ~60 per 10^8 cells transferred early on to 200 per 10^8 at the later time points. If the A_{492} reading for a mAb binding mouse CYT was within 20% of the reading for the mAb binding pigeon CYT, then the mAb was considered to bind both Ags equally (mouse CYT = pigeon CYT). The numbers of mAbs tested at each time point were as follows: day 8, n = 17; day 10, n = 36; day 14, n = 33; day 21, n = 11; day 45, n = 29; and day 90, n = 9.

otide 5′ of CDR3), possibly a result of somatic mutation. It is also possible that this V_H gene is distinct from the gene *19.1.2*, since they are both members of the large and closely related J558 family of V_H genes (30). While the V_κ gene encoding the L chain of mAb 5H11 differs from the V_κ gene *R9*, it appears to derive from the V_κ gene *2G5* (Fig. 4*B*), which is another common gene used in the response to mouse CYT (17). The single nucleotide difference in codon 9 could be due to an error in the PCR (31) or to somatic mutation.

mAbs Elicited Early in the B-Cell Responses to Other CYT Are Not Heteroclitic for Mouse CYT. The cross-reaction of mouse CYT-specific B cells with pigeon CYT could be explained chemically by the fact that these two CYTs share the critical aspartic acid residue at position 62 and have identical residues at other positions in that vicinity (Table 2). Chicken and yeast iso-2 CYTs also have an aspartic acid residue at position 62 (Table 2). To determine whether the early B-cell responses to these Ags were also mouse CYT-specific, we analyzed the specificity of mAbs elicited against chicken and yeast iso-2 CYTs that we had prepared in a previous study (17). Only one of four mAbs to chicken CYT was heteroclitic for the self Ag, and none of six mAbs to yeast iso-2 CYT was heteroclitic. These results correlate with the extent to which we found that chicken and yeast CYTs cross-react with anti-mouse CYT mAbs: 19% of these mAbs bound chicken CYT with reduced affinity, and none of them bound yeast iso-2 CYT. In contrast, 70% of anti-mouse CYT mAbs bound pigeon CYT.

DISCUSSION

Previously we found that in BALB/c mice primed with OVA and challenged with various CYT-OVA conjugates, the frequency of B lymphocytes responding early (day 4) to the self Ag, mouse CYT, was 1.5- to 3-fold higher than the frequency of B cells responding to closely related CYTs and >10-fold higher than the frequency of cells responding to more foreign CYTs (17). Since the closely related CYTs (rabbit and pigeon CYTs) cross-react with a large number of anti-mouse CYT mAbs, we hypothesized that the B cells activated early by these Ags are actually specific for the self Ag, mouse CYT. In this report we provided strong evidence that the CYT-specific B cells responding early to pigeon CYT-OVA are the same population of B cells used in the response to mouse CYT. Thus, mAbs produced by the pigeon CYT-activated B cells are heteroclitic for mouse CYT, and the V_H and V_κ genes expressed by those cells are dominant in the B-cell response to mouse CYT. The extreme variability in CDR3 of mouse CYT-specific mAbs that derive from the V_H gene *19.1.2* does not allow identity in CDR3 to be used to relate these two populations of cells. However, for these to be different cell populations, one would have to argue that mouse CYT cannot activate the B cells responding early to pigeon CYT, even though the self Ag binds the Ag receptors on those cells with higher affinity than does pigeon CYT. Furthermore, the variability in CDR3 suggests that Ag recognition is largely, if not entirely, encoded within the V gene segments.

As the response to pigeon CYT-OVA progressed, the mAbs produced by the B cells became less heteroclitic for mouse CYT and more specific for pigeon CYT. This change appears to be due to recruitment of B cells expressing different combinations of V_H and V_κ genes and, possibly, to somatic mutation of the V genes in B cells that are used in response to mouse CYT. Of the three mAbs from late in the response to pigeon CYT that we sequenced, two derive from different V genes than any of 29 mouse CYT-specific mAbs that we have sequenced, most of which derive from the V_H gene *19.1.2* and the V_κ gene *R9*. The third mAb derives from a V_κ gene that is frequently used in the response to mouse CYT and may derive from the dominant V_H gene *19.1.2*. While there are 10 nucleotide differences in V_H from the germ-line gene, a comparable number of differences has been seen in secondary mAbs elicited against mouse CYT (17), and this number is within the

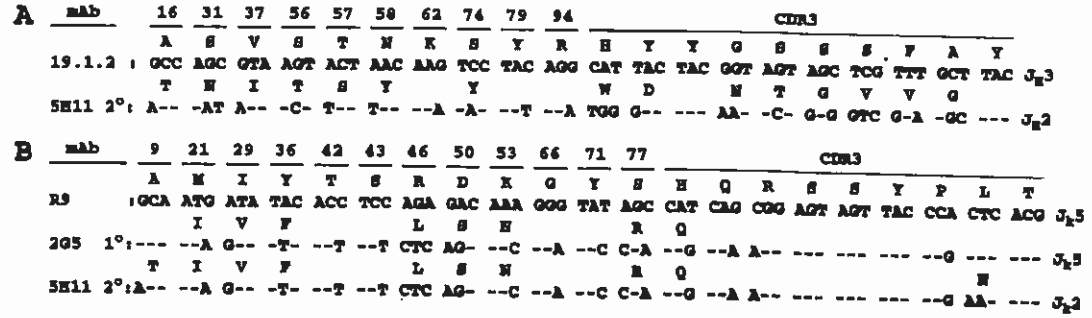


FIG. 4. cDNA and deduced amino acid sequences of the pigeon CYT-specific mAb 5H11 H chain (*A*) and L chain (*B*). The 5H11 hybridoma was prepared 3 months after injecting BALB/c mice with pigeon CYT-OVA. See the legend to Fig. 2 for details of the sequence comparisons shown here. The L chain cDNA sequence of the mouse CYT-specific mAb 2G5 is also given. Like *R9*, the V_κ gene *2G5* is frequently used in response to mouse CYT (17). The GenBank accession numbers for 5H11 mAb H and L chain sequences, respectively, are L42273 and L42274.

range of differences that can be expected as a result of somatic mutation (32). Further study is required to determine to what extent, if any, somatic mutation is a mechanism for the change in specificity during the response to pigeon CYT.

In addition to pigeon CYT, two other CYTs that share the critical Asp-62 residue with mouse CYT, chicken and yeast iso-2 CYTs, were studied to determine if they, too, would elicit a heteroclitic response. Although chicken CYT is very similar to pigeon CYT, differing only at residues 89, 92, 100, and 103 (Table 2), it activated ≈70% fewer B cells than did pigeon CYT early in the response and only one of four mAbs tested was heteroclitic for mouse CYT. Since residue 89 is not involved in mAb binding to mouse CYT, one or more of the amino acid changes at positions 92, 100, and 103 must be responsible for the differences in autoreactive B-cell activation between pigeon and chicken CYTs. Yeast iso-2 CYT did not activate any of the mouse CYT-specific B cells and is even more different than chicken CYT from pigeon and mouse CYTs (Table 2).

Our finding that the B-cell response to pigeon CYT initiates by activating B cells specific for the self Ag, mouse CYT, supports the hypothesis that autoreactive B cells can be used in response to foreign Ags (13, 14). While it is possible that some of these autoreactive B cells may undergo somatic mutation and become more specific for the foreign Ag, other B cells expressing different combinations of *V* genes are recruited as the response progresses. Thus, the autoreactive B cells serve as a first line of defense (14). The recognition of mouse CYT by autoreactive B cells is very specific; a single conservative amino acid substitution at position 62 blocks B-cell recognition. This contrasts with other autoreactive B cells that are polyreactive (33). Those B cells may be used more frequently in the response to foreign Ags than we have observed for autoreactive B cells specific for mouse CYT.

We are grateful to Drs. E. Margoliash and B. Nall for their gifts of CYTs. J.M.M. was supported by training grant AI-07313 from the U.S. Public Health Service, National Institutes of Health. This work was funded by Grant 3282 from The Council for Tobacco Research, U.S.A.-Inc. (to R.J.) and Grant R01-GM12702 from the National Institutes of Health (to F.S.).

1. Nossal, G. J. V. (1994) *Cell* **76**, 229–239.
2. Nemazee, D. A. & Bürki, K. (1989) *Nature (London)* **337**, 562–566.
3. Hartley, S. B., Crosbie, J., Brink, R., Kantor, A. B., Basten, A. & Goodnow, C. C. (1991) *Nature (London)* **353**, 765–769.
4. Goodnow, C. C., Crosbie, J., Adelstein, S., Lavoie, T. B., Smith-Gill, S. J., Brink, R. A., Pritchard-Briscoe, H., Wotherspoon, J. S., Loblay, R. H., Raphael, K., Trent, R. J. & Basten, A. (1988) *Nature (London)* **334**, 676–682.
5. Erikson, J., Radic, M. Z., Camper, S. A., Hardy, R. R., Carmack, C. & Weigert, M. (1991) *Nature (London)* 349, 331–334.
6. Goodnow, C. C., Crosbie, J., Jorgensen, H., Brink, R. A. & Basten, A. (1989) *Nature (London)* **342**, 385–391.
7. Adelstein, S., Pritchard-Briscoe, H., Anderson, T. A., Crosbie, J., Gammon, G., Loblay, R. H., Basten, A. & Goodnow, C. C. (1991) *Science* **251**, 1223–1225.
8. Prabhakar, B. S., Saegusa, J., Onodera, T. & Notkins, A. L. (1984) *J. Immunol.* **133**, 2815–2817.
9. Portnoï, D., Freitas, A., Holmberg, D., Bandeira, A. & Coutinho, A. (1986) *J. Exp. Med.* **164**, 25–35.
10. Grabar, P. (1975) *Clin. Immunol. Immunopathol.* **4**, 453–466.
11. Souroujon, M., White-Scharf, M. E., André-Schwartz, J., Gefter, M. L. & Schwartz, R. S. (1988) *J. Immunol.* **140**, 4173–4179.
12. Avrameas, S. & Ternynck, T. (1993) *Mol. Immunol.* **30**, 1133–1142.
13. Cunningham, A. J. & Steele, E. J. (1979) in *B Lymphocytes in the Immune Response*, eds. Cooper, M., Mosier, D., Scher, A. & Vitetta, E. (Elsevier–North Holland, Amsterdam), pp. 187–201.
14. Stollar, B. D. (1991) *Mol. Immunol.* **28**, 1399–1412.
15. Naparstek, Y., André-Schwartz, J., Manser, T., Wysocki, L. J., Breitman, L., Stollar, B. D., Gefter, M. & Schwartz, R. S. (1986) *J. Exp. Med.* **164**, 614–626.
16. Radic, M. Z. & Weigert, M. (1994) *Annu. Rev. Immunol.* **12**, 487–520.
17. Minnerath, J. M., Mueller, C. M., Buron, S. & Jemmerson, R. (1995) *Eur. J. Immunol.* **25**, 784–791.
18. Carlson, S. S., Mross, G. A., Wilson, A. C., Mead, R. T., Wolin, L. D., Bowers, S. F., Foley, N. T., Muijsers, A. O. & Margoliash, E. (1977) *Biochemistry* **16**, 1437–1442.
19. Hickey, D. R., Jayaraman, K., Goodhue, C. T., Shah, J., Fingar, S. A., Clements, J. M., Hosokawa, Y., Tsunasawa, S. & Sherman, F. (1991) *Gene* **105**, 73–81.
20. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987) *Methods Enzymol.* **154**, 367–382.
21. Margoliash, E., Schejter, A., Koshy, T. I., Luntz, T. L. & Garber, E. A. E. (1990) in *Proceedings of the International Symposium on Bioenergetics: Molecular Biology, Biochemistry and Pathology*, ed. Kim, C. H. (Plenum, NY), pp. 125–145.
22. Klinman, N. R. (1969) *Immunochemistry* **6**, 747–759.
23. Kearney, J. F., Radbruch, A., Liesegang, B. & Rajewsky, K. (1979) *J. Immunol.* **123**, 1548–1550.
24. Goshorn, S. C., Retzel, E. & Jemmerson, R. (1991) *J. Biol. Chem.* **266**, 2134–2142.
25. Borden, D. & Margoliash, E. (1976) in *Handbook of Biochemistry and Molecular Biology, Proteins*, ed. Fasman, G. D. (Chemical Rubber Co., Cleveland), Vol. 4, pp. 268–279.
26. Akolkar, P. N., Sikder, S. K., Bhattacharya, S. B., Liao, J., Gruezo, F., Morrison, S. L. & Kabat, E. A. (1987) *J. Immunol.* **138**, 4472–4479.
27. Even, J., Griffiths, G. M., Berek, C. & Milstein, C. (1985) *EMBO J.* **4**, 3439–3445.
28. Roth, M. S., Weiner, G. J., Allen, E. A., Terry, V. H., Harnden, C. E., Boehnke, M., Kaminski, M. S. & Ginsburg, D. (1990) *J. Immunol.* **145**, 768–777.
29. Minnerath, J. M., Crump, B., Margoliash, E. & Jemmerson, R. (1995) *Mol. Immunol.* **32**, 795–803.
30. Brodeur, P. H. & Riblet, R. (1984) *Eur. J. Immunol.* **14**, 922–930.
31. Jacob, J., Przylepa, J., Miller, C. & Kelsoe, G. (1993) *J. Exp. Med.* **178**, 1293–1307.
32. Clarke, S. H., Huppi, K., Ruezinsky, D., Staudt, L., Gerhard, W. & Weigert, M. (1985) *J. Exp. Med.* **161**, 687–704.
33. Schwartz, R. S. & Stollar, B. D. (1994) *Immunol. Today* **15**, 27–32.

247

HIV infection

Human immunodeficiency virus (HIV) was discovered during the 1980's. If untreated, HIV infection eventually results in a disease known as acquired immune-deficiency syndrome (AIDS). The life cycle of HIV in the human cystem is summarized on the next page. At step 10, the viral coat protein SU or gp120 first binds the human cell surface molecule CD4 and then subsequently binds another surface molecule CCR5 to initiate the fusion process between the vrus and the human cell. Since the first human cell surface receptor is CD4, the primary target of HIV is thus the $CD4^+$ or helper T cells.

Eventually, the dsDNA derived from the cDNA synthesized from the viral RNA will be integrated into the human cell chromosome. The virus then takes over the human cell machinery and uses it as a factory to manufacture many copies of the virus. Many helper T cells are fused together to form a syncytium and finally die off. As the disease progresses, the number of $CD4^+$ cells reduces rapidly. The biological processes requiring various T_h cells grind to a halt. Antibodies are not produced effectively, macrophages not stimulated, etc. That is the reason why the disease entity is known as AIDS.

To understand the initial entry of the virus into human cells, the structure of SU or gp120 has been studied extensively. A fragment of it, consisting of about 65 % of the entire molecule, has been crystallized and its three dimensional structure determined (Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA, Sodroski JG (1995) The antigenic structure of the HIV go120 envelope glycoprotein. Nature, 393, 705-711). There is a pocket on its surface that binds the D1 region of CD4 very tightly (Xie H, Ng D, Savinov SN, Kwong PD, Wyatt R, Smith AB, III, Hendrickson WA (2007) Structure-activity relationships in the binding at chemically derivatized CD4 to gp120 human immunodeficiency virus. J Med Chem, 50, 4898-4908). As a result, CD4 can no longer bind to the A2 and B2 regions of the MHC class II molecule thus disrupting the interaction between APC and helper T cells.

Various medications have been developed during the past couple of decades to disrupt the life cycle of the HIV. That includes:

Fusion inhibitors

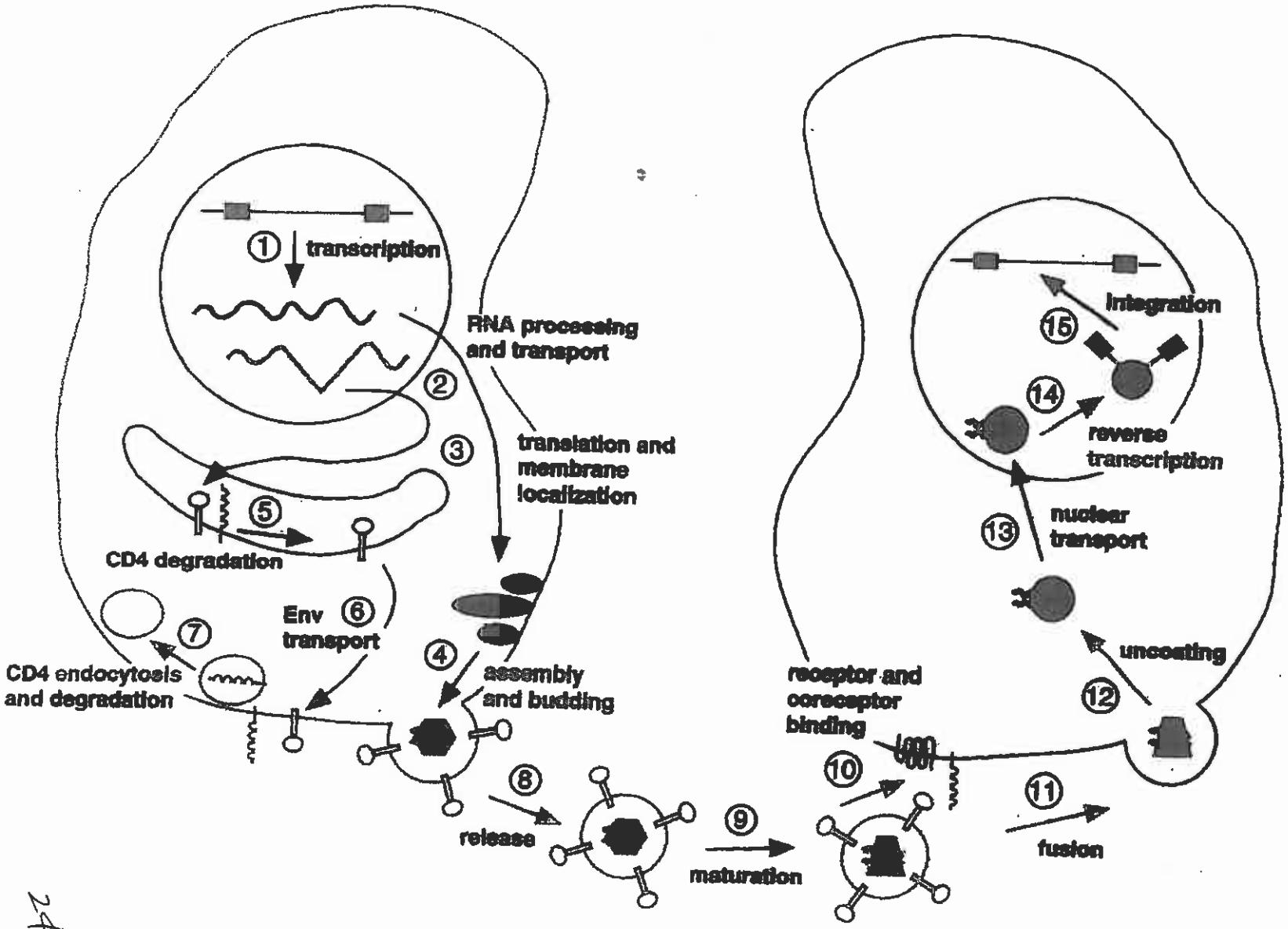
Reverse transcriptase inhibitors

Integrase inhibitors

Protease inhibitors

Some of these medications are combined and others taken sequentially. As a result, the survival rate of HIV infection has been greatly increased. However, vaccination development has been slow but considered as essential (Wu TT, Johnson G (2004) HIV vaccine candidates. Drugs of Today, 40, 949-955)

HIV-1 LIFE CYCLE



mutations (DDPAC (2 brains), ManF23 and Aus1) and 3 were from families with the P301L point mutant (FTD003 (2 brains) and HFTD1). Reverse transcription was performed using the Superscript preamplification kit (Life Technologies) on 1–4 μg of brain RNA with an oligo(dT) primer. PCR was performed between exon 9 (forward, 5′-ATCGCAGCGGCTACAGCAG-3′) and exon 11 (reverse, 5′-TGGTTTATGATGGATGTTGCC-3′) and between exon 9 and exon 13 (reverse 5′-TCTTGGCTTTGGCGTTCTC-3′). In each case, the 5′ end of the forward amplification primer was labelled with TET (Perkin Elmer) to allow detection by an ABI377 automated sequencer. Preliminary PCRs (not shown) were performed using a range of amplification cycles (18–37) to determine the optimum number of cycles for this analysis. Based on these results, we used 32 cycles in subsequent experiments. After amplification, PCR products were analysed on an ABI377 automated sequencer (Perkin Elmer) where they resolved into two major fragments (327 and 418 bp, exons 9–11; 487 and 578 bp, exon 9–13) corresponding to *tau* transcripts with and without exon 10. The identity of each band was confirmed by sequence analysis. The molar ratio of exon 10^+ to exon 10^- RNA was determined using Genescan software. Three independent PCRs (for both exons 9–11 and 9–13) were used to determine the mean and s.d. of the ratio for each brain. Results (not shown) were essentially identical when exon $10^{+/-}$ ratios were estimated by densitometric analysis of PCR products on agarose gels using a Kodak DC120 camera kit and ID Image Gel densitometry software.

Exon-trapping analysis of exon-10 splicing. Mutant and wild-type versions of *tau* exon 10 were amplified from the DNA of patients with each of the three different splice mutations (+13, +14 and +16) and from normal individuals. PCR products contained exon 10 and ~40 bp of flanking intron sequence at either end. PCR products were cloned into the splicing vector pSPL3b using *Xho*I and *Pst*I sites incorporated into the amplification products. Mutant and wild-type constructs were identified by sequence analysis. For exon trapping, the exon-trapping system of Life Technologies was used. Briefly, COS-7 cells were transfected in duplicate with 1 μg each construct using LipofectACE reagent (Life Technologies). Cells were collected 24 h post-transfection and RNA prepared using the Trizol reagent (Life Technologies). First-strand synthesis and nested PCR were done using reagents supplied with the system and conditions described in manufacturer's instructions, except that *Bst*X1 digestion of primary PCR products was excluded. To verify that the RT-PCR was quantitative, different amounts of primary PCR template (1–5 μl) were used and the total number of amplification cycles was varied (30–35 cycles). PCR products were analysed on 3% Metaphore (FMC) gels. RT-PCR products (Fig. 3) had their identities confirmed by sequencing.

Received 9 April; accepted 1 June 1998.

1. Foster, N. L. *et al.* Frontotemporal Dementia and Parkinsonism Linked to Chromosome 17: A Consensus Statement. *Annu. Neurol* **41**, 706–715 (1997).
2. Wilhelmsen, K. C., Lynch, T., Pavlou, E. & Nygaard, T. G. Localization of disinhibition-dementia-parkinsonism-amyotrophy complex to 17q21-22. *Am. J. Hum. Genet.* **55**, 1159–1165 (1994).
3. Baker, M. *et al.* Localization of frontotemporal dementia with parkinsonism in an Australian kindred to chromosome 17q21-22. *Annu. Neurol* **42**, 794–798 (1997).
4. Froelich, S. *et al.* Mapping of a disease locus for familial rapidly progressive frontotemporal dementia to chromosome 17q12-21. *Am. J. Med. Genet.* **74**, 380–385 (1997).
5. Morrell, J. *et al.* Familial multiple system tauopathy with presenile dementia localized to chromosome 17. *Am. J. Hum. Genet.* **61**, 1131–1138 (1997).
6. Wijker, M. *et al.* Localization of the gene for rapidly progressive autosomal dominant parkinsonism and dementia with pallido-ponto-nigral degeneration to chromosome 17q21. *Hum. Mol. Genet.* **5**, 151–154 (1996).
7. Heutink, P. *et al.* Hereditary fronto-temporal dementia is linked to chromosome 17q21-22. A genetic and clinico-pathological study of three Dutch families. *Annu. Neurol.* **41**, 150–159 (1997).
8. Yamaoka, L. H. *et al.* Linkage of frontotemporal dementia to chromosome 17- clinical and neuropathological characterization of phenotype. *Am. J. Hum. Genet.* **59**, 1306–1312 (1996).
9. Dark, F. A family with autosomal dominant, non-Alzheimer's presenile dementia. *Austr. N. Z. J. Psychiat.* **31**, 139–144 (1997).
10. Constantinidis, J., Richard, J. & Tissot, R. Pick's disease: Histological and clinical classification. *Eur. Neurol.* **11**, 208–217 (1974).
11. Andreadis, A., Brown, W. M. & Kosik, K. S. Structure and novel exons of the human *tau* gene *Biochemistry* **31**, 10626–10633 (1992).
12. Spillantini, M. G., Bird, T. D. & Ghetti, B. Frontotemporal dementia and parkinsonism linked to chromosome 17: A new group of tauopathies. *Brain Path.* **8**, 387–402 (1998).
13. Goedert, M., Spillantini, M. G., Potter, M. C., Ulrich, J. & Crowther, R. A. Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein *tau* containing four tandem repeats: differential expression of tau protein mRNAs in human brain. *EMBO J.* **8**, 393–399 (1989).
14. Spillantini, M. G. *et al.* Familial multiple system tauopathy with presenile dementia: a disease with abundant neuronal and glial tau filaments. *Proc. Natl Acad. Sci. USA* **94**, 4113–4118 (1997).
15. Goedert, M. *et al.* Assembly of microtubule-associated protein *tau* into Alzheimer-like filaments induced by sulphated glycosaminoglycans. *Nature* **383**, 550–553 (1996).
16. Goedert, M., Spillantini, M. G., Jakes, R., Rutherford, D. & Crowther, R. A. Multiple isoforms of human microtubule-associated protein *tau*: sequence and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron* **3**, 519–526 (1989).
17. The Lund and Manchester groups. Clinical and neuropathological criteria for frontotemporal dementia. *J. Neurol Neurosurg Psychiat* **57**, 416–418 (1994).
18. Butner, K. A. Kirschner, M. W. J. *Tau* protein binds to microtubules through a flexible array of distributed weak sites. *J. Cell Biol.* **115**, 717–730 (1991).
19. Reed, L. A. *et al.* Autosomal dominant dementia with widespread neurofibrillary tangles. *Annu. Neurol* **42**, 564–572 (1997).
20. McGeer, P. L., Schwab, C., McGeer, E. G., Haddock, R. L. & Steele, J. C. Familial nature and continuing morbidity of the amyotropic lateral sclerosis-parkinsonism dementia complex of Guam. *Neurology* **49**, 400–409 (1997).
21. Goedert, M. Tau protein and the neurofibrillary pathology of Alzheimer's disease. *Trends Neurosci.* **16**, 460–465 (1993).
22. Church, D. M. *et al.* Isolation of genes from complex sources of mammalian genomic DNA using exon amplification *Nature Genet* **6**, 98–105 (1994).
23. Eperon, L. P., Graham, I. R., Griffiths, A. D. & Eperon, I. C. Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell* **54**, 393–401 (1988).
24. Kuo, H.-C., Nasim, F.-U. H. & Grabowski, P. J. Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. *Science* **1251**, 1045–1050 (1991).
25. Dickson, D. Neurodegenerative diseases with cytoskeletal pathology: a biochemical classification. *Annu. Neurol.* **42**, 541–543 (1997).
26. Conrad, C. *et al.* Genetic evidence of the involvement of τ in progressive supranuclear palsy. *Annu. Neurol.* **41**, 277–281 (1997).
27. Higgins, J. J., Litvan, I., Pho, L. T., Li, W. & Nee, L. E. Progressive supranuclear palsy is in linkage disequilibrium with the τ and not the α-synuclein gene. *Neurology* **50**, 270–273 (1998).
28. Freier, S. M. *et al.* Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl Acad. Sci. USA* **83**, 9373–9377 (1986).

Acknowledgements. This work was supported by an NINDS (RO1) grant to M.H, an NIA (MCSDA) grant to T.L. and NIA (P50) grants to A.G. and I.C.M. and to M.H., J.H. and R.C.P. Additional support was provided by the Mayo Foundation (M.H., J.H.), the International Foundation for Alzheimer Research, the Dutch Brain Research Foundation (P.H.) and Judith Mason (P. Dodd). A.G. is the recipient of an NIH career development award (NIA). C.L.L. was a Washington University Alzheimer's Disease Research Center postdoctoral fellow. P.R. is the recipient of a TMG of EU grant. J.M.K. is the recipient of an NSADA award. T.L. is the recipient of Irving Scholar, NARSAD and Parkins's Disease Foundation awards. We thank the Mayo Clinic Molecular Biology Core Facility for sequencing and acknowledge the support of the Dutch Brain Bank (W.K., R.R.), the Michigan ADRC Brain Bank and the Columbia University Brain Bank. The participation of the families has been crucial.

Correspondence and requests for materials should be addressed to M.H. (e-mail: hutton michael@mayo.edu) or P.H. (e-mail: heutink@kgen.fgg.eur.nl).

The antigenic structure of the HIV gp120 envelope glycoprotein

Richard Wyatt*, Peter D. Kwong†, Elizabeth Desjardins*, Raymond W. Sweet‡, James Robinson§, Wayne A. Hendrickson† & Joseph G. Sodroski*‖

** Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Department of Pathology, Harvard Medical School, and ‖ Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts 02115, USA*

† Department of Biochemistry and Molecular Biophysics, Howard Hughes Medical Institute, Columbia University, New York, New York 10032, USA

‡ SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, Pennsylvania 19406-0939, USA

§ Department of Pediatrics, Tulane University Medical Center, 1430 Tulane Avenue, New Orleans, Louisiana 70112, USA

The human immunodeficiency virus HIV-1 establishes persistent infections in humans which lead to acquired immunodeficiency syndrome (AIDS). The HIV-1 envelope glycoproteins, gp120 and gp41, are assembled into a trimeric complex that mediates virus entry into target cells[1]. HIV-1 entry depends on the sequential interaction of the gp120 exterior envelope glycoprotein with the receptors on the cell, CD4 and members of the chemokine receptor family[2–4]. The gp120 glycoprotein, which can be shed from the envelope complex, elicits both virus-neutralizing and non-neutralizing antibodies during natural infection. Antibodies that lack neutralizing activity are often directed against the gp120 regions that are occluded on the assembled trimer and which are exposed only upon shedding[5,6]. Neutralizing antibodies, by contrast, must access the functional envelope glycoprotein complex[7] and typically recognize conserved or variable epitopes near the receptor-binding regions[8–11]. Here we describe the spatial organization of conserved neutralization epitopes on gp120, using epitope

Nature © Macmillan Publishers Ltd 1998

maps in conjunction with the X-ray crystal structure of a ternary complex that includes a gp120 core, CD4 and a neutralizing antibody[12]. A large fraction of the predicted accessible surface of gp120 in the trimer is composed of variable, heavily glycosylated core and loop structures that surround the receptor-binding regions. Understanding the structural basis for the ability of HIV-1 to evade the humoral immune response should assist in the design of a vaccine.

The amino-acid sequence of human and simian immunodeficiency virus gp120 glycoproteins consists of five variable regions (V1–V5) interposed among more conserved regions[13]. Variable regions V1–V4 form exposed loops anchored at their bases by disulphide bonds[14].

Neutralizing antibodies recognize both variable and conserved gp120 structures. The V2 and V3 loops contain epitopes for strain-restricted neutralizing antibodies[15–17]. More broadly neutralizing antibodies recognize discontinuous, conserved epitopes in three regions of the gp120 glycoprotein (Table 1). In HIV-1-infected humans, the most abundant of these are directed against the CD4-binding site (CD4BS) and block gp120–CD4 interaction[8,9]. Less common are antibodies against epitopes induced or exposed upon CD4 binding (CD4i)[18]. Both CD4i and V3 antibodies disrupt the binding of gp120–CD4 complexes to chemokine receptors[10,11]. A third gp120 neutralization epitope is defined by a unique monoclonal antibody, 2G12 (ref. 19), which does not efficiently block receptor binding[11].

In the accompanying Article[12], we report the X-ray crystal structure of an HIV-1 gp120 core in a ternary complex with two-domain soluble CD4 and the Fab fragment of the CD4i antibody 17b. The gp120 core lacks the V1/V2 and V3 variable loops, as well as amino- and carboxy-terminal sequences, which interact with the gp41 glycoprotein[6], and is enzymatically deglycosylated[12,20]. Despite these modifications, the gp120 core binds CD4 and antibodies against CD4BS and CD4i epitopes[20,21] and thus retains structural integrity. The gp120 core is composed of an inner domain, an outer domain and a third element, the 'bridging sheet'[12] (Fig. 1a). All three structural elements contribute, either directly or indirectly, to CD4 and chemokine-receptor binding[12]. We now analyse the organization of the surface of the gp120 core in light of the known antibody responses directed against this exposed viral glycoprotein.

Although generally well conserved compared with the five variable regions, some variability in the surface of the gp120 core is evident when the sequences of all primate immunodeficiency viruses are analysed. This variability is disproportionately associated with the surface of the outer domain proximal to the V4 and V5 regions and removed from the receptor-binding regions (Fig. 1a–c). The $\mathcal{L}$A, $\mathcal{L}$C, $\mathcal{L}$D and $\mathcal{L}$E surface loops[12] contribute to the variability of this surface. The potential *N*-linked glycosylation sites present in the gp120 core are concentrated in this variable half of the protein (Fig. 1b, c). The only conserved residues apparent on this relatively variable surface are asparagine 356 and threonine/serine 358, which constitute a complex carbohydrate addition site within the $\mathcal{L}$E loop (Fig. 1b, c). As most carbohydrate moieties may appear as 'self' to the immune system, the extensive glycosylation of the outer domain surface may render it less visible to immune surveillance. This helps to explain why antibodies directed against this gp120 surface have been identified so infrequently.

The receptor-binding regions retained in the gp120 core are well conserved among primate immunodeficiency viruses[12]. Also highly conserved is the surface of the inner domain spanned by the α1 helix and located opposite the variable surface described above (Fig. 1d). This surface is likely to interact with gp41 and/or with N-terminal gp120 segments absent from the gp120 core. This inner domain surface and the receptor-binding regions are devoid of glycosylation.

In conjunction with prior mutagenic and antibody competition analyses[5,6,18–20,22] the gp120 core structure reveals the spatial positioning of the conserved gp120 neutralization epitopes. Although the principal variable loops are either absent (V1/V2 and V3) or poorly resolved (V4) in the gp120 core structure, their approximate positions can be deduced (Fig. 2a). The conserved gp120 neutralization epitopes are discussed in relation to these variable loops and the variable, glycosylated core surface.

CD4i epitopes. The gp120 epitope recognized by the CD4i antibody 17b can be directly visualized in the crystallized ternary complex[12] (Fig. 2b, c). Strands from the gp120 fourth conserved (C4) region and the V1/V2 stem contribute to an antiparallel β-sheet (the 'bridging sheet'; Fig. 1a) that contacts the antibody. Most gp120 residues previously implicated in the formation of the CD4i epitopes[18] (Table 1) are located either within this β-sheet or in nearby structures. With the exception of Thr 202 and Met 434, the gp120 residues in contact with the 17b Fab are highly conserved among HIV-1 isolates (Figs 1c, 2a). The prominent ('male') CDR3 loop of the 17b heavy chain dominates the contacts with gp120, with

Table 1 Conserved epitopes for neutralizing antibodies identified on the gp120 core

Competition group*	Examples of monoclonal antibodies	gp120 amino acids†	Probable mechanism of virus neutralization	Properties	Selected refs
CD4-binding site (CD4BS)	F105 15e 21h 1125h 448D 39.3 IgG1b12 830D	Asn 88 (13), Asp 113 (50), Lys 117 (25), Ser 256 (75) Thr 257 (75), Asn 262 (63), Ala 265 (13), Asp 368 (100), Glu 370 (100), Tyr 384 (13), Lys 421 (50), Trp 427 (25), Asp 457 (13), Pro 470 (25), Asp 474 (13), Met 475 (13), Asp 477 (63), Asp/Leu/Tyr 482/483/484 (25)	Interference with gp120–CD4 binding	CD4BS antibodies compete with CD4 and with antibodies against CD4i epitopes	8, 9, 22
CD4-induced epitopes (CD4i)	17b 48d	Asn 88, Lys 117, Lys 121, Lys 207, Ser 256, Thr 257, Asn 262, ΔV3, Glu 370, Glu 381, Phe 382, Arg 419, Ile 420, Lys 421, Gln 422, Ile 423, Trp 427, Tyr 435, Pro 438, Met 475	Interference with chemokine-receptor binding	CD4 binding increases exposure of the epitopes as a result of movement of the V2 variable loop	18; C. Rizzuto and J.G.S., submitted
2G12	2G12	Asn 295, Thr 297, Ser 334, Asn 386, Asn 392, Asn 397	Unknown	Antibody binding is dependent upon proper *N*-linked glycosylation	19

* The gp120 competition groups are defined as in ref. 5.
† The gp120 amino acids are numbered according to the sequence of the HXBc2 (IIIB) gp120 glycoprotein, where residue 1 is the methionine at the amino terminus of the signal peptide. Changes in the amino acids listed resulted in significant reduction in antibody binding to the gp120 glycoprotein (refs 18, 19, 22). Numbers in parentheses indicate the percentage of CD4BS antibodies examined whose binding is decreased by changes in the indicated residue.

Nature © Macmillan Publishers Ltd 1998

additional contacts through the heavy chain CDR2 (ref. 12). Unusually, there are minimal 17b light-chain contacts, leaving a large gap between the gp120 core and most of the 17b light-chain surface. In the complete gp120 glycoprotein, this gap is probably occupied by the V3 loop. This is consistent with the position and orientation of the V3 base on the gp120 core structure[12], the effect of V3 deletions on the binding of CD4i antibodies in the absence of soluble CD4 (ref. 21), the competition of some V3-directed antibodies with CD4i antibodies[5], and the ability of both antibody groups to block chemokine-receptor binding[10,11]. The chemokine-receptor-binding region of gp120 probably consists of elements near or within the 'bridging sheet' and the V3 loop (Fig. 1a), a model that is supported by recent mutagenic analysis (C. Rizzuto *et al.*, submitted).

The V2 loop probably resides on the side of the 17b epitope opposite the V3 loop (Fig. 2a). The V1/V2 loops, which vary from 57 to 86 residues in length[13], are dispensable for HIV-1 replication[21,23], but decrease the sensitivity of viruses to neutraliza-

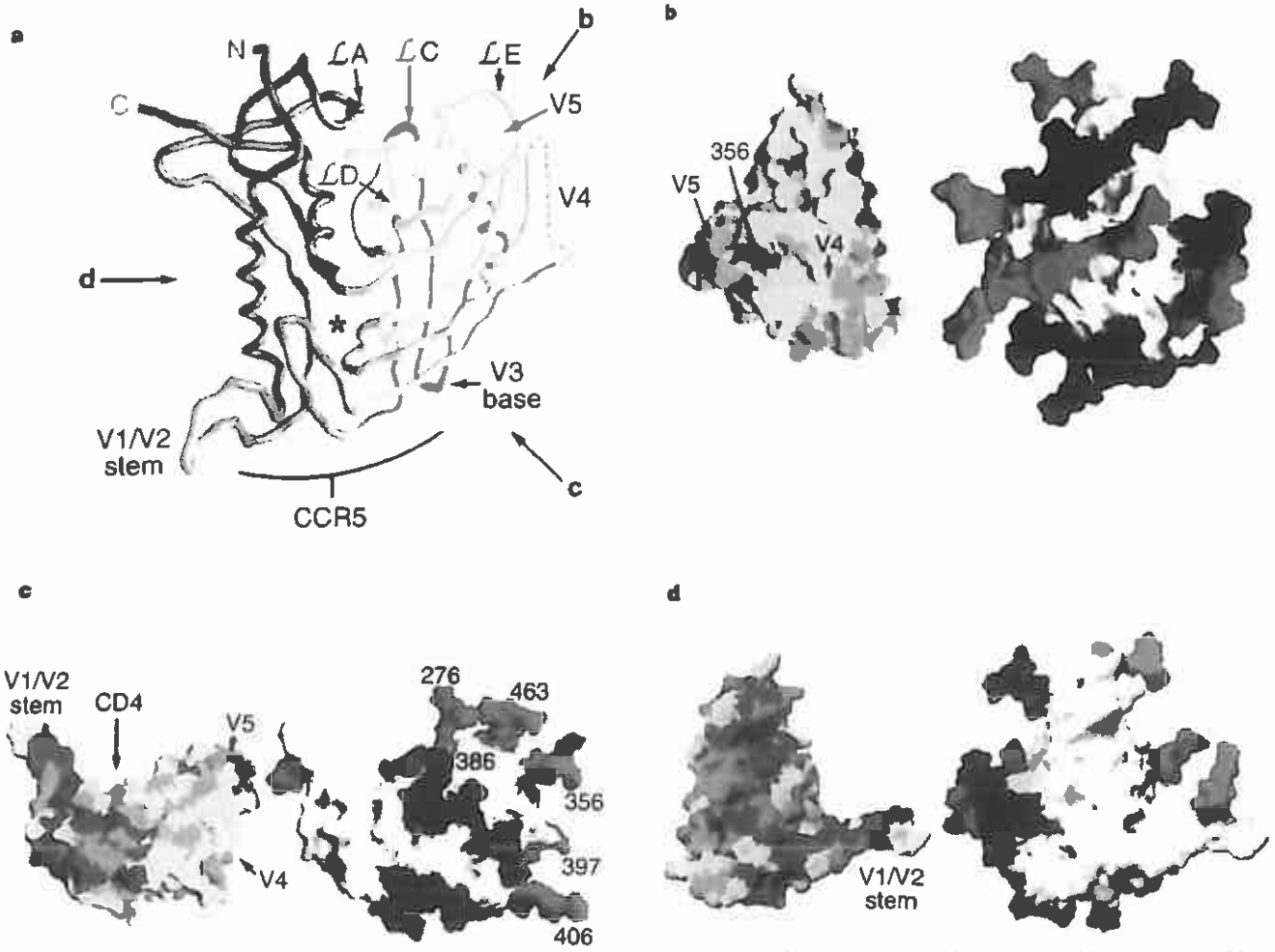


Figure 1 Structure and orientation of the HIV-1 gp120 core. **a**, A Cα tracing of the gp120 core, which was crystallized in a ternary complex with two-domain soluble CD4 and the Fab fragment of the 17b antibody[12], is shown. The gp120 core is seen from the perspective of CD4, and is oriented with the viral membrane at the top of the figure and the target cell membrane at the bottom. The inner gp120 domain is shown in red and the outer domain in yellow; the 'bridging sheet' is orange. The N and C termini of the truncated gp120 core are labelled, as are the positions of structures related to the gp120 variable regions V1–V5. The αA, αC, αD and αE surface loops[12] are shown. The position of the Phe 43 cavity involved in CD4 binding is indicated by an asterisk. The gp120 surface implicated in binding to the CCR5 chemokine receptor (C. Rizzuto and J.G.S., submitted) is shown. The perspectives shown in **b**–**d** are indicated. **b**, View of the molecular surface of the gp120 outer domain, from the perspective indicated in **a**. The molecular surface on the left is coloured according to the variability observed in gp120 residues among primate immunodeficiency viruses: red, residues conserved among all primate immunodeficiency viruses; orange, residues conserved in all HIV-1 isolates; yellow, residues exhibiting some variation among HIV-1 isolates; and green, residues showing significant variability among HIV-1 isolates (see Methods). The variability of the gp120 surface is underestimated here because the V4 variable loop, which is not resolved in the structure, contributes to this surface (approximate location is indicated). The position of the V5 region is shown. Also note the highly conserved glycosylation site (Asn 356 and Thr/Ser 358) within the αE loop, between the V5 and V4 regions. On the right, the V4 loop and the carbohydrates are modelled (see Methods). The complex carbohydrate addition sites used in mammalian cells[14] are coloured light blue, and the high-mannose sites are dark blue. The gp120 protein surface is in white. **c**, View of the gp120 molecular surface that faces the target cell. Variability is indicated on the left, using the same colour scheme as in **b**. Note the clear demarcation between the conserved surface, which has been implicated in the formation of CD4i epitopes[18] and in chemokine-receptor binding (C. Rizzuto and J.G.S., submitted), and the variable surface of the outer domain. The recessed binding site for CD4 is indicated, flanked by the V1/V2 stem, which is labelled. The V4 loop and the carbohydrates are modelled on the right (colouring as in **b**). Particular carbohydrates referred to in the text are labelled. **d**, View of the molecular surface of the gp120 core inner domain. Variability is indicated on the left by the colour scheme used in **b**. The CD4-binding site is on the right; the protruding V1/V2 stem is indicated. The conserved molecular surface, which is associated with the inner domain of the gp120 core, is devoid of known *N*-linked glycosylation sites. These are modelled on the right, which is coloured as in **b**.

Nature © Macmillan Publishers Ltd 1998

tion by antibodies against V3 and CD4i epitopes[23]. The latter effect is mediated primarily by the V2 loop[21], suggesting that part of the V2 loop folds back along the V1/V2 stem to mask the 'bridging sheet' and adjacent V3 loop. The proximity of the V2 and V3 loops is supported by the observation that, in monkeys infected with simian–human immunodeficiency viruses (SHIVs), neutralizing antibodies are raised against discontinuous epitopes with V2 and V3 components (B. Etemad-Moghadam *et al.*, submitted). The CD4i epitopes are probably masked by the flanking V2 and V3 loops, requiring the evolution of antibodies with protruding ('male') complementarity-determining regions (CDRs) to access these conserved epitopes. It has been suggested that CD4 binding repositions the V1/V2 loops, thus exposing the CD4i epitopes[21]. The presence of contacts between the V1/V2 stem and CD4 in the crystal structure[12] is consistent with this model.

CD4BS epitopes. CD4 makes several contacts within a recessed pocket on the gp120 surface. The gp120–CD4 interface includes two cavities, one water-filled and bounded equally by both proteins,

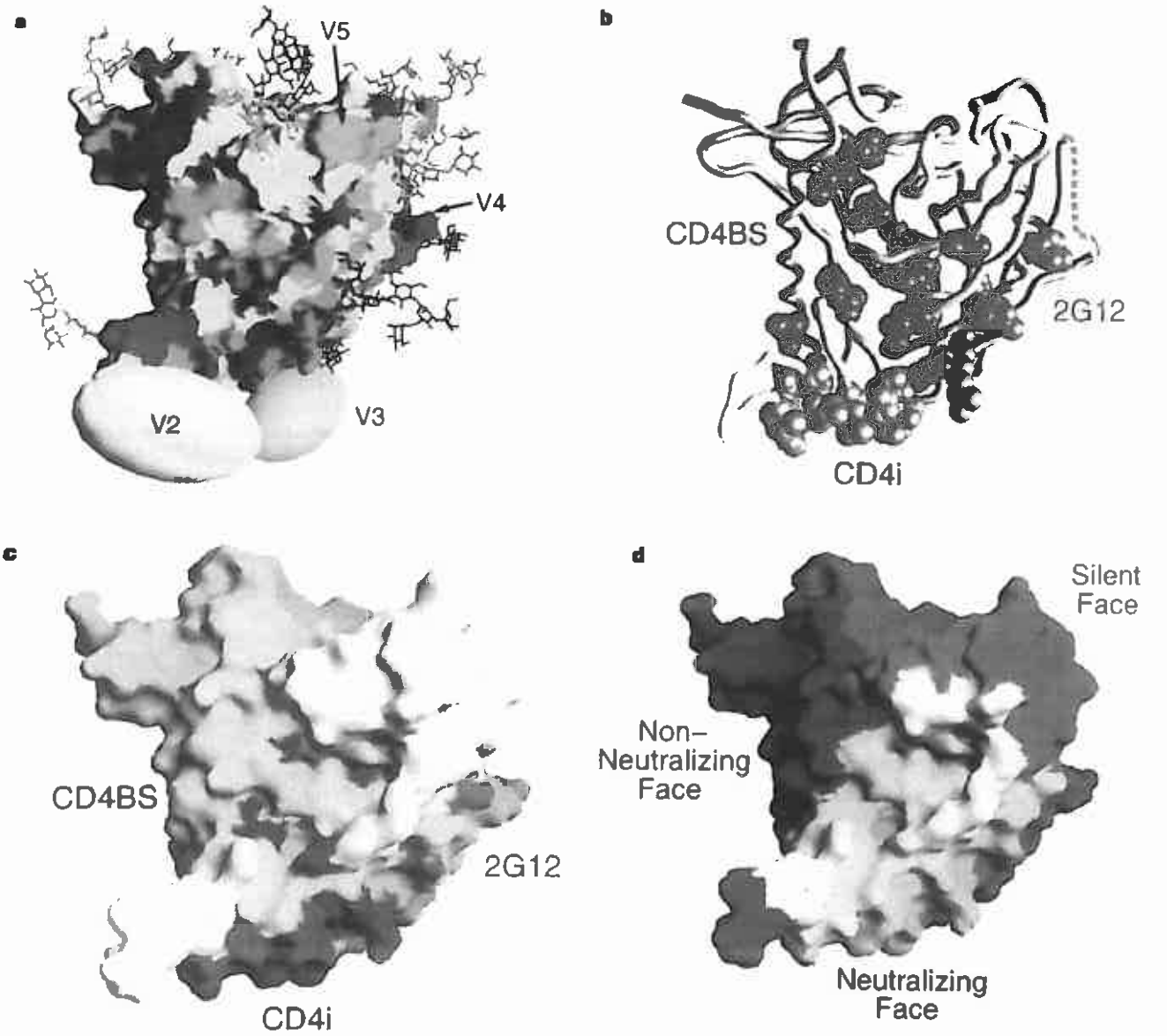


Figure 2 The spatial relationship of epitopes on the HIV-1 gp120 glycoprotein. **a**, The molecular surface of the gp120 core is shown, using the same perspective as in Fig. 1a. The modelled N-terminal gp120 core residues, V4 loop and carbohydrate structures are included. The variability of the molecular surface is indicated (colour scheme as in Fig. 1b). The modelled carbohydrates are shown in light blue (complex sugars) or dark blue (high-mannose sugars). The approximate locations of the V2 and V3 variable loops are indicated. Note the well-conserved surfaces near the 'Phe 43' cavity and the chemokine-receptor-binding site (Fig. 1a). **b**, Cα tracing of the gp120 core, oriented as in Fig. 1a. The gp120 residues within 4 Å of the 17b CD4i antibody are shown in green; those implicated in the binding of CD4BS antibodies[22] are in red. Changes in these residues significantly affect the binding of at least 25% of the CD4BS antibodies listed in Table 1. The residues implicated in 2G12 binding[19] are shown in blue. The V4 variable loop, which contributes to the 2G12 epitope[19], is indicated by dotted lines. **c**, The molecular surface of the gp120 core, oriented and coloured as in **b**. **d**, Approximate locations of the faces of the gp120 core, defined by the interaction of gp120 and antibodies. The molecular surface accessible to neutralizing ligands (CD4 and CD4BS, CD4i and 2G12 antibodies) is shown in white. The neutralizing face of the complete gp120 glycoprotein includes the V2 and V3 loops, which are found adjacent to the surface shown (**a**). The approximate location of the gp120 face that is poorly accessible on the assembled envelope glycoprotein trimer and therefore elicits only non-neutralizing antibodies[5,6] is shown in magenta. The approximate location of an immunologically 'silent' face of gp120, which roughly corresponds to the highly glycosylated outer domain surface, is in blue.

Nature © Macmillan Publishers Ltd 1998

the other extending into the gp120 interior and contacting CD4 only at Phe 43 (Fig. 1a)[12]. Table 1 and Fig. 2b, c show the gp120 residues implicated in the formation of CD4BS epitopes recognized by eight representative antibodies. CD4BS epitopes are uniformly disrupted by changes in Asp 368 and Glu 370 (ref. 22), which surround the opening of the 'Phe 43 cavity'. These residues are located on a ridge at the intersection of the two receptor-binding gp120 surfaces, consistent with competition studies suggesting that CD4BS epitopes overlap both the CD4i epitopes and the binding site for CD4 (refs 5, 18). The location of the gp120 residues implicated in formation of the CD4BS epitopes suggests that important elements of the CD4-binding surface of gp120 are accessible to antibodies.

Some CD4BS antibodies, like IgG1b12, are particularly potent at neutralizing HIV-1 (ref. 24). IgG1b12 binding is disrupted by gp120 changes that affect the binding of other CD4BS antibodies but, atypically, is sensitive to changes in the V1/V2 stem–loop structure[25]. The observation that some well-conserved residues in the gp120 V1/V2 stem contact CD4 (ref. 12) raises the possibility that this protruding structure also contributes to the IgG1b12 epitope. This might increase the ability of the antibody to access the assembled envelope glycoprotein trimer, thus increasing neutralizing capability.

Although the CD4BS epitopes and the CD4-binding site overlap, several observations demonstrate that the binding of CD4BS antibodies differs from that of CD4. Changes in Trp 427, a gp120 residue that contacts both the Phe 43 cavity and CD4, uniformly disrupt CD4 binding but affect the binding of only some CD4BS antibodies (Table 1). Conversely, some changes in other cavity-lining gp120 residues, Ser 256 and Thr 257, affect the binding of CD4BS antibodies more than the binding of CD4 (ref. 22). As the recessed position of Ser 256 and Thr 257 in the current crystal structure (Fig. 2b, c) makes direct contacts with antibody unlikely, either the effects of changes in these residues are indirect or the CD4BS antibodies recognize a gp120 conformation that differs from the CD4-bound state. With respect to the latter possibility, it is interesting that several of the residues implicated in the integrity of the CD4BS epitopes are located in the interface between the inner and outer gp120 domains. CD4BS antibodies might recognize a gp120 conformation in which the spatial relationship between the domains is altered compared with the CD4-bound state, thus allowing better surface exposure of these residues. Differences between the CD4BS epitopes and the CD4-binding site create opportunities for neutralization escape[22]. The gp120 residues surrounding the Phe 43 cavity are highly conserved among primate immunodeficiency viruses (Fig. 2a), but the observed modest variation in adjacent surface-accessible residues (for example, Pro 369, Thr 373 and Lys 432) could account for decreased recognition of the gp120 glycoprotein from some geographic clades of HIV-1 by CD4BS antibodies[25]. Additional potential for variation near or within the CD4BS epitopes is created by the unusual water-filled cavity in the gp120–CD4 binding interface, because CD4 binding can apparently tolerate change in the gp120 residues contacting this cavity[12].

The recessed nature of the CD4-binding pocket on gp120 (Fig. 1c) may delay the generation of high-affinity antibodies against the CD4BS epitopes and may afford opportunities to minimize the antiviral efficacy of such antibodies once they are elicited. The degree of recession is probably much greater on the full-length glycosylated gp120 than is evident on the crystallized gp120 core. The recessed pocket is flanked on one side by the V1/V2 stem–loop structure. The characterization of HIV-1 escape mutants from the IgG1b12 CD4BS antibody and the mapping of several V2 conformational epitopes support a model in which the V2 loop folds back along the V1/V2 stem, with V2 residues 183–188 proximal to Asp 368 and Glu 370. This model is consistent with observations that V1/V2 changes, in combination with V3 changes, can alter the exposure of the adjacent CD4BS epitopes, particularly on the assembled trimer[26]. The high temperature factors associated with the V1/V2 stem[12] imply flexibility in this protruding element (Fig. 1c, d), expanding the potential range of space occupied by the V1/V2 stem–loop structure. This could increase masking of the adjacent CD4BS and CD4i gp120 epitopes and divert antibody responses towards the variable loops.

Glycosylation may modify the interaction of antibodies with CD4BS epitopes. The $\mathcal{L}$D loop, on the rim of the CD4-binding pocket opposite the V1/V2 stem, contains a well-conserved glycosylation site, Asn 276 (Fig. 1c). Changes in this site and at the adjacent Ala 281 have been associated with escape from the neutralizing activity of patient sera[27] and have been seen in SHIVs extensively passaged in monkeys[28]. Another conserved glycosylation site at Asn 386 lies adjacent to both CD4BS and CD4i epitopes (Fig. 1c) and could diminish antibody responses against those sites. Additionally, in various HIV-1 strains, carbohydrates are added to the V2 loop segment (residues 186–188) thought to be proximal to the CD4BS epitopes.

The 2G12 epitope. The integrity of the 2G12 epitope is disrupted by changes in gp120 glycosylation, by either glycosidase treatment or mutagenic alteration of specific *N*-linked carbohydrate-addition sites[19]. These sites are located on the relatively variable surface of the gp120 outer domain, opposite to and approximately 25 Å away from the CD4-binding site (Fig. 2b, c). The gp120 glycoprotein synthesized in mammalian cells exhibits a dense concentration of high-mannose sugars in this region (Fig. 2a). Even in the enzymatically deglycosylated gp120 core, carbohydrate residues constitute much of this surface. 2G12 probably binds at least in part to these

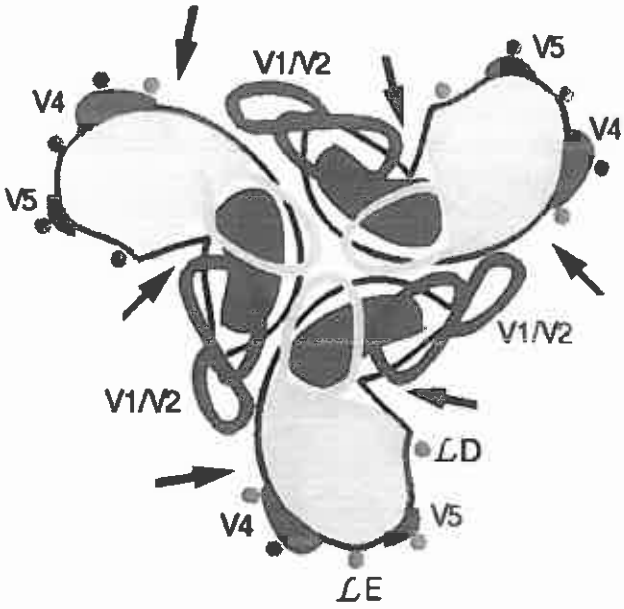


Figure 3 A likely arrangement of the HIV-1 gp120 glycoproteins in a trimeric complex. The gp120 core was organized into a trimeric array, based on the criteria discussed in the text. The perspective is from the target-cell membrane, like that in Fig. 1c. The CD4-binding pockets are indicated by black arrows, and the conserved chemokine-receptor-binding regions are in red. Areas shaded in light green indicate the more variable, glycosylated surfaces of the gp120 cores. The approximate locations of the 2G12 epitopes are indicated by blue arrows; those of the V3 loops (yellow) and V4 regions (green) are indicated. The positions of the V5 regions (green) and some complex-carbohydrate addition sites (asparagines 276, 463, 356, 397 and 406) (blue dots) are shown. The approximate locations of the large V1/V2 loops, centred on the known positions of the V1/V2 stems, are indicated (green). On one of the gp120 subunits, the positions of the $\mathcal{L}$D and $\mathcal{L}$E loops are indicated. The distance of each of the gp120 monomers from the 3-fold symmetry axis is arbitrary.

Nature © Macmillan Publishers Ltd 1998

carbohydrates, explaining the surprising conservation of the 2G12 epitope despite the variability of the underlying protein surface, which includes the stem of the V3 loop and the V4 variable region. The inclusion of carbohydrate in the epitope might also explain the apparent rarity with which these antibodies are generated. The localization of the 2G12 epitope is consistent with previous studies indicating that 2G12 forms a unique competition group[5,19] and does not interfere with the binding of monomeric gp120 to either CD4 or chemokine receptors[11]. As the 2G12 epitope is predicted to be oriented towards the target cell upon CD4 binding (see below), the antibody may sterically impair interactions of the oligomeric envelope glycoprotein complex with host cell moieties.

Possible orientations of the exterior glycoproteins in the trimer are significantly constrained by the requirement that observed and deduced binding sites for receptors and neutralizing antibodies, sites of *N*-linked glycosylation, and variable structures be exposed on the surface of the assembled complex. The two-domain CD4 in the ternary complex structure was aligned to the structure of four-domain CD4 (ref. 29) to orient the trimer model with respect to the target cell membrane. The predictions of such a model (Fig. 3) are: (1) the chemokine-receptor-binding sites are clustered at the vertex of the trimer predicted to be closest to the target cell; (2) both variable and conserved neutralization epitopes are concentrated on the half of gp120 facing the target cell; (3) possibilities for inter-subunit interactions among the variable structures that could help mask conserved neutralization epitopes are created; (4) the subset of gp120 glycosylation sites to which complex carbohydrates are added in mammalian cells[14] is well exposed on the outer periphery of the trimer; (5) the highly conserved surface near the α1 helix is available for gp41 and/or gp120 protein interactions within the trimers; and (6) the surface of the assembled envelope glycoprotein complex is roughly hemispherical, minimizing the surface area of the viral spike that is potentially exposed to antibodies.

In summary, the X-ray crystal structure of the gp120 core/two-domain CD4/17b Fab complex provides a framework for visualizing important interactions between HIV-1 and the humoral immune system. Previous antibody competition analyses indicated that the gp120 surface buried in the assembled trimer elicits non-neutralizing antibodies[5,6]. By contrast, the binding sites for neutralizing antibodies cluster on a different gp120 surface[5]. Our structural studies support the existence of non-neutralizing and neutralizing faces of gp120, and reveal another, immunologically 'silent' face of the glycoprotein (Fig. 2d). This outer domain surface, together with the principal variable loops, contributes to the large fraction of the gp120 surface that is protected against antibody responses by a dense array of carbohydrates and by the capacity for variation. The conserved receptor-binding regions of gp120 represent attractive targets for immune intervention. However, the elicitation of antibodies against these conformation-dependent structures is inefficient. As the gp120 epitopes near the receptor-binding regions span the inner and outer domains, interdomain conformational shifts may decrease their representation in the immunogen pool. The recessed nature of the CD4-binding site probably contributes to its poor immunogenicity. The sequential recognition of two receptors by primate immunodeficiency viruses allows the conserved elements of the chemokine-receptor-binding site to be created or exposed only after CD4 binding has occurred. At that point, it is likely that the proximity of the chemokine-receptor-binding site to the cell membrane sterically limits antibody binding. The evolution of primate immunodeficiency viruses that persist successfully, despite the host immune response, presents a challenge to vaccine development. An understanding of the structures of the relevant gp120 epitopes should assist efforts to overcome these hurdles. □

Methods

Graphics. Molecular graphics were produced using Midas-Plus (University of California, San Francisco) and GRASP[10].

Assignment of variability. Variability in gp120 residues was assessed using an alignment of sequences derived from ~400 HIV-1, HIV-2 and simian immunodeficiency viruses[13]. Residues were assigned variability indices and colour coded as follows: red, conserved in all primate immunodeficiency viruses; orange, conserved in all HIV-1, including groups M and O and chimpanzee isolates; yellow, some variation among HIV-1 isolates (divergence from the consensus sequence in 1–8 of the 12 HIV-1 groups examined); green, variable among HIV-1 isolates (divergence from the consensus sequence in ≥9 of the 12 HIV-1 groups examined).

Molecular modelling. Residues 88, 89 and 397–409, which are disordered in the ternary complex crystals[12], were built manually using the program TOM. For the V4 loop (residues 397–409), a dominant constraint was the distance between the ordered residues 396 and 410 (Cα–Cα distance of 26.88 Å). For the carbohydrate, examination of the *N*-linked carbohydrate in several crystal structures (for example, 1fc2, 1gly, 1lte) showed that the core common to both high-mannose and complex *N*-linked sugars, $(NAG)_2(MAN)_3$, did not differ greatly in conformation after alignment of the first NAG (*N*-acetylglucosamine) (MAN, mannose). This core, which represents roughly half the total glycosylation for a typical *N*-linked site, was built onto each of the 18 consensus *N*-linked glycosylation sites found on the HXBc2 gp120 core. The stereochemistry of this initial model was refined using simulated annealing in XPLOR. Briefly, the model was heated to temperatures of between 2,500 and 3,500 K, and slow cooled in steps of 25 to 300 K. At each step, molecular dynamics were performed with the core gp120 fixed, allowing only the modelled residues and carbohydrate (including any attached Asn) to move. In three separate runs, performing molecular dynamics for 5 fs per step, all steric clashes could be removed and the geometry idealized, with an average root-mean-square (r.m.s.) of carbohydrate movement of only ~3.5 Å. Four subsequent runs were made using dynamic times of between 50–75 fs per step. The carbohydrate positions obtained from these runs differed substantially from those in the starting model (average carbohydrate r.m.s. difference of ~8 Å). Two of the models from these longer annealings were more similar to each other than to the rest (r.m.s. differences in carbohydrate of ~4 Å versus ~8 Å for all other models): one had been heated to 3,500 K, with dynamics of 75 fs per step; the other (see figures) was heated to only 2,500 K, with dynamics of 50 fs per step. In general, the r.m.s. movement of the NAG sugars was roughly half the r.m.s. movement of the MAN sugars, reflecting greater conformational flexibility away from the protein surface.

Received 15 May; accepted 27 May 1998.

1. Allen, J. *et al.* Identification of the major envelope glycoprotein product of HTLV-III. *Science* **228**, 1091–1094 (1983).
2. Dalgleish, A. G. *et al.* The CD4 (T4) antigen is essential component of the receptor for the AIDS retrovirus. *Nature* **312**, 763–767 (1984).
3. Klatzmann, D. *et al.* T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* **312**, 767–768 (1984).
4. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. HIV-1 entry cofactor: Functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872–877 (1996).
5. Moore, J. P. & Sodroski, J. Antibody cross-competition analysis of the human immunodeficiency virus type 1 gp120 exterior envelope glycoprotein. *J. Virol.* **70**, 1863–1872 (1996).
6. Wyatt, R. *et al.* Analysis of the interaction of the human immunodeficiency virus type 1 (HIV-1) gp120 envelope glycoprotein with the gp41 transmembrane glycoprotein. *J. Virol.* **71**, 9722–9731 (1997).
7. Sattentau, Q. J. & Moore, J. P. Human immunodeficiency virus type 1 neutralization is determined by epitope exposure on the gp120 oligomer. *J. Exp. Med.* **182**, 185–196 (1995).
8. Posner, M. *et al.* An IgG human monoclonal antibody which reacts with HIV-1 gp120, inhibits virus binding to cells, and neutralizes infection. *J. Immunol.* **146**, 4325–4332 (1991).
9. Ho, D. *et al.* Conformational epitope on gp120 important in CD4 binding and human immunodeficiency virus type 1 neutralization identified by a human monoclonal antibody. *J. Virol.* **65**, 489–493 (1991).
10. Wu, L. *et al.* CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CR5. *Nature* **384**, 179–183 (1996).
11. Trkola, A. *et al.* CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature* **384**, 184–187 (1996).
12. Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648–659 (1998).
13. Myers, G. *et al.* *Human Retroviruses and AIDS. A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences* (Los Alamos National Laboratory, 1996).
14. Leonard, C. *et al.* Assignment of interchain disulfide bonds and characterization of potential glycosylation sites of the type 1 human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J. Biol. Chem.* **265**, 10373–10382 (1990).
15. Fung, M. S. C. *et al.* Identification and characterization of a neutralization site within the second variable region of human immunodeficiency virus type 1 gp120. *J. Virol.* **66**, 848–856 (1992).
16. Putney, S. *et al.* HTLV-III/LAV-neutralizing antibodies to an *E. coli*-produced fragment of the virus envelope. *Science* **234**, 1392–1395 (1986).
17. Rusche, J. R. *et al.* Antibodies that inhibit fusion of human immunodeficiency virus-infected cells bind a 24-amino-acid sequence of the viral envelope gp120. *Proc. Natl Acad. Sci. USA* **85**, 3198–3202 (1988).
18. Thali, M. *et al.* Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J. Virol.* **67**, 3978–3988 (1993).

Nature © Macmillan Publishers Ltd 1998

255

19. Trkola, A. *et al.* Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J. Virol.* **70**, 1100–1108 (1996).
20. Binley, J. *et al.* Analysis of the interaction of antibodies with a conserved, enzymatically deglycosylated core of the HIV-1 gp120 envelope glycoprotein. *AIDS Res. Hum. Retrovir.* **14**, 191–198 (1997).
21. Wyatt, R. *et al.* Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding. *J. Virol.* **69**, 5723–5733 (1995).
22. Thali, M. *et al.* Discontinuous, conserved neutralization epitopes overlapping the CD4 binding region of the HIV-1 gp120 envelope glycoprotein. *J. Virol.* **66**, 5635–5641 (1992).
23. Cao, J. *et al.* Replication and neutralization of human immunodeficiency virus type 1 lacking the V1/V2 variable loops of the gp120 envelope glycoprotein. *J. Virol.* **71**, 9808–9812 (1997).
24. Roben, P. *et al.* Recognition properties of a panel of human recombinant Fab fragments to the CD4 binding site of gp120 that show differing abilities to neutralize human immunodeficiency virus type 1. *J. Virol.* **68**, 4821–4848 (1994).
25. Moore, J. P. *et al.* Exploration of antigenic variation in gp120 from clades A through F of human immunodeficiency virus type 1 by using monoclonal antibodies. *J. Virol.* **68**, 8350–8364 (1994).
26. Wyatt, R. *et al.* Functional and immunologic characterization of human immunodeficiency virus type 1 envelope glycoproteins containing deletions of the major variable regions. *J. Virol.* **67**, 4557–4565 (1993).
27. Watkins, B. A. *et al.* Immune escape by human immunodeficiency virus type 1 from neutralizing antibodies: evidence for multiple pathways. *J. Virol.* **67**, 7493–7500 (1993).
28. Karlsson, G. *et al.* Characterization of molecularly cloned simian-human immunodeficiency viruses causing rapid CD4$^+$ lymphocyte depletion in rhesus monkeys. *J. Virol.* **71**, 4218–4225 (1997).
29. Wu, H., Kwong, P. D. & Hendrickson, W. A. Dimeric association and segmental variability in the structure of human CD4. *Nature* **387**, 527–530 (1997).
30. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Acknowledgements. We thank Y. McLaughlin and S. Farnum for manuscript preparation and M. Farzan for discussion. This research was supported by the NIH, the Howard Hughes Medical Institute, the American Foundation for AIDS Research and the Aaron Diamond Foundation. This work was made possible by gifts from the late William McCarty-Cooper, from the G. Harold and Leila Y. Mathers Foundation, from the Friends 10, and from Douglas and Judi Krupp.

Correspondence and requests for materials should be addressed to J.G.S.

Dynein arms are oscillating force generators

Chikako Shingyoji, Hideo Higuchi*†, Misako Yoshimura, Eisaku Katayama‡ & Toshio Yanagida*§

Department of Biological Sciences, Graduate School of Science, University of Tokyo, Hongo, Tokyo 113-0033, Japan
** Yanagida BioMotron Project, ERATO, JRDC, Senba-Higashi 2-4-14, Mino, Osaka 562-0035, Japan*
‡ Department of Fine Morphology, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-0071, Japan
§ Department of Physiology, Osaka University Medical School, Suita, Osaka 565-0871, Japan

Eukaryotic flagella beat rhythmically[1]. Dynein is a protein that powers flagellar motion, and oscillation may be inherent to this protein[2–5]. Here we determine whether oscillation is a property of dynein arms themselves or whether oscillation requires an intact axoneme[6], which is the central core of the flagellum and consists of a regular array of microtubules. Using optical trapping nanometry[7,8], we measured the force generated by a few dynein arms on an isolated doublet microtubule. When the dynein arms on the doublet microtubule contact a singlet microtubule and are activated by photolysis of caged ATP[8], they generate a peak force of ~6 pN and move the singlet microtubule over the doublet microtubule in a processive manner. The force and displacement oscillate with a peak-to-peak force and amplitude of ~2 pN and ~30 nm, respectively. The geometry of the interaction indicates that very few (possibly one) dynein arms are needed to generate the oscillation. The maximum frequency of the oscillation at 0.75 mM ATP is ~70 Hz; this frequency decreases as the ATP concentration decreases. A similar oscillatory force is also generated by inner dynein arms alone on doublet microtubules that are depleted of outer dynein arms. The oscillation of the dynein arm may be a basic mechanism underlying flagellar beating.

A crucial step in our study was the preparation of functionally intact dynein arms. ATP was photoreleased to activate demembranated and elastase-treated[4] sea-urchin sperm flagella. This activation caused disintegration of the axoneme into individual doublet microtubules. This procedure exposed fresh, functional dynein arms[9] on the doublets (Fig. 1a). When a suspension of singlet microtubules was added, the singlet microtubules attached themselves to the doublets and, after flash photolysis of caged ATP, moved along the doublets. Figure 1c shows two microtubules attached to the dynein arms of a doublet. Using polarity-marked microtubules, we confirmed that the sliding movement was always towards the plus (fast-growing) end of the microtubule. This is consistent with the known polarity of force generation by dynein[9–12], indicating that the movement was caused by the dynein arms. The maximum velocity, obtained ~100 ms after flash photolysis, was 6.4 μm s^{-1} (see below). This is comparable to

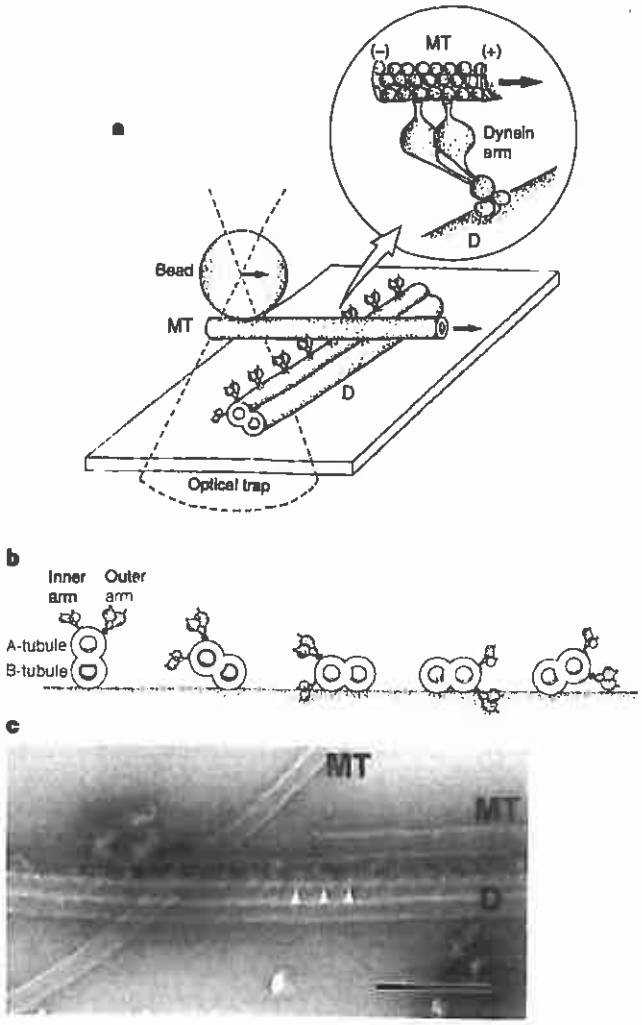


Figure 1 Interactions of singlet microtubules with dynein arms present on doublet microtubules. **a**, Schematic diagram of exposed dynein arms, on a doublet microtubule (D), interacting with a singlet microtubule (MT). The singlet microtubule is manipulated by means of an optically trapped streptavidine-coated bead. A two-headed arm[9] pulls the singlet microtubule in the direction of its plus (+) end, causing the bead to move away from the centre of the trap force (black arrows). **b**, Possible orientations of doublet microtubules on the glass surface (bottom) with outer and inner dynein arms. Dynein arms pointing upwards may be capable of interaction with a singlet microtubule (not shown). **c**, Electron micrograph showing two singlet microtubules (MT) interacting with dynein arms (white arrowheads) on a doublet microtubule (D). Scale bar, 100 nm.

† Present address: Department of Metallurgy, Faculty of Engineering, Tohoku University, Aramaki-aza-Aoba, Sendai 980-8579, Japan.

Nature © Macmillan Publishers Ltd 1998

Structure−Activity Relationships in the Binding of Chemically Derivatized CD4 to gp120 from Human Immunodeficiency Virus

Hui Xie,[#,▽] Danny Ng,[§,||] Sergey N. Savinov,[§,⊥] Barna Dey,[‡] Peter D. Kwong,[‡] Richard Wyatt,[‡] Amos B. Smith, III,[§] and Wayne A. Hendrickson*[,†]

Department of Pharmacology and Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, Pennsylvania 19104, Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, Maryland 20815-6789, and Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032

Received May 15, 2007

The first step in HIV infection is the binding of the envelope glycoprotein gp120 to the host cell receptor CD4. An interfacial "Phe43 cavity" in gp120, adjacent to residue Phe43 of gp120-bound CD4, has been suggested as a potential target for therapeutic intervention. We designed a CD4 mutant (D1D2F43C) for site-specific coupling of compounds for screening against the cavity. Altogether, 81 cysteine-reactive compounds were designed, synthesized, and tested. Eight derivatives exceeded the affinity of native D1D2 for gp120. Structure−activity relationships (SAR) for derivatized CD4 binding to gp120 revealed significant plasticity of the Phe43 cavity and a narrow entrance. The primary contacts for compound recognition inside the cavity were found to be van der Waals interactions, whereas hydrophilic interactions were detected in the entrance. This first SAR on ligand binding to an interior cavity of gp120 may provide a starting point for structure-based assembly of small molecules targeting gp120−CD4 interaction.

Introduction

Human immunodeficiency virus 1 (HIV-1[a]) is the primary cause of acquired immunodeficiency syndrome (AIDS).[1,2] Among 20 antiretroviral drugs that have been approved by FDA for clinical treatment of AIDS,[3] nearly all target either the reverse transcriptase or the protease of HIV-1. Currently, the continuing emergence of drug-resistant viruses and drug toxicity problems, as well as the absence of an effective AIDS vaccine, motivates the search for novel antiretroviral drugs.

Elements of virus entry are attractive antiviral targets because they are at the first step of the HIV-1 life cycle, preceding cellular infection. The entry of the virus is mediated by specific interactions between viral envelope glycoproteins (gp120 and gp41) and host cell surface receptors (CD4 and CCR5/CXCR4). HIV-1 virus first attaches to the host cell surface through recognition between gp120 and CD4.[4,5] CD4 binding induces extensive structural rearrangements in gp120, resulting in the exposure of a binding surface for the second host cell chemokine receptor, CCR5 or CXCR4,[6,7] and exposure of new epitopes on gp41.[8,9] The ensuing engagement of gp120 with the chemokine receptor triggers further conformational changes, which ultimately lead to the fusion of viral membrane with the host cell membrane.[10−12]

The interface between gp120 and CD4 is highly conserved among different HIV-1 isolates, as evidenced by multiple X-ray crystal structures of core gp120 proteins derived from different isolates in complex with D1D2 (the first two immunoglobulin-like domains of CD4) and a Fab fragment of antibody, 17B or X5.[13−15] In these complexes, the D1 domain of CD4 binds in a depression on the core gp120 formed by residues from all three domains of gp120, including the inner domain, the outer domain, and the bridging β-sheet. The binding region in CD4 is centered on residues 40−48 of the D1 domain, a stretch corresponding to the second complementarity-determining region (CDR2) of immunoglobulins. Among these residues, Phe43 alone contributes 23% of the total interactions.[14] At the gp120−CD4 interface in all structures, a deep hydrophobic cavity enclosed by conserved gp120 residues has been identified. The CD4 residue Phe43 sits right at the opening of this cavity, acting as a lid, and is the only cavity-interacting residue in CD4. This cavity is thus named "the Phe43 cavity"[13,14] and has been suggested as a potential target for therapeutic intervention.[13,14,16] Biochemical,[16−20] structural,[21,22] and thermodynamic[23] studies indicate that the Phe43 cavity-bearing CD4-bound gp120 adopts a conformation more rigidified than that of free gp120, and this conformation of gp120 has been estimated to be only sampled by a small fraction (12−15%) of free gp120.[24]

Most of the potent gp120−CD4 inhibitors identified to date are proteins or peptides.[25] A few gp120-directed small-molecular entry inhibitors, e.g., BMS-378806, BMS-488043,[26−29] and NBD-556,[30] have been identified by high-throughput screening, although their modes of action remain controversial.[31−33]

* To whom correspondence should be addressed. Address: Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032. Phone: (212) 305-3456. Fax: (212) 305-7379. E-mail: wayne@convex.hhmi.columbia.edu.

[#] Department of Pharmacology and Department of Biochemistry and Molecular Biophysics, Columbia University.

[▽] Present address: Elusys Therapeutics, 25 Riverside Drive, P.O. Box 102, Pine Brook, NJ 07058.

[§] University of Pennsylvania.

[||] Present address: Medicinal Chemistry, FibroGen Inc., 225 Gateway Boulevard, South San Francisco, CA 94080.

[⊥] Present address: Department of Chemistry, Purdue University, West Lafayette, IN 47907.

[‡] National Institute of Allergy and Infectious Diseases, National Institutes of Health.

[†] Howard Hughes Medical Institute and Department of Biochemistry and Molecular Biophysics, Columbia University.

[a] Abbreviations: HIV-1, human immunodeficiency virus 1; AIDS, acquired immunodeficiency syndrome; CDR, complementarity-determining region; D1D2, two N-terminal IgG-like domains of CD4; IC_{50}, the half-maximal inhibitory concentration; D1D2F43C-"compound X", modification product of D1D2F43C by compound X; HTS, high-throughput screening; WT, wild type; FL, full length; DTNB, 5,5′-dithiobis(2-nitrobenzoic acid); NEM, *N*-ethylmaleimide; ELISA, enzyme-linked immunosorbent assay.

10.1021/jm070564e CCC: $37.00 © 2007 American Chemical Society
Published on Web 09/06/2007

257

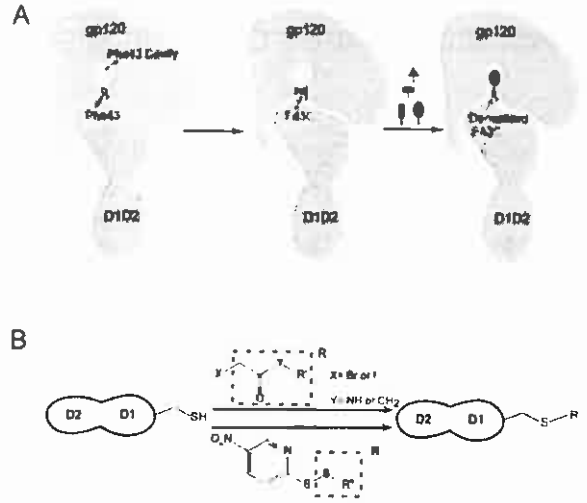


Figure 1. Screen for **Phe43** cavity-targeting compounds using derivatized CD4. (A) Design of modified D1D2F43C for targeting the Phe43 cavity in gp120. The complex of gp120 and the D1D2 domains of CD4 are drawn schematically with gp120 and D1D2 colored in gray and salmon, respectively. Residue Phe43, right at the entrance of the Phe43 cavity, is mutated to a chemically reactive cysteine for specific conjugation of a library of cysteine-reactive compounds (in green). The generated D1D2F43C derivatives are then screened for their affinity for gp120. (B) Modification of F43C of D1D2 by haloacetamides or halopropanones (shown on top of the arrows, Y = NH or CH_2, respectively) or 5-nitro-2-pyridinesulfenyl reagents (shown below the arrows).

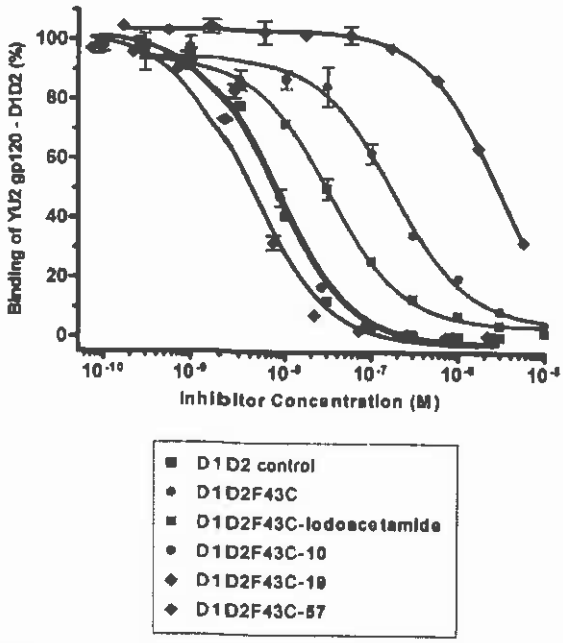


Figure 2. Representative curves for the inhibition of gp120−CD4 binding by D1D2 control, D1D2F43C, and D1D2F43C derivatives (D1D2F43C-iodoacetamide, D1D2F43C-**10**, D1D2F43C-**19**, and D1D2F43C-**57**) in a competition ELISA assay. Curves were generated from binding data using Prism 4 (GraphPad Software). The derived IC_{50} values were 7.31 ± 1.07, 206 ± 21, 33.3 ± 5.5, 7.76 ± 0.83, 3340 ± 510, and 4.14 ± 0.03 nM, respectively.

Possible reasons for the difficulty in identifying a small-molecule inhibiting gp120−CD4 interaction include the unusually large conformational flexibility of gp120 and the intrinsic difficulty of small-molecule drugs to target a complex protein−protein interface,[34] such as that between gp120 and CD4. A highly conserved interfacial hydrophobic pocket like the Phe43 cavity in gp120,[13,14,35] on the other hand, could be a good binding site for small molecules that block protein−protein binding either through a direct steric effect or through an allosteric mechanism. Moreover, such a site should be less prone to resistance-conferring mutations. Thus, in this work, instead of exploring the entire gp120−CD4 interface per se, we characterized the Phe43 cavity as a potential drug-targeting site. Because a large entropic penalty is associated with stabilization of the Phe43 cavity in gp120,[23] direct identification of cavity-targeting small molecules is difficult. By using chemically derivatized CD4 molecules, we could specifically target druglike fragments to the cavity and use them to characterize cavity properties. Herein, we report the construction of 81 modified CD4 molecules and the first structure−activity relationships for the interactions of the Phe43 cavity-targeting probes with gp120.

Results

Design Rationale. The exact location of residue Phe43 of CD4 at the entrance of the cavity makes this locale the best attachment point for cavity-targeting compounds. Cysteine, among all natural amino acids, best enables chemoselective conjugation. Thus, we used a two-domain CD4 (D1D2) scaffold with **Phe43** replaced by a cysteine (D1D2F43C) as the template for site-specific **delivery** of diverse probes to the cavity (Figure 1). D1D2F43C, containing all the essential elements of CD4 in binding gp120 except for the Phe43 side chain, stabilizes gp120 in the cavity-bearing conformation but binds to gp120 much more weakly than CD4 (IC_{50}(D1D2F43C) = 206 ± 21 nM vs IC_{50}(D1D2) = 7.31 ± 1.07 nM, Figure 2). This substitution allows the compatibility of a chemical group and the cavity to be easily judged by comparing the gp120-binding affinities of the derivatized D1D2F43C proteins to that of D1D2F43C (Figure 1).

We first tested each of three commonly used classes of sulfhydryl-specific reagents [i.e., iodoacetamide, 5,5′-dithiobis-(2-nitrobenzoic acid) (DTNB), and *N*-ethylmaleimide (NEM)] to identify the best thiol-reacting module for linking compounds to Cys43. Although all three compounds were found to modify D1D2F43C with satisfactory completeness (>98%, 80%, and >98%, respectively) (Supporting Information), iodoacetamide was the only reagent that produced a CD4 derivative with gp120-binding affinity higher than that of the unmodified template D1D2F43C, demonstrated by a 6-fold lower IC_{50} value (Figure 2 and Supporting Information). Consequently, haloacetamide was chosen as the primary starting module for construction of the first thiol-reactive compound library, which included 41 bromoacetamides (compounds **1** (SNS-1) to **41** (SNS-41)) and 7 5-nitro-2-pyridyldisulfides (compounds **42** (SNS-42) to **48** (SNS-48)), designed through a computer-assisted molecular complementarity search using GrowMol.[36,37] On the basis of the screening results from the first library, we further designed a second library of 42 bromoacetamides, 1 bromopropanone, 1 bromoethylamine, and 2 5-nitro-2-pyridyldisulfides (compounds **49** (DN-10) to **94** (DN-271)) to complete the analysis of structure−activity relationships of cavity-filling probes (Table 1 in Supporting Information).

All D1D2F43C derivatives were generated employing nucleophilic reactions between thiolate anions of D1D2F43C and electrophiles of the two compound libraries synthesized (Figure 1). The derivatized proteins were named D1D2F43C-"compound" according to the compounds used to derive them; e.g., D1D2F43C-**1** is the product of the modification of D1D2F43C by compound **1**. A "D1D2 control" was prepared as a modifica-

Table 1. Summary of Modification of D1D2F43C by Both Compound Libraries[a]

	initial library		second library		total	
cysteine-modifying compound	reacted, no.	unreacted, no.	reacted, no.	unreacted, no.	reacted, no.	unreacted, no.
haloacetamides	39	2	35	9	74	11
disulfides	6	1	1	1	7	2
total	45	3	36	10	81	13

[a] Reacted: completeness, ≥80%. Unreacted: completeness, <80%.

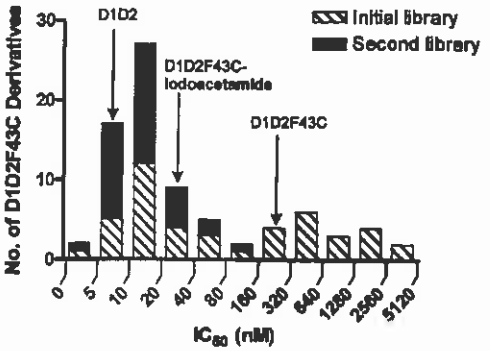


Figure 3. Distribution of the IC_{50} values of the D1D2F43C derivatives generated from the initial and second libraries.

tion control by using compound **40** to "mock-modify" wild type (WT) D1D2 in the same way as described above.

Overview of Binding Activities of CD4 Derivatives. Out of 94 compounds synthesized, 81 modified D1D2F43C with >80% completeness (Table 1 and Supporting Information). The resulting 81 derivatives were tested for their abilities to inhibit the interaction of gp120 with D1D2 in a competition ELISA assay (Figure 2), from which an IC_{50} value (the concentration for obtaining 50% inhibition of gp120−CD4 binding) was determined for each of the derivatives. Because the D1D2F43C derivatives bind gp120 competitively with D1D2 and the IC_{50} value of the D1D2 control is in good agreement with the reported K_d of D1D2 and gp120,[38,39] the measured IC_{50} values were regarded as the estimations of affinities between the derivatives and gp120.

The derivatives from the initial library had widely spread IC_{50} values, whereas those from the second library, optimized from results of initial libraries, all bound gp120 better than D1D2F43C (Figure 3). Furthermore, 90% of the proteins derivatized from the second library had IC_{50} values lower than that of D1D2F43C-iodoacetamide, the parent derivative for all haloacetamide-derived proteins. Out of all derivatives, eight had IC_{50} values lower than that of D1D2 (7.3 nM). The best derivative was D1D2F43C-**57**, with an IC_{50} value almost half that of D1D2 (Figure 2). The affinity of the derivatized D1D2F43C proteins for gp120, however, reached a plateau with an IC_{50} value of around 4 nM even after extensive optimization.

Validation of the Binding Sites for CD4 Derivatives. We confirmed that the chemical probes presented by CD4 derivatives bind into the Phe43 cavity of gp120 by comparing the affinities of the derivatives for gp120 to their affinities for a Phe43 cavity-filled mutant of gp120. The cavity-filled gp120 mutant, namely, gp120 S375W/T257S, has the cavity-lining residue Ser375 mutated to a tryptophan, which fills the cavity (Figure 4A).[20,22,50] The other mutation, T257S, was incorporated to stabilize gp120 S375W. Unmodified D1D2 and D1D2F43C, which do not bind into the cavity, bound the cavity-filled mutant of gp120 equally well as WT gp120, as indicated by the identical IC_{50} values in competition assays (Figure 4B,C). In contrast, all except one of the D1D2F43C derivatives, regardless of their affinities for gp120, exhibited dramatically lower affinities (4- to 150-fold higher IC_{50} values) for gp120 S375W/T257S than for WT gp120 (Figure 4B,C), which is consistent with the Phe43 cavity being the binding sites for the chemical groups linked to Cys43 in these derivatives. Among the derivatives with nanomolar IC_{50} values, the molecular weights of the compounds were found to correlate well with the loss of affinity for cavity-filled gp120 (data not shown). D1D2F43C-iodoacetamide, the only derivative with equal affinity for gp120 S375W/T257S and WT gp120, had only a small acetamide moiety attached to residue Cys43, which was not expected to protrude into the cavity based on our modeling results (data not shown).

SAR of Binding of CD4-Attached Compounds to gp120: *N*-Alkylacetamide Derived D1D2F43C. Nine derivatives were evaluated for their gp120-binding affinities to study the effect of the addition of alkyl groups at the acetamide nitrogen of the parent derivative D1D2F43C-iodoacetamide (Table 2). Small alkyl groups such as isopropyl (**1**) reduced the affinity greatly (5-fold), whereas larger alkyl groups such as isobutyl (**2**) and 6-hydroxyhexyl (**9**) did not have much affect on the affinity. An even bulkier cyclic group [i.e., the cyclohexyl group (**3**)] increased the binding affinity more than 2-fold with an IC_{50} of 13 nM. This positive effect was further enhanced when an extra methylene linker was inserted between the acetamide nitrogen and the cyclohexyl group to yield the cyclohexylmethyl group (**4**). Addition of even bigger groups (**6**, **56**, and **8**) reduced the binding affinities drastically to micromolar levels, suggesting that the volume limit of the cavity had been reached.

SAR: *N*-Arylacetamide Derived D1D2F43C. Unlike alkyl groups, aryl groups were much more favored when attached to the acetamide nitrogen. The highest binding affinity in this group (Table 3) was observed with D1D2F43C-**10** (IC_{50} = 7.76 nM), which has a phenyl group at the acetamide nitrogen. Insertion of up to three methylene linkers between the nitrogen and phenyl groups (**28**, **40**, and **51**) decreased binding affinity slightly with an ethylene group being the best linker. Interestingly, branching at the benzylic position resulted in a drastic (5- to 10-fold) loss in affinity (**29** and **30** compared to **28**). Similar to the phenyl group, a naphthyl group (**14**) at the acetamide nitrogen was also favorable, but the insertion of additional linkers (**32**, **64**, **31**, and **66**) between the naphthyl unit and the acetamide nitrogen attachment was much less tolerated.

In light of the preference for phenyl- and naphthyl-like groups in the cavity, we also synthesized compounds containing five-membered aromatic rings linked to the acetamide. The findings (Table 4) were very similar except for compound **67**, the derivative of which has affinity 3-fold lower than those of the derivatives of **10** and **87**, suggesting possibly unfavorable interaction between gp120 and the sulfur atom at position 2 in thiophen.

SAR: Substituted *N*-Phenylacetamide Derived D1D2F43C. Extensive substitutions on the phenyl group of compound **10** were employed both to screen for higher-affinity binders compared to the parent compound **10** and to characterize further the chemical preferences of the Phe43 cavity. Eighteen compounds were synthesized with various substitutions at the para position (Table 5A). In general, small substitutions ranging from

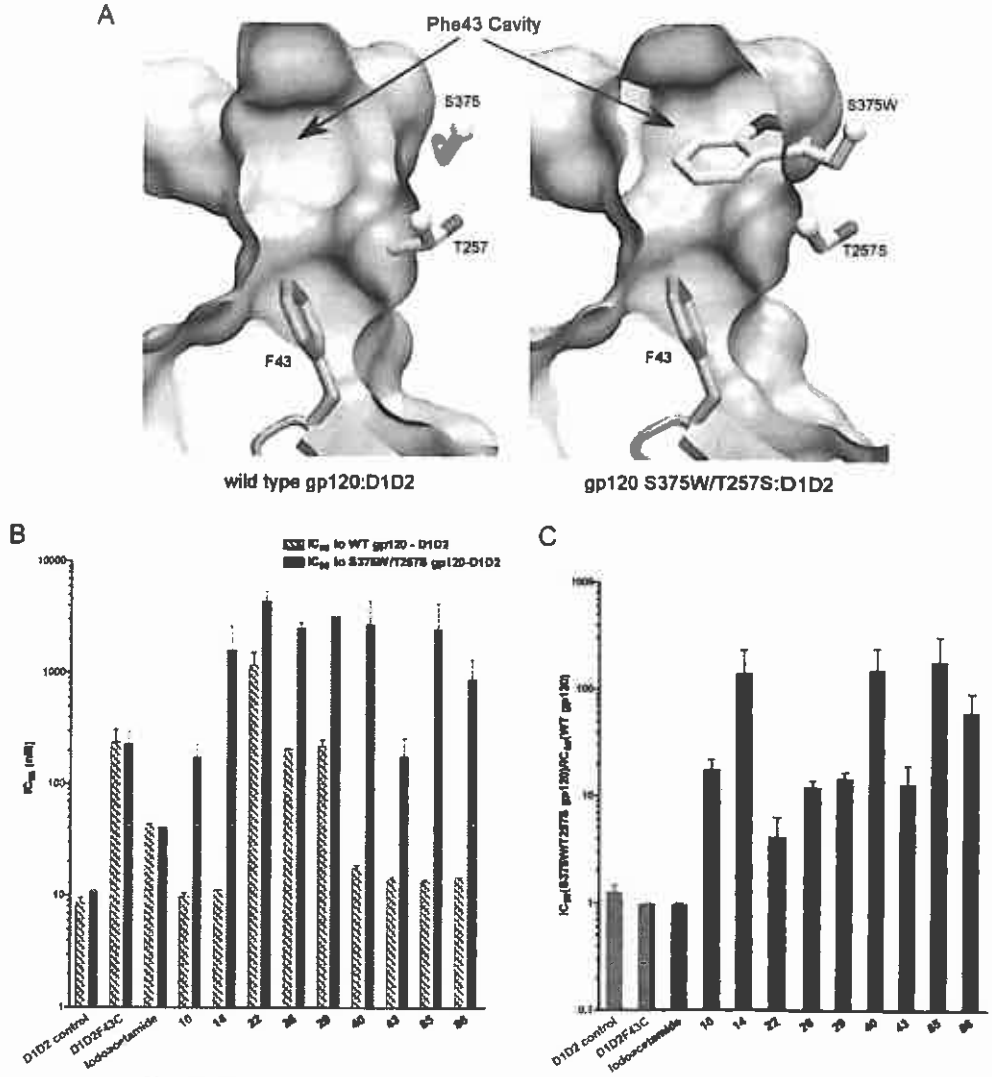


Figure 4. Binding of D1D2 conjugates to cavity-filled gp120. (A) Sliced-open surface representations of the Phe43 cavities of wild type core gp120 (left, PDB code 1RZJ) and cavity-filled core gp120 mutant S375W/T257S (right, PDB code 2NXZ) bound to CD4. CD4 and gp120 molecules are colored in gray and salmon, respectively. Side chains and Cα atoms of gp120 residues 375 and 257, as well as those of CD4 residue F43, are shown in ball-and-stick model in the coloring scheme: CD4 carbon atoms (salmon), gp120 carbon atoms (gray), nitrogen atom (blue) and oxygen atoms (red). Surfaces in both panels were calculated based on the respective gp120 model in which the side chains of residues 375 and 257 except for Cβ atom were removed. Note that the cavity has a narrow entrance close to the tip of phenyl ring of F43 and that a water channel is located left to the cavity. (B) Side by side comparisons of IC_{50} values of D1D2 variants on the binding of D1D2 to YU2 WT (wild type) gp120 to their IC_{50} values on the binding of D1D2 to YU2 S375W/T257S gp120. (C) Ratios of two IC_{50} values.

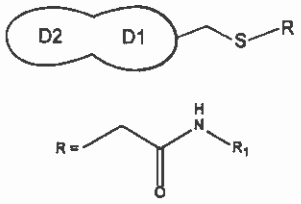
methyl (**73**) to ethoxy (**74**) groups had limited effect on the activities, but larger groups such as isopropoxy (**22**) and benzyloxy (**26**) groups reduced the affinities significantly. The highest binding affinities of this whole study were observed when compound **10** was substituted at the para position by a nitro (**57**, 4.14 nM), an isopropyl (**12**, 4.84 nM), or an ethyl (**76**, 5.87 nM) group. Interestingly, derivatives with hydroxy-containing substituents at the para position invariably showed IC_{50} values higher than those of similar derivatives with the hydroxy groups replaced by the methyl groups (**59** and **73**; **77** and **76**). This finding indicates a disfavoring of electron-donating groups at the para position. An isopropanol has been reported to occupy the Phe43 cavity in the refined 2.2 Å structure of HXBc2 gp120:D1D2:17b Fab,[13] suggesting a possible starting point for the design of cavity-binding compounds. Substitution of **10** at the para position by an isopropanol group (**94**) led to slightly reduced affinity, indicating disfavor of the isopropanol group at this orientation.

Eight groups were tested for meta-position substitution of the phenyl ring in compound **10**. Five of these (methyl, hydroxy, nitro, 2-hydroxyethoxy, and benzyloxy groups) were also tested at the para and ortho positions. A much higher tolerance for bulkier groups was noticed at the meta position compared to the para position (Table 5B). Smaller groups at the meta position, on the other hand, bound comparably well or somewhat worse (nitro) than at the para position.

None of seven different substitutions at the ortho position of the phenyl group was able to enhance the affinity (Table 5C). Interestingly, the smallest group tested, methyl (**78**), was the second to the least favored at this position with an IC_{50} of 24.5 nM; yet a much bulkier group benzyloxy (**85**) performed best at this position with an IC_{50} of 8.87 nM, comparable to that of the parent compound **10**.

SAR: Role of the Acetamide Moiety. Modification of D1D2F43C by the simplest haloacetamide-based compound that we tested (iodoacetamide) increased the binding affinity for

Table 2. Chemical Structures of Cysteine-Modifying Compounds (R as *N*-Alkyacetamide) and IC_{50} Values of the Corresponding Derivatized D1D2F43C for Inhibition of the Binding of D1D2 to YU2 gp120[a]



Compound	R_1	IC_{50} ± SD of Derivatized D1D2 (nM)
Iodoacetamide	H	33.3 ± 5.5
1		184 ± 39
2		29.4 ± 2.2
9		33.0 ± 9.7
3		13.0 ± 2.1
4		7.79 ± 2.08
5		17.0 ± 2.5
6		223 ± 49
56		42.3 ± 5.9
8		1680 ± 500

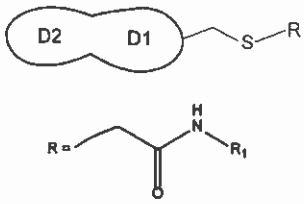
[a] Corrected IC_{50} values are presented as the mean ± SD values from two to three independent experiments.

gp120 by 6-fold (Figure 2). On the basis of our modeling result (data not shown), the acetamide moiety binds at the vestibule to the Phe43 cavity, which is the binding site for residue Phe43 of WT CD4. The 6-fold affinity increase suggests a good complementarity of the acetamide group to the vestibule, possibly due to the relatively small size as well as rigid and nonbranching shape. The narrower shape of the acetamide moiety compared to the phenyl ring of Phe43, however, probably limits the ability to make extensive interactions with gp120 around the cavity opening, in contrast to Phe43, as suggested by affinity of D1D2 that is 5-fold higher than that of D1D2F43C-iodoacetamide.

We used compound **10** as a positive background to identify the gp120-binding elements of the acetamide group. Replacement of the acetamide nitrogen by a carbon (**71**) decreased the binding affinity by 3-fold (Table 6), indicating that the acetamide nitrogen plays a significant role in the interaction of the acetamide moiety with gp120, probably through an interaction for which the methylene carbon cannot substitute, such as the hydrogen bonding. Compound **69**, in which the carbonyl oxygen of the acetamide moiety was removed, failed to modify D1D2F43C (Supporting Information). Therefore, the significance of the carbonyl oxygen in the acetamide is unclear. Some indication may be derived from comparison of **71** with a disulfide compound **63**, which resembles **71** without the carbonyl oxygen (Table 6). The similar IC_{50} values of these two derivatives suggest either that the carbonyl oxygen is not important for the interaction or that its deletion is compensated in some way by substitution of sulfur for carbon.

SAR: Mixed-Disulfide D1D2F43C Derivatives. Only a small series of disulfide compounds were investigated in this study because of their limited solubility and poor reactivity. All

Table 3. Chemical Structures of Cysteine-Modifying Compounds (R as *N*-Arylacetamide) and IC_{50} Values of the Corresponding Derivatized D1D2F43C for Inhibition of the Binding of D1D2 to YU2 gp120[a]



Compound	R_1	IC_{50} ± SD of Derivatized D1D2 (nM)
10		7.76 ± 0.83
28		15.2 ± 0.7
29		88.7 ± 1.5
30		166 ± 22
40		10.6 ± 2.9
51		16.4 ± 2.6
14		10.4 ± 0.8
32		75.7 ± 11.2
64		69.3 ± 1.8
31		12.2 ± 0.8
66		146 ± 4

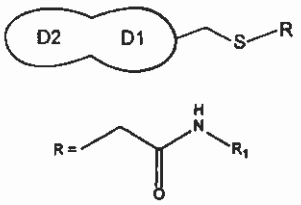
[a] Corrected IC_{50} values are presented as the mean ± SD values from two to three independent experiments.

derivatives contain aryl groups that were linked to Cys43 through disulfide bonds (Table 7). The incorporation of most of these aryl groups (6 out of 7) significantly increased activities of derivatives in binding gp120 when compared with D1D2F43C. Similar to the SAR in the substituted *N*-phenylacetamide derived D1D2F43C, addition of an isopropyl group but not other bulkier groups at the para position of the benzyl group in compound **42** slightly enhanced affinity (**43**).

Discussion

Screen for Phe43 Cavity-Targeting Inhibitors Using Derivatized CD4. The accessible surface of gp120 is thought to be largely composed of both highly variable and heavily glycosylated regions,[14,16] and the conserved Phe43 cavity is thought to be poorly formed in free gp120.[24] Hits identified in high-throughput screens against gp120 are likely to recognize variable regions, and thus, the resulting compounds are expected to be highly susceptible to resistance because of high mutation rates. In contrast, the emergence of resistant viruses to inhibitors targeting the conserved Phe43 cavity should be rare, if at all possible. The Phe43 cavity is thought to exist preferably in the relatively rigidified gp120 state when bound to CD4, and it is disrupted in the free SIV core gp120[21] and the HIV gp120 complex with Fab b12.[22] Mutation of the cavity-lining residue Ser375 to a bulky tryptophan fills the Phe43 cavity and drives gp120 into a conformation closer to that of the CD4-bound state.[20] The large entropy change associated with the transition

Table 4. Chemical Structures of Cysteine-Modifying Compounds (R as *N*-Thiophen(furan)acetamide) and IC_{50} Values of the Corresponding Derivatized D1D2F43C for Inhibition of the Binding of D1D2 to YU2 gp120[a]



Compound	R_1	IC_{50} ± SD of Derivatized D1D2 (nM)
87		7.40 ± 1.04
67		18.2 ± 1.8
68		13.1 ± 0.5
53		18.3 ± 2.1
55		18.8 ± 0.5

[a] Corrected IC_{50} values are presented as the mean ± SD values from two to three independent experiments.

of free gp120 to CD4-bound gp120[23] undoubtedly determines that gp120 poorly populates cavity-bearing conformations,[24] which reduces the chance of identifying hits targeting the Phe43 cavity in conventional high-throughput screening (HTS).

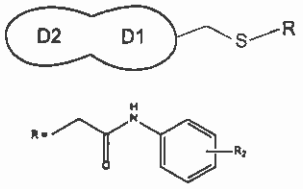
New methods for identifying lead compounds or fragments of lead compounds as inhibitors of protein−protein interactions have emerged recently as alternatives to HTS. These include structure-based screening by NMR techniques and X-ray crystallography, in silico screening, phage display, rational discovery, and tethering.[40,41] One of these approaches, tethering,[42] identifies small-molecule leads by screening the compounds covalently linked to the targeting site. This method, however, is unlikely to be applicable for identifying inhibitors for gp120 because of the high intrinsic flexibility of this target. Moreover, highly glycosylated gp120 does not readily permit the use of mass spectrometry in identifying suitable compounds in the tethering method. Unlike other approaches, our method focuses on the screening of the compounds that are covalently attached to a protein ligand (CD4) against the target protein (gp120). CD4, as the natural ligand for gp120, stabilizes formation of the Phe43 cavity in gp120. In addition, using a gp120 ligand such as CD4 as the ligand-attachment point in our method enables the quantitative determination of structure−activity relationships on the small-molecule fragments, which would otherwise bind gp120 too weakly for accurate affinity determination using traditional binding assays. A parallel study recently proposed a similar idea; the protocol entailed assembly of a library of peptide−small molecule hybrids via cysteine modification for studying SAR between the target protein and small molecules;[43] however, to date, no screening results have been reported.

Although our results are fully consistent with binding small molecules into the specific target site, surely the CD4 scaffold to which these chemicals are attached must also contribute to the binding properties. Scaffold constraints also affect binding properties in more conventional combinatorial chemistry approaches, but in our approach constraints from the larger D1D2 moiety might have a more substantial impact. Various aspects of the data nevertheless indicate that the SAR is indeed reflective of intrinsic molecular complementarity between the compounds and the cavity, at least for those derivatives with higher affinities than the unmodified template, D1D2F43C. The observed binding affinities for gp120 are widely distributed for the various chemically derivatized CD4 molecules, and these affinities are drastically lowered for cavity-filled gp120. This proves that the binding affinities of the derivatives to gp120 are sensitive to the interaction between the template-attached cavity-probing compounds and the cavity. Moreover, in the one possibility for comparison, this sensitivity is independent of the scaffold. That is, the enhancement (4-fold) of gp120-binding affinity upon addition of a cavity-targeting phenyl group to the vestibule-binding template D1D2F43C-iodoacetamide (vs D1D2F43C-**10**) is found to be the same as that gained (4-fold) through addition of a phenyl group to the existing phenyl group equivalent of that of Phe43 in a CD4 miniprotein mimetics F23.[35] D1D2F43C and the miniprotein scaffold are radically different in size, have different vestibule-binding groups, and have substantially different intrinsic affinities; thus, the high similarity in relative binding characteristics suggests that high-affinity derivatives probably bind the cavity with minimum constraints from the template. Therefore, the SAR derived from this study likely reflects true binding preferences of the Phe43 cavity.

Characterization of the Phe43 Cavity. The Phe43 cavity has been observed in the crystal structures of CD4-bound core gp120[13−15] to have an entrance narrower than the main body of the cavity (Figure 4A). Modeling of the acetamide group onto the Cys43 of D1D2 positioned the acetamide nitrogen close to the narrow cavity entrance of core gp120 (data not shown). Disfavoring of branching at the first carbon linked to the acetamide nitrogen, such as seen with compounds **1**, **29**, and **30** (Tables 2 and 3), suggests that the narrow cavity entrance is also present in the full-length (FL) gp120 used in this SAR study and that the acetamide nitrogen atom is near this constriction. Hydrophilic interaction, such as a hydrogen bond, is likely to be part of the recognition mechanism at this site, as discussed earlier. In the disulfide-linked derivatives, branching at the carbon next to the sulfur also had a negative effect on binding affinity (**44**, Table 6). Although the branching point is further away from the cavity than the acetamide nitrogen, the large branched phenyl group in **44** could potentially reach the cavity entrance and cause unfavorable steric clashes. The repulsions caused by the branching at the cavity entrance could be partially released by insertion of a methylene linker between the acetamide nitrogen and the branching groups, as demonstrated by compounds **6** and **56** (Table 2). Preference of aromatic groups over aliphatic groups by the cavity (Tables 2−7) is also in agreement with the presence of a narrow entrance to cavity where the rigid and planar shapes of aromatic groups should be preferred. Additional methylene linkers between the nitrogen atom and aromatic groups, unlike in the case of the alkyl groups, affected binding adversely (Tables 3 and 4), albeit only slightly. One plausible explanation is that while the phenyl group fits well at the narrow cavity entrance, the aromatic ring can no longer maintain all the favorable van der Waals contacts once positioned in the broader space inside the cavity because of the insertion of methylene linkers.

The para position was found to be the best position to substitute the aromatic groups, such as the phenyl ring in compound **10** (Table 5), for increase of gp120-binding affinity. The best substituents found are nitro (**57**) and isopropyl (**12**) groups. Although the groups possess different electronic proper-

Table 5. Chemical Structures of Cysteine-Modifying Compounds (R as Substituted *N*-Phenylacetamide) and IC_{50} Values of the Corresponding Derivatized D1D2F43C for Inhibition of the Binding of D1D2 to YU2 gp120[a]



A. para substituents

Compound	R_2	IC_{50} ± SD of Derivatized D1D2 (nM)
73		7.07 ± 1.51
59		11.4 ± 0.2
50		7.08 ± 1.24
76		5.87 ± 0.17
77		10.6 ± 3.0
12		4.84 ± 0.83
13		8.14 ± 0.38
57		4.14 ± 0.03
54		9.87 ± 0.67
94		10.9 ± 1.3
74		7.23 ± 0.14
22		1500 ± 330
88		16.4 ± 0.7
27		1210 ± 340
23		637 ± 202
17		402 ± 68
16		46.2 ± 1.3
26		290 ± 84

B. meta substituents

Compound	R_2	IC_{50} ± SD of Derivatized D1D2 (nM)
81		6.59 ± 0.89
82		9.82 ± 1.92
79		7.12 ± 1.3
93		11.4 ± 0.1
89		14.9 ± 0.4
92		22.0 ± 0.4
84		14.3 ± 0.25
91		36.2 ± 6.5

C. ortho substituents

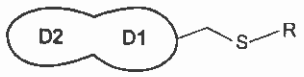
Compound	R_2	IC_{50} ± SD of Derivatized D1D2 (nM)
78		24.5 ± 1.0
83		9.06 ± 0.41
80		17.8 ± 4.0
86		12.9 ± 1.4
90		17.6 ± 1.7
15		51.9 ± 6.3
85		8.87 ± 1.09

[a] Corrected IC_{50} values are presented as the mean ± SD values from two to three independent experiments.

ties but similarities in molecular shape, the similar effects on binding of these two groups suggest that they may enhance binding solely through van der Waals interaction with cavity residues. This finding, together with the adverse effect on the affinity of many hydrophilic substituents on the aromatic ring (Tables 4 and 5), indicates that van der Waals interactions may dominate the recognition of the tethered compounds within the cavity. This conclusion is consistent with the hydrophobic nature of the cavity seen in the crystal structures of core gp120.[13,14] The finding that an isopropanol molecule occupies the cavity in the crystal structure of core gp120[13] is not inconsistent with this observed preference for hydrophobic groups in the cavity because the hydroxy group of isopropanol is hydrogen-bonded to two water molecules but not to any residue in gp120.

Although substitutions at the meta and ortho positions of the phenyl ring in compound **10** did not result in the best binding to gp120, much larger groups than those allowed at the para position can be substituted at these positions without affecting binding affinity (Table 5). This finding suggests that the meta- or ortho-substituted phenyl groups may bind in the cavity in differ-

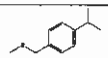
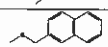
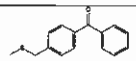
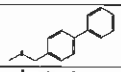
Table 6. Chemical Structures of Cysteine-Modifying Compounds (Role of the Acetamide Moiety) and IC_{50} Values of the Corresponding Derivatized D1D2F43C for Inhibition of the Binding of D1D2 to YU2 gp120[a]



Compound	R	IC_{50} ± SD of Derivatized D1D2 (nM)
10		7.76 ± 0.83
63		24.7 ± 8.6
71		26.4 ± 3.9

[a] Corrected IC_{50} values are presented as the mean ± SD values from two to three independent experiments.

Table 7. Chemical Structures of Cysteine-Modifying Compounds (Mixed-Disulfide Compounds) and IC_{50} Values of the Corresponding Derivatized D1D2F43C for Inhibition of the Binding of D1D2 to YU2 gp120[a]

Compound	R	IC_{50} ± SD of Derivatized D1D2 (nM)
42		11.5 ± 1.8
43		9.14 ± 1.99
44		1380 ± 100
45		14.3 ± 0.6
47		27.0 ± 1.6
48		14.9 ± 0.0
63		24.7 ± 8.6

[a] Estimated IC_{50} values are presented as the mean ± SD values from two to three independent experiments. Correction for compound absorbance was not possible here because of absorbance from the 5-nitro-2-pyridyl leaving group.

ent modes from the para-substituted phenyl groups. In addition, the binding may involve expansion of the cavity region in gp120.

As discussed, the Phe43 cavity is thought to be conserved in gp120 because it is composed of highly conserved residues across HIV-1.[13,14] However, whether the structure of the cavity in core gp120 resembles that in the FL gp120 was not clear. Our SAR study of the binding of the probes to the Phe43 cavity in FL gp120 now reveals a close similarity (narrow entrance and hydrophobic surface) between the cavities in FL and core gp120 and thus suggests that the core gp120 structure holds great promise for development of structure-based inhibitors targeting FL gp120.

In addition, our SAR analysis implies significant plasticity in gp120. All derivatives except for D1D2F43C-iodoacetamide were found to bind inside the Phe43 cavity, despite the fact that the derivatization groups on some of the derivatives are much larger than the size of the cavity seen in the structures of core gp120.[13-15] Whether all parts of each compound could fit completely in the cavity is not clear, but it is reasonable to believe that gp120, even when bound to CD4, has substantial plasticity to undergo necessary structural rearrangements around the Phe43 cavity to accommodate bulky compounds, such as **85** (Table 5C), a derivative that maintains high affinity for gp120 (IC_{50} = 9 nM).

Future Development and Therapeutic Applications. Although gp120 retains significant plasticity during ligand binding in the presence of the CD4 scaffold, gp120 adaptability is nevertheless restricted; thus, potentially favorable interactions of compounds bound into the intrinsically flexible Phe43 cavity may not be realized fully. This proposal may explain the IC_{50} plateau of 4 nM reached after optimization of the derivatives. The potential penalty from occupying a cavity that is intrinsically flexible but stabilized by the D1D2 scaffold may prevent a further increase in affinity of derivatized D1D2F43C. A smaller CD4 scaffold such as a mimetic may comprise a more reasonable template for further optimization of cavity-binding fragments because it may structure gp120 to a lesser degree.

When bound to CD4 derivatives, gp120 probably adopts a rigidified conformation close to that of the CD4-bound form. CD4 derivatives, however, may not fully activate gp120 into a chemokine-receptor binding conformation because a cavity-binding CD4-mimetic CD4M33 (a scyllatoxin-based miniprotein with a biphenylalanine residue at the position corresponding to Phe43 of CD4) has been shown to promote chemokine-receptor binding 6-fold less than does CD4[44] and to induce 40% less entropic change in gp120 compared to F23, which has a phenylalanine replacing biphenylalanine in CD4M33.[35] Identification of cavity-lining residues critical for the inhibitory effect of BMS-368806, a small molecule that inhibits HIV-1 entry most likely by blocking CD4-induced conformational changes in HIV glycoproteins,[31] further supports that cavity-targeting CD4 derivatives may serve as entry inhibitors both by preventing viral attachment to CD4 and by partially precluding chemokine-receptor induced viral fusion. The immediate clinical application of the cavity-targeting fragments identified in this study would include incorporation of these moieties into current CD4-based therapeutic agents, such as CD4–IgG_2[45-47] and CD4–Ig,[48] for possibly better clinical efficacy.

Conclusion

We have designed and constructed a library of chemically derivatized CD4 and carried out a structure–activity study of these compounds binding to the Phe43 cavity in gp120. The Phe43 cavity in FL gp120 was found to be similar to that observed in the crystal structure of core gp120. It is hydrophobic and plastic in nature and has a narrow entrance, where hydrophilic interactions may play a significant role in binding ligands. Aromatic groups attached to an acetamide moiety and substituted with small groups at the para position were found to be most preferred by the cavity. The detailed characterization of the cavity and identification of the chemical preferences for cavity-binding ligands should benefit further development of the identified compounds into efficient gp120–CD4 inhibitors by incorporating them into either smaller protein-like scaffolds (e.g., miniproteins) or small molecules by fragment assembly. The unique utilization of a reactive ligand template (D1D2F43C) in screening compounds for targeting a receptor site (the Phe43 cavity in gp120) at the ligand–receptor interface may be applied to other systems where the stabilization of the targeting site on the receptor requires the presence of a ligand.

Experimental Section

Thiol-Reactive Reagents. Iodoacetamide, 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB), and *N*-ethylmaleimide (NEM) were purchased from Sigma-Aldrich. All other cysteine-modification compounds were synthesized in one to seven steps using commercially available compounds. Experimental details of the syntheses and characterization of these compounds are included in the Supporting Information.

Protein Reagents. The gp120 antibody 17b was produced in ascites and purified by Strategic BioSolutions (DE). Full-length YU2 gp120 was produced in Drosophila S2 cells as described previously[49] and purified similarly as described before.[13,14] WT and S375W/T257S YU2 were expressed in serum-free medium by transient transfection of HEK293T cells and were purified by either F105 (WT YU2) or 17b (S375W/T257S YU2) affinity columns as previously described.[50] The numbering of gp120 amino acid residues is based on the sequence of the prototypic HXBc2 strain.[51]

Preparation of Recombinant D1D2 and D1D2 Mutants. Recombinant two-domain CD4 (D1D2, residues 1–183) was cloned into NcoI and XhoI sites of vector pET24d (Novagen). D1D2 was

expressed as inclusion bodies in Rosetta (DE3) cells (Novagen) through leaky expression of T7lac promoter without IPTG induction in SuperBroth medium (BIO 101, Inc.) with addition of 0.2% glucose and 20 mM $MgCl_2$[52] at 37 °C for 24 h. The inclusion bodies were isolated from cells by sonication and centrifugation and were then washed three times with 2% Triton-X 100, 2 M urea, 5 mM EDTA, and 5 mM DTT in Tris·HCl buffer, pH 7.5. Purified inclusion bodies were solubilized by 6 M guanidine·HCl, 5 mM EDTA, 20 mM Tris-HCl, pH 7.5, 10 mM DTT, and 0.5 mM PMSF. Solubilized D1D2 proteins were refolded at 0.5 mg/mL in a refolding solution optimized based on the no. 10 condition of Foldit kit (Hampton Research): 50 mM Tris-HCl, pH 8, 10 mM NaCl, 1 mM KCl, 1 mM EDTA, 440 mM sucrose, 2 mM cysteine, and 0.5 mM cystine. Impurities and oligomeric D1D2 were removed by passing refolded D1D2 proteins through Q and SP Sepharose Fast Flow resins (Amersham) in batch mode at pH 10.5 and pH 6.2, respectively. As the last step, a size exclusion column Superdex 200 26/60 (Amersham) was used to separate any residual misfolded oligomeric D1D2 from monomeric D1D2. Purified soluble D1D2 contains residues 1−183 of CD4 and an additional glycine from the cloning vector at the N-terminus. D1D2F43C was created by site-directed mutagenesis and prepared similarly as WT D1D2.

Preparation of Derivatized D1D2F43C. Iodoacetamide, DTNB, and NEM were dissolved and stored as 20 mM solution in 0.5 M Na/K phosphate, pH 7.4. All other compounds including both bromo compounds and mixed-disulfide compounds were dissolved in DMF, DMSO, or ethanol as 20−60 mM stock. D1D2F43C protein at 1−3 mg/mL was first reduced by 2 mM dithiothreitol (DTT). Excessive DTT was then removed by PD-10 columns (Amersham), and at the same time the proteins were exchanged into proper reaction buffer: 0.1 M phosphate buffer, pH 7.4, 0.1 M NaCl, 1 mM EDTA, and 0.1 mM DTT for all bromo compounds, iodoacetamide, and DTNB; 0.1 M phosphate buffer, pH 7.4, 0.1 M NaCl, and 1 mM EDTA for 5-nitro-2-pyridinesulfenyl compounds; and 0.1 M phosphate buffer, pH 7.0, 0.1 M NaCl, 1 mM EDTA, and 0.1 mM DTT for NEM.[53] The thiol-reactive compounds were diluted into protein solutions to reach a final reaction concentration of 2 mM, 10 times excess of the free thiol groups in the reaction. The reactions were allowed to continue at 25 °C for 2 h in the dark. The final products, namely, derivatized D1D2F43C proteins, were separated from small molecules by passage through desalting columns (PD-10 columns, Amersham) followed by solvent exchange in Amicon Ultra-4 5K concentrators (Millipore) with reaction buffer 3 times. The completeness of the modifications was examined using protein mass spectrometry (MALDI-TOF, Columbia University Protein Core and DNA Sequencing Facility) as well as peptide mass spectrometry of trypsin-digested proteins (Columbia University Protein Core and DNA Sequencing Facility). This latter step was also used to confirm the correct site-directed incorporation of the compounds.

Concentration Determination of Derivatized D1D2F43C Proteins. The "estimated concentrations" of proteins were calculated from 280 nm absorbance of protein samples in PBS using the theoretic extinction coefficient of D1D2F43C. The concentrations of D1D2F43C variants derived from the halo compounds were then corrected by the corresponding correction factors:

$$C_{\epsilon,R}^{\text{measured}} = \frac{\epsilon_{\text{D1D2F43C}}^{\text{calcd}}}{\epsilon_{\text{D1D2F43C}}^{\text{calcd}} + \epsilon_{\text{R}}^{\text{measured}}}$$

where $\epsilon_{\text{D1D2F43C}}^{\text{calcd}}$ is the theoretic extinction coefficient of D1D2F43C and $\epsilon_{\text{R}}^{\text{measured}}$ is the experimentally determined extinction coefficient of thiol-reacting compounds in PBS. The "estimated concentration" was used for derivatized D1D3F43C from all other compounds.

The concentrations of a selected group of D1D2F43C derivatives were also determined by amino acid compositional analysis (Keck Biotechnology Resource Laboratory, Yale University) and were found to agree well with the results obtained by the UV absorbance method discussed above.

Competition ELISA Assay. The abilities of derivatized D1D2F43C proteins to bind gp120 and inhibit gp120−CD4 interaction were evaluated using a competition ELISA (enzyme-linked immunosorbent assay). Briefly, Immuno 2HB plates (Thermo LabSystem) were coated with 100 μL of 4 μg/mL recombinant D1D2 in coating buffer (10 mM phosphate buffer, pH 7.4, and 0.15 M NaCl) overnight at 4 °C. The plates were then blocked by 3% bovine serum albumin (BSA) (CalBiochem) in PBS (300 μL per well) for 2 h at 25 °C. Fifty nanograms of YU2 gp120 diluted in 100 μL of 3% BSA-PBS was added to the plates in the absence or presence of one of the D1D2 variants (either D1D2 or a D1D2F43C derivative) at various ranges of concentrations and incubated for 90 min at 25 °C. After removal of unbound gp120 by a four-time wash with PBST (0.05% Tween-20 in PBS), the bound gp120 was detected by a gp120 antibody 17b (100 μL, 1 μg/mL), which was further probed with a peroxidase-conjugated donkey antihuman antibody (Jackson ImmunoResearch, 1:20000 dilution, 100 μL). 3,3′,5,5′-Tetramethylbenzidine (Sigma) was used as the substrate for peroxidase, and the optical density (OD) was read at 450 nm. The binding of the residual gp120 to plate-bound D1D2 was calculated by using the following formula:

$$\text{binding (\%)} = 100 \times \frac{\text{OD}_{\text{gp120-competitor}} - \text{OD}_{\text{background}}}{\text{OD}_{\text{gp120}} - \text{OD}_{\text{background}}}$$

IC_{50} values were obtained by nonlinear regression fitting of the binding results by Prism 4 (GraphPad Software) using the formula of one-site competition as shown below:

$$\text{binding (\%)} = \text{bottom} + \frac{\text{top} - \text{bottom}}{1 + 10^{(\log(\text{concentration}) - \log(\text{IC}_{50}))}}$$

Acknowledgment. We thank Mary Ann Gawinowicz and Yelena Milgrom for mass spectrometry and N-terminal sequencing; Joseph Sodroski and Irwin Chaiken for discussions on experimental designs; Joseph Sodroski, Shi-Hua Xiang, and Christine Winter for providing reagents; and Joel Courter for help with organizing supplemental data. This research was supported in part by National Institutes of Health Grant P01-GM 056550.

Supporting Information Available: Experimental details of the syntheses, characterization results of thiol-reacting compounds, and comprehensive information of all derivatized D1D2. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Barre-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Vezinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **1983**, *220* (4599), 868−871.

(2) Gallo, R. C.; Salahuddin, S. Z.; Popovic, M.; Shearer, G. M.; Kaplan, M.; Haynes, B. F.; Palker, T. J.; Redfield, R.; Oleske, J.; Safai, B.; White, G.; Foster, P.; Markham, P. D. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* 1984, *224* (4648), 500−503.

(3) De Clercq, E. Emerging anti-HIV drugs. *Expert Opin. Emerging Drugs* **2005**, *10* (2), 241−273.

(4) Klatzmann, D.; Champagne, E.; Chamaret, S.; Gruest, J.; Guetard, D.; Hercend, T.; Gluckman, J. C.; Montagnier, L. T-Lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* **1984**, *312* (5996), 767−768.

(5) Dalgleish, A. G.; Beverley, P. C.; Clapham, P. R.; Crawford, D. H.; Greaves, M. F.; Weiss, R. A. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **1984**, *312* (5996), 763−767.

(6) Trkola, A.; Dragic, T.; Arthos, J.; Binley, J. M.; Olson, W. C.; Allaway, G. P.; Cheng-Mayer, C.; Robinson, J.; Maddon, P. J.; Moore, J. P. CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature* **1996**, *384* (6605), 184−187.

(7) Wu, L.; Gerard, N. P.; Wyatt, R.; Choe, H.; Parolin, C.; Ruffing, N.; Borsetti, A.; Cardoso, A. A.; Desjardin, E.; Newman, W.; Gerard, C.; Sodroski, J. CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature* **1996**, *384* (6605), 179−183.
(8) Dimitrov, A. S.; Xiao, X.; Dimitrov, D. S.; Blumenthal, R. Early intermediates in HIV-1 envelope glycoprotein-mediated fusion triggered by CD4 and co-receptor complexes. *J. Biol. Chem.* **2001**, *276* (32), 30335−30341.
(9) Furuta, R. A.; Wild, C. T.; Weng, Y.; Weiss, C. D. Capture of an early fusion-active conformation of HIV-1 gp41. *Nat. Struct. Biol.* **1998**, *5* (4), 276−279.
(10) Lu, M.; Blacklow, S. C.; Kim, P. S. A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nat. Struct. Biol.* **1995**, *2* (12), 1075−1082.
(11) Weissenhorn, W.; Dessen, A.; Harrison, S. C.; Skehel, J. J.; Wiley, D. C. Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **1997**, *387* (6631), 426−430.
(12) Chan, D. C.; Fass, D.; Berger, J. M.; Kim, P. S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **1997**, *89* (2), 263−273.
(13) Kwong, P. D.; Wyatt, R.; Majeed, S.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure* **2000**, *8* (12), 1329−1339.
(14) Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **1998**, *393* (6686), 648−659.
(15) Huang, C. C.; Tang, M.; Zhang, M. Y.; Majeed, S.; Montabana, E.; Stanfield, R. L.; Dimitrov, D. S.; Korber, B.; Sodroski, J.; Wilson, I. A.; Wyatt, R.; Kwong, P. D. Structure of a V3-containing HIV-1 gp120 core. *Science* **2005**, *310* (5750), 1025−1028.
(16) Wyatt, R.; Kwong, P. D.; Desjardins, E.; Sweet, R. W.; Robinson, J.; Hendrickson, W. A.; Sodroski, J. G. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **1998**, *393* (6686), 705−711.
(17) Sattentau, Q. J.; Moore, J. P.; Vignaux, F.; Traincard, F.; Poignard, P. Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding. *J. Virol.* **1993**, *67* (12), 7383−7393.
(18) Thali, M.; Moore, J. P.; Furman, C.; Charles, M.; Ho, D. D.; Robinson, J.; Sodroski, J. Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120−CD4 binding. *J. Virol.* **1993**, *67* (7), 3978−3988.
(19) Thali, M.; Furman, C.; Ho, D. D.; Robinson, J.; Tilley, S.; Pinter, A.; Sodroski, J. Discontinuous, conserved neutralization epitopes overlapping the CD4-binding region of human immunodeficiency virus type 1 gp120 envelope glycoprotein. *J. Virol.* **1992**, *66* (9), 5635−5641.
(20) Xiang, S. H.; Kwong, P. D.; Gupta, R.; Rizzuto, C. D.; Casper, D. J.; Wyatt, R.; Wang, L.; Hendrickson, W. A.; Doyle, M. L.; Sodroski, J. Mutagenic stabilization and/or disruption of a CD4-bound state reveals distinct conformations of the human immunodeficiency virus type 1 gp120 envelope glycoprotein. *J. Virol.* **2002**, *76* (19), 9888−9899.
(21) Chen, B.; Vogan, E. M.; Gong, H.; Skehel, J. J.; Wiley, D. C.; Harrison, S. C. Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature* **2005**, *433* (7028), 834−841.
(22) Zhou, T.; Xu, L.; Dey, B.; Hessell, A. J.; Van Ryk, D.; Xiang, S. H.; Yang, X.; Zhang, M. Y.; Zwick, M. B.; Arthos, J.; Burton, D. R.; Dimitrov, D. S.; Sodroski, J.; Wyatt, R.; Nabel, G. J.; Kwong, P. D. Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **2007**, *445* (7129), 732−737.
(23) Myszka, D. G.; Sweet, R. W.; Hensley, P.; Brigham-Burke, M.; Kwong, P. D.; Hendrickson, W. A.; Wyatt, R.; Sodroski, J.; Doyle, M. L. Energetics of the HIV gp120−CD4 binding reaction. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (16), 9026−9031.
(24) Yuan, W.; Bazick, J.; Sodroski, J. Characterization of the multiple conformational states of free monomeric and trimeric human immunodeficiency virus envelope glycoproteins after fixation by cross-linker. *J. Virol.* **2006**, *80* (14), 6725−6737.
(25) Vermeire, K.; Schols, D. Anti-HIV agents targeting the interaction of gp120 with the cellular CD4 receptor. *Expert Opin. Invest. Drugs* **2005**, *14* (10), 1199−1212.
(26) Lin, P. F.; Blair, W.; Wang, T.; Spicer, T.; Guo, Q.; Zhou, N.; Gong, Y. F.; Wang, H. G.; Rose, R.; Yamanaka, G.; Robinson, B.; Li, C. B.; Fridell, R.; Deminie, C.; Demers, G.; Yang, Z.; Zadjura, L.; Meanwell, N.; Colonno, R. A small molecule HIV-1 inhibitor that targets the HIV-1 envelope and inhibits CD4 receptor binding. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (19), 11013−11018.
(27) Guo, Q.; Ho, H. T.; Dicker, I.; Fan, L.; Zhou, N.; Friborg, J.; Wang, T.; McAuliffe, B. V.; Wang, H. G.; Rose, R. E.; Fang, H.; Scarnati, H. T.; Langley, D. R.; Meanwell, N. A.; Abraham, R.; Colonno, R. J.; Lin, P. F. Biochemical and genetic characterizations of a novel human immunodeficiency virus type 1 inhibitor that blocks gp120−CD4 interactions. *J. Virol.* **2003**, *77* (19), 10528−10536.
(28) Wang, T.; Zhang, Z.; Wallace, O. B.; Deshpande, M.; Fang, H.; Yang, Z.; Zadjura, L. M.; Tweedie, D. L.; Huang, S.; Zhao, F.; Ranadive, S.; Robinson, B. S.; Gong, Y. F.; Ricardi, K.; Spicer, T. P.; Deminie, C.; Rose, R.; Wang, H. G.; Blair, W. S.; Shi, P. Y.; Lin, P. F.; Colonno, R. J.; Meanwell, N. A. Discovery of 4-benzoyl-1-[(4-methoxy-1*H*-pyrrolo[2,3-*b*]pyridin-3-yl)oxoacetyl]-2-(*R*)-methylpiperazine (BMS-378806): a novel HIV-1 attachment inhibitor that interferes with CD4-gp120 interactions. *J. Med. Chem.* **2003**, *46* (20), 4236−4239.
(29) Ho, H. T.; Fan, L.; Nowicka-Sans, B.; McAuliffe, B.; Li, C. B.; Yamanaka, G.; Zhou, N.; Fang, H.; Dicker, I.; Dalterio, R.; Gong, Y. F.; Wang, T.; Yin, Z.; Ueda, Y.; Matiskella, J.; Kadow, J.; Clapham, P.; Robinson, J.; Colonno, R.; Lin, P. F. Envelope conformational changes induced by human immunodeficiency virus type 1 attachment inhibitors prevent CD4 binding and downstream entry events. *J. Virol.* **2006**, *80* (8), 4017−4025.
(30) Zhao, Q.; Ma, L.; Jiang, S.; Lu, H.; Liu, S.; He, Y.; Strick, N.; Neamati, N.; Debnath, A. K. Identification of *N*-phenyl-*N*′-(2,2,6,6-tetramethyl-piperidin-4-yl)-oxalamides as a new class of HIV-1 entry inhibitors that prevent gp120 binding to CD4. *Virology* **2005**, *339* (2), 213−225.
(31) Si, Z.; Madani, N.; Cox, J. M.; Chruma, J. J.; Klein, J. C.; Schon, A.; Phan, N.; Wang, L.; Biorn, A. C.; Cocklin, S.; Chaiken, I.; Freire, E.; Smith, A. B., III; Sodroski, J. G. Small-molecule inhibitors of HIV-1 entry block receptor-induced conformational changes in the viral envelope glycoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (14), 5036−5041.
(32) Madani, N.; Perdigoto, A. L.; Srinivasan, K.; Cox, J. M.; Chruma, J. J.; LaLonde, J.; Head, M.; Smith, A. B., III; Sodroski, J. G. Localized changes in the gp120 envelope glycoprotein confer resistance to human immunodeficiency virus entry inhibitors BMS-806 and #155. *J. Virol.* **2004**, *78* (7), 3742−3752.
(33) Schon, A.; Madani, N.; Klein, J. C.; Hubicki, A.; Ng, D.; Yang, X.; Smith, A. B., III; Sodroski, J.; Freire, E. Thermodynamics of binding of a low-molecular-weight CD4 mimetic to HIV-1 gp120. *Biochemistry* **2006**, *45* (36), 10973−10980.
(34) Cochran, A. G. Antagonists of protein−protein interactions. *Chem. Biol.* **2000**, *7* (4), R85−R94.
(35) Huang, C. C.; Stricher, F.; Martin, L.; Decker, J. M.; Majeed, S.; Barthe, P.; Hendrickson, W. A.; Robinson, J.; Roumestand, C.; Sodroski, J.; Wyatt, R.; Shaw, G. M.; Vita, C.; Kwong, P. D. Scorpion-toxin mimics of CD4 in complex with human immunodeficiency virus gp120 crystal structures, molecular mimicry, and neutralization breadth. *Structure* **2005**, *13* (5), 755−768.
(36) Bohacek, R. S.; McMartin, C. Multiple highly diverse structures complementary to enzyme binding sites: results of extensive application of a de novo design method incorporating combinatorial growth. *J. Am. Chem. Soc.* **1994**, *116* (13), 5560−5571.
(37) Ripka, A. S.; Satyshur, K. A.; Bohacek, R. S.; Rich, D. H. Aspartic protease inhibitors designed from computer-generated templates bind as predicted. *Org. Lett.* **2001**, *3* (15), 2309−2312.
(38) Arthos, J.; Deen, K. C.; Chaikin, M. A.; Fornwald, J. A.; Sathe, G.; Sattentau, Q. J.; Clapham, P. R.; Weiss, R. A.; McDougal, J. S.; Pietropaolo, C.; Axel, R.; Trunch, A.; Maddon, P. J.; Sweet, R. W. Identification of the residues in human CD4 critical for the binding of HIV. *Cell* **1989**, *57* (3), 469−481.
(39) Smith, D. H.; Byrn, R. A.; Marsters, S. A.; Gregory, T.; Groopman, J. E.; Capon, D. J. Blocking of HIV-1 infectivity by a soluble, secreted form of the CD4 antigen. *Science* **1987**, *238* (4834), 1704−1707.
(40) Gadek, T. R. Strategies and methods in the identification of antagonists of protein−protein interactions. *BioTechniques* **2003**, Suppl., 21−24.
(41) Carr, R.; Jhoti, H. Structure-based screening of low-affinity compounds. *Drug Discovery Today* **2002**, *7* (9), 522−527.
(42) Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A. Site-directed ligand discovery. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (17), 9367−9372.
(43) Smith, A. B., III; Savinov, S. N.; Manjappara, U. V.; Chaiken, I. M. Peptide−small molecule hybrids via orthogonal deprotection−chemoselective conjugation to cysteine-anchored scaffolds. A model study. *Org. Lett.* **2002**, *4* (23), 4041−4044.
(44) Martin, L.; Stricher, F.; Misse, D.; Sironi, F.; Pugniere, M.; Barthe, P.; Prado-Gotor, R.; Freulon, I.; Magne, X.; Roumestand, C.; Menez, A.; Lusso, P.; Veas, F.; Vita, C. Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. *Nat. Biotechnol.* **2003**, *21* (1), 71−76.

(45) Allaway, G. P.; Davis-Bruno, K. L.; Beaudry, G. A.; Garcia, E. B.; Wong, E. L.; Ryder, A. M.; Hasel, K. W.; Gauduin, M. C.; Koup, R. A.; McDougal, J. S.; Maddon, P. J. Expression and characterization of CD4-IgG2, a novel heterotetramer that neutralizes primary HIV type 1 isolates. *AIDS Res. Hum. Retroviruses* **1995**, *11* (5), 533–539.

(46) Gauduin, M. C.; Allaway, G. P.; Maddon, P. J.; Barbas, C. F., 3rd; Burton, D. R.; Koup, R. A. Effective ex vivo neutralization of human immunodeficiency virus type 1 in plasma by recombinant immunoglobulin molecules. *J. Virol.* **1996**, *70* (4), 2586–2592.

(47) Trkola, A.; Pomales, A. B.; Yuan, H.; Korber, B.; Maddon, P. J.; Allaway, G. P.; Katinger, H.; Barbas, C. F., 3rd; Burton, D. R.; Ho, D. D.; et al. Cross-clade neutralization of primary isolates of human immunodeficiency virus type 1 by human monoclonal antibodies and tetrameric CD4-IgG. *J. Virol.* **1995**, *69* (11), 6609–6617.

(48) Arthos, J.; Cicala, C.; Steenbeke, T. D.; Chun, T. W.; Dela Cruz, C.; Hanback, D. B.; Khazanie, P.; Nam, D.; Schuck, P.; Selig, S. M.; Van Ryk, D.; Chaikin, M. A.; Fauci, A. S. Biochemical and biological characterization of a dodecameric CD4–Ig fusion protein: implications for therapeutic and vaccine strategies. *J. Biol. Chem.* **2002**, *277* (13), 11456–11464.

(49) Culp, J. S.; Johansen, H.; Hellmig, B.; Beck, J.; Matthews, T. J.; Delers, A.; Rosenberg, M. Regulated expression allows high level production and secretion of HIV-1 gp120 envelope glycoprotein in Drosophila Schneider cells. *Biotechnology (N. Y.)* **1991**, *9* (2), 173–177.

(50) Dey, B.; Pancera, M.; Svehla, K.; Shu, Y.; Xiang, S. H.; Vainshtein, J.; Li, Y.; Sodroski, J.; Kwong, P. D.; Mascola, J. R.; Wyatt, R. Characterization of human immunodeficiency virus type 1 monomeric and trimeric gp120 glycoproteins stabilized in the CD4-bound state: antigenicity, biophysics, and immunogenicity. *J. Virol.* **2007**, *81* (11), 5579–5593.

(51) Korber, B.; Foley, F.; Kuiken, C.; Pillai, S.; Sodroski, J. Numbering Positions in HIV Relative to HXBc2. In *Human Retroviruses and AIDS*; Korber, B., Kuiken, C. L.; Foley, B., Hahn, B., McCutchan, F., Mellors, J. W., Sodroski, J., Eds.; Los Alamos National Laboratory: Los Alamos, NM, 1998; pp III-102 through IV-103.

(52) Osburne, M. S.; Neidhardt, E. A.; Godoy, J. E.; van Schravendijk, M. R.; Grossman, T. H. Production of secreted, soluble human two-domain CD4 protein in *Escherichia coli*. *J. Immunol. Methods* **1999**, *224* (1–2), 19–24.

(53) Coligan, J. E.; Dunn, B. M.; Ploegh, H. L.; Speicher, D. W.; Wingfield, P. T. *Current Protocols in Protein Science*; John Wiley & Sons: 2001.

JM070564E

Drugs of Today 2004, 40 (11): 949-955
Copyright © 2004 PROUS SCIENCE
CCC: 0025-7656/2004

HIV VACCINE CANDIDATES

Tai Te Wu[1] and George Johnson[2]

[1]Department of Biochemistry, Molecular Biology and Cell Biology, and
[2]Department of Biomedical Engineering, Northwestern University, Evanston, Illinois, USA

CONTENTS

Summary 949
Introduction 949
Nearly invariant segments of HIV-1 surface protein gp120 951
Locations in three-dimensional structure of HIV-1 surface protein gp120 951
Possible HIV vaccine candidates 951
Time course of useful antibodies 952
Discussion 953
References 954

Summary

Amino acid sequences of the HIV-1 surface protein gp120 show extensive variations. Some of the peptide or DNA vaccines using the V3-loop of gp120 can trigger antibody production against some isolates of HIV, but these antibodies are not protective against other isolates. Thus, more conserved peptides from other HIV proteins have also been tried as vaccines. However, antibodies specific for these peptides cannot prevent the normal entry of HIV into the host cells, followed by integration of viral genomes into host chromosomes. Therefore, we decided to make a careful analysis of the aligned amino acid sequences of gp120 stored at Los Alamos National Laboratory. Unexpectedly, we found that there are several, not just one or two, segments of seven or more consecutive and nearly invariant amino acid residues present in the C1, C2 and C5 regions of this protein. Furthermore, four of these segments are physically close to each other in the known three-dimensional structure of the core portion of gp120. The remaining segments are not in that crystallized region; however, two of these are covalently linked with a disulfide bond. Therefore, we propose that these segments may be combined to form the basis of possible HIV vaccine candidates. In addition, it is necessary to incorporate the processed peptide from gp120 or other proteins and to understand the time-dependent effects of the highly variable loops of gp120 on triggering antibody production in human patients. © 2004 Prous Science. All rights reserved.

Correspondence: Tai Te Wu, Room E348, Technological Institute, 2145 Sheridan Road, Evanston, IL 60208, USA.
Tel.: (847) 491-7849; FAX: (847) 491-4928.
E-mail: t-wu@northwestern.edu

Introduction

In 1981, three rare diseases, Kaposi's sarcoma, *Pneumocystis carnii* pneumonia and mucosal can-

didiasis, were noted in previously healthy homosexual males (1, 2). Acquired immunodeficiency syndrome (AIDS) was identified as a clinical entity the following year (3, 4). While type C virus particles were discovered in human T-cell leukemia in 1981 (5, 6), a different type of virus was isolated from patients with AIDS in 1983 (7, 8). These two types of viruses are pathologically different: human immunodeficiency virus (HIV) is cytotoxic, whereas human T-cell leukemia virus (HTLV) is capable of transforming cells (9). HIV is a retrovirus containing two usually identical positive strands of RNA molecules coding for fifteen proteins (10). During the past two decades, numerous variants of HIV have been isolated from all over the world. HIV-1 is more virulent, while HIV-2 is more closely related to simian immunodeficiency virus (SIV). Based on their amino acid sequences, they are classified into different clades. Thousands of complete and partial nucleotide and amino acid sequences have been stored in the GenBank (http://www.ncbi.nlm.nih.gov), while several hundred nearly complete amino acid sequences are aligned at the Los Alamos National Laboratory (http://www.lanl.gov).

In a careful longitudinal study of a patient infected with HIV-1 for about 10 years, several different amino acid sequences were initially identified (11). As the disease progressed, these sequences kept on changing. While there was no clear-cut pattern of amino acid substitutions as a function of time, some of the amino acid sequences in the V3-loop region appeared to be similar to those present in several of the third complementarity determining regions of human antibody heavy chains (12). This process of humanization of the V3-loop might be a useful mechanism for HIV to avoid immune surveillance by the host.

Since the first step of HIV infection involves the binding of the gp120 surface protein with host cell surface proteins CD4 and CCR5 (or CXCR4), an efficient HIV vaccine candidate should presumably block this binding. The protective antibodies thus induced should bind gp120 of all different clades of HIV very tightly and promote the engulfing and digestion of the entire viral particle by host cells. Hopefully, this process of internalization of HIV would not mimic the normal process of fusion and infection (9). Thus, we have to analyze the amino acid sequences of gp120 carefully to look for nearly invariant segments of sufficient lengths. These segments, if identified, should be on the surface of the three-dimensional structure of gp120 so that host antibodies can interact with them (13).

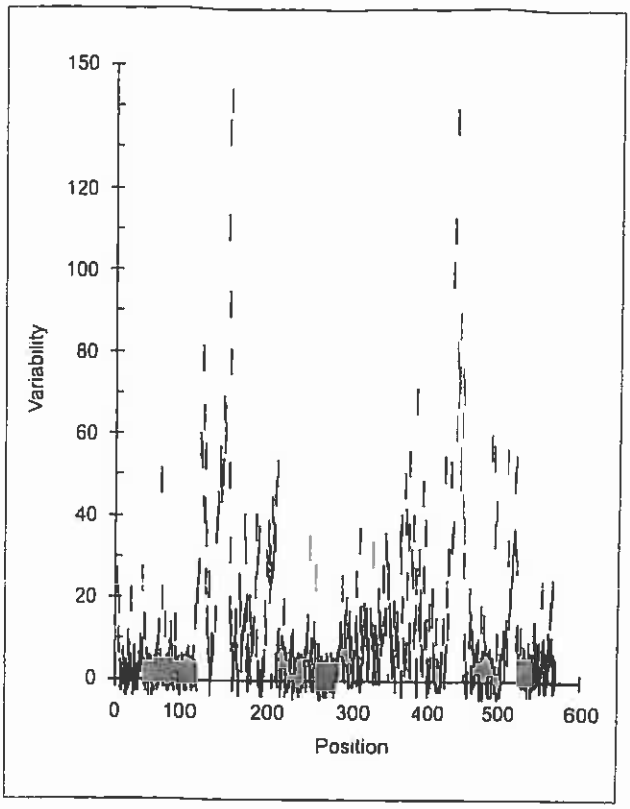


Fig. 1. Variability plot for 383 surface protein gp120 sequences.

Researchers at Los Alamos National Laboratory have aligned most of the nearly complete amino acid sequences of gp120. The total number of these sequences has been increasing steadily over the years. However, our analysis is relatively insensitive to this number as long as it is over 200 and as long as the collection consists of sequences from all clades. We have used 383 amino acid sequences for our present sequence analysis, and they are numbered from position 1 to 570. The invariant Cys residues are aligned with disulfide bonds between positions 25 and 45, 90 and 218, 97 and 208, 102 and 147, 232 and 263, 242 and 253, 314 and 358, 411 and 490, and 418 and 460. Based on the variability plot (14) (Fig. 1), surface protein gp120 can be divided into five constant and five variable regions: C1 (1 to 102), V1 (103 to 147), V2 (148 to 207), C2 (208 to 314), V3 (315 to 357), C3 (358 to 418), V4 (419 to 459), C4 (460 to 501), V5 (502 to 529) and C5 (530 to 570).

Even though there are extensive amino acid residue substitutions in various gp120 sequences, some segments remain relatively constant. Indeed, the presumed processed peptide from gp120 binding major histocompatibility complex (MHC) class II

molecules is located in the C4 region and is usually assumed to be either of the following: KQIINM–WQE–VGKAMYA or KQIINM–WQE–VGKQMYAP PIR–GQIR. Therefore, we decided to look for segments of seven or more consecutive, nearly invariant amino acid residues in gp120. Nearly invariant residues are defined as residues at any given position among all 383 sequences occurring more frequently than 95% of the time (14). Seven amino acid residues in a stretched out configuration can roughly span the antibody-combining site.

Nearly invariant segments of HIV-1 surface protein gp120

Aligned amino acid sequences of 383 HIV-1 surface proteins from Los Alamos National Laboratory were placed into a 383 x 570 table, with each sequence listed in a row. Columns were used to identify the amino acid positions from the N-terminal to the C-terminal. Computer programs were then written to count the number of different amino acid residues at each position, and to identify the most common one. If that residue occurred more than 95% of the time, it was considered as nearly invariant. If there were seven consecutive nearly invariant residues, that peptide was identified as a nearly invariant segment. By this simple logical approach, we discovered three such segments in the C1 region, three in the C2 region and one in the C5 region (Table I).

Segments #II and #III are covalently linked by a disulfide bond, and so are segments #IV and #V. Segment #VI is only one amino acid residue, being mostly K or R, away from #V.

This analysis did not locate the presumed gp120 processed peptide in the C4 region, suggesting that the processed peptide might be somewhat more varied than the segments listed above. In fact, the presence of these nearly invariant segments seems contradictory to the general belief that HIV-1 surface proteins are more varied than the amino acid sequences of influenza A virus hemagglutinins, where no such nearly invariant segments are present.

Table I. Nearly invariant segments of HIV-1 surface protein gp120.

#	Positions	Sequence	Region
I	4 to 14	WVTVYYGVPVW	C1
II	23 to 29	LFCASDA	C1
III	44 to 50	ACVPTDP	C1
IV	226 to 232	PIPIHYC	C2
V	261 to 267	VQCTHGI	C2
VI	269 to 282	PVVSTQLLL–NGSL	C2
VII	536 to 543	ELYKYKVV	C5

Locations in three-dimensional structure of HIV-1 surface protein gp120

Part of gp120 has been crystallized and its three-dimensional structure determined by x-ray diffraction studies (13). However, that core structure consists of only about 65% of the entire protein, lacking about 60 amino acid residues from the N-terminal and about 30 amino acid residues from the C-terminal and the V1-, V2- and V3-loops. Furthermore, the V4-loop, although present in the crystal, is not visible, probably due to thermal fluctuations (15). As a result, the nearly invariant segments #I, #II and #III listed in Table I are not included in the known three-dimensional structure.

Due to the presence of a disulfide bond between segments #IV and #V, and to the separation of only one amino acid residue between segments #V and #VI, these three segments are physically close to each other. Interestingly, even though segment #VII is sequentially far away, it is physically close to segments #IV, #V and #VI in the three-dimensional structure. In fact, segments #IV, #V, #IV and #VII form one of the bridges between the inner and outer domains of gp120, as well as the back wall of the CD4 binding site (13).

Possible HIV vaccine candidates

Various combinations of the above seven nearly invariant segments in HIV-1 surface protein gp120 can thus suggest several possible HIV vaccine candidates. A few possibilities are presented in Figure 2, and many others can be designed. In addition, the presumed processed peptide of gp120 could be covalently linked to the N-terminal of these candidates.

The lengths of the linkers in the constructs presented in Figure 2 can be adjusted to mimic the native configuration of these peptides in the gp120 three-dimensional structure (13). The native configurations of constructs (A) and (B) have not yet been determined experimentally. Further restriction on these two constructs can be imposed when they become known.

Determining whether processed peptides from proteins other than gp120 could be more efficient in triggering antibody production requires detailed experimentation. In any case, these processed peptides could also be covalently linked to the N-terminals of the above HIV vaccine candidates.

A) Segments #II and #III:

LFCASDA
|
ACVPTDP

B) Segments #I, #II and #III:

WVTVYYGVPVW – (linker) – LFCASDA
|
ACVPTDP

C) Segments #IV and #V:

PIPIHYC
|
VQCTHGI

D) Segments #IV and #V with additional covalent bond:

FEPIPIHYC
|
VQCTHGIK

E) Segments #IV, #V and #VI:

PIPIHYC
|
VQCTHGIKPVVSTQLLL–NGSL

F) Segments #IV, #V and #VI with additional covalent bond:

FEPIPIHYC
|
VQCTHGIKPVVDSTQLLL–NGSL

G) Segments #IV, #V, #VI and #VII:

PIPIHYC
|
VQCTHGIKPVVSTQLLL–NGSL–(linker)–ELYKYKVV

H) Segments #IV, #V, #VI and #VII with additional linker:

PIPIHYC
|
VQCTHGIKPVVSTQLLL–NGSL–(linker)–ELYKYKVV
|
(linker)

I) Segments #IV, #V, #VI and #VII with additional covalent bond and linker:

FEPIPIHYC
|
VQCTHGIKPVVSTQLLL–NGSL–(linker)–ELYKYKVV
|
(linker)

Fig. 2. Suggested combinations of nearly invariant segments of HIV-1 surface protein gp120 as possible HIV vaccine candidates.

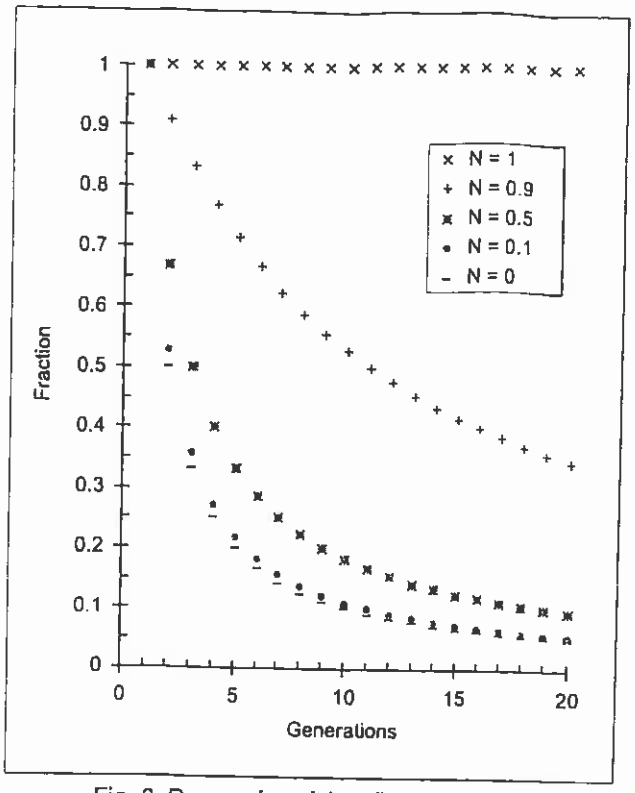


Fig. 3. Decay of useful antibody fraction.

Time course of useful antibodies

One of the major difficulties in designing HIV vaccine candidates is that the amino acid sequence of gp120 changes continuously, thus triggering the production of different antibodies by the patient as the disease progresses. The five variable loops, V1 to V5, may even be more efficient as antigenic epitopes to stimulate the patient's immune system than any of the designed vaccine candidates. In order to understand the time course of the fraction of useful antibodies present in the patient, the following simple model should be considered.

Let N be the percentage of invariant epitopes of gp120. Then 1 – N is the percentage of varying antigenic epitopes. For simplicity, let G be the generation of mutations occurring simultaneously on the varying antigenic epitopes. Then, at the beginning of HIV-1 infection, the patient can produce antibodies against all epitopes, and they are all useful. Therefore, the infection subsides. However, the viral genome integrates into the host chromosome. After one generation of mutations on all the varying epitopes, the old antibodies against the initial varying epitopes become useless. The fraction of useful antibodies, F, can be calculated as a function of the generation of mutations, G, as follows:

$$F = 1 / (G \times (1 - N) + N)$$

271

In Figure 3, this fraction is plotted against generations for various values of N. It is clearly indicated that for N = 1, the fraction of useful antibodies will remain at 1, or 100%. However, even for N = 0.9, that fraction decays rapidly. In the actual situation of HIV-1 surface protein as discussed above, it is probably the case that N = 0.1 or less. After about five generations of mutations, the fraction of useful antibodies is reduced to about 20%, not much different from N = 0, even though the total amount of such antibodies may be increased to a certain extent.

Based on this simple analytical model, any useful HIV vaccine candidate must therefore stimulate the production of a sufficient amount of protective antibodies in the individual so that all invading HIV particles are eliminated initially.

Discussion

Currently, there are three groups of drugs that are very effective in the treatment of AIDS: reverse transcriptase inhibitors (16, 17), protease inhibitors (18) and fusion inhibitors (19). Reverse transcriptase inhibitors were first developed soon after the detailed life cycle of HIV became known. Its combination with protease inhibitors resulted in the massive elimination of HIV from patients' blood. However, host cells with integrated viral genome could not be eliminated. The relatively recent development of fusion inhibitors would hopefully prevent the initiation of HIV infection by preventing HIV from fusing to host cells. The down side of all these drugs designed to treat AIDS is that they are very expensive, especially for patients in developing countries where HIV infection is epidemic, and they require long-term maintenance administration.

For this reason, various protein, peptide and DNA vaccines have been tried to combat HIV infection. The first proposed peptide vaccine candidate consisted of a few amino acid residues, LERLLL, at the C-terminal end of gp41 based on the observation of some sequence homology of that region to interleukin 2 (20). However, as more sequences became available, substantial amino acid variations were present in that region. Furthermore, that segment was not exposed on the surface of HIV.

Based on the finding that a 15-amino-acid residue segment in two of the 383 sequences of the V3-loops of gp120 analyzed in the present study, RIQRGPGRAFVTIGK, might be recognized by CD8+ cytotoxic lymphocytes from mouse (21), several research groups decided that amino acid sequences from the V3-loop region of gp120 could be good vaccine candidates (22, 23). Sometimes, these peptides were covalently linked to the presumed processed peptide discussed before, giving rise to the so-called C4-V3 construct (24). However, by definition, the V3 region of gp120 is highly varied, with several thousand different amino acid sequences from different isolates of HIV. It was rather uncertain as to which V3 amino acid sequence or sequences should be used for vaccination.

Other peptide vaccine candidates from HIV-1 surface protein consisted of various combinations of the five V-loops (25), the entire surface protein, a combination of gp120 and gp41, the tetramer of ELDKWA in the HR2 of gp41 (26), etc. Some of the peptide loop structures have been converted to DNA sequences for DNA vaccination.

Due to the extensive amino acid substitutions in various HIV-1 surface proteins, others have tried relatively conserved peptides from p17 (27) and from p24 (28). However, these proteins are not on the surface of HIV. Various combinations of Env and Gag polyprotein segments have also been investigated (29). Unfortunately, in general, the results of such HIV vaccine candidates have not been very encouraging, primarily due to sequence variations. However, effective vaccines are probably the only hope for halting the continuous spread of this pandemic infection to more communities around the entire world.

On the other hand, the combined effort of many research groups during the past two decades has provided valuable information on the nature of HIV and AIDS, especially in relation to the massive amounts of sequence determination of the virus isolated from all over the world. In the present analysis, aligned amino acid sequences prepared by the scientists at Los Alamos National Laboratory illustrate to us that we may still identify HIV vaccine candidates constructed from nearly invariant peptides of the major surface protein gp120. These observations, together with the conformational architecture deduced from the crucial three-dimensional structure of the core portion of gp120 (13), have enabled various constructs to be proposed.

Effective HIV vaccine candidates could hopefully arise from the following factors: 1) nearly invariant segments of seven or more amino acid residues in gp120; 2) several such segments in close physical proximity; 3) location on the surface of gp120; 4) identification of the processed peptide; 5) triggering of the effective production of large amounts of protective antibodies; 6) prevention of

the normal fusion of HIV to host cells to initiate infection; and 7) elimination of the integration of the HIV genome into the host chromosome. However, these characteristics are theoretical at present and must be verified experimentally. To be clinically successful, protective antibodies must be produced in effective amounts to get rid of all HIV particles initially introduced into any individual.

References

1. Hymes, K.B., Cheung, T., Greene, J.B. et al. *Kaposi's sarcoma in homosexual men – A report of eight cases.* Lancet 1981, 2 (8247): 598-600.
2. Gottlieb, M.S., Schroff, R., Schanker, H.M. et al. *Pneumocystis carnii pneumonia and mucosal candidiasis in previously healthy homosexual men: Evidence of a new acquired cellular immunodeficiency.* New Eng J Med 1981, 305: 1425-31.
3. Zakowski, P., Fligiel, S., Berlin, G.W., Johnson, L. Jr. *Disseminated Mycobacterium avium-intracellulare infection in homosexual men dying of acquired immunodeficiency.* JAMA 1982, 248: 2980-2.
4. Quagliarello, V. *The acquired immunodeficiency syndrome: Current status.* Yale J Biol Med 1982, 55: 443-52.
5. Hinuma, Y., Nagata, K., Hanaoka, M. et al. *Adult T-cell leukemia: Antigen in an ATL cell line and detection of antibodies to the antigen of human sera.* Proc Natl Acad Sci USA 1981, 78: 6476-80.
6. Miyoshi, I., Kubonishi, I., Yoshimoto, S. et al. *Type C virus particle in a cord T-cell line derived by co-cultivating normal human cord leukocytes and human leukemic T cell.* Nature 1981, 284: 770-1.
7. Chermann, J.C., Barre-Sinoussi, F., Dauguet, C. et al. *Isolation of a new retrovirus in a patient at risk for acquired immunodeficiency syndrome.* Antibiot Chemother 1983, 32: 48-53.
8. Barre-Sinoussi, F., Chermann, J.C., Rey, F. et al. *Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immunodeficiency syndrome (AIDS).* Science 1983, 220: 868-71.
9. Levy, J.A. HIV and Pathogenesis of AIDS, 2nd ed. ASM Press: Washington 1998.
10. Frankel, A.D., Young, J.A.T. *HIV-1: Fifteen proteins and an RNA.* Ann Rev Biochem 1998, 67: 1-25.
11. Janssens, W., Nkengasong, J., Heyndrickx, L. et al. *Intrapatient variability of HIV type I group O ANT70 during a 10-year follow-up.* AIDS Res Hum Retroviruses 1999, 15: 1325-32.
12. Johnson, G., Wu, T.T. *The Kabat database and a bioinformatics example.* In: Antibody Engineering: Methods and Protocols. B.K.C. Lo (Ed.). Humana Press: Totowa 2003, Chapter 2.
13. Wyatt, R., Kwong, P.D., Desjardina, E. et al. *The antigenic structure of the HIV gp120 envelope glycoprotein.* Nature 1998, 393: 705-11.
14. Wu, T.T., Kabat, E.A. *An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity.* J Exp Med 1970, 132: 211-50.
15. Yamaguchi-Kabata, Y., Gojobori, T. *Reevaluation of amino acid variability of the human immunodeficiency virus type 1 gp120 envelope glycoprotein and prediction of their discontinuous epitopes.* J Virol 2000, 74: 4335-50.
16. Bhana, N., Ormrod, D., Perry, C.M., Figgitt, D.P. *Zidovudine: A review of its use in the management of vertically-acquired pediatric HIV infection.* Paediatr Drugs 2002, 4: 515-53.
17. Corbett, J.W., Rodgers, J.D. *Discovery of second generation quinazolinone non-nucleotide reverse transcriptase inhibitors of HIV-1.* Prog Med Chem 2002, 40: 63-105.
18. Smith, A.M. III, Cantin, L.D., Pasternak, A. et al. *Design, synthesis and biological evaluation of monopyrrolinone-based HIV-1 protease inhibitors.* J Med Chem 2003, 46: 1831-1844.
19. Kilby, J.M., Lalezari, J.P., Eron, J.J. et al. *The safety, plasma pharmacokinetics, and antiviral activity of subcutaneous enfuvirtide (T-20), a peptide inhibitor of gp41-mediated virus fusion, in HIV-infected adults.* AIDS Res Hum Retroviruses 2002, 18: 685-93.
20. Reiher, W.E. III, Blalock, J.E., Brunck, T.K. *Sequence homology between acquired immunodeficiency syndrome virus envelope protein and interleukin 2.* Proc Natl Acad Sci USA 1986, 83: 9188-92.
21. Takahashi, H., Cohen, J., Hosmalin, A. et al. *An immunodominant epitope of the HIV gp160 envelope glycoprotein recognized by class I MHC molecule-restricted murine cytotoxic T lymphocytes.* Proc Natl Acad Sci USA 1988, 85: 3105-9.
22. Defoort, J.-P., Nardelli, B., Huang, W., Ho, D.D., Tam, J.P. *Macromolecular assemblage in the design of a synthetic AIDS vaccine.* Proc Natl Acad Sci USA 1992, 89: 3879-83.
23. Kozlowski, S., Corr, M., Takeshita, T. et al. *Serum angiotensin-1 converting enzyme activi-*

ty processes a human immunodeficiency virus I gp160 peptide for presentation by major histocomapatibility complex class I molecules. J Exp Med 1992, 175: 1417-22.

24. Ahlers, J.D., Takeshita, T., Pendleton, C.D., Berzofzky, J.A. *Enhanced immunogenicity of HIV-1 vaccine construct by modification of the native peptide sequence.* Proc Natl Acad Sci USA 1997, 94: 10856-61.
25. Carlos, M.P., Anderson, D.E., Gardner, M.B., Torres, J.V. *Immunogenicity of a vaccine preparation representing the variable regions of the HIV type I envelope glycoprotein.* AIDS Res Hum Retroviruses 2000, 16: 153-61.
26. Xiao, Y., Zhao, Y., Lu, Y., Chen, Y.H. *Epitope-vaccine induces high levels of ELDKWA-epitope-specific neutralizing antibodies.* Immunol Invest 2000, 29: 41-50.
27. Naylor, P.H., Naylor, C.W., Badamchian, M. et al. *Human immunodeficiency virus contains an epitope immunoreactive with thymosin alpha 1 and the 30-amino acid synthetic p17 group-specific antigen peptide HGP-30.* Proc Natl Acad Sci USA 1987, 84: 2951-5.
28. Nishino, Y., Kameoka, M., Okada, Y. et al. *In vivo induction of human immunodeficiency virus type 1-specific cytotoxic T lymphocytes and delay-type hypersensitivity by a 23-amino acid peptide from the highly conserved region of major core protein p24.* Vaccine 1994, 12: 485-91.
29. Okuda, K., Kaneko, T., Yamakawa, T. et al. *Strong immunogenicity of a multicomponent peptide vaccine developed with the branched lysine oligopeptide method for human immunodeficiency virus infection.* J Mol Recognit 1993, 6: 101-9.

The surface protein SU or gp120 of HIV mutates rapidly. In one study, the infected individual had more than ten different gp120 amino acid sequences in his blood. Furthermore, as time went one, these sequences kept on changing. That individual could produce antibodies against these different gp120 proteins, but earlier antibodies became totally useless for later HIV. It is for this reason that so far vaccination against HIV infection has not been successful.

This situation is similar to influenza viruses. Their hemagglutinin molecule also mutates rapidly, but not as extensively as gp120 of HIV. Therefore, yearly vaccination against prevalent strains of influenza viruses has been reasonably successful.

Interestingly, on aligning the available complete amino acid sequences of gp120 as done by the scientists at Los Alamos National Laboratory, one can identify stretches of nearly invariant segments for seven or more consecutive residues. Four of the seven such segments are in the portion of gp120 with known three-dimensional structure. These four segments are physically close to each other. Indeed they form the back wall of the CD4 binding site. For detailed analysis, please see Lessons # 1.1, 1.2, 1.3, 1nd 1.4 at the following website:

http://www.northwestern.edu/bme/bme351

In principle, it will be possible to genetically engineer an artificial protein consisting of these four segments and use it for vaccination against HIV infection.

While active immunization has been very successful during the past couple of centuries, we will undoubtedly encounter new diseases and new methods where vaccination will play viral roles. The pioneering work of Coley using bacterial extracts to treat cancer patients by stimulating their immune systems took place over a century ago. More recent studies of his method may involve the interleukins, various CD molecules, cytokines, etc. To a certain extent, we are still puzzled by its therapeutic value. Eventually, the detailed understanding of various macro-molecules and cells involved in immunochemistry may provide us with an answer. Vaccinations against cancers, new and old infections, autoimmune diseases, etc. require careful analysis. To conclude this discussion, a simple mathematical model of active immunization will be presented in the last section.

Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody VRC01

Tongqing Zhou,[1] Ivelin Georgiev,[1*] Xueling Wu,[1*] Zhi-Yong Yang,[1*] Kaifan Dai,[1] Andrés Finzi,[2] Young Do Kwon,[1] Johannes F. Scheid,[3] Wei Shi,[1] Ling Xu,[1] Yongping Yang,[1] Jiang Zhu,[1] Michel C. Nussenzweig,[3] Joseph Sodroski,[2,4] Lawrence Shapiro,[1,5] Gary J. Nabel,[1] John R. Mascola,[1] Peter D. Kwong[1]†

During HIV-1 infection, antibodies are generated against the region of the viral gp120 envelope glycoprotein that binds CD4, the primary receptor for HIV-1. Among these antibodies, VRC01 achieves broad neutralization of diverse viral strains. We determined the crystal structure of VRC01 in complex with a human immunodeficiency virus HIV-1 gp120 core. VRC01 partially mimics CD4 interaction with gp120. A shift from the CD4-defined orientation, however, focuses VRC01 onto the vulnerable site of initial CD4 attachment, allowing it to overcome the glycan and conformational masking that diminishes the neutralization potency of most CD4-binding-site antibodies. To achieve this recognition, VRC01 contacts gp120 mainly through immunoglobulin V-gene regions substantially altered from their genomic precursors. Partial receptor mimicry and extensive affinity maturation thus facilitate neutralization of HIV-1 by natural human antibodies.

Successful vaccine development often takes advantage of clues from humoral responses elicited by natural infection. For HIV-1, neutralizing antibody responses elicited within the first year or two of infection are generally strain-specific (*1*) and thus provide limited insight into vaccine design (*2*). A few monoclonal antibodies from HIV-1–infected individuals, however, are broadly neutralizing, and an effort has been made to facilitate vaccine design by defining their structures (*3*, *4*).

The well-studied broadly neutralizing antibodies to HIV-1—2G12, 2F5, 4E10, and b12—have unusual characteristics that have posed barriers to eliciting similar antibodies in humans (*5*). Thus, in addition to having broad capacity for neutralization, an appropriate antibody should be present in high enough titers in humans to suggest that it can be elicited in useful concentrations. We and others have screened cohorts of sera from infected individuals to find broadly neutralizing responses that are detectable in a substantial percentage of subjects (*6*–*10*). One serum response that satisfies these criteria has been mapped to the site on the HIV-1 gp120 envelope (Env) glycoprotein that binds to the CD4 receptor (*8*).

Although potentially accessible, the CD4-binding site is protected from humoral recognition through glycan and conformational masking (*11*). The identification of monoclonal antibodies against this site is described in a companion manuscript (*12*). In brief, we created resurfaced, conformationally stabilized probes, with antigenic specificity for the initial site of CD4 attachment to gp120 (*13*). This site, a conformationally invariant subset of the CD4-binding surface, is vulnerable to antibody-mediated neutralization (*13*), and we used probes specific for this site to identify antibodies that neutralize most viruses (*12*). In this work, we analyzed the crystal structure for one of these antibodies, VRC01, in complex with

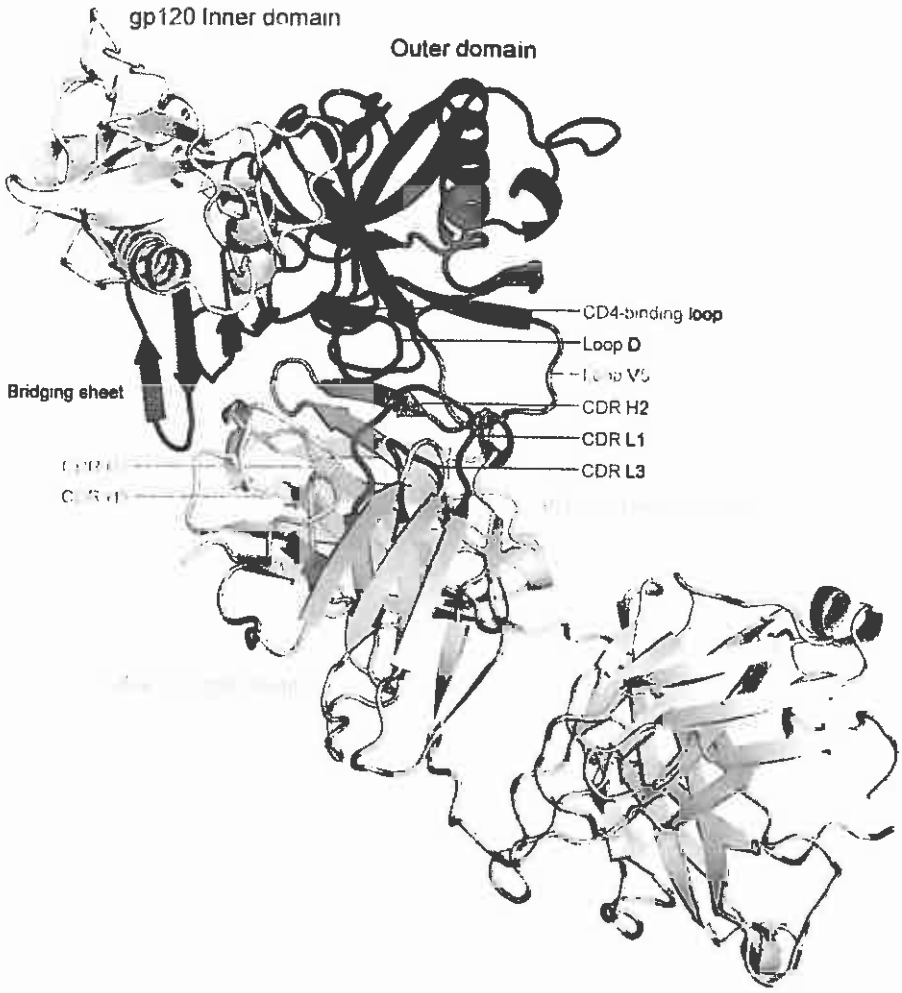


Fig. 1. Structure of antibody VRC01 in complex with HIV-1 gp120. Atomic-level details for broad and potent recognition of HIV-1 by a natural human antibody are depicted with polypeptide chains in ribbon representations. The gp120 inner domain is shown in gray, the bridging sheet in blue, and the outer domain in red, except for the CD4-binding loop (purple), the D loop (brown), and the V5 loop (orange). The light chain of the antigen-binding fragment (Fab) of VRC01 is shown in light blue, with complementarity-determining regions (CDRs) highlighted in dark blue (CDR L1) and marine blue (CDR L3). The heavy chain of Fab VRC01 is shown in light green, with CDRs highlighted in cyan (CDR H1), green (CDR H2), and pale yellow (CDR H3). Both light and heavy chains of VRC01 interact with gp120: The primary interaction surface is provided by the CDR H2, with the CDR L1 and L3 and the CDR H1 and H3 providing additional contacts.

Downloaded from www.sciencemag.org on December 23, 2011

an HIV-1 gp120 core. We deciphered the basis of VRC01 neutralization, identified mechanisms of natural resistance, showed how VRC01 minimizes such resistance, examined potential barriers to elicitation, and defined the role of affinity maturation in gp120 recognition.

Similarities of Env recognition by CD4 and VRC01 antibody. To gain a structural understanding of VRC01 neutralization, we crystallized the antigen-binding fragment (Fab) of VRC01 in complex with an HIV-1 gp120 from the clade A/E recombinant 93TH057 (*14*, *15*). The crystallized gp120 consisted of its inner domain–outer domain core, with truncations in the variable loops V1/V2 and V3 as well as the N- and C-termini, which are regions known to extend away from the main body of the gp120 envelope glycoprotein (*16*). Diffraction to 2.9 Å resolution was obtained from orthorhombic crystals, which contained four copies of the VRC01-gp120 complex per asymmetric unit, and the structure was solved by means of molecular replacement and refined to a crystallographic R value of 19.7% (Fig. 1 and table S1) (*17*).

The interaction surface between VRC01 and gp120 encompasses almost 2500 Å^2, with 1244 Å^2 contributed by VRC01 and 1249 Å^2 by gp120 (*18*). On VRC01, both heavy chain (894 Å^2) and light chain (351 Å^2) contribute to the contact surface (table S2), with the central focus of binding on the heavy chain–second complementarity–determining region (CDR H2). Over half of the interaction surface of VRC01 (644 Å^2) involves CDR H2, a mode of binding that is reminiscent of the interaction between gp120 and the CD4 receptor; CD4 is a member of the V-domain class of the immunoglobulin superfamily (*19*), and the CDR2-like region of CD4 is a central focus of gp120 binding (Fig. 2A and table S3) (*20*). For CD4, the CDR2-like region forms antiparallel, intermolecular hydrogen bonds with residues 365_{gp120} to 368_{gp120} of the CD4-binding loop of gp120 (*20*) (Fig. 2B); with VRC01, one hydrogen-bond is observed between the carbonyl oxygen of $Gly54_{VRC01}$ and the backbone nitrogen of $Asp368_{gp120}$. This hydrogen bond occurs at the loop tip, an extra residue relative to CD4 is inserted in the strand, and the rest of the potential hydrogen bonds are of poor geometry or distance (Fig. 2C and table S4). Other similarities and differences with CD4 are found: Of the two dominant CD4 residues ($Phe43_{CD4}$ and $Arg59_{CD4}$) involved in interaction with gp120, VRC01 mimics the arginine interaction but not the phenylalanine one (Fig. 2, B and C). Lastly, substantial correlation was observed between gp120 residues involved in binding VRC01 and CD4 (fig. S1).

Superposition of the gp120 core in its VRC01-bound form with gp120s in other crystalline lattices and bound by other ligands indicates a CD4-bound conformation [Protein Data Bank (PDB) ID number 3JWD] (*16*) to be most closely related in structure, with a Cα-root-mean-square deviation of 1.03 Å (table S5). Such superposition of gp120s from CD4- and VRC01-bound conformations brings the N-terminal domain of CD4 and the heavy chain–variable domain of VRC01 into close alignment (Fig. 2), with 73% of the CD4 N-terminal domain volume overlapping with VRC01 (*21*). This domain overlap is much higher than observed with the heavy chains of other CD4-binding-site antibodies, such as b12, b13, or F105 (table S6). However, when the VRC01 heavy chain is superimposed—on the basis of conserved framework and cysteine residues—on CD4 in the CD4-gp120 complex,

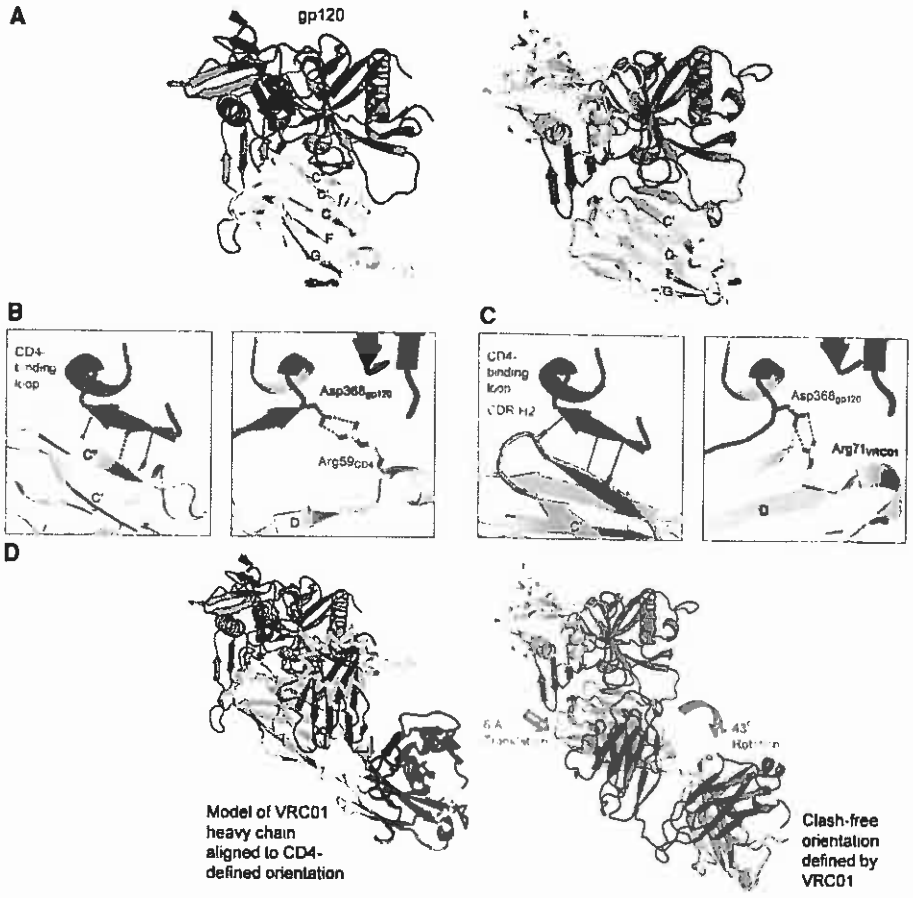


Fig. 2. Structural mimicry of CD4 interaction by antibody VRC01. VRC01 shows how a double-headed antibody can mimic the interactions with HIV-1 gp120 of a single-headed member of the immunoglobulin superfamily, such as CD4. (A) Comparison of HIV-1 gp120 binding to CD4 (N-terminal domain) and VRC01 (heavy chain–variable domain). Polypeptide chains are depicted in ribbon representation for (right) the VRC01 complex and (left) the CD4 complex with the lowest gp120 root mean square deviation (table S5). The CD4 complex (3JWD) (*16*) is indicated in yellow for CD4 and red for gp120, except for the CDR-binding loop (purple). The VRC01 complex is colored as in Fig. 1. Immunoglobulin domains are composed of two β sheets, and the top sheet of both ligands is labeled with the standard immunoglobulin-strand topology (strands G, F, C, C′, C″). (B and C) Interface details for (B) CD4 and (C) VRC01. Close-ups are shown of critical interactions between the CD4-binding loop (purple) and the C″ strand as well as between $Asp368_{gp120}$ and either $Arg59_{CD4}$ or $Arg71_{VRC01}$. Hydrogen bonds with good geometry are depicted by blue dotted lines, and those with poor geometry are depicted in gray. Atoms from which hydrogen bonds extend are depicted in stick representation and colored blue for nitrogen and red for oxygen. In the left panel of (C), the β15-strand of gp120 is depicted to aid comparison with (B), although because of the poor hydrogen-bond geometry, it is only a loop. (D) Comparison of VRC01- and CD4-binding orientations. Polypeptides are shown in ribbon representation, with gp120 colored the same as in (A) and VRC01 depicted with heavy chain in dark yellow and light chain in dark gray. When the heavy chain of VRC01 is superimposed onto CD4 in the CD4-gp120 complex, the position assumed by the light chain evinces numerous clashes with gp120 (left). The VRC01-binding orientation (right) avoids clashes by adopting an orientation rotated by 43° and translated by 6 Å.

[1]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH), Bethesda, MD 20892, USA. [2]Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Department of Pathology, Division of AIDS, Harvard Medical School, Boston, MA 02115, USA. [3]Laboratory of Molecular Immunology and Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065, USA. [4]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115, USA. [5]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: pdkwong@nih.gov

Downloaded from www.sciencemag.org on December 23, 2011

277

clashes are found between gp120 and the entire top third of the VRC01 variable light chain (Fig. 2D) (*22*). In its complex with gp120, VRC01 rotates 43° relative to the CD4-defined orientation and translates 6 Å away from the bridging sheet, to a clash-free orientation that mimics many of the interactions of CD4 with gp120, although with considerable variation. Analysis of electrostatics shows that the interaction surfaces of VRC01 and CD4 are both quite basic, although the residue types of contacting amino acids are distinct (fig. S2). Thus, although VRC01 mimics CD4 binding to some extent, considerable differences are observed.

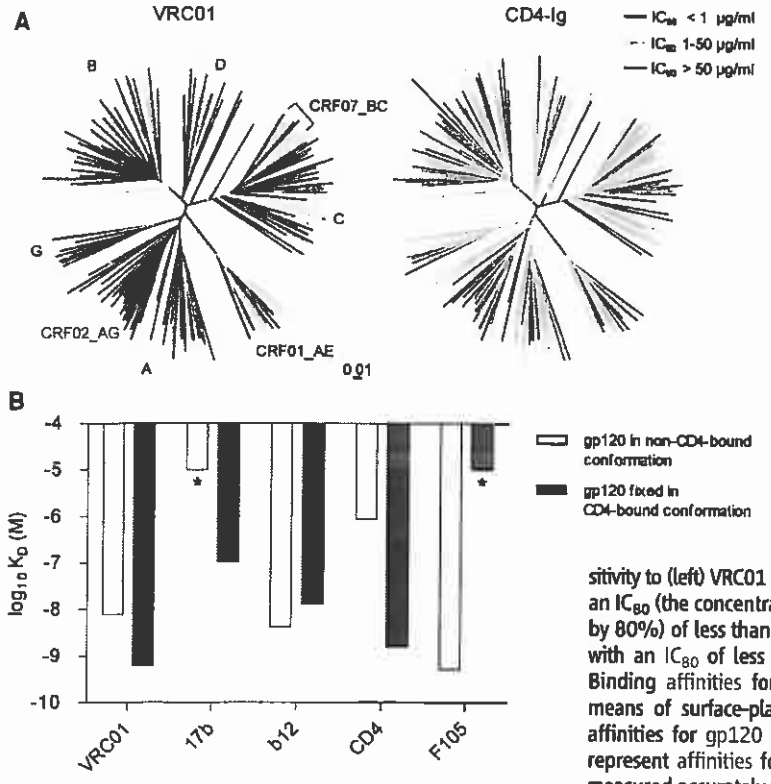
Structural basis of VRC01 breadth and potency. When CD4 is placed into an immunoglobulin context by fusing its two N-terminal domains to a dimeric immunoglobulin constant region, it achieves reasonable neutralization. VRC01, however, neutralizes more effectively (Fig. 3A) (*12*). To understand the structural basis for the exceptional breadth and potency of VRC01, we analyzed its interactive surface with gp120. VRC01 focuses its binding onto the conformationally invariant outer domain, which accounts for 87% of the contact surface area of VRC01 (table S7). The 13% of the contacts made with flexible inner domain and bridging sheet are noncontiguous, and we judged these not to be critical for binding. In contrast, CD4 makes 33% of its contacts with the bridging sheet, and many of these interactions are essential (*20*). The reduction in inner domain and bridging sheet interactions by VRC01 is accomplished primarily by a 6 Å translation relative to CD4, away from these regions; critical contacts such as made by $Phe43_{CD4}$ to the nexus of the bridging sheet–outer domain are not found in VRC01, whereas those to the outer domain (such as $Arg59_{CD4}$) are mimicked by VRC01.

To determine the affinity of VRC01 for gp120 in CD4-bound and non–CD4-bound conformations, we used surface-plasmon resonance spectroscopy to measure the affinity of VRC01 and other gp120-reactive antibodies and ligands to two gp120s: a β4-deletion developed by Harrison and colleagues that is restrained from assuming the CD4-bound conformation (*23*) or a disulfide-stabilized gp120 core that is largely fixed in the CD4-bound conformation in the absence of CD4 itself (*13*) (Fig. 3B and fig. S3). VRC01 showed high affinity to both CD4-bound and non–CD4-bound conformations, which is a property shared by the broadly neutralizing b12 antibody (*13*). In contrast, antibodies F105 and 17b as well as soluble CD4 showed strong preference for either one, but not both, of the conformations.

To assess the binding of VRC01 in the context of the functional viral spike, we examined its ability to neutralize variants of HIV-1 with gp120 changes that affect the ability to assume the CD4-bound state. Two of these mutations, $His66Ala_{gp120}$ and $Trp69Leu_{gp120}$, are less sensitive (*24*), whereas a third, $Ser375Trp_{gp120}$, is more sensitive to neutralization by CD4 (*24*, *25*). VRC01 neutralized all three of these variant HIV-1 viruses with similar potency (Fig. 3C), suggesting that VRC01 recognizes both CD4-bound and non–CD4-bound conformations of the viral spike. This diversity in recognition allows VRC01 to avoid the conformational masking that hinders most CD4-binding-site ligands (*26*) and to potently neutralize HIV-1 (*27*).

Precise targeting by VRC01. Prior analysis of effective and ineffective CD4-binding-site antibodies suggested that precise targeting to the vulnerable site of initial CD4 attachment is required to block viral entry (*11*, *28*). This site represents the outer domain contact for CD4 (*13*). Analysis of the VRC01 interaction with gp120 shows that it covers 98% of this site (Fig. 4, A and B, and fig. S4), comprising 1089 $Å^2$ on the gp120 outer domain, which is about 50% larger than the 730 $Å^2$ surface covered by CD4. The VRC01 contact surface outside the target site is largely limited to the conformationally invariant outer domain and avoids regions of conformational flexibility. This concordance of binding is much greater than for ineffective CD4-binding-site antibodies as well as for those that are partially effective, such as antibody b12 (*11*, *13*) (fig. S4).

The outer domain–contact site for CD4 is shielded by glycan (*11*, *20*). Contacts by the VRC01 light chain ($Tyr28_{VRC01}$ and $Ser30_{VRC01}$) are made with the protein-proximal N-acetylglucosamine from the N-linked glycan at residue



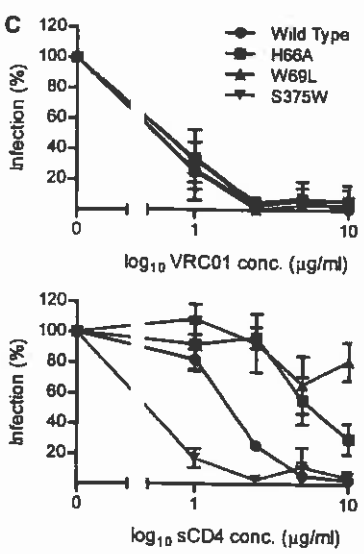


Fig. 3. Structural basis of antibody VRC01 neutralization breadth and potency. VRC01 displays remarkable neutralization breadth and potency, a consequence in part of its ability to bind well to different conformations of HIV-1 gp120. (**A**) Neutralization dendrograms. The genetic diversity of current circulating HIV-1 strains is displayed as a dendrogram, with locations of prominent clades (such as A, B, and C) and recombinants (such as CRF02_AG) labeled. The strains are colored by their neutralization sensitivity to (left) VRC01 or (right) CD4. VRC01 neutralizes 72% of the tested HIV-1 isolates with an IC_{80} (the concentration of a substance required to inhibit the activity of another substance by 80%) of less than 1 µg/ml; in contrast, CD4 neutralizes 30% of the tested HIV-1 isolates with an IC_{80} of less than 1 µg/ml (table S14). (**B**) Comparison of binding affinities (K_d). Binding affinities for VRC01 and various other gp120-reactive ligands as determined by means of surface-plasmon resonance are shown on a bar graph. White bars represent affinities for gp120 restrained from assuming the CD4-bound state (*23*), and black bars represent affinities for gp120 fixed in the CD4-bound state (*13*). Binding too weak to be measured accurately is shown as with an asterisk and bar at 10^{-5} M K_d. (**C**) Neutralization of viruses with altered sampling of the CD4-bound state. Mutant $S375W_{gp120}$ favors the CD4-bound state, whereas mutants $H66A_{gp120}$ and $W69L_{gp120}$ disfavor this state. Neutralization by VRC01 (top) is similar for wild type (WT) and all three mutant viruses, whereas neutralization by CD4 (bottom) correlates with the degree to which gp120 in the mutant viruses favors the CD4-bound state.

Downloaded from www.sciencemag.org on December 23, 2011

276_{gp120} (*29*). Thus instead of being occluded by glycan, VRC01 makes use of a glycan for binding. Other potential glycan interactions may occur with different strains of HIV-1 because the VRC01 recognition surface on the gp120–outer domain extends further than that of the functionally constrained CD4 interaction surface, especially into the loop D and the often-glycosylated V5 region (fig. S5).

Natural resistance to antibody VRC01. In addition to conformational masking and glycan shielding, HIV-1 resists neutralization by antigenic variation. In a companion manuscript, we show that of the 190 circulating HIV-1 isolates tested for sensitivity to VRC01, 173 were neutralized and 17 were resistant (*12*). To understand the basis of this natural resistance to VRC01, we analyzed all 17 resistant isolates by threading their sequences onto the gp120 structure (fig. S5). Variation was observed in the V5 region in resistant isolates, and this variation—along with alterations in gp120 loop D—appeared to be the source of most natural resistance to VRC01 (Fig. 4C and figs. S5 and S6).

Because substantial variation exists in V5, structural differences in this region might be expected to result in greater than 10% resistance. The lower observed frequency of resistance suggests that VRC01 employs a recognition mechanism that allows for binding despite V5 variation. Examination of VRC01 interaction with V5 shows that VRC01 recognition of V5 is considerably different from that of CD4 (fig. S7), with $Arg61_{VRC01}$ in the CDR H2 penetrating into the cavity formed by the V5 and β24 strands of gp120 (fig. S8). The V5 loop fits into the gap between heavy and light chains; thus, by contacting only the more conserved residues at the loop base, VRC01 can tolerate variation in the tip of the V5 loop (Fig. 4D).

Unusual VRC01 features and contribution to recognition. We examined the structure of VRC01 for special features that might be required for its function. A number of unusual features were apparent, including a high degree of affinity maturation, an extra disulfide bond, a site for N-linked glycosylation, a two-amino-acid deletion in the light chain, and an extensively matured binding interface between VRC01 and gp120 (Fig. 5 and fig. S9). We assessed the frequency with which these features were found in HIV-1 Env-reactive antibodies (appendix S1) or in human antibody-antigen complexes (fig. S10 and tables S8 and S9) and measured the effect of genomic reversion of these features on affinity for gp120 and neutralization of virus (Fig. 5, A to D, and table S10).

Higher levels of affinity maturation have been reported for HIV-1–reactive antibodies in general (*30*) and markedly higher levels for broadly neutralizing ones (*31*). These maturation levels could be a by-product of the persistent nature of HIV-1 infection and may not represent a functional requirement. Removal of the N-linked glycosylation or the extra disulfide bond, which connects CDR H1 and H3 regions of the heavy chain, had little effect on binding or neutralization (Fig. 5, A and B, and table S10). Insertion of two amino acids to revert the light chain deletion had moderate effects, which were larger for an Ala-Ala insertion (50-fold decrease in binding affinity) versus a Ser-Tyr insertion (fivefold decrease in affinity), which mimics the genomic sequence (Fig. 5C and table S10). Lastly, reversion of the interface was examined with either single-, four-, seven- or 12-mutant reversions. For the single-mutant reversions of the interface to the genomic antibody sequence, all 12 mutations had minor effects [most with a less than twofold effect on the dissociation constant (K_d), with the largest effect for a Gly54Ser change

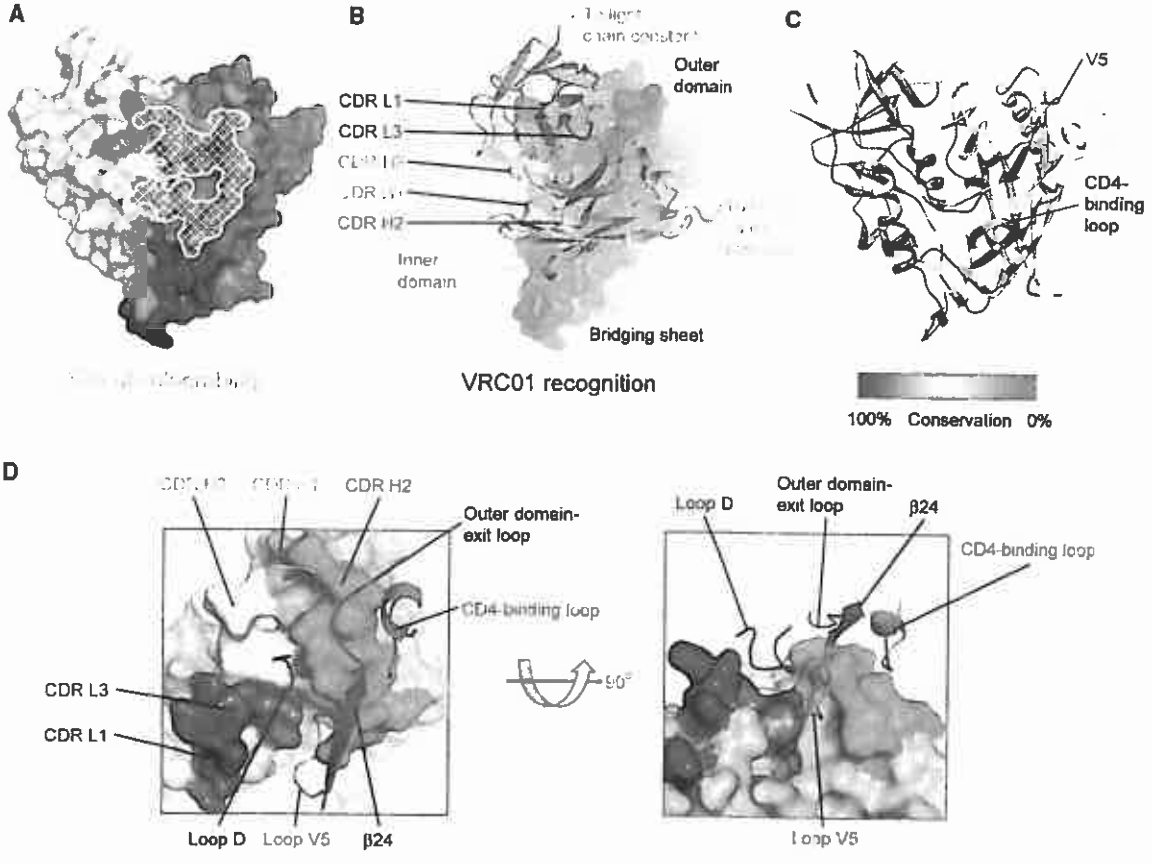


Fig. 4. Natural resistance to antibody VRC01. VRC01 precisely targets the CD4-defined site of vulnerability on HIV-1 gp120. Its binding surface, however, extends outside of the target site, and this allows for natural resistance to VRC01 neutralization. (**A**) Target site of vulnerability. The CD4-defined site of vulnerability is the initial contact surface of the outer domain of gp120 for CD4 and comprises only two thirds of the contact surface of gp120 for CD4 (*13*). The molecular surface of HIV-1 gp120 has been colored according to its underlying domain substructure: red for the conformationally invariant outer domain, gray for the inner domain, and blue for the highly mobile bridging sheet. Regions of the gp120 surface that interact with VRC01 have been colored green, with the CD4-defined site of vulnerability outlined in yellow. The view shown here is rotated 90° about the horizontal from the view in Figs. 1 and 2. (**B**) VRC01 recognition. The molecular surface of gp120 in the VRC01 bound conformation is colored as in (A). The variable domains of VRC01 are shown in ribbon representation, with the heavy and light chains colored as in Fig. 1 and extensions to constant regions indicated. (**C**) Antigenic variation. The polypeptide backbone of gp120 is colored according to sequence conservation: blue if conservation is high and red if conservation is low. (**D**) Molecular surface of VRC01 and select interactive loops of gp120. Variation at the tip of the V5 loop is accommodated by a gap between heavy and light chains of VRC01.

Downloaded from www.sciencemag.org on December 23, 2011

279

having a K_d of 20.2 nM] (table S10). Larger effects were observed with multiple (four, seven, or 12) changes (Fig. 5D and table S10). Thus, although VRC01 has a number of unusual features, no single alteration to genomic sequence substantially altered binding or neutralization.

Elicitation of VRC01-like antibodies. The probability for elicitation of a particular antibody is a function of each of the three major steps in B cell maturation: (i) recombination to produce nascent antibody heavy and light chains from genomic V_H-D-J and $V_{\kappa/\lambda}$-J precursors, (ii) deletion of auto-reactive antibodies, and (iii) maturation through hypermutation of the variable domains to enhance antigen affinity. For the recombination step, a lack of substantial CDR L3 and H3 contribution to the VRC01-gp120 interface (table S2) indicates that specific $V_{\kappa/\lambda}$-J or V_H(D)J recombination is not required (*32*) (fig. S11). The majority of recognition occurs with elements encoded in single genomic elements or cassettes, suggesting that specific joining events between them are not required. Within the V_H cassette, a number of residues associated with the IGHV1-02*02 precursor of VRC01 interact with gp120; many of these are conserved in related genomic V_Hs, some of which are of similar genetic distance from VRC01 (fig. S12). These results suggest that appropriate genomic precursors for VRC01 are likely to occur at a reasonable frequency in the human antibody repertoire.

Recombination produces nascent B cell–presented antibodies that have reactivities against both self and nonself antigens. Those with autoreactivity are removed through clonal deletion. With many of the broadly neutralizing antibodies to HIV-1, such as 2G12 (glycan reactive) (*33*, *34*), 2F5, and 4E10 (membrane reactive) (*35*, *36*), this appears to be a major barrier to elicitation. Although this remains to be characterized for genomic revertants and maturation intermediates, no autoreactivity has so far been observed with VRC01.

The third step influencing the elicitation of VRC01-like antibodies is affinity maturation, which is a process that involves the hypermutation of variable domains combined with affinity-based selection that occurs during B cell maturation in germinal centers (*37*). In the case of VRC01, 41 residue alterations were observed from the genomic V_H gene and 25 alterations from the V_κ gene (including a deletion of two residues) (fig. S13) (*38*). To investigate the effect of affinity maturation on HIV-1 gp120 recognition, we reverted the V_H and V_κ regions of VRC01, either individually or together, to the sequences of their genomic precursors. We tested the affinity and neutralization of these reverted antibodies (Fig. 6A) and combined these data with the genomic reversion data obtained while querying the unusual molecular features of VRC01 (previous section) (Fig. 6B).

No antibodies containing V_H and V_κ regions, which were fully reverted to their genomic precursors, bound gp120 or neutralized virus (*39*). Binding affinity and neutralization showed significant correlations with the number of affinity-matured residues ($P < 0.0001$). Binding to stabilized gp120 did not correlate well with other types of gp120 or to neutralization (table S11), which is related in part to greater retention of binding to VRC01 variants with genomically reverted V_κ regions. Extrapolation of the correlation to the putative genomic V gene sequences predicted binding affinities of 0.7 ± 0.4 μM K_d for gp120 stabilized in the CD4-bound conformation and substantially weaker affinites for nonstabilized gp120s (Fig. 6B and fig. S14).

No single affinity maturation alteration appeared to affect affinity by more than tenfold, suggesting that affinity maturation occurs in multiple small steps, which collectively enable tight binding

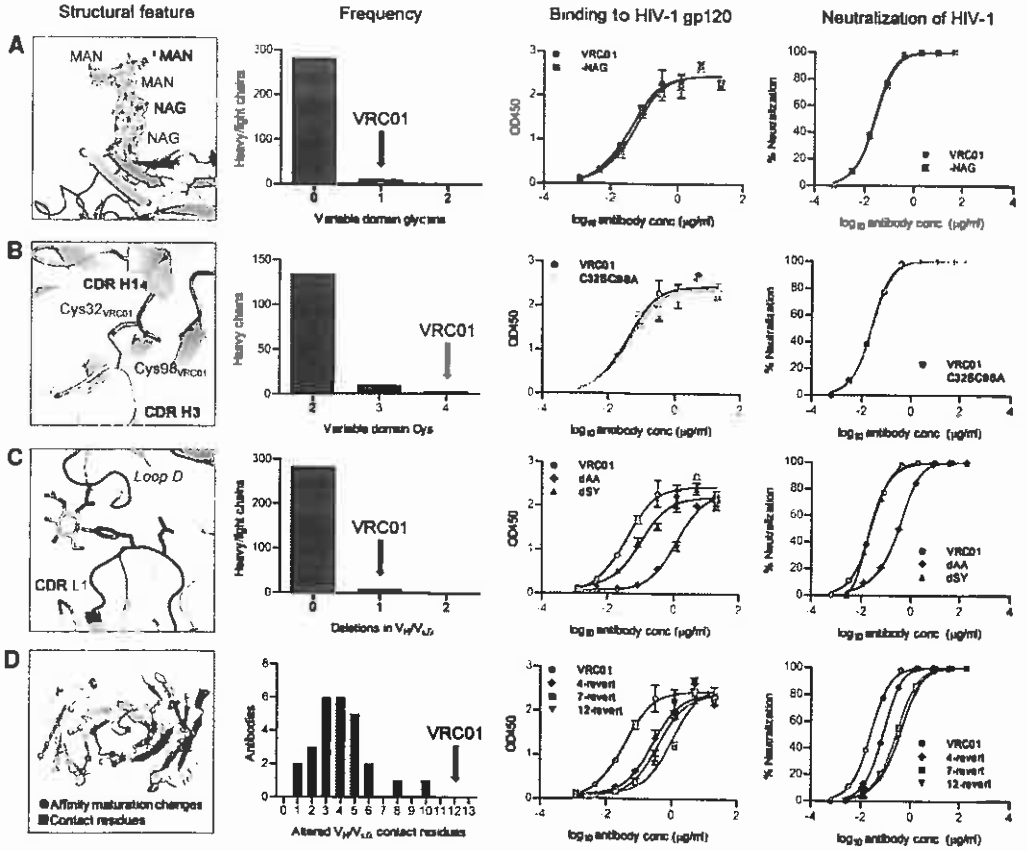


Fig. 5. Unusual VRC01 features. The structure of VRC01 displays a number of unusual features, which if essential for recognition might inhibit the elicitation of VRC01-like antibodies. (**A** to **D**) Unusual features of VRC01 are shown structurally (far left), in terms of frequency as a histogram with other antibodies (second from left), and in the context of affinity and neutralization measurements after mutational alteration (second from right and right). Affinity measurements were made by means of enzyme-linked immunosorbent assay to the gp120 construct used in crystallization (93TH057), and neutralization measurements were made with a clade A HIV-1 strain Q842.d12. Additional binding and neutralization experiments are reported in table S10. (A) N-linked glycosylation. The conserved tri-mannose core is shown with observed electron density, along with frequency and effect of removal on affinity and neutralization. (B) Extra disulfide. Variable heavy domains naturally have two Cys, linked by a disulfide; VRC01 has an extra disulfide linking CDR H1 and H3 regions. This occurs rarely in antibodies, but its removal through mutation to Ser/Ala has little effect on affinity or neutralization. (C) CDR L1 deletion. A two-amino-acid deletion in the CDR L1 prevents potential clashes with loop D of gp120. Such deletions are rarely observed; reversion to the longer loop may have a moderate effect on gp120 affinity. (D) Somatically altered contact surface. (Left) The VRC01 light chain is shown in violet, and the heavy chain is in green. Residues altered by affinity maturation are depicted with "balls," and contacts with HIV-1 gp120 are colored red. Approximately half of the contacts are altered during the maturation process. Analysis of human antibody-protein complexes in PDB shows this degree of contact surface alteration is rare; reversion of each of the contact sites to the genome sequence has little effect (table S10), although in aggregate the effect on affinity is larger.

Downloaded from www.sciencemag.org on December 23, 2011

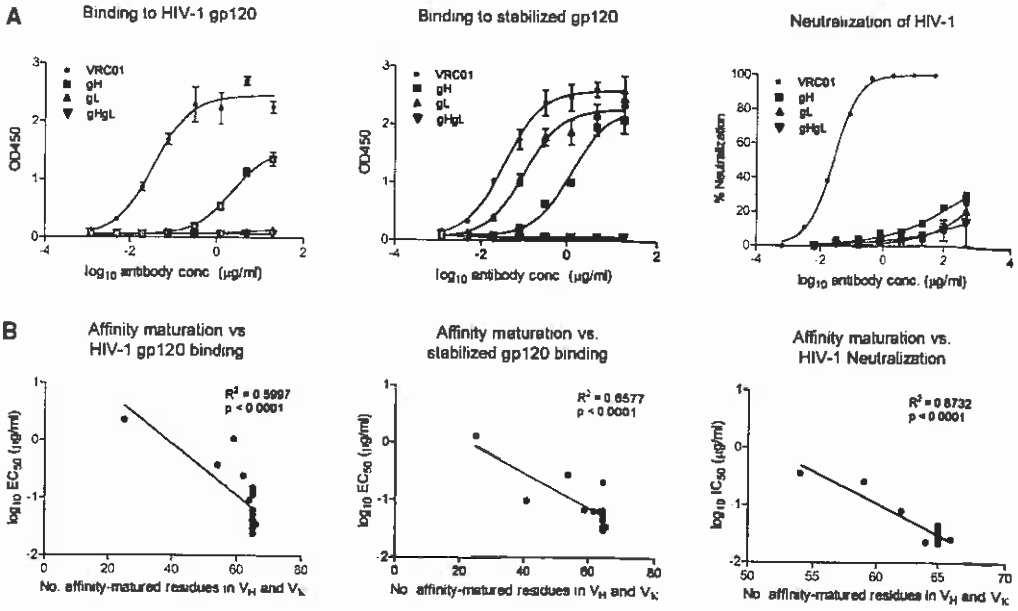


Fig. 6. Somatic maturation and VRC01 affinity. Hypermutation of the variable domain during B cell maturation allows for the evolution of high-affinity antibodies. With VRC01, this enhancement to affinity occurs principally through the alteration of noncontact residues, which appear to reform the genomic contact surface from affinity too low to measure to a tight (nanomolar) interaction. (**A**) Effect of genomic reversions. The V_H- and V_K-derived regions of VRC01 were reverted to the sequences of their closest genomic precursors, expressed as immunoglobulins, and tested for binding as V_H- and V_K-revertants (gHgL), as a V_H-only revertant (gH), or as a V_K-only revertant (gL) to the gp120 construct used in crystallization (93TH057) or to a stabilized HXBc2 core (*13*). These constructs were also tested for neutralization of a clade A HIV-1 strain Q842.d12. Additional neutralization experiments with clade B and C viruses are reported in (*15*). (**B**) Maturation of VRC01 and correlation with binding and neutralization. Affinity and neutralization measurements for the 19 VRC01 mutants created during the structure-function analysis of VRC01 were analyzed in the context of their degree of affinity maturation. Significant correlations were observed, with extrapolation to V_H- and V_K-genomic precursors suggesting greatly reduced affinity of the initial genomic arrangement for gp120.

to HIV-1 gp120. When the effects of VRC01 affinity maturation reversions are mapped to the structure of the VRC01-gp120 complex, they are broadly distributed throughout the VRC01 variable domains rather than focused on the VRC01-gp120 interface. Noncontact residues therefore appear to influence the interface with gp120 through indirect protein-folding effects. Thus, for VRC01 the process of affinity maturation entails incremental changes of the nascent genomic precursors to obtain high-affinity interaction with the HIV-1 Env surface.

Receptor mimicry and affinity maturation. The possibility that antibodies use conserved sites of receptor recognition to neutralize viruses effectively has been pursued for several decades. The recessed canyon on rhinovirus that recognizes the unpaired terminal immunoglobulin domains of intercellular adhesion molecule–1 highlights the role that a narrow canyon entrance may play in such occlusion of bivalent antibody-combining regions (*40*), although framework recognition can in some instances permit entry (*41*). Partial solutions such as those presented by antibody b12 (neutralization of ~40% of circulating isolates) (*12*, *13*) or by antibody HJ16 (neutralization of ~30% of circulating isolates) (*42*), a recently identified CD4-binding-site antibody, may allow recognition of some HIV-1 isolates.

With VRC01, the potency and breadth of neutralization (over 90%) suggest a more general solution. It remains to be seen how difficult it will be to guide the elicitation of VRC01-like antibodies from genomic rearrangement, through affinity maturation, to broad and potent neutralization of HIV-1. Accumulating evidence suggests that the VRC01-defined mode of recognition is used by other antibodies (*12*). These findings suggest that VRC01 is not an isolated example and probably provides a template for a general mode of recognition. The structure-function insights of VRC01 described here thus provide a foundation for rational vaccine design that is based not only on the particular mode of antibody-antigen interaction but also on defined relationships between genomic antibody precursors, somatic hypermutation, and interface-recognition elements.

References and Notes

1. R. A. Weiss *et al.*, *Nature* **316**, 69 (1985).
2. J. R. Mascola, D. C. Montefiori, *Annu. Rev. Immunol.* **28**, 413 (2010).
3. D. R. Burton, *Nat. Rev. Immunol.* **2**, 706 (2002).
4. D. R. Burton *et al.*, *Nat. Immunol.* **5**, 233 (2004).
5. Antibodies 2F5 and 4E10 require nonspecific membrane interaction (*35*, *36*), antibody 2G12 recognizes carbohydrate and is domain swapped (*33*, *34*), and antibody b12 was originally derived by means of phage display and has heavy-chain-only recognition (*13*).
6. A. K. Dhillon *et al.*, *J. Virol.* **81**, 6548 (2007).
7. Y. Li *et al.*, *J. Virol.* **83**, 1045 (2009).
8. Y. Li *et al.*, *Nat. Med.* **13**, 1032 (2007).
9. J. M. Binley *et al.*, *J. Virol.* **82**, 11651 (2008).
10. D. N. Sather *et al.*, *J. Virol.* **83**, 757 (2009).
11. L. Chen *et al.*, *Science* **326**, 1123 (2009).
12. X. Wu *et al.*, *Science* **329**, 856 (2010).
13. T. Zhou *et al.*, *Nature* **445**, 732 (2007).
14. The Env components of this strain derive from Clade E HIV-1 lineages (*43*).
15. Materials and methods are available as supporting material on *Science* Online.
16. M. Pancera *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1166 (2010).
17. The four independent copies of the VRC01-gp120 complex in the asymmetric unit resembled each other closely for the antibody variable domain–gp120 components, with an Cα-root-mean-square deviation of less than 0.2 Å. Elbow variation, however, between variable and constant domains was apparent, and we found one copy (molecule 1) to be more ordered than the others. In the figures, we display molecule 1; see fig. S15 for a comparison of all of the molecules in the asymmetric unit.
18. Surface areas of interaction reported in this paper were determined with the program PISA, as implemented in CCP4 (*44*). Values were about 20% higher than those reported previously for the gp120-CD4 complex (*20*), which were obtained using the program MS (*45*).
19. P. J. Maddon *et al.*, *Cell* **42**, 93 (1985).
20. P. D. Kwong *et al.*, *Nature* **393**, 648 (1998).
21. The overlap of molecular volumes was calculated by comparing the separate volumes of interacting domains with the combined volume of these domains after gp120 superposition.
22. The relative orientation of the light-chain variable domain and the heavy-chain variable domain of VRC01 is similar to that of other antibodies (fig. S16).
23. S. Rits-Volloch, G. Frey, S. C. Harrison, B. Chen, *EMBO J.* **25**, 5026 (2006).
24. A. Finzi *et al.*, *Mol. Cell* **37**, 656 (2010).
25. S. H. Xiang *et al.*, *J. Virol.* **76**, 9888 (2002).
26. P. D. Kwong *et al.*, *Nature* **420**, 678 (2002).
27. In a companion paper (*12*), we show that VRC01 binding induces 17b and CCR5 binding in the context of monomeric gp120 with an unusual entropy signature characteristic of transiting to the CD4-bound state (*46*); neutralization data, however, show that VRC01 does not induce 17b or CCR5 binding in the context of the viral spike. This difference probably arises from the more constrained gp120 conformation in the trimeric spike. Thus, although VRC01 is able to induce large conformational changes in monomeric HIV-1 gp120 that resemble those induced by CD4, VRC01 interaction with gp120 does not depend on these conformational changes.
28. X. Wu *et al.*, *J. Virol.* **83**, 10892 (2009).
29. Endo H treatment of gp120 for crystallization removed all but the protein-proximal N-acetyl-glucosamine and potential 1,6-fucose from sites of N-linked glycosylation. Examination of the VRC01 interactions with the N-acetyl-glycosamine at residue 276_{gp120} shows that both 1,4 additions and 1,6 additions are tolerated.
30. J. F. Scheid *et al.*, *Nature* **458**, 636 (2009).
31. C. C. Huang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2706 (2004).
32. Four residues are provided by the CDR H3, $Asp99_{VRC01}$ to $Trp100B_{VRC01}$, with a combined interaction surface of 123 Å^2 (tables S3 and S12). These four residues are probably contributed by the D segment (IGHD3-16*02), and none of them appears critical to VRC01 recognition because changes are observed in two of these residues in the closely related

Downloaded from www.sciencemag.org on December 23, 2011

281

broadly neutralizing antibody VRC03, which was one of two antibodies we isolated along with VRC01 (*12*). Meanwhile, three residues are provided by the CDR L3—Tyr91$_{VRC01}$, Glu96$_{VRC01}$, and Phe97$_{VRC01}$—with a combined interaction surface of 190 Å^2 (tables S3 and S13). These three residues lie at the junction between V and J genes. They make important hydrophobic interactions with loop D of gp120, and two of them are conserved between VRC01 and VRC03. Although it is difficult to know how precisely the CDR L3 needs to be aligned, with only three contact residues variation at the V_K-J gene junction should provide sufficient diversity for it to be represented in the repertoire.

33. R. W. Sanders *et al.*, *J. Virol.* **76**, 7293 (2002).
34. D. A. Calarese *et al.*, *Science* **300**, 2065 (2003).
35. G. Ofek *et al.*, *J. Virol.* **78**, 10724 (2004).
36. B. F. Haynes *et al.*, *Science* **308**, 1906 (2005).
37. C. Berek, A. Berger, M. Apel, *Cell* **67**, 1121 (1991).
38. Analysis of the HIV-1 Env-reactive antibody repertoire from infected individuals shows increased levels of affinity maturation (*30*). Analysis of a subset of this data (appendix S1) containing 147 heavy and 147 light chains from HIV-1 Env-reactive antibodies reveals an average of 15 alterations (30 maximum) for the heavy chain and an average of 8.6 alterations (22 maximum) for the light chain (fig. S13). In terms of the subset of HIV-1 Env-reactive antibodies that are broadly neutralizing (such as 2G12, 2F5, 4E10, and b12), antibodies b12 and 2G12 have 45 and 51 changes, respectively, relative to nearest genomic precursors in their V_H and J segments of the heavy chain (*31*).
39. Similar significant reductions in affinity have been observed with reversion of other broadly neutralizing antibodies to HIV-1 to putative genomic sequences (*47*–*49*); these observations have led to the suggestion that the dramatically reduced germline affinity for gp120 might hinder the initiation of affinity maturation of these antibodies (*50*). That is, if the affinity for gp120 of the genomic precursor of a broadly neutralizing antibody were below the threshold required for the nascent B cell to mature, then maturation would either not occur or would need to occur in response to a different immunogen. This lack of guided initiation of the maturation process may provide an explanation for the absence of such broadly neutralizing antibodies in the first few years of infection. Conversely, the introduction of modified gp120s with affinity to genomic precursors and affinity maturation intermediates could provide a mechanism by which to elicit antibodies like VRC01.
40. M. G. Rossmann, *J. Biol. Chem.* **264**, 14587 (1989).
41. T. J. Smith, E. S. Chase, T. J. Schmidt, N. H. Olson, T. S. Baker, *Nature* **383**, 350 (1996).
42. D. Corti *et al.*, *PLoS ONE* **5**, e8805 (2010).
43. J. P. Anderson *et al.*, *J. Virol.* **74**, 10752 (2000).
44. E. J. Dodson, M. Winn, A. Ralph, *Methods Enzymol.* **277**, 620 (1997).
45. M. L. Connolly, *J. Mol. Graph.* **11**, 139 (1993).
46. D. G. Myszka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9026 (2000).
47. X. Xiao, W. Chen, Y. Feng, D. S. Dimitrov, *Viruses* **1**, 802 (2009).
48. X. Xiao *et al.*, *Biochem. Biophys. Res. Commun.* **390**, 404 (2009).
49. M. Pancera *et al.*, *J. Virol.* **84**, 8098 (2010).
50. D. S. Dimitrov, *MAbs* **2**, 347 (2010).
51. T.Z., I.G., Z.Y., J.S., L.S., G.J.N., J.R.M., and P.D.K. designed the research; T.Z., I.G., X.W., Z.Y., K.D. A.F., W.S., L.X., Y.Y., and J.Z. performed the research; X.W., Y.K., J.F.S., M.C.N., and J.R.M. contributed new reagents or reference data; T.Z., I.G., X.W., J.S., L.S., G.J.N., J.R.M., and P.D.K. analyzed the data; and T.Z., I.G., J.S., L.S., G.J.N., J.R.M., and P.D.K. wrote the paper, on which all authors commented. We thank I. A. Wilson and members of the Structural Biology Section and Structural Bioinformatics Core, Vaccine Research Center, for discussions and comments on the manuscript; the staff of sector 22 for assistance with data collection; and J. Stuckey for assistance with figures. Support for this work was provided by the Intramural Research Program of NIH and by grants from NIH and from the International AIDS Vaccine Initiative's Neutralizing Antibody Consortium. Use of sector 22 (Southeast Region Collaborative Access team) at the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract W-31-109-Eng-38. Coordinates and structure factors for the VRC01-gp120 complex have been deposited with the PDB under accession code 3NGB.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1192819/DC1
Materials and Methods
Figs. S1 to S16
Tables S1 to S14
References
Appendix S1

25 May 2010; accepted 1 July 2010
Published online 8 July 2010;
10.1126/science.1192819
Include this information when citing this paper.

Downloaded from www.sciencemag.org on December 23, 2011

Gamma-Ray Emission Concurrent with the Nova in the Symbiotic Binary V407 Cygni

The Fermi-LAT Collaboration*†

Novae are thermonuclear explosions on a white dwarf surface fueled by mass accreted from a companion star. Current physical models posit that shocked expanding gas from the nova shell can produce x-ray emission, but emission at higher energies has not been widely expected. Here, we report the Fermi Large Area Telescope detection of variable γ-ray emission (0.1 to 10 billion electron volts) from the recently detected optical nova of the symbiotic star V407 Cygni. We propose that the material of the nova shell interacts with the dense ambient medium of the red giant primary and that particles can be accelerated effectively to produce π^0 decay γ-rays from proton-proton interactions. Emission involving inverse Compton scattering of the red giant radiation is also considered and is not ruled out.

V407 Cygni (V407 Cyg) is a binary system consisting of a Mira-type pulsating red giant (RG) with a white dwarf (WD) companion; these properties place it among the class of symbiotic binaries (*1*). Though one of the more active symbiotic systems, V407 Cyg has historically shown an optical spectrum in quiescence dominated by the Mira-like RG (M6 III) and only weak emission lines [see, for example, (*2*)]. Its infrared continuum (consistent with a dusty wind) and maser emission (*3*) are detected at levels similar to other symbiotic Miras [for instance, R Aqr (*4*)]. One outstanding anomaly of V407 Cyg is a strong Li I λ6707 line indicative of an overabundance of Li relative to normal Mira RGs (*5*, *6*). Based on the 745-day pulsation period of the RG (*7*) and the Mira period-luminosity relation (*8*), we adopt the distance $D = 2.7$ kpc, estimated as the mean derived from photometry in three near-infrared bands, assuming an extinction $E_{B-V} = 0.57$ (*2*).

A nova outburst from V407 Cyg was detected on 10 March 2010 (*9*); it had a magnitude of ~6.9 in an unfiltered charge-coupled device image obtained at 19:08 UT. Subsequent densely sampled observations show that the outburst was followed by a smooth decay, though the precise epoch of the nova is formally uncertain by up to 3 days, due to the time gap from the pre-outburst image (Fig. 1). Monitoring of the source over the past 2 years indicates pre-outburst magnitude values in the range of 9 to 12 [see the supporting online material (SOM)]. V407 Cyg has been monitored optically for decades and has shown earlier signs of optical brightening on month-long time scales by one to two magnitudes in the B and V bands (around 1936 and 1998) from typical V-band magnitudes of 13 to 16 (*2*, *10*, *11*), but the magnitude of the recent nova was unprecedented.

Here, we report on a high-energy γ-ray source (Fig. 2) positionally coincident with V407 Cyg detected after the nova (*12*) during routine automated processing of all-sky monitoring data from the Fermi Large Area Telescope (LAT) (*13*). A γ-ray light curve (1-day time bins) of this source generated from an analysis of all LAT data reveals that the first significant detection (4.3σ) was, in fact, on 10 March, indicating that the γ-ray activity began on the same day as the reported optical maximum of V407 Cyg (Fig. 1; see also SOM). The observed 10 March flux is up to a factor of 3 larger than the 1-day upper limits (unless otherwise noted, 95% confidence limits are reported throughout) on the pre-outburst days. To further isolate the onset of detectable γ-ray emission, we divided the 10 March data into 6-hour intervals,

*All authors with their affiliations appear at the end of this paper.
†To whom correspondence should be addressed. E-mail: Teddy.Cheung.ctr@nrl.navy.mil (C.C.C.); Adam.Hill@obs.ujf-grenoble.fr (A.B.H.); Pierre.Jean@cesr.fr (P.J.); srazzaque@ssd5.nrl.navy.mil (S.R.); Kent.Wood@nrl.navy.mil (K.S.W.)

Mathematical model of active immunity

Let V denote the concentration of viruses, and A the concentration of antibodies against such viruses. Initially, an individual is infected by a small number of viruses, V_0, and has a small amount of antibodies, A_0. The viruses will increase in number by multiplication, but be reduced by binding to antibodies. Instead of the detailed biological processes of the immune system, we consider the following drastically simplified equations:

$$dV / dt = \alpha V - \beta V A$$

$$dA / dt = \gamma V A$$

The constants α, β, and γ are all positive. The initial conditions are at $t = 0$, $V = V_0$ and $A = A_0$. These two non-linear equations can be solved numerically. V will steadily increase until sufficient antibodies are produced to fight against the viruses. Then V will eventually decrease and become zero. At the same time, A will gradually increase and reach a steady state. This result has been illustrated in standard immunology textbooks.

On the other hand, we can combine these two equations to provide a concise result between V and A. The second equation gives:

$$V = (1 / \gamma A)\ dA / dt$$

On substituting this into the first equation, we get:

$$dV / dt = (\alpha / \gamma A)\ dA / dt - (\beta / \gamma)\ dA / dt$$

This equation can be solved to give:

$$V = (\alpha / \gamma)\ \ln A - (\beta / \gamma)\ A + \text{constant}$$

The constant can be determined by the initial conditions. Finally, we get an equation relating V to A as follows:

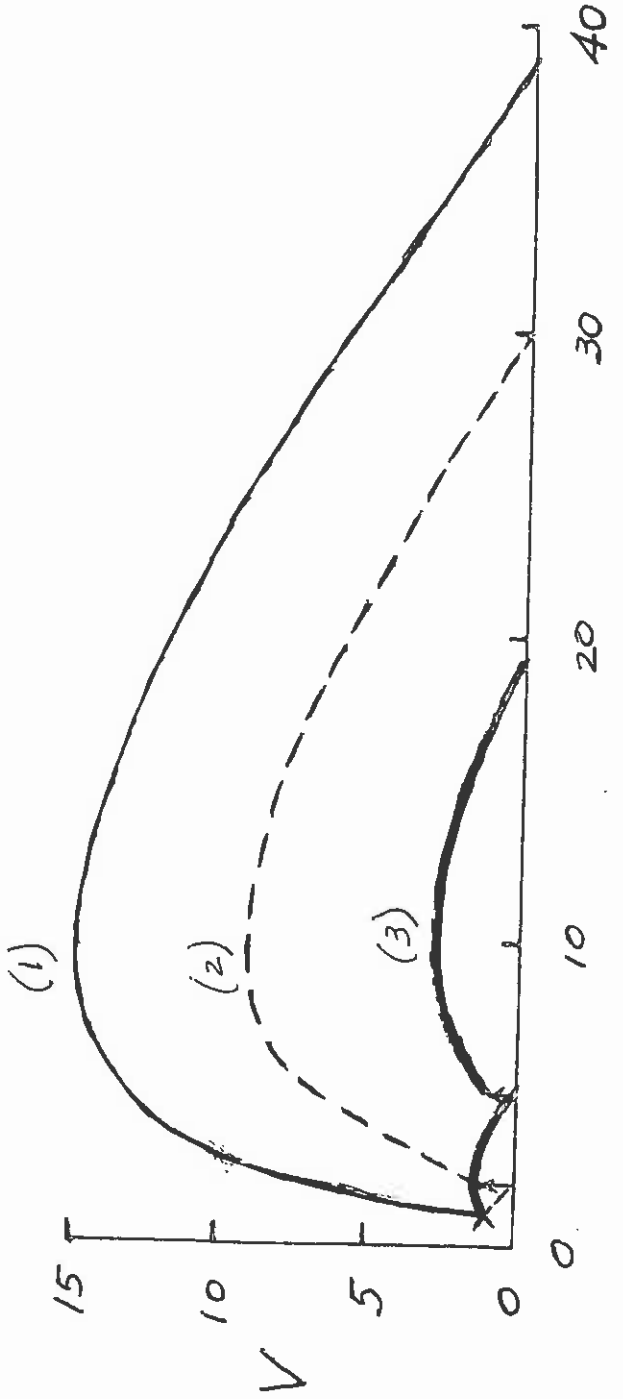
$$V = V_0 + (\alpha / \gamma) \ln (A / A_0) + (\beta / \gamma) (A - A_0)$$

We can plot V against A to get an indication of how the disease progresses. Three different situations will be considered: (1) infection with a virulent virus without previous vaccination, (2) infection with the virulent virus after vaccination with a dead virus, and (3) infection with a virulent virus after vaccination with an attenuated virus. For simplicity, let $V_0 = 1$, $A_0 = 1$, and $\beta / \gamma = 1$. For the virulent virus, assume $\alpha / \gamma = 10$. For dead virus, $\alpha / \gamma = 0$. For the attenuated virus, assume $\alpha / \gamma = 2$.

V is plotted against A on the next page showing the three different situations. In (1), V reaches a value of about 15 before coming down. In (2), the maximum value of V is around 9.1. In (3), V only gets to roughly 2.9 at the peak.

If we assume that when V reaches a value of 10, the patient will die. Then in (1), the patient dies from the infection of the virulent virus. On the other hand, in (2), the patient gets very sick after infecting with the virulent virus but recovers. In (3), the patient may not even know that he has been subsequently infected with the virulent virus. This is known as a sub-clinical infection.

This simple mathematical model illustrates the usefulness of vaccination. Vaccinating with attenuated viruses are more efficient than vaccinating with dead viruses. However, one has to be very careful about such decisions. Recently, in HIV endemic areas in Africa, vaccination with attenuated poliomyel viruses resulted in many deaths due to the low efficiency of the children's immune systems. They should be vaccinated with dead poliomyelitis viruses.



A