Lesson #7

June 24, 2011

## ENGINEERING DESIGN OF THERAPEUTIC ANTIBODIES (Tai Te Wu)

### Introduction

Therapeutic antibodies are designed using both genetic and protein engineering methods. It requires the detailed knowledge of analytical engineering method on one hand and of molecular biology on the other hand. Many of them have been approved by the Food and Drug Administration (FDA) during the past decades for the treatment of specific diseases including cancer, auto-immune malfunctions, viral infections, etc. In addition, many more artificial antibodies have been constructed for diagnostic purposes.

The present course provides an over view of the historical development of this rapidly developing field. It illustrates how engineering and biomedical sciences can work together to generate new tools for the treatment of diseases. In particular, most cancers are treated by surgical removal, if possible, followed by chemotherapy and radiation therapy. With the advancement of therapeutic antibodies, a new modality known as immunotherapy has been brought to full operation.

We need to trace the meticulous approaches of many biologists over the years of defining what antibodies are. After their amino acid and nucleotide sequences have been determined, engineers then applied simple analytical approaches to decipher the information content of these sequences. The joint effort became extremely successful in the detailed understanding of how antibodies bind antigens with precise interactions, based on three dimensional structures of such complexes.

At the genetic level, a new mechanism known as DNA re-arrangement was discovered. At the protein level, detailed analytical studies involving the folding of the backbone structure became essential. The precise delineation of the antibody combining site segments provided the guideline of how recombinant molecules at both the DNA and protein levels should be constructed. With the combination of all such knowledge together with modern day biotechnology methods eventually made the production of therapeutic antibody possible.

Currently, further engineering design of these therapeutic antibodies can even improve their antigen binding affinities beyond what natural selection of antibodies in humans or animals has dictated. Based on these approaches, therapeutic antibodies become one of the most rapidly growing medications in the pharmaceutical market.

Topics to be covered are:

1. Discovery of vaccination
2. Concept of antibodies
3. Use of immunized horse serum to treat infection

4. Gamma globulin fraction of serum containing antibodies
5. Bence Jones protein and multiple myeloma
6. Basic structure of an antibody molecule
7. Determination of amino acid sequences of antibody light chains
8. Variability plot and complementarity determining regions (CDR's)
9. Similar studies for antibody heavy chains
10. Hydridoma technology
11. Three dimensional structure of antibody-antigen complex
12. Genetic mechanism of generating different antibody genes
13. Folding of protein backbone structures
14. Prediction of three dimensional structure of CDR's
15. Engineering design of therapeutic antibodies
16. The first monoclonal therapeutic antibody
17. Chimeric therapeutic antibody
18. Humanized mouse therapeutic antibody
19. Human therapeutic antibody
20. Improve binding affinity of therapeutic antibody
21. Ultimate goal of therapeutic antibodies
22. Limitations of therapeutic antibodies

**Discovery of vaccine**

Initially, during the early 1700's in Asia, the scabs from patients infected with smallpox were collected, dried and pulverized. This powder was either snuffed or introduced into small skin cuts. It was supposed to prevent smallpox infection, but could also result in full blown infections. This practice was soon abandoned.

The classical case of immunizing humans with cowpox against smallpox infection was described in detail in W. Osler, The Principles and Practice of Medicine, 1918, D. Appleton and Company, New York and London, p. 326. Chronologically, the events took place as follows:

In 1774, Jesty successfully vaccinated his wife and two sons.

In 1791, Plett also successfully vaccinated his three sons.

On May 14, 1796, Jenner, following the suggestion of successful vaccination received by a young girl, took the pustule from Nelmes infected with cowpox, inoculated into an eight year old boy Phipps. Phipps was subsequently inoculated with the pustule from a smallpox patient, and did not come down with smallpox. The result was published in 1798. That experiment had no control to prove that the pustule from the smallpox patient was indeed infectious.

On August 16, 1802, Waterhouse inoculated nineteen boys with cowpox. On Nov. 9, 1802, twelve of them were then inoculated with smallpox and none of them came down with the disease. He also did a control experiment by inoculating two unvaccinated boys with smallpox and both of them came down with the disease. By the way, such control experiments are no longer allowed, even though they are essential to verify the disease agent.

As clearly illustrated here, folk medicine cannot be discarded completely. Indeed, the tradition among cow farmers to protect their family against smallpox was probably widely accepted, and predated the medical profession by more than twenty years. Even before 1774, there were other less reliable approaches of vaccination against smallpox (Bibel DJ (1988) Milestones in Immunology, Science Tech Publishers). The use of cowpox, essentially non-virulent in humans, to protect against the highly virulent smallpox infection started the development of immunology.

Jenner has been designated as the father of immunology, since he published his result in 1798 (An inquiry into the cause and effects of the variolae vaccinae, a disease discovered in some of the western counties of England, particularly Gloucestershire and known by the name of cowpox. London, Sampson Low. 75 p.)

Of course, at that time, it was not understood why this process of vaccination worked. It was postulated that somehow the human body reacted to the introduction of the essentially non-virulent cowpox virus by producing some substance which could eliminate the virulent smallpox virus. That substance is subsequently known as antibody.

**Concept of antibodies**

It was soon suspected that the protective effect of cowpox vaccination resulted from the human producing a substance designated as antibody capable of binding the cowpox virus and also with the ability of cross-reacting with the smallpox virus. This was verified by Behring and Kitasato in 1890 (von Behring E, Kitasato S (1890) Ueber der Zustandekommen der Diphthrie-Immunitat und der Tetanus-Immunitat bie Thieren, Dysch Med Wochsuschr, 16, 1113-1114). They immunized animals with diphtheria toxin or tetanus toxin, and discovered the serum from these animals could protect other animals from fatal doses of such toxins.

They immunized rabbits, for example, against tetanus so that these rabbits could tolerate twenty times the lethal amount of tetanus bacilli or tetanus toxin. Small amounts of the blood or serum from such rabbits were then injected into abdominal cavities of mice. They were sufficient to protect the mice from fatal doses of tetanus bacilli. In other words, the blood or serum of the immunized rabbits contained substances capable of destroying the tetanus bacteria. They could also protect against tetanus toxin.

Similar studies were carried out for diphtheria bacteria and toxin. In addition to rabbits, they also experimented with cows, calves, horses and sheep.

**Use of immunized horse serum to treat infections**

Indeed, horses were subsequently used for the production of large amounts of anti-diphtheria serum for the treatment of human patients. Unfortunately, due to the presence of numerous other substances in the horse serum, sometimes the side effect could be life threatening.

While vaccination involved active immunization, the use of immune serum started the era of passive immunization. The active ingredients in the immune serum had to be isolated in order to gain a precise understanding of how they functioned. Furthermore, the undesirable side-effects known as serum sickness could be completely avoided.

However, around the end of 19$^{th}$ century and the beginning of the 20$^{th}$ century, immunized horse serum was the only effective treatment against diphtheria infection in humans. Such horse serum contained both anti-bacteria and anti-toxin components and saved many lives, even though serum sickness was a major concern.

Indeed during the influenza pandemic around 1918, similar approaches were attempted to treat influenza infection but turned out to be ineffective. Instead, immunized horse serum was used to treat pneumonia, a major complication of influenza infection, with some success.

Currently, influenza is prevented by vaccination, and diphtheria treated by antibiotics discovered after the 1920's. During the past century, human infections by micro-organisms have exploded. Medical remedies involve active immunization, passive immunization, antibiotics, anti-viral medications, etc.

At the same time, the use of immunized horse serum has undergone major developments with the basic understanding of the properties of antibodies. This approach of passive immunization has reached at the level of molecular structures, genetic engineering, protein engineering, detailed design, etc.

It is thus essential to isolate the active components of the immunized horse serum, or antibodies specific for defined infections.

**Gamma globulin fraction of serum containing antibodies**

The first attempt of trying to isolate the active ingredient of the immunized horse serum was carried out by L. B. Felton in 1924 from horses immunized against pneumonia. He diluted the horse serum with water at a ratio of 1 to 10, and noticed a white precipitate formed. His white powder was then used to treat patients with pneumonia successfully without the side effect of serum sickness. However, it was not known how much potency of the original immunized horse serum was retained.

Using physical methods, rabbit antibodies were identified in the gamma globulin fraction of serum while the horse antibodies appeared as an isolated peak between the beta and gamma fractions of the serum (Tiselius A, Kabat EA (1939) An electrophoretic study of immune sera and purified antibody preparations. J Exp Med, 69, 119-131). Indeed, the gamma globulin fraction of serum is probably the first therapeutic antibody ever used in humans, although it contains thousands of different antibodies. Most of the unwanted components of the original immunized horse serum or serum from other immunized animals are thus removed without any loss of its therapeutic potency.

Gamma globulin is usually prepared from pooled sera of 3,000 to 10,000 or even 100,000 healthy donors, thus presumably containing all sorts of probably millions or more different antibodies, although the concentration of each individual antibody can be very small. Current methods of preparation have eliminated all viral contaminations. Most of the intravenous preparations contained pooled immunoglobulins IgG and some IgA as originally noted by Tiselius and Kabat. Immunoglobulins IgM have much larger molecular weights and migrate in the region between beta and gamma globulins.

Gamma globulins have been used to prevent various bacterial and viral infections such as those common in third world countries. Of course, the donors for the pooled sera should be from geographical locations where such infections are prevalent.

A special preparation known as intravenous immunoglobulin (IVIG) can neutralize pathogens, microbial toxins and superantigens, modulate T cell functions and antibody production, stimulate antibody-dependent cellular cytotoxicity, block cellular receptors for pathogens, promote opsonization and phagocytosis, balance the cytokine network, recruit pathogens into lymphoid organs, etc. FDA has approved its use for allogeneic bone marrow transplant, chronic lymphocytic leukemia, idiopathic thrombocytopenic purpura, pediatric HIV, primary immunodeficiencies, Kawasaki disease, chronic inflammatory demyelinating polyneuropathy, special kidney transplantation, etc. However, the basic mechanisms of IVIG remain unclear.

# AN ELECTROPHORETIC STUDY OF IMMUNE SERA AND PURIFIED ANTIBODY PREPARATIONS*†

By ARNE TISELIUS, Ph.D., and ELVIN A. KABAT, Ph.D.

*(From the Institute of Physical Chemistry, University of Upsala, Upsala, Sweden)*

(Received for publication, September 20, 1938)

Studies on the electrophoresis of serum have shown that besides the albumin normal sera possess three separate globulin components differing in mobility and designated as $\alpha$, $\beta$, and $\gamma$ (1). The relationship of antibodies to these components is of considerable importance. Differences in the antibody globulin formed in various animal species have already been found by ultracentrifugal studies (2, 3) and in electrophoresis (4) as well as by immunological means (5). It has also been shown (1) that in an antiserum to crystalline egg albumin, the antibody migrated with the slowest ($\gamma$) component and could be isolated in one of the cells of the apparatus and analyzed for antibody content. The work reported in this communication is an attempt to study and compare the electrochemical properties of the antibody in the original sera with the purified antibody preparations described in (6) and to measure the isoelectric points of these purified preparations. It is also of importance to make use of and correlate the results obtained using both the mutually independent electrophoretic and ultracentrifugal methods since molecular weight homogeneity does not necessarily mean electrochemical homogeneity and *vice versa*. Since the methods previously described for determining the concentration of the various components are not sufficiently precise, especially in systems of several components, the Lamm scale method (7) has been used with the electrophoresis apparatus (8) and has provided a quantitative method of following changes in the various serum components on removal of antibody.

* The expenses of this investigation were defrayed by grants from the Andersson and Nobel Foundations.

† A preliminary note was published in *Science*, 1938, **87**, 312.



7

## *Methods*

The new Tiselius electrophoresis apparatus (8), in which sections of the U tube may be moved with respect to one another and cut off the column of solution into four parts to permit isolation of various components, was used. The apparatus is designed to carry strong currents, which makes possible adequate resolution of the components of serum even in the presence of large amounts of salt. Two methods of observation of the migrating boundaries were applied, Toepler's *Schlieren* method (9) and the Lamm scale method (7), both depending on refractive index changes in the solution due to the migrating boundary. The former of these in which each component appears as a black band in the U tube image in the focus of the camera, provides a convenient way of following the experiment by direct visual observation. In the latter method an equidistant scale placed behind the U tube is photographed at intervals through the solution. These photographs are compared microscopically with those from a reference scale photographed before the current is started. The displacements of the scale lines, $Z$, from their positions on the reference scale are plotted as ordinates against the corresponding positions in the U tube cell as abscissae yielding a curve in which each electrochemically distinct component in the solution will have its own peak (Fig. 1) provided the differences in mobility are sufficiently large and the experiment run long enough to produce adequate separation. Since $Z = k \frac{dc}{dx}$, where $k$ depends on the refractive index increment of the migrating substance and on known apparatus constants, it is evident that integration of the entire individual curve corresponding to a single peak will give the concentration, $c$, of the substance having the mobility of that peak. One may thus obtain from the same curve the mobility and concentration of the electrochemically different proteins in the solution. A micro modification of the apparatus in which 2 ml. of solution could be used was found convenient for the isoelectric point measurements on the purified antibody preparations.

In practice, however, since only data on the relative concentrations of the various components were desired and since nitrogen content and the area under each peak were related to protein concentration, it was not necessary to calculate absolute protein concentration in each case. A base line for area measurements was determined by comparing the reference scale and photograph in the adjacent cells where no boundaries appeared.

It was also found convenient to compensate the boundary out from under the glass plates by a clockwork mechanism lowering an ebonite rod into the solution (8) before starting the current and then measuring the distance moved by an eyepiece scale fixed to the camera, using the *Schlieren* method, and in this manner to calculate the mobility.

All experiments were conducted in buffer solutions containing 0.15 M NaCl + 0.02 M total phosphates at the desired pH. In the more acid range, the saline was buffered with acetate buffer in which the NaAc concentration was 0.02 M. This

high salt concentration was found necessary since some of the horse and pig antibodies precipitated from solution if the usual 0.1 ionic strength buffers were employed. This procedure also enables the direct comparison of mobilities at different pH since the total ionic strength of the solution is but slightly changed by variation of the buffer and has the additional advantage of permitting the direct analysis of the solutions for antibody nitrogen by the absolute quantitative precipitin (10) and agglutinin (11) methods. The high salt concentration, however, does not permit the use of as high voltages as in 0.1 ionic strength buffer. A potential of 120 volts was usually employed giving a current of about 26 to 28 milliamperes and a potential gradient of about 3.4 volts per cm. with the ordinary apparatus. Higher voltages increased the risk of heat convection currents.

For the scale method experiments it was found advisable to use a 1:4 diluted serum to obtain optimum scale line displacements. In the case of unfractionated horse sera, however, it was not possible to determine concentration with the scale method since the curves did not come within a reasonable distance of the base line between components. This seemed also to be the case with the pig serum studied, but in rabbit and monkey antisera determination of concentration was quite satisfactory. These differences between horse and rabbit sera were most marked in higher concentrations and may perhaps indicate partial interaction or compound formation between the various globulin components of horse antisera (*cf.* Kendall, 12).

The usual procedure for studying the distribution of antibody in serum was to dilute the unabsorbed serum 1:4 and dialyze against buffer overnight. Another sample of the same serum was absorbed by addition of the proper amount of antigen and the combined supernatant and washings of the precipitate were diluted to the same volume as the unabsorbed serum and also dialyzed against buffer. In this manner the other components of the serum were present in unchanged concentration.

## RESULTS

Table I is a summary of the mobility data obtained for the antisera from different species in the saline phosphate buffer mixture at pH 7.72 ± 0.02. It will be observed that the mobility of all the components is less than in ordinary 0.1 ionic strength buffer (*cf.* 1). It can also be noted that in the horse antipneumococcus sera a new component migrating between $\beta$ and $\gamma$ was present. That this component disappeared on removal of the antibody (20.7 per cent of the total protein) can be seen from the *Schlieren* photograph (Fig. 2) of 902, before and after absorption of the antibody. No new component was observed in the other sera at the same pH. The mobility values, as measured for the slower components, would seem

to be subject to a somewhat greater experimental error in high salt concentration. In some cases, especially in concentrated solutions, the distance migrated did not correspond in each side of the U tube. This is a limitation of the method since not all of the components can be observed in more dilute solutions. As it was usually possible to get sharper bands on the positive side of the U tube, *i.e.* protein migrating into buffer, these values were used to calculate the mobilities listed in Table I.

TABLE I

*Electrophoretic Mobilities of the Components of Sera from Various Animal Species at pH 7.72 ± 0.02 and in Buffer Containing 0.15 M NaCl + 0.02 M $PO_4^{\equiv}$ Temperature + 0.5°*

| Species | Serum | Antibody | u × 10⁵ cm.² sec.⁻¹ volt⁻¹ | | | | |
|---|---|---|---|---|---|---|---|
| | | | Albumin | α | β | Antibody | γ |
| | | per cent | | | | | |
| Horse anti-pneumococcus I serum | 902 I undiluted | 20.7 | −5.5 | −3.7 | −3.0 | −2.1 | −0.9 |
| | 902 II undiluted | 21.8 | −5.7 | −4.0 | −3.2 | −2.2 | −1.3 |
| Normal horse | | | −5.7 | −3.8 | −3.1 | — | −1.0 |
| Pig antipneumococcus I serum | W undiluted | 1.6* | −5.7 | −3.5 | −2.7 | — | −1.1 |
| Rabbit anti-egg albumin serum | 431-5 1:4 diluted | 36.4 | −6.0 | −3.6 | −2.9 | — | −1.1 |
| Monkey antipneumococcus III serum | 8-58 1:4 diluted | 6.6 | −5.2 | −4.3 | −3.0 | — | −0.7 |

* This serum showed 9.7 per cent of a heavy component in the ultracentrifuge.

Table II is a summary of the results obtained by electrophoresis of rabbit and monkey antipneumococcus and anti-egg albumin sera before and after removal of the antibody. The type of curves obtained in the case of anti-egg albumin serum 431-5 is shown in Fig. 1 and the marked decrease in the $\gamma$ component in the absorbed serum is evident. The absence of antibody in the $\alpha$ and $\beta$ components was also directly established in the case of a rabbit antiserum to crystalline horse serum albumin by isolating the albumin and $\alpha$ and $\beta$ components and demonstrating the absence of antibody with crystalline serum albumin.

Table III gives the mobility and isoelectric point measurements on

antibody preparations from different animal species. The details of the preparations are included in (3). All the preparations showed but one *Schlieren* band at each pH and the horse antibody preparation

TABLE II

*Distribution of Antibody in Rabbit and Monkey Antisera*

| Species | Serum | Antibody by analysis | | Total area under curve | Area $\gamma$ component unabsorbed | Area $\gamma$ component absorbed | Antibody area by difference | $\gamma$ component in unabsorbed serum | Antibody in $\gamma$ component | Antibody in original serum |
|---|---|---|---|---|---|---|---|---|---|---|
| | | *per cent* | | *sq. cm.* | *sq. cm.* | *sq. cm.* | *sq. cm.* | *per cent* | *per cent* | *per cent* |
| Rabbit anti-egg albumin | 431-5 | 36.4 | Positive side | 229.4 | 119.8 | 34.4 | 85.4 | 52.2 | 71.2* | 37.2 |
| | | | Negative side | 252.4 | 125.1 | 29.5 | 95.6 | 49.6 | 76.4* | 37.9 |
| Rabbit anti-egg albumin | 446 | 12.3 | Positive side | 94.9 | 31.1 | 18.6 | 12.5 | 32.8 | 40.2 | 13.2 |
| | | | Negative side | 111.5 | 32.1 | 16.7 | 16.1 | 28.8 | 48.0 | 13.8 |
| Rabbit antipneumococcus Type I | 517 | 18.6 | Negative side†‡ | 143.4 | 80.5 | 28.2 | 52.3 | 56.0 | 35.0 | 19.6 |
| Monkey antipneumococcus Type III | 8-58 | 6.6 | Positive side | 206.1 | 53.7 | 31.7 | 22.0 | 26.1 | 41.0 | 10.7 |
| | | | Negative side | 217.6 | 46.4 | 28.7 | 17.7 | 21.3 | 38.2 | 8.1 |

* Isolation of the component from this serum by electrophoresis yielded a solution containing 76 per cent antibody by analysis.

† Positive side results not used because of unsatisfactory reference scale; base line could not be accurately drawn.

‡ A different dilution was used for absorbed and unabsorbed sera; values given are corrected to the same dilution.

showed only one component by the scale method. In the case of the purified antibodies the distance moved in both the positive and negative sides of the U tube was practically the same and the values given are averages. The mobility pH curves are shown in Fig. 3. Insuffi-

cient monkey antibody was available for an isoelectric point determination. The mobility values obtained for the purified horse, pig,

TABLE III

*Mobilities of Purified Antibodies at Various pH*

$u \times 10^5$ cm.$^2$ sec.$^{-1}$ volt$^{-1}$

| pH | Antipneumococcus | | | | | Anti-egg albumin |
|---|---|---|---|---|---|---|
| | | | | Rabbit | | |
| | Horse 902 E Salt dissociated | Cow $D_3$ Salt dissociated | Pig $W_1B$ Salt dissociated | $456_2$ A Salt dissociated | $456_3$ B Barium dissociated | 431-5 γ Isolated by electrophoresis |
| 3.18 | +2.21 | | | | | |
| 4.02 | +1.12 | +1.43 | +2.15 | | | |
| 4.81 | −0.63 | 0.00 | +0.49 | +1.04 | +1.10 | +0.95 |
| 5.86 | −1.12 | −0.98 | −0.62 | 0.00 | 0.00 | 0.00 |
| 6.70 | −1.34 | −1.20 | −1.02 | −0.86 | −0.60 | −0.75 |
| 7.72 | −1.70 | −1.31 | −1.30 | −0.95 | −0.85 | −1.20 |
| Isoelectric point pH | 4.4 | 4.8 | 5.1 | 5.8(5) | 5.8(5) | 5.8(5) |

TABLE IV

*Electrophoresis of Purified Horse Antibody Showing Degraded Material in Centrifuge (Cf. 3) at pH 7.72*

| Preparation No. | Antibody | Number of components | $u \times 10^5$ | Approximate concentration of degraded material in electrophoresis Per cent of total |
|---|---|---|---|---|
| | *per cent* | | | |
| 9093 B | 86 | 2 | −1.6<br>−0.3 | 46 |
| 902 K | 49.4 | 2 | −1.7<br>−0.7(5) | 22 |
| Sn B | 30 | 2 | −1.8(5)<br>−0.9(5) | |

and rabbit antibodies were the same as that observed in serum at the same pH within experimental error.

As noted in Table I, pig W serum showed only four components at pH 7.72. From Table III and Fig. 3 it will be seen that at pH 7.72, the mobility of pig antibody is practically the same as that of $\gamma$ component and hence they would not be separated at this pH. However, since the isoelectric point of pig antibody is much lower than that of $\gamma$ component, a separation was possible by running the pig W serum at pH 5.86 (the isoelectric point of $\gamma$ component) and five components were actually observed at this pH.

The production of inhomogeneous degraded components observed in the centrifuge (3) in several horse sera is paralleled in electrophoresis by the production of a new component of lower mobility than the heavy antibody. Table IV gives the mobility values and concentration of the components. 902 K is an antibody solution prepared from a bleeding from horse 902 a year after the bleeding from which 902 E (Table III) was made (*cf.* 3). This component seems to be present in a concentration about that shown in the ultracentrifuge.

## DISCUSSION

The results now obtainable with the new electrophoresis technique (8) make it possible to obtain a more quantitative picture of the relationships of immune substances to the serum globulin and more detailed information about the electrochemical properties of immune sera and purified antibodies. Table I indicates that the mobilities of the various normal components in sera from several animal species are approximately the same. The immune horse serum, however, showed the presence of a new component, all of which seemed to be antibody, since it was absent on electrophoresis of the same serum after absorbtion of the antibody as shown in the photograph, Fig. 2. In the ultracentrifuge this serum also showed a heavy component which was the bearer of antibody activity (3). Various species of antibodies showing a heavy component in the ultracentrifuge, however, may not have a mobility sufficiently different from one of the other components at any definite pH to be observed as a distinct component in electrophoresis. Thus pig serum W containing 9.7 per cent of a heavy component in the ultracentrifuge showed only four components at pH 7.72. The purified antibody from this serum had a mobility so close to that of the $\gamma$ component as to make effective

separation at this pH impossible. By selecting a pH at which the difference was more pronounced, the pig serum could be shown to contain five components.

The other group of sera in which the antibody was found to have the same molecular weight as the $\gamma$ globulin fraction showed no new component. A comparison of the mobility pH curves (Fig. 3) shows no differences between rabbit antibody solutions prepared by salt dissociation (456$_2$ A), barium dissociation (456$_1$ B), or electrophoretic methods (431-5 $\gamma$). The rabbit and monkey antisera, because of the wide separation of the peaks, were easily adapted to quantitative measurements. Table II and Fig. 1 give a comparison of the change in the electrophoresis scale diagram due to the removal of antibody. The values for percentage of antibody obtained by analysis for antibody nitrogen and total nitrogen (10, 11) (Table II, column 3) and those obtained by integration of the electrophoresis diagrams of the unabsorbed and absorbed sera (last column Table II) are well within experimental error and indicate that all of the antibody is contained in the $\gamma$ globulin fraction. It will be seen that both sides of the U tube are in good agreement for percentage of antibody, but that the total area on the negative side is always greater than that on the positive side. This is perhaps due to the so called $\delta$ boundary (1) observed in most of these sera. The $\delta$ boundary is now thought to be a shift in total protein concentration due to a discontinuity in buffer concentration and conductivity formed near the starting position of the protein boundary in concentrated solutions (13). It has not been included in the curves. The agreement between the two sides in itself indicates that the occurrence of a $\delta$ boundary does not alter the relative concentration of the other components.

In those antisera in which the antibody occurs in the slowest migrating fraction, the electrophoresis method could be used to isolate fairly pure antibody. It is of great importance to select antisera in which a very high percentage of the $\gamma$ component is antibody. On immunization there occurs an increase in total serum protein, as well as in total globulin and antibody (*cf.* for example 14); therefore determination of the amount of antibody nitrogen would not be a sufficient guide to purification by the electrophoresis method. The agreement between the percentage of antibody in the $\gamma$ globulin

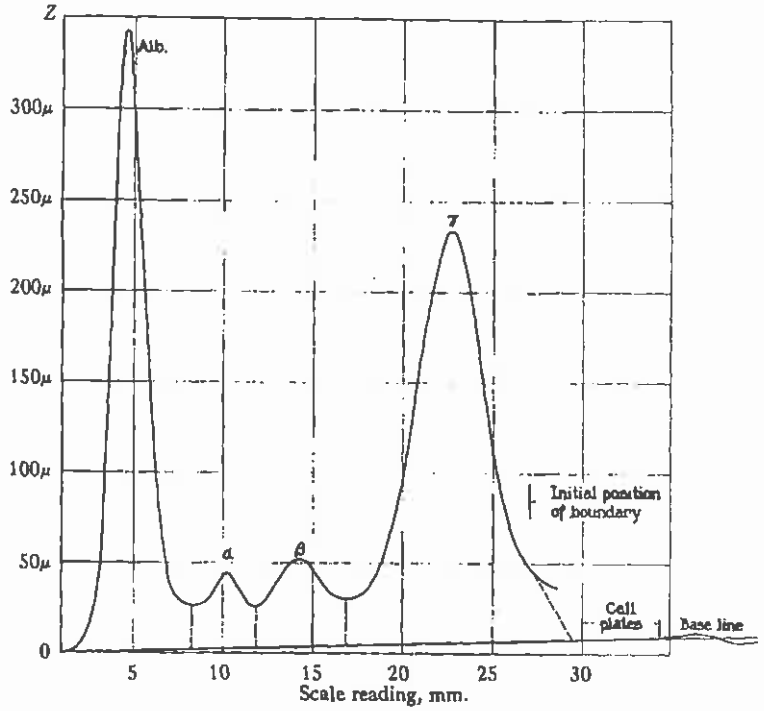



FIG. 1 *a*

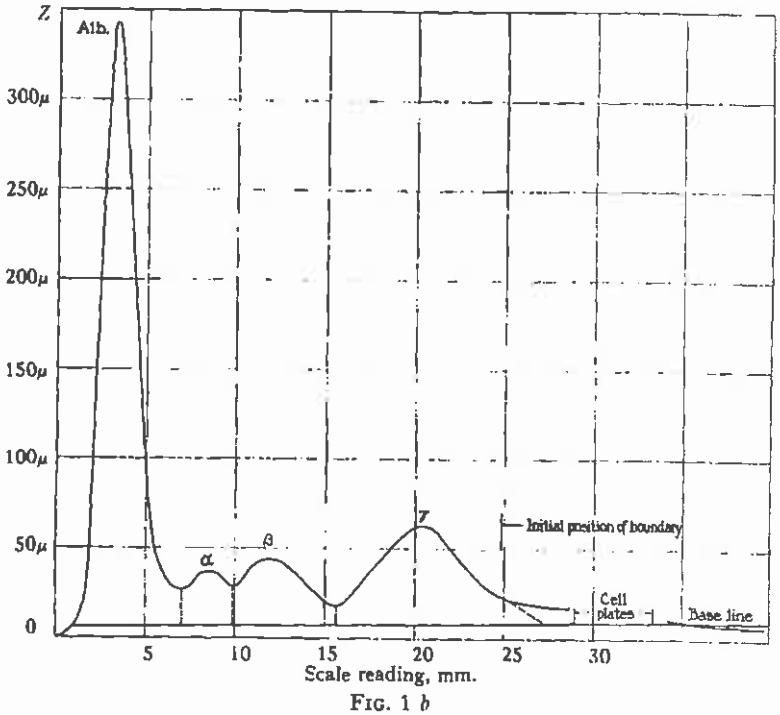



FIG. 1 *b*

FIG. 1. Electrophoresis scale diagram of anti-egg albumin rabbit serum 431-5 1:4 before (*a*) and after (*b*) absorption of the antibody.



15

fraction of 431-5 (Table II) and that found by analysis after electrophoretic purification suggests the advisability of a preliminary run by the electrophoresis method as an indication of the suitability of the method for purification of the antibody from a particular antiserum. A test bleeding from the animal could be taken and the four times diluted serum run in the electrophoresis apparatus before and after removal of the antibody. From the scale photograph (Fig. 1), the percentage of antibody in the $\gamma$ component could be

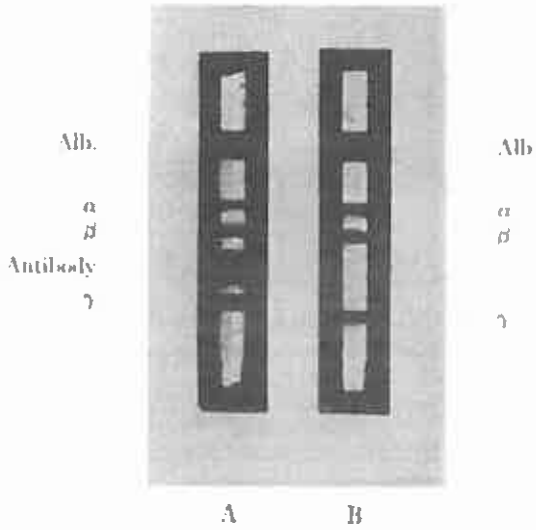



FIG. 2. Electrophoretic *Schlieren* photograph of unabsorbed (*a*) and absorbed (*b*) antipneumococcus horse serum.

calculated and the advisability of further immunization or of immediate purification of the $\gamma$ component decided. For this preliminary test, at most 5 ml. of serum would be required for the ordinary size electrophoresis apparatus (8) but with a new cell of the same length but of smaller cross-section area now being built, the required volume of serum would be reduced to about 2 ml. The method should be most useful for obtaining purified antibodies to protein antigens since the salt dissociation methods are not applicable. In those cases in which the antibody is not in the slowest component the method is too

laborious (such as horse antibody). The purified $\gamma$ component is usually isolated in lower concentration than in the original serum due to the lack of absolute homogeneity of the globulin components (1). The advantage of the electrophoretic method is that it is a very mild method and permits isolation of proteins which retain their native state.

The isoelectric points of the different antibody preparations and the pH mobility curves are included in Table III and Fig. 3. The fact

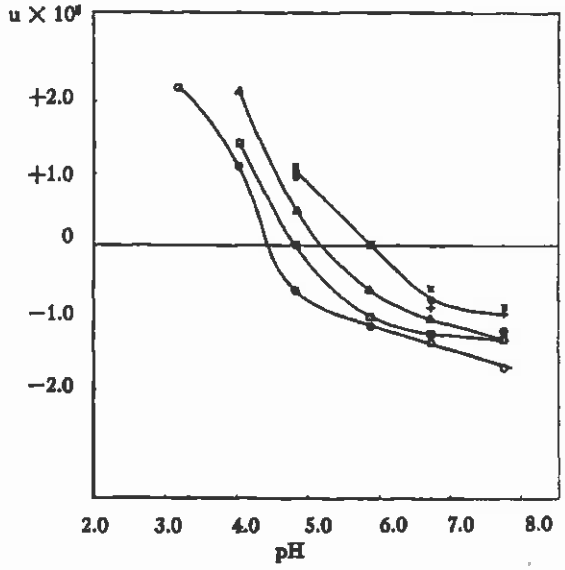



FIG. 3. Mobility of purified antibodies at different pH.
○, horse. □, cow. △, pig.
Rabbit 456₂ A +; 456₂ B ×; 431-5 ⊕.

that these preparations showed only one *Schlieren* band does not indicate absolute homogeneity but merely that the purified preparations were no less uniform than any of the components observed in serum. The value for the isoelectric point of horse antibody is even somewhat lower than that given by Tiselius (4). Both these values are definitely lower than those given by Felton (15) and by Chow and Goebel (16). The methods of purification used by both of these

workers were non-specific precipitation methods. As has been noted in the experimental section, the inability to integrate the scale diagrams in concentrated horse sera because of the poor separation of the globulin peaks may possibly be an indication of loose compound formation in these globulins and that the concentration methods of isoelectric precipitation used by these authors (15, 16) resulted in the precipitation of an addition compound of higher isoelectric point (*cf.* also Kendall, 12).

Additional evidence that the washed antigen-antibody precipitates (*cf.* 10, 11) are free from non-specific protein was obtained by dissolving a washed specific precipitate of egg albumin-anti-egg albumin in excess antigen and running in the electrophoresis apparatus. Two components were observed, one of excess antigen and another of an antigen-antibody compound. There was no component corresponding to ordinary $\gamma$ globulin, indicating the absence of the corresponding non-specific $\gamma$ component in the original washed precipitate. A detailed study of the inhibition zone in electrophoresis has been completed.

The data concerning the degradation of horse antibody in the animal on continued immunization (3) indicate that it is also accompanied by a definite change in mobility as noted by the presence of two antibody components in a later bleeding from an animal which had previously shown one component (Table IV). Thus $909_1$ B showed two components in electrophoresis, both in appreciable concentration and a comparison of the original serum absorbed and unabsorbed showed that some of the slower $\gamma$ component was removed on absorption of the antibody. Similarly in horse antibody 902 E (Table III) prepared from an earlier bleeding, only one component was observed in electrophoresis and in the centrifuge (3), while in 902 K (Table IV), prepared in the same way from a bleeding taken after another year of immunization, degraded material was present in the centrifuge and a new component of lower mobility was present in electrophoresis. This is probably a manifestation of a process of degradation and removal of antibody protein in the horse after prolonged immunization.

## SUMMARY

1. Antibody produced in the horse migrates as a new serum component between the $\beta$ and $\gamma$ components, whereas rabbit antibody

is electrophoretically identical with the $\gamma$ globulin component of the serum.

2. In rabbit and monkey antisera the percentage of antibody in the serum and in the $\gamma$ globulin fraction can be determined by integration of the electrophoresis diagrams of unabsorbed and absorbed sera. Antibody solutions of high purity can be obtained by electrophoretic isolation of the $\gamma$ globulin of rabbit antisera in which the percentage of antibody to total $\gamma$ globulin is high.

3. The isoelectric points of pig, cow, horse, and rabbit antibodies have been determined.

4. In horse sera prolonged immunization is accompanied by the formation of another antibody component of lower mobility.

We wish to thank Professor The Svedberg for his interest in this work and Dr. Michael Heidelberger for his valuable suggestions and assistance in supplying the various antisera and antibody preparations.

## BIBLIOGRAPHY

1. Tiselius, A., *Biochem. J.*, 1937, **31**, 1464.
2. Heidelberger, M., and Pedersen, K. O., *J. Exp. Med.*, 1937, **65**, 393.
3. Kabat, E. A., and Pedersen, K. O., *Science*, 1938, **87**, 372; Kabat, E. A., *J. Exp. Med.*, 1939, **69**, 103.
4. Tiselius, A., *J. Exp. Med.*, 1937, **65**, 641.
5. Heidelberger, M., and Kendall, F. E., *J. Exp. Med.*, 1933, **57**, 373. Heidelberger, M., Kendall, F. E., and Scherp, H. W., *J. Exp. Med.*, 1936, **64**, 559.
6. Heidelberger, M., and Kendall, F. E., *J. Exp. Med.*, 1936, **64**, 161. Heidelberger, M., and Kabat, E. A., *J. Exp. Med.*, 1938, **67**, 181.
7. Lamm, O., *Nova Acta Regiae Soc. Scient. Upsaliensis, IV*, 1937, **10**, 6; *Z. physikal. Chem., Abt. A*, 1928, **138**, 313; *Abt. A*, 1929, **143**, 177.
8. Tiselius, A., *Tr. Faraday Soc.*, 1937, **33**, 524.
9. Tiselius, A., Pedersen, K. O., and Eriksson-Quensel, I.-B., *Nature*, 1937, **139**, 546.
10. Heidelberger, M., Sia, R. H. P., and Kendall, F. E., *J. Exp. Med.*, 1930, **52**, 477. Heidelberger, M., Kendall, F. E., and Soo Hoo, C. M., *J. Exp. Med.*, 1933, **58**, 137. Heidelberger, M., and Kendall, F. E., *J. Exp. Med.*, 1935, **61**, 359.
11. Heidelberger, M., and Kabat, E. A., *Proc. Soc. Exp. Biol. and Med.*, 1934, **31**, 595; *J. Exp. Med.*, 1934, **60**, 643.
12. Kendall, F. E., *J. Clin. Inv.*, 1937, **16**, 921.
13. Tiselius, A., and Svensson, H., to be published.
14. Boyd, W. C., and Bernard, H., *J. Immunol.*, 1937, **33**, 111.
15. Felton, L. D., *J. Immunol.*, 1931, **21**, 241, and earlier papers.
16. Chow, B. F., and Goebel, W. F., *J. Exp. Med.*, 1935, **62**, 179.

# Intravenous immunoglobulin

From Wikipedia, the free encyclopedia

**Intravenous immunoglobulin** (IVIG) is a blood product administered intravenously. It contains the pooled IgG (immunoglobulin (antibody) G) extracted from the plasma of over one thousand blood donors. IVIG's effects last between 2 weeks and 3 months. It is mainly used as treatment in three major categories:

- Immune deficiencies such as X-linked agammaglobulinemia, hypogammaglobulinemia (primary immune deficiencies), and acquired compromised immunity conditions (secondary immune deficiencies) featuring low antibody levels.
- Autoimmune diseases e.g. Immune thrombocytopenia ITP and Inflammatory diseases e.g. Kawasaki disease.
- Acute infections.

## Contents


- 1 IVIG use
- 2 Mechanism of action
- 3 IVIG notes
- 4 IVIG dose
  - 4.1 FDA-approved indications
  - 4.2 In Phase III testing in the US (as of December 2008)
  - 4.3 Off-label uses
- 5 Complications and side effects
- 6 References


## IVIG use

IVIG is given as a plasma protein replacement therapy (IgG) for immune deficient patients who have decreased or abolished antibody production capabilities. In these immune deficient patients, IVIG is administered to maintain adequate antibody levels to prevent infections and confers a passive immunity. Treatment is given every 3–4 weeks. In the case of patients with autoimmune disease, IVIG is administered at a high dose (generally 1-2 grams IVIG per kg body weight) to attempt to decrease the severity of the autoimmune diseases such as Dermatomyositis. Currently, IVIg is being increasingly used off-label in a number of pathological conditions; the increasing world-wide usage of IVIg may lead to shortages of this beneficial drug.

IVIG is useful in some acute infection cases such as pediatric HIV infection and Guillain–Barré syndrome.[1]

## Mechanism of action

The precise mechanism by which IVIG suppresses harmful inflammation has not been definitively

established but is believed to involve the inhibitory Fc receptor.[2][3] The actual primary target(s) of IVIG in autoimmune disease are still unclear, however. IVIG may work via a multi-step model where the injected IVIG first forms a type of immune complex in the patient.[4] Once these immune complexes are formed, they interact with activating Fc receptors on dendritic cells[5] which then mediate anti-inflammatory effects helping to reduce the severity of the autoimmune disease or inflammatory state.

Additionally, the donor antibody may bind directly with the abnormal host antibody, stimulating its removal. Alternatively, the massive quantity of antibody may stimulate the host's complement system, leading to enhanced removal of all antibodies, including the harmful ones. IVIG also blocks the antibody receptors on immune cells (macrophages), leading to decreased damage by these cells, or regulation of macrophage phagocytosis.

IVIG may also regulate the immune response by reacting with a number of membrane receptors on T cells, B cells, and monocytes that are pertinent to autoreactivity and induction of tolerance to self.[6]

A recent report stated that IVIG application to activated T cells leads to their decreased ability to engage microglia. As a result of IVIG treatment of T cells, the findings showed reduced levels of tumor necrosis factor-alpha and interleukin-10 in T cell-microglia co-culture. The results add to the understanding of how IVIG may affect inflammation of the central nervous system in autoimmune inflammatory diseases.[7]

## IVIG notes

- IVIG is an infusion of IgG antibodies only. Therefore, peripheral tissues that are defended mainly by IgA antibodies, such as the eyes, lungs, gut and urinary tract are not fully protected by the IVIG treatment.
- XLA patients are immune to the most virulent adverse effect, anaphylactic shock, as they do not have the antibodies to react against the treatment. Anaphylactic shock has a higher chance to occur in IgA deficient patients which do have other antibody types.
- In case of recurring side effects, it is recommended to slow the pace of the IVIG administration and to reduce the dosage. It is also advisable to change IVIG brand, as some people react against a specific brand.
- If the patient is diabetic, he should take into consideration the medium in which the antibodies are solubilized in the IVIG treatment, as some brand solubilize antibodies with high concentrated sugars (such as sucrose and maltose).
- U.S. Food and Drug Administration (FDA) guidelines for IVIG state the product should be:
  - Prepared out of at least 1,000 different human donors.
  - All four IgG subgroups (1-4) should be present.
  - The IgG should maintain biological activity and lifetime of at least 21 days.
  - Does not contain samples which are HIV, hepatitis B, hepatitis C positive.
  - Screened and treated in a manner that destroys viruses.
- IVIG is also considered a modulator of the immune system and was shown to be beneficial in treating numerous autoimmune diseases such as relapsing and remitting multiple sclerosis (MS), myasthenia gravis, pemphigus, polymyositis (PM), dermatomyositis (DM), Wegener's granulomatosis (WG), Churg-Strauss syndrome, chronic inflammatory demyelinating polyneuropathy (CIDP) and more.
- IVIG can be given to pregnant women.

21

- IVIG is also used as a treatment for unexplained recurring miscarriages. The effectiveness of the therapy is controversial.
- IVIG cost is climbing and well over $50/g. ($10,000 for a 100 kg (220 lbs) person at 2g/kg)

# IVIG dose

Dosage of IVIG is dependent on indication.

For primary immune dysfunction 100 to 400 mg/kg of body weight every 3 to 4 weeks is implemented.

For neurological and autoimmune diseases 2 grams per kilogram of body weight is implemented for three to six months over a five day course once a month. Then maintenance therapy of 100 to 400 mg/kg of body weight every 3 to 4 weeks follows.

## FDA-approved indications

- Allogeneic bone marrow transplant
- Chronic lymphocytic leukemia
- Idiopathic thrombocytopenic purpura
- Pediatric HIV
- Primary immunodeficiencies
- Kawasaki disease
- Chronic inflammatory demyelinating polyneuropathy (CIDP). Only the "Gamunex" brand manufactured by Talecris is approved for CIDP (in 2008), under the U.S. Orphan Drug law provisions
- Kidney transplant with a high antibody recipient or with an ABO incompatible donor

In 2004 the FDA approved the Cedars-Sinai IVIG Protocol which has been 90-95% successful in removing antibodies from the blood of kidney transplant recipients so that they can accept a living donor kidney from any healthy donor no matter blood type (ABO incompatible) or tissue match.

## In Phase III testing in the US (as of December 2008)

- Alzheimer's Disease

## Off-label uses

- Autism
- Chronic fatigue syndrome
- Clostridium difficile colitis
- Dermatomyositis and polymyositis
- Graves' ophthalmopathy
- Guillain-Barré syndrome
- Muscular Dystrophy
- Inclusion body myositis
- Lambert-Eaton syndrome
- Lupus erythematosus

- Multifocal motor neuropathy
- Multiple sclerosis
- Myasthenia gravis
- Neonatal alloimmune thrombocytopenia
- Parvovirus B19
- Pemphigus
- Post-transfusion purpura
- Renal transplant rejection
- Spontaneous Abortion/Miscarriage
- Stiff person syndrome
- Opsoclonus Myoclonus
- Severe sepsis and septic shock in critically ill adults[8]
- Toxic epidermal necrolysis
- In chronic lymphocytic leukemia and multiple myeloma, as well as various rare deficiencies of immunoglobulin synthesis (e.g. X-linked agammaglobulinemia, hypogammaglobulinemia), IVIG is administered to maintain adequate immunoglobulin levels to prevent infections.

## Complications and side effects

Although routine use of IVIG is common practice, sometimes for long term treatments, and is considered safe, complications of IVIG therapy are known and include:

- headache
- dermatitis - usually peeling of the skin of the palms and soles.
- infection (such as HIV or viral hepatitis) by contaminated blood product; there is also an as yet unknown risk of contracting variant CJD (vCJD) however the process whereby the product is extracted shows that the contaminants are usually not present in the product.
- pulmonary edema from fluid overload, due to the high colloid oncotic pressure of IVIG
- allergic/anaphylactic reactions; for example, anaphylactic shock, especially in IgA deficient patients, who by definition can still produce IgG antibodies (IgA deficient patients are more likely to produce IgG against the IVIG administration than normal patients).
- damage such as hepatitis caused directly by antibodies contained in the pooled IVIG.
- acute renal failure[9]
- venous thrombosis
- aseptic meningitis

## References

1. ^ Meythaler, R.G. Miller, J.T. Sladky and J.C. Stevens, R.A.C. Hughes, E.F.M. Wijdicks, R. Barohn, E. Benson, D.R. Cornblath, A. F. Hahn, J.M., "Practice parameter: Immunotherapy for Guillain–Barré syndrome: Quality Standards Subcommittee of the American Academy of Neurology", Neurology 2003;61;736-740. Download from http://www.neurology.org /cgi/reprint/61/6/736.pdf.
2. ^ Gern JE (August 2002). "Antiinflammatory Activity of IVIG Mediated through the Inhibitory FC Receptor" (http://pediatrics.aappublications.org /cgi/content/full/110/2/S1/467-b) . *Pediatrics* **110** (2): 467–8. doi:10.1542/peds.110.2.S1.467-b (http://dx.doi.org /10.1542%2Fpeds.110.2.S1.467-b) .

23

http://pediatrics.aappublications.org/cgi/content/full/110/2/S1/467-b.

3. ^ Nimmerjahn F, Ravetch JV (January 2007). "The antiinflammatory activity of IgG: the intravenous IgG paradox". *J. Exp. Med.* **204** (1): 11–5. doi:10.1084/jem.20061788 (http://dx.doi.org/10.1084%2Fjem.20061788) . PMC 2118416 (http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=2118416) . PMID 17227911 (http://www.ncbi.nlm.nih.gov/pubmed/17227911) .
4. ^ Clynes R (January 2005). "Immune complexes as therapy for autoimmunity". *J. Clin. Invest.* **115** (1): 25–7. doi:10.1172/JCI23994 (http://dx.doi.org/10.1172%2FJCI23994) . PMC 539209 (http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=539209) . PMID 15630438 (http://www.ncbi.nlm.nih.gov/pubmed/15630438) .
5. ^ Siragam V, Crow AR, Brinc D, Song S, Freedman J, Lazarus AH (June 2006). "Intravenous immunoglobulin ameliorates ITP via activating Fc gamma receptors on dendritic cells". *Nat. Med.* **12** (6): 688–92. doi:10.1038/nm1416 (http://dx.doi.org/10.1038%2Fnm1416) . PMID 16715090 (http://www.ncbi.nlm.nih.gov/pubmed/16715090) .
6. ^ Bayry J, Thirion M, Misra N, *et al.* (October 2003). "Mechanisms of action of intravenous immunoglobulin in autoimmune and inflammatory diseases". *Neurol. Sci.* **24 Suppl 4**: S217–21. doi:10.1007/s10072-003-0081-7 (http://dx.doi.org/10.1007%2Fs10072-003-0081-7) . PMID 14598046 (http://www.ncbi.nlm.nih.gov/pubmed/14598046) .
7. ^ Janke AD, Yong VW (April 2006). "Impact of IVIg on the interaction between activated T cells and microglia". *Neurol. Res.* **28** (3): 270–4. doi:10.1179/016164106X98143 (http://dx.doi.org/10.1179%2F016164106X98143) . PMID 16687052 (http://www.ncbi.nlm.nih.gov/pubmed/16687052) .
8. ^ Laupland KB, Kirkpatrick AW, Delaney A (December 2007). "Polyclonal intravenous immunoglobulin for the treatment of severe sepsis and septic shock in critically ill adults: a systematic review and meta-analysis". *Crit. Care Med.* **35** (12): 2686–92. doi:10.1097/01.CCM.0000295312.13466.1C (http://dx.doi.org/10.1097%2F01.CCM.0000295312.13466.1C) . PMID 18074465 (http://www.ncbi.nlm.nih.gov/pubmed/18074465) .
9. ^ Aaron Graumann, MS-IV And Edward T. Zawada Jr, MD (March 2010). "Case Report: Acute Renal Failure After Administering Intravenous Immunoglobulin" (http://www.postgradmed.com/index.php?article=2131#) . *Postgraduate Medicine.* **122** (2): 142–147. doi:10.3810/pgm.2010.03.2131 (http://dx.doi.org/10.3810%2Fpgm.2010.03.2131) . PMID 20203465 (http://www.ncbi.nlm.nih.gov/pubmed/20203465) . http://www.postgradmed.com/index.php?article=2131#.

Retrieved from "http://en.wikipedia.org/wiki/Intravenous_immunoglobulin"
Categories: Glycoproteins | Antibodies | Immune system | Medical treatments

- This page was last modified on 21 May 2011 at 15:14.
- 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.

**Bence Jones proteins and multiple myeloma**

In order to survive various viral and bacterial infections, every human needs to produce thousands or even millions of different antibodies to selectively bind such foreign invaders or antigens. As a result, it is essentially impossible to purify one specific antibody from such a mixture in the human serum, so that biochemists can analyze its properties in detail.

Even earlier during the 19th century, in 1848, another physician Bence Jones received samples of urine from his two colleagues, Watson and MacIntyre, treating a patient suffering from mollities ossium (now known as multiple myeloma), a cancer difficult to treat. He noticed that on heating the urine from such patients, a new precipitate appeared (Bence Jones H (1848) On a new substance occurring in the urine of a patient with mollities ossium. Phil Trans R Soc Lond, 138, 55-62). After thorough chemical analysis, Bence Jones concluded that the new substance was a protein. He also noticed that protein was soluble in boiling water. The new substance was subsequently named after him as Bence Jones protein.

The origin of that protein was unknown during the 19th century. As a result, even though the protein was pure, no biochemist bothered to analyze it further. They never sequenced it, except probably realized that such proteins from different patients were not the same.

Subsequent electrophoretic studies of the serum gamma globulin fraction from these patients exhibited in most cases a large sharp peak located at different positions in different patients. That unknown protein was designated as the myeloma protein with fairly large molecular weight. Again they were essentially pure. Due to their unknown functions, biochemists have left such proteins alone without further studies. Instead, Bence Jones proteins and myeloma proteins were basically used as a diagnostic test for multiple myeloma.

It took many more years to discover the connection between Bence Jones protein and myeloma protein by a graduate student

PHILOSOPHICAL
TRANSACTIONS:

# On a New Substance Occurring in the Urine of a Patient with Mollities Ossium

Henry Bence Jones

*Phil. Trans. R. Soc. Lond.* 1848 **138**, 55-62
doi: 10.1098/rstl.1848.0003

**References** **Article cited in:**
http://rstl.royalsocietypublishing.org/content/138/55.citation#related-urls

**Email alerting service** Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. Lond.* go to: **http://rstl.royalsocietypublishing.org/subscriptions**

This journal is © 1848 The Royal Society

## III. *On a new substance occurring in the Urine of a patient with Mollities Ossium.*

*By* Henry Bence Jones, *M.A., F.R.S., Physician to St. George's Hospital.*

Received February 25,—Read April 22, 1847.

On the 1st of November 1845 I received from Dr. Watson the following note, with a test tube containing a thick, yellow, semi-solid substance:—"The tube contains urine of very high specific gravity; when boiled it becomes highly opake; on the addition of nitric acid it effervesces, assumes a reddish hue, becomes quite clear, but, as it cools, assumes the consistence and appearance which you see: heat reliquifies it. What is it?"

A few hours afterwards a specimen of the same urine, passed by a grocer forty-seven years of age, who had been out of health for thirteen months, was sent to me by Dr. MacIntyre. He being in attendance on the case with Dr. Watson, had two days previously first observed the peculiar reactions of the urine.

The specimen of urine was slightly acid; specific gravity 1034·2; it contained a sediment consisting of crystalline phosphate of lime, oxalate of lime, and cylinders of fibrin. The urine became thick with heat from a deposit of phosphates, but cleared with a drop of acid. It gave no precipitate with an excess of nitric acid, unless left to stand, or unless heated and left to cool, when it became solid. This solid redissolved by heat, and again formed on cooling. Continued boiling with strong nitric acid evolved but little gas, and did not quickly hinder this reaction. Hydrochloric acid gave the same solid precipitate, soluble by heat. Strong acetic acid gave only a slight precipitate, which redissolved by heat. Caustic potash and sulphate of copper gave a splendid bright blue, clear liquid, passing over when heated to claret colour.

516·84 grains evaporated to dryness *in vacuo* over sulphuric acid, gave 48·37 grains solid residue =93·58 per 1000 urine.

November 3rd.—I received more urine from the same patient. It gave a greater sediment, consisting of urate of ammonia, and some amorphous phosphate of lime, and some coarse cylinders. Filtered, specific gravity =1043·2. The reactions were the same as before.

521·39 grains gave 65·85 grains solid residue =126·30 per 1000 urine.

All the phosphoric acid was precipitated in combination with lime by adding chloride of calcium and ammonia. Phosphate of lime =5·68 per 1000 urine.

November 7th.—Water from the same patient contained some sediment of crystalline phosphate, some granular and laminar phosphate of lime: some coarse cylinders



The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions of the Royal Society of London.* www.jstor.org

of fibrin, very coarse. The urine was slightly alkaline from ammonia; it gave a plentiful precipitate with alcohol; coagulated firmly with heat, very perfectly with a drop or two of acetic acid. Ferro-prussiate of potash gave no immediate precipitate; in less than half an hour a considerable precipitate had formed, which was soluble in liq. potassæ.

November 8th.—I saw multitudes of phosphate of lime crystals: no cylinders: few octohedral crystals. The urine coagulated with heat even when rendered feebly alkaline by liq. potassæ or liq. ammoniæ: if these were in excess coagulation by heat did not take place.

November 9th.—The urine was loaded with urates. Not nearly so readily coagulable by heat as it had been the two previous days; it bore today brisk and prolonged boiling: specific gravity =1037·2.

November 15th.—I saw the patient. He stated that he usually passed about 35℥ of urine daily.

November 18th.—The urine had a specific gravity = 1039·6 and was acid. It contained much urate of ammonia, phosphate of lime, and oxalate of lime; and bore long-continued boiling without coagulating. The urine was filtered and the following analysis was made:—

519·54 grains evaporated to dryness *in vacuo* over sulphuric acid, gave,—

56·78 grains of solid residue =109·28 per 1000 urine.

Ash =6·02 grains =11·58 per 1000 urine.

This ash dissolved in water, filtered, precipitated by nitrate of silver and nitric acid, gave—

Fused chloride of silver =48·84 grains=1·99 grain ch. sod. =3·83 per 1000 urine.
Sulphate of baryta =1·47 grain =1·10 grain sulphate of potash =2·11 per 1000 urine.

519·62 grains of urine precipitated by alcohol—

Precipitate =37·16 grains =71·51 per 1000 urine.

29·26 grains of this precipitate gave,—

1·47 grain ash (=5·01 per cent.) =3·58 per 1000 urine.

519·62 grains treated with alcohol—

Dissolved =19·72 grains=37·93 per 1000.

Ash . . = 4·17 grains = 8·03 per 1000.

1024·48 grains precipitated by strong acetic acid—

Uric acid =·99 grain =·96 per 1000 urine.

Filtered fluid precipitated by ammonia,—

Earthy phosphate =1·23 grain =1·20 per 1000.

Hence,

| | |
|---|---|
| Water . . . . . | =890·72 per 1000 urine. |
| Total solid residue . . . . | =109·28 |
| Total ash . . . . . | = 11·58 |

Substance soluble in alcohol . . . . . =37·93 per 1000 urine.
Ash . . . . . = 8·03
Whence urea and alcoholic extract . =29·90

Substance insoluble in alcohol . . . . =71·51 per 1000 urine.
Ash . . . . . = 3·58
Uric acid . . . . . = ·96
Whence new substance . . . . . =66·97

Earthy phosphate . . . . . = 1·20 per 1000 urine.
Chloride of sodium . . . . . = 3·83
Sulphate of potash . . . . . = 2·10
Thence phosphate of soda . . . . = 4·45, because total ash=11·58.

Hence,

| | |
|---|---|
| Water . . . . . | =890·72 |
| New substance . . | = 66·97 |
| Urea . . . . . | = 29·90 |
| Uric acid . . . . | = ·96 |
| Earthy phosphate . | = 1·20 |
| Chloride of sodium | = 3·83 |
| Sulphate of potash | = 2·10 |
| Alkaline phosphate | = 4·45 |
| | 1000·13 |

January 6th.—I was given by Dr. MacIntyre the water passed on the 27th, 29th and 30th of December, all strongly acid, containing much urate of ammonia, oxalate of lime and phosphate of lime.

In the water of December the 27th, I saw some dead spermatozoa. The urine when filtered had a specific gravity of 1031·3.

515·37 grains precipitated by chloride of calcium and ammonia. Phosphate of lime =4·15 grains =8·05 per 1000 urine.

1018·85 grains, precipitated by ammonia. Earthy phosphates =1·47 grain =1·44 per 1000.

In that of December 29th, the specific gravity when filtered was =1037·9.

518·69 grains, precipitated as above. Phosphate of lime=4·27 grains=8·24 per 1000.

1027·71 grains, precipitated as above. Earthy phosphate =1·62 grain =1·57 per 1000.

In that of December 30th, specific gravity =1042·7, filtered. No fibrinous cylinders. The urine remained acid for a month in an open vessel.

521·05 grains, precipitated as above. Phosphate of lime =6·17 grains =11·85 per 1000.

717·04 grains, precipitated by ammonia. Earthy phosphate =1·23 grain =1·72 per 1000.



29

January 2nd.—The patient died. The following day I saw that the bony structure of the ribs was cut with the greatest ease, and that the bodies of the vertebræ were capable of being sliced off with the knife*. For an account of the structure of the bone, see a paper by Mr. Dalrymple in the third number of the Dublin Journal, August 1846.

The variation in the coagulability by heat observed in these different examinations of the urine might possibly have arisen from the variation in the acidity of the secretion. Free acid might have hindered the coagulation. The only other supposition is that the composition of this new substance varied slightly at different times, and, on the whole, this seems not unlikely to have been the case.

The variations in the amount of earthy and alkaline phosphates in this case may be arranged in the following Table:—

| | | Earthy phosphates. | Total phosphates. |
|---|---|---|---|
| 1st analysis. | Urine, specific gravity =1043·2 | | 5·68 per 1000 urine. |
| 2nd analysis. | Urine, specific gravity =1039·6 | 1·20 | 5·65 per 1000 urine. |
| 3rd analysis. | Urine, specific gravity =1031·3 | 1·43 | 8·05 per 1000 urine. |
| 4th analysis. | Urine, specific gravity =1037·9 | 1·57 | 8·24 per 1000 urine. |
| 5th analysis. | Urine, specific gravity =1042·7 | 1·72 | 11·85 per 1000 urine. |

*Reactions of this new substance precipitated from the Urine by Alcohol, well-washed, dried and powdered.*

The substance slowly but entirely dissolved when thrown into cold water.

It was much more readily dissolved by boiling water; if the boiling was continued, at first no appearance of coagulation was evident, but, after ten minutes, in repeated experiments a gelatinous coagulation took place. By continuing the boiling, adding small quantities of water from time to time, the coagulum in an hour entirely redissolved.

If water was poured on the powdered substance it sometimes caked together, and on boiling became horny, and resisted the action even of long-continued heat.

The substance was soluble in caustic potash, and, after long standing, at the ordinary temperature, or at 140°. When the alkaline solution was neutralized by acetic acid, no precipitate formed, but if an excess of acid was added, a plentiful precipitate fell, which was insoluble in more acetic acid, but was readily soluble by heat. If a great excess of acetic acid was added the precipitate was dissolved.

The solution of this substance in water gave an immediate precipitate with nitric acid, which entirely and readily dissolved when heated. Boiling caused no precipitation. On cooling, the precipitate again was formed. If the watery solution was acidulated with acetic acid, an immediate white precipitate fell on the addition of ferro-prussiate of potash. The precipitate was soluble in caustic potash.

* The pericardium contained an ounce or two of fluid, which gave a precipitate immediately by heat and acid. The coagulum did not redissolve when heated. The fluid was feebly alkaline; specific gravity 1015·3. Heat and acid coagulated the blood, and on heating the acid coagulum it did not redissolve. The kidneys were healthy to the naked eye and to the microscope.

The watery solution gave a precipitate with sulphate of copper and with bichloride of mercury; each precipitate readily dissolved when acetic acid was added. The precipitate with sulphate of copper was soluble in caustic potash, forming a splendid deep blue solution, which became more claret-coloured when heated.

Strong hydrochloric acid dissolved the substance, giving a splendid purple blue solution.

Dissolved in caustic potash and boiled, a deep inky blackness was produced by dropping acetate of lead into the solution.

*On the Analysis of this Substance.*

The substance was prepared in the following way. The urine was mixed with a large excess of alcohol. The precipitate which formed was thrown on a filter, and well-washed with cold alcohol until it passed through, leaving no stain when evaporated on platinum.

It was then dried *in vacuo*, finely powdered, and treated with æther at the boiling temperature, until no traces of fatty matter were left on evaporation of the æther. The substance was long dried *in vacuo* over sulphuric acid.

6·72 grains of the substance so prepared gave 0·19 ash =2·8 per cent.

Of second preparation 4·59 grains of the substance gave 0·13 ash =2·9 per cent.

4·10 grains, minus ash, burnt with chromate of lead, gave 7·77 carbonic acid and 2·64 water.

C =51·68
H= 7·15

4·12 grains, minus ash, burnt as before, gave 7·78 carbonic acid and 2·63 water.

C =51·50
H= 7·09

Of another preparation,

3·57 grains, minus ash, gave 6·84 carbonic acid and 2·24 water.

C =52·25
H= 6·97

4·12 grains, minus ash, dried at 280° Fahr., gave 7·90 carbonic acid and 2·66 water.

C =52·29
H= 7·17

Another preparation,

3·83 grains of the substance, minus ash, gave 7·33 carbonic acid and 2·47 water.

C =52·16
H= 7·16

4·42 grains of the substance, minus ash, gave 8·46 carbonic acid.

C =52·19



31

3·57 grains, minus ash, gave 6·84 carbonic acid and 2·24 water.

C = 52·25
H = 6·97

2·85 grains of the substance of the 1st preparation, minus ash, gave 6·89 ammonio-chloride of platinum.

N = 15·24

4·37 grains of the 2nd preparation, minus ash, gave 10·27 grains ammonio-chloride of platinum,

N = 14·81

Hence,

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| C . . | 51·68 | 51·50 | 52·25 | 52·29 | 52·16 | 52·19 | 52·25 |
| H . . | 7·15 | 7·09 | 6·97 | 7·17 | 7·16 | | 6·97 |
| N . . | 15·24 | | | 14·81 | | | |
| O. | | | | | | | |

5 grains of the substance, mixed with 50 grains of pure nitre, free from chloride and sulphate, heated gently in a platinum crucible, inflamed (some was lost); the residue, dissolved in water acidulated with pure nitric acid, gave no precipitate with nitrate of silver; but gave a very plentiful precipitate with chloride of barium, insoluble in nitric acid, when boiled. After being filtered it gave a cloud with ammonia, soluble in acid, reprecipitated by ammonia.

Previously to determining the sulphur in this substance, I determined the quantity present in some dried albumen from an egg. Treated with hydrochloric acid, it gave scarcely a trace of precipitate with chloride of barium.

9 grains of dry albumen, fused with 45 grains of nitre, and the same quantity of pure carbonate of potash, gave ·83 grain of sulphate of baryta.

= ·11 grain of sulphur.
= 1·27 per cent. sulphur.

9·56 grains of the new substance, fused with 45 grains of nitre, and the same quantity of carbonate of potash, gave 1·01 grain of sulphate of baryta.

= ·139 grain of sulphur.
= 1·45 per cent. sulphur.

Some of the substance, prepared at a different time, was next examined.

20·67 grains, when treated with boiling dilute hydrochloric acid for about half an hour, gave . . . . . ·17 grain of sulphate of baryta.

= ·023 grain of sulphur.
= ·09 per cent. sulphur.

The clear liquid, after separation of the sulphate of baryta, precipitated by ammonia, filtered, washed with boiling water, burnt, redissolved in hydrochloric acid, and reprecipitated by ammonia, gave ·16 grain of phosphate of baryta; redissolved in

hydrochloric acid, precipitated with sulphate of soda = ·059 grain of baryta.
= ·022 grain of phosphorus.
=0·1 per cent. phosphorus.

13·53 grains of the substance, fused with 69 grains of nitre, and the same quantity of carbonate of potash, gave 1·17 grain of sulphate of baryta.

=0·1614 grain of sulphur.
=1·19 per cent. sulphur.

The clear liquid precipitated by ammonia—

Phosphate of baryta . . . . . =·29
Phosphorus . . . . . =·041
Sulphate of baryta . . . . . =·16=·10 baryta
Hence phosphorus . . . . . =·30 per cent.

11·88 grains of the substance, fused with 55 grains of nitre and the same quantity of carbonate of potash, gave ·97 grain of sulphate of baryta.

= ·1338 grain sulphur.
=1·12 per cent. sulphur.

The clear liquid precipitated by ammonia—

Phosphate of baryta . . . . . =·25.
Phosphorus . . . . . =·0349.
Sulphate of baryta . . . . . =·14=·9 baryta.
Hence phosphorus . . . . . =·29 per cent.

There is, then, deducting the sulphur and phosphorus dissolved out by the boiling hydrochloric acid,—

| | 1st. | 2nd. | 3rd. |
|---|---|---|---|
| Sulphur, per cent. . . . . | 1·36 | 1·09 | 1·03 |
| Phosphorus, per cent. . . . | | ·20 | ·19 |

The presence of sulphur and phosphorus in this new substance proves that it is not an oxide of protein. The solubility of this body in water might lead to the comparison of it with the hydrated tritoxide of protein of MULDER, which is also soluble in water; but according to MULDER, this hydrated tritoxide of protein gives no precipitate on the addition of ferro-prussiate of potash, and it contains no sulphur or phosphorus, while this new substance gives both sulphur and phosphorus, and also a precipitate with ferro-prussiate of potash. In these respects there is a most marked and essential difference between these substances. The peculiar reaction of this new body with nitric acid, the solubility of the nitric acid precipitate when heated, I find to be also common to the so-called hydrated tritoxide of protein prepared from the inflammatory crust of the blood; from albumen of the blood by chlorine; and from albumen of the egg. Lastly, this peculiar reaction with nitric acid hinders all possibility of confusing this new substance with albumen. Indeed ordinary albumen may be separated from this new substance by adding nitric acid,

boiling and filtering whilst hot; on cooling, the hydrated oxide will be precipitated from the filtered liquid, and it will again be redissolved by heat, whilst the albumen will remain on the filter.

*Conclusions.*

1st. The ultimate analysis of this substance may be represented by $C_{48}$, $H_{38}$, $N_6$, $O_{18}$, or by $C_{40}$, $H_{31}$, $N_5$, $O_{15}$; according as protein is $=C_{48}$, $H_{37}$, $N_6$, $O_{15}$, or $C_{40}$, $H_{30}$, $N_5$, $O_{12}$.

| | 1st. | 2nd. | 3rd. | 4th. | 5th. | 6th. | 7th. | $C_{48}$, $H_{38}$, $N_6$, $O_{18}$; or $C_{40}$, $H_{31}$, $N_5$, $O_{15}$. Reckoned. | Reckoned. |
|---|---|---|---|---|---|---|---|---|---|
| C | 51·68 | 51·50 | 52·25 | 52·29 | 52·16 | 52·19 | 52·25 | 52·10 | 52·00 |
| H | 7·15 | 7·09 | 6·97 | 7·17 | 7·16 | | 6·97 | 6·70 | 6·85 |
| N | 15·24 | | | 14·81 | | | | 15·17 | 15·15 |
| O | | | | | | | | 26·00 | 25·99 |

The sulphur and phosphorus in this substance, per cent., gave

| | | | |
|---|---|---|---|
| Sulphur . . . . . | =1·36 | 1·09 | 1·03 per cent. |
| Phosphorus . . . . | = | ·20 | ·19 per cent. |

Hence it is an oxide of albumen, and from the ultimate analysis, it is the hydrated deutoxide of albumen.

2nd. In the above case of mollities ossium 66·97 parts of this hydrated deutoxide of albumen were passing out of the body in every 1000 parts of the urine. Hence, therefore, there was as much of this peculiar albuminous substance in the urine as there is ordinary albumen in healthy blood. So far, then, as the albumen alone is concerned, each ounce of urine passed was equivalent to an ounce of blood lost.

3rd. The peculiar characteristic of this hydrated deutoxide of albumen was its solubility in boiling water, and the precipitate with nitric acid being dissolved by heat and reformed when cold. By this reaction a similar substance in small quantity may be detected in pus and in the secretion from the vesiculæ seminales.

4th. This substance must again be looked for in acute cases of mollities ossium. The reddening of the urine on the addition of nitric acid might perhaps lead to the rediscovery of it; when found, the presence of chlorine in the urine, of which there was a suspicion in the above case, should be a special subject of investigation, as it may lead not only to the explanation of the formation of this substance, but to the comprehension of the nature of the disease which affects the bones.

Lastly, I am much indebted to Dr. Watson and Dr. MacIntyre for enabling me to follow out this case, and also to Professor Fownes for the use of his laboratory at University College.

**Basic structure of an antibody molecule**

Due to the vast number of different antibodies present in each human, biochemists had a hard time in trying to purify one of them out of million different ones. Since the myeloma protein fractionated with gamma globulin of the serum, it was thus suspected that it may be related to antibodies. Thus it was essential to try to decipher the basic structure of the simplest antibody molecule.

Two crucial studies were undertaken during 1959. The combination of these findings eventually provided the basic Y-shaped structure of antibodies.

Rabbit gamma-globulins were relatively homogeneous in size. One might then use a collection of different rabbit antibodies with similar binding affinities for a specific antigen. Porter tried to investigate the possibility of reducing the size of gamma-globulins but retaining the same binding affinity with enzymatic digestion of crystalline papain (Porter RR (1959) The hydrolysis of rabbit γ-globulin and antibodies with crystalline papain. Biochem J, 73, 119-126). He obtained three fractions: I, II and III. Fractions I and II have similar molecular weights and could bind the antigen, while fraction III was larger and had no antigen binding ability. Currently, fractions I and II are designated as Fab, and fraction III Fc. Therefore, the ratio of Fab to Fc is 2 to 1.

Also in 1959, Edelman tried to dissociate gamma-globulin with sulfhydryl compounds (Edelman GM (1959) Dissociation of γ-globulin. J Amer Chem Soc, 81, 3155-3156). He concluded that human gamma-globulin contained subunits linked at least in part by disulfide bonds. Subsequently, he designated the two components as heavy and light chains, with the light chains having similar sizes to the Bence Jones proteins from the same patient. The rumor was that Edelman was a graduate student at that time. His advisor did not consider Edelman's finding of any significance and refused to publish with Edelman. Subsequently, Porter and Edelman were awarded the Nobel Prize while Edelman's advisor never got that Prize.

These results therefore suggested that Bence Jones protein and myeloma protein from a patient suffering from multiple myeloma represented one specific antibody molecule with the light chain being filtered out by the kidney. Consequently, while biochemists had no luck in purifying one antibody from a mixture of millions of different ones, the cancerous process of multiple myeloma did the isolation of one specific one. Indeed, multiple myeloma is the cancer of one specific antibody producing cell.

of iodine. Coupling in the 3:5 positions occurs only slightly or not at all.

3. This reaction has been utilized as the basis of a method for the localization of $^{131}$I in the thyroxine molecule.

4. The method is applicable to biosynthesized [$^{131}$I]thyroxine.

I am indebted to the Rockefeller Trust of the Medical School and to the Medical Research Council for facilities for carrying out this work.

I wish to thank Miss Diana Walmsley for technical assistance.

## REFERENCES

Barac, G. & Morren, H. (1953). *Bull. Soc. Chim. biol., Paris*, **35**, 299.

Bowden, C. H., Maclagan, N. F. & Wilkinson, J. H. (1955). *Biochem. J.* **59**, 93.

Gross, J. & Leblond, C. P. (1950). *J. biol. Chem.* **184**, 489.

Komant, W. (1950). *Hoppe-Seyl. Z.* **286**, 229.

Michel, R., Roche, J. & Tata, J. (1952). *Bull. Soc. Chim. biol., Paris*, **34**, 466.

Saunders, K. H. (1949). *The Aromatic Diazo-compounds*, p. 218. London: Edward Arnold and Co.

# The Hydrolysis of Rabbit γ-Globulin and Antibodies with Crystalline Papain

By R. R. PORTER

*National Institute for Medical Research, Mill Hill, London, N.W.7*

(*Received* 23 *February* 1959)

The molecular size of rabbit γ-globulin is such that any direct attempt to relate structure to biological activity is not feasible at present. An alternative approach is to degrade antibody molecules in such a way that activity will persist in smaller fragments and so to reduce the structural problems involved. In an earlier investigation of this type (Porter, 1950*b*) it was found that, if rabbit γ-globulin containing antiovalbumin was digested with crude papain, a fragment with a molecular weight of about 40 000 could be produced which retained the ability to combine specifically with ovalbumin though it would no longer form a precipitate. It seemed probable that this fragment contained an antibody-combining site.

In that work, the amounts of crude enzyme used relative to γ-globulin were large, making subsequent investigation of the products of digestion rather difficult. As crystalline papain can now be prepared easily (Kimmel & Smith, 1954) and fractionation techniques have improved, these earlier experiments have been repeated. It has been found that γ-globulin is split by papain into three large pieces with very little release of amino acids or small peptides. If the γ-globulin contains antibody against any of the several antigens investigated, then two of these pieces retain combining though not precipitating power. The third piece, which may be readily crystallized, has no antibody activity but it has most of the antigenic sites of the original molecule. The isolation and properties of these three fractions will be described.

A preliminary account of this work has been given (Porter, 1958*a*).

## EXPERIMENTAL

### *Materials*

*Antisera.* Rabbit antiovalbumin, antibovine serum albumin and antihuman serum albumin were prepared by intravenous injection of the alum-precipitated protein (Porter, 1955).

Rabbit antipneumococci polysaccharide type 3 was prepared by injection of a suspension of the formalin-killed bacteria in 0·9 % NaCl solution.

Goat antirabbit γ-globulin was prepared by intravenous injection of alum precipitate and subcutaneous injection of γ-globulin with adjuvant (Freund & McDermott, 1942).

Rat antiserum was prepared according to the following schedule. The rats were given two intramuscular injections of protein and adjuvant into different thighs, with 1 week's interval between injection. Five weeks later they were given three injections, with 3-day intervals between injections, of alum-precipitated protein into the tail vein. They were bled by heart puncture 6 days after the last injection.

*Rabbit γ-globulin.* This was prepared either by chromatography, with diethylaminoethylcellulose (Sober, Gutter, Wyckoff & Peterson, 1956), or by $Na_2SO_4$ precipitation according to the method of Kekwick (1940).

*Crystalline papain.* This was prepared from crude enzyme powder purchased from Hopkin and Williams Ltd., London. The enzyme was crystallized once as the free enzyme and twice as the inactive Hg dimer (Kimmel & Smith, 1954). It was freeze-dried and stored as the dimer.

### *Methods*

*Quantitative estimation of precipitating antibody.* This was carried out according to Kabat & Mayer (1948).

*Estimations of inhibitory power.* These were made by quantitative antibody assay in the presence of the inhibitor

or by measure of the delay in precipitation caused by the inhibitor with the antibody and antigen mixed in optimum proportions.

*Chromatography.* Diethylaminoethylcellulose and carboxymethylcellulose were prepared according to Peterson & Sober (1956). Column size was 2·4 cm. (diam.) by 30–35 cm. (ht.); volume of mixing chamber was 1200 ml. Buffers used on carboxymethylcellulose columns were 0·01 M-sodium acetate, pH 5·5, with gradient to 0·9 M-sodium acetate, pH 5·5. In the refractionation of fraction I on diethylaminoethylcellulose, the following system was used: 0·01 M-sodium phosphate, pH 6·4, with gradient to 0·2 M-sodium phosphate, pH 6·2. All buffers were saturated with toluene. The pH was measured with an EIL Direct Reading pH meter (Electronic Instruments Ltd.).

*Enzyme digestion.* γ-Globulin (150 mg.) and Hg papain (1·5 mg.) were dissolved in 10 ml. of buffer (0·1 M-sodium phosphate, pH 7·0, 0·01 M-cysteine, 2 mM-ethylenediaminetetra-acetate). This solution was incubated at 37° for 16 hr. in the presence of toluene. It was then dialysed against water with vigorous stirring and several changes of the outer liquid over 48 hr. This procedure, which removed the cysteine and ethylenediaminetetra-acetate, and facilitated oxidation, inactivated the enzyme. *N*-Ethyl maleimide was also used to inactivate the enzyme but as there appeared to be no advantage this was not continued. The non-diffusible digestion products were either freeze-dried or fractionated directly by chromatography after dialysis against acetate buffer, pH 5·5.

*Protein determinations.* Protein concentrations were determined by reading the absorption at 280 and 260 mμ in a 1 cm. cell in a Unicam SP. 500 spectrophotometer.

*Radioactivity measurements.* Injection of hydrolysed algal [$^{14}$C]protein and counting of the γ-globulin fraction were carried out as described previously (Askonas, Humphrey & Porter, 1956).

*Analytical methods.* Amino acid analysis was by the method of Moore & Stein (1951) as modified by McDermott & Pace (1957).

*N*-Terminal amino acids were estimated by the fluorodinitrobenzene technique (Porter, 1957*a*). Hexose as 'glucose' was estimated by the anthrone method of Mokrasch (1954). Hexosamine as 'glucosamine' was estimated, after separation from amino acids on a cation-exchange resin (Boas, 1953), by a modification of the method of Elson & Morgan (1933).

## RESULTS

When a digestion mixture, prepared as described above, was dialysed against water at 2° a precipitate formed which appeared to be crystalline. If, however, the dialysis was against 0·04 N-acetic acid there was no precipitation and the recovery of the non-diffusible digestion products could be estimated either by measuring the absorption at 280 mμ or by freeze-drying and weighing the dry powder. This has been done in a number of experiments, and by either method the recovery of γ-globulin protein taken has fallen in the range 85–95 %. In view of the probable handling losses in dialysis and freeze-drying, it is considered that the higher figure is the most accurate. When such a digest was examined in the ultracentrifuge in 0·1 M-phosphate, pH 6·7, some crystals again formed on dialysis but the supernatant showed only one peak ($S_{20,w}$ 3·5). As γ-globulin has $S_{20,w}$ 6·5, it was clear that the protein had been split into large fragments of similar size with very little production of diffusible peptides. Attempts to fractionate this mixture by zone electrophoresis were not successful, but resolution could be achieved by chromatography on carboxymethylcellulose. Acetate buffer, pH 5·5, was chosen because under these conditions most of the carboxyl groups of the ion-exchanger are dissociated; if the digest was brought nearer to neutrality crystals formed, indicating a low solubility of at least one component. To help to keep all the material in solution chromatography was carried out at room temperature (20–23°). With a gradient of increasing salt concentration at this pH, three components could be resolved which have been named fractions I, II and III in order of elution from the column (Fig. 1).

If the gradient on the column was reduced, fractions II and III were more spread and III 'tailed' badly, but there was no suggestion of any further resolution. Fraction I appeared very close to the solvent front and this was re-run on a diethylaminoethylcellulose column at pH 6·4. No fractionation was obtained but again with a slow gradient there was considerable tailing. By these limited criteria the three fractions appeared to be single components and to be the only significant products of the digestion of γ-globulin by papain.

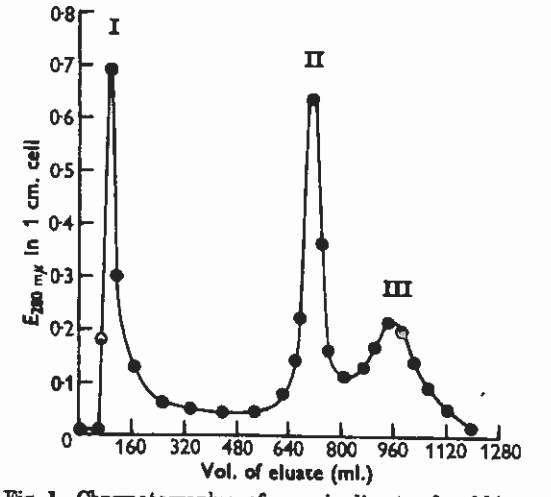



Fig. 1. Chromatography of papain-digest of rabbit γ-globulin on carboxymethylcellulose. Weight of digest 150 mg. Column 30 cm. × 2·4 cm. diam. Volume of mixing chamber 1200 ml. Gradient from 0·1 M-sodium acetate, pH 5·5, to 0·9 M-sodium acetate, pH 5·5, commencing at 200 ml. eluate volume.

Results with shorter times of hydrolysis under the same conditions suggest that the splitting is complete in very much less than 16 hr., the period which has always been used. It follows that these three fractions are exceptionally resistant to further digestion by papain.

Yields of the three fractions were measured by summing the absorption at 280 mμ in each peak. The ratios of yield varied somewhat from experiment to experiment but averaged (I:II:III) 1:0·8:0·9, and total recovery from the column was 85–90 %. When re-run, fractions I and III were recovered in about 95 % yield and fraction II in about 85 % yield. The absorptions of peaks I, II and III at 280 mμ at a concentration of 1 mg./ml. in water or 0·02N-acetic acid were 1·4, 1·4 and 1·0 respectively. If the relative yields are corrected for the lower recovery of fraction II, and the lower specific absorption at 280 mμ of fraction III, then the corrected relative yields (I:II:III) are 1:0·9:1·25 by wt. In some experiments the fractions were concentrated by pressure-dialysis in the cold against water or 0·02N-acetic acid; insoluble material was discarded and the solution freeze-dried and weighed. The yields of I and III were similar to those above but II was somewhat lower. Considerable error can occur because of denaturation in dilute solution, and variable losses on chromatography, but it is considered that the average yields calculated from chromatography approximate to the yields of the fractions produced in the digestion.

Fraction III is readily identifiable as the material with a low solubility near neutrality. It may be crystallized and recrystallized by dialysing a solution in 0·02N-acetic acid against sodium phosphate buffer, pH 6·0–7·0, at 2°. The crystals are diamond-shaped plates, often of considerable size but thin and easily broken (Fig. 2).

*Molecular weights.* The three fractions were studied in the ultracentrifuge (see Addendum) and the results for normal γ-globulin are summarized in Table 1. The sum of the molecular weights of the three fragments is very close to the molecular weight of the original γ-globulin. This is in agreement with the high recovery of non-diffusible digestion products. The relative sizes of the three fragments were (I:II:III) 1:1·05:1·6, which is in approximate though not exact agreement with the calculated relative yields of the fragments.

*Amino acid analysis.* Amino acid analysis was carried out by Mr E. E. McDermott and Dr J. Pace at the Cereals Research Association, St Albans. The results are given in Table 2 for fractions prepared from γ-globulin from non-immune serum. The most striking feature of the analysis is the great similarity of the amino acid content of fractions I and II. With the exception of cystine,

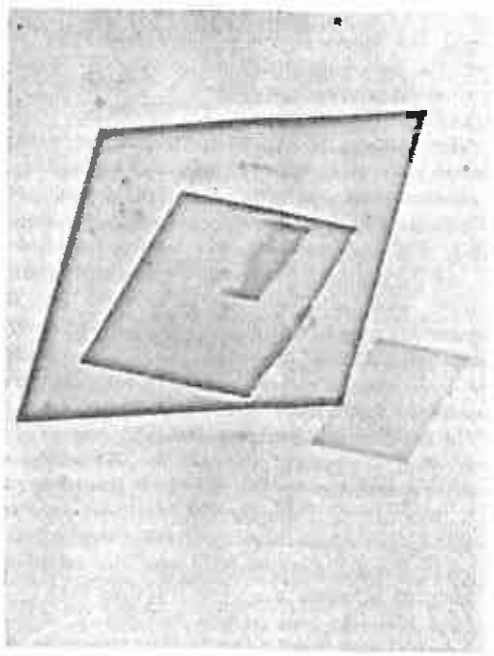


Fig. 2. Crystals of fraction III. Magnification × 500.

Table 1. *Molecular weights of fragments and γ-globulins from non-immune serum*

For details see Addendum.

| | |
|---|---|
| γ-Globulin | 188 000 |
| I | 50 000 |
| II | 53 000 |
| III | 80 000 |

Table 2. *Amino acid analysis of fragments from non-immune serum*

| Amino acid | N (percentage of total N) I | II | III |
|---|---|---|---|
| Aspartic acid | 5·66 | 6·01 | 6·18 |
| Threonine | 10·42 | 11·56 | 5·61 |
| Serine | 8·77 | 9·57 | 7·34 |
| Glutamic acid | 5·99 | 6·18 | 7·83 |
| Proline | 5·08 | 5·08 | 6·68 |
| Glycine | 7·57 | 6·93 | 3·18 |
| Alanine | 5·65 | 6·12 | 2·82 |
| Valine | 7·38 | 7·58 | 5·96 |
| Methionine | 0·64 | 0·75 | 1·48 |
| Isoleucine | 2·20 | 2·26 | 3·43 |
| Leucine | 5·17 | 5·39 | 5·19 |
| Tyrosine | 3·44 | 3·58 | 2·33 |
| Phenylalanine | 2·53 | 2·65 | 2·76 |
| Histidine | 1·59 | 1·87 | 5·23 |
| Lysine | 6·03 | 6·84 | 9·95 |
| Arginine | 5·94 | 6·22 | 13·93 |
| Tryptophan | | Not estimated | |
| ½ Cystine | 2·80 | 2·22 | 1·18 |

differences rarely exceed 10 % and, as the results quoted are taken from a single analysis, this is close to the experimental error. In contrast, fraction III shows considerable differences, most marked in the contents of threonine, proline, glycine, alanine, methionine, isoleucine, tyrosine, cystine and the three basic amino acids, the differences in several cases being more than 100 %.

These results have been recalculated in terms of amino acid residues per mol. of fraction and per mol. of $\gamma$-globulin. It has been assumed that the ash-free, moisture-free fractions have the same nitrogen content of 16 % as found for $\gamma$-globulin by Smith, McFadden, Stockell & Buettner-Janesch (1955). In view of the difference in basic amino acid content the figure may be rather high for fractions I and II and low for fraction III. The results of this calculation are given in Table 3. The amino acid residues accounted for by the three fractions range from 83 to 116 % of the figures for whole $\gamma$-globulin. These discrepancies are probably explicable by errors in the different estimations, the assumption of 16 % of nitrogen in each fraction and by the peptide material lost in the dialysis. The overall recovery of amino acid residues in this calculation is 94 %.

Carbohydrate estimations were made by Dr H. R. Perkins of this Institute, and the results with whole $\gamma$-globulin and the fractions are given in Table 4. About two-thirds of the carbohydrate of the original molecule is found with fraction III, one third with fraction I and a small amount, which may not be significant, is with fraction II. It seems likely therefore that the carbohydrate moiety of $\gamma$-globulin is in the two pieces, the larger with that part of the molecule which corresponds to fraction III and the smaller with the part corresponding to fraction I. If the total recovery of carbohydrate in the three fractions is compared with that present in the original, the yields are approximately 110 and 120 % for hexose and hexosamine respectively.

*N*-Terminal assay has been carried out on fractions I and II, and in both alanine was found to be the main terminal amino acid (about 0·9 mol./50 000), together with smaller amounts of aspartic acid (0·2 mol./50 000) and trace amounts of serine and threonine (together about 0·1 mol./50 000). This is a very similar result to that obtained for whole rabbit $\gamma$-globulin (Porter, 1950*a*; McFadden & Smith, 1955), except that the *N*-terminal amino acid content is more than three times as great as in the whole globulin. It is also in approximate agreement with the results of the digestion of rabbit antiovalbumin with papain powder, when an immunologically active fraction containing one *N*-terminal alanine per 40 000 was reported (Porter, 1950*a*). The significance of finding *N*-terminal alanine in both I and II is not certain, as there are about 110 alanine residues per molecule of $\gamma$-globulin. An attempt to determine the *N*-terminal sequence was unsatisfactory owing to shortage of material, but it appeared that the sequence of I was alanylaspartyl and of II alanyl-leucyl. If this is correct, it suggests that II may derive from the *N*-terminal part of the molecule, as the sequence there is alanyl-leucylvalylaspartylglutamyl (Porter, 1950*a*; McFadden & Smith, 1955).

*Immunological properties.* The three fractions were prepared from digests of $\gamma$-globulin which had been obtained from rabbit antisera against ovalbumin, bovine serum albumin, human serum albumin and antipneumococci polysaccharide type 3. None would precipitate with the corresponding antigen but fractions I and II prepared from $\gamma$-globulin containing antiprotein antibodies inhibited the precipitation of the antigen by the homologous antiserum. This effect is specific; for

Table 3. *Recovery of amino acid residues*

| Amino acid | Total no. of residues in I+II+III* | Total no. of residues in whole $\gamma$-globulin† | Recovery (%) |
|---|---|---|---|
| Aspartic acid | 126 | 137 | 93 |
| Threonine | 180 | 207 | 87 |
| Serine | 175 | 196 | 89 |
| Glutamic acid | 144 | 150 | 95 |
| Proline | 121 | 138 | 87 |
| Glycine | 115 | 140 | 83 |
| Alanine | 95 | 111 | 86 |
| ½ Cystine | 40 | 44 | 91 |
| Valine | 144 | 158 | 91 |
| Methionine | 22 | 19 | 116 |
| Isoleucine | 58 | 63 | 92 |
| Leucine | 110 | 111 | 100 |
| Tyrosine | 63 | 68 | 92 |
| Phenylalanine | 56 | 62 | 90 |
| Lysine | 84 | 83 | 100 |
| Histidine | 23 | 19 | 120 |
| Arginine | 66 | 54 | 120 |
| Tryptophan | | Not estimated | |
| Total | 1582 | 1680 | 94 |

* Calculated from molecular weights and amino acid analysis, assuming a nitrogen content of 16 % for each fraction.

† Calculated from Smith, McFadden, Stockell & Buettner-Janesch (1955), assuming a molecular weight of 188 000.

Table 4. *Carbohydrate content of $\gamma$-globulin and fractions*

| | Hexosamine (%, as glucosamine) | Hexose (%) | Total carbohydrate (%) |
|---|---|---|---|
| $\gamma$-Globulin | 0·6 | 0·5 | 1·1 |
| I | 0·5 | 0·6 | 1·1 |
| II | 0·1 | 0·15 | 0·25 |
| III | 1·3 | 0·9 | 2·2 |

example, I and II from antihuman serum albumin γ-globulin had no effect when added to bovine serum albumin and its antiserum; fraction III, on the contrary, had no effect whatever its source or whatever the test system. Testing fraction III is difficult owing to its low solubility near neutrality, where antibody–antigen precipitation is carried out, but it is soluble to about 0·3 mg./ml. at 37° at pH 7·2 and it could be shown that it had less than one-thirtieth of the activity of the corresponding fractions I and II.

Quantitative estimates of the inhibitory power of I and II are shown in Fig. 3. In order to assess this on a molar basis it has been assumed that the fractions would have the same proportion of active molecules as the proportion of antibody in the γ-globulin taken, and the weight of each fraction has been corrected accordingly when plotting the graph. On this basis it can be seen that approximately equal weights of inhibitor are required for complete inhibition of precipitation, i.e. a molar ratio of inhibitor to antibody of 3·5:1.

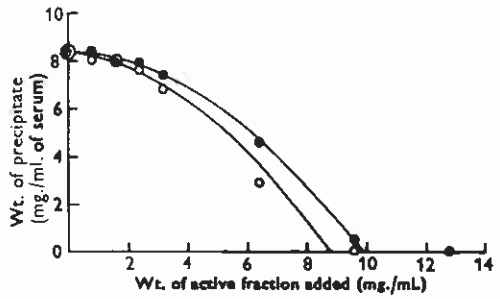



Fig. 3. Inhibition of precipitation of human serum albumin–antihuman serum albumin. Increasing weights of fractions I and II from antihuman serum albumin digest were added to antibody–antigen in equivalence. Weights of fractions were corrected for the presence of inactive material (see text). ●, Fraction I; ○, fraction II.

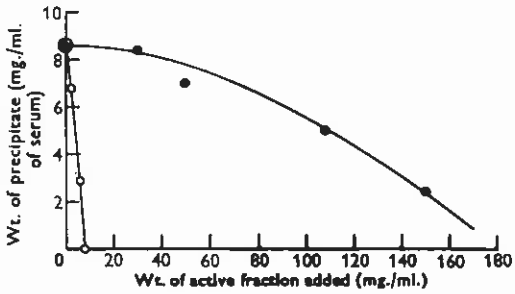



Fig. 4. Inhibition of precipitation of human serum albumin–antihuman serum albumin and pneumococci polysaccharide type 3 antipolysaccharide. ●, Fraction II from antipolysaccharide γ-globulin; ○, fraction II from antiserum albumin γ-globulin.

In contrast with these results none of the fractions from antipolysaccharide γ-globulin appeared to have any power to inhibit the combination of polysaccharide with its antiserum. However, when much higher ratios of inhibitor to antibody were used, inhibition of precipitation occurred with fraction I or II. The difference between the antiprotein and antipolysaccharide fractions appears to be quantitative rather than qualitative, as originally stated (Porter, 1958*a*). The results with fraction II from antipneumococci and antihuman serum albumin γ-globulin are shown in Fig. 4. The inhibitory power of the former is about 4 % of that of the latter, on a molar basis. Fraction III again appeared to be completely inactive in qualitative tests, but in this case the solubility limitation is greater and it could only be concluded that III has certainly less than half the activity of I or II.

*Antigenic activity.* The power of the fractions to precipitate with goat antirabbit γ-globulin serum was now tested and I and II showed neither precipitation nor inhibitory activity. Fraction III precipitated 70 % of the antibody precipitated by γ-globulin.

Rat antiserum was prepared against rabbit γ-globulin and fractions I, II and III. Groups of three rats were used for each antigen and, after the immunization course described above, the serum of each group was pooled, and gave antibody contents of 4·2, 7·8, 8·1 and 4·0 mg. of antibody/ml. for γ-globulin, I, II and III respectively. All the fractions were at least as effective antigens as the original molecule, and I and II were the most effective, but as only three animals were used per group the difference may not be significant.

When rat anti-γ-globulin was tested with the fractions, rather different results from those with

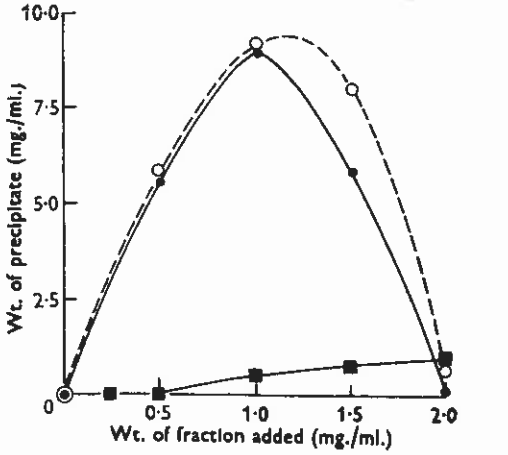



Fig. 5. Precipitation curves of I (●), II (○) and III (■) with rat anti-II serum.

goat antiserum were obtained. Fraction III precipitated 50 % of the antibody, and I and II 15 % each. The antigenic specificities of the fractions were compared by measuring the precipitates formed by I, II and III with rat anti-II serum. The precipitation curves are shown in Fig. 5. It can be seen that the curves for I and II are almost identical, but III gives almost no precipitate. This again emphasizes the great similarity of I and II and their sharp distinction from III.

*Synthesis of the fraction* in vivo. In view of the distinctive characters of the three fractions the possibility was considered that they might be synthesized separately and joined into the whole molecule in a separate step. If this were to happen the fractions would probably not all be synthesized from the same pool of amino acids at the same time, and hence if radioactive amino acids were injected into a rabbit the incorporation of radioactivity after a short time interval would vary from fraction to fraction. Hydrolysed algal [$^{14}$C]protein was used as the source of labelled amino acids; 240 $\mu$c was injected intravenously, and the rabbit was bled after 1 and 4 hr. The plasma $\gamma$-globulin is not labelled until 30 min. after injection (Askonas *et al.* 1956). The $\gamma$-globulin was prepared and hydrolysed with papain, and the fractions were isolated, precipitated with trichloroacetic acid, and counted, at infinite thickness on a 1 cm.$^2$ disk. The specific activities are given in Table 5 and it can be seen that there is no significant difference between the different fractions. There is therefore no evidence to suggest that these fractions or any large parts of them are synthesized independently.

## DISCUSSION

The results show that papain-digestion of rabbit $\gamma$-globulin causes limited and highly selective splitting to give three large fragments and very few small peptides. As papain shows rather a wide specificity in the digestion of the $\beta$ chain of insulin (Sanger, Thompson & Kitai, 1955), it is apparent that the structure of the native molecule must be such that many potentially hydrolysable bonds are protected by steric and other factors. Ultracentrifuge studies of the splitting of $\gamma$-globulins from different species, by a variety of enzymes (Petermann & Pappenheimer, 1941; Petermann, 1942, 1946), have all shown that a small number of fragments are the main products in each case. It is possible that in $\gamma$-globulins only small parts of the peptide chain are accessible to proteolytic enzymes, so that even though different enzymes break different bonds in these vulnerable sections the principal large digestion products are similar.

Our results with the papain-digestion of rabbit $\gamma$-globulin suggest that the single polypeptide chain (Porter, 1950*a*; McFadden & Smith, 1955) has been split into three distinct sections, which together comprise some 90 % of the original molecule. However, fractions I and II are so similar that the question arises whether they could be derived very largely from the same section of the chain. They are almost indistinguishable in *N*-terminal amino acid, amino acid analysis, molecular weight, antigenic specificity and, if derived from antibody, in their antigen-binding capacity. They differ in chromatographic behaviour and carbohydrate and cystine content.

There are at least three possible ways of splitting a single chain to give results such as this, and they are illustrated in Fig. 6: A shows the obvious split into three distinct sections, B the production of two fractions from the same section of chain and C the result if $\gamma$-globulin consists of two types of molecule, such as euglobulin and pseudoglobulin, one of which gives rise to I+III and the other to II+III. If either B or C were correct then the yield of III should exceed the sum of the yield of I and II. In fact the yield of each of the three fractions is very similar, as would be expected in A. In B and C the overall recovery, in view of the molecular weights found, would be 130 000/188 000, i.e. 70 %, considerably lower than the experimental figure of about 90 %, which again is

Table 5. *Specific activities of fractions from $\gamma$-globulin after injection of algal* [$^{14}$C]*amino acids*

Standard error was less than 5 %.

| Time after injection (hr.) | Counts/min. (on 1 cm.$^2$ disk at infinite thickness) I | II | III |
|---|---|---|---|
| 1 | 22 | 25 | 25 |
| 4 | 80 | 84 | 83 |

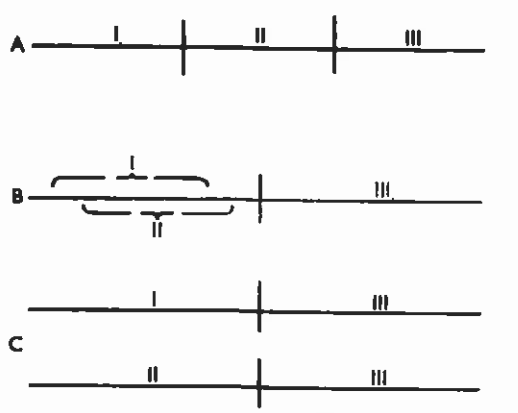



Fig. 6. Different ways of splitting rabbit $\gamma$-globulin to give three fractions of approximately equal size. For explanation see Discussion.

41

that which would be expected if A were correct. Similarly, with the individual amino acid residues, the recoveries would be only 70%, with the possibility of big variations in individual cases as much more material would be lost. In fact the recovery of residues is 90–95%, and the variation 83–120% in the recoveries of individual amino acids is close to the range expected from the errors in the different measurements and the assumptions made in the calculation.

Further, in B it would be expected that if I and II were digested further by papain, either interconversion would occur or both would be reduced to a slightly smaller common product. In fact both appear to be stable to a further 16 hr. digestion with papain at 37°, being unchanged in chromatographic behaviour and other properties. The simplest explanation of our results therefore seems to be that γ-globulin has been split into three distinct fractions, as shown in A. It is probable that the inhibitor described in the earlier work (Porter, 1950*b*) was a mixture of fractions I and II, and that fraction III was prevented from crystallizing by the impurities introduced with the crude enzyme preparation.

It follows that two large sections of γ-globulin (each about 30% of the whole) are extremely similar in chemical, physical and biological properties. The finding of such close agreement between the amino acid analysis of these two fragments and also an almost identical antigenic specificity suggests that there may be almost a repeat of the amino acid sequence and configuration. This is in contrast with the properties of fraction III, which differs in every respect, so that there appears to be a large repeating unit (I or II) joined to a larger section (III) of entirely different character. This unusual make-up of the γ-globulin molecule is presumably related to its antibody activity. The similarity of the pieces with mol.wt., 50 000, where the antibody-combining sites may be very much smaller (Kabat, 1956), raises the question whether large sections are required to maintain the structural integrity of a small combining site or whether antibody-combining sites may occur anywhere in these pieces.

The big quantitative difference between the inhibitory power of fractions I and II, when derived from antiprotein or antipolysaccharide antibodies, may reflect important differences between the two types of combining site, but perhaps more probably arises from differences in the speed and mechanism of precipitation between the two systems.

The significance of the part of the γ-globulin molecule represented by fraction III in antibody–antigen reactions is not known. The ease of crystallization could be taken as evidence of greater identity of structure among individual molecules in III than in the whole molecule, which appears to be complex by all available physical and biological data (see Porter, 1958*b*). Preliminary electrophoretic examination, however, suggests that it is as complex as the original molecule by this criterion. Most of the antigenic sites of γ-globulin appear to be on III, but as these sites can only be defined in relation to a given antiserum (Porter, 1957*b*), variable results might be expected, and there is in fact a difference in this respect between the goat and rat anti-γ-globulin serum. Fraction III is remarkably stable. After precipitation by 5% trichloroacetic acid at room temperature, washing with ethanol and ether and drying for 1 hr. at 110°, it will redissolve in 0·05N-acetic acid and it retains its power specifically to precipitate goat antirabbit γ-globulin.

The finding that γ-globulin appears to be built of three sections, one of exceptional stability and the other two containing the antibody-combining sites, is reminiscent of Pauling's (1940) theory of antibody formation, in which it is suggested that antibody molecules consist of a rigid centrepiece and two flexible ends capable of taking up configurations complementary to the antigen and hence forming antibody-combining sites on these flexible parts. Pauling further suggested that the flexibility might be due to a high content of proline residues. There is no evidence on the relative positions of fractions I, II and III in the whole molecule, except that II may be from the *N*-terminal end. Nor is there any evidence on the essential feature of Pauling's theory, that the amino acid sequence of all antibodies is identical and that the different antibody-combining sites are formed only by refolding of the same polypeptide chain. The proline content of I and II is less than that of III, whereas the cystine content, often an important feature in determining the stability of a protein molecule, is lower in III than in I and II.

The equal rate of incorporation of radioactive amino acids into the different fractions is in agreement with the view that the whole molecule is synthesized simultaneously from the same pool of amino acids.

## SUMMARY

1. Rabbit γ-globulin, when digested by crystalline papain, gives three fragments which together form 90% or more of the original molecule.

2. If the γ-globulin contains antibodies, two fragments (I and II), of molecular weight 50 000–55 000, retain the power to combine with the antigen. The third fragment (III), molecular weight about 80 000, crystallizes easily and has much of the antigenic specificity of the original molecule.

3. I and II are extremely similar in chemical and biological properties, and III differs very widely in all respects. This has led to the suggestion that rabbit $\gamma$-globulin is formed of two pieces with very similar structure joined to a third piece of quite different character.

I wish to thank Dr J. H. Humphrey for the gift of goat antiserum and advice on the immunization of rats, Mr E. E. McDermott and Dr J. Pace for their invaluable assistance in carrying out the amino acid analysis, Dr H. R. Perkins for the carbohydrate analysis and Mr M. Young for the photographs of the crystals. I also wish to thank Miss Hilary Morgan for her excellent technical assistance.

## REFERENCES

Askonas, B. A., Humphrey, J. H. & Porter, R. R. (1956). *Biochem. J.* **63**, 412.
Boas, N. F. (1953). *J. biol. Chem.* **204**, 553.
Elson, L. A. & Morgan, W. T. J. (1933). *Biochem. J.* **27**, 1824.
Freund, J. & McDermott, K. (1942). *Proc. Soc. exp. Biol., N.Y.*, **49**, 548.
Kabat, E. A. (1956). *J. Immunol.* **77**, 377.
Kabat, E. A. & Mayer, M. M. (1948). *Experimental Immunochemistry*, p. 59. Springfield, Illinois: C. C. Thomas.
Kekwick, R. A. (1940). *Biochem. J.* **34**, 1248.
Kimmel, J. R. & Smith, E. L. (1954). *J. biol. Chem.* **207**, 515.
McDermott, E. E. & Pace, J. (1957). *Brit. J. Nutr.* **11**, 446.
McFadden, M. L. & Smith, E. L. (1955). *J. biol. Chem.* **214**, 185.
Mokrasch, L. C. (1954). *J. biol. Chem.* **208**, 55.
Moore, S. & Stein, W. H. (1951). *J. biol. Chem.* **192**, 663.
Pauling, L. (1940). *J. Amer. chem. Soc.* **62**, 2643.
Petermann, M. L. (1942). *J. biol. Chem.* **144**, 607.
Petermann, M. L. (1946). *J. Amer. chem. Soc.* **68**, 106.
Petermann, M. L. & Pappenheimer, A. M., jun. (1941). *J. phys. Chem.* **45**, 1.
Peterson, E. A. & Sober, H. A. (1956). *J. Amer. chem. Soc.* **77**, 751.
Porter, R. R. (1950*a*). *Biochem. J.* **46**, 473.
Porter, R. R. (1950*b*). *Biochem. J.* **46**, 479.
Porter, R. R. (1955). *Biochem. J.* **59**, 405.
Porter, R. R. (1957*a*). *Methods in Enzymology*, vol. 4, p. 221. Ed. by Colowick, S. P. & Kaplan, N. O. New York: Academic Press Inc.
Porter, R. R. (1957*b*). *Biochem. J.* **66**, 677.
Porter, R. R. (1958*a*). *Nature, Lond.*, **182**, 670.
Porter, R. R. (1958*b*). *Folia biol., Prague*, **4**, 310.
Sanger, F., Thompson, E. O. P. & Kitai, R. (1955). *Biochem. J.* **59**, 509.
Smith, E. L., McFadden, M. L., Stockell, A. & Buettner-Janesch, V. (1955). *J. biol. Chem.* **214**, 197.
Sober, H. A., Gutter, F. J., Wyckoff, M. M. & Peterson, E. A. (1956). *J. Amer. chem. Soc.* **78**, 756.

## ADDENDUM

# Ultracentrifugal Examination of Digestion Products from Rabbit $\gamma$-Globulin

By P. A. CHARLWOOD
*The National Institute for Medical Research, Mill Hill, London, N.W. 7*

(*Received* 23 *February* 1959)

Sedimentation-velocity experiments were carried out at 59 780 rev./min. in a Spinco model E ultracentrifuge, the technique and calculation of results being as described by Charlwood (1952, 1955). Molecular-weight determinations were by a method suggested originally by Archibald (1947), the procedure used being that of Charlwood (1957). For these determinations the ultracentrifuge was run at 8210 or 12 590 rev./min. according to the sample being studied. Protein concentrations were measured in terms of the refractive increment between the protein solution and buffer against which it had been dialysed until in equilibrium. The differential refractometer used was of the type introduced by Cecil & Ogston (1951). A cell with a Kel-F (polychlorotrifluoroethylene) plastic centrepiece was used for most experiments in acetate buffer. In all calculations the partial specific volume was taken arbitrarily as 0·735, because no experimental values were available.

When digests of the $\gamma$-globulin were dialysed against sodium phosphate buffer (pH 6·8; $I$ 0·2), some material (later identified with fraction III) precipitated, but the soluble part, on examination at top speed in the ultracentrifuge, showed a single, fairly symmetrical peak of sedimentation coefficient about 3·5s. The results of some of the experiments made with the isolated fractions are shown in Table 1, together with the corresponding figures for untreated $\gamma$-globulin. In all cases, with a possible exception referred to below, preparations were monodisperse in the ultracentrifuge.

The sedimentation coefficients of all three fractions were approximately the same ($\sim$3·5s) at the concentrations used in most of the experiments (0·65–0·82%), and measurements made at lower concentrations, but not given in Table 1, showed that the concentration-dependence was small (not more than 3–4% for a change in protein concentration of 1%, w/v). Molecular weights measured

Table 1. *Molecular weights of digestion products from rabbit γ-globulins*

Rabbit γ-globulin was prepared by sodium sulphate fractionation of serum (Kekwick, 1940).

| Source | Sample | Buffer | Refractive increment of protein solution | $S_{20,w}$ | Molecular wt. ± S.D. |
|---|---|---|---|---|---|
| Rabbit γ-globulin (batch 1) | Whole globulin | Phosphate (pH 6·8; *I* 0·2) | 0·00149 | 6·50 s | 187 600 ± 1400 |
| Rabbit γ-globulin (batch 2) | Fraction I | Phosphate (pH 6·8; *I* 0·2) | 0·00147 | 3·59 s | 49 600 ± 1800 |
| | Fraction II | Phosphate (pH 6·8; *I* 0·2) | 0·00138 | 3·55 s | 52 600 ± 2600 |
| | Fraction III | Acetate (pH 4·0; *I* 0·2) | 0·00149 | 3·40 s | 80 200 ± 9100 |
| Rabbit γ-globulin (batch 3) | Fraction I | Phosphate (pH 6·8; *I* 0·2) | 0·00132 | — | 50 400 ± 2000 |
| | Fraction II | Phosphate (pH 6·8; *I* 0·2) | 0·00118 | — | 55 700 ± 2000 |
| | Fraction III | Acetate (pH 4·5; *I* 0·2) | 0·00133 | — | 80 200 ± 1600 |

at the finite concentrations indicated should therefore not differ greatly from values which would be obtained by extrapolation to infinite dilution.

The molecular weights quoted in Table 1 were each derived by taking the mean value calculated from five photographic exposures made at 8- or 16-min. intervals. The standard deviations of individual values about the corresponding mean are given in the last column as a measure of internal consistency. Comparisons between means must in addition take into account the uncertainty residing in refractometer measurements, which amounted to about ±0·6%, but which could affect a molecular weight to the extent of about ±0·9%. Agreement between different measurements on corresponding fractions is considered to be reasonably good when allowance is made for these sources of variation. The only unsatisfactory set of measurements in Table 1 refers to fraction III at pH 4·0, where very erratic results were obtained, with some suggestion of a downward trend with increasing time. Moreover, there was a detectable asymmetry of the peak obtained in the ultracentrifuge at high speed, although the sedimentation coefficient of the maximum ordinate (3·40 s) was lower than for similar preparations at pH 4·5, which gave about 3·7 s. It was concluded that complications, which may include some aggregation and some molecular-shape change, make pH 4·0 unsuitable for investigation of fraction III.

The sum of the molecular weights of the three fractions is in good agreement with the molecular weight of whole γ-globulin, the figure for which corresponds fairly well with that most commonly accepted (170 000; see Porter, 1959).

## REFERENCES


Archibald, W. J. (1947). *J. phys. Chem.* **51**, 1204.
Cecil, R. & Ogston, A. G. (1951). *J. sci. Instrum.* **28**, 253.
Charlwood, P. A. (1952). *Biochem. J.* **51**, 113.
Charlwood, P. A. (1955). *Canad. J. Chem.* **33**, 1043.
Charlwood, P. A. (1957). *Trans. Faraday Soc.* **53**, 871.
Kekwick, R. A. (1940). *Biochem. J.* **34**, 1248.
Porter, R. R. (1959). In *The Plasma Proteins*. Ed. by Putnam, F. W. New York: Academic Press Inc.


# The *N*-Acetylation and Estimation of Hexosamines


By G. A. LEVVY and A. McALLAN
*Rowett Research Institute, Bucksburn, Aberdeenshire*


(*Received 5 March* 1959)

*N*-Acetylglucosamine is usually determined by Morgan & Elson (1934) colour reaction, as modified by Aminoff, Morgan & Watkins (1952). A more recent modification by Reissig, Strominger & Leloir (1955) is greatly superior on grounds of sensitivity, convenience and reproducibility, and in addition is unaffected by small quantities of acid or alkali in the test sample. Glucosamine in relatively large amounts does not interfere in the Morgan–Elson reaction.

The standard procedure for the estimation of glucosamine calls for preliminary treatment with alkaline acetylacetone (Elson & Morgan, 1933; Rondle & Morgan, 1955); under the usual conditions *N*-acetylglucosamine interferes to some extent. Interference by other tissue constituents is greater than in the *N*-acetylglucosamine determination. Conversion of glucosamine into *N*-acetylglucosamine as a prelude to assay by the Morgan–Elson reaction offers certain advantages

nally present was further demonstrated by a Van Slyke amino nitrogen assay.

Acid hydrolysis yielded two aglycons which were separated by chromatography; I, m.p. 111–115°, $C_{21}H_{33}N$, ultraviolet absorption at 229–230 mμ (ε 15,260); 234 mμ (ε 16,700), 243 mμ shld. (ε 12,190); II, m.p. 168–173°, $C_{21}H_{35}NO$. Compound II has now been identified as 20α-amino-3β-hydroxy-5-pregnene[2] and, therefore, I corresponds to 20α-amino-3,5-pregnadiene. The sugar fraction of the hydrolysis was shown to be D-glucose by formation of its osazone, by paper chromatography and by its oxidation to potassium gluconate and subsequent formation of the characteristic aldobenzimidazole.[3] From the above evidence and from other considerations it seemed quite probable that the new steroid alkaloid was 20α-amino-3β-hydroxy-5-pregnene β-D-glucoside. This has been confirmed by synthesis.

20α-Amino-3β-hydroxy-5-pregnene was converted to 3β-hydroxy-20α-trifluoroacetamido-5-pregnene, m.p. 199–201°. Reaction with acetobromoglucose gave the 3β-D-glucoside acetate, m.p. 200–205°. Alkaline hydrolysis removed the blocking groups and the resulting base, m.p. 285–287°, yielded a hydrochloride, m.p. 257–259°, which was identical in every respect with isolated hypotensively active material. Several derivatives and compounds related to this substance have been investigated and these together with more complete details of the above experiments will be reported later.

(2) P. L. Julian, E. W. Meyer and H. C. Printy, THIS JOURNAL, 70, 887 (1948). See also V. Cerny, L. Lobler and F. Sorm, *Coll. Czech.*, 22, 76 (1957), for the stereochemistry of the C-20 amines.

(3) S. Moore and K. P. Link, *J. Biol. Chem.*, 133, 293 (1940).

RESEARCH DEPARTMENT DANIEL DICKEL
CIBA PHARMACEUTICAL PRODUCTS, INC. ROBERT LUCAS
SUMMIT, N. J. H. B. MACPHILLAMY

RECEIVED MAY 7, 1959

## DISSOCIATION OF γ-GLOBULIN

*Sir:*

Reaction of human γ-globulin with sulfhydryl compounds, sulfite, or performic acid resulted in marked diminution in the sedimentation coefficient and molecular weight. Efficient reduction required the presence of denaturing agents. Because the products were insoluble after removal of the denaturing agent, molecular weights were determined in urea solutions[1] by the Archibald principle[2]

$$q_a \equiv \frac{RT(\partial c/\partial r)_{r_a}}{\omega^2 r_a} = M_{app}(1 - \bar{v}\rho)[(c_a - c_0) + c_0]$$

$R$ is the gas constant, $T$ the absolute temperature, $(\partial c/\partial r)_{r_a}$ the concentration gradient at the meniscus $r_a$, $\omega$ the constant angular velocity, $M_{app}$ the apparent molecular weight of solute calculated as if its activity coefficient were unity, $\bar{v}$ the partial specific volume, $\rho$ the density of the solvent, $c_a$ the concentration at $r_a$, and $c_0$ the initial concentration. Plotting $q_a$ against $(c_a - c_0)$ yields $M_{app}(1 - \bar{v}\rho)$ as the slope of a least squares line. This allows calculation of a molecular weight average which initially emphasizes the heavier components of a polydisperse solution.[3]

Fifteen milligrams of human γ-globulin was reduced with 5 ml. of 0.1 $M$ β-mercaptoethylamine·HCl in 6 $M$ urea at room temperature for four hours. Some samples then were dialyzed against a large volume of 6 $M$ urea that was 0.02 $M$ in iodoacetamide. The products appeared as a polydisperse peak in the ultracentrifuge. Corrected for the density and viscosity of urea, the $s^0_{20w}$ of reduced iodoacetamide-treated γ-globulin was 2.3$S$. In Table I are the measured values of $M_{app}(1 - \bar{v}\rho)$ and $M_{app}$ values calculated assuming $\bar{v} = 0.74$.[4] Simultaneous determinations of $M_{app}(1 - \bar{v}\rho)$ with solutions in $D_2O$ and $H_2O$ that were 6 $M$ in urea and 0.1 $M$ in β-mercaptoethylamine·HCl, allowed estimation of $\bar{v} = 0.71$. Using this value, and $\rho = 1.097$ for the $H_2O$ solution yields $M_{app.} = 42{,}000$.

TABLE I

EFFECT OF VARIOUS REAGENTS ON $M_{app}(1 - \bar{v}\rho)$ VALUES OF HUMAN γ-GLOBULIN

(Fraction II of Cohn, Lederle lot C 543[a])

| Solvent | $M_{app}(1 - \bar{v}\rho)$ ± standard deviation | $M_{app}$ |
|---|---|---|
| 0.2 $M$ KCl | $(4.8 \pm 0.1) \times 10^4$ | 192,000 |
| 6 $M$ urea + 0.2 $M$ KCl | $(3.0 \pm 0.3) \times 10^4$ | 158,000 |
| 0.1 $M$ MEA[b] + 0.2 $M$ KCl | $(3.5 \pm 0.1) \times 10^4$ | 140,000 |
| 0.1 $M$ MEA + 6 $M$ urea + 0.2 $M$ KCl | $(0.93 \pm 0.07) \times 10^4$ | 48,000 |
| Reduced in 8 $M$ urea + 0.1 $M$ MEA, next dialyzed against 6 $M$ urea + 0.02 $M$ iodoacetamide then 6 $M$ urea + 0.2 $M$ KCl | $(0.92 \pm 0.05) \times 10^4$ | 48,000 |

[a] This sample contained a small amount of heavy material sedimenting faster than the main 7$S$ component. [b] MEA = β-mercaptoethylamine·HCl.

A pathological macroglobulin reduced in 6 $M$ urea had a value of 41,000 for $M_{app}$. Macroglobulins have been found[5] to dissociate in mercaptoethanol solutions without urea to sub-units of about the same molecular weight as normal γ-globulin.

Reaction of human γ-globulin with sulfite[6] in urea[7] yielded a water soluble aggregated S-sulfoprotein. In tris-(hydroxymethyl)-aminomethane buffer, $p$H 8, made 6 $M$ in urea, the smallest non-dialyzable component had a value of $0.81 \times 10^4$ for $M_{app}(1 - \bar{v}\rho)$. Performic acid oxidation[8] of human γ-globulin gave a product with $M_{app} = 32{,}000$ that was slightly soluble in citrate buffer, $p$H 10.9, in the absence of urea.

These findings suggest that human γ-globulin contains subunits linked at least in part by disulfide

(1) R. Trautman and C. F. Crampton, Abstracts of Papers, 130th Meeting, American Chemical Society, Sept. 1956, 0-C; THIS JOURNAL, in press.

(2) R. Trautman, *Biochim. et Biophys. Acta*, 28, 417 (1958).

(3) D. A. Yphantis, *J. Phys. Chem.*, in press.

(4) J. L. Oncley, G. Scatchard and A. Brown, *J. Phys. Colloid Chem.*, 51, 184 (1947).

(5) H. F. Deutsch and J. I. Morton, *J. Biol. Chem.*, 231, 1107 (1958).

(6) J. M. Swan, *Nature*, 180, 643 (1957).

(7) J. F. Pechère, G. H. Dixon, R. H. Maybury and H. Neurath, *J. Biol. Chem.*, 233, 1364 (1958).

(8) C. H. W. Hirs, *ibid.*, 219, 611 (1956)

45

bonds. The possibility that linkages other than disulfide bonds are involved has not been excluded.

THE ROCKEFELLER INSTITUTE
NEW YORK, N. Y. GERALD M. EDELMAN

RECEIVED APRIL 24, 1959

## THE HOMOPOLYMERIZATION OF MONOISOCYANATES

*Sir:*

Beyond the formation of cyclic dimers[1] and trimers[2] from monoisocyanates (Reactions I and II in Fig. 1) and the suggestion that cyamelide[3] is a linear polyisocyanic acid, there has been no evidence to indicate the possibility of obtaining addition polymers from monofunctional compounds containing >C=N– groups. I have recently found that monoisocyanates can be polymerized to linear high molecular weight polymers in a manner similar to vinyl compounds in accordance with the Reaction III in Fig. 1. These polymers may be regarded as N-substituted "1-nylons."

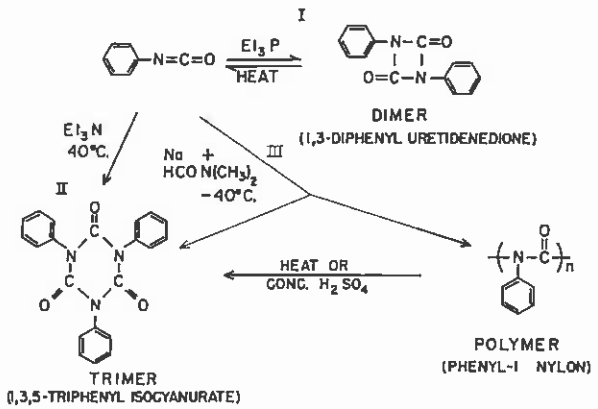



Fig. 1.—Reactions of phenyl isocyanate.

The polymerization takes place at low temperatures (−20° to −100°) in polar solvents, such as N,N-dimethylformamide, N,N-dimethylacetamide and triethylamine, through the use of anionic initiators. Ethyl isocyanate and phenyl isocyanate, examples of an aliphatic and aromatic isocyanate, were polymerized in the manner to be described.

Ethyl isocyanate, 25 ml., and dry triethylamine, 25 ml., both previously cooled to −40°, were stirred together rapidly under dry nitrogen, and further cooled to −100° with liquid nitrogen. Next 10 ml. of a sodium catalyst solution in N,N-dimethylformamide[4] was added to the rapidly stirring solution during about 30 seconds. A white, fibrous solid precipitated immediately upon contact of the catalyst with the monomer solution. The solid was filtered off, washed with methanol and dried to give 8.5 g. (39%) of ethyl 1-Nylon polymer (m.p. 250° with decomposition; found: C, 50.76; H, 6.88; N, 19.34). From the filtrate the cyclic trimer, 1,3,5-triethyl isocyanurate, was isolated and identified by its solubility and melting (m.p. 94°) characteristics.

Ethyl 1-Nylon was soluble in trifluoroacetic acid (TFA) and concentrated sulfuric acid. Clear films were obtained when TFA solutions were air dried and extracted with methanol at room temperature. Prolonged standing of the polymer in TFA resulted in degradation: a change in the inherent viscosity at 0.5% concentration from 0.3 to 0.04 occurred in two days. Light scattering studies on the aged solution gave a molecular weight value of 1600–3000 (*i.e.*, a D.P. of 23 to 42).[5] The polymer was further characterized by its distinct infrared spectrum where the only significant bands were those corresponding to carbonyl at 5.85$\mu$, N,N-disubstituted amide at 7.4$\mu$ and a conspicuous absence of isocyanate and NH bands.[5]

With phenyl isocyanate, a polymer (m.p. 197° with decomposition) was obtained in 86% yield in N,N-dimethylformamide using the sodium catalyst[4a] at −40°. The polymer was soluble in concentrated sulfuric acid, but insoluble in TFA and chlorinated hydrocarbons. Its infrared spectrum was consistent with a phenyl 1-Nylon structure and quite distinct from either the dimer or trimer of phenyl isocyanate.

(5) The author wishes to thank Drs. D. Akeley and R. Zbinden for the light scattering data and infrared analysis, respectively.

PIONEERING RESEARCH LABORATORY
TEXTILE FIBERS DEPARTMENT VICTOR E. SHASHOUA
E. I. DU PONT DE NEMOURS & CO., INC.
WILMINGTON, DELAWARE

RECEIVED APRIL 28, 1959

## STEROIDS. CXXII.[1] 6α-FLUORO-16α-METHYLCORTICOIDS—SYNTHESIS AND BIOLOGICAL ACTIVITY

*Sir:*

Among the most recent chemical modifications of the hydrocortisone and prednisolone molecule—resulting in increased anti-inflammatory activity and diminution of salt retention—have been introduction of a 6α-fluorine atom[1,2,3] *or* addition of a methyl group in the 16α-position.[4,5] We now wish to report the synthesis and preliminary biological evaluation of analogs of the most important cortical hormones combining *both* of these structural features.

The preparation in this Laboratory of 6α-fluoro-16α-methylhydrocortisone acetate (I) *via* 6α-fluoro-16α-methyl "Substance S" already has been recorded.[6] Dehydration of I with mesyl chloride in

(1) A. W. Hofmann, *Ber.*, **3**, 765 (1870).

(2) D. H. Chadwick and T. C. Allen, U. S. Patent 2,733,254 (1956).

(3) W. Kern, H. Paul and W. Mehren, *Makromol. Chem.*, **14**, 146 (1954).

(4) (a) The catalyst solution was prepared by adding 10 drops of a 50% sodium dispersion in xylene (du Pont Electrochemicals Department) to 50 ml. of dry N,N-dimethylformamide. The mixture was stirred for 1 hour before use. (b) The use of this combination as a catalyst for the anionic polymerization for vinyl compounds was demonstrated by L. Grandine, E. I. du Pont de Nemours & Company (private communication).

(1) Paper CXXI, A. Bowers, E. Denot, M. B. Sánchez and H. J. Ringold, *Tetrahedron*, in press.

(2) A. Bowers and H. J. Ringold, THIS JOURNAL, **80**, 4423 (1958).

(3) J. A. Hogg, G. B. Spero, J. L. Thompson, B. J. Magerlein, W. P. Schneider, D. H. Peterson, O. K. Sebek, H. C. Murray, J. C. Babcock, R. L. Pederson and J. A. Campbell, *Chemistry and Industry*, 1002 (1958).

(4) G. E. Arth, J. Fried, D. B. R. Johnston, D. R. Hoff, L. H. Sarett, R. H. Silber, H. C. Stoerk and C. A. Winter, THIS JOURNAL, **80**, 3161 (1958).

(5) E. P. Oliveto, R. Rausser, L. Weber, A. L. Nussbaum, W. Gebert, C. T. Coniglio, E. B. Hershberg, S. Tolksdorf, M. Eisler, P. L. Perlman and M. M. Pechet, *ibid.*, **80**, 4431 (1958).

(6) J. A. Edwards, A. Zaffaroni, H. J. Ringold and C. Djerassi, *Proc. Chem. Soc.*, 87 (1959). All intermediates in that synthesis were

**Determination of amino acid sequences of antibody light chains**

Once the importance of Bence Jones proteins and myeloma proteins from multiple myeloma patients were realized, it became essential to determine the amino acid sequences of Bence Jones proteins as a first step towards the understanding of the detailed structures of antibodies.

Indeed in 1965, the first three amino acid sequences of Bence Jones proteins were determined by Hilschmann and Craig (Hilschmann N, Craig LC (1965) Amino acid sequence studies with Bence-Jones proteins. Proc Natl Acad Sci USA, 53, 1403- 1409). At that time, amino acid sequence determination was quite tedious so that not their entire sequences were reported in this pioneering study. Roy was sequenced completely, while Ag and Cum only partially. Of course, years later, all three were sequenced completely.

They proposed that the C-terminal half of Bence Jones proteins might exist as constant regions of antibody light chains, except for amino acid substitutions at one or two positions. On the other hand, the N-terminal half were quite different and consisted of the variable regions of the antibody light chains. They refrained from drawing conclusions of how the variable regions might be generated, suggesting the process could involve cross-overs, mutations, etc.

After this initial amino acid sequencing of Bence Jones proteins, numerous investigators started todetermine the amino acid sequences of of other human Bence Jones proteins as well the light chains of myeloma proteins from mice. A few years later, massive effort were also initiated to determine the amino acid sequences of heavy chains of myeloma proteins.

When the technique of DNA sequencing was established soon afterwards, nucleotide sequences for light and heavy chains of antibodies from humans and other species were also carried out on massive scale. As discussed later, the intricate mechanism of generating millions of different antibodies from a few gene segments was finally deciphered at the nucleotide sequence level.

This opened up the entire field of analyzing the amino acid and nucleotide sequences of antibodies. Currently, vast amounts of sequence data on antibodies and related proteins are available in the GenBank database. One of the major questions is how antibodies can be so specific in binding designated antigens. The idea of lock and key between antibody and antigen plays a vital role in the eventual engineering design of therapeutic antibodies.

[19] Strumeyer, D. H., W. N. White, and D. E Koshland, Jr., these PROCEEDINGS, 50, 931 (1963).
[20] Kielley, W. W., and L. B. Bradley, *J. Biol. Chem.*, 218, 653 (1956).
[21] Lai, C. Y., O. Tchola, T. Cheng, and B. L. Horecker, *J. Biol. Chem.*, 240, 1347 (1965).

---

# *AMINO ACID SEQUENCE STUDIES WITH BENCE-JONES PROTEINS*

BY NORBERT HILSCHMANN* AND LYMAN C. CRAIG

THE ROCKEFELLER INSTITUTE

*Communicated April 12, 1965*

Precise structural studies with Bence-Jones proteins have assumed much greater significance following the suggestion by Edelman and Poulik[1] that they are polypeptide chains associated with the γ-globulins of patients with multiple myeloma and the confirmation of this suggestion by the findings[2] that these myeloma globulins after reduction and alkylation give bands in starch gel electrophoresis which are at the same position as those given by the Bence-Jones protein (which has been similarly reduced and alkylated) from the same patient.

This concept has been further strengthened by the findings of Edelman and Gally,[3] Putnam,[4] Schwartz and Edelman,[5] and Mannik and Kunkel.[6] It is now accepted that the Bence-Jones proteins correspond in most instances to the light chains of the myeloma protein from the same patient. The light chains are known[7] to be important in antibody structure. Structural differences between different Bence-Jones proteins might not only reveal certain features responsible for antibody specificity but might also yield more detailed information concerning the genetic mechanism which determines the synthesis of light chains of antibodies.

A study of the amino acid sequence of the Bence-Jones proteins from certain selected patients was therefore undertaken. The patients were selected on the basis of minimal heterogeneity with respect to preliminary chemical studies[8] and the physical and immunological studies carried out in the laboratories of Dr. H. G. Kunkel and Dr. G. M. Edelman of this institute.

The most favorable case (Roy) had been shown by Harboe, Osterland, Mannik, and Kunkel[9] to be of the immunological type I (genetic type a+, b−, 1−). It was studied rather intensively from the standpoint of purity by a number of fractionation techniques including countercurrent distribution under dissociating conditions[8] and the amino acid composition established with extensively fractionated material.

While the initial purity studies were in progress, preliminary investigations were under way on the best conditions for enzymatic splitting. For this purpose, protein prepared by ammonium sulfate fractionation only was employed. These results are reported in the present paper. While confirmation by repetition on material purified by countercurrent distribution is planned, the fragmentation results thus far are entirely self-consistent and lead within the experimental error to the over-all analytical composition established on the highly fractionated protein.

*Materials and Methods.*—The enzymes used were obtained from Worthington. The trypsin was pretreated.[10] The digestions (protein:enzyme ratios of 100:1) were made at 38° on 1–1.5 gm of lyophylized protein which had been fully reduced with mercaptoethanol and alkylated with iodoacetic acid in 10 *M* urea according to the directions of Anfinsen and Haber.[11] In the case of

trypsin and chymotrypsin the splitting was made in 100 ml of solution in a pH stat at pH 8 under nitrogen. For the tryptic digestion the uptake of alkali was more rapid during the first 3 hr, then became slower, and by 5 hr had reached 96% of that expected from the lysine and arginine content of the protein. During the digestion the originally clear solution formed a precipitate which amounted to about $^1/_3$ of the total weight of the protein.

In the case of chymotrypsin the originally clear solution gave no precipitate. The digestion was interrupted after $3^1/_2$ hr.

The pepsin digestion was carried out in 30% acetic acid at 38° for 18 hr. The solution, turbid at the beginning, gave more insoluble material as the digestion proceeded.

The isolation of the peptides from the digest was accomplished by an initial fractionation with a column packed with the anion exchange resin Dowex 1 × 2,[13] as shown in Figure 1. A gradient elution system was used in which the initial buffer was an aqueous solution containing 1% pyridine, 1% alpha picoline, 1% 2,4-lutidine, and sufficient acetic acid to reach a pH of 8.5. The relatively linear decrease in pH throughout the elution as shown in Figure 1 was accomplished by progressive addition of acetic acid to the stirred solution in the buffer reservoir in the well-known way. The concentrations of acetic acid used for this purpose were changed at various times during the elution as shown along the abscissa of the pattern. The column used was 150 cm in length and 2 cm in diameter. The buffer reservoir held 1 liter of buffer. Fractions containing 10 ml were collected. A ninhydrin[14] color determination was made on each fraction and again following alkaline hydrolysis, but only the latter are shown in Figure 1.

It will be noted that the chart is discontinuous at tubes 430 and 465. This is because the particular run was not good in this section due to buffer change. A short section is therefore substituted from a second more favorable run in order to simplify the presentation.

The whole digest was placed at the top of the column even though, except for chymotrypsin, it was a suspension. The suspension was mixed with a little of the resin to prevent blocking. In the case of pepsin the digest was brought to dryness by rotary evaporation and lyophilization before being suspended in the buffer. As the pH of the buffer decreased, the precipitate became soluble and was then carried into the column, emerging in the later fractions.

Each separate peak or a cut of it was recovered by rotary evaporation and rechromatographed on either Dowex 50 × 2 or phosphorylated cellulose[14] or both with pyridine formate or pyridine acetate as the eluting buffer until a single peak was obtained which gave a satisfactory amino acid analysis. A certain solubility difficulty, encountered with the peptides in peaks 14 and 15 of Figure 1, could be overcome by adding n-propanol to the buffer up to a concentration of 30%. This permitted the largest tryptic peptide containing 40 residues to be separated as a single peak.

Material recovered from each peak was hydrolyzed at 108° for 24 hr with 6 *N* hydrochloric acid in evacuated sealed glass tubes. The amino acid composition was then determined with a recording Moore-Spackman-Stein amino acid analyzer.[15] A peptide was considered of satisfactory purity only when its analytical composition was consistent with whole ratios of amino acid residues and a reasonable molecular size.

*Results.*—Amino acid analysis of the purified protein from Roy[8] showed 14 lysine residues and 5 arginine residues. Since neither of these amino acids is at the carboxyl terminal position, complete digestion with trypsin should result in a total of 20 peptides. These 20 peptides have all been isolated. Their composition is given in Table 1. In the right-hand columns of this table are listed the sum of the individual amino acids in each peptide and that obtained[8] by analysis of the protein after purification by countercurrent distribution. Thus, all the expected peptides can be accounted for to give a total of 212 amino acid residues. (Certain of the analytical data for the intact protein, as well as for T18, indicate this peptide to contain an additional glutamic acid residue. If so, the protein would contain 213 residues.)

The peptide number arbitrarily assigned during the work is given at the top horizontal row of Table 1, and the cut or peak number from Figure 1 is given in the second row. Edman degradations were made on each peptide with the reaction modification of Konigsberg and Hill[16] and the subtractive method of Hirs, Moore,

TABLE 1

**Peptide compos tion of the tryptic peptides from Bence-Jones protein Cummings. The amino terminal amino acid is indicated by an asterisk.**

| Peptide number | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 | T9 | T10 | T11 | T12 | T13 | T14 | T15 | T16 | T17 | T18 | T19 | T20 | Number of residues in | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cut number | 1 | 1 | 2 | 3 | 3 | 4 | 5 | 6 | 7 | 8 | 8 | 9 | 9 | 10 | 11 | 12 | 13 | 15 | 15 | 14 | Sum of T1-T20 | Protein |
| Lys | | 1 | 1 | | 1 | 1 | | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | | 1 | 1 | 1 | | 14 | 14 |
| His | | | | | 1* | | | | | | | | 1 | | | | | | | | 2 | 2 |
| Arg | 1 | | | 1 | | | 1 | | | | | 1 | | | | 1 | | | | | 5 | 5 |
| Cys | | | | | | | | | | | | | 1 | | | 1 | | 1 | 1 | 1 | 5 | 5 |
| Asp | | | | 1 | | | | | 1 | 1* | 1 | 2* | | 1 | 1 | 2 | 4 | 4 | 2 | | 20 | 20 |
| Thr | | | | | | | | | | 3 | | 1 | 2 | | 1* | 1 | 1 | 6 | 2 | | 17 | 17 |
| Ser | | | | 1* | | | 1 | | | 5 | 1 | 5 | 2 | | 2 | 2* | 4 | 4 | 2 | | 29 | 29 |
| Glu | | | | | | 1 | 1 | 1* | | | | 2 | 2 | 1 | 2 | | 5 | 4 | 4 | 1 | 24 | 24 |
| Pro | | | 1 | | | | 1 | | | | | 1 | 1 | | 3 | 1 | | 2 | 1 | | 11 | 11 |
| Gly | | | | | | | 1 | | | | | 1 | 1 | | | 1 | 1 | 6 | 1 | 1* | 13 | 13 |
| Ala | | | 1* | | | | 1 | 1 | | | 1 | 1 | 1 | 1* | 2 | 1 | 1 | 1 | 1 | | 13 | 13 |
| Val | | | | | | 1* | 1 | | 1* | | | 1 | 2 | | 2 | 2 | 2* | | 1 | | 13 | 13 |
| Met | | | | | | | | | | | | 1 | | | | | | | | | 1 | 1 |
| Ileu | | | | | | | | | | | 1 | 1 | | | 1 | | | 2 | 3 | | 8 | 8 |
| Leu | | | | | | | 1* | | | 3 | 2* | 1 | 2* | | 1 | 2 | 1 | 3 | 1 | | 17 | 17 |
| Tyr | | | | | | | | | | 1 | 1 | | 1 | 1 | | 1 | | 2 | 1 | | 8 | 8 |
| Phe | | | | 1 | | | | | 1 | | | | | | 2 | 1 | | 4* | 1 | | 10 | 10 |
| Try | | | | | | 1 | | | | | | | | | | | | | 1 | | 2 | 2 |
| Number of residues | 1 | 1 | 3 | 4 | 2 | 4 | 8 | 3 | 4 | 14 | 8 | 18 | 17 | 5 | 18 | 16 | 20 | 40 | 23 | 3 | 212 | 212 |

and Stein.[17] The amino terminal amino acid of each peptide is designated in Table 1 by an asterisk.

The carboxyl terminal amino acid in each peptide was determined enzymatically with carboxypeptidase B[18] (Worthington) or in a few cases by hydrazinolysis.[19] The amino acid split off (or remaining) was quantitatively determined in the amino acid analyzer. The carboxyl terminal amino acid determined in this way for every peptide in Table 1 was lysine or arginine, as would be expected from the specificity of trypsin except for T20. It has cysteine as the carboxyl terminal amino acid.

The elution profiles for the chymotryptic and peptic digests indicated greater complexity than in Figure 1. In spite of this, cuts were taken and refractionated on Dowex 50 × 2 or phosphorylated cellulose until satisfactory analyses were obtained with the most favorable final cut. Figure 2 shows only those peptides which have proved important in arranging the order of the tryptic peptides, but a considerable number of others of satisfactory purity were also isolated. They are not discussed here because of lack of space, but some of them add support to the assignments given.

On the basis of the data we have thus far accumulated, the tryptic peptides can be ordered with considerable certainty as given in Figure 2. Here the tryptic peptides are designated by T1, T2, etc., with this designation placed above the carboxyl terminal residue. A line with Ch 1, Ch 2, etc., gives the position of the corresponding chymotryptic peptide and a line with P1, P 2, etc., for the peptic peptides. A rectangle enclosing a residue indicates that position to be known in the

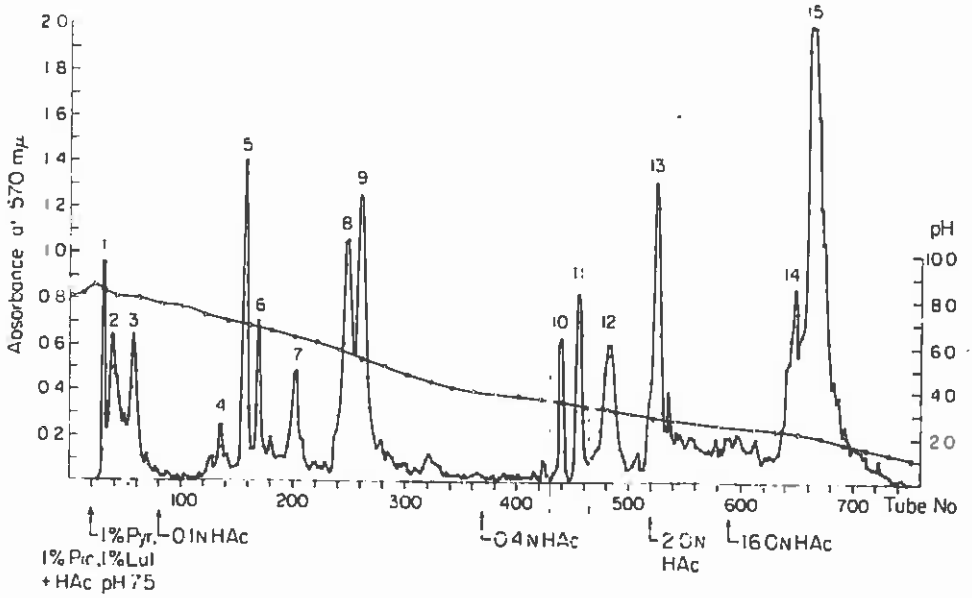



FIG. 1.—Elution pattern of the ion exchange chromatography of the tryptic digest of Bence-Jones protein Roy. ●—● = absorbance. O—O = pH.

chain. Those not enclosed are undetermined sequences, although their position in the chain is restricted by the peptides which contain them.

This partial formula is consistent with the findings of Putnam[20] that aspartic acid is the amino terminal residue found only in Bence-Jones proteins of antigenic group I. Direct determination of the amino terminal residue of intact protein (Roy) also gave confirmation.[21] It is further consistent with the results of Milstein[22] regarding the carboxyl terminal residue. Moreover, the peptic peptide isolated by him

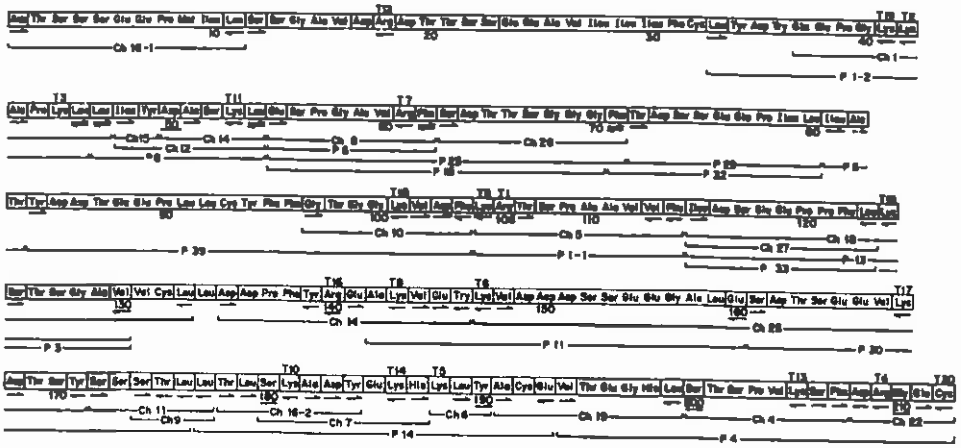


FIG. 2.—Proposed order of the tryptic peptides and partial amino acid sequence for the Bence-Jones protein Roy. A residue enclosed indicates the position is established in the chain. The symbols → and — indicate an Edman and carboxypeptidase or hydrazinolysis determination, respectively. *T 1*, *T 2*, etc. designates a tryptic peptide, *Ch* a chymotryptic peptide, and *P* a peptic peptide. The sequences in *T 12* can be assigned by analogy from the work of Titani and Putnam[22] on Ag.

51

which contains an S-S bridge must, from its sequences, connect the cysteine residues in positions 132 and 192 (Roy numbering).

*Bence-Jones Protein Cummings:* In order to demonstrate differences in the primary sequences of different Bence-Jones proteins, several others from different patients were studied simultaneously. The study of one of these (Cummings), although much less advanced than that for Roy, is sufficiently clear to be of great interest in relation to Roy. This protein (antigenic type I, genetic type a−, b+, 1−)[9] showed an individual antigenic specificity which differed markedly from Roy[28] and also gave an amino acid analysis considerably different. It was therefore only to be expected that marked differences in the primary structure would be found.

So far 18 out of the 21 expected tryptic peptides have been isolated and characterized in terms of amino acid composition. They are listed in Table 2. Ten of these (Tc20, Tc4, Tc5, Tc14, Tc10, Tc17, Tc6, Tc8, Tc15, Tc13) have their counterparts in Roy. Tc13 differs only in a single amino acid replacement (Val for Leu). Moreover, evidence was obtained from the chymotryptic peptides that these all had the same order as in Roy. Tc16, a peptide not given in Table 2 because its amino acid analysis still indicated impurity, could be located by the isolation of a chymotryptic peptide which contains most of the sequence. It seems likely, therefore, that the entire sequence in Roy from 106 to 212 is also present in Cummings except for the single Val-Leu replacement at position 189.

The amino terminal amino acid in Cummings was found to be glutamic acid.[21] This terminal residue was also found in a peptide, Tc12, whose purity is not yet entirely satisfactory but which contains a cysteine and one of the two methionines. Roy contains only one methionine.

TABLE 2

Peptide composition of the tryptic peptides from Bence-Jones protein Roy. The amino terminal amino acid is indicated by an asterisk.

| | Tc 1 | Tc 4 | Tc 5 | Tc 6 | Tc 7 | Tc 8 | Tc 9 | Tc 10 | Tc 13 | Tc 14 | Tc 15 | Tc 17 | Tc 18a | Tc 18b | Tc 18c | Tc 18d | Tc 20 | Tc 19+2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lys | | | 1 | 1 | | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | | | 1 | | 1 |
| His | | | 1 | | | | | | 1 | | | | | | | | | |
| Arg | 1 | 1 | | | 1 | | 1 | | | | | | | 1 | 1 | | | |
| Asp | | 1 | | | 1 | | | 1 | | 1 | 1 | 4 | 1 | | 1 | | | 4 |
| Thr | | | | | | | | 3 | 2 | | 1 | 1 | 2 | | | 2 | | 1 |
| Ser | | 1 | | | 1 | | | 5 | 2 | | 2 | 4 | 3 | 1 | | | | 4* |
| Glu | | | | 1 | | 1 | 1 | | 2 | 1 | 2 | 5 | | | 3 | 2 | 1 | 2 |
| Pro | | | | | 1 | | | | 1 | | 3 | | | | | 1 | | |
| Gly | | | | | 1 | | | | 1 | | | 1 | 3 | | 1 | 2 | 1 | 2 |
| Ala | | | | | 1* | 1 | | | 1 | 1 | 2 | 1 | | | 1 | | | |
| Half Cys | | | | | | | | | 1 | | | | | | 1 | | 1 | |
| Val | | | | 1 | 1 | | | | 3* | | 2 | 2 | | | 3* | | | |
| Met | | | | | | | | | | | | | | | 1 | | | |
| Ileu | | | | | | | 1 | | | | 1 | | | 1* | | 1 | | |
| Leu | | | | | | | 1* | 3 | 1 | | 1 | 1 | 1 | | | 1* | | 4 |
| Tyr | | | | | | | | 1 | 1 | 1 | | | | | 2 | 1 | | 2 |
| Phe | | 1 | | | | | | | | | 2 | | 2* | | | 1 | | |
| Try | | | | 1 | | | | | | | | | | | | | | 1 |

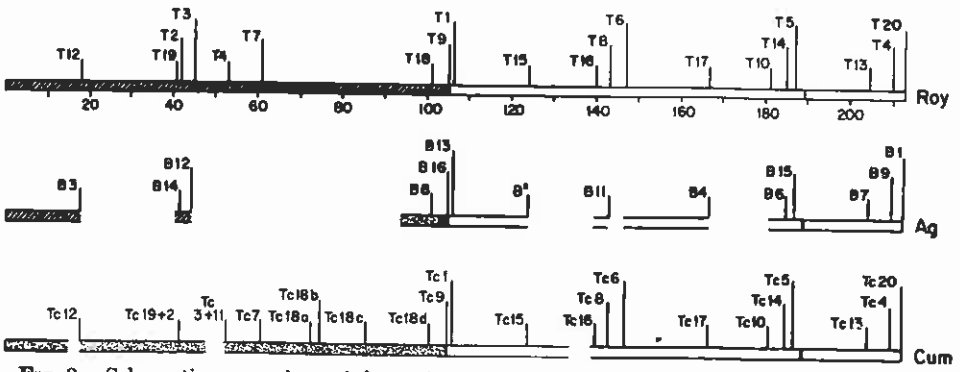



FIG. 3.—Schematic comparison of the positions of the known tryptic peptides from three different patients, Roy, Ag, and Cummings. ═══ = Regions of identity. ▨▨▨ = Regions of identity for Roy and Ag only. ▩▩▩ = Regions of identity for Cummings and Ag only. Residue 189 Val in Ag and Cum., Leu in Roy. $B^2$ in personal communication from Dr. Putnam.

The overlapping chymotryptic and peptic peptides isolated from Cummings are sufficient to establish the order Tc12-Tc19- and -Tc7-Tc18a-Tc18b-Tc18c-Tc18d-Tc9-Tc1-Tc15-. Although the amino acid composition of these peptides is markedly different from any of those from Roy, certain similarities between these corresponding regions can be seen. For instance, in both cases in the amino terminal 11 residues of T18 and Tc18a, the first residue is phenylalanine and the last sequence is PheThr. The amino acids enclosed are the same except for a Thr-Ser exchange.

Both proteins have a tryptophan-containing peptide which also contains Leu, Tyr, and Asp, in corresponding positions. Another similarity occurs in the 95–101 region of Roy except for a Glu-Gly replacement. T9 of Roy is substituted in Cummings by Tc9 which has the sequence -Leu Glu Ileu Arg-. Similarities of this kind were noted when the $\alpha$ and $\beta$ chains of human hemoglobin[24, 25] were compared.

The two cysteines in the amino terminal half of Cummings are confined to approximately the same position as in Roy. Thus, the approximate ring structures imposed by the S-S bridges suggested by Milstein[23] may be similar.

*General nature of sequence variation:* There are presently available data from three different Bence-Jones proteins (Roy, Cummings, and Ag[26, 27]) on which to base comparisons, although establishment of the sequences in Cummings and Ag are less advanced. A comparison of these can be shown graphically as in Figure 3. The positions of the known tryptic peptides in Roy and Cummings for residues 106–212 (Roy numbering) seem to be identical. Eight of the tryptic peptides reported for Ag (B13, B11, B4, B6, B15, B7, B9, B1)[26, 27] have the same amino acid composition as counterparts in Roy or Cummings, and it is likely that these have the same order. This supports the conjecture that this part of the chain in Bence-Jones proteins of Type I are identical or nearly so with perhaps single replacements of the type found at sequence 189 (Val-Leu) (Roy numbering). It should be noted that the proteins compared here are not of the same genetic type.

On the other hand, the tryptic peptides of the amino terminal half of Roy have no counterpart among those found for Cummings. In spite of this fact, the positions of two of the remaining tryptic peptides (B14, B12) reported for Ag[26, 27] can be assigned by analogy. B3 was shown to be the amino terminal peptide. From its composition and sequence it is identical with T12 of Roy. B8 and B16 cannot be

found in Roy. However, there is a counterpart in a chymotryptic product of Tc18d and in Tc9 of Cummings, except for a Lys-Arg interchange in Tc9.

The peculiar distribution of these peptides of Ag between Roy and Cummings in the amino terminal half of the protein clearly indicates that the parts in common between Bence-Jones proteins show a considerable amount of variation. Recently, Smithies[28, 29] has proposed a hypothesis for the variability of $\gamma$-globulin composition. According to this hypothesis, individual differences in immunoglobulins might be explained by crossing over between genes determining the synthesis of these proteins. This crossover might even occur as a somatic mutation. Data presented here are too fragmentary for drawing valid conclusions. Further investigations along these lines, however, would seem to be very interesting, especially in connection with the problem of somatic mutations.

The technical assistance of Penelope Ryan, Connie Mahoney, and Gerhard Bitterlich is gratefully acknowledged.

This work was supported in part by a grant (A-02493) from the National Institutes of Health.

* Recipient of a two-year fellowship from the Fritz Thyssen Fund. Present address: Medizinische Forschungsanstalt der Max-Planck-Gesellschaft, Göttingen, Germany.

[1] Edelman, G. M., and M. D. Poulik, *J. Exptl. Med.*, 113, 861 (1961).
[2] Poulik, M. D., and G. M. Edelman, *Nature*, 191, 1 (1961).
[3] Edelman, G. M., and J. A. Gally, *J. Exptl. Med.*, 116, 207 (1962).
[4] Putnam, F. W., *Biochim. Biophys. Acta*, 63, 539 (1962).
[5] Schwartz, J. H., and G. M. Edelman, *J. Exptl. Med.*, 118, 41 (1963).
[6] Mannik, M., and H. G. Kunkel, *J. Exptl. Med.*, 116, 859 (1962).
[7] Edelman, G. M., and J. A. Gally, these PROCEEDINGS, 51, 846 (1964).
[8] Hilschmann, N., and L. C. Craig, *Biochemistry*, 4, 5 (1965).
[9] Harboe, M., C. K. Osterland, M. Mannik, and H. G. Kunkel, *J. Exptl. Med.*, 116, 719 (1962).
[10] Redfield, R. P., and C. B. Anfinsen, *J. Biol. Chem.*, 221, 385 (1956).
[11] Anfinsen, C. B., and E. Haber, *J. Biol. Chem.*, 236, 1361 (1961).
[12] Rudloff, V., and G. Braunitzer, *Z. Physiol. Chem.*, 323, 129 (1961).
[13] Moore, S., and W. H. Stein, *J. Biol. Chem.*, 211, 907 (1954).
[14] Canfield, R. E., and C. B. Anfinsen, *J. Biol. Chem.*, 238, 2684 (1963).
[15] Spackman, D. H., W. H. Stein, and S. Moore, *Anal. Chem.*, 30, 1190 (1958).
[16] Konigsberg, W., and R. J. Hill, *J. Biol. Chem.*, 237, 2547 (1962).
[17] Hirs, C. H. W., S. Moore, and W. H. Stein, *J. Biol. Chem.*, 235, 633 (1960).
[18] Guidotti, G., *Biochim. Biophys. Acta*, 42, 177 (1960).
[19] Akabori, S., K. Ohno, and K. Narita, *Bull. Chem. Soc. (Japan)*, 25, 214 (1952).
[20] Putnam, F. W., S. Migita, and C. W. Easley, *Protides Biol. Fluids, Proc. Colloq.*, 10, 93 (1962).
[21] Determined by the isocyanate method of G. R. Stark and D. G. Smyth [*J. Biol. Chem.*, 238, 214 (1963)].
[22] Milstein, C., *J. Mol. Biol.*, 9, 846 (1964).
[23] Personal communication from Dr. Kunkel.
[24] Konigsberg, W., G. Guidotti, and R. J. Hill, *J. Biol. Chem.*, 236, PC 55 (1961): Konigsberg, W., J. Goldstein, and R. J. Hill, *J. Biol. Chem.*, 238, 2028 (1963).
[25] Braunitzer, G., R. Gehring-Müller, N. Hilschmann, K. Hilse, G. Hobom, V. Rudloff, and B. Wittmann-Liebold, *Z. Physiol. Chem.*, 325, 283 (1961).
[26] Putnam, F. W., C. W. Easley, and J. W. Helling, *Biochim. Biophys. Acta*, 78, 231 (1963).
[27] Titani, K., and F. W. Putnam, *Science*, 147, 1304 (1965).
[28] Smithies, O., *Nature*, 199, 1231 (1963).
[29] Smithies, O., in *Human Genetics*, Cold Spring Harbor Symposia on Quantitative Biology, vol. 29 (1964), p. 309.

**Variability plot and the complementarity determining regions (CDR's)**

Five years later, there were about 80 complete or partial known amino acid sequences of the variable regions of Bence Jones proteins and myeloma light chains available in the literature from humans and mice. A simple logical argument suggested that there should be extensive amino acid residue changes at positions in these variable regions in order to bind different antigens. With the assumption that all variable regions should fold similarly, the variable region amino acid sequences were thus aligned with Cys at identical positions. In addition, some other amino acid residues could also be aligned, such as Trp, Gly ( ) Gly, etc.

Previously, some segments of amino acid residues were noted to be quite variable. However, it was essential to provide a quantitative measure rather than hand waiving. After alignment, for each position, a number known as variability was calculated. It was defined as the ratio of the number of different amino acid residues at that position divided by the frequency of the most common amino acid residues. Theoretically, for an invariant position, variability is equal to 1 / 1 = 1. On the other hand, for a theoretical maximum, variability is equal to 20 / 0.05 = 400. Of course this theoretical maximum was never reached, but provides a sufficiently large range, from 1 to 400, to look for segments in the variable region of antibody light chains with relatively high number of amino acid residue substitutions.

This idea of plotting variability against positions of the variable region of antibody light chains was the joint effort of a world famous immunologist and a person trained in medicine, engineering, and applied sciences (Wu TT, Kabat EA (1970) An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J Expt Med, 132, 211-250; also re-published in 2008 as a Pillars Article. J. Immunol, 180, 7057-7096).

This study also pioneered the idea of bio-informatics. In other words, the information content of aligned amino acid sequences of not only Bence Jones proteins but also any other protein could provide useful clues to the functions of these proteins not available from the amino acid sequence of one of them.

In the plot of variability for the variable regions of antibody light chains, three peaks became apparent. They were initially designated as hyper-variable regions forming the part of the antibody combining site. Currently, these short segments are known as complementarity determining regions (CDR's).

J. I. (2008), 180, 7057-7096

# AN ANALYSIS OF THE SEQUENCES OF THE VARIABLE REGIONS OF BENCE JONES PROTEINS AND MYELOMA LIGHT CHAINS AND THEIR IMPLICATIONS FOR ANTIBODY COMPLEMENTARITY*

By TAI TE WU, Ph.D., and ELVIN A. KABAT, Ph.D

*(From the Departments of Microbiology, Neurology, and Human Genetics and Development, College of Physicians and Surgeons, Columbia University, and the Neurological Institute, Presbyterian Hospital, New York 10032; the Biomathematics Division, Graduate School of Medical Sciences, Cornell University and the Sloan-Kettering Institute, New York 10021)*

(Received for publication 26 March 1970)

The extraordinary versatility of the antibody-forming mechanism in producing an almost limitless number of specific receptor sites complementary for almost any molecular conformation of matter within a size range (1–3) represented by a hexa- or heptasaccharide as an upper and a mono- or disaccharide as a lower limit, is almost certainly related to the unique structural features of immunoglobulins and differentiates them from all other known proteins. These antibody-combining sites are formed as a consequence of the interaction of two polypeptide chains, a light and a heavy chain (2, 4, 5). The antibodies usually formed to various antigens often represent heterogeneous populations of immunoglobulin molecules of different classes, subclasses, and genetic variants and also show specificities toward different antigenic determinants (1, 2, 6, 7). In some instances, however, relatively homogeneous populations of antibodies with respect to many of these properties have been obtained. Among these have been human antibodies to dextran and levan (8, 9) and rabbit antibodies to the group-specific carbohydrate of streptococcus (10–12), antibodies to the Type III–specific capsular polysaccharide of pneumococcus (13, 14), rabbit antihapten (15), and specimens of antibodies and of Fab′ fragments which crystallized (Nisonoff et al., in references 16, 17), but sequence data on these are not yet available.

The large body of sequence data related to immunoglobulin structure comes from the analysis of urinary Bence Jones proteins and from the monoclonal immunoglobulins found in large amounts in the sera of patients with multiple myeloma and Waldenström macroglobulinemia (16, 18). While a substantial body of evidence was available relating these proteins to immunoglobulins, the recent demonstration that many myeloma globulins have specific ligand-binding properties like those of many antibodies provides increasing confidence that myeloma globulins represent homogeneous populations of antibody molecules (16, 18–27). The ability to produce in BALB/c

* Aided by grants from the National Science Foundation (GB-8341) and the National Cancer Institute (CA-08748), and a general Research Support Grant of the U. S. Public Health Service.



Downloaded from www.jem.org on November 2, 2004

mice myelomas and macroglobulinemias (28) which produce myeloma globulins and Bence Jones proteins like those in the human, provides a source of data from which important evolutionary trends can be inferred.

Thus the extensive sequence data on Bence Jones proteins, which are considered to be light chains of myeloma globulins and Waldenström macroglobulins (29), and on various light and heavy chains, provide information clearly pertinent to the problem of the elucidation of the structure of antibody-combining sites.

The unique finding that distinguishes the immunoglobulins from all other proteins is that the N-terminal half of the light chains and the N-terminal quarter of the heavy chains vary in sequence in samples obtained from individual monoclonal immunoglobulins and that indeed no two such variable regions of any chain and no two myeloma immunoglobulins or Bence Jones proteins have thus far been found to be identical in sequence (30). The constant region, however, is essentially no different from other proteins in that the variation in the amino acids found at any position is ascribable to species and class variations or to genetic variants such as Inv factors. By comparison of sequence data on the variable and constant regions of Bence Jones proteins with amino acid composition of purified human antibodies, it could be shown that most of the compositional variation could only originate in the variable region (see Kabat in reference 18).

From sequence data, a variety of hypotheses have been advanced (7, 31–35) to explain the structural basis of antibody complementarity. All of these are selective theories, i.e. they consider that the information for complementarity is essentially built into the primary sequence of each chain and that a given antigen only triggers the biosynthesis of those antibody molecules having complementary receptor sites. There are two types of selective theories: germ line theories (36) and somatic mutation theories (37–39). At present no hypothesis is generally accepted. Excellent reviews (see above) are available.

The present communication is an extension of earlier efforts from this laboratory (18, p. 87, and 40–43) to locate more precisely those portions of the variable region which are directly responsible for antibody complementarity, that is which make direct contact with the antigenic determinant, and to explain the unique capacity of these proteins to have so many complementary regions.

As in the earlier studies, all human κ, human λ, and mouse κ Bence Jones protein and light chain sequences are aligned for maximum homology (44) and all variable regions are considered as a unit and compared with the constant regions. These earlier studies had called attention to the following:

(*a*) The variable regions had few if any species-specific positions while the constant regions of the human and mouse proteins had 36 species-specific amino acid substitutions per 107 residues (40, 45). A species-specific position is defined as one at which the amino acid residues in the mouse proteins differ from those in the human proteins.

(*b*) When the invariant residues of these two regions were compared, the latest tabulation (45) showed the variable regions to have 10 invariant and almost invariant glycines and no invariant alanines, leucines, valines, histi-

Downloaded from www.jem.org on November 2, 2004

dines, lysines, or serines while the constant regions had 3 each of invariant alanine, leucine, and valine, and 2 invariant histidines, 2 invariant lysines, and 5 invariant serines. It was suggested that the invariant glycines were important in contributing to the flexibility needed by the variable region in accommodating the numerous substitutions (41, 43) at the variable positions. It was also suggested that the invariant glycines near the end of the variable region at positions 99 and 101, plus the almost invariant glycine at position 100, provided a pivot upon which the complementarity-determining regions might move to make better contact with the antigenic determinant (43; 18, p. 87) just as the walls of the lysozyme site have been shown to adjust somewhat to accommodate its hexasaccharide substrate (46). The hydrophobic residues in the constant region were hypothesized to be involved in noncovalent bonding to the heavy chain.

(*c*) From an examination of sequences of the κI, κII, and κIII subgroups (Hood et al. in reference 16) (47, 48) of the human Bence Jones proteins in which many of the proteins in a subgroup had an identical sequence for the first 20–24 residues, it was postulated that there are two kinds of residues in the variable regions, those making direct contact with the antigenic determinant (complementarity determining) and those which are involved only in three-dimensional folding (42). The latter would be expected to have less stringent requirements, and more mutation noise would be permitted than with the complementarity-determining residues. This distinction led to the inspection of the sequences for short stretches showing very high variability and two of these were identified: the most variable beginning at residue 89 and ending at 97, the other running from residue 24 through 34. Each of these two unusually highly variable regions began after an invariant half-cystine and was followed by an invariant phenylalanine (residue 98) and an invariant tryptophane (residue 35) respectively. It is of interest that the two regions are brought close together by the S—S bond $I_{23}$–$II_{88}$ (45). Milstein (47), Milstein and Pink (7), and Franěk (49) have also called attention to the highly variable positions in these regions and Franěk (49) has noted an additional highly variable region around residues 52–55. It was hypothesized (45) that these first two regions might represent the complementarity-determining regions and that complementarity might be acquired by the insertion of small linear sequences into the light and heavy chains by some episomal or other insertion mechanism. It is striking that the differences in chain length seen in the Bence Jones proteins are confined to these two regions of the chain. The remaining portions of each chain would be essentially under the control of structural genes. The inserted sequences would be drawn from a large but finite set and either inserted under the influence of antigen, if antibody-forming cells are multipotent, or individual sequences might be distributed to immunoglobulin-

Downloaded from www.jem.org on November 2, 2004

forming cells during differentiation if the capacity of individual cells to synthesize antibody is restricted.

This working hypothesis offers several advantages:

(*a*) It is capable of providing the evolutionary stability and accounts for the universality of the antibody-forming mechanism throughout the vertebrates. Germ line theories (34–36) postulate one gene for each of the thousand or more variable regions (30). This would be expected to result in divergence during evolution since the loss by mutation of any one variable region would only minimally affect the capacity to form antibody and survival; thus individuals and populations lacking certain variable regions would arise.

(*b*) It offers a substantial simplification to the problem of producing a very large number of complementary sites. While it is known that in all proteins with specific receptors the site is formed by residues from widely separated portions of the chain, these sites are all formed by single chains. Thus, forming a three-dimensional site must involve residues from various regions. The antibody site being formed by a heavy and a light chain need not necessarily be so restricted.

Since much additional data on the light chains and a number of heavy chain sequences have been accumulated, the present communication represents a further attempt at analyzing the unique features of the variable regions of immunoglobulin chains. Among aspects considered are the role of glycine, invariant residues, and hydrophobicity patterns, and the highly variable portions, with a view to localizing the regions responsible for complementarity and evaluating various theories in terms of evolutionary origin and perpetuation of the antibody-forming mechanism.

*Sequence Data Employed*—Complete and partial sequence data have been published on 77 Bence Jones proteins and immunoglobulin light chains as well as on a number of heavy chains. Data were available on 24 human κI, 4 human κII, 17 human κIII, 10 human λI, 2 human λII, 6 human λIII, 5 human λIV, 2 human λV, 2 mouse κI, and 5 mouse κII proteins.[1]

The original light chain sequence data may be found in the following references.

HBJ 98: Baglioni, C. 1967. *Biochem. Biophys. Res. Commun.* **26**:82.

Eu: Cunningham, B. A., P. D. Gottlieb, W. H. Konigsberg, and G. M. Edelman. 1968. *Biochemistry.* **7**:1983.

Mil (human κII): Dreyer, W. J., W. R. Gray, and L. Hood. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32**:353.

Hac, Dob, Pal: Grant, A., and L. Hood. Unpublished work.

Roy, Cum: Hilschman, N., and L. C. Craig. 1965. *Proc. Nat Acad. Sci. U. S. A.* **53**:1403; Hilschmann, N. 1967. *Hoppe-Seyler's Z. Physiol. Chem.* **348**:1077; Hilschmann, N., H. U. Barnikol, M. Hess, B. Langer, H. Ponstingl, M. Steinmetz-Kayne, L. Suter, and S. Watanabe. 1968. *Fed. Eur. Biochem. Soc. Symp., 5th.* In press.

[1] The World Health Organization has recently changed the notation of subgroups so that human κII in this paper will become human κIII and human κIII will become human κII.

Downloaded from www.jem.org on November 2, 2004

HS 78, HS 92, HS 94, HS 68, HS 70, HS 77, HS 86, HS 24: Hood, L., and D. Ein. 1968. *Nature (London)*. **220**:764.

HBJ 7, HBJ 11, HBJ 2, HBJ 8: Hood, L., W. R. Gray, and W. J. Dreyer. 1966. *J. Mol. Biol.* **22**:179.

MBJ 41, MBJ 70, MBJ 6: Hood, L., W. R. Gray, and W. J. Dreyer. 1966. *Proc. Nat'l Acad. Sci. U. S. A.* **55**:826.

HBJ 10, HBJ 1, HBJ 4, HBJ 6, HBJ 5, HS 4, HBJ 12, HS 6, HBJ 15: Hood, L., W. R. Gray, B. G. Sanders, and W. J. Dreyer. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32**:133.

Ste: Edman, P., and A. G. Cooper. 1968 *Fed. Eur. Biochem Soc. Letters.* **2**:33; Hood, L., and D. W. Talmage. 1969. *In* Developmental Aspects of Antibody Formation and Structure. Prague. In press.

Lay, Mar, Ioc, Wag, How, Koh: Kaplan, A. P. and H. Metzger. 1969. *Biochemistry.* **10**: 3944.

New, III, Mil (human λIV): Langer, B., M. Steinmetz-Kayne, and N. Hilschmann. 1968. *Hoppe-Seyler's Z. Physiol. Chem.* **349**:945.

BJ, Ker: Milstein, C. 1966. *Biochem. J.* **101**:352.

Rad, Fr4: Milstein, C. 1967. *Nature (London)* **216**:330.

X: Milstein, C. 1968. *Biochem. J.* **110**:631.

Bel, Man, B6: Milstein, C. 1968. *Fed. Eur. Biochem. Soc. Symp. on γ-globulin*, Prague.

Day, MBJ46, Roy: Atlas of Protein Sequence and Structure, M. O. Dayhoff, Editor. 1969.

Mz: Milstein, C., B. Frangione, and J. R. L. Pink. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32**:31.

Ale, Car, Dee: Milstein, C., C. P. Milstein, and A. Feinstein. 1969. *Nature (London)* **221**:151.

Cra, Pap, Lux, Mon, Con, Tra, Nig, Win, Gra, Cas, Smi: Niall, H., and P. Edman. 1967. *Nature (London)* **216**:262.

MOPC 149, AdjPC 9, MOPC 157: Perham, R., E. Appella, and M. Potter. 1966. *Science (Washington)* **154**:391.

Kern: Ponstingl, H., M. Hess, and N. Hilschmann. 1968. *Hoppe-Seyler's Z. Physiol. Chem.* **349**:867.

Tew: Putnam, F. W. 1969. *Science (Washington)*. **163**:633.

Ag, Ha, Bo, Sh: Putnam, F. W., K. Titani, M. Wikler, and T. Shinoda. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32**:9; Titani, K., T. Shinoda, and F. W. Putnam. 1969. *J. Biol. Chem.* **244**:3550.

TI: Suter, L., H. U. Barnikol, S. Watanabe, and N. Hilschmann. 1969. *Hoppe-Seyler's Z. Physiol. Chem.* **350**:275.

The accumulation of such large numbers of sequences makes it possible to use statistical criteria in defining the types of residues. Thus in earlier studies, an invariant residue was rigidly defined, e.g., a position at which all samples showed the same amino acid residue sometimes allowing a single exception. The definition of an invariant residue used in this paper is taken as a position at which 88–90% or more of the samples contain the same amino acid. This may allow potential functions to be recognized despite possible errors or uncertainties in sequence, or occasional substitutions compatible with function.

A summary of the sequence data is provided in Table I which lists the amino acids found at any position in any subgroup of human κ-, human λ-, and mouse κ-chains, the number of times each occurs, and the total number of sequences

Downloaded from www.jem.org on November 2, 2004

TABLE I

*Amino Acids Found at each Position in the Variable Region of the Various Subgroups of Human κ-, Human λ- and Mouse κ-Bence Jones Proteins*

| Position | No. of Protein Sequences Studied | Amino Acids | Total | Human Kappa I | Human Kappa II | Human Kappa III | Human Lambda I | Human Lambda II | Human Lambda III | Human Lambda IV | Human Lambda V | Mouse Kappa I | Mouse Kappa II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 74 | Glu | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | --- | 73 | 23 | 3 | 17 | 9 | 2 | 6 | 4 | 2 | 2 | 5 |
| 1 | 74 | Lys | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | PCA | 17 | 0 | 0 | 0 | 8 | 2 | 6 | 0 | 0 | 0 | 1 |
| | | Asp | 29 | 18 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 4 |
| | | Asx | 5 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Glu | 15 | 0 | 0 | 14 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Glx | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | --- | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 2 | 0 | 0 |
| 2 | 73 | Ile | 48 | 22 | 3 | 16 | 0 | 0 | 0 | 0 | 0 | 2 | 5 |
| | | Tyr | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 |
| | | Val | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Met | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 19 | 0 | 0 | 0 | 9 | 2 | 6 | 0 | 2 | 0 | 0 |
| 3 | 72 | Ile | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Pro | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| | | Leu | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Val | 33 | 0 | 3 | 16 | 8 | 0 | 0 | 1 | 0 | 0 | 5 |
| | | Ala | 9 | 0 | 0 | 0 | 1 | 2 | 5 | 1 | 0 | 0 | 0 |
| | | Asp | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Glu | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| | | Gln | 21 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| | | Glx | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 71 | Leu | 43 | 3 | 1 | 14 | 9 | 2 | 6 | 4 | 1 | 0 | 3 |
| | | Val | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Met | 26 | 20 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| 5 | 70 | Ala | 2 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| | | Thr | 67 | 23 | 3 | 17 | 9 | 2 | 3 | 4 | 1 | 2 | 3 |
| | | Ser | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 6 | 69 | Gln | 63 | 20 | 3 | 16 | 8 | 1 | 6 | 4 | 1 | 1 | 3 |
| | | Glx | 6 | 3 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 7 | 63 | Pro | 20 | 0 | 0 | 0 | 9 | 2 | 6 | 3 | 0 | 0 | 0 |
| | | Thr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 41 | 22 | 2 | 14 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| | | Asp | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

Downloaded from www.jem.org on November 2, 2004

TABLE I- *Continued*

| Position | No. of Protein Sequences Studied | Amino Acids | Total | Human Kappa | | | Human Lambda | | | | | Mouse Kappa | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | I | II | III | I | II | III | IV | V | I | II |
| 8 | 64 | Pro | 58 | 22 | 3 | 15 | 9 | 2 | 0 | 3 | 1 | 1 | 2 |
| | | Ala | 6 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 0 |
| 9 | 63 | Leu | 3 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ala | 7 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 1 | 0 | 2 |
| | | Thr | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 41 | 21 | 0 | 0 | 9 | 2 | 6 | 2 | 0 | 1 | 0 |
| | | Gly | 10 | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asx | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 63 | Phe | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Thr | 17 | 3 | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 25 | 18 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| | | --- | 20 | 0 | 0 | 0 | 9 | 2 | 6 | 2 | 1 | 0 | 0 |
| 11 | 63 | Leu | 43 | 22 | 4 | 14 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| | | Val | 15 | 0 | 0 | 0 | 6 | 0 | 6 | 2 | 1 | 0 | 0 |
| | | Ala | 5 | 0 | 0 | 0 | 3 | 2 | 0 | 0 | 0 | 0 | 0 |
| 12 | 61 | Pro | 4 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ala | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Ser | 56 | 21 | 0 | 13 | 9 | 2 | 6 | 2 | 1 | 1 | 1 |
| 13 | 61 | Leu | 12 | 0 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Val | 11 | 2 | 4 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 2 |
| | | Met | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ala | 23 | 19 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Gly | 14 | 0 | 0 | 0 | 6 | 2 | 6 | 0 | 0 | 0 | 0 |
| 14 | 61 | Ala | 6 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Thr | 9 | 0 | 4 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Ser | 46 | 21 | 0 | 13 | 0 | 2 | 6 | 2 | 0 | 1 | 1 |
| 15 | 61 | Pro | 36 | 0 | 4 | 13 | 8 | 2 | 6 | 2 | 0 | 0 | 1 |
| | | Leu | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| | | Val | 20 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asx | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | 61 | Arg | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Gly | 60 | 20 | 4 | 13 | 9 | 2 | 6 | 2 | 1 | 1 | 2 |
| 17 | 61 | Asp | 23 | 21 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Glu | 17 | 0 | 4 | 11 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| | | Gln | 19 | 0 | 0 | 0 | 8 | 2 | 5 | 2 | 1 | 0 | 1 |
| | | Glx | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

Downloaded from www.jem.org on November 2, 2004

TABLE I—*Continued*

| Position | No. of Protein Sequences Studies | Amino Acids | Total | Human Kappa I | Human Kappa II | Human Kappa III | Human Lambda I | Human Lambda II | Human Lambda III | Human Lambda IV | Human Lambda V | Mouse Kappa I | Mouse Kappa II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 18 | 61 | Pro | 4 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Lys | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ala | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Arg | 40 | 21 | 0 | 13 | 4 | 0 | 0 | 0 | 0 | 1 | 1 |
| | | Thr | 5 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 1 | 0 | 0 |
| | | Ser | 9 | 0 | 0 | 0 | 1 | 2 | 5 | 0 | 0 | 0 | 1 |
| | | Gly | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19 | 53 | Ile | 7 | 1 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 0 |
| | | Val | 28 | 17 | 0 | 0 | 6 | 2 | 0 | 0 | 1 | 1 | 1 |
| | | Ala | 18 | 0 | 4 | 10 | 1 | 0 | 0 | 2 | 0 | 0 | 1 |
| 20 | 53 | Ile | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Val | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Ala | 3 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Arg | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Thr | 40 | 17 | 0 | 8 | 6 | 2 | 6 | 0 | 0 | 0 | 1 |
| | | Ser | 7 | 0 | 4 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| 21 | 43 | Ile | 30 | 14 | 4 | 0 | 5 | 2 | 0 | 3 | 1 | 0 | 1 |
| | | Leu | 12 | 1 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| | | Val | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | 42 | Ala | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Thr | 19 | 14 | 0 | 0 | 0 | 0 | 0 | 3 | 1 | 1 | 0 |
| | | Ser | 22 | 0 | 4 | 9 | 5 | 2 | 0 | 0 | 0 | 0 | 2 |
| 23 | 30 | Cys | 30 | 9 | 3 | 4 | 5 | 2 | 0 | 3 | 1 | 1 | 2 |
| 24 | 26 | Arg | 11 | 1 | 3 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| | | Thr | 2 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 6 | 0 | 0 | 0 | 4 | 0 | 0 | 2 | 0 | 0 | 0 |
| | | Gly | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Gln | 4 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Glx | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25 | 25 | Ala | 13 | 6 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| | | Ser | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Gly | 10 | 0 | 0 | 0 | 4 | 2 | 0 | 3 | 1 | 0 | 0 |
| 26 | 24 | Thr | 2 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 16 | 6 | 2 | 4 | 1 | 0 | 0 | 0 | 0 | 1 | 2 |
| | | Gly | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asp | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 1 | 0 | 0 |

Downloaded from www.jem.org on November 2, 2004

TABLE I—*Continued*

| Position | No. of Protein Sequences Studied | Amino Acids | Total | Human Kappa I | Human Kappa II | Human Kappa III | Human Lambda I | Human Lambda II | Human Lambda III | Human Lambda IV | Human Lambda V | Mouse Kappa I | Mouse Kappa II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27 | 23 | Lys | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Ser | 5 | 0 | 0 | 0 | 3 | 1 | 0 | 0 | 1 | 0 | 0 |
| | | Asn | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Glu | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Gln | 13 | 5 | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| | | Glx | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| a | 22 | Ser | 5 | 0 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Asn | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | --- | 16 | 6 | 0 | 1 | 3 | 1 | 0 | 2 | 1 | 1 | 1 |
| b | 22 | Leu | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Val | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | --- | 19 | 6 | 0 | 4 | 3 | 1 | 0 | 2 | 1 | 1 | 1 |
| c | 22 | Leu | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | --- | 20 | 6 | 0 | 4 | 3 | 1 | 0 | 2 | 1 | 1 | 2 |
| d | 22 | Thr | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 3 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Asp | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | --- | 17 | 6 | 1 | 4 | 0 | 0 | 0 | 2 | 1 | 1 | 2 |
| e | 22 | Asp | 2 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Asn | 3 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asx | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | --- | 16 | 6 | 1 | 4 | 0 | 0 | 0 | 2 | 1 | 1 | 1 |
| f | 22 | Val | 2 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Gly | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Glx | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | --- | 16 | 6 | 0 | 3 | 2 | 0 | 0 | 2 | 1 | 1 | 1 |
| 28 | 22 | Ile | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Leu | 4 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 1 | 0 | 0 |
| | | Val | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Met | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Thr | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 3 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Gly | 2 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | His | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Asp | 4 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asx | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

Downloaded from www.jem.org on November 2, 2004

TABLE I—*Continued*

| Position | No. of Protein Sequences Studied | Amino Acids | Total | Human Kappa I | Human Kappa II | Human Kappa III | Human Lambda I | Human Lambda II | Human Lambda III | Human Lambda IV | Human Lambda V | Mouse Kappa I | Mouse Kappa II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 29 | 21 | Ile | 8 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| | | Arg | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Ser | 3 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Gly | 4 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 1 |
| | | Glu | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Asp | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asx | 2 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 30 | 21 | Ile | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Lys | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Ser | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Gly | 5 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| | | Asp | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Asn | 8 | 3 | 0 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asx | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 31 | 20 | Tyr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Phe | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Lys | 2 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Thr | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 7 | 1 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 1 | 1 |
| | | His | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asn | 4 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Asx | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 32 | 20 | Trp | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Tyr | 10 | 2 | 2 | 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Phe | 5 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Leu | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Thr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Asp | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Asn | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 33 | 18 | Leu | 11 | 4 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Val | 4 | 0 | 0 | 0 | 2 | 1 | 0 | 1 | 0 | 0 | 0 |
| | | Met | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Ala | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Ser | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 34 | 18 | Tyr | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ala | 6 | 1 | 0 | 4 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Ser | 3 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| | | Asp | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asn | 4 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Asx | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

Downloaded from www.jem.org on November 2, 2004

TABLE I—*Continued*

| Position | No. of Protein Sequences Studied | Amino Acids | Total | Human Kappa I | Human Kappa II | Human Kappa III | Human Lambda I | Human Lambda II | Human Lambda III | Human Lambda IV | Human Lambda V | Mouse Kappa I | Mouse Kappa II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 35 | 17 | Trp | 17 | 4 | 2 | 4 | 2 | 1 | 0 | 1 | 1 | 1 | 1 |
| 36 | 17 | Tyr | 13 | 4 | 2 | 4 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| | | Phe | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| | | Leu | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | His | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 37 | 16 | Leu | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Gln | 12 | 3 | 0 | 4 | 2 | 1 | 0 | 1 | 1 | 0 | 0 |
| | | Glx | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 38 | 16 | His | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Gln | 12 | 3 | 1 | 4 | 1 | 1 | 0 | 1 | 1 | 0 | 0 |
| | | Glx | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 39 | 16 | Leu | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Lys | 9 | 2 | 2 | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| | | Arg | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Gly | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | His | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Asx | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 40 | 16 | Pro | 15 | 3 | 1 | 4 | 2 | 1 | 0 | 1 | 1 | 1 | 1 |
| | | Ala | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 41 | 16 | Lys | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Gly | 15 | 2 | 2 | 4 | 2 | 1 | 0 | 1 | 1 | 1 | 1 |
| 42 | 16 | Lys | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Arg | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Thr | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Gln | 7 | 0 | 1 | 4 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| | | Glx | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 43 | 16 | Pro | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Ala | 11 | 3 | 0 | 4 | 2 | 1 | 0 | 0 | 1 | 0 | 0 |
| | | Thr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Ser | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Gln | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 44 | 16 | Ile | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Pro | 15 | 3 | 2 | 4 | 2 | 1 | 0 | 1 | 1 | 0 | 1 |
| 45 | 16 | Leu | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| | | Lys | 9 | 3 | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 1 | 1 |
| | | Arg | 3 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Glx | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Downloaded from www.jem.org on November 2, 2004

TABLE I—*Continued*

| Position | No. of Protein Sequences Studied | Amino Acids | Total | Human Kappa I | Human Kappa II | Human Kappa III | Human Lambda I | Human Lambda II | Human Lambda III | Human Lambda IV | Human Lambda V | Mouse Kappa I | Mouse Kappa II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 46 | 14 | Ile | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Leu | 12 | 2 | 2 | 2 | 2 | 1 | 0 | 1 | 1 | 0 | 1 |
| | | Arg | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 47 | 14 | Leu | 11 | 3 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 1 | 1 |
| | | Val | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 |
| 48 | 14 | Ile | 12 | 2 | 2 | 1 | 2 | 1 | 0 | 1 | 1 | 1 | 1 |
| | | Met | 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 49 | 14 | Tyr | 13 | 3 | 2 | 2 | 2 | 0 | 0 | 1 | 1 | 1 | 1 |
| | | Phe | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 50 | 14 | Leu | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Val | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Lys | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ala | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| | | Arg | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Thr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Gly | 3 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| | | Asp | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Glu | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 51 | 14 | Leu | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Val | 2 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Ala | 5 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Arg | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Thr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Gly | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | His | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Asp | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 52 | 14 | Ser | 10 | 3 | 2 | 2 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| | | Asp | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asn | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Glu | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 53 | 14 | Tyr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Lys | 3 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 4 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Asn | 4 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| | | Glu | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Glx | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 54 | 14 | Leu | 4 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Arg | 9 | 0 | 2 | 2 | 2 | 1 | 0 | 1 | 1 | 0 | 0 |
| | | Gln | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

Downloaded from www.jem.org on November 2, 2004

TABLE I—*Continued*

| Position | No. of Protein Sequences Studied | Amino Acids | Total | Human Kappa I | Human Kappa II | Human Kappa III | Human Lambda I | Human Lambda II | Human Lambda III | Human Lambda IV | Human Lambda V | Mouse Kappa I | Mouse Kappa II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 55 | 16 | Pro | 5 | 0 | 0 | 0 | 2 | 1 | 0 | 1 | 1 | 0 | 0 |
| | | Ala | 6 | 0 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Gly | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Asx | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Glu | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 56 | 16 | Ala | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Thr | 5 | 1 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 10 | 1 | 2 | 0 | 2 | 1 | 0 | 1 | 1 | 1 | 1 |
| 57 | 16 | Thr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Gly | 15 | 3 | 2 | 4 | 2 | 1 | 0 | 0 | 1 | 1 | 1 |
| 58 | 16 | Ile | 6 | 0 | 0 | 3 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| | | Val | 9 | 3 | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| | | Thr | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 59 | 16 | Pro | 16 | 3 | 2 | 4 | 2 | 1 | 0 | 1 | 1 | 1 | 1 |
| 60 | 16 | Val | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Lys | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Ala | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Ser | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asp | 8 | 0 | 1 | 3 | 2 | 1 | 0 | 0 | 1 | 0 | 0 |
| | | Asn | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Glu | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 61 | 16 | Arg | 16 | 3 | 2 | 4 | 2 | 1 | 0 | 1 | 1 | 1 | 1 |
| 62 | 16 | Ile | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Phe | 15 | 3 | 2 | 4 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| 63 | 16 | Ile | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 15 | 2 | 2 | 4 | 2 | 1 | 0 | 1 | 1 | 1 | 1 |
| 64 | 16 | Ala | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Gly | 15 | 3 | 2 | 4 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| 65 | 16 | Thr | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 15 | 2 | 2 | 4 | 2 | 1 | 0 | 1 | 1 | 1 | 1 |
| 66 | 16 | Lys | 3 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Arg | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Ser | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| | | Gly | 10 | 3 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 67 | 16 | Phe | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 15 | 2 | 2 | 4 | 2 | 1 | 0 | 1 | 1 | 1 | 1 |

Downloaded from www.jem.org on November 2, 2004

TABLE I—*Continued*

| Position | No. of Protein Sequences Studied | Amino Acids | Total | Human Kappa I | Human Kappa II | Human Kappa III | Human Lambda I | Human Lambda II | Human Lambda III | Human Lambda IV | Human Lambda V | Mouse Kappa I | Mouse Kappa II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 68 | 16 | Gly | 15 | 3 | 2 | 4 | 2 | 0 | 0 | 1 | 1 | 1 | 1 |
| | | Asn | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 69 | 16 | Ala | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Thr | 11 | 3 | 2 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Ser | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | His | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Asp | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 70 | 16 | Thr | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 |
| | | Ser | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asp | 9 | 2 | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| | | Asx | - | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Glu | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 71 | 16 | Tyr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Phe | 10 | 3 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Ala | 5 | 0 | 0 | 0 | 2 | 1 | 0 | 1 | 1 | 0 | 0 |
| 72 | 16 | Thr | 11 | 3 | 2 | 4 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Ser | 5 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| 73 | 16 | Phe | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Leu | 14 | 1 | 2 | 4 | 2 | 1 | 0 | 1 | 1 | 1 | 1 |
| 74 | 16 | Lys | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ala | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Thr | 12 | 3 | 1 | 4 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| | | Gly | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asn | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 75 | 16 | Ile | 15 | 3 | 2 | 4 | 2 | 0 | 0 | 1 | 1 | 1 | 1 |
| | | Val | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 76 | 16 | Thr | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Ser | 13 | 3 | 2 | 4 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| | | His | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 77 | 17 | Pro | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Arg | 6 | 0 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 4 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Gly | 6 | 1 | 0 | 0 | 2 | 1 | 0 | 1 | 1 | 0 | 0 |
| 78 | 17 | Leu | 13 | 4 | 1 | 4 | 2 | 1 | 0 | 0 | 0 | 1 | 0 |
| | | Val | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Met | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Ala | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |

Downloaded from www.jem.org on November 2, 2004

TABLE I—*Continued*

| Position | No. of Protein Sequences Studied | Amino Acids | Total | Human Kappa I | Human Kappa II | Human Kappa III | Human Lambda I | Human Lambda II | Human Lambda III | Human Lambda IV | Human Lambda V | Mouse Kappa I | Mouse Kappa II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 79 | 17 | Arg | 3 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Glu | 5 | 0 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Gln | 6 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| | | Glx | 3 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 80 | 17 | Pro | 9 | 4 | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ala | 3 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| | | Thr | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| | | Glx | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 81 | 17 | Val | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Ala | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Gly | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asp | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asx | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Glu | 10 | 2 | 1 | 4 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |
| | | Glx | 2 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 82 | 17 | Asp | 14 | 4 | 1 | 4 | 2 | 0 | 0 | 1 | 1 | 1 | 0 |
| | | Asx | 3 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 83 | 18 | Ile | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Phe | 8 | 2 | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Val | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Thr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Glu | 5 | 0 | 0 | 0 | 2 | 0 | 0 | 2 | 1 | 0 | 0 |
| | | Glx | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 84 | 18 | Val | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Ala | 16 | 4 | 1 | 4 | 2 | 1 | 0 | 2 | 1 | 0 | 1 |
| | | Gly | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 85 | 18 | Val | 6 | 0 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Met | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Thr | 4 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | His | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asp | 5 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 1 | 1 | 0 |
| | | Asx | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 86 | 20 | Tyr | 20 | 6 | 2 | 4 | 2 | 1 | 0 | 2 | 1 | 1 | 1 |
| 87 | 20 | Tyr | 16 | 6 | 2 | 3 | 2 | 1 | 0 | 1 | 1 | 0 | 0 |
| | | Phe | 3 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| | | His | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 88 | 21 | Cys | 21 | 6 | 2 | 4 | 3 | 1 | 0 | 2 | 1 | 1 | 1 |

Downloaded from www.jem.org on November 2, 2004

TABLE I—*Continued*

| Position | No. of Protein Sequences Studied | Amino Acids | Total | Human Kappa I | Human Kappa II | Human Kappa III | Human Lambda I | Human Lambda II | Human Lambda III | Human Lambda IV | Human Lambda V | Mouse Kappa I | Mouse Kappa II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 89 | 22 | Leu | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Met | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ala | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Asn | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Gln | 15 | 6 | 2 | 4 | 1 | 0 | 0 | 2 | 0 | 0 | 0 |
| | | Glx | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 90 | 22 | Met | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ala | 3 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Thr | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Ser | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| | | Gln | 13 | 6 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Glx | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 91 | 22 | Trp | 5 | 0 | 0 | 0 | 3 | 0 | 0 | 2 | 0 | 0 | 0 |
| | | Tyr | 12 | 5 | 1 | 4 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| | | Phe | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ala | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Arg | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Ser | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 92 | 21 | Leu | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Val | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Lys | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Ala | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | Gly | 3 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asp | 9 | 4 | 0 | 0 | 2 | 0 | 0 | 2 | 1 | 0 | 0 |
| | | Asn | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Glu | 3 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 93 | 21 | Tyr | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Thr | 4 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Ser | 7 | 1 | 0 | 2 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| | | Gly | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | His | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asp | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asn | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asx | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Glu | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Gln | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 94 | 21 | Ile | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Leu | 5 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Val | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | Met | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Ala | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Arg | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 8 | 0 | 1 | 4 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |
| | | Asp | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asx | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

Downloaded from www.jem.org on November 2, 2004

TABLE I—*Continued*

| Position | No. of Protein Sequences Studied | Amino Acids | Total | Human Kappa I | Human Kappa II | Human Kappa III | Human Lambda I | Human Lambda II | Human Lambda III | Human Lambda IV | Human Lambda V | Mouse Kappa I | Mouse Kappa II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 95 | 21 | Pro | 14 | 5 | 3 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| | | Leu | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Thr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Ser | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Gly | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Asx | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 96 | 21 | Trp | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| | | Ile | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Tyr | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Phe | 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Pro | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Leu | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Lys | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Arg | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Thr | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asn | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Asx | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Gln | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | --- | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| 97 | 20 | Phe | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Pro | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Met | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Ala | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Thr | 12 | 4 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| | | His | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | --- | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| a | 19 | Val | 4 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 1 | 0 | 0 |
| | | Ala | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | --- | 14 | 5 | 3 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| b | 20 | Ile | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Leu | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | | Val | 4 | 0 | 0 | 0 | 2 | 1 | 0 | 1 | 0 | 0 | 0 |
| | | --- | 14 | 5 | 3 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 98 | 20 | Phe | 20 | 5 | 3 | 4 | 2 | 1 | 0 | 2 | 1 | 1 | 1 |
| 99 | 20 | Gly | 20 | 5 | 3 | 4 | 2 | 1 | 0 | 2 | 1 | 1 | 1 |
| 100 | 20 | Pro | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Gly | 11 | 2 | 1 | 0 | 2 | 1 | 0 | 2 | 1 | 1 | 1 |
| | | Gln | 8 | 2 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 101 | 19 | Gly | 19 | 4 | 3 | 4 | 2 | 1 | 0 | 2 | 1 | 1 | 1 |

Downloaded from www.jem.org on November 2, 2004

TABLE I—*Concluded*

| Position | No. of Protein Sequences Studied | Amino Acids | Total | Human Kappa I | Human Kappa II | Human Kappa III | Human Lambda I | Human Lambda II | Human Lambda III | Human Lambda IV | Human Lambda V | Mouse Kappa I | Mouse Kappa II |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 102 | 19 | Thr | 18 | 4 | 3 | 3 | 2 | 1 | 0 | 2 | 1 | 1 | 1 |
| | | Ser | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 103 | 19 | Lys | 14 | 4 | 1 | 3 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| | | Arg | 3 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | | Asn | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Gln | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 104 | 22 | Leu | 15 | 3 | 2 | 3 | 1 | 1 | 0 | 2 | 1 | 1 | 1 |
| | | Val | 7 | 3 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 105 | 22 | Thr | 6 | 0 | 0 | 0 | 2 | 1 | 0 | 2 | 1 | 0 | 0 |
| | | Asp | 4 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Glu | 12 | 3 | 4 | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 106 | 22 | Ile | 12 | 3 | 4 | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| | | Phe | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Leu | 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Val | 7 | 1 | 0 | 0 | 2 | 1 | 0 | 2 | 1 | 0 | 0 |
| a | 22 | Leu | 6 | 0 | 0 | 0 | 2 | 1 | 0 | 2 | 1 | 0 | 0 |
| | | --- | 16 | 6 | 4 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 107 | 22 | Lys | 15 | 6 | 3 | 4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| | | Arg | 3 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | Ser | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| | | Gly | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |

studied at the given position. Only data for which the sequence has been clearly assigned by the various authors have been included.

*The Role of Glycine*—It has been suggested that glycine plays a unique role in the structure of the variable region of immunoglobulin light chains (18, p. 87; 41–43, 45). Jukes (50) and Welscher (51) have generally agreed with this. A further careful analysis becomes essential for the understanding of the function of the glycines in the over-all structure and in relation to antibody-combining sites.

The basic property that differentiates glycine from all other amino acids structurally is the absence of a side chain. As a result, glycine can have many sterically allowable configurations. This has been verified experimentally in the case of lysozyme (46, 52) and tosyl-$\alpha$-chymotrypsin (53). The two angles, $\Phi$ and $\psi$ (54), which specify the conformation of the backbone of an amino acid have been calculated for each of the amino acids from the known tertiary structures of lysozyme (46, 52) and of tosyl-$\alpha$-chymotrypsin (53). A typical plot of the permissible angles of the glycine as compared with the alanine residues is shown in Fig. 1. The allowable configurations of alanine are mostly

Downloaded from www.jem.org on November 2, 2004

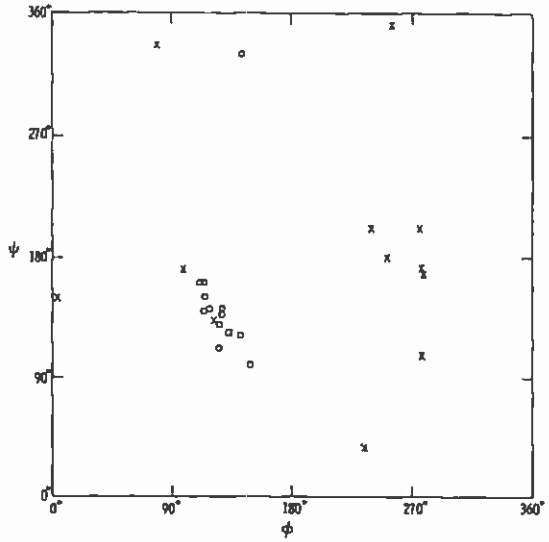



FIG. 1 *a*

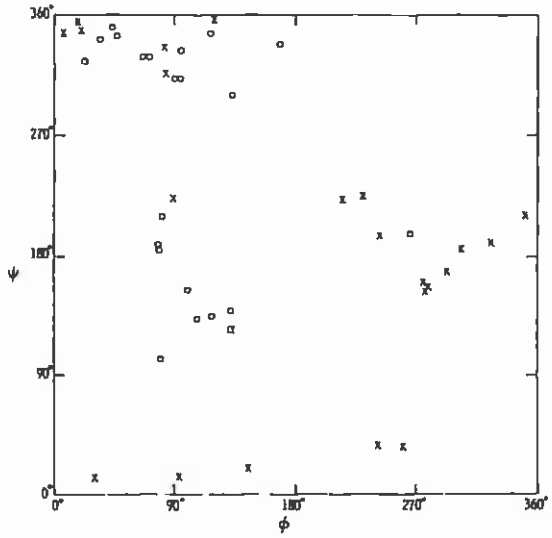



FIG. 1 *b*

FIG. 1. Comparison of $\phi$ and $\psi$ for glycine (×) and alanine (○) based on their occurrence in (*a*) lysozyme (data from references 46, 52), and in (*b*) tosyl-α-chymotrypsin (calculated from data given in reference 53). $\alpha$ indicates the values of $\phi$ and $\psi$ for the α-helix.

Downloaded from www.jem.org on November 2, 2004

clustered near the $\alpha$-helical region of lysozyme (Fig. 1 *a* and reference 55), while those of glycine are widely distributed. Comparison with similar maps for other amino acids in lysozyme also shows them to be more restricted. The data for tosyl-$\alpha$-chymotrypsin also show that glycine may have many more conformations (Fig. 1 *b*). This unique property of glycine thus may permit relative motion of the chains attached to the two ends of the molecule. With immuno-

TABLE II

*Frequencies of Glycine Residues at Various Positions in the Variable Region of Light Chains*

| Position | Human Kappa | | | Human Lambda | | | | | Mouse Kappa | | Total | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | I | II | III | I | II | III | IV | V | I | II | | |
| 9 | | | 10/15 | | | | | | | | 10/63 | 16 |
| 13 | | | | 6/9 | 2/2 | 6/6 | | | | | 14/61 | 23 |
| 16 | 20/21 | 4/4 | 13/13 | 9/9 | 2/2 | 6/6 | 2/2 | 1/1 | 1/1 | 2/2 | 60/61 | 99 |
| 24 | | | | | | | 1/3 | | | | 1/26 | 4 |
| 25 | | | | 4/4 | 2/2 | | 3/3 | 1/1 | | | 10/25 | 40 |
| 26 | | | | 2/3 | | | | | | | 2/24 | 8 |
| 27f | | | | 1/3 | | | | | | | 1/22 | 5 |
| 28 | | 1/2 | | | 1/1 | | | | | | 2/22 | 9 |
| 29 | | | | 2/2 | | | 1/2 | | | 1/2 | 4/21 | 19 |
| 30 | | 2/2 | 1/4 | | | | | 1/1 | 1/1 | | 5/21 | 24 |
| 39 | 1/3 | | | | | | | | | | 1/16 | 6 |
| 41 | 2/3 | 2/2 | 4/4 | 2/2 | 1/1 | | 1/1 | 1/1 | 1/1 | 1/1 | 15/16 | 94 |
| 50 | | | 1/2 | | 1/1 | | | 1/1 | | | 3/14 | 21 |
| 51 | | 1/2 | | | | | | | | | 1/14 | 7 |
| 55 | | | | | | | | | | 1/1 | 1/16 | 6 |
| 57 | 3/3 | 2/2 | 4/4 | 2/2 | 1/1 | | | 1/1 | 1/1 | 1/1 | 15/16 | 94 |
| 64 | 3/3 | 2/2 | 4/4 | 1/2 | 1/1 | | 1/1 | 1/1 | 1/1 | 1/1 | 15/16 | 94 |
| 66 | 3/3 | 2/2 | 4/4 | | | | | | | 1/1 | 10/16 | 62 |
| 69 | 3/3 | 2/2 | 4/4 | 2/2 | | | 1/1 | 1/1 | 1/1 | 1/1 | 15/16 | 94 |
| 74 | | | | 1/2 | | | | | | | 1/16 | 6 |
| 77 | 1/4 | | | 2/2 | 1/1 | | 1/1 | 1/1 | | | 6/17 | 35 |
| 81 | | | | 1/2 | | | | | | | 1/17 | 6 |
| 84 | | 1/2 | | | | | | | | | 1/18 | 6 |

Downloaded from www.jem.org on November 2, 2004

TABLE II—*Concluded*

| Position | Human Kappa I | II | III | Human Lambda I | II | III | IV | V | Mouse Kappa I | II | Total | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 92 | | | 3/4 | | | | | | | | 3/21 | 14 |
| 93 | | | 1/4 | | | | | | | | 1/21 | 5 |
| 95 | | | | | | | | 1/1 | | | 1/21 | 5 |
| 99 | 5/5 | 3/3 | 4/4 | 2/2 | 1/1 | | 2/2 | 1/1 | 1/1 | 1/1 | 20/20 | 100 |
| 100 | 2/5 | 1/3 | | 2/2 | 1/1 | | 2/2 | 1/1 | 1/1 | 1/1 | 11/20 | 55 |
| 101 | 4/4 | 3/3 | 4/4 | 2/2 | 1/1 | | 2/2 | 1/1 | 1/1 | 1/1 | 19/19 | 100 |
| 107 | | | | 1/2 | | | | 1/1 | | | 2/18 | 11 |

Fractions represent the number of instances in which glycine occurs to the total number of proteins studied at the given position for each subgroup. When values have not been given, glycine has not been reported at that position in the subgroup.

globulins, flexibility of the protein backbone can be one of the major factors that permit substitution of various amino acids at the variable positions; it also may allow movement of the site to make most favorable contact in combining with an antigenic determinant. Though a glycine residue confers maximum flexibility over all other amino acids, it might also arise from a random mutation in which the difference between glycine and other amino acids is not adverse for the over-all structure. In addition, some glycines might be complementarity determining. These latter two kinds of glycines must be distinguished from the first.

For the variable region of the light chains of human and mouse immunoglobulins and of Bence Jones proteins, alignment of amino acid sequences serves to identify the glycines which may be conferring flexibility as shown in Table II. All the glycines are listed. The frequencies, expressed as per cent, can roughly be divided into three categories:

*A. 94–100%*: Positions 16, 41 (or 39), 57, 64, 68, 99, and 101. Since glycine occurs at these positions in nearly all the proteins studied, it must have a fundamental structural significance and has been preserved in the evolution to man and mouse. These glycines are assumed to confer flexibility unique to antibodies.

It is of interest that glycine occurs at position 41 in 15 of 16 proteins studied (94%). The sequence at residues 39, 40, and 41 is Lys-Pro-Gly in 14 cases and Lys-Ala-Gly in one case. In the single exception, a human κI protein Ag, the

Downloaded from www.jem.org on November 2, 2004

sequence at 39, 40, and 41 is Gly-Pro-Lys.[1] Thus if the reported sequence is correct, the glycine at residue 39 might well serve the same function as the glycine at position 41. The two positions 41 and 39 may thus provide one invariant glycine (100%).

*B. 35–62%*: Positions 25, 66, 77, and 100. A careful examination indicates that glycines at these positions are at least group specific. Thus, within a group ($\kappa$ or $\lambda$), they could serve the same function as the glycines of category *A*.

*C. 4–24%*: Positions 9, 13, 24, 26, 27f, 28, 29, 30, 50, 51, 55, 74, 81, 84, 92, 93, 95, and 107. These glycine residues are at variable positions. They therefore play a distinctly different role from those of the other two categories. They might either be related to antibody complementarity or, if not involved in the site itself, could have arisen from random mutation and be nevertheless compatible with three-dimensional folding.

Thus there are about 8 (human $\kappa$ and mouse $\kappa$) to 10 (human $\lambda$) glycines in the variable region of the light chains of all proteins for which sequence data are available at these positions. They are as follows:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Human $\kappa$: Positions 16, | 41 (or 39), | 57, | 64, | 66, | 68, | 99, | and 101. |
| Human $\lambda$: Positions 16, | 25, 41, | 57, | 64, | 68, | 77, | 99, 100, | and 101. |
| Mouse $\kappa$: Positions 16, | 41, | 57, | 64, | 68, | | 99, 100, | and 101. |

Positions 99, 100, and 101 have been postulated to function as a pivot permitting the combining regions of the light and heavy chains to make most favorable contact with the antigenic determinant (18, p. 87; and 41, 43). Examination of the heavy chain sequences reported to date (56, 57) shows Gly-Gln-Gly at positions 112, 113, and 114 in two instances (He and Daw), Gly-Arg-Gly in one (Cor) and Gly-Gly at positions 114 and 115 in Eu; but in the latter protein two gaps have been placed at positions 108 and 109 in aligning with He for maximum homology. Thus these glycines could also be functionally and positionally equivalent.

*Invariant Residues*—Earlier comparisons of the invariant residues of the variable and constant regions were based entirely on those positions at which only a single amino acid occurred. As more data accumulated this number diminished, until there are now only 11 such positions in the variable region: Gln 6, Cys 23, Trp 35, Pro 59, Arg 61, Asp (Asx) 82, Tyr 86, Cys 88, Phe 98, Gly 99, and Gly 101. However, if one accepts as essentially invariant those positions at which more than 88–90% of the proteins studied have the same amino acid at a given position, this number increases to 29. A comparison of these residues with those in the constant region is given in Table III. This procedure allows for possible errors as well as for the ability of some residues to substitute for others at a given position. The difference between the con-

[1] The sequence of Roy was originally reported as Gly-Pro-Lys but has been changed to Lys-Pro-Gly (see references to Roy in sequence data).

Downloaded from www.jem.org on November 2, 2004

stant and variable regions is not much different than originally appeared (41) except that there is one invariant alanine and one invariant leucine in the variable region. However, the difference between the two regions is still quite clear, the variable region having in addition no invariant valine, lysine, and

TABLE III

*Comparison of the Invariant Residues of the Variable with those of the Constant Region in Human κ-, Human λ-, and Mouse κ-Bence Jones Proteins*

| Amino Acid | Variable Region | Constant Region |
|---|---|---|
| Gly | 7* | 0 |
| Ala | 1 | 3 |
| Leu | 1 | 3 |
| Val | 0 | 3 |
| Lys | 0 | 2 |
| His | 0 | 2 |
| Ile | 1 | 0 |
| Ser | 3 | 5 |
| Glu | 0 | 2 |
| Gln | 2 | 0 |
| Arg | 1 | 0 |
| Pro | 3 | 4 |
| Tyr | 2 | 3 |
| Cys | 2 | 3 |
| Phe | 2 | 2 |
| Trp | 1 | 1 |
| Thr | 2 | 1 |
| Asp | 1 | 1 |
| Total | 29 | 39 |

* Not including positions 25, 66, 77, and 100.

Invariant residues are those in which 88-90% of the proteins analyzed contain the same amino acid residue at a given position.

histidine, while the invariant residues in the constant region include three alanines, three leucines, three valines, two lysines, and two histidines.

*Hydrophobicity Distribution of the Invariant Residues of the Variable Region*—A parameter, $H\emptyset_{ave}$, based on the free energies of transfer of amino acid side chains from an organic to an aqueous environment has been introduced by Tanford (58) and applied by Bigelow (59) to the study of various proteins. $H\emptyset_{ave}$ is expressed in kilocalories per residue and varies from 3.00 for Trp to 0.45 for Thr and is very small, zero, or negative for Gly, Ser, His, Asp, Glu, Asn, and Gln; these have been taken as zero. These values have been used in

Downloaded from www.jem.org on November 2, 2004

an examination of the invariant residues of the variable region. Table IV summarizes the findings and tabulates $H\emptyset_{ave}$ for the invariant residues and for the occasional other substituents found. In addition we have included data on two positions, 46 and 48, in which another substituent occurred with slightly lower frequency than that required by the definition of an invariant residue, and on three other positions 8, 12, and 37 in which the other substituents were confined to a single subgroup.

TABLE IV

*Hydrophobicity, $H\emptyset_{ave}$, of the Invariant Residues and almost Invariant Residues of the Variable Region*

| Position | Amino Acids | | Hydrophobicity, $H\emptyset_{ave}$ | |
|---|---|---|---|---|
| | Invariant | Other | Invariant | Other |
| 5 | Thr 67* | Ala 2, Ser 1 | 0.45 | 0.75, 0.00 |
| 6 | Gln 63 | | 0.00 | |
| 16 | Gly 60 | Arg 1 | 0.00 | 0.75 |
| 23 | Cys 30 | | 1.00 | |
| 35 | Trp 17 | | 3.00 | |
| 38 | Gln 12, Glx 3 | His 1 | 0.00 | 0.00 |
| 40 | Pro 15 | Ala 1 | 2.60 | 0.75 |
| 41(39) | Gly 15 | Lys 1 | 0.00 | 1.50 |
| 44 | Pro 15 | Ile 1 | 2.60 | 2.95 |
| 49 | Tyr 13 | Phe 1 | 2.85 | 2.65 |
| 57 | Gly 15 | Thr 1 | 0.00 | 0.45 |
| 59 | Pro 16 | | 2.60 | |
| 61 | Arg 16 | | 0.75 | |
| 62 | Phe 15 | Ile 1 | 2.65 | 2.95 |
| 63 | Ser 15 | Ile 1 | 0.00 | 2.95 |
| 64 | Gly 15 | Ala 1 | 0.00 | 0.75 |
| 65 | Ser 15 | Thr 1 | 0.00 | 0.45 |
| 67 | Ser 15 | Phe 1 | 0.00 | 2.65 |
| 68 | Gly 15 | Asn 1 | 0.00 | 0.00 |
| 73 | Leu 14 | Phe 2 | 2.40 | 2.65 |
| 75 | Ile 15 | Val 1 | 2.95 | 1.70 |
| 82 | Asp 14, Asx 3 | | 0.00 | |
| 84 | Ala 16 | Gly 1, Val 1 | 0.75 | 0.00, 1.70 |
| 86 | Tyr 20 | | 2.85 | |
| 88 | Cys 21 | | 1.00 | |
| 98 | Phe 20 | | 2.65 | |
| 99 | Gly 20 | | 0.00 | |
| 101 | Gly 19 | | 0.00 | |
| 102 | Thr 18 | Ser 1 | 0.45 | 0.00 |

Downloaded from www.jem.org on November 2, 2004

TABLE IV—*Concluded*

Borderline for almost invariant.

| Position | Amino Acids | | Hydrophobicity, $H\phi_{ave}$ | |
|---|---|---|---|---|
| | Invariant | Other | Invariant | Other |
| 46 | Leu 12 | Ile 1, Arg 1 | 2.40 | 2.95, 0.75 |
| 48 | Ile 12 | Met 2 | 2.95 | 1.30 |

Invariant with a subgroup exception.

| Position | Amino Acids | | Hydrophobicity, $H\phi_{ave}$ | |
|---|---|---|---|---|
| | Invariant | Other | Invariant | Other |
| 8 | Pro 58 | Ala 6 (Human λ III) | 2.60 | 0.75 |
| 12 | Ser 56 | Pro 4 (Human κ II) | 0.00 | 2.60 |
| | | Ala 1 (Mouse κ II) | | 0.75 |
| 37 | Glu 12, Glx 2 | Pro 2 (Human κ II) | 0.00 | 2.60 |

* Numbers next to the residue represent the number of samples in which the residue occurred.

It is of interest that 15 of the invariant residues have values of 0.00 or 0.45 and that 10 have values of 2.40–3.00, leaving only 4 residues of intermediate hydrophobicity: 2 half-cystines, 1 arginine, and 1 alanine. In most instances the other substituents reported as replacements at these positions generally had $H\phi_{ave}$ values not too different from the major substituent, but at positions 40, 41, 63, 67, 75, and 84 changes of over one unit were seen. (Position 41 may not be significant since, as discussed earlier, the exception had glycine in position 39.) The two borderline invariant residues both showed substantial $H\phi_{ave}$ differences and the three subgroup specific residues also varied substantially in $H\phi_{ave}$.

Of the 35 invariant residues in the constant region, 11 have values of 0.00 or 0.45, 13 have values of 2.40 to 3.00, and 11 have values of 0.75 to 1.70 (Table V). Thus the constant region has invariant residues which appear to be relatively uniformly distributed with respect to $H\phi_{ave}$ while in the variable region they are generally either very high or zero.

The average hydrophobicity for invariant residues of the variable region is about 1.09 while that of invariant residues of the constant region is 1.39.

Welcher (51) has computed $H\phi_{ave}$ for the entire light chain from sequence data on Bence Jones proteins and obtains values ranging from 0.970 to 1.04 kcal/residue, values in the same range as those for chains of other proteins. The variable regions range from 0.930 to 1.11 while the constant region values were

Downloaded from www.jem.org on November 2, 2004

0.950 for mouse κ, 0.970 for human κ, and 1.02 for human λ. Thus the over-all hydrophobicities of the two regions are no different; Welcher also reported no difference between the two regions with respect to the pattern of nonpolar positions, while our data show a substantial difference in $H\emptyset_{ave}$ of the invariant residues of the two regions. Moreover, the average hydrophobicity of the entire

TABLE V

*Hydrophobicity, $H\emptyset_{ave}$, of the Invariant Residues of the Constant Region*

| Position | Amino Acid | Hydrophobicity $H\emptyset_{ave}$ | Position | Amino Acid | Hydrophobicity $H\emptyset_{ave}$ |
|---|---|---|---|---|---|
| 111 | Ala | 0.75 | 149 | Lys | 1.50 |
| 112 | Ala | 0.75 | 151 | Asp | 0.00 |
| 113 | Pro | 2.60 | 168 | Ser | 0.00 |
| 115 | Val | 1.70 | 173 | Tyr | 2.85 |
| 118 | Phe | 2.65 | 176 | Ser | 0.00 |
| 119 | Pro | 2.60 | 177 | Ser | 0.00 |
| 120 | Pro | 2.60 | 179 | Leu | 2.40 |
| 121 | Ser | 0.00 | 181 | Leu | 2.40 |
| 123 | Glu | 0.00 | 189 | His | 0.00 |
| 125 | Leu | 2.40 | 192 | Tyr | 2.85 |
| 130 | Ala | 0.75 | 194 | Cys | 1.00 |
| 133 | Val | 1.70 | 197 | Thr | 0.45 |
| 134 | Cys | 1.00 | 198 | His | 0.00 |
| 139 | Phe | 2.65 | 203 | Ser | 0.00 |
| 140 | Tyr | 2.85 | 207 | Lys | 1.50 |
| 141 | Pro | 2.60 | 213 | Glu | 0.00 |
| 146 | Val | 1.70 | 214 | Cys | 1.00 |
| 148 | Trp | 3.00 | | | |

constant region is about 0.98, significantly lower than 1.39 for the hydrophobicity of its invariant residues.

*Variability*—In considering the nature of the variable region, it is of importance to ascertain whether the variability is uniformly distributed or is confined to small segments of the variable regions. Thus, a quantity is defined for each amino acid position in the sequence

$$\text{Variability} = \frac{\text{Number of different amino acids at a given position}}{\text{Frequency of the most common amino acid at that position}} \quad [1]$$

in which the denominator is the number of times the most common amino acid occurs divided by the total number of proteins examined. Thus at position 7 (Table I) 63 proteins were studied, serine occurred 41 times and 4 different

Downloaded from www.jem.org on November 2, 2004

amino acids, Pro, Thr, Ser, and Asp, have been reported. The frequency of the most common is 41/63 = 0.65 and the variability is then 4/0.65 = 6.15. When there was uncertainty as to the number of amino acids, as in instances in which Glx or Asx has been reported, the extreme values of variability have been computed. For this equation an absolutely invariant residue would have a value of 1 while the theoretical upper limit for 20 amino acids randomly occurring would be 400.

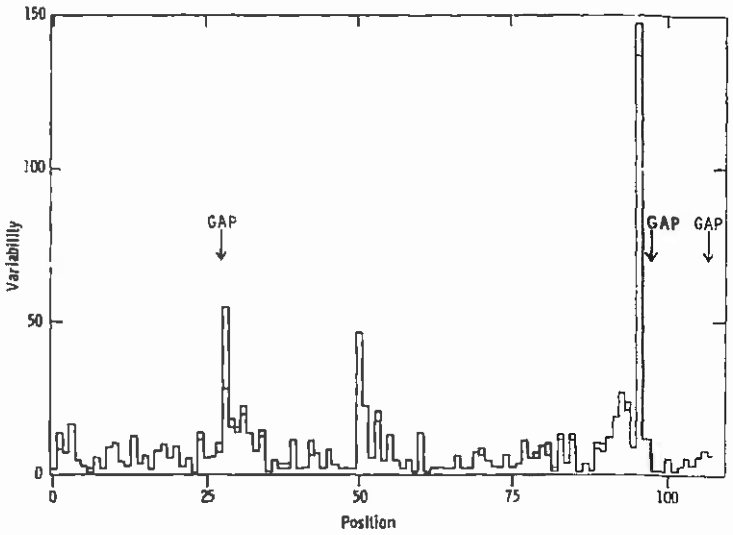



FIG. 2. Variability at different amino acid positions for the variable region of the light chains. GAP indicates positions at which insertions have been found.

Plotting variability against position for the 107 residues of the variable region (Fig. 2) shows three main peaks in the regions of residues 28, 50, and 96; two of these, 28 and 96, are the highly variable regions (7, 45, 47, 49) in which insertions occur, while position 50 has not been associated with an insertion. Franěk (49) has previously noted the high variability around position 50. The stretches of amino acid residues showing this high variability are 24–34, 50–56, and 89–97. The first and third regions begin after an invariant Cys and are followed by an invariant Trp and Phe respectively, at positions 35 and 98. The second region begins after an almost invariant position 49 (Tyr 13/14, Phe 1/14) and is followed by an invariant position 57 (Gly 15/16, Thr 1/16) (Table I).

The over-all sequence data were also examined to ascertain whether amino acid substitutions at each position were reflected in changes in hydrophobicity

Downloaded from www.jem.org on November 2, 2004

confined to certain stretches of the variable region. The findings which are not plotted showed that the same three regions associated with high variability were those with the greatest variation in $H\emptyset_{ave}$.

Since a portion of the variability at many positions is group- or subgroup-specific and is therefore generally not complementarity determining, variability as defined in equation [1] was computed for the individual subgroups for which sufficient sequence data were available. A plot for κI showed high variability in the stretches 24–34 and 92–96 and at residues 53 and 56. κIII shows high variability at residue 96. Data are generally insufficient even for these two subgroups and the data for other subgroups do not permit such an analysis; the λI subgroup showed unusually high variability at position 18.

*Classification of Variability at Individual Positions*—The sequence data available at each position (Table I) were examined in an attempt to classify the position with respect to its role in the over-all structure. Invariant residues already considered are not included. The following categories were set up. (*a*) Invariant except for one subgroup. (*b*) κ- vs. λ-specific. (*c*) κ- vs. λ-specific with some subgroup variations. (*d*) Variation in κ- and λ-subgroups. (*e*) Variation in κ-subgroups with λ relatively constant. (*f*) Variation in λ-subgroups with κ relatively constant. (*g*) Unaccountable variability. (*h*) Insufficient data. (*i*) Possible species specificity. It has not been possible to set up absolute criteria for each of these classes and some difficulties were encountered in making these assignments, especially at positions for which the sequence data were relatively sparse. Occasional substitutions compatible with point mutations have often been neglected.

Examinations of each of these categories permit some interesting inferences to be made:

(*a*) There are five positions, 8, 12, 15, 37, and 54, which are essentially invariant except that in one subgroup another amino acid occurs. Thus at position 8, 58 proteins have Pro while 6, all of which belong to the λIII subgroup, contain Ala. At the other positions, additional amino acids sometimes occur in individual proteins. These positions are considered to be part of the basic skeleton of the variable region under the control of structural genes, the subgroup differences being the result of permissible mutations.

(*b*) Six positions, 7, 33, 71, 83, 105, and 106, show predominantly κ- vs. λ-specificity. Thus at position 7, 41 human κ- and mouse κ-proteins have Ser while 20 human λ-specimens have Pro. Two exceptions occur, a κII protein with Thr instead of Ser, and a λV with Asp instead of Pro.

(*c*) At five additional positions, 1, 2, 13, 25, and 27, evidence of κ- vs. λ-specificity persists but a substitution may occur in one or more subgroups. Thus at position 1, 16 human λ-proteins of subgroups λI, λII, λIII have PCA, while λIV and λV have no amino acid, and 31 human and mouse κI and κII proteins have Asp (5 additional have Asx) while 14 human κIII specimens have Glu

Downloaded from www.jem.org on November 2, 2004

(1 additional with Glx). There are several exceptions: one human κIII with Lys and one with Asp and one mouse κII with PCA.

The 11 positions in categories *b* and *c* are of importance because they show that κ- and λ-specificity is not exclusively a property of the constant region but involves the variable region as well. These findings are consistent with the data indicating that the variable and constant regions of κ-chains always go together as do the variable and constant regions of λ-chains. They are also in accord with the immunochemical studies of Ruffilli and Baglioni (60) who showed that λ-specificity extended into the variable region, and with those of Ruffilli (61) that determinants in the variable region of κ cross-react with anti λ-sera. They also suggest a continuous evolutionary association for the variable and constant regions of the κ-chains as well as for the variable and constant regions of the λ-chains. The regions showing κ- and λ-specificity appear to be distributed at the beginning and end of the variable region.

(*d–f*) These three categories are instances of variation in composition ascribable to subgroup variation. At 9 positions, 3, 14, 18, 19, 22, 29, 42, 51, and 79, variation ascribable to subgroups is seen in both κ and λ chains. In an additional 10 positions, 4, 9, 10, 17, 20, 21, 55, 56, 77, and 85, subgroup variation is predominantly in κ-chains with λ relatively constant, and at 15 positions, 11, 26, 39, 47, 52, 66, 69, 72, 76, 78, 80, 89, 95, 97, and 107, the subgroup variation seems to involve λ-chains with κ relatively constant. The larger number of residues placed in category *f* is probably a consequence of the classification of λ-chains into five subgroups while κ-chains are only divided into three subgroups. Moreover the number of samples of λ-chains is fewer than for κ-chains so that other types of variation may be masked.

(*g*) Unaccountable variability. At 8 positions, 28, 30–32, 93, 94, 96, and 103, the variation within each subgroup appears to be greater than can be accounted for on any known basis, especially for those subgroups for which a sufficient number of samples have been examined. It is of interest that except for residue 103 these positions are clustered in the two regions of highest variability (45) and would be brought into close proximity by the disulfide bond $I_{23}$–$II_{88}$. It is postulated that these residues may be complementarity determining and actually be involved in making contact with the antigenic determinant. These residues are also very close to the position at which insertions occur (Table I).

(*h*) Insufficient data. At 15 positions, 24, 34, 36, 43, 45, 53, 58, 70, 74, 81, 87, 90–92, and 104, not enough sequences are available to assign them clearly to one of the other categories. Five of the residues, 24, 34, 90–92, occur in the two regions of highest variability close to most of those with unaccountable variability. The others are fairly well spread and only one, 53, occurs in the third highly variable region. Position 18, although placed in group *d* shows an extraordinary variability in the λI subgroup.

(*i*) Species-specific residues. At positions 50 and 60 the two mouse Bence

Downloaded from www.jem.org on November 2, 2004

Jones proteins examined have residues which do not correspond to any reported for the human proteins. Thus these two positions are still classifiable as species-specific. Position 96, with its extraordinary variability, although placed in category *g* might also technically be called species specific since the two mouse samples both contain Trp which is absent in the human samples. At all three positions the human proteins show substantial variability in the number of substitutions which can occur—9 at position 50, 5 at position 60, and 11 at position 96. It thus seems very unlikely that the apparent species specificity of these positions will persist when more mouse sequences are available.

## DISCUSSION

The ability to subject the large amount of sequence data on human and mouse Bence Jones proteins and light chains to a statistical analysis has supported the earlier conclusion about the lack of species specificity in the variable regions of human and mouse Bence Jones proteins (40). The data now available indicate at most 2 or 3 such residues in the variable region as compared with 36 in the constant region. Data in other species, although limited to the first few amino terminal residues, tend to support this. The rabbit has always been considered an important exception to the view that there was little if any species specificity in the variable region, since rabbit light chains had Ala and Ile as N-terminal residues. However, the demonstration by Hood et al. (62) that the N-terminal sequence of a homogeneous rabbit antibody to the C carbohydrate of the streptococcus was Ala-Asp-Val-Val-Met-Thr-Glu-Thr-Pro-Ala-Ser-Val indicates that the rabbit has merely added an N-terminal Ala to the N-terminal Asp, the other residues then being essentially similar to those in human light chains. Thus the rabbit is not an exception to the basic evolutionary unity of light chains.

The data now available (Table II) amply support the earlier suggestions for the role of the invariant glycines of the variable region both in conferring flexibility (Fig. 1) to permit substitutions at the variable positions, and for the glycines at positions 99 and 101, together with the frequently found glycine at residue 100, in functioning as a pivot to permit optimal fitting around the antigenic determinant (18, p. 87; and 41, 43). This postulate is now strongly supported by the finding of two glycines in an analogous region of the heavy chain. The heavy chain sequences thus far available also indicate that there may be other invariant glycines in the variable region (56, 57).

The data in Table IV show an unusual distribution of invariant residues in the variable region with respect to $H\emptyset_{ave}$ in that, with a few exceptions, these residues have either a high or a low or zero value, while $H\emptyset_{ave}$ values appear to be uniformly distributed throughout the invariant residues of the constant region (Table V). The significance and structural implications of this are not clear. This finding is not evident when one examines only the over-all hydro-

Downloaded from www.jem.org on November 2, 2004

phobicity, which does not differ significantly for the two halves of the chain (51).

The variability at each position shows that variability is concentrated in three regions of the molecule (49), of which two have also been noted by other workers (7, 45, 47). Examination of the basis for the variability at each position shows that variation ascribable to $\kappa$ and $\lambda$ or to their subgroups does not readily explain all of the variation. Seven of the eight positions at which the variability cannot be otherwise accounted for occur in two of these regions. Of the 15 positions for which sufficient sequence data were lacking to permit clear assignment, 5 occur in the same 2 regions—24–34 and 89–97 and those are the 2 regions at which additional residues are found (Table I). Much more sequence data will be required to permit unequivocal elucidation of the role of each variable position.

For the moment, if one accepts (*a*) the tendency of the positions of unaccountably high variability to be concentrated in the two short stretches of the chain at which insertions are found, (*b*) the finding that subgroup and group specificity occurs in the variable region and indeed probably over substantial portions of it, one might formulate the following working hypothesis extending the earlier concept (42, 45): The light chains of immunoglobulins except for the regions of unaccountable variability, 24–34 and 89–97, are governed by a number of structural genes, each chain being the product of two linked genes, one for the variable and one for the constant region (Hood et al. in reference 16). These structural genes are free to mutate and are limited only by the requirement that their product be capable of assuming the proper three-dimensional structure to permit an antibody site to be formed. By hypothesis the complementarity-determining residues are considered to be the result of the insertion into the DNA of the two short linear sequences, 24–34 and 89–97, which specifically determine what kind of antibody site will be formed. In the light chain the two insertions would be brought into close proximity by the disulfide bond $I_{23}$–$II_{88}$. A similar type of insertion would be made in the heavy chain, but thus far there is evidence for only one region of high variability (56, 57).

An insertion mechanism involving only short linear sequences provides a substantial simplification of the problem of providing a seemingly limitless number of complementary sites without the use of very large amounts of DNA. It would also be more likely to provide the necessary evolutionary stability and universality for the antibody-forming mechanism which cannot be adequately accounted for by the germ line hypothesis (36) of one gene for each variable region since such a system would diverge on an evolutionary time scale. The precision with which the insertion would have to be accomplished, (e.g. changes in length of one or two nucleotides would result in nonsense) in itself would tend to prevent evolutionary divergence since failures in the insertional mechanism

Downloaded from www.jem.org on November 2, 2004

would completely destroy the ability to form any antibody and individuals with such a defect would probably not survive. It eliminates difficulties in the translocation hypothesis which provides for the joining of one of many V genes with a C gene (33) and yet maintains the exclusive association of $V_\kappa$ with $C_\kappa$ and $V_\lambda$ with $C_\lambda$ genes. Indeed the finding that $\kappa$- and $\lambda$-specificity extends into the variable region supports the concept of one polypeptide chain resulting from the action of closely linked V and C genes. This also is consistent with the allotypic studies on $\gamma$G, $\gamma$M and $\gamma$A (63–65). Moreover it has the additional merit of providing a fixed location for the antibody site while other theories (32, 34, 35, 39, 66) permit it to be formed by different portions of the variable region for various antibodies. Indeed Eisen (16) considers that three different sites, each of a different specificity, might be formed by a given $V_L$ and $V_H$ pair. This would make antibody sites completely different from all other sets of specific receptors known. The present hypothesis considers the site as involving a small fixed region of the molecule with specificity determined by the differences in side chains of complementarity-determining residues. This insertion hypothesis readily accounts for the variations in length occurring in these regions; no other hypothesis has explicitly considered this or adequately accounted for it.

On the basis of this hypothesis which ascribes only a role in three-dimensional folding to the first 23 N-terminal amino acid residues of the light chain, all recombinational theories of antibody formation (66, 67) become uninformative since they are based exclusively on sequence data in this region and thus are probably not dealing with a region involving antibody complementarity.

The present working hypothesis of linear regions determining antibody complementarity will stand or fall when adequate numbers of sequences for human and mouse light (and heavy) chains have been worked out.

The hypothesis makes certain predictions which also can be used to test its validity. One of these, since the insertion is hypothesized to determine complementarity, is that antibody molecules of a given specificity and with a uniform site can occur in any class or subclass of immunoglobulin by the insertion of the given short linear sequences. The finding (8) that human antidextran of $\alpha$-(1 $\rightarrow$ 6) specificity may occur in $\gamma$A, $\gamma$M, $\gamma$G2, and sometimes in $\gamma$G1 immunoglobulins and may have $\kappa$- or $\lambda$-chains is consistent with but does not provide conclusive evidence for this hypothesis, since the human antidextran still represents mixtures with heterogeneous sizes of combining sites. More important, however, may be the findings of Pincus et al. (14) that rabbit type III and type VIII antibodies to the pneumococcal polysaccharide, which gave a straight line of slope not significantly different from 1.0 in a Sips plot for binding of an octasaccharide and could thus be considered quite homogeneous with respect to their antibody-combining sites, showed many light chains and several heavy chains on acrylamide gel electrophoresis. Thus a mixture of structurally

Downloaded from www.jem.org on November 2, 2004

different antibody molecules could have the same binding affinity and therefore probably have very similar or even identical combining sites. It would be especially important to determine whether such mixtures of antibodies also belonged to different classes and had antigenically different light and heavy chains.

Further studies on homogeneous antibodies of a given specificity but of different classes or subclasses would be predicted to show similar sequences in the insertional regions if their sites were identical. Conversely, antibodies of a given specificity but showing differences in the degree of cross-reactivity with related antigens would be expected to show smaller differences in their insertion regions than would antibodies of totally unrelated specificities. Such data may soon become available as sequences on myeloma proteins with antibody activity and on homogeneous antibodies are accumulated.

While such findings would provide strong support for the hypothesis, it should be borne in mind that the contour of an antibody site having a given specificity or binding affinity for a given determinant could conceivably be formed by several kinds of patterns of amino acid sequences. Under such circumstances, antibodies which were homogeneous in binding affinity but were mixtures of molecules with different sequences in the hypothesized insertional regions would be expected not to give unique sequences in these regions. Moreover, the possibility exists that binding could be influenced to some extent by different residues adjacent to a site but not in themselves complementarity determining.

The data on idiotypic specificity of myeloma globulins (68) and antibodies (69–73) are compatible with the insertion model and with the over-all concept of antibody structure proposed. Thus idiotypic determinants which are found in the variable regions would represent antigenic determinants formed by patterns of amino acid sequences involving some of the side chains of residues from the inserted regions—namely those forming the exterior portions of the site but also including some of the residues involved in three-dimensional folding and belonging to various subgroups, etc. This could give rise to a large number of determinants generally not related to specificity but influenced by or indeed partly created by the sequence of site-determining residues. In many instances in which immunodominant groups of the idiotypic determinants were from those of the inserted regions, one might expect the same idiotypic specificity to be manifested in several classes of immunoglobulins; this has been shown to be the case for $\gamma$M and $\gamma$G antibodies from the same rabbit (71). Thus the findings on idiotypic specificity provide further support for the uniqueness and universality of the antibody-forming mechanism.

Several models for insertion of information into DNA have been recognized. One of these is the episomal model (74, 75), and another involves two recombinations as in P1 phage transduction (76, 77). A self-perpetuating episome

Downloaded from www.jem.org on November 2, 2004

containing a large number of short nucleotide sequences the incorporation of any one of which into the structural genes of the light and heavy chains to provide the information for a given antibody complementarity provides a tempting mechanism. Such incorporation could be during embryogenesis, by which each cell receives the proper nucleotide sequences to program its structural genes for a given antibody specificity. Alternatively, if a cell is multipotent, it could have a number of sequences or the entire episome and the insertion of the proper sequence could be accomplished in some unknown manner after antigenic stimulation. While the evidence shows that one cell produces one kind of antibody at a time, and that myeloma cells produce populations of molecules which bind ligands in a homogeneous manner, such cells would already be programmed. The hypothesized programming could possibly result from the cell-cell interactions for which some evidence has been advanced. Systems of this type include transfer of information from macrophages to lymphoid cells (78) and two cell interactions such as thymus-bone marrow, etc. (16, p. 431, 79-81). There is no basis at present for any more than a passing reference to these as possibilities.

Placing the information for site complementarity in an independent self-duplicating mechanism would provide the evolutionary stability needed. It would lead to the universality which the antibody-forming system has clearly manifested, because the requirements for successful insertion would be relatively more stringent so that changes in the insertion material would have a higher frequency for completely destroying the capacity to form antibodies and such individuals would probably not survive.

There are difficulties in applying the episomal model to the antibody complementarity problem. The insertion in the antibody case would have to be by a recombination mechanism involving overlapping sequences on each side as in the P1 phage transduction, rather than as in the Campbell model. However, the degree of overlap on each side of the postulated insertions is very small, e.g., an invariant Cys at the beginning and invariant Trp 35 and Phe 98 at the end. Moreover episomes have not been found to date except in bacteria, although several possibly relevant systems in eukaryotes have been noted (74). It is also difficult to see how insertions to produce a given antibody specificity could be programmed for both the light and heavy chains since their contributions to binding are generally quite different.

The insertion model does not necessarily distinguish between germ line and somatic theories, for generating the diversity necessary to provide a large number of sites. Such diversity could be obtained by the existence of a large dictionary of insertions each of which determined a given specificity, e.g. by a germ line theory, or by the recombination of the nucleotides within the insertion material of a relatively small number of different sequences to produce a large number of recombinant sequences. This latter alternative might result in the

Downloaded from www.jem.org on November 2, 2004

formation of site-determining sequences which were not exactly the same for a given antibody in different species or in different individuals. Should sequence data on homogeneous antibodies to a given antigenic determinant formed in various species show that the same sequence is complementarity determining, the dictionary model would be favored over the recombinational version. Similarly, if homogeneous antibodies to a given determinant in the same individual but belonging to the various classes ($\gamma$G, $\gamma$M, $\gamma$A) or subclasses ($\gamma$G1, $\gamma$G2, etc.) of immunoglobulins have the same complementarity determining residues, a dictionary model would also be favored.

The suggestions put forward are admittedly speculative as is the case at the moment with all other hypotheses. They do however present the problem of the antibody-combining site and of immunoglobulin structure in a different way, but in a way which is susceptible to verification or disproof by further data. Whether the model proposed ultimately stands or falls, the effort to assign each amino acid residue in the variable region a definite role by the statistical analysis employed should prove useful.

SUMMARY

In an attempt to account for antibody specificity and complementarity in terms of structure, human $\kappa$-, human $\lambda$-, and mouse $\kappa$-Bence Jones proteins and light chains are considered as a single population and the variable and constant regions are compared using the sequence data available. Statistical criteria are used in evaluating each position in the sequence as to whether it is essentially invariant or group-specific, subgroup-specific, species-specific, etc.

Examination of the invariant residues of the variable and constant regions confirms the existence of a large number of invariant glycines, no invariant valine, lysine, and histidine, and only one invariant leucine and alanine in the variable region, as compared with the absence of invariant glycines and presence of three each of invariant alanine, leucine, and valine and two each of invariant lysine and histidine in the constant region. The unique role of glycine in the variable region is emphasized. Hydrophobicity of the invariant residues of the two regions is also evaluated. A parameter termed variability is defined and plotted against the position for the 107 residues of the variable region. Three stretches of unusually high variability are noted at residues 24–34, 50–56, and 89–97; variations in length have been found in the first and third of these. It is hypothesized that positions 24–34 and 89–97 contain the complementarity-determining residues of the light chain—those which make contact with the antigenic determinant. The heavy chain also has been reported to have a similar region of very high variability which would also participate in forming the antibody-combining site. It is postulated that the information for site complementarity is contained in some extrachromosomal DNA such as an episome and is incorporated by insertion into the DNA of

Downloaded from www.jem.org on November 2, 2004

the structural genes for the variable region of short linear sequences of nucleotides. The advantages and disadvantages of this hypothesis are discussed.

## BIBLIOGRAPHY


1. Kabat, E. A. 1966. The nature of an antigenic determinant. *J. Immunol.* **97**:1.
2. Kabat, E. A. 1968. Structural Concepts in Immunology and Immunochemistry. Holt, Rinehart & Winston, Inc., New York.
3. Goodman, J. W. 1969. Immunochemical specificity: Recent conceptual advances. *Immunochemistry.* **6**:139.
4. Cohen, S., and R. R. Porter. 1964. Structure and biological activity of immunoglobulins. *Advan. Immunol.* **4**:287.
5. Kabat, E. A. 1966. Structure and heterogeneity of antibodies. *Acta haematol.* **36**:198.
6. Cohen, S., and C. Milstein. 1967. Structure and biological properties of immunoglobulins. *Advan. Immunol.* **7**:1.
7. Milstein, C., and J. R. L. Pink. 1970. Structure and evolution of immunoglobulins. *Prog. Biophys. Mol. Biol.* **21**:209.
8. Yount, W. J., M. M. Dorner, H. G. Kunkel, and E. A. Kabat. 1968. Studies on human antibodies. VI. Selective variations in subgroup composition and genetic markers. *J. Exp. Med.* **127**:633.
9. Dorner, M. M., W. J. Yount, and E. A. Kabat. 1969. Studies on human antibodies. VII. Acrylamide gel electrophoresis of purified human antibodies and myeloma proteins, their heavy and light chains. *J. Immunol.* **102**:273.
10. Fleischman, J. B., D. G. Braun, and R. M. Krause. 1968. Streptococcal group-specific antibodies: Occurrence of a restricted population following secondary immunization. *Proc. Nat. Acad. Sci. U. S. A.* **60**:134.
11. Braun, D. G., and R. M. Krause. 1968. The individual antigenic specificity of antibodies to streptococcal carbohydrates. *J. Exp. Med.* **128**:969.
12. Eichmann, K., H. Lackland, L. Hood, and R. M. Krause. 1970. Induction of rabbit antibody with molecular uniformity after immunization with group C streptococci. *J. Exp. Med.* **131**:207.
13. Pappenheimer, A. M., Jr., W. P. Reed, and R. Brown. 1968. Quantitative studies of the specificity of anti-pneumococcal polysaccharide antibodies, types III and VIII. III. Binding of a labeled oligosaccharide derived from S8 by anti-S8 antibodies. *J. Immunol.* **100**:1237.
14. Pincus, J. H., E. Haber, M. Katz, and A. M. Pappenheimer, Jr. 1968. Antibodies to pneumococcal polysaccharides: Relation between binding and electrophoretic heterogeneity. *Science* (*Washington*) **162**:667.
15. Brenneman, L., and S. J. Singer. 1968. The generation of antihapten antibodies with electrophoretically homogeneous L chains. *Proc. Nat. Acad. Sci. U. S. A.* **60**:258.
16. Frisch, L., editor. 1967. Antibodies. *Cold Spring Harbor Symp. Quant. Biol.* **32**:1–603.
17. Rossi, G., T. K. Choi, and A. Nisonoff. 1969. Crystals of fragment $F_{ab}'$: Preparation from pepsin digests of human IgG myeloma proteins. *Nature* (*London*) **223**:837.


Downloaded from www.jem.org on November 2, 2004

18. Killander, J., editor. 1967. Gamma Globulins. *Nobel Symp.* **3**:17–643.
19. Beaumont, J. L. 1967. Une spécificité commune aux α- et β-lipoprotéines du sérum révélée par un autoanticorps de myélome-L'antigène Pg. *C. R. Acad. Sci Ser. D.* **264**:185.
20. Eisen, H. N., E. S. Simms, and M. Potter. 1968. Mouse myeloma proteins with antihapten antibody activity. The protein produced by plasma cell tumor MOPC-315. *Biochemistry.* **7**:4196.
21. Metzger, H., and M. Potter. 1968. Affinity site labeling of a mouse myeloma protein which binds dinitrophenyl ligands. *Science (Washington).* **162**:1398.
22. Potter, M., and M. A. Leon. 1968. Three IgA myeloma immunoglobulins from BALB/c mouse:Precipitation with pneumococcal C polysaccharide. *Science (Washington).* **162**:369.
23. Schubert, D., A. Jobe, and M. Cohn. 1968. Mouse myeloma producing precipitating antibody to nucleic acid bases and/or nitrophenyl derivatives. *Nature (London).* **220**:882.
24. Ashman, R. F., and H. Metzger. 1969. A Waldenström macroglobulin which binds nitrophenyl ligands. *J. Biol. Chem.* **244**:3405.
25. Cohn. M., G. Notani, and S. A. Rice. 1969. Characterization of the antibody to the C-carbohydrate produced by a transplantable mouse plasmacytoma. *Immunochemistry.* **6**:111.
26. Capra, J. D., D. Wertheimer, W. J. Yount, and H. G. Kunkel. 1969. Monoclonal γG anti-γ globulins in hypergammaglobulinemic purpura. *Clin. Res.* **17**:351.
27. American Association of Immunologists Symposium on Experimental Approaches to Homogeneous Antibody Populations. 1970. *Fed. Proc.* **29**:55–91.
28. Potter, M. 1967. The plasma cell tumors and myeloma proteins of mice. *Methods Cancer Res.* **2**:106.
29. Edelman, G. M., and J. A. Gally. 1962. The nature of Bence Jones proteins. Chemical similarities to polypeptide chains of myeloma globulins and normal γ-globulins. *J. Exp. Med.* **116**:207.
30. Quattrocchi, R., D. Cioli, and C. Baglioni. 1969. A study of immunoglobulin structure. III. An estimate of the variability of human light chains. *J. Exp Med.* **130**:401.
31. Lennox, E. S., and M. Cohn. 1967. Immunoglobulins. *Annu. Rev. Biochem.* **36**:365.
32. Cohn, M. 1968. The molecular biology of expectation. *In* Nucleic Acids in Immunology. O. J. Plescia, and W. Braun, editors. Springer-Verlag, New York. 671.
33. Edelman, G. M., and W. E. Gall. 1969. The antibody problem. *Annu. Rev. Biochem.* **38**:415.
34. Hood, L., and D. W. Talmage. 1969. On the mechanism of antibody diversity: Evidence for the germ line basis of antibody variability. *In* Symposium on Developmental Aspects of Antibody Formation and Structure, Prague. In press.
35. Hood, L., and D. W. Talmage. 1970. Mechanism and antibody diversity: Germ line basis for variability. *Science (Washington)* **168**:325.
36. Dreyer, W. J., and J. C. Bennett. 1965. The molecular basis of antibody formation: A paradox. *Proc. Nat. Acad. Sci. U. S. A.* **54**:864.
37. Lederberg, J. 1959. Genes and antibodies. *Science (Washington).* **129**:1649.

Downloaded from www.jem.org on November 2, 2004

38. Burnet, M. 1966. A possible genetic basis for specific pattern in antibody. *Nature (London)* **210**:1308.
39. Brenner, S., and C. Milstein. 1966. Origin of antibody variation. *Nature (London)*. **211**:242.
40. Kabat, E. A. 1967. The paucity of species-specific amino acid residues in the variable regions of human and mouse Bence Jones proteins and its evolutionary and genetic implications. *Proc. Nat. Acad. Sci. U. S. A.* **57**:1345.
41. Kabat, E. A. 1967. A comparison of invariant residues in the variable and constant regions of human K, human L and mouse K Bence Jones proteins. *Proc. Nat. Acad. Sci. U. S. A.* **58**:229.
42. Kabat, E. A. 1968. Unique features of the variable regions of Bence Jones proteins and their possible relation to antibody complementarity. *Proc. Nat. Acad. Sci. U. S. A.* **59**:613.
43. Kabat, E. A. 1969. Antibody complementarity and light chain structure. *In* Symposium on the Evolution of Immune Response. Oct. 20–22, 1969. U.S. Argonne National Laboratories, Lemont, Ill.
44. Fitch, W. M., and E. Margoliash. 1967. Construction of phylogenetic trees. *Science (Washington)*. **155**:279.
45. Kabat, E. A. 1970. Heterogeneity and structure of antibody-combining sites. *Ann. N. Y. Acad. Sci.* **169**:43.
46. Phillips, D. C. 1967. The hen egg-white lysozyme molecule. *Proc. Nat. Acad. Sci. U. S. A.* **57**:484.
47. Milstein, C. 1967. Linked groups of residues in immunoglobulin K chains. Nature (*London*). **216**:330.
48. Niall, H., and P. Edman. 1967. Two structurally distinct classes of kappa-chains in human immunoglobulins. *Nature (London)*. **216**:262.
49. Franěk, F. 1969. The character of variable sequences in immunoglobulins and its evolutionary origin. *In* Symposium on Developmental Aspects of Antibody Formation and Structure, Prague. In press.
50. Jukes, T. H. 1969. Evolutionary pattern of specificity regions in light chains of immunoglobulins. *Biochem. Genet.* **3**:109.
51. Welscher, H. D. 1969. Correlations between amino acid sequence and conformation of immunoglobulin light chains. I. Hydrophobicity and fraction charge. *Int. J. Protein Res.* **1**:235. II. Sequence comparison and the pattern of nonpolar residues. *Int. J. Protein Res.* **1**:267.
52. Blake, C. C. F., A. G. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma. 1967. On the conformation of the hen egg-white lysozyme molecule. *Proc. Roy. Soc. Ser. B. Biol. Sci.* **167**:365.
53. Birktoft, J. J., B. W. Matthews, and D. M. Blow. 1969. Atomic coordinates of tosyl-α-chymotrypsin. *Biochem. Biophys. Res. Commun.* **36**:131.
54. Edsall, J. T., P. L. Flory, J. C. Kendrew, A. M. Liquori, G. Némethy, G. N. Ramachandran, and H. A. Scheraga. 1966. A proposal of standard conventions and nomenclature for the description of polypeptide conformations. *Biopolymers* **4**:121.
55. Ramachandran, G. N., and V. Sasisekharan. 1968. Conformation of polypeptides and proteins. *Advan. Protein Chem.* **23**:283.

Downloaded from www.jem.org on November 2, 2004

56. Press, E. M., and N. M. Hogg. 1969. Comparative study of two immunoglobulin G Fd-fragments. *Nature* (*London*). **223**:807.
57. Cunningham, B. A., M. N. Pflumm, U. Rutishauser, and G. M. Edelman. 1969. Subgroups of amino acid sequences in the variable regions of immunoglobulin heavy chains. *Proc. Nat. Acad. Sci. U. S. A.* **64**:997.
58. Tanford, C. 1962. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Amer. Chem. Soc.* **84**:4240.
59. Bigelow, C. C. 1967. On the average hydrophobicity of proteins and the relation between it and protein structure. *J. Theor. Biol.* **16**:187.
60. Ruffilli, A., and C. Baglioni. 1967. Subgroups of L type Bence Jones proteins. *J. Immunol.* **98**:874.
61. Ruffilli, A. 1968. The antigenic determinants of the variable half of a Bence Jones protein of type L. *J. Immunol.* **100**:201.
62. Hood, L., H. Lackland, K. Eichmann, J. Ohms, and R. M. Krause. 1970. Amino acid sequences from rabbit antibody light chains and gene evolution. In preparation.
63. Todd, C. W. 1963. Allotypy in rabbit 19S protein. *Biochem. Biophys. Res. Commun.* **11**:170.
64. Todd, C. W. 1966. Discussion. *J. Cell. Physiol.* **67**(Suppl. *1*):95.
65. Koshland, M. E., J. J. Davis, and N. J. Fujita. 1969. Evidence for multiple gene control of a single polypeptide chain: The heavy chain of rabbit immunoglobulin. *Proc. Nat. Acad. Sci. U. S. A.* **63**:1274.
66. Smithies, O. 1963. Gamma-globulin variability: A genetic hypothesis. *Nature* (*London*). **199**:1231.
67. Edelman, G. M., and J. A. Gally. 1967. Somatic recombination of duplicated genes: An hypothesis on the origin of antibody diversity. *Proc. Nat. Acad. Sci. U. S. A.* **57**:353.
68. Slater, R. J., S. M. Ward, and H. G. Kunkel. 1955. Immunological relationships among the myeloma proteins. *J. Exp. Med.* **101**:85.
69. Kunkel, H. G., M. Mannik, and R. C. Williams. 1963. Individual antigenic specificities of isolated antibodies. *Science* (*Washington*) **140**:1218.
70. Oudin, J., and M. Michel. 1963. Une nouvelle forme d'allotypie des globulines $\gamma$ de sérum de lapin apparemment liéé à la jonction et à la specificité anticorps. *C. R. Acad. Sci.* **257**:805.
71. Oudin, J., and M. Michel. 1969. Idiotypy of rabbit antibodies. I. Comparison of idiotype of antibodies against *Salmonella typhi* with that of antibodies against other bacteria in the same rabbit, or of antibodies against *Salmonella typhi* in various rabbits. *J. Exp. Med.* **130**:595. II. Comparison of idiotypy of various kinds of antibodies formed in the same rabbit against *Salmonella typhi*. *J. Exp. Med.* **130**:619.
72. Daugharty, H., J. E. Hooper, A. B. MacDonald, and A. Nisonoff. 1969. Quantitative investigations of idiotypic antibodies. I. Analysis of precipitating antibody populations. *J. Exp. Med.* **130**:1047.
73. Hurez, D., G. Meshaka, C. Mihaesco, and M. Seligmann. 1968. The inhibition by normal $\gamma$G-globulins of antibodies specific for individual $\gamma$G myeloma proteins. *J. Immunol.* **100**:69.

Downloaded from www.jem.org on November 2, 2004

74. Campbell, A. M. 1962. Episomes. *Advan. Genet.* **11**:101.
75. Campbell, A. M. 1969. Episomes. Harper & Row, Publishers, New York.
76. Lennox, E. S 1955. Transduction of linked genetic characters of the host by bacteriophage P1. *Virology*. **1**:190.
77. Ozeki, H., and H. Ikeda. 1968. Transduction mechanisms. *Annu. Rev. Genet.* **2**:245.
78. Adler, F. L., M. Fishman, and S. Dray. 1966. Antibody formation initiated *in vitro*. III. Antibody formation and allotypic specificity directed by ribonucleic acid from peritoneal exudate cells. *J. Immunol.* **97**:554.
79. Miller, J. F. A. P., and G. F. Mitchell. 1968. Cell to cell interaction in the immune response. I. Hemolysin-forming cells in neonatally thymectomized mice reconstructed with thymus or thoracic duct lymphocytes. *J. Exp. Med.* **128**:801.
80. Mitchell, G. F., and J. F. A. P. Miller. 1968. Cell to cell interaction in the immune response. II. The source of hemolysin-forming cells in irradiated mice given bone marrow and thymus or thoracic duct lymphocytes. *J. Exp. Med.* **128**:821.
81. Claman, H. N., E. A. Chaperon, and R. F. Triplett. 1966. Immunocompetence of transferred thymus-marrow cell combinations. *J. Immunol.* **97**:828.

Downloaded from www.jem.org on November 2, 2004

**Similar studies for antibody heavy chains**

While the antibody light chains can be isolated relatively easily from the urines of patients suffering from multiple myeloma, the corresponding antibody heavy chains are more difficult to purify. The myeloma protein has to be first isolated from the serum of these patients. Then the disulfide bonds have to be broken and the heavy chain separated from the light chain. Several investigators determined the antibody heavy chain sequences around 1970, i.e. about five years after the amino acid sequences of Bence Jones proteins or antibody light chains.

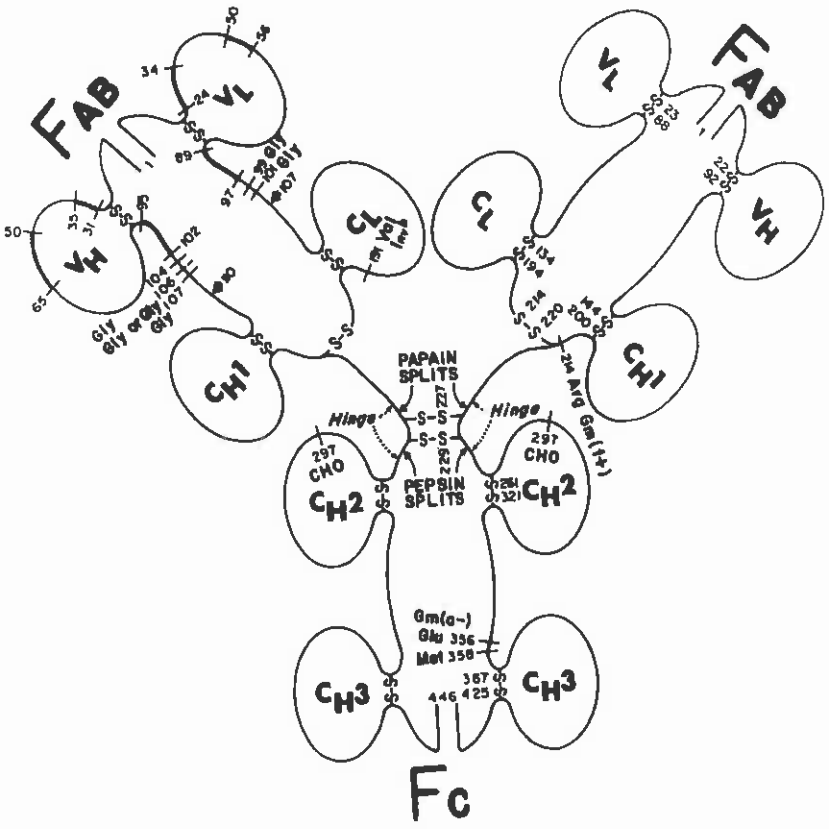
A similar variability plot for the antibody heavy chain variable region amino acid sequences also identified three peaks. It was thus proposed that these six segments or complementarity determining regions (CDRs), three each from the light and heavy chains, constituted the antibody combining site. It was like a six-fingered hand holding the antigen molecule. If each CDR could have ten different amino acid sequences, the random assortment of these six CDRs could easily generate $10^6$ different antibodies. Indeed, there are many more different CDR sequences. As a result, the number of different antibodies can be much larger than one million.

The variable regions of antibody light and heavy chains can thus be divided into seven segments: four framework regions (FR1, FR2, FR3 and FR$), and three complementarity determining regions (CDR1, CDR2 and CDR3), precisely located at the following positions (using the Kabat numbering system):

| Segment | Light chain | Heavy chain |
|---|---|---|
| FR1 | 1 – 23 | 1 – 30 |
| CDR1 | 24 – 34 | 31 – 35 |
| FR2 | 25 – 49 | 36 – 49 |
| CDR2 | 50 – 56 | 50 – 65 |
| FR3 | 57 – 88 | 66 – 94 |
| CDR3 | 89 – 97 | 95 – 102 |
| FR4 | 98 – 107 | 103 – 113 |

For more precise descriptions, see Wu TT (2001) Analytical Molecular Biology, Kluwer Academic Publishers.

The Y-shaped antibody molecule can thus be represented by the figure shown on the next page.


FAB
VL
VH
CL
CH1
CH2
CH3
FAB
VL
VH
CL
CH1
PAPAIN SPLITS
Hinge
Hinge
S-S
S-S
PEPSIN SPLITS
297 CHO
297 CHO
2nd Avg Gm (1+)
Gm(a-)
Glu 356
Met 358
FC

**Hybridoma technology**

While the theoretical prediction of the antibody combining site being formed by six completementarity determining regions appeared very attractive, Eventual verification by experimental studies would be essential. Several of the Fab fragments of myeloma proteins from both human and mouse have been crystallized, and their CDRs were indeed located on one side of these molecules. Unfortunately, what antigens these myeloma proteins would bind was completely unknown.

In 1975, Kohler and Milstein (Kohler G, Milstein C (1975) Continous cultures of sudes cells secreting antibody of predefined specificity. Nature, 256- 258) discovered an ingenious method of fusing myeloma cells with B lymphocytes secreting specific antibodies. These cells are now known as hybridoma cells and form the basic technique of generating monoclonal antibodies in large quantity.

While the fusing partners would each produce a light and a heavy chain, the combination of these two cells into one ended up in producing mixtures of light and heavy combinations. Currently, some of the myeloma cell lines used for hybridoma production are non-secretors, i.e. they do not produce complete light and heavy chains. Therefore, only the required antibody from the selected B lymphocytes will be produced. Unfortunately, most of the hybridoma cell lines are made with mouse B lymphocytes. Somehow, hybridomas made from human B lymphocytes are somewhat unstable. There are probably also some other reasons requiring containment.

Thus the stage was ste to perform the crucial experiment for the verification of the theoretical predictions of the antibody combining site being formed by all six CDRs. That required detailed X-ray diffraction studies of protein complexes at high resolution. Finally it was achieved in 1986 as discussed in the next section.

Reprinted with permission from *Nature* 256 (5517): 495–497 (1975). Copyright © 1975 Macmillan Publishers Ltd.

*Nature Vol. 256 August 7 1975* 495

# Continuous cultures of fused cells secreting antibody of predefined specificity

THE manufacture of predefined specific antibodies by means of permanent tissue culture cell lines is of general interest. There are at present a considerable number of permanent cultures of myeloma cells[1-3] and screening procedures have been used to reveal antibody activity in some of them. This, however, is not a satisfactory source of monoclonal antibodies of predefined specificity. We describe here the derivation of a number of tissue culture cell lines which secrete anti-sheep red blood cell (SRBC) antibodies. The cell lines are made by fusion of a mouse myeloma and mouse spleen cells from an immunised donor. To understand the expression and interactions of the Ig chains from the parental lines, fusion experiments between two known mouse myeloma lines were carried out.

Each immunoglobulin chain results from the integrated expression of one of several *V* and *C* genes coding respectively for its variable and constant sections. Each cell expresses only one of the two possible alleles (allelic exclusion; reviewed in ref. 3). When two antibody-producing cells are fused, the products of both parental lines are expressed[4,5], and although the light and heavy chains of both parental lines are randomly joined, no evidence of scrambling of *V* and *C* sections is observed[4]. These results, obtained in an heterologous system involving cells of rat and mouse origin, have now been confirmed by fusing two myeloma cells of the same mouse strain, and provide the background for the derivation and understanding of antibody-secreting hybrid lines in which one of the parental cells is an antibody-producing spleen cell.

Two myeloma cell lines of BALB/c origin were used. P1Bul is resistant to 5-bromo-2'-deoxyuridine[4], does not grow in selective medium (HAT, ref. 6) and secretes a myeloma protein, Adj PC5, which is an IgG2A (κ), (ref. 1). Synthesis is not balanced and free light chains are also secreted. The second cell line, P3-X63Ag8, prepared from P3 cells[2], is resistant to 20 μg ml$^{-1}$ 8-azaguanine and does not grow in HAT medium. The protein secreted (MOPC 21) is an IgG1 (κ) which has been fully sequenced[7-9]. Equal numbers of cells from each parental line were fused using inactivated Sendai virus[5] and samples containing $2 \cdot 10^5$ cells were grown in selective medium in separate dishes. Four out of ten dishes showed growth in selective medium and these were taken as independent hybrid lines, probably derived from single fusion events. The karyotype of the hybrid cells after 5 months in culture was just under the sum of the two parental lines (Table 1). Figure 1 shows the isoelectric focusing[10] (IEF) pattern of the secreted products of different lines. The hybrid cells (samples *c–h* in Fig. 1) give a much more complex pattern than either parent (*a* and *b*) or a mixture of the parental lines (*m*). The important feature of the new pattern is the presence of extra bands (Fig. 1, arrows). These new bands, however, do not seem to be the result of differences in primary structure; this is indicated by the IEF pattern of the products after reduction to separate the heavy and light chains (Fig. 1*B*). The IEF pattern of chains of the hybrid clones (Fig. 1*B*, *g*) is equivalent to the sum of the IEF pattern (*a* and *b*) of chains of the parental clones with no evidence of extra products. We conclude that, as previously shown with interspecies hybrids[4,5], new Ig molecules are produced as a result of mixed association between heavy and light chains from the two parents. This process is intracellular as a mixed cell population does not give rise to such hybrid molecules (compare *m* and *g*, Fig. 1*A*). The individual cells must therefore be able to express both isotypes. This result shows that in hybrid cells the expression of one isotype and idiotype does not exclude the expression of another: both heavy chain isotypes (γ1 and γ2a) and both $V_H$ and both $V_L$ regions (idiotypes) are expressed. There are no allotypic markers for the $C_\kappa$ region to provide direct proof for the expression of both parental $C_\kappa$ regions. But this is indicated by the phenotypic link between the *V* and *C* regions.

Figure 1*A* shows that clones derived from different hybridisation experiments and from subclones of one line are indistinguishable. This has also been observed in other experiments (data not shown). Variants were, however, found in a survey of 100 subclones. The difference is often associated with changes

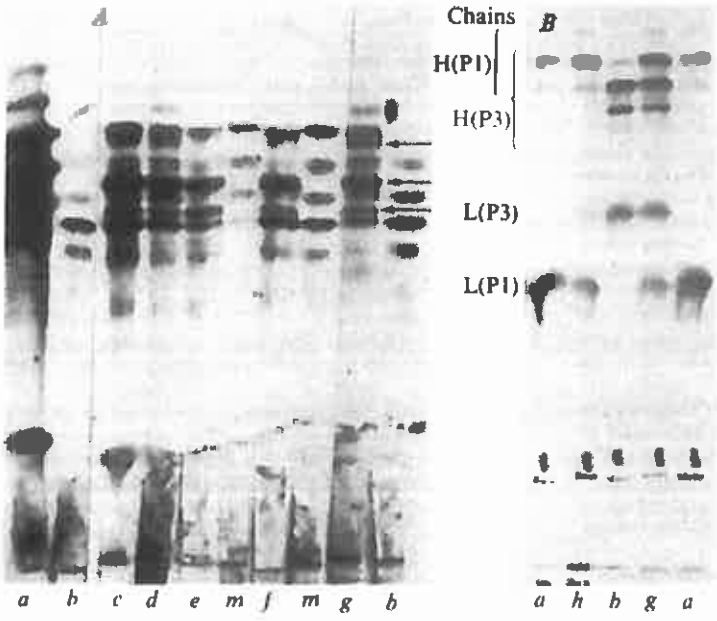



Fig. 1 Autoradiograph of labelled components secreted by the parental and hybrid cell lines analysed by IEF before (*A*) and after reduction (*B*). Cells were incubated in the presence of $^{14}$C-lysine[11] and the supernatant applied on polyacrylamide slabs. *A*, pH range 6.0 (bottom) to 8.0 (top) in 4 M urea. *B*, pH range 5.0 (bottom) to 9.0 (top) in 6 M urea; the supernatant was incubated for 20 min at 37 °C in the presence of 8 M urea, 1.5 M mercaptoethanol and 0.1 M potassium phosphate pH 8.0 before being applied to the right slab. Supernatants from parental cell lines in: *a*, P1Bul; *b*, P3-X67Ag8; and *m*, mixture of equal number of P1Bul and P3-X67Ag8 cells. Supernatants from two independently derived hybrid lines are shown: *c–f*, four subclones from Hy-3, *g* and *h*, two subclones from Hy-B. Fusion was carried out[4,5] using $10^6$ cells of each parental line and 4,000 haemagglutination units inactivated Sendai virus (Searle). Cells were divided into ten equal samples and grown separately in selective medium (HAT medium, ref. 6). Medium was changed every 3 d. Successful hybrid lines were obtained in four of the cultures, and all gave similar IEF patterns. Hy-B and Hy-3 were further cloned in soft agar[12]. L, Light; H, heavy.

97

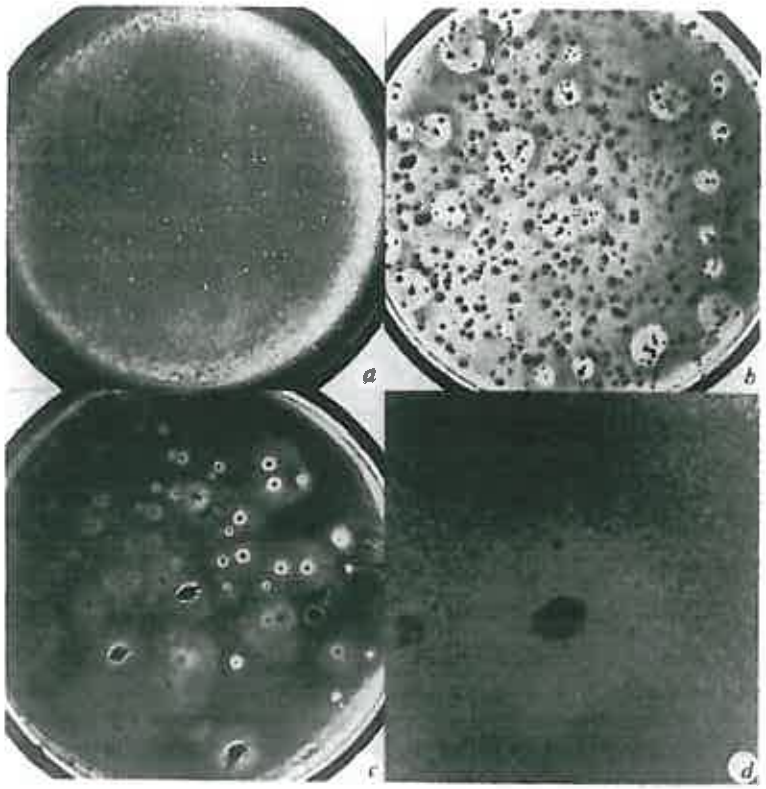


Fig. 2 Isolation of an anti-SRBC antibody-secreting cell clone. Activity was revealed by a halo of haemolysed SRBC. Direct plaques given by: *a*, 6,000 hybrid cells Sp-1; *b*, clones grown in soft agar from an inoculum of 2,000 Sp-1 cells; *c*, recloning of one of the positive clones Sp-1/7; *d*, higher magnification of a positive clone. Myeloma cells ($10^7$ P3-X67A g8) were fused to $10^8$ spleen cells from an immunised BALB/c mouse. Mice were immunised by intraperitoneal injection of 0.2 ml packed SRBC diluted 1:10, boosted after 1 month and the spleens collected 4 d later. After fusion, cells (Sp-1) were grown for 8 d in HAT medium, changed at 1–3 d intervals. Cells were then grown in Dulbecco modified Eagle's medium, supplemented for 2 weeks with hypoxanthine and thymidine. Forty days after fusion the presence of anti-SRBC activity was revealed as shown in *a*. The ratio of plaque forming cells/total number of hybrid cells was 1/30. This hybrid cell population was cloned in soft agar (50% cloning efficiency). A modified plaque assay was used to reveal positive clones shown in *b–d* as follows. When cell clones had reached a suitable size, they were overlaid in sterile conditions with 2 ml 0.6% agarose in phosphate-buffered saline containing 25 μl packed SRBC and 0.2 ml fresh guinea pig serum (absorbed with SRBC) as source of complement. *b*, Taken after overnight incubation at 37 °C. The ratio of positive/total number of clones was 1/33. A suitable positive clone was picked out and grown in suspension. This clone was called Sp-1/7, and was recloned as shown in *c*; over 90% of the clones gave positive lysis. A second experiment in which $10^6$ P3-X67Ag8 cells were fused with $10^8$ spleen cells was the source of a clone giving rise to indirect plaques (clone Sp-2/3-3). Indirect plaques were produced by the addition of 1:20 sheep anti-MOPC 21 antibody to the agarose overlay.

in the ratios of the different chains and occasionally with the total disappearance of one or other of the chains. Such events are best visualised on IEF analysis of the separated chains (for example, Fig. 1*h*, in which the heavy chain of P3 is no longer observed). The important point that no new chains are detected by IEF complements a previous study[4] of a rat–mouse hybrid line in which scrambling of *V* and *C* regions from the light chains of rat and mouse was not observed. In this study, both light chains have identical $C_\kappa$ regions and therefore scrambled $V_L$–$C_L$ molecules would be undetected. On the other hand, the heavy chains are of different subclasses and we expect scrambled $V_H$–$C_H$ to be detectable by IEF. They were not observed in the clones studied and if they occur must do so at a lower frequency. We conclude that in syngeneic cell hybrids (as well as in interspecies cell hybrids) *V*–*C* integration is not the result of cytoplasmic events. Integration as a result of DNA translocation or rearrangement during transcription is also suggested by the presence of integrated mRNA molecules[11] and by the existence of defective heavy chains in which a deletion of *V* and *C* sections seems to take place in already committed cells[12].

The cell line P3-X63Ag8 described above dies when exposed to HAT medium. Spleen cells from an immunised mouse also die in growth medium. When both cells are fused by Sendai virus and the resulting mixture is grown in HAT medium, surviving clones can be observed to grow and become established after a few weeks. We have used SRBC as immunogen, which enabled us, after culturing the fused lines, to determine the presence of specific antibody-producing cells by a plaque assay technique[13] (Fig. 2*a*). The hybrid cells were cloned in soft agar[14] and clones producing antibody were easily detected by an overlay of SRBC and complement (Fig. 2*b*). Individual clones were isolated and shown to retain their phenotype as almost all the clones of the derived purified line are capable of lysing SRBC (Fig. 2*c*). The clones were visible to the naked eye (for example, Fig. 2*d*). Both direct and indirect plaque assays[13] have been used to detect specific clones and representative clones of both types have been characterised and studied.

The derived lines (Sp hybrids) are hybrid cell lines for the following reasons. They grow in selective medium. Their karyotype after 4 months in culture (Table 1) is a little smaller than the sum of the two parental lines but more than twice the chromosome number of normal BALB/c cells, indicating that the lines are not the result of fusion between spleen cells. In addition the lines contain a metacentric chromosome also present in the parental P3-X67Ag8. Finally, the secreted immunoglobulins contain MOPC 21 protein in addition to new, unknown components. The latter presumably represent the chains derived from the specific anti-SRBC antibody. Figure 3*A* shows the IEF pattern of the material secreted by two such Sp hybrid clones. The IEF bands derived from the parental P3 line are visible in the pattern of the hybrid cells, although obscured by the presence of a number of new bands. The pattern is very complex, but the complexity of hybrids of this type is likely to result from the random recombination of chains (see above, Fig. 1). Indeed, IEF patterns of the reduced material secreted by the spleen–P3 hybrid clones gave a simpler pattern of Ig chains. The heavy and light chains of the P3 parental line became prominent, and new bands were apparent.

The hybrid Sp-1 gave direct plaques and this suggested that it produces an IgM antibody. This is confirmed in Fig. 4 which shows the inhibition of SRBC lysis by a specific anti-IgM

Table 1 Number of chromosomes in parental and hybrid cell lines

| Cell line | Number of chromosomes per cell | Mean |
|---|---|---|
| P3-X67Ag8 | 66,65,65,65,65 | 65 |
| P1Bul | Ref. 4 | 55 |
| Mouse spleen cells | – | 40 |
| Hy-B (P1-P3) | 112,110,104,104,102 | 106 |
| Sp-1/7-2 | 93,90,89,89,87 | 90 |
| Sp-2/3-3 | 97,98,96,96,94,88 | 95 |

*Nature Vol 256 August 7 1975* 497

antibody. IFF techniques usually do not reveal 19S IgM molecules. IgM is therefore unlikely to be present in the unreduced sample *a* (Fig. 3*B*) but μ chains should contribute to the pattern obtained after reduction (sample *a*, Fig. 3*A*).

The above results show that cell fusion techniques are a powerful tool to produce specific antibody directed against a predetermined antigen. It further shows that it is possible to isolate hybrid lines producing different antibodies directed against the same antigen and carrying different effector functions (direct and indirect plaque).

The uncloned population of P3 spleen hybrid cells seems quite heterogeneous. Using suitable detection procedures it should be possible to isolate tissue culture cell lines making different classes of antibody. To facilitate our studies we have used a myeloma parental line which itself produced an Ig. Variants in which one of the parental chains is no longer expressed seem fairly common in the case of P1–P3 hybrids (Fig. 1*h*). Therefore selection of lines in which only the specific antibody chains are expressed seems reasonably simple. Alternatively, non-producing variants of myeloma lines could be used for fusion.

We used SRBC as antigen. Three different fusion experiments were successful in producing a large number of antibody-producing cells. Three weeks after the initial fusion, 33/1,086

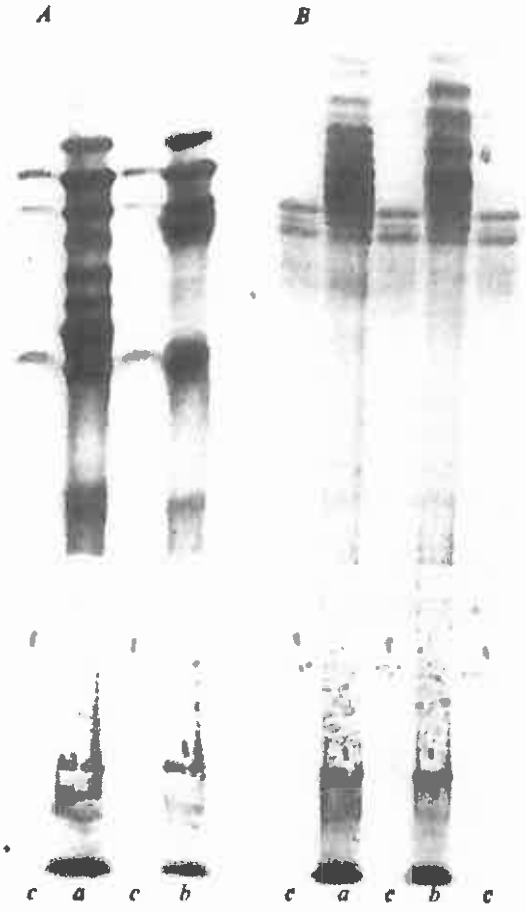



**Fig. 3** Autoradiograph of labelled components secreted by anti-SRBC specific hybrid lines. Fractionation before (*B*) and after (*A*) reduction was by IEF. pH gradient was 5.0 (bottom) to 9.0 (top) in the presence of 6 M urea. Other conditions as in Fig. 1. Supernatants from: *a*, hybrid clone Sp-1/7-2; *b*, hybrid clone Sp-2/3-3; *c*, myeloma line P3-X67Ag8.

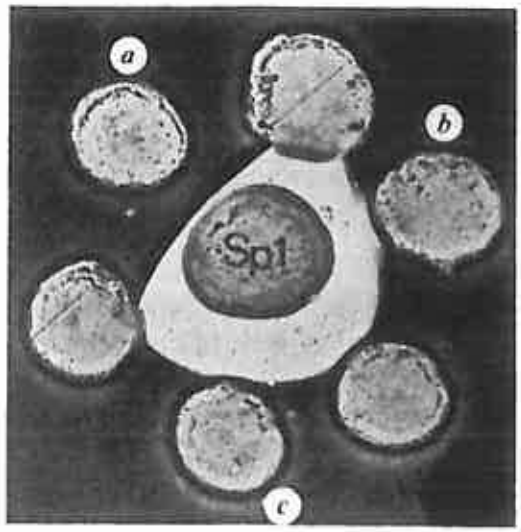



**Fig. 4** Inhibition of haemolysis by antibody secreted by hybrid clone Sp-1/7-2. The reaction was in a 9-cm Petri dish with a layer of 5 ml 0.6% agarose in phosphate-buffered saline containing 1/80 (v/v) SRBC. Centre well contains 2.5 μl 20 times concentrated culture medium of clone Sp-1/7-2 and 2.5 μl mouse serum. *a*, Sheep specific anti-mouse macroglobulin (MOPC 104E, Dr Feinstein). *b*, sheep anti-MOPC 21 (P3) IgG1 absorbed with Adj PC-5. *c*, sheep anti-Adj PC-5 (IgG2a) absorbed with MOPC 21. After overnight incubation at room temperature the plate was developed with guinea pig serum diluted 1:10 in Dulbecco's medium without serum.

clones (3%) were positive by the direct plaque assay. The cloning efficiency in the experiment was 50%. In another experiment, however, the proportion of positive clones was considerably lower (about 0.2%). In a third experiment the hybrid population was studied by limiting dilution analysis. From 157 independent hybrids, as many as 15 had anti-SRBC activity. The proportion of positive over negative clones is remarkably high. It is possible that spleen cells which have been triggered during immunisation are particularly successful in giving rise to viable hybrids. It remains to be seen whether similar results can be obtained using other antigenes.

The cells used in this study are all of BALB/c origin and the hybrid clones can be injected into BALB/c mice to produce solid tumours and serum having anti-SRBC activity. It is possible to hybridise antibody-producing cells from different origins[4,5]. Such cells can be grown *in vitro* in massive cultures to provide specific antibody. Such cultures could be valuable for medical and industrial use.

G. KÖHLER
C. MILSTEIN

*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK*

Received May 14; accepted June 26, 1975.

1 Potter, M., *Physiol. Rev.*, 52, 631-719 (1972).
2 Horibata, K., and Harris, A. W., *Exp. Cell Res.*, 60, 61-70 (1970).
3 Milstein, C., and Munro, A. J., in *Defence and Recognition* (edit. by Porter, R. R.), 199-228 (MTP Int. Rev. Sci., Butterworth, London, 1973).
4 Cotton, R. G. H., and Milstein, C., *Nature*, 244, 42-43 (1973).
5 Schwaber, J., and Cohen, E. P., *Proc. natn. Acad. Sci. U.S.A.*, 71, 2203-2207 (1974).
6 Littlefield, J. W., *Science*, 145, 709 (1964).
7 Svasti, J., and Milstein, C., *Biochem. J.*, 128, 427-444 (1972).
8 Milstein, C., Adetugbo, K., Cowan, N. J., and Secher, D. S., *Progress in Immunology*, II, 1 (edit. by Brent, L., and Holborow, J.), 157-168 (North-Holland, Amsterdam, 1974).
9 Harris, H., and Watkins, J. F., *Nature*, 205, 640-646 (1965).
10 Awdeh, A. L., Williamson, A. R., and Askonas, B. A., *Nature*, 219, 66-67 (1968).
11 Milstein, C., Brownlee, G. G., Cartwright, E. M., Jarvis, J. M., and Proudfoot, N. J., *Nature*, 252, 354-359 (1974).
12 Frangione, B., and Milstein, C., *Nature*, 244, 597-599 (1969).
13 Jerne, N. K., and Nordin, A. A., *Science*, 140, 405 (1963).
14 Cotton, R. G. H., Secher, D. S., and Milstein, C., *Eur. J. Immun.*, 3, 135-140 (1973).

99

**Three dimensional structure of antibody-antigen complex**

The three dimensional structure of chicken egg-white lysozyme was determined precisely many years ago. Lysozyme was injected into mouse and the B lymphocytes from the mouse spleen were used to generate a number of hybridoma cell lines. It was essential that the binding affinity of the anti-lysozyme antibody with lysozyme should be fairly high so that they could be crystallized together. One of these was designated as D1.3 (Amit AG, Mariuzza RA, Phillips SEV, Poljak RJ (1986) Three dimentional structure of an antigen-antibody complex at 2.8 A resolution. Science, 233, 747-753).

The Fab fragment of the anti-lysozyme antibody D1.3 was crystallized with the entire lysozyme molecule. At this high resolution of 2.8 A, it was possible to identify the interaction between the amino acid residues of the antibody and those of the antigen. These authors summarized the result as follows using the Kabat numbering for the light and heavy chains of antibodies:

| Segments | Antibody residues | Lysozyme residues |
|---|---|---|
| CDRL1 | His 30 | Leu 129 |
| | Tyr 32 | Leu 23, Gln 121, Ile 124 |
| FRL2 | Tyr 49 | Gly 22 |
| CDRL2 | Tyr 50 | Asp 18, Asn 19, Leu 25 |
| CDRL3 | Phe 91 | Gln 121 |
| | Trp 92 | Gln 121, Ile 124 |
| | Ser 93 | Gln 121 |
| FRH1 | Thr 30 | Lys 116, Gly 117 |
| CDRH1 | Gly 31 | Lys 116, Gly 117 |
| | Tyr 32 | Lys 116, Gly 117 |
| CDRH2 | Trp 52 | Gly 117, Thr 118, Asp 119 |
| | Gly 53 | Gly 117 |
| | Asp 54 | Gly 117 |
| CDRH3 | Arg 96 | Arg 21, Gly 22, Tyr 23 |

| | |
|---|---|
| Asp 97 | Gly 22, Tyr 23, Ser 24, asn 27 |
| Tyr 98 | Thr 118, Asp 119, Val 120, Gln 121 |
| Arg 99 | Asn 19, Gly 22 |

As clearly indicated, all six CDRs are involved in binding the antigen, lysozyme. The two amino acid residues in the framework regions of the anti-lysozyme antibody are immediately adjacent to the CDRs.

In this case, the lysozyme molecule retains its original three dimensional configuration amd its binding to the anti-lysozyme antibody is simply in the lock and key form. Subsequent studies of many other antigen-antibody complexes further verified the importance of all six CDRs. In some cases, either the antigen or the antibody may change the structure slight to produce induced fit for better binding.

So the stage is set for the engineering design of therapeutic antibodies. One requires the amino acid sequences of these six CDRs and understands how they are folded. However, before undertaking this endeavor, one needs to understand how the human system can generate various antibodies at the genetic level. This can provide some insight as to what amino acid sequences are considered as foreign to humans and can thus be rejected when introduced into human patients.

# Three-Dimensional Structure of an Antigen-Antibody Complex at 2.8 Å Resolution

A. G. AMIT, R. A. MARIUZZA, S. E. V. PHILLIPS, R. J. POLJAK

**The 2.8 Å resolution three-dimensional structure of a complex between an antigen (lysozyme) and the Fab fragment from a monoclonal antibody against lysozyme has been determined and refined by x-ray crystallographic techniques. No conformational changes can be observed in the tertiary structure of lysozyme compared with that determined in native crystalline forms. The quaternary structure of Fab is that of an extended conformation. The antibody combining site is a rather flat surface with protuberances and depressions formed by its amino acid side chains. The antigen-antibody interface is tightly packed, with 16 lysozyme and 17 antibody residues making close contacts. The antigen contacting residues belong to two stretches of the lysozyme polypeptide chain: residues 18 to 27 and 116 to 129. All the complementarity-determining regions and two residues outside hypervariable positions of the antibody make contact with the antigen. Most of these contacts (10 residues out of 17) are made by the heavy chain, and in particular by its third complementarity-determining region. Antigen variability and antibody specificity and affinity are discussed on the basis of the determined structure.**

THE BINDING OF FOREIGN ANTIGENS TO COMPLEMENTARY structures on the surface of B and T lymphocytes represents the initial step in the sequence of events leading to activation of the immune system. The receptor molecule on the surface of B lymphocytes responsible for antigen recognition is membrane immunoglobulin. A mature B cell produces and inserts into its plasma membrane only limited amounts of a single kind of immunoglobulin. Contact with antigen results in the expansion of B cell clones specific for that antigen and in their differentiation into plasma cells capable of producing and secreting large amounts of antibody of the same specificity (monoclonal antibody).

Antibody molecules of the immunoglobulin G (IgG) class, the most abundant in normal serum, are composed of two identical light (L) and two identical heavy (H) polypeptide chains. The amino terminal regions of the H and L chains, termed $V_H$ and $V_L$, are each about 110 amino acids long and have variable (and homologous) amino acid sequences. The constant (C) half of the L chain, $C_L$, and the constant regions $C_H1$, $C_H2$, and $C_H3$ of the H chain, each about 100 amino acids long, have homologous sequences that belong to one of a few classes (κ and λ for L chains; μ, δ, γ, ε, and α for H chains). The $V_H$ and $V_L$ regions each contain three hypervariable or complementarity-determining regions (CDR1, CDR2, and CDR3) responsible for antigen recognition. These are flanked by less variable (FR1, FR2, FR3, and FR4) "framework" regions (*1*).

Present understanding of the three-dimensional structure of antibody combining sites is based on x-ray diffraction studies of myeloma immunoglobulins as reviewed (*2*). These have shown that the conformation of combining sites is determined by the amino acid sequences, unique to each different antibody, of the CDR's. The structures of two complexes of antigen-binding fragments (Fab) of myeloma immunoglobulins with small ligands have also been determined (*3, 4*). Although these studies resulted in useful models for ligand-antibody interactions, they are insufficient to establish unequivocally the precise size and shape of antibody combining sites, the nature and extent of antigen-antibody interactions, and the occurrence of possible conformational changes (if any) in the antibody after antigen binding. In addition the precise structure of antigenic determinants on protein molecules remains to be determined (*5*). Equally important are questions concerning the nature of possible conformational changes in the complexed antigen and the effect of single amino acid substitutions on antigenic specificity and antigen recognition by the antibody.

We have recently determined the three-dimensional structure of an antigen-antibody complex, one between lysozyme and the Fab fragment of a monoclonal antibody to hen egg white lysozyme, at 6 Å resolution (*6*). We have since extended the resolution of the x-ray structure determination to 2.8 Å, and now present a complete description of antigen-antibody interactions in the complex.

**Structure determination.** The production of hybrid cell lines secreting murine monoclonal antibody to hen egg white lysozyme, and the purification, crystallization (*7*), and 6 Å resolution crystal structure determination (*6*) of the complex between Fab D1.3 and lysozyme have been described. Crystals grown from solutions containing 15 to 20 percent polyethylene glycol 8000 at *p*H 6.0 are monoclinic, space group $P2_1$, with $a$ = 55.6, $b$ = 143.4, $c$ = 49.1 Å, β = 120.5°, and one molecule of complex per asymmetric unit.

Three heavy atom isomorphous derivatives were prepared with $(NH_4)_2PtCl_4$, $K_3F_5UO_2$, and *p*-hydroxymercuribenzenesulfonate. X-ray intensities were measured to 2.8 Å resolution with the use of a four-circle automatic diffractometer. Heavy atom sites were refined in alternate cycles of phasing and refinement (*8*); isomorphous phases, including anomalous scattering contributions (*9*), were calculated. The mean figure of merit (*10*) to 2.8 Å resolution was 0.47 for 15592 reflections. The electron density map calculated from these data was not readily interpretable, presumably because of lack of isomorphism of the heavy atom derivatives affecting phase determination at high resolution. The phases were further refined by a density modification technique (*11*) with a molecular envelope traced from the Fab-lysozyme model determined at 6 Å resolution (*6*). The resulting phases depend only on the observed data and the overall shape and position of the complex, but are independent of the detailed conformation of the previous model (*6*). The resulting

A. G. Amit, R. A. Mariuzza, and R. J. Poljak are in the Département d'Immunologie, Institut Pasteur, 75724 Paris Cedex 15, France. S. E. V. Phillips is in the Astbury Department of Biophysics, University of Leeds, Leeds, United Kingdom.

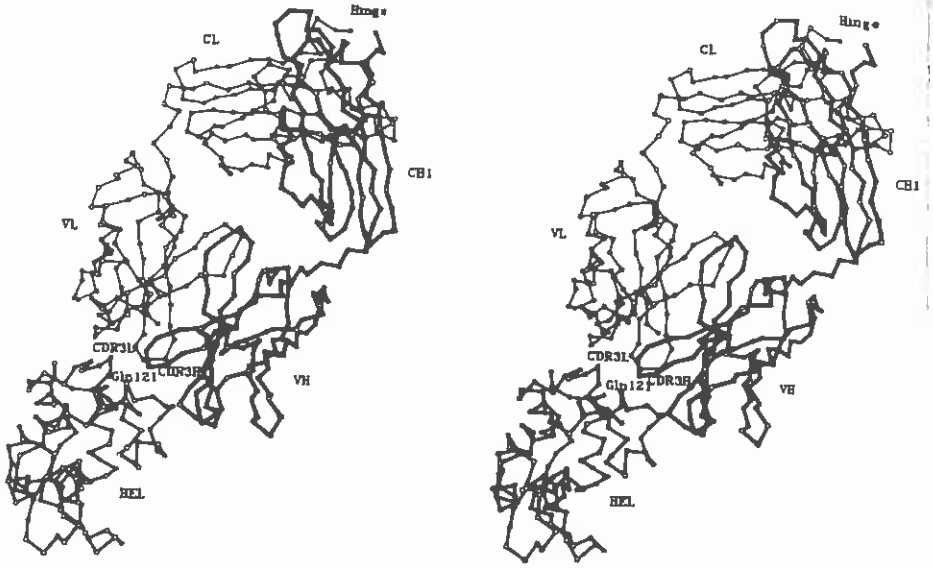



Fig. 1. Stereo diagram of the Cα skeleton of the complex. Fab is shown (upper right) with the heavy and light chains with thick and thin bonds, respectively. The lysozyme active site is the cleft containing the label HEL. Antibody-antigen interactions are most numerous between lysozyme and the heavy chain CDR loops.

electron density map was much improved, and an atomic model was fitted to it on an Evans and Sutherland PS300 interactive graphics system with the use of the program FRODO (*12*). The amino acid sequence of Fab D1.3 was derived from the corresponding light and heavy chain complementary DNA (cDNA) sequences (*13*). Of the 562 amino acid residues in the complex, 24 of those in the constant regions could not be located in the initial map. The atomic coordinates were submitted to alternate cycles of restrained crystallographic least-squares refinement (*14*) and model building. The model was checked in the later stages of refinement by sequentially omitting segments of the polypeptide chain (up to 20 percent of the total) and rebuilding them in maps phased from the remainder of the structure in combination with isomorphous replacement data (*15*). All residues have now been located, and the current crystallographic $R$ factor is 0.28 for all data in the 20 to 2.8 Å resolution range. ($R = \Sigma \mid |Fo| - |Fc| \mid / \Sigma |Fo|$, where $Fo$, $Fc$ are the observed and calculated structure factors of x-ray reflections.) No attempt was made to locate solvent molecules. Two isotropic temperature factors were used for each residue, one for the main chain atoms, and another for the side chain atoms. Stereochemical restraints were adjusted to give a standard deviation in C–G bonds of ±0.03 Å. No restraints were applied between residues across the antibody-antigen interface. Atomic coordinates will be deposited at Brookhaven Protein Data Bank after higher resolution and crystallographic refinement.

**Conformation of the complexed antigen and of the Fab.** The overall structure of the complex at 2.8 Å resolution (Fig. 1) confirms the results of the 6 Å resolution study (*6*). The assignment of the H and L polypeptide chains of Fab is unchanged. The closely packed β sheets are seen in Fab as are the helical and β-sheet structures surrounding the active site in lysozyme. The Fab appears in an almost fully extended conformation, with a definite separation between the variable (V) and constant (C) domains. With the exception of this difference in quaternary structure, Fab D1.3 compares closely to other known Fab's (*4, 16*), except in the CDR loops. Predicted structures for D1.3 (*17*) based on other Fab's also agree well with the determined structure in the framework β-sheet regions and in some, but not all, of the CDR loops. The relative disposition of the variable subunits of the H chain ($V_H$) and of the L chain ($V_L$), is unaltered, indicating no change in quaternary structure in the V domain resulting from antigen binding. Since the crystal structure of the unliganded Fab D1.3 has not been determined, detailed changes in antibody conformation remain to be verified. However, the similarity with other Fab structures suggests that possible conformational changes would be small. This observation is in agreement with that made by nuclear magnetic resonance (NMR) on the unliganded and hapten-liganded (dinitrophenol) mouse myeloma protein MOPC315 (*18*).

A least-squares fit of Cα atoms of lysozyme in the complex and native lysozyme refined at 1.6 Å in its tetragonal crystal form (*19*) gives a root-mean-square (rms) deviation of 0.64 Å between the two (see Fig. 2). Since the error in atomic positions in the complex can be estimated (*20*) to be approximately 0.6 Å, the difference is not significant. Furthermore, the largest changes (up to 1.6 Å) occur in regions remote from antibody contacts. Similar comparisons of native tetragonal lysozyme with other crystal forms gave rms deviations of 0.88 Å with triclinic lysozyme refined from x-ray and neutron diffraction data (*21*) and 0.46 Å for orthorhombic lysozyme determined at physiological temperature (*22*). Some differences in side chain conformation are observed between tetragonal and complexed lysozyme, but close examination with computer graphics revealed these to be similar to differences observed between different crystal structures of native lysozyme. Thus, complex formation with antibody D1.3 produces no more distortion of the structure of lysozyme than does crystallization.

**The antigen-antibody interface.** The interface between antigen and antibody extends over a large area with maximum dimensions of about 30 by 20 Å (Figs. 3 and 4). The antibody combining site appears as an irregular, rather flat surface with protuberances and depressions formed by the amino acid side chains of the CDR's of $V_H$ and $V_L$. In addition, there is a small cleft between the third CDR's of $V_H$ and $V_L$, corresponding to the binding site characterized in hapten-antibody complexes (*3, 4*). The cleft accepts the side chain Gln 121 of lysozyme although this is not the center of the antigen-antibody interface (Fig. 3).

The lysozyme antigenic determinants recognized by D1.3 are made up of two stretches of polypeptide chain, comprising residues 18 to 27 and 116 to 129, distant in the amino acid sequence but adjacent on the protein surface. All six CDR's interact with the antigen and in all, 16 antigen residues make close contacts with 17 antibody residues (Tables 1 and 2). Two antibody contacting residues, $V_L$ Tyr 49 and $V_H$ Thr 30, are just outside segments commonly defined as CDR's [sequence numbers are as in Kabat *et al.* (*1*) except for $V_H$ CDR3; see Tables 2 and 3]. $V_H$ Thr 30 is a constant or nearly constant residue in mouse H chain subgroups I and II, as is $V_L$ Tyr 49 in mouse kappa chains. While the interaction of $V_L$ Tyr 49 with antigen is relatively weak (one van der Waals

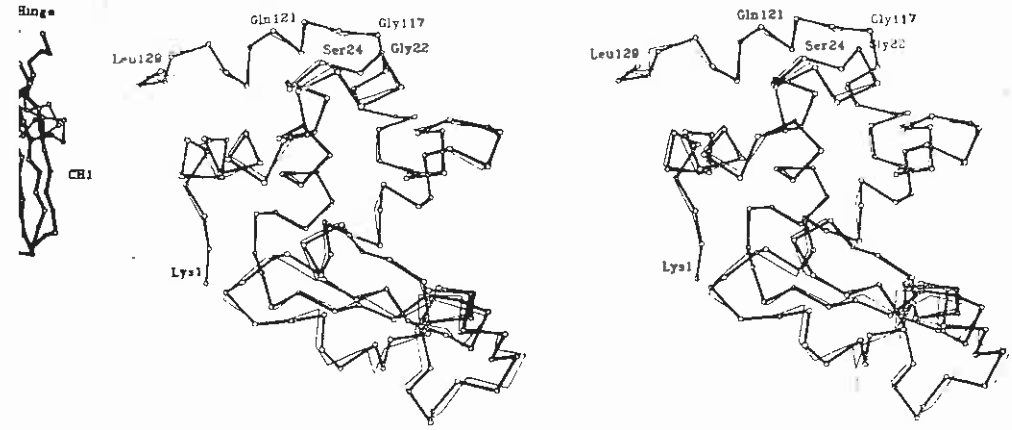



Fig. 2. The Cα skeleton of lysozyme in the complex (thick trace) superimposed by least squares on that of native lysosyme in the tetragonal crystal form (thin trace). The interface to Fab is at the top, and no significant conformation change is apparent in this region. Greater differences, although still not significant at this resolution, occur at the bottom of the molecule.

contact between its aromatic side chain and Cα of Gly 22 of lysozyme), there is a strong hydrogen bond between the hydroxyl group of $V_H$ Thr 30 and the carbonyl oxygen of Lys 116 of lysozyme. This specific interaction involving an invariant antibody residue demonstrates that the functional distinction between "framework" (FR) and CDR residues, although largely maintained, is not absolute. The interacting surfaces are complementary, with protruding side chains of one lying in depressions of the other (Fig. 3) in common with other known protein-protein interactions (*23*). There are many van der Waals interactions interspersed with hydrogen bonds. This is most striking for the side chain of Gln 121, which penetrates deeply into the Fab, surrounded by three aromatic side chains, $V_L$ Tyr 32 and Trp 92 and $V_H$ Tyr 101 (Figs. 3, 4, and 5). Its amide nitrogen forms a strong, buried hydrogen bond to the main chain carbonyl oxygen of $V_L$ Phe 91 (Fig. 5 and Table 3). The adjacent $V_H$ Tyr 101 extends to the surface of lysozyme, its terminal hydroxyl group forming hydrogen bonds to the main chain nitrogens of Val 120 and Gln 121, and to Oδ1 of Asp 119. Many hydrogen bonds occur between the side chains of the antigen and the main polypeptide chain of the antibody, and vice versa (Table 3). Hydrogen bonds between main polypeptide chain atoms, similar to those in β-sheet structures, occur between Lys 116 of lysozyme and $V_H$ Gly 31, and between Gly 117 and $V_H$ Gly 53, where the lack of side chains allows close approach. There are many side chain–side chain close interactions forming, together with the ones described above, a tightly packed interface which mostly excludes solvent.

Although the antigen-antibody interface involves all six CDR's of the Fab, there are more interactions with $V_H$ than with $V_L$ CDR's, and with $V_H$ CDR3 in particular (Tables 1 to 3). The geometrical

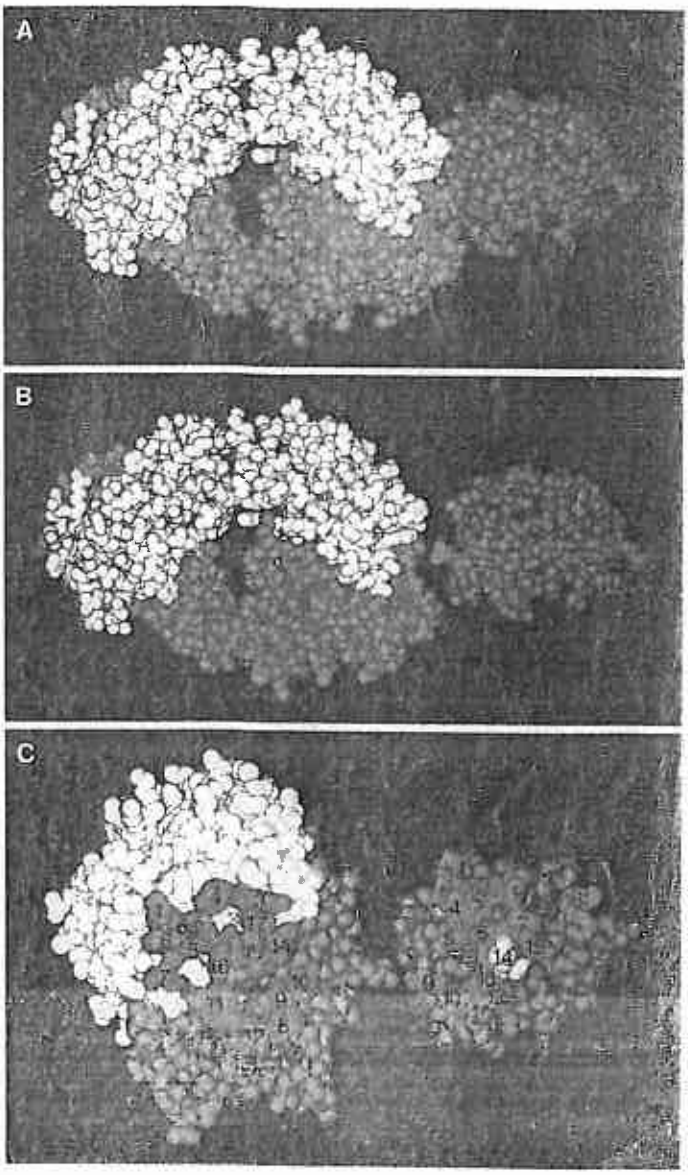
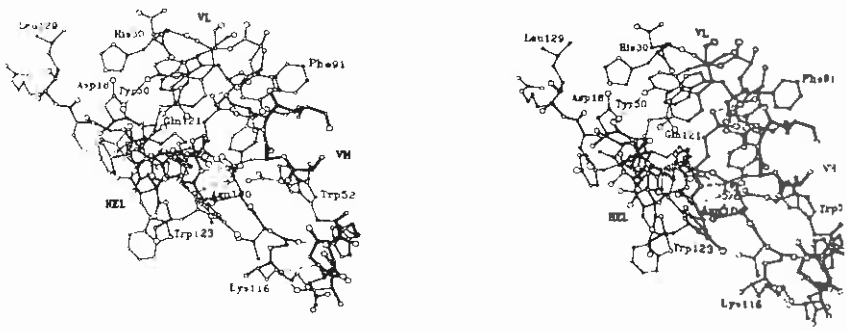



Fig. 3. Space filling representation of Fab D1.3 and lysozyme. (A) Antigen-antibody complex structure as determined in this work. The antibody H chain is shown in blue, the L chain in yellow, lysozyme in green, and Gln 121 in red. (B) The Fab and lysozyme models have been pulled apart to indicate protuberances and depressions of each fit in complementary surface features of the other. Compare with (A) above. At the top of the interface, protruding $V_L$ residues His 30 and Tyr 32 fit into a depression in lysozyme, between residues Ile 124 and Leu 129 (see also Table 1). Below the Gln 121, in red, a protuberance of lysozyme consisting of residues around Thr 118 fits into a surface depression formed by $V_H$ residues of CDR1 and CDR2 ($V_H$ Trp 52 can be seen at the bottom of this depression). (C) End-on views of the antibody combining site (left) and the antigenic markers of lysozyme recognized by antibody D1.3, formed from (B) above, by rotating each of the molecules approximately 90° about a vertical axis. Contacting residues on the antigen and antibody are shown in red, except for Gln 121 shown in light purple. L chain residues that contact the antigen are labeled 1 (His 30), 2 (Tyr 32), 3 (Tyr 49), 4 (Tyr 50), 5 (Phe 91), 6 (Trp 92), and 7 (Ser 93). H chain residues that contact the antigen are labeled 8 (Thr 30), 9 (Gly 31), 10 (Tyr 32), 11 (Trp 52), 12 (Gly 53), 13 (Asp 54), 14 (Arg 99), 15 (Asp 100), 16 (Tyr 101), and 17 (Arg 102). see Table 1. Lysozyme residues that contact the antibody are labeled 1 (Asp 18), 2 (Asn 19), 3 (Arg 21), 4 (Gly 22), 5 (Tyr 23), 6 (Ser 24), 7 (Leu 25), 8 (Asn 27), 9 (Lys 116), 10 (Gly 117), 11 (Thr 118), 12 (Asp 119), 13 (Val 120), 14 (Gln 121), 15 (Ile 124), and 16 (Leu 129). Gln 121 fits into the antibody surface pocket surrounded by $V_L$ and $V_H$ residues 2, 5, 6, 7, and 16 (Table 1).

Fig. 4. Stereo diagram of the antibody-antigen interface in a similar orientation to Fig. 1. All atoms are shown for those residues involved in the interaction. Heavy and light main chains are indicated by thick and thin bonds, respectively, and hydrogen bonds by dotted lines. Lysozyme residues broadly lie below the diagonal from top left to lower right of the diagram.




center of the surface lies near $V_H$ CDR3, and is occupied by the side chain of $V_H$ Asp 100, which forms H bonds to the side chains of Ser 24 and Asn 27 of lysozyme. Of the antibody hypervariable regions, $V_L$ CDR2 contributes the least to antigen binding. A large number of antibody side chains in the interface (9 out of 15 if we exclude Gly residues) are aromatic, thus presenting large areas of hydrophobic surface to the antigen; in addition, some of them such as $V_L$ Tyr 50 and $V_H$ Tyr 101 participate in hydrogen bonding with the antigen via their polar atoms. In all, 748 Å$^2$ or about 11 percent of the solvent-accessible surface (*24*) of lysozyme is buried on complex formation, together with 690 Å$^2$ for the antibody.

**Antigen variability and antibody specificity.** The fine specificity of monoclonal antibody D1.3 for other avian lysozymes shows its ability to distinguish a single amino acid change in the antigen, at position 121. Fab D1.3 binds hen egg white lysozyme with an equilibrium affinity constant of $4.5 \times 10^7 M^{-1}$ (*25*). Bobwhite quail lysozyme, with four amino acid sequence differences (*26*) from hen lysozyme but none in the interface with Fab D1.3, binds with similar affinity (*25*). The binding of antibody D1.3 to the lysozymes of partridge [three amino acid differences (*26*)], California quail [four amino acid differences (*26*)], Japanese quail [six amino acid differences (*27*)], turkey [seven amino acid differences (*28*)], and pheasant and guinea fowl [ten amino acid differences each (*29*)] is undetectable ($K_A < 1 \times 10^5 M^{-1}$) with the enzyme-linked immunoabsorption assay used in our laboratory. These lysozymes differ from hen lysozyme in the amino acid residue at position 121, which makes close contacts with the antibody. Except for Japanese quail and pheasant lysozymes, all have Gln replaced by His.

Table 1. Antibody residues involved in contact with lysozyme. Sequence positions are numbered as in Kabat *et al.* (*1*) except for $V_H$ CDR3, where the numbers of Kabat *et al.* (*1*) are given in parentheses.

| Antibody residues | | Lysozyme residues in contact |
|---|---|---|
| Light chain | | |
| CDR1 | His 30 | Leu 129 |
| | Tyr 32 | Leu 25, Gln 121, Ile 124 |
| FR2 | Tyr 49 | Gly 22 |
| CDR2 | Tyr 50 | Asp 18, Asn 19, Leu 25 |
| CDR3 | Phe 91 | Gln 121 |
| | Trp 92 | Gln 121, Ile 124 |
| | Ser 93 | Gln 121 |
| Heavy chain | | |
| FR1 | Thr 30 | Lys 116, Gly 117 |
| CDR1 | Gly 31 | Lys 116, Gly 117 |
| | Tyr 32 | Lys 116, Gly 117 |
| CDR2 | Trp 52 | Gly 117, Thr 118, Asp 119 |
| | Gly 53 | Gly 117 |
| | Asp 54 | Gly 117 |
| CDR3 | Arg 99 (96) | Arg 21, Gly 22, Tyr 23 |
| | Asp 100 (97) | Gly 22, Tyr 23, Ser 24, Asn 27 |
| | Tyr 101 (98) | Thr 118, Asp 119, Val 120, Gln 121 |
| | Arg 102 (99) | Asn 19, Gly 22 |

A computer graphics analysis indicates that a His residue could be placed in the interface, in the space occupied by Gln 121, with small displacements of the contacting antibody side chains, maintaining the H bonds made by Gln 121. Conformational energy calculations (*30*) confirm this possibility, the total energy being little changed on substitution of His for Gln 121. The buried hydrogen bond is maintained with good geometry, and only very small shifts of neighboring groups are necessary to accommodate the mutation. This seems to rule out steric hindrance in explaining the absence of complex formation when His occurs at position 121. Other possible explanations for the effect of this amino acid substitution include the following. (i) His 121 could be charged, and consequently unstable in the hydrophobic pocket occupied by Gln 121; (ii) its side chain may have a different orientation from that of Gln, forming, for example, a salt bridge with Asp 119; and (iii) substitution of His for Gln at position 121 may induce a local change of conformation in the polypeptide backbone making the antigenic determinant unrecognizable by the antibody. Not enough information is available to decide on the relative importance of these factors. Nevertheless, the

Table 2. Lysozyme residues in contact with antibody.

| Lysozyme residues | Antibody residues in contact (No.) | Lysozyme residues | Antibody residues in contact (No.) |
|---|---|---|---|
| Asp 18 | 1 L chain | Lys 116 | 3 H |
| Asn 19 | 2 H, L | Gly 117 | 6 H |
| Arg 21 | 1 H | Thr 118 | 2 H |
| Gly 22 | 4 H(3), L | Asp 119 | 2 H |
| Tyr 23 | 2 H | Val 120 | 1 H |
| Ser 24 | 1 H | Gln 121 | 5 H(1), L(4) |
| Leu 25 | 1 L | Ile 124 | 2 L |
| Asn 27 | 1 H | Leu 129 | 1 L |

Table 3. Hydrogen bonded interactions between antibody and lysozyme. Sequence positions are numbered as in Kabat *et al.* (*1*) except for $V_H$ CDR3 where the numbers of Kabat *et al.* (*1*) are given in parentheses.

| | Antibody residue | | Lysozyme residue | |
|---|---|---|---|---|
| Light chain | | | | |
| | Nε2 | His 30 | O | Leu 129 |
| | Oη | Tyr 50 | Oδ1 | Asp 18 |
| | O | Phe 91 | Nε2 | Gln 121* |
| Heavy chain | | | | |
| | Oγ1 | Thr 30 | O | Lys 116* |
| | N | Gly 31 | O | Lys 116 |
| | N | Gly 53 | O | Gly 117* |
| | Nη1 | Arg 99 (96) | O | Gly 22 |
| | Oδ1 | Asp 100 (97) | Nδ2 | Asn 27 |
| | Oδ2 | Asp 100 (97) | Oγ | Ser 24 * |
| | Oη | Tyr 101 (98) | N | Val 120 |
| | Oη | Tyr 101 (98) | N | Gln 121 |
| | Oη | Tyr 101 (98) | Oδ1 | Asp 119 |

*Denotes the closest interactions (distances ≤ 2.5 Å)

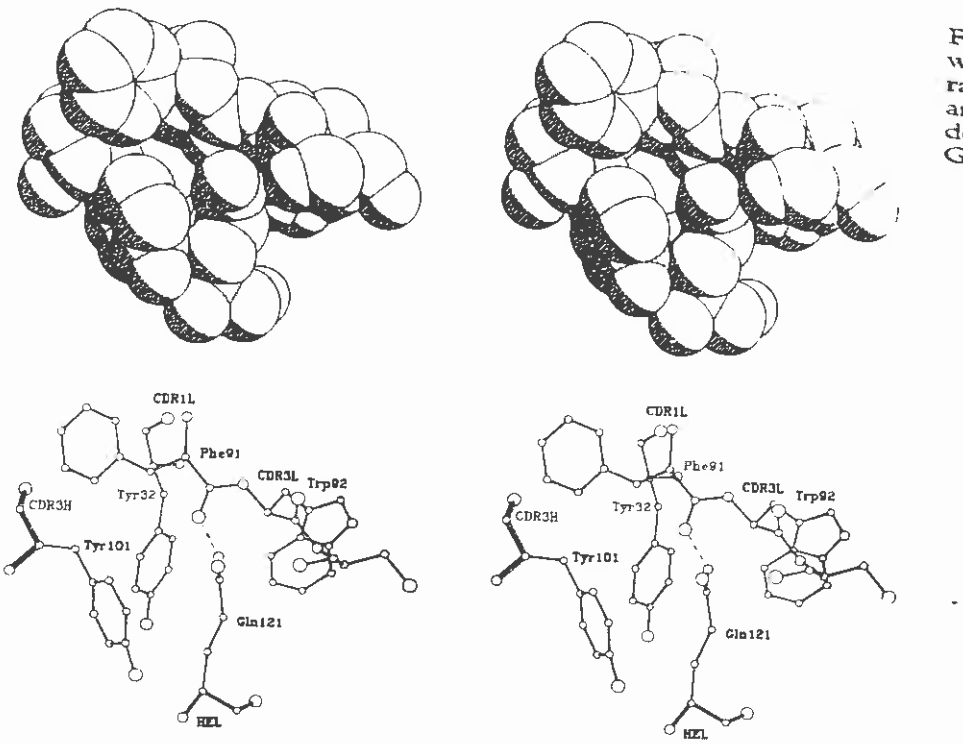



Fig. 5. Stereo view of the environment of Gln 121 with (above) atoms drawn with their van der Waals radii, showing the close packing of the three antibody aromatic rings around the antigen side chain. The dotted line indicates the hydrogen bond from Nε2 of Gln to the main chain carbonyl oxygen of $V_L$ Phe 91.

fact that the His residue is not induced to fit into the interface position occupied by Gln 121 is in agreement with a "lock and key" model (see below) of complex formation between conformationally stable antigen and antibody structures.

Japanese quail lysozyme has an Asn at position 121 and the additional differences Asn 19 → Lys and Arg 21 → Gln in the antigen-antibody interface. Asn 121 would be unable to form hydrogen bonds as strong as those of Gln 121. In addition, replacement of Asn 19 by Lys causes the loss of weak interactions between Oδ1 of Asn 19 and N of $V_H$ Arg 102. The positively charged Lys 19 side chain would be repelled by $V_H$ Arg 102 and would probably remain outside the interface, further reducing packing efficiency. Only main chain atoms of Arg 21 make contact with Fab and the side chain is external; therefore changes at this position in the antigen surface are probably not detrimental to complex formation.

The equilibrium affinity constant of Fab D1.3 binding of hen lysozyme is $4.5 \times 10^7 M^{-1}$. Other monoclonal antibodies to lysozyme that we (*25*) and others (*31*) have obtained and characterized show similar affinity constants for the homologous lysozyme antigen. Moreover, the determined equilibrium constants of protein antigens with their specific antibodies range from $10^5 M^{-1}$ to $10^{10} M^{-1}$ (*32*). Thus, D1.3 is a typical antibody of the monoclonal response in BALB/c mice and one of an about average affinity constant in immune responses to protein antigens in general.

Comparison of evolutionarily related proteins has been used to identify antigenic sites in proteins such as lysozyme (*5, 25, 33*). The detection of antigenic determinants by these fine specificity studies is biased toward the recognition of evolutionarily variable residues, such as Gln 121 by antibody D1.3. As our results show, such analyses are limited in defining antigenic determinants and, in particular, in defining the area of the antigen-antibody interaction, or even its center. Antigenic determinants have also been localized by measuring the reactivity of natural or synthetic peptides corresponding to different parts of the sequence of the protein with antibodies to the protein. This method cannot identify, or it can identify only partially, noncontinuous determinants such as those recognized by D1.3. Furthermore, given the large size of an antibody combining site (about 690 Å$^2$ of accessible surface area in our study) plus the fact that only a small portion of the surface of a globular protein is made up of linear arrays of residues, the probability that all of the antigen residues contacted by a given antibody come from the same continuous segment of polypeptide chain is very low (*34*). Thus, most antibodies to native protein molecules probably recognize noncontinuous determinants.

In the three-dimensional structure of the antigen-Fab complex presented in this article, the axes of the V and C domains of Fab make an angle close to 180°. This gives an extended conformation, with the V and C domains further apart than they would be if that angle were smaller. This observation is not in agreement with hypotheses (*35*) in which the liganded antibody molecule is postulated to assume a more rigid conformation with an "elbow bending" angle (between the axes of the V and C domains) close to 120°. In fact, the angle observed in the lysozyme–Fab D1.3 complex corresponds to that postulated to occur in unliganded Fab's. Thus, the allosteric model of antibodies (*35, 36*) in which antigen binding induces changes in quaternary structure resulting in closer contacts across V and C domains is not consistent with the structure of this complex.

No other change of conformation in the antibody or antigen can be established by the present analysis. The classical "lock and key" metaphor (*37*) is an adequate simplification to describe the interaction of lysozyme and antibody D1.3. It implies that somatic recombination of the germline gene repertoire provides all the complementary antibody templates necessary to bind all possible antigens. These combining site templates preexist and are basically unaltered in binding their specific antigens. The lysozyme–D1.3 binding is accomplished by van der Waals and hydrogen bonding interactions, and the number of contacts is of the order of that seen in other protein-protein systems, with similar implications for the specificity and the energetics of the interacting molecules (*23*). Although the fit of the antigen-antibody contacting surface is remarkably good, there are some imperfections in the form of holes. One of these holes, between $V_H$ residues 52 and 100 and lysozyme residues 24 and 118, is probably filled by a water molecule, hydrogen-bonded to the N of lysozyme Gly 117, as suggested by an

electron density peak at that location. Other holes appear not to be filled by water molecules. This observation might explain the occurrence of heteroclitic antibodies, that is, antibodies that have higher affinities for heterologous, closely related antigens, which would fill those holes and provide a tighter association. It might also explain an improved fit between antigen and antibody by somatic mutations in the antibody genes. For example, in antibody D1.3, further increases in affinity could be achieved by amino acid changes which would permit salt links and hydrogen bonding to lysozyme residues, such as Arg 21 and Thr 118, whose polar parts are unbonded in the interface.

Complete immunoglobulin V region genes are generated by somatic recombination of gene segments during B lymphocyte differentiation (*38*). The $V_H$ polypeptide chain is encoded by three gene segments, $V_H$, D (diversity), and $J_H$ (joining). A complete H chain V gene is generated by $V_H$–D and D–$J_H$ joinings (*39*). The core portion of $V_H$ CDR3, composed of 1 to 13 residues, is encoded by the D segment. As the $V_H$–D and D–$J_H$ boundaries fall with the $V_H$ CDR3, the combinatorial joining contributes to the high sequence variability of this CDR3, higher indeed than that of other CDR's. This variability suggests an essential role for $V_H$ CDR3 in antigen recognition (*40*), which we now confirm for antibody D1.3. Specifically, our three-dimensional model shows that (i) the physical center of the lysozyme-antibody interface is at $V_H$ CDR3; (ii) this CDR makes a proportionately greater contribution to the formation of the complementary antigen-binding surface of the combining site than the other CDR's, as measured by the number of van der Waals contacts and hydrogen bonds (6 out of 12; see Table 3) it makes with lysozyme. All four $V_H$ CDR3 residues (99 to 102) in contact with antigen are encoded by the D segment, thus illustrating its critical role in the generation of functionally different combining sites.

Our structural model permits us to evaluate the contribution of the imprecise joining of gene segments at the $V_L$–$J_L$, $V_H$–$D_H$, and $D_H$–$J_H$ junctions to the antigen contacts at the binding site. In mouse kappa chains, the somatic recombination process between $V_L$ and $J_L$ generates sequence diversity at position 96 (*41*). The interactions of $V_L$ CDR3 with antigen are illustrated in Figs. 3 to 6. Only atoms of residues 91 to 93 are within 4 Å of neighboring lysozyme atoms; the V–J junction at Arg 96 is relatively distant (>6 Å) from the interface. Thus neither residue 96 nor $J_\kappa$-encoded residues participate directly in contacts with antigen in antibody D1.3. The interactions of $V_H$ CDR3 with lysozyme along with the locations of the V–D and D–J boundaries are shown in Fig. 6. The side chains of $V_H$ Arg 99, Asp 100, and Tyr 101 form a number of hydrogen bonds with several lysozyme residues (Table 3), while

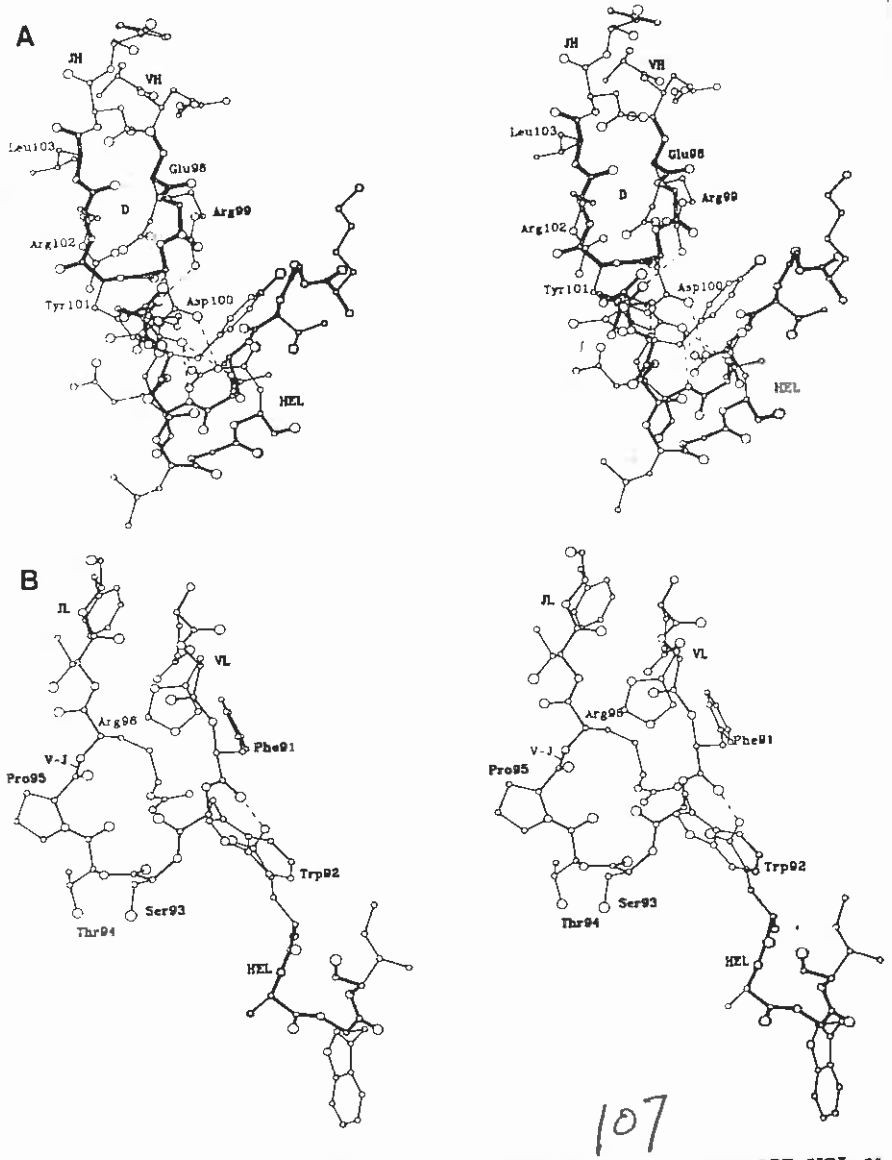



Fig. 6. (A) Stereo diagram of $V_H$ CDR3 with interacting residues from lysozyme shown below and to the right. Thick bonds indicate residues coded by the D gene segment through which all CDR3 contacts to the antigen are made. (B) Similar view to (A) for the $V_L$ CDR3. The V–J boundary is indicated, showing lack of involvement in the interaction of residues coded by the $J_L$ gene segment.

guanidino group atoms of Arg 102 are in van der Waals contact with lysozyme Asn 19 and Gly 22. These are all D-segment encoded residues; thus, neither $J_H$ residues nor those that might arise from imprecise joining at the D–J junction (*42*) contribute directly to antigen contacts made by antibody D1.3.

It is noteworthy that most lysozyme residues at the interface contact only one of the antibody chains: Asp 18, Leu 25, Ile 124, and Leu 129 contact only the L chain, while Arg 21, Tyr 23, Ser 24, Asn 27, Lys 116, Gly 117, Thr 118, Asp 119, and Val 120 contact only the H chain (Table 2). Thus, pairing of an L or H chain with different H or L chains could generate antibodies that will bind antigenic variants at positions contacting only one of the antibody chains.

The three-dimensional model obtained from the antigen-antibody complex presented here differs in important respects from that obtained in the study of the three-dimensional structure of hapten-antibody complexes (*3, 4*). In these latter studies no conformational change in the Fab's was detected after hapten binding. The lysozyme-Fab complex suggests a similar result except in the relative disposition of the V and C domains, as discussed above. However, in the hapten-Fab studies and, even more, in those of the less specific ligand binding to an L-chain dimer (*43*), a cavity or pocket surrounded by the CDR's of $V_H$ and $V_L$ was the most relevant structural feature of the antibody combining site. In the antigen-Fab model presented here the combining site appears as a large, irregular, and rather flat surface with protrusions and depressions formed by the amino acid side chains of the CDR's of the antibody. The complementarity of shape between antigen and antibody in the interface is striking: protruding side chains of one lie in depressions of the other. In addition, two antibody FR residues also closely contact the antigen.

Antigenically distinguishing regions or idiotypes (*44*) of monoclonal immunoglobulins have been widely used to characterize immune responses against different antigens (*45*). Combining site residues and idiotypic antigenic determinants are closely associated and may partially overlap (*46*), since interactions between idiotype (on antibody 1) and an antibody against the idiotype can often be competitively inhibited by the ligand for which antibody 1 is specific. As antigenic determinants, idiotypes should be exposed residues of antibody molecules. However, in the three-dimensional models of hapten-Fab complexes (*3, 4*) in which the hapten is buried in a cavity or groove at the combining site, the correlation between idiotypes and combining site residues is unclear. This is because (i) only a small area of the combining site is engaged in interactions with the ligand; (ii) the hapten contacts only those CDR residues that line the combining site cavity or groove in which it is buried, and not the outlying CDR residues which should be available for recognition by an antibody to an idiotype. Thus for example, to interpret the results of studies of hybridoma (*47*) and mouse myeloma antibodies (*48*) against $\alpha$-(1 → 6 dextrans), it has been postulated (*46*) that hapten binding alters the conformation of idiotype so that it can no longer react with an antibody to the idiotype. In the model that we present, the correlation between combining site residues and idiotypes becomes much clearer. First, given the area of antigen-antibody interactions and the location of the antibody interacting residues, antigen binding should sterically hinder the reaction between combining site idiotypes and antibodies against them. No conformational change is involved. Second, it can be seen that antibodies against idiotypes will recognize some of the most exposed residues of the $V_H$ and $V_L$ CDR's, whereas other, less exposed CDR residues would have no direct influence in promoting that binding. In fact, the contacts between idiotypes and antibodies against them could be approximated by the antibody-antigen contacts (respectively) between two complementary, irregular but rather flat, surfaces of the type described in this article.

## REFERENCES AND NOTES


1. E. A. Kabat, T. T. Wu, H. Bilofsky, M. Reid-Miller, H. Perry, *Sequences of Proteins of Immunological Interest* (U.S. Department of Health and Public Services, Washington, DC, 1983), vol. 1.
2. L. M. Amzel and R. J. Poljak, *Annu. Rev. Biochem.* 48, 961 (1979); D. R. Davies and H. A. Metzger, *Annu. Rev. Immunol.* 1, 87 (1983).
3. L. M. Amzel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 71, 1427 (1974).
4. D. M. Segal *et al.*, *ibid.*, p. 4298.
5. D. C. Benjamin *et al.*, *Annu. Rev. Immunol.* 2, 67 (1984).
6. A. G. Amit *et al.*, *Nature (London)* 313, 156 (1985).
7. R. A. Mariuzza *et al.*, *J. Mol. Biol.* 170, 1055 (1983).
8. R. E. Dickerson, J. C. Kendrew, B. E. Strandberg, *Acta Crystallogr.* 14, 1188 (1961); D. M. Blow and B. W. Matthews, *ibid.* A29, 56 (1973).
9. A. C. T. North, *ibid.* 18, 212 (1959).
10. D. M. Blow and F. H. C. Crick, *ibid.* A29, 56 (1973).
11. R. W. Schevitz, A. D. Podjarny, M. Zwick, J. J. Hughes, P. B. Sigler, *ibid.* A37, 669 (1981).
12. T. A. Jones, *J. Appl. Cryst.* 11, 268 (1978).
13. M. Verhoeyen, C. Berek, J. M. Jarvis, G. Winter, personal communication.
14. A. Jack and M. Levitt, *Acta Crystallogr.* A34, 931 (1978).
15. M. G. Rossmann and D. M. Blow, *ibid.* 14, 631 (1961); D. W. Rice, *ibid.* A37, 491 (1981).
16. R. J. Poljak *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 70, 3305 (1973); M. Marquart, J. Deisenhofer, R. Huber, *J. Mol. Biol.* 141, 369 (1980).
17. C. Chothia *et al.*, *Science* 233, 755 (1986).
18. R. A. Dwek *et al.*, *Nature (London)* 266, 31 (1977).
19. C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips, V. R. Sarma, *ibid.* 206, 757 (1965).
20. V. Luzzati, *Acta Crystallogr.* 5, 802 (1952).
21. M. Ramanadham, L. C. Sieker, L. N. Jensen, personal communication; S. A. Mason, G. A. Bentley, G. J. McIntyre, in *Neutrons in Biology*, B. P. Schoenborn, Ed. (Plenum, New York, 1984), p. 323.
22. J. Berthou, A. Lifchitz, P. Artymiuk, P. Jollès, *Proc. R. Soc. London* B217, 471 (1983).
23. C. Chothia and J. Janin, *Nature (London)* 256, 705 (1975).
24. B. Lee and F. M. Richards, *J. Mol. Biol.* 55, 379 (1971).
25. M. Harper, F. Lema, G. Boulot, R. J. Poljak, *Mol. Immunol.*, in press.
26. E. M. Prager, N. Arnheim, G. A. Mross, A. C. Wilson, *J. Biol. Chem* 247, 2905 (1972); I. M. Ibrahimi, J. Eder, E. M. Prager, A. C. Wilson, R. Arnon, *Mol. Immunol* 17, 37 (1980); I. M. Ibrahimi, E. M. Prager, T. J. White, A. C. Wilson, *Biochemistry* 13, 2736 (1979).
27. M. Kaneda, I. Kato, N. Tominaga, K. Titani, K. J. Narita, *J. Biochem Tokyo* 66, 747 (1969).
28. J. N. LaRue and J. C. Speck, *J. Biol. Chem* 245, 1985 (1970).
29. P. Jollès and J. Jollès, *Mol. Cell. Biochem.* 63, 165 (1984).
30. M. Levitt, thesis, University of Cambridge (1974).
31. N. Sakato, H. Fujio, T. Amano, *Biken J.* 14, 405 (1971); I. Pecht, E. Maron, R. Arnon, M. Sela, *Eur. J. Biochem* 19, 368 (1971); T. Kobayashi *et al.*, *Mol Immunol* 19, 619 (1982).
32. F. Karush, in *Comprehensive Immunology*, R. A. Good and S. B. Day, Eds. (Plenum, New York, 1978), vol. 5, p. 85.
33. S. J. Smith-Gill, T. B. Lavoie, C. R. Mainhart, *J. Immunol.* 133, 384 (1984); D. W. Metzger, L. K. Ch'ng, A. Miller, E. E. Sercarz, *Eur. J. Immunol.* 14, 87 (1984).
34. S. J. Leach, *Biopolymers* 22, 425 (1983); P. E. E. Todd, I. J. East, S. J. Leach, *Trends Biochem. Sci.* 7, 212 (1982); J. A. Berzofsky, *Science* 229, 932 (1985).
35. R. J. Poljak, L. M. Amzel, B. L. Chen, R. P. Phizackerley, F. Saul, *Proc. Natl. Acad. Sci. U.S.A.* 71, 3440 (1974); R. Huber, J. Deisenhofer, P. M. Colman, M. Matsushima, W. Palm, *Nature (London)* 264, 415 (1976).
36. H. Metzger, *Contemp. Top. Mol. Immunol.* 7, 191 (1978).
37. E. Fischer, *Chem. Ber.* 27, 2985 (1894); P. Ehrlich, *Proc. R. Soc. London Biol.* 66, 424 (1900).
38. S. Tonegawa, *Nature (London)* 302, 575 (1983); T. Honjo, *Annu. Rev. Immunol.* 1, 499 (1983).
39. P. Early, H. Huang, M. Davis, K. Calame, L. Hood, *Cell* 19, 981 (1980); H. Sakano, R. Maki, Y. Kurosawa, W. Roeder, S. Tonegawa, *Nature (London)* 286, 676 (1980).
40. J. Schilling, B. Clevinger, J. M. Davie, L. Hood, *Nature (London)* 283, 35 (1980).
41. H. Sakano, K. Huppi, G. Heinrich, S. Tonegawa, *ibid.* 280, 288 (1979); E. E. Max, J. C. Seidman, P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* 76, 3450 (1979); M. Weigert *et al.*, *Nature (London)* 283, 497 (1980).
42. Y. Kurosawa *et al.*, *Nature (London)* 290, 565 (1981).
43. A. B. Edmundson and K. R. Ely, *Ann. Inst. Pasteur Immunol.* 136C, 276 (1985); ———, J. N. Herron, *Mol. Immunol* 21, 561 (1984).
44. J. Oudin and M. Michel, *C. R. Seances Acad. Sci. Paris* 257, 805 (1963); H. G. Kunkel, M. Mannik, R. C. Williams, *Science* 140, 1218 (1963).
45. Reviewed in J. E. Hopper and A. Nisonoff, *Adv. Immunol.* 13, 58 (1971); J. D. Capra and J. M. Kehoe, *ibid.* 20, 1 (1975); K. Eichmann, *ibid.* 26, 195 (1978); K. Rajewsky and T. Takemori, *Annu. Rev. Immunol* 1, 569 (1983).
46. E. A. Kabat, in *The Biology of Idiotypes*, M. I. Greene and A. Nisonoff, Eds. (Plenum, New York, 1984), pp. 3–17.
47. J. Sharon, E. A. Kabat, S. L. Morrison, *Mol. Immunol* 19, 375 (1982).
48. M. Weigert, W. C. Raschke, D. Carson, M. Cohn, *J. Exp. Med.* 139, 137 (1974).


49. Supported by grants from Institut Pasteur, C.N.R.S., INSERM, and a short-term EMBO Fellowship (ASTF 4777) (to S.F.V.P.). We thank M. Verhoeyen for making the $V_H$ and $V_L$ amino acid sequences of antibody D1.3 available to us before publication, C. Wilmot for available surface area and energy calculations, G. Bricogne for help in phase refinement calculations, G. Boulot for purification of IgG1 D1.3 and its Fab, F. Gauthier for expert secretarial help, the Laboratoire d'Utilisation du Rayonnement Synchrotron, C.N.R.S., Orsay, for the use of a graphics system.

16 April 1986; accepted 27 June 1986

**Genetic mechanism of generating various antibody genes**

During the 1970's, Tonegawa and his collaborators tried to understand why mouse embryos did not produce antibodies. In 1978, they determined the DNA sequences of both adult and fetal mouse lambda light chains and discovered a novel mechanism at the DNA level to generate various antibody genes (Bernard O, Hozumi N, Tonegawa S (1978) Sequences of mouse immunoglobulin light chain genes before and after somatic changes. Cell, 15, 1133-1144). Two separate genomic DNA segments, V and J, were joined together by DNA re-arrangement. The amino acid segment coded by the J minigene was previously identified as the switch peptide (Milstein C (1967) Linked groups of residues in immunoglobulin chains. Nature, 216, 330-332).

They summarized the gene arrangements are as follows:

Adult:

5' untranslated region, L, I-small, V linked with J, I-large, C, 3' untranslated region

Fetus:

5' untranslated region, L, V, large gap, J, I-large, C, 3' untranslated region

Subsequent studies have discovered that the large gap in fetal DNA consisted of around 100 kilobases. Also, there are about 100 V genes and around 5 different J minigenes for bothe the kappa and lambda light chains.

The joining of V gene to the J minigene required special sequences adjacent to them known as recombination signal sequences (RSS).

In the case of heavy chains, the situation is slightly more complicated:

Adult:

5' untranslated region, L, I-small, V linked with D and with J, I-large, C, 3' untranslated region

Fetal:

5' untranslated region, L, I-small, V large gap, D, large gap, J, I-large, several C's, 3' untranslated region

For heavy chains there are also about 100 V genes, 50 D minigenes and 6 J minigenes in the human system. Two re-arrangements are required, i.e. V to D and D to J.

The formation of CDRH3 becomes very complicated. Due to the special features of RSS at the ends of V genes, D minigenes and J minigenes, several different re-arrangements are possible:

1. Extra nucleotides can be placed between V gene and D minigene, as well as between D minigene and J minigene. If these extra nuleotides are G,C-rich, they are known a N segment.

2. If the extra nucleotides are A,T-rich, they are known as P segments. In most cases, they are general designated as N/P-segment.

3. The RSS at the end of V gene in the heavy chain has the 7-23-9 pattern. Those at both ends of the D minigene are 9-12-7 and 7-12-9. In front of the J minigene, the RSS is 9-23-7. Due to the symmetry of RSS around D minigene, the D minigene can be inverted in the process of re-arrangement.

4. Varying number of nucleotides in the N/P segments between V and D and D and J, with the sum being divisible hy three, can shift the reading frame of the D minigene.

5. Some longer D minigenes may have an internal 7-mer, thus D-D joining becomes possible.

It is probably for these reasons that CDRH3 plays a unique role in conferring antibody specificity. All antibodies with different specificities have different CDRH3 sequences, while some of the other five CDRs can be identical in amino acid sequences. In designing new therapeutic antibodies, it will be essential to characterize CDRH3 first.

Cell, Vol. 15, 1133–1144, December 1976, Copyright © 1976 by MIT

# Sequences of Mouse Immunoglobulin Light Chain Genes before and after Somatic Changes

**Ora Bernard, Nobumichi Hozumi and Susumu Tonegawa**
Basel Institute for Immunology
Grenzacherstrasse 487
Postfach, CH-4005 Basel 5
Switzerland

## Summary

**We have determined the nucleotide sequences of the germ line gene as well as a corresponding somatically mutated and rearranged gene coding for a mouse immunoglobulin $\lambda_I$ type light chain. These sequencing studies were carried out on three Eco RI-DNA fragments which had been cloned from BALB/c mouse embryos or a $\lambda_I$ chain-secreting myeloma, H2020. The embryonic DNA clone Ig 99λ contains two protein-encoding segments, one for the majority of the hydrophobic leader (L) and the other for the rest of the leader and the variable (V) region of the $\lambda_0$ chain (Cohn et al., 1974); these segments are separated by a 93 base pair (bp) intervening sequence (I-small). The coding of the V region ends with His at residue 97. The second embryonic DNA clone Ig 25λ includes a 39 bp DNA segment (J) coding for the rest of the conventionally defined V region (that is, up to residue 110), and also contains the sequence coding for the constant (C) region approximately 1250 untranslated bp (I-large) away from the J sequence. The J sequence is directly linked with the V-coding sequence in the myeloma DNA clone, Ig 303λ, which has the various DNA segments arranged in the following order: 5′ untranslated region, L, I-small, V linked with J, I-large, C, 3′ untranslated sequence. The Ig 303λ V DNA sequence codes for the V region synthesized by the H2020 myeloma and is different from the Ig 99λ V DNA sequence by only two bases. No silent base change was observed between the two DNA clones for the entire sequence spanning the 5′ untranslated regions and the V-coding segments. These results confirm the previously drawn conclusion that an active complete $\lambda_I$ gene arises by somatic recombination that takes place at the ends of the V-coding DNA segment and the J sequence. No sequence homology was observed at or near the sites of the recombination.**

## Introduction

An immunoglobulin polypeptide chain is composed of two parts: a highly variable amino terminal region (V region) and a relatively constant carboxy terminal region (C region). The DNA segments coding for the two regions (V DNA and C DNA) are distant in embryonic cells and are rearranged by recombination that takes place in the precursors of lymphocytes (Hozumi and Tonegawa, 1976; Tonegawa et al., 1976; Brack et al., 1978; Lenhard-Schuller et al., 1978).

We have previously cloned, from both BALB/c mouse embryos and myeloma cells, various DNA fragments containing part or all of the λ or κ chain genes. Characterization of these DNA clones by R loop mapping, heteroduplex analysis and restriction enzyme mapping provided direct evidence for somatic recombination (Brack and Tonegawa, 1977; Tonegawa et al., 1977; Brack et al., 1978; Lenhard-Schuller et al., 1978). The same studies revealed a surprising feature of gene organization: somatic recombination does not generate a DNA stretch containing contiguously arranged V and C DNA segments. In a complete $\lambda_I$ chain gene expressed in myelomas, the $V_{\lambda I}$ and $C_{\lambda I}$ DNA segments are 1.2 kb apart, while in an active complete κ chain gene, the $V_\kappa$ and $C_\kappa$ DNA segments are 2.9 kb apart (Brack and Tonegawa, 1978; Lenhard-Schuller et al., 1978). In embryonic DNA, a short DNA segment called J (~50 bp) lies 1.2 kb upstream (that is, toward the 5′ end of the coding strand) from the $C_{\lambda I}$ DNA. The J sequence is homologous to a sequence near the V-C junction of a $\lambda_I$ mRNA (Brack et al., 1978). The recombination seems to take place at or near the 3′ end of the $V_{\lambda I}$ DNA and at or near the J sequence, generating a complete $\lambda_I$ gene composed of various DNA segments in the following order: 5′ V, J, 1.2 kb intervening sequence, C 3′ (Brack et al., 1978). The precise locations of the recombination sites, however, can only be revealed by determination of the DNA sequences.

Mouse λ chains are classified into the two subtypes $\lambda_I$ and $\lambda_{II}$. Amino acid sequences of the two subtype chains are different at ~10–12 positions in the V region and at about 25 positions in the C region (Weigert et al., 1970; Dugan et al., 1973). Total embryonic DNA gives two Eco RI fragments of 4.8 and 3.6 kb, which carry sequences homologous to the V part of a $\lambda_I$ hybridization probe (Tonegawa et al., 1977; Brack et al., 1978). We have previously cloned the 4.8 kb fragment and identified its $V_\lambda$ DNA as that of a $\lambda_{II}$ type by DNA sequencing. These studies revealed a 93 bp intervening sequence separating a DNA segment coding for the majority of the hydrophobic leader residues from the V DNA segment (Tonegawa et al., 1978). We would like to know whether an equivalent intervening sequence is present in the embryonic $V_{\lambda I}$ DNA, and if so, whether this intervening sequence persists in the expressed complete $\lambda_I$ gene in myeloma cells.

The BALB/c mouse genome carries no more

than a few copies of $V_{\lambda I}$ DNA segments per haploid genome (Tonegawa, 1976; Honjo et al., 1976). Since seven different (but highly homologous) $V_{\lambda I}$ amino acid sequences are known and more are very likely to exist (Weigert et al., 1970), we believe that a large proportion of the V region diversity is generated by a somatic process such as somatic mutation. Based on the frequent recurrence among myeloma $\lambda_I$ chains, Cohn et al. (1974) postulated that the V region represented by MOPC 104E is the product of a putative germ line V DNA segment. Sequencing of an embryonic $V_{\lambda I}$ clone will test this proposal directly. Furthermore, comparison of the embryonic V DNA sequence with that of a somatically derived myeloma V DNA may provide us with some information relevant to the mechanism of the somatic change.

To study the problems raised above and to analyze the fine structure of an immunoglobulin gene and the flanking regions, we have determined the nucleotide sequences of the relevant regions of three $\lambda_I$ DNA clones: an embryonic $V_{\lambda I}$ (Ig 99λ), an embryonic $C_{\lambda I}$ (Ig 25λ) and a myeloma $V_{\lambda I}$ plus $C_{\lambda I}$ clone (Ig 303λ).

## Results

The basic characteristics of the three $\lambda_I$ DNA clones are summarized in Table 1. It should be noted that the clone Ig 303λ was isolated from DNA prepared from H2020 myeloma, and contains an Eco RI-DNA fragment that is absent from embryonic DNA (Brack and Tonegawa, 1977).

### Restriction Enzyme Mapping of the Cloned DNA

Detailed restriction enzyme maps were constructed to sequence the cloned DNA. Our previous R loop mapping by electron microscopy showed the locations of the regions homologous to H2020 $\lambda_I$ mRNA (Brack et al., 1978). These locations are indicated in Figure 1. We have also previously determined the regions of homology among the cloned mouse DNA fragments by heteroduplex mapping (Brack et al., 1978). Thus about one third of Ig 303λ on one end is homologous to one end of Ig 99λ, and about two thirds of Ig 303λ on the other end is homologous to one end of Ig 25λ. Ig 99λ and Ig 25λ exhibited no detectable homology. The results of these analyses were used as guides in map construction.

The mapping was performed in two steps. First, isolated mouse Eco RI DNA fragments were digested with restriction enzymes such as Bam HI, Xba I and Kpn I, which cleave DNA relatively infrequently (Figure 1A). Digestion products of Ig 303λ and one of the two embryonic clones were fractionated by electrophoresis on a single slab gel to identify DNA fragments of a common size. Digestion products were also analyzed by the Southern gel blotting technique (Southern, 1975) to identify the DNA fragments containing sequences homologous to H2020 $\lambda_I$ mRNA. DNA fragments containing the ends of the Eco RI fragment were also identified by comparing the electrophoretic patterns generated by single (enzyme X) and double (enzyme X and Eco RI) digestion of the whole $\lambda gt_{WES}$-Ig phage DNA. These procedures allowed unique ordering of most of the DNA fragments. When ambiguity existed, fragment order was determined or confirmed by double enzyme digestion (enzyme X and enzyme Y). Figure 1A summarizes the results obtained with this approach.

In the second step, the DNA fragments containing sequences homologous to H2020 $\lambda_I$ mRNA were labeled at 5′ ends with $\gamma$-$^{32}$P-ATP using polynucleotide kinase and cleaved with restriction enzymes to yield two fragments labeled at only one of the 5′ ends. These fragments were isolated and partially digested with enzymes such as Alu I, Hae III, Hinf I and others which cut DNA relatively frequently. The digestion products were fractionated by polyacrylamide gel electrophoresis and enzyme cleavage sites were deduced (Smith and Birnstiel, 1976). In most of the cases, cleavage

Table 1. Characteristics of λ Light Chain DNA Clones

| Clone | DNA Source | Size of Cloned DNA (kb) | Type of λ DNA Contained | References |
|---|---|---|---|---|
| Ig 99λ | Embryo 3.2–3.8 kb | 3.5 | $V_{\lambda I}$ | Brack et al. (1978) |
| Ig 303λ | Myeloma H 2020 7.0–7.6 kb | 7.4 | $V_{\lambda I}$ + $C_{\lambda I}$ | Brack and Tonegawa (1977) |
| Ig 25λ | Embryo 8.2–8.8 kb | 8.6 | $C_{\lambda I}$ | Brack et al. (1978) |
| Ig 13λ | Embryo 4.5–5.2 kb | 4.8 | $V_{\lambda II}$ | Tonegawa et al. (1977) Tonegawa et al. (1978) |

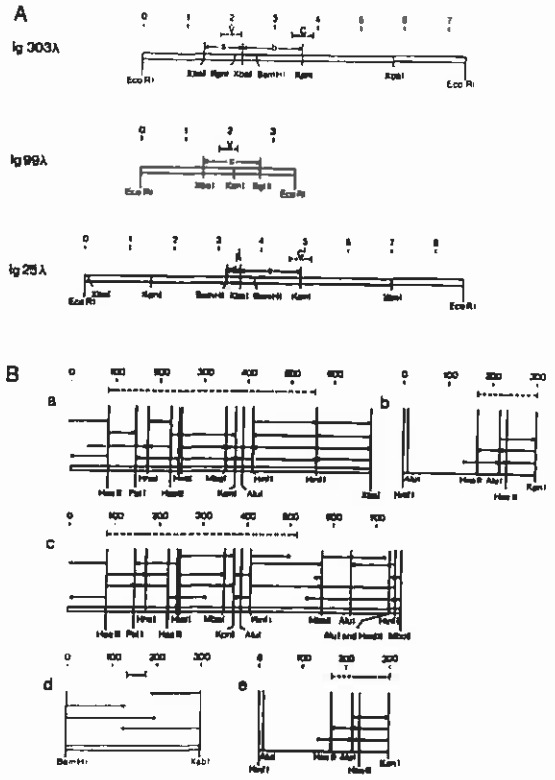


Figure 1. Restriction Enzyme Maps and Sequencing Strategies of the Three $\lambda_I$ DNA Clones

(A) Cleavage site maps of the cloned Eco RI-DNA fragments are shown together with the regions homologous to mRNA. The latter were determined by R loop mapping and Southern gel blotting analysis. (a) The fragment used to sequence the V DNA and the flanking regions of Ig 303λ. (b) The fragment used to sequence the 5′ end region of the C DNA of Ig 303λ. (c) The fragment used to sequence the V DNA and the flanking regions of Ig 99λ. (d) The fragment used to sequence the J DNA and the flanking regions of Ig 25λ. (e) The fragment containing the sequence of the 5′ end of the C DNA of Ig 25λ. Fragment sizes are given in kb.

(B) Detailed restriction enzyme maps of all or part of the DNA fragments (a–e) indicated in (A) are shown together with the regions sequenced. The dotted lines indicate regions homologous to mRNA. Fragment sizes are given in bp.

sites determined by the partial digestion were confirmed by complete digestion. Figure 1B summarizes the results obtained with this approach.

### DNA Sequencing Strategy

For DNA sequencing of Ig 303λ, whole phage DNA was digested with Xba I and fractionated by 5% polyacrylamide gel electrophoresis (Figure 1A). Xba I digestion was most useful for our purposes because it does not digest λ phage arms and cleaves the inserted DNA at three positions. The 0.9 kb fragment obtained in this way contains sequences homologous to the V region of H2020 $\lambda_I$ mRNA (Figure 1A, fragment a). Fragment a was further digested with the restriction enzymes indicated in Figure 1Ba, and then labeled at the 5′ ends with $\gamma$-$^{32}$P-ATP using polynucleotide kinase. The labeled DNA was either cut with another restriction enzyme to give two fragments labeled only at one end, or it was subjected to strand separation. For sequencing the C DNA segment of Ig 303λ, whole $\lambda gt_{WES}$-Ig 303λ DNA was digested with Xba I. The largest fragment (3.6 kb) was then digested with Kpn I to give fragment b (Figure 1A). This fragment was further digested with the restriction enzymes indicated in Figure 1Bb, and sequenced as described below.

Whole $\lambda gt_{WES}$-Ig 99λ DNA was digested with Bgl II and Xba I, and the 1.4 kb DNA fragment containing the V DNA segment was isolated (Figure 1A, fragment c). This fragment was digested with the restriction enzymes indicated in Figure 1Bc and then labeled at the 5′ ends for DNA sequencing.

Whole $\lambda gt_{WES}$-Ig 25λ was digested with Xba I. A 3.5 kb DNA fragment which contains the J sequence and extends from 0.05–3.5 kb from the left Eco RI site was isolated by electrophoresis. This fragment was further digested with Bam HI and the 5′ ends were labeled. A 300 base long fragment containing the J sequence was isolated (Figure 1A, fragment d), and the strands separated by electrophoresis on a 5% polyacrylamide gel were sequenced (Figure 1Bd). The C DNA segment of Ig 25λ is present in the 3.6 kb fragment of the Xba I digest. This fragment was further digested with Kpn I to give fragment e (Figure 1A). The sequencing strategy is outlined in Figure 1Be and was the same as for the corresponding region of Ig 303λ.

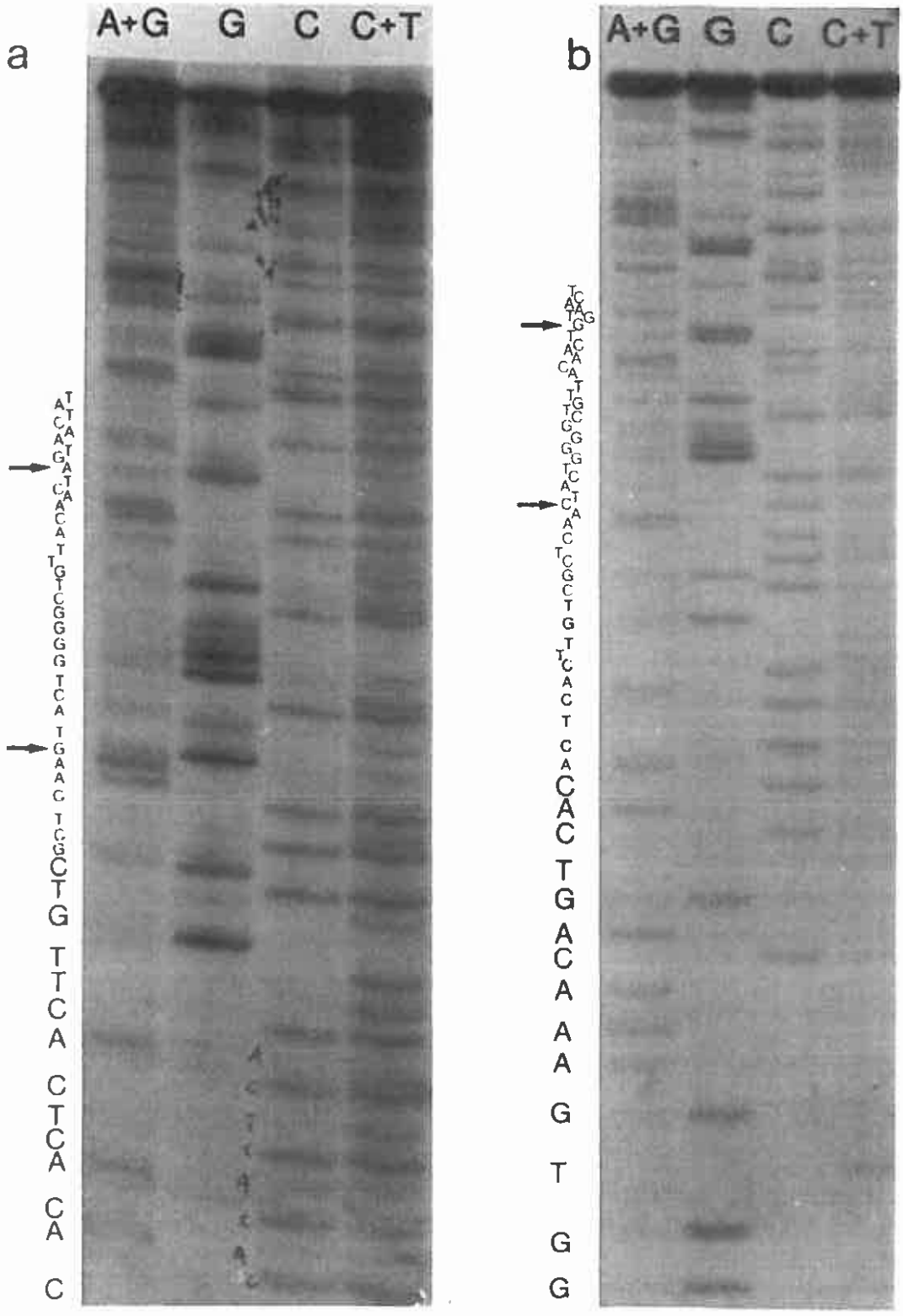
Using 40 cm long gels and the three polyacrylamide concentrations (12, 20 and 24%), we were able to read sequences up to about 200 nucleotides. Typical examples of the sequencing gels are shown in Figure 2. As shown in Figure 1B, gel sequences presented in this report were determined at least twice and, in the case of doubtful assignments, up to four times.

### Nucleotide Sequences

DNA sequences around the V segments of Ig 303λ and Ig 99λ, around the J segment of Ig 25λ, and around the C segments of Ig 303λ and Ig 25λ are shown in Figure 3. The determined DNA sequences and the amino acid sequences derived from them are presented.

#### *Ig 99λ*

The DNA sequence of Ig 99λ corresponds to a $V\lambda_I$ region and more specifically to the V region of the $\lambda_{II}$ chains (Cohn et al., 1974). The coding beings with the first amino acid of the hydrophobic leader sequences, Met at position −19; continues to Ser at position −5 (or Gly at position −4; see below); is


a
A+G
G
C
C+T
b
A+G
G
C
C+T

interrupted by a 93 bp intervening sequence; resumes with Gly at position −4 (or Ala at position −3; see below); and ends abruptly with His at position 97. The leader amino acid sequence predicted by the determined DNA sequence is exactly that determined by protein chemistry (Burstein and Schechter, 1977b).

In the V region, the amino acid sequence predicted by the DNA sequence fits that determined by protein chemistry (Cesari and Weigert, 1973). Furthermore, the DNA sequence clarifies several ambiguous assignments made by protein chemistry due to difficulties in distinguishing between Glu and Gln or Asp and Asn. Thus according to the DNA sequences of both the Ig 99λ and Ig 303λ, Glu is found at positions 7, 16, 40, 83 and 85, while Gln is at positions 6, 39 and 81. Likewise, Asp is at positions 71 and 84, while Asn is at positions 33, 36, 54, 55 and 96. The above residue assignments are in partial agreement with those reported by Burstein and Schechter (1977a), while they are at variance with the sequence reported by Appella (1971) at positions 7, 40 and 71.

The exact boundaries of the 93 bp intervening sequence are ambiguous because the quadruplet CAGG followed by T or G is repeated at the opposite boundary regions. Thus coding may be interrupted either within the Ser codon TCA (position −5) or within or at the end of the Gly codon GGT (position −4). Note that the triplets GGT and GGG at the opposite ends of the intervening sequence are synonymous and both code for Gly.

***Ig 303λ***

As expected, the DNA sequence of the V coding segment corresponds to the V region of the H2020 λ chain (Weigert et al., 1970). As in Ig 99λ, coding beings with Met at position −19 and is interrupted by a 93 bp intervening sequence; unlike the case of Ig 99λ, the coding continues beyond His at position 97 until Gly at position 110, where it ends again abruptly. Almost all the DNA sequences of the two clones Ig 99λ and Ig 303λ are identical in the entire stretch extending from the 5′ end untranslated region, through the leader, through the intervening region and through the V coding region, to the His codon at position 97. Only two exceptions to this almost perfect sequence identity were found, and these were in the codons for residues 25 and 32. The determined triplet codons of Ig 99λ are AGT for Ser at both positions, while those of Ig 303λ are ACT for Thr at position 25 and GGT for Gly at position 32 (Figure 2). In contrast, in the more than 160 bp that follow the His codon CAT at position 97, Ig 99λ and Ig 303λ have no sequence resemblance.

Approximately 1.2 kb away from the V DNA segment (Brack and Tonegawa, 1977), coding of the C region starts with Gly at position 110 and continues without interruption until Gly at position 156. The nucleotide sequence beyond this position is not dealt with in this paper. The precise boundaries of the 1.2 kb intervening sequence are again ambiguous because of the directly repeated Gly codons, GGT at the end of the expanded myeloma V DNA segment and GGC at the beginning of the C DNA segment. Due to the degeneracy of Leu codons, it is even possible that the coding interruption occurs between the second and third nucleotides of the codon CTA for Leu at position 109.

***Ig 25λ***

As indicated above, the Ig 303λ V DNA segment codes for amino acids until Gly at position 110, while coding by the Ig 99λ V DNA segment ends with His at position 97. The 39 base long sequence necessary for encoding the extra 13 residues at the end of the Ig 303 V region was discovered in the J segment of Ig 25λ. Coding in the J segment starts with Trp at position 98 and ends with Gly at position 110. The region preceding the J segment has no sequence resemblance with the corresponding regions of either Ig 303λ or Ig 99λ. In contrast, the sequence in the region following the J segment is identical with that of Ig 303λ, at least for the first 130 bp shown in Figure 3. The only exception was found at 48 bp away from the first nucleotide immediately following the triplet GGT for Gly at position 110 (indicated by an asterisk in Figure 3). We observed A in Ig 303λ and G in Ig 25λ. We do not know the exact reason for this single base difference. We speculate that it may be due to residual allele heterogeneity among individual BALB/c mice. Alternatively, the single base change might have occurred either during propagation of H2020 myeloma cells or during propagation of the DNA clones in *E. coli*.

Partial sequences of the Ig 25λ C DNA segment

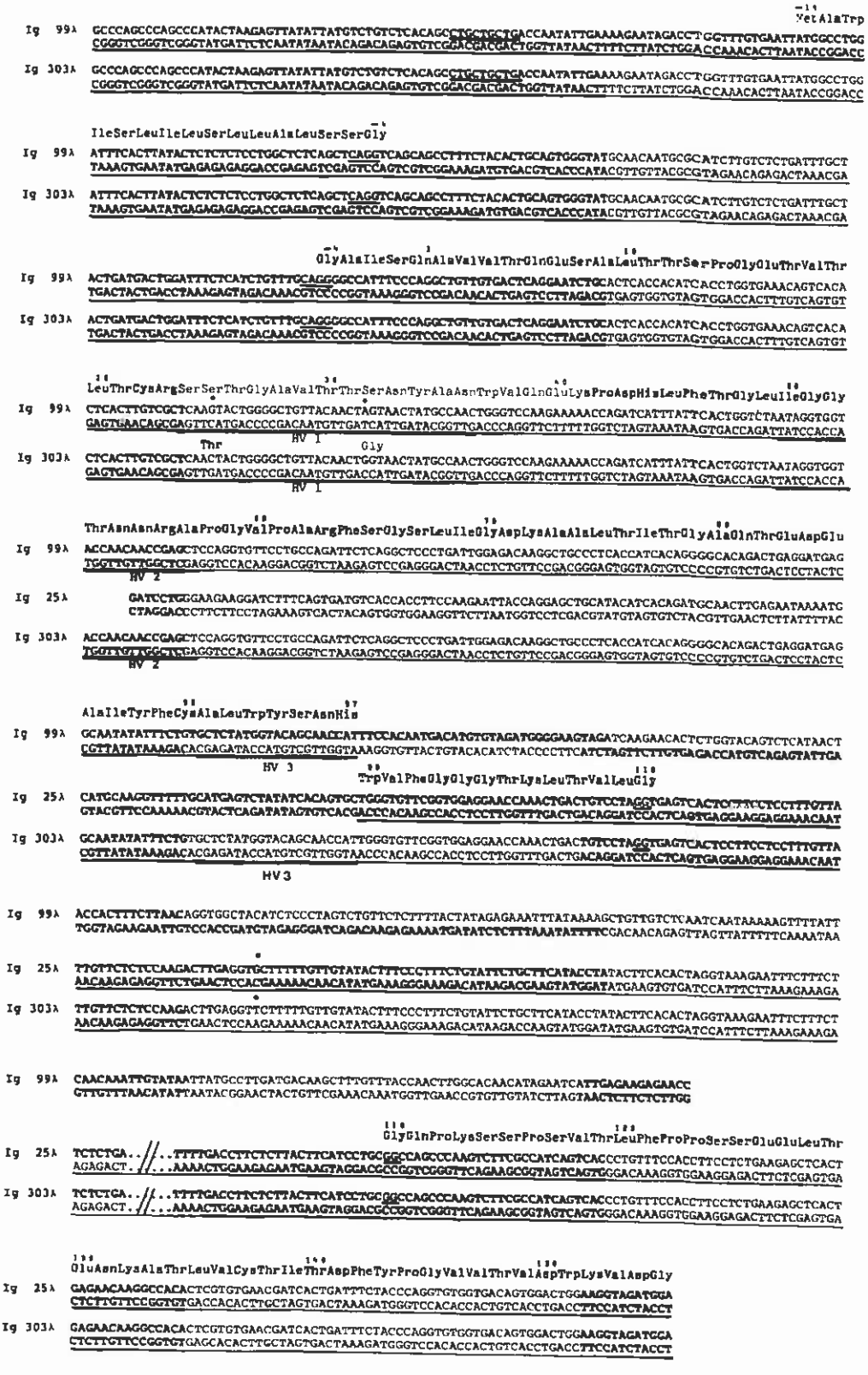
Figure 2. Autoradiogram of Sequencing Gels Illustrating Minor Sequence Differences in the V Coding Part of Ig 99λ and Ig 303λ

(a) Fragment c (Figure 1A) isolated from Ig 99λ was also digested with Hinf I, labeled at its 5′ ends and separated on an 8% acrylamide gel. The largest fragment was digested with Mbo I and the larger of the two resulting fragments was sequenced. The sequence indicated in the figure codes for amino acids 19–34 (Figure 3).

(b) Fragment a (Figure 1A) isolated from Ig 303λ was digested with Hinf I and labeled at its 5′ ends. After fractionation on an 8% acrylamide gel, the second largest DNA fragment was isolated and digested into two fragments with Mbo I. The larger of the two fragments was sequenced from the Hinf I cleavage end (Figure 1Ba). The nucleotide sequence that appears in the autoradiogram and is indicated by letter labels codes for amino acids 15–33 (Figure 3).

In both (a) and (b), the arrows indicate nucleotides that are different in Ig 99λ and Ig 303λ.

115


Ig 99λ
Ig 303λ
Ig 25λ
HV 1
HV 2
HV 3
HV3

and the region immediately preceding the C DNA segment were also determined and are shown in Figure 3. The sequences are identical with those of the corresponding regions of Ig 303λ.

## Discussion

### Precise Position and Characteristics of the Recombination Sites

Our previous heteroduplex and R loop mapping studies predicted that recombination takes place near the 3′ end of the embryonic V DNA segment and near the J segment that lies 1.2 kb upstream (that is, toward the 5′ side) from the embryonic C DNA segment (Brack et al., 1978). The present DNA sequencing studies confirmed this prediction and determined the recombination site to within two phosphodiester bonds. Thus in embryonic DNA, recombination takes place at the 3′ side of the His codon CAT (position 97) on Ig 99λ and at the 5′ side of the Trp codon TGG (position 98) on Ig 25λ. Alternatively, it takes place at the 3′ side of T following the His codon and at the 5′ side of G comprising the second nucleotide of the Trp codon. In either case, the recombination event attaches the embryonic V DNA segment directly to the J segment, generating an expanded V DNA segment observed in the myeloma clone.

A striking feature of the sequences around the recombination sites of the two embryonic clones is the lack of even a short homology (except for a single base homology of T). In fact, had there been a longer homology, it would not have been possible to define the recombination sites with the accuracy described above. The lack of sequence homology in the V-J recombination sites contrasts with the presence of 15 bp core homology at the *att* sites for the site-specific recombination of phage λ (Landy and Ross, 1977). In this case, although the biological significance of the common core remains unclear, its existence might be taken as confirmation of the suggestion that the core homology serves as the recognition signal for the site-specific recombination (Shulman and Gottesman, 1973). Lack of a core homology at the V-J recombination sites rules out any mechanism in whcih a sequence homology has a key role in the site determination. In this respect, the V-J recombination bears some resemblance to the recombinations accompanying mobilization of the transposable elements in procaryotic systems (for reviews, see Cohen, 1976; Starlinger and Saedler, 1976; Bukhari, Shapiro and Adhya, 1977). There is increasing evidence suggesting that at least a major part of the site recognition in the procaryotic *rec*-independent recombinations is not carried out by sequence homology, as might have been predicted by the Campbell model (Campbell, 1962; L. Johnsrud, M. P. Calos and J. H. Miller, personal communication).

In procaryotic systems, the degree of site specificity on the recipient genomes varies widely from one case to another, with phage Mμ and phage λ exhibiting the least and the most restricted specificity. In the case of recombination involving the immunoglobulin DNA segments, one would expect that the specificity of the site recognition is high so that efficient generation of a properly rearranged, complete immunoglobulin gene is assured in lymphocyte precursors. Our recent studies suggest, however, that the V or the J-C segment occasionally interacts with DNA segments elsewhere in the genome, generating apparently "aberrant" recombinants (Lenhard-Schuller et al., 1978). This may be an indication that the site recognition system used by V-J recombination has specificity less stringent than its involvement in the generation of an active complete gene might imply.

Does V-J recombination always take place in as narrowly defined regions as the present studies suggest? The apparent precision of the V-J recombination sites should be reconsidered in the light of the unique nature of immunoglobulin polypeptide chains. As is discussed below, we believe that the entire population of the BALB/c $V_\lambda$ DNA segments arises by somatic alteration (perhaps by

Figure 3. Nucleotide Sequences of the Three $\lambda_1$ DNA Clones

Determined nucleotide sequences of Ig 99λ, Ig 303λ and Ig 25λ, aligned in phase with the protein-encoding sequences, are shown together with the amino acid sequences predicted from the triplet codons. Nucleotide sequences are underlined where a pair of clones exhibit an identical sequence. Where sequence differences occur, they are indicated by an asterisk(*). The hypervariable regions HV1, HV2 and HV3 are underlined with thick lines. Dotted lines with a double slash indicate part of the 1250 bp intervening sequences for which no nucleotide sequence data are reported in this paper. In 99λ, coding begins with the first amino acid of the hydrophobic leader (Met at position −19), is interrupted at Ser (−5)/Gly(−4) by a 93 bp intervening sequence and resumes with Gly(−4)/Ala(−3). The V region coded in Ig 99λ is that of the $\lambda_1$ chain, but coding ends prematurely with His at position 97. A 39 bp DNA segment (J sequence) coding for the rest of the conventionally defined V region—that is, a peptide starting with Trp (98) and ending with Gly (110)—was found in Ig 25λ. The J sequence is followed by an approximately 1250 bp intervening sequence that is followed in turn by the C region coding sequence beginning with Gly (110)/Gln(111). Except for the two nucleotides in HV1, the nucleotide sequence of Ig 303λ is identical to that of Ig 99λ in the entire DNA segment extending from the 5′ untranslated region to the His (97) codon, after which it switches to the sequence of Ig 25λ except in the case of the one base pair indicated. The V region encoded in Ig 303λ is that of the H2020 λ chain. Short sequence repeats at the ends of the intervening sequences as well as the CTG repeats toward the 5′ ends are indicated by underlining between the two complementary DNA strands (see the text).

mutations) from a single germ line $V_{\lambda I}$ DNA segment. As indicated in Figure 3, the V-J recombination site is immediately adjacent to one of the three regions (designated HV3) where the putative somatic alterations are frequent (Cohn et al., 1974). It thus remains possible that at least part of the somatic sequence change in HV3 is brought about by somatic recombinations that take place somewhere within the HV3 coding region and in a region to the left of the J segment. Such recombinations will generate a series of HV3-coding DNA composed partly of germ line V DNA sequence and partly of the sequence preceding the J segment. Although the available amino acid sequences of the $V_{\lambda I}$ regions do not support this notion, it nevertheless remains a possibility in view of the paucity of the data (Weigert et al., 1970).

**Somatic Generation of $V_\lambda$ DNA Diversity**

Weigert et al. (1970) have analyzed the amino acid sequences of $\lambda_I$ chains isolated from eighteen independently isolated BALB/c mouse plasmacytomas. Twelve of the eighteen $\lambda_I$ chains have the same amino acid sequence, called $\lambda_0$ (Cohn et al., 1974). The amino acid sequences of the other six $\lambda_I$ chains differ from the $\lambda_0$ type by one to three residues, and all the differences occur in the three regions termed hypervariable. Previous hybridization studies indicated that the entire population of $V_{\lambda I}$ regions represented by BALB/c mouse plasmacytomas arise from no more than a few copies of germ line V DNA segments (Tonegawa, 1976; Honjo et al., 1976). There must be a somatic mechanism by which the coding sequence of the germ line $V_\lambda$ DNA is diversified in clones of immunoglobulin-producing lymphocytes. Our subsequent restriction enzyme mapping of whole cell DNA confirmed this notion and strongly suggested that there is only a single copy of $V_{\lambda I}$ DNA per haploid genome (Brack et al., 1978; M. Hirama and S. Tonegawa, unpublished observations). Since it is highly improbable that the somatic mechanism will generate identical V regions at a frequency as high as one out of three cases, Cohn et al. (1974) postulated that the V region of the $\lambda_0$ type chain is a product of the germ line V DNA. The present DNA sequence of embryonic $V_\lambda$ DNA of Ig 99λ confirmed this postulate.

Plasmacytoma H2020 synthesizes a λ chain whose amino acid sequence is different from the $\lambda_0$ type chain sequence at positions 25 and 32 (Weigert et al., 1970). Since Ig 303λ was cloned from an H2020 DNA fraction that had been enriched for the somatically rearranged active λ gene, we expected its V DNA segment to code for the H2020 V region. As shown in Figure 3, this expectation was fulfilled. The DNA sequences of Ig 99λ and Ig 303λ show that in each of the two positions where the $\lambda_0$ and the H2020 V regions are different, the involved triplet codons are related by a single base difference. The DNA sequences of the two clones are identical except for these two base positions until the His codon at position 97. We observed neither silent base changes in the protein-encoding regions, nor any difference in the 5′ noncoding regions or in the 93 nucleotide long intervening sequences (see below for additional discussion of intervening sequences). These results are compatible with the notion that lymphocytes have a special enzymatic mechanism that preferentially introduces mutations to the hypervariable regions (Brenner and Milstein, 1966). The results, however, are also compatible with the hypothesis that mutations occur indiscriminately along the entire V DNA segment at a relatively low rate, and only those V DNAs that have acquired missense mutations in hypervariable regions are selected at the cellular level. In this case, our results show that the rate of mutation should be sufficiently low that the appearance of neutral base changes throughout the V DNA segment is infrequent.

**Multiple Intervening Sequences in Both Embryo and Plasmacytoma λ Chain Genes**

The sequence of Ig 99λ DNA revealed a 93 bp long intervening sequence toward the amino terminal end of the embryonic $V_{\lambda I}$ DNA segment. The position and length of this intervening sequence are identical with those found in the embryonic $V_{\lambda II}$ DNA segment (Tonegawa et al., 1978), although the two sequences are different at four nucleotide positions. We term the two intervening sequences $I_{\lambda I}$-small and $I_{\lambda II}$-small. $I_{\lambda I}$-small is also present in Ig 303λ, indicating that somatic rearrangement at the DNA level does not eliminate the noncoding sequence. The 1.2 kb intervening sequence present in Ig 25λ (termed $I_{\lambda I}$-large) also remains in the active rearranged complete $\lambda_I$ chain gene represented by Ig 303λ. Sequencing of the C DNA segment did not reveal any additional intervening sequence in the region extending from position 111 to position 155. Our preliminary sequence data for the rest of the C DNA segment indicate that no additional coding interruption occurs in this part of the molecule (O. B. and S. T., unpublished data). Thus it seems that a complete $\lambda_I$ gene has two intervening sequences, $I_{\lambda I}$-small and $I_{\lambda I}$-large, in the translated portion of the gene. From the size of the λ mRNA (Tonegawa, 1976) and from the known anatomy of the κ mRNA (Milstein et al., 1974), we expect that there are about 150 and 200 base long untranslated regions at the 5′ end and the 3′ end of a λ mRNA molecule, respectively. R loop mapping of both Ig 99λ and Ig 303λ did not give any hint of

additional intervening sequences in these untranslated regions (Brack et al., 1978). The presence of short (<100 base) intervening sequences, however, would not be revealed by this procedure. In this respect, it is interesting to note that in both Ig 99λ and Ig 303λ, three repeats of CTG are present about 50 nucleotides upstream from the translation initiation codon AUG for Met at position -19 (Figure 3). The same nonanucleotide was present at the equivalent position in the embryonic $V_{\lambda II}$ clone, Ig 13λ (Tonegawa et al., 1978). A cloned rabbit β-globin DNA fragment (A. Efstratiadis, T. Maniatis and E. Lacy, personal communication) and a cloned hen ovalbumin DNA fragment (Breathnach et al., 1978) also carry a similar repeat of CTG about 60 and 35 nucleotides upstream from the AUG initiation codon, respectively. In the case of the ovalbumin DNA clone, the CTG repeat is present within the 5′ end-most intervening sequence and near its junction with the adjacent DNA sequence which is expressed in the mature ovalbumin mRNA. It is tempting to speculate that the CTG repeats have a common role in all these gene systems.

Intervening sequences are thought to be eliminated on the RNA level by a splicing mechanism (Klessig, 1977; Berget, Moore and Sharp, 1977). While this remains to be demonstrated in the case of the mouse $\lambda_I$ chain gene, evidence in support of this notion is accumulation for other gene systems (Knapp et al., 1978; O'Farrell et al., 1978; Tilghman et al., 1978). What are the recognition signals for the putative RNA splicing enzymes? Probable candidates are hairpin structures and/or base sequences at or near the boundaries of protein-encoding and intervening sequences. We have previously reported that a hairpin structure of considerable stability can be formed around the $V_{\lambda II}$-small intervening sequences (Tonegawa et al., 1978). While a similar hairpin structure can also be formed around the $V_{\lambda I}$-small intervening sequences, nucleotide sequences around the larger intervening sequences do not allow the formation of a hairpin structure of analogous stability. As we have previously pointed out (Tonegawa et al., 1978), however, we should keep in mind the possibility that a sequence lying elsewhere in the RNA transcript may hold the ends of adjacent protein-encoding sequences in juxtaposition and facilitate RNA splicing.

Do the nucleotide sequences at the junction of protein-encoding and intervening sequences exhibit unusual features? The sequences of the $\lambda_I$ and $\lambda_{II}$ DNA at the splice points are listed in Figure 4. We notice the following sequence features. First, short nucleotide sequences are directly repeated at the ends of a given intervening sequence. Thus a

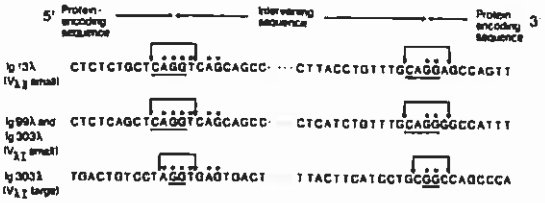



Figure 4. Sequence Features at the Boundaries of Protein-Encoding and Intervening Sequences

Three intervening sequences of the mouse $\lambda_I$ (present paper) and $\lambda_{II}$ (Tonegawa et al., 1978) genes are compared. Short sequences repeated at the opposite ends of an intervening sequence are underlined. Short sequences that are common among the three intervening sequences and are present at or near either their 5′ or 3′ ends are indicated by dots over the nucleotides. Brackets indicate the maximal and minimal boundaries of the intervening sequences.

tetranucleotide CGAA is repeated at opposite ends of both $V_{\lambda I}$-small and $V_{\lambda II}$-small intervening sequences, while a dinucleotide GG is repeated at the ends of the $V_{\lambda I}$ large intervening sequence. It is partly for this reason that unique identification of the exact splice points is impossible (see above). Second, an oligonucleotide AGGTXAG appears at the 5′ end of the three intervening sequences, while no common sequence other than a dinucleotide GG is present at the 3′ ends. These sequence features apply to other genes carrying intervening sequences, such as mouse and rabbit β-globin gene (J. Van den Berg, A. Van Ooyen, N. Mantei, A. Schambäck, G. Grosveld, R. A. Flavell and C. Weissmann, personal communication), hen ovalbumin genes (Breathnach et al., 1978), and SV40 early and late genes (Reddy et al., 1978; Ghosh et al., 1979). In all cases, 1–4 nucleotides are repeated at the opposite ends of a given intervening sequence. In most cases, either AGGTXAG or sequences related to it appear at the 5′ end. We believe that these sequences constitute at least part of the recognition signals for the splicing enzymes.

The presence of short direct repeats at the boundaries of all intervening sequences may be a requirement for the enzymatic reaction leading to RNA splicing. This is reminiscent of an analogous situation at the ends of insertion elements and transposons. In all cases investigated so far, these procaryotic DNA elements are flanked by a direct repeat of five or nine bases at the site of insertion (Grindley, 1978; Calos, Johnsrud and Miller, 1978; Rosenberg et al., 1978; L. Johnsrud, M. P. Calos and J. H. Miller, personal communication). It is possible that the short repeats at the boundaries of intervening sequences are evolutionary relics of their initial insertion events, which might have occurred by a mechanism similar to transposition of the procaryotic insertion elements. If this is the case, however, the origin of intervening sequences

is not in accord with a proposal made previously. We hypothesized that the intervening sequences are there not because they were inserted into a preexisting gene, but because genes are often created in evolution as a result of the emergence of a pair of RNA splice points at two separate DNA sites within a preexisting transcription unit (Brack et al., 1978; Tonegawa et al., 1978).

### Comparison of Nucleotide Sequences of Ig 99λ and Ig 13λ

Mouse λ chains are classified into the two subtypes $\lambda_I$ and $\lambda_{II}$. Amino acid sequences of the two subtype chains are different at ~10–12 positions in the V region and ~25 positions in the C region (Weigert et al., 1970; Dugan et al., 1973). Total embryonic DNA gives two Eco RI fragments of 4.8 and 3.6 kb carrying sequences homologous to the V part of a $\lambda_I$ hybridization probe (Tonegawa et al., 1977; Brack et al., 1978). We have previously cloned the 4.8 kb fragment and identified its $V_\lambda$ DNA as that of a $\lambda_{II}$ type by DNA sequencing (Tonegawa et al., 1978). Present identification of Ig 99λ (contains the 3.6 kb fragment) as a clone carrying a $V_{\lambda I}$ DNA sequence confirmed the notion that the two λ subtypes are encoded by separate germ line V DNA segments. Although a $C_{\lambda II}$ DNA segment has not been identified (it does not appear in a Southern gel blot assayed with a $\lambda_I$ hybridization probe, presumably because of the lack of sufficient sequence homology between the two C DNA segments), it is almost certain that such a DNA segment exists as a separate entity other than the $C_{\lambda I}$ DNA segment.

Figure 5 illustrates the difference in the DNA sequences of Ig 99λ and Ig 13λ around the V DNA segments. Comparison was made for ~350 base long protein-encoding regions, the 5′ end noncoding region, the 93 base long small intervening sequences and the 3′ end noncoding regions. The cumulative length of the three noncoding regions amounts to about 380 bp. As shown in Figure 5, the basic architecture of the two V DNA segments and the flanking regions is very similar in the two genes. A minor difference was observed at the end of the V-coding DNA segment. As mentioned above, Ig 99λ codes for the $V_{\lambda I}$ region up to position 97, while we have previously reported that Ig 13λ codes for the $V_{\lambda II}$ region up to position 98 (Tonegawa et al., 1978). We do not yet know whether the difference in the size of the V-coding region between the two DNA clones reflects a true difference in the distribution of coding sequences between V and J DNA segments in the two λ chain genes. An alternative possibility is that in the $\lambda_{II}$ gene, two Phe codons (which are not necessarily identical) appear at both the end of the V DNA segment and the beginning of the putative $J_{\lambda II}$ segment (J segment for $\lambda_{II}$ chain).

In the coding regions, we observed differences in 12 nucleotides. Ten were missense changes, causing assignment of different amino acids at the corresponding positions of the two subtype V regions. The other two differences were silent changes in the third base of triplet codons. In no case was more than one base difference observed within a corresponding pair of codons. In the noncoding regions, seventeen differences were observed within the approximately 400 bp long stretches that have been sequenced. Of these, three were in the 5′ noncoding region (85 bp), four were in the small intervening sequences (93 bp) and ten were in the 3′ noncoding region (200 bp). In the 3′ noncoding regions, six differences were due to three base long deletions or insertions occurring at two different sites. In this respect, it is interesting to note that the 3′ noncoding regions are rich in a short stretch of identical nucleotides,

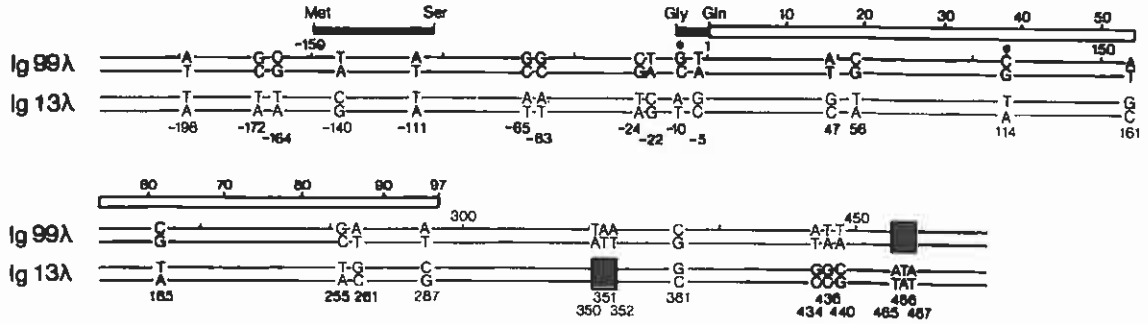



Figure 5. Schematic Illustration Comparing Nucleotide Sequences of Ig 99λ and Ig 13λ

Numbers indicate either nucleotide positions or amino acid positions. For nucleotide positions, number 1 is given to the first nucleotide of the triplet codon corresponding to the cyclized glutamine found at the amino terminus of mature light chains. Where the two sequences are different, the nucleotides involved are indicated together with their positions. Asterisks (*) indicate neutral base changes in the protein-encoding regions. The leader-encoding part is indicated with a black bar, while the mature V region-encoding part is indicated by open bars. Filled boxes indicate deletions or insertions.

particularly of T and A. Except for the 3′ end noncoding regions (where the density of base differences may be slightly higher than in other regions), the variation density is no higher in the noncoding regions than in the protein-encoding regions. The noncoding regions in question are considered to be transcribed into pre mRNA. The functions of the noncoding transcribed region may require a rather rigorous secondary structure, and this may be the primary reason why the nonprotein-encoding regions appear to be as conserved as the protein-encoding regions.

## Experimental Procedures

### Materials

$\gamma$-$^{32}$P-ATP (spec. act. 3000–5000 Ci/mmole) was purchased from The Radiochemical Centre (Amersham, England). T4 polynucleotide kinase was purified from T4-infected cells by the modified method described by Richardson (1965). Bacterial alkaline phosphatase was obtained from Worthington Biochemicals, and was used after removing the ammonium sulfate by centrifugation and dissolving the enzyme in 10 mM Tris (pH 7.4). Restriction endonucleases Eco RI (Greene, Betlach and Boyer, 1974), Bam HI (Wilson and Young, 1975) and Bgl II (Bickel, Pirrotta and Imber, 1977) were prepared according to published methods. Hinf I was prepared according to unpublished methods of R. J. Roberts (personal communication). All other restriction enzymes were purchased from New England Biolabs.

Phage $\lambda gt_{wes}$ Ig 25λ, 99λ and 303λ were grown in an EK2 host DP50 (Sull$^+$, SuIII$^+$) (Leder, Tiemeier and Enquist, 1977) in a P3 facility in accordance with the NIH Guidelines issued in June 1976.

### Preparation of Eco RI Inserts

Phage DNA was digested with Eco RI and the fragments were separated as previously described (Hozumi and Tonegawa, 1976). DNA was eluted from the gel by passing the agarose gel slice through a 22 gauge needle and centrifuging the crushed gel in a Beckman SW41 rotor at 35,000 rpm for 30 min. The DNA was precipitated from the supernatant by the addition of 2 vol of ethanol and dissolved in 10 mM Tris (pH 7.4).

### Preparative Gel Electrophoresis

DNA fragments for preparation of restriction enzyme maps and for sequencing were obtained from digestion of 303λ and 25λ DNA with Xba I and Eco RI. 2 mg of phage DNA were digested for 16 hr at 37°C with 100 U of Xba I in 0.15 M NaCl, 6 mM Tris–HCl (pH 7.9), 6 mM $MgCl_2$ and 6 mM 2-mercaptoethanol. For Eco RI digestion, the concentration of Tris buffer was adjusted to 0.1 M with 2 M Tris (pH 7.2) and the DNA was incubated with 1000 U of Eco RI for 1 hr. DNA was then concentrated twice with equal volumes of butanol-2 (Stafford and Bieber, 1975), dialyzed against 10 mM Tris (pH 7.4) for 4 hr and loaded onto a 5% acrylamide gel (slot size 10 × 0.5 cm). The mouse DNA insert fragments were extracted as described (Maxam and Gilbert, 1977), except that the eluate was passed through a glass-wool layer in a 5 ml syringe before ethanol precipitation. Ig 99λ DNA was digested with Eco RI, Xba I and Bgl II to yield three mouse DNA fragments (Figure 1) which were separated and eluted as described above.

### 5′ Terminal Labeling of Restriction Endonuclease Fragments

DNA fragments (10 μg) were incubated with 1 U of alkaline phosphatase for 1 hr in 50 mM Tris–HCl (pH 9), 10 mM $MgCl_2$, 6 mM dithiothreitol, and then extracted 3 times with phenol saturated with water. Residual phenol was removed by ether extraction followed by two ethanol precipitations. The DNA fragments were incubated at 37°C for 1 hr in 100 μl of 50 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 0.085 mM spermidine, 0.5 mCi $\gamma$-$^{32}$P-ATP and several units of T4 polynucleotide kinase. The reaction was stopped by adjusting the ammonium acetate concentration to 2 M. The majority of the free $\gamma$-$^{32}$P-ATP was removed by two successive ethanol precipitations. The fragments were separated on an 8% acrylamide gel and eluted as indicated above.

### DNA Sequencing

DNA fragments labeled at their 5′ end were cut with another restriction enzyme or the strands were separated as described by Maxam and Gilbert (1977). DNA sequencing was also performed according to the procedure described by Maxam and Gilbert (1977). For sequencing long fragments (more than 150 nucleotides), the DNA samples were loaded onto 0.5 × 300 × 400 mm, 6, 20 and 24% acrylamide gels.

### Acknowledgments

We thank Allan Maxam and Walter Gilbert for their advice on sequencing procedures and Richard Pink for his critical reading of the manuscript. The excellent technical help of Ms. E. Hansson and Mr. G. Dastoornikoo is highly appreciated, and we thank Ms. K. Perret for preparing the manuscript.

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.

Received September 26, 1978

### References

Appella, E. (1971). Amino acid sequences of two mouse immunoglobulin lambda chains. Proc. Nat. Acad. Sci. USA *68*, 590–594.

Berget, S. M., Moore, C. and Sharp, P. A. (1977). Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc. Nat. Acad. Sci. USA *74*, 3171–3175.

Bickle, T. A., Pirrotta, V. and Imber, R. (1977). A simple, general procedure for purifying restriction endonucleases. Nucl. Acids Res. *4*, 2561–2572.

Brack, C. and Tonegawa, S. (1977). Variable and constant parts of the immunoglobulin light chain gene of a mouse myeloma cell are 1250 nontranslated bases apart. Proc. Nat. Acad. Sci. USA *74*, 5652–5656.

Brack, C., Hirama, M., Lenhard-Schuller, R. and Tonegawa, S. (1978). A complete immunoglobulin gene is created by somatic recombination. Cell *15*, 1–14.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978). The ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-nitron boundaries. Proc. Nat. Acad. Sci. USA, in press.

Brenner, S. and Milstein, C. (1966). Origin of antibody variation. Nature *211*, 242–243.

Bukhari, A. I., Shapiro, J. A. and Adhya, S., eds. (1977). DNA Insertion Elements, Plasmids and Episomes (New York: Cold Spring Harbor Laboratory).

Burstein, Y. and Schechter, I. (1977a). Glutamine as a precursor to N-terminal pyrolid-2-one-5-carboxylic acid in mouse immunoglobulin λ-type light chains. Biochem. J. *165*, 347–354.

Burstein, Y. and Schechter, I. (1977b). Amino acid sequence of the $NH_2$-terminal extra piece segments of the precursors of mouse immunoglobulin $\lambda_1$-type and κ-type light chains. Proc. Nat. Acad. Sci. USA *74*, 716–720.

Calos, M. P., Johnsrud, L. and Miller, J. H. (1978). DNA sequence at the integration sites of the insertion element IS1. Cell *13*, 411–418.

Campbell, A. M. (1962). Episomes. Adv. Genet. *11*, 101–145.

Cesari, I. M. and Weigert, M. (1973). Mouse lambda-chain sequences. Proc. Nat. Acad. Sci. USA *70*, 2112–2116.

Cohen, S. N. (1976). Transposable genetic elements and plasmid evolutions. Nature *263*, 731–738.

Cohn, M., Blomberg, B., Geckler, W., Raschke, W., Riblet, R. and Weigert, M. (1974). First order considerations in analyzing the generator of diversity. I. In The Immune System, Genes, Receptors and Signals, E. E. Sercarz, A. R. Williams and C. F. Fox, eds. (New York: Academic Press), pp. 89–117.

Dugan, E. S., Bradshaw, R. A., Simms, E. S. and Eisen, H. N. (1973). Amino acid sequence of the light chain of a mouse myeloma protein (MOPC-315). Biochemistry *12*, 5400–5416.

Ghosh, P. K., Reddy, V. B., Swinscoe, J., Lebowitz, P. and Weissman, S. U. (1979). The heterogeneity and 5′ terminal structures of the late RNAs of simian virus 40. J. Mol. Biol., in press.

Greene, P. J., Betlach, M. C. and Boyer, H. W. (1974). The EcoRI restriction endonuclease. In Methods in Molecular Biology, *1*, R. B. Wickner, ed. (New York: Marcel Dekker), pp. 87–111.

Grindley, N. D. F. (1978). IS1 insertion generates duplication of a nine base pair sequence at its target site. Cell *13*, 419–426.

Honjo, T., Packman, S., Swan, D. and Leder, P. (1976). Quantitation of constant and variable region genes for mouse immunoglobulin λ chains. Biochemistry *15*, 2780–2785.

Hozumi, N. and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proc. Nat. Acad. Sci. USA *73*, 3628–3632.

Klessig, D. F. (1977). Two adenovirus mRNAs have a common 5′ terminal leader sequence encoded at least 10 kb upstream from their main coding regions. Cell *12*, 9–21.

Knapp, G., Beckmann, J. S., Johnson, P. F., Fuhrman, S. A. and Abelson, J. (1978). Transcription and processing of intervening sequences in yeast tRNA genes. Cell *14*, 221–236.

Landy, A. and Ross, W. (1977). Viral integration and excision: structure of the lambda att sites. Science *197*, 1147–1160.

Leder, P., Tiemeier, D. and Enquist, L. (1977). EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: the $\lambda gt_{WES}$ system. Science *196*, 175–177.

Lenhard-Schuller, R., Hohn, B., Brack, C., Hirama, M. and Tonegawa, S. (1978). DNA clones containing mouse immunoglobulin κ chain genes isolated by in vitro packaging into phage λ coats. Proc. Nat. Acad. Sci. USA *75*, in press.

Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Nat. Acad. Sci. USA *74*, 560–564.

Milstein, C., Brownlee, G. G., Cartwright, E. M., Jarvis, J. M. and Proudfoot, N. J. (1974). Sequence analysis of immunoglobulin light chain messenger RNA. Nature *252*, 354–359.

O'Farrell, P. Z., Cordell, B., Valenzuela, P., Rutter, W. J. and Goodman, H. M. (1978). Structure and processing of yeast precursor tRNAs containing intervening sequences. Nature *274*, 438–445.

Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L. and Weissman, S. M. (1978). The genome of simian virus 40. Science *200*, 494–502.

Richardson, C. C. (1965). Phosphorylation of nucleic acid by an enzyme from T4 bacteriophage-infected Escherichia coli. Proc. Nat. Acad. Sci. USA *54*, 158–165.

Rosenberg, M., Court, D., Wulff, D. L., Shimatake, H. and Brady, C. (1978). The relationship between function and DNA sequence in an intercistronic regulatory region in phage λ. Nature *272*, 414–423.

Shulman, M. J. and Gottesman, M. E. (1973). Attachment site mutants of bacteriophage lambda. J. Mol. Biol. *81*, 461–482.

Smith, H. O. and Birnstiel, M. L. (1976). A simple method for DNA restriction site mapping. Nucl. Acids Res. *3*, 2387–2398.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503–517.

Stafford, D. W. and Bieber, D. (1975). Concentration of DNA solutions by extraction with 2-butanol. Biochim. Biophys. Acta *378*, 16–21.

Starlinger, P. and Saedler, H. (1976). IS-elements in microorganisms. Curr. Topics Microbiol. Immunol. *75*, 111–152.

Tilghman, S. M., Curtis, P. J., Tiemeier, D. C., Leder, P. and Weismann, C. (1978). The intervening sequence of a mouse β-globin gene is transcribed within the 15S β-globin mRNA precursor. Proc. Nat. Acad. Sci. USA *75*, 1309–1313.

Tonegawa, S. (1976). Reiteration frequency of immunoglobulin light chain genes: further evidence for somatic generation of antibody diversity. Proc. Nat. Acad. Sci. USA *73*, 203–207.

Tonegawa, S., Hozumi, N., Matthyssens, G. and Schuller, R. (1976). Somatic changes in the content and context of immunoglobulin genes. Cold Spring Harbor Symp. Quant. Biol. *41*, 877–889.

Tonegawa, S., Brack, C., Hozumi, N. and Schuller, R. (1977). Cloning of an immunoglobulin variable region gene from mouse embryo. Proc. Nat. Acad. Sci. USA *74*, 3518–3522.

Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. and Gilbert, W. (1978). Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. Proc. Nat. Acad. Sci. USA *75*, 1485–1489.

Weigert, M. G., Cesari, I. M., Yonkovich, S. J. and Cohn, M. (1970). Variability in the lambda light chain sequences of mouse antibody. Nature *228*, 1045–1047.

Wilson, G. A. and Young, F. E. (1975). Isolation of a sequence-specific endonuclease (Bam I) from Bacillus amyloliquefaciens H. J. Mol. Biol. *97*, 123–125.

## Note Added in Proof

Nobumichi Hozumi's present address is: The Ontario Cancer Institute, 500 Sherbourne Street, Toronto, Ontario, Canada.

**Folding of protein backbone structure**

In the same year that Hilschmann and Craig determined the first three amino acid sequences of Bence Jones proteins, i.e. 1965, the protein backbone structure was analyzed in detail for the first time (Ramakrishnan C, Ramachandran GN (1965) Stereochemical criteria for peptide and protein chain conformation. Biophys J, 5, 909-933). For every amino acid residue between two nearly planar peptide bonds, the rotation around the N-CA bond was designated as $\phi$ and that around the CA-N bond by $\phi'$ so that all possible configurations of the amino acid residue can be represented in a two dimensional plot of $\phi$ vs. $\phi'$ from $0^{\circ}$ to $360^{\circ}$. Over the years, this simple diagram has been referred to as the Ramachandran plot. The angle $\phi'$ has been replaced by $\psi$ and the range by $-180^{\circ}$ to $+180^{\circ}$. However, the diagram remains the same.

Assuming non-bonding atoms as hard spheres, certain combinations of the $\phi$ and $\psi$ angles are thus not possible due to steric hindrance. Except for glycine residue, most of the combinations are indeed not allowed. The most important atom involved in such studies is hydrogen not visible in protein three dimensional studies by X-ray diffraction.

Atomic coordinates of protein atoms from X-ray diffraction analysis can be used to calculate the $\phi$ and $\psi$ angles precisely. If that combination falls into regions of the Ramachandran plot with steric hindrance, the atomic coordinates may be in error. The major problem is that X-ray diffraction studies routinely ignore the presence of hydrogen atoms. Furthermore, the resolution is usually not good enough to position other atoms unambiguously.

Thus the incorporation of steric hindrance studies into X-ray diffraction analysis can greatly improve the accuracy of the three dimensional structures of proteins.

Conversely, using the bond distances and bond angles of a standard peptide unit, a given $\phi$ and $\psi$ angle combination can define the atomic coordinates precisely. Therefore, for a peptide segment, predicting the $\phi$ and $\psi$ angles of each amino acid residue in that segment can generate a detailed three dimensional structure for that segment.

# STEREOCHEMICAL CRITERIA FOR POLYPEPTIDE AND PROTEIN CHAIN CONFORMATIONS

## II. Allowed Conformations for a Pair of Peptide Units

C. RAMAKRISHNAN *and* G. N. RAMACHANDRAN

*From the Centre of Advanced Study in Biophysics, University of Madras, Madras, India, and the Biophysics Research Division, University of Michigan, Ann Arbor*

ABSTRACT The conformation of a polypeptide or protein chain may be specified by stating the orientations of the two linked peptide residues at each alpha carbon atom in the chain, namely the two dihedral angles $\phi$, $\phi'$ about the single bonds N—$\alpha$C and $\alpha$C—C′ from a defined standard conformation. By using certain criteria of minimum contact distances between the various atoms, the allowed anges of ($\phi$, $\phi'$) have been worked out for three values of the angle N-$\alpha$C-C′ ($\tau$), namely 105, 110, and 115° for non-glycyl, and 110 and 115° for glycyl residues. The theory is compared with all the available crystallographic data (up to early 1965) on simple (di- and tri-) peptides, cyclic peptides, polypeptide and protein structures, and the observed data fully support the conclusions from theory. The effect of the gamma carbon atom, in its three possible positions, is also discussed, and is found to alter the outer limits of the allowed region of ($\phi$, $\phi'$) only slightly. The paper contains exhaustive references to the published data on these structures, using x-ray diffraction.

### INTRODUCTION

The conformation of the backbone of a polypeptide or protein chain can be completely specified by giving the relative orientations of the two linked peptide groups at each $\alpha$-carbon atom. When the relative orientations of the peptide group are the same at every $\alpha$-carbon atom, then the chain takes up a regular helical folding. In Part I (Ramakrishnan, 1964) methods were described of calculating the parameters of such a helix, mainly the number of residues per turn, $n$, and the unit translation of the residue along the helix, $h$, from a knowledge of the geometry of the peptide group and the relative orientations of the linked peptide groups. However, not all possible orientations will be stereochemically allowed, because of the short contacts between the atoms of the adjacent residues. A study of this type has been made using certain definite stereochemical criteria and the preliminary results corresponding to the angle N-$\alpha$C-C′ ($\tau$) = 100° have already been published (Sasisekharan, 1962;

Ramachandran, Ramakrishnan, and Sasisekharan, 1963*a* and *b*). This paper describes the fuller details of the results for this value of the angle, $\tau$, and also for two other values of $\tau$, namely 105 and 115°. In addition, a complete survey of the structures of di- and tripeptides, cyclic peptides, and polypeptides has also been made to check the predicted ranges of allowed conformations. These results are discussed in addition to a study of the conformations which occur in myoglobin.

It may be mentioned that although studies of this type have been reported so far in the literature (Huggins, 1943; Bragg, Kendrew, and Perutz, 1950; Pauling, Corey, and Branson, 1951; Low and Baybutt, 1952; Low and Grenville-Wells 1953; Donohue, 1953; Lindley and Rollett, 1955; Shimanouchi and Mizushima, 1955; Scheraga, 1960), they have not been definitive. Mizushima and Shimanouchi (1961) have taken into account the restrictions or barriers due to the rotation about the two single bonds meeting at the $\alpha$-carbon atoms and have preferred nine conformations for the polypeptide chain. More recently, De Santis *et al.* (1965) have considered this problem from potential energy considerations, and Nemethy and Scheraga (1965) have obtained interesting results regarding extended chains and the formation of closed loops containing S-S bridges.

The notation used here for representing the relative orientation of the linked peptide groups is the same as that described in Part I (Ramakrishnan, 1964). The rotations of the two peptide groups are denoted by two dihedral angles ($\phi$, $\phi'$) respectively, these being measured from an initial standard conformation $\phi = \phi' = 0°$, in which the planes of the two peptide groups lie in the plane N-$\alpha$C-C′ containing the two axes of rotation, and the NH groups of the two residues are pointing towards each other.[1]

## DETERMINATION OF ALLOWED CONFORMATIONS

In order to determine the allowed conformations, the contact distances between the atoms in the adjacent residues have to be examined using criteria for minimum van der Waals contact distances. This is best done by first working out the positions of

[1] At a recent conference of some of the representative workers in this field held in Bethesda, it was decided to denote the two dihedral angles about the bonds N—$C_\alpha$ and $C_\alpha$—C′ by $\phi$ and $\psi$ respectively, the sense of rotation being the same as that adopted here. The fully extended chain, with N—H and C′=O *trans* with respect to one another, is to be taken as the standard conformation with $\phi = \psi = 0$.

It is readily seen that the new $\phi$ is the same as the old $\phi$, but that $\psi = 180° + \phi'$ (and $\phi' = 180° + \psi$). All the data reported here are thus readily converted into the ($\phi$, $\psi$) coordinates. The diagrams in Figs. 2, 3, and 6 have to be shifted up by 180°, or half the total length, along the vertical direction. In particular, the right- and left-handed alpha helices will have ($\phi$, $\psi$) equal to (133°, 123°) and (227°, 237°). As before, a helix with ($-\phi$, $-\psi$) will be inverse to one with ($\phi$, $\psi$); *i.e.*, it will be of opposite sense, but having the same number of turns per unit.

This paper was finalized well before this meeting, and so the older conventions are adopted here. However, it is proposed to use the new notation in the following papers in this series.



125

the atoms in the two residues for various values of $(\phi, \phi')$. Actually, the coordinates of the atoms in the two residues were calculated with respect to a suitably chosen fixed coordinate system, from which the various contact distances were calculated for the various values of $(\phi, \phi')$. The following fixed system of coordinates was found to be convenient for this purpose.

The $\alpha$-carbon atom $C_2$ at which the two residues [$C_1$—$C_1'O_1$—$N_1H_1$—$C_2$] and [$C_2$—$C_2'O_2$—$N_2H_2$—$C_3$] are linked together is taken to be the origin of coordinates of a system of axes $OXYZ$ defined as follows. The plane $N_1C_2C_2'$, which remains unchanged whatever be the rotation of the two groups, is taken as the $XY$ plane with the $Z$ axis upwards. The direction $C_1C_2$, which lies in the $XY$ plane for $\phi = 0°$, is taken as the $Y$ axis. The $Z$ axis is the third perpendicular direction, such that the $X$, $Y$, and $Z$ axes form a right-handed system of coordinates. This is shown in Fig. 1.

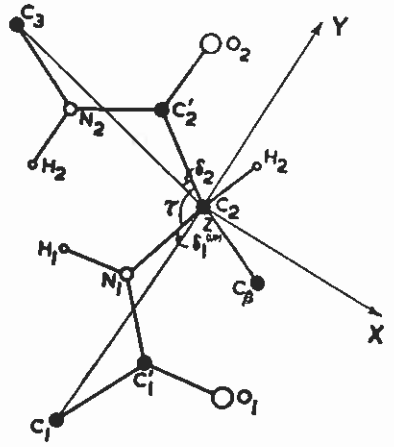



FIGURE 1 The two peptide groups in the initial conformation $\phi = \phi' = 0°$. The coordinate axes $X$, $Y$, and $Z$ used in the present study are also marked.

Throughout this study, the peptide group has been taken to be planar with NH and CO groups in the *trans* configuration and having the dimensions given by Corey and Pauling (1953).

If $x$, $y$, $z$ and $x'$, $y'$, $z'$ are the coordinates of an atom before and after rotation, the relations connecting these are given by (Whittaker, 1952)

$$\begin{aligned} x' &= (a^2 + b^2 - c^2 - d^2)x + 2(bc - ad)y + 2(bd + ac)z \\ y' &= 2(bc + ad)x + (a^2 - b^2 + c^2 - d^2)y + 2(cd - ab)z \\ z' &= 2(bd - ac)x + 2(cd + ab)y + (a^2 - b^2 - c^2 + d^2)z \end{aligned} \tag{1a}$$

where

$$a = \cos(\omega/2), \quad b = L\sin(\omega/2), \quad c = M\sin(\omega/2), \quad d = N\sin(\omega/2) \tag{1b}$$

$\omega$ being the angle of rotation and $L$, $M$, $N$, the direction cosines of the axis of rotation with respect to the chosen system of coordinates. The sign of $\omega$ is taken

to be positive, when the rotation is counterclockwise viewing along the axis towards the origin.

The angles of rotation and direction cosines of the axes of rotation corresponding to the two peptide groups about the axes $N_1$—$C_2$ and $C_2$—$C_2'$ are given in Table I.

The coordinates of the atoms $C_1$, $C_1'$, $O_1$, and $H_1$ were evaluated at intervals of 10° of $\phi$ and the coordinates of the atoms $O_2$, $N_2$, $H_2$, and $C_2$ were evaluated at intervals of 10° for $\phi'$, using the relations (1) for three values of the angle $\tau$, namely, 105, 110, and 115°. In fact, all the calculations corresponding to $\tau = 110$ and 115° were carried out using a desk calculator, while those corresponding to $\tau = 105°$ were carried out much later using an Elliott-803 electronic computer.

The next step towards working out the allowed conformations is to choose a set of permitted minimum contact distances between the different types of atoms. Only those conformations ($\phi$, $\phi'$) which do not have any of the contact distances less than these minimum values now become allowed. Two such sets are given in Table II, termed "normally allowed" and "outer limit" contact distances. These are

TABLE I

ANGLES OF ROTATION AND THE DIRECTION COSINES OF THE AXES OF ROTATION USED IN THE RELATIONS (1)

| Axis of rotation | Angle of rotation | Direction cosines of the axes of rotation | | |
|---|---|---|---|---|
| | | L | M | N |
| $N_1$—$C_2$ | $\phi$ | $-\sin \delta_1$ | $-\cos \delta_1$ | 0 |
| $C_2$—$C_2'$ | $\phi'$ | $-\sin \delta_1'$ | $\cos \delta_1'$ | 0 |

where $\delta_1 = C_1\hat{C}_2N_1$, $\delta' = 180° - (\tau + \delta_1)$ and $\tau = N_1\hat{C}_2C_2'$

TABLE II

MINIMUM CONTACT DISTANCES ASSUMED

| Contact | Normally allowed | Outer limit |
|---|---|---|
| | *A* | *A* |
| C ··· C | 3.20 | 3.00 |
| C' ··· C' | 2.95 | 2.90 |
| C ··· O | 2.80 | 2.70 |
| C ··· N | 2.90 | 2.80 |
| C ··· H | 2.40 | 2.20 |
| O ··· O | 2.70 | 2.60 |
| O ··· N | 2.70 | 2.60 |
| O ··· H | 2.40 | 2.20 |
| N ··· N | 2.70 | 2.60 |
| N ··· H | 2.40 | 2.20 |
| H ··· H | 2.00 | 1.90 |



127

essentially the same as those given by Ramachandran, Ramakrishnan, and Sasisekharan (1963) arrived at from a study of the contact distances observed in structures of various organic compounds. Contact distances in between the two limits have also been observed in actual crystals, but not as frequently as the normally allowed values.

In order that the results can be applied to actual polypeptide and protein structures, the contact distances of the side chain atoms with those of the backbone atoms must also be considered. However, in this study, only the contact distances of the backbone atoms with the $\beta$-carbon atom were considered (see Appendix for the effect of $C_\gamma$).

The position of the $\beta$-carbon atom was fixed (using a stereographic projection technique) in such a way that the distance $C_\alpha$—$C_\beta$ is 1.54 A and that the angles $N_1$-$C_\alpha$-$C_\beta$ and $C_2'$-$C_\alpha$-$C_\beta$ had tetrahedral values (119°30′). Of the two possible positions of $C_\beta$, the one corresponding to the L-type of residues has been used in all these studies and hence the results pertain to L-amino acid residues. However, for D-amino acids, the corresponding allowed conformations will be $(-\phi, -\phi')$, where $(\phi, \phi')$ refers to L-amino acids.

Since the positions of the atoms $N_1$, $C_2'$ (which lie on the axis of rotation) and $C_\beta$ do not change whatever the angle of rotation of the two residues is, only the following contact distances were calculated initially: (*a*) from $C_\beta$ and $C_2'$ to $C_1$, $C_1'$, $O_1$ and $H_1$ for values of $\phi$ from 0 to 360° at intervals of 10°; and (*b*) from $C_\beta$ and $N_1$, to $O_2$, $N_2$, $H_2$, and $C_3$ for values of $\phi'$ for 0 to 360° at intervals of 10°. These calculations were carried out corresponding to all three values of $\tau$, namely, 105, 110, and 115°. When the ranges of $\phi$ and $\phi'$ allowed by each contact, using the minimum contact distances given in Table III, are combined, the resulting ranges given in the first half (*a*) of Table III are obtained.

Within the permitted ranges of $\phi$ and $\phi'$ given in Table III, the contact distances between the hydrogen atom, H, attached to the $\alpha$-carbon atom and the atoms of the backbone were also examined and it was found that for all the three values of $\tau$, no further restrictions arise due to these contact distances.

When the remaining contact distances *between* the atoms occurring in the *backbones* of the *two* residues (the positions of which change during the rotations $\phi$ or $\phi'$) were examined in the permitted range given in Table III, contact distances less than the minimum allowed values were found to occur for some values of $(\phi, \phi')$ and thus some more conformations became disallowed on account of these short contacts. The resulting diagrams showing the regions of stereochemically allowed conformations are shown in Figs. 2*a*, *b*, and *c* corresponding to the three values of $\tau$. In all these figures, the normally allowed regions are shown by continuous lines, while the boundaries of the region allowed by the outer limit values are shown by broken lines.

In order that the results can be applied to the conformations occurring with glycyl

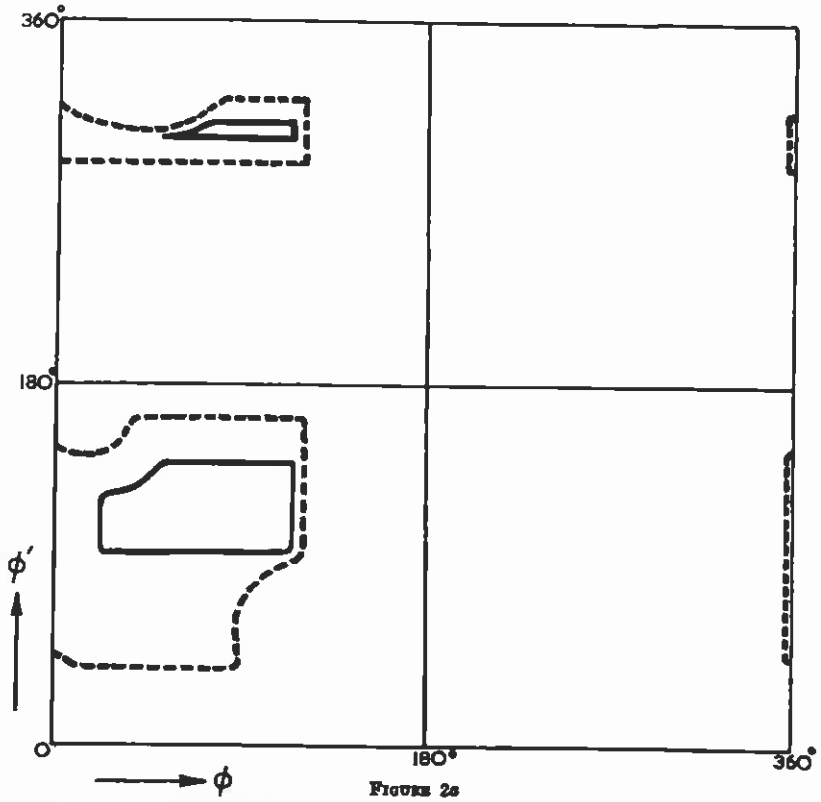



FIGURE 2a

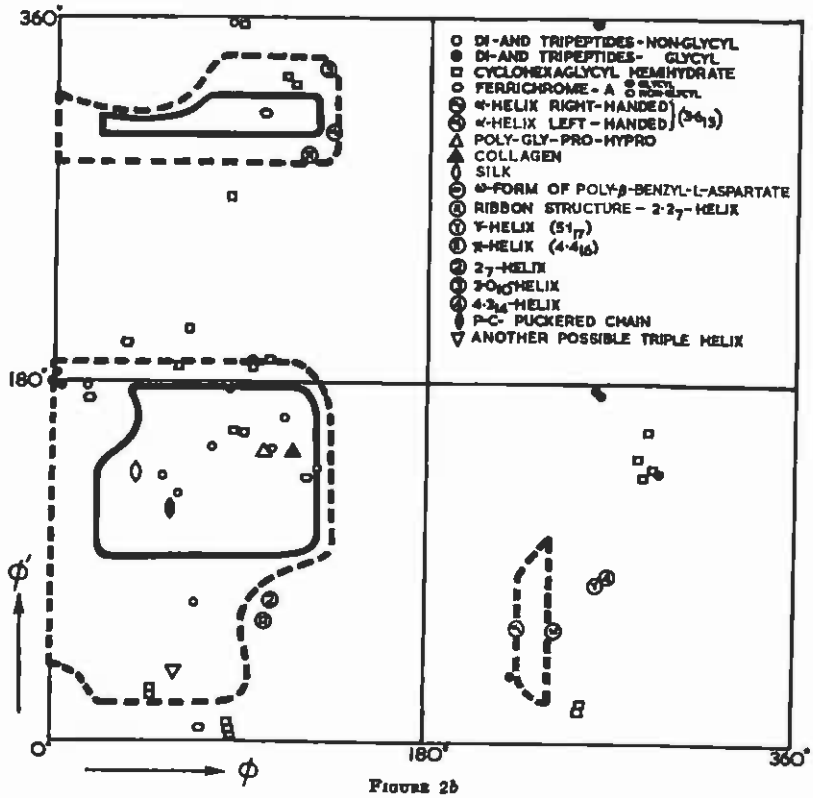



FIGURE 2b

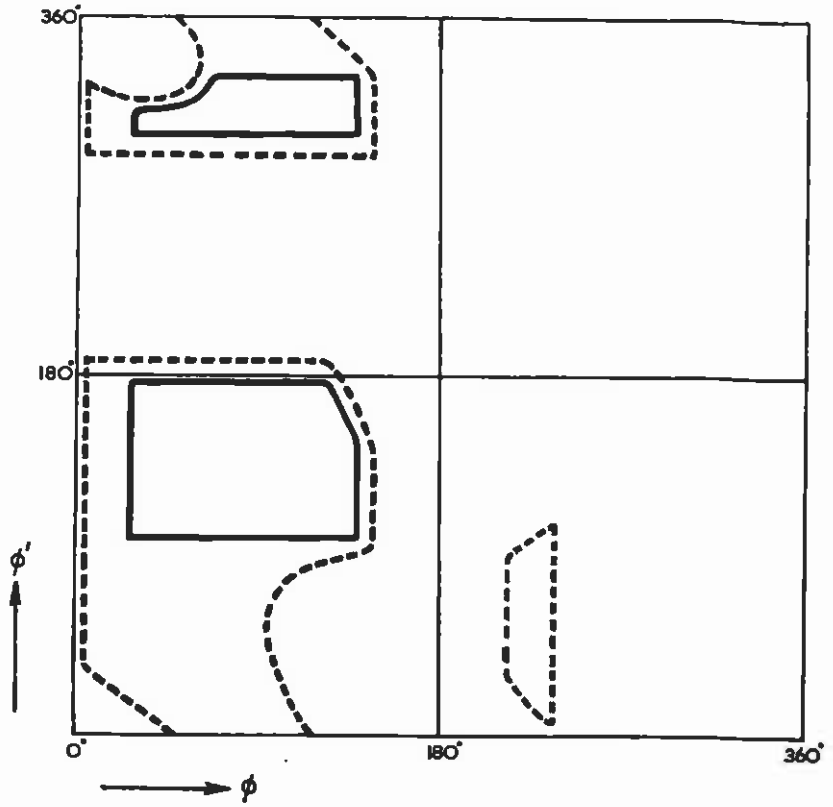



FIGURE 2c

FIGURE 2 The normally allowed (———) and the outer limit (– – – –) regions of $(\phi, \phi')$ for (*a*) $\tau = 105°$, (*b*) $\tau = 110°$, (*c*) $\tau = 115°$. The conformations of known simple peptides, polypeptides, and proteins are also marked in Fig. 2*b*.

residues (*i.e.* where there is no $\beta$-carbon atom), the contract distances between the hydrogen atoms attached to the $\alpha$-carbon atom and the various backbone atoms were evaluated for $\tau = 110$ and 115°. The two hydrogen atoms were fixed at tetrahedral directions to $N_1$—$C_2$ and $C_2$—$C_2'$ and at a distance of 1.0A from $C_2$. It was found that both for $\tau = 110$ and 115°, for all values of $\phi$ and $\phi'$, there were no short contacts between the hydrogen atoms attached to the $\alpha$-carbon atom and the backbone atoms. As a result, the permitted ranges of $\phi$ and $\phi'$ are larger for glycyl residues than for non-glycyl residues. The permitted ranges are given in the second half (*b*) of Table III. The regions of allowed conformations for glycyl residues when all the contact distances were taken into account are shown in Figs. 3*a* and *b*, where shaded regions are those allowed by only the outer limits, but not the normally allowed contact distances.

It may be mentioned that for glycyl residues the contact distances between any

TABLE III

RANGES OF $\phi$ AND $\phi'$ ALLOWED BY THE CONTACT DISTANCES BETWEEN THE ATOMS $N_1$, $C_3'$, $C_\beta$ AND THE REMAINING BACKBONE ATOMS IN THE TWO RESIDUES FOR NON-GLYCYL RESIDUES AND BETWEEN THE ATOMS $N_1$, C, $H_2'$, AND THE REMAINING BACKBONE ATOMS FOR GLYCYL RESIDUES

| | Normally allowed range | | Outer limit range | |
|---|---|---|---|---|
| | $\phi$ | $\phi'$ | $\phi$ | $\phi'$ |
| (*a*) Non-glycyl residues | | | | |
| 105° | 22 to 115° | 95 to 141° and 301 to 309° | 0 to 120° and 357 to 360° | 38 to 163° and 289 to 322° |
| 110° | 22 to 127° | 92 to 179° and 300 to 320° | 0 to 135° and 225 to 241° | 20 to 190° and 289 to 338° |
| 115° | 27 to 139° | 98 to 177° and 301 to 330° | 4 to 147° and 213 to 237° | 0 to 188° and 290 to 360° |
| (*b*) Glycyl residues | | | | |
| 110° | 0 to 127° and 233 to 360° | 40 to 320° | 0 to 135° and 225 to 360° | 22 to 338° |
| 115° | 0 to 139° and 221 to 360° | 30 to 330° | 0 to 147° and 213 to 360° | No restriction |

two atoms of the backbone in the adjacent residues for any conformation ($\phi$, $\phi'$) is the same as for the inverse conformation ($-\phi$, $-\phi'$) [as can be seen from the relations (1), where, if $x'$, $y'$, and $z'$ are the coordinates of a backbone atom for $\phi$ (or $\phi'$) then the coordinates of the same atom for $-\phi$ (or $-\phi'$) are $x'$, $y'$, and $-z'$].

The results shown in Figs. 2 and 3 will be useful in investigating the conformation at an $\alpha$-carbon atom when the two dihedral angles $\phi$ and $\phi'$ are known. For polypeptide and protein chains, the parameters that are obtainable from X-ray diffraction studies are the number of residues per turn, $n$, and the unit translation, $h$, (Ramachandran, 1960) along the axis of the helix. Hence, the stereochemically allowed combinations of $n$ and $h$ were evaluated using the detailed results of the calculations described in Part I (Ramakrishnan, 1964) and are represented graphically in Figs. 4*a*, *b*, and *c* for the three values of $\tau$. In these figures also, the regions of ($n$, $h$) allowed by the normally allowed contact distances are shown by continuous line boundaries, while those allowed by outer limit values are shown by broken lines. The origins in these figures correspond to $n = \pm 2$, and $h = 0$ A. The positive and



131

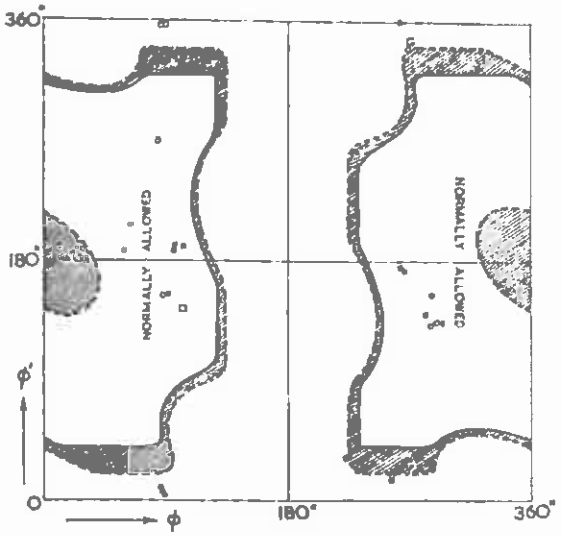



FIGURE 3*a*

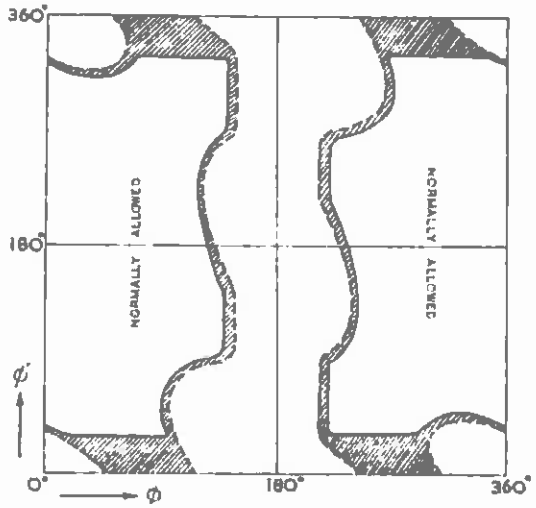



FIGURE 3*b*

FIGURE 3 Allowed regions of $(\phi, \phi')$ for (*a*) $\tau = 110°$, (*b*) $\tau = 115°$, for glycyl residues. The shaded regions represent conformations which are allowed only by the outer limit contact distances. The glycyl conformations observed in simple peptides are also marked in Fig. 3*a*: ● = Di- and Tripeptides; ○ = Cyclic peptides; □ = Polyglycine II.

negative directions of the $X$ axes correspond respectively to positive and negative values of $n$, but $h$ is always taken to be positive. In other words, plots corresponding to all right-handed helices lie to the right of the $Y$ axis and similarly points representing all left-handed helices lie to the left of the $Y$ axis. In Fig. 4*b*, which corresponds to $\tau = 110°$, the various observed and postulated polypeptide and protein structures are also plotted.

## DISCUSSION OF THE RESULTS

From Figs. 2*a*, *b*, and *c*, it can be seen that the general nature of the regions of allowed conformations ($\phi$, $\phi'$) is the same for the three values of $\tau$. An interesting feature is that the region around (220°, 60°), where the left-handed $\alpha$-helix occurs, is not allowed for $\tau = 105°$, while it is allowed for $\tau = 110$ and 115°.

In order to have a quantitative estimate of the results, the area under the allowed regions may be expressed as a percentage of the total area of the plane $\phi$—$\phi'$. These are given in Table IV. There is an appreciable increase in the range of allowed regions with increasing $\tau$. Also the allowed region is much larger for glycyl residues than when a $\beta$-carbon atom is present (about 6 times for normally allowed and 2.5 times for outer limit for $\tau = 110°$). Thus there is more freedom of rotation about the single bonds when the amino acid residue at the $\alpha$-carbon atom is glycine than when it contains a $\beta$-carbon atom.

Since the $\alpha$-carbon atom is not an asymmetric carbon atom in the case of glycine, no D or L isomer is possible. This is reflected clearly in Figs. 3*a* and *b*, which have the symmetry of a center of inversion about (180°, 180°).

In order to verify how far the above results from theory agree with observation, the parameters $\phi$ and $\phi'$ were evaluated from the available crystal structure data on di- and tripeptides, cyclic peptides, and polypeptides. These were calculated on an IBM 1620 computer. In addition, the two values of $\phi'$ ($\phi_1'$ and $\phi_2'$), which alone are relevant in the case of amino acid structures, were also evaluated.

TABLE IV

AREA OF ALLOWED CONFORMATIONS EXPRESSED AS PERCENTAGE OF THE TOTAL AREA OF THE ($\phi - \phi'$) PLANE

| $\tau$ | Normally allowed | Outer limit |
|---|---|---|
| | *per cent* | *per cent* |
| (*a*) Non-glycyl residues | | |
| 105° | 3.5 | 12.7 |
| 110° | 7.7 | 20.3 |
| 115° | 8.8 | 24.2 |
| (*b*) Glycyl residues | | |
| 110° | 45.1 | 57.4 |
| 115° | 56.7 | 66.8 |



133

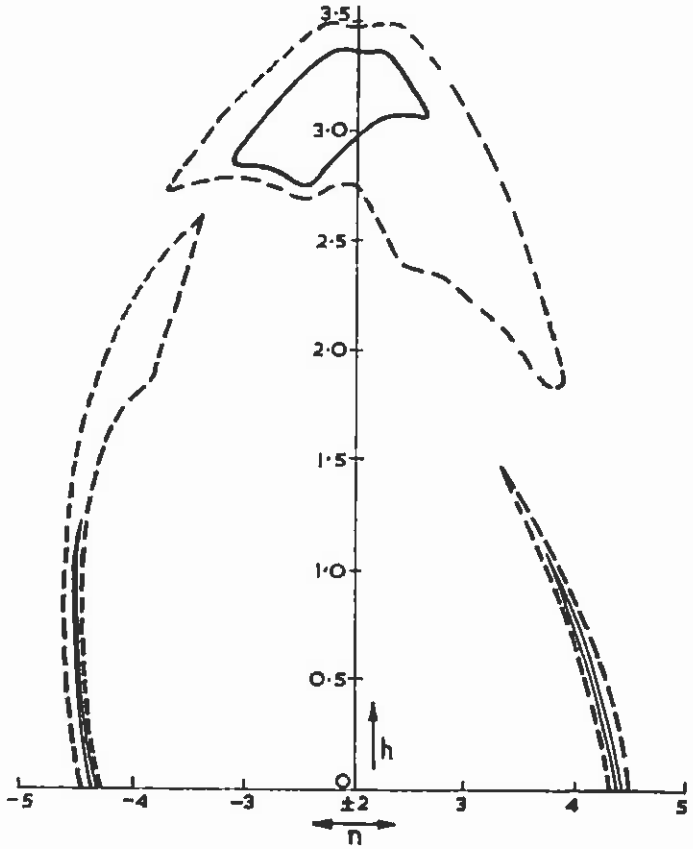



FIGURE 4*a*

*Amino Acids.* The two values $\phi_1'$, $\phi_2'$ for the various known crystal structures of amino acids are given in Table V along with the values of the angle N-$\alpha$C-C′ ($\tau$). In general, $\phi'$ lies between ± 30° and 180° ± 30°, exceptions being L-arginine HCl (Mazumdar, 1964), DL-aspartic acid (Dawson, 1953), L-glutamic acid (Hirokawa, 1955), L-tyrosine HCl and L-tyrosine HBr (Srinivasan, 1959*a* and *b*).

*Di- and Tripeptides.* In this case, for the N-terminal residues, $\phi$ is the only relevant parameter, since the positions of the hydrogen atoms attached to $N_0$ are not usually available. For the C-terminal residues, there is one value of $\phi$ and two values of $\phi'$ corresponding to the two oxygen atoms of the $C_0^{O^-}$ group. For the other residues (middle), there is one value of $\phi$ and one value of $\phi'$. These are shown in Fig. 5, where the $\alpha$-carbon atom under consideration is $C_1$. The results are given in Table VI along with the value of the angle N-$\alpha$C-C′ ($\tau$).

Considering the N-terminal residues, for which only $\phi'$ is relevant, the observed

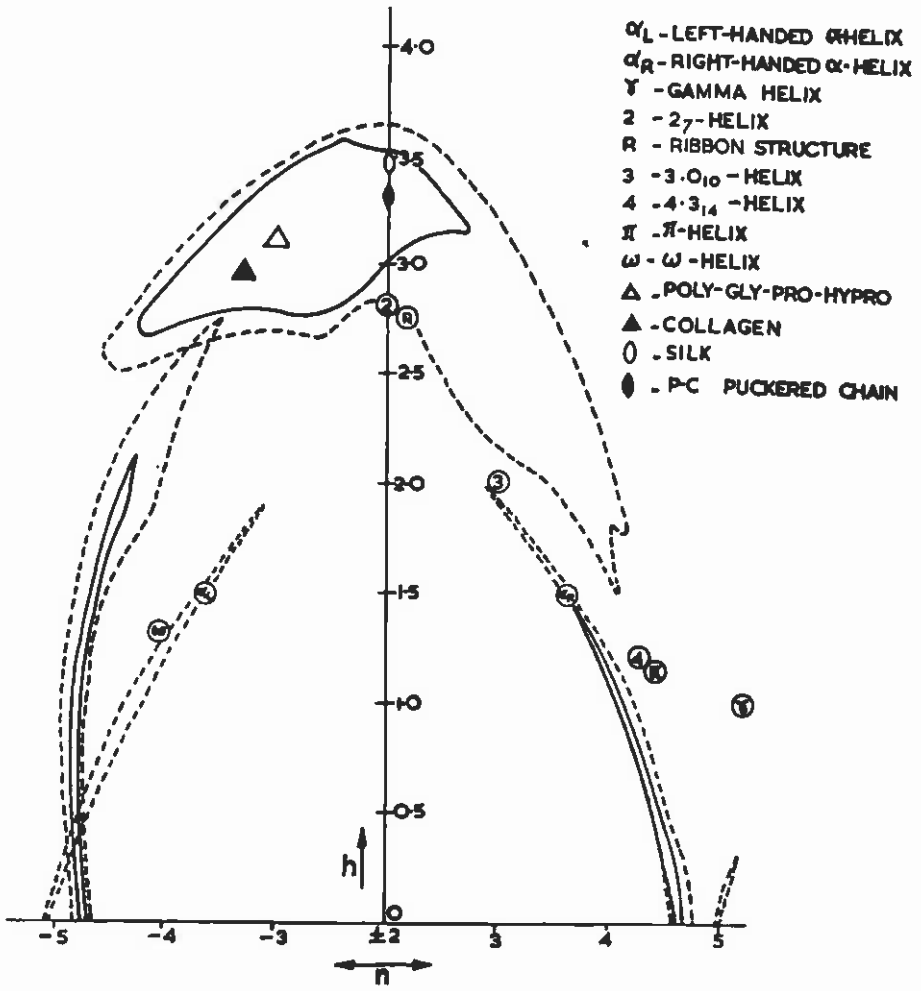



FIGURE 4*b*

values of $\phi'$ lie well within the allowed region mentioned in Table III both in the case of glycyl and non-glycyl residues. In addition, $\phi'$, in general, lies between $180° \pm 30°$ agreeing well with the trend observed in amino acid structures.

In the case of C-terminal and middle residues, the conformations are plotted in Fig. 2*b*, which also contains the allowed regions of $(\phi, \phi')$ for $\tau = 110°$. For better comparison the conformations at the α-carbon atoms corresponding to glycyl residues are also plotted in Fig. 3*a* which gives the allowed regions of $(\phi, \phi')$ for such residues for $\tau = 110°$.

Both for glycyl and non-glycyl residues, almost all the conformations lie within the allowed region. In the case of non-glycyl residues it is very interesting to note



135

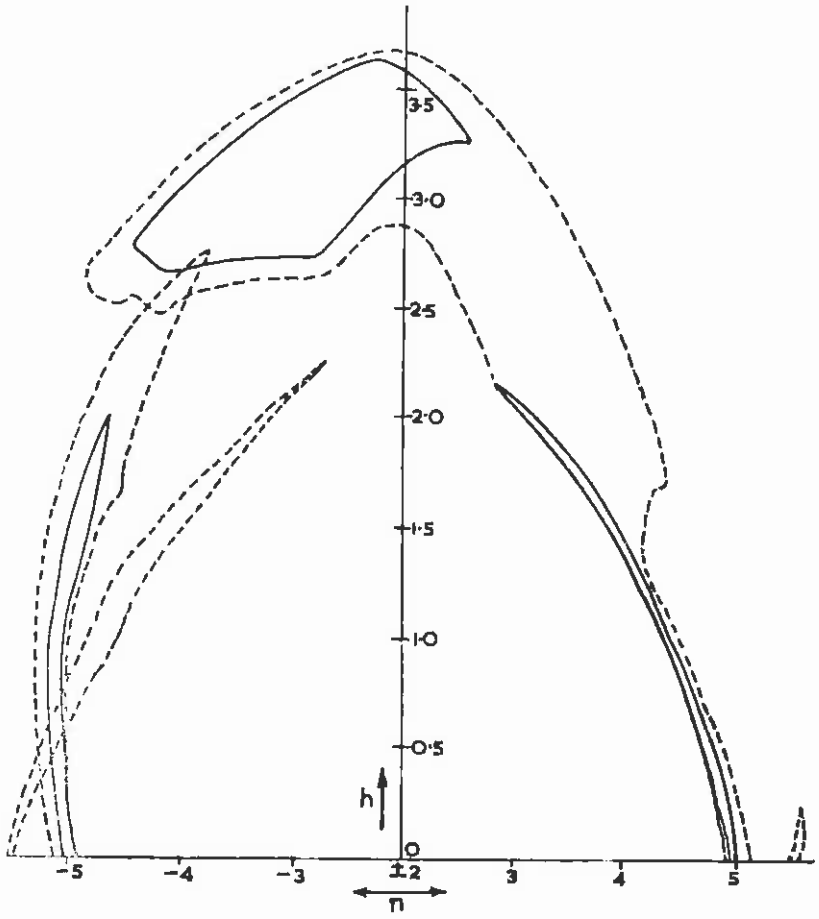



FIGURE 4c

FIGURE 4 The normally allowed (———) and the other limit (– – – –) regions of $(n, h)$ for $(a)$ $\tau = 105°$, $(b)$ $\tau = 110°$, $(c)$ $\tau = 115°$. The known polypeptide and protein structures are also marked in Fig. 4$b$—$\alpha_L$ = left-handed $\alpha$-helix; $\alpha_R$ = right-handed $\alpha$-helix; (7) = gamma helix; (2) = $2_7$-helix; (R) = ribbon structure; (3) = $3.0_{10}$-helix; (4) = $4.3_{14}$-helix; ($\pi$) = $\pi$-helix; ($\omega$) = $\omega$-helix; Δ = poly-gly, -pro, -hypro; ▲ = collagen; ○ = silk; ● = P-C puckered chain.

that six out of nine conformations listed in Table VI lie well within the normally allowed region and two more lie within the outer limit but outside the normally allowed region. The one conformation that lies outside the allowed region is that of the cysteine residue of the compound glutathione. But the value of $\tau$ in this case is 115° and so the conformation comes into the allowed region corresponding to $\tau = 115°$ (Fig. 2$c$). In addition, the value of $\phi'$ in many cases lies close to 150°, which agrees well with the trend observed in amino acids.

TABLE V

OBSERVED VALUES OF $\phi'$ IN AMINO ACIDS. THE VALUE OF THE ANGLE N-$\alpha$C-C' ($\tau$) IS ALSO GIVEN

| Amino acid | $\tau$ | $\phi_1'$ | $\phi_2'$ | Reference |
|---|---|---|---|---|
| | *degrees* | *degrees* | *degrees* | |
| DL-Alanine | 108.4 | 162.9 | −15.3 | 1 |
| L-Arginine (in L-arginine dihydrate) | 110.9 | −169.3 | 12.1 | 2 |
| *L-Arginine (in L-arginine HBr, $H_2O$) | 109.5 | −173.6 | 2.8 | 3 |
| L-Arginine (in L-arginine HBr, $H_2O$) | 110.2 | −151.2 | 23.9 | 3 |
| *L-Arginine (in L-arginine HCl, $H_2O$) | 107.8 | −174.2 | 1.7 | 3 |
| L-Arginine (in L-arginine HCl, $H_2O$) | 109.5 | −154.3 | 22.6 | 3 |
| *L-Arginine (in L-arginine HCl) | 109.4 | 129.2 | −45.7 | 3 |
| L-Arginine (in L-arginine HCl) | 108.0 | 139.6 | −41.7 | 3 |
| Asparagine (in asparagine monohydrate) | 110.1 | −168.8 | 6.1 | 4 |
| DL-Aspartic acid (in DL-aspartic acid HCl) | 109.7 | −137.8 | 42.4 | 5 |
| L-Cysteine (in L-cysteine HCl) | 105.2 | 169.1 | −5.1 | 6 |
| L-Cysteine (in S-methyl-L-cysteine sulphoxide) | 111.1 | 176.7 | −9.8 | 7 |
| L-Cysteine (in hexagonal L-cysteine) | 108.4 | −165.8 | 12.3 | 8 |
| L-Cystine (in L-cystine dihydrobromide) | 108.1 | −179.2 | 0.9 | 9 |
| L-Glutamic acid | 108.3 | 134.2 | −31.9 | 10 |
| DL-Glutamic acid (in DL-glutamic acid HCl) | 110.7 | −162.4 | 18.0 | 11 |
| L-Glutamine (in glutathione) | 110.6 | 170.5 | −12.5 | 12 |
| L-Glutamine | 110.8 | 160.5 | −11.4 | 13 |
| Glycine (in bis glycino copper monohydrate) | 111.3 | 171.7 | −6.1 | 14 |
| $\alpha$-Glycine | 111.8 | −160.9 | 18.4 | 15 |
| $\beta$-Glycine | 110.8 | −152.7 | 23.8 | 16 |
| $\gamma$-Glycine | 110.8 | −165.0 | 11.7 | 17 |
| Histidine (in dihistidino zinc pentahydrate) | 110.8 | 176.4 | −3.9 | 18 |
| Histidine (in dihistidino zinc dihydrate) | 109.0 | −168.2 | 9.1 | 19 |
| Histidine (in histidine HCl, $H_2O$) | 109.4 | 180.0 | 0.0 | 20 |
| Hydroxy-L-proline II | 110.7 | −177.0 | 2.0 | 21 |
| D-Isoleucine (in D-isoleucine HBr) | 108.5 | 165.7 | −20.6 | 22 |
| L-Leucine (in L-leucine HBr) | 110.1 | 160.4 | −12.3 | 23 |
| L-lysine (in L-lysine HCl) | 109.7 | 160.1 | −17.2 | 24 |
| $\alpha$-Methionine | 112.1 | 146.3 | −30.4 | 25 |
| $\beta$-Methionine | 110.0 | 150.6 | −31.6 | 25 |
| DL-Norleucine | 109.0 | 144.6 | −24.4 | 26 |
| Proline (in copper proline dihydrate) | 109.7 | −169.0 | 11.0 | 27 |
| DL-Serine | 110.1 | −176.3 | 1.2 | 28 |
| L-Threonine | 110.4 | 153.9 | −28.8 | 29 |
| Tryptophan | 111.9 | 175.8 | 21.4 | 30 |
| L-Tyrosine (in L-tyrosine HCl) | 110.1 | 142.1 | −25.4 | 31 |
| L-Tyrosine (in L-tyrosine HBr) | 108.4 | 143.9 | −30.4 | 32 |
| L-Valine (in L-valine HBr) | 108.1 | 168.4 | −14.7 | 33 |
| L-Valine (in L-valine HCl) | 105.9 | 172.2 | −8.9 | 34 |
| L-Valine (in L-valine HCl, $H_2O$) | 108.2 | −173.1 | 5.3 | 35 |

*In the structures of L-arginine HBr, L-arginine HCl, $H_2O$, and L-arginine HCl, there are two molecules per asymmetric unit.

References: (1) Donohue (1951); (2) Karle and Karle (1964); (3) Work done in Madras, Mazumdar (1964); (4) Kartha and De Vries (1961); (5) Dr. B. Dawson, private communication; (6) Work done in

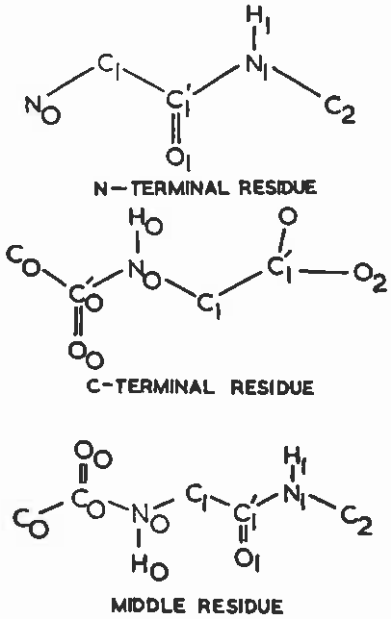



FIGURE 5 The environment around the α-carbon atom $C_1$ in the case of N-terminal, C-terminal, and middle residues of peptides.

Considering the glycyl residues, almost all of them lie within the allowed region. Also in the case of C-terminal glycyl residues of the compounds Cys-Gly NaI and Gly-Phe-Gly and in the case of middle and C-terminal residues in Gly-Gly-Gly copper chloride, the observed conformations are such that they are allowed only for glycyl residues.

*Cyclic Peptides.* Recently, the crystal structures of two cyclic hexapeptides, cyclohexaglycyl hemihydrate (Karle and Karle, 1963), and Ferrichrome A (Zalkin, Foster, and Templeton, 1964) have been reported. Of these, the former contains four molecules per asymmetric unit and offers twenty four glycyl conformations for investigation. The latter hexapeptide (Ferrichrome A) with the sequence Gly-Ser-Ser-Orn-Orn-Orn (Orn standing for ornithine) gives six conformations, one glycyl and five non-glycyl. The parameters ($\phi$, $\phi'$) for these two hexapeptides are given in

---

Madras; (7) Hine (1962); (8) Oughton and Harrison (1959); (9) Peterson *et al.* (1960); (10) Hirokawa (1955); (11) Dawson (1953); (12) Wright (1958); (13) Cochran and Penfold (1952); (14) Freeman *et al.* (1964*b*); (15) Marsh (1958); (16) Itaka (1960); (17) Itaka (1961); (18) Harding and Cole (1963); (19) Kretsinger and Cotton (1963); (20) Donohue and Caron (1964); (21) Donohue and Trueblood (1952); (22) Trommel and Bijvoet (1954); (23) Subramanian (1965); (24) Wright and Marsh (1962); (25) Mathieson (1952); (26) Mathieson (1953); (27) Mathieson and Welsh (1952); (28) Shoemaker *et al.* (1953); (29) Shoemaker *et al.* (1951); (30) Work done in Madras; (31) Srinivasan (1959*a*); (32) Srinivasan (1959*b*); (33) Parthasarathy (1962); (34) Work done in Madras; (35) Work done in Madras.

TABLE VI

THE CONFORMATIONS ($\phi$, $\phi'$) OBSERVED IN DI- AND TRIPEPTIDES. THE ANGLE N-$\alpha$C-C′ AT THE CORRESPONDING $\alpha$-CARBON ATOMS ARE ALSO LISTED

| Peptides | | $\tau$ | $\phi$ | $\phi_1'$ | $\phi_2'$ | Reference* |
|---|---|---|---|---|---|---|
| | | *degrees* | *degrees* | *degrees* | *degrees* | |
| N-acetyl glycine | | 110.4 | 1.5 | −176.4 | 3.3 | 1 |
| (*a*) Dipeptides | | | | | | |
| β-Gly-Gly | N-term-Gly | 110.0 | — | −155.4 | — | 2 |
| | C-term-Gly | 110.7 | −0.8 | −180.0 | 4.7 | 2 |
| Gly-L-Asp | N-term-Gly | 111.2 | — | −171.4 | — | 3 |
| | C-term-Asp | 109.5 | 69.0 | −116.7 | 69.2 | 3 |
| Gly-L-Tyr | N-term-Gly | 112.1 | — | 173.6 | — | 4 |
| | C-term-Tyr | 111.4 | 76.6 | 146.6 | −30.9 | 4 |
| Gly-L-Tyr | N-term-Gly | 112.1 | — | 157.4 | — | 5 |
| | C-term-Tyr | 112.0 | 106.3 | 146.7 | −23.7 | 5 |
| N-N′ diglycyl cystine | N-term-Gly | 109.0 | — | −137.1 | — | 6 |
| | C-term-Cys | 109.5 | 16.5 | 178.8 | −3.7 | 6 |
| Cys-Gly sodium iodide | N-term-Cys | 111.5 | — | 171.7 | — | 7 |
| | C-term-Gly | 125.7 | −136.8 | −147.6 | 33.1 | 7 |
| Glutathione (γ-Glu-Cys-Gly) | Middle-Cys | 115.0 | 89.3 | −2.8 | — | 8 |
| | C-term-Gly | 109.5 | 95.5 | −169.1 | 10.0 | 8 |
| *p*-tosyl-L-Pro-L-hypro | N-term-Pro | 106.4 | — | 163.8 | — | 9 |
| | C-term-Hypro | 117.8 | 128.5 | 136.6 | −43.6 | 9 |
| L-Thr-L-Phe | N-term-Thr | 106.4 | — | 140.9 | — | 10 |
| | C-term-Phe | 107.8 | 61.4 | 123.5 | −67.0 | 10 |
| L-Leu-Gly | N-term-Leu | 109.1 | — | −144.0 | — | 11 |
| | C-term-Gly | 112.3 | 86.4 | 176.6 | 2.9 | 11 |
| (*b*) Tripeptides | | | | | | |
| Gly-Gly-Gly-$CuCl_2$ | N-term-Gly | 107.9 | — | 169.4 | — | 12 |
| | Middle-Gly | 111.1 | −65.6 | 133.4 | — | 12 |
| | C-term-Gly | 110.7 | −94.0 | 173.3 | −4.7 | 12 |
| Gly-L-Phe-Gly | N-term-Gly | 109.9 | — | 126.8 | — | 13 |
| | Middle-Phe | 107.0 | 52.9 | 132.7 | — | 13 |
| | C-term-Gly | 114.7 | −96.5 | 175.4 | −0.9 | 13 |
| L-Leu-L-Pro-Gly | N-term-Leu | 107.9 | — | 152.7 | — | 14 |
| | Middle-Pro | 111.3 | 112.2 | 162.5 | — | 14 |
| | C-term-Gly | 114.5 | 3.7 | 178.1 | 0.5 | 14 |

*(1) Donohue and Marsh (1962); (2) Hughes and Moore (1949); (3) Pasternak *et al.* (1954); (4) Smith and Wiebenga (1953); (5) Pasternak (1956); (6) Yakel and Hughes (1954); (7) Dyer (1951); (8) Wright (1958); (9) Fridrichsons and Mathieson (1962); (10) Work done in Madras; (11) Work done in Madras; (12) Freeman *et al.* (1964*a*); (13) Marsh and Glusker (1961); (14) Leung and Marsh (1958).



139

Table VII[a] and are plotted in Fig. 2*b*. The glycyl conformations are also plotted in Fig. 3*a*.

In the case of cyclohexaglycyl hemihydrate, a majority of the conformations lie within the allowed regions. However, a few of them (1, 5, 7, 10, 12, 13, 15) are clustered around 0° (or 360°) of $\phi'$. Although these conformations are not allowed for $\tau = 110°$, they do become allowed for higher values of $\tau$ (see Fig. 3*b*). It can be

TABLE VII

OBSERVED CONFORMATIONS ($\phi$, $\phi'$) IN CYCLOHEXAGLYCYL HEMIHYDRATE AND IN FERRICHROME A

| | Residue | $\tau$ | $\phi$ | $\phi'$ |
|---|---|---|---|---|
| | | *degrees* | *degrees* | *degrees* |
| Cyclohexaglycyl hemihydrate* | | | | |
| | Gly (1) | 112.8 | 85.6 | 9.9 |
| | Gly (2) | 111.2 | 111.6 | −29.2 |
| | Gly (3) | 111.7 | 85.4 | −88.9 |
| | Gly (4) | 112.6 | 112.7 | −29.3 |
| | Gly (5) | 110.8 | 86.6 | 6.9 |
| | Gly (6) | 113.0 | 115.6 | −33.0 |
| | Gly (7) | 112.1 | 86.8 | 5.8 |
| | Gly (8) | 111.5 | 114.5 | −32.6 |
| | Gly (9) | 110.2 | 49.4 | 24.3 |
| | Gly (10) | 116.2 | 89.7 | −3.8 |
| | Gly (11) | 110.0 | 48.2 | 26.3 |
| | Gly (12) | 115.4 | 86.9 | −1.6 |
| | Gly (13) | 114.9 | −103.1 | 16.9 |
| | Gly (14) | 102.4 | 87.3 | 155.1 |
| | Gly (15) | 113.7 | −102.9 | 18.1 |
| | Gly (16) | 105.3 | −71.8 | 155.9 |
| | Gly (17) | 105.3 | −71.8 | 154.9 |
| | Gly (18) | 109.5 | −76.4 | 141.0 |
| | Gly (19) | 109.7 | −72.3 | 133.3 |
| | Gly (20) | 105.3 | −68.2 | 135.3 |
| | Gly (21) | 110.2 | 65.0 | −153.3 |
| | Gly (22) | 108.3 | 60.1 | −171.8 |
| | Gly (23) | 110.0 | 95.7 | −171.1 |
| | Gly (24) | 109.5 | 104.1 | −169.2 |
| Ferrichrome A‡ | | | | |
| | Gly | 114.7 | −98.1 | −1.2 |
| | Ser 1 | 114.2 | 123.7 | 132.3 |
| | Ser 2 | 104.0 | 17.3 | 171.5 |
| | Orn 1 | 113.5 | 71.6 | 6.6 |
| | Orn 2 | 113.2 | 101.4 | −47.2 |
| | Orn 3 | 106.1 | 34.8 | −160.7 |

*Karle and Karle (1963); marked by the symbol, □, in Fig. 2*b*.
‡Zalkin *et al.* (1963); marked by the open hexagon in Fig. 2*b*.

[a] The coordinates of the atoms in the cyclic hexapeptide Ferrichrome A has kindly been furnished by Dr. Zalkin to us prior to publication.

seen from Table VII that for all these conformations, the value of $\tau$ is systematically greater than 110° and thus these conformations also become allowed, under the short contact criteria adopted in this paper.

In the case of the cyclic peptide Ferrichrome A, the conformations corresponding to Ser 1, Ser 2, and Orn 1 lie well within the allowed region. Though the conformation corresponding to Orn 2 apparently lies outside the allowed region, it is allowed on account of the higher value of $\tau$ (114°) at the $\alpha$-carbon atom. The corresponding Orn 3 has a short contact $C_\beta \ldots N_2$ according to the present study, but in the actual structure this has a value of 2.96 A, which is greater than the minimum value given in Table II. Thus this conformation is also allowed. The conformation corresponding to the glycyl residue has a value of 115° for $\tau$ and hence becomes allowed (Fig. 3*b*). An interesting feature of this conformation is that it is allowed only for a glycyl residue. The value of $\phi'$ for this conformation lies close to 360° which is also the case with many conformations in the previous cyclic peptide.

*Peptides in the Non-Helical Regions of Myoglobin.* The structure of the crystalline protein myoglobin contains long segments of $\alpha$-helices linked by non-helical peptide segments (Kendrew *et al.* 1960, 1961). The conformations at the $\alpha$-carbon atoms in these non-helical regions have been plotted by Dr. H. C. Watson of Medical Research Council, Cambridge, England, who has kindly made the plot available to the authors prior to publication. Most of the conformations lie within the allowed regions for $\tau = 110°$, with a clustering of conformations around that of the right-handed $\alpha$-helix. A few are allowed only for $\tau > 110°$. In fact, two conformations lie far outside the allowed limits, namely those of the residues EF3 and HC2. On examination, it is found that the short contact responsible for disallowing those two involve the $\beta$-carbon atom, so that they will become allowed if they were glycyl residues. When this fact was communicated to Dr. Watson, it was learned that these were in fact glycyl residues, as expected. The others that lie just outside the allowed regions have $\phi'$ close to 0° (or 360°) and hence are likely to have a value of $\tau$ greater than 110°, in which case they also come within the allowed region.[3]

*Polypeptides and Proteins.* The conformations of the various structures proposed for polypeptides and fibrous proteins have been calculated from the reported coordinates of the atoms and are given in Table VIII along with the data on the number of residues per turn, $n$, and the unit translation, $h$, along the axis of the axis of the helix. These are plotted in Fig. 2*b* and also in Fig. 4*b*, which gives the allowed combinations of $(n, h)$ for $\tau = 110°$.

All the structures except four ($\gamma$-helix, $4.3_{14}$-helix, $2_7$-helix, and $2.2_7$-helix) lie within the allowed regions, some well within the normally allowed regions and some others only within the outer limit region. These structures are discussed below.

[3] The $\phi - \phi'$ diagram for myoglobin will be published shortly by Dr. Kendrew and Dr. Watson (private communication).



141

TABLE VIII

OBSERVED AND POSTULATED CONFORMATIONS ($\phi$, $\phi'$) FOR POLYPEPTIDE AND PROTEIN STRUCTURES. THE PARAMETERS $n$ AND $h$ ARE ALSO LISTED

| Polypeptide or protein | $\tau$ | $\phi$ | $\phi'$ | $n$ | $h$ | Reference* |
|---|---|---|---|---|---|---|
| | *degrees* | *degrees* | *degrees* | | *A* | |
| $\alpha$-helix ($3.6_{13}$) | 109.5 | 133.0 | −57.2 | 3.615 | 1.495 | 1 |
| $\gamma$-helix ($5.1_{17}$) | 110.1 | −96.3 | 78.0 | 5.143 | 0.98 | 2 |
| $2_7$a-helix | 111.3 | 105.1 | 69.5 | 2.000 | 2.80 | 1 |
| $2.2_7$-helix (ribbon structure) | 111.6 | 101.9 | 59.2 | 2.169 | 2.75 | 3 |
| $3.0_{10}$-helix | 111.6 | 130.7 | −25.7 | 3.000 | 2.00 | 3 |
| $4.3_{14}$-helix | 100.5 | −91.9 | 91.7 | 4.337 | 1.20 | 3 |
| $\pi$-helix ($4.4_{16}$) | 114.9 | 122.9 | −69.7 | 4.40 | 1.15 | 4 |
| Polyglycine II | 109.1 | 102.0 | 145.8 | −3.00 | 3.10 | 5 |
| Poly-L-proline II | 110.0 | 102.8 | 145.9 | −3.00 | 3.12 | 6 |
| Poly-L-proline II | 108.8 | 103.7 | 145.1 | −3.00 | 3.12 | 7 |
| Poly-L-hydroxyproline A | 105.5 | 103.1 | 147.6 | −3.00 | 3.05 | 8 |
| $\omega$-form of poly-$\beta$-benzyl-L-aspartate | 109.9 | −115.6 | 55.4 | −4.00 | 1.325 | 9 |
| Collagen type helix | 110 | 116 | 145 | −3.28 | 2.95 | 10 |
| Silk | 110 | 40 | 135 | 2.00 | 3.45 | 11 |
| P-C puckered chain | 110 | 57 | 118 | 2.00 | 3.3 | 12 |

*(1) Bamford *et al.* (1956); (2) Pauling and Corey (1951); (3) Donohue (1953); (4) Low and Grenville-Wells (1953); (5) Crick and Rich (1955); (6) Sasisekharan (1959*a*); (7) Cowan and McGavin (1955); (8) Sasisekharan (1959*b*); (9) Bradbury *et al.* (1962); (10) Ramachandran (1963); (11) Marsh *et al.* (1955); (12) Pauling and Corey (1953).

*α-helix.* Both the right-handed α-helix (132°, -57°) and the left-handed α-helix (-132°, 57°) lie within the outer limit regions only. However, slight perturbations such as changing the value of $\tau$ would bring the right-handed helix into the normally allowed region but the left-hand α-helix is always outside the normal limits although it is just allowed by the outer limits. Also the right-handed α-helix can accommodate L-proline towards an end, since $\phi' \approx 120°$, while the left-handed α-helix cannot. Even in the former case, because the planes of the peptide groups are nearly vertical, short contacts arise between the imino carbon atom, C, and the atoms of the backbone of the residue on top of it, if it is in the middle. It has been found that the two α-helical portions adjoining the proline ring must be inclined by at least 35° in order that these short contacts may be relieved. Even in this case, there arise two short contacts $C_2' \ldots C_1' = 2.61$ A and $C_2' \ldots O_1 = 2.45$ A between atoms in the adjacent residues and these can be relieved only by distorting the planarity of the peptide group.

*γ-helix.* The conformation of the right-handed γ-helix (−96°, 78°) lies well outside the allowed region and hence is not likely to be observed.

*$2_7a$-helix.* This type of folding originally proposed by Huggins (1943) and designated as $2_7$-helix by Bragg *et al.* (1950) has the conformation (105°, 70°) and lies apparently outside the allowed region. But the short contact in this case (as well as in the case of the $2.2_7$-helix to be discussed next) is $O_1 \cdots H_2$ (1.93A). Since this distance forms part of a possible intrachain hydrogen bond with $N_2 \cdots O_1$ = 2.72 A and $H_2\hat{N}_2O_1 = 28°30'$, this is really not a short contact.

*$2.2_7$-helix (ribbon structure).* This structure proposed by Donohue (1953) has the conformation (102°, 59°) and this also apparently lies outside the allowed region. As in the previous case, the short contact $O_1 \cdots H_2$ (1.88 A) forms part of a possible intrachain hydrogen bond and so this conformation becomes allowed. In fact a similar type of folding has been observed by Kakudo *et al.* (1963) in a tetrapeptide.

*$3.0_{10}$-helix.* This structure proposed by Bragg *et al.* (1950) and Donohue (1953) has the conformation (131°, −26°) for the right-handed helix and lies close to the $\alpha$-helix conformation, within the outer limit region. This structure is also quite likely to occur.

*$4.3_{14}$-helix.* This structure, also proposed by Donohue (1953), has the conformation (−92°, 92°) for the right-handed helix and lies close to the $\gamma$-helix, outside the allowed regions and hence is not likely to occur.

*$4.4_{16}$-helix ($\pi$).* The two slightly different forms of $\pi$-helices proposed by Low and Grenville-Wells (1953), one with $n = 4.3$ and $h = 1.14$ A and the other with $n = 4.4$ and $h = 1.15$ A, have the conformations (124°, −69°) and (123°, −70°) lying close to the conformation of $\alpha$-helix and are quite likely to occur.

*Poly-Gly, -Pro, -Hypro.* The conformations of these helices (left-handed) lie close to (100°, 150°) well within the normally allowed region. In the case of poly-L-proline and poly-L-hydroxyproline, the value of $\phi$ is necessarily close to 120° for the imino residues to be accommodated. Also, the right-handed counterparts of these cannot accommodate imino acid residues, since the value of $\phi$ is close to −120° in these cases.

*$\omega$-Form of Poly-$\beta$-Benzyl-L-Aspartate.* The structure of the $\omega$-form of poly-$\beta$-benzyl-L-aspartate has been determined by Bradbury *et al.* (1962) to be a left-handed helix with a fourfold screw axis and with a folding similar to the $4_{18}$-helix proposed by Bragg *et al.* (1950). The conformation of this helix is (−116°, 55°) and this lies just outside the allowed region. However, by making the peptide groups non-planar, the short contact that disallows this conformation, namely $C_\beta \cdots C_1'$, can be made to have a value of 3.13 A, which is allowed by the outer limit. Thus this structure, in the form proposed, also becomes allowed.

*Silk.* Silk has the conformation close to (40°, 135°) which lies well within the allowed region. The structure is slightly buckled with $h = 3.45$ A, smaller than that of a fully extended chain ($h = 3.63$ A). This is due to the presence of side chains like alanine, etc. In fact, the conformations of other $\beta$-forms of polypeptides

like polyglycine, β-poly-L-alanine, β-poly-γ-methyl-L-glutamate have conformations close to silk and are also slightly bucked.

*P-C Puckered Chain.* This chain conformation proposed by Pauling and Corey (1953) for the β-form of proteins has the conformation (57°, 118°), close to that of silk. But this is much more buckled and the value of $h$ in these cases is 3.3 A, less than that for silk (3.45 A). This is fully allowed, but this chain cannot accommodate proline and this has led to the suggestion of an alternative puckered chain (Pauling and Corey, 1953; Ramachandran, 1962; Sasisekharan, 1962).

*Collagen.* The polypeptide chains of the collagen helix ($n = -3.28$ and $h = 2.95$ A) has the conformation (116°, 145°). Because $\phi \simeq 120°$, imino acid residues can be readily accommodated. The value of $\phi'$ close to 180° is also a natural occurrence in the case of amino acids. Consequently, the fact that this conformation provides enough sites for imino acids and that it has a natural value of $\phi'$ suggest that, even without internal hydrogen bonding, the collagen chain conformation is likely to be quite stable.

The conformation of collagen lies within the normally allowed region. Another interesting feature is that if $\phi'$ is kept close to 120°, the value of $h$ does not vary from 2.9 A by more than 5 per cent within the allowed region, thus explaining the relative inextensibility of collagen.

## CONCLUDING REMARKS

Thus the conformations observed in simple and cyclic peptides are seen to confirm the predictions of the theory, thereby proving the validity of the stereochemical criteria assumed in this study.

Finally, the results obtained in our study agree well with those obtained by others from considerations of potential energy. Thus, Mizushima and Shimanouchi (1961), from a study of the internal rotation of simple molecules have preferred nine orientations. In our notation, these correspond to $\phi = 0$, 120, and 240°, $\phi' = 60$, 180, and 300°. These are given in Table IX, along with the values of the unit twist $t$ and the unit translation $h$. An examination of these in the light of the present study shows that all except *d*, *e*, and *i* lie within the allowed region. Even among these conformations, *i* becomes allowed because of the formation of a hydrogen bond of the type $N_2$-$H_2$ . . . $O_1$, which offsets the short contact, thus leaving the conformation *d* and *e* alone to be really disallowed.

Recently, De Santis *et al.* (1963) have attempted a study of the stability of helical conformations of linear polypeptide chains from potential energy considerations. Although there is a broad similarity between their potential energy contours and the boundaries of our diagrams, there are some differences, for example a deep potential minimum corresponding to about (120°, 300°) and another one at about (160°, 60°). Neither of these are particularly populated in Fig. 2*b*, showing that

TABLE IX

PREFERRED ORIENTATIONS OF POLYPEPTIDE CHAIN IN OUR NOTATION (*ACCORDING TO MIZUSHIMA AND SHIMANOUCHI*)

| Conformation | Twist angles | | Unit twist | No. of residues per turn | Residue height |
|---|---|---|---|---|---|
| | $\phi$ | $\phi'$ | $t$ | $n$ | $h$ |
| | *degrees* | | *degrees* | | *A* |
| a | 0 | 180 | 180 | 2.00 | 3.63 |
| b | 0 | −60 | −96 | −3.75 | 2.45 |
| c | 0 | 60 | 96 | 3.75 | 2.45 |
| d | 240 | 180 | 84 | 4.28 | 2.72 |
| e | 240 | −60 | 176 | 2.05 | 2.63 |
| f | 240 | 60 | −90 | −4.00 | 1.16 |
| g | 120 | 180 | −84 | −4.28 | 2.72 |
| h | 120 | −60 | 90 | 4.00 | 1.16 |
| i | 120 | 60 | −176 | −2.05 | 2.63 |

for a single pair of residues these are not specially favored. The potential energy data that are fed in such studies may have to be further revised.

## APPENDIX

While evaluating the contact distances for non-glycyl residues, only the $\beta$-carbon atom of the side chain was taken into account. It has been possible to extend the study to working out the effect of the $\gamma$-carbon atom on the conformation of the two residues. This was made possible since it has been found in Madras, from an analysis of the observed orientation of the side chains in the structures of various amino acids and peptides, that there are three preferred positions for the $\gamma$-carbon atom.

The position of $C_\gamma$ may be represented by a parameter $\psi$ which measures the rotation of $C_\gamma$ about the $C_\alpha$—$C_\beta$ bond from a standard initial conformation ($\psi = 0°$) in which the atom $C_\gamma$ lies in the plane N-$C_\alpha$-$C_\beta$ and with atoms N and $C_\gamma$ being in the *cis* configuration with respect to the $C_\alpha$—$C_\beta$ bond. Viewing from $C_\beta$ to $C_\alpha$, $\psi$ is taken to be positive for an counterclockwise rotation. With this convention, the three positions of $C_\gamma$ are found to lie close to values of 60, 180, and 300° for $\psi$, and these are indicated by indices 1, 2, and 3 respectively.

The contact distances between the $\gamma$-carbon atom in these three positions and the backbone atoms were calculated in the range of ($\phi$, $\phi'$) allowed by the $\beta$-carbon atom (given in Table III) for $\tau = 110°$. The alterations in the outer limit boundaries thus worked out are shown in Fig. 6, where the shifts in the boundary due to the $\gamma$-carbon atom in the three positions are shown by arrows.

In all the three cases, there are some additional restrictions in the allowed ranges of either $\phi$ or $\phi'$, or both, as can be seen from Fig. 6. For the first position of $C_\gamma$ ($\psi = 60°$), there are restrictions on both $\phi$ and $\phi'$, while for the second position ($\psi = 180°$), the restrictions are on $\phi'$ only and for the third position ($\psi = 300°$), on $\phi$ only.

Fig. 6 also contains the conformations of the right-handed and left-handed $\alpha$-helices marked in it. It is interesting to note that for the first position $C_\gamma$ ($\psi = 60°$) the con-



145

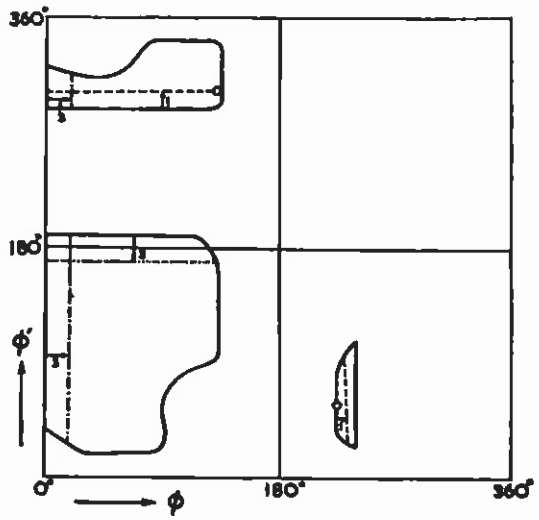



FIGURE 6 The shifts in the boundaries (indicated by →) of the allowed regions (outer limit) taking into account the restrictions due to γ-carbon atom in the three positions: ——— $C_\beta$ alone; – – – with $C_\gamma$ in position 1; ..... $C_\gamma$ in position 2; and –.–.–.–. $C_\gamma$ in position 3. The conformations of the right- and left-handed α-helices are also marked.

formation of the right-handed α-helix lies just on the boundary of the outer limit region, while the conformation of the left-handed α-helix goes outside the outer limit. In the other two cases, both are only within the outer limits and there is no obvious case for either the right-handed or the left-handed α-helix to be more likely. In the analysis made on amino acid side group conformations (Lakshminarayanan, unpublished), both L-glutamic acid and L-glutamic acid HCl have the γ-carbon atom in position 1, which may explain why poly-L-glutamic acid has a hight-handed helix. However, the three positions are found to occur fairly evenly distributed in different amino acids, and this explanation will not hold good for the very common occurrence of the right-handed alpha helix in preference to the left-handed one in various polypeptides.

Our thanks are due to Dr. V. Sasisekharan for the discussions which we had with him. One of the authors (C.R.) wishes to thank the Ministry of Scientific Research and Cultural Affairs and The Council of Scientific and Industrial Research, India, for financial support. The other author (G.N.R.) is grateful to the United States National Science Foundation for a Senior Visiting Fellowship, which enabled him to spend a year in the University of Michigan. We are very grateful to Professor J. L. Oncley for his interest in these studies.

Part I of this article appeared in *Proc. Indian Acad. Sc.*, 1964, **A59**, 327.

*Received for publication, July 19, 1965.*

## REFERENCES

BAMFORD, C. H., ELLIOTT, A., and HANBY, W. E., 1956, Synthetic Polypeptides, New York, Academic Press, 260.

BRADBURY, E. M., BROWN, L., DOWNIE, A. R., ELLIOTT, A., FRASER, R. D. B., and HANBY, W. E., 1962, *J. Mol. Biol.*, **5**, 230.

BRAGG, W. L., KENDREW, J. C., and PERUTZ, M. F., 1950, *Proc. Roy. Soc., London, Series A*, **203**, 321.
COCHRAN, W., and PENFOLD, B. R., 1952, *Acta Cryst.*, **5**, 644.
COREY, R. B., and PAULING, L., 1953, *Proc. Roy. Soc., London, Series B*, **141**, 10.
COWAN, P. M., and MCGAVIN, S., 1955, *Nature*, **176**, 501.
CRICK, F. H. C., and RICH, A., 1955, *Nature*, **176**, 780.
DAWSON, B., 1953, *Acta Cryst.*, **6**, 81.
DE SANTIS, P., GIGLIO, E., LIQUORI, A. M., and RIPAMONTI, A., 1965, *Nature*, **206**, 456.
DONOHUE, J., 1951, *J. Am. Chem. Soc.*, **72**, 949.
DONOHUE, J., 1953, *Proc. Nat. Acad. Sc.*, **39**, 205.
DONOHUE, J., and CARON, A., 1964, *Acta Cryst.*, **17**, 1178.
DONOHUE, J., and MARSH, R. E., 1962, *Acta Cryst.*, **15**, 941.
DONOHUE, J., and TRUEBLOOD, K. N., 1952, *Acta Cryst.*, **5**, 419.
DYER, H. B., 1951, *Acta Cryst.*, **4**, 42.
FREEMAN, H. C., ROBINSON, G., and SCHOONE, J. C., 1964*a*, *Acta Cryst.*, **17**, 719.
FREEMAN, H. C., SNOW, M. R., NITTA, I., and TOMITA, K., 1964*b*, *Acta Cryst.*, **17**, 1463.
FRIDRICHSONS, J., and MATHIESON, A. MCL., 1962, *Acta Cryst.*, **15**, 569.
HARDING, M. M., and COLE, S. J., 1963, *Acta Cryst.*, **16**, 643.
HINE, R., 1962, *Acta Cryst.*, **15**, 635.
HIROKAWA, S., 1955, *Acta Cryst.*, **8**, 637.
HUGGINS, M. L., 1943, *Chem. Rev.*, **32**, 195.
HUGHES, E. W., and MOORE, W. J., 1949, *J. Am. Chem. Soc.*, **71**, 2618.
ITAKA, Y., 1960, *Acta Cryst.*, **13**, 35.
ITAKA, Y., 1961, *Acta Cryst.*, **14**, 1.
KAKUDO, M., SASADA, Y., KATSUBE, Y., SAKAKIBARA, S., and AKABORI, S., 1963, in Aspects of Protein Structure, (G. N. Ramachandran, editor), London, Academic Press, 69.
KARLE, I. L., and KARLE, J., 1963, *Acta Cryst.*, **16**, 969.
KARLE, I. L., and KARLE, J., 1964, *Acta Cryst.*, **17**, 835.
KARTHA, G., and DE VRIES, A., 1961, *Nature*, **192**, 862.
KENDREW, J. C., DICKERSON, R. E., STRANDBERG, B. E., HART, R. G., DAVIES, D. R., PHILLIPS, D. C., and SHORE, V. C., 1960, *Nature*, **185**, 422.
KENDREW, J. C., WATSON, H. C., STRANDBERG, B. E., DICKERSON, R. E., PHILLIPS, D. C., and SHORE, V. C., 1961, *Nature*, **190**, 666.
KRETSINGER, R. H., and COTTON, F. A., 1963, *Acta Cryst.*, **16**, 651.
LEUNG, Y. C., and MARSH, R. E., 1958, *Acta Cryst.*, **11**, 17.
LINDLEY, H., and ROLLETT, J. S., 1955, *Biochim. et Biophysica Acta*, **18**, 183.
LOW, B. W., and BAYBUTT, R. B., 1952, *J. Am. Chem. Soc.*, **74**, 5806.
LOW, B. W., and GRENVILLE-WELLS, H. J., 1953, *Proc. Nat. Acad. Sc.*, **39**, 785.
MARSH, R. E., 1958, *Acta Cryst.*, **11**, 654.
MARSH, R. E., and GLUSKER, J. P., 1961, *Acta Cryst.*, **14**, 1110.
MARSH, R. E., COREY, R. B., and PAULING, L., 1955, *Acta Cryst.*, **8**, 710.
MATHIESON, A. MCL., 1952, *Acta Cryst.*, **5**, 332.
MATHIESON, A. MCL., 1953, *Acta Cryst.*, **6**, 399.
MATHIESON, A. MCL., and WELSH, H. K., 1952, *Acta Cryst.*, **5**, 599.
MAZUMDAR, S. K., 1964, Ph.D. Thesis, Madras University.
MIZUSHIMA, S., and SHIMANOUCHI, T., 1961, *Advanc. Enzymol.*, **23**, 1.
NÉMETHY, G., and SCHERAGA, H. A., 1965, *Biopolymers*, **3**, 155.
OUGHTON, B. M., and HARRISON, P. M., 1959, *Acta Cryst.*, **12**, 396.
PARTHASARATHY, R., 1962, Ph.D. Thesis, Madras University.
PASTERNAK, R. A., 1956, *Acta Cryst.*, **9**, 341.
PASTERNAK, R. A., KATZ, L., and COREY, R. B., 1954, *Acta Cryst.*, **7**, 225.
PAULING, L., and COREY, R. B., 1951, *Proc. Nat. Acad. Sc.*, **37**, 235.
PAULING, L., and COREY, R. B., 1953, *Proc. Nat. Acad. Sc.*, **39**, 253.



147

PAULING, L., COREY, R. B., and BRANSON, H. R., 1951, *Proc. Nat. Acad. Sc.*, **37**, 205.
PETERSON, J., STEINRAUF, L. K., and JENSON, L. H., 1960, *Acta Cryst.*, **13**, 104.
RAMACHANDRAN, G. N., 1960, *Proc. Indian Acad. Sc.*, **A52**, 240.
RAMACHANDRAN, G. N., 1962, *in* Collagen, (N. Ramanathan, editor), New York, John Wiley & Sons, 1.
RAMACHANDRAN, G. N., 1963, *in* Aspects of Protein Structure, (G. N. Ramachandran, editor), London, Academic Press, 39.
RAMACHANDRAN, G. N., RAMAKRISHNAN, C., and SASISEKHARAN, V., 1963*a*, *in* Aspects of Protein Structure, (G. N. Ramachandran, editor), London, Academic Press, 121.
RAMACHANDRAN, G. N., RAMAKRISHNAN, C., and SASISEKHARAN, V., 1963*b*, *J. Mol. Biol.*, **7**, 95.
RAMAKRISHNAN, C., 1964, *Proc. Indian Acad. Sc.*, **A59**, 327.
SASISEKHARAN, V., 1959*a*, *Acta Cryst.*, **12**, 897.
SASISEKHARAN, V., 1959*b*, *Acta Cryst.*, **12**, 903.
SASISEKHARAN, V., 1962, *in* Collagen, (N. Ramanathan, editor), New York, John Wiley and Sons, 39.
SCHERAGA, H. A., 1960, *J. Am. Chem. Soc.*, **82**, 3847.
SHIMANOUCHI, T., and MIZUSHIMA, S., 1955, *J. Chem. Physics*, **23**, 707.
SHOEMAKER, D. P., DONOHUE, J., SCHOMAKER, V., and COREY, R. B., 1951, *J. Am. Chem. Soc.*, **72**, 2328.
SHOEMAKER, D. P., BARREAU, R. E., and DONOHUE, J., 1953, *Acta Cryst.*, **6**, 241.
SMITH, D. W., and WIEBENGA, E. H., 1953, *Acta Cryst.*, **6**, 531.
SRINIVASAN, R., 1959*a*, *Proc. Indian Acad. Sc.*, **A49**, 340.
SRINIVASAN, R., 1959*b*, *Proc. Indian Acad. Sc.*, **A50**, 19.
SUBRAMANIAN, E., 1965, Ph.D. Thesis, Madras University.
TROMMEL, J., and BIJVOET, J. M., 1954, *Acta Cryst.*, **7**, 703.
WHITTAKER, E. T., A Treatise on Analytical Dynamics of Particles and Rigid Bodies, Cambridge, Cambridge University Press, 1952, 8.
WRIGHT, D. A., and MARSH, R. E., 1962, *Acta Cryst.*, **15**, 54.
WRIGHT, W. B., 1958, *Acta Cryst.*, **11**, 632.
YAKEL, H. L., and HUGHES, E. W., 1954, *Acta Cryst.*, **7**, 291.
ZALKIN, A., FOSTER, J. D., and TEMPLETON, H., 1964, *Science*, **146**, 261.

**Prediction of three dimensional folding of CDR's**

The framework regions of antibody molecules have very similar three dimensional structures. Therefore, different antibodies have characteristic structures localized in their CDR's. For given amino acid sequences of the six CDR's, a predictive method has been developed to predict the detailed three dimensional structure of the antibody combining site (Stanford JM, Wu TT (1981) A predictive method for determining possible three-dimensional folding of immunoglobulin backbones around antibody combining sites. J Theor Biol, 88, 421-439).

The basic approach involved the following steps:

1. Compare every three consecutive amino acid residues in each CDR with sequences in reference proteins with known three dimensional structures.

2. If similar tri-peptide sequences are found, the $\phi$ and $\psi$ angles of the middle amino acid residues in reference proteins are recorded.

3. Thus several possible $\phi$ and $\psi$ choices are available for each amino acid residue in all six CDR's.

4. Each CDR has fix locations and orientations for it N and C terminal residues in order to fit onto the framework region.

5. From the combinations of $\phi$ and $\psi$ choices of all positions of each CDR, several of them may have no steric hindrance of the backbone atoms including C$\beta$. Each angle is then adjusted within $30^{\circ}$ so that the N and C terminal ends are precisely fitted onto the framework structure.

6. Since the antibody combining site is very compact, predictions of the folding of the CDR's are carried sequentially starting with CDRH3, followed by CDRL3, CDRH1, CDRL1, CDRH2, and CDRL2, i.e. from the center of antibody combining site going outwards. Otherwise, there would be no space to fit all the CDR's

7. Taking all the six CDR's together, usually only one possible three dimensional structure of the antibody combining site is possible with no steric hindrance of the backbone atoms including C$\beta$.

Finally, side chain atoms can then be incorporated. In most cases, there are reasonable degrees of freedom for the rotation around the C$\beta$-C$\gamma$ bond (Hovis JG, Wu TT (1985) Prediction of the side chain configurations. In Methodologies in Studies of Protein Configuration. Ed. T. T. Wu, Van Nostrand Reinhold Co., New York).

*J. theor. Biol.* (1981) **88**, 421–439

# A Predictive Method for Determining Possible Three-dimensional Foldings of Immunoglobulin Backbones Around Antibody Combining Sites

JANETTE M. STANFORD

*Center for Naval Analyses, Alexandria, VA* 22311, *U.S.A.*

AND TAI TE WU

*Departments of Biochemistry & Molecular Biology and Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL* 60201, *U.S.A.*

(*Received* 10 *March* 1980, *and in revised form* 6 *August* 1980)

A method has been developed to replace segments of a protein backbone by other segments with standard peptide units and given $(\phi, \psi)$ angles. Restrictions were imposed by fixed orientations of the N- and C-terminals and by the absence of steric hindrance between non-bonding atoms. This procedure was applied to the prediction of three dimensional backbone conformations of antibody combining sites based on the amino acid sequences of the six complementarity determining regions. In particular, the detailed backbone structure around the combining site of immunoglobulin MOPC-315 was predicted. Atomic co-ordinates of all non-hydrogen backbone atoms including $C_\beta$ of the predicted structure were given so that future comparisons with experimental data and other predictions could be made.

## 1. Introduction

Anfinsen & Redfield (1956) proposed that the amino acid sequence of a protein contained all the necessary information for the specification of its three-dimensional folding, presumably by acquiring the state of lowest free energy in a given environment (Anfinsen & Scheraga, 1975). Many predictive methods based on this principle (Anfinsen & Redfield, 1956) were subsequently developed (see, for example, Schulz *et al.*, 1974; Wu, Fitch & Margoliash, 1974; Schulz, 1977; Sternberg & Thornton, 1978; Chou & Fasman, 1978). One of these methods (Kabat & Wu, 1972; Wu & Kabat, 1973) attempted to predict $(\phi, \psi)$ angles rather than to locate $\alpha$-helices, $\beta$-sheets and $\beta$-bends. This method has been successfully applied to small peptides (Honig *et al.*, 1973; Johnson & Wu, 1976) to give detailed three-dimensional structures of their backbones.

Immunoglobulins are a distinct collection of proteins in the sense that the three-dimensional structures of their variable regions can be divided into two



0022-5193/81/030421+19 $02.00/0 © 1981 Academic Press Inc. (London) Ltd.

separate parts (see, for example, Kabat, 1978). The framework (FR) portions of different immunoglobulin variable regions have essentially a similar tertiary structure except for some minor variations. Their antibody combining sites formed by the six complementarity determining regions (CDR) are all different. Therefore, the problem of predicting three dimensional structures of immunoglobulin variable regions is reduced to the determination of the foldings of the six CDRs or hypervariable segments (Wu & Kabat, 1970; Kabat & Wu, 1971) which are relatively short. For position numbering and locations of FRs and CDRs, please refer to Kabat, Wu & Bilofsky, 1979.

In the present study, a detailed account will be given to show how a short peptide segment can be fitted onto an existing protein framework, starting with ($\phi$, $\psi$) selection (Kabat & Wu, 1972; Wu & Kabat, 1973). Among the possible ($\phi$, $\psi$) choices, restrictions are imposed by the requirement that the orientations of and distances between N- and C-terminals of the segment must agree with those of the framework, and that steric hindrances between non-bonding atoms must be avoided. Specifically, the backbone structures of the six CDRs of immunoglobulin MOPC-315 will be analyzed.

## 2. Materials and Methods

### (A) SELECTION OF IMMUNOGLOBULINS

In order to predict the three-dimensional folding of the backbone of an immunoglobulin around the antibody combining site, the amino acid sequence for the six complementarity determining regions must have been determined. There are several immunoglobulins which satisfy this requirement (Kabat, Wu & Bilofsky, 1979); however, extensive experimental work has been done on MOPC-315 to determine its contact residues (Eisen, Simms & Potter, 1968; Haimovich *et al.*, 1972; Klostergaard *et al.*, 1977) and the size of its binding site (Sutton *et al.*, 1977; Pecht, 1975). Immunoglobulin MOPC-315 was thus used in the present predictive study.

The atomic co-ordinates of the framework region of Newm (Poljak *et al.*, 1971; Poljak *et al.*, 1973; Poljak *et al.*, 1975; Poljak *et al.*, 1977; Saul, Amzel & Poljak, 1978; Protein Data Bank, 1979) were used as a model for the invariant framework, since it is the only immunoglobulin whose detailed structure, including both the light and the heavy chain variable regions, is available to the public through the Protein Data Bank (c/o Dr T. F. Koetzle, Brookhaven National Laboratory, Uptown, NY 11973, U.S.A.). Since the second CDR in Newm's light chains (CDR-L2) is deleted, this local structure was substituted with that of REI (Epp *et al.*, 1974; Protein Data Bank),



and the loops in the framework of the light chain which border on the antibody combining site (Kabat, 1978) had to be adjusted slightly.

### (B) SELECTION OF ($\phi$, $\psi$) ANGLES

The empirical procedure developed previously (Kabat & Wu, 1972; Wu & Kabat, 1973, Wu, Fitch & Margoliash, 1974) was used with the following modifications. Since immunoglobulins do not contain a substantial amount of $\alpha$-helices (e.g. 2% for Newm, 5% for REI and 10% for Mcg), it would be inaccurate to predict their structures based on reference proteins which contain essentially $\alpha$-helices. Therefore, the reference proteins (Wu & Kabat, 1973; Protein Data Bank) are divided into three groups (Levitt & Crothia, 1976; for simplicity, we combined helix+sheet and helix/sheet into one group): those which mainly contain $\beta$-sheets ($\beta$-sheet proteins), those which mainly contain $\alpha$-helices (helix proteins), and those which do not contain a majority of either $\beta$-sheets or $\alpha$-helices (intermediate proteins). The $\beta$-sheet reference proteins, e.g. tosyl-$\alpha$-chymotrypsin, chymotrysinogen, concanavalin A, tosyl-elastase, REI, rubredoxin, etc., were preferentially used as the data base in the present study. However, if no ($\phi$, $\psi$) choices were available for certain tripeptides or no combination of ($\phi$, $\psi$) choices were adequate, intermediate, e.g. carboxypeptidase A, lysozyme, ribonuclease, subtilisin, staphlococcal nuclease, calf cytochrome $b_5$, papain, basic pancreatic trypsin inhibitor, insulin, thermolysin, high potential iron protein, dogfish apolactate dehydrogenase, alcohol dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase, etc., and helix, e.g. myoglobin, horse hemoglobin, human hemoglobin, lamprey hemoglobin, cytochromes *c*, etc., reference proteins were then used.

Since several ($\phi$, $\psi$) choices were usually possible for each amino acid residue, each CDR had one to two thousand different possible structures, each structure corresponding to a particular combination of the ($\phi$, $\psi$) choices. Additional restrictions had to be imposed to reduce the number of choices (Wu & Kabat, 1973; Honig *et al.*, 1973; Johnson & Wu, 1976). These restrictions were provided by the requirement that the CDRs had to be able to connect onto the framework, whose structures was precisely determined by X-ray diffraction studies of Newm and REI.

### (C) RESTRICTIONS IMPOSED BY THE FRAMEWORK AND STERIC HINDRANCE

Before a CDR could hook onto the framework, it had to possess the proper distance between its N- and C-terminal residues. Furthermore, the

spacial orientation of these residues were also determined by the framework.

These restrictions can best be illustrated by analyzing one CDR, for example, the third CDR of the light chain (CDR-L3). CDR-L3 extends from residue 89 to 97 (Kabat, Wu & Bilofsky, 1979), so that residues 88 and 98 are the framework residues immediately preceeding and following CDR-L3 respectively. The atomic co-ordinates and ($\phi$, $\psi$) angles for residues 88 and 98 have been determined by X-ray diffraction analysis for Newm (Poljak *et al.*, 1971; Poljak *et al.*, 1973; Poljak *et al.*, 1975; Poljak *et al.*, 1977; Protein Data Bank). Fixing $\psi$ of residue 88 also sets the locations of the NH and $C_\alpha$ of residue 89 in CDR-L3, and similarly, fixing $\phi$ of residue 98 sets the locations of the CO and $C_\alpha$ of residue 97 in CDR-L3. The distance between the $C_\alpha$s of residues 89 and 97 in the predicted structure of CDR-L3 had to be close to that dictated by the framework. If the distances between the predicted and experimental locations of $C_\alpha$ of residue 98 (also 99 and 100) differed by more than about 5 Å, the set of ($\phi$, $\psi$) choices was discarded.

The crystallographically determined ($\phi$, $\psi$) angles of residue 88 in the framework were incorporated with any set of ($\phi$, $\psi$) choices for CDR-L3 so that the orientational restriction of N-terminal of CDR was precisely satisfied. In order to satisfy the orientation of the C-terminal of CDR required by the fixed framework, the known ($\phi$, $\psi$) angles of residues 98, 99 and 100 were incorporated with ($\phi$, $\psi$) choices of CDR-L3 so that the atomic co-ordinates of the $C_\alpha$ atoms of these three residues could be calculated for each predicted structure. The distances between the predicted and actual locations of these $C_\alpha$s were squared and summed (least square criterion). If this number was greater than 75 Å$^2$, the set of ($\phi$, $\psi$) choices was discarded, since it would be unlikely for these choices that the least square criterion could be reduced by adjusting ($\phi$, $\psi$) choices within ±30° to less than 1·5 Å$^2$.

The above restrictions easily reduced the original one to two thousand angle combinations for each CDR to about five possibilities. For each of these few combinations, the ($\phi$, $\psi$) angles were then adjusted individually within ±30° in order to reduce the value of the sum of least squares. This procedure involved "walking" in a $2n$-dimensional space where n was the number of amino acid residues in the peptide segment. Each $\phi$ or $\psi$ angle of the residue of the peptide segment was individually varied by ±30° from the initial choice, and the sums of squares were calculated for each 5° interval and tabulated (see Table 1). The change in $\phi$ or $\psi$ angle which gave the least sum was then used for the next iteration (see Table 2). In the mean time, the non-bonding distances between the backbone atoms including NH, $C_\alpha$, H,



TABLE 1

*Least square criterions for the initial set of ($\phi$, $\psi$) angles with various individual increments of the $\phi$ or $\psi$ angle*

| $\phi$ or $\psi$ | Residue number | Independent Increments of $\phi$ or $\psi$ Angle | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | −30° | −25° | −20° | −15° | −10° | −5° | 0° | +5° | +10° | +15° | +20° | +25° | +30° |
| $\phi$ | 89 | 47·6 | 39·5 | 33·4 | 29·6 | 27·9 | 28·3 | 31·0 | 35·9 | 42·8 | 51·9 | 62·9 | 75·9 | 90·7 |
| $\psi$ | 89 | 52·0 | 42·7 | 35·6 | 30·9 | 28·6 | 28·6 | 31·0 | 35·8 | 42·8 | 52·2 | 63·7 | 77·3 | 93·0 |
| $\phi$ | 90 | 43·6 | 36·7 | 31·7 | 28·6 | 27·4 | 28·2 | 31·0 | 35·7 | 42·3 | 50·8 | 61·0 | 73·0 | 86·5 |
| $\psi$ | 90 | 60·0 | 49·3 | 40·9 | 34·9 | 31·2 | 29·9 | 31·0 | 34·5 | 40·4 | 48·6 | 59·1 | 71·8 | 86·5 |
| $\phi$ | 91 | 71·3 | 57·1 | 45·7 | 37·4 | 32·2 | 30·0 | 31·0 | 35·1 | 42·3 | 52·5 | 65·7 | 81·7 | 100·4 |
| $\psi$ | 91 | 40·8 | 31·0 | 24·4 | 21·1 | 21·1 | 24·4 | 31·0 | 40·8 | 53·8 | 69·9 | 88·9 | 110·7 | 135·1 |
| $\phi$ | 92 | 36·5 | 27·9 | 22·4 | 19·9 | 20·5 | 24·2 | 31·0 | 40·8 | 53·5 | 69·1 | 87·3 | 108·2 | 131·4 |
| $\psi$ | 92 | 62·8 | 55·3 | 48·6 | 42·7 | 37·8 | 33·9 | 31·0 | 29·1 | 28·3 | 28·5 | 29·7 | 32·0 | 35·3 |
| $\phi$ | 93 | 59·7 | 52·6 | 46·5 | 41·2 | 36·8 | 33·4 | 31·0 | 29·6 | 29·2 | 29·8 | 31·3 | 33·9 | 37·5 |
| $\psi$ | 93 | 184·9 | 142·4 | 106·3 | 76·9 | 54·5 | 39·2 | 31·0 | 30·2 | 36·6 | 50·2 | 71·0 | 98·7 | 133·2 |
| $\phi$ | 94 | 160·6 | 123·9 | 93·0 | 68·1 | 49·4 | 37·0 | 31·0 | 31·5 | 38·5 | 51·9 | 71·6 | 97·5 | 129·4 |
| $\psi$ | 94 | 214·1 | 170·0 | 131·0 | 97·4 | 69·4 | 47·2 | 31·0 | 20·9 | 17·1 | 19·4 | 28·0 | 42·6 | 63·4 |
| $\phi$ | 95 | 183·1 | 147·2 | 115·2 | 87·4 | 64·0 | 45·2 | 31·0 | 21·7 | 17·2 | 17·7 | 23·0 | 33·3 | 48·3 |
| $\psi$ | 95 | 99·7 | 82·6 | 67·7 | 54·9 | 44·5 | 36·5 | 31·0 | 28·0 | 27·5 | 29·6 | 34·2 | 41·2 | 50·7 |
| $\phi$ | 96 | 132·7 | 107·4 | 85·2 | 66·4 | 51·0 | 39·2 | 31·0 | 26·6 | 25·9 | 29·0 | 35·8 | 46·3 | 60·4 |
| $\psi$ | 96 | 66·3 | 59·2 | 52·6 | 46·4 | 40·7 | 35·6 | 31·0 | 27·1 | 23·8 | 21·2 | 19·3 | 18·1 | 17·7 |
| $\phi$ | 97 | 75·6 | 65·7 | 56·8 | 48·8 | 41·8 | 35·8 | 31·0 | 27·4 | 24·9 | 23·6 | 23·5 | 24·6 | 26·9 |
| $\psi$ | 97 | 20·8 | 21·1 | 21·9 | 23·3 | 25·3 | 27·9 | 31·0 | 34·7 | 38·8 | 43·4 | 48·5 | 53·9 | 59·7 |

TABLE 2

*Least square criterions after the first iteration of the initial ($\phi$, $\psi$) angles with various individual increments of the $\phi$ or $\psi$ angle*

| $\phi$ or $\psi$ | Residue Number | Independent Increments of $\phi$ or $\psi$ Angle −30° | −25° | −20° | −15° | −10° | −5° | 0° | +5° | +10° | +15° | +20° | +25° | +30° |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\phi$ | 89 | 40·2 | 32·2 | 25·9 | 21·1 | 18·1 | 16·8 | 17·2 | 19·3 | 23·0 | 28·5 | 35·6 | 44·2 | 54·4 |
| $\psi$ | 89 | 47·3 | 37·6 | 29·7 | 23·7 | 19·5 | 17·4 | 17·2 | 18·9 | 22·6 | 28·2 | 35·7 | 45·0 | 56·1 |
| $\phi$ | 90 | 40·4 | 32·8 | 26·6 | 21·9 | 18·8 | 17·2 | 17·2 | 18·7 | 21·8 | 26·5 | 32·7 | 40·3 | 49·3 |
| $\psi$ | 90 | 45·6 | 36·0 | 28·3 | 22·5 | 18·7 | 16·9 | 17·2 | 19·4 | 23·7 | 30·0 | 38·2 | 48·3 | 60·2 |
| $\phi$ | 91 | 55·9 | 43·2 | 32·8 | 25·0 | 19·8 | 17·2 | 17·2 | 19·8 | 25·0 | 32·9 | 43·2 | 56·0 | 71·1 |
| $\psi$ | 91 | 58·8 | 44·9 | 33·7 | 25·3 | 19·7 | 17·0 | 17·2 | 20·2 | 26·2 | 35·0 | 46·5 | 60·8 | 77·6 |
| $\phi$ | 92 | 55·5 | 42·6 | 32·2 | 24·4 | 19·2 | 16·8 | 17·2 | 20·2 | 26·0 | 34·4 | 45·5 | 59·0 | 75·0 |
| $\psi$ | 92 | 32·5 | 28·1 | 24·5 | 21·5 | 19·3 | 17·9 | 17·2 | 17·9 | 18·1 | 19·7 | 22·1 | 25·2 | 29·0 |
| $\phi$ | 93 | 31·7 | 27·6 | 24·1 | 21·3 | 19·2 | 17·8 | 17·2 | 17·3 | 18·1 | 19·6 | 21·9 | 24·9 | 28·6 |
| $\psi$ | 93 | 136·8 | 100·2 | 70·0 | 46·4 | 29·7 | 19·9 | 17·2 | 21·4 | 32·7 | 50·9 | 75·9 | 107·5 | 145·4 |
| $\phi$ | 94 | 122·2 | 89·5 | 62·7 | 42·0 | 27·4 | 19·1 | 17·2 | 21·6 | 32·4 | 49·4 | 72·5 | 101·6 | 136·5 |
| $\psi$ | 94 | 137·8 | 103·1 | 74·0 | 50·7 | 33·4 | 22·2 | 17·2 | 18·4 | 25·9 | 39·6 | 59·3 | 84·9 | 116·3 |
| $\phi$ | 95 | 114·0 | 86·4 | 63·1 | 44·4 | 30·5 | 21·4 | 17·2 | 17·9 | 23·5 | 34·0 | 49·4 | 69·4 | 93·9 |
| $\psi$ | 95 | 79·2 | 62·0 | 47·3 | 35·4 | 26·4 | 20·3 | 17·2 | 17·0 | 19·8 | 25·6 | 34·3 | 45·9 | 60·3 |
| $\phi$ | 96 | 101·9 | 78·1 | 58·0 | 41·7 | 29·4 | 21·2 | 17·2 | 17·3 | 21·6 | 30·0 | 42·6 | 59·1 | 79·4 |
| $\psi$ | 96 | 33·9 | 29·6 | 25·9 | 22·8 | 20·3 | 18·4 | 17·2 | 16·6 | 17·7 | 19·5 | 22·4 | 26·2 | 30·9 |
| $\phi$ | 97 | 39·4 | 33·4 | 28·2 | 23·9 | 20·7 | 18·4 | 17·2 | 16·9 | 17·7 | 19·5 | 22·4 | 26·2 | 30·9 |
| $\psi$ | 97 | 27·5 | 24·4 | 21·8 | 19·8 | 18·4 | 17·5 | 17·2 | 17·4 | 18·3 | 19·7 | 21·7 | 24·2 | 27·3 |



$C_\beta$ and CO were then calculated for each CDR and were not allowed to be less than the minimum contact distances based on normally accepted van der Waal radii (Dickerson & Geis, 1969). In addition, steric hindrances between the CDRs and between the CDR and the framework atoms were also avoided. This process was then repeated until the sum of squares was reduced to less than 1·5 Å$^2$. In actual cases, not all combinations of initial ($\phi$, $\psi$) choices with reasonable structures could be so adjusted. Sometimes, a combination of ($\phi$, $\psi$) angles with larger initial sum of squares might result in a structure with lower final sum of squares. A schematic illustration of the results from this iterative procedure is shown in Fig. 1 for CDR-L3 of immunoglobulin MOPC-315, where the initial and final structures are indicated.




FIG. 1. Schematic diagram of the initial and final fits of CDR-L3 to the framework. The open circles with the primed numbers represent the Cs of framework residues from Newm whose positions are fixed. The filled circles with the unprimed numbers represent the Cs of residues in CDR-L3 of immunoglobulin MOPC-315 (positions 89–97) and four adjacent framework residues (positions 88, 98, 99 and 100) which are used to establish the fit of CDR-L3. The initial fit (top figure) is reasonable since the Cs of residues 98–100 are in the proximity of residues 98′ to 100′ respectively. After the refinement procedure, the fit to the framework (bottom figure) is quite good as is reflected in the highly similar locations of residues 98–100 and 98′ to 100′ respectively.

The above refinement method of fitting CDRs into framework portions is self-consistent in the sense that the experimentally known three-dimensional CDR structures of Newm and REI (Protein Data Bank) can be predicted by this method with good accuracy, in the sense that the predicted locations of all $C_\alpha$-atoms are within 0·5 Å of the experimentally determined locations.

153

## 3. Results

In order to illustrate the present findings, CDR-L3 is again used as an example to show how one of the final structures was obtained. CDR-L3 of MOPC-315 contains nine amino acid residues extending from position 89 to 97. Table 3 lists all the $(\phi, \psi)$ choices for these residues; there are a total of 1728 possible combinations. Together with the $(\phi, \psi)$ angles determined by X-ray diffraction studies at positions 88, 98, 99 and 100, one set of $(\phi, \psi)$ angle combinations for CDR-L3 (underlined in Table 3) produced a least square criterion of 31·0 Å$^2$ which could be further reduced to 1·3 Å$^2$ by adjusting the $(\phi, \psi)$ angles. The procedure of adjusting this set of $(\phi, \psi)$ angles within ±30° is illustrated below.

TABLE 3

*The $(\phi, \psi)$ choices for CDR-L3 of immunoglobulin MOPC-315, together with $(\phi, \psi)$ angles for residues 88, 98, 99 and 100 determined crystallographically for Newm given in parenthesis. The $(\phi, \psi)$ choices used for model construction are underlined. The adjusted $(\phi, \psi)$ angles used for fitting CDR-L3 onto the framework are also listed*

| Amino acid position and residue | $(\phi, \psi)$ choices | Adjusted $(\phi, \psi)$ angles |
|---|---|---|
| 88 Cys | (−120·8, 131·1) | −120·8, 118 |
| 89 Ala | −69·7, −13·0; −157·7, 144·5; −81·3, 102·0 | −103, 125 |
| 90 Leu | −109·6, 107·6; −72·6, −24·7; −158·4, 169·4 | −168, −178 |
| 91 Trp | −119·5, 142·5; −98·7, −52·7 | −125, 132 |
| 92 Phe | −102·5, 173·5; −94·0, 3·0 | −120, −179 |
| 93 Arg | −139·5, 149·0; −67·5, −43·0 | −71, −42 |
| 94 Asn | −60.7, −55.7; −112·5, 117·8 | −61, −72 |
| 95 His | −159·0, 139·0; −66·3, −54·8; −100·0, 15·0 | −161, 133 |
| 96 Phe | −124·8, 142·4; −53·0, −28·0 | −58, 2 |
| 97 Val | −70·7, 121·3; −62·5, −37·5 | −41, 147 |
| 98 Phe | (−118·7, 116·1) | −89, 116·1 |
| 99 Gly | (−101·4, 97·4) | −101·4, 97·4 |
| 100 Gly | (−100·2, 141·3) | −100·2, 141·3 |

For numbering, please refer to Kabat *et al.* (1979).

On varying $\phi$ of residue 92, for example, alone in 5° intervals for ±30° from its original value of −102·5°, the least square criterion reached a minimum of about 19·9 Å$^2$ for $\Delta\phi$ being −15° (Fig. 2). Similarly, on varying $\psi$ of residue 92 alone, a minimum occurred around $\Delta\psi$ being +10° (Fig. 2). Thus, by adjusting each $\phi$ or $\psi$ angle independently, a table of least square criterion could be constructed, consisting of all least square criterion values



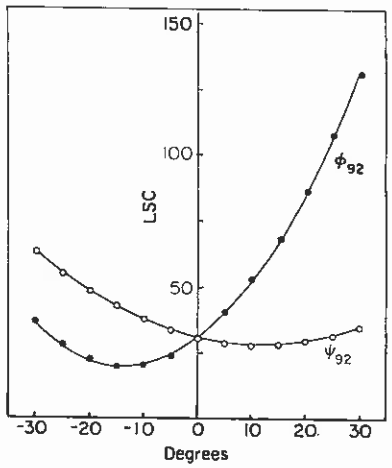



FIG. 2. Graphical representation of part of the first iteration in the refinement fitting procedure for CDR-L3. The abscissa represents the number of degrees added to or subtracted from a $\phi$ and $\psi$ angle, and the ordinate represents the least square criterion (LSC) values resulting from the modified conformations of CDR-L3. The curve with the filled dots shows the variation in LSC as $\phi$ of residue 92 is varied within ±30°, and the curve with the open dots shows the LSC variations as $\psi$ of residue 92 is similarly varied.

in 5° intervals (Table 1). On scanning this table, one noticed that the lowest values reached were 17·1 Å$^2$ or 17.2 Å$^2$ by adding 10° to the initial $\psi$ value of residue 94 or to the initial $\phi$ value of residue 95 respectively. Indeed, in many cases, the change in the $\psi$ value of one residue and that in the $\phi$ value of the next residue produced similar results. Thus, the $\psi$ of residue 94 was increased from −55·7° to −51° and $\phi$ of residue 95 from −159·0° to −154°, the increment in each being half of 10° or 5°. This adjustment reduced the least square criterion from 31·0 Å$^2$ to 17.2 Å$^2$.

This procedure was of course not unique. One might, for example, decrease the $\psi$ of residue 91 and $\phi$ of residue 92 to reduce the least square criterion to around 20 Å$^2$. However, all such adjustments eventually reached final sets of $(\phi, \psi)$ angles being within ±10°.

After the first adjustment, one could then generate a new table listing all least square criterion values using the new set of $(\phi, \psi)$ angles as initial choices (Table 2). This table indicated that 17·2 Å could be reduced to 16·6 Å$^2$ by changing $\psi$ of residue 96 from −28·0° to −21°. It became obvious that the reduction in least square criterion value in the second step of adjustment was not as drastic as the first. Indeed, for certain initial $(\phi, \psi)$ choices, the least square criterion value could not be reduced to less than about 10 Å$^2$, and such choices had to be discarded.

On repeating this iterative procedure numerous times, a final set of adjusted ($\phi$, $\psi$) angles was obtained (Table 3). A least square criterion of 1·3 Å$^2$ was reached. Further reduction was possible; however, in view of the experimental errors involved in determining the $C_\alpha$-atom locations in the framework portions of the protein, such reduction might not be justified.

During this entire process of iteration, steric hindrance within CDR-L3 and between CDR-L3 and framework (with minor adjustments) atoms were avoided. Based on the final set of ($\phi$, $\psi$) angles (Table 3), the atomic co-ordinates of the backbone atoms of CDR-L3 of MOPC-315 were calculated and are listed in Table 4.

The same procedure was repeated for CDR-H3, CDR-H1, CDR-L1, CDR-H2 and CDR-L2, in that order, since CDR-L3 and CDR-H3 were located near the center of the combining site while CDR-H2 and CDR-L2 were on the periphery. Any steric hindrance between the CDRs could most easily be avoided by adjusting the more peripheral ones. CDR-L2 required special consideration since it is deleted in Newm. Residues 50–56 of CDR-L2 were first fitted onto the framework of REI, and then residues 47–61 were fitted onto the framework of Newm. Since Newm has this deletion, its framework loop consisting of residues 62–75 bordering on the combining site is located more towards the center of the site as compared with other immunoglobulins (Kabat, 1978). Therefore, slight adjustments would be required to move this loop outwards in the case of immunoglobulin MOPC-315. In order to avoid some minor steric hindrances, three amino acid residues (47 Trp, 48 Leu, 49 Gly; see Table 5) instead of the usual one residue in front of CDR-H2 were used for adjustment. The final adjusted ($\phi$, $\psi$) angles for all CDRs (except CDR-L3) and some adjacent residues are listed in Table 5, and the atomic coordinates of the backbone atoms of these CDRs in Table 6. The least square criterion values for the finally adjusted CDRs L1, L2, L3, H1, H2 and H3 were 1·2, 0·3, 1·3, 1·4, 1·4 and 0·9 Å$^2$ respectively.

The predicted structure of the backbone of immunoglobulin MOPC-315 combining site is shown in Fig. 3, where only $C_\alpha$ and $C_\beta$ atoms are indicated. This view is looking straight down into the cavity. A stereo-pair of the site is shown in Fig. 4.

## 4. Discussion

Since the three-dimensional structure of the combining site of immunoglobulin MOPC-315 has not been analyzed by X-ray diffraction studies, the present predicted model could only be compared with other less detailed experimental data. It was found that Tyr 32 of CDR-L1 was in contact with the hapten 2,4-dinitrophenol, that Lys 52 of CDR-H2 was peripheral to the

TABLE 4

*Atomic co-ordinates of non-hydrogen backbone atoms for CDR-L3 of MOPC-315. These co-ordinates can be used directly for fitting onto the Newm framework co-ordinates provided by the Protein Data Bank*

| Amino acid position and atom | x | y | z |
|---|---|---|---|
| 89 N | −36·6 | 22·3 | 0·9 |
| CA | −37·9 | 22·7 | 1·6 |
| CB | −37·5 | 23·6 | 2·8 |
| C | −38·7 | 23·5 | 0·6 |
| O | −38·3 | 24·5 | 0·0 |
| 90 N | −40·0 | 23·0 | 0·5 |
| CA | −41·0 | 23·6 | −0·4 |
| CB | −40·6 | 23·3 | −1·8 |
| C | −42·3 | 23·0 | −0·3 |
| O | −42·5 | 22·2 | 0·9 |
| 91 N | −43·3 | 23·5 | −0·8 |
| CA | −44·7 | 23·0 | −0·7 |
| CB | −45·6 | 24·2 | −0·2 |
| C | −45·2 | 22·5 | −2·0 |
| O | −45·0 | 23·1 | −3·1 |
| 92 N | −45·8 | 21·3 | −1·9 |
| CA | −46·4 | 20·6 | −3·1 |
| CB | −45·7 | 19·2 | −3.2 |
| C | −47·9 | 20·4 | −2·9 |
| O | −48·5 | 20·8 | −1·9 |
| 93 N | −48·5 | 19·7 | −3·9 |
| CA | −49·9 | 19·5 | −3·9 |
| CB | −50·3 | 18·9 | −5·3 |
| C | −50·2 | 18·4 | −2·8 |
| O | −51·2 | 18·5 | −2·1 |
| 94 N | −49·4 | 17·4 | −2·8 |
| CA | −49·5 | 16·3 | −1·9 |
| CB | −48·3 | 15·3 | −2·1 |
| C | −49·4 | 16·8 | −0·4 |
| O | −50·4 | 16·8 | 0·3 |
| 95 N | −48·2 | 17.2 | 0·0 |
| CA | −48·0 | 17·7 | 1·3 |
| CB | −47·8 | 16·6 | 2·3 |
| C | −46·7 | 18·6 | 1·3 |
| O | −45·6 | 18·1 | 0·7 |
| 96 N | −46·8 | 19·7 | 1·9 |
| CA | −45·6 | 20·7 | 2·0 |
| CB | −46·1 | 21·9 | 2·8 |
| C | −44·5 | 20·0 | 2·7 |
| O | −43·4 | 20·6 | 2·9 |
| 97 N | −44·7 | 18·7 | 3·0 |
| CA | −43·7 | 17·9 | 3·7 |
| CB | −44·0 | 16·4 | 3·5 |
| C | −42·3 | 18·2 | 3·1 |
| O | −42·2 | 18·5 | 1·9 |

TABLE 5

*The final adjusted ($\phi$, $\psi$) angles for all CDR residues of MOPC-315 together with some adjacent residues given in parentheses. Values for CDR-L3 are omitted since they are listed in Table 1*

| CDR-L1 | ($\phi$, $\psi$) | CDR-L2 | ($\phi$, $\psi$) | CDR-H1 | ($\phi$, $\psi$) | CDR-H2 | ($\phi$, $\psi$) | CDR-H3 | ($\phi$, $\psi$) |
|---|---|---|---|---|---|---|---|---|---|
| (23 Cys) | −130·1, 133 | (46 Gly) | −86, 106 | (30 Thr) | −45·3, −96 | (47 Trp) | −83·6, 137 | (94 Gly) | −113·2, 154 |
| 24 Arg | −155, 131 | (47 Leu) | −72, −37 | 31 Ser | 38, −112 | (48 Leu) | −46, −73 | 95 Asp | −116, 28 |
| 25 Ser | −72, −41 | (48 Ile) | −152, 133 | 32 Gly | −122, 146 | (49 Gly) | −170, −158 | 96 Asn | −53, 138 |
| 26 Ser | −75, 151 | (49 Gly) | −125, 149 | 33 Tyr | −109, 152 | 50 Phe | −112, 141 | 97 Asp | −61, −56 |
| 27 Thr | −112, −16 | 50 Gly | −67, −33 | 34 Phe | −177, 176 | 51 Ile | −95, 150 | 98 His | −165, 88 |
| 27D Gly | −79, −41 | 51 Thr | −79, −38 | 35 Trp | −57, 119 | 52 Lys | −134, 137 | 99 Leu | −143, 162 |
| 27E Ala | −107, 154 | 52 Ser | −144, 111 | 35A Asn | −110, 34 | 53 Tyr | −66, −28 | 100 Tyr | −83, 32 |
| 27F Val | −63, −47 | 53 Asp | −100, −31 | (36 Trp) | −136, 91 | 54 Asp | −107, −7 | 100A Phe | −90, 138 |
| 28 Thr | −138, 139 | 54 Arg | −76, 69 | (37 Ile) | −108·9, 131·6 | 55 Gly | −35, 160 | 101 Asp | −62, 133 |
| 29 Thr | −118, 137 | 55 Ala | −75, 122 | (38 Arg) | −120·5, 164·4 | 56 Ser | −81, 128 | 102 Tyr | −72, −70 |
| 30 Ser | 53, 82 | 56 Pro | −62, 142 | | | 57 Asx | −113, −3 | (103 Trp) | −118, 127·8 |
| 31 Asn | −66, 108 | (57 Gly) | 65, 15 | | | 58 Tyr | −71, 1 | (104 Gly) | −132·2, 96·5 |
| 32 Tyr | −157, 151 | (58 Val) | −93, 137 | | | 59 Gly | −96, 142 | (105 Gln) | 17·4, 148·3 |
| 33 Ala | −69, 99 | (59 Pro) | −77, 166 | | | 60 Asx | −90, 162 | | |
| 34 Asn | −124, 91 | (60 Val) | −98, −8 | | | 61 Pro | −77, 122 | | |
| (35 Trp) | −142, 140 | (61 Arg) | −57, −17 | | | 62 Ser | −159, 165 | | |
| (36 Ile) | −149, 123 | | | | | 63 Leu | −72, −22 | | |
| (37 Gln) | −30·1, 122·8 | | | | | 64 Lys | −69, 159 | | |
| | | | | | | 65 Asn | −46, −22 | | |
| | | | | | | (66 Arg) | −70, 64 | | |
| | | | | | | (67 Val) | −109, 92·7 | | |
| | | | | | | (68 Ser) | −106·5, 134·2 | | |

For numbering, please refer to Kabat *et al.* (1979).



TABLE 6

*Atomic co-ordinates of non-hydrogen backbone atoms for all CDRs (except CDR-L3) of MOPC-315. For co-ordinates of CDR-L3 atoms, please see Table 2*

| | *x* | *y* | *z* | | *x* | *y* | *z* |
|---|---|---|---|---|---|---|---|
| CDR-L1 | | | | CDR-L1 | | | |
| 24 N | −38·3 | 15·7 | −6·6 | 28 N | −39·3 | 17·5 | −12·6 |
| CA | −39·6 | 16·0 | −7·3 | CA | −39·4 | 18·4 | −11·4 |
| CB | −39·4 | 15·8 | −8·8 | CB | −38·4 | 19·5 | −11·6 |
| C | −40·6 | 15·0 | −6·7 | C | −40·8 | 19·0 | −11·4 |
| O | −40·4 | 13·8 | −6·6 | O | −41·4 | 19·4 | −12·4 |
| 25 N | −41·8 | 15·6 | −6·4 | 29 N | −41·3 | 19·1 | −10·2 |
| CA | −42·9 | 14·9 | −5·8 | CA | −42·7 | 19·7 | −9·9 |
| CB | −43·9 | 15·9 | −5·2 | CB | −43·5 | 18·6 | −9·2 |
| C | −43·6 | 14·1 | −6·9 | C | −42·5 | 20·9 | −9·0 |
| O | −44·0 | 12·9 | −6·7 | O | −41·7 | 20·9 | −8·0 |
| 26 N | −43·7 | 14·7 | −8·1 | 30 N | −43·3 | 21·9 | −9·3 |
| CA | −44·3 | 14·1 | −9·2 | CA | −43·3 | 23·2 | −8·6 |
| CB | −44·6 | 15·2 | −10·3 | CB | −43·9 | 23·0 | −7·2 |
| C | −43·3 | 13·1 | −9·8 | C | −41·9 | 23·7 | −8·5 |
| O | −42·1 | 13·2 | −9·7 | O | −41·2 | 23·5 | −7·5 |
| 27 N | −43·9 | 12·1 | −10·5 | 31 N | −41·4 | 24·3 | −9·6 |
| CA | −43·2 | 11·0 | −11·1 | CA | −40·1 | 24·9 | −9·7 |
| CB | −43·8 | 9·6 | −10·6 | CB | −39·9 | 25·5 | −11·1 |
| C | −43·4 | 11·1 | −12·6 | C | −40·0 | 26·1 | −8·7 |
| O | −42·6 | 10·4 | −13·4 | O | −40·6 | 27·2 | −8·7 |
| 27DN | −44·4 | 11·9 | −13·0 | 32N | −39·2 | 25·8 | −7·6 |
| CA | −44·7 | 12·1 | −14·4 | CA | −38·9 | 26·8 | −6·6 |
| C | −43·7 | 13·1 | −15·0 | CB | −40·1 | 26·8 | −5·5 |
| O | −43·3 | 13·0 | −16·2 | C | −37·6 | 26·4 | −5·8 |
| 27EN | −43·4 | 14·1 | −14·2 | O | −37·2 | 25·3 | −5·8 |
| CA | −42·5 | 15·2 | −14·6 | 33 N | −37·0 | 27·5 | −5·3 |
| CB | −43·1 | 16·5 | −14·2 | CA | −35·7 | 27·4 | −4·6 |
| C | −41·2 | 15·0 | −13·8 | CB | −35·2 | 28·8 | −4·3 |
| O | −41·2 | 14·4 | −12·7 | C | −35·9 | 26·6 | −3·3 |
| 27FN | −40·1 | 15·6 | −14·4 | O | −36·4 | 27·2 | −2·3 |
| CA | −38·8 | 15·5 | −13·9 | 34 N | −35·6 | 25·4 | −3·3 |
| CB | −37·8 | 16·2 | −14·9 | CA | −35·8 | 24·5 | −2·1 |
| C | −38·8 | 16·2 | −12·5 | CB | −36·8 | 23·4 | −2·4 |
| O | −38·2 | 15·8 | −11·5 | C | −34·4 | 23·8 | −1·8 |
| | | | | O | −34·1 | 22·7 | −2·4 |

TABLE 6 cont.

| CDR-L2 | | | | |
|---|---|---|---|---|
| 50 | N | −36·4 | 30·5 | −1·2 |
| | CA | −37·4 | 31·0 | −2·2 |
| | C | −37·8 | 32·5 | −1·9 |
| | O | −38·0 | 33·2 | −2·9 |
| 51 | N | −37·7 | 32·8 | −0·6 |
| | CA | −38·0 | 34·2 | −0·2 |
| | CB | −38·3 | 34·2 | 1·3 |
| | C | −36·8 | 35·1 | −0·4 |
| | O | −36·9 | 36·2 | −0·8 |
| 52 | N | −35·6 | 34·4 | −0·2 |
| | CA | −34·3 | 35·1 | −0·3 |
| | CB | −33·8 | 35·6 | 1·0 |
| | C | −33·3 | 34·1 | −0·9 |
| | O | −32·9 | 33·2 | −0·2 |
| 53 | N | −32·9 | 34·3 | −2·1 |
| | CA | −31·9 | 33·5 | −2·8 |
| | CB | −32·4 | 33·3 | −4·3 |
| | C | −30·6 | 34·2 | −2·8 |
| | O | −29·5 | 33·5 | −2·8 |
| 54 | N | −30·6 | 35·5 | −2·8 |
| | CA | −29·4 | 36·3 | −2·7 |
| | CB | −29·8 | 37·8 | −3·1 |
| | C | −28·8 | 36·3 | −1·3 |
| | O | −28·8 | 37·3 | −0·6 |
| 55 | N | −28·3 | 35·1 | −1·0 |
| | CA | −27·7 | 34·9 | 0·4 |
| | CB | −27·5 | 33·4 | 0·6 |
| | C | −26·3 | 35·6 | 0·4 |
| | O | −25·4 | 35·2 | −0·4 |
| 56 | N | −26·2 | 36·4 | 1·4 |
| | CA | −24·9 | 37·2 | 1·6 |
| | CB | −25·1 | 38·2 | 2·8 |
| | C | −23·8 | 36·2 | 2·0 |
| | O | −24·0 | 35·2 | 2·7 |

| CDR-H1 | | | | |
|---|---|---|---|---|
| 31 | N | −44·2 | 39·8 | 13·1 |
| | CA | −43·8 | 39·5 | 11·7 |
| | CB | −42·4 | 38·9 | 11·7 |
| | C | −44·7 | 38·5 | 11·1 |
| | O | −45·9 | 38·8 | 10·8 |
| 32 | N | −44·2 | 37·3 | 10·8 |
| | CA | −44·9 | 36·2 | 10·2 |
| | C | −44·9 | 35·0 | 11·2 |
| | O | −43·9 | 34·8 | 11·9 |
| 33 | N | −46·0 | 34·2 | 11·2 |
| | CA | −46·1 | 33·0 | 12·0 |
| | CB | −47·5 | 33·1 | 12·7 |
| | C | −46·1 | 31·8 | 11·1 |
| | O | −46·4 | 31·8 | 9·9 |
| 34 | N | −45·7 | 30·7 | 11·7 |
| | CA | −45·5 | 29·4 | 11·1 |
| | CB | −44·5 | 29·5 | 10·0 |
| | C | −45·1 | 28·4 | 12·1 |
| | O | −44·8 | 28·7 | 13·3 |
| 35 | N | −45·0 | 27·1 | 11·6 |
| | CA | −44·6 | 26·0 | 12·4 |
| | CB | −44·6 | 24·7 | 11·6 |
| | C | −43·2 | 26·2 | 13·0 |
| | O | −42·2 | 26·4 | 12·3 |
| 35A | N | −43·2 | 26·2 | 14·4 |
| | CA | −41·9 | 26·5 | 15·1 |
| | CB | −42·2 | 27·6 | 16·2 |
| | C | −41·5 | 25·2 | 15·8 |
| | O | −41·0 | 25·2 | 16·9 |

TABLE 6 cont.

| CDR-H2 | | | | |
|---|---|---|---|---|
| 50 | N | −48·9 | 21·8 | 10·5 |
| | CA | −49·4 | 23·2 | 10·2 |
| | CB | −50·3 | 23·1 | 8·9 |
| | C | −50·2 | 23·8 | 11·3 |
| | O | −50·9 | 23·1 | 12·0 |
| 51 | N | −49·9 | 25·1 | 11·5 |
| | CA | −50·6 | 25·9 | 12·6 |
| | CB | −49·6 | 26·9 | 13·1 |
| | C | −51·8 | 26·5 | 12·0 |
| | O | −51·9 | 26·9 | 10·8 |
| 52 | N | −52·8 | 26·7 | 12·9 |
| | CA | −54·0 | 27·4 | 12·6 |
| | CB | −55·2 | 26·3 | 12·6 |
| | C | −54·4 | 28·4 | 13·7 |
| | O | −54·2 | 28·2 | 14·9 |
| 53 | N | −54·8 | 29·6 | 13·2 |
| | CA | −55·2 | 30·7 | 14·1 |
| | CB | −55·6 | 31·9 | 13·2 |
| | C | −56·4 | 30·3 | 14·9 |
| | O | −56·6 | 30·8 | 16·0 |
| 54 | N | −57·2 | 29·4 | 14·4 |
| | CA | −58·4 | 28·9 | 15·0 |
| | CB | −59·6 | 28·9 | 14·0 |
| | C | −58·2 | 27·5 | 15·5 |
| | O | −59·0 | 26·9 | 16·2 |
| 55 | N | −57·0 | 26·9 | 15·1 |
| | CA | −56·6 | 25·6 | 15·4 |
| | C | −57·0 | 25·3 | 16·9 |
| | O | −57·3 | 26·2 | 17·7 |
| 56 | N | −57·1 | 24·0 | 17·1 |
| | CA | −57·5 | 23·4 | 18·4 |
| | CB | −57·9 | 22·0 | 18·3 |
| | C | −56·3 | 23·5 | 19·4 |
| | O | −55·2 | 23·0 | 19·1 |
| 57 | N | −56·6 | 24·1 | 20·6 |
| | CA | −55·6 | 24·3 | 21·6 |
| | CB | −55·6 | 25·7 | 22·0 |
| | C | −56·0 | 23·4 | 22·8 |
| | O | −55·3 | 23·4 | 23·8 |

| CDR H2 | | | | |
|---|---|---|---|---|
| 58 | N | −57·2 | 22·8 | 22·7 |
| | CA | −57·7 | 21·9 | 23·7 |
| | CB | −59·2 | 21·5 | 23·3 |
| | C | −56·9 | 20·6 | 23·8 |
| | O | −57·2 | 19·7 | 24·6 |
| 59 | N | −55·9 | 20·6 | 23·0 |
| | CA | −54·9 | 19·5 | 23·0 |
| | C | −53·7 | 19·8 | 23·8 |
| | O | −53·2 | 20·9 | 23·9 |
| 60 | N | −53·3 | 18·7 | 24·5 |
| | CA | −52·1 | 18·8 | 25·4 |
| | CB | −52·3 | 17·7 | 26·5 |
| | C | −50·8 | 18·5 | 24·6 |
| | O | −50·9 | 17·9 | 23·5 |
| 61 | N | −49·7 | 18·9 | 25·2 |
| | CA | −48·4 | 18·7 | 24·6 |
| | CB | −47·4 | 19·6 | 25·4 |
| | C | −47·9 | 17·2 | 24·8 |
| | O | −47·9 | 16·7 | 25·9 |
| 62 | N | −47·7 | 16·6 | 23·7 |
| | CA | −47·2 | 15·2 | 23·6 |
| | CB | −48·4 | 14·3 | 23·7 |
| | C | −46·5 | 15·0 | 22·3 |
| | O | −46·6 | 15·8 | 21·3 |
| 63 | N | −45·8 | 13·9 | 22·2 |
| | CA | −45·0 | 13·5 | 21·0 |
| | CB | −44·1 | 12·3 | 21·4 |
| | C | −46·0 | 13·0 | 19·9 |
| | O | −45·6 | 13·0 | 18·7 |
| 64 | N | −47·2 | 12·7 | 20·4 |
| | CA | −48·3 | 12·2 | 19·5 |
| | CB | −49·4 | 11·7 | 20·3 |
| | C | −48·7 | 13·4 | 18·6 |
| | O | −48·5 | 14·5 | 19·0 |
| 65 | N | −49·4 | 13·0 | 17·5 |
| | CA | −49·9 | 14·0 | 16·5 |
| | CB | −50·8 | 13·2 | 15·6 |
| | C | −50·7 | 15·1 | 17·3 |
| | O | −50·9 | 16·2 | 16·8 |

TABLE 6 cont.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CDR-H3 | | | | | CDR-H3 | | | |
| 95 | N | −39·7 | 31·7 | 10·0 | 100 | N | −32·7 | 39·1 | 7·9 |
| | CA | −40·4 | 31·6 | 8·7 | | CA | −31·4 | 38·9 | 8·4 |
| | CB | −41·8 | 31·1 | 8·9 | | CB | −30·5 | 39·8 | 8·6 |
| | C | −40·5 | 33·0 | 8·0 | | C | −31·6 | 37·9 | 9·7 |
| | O | −41·3 | 33·3 | 7·2 | | O | −30·7 | 37·8 | 10·5 |
| 96 | N | −39·5 | 33·8 | 8·4 | 100A | N | −32·7 | 37·2 | 9·8 |
| | CA | −39·3 | 35·2 | 7·9 | | CA | −33·1 | 36·4 | 11·0 |
| | CB | −38·0 | 35·8 | 8·4 | | CB | −34·6 | 36·4 | 11·2 |
| | C | −39·3 | 35·2 | 6·4 | | C | −32·6 | 34·9 | 10·8 |
| | O | −38·7 | 34·3 | 5·8 | | O | −32·8 | 34·4 | 9·7 |
| 97 | N | −40·1 | 36·1 | 5·8 | 101 | N | −32·1 | 34·4 | 11·8 |
| | CA | −40·2 | 36·3 | 4·4 | | CA | −31·5 | 33·0 | 11·8 |
| | CB | −41·1 | 37·5 | 4·1 | | CB | −30·9 | 32·7 | 13·2 |
| | C | −38·8 | 36·6 | 3·8 | | C | −32·7 | 32·0 | 11·6 |
| | O | −38·4 | 35·8 | 2·9 | | O | −33·8 | 32·1 | 12·1 |
| 98 | N | −38·2 | 37·6 | 4·3 | 102 | N | −32·4 | 31·1 | 10·6 |
| | CA | −36·8 | 38·0 | 3·8 | | CA | −33·4 | 30·1 | 10·2 |
| | CB | −37·0 | 38·7 | 2·5 | | CB | −32·8 | 29·4 | 9·0 |
| | C | −36·2 | 38·9 | 4·9 | | C | −33·5 | 29·1 | 11·4 |
| | O | −36·4 | 40·2 | 4·8 | | O | −34·6 | 29·0 | 12·0 |
| 99 | N | −35·6 | 38·3 | 5·8 | | | | | |
| | CA | −34·9 | 39·1 | 6·9 | | | | | |
| | CB | −35·8 | 39·1 | 8·1 | | | | | |
| | C | −33·6 | 38·4 | 7·3 | | | | | |
| | O | −33·4 | 37·2 | 6·9 | | | | | |

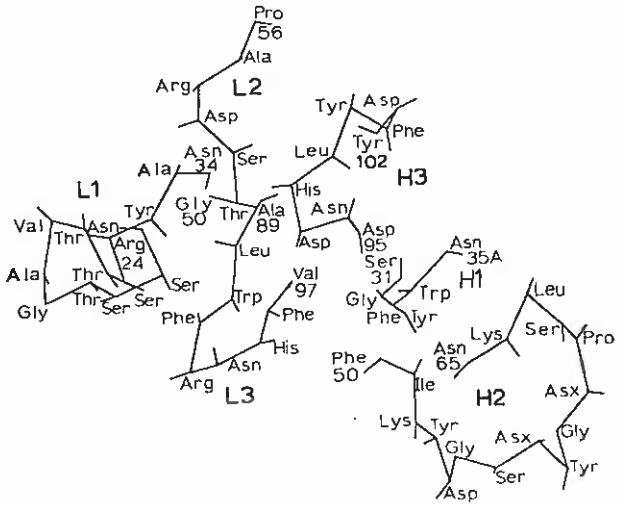



FIG. 3. View of the antibody combining site of immunoglobulin MOPC-315. The various CDRs are indicated together with the amino acid residues. Only $C_\alpha$ and $C_\beta$ atoms are shown. This view is straight down into the binding cavity.



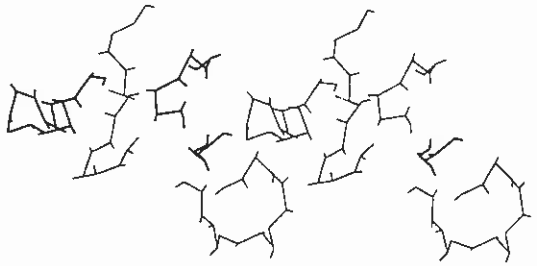


FIG. 4. Stereo-pair of the combining site as shown in Fig. 3. The same view is given so that comparing with Fig. 3, one can identify the locations of CDR's and amino acid residues.

hapten binding site, and that a Trp residue was in the proximity of the site (Eisen, Simms & Potter, 1968; Goetzl & Metzger, 1970*a*, *b*; Haimovich, Givol & Eisen, 1970; Givol *et al.*, 1971; Glaser & Singer, 1971; Haimovich *et al.*, 1972; Rockey *et al.*, 1972; Inbar, Givol & Hochman, 1973; Klostergaard *et al.*, 1977). Furthermore, the binding site was found to be asymmetric with a positively charged amino acid residue at the entrance and to have a depth of approximately 11 Å, an opening of about 9 Å × 10 Å and a bottom only about 3 Å wide (Haimovich *et al.*, 1972; Haselkorn *et al.*, 1974; Pecht, 1975; Dwek *et al.*, 1977; Sutton *et al.*, 1977). All these experimental results could be accomodated by the present model (Figs 3 and 4).

The combining site of MOPC-315 has been predicted previously by Padlan *et al.* (1976) and modified by Dwek *et al.* (1977), on the basis that the CDR structures of Mcg (Edmunson *et al.*, 1974), Newm (Poljak *et al.*, 1974), McPC-603 (Segal *et al.*, 1974) and REI (Epp *et al.*, 1974) determined by X-ray diffraction studies could be used directly for MOPC-315 as long as the lengths of the CDRs were the same. Our present predictive method may therefore be of more general utility since the lengths of CDRs of different immunoglobulins vary extensively (Kabat, Wu & Bilofsky, 1979) while only a few tertiary structures of CDRs have been determined by X-ray crystallography (Protein Data Bank). Roughly speaking, CDR-H1 and CDR-H3 in the model of Padlan *et al.* (1976) looked similar to the present predictions, probably due to the large distances between the points of attachment to the framework and the small number of amino acid residues in these two CDRs so that both were quite extended. Otherwise, the remaining four CDRs were all quite different.

It must be emphasized that the configurations of the side chains (Dwek *et al.*, 1977) and the influence of the antigen on the antibody conformation are beyond the present predictive method, though the side chains, especially those pointing into the antibody combining site, may have some degree of

freedom to change configurations. Furthermore, as more three-dimensional structures (Protein Data Bank) are determined by X-ray diffraction studies, the number of possible ($\phi$, $\psi$) choices for each position will be increased and possibly other sterically allowable foldings of various CDR's may be found. In addition, the framework structure may be slightly different from the standard one (Wang, Yoo & Sax, 1979). However, the present method represents a quantitative and logical approach to predicting three dimensional foldings of backbones around antibody combining sites on the basis of the original principle of Anfinsen & Redfield (1956). The validity of the presently predicted conformation for MOPC-315's binding site must ultimately be judged on the basis of its agreement with X-ray diffraction analysis (Aschaffenburg *et al.*, 1979). Any difference may then be used to improve upon our predictive method.

This work was supported in part by NIH Grant 5-R01-GM21842, and NIH Contracts N01-RR-4-2147 and N01-RR-8-2158 (with PROPHET computer system). TTW was a Research Career Development Awardee, NIH 5-K01-AI70497.

## REFERENCES

Anfinsen, C. B. & Redfield, R. R. (1956). *Adv. Protein Chem.* **11,** 1.
Anfinsen, C. B. & Scheraga, H. A. (1975). *Adv. Protein Chem.* **29,** 205.
Aschaffenburg, R., Phillips, D. C., Rose, D. R., Sutton, B. J., Dower, S. K. & Dwek, R. A. (1979). *Biochem. J.* **181,** 497.
Chou, P. Y. & Fasman, G. D. (1978). *Ann. Rev. Biochem.* **47,** 251.
Dickerson, R. E. & Geis, I. (1969). *The Structure and Action of Proteins.* Menlo Park, CA: W. A. Benjamin, Inc.
Dwek, R. A., Wain-Hobson, S., Dower, S., Gettens, P., Sutton, B., Perkins, S. J. & Givol, D. (1977). *Nature* **266,** 31.
Edmundson, A. B., Ely, K. R., Girling, R. L., Abola, E. E., Schiffer, M., Westholm, F. A., Fausch, M. D. & Deutsch, H. F. (1974). *Biochemistry* **13,** 3816.
Eisen, J. N., Simms, E. S. & Potter, M. (1968). *Biochemistry* **7,** 4126.
Epp, O., Coleman, P., Fehlhammer, H., Bode, W., Schiffer, M., Huber, R. & Palm, W. (1974). *Eur. J. Biochem.* **45,** 513.
Givol, D., Strausbauch, P. H., Hurwitz, E., Wilchek, M., Haimovich, J. & Eisen, H. N. (1971). *Biochemistry* **10,** 3461.
Glaser, M. & Singer, S. J. (1971). *Proc. natn. Acad. Sci. U.S.A.* **68,** 2477.
Goetzl, E. J. & Metzger, H. (1970*a*). *Biochemistry* **9,** 1267.
Goetzl, E. J. & Metzger, H. (1970*b*). *Biochemistry* **9,** 3862.
Haimovich, J., Givol, D. & Eisen, H. N. (1970). *Proc. natn. Acad. Sci. U.S.A.* **67,** 1656.
Haimovich, J., Eisen, H. N., Hurwitz, E. & Givol, D. (1972). *Biochemistry* **11,** 2389.
Haselkorn, D., Friedman, S., Givol, D. & Pecht, U. (1974). *Biochemistry* **13,** 2210.
Honig, B., Kabat, E. A., Katz, L., Levinthal, C. & Wu, T. T. (1973). *J. mol. Biol.* **80,** 277.
Inbar, D., Givol, D. & Hochman, J. (1973). *Biochemistry* **12,** 4541.
Johnson, M. E. & Wu, T. T. (1976). *J. theor. Biol.* **60,** 183.
Kabat, E. A. (1978). *Adv. Protein Chem.* **32,** 1.
Kabat, E. A. & Wu, T. T. (1971). *Ann. New York Acad. Sci.* **190,** 382.



Kabat, E. A. & Wu, T. T. (1972). *Proc. natn. Acad. Sci. U.S.A.* **69,** 960.
Kabat, E. A., Wu, T. T. & Bilofsky, H. (1979). *Sequences of Immunoglobulin Chains.* Bethesda, MD: NIH Publication No. 80–2008.
Levitt, M. & Crothia, C. (1976). *Nature* **261,** 552.
Klostergaard, J., Grossberg, A. L., Krausz, L. M. & Pressman, D. (1977). *Immunochem.* **14,** 37.
Padlan, E. A., Davies, D. R., Pecht, I., Givol, D. & Wright, C. (1976). *Cold Spring Harbor Symp. Quant. Biol.* **41,** 627.
Pecht, I. (1975). In *Protein–Ligand Interactions,* (H. Sund & G. Blauer, eds), p. 356. Berlin: Walter de Gruyter.
Poljak, R. J., Amzel, L. M., Avey, H. P., Becka, L. N., Goldstein, D. J. & Humphrey, R. L. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36,** 421.
Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerley, R. P. & Saul, F. (1973). *Proc. natn. Acad. Sci. U.S.A.* **70,** 3305.
Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P. & Saul, F. (1974). *Proc. natn. Acad. Sci. U.S.A.* **71,** 3440.
Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P. & Saul, F. (1975). *Proc. Roy. Soc., Lond.* **B272,** 43.
Poljak, R. J., Makashima, Y., Chen, B. L. & Konigsberg, W. (1977). *Biochemistry* **16,** 3413.
Rockey, J. H., Montgomery, P. C., Underdown, B. J. & Dorrington, K. L. (1972). *Biochemistry* **11,** 3172.
Saul, F. A., Amzel, L. M. & Poljak, R. J. (1978). *J. biol. Chem.* **253,**585.
Schulz, G. E. (1977). *Angew. Chem. Int. Ed.* **16,** 23.
Schulz, G. E., Barry, C. D., Friedman, J., Chou, P. Y., Fasman, G. D., Finkelstein, A. V., Lim, V. I., Ptitsyn, O. B., Kabat, E. A., Wu, T. T., Levitt, M., Robson, B. & Nagano, K. (1974). *Nature* **250,** 140.
Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. (1974). *Proc. natn. Acad. Sci. U.S.A.* **71,** 4298.
Sternberg, M. J. E. & Thornton, J. M. (1978). *Nature* **271,** 15.
Sutton, B. J., Gettins, P., Givol, D., Marsh, D. Wain-Hobson, S., William, K. J. & Dwek, R. A. (1977). *Biochem. J.* **165,** 177.
Wang, B. C., Yoo, C. S. & Sax, M. (1979). *J. mol. Biol.* **129,** 657.
Wu, T. T., Fitch, W. M. & Margoliash, E. (1974). *Ann. Rev. Biochem.* **43,** 539.
Wu, T. T. & Kabat, E. A. (1970). *J. exp. Med.* **132,** 211.
Wu, T. T. & Kabat, E. A. (1973). *J. mol. Biol.* **75,** 13.

**Engineering design of therapeutic antibodies**

For an antigen with known three-dimensional structure, it is then possible to estimate the required amino acid sequence most likely to interact with a certain part of the surface of the antigen. Amino acid residues in that sequence can be hydrophobic, hydrophilic, positively or negatively charged. The CDRH3 will play the post important role of binding. For cavities on the antigen surface, CDRH3 would be longer, and for protrusions shorter. As shown on the next two pages, increasing the length of CDRH3 results in a more extended form of that complementarity determining region.

Another important factor is that the chosen amino acid sequence for CDRH3 should not be antigenic to humans. One of the possible guide-lines is that sequence may be present in an existing human protein. Currently with the human genome project essentially completed and most of the known human proteins available in the GenBank database, the BLAST programs can use for such matches.

The predictive method for predicting folding of CDR's as outlined in the previous section can then be used to determine the three-dimentionsl structure of that chosen CDRH3. It should fit well to one special part of the surface of the antigen molecule. Afterwards, other CDR's may also be estimated. In these cases, most likely they would improve the binding affinity of the antibody but not change the specificity.

This approach requires massive amounts of computation and model fitting. After the six CDR's have been decided on, the production of the antibody molecule needs to make use of genetic engineering. The amino acid sequences of the CDR's are required to be converted to nucleotide sequences which are then ligated to the framework nucleotide sequences of one of the existing human antibodies. The entire segments of the light and heavy chain variable region gene segments are inserted into plasmid molecules containing the constant region gene segments of the light and heavy chains respectively.

While the above mentioned approach is complete feasible, commercial companies have tried various methods to short circuit the basic process of finding the required CDR sequences. Indeed, currently, most of the pharmaceutical companies have developed simpler selecting processes to identify the required therapeutic antibodies. They consist of several different categories to be discussed subsequently:

1. Mouse
2. Chimeric
3. Humanized
4. Human

. Stereo view of CDRH3 with five amino acid residues.

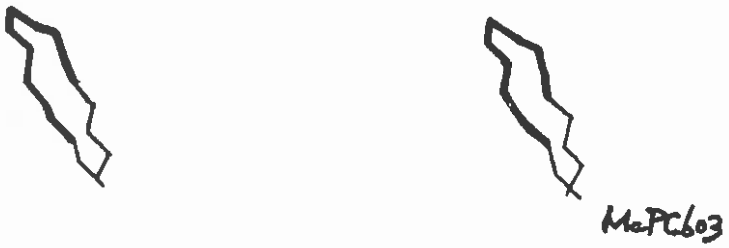
Stereo view of CDRH3 with eight amino acid residues.

Stereo view of CDRH3 with six amino acid residues.

Stereo view of CDRH3 with nine amino acid residues.

Stereo view of CDRH3 with seven amino acid residues.

Stereo view of CDRH3 with ten amino acid residues.




Stereo view of CDRH3 with eleven amino acid residues.

Stereo view of CDRH3 with twelve amino acid residues.

Stereo view of CDRH3 with fifteen amino acid residues.

Stereo view of CDRH3 with seventeen amino acid residues.

Protein Engineering vol.13 no.12 pp.819–824, 2000

# WAM: an improved algorithm for modelling antibodies on the WEB

**Nicholas R.J.Whitelegg and Anthony R.Rees[1]**

Centre for Protein Analysis and Design,University of Bath, Claverton Down, Bath BA2 7AY, UK
[1]To whom correspondence should be addressed. Present address: Syntem, Parc Scientifique Georges Besse, 30000 Nimes, France. E-mail: arrees@compuserve.com

**An improved antibody modelling algorithm has been developed which incorporates significant improvements to the earlier versions developed by Martin *et al.* (1989, 1991), Pedersen *et al.* (1992) and Rees *et al.* (1996) and known as AbM (Oxford Molecular). The new algorithm, WAM (for Web antibody modelling), has been launched as an online modelling service and is located at URL *http://antibody.bath.ac.uk*. Here we provide a summary only of the important features of WAM. Readers interested in further details are directed to the website, which gives extensive background information on the methods employed. A brief description of the rationale behind some of the newer methodology (specifically, the knowledge-based screens) is also given.**
*Keywords*: antibody modelling algorithm/knowledge-based screens/WAM/WEB

## Introduction

Predicting the structure of the antibody variable region from sequence has been a subject of considerable activity since the work of Kabat and Wu (1972) (prediction of a kappa light chain), Padlan *et al.* (1976), Stanford and Wu (1981) and Feldmann *et al.* (1981). Following the essentially 'homology'-based predictions of these early approaches, methods were developed that introduced more rule-based procedures exemplified by our own work (Darsley and Rees, 1985; de la Paz *et al.*, 1986; Martin *et al.*, 1989, 1991; Rees *et al.*, 1990, 1996; Pedersen *et al.*, 1992) and that of Chothia, Lesk and colleagues (Chothia and Lesk, 1987; Chothia *et al.*, 1989, 1992; Tomlinson *et al.*, 1995; AlLazikani *et al.*, 1997), who, building on the observations of Wu and Kabat (1970), Kabat and Wu (1972), Kabat *et al.* (1977) and Padlan and Davies (1975), developed the concept of canonical classes for certain of the variable-region CDRs. Since then, attention has focused on the more difficult problem of non-canonical CDRs, of which the antibody heavy chain CDR3 (hereafter referred to as H3) is the most non-conformist.

Antibody modelling has an advantage over protein modelling in general in that only the Fv needs to be modelled, the constant region being conserved. Further, the majority of the Fv itself, the framework, is highly conserved in structure between different antibodies and can be modelled using the most sequence-homologous known framework (see Pedersen *et al.*, 1994). For CDR modelling, five of the six CDRs (all except H3) frequently fall into one of between one and 10 *canonical* classes, a set for each CDR. Members of a canonical class all have approximately the same backbone conformation. This is determined by the loop length and the presence of a number of key residues, both in the CDR and in the framework (Foote and Winter, 1992), which hold the CDR in a given conformation by hydrogen bonding, electrostatic and/or hydrophobic interactions. So, to model an unknown CDR, the sequence is examined, the appropriate canonical class assigned and the most sequence-homologous known CDR used. For each loop except L2, a few examples fall outside existing canonical classes, and, along with the H3 loop, must be modelled in other ways (see below). It is likely that further canonical classes will be revealed as more crystal structures are solved, although to date no strictly 'canonical' classification has been possible for H3.

The difficulty of modelling the H3 loop is due to the extensive variability in both sequence and structure between different antibodies. There are essentially three approaches: *knowledge-based* methods, such as database searching, where the closest matching database loop [either from antibodies or from the entire Protein Data Bank (PDB) (Berman *et al.*, 2000)] in sequence and length is used as the model (e.g. as in ABGEN; Mandal *et al.*, 1996) or *ab initio* methods, such as the CONGEN conformational search (Bruccoleri and Karplus, 1987), or a combination of both (Martin *et al.*, 1991). More recent analysis suggests that eventually, knowledge-based methods will provide the degree of accuracy required for most applications of modelling, such as antibody humanization (Reichmann *et al.*, 1988; Pedersen *et al.*, 1994; Roguska *et al.*, 1994, 1996). However, for the time being, a combination of empirical and *ab initio* methods is likely to be necessary where accuracy approaching that of a medium-resolution X-ray structure is required.

## Methods and results

The algorithm used within WAM (Figure 1) is based on that of AbM (Martin *et al.*, 1989, 1991; Pedersen *et al.*, 1992; Rees *et al.*, 1996). The original implementation, as found in the software package AbM (Oxford Molecular/MSI Pharmacopeia), is summarized below:

1. The framework (both backbone and side chains) and canonical loop backbones are built using the most sequence-homologous known structures.
2. The canonical loop side chains are built using an iterative placement algorithm to achieve the global minimum energy conformation, implemented within CONGEN.
3. For each non-canonical loop, a series of possible conformations are constructed using either a PDB database search or a combined database/conformational search (in which the apex is rebuilt using CONGEN), depending on the loop length.
4. The conformations generated above are energy-screened using Eureka, a solvent-modified version of the Valence Force Field (VFF) (Dauber-Osguthorpe *et al.*, 1988).

**Abbreviations:** CDR, complementarity determining region; Fv, antibody fragment containing the variable region; L1-3 and H1-3, light- and heavy-chain CDRs, respectively; VFF, valence force field; r.m.s.d., root mean square deviation.

© Oxford University Press



163

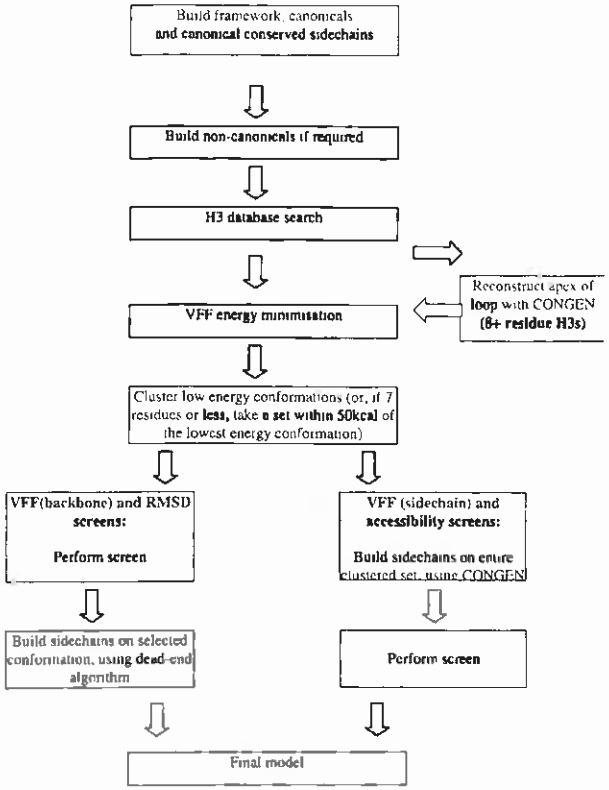



**Fig. 1.** Flow chart of the algorithm.

5. The final conformation is selected from the five lowest energy conformations by comparing the torsion angles with those observed in the original set of loops extracted from the PDB. The conformation with the closest set of torsion angles is selected as the final model.

In the following sections, we outline the current procedure used within WAM and indicate where the protocol differs from AbM.

*Framework regions and canonical CDRs*

The framework and canonical backbones are built as in AbM. Following the observation that chi-1 torsion angles of certain, but not all, residue side chains in canonical loops are conserved (see Web site), the CONGEN iterative side chain placement algorithm (Bruccoleri and Karplus, 1987), used in the original AbM, was replaced by a procedure in which the side chains are built using a chi-1 biased maximum-overlap procedure. This attempts to overlap as many as possible of the atoms of the side chain being modelled with the template side chain, while avoiding steric clashes. The non-conserved side chains are excluded at this stage and built later, at the same time as the non-conserved H3 side chains (see below).

Most of the algorithm developments relate to modelling of non-canonical CDRs, in particular H3. Analysis of antibody structures has revealed the existence of certain preferred interactions across the loop (Rees *et al.*, 1996) and, further, that the majority of H3 loops have a kinked backbone conformation, involving the four C-terminal residues, that can be defined by a set of sequence-based rules (Shirai *et al.*, 1996; Morea *et al.*, 1998). For loops of up to 11 residues, we have observed that the majority show semi-conserved conformations (up to 2.5 Å r.m.s.d. between individual examples) and solvent accessibility profiles (Whitelegg, 1998). This has allowed a modification of the published 'kinked' rules for H3 loops leading to an improved prediction. The analysis on which this improvement is based is described in more detail below.

For other non-canonical examples of L1, L2, L3, H1 and H2, further small but significant changes to the original AbM algorithm have been made. First, construction of these CDRs uses only the C-alpha to C-alpha database search, replacing the more complicated combined database/conformational search algorithm (CAMAL; Martin *et al.*, 1989, 1991). This has been made possible because the number of antibody structures is now much greater than when the CAMAL algorithm was introduced, where a conformational search procedure was found to be essential to saturate the available CDR conformational space. A series of benchmarks (see Website) using the database method has revealed that no deterioration in accuracy of these non-canonical CDRs is seen but considerable savings in computer time are achieved. Second, the stand-alone implementation of the VFF (Dauber-Osguthorpe *et al.*, 1988) is used to screen the energies of the conformations generated, rather than the modified Eureka implementation used in AbM.

*Modified algorithm for H3 loops*

H3 loops predicted to belong to the 'kinked' family on the basis of sequence rules are built using the modified algorithm presented below, using either CAMAL (eight residue loops or longer) or a Calpha–Calpha database search (seven residue loops or shorter). Loops falling outside the 'kinked' definition are built using a further modification in which the full range of H3 loops is used to generate the database search constraints. The final set of conformations is then screened by energy, rather than the newer knowledge-based methods (see below).

*Database search.*

The Calpha–Calpha database search (Martin *et al.*, 1989) is performed, with two modifications. First, only the set of 'kinked' H3 crystal structures is used to derive the distance constraints, in order to bias the database hits in favour of 'kinked' structures. Second, the database hits are subjected to 45 rounds of steepest-descent minimization after grafting on to the framework. Minimization, not included in AbM, has been introduced to avoid the situation where some low r.m.s.d. database loops are rejected because they sit above an energy threshold after construction by the CONGEN procedure.

*Conformational search and energy minimization.*

The CONGEN (Bruccoleri and Karplus, 1987) and chain closure (Go and Scheraga, 1970) algorithms are used (as in AbM) to rebuild the central five-residue segment of each database loop after its excision. However, the definition of 'central' is not fixed. For each candidate, the rebuild region is moved to contain those residues with the highest solvent exposure; for example, investigations of known 'kinked' H3 structures has revealed that the most exposed regions lie towards the N-terminal regions of these loops.

A preliminary set of side chain conformations are built, using the CONGEN iterative algorithm, to enable an initial screen of loop candidates to be carried out; previous work (Whitelegg, 1998) has shown that building the final side chain conformations on loops before they have been fully minimized can lead to steric clashes.

Each loop conformation is then subjected to 15 rounds of steepest-descent minimization by VFF to relieve any residual high energy in the loops and the conformations are clustered

to remove duplicates using a modification of the original AbM procedure.

*Clustering.*
In the first step, all conformations within a starting value, typically within 0.5Å r.m.s.d. of the first conformation, are 'clustered' with the first conformation; the first conformation to be excluded from the first conformational set forms the basis of the second cluster where all conformations within 0.5 r.m.s.d. of it are clustered with that conformation, and so on. Then the median conformation of each initial cluster is calculated and clusters are merged if their median conformations (as opposed to an arbitrary first conformation of the cluster, as in AbM) are within the clustering r.m.s.d./torsion resolution (a step value higher than the previous stage). This is repeated until a target number of clusters, dependent on the initial number of conformations, are produced. Where structures are modelled using the database search only, a clustering process is already included. However, an energy screen is performed so that only conformations within 50 kcal of the lowest energy of the set are retained.

*Final screening.*
Three methods are used to screen the clustered sets of conformations:
1. an accessibility profile is calculated and used;
2. a screen based on r.m.s.d. to the known 'kinked' structures of the particular length concerned is performed;
3. a simple VFF energy screen is performed on the resulting models, with or without side chains.

*Accessibility profiles.*
It has been observed (Whitelegg, 1998) among 'kinked' antibody H3 crystal structures ranging in length from 7 to 14 residues that certain residues are frequently exposed (>30% relative accessibility), whereas others are persistently buried (<30% relative accessibility). Using this accessibility 'pattern', a screen for the conformations generated has been devised. First, side chains are built on each clustered conformation (see below). Each subsequent conformation is scored based on the amount that the accessibility of each key residue deviates from the mean observed accessibility in crystal structures, divided by the standard deviation as a measure of the certainty of the observed accessibilities:

$$\text{score} = \Sigma_{\text{key res}}(O - M)^2/\sigma^f$$

where 'key res' are the key residues defining the accessibility pattern, $O$ is the observed accessibility for each residue in the model, $M$ is the mean accessibility for that residue and that loop length in the crystal structures, $\sigma$ is the standard deviation and $f$ is the 'standard deviation factor' (sdf), a power by which the standard deviation is raised to influence its effect on the score: an sdf of 2 has been found to be most effective (Whitelegg, 1998).

The accessibility screen is an effective test for a 'kink' (Shirai *et al.*, 1996), since the features of the 'kink' include a buried H3:C – 2 (C-terminus minus 2) residue and an exposed H3:C-terminal residue, but it also conveys the additional information of the solvent exposure of the central part of the loop and whether or not the N-terminal residue is buried.

*r.m.s.d. screen.*
This screen is based on similarity, as measured by r.m.s.d., to the known H3 structures having the same length as the modelling candidate. It has been observed that H3 loops with the same or similar lengths have similar conformations in many cases (described in detail on the Website; first seen by de la Paz *et al.*, 1986). A similarity score for each of the clustered conformations is calculated:

$$\text{score} = \Sigma_{\text{all known structs}}(1/d_x)\, m_x$$

where $d_x$ is the RMS deviation between the model and structure $x$ and $m_x$ is the sequence similarity (using the Dayhoff mutation matrix, MDM) between the sequence of the model and structure $x$.

*H3 and canonical non-conserved side chain modelling.*
This step is required since certain side chain conformations in canonical non-H3 CDRs and all of the H3 side chains are non-conserved. H3 and canonical non-conserved side chains (as well as all side chains on any non-canonical non-H3 loops) are built simultaneously on the final selected backbone (for the r.m.s.d. and VFF backbone only screens) or the entire clustered set (for the accessibility and VFF side chain screens), using either:
1. Our own implementation of the dead-end elimination algorithm (Lasters *et al.*, 1995). This attempts to find a global minimum by successively eliminating rotamers from a rotamer library (Ponder and Richards, 1987) specially developed by the Desmet group (De Maeyer *et al.*, 1997), which have a higher lowest energy of interaction with their environment than the highest energy of interaction of the other rotamers of that residue. The residues of the H3 kink have been found to be conserved in chi-1 (N.R.J.Whitelegg and A.R.Rees, unpublished results) and are therefore pre-built using the standard chi-1 values for those residues. For the chi-2 of the kink residues, depending on the residue type, either the extended conformation or the mean of frequently observed (64–88%) chi-2 values is used.
2. The CONGEN side chain construction algorithm (Bruccoleri and Karplus, 1987), which iteratively builds all the side chains in their lowest possible energy conformation with respect to the molecular environment until energy convergence is reached.

The preferred method depends on the screen used. For the r.m.s.d. and VFF backbone screens (where backbone energy only is being used as the screen and therefore only one conformation needs to have side chains built), the dead-end elimination algorithm is used. For VFF with side chains or the accessibility screen, side chains need to be built on the entire set of clustered conformations. Therefore, CONGEN is used in preference to the dead-end elimination algorithm since, although slightly less accurate and less theoretically sound, it is better able to deal with the processing of multiple conformations in a timely fashion.

Table I shows the results from modelling 19 test cases of known crystal structure using the three possible methods available in WAM to screen the final conformations:
1. energy screen on the final, post-clustering set of conformations, including side chains;
2. accessibility profile screen;
3. the r.m.s.d. screen.
It can be seen that the energy-based VFF method is not particularly discriminating, selecting a conformation of 2.0 Å or less in only nine out of 19 examples. This was the screening method used in AbM and although only four of the 19 antibodies modelled here were also modelled using the AbM package, the results show essentially the same trend. By


165

**Table I.** R.m.s.d.s of final CDR-H3 models using different screening methods

| Structure | Length | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|---|
| 1mlb | 7 | 1.1 | 1.1 | 0.4 | 0.7 | 1.6 | – | 0.8, 0.6, 1.1, 0.8, 1.0, 0.5 |
| 1fgn | 8 | 1.7 | 1.6 | 1.6 | 1.3 | 0.8 | – | 0.9, 0.4, 1.2, 1.1, 0.9, 1.6 |
| 1vfa | 8 | 1.8 | 1.6 | 0.9 | 1.1 | 1.4 | – | 0.5, 0.9, 0.9, 1.0, 0.7, 1.3 |
| 2fbj | 9 | 2.4 | 1.9 | 1.6 | 0.9 | 1.9 | 1.7 | 1.7, 1.7, 2.3, 2.0, 1.2, 1.7 |
| 1igt | 9 | 3.8 | 1.4 | 0.9 | 1.3 | 1.1 | – | 1.5, 1.8, 0.9, 0.9, 1.0, 1.4 |
| 1for | 10 | 2.6 | 2.4 | 1.3 | 1.4 | 2.0 | 1.5 | 1.4, 1.0, 1.5, 0.9, 1.0, 1.4 |
| 2h1p | 11 | 1.6 | 2.1 | 1.8 | 1.6 | 3.0 | – | 1.3, 1.2, 1.7, 1.2, 1.6, 2.2 |
| 1mcp | 11 | 1.8 | 2.6 | 1.9 | 1.4 | 3.0 | – | 1.9, 1.2, 0.8, 0.8, 1.4, 2.6 |
| 1fpt | 11 | 2.6 | 2.1 | 2.3 | 1.5 | 3.1 | – | 0.8, 1.2, 1.3, 0.9, 0.8, 2.8 |
| 1igm | 12 | 1.3 | 3.8 | 1.2 | 1.3 | 2.2 | 2.6 | 1.3, 0.9, 1.3, 0.9, 0.9, 1.9 |
| *2+ non-canonical* | | | | | | | | |
| 1tet | 7 | 1.6 | 1.6 | 1.7 | 1.6 | 1.8 | – | 1.9, 1.3, 0.8, 1.2, 1.7, 1.9 |
| 1cgs | 7 | 1.6 | 1.6 | 1.5 | 1.2 | 1.6 | 2.8 | 1.3, 0.8, 1.6, 0.9, 1.0, 1.8 |
| 1a6t | 8 | 0.7 | 1.1 | 0.7 | 0.7 | 1.5 | – | 1.0, 0.7, 1.0, 1.3, 1.9, 0.9 |
| 1kem | 8 | 2.4 | 1.3 | 2.7 | 1.3 | 2.7 | – | 1.3, 0.9, 1.8, 1.2, 1.8, 2.4 |
| 1jhl | 9 | 2.4 | 2.0 | 0.9 | 1.4 | 1.1 | – | 1.7, 0.8, 0.7, 1.4, 1.6, 1.2 |
| 2jel | 9 | 3.1 | 3.0 | 1.7 | 1.8 | 2.9 | – | 1.4, 1.0, 0.7, 0.8, 1.2, 2.7 |
| 1clz | 10 | 3.1 | 3.1 | 1.8 | 1.5 | 1.7 | – | 1.2, 1.3, 1.5, 0.9, 0.9, 2.8 |
| 1dba | 10 | 3.2 | 3.2 | 1.6 | 1.5 | 2.4 | – | 1.7, 1.2, 1.2, 1.2, 1.8, 1.8 |
| 1kb5 | 10 | 3.2 | 2.6 | 2.7 | 1.6 | 2.0 | – | 1.1, 1.2, 1.6, 1.6, 2.6, 3.4 |

Key: r.m.s.d. values of the final models: A, by the VFF energy screen; B, by the accessibility screen; C, by the r.m.s.d. screen; D is the lowest r.m.s.d. model of the entire set of constructed conformations; E is the r.m.s.d. between the crystal structure under investigation and the closest other known structure by the Dayhoff mutation matrix; F is the r.m.s.d. of the final model using the original AbM procedure where this was carried out (Whitelegg, 1998); G is the backbone plus buried side chain r.m.s.d.s using the dead-end elimination method and the r.m.s.d. screen, for *all* loops (values for non-canonical loops are underlined). All other r.m.s.d. values are backbone only. Note: r.m.s.d. is measured on N, Cα, C and Cβ atoms and is performed without fitting the CDR-H3 to the crystal loop.

contrast, while overall the accessibility screen does little better than VFF, with a conformation of ≤2.0 Å in 10 out of 19 examples, it is notably effective at short (nine or less) H3 lengths, selecting a conformation of ≤2.0 Å in 10 out of 11 examples. This is because for shorter loops, a greater percentage of the loop consists of residues with conserved accessibility values, whereas for longer loops, there are generally more apex residues with variable accessibility between individual examples.

The most successful screen is the r.m.s.d. screen. This method performs well, selecting a conformation of ≤2.0 Å r.m.s.d. (measured on N, Cα, C, Cβ atoms) in 16 out of 19 examples. On the whole, the r.m.s.d. screen performs better than merely selecting the closest crystal structure loop of that length by sequence similarity using the Dayhoff mutation matrix. Those structures which perform poorly using the r.m.s.d. screen (e.g. 1kem and 1kb5) are examples of structures which deviate considerably from the average conformation for that length, for sequence-dependent steric reasons (see Website).

Buried side chains are also modelled well using a combination of the r.m.s.d. screen and the dead-end elimination method for side chain construction, with a backbone plus buried side chain r.m.s.d. of ≤2.0 Å for non-H3 loops in almost all cases (one exception, CDR-H2 in 1kb5, is a non-canonical loop with backbone r.m.s.d. of 2.2 Å; in the other, CDR-L3 in 2fbj, the value is entirely due to a poorly placed tryptophan which is only semi-buried and is close to a number of exposed side chains with high r.m.s.d.). The r.m.s.d. values for CDR-H3 mirror those for backbone only, in that well-modelled backbones lead to well-modelled side chains whereas mild inaccuracies in the backbone are exacerbated in side chain construction (e.g. r.m.s.d. values of 1.7 and 2.7 Å, respectively, for the CDR-H3 of 2jel). In all these instances the poorly modelled side chains correspond to regions of the loop where the backbone deviation is the greatest.

*Conclusion*

WAM offers an improved algorithm for CDR-H3 modelling. The results presented here show that those knowledge-based approaches applied have been more effective than the original energy-based approaches. For shorter H3 loops of up to nine residues, accurate models (≤1.7 Å r.m.s.d. for nine out of 11 examples) can now be constructed, while for loops of 10–12 residues a slight increase in the range of r.m.s.d. values is seen (1.3–2.7 Å r.m.s.d.). While the accessibility and the r.m.s.d. methods provide equivalent accuracies for kinked loops of up to nine residues, the accessibility score method is preferred since it is better able to screen those antibodies containing outlying H3 conformations. By contrast, the r.m.s.d. score is the method of choice for 10–12 residue H3s where fewer residues have a well-defined accessibility profile.

There are some remaining problems to be solved: first, for loops longer than 12 residues there are too few database examples for either knowledge-based method to show significant improvement over an energetic screen such as VFF. Furthermore, the apices of the longer H3 loops vary between the known structures (perhaps required for conformational change during binding), reducing the inherent advantages of a consensus, knowledge-based approach. For these cases, methods that explicitly take account of flexibility (such as molecular dynamics simulation) will need to be developed. Second, those kinked H3 loops of atypical conformation (such as 1kem and 1kb5 above) cannot be constructed accurately by the r.m.s.d. screen. Until such time as examples of similar conformation appear in the PDB, allowing sequence–structure relationships to be devised (this also applies to non-kinked

H3s for which there are only a small number known at present), it will only be possible to have moderate confidence in the models produced. Having said that, our own work (see the Website) and that of Shirai's group (Shirai *et al.*, 1999) have allowed the definition of some loose rules for the kinked outliers.

The other remaining problem relates to the small number of non-canonical CDRs within the L1–L3 and H1, H2 sets. Studies by Chothia and co-workers (AlLazikani *et al.*, 1997) and our own examination of the newer structures deposited in the Protein Data Bank suggest that most of the exceptions will actually probably turn out to be members of new canonical classes.

## Acknowledgements

WAM contains modules developed by the former members of this research group Andrew Martin, Jan Pedersen, Andrew Henry and Steve Searle, whose contributions are acknowledged. We also thank Oxford Molecular for their cooperation in this venture. We would thank the CRC for funding to allow this work to take place as part of a larger project.

## References

AlLazikani,B., Lesk,A.M. and Chothia,C. (1997). *J. Mol. Biol.*, **273**, 927–948.
Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000). *Nucleic Acids Res.*, **28**, 235–242.
Bruccoleri,R.E. and Karplus,M. (1987) *Biopolymers*, **26**, 137–168.
Chothia,C. and Lesk,A.M. (1987) *J. Mol. Biol.*, **196**, 901–917.
Chothia,C. *et al.* (1989) *Nature*, **342**, 877–883.
Chothia,C., Lesk,A.M., Gherardi,E., Tomlinson,I.M., Walter,G., Marks,J.D., Llewellyn,M.B. and Winter,G. (1992) *J. Mol. Biol.*, **227**, 799–817.
Darsley,M.J. and Rees,A.R. (1985) *EMBO J.* **4**, 383–392.
Dauber-Osguthorpe,P., Roberts,V.A., Osguthorpe,D.J., Wolff,J., Genest,M. and Hagler A.T. (1988) *Proteins*, **4**, 31–47.
de la Paz,P., Sutton,B.J., Darsley,M.J. and Rees,A.R. (1986) *EMBO J.*, **5**, 415–425.
De Maeyer,M., Desmet,J. and Lasters,I. (1997) *Fold. Des.*, **2**, 53–66.
Feldmann,R.J., Potter,M. and Glaudemans,C.P.J. (1981) *Mol. Immunol.*, **18**, 683–698.
Foote,J. and Winter,G. (1992) *J. Mol. Biol.*, **224**, 487–499.
Go,N. and Scheraga,H.A. (1970) *Macromolecules*, **3**, 178–187.
Kabat E.A and Wu,T.T. (1972) *Proc. Natl Acad. Sci. USA* **69**, 960–964.
Kabat,E.A., Wu,T.T. and Bilofsky,H. (1977) *J. Biol. Chem.*, **252**, 6609–6616.
Lasters,I., De Maeyer,M. and Desmet,J. (1995) *Protein Eng.*, **8**, 815–822.
Mandal, C., Kingery,B.D., Anchin,J.M., Subramaniam,S. and Linthicum,D.S. (1996) *Nature Biotechnol.*, **14**, 323.
Martin,A.C.R., Cheetham,J.C. and Rees,A.R. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 9268–9272.
Martin,A.C.R., Cheetham,J.C. and Rees,A.R. (1991) *Methods Enzymol.*, **203**, 121–153.
Morea,V., Tramontano,A., Rustici,M., Chothia,C. and Lesk,A.M. (1998) *J. Mol. Biol.*, **275**, 269–294.
Padlan,E.A. and Davies,D.R. (1975) *Proc. Natl Acad. Sci. USA*, **72**, 819–823.
Padlan,E.A., Davies,D.R., Pecht,I., Givol,D. and Wright,C. (1976) *Cold Spring Harbor Symp. Quant. Biol.*, **41**, 627–637.
Pedersen,J.T., Searle,S.M.J., Henry,A.H. and Rees,A.R. (1992) *Immunomethods*, **1**, 126.
Pedersen,J.T., Henry,A.H., Searle,S.M.J., Guild,B.C., Roguska,M. and Rees,A.R. (1994) *J. Mol. Biol.*, **235**, 939–973.
Ponder,J.W. and Richards,F.M. (1987) *J. Mol. Biol.*, **193**, 775–791.
Rees,A.R., Martin,A.C.R., Webster,D., Cheetham,J.C. and Roberts,S. (1990) *Biophys. J.* **57**, A384.
Rees,A.R., Searle,S.M.J., Henry,A.H.and Pedersen,J.T. (1996) In Sternberg M.J.E. (ed.). *Protein Structure Prediction*. Oxford University Press, Oxford, 141–172.
Reichmann,L., Clark,M., Waldmann,H. and Winter,G. (1988) *Nature*, **332**, 323–327.
Roguska,M.A. *et al.* (1994) *Proc. Natl Acad. Sci. USA*, **91**, 969–973.
Roguska,M.A. *et al.* (1996) *Protein Eng.*, **9**, 895–904.
Shirai,H., Kidera,A. and Nakamura,H. (1996) *FEBS Lett.*, **399**, 1–8.
Shirai,H., Kidera,A. and Nakamura,H. (1999) *FEBS Lett.*, **455**, 188–197.
Stanford,J.M. and Wu,T.T (1981) *J. Theor. Biol.*, **88**, 421–439.
Tomlinson,I.M., Cox,J.P.L., Gherardi,E., Lesk,A.M. and Chothia,C. (1995) *EMBO J.*, **14**, 4628–4638.
Whitelegg,N.R.J. (1998) PhD Thesis University of Bath.
Wu,T.T. and Kabat,E.A. (1970) *J. Exp. Med.* 132, 211–250.

*Received July 17, 2000, revised October 19, 2000; accepted October 23, 2000*

## Appendix. Worked example

A worked example of a typical WAM modelling run is presented below. The following sequence (CDRs in bold) is aligned (see web page) and entered into the interactive online WAM interface.

| | |
|---|---|
| Light chain | DVVMTQIPLSLPVNLGDQASISC**RASQSISNNLN**<br>WYLQKPGQSPKLLMY**YATSLAD**GVPDRFSGSGSGTD<br>FTLKISRVEAEDLGIYFC**QHFWS TPKT**FGGGTKLEIKRA |
| Heavy chain | RVQLLESGAELMKPGASVQISCKAT**GFEFAKYWMS**<br>WVKERPGHGLEWIG**YIANAGGSTY**YREKFKGKATFT<br>RDTSSNTAYMQLSSLTSEDSAVYYCT**RSGNYAMDY**<br>WGQG TSVTVS |

### *Framework and canonicals*

The first stage is to determine:

1. the most sequence-homologous framework for each of the light and heavy chains;
2. the canonical class of each loop;
3. the most sequence-homologous database loop of that class.

For this example, the most homologous light chain framework is 1dba, whereas for the heavy chain it is 1cgs.
The canonicals are as shown.

| CDR | Sequence[a] | Structure used[b] | Class[c] (or H3 type) |
|---|---|---|---|
| L1 | RASQSISN·····NLN | 1mlb (RASQSISN·····NLH) | 2 |
| L2 | YATSLAD | 1fgn (YATSLAD) | 1 |
| L3 | QHFWS··TPKT | 1vfa (QHFWS··TPRT) | 1 |
| H1 | GFEFAK··YWMS | 2fbj (GFDFSK··YWMS) | 1 |
| H2 | YIANAGG··STY | 1igt (YISNGGG··STY) | 3 |
| H3 | (R)SGNYA·····MDY | | Kinked |

[a]For H3, the residue immediately preceding the CDR (R), is also shown since it determines the presence of the kink.
[b]Most sequence-homologous member of the appropriate class. Side chain positions with conserved chi-1 values between individual members of the class are highlighted.
[c]According to Chothia and Lesk (1987) and subsequent publications by Chothia and co-workers (see Methods).

Thus, the light chain framework is built using that of 1dba, resequencing residues using maximum overlap if necessary; and the heavy chain framework is built using that of 1cgs. The two chains are combined into a single model by fitting on an average β-barrel generated from the known Fv structures. The frameworks of loops L1 to H2 are deleted and rebuilt using the canonicals from 1mlb, 1fgn, 1vfa, 2fbj and 1igt, respectively. Finally, the conformationally conserved canonical side chains (N.R.J.Whitelegg and A.R.Rees, unpublished results) are built, using maximum-overlap resequencing.



167

*The H3 loop: construction*

The eight-residue H3 loop in this instance is preceded by Arg and features Asp at the residue preceding the C-terminal, specifying a kinked conformation (Shirai *et al.*, 1996). Therefore, CAMAL (the combined database/conformational search) is used, in this instance with C-alpha to C-alpha distance constraints derived from kinked H3 structures only. For this example, the database search generated 151 hits, while after rebuilding the apex using CONGEN, a total of 14 108 conformations were produced. On clustering this was reduced to 69.

*The H3 loop: final screen*

This is an eight-residue kinked CDR and therefore the r.m.s.d. screen is the method of choice, as it is a relatively short loop and the known structures have a well-defined profile. Since side chains are required to be in place for this screen, side chains need to be built on *all* the clustered conformations. As CONGEN is a quicker method than dead-end elimination, this is the default algorithm here, although the user may alter this if it is decided that a slight time penalty can be taken in order to achieve a more theoretically sound model. (Note that if the r.m.s.d. screen was used, the screening requires only the backbone conformation and therefore only the final backbone, selected by the screen, would require side chain construction. Therefore dead-end elimination would be the default method.)

All non-conserved side chains (H3 and canonical CDRs) are built simultaneously using CONGEN, which iteratively places side chains at each residue until an acceptable global minimum is reached according to energy convergence. After side chain construction each conformation is energy-minimized (by steepest descent, using VFF) to relieve steric clashes and the accessibility screen performed on each model to select the final conformation, which is that most closely matching the typical accessibility profile of eight-residue kinked H3s.

The final model is then made available as a PDB file for visualization by any standard graphics software package.

| Monthly Salary Rate | Sept 1-15<br>$7,173 | Sept 16 - 30<br>$7,173 | October<br>$14,346 | November<br>$14,346 | December<br>$14,346 | January<br>$14,346 | February<br>$14,346 | March<br>$14,346 | April<br>$14,346 | May<br>$14,346 | June 1 - 15<br>$7,173 | June 16-30<br>$7,173 | July<br>$14,346 | August<br>$14,346 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 620-4731000-60023911 | 0.0000% | 0.0000% | 28.9996% | 28.9996% | 28.9996% | 28.9996% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | | | | |
| 620-4731000-60023399 | 0.0000% | 0.0000% | 6.2473% | 6.2473% | 6.2473% | 6.2473% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | | | | |
| 650-4731000-60029641-02 | 0.0000% | 0.0000% | 4.5137% | 4.5137% | 4.5137% | 4.5137% | 4.5137% | 4.5137% | 4.5137% | 4.5137% | 0.0000% | | | | |
| 650-4731000-60029641-04 | 0.0000% | 0.0000% | 4.5137% | 4.5137% | 4.5137% | 4.5137% | 4.5137% | 4.5137% | 4.5137% | 4.5137% | 0.0000% | | | | |
| 110-4731001 OVER THE CAP | 0.0000% | 0.0000% | 7.1106% | 7.1106% | 7.1106% | 7.1106% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | | | | |
| 110-4731001 | 100.0000% | 39.8600% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 41.6850% | 41.6850% | 41.6850% | 41.6850% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | |
| 450-4700000-40005020 | 0.0000% | 60.1400% | 48.6150% | 48.6150% | 48.6150% | 48.6150% | 49.2875% | 49.2875% | 49.2875% | 49.2875% | 100.0000% | 100.0000% | 100.0000% | 100.0000% | |
| Total | 100.0000% | 100.0000% | 100.0000% | 100.0000% | 100.0000% | 100.0000% | 100.0000% | 100.0000% | 100.0000% | 100.0000% | 100.0000% | 100.0000% | 100.0000% | 100.0000% | |
| Balance | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | |
| 620-4731000-60023911 | $ - | $ - | $ 4,160.32 | $ 4,160.32 | $ 4,160.32 | $ 4,160.32 | $ - | $ - | $ - | $ - | $ - | $ - | $ - | $ - | $ 16,641.26 |
| 620-4731000-60023399 | $ - | $ - | $ 896.25 | $ 896.25 | $ 896.25 | $ 896.25 | $ - | $ - | $ - | $ - | $ - | $ - | $ - | $ - | $ 3,585.00 |
| 650-4731000-60029641-02 | $ - | $ - | $ 647.55 | $ 647.55 | $ 647.55 | $ 647.55 | $ 647.55 | $ 647.55 | $ 647.55 | $ 647.55 | $ - | $ - | $ - | $ - | $ 5,180.36 |
| 650-4731000-60029641-04 | $ - | $ - | $ 647.55 | $ 647.55 | $ 647.55 | $ 647.55 | $ 647.55 | $ 647.55 | $ 647.55 | $ 647.55 | $ - | $ - | $ - | $ - | $ 5,180.36 |
| 110-4731001 OVER THE CAP | $ - | $ - | $ 1,020.09 | $ 1,020.09 | $ 1,020.09 | $ 1,020.09 | $ - | $ - | $ - | $ - | $ - | $ - | $ - | $ - | $ 4,080.38 |
| 110-4731001 | $ 7,173.00 | $ 2,859.18 | $ - | $ - | $ - | $ - | $ 5,980.18 | $ 5,980.18 | $ 5,980.18 | $ 5,980.18 | $ - | $ - | $ - | $ - | $ 33,952.88 |
| 450-4700000-40005020 | $ - | $ 4,313.88 | $ 6,974.36 | $ 6,974.36 | $ 6,974.36 | $ 6,974.36 | $ 7,070.84 | $ 7,070.84 | $ 7,070.84 | $ 7,070.84 | $ 7,173.06 | $ 7,173.06 | $ 14,346.11 | $ 14,346.11 | $ 103,533.01 |
| | $ 7,173.00 | $ 7,173.06 | $ 14,346.11 | $ 14,346.11 | $ 14,346.11 | $ 14,346.11 | $ 14,346.11 | $ 14,346.11 | $ 14,346.11 | $ 14,346.11 | $ 7,173.06 | $ 7,173.06 | $ 14,346.11 | $ 14,346.11 | $ 172,153.25 |

APPLICATIONS NOTE Vol. 24 no. 17 2008, pages 1953–1954
doi:10.1093/bioinformatics/btn341

*Structural bioinformatics*

# PIGS: automatic prediction of antibody structures

Paolo Marcatili[1], Alessandra Rosi[1] and Anna Tramontano[1,2,*]

[1]Department of Biochemical Sciences and [2]Istituto Pasteur Fondazione Cenci Bolognetti, Sapienza University, P.le A. Moro 5, 00185 Rome, Italy

Received on April 19, 2008; revised on June 17, 2008; accepted on July 2, 2008

Advance Access publication July 19, 2008

Associate Editor: Alfonso Valencia

**ABSTRACT**

**Summary:** We describe a web server for the automatic prediction of immunoglobulin variable domains based on the canonical structure model. The server is user-friendly and flexible. It allows the user to select the templates for the frameworks and the loops using different strategies. The final output is a full-fledged 3D model of the variable domains of the target immunoglobulin.

**Availability:** The server is openly accessible to academic users at the address: http://arianna.bio.uniroma1.it/pigs. It does not require registration and there is no limit to the number of sequences that can be submitted.

**Contact:** anna.tramontano@uniroma1.it

## 1 INTRODUCTION

Immunoglobulins are key players of the immune response and their overall structure is reasonably well conserved. They are composed of two heavy and two light chains that contain four and two domains with a similar fold, respectively (Chothia *et al.*, 1989). Antibodies bind their cognate antigen using the tip of the first domains of each chain (VL and VH). From a structural point of view, the antigen binding site is formed by six loops, three from the light (L1, L2, L3) and three from the heavy chain (H1, H2, H3) named according to their order of appearance in the amino acid sequence.

The structure of the main chain of five of these loops can be predicted quite accurately by taking into account the position and identity of a few specific key amino acids (Chothia and Lesk, 1987; Chothia *et al.*, 1989; Tramontano *et al.*, 1990; Webster and Rees, 1995). For example, either five or six residues form the L3 loop. The five residue L3 loops all have similar main chain conformation (within 0.2 Å). The six residue L3 loops can only take one of two possible conformations depending upon the position of a proline residue within the loop. The case of the third loop of the heavy chain (H3) is more complex. Extensive analysis of this loop in the many available structures has demonstrated that the conformation of the 10 residues closer to the framework (the *torso*) can be predicted in a similar fashion as for the other loops, while the remaining do not seem to follow identifiable sequence rules (Morea *et al.*, 1998; Shirai *et al.*, 1996).

The advanced understanding of the sequence to structure relationship in this important class of molecules makes it possible to predict their structure quite accurately and automatically. The crucial steps in the prediction are the correct alignment of the target sequence with those of immunoglobulins of known structure, the identification of the limits of the hypervariable loops (where insertions and deletions occur) and of the key residues determining their conformation. The alignment has to follow immunoglobulin specific rules and cannot by obtained by classical dynamic programming methods, because insertions and deletions can only occur at very specific positions and some important conserved residues, for example, two bonded cysteine and a tryptophan, need to be aligned. Nevertheless, rule-based techniques for the alignment, the identification of the canonical structures and the detection of the appropriate templates for the loops can be implemented and automated.

A server, named WAM (Webster and Rees, 1995), mostly based on the rules described earlier, is already available for immunoglobulin structure prediction. However, its usage by the academic community at large is limited by a number of factors. Users, who need to register via fax, are restricted to five sequences per month, which is a rather low threshold in the genomic era. Furthermore, the server is rather rigid: it does not allow any input for the selection of the template structures and it only works if the sequence spans the precise boundaries of the domain. Finally the alignment often requires manual intervention.

## 2 THE PIGS SERVER

We have developed PIGS (prediction of immunoglobulin structure), a tool to build the structure of immunoglobulins available to the academic community. The PIGS server is flexible and user-friendly and relies upon a database of known immunoglobulin structures and of their structural alignment that is regularly updated. The user only needs to input the sequence of the variable chains of the antibody of interest and the program will display a list of putative templates for both the loops and the framework for each chain, together with other useful information (Fig. 1).

The user can either manually select the templates, or automatically select one of four possible strategies:

*Same antibody:* select the known structure that can provide a template for both the heavy and light chain, even if a different template with a higher sequence identity exists for one of the chains. If two different antibodies are selected for each chain, the program needs to reconstruct the complete molecule by matching residues known to be conserved at the VL–VH interface and this can introduce more errors than taking the two chains from the same antibody.

*Same canonical structure:* select the template having loops with the same canonical structure of the target even if a different template

*To whom correspondence should be addressed.

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

Downloaded from http://bioinformatics.oxfordjournals.org/ at Northeastern University Libraries on January 11, 2012

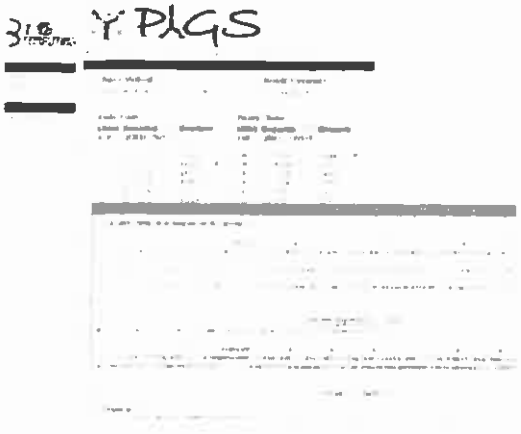


Fig. 1. The main page of PIGS, displayed after the user has uploaded the target sequences. The template structures for each chain can be selected manually or according to a predefined strategy. The numbers in the 'Loop' column indicate the canonical structure of the loop, with the blue color indicating the same canonical structure as the target immunoglobulin chain. The alignment with each template structure can be viewed and edited in the pop-up window by clicking on the %id figure.

with a higher sequence identity exists for one or both chains. If a chain with different canonical structures is selected, the program needs to reconstruct the complete chain by building a 'chimera' taking the loops from an antibody with the matching canonical structures and this can introduce more errors than taking the loops from the same chain used to model the framework.

*Same antibody and canonical structure:* select an antibody structure that can be used as a template for both Vl and Vh and where the canonical structures of the loops are the same as those of the target even if a different template with a slightly better sequence identity exists for one or both chains.

*Best L and H chain:* select the two chains with highest sequence identity with the corresponding chains of the target and, if needed, pack the two chains together and take the loops from a different structure.

The user can also select whether he/she prefers to obtain just the main chain coordinates ('Backbone only'), the coordinates of the main chain and of all the side chains conserved with respect to the template ('Transfer conserved residues') or the complete structure. In the latter case ('Transfer conserved + SCWRL'), the conformation of the conserved side chains is retained and the remaining ones are reconstructed using SCWRL (Bower *et al.*, 1997).

No extensive benchmarking has ever been performed to assess which of the strategies described earlier works best. In our own experience, the 'Same antibody' and 'Transfer conserved + SCWRL' are preferable, and these are the default parameters of the server. We are presently performing a large scale benchmarking whose results will be made available via the PIGS server.

Once the choice has been made, the user can select the 'build' option and obtain the predicted structure of the antibody in PDB format. The model can be visualized via JMol, where a few options allow the user to identify key parts of the molecule.

The reliability of the canonical structure method has been proven over and over again, so there is no need to further demonstrate it. We just show as an example, in Table 1 the r.m.s.d. deviation of the main chain atoms between the models obtained by the completely automatic procedure on three antibody structure recently deposited in the protein structure database and their experimental structures.

Table 1. Comparison between the predicted and experimental main chain structures of four recently solved immunoglobulin structures

| r.m.s.d. | Region | 2HWZ | 2ADF | 2R29 | 2ZCH |
|---|---|---|---|---|---|
| All | All | 1.11 | 1.08 | 1.42 | 1.16 |
| FW | FW | 0.94 | 0.88 | 1.22 | 0.95 |
| Loops | FW | 1.75 | 1.75 | 2.12 | 1.81 |
| L | L | 1.17 | 0.48 | 1.14 | 0.6 |
| FW L | FW L | 1.03 | 0.48 | 0.94 | 0.47 |
| L loops | FW L | 1.90 | 0.48 | 1.90 | 1.07 |
| L1 (length) | FW L | 1.63 (6) | 0.31 (7) | 2.29 (11) | 1.24 (11) |
| L2 (length) | FW L | 0.91 (3) | 0.32 (3) | 1.05 (3) | 0.44 (3) |
| L3 (length) | FW L | 2.43 (6) | 0.68 (5) | 1.39 (6) | 0.94 (6) |
| H | H | 0.74 | 1.01 | 1.50 | 1.21 |
| FW H | FW H | 0.52 | 0.61 | 1.44 | 0.73 |
| H loops | FW H | 1.38 | 1.97 | 1.90 | 2.44 |
| H1 (length) | FW H | 0.8 (9) | 0.85 (7) | 0.81 (7) | 0.56 (7) |
| H2 (length) | FW H | 1.17 (3) | 0.88 (4) | 1.14 (4) | 0.49 (4) |
| H3 (length) | FW H | 1.66 (10) | 2.52 (9) | 2.47 (7) | 3.06 (13) |

No refinement protocol has been applied. Values are the r.s.m.d. in A of the region in the first column after optimal superposition of the region indicated in the second. FW, Framework; L, Light chain; H, Heavy chain.

## 3 CONCLUSIONS

The prediction of antibody structures is not only important, but feasible at a level of accuracy much higher than for any other protein type. The level of understanding of the sequence structure relationship in this class of molecules is sufficiently advanced so that automatic easy to use methods can be developed and employed for the molecular analysis of the 'immunome'. The purpose of the PIGS server described here is to enable scientists to tackle the open problem of the molecular basis of the specificity of antibodies at large and/or to obtain data useful in the context of specific biological problems.

## ACKNOWLEDGEMENTS

*Funding*: PIGS is supported by the EU contract number LHSG-CT-2003-503265, and AIRC BICG Project.

*Conflict of Interest*: none declared.

## REFERENCES

Bower,M. *et al.* (1997) Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.*, **267**, 1268–1282.

Chothia,C. and Lesk,A. (1987) Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.*, **196**, 901–917.

Chothia,C. *et al.* (1989) Conformations of immunoglobulin hypervariable regions. *Nature*, **342**, 877–883.

Morea,V. *et al.* (1998) Conformations of the third hypervariable region in the VH domain of immunoglobulins. *J. Mol. Biol.*, **275**, 269–294.

Shirai,H. *et al.* (1996) Structural classification of CDR-H3 in antibodies. *FEBS Lett.*, **399**, 1–8.

Tramontano,A. *et al.* (1990) Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. *J. Mol. Biol.*, **215**, 175–182.

Webster,D.M. and Rees,A.R. (1995) Molecular modeling of antibody-combining sites. *Methods Mol. Biol.*, **51**, 17–49.



Downloaded from http://bioinformatics.oxfordjournals.org/ at Northeastern University Libraries on January 11, 2012

171

**The first monoclonal therapeutic antibody**

Following the development of the hybridoma technology pioneered by Kohler and Milstein, various mouse monoclonal antibodies against human proteins were developed. One of these, OKT3 (Kung P, Goldstein G, Reinherz EL, W, Schlossman SF (1979) Monoclonal antibodies defining distinctive human T cell surface antigens. Science, 206, 347-349), was approved by the Food and Drug Administration (FDA) for therapeutic use in 1986, becoming the first FDA approved monoclonal therapeutic antibody. It is a mouse IgG2a antibody specific for human CD3 epsilon chain.

CD3 epsilon chain is one of the components of the T cell receptor complexes on both helper and cytotoxic T lymphocytes. Initially, OKT3 stimulates the activity of T cells, but eventually leads to blockage and apoptosis of T cells. Therefore, it is an immunosuppressant used in acute and glucocorticoid resistant rejections of allogeneic renal, heart and lung transplantations.

As an immunosuppressant, OKT3 minimizes the production of anti-mouse protein antibodies in human patients. Thus it can be infused daily for about one week without adverse effects. However, due to its initial stimulatory effect of T cells, glucocorticoid, acetaminophen, and diphenhydramine are given before the infusion to minimize the cytokine release syndrome, Eventually, it will lead to anaphylactic reactions against mouse proteins and should be discontinued.

Subsequently, most of the monoclonal therapeutic antibodies are modified mouse monoclonal antibodies, containing less amounts of mouse protein to reduce allergic reaction to foreign proteins in human patients. These require some additional genetic engineering manipulations. As discussed in following sections, the basic findings of the existence of variable and constant regions of antibody light and heavy chains, the location of the complementarity determining regions in the variable regions, the similar structures of various frame region three dimensional structures of different antibodies, etc. will play vital roles in the design of monoclonal therapeutic antibodies.

Finally, new molecular biology techniques have been developed to generate monoclonal human therapeutic antibodies without any component of mouse proteins.

suggests that the observed regenerated synapses are functional. This suggestion is further supported by the observation that a second lesion of the spinal cord just proximal to the first, in animals that had recovered the ability to swim, resulted in an immediate paralysis persisting for at least 24 days. The return of relatively normal locomotor function in the lamprey despite grossly aberrant patterns of regenerated synaptic connections supports similar findings in the goldfish (*13*). Despite the different spatial location of the regenerated junctions in the cord, the presynaptic elements may be still connecting to different regions of their normal postsynaptic target neurons.

The increase in branching may be related to the relatively short distance that regenerated neurites grow distal to the lesion. It has been hypothesized that neurons are genetically programmed to produce a relatively fixed surface area of plasma membrane (*14*). If this area is produced by profuse branching, the critical point at which growth stops may be reached closer to the lesion than if elongating growth were confined to a single neurite.

The formation of scar tissue (for example, ependymal or glial or both) has been implicated in producing a local physical barrier to the growth of nerve fibers across a lesion in the central nervous system (*15*). Scar tissue in the lamprey spinal cord appears to be composed principally of proliferating loosely aggregated ependymal cells. We found no collagen within the wound, although substantial quantities surround the normal cord (*16*). The resulting extracellular spaces between ependymal cells may provide channels for regenerating neurites to penetrate the lesion. A similar suggestion has been made for successful spinal regeneration in the newt (*17*).

We have demonstrated that axons from identified regenerating neurons grow through a spinal lesion and form new synapses caudal to the injury. This process is accompanied by functional recovery in the lamprey. The ability to follow growth and synaptic regeneration in identified reticulospinal neurons of the lamprey may permit this preparation to serve as a model for examining the cellular factors concerned with functional recovery of the injured vertebrate CNS.

MALCOLM R. WOOD
MELVIN J. COHEN
*Department of Biology,*
*Yale University,*
*New Haven, Connecticut 06520*

### References and Notes

1. E. Puchala and W. F. Windle, *Exp. Neurol.* **55**, 1 (1977).
2. D. Hooker, *J. Exp. Zool.* **55**, 23 (1930); J. Piatt, *ibid.* **129**, 177 (1955); W. Kirsche, *Z. Mikrosk. Anat. Forsch.* **62**, 521 (1956).
3. D. Hooker, *J. Comp. Neurol.* **56**, 277 (1932); J. J. Bernstein and M. E. Bernstein, *Exp. Neurol.* **24**, 538 (1969).
4. K. Maron, *Folia Biol. (Krakow)* **7**, 179 (1959).
5. C. M. Rovainen, *J. Comp. Neurol.* **168**, 545 (1976).
6. M. E. Selzer, *J. Physiol. (London)* **277**, 395 (1978).
7. C. Rovainen, in *Neurobiology of the Mauthner Cell*, D. Faber and H. Korn, Eds. (Raven, New York, 1978), p. 245.
8. W. O. Wickelgren, *J. Physiol. (London)* **270**, 89 (1977).
9. G. L. Ringham, *ibid.* **251**, 395 (1975).
10. B. N. Christensen, *J. Neurophysiol.* **39**, 197 (1976); D. S. Smith, U. Jarlfors, R. Beranek, *J. Cell Biol.* **46**, 199 (1970).
11. K. H. Pfenninger and C. M. Rovainen, *Brain Res.* **72**, 1 (1974).
12. C. M. Rovainen, *J. Comp. Neurol.* **154**, 189 (1974); *ibid.*, p. 207.
13. J. J. Bernstein and J. B. Gelderd, *Exp. Neurol.* **41**, 402 (1973).
14. M. Devor and G. E. Schneider, *Colloq. Inst. Natl. Santé Rech. Med.* **43**, 191 (1975).
15. C. D. Clemente, *Int. Rev. Neurobiol.* **6**, 257 (1964).
16. B. Bertolini, *J. Ultrastruct. Res.* **11**, 1 (1964); T. Nakao, *J. Comp. Neurol.* **183**, 429 (1979); R. Schultz, E. C. Berkowitz, D. C. Pease, *J. Morphol.* **98**, 251 (1956).
17. R. H. Nordlander and M. Singer, *J. Comp. Neurol.* **180**, 349 (1978).
18. This work was supported by NIH grant PHY 5 RO1 NS08996-09S1 and NIH Spinal Injury Center grant 2 P50 NS10174-07. The skilled technical assistance of A. Breslin is greatly appreciated.

13 August 1979

## Monoclonal Antibodies Defining Distinctive Human T Cell Surface Antigens

Abstract. *Three novel monoclonal antibodies (designated OKT1, OKT3, and OKT4) were generated against surface determinants of human peripheral T cells. Both OKT1 and OKT3 reacted with all human peripheral T cells and 5 to 10 percent of thymocytes but differed in their reactivities with T cell lines. By contrast, OKT4 reacted with 55 percent of human peripheral T cells and 80 percent of thymocytes. All three antibodies were selective for T cells in the peripheral blood in that they did not react with normal B cells, null cells, monocytes, or granulocytes.*

There is now ample evidence from work with mouse and man that lymphocytes, which appear relatively homogeneous morphologically, consist of subclasses of cells with differing functions and that these subclasses have distinctive cell surface antigens that can be identified with appropriate antiserums (*1*, *1a*). In man, such analyses of T cell surface antigens have been made with spontaneous autoantibodies, with heteroantiserums, and with serendipitously discovered markers such as rosetting of sheep erythrocytes (E) and bindings of the Fc (crystallizable fragment) of immunoglobulins (*2*). These techniques have provided valuable scientific insights but have been of limited value in clinical medicine because of the heterogeneity of antibody preparations obtained, frequent low titers, and the multiple absorptions required for production of specific heteroantiserums.

The hybridoma technique of Kohler and Milstein (*3*) provides a new approach to this problem whereby the multispecific responses to complex antigens are reduced to a series of monospecific responses by cloning. A number of hybridomas secreting monoclonal antibodies to various human cell surface antigens have been described (*3*, *4*). We report our methods for preparing monoclonal antibodies to the cell surface antigens of human T cells by immunizing mice, fusing their spleen cells with a mouse myeloma

Table 1. Immunofluorescent binding data from hybridoma supernatants in two fusion experiments. Single cell suspensions from mice immunized with human peripheral $E^+$ cells were mixed with P3 × 63Ag8U, myeloma cells and fused with PEG as described (*3*). After the fusion, cells were washed once in a mixture of DME, 10 percent IFS, and 10 percent IHS, and were resuspended in HAT medium (*5*, *12*). In experiment 1, cells were distributed in 1-ml portions into 100 wells, and in experiment 2, into 384 wells of Costar culture plates (Costar catalog No. 3524). Seven to 10 days after fusion, 0.5 ml of HAT medium (*12*) was added to each well. Supernatant from each well showing hybridoma growth was incubated with $E^+$ and $E^-$ human mononuclear cells and binding was determined by an indirect immunofluorescence assay (*7*). In each experiment supernatants from only two wells showed specific reactivities for $E^+$ cells and these hybridomas were studied further.

| Number of wells showing growth | Binding of supernatant with | |
|---|---|---|
| | $E^+$ cells | $E^-$ cells |
| | *Experiment 1* | |
| 50 | − | Not tested |
| 23 | + | + |
| 2 | + | − |
| | *Experiment 2* | |
| 277 | − | Not tested |
| 32 | + | + |
| 2 | + | − |


0036-8075/79/1019-0347$00.50/0 Copyright © 1979 AAAS

line, and cloning the resultant hybridomas. Furthermore, we emphasize the method that we used successfully to screen for antibodies with binding specificity for T cells, and describe the production of three hybridomas secreting monoclonal antibodies recognizing distinctive human T cell surface antigens with differing cellular representations.

Female mice (8 weeks old) of strains BALB/cJ or $CAF_1$ (Jackson Laboratory) were immunized intraperitoneally with $2 \times 10^7$ E rosette-purified peripheral T cells in PBS at 14-day intervals (*5*). Four days after the third immunization the spleens were removed and a suspension of single cells was prepared. Cell fusions were carried out according to the procedure developed by Kohler and Milstein (*3*). For each fusion, $1 \times 10^8$ splenocytes were fused in 35 percent PEG in 5 percent DMSO and RPMI 1640 with $2 \times 10^7$ P3 × 63Ag8$U_1$ myeloma cells (Table 1).

After cell fusion, the cells were distributed into 100 to 400 wells and cultured in selective medium (see Table 1). The wells were observed regularly under an inverted phase microscope and as cells grew up supernatants from those wells were harvested and tested for binding to human peripheral lymphocytes separated into E rosette-positive ($E^+$) and E rosette-negative ($E^-$) populations (*6*). Binding was detected both by radioimmunosassay and indirect immunofluorescence techniques (*1a*, *7*).

Table 1 summarizes the results from two experiments. In each case the majority of supernatants failed to react with $E^+$ cells and the contents of these wells were discarded. A fair proportion of wells in each experiment yielded supernatants that bound both $E^+$ and $E^-$ cells. Although these wells may have contained antibodies to potentially interesting cell surface molecules they did not meet our criteria for the selection of T cell-specific antibodies and were not studied further. In each of two experiments two wells contained hybridomas secreting T cell-specific monoclonal antibodies. These were cloned and recloned by the limiting dilution method (*8*) to ensure that a single hybridoma was indeed secreting a monoclonal antibody with T cell specificity. That clones were formed was further suggested by the persistence of specific staining patterns in numerous subclones and the restriction of each monoclonal antibody to a single immunoglobulin subclass. The resultant clones were then given the designations OKT1 through OKT4, and larger amounts of antibody were obtained by injecting $10^7$ hybridoma cells into mice primed with 0.5 ml of pristane (Aldrich) 7 days previously. The subsequent malignant ascites from these mice were then harvested and used as reagents to characterize the binding properties of the monoclonal antibodies on different human cell populations.

The results are summarized in Fig. 1 and Table 2. We found that OKT1 and OKT2, which were obtained from the same fusion, were both of the subclass immunoglobulin $G_1$ ($IgG_1$) and had identical patterns of reactivity; they probably represent daughter clones from a single parent clone in the immune spleen. The antibodies OKT1 and OKT3 had similar patterns of reactivity with peripheral T cells and thymocytes. However, OKT3 was of the $IgG_2$ subclass and the antibodies were also clearly distinguishable by complement-fixing properties (nega-

Table 2. Cellular representation and other characteristics of three monoclonal antibodies specific for human T cells. Antibodies OKT1 and OKT3 showed similar patterns of reactivity on peripheral cells and thymocytes but were distinguishable by their differing reactivities with T cell lines, their immunoglobulin class, and their complement-binding properties. The binding pattern of OKT4 was distinct from that of OKT1 and OKT3.

| Tests | Monoclonal antibodies | | |
|---|---|---|---|
| | OKT1 | OKT3 | OKT4 |
| Percentage of reactivity with | | | |
| Peripheral T cells (ten samples) | > 95 | > 95 | 55 |
| Peripheral B cells (ten samples) | < 2 | < 2 | < 2 |
| Peripheral null cells (ten samples) | < 2 | < 2 | < 2 |
| Thymocytes* (eight samples) | 5 to 10 | 5 to 10 | 80 |
| Reactivity with | | | |
| B cell lines† | – | – | – |
| T cell lines† | | | |
| HJD-1 | + | (±) | – |
| CEM | + | – | + |
| Laz 191 | + | – | – |
| HM1 | + | – | – |
| IgG subclass | $IgG_1$ | $IgG_2$ | $IgG_2$ |
| Complement fixation | – | + | + |
| Antibody titer (fluorescence binding) | 30,000 | 100,000 | 50,000 |

*From patients aged 2 months to 18 years. †Obtained from H. Lazarus, Sidney Farber Cancer Institute. The B cell lines Laz 256, 156, 007, and SB were obtained by Epstein-Barr virus transformation of human peripheral B cells; the T cell lines HJD-1, CEM, Laz 191, and HM1 were established from leukemic patients.

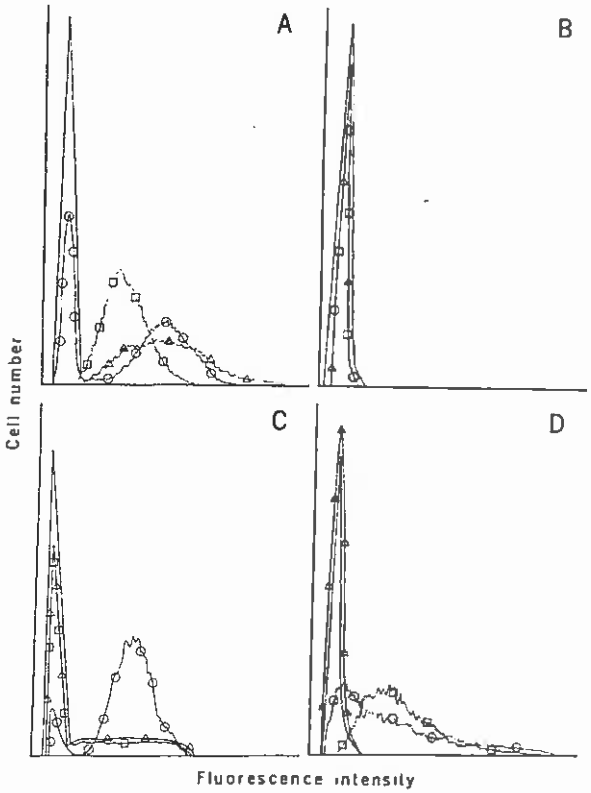



Fig. 1. Cytofluorograf (Ortho Instruments) analysis of fluorescent staining of human cells incubated with ascites from OKT1 (□), OKT3 (△), OKT4 (○), or control ascites induced with P3 × 63Ag8$U_1$ myeloma cells (—). (A) With $E^+$ peripheral blood mononuclear cells; (B) with $E^-$ peripheral blood mononuclear cells; (C) with human thymocytes; and (D) with CEM cells, a human T cell line. Each antibody showed a distinctive pattern of staining with regard to cell distribution and fluorescent intensity.





175

tive for OKT1 and positive for OKT3) and by their differing patterns of reactivity with T cell lines. The binding pattern of OKT4 was quite distinct from OKT1 and OKT3; OKT4 bound only 55 percent of peripheral T cells and 80 percent of thymocytes. Furthermore, the pattern of binding of OKT4 to T cell lines was distinct from that of OKT1 and OKT3. Detailed studies demonstrated that OKT1, OKT3, and OKT4 were all nonreactive with peripheral B cells, monocytes, granulocytes, null cells, and B cell lines (*9*).

We have previously shown that the OKT1$^+$ population of peripheral lymphocytes are indeed the cells responsible for the T cell functions of mitogen reactivity, mixed lymphocyte culture responsivity, and proliferate response to soluble antigens. Furthermore we have shown that OKT1 defines a minor population of thymocytes responding in mixed lymphocyte culture but lacking mitogen responsivity (*9*). Preliminary studies suggest that OKT4$^+$ peripheral cells contain the helper T cell population.

Our present studies demonstrate the generation of monoclonal antibodies identifying antigens restricted to functional subclasses of T cells. Although our procedures entailed immunizing mice with the entire range of antigens represented on human T cells, we were able to retrieve hybrid cells producing monoclonal antibody reacting selectively with T cell subpopulations. Such antibodies should be invaluable for both diagnosis (*10*) and therapy (*11*) in clinical medicine.

PATRICK C. KUNG
GIDEON GOLDSTEIN
*Ortho Pharmaceutical Corporation, Raritan, New Jersey 08869*
ELLIS L. REINHERZ
STUART F. SCHLOSSMAN
*Division of Tumor Immunology, Sidney Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115*

### References and Notes

1. H. Cantor and E. A. Boyse, *Cold Spring Harbor Symp. Quant Biol.* **41**, 23 (1977); L. Chess and S. F. Schlossman, *Contemp. Top. Immunobiol.* **7**, 363 (1977); E. L. Reinherz and S. F. Schlossman, *J. Immunol.* **122**, 1335 (1979).
1a. R. L. Evans, H. Lazarus, A. C. Penta, S. F. Schlossman, *J. Immunol.* **129**, 1423 (1978).
2. A. J. Strelkauskas, V. Schauf, B. S. Wilson, L. Chess, S. F. Schlossman, *ibid.* **120**, 1278 (1978); L. Moretta, M. Ferrarini, A. Mingari, A. Moretta, S. R. Webb, *ibid.* **117**, 2171 (1976); J. H. Wybran and H. H. Fudenberg, *Trans. Assoc. Am. Phys.* **84**, 239 (1971).
3. G. Kohler and C. Milstein, *Nature (London)* **256**, 495 (1975).
4. C. J. Barnstable, W. F. Bodmer, G. Brown, G. Galfa, C. Milstein, A. F. Williams, *Cell* **14**, 9 (1978); M. M. Trucco, J. W. Stocker, R. Ceppellini, *Nature (London)* **273**, 666 (1978); H. Koprowski, Z. Steplewski, D. Herlyn, M. Herlyn, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3405 (1978).
5. Abbreviations: PBS, phosphate-buffered saline; DMA, Dulbecco modified minimum essential medium; PEG, polyethylene glycol; IFS, heat-inactivated fetal calf serum; IHS, heat-inactivated horse serum; DMSO, dimethyl sulfoxide; and HAT, hypoxanthine, aminopterin, and thymidine.
6. N. F. Mendes, M. C. A. Tolnai, N. P. A. Silveira, R. B. Gilbertson, R. S. Mitzgar, *J. Immunol.* **111**, 860 (1973).
7. R. Levy and J. Dilley, *ibid.* **119**, 394 (1977).
8. R. Levy, J. Dilley, L. A. Lampson, *Cur. Top. Microbiol. Immunol.* **81**, 164 (1978).
9. E. L. Reinherz, P. C. Kung, G. Goldstein, S. F. Schlossman, *J. Immunol.* **123**, 1233 (1979); *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4061 (1979).
10. E. L. Reinherz, L. M. Nadler, D. S. Rosenthal, W. C. Moloney, S. F. Schlossman, *Blood* **53**, 1066 (1979).
11. A. B. Cosini, H. H. Wortis, F. L. Delmonico, P. S. Russell, *Surgery* **80**, 155 (1976); R. L. Edelson *et al.*, *Cancer Chemother. Rep.* **63**, 675 (1979).
12. D. H. Margulies, W. M. Kuel, M. D. Scharff, *Cell* **8**, 405 (1976).
13. We thank M. D. Scharff (Albert Einstein College of Medicine) for providing the P3 × 63Ag8U.1 myeloma cell line for cell fusion; A. Penta, C. O'Brien, J. McDowell, K. Singer, M. A. Talle, and M. DeMaria for technical assistance; and L. M. Grappi and C. Brain for secretarial assistance. The research was supported in part by NIH training grant CA 09172 and NIH grants AI 12069, AI 19589, and CA 06516.

7 May 1979; revised 20 June 1979

## Ciliary Membrane Alterations Occurring in Experimental *Mycoplasma pneumoniae* Infection

Abstract. *Experimental infection of hamster ciliated tracheal epithelium in organ culture with virulent* Mycoplasma pneumoniae *resulted in the deterioration of ciliary necklaces and an altered distribution of membrane-associated particles on the shafts of the affected cilia. To our knowledge this is the first report of an altered disposition of ciliary membrane-associated particles in response to a specific infectious agent.*

Infections due to *Mycoplasma pneumoniae* are a major cause of respiratory disease in children and young adults (*1*). The general clinical spectrum of the disease ranges from upper respiratory tract infections, bronchitis, and tracheitis to primary atypical pneumonia (*1*). The disease is self-limiting and rarely fatal (*2*).

Investigations of the cellular basis of experimental and natural *M. pneumoniae* disease (*2–4*) have revealed ciliostasis and degenerating abnormal epithelial cells. Ultrastructural and biochemical studies also have demonstrated that in order to be fully pathogenic the organism must be specifically oriented to the host cell epithelial surface by means of a specialized terminal organelle (*4*). With the use of the freeze-etch technique, a new facet of the pathogenesis of experimental *M. pneumoniae* disease has been revealed. In these experiments, freeze-etching of hamster tracheal epithelium after incubation in organ culture with virulent *M. pneumoniae* demonstrated marked configurational alterations of the membrane-associated particles of the cilia (*5*).

Inspection of the freeze-etch replicas showed that control epithelial cells maintained in organ culture up to 48 hours retained ultrastructural integrity with their in vivo correlates (Fig. 1, A and B). A particularly prominent feature revealed by freeze-etching was the ciliary necklace, a complex of membrane-associated particles at the base of each cilium (*6*). The ciliary necklaces of epithelial cell cilia from tracheae fixed immediately after resection were comprised of generally five rows of evenly spaced particles at the base of each cilium which extended up the shaft a distance of approximately 0.2 μm (Fig. 1A). Other membrane-associated particles appeared scattered randomly and uniformly over the fractured surfaces of the cilia (Fig. 1A). This configuration of ciliary necklace and membrane-associated particles was reliably retained with minimum variability in control organ culture specimens for up to 48 hours (Fig. 1B). The membrane-associated particles of the ciliary necklaces and on the shafts of the cilia measured approximately 100 Å in diameter in both freshly fixed and organ culture specimens.

Replicas from the experimentally infected specimens showed that the highly organized ciliary necklaces were severely altered within the first 24 hours. In some cases, roughly organized remnants of the necklaces were observed; however, the general view of this area illustrated the complete disarray of the necklace particles and, in many cases, the ciliary necklace was completely absent (Fig. 1C). Membrane-associated particles observed over the ciliary shaft in infected specimens also measured approximately 100 Å in diameter and appeared clumped and unevenly distributed.

Our studies have indicated that a relatively heavy inoculum of the pathogen is necessary to produce disorganization and decreased frequency of ciliary beat in vitro. In the infected tissues, *M. pneumoniae* is observed frequently inter-


0036-8075/79/1019-0349$00.50/0 Copyright © 1979 AAAS

# Muromonab-CD3

From Wikipedia, the free encyclopedia

**Muromonab-CD3** (trade name **Orthoclone OKT3**, marketed by Janssen-Cilag) is an immunosuppressant drug given to reduce acute rejection in patients with organ transplants.[1][2] It is a monoclonal antibody targeted at the CD3 receptor, a membrane protein on the surface of T cells. It was the first monoclonal antibody to be approved for clinical use in humans.[2]

**Muromonab-CD3** ?

| | |
|---|---|
| **Monoclonal antibody** | |
| **Type** | Whole antibody |
| **Source** | Mouse |
| **Target** | CD3ε |
| **Clinical data** | |
| **Pregnancy cat.** | C(US) |
| **Legal status** | ℞ Prescription only |
| **Routes** | intravenous |
| **Pharmacokinetic data** | |
| **Bioavailability** | N/A |
| **Identifiers** | |
| **CAS number** | 140608-64-6 |
| **ATC code** | L04AA02 |
| **ChEMBL** | CHEMBL1201608 ✘ |
| **Chemical data** | |
| **Formula** | $C_{6460}H_{9946}N_{1720}O_{2043}S_{56}$ |
| **Mol. mass** | 146.091 kDa |

✘(what is this?) (verify)

## Contents


- 1 History
- 2 Indications
- 3 Pharmacodynamics and chemistry
- 4 Adverse effects
- 5 Contraindications
- 6 Etymology
- 7 References


## History

Muromonab-CD3 was approved by the U.S. Food and Drug Administration (FDA) in 1986,[3] making it the first monoclonal antibody to be approved anywhere as a drug for humans. In the European Communities, it was the first drug to be approved under the directive 87/22/EWG, a precursor of the European Medicines Agency (EMEA) centralised approval system in the European Union. This process included an assessment by the Committee for Proprietary Medicinal Products (CPMP, now CHMP), and a subsequent approval by the national health agencies; in Germany, for example, in 1988 by the Paul Ehrlich Institute in Frankfurt.

## Indications

Muromonab-CD3 is approved for the therapy of acute, glucocorticoid resistant rejection of allogeneic renal, heart and liver transplants.[4] Unlike the monoclonal antibodies basiliximab and daclizumab, it is not approved for prophylaxis of transplant rejection, although a 1996 review has found it to be safe for that purpose.[3]

It has also been investigated for use in treating T-cell acute lymphoblastic leukemia.[5]

## Pharmacodynamics and chemistry

177

T cells recognise antigens primarily via the T cell receptor. This receptor needs various co-receptors to function, one of which is CD3. The T cell receptor-CD3 complex transduces the signal for the T cell to proliferate and attack the antigen.

Muromonab-CD3 is a murine (mouse) monoclonal IgG2a antibody which was created using hybridoma technology. It binds to the T cell receptor-CD3-complex (specifically the CD3 epsilon chain) on the surface of circulating T cells, initially leading to an activation, but subsequently inducing blockage and apoptosis of the T cells. This protects the transplant against the T cells.[2][4] After application of muromonab-CD3, normal T cell function is said to be restored within a week. When administered for transplant induction, the drug is administered daily thereafter for up to 7 days.[6]

Newer monoclonal antibodies with the same mechanism of action include otelixizumab, teplizumab and visilizumab. They are being investigated for the treatment of other conditions like Crohn's disease, ulcerative colitis and type 1 diabetes.

## Adverse effects

Especially during the first infusion, the binding of muromonab-CD3 to CD3 can activate T cells to release cytokines like tumor necrosis factor and interferon gamma. This cytokine release syndrome, or CRS, includes side effects like skin reactions, fatigue, fever, chills, myalgia, headaches, nausea and diarrhea,[7] and could lead to life-threatening conditions like apnoea, cardiac arrest, and flash pulmonary edema.[4] To minimize the risk of CRS and to offset some of the minor side effects patient experience, glucocorticoids (such as methylprednisolone), acetaminophen, and diphenhydramine are given before the infusion.[8]

Other adverse effects include leucopenia, as well as an increased risk for severe infections and malignancies typical of immunosuppressive therapies. Neurological side effects like aseptic meningitis and encephalopathy have been observed. Possibly, they are also caused by the T cell activation.[4]

Repeated application can result in tachyphylaxis (reduced effectiveness) due to the formation of anti-mouse antibodies in the patient, which accelerates elimination of the drug. It can also lead to an anaphylactic reaction against the mouse protein,[2] which may be difficult to distinguish from a CRS.

## Contraindications

Except under special circumstances, the drug is contraindicated for patients with an allergy against mouse proteins, as well as patients with uncompensated heart failure, uncontrolled arterial hypertension or epilepsy. It should not be used during pregnancy or lactation.[2][4]

## Etymology

Muromonab-CD3 was developed before the WHO nomenclature of monoclonal antibodies took effect, and consequently its name does not follow this convention. Instead, it is a contraction from "*mur*ine *mon*oclonal *a*nti*b*ody targeting *CD3*".[2]

## References

1. ^ Midtvedt K, Fauchald P, Lien B, *et al.* (February 2003). "Individualized T cell monitored administration of ATG versus OKT3 in steroid-resistant kidney graft rejection" (http://www.blackwell-synergy.com/openurl?genre=article&sid=nlm:pubmed&issn=0902-0063&date=2003&volume=17&issue=1&spage=69) . *Clinical transplantation* **17** (1): 69–74. PMID 12588325 (http://www.ncbi.nlm.nih.gov/pubmed/12588325) . http://www.blackwell-synergy.com/openurl?genre=article&sid=nlm:pubmed&issn=0902-0063&date=2003&volume=17&issue=1&spage=69.
2. ^ *a b c d e f* Mutschler, Ernst; Gerd Geisslinger, Heyo K. Kroemer, Monika Schäfer-Korting (2001) (in German). *Arzneimittelwirkungen* (8 ed.). Stuttgart: Wissenschaftliche Verlagsgesellschaft. p. 937. ISBN 3-8047-1763-2.
3. ^ *a b* Smith, SL (1996). "Ten years of Orthoclone OKT3 (muromonab-CD3): a review". *Journal of transplant coordination : official publication of the North American Transplant Coordinators Organization (NATCO)* **6** (3): 109–19; quiz 120–1. PMID 9188368 (http://www.ncbi.nlm.nih.gov/pubmed/9188368) .
4. ^ *a b c d e* "Orthoclone OKT3" (http://www.drugs.com/mmx/orthoclone-okt3.html) . *Professional Drug Information*. Drugs.com. http://www.drugs.com/mmx/orthoclone-okt3.html. Retrieved 3 January 2010.
5. ^ Gramatzki M, Burger R, Strobel G, *et al.* (March 1995). "Therapy with OKT3 monoclonal antibody in refractory T cell acute lymphoblastic leukemia induces interleukin-2 responsiveness". *Leukemia* **9** (3): 382–90. PMID 7885036 (http://www.ncbi.nlm.nih.gov/pubmed/7885036) .
6. ^ Bhorade, S. M., Stern, E. (2009). Immunosuppression for Lung Transplantation. Proc Am Thorac Soc 6: 47-53
7. ^ Abramowicz; Schandene, L; Goldman, M; Crusiaux, A; Vereerstraeten, P; De Pauw, L; Wybran, J; Kinnaert, P et al. (1989). "Release of tumor necrosis factor, interleukin-2, and gamma-interferon in serum after injection of OKT3 monoclonal antibody in kidney transplant recipients". *Transplantation* **47** (4): 606–8. doi:10.1097/00007890-198904000-00008 (http://dx.doi.org/10.1097%2F00007890-198904000-00008) . PMID 2523100 (http://www.ncbi.nlm.nih.gov/pubmed/2523100) .
8. ^ Bhorade, S. M., Stern, E. (2009). Immunosuppression for Lung Transplantation. Proc Am Thorac Soc 6: 47-53


Retrieved from "http://en.wikipedia.org/wiki/Muromonab-CD3"
Categories: Immunosuppressants


- This page was last modified on 6 July 2011 at 03:32.
- 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.


179

**Chimeric therapeutic antibody**

FDA approved the first chimeric antibody rituximab in 1997 after three years of testing (Reff ME, Carner K, Chambers KS, Chinn PC, Leonard JE, Raab R, Newman RA, Hanna N, Anderson DR (1994) Depletion of B cells in vivo by a chimeric mouse human monoclonal antibody to CD20. Blood, 83, 435-445), even though the idea of joining mouse antibody variable regions to human constant regions was successfully achieved in 1984 (Morrison SL, Johnson MJ, Herzenberg LA, Oi VT (1984) Chimeric human antibody molecules: Mouse antigen binding domains with human constant region domains. Proc Natl Acad Sci, USA, 81, 8651-8655). This process reduced the amount of mouse protein in the antibody to be injected into human patients, and retained the antibody specificity against the target antigen.

The basic idea makes use of the pioneering discovery of Hilschmann and Craig that Bence Jones proteins can be divided into two halves, an N-terminal variable region and a C-terminal constant region. Similarly, antibody heavy chains can also be divided into variable and constant regions.

The mouse light chain variable region gene was isolated from a hybridoma cell line with desired specificity and inserted into a plasmid molecule. The light chain constant region gene from a human antibody was also inserted into the same plasmid with proper orientation. This genetically engineered plasmid molecule using recombinant DNA technique can then be internalized into a non-secretor myeloma cell line. Similarly, another plasmid molecule was used to carry the mouse heavy chain variable region from the same hybridoma cell line and the heavy chain constant regions of the human antibody. That recombinant plasmid molecule was inserted into the same non-secreting myeloma cell line.

This process involves the isolation of four gene segments, two from the mouse hybridoma cell line and two from a human antibody. The genetic manipulation is relatively straight forward of putting these gene segments into plasmid molecules in proper orientations. It was clearly illustrated by Morrison et al. mentioned above.

The amount of mouse protein in the chimeric antibody is only roughly about one third of the entire molecule of the therapeutic antibody. Therefore, the chance of being rejected by the human patient is also corresponding reduced. Roughly, mouse antibodies have half-lives of about one to four days in human patients, while chimeric antibodies of about four to fifteen days.

The mechanisms of action of chimeric therapeutic antibodies are, however, very complicated. As discussed by the original authors for rituximab mentioned above, they may involve the activation of the complement system, the recruitment of killer cells, sensitization to apoptotic signals, etc. In most of therapeutic uses of rituximab, its half-live ranges from 2.5 to 7.5 days. Therefore, weekly infusion is needed. In addition, due to its antigenicity, in most cases, only four infusions can be tolerated by human patients. Therefore, further refinement of therapeutic antibodies from mouse hybridoma cell lines will be essential for prolonged treatment of patients.

# Depletion of B Cells In Vivo by a Chimeric Mouse Human Monoclonal Antibody to CD20

By Mitchell E. Reff, Kristin Carner, Karen S. Chambers, Paul C. Chinn, John E. Leonard, Ron Raab, Roland A. Newman, Nabil Hanna, and Darrell R. Anderson

**Murine monoclonal antibody 2B8 specifically recognizes the CD20 phosphoprotein expressed on the surface of normal B lymphocytes and B-cell lymphomas. The light- and heavy-chain variable regions of 2B8 were cloned, after amplification by the polymerase chain reaction, into a cDNA expression vector that contained human IgG1 heavy chain and human κ-light chain constant regions. High-level expression of chimeric-2B8 antibody (C2B8) was obtained in Chinese hamster ovary cells. Purified C2B8 exhibited antigen binding affinity and human-tissue reactivity similar to the native murine antibody. In vitro studies showed the ability of C2B8 to bind human C1q, mediate complement-dependent cell lysis of human B-lymphoid cell lines, and lyse human target cells through antibody-dependent cellular cytotoxicity. Infusion of macaque cynomolgus monkeys with doses ranging from 1.6 mg/kg to 6.4 mg/kg resulted in greater than 98% depletion of peripheral blood (PB) B cells and 40% to 70% depletion of lymph node B cells. Recovery of PB B cells usually started at 2 weeks after treatment and required 60 to greater than 90 days to reach normal levels. As much as 95% depletion of B cells in peripheral lymph nodes and bone marrow was observed following weekly injections of 16.8 mg/kg antibody. No toxicity was observed in any of the animals. These results offer the possibility of using an "immunologically active" chimeric anti-CD20 antibody as an alternative approach in the treatment of B-cell lymphoma.**

*© 1994 by The American Society of Hematology.*

B LYMPHOCYTES ARISE from pluripotent stem cells and differentiate to fully mature antibody-secreting plasma cells.[1] The human B-lymphocyte–restricted differentiation antigen Bp35 (CD20), is a cell surface nonglycosylated hydrophobic phosphoprotein of 35 kD, and is expressed during early pre-B–cell development just before the expression of cytoplasmic μ-heavy chains.[2] Uncommitted hematopoietic-precursor stem cells do not express CD20. CD20 is expressed consistently until the plasma cell stage and regulates a step in the activation process that is required for cell-cycle initiation and differentiation.[3-7]

The CD20 molecule expressed on neoplastic B cells provides a promising target for therapy of B-cell lymphomas and leukemia.[8] It is especially suitable as a target for antibody-mediated therapy because of accessibility and sensitivity of hematopoietic tumors to lysis by way of immune-effector mechanisms. The CD20 antigen is not normally shed from the cell surface, and there are no detectable serum levels of soluble CD20, which might block targeting of antibody to lymphomas.[9] CD20 does not internalize after binding to antibody.[10] Earlier studies have shown that partial clinical responses were achieved in lymphoma patients using high doses of murine anti-CD20 monoclonal (MoAb).[8]

We have identified a high-affinity murine anti-CD20 MoAb and constructed a mouse/human chimeric antibody that is expressed at high levels in mammalian cells. We have shown that the chimeric MoAb has specificity and affinity equivalent to the native-murine parent. However, unlike the murine antibody, the chimeric antibody binds human C1q, and mediates complement-dependent cell lysis (CDCC) in the presence of human complement, and antibody-dependent cellular cytotoxicity (ADCC) with human effector cells. Chimeric anti-CD20 antibody, when administered to cynomolgus monkeys, caused a marked B-cell depletion in peripheral blood (PB), bone marrow (BM), and lymphatic tissue.

## MATERIALS AND METHODS

*Production of murine anti-CD20 MoAb.* Mice from the BALB/c strain were immunized weekly with the human lymphoblastoid cell line SB.[11] Spleens were taken from mice with high serum titers of anti-CD20 antibodies, and the splenocytes fused with the mouse myeloma SP2/0.[12] All assays for CD20 reactivity were accomplished by radio immunoassay. Briefly, purified anti-CD20 B1 (Coulter Corp, Hialeah, FL) was radiolabeled with $^{125}I$ by the iodobead method.[13] Hybridomas were screened by coincubation with $^{125}I$-B1 (10 ng) in 1% bovine serum albumin (BSA), phosphate-buffered saline (PBS), and 100,000 SB cells. After incubation for 1 hour at room temperature, the cells were harvested by transferring to 96-well filter plates (V&P Scientific, San Diego, CA) and washed thoroughly. Duplicate wells containing unlabeled B1 and wells containing no inhibiting antibody were used as positive and negative controls, respectively. Wells containing greater than 50% inhibitions were expanded and cloned. The antibody showing the highest reactivity was derived from the cloned cell line designated 2B8.

*Construction of the chimeric anti-CD20 Ig DNA expression vector.* RNA was isolated from the 2B8 mouse hybridoma cell[14] from which single-stranded cDNA was prepared. Mouse Ig light-chain variable-region DNA was amplified from the cDNA by the polymerase chain reaction (PCR) using a set of DNA primers with homology to mouse light-chain signal sequences at the 5' end and mouse light-chain J region at the 3' end (see Fig 1 for primers). This DNA fragment was inserted directly into an expression plasmid in front of the human κ-light–chain constant domain and the entire construct sequenced. Two clones from two separate PCR reactions were sequenced to assure that no PCR errors occurred. The κ gene in the expression plasmid contained a synthetic signal sequence, unique restriction sites for the insertion of a light-chain variable domain, which maintain the reading frame and conserve the amino acids (Fig 2), and the human κ constant region (amino acid 108-214 Kabat numbering[15]). A schematic drawing of the expression plasmid can be found in Fig 3. The mouse light-chain variable region from 2B8 is in the mouse κ VI family.[15] Similarly, the mouse

*From IDEC Pharmaceuticals Corp, San Diego, CA.*

*Submitted April 5, 1993; accepted September 21, 1993.*

*Address reprint requests to Darrell R. Anderson, PhD, IDEC Pharmaceuticals Corporation, 11011 Torreyana Rd, San Diego, CA 92121.*

*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. section 1734 solely to indicate this fact.*

*© 1994 by The American Society of Hematology.*

*0006-4971/94/8302-0021$3.00/0*

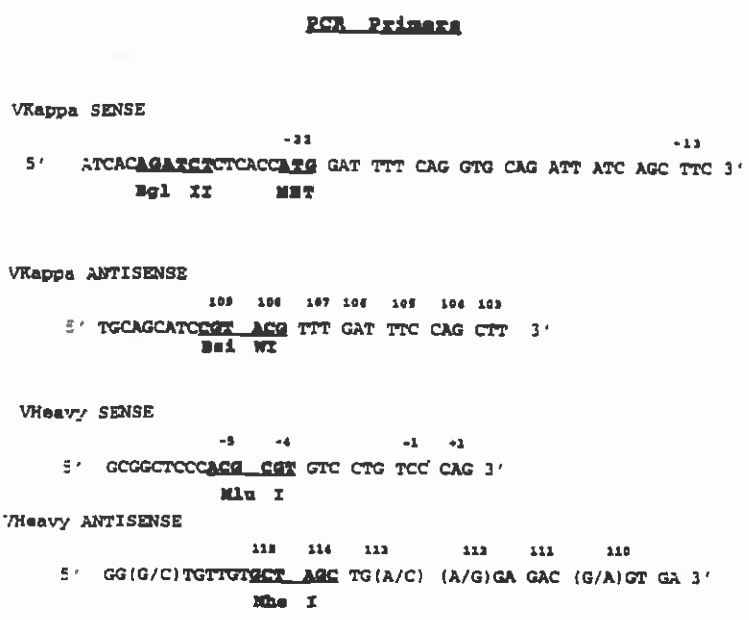



**Fig 1. DNA primers utilized in the PCR reactions to amplify the light-chain (VKappa) and heavy-chain (VHeavy) variable-region DNA.**

heavy-chain variable region was isolated from the 2B8 hybridoma and inserted in front of human IgG1 constant domains and sequenced (see Fig 1 for primers). The γ-1 gene construct in the expression plasmid contained a synthetic-leader sequence, unique restriction sites for insertion of a heavy-chain variable domain (Fig 2), and the human γ-1 constant region (amino acids 114 through 478 Kabat numbering[15]). The mouse heavy-chain variable region from 2B8 is in the mouse VH 2B family.[15] The synthetic signal sequences were derived from consensus signal sequences of Ig genes.[13]

*Creation of C2B8 producing transfectomas.* Chinese hamster ovary (CHO) cells (DG44)[16] were grown in SSFM II minus hypoxanthine and thymidine (GIBCO). Approximately 25 μg of plasmid DNA was inserted into $4 \times 10^6$ CHO cells, by electroporation using a BTX 600 electroporation device (BTX, San Diego, CA). The plasmid DNA had been previously restricted with the enzyme *Not* 1, which separates the genes expressed in mammalian cells from the portion used to grow the plasmid in bacteria (see Fig 3). Conditions for electroporation were 210 V, 400 microfaradays, 13 ω. Each electroporation was plated into six 96-well dishes (about 7,000 cells/well). Dishes were fed with media containing G418 (Geneticin; GIBCO) at 400 μg/mL active compound 2 days after electroporation, and thereafter every second or third day until colonies arose. Supernatant from colonies was assayed for the presence of chimeric Ig by an enzyme-linked immunosorbent assay specific for human antibody. Colonies producing the highest amount of Ig were expanded and plated out into 96-well dishes containing media plus methotrexate (5 nmol/L) and fed every 2 or 3 days. Supernatant was assayed as above and colonies producing the highest amount of Ig were expanded. Chimeric antibody (C2B8) was purified from supernatant using protein-A affinity chromatography.

*Immunoreactivity of C2B8.* Purified antibodies produced by CHO cells were radiolabeled with $^{125}I$ or with fluorescein isothiocyanate (FITC). They were tested by direct binding to various CD20$^+$ target cells and detected either by gamma counter or by flow cytometry using a FACScan analyzer (Becton Dickinson). The B-lymphoblastoid SB cell line was used as a positive control to examine direct binding of C2B8. The CD20$^-$ T-cell leukemia line, HSB, was used as a negative control.[17] Competitive inhibition of binding of C2B8 to these cells was studied and compared with that observed for murine anti-CD20 antibodies 2B8 and B1.[4]

*In vitro effector function of C2B8.* A whole cell cytometry method for detection of human C1q binding was developed using the CD20$^+$ cell line SB. In this assay, chimeric-C2B8 and murine-2B8 antibodies are incubated for 30 minutes at room temperature followed by washing with PBS (pH 7.4), 1% BSA. After washing, the cells are incubated for an additional 30 minutes with 10 μg/mL C1q (Quidel, San Diego, CA), previously labeled with FITC. The cells were washed again and fixed with 0.15 mol/L NaCl, 1% paraformaldehyde. Negative-binding histograms were determined by SB cells incubated with irrelevant human IgG1 followed by addition of C1q-FITC and by incubation of C1q-FITC with SB cells in the absence of antibody.

Complement-dependent cytotoxicity and ADCC effector assays were accomplished by using $^{51}Cr$-labeled SB cells that were exposed to antibodies and human serum, or activated PB effector cells (effector:target = 100:1), for a period of 4 hours at 37°C.[18] Control studies were performed simultaneously with the CD20$^-$ cell line HSB. The percentage of specific lysis was calculated according to the following formula:

$$\% \text{ Specific Lysis} = \frac{\text{Experimental} - \text{Spontaneous Release}}{\text{Maximum} - \text{Spontaneous Release}} \times 100$$

Maximum release was determined by lysis in 0.1% Triton X-100. Assays used 10,000 SB (CD20$^+$) cells or 10,000 HSB (CD20$^-$) cells. Results were obtained in triplicate and averaged for both cell lines.

*Fluorescent antibody staining of primate lymphoid cell population.* After removal of plasma, primate blood cells were washed twice in Hanks' Balanced Salt Solution and resuspended in a plasma-equivalent volume of fetal bovine serum. A 0.2-mL volume of the blood preparation was distributed to each of four 15-mL conical centrifuge tubes. Human-specific fluorescent-labeled MoAbs, which corecognize monkey-lymphocyte surface markers CD2 (AMAC, Westbrook, ME) and CD20 (Becton Dickinson), were

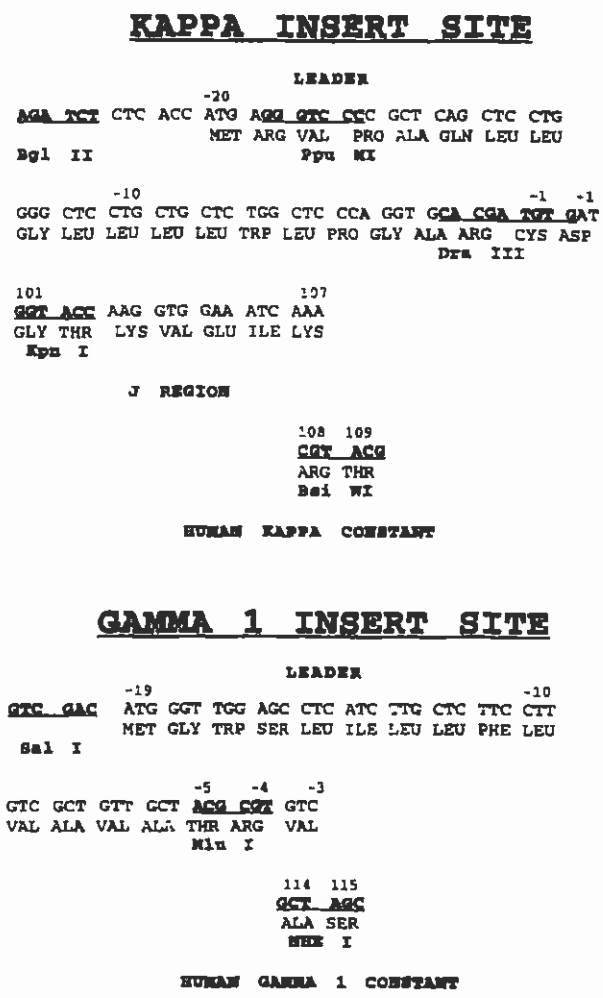



**Fig 2. DNA sequence of the Ig vector indicating restriction endonuclease sites where variable-region DNA can be cloned. The PCR-DNA fragments of anti-CD20 were cloned into the *Bgl* II and *Bs*WI (Kappa) and *Mlu* I and *Nhe* I (Gamma 1) sites, respectively.**

added to two of the tubes for identifying T- and B-lymphocyte populations. C2B8 antibody bound to monkey B-cell surface CD20 was measured in the third tube using polyclonal goat-antihuman IgG coupled with phycoerythrin (AMAC). This reagent was preabsorbed on a monkey IgG-Sepharose column to eliminate background staining of monkey surface-IgG$^+$ B cells. A fourth sample was included for autofluorescence. Cells were incubated with antibodies for 30 minutes, washed and fixed with 0.5 mL of fixation buffer (0.15 mol/L NaCl, 1% paraformaldehyde), and analyzed on a Becton Dickinson FACScan instrument. Fluorescence measurements reflected only gated lymphocyte events.

*Depletion of B cells from primate PB.* Four cynomolgus monkeys ranging in weight from 4.5 to 7 kg received doses of C2B8 corresponding to 0.01 mg/kg, 0.1 mg/kg, 0.4 mg/kg, and 1.6 mg/kg each day for 4 consecutive days. Therefore, the dose range was 0.04 to 6.4 mg/kg body weight. The antibody was administered by intravenous infusion with blood samples drawn before each infusion. Additional blood samples beginning 24 hours after the last injection were taken at days 1, 2, 4, 8, 15, 29, and subsequently thereafter at biweekly intervals until completion of the study at day 90. Approximately 5 mL of whole blood from each animal was centrifuged at 2,000 rpm for 5 minutes. The pellet containing PB leukocytes and red blood cells was resuspended in fetal calf serum for staining with fluorescent antibodies.

*Depletion of B cells from primate lymph nodes.* Four cynomolgus monkeys ranging in weight from 4 to 6 kg were divided into two groups of two monkeys. One group was injected intravenously with a dose of 1.6 mg/kg (0.4 mg/kg each day for 4 consecutive days). The animals in the other group were injected with a single dose of 6.4 mg/kg. Animals were bled after the last injection at days 1, 2, 4, 8, 15 and 22, and blood was processed the same as before for fluorescent antibody staining. Lymph node biopsies from the inguinal area of the leg were taken at days 15 and 29 with cell preparations stained for quantitation of lymphocyte populations by flow cytometry. A lymph node biopsy was also taken from a normal animal injected with saline. The percent of B lymphocytes in the total lymphocyte fraction, detected by flow cytometry in the normal animal, was used as a reference for comparing the percentage of B lymphocytes in the lymph nodes of the treated animals. Depletion of lymph nodes was quantitated by assuming the fraction of B lymphocytes in normal saline-treated animals to be approximately 40%. Reduced amounts of B cells found in treated lymph nodes were calculated as percent depleted B cells by the following formula:

$$\% \text{ Depletion} = 100 - \frac{\% \text{ B Cells in Treated Lymph Node}}{\% \text{ B Cells in Saline Control}}$$

*High-dose in vivo pharmacology study.* Four monkeys were infused with 16.8 mg/kg C2B8 at weekly intervals over a period of 4 consecutive weeks. At 22 days beyond the end of the dosing period, two animals were anesthetized for removal of BM and lymph node biopsies. At 36 days beyond the end of the dosing period, the two remaining animals were anesthetized for BM and lymph node biopsies. Both sets of tissues were stained for the presence of B lymphocytes using the CD20 marker Leu-16 by flow cytometry.[19]

*Evaluation of animal systemic and tissue toxicity.* All animals used in the in vivo studies were routinely examined for toxic side effects during the course of the studies by routine blood chemistry, urinalysis, and physical examination including weight loss and loss of appetite. Two of the four animals in the multiple high-dose study were killed at the end of the study at 22 days after final dosing. Body tissues were thoroughly examined for evidences of toxicity and tissue damage.

## RESULTS

*Construction and expression of chimeric 2B8 antibody.* The light and heavy chain variable regions from the murine 2B8 anti-CD20 antibody were cloned using PCR amplification and inserted directly into a cDNA mammalian chimeric antibody expression vector. The expression vector (TCAE see Fig 3) contains four separate eukaryotic cDNA genes arranged in a tandem head to tail fashion, each containing its own promoter and polyadenylation region. The four genes are the human $\kappa$ gene (allotype Km3),[15] the human $\gamma$-1 constant gene (allotype z, a),[15] a neomycin phosphotransferase gene (NEO),[20] and a mouse dihydrofolate reductase (DHFR) gene.[21] The NEO gene is used for selecting mammalian cell clones that have integrated the gene and are resistant to the antibiotic G418.[19] The DHFR gene is used for gene amplification in the presence of the competitive substrate inhibitor methotrexate.[22]

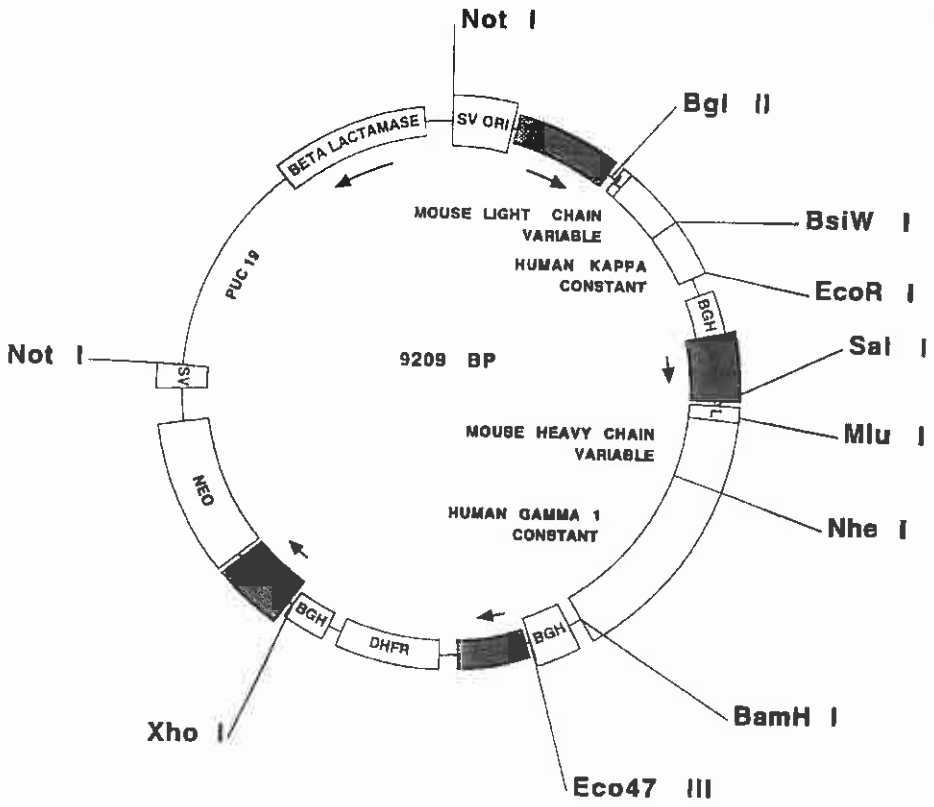



Fig 3. Schematic representation of the DNA expression vector tandem chimeric-antibody expression used to express anti-CD20 in CHO cells. L, leader; CMV, cytomegalovirus promoter; BETA, mouse β-globin major promoter; BGH, bovine growth hormone polyadenylation; SV, SV40 early polyadenylation; SV ORI, SV40 origin.

After electroporation into CHO cells and selection in G418, drug-resistant clones were isolated from cells that secreted 2 $\mu g/10^6$ cells/d. Cell lines have been derived in 5 nmol/L methotrexate in a single amplification step that secrete stable Ig levels at 30 $\mu g/10^6$ cells/d.

*Antibody specificity and immunoreactivity.* Chimeric anti-CD20 was purified by protein-A affinity chromatography and estimated to be greater than 95% pure by electrophoresis on polyacrylamide gels. It was first tested in direct and competitive binding assays and compared with murine 2B8, where it was shown to be of comparable affinity and specificity on CD20$^+$ SB cell line. The apparent-affinity constant (Kap) was determined by direct binding of radiolabeled C2B8 and compared with radiolabeled murine 2B8 by Scatchard analysis. The Kap of C2B8 was estimated as $5.2 \times 10^{9}$ mol/L (Fig 4), compared with a Kap of approximately $3.5 \times 10^{9}$ mol/L for the murine antibody. Direct competition by radioimmunoassay was used to confirm both the specificity and retention of immunoreactivity of the chimeric antibody by comparing its ability to effectively compete with binding of native murine antibody. Data illustrated in Fig 5 shows that virtually equivalent amounts of chimeric and native murine antibody were required to inhibit the binding of the parent murine-2B8 antibody to CD20 sites on SB cells.

*In vitro functional characterization of C2B8 antibody.* C2B8 was evaluated for human C1q binding by flow cytometry, as shown in Fig 6. In this assay, when C2B8 was incubated with SB cells followed by the addition of fluorescein-labeled C1q, a significant increase in fluorescent intensity was observed. In contrast, under the same conditions, murine-2B8 antibody, as well as an irrelevant human IgG1 (data not shown), failed to bind human C1q.

C2B8 was then tested for its ability to lyse B-lymphoid cell lines in the presence of human serum as a source of complement. Approximately 50% of the SB target cells (Fig 7) were lysed in the presence of a 2.2 $\mu g$ concentration of C2B8 antibody and a 1:4 dilution of human complement (Pel Freeze, Rogers, AK). In contrast, no significant lysis was observed in experiments using CD20$^+$ SB cells incubated with murine-2B8 antibody, or CD20$^-$ HSB cells incubated with C2B8 antibody. A summary of data from an ADCC assay using C2B8 and activated human effector cells at a 100:1 effector:target ratio is presented in Fig 8. Approximately 50% of the antigen$^+$ SB cells were lysed during the 4-hour incubation period. In contrast, CD20$^-$ HSB cells were not lysed nor were CD20$^+$ cells incubated with murine-2B8 antibody.

*In vivo depletion of PB B lymphocytes.* Cynomolgus monkeys received doses of C2B8 corresponding to 0.01, 0.1, 0.4, and 1.6 mg/kg daily for a total of 4 consecutive days resulting in total doses ranging from 0.04 to 6.4 mg/kg. The data in Fig 9 shows that doses of 0.4 to 6.4 mg/kg depleted greater than 95% of PB B cells for as long as 8 days postinfusion. The lowest dose (0.04 mg/kg) effectively depleted greater than 50% of the peripheral B cells. It is also apparent

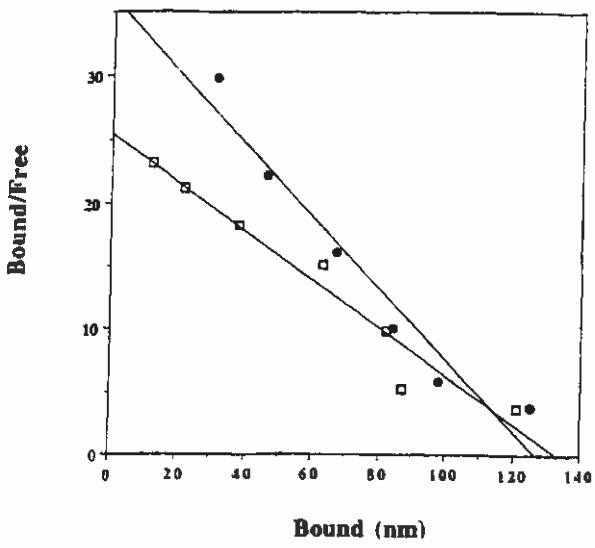
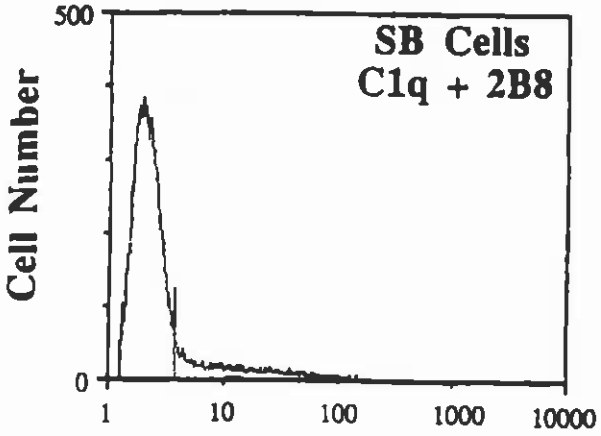



Fig 4. Determination of Kap for chimeric antibodies produced in CHO transfectoma. Antibodies produced in the CHO cell and the original murine cell line were purified, radiolabeled with $I^{125}$, and evaluated by direct binding to CD20 antigen expressed on 10,000 human SB cells. Apparent binding affinities were determined by Scatchard plots. (●) 2B8; (□) C2B8.

from the CD20 marker, Leu-16, that saturation of the CD20 antigen on the B-lymphocyte population in PB was not achieved with this dose of C2B8 antibody (Table 1). Table 1 compares the percentage of cells coated with C2B8 and the percentage of $CD20^+$ cells. The percentage of C2B8-coated cells at the lowest dose is basically identical to the $CD20^+$ cells at 24 hours. By day 2, the percent of C2B8-coated cells is only half that of the population of $CD20^+$ cells. Total saturation at 24 hours would have blocked binding of Leu-16 to these cells, as both reagents compete with each other for the same site on SB cells. At 0.04 and 0.4 mg/kg dose levels, B lymphocytes coated with C2B8 antibody were detected in the circulation during the initial 4 days after antibody infusion. By day 8, C2B8 antibody-coated cells at all doses were undetectable in PB. The levels of C2B8-coated cells were very low at the higher doses, presumably because of the absence of B cells by 24 hours. After day 8, the B cells began



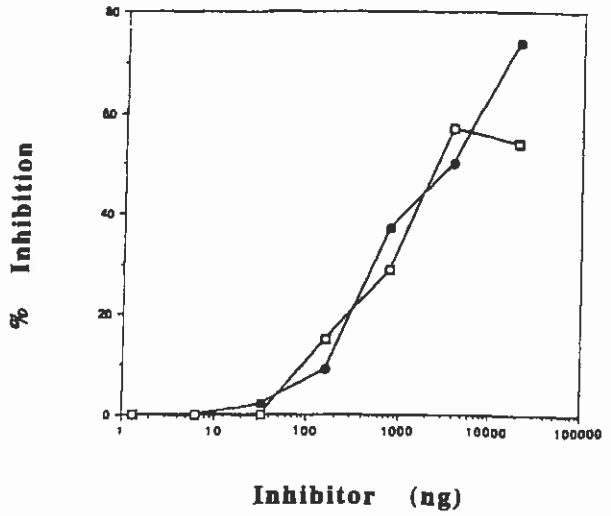




Fig 5. Immunoreactivity of purified antibodies produced by CHO transfectoma. Unlabeled chimeric 2B8 was incubated at various concentrations in the presence of 10,000 B-lymphoid cells and 10 ng of $I^{125}$-labeled native-murine 2B8. The amounts of chimeric antibody required to directly compete with the radiolabeled murine anti-CD20 antibody were compared with similar amounts of unlabeled murine antibody. Results indicate minimal loss of inhibiting activity due to chimerization. (●) 2B8; (□) C2B8.

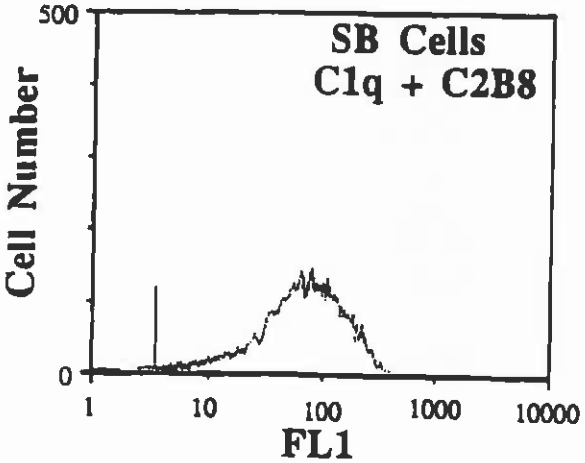



Fig 6. Binding of fluorescent-C1q to Chimeric MoAb C2B8 associated with antigen on the surface of human B-lymphoid cells detected by flow cytometry. Purified-C2B8 and murine-2B8 antibodies were incubated with human SB cells, washed to remove unattached antibody and incubated with fluoresceinated C1q. Cells were washed again, fixed and C1q fluorescence measured by flow cytometry. Controls were C1q-FITC and C1q-FITC plus irrelevant human IgG1,κ.

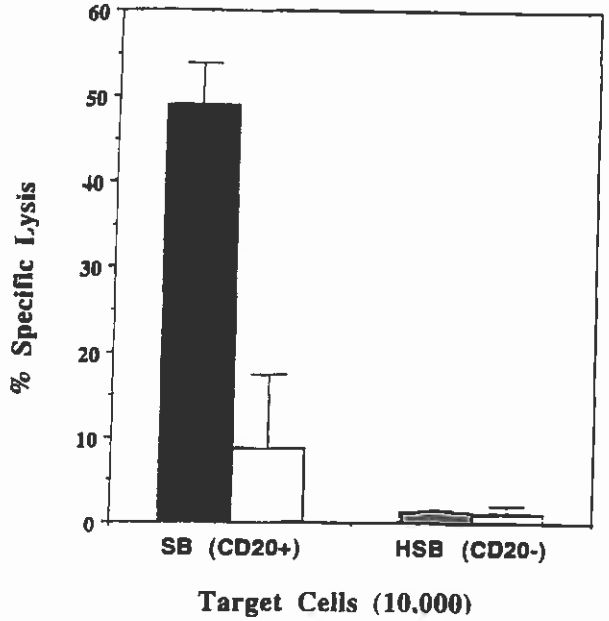



Fig 7. Chimeric C2B8 (■) was compared with murine 2B8 (□) for complement-related lysis in a concentration of 2.2 μg/mL. Ten thousand SB cells ($CD20^+$) were labeled with $Cr^{51}$ followed by incubation in the presence of anti-CD20 antibodies and human complement was used at a 1:4 dilution (Pel Freeze). $CD20^-$ control cell line was 10,000 HSB cells.

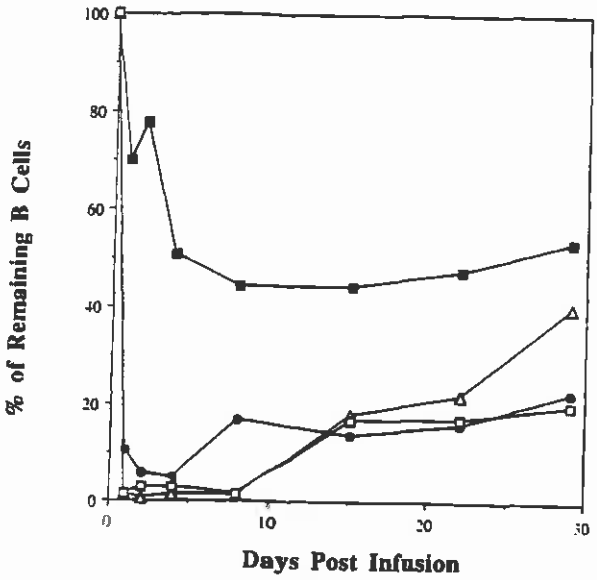



Fig 9. Depletion of PB B lymphocytes in cynomolgus primates infused with four daily injections of C2B8 antibody purified from CHO transfectoma. Fluorescein-labeled MoAb with specificity for human B lymphocytes (Leu-16, anti-CD20; Becton Dickinson) was used to stain and detect the B-lymphocyte population. Dose: (■) 0.04 mg/kg; (●) 0.4 mg/kg; (△) 1.6 mg/kg; (□) 6.4 mg/kg.

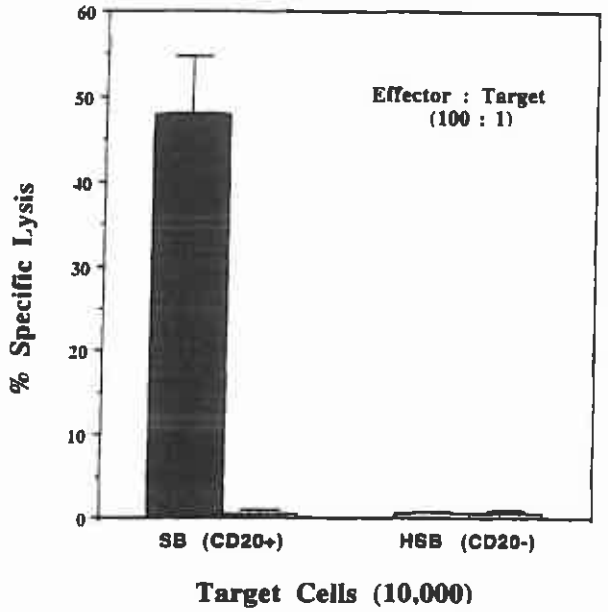



Fig 8. ADCC with in vitro human effector cells comparing chimeric-C2B8 and murine-2B8 lysis of human lymphoid target cells. C2B8 (■) and 2B8 (□) concentration with target cells was 3.9 μg/mL. Human peripheral lymphocytes were used as effector cells in a 100:1 ratio to target cells previously loaded with $Cr^{51}$. $CD20^-$ control cells were HSB at the same effector to target ratio.

a slow recovery. This occurred more rapidly in some monkeys than others and was seemingly dose independent.

*In vivo depletion of B lymphocytes in peripheral lymph nodes.* The objective of this study was to determine the effects of the two highest C2B8 doses that previously depleted greater than 98% of the B cells in the PB on the lymph node B-cell populations. The results summarized in Table 2 are obtained from lymph node biopsies of four animals infused with 1.6 mg/kg or 6.4 mg/kg of C2B8 and a single control animal injected with saline. Examination of lymphatic tissue removed on days 15 and 29 after administration of 1.6 and 6.4 mg/kg shows effective depletion of B lymphocytes (Table 2). B-cell depletion varied widely in individuals from both dose groups ranging from 34% to 78%, based on calculations using 39.5% B cells present in the inguinal node of a saline-treated animal. In the most highly depleted individuals, a major increase in the percentage of T cells was observed, confirming the depletion of B cells. In each of the antibody doses used (Table 2), complete depletion of the B cells in the lymphatic tissue with C2B8 still was not achieved. Furthermore, day 15 and day 29 B-cell populations were quite similar in each animal indicating most of the depletion occurred before day 15, with very little regeneration within the first 30 days.

Using the data derived from these studies, a high-dose pharmacology/toxicology study of C2B8 in cynomolgus monkeys was conducted. This protocol was designed to evaluate the toxicity (if any) associated with the administration of C2B8, as well as the efficacy of B-cell depletion from lymph nodes and BM. Also, because the majority of lymph node B cells were depleted within 15 days after antibody

**Table 1. Coating of CD20 Antigen Positive B Lymphocytes in PB of Cynomolgus Monkeys Treated With Various Doses of C2B8 Antibody**

| Monkey | Dose (mg/kg) | Day | Percent of Total Lymphocytes: Leu-16+ B Cells* | C2B8+ Cells† | CD2+ T Cells‡ |
|---|---|---|---|---|---|
| 672 | 0.04 | Pre | 16.7 | 0 | 74.7 |
| | | 1 | 11.7 | 11 | 74.7 |
| | | 2 | 13.0 | 6.3 | 80.6 |
| | | 4 | 8.5 | 2.6 | 80.6 |
| | | 8 | 7.4 | 0.2 | 82.7 |
| 848 | 0.4 | Pre | 12.4 | 0 | 80.1 |
| | | 1 | 0.1 | 1.3 | 85.7 |
| | | 2 | 0.1 | 0.7 | 87.8 |
| | | 4 | 0.1 | 0.6 | 90.7 |
| | | 8 | 1.6 | 0.5 | 88.5 |
| 660 | 1.6 | Pre | 28.3 | 0 | 72.9 |
| | | 1 | 0.4 | 0.4 | 94.0 |
| | | 2 | 0.1 | 0.2 | 95.6 |
| | | 4 | 0.1 | 0.4 | 96.2 |
| | | 8 | 0.0 | 0.3 | 96.0 |
| 427 | 6.4 | Pre | 7.1 | 0 | 84.7 |
| | | 1 | 0.0 | 0.1 | 92.1 |
| | | 2 | 0.0 | 0.2 | 92.0 |
| | | 4 | 0.0 | 0.2 | 94.1 |
| | | 8 | 0.0 | 0.1 | 91.3 |

* Determined as a percentage of the total gated lymphocyte population by flow cytometry using the CD20 marker antibody, Leu-16–FITC.

† Determined as a percentage of the total gated lymphocyte population by detecting the percentage of C2B8-coated cells using monkey IgG absorbed goat antihuman-phycoerythrin (RPE).

‡ Determined as a percentage of the total gated lymphocytes using the CD2 specific antibody T-cell marker (antihuman CD2-RPE, AMAC).

treatment, a weekly dosing regimen was adopted. In this study, four animals were given four weekly doses of approximately 16.8 mg/kg C2B8. At the completion of the dosing schedule, lymph node and BM specimens were removed and analyzed by flow cytometry for the presence of $CD20^+$ B cells. Two animals were examined at 22 days after the last dose and the other two examined at 36 days. Figure 10 illustrates the results of a typical flow cytometry experiment with one of the monkeys used in the study. Two normal animals treated with saline were examined for comparible lymph node and BM tissues. Figure 11 histograms compare T-cell populations in normal animals and depleted animals from both lymph node and BM tissue specimens. Table 3 summarizes the results of all the animals in the study. It can be seen that lymph nodes from both animals evaluated at 22 days after cessation of treatment contained from 5.3% to 6.3% B cells as compared with 39.5% in control lymph nodes (Table 3). Similarly, in the BM of these animals, the levels of $CD20^+$ cells ranged from 3% to 4.3%, as compared with 16.6% in the normal animal. In animals evaluated at 36 days after cessation of treatment, studies showed that one of two animals had approximately 12% B cells in the lymph node and 4.4% B cells in BM. The other animal had approximately 5% and 0.8% B cells in the lymph node and BM, respectively. T-cell populations generally increased dramatically with the absence of B cells in lymph nodes, ranging from 85% to 90%. T-cell populations in the BM of depleted animals ranged from slightly below normal at 25.6% to slightly over 46%. The normal control animal 908 BM sample had 29.8% T cells.

All animals used in the in vivo studies were routinely examined for symptoms of toxicity during the course of the studies by routine blood chemistry, urinalysis, and physical examination. Two of the four animals in the multiple high-dose study were killed at the end of the study at 22 days after final dosing. Body tissues were thoroughly examined for evidences of toxicity and tissue damage. No weight loss or other visible effects of antibody treatment other than loss of B cells were noted.

## DISCUSSION

MoAbs have been used for therapy of hematologic malignancies. Radiolabeled, toxin-conjugated, and unmodified MoAbs have all been tested clinically.[8,23-25] Some of the problems associated with MoAb therapy of neoplasms include specificity of the antibody for tumor cells as opposed to normal tissues, heterogeneity of tumor-specific antigen expression, low distribution of antibody in solid neoplasms, human antimouse antibody responses, which often limit the usefulness of murine antibodies for repeated therapy, and toxicity of radiolabeled or toxin-conjugated antibodies.[23]

We have expressed at high levels a chimeric mouse/human MoAb to CD20, a cell-surface phosphoprotein expressed on cells of the B-cell lineage. We have shown that the chimeric antibody, C2B8, has specificity and affinity equivalent to the mouse MoAb from which it was derived. The chimeric antibody has shown human effector functions using in vitro assays for C1q binding, complement-mediated cell lysis and antibody-dependent cellular-mediated cytotoxicity. Unlike its mouse counterpart (data not shown),

**Table 2. T- and B-Lymphocyte Populations in Lymph Node Biopsies Taken From C2B8-Treated Cynomolgus Monkeys, 15 and 29 Days After Treatment**

| Monkey | Dose (mg/kg) | Day | % CD2+ * | % CD20+ † | % Depletion‡ |
|---|---|---|---|---|---|
| 425 | Saline | ND | ND | ND | 0 |
| 684 | 1.6 | 15 | 76.9 | 22.6 | 42.9 |
| 688 | 1.6 | 15 | 83.2 | 14.5 | 63.3 |
| 692 | 6.4 | 15 | 74.1 | 23.9 | 39.5 |
| 696 | 6.4 | 15 | 74.1 | 8.7 | 77.9 |
| 425 | Saline | 29 | 52.1 | 39.5 | 0 |
| 684 | 1.6 | 29 | 61.6 | 26.0 | 34.2 |
| 688 | 1.6 | 29 | 84.1 | 14.6 | 63.0 |
| 692 | 6.4 | 29 | 66.9 | 21.4 | 45.8 |
| 696 | 6.4 | 29 | 84.1 | 12.9 | 67.3 |

Abbreviation: ND, not determined.

* Determined by flow cytometry with antihuman CD2-RPE (AMAC).

† Determined by flow cytometry with Leu-16 antihuman CD20-FITC (Becton Dickinson).

‡ Values calculated assuming 39.5% B cells from a normal animal treated with saline and biopsied at day 29 after last saline injection. Day 15 and day 29 treated animals all calculated based on 39.5% B cells in normal lymph node.

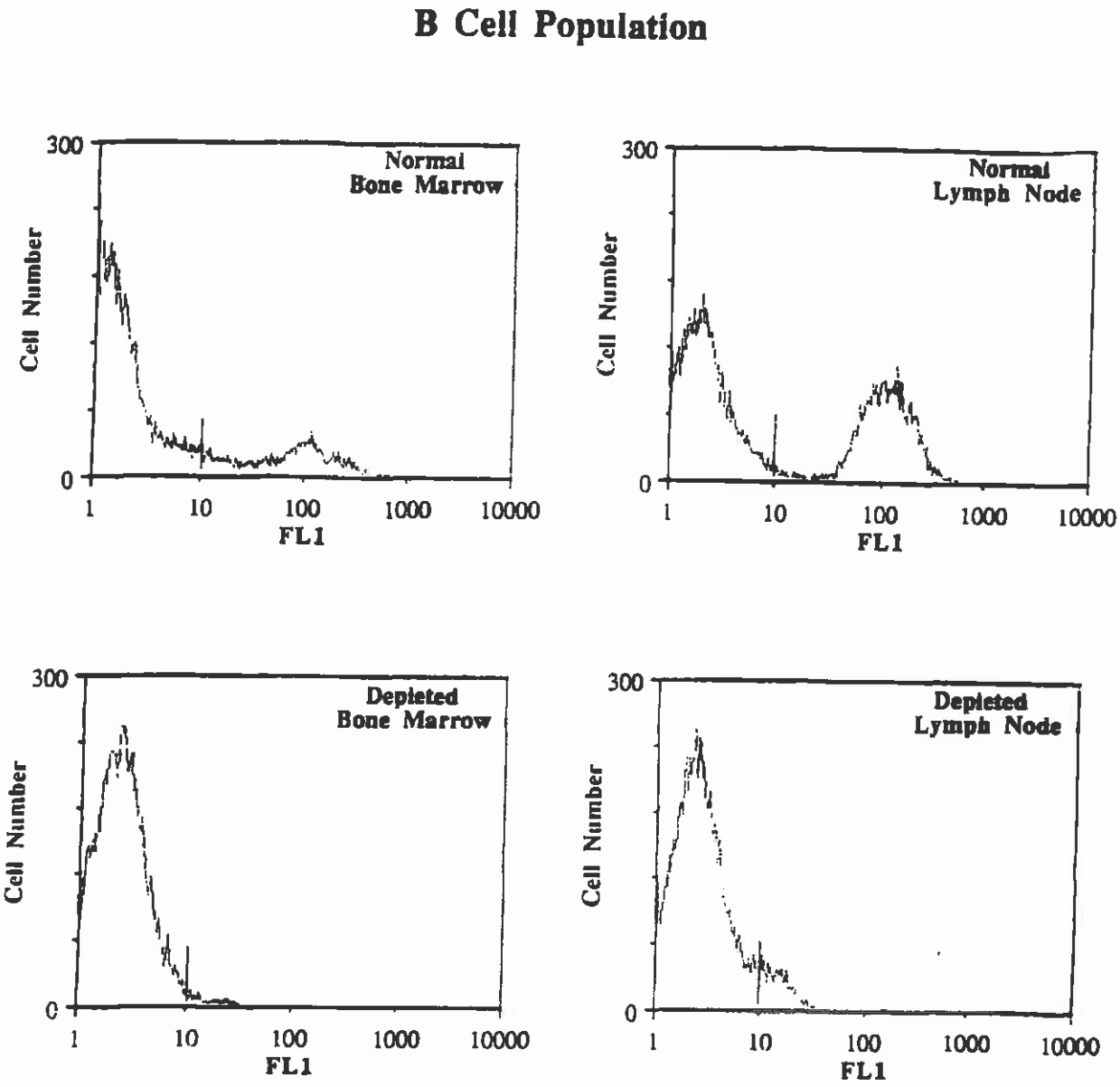



**Fig 10. Depletion of monkey lymph node and BM B lymphocytes after infusion of four weekly 16.8 mg/kg doses of C2B8. Chimeric C2B8 was injected into four monkeys: 677, 892, 681, and 884. Inguinal lymph nodes were surgically removed for flow cytometry cell analysis along with BM aspirates. Monkey-681 B-lymphocyte BM and lymph node histograms compared with nondepleted animals 425 (lymph node) and 908 (BM), injected with saline. C2B8-treated monkey 681 was examined at 38 days after last injection.**

a low dose of chimeric anti-CD20 leads to long-term peripheral B-cell depletion in primates. Furthermore, significant depletion of B-cell populations was achieved in peripheral lymph nodes and BM when repetitive high doses were administered.

We have chosen CHO transfectomas to produce antibody for future clinical trials because of the previous use in humans of other CHO-derived recombinant proteins, and because we are able to derive CHO cells that secrete 80 to 100 $\mu g/10^6$ cells/d. These transfectomas are capable of producing greater than 1 g/L in short fermentation-tank studies.

C2B8 antibody has shown in vivo characteristics that support its use in treatment of B-cell malignancies. Tumors of this type are particularly accessible and readily susceptible to immunotherapy.[26] The CD20 target is strongly expressed on most B-lymphocyte tumors and is not shed into the circulation nor expressed in nonhematopoietic tissues, where it could interfere with antibody therapy. The concern for immunoreactivity against mouse/human chimeric antibodies is reduced because of the immunocompromised nature of lymphoma patients[27,28] and perhaps aided further by the human constant region.[29] We have not yet compared di-

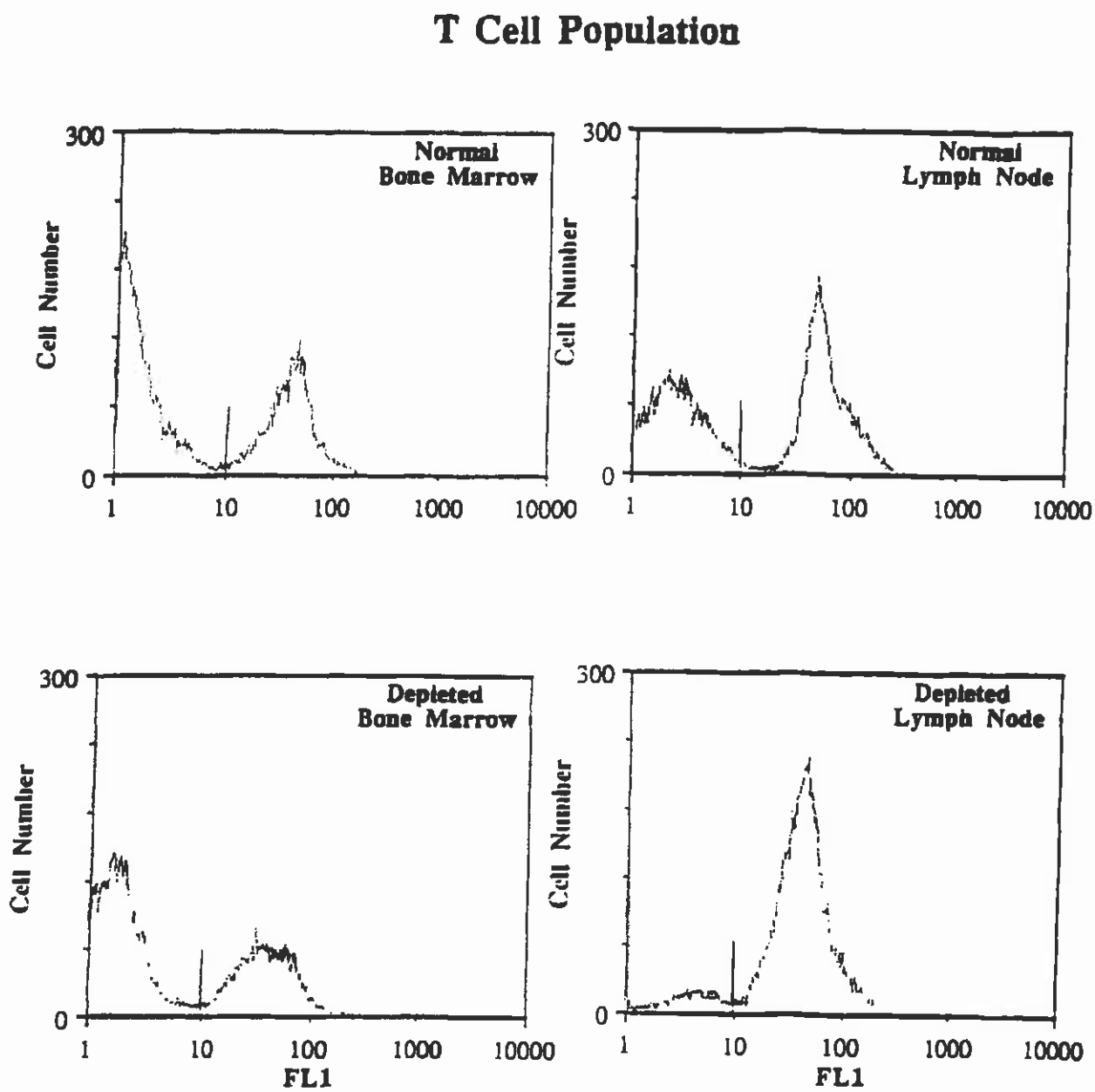



Fig 11. T-cell populations in monkey lymph node and BM tissues after infusion of four weekly 16.8 mg/kg doses of C2B8. Chimeric C2B8 was injected into four monkeys: 677, 681, 684, and 892. Inguinal lymph nodes were surgically removed from anesthetized animals for immunohistochemical staining and flow cytometry cell analysis along with BM aspirates. Monkey-681 T-lymphocyte BM and lymph node histograms compared with nondepleted animals 425 (lymph node) and 908 (BM), injected with saline. C2B8-treated monkey 681 was examined at 36 days after last injection.

rectly in monkeys the immunogenicity of the murine and chimeric antibodies. We have analyzed all antibodies produced in these animals against C2B8 and found them to be exclusively anti-idiotypic and antijoining region specific. None of these antibodies cross-react to other human or murine antibodies; however, immunogenicity still remains an issue for further investigation.

Despite the fact that all animals were dramatically depleted of peripheral B lymphocytes during the first week after treatment, there have been no adverse effects observed in any of the treated animals. In our studies, all antibody-treated monkeys have shown no apparent depletion of other cell types; however, the lack of suitable marker reagents for monkey hematopoietic tissues other than B- or T-lymphocytes, have made it difficult to establish this claim with certainty. Lymph nodes in B-cell–depleted animals become al-

Table 3. T- and B-Lymphocyte Populations in Lymph Node and BM Biopsies Taken From Cynomolgus Monkeys Treated with Four Weekly Doses (16.8 mg/kg) of C2B8

| Monkey | % CD2+ * | % CD20+ † | Day‡ | Dose (mg/kg) | % Depletion§ |
|---|---|---|---|---|---|
| Inguinal lymph node | | | | | |
| 425 | 52.1 | 39.5 | 29 | Saline | 0 |
| 677 | 90 | 5.3 | 22 | 16.8 | 86.6 |
| 892 | 91 | 6.3 | 22 | 16.8 | 84.1 |
| 681 | 89.9 | 5.0 | 36 | 16.8 | 87.4 |
| 884 | 86.4 | 12.3 | 36 | 16.8 | 68.9 |
| Bone marrow | | | | | |
| 677 | 46.7 | 4.3 | 22 | 16.8 | 74.1 |
| 892 | 41.8 | 3.0 | 22 | 16.8 | 81.9 |
| 681 | 35.3 | 0.8 | 36 | 16.8 | 95.2 |
| 884 | 25.6 | 4.4 | 36 | 16.8 | 73.5 |
| 908 | 29.8 | 16.6 | 36 | Saline | 0 |

* Determined by flow cytometry with antihuman CD2-RPE (AMAC). Values are percentages of total lymphocytes.

† Determined by flow cytometry with Leu-16 antihuman CD20-FITC (Becton Dickinson). Values are percentages of total lymphocytes.

‡ Days after final weekly injection.

§ Values are calculated assuming 39.5% of total lymphocytes in animal 425 treated with saline as B-cell fraction in a normal lymph node. Percent depletion represents the decreased fraction of B cells compared with the normal lymph node, found in the treated animals. The percent depletion of BM calculated in treated animals was based on fraction of B cells in normal BM sample from saline animal 908 as 16.6% B cells.

most entirely populated with T cells. All nonsacrificed animals in the studies eventually recovered their B-cell populations, indicating that monkey pluripotent stem cells are not affected. This is further supported by our in vitro studies on human BM stem cells (data not shown), which demonstrated that the CD20 antigen is not expressed on stem cells. In addition, other studies in humans that have used anti-CD20 antibodies to purge BM have seen no adverse effects on BM reconstitution.[30]

The application of C2B8 to human lymphoma therapy will provide a much greater challenge than the primate model described here. Human disease will likely present excessive tumor burdens within lymphatic tissues involving bulky fibroidal lesions with markedly decreased accessibility.[23] Our present plans are to determine how effective C2B8 is in reducing tumor bulk or in erradicating micrometastases in patients suffering from non-Hodgkin's B-cell lymphoma.

## ACKNOWLEDGMENT

The authors thank Mabel Hernandez and Nora Morris for preparing the manuscript, and Robert McCoobery and James Alberts for excellent technical assistance. We are grateful to Ken Woolcott and Dr William Rastetter for providing useful critical comments to the manuscript.

## REFERENCES

1. Calvert JE, Maruyama S, Tedder TF, Webb CF, Cooper MD: Cellular events in the differentiation of antibody secreting cells. Semin Hematol 21:226, 1984

2. Valentine MA, Meier KE, Rossie S, Clark EA: Phosphorylatin of the CD20 phosphoprotein in resting B lymphocytes. J Biol Chem 264:11282, 1989

3. Nadler LM, Korsmeyer SJ, Anderson KC, Boyd AW, Slaughenhoupt B, Park E, Jansen J, Coral F, Mayer RJ, Dalla SE, Ritz J, Schlossman SF: B cell origin of non-T cell acute lymphoblastic leukemia. J Clin Invest 74:332, 1984

4. Tedder TF, Boyd AW, Freedman AS, Nadler LM, Schlossman SF: The B cell surface molecule B1 is functionally linked with B cell activation and differentiation. J Immunol 135:973, 1985

5. Tedder TF, Forsgren A, Boyd AW, Nadler LM, Schlossman SF: Antibodies reactive with the B1 molecule inhibit cell cycle progression but not activation of human B lymphocytes. Eur J Immunol 16:881, 1986

6. Golay JT, Clark EA, Beverley PCL: The CD20 (Bp35) antigen is involved in activation of B cells from the $G_0$ to the $G_1$ phase of the cell cycle. J Immunol 135:3795, 1985

7. Clark EA, Shu G, Ledbetter JA: Role of the Bp35 cell surface polypeptide in human B-cell activation. Proc Natl Acad Sci USA 82:1766, 1985

8. Press OW, Appelbaum F, Ledbetter JA, Martin PJ, Zarling J, Kidd P, Thomas ED: Monoclonal antibody 1F5 (anti-CD20) serotherapy of human B-cell lymphomas. Blood 69:584, 1987

9. Einfeld DA, Brown JP, Valentine MA, Clark EA, Ledbetter JA: Molecular cloning of the human B cell CD20 receptor predicts a hydrophobic protein with multiple transmembrane domains. EMBO J 7:711, 1988

10. Liu AY, Robinson R Jr, Murray ED, Ledbetter JA, Hellström I, Hellström KE: Production of a mouse-human chimeric monoclonal antibody to CD20 with potent Fc-dependent biologic activity. J Immunol 139:3521, 1987

11. Adams RA, Flowers A, Davis BJ: Direct implantation and serial transplantation of human acute lymphoblastic leukemia in hamsters, SB-2. Cancer Res 28:1121, 1968

12. Shulman M, Wilde CD, Kohler G: A better cell line for making hybridomas secreting specific antibodies. Nature 276:269, 1978

13. Markwell MAK: A new solid-state reagent to iodinate proteins I. conditions for the efficient labeling of anti-serum. Anal Biochem 125:427, 1982

14. Chomczynki P, Nicoletta S: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156, 1987

15. Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C: Sequences of Proteins of Immunological Interest (ed 5). NIH Publication No. 91-3242. Bethesda, MD, National Institutes of Health, 1991

16. Urlaub G, Mitchell PJ, Kas E, Chasin LA, Funanage VL, Myoda T, Hamlin J: Effect of gamma rays at the dihydrofolate reductast locus: Deletions and inversions. Somat Cell Mol Genet 16: 555, 1986

17. Adams RA: Formal discussion: The role of transplantation in the experimental investigation of human leukemia and lymphoma. Cancer Res 27:2479, 1967

18. Brunner KT, Mauel J, Cerottini JC, Chapuis B: Quantitative assay of the lytic action of immune lymphoid cells on $^{51}$Cr-labelled allogeneic target cells *in vitro:* inhibition by isoantibody and by drugs. Immunology 14:181, 1968

19. Ling NR, Maclennan ICM, Mason DY: B-cell and plasma cell antigens: New and previously defined clusters, in McMichael AJ (ed): Leucocyte Typing III: White Cell Differentiations Antigens. Oxford, UK, Oxford University, 1987, p 302

20. Colbere-Garapin F, Horodniceanu F, Kourilsky P, Garaphin AC: A new dominant hybrid selective marker for higher eukaryotic cells. J Mol Biol 150:1, 1981

21. Subramani S, Mulligan R, Berg P: Expression of the mouse dihydrofolate reductase complementary deoxyribonucleic acid in simain virus 40 vectors. Mol Cell Biol 9:854, 1981

22. Kaufman R, Sharp PA: Amplification and expression of sequences cotransfected with a modular dihydrofolate reductase complementary DNA gene. J Mol Biol 159:601, 1982

23. Grossbard ML, Press OW, Appelbaum FR, Bernstein ID, Nadler LM: Monoclonal antibody-based therapies of leukemia and lymphoma. Blood 80:863, 1992

24. Grossbard ML, Freedman AS, Ritz J, Coral F, Goldmacher VS, Eliseo L, Spector N, Dear K, Lambert JM, Blattler WA, Taylor JA, Nadler LM: Serotherapy of B-cell neoplasms with anti-B4-blocked ricin: A phase I trial of daily bolus infusion. Blood 79:576, 1992

25. Press OW, Eary JF, Badger CC, Appelbaum F, Levy R, Miller R, Brown S, Nelp WB, Krohn KA, Fisher D, DeSantes K, Porter B, Kidd P, Thomas ED, Bernstein ID: Treatment of refractory non-Hodgkin's lymphoma with radiolabeled MB-1 (anti-CD37) antibody. J Clin Oncol 7:1027, 1989

26. Nadler LM, Stashenko P, Hardy R, Kaplan WD, Button LN, Kufe DW, Antman KH, Schlossman SF: Serotherapy of a patient with a monoclonal antibody directed against a human lymphoma-associated antigen. Cancer Res 40:3147, 1980

27. Meeker TC, Lowder J, Maloney DG, Miller RA, Thielemans K, Warnke R, Levy R: A clinical trial of anti-idiotype therapy for B-cell malignancy. Blood 65:1349, 1985

28. Brown SL, Miller RA, Horning SJ, Czerwinski D, Hart SM, McElderry R, Basham T, Warnke RA, Merigan TC, Levy R: Treatment of B-cell lymphomas with anti-idiotype antibodies alone and in combination with alpha interferon. Blood 73:651, 1989

29. Hale G, Clark MR, Marcus R, Winter G, Dyer MJS, Phillips JM, Riechmann W: Remission induction in non-Hodgkin lymphoma with reshaped human monoclonal antibody campath-1H. Lancet 332:1394, 1988

30. Anderson KC, Ritz J, Takvorian T, Coral F, Daley H, Gorgone BC, Freedman AS, Canellos GP, Schlossman SF, Nadler LM: Hematologic engraftment and immune reconstitution post-transplantation with anti-B1 purged autologous bone marrow. Blood 69:597, 1987

*Proc. Natl. Acad. Sci. USA*
Vol. 81, pp. 6851–6855, November 1984
Immunology

# Chimeric human antibody molecules: Mouse antigen-binding domains with human constant region domains

(transfection/protoplast fusion/calcium phosphate transfection/intronic controlling elements/transfectoma)

SHERIE L. MORRISON*, M. JACQUELINE JOHNSON†, LEONARD A. HERZENBERG†, AND VERNON T. OI‡

*Department of Microbiology and the Cancer Center, Institute for Cancer Research, College of Physicians and Surgeons, Columbia University, New York, NY 10032; †Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305; and ‡Becton–Dickinson Monoclonal Center, 2375 Garcia Avenue, Mountain View, CA 94043

*Contributed by Leonard A. Herzenberg, August 1, 1984*

**ABSTRACT** We have created mouse–human antibody molecules of defined antigen-binding specificity by taking the variable region genes of a mouse antibody-producing myeloma cell line with known antigen-binding specificity and joining them to human immunoglobulin constant region genes using recombinant DNA techniques. Chimeric genes were constructed that utilized the rearranged and expressed antigen-binding variable region exons from the myeloma cell line S107, which produces an IgA ($\kappa$) anti-phosphocholine antibody. The heavy chain variable region exon was joined to human IgG1 or IgG2 heavy chain constant region genes, and the light chain variable region exon from the same myeloma was joined to the human $\kappa$ light chain gene. These genes were transfected into mouse myeloma cell lines, generating transformed cells that produce chimeric mouse–human IgG ($\kappa$) or IgG ($\kappa$) anti-phosphocholine antibodies. The transformed cell lines remained tumorigenic in mice and the chimeric molecules were present in the ascitic fluids and sera of tumor-bearing mice.

The capability to transfer immunoglobulin genes into lymphoid cells where they produce protein in quantities sufficient for structural studies (1–3) provides us with the opportunity to generate and characterize novel immunoglobulin molecules. Cloned variable (*V*) region genes from mouse or rat hybridoma cell lines can be ligated to human constant (*C*) region genes and we would expect that these chimeric genes can be transfected into mouse myeloma cells, which then will produce novel human antibody molecules. We would thus produce antibodies that are largely human but which have antigen-binding specificities generated in mice. The additional potential for *in vitro* manipulation and alteration of both the antigen-binding site and the structures correlated with biological effector functions of these antibody molecules using recombinant DNA techniques would introduce a powerful approach for further understanding antibody structure, function, and immunogenetics.

As we show here, both chimeric mouse heavy chain *V* region exon ($V_H$)–human heavy chain *C* region genes and chimeric mouse light chain *V* region exon ($V_\kappa$)–human $\kappa$ light chain gene constructs are expressed when transfected into mouse myeloma cell lines. When both chimeric heavy and light chain genes are transfected into the same myeloma cell, an intact tetrameric ($H_2L_2$) chimeric antibody is produced. In this study we used $V_H$ and $V_\kappa$ exons from the mouse phosphocholine (*P*Cho)-binding antibody-producing S107 myeloma cell line (4, 5). Chimeric mouse–human anti-*P*Cho antibodies were produced in culture by appropriate transfected cell lines or by "transfectomas" obtained when such cell lines are injected into mice.

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.

## MATERIALS AND METHODS

**Chimeric Genes.** The cloned S107 $V_H$ and S107 $V_\kappa$ genes were gifts from Matthew Scharff (Albert Einstein College of Medicine, Bronx, NY). The S107 $V_H$ gene was spliced to human IgG1 and IgG2 *C* region genes by using *Sal* I linkers as shown in Fig. 1*A*. Both constructs were inserted into the vector pSV2ΔH-gpt (1, 6). The S107 $V_\kappa$ gene was spliced to the human $\kappa$ gene at a unique *Hin*dIII site located in the large intron between the $\kappa$ light chain joining and *C* ($J_\kappa$ and $C_\kappa$) region exons as shown in Fig. 1*B*. This chimeric light chain gene construct was inserted into both PSV2ΔH-gpt and pSV2-neo plasmid vectors (7).

**Transfection.** Protoplast fusion and calcium phosphate precipitation techniques (1, 8, 9) were used to transfect these chimeric immunoglobulin genes into the J558L myeloma cell line (a λ light chain-producing mouse myeloma cell line) and an immunoglobulin nonproducing derivative of the P3 myeloma cell line. Mycophenolic acid (GIBCO) was used for selection of cells transfected with pSV2Δ-gpt vectors as described (1, 3). G418 (GIBCO) at 1.0 mg/ml was used for selection of cells transfected with pSV2-neo vectors (7).

Light and heavy chain chimeric immunoglobulin genes were transfected sequentially by protoplast fusion using G418 selection for the chimeric light chain gene vector and mycophenolic acid for the chimeric heavy chain gene vector. The protoplast fusion transfection procedure used was as described (1).

Transfection using the calcium phosphate precipitation procedure was done by transfecting a mixture of 40 $\mu$g each of the chimeric light and chimeric heavy chain pSV2ΔH-gpt vectors into $5 \times 10^6$ cells. Mycophenolic acid was used to select for transformed cell lines as described (1).

**Antigen Binding.** *P*Cho binding of antibody secreted into the culture supernates of transfected cell lines was analyzed by using a solid-phase radioimmunoassay described previously (10). *P*Cho-keyhole limpet hemocyanin antigen was bound to 96-well polyvinyl plates; binding of chimeric anti-*P*Cho antibodies was detected by using $^{125}$I-labeled protein A or $^{125}$I-labeled-anti-human IgG. *P*Cho-binding antibodies in biosynthetically labeled culture supernates and cell lysates of transfected cell lines also were analyzed by binding the biosynthetically labeled antibody to *P*Cho-coupled Sepharose 4B (Pharmacia) and then eluting the bound antibody with *P*Cho-hapten. The bound and eluted antibody was examined by $NaDodSO_4$/polyacrylamide gel electrophoresis ($NaDodSO_4$/PAGE). Biosynthetic-labeling procedures were as described (11).

**Idiotope Analysis.** Three hybridoma anti-idiotope antibodies, also kindly provided by Matthew Scharff and Angela

Abbreviations: V, variable; C, constant; J, joining; $_H$, heavy chain; $_L$, light chain; *P*Cho, phosphocholine; kb, kilobase(s); NEPHGE, nonequilibrium pH gradient electrophoresis.

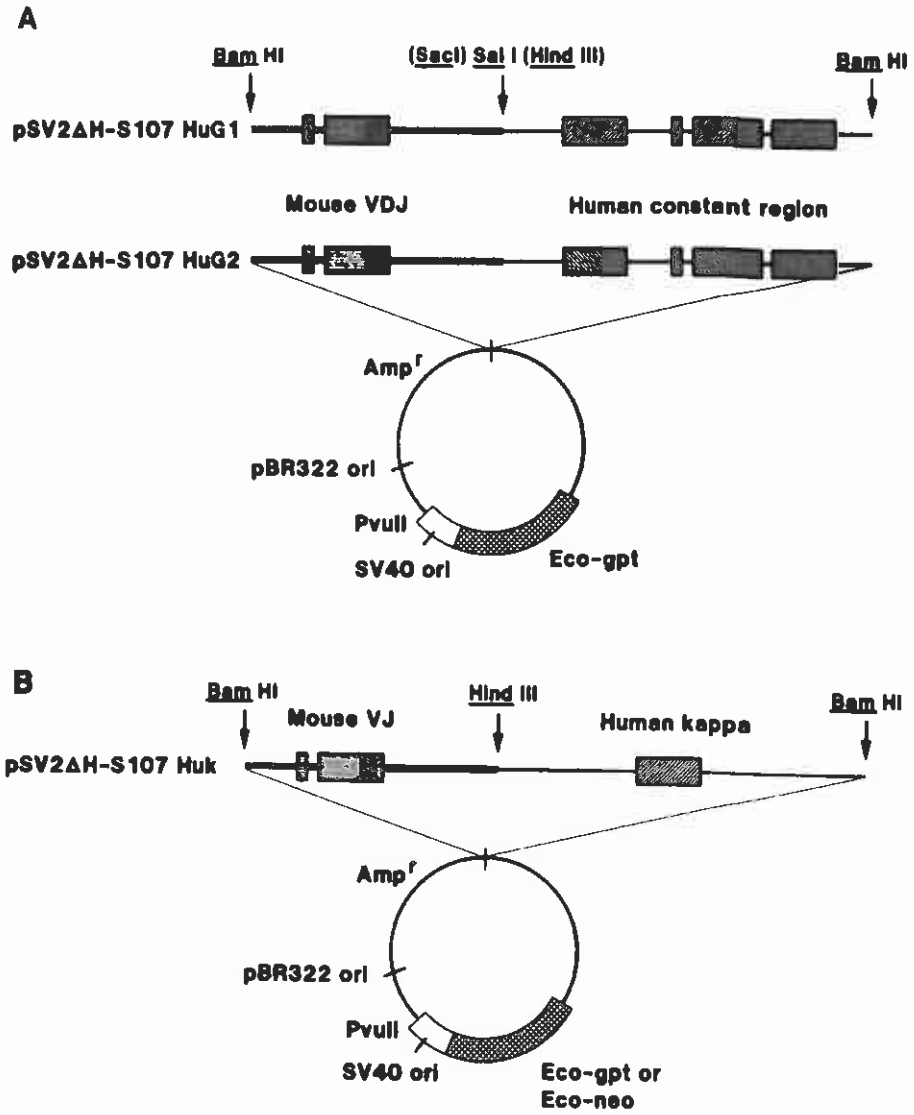



FIG. 1. Schematic diagrams (not drawn to scale) of the chimeric mouse–human heavy chain gene vector (*A*) and the chimeric light chain gene vectors (*B*). DNA fragment sizes are as follows: *Hin*dIII–*Bam*HI human IgG1 (HuG1) or IgG2 (HuG2) heavy chain gene, 7 kilobases (kb); mouse S107 $V_H$ gene, 4.5 kb; *Hin*dIII–*Bam*HI human $C_\kappa$ gene (HuK), 10 kb; and mouse S107 $V_\kappa$ gene, 3.5 kb. The *Hin*dIII sites in the human IgG1 and IgG2 heavy chain genes were ligated to the *Sac* I site of the mouse S107 $V_H$ gene with *Sal* I linkers.

Guisti (Albert Einstein College of Medicine), were used to analyze the $V_H$–$V_L$ (L is light chain) domain structure of the chimeric human anti-*P*Cho antibodies. These antibodies (TC102.1.2, T139.2, and T156.1.1), recognizing three independent idiotopes (12), were used to immunoprecipitate biosynthetically labeled material eluted with *P*Cho from the *P*Cho-Sepharose 4B matrix. Immunoprecipitates were analyzed by $NaDodSO_4$/PAGE.

**Immunoglobulin Chain Composition.** Monoclonal anti-human IgG and anti-human κ antibodies (Becton–Dickinson) were used to immunoprecipitate biosynthetically labeled chimeric human anti-*P*Cho antibodies for analyses using two-dimensional nonequilibrium pH gradient PAGE (NEPHGE) (11, 13). *P*Cho-coupled Sepharose 4B also was used for affinity purification.

**Immunoglobulin Heavy Chain Glycosylation.** Tunicamycin (Calbiochem–Behring) was used to inhibit asparagine-linked glycosylation of biosynthetically labeled antibody from mouse cell lines producing mouse–human chimeric immunoglobulins (11). *P*Cho-binding antibody from tunicamycin-treated cells was analyzed by $NaDodSO_4$/PAGE. Procedures for tunicamycin treatment were as described (11).

**Chimeric Mouse–Human Antibody Production in Mice.** Transfected J558L cells producing chimeric mouse–human antibody were injected intraperitoneally into BALB/c mice ($10^6$ cells per mouse). Sera and ascitic fluids from tumor-bearing mice were analyzed for human anti-*P*Cho antibody by a solid-phase radioimmunoassay (10) and by immunoelectrophoresis using a polyclonal anti-human IgG antiserum.

## RESULTS

We obtained expression of chimeric mouse *V* region–human *C* region genes in transfected J558L and the immunoglobulin nonproducing P3 myeloma cell lines. When both light chain and heavy chain chimeric genes (see Fig. 1) were expressed in the same cell, tetrameric ($H_2L_2$) antigen-binding antibodies were obtained. Biosynthetically labeled antibody molecules secreted by J558L cells expressing chimeric genes were bound and hapten-eluted from *P*Cho-Sepharose 4B. Autoradiograms of two-dimensional NEPHGE analyses (Fig. 2) show immunoglobulin polypeptide chains of the ex-

193

pected charge and relative molecular mass. Identical two-dimensional gel analysis results were obtained with immunoprecipitates by using monoclonal anti-human κ and IgG antibodies. Similar results were obtained when the chimeric IgG2 (κ) antibodies produced in the transfected nonproducing P3 cell line were analyzed. Since the recipient P3 cell line does not produce endogenous immunoglobulin polypeptide chains, only the chimeric mouse–human heavy and light chains were seen (data not shown). *P*Cho binding by the chimeric antibody produced in the J558L cell line required the specific association of both the $V_H$ and $V_L$ domains of the S107 myeloma protein connected to human *C* region polypeptides. Antibody secreted by transfected J558L cells expressing only the chimeric heavy chain and the endogenous J558L light chain did not bind *P*Cho (data not shown).

Further verification of the appropriate polypeptide folding of the mouse $V_H$ and $V_L$ domains attached to human *C* region polypeptide chains was done by analyses detecting the presence of idiotopes known to occur on the parental S107 *P*Cho-binding antibody molecule. Three monoclonal anti-idiotope antibodies—two recognizing idiotopes requiring the association of light and heavy *V* region domains (TC102.1.2 and T139.2) and the third (T156.1.1) recognizing an epitope present on the heavy chain *V* region domain (12)—were found to react with the mouse–human chimeric anti-*P*Cho antibodies (see Fig. 3). This is good evidence that the mouse S107 antigen-binding domains have folded into their intended structures.

The conformation of IgG heavy chain *C* regions depends on the presence of the asparagine-linked carbohydrate moeity in the $CH_2$ domain of the molecule (4). The loss of this carbohydrate chain affects profoundly the overall domain structure of this part of the immunoglobulin molecule. Concomitant with this structural change, the catabolism rate of the molecule is increased and biological effector functions such as complement-fixation are lost (14). Glycosylation of the mouse–human chimeric antibodies in mouse myeloma cells was deduced by determining the molecular masses of antibodies synthesized in the presence and absence of tunicamycin, an antibiotic inhibitor of asparagine-linked glycosylation. Fig. 4 is an autoradiogram of $NaDodSO_4$/PAGE analysis of glycosylated and nonglycosylated chimeric heavy and light chains produced in transfected mouse myeloma cells. Though we cannot be certain that in the absence of tunicamycin the appropriate asparagine residue in the $CH_2$ domain is glycosylated, clearly, the lower molecular mass of the heavy chain synthesized in the presence of tunicamycin is as expected if a single N-linked carbohydrate were absent

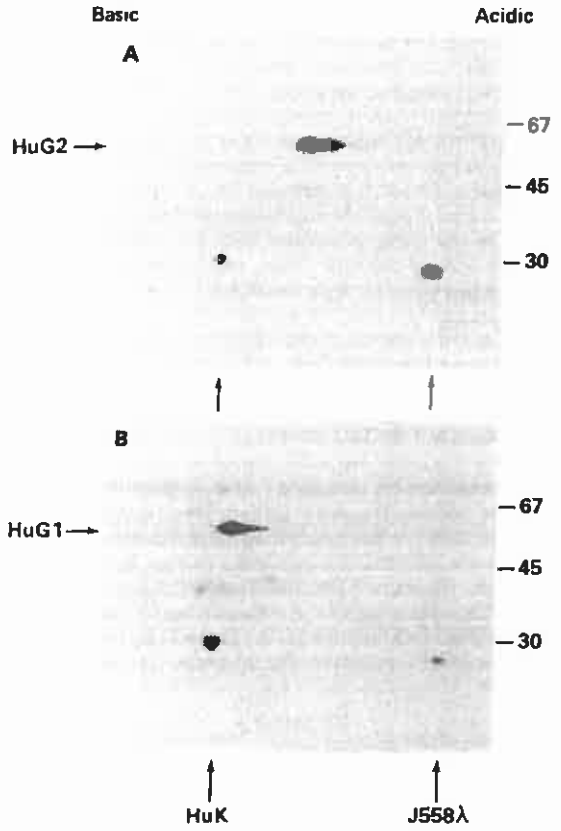



FIG. 2. Autoradiographs of two-dimensional NEPHGE gels of mouse–human chimeric IgG2 (κ) anti-PC antibody (*A*) and mouse–human chimeric IgG1 (κ) anti-PC antibody (*B*). Antibodies synthesized by transfected J558L cells were affinity purified with *P*Cho-Sepharose 4B. The electrophoretic mobility of the J558L λ chain had been determined previously (1). Size markers are shown in kDa.

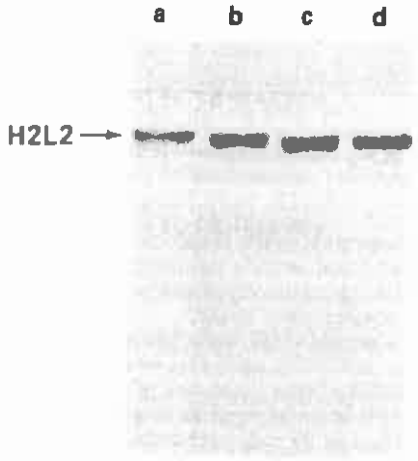



FIG. 3. Autoradiograph of nonreduced $NaDodSO_4$/PAGE analyses of *P*Cho-binding material from a chimeric IgG2 (κ) antibody-producing J558L cell line immunoprecipitated with *Staphylococcus aureus* protein A (lane a), monoclonal anti-human κ antibody (lane b), and anti-idiotope antibodies TC102.1.2 and T156.1.1 (lanes c and d).

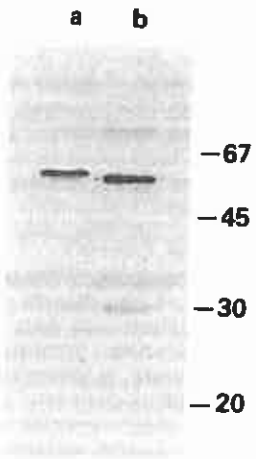



FIG. 4. Autoradiograph of reduced $NaDodSO_4$/PAGE analyses of chimeric IgG2 (κ) antibody synthesized in the presence (lane a) and absence (lane b) of tunicamycin. Size markers are shown in kDa.

FIG. 5. Immunoelectrophoresis of ascitic fluid from a BALB/c mouse bearing a J558L transfectoma producing chimeric IgG2 (κ) anti-*P*Cho antibodies developed with a polyclonal anti-human IgG antiserum.

from the polypeptide chain. Note that the light chain band is not affected by tunicamycin treatment. From these data, we conclude that the mouse myeloma cell appropriately glycosylates the human heavy chain.

When transfected J558L cells producing the human IgG2 (κ) chimeric anti-*P*Cho antibody were grown as an ascitic tumor in BALB/c mice, analyses of sera and ascitic fluids showed the presence of significant quantities of anti-*P*Cho-binding antibodies by radioimmunoassay. Fig. 5 shows the results of immunoelectrophoretic analysis of ascitic fluids from these transfectoma-bearing mice. Polyclonal anti-human antiserum (a gift of E. A. Kabat, Columbia University, New York) was used to demonstrate the presence of human immunoglobulin. Based on our previous experience with mouse hybridoma-antibody production in mice, we conclude that the amounts of immunoglobulin visualized by this immunoelectrophoretic analysis are similar to the levels seen with lower-producing mouse hybridomas *in vivo*.

Transfectomas producing chimeric heavy and light chains were produced either by cotransfection of both genes using calcium phosphate precipitation or by sequential protoplast fusion using vectors with different drug markers. For protoplast fusion, the heavy chains were introduced first on a plasmid expressing the *Eco-gpt* gene. One heavy chain-producing cell line then was transfected by protoplast fusion with a plasmid containing the light chain gene and the neo marker. Transformants resistant to both mycophenolic acid and G418 were assayed for their production of anti-*P*Cho antibodies. Of 77 transformants analyzed, only 7 were positive for antibody by radioimmunoassay and only 2 of these were found to be producing significant quantities of κ light chain. In cotransfection experiments using the calcium phosphate precipitation protocol, the same phenomenon was observed. Analyses of a large number of independent transformants revealed that a minority (<10%) of the transfected cell lines produced both chimeric heavy and light chain polypeptides. Since our own unpublished data using the calcium phosphate precipitation procedure demonstrate that two intact mouse immunoglobulin genes cotransfected into the same cell, using the same protocol as described here, coexpress both gene products in the majority of transformed cell lines, these results suggest that appropriate transcriptional or translational controlling elements are absent in the chimeric light chain gene construction or that required transcriptional or translational factors are absent in the mouse cell lines transfected here.

## DISCUSSION

The opportunity to use recombinant DNA techniques to construct novel antibody genes and then to produce antibody molecules in mammalian lymphoid cells transfected with these genes in appropriate vectors provides a new approach to understanding the structure, function, and immune properties of antibody molecules. The use of chimeric mouse–human antibody gene constructs permits us to study "near-human" antibodies with desired antigen-binding specificities. For clinical applications, human antibody molecules constructed by using recombinant DNA techniques should complement human–human hybridoma antibodies as useful immunotherapeutic and diagnostic reagents. The use of human or near-human antibodies might decrease or eliminate the immunogenicity of antibodies used *in vivo*, relative to monoclonal mouse hybridoma antibodies currently being used (15). Further, since the same *V* region can be joined to any *C* region, it is possible to use the *C* region of an appropriate human heavy chain isotype that will exhibit desired biological effector functions.

The use of mammalian lymphoid cells to produce these antibodies rather than a prokaryotic expression system assures that any post-translational modifications of the antibody molecules as they are synthesized, processed, folded, and assembled in the eukaryotic endoplasmic reticulum are carried out correctly. For example, the immunoglobulins encoded by the transfected mouse–human chimeric immunoglobulin genes in both mouse recipient cells were glycosylated, bound antigen, and presumably would be treated as normal immunoglobulins if injected into humans. This obviates the need for any post-synthetic *in vitro* modification or assembly that is required with prokaryotic synthesis of immunoglobulin polypeptides to produce functional antibody molecules (16). Furthermore, we have shown that transfectomas produce useable amounts of human antibodies either in tissue culture or in mice.

It is now feasible to shuffle exons or carry out other kinds of directed mutagenesis to explore the human antibody molecules *C* region structures required for carrying out diverse antibody functions. It is conceivable that an altered antibody molecule more effective in specific functions than naturally occurring antibody molecule can be created. In a similar manner, it should be possible to modify *V* regions to alter their interactions with antigen. Even more exciting would be to construct antigen-binding sites *de novo*.

A limitation yet to be overcome is the low frequency of transformants that produce chimeric light chains. The chimeric light chain gene construct (see Fig. 1*B*) has the mouse $V_\kappa$ gene coupled to the presumed human intronic controlling element (ICE) or "enhancer"—that is, this construct contains the human intron with a sequence homologous to the established mouse ICE sequence (17). In contrast, the heavy chain chimeric mouse–human polypeptide is expressed efficiently. In the latter construct (see Fig. 1*A*), the ICE sequence is mouse-derived. Thus, the lower expression frequency of the chimeric light chain gene may reflect a species preference for the effective interaction of the ICE sequence with factors or elements controlling immunoglobulin expression in mouse lymphoid cells. This suggestion is consistent with our earlier finding that the mouse light chain gene, presumably due to its ICE sequence, is not expressed efficiently in a rat myeloma cell line (1). Construction of chimeric genes using entire mouse intronic sequences and elimination of most human intronic sequences, to determine if these will be expressed efficiently, await further investigation.

In this work we have taken the initial steps toward constructing and efficiently expressing unique immunoglobulin molecules. We expect that this technique will permit an increased understanding of the structural and dynamic requirements of antibody functions. It also will facilitate development of antibody molecules for immunodiagnosis and immunotherapy.

We thank Letitia Wims and Linda Nakamura Roark, who provided invaluable assistance with these experiments. We also thank Olivia Gagliani for her patience in preparing this manuscript. This research was supported in part by Grants K04 AI00408, AI 19042, AI 08917, CA 16858, CA 22736, CA 13969, and CA 04681 from the National Institutes of Health, Grant IMS-360 from the American Cancer Society, and Grant SPO 13712-02-00 from the Becton–Dickinson Monoclonal Center.

1. Oi, V. T., Morrison, S. L., Herzenberg, L. A. & Berg, P. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 825–829.

195

2. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. (1983) *Cell* **33**, 717–728.
3. Ochi, A., Hawley, R. G., Hawley, T., Shulman, M. J., Traunecker, A., Kohler, G. & Hozumi, N. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6351–6355.
4. Kwan, S. P., Rudikoff, S., Seidman, J. G., Leder, P. & Scharff, M. D. (1981) *J. Exp. Med.* **153**, 1366–1370.
5. Clarke, C., Berenson, J., Governman, J., Boyer, P. D., Crews, S., Siu, G. & Calame, K. (1982) *Nucleic Acids Res.* **10**, 7731–7749.
6. Mulligan, R. C. & Berg, P. (1980) *Science* **209**, 1422–1427.
7. Mulligan, R. C. & Berg, P. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2072–2076.
8. Sandri-Goldin, R. M., Goldin, A. L., Levine, M. & Glorioso, J. C. (1981) *Mol. Cell. Biol.* **1**, 743–752.
9. Chu, G. & Sharp, P. A. (1981) *Gene* **13**, 197–202.
10. Oi, V. T. & Herzenberg, L. A. (1979) *Mol. Immunol.* **16**, 1005–1017.
11. Oi, V. T., Bryan, V. M., Herzenberg, L. A. & Herzenberg, L. A. (1980) *J. Exp. Med.* **151**, 1260–1274.
12. Desaymaid, C., Guist, A. M. & Sharff, M. D. (1984) *Mol. Immunol.*, in press.
13. O'Farrell, P. Z., Goodman, H. M. & O'Farrell, P. H. (1977) *Cell* **12**, 1133–1142.
14. Nose, M. & Wigzell, H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6632–6636.
15. Miller, R. A. & Levy, R. (1981) *Lancet* **I**, 226–229.
16. Boss, M. A., Kenton, J. H., Wood, C. R. & Emtage, J. S. (1984) *Nucleic Acids Res.* **12**, 3791–3806.
17. Morrison, S. L. & Oi, V. T. (1984) *Annu. Rev. Immunol.* **2**, 239–256.

**Humanized mouse therapeutic antibody**

As mentioned before, Wu and Kabat discovered that the antibody combining site is formed by six complementarity determining regions (CDR's), three each from the light and heavy chain variable regions. Therefore, there is no need to use the entire variable regions of the selected antibody from mouse hybridoma cell line. To bind the specific antigen, only these six CDR's will be required. Therefore an attempt was made to graft these mouse CDR's into the frame regions of a human antibody molecule (Jones PT, Dear PH, Foote J, Neuberger MS, Winter G (1986) Replacing the complementarity-determining regions in a human antibody with those from a mouse. Nature, 321, 522-525). They discovered that the humanized mouse antibody retained nearly the same binding affinity to the antigen, 4-hydroxy-3-nitrophenacetyl caproic acid.

The detailed construction of the variable regions of the light and heavy chain variable regions of the humanized mouse therapeutic antibody was somewhat complicated. Once the required amino acid sequences of the CDR's and FR's were selected, the corresponding nucleotides are designed using the most efficient codons. Short segments of the coding strand were synthesized, together with overlapping segments of the complementary strand. Then they were annealed together and ligated to form complete double stranded DNA molecules of the entire variable region. Synthesis of the entire variable region nucleotide sequence would be more cumbersome.

The frame work region of the select human antibody should have similar amino acid sequence to that of the original antibody from the mouse hybridoma cell line. This would reduce the distortion of joining the CDR's to the frame work portion of the variable region molecule. After the entire variable region gene was constructed, it was then inserted into a plasmid, as in the case of constructing chimeric therapeutic antibodies.

Again, many years passed until the first humanized mouse therapeutic antibody was approved by FDA. That antibody, daclizumab, was synthesized in 1989 (Queen C, Schneider WP, Selick HE, Payne PW, Landolfi NF, Duncan JF, Avdalovic NM, Levitt M, Junghans RP Waldmann TA (1989) A humanized antibody that binds the interleukin 2 receptor. Proc Natl Acad Sci USA, 86, 10029-10033), and approved for human use in 1997. For clinical use, each therapeutic antibody has to be subjected to strict tests for its toxicity, usefulness, specificity, etc.

For comparison, the half-lives of the above mentioned three types of therapeutic antibodies in human patients are roughly as follows:

| Therapeutic antibody | Type | Half-life |
|---|---|---|
| OKT3 | Mouse | 1 day |
| Rituximab | Chimeric | 2.5 – 7.5 days |
| Daclizimab | Humanized | 20 days |

differed between NMDA and KA (Fig. 3c, *d*), and in individual spinal cord neurones KA-evoked increases in $[Ca^{2+}]_i$ were always much smaller than those evoked by NMDA. These experiments suggest that $Na^+$ is a poor trigger for inducing an increase in $[Ca^{2+}]_i$, since in several neurones the inward ($Na^+$) current activated by KA produced no detectable arsenazo III signal. However, our results do not exclude the possibility that $Ca^{2+}$ influx through ion channels activated by NMDA triggers release of $Ca^{2+}$ from intracellular stores[27], contributing further to the NMDA-evoked arsenazo III signals reported here. Although the present results suggest a high $Ca^{2+}$ permeability of NMDA-receptor-activated channels (Fig. 3), the net flux of monovalent cations (that is, conductance) decreases in the presence of $Ca^{2+}$. This reflects interactions between permeant ions within the channel with $Ca^{2+}$ acting as both a permeant ion and as a blocker of monovalent cation flux[25,26,28].

The experiments reported here provide evidence for an agonist-triggered increase in $[Ca^{2+}]_i$ in mammalian spinal cord neurones. Previously, ion-sensitive microelectrodes were used to measure changes in intracellular ionic activity triggered by excitatory amino acids in frog motoneurones[9]. The latter experiments suggested an increase in both $[Na^+]_i$ and $[Ca^{2+}]_i$ during perfusion with L-glutamate but the results were difficult to interpret clearly as (1) neurones were not voltage-clamped and thus it is difficult to separate the relative contributions of $Ca^{2+}$ influx via voltage-dependent calcium channels and agonist-activated channels, and (2) L-glutamate is a mixed agonist that acts at multiple subtypes of excitatory amino-acid receptor[2,6,7].

The response to NMDA-receptor activation thus provides a second source of calcium flux, distinct from that resulting from conventional voltage-dependent calcium channels, which may have important long-term effects on excitability. Our finding that the ion channels linked to the NMDA receptor subtype are more permeable to $Ca^{2+}$ than those linked to KA receptors, has implications for the role of excitatory amino-acid receptors in CNS function. It is possible that $Ca^{2+}$ influx activated by NMDA receptors underlies the synaptic plasticity generating long-term potentiation, as the latter is prevented by intracellular injection of EGTA to chelate $Ca^{2+}$ (ref. 29), or by blocking NMDA receptors with selective antagonists[30]. For example, $Ca^{2+}$ influx localized at transmitter-operated ion channels could have a role in organizing and regulating postsynaptic structures in an appropriate spatial relation to transmitter-releasing presynaptic terminal boutons, and it is important to consider that $Ca^{2+}$ influx occurring at NMDA receptors located on dendritic spines might produce an especially large but localized elevation in intracellular $Ca^{2+}$ concentration, due to restriction of $Ca^{2+}$ diffusion along the narrow shaft of the spine. In addition, our results have some bearing on the mechanisms of desensitization of NMDA receptors, as the link that has been demonstrated between $[Ca^{2+}]_i$ and desensitization of nicotinic receptors at the neuromuscular junction[31,32] may occur also for other receptor-ionophore complexes. Thus our results may help to explain the similar desensitization evoked by either large doses of NMDA or depolarizing voltage jumps[7], which trigger $Ca^{2+}$ entry through NMDA channels and voltage-dependent calcium channels, respectively.

Received 3 January, accepted 1 April 1986.


1. Krogsgaard-Larsen, P., Honoré, T., Hansen, J. J., Curtis, D. R. & Lodge, D. *Nature* **284**, 64–66 (1980).
2. Watkins, J. C. & Evans, R. H. *A. Rev. Pharmac. Tox.* **21**, 165–205 (1981).
3. McLennan, H. *Prog. Neurobiol.* **20**, 251–271 (1983).
4. Nowak, L., Bregestovski, P., Ascher, P., Herbet, A. & Prochiantz, A. *Nature* **307**, 462–465 (1984).
5. Mayer, M. L., Westbrook, G. L. & Guthrie, P. B. *Nature* **309**, 261–263 (1984).
6. Mayer, M. L. & Westbrook, G. L. *J. Physiol., Lond.* **354**, 29–53 (1984).
7. Mayer, M. L. & Westbrook, G. L. *J. Physiol., Lond.* **361**, 65–90 (1985).
8. Dingledine, R. *J. Physiol., Lond.* **343**, 385–405 (1983).
9. Bührle, C. P. & Sonnhof, U. *Pflügers Arch. ges. Physiol.* **396**, 154–162 (1983).
10. Zanotto, L. & Heinemann, U. *Neurosci. Lett.* **35**, 79–84 (1983).
11. Pumain, R. & Heinemann, U. *J. Neurophysiol.* **53**, 1–16 (1985).
12. Lansman, J. B., Hess, P. & Tsien, R. W. *J. gen. Physiol.* (in the press).
13. Ault, B., Evans, R. H., Francis, A. S., Oakes, D. J. & Watkins, J. C. *J. Physiol., Lond.* **307**, 413–428 (1980).
14. Crunelli, V. & Mayer, M. L. *Brain Res.* **311**, 392–396 (1984).
15. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).
16. Cull-Candy, S. G. & Ogden, D. C. *Proc. R. Soc.* **B224**, 367–373 (1985).
17. Hagiwara, S. & Byerly, L. *A. Rev. Neurosci.* **4**, 69–125 (1981).
18. Smith, S. J., MacDermott, A. B. & Weight, F. F. *Nature* **304**, 350–352 (1983).
19. Gorman, A. L. F. & Thomas, M. V. *J. Physiol., Lond.* **308**, 259–285 (1980).
20. Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315–321 (1984).
21. Sladeczek, F., Pin, J. P., Récasens, M., Bockaert, J. & Weiss, S. *Nature* **317**, 717–719 (1985).
22. Schoffelmeer, A. M. N. & Mulder, A. H. *J. Neurochem.* **40**, 615–621 (1983).
23. Evans, R. H. & Watkins, J. C. *J. Physiol., Lond.* **277**, 57P (1977).
24. Nowak, L. M. & Ascher, P. *Soc. Neurosci. Abstr.* **10**, 23 (1984).
25. Mayer, M. L. & Westbrook, G. L. *Soc. Neurosci. Abstr.* **11**, 785 (1985).
26. Ascher, P. & Nowak, L. *J. Physiol., Lond. Proc.* (in the press).
27. Fabiato, A. & Fabiato, F. *Ann. N.Y. Acad. Sci.* **307**, 491–522 (1978).
28. Nowak, L. M. & Ascher, P. *Soc. Neurosci. Abstr.* **11**, 953 (1985).
29. Lynch, G., Larson, J., Kelso, S., Barrionuevo, G. & Schottler, F. *Nature* **305**, 719–721 (1983).
30. Collingridge, G. L., Kehl, S. J. & McLennan, H. *J. Physiol., Lond.* **334**, 33–46 (1983).
31. Parsons, R. L. in *Calcium in Drug Action* (ed. Weiss, G. B.) 289–314 (Plenum, New York, 1978).
32. Miledi, R. *Proc. R. Soc.* **B209**, 447–452 (1980).
33. Adams, D. J., Dwyer, T. M. & Hille, B. *J. gen. Physiol.* **75**, 493–510 (1980).
34. Edwards, C. *Neuroscience* **7**, 1335–1366 (1982).


# Replacing the complementarity-determining regions in a human antibody with those from a mouse

**Peter T. Jones, Paul H. Dear, Jefferson Foote, Michael S. Neuberger & Greg Winter**

Laboratory of Molecular Biology, Medical Research Council, Hills Road, Cambridge CB2 2QH, UK

**The variable domains of an antibody consist of a β-sheet framework with hypervariable regions (or complementarity-determining regions—CDRs) which fashion the antigen-binding site. Here we attempted to determine whether the antigen-binding site could be transplanted from one framework to another by grafting the CDRs. We substituted the CDRs from the heavy-chain variable region of mouse antibody B1–8, which binds the hapten NP-cap (4-hydroxy-3-nitrophenacetyl caproic acid; $K_{NP\text{-}cap} = 1.2\ \mu M$), for the corresponding CDRs of a human myeloma protein. We report that in combination with the B1–8 mouse light chain, the new antibody has acquired the hapten affinity of the B1–8 antibody ($K_{NP\text{-}cap} = 1.9\ \mu M$). Such 'CDR replacement' may offer a means of constructing human monoclonal antibodies from the corresponding mouse monoclonal antibodies.**

The three-dimensional structures of several immunoglobulins show that the variable domains consist of two β-sheets pinned together by a disulphide bridge, with their hydrophobic faces packed together[1–3]. The individual β-strands are linked by loops which at one tip of the β-sheet may fashion a binding pocket for small haptens[1,2]. Sequence comparisons among heavy- and light-chain variable domains ($V_H$ and $V_L$ respectively) reveal that each domain has three CDRs flanked by four relatively conserved regions (framework regions—FRs)[4]. As seen in the structure of the human myeloma protein NEWM (Fig. 1), the CDRs include each of the three main loops. Often the CDRs also include the ends of the β-strands, suggesting that side chains at the ends of the β-strands may help to fix the conformation or orientation of the loops. The framework regions form the bulk of the β-sheet, although for example in the $V_H$ domain of NEWM, FR1 includes part of the loop between the two β-sheets and CDR2 not only forms a loop but a complete β-strand (Fig. 1). The structure of the β-sheet framework is similar in different antibodies, as the packing of different side chains is accommodated by slight shifts between the two β-strands[5]. Furthermore, the packing together of $V_L$ and $V_H$ FRs is conserved[6], therefore the orientation of $V_L$ with respect to $V_H$ is fixed. We wondered whether the FRs represent a simple β-sheet scaffold on which new binding sites may be built, and

©1986 Nature Publishing Group

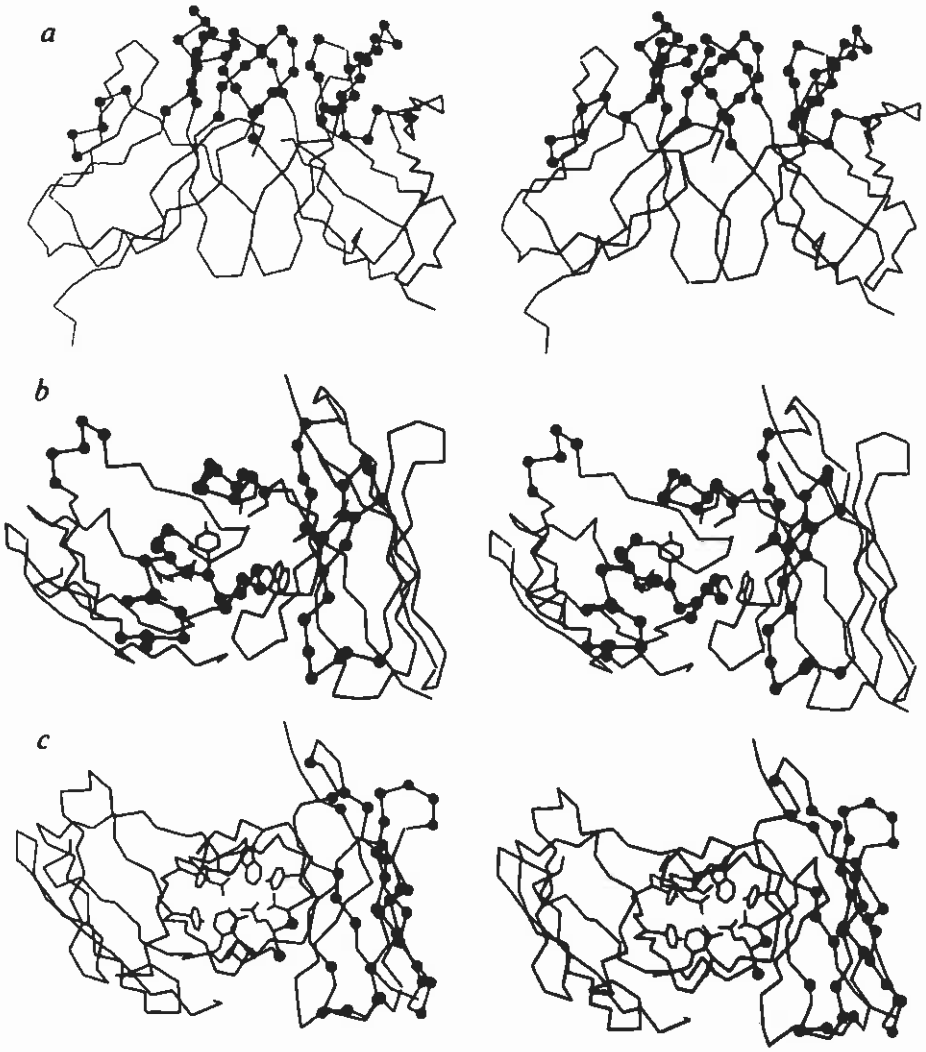



**Fig. 1** Stereo pairs of the $V_H$ (right) and $V_L$ (left) domains of the human myeloma protein NEWM[1,8] generated using the computer graphics program FRODO[25]. The tracings indicate the backbone of $C^\alpha$ atoms for the framework regions. *a*, The $C^\alpha$ atoms of the CDRs (●) cluster at the tip of the variable domain. *b*, A view into the hapten binding pocket with the CDRs in the order (clockwise from noon): $V_H$ CDR3, CDR1 and CDR2, and $V_L$ CDR3, CDR1 and CDR2. The side chains lining the binding pocket ($V_L$ A 28, N 30, Y 90, S 93, R 95; $V_H$ W 47, Y 50, F 52, I 100, A 101) lie almost entirely in the CDRs. *c*, The $C^\alpha$ atoms in the NEWM $V_H$ domain are marked (●) where side chains in the mouse B1-8 $V_H$ domain are different. The side chains ($V_L$ Y 35, Q 37, A 42, P 43, Y 86, F 99; $V_H$ V 37, Q 39, L 45, Y 94, W 107) involved in packing $V_H$ and $V_L$ framework regions are traced. In the $V_H$ domain all these side chains are conserved in mouse B1-8.

whether the structure of the CDRs (and antigen binding) is therefore independent of the FR context. To answer these questions experimentally, we have grafted the CDRs from one antibody to another, to determine whether antigen binding transfers with the CDRs.

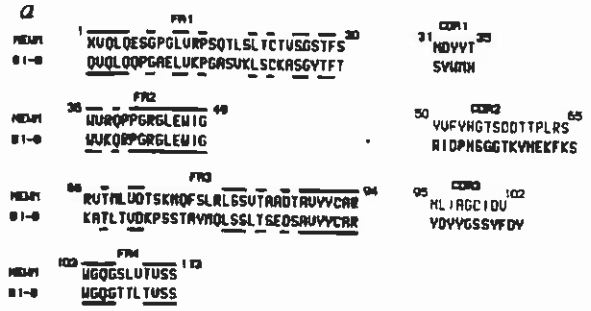
We grafted the CDRs from the $V_H$ domain of the mouse monoclonal antibody B1-8 (ref. 7) into the $V_H$ domain of the human myeloma protein NEWM, whose crystallographic structure is known[1,8]. The $V_H$ domain of the CDR donor (B1-8) is attached to a $\mu$ constant region and associated with a mouse $\lambda$1 light chain, and the antibody is directed against the hapten NP-cap. Both the $V_H$ and $V_L$ domains seem to have a role in determining the affinity of the antibody for NP-cap as the substitution of either domain by other, often highly related variable domains can destroy hapten binding (refs 7, 9 and M.S.N., unpublished results). In the $V_H$ domain, each of the CDRs has been implicated in NP-cap binding[10], but the class of constant domains attached to $V_H$ does not seem to affect binding of hapten[11,12]. The CDRs from the $V_H$ domain of antibody B1-8 (ref. 13) are longer than the CDRs which they replace in NEWM[4] and this may give rise to a deeper binding pocket.

Most of the residues conserved between the $V_H$ domains of B1-8 and NEWM are located in FR2, FR4 and the carboxy-terminal third of FR3 (Fig. 2*a*) and largely form the region of $\beta$-sheet which is packed against the light chain. Therefore, it might be expected that the $V_H$ domain of B1-8 (hereafter abbreviated to $MV_{NP}$) and the hybrid B1-8/NEWM domain ($HuV_{NP}$) would dock in a similar manner with the mouse $V_L$ domain to form the antigen-combining site[6]. The more variable FR1 and N-terminal two-thirds of FR3 form the other $\beta$-sheet which is exposed to solvent (Fig. 1*c*).

The gene encoding the $HuV_{NP}$ domain was constructed by gene synthesis (Fig. 2*b*). We then constructed a plasmid, pSV-$HuV_{NP}H\varepsilon$, in which the $HuV_{NP}$ domain is linked to a human $\varepsilon$ constant region, and cloned into a pSV2gpt-derived vector[14]. The plasmid DNA was introduced into cells of the J558L mouse myeloma by spheroplast fusion. J558L secretes $\lambda$1 light chains which have been shown to associate with heavy chains containing a $MV_{NP}$ variable domain, to create a binding site for NP-cap or the related hapten NIP-cap (3-iodo-4-hydroxy-5-nitrophenylacetyl caproic acid)[7]. As the plasmid pSV-$HuV_{NP}H\varepsilon$ contains the *gpt* marker (encoding guanine phosphoribosyltransferase), stably transfected myeloma cells could be selected in medium containing mycophenolic acid[14]; transfectants would be expected to secrete an antibody ($HuV_{NP}$-IgE) with a heavy chain composed of a $HuV_{NP}$ variable domain and human $\varepsilon$ constant regions, and the $\lambda$1 light chain of the J558L myeloma. The culture supernatants of several *gpt*$^+$ clones were assayed by radioimmunoassay and found to contain NIP-cap-binding antibody. The antibody secreted by one such clone was purified from the culture supernatant by affinity chromatography on NIP-cap–Sepharose, and by SDS-polyacrylamide gel electrophoresis the protein was indistinguishable from the mouse chimaeric $MV_{NP}$-IgE (ref. 12) (results not shown). The $HuV_{NP}$-IgE antibody competes effectively with $MV_{NP}$-IgE for binding to both anti-human $\varepsilon$ (Fig. 3*a*) and NIP-cap coupled to bovine serum albumin (NIP-BSA) (Fig. 3*b*).

The affinities of $HuV_{NP}$-IgE for NP-cap and NIP-cap were

©1986 Nature Publishing Group



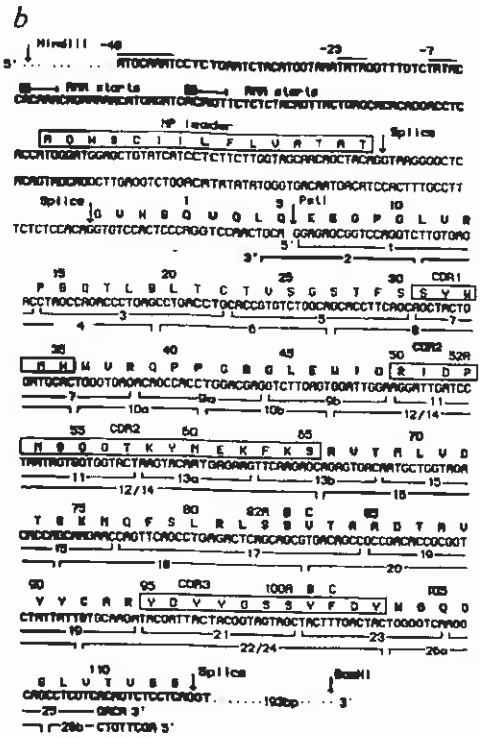


**Fig. 2** *a*, Amino-acid sequence of the $V_H$ domain of the human myeloma protein NEWM compared with that of the mouse B1-8 (anti-NP) antibody, and divided into FRs and CDRs according to Kabat *et al.*[4]. Conserved residues are marked with a line above and below the residue. *b*, Amino-acid and nucleotide sequence of the $HuV_{NP}$ gene, in which the CDRs from the $V_H$ domain of the mouse B1-8 antibody alternate with the FRs of human NEWM protein. The gene was constructed by replacing a section of the $MV_{NP}$ gene in the vector pSV-$V_{NP}$ (ref. 12) with a synthetic fragment encoding the $HuV_{NP}$ domain. Thus the 5′ and 3′ noncoding sequences, the leader sequence, the leader–variable region intron, five N-terminal and four C-terminal amino acids are derived from the $MV_{NP}$ gene; the rest of the coding sequence is derived from the synthetic $HuV_{NP}$ fragment. The oligonucleotides are aligned with the corresponding portions of the $HuV_{NP}$ gene. For convenience in cloning, the ends of oligonucleotides 25 and 26b form a *Hin*dII site followed by a *Hin*dIII site, and the sequences of the 25/26b oligonucleotides therefore differ from the $HuV_{NP}$ gene.

**Methods.** The synthetic gene for $HuV_{NP}$ was constructed as a *Pst*I–*Hin*dIII fragment. The nucleotide sequence was derived from the protein sequence using the computer program ANALYSEQ[26] with optimal codon usage taken from the sequences of mouse constant-region genes. The oligonucleotides used in synthesis (1–26b, 28 in total) vary in size from 14 to 59 bases and were made on a Biosearch SAM or an Applied Biosystems machine, and purified on 8 M urea/polyacrylamide gels[27]. The oligonucleotides were assembled in eight single-stranded blocks (A–D and A′–D′) containing oligonucleotides 1, 3, 5 and 7 (block A), 2, 4, 6 and 8 (block A′), 9, 11, 13a and 13b (block B), 10a, 10b and 12/14 (block B′), 15 and 17 (block C), 16 and 18 (block C′), 19, 21, 23 and 25 (block D), and 20, 22/24, 26a and 26b (block D′). In a typical assembly, for example of block A, 50 pmol of oligonucleotides 1, 3, 5 and 7 were phosphorylated at the 5′ end with $T_4$ polynucleotide kinase and mixed with 5 pmol of the terminal oligonucleotide which had been phosphorylated with 5 μCi of [γ-$^{32}$P]ATP (Amersham; 3,000 Ci mmol$^{-1}$). These oligonucleotides were annealed by heating to 80 °C and cooling to room temperature over 30 min with unkinased oligonucleotides 2, 4 and 6 as splints in 150 μl of 50 mM Tris-HCl *p*H 7.5, 10 mM $MgCl_2$. For the ligation, ATP (1 mM) and dithiothreitol (10 mM) were added, together with 50 units of $T_4$ DNA ligase (Anglian Biotechnology Ltd), and the mixture was incubated for 30 min at room temperature. EDTA was added to 10 mM, the sample extracted with phenol, precipitated from ethanol, dissolved in 20 μl of water and boiled for 1 min with an equal volume of formamide dyes. Then the sample was loaded onto a thin (0.3 mm) 8 M urea/10% polyacrylamide gel[27] and a band of the expected size detected by autoradiography and eluted by soaking. The two full-length single strands were assembled from A–D and A′–D′ using splint oligonucleotides; thus, A–D were annealed and ligated in 30 μl as above with 100 pmol each of oligonucleotides 10a, 16 and 20 as splints, then incubated overnight (A′–D′ were constructed with oligonucleotides 7, 13b and 17 as splints). After phenol/ether extraction blocks A–D were annealed with blocks A′–D′, small amounts were cloned in the vector M13 mp18 (ref. 28) then cut with *Pst*I and *Hin*dIII, and the gene sequenced by the dideoxy technique[29]. The $MV_{NP}$ gene was transferred as a *Hin*dIII–*Bam*HI fragment from the vector pSV-$V_{NP}$ (ref. 12) to the vector M13mp8 (ref. 30). To facilitate the replacement of $MV_{NP}$ coding sequences by the synthetic $HuV_{NP}$ fragment, three *Hin*dII sites were removed from the 5′ noncoding sequence by site-directed mutagenesis, and a new *Hin*dII site subsequently introduced near the end of FR4. By cutting the vector with *Pst*I and *Hin*dII, most of the $V_{NP}$ coding sequence falls out and the synthetic fragment could be introduced as a *Pst*I–*Hin*dII fragment. The sequence at the *Hin*dII site was corrected to give NEWM FR4 by site-directed mutagenesis. The *Hin*dIII–*Bam*HI fragment, now carrying the $HuV_{NP}$ gene, was excised from M13 and cloned back into pSV-$V_{NP}$ to replace the $MV_{NP}$ gene (and yield the vector pSV-$HuV_{NP}$). Finally, the heavy-chain constant domains of human IgE (ref. 31) were introduced as a *Bam*HI fragment to yield the vector pSV-$HuV_{NP}H_\varepsilon$, which was transfected into the myeloma line J558L by spheroplast fusion. The sequence of the $HuV_{NP}$ gene in pSV-$HuV_{NP}H\varepsilon$ was checked by re-cloning the *Hin*dIII–*Bam*HI fragment back into M13mp8 (ref. 30).

**Table 1** Affinity of $HuV_{NP}$-IgE and $MV_{NP}$-IgE for the haptens NP-cap and NIP-cap

| | $K_{NP\text{-}cap}$ (μM) | $K_{NIP\text{-}cap}$ (μM) |
|---|---|---|
| $MV_{NP}$-IgE | 1.2±0.1 | 0.02±0.01 |
| $HuV_{NP}$-IgE | 1.9±0.2 | 0.07±0.02 |

The affinity of $HuV_{NP}$-IgE and $MV_{NP}$-IgE for NP-cap was determined by fluorescence quenching with excitation at 295 nm and emission observed at 340 nm (ref. 22). Antibody solutions were diluted to 100 nM in phosphate-buffered saline, filtered (0.45 μm-pore cellulose acetate) and titrated with NP-cap in the range 0.2–20 μM. As a control, the mouse D1-3 antibody[23], which does not bind hapten, was titrated in parallel. A decrease in the ratio of the fluorescence of $HuV_{NP}$-IgE or $MV_{NP}$-IgE (as appropriate) to that of the D1-3 antibody was taken as being proportional to NP-cap occupancy of the antigen-binding sites. The maximum quench was ~40% for both $HuV_{NP}$-IgE and $MV_{NP}$-IgE, and hapten dissociation constants were determined from least-squares fits of triplicate data sets to a hyperbola. The concentration of NIP-cap was varied from 10 to 300 nM, and ~50% quenching of fluorescence was observed at saturation. As the antibody concentrations were comparable to the values of the dissociation constants, data were fitted by least-squares to an equation[24] describing tight binding inhibition.

then measured directly using the fluorescence quench technique, and compared with those of $MV_{NP}$-IgE (Table 1). The antibodies $HuV_{NP}$-IgE and $MV_{NP}$-IgE have similar affinities for either hapten (NP-cap or NIP-cap), and although the affinity of $HuV_{NP}$-IgE for both haptens is slightly lower than that of $MV_{NP}$-IgE (2–3-fold, 0.3–0.6 kcal mol$^{-1}$), the difference in affinity is less than expected for loss of either a hydrogen bond or van der Waals' contact from the active site of an enzyme[15,16]. Thus, it seems that binding affinity and specificity for hapten can be conferred on a human antibody by grafting in the CDRs from an appropriate mouse antibody.

Is this result likely to be general? This would assume (1) that antigen usually binds to the CDRs, and any contacts to the FRs are made to the polypeptide backbone or to conserved side chains, and (2) that substitutions in the FRs do not usually affect the conformation of the CDR loops. These assumptions seem reasonable: thus, in the structure of a complex of the D1-3 antibody with lysozyme (R. A. Mariuzza, S. Phillips and R. J. Poljak, personal communication) most contacts to the lysozyme are made by the CDRs, but there is also a hydrogen bond in FR1 of the $V_H$ domain from the β-OH of Thr 30 (often conserved or replaced by Ser). Similarly, the conformation of CDR loops

©1986 Nature Publishing Group

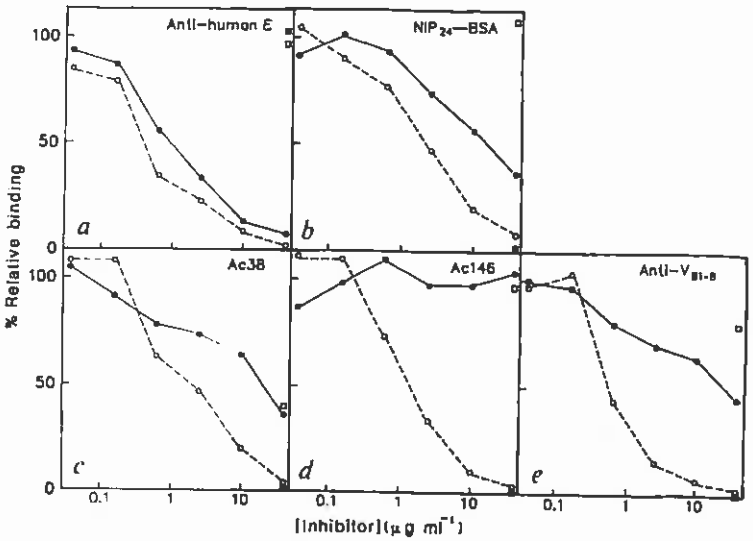



Fig. 3 Comparison of $HuV_{NP}$ and $MV_{NP}$ IgEs in binding inhibition assays. Various concentrations of $HuV_{NP}$-IgE (●) and $MV_{NP}$-IgE (○) were used to compete the binding of radiolabelled $MV_{NP}$-IgE to polyvinyl microtitre plates that had been coated with *a*, sheep anti-human ε antiserum (Seward Laboratory); *b*, $(NIP\text{-}cap)_{24}$-BSA; *c*, Ac38 anti-idiotypic antibody; *d*, Ac146 anti-idiotypic antibody; *e*, rabbit anti-$MV_{NP}$ antiserum. Binding was also carried out in the presence of $MV_{NP}$-IgM antibody JW1/2/2 (ref. 32) (■) as well as in the presence of JW5/1/2 (□), which is an IgM antibody that differs from JW1/2/2 at 13 residues mainly located in $V_H$ CDR2 (M.S.N., unpublished results). Values of binding are relative to the binding in the absence of inhibitor.

between β-strands depends on loop size and specific interactions of the loop back to the β-sheet. However, in the same class of variable domains ($V_H$, $V_\kappa$ or $V_\lambda$) these interactions are usually conserved (ref. 5 and A. M. Lesk and C. Chothia, personal communication).

While human monoclonal antibodies have therapeutic potential in human disease, they can be difficult to prepare[17] and treatment of patients with mouse monoclonal antibodies often increases the titre of circulating antibody against the mouse immunoglobulin[18]. As chimaeric antibodies containing human constant domains[12,19,20] and variable domains made by grafting mouse CDRs into human FRs, could have therapeutic potential, we wondered whether the $HuV_{NP}$-IgE antibody loses antigenic determinants associated with the $MV_{NP}$ variable region (idiotopes). The binding of $HuV_{NP}$-IgE and $MV_{NP}$-IgE to both monoclonal and polyclonal anti-idiotypic antibodies directed against the $MV_{NP}$ domain was examined by using inhibition assays. As shown in Fig. 3*d*, the $HuV_{NP}$-IgE antibody has lost the $MV_{NP}$ idiotypic determinant recognized by antibody Ac146 (ref. 21). Furthermore, $HuV_{NP}$-IgE also binds the antibody Ac38 (ref. 21) less well (Fig. 3*c*), therefore it is not surprising that $HuV_{NP}$-IgE has lost many of the determinants recognized by a polyclonal rabbit anti-idiotypic antiserum (Fig. 3*e*). While the loss of idiotypic determinants that accompanies 'humanizing' of the $V_H$ region is reassuring in view of potential therapeutic applications, it does suggest that the recognition of the hapten and of anti-idiotypic antibodies is not equivalent. Thus the $HuV_{NP}$-IgE antibody retains hapten binding but has lost idiotypic determinants, indicating that the immunoglobulin uses different sites to bind hapten and anti-idiotypic antibodies. It appears, therefore, that both FR and CDR side chains form the binding site for these anti-idiotopes, but mainly CDR side chains interact with hapten.

We thank C. Milstein for suggesting this project, K. Rajewsky and M. Reth for the anti-idiotypic antibodies Ac38 and Ac146, and A. M. Lesk, C. Chothia, R. J. Leatherbarrow and C. Milstein for helpful discussions. J.F. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research.

Received 17 February; accepted 17 March 1986.

1. Poljak, R. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **70**, 3305-3310 (1973).
2. Segal, D. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **71**, 4298-4302 (1974).
3. Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. *J. molec. Biol.* **141**, 369-391 (1980).
4. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. in *Sequences of Proteins of Immunological Interest* (U.S. Department of Health and Human Services, 1983).
5. Lesk, A. M. & Chothia, C. *J. molec. Biol.* **160**, 325-342 (1982).
6. Chothia, C., Novotny, J., Bruccoleri, R. & Karplus, M. *J. molec. Biol.* **186**, 651-663 (1985).
7. Reth, M., Hämmerling, G. J. & Rajewsky, K. *Eur. J. Immun.* **8**, 393-400 (1978).
8. Saul, F. A., Amzel, M. & Poljak, R. J. *J. biol. Chem.* **253**, 585-597 (1978).
9. Brüggemann, M., Radbruch, A. & Rajewsky, K. *EMBO J.* **1**, 629-634 (1982).
10. Reth, M., Bothwell, A. L. M. & Rajewsky, K. in *Immunoglobulin Idiotypes and Their Expression* (eds Janeway, C., Wigzell, H. & Fox, C. F.) 169-178 (Academic, New York, 1981).
11. Neuberger, M. S. & Rajewsky, K. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1138-1142 (1981).
12. Neuberger, M. S. *et al. Nature* **314**, 268-270 (1985).
13. Bothwell, A. L. M. *et al. Cell* **24**, 625-637 (1981).
14. Mulligan, R. C. & Berg, P. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2072-2076 (1981).
15. Fersht, A. R. *et al. Nature* **314**, 235-238 (1985).
16. Fersht, A. R., Wilkinson, A. J., Carter, P. & Winter, G. *Biochemistry* **24**, 5858-5861 (1985).
17. Boyd, J. E., James, K. & McClelland, D. B. L. *Trends Biotechnol.* **2**, 70-77 (1984).
18. Shawler, D. L., Bartholomew, R. M., Smith, L. M. & Dillman, R. O. *J. Immun.* **135**, 1530-1535 (1985).
19. Morrison, S. L., Johnson, M. J., Herzenberg, L. A. & Oi, V. T. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6851-6855 (1984).
20. Boulianne, G. L., Hozumi, N. & Shulman, M. J. *Nature* **312**, 643-646 (1984).
21. Reth, M., Imanishi-Kari, T. & Rajewsky, K. *Eur. J. Immun.* **9**, 1004-1013 (1979).
22. Eisen, H. N. *Meth. med. Res.* **10**, 115-121 (1964).
23. Mariuzza, R. A. *et al. J. molec. Biol.* **170**, 1055-1058 (1983).
24. Segal, I. H. in *Enzyme Kinetics*, 73-74 (Wiley, New York, 1975).
25. Jones, T. A. in *Computational Crystallography* (ed. Sayre, D.) 303-310 (Clarendon, Oxford, 1982).
26. Staden, R. *Nucleic Acids Res.* **12**, 521-538 (1984).
27. Sanger, F. & Coulson, A. *FEBS Lett.* **87**, 107-110 (1978).
28. Yanisch-Perron, C., Vieira, J. & Messing, J. *Gene* **33**, 103-119 (1985).
29. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
30. Messing, J. & Vieira, J. *Gene* **19**, 269-276 (1982).
31. Flanagan, J. G. & Rabbitts, T. H. *EMBO J.* **1**, 655-660 (1982).
32. Neuberger, M. S., Williams, G. T. & Fox, R. O. *Nature* **312**, 604-608 (1984).

# Regulation of human insulin gene expression in transgenic mice

**Richard F Selden*, Marek J. Skoślewicz†, Kathleen Burke Howie*, Paul S. Russell† & Howard M. Goodman***

Departments of *Molecular Biology and †Surgery, Massachusetts General Hospital, and Departments of *Genetics and †Surgery, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

**Insulin is a polypeptide hormone of major physiological importance in the regulation of fuel homeostasis in animals (reviewed in refs 1, 2). It is synthesized by the β-cells of pancreatic islets, and circulating insulin levels are regulated by several small molecules, notably glucose, amino acids, fatty acids and certain pharmacological agents. Insulin consists of two polypeptide chains (A and B, linked by disulphide bonds) that are derived from the proteolytic cleavage of proinsulin, generating equimolar amounts of the mature insulin and a connecting peptide (C-peptide). Humans, like most vertebrates, contain one proinsulin gene[3,4], although several species, including mice[5] and rats[6,7], have two highly homologous insulin genes. We have studied the regulation of serum insulin**

© 1986 Nature Publishing Group

*Proc. Natl. Acad. Sci. USA*
Vol. 86, pp. 10029–10033, December 1989
Immunology

# A humanized antibody that binds to the interleukin 2 receptor

(chimeric antibody/antibody affinity/autoimmune disease)

CARY QUEEN*, WILLIAM P. SCHNEIDER*, HAROLD E. SELICK*†, PHILIP W. PAYNE*, NICHOLAS F. LANDOLFI*, JAMES F. DUNCAN*‡, NEVENKA M. AVDALOVIC*, MICHAEL LEVITT§, RICHARD P. JUNGHANS¶, AND THOMAS A. WALDMANN¶

*Protein Design Labs, 3181 Porter Drive, Palo Alto, CA 94304; §Department of Cell Biology, Stanford University, Stanford, CA 94305; and ¶Metabolism Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

*Contributed by Thomas A. Waldmann, August 30, 1989*

**ABSTRACT** **The anti-Tac monoclonal antibody is known to bind to the p55 chain of the human interleukin 2 receptor and to inhibit proliferation of T cells by blocking interleukin 2 binding. However, use of anti-Tac as an immunosuppressant drug would be impaired by the human immune response against this murine antibody. We have therefore constructed a "humanized" antibody by combining the complementarity-determining regions (CDRs) of the anti-Tac antibody with human framework and constant regions. The human framework regions were chosen to maximize homology with the anti-Tac antibody sequence. In addition, a computer model of murine anti-Tac was used to identify several amino acids which, while outside the CDRs, are likely to interact with the CDRs or antigen. These mouse amino acids were also retained in the humanized antibody. The humanized anti-Tac antibody has an affinity for p55 of $3 \times 10^9$ $M^{-1}$, about 1/3 that of murine anti-Tac.**

The cellular receptor for the lymphokine interleukin 2 (IL-2) plays an important role in regulation of the immune response (reviewed in ref. 1). The complete IL-2 receptor (IL-2R) consists of at least two IL-2-binding peptide chains: the p55 or Tac peptide (2, 3), and the recently discovered p75 peptide (4, 5). Identification and characterization of the p55 peptide were facilitated by the development of a monoclonal antibody, anti-Tac, which binds to human p55 (2). The p55 peptide was found to be expressed on the surface of T cells activated by an antigen or mitogen but not on resting T cells. Treatment of human T cells with anti-Tac antibody strongly inhibits their proliferative response to antigen or to IL-2 by preventing IL-2 binding (3, 6).

These results suggested that anti-IL-2R antibodies would be immunosuppressive when administered *in vivo*. Indeed, injection of an anti-IL-2R antibody into mice and rats greatly prolonged survival of heart allografts (7, 8). Anti-IL-2R was also effective in rats against experimental graft-versus-host disease (9). In animal models of autoimmune disease, an anti-IL-2R antibody alleviated insulitis in nonobese diabetic mice and lupus nephritis in NZB × NZW mice (10). Anti-Tac itself was highly effective in prolonging survival of kidney allografts in cynomolgus monkeys (11).

In human patients, the specificity of anti-Tac for activated T cells might give it an advantage as an immunosuppressive agent over OKT3 (monoclonal anti-CD3), which is effective in treating kidney transplant rejection (12), but which suppresses the entire peripheral T-cell population. In fact, in phase I clinical trials for kidney transplantation, prophylactic administration of anti-Tac significantly reduced the incidence of early rejection episodes, without associated toxicity (13). Furthermore, treatment with anti-Tac induced temporary partial or complete remission in three of nine patients with Tac-expressing adult T-cell leukemia (14). However, as a murine monoclonal antibody, anti-Tac elicits a strong human antibody response against itself, as does OKT3 (15). This response would prevent its long-term use in treating autoimmune conditions or suppressing organ transplant rejection.

The immune response against a murine monoclonal antibody may potentially be reduced by transforming it into a chimeric antibody. Such antibodies, produced by methods of genetic engineering, combine the variable (V) region binding domain of a mouse (or rat) antibody with human antibody constant (C) regions (16–18). Hence, a chimeric antibody retains the binding specificity of the original mouse antibody but contains less amino acid sequence foreign to the human immune system. Chimeric antibodies have been produced against a number of tumor-associated antigens (19–21). In some but not all cases, the chimeric antibodies have mediated human complement-dependent cytotoxicity (CDC) or antibody-dependent cellular cytotoxicity (ADCC) more efficiently than the mouse antibodies (21).

When the murine antibody OKT3 is used in human patients, much of the resulting antibody response is directed against the V region of OKT3 rather than the C region (15). Hence, chimeric antibodies in which the V region is still nonhuman may not have sufficient therapeutic advantages over mouse antibodies. To further reduce the immunogenicity of murine antibodies, Winter and colleagues constructed "humanized" antibodies in which only the minimum necessary parts of the mouse antibody, the complementarity-determining regions (CDRs), were combined with human V region frameworks and human C regions (22–25). We report here the construction of chimeric and humanized anti-Tac antibodies.‖ For the humanized antibody, sequence homology and molecular modeling were used to select a combination of mouse and human sequence elements that would reduce immunogenicity while retaining high binding affinity.

## MATERIALS AND METHODS

**Construction of Plasmids.** cDNA cloning was by the method of Gubler and Hoffman (26), and sequencing was by the dideoxy method (27). The plasmid pV$\kappa$1 (Fig. 1*A*) was constructed from the following fragments: an approximately 4550-base-pair (bp) *Bam*HI–*Eco*RI fragment from the plas-

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: IL-2R, interleukin 2 receptor; CDR, complementarity-determining region; CDC, complement-dependent cytotoxicity; ADCC, antibody-dependent cellular cytotoxicity; V, variable; J, joining; C, constant.

†Present address: Biospan, 440 Chesapeake Drive, Redwood City, CA 94063.

‡Present address: Beckman Instruments, 1050 Page Mill Road, Palo Alto, CA 94304.

‖The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M28250 and M28251).

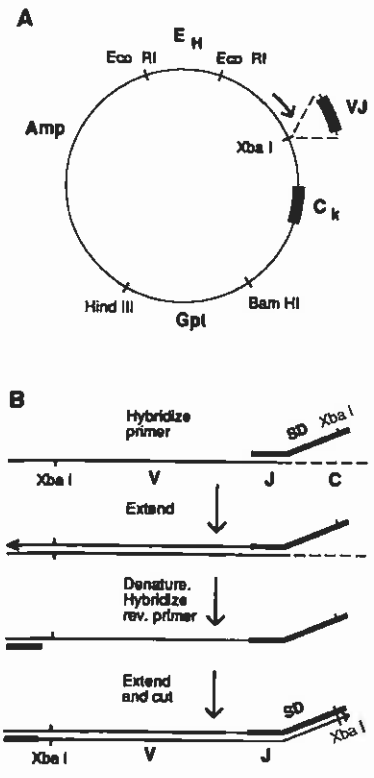



FIG. 1. (*A*) Schematic diagram of the plasmids pVκ1 and pLTac. Light chain exons are shown as boxes. An arrow indicates the direction of transcription from the κ promoter. $E_H$, heavy chain enhancer. Not drawn to scale. (*B*) Schematic diagram of the method used to excise the *V–J* region. SD, splice donor sequence; rev. primer, reverse primer.

mid pSV2gpt (28) containing the *amp* and *gpt* genes; an 1800-bp *Eco*RI–*Bgl* II fragment from pKcatH (29) containing the heavy chain enhancer and κ promoter; and a 1500-bp *Eco*RI–*Xba* I fragment containing the human $C_\kappa$ region (30). Similarly, pVγ1 was constructed starting from a 4850-bp *Bam*HI–*Eco*RI fragment of the plasmid pSV2hph (a gift of A. Smith, A. Miyajima, and D. Strehlow, Stanford University), which is analogous to pSV2gpt except that the *gpt* gene is replaced by the *hyg* gene (31). This fragment was combined with the *Eco*RI–*Bgl* II fragment from pKcatH and a 2800-bp *Hin*dIII–*Pvu* II fragment containing the human γ1 constant region, isolated from a phage kindly provided by L. Hood (32). In each case, the fragments were combined by standard methods (ref. 33, pp. 390–401), with an *Xba* I linker inserted between the κ promoter fragment and the 5′ end of the C region fragment.

**Construction of Chimeric Genes.** *Eco*RI fragments containing the anti-Tac light and heavy chain cDNAs were separately inserted into the *Eco*RI site of the phage M13mp11D, a variant of M13mp11 (34) in which the *Eco*RI and *Xba* I sites of the polylinker were filled in and joined. The resulting phage, in which the 5′ ends of the cDNAs abutted the *Xba* I site, were respectively denoted M13L and M13H. The V–J (J, joining) segments of the cDNAs, followed by splice donor signals, were precisely excised from these phage, using a double-priming scheme (Fig. 1*B*). For the light chain, the following primer was synthesized (Applied Biosystems model 380B DNA synthesizer) and purified by gel electrophoresis: 5′-CCAGAATTCTAGAAAAGTGTACTTAC-GTTTCAGCTCCAGCTTGGTCCC-3′. From the 3′ end, the first 22 residues of the primer are the same as the last 22 residues of the $J_\kappa 5$ segment (noncoding strand). The next 16 nucleotides are the same as the sequence that follows $J_\kappa 5$ in mouse genomic DNA and therefore includes a splice donor signal. The final 10 nucleotides of the oligonucleotide include an *Xba* I site.

We hybridized this oligonucleotide to M13L and extended it with the Klenow fragment of DNA polymerase. The DNA was heat-denatured, hybridized with an excess of the "reverse primer" 5′-AACAGCTATGACCATG-3′, again extended with Klenow DNA polymerase, and digested with *Xba* I. The digested DNA was run on a gel, and an approximately 400-bp fragment was excised and inserted into the *Xba* I site of pVκ1. Sequencing showed that the fragment consisted of the V–J region of the light chain cDNA followed by the splice donor "tail," as expected (Fig. 1*B*), and pLTac, a clone with the appropriate orientation, was chosen. In an analogous fashion, the heavy chain V–J segment, followed by the mouse $J_H2$ splice donor sequence, was excised from M13H and inserted into the *Xba* I site of pVγ1 to yield pGTac.

**Computer Analysis.** Sequences were manipulated and homology searches were performed with the MicroGenie Sequence Analysis Software (Beckman). The molecular model of the anti-Tac V region was constructed with the ENCAD program (35) and examined with the MIDAS program (36) on an IRIS 4D-120 graphics workstation (Silicon Graphics).

**Construction of Genes for Humanized Antibody.** Nucleotide sequences were selected that encoded the protein sequences of the humanized light and heavy chain V regions including signal peptides (*Results*), generally utilizing codons found in the mouse anti-Tac sequence. These nucleotide sequences also included the same splice donor signals used in the chimeric genes and an *Xba* I site at each end. For the heavy chain V region, four overlapping 120- to 130-nucleotide-long oligonucleotides were synthesized that encompassed the entire sequence on alternating strands. The oligonucleotides were phosphorylated with polynucleotide kinase, annealed, extended with T4 DNA polymerase, cut with *Xba* I, and ligated into the *Xba* I site of pUC19 (34), using standard reaction conditions. An insert with the correct sequence was recloned in pVγ1. The humanized light chain V region was constructed similarly..

**Transfections.** For each antibody constructed, the light chain plasmid was first transfected into Sp2/0 mouse myeloma cells (ATTC CRL 1581) by electroporation (Bio-Rad Gene Pulser) and cells were selected for *gpt* expression (28). Clones secreting a maximal amount of light chain, as determined by ELISA, were transfected with the heavy chain plasmid and cells were selected for hygromycin B resistance (31). Clones secreting a maximal amount of complete antibody were detected by ELISA. The clones were used for preparation of chimeric and humanized antibodies.

**Antibody Purification.** Medium from confluent cells was passed over a column of staphylococcal protein A-Sepharose CL-4B (Pharmacia), and antibody was eluted with 3 M $MgCl_2$. Antibody was further purified by ion-exchange chromatography on BakerBond ABx (J. T. Baker). Final antibody concentration was determined, assuming that 1 mg/ml has an $A_{280}$ of 1.4. Anti-Tac antibody itself was purified as described (2).

**Affinity Measurements.** Affinities were determined by competition binding. HuT-102 human T-lymphoma cells (ATTC TIB 162) were used as source of p55 Tac antigen. Increasing amounts of competitor antibody (anti-Tac, chimeric, or humanized) were added to 1.5 ng of radioiodinated (Pierce Iodo-Beads) tracer anti-Tac antibody (2 μCi/μg; 1 Ci = 37 GBq) and incubated with $4 \times 10^5$ HuT cells in 0.2 ml of binding buffer (RPMI 1040 medium with 10% fetal calf serum, human IgG at 100 μg/ml, 0.1% sodium azide) for 3 hr at room temperature. Cells were washed and pelleted, and their radioactivities were measured, and the concentrations of bound and free tracer antibody were calculated. The affinity of mouse anti-Tac was determined by Scatchard plot analy-

sis, using anti-Tac itself as the competitor. Then the affinities of the chimeric and humanized antibodies were each calculated according to the formula $[X] - [\text{anti-Tac}] = (1/K_a) - (1/K_a)$, where $K_a$ is the affinity of anti-Tac ($9 \times 10^9$ $M^{-1}$), $K_x$ is the affinity of the competitor X, [ ] indicates the concentration of competitor antibody at which bound/free tracer binding is $R_0/2$, and $R_0$ is maximal bound/free tracer binding (37).

## RESULTS

**Cloning of Light and Heavy Chain cDNA.** A cDNA library in λgt10 was prepared from anti-Tac hybridoma cells and screened with oligonucleotide probes for the mouse κ and γ2a constant regions. The cDNA inserts from four κ-positive and four γ2a-positive phage were subcloned in M13mp19. Partial sequencing showed that two of the κ isolates had one sequence, and the other two had another sequence. In one pair, a $V_\kappa$ gene segment was joined to the $J_\kappa 2$ segment out of its reading frame. In addition, the conserved cysteine at position 23 was absent from this V segment, and the sequences of the two isolates differed slightly. Presumably, these clones were the result of an aberrant joining event in one κ allele, which continued to undergo somatic mutation after the formation of the hybridoma.

The V–J segments of the other pair of κ clones were sequenced completely and were identical. This light chain uses the $J_\kappa 5$ segment. Partial sequencing of the four γ2a clones showed they were all from the same gene. The V–J segments of two were sequenced completely and were identical. This heavy chain uses the $J_H 2$ segment and is of subgroup II (38). The DNA sequences have been deposited with GenBank;[‖] the deduced protein sequences are shown in Fig. 2. As both alleles of the κ light chain were accounted for and only one heavy chain sequence was detected, we tentatively assigned these sequences to the anti-Tac antibody genes.

**Construction of Chimeric Genes.** Plasmid vectors were prepared for the construction and expression of chimeric light and heavy chain genes. The plasmid pVκ1 (Fig. 1*A*) contains the human genomic $C_\kappa$ segment, including 336 bp of the preceding intron and the poly(A) signal. It also contains the promoter sequence from the MOPC 41 κ gene and the heavy chain enhancer sequence, which synergize to form a very strong transcriptional unit (29). There is a unique *Xba* I site between the promoter and the intron. A similar plasmid, pVγ1, was prepared by using the human $C_\gamma 1$ region in place of the $C_\kappa$ region. In that case, the region inserted between the *Xba* I and *Bam*HI sites extended from about 210 bp 5′ of the $C_H 1$ exon to beyond the $C_H 3$ exon.

Our strategy was to insert the V–J region from the anti-Tac κ cDNA, followed by a splice donor signal, at the *Xba* I site of pVκ1 to construct the plasmid pLTac. Doing so created a chimeric κ gene with a short synthetic intron between the mouse V–J and human $C_\kappa$ segments (Fig. 1*A*). For this purpose, we used a form of double primer-directed mutagenesis (*Materials and Methods*; Fig. 1*B*). Similarly, the V–J region from the anti-Tac γ2a heavy chain cDNA, followed by a splice donor signal, was inserted into the *Xba* I site of pVγ1. The resulting plasmid, pGTac, contained a chimeric heavy chain gene, with a synthetic intron between the mouse V–J and human $C_\gamma 1$ segments.

**Construction of a Humanized Anti-Tac Antibody.** In selecting a human antibody to provide the variable region framework for the humanized anti-Tac antibody, we reasoned that the more homologous the human antibody was to the original anti-Tac antibody, the less likely would combining the anti-Tac CDRs with the human framework be to introduce distortions into the CDRs. The anti-Tac heavy chain sequence was therefore compared by computer with all the human heavy chain sequences in the National Biomedical Research Foundation Protein Identification Resource (release 15). The heavy chain V region of the Eu antibody (of human heavy chain subgroup I; ref. 38) was 57% identical to the anti-Tac heavy chain V region (Fig. 2*B*); all other complete $V_H$ regions in the data bank were 30–52% identical. However, no one human light chain V region was especially homologous to the anti-Tac light chain. We therefore chose to use the Eu light chain (of human light chain subgroup I; ref. 38) together with the Eu heavy chain to supply the framework sequences for the humanized antibody. The CDRs in the humanized antibody were of course chosen to be identical to the anti-Tac CDRs (Fig. 2).

A computer program was used to construct a plausible molecular model of the anti-Tac V domain (Fig. 3), based on homology to other antibody V domains with known crystal structure and on energy minimization. Graphic manipulation shows that a number of amino acid residues outside of the CDRs are in fact close enough to them to either influence their conformation or interact directly with antigen. When these residues differ between the anti-Tac and Eu antibodies, the residue in the humanized antibody was chosen to be the anti-Tac residue rather than the Eu residue. This choice was made for residues 27, 30, 48, 67, 68, 98, and 106 in the humanized heavy chain, and for 47 and 59 in the humanized light chain (Figs. 2 and 3; amino acids shown in blue in Fig. 3), although we now consider the light chain residue 59, which was chosen on the basis of an earlier model, to be doubtful. In this way, we hoped to better preserve the precise structure of the CDRs at the cost of possibly making the humanized antibody slightly less "human."

Different human light or heavy chain V regions exhibit strong amino acid homology outside of the CDRs, within the framework regions. However, a given V region will usually

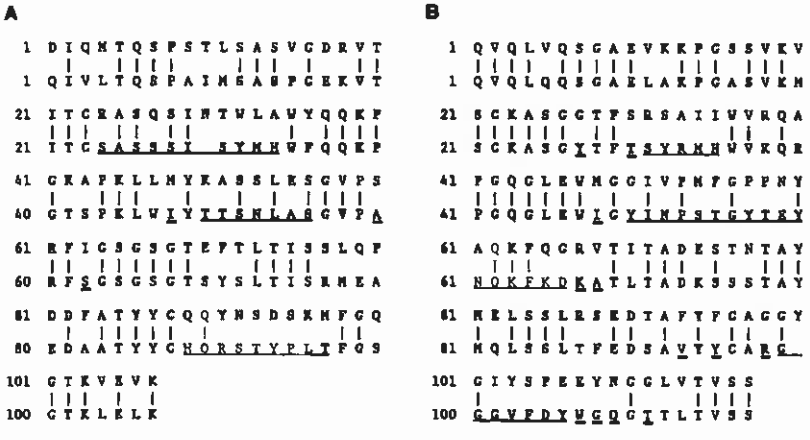



FIG. 2. Amino acid sequences of the humanized anti-Tac light (*A*) and heavy (*B*) chains. The sequences of the Eu antibody light and heavy chains (upper lines) are shown aligned above the mouse anti-Tac light and heavy chain sequences (lower lines), with a | indicating identity of amino acids. The three CDRs in each chain are underlined, and the other mouse amino acids used in the humanized antibody are double underlined. Hence, the humanized sequences are the same as the upper (Eu) sequences, except where the amino acid is underlined or double underlined.

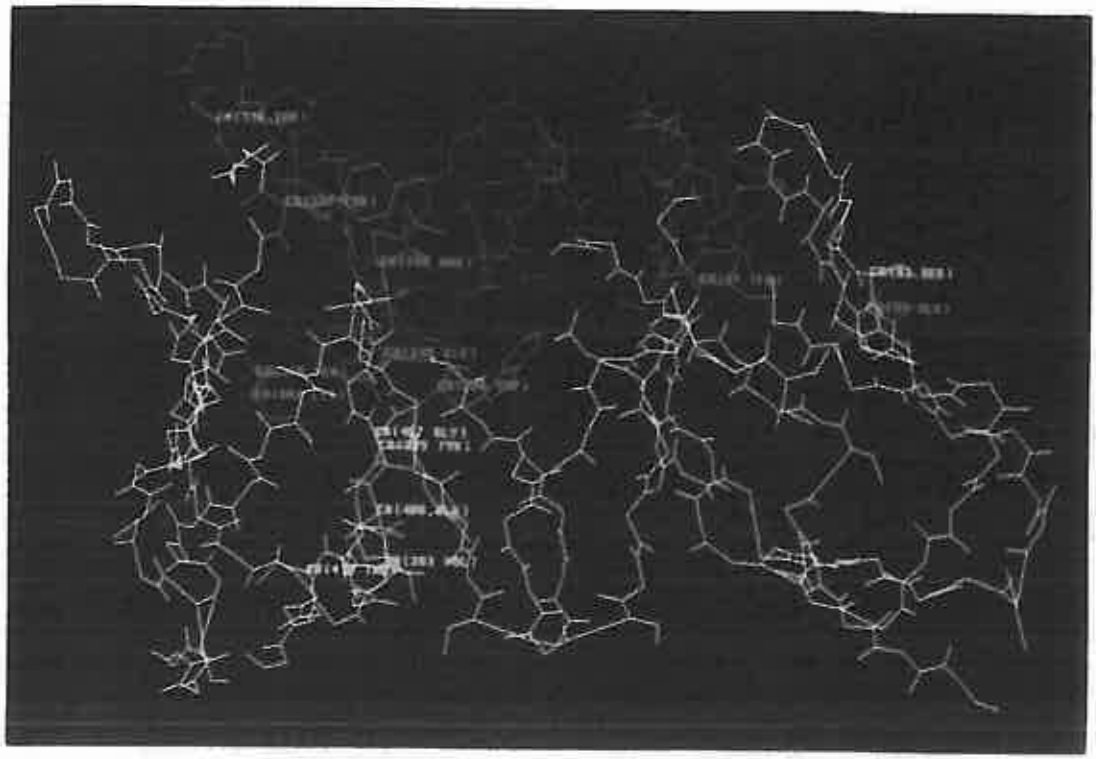


FIG. 3. Model of the mouse anti-Tac antibody V region, generated with the ENCAD program and displayed with the MIDAS program. Amino acids in the CDRs are shown in red; amino acids potentially interacting with the CDRs are shown in blue; other mouse amino acids used in the humanized antibody are shown in yellow, as described in the text. Thus, all amino acids transferred from the anti-Tac sequence to the humanized antibody are shown in red, blue, or yellow. Residue 1 is the first amino acid of $V_\kappa$; residue 301 is the first amino acid of $V_H$.

contain exceptional amino acids, atypical of other human V regions, at several framework positions. The Eu antibody contains such unusual residues at positions corresponding to 93, 95, 98, 106, 107, 108, and 110 of the humanized heavy chain and 47 and 62 of the light chain (Fig. 2), as determined by visual comparison of the Eu heavy and light chain V regions with other human V regions of subgroup I (38). The Eu antibody contains several other unusual residues, but at the listed positions, the murine anti-Tac antibody actually has a residue much more typical of human sequences than does Eu. At these positions, we therefore chose to use the anti-Tac residue rather than the Eu residue in the humanized antibody, to make the antibody more generically human. Some of these residues had already been selected because of their proximity to the CDRs, as described above (the remaining ones are shown in yellow in Fig. 3).

These criteria allowed the selection of all amino acids in the humanized antibody V regions as coming from either anti-Tac or Eu (Fig. 2). DNA segments encoding the desired heavy and light chain amino acid sequences were synthesized. These DNA segments also encoded typical immunoglobulin signal sequences for processing and secretion, and they contained splice donor signals at their 3' end. The light and heavy chain segments were cloned, respectively, in pVκ1 and pVγ1 to form the plasmids pHuLTac and pHuGTac.

Properties of Chimeric and Humanized Antibodies. Sp2/0 cells, a nonproducing mouse myeloma line, were transfected sequentially with pLTac and pGTac (chimeric genes) or with pHuLTac and pHuGTac (humanized genes). Cell clones were selected first for antibiotic resistance and then for maximal antibody secretion, which reached 3 $\mu g/10^6$ cells per 24 hr. S1 nuclease mapping of RNA extracted from the cells transfected with pLTac and pGTac showed that the synthetic introns between the V and C regions (Fig. 1*A*) were correctly spliced (data not shown). Antibody was purified from the culture medium of cells producing the chimeric or humanized antibody. When analyzed by reducing SDS/polyacrylamide gel electrophoresis, the antibodies showed only two bands, having the expected molecular weights 50,000 and 25,000.

Flow cytometry showed that the chimeric and humanized antibodies bound to Hut-102 and CRII.2 cells, two human T-cell lines that express the p55 chain of the IL-2R, but not to CEM and other cell lines that do not express the IL-2R. To determine the binding affinity of the chimeric and humanized antibodies, their ability to compete with labeled mouse anti-Tac for binding to Hut-102 cells was determined. The affinity of chimeric anti-Tac was indistinguishable from that of anti-Tac (data not shown), as expected from the fact that their entire V regions are identical. The affinity of humanized anti-Tac for membrane-bound p55 was $3 \times 10^9$ $M^{-1}$, about 1/3 the measured affinity of $9 \times 10^9$ $M^{-1}$ of anti-Tac itself (Fig. 4).

## DISCUSSION

Because monoclonal antibodies can be produced that are highly specific for a wide variety of cellular targets, antibody therapy holds great promise for the treatment of cancer, autoimmune conditions, and other diseases. However, this promise has not been widely realized, largely because most monoclonal antibodies, which are of mouse origin, are immunogenic when used in human patients and are ineffective at recruiting human immune effector functions such as CDC and ADCC. A partial solution to this problem is the use of chimeric antibodies (16), which combine the V region binding domains of mouse antibodies with human antibody C regions. Initially, chimeric antibodies were constructed by combining genomic clones of the V and C region genes. However, this method is very time consuming because of the difficulty of genomic cloning, especially from tetraploid hybridomas.

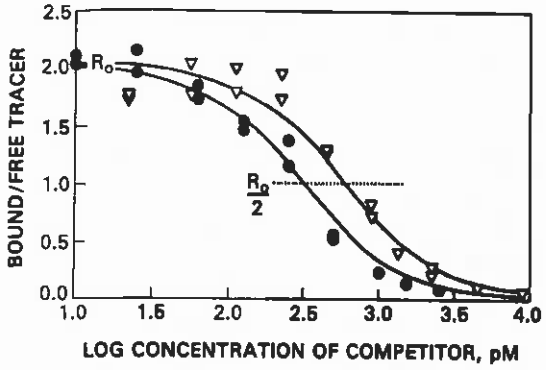



FIG. 4. Competitive binding of labeled anti-Tac tracer to Hut-102 cells. Duplicate samples are shown. ●, Mouse anti-Tac competitor; ▽, humanized anti-Tac competitor.

More recently, cDNA clones of the V and C regions have been combined, but this method is also tedious because of the need to join the V and C regions precisely (20, 21). Here we show that the V region from a readily obtainable cDNA clone can be easily joined to a human genomic C region, which need only be cloned once, by leaving a synthetic intron between the V and C regions. When linked to suitable transcriptional regulatory elements and transfected into an appropriate host cell, such chimeric genes produce antibody at a high level.

Chimeric antibodies represent an improvement over mouse antibodies for use in human patients, because they are presumably less immunogenic and sometimes mediate CDC or ADCC more effectively (21). For example, chimeric anti-Tac mediates ADCC with activated human effector cells, whereas murine anti-Tac does not (unpublished data). However, the mouse V region can itself be highly immunogenic (15). Winter and colleagues therefore took the further, innovative, step of combining the CDRs from a mouse (or rat) antibody with the framework region from a human antibody (22–25), thus reducing the xenogeneic elements in the humanized antibody to a minimum. Unfortunately, in some cases the humanized antibody had significantly less binding affinity for antigen than did the original mouse antibody. This is not surprising, because transferring the mouse CDRs from the mouse framework to the human framework could easily deform them.

In humanizing the anti-Tac antibody, which binds to the p55 chain of the human IL-2R, we have introduced two ideas that may have wider applicability. First, the human framework was chosen to be as homologous as possible to the original mouse antibody to reduce any deformation of the mouse CDRs. Second, computer modeling was used to identify several framework amino acids in the mouse antibody that might interact with the CDRs or directly with antigen, and these amino acids were transferred to the human framework along with the CDRs. The resulting humanized antibody has a high affinity, $3 \times 10^9$ $M^{-1}$, for its antigen. Further work is needed to determine to what extent the choice of human framework and the preservation of particular mouse amino acids in fact contributed to the affinity of the humanized antibody. The extent to which humanization eliminates immunogenicity will need to be addressed in clinical trials, where humanized anti-Tac will be administered to patients with Tac-expressing lymphomas or selected autoimmune diseases or to patients receiving organ transplants.


1. Waldmann, T. A. (1989) *Annu. Rev. Biochem.* **58**, 875–911.
2. Uchiyama, T., Broder, S. & Waldmann, T. A. (1981) *J. Immunol.* **126**, 1393–1397.
3. Leonard, W. J., Depper, J. M., Uchiyama, T., Smith, K. A., Waldmann, T. A. & Greene, W. C. (1982) *Nature (London)* **300**, 267–269.
4. Tsudo, M., Kozak, R. W., Goldman, C. K. & Waldmann, T. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9694–9698.
5. Sharon, M., Klausner, R. D., Cullen, B. R., Chizzonite, R. & Leonard, W. J. (1986) *Science* **234**, 859–863.
6. Depper, J. M., Leonard, W. J., Robb, R. J., Waldmann, T. A. & Greene, W. C. (1983) *J. Immunol.* **131**, 690–696.
7. Kirkman, R. L., Barrett, L. V., Gaulton, G. N., Kelley, V. E., Ythier, A. & Strom, T. B. (1985) *J. Exp. Med.* **162**, 358–362.
8. Kupiec-Weglinski, J. W., Diamantstein, T., Tilney, N. L. & Strom, T. B. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2624–2627.
9. Volk, H.-D., Brocke, S., Osawa, H. & Diamantstein, T. (1986) *Clin. Exp. Immunol.* **66**, 126–131.
10. Kelley, V. E., Gaulton, G. N., Hattori, M., Ikegami, H., Eisenbarth, G. & Strom, T. B. (1988) *J. Immunol.* **140**, 59–61.
11. Reed, M. H., Shapiro, M. E., Strom, T. B., Milford, E. L., Carpenter, C. B., Weinberg, D. S., Reimann, K. A., Letvin, N. L., Waldmann, T. A. & Kirkman, R. L. (1989) *Transplantation* **47**, 55–59.
12. Ortho Multicenter Transplant Study Group (1985) *N. Engl. J. Med.* **313**, 337–342.
13. Kirkman, R. L., Shapiro, M. E., Carpenter, C. B., Milford, E. L., Ramos, E. L., Tilney, N. L., Waldmann, T. A., Zimmerman, C. E. & Strom, T. B. (1989) *Transplant. Proc.* **21**, 1766–1768.
14. Waldmann, T. A., Goldman, C. K., Bongiovanni, K. F., Sharrow, S. O., Davey, M. P., Cease, K. B., Greenberg, S. J. & Longo, D. L. (1988) *Blood* **72**, 1805–1816.
15. Jaffers, G. J., Fuller, T. C., Cosimi, A. B., Russell, P. S., Winn, H. J. & Colvin, R. B. (1986) *Transplantation* **41**, 572–578.
16. Morrison, S. L., Johnson, M. J., Herzenberg, L. A. & Oi, V. T. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6851–6855.
17. Boulianne, G. L., Hozumi, N. & Shulman, M. J. (1984) *Nature (London)* **312**, 643–646.
18. Neuberger, M. S., Williams, G. T., Mitchell, E. B., Jouhal, S. S., Flanagan, J. G. & Rabbitts, T. H. (1985) *Nature (London)* **314**, 268–270.
19. Sun, L. K., Curtis, P., Rakowicz-Szulczynska, E., Ghrayeb, J., Chang, N., Morrison, S. L. & Koprowski, H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 214–218.
20. Whittle, N., Adair, J., Lloyd, C., Jenkins, L., Devine, J., Schlom, J., Raubitschek, A., Colcher, D. & Bodmer, M. (1987) *Protein Eng.* **1**, 499–505.
21. Liu, A. Y., Robinson, R. R., Hellstrom, K. E., Murray, E. D., Jr., Chang, C. P. & Hellstrom, I. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3439–3443.
22. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. (1986) *Nature (London)* **321**, 522–525.
23. Verhoeyen, M., Milstein, C. & Winter, G. (1988) *Science* **239**, 1534–1536.
24. Reichmann, L., Clark, M., Waldmann, H. & Winter, G. (1988) *Nature (London)* **332**, 323–327.
25. Hale, G., Dyer, M. J. S., Clark, M. R., Phillips, J. M., Marcus, R., Riechmann, L., Winter, G. & Waldmann, H. (1988) *Lancet* **i**, 1394–1399.
26. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
28. Mulligan, R. C. & Berg, P. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2072–2076.
29. Garcia, J. V., Bich-Thuy, L. T., Stafford, J. & Queen, C. (1986) *Nature (London)* **322**, 383–385.
30. Hieter, P. A., Max, E. E., Seidman, J. G., Maizel, J. V., Jr., & Leder, P. (1980) *Cell* **22**, 197–207.
31. Sugden, B., Marsh, K. & Yates, J. (1985) *Mol. Cell. Biol.* **5**, 410–413.
32. Ellison, J. W., Berson, B. J. & Hood, L. E. (1982) *Nucleic Acids Res.* **10**, 4071–4079.
33. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
34. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33**, 103–119.
35. Levitt, M. (1983) *J. Mol. Biol.* **168**, 595–617.
36. Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988) *J. Mol. Graphics* **6**, 13–27.
37. Berzofsky, J. A. & Berkower, I. J. (1984) in *Fundamental Immunology*, ed. Paul W. E. (Raven, New York), pp. 595–644.
38. Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. (1987) *Sequences of Proteins of Immunological Interest* (National Institutes of Health, Bethesda, MD).

**Human therapeutic antibody**

All the above mentioned therapeutic antibodies contain mouse proteins. Eventually, the human patient will develop antibodies against them. Thus their usefulness is limited to a relatively short period of time as summarized here:

| Therapeutic antibody | Type | Length of therapy |
|---|---|---|
| OKT3 | Mouse | 1 week |
| Rituzimab | Chimeric | 4 weeks |
| Daclizumab | Humanized | 10 weeks |

In most cases, they are used in conjunction with chemotherapy or radiation therapy. While human antibodies have a half-life of about 24 days in humans somewhat longer than humanized mouse therapeutic antibodies, they are not antigenic in humans and can be used indefinitely for therapy. Therefore various approaches have been developed to manufacture completely human therapeutic antibodies. Among many different methods, the following four are usually favored:

1. Introduction of human immunoglobulin genes into transgenic mice (Bruggemann M, Casky HM, Teale C, Waldmann H, Williams GT, Surani MA, Neubergen MS (1989) A repertoire of monoclonal antibodies with human heavy chains from transgenic mice. Proc Natl Acad Sci USA, 86, 6709-6713; Bruggemann N, Taussig M (1997) Production of human antibody reportoires in transgenic mice. Curr Opin Biotech, 7, 455-458).

2. Displaying antibody Fab on filamentous phages (Hoogenboom HR, Griffiths AD, Hohnson KS, Chiswell DJ, Hudson P, Winter G (1991) Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. Nucl Acids Res, 19, 4133-4137; Clackson T, Hoogwnboom HR, Griffiths AS, Winter G (1991) Making antibody fragments using phage display libraries. Nature, 352, 624-628; Nissim A, Hoogenboom HR, Tomlinson IM, Flynn G, Midgley C, Lane D, Winter G (1994) Antibody fragments from a 'single pot' phage display ;ibrary as immunochemical reagent. EMBO J, 13, 692-698).

3. Expression of antibodies in *Escherichia coli* periplasm (Simmons LC, Reilly D, Klimowaki L, Raju TS, Meng G, Sims P, Hong K, Shields RL, Damico LA, Rancatore P, Yansura DG (2002) Expression of full-length immunoglobulins in *Escherichia coli*: rapid and efficient production of aglycosylated antibodies. J Immunol Meth, 153, 133-147).

4. Production of antibodies in human hybridoma Schilbach K, Kreyenberg H, Gieselhart A, Niethammer D, Handgretinger R (2004) Tissue Antigens, 63, 122-131).

One of the first FDA approved, in 2002, completely human therapeutic antibody is adalimumab selected by the phage display library method against tumor necrosis factor α and used primarily for rheumatoid arthritis and other autoimmune diseases (Kempeni J (1999) Preliminary results of early clinical trials with the fully human anti-TNFα monoclonal antibody D2E7. Ann Rheum Dis, 58, 170-172; also see Adalimumab – Wikipedia). It has a half-life of about 14 days in human patients, somewhat less than the usual 24 days for most human antibodies, but can be tolerated by human patients indefinitely. It can be injected subcutaneously every two weeks for one or two years.

Currently, there are about 20 to 30 FDA approved therapeutic antibodies of various types as mentioned above. They are used to treat different types of cancers, transplantation rejections, autoimmune diseases, etc. There are several hundred more under development.

Additional refinements and new methodologies of generating more therapeutic antibodies for passive immunization are being developed. Such medications are thus known as the magic bullets hitting the target.

*Proc. Natl. Acad. Sci. USA*
Vol. 86, pp. 6709–6713, September 1989
Immunology

# A repertoire of monoclonal antibodies with human heavy chains from transgenic mice

(human immunoglobulin/DNA rearrangement/antibody engineering)

MARIANNE BRÜGGEMANN*, HELENE M. CASKEY†, CATHERINE TEALE*, HERMAN WALDMANN‡, GARETH T. WILLIAMS†, M. AZIM SURANI*, AND MICHAEL S. NEUBERGER†

*Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, England; ‡Department of Pathology, University of Cambridge, Cambridge CB2 2QQ, England; and †Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

*Communicated by M. F. Perutz, May 11, 1989 (received for review April 18, 1989)*

**ABSTRACT** **The introduction of human immunoglobulin gene segments in their unrearranged configuration into the germ line of mice might allow the production of a repertoire of human antibodies. Such transgenic mice could be used for the production of human monoclonal antibodies against human antigens. To test the feasibility of this approach, mice were created that carry a human heavy-chain minilocus comprising unrearranged immunoglobulin variable, diversity, and joining elements linked to a human $\mu$-chain gene. The gene segments of this minilocus are rearranged in a large proportion of cells in thymus and spleen but not in nonlymphoid tissue. Some 4% of the B lymphocytes synthesize human $\mu$ chains resulting in a serum titer of about 50 $\mu$g of transgenic IgM antibody per ml. Hybridomas were established from the transgenic mice that stably secreted several micrograms of antibodies containing human $\mu$ heavy chains per milliliter.**

The preparation of monoclonal antibodies of rodent origin is an established and widely used procedure (1). Their therapeutic application would be greatly assisted by the availability of monoclonal antibodies of human origin as these are likely to exhibit greatly reduced immunogenicity in humans. However, two types of problem beset the preparation of such antibodies (2). (*i*) Most immortalized human cell lines or hybridomas do not stably express large amounts of antibody. (*ii*) *In vivo* immunization of humans is not feasible for many antigens; *in vitro* priming is usually inefficient.

Antibodies are composed of variable (V) and constant (C) domains with the V domains defining the antigen specificity. Recently, a considerable advance has been achieved by making chimeric antibodies composed of mouse V domains linked to human C domains; these retain the antigen-binding specificity of the original mouse antibody but over two-thirds of the sequence of the human chimeric antibody is of human origin (3–5). Further improvements have been made by exploiting the fact that the hypervariable regions of the V domains are—at least in several cases—sufficient to determine antigen specificity; antibodies have been prepared in which only the hypervariable regions are of rodent origin (6, 7).

The genes encoding the V domains of the immunoglobulin heavy (IgH) chain are assembled during lymphoid differentiation by rearrangement of constituent $V_H$, diversity (D), and joining heavy-chain ($J_H$) segments (8). In this work, we create transgenic mice that carry a human IgH minilocus containing $V_H$, D, and $J_H$ segments linked to a $C_\mu$ gene. The mice rearrange this minilocus in their lymphoid tissue and contain B lymphocytes that make IgM molecules composed of human $\mu$ chains. This approach may allow the preparation of a repertoire of entirely human monoclonal antibodies that lack any sequences of rodent origin.

## MATERIALS AND METHODS

**Construction of the Minilocus and of Transgenic Mice.** The minilocus was assembled in a pUC12 derivative that contained an *Sph* I linker in the *Nar* I site. The mouse $C_\mu$ membrane exons were inserted as a *Hin*dIII–*Sph* I fragment and the *Sph* I site then converted to a *Bgl* II site by linker insertion. The $C_\mu$ region (9) was completed by inserting the human $S_\mu$ and $C_\mu$ 1–4 exons as an *Xba* I fragment into the polylinker. For the V domain, an *Eco*RI fragment containing the human $V_H26$ segment (10) was cloned into a pUC18 derivative that included a *Sac* I–*Kpn* I fragment that spans the mouse $V_H186$-2 gene (11). A separate pUC12 derivative was created with the mouse IgH enhancer in the *Xba* I site and a *Bgl* II fragment that spans the human $J_H$ locus in the *Bam*HI site. Following the use of linkers, the fragment containing the $V_H$ segments and the fragment containing the enhancers and $J_H$ segments were both introduced into the pUC12 derivative that included $C_\mu$. For the creation of the D segments, an M13tg131 clone that contained the 221-base-pair (bp) *Xho* I–*Sac* I mouse D-Q52 fragment was mutagenized using oligonucleotides to create D segments (TTTATTACTACGGTAGCAGCTAC and TCTACTATGGTTACGAC) that we designate D′-FL16 and D′-SP2. The mutagenized D segments were excised as *Bam*HI–*Bgl* II fragments from the M13tg131 polylinker, joined in tandem with the mouse D-Q52 and inserted into the unique *Bam*HI site of the pUC12 derivative containing the $V_H$, $J_H$, and $C_\mu$ genes to yield the final minilocus (Fig. 1). This plasmid containing the minilocus was linearized with *Bgl* II and introduced into the pronucleus of fertilized (C57BL/6 × CBA)$F_1$ mice as described (12).

**Analysis of DNA and RNA.** DNA from cell lines, from tissues, and from spleen cells that had been cultured for 3 days in medium containing 50 $\mu$g of bacterial lipopolysaccharide (LPS) per ml was extracted by phenol extraction following proteinase K digestion. Southern blots were hybridized with probes made by random oligonucleotide priming (13). Cytoplasmic RNA was prepared by phenol extraction of lysates made by use of Nonidet P-40 and analyzed by using glyoxal gels and the Northern blot procedure of Thomas (14).

**Preparation of Hybridomas and Analysis of Immunoglobulin.** Hybridomas were made by fusion of spleen cells with the NS0 plasmacytoma. In some fusions, LPS was included in the selective medium to stimulate class-switching of the endogenous IgH locus. The enzyme-linked immunosorbent assay (ELISA) for human $\mu$ chain was carried out by incu-

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: V, C, D, and J regions, variable, constant, diversity, and joining regions of immunoglobulin; IgH, immunoglobulin heavy chain; LPS, bacterial lipopolysaccharide.



209

bating samples in wells of a microtiter plate to which had been bound a monoclonal anti-human $\mu$ antibody (Binding Site, Birmingham, UK). Human $\mu$ determinants retained in the well were then detected by using a biotinylated goat anti-human $\mu$ antiserum (Amersham) and peroxidase-conjugated streptavidin. The assay was calibrated by using a human chimeric IgM antibody (15). Immunofluorescence as well as immunoprecipitations of biosynthetically labeled samples were carried out by using the biotinylated anti-human $\mu$ antiserum in the presence of an unlabeled mouse IgM antibody as competitor essentially as described (5, 15).

## RESULTS

**Transgenic Mice Carrying the Human Mini-IgH Locus.** A plasmid (Fig. 1) was assembled that contains the human $J_H$ cluster and $\mu$ C region ($C_\mu$) linked to two germ-line $V_H$ gene segments and four D segments. For the $V_H$ segments, we chose the human $V_H26$ and the mouse $V_H186$-2 genes (10, 11). The mouse $V_H$ gene was included as it is known to be functional, whereas we were not initially sure that the same was true of the human $V_H26$. Two of the D segments (D'-SP2 and D'-FL16) were generated by site-directed mutagenesis of the mouse D-Q52 element, since sequences corresponding to these D segments have been described in human heavy chains (16). The minilocus also included mouse and human D-Q52 elements. Since our human $C_\mu$ clone did not extend sufficiently far downstream of $C_\mu4$, we used $\mu$ membrane exons of mouse origin. The minilocus includes a 1-kilobase (kb) *Xba* I fragment containing the mouse IgH enhancer as well as the 5' end of the human IgH enhancer.

The linearized minilocus plasmid was injected into the pronucleus of fertilized mouse eggs. Of 32 mice born, 12 carried the minilocus transgene as judged by Southern blot analysis of tail DNA. Most of the work described here was carried out on descendants of founders F17, F19, and F29, which have 3–5 copies of the transgene, as well as offspring of founder F23, which transmits about 50 copies.

**The Minilocus Rearranges In Spleen and Thymus.** To test whether the gene segments in the transgenic minilocus were able to undergo rearrangement, DNA was prepared from different organs and, following digestion with *Eco*RI, analyzed by Southern blot hybridization. The results (Fig. 2*A*) reveal that the locus has rearranged at high frequency in spleen and thymus of the low-copy-number F17 transgenic mice but not in the nonlymphoid tissues. Comparison of spleen and thymus DNAs from transgenic offspring of founders F17 and F24 reveals a striking similarity in their pattern of rearrangements (Fig. 2*B*). In contrast to the mice bearing a low transgene copy number, offspring of founder F23 (the high-copy-number founder) show a greatly reduced frequency of rearrangement of the minilocus.

A detailed analysis of the different types of rearrangement must await molecular cloning, although some information can be obtained by Southern blot analysis. Any rearrangement of a D element to a $J_H$ segment occurring by a looping-out mechanism will delete the *Bam*HI site. Analysis of DNA hybridized with a vector probe (Fig. 2*C*) suggests that a substantial proportion of the rearrangements involves D-$J_H$ integration as loss of the *Bam*HI site leads to an increase in the size of the hybridizing fragment in an *Fsp* I + *Bam*HI digest. Similarly, if the minilocus rearrangement involves one of the two $V_H$ segments, then there should be an alteration in the size of the *Sca* I fragment that hybridizes with the vector probe. The results (Fig. 2*D*) suggest that only a minority of the rearrangements involves the $V_H$ genes as most of the hybridizing DNA in the *Sca* I digest is unrearranged. Comparison of the results obtained with DNA from thymus and from LPS-activated spleen suggests that the pattern of rearrangements in B and T lymphocytes is similar.

**Mouse B Cells Making Human $\mu$ Chains.** To ascertain whether the DNA rearrangements of the minilocus could lead to the synthesis of antibodies comprising human $\mu$ heavy chains, serum immunoglobulin was analyzed in an ELISA assay. Analysis of sera from the 12 founder mice revealed that 7 were positive for human $\mu$ chains and had transgenic IgM concentrations in the range 1–100 $\mu$g/ml. It was notable that the high-copy-number founder F23 does contain antibodies with human $\mu$ chains in its serum (about 5 $\mu$g/ml) despite the fact that most of the copies of the transgene do not rearrange. Immunofluorescence analysis was performed to discover the proportion of B lymphocytes making human $\mu$ chains. The results (Fig. 3) show that about 4% of mitogen-activated B cells from an F17 offspring stain in the cytoplasm for human $\mu$ chain; this is similar to the proportion of unstimulated B cells that stain for human $\mu$ chain on the cell surface (not shown).

**Hybridomas from the Transgenic Mice.** Splenic hybridomas were prepared from the transgenic mice to allow a more detailed analysis of the expressed immunoglobulin. After fusion, cells were incubated in selective medium under conditions of limiting dilution. ELISA analysis of culture supernatants indicated that 3% of the hybrids secreted an antibody containing human $\mu$ chains. Following cloning by limiting dilution, the concentration of human IgM antibody in the culture supernatant was estimated as 0.5–5 $\mu$g/ml. Northern blot analysis of RNA from several cloned hybrids (Fig. 4*A*) revealed the presence of human $\mu$-chain mRNA and evidence that the human $V_H26$ and the mouse $V_H186$-2 were being utilized with an apparent preference for $V_H186$-2 although one hybrid (NW497/16.20) appears to express both $V_H$ genes. Short $V_H186$-2 transcripts were observed in several of the hybrids that made human $\mu$ chains; these short transcripts may be analogous to the germ-line $V_H$ transcripts

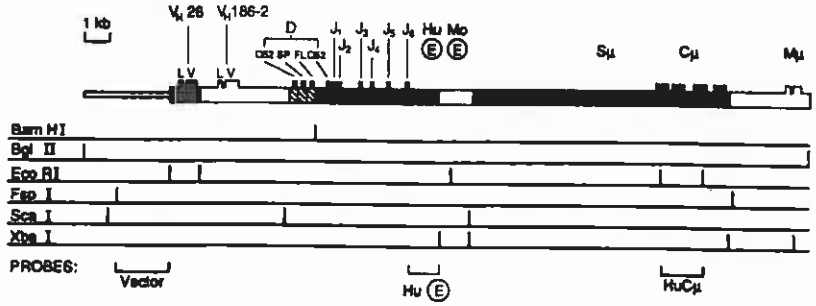



FIG. 1. Minilocus. Human sequences are depicted with thick filled lines, vector sequences with thin lines, and mouse sequences with unfilled lines. The D elements that were created by site-directed mutagenesis of mouse D-Q52 are hatched. Locations of exons, $\mu$ switch region ($S_\mu$), and the mouse and 5' end of the human enhancer elements (Ⓔ) but not the human pseudo-$J_H$ segments are indicated. The derivation of probes used for hybridization is also shown.

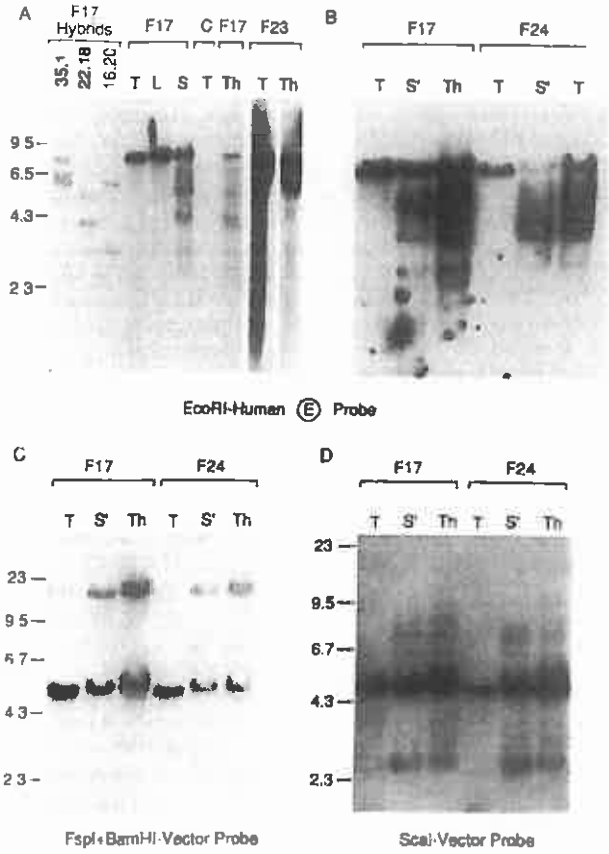



FIG. 2. Southern blot analysis of rearrangements. (*A*) DNA from tail (T), liver (L), spleen (S), and thymus (Th) of a transgenic daughter of F17 as well as tissues from a transgenic and a nontransgenic control (C) offspring of F23 was digested with *Eco*RI and hybridized with 1-kb *Bal* I–*Bgl* II probe from the 5′ end of the human IgH enhancer (see Fig. 1). Also included is DNA from cloned human $\mu$-expressing hybridomas NW499/35.1, NW499/22.18, and NW499/16.20 that were derived from another transgenic F17 offspring. (*B*) Comparison of minilocus rearrangements in tail (T), thymus (Th), and LPS-activated spleen cells (S′) from a daughter of F24 and from a granddaughter of F17; spleen cells were cultured for 3 days in the presence of LPS prior to preparing DNA from viable cells. The DNA was digested and probed as described for *A*. (*C*) Blot to detect rearrangements involving the D segments. DNA from tissues was digested with *Fsp* I + *Bam*HI and hybridized with a vector probe (pUC18 *Fsp* I + *Eco*RI; see Fig. 1). (*D*) Blot to detect rearrangements involving the $V_H$ segments. DNA was digested with *Sca* I and hybridized with the vector probe. Size markers (kb) are provided by a *Hin*dIII digest of $\lambda$ DNA.

observed in pre-B cells (8). The abundance of human $\mu$-chain mRNA in the hyridomas from the transgenic mice is similar if not greater than that found in plasmacytoma transfectants that secrete about 2 mg of chimeric human IgM per liter (Fig. 4*B*). ELISA of culture supernatants from the cloned hybrids that were positive for human $\mu$ chain and that had been obtained in the absence of LPS revealed that about 80% were also positive for mouse $\mu$ chain.

Biosynthetic labeling was carried out to confirm that the hybrids did indeed secrete an antibody with human $\mu$ chains of the expected size; immunoprecipitation was carried out by using an anti-human $\mu$ antiserum (Fig. 5). No band was obtained by using a control cell line transfectant (JW377) that secreted mouse but not human IgM. Four of the transgenic hybrids (lines NW499/16.20, NW499/31.10, NW500/24.2, and NW478/35.5) simultaneously express mouse and human $\mu$ chains, and the immunoprecipitates are likely to contain mixed IgM molecules composed of $\mu$ chains from both

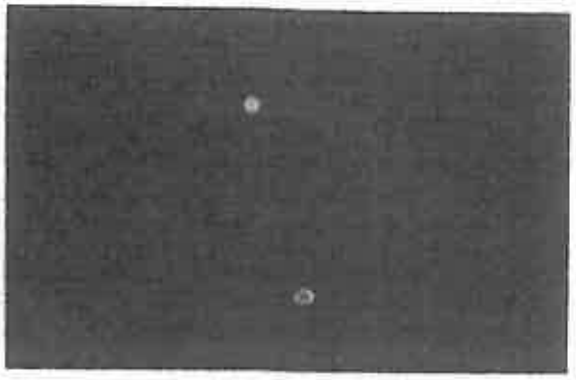


FIG. 3. Cytoplasmic immunofluorescence analysis of spleen cells from an F17 offspring that have been cultured in the presence of LPS for 3 days prior to staining with anti-human $\mu$ antiserum.

species. However, the other three hybrids (NW499/23.18, NW499/35.14, and NW497/21.1) do not make mouse IgM and the heavy-chain band is solely due to human $\mu$ chains.

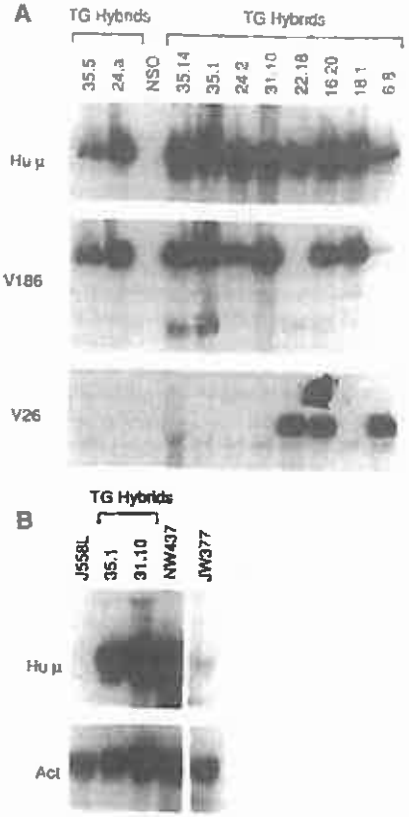



FIG. 4. Northern blot analysis of total cytoplasmic RNA from hybridomas. (*A*) Comparison of $V_H$ usage by different hybrids. The blot was sequentially hybridized to probes specific for human $C_\mu$ (Hu $\mu$), $V_H$186-2, and $V_H$26. All of the hybridomas secreted antibodies containing human $\mu$ chains. Hybridomas NW478/35.5 and NW475/24a were from an F29 descendant, NW497/16.20, NW497/18.1, and NW497/6.8 were from an F19 offspring, and all other hybrids were obtained from the F17 line. (*B*) Comparison of the abundance of human $\mu$-chain mRNA in F17 transgenic hybrids NW499/35.1 and NW499/31.10 with that in a transfectant (NW437) of plasmacytoma NS0 that secretes 2 mg of a chimeric human IgM antibody per liter; plasmacytoma J558L that expresses no IgH chain and its transfectant JW377 that secretes a mouse IgM antibody provide controls. The blot was rehybridized with a mouse $\beta$-actin probe (Act).

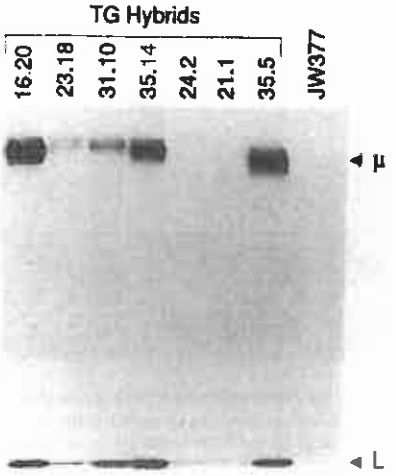



FIG. 5. Analysis by SDS/polyacrylamide gel electrophoresis of biosynthetically labeled antibodies secreted by hybridomas from the transgenic mice; positions of $\mu$ and light (L) chains are marked. Antibodies were purified by immunoprecipitation using anti-human $\mu$ antiserum. The transfectant JW377, which secretes a mouse IgM antibody, provides a control. All of the transgenic hybrids were derived from mice of the F17 line except NW497/21.1 (F19) and NW478/35.5 (F29).

## DISCUSSION

The transgenic mice described here contain B lymphocytes that make antibodies with human IgH chains. Hybridomas from these mice are stable with respect to the secretion of reasonable levels of antibody. It should be possible in the future to make a repertoire of fully human antibodies that, unlike chimeric antibodies or antibodies with grafted hypervariable regions, will have no sequences of foreign origin, by using mice carrying heavy and light chain miniloci. Such mice will not be tolerant to most human determinants and could be used to make human antibodies against human antigens.

The human IgH gene segments introduced into the mouse genome in an unrearranged configuration rearrange in the lymphoid organs; this is similar to what has been found for the light-chain gene segments from other species (17, 18). The pattern of minilocus rearrangements is very similar in spleen and thymus; this contrasts with the endogenous locus, where $V_H$ to $DJ_H$ rearrangements occur only in B cells (8). The proximity of the germ-line $V_H$ segments in the minilocus to the IgH enhancer may account for their rearrangement in T cells (12) and may also render them insensitive to the normal feedback regulation of rearrangement that mediates allelic exclusion (8). Further studies of transgenic miniloci are likely to shed light on the regulation of the rearrangement of V-region gene segments.

The work described here is obviously only a step in the direction of creating the ideal transgenic mouse for the preparation of human monoclonal antibodies. We have concentrated on an IgH locus that includes only a $\mu$-chain C region. The major remaining problems are focused on (*i*) the avoidance of B cells that express transgenic and endogenous IgH chains, (*ii*) the levels of transgene expression, and (*iii*) the effective functional repertoire size.

It would be useful to have transgenic mice that have nonfunctional endogenous immunoglobulin gene loci so that they can only make human antibodies. However, other strategies can also be used to obtain hybridomas that express a transgenic human IgH chain but not an endogenous mouse chain. One could select for loss variants of transgenic hybridomas that no longer retain a functional mouse IgH allele but still express an IgH chain encoded by the minilocus. Alternatively, if the minilocus becomes accessible to recombinase earlier in lymphoid ontogeny than do the endogenous loci, then the feedback regulation of immunoglobulin gene rearrangement could ensure the allelic exclusion of the endogenous mouse genes (8). We do not see such allelic exclusion; either the minilocus itself is not responsive to the feedback regulation of rearrangement or the transgenic IgH chains are expressed at too low a level to prevent rearrangement of the endogenous IgH loci. A high level of expression of the minilocus is also needed to obtain a large amount of antibody from hybridomas as well as probably being necessary to allow the selection by antigen of B cells that make high-affinity antibodies. The amount of transgenic IgM antibody secreted by the hybridomas, though similar to that obtained from transfectants, is about 1/10th that of endogenous mouse IgM. This may be attributable to the existence of as yet unidentified transcription elements in the IgH locus that have not been included in the transgene.

The level of expression and avoidance of cells expressing two heavy chains do not constitute overwhelming obstacles, but the size of the functional repertoire that can be provided by the minilocus is a critical consideration. The diversity of IgH chain expression is normally achieved by choice of the combination of $V_H$, D, and $J_H$ segment used, by N-region insertion, and by somatic mutation (19). We do not know to what extent these various mechanisms contribute to the diversity of IgH chains encoded by the minilocus. Preliminary experiments indicate that it is possible to obtain hybridomas at low frequency from immunized transgenic mice that secrete antibodies composed of human (but not mouse) $\mu$ chains and that are directed against the sheep erythrocyte immunogen. However, useful determination of the functional repertoire size must await transgenic mice bearing heavy-chain and light-chain miniloci.

We thank Sheila Barton and Mike Norris for help with creating transgenic mice, P. Leder and T. Rabbitts for gifts of DNA clones, and Victor Tybulewicz and Cesar Milstein for helpful discussions. M.B. is the recipient of a Leukemia Society of America special fellowship. The work was supported in part by grants from the Medical Research Council.

1. Köhler, G. & Milstein, C. (1975) *Nature (London)* **256**, 495–497.
2. Borrebaeck, C. A. K. (1988) *Immunol. Today* **9**, 355–359.
3. Boulianne, G. L., Hozumi, N. & Shulman, M. J. (1984) *Nature (London)* **312**, 643–646.
4. Morrison, S. L., Johnson, M. J., Herzenberg, L. A. & Oi, V. T. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6851–6855.
5. Neuberger, M. S., Williams, G. T., Mitchell, E. B., Jouhal, S. S., Flanagan, J. G. & Rabbitts, T. H. (1985) *Nature (London)* **314**, 268–270.
6. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. (1986) *Nature (London)* **321**, 522–525.
7. Riechmann, L., Clark, M., Waldmann, H. & Winter, G. (1988) *Nature (London)* **332**, 323–327.
8. Reth, M. & Leclercq, L. (1987) in *Molecular Genetics of Immunoglobulin*, New Comprehensive Biochemistry, eds. Calabi, F. & Neuberger, M. S. (Elsevier, Amsterdam), Vol. 17, pp. 111–134.
9. Ravetch, J. V., Siebenlist, U., Korsmeyer, S., Waldmann, T. & Leder, P. (1981) *Cell* **27**, 583–591.
10. Matthyssens, G. & Rabbitts, T. H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6561–6565.
11. Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. & Baltimore, D. (1981) *Cell* **24**, 625–637.
12. Reik, W., Williams, G. T., Barton, S., Norris, M., Neuberger, M. S. & Surani, M. A. (1987) *Eur. J. Immunol.* **17**, 465–469.
13. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
14. Thomas, P. S. (1983) *Methods Enzymol.* **100**, 255–266.
15. Brüggemann, M., Williams, G. T., Bindon, C. I., Clark, M. R., Walker, M. R., Jefferis, R., Waldmann, H. & Neuberger, M. S. (1987) *J. Exp. Med.* **166**, 1351–1361.

16. Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. (1987) ***Sequences of Proteins of Immunological Interest*** (U.S. Dept. Health and Human Services, National Institutes of Health, Washington, DC).
17. Bucchini, D., Reynaud, C.-A., Ripoche, M.-A., Grimal, H., Jami, J. & Weill, J.-C. (1987) *Nature (London)* **326**, 409–411.
18. Goodhardt, M., Cavelier, P., Akimenko, M. A., Lutfalla, G., Babinet, C. & Rougeon, F. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4229–4233.
19. Tonegawa, S. (1983) *Nature (London)* **302**, 575–581.

# Production of human antibody repertoires in transgenic mice

Marianne Brüggemann* and Michael J Taussig

Transgenic mice have been created that carry human immunoglobulin heavy and light chain genes in germline configuration and that have the corresponding endogenous genes silenced. The transgenes are either minigene constructs or large, almost authentic, transloci on yeast artificial chromosomes and undergo B-cell-specific DNA rearrangement and hypermutation in the mouse lymphoid tissue. Monoclonal antibodies with good affinities for human antigens have been obtained after immunisation. These mice may be a future source of human antibodies for therapy.

**Addresses**
Laboratory of Developmental Immunology and Translocus Therapeutics, Ltd., The Babraham Institute, Babraham, Cambridge CB2 4AT, UK
*e-mail: Marianne.Bruggemann@bbsrc.ac.uk

**Current Opinion in Biotechnology** 1997, 8:456–458

http://biomednet.com/elecref/0958166900800455

© Current Biology Ltd ISSN 0958-1669

**Abbreviations**

| | |
|---|---|
| **C** | constant region |
| **D** | diversity segment |
| **H** | heavy chain |
| **J** | joining segment |
| **L** | light chain |
| **mAbs** | monoclonal antibodies |
| **Mb** | megabase pairs |
| **V** | variable region |
| **YAC** | yeast artificial chromosome |

## Introduction

Research on transgenic animals to identify the mechanism of DNA rearrangement in Ig genes was first described some 10 years ago (see [1••] for review; see [2•] for comparison with the phage display system). Initially, small gene constructs from chicken or rabbit that contained light (L) chain gene segments in germline configuration were used. The approach used to analyse heavy (H) chain rearrangement concentrated on human genes. The first minigene constructs, obtained by subcloning individual segments in an artificially compact configuration, contained human as well as mouse variable region (V), diversity segment (D), joining segment (J) and constant region (C) gene segments. The experiments demonstrated that not only H chain rearrangement of mouse segments could be obtained but also rearrangement and expression of the human genes. This result and the extension in size of H and L chain regions, the largest greater than one million base pairs (>1 Mb) on yeast artificial chromosome (YACs) [3•,4••], showed that human Ig could be expressed in transgenic mice. Although the results of producing human antibodies were encouraging, drawbacks were encountered with suboptimal expression and interference from the endogenous mouse antibody machinery. A step forward in obtaining a predominantly human antibody repertoire in a transgenic mouse was achieved by selective crossing of the transgenic mouse with knock-out strains in which the endogenous IgH and Igκ loci had been silenced. This approach led to human antibody production in the milligram per milliliter range in mouse serum and showed that it was possible, after immunisation, to obtain a variety of high affinity human anti-human monoclonal antibodies (mAbs).

## Introduced human Ig regions

The approaches to obtain and express human Ig transgenes (listed in Table 1) followed broadly two strategies: the assembly of minigene constructs and their injection into mouse eggs, and the transfer of larger, almost authentic, transloci into embryonic stem cells and derivation of mice. As shown in Figure 1 for human IgH, the use of larger regions has the advantages that many more V genes can be included on H and L chain transloci, and also that by adding individual H chain C regions to improve effector functions, human $IgG_1$ or $IgG_2$ can be obtained after class switching [4••,5,6,7•]. Strategies for building large loci in near authentic configurations make use of YACs and the yeast machinery for site-specific extension of core YAC regions. In this way, human IgH, Igκ and Igλ transloci of a few hundred kilobase pairs (kb) up to 1300 kb have been assembled [3•,4••,8•]. With the YAC recombination strategy it was possible to add nearly 100 V gene segments to H and L chain loci, with perhaps half of them regarded as functional and many shown to be expressed. Although the provision of a large number of V genes may have an advantage in primary repertoire formation (V–D–J joining), it is far from clear that high affinity antibodies (dependent on somatic hypermutation) can be produced simply by creating larger V gene transloci.

## Silencing mouse antibody production

A major improvement in human antibody production was achieved when transgenic mice carrying H or L chain transloci were crossed with mice in which the endogenous loci were silenced by homologous integration (reviewed in [9]). Animals with inactivated endogenous IgH and Igκ loci showed a dramatic increase in human antibody production, which made it possible to use such strains for the efficient production of human antibody repertoires essentially without interference by mouse Ig. Gene silencing was achieved either by integration of a selectable marker into an H or L chain C region or by deletion of the J segments and marker integration. These gene manipulation strategies prohibited endogenous rearrangement or C region expression, which for the H chain resulted in B-cell-deficient mice and for κ L chain led to a λ-producing strain. With the integration of human

**Table 1**

**Mice with human Ig translocl in knock-out background: size, variable genes and serum titer.**

| Translocus | Human Ig chains | Size (kb) | Functional V genes | Titer (μg ml$^{-1}$)* | References |
|---|---|---|---|---|---|
| Plasmid-based transgenes | | | | | |
| IgH loci | | | | | |
| HulgH | μ | 25 | 2 | 300 | [10,13,20] |
| HulgH$^{Cos}$ | μ | 100 | 2 | 350 | [2•,10,13] |
| HC1 | μ, γ1 | 61 | 1 | | [22] |
| HC2 | μ, γ1 | 80 | 4 | 100 | [5,6] |
| Igκ loci | | | | | |
| KC1 | κ | 24 | 1 | | [22] |
| KCo4 | κ | 43 | 4 | 100 | [6] |
| YAC-based transgenes | | | | | |
| IgH loci | | | 2 | 0.2 | [23] |
| J1.3 | μ | 85 | 5 | 180 | [1••,16] |
| HulgH | μ | 240 | 5 | 350 | [12] |
| yH1 | μ | 220 | ~40 | 700 (IgM), 600 (IgG) | [4••] |
| yH2 | μ, γ2 | 1020 | | | |
| Igκ loci | | | | | |
| yK1 | κ | 170 | 2 | | [12] |
| Hulgκ YAC | κ | 300 | 2 | | [24] |
| HucosIgκ Yac | κ | 1300 | ~80 | 750 | [3•] |
| KCo5 | κ | 450 | ~26 | | [7•] |
| yK2 | κ | 800 | ~25 | 800 | [4••] |

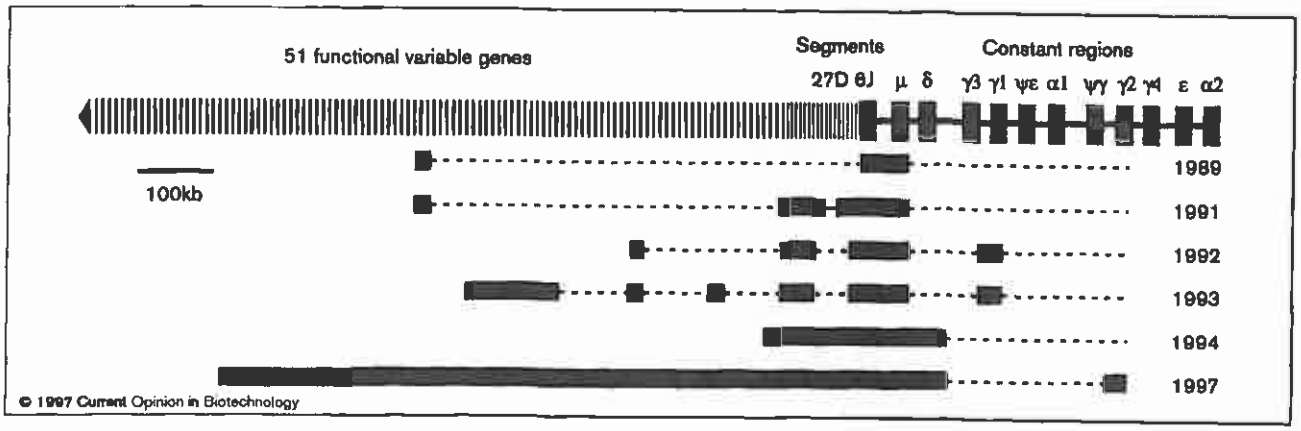
*The approximate (usually mean) titer of serum antibody containing human Ig is shown for the different transgenic lines. Cos, cosmid; H, heavy chain; Hu, human; V, variable region; YAC, yeast artificial chromosome.

translocl, B-cell development and antibody expression were restored to near normal levels. A disadvantage of the gene silencing approaches, however, was that the V and C region gene clusters were still present. This is especially problematic as it was discovered somewhat later that these genes could be used in *trans*-switching or *trans*-splicing events and possibly in *trans*-location events, showing that the remaining mouse genes were utilised [6,10,11]. In that way, relatively rare events produced chimaeric mouse–human Igs. In order to avoid this possibility mouse strains with deleted H and L chain loci need to be established.

## B-cell development and transgene expression

In the initial experiments, human transgene expression was compared with endogenous antibody production and showed, at best, that only a few percent of B cells expressed human Ig chains and that human Ig levels in serum were rarely more than a few micrograms per milliliter [1••]. The rather low level of human antibody

**Figure 1**




A schematic view of the human IgH locus and the derived translocus strains showing increasing transgene regions which were assembled and expressed between 1989 and 1997 (reviewed in [1••,4••]). The human IgH locus is 1.5 Mb in size with a region of ~1000 Kb accommodating ~95 $V_H$ genes, of which 51 are regarded as functional [18,19]. The D–J region, accommodating 27 D and 6 J segments, is ~100 kb and the region accommodating the nine functional constant regions is ~350 kb in size. The dotted lines indicate transgene assembly.

production in normal transgenic mice was interpreted as being due to inefficient competition of the human genes with the mouse antibody machinery. An encouraging result was that human antibody production in hybridoma cells could be as efficient as the secretion of mouse Ig. Extensive analyses of strains producing human H and κ L chains crossed into the endogenous mouse knock-out background revealed that B-cell development could be rescued at up to about half the normal levels. This is reflected in a somewhat reduced spleen size in the transgenic H chain knock-out animals and in a lower number of human IgM+ cells (compared with mouse IgM+ cells in normal mice) in spleen, Peyer's patches and bone marrow [6,7•,10,12]. Fluorescence activated cell sorter (FACS) analysis highlighted the efficient transgene expression in B220+ B cells, where a high proportion expressed human H and κ L chain simultaneously [4••,6,12]. Although there is a lack of comparative analysis it seems that the transloci mice harbouring larger transgenic regions appear to perform better than the human minilocus strains. This holds true for spleen size and B-cell development in general and also for human Ig levels. A reason for this could be that as yet unknown transgene activation signals are accommodated on the large transferred regions.

## Antibody responses and repertoire

Human antibody producing mice with four features—human IgH/Igκ transloci and endogenous IgH/Igκ knock-outs—are particularly useful for evaluating immune responses [4••,6,7•,12]. The immunisation and antibody production procedures are identical to those used in establishing antibody responses in normal mice and high affinity human mAbs have been obtained. As illustrated in Table 2, an impressive range of IgM and IgG antibodies specific for cells, proteins or haptens has been produced; particularly noteworthy is the production of human mAbs against human antigens. There appears to be an advantage of human IgG antibodies over IgM antibodies in that the affinity of IgG may be superior; this advantage could indicate that IgGs are a more effective substrate for somatic hypermutation and selection [1••,4••,6,10]. Obtaining high affinity human antibodies implies that the mouse recombination and hypermutation mechanism can be fully operative on the introduced transloci. There is a high expectation that mice with these transloci will be suitable for the production of human mAbs for clinical diagnostic and therapeutic applications. For example, human anti-CD4 antibodies, with their reduced immunogenicity and longer half-life compared to conventional rodent antibodies, may be invaluable as immunosuppressants in the treatment of autoimmune disease and graft rejection [6,7•].

Critical issues for the production of human mAbs is the diversity generated at the junctions—with the terminal deoxynucleotide transferase being particularly active during heavy chain recombination—and the extent of somatic hypermutation obtained in the human V genes.

**Table 2**

**Antigen-specific monoclonal antibodies from mice with human Ig transloci.**

| Translocus | Antigen | Human Ig chains* | Affinity | Reference |
|---|---|---|---|---|
| HuIgH | Phenyloxazolone | μ | 60 and 160 nM | [10] |
| | Phycoerythrin | | | |
| | Lysozyme | | | |
| | Human lymphoma cells | | | |
| HuIgH$^{Cos}$ | Phenyloxazolone | μ | | [10] |
| | Lysozyme | | | |
| | Human lymphoma cells | | | |
| γH1/γK1 | Tetanus toxin | μ+κ | | [12] |
| HC2/KCo4 | Human CD4 | γ1+κ | 12 nM | [6] |
| HC2/KCo5 | Human CD4 | γ1+κ | 0.03–0.4 nM | [7•] |
| γH2/γK2 | Human IL8 | γ2+κ | 0.16–0.8 nM | [4••] |
| | Human EGFR | | 0.03–0.4 nM | |
| | Human TNFα | | 0.2–0.8 nM | |

*The HuIgH and HuIgH$^{Cos}$ transgenes were analysed in mice with disrupted endogenous IgH locus but capable of producing endogenous Igκ chains. The other mouse lines contained transloci encoding both human IgH and Igκ on mouse backgrounds disrupted for the corresponding endogenous loci. The mice that produce human IgG also produce human IgM but only the results for the specific IgG antibodies are shown. Cos, cosmid; EGFR, epidermal growth factor receptor; H, heavy chain; Hu, human; IL, interleukin.

In most of the transloci mice only a limited number of V genes are provided; nevertheless, extensive diversity is generated for the heavy chain by use of the D segment and the imprecise nature of the rearrangements. This diversity results from the addition of non-germline encoded nucleotides in the conventional V–D–J joining process [4••,5,10,13–15]. Hypermutation has been described for minigene constructs as well as for large YAC-based transloci, but the extent of mutational changes appears to be less than that observed in endogenous mouse genes [5,10,16]. One reason for this could be that the extent of hypermutation is directly linked to the size of the transferred locus. Alternatively, the addition of γ constant regions to allow μ to γ switching may favourably bias hypermutation (reviewed in [17]). Interestingly, the fact that mouse H and L chain loci accommodate important 3′ regulatory sequences suggests that the addition of sequences or 3′ extension of translocus regions may prove critical to achieving optimal human antibody expression in mice.

## Conclusions

Transgenic mice carrying human H and L chain regions in germline configuration over 1 Mb in size have been produced. The human Ig genes are rearranged, expressed and hypermutated in the mouse background. These processes are not well understood and they operate with reduced efficiency in 'normal' transgenic mice. When human H and L chain transloci are expressed in endogenous IgH and Igκ knock-out strains, however, the efficiency is dramatically improved. In immunisation procedures, a variety of specific, high-affinity human mAbs has been obtained. This implies that antibodies from the human transloci Ig mice will, in the future, replace the use of rodent antibodies for clinical therapeutic applications. Focusing on future approaches to secure optimal expression of the human genes, the aim will be to transfer authentic human loci into an animal in which the endogenous loci have been deleted.

## Acknowledgements

We are grateful for the generous support provided by the Babraham Institute and for grant support for Marianne Brüggemann's laboratory from the Medical Research Council, the Cancer Research Campaign and the Leverhulme Trust.

## References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. •• Brüggemann M, Neuberger MS: **Strategies for expressing human antibody repertoires in transgenic mice.** *Immunol Today* 1996, 17:391-397.

A summary of the events and results in producing human antibody repertoires in transgenic mice with maps of the translocus regions.

2. • Neuberger M, Brüggemann M: **Mice perform a human repertoire.** *Nature* 1997, 386:25-26.

A comparison of two equally important strategies to produce human antibodies: transgenic mice, *in vivo*, and the phage display system, *in vitro*.

3. • Zou X, Xian J, Davies NP, Popov AV, Brüggemann M: **Dominant expression of a 1.3 Mb human Igκ locus replacing mouse light chain production.** *FASEB J* 1996, 10:1227-1232.

This paper shows for the first time that complete megabase size regions can be transferred and dominantly expressed in transgenic mice. The expression of human Igκ transgenes can lead to a down regulation of mouse Igλ light chain expression.

4. •• Mendez MJ, Green LL, Corvalan JRF, Jia X-C, Maynard-Currie CE, Yang X-D, Gallo ML, Louis DM, Lee DV, Erickson KL *et al.*: **Functional transplant of megabase human immunoglobulin loci recapitulates human antibody response in mice.** *Nat Genet* 1997, 15:146-156.

Large regions perform better: with 66 $V_H$ and 32 $V_K$ segments, human antibody expression in endogenous knock-out mice closely resembles that seen in humans with respect to rearrangement, assembly and repertoire. A variety of specific human antibodies with high affinity have been produced.

5. Taylor LD, Carmack CE, Huszar D, Higgins KM, Mashayekh R, Sequar G, Schramm SR, Kuo C-C, O'Donnell SL, Kay RM *et al.*: Human immunoglobulin transgenes undergo rearrangement, somatic mutation and class switching in mice that lack endogenous IgM. *Int Immunol* 1993, 6:579-591.

6. Lonberg N, Taylor LD, Harding FA, Trounstine M, Higgins KM, Schramm SR, Kuo C-C, Mashayekh R, Wymore K, McCabe JG *et al.*: **Antigen-specific human antibodies from mice** comprising four distinct genetic modifications. *Nature* 1994, 368:856-859.

7. • Fishwild DM, O'Donnell SL, Bengoechea T, Hudson DV, Harding F, Bernhard SL, Jones D, Kay RM, Higgins KM, Schramm SR, Lonberg N: **High-avidity human Igκ monoclonal antibodies from a novel strain of minilocus transgenic mice.** *Nat Biotechnol* 1996, 14:845-851.

Production of high affinity human anti-CD4 antibodies after class switching.

8. • Popov AV, Bützler C, Frippiat J-P, Lefranc MP, Brüggemann M: **Assembly and extension of yeast artificial chromosomes to build up a large locus.** *Gene* 1996, 177:195-205.

Of general interest for researchers who wish to construct or extend large regions on YACs. The method provides a positive/negative selection strategy to obtain authentic regions by site-specific integration.

9. Jakobovits A: **Production of fully human antibodies by transgenic mice.** *Curr Opin Biotechnol* 1995, 6:561-566.

10. Wagner SD, Popov AV, Davies SL, Xian J, Neuberger MS, Brüggemann M: **The diversity of antigen-specific monoclonal antibodies from transgenic mice bearing** human immunoglobulin gene mini loci. *Eur J Immunol* 1994, 24:2672-2681.

11. Knight KL, Kingzette M, Crane MA, Zhai S-K: **Transchromosomally derived Ig** heavy chains. *J Immunol* 1995, 155:684-691.

12. Green LL, Hardy MC, Maynard-Currie CE, Tsuda H, Louie DM, Mendez MJ, Abderrahim H, Noguchi M, Smith DH, Zeng Y, David *et al.*: **Antigen-specific human** monoclonal **antibodies from mice engineered with human Ig heavy and light chain YACs.** *Nat Genet* 1994, 7:13-21.

13. Wagner SD, Williams GT, Larson T, Neuberger MS, Kitamura D, Rajewsky K, Xian J, Brüggemann M: **Antibodies generated from human** immunoglobulin **miniloci in transgenic mice.** *Nucleic Acids Res* 1994, 22:1389-1393.

14. Tuaillon N, Miller AB, Tucker PW, Capra JD: Analysis **of directed and inverted DJH rearrangements** in a human Ig heavy chain **transgenic minilocus.** *J Immunol* 1995, 154:6453-6465.

15. Harmer U, Mageed RA, Kaminski A, Charles P, Brüggemann M, Mackworth-Young CG: Chimaeric monoclonal **antibodies encoded by the human $V_H26$ gene from naive transgenic mice display a wide range of antigen binding specificities.** *Immunology* 1996, 88:174-182.

16. Wagner SD, Gross G, Cook GP, Davies SL, Neuberger MS: **Antibody expression** from **the core region of the human IgH locus reconstructed in transgenic mice using bacteriophage P1 clones.** *Genomics* 1996, 35:405-414.

17. Harding FA, Lonberg N: **Class** switching **in human immunoglobulin transgenic mice.** *Ann NY Acad Sci* 1995, 764:536-547.

18. Hofker MH, Walter MA, Cox DW: **Complete** physical **map of the human** immunoglobulin **heavy chain** constant region **gene complex.** *Proc Natl Acad Sci USA* 1989, 86:5567-5571.

19. Cook GP, Tomlinson IM: **The human immunoglobulin $V_H$ repertoire.** *Immunol Today* 1995, 5:237-242.

20. Brüggemann M, Caskey HM, Teale C, Waldmann H, Williams GT, Surani MA, Neuberger MS: A repertoire of monoclonal **antibodies with human heavy-chains from transgenic mice.** *Proc Natl Acad Sci USA* 1989, 86:6709-6713.

21. Brüggemann M, Spicer C, Buluwela L, Rosewell I, Barton S, Surani MA, Rabbitts TH: Human antibody production **in transgenic mice:** expression **from 100 kb of the human IgH locus.** *Eur J Immunol* 1991, 21:1323-1326.

22. Taylor LD, Carmack CE, Schramm SR, Mashayekh R, Higgins KM, Kuo C-C, Woodhouse C, Kay RM, Lonberg N: **A transgenic mouse that expresses a diversity of** human **sequence heavy and light chain immunoglobulins.** *Nucleic Acids Res* 1992, 20:6287-6295.

23. Choi TK, Hollenbach PW, Pearson BE, Ueda RM, Weddell GN, Kunahara CG, Woodhouse CS, Kay RM, Loring JF: **Transgenic mice containing a human heavy chain immunoglobulin gene fragment cloned in a yeast** artificial **chromosome.** *Nat Genet* 1993, 4:117-123.

24. Davies NP, Rosewell IR, Richardson JC, Cook GP, Neuberger MS, Brownstein BH, Norris ML, Brüggemann M: **Creation of mice expressing human antibody light chains by introduction of a yeast artificial chromosome containing the core region of the human immunoglobulin κ locus.** *Bio-Technology* 1993, 11:911-914.

217

© 1991 Oxford University Press

Nucleic Acids Research, Vol. 19, No. 15 4133–4137

# Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains

Hennie R.Hoogenboom[1], Andrew D.Griffiths[1], Kevin S.Johnson[2], David J.Chiswell[2], Peter Hudson[4] and Greg Winter[1, 3*]

[1]Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, [2]Cambridge Antibody Technology Ltd, Daly Research Laboratories, Babraham Hall, Cambridge CB2 4AT, [3]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK and [4]CSIRO Division of Biomolecular Engineering, 343 Royal Parade, Parkville, Victoria 3052, Australia

Received May 12, 1991; Revised and Accepted June 17, 1991

## ABSTRACT

**The display of proteins on the surface of phage offers a powerful means of selecting for rare genes encoding proteins with binding activities. Recently we found that antibody heavy and light chain variable (V) domains fused as a single polypeptide chain to a minor coat protein of filamentous phage fd, could be enriched by successive rounds of phage growth and panning with antigen. This allows the selection of antigen-binding domains directly from diverse libraries of V-genes. Now we show that heterodimeric Fab fragments can be assembled on the surface of the phage by linking one chain to the phage coat protein, and secreting the other into the bacterial periplasm. Furthermore by introducing an amber mutation between the antibody chain and the coat protein, we can either display the antibody on phage using *supE* strains of bacteria, or produce soluble Fab fragment using non-suppressor strains. The use of Fab fragments may offer advantages over single chain Fv fragments for construction of combinatorial libraries.**

## INTRODUCTION

The fd and M13 phages are members of the Ff filamentous, single-stranded DNA phages that infect male *Escherichia coli* cells. Adsorption to the host sex pili is mediated by a minor coat protein (gene 3 protein: g3p) displayed at one tip of the virion. Each virion appears to contain three copies of g3p (1). The amino-terminal domains, which form knob-like structures extending from the phage particle, are responsible for binding the phage to the F-pilus, whilst the C-terminal region is anchored in the phage coat (2). Peptides have been displayed at the surface of phage by fusion to the N-terminus of g3p (3), and phage with binding activities isolated from random peptide libraries after rounds of growth and panning (4, 5, 6).

Folded proteins have also been displayed as fusions with g3p (7, 8). Thus, genes encoding binding activities can be selected directly by panning of the recombinant phage. For example, rare antigen-binding phage have been selected from diverse repertoires of antibody genes from immunised animals and the stronger binders enriched by panning (9). In principle the use of phage display libraries from naive antibody genes might allow antibodies to be made entirely *in vitro*: higher affinity variants might be made by random mutation of the genes, so by-passing human or animal immunisation (10).

Antibodies are heterodimeric proteins: the heavy and light chain variable (V) domains combine to make the antigen-binding site. For display on the surface of phage we previously linked both domains into a single polypeptide (7, 9), via a flexible polypeptide chain as a single chain Fv fragment (scFv) (11, 12). We now sought to display the antibodies as heterodimeric Fab fragments, to allow the two chains to be readily reassorted in combinatorial libraries (see Discussion). Since in phage assembly, the coat proteins are exported into the bacterial periplasm, and antibody heavy and light chains can be directed into the bacterial periplasm by a signal sequence (13, 14), we decided to link either the heavy or light chain to the phage g3p with a signal sequence, and co-express the complementary chain into the periplasm. Since antibody heavy and light chains fragments are assembled together in the bacterial periplasm (14), the pairings of heavy and light chains of each bacterium should be retained on the surface of the phage.

## MATERIALS AND METHODS

### E.coli strains

TG1 :K12, Δ(*lac-pro*), *sup*E, *thi*, *hsdD5*/F'*traD*36, *proA*$^+$B$^+$, *lacI*$^q$, *lacZ*ΔM15

HB2151 :K12, *ara*, Δ(*lac-pro*), *thi*/F'*proA*$^+$*B*$^+$, *lacI*$^q$Z*ΔM*15

* To whom correspondence should be addressed

### Oligonucleotides

G3FUFO, 5'-CAG TGA ATT CTT ATT AAG ACT CCT TA-T TAC GCA GTA TGT TAG C; G3FUBA, 5'-TGC GAA G-CT TTG GAG CCT TTT TTT TTG GAG ATT TTC AAC G; VH1BACKAPA, 5'-CAT GAC CAC AGT GCA CAG GT(C/G) (A/C)A(A/G) CTG CAG (C/G)AG TC(A/T) GG; VH1BACK-SFI15, 5'-CAT GCC ATG ACT CGC GGC CCA GCC GG-C CAT GGC C(C/G)A GGT (C/G)(A/C)A (A/G)CT GCA G(C/G)A GTC (A/T)GG; FABNOTFOH, 5'-CCA CGA TTC TGC GGC CGC TGA AGA TTT GGG CTC AAC TTT CTT GTC GAC; FABNOTFOK, 5'-CCA CGA TTC TGC GGC C-GC TGA CTC TCC GCG GTT GAA GCT CTT TGT GAC; MVKBAAPA, 5'-CAC AGT GCA CTC GAC ATT GAG C-TC ACC CAG TCT CCA; MVKBASFI, 5'-CAT GAC CAC GCG GCC CAG CCG GCC ATG GCC GAC ATT GAG CTC ACC CAG TCT CCA; VK3F2NOT, 5'-TTC TGC GGC CGC CCG TTT CAG CTC GAG CTT GGT CCC. Restriction sites are underlined.

### Vector constructions

The two vectors used for expression are depicted in Figure 1. The vector fd-tet-DOG1 (Figure 1A) (9) is derived from fd-CAT1 (7) in turn derived from fd-tet (15). fd-tet-DOG1 has *Apa*LI and *Not*I restriction sites (Figure 1C) for cloning antibody genes as fusions to the N-terminus of g3p. The phagemid pHEN1 (Figure 1B) is a derivative of pUC119 (16): the coding region of g3p from fd-tet-DOG1, including signal peptide and cloning sites, was amplified by PCR, using primers G3FUFO and G3FUBA (which contain *Eco*RI and *Hin*dIII sites respectively), and cloned as a *Hin*dIII-*Eco*RI fragment into pUC119. The *Hin*dIII-*Not*I fragment encoding the g3p signal sequence was then replaced by a pelB signal peptide (13) with an internal *Sfi*I site, allowing antibody genes to be cloned as *Sfi*I-*Not*I fragments. A peptide tag (17, 18) was introduced directly after the *Not*I site by cloning an oligonucleotide cassette, and followed by an amber codon introduced by site-directed mutagenesis (19) (Figure 1D).

A range of constructs (see Figure 2) were made from a clone (essentially construct II in pUC19) designed for expression in bacteria of a soluble Fab fragment (13) from the mouse anti-phOx (2-phenyl-5-oxazolone) antibody NQ10.12.5 (20). In construct II, the V-regions are derived from NQ10.12.5 and attached to human C$\kappa$ and CH1 ($\gamma$1 isotype) constant domains: the C-terminal cysteine residues, which normally form a covalent link between light and heavy antibody chains, have been deleted from both the constant domains (M.Dreher and E.S.Ward unpublished). To clone heavy and light chain genes together as Fab fragments (construct II) or as separate chains (constructs III and IV) for phage display, DNA was amplified from construct II by PCR to introduce a *Not*I restriction site at the 3' end, and at the 5' end either an *Apa*LI site (for cloning into fd-tet-DOG1) or *Sfi*I site (for cloning into pHEN1). The primers FABNOTFOK with VH1BACKAPA (or VH1BACKSFI15) were used for PCR amplificatiion of genes encoding Fab fragments (construct II), the primers FABNOTFOH with VH1BACKAPA (or VH1BA-CKSFI15) for heavy chains (construct III), and the primers FABNOTFOK and MVKBAAPA (or MVKBASFI) for light chains (construct IV).

The single-chain Fv version of NQ10.12.5 (construct I) has the heavy (VH) and light chain (V$\kappa$) variable domains joined by a flexible linker $(Gly_4Ser)_3$ (12) and was constructed from construct II by 'splicing by overlap extension' (21) as in Refs. 9 and 22. The assembled genes were reamplified with primers VK3F2NOT and VH1BACKAPA (or VH1BACKSFI15) to append restriction sites for cloning into fd-tet-DOG1 (*Apa*LI-*Not*I) or pHEN1 (*Sfi*I-*Not*I).

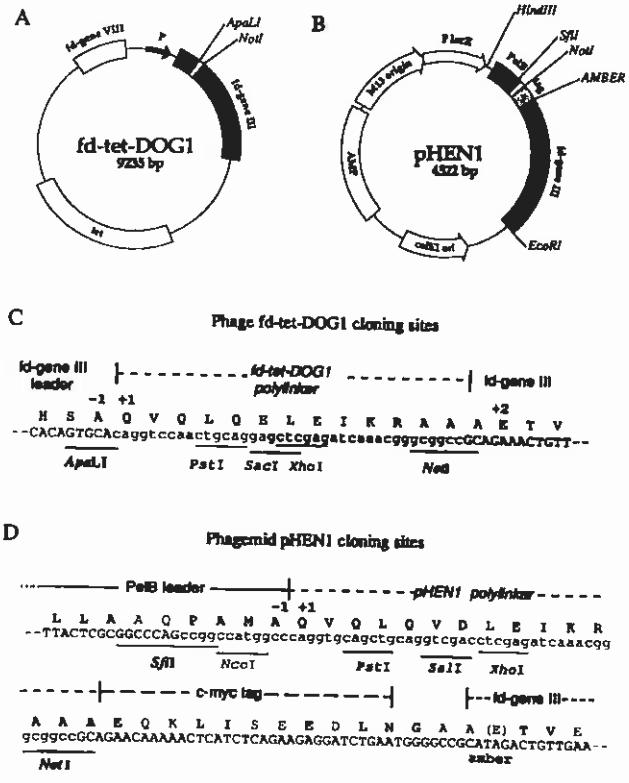



**Figure 1.** Phage and phagemid vectors used for display of antibody fragments on the surface of phage. (A) Structure of vector fd-tet-DOG1. (B) Structure of vector pHEN1. (C) Sequence of fd-tet-DOG1 cloning sites. (D) Sequence of pHEN1 cloning sites.

**Table 1.** Overview of phOx-BSA ELISA results of phage and phagemid constructions.

| | Phage/Phagemid† | Helper phage | Binding to phOx* | Chain(s) displayed# | Chain as gene III fusion# | Soluble chain(s)§ |
|---|---|---|---|---|---|---|
| A | fd-tet-DOG1 | | non binding | none | | |
| | fd-tet-DOG1-I | | binding | scFv | scFv | |
| | fd-tet-DOG1-II | | binding | Fab | light chain | heavy chain |
| | pHEN1 | VCSM13 | non binding | none | | |
| | pHEN1-I | VCSM13 | binding | scFv | scFv | |
| | pHEN1-II | VCSM13 | binding | Fab | light chain | heavy chain |
| B | pHEN1-I (HB2151) | | binding | | | scFv |
| | pHEN1-II (HB2151) | | binding | | | Fab |
| C | fd-tet-DOG1-III | | non binding | heavy chain | heavy chain | |
| | fd-tet-DOG1-IV | | non binding | light chain | light chain | |
| | pHEN1-III (HB2151) | VCSM13 | non binding | none | | heavy chain |
| | pHEN1-III (HB2151) | fd-tet-DOG1-IV | binding | Fab | light chain | heavy chain |
| | pHEN1-IV (HB2151) | VCSM13 | non binding | none | | light chain |
| | pHEN1-IV (HB2151) | fd-tet-DOG1-III | binding | Fab | heavy chain | light chain |

*Phage were considered to be 'binding' if $OD_{405}$ of sample was at least 10-fold greater than background in ELISA; †*E.coli* TG1 was used for the growth of the phage unless the use of *E.coli* HB2151 is specifically indicated; #Information deduced from genetic structure and in accordance with binding data; §Result confirmed experimentally by Western blot (for Fab, see Figure 3).

### Rescue of phage and phagemid particles

Constructs I–IV (Figure 2) were introduced into both fd-tet-DOG1 and pHEN1. Phage fd-tet-DOG1 (and fd-tet-DOG1-I, II, III or IV) was taken from the supernatant of infected *E.coli* TG1 after shaking at 37°C overnight in 2×TY medium (23) with 12.5 μg/ml tetracycline, and used directly in ELISA. Phagemid pHEN1 (and pHEN1-I and II) in *E.coli* TG1 (*supE*) were grown overnight in 2 ml 2×TY medium, 100 μg/ml ampicillin, and 1% glucose (without glucose, expression of g3p prevents later superinfection by helper phage). 10 μl of the overnight culture was used to inoculate 2 ml of 2×TY medium, 100 μg/ml ampicillin, 1% glucose, and shaken at 37°C for 1 hour. The cells were washed and resuspended in 2×TY, 100 μg/ml ampicillin, and phagemid particles rescued by adding 2 μl ($10^8$ pfu) VCSM13 helper phage (Stratagene). After growth for one hour, 4 μl kanamycin (25 mg/ml) was added, and the culture grown overnight. The phagemid particles were concentrated 10-fold for ELISA by precipitation with polyethylene glycol (as in Ref. 7).

For assembly of heavy and light chains expressed from different vectors, phagemid (pHEN1-III or IV) was grown in *E.coli* HB2151 (a non-suppressor strain) to allow production of soluble chains, and rescued as above except using helper phage with partner chains as fusions to g3p ($10^9$ TU fd-tet-DOG1-IV or III respectively) and 2 μl tetracycline (12.5 mg/ml) in place of kanamycin.

### Induction of soluble scFv and Fab

*E.coli* HB2151 was infected with pHEN phagemid (pHEN1-I or II), and plated on TYE (23), 100 μg/ml ampicillin plates. Colonies were shaken at 37°C in 2×TY medium, 100 μg/ml ampicillin, 1% glucose to $OD_{550}$ = 0.5 to 1.0. Cells were pelleted, washed once in 2×TY medium, resuspended in medium with 100 μg/ml ampicillin, 1 mM isopropyl β-D-thiogalactoside (IPTG), and grown for a further 16 hours (18). Cells were pelleted and the supernatant, containing the secreted chains, used directly in ELISA.

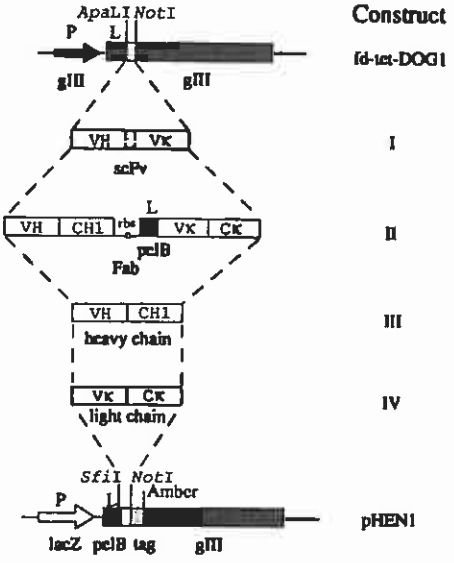



**Figure 2.** The antibody constructs cloned into fd-tet-DOG1 and pHEN1 for display on the surface of phage. Constructs I, II, III, and IV were cloned into both fd-tet-DOG1 (as *ApaLI-NotI* fragments) and pHEN1 (as *SfiI-NotI* fragments). All the constructs contained the heavy chain (VH) and light chain (V*x*) variable regions of the mouse anti-phOx antibody NQ10.12.5. The constant domains were human C*x* and CH1 (γ1 isotype).

### ELISA

Detection of phage binding to 2-phenyl-5-oxazolone (phOx) was performed as in Ref. 9: 96-well plates were coated with 10 μg/ml phOx-BSA or 10 μg/ml BSA in PBS overnight at room temperature, and blocked with PBS containing 2% skimmed milk powder. Phage(mid) supernatant (50 μl) mixed with 50 μl PBS containing 4% skimmed milk powder was added to the wells and assayed as described in Ref. 7. To detect binding of soluble scFv or Fab fragments secreted from pHEN1, the c-myc peptide tag was detected as in Refs. 17 and 18.

### Western blot

Western blots were essentially as in Ref. 24. For detection of secreted proteins, 10 μl supernatant of induced cultures were subjected to SDS–PAGE (25) and proteins transferred by electroblotting to Immobilon-P (Millipore). Soluble heavy and light chain were, detected with goat polyclonal anti-human Fab antiserum (Sigma) and peroxidase conjugated rabbit anti-goat immunoglobulin (Sigma), each at a dilution of 1:1000. The tagged V*x* domain was detected with 9E10 antibody (1:1000) and peroxidase conjugated goat anti-mouse immunoglobulin (Fc specific) (1:1000) (Sigma) as in Ref. 18 or with a peroxidase labeled anti-human C*x* antiserum (Dako). 3,3′-diaminobenzidine (DAB; Sigma) was used as peroxidase substrate (26).

## RESULTS

### Vectors for display of scFv and Fab fragments on phage

We have used phage and phagemid vectors to display antibody fragments on the surface of filamentous phage. The phage vector, fd-tet-DOG1 (Ref. 9; Figure 1A and C) is based on the vector fd-tet (15) and has restriction sites (*Apa*LI and *Not*I) for cloning antibody (or other protein) genes for expression as fusions to the N-terminus of the phage coat protein g3p. Transcription of the antibody-g3p fusions in fd-tet-DOG1 is driven from the gene III promoter and the fusion protein targetted to the periplasm by

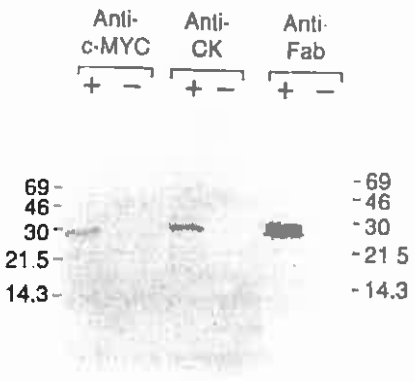



**Figure 3.** Western blot of supernatant taken from pHEN1-II (+) or pHEN1 (−) cultures in *E.coli* HB2151, showing secretion of Fab fragment from pHEN1-II only. The anti-human Fab detects both heavy and light chain. Due to the attached c-myc tag, the light chain, highlighted by both anti-c-myc tag and anti-human C*x* antisera, is slightly larger (calculated $M_r$ 24625) than the heavy chain (calculated $M_r$ 23145).

means of the g3p leader (Figure 1C). Fab and scFv fragments of NQ10.12.5 cloned into fd-tet-DOG1 for display were shown to bind to phOx-BSA (but not BSA) by ELISA (Table 1A).

The phagemid vector, pHEN1 (Figure 1B and D), is based on pUC119 (16) and contains restriction sites (*SfiI* and *NotI*) for cloning the fusion proteins. Here the transcription of antibody-g3p fusions is driven from the inducible lacZ promoter and the fusion protein targetted to the periplasm by means of the pelB leader (13). Phagemid was rescued with VCSM13 helper phage in 2×TY medium containing no glucose or IPTG: under these conditions the natural leakiness of the promoter provided sufficient expression of antibody-g3p. Increased expression by IPTG induction led to cell death due to the toxicity of the fusion protein. Fab and scFv fragments of NQ10.12.5 cloned into pHEN1 for display were shown to bind to phOx-BSA (but not BSA) by ELISA (Table 1A).

The phagemid pHEN1 has the advantage over phage fd-tet-DOG1 in that antibody can be produced either for phage display (by growth in *sup*E strains of *E.coli*) or as a tagged soluble fragment (by growth in non-suppressor strains), as we introduced a peptide tag (17, 18) and amber codon between the antibody and g3p. Secretion of soluble Fab fragments from pHEN1-II or scFv fragments from pHEN1-I was demonstrated after growth in *E.coli* HB2151 and induction with IPTG using Western blots (Figure 3). With the scFv, the fragments were detected using the 9E10 anti-myc tag antibody (17, 18) (data not shown). With the Fab, only the light chain was detected by 9E10 (or anti-human Cκ) antibody, as expected, while the anti-human Fab antiserum detected both heavy and light chains. Binding of the soluble scFv and Fab fragments to phOx-BSA (but not to BSA) was also demonstrated by ELISA (Table 1B). Thus scFv and Fab fragments can be displayed or secreted as soluble fragments from the same phagemid vector.

### Separate vectors to encode Fab heavy and light chains

The heavy and light chains of Fab fragments can be encoded together in the same vector (see above) or in different vectors. To demonstrate this we cloned the heavy chain (construct III) into pHEN1 (to provide soluble fragments) and the light chain (construct IV) into fd-tet-DOG1 (to make the fusion with g3p). The phagemid pHEN1-III, grown in *E.coli* HB2151 (non-suppressor) was rescued with fd-tet-DOG1-IV phage, and phage(mid) shown to bind to phOx-BSA, but not to BSA (Table 1C). This demonstrates that soluble heavy chain is correctly associating with the light chain anchored to the g3p, since neither heavy chain nor light chain alone bind antigen (Table 1C).

Similar results were obtained in the reverse experiment (with phagemid pHEN1-IV and fd-tet-DOG1-III phage) in which the light chain was produced as a soluble molecule and the heavy chain anchored to g3p (Table 1C). Hence a Fab fragment is assembled on the surface of phage by fusion of either heavy or light chain to g3p, provided the other chain is secreted using the same or another vector (Figure 4).

The resulting phage population is a mixture of fd phage and rescued phagemid. The ratio of the two types of particle was assessed by infecting log phase *E.coli* TG1 and plating on TYE plates with either 15 μg/ml tetracycline (to select for fd-tet-DOG1) or 100 μg/ml ampicillin (to select for pHEN1). The titre of fd-tet-DOG1 phage was $5\times10^{11}$ TU/ml and the titre of pHEN1 $2\times10^{10}$ TU/ml, indicating a packaging ratio of 25 phage per phagemid.

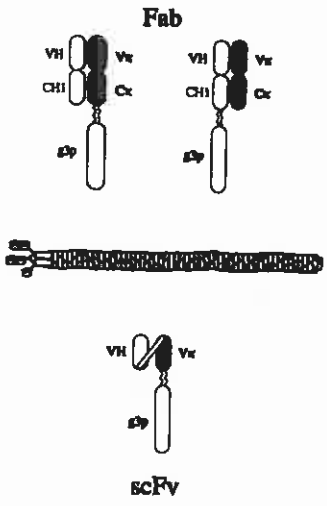



Figure 4. Three ways of displaying antibody fragments on the surface of phage by fusion to g3p.

## DISCUSSION

The antigen-binding site of an antibody is formed by two domains, the heavy (VH) and light (VL) chain variable domains, which are on different polypeptide chains. These two variable domains can be expressed on the same polypeptide if they are joined artificially by a flexible linker (11, 12) to form single-chain Fv fragments (scFv). Previously we demonstrated that scFv antibody fragments can be displayed on the surface of fd phage by fusion to the amino terminus of g3p, that these 'phage antibodies' bind antigen (7), and that rare phage can be selected from large libraries (7, 9). However, scFv fragments often have affinities lower than the parent antibody (27).

Here we have demonstrated an alternative strategy involving display of the heterodimeric antibody Fab fragments on the surface of phage. One of the chains is fused to g3p and the other is secreted in soluble form into the periplasmic space of the *E.coli* where it associates non-covalently with the g3p fusion, and binds specifically to antigen. Either the light or heavy chain can be fused to the g3p: they are displayed on the phage as Fab fragments and bind antigen (Figure 4). We have described both phage and phagemid vectors for surface display: phagemids are probably superior to phage vectors for creation of large phage display libraries in view of their higher transfection efficiencies — two to three orders of magnitude higher, allowing larger libraries to be constructed. Our phagemid vector, pHEN1 also allows the expression of soluble Fab fragments in non-suppressor *E.coli*.

We have also demonstrated that heavy and light chains encoded on the same vector (construct II), or on different vectors (constructs III and IV) can be displayed as Fab fragments. This offers two distinct ways of making random combinatorial (28) libraries for display. Libraries of heavy and light chain genes, amplified by PCR, could be randomly linked by a 'PCR assembly' process (9, 22) based on 'splicing by overlap extension' (21), cloned into phage(mid) display vectors and expressed from the same promoter as part of the same transcript (construct II) as above, or indeed from different promoters as separate transcripts. Here the phage(mid) vector encodes and displays both chains. For a combinatorial library of $10^7$ heavy chains and $10^7$

light chains, the potential diversity of displayed Fab fragments ($10^{14}$) is limited by the transfection efficiency of bacterial cells by the vector (about $10^9$ clones per $\mu$g cut and ligated plasmid at best) (29).

Alternatively, libraries of heavy and light chains could be cloned into different vectors for expression in the same cell, with a phage vector encoding the g3p fusion and a phagemid encoding the soluble chain. The phage acts as a helper, and the infected bacteria produce both packaged phage and phagemid. Each phage or phagemid displays both chains but encodes only one chain and thus only the genetic information for half of the antigen-binding site. However, the genes for both antibody chains can be recovered separately by plating on the selective medium, suggesting a means by which mutually complementary pairs of antigen binding heavy and light chain combinations could be selected from random combinatorial libraries. For example, a light chain repertoire on fd phage could be used to infect cells harbouring a library of soluble heavy chains on the phagemid. The affinity purified phagemid library could then be used to infect *E.coli*, rescued with the affinity purified phage library, and the new combinatorial library subjected to a round of selection. Thus, antibody heavy and light chain genes are reshuffled after each round of purification. Finally, after several rounds, infected bacteria could be plated and screened individually for antigen-binding phage. Such 'dual' combinatorial libraries are potentially more diverse than those encoded on a single vector: by combining separate libraries of $10^7$ heavy chain and $10^7$ light chain phage(mid)s, the diversity of displayed Fab fragments (potentially $10^{14}$) is limited only by the number of bacteria ($10^{12}$ per litre). More simply, the use of two vectors should also facilitate the construction of 'hierarchical' libraries, in which a fixed heavy or light chain is paired with a library of partners (9), offering a means of 'fine-tuning' antibody affinity and specificity.

In conclusion the display of heterodimeric Fab fragments on the surface of phage offers new possibilities for making antibodies from diverse libraries of genes, and our strategy may serve as a model for the display of other multichain polypeptides.

## ACKNOWLEDGEMENTS

We thank M.Dreher and E.S.Ward for providing us with vectors, T.Clackson for advice on 'splicing by overlap extension', T.Bonnert for help with computer graphics and M.Hobart for sheep anti-M13 antibody. H.R.H. was supported by the D.Collen Research Foundation, Leuven, and the European Molecular Biology Organisation, and A.D.G. by the Cancer Research Campaign.

## REFERENCES

1. Glaser-Wuttke,G., Keppner,J. and Rasched,I. (1989) *Biochim. Biophys. Acta* **985**, 239–247.
2. Crissman,J.W. and Smith,G.P. (1984) *Virology* **132**, 445–455.
3. Parmley,S.F. and Smith,G.P. (1988) *Gene* **73**, 305–318.
4. Scott,J.K. and Smith,G.P. (1990) *Science* **249**, 386–390.
5. Devlin,J.J., Panganiban,L.C. and Devlin,P.E. (1990) *Science* **249**, 404–406.
6. Cwirla,S.E., Peters,E.A., Barrett,R.W. and Dower,W.J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6378–6382.
7. McCafferty,J., Griffiths,A.D., Winter,G. and Chiswell,D.J. (1990) *Nature* **348**, 552–554.
8. Bass,S., Greene,R. and Wells,J.A. (1990) *Proteins* **8**, 309–314.
9. Clackson,T.P., Hoogenboom,H.R., Griffiths,A.D. and Winter,G. (1991) *Nature* in press.
10. Winter,G. and Milstein,C. (1991) *Nature* **349**, 293–299.
11. Bird,R.E., Harriman,K.D., Jacobson,J.W., Johnson,S., Kaufman,B.M., Lee,S.-M., Lee,T., Pope,S.H., Riordan,G.S. and Whitlow,M. (1988) *Science* **242**, 423–426.
12. Huston,J.S., Levinson,D., Mudgett-Hunter,M., Tai,M.-S., Novotny,J., Margolies,M.N., Ridge,R.J., Bruccoleri,R.E., Haber,E., Crea,R. and Oppermann,H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5879–5883.
13. Better,M., Chang,C.P., Robinson,R.R. and Horwitz,A.H. (1988) *Science* **240**, 1041–1043.
14. Skerra,A. and Plückthun,A. (1988) *Science* **240**, 1038–1040.
15. Zacher,A.N., Stock,C.A., Golden,J.W. and Smith,G.P. (1980) *Gene* **9**, 127–140.
16. Vieira,J. and Messing,J. (1987) *Methods in Enzymol.* **153**, 3–11.
17. Munroe,S. and Pelham,H. (1986) *Cell* **46**, 291–300.
18. Ward,E.S., Gussow,D., Griffiths,A.D., Jones,P.T. and Winter,G. (1989) *Nature* **341**, 544–546.
19. Taylor,J.W., Ott,J. and Eckstein,F. (1985) *Nucl. Acids Res.* **13**, 8764–8785.
20. Griffiths,G.M., Berek,C., Kaartinen,M. and Milstein,C. (1984) *Nature* **312**, 271–275.
21. Horton,R.M., Hunt,H.D., Ho,S.N., Pullen,J.K. and Pease,L.R. (1989) *Gene* **77**, 61–68.
22. Clackson,T., Gussow,D. and Jones,P.T. (1991) In McPherson,J., Taylor,G.R. and Quirke,P. (eds), *PCR, A Practical Approach*. IRL Press, Oxford (in press).
23. Miller,J.H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
24. Towbin,H.T., Staehelin,T. and Gordon,J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4350–4354.
25. Laemmli,U.K. (1970) *Nature* **227**, 680–685.
26. Harlow,E. and Lane,D. (1988) *Antibodies. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
27. Bird,R.E. and Walker,B.W. (1991) *TIBTECH* **9**, 132–137.
28. Huse,W.D., Sastry,L., Iverson,S.A., Kang,A.S., Alting-Mees,M., Burton,D.R., Benkovic,S.J. and Lerner,R.A. (1989) *Science* **246**, 1275–1281.
29. Dower,W.J., Miller,J.F. and Ragsdale,C.W. (1988) *Nucl. Acids Res.* **16**, 6127–6145.

6 Trinchieri, G. *et al. J. exp. Med.* **160**, 1147-1169 (1984).
7. Tigges, M. A., Casey, L. S. & Koshland, M. E. *Science* **243**, 781-786 (1989)
8. Takacs, L., Osawa, H. & Diamantstein, T. *Eur. J. Immun.* **14**, 1152 1156 (1984)
9 Jenkinson, E. J., Kingston, R. & Owen, J. J. T. *Nature* **329**, 160-162 (1987)
10. Tenton, L., Longo, D. L. Zuniga-Pflucker, J. C., Wing, C. & Kruisbeek, A. M. *J. exp. Med.* **168**, 1741-1747 (1988)
11. Plum, J. & de Smedt, M. *Eur. J. Immun.* **18**, 795-799 (1988)
12. MacDonald, H. R. *et al. Immunol. Rev.* **104**, 157-182 (1988)
13 Doetschman, T. *et al. Nature* 330, 576-578 (1987)
14. Thomas, K. R. & Capecchi, M. R. *Cell* **51**, 503-512 (1987)
15. Thompson, S., Clarke, A. R., Pow, A. M., Hooper, M. L. & Melton, D. W. *Cell* **56**, 313-321 (1989)
16 Zijlstra, M., Li, E., Sajjadi, F., Subramani, S. & Jaenisch, R. *Nature* **342**, 435-438 (1989)
17 Fuse, A. *et al. Nucleic Acids Res.* **12**, 9323-9331 (1984)
18. Zurawski, S. M. & Zurawski, G. *EMBO J.* **7**, 1061-1069 (1988).
19 Evans, M. J. & Kaufman, M. H. *Nature* **292**, 154-156 (1981)
20 Shortman, K., Egerton, M., Spangrude, G. J. & Scollay, R. *Semin. Immun.* **2**, 3-12 (1990)
21. von Boehmer, H. & Kisielow, P. *Science* **248**, 1369-1372 (1990)
22 Weinberg, K. & Parkman, R. *New Engl. J. Med.* **322**, 1718-1723 (1990)
23. Chen-Bettecken, U., Weckar, E. & Schimpl, A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7384-7388 (1985).
24. Pahwa, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 5069-5073 (1989)
25 Hooper, M. *et al. Nature* **326**, 292-295 (1987)
26. Smith, A. G. & Hooper, M. L. *Devl. Biol.* **121**, 1-9 (1987).
27 Williams, R. L. *et al. Nature* **336**, 684-687 (1988)
28 Kubo, R., Born, W., Kappler, J., Marrack, P. & Pigeon, M. *J. Immun.* **142**, 2736-2742 (1989)
29 Molinaro, G. A., Maron, E., Eby, W. C. & Dray, S. *Eur. J. Immun.* **5**, 771-774 (1975)

ACKNOWLEDGEMENTS. We thank M. Hooper for ES cells, R. Swoboda for IL-2 DNA clones, H. Haber, R. Mitnacht and K. Borschert for assistance, and S. Siddell for critical reading of the manuscript. These studies were supported by the Deutsche Forschungsgemeinschaft and by the Fonds der Chemischen Industrie.

# Making antibody fragments using phage display libraries

**Tim Clackson*, Hennie R. Hoogenboom†, Andrew D. Griffiths† & Greg Winter*†**

*MRC Laboratory of Molecular Biology and †Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, UK

**To by-pass hybridoma technology and animal immunization, we are trying to build antibodies in bacteria by mimicking features of immune selection[1]. Recently we used fd phage[2] to display antibody fragments fused to a minor coat protein[3,4], allowing enrichment of phage with antigen[3]. Using a random combinatorial library of the rearranged heavy (VH) and kappa (Vκ) light chains[5-8] from mice immune to the hapten 2-phenyloxazol-5-one (phOx), we have now displayed diverse libraries of antibody fragments on the surface of fd phage. After a single pass over a hapten affinity column, fd phage with a range of phOx binding activities were detected, at least one with high affinity (dissociation constant, $K_d = 10^{-9}$ M). A second pass enriched for the strong binders at the expense of the weak. The binders were encoded by *V* genes similar to those found in anti-phOx hybridomas but in promiscuous combinations (where the same *V* gene is found with several different partners). By combining a promiscuous VH or Vκ gene with diverse repertoires of partners to create hierarchical libraries, we elicited many more pairings with strong binding activities. Phage display offers new ways of making antibodies from *V*-gene libraries, altering V-domain pairings and selecting for antibodies with good affinities.**

We used the polymerase chain reaction (PCR)[9] to amplify the VH and Vκ genes from the spleen messenger RNA of mice immunized with phOx, and also developed a 'PCR assembly' process[10] to link these genes together randomly for expression as single-chain Fv (scFv) fragments[11,12] (Fig. 1*a–c*). The assembled genes were cloned in a single step into the vector fdDOG1 (Fig. 1*e*) for display as a fusion with the fd gene III coat protein. This initial library of $2 \times 10^5$ clones seemed to be diverse (Fig. 1*d*), and sequencing revealed the presence of most VH groups[13] and Vκ subgroups[14] (data not shown). None of the 568 clones tested bound to phOx as detected by enzyme-linked immunosorbent assay (ELISA).

The library of phages was passed down a phOx affinity column (Table 1*a*), and eluted with hapten. Of the eluted clones, 13%

TABLE 1 Affinity selection of hapten-binding phage

| | Clones binding to phOx* | | | |
|---|---|---|---|---|
| | Precolumn | After first round | After second round | After third round |
| (*a*) Random combinatorial libraries | | | | |
| phOx-immunized mice | 0/568 (0%) | 48/376 (13%) | 175/188 (93%) | — |
| Unimmunized mice | — | — | 0/388 (0%) | — |
| (*b*) Hierarchical libraries | | | | |
| VH-B/Vκ-*rep* library | 6/190 (3%) | 348/380 (92%) | — | — |
| VH-rep/Vκ-d library | 0/190 (0%) | 23/380 (7%) | — | — |
| (*c*) Fractionation of VH-B/Vκ-d and VH-B/Vκ-b phage* | | | | |
| Mixture of clones | 88/1,896 (4.6%) (44/1,740 (2.5%)†) | 55/95 (57.9%) | 1,152/1,156 (99.7%) | 1,296/1,299 (99.8%) |

Selection of phage with hapten-binding activities from the random combinatorial and hierarchical libraries (*a* and *b*, respectively), and fractionation of clones with different affinities for phOx (*c*). For the random combinatorial libraries fdDOG1 RF was extensively digested with *Not*I and *Apa*LI, purified by electroelution[24] and 1 μg ligated to 0.5 μg (5 μg for the hierarchical libraries) of the assembled *scFv* genes in 1 ml with 8,000 units T4 DNA ligase (New England Biolabs) overnight at 16 °C. Purified ligation mix was electroporated in six aliquots into MC1061 cells[25] and plated on NZY medium[24] with 15 μg ml$^{-1}$ tetracycline, in 243 × 243 mm dishes (Nunc); 90–95% of clones contained *scFv* genes by PCR screening (see legend to Fig. 1). Colonies were scraped into 50 ml 2 × TY medium[26] and shaken at 37 °C for 30 min. Liberated phage were precipitated twice with polyethylene glycol and resuspended to $10^{12}$ transducing units (TU) ml$^{-1}$ in water (titred as in ref. 3). For affinity selection, a 1-ml column of phOx-BSA–Sepharose[27] M. Dreher and C. Milstein, unpublished results) was washed with 300 ml PBS, and 20 ml PBS containing 2% skimmed milk powder (MPBS). Phage ($10^{12}$ TU) were loaded in 10 ml MPBS, washed with 10 ml MPBS and finally 200 ml PBS. The bound phage were eluted with 5 ml 1 mM 4-ε-amino-caproic acid methylene 2-phenyl-oxazol-5-one (phOx-CAP). About $10^8$ TU eluted phage were amplified by infecting 1 ml log phase *E. coli* TG1 (ref. 28) and plating as above. For a further round of selection, colonies were scraped into 10 ml 2 × TY medium and then processed as above. For the hierarchical libraries, VH-B and Vκ-d genes were individually recloned, then assembled with the VH or Vκ repertoires. For the fractionation of clone VH-B/Vκ-d, $7 \times 10^{10}$ TU phage in the ratio 20 VH-B/Vκ-b:1 VH-B/Vκ-d were loaded onto a phOx-BSA-Sepharose column in 10 ml MPBS and eluted as above. Eluted phage were used to reinfect *E. coli* TG1, and phage produced and harvested as before. About $10^{11}$ TU of phage were loaded onto a second affinity column and the process repeated to give a total of three column passes. Dilutions of eluted phage at each stage were plated in duplicate and probed separately[24] with oligonucleotides specific for Vκ-b (5′-GAGCGGGTAACCACTGTACT) or Vκ-d (5′-GAATGGTATAGTACTACCCT).

* In (*c*), numbers refer to VH-B/Vκ-d.

† Numbers after three reinfections and cycles of growth. This control, omitting the column steps, confirms that a spurious growth or infectivity advantage was not responsible for the enrichment of clone VH-B/Vκ-d.

bound to phOx, and ranged from poor to strong binding in ELISA. We sequenced 23 of these hapten-binding clones and found eight different $V_H$ genes (A–H) in a variety of pairings with seven different $V_\kappa$ genes (a–g) (Fig. 2*a*). Most of the domains, such as VH-B and V$\kappa$-d, were able to bind hapten with any of several partners[15]. The probability of finding multiple partners for a given chain should depend mainly on the inherent promiscuity of the chain and on the number of available partners and competing chains. Two other examples of promiscuous pairings have been noted in random combinatorial libraries made in λ phage[6,8], so this may prove to be a feature of small combinatorial libraries from immunized animals.

The sequences of the *V* genes were related to those seen in the secondary response to phOx, but with differences (Fig. 2*b*). Thus most phOx hybridomas from the secondary response use somatically mutated derivatives of three types of $V_\kappa$ genes, $V_\kappa$ox1, '$V_\kappa$ox-like' and $V_\kappa$45.1 genes[16]. These can pair with $V_H$ genes from several groups, but $V_\kappa$ox1 more commonly pairs with the $V_H$ox1 gene ($V_H$ group 2; ref. 13). $V_\kappa$ox1 genes are always, and $V_\kappa$ox-like genes often, found in association with heavy chain genes (including $V_H$ox1) that encode a short five-residue CDR3, with the sequence motif Asp-X-Gly-X-X (where X is any amino acid[16]), in which the central glycine creates a cavity for phOx (ref. 17). In our library, nearly all of the $V_H$ genes belonged to group 1, and most of the $V_\kappa$ genes were ox-like and associated with $V_H$ genes encoding a five-residue CDR3 motif Asp/Asn-X-Gly-X-X (Fig. 2*b*). $V_\kappa$ox1 and $V_H$ox1 were found only once ($V_\kappa$–f and VH-E), and not in combination with each other: indeed $V_\kappa$-f does not encode the Trp 91 involved in phOx binding[17] and was paired with a $V_H$ gene ($V_H$-C) encoding a six-residue CDR3.

The promiscuity of the VH-B and V$\kappa$-d domains prompted us to force further pairings, by assembling these genes with the entire repertoires of either $V_\kappa$ or $V_H$ genes from the same immunized mice. The resulting hierarchical libraries, (VH-B × $V_\kappa$-*rep* and $V_H$-*rep* × V$\kappa$-d), each with $4 \times 10^7$ members, were subjected to a round of selection and hapten-binding clones isolated (Table 1*b*). Most were strong binders by ELISA

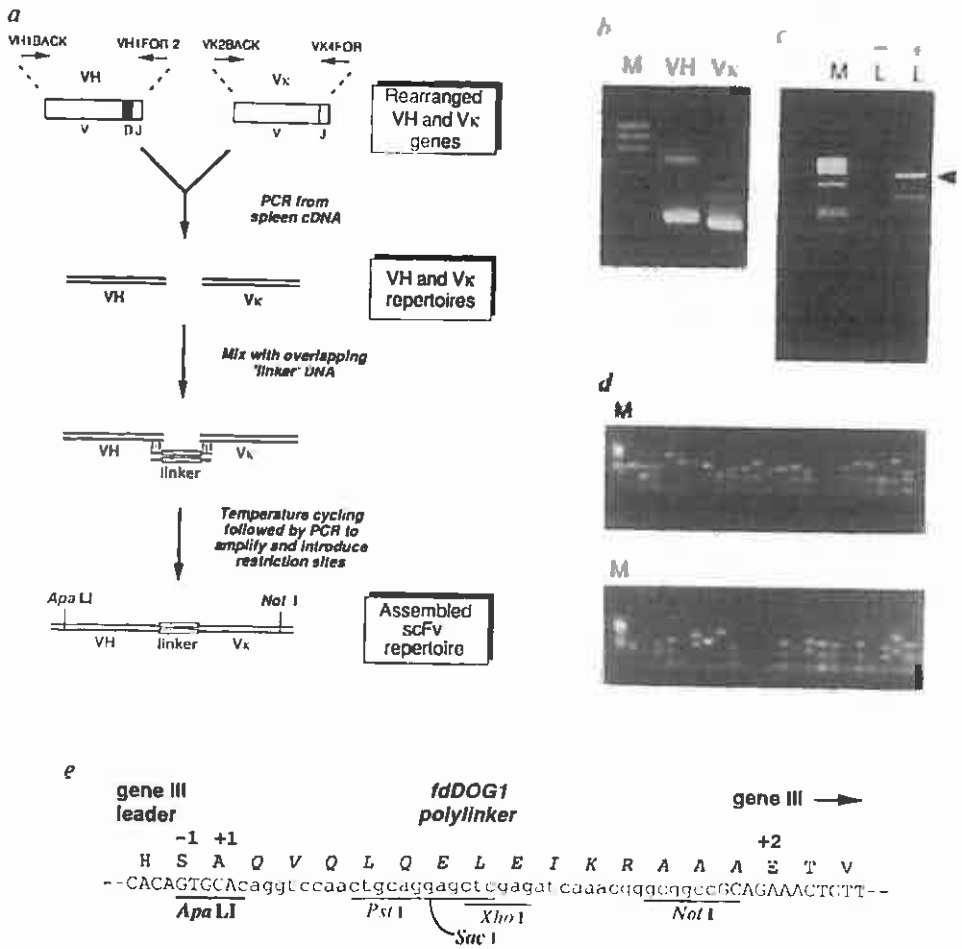


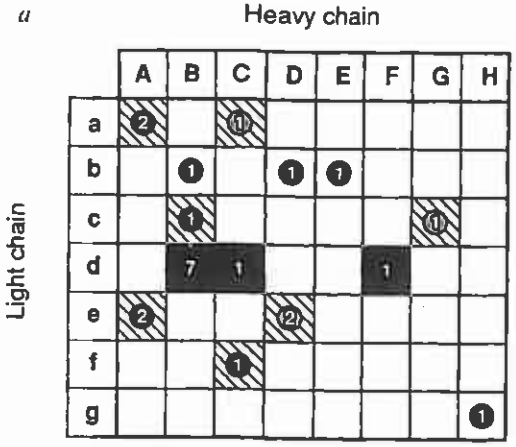

FIG. 1 PCR assembly of scFv library. *a*, $V_H$ and $V_\kappa$ genes are separately amplified, then mixed with a linker fragment that overlaps them both. The linker (93 base pairs) encodes the short peptide, $(Gly_4Ser)_3$, which links $V_H$ and $V_\kappa$ in scFvs (ref. 11). Cycles of annealing-denaturation, followed by reamplification of the mixture, generate a random combinatorial cassette of $V_H$ and $V_\kappa$ genes joined in-frame for expression. *b*, $V_H$ and $V_\kappa$ gene repertoire PCR products from the immunized mice analysed by electrophoresis on agarose (1%) gel. *c*, PCR assembly of *scFv* gene repertoires with linker (+L) or without (−L); arrow indicates assembled repertoire. M is DNA marker ΦX174 replicative form DNA digested with *Hae*III. *d*, Diversity of library as seen by *Bst*NI fingerprinting of individual clones. *e*, Sequence of *fd* gene III around the signal peptide cleavage site in fdDOG1.

METHODS. For the random combinatorial libraries, cytoplasmic RNA was isolated[29] from the pooled spleens of either 5 male BALB/c mice boosted 8 weeks after primary immunization with phOx coupled to chicken serum albumin[27], or of 5 unimmunized mice. The cDNA was made with avian myoblastosis virus reverse transcriptase (Anglian Biotech)[30] and primers that straddle the junction between the variable and constant regions of γ heavy chains and κ light chains (C. Marks, unpublished data). $V_H$ and $V_\kappa$ repertoires were amplified from the cDNA with 25 cycles of PCR (94 °C for 1 min, 60 °C for 1 min, 72 °C for 2 min) using Vent polymerase (New England Biolabs) and the primers VH1BACK (ref. 19) and VH1FOR-2 (ref. 31) or the primers VK2BACK and VK4FOR. The linker DNA was similarly amplified from pSW2scD1.3 (ref. 3) using primers LINKFOR and LINKBACK (complementary to VK2BACK and VH1FOR-2 respectively). After gel purification, 1 μg each of the $V_H$ and $V_\kappa$ products were mixed with 300 ng linker in a 25 μl PCR reaction mix without primers, and cycled 7 times (94 °C 2 min, 72 °C 4 min) to join the fragments, then amplified for 20 cycles (94 °C 1.5 min, 72 °C 2.5 min) using 25 pmol each VH1BACK and VK4FOR primers. Finally, the assembled products were gel-purified and reamplified with VH1BACK-ApaLI and VK4FOR-NotI ('tagged' versions of the original primers) to append restriction sites. Products (1–5 μg) were extensively digested with *Apa*LI and *Not*I for cloning into fdDOG1. Recombinant colonies were screened by PCR[32] with the primers VH1BACK and VK4FOR, followed by digestion with the frequently cutting enzyme *Bst*NI. Primers: VK2BACK 5′-GACATTGAGCTCACCCAGTCTCCA; VK4FOR, an equimolar mix of 5′-CCGTTTGATTTCCAGCTTGGTGCC, 5′-CCGTTTTATTTCCAGCTTGGTCCC, 5′-CCGTTTTATTTCCAACTTTGTCCC and 5′-CCGTTTCAGCTCCAGCTTGGTCCC; LINKFOR, 5′-TGGAGACTGGGTGAGCTCAATGTC; LINKBACK, 5′-GGGACCACGGTCACCGTCTCCTCA; VH1BACK-ApaLI, as VH1BACK (ref. 19) but with 5′-CATGACCACAGTGCAC added at the 5′ end; VK4FOR-NotI, as VK4FOR but with 5′-GAGTCATTCTGCGGCCGC similarly added (restriction sites underlined).

FIG. 2 *a*, Matrix of $V_H$ and $V_\kappa$ genes identified in phOx-binding clones selected from random combinatorial library. The number of clones found with each combination is shown. The binding to phOx–BSA, as judged by the ELISA signal, seemed to vary (marked by shading): no binding was seen to BSA alone. Optical density at 405 nm: 0.2–0.9, dotted box; 0.9–2.0, hatched box; >2.0, solid box. *b*, Encoded protein sequences of phOx-binding clones. Sequences of phOx-binding clones isolated (single-letter amino-acid code) after one round of selection of the random combinatorial library, with pairings as above, or the hierarchical library. Note that the first eight or seven residues, and the last nine or eleven residues, of the $V_\kappa$ or $V_H$ genes, respectively, are encoded by the PCR primers. Classifications into $V_H$ groups[13] and $V_\kappa$ subgroups[14], and the position of residue 91 encoded by the $V_\kappa$ genes (#), are indicated. The relationship to genes from the hybridoma analysed secondary response to phOx (ref. 16) is also shown; all of the $V_\kappa$ genes are 'ox-like', apart from those marked * with an asterisk, which are $V_\kappa$ox1, and the only example of $V_H$ group 2 ($V_H$-E) is $V_H$ox1. The intensity of ELISA signals from the hierarchical libraries, corrected relative to the signal from control phage, are indicated: Optical density at 405 nm 0.9–2.0 (++), >2.0 (+++). Multiple isolations of sequences are marked, and sequences ($V_H$-B and $V_\kappa$-c) isolated from the random combinatorial library, and also the hierarchical libraries, are shown in italics. The $V_H$-B/$V_\kappa$-d and $V_H$-B/$V_\kappa$-c pairings gave similar signals (after correction of ELISA) when recovered from either combinatorial or hierarchical libraries.

METHODS. We screened for binding of the phage to hapten by ELISA: 96-well plates were coated with 10 $\mu g\ ml^{-1}$ phOx-BSA[27] or 10 $\mu g\ ml^{-1}$ BSA in PBS overnight at room temperature. Colonies of phage-transduced bacteria were inoculated into 200 $\mu l$ 2×TY medium[28] with 12.5 $\mu g\ ml^{-1}$ tetracycline in 96-well plates ('cell wells', Nuclon) and grown with shaking (300 r.p.m.) for 24 h at 37 °C. At this stage, cultures were saturated and phage titres were reproducible ($10^{10}$ TU $ml^{-1}$). Phage supernatant (50 $\mu l$), mixed with 50 $\mu l$ PBS containing 4% skimmed milk powder, was then added to the coated plates. Further details given in ref. 3. To sequence the clones, template was prepared[24] from the supernatants of 10-ml cultures grown for 24 h, and sequenced using the dideoxy method[33] and a Sequenase kit (USB), with primer LINKFOR for the $V_H$ genes and primer fdSEQ1 (5'-GAATTTTCTGTATGAGG) for the $V_\kappa$ genes.

*a*




*b*

$V_H$ sequences

From combinatorial library:

| | CDR1 | CDR2 | CDR3 | Group | ELISA signal |
|---|---|---|---|---|---|
| A | QVQLQQSGAELARPGASVKMSCKASGYTFT SYTMH WVKQRPGQGLEWIG YINPSSGYTNYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVYYCAN RYGAY WGQGTTVTVSS x4 | 1 | |
| B | QVKLQQSGAELARPGASVKMSCKASGYTFT RDMMH WLKQRPGQGLEWIG YINPSTGYTEYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVTYCAR NYGLY WGQGTTVTVSS x9 | 1 | |
| C | QVQLQQSGPELVKPGASVKMSCKASGYTFT SYVMH WVKQKPGQGLEWIG YINPYNDGTKYNEKFKG KATLTSDKSSSTAYMELSSLTSEDSAVYYCAI YRSFPY WGQGTTVTVSS x3 | 1 | |
| D | QVQLQQSGPELVKPGASVKISCKASGYSFT GYFMN WVKQSHGKSLEWIG RINPYNGDTFYNQKFKG KATLTVDKSSSTAHMELLSLTSEDSAVYYCVG ITTRFAY WGQGTTVTVSS x3 | 1 | (see |
| E | QVQLQESGPGLVAPSQSLSITCTVSGFSLT SYGVH WVRQPPGRGLEWLG VIWAGGSTNYNSALMS RLSISKDNSKSQVFLKMNSLQTDDTAMYYCAR DRGDY WGQGTTVTVSS | 2 | Fig.2a) |
| F | QVQLQQSGPELAKPGASVKMSCKASGYTFT SYLMH WVKQRPGQGLKWIG YINPSTGYTEYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR DYGYY WGQGTTVTVSS | 1 | |
| G | QVKLQQSGAELVRPGASVKLSCKASGYTFT RYLMH WVRQRPGQGLEWIG YINPSTGYTEYNQKFKD EATLTADKSSNTAYMQLSSLTSEDSAVYYCAR DYGYY WGQGTTVTVSS | 1 | |
| H | QVQLQQSGPELMKPGASVKISCKASGYSFS RNYMH WVKQSHGKSLEWIG YIAPFNGGTTYNQKFKG KATLTVDRSSSTAYMHLSSLTSEDSAVYYCAT DYGRD WGQGTTVTVSS | 1 | |

From hierarchical library $V_H$-rep × $V_\kappa$-d.

| | | | | |
|---|---|---|---|---|
| I | QVKLQQSGPELARPGVSVRMSCKASGYTFT SYAMH WVKQSQSKSLEWIG VISTYNGNTNYNQKFKG KATMTVDKSSSTAYMELARLTSEDSAIYYCAR DYGDY WGQGTTVTVSS | 1 | +++ |
| J | QVKLQQSGAELARPGASVKMSCKASGYTFT RYTMH WVKQRPGQGLEWIG YINPSSGYTNYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR DRGAY WGQGTTVTVSS | 1 | +++ |
| K | QVKLQQSGAELAKPGASVKMSCKASGYTFT RDWGH WVKQRPGQGLEWIG YINPSTGYTEYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR NYGLY WGQGTTVTVSS x3 | 1 | +++ |
| L | QVQLQQSGLELARPGASVKMSCKASGYTFT NYLMH WVKQRPGQGLEWIG YINPSTGYTEYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR DYGYY WGQGTTVTVSS x2 | 1 | ++ |
| M | QVKLQQSGAELARPGASVKMSCKASGYTFT NYMMH WVKQRPGQGLEWIG YINPSTGYTEYNQKFKD KATLTADKSSSTAYMQLSSLTSDDSAVYYCAR DYGYY WGQGTTVTVSS | 1 | +++ |
| N | QVQLQQSGAELVKPGASVKLSCKTSGYTFT SYTMH WVKQRPGQGLEWIG YINPSSGYTNYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR DYGYY WGQGTTVTVSS | 1 | ++ |
| O | QVQLQQSGAELAKPGASVKMSCEASGYTFT SHLMH WVKQRPGQGLEWIG YINPRTGYTEYNQKFKD KATFTADKSSSTAYMQLSSLTSEDSAVYYCAR DYGAY WGQGTTVTVSS | 1 | ++ |
| P | QVKLQQSGAELAKPGASVKMSCKASGYTFT SYMMH WVKQRPGQGLEWIG YINPSTGYTEYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR DYGYY WGQGTTVTVSS | 1 | ++ |
| Q | QVKLQQSGAELAKPGASVKMSCKATGYTFT SYLMH WVKQRPGQGLEWIG YINPSTGYTEYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR DYGYY WGQGTTVTVSS | 1 | +++ |
| R | QVQLQQSGAELARPGASVKMSCKASGYTFT SYVMH WVKQRPGQGLEWIG YINPSSGYTNYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR DYGYY WGQGTTVTVSS | 1 | +++ |
| S | QVQLQQSGAELARPGASVKMSCKASGYTFT TFLMH WLKQRPGQGLEWIG YINPSTGYTEYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR NYGIY WGQGTTVTVSS | 1 | ++ |
| T | QVKLQQSGAELARPGASVKMSCKASGYTFT SYTMH WVKQRPGQGLEWIG YINPSTGYTEYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR DYGYY WGQGTTVTVSS x2 | 1 | +++ |
| U | QVKLQQSGAELARPGASVKMSCKASGYTFT SYTMH WVKQRPGQGLEWIG YINPTTGYTEYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR DYGYY WGQGTTVTVSS | 1 | +++ |
| *B* | *QVKLQQSGAELAKPGASVKMSCKASGYTFT RDMMH WLKQRPGQGLEWIG YINPSTGYTEYNQKFKD KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR NYGLY WGQGTTVTVSS* | 1 | +++ |

$V_\kappa$ sequences

From combinatorial library:

| | CDR1 | CDR2 | CDR3 | | |
|---|---|---|---|---|---|
| a | DIELTQSPSSLSASLGERVSLTC RASQEISGYLS WLQQKPDGSIKRLIY AASTLES GVPKRFSGSRSGSDYSLTISSLESEDFADIYC LQYASYPT FGAGTKLEIKRA x3 | V | |
| b | DIELTQSPAIMSASPGEKVTMTC RASSSVSSSYLH WYQQKSGASPKVWIY STSNLAS GVPARFSGSGSGTSYSLTISSVEAEDAATYYC QQYSGYPLT FGAGTKLEIKRA x3 | IV | |
| c | DIELTQSPTTMAASPGEKITITC SASSSISSNYLH WYQQKPGFSPKLLIY RTSNLAS GVPARFSGSGSGTSYSLTIGTMEAEDVATYYC QQGSSIPLT FGAGTKLEIKRA x2 | IV | (see |
| d | DIELTQSPTTMAASPGEKITITC SASSSISSNYLH WYQQKPGFSPKLLIS RTSNLAS GVPARFSGSGSGTSYSLTIGTMEAEDVATYYC QQGSTIPFT FGSGTKLEIKRA x9 | IV | Fig.2a) |
| e | DIELTQSPAIMSASPGEKVTITC SASSSVNYMH WFQQKPGTSPKLWIY STSNLAS GVPTRFSGSGSGTSYSLTISRMEAEDAATYYC QQRSSYPFT FGSGTKLEIKRA x9 | IV | |
| f | DIELTQSPAIMSAFPGEKVTMTC SASSSVSYMH WYQQKSGTSPKRWIY DTSKLAS GVPARFSGSGSGTSYSLTISSMEAEDAATYYC QQFSSNPLT FGAGTKLELKRA | VJ | |
| g | DIELTQSPAIMSASPGEKVTMTC SASSSINYMH WYQQKPGASPKRWIY DTSKLAS GVPARFSGSGSGTSYSLTISSMEAEDAATYYC HQRNSYPWT FGGGTKLEIKRA | VI | |

From hierarchical library $V_H$-B × $V_\kappa$-rep

| | | | | |
|---|---|---|---|---|
| h | DIELTQSPAIMSASPGEKVTMTC SASSSVSYMH WYQQKSGTSPKRWIY DTSKLAS GVPARFSGSGSGTSYSLTISSMEAEDAATYYC QQWSSNPLT FGAGTKLEIKRA x4 | VI * | +++ |
| i | DIELTQSPAIMSASPGEKVTITC SASSSVSYIH WFQQKPGTSPKLWIY STSNLAS GVPARFSGSGSGTSYSLTISRMEAEDAATYYC QQYHSYPLT FGAGTKLEIKRA | VI | +++ |
| j | DIELTQSPTTMAASPGEKITITC SASSSISSNYLH WYQQKPGFSPKLLIY RTSNLAS GVPARFSGSGSGTSYSLTIGTMEAEDVATYYC QQGSSIPLT FGGGTKLEIKRA | IV | ++ |
| k | DIELTQSPTTMAASPGDMITITC SATSSISSNYLH WYQQKPGFSPKLLIY RTSNLAS GVPPRFSGSGSGTSYSLTIGAMEAEDVATYYC QQGSSIPYT FGAGTKLEIKRA | IV | +++ |
| l | DIELTQSPTTMAASPGEKITITC SASSSISSNYLH WYQQKPGFSPKLLIY RTSNLAS GVPARFSGSGSGTSYSLTIGTMEAEDVATYYC QQGSSIPYT FGGGTKLEIKRA | IV | ++ |
| m | DIELTQSPTTMAASPGEKITITC SASSSISSNHLH WYQQKPGFSPKLLIY RTSNLAS GVPARFSGSGSGTSYSLTIGTMEAEDVATYYC QQGSSIPYT FGGGTKLEIKRA | IV | +++ |
| n | DIELTQSPTTMAASPGEKITITC SASSSISSNYLH WYQQKPGFSPKLLIY RTSNLAS GVPARFSGSGSGTSYSLTIGTMEAEDVATYYC QQGSGIPYT FGGGTKLEIKRA | IV | ++ |
| o | DIELTQSPAIMAASPGEKITITC SASSSISSNYLH WYQQKPGFSPKLLIY RTSNLAS GVPARFSGSGSGTSYSLTIGTMEAEDVATYYC QQGSSIPFT FGGGTKLEIKRA | IV | +++ |
| p | DIELTQSPAIMSASPGEKVTMTC SASSSVSYMH WYQQKSGTSPKRWIY DTSKLAS GVPARFSGSGSGTSYSLTISSMEAEDVATYYC QQWSSNPLT FGAGTKLEIKRA x2 | IV | +++ |
| q | DIELTQSPAIMSASPGDKVTLTC SASSSVRYVN WFQQKSGTSPKRWIY DTSKLAS GVPARFSGSGSGTSYSLTISSMEAEDAATYYC QQWTSNPPT FGGGTKLEIKRA | VI * | +++ |
| r | DIELTQSPAIMSASPGEKVTMTC SASSSVSYMH WYQQKSGTSPKRWIY DTSKLAS GVPARFSGSGSGTSYSLTISSMEAEDAATYYC QQWTSNPPT FGGGTKLEIKRA | VI * | +++ |
| s | DIELTQSPAIMSASPGEKVTMTC RASSSVTSSYLN WYQQKSGASPKLWVY STSNLAS GVPARFSGSGSGTSYSLTISSVEAEDAATYYC QQWTSNALT FGAGTKLEIKRA | VI * | ++ |
| t | DIELTQSPAIMSASPGEKVTMTC RASSSVSSSYLN WYQQKSGASPKLWIY STSNLAS GVPARFSGSGSGTSYSLTISSVEAEDAATYYC QQYSGYPLT FGAGTKLEIKRA | IV | +++ |
| u | DIELTQSPAIMSASPGEKVTMTC RASSSVSSSYLH WYQQKSGASPKLWIY STSNLAS GVPARFSGSGSGTSYSLTISSVEAEDAATYYC QQYSGYPLT FGAGTKLEIKRA | IV | +++ |
| v | DIELTQSPAIMSASPGEKVTMTC RASSSVSSSYLH WFQQKSGASPKLWIY STSNLPS GVPARFSGSGSGTSYSLTISSVEAEDAATYYC QQYSGYPLT FGGGTKLEIKRA | IV | ++ |
| *a* | *DIELTQSPTTMAASPGEKITITC SASSSISSNYLH WYQQKPGFSPKLLIY RTSNLAS GVPARFSGSGSGTSYSLTIGTMEAEDVATYYC QQGSSIPLT FGAGTKLEIKRA x3* | IV | ++ |



225

(Fig. 2*b*). By sequencing 23 clones from each library, we identified 14 new partners for VH-B and 13 for Vκ-d; apart from VH-B and Vκ-c, none of the previous VH-B or Vκ-d partners (or indeed other partners) cloned and sequenced from the random combinatorial library was isolated again. These features are consistent with the much larger number of available partners ($4\times10^7$) for the VH-B (or Vκ-d) domain, and the promiscuous nature of the domain. The *Vκ* genes were mainly ox-like and the *VH* genes mainly group 1, but the only examples of *Vκ*ox1 (Vκ-h, −p, −q and −r) encode Trp 91, and the VH-CDR3 motif Asp-X-Gly-X-X now predominates. Thus some features of the phOx hybridomas seem to emerge more strongly in the hierarchical library. The new partners differ from each other mainly by small alterations in the CDRs, indicating that much of the subtle diversity had not been tapped by the original random combinatorial library. More generally we find that a range of related antibodies can be made by keeping one of the partners fixed and varying the other, and this could be invaluable for fine tuning of antibody affinity and specificity.

To determine the range of antibody affinities for phOx, we recloned the combinations of VH-B with Vκ-b and Vκ-d (which gave weak and strong binding signals to phOx in ELISA) for secretion as soluble scFv fragments (Fig. 3, legend). Fluorescence quench titrations determined the $K_d$ of VH-B/Vκ-d for phOx-GABA as $1.0\times10^{-8}$ M (Fig. 3*a*), indicating that antibodies with affinities representative of the secondary response can be selected from phage display libraries. Indeed of anti-phOx hybridomas from the secondary response, only two (out of 11 characterized) secrete antibodies of a higher affinity than VH-B/Vκ-d (ref. 16). The $K_d$ of VH-B/Vκ-b for phOx-GABA was determined as $1.8\times10^{-5}$ M (Fig. 3*b*); thus phage bearing scFv fragments with weak affinities can also be selected with antigen, probably because of the avidity of the multiple antibody heads on the phage.

A second round of selection of the original, random combinatorial library from immune mice resulted in 93% of eluted clones binding phOx (Table 1*a*). Most of these clones were Vκ-d combinations, and bound strongly to phOx in ELISA (data not shown). Few weak binders were seen. This suggested that affinity chromatography had not only enriched for binders, but also for the best. To confirm this we mixed the phage VH-B/Vκ-d with a 20-fold excess of the phage VH-B/Vκ-b and subjected the mixture to rounds of selection: after only two rounds, essentially all the eluted phage were VH-B/Vκ-d (Table 1*c*).

We also constructed a random combinatorial library ($2\times10^6$ members) from unimmunized mice, but found no phOx-binding clones after two rounds of selection (Table 1*a*). Immunization therefore seems to be necessary to create and/or enrich for VH or Vκ domains with at least some of the features required for hapten binding. With libraries of this size ($\sim10^6$ members), such domains need to be represented at a high frequency to reconstitute a binding site[1], and immunization ensures this by biasing the spleen lymphoid cell population heavily towards messenger RNA-rich blast cells making specific antibody (R. Hawkins and G.W., unpublished data). With larger libraries ($>10^7$) now accessible using selection[3] rather than screening[5-8], immunization may be unnecessary for the isolation of antibody fragments. It has been estimated that a library of $10^7$ different antibodies will probably recognize >99% of epitopes with a dissociation constant of $\geq10^{-5}$ M (ref. 18), and we have shown here that we can recover antibody fragments with such affinities

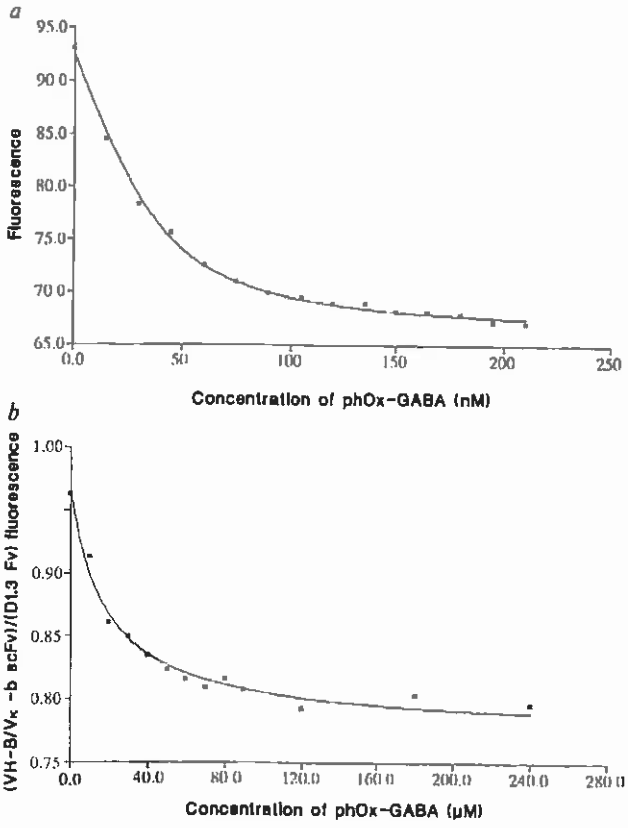



FIG. 3 Fluorescence quench titration of soluble scFv fragments. *a*, The $K_d$ ($1.0\pm0.2\times10^{-8}$ M) for clone VH-B/Vκ-d was determined by fluorescence quench titration[34] of purified scFv (100 nM) with 4-γ-amino-butyric acid methylene 2-phenyl-oxazol-5-one (phOx-GABA). Excitation was at 280 nm, emission was monitored at 340 nm and the $K_d$ calculated as in refs 35 and 36. All values were calculated with standard errors included. The $K_d$ was determined to be $1.0\pm0.2\times10^{-8}$ M with $0.38\pm0.05$ binding sites per scFv molecule. *b*, For measurement of the $K_d$ of the low affinity clone VH-B/Vκ-b, 2 μM purified scFv protein was titrated with phOx-GABA as above. But to minimize light absorption by the higher concentrations of phOx-GABA required, excitation was at 260 nm and emission was monitored at 304 nm. In addition, the fluorescence values were divided by those from a parallel titration of the lysozyme binding Fv fragment D1.3 (ref. 31). The $K_d$ was calculated as described in refs 34 and 36 and determined to be $1.8\pm0.3\times10^{-5}$ M, with a fractional quench of $0.20\pm0.01$.

METHODS. Clones VH-B/Vκ-b and VH-B/Vκ-d were reamplified with VK4FOR-*Not*I and VH1BACK-*Sfi*I (5′-CATGCCATGACTCGCGGCCCAGCCGGCCATGG-CC(G/C)AGGT(C/G)(A/C)A(A/G)CTGCAG(C/G)AGTC(A/T)GG-3′), a primer that introduces an *Sfi*I site (underlined) at the 5′ end of the VH gene. VH-B/Vκ-d was cloned into the phagemid pJM1 (A.D.G. and J. Marks, unpublished results) as an *Sfi*I–*Not*I cassette, downstream of the pelB leader for periplasmic secretion[37], with a C-terminal peptide tag for detection[31,38], and under the control of a $\lambda P_L$ promoter[39]. Cultures (1.0 l) of *Escherichia coli* N4830-1 (ref. 40) harbouring each phagemid were induced[26] and supernatants precipitated with 50% ammonium sulphate. The resuspended precipitate was dialysed into PBS, 0.2 mM EDTA (PBSE), loaded onto a 1.5-ml column of phOx:Sepharose[41] and the column washed sequentially with 100 ml PBS; 100 ml 0.1 M Tris-HCl, 0.5 M NaCl pH 8.0; 10 ml 50 mM citrate, pH 5.0; 10 ml 50 mM citrate, pH4.0 and 20 ml 50 mM glycine, pH 3.0. The scFv fragment was eluted with 50 mM glycine, pH 2.0, neutralized with Tris base and dialysed against PBSE. VH-B/Vκ-b was cloned into a phagemid vector (A.D.G., unpublished results) based on pUC119 (ref. 42) encoding identical signal and tag sequences to pJM1, and expression induced at 30 °C in a 10-culture of *E. coli* TG1 (ref. 28) harbouring the phagemid, as in ref. 43. The low affinity of clone VH-B/Vκ-b made its purification on phOx-Sepharose impossible. Therefore after concentration by ultrafiltration (Filtron, Flowgen) the supernatant (100 ml of 600 ml) was loaded onto a 1-ml column of protein A–Sepharose coupled[44] to the monoclonal antibody 9E10 that recognizes the C-terminal peptide tag[31,38]. The column was washed with 200 ml PBS and 50 ml PBS, 0.5 M NaCl. The scFv fragment was eluted with 100 ml 0.2M glycine, pH 3.0, with neutralization and dialysis as before.

from phage display libraries. The antibody fragments could be rebuilt from their genes into complete antibodies, and expressed in myeloma cells if required, as described in ref. 19.

It may be possible to retain the original VH/Vκ pairings of the splenocytes, as in hybridoma technology. In principle, PCR assembly could be used to construct such 'natural' libraries, if the *V* genes from individual cells could be amplified and assembled in capsules. More immediately, affinity selection from combinatorial and hierarchical libraries promises an attractive route to high-affinity antibodies, in particular those from humans that are difficult to produce by hybridoma technology[1]. But the use of phage display libraries is not limited to antibodies: it offers a powerful and general method to change and refine the properties of any other protein[20] or peptides[21-23] that can be displayed on the phage surface. □

Received 19 March; accepted 9 July 1991.


1. Winter, G. & Milstein, C. *Nature* **349**, 293–299 (1991).
2. Smith, G. P. *Science* **228**, 1315–1317 (1985).
3. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. *Nature* **348**, 552–554 (1990).
4. Hoogenboom, H. R. *et al. Nucleic Acids Res.* **19**, 4133–4137 (1991).
5. Huse, W. D. *et al. Science* **246**, 1275–1281 (1989).
6. Caton, A. J. & Koprowski, H. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6450–6454 (1990); correction, **88**, 1590 (1991).
7. Mullinax, R. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 8095–8099 (1990).
8. Persson, M. A. A., Caothien, R. H. & Burton, D. R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2432–2436 (1991).
9. Saiki, R. K. *et al. Science* **239**, 487–491 (1988).
10. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. & Pease, L. R. *Gene* **77**, 61–68 (1989).
11. Huston, J. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 5879–5883 (1988).
12. Glockshuber, R., Malia, M., Pfitzinger, I. & Pluckthun, A. *Biochemistry* **29**, 1362–1367 (1990).
13. Dildrop, R. *Immun. Today* **5**, 85–86 (1984).
14. Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. *Sequences of Proteins of Immunological Interest* (US Department of Health and Human Services, US Government Printing Office, 1987).
15. Hudson, N. W., Mudgett-Hunter, M., Panka, D. J. & Margolies, M. N. *J. Immun.* **139**, 2715–2723 (1987).
16. Berek, C., Griffiths, G. M. & Milstein, C. *Nature* **316**, 412–418 (1985).
17. Alzari, P. M. *et al. EMBO J.* **9**, 3807–3814 (1990).
18. Perelson, A. S. *Immunol. Rev.* **110**, 5–33 (1989).
19. Orlandi, R., Güssow, D. H., Jones, P. T. & Winter, G. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3833–3837 (1989).
20. Bass, S., Greene, R. & Wells, J. A. *Proteins* **8**, 309–314 (1990).
21. Cwirla, S. E., Peters, E. A., Barrett, R. W. & Dower, W. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6378–6382 (1990).
22. Devlin, J. J., Panganiban, L. C. & Devlin, P. E. *Science* **249**, 404–406 (1990).
23. Scott, J. K. & Smith, G. P. *Science* **249**, 386–390 (1990).
24. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1989).
25. Dower, W. J., Miller, J. F. & Ragsdale, C. W. *Nucleic Acids Res.* **16**, 6127–6145 (1988).
26. Nagai, K. & Thøgersen, H. C. *Meth. Enzym.* **153**, 461–481 (1987).
27. Mäkelä, O., Kaartinen, M., Pelkonen, J. L. T. & Karjalainen, K. *J. exp. Med.* **148**, 1644–1660 (1978).
28. Gibson, T. *thesis*, Univ. Cambridge (1984).
29. Gherardi, E., Pannell, R. & Milstein, C. *J. Immunol. Meth.* **126**, 61–68 (1990).
30. Clackson, T., Güssow, D. & Jones, P. T. in *PCR: a Practical Approach* (eds McPherson, M. J., Taylor, G. R. & Quirke, P.) (IRL, Oxford, in the press).
31. Ward, E. S., Güssow, D., Griffiths, A. D., Jones, P. T. & Winter, G. *Nature* **341**, 544–546 (1989).
32. Güssow, D. & Clackson, T. *Nucleic Acids Res.* **17**, 4000 (1989).
33. Sanger, F., Nicklen, S. & Coulson, H. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
34. Eisen, H. N. *Meth. med. Res.* **10**, 115–121 (1964).
35. Segal, I. *Enzyme Kinetics* 73–74 (Wiley-Interscience, New York, 1975).
36. Foote, J. & Milstein, C. *Nature* (in the press).
37. Better, M., Chang, C. P., Robinson, R. R. & Horwitz, A. H. *Science* **240**, 1041–1043 (1988).
38. Munro, S. & Pelham, H. R. B. *Cell* **46**, 291–300 (1986).
39. Shimatake, H. & Rosenberg, M. *Nature* **292**, 128–132 (1981).
40. Gottesman, M. E., Adhya, S. & Das, A. *J. molec. Biol.* **140**, 57–75 (1980).
41. McManus, S. & Riechmann, L. *Biochemistry* **30**, 5851–5857 (1991).
42. Vieira, J. & Messing, J. *Meth. Enzym.* **153**, 3–11 (1987).
43. De Bellis, D. & Schwartz, I. *Nucleic Acids Res.* **18**, 1311 (1990).
44. Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1988).


ACKNOWLEDGEMENTS. We thank J. McCafferty for help in constructing the vector fdDOG1; J. Marks for the vector pJM1; C. Marks for oligonucleotide primers for cDNA synthesis and for D1.3 Fv protein; E. Gherardi for spleen mRNA; J. Foote for help with fluorescence quench titrations; and M. Dreher, L. Riechmann, J. Jarvis, D. Chiswell, A. R. Fersht and C. Milstein for advice. Supported by the MRC, the D. Collen Research Foundation, Leuven, and the European Molecular Biology Organisation (H.R.H), and by the Louis Jeantet Foundation (A.D.G. and T.C.).

# Phosphorylation-regulated $Cl^-$ channel in CHO cells stably expressing the cystic fibrosis gene

**Joseph A. Tabcharani*, Xiu-Bao Chang†, John R. Riordan† & John W. Hanrahan*‡**

* Department of Physiology, McGill University, 3655 Drummond Street, Montréal, Québec H3G 1Y6, Canada
† Research Institute, Hospital for Sick Children, and University of Toronto, Toronto, Ontario M5G 1X8, Canada

**A CYCLIC AMP-stimulated choride conductance appears when the cystic fibrosis gene is expressed in non-epithelial cells by infection with recombinant viruses[1,2]. Cyclic AMP-stimulated conductance in this system is mediated by the same ohmic, low-conductance $Cl^-$ channel as in human secretory epithelia[2-4], but control of this channel by phosphorylation has not been directly demonstrated. Here we report the appearance of the low-conductance $Cl^-$ channel in Chinese hamster ovary cells after stable transfection with the cystic fibrosis gene. The channel is regulated on-cell by membrane-permeant analogues of cAMP and off-cell by protein kinases A and C and by alkaline phosphatase. These results are further evidence that the cystic fibrosis transmembrane regulator is a $Cl^-$ channel which can be activated by specific phosphorylation events and inactivated by dephosphorylation; they reveal an unsuspected synergism between converging kinase regulatory pathways.**

The coding sequence of the cystic fibrosis transmembrane regulator (CFTR) was cloned behind the metallothionein promoter of a plasmid that also contained a mutant dihydrofolate reductase gene, driven by the simian virus 40 early promoter (Fig. 1*a*). Stably transformed colonies were selected with methotrexate after calcium phosphate transfection of Chinese hamster ovary (CHO)-K1 cells. CFTR-expressing variants were chosen for further study on the basis of their capacity to produce a protein of the same apparent size as that present in T84 cell membranes in western blots probed with monoclonal antibodies against CFTR (Fig. 1*b*). CFTR protein was localized in a highly enriched plasma membrane vesicle fraction. In variants containing nearly the same amount of CFTR as T84 cells, cAMP-regulated chloride permeability, as monitored by $^{125}I$ efflux, was indistinguishable from that in T84 cells (Fig. 1*c*).

Patch-clamp recording was used to identify the channel responsible for cAMP-stimulated $^{125}I$ efflux. Channels became active in cell-attached patches after a lag of $69 \pm 26$ seconds when cells were exposed to membrane-permeant derivatives of cAMP; this was reversed by washing cAMP from the bath ($n = 7$, Fig. 2*a*). The channel was observed in 80% of all seals during cAMP stimulation (225/282) at an average density of between five and ten channels per patch. By contrast, it was recorded only once in 55 patches on unstimulated, CFTR-transfected cells and was never observed on cAMP-stimulated CHO cells that had been transfected with vector alone (0/31). Figure 2*b* shows that open probability was relatively independent of voltage, despite increased flickering at hyperpolarized potentials. Flickering was not observed using excised patches (see below), therefore these brief closures may reflect voltage-dependent, fast channel block by some anion in the cytosol. The current–voltage relationship rectified slightly in the outward direction during cell-attached recordings (Fig. 2*c*), but was linear ($r^2 = 0.9997$) when patches were excised and bathed symmetrically with 154 mM $Cl^-$ (data not shown). In cell-attached patches the reversal potential ($E_{rev}$) was near the membrane potential ($0.6 \pm 0.3$ mV applied potential) and the slope conductance at $E_{rev}$ was $9.6 \pm 0.5$ pS ($n = 5$). The $E_{rev}$ shifted to $+32.4 \pm 2.0$ mV when the pipette solution contained 110 mM sodium gluconate and

‡ To whom correspondence should be addressed.

The EMBO Journal vol.13 no.3 pp.692–698, 1994

# Antibody fragments from a 'single pot' phage display library as immunochemical reagents

**Ahuva Nissim[1], Hennie R.Hoogenboom[2], Ian M.Tomlinson[1,3], Greg Flynn[4], Carol Midgley[5], David Lane[5] and Greg Winter[1,3]**

[1]MRC Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, [2]Cambridge Antibody Technology, The Science Park, Melbourn, Cambridgeshire SG8 6EJ, [3]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK, [4]Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA and [5]Department of Biochemistry, University of Dundee, Dundee DD1 4HN, UK

Communicated by G.Winter

**The display of repertoires of antibody fragments on the surface of filamentous bacteriophage offers a new way of making antibodies with predefined binding specificities. Here we explored the use of this technology to make immunochemical reagents to a range of antigens by selection from a repertoire of $>10^8$ clones made *in vitro* from human V gene segments. From the same 'single pot' repertoire, phage were isolated with binding activities to each of 18 antigens, including the intracellular proteins p53, elongation factor EF-1$\alpha$, immunoglobulin binding protein, rhombotin-2 oncogene protein and sex determining region Y protein. Both phage and scFv fragments secreted from infected bacteria were used as monoclonal and polyclonal reagents in Western blots. Furthermore the monoclonal reagents were used for epitope mapping (a new epitope of p53 was identified) and for staining of cells. This shows that antibody reagents for research can be readily derived from 'single pot' phage display libraries.**

*Key words*: immunochemical reagents/immunoglobulins/phage/scFv

## Introduction

Animal immunization has provided a wealth of valuable antisera and monoclonal antibodies as research reagents. However, the display of small peptides (Smith, 1985) and folded proteins (Bass *et al.*, 1990; McCafferty *et al.*, 1990) on filamentous phage, offers another source of such reagents. Thus antibody fragments have been made in bacteria by selection from repertoires of variable domains displayed on the surface of filamentous bacteriophage (for review see Hoogenboom *et al.*, 1992; Marks *et al.*, 1992a). Heavy and light chain variable domains are displayed by fusion to the viral coat protein, allowing phage with antigen binding activities (and encoding the antibody fragments) to be selected by panning on antigen. The selected phage can be grown after each round of panning and selected again, and rare phage ($<1/10^7$) isolated over several rounds of panning (McCafferty *et al.*, 1990). By interposing an amber mutation between the antibody fragment and coat protein, the fragments can also be secreted from infected bacteria as soluble fragments (Hoogenboom *et al.*, 1991). As in the immune system, selected antibodies from the primary library can be mutated, and those with improved affinities isolated (Hawkins *et al.*, 1992; Marks *et al.*, 1992b).

The repertoires of fragments, displayed on phage as single chain Fv fragments (McCafferty *et al.*, 1990) or Fab fragments (Hoogenboom *et al.*, 1991), are encoded by rearranged V genes. Diverse repertoires of heavy ($V_H$) and light ($V_L$) chain V genes can be derived from populations of lymphocytes (Sastry *et al.*, 1989; Ward *et al.*, 1989) using the polymerase chain reaction (PCR) (Saiki *et al.*, 1985; Orlandi *et al.*, 1989), and paired at random to encode diverse repertoires of fragments (Huse *et al.*, 1989). Immunization enriches the mRNA for $V_H$ and $V_L$ genes encoding antigen binding activities (Burton, 1991; Winter and Milstein, 1991), but requires a different library for each antigen.

By contrast, the use of larger and more diverse repertoires, either from V genes rearranged *in vivo* (Marks *et al.*, 1991a) or *in vitro* (Hoogenboom and Winter, 1992), can allow the isolation of antibodies of different binding specificities, without immunization, from the same library, as proposed by Milstein (1990). Such libraries have been used to produce antibodies against haptens, and foreign and self antigens (Marks *et al.*, 1991a; Hoogenboom and Winter, 1992; Griffiths *et al.*, 1993). Here we have explored the use of the same 'single pot' library to isolate antibodies with specificities that have proved difficult by hybridoma technology, for example against highly conserved intracellular proteins, and to use the phage and the encoded scFv fragments as immunochemical reagents.

## Results

### *Selection of clones of different specificities*

We built diverse repertoires of rearranged $V_H$ genes *in vitro* from a bank of 50 cloned human $V_H$ gene segments (Tomlinson *et al.*, 1992) and random nucleotide sequence encoding CDR3 lengths of 4–12 residues. The library was subjected to four or five rounds of selection against 18 antigens, including haptens, foreign and self-antigens, as well as intracellular antigens (Figure 1). Phage or secreted scFv fragments were screened for binding to the selecting antigen by ELISA, as a 'polyclonal' population and also as isolated clones. In all cases, phage with binding activities were detected to the selecting antigen. For some clones (see legend to Table I), binding activities were detected as phage but not as scFv fragments using supernatants from infected bacteria.

To obtain a diverse set of fragments for each antigen, clones with binding activities were screened for $V_H$ family by hybridization, and for CDR3 length by PCR (see Materials and methods). Different clones were selected for sequencing. A range of CDR3 loop sizes were detected for several antigens, for example, loops of four, five, nine and



© Oxford University Press

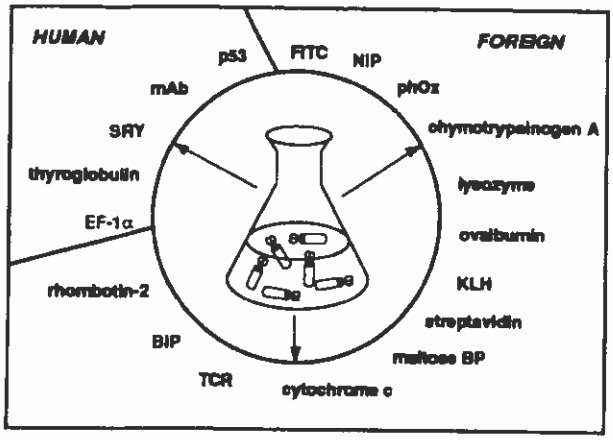



**Fig. 1.** Specificities from a 'single pot' human synthetic library. FITC, fluorescein isothiocyanate; NIP, 4-hydroxy-5-iodo-3-nitrophenylacetyl; phOx, 2-phenyl-5-oxazolone; KLH, haemocyanin from keyhole limpet; maltose BP, maltose binding protein; TCR, soluble chimeric murine T cell receptor; BiP, recombinant rat immunoglobulin binding protein; EF-1α, human elongation factor-1α; SRY, human sex-determining region Y protein; mAb, anti-erythrocyte rhesus D antibody Oak-3; p53, human tumour supressor protein p53.

ten residues for the hapten NIP. Over all the antigens, $V_H$ segments were found from each of the $V_H$ families except the small $V_H2$ family, as were all the CDR3 loop lengths represented in the library (Table I). Eight residue CDR3 loops were frequent. As *in vivo* (Zouali and Theze, 1991), segments of the $V_H3$ family (35/54) were well represented.

### *Western blots*

Phage and scFv fragments were used as 'polyclonal' and monoclonal reagents for Western blots. 'Polyclonal' phage, prepared by PEG precipitation, were used in Western blotting (Towbin *et al.*, 1979) to detect a panel of purified antigens on nitrocellulose after PAGE (Figure 2A); anti-EF-1α phage were also used to detect EF-1α in COS cell lysates (not shown). Similarly 'polyclonal' scFv fragments, concentrated 10-fold by dialysis against PEG from the bacterial supernatants, were used to detect the panel of antigens (Figure 2B); and to detect p53 in lysates of the breast cancer (T47D) cells with high levels of mutant human p53 protein (Bartek *et al.*, 1990). p53 was not detected in lysates of a human osteosarcoma cell line (SAOS-2) with a genetic deletion of both p53 genes (Masuda *et al.*, 1987) (Figure 3).

Monoclonal scFv fragments derived from segments of the

**Table I.** Deduced $V_H$-CDR3 sequences and germline origin of selected antigen-specific synthetic antibodies isolated from a single pot human synthetic library

| Self | CDR3 sequences | $V_H$ segment |
|---|---|---|
| anti-RhD mAb (Oak-3) | RRPTGHWH | DP-53 |
| | RHNGNWD | DP-53 |
| thyroglobulin | VARYNMYIPP | DP-49 |
| | SKGAWFVRPP | DP-49 |
| | MKSSARPV | DP-66 |
| EF-1α | VYPFRBTN | DP-51 |
| | INGKKFDY | DP-73 |
| p53 | NQNV | DP-45 |
| | NQHV | DP-45 |
| | FTRKRD (#30) | DP-53 |
| SRY | QNLHNNQLV | DP-2 |
| | RTPAVLSQB RNV | DP-42 |
| | MAIQT | DP-47 |
| | SWRGLMM | DP-70 |

| Haptens | CDR3 sequences | $V_H$ segment |
|---|---|---|
| NIP | VMSSADGHR | DP-38 |
| | RGMN | DP-42 |
| | KGGFD | DP-42 |
| | GGTR | DP-42 |
| | SGRD | DP-42 |
| | RRGPIGAYYP | DP-47 |
| phOx | TRAYRFDY | DP-47 |
| | LRNFSFDY | DP-47 |
| | RNAWGLIAARRT | DP-53 |
| FITC | SSKWSMRN | DP-7 |
| | ALQGPAYST | DP-35 |
| | EHRTSIPT | DP-42 |
| | LRARPV | DP-67 |

| Foreign | CDR3 sequences | $V_H$ segment |
|---|---|---|
| rhombotin-2 | QVYQA | DP-46 |
| | YTRKRFRRPER | DP-47 |
| recombinant | TRHARFDY | DP-14 |
| rat BiP | TWPTK (#17) | DP-32 |
| | GYPTPFNY (#2) | DP-33 |
| | GKRYFTK | DP-42 |
| | TRRSRFDY | DP-67 |
| | GRHSRFDY | DP-67 |
| mouse TCR | GKGPSYTAILSF | DP-10 |
| | SYYHLFDY | DP-46 |
| | GFAYIFDY | DP-74 |
| maltose BP | NWRNSFPQ | DP-14 |
| | SSNPWK | DP-14 |
| | ELMYFV | DP-48 |
| | TPIHRRRQFNTG | DP-53 |
| | YRTAHPL | DP-67 |
| turkey egg | SYNEIVPI | DP-3 |
| lysozyme | NYQLBVPQ | DP-3 |
| cytochrome *c* | RIGTPFDY | DP-3 |
| KLH | LRHGF | DP-45 |
| | RRHRAFDY | DP-47 |
| ovalbumin | GWNASD | DP-7 |
| chymotrypsinogen A | KKBI | DP-58 |
| | SKBI | DP-58 |
| streptavidin | DWRMIEG | DP-49 |
| | QWEGIRS | DP-69 |

The encoded sequences of $V_H$-CDR3, the third complementarity determining region of the antibody heavy chain are given. Nomenclature of the heavy chain germline segments is as in Tomlinson *et al.* (1992). The specificities were detected as phage and also in bacterial supernatants as soluble scFv fragments, except for anti-thyroglobulin, EF-1α, SRY, turkey egg lysozyme, ovalbumin, chymotrypsinogen A and streptavidin, which were initially detected only as phage antibodies. B represents translation of amber codon to glutamine in *E.coli* TG-1 *supE* strain. Abbreviations as in Figure 1.

229

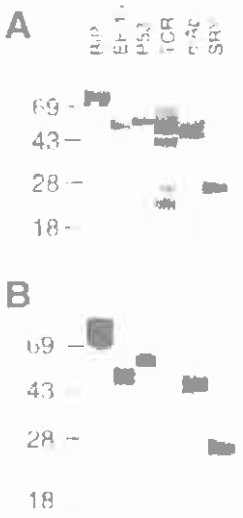



**Fig. 2.** Western blotting with polyclonal phage and fragments antibodies. 2 $\mu$g of purified antigens were electrophoresed under reducing conditions on 10% acrylamide gels, electroblotted to nitrocellulose and hybridized with polyclonal phage antibodies (A) or polyclonal scFvs (B). The expected molecular weights of BiP, EF-1$\alpha$, p53, TCR, mAb and SRY are respectively 78, 53, 53, 43, 50 ($\gamma$3 heavy chain) and 26 kDa.

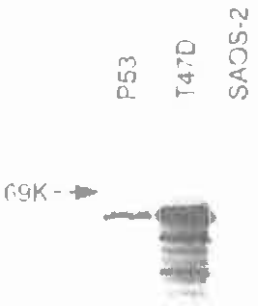



**Fig 3.** Detection of p53 in cell lysates by 'polyclonal' scFv fragments. Human p53 (2 $\mu$g) purified from SF9 insect cells infected with a recombinant baculovirus encoding wild type human p53 (Luckow and Summers, 1989; C.A.Midgley, unpublished), cell lysates ($10^5$ cells) of T47D human breast cancer cell line (overexpresses p53) and SAOS-2 human osteosarcoma cell line (in which both p53 genes are deleted).

human $V_H3$ family (and therefore binding to protein A) (Sasso *et al.*, 1991) were chosen for use as reagents, as they could be purified by binding to protein A–Sepharose (Hoogenboom and Winter, 1992), and as reagents readily detected with horseradish peroxidase (HRP)–protein A. (The fragments of other families, or indeed the $V_H3$ family, could be detected by their C-terminal peptide tag using the mouse monoclonal antibody 9E10 followed by HRP–goat anti-mouse IgG.) Analysis of the scFv fragments by gel filtration, for example those described here against p53 and BiP, indicated that the fragments were partly aggregated, as shown previously with scFv fragments derived from phage display repertoires (Griffiths *et al.*, 1993) or from monoclonal antibodies (Holliger *et al.*, 1993).

In lysates of COS cells, a purified monoclonal scFv

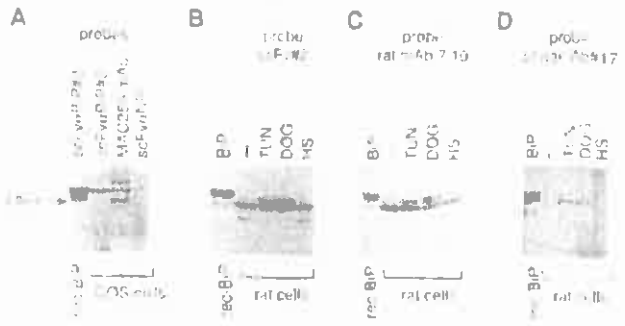



**Fig. 4.** Detection of hsp70 proteins by monoclonal anti-BiP scFv fragments and phage. (A) Detection of recombinant rat BiP and BiP in COS cell lysates with anti-BiP scFv#2, with controls of MAC256 monoclonal antibody (recognizing a KDEL motif shared by luminal ER proteins) and a scFv (directed against the hapten NP). (B–D) Specificity of anti-BiP scF fragments (B and C) or phage (D). Rat cells (RAT-1) were incubated for 19 h without drugs (–), or with 0.5 $\mu$g/ml tunicamycin (TUN), or with 10 mM 2-dioxy-D-glucose (DOG), or were heat shocked for 1 h at 42°C (HS). scFv#2 and rat mAb 7.10 recognize the hsp70 family. However, phage #17 reacts only with BiP, induced by TUN or DOG treatment but not by HS.

fragment (scFv$\alpha$BIP#2) against recombinant rat BiP (immunoglobulin binding protein, and identical to glucose regulated protein grp78) (Munro and Pelham, 1986) detected a band of ~70 kDa, corresponding to 'cognate' cytosolic stress-70 proteins (hsc70) (Figure 4A). Likewise in lysates of rat fibroblast (RAT-1) cells (Figure 4B) the reagent mainly detected a single band of ~70 kDa, as does the control 7.10 mAb (Palter *et al.*, 1986) which is known to bind to several members of the hsp70 protein family including 'cognate' hsc70, heat inducible hsp70 and BiP (Figure 4C). The scFv fragments appear to be as sensitive as monoclonal antibodies as reagents. A solution of 10 $\mu$g/ml of scFv#2 reagent readily detected 15 ng recombinant BiP spotted onto nitrocellulose, resulting in strong staining by ECL.

BiP is present at low levels in fibroblasts; it is not heat inducible but the rate of its synthesis is increased when cells are starved for glucose or when induced by glycosylation inhibitors such as tunicamycin or 2-dioxy-D-glucose (Pouyssegur *et al.*, 1977; Olden *et al.*, 1979). When RAT-1 cells were treated, both the scFv reagent and the 7.10 mAb detected the appearance of a second band of ~80 kDa, presumably BiP (Figure 4B and C). Thus the scFv fragments appear to have a similar specificity to mAb 7.10. However, we also isolated more specific fragments that recognized only BiP, and not hsc70 (or the heat inducible hsp70) (Figure 4D), as illustrated here with a monoclonal phage reagent #17.

### *Epitope mapping*

The epitope on p53 was mapped using monoclonal phage or scFv as a reagent (#30, CDR3: FTRKRD) Both phage and scFv fragment detected recombinant human p53 expressed in bacteria (Figure 5A), and also a series of p53 proteins deleted at the C-terminus (Figure 5B). This localized the epitope to the N terminus of p53. Mapping of two p53 proteins deleted at the N-terminus (Figure 5C) indicates that the epitope lies between residues 27 and 44 since a p53 fragment encoding amino acids 44–393 was not detected. The results were confirmed with the soluble scFv fragment reagent (not shown). The localization of epitope is consistent with the observation that the binding of the scFv fragment is not blocked by the DO-1 antibody (not shown), which

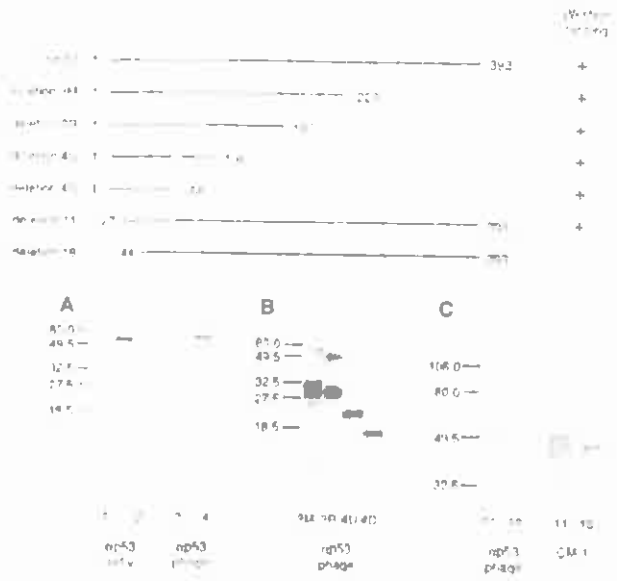



**Fig. 5.** Mapping the epitope in p53 recognized by an scFv fragment. The figure shows the portions of the p53 fragments expressed in bacteria and their reactivity with an anti-human p53 scFv fragment or corresponding phage on Western blots. (**A**) Western blot of lysate from induced *E.coli* containing pT7-7 vector only (lanes 1 and 3) or pT7-7Hup53 which direct the expression of full length human p53 (lanes 2 and 4). (**B** and **C**) Western blots of lysate from induced *E.coli* containing 3' (**B**) or 5' (C) truncated pT7-7Hup53. Anti-p53 phage antibodies detected all of the deleted proteins except deletion 18. CM-1 rabbit polyclonal anti-human p53 antibodies were used as positive control in (C).

recognizes an immunodominant site within the first 27 amino acids of the protein (D.Lane, unpublished data).

### *Immunocytochemistry*

Monoclonal anti-p53 scFv fragments (#30) were used to stain the breast cancer cell line T47D. Clear nuclear staining of the T47D cells was seen when the scFv was used at a concentration of 10 $\mu$g/ml. At this concentration only a very low level of non-specific staining was seen when the same reagents were applied to the human SAOS-2 cells (Figure 6) or mouse C6 cells (and indicating that the scFv fragment does not bind mouse p53). We failed to find conditions for specific staining of the nucleus with the anti-p53 phage; presumably the phage particle is too large to penetrate.

## Discussion

We have shown that phage with a range of binding specificities, can be isolated from a 'single pot' phage antibody library (Figure 1) within a few days and without animal immunization. We could isolate binding specificities to haptens, foreign and self-antigens, including secreted proteins (lysozyme and ovalbumin), cell surface proteins (T cell receptor), intracellular proteins from the cytoplasm (thyroglobulin, EF-1$\alpha$), proteins from the lumen of the endoplasmic reticulum (BiP) and proteins from the nucleus (p53, SRY and rhombotin-2).

Phage and scFv fragments from the single pot library could be used without further affinity maturation as immunological reagents for Western blotting, epitope mapping and cell staining. We also used phage and scFv fragments in monoclonal and polyclonal formats. Polyclonal reagents are

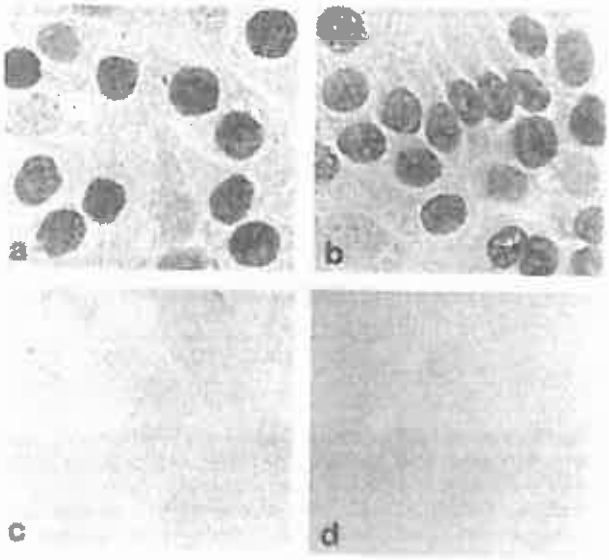



**Fig. 6.** Immunocytochemistry using monoclonal scFv. Staining of T47D (a and b) and SAOS-2 cells (c and d). (a) and (c) show staining with monoclonal anti-p53 scFv fragment, while (b) and (d) show staining by mouse monoclonal anti-p53 antibody (DO-1).

simpler to prepare, as the presence of binding phage can be readily detected by ELISA, and the phage population (or soluble scFv fragments) can then be used directly, or after affinity purification. In particular, scFv fragments derived from segments of the human $V_H3$ family (and therefore binding to protein A) could be readily purified with protein A–Sepharose, and as reagents readily detected with protein A–HRP conjugates.

Although for each of the antigens we could isolate phage clones with binding activities, we were sometimes unable to detect binding of the encoded soluble scFv fragments using bacterial supernatants. This may reflect the amplified detection of the phage by virtue of the 3000 subunits of the phage 'tail': if so, a more sensitive assay should be used for the soluble fragments. It might also reflect the greater binding avidities afforded by multivalent display on phage, and the moderate binding affinities of the monomeric scFv fragments from primary phage repertoires of $10^7$–$10^8$ clones (Marks *et al.*, 1991a; Hoogenboom and Winter, 1992; Griffiths *et al.*, 1993).

We found that the scFv fragments in supernatants could be more readily detected by concentrating the supernatant, or by 'multimerization' of the fragments during purification. Although scFv fragments can associate to form dimers in bacterial supernatants (Griffiths *et al.*, 1993; Holliger *et al.*, 1993), and the dimers have greater binding avidities to solid phase antigen (Griffiths *et al.*, 1993), purification can also drive the formation of multimers (Figure 7). Presumably multimerization is promoted by dissociation of $V_H$ and $V_L$ domains of the same chain, followed by pairing with another chain, especially during acid elution and neutralization in concentrated solutions. Indeed the purified scFv fragments used to detect p53 and BiP in Western blotting, epitope mapping and cell staining were partly aggregated (Figure 7).

Aggregation of 'moderate affinity' scFv fragments is probably important for their use as immunological reagents as it enhances the avidity of binding and slows the rate of dissociation during washing. Clearly antigen should be provided in multimeric form, for example immobilized on



231

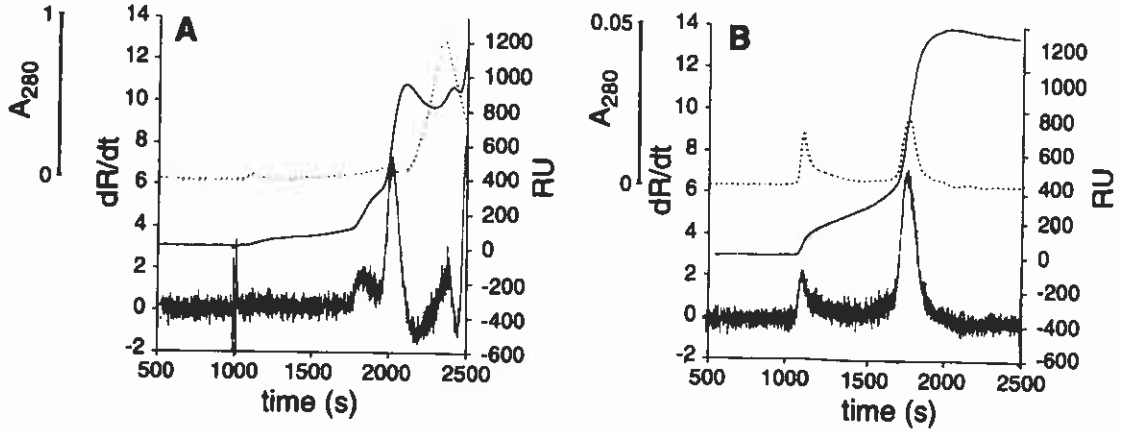



Fig. 7. Combined FPLC/BIAcore analysis of anti-BiP #2 scFv. Bacterial culture supernatant (A) and protein A affinity purified anti-BiP scFv #2 (B) were fractionated on a Superdex 200 column, with the eluent monitored by both UV 280 nm absorbance (upper, dotted line) and on-line BIAcore with immobilized 9E10 on the sensor chip (middle curve, light solid line). FPLC standards: of hen egg lysozyme (2455 s; $M_r$ 14.3 kDa), hen ovalbumin (1900 s; 45 kDa), BSA monomer (1780 s; 68 kDa) and BSA dimer (1590 s; 136 kDa), were used to calibrate the FPLC column. The derivatized sensorgram (dR/dt, lower, heavy solid line) illustrates the rate of changes in mass of protein bound to the sensor chip as a function of time. Note that the various *y*-axes have been scaled with zero at different levels for clarity. In (A), the main peak of binding near 2020 s corresponds to monomeric scFv, with a small peak of more aggregated material at 1830 s. The disturbance near 1000 s is an artifact caused by pump filling; the exclusion volume corresponds to 1100 s on the traces shown. In (B) there is substantial material at or near the exclusion volume (1120 s); the main peak (1750 s) most closely corresponds to a mixture of trimer and tetramer, with only minor amounts of dimer and monomer.

nitrocellulose as in Western blots, or in fixed cells, and at high density to allow crosslinking with multimeric antibody. Furthermore the use of high density antigen should also facilitate the rebinding of multimeric and even monomeric fragments, slowing the net rate of dissociation from the solid phase. It would therefore be desirable to drive the self-aggregation of scFv fragments, for example by use of dimerization peptides (Pack and Pluckthun, 1992) or by shortening the flexible linker (Holliger *et al.*, 1993). However, scFv fragments are also aggregated by binding to protein A or to the bivalent mouse monoclonal antibody 9E10.

Not only have the single pot libraries provided a range of immunochemical reagents, but monoclonal specificities that are difficult to make by hybridoma technology, for example against the elongation factor 1α (EF-1α) (Merrick *et al.*, 1993) and immunoglobulin heavy chain binding protein (BiP). BiP is located in the lumen of the endoplasmic reticulum (see Gething and Sambrook, 1992), and is a member of a family of hsp70-related proteins that includes cytosolic 'cognate' hsc70 (p73) and heat inducible hsp70 (p72). BiP binds to partially folded or unassembled proteins, blocking their transport through the endoplasmic reticulum, and is highly conserved between species. Only one monoclonal antibody specific for BiP (Bole *et al.*, 1986), has been reported, and other less specific monoclonal antibodies have been used as reagents. For example, rat monoclonal antibody (7.10) raised against *Drosophila* hsp70 also binds to BiP (and to hsp70s and hsp-70-related proteins from a wide range of species) (Palter *et al.*, 1986), and antibodies against the C-terminal (KDEL) tetrapeptide of BiP, also bind to other KDEL proteins (including grp94 and protein disulfide isomerase) (Pelham, 1989). Presumably antibodies against BiP are retained in the lumen of the endoplasmic reticulum of B-cells and are difficult to display or secrete (Bole *et al.*, 1986). However, from the single pot library, we could readily isolate an anti-BiP reagent that reacts specifically with rat BiP in rat cell lysates, and also a more cross-reactive reagent that recognizes hsc70 (and with similar specificity to rat 7.10).

There is also a hint that the phage may recognize different epitopes, or at least those with a different bias, from monoclonal antibodies. For example, mouse antibodies to human p53 bind to two main immunodominant regions of the protein, at the N-terminal region and at the C-terminal region (Schlichtholz *et al.*, 1992). The immunodominant region lies between amino acids 1 and 25 at the N terminus or amino acids 370 and 381 at the C terminus (D.Lane, unpublished data); the epitope recognized by the anti-p53 scFv fragment appears to lie between residues 27 and 44. This portion of sequence differs significantly between mouse and human p53, and indeed as predicted the scFv reagent does not recognize mouse p53 in cell staining (or indeed by ELISA on purified mouse p53—not shown). However, it remains to be seen whether in general the phage libraries lead to a different spectrum of epitopes from natural immune systems. If so, such reagents could prove invaluable for mapping the functions of proteins.

In conclusion, single pot libraries are capable of providing immunochemical reagents of high specificity, and in future higher affinity fragments may emerge from larger libraries (Perelson and Oster, 1979; Waterhouse *et al.*, 1993), or by affinity maturation of selected antibodies by chain shuffling (Marks *et al.*, 1992b). The use of these libraries appears to open new possibilities for immunochemistry.

## Materials and methods

### *Library construction*

The human synthetic libraries (HSL) were constructed as in Hoogenboom and Winter (1992) using PCR primers designed to introduce $V_H$-CDR3 of random sequence varying in length from 4 to 12 residues. All 49 germline $V_H$ segments amplified in Hoogenboom and Winter (1992) were used except for DP-44 (as this produces an identical product to DP-45; see Tomlinson *et al.*, 1992), and two additional $V_H2$ segments (DP-26 and DP-28) were incorporated. Note that several of these segments are not seen as rearranged genes, and lie outside the main $V_H$ locus (unpublished data). DNA template encoding each of the 50 germline $V_H$ segments was individually amplified using the $V_H$ family based primers (Marks *et al.*, 1991a) VHBACKSfi and HSLP4-HSLP12 (Table II) in a volume of 50 μl with 250 μM dNTPs, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris–HCl (pH 8.8), 2 mM $MgCl_2$, 100 μg/ml BSA and 1 μl (1 unit) of Taq DNA

**Table II.** Oligonucleotide primers used to create the synthetic library

| | |
|---|---|
| HSLP | 5'-GAC CAG GGT ACC TTG GCC CCA [(A/C)NN]$_n$ *TCT TGC ACA GTA ATA CAC GGC CGT GTC* |
| JHSAL | 5'-GCC TGA ACC GCC TCC ACC AGT CGA CAC GGT GAC CAG GGT ACC TTG GCC CCA |
| CDRFOR | 5'-CAG GGT ACC TTG GCC CCA |
| CDRBACK | *5'-GTG TAT TAC TGT GCA AGA* |

Each germline $V_H$ segment (Tomlinson *et al.*, 1992) was amplified using $V_H$ family based primers (Marks *et al.*, 1991a) VHBACKSfi and HSLP. The HSLP primers were designed to introduce a J segment and a random D segment of 4–12 residues (A/C, N, N)$_n$ where $n = 4-12$. The 3' portion of the oligonucleotide was designed to anneal with FR3 of the germline $V_H$ gene segments. The JHSAL primers were designed to introduce a *SalI* restriction site (underlined) to the 3' end of each amplified germline $V_H$ segment for cloning, and the CDRFOR and CDRBACK primers to amplify CDR3. Regions of primers based in the J segment are marked in bold, and in italics for regions based in FR3

polymerase. Template was provided as $10^6$ bacteria of a stock infected with recombinant M13 encoding each $V_H$ segment. 25 cycles (94°C for 1 min, 55°C for 1 min and 72°C for 1.5 min) were used, with limiting VHBACKSfi and HSLP primers (0.25 pmol/$\mu$l) to exhaust the primers.

In a second PCR step, a *SalI* restriction site was introduced at the 3' end of the $J_H$ segment. 30 pmol of the same VHBACKSfi oligonucleotide and 30 pmol of JHSAL (Table II) were added, and the PCR continued for additional 15 cycles. After checking that a band of the appropriate size was observed on agarose gel electrophoresis for each of the 450 samples ($9 \times 50$), the PCR products of each of the amplifications encoding the nine different CDR3 loop lengths were pooled, cut with *NcoI* and *SalI*, and cloned into *NcoI–XhoI*-cut pHEN1-Vλ3 (Hoogenboom and Winter, 1992) to provide nine phagemid libraries, each of at least $10^7$ different clones.

Each library was rescued separately with helper phage VCS-M13 (Stratagene), then pooled with each other, and with two earlier libraries (Hoogenboom and Winter, 1992), each of at least $10^7$ different clones and encoding loops of five or eight residues (the latter with five random and C-terminal Phe-Asp-Tyr), to create a single pot library of $>10^8$ clones.

### *Selection of phage library*

Phage were panned for binding using immunotubes (Nunc; Maxisorp) coated with each antigen overnight at room temperature (Marks *et al.*, 1991; Griffiths *et al.*, 1993). The library was subjected to four or five rounds of selection on the various antigens.

The following antigens were coated at 100 $\mu$g/ml: human mAb ($\gamma$3, $\kappa$) anti-erythrocyte rhesus (D) monoclonal antibody (Oak-3, from Barbara D.Gorick; Bye *et al.*, 1992); human thyroglobulin (Hu-Thy; Scipac); 2-phenyl-5-oxazolone (phOx)-BSA; 4-hydroxy-5-iodo-3-nitrophenylacetyl (NIP)-BSA; turkey egg-white lysozyme (TEL; Sigma); horse heart cytochrome *c* (Sigma); keyhole limpet haemocyanin (KLH; Sigma); chicken egg ovalbumin (Ovalbumin; Sigma); α-chymotrypsinogen A (type II from bovine pancreas; Sigma); bovine serum albumin–L-fluorescein isothiocyanate (FITC-BSA; Sigma) and streptavidin from *Streptomyces avidinii* (Sigma). Coating was in 50 mM $NaHCO_3$ pH 9.6 (Oak-3, TEL, KLH, ovalbumin and chymotrypsinogen A), or PBS (human thyroglobulin, NIP, phOx, cytochrome *c*, FITC-BSA and streptavidin).

The following antigens were coated at 20 $\mu$g/ml: recombinant rat BiP (see below); human elongation factor 1α (EF-1α, from P.Kristensen; Merrick *et al.*, 1993); human p53 isolated from SF9 insect cells infected with a recombinant baculovirus encoding wild type human p53 (Luckow and Summers, 1989; C.A.Midgley, T.Hupp and D.Lane, unpublished); soluble chimeric murine T cell receptor (TCR, from K.Karjalainen: Weber *et al.*, 1992); protein encoded by the human testis determining gene, SRY (sex-determining region Y) from V.R.Harley (Harley *et al.*, 1992); fusion of rhombotin-2 to maltose binding protein (maltose BP) and recombinant mouse rhombotin-2 from T.Rabbitts (Foroni *et al.*, 1992). Coating was in PBS (BiP, TCR) or 50 mM $NaHCO_3$ pH 9.6 (p53, EF-1α, SRY, maltose BP and rhombotin-2).

The rat BIP was engineered with six N-terminal histidine residues and a factor X cleavage site preceding the mature BiP polypeptide (His-BiP). Oligonucleotide primers were used to amplify the coding region of BiP from the plasmid R76 (from H.Pelham). The amplified product was inserted into the *BamHI–HindIII* restriction sites of pQE40 vector (Qiagen) encoding the His6 tail, then transformed into *Escherichia coli* strain XL-1 Blue (Stratagene). Expression and purification of His6-tagged protein over Ni-NTA resin were performed by protocols provided by the manufacturer (Qiagen). After eluting from the Ni-NTN resin with 250 mM imidazole, 200 mM NaCl, 50 mM HEPES, pH 7.0, the His-BiP was further purified by ATP-agarose affinity chromatography (Flynn *et al.*, 1989). His-BiP was >95% pure.

### *Screening and sequencing of clones*

As described (Marks *et al.*, 1991; Hoogenboom and Winter, 1992; Griffiths *et al.*, 1993), phage were rescued from single ampicillin-resistant colonies of infected (suppressor) *E.coli* TG-1 using the helper phage VCS-M13 (Stratagene); soluble scFv fragments were induced from single colonies of infected (non-suppressor) *E.coli* HB2151 (Hoogenboom *et al.*, 1991) by IPTG. Bacterial supernatants containing phage or scFv fragments were screened for binding to antigen by ELISA. The binding specificity was checked by ELISA using plates coated with one or more other antigens: phage isolated by selection on 20 $\mu$g/ml antigens were screened against all the other antigens of this set. For the fusion proteins chimeric TCR–murine Cα and rhombotin–2-maltose BP, the specificity of the selected phage antibodies was further examined by ELISA using plates coated with mouse IgM, κ antibody and maltose BP, respectively.

Phage with binding activities were screened for CDR3 length by PCR and for $V_H$ family by hybridization. Individual phagemid colonies were amplified using $\gamma$-$^{32}$P-labelled CDR-BACK and CDR-FOR primers (Table II). The size of the amplified CDR3 was followed by gel electrophoresis using 7.6 M urea–8% polyacrylamide gels. Template from the entire phage library was used to produce a ladder of CDR3 lengths. Colonies were probed as in Tomlinson *et al.* (1992) for $V_H$ family using the family-specific oligonucleotide probes (Marks *et al.*, 1991b). Different clones, as shown by different CDR lengths in combination with $V_H$ segments of different families were sequenced by the dideoxy method (Sanger *et al.*, 1977) using DyeDeoxy chain termination (Applied Biosystems Inc.) and an Applied Biosystems 373A DNA sequencer. The sequences were analysed using SeqEd (Applied Biosystems Inc.) and MacVector 3.5 (IBI Kodak, New Haven, CT).

### *Preparation of monoclonal and polyclonal reagents*

For monoclonal scFv fragments, supernatants from 1 l cultures of infected HB2151 bacteria induced overnight with IPTG at 30°C (Griffiths *et al.*, 1993) were filtered with Sterivex-HV 0.45 $\mu$m filter unit (Millipore) and then purified on protein A–Sepharose (Pharmacia), as described in Hoogenboom and Winter (1992). Monoclonal phage was prepared by PEG precipitation of the supernatant of infected bacteria (Marks *et al.*, 1991). Polyclonal reagents were prepared from an aliquot of infected bacteria after the last round of selection. However, for the polyclonal scFv fragments, the supernatant was concentrated ~10-fold by dialysis against dry granular PEG-6000.

### *Western blotting*

Polyclonal reagents were used to detect purified antigen on Western blots as follows. 2 $\mu$g of each purified antigen were run on 10% polyacrylamide gel and then electroblotted. Filters were blocked for 1 h at room temperature in 5% Marvel/PBS/0.2% Tween 20 (for scFv fragments) or 10% Marvel/PBS/0.5% Tween 20 (for phage). The scFv fragments (1:2 dilution of 10 × concentrated supernatant) or phage ($10^{11}$ TU/ml) were added and incubated overnight with gentle shaking at 4°C. After washing with PBS–0.2% Tween 20 (or PBS–0.5% Tween 20 for phage), binding of scFv fragments was detected with a mixed reagent, HRP–protein A (Sigma P8651) mixed with the mouse monoclonal antibody 9E10, and followed by HRP–goat anti-mouse Ig (Sigma). Phages were detected with sheep anti-M13 1:1000 (provided by Cambridge Antibody Technology), followed by HRP–anti-goat antibody (Sigma). Peroxidase activity was detected using an ECL kit from Amersham.

Anti-p53 polyclonal scFv fragments (~10 $\mu$g/ml) were likewise used to detect p53 in T47D breast cancer cell lysates, and polyclonal anti-EF-1α phage antibodies ($10^{11}$ TU/ml) to detect EF-1α in COS cell lysates. Lysates were fractionated on 10% SDS–polyacrylamide gels (Laemmli, 1970), and electrophoretically transferred onto nitrocellulose (Towbin *et al.*, 1979).

For detection of BiP in cell lysates with monoclonal reagents, cell lysates were prepared from $5 \times 10^6$ COS cells grown at 37°C in DMEM/10% FCS, or from $5 \times 10^6$ RAT-1 fibroblast cells (Munro and Pelham, 1986; Napier *et al.*, 1992). The RAT-1 cells were treated with 0.5 $\mu$g/ml tunicamycin, 10 mM 2-dioxy-D-glucose or heat shock (Munro and Pelham, 1986). The filters were treated as above, and BiP detected either with 20



233

μg/ml monoclonal α-BiP scFv reagent followed by 10 μg/ml HRP–protein A; or with $10^{11}$ TU/ml of α-BiP phage (as above, except that the HRP was detected by 3,3′-diaminobenzidine in 0.02% cobalt sulfate). The rat 7.10 α-BiP mAb (ascites diluted 1:500; Cambridge Research Biochemicals), and the rat anti-KDEL mAb from J.Butcher (Napier *et al.*, 1992) were detected with 10 μg/ml anti-rat HRP.

### *Epitope mapping*

Plasmid pT7-7 Hup53 directs the expression of human p53 and deletion mutants truncated at the N- or C-terminus (Vojtesek *et al.*, 1992) in *E.coli* BL21 (DE3) cells. Total lysates from log phase cultures induced with IPTG were blotted much as above, except using 15% SDS–polyacrylamide gels. CM-1 polyclonal rabbit serum (1:300) against human p53, monoclonal α-p53 scFv fragment (10 μg/ml), or monoclonal α-p53 phage ($10^{11}$ TU/ml) were used to detect p53 using 10 μg/ml HRP–protein A (Sigma P8651), or (for the phage) 1:500 rabbit anti-F1 phage anti-serum followed by HRP–swine anti-rabbit Ig (Dako). Peroxidase activity was detected using an ECL kit from Amersham.

### *Immunocytochemistry*

Cell staining was performed as in Vojtesek *et al.* (1992). Briefly, T47D cells containing high levels of mutant human p53 protein, and SAOS-2 cells which contain a genetic deletion of both p53 genes, were passaged on plastic dishes and prefixed in cold methanol/acetone (1:1 by volume). A 2-fold dilution series (in PBS/10% fetal calf serum) of scFv fragment was spotted onto the fixed cells as 5 μl drops and incubated at room temperature for 2 h. The cells were washed three times with PBS then incubated with HRP-conjugated protein A (Sigma, 1 μg/ml). Alternatively, 9E10 was used followed by HRP–goat anti-mouse IgG (Sigma). The peroxidase activity was detected with 3,3′-diaminobenzidine in 0.03% nickel sulfate.

### *Sizing of scFv fragments by combined FPLC/BIAcore analysis*

Bacterial cultures of anti-BiP scFv #2 supernatants (0.5 ml of 0.2 μm-filtered) and protein A affinity-purified fragments (100 μl of 1.6 mg/ml) were analysed as described by Griffiths *et al.* (1993). The samples were fractionated on a Superdex S200 column (Pharmacia) and were monitored both by absorbance at 280 nm and on-line by BIAcore with immobilized 9E10 on the sensor chip (Johnsson *et al.*, 1991). The buffer was PBS containing 0.2 mM EDTA and 0.05% $NaN_3$. The BIAcore reading was corrected for the baseline, and also then calculated as dR/dt, the rate of change of the signal.

## Acknowledgements

We would like to thank S.Munro, G.Walter, A.D.Griffiths for advice, T.D.Prospero for FPLC/BIAcore analysis, and many colleagues for kindly providing antigens, as acknowledged in Materials and methods. A.N. is funded by Human Frontier Science Programme Organization (HFSPO), and C.M. and D.L. by the Cancer Research Campaign.

## References

Bartek,J., Iggo,R., Gannon,J. and Lane,D.P. (1990) *Oncogene*, **5**, 893–899.

Bass,S., Greene,R. and Wells,J.A. (1990) *Proteins*, **8**, 309–314.

Bole,D.G., Hendershot,L.M. and Kearney,J.F. (1986) *J. Cell Biol.*, **102**, 1558–1566.

Burton,D.R. (1991) *Trends Biotechnol.*, **9**, 168–175.

Bye,J.M., Carter,C., Cui,Y., Gorick,B.D., Songsivilai,S., Winter,G., Hughes-Jones,N.C. and Marks,J.D. (1992) *J. Clin. Invest.*, **90**, 2481–2490.

Flynn,G.F., Chappell,T.G. and Rothman,J.E. (1989) *Science*, **245**, 385–390.

Foroni,L., Boehm,T., White,A., Forster,A., Sherrington,P., Liao,X.B., Brannan,C.I., Jenkins,N.A., Copeland,N.G. and Rabbitts,T.H. (1992) *J. Mol. Biol.*, **226**, 747–761.

Gething,M.J. and Sambrook,J. (1992) *Nature*, **355**, 33–45.

Griffiths,A.D. *et al.* (1993) *EMBO J.*, **12**, 725–734.

Harley,V.R., Jackson,D.I., Hextall,P.J., Hawkins,J.R., Berkovitz,G.D., Sockanathan,S., Lovell-Badge,B. and Goodfellow,P.N. (1992) *Science*, **255**, 453–456.

Hawkins,R.E., Russell,S.J. and Winter,G. (1992) *J. Mol. Biol.*, **226**, 889–896.

Holliger,P., Prospero,T. and Winter,G. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6444–6448.

Hoogenboom,H.R. and Winter,G. (1992) *J. Mol. Biol.*, **227**, 381–388.

Hoogenboom,H.R., Griffiths,A.D., Johnson,K.S., Chiswell,D.J., Hudson,P. and Winter,G. (1991) *Nucleic Acids Res.*, **19**, 4133–4137.

Hoogenboom,H.R., Marks,J.D., Griffiths,A.D. and Winter,G. (1992) *Immunol. Rev.*, **130**, 41–68.

Huse,W.D., Sastry,L., Iverson,S.A., Kang,A.S., Alting,M.M., Burton,D.R., Benkovic,S.J. and Lerner,R.A. (1989) *Science*, **246**, 1275–1281.

Johnsson,B., Löfås,S. and Lindqvist,G. (1991) *Anal. Biochem.*, **198**, 268–277.

Laemmli,U. (1970) *Nature*, **227**, 680–685.

Luckow,V.A. and Summers,M.D. (1989) *Bio/Technology*, **6**, 47–55.

Marks,J.D., Hoogenboom,H.R., Bonnert,T.P., McCafferty,J., Griffiths,A.D. and Winter,G. (1991a) *J. Mol. Biol.*, **222**, 581–597.

Marks,J.D., Tristrem,M., Karpas,A. and Winter,G. (1991b) *Eur. J. Immunol.*, **21**, 985–991.

Marks,J.D., Hoogenboom,H.R., Griffiths,A.D. and Winter,G. (1992a) *J. Biol. Chem.*, **267**, 16007–16010.

Marks,J.D., Griffiths,A.D., Malmqvist,M., Clackson,T., Bye,J.M. and Winter,G. (1992b) *Bio/Technology*, **10**, 779–783.

Masuda,H., Miller,C., Koeffler,H.P., Battifora,H. and Cline,M.J. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 7716–7719.

McCafferty,J., Griffiths,A.D., Winter,G. and Chiswell,D.J. (1990) *Nature*, **348**, 552–554.

Merrick,W.C. (1979) *Methods Enzymol.*, **60**, 108–123.

Midgely,C.A., Fisher,C.J., Bartek,J., Vojtesek,B., Lane,D. and Barnes,D.M. (1992) *J. Cell Sci.*, **101**, 183–189.

Milstein,C. (1990) *Proc. R. Soc. Lond. Biol.*, **239**, 1–16.

Munro,S. and Pelham,H.R.B. (1986) *Cell*, **46**, 291–300.

Napier,R.M., Fowke,L.C., Hawes,C., Lewis,M. and Pelham,H.R.P. (1992) *J. Cell Sci.*, **102**, 261–271.

Olden,K., Pratt,R.M., Jaworski,C. and Yamamda,K.M. (1979) *Proc. Natl Acad. Sci. USA*, **76**, 791–795.

Orlandi,R., Gussow,D.H., Jones,P.T. and Winter,G. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 3833–3837.

Pack,P. and Pluckthun,A. (1992) *Biochemistry*, **31**, 1579–1584.

Palter,K.B., Wanatable,M., Stinson,L., Mahowald,A.P. and Craig,E.A. (1986) *Mol. Cell. Biol.*, **6**, 1187–1203.

Pelham,H.R.B. (1989) *Annu. Rev. Cell Biol.*, **5**, 1–23.

Perelson,A.S. and Oster,G.F. (1979) *J. Theor. Biol.*, **81**, 645–670.

Pouyssegur,J., Shiu,R.P.C. and Pastan,I. (1977) *Cell*, **11**, 941–947.

Saiki,R.K., Scharf,S., Faloona,F., Mullis,K.B., Horn,G.T., Erlich,H.A. and Arnheim,N. (1985) *Science*, **230**, 1350–1354.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.

Sasso,E.H., Silverman,G.J. and Mannik,M. (1991) *J. Immunol.*, **147**, 1877–1883.

Sastry,L., Alting,M.M., Huse,W.D., Short,J.M., Sorge,J.A., Hay,B.N., Janda,K.D., Benkovic,S.J. and Lerner,R.A. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 5728–5732.

Schlichtholz,B., Legros,Y., Gillet,D., Gaillard,C., Marty,M., Lane,D., Calvo,F. and Soussi,T. (1992) *Cancer Res.*, **52**, 6380–6384.

Smith,G.P. (1985) *Science*, **228**, 1315–1317.

Tomlinson,I.M., Walter,G., Marks,J.D., Llewelyn,M.B. and Winter,G. (1992) *J. Mol. Biol.*, **227**, 776–798.

Towbin,H., Staehelin,R. and Gordon,J. (1979) *Proc. Natl Acad. Sci. USA*, **76**, 4350–4354.

Vojtesek,B., Bartek,J., Midgley,C.A. and Lane,D.P. (1992) *J. Immunol. Methods*, **151**, 237–244.

Ward,E.S., Gussow,D., Griffiths,A.D., Jones,P.T. and Winter,G. (1989) *Nature*, **341**, 544–546.

Waterhouse,P., Griffiths,A.D., Johnson,K.S. and Winter,G. (1993) *Nucleic Acids Res.*, **21**, 2265–2266.

Weber,S., Traunecker,A., Oliveri,F., Gerhard,W. and Karjalainen,K. (1992) *Nature*, **356**, 793–796.

Winter,G. and Milstein,C. (1991) *Nature*, **349**, 293–299.

Zouali,M. and Theze,J. (1991) *J. Immunol.*, **146**, 2855–2864.

*Received on September 7, 1993*

ELSEVIER

Journal of Immunological Methods 263 (2002) 133–147

JIM
Journal of Immunological Methods

www.elsevier.com/locate/jim

Recombinant Technology

# Expression of full-length immunoglobulins in *Escherichia coli*: rapid and efficient production of aglycosylated antibodies

Laura C. Simmons [a,1], Dorothea Reilly [b,1], Laura Klimowski [a,2], T. Shantha Raju [c], Gloria Meng [d], Paul Sims [d], Kyu Hong [d], Robert L. Shields [e], Lisa A. Damico [f], Patricia Rancatore [g], Daniel G. Yansura [a,*]

[a] *Department of Molecular Biology, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA*
[b] *Department of Cell Culture and Fermentation Research and Development, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA*
[c] *Department of Analytical Chemistry, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA*
[d] *Department of Bioanalytical Technology, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA*
[e] *Department of Immunology, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA*
[f] *Department of Pharmacokinetics and Metabolism, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA*
[g] *Department of Recovery Sciences, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA*

Received 2 September 2001; received in revised form 14 February 2002; accepted 14 February 2002

**Abstract**

Many research and clinical applications require large quantities of full-length antibodies with long circulating half-lives, and production of these complex multi-subunit proteins has in the past been restricted to eukaryotic hosts. In this report, we demonstrate that efficient secretion of heavy and light chains in a favorable ratio leads to the high-level expression and assembly of full-length IgGs in the *Escherichia coli* periplasm. The technology described offers a rapid, generally applicable and potentially inexpensive method for the production of full-length therapeutic antibodies, as verified by the expression of several humanized IgGs. One *E. coli*-derived antibody in particular, anti-tissue factor IgG1, has been thoroughly evaluated and has all of the expected properties of an aglycosylated antibody, including tight binding to antigen and the neonatal receptor. As predicted, the protein lacks binding to C1q and the FcγRI receptor, making it an ideal candidate for research purposes and therapeutic indications where effector functions are either not required or are actually detrimental. In addition, a limited chimpanzee study suggests that the *E. coli*-derived IgG1 retains the long circulating half-life of mammalian cell-derived antibodies. © 2002 Elsevier Science B.V. All rights reserved.

*Keywords:* *E. coli* expression; Antibody production; Aglycosylated antibodies

## 1. Introduction

Monoclonal antibodies have recently become promising therapeutic proteins (King and Adair, 1999), in addition to their conventional use as research tools. These proteins can be targeted to almost any extracellular or cell surface protein and, through the

* Corresponding author. Tel.: +1-650-225-1412; fax: +1-650-225-6412.
*E-mail address:* yansura@gene.com (D.G. Yansura).
[1] LCS and DR contributed equally to this work.
[2] Current address: Department of Pharmacology and Toxicology, University of Utah, Salt Lake City, UT 84112, USA.

0022-1759/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0022-1759(02)00036-4

simple act of binding, promote the blocking or activation of specific biochemical steps (Hori, 1991; Kim et al., 1993). Additionally, antibodies can couple their antigen to natural effector functions (Dyer et al., 1989; Reff et al., 1994) or perhaps provide another activity through conjugation (Hinman et al., 1993; Liu et al., 1996). One significant advantage of full-length IgGs is their long circulating half-life in mammals (Mariani and Strober, 1990), the result of both a large molecular size preventing clearance in the kidneys, and the ability of these proteins to avoid proteolysis in the endothelium by using a salvage pathway (Junghans, 1997). This pathway plays a major role in the slow clearance of IgGs and depends on binding of the immunoglobulin Fc-domain to the neonatal receptor (FcRn) (Simister and Mostov, 1989).

Full-length monoclonal antibodies have traditionally been produced in mammalian cell culture due to their original hybridoma source and due to the complexity of the molecule (King, 1998). More recently, transgenic plants offer an alternative route and potentially a more economical system of production (Conrad and Fiedler, 1998; Giddings et al., 2000). Generally, *Escherichia coli* is the host system of choice for the expression of antibody fragments such as Fvs, scFvs, Fabs or $F(ab')_2$s (Pluckthun et al., 1996; Kipriyanov and Little, 1999). These fragments can be made relatively quickly in large quantities with the retention of antigen binding activity. However, because antibody fragments lack the Fc domain, they do not bind the FcRn receptor and are cleared quickly; thus, they are only occasionally suitable as therapeutic proteins (Knight et al., 1995). Attempts to extend the heavy chain beyond the hinge have been successful with the expression of a $C_H2$-deleted antibody (Lo et al., 1992). The addition of the $C_H3$ domain promotes dimerization of the antigen binding arms, but the lack of the $C_H2$ voids any chance of the protein binding the FcRn receptor (Yasmeen et al., 1976). Full-length antibody chains can also be expressed in *E. coli* as insoluble aggregates and then refolded in vitro (Boss et al., 1984; Cabilly et al., 1984), but the complexity of this method limits its usefulness.

Although the assembly of full-length antibodies in *E. coli* has not been reported, two reasons suggested the feasibility of this approach. First, fully functional antibody fragments through the hinge are routinely expressed as $F(ab')_2$s (Rodrigues et al., 1993; Koumenis et al., 2000), and second, the Fc fragment including the hinge has also been successfully produced as a dimer in *E. coli* (Kim et al., 1994). The aglycosylated Fc fragment bound to immobilized protein A, and displayed equivalent in vivo clearance to a full-length mammalian-derived glycosylated IgG. This data correlates well with previous studies showing that glycosylated and aglycosylated IgGs have equivalent in vitro FcRn binding and in vivo half-life in mammals (Tao and Morrison, 1989; Hobbs et al., 1992). Since both halves of a full-length antibody can be expressed and assembled independently in *E. coli*, the folded full-length molecule itself could likely be produced, with advances in technology.

This study demonstrates that a full-length anti-tissue factor (αTF) IgG1 (Presta et al., 2001) can be expressed at high levels in *E. coli*, produced in large-scale fermentors, and captured on immobilized protein A. In vitro, the purified protein binds tightly to both the antigen and the FcRn receptor and, as predicted, fails to bind C1q or FcγRI indicating a lack of effector functions. Additionally, the circulating half-life of the *E. coli*-derived IgG1, determined in chimpanzees, appears similar to glycosylated αTF IgG2 and IgG4 produced in mammalian cells.

## 2. Materials and methods

### 2.1. Plasmid construction

Expression cassettes were cloned into the framework of pBR322 (Sutcliffe, 1978) at the EcoRI site using standard methods (Chang et al., 1987). All constructions contained at least one *phoA* promoter (Kikuchi et al., 1981) and at least one lambda $t_0$ transcriptional terminator (Scholtissek and Grosse, 1987). Additionally, the STII signal sequence (Picken et al., 1983) or silent codon variants thereof (Simmons and Yansura, 1996) preceded the coding sequence for both light and heavy chains.

### 2.2. Small scale inductions and SDS-PAGE

Transformants from strain 33D3 (W3110 $kan^R$ ΔfhuA (ΔtonA) ptr3 lacIq lacL8 ompTΔ (nmpc-fepE)-

degP) were inoculated into 5 ml Luria–Bertani medium plus carbenicillin (50 ug/ml) and grown overnight at 30 °C. Each culture was then diluted (1:50 or 1:100) into C.R.A.P. phosphate-limiting media (3.57 g $(NH_4)_2SO_4$, 0.71 g Na citrate–$2H_2O$, 1.07 g KCl, 5.36 g Yeast Extract (certified), 5.36 g HycaseSF-Sheffield, pH adjusted with KOH to 7.3, volume adjusted to 872 ml with deionized $H_2O$ and autoclaved; cooled to 55 °C and supplemented with 110 ml 1 M MOPS pH 7.3, 11 ml 50% glucose, 7 ml 1 M $MgSO_4$) plus carbenicillin (50 ug/ml) and grown for approximately 24 h at 30 °C. Whole cell lysates from induced cultures (equivalent O.D. 600) were prepared and analyzed by SDS-PAGE (12% Tris–Glycine gels; NOVEX, San Diego, CA). Non-reduced samples were processed with 10 mM iodoacetic acid (Sigma I-2512) to prevent disulfide shuffling. The gels were then used for immunoblots.

### 2.3. Immunoblots

Following SDS-PAGE, samples from induced cultures were analyzed by immunoblot using standard methods. The protein bands bound by the anti-Fab antibody (peroxidase-conjugated goat IgG fraction to human IgG Fab; CAPPEL #55223) were visualized using an Amersham Pharmacia Biotech ECL detection kit and exposing the membrane to X-ray film.

### 2.4. N-terminal sequence analysis of small scale produced proteins

N-terminal sequence analysis of samples from small scale induced cultures were performed as described previously (Simmons and Yansura, 1996).

### 2.5. Fermentation

A 500 ml LB culture, grown for approximately 16 h at 30 °C with shaking, was used to inoculate a 10-l fermentor. The fermentor medium initially contained 8.9 mM glucose, 14.3 mM $MgSO_4$, 143 μM $FeCl_3$,46 μM $ZnSO_4$, $CuSO_4$, $H_3BO_3$, 42 μM $CoCl_2$, $NaMoO_4$, $MnSO_4$, either 15–20 mg/l tetracycline or 50–70 mg/l ampicillin, 40.6 mM $(NH_4)_2SO_4$, 16 mM $K_2HPO_4$, 10 mM $NaH_2PO_4$, 3.6 mM sodium citrate, 12 mM $KH_2PO_4$, 29 g/l casein hydrolyasate, and 14.3 g/l yeast extract. Fermentations were performed at 30 °C, 650 rpm, and 2 vvm aeration. Concentrated ammonium hydroxide was added to maintain the pH at 7.0.

Following inoculation of the fermentor, the culture was grown to high cell densities using a computer-based algorithm to feed a concentrated glucose solution. When the culture reached approximately 20 OD550, 1250 ml of a salt solution containing 12.5 g $(NH_4)_2SO_4$, 32.5 g $K_2HPO_4$, 16.25 g $NaH_2PO_4$, 2.5 g sodium citrate dihydrate, 18.75 g $KH_2PO_4$, 540 mg $FeCl_3$, 80mg $ZnSO_4$, $CuSO_4$, 70mg $CoCl_2$, $NaMoO_4$, 20 mg $H_3BO_3$, and 50 mg $MnSO_4$, was slowly fed to the culture. At a cell density of approximately 40 OD550, 100 ml of 1 M magnesium sulfate was added. Fermentations were typically continued for 70–80 h.

### 2.6. Purification

Fermentation cell paste was diluted 1:5 (w/v) in 20 mM sodium phosphate pH 7.4, 0.14 M NaCl, then lysed using an M110Y microfluidizer (Microfluidics, Newton, MA). The cell lysate was clarified by centrifugation (4300 × *g*, 30 min) and polyethylene imine (BASF, Rensselaer, NY) was added to the supernatant to a final concentration of 0.2%, followed by a second centrifugation. After filtration (0.2 μm), the supernatant was applied to a Protein A affinity resin (Prosep A Millipore, Bedford, MA) and eluted using 0.1 M acetic acid pH 2.9. The Protein A pool was conditioned by the addition of urea to a final concentration of 2 M, adjusted to pH 5.5, then diluted with purified water and applied to SP Sepharose FF (Amersham Pharmacia Biotech, Uppsala, Sweden). The remaining process followed a previously described method for the purification of CHO-derived antibodies (Presta et al., 2001).

### 2.7. Chemical analyses

MALDI-TOF-MS, cation-exchange chromatography, amino acid and N-terminal analysis were performed on the purified *E. coli*-derived αTF using standard methods. Size exclusion chromatography was performed on purified samples using a 7.8 mm × 30 cm TosoHaas TSKgel G3000SWXL column. The column was run at room temperature with

a mobile phase of 50 mM sodium phosphate buffer, pH 6.1, at a flow rate of 0.5 ml/min. Protein was detected at 280 nm.

### 2.8. Antigen binding

Antigen binding was determined using an ELISA, with soluble tissue factor (TF) as the coat and peroxidase-conjugated goat anti-human $F(ab')_2$ (Jackson ImmunoResearch, West Grove, PA) as the detecting antibody.

### 2.9. C1q, FcRn and FcγRI binding

The binding of purified *E. coli*-produced αTF to human C1q, human FcRn and the FcγRI receptor was determined using previously described assays (Idusogie et al., 2000; Shields et al., 2001).

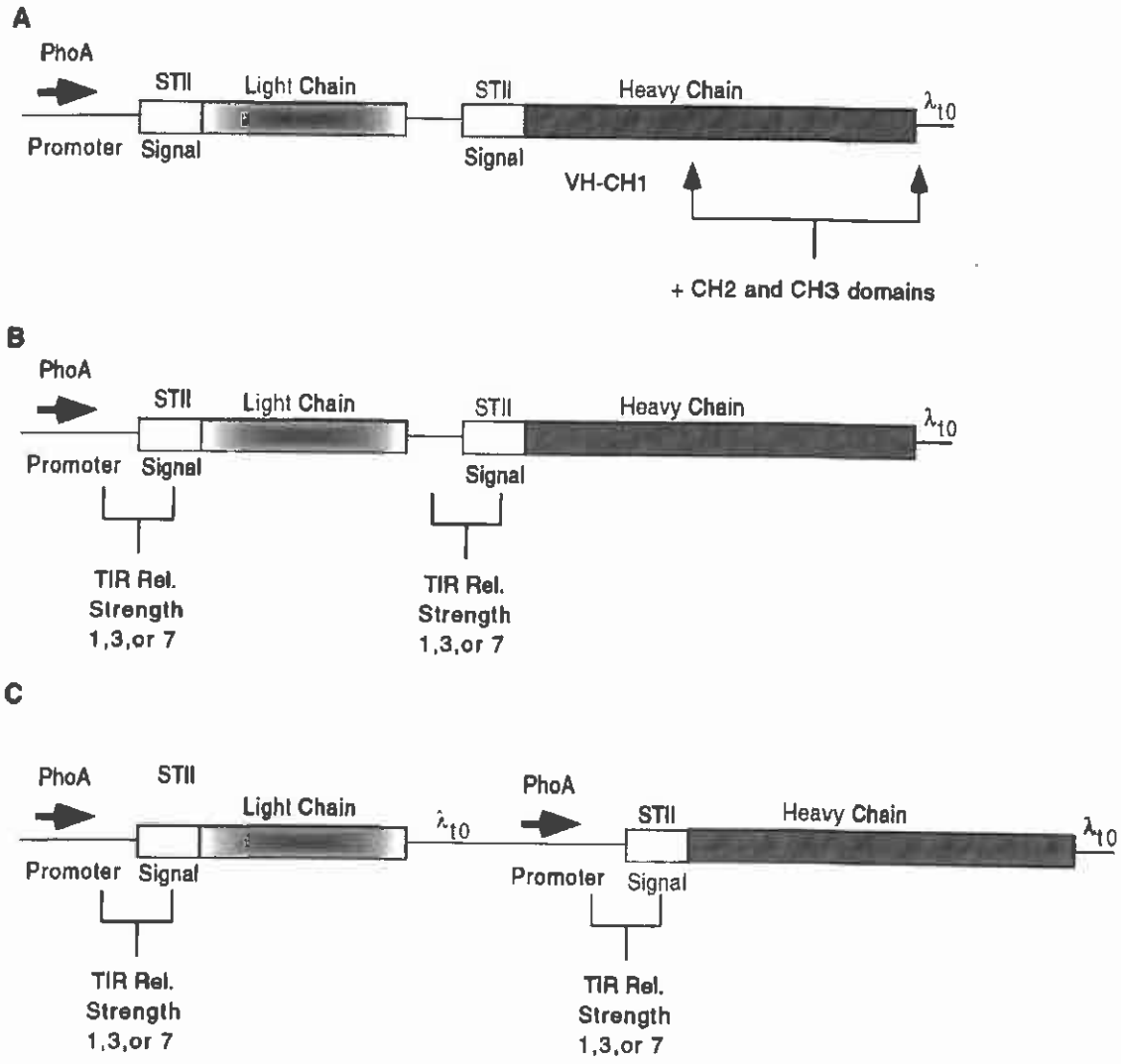



Fig. 1. Schematic of the expression plasmids. (A) polycistronic full-length expression plasmid derived from the published vector, pAK19 (Carter et al., 1992a,b). (B) Polycistronic full-length expression plasmids with varied TIR combinations. (C) Two cistron full-length antibody expression plasmids with varied TIR combinations.

### 2.10. Pharmacokinetics study

After testing negative for the presence of αTF antibodies, three chimpanzees each received a single IV bolus dose of αTF antibody (IgG1 *E. coli*, IgG2 CHO, or IgG4 CHO) at 0.10 mg/kg ($n=1$ for each antibody). Plasma samples were collected as follows: 30 and 15 min predose; 2, 15, 30 min; 1, 2, 3, 4, 6 and 12 h; 1, 2, 4, 7, 14, 21 and 28 days post IV bolus dose. Samples were assayed for αTF antibody (ATF) content

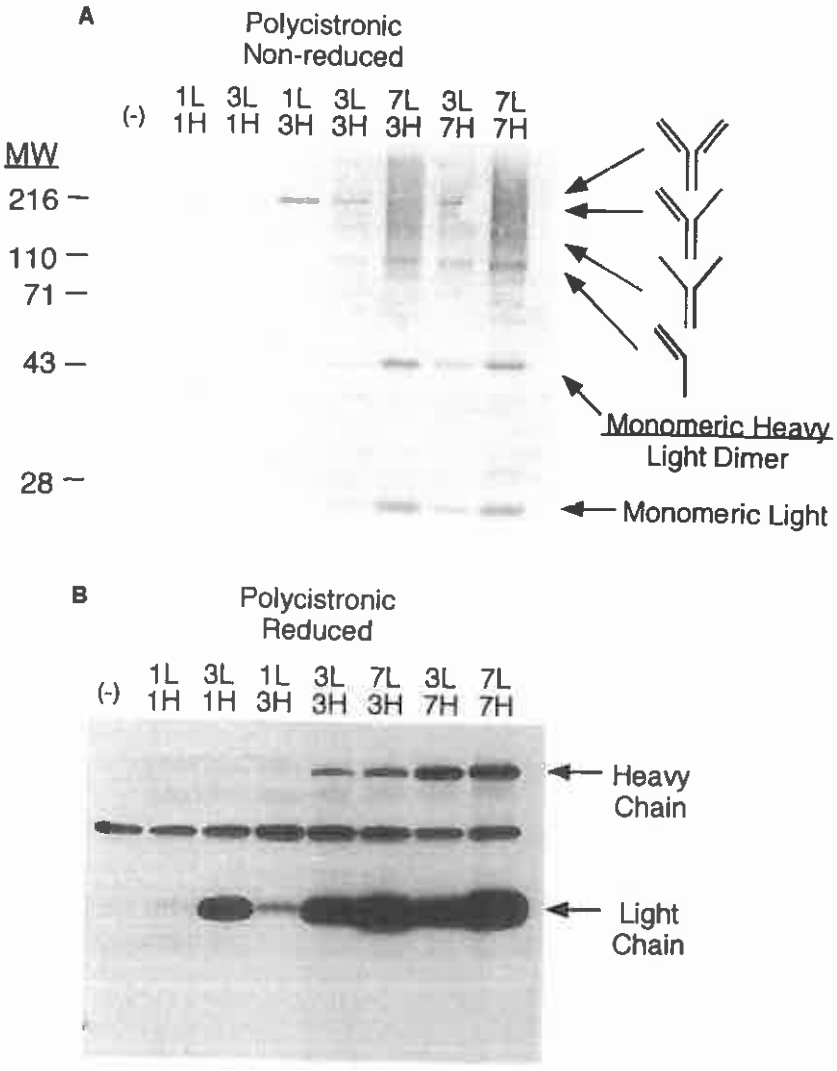



Fig. 2. Expression of full-length αTF IgGl using the polycistronic vectors. Whole cell lysates were analyzed by SDS-PAGE immunoblot following induction. (A) Non-reduced samples. The top arrow indicates folded full-length IgG1 with all disulfides formed. Monomeric light chain, and heavy chains, light chain dimer and other species, presumably lacking one or more disulfide bonds, are also indicated. The partially assembled species were verified by N-terminal analysis and mass spectrometry. (B) Reduced samples. The long exposure necessary to detect the heavy chain in the reduced samples also results in the detection of an *E. coli* protein. This band appears in all lanes, including the negative control. Listed above each lane is the relative TIR strength for light ("L") and heavy ("H") chain. (-): Induced cells harboring only the background vector, pBR322.

by ELISA, using TF as a coat and an anti-Fc monoclonal antibody as a detecting antibody. The limit of quantification was 0.102 μg/ml in chimpanzee plasma.

The plasma ATF versus time profiles were fit to a one compartment elimination profile to determine volume of distribution, clearance and terminal half-life.

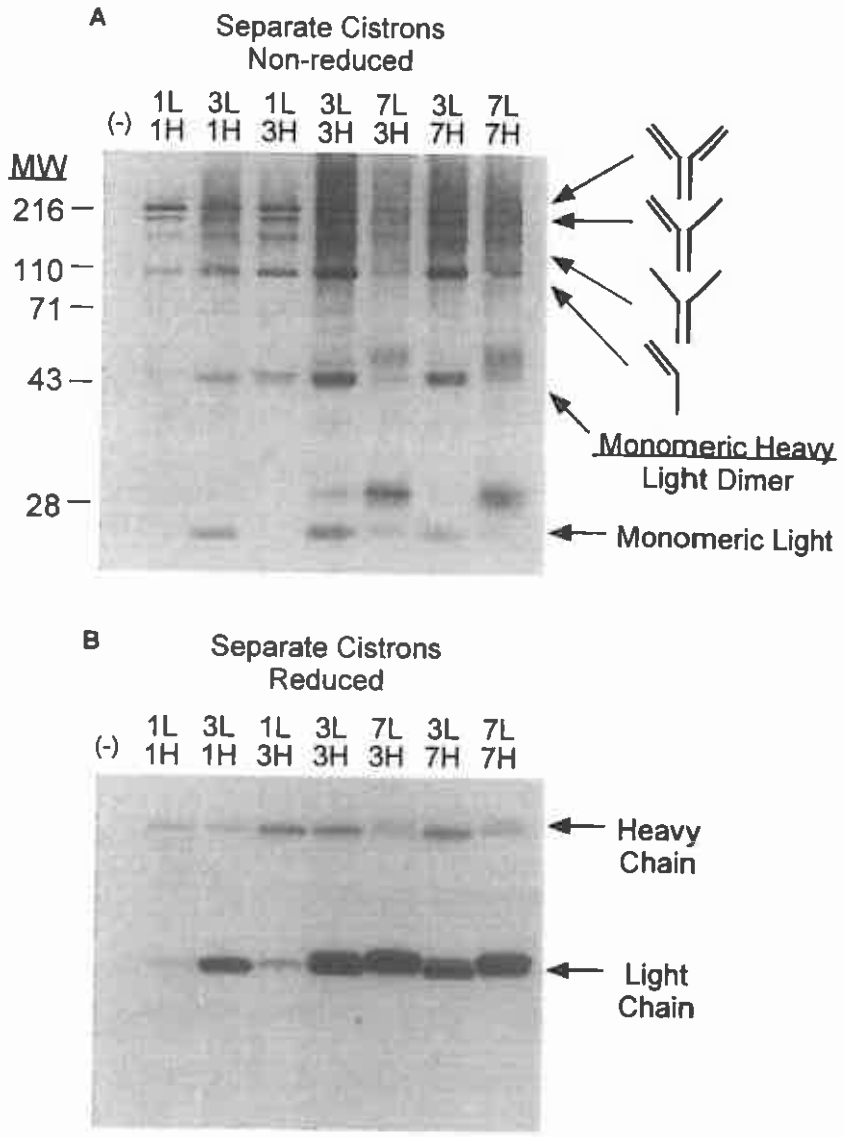



Fig. 3. Expression of full-length αTF IgG1 using the two cistron vectors. Whole cell lysates were analyzed by SDS-PAGE immunoblot following induction. (A) Non-reduced samples. The top arrow indicates folded full-length IgG1 with all disulfides formed. Monomeric light chain and heavy chains, light chain dimer and other species, presumably lacking one or more disulfide bonds, are also indicated. The partially assembled species were verified by N-terminal analysis and mass spectrometry. (B) Reduced samples. Since heavy chain was expressed at higher levels in these samples than in the polycistronic samples, only a short exposure was required when the immunoblot was developed. As a result, the additional *E. coli* band noted in Fig. 2 is not visible. Listed above each lane is the relative TIR strength for light ("L") and heavy ("H") chain. (-): Induced cells harboring only the background vector, pBR322.

## 3. Results

### 3.1. Attempts with a published vector

Initial attempts to produce full-length antibodies in *E. coli* focused on using the previously described Fab′ polycistronic expression vector pAK19 (Carter et al., 1992a). The $V_L$ and $V_H$ regions of the pAK19 Fab′ were replaced with those of the αTF antibody (Presta et al., 2001), and heavy chain was extended to encode the $C_H2$ and $C_H3$ domains of a human IgG1 immunoglobulin (Fig. 1A). The construct was transformed into a periplasmic protease deficient host, and grown in a phosphate-limiting media. Non-reduced whole cell lysates were analyzed by immunoblot using an anti-human Fab conjugated antibody as a probe; several bands were observed, including one having the molecular size of a full-length antibody. However, N-terminal amino acid analysis of the two chains from reduced samples revealed both processed and unprocessed light and heavy chain (data not shown).

This secretion problem was corrected by lowering translation levels using a series of translation initiation regions (TIRs) (Simmons and Yansura, 1996). Several constructs were prepared using different combinations of TIR strengths (Fig. 1B) for light and heavy chains, and the expression analyzed as previously described. Non-reduced whole cell lysates showed varying levels of putative full-length antibody as well as intermediate species (Fig. 2A). The optimal constructs for producing full-length antibody were 1-light 3-heavy followed by 3-light 7-heavy. Additionally, samples from the 7-light 3-heavy and 7-light 7-heavy constructs showed considerable smearing, indicative of antibody aggregation. Analysis of the reduced samples revealed a large excess of light chain over heavy for all constructs, even taking into consideration that the anti-human Fab antibody detects light chain more readily than heavy chain (Fig. 2B). Accumulation of precursor light chain was also apparent for the 7-light 7-heavy construct. Notably, the two constructs producing the greatest accumulation of full-length antibody

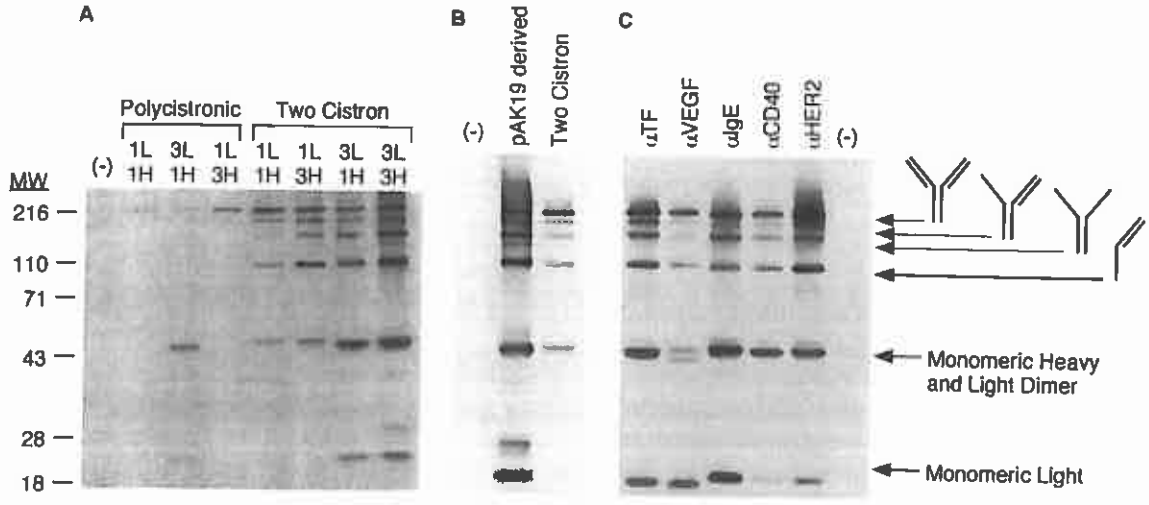



Fig. 4. Comparison of the two cistron vs. the polycistronic systems. Non-reduced whole cell lysates were analyzed by SDS-PAGE immunoblot following induction. (A) Comparison of αTF IgG1 expression using the polycistronic vectors and the two cistron vectors. Listed above each lane is the relative TIR strength for light ("L") and heavy ("H") chain. (-): Induced cells harboring only the background vector, pBR322. (B) Comparison of αTF IgG1 expression using the pAK19-derived polycistronic vector (Carter et al., 1992a) and the two cistron vector with TIR relative strengths of 1-light 1-heavy. (C) Expression of additional antibodies using the two cistron vector with relative TIR strengths of 1-light 1-heavy. The top arrow indicates folded full-length IgG1 with all disulfides formed. Monomeric light chain and heavy chains, light chain dimer and other species, presumably lacking one or more disulfide bonds, are also indicated. The partially assembled species were verified by N-terminal analysis and mass spectrometry.

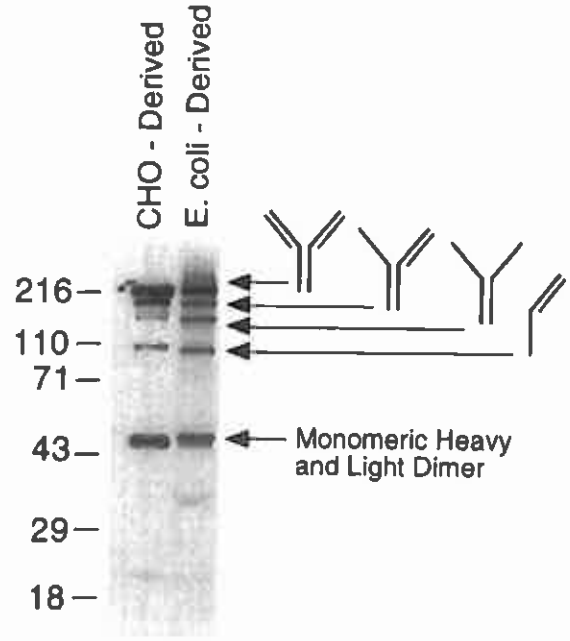



Fig. 5. Immunoblot comparison of *E. coli* vs. CHO-derived αTF IgG1. Non-reduced samples were analyzed by SDS-PAGE immunoblot. Shown are whole cell lysates for the *E.coli* sample and tissue culture supernatant for the CHO sample.

displayed the lowest light to heavy ratio, while the two constructs producing significant aggregation appeared to correlate with a large excess of light chain and the accumulation of precursor chains.

### 3.2. Two cistron system

The data from the polycistronic constructs suggested that lower light to heavy ratios might improve the accumulation of folded full-length antibody. However, the polycistronic design makes it difficult to control each chain individually since the translation of the second gene (heavy) is dependent on the light chain TIR, the translational coupling between the two genes, and finally, the heavy chain TIR. Placing each chain under the control of its own promoter addresses this problem. In a two cistron design, translational initiation of each chain is only dependent on its own TIR and the ratio of light to heavy chain can be more readily adjusted.

A series of constructs was prepared to determine the optimal TIR strengths for light and heavy chain using a two promoter system (Fig. 1C). The expression from these plasmids was then analyzed as previously described (Fig. 3A and B). Non-reduced samples of whole cell lysates produced similar antibody related bands to those observed from the polycistronic constructs, including a putative full-length species, with the optimal TIR combinations being 1-light 1-heavy and 1-light 3-heavy. Although other combinations displayed a full-length antibody band under non-reduced conditions, significant smearing caused by antibody aggregation was also apparent. Reduced samples showed a more even distribution of light and heavy chains than that of the polycistronic series. In particular, the optimal constructs for producing full-length antibody displayed almost equal band intensities for light and heavy chains, while the remainder produced excess light, a factor associated with aggregation under non-reduced conditions. Addi-

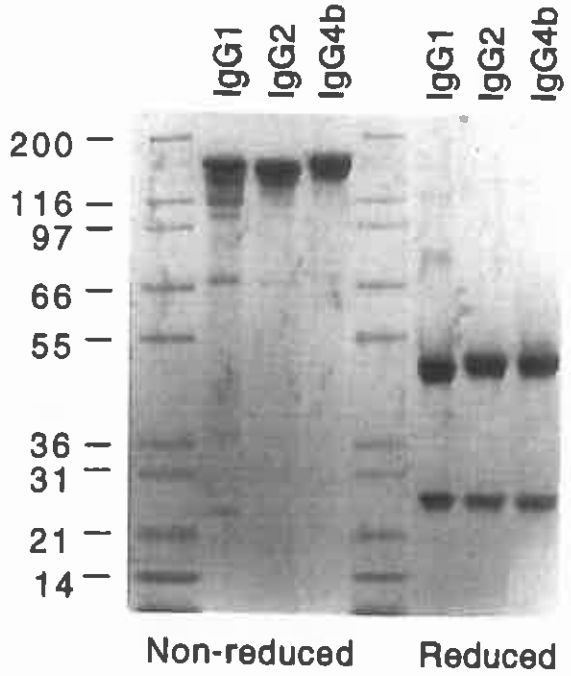



Fig. 6. SDS-PAGE analysis of purified protein. Following purification, non-reduced and reduced samples of the *E. coli*-derived IgG1 and the mammalian-produced IgG2 and IgG4b were analyzed on a NOVEX NuPAGE 4–12% gel using MOPS buffer. The gel was stained with Coomassie blue following electrophoresis. CHO-derived IgG4b is an engineered variant of IgG4 (Ser H241 Pro in the hinge region) for improved formation of inter-heavy chain disulfide bonds (Presta et al., 2001).

tionally, unprocessed light and heavy chains were observed for some constructs (3-light 3-heavy, 7-light 3-heavy and 7-light 7-heavy).

### 3.3. Comparison of the two cistron versus the polycistronic systems

A direct comparison of the separate cistron and polycistronic systems illustrates the advantage of improved control over the light to heavy chain ratio. First, non-reduced samples of whole cell lysates were prepared using plasmids with optimized TIR constructions from each system. The results demonstrate that even when the polycistronic TIRs are adjusted, the separate cistron constructs still produce considerably more full-length antibody than their polycistronic counterparts (Fig. 4A).

A second analysis involved a comparison of non-reduced whole cell lysate samples from the two cistron plasmid (1-light 1-heavy) with the construct derived from the published Fab′ vector (Fig. 4B). An induced full-length antibody band is observed in both samples; however, the expression level increases dramatically using the two cistron plasmid and the smearing associated with antibody aggregation is significantly reduced.

This two cistron technology was further tested by constructing expression plasmids for additional antibodies. As illustrated in Fig. 4C, antibodies against VEGF (higher affinity variant of the published humanized antibody; Presta et al., 1997), IgE (Presta et al., 1993), CD40 (humanized version of the murine antibody; Francisco et al., 2000) and HER-2 (Carter et al., 1992b) were also successfully expressed.

### 3.4. Scale-up and purification

An appropriate sized band on the immunoblots suggested full-length IgG1 expression and assembly. However, further analysis was required for a conclusive identification of this product. Initially, an immunoblot comparison was done using the *E. coli*-derived

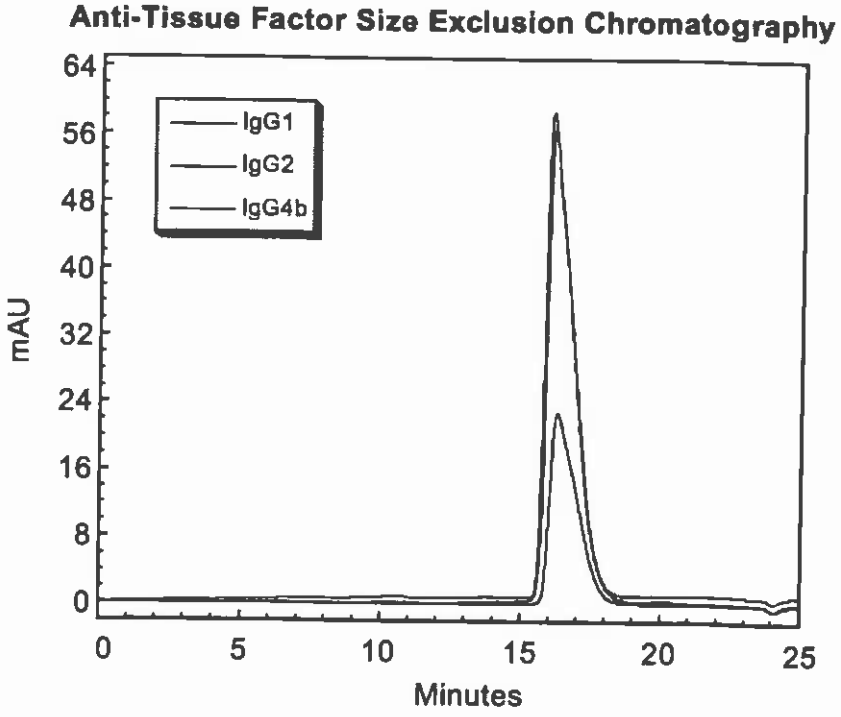



Fig. 7. Size exclusion chromatography of purified protein. Following purification, *E. coli*-derived IgG1 and the mammalian-produced IgG2 and IgG4b were analyzed by size exclusion chromatography. CHO-derived IgG4b is an engineered variant of IgG4 (Ser H241 Pro in the hinge region) for improved formation of inter-heavy chain disulfide bonds (Presta et al., 2001).

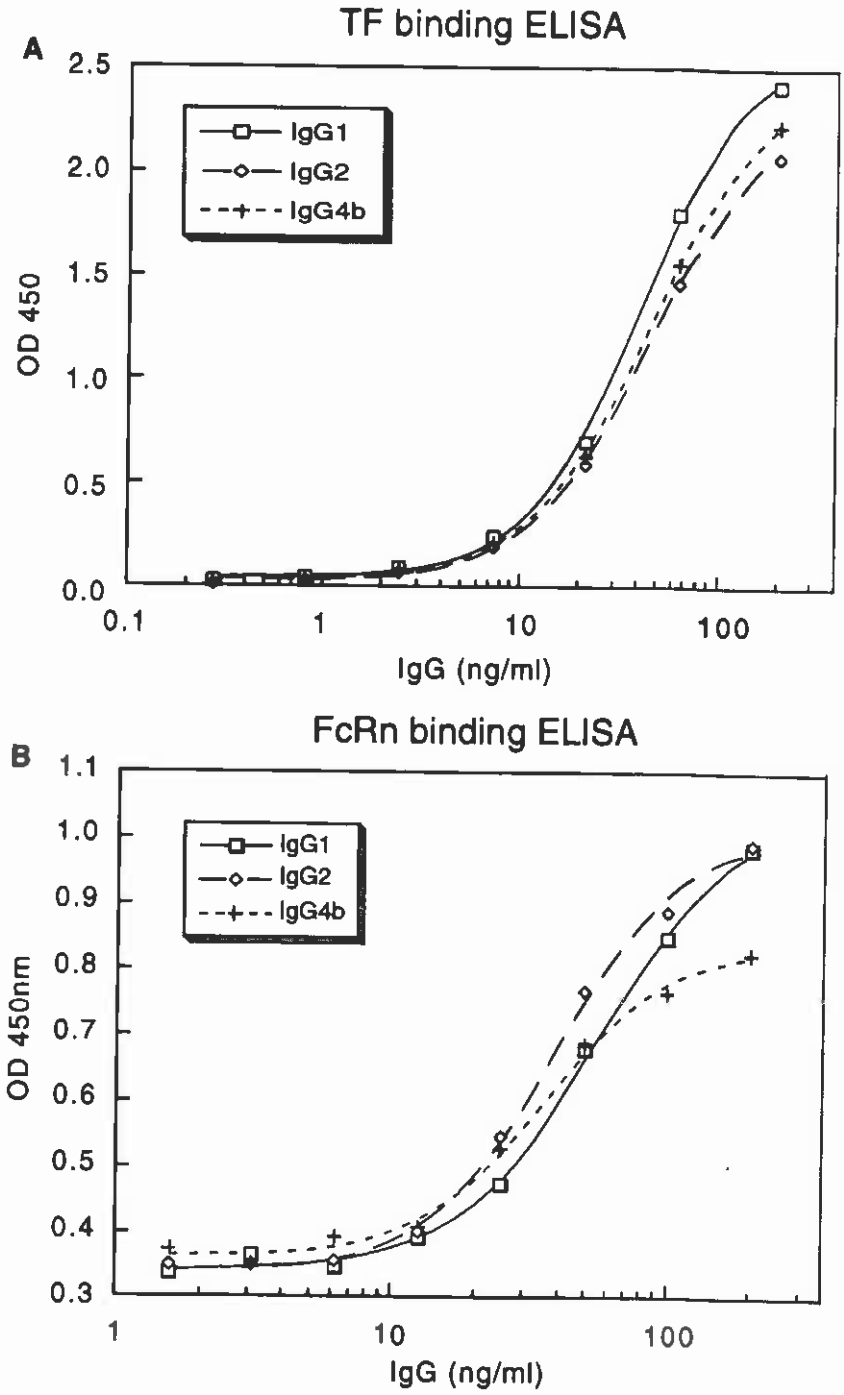



Fig. 8. Tissue factor and neonatal receptor binding assays. (A) Binding of purified *E. coli* derived αTF IgG1 and mammalian produced control αTF antibodies to soluble tissue factor. CHO-derived IgG4b is an engineered variant of IgG4 (Ser H241 Pro in the hinge region) for improved formation of inter-heavy chain disulfide bonds (Presta et al., 2001). (B) Binding of the αTF antibodies to the neonatal receptor, FcRn.

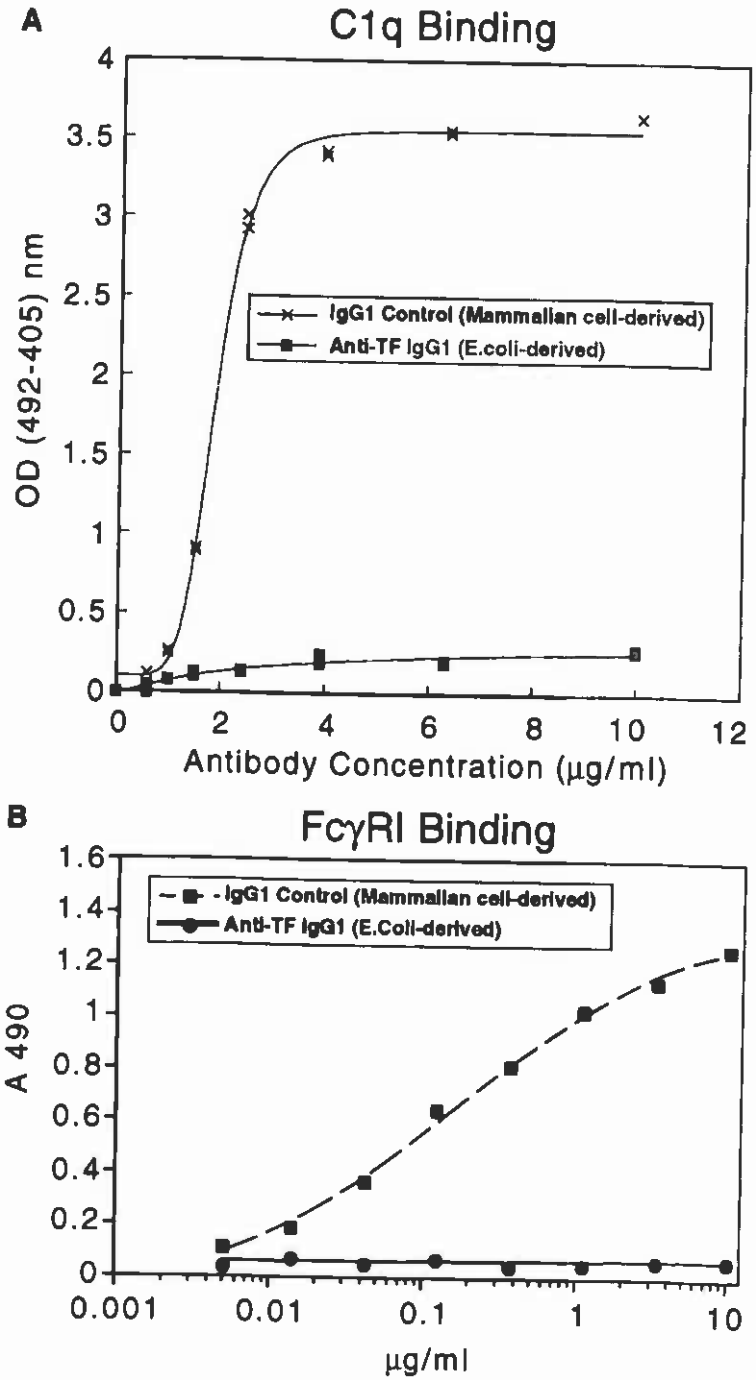



Fig. 9. C1q and FcγRI binding assays. (A) Binding of a mammalian-produced control antibody, Rituxan (Idusogie et al., 2000), to C1q. No binding was detected for the *E. coli*-derived αTF IgG1. (B) Binding of mammalian-produced αTF IgG1 to FcγRI. No binding was detected for the *E. coli*-derived αTF IgG1.

sample and an αTF IgG1 standard secreted by mammalian cells. As shown in Fig. 5, the putative fully assembled antibody from *E. coli* co-migrates with the completely assembled product from mammalian cells. Notably, the samples also show similar banding patterns, demonstrating that the presence of lower molecular weight species are not due to issues specific to *E. coli* antibody production. While the comparison with the CHO standard provided additional support that this *E. coli* system was producing fully assembled antibody, the next step in the process was to extensively test purified material obtained from fermentation paste.

Fermentations were performed using the best polycistronic (1-light 3-heavy) and separate cistron (1-light 1-heavy) constructs. The large-scale production of full-length antibody was analyzed using the AME5-RP assay (Battersby, in preparation). This quantitative assay, calibrated with an αTF aglycosylated IgG1, utilizes an immobilized Fab to first capture protein complexes possessing a light chain. A reversed phase separation of the captured species follows, distinguishing the majority of the partially assembled products from a peak containing primarily fully folded IgG1. In contrast, data from Protein G or Protein A assays include quantitation of the partially assembled species; thus, the results from the AME5-RP assay provide a more accurate calculation of yield. Samples from the polycistronic fermentation gave titers of 18 mg/l, while the separate cistron construct resulted in titers as high as 134 mg/l using the same host strain, and 156 mg/l in an alternate strain.

Following fermentation, antibody from the soluble cell lysate, where the majority of folded full-length antibody is found, was captured on protein A resin, and further purified by chromatography using SP sepharose followed by Q sepharose. An SDS-PAGE gel of this material is shown in Fig. 6 along with the purified αTF IgG2 and IgG4b produced in mammalian cells (Presta et al., 2001).

### 3.5. Characterization of the full-length IgG1

The purified αTF IgG1 was then characterized in terms of its chemical and biochemical proterties. Size exclusion chromatography (Fig. 7) of the *E. coli*-derived IgG1 and the mammalian-produced IgG2 and IgG4b each showed the expected molecular weight for a full-length intact antibody. Additionally, mass spectrometry, cation exchange chromatography, amino acid analysis and N-terminal amino acid analysis all produced the expected results. The purified *E. coli* IgG1 bound tissue factor and the FcRn receptor with similar affinity to that of αTF IgG2 and IgG4 produced in mammalian cells (Fig. 8A and B). However, as expected, the *E. coli*-produced aglycosylated IgG1 failed to bind C1q and the FcγRI receptor while mammalian-derived IgG1control antibodies bound these proteins tightly (Fig. 9A and B), suggesting that the *E. coli*-derived IgG1 lacks effector functions.

### 3.6. Pharmacokinetic properties in chimpanzees

The *E. coli*-derived full-length αTF was subject to a single IV bolus dose chimpanzee pharmacokinetic (PK) study, along with control αTF produced from mammalian cells. No obvious differences in PK parameter estimates between the antibody made in *E. coli* and the antibodies made in mammalian cells were observed (Fig. 10). While it should be noted that

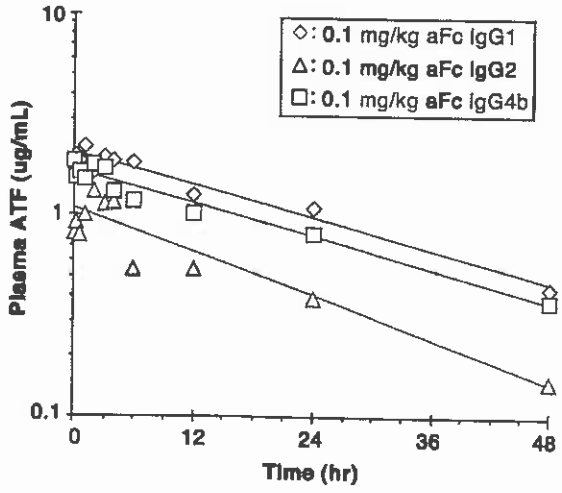



Fig. 10. Pharmacokinetic study of the purified *E. coli*-derived αTF IgG1. The pharmacokinetic properties of the *E. coli*-derived αTF IgG1 and mammalian produced control antibodies were determined in chimpanzees. The results represent a single experiment for each antibody. Calculated $T_{1/2}$ (day) values: 0.94 for *E. coli*-derived IgG1; 0.93 for CHO-derived IgG4b (Ser H241 Pro-mutation in hinge region) (Presta et al., 2001); 0.69 for CHO-derived IgG2 (Presta et al., 2001).

this experiment involved only a single chimpanzee for each antibody, these results are consistent with published data comparing glycosylated versus aglycosylated IgG1 antibodies (Tao and Morrison, 1989; Dorai et al., 1991; Hand et al., 1992; Baez et al., 2000). The authors in each study produced aglycosylated antibodies by genetic alteration, as opposed to tunicamycin treatment, and reported that the pharmacokinetic properties remained comparable to those of the glycosylated antibody.

## 4. Discussion

This report demonstrates that aglycosylated full-length antibodies can be successfully expressed and assembled in *E. coli* using a two cistron system with optimized light and heavy chain translation levels. Applying this approach, we were able to overcome the two primary problems identified using the traditional polycistronic design. First, simply adding the $C_H2$ and $C_H3$ domains to a typical vector described in the literature for the production of antibody fragments leads to inefficient secretion. Adjustments of the TIRs eliminated the secretory block; nonetheless, the expression level of heavy chain was still low in comparison to the level of light chain, highlighting the second problem associated with a polycistronic design. This design created an inherent gene expression imbalance due to inadequate translational control over the second gene on the mRNA. Thus, these constructs produce an overwhelming excess of light chain, the first gene on the message, relative to heavy chain, the second gene. However, with the two promoter system and optimized TIRs, heavy chain is efficiently secreted and a more desirable ratio of light to heavy chain is obtained, allowing the individual chains to then assemble into a correctly folded tetrameric complex. Interestingly, the ratio with the new system was closer to 1, corresponding to the light to heavy chain ratio noted in mammalian cell hosts (Scharff and Laskov, 1970; Bibila and Flickinger, 1991). This approach has proven successful for the expression in *E. coli* of several full-length antibodies including αTF, αVEGF, αCD40, αIgE, and αHER2.

Moreover, this technology is not limited to small-scale experimental work. The titers of *E. coli*-derived αTF produced in fermentors were approximately 130–150 mg/l of assembled antibody. Following purification, the IgG1 underwent a thorough chemical analysis and all results were consistent with expected properties. In addition, the *E. coli*-produced antibody strongly bound the antigen and the neonatal receptor; yet, as predicted, no binding was detected to C1q or the FcγRI receptor. Importantly, the *E. coli*-derived antibody appears to retain the desirable long half-life of a full-length antibody as demonstrated by a limited chimpanzee pharmacokinetic study.

Production of antibodies in *E. coli* offers several advantages over existing methods. First, ease of scale up and shorter fermentation run times using *E. coli* reduces the duration from plasmid construction to purified protein. A process that routinely takes several months in mammalian systems can take as little as one month with *E. coli*. Second, fermentation capital costs for therapeutic proteins are lower for bacterial production systems. And, third, the data suggest that *E. coli*-produced antibodies lack effector functions, a property desirable in many therapeutic indications that only require a blocking function by the antibody (Presta et al., 1993; Thompson et al., 1999; Armour et al., 1999; Presta et al., 2001). Indeed, for some indications, effector functions may have a deleterious impact on treatment (Isaacs et al., 1996; Friend et al., 1999; Reddy et al., 2000). Thus, *E. coli* offers a new route for the production of many therapeutic full-length antibodies and, in the less common scenario where killing of the target cell is desired, other techniques, such as coupling to a chemical toxin, could be employed.

The technology described serves as a significant step in the process of efficiently producing full-length aglycosylated antibodies in *E. coli*. The yields are presently high enough to routinely and rapidly supply material for research purposes, and the use of these proteins as therapeutic agents should be forthcoming with further increases in titer.

## Acknowledgements

We thank Brad Snedecor, Jennifer Gaunce, Nancy McFarland, John Joly, and Jane Gunson for supporting the scale-up of this technology; Bridget Currell for the sequence confirmation of the constructions; William Henzel, Jake Tropea, Adrianne Kishiyama,

Wendy Shillinglaw, and Tony Moreno for N-terminal analysis of proteins; Jeff Gorrell, Amy Lim, and Mark Iverson for help with protein purification; John Battersby for developing the AME5-RP assay; Tami Nyberg, Sandra Sandell, and PSAO for the AME5-RP results; and Jim Ligos for preparing the figures.

## References

Armour, K.L., Clark, M.R., Hadley, A.G., Williamson, L.M., 1999. Recombinant human IgG molecules lacking Fcγ receptor I binding and monocyte triggering activities. Eur. J. Immunol. 29, 2613.

Baez, J., Russell, D., Craig, J., 2000. Corn seed production of therapeutic proteins moves forward. BioPharm, 50 August.

Bibila, T., Flickinger, M.C., 1991. A structured model for monoclonal antibody synthesis in exponentially growing and stationary phase hybridoma cells. Biotechnol. Bioeng. 37, 210.

Boss, M.A., Kenten, J.H., Wood, C.R., Emtage, J.S., 1984. Assembly of functional antibodies from immunoglobulin heavy and light chains synthesized in *E. coli*. Nucleic Acids Res. 12, 3791.

Cabilly, S., Riggs, A.D., Pande, H., Shively, J.E., Holmes, W.E., Rey, M., Perry, L.J., Wetzel, R., Heyneker, H.L., 1984. Generation of antibody activity from immunoglobulin polypetide chains produced in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 81, 3273.

Carter, P., Kelley, R.F., Rodrigues, M.L., Snedecor, B., Covarrubias, M., Velligan, M.D., Wong, W.L., Rowland, A.M., Kotts, C.E., Carver, M.E., 1992a. High level *Escherichia coli* expression and production of a bivalent humanized antibody fragment. Bio/Technology 10, 12.

Carter, P., Presta, L., Gorman, C.M., Ridgeway, J.B., Henner, D., Wong, W.L., Rowland, A.M., Kotts, C., Carver, M.E., Shepard, H.M., 1992b. Humanization of an anti-p185HER2 antibody for human cancer therapy. Proc. Natl. Acad. Sci. U. S. A. 89, 4285.

Chang, C., Rey, M., Bochner, B., Heyneker, H., Gray, G., 1987. High-level secretion of human growth hormone by *Escherichia coli*. Gene 55, 189.

Conrad, U., Fiedler, U., 1998. Compartment-specific accumulation of recombinant immunoglobulins in plant cells: an essential tool for antibody production and immunomodulation of physiological functions and pathogen activity. Plant Mol. Biol. 38, 101.

Dorai, H., Mueller, B.M., Reisfeld, R.A., Gillies, S.D., 1991. Aglycosylated chimeric mouse/human IgG1 antibody retains some effector function. Hybridoma 10, 211.

Dyer, M.J.S., Hale, G., Hayhoe, F.G.J., Waldmann, H., 1989. Effects of CAMPATH-1 antibodies in vivo in patients with lymphoid malignancies: influence of antibody isotype. Blood 73, 1431.

Francisco, J.A., Donaldson, K.L., Chace, D., Siegall, C.B., Wahl, A.F., 2000. Agonistic properties and in vivo activity of the anti-CD40 antibody SGN-14. Cancer Res. 60, 3225.

Friend, P.J., Hale, G., Chatenoud, L., Rebello, P., Bradley, J., Thiru, S., Phillips, J.M., Waldmann, H., 1999. Phase I study of an engineered aglycosylated humanized CD3 antibody in renal transplant rejection. Transplantation 68, 1632.

Giddings, G., Allison, G., Brooks, D., Carter, A., 2000. Transgenic plants as factories for biopharmaceuticals. Nat. Biotechnol. 18, 1151.

Hand, P.H., Calvo, B., Milenic, D., Yokota, T., Finch, M., Snoy, P., Garmestani, K., Gansow, O., Schlom, J., Kashmiri, S.V.S., 1992. Comparative biological properties of a recombinant chimeric anti-carcinoma mAb and a recombinant aglycosylated variant. Cancer Immunol. Immunother. 35, 165.

Hinman, L.M., Hamann, P.R., Wallace, R., Menendez, A.T., Durr, F.E., Upeslacis, J., 1993. Preparation and characterization of monoclonal antibody conjugates of the calicheamicins: a novel and potent family of antitumor antibiotics. Cancer Res. 53, 3336.

Hobbs, S.M., Jackson, L.E., Hoadley, J., 1992. Interaction of aglycosyl immunoglobulins with the IgG Fc transport receptor from neonatal rat gut: comparison of deglycosylation by tunicamycin treatment and genetic engineering. Molec. Immunol. 29, 949.

Hori, A., 1991. Suppression of solid tumor growth by immunoneutralising monoclonal antibody against human basic fibroblast growth factor. Cancer Res. 51, 6180.

Idusogie, E.E., Presta, L.G., Gazzano-Santoro, H., Totpal, K., Wong, P.Y., Ultsch, M., Meng, Y.G., Mulkerrin, M.G., 2000. Mapping of the C1q binding site on Rituxan, a chimeric antibody with a human IgG1 Fc. J. Immunol. 164, 4178.

Isaacs, J.D., Wing, M.G., Greenwood, J.D., Hazleman, B.L., Hale, G., Waldmann, H., 1996. A therapeutic human IgG4 monoclonal antibody that depletes target cells in humans. Clin. Exp. Immunol. 106, 427.

Junghans, R.P., 1997. Finally! The brambell receptor (FcRB). Mediator of transmission of immunity and protection from catabolism for IgG. Immunol. Res. 16, 29.

Kikuchi, Y., Yamasaki, M., Tamura, G., 1981. The nucleotide sequence of the promoter and the amino terminal region of alkaline phosphatase structural gene (phoA) of Escherichia coli. Nucleic Acids Res. 9, 5671.

Kim, K.J., Li, B., Winer, J., Armanini, M., Gillett, N., Phillips, H.S., Ferrara, N., 1993. Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumor growth in vivo. Nature 362, 841.

Kim, J., Tsen, M., Ghetie, V., Ward, E.S., 1994. Identifying amino acid residues that influence plasma clearance of murine IgG1 fragments by site-directed mutagenesis. Eur. J. Immunol. 24, 542.

King, D.J., 1998. Applications and Engineering of Monoclonal Antibodies. Taylor and Francis, Philadelphia, PA, p. 161.

King, D.J., Adair, J.R., 1999. Recombinant antibodies for the diagnosis and therapy of human disease. Curr. Opin. Drug Discovery Dev. 2, 110.

Kipriyanov, S.M., Little, M., 1999. Generation of recombinant antibodies. Mol. Biotechnol. 12, 173.

Knight, D.M., Wagner, C., Jordan, R., McAleer, M.F., DeRita, R., Fass, D.N., Coller, B.S., Weisman, H.F., Ghrayeb, J., 1995. The immunogenicity of the 7E3 murine monoclonal Fab antibody fragment variable region is dramatically reduced in humans by substitution of human for murine constant regions. Mol. Immunol. 32, 1271.

Koumenis, I.L., Shahrokh, Z., Leong, S., Hsei, V., Deforge, L., Zapata, G., 2000. Modulating pharmacokinetics of an anti-interleukin-8 $F(ab')_2$ by amine-specific PEGylation with preserved bioactivity. Int. J. Pharm. 198, 83.

Liu, C., Tadayoni, B.M., Bourret, L.A., Mattocks, K.M., Derr, S.M., Widdison, W.C., Kedersha, N.L., Ariniello, P.D., Goldmacher, V.S., Lambert, J.M., Blattler, W.A., Chari, R.V.J., 1996. Eradication of large colon tumor xenografts by targeted delivery of maytansinoids. Proc. Natl. Acad. Sci. U. S. A. 93, 8618.

Lo, K., Roy, A., Foley, S.F., Coll, J.T., Gillies, S.D., 1992. Expression and secretion of an assembled tetrameric CH2-deleted antibody in *E. coli*. Hum. Antibodies Hybridomas 3, 123.

Mariani, G., Strober, W., 1990. Immunoglobulin metabolism. In: Metzger, H. (Ed.), Fc Receptors and the Action of Antibodies. American Society of Microbiology, Washington, DC, p. 94.

Picken, R.N., Mazaitis, A.J., Maas, W.K., Rey, M., Heyneker, H., 1983. Nucleotide sequence of the gene for heat-stable enterotoxin II of *Escherichia coli*. Infect. Immun. 42, 269.

Pluckthun, A., Krebber, A., Krebber, C., Horn, U., Knupfer, U., Wenderoth, R., Nieba, L., Proba, K., Riesenberg, D., 1996. Producing antibodies in *Escherichia coli*: from PCR to fermentation. Antibody Engineering: A Practical Approach. Oxford Univ. Press, New York, New York, p. 203.

Presta, L.G., Lahr, S.J., Shields, R.L., Porter, J.P., Gorman, C.M, Fendley, B.M., Jardieu, P.M., 1993. Humanization of an antibody directed against IgE. J. Immunol. 151, 2623.

Presta, L.G., Chen, H., O'Conner, S.J., Chisholm, V., Meng, Y.G., Krummen, L., Winkler, M., Ferrara, N., 1997. Humanization of an anti-vascular endothelial growth factor monoclonal antibody for the therapy of solid tumors and other disorders. Cancer Res. 57, 4593.

Presta, L., Sims, P., Meng, Y.G., Moran, P., Bullens, S., Bunting, S., Schoenfeld, J., Lowe, D., Lai, J., Rancatore, P., Iverson, M., Lim, A., Chisholm, V., Kelley, R.F., Riederer, M., Kirchofer, D., 2001. Generation of a humanized, high-affinity anti-tissue factor antibody for use as a novel antithrombotic therapeutic. Thromb. Haemostasis 85, 379.

Reddy, M.P., Kinney, C.A., Chaikin, M.A., Payne, A., Fishman-Lobell, J., Tsui, P., DalMonte, P.R., Doyle, M.L., Brigham-Burke, M.R., Anderson, D., Reff, M., Newman, R., Hanna, N., Sweet, R.W., Truneh, A., 2000. Elimination of Fc receptor-dependent effector functions of a modified IgG4 monoclonal antibody to human CD4. J. Immunol. 164, 1925.

Reff, M.E., Carner, K., Chambers, K.S., Chinn, P.C., Leonard, J.E., Raab, R., Newman, R.A., Hanna, N., Anderson, D.R., 1994. Depletion of B cells in vivo by a chimeric mouse human monoclonal antibody to CD20. Blood 83, 435.

Rodrigues, M.L., Snedecor, B., Chen, C., Wong, W.L., Garg, S., Blank, G.S., Maneval, D., Carter, P., 1993. Engineering Fab' fragments for efficient $F(ab')_2$ formation in Escherichia coli and for improved in vivo stability. J. Immunol. 151, 6954.

Scharff, M.D., Laskov, R., 1970. Synthesis and assembly of immunoglobulin polypeptide chains. Progr. Allergy 14, 37.

Scholtissek, S., Grosse, F., 1987. A cloning cartridge of lambda to terminator. Nucleic Acids Res. 15, 7.

Shields, R.L., Namenuk, A.K., Hong, K., Meng, Y.G., Rae, J., Briggs, J., Xie, D., Lai, J., Stadlen, A., Li, B., Fox, J.A., Presta, L.G., 2001. High-resolution mapping of the binding site on human IgG1 for FcγRI, FcγRII, FcγRIII and FcRn and design of IgG1 variants with improved binding to the FcγR. J. Biol. Chem. 276, 6591.

Simister, N.E., Mostov, K.E., 1989. An Fc receptor structurally related to MHC class I antigens. Nature 337, 184.

Simmons, L.C., Yansura, D.G., 1996. Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*. Nat. Biotechnol. 14, 629.

Sutcliffe, J.G., 1978. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. 43, 77.

Tao, M.H., Morrison, S.L., 1989. Studies of aglycosylated chimeric mouse-human IgG. Role of carbohydrate in the structure and effector functions mediated by the human IgG constant region. J. Immunol. 143, 2595.

Thompson, J.E., Vaughan, T.J., Williams, A.J., Wilton, J., Jonhson, K.S., Bacon, L., Green, J.A., Field, R., Ruddock, S., Martins, M., Pope, A.R., Tempest, P.R., Jackson, R.H., 1999. A fully human antibody neutralising biologically active human TGFbeta2 for use in therapy. J. Immunol. Methods 227, 17.

Yasmeen, D., Ellerson, J.R., Dorrington, K.J., Painter, R.H., 1976. The structure and functions of immunoglobulin domains. J. Immunol. 116, 518.

K. Schilbach*
H. Kreyenberg*
A. Geiselhart
D. Niethammer
R. Handgretinger

# Cloning of a human antibody directed against human neuroblastoma cells and specific for human translation elongation factor 1α

**Key words:**
antibodies; elongation factor 1α; immunity; monoclonal; natural; neuroblastoma

**Acknowledgments:**
This work was supported by the fortune/Profile Program, grant F1273213. GenBank accession number of B.12.6.22 heavy chain CDR3 is AY374141.

**Abstract:** Human sera have shown antitumor effects mediated by tumor-specific immunoglobulin M (IgM) antibodies. Most people who have cytotoxic serum are in good health and show no evidence of exposure to tumor antigens. We characterized the serum of a healthy female adult that was highly lytic to a neuroblastoma cell line via IgM-activated complement (>60% of malignant cells were killed during the 60-min assay). Complement-dependent lysis was not mediated by other classes of serum antibodies (data not shown) which is consistent with the findings of Ollert et al. To identify the target antigen on neuroblastoma cells, we fractionated neuroblastoma cell lysates by ion-exchange chromatography. In the fraction that showed maximal IgM binding, the dominant protein was identified as the 47-kDa translational elongation factor 1α (eEF1α). We used the donor's B-cells to create hybridomas producing the antibody (B12.6.22) that bound to neuroblastoma cells and mediated cytotoxicity. This antibody recognized eEF1α in a specific manner. Sequence analysis of the heavy chain of B12.6.22 showed usage of VH3-23 and JH6 gene segments, with no somatic mutation. The structural similarity of B12.6.22 to antibodies of the innate immune system supports the assumption that natural antibodies are a potential source of therapeutic antibodies.

**Authors' affiliations:**

K. Schilbach[1,*],
H. Kreyenberg[1,*],
A. Geiselhart[1],
D. Niethammer[1],
R. Handgretinger[2]

[1]Children's Hospital, University of Tuebingen, Department of Hematology and Oncology I, Tuebingen, Germany

[2]St Jude Children's Research Hospital, Department of Hematology-Oncology, Memphis, TN, USA

**Correspondence to:**
Karin Schilbach
Children's Hospital
University of Tuebingen
Department of Hematology and Oncology I
Hoppe-Seyler Str. 1
72076 Tuebingen, Germany
e-mail: knschilb@med.uni-tuebingen.de

Neuroblastoma is one of the most common tumors of childhood. It occurs at an early age and 50% of the patients are less than 2 years old. In infants, the disease tends to have a relatively benign course and may regress spontaneously. In contrast, children more than 2 years of age usually have metastatic disease and a poor prognosis. The current treatment of these children is very aggressive; most receive intensive chemotherapy and radiotherapy with bone marrow or stem cell rescue. However, the cure rate remains low, especially for stage IV metastatic neuroblastoma (1). Efforts have been made to develop immunotherapies that are targeted specifically to neuroblastoma cells and are minimally toxic to normal cells. Such potential agents have included antibodies directed against the ganglioside

Received 11 April 2003, revised 25 July 2003, accepted for publication 30 September 2003

Copyright © Blackwell Munksgaard 2004
*Tissue Antigens.*

*Tissue Antigens* 2004: 63: 122–131
Printed in Denmark. All rights reserved

*These authors contributed equally to the work.



251

GD2 or against the cell adhesion molecule N-CAM (2–6). Although these antibodies are highly specific, their severe side effects are dose limiting and prevent their optimal use (7–12). Some of these side effects are provoked by murine antibody or by the murine part of a chimeric antibody. The use of human antibodies should diminish these adverse effects.

Human sera contain antibodies that induce complement-mediated lysis and apoptosis of neuroblastoma cells. This phenomenon was first observed in the sera of pregnant women (13). It was described as ubiquitous in the sera of healthy children aged more than 1 year, whereas its prevalence is low in the sera of patients with neuroblastoma (14, 15). Neuroblastoma-specific immunoglobulin M (IgM) is detectable in the serum of 30% of healthy adults (16, 17). The therapeutic potential of these sera was demonstrated by the effect of total serum IgM on tumor growth in an animal model (18). A phase I/II study showed that immunotherapy for children with neuroblastoma could be accomplished feasibly and safely by exchange transfusion with donor plasma containing anti-neuroblastoma antibody; this method significantly increased serum toxicity to neuroblastoma cells (19). Both approaches, however, used total serum IgM or even unprocessed serum, and therefore the proportion of neuroblastoma-specific antibodies was smaller than that of non-specific immunoglobulins.

We sought to develop a non-chimeric human monoclonal antibody that is highly specific for human neuroblastoma cells and that produces no dose-limiting toxic effects. Here, we report the establishment of a hybridoma cell line producing a human monoclonal IgM antibody that binds specifically to human neuroblastoma cells and induces complement-mediated lysis. The B-cell donor was a healthy adult female whose serum showed high natural humoral cytotoxicity to human neuroblastoma cell lines. Western blot analysis showed that the clonal hybridoma supernatant reacts with the same 47 kDa protein targeted by the serum of the B-cell donor: translational elongation factor 1α (eEF1α).

## Materials and methods

### Human tumor cell lines

The human neuroblastoma cell line SK-N-SH, the human Burkitt lymphoma cell line Raji, the human erythroleukemia cell line K562, and the M-5 variant of the non-secretor murine myeloma cell line SP2/0 were obtained from the American Type Culture Collection (Rockville, MD). The neuroblastoma cell line SK-N-LO was kindly provided by Dr L. Helson (Sloan Kettering Institute, NY). The neuroblastoma cell line LS was established in our Hematology Research Laboratory I (20) as were the human rhabdoid tumor cell lines KD and WA (21). All cell lines were cultured as instructed by the supplier and were tested routinely for mycoplasma (Mycoplasma-PCR KIT, Roche Diagnostics, Mannheim, Germany).

### Cytotoxicity assay

The cytotoxicity of sera to target cells was measured by time-resolved fluorometric assay as described by Blomberg et al. (22). Target cells ($1 \times 10^6$ cells/ml) were incubated for 30 min at 37°C with 3 μl/ml of acetoxymethyl ester of 2.2′:6′.2″-terpyridine-6.6″ dicarboxylate (BATDA) stock solution (Wallac, Turku, Finland) in 1–3 ml of culture medium. Cells were then washed five times with culture medium at 37°C, and $5 \times 10^3$ labeled target cells/100 μl were plated into 96-well U-bottom plates (Costar, Cambridge, MA). Pooled blood group AB serum (50 μl) was added to provide complement, and the assay was completed with serial dilutions of sera or supernatants (50 μl). Spontaneous release of 2.2′:6′.2″-terpyridine-6.6″ dicarboxylic acid (TDA) was determined by incubating target cells with culture medium, and maximum release was measured after lysing the cells with 1% Triton X-100 solution. Tests were performed in triplicate. After mixing, the plates were centrifuged (2 min at $100 \times g$) and incubated for 60 min at 37°C. About 20 μl of cleared supernatant ($400 \times g$, 2 min) was incubated with 180 μl $Eu^{3+}$ solution (Wallac) for 10 min, and the fluorescence intensity of the EuTDA chelates formed was measured in a time-resolved VICTOR fluorometer (Wallac). The percent specific release was calculated as:

$$\text{Specific release (\%)} = 100 \times \frac{\text{Experimental release} - \text{spontaneous release}}{\text{Maximum release} - \text{spontaneous release}}$$

### Media, supplements, and antibodies

The growth medium for hybridomas was RPMI 1640$^+$: RPMI 1640 (Seromed Biochrom, Berlin, Germany) supplemented with 10% heat-inactivated human serum (Sigma, Taufkirchen, Germany), HEPES (25 mM; Seromed Biochrom), interleukins 2 and 4 (IL-2/IL-4, 200 IU/ml each; Chiron, Ratingen, Germany), 200 mM L-glutamine, 0.25% glucose, 1 mM pyruvate, 20 μM β-mercaptoethanol (all from Sigma), and non-essential amino acids (Seromed Biochrom).

### Cell hybridization and cloning of hybridoma B12.6.22

Monoclonal antibody to human neuroblastoma cell lines was produced by standard hybridoma technology. Briefly, peripheral

mononuclear cells from a healthy volunteer ($4–5 \times 10^6$ cells/ml of growth medium) were stimulated for 4 days with *Salmonella typhimurium* mitogen (10 μg/ml; PAN-Systems, Aidenbach, Germany) or muramyl dipeptide (10 μg/ml; Sigma), and IL-2/IL-4 (each 10 μg/ml; Chiron) before the fusion with the M-5 variant of the non-secretor murine myeloma cell line SP2/0. Hybridomas were selected by growth in RPMI $1640^+$ containing hypoxanthine, aminopterin, and thymidine (final concentration 0.1 mM, 0.1 mM, and 16 μM, respectively, all from Sigma) and 10 IU/ml IL-6 (Chiron). IgM-producing clones were identified by dot-blot technique: Supernatant (1 μl) was dotted on a nitrocellulose membrane (Hybond C pure, Amersham Pharmacia Biotech Europe, Freiburg, Germany) and incubated with phosphate-buffered saline (PBS) (Dulbecco's phosphate-buffered 0.8% NaCl solution; components are 0.20 g/l KCl, 0.20 g/l $KH_2PO_4$, 8.0 g/l NaCl, and 1.15 g/l $Na_2HPO_2$) supplemented with 10% fetal calf serum (FCS). The membrane was incubated with diluted (1 : 1000 in PBS/10% FCS) peroxidase-conjugated goat anti-human IgG (Fc) or goat antihuman IgG + IgM heavy and light chains (H + L) (Dianova, Hamburg, Germany), for 60 min, at room temperature on a shaker. After washing aminoethylcarbonate was added, and the peroxidase reaction was stopped by adding distilled water. Stability and binding specificity was checked by applying clonal supernatant to membranes which where prehybridized with diluted (1 : 100 in PBS) lysate of neuroblastoma cells.

### Flow cytometry

Viable neuroblastoma cells ($5 \times 10^4$ cells per test) were incubated with clonal hybridoma supernatant (0.1, 1.0, or 10 μl) on ice, bound IgM was detected with anti-human IgM-FITC (fluorescein) (Pharmingen, Hamburg, Germany), and cells were analyzed on a FACScan cytometer (Becton Dickinson, Heidelberg, Germany). Serum from a donor with a high cytotoxic potential against neuroblastoma cells in its IgM fraction was used as positive control.

### Separation of serum IgM from other immunoglobulin classes

Human serum antibodies were separated by size-exclusion chromatography (SEC) on a 26/60 Superdex 200 column (Amersham Pharmacia Biotech Europe) in Dulbecco's phosphate-buffered 0.8% NaCl solution with a flow rate of 2 ml/min. The resolution of the column was sufficient to enrich pentameric IgM antibodies as well as the dimeric serum antibodies. Both the fractions were assayed for binding and cytotoxicity without further concentration.

### Enzyme-linked immunosorbent assay

Microtiter plates (MaxiSorp, Nunc, Wiesbaden, Germany) were coated with diluted (1 : 1000) lysate of neuroblastoma cells, incubated with 0.5% bovine serum albumin (BSA) in PBS to block non-specific binding, and then incubated with 1 : 100 diluted serum from a positive or negative donor. Antigen-bound IgM was detected with peroxidase-conjugated goat anti-human IgM (Dianova, Hamburg, Germany) and stained with *o*-phenylenediamine (0.4 mg/ml) (Sigma) supplemented with 30% $H_2O_2$ (0.2 μl/ml) (Merck, Darmstadt, Germany). The reaction was stopped after 5–15 min by adding sulfuric acid (2.5 M). The intensity of staining was measured in a Milenia enzyme-linked immunosorbent assay (ELISA) reader (Diagnostic Products Corporation, Bad Nauheim, Germany).

### Biochemical characterization of the antigen

Cells from the neuroblastoma cell line LS were lyzed in buffer containing 20 mM Tris, 2 mM ethylenediaminetetraacetic acid, 0.2% Triton X-100, and 2 mM Dithiothreitol (DTT) (pH 7.4) at 4°C. After 30 min, samples were centrifuged for 10 min at 15,000 × *g*. Diluted lysate was transferred to an anion-exchange column (HiTrap Anion Q, Amersham Pharmacia Biotech Europe) in 2 mmol/l $NaPO_4$, pH 7.4. Bound proteins were fractionated by linear-gradient elution (0–1 M NaCl). Each fraction was tested for the antigen by ELISA. Positive fractions were pooled, concentrated (Centriprep 3000, Amicon Inc., Beverly, MA), and analyzed by SDS-PAGE.

### Western blot analysis

Cell extracts were fractionated on 12.5% SDS-polyacrylamide gel in a minigel system (Biometra, Novex Electrophoresis, Frankfurt, Germany) and transferred to a polyvinylidene fluoride membrane (Millipore, Eschborn, Germany) at 50 V for 120 min in a Tris/methanol transfer buffer according to established procedures. Blots were immunostained with supernatant of clone B12.6.22 or with diluted serum (1 : 100) from a donor with a high natural level of humoral anti-neuroblastoma activity. Blots were then incubated with a peroxidase-conjugated goat anti-human IgM (Boehringer Ingelheim, Germany). Soluble BM Blue POD Substrate® (3,3′-5,5′-tetramethylbenzidine ready-to-use solution) (Roche Diagnostics, Mannheim, Germany) was added to produce a color reaction.

### Nucleotide sequencing

Purified polymerase chain reaction (PCR) products were sequenced directly in both directions by using the Dye Terminator Cycle

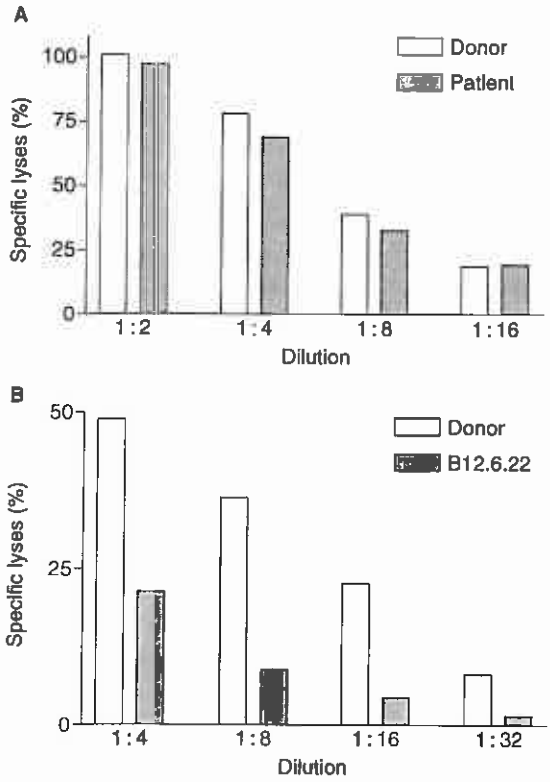



*Fig. 1.* **Cytotoxicity in BATDA release assays.** Results are presented as a percentage of the specific lysis of neuroblastoma cells induced by donor serum diluted 1:2. (A) Sera of B-cell donor and patient against LS neuroblastoma cells. The donor was a healthy adult female, and the patient had experienced 12 years of complete remission after a diagnosis of stage IVS neuroblastoma during infancy. (B) Cytotoxicity donor's serum compared to mAb B12.6.22. Serial dilutions (1:4–1:32) of a 10-fold concentrate of B12.6.22 supernatant were tested. The concentration of monoclonal antibody B12.6.22 was <1 μg/ml in undiluted concentrated supernatant. The cytotoxicity of unconcentrated supernatant was 14.4% that of donor serum diluted 1:4 (data not shown).

Sequencing Ready Reaction Kit with AmpliTaq DNA Polymerase (Applied Biosystems, Darmstadt, Germany). The sequencing primers were the same as those used for PCR reactions. Sequencing reactions were analyzed on the 310 ABI-automated nucleic acid sequencer (Applied Biosystems). All electropherograms were re-analyzed by visual inspection to check for ambiguous base readouts.

## Results

### Selection of fusion partners

Human monoclonal antibodies to neuroblastoma cells were produced by cloning SP2/0 hybridomas of the B cells of a healthy adult donor, whose serum showed high natural humoral anti-neuroblastoma activity. This donor was identified by assaying approximately 6400 individual sera by BATDA release for cytotoxicity to the neuroblastoma cell line LS. Only 1–2% of these sera induced specific cell killing greater than 60% at a 1:4 final dilution. Diluted sera allow more accurate discrimination between high and low natural humoral cytotoxicity (Geiselhart and Spring, unpublished). The serum of the selected B-cell donor showed a lytic activity of 60% at a 1:4 dilution in the cytotoxicity assay (Fig. 1), whereas 88% of the tested sera had a lytic activity of 10% or less at the same ratio. The frequency of fusion was $10^{-2}$–$10^{-3}$. We tested the IgM production of 20,000 polyclonal hybridoma cultures generated in this manner and identified 524 (2.6%) that produced IgM.

### FACS analysis

Supernatants of the 524 IgM-positive polyclonal hybridoma cultures were tested for specific binding to the neuroblastoma cell line LS on a FACScan cytometer. Those that showed high specific binding of IgM to the LS cells were cloned and recloned by limiting dilution. The clones with the highest rates of proliferation were retested by FACS analysis for specific binding to two additional neuroblastoma cell lines (SK-N-LO, SK-N-SH), to the erythroleukemia line K562, to the melanoma cell line M21, and to the human rhabdoid cell line KD. Highly lytic serum from a healthy donor was used as a positive control, and unspecific IgG and IgM from non-cytotoxic cell culture supernatants or non-toxic serum was used as negative control. The supernatant of clone B12.6.22 bound the erythroid leukemia cell line K562 minimally or not at all, but it bound the melanoma cell line M21 and the neuroblastoma cell lines SK-N-SH, SK-N-LO, and LS to the same extent (Fig. 2). It also shows trace reactivity with the human rhabdoid tumor cell line KD (Fig. 2).

### Sequence analysis of the IgM heavy chain

The monoclonal character of the cell line B12.6.22 was validated by PCR amplification and subsequent sequencing of the rearranged IgM heavy chain gene. The DNA was tested with primers specific for $V_H$-family framework region 1 (FR1) in combination with consensus joining primer $J_H$-cons (23) and, for more detailed analysis, with primers specific for JH1 to JH6 (24). Both the PCR amplification and the subsequent sequence analysis showed a monoclonal $V_H3$-23 $J_H6$ rearrangement that involved a D2-21 segment. Junction analysis (Fig. 3) was performed with the ImMunoGeneTics (IMGT) Junction Analysis tool (25).

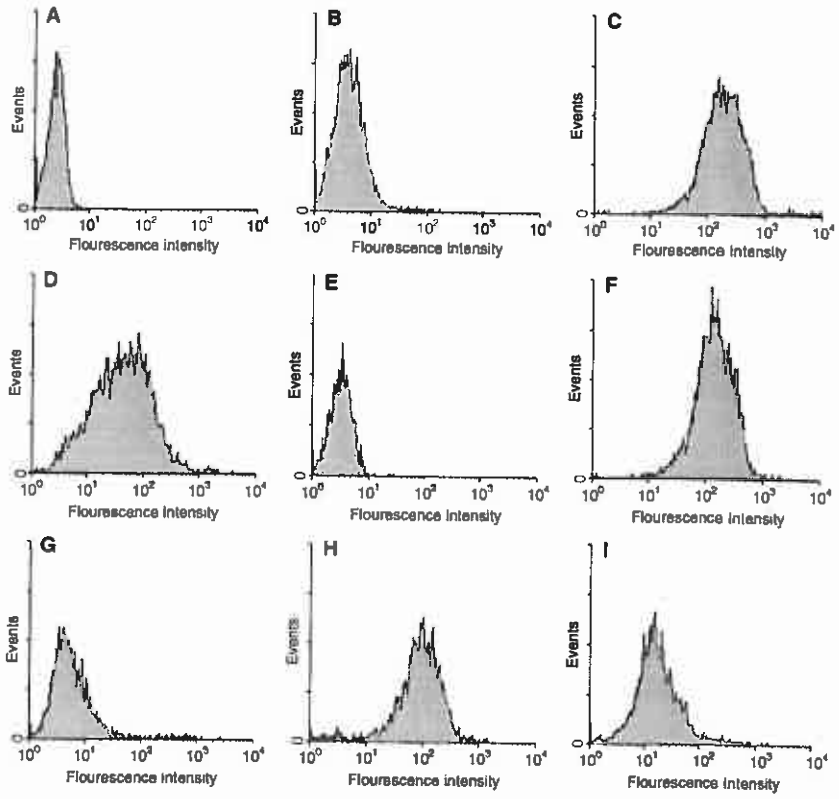



*Fig. 2.* **Representative fluorescence-activated cell sorter analysis of the sera of B-cell donor and neuroblastoma patient (*see* Fig. 1) and supernatant containing B12.6.22 for specific cell binding.** Panels A-F show binding to LS neuroblastoma cells. (A) Control 1 (secondary antibody only); (B) control 2 (pooled human serum); (C) donor serum shows high natural anti-neuroblastoma activity; (D) serum of patient in complete remission of stage IVS neuroblastoma for 12 years shows high natural anti-neuroblastoma activity; (E) Serum of patient with active neuroblastoma (F) Binding of B12.6.22 supernatant to LS neuroblastoma cells; (G) Binding of B12.6.22 supernatant to K562 erythroleukemia cells; (H) Binding of B12.6.22 supernatant to M21 melanoma cells; (I) Binding of B12.6.22 supernatant to KD rhabdomyosarcoma cells.

### Characterization of the tumor antigen by Western blot analysis

To identify the antigen recognized by IgM B12.6.22 antibody, we analyzed lysates of neuroblastoma cells and lysates of the rhabdoid cell line KD in comparison with the enriched antigen (*Material and methods*), which was recovered by anion-exchange chromatography (Fig. 4). The antigen was not pure, but staining revealed fewer separate bands in the (lane 4) than in unfractionated, unprocessed lysate of LS (lane 2) and KD (lane 3) cells. When hybridized with the blot, the serum proteins of the healthy donor recognized both a 85-kDa band and a 47-kDa band in LS and KD lysates. The supernatant of B12.6.22 culture strongly recognized a 47-kDa protein in LS lysate but recognized this protein only weakly in KD lysate. A high molecular-weight band was also recognized in the LS lysate but to a much smaller extent. Therefore, B12.6.22-generated IgM antibody

**A**

V-Region ($V_H3$) N1 D-Region (D2-21) J-Region ($J_H6$)

TGTGCGAAAG-CCCGAC-TGGAGGTGACTAACA-ATTACTACTACTACTACGGTATGGACGCTGG

**B**

```
  1 yilpptrptd kplrlplqgv ykiggigtvp vgrvetgvlk pgmvvtfapv nvttevkave
 61 mhhealseal prdnvgfnvk nvsvkdvrrg nvagdskndl pmeaagftaq viilnhlgqi
121 sagyapvldc htahiackfa elkekiavla ksrktaeflk sgdaaivdmv pskpmcveaf
181 sdypplgrfa vrdmrqtvav gvikavdkka agagkvtksa qkaqkak
```

*Fig. 3.* **(A) Partial sequence of the heavy chain junctional region (H-CDR3) showing the complete V-D-J rearrangement of B12.6.22.** The D and J regions were directly linked, and no N2 element was identified. (B) Peptide sequence of the human elongation factor 1α. Sequences of analyzed peptide fragments are highlighted.

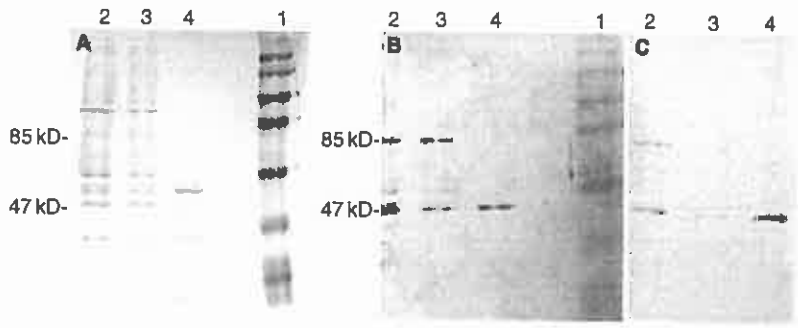



*Fig. 4.* **Detection of antigen by immunoglobulin M (IgM) in serum with high natural anti-neuroblastoma activity and in supernatant containing B12.6.22.** (A) Lysates of LS neuroblastoma cells (lane 2) and KD rhabdoid cells (lane 3), extract of neuroblastoma cell lysate after passage over anionic exchange column (lane 4), and size marker (lane 1) are shown in SDS-PAGE gel after electrophoresis under non-reducing conditions and staining with Coomassie blue to differentiate bands. (B) Nitrocellulose-blotted proteins of gel described in A, hybridized with donor and probed with peroxidase-labeled anti-human IgM. (C) Nitrocellulose-blotted proteins of gel described in A, hybridized with B12.6.22-containing supernatant and probed with peroxidase-labeled anti-human IgM.

recognizes a specific epitope on neuroblastoma cells that is clearly expressed less strongly on KD cells. Both unfractionated serum antibodies and monoclonal B12.6.22 recognize the 47-kDa band in the fraction enriched for the antigen, which shows that the antigen was not damaged during the enrichement procedure.

### Protein analysis

The 47-kDa protein seperated from the crude neuroblastoma cell lysate by anion-exchange chromatography was finally purified by SDS-PAGE. After staining with Coomassie blue, the protein band was cut out and further analysis was performed by custom service (TopLab, Martinsried, Germany). Briefly, after trypsin digest, the resulting fragments were separated by reverse-phase chromatography on a HP 1090 chromatograph (Hewlett-Packard). The resulting peptide pattern was identical to eEF1α. The peptides were then sequenced on a PE Procise 492 sequencer (Applied Biosystems). All internal sequences showed identity to trypsinized fragments of the human eEF1α (Fig. 3) (26).

### Cytotoxicity assay

We used a BATDA release assay to measure the specific lysis of neuroblastoma cells by the B12.6.22 monoclonal antibody (Fig. 1). B12.6.22 supernatant was concentrated 10-fold by protein filtration, and the assay used a maximum 1:4 dilution of this concentrate because of the limited quantity of supernatant. The 1:4 dilution exerted a level of cytotoxicity 20% that of highly lytic donor serum diluted 1:2. In a custom IgM ELISA, the concentration of monoclonal antibody B12.6.22 in supernatant was detectable but below the range of standard (<1 μg/ml). The cytotoxicity of pooled human group AB serum, which was used as the background control value in the cytotoxicity assays was zero.

### Relation of cytotoxicity to IgM binding to tumor cell lysate and to eEF1α

In wells coated with neuroblastoma antigens, we quantified the IgM binding of nine sera that were consistently highly cytotoxic (positive sera) and nine sera that were consistently non-cytotoxic (negative sera) to neuroblastoma cells. The wells of 96-well plates were coated with the neuroblastoma cell line LS lysate or with eEF1α. BSA (or human serum albumin (HSA)) was used as a blocking agent. The binding of pooled serum was used as a reference and defined as zero. Analysis of the results by unpaired *t*-test showed that positive and negative sera differed significantly in the binding of LS lysate ($P=0.0001$) and in the binding of eEF1α ($P=0.0014$). Although some sera that bound strong to LS lysate showed low or moderate binding to eEF1α and vice versa, when unspecific binding was prevented by treating the plates with BSA or HSA, prior to incubation with sera. However, the correspondence between cytotoxicity and antigen binding was evident (Table 1).

## Discussion

In this study, we selected a healthy donor whose serum lyzes neuroblastoma cells with high efficiency via complement activation *in vitro*. The fractionation of serum antibodies by SEC revealed that the IgM fraction mediates this cytotoxic effect exclusively and not the dimeric antibody classes (data not shown), which is in accordance

**Quantification of bound serum immunoglobulin M (IgM) to neuroblastoma cell lysate and eEF1Aα**

| | OD 490–630 | | |
|---|---|---|---|
| | LS lysate[a] | eEF1α[b] | eEF1α[c] |
| Cytotoxic sera | | | |
| 1 | 1.1340 | 1.1055 | 1.3670 |
| 2 | 0.1040 | 0.1000 | 0.5855 |
| 3 | 0.7860 | 0.5925 | 0.9890 |
| 4 | 0.6910 | 0.5675 | 1.4550 |
| 5 | 0.3190 | 0.5240 | 1.0125 |
| 6 | 0.3180 | 0.2995 | 0.7695 |
| 7 | 0.9660 | 1.0055 | 1.1295 |
| 8 | 0.1690 | 0.2445 | 0.4050 |
| 9 | 0.4880 | 0.5365 | 0.8395 |
| Non-cytotoxic sera | | | |
| 10 | −0.0450 | 0.2435 | 0.6820 |
| 11 | −0.0420 | 0.0265 | 1.0615 |
| 12 | −0.0460 | 0.1525 | 0.3215 |
| 13 | −0.0290 | 0.1985 | 0.7230 |
| 14 | −0.0670 | 0.1245 | 0.0970 |
| 15 | −0.0400 | 0.0505 | 0.0310 |
| 16 | −0.0410 | 0.0705 | 0.0315 |
| 17 | −0.0420 | 0.0685 | 0.2211 |
| 18 | −0.0270 | 0.1125 | 0.1550 |

eEF1α, elongation factor 1α. Quantification of serum IgM binding to NB cell lysate or eEF1α. Nine cytotoxic and nine non-cytotoxic sera were tested. Tests were performed in duplicate and respective mean values are presented. Pooled antibody serum was taken as control, and its staining intensity defined as zero. Test I demonstrates IgM binding to LS lysate, test II and III binding to eEF1α, whereby plates were blocked with human serum albumin (test II) and bovine serum albumin (test I and III).

[a]Test I.

[b]Test II.

[c]Test III.

*Table 1*

with the published results (17). In addition, IgM antibodies of this specimen bind to melanoma, a tumor which, like the neuroblastoma, originates in the epidermal primordium during ontogenesis, and they also show a trace reaction with rhabdoid cell lines. However, no binding to leukemic cells and no complement activation was observed in their case, either in FACS analysis or in cytotoxicity assays.

Our goal was to isolate IgM antibodies mediating specific cytotoxic effects. To this end, we fused the B cells of the selected donor with SP2/0 cells to generate a hybridoma. Screening of the IgM-producing hybridomas thus generated was performed by fluorescence-activated cell sorter (FACS) analysis, whereby one antibody (clone B12.6.22) was identified that revealed exactly the same binding pattern as the donor's unprocessed serum IgM.

Because the corresponding antigen was difficult to identify after SDS-PAGE and Western blotting of the crude cell lysate, the target was enriched by chromatography with an anion-exchange matrix, which binds the target at physiological pH and low salt concentration (below 0.01 M). One eluted fraction showed high binding to serum IgM with a simultaneously low protein concentration, which indicates a high portion of antigenic protein. One main protein of this fraction has a molecular mass of approximately 47 kDa, and both the donor serum and the isolated antibody B12.6.22 recognized the corresponding band in Western blot experiments. After electrophoresis, the peptide revealed sufficient purity for sequence analysis and was subsequently identified as the peptide elongation factor eEF1α, a virtually ubiquitous protein. This factor transfers aminoacyl-t-RNA to the ribosome and is released upon hydrolysis of guanosine 5′ triphosphate (GTP) (27). Thus, eEF1α plays a key role in protein synthesis. Additionally, this factor also influences the cytoskeleton organization. It has been described as an actin-binding protein in many divergent species, including plants. The characteristic cytosolic or perinuclear localization of eEF1α does not exclude it as an



257

antibody target (28), because other cytosolic proteins, such as β-tubulin isoforms, are targeted by serum IgM and IgG antibodies (29). However, the preferential binding of the antibody to neuroblastoma cells remains unexplained. We speculate that the expression of different tissue-specific (30) or development-dependent isoforms of eEF1α may be implicated (31). Overexpression may also contribute. Edmonds et al. (32) for example, reported an overexpression of eEF1α in metastatic rat mammary adenocarcinoma. The increased expression of eEF1α is also related to increased cell proliferation, oncogenic transformation, delayed cell senescence, and metastasis (33, 34). Both RT-PCR and quantitative real-time PCR analysis with β2-microglobulin as internal reference revealed that eEF1α was transcribed at similar levels in the neuroblastoma cell line as in peripheral blood mononuclear cells of healthy donors (*data not shown*). However, the levels of mRNA do not define the actual protein expression and the implications of an aberrantly expressed protein.

Ditzel et al. (35) reported the existence of anti-EF1a antibodies in the serum of patients suffering from Felty's Syndrome, an autoimmune disorder which leads to the peripheral destruction of neutrophils. This indicates that the presence of these antibodies is associated with the potential to react with autoantigens.

To investigate whether anti-eEF1α antibodies could be implicated in the cytotoxicity of the other positive donor sera, the binding of cytotoxic and non-cytotoxic sera to (a) neuroblastoma lysates and (b) enriched eEF1α protein fraction was assayed. In three individual assays, the mean quantity of bound IgM was significantly greater ($P < 0.01$, unrelated $t$-test) in cytotoxic than in non-cytotoxic sera, irrespective of whether the target was lyzed cell extract or enriched for eEF1α. However, not all toxic sera were found to bind to the eEF1α, indicating that eEF1α is not the only target.

Little is known about the nature and frequency of natural IgM antibodies which are cytotoxic for human neuroblastoma cells. It is also still unclear whether they are of high or low affinity/avidity, whether they are of monoclonal, pauci-, oligo-, or polyclonal origin, or whether they have to act synergistically to mediate a measurable cytotoxic effect. As IgM binding of cytotoxic and non-cytotoxic sera is clearly divergent, but not absolutely stringent, we envisage cytotoxicity of sera as a multifactorial event in which cytotoxicity is predominantly associated with specific binding to eEF1α. In an approach similar to that described here, David et al. (36, 37) clearly demonstrated a protein of 260 kDa as a target of sera toxic to neuroblastoma cells. This protein is certainly one such additional antigenic target and there are probably more targets on neuroblastoma cells that remain to be elucidated. In the course of our experiments, we did not recognize any interaction with a factor of similar size or with potential fragments of this factor. The target identified by David et al. (36, 37) was reported to be very fragile and to undergo rapid autodegradation into two fragments. These fragments do not correspond to the molecular mass of eEF1α.

Sequence analysis of the heavy chain of the B12.6.22 antibody revealed the involvement of $V_H$ loci 3–23, D loci 2–21, and J locus 6. The expression of $V_H$ 3–23 has been reported to be associated with natural antibodies, which often react with autoantigens (38), but it has also been observed in antibodies induced by foreign antigens (39). However, the involvement of gene segments close to the germline configuration is a strong indication that B12.6.22 is a natural autoantibody (40, 41). Neither the $V_H$ nor the $J_H$ segment show any somatic mutation, and only one mutation was observed within the D segment. Most of these natural antibodies reveal a broad antigen recognition pattern and thus lack specificity. There is evidence, however, that a high frequency of Trp within the H-CDR3 region corresponds to reactivity with multiple antigens and that a low frequency of Trp increases the binding specificity (42). In fact, the H-CDR3 of B12.6.22 contains no Trp residue, which is consistent with its observed binding specificity to eEF1α.

Natural auto-reactive antibodies differ conspicuously from autoimmune antibodies and from antibodies induced by exogenous antigens in their early development and their possible association with homeostasis. They may be involved in the growth control of malignant cells, as some recognized antigens may be restricted to certain tissues or tumors. These antibodies belong to the innate immune system, and their recognition patterns remain stable throughout life (43–45). A characteristic feature of the innate immune system is that its B cells (B1 cells) are neither activated by antigen exposure nor by cross-linking of their Ig receptor. Antibody production is thus increased by antigen exposure in the acquired immune system but not in the innate immune system (46). As a result, the level of innate antibodies is reduced by depletion after antigen exposure (47, 48). Accordingly, in patients with neuroblastoma, titers of anti-neuroblastoma antibodies are observed to decline during active disease (15, 16, 18) but may increase during disease remission. As an example, Fig. 1 shows high cytotoxicity of a patients serum 12 years after spontaneous remission.

Extensive screening of the binding repertoire of innate antibodies would provide insight into the structure of their targets, and this knowledge could serve as a basis for the development of highly specific, antibody-based immunotherapy. Although we found the isolation of target-specific antibodies to be a feasible method in the case of B12.6.22 and its recognized antigen eEF1α, optimization of the method is required. The generation of monoclonal antibodies by the hybridoma technique is cumbersome and time-consuming, and its results are limited. A critical factor is the small number of clones that recognize a specific target and the poor yield of antibodies produced by the clones; these factors restricted our work to the analytical scale.

Enrichment technologies using phage display libraries (49), ribosome display (50), bacterial display (51), yeast display (52), or even fully synthetic antibody libraries (53) might be an efficient alternative to classical hybridoma technology.

## References


1. Berthold F, Hero B. Neuroblastoma: current drug therapy recommendations as part of the total treatment approach. *Drugs* 2000: **59**: 1261–77.
2. Cheung NK, Lazarus H, Miraldi FD, Abramowsky CR, Kallick S, Saarinen UM et al. Ganglioside GD2 specific monoclonal antibody 3F8: a phase I study in patients with neuroblastoma and malignant melanoma. *J Clin Oncol* 1987: **5**: 1430–40.
3. Patel K, Rossell RJ, Pemberton LF, Cheung NK, Walsh FS, Moore SE et al. Monoclonal antibody 3F8 recognises the neural cell adhesion molecule (NCAM) in addition to the ganglioside GD2. *Br J Cancer* 1989: **60**: 861–6.
4. Mujoo K, Cheresh DA, Yang HM, Reisfeld RA. Disialoganglioside GD2 on human neuroblastoma cells: target antigen for monoclonal antibody-mediated cytolysis and suppression of tumor growth. *Cancer Res* 1987: **47**: 1098–104.
5. Mueller BM, Romerdahl CA, Gillies SD, Reisfeld RA. Enhancement of antibody-dependent cytotoxicity with a chimeric anti-GD2 antibody. *J Immunol* 1990: **144**: 1382–6.
6. Bourne SP, Patel K, Walsh F, Popham CJ, Coakham HB, Kemshead JT. A monoclonal antibody (ERIC-1), raised against retinoblastoma, that recognizes the neural cell adhesion molecule (NCAM) expressed on brain and tumours arising from the neuroectoderm. *J Neurooncol* 1991: **10**: 111–9.
7. Saleh MN, Khazaeli MB, Wheeler RH, Allen L, Tilden AB, Grizzle W et al. Phase I trial of the chimeric anti-GD2 monoclonal antibody ch14.18 in patients with malignant melanoma. *Hum Antibodies Hybridoma* 1992: **3**: 19–24.
8. Handgretinger R, Baader P, Dopfer R, Klingebiel T, Reuland P, Treuner J et al. A phase I study of neuroblastoma with the anti-ganglioside GD2 antibody 14. G2a. *Cancer Immunol Immunother* 1992: **35**: 199–204.
9. Handgretinger R, Anderson K, Lang P, Dopfer R, Klingebiel T, Schrappe M et al. A phase I study of human/mouse chimeric antiganglioside GD2 antibody ch14.18 in patients with neuroblastoma. *Eur J Cancer* 1995: **31A**: 261–7.
10. Uttenreuther-Fischer MM, Huang CS, Reisfeld RA, Yu AL. Pharmacokinetics of anti-ganglioside GD2 mAb 14G2a in a phase I trial in pediatric cancer patients. *Cancer Immunol Immunother* 1995: **41**: 29–36.
11. Yu AL, Uttenreuther-Fischer MM, Huang CS, Tsui CC, Gillies SD, Reisfeld RA et al. Phase I trial of a human-mouse chimeric anti-disialoganglioside monoclonal antibody ch14.18 in patients with refractory neuroblastoma and osteosarcoma. *J Clin Oncol* 1998: **16**: 2169–80.
12. Cheung NK, Kushner BH, Cheung IY, Kramer K, Canete A, Gerald W et al. Anti-G (D2) antibody treatment of minimal residual stage 4 neuroblastoma diagnosed at more than 1 year of age. *J Clin Oncol* 1998: **16**: 3053–60.
13. Bolande RP, Mayer DC. The cytolysis of human neuroblastoma cells by a natural IgM 'antibody'-complement system in pregnancy serum. *Cancer Invest* 1990: **8**: 603–11.
14. Fukuda M, Nozaki C, Ishiguro Y, Horibe K. Natural antibody against neuroblastoma among Japanese children with or without neuroblastoma. *Cancer* 1999: **86**: 2166–70.
15. Fukuda M, Nozaki C, Ishiguro Y, Horibe K. Distribution of natural antibody against human neuroblastoma among children with or without neuroblastoma. *Med Oncol* 2001: **36**: 147–8.
16. Rodman TC, To SE, Sullivan JJ, Winston R. Innate natural antibodies. Primary roles indicated by specific epitopes. *Hum Immunol* 1997: **55**: 87–95.
17. Ollert MW, David K, Schmitt C, Hauenschild A, Bredehorst R, Erttmann R et al. Normal human serum contains a natural IgM antibody cytotoxic for human neuroblastoma cells. *Proc Natl Acad Sci USA* 1996: **93**: 4498–503.
18. Engler S, Thiel C, Forster K, David K, Bredehorst R, Juhl H. A novel metastatic animal model reflecting the clinical appearance of human neuroblastoma: growth arrest of orthotopic tumors by natural, cytotoxic human immunoglobulin M antibodies. *Cancer Res* 2001: **61**: 2968–73.
19. Schmitt C, David K, Hiller J, Schrum J, Bredehorst R, Vogel CW et al. Natural human IgM-antibodies in neuroblastoma therapy: preliminary findings of a phase I/II clinical trial. *Klin Padiatr* 1999: **211**: 314–8.
20. Rudolph G, Schilbach-Stuckle K, Handgretinger R, Kaiser P, Hameister H. Cytogenetic and molecular characterization of a newly established neuroblastoma cell line LS. *Hum Genet* 1991: **86**: 562–6.
21. Handgretinger R, Kimmig A, Koscielnak E, Schmidt D, Rudolph G, Wolburg H et al. Establishment and characterization of a cell line (Wa-2) derived from an extrarenal rhabdoid tumor. *Cancer Res* 1990: **50**: 2177–82.
22. Blomberg K, Hautala R, Lovgren J, Mukkala VM, Lindqvist C, Akerman K. Time-resolved fluorometric assay for natural killer activity using target cells labelled with a fluorescence enhancing ligand. *J Immunol Methods* 1996: **193**: 199–206.
23. Pongers-Willemse MJ, Verhagen OJ, Tibbe GJ, Wijkhuijs AJ, de HV, Roovers E et al. Real-time quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia using junctional region specific TaqMan probes. *Leukemia* 1998: **12**: 2006–14.
24. Verhagen OJ, Willemse MJ, Breunis WB, Wijkhuijs AJ, Jacobs DC, Joosten SA et al. Application of germline IGH probes in real-time quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia. *Leukemia* 2000: **14**: 1426–35.




259

25. Lefranc MP. MGT, the international ImMunoGeneTics database. *Nucleic Acids Res* 2003: 31: 307–10.
26. Opdenakker G, Cabeza-Arvelaiz Y, Fiten P, Dijkmans R, Van Damme J, Volckaert G et al. Human elongation factor 1 alpha: a polymorphic and conserved multigene family with multiple chromosomal localizations. *Hum Genet* 1987: 75: 339–44.
27. Kaziro Y. The role of guanosine 5'-triphosphate in polypeptide chain elongation. *Biochim Biophys Acta* 1978: 505: 95–127.
28. Kjaer S, Wind T, Ravn P, Ostergaard M, Clark BF, Nissim A. Generation and epitope mapping of high-affinity scFv to eukaryotic elongation factor 1A by dual application of phage display. *Eur J Biochem* 2001: 268: 3407–15.
29. Prasannan L, Misek DE, Hinderer R, Michon J, Geiger JD, Hanash SM. Identification of beta-tubulin isoforms as tumor antigens in neuroblastoma. *Clin Cancer Res* 2000: 6: 3949–56.
30. Lee S, Francoeur AM, Liu S, Wang E. Tissue-specific expression in mammalian brain, heart, and muscle of S1, a member of the elongation factor-1 alpha gene family. *J Biol Chem* 1992: 267: 24064–8.
31. Khalyfa A, Bourbeau D, Chen E, Petroulakis E, Pan J, Xu S et al. Characterization of elongation factor-1A (eEF1A-1) and eEF1A-2/S1 protein expression in normal and wasted mice. *J Biol Chem* 2001: 276: 22915–22.
32. Edmonds BT, Wyckoff J, Yeung YG, Wang Y, Stanley ER, Jones J et al. Elongation factor-1 alpha is an overexpressed actin binding protein in metastatic rat mammary adenocarcinoma. *J Cell Sci* 1996: 109: 2705–14.
33. Taniguchi S, Miyamoto S, Sadano H, Kobayashi H. Rat elongation factor 1 alpha: sequence of cDNA from a highly metastatic fos-transferred cell line. *Nucleic Acids Res* 1991: 19: 6949.
34. Pencil SD, Toh Y, Nicolson GL. Candidate metastasis-associated genes of the rat 13762NF mammary adenocarcinoma. *Breast Cancer Res Treat* 1993: 25: 165–74.
35. Ditzel HJ, Masaki Y, Nielsen H, Farnaes L, Burton DR. Cloning and expression of a novel human antibody-antigen pair associated with Felty's syndrome. *Proc Natl Acad Sci USA* 2000: 97: 9234–9.
36. David K, Ehrhardt A, Ollert MW, Erttmann R, Bredehorst R, Vogel CW. Expression of a 260 kDa neuroblastoma surface antigen, the target of cytotoxic natural human IgM: correlation to MYCN amplification and effects of retinoic acid. *Eur. J.Cancer* 1997: 33: 1937–41.
37. David K, Ollert MW, Vollmert C et al. Human natural immunoglobulin M antibodies induce apoptosis of human neuroblastoma cells by binding to a Mr 260,000 antigen. *Cancer Res* 1999: 59: 3768–75.
38. Dersimonian H, McAdam KP, Mackworth-Young C, Stollar BD. The recurrent expression of variable region segments in human IgM anti-DNA autoantibodies. *J Immunol* 1989: 142: 4027–33.
39. Ikematsu H, Kasaian MT, Schettino EW, Casali P. Structural analysis of the VH-D-JH segments of human polyreactive IgG mAb. Evidence for somatic selection. *J Immunol* 1993: 151: 3604–16.
40. Grandien A, Modigliani Y, Freitas A, Andersson J, Coutinho A. Positive and negative selection of antibody repertoires during B-cell differentiation. *Immunol Rev* 1994: 137: 53–89.
41. Schwartz RS, Stollar BD. Heavy-chain directed B-cell maturation: continuous clonal selection beginning at the pre-B cell stage. *Immunol Today* 1994: 15: 27–32.
42. Aguilera I, Melero J, Nunez-Roldan A, Sanchez B. Molecular structure of eight human autoreactive monoclonal antibodies. *Immunology* 2001: 102: 273–80.
43. Lacroix-Desmazes S, Kaveri SV, Mouthon L, Ayouba A, Malanchere E, Coutinho A et al. Self-reactive antibodies (natural autoantibodies) in healthy individuals. *J Immunol Methods* 1998: 216: 117–37.
44. Lacroix-Desmazes S, Mouthon L, Coutinho A, Kazatchkine MD. Analysis of the natural human IgG antibody repertoire: life-long stability of reactivities towards self antigens contrasts with age-dependent diversification of reactivities against bacterial antigens. *Eur J Immunol* 1995: 25: 2598–604.
45. Mouthon L, Haury M, Lacroix-Desmazes S, Barreau C, Coutinho A, Kazatchkine MD. Analysis of the normal human IgG antibody repertoire. Evidence that IgG autoantibodies of healthy adults recognize a limited and conserved set of protein antigens in homologous tissues. *J Immunol* 1995: 154: 5769–78.
46. Baumgarth N, Herman OC, Jager GC, Brown L, Herzenberg LA, Herzenberg LA. Innate and acquired humoral immunities to influenza virus are mediated by distinct arms of the immune system. *Proc Natl Acad Sci USA* 1999: 96: 2250–5.
47. Rodman TC, Winston R, Sullivan JJ, Yan XJ, Chiorazzi N. An innate natural antibody is reactive with a cryptic sequence of lactoferrin exposed on sperm head surface. *Proc Soc Exp Biol Med* 1997: 216: 404–9.
48. Re MC, Furlini G, Vignoli M, Ramazzotti E, Roderigo G, De R et al. Effect of antibody to HIV-1 Tat protein on viral replication in vitro and progression of HIV-1 disease in vivo. *J Acquir Immune Defic Syndr Hum Retrovirol* 1995: 10: 408–16.
49. Scott JK, Smith GP. Searching for peptide ligands with an epitope library. *Science* 1990: 249: 386–90.
50. Hanes J, Pluckthun A. In vitro selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci USA* 1997: 94: 4937–42.
51. Georgiou G, Stathopoulos C, Daugherty PS, Nayak AR, Iverson BL, Curtiss R. III Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nat Biotechnol* 1997: 15: 29–34.
52. Kieke MC, Cho BK, Boder ET, Kranz DM, Wittrup KD. Isolation of anti-T cell receptor scFv mutants by yeast surface display. *Protein Eng* 1997: 10: 1303–10.
53. Knappik A, Ge L, Honegger A, Pack P, Fischer M, Wellnhofer G et al. Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J Mol Biol* 2000: 296: 57–86.

*Ann Rheum Dis* 1999;58:(Suppl I) I70–172

# Preliminary results of early clinical trials with the fully human anti-TNFα monoclonal antibody D2E7

Joachim Kempeni

Current pharmacological treatments for rheumatoid arthritis (RA), including non-steroidal anti-inflammatory drugs (NSAIDs), disease modifying anti-rheumatic drugs (DMARDs), and corticosteroids, have been moderately successful in alleviating the discomforts associated with swollen, painful joints. However, conventional medical approaches to treatment have had little or no impact on the disease course of RA.[1] Innovative strategies, particularly those based on new concepts in the immunobiology of RA, are being developed to target cellular inflammatory mechanisms and potentially prevent disease progression. The more promising of these treatments seem to be those that block the effects of tumour necrosis factor (TNF) α, because this proinflammatory cytokine seems to play a central part in the immunopathogenesis of RA.[2 3]

## Anti-TNF treatments

Biological agents such as antibodies and soluble TNF receptors that bind TNF α with high specificity neutralise its activity and have been developed for use as therapeutic agents. Several are currently being evaluated in patients with RA (table 1). Infliximab, cA2, is a chimeric monoclonal antibody (mAb) that consists of the variable region of a murine anti-TNF mAb coupled to the constant region of human IgG1κ.[4 5] The resulting construct is approximately two thirds human. CDP571 is a humanised mAb consisting of the complementarity determining regions of a murine anti-TNF mAb grafted into a human immunoglobulin (IgG4κ).[6] This mAb is approximately 95% human. Etanercept is a fusion protein consisting of two recombinant p75 TNF receptors attached to the Fc portion of a human IgG1.[7] Although this construct consists of two independent elements, which themselves contain 100% human peptide sequences, they are arranged in an unnatural configuration.

The duration of the therapeutic efficacy of these TNF antagonists may be limited by an immune response to their non-human elements or artificially fused human sequences. The development of antibodies to these biological agents could reduce their half life, thereby decreasing efficacy.[8] In addition, an immune response could result in adverse events from the formation of immune complexes or the development of hypersensitivity. For these reasons, an antibody that is fully human may have greater therapeutic potential.

**Department of Clinical Oncology/Immunology, Knoll AG-BASF Pharma, Knollstrasse, D-67008 Ludwigshafen, Germany**

Correspondence to:
Dr J Kempeni.

*Table 1 Anti-TNF α treatments*

| *Name* | *Description* |
|---|---|
| D2E7 | Fully human anti-TNF α mAb |
| CDP571 | Humanised anti-TNF α mAb (murine CDR) |
| Etanercept | Fusion protein consisting of two recombinant p75 TNF α receptors and the Fc portion of human IgG1 |
| Infliximab (cA2) | Chimeric murine/human anti-TNF-α mAb |

TNF=tumour necrosis factor; mAb=monoclonal antibody; IgG=immunoglobulin G; CDR=complementarity determining region.

## Development of a fully human anti-TNF antibody

BASF Pharma set out to develop a fully human, anti-TNF mAb structurally identical to naturally occurring human antibodies and therefore less likely to engender an immune response in the recipient. The generation of a fully human anti-TNF mAb required bioengineering techniques[9] that mimic immune selection in humans and contrast with existing means of "humanising" murine monoclonal antibodies in that the antibodies derived are completely human. The result of this effort is D2E7, a new class of anti-TNF mAb, which may have advantages in minimising antigenicity in humans.

## Preclinical pharmacology

The efficacy and safety of D2E7 were evaluated in a number of experimental systems. The ability of D2E7 to neutralise TNF bioactivity was demonstrated in three different in vitro cell systems. Additionally, the effectiveness of D2E7 in preventing polyarthritis was shown in a transgenic mouse model that mimics the clinical and histopathological progression of RA in humans.[10] D2E7 treated mice showed no clinical signs of arthritis during the 11 week study period. Microscopical examinations of the ankle joints of the animals showed no histopathological changes. In contrast, control mice developed severe arthritis with cartilage destruction and bone erosion.

## Preliminary clinical data

There are important similarities in the designs of the early clinical trials assessing D2E7 in RA. All studies enrolled patients with an established diagnosis of RA who also had active disease, as evidenced by having a combination of swollen and tender joints, increased concentrations of acute phase reactants, and prolonged early morning stiffness. In addition, all trials involved RA patients with long disease duration and a history of failure of several DMARDs. During the trials, patients were allowed to continue stable doses of NSAIDs and corticosteroids. Efficacy was assessed

using composite criteria, such as the American College of Rheumatology improvement criteria (ACR 20)[11] and the Disease Activity Score (DAS).[12] Such criteria, which require improvement in multiple variables, are more stringent than is improvement in one or only selected clinical variables. For example, to be classified as a responder according to ACR 20 criteria, patients must demonstrate: (1) greater than or equal to 20% improvement in swollen joint count; (2) greater than or equal to 20% improvement in tender joint count and; (3) at least 20% improvement in three of five other measures (patient global assessment of disease activity, physician global assessment of disease activity, patient assessment of pain, an acute phase reactant (for example, erythrocyte sedimentation rate (ESR) or C reactive protein), and a measure of disability (for example, the Health Assessment Questionnaire)). The DAS is a composite score of tender joints, swollen joints, ESR and patient's assessment of disease activity as measured on a visual analogue scale.

Preliminary results of early trials of D2E7 are available (table 2). At the time of this review, some of the studies referenced have only been published in abstract form. In a randomised, double blind, placebo controlled phase I study, 120 patients were treated with single doses of D2E7 given in an ascending fashion ranging from 0.5 to 10 mg/kg.[13] Patients were treated in cohorts of 24 (18 patients were treated with D2E7 and six received placebo). After a wash out period of three weeks, the patients received a single dose of D2E7 or placebo by intravenous injection over 3–5 minutes. Patients were followed up for at least four weeks to determine the pharmacokinetics of D2E7, as well as to evaluate the safety and clinical efficacy of the compound in terms of onset, duration and magnitude of response. Positive response was defined as a decrease of at least 1.2 (compared with baseline) in the DAS. All parameters needed to calculate the response criteria as defined by the ACR 20 were also measured. Patients in whom the effect of D2E7 had declined below response status by week 4 entered an open label extension study. However, patients who maintained a response at week 4 were continued without retreatment until their response status was lost. Thereafter, these patients could also continue in the extension study.

The data from this first therapeutic trial in humans were very encouraging. In the three highest dose groups, 40–70% of patients achieved DAS and ACR 20 response status at 24 hours to 29 days of treatment. The therapeutic effects became evident within 24 hours to one week after D2E7 administration and reached the maximum effect after 1–2 weeks, with dose response reaching a plateau at 1 mg/kg D2E7. In contrast, only 19% of patients taking placebo achieved response status. Single doses of D2E7 were well tolerated and the dose increment scheme was followed as planned reaching the maximum dose of 10 mg/kg without any evidence of clinically relevant or dose related adverse effects. Pharmacokinetic parameters were calculated for a total of 89 patients from all dose groups. The systemic drug exposure (AUC) increased proportionally with increased dose. The mean total serum clearance was 0.180 to 0.271 ml/min, and the steady state volume of distribution ranged from 0.063 to 0.076 l/kg indicating that distribution of D2E7 was mostly in the intravascular space. The estimated mean terminal half life was 11.6 to 13.7 days.

Patients who participated in the open label extension study received a second blinded dose identical to their first dose (medication was given after a minimum period of four weeks and only after loss of their initial response status).[14] From the third dose onwards, all patients were given active drug (that is, the placebo patients received D2E7 doses according to their dose group). D2E7 was administered every two weeks until responses could be rated as "good", defined as an absolute DAS of < 2.4. To measure the duration of the good response, these patients were retreated only upon disease flare up. To keep as many patients as possible in the study for the long term evaluation of safety, patients who did not respond well after 0.5 or 1 mg/kg received higher doses of up to a maximum of 3 mg/kg. Doses were kept constant in the patients on 3, 5 or 10 mg/kg. Treatment lasting several years is intended. The response in the DAS over time demonstrated sustained therapeutic effects and some continuing improvement after multiple infusions of D2E7. Response rates of more than 80% have been achieved with a mean dosing interval of 2.5 weeks. After six months, 86% of patients continued to receive treatment with D2E7 indicating that long term intravenous treatment with D2E7 in the dose range from 0.5 to 10 mg/kg was well tolerated.

The safety and efficacy of weekly subcutaneous administration of 0.5 mg/kg D2E7 was evaluated in 24 patients with active RA in a phase I placebo controlled trial.[15] After a wash

*Table 2 Early trials of D2E7 in rheumatoid arthritis*

| *Trial design* | *Patients* | *D2E7 dosing schedule* | *Concurrent DMARD* | *Maximum ACR 20 response* |
|---|---|---|---|---|
| Randomised, double blind, placebo controlled | n=120; 83% RF+; mean disease duration=12 years; mean DMARDs failed=3.6 | Single and multiple iv injections, ascending doses ranging from 0.5 to 10 mg/kg | No | 78% |
| Randomised, double blind, placebo controlled | n=24; mean disease duration=10 years; mean DMARDs failed=3.4 | Weekly 0.5 mg/kg sc injections | No | 70% |
| Randomised, double blind, placebo controlled | n=54; 87% RF+; mean disease duration=11 years; mean DMARDs failed (including MTX)=3.6 | Single iv or sc injection of 1 mg/kg | MTX | 67%* |

RF+=rheumatoid factor positive; DMARD=disease modifying anti-rheumatic drug; MTX= methotrexate; iv=intravenous; sc=subcutaneous. *With subcutaneous administration.

out period of three weeks, patients were treated with subcutaneous D2E7 or placebo for three months. The dose of D2E7 was increased to 1 mg/kg subcutaneously weekly for non-responders or those losing their responder status. Blood samples were collected to determine D2E7 plasma concentrations. All responding patients continued in an open label extension of this study. Based on preliminary data, plasma concentrations of D2E7 after multiple subcutaneous doses were comparable to those achieved with intravenous administration. Up to 78% of patients achieved a DAS/ACR 20 response after three months of treatment with subcutaneous D2E7. With the exception of mild and transient injection site reactions, adverse events occurred with the same frequency and distribution in the D2E7 and placebo groups. The investigators concluded that D2E7 given subcutaneously was safe and as effective as when administered intravenously demonstrating that subcutaneous self administration is a promising approach for D2E7 delivery.

Monotherapy is often inadequate to control arthritic symptoms and rapid progression of RA. D2E7 (1 mg/kg as a single subcutaneous or intravenous injection) was evaluated in a randomised, double blind, placebo controlled trial in patients whose stable dose of methotrexate was insufficient to control symptoms. An ACR 20 response was seen in 67% and 72% of patients receiving D2E7 by subcutaneous and intravenous injection, respectively. The safety profile of single dose D2E7 administration was comparable to that of placebo.

Collectively, these early data suggest that the fully human anti-TNFα mAb D2E7 is safe and effective as monotherapy or in combination with methotrexate when administered by single and multiple intravenous and subcutaneous injections. Additional studies are underway to further define optimal use of this novel treatment.

1 Brooks P. Clinical management of rheumatoid arthritis. Lancet 1993;341:286–90.
2 Maini R, Brennan F, Williams R, *et al.* TNF-a in rheumatoid arthritis and prospects of anti-TNF therapy. Clin Exp Rheumatol 1993;11 (suppl 8):S173–5.
3 Feldmann M, Brennan F, Chu C, *et al.* Does TNF-a have a pivotal role in the cytokine network in rheumatoid arthritis? In: Friers W, ed. *Tumor necrosis factor: molecular and cellular biology and clinical relevance.* Basel: Karger, 1993:144–52.
4 Elliott MJ, Maini RN, Feldmann M, *et al.* Treatment of rheumatoid arthritis with chimeric monoclonal antibodies to tumor necrosis factor alpha. Arthritis Rheum 1993;36: 1681–90.
5 Scallon B, Moore M, Trinh H, Knight D, Ghrayeb J. Chimeric anti-TNF-alpha monoclonal antibody cA2 binds recombinant transmembrane TNF-alpha and activates immune effector functions. Cytokine 1995; 7:251–9.
6 Rankin EC, Choy EH, Kassimos D, *et al.* The therapeutic effects of an engineered human anti-tumour necrosis factor alpha antibody (CDP571) in rheumatoid arthritis. Br J Rheumatol 1995; 34:334–42.
7 Moreland LW, Baumgartner SW, Schiff MH, *et al.* Treatment of rheumatoid arthritis with a recombinant human tumor necrosis factor receptor (p75)-Fc fusion protein. N Engl J Med 1997;337:141–7.
8 Kavanaugh AF. Anti-tumor necrosis factor-alpha monoclonal antibody therapy for rheumatoid arthritis. Rheum Dis Clin North Am 1998;24:593–614.
9 Jespers L, Roberts A, Mahler S, Winter G, Hoogenboom H. Guiding the selection of human antibodies from phage display repertoires to a single epitope of an antigen. Biotechnology (NY) 1994;12:899–903.
10 Keffer J, Probert L, Cazlaris H, *et al.* Transgenic mice expressing human tumour necrosis factor: a predictive genetic model of arthritis. EMBO J 1991;10:4025–31.
11 Felson D, Anderson J, Boers M, *et al.* The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical studies. Arthritis Rheum 1995;38:727–35.
12 van Gestel A, Prevoo M, Van'T Hoff M, *et al.* Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Arthritis Rheum 1996;39:34–40.
13 van de Putte L, van Riel P, den Broeder A, *et al.* A single dose placebo-controlled phase I study of the fully human anti-TNF antibody D2E7 in patients with rheumatoid arthritis. Arthritis Rheum 1998;41 (suppl):57.
14 Rau R, Sander O, den Broeder A, *et al.* Long term efficacy and tolerability of multiple iv doses of the fully human anti-TNF antibody D2E7 in patients with rheumatoid arthritis. Arthritis Rheum 1998;41 (suppl):55.
15 Schattenkirchner M, Krüger K, Sander O, *et al.* Efficacy and tolerability of weekly subcutaneous injections of the fully human anti-TNF antibody D2E7 in patients with rheumatoid arthritis. Arthritis Rheum 1998;41 (suppl):57.

# Adalimumab

From Wikipedia, the free encyclopedia

**Adalimumab** (HUMIRA, Abbott) is the third TNF inhibitor, after infliximab and etanercept, to be approved in the United States. Like infliximab and etanercept, adalimumab binds to TNFα, preventing it from activating TNF receptors; adalimumab was constructed from a fully human monoclonal antibody, while infliximab is a mouse-human chimeric antibody and etanercept is a TNF receptor-IgG fusion protein. TNFα inactivation has proven to be important in downregulating the inflammatory reactions associated with autoimmune diseases. As of 2008 adalimumab has been approved by the FDA for the treatment of rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis, Crohn's disease, moderate to severe chronic psoriasis and juvenile idiopathic arthritis.

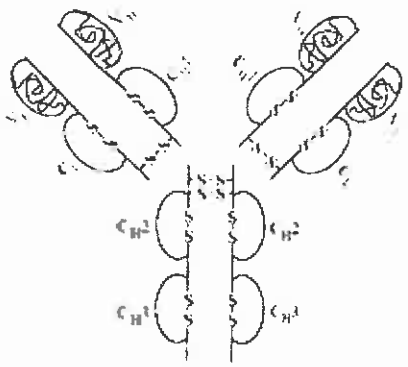
However, because TNFα is part of the immune system that protects the body from infection, prolonged treatment with adalimumab may slightly increase the risk of developing infections.

HUMIRA ("Human Monoclonal Antibody in Rheumatoid Arthritis") is marketed in both preloaded 0.8 mL syringes and also in preloaded pen devices (called Humira Pen), both injected subcutaneously, typically by the patient at home. It cannot be administered orally, because the digestive system would destroy the drug.

Adalimumab was the first fully human monoclonal antibody drug approved by the FDA. It was derived from phage display,[1] and was discovered through a collaboration between BASF Bioresearch Corporation (Worcester, Massachuetts, a unit of BASF) and Cambridge Antibody Technology as D2E7,[2] then further manufactured at BASF Bioresearch Corporation and developed by BASF Knoll (BASF Pharma) and, ultimately, manufactured and marketed by Abbott Laboratories after the acquisition of BASF Pharma by Abbott.

In 2009, HUMIRA had over $5bn in annual sales[3]

**Adalimumab** ?



| | |
|---|---|
| **Monoclonal antibody** | |
| **Type** | Whole antibody |
| **Source** | Human |
| **Target** | TNF alpha |
| **Clinical data** | |
| **Pregnancy cat.** | C(AU) B(US) |
| **Legal status** | POM (UK) ℞-only (US) |
| **Routes** | Subcutaneous |
| **Pharmacokinetic data** | |
| **Bioavailability** | 64% (subcutaneous), 0% (oral) |
| **Half-life** | 10–20 days. |
| **Identifiers** | |
| **CAS number** | 331731-18-1 |
| **ATC code** | L04AB04 |
| **DrugBank** | BTD00049 |
| **UNII** | FYS6T7F842 ✓ |
| **KEGG** | D02597 |
| **ChEMBL** | CHEMBL1201580 |
| **Chemical data** | |
| **Formula** | $C_{6428}H_{9912}N_{1694}O_{1987}S_{46}$ |
| **Mol. mass** | 144190.3 g/mol |

## Contents


- 1 History
- 2 Notable clinical and marketing milestones
- 3 Indications
  - 3.1 Rheumatoid arthritis
  - 3.2 Psoriatic arthritis
  - 3.3 Ankylosing spondylitis
  - 3.4 Crohn's disease
  - 3.5 Ulcerative Colitis
  - 3.6 Plaque psoriasis
  - 3.7 Juvenile Idiopathic Arthritis
- 4 Safety
- 5 Royalty litigation
- 6 Similar agents
- 7 References
- 8 External links


## History

Adalimumab was discovered as a result of the collaboration between BASF Bioresearch Corporation (Worcester, Massachuetts, a unit of BASF) and Cambridge Antibody Technology which began in 1993.[4]

The drug candidate was discovered initially using CAT's phage display technology and named D2E7.[2] The key components of the drug were found by guiding the selection of human antibodies from phage display repertoires to a single epitope of an antigen TNF alpha.[5] The ultimate clinical candidate, D2E7, was created and manufactured at BASF Bioresearch Corporation and taken through most of the drug development process by BASF Knoll, then further development, manufacturing and marketing by Abbott Laboratories, after Abbott acquired the pharmaceutical arm of BASF Knoll.[6]

## Notable clinical and marketing milestones

- 1999: Preliminary results of early clinical trials with the fully human anti-TNFalpha monoclonal antibody D2E7[2]
- 2001, June: Results from ARMADA, a double-blind, placebo-controlled clinical trial involving 271 patients with active rheumatoid arthritis despite treatment with methotrexate are announced. Among the results are that 50% of patients show a 50% improvement in ACR score.[7]
- 2002: Broke ground on a new state-of-the-art biologics manufacturing facility.[3]
- 2002: Adalimumab results from five separate trials show that it is effective at reducing signs and symptoms of rheumatoid arthritis. In these studies, adalimumab had a rapid onset of action and sustained efficacy. Furthermore, adalimumab was safe and effective when given alone or in combination with MTX as a subcutaneous injection.[8]

265

- 2002, December 31: HUMIRA approved by FDA for treatment of rheumatoid arthritis.[3]
- 2003: Launched HUMIRA for rheumatoid arthritis and continued clinical studies for additional indications.[3]
- 2005: Launched HUMIRA for psoriatic arthritis. Exceeded $1 billion in annual sales for the first time.[3]
- 2005, 10 December: Eisai Submits New Drug Application for Rheumatoid Arthritis Drug Adalimumab (D2E7) in Japan.[9]
- 2006: Submitted HUMIRA for the Crohn's disease indication and launched it for AS. Exceeded $2 billion in annual sales.[3]
- 2007: Launched HUMIRA for Crohn's disease in the United States,[3] submitted HUMIRA for global regulatory approval for psoriasis — the fifth new HUMIRA disease indication at this time, achieved more than $3 billion in worldwide HUMIRA sales.[10]
- 2007, 10 December: Abbott Opens New Biotechnology Manufacturing Facility in Puerto Rico[11]
- 2009, 10 June: Five-Year Data Demonstrate Initial Use of HUMIRA Plus Methotrexate May Prevent Further Joint Damage in Early Rheumatoid Arthritis Patients [12]

# Indications

Adalimumab, like its TNF inhibitor competitors, infliximab and etanercept, has proven versatile in its effectiveness in treating several conditions.

## Rheumatoid arthritis

Adalimumab has been shown to reduce the signs and symptoms of moderate-to-severe rheumatoid arthritis (RA) in adults. It has also been shown to have efficacy in moderate to severe polyarticular juvenile idiopathic arthritis (JIA) in children 4 years of age and older, and is approved for use in the treatment of that condition. In RA it can be used alone or with methotrexate or similar medicines.

## Psoriatic arthritis

Adalimumab (HUMIRA, Abbott Laboratories) is a fully human TNF-alpha monoclonal antibody approved for the treatment of rheumatoid arthritis and undergoing trials for use in treating other conditions, including psoriasis and psoriatic arthritis.[13]

## Ankylosing spondylitis

Adalimumab has been shown to reduce the signs and symptoms of, and is approved for treatment of, ankylosing spondylitis (AS) in adults.

## Crohn's disease

Adalimumab has been shown to reduce the signs and symptoms[14] of, and is approved for treatment of, moderate to severe Crohn's disease.

### Ulcerative Colitis

Adalimumab may be effective and well tolerated in Ulcerative Colitis. Its efficacy in maintaining clinical remission needs to be confirmed in a randomized controlled trial. [15]

### Plaque psoriasis

Adalimumab has been shown to treat moderate to severe chronic (lasting a long time) plaque psoriasis (Ps) in adults who have the condition in many areas of their body and who may benefit from taking injections or pills (systemic therapy) or phototherapy (treatment using ultraviolet light alone or with pills).[16] Adalimumab has been shown to be effective therapy when used either continuously or intermittently in patients with moderate to severe psoriasis.

### Juvenile Idiopathic Arthritis

Adalimumab has been shown to reduce the signs and symptoms of moderate to severe polyarticular juvenile idiopathic arthritis (JIA) in children 4 years of age and older. It is commonly used in combination with methotrexate and administered via a 20 mg pre-filled pediatric syringe, 40 mg pre-filled syringe or a 40 mg pen.[17]

## Safety

Because adalimumab suppresses TNF, which is part of the immune system, latent infections, such as tuberculoses, can be reactivated, and the immune system may be unable to fight new infections. This has led to fatal infections.[18]

According to the product labeling, after a number of studies and reports of adverse events in patients receiving adalimumab, including serious and sometimes fatal blood disorders; serious infections including TB (tuberculosis) and infections caused by viruses, fungi, or bacteria; rare reports of lymphoma[2] (http://www.fdable.com/advanced_aers_query/c92ac23d3e2c43d72f398360cda8ce5c) and solid tissue cancers; rare reports of serious liver injury; and rare reports of demyelinating central nervous system disorders); rare reports of cardiac failure; the U.S. Food and Drug Administration issued a black box warning to doctors appearing in the product labeling of adalimumab and the other TNF drugs instructing them to screen and monitor potential patients more carefully.[19]

## Royalty litigation

In March 2003, British company Cambridge Antibody Technology (CAT), stated its wish to "initiate discussions regarding the applicability of the royalty offset provisions for HUMIRA" with Abbott Laboratories in the High Court of London. In November 2004, the trial began, and in December 2004, the Judge, The Hon. Mr Justice Laddie, ruled for CAT. In an unusual step, a draft of the judgement was not made available in advance.

A short version of the full statement of the proceedings was released.[20] In it Justice Laddie remarked, "Abbott was in error when it made its first royalty payment to CAT calculated on the basis that only 2% of the Net Sales was due. It should have calculated on the basis of the full royalty of just over 5% and should

have paid and continued to pay CAT accordingly." Justice Laddie went on to observe "...that the construction advanced by Abbott does violence to the language of the agreements, renders them obscure and makes little or no commercial sense. For this reason CAT wins the action.".[21]

Abbott was required to pay CAT US$255m, some of which was to be passed to its partners in development.[22] Of this sum, the Medical Research Council received US$191m, and in addition, Abbott was asked to pay the MRC a further US$7.5m over five years from 2006, providing that HUMIRA remains on the market. The MRC also is to receive a further £5.1m (sterling) in respect of past royalties.[23]

## Similar agents

- Infliximab
- Etanercept

## References


1. ^ Brekke OH , Sandlie I (January 2003). "Therapeutic antibodies for human diseases at the dawn of the twenty-first century" (http://www.nature.com/nrd/journal/v2/n1/full/nrd984.html) . *Nat Rev Drug Discov* **2** (1): 52–62. doi:10.1038/nrd984 (http://dx.doi.org/10.1038%2Fnrd984) . http://www.nature.com/nrd/journal/v2/n1/full/nrd984.html.
2. ^ *a b c* Kempeni J (January 1999). "Preliminary results of early clinical trials with the fully human anti-TNFα monoclonal antibody D2E7" (http://ard.bmj.com/content/58/suppl_1/I70.full) . *Ann Rheum Dis* **58** (suppl 1): I70. doi:10.1136/ard.58.2008.i70 (http://dx.doi.org/10.1136%2Fard.58.2008.i70) . PMC 1766582 (http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=1766582) . PMID 10577977 (http://www.ncbi.nlm.nih.gov/pubmed/10577977) . http://ard.bmj.com/content/58/suppl_1/I70.full.
3. ^ *a b c d e f g* http://www.abbott.com/static/content/microsite/annual_report/2006/humira.html
4. ^ [1] (http://www.cambridgeantibody.com/home/news_and_resources/news_archive/2003/cambridge_antibody_technology__humira) Cambridge Antibody Technology website
5. ^ Jespers LS, Roberts A, Mahler SM, Winter G, Hoogenboom HR (September 1994). "Guiding the selection of human antibodies from phage display repertoires to a single epitope of an antigen". *Biotechnology (N.Y.)* **12** (9): 899–903. PMID 7521646 (http://www.ncbi.nlm.nih.gov/pubmed/7521646) .
6. ^ http://www2.basf.us/corporate/news2000/newsknoll_pharma_121500.html
7. ^ http://www.prnewswire.com/cgi-bin/stories.pl?ACCT=104&STORY=/www/story/11-12-2001/0001613559&EDATE=
8. ^ Rau R (January 2002). "Adalimumab (a fully human anti-tumour necrosis factor α monoclonal antibody) in the treatment of active rheumatoid arthritis: the initial results of five trials" (http://ard.bmj.com/content/61/suppl_2/ii70.full) . *Ann Rheum Dis* **61** (Suppl 2): ii70. doi:10.1136/ard.61.suppl_2.ii70 (http://dx.doi.org/10.1136%2Fard.61.suppl_2.ii70) . PMC 1766697 (http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=1766697) . PMID 12379628 (http://www.ncbi.nlm.nih.gov/pubmed/12379628) . http://ard.bmj.com/content/61/suppl_2/ii70.full.
9. ^ http://www.thefreelibrary.com/Eisai+Submits+New+Drug+Application+for+Rheumatoid+Arthritis+Drug...-a0140148456
10. ^ http://www.abbott.com/static/content/microsite/annual_report/2007/overview_milestones.html
11. ^ http://www.abbott.com/global/url/pressRelease/en_US/60.5:5/Press_Release_0439.htm
12. ^ http://www.abbott.com/global/url/pressRelease/en_US/60.5:5/Press_Release_0737.htm
13. ^ Scheinfeld, N (2003). "Adalimumab (HUMIRA): a review". *J Drugs Dermatol* **2** (4): 375–7. PMID 12884458 (http://www.ncbi.nlm.nih.gov/pubmed/12884458) .
14. ^ Podolsky, Daniel K. (August 2002). "Inflammatory bowel disease" (http://content.nejm.org/cgi/content/extract/347/6/417) . *N Engl J Med* **346** (6): 417–29. doi:10.1056/NEJMra020831 (http://dx.doi.org/10.1056%2FNEJMra020831) . PMID 12167685 (http://www.ncbi.nlm.nih.gov/pubmed/12167685) .

http://content.nejm.org/cgi/content/extract/347/6/417.
15. ^ Fiorino G, Peyrin-Biroulet L, Repici A, Malesci A, Danese S (January 2011). "Adalimumab in ulcerative colitis: hypes and hopes" (http://informahealthcare.com/doi/abs/10.1517/14712598.2011.541435) . *Expert Opin Biol Ther* **11** (1): 109–16. doi:10.1517/14712598.2011.541435 (http://dx.doi.org/10.1517%2F14712598.2011.541435) . PMID 21133817 (http://www.ncbi.nlm.nih.gov/pubmed/21133817) . http://informahealthcare.com/doi/abs/10.1517/14712598.2011.541435.
16. ^ Croom, Katherine F; McCormack, Paul L (2009). "Adalimumab". *Am J Clin Dermatol* **10** (1): 43–50. doi:10.2165/0128071-200910010-00008 (http://dx.doi.org/10.2165%2F0128071-200910010-00008) . PMID 19170412 (http://www.ncbi.nlm.nih.gov/pubmed/19170412) .
17. ^ http://www.rxabbott.com/pdf/humira.pdf
18. ^ Karen M. Frank, et al, Investigation of the cause of death in a gene-therapy trial, N Engl J Med 361:161, July 9, 2009
19. ^ FDA label - http://www.rxabbott.com/pdf/humira_medguide.pdf
20. ^ http://www.chandb.com/MBE/assessment_041220_CaT_Abbott_judgement_summary.doc
21. ^ "Biotech firm wins royalty fight" (http://news.bbc.co.uk/1/hi/business/4112235.stm) . *BBC News*. 2004-12-20. http://news.bbc.co.uk/1/hi/business/4112235.stm. Retrieved 2010-04-23.
22. ^ Murray-, Rosie (2005-10-27). "Drug maker CAT surges after royalty settlement" (http://www.telegraph.co.uk/finance/2924694/Drug-maker-CAT-surges-after-royalty-settlement.html) . *The Daily Telegraph* (London). http://www.telegraph.co.uk/finance/2924694/Drug-maker-CAT-surges-after-royalty-settlement.html. Retrieved 2010-04-23.
23. ^ http://www.mrctechnology.org/resources/PR/MRC_Humira_PRESS_RELEASE.pdf

golimumab monthly

## External links

- HUMIRA Official site (http://www.humira.com)
- myHUMIRA Official site (http://www.myhumira.com)
- Google Timeline Search (http://www.google.com/views?hl=en&safe=off&q=humira++view%3Atimeline&btnG=Search)
- Humira Adverse Events Reported to the FDA (http://www.drugcite.com?q=HUMIRA)

Retrieved from "http://en.wikipedia.org/wiki/Adalimumab"
Categories: TNF-alpha inhibitors | Monoclonal antibodies | Immunosuppressants

- This page was last modified on 1 July 2011 at 02:12.
- 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.

269

**Improve binding affinity of therapeutic antibody**

All the above mentioned methods of constructing therapeutic antibodies depend on the ability of the mouse or human to generate antibodies on their own. Many different antibodies are usually produced against one specific antigen with varying binding affinities. In most cases, the higher binding affinity ones are selected. However, they may still not have the maximum possible binding affinity.

Detailed analysis of the antigen-antibody complex can provide useful information about the contacting residues between these two macro-molecules. Locally, it is thus possible to improve these contacting amino acid residues to improve their charge-charge interaction or hydrophobic interaction. Minor changes of one or two amino acid residues in the six CDR's especially the CDRH3, may thus increase the binding affinity (Lippow SM, Wittrup KD, Tidor B (2007) Computational design of antibody-affinity improvement beyond *in vivo* maturation. Nature Biotech, 25, 1161-1176).

Several antibodies were analyzed by Lippow and Wittrup. Among them, one was a chimeric therapeutic antibody, cetuximab (Mendelsohn J (1989) Anti-EGF receptor monoclonal antibodies: biological studies and potential clinical applications. Tran Am Clin Climatol Assoc, 100, 31-38). By changing the following three amino acid residues, the binding affinity was increased by about ten-fold:

| | | |
|---|---|---|
| L26 | (CDRL1) | Ser to Asp |
| L31 | (CDRL1) | Thr to Glu |
| L93 | (CDRL3) | Asn to Ala |

Two of the amino acid residues in CDRL1, Ser and Thr, were substituted with two charged amino acid residues, Asp and Glu. CDRH3 was not altered. As explained before, CDRH3 defines the fine specificity of the antibody while the other five CDR's can improve the binding affinity.

They also studied an anti-lysozyme antibody and managed to increase its binding affinity by about 140-fold with the change of several amino acid residues in CDRL3, CDRH1, CDRH2, CDRH3 and one amino acid residue in FRH1. Of course, this has no relevance to therapeutic antibodies.

nature biotechnology

© 2007 Nature Publishing Group http://www.nature.com/naturebiotechnology

# Computational design of antibody-affinity improvement beyond *in vivo* maturation

Shaun M Lippow[1,4], K Dane Wittrup[1,2] & Bruce Tidor[2,3]

**Antibodies are used extensively in diagnostics and as therapeutic agents. Achieving high-affinity binding is important for expanding detection limits, extending dissociation half-times, decreasing drug dosages and increasing drug efficacy. However, antibody-affinity maturation *in vivo* often fails to produce antibody drugs of the targeted potency[1], making further affinity maturation *in vitro* by directed evolution or computational design necessary. Here we present an iterative computational design procedure that focuses on electrostatic binding contributions and single mutants. By combining multiple designed mutations, a tenfold affinity improvement to 52 pM was engineered into the anti–epidermal growth factor receptor drug cetuximab (Erbitux), and a 140-fold improvement in affinity to 30 pM was obtained for the anti-lysozyme model antibody D44.1. The generality of the methods was further demonstrated through identification of known affinity-enhancing mutations in the therapeutic antibody bevacizumab (Avastin) and the model anti-fluorescein antibody 4-4-20. These results demonstrate computational capabilities for enhancing and accelerating the development of protein reagents and therapeutics.**

Computational design depends critically on two capabilities: accurate energetic evaluation and thorough conformational search. Previous work has addressed many problems related to the design of improved protein-protein binding affinity, such as the design of stable protein folds[2–4], binding pockets for peptides and small molecules[5–7], altered protein-protein specificity[8–12] and altered enzymatic activity[13–15]. The design of improved binding affinity has met with limited success, however[16–19]. Challenges for protein-protein affinity design include conformational change upon binding, interfacial trapped water molecules, polar and charged side chains and the trade-off of protein-solvent with protein-protein interactions from the unbound to bound state. Fine free-energy discrimination for redesign from nanomolar to picomolar affinities is a particular challenge.

A robust design strategy should both produce a considerable fraction of designs that are successful when tested experimentally and yield substantial improvements across multiple systems. Although there are potentially many mutations that confer improved binding affinity for a particular interaction, calculations need identify only a subset to be successful. Our approach uses thorough optimization techniques that exhaustively rank-order the best solutions in a discretized search space. Although some of these solutions are expected to be improved designs, others will be unsuccessful but may be useful in learning about deficiencies in the energy functions, search procedures or other methodology.

First we attempted to redesign the model antibody D1.3 for improved binding to its antigen, hen egg-white lysozyme. Single mutations at each of 60 complementarity-determining region (CDR) positions to the 20 common side chains, excluding proline and cysteine, were designed using a physics-based energy function and a hierarchical search procedure. (Physics-based refers to an energy function constructed from and parameterized by theoretical and experimental models of the underlying physical interactions; by contrast, knowledge-based refers to potentials statistically derived from observational data, which often take the form of structure databases.) A striking feature was that predictions for improved binding were dominated by mutations to large amino acids (**Supplementary Fig. 1** online), many of which exhibited improved van der Waals packing interactions that outweighed disfavored net Poisson-Boltzmann continuum electrostatic solvation and interaction. Our confidence in these designs was low because many had unsatisfied hydrogen bonding or a predicted packing improvement of a magnitude uncharacteristically large for a single amino-acid substitution. Nevertheless, 17 single mutations, most with calculated improved total binding free energy, were selected for experimental binding-affinity measurement. Only three mutations improved affinity (**Supplementary Table 1** online), with 2.4-fold improvement for the best single mutant (**Supplementary Fig. 2** online). We found that the calculated electrostatic term for binding was a better predictor for improvement than the total calculated binding free energy, that improvements from mutations to larger amino acids were mostly not realized and that avoiding potentially destabilizing mutations was important. This led us to ask: (i) Could electrostatics-based predictions alone be used to design binding-affinity improvements? (ii) Is there a physics-based explanation for the seemingly inaccurate calculated packing improvements?

In our second attempt, we used only the electrostatic term of the computed binding free energy to predict improvements in affinity but kept the original design procedure and full energy function for side-chain conformational search. Because there were few computed

[1]Department of Chemical Engineering, [2]Department of Biological Engineering, [3]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. [4]Present address: Codon Devices, Inc., One Kendall Square, Building 300, Cambridge, Massachusetts 02139, USA. Correspondence should be addressed to B.T. (tidor@mit.edu) or K.D.W. (wittrup@mit.edu).

Received 20 March; accepted 9 August; published online 23 September 2007; doi:10.1038/nbt1336

© 2007 Nature Publishing Group http://www.nature.com/naturebiotechnology

opportunities for improving D1.3 based on electrostatics, new antibodies were chosen to explore both this method and the possibility that D1.3 is anomalous.

Our second target for redesign was the antibody D44.1. It binds lysozyme, facilitating experiments, but its epitope is different from that of D1.3. In addition, D44.1 has low nanomolar affinity, maintaining the challenge of nanomolar-to-picomolar affinity maturation. Single mutations were designed at all D44.1 CDR positions, and then mutations were ranked by the electrostatic binding free-energy term. In contrast to the D1.3 design, there were many computed opportunities for electrostatic improvement. We measured experimental binding affinity for the nine largest-magnitude predictions, choosing no more than two mutations per position. Six of the nine mutants bound tighter than wild type, and the best mutation, L92 Asn-to-Ala, exhibited an eightfold improvement (**Table 1a** and **Fig. 1a**). In addition, following the original design procedure, two mutations predicted to increase packing interactions were tested; as expected, neither mutation led to an improvement in affinity (**Table 1a**).

The successful D44.1 single mutations were combined and tested experimentally. The six favorable mutations span five positions, where L32 and L92 are the only positions in direct contact with each other in the wild-type crystal structure. Contrary to prediction, this double mutant is not as improved as the L92 single mutant alone. The mutations at the other three positions are additive with L92, yielding a quadruple mutant with 43-pM affinity, 100-fold improved over wild type (Table 1a and **Fig. 1a**). **Figure 1b–d** displays the designed interactions. Measurements of dissociation and association rate-constants for D44.1 and the high-affinity quadruple mutant revealed that both kinetic steps were improved, including a 23-fold slower off-rate (**Table 1b**). Each ratio of $k_{off}$ to $k_{on}$ is in agreement with the independently measured $K_d$. Increasing ionic strength from 167 mM to 1.67 M screens the designed quadruple-mutant interaction by approximately fourfold, in agreement with calculation (+0.9 kcal/mol), whereas the affinity of the wild-type antibody is marginally improved at high salt, in contrast to calculation (+0.3 kcal/mol; **Table 1b**).

Two sets of calculations were made in an attempt to further improve the high-affinity D44.1 quadruple mutant. First, single mutations were designed based on the predicted structure of the quadruple mutant. These calculations reiterated many predictions seen in the original D44.1 design that had been ranked lower and not tested experimentally, as

**Table 1 Predicted and experimental antibody mutations**

(a) Predicted and experimental binding affinities in D44.1

| Position | Mutation | $K_d^{WT}/K_d^{mutant}$ | $\Delta\Delta G_{exp}$[a] | $\Delta\Delta G_{calc,\ total}$[b] | $\Delta\Delta G_{calc,\ elec}$[c] |
|---|---|---|---|---|---|
| D44.1 ($K_d^{WT}$ = 4.4 ± 0.5 nM) | | | | | |
| Single mutations, selected based on electrostatic energy | | | | | |
| L32 Asn | Gly | 4.2 ± 0.5 | −0.86 ± 0.07 | +0.99 | −1.03 |
| L92 Asn | Ala | 8.3 ± 1.3 | −1.25 ± 0.10 | −0.02 | −0.86 |
| H28 Thr | Asp | 1.29 ± 0.06 | −0.15 ± 0.03 | −0.30 | −0.26 |
| H31 Thr | Ala | 0.47 ± 0.07 | +0.45 ± 0.09 | +0.31 | −0.97 |
| H31 Thr | Val | 0.41 ± 0.09 | +0.53 ± 0.13 | −2.06 | −0.36 |
| H57 Ser | Ala | 1.9 ± 0.5 | −0.37 ± 0.16 | −0.59 | −1.44 |
| H57 Ser | Val | 2.3 ± 0.4 | −0.49 ± 0.12 | −2.29 | −1.34 |
| H58 Thr | Asp | 2.58 ± 0.16 | −0.56 ± 0.04 | −0.59 | −0.53 |
| H65 Lys | Asp | 0.97 ± 0.12 | +0.02 ± 0.07 | −0.38 | −0.42 |
| Single mutations, selected based on total energy | | | | | |
| L32 Asn | Tyr | 1.0 ± 0.2 | +0.00 ± 0.13 | −0.29 | +2.59 |
| H31 Thr | Trp | 0.8 ± 0.3 | +0.13 ± 0.2 | −3.75 | +1.56 |
| Combinations of single mutations | | | | | |
| (L32)G, (L92)A | | 3.2 ± 0.3 | −0.69 ± 0.05 | +1.23 | −1.68 |
| (H57)V, (H58)D | | 5.3 ± 0.4 | −0.99 ± 0.04 | −2.88 | −1.87 |
| (L32)G, (L92)A, (H57)V, (H58)D | | 18.7 ± 1.9 | −1.74 ± 0.06 | −1.65 | −3.55 |
| (L92)A, (H57)V, (H58)D | | 57 ± 5 | −2.40 ± 0.05 | −2.90 | −2.73 |
| (L92)A, (H28)D, (H57)V, (H58)D[d] | | 102 ± 3 | −2.74 ± 0.02 | −3.18 | −2.96 |

(b) Characterization of D44.1 quadruple mutant

| | $k_{off}$ ($10^{-3}$ $s^{-1}$) | $k_{on}$ ($10^6$ $M^{-1}s^{-1}$) | $k_{off}/k_{on}$ (nM) | $K_d$ (nM) | High-salt $K_d$ (nM) |
|---|---|---|---|---|---|
| D44.1 | 10.7 ± 1.0 | 2.52 ± 0.13 | 4.3 ± 0.5 | 4.4 ± 0.5 | 3.0 ± 0.8 |
| Quad[d] | 0.466 ± 0.014 | 8.4 ± 1.7 | 0.055 ± 0.012 | 0.0430 ± 0.0013 | 0.18 ± 0.08 |

(c) Double mutant cycle added to D44.1 quadruple mutant

| Position(s) | Mutation(s) | $K_d^{WT}/K_d^{mutant}$ | $\Delta\Delta G_{exp}$[a] |
|---|---|---|---|
| Quad[d] | | 102 ± 3 | −2.74 ± 0.02 |
| H35 Glu | Ser | < 0.04 | > +2 |
| H99 Gly | Asp | 1.41 ± 0.17 | −0.20 ± 0.07 |
| H35 Glu + H99 Gly[e] | Ser + Asp | 145 ± 17 | −2.95 ± 0.07 |

(d) Predicted and experimental binding affinities in cetuximab (Erbitux)

| Position | Mutation | $K_d^{WT}/K_d^{mutant}$ | $\Delta\Delta G_{exp}$[a] | $\Delta\Delta G_{calc,\ total}$[b] | $\Delta\Delta G_{calc,\ elec}$[c] |
|---|---|---|---|---|---|
| cetuximab ($K_d^{WT}$ = 0.49 ± 0.06 nM) | | | | | |
| L26 Ser | Asp | 1.4 ± 0.7 | −0.2 ± 0.3 | −0.40 | −0.38 |
| L31 Thr | Glu | 2.4 ± 0.8 | −0.53 ± 0.19 | −0.50 | −0.48 |
| L93 Asn | Ala | 3.5 ± 1.4 | −0.7 ± 0.2 | +1.06 | −0.75 |
| H56 Asn | Ala | 1.1 ± 0.4 | −0.04 ± 0.2 | +2.25 | −0.27 |
| H61 Thr | Glu | 1.1 ± 0.3 | −0.07 ± 0.17 | −0.77 | −0.77 |
| Combinations of single mutations | | | | | |
| (L31)E, (L93)A | | 4 ± 2 | −0.8 ± 0.3 | +0.51 | −1.28 |
| (L26)D, (L31)E, (L93)A | | 9.5 ± 0.3 | −1.33 ± 0.02 | +0.13 | −1.64 |
| (L26)D, (L31)E, (L93)A, (H56)A, (H61)E | | 3 ± 3 | −0.7 ± 0.5 | +1.69 | −2.61 |

Experimental error is the s.d. from two or more (typically three) independent trials. WT, wild type.
[a]Experimental binding affinity relative to WT (kcal/mol). [b]Calculated binding affinity relative to WT, all energy function terms (kcal/mol). [c]Electrostatic term of calculated binding affinity relative to WT (kcal/mol). [d]D44.1 quadruple mutant. 'Quad': (L92)A, (H28)D, (H57)V, (H58)D. [e]D44.1 hexluple mutant. 'Hex': (L92)A, (H28)D, (H57)V, (H58)D, (H35)S, (H99)D.

© 2007 Nature Publishing Group http://www.nature.com/naturebiotechnology

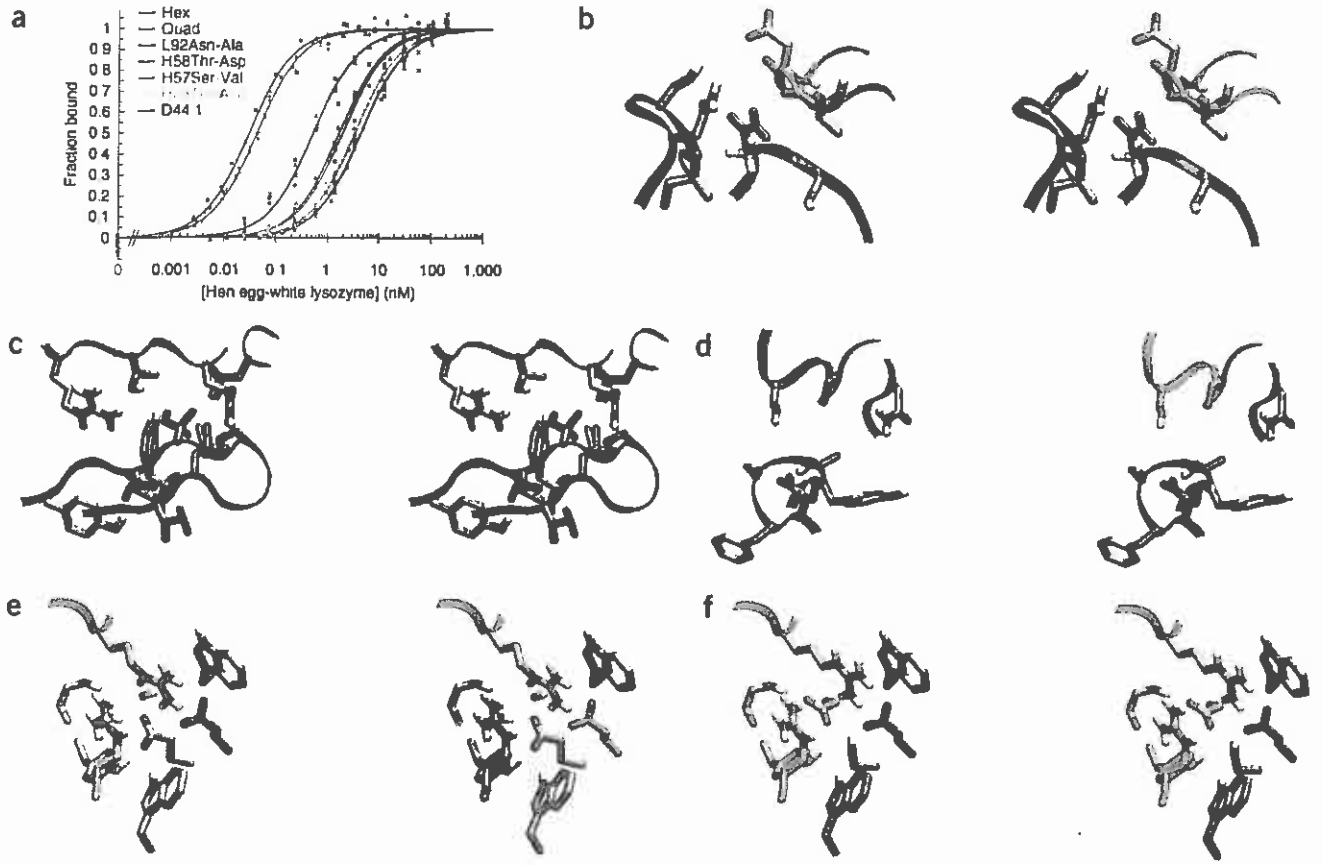



**Figure 1** Designed high-affinity mutations in D44.1. (a) Experimental binding affinities, from left to right: 140-fold hex, 100-fold quad, L92 Asn-to-Ala, H58 Thr-to-Asp, H57 Ser-to-Val, H28 Thr-to-Asp, wild type. For each variant, different symbol shapes distinguish independent measurements. (b–d) Predicted structures for single mutations; green ribbon, lysozyme backbone; magenta ribbon, antibody backbone; atom color, design; gray, wild type. L92 Asn-to-Ala (b). H57 Ser-to-Val, top; H58 Thr-to-Asp, bottom (c). H28 Thr-to-Asp (d). (e) Wild type in region of double mutation. The top-center arginine is from lysozyme and all other residues are from the antibody. (f) Designed double mutation: H99 Gly-to-Asp, H35 Glu-to-Ser. The asparagine is predicted to displace a structural water molecule.

well as predictions for mutating position L32 that had been found experimentally not to be additive. Second, double mutations were designed at all 93 pairs of contacting positions, and triple mutations were designed at positions H32, H98 and H100, based on cooperative packing at these positions. The double- and triple-mutant designs were filtered for favorable, cooperative predictions, requiring the energy of the double or triple mutation to exceed the energy of each of its component single mutations and the sum of the single mutation energies. Four single mutations, three double mutations and one triple mutation were selected for experimental testing in the quadruple-mutant context (**Supplementary Table 2** online). Only the H35+H99 double mutant (hextuple mutant from wild type) was improved relative to the quadruple mutant, at 30-nM affinity, 140-fold improved over wild type (**Table 1c** and **Fig. 1a**). Measurement of the individual H35 and H99 mutations revealed high cooperativity (**Table 1c**), consistent with the predicted salt-bridge and hydrogen-bond rearrangements (**Fig. 1e,f**). The diminished success rate of this subsequent design round may be due to smaller-magnitude predictions, antibody destabilization or design in a modeled rather than an experimentally determined structure.

Next, we applied our electrostatics-based methods to the anticancer therapeutic antibody cetuximab (Erbitux), which binds epidermal growth factor receptor (EGFR) to block ligand binding[20]. Calculations revealed nine positions with opportunities for affinity-enhancing single mutations based on improved electrostatics. The five mutations of largest magnitude of predicted affinity improvement were selected for experimental testing in the single-chain antibody format, with one mutation per position, and no two positions in close proximity. The EGFR extracellular domain (EGFR-ECD) mutant used in the assays contains a point mutation that likely interferes with the H56 mutation. Three of the other four mutants bind EGFR tighter than does cetuximab; these three mutations were combined to produce a triple mutant with tenfold overall improvement, from 490 pM to 52 pM (**Table 1d** and **Fig. 2**).

The transferability and utility of these electrostatics-based methods were further demonstrated by designs for which published data validate the predictions. One of our test cases was the antibody 4-4-20, which binds its small-molecule hapten, fluorescein, with 1.2-nM affinity, and was previously engineered using directed evolution to yield the antibody 4M5.3, with over 1,000-fold affinity improvement and 14 mutations[21]. Our designed single mutations in 4-4-20 revealed opportunities for improvement based on computed electrostatics at nine positions. Two of the predicted mutations, H31 Asp-to-His and H101 Ser-to-Ala, have already been shown to improve binding affinity as single mutations in 4M5.3 (ref. 22). Interestingly, the design missed the H102 Tyr-to-Ser mutation in 4M5.3 because of the rigid backbone



273

© 2007 Nature Publishing Group http://www.nature.com/naturebiotechnology

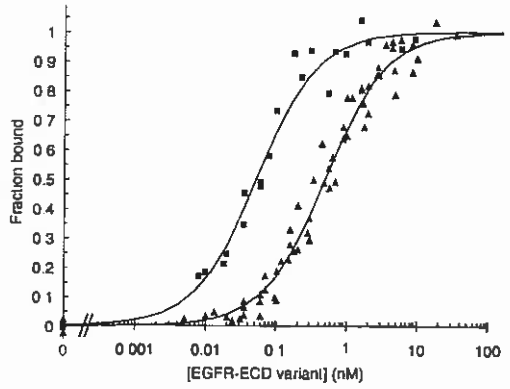



**Figure 2** Designed high-affinity cetuximab mutant. Experimental binding affinity titrations as displayed for cetuximab (triangles) and tenfold improved triple-mutant (squares).

constraint. Another test case was bevacizumab (Avastin), a therapeutic antibody that binds vascular endothelial growth factor (VEGF)[23], where we found five positions suitable for electrostatics-based improvement. The mutation H28 Thr-to-Asp and mutations at H31 and H101 are found in a published high-affinity variant of bevacizumab[24]. Our other electrostatics-based predictions in both test systems remain to be tested.

The single mutations designed and experimentally validated in this work lead to enhanced binding affinity through one of two electrostatic mechanisms. In one mechanism, a poorly satisfied polar group (a polar residue calculated to lose more free energy from desolvation than is recovered by interaction) is mutated to a hydrophobic residue. These mutations account for the majority of improved energetics. In a second mechanism—the addition of a charged residue—net charge is changed to increase electrostatic interaction, often at the periphery of the antibody-antigen interface where desolvation is minimal. Unlike previous work using electrostatics to guide design[25,26], our methods explicitly model the mutation, calculate a binding free energy relative to wild type, include positions that are partially or fully buried upon binding and avoid opportunities where the mutation is predicted to destabilize the mutant protein.

The modeled structures result from optimization with the full energy function, yet the van der Waals and nonpolar solvation energies are then discarded to predict improvement based on only the net electrostatic term. A potential problem is that different or wrong van der Waals parameters would produce altered equilibrium distances and hence altered electrostatic energies, particularly for short-range interactions. The specific mutations predicted and tested here may be less susceptible to this potential problem because they involve the calculation of solvation and medium-range electrostatic energies, which are not especially sensitive to precise atom locations. The energy for substitution of an unsatisfied polar group is dominated by its desolvation penalty, and the addition of a charged group at the interface periphery is not sensitive to precise side-chain placement. However, designs involving the introduction of a new short-range electrostatic contact, such as a new hydrogen bond, could be susceptible to this potential problem. Further work is necessary, but it could be that the energetic terms are actually appropriate for identifying low-energy structures for any sequence, but not for comparing energetics between sequences. Indeed, we think that the packing and nonpolar interactions are not balanced in some situations for accurate comparison of one amino acid to another, as described next.

We investigated the underlying physical model to address the seemingly inaccurate calculated packing interactions. The majority of problematic designs were at the binding-site periphery, where mutation to a larger amino acid was predicted to be favorable due to increased intermolecular van der Waals interactions. In some cases, hydrogen-bonding groups were buried, but the unfavorable electrostatic term was outweighed by improved packing. In principle, a larger amino acid at the interface periphery will exhibit increased protein-protein interactions in the bound state, counteracted by increased protein-water interactions in the unbound state. However, these offsetting energetics are calculated asymmetrically, with an atomistic

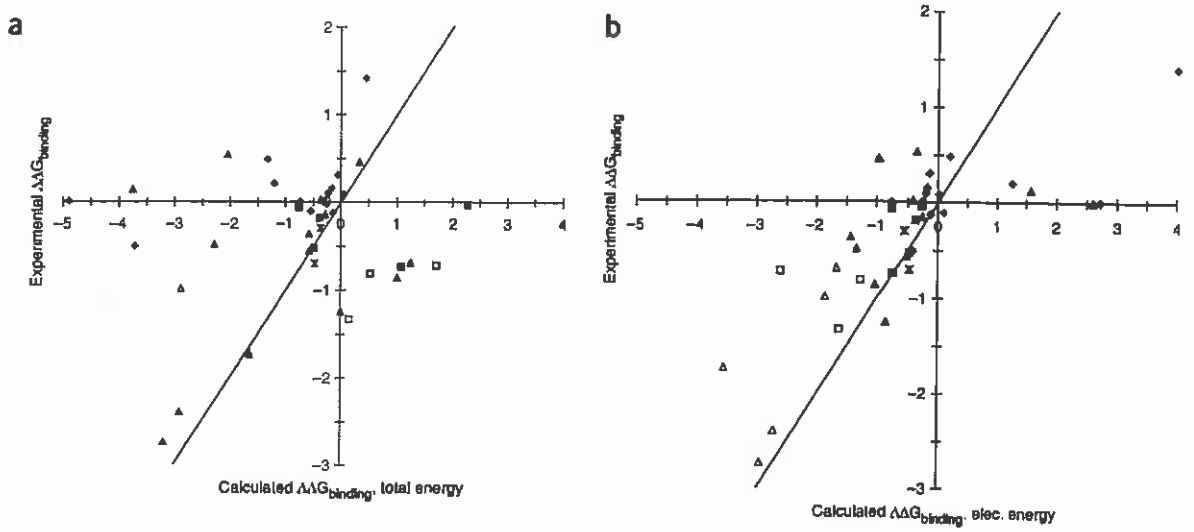



**Figure 3** Comparison of calculated and experimental binding free energies. Free energies are in kcal/mol relative to wild type, with the y = x line included to aid interpretation. Diamonds, D1.3 mutants; filled triangles, D44.1 single mutants; open triangles, D44.1 combination mutants; filled squares, cetuximab single mutants; open squares, cetuximab combination mutants; asterisks, 4-4-20 mutants. (a) Calculated total free energy. (b) Calculated electrostatic free-energy term.

© 2007 Nature Publishing Group http://www.nature.com/naturebiotechnology

Lennard-Jones potential for the protein-protein interactions and a simple solvent-accessible surface area nonpolar term for the protein-water interactions. The protein-water nonpolar interaction should be a function of the detailed geometry, including volume effects, as well as the particular protein atom types present. We implemented and parameterized an atomically detailed nonpolar term based on work by Levy and coworkers (**Supplementary Fig. 3** online)[27], and found that it reduced the magnitude of favorable prediction for many of the counterintuitive designs to larger side chains (**Supplementary Fig. 4** online). In addition, unbound-state conformational search attenuated predictions for larger side chains by introducing an energetic cost to adopt the binding conformation (data not shown). Nevertheless, some incorrect predictions remain, and future work should address this issue.

This work presents a computational alternative to directed evolution for affinity maturation. Directed evolution is adept at accumulating successive, additive mutations, but (with the exception of large-scale shuffling) is less well suited for selecting variants whose encoding DNA is further from wild type. Experimental libraries generated using error-prone PCR generally do not cover all single amino-acid mutations, let alone all pairs or greater combinations of mutations, as 13 of the 19 possible single mutations require more than one base-pair change, on average. Some classes of mutation require two base-pair changes—for example, mutation to either negatively charged residue from any codon of 10 of the 18 non-negatively-charged side chains. Of the 12 single mutations found to improve D1.3, D44.1 or cetuximab, ten required two base-pair changes and would therefore have been substantially more difficult to identify by a method that relies on error-prone PCR. Also, the H35+H99 cooperative double mutation required concerted amino-acid mutation and three total base-pair changes. Computation has the capability to search a vastly larger space than that accessible to either *in vivo* maturation or experimental selection techniques, potentially discovering larger and more beneficial evolutionary steps.

Our results present several design lessons. We find that computed electrostatics alone is a better predictor for improved binding than is computed total free energy. Electrostatics-based predictions yielded fewer false positives, more true positives and a >60% success rate for single mutations from wild type (**Fig. 3**). Predictions based on improved total free energy were dominated by mutations to larger amino acids that did not experimentally improve affinity; incorporation of improved nonpolar hydration models may improve accuracy of predicted packing changes[28–30]. We find that designing single and double mutations allows for in-depth conformational search and avoids having any particular design flaw spoiling all results. In addition, avoiding destabilizing mutations based on calculated folding stability was important. Crystal structure resolution did not have a substantial effect, as the D44.1 and cetuximab structures are of 2.5- and 2.8-Å resolution, respectively, whereas the D1.3 structure is of 1.8-Å resolution. Calculations in D1.3 showed few opportunities for electrostatics-based improvement, and accumulated evidence indicates that D1.3 is the anomaly, possibly because of the combination of the many large side chains and buried water molecules at the antibody-antigen interface.

Our results demonstrate capabilities for improving protein binding affinity using computational design. Maturation of the model system D44.1 by 140-fold to 30 pM, maturation of the therapeutic antibody cetuximab by tenfold to 52 pM and identification of known mutations in 4-4-20 and bevacizumab together indicate that our method should be effective for other antibodies and protein interactions. Computational design holds the promise of far greater exploration of sequence space than is possible experimentally, enabling rapid and inexpensive protein improvement.

## METHODS

**Computational design.** The design approach used a two-stage hierarchical procedure. First, conformational search was simplified by assuming a rigid protein backbone and allowing only discrete side-chain rotamers. The physics-based energy function is pairwise-decomposable, permitting application of dead-end elimination and A* search algorithms. For each protein sequence, we found its global minimum energy conformation (GMEC), and, if this energy was within a cut-off of the wild-type GMEC energy, then a continued list of lowest-energy structures was found. Second, we reevaluated the lowest-energy structures of each sequence using more accurate, yet more computationally demanding models, including Poisson-Boltzmann continuum electrostatics, continuum solvent van der Waals, unbound-state side-chain conformation search and minimization. Structures were ranked again based on these latter calculations. Binding energy was initially predicted from the bound-state conformation and a rigid binding model. The unbound state search, when applied, approximated flexible binding and estimated a deformation penalty that offset binding. Changes to protein-fold stability were approximated from the energetic difference between the folded state and isolated model compounds. Detailed computational methods are provided in **Supplementary Methods** online.

**Experiments.** The single-chain format of antibodies were displayed on the surface of yeast (**Supplementary Table 3** online). Binding affinities were measured by incubating different vessels of antibody-displaying cells with varying antigen concentration. Secondary reagents were used to detect antibody-antigen complexes and analyzed with flow cytometry. Detailed experimental methods are provided in **Supplementary Methods** online.

*Note: Supplementary information is available on the Nature Biotechnology website.*

ACKNOWLEDGMENTS

We thank S.L. Sazinsky for the gift of the 404SG material, and D. Lipovsek and R.T. Sauer for comments on the manuscript. This work was supported by a National Science Foundation Graduate Fellowship to S.M.L. and grants from the National Institutes of Health (CA96504 and GM65418).

AUTHOR CONTRIBUTIONS

B.T. oversaw all computational aspects of the work, and K.D.W. oversaw all experimental aspects of the work. S.M.L. developed and adopted the design methods and software and carried out all computational and experimental studies. The authors as a group interpreted the results of the calculations and selected the mutants to create experimentally. S.M.L. drafted the manuscript, and all authors contributed to its editing.

Published online at http://www.nature.com/naturebiotechnology
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Foote, J. & Eisen, H.N. Kinetic and affinity limits on antibodies produced during immune responses. *Proc. Natl. Acad. Sci. USA* **92**, 1254–1256 (1995).
2. Desjarlais, J.R. & Handel, T.M. De novo design of the hydrophobic cores of proteins. *Protein Sci.* **4**, 2006–2018 (1995).
3. Dahlyat, B.I. & Mayo, S.L. De novo protein design: fully automated sequence selection. *Science* **278**, 82–87 (1997).
4. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
5. Reina, J. *et al.* Computer-aided design of a PDZ domain to recognize new target sequences. *Nat. Struct. Biol.* **9**, 621–627 (2002).
6. Looger, L.L., Dwyer, M.A., Smith, J.J. & Hellinga, H.W. Computational design of receptor and sensor proteins with novel functions. *Nature* **423**, 185–190 (2003).
7. Cochran, F.V. *et al.* Computational de novo design and characterization of a four-helix bundle protein that selectively binds a nonbiological cofactor. *J. Am. Chem. Soc.* **127**, 1346–1347 (2005).
8. Havranek, J.J. & Harbury, P.B. Automated design of specificity in molecular recognition. *Nat. Struct. Biol.* **10**, 45–52 (2003).
9. Bolon, D.N., Grant, R.A., Baker, T.A. & Sauer, R.T. Specificity versus stability in computational protein design. *Proc. Natl. Acad. Sci. USA* **102**, 12724–12729 (2005).



275

10. Ashworth, J. *et al.* Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441**, 656–659 (2006).
11. Joachimiak, L.A., Kortemme, T., Stoddard, B.L. & Baker, D. Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J. Mol. Biol.* **361**, 195–208 (2006).
12. Ambroggio, X.I. & Kuhlman, B. Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J. Am. Chem. Soc.* **128**, 1154–1161 (2006).
13. Dwyer, M.A., Looger, L.L. & Hellinga, H.W. Computational design of a biologically active enzyme. *Science* **304**, 1967–1971 (2004).
14. Lassila, J.K., Keeffe, J.R., Oelschlaeger, P. & Mayo, S.L. Computationally designed variants of *Escherichia coli* chorismate mutase show altered catalytic activity. *Protein Eng. Des. Sel.* **18**, 161–163 (2005).
15. Korkegian, A., Black, M.E., Baker, D. & Stoddard, B.L. Computational thermostabilization of an enzyme. *Science* **308**, 857–860 (2005).
16. Dixon, R.W. *et al.* Theoretical and experimental studies of biotin analogues that bind almost as tightly to streptavidin as biotin. *J. Org. Chem.* **67**, 1827–1837 (2002).
17. Song, G. *et al.* Rational design of intercellular adhesion molecule-1 (ICAM-1) variants for antagonizing integrin lymphocyte function-associated antigen-1-dependent adhesion. *J. Biol. Chem.* **281**, 5042–5049 (2006).
18. Sood, V.D. & Baker, D. Recapitulation and design of protein binding peptide structures and sequences. *J. Mol. Biol.* **357**, 917–927 (2006).
19. Clark, L.A. *et al.* Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design. *Protein Sci.* **15**, 949–960 (2006).
20. Sato, J.D. *et al.* Biological effects in vitro of monoclonal antibodies to human epidermal growth factor receptors. *Mol. Biol. Med.* **1**, 511–529 (1983).
21. Boder, E.T., Midelfort, K.S. & Wittrup, K.D. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc. Natl. Acad. Sci. USA* **97**, 10701–10705 (2000).
22. Midelfort, K.S. & Wittrup, K.D. Context-dependent mutations predominate in an engineered high-affinity single chain antibody fragment. *Protein Sci.* **15**, 324–334 (2006).
23. Presta, L.G. *et al.* Humanization of an anti-vascular endothelial growth factor monoclonal antibody for the therapy of solid tumors and other disorders. *Cancer Res.* **57**, 4593–4599 (1997).
24. Chen, Y. *et al.* Selection and analysis of an optimized anti-VEGF antibody: crystal structure of an affinity-matured Fab in complex with antigen. *J. Mol. Biol.* **293**, 865–881 (1999).
25. Joughin, B.A., Green, D.F. & Tidor, B. Action-at-a-distance interactions enhance protein binding affinity. *Protein Sci.* **14**, 1363–1369 (2005).
26. Selzer, T., Albeck, S. & Schreiber, G. Rational design of faster associating and tighter binding protein complexes. *Nat. Struct. Biol.* **7**, 537–541 (2000).
27. Levy, R.M., Zhang, L.Y., Gallicchio, E. & Felts, A.K. On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. *J. Am. Chem. Soc.* **125**, 9523–9530 (2003).
28. Gallicchio, E., Zhang, L.Y. & Levy, R.M. The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. *J. Comput. Chem.* **23**, 517–529 (2002).
29. Gallicchio, E. & Levy, R.M. AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* **25**, 479–499 (2004).
30. Wagoner, J.A. & Baker, N.A. Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. *Proc. Natl. Acad. Sci. USA* **103**, 8331–8336 (2006).

© 2007 Nature Publishing Group http://www.nature.com/naturebiotechnology

Trans Am Clin Climatol Assoc 1989, 100, 31-38
(Climatological)

# ANTI-EGF RECEPTOR MONOCLONAL ANTIBODIES: BIOLOGICAL STUDIES AND POTENTIAL CLINICAL APPLICATIONS*,**

JOHN MENDELSOHN*

NEW YORK

## INTRODUCTION

Cancer cells appear to utilize three methods to bypass requirements for growth factors: (1) autocrine secretion of growth factors, (2) increased expression of receptors for growth factors, and (3) activation in an autonomous fashion of biochemical pathways normally regulated by growth factors. In the case of the EGF receptor, autocrine stimulation may be mediated by TGF alpha, which binds to the EGF receptor with affinity and activity comparable to EGF (1). Surveys of a large number of human tumor cell lines and primary tumor specimens have demonstrated increased expression of messenger RNA for the EGF receptor and TGF alpha, and the presence of increased amounts of receptor and ligand are commonly observed (2–8). Although these observations provide circumstantial evidence for a potential autocrine pathway in tumor cells and a number of experimental studies involving transfection of EGF receptors or TGF alpha support this hypothesis, the level of autocrine stimulation in primary human tumors remains to be more clearly defined.

Our laboratory is exploring the possibility that monoclonal antibodies (mAbs) against the EGF receptor may be useful for detecting tumors expressing high numbers of EGF receptors, and for preventing proliferation of tumor cells regulated by TGF alpha or EGF. We utilized the A431 human vulvar epidermoid carcinoma cell line as a model system (Figure 1) because these cells express very high levels of EGF receptors, in the range of $2 \times 10^6$ per cell (9). Subsequently, it has been shown that A431 cells express messenger RNA for TGF alpha (10).

## EXPERIMENTAL RESULTS

We produced a panel of mAbs against the human EGF receptor. Two of these, 225 IgG1 and 528 IgG2a, are similar in that they bind to the

* From the Department of Medicine and Laboratory of Receptor Biology, Memorial Sloan-Kettering Cancer Center and Cornell University Medical College, New York, NY 10021

** Supported in part by grants CA37641 and CA-42060 from the National Cancer Institute



277

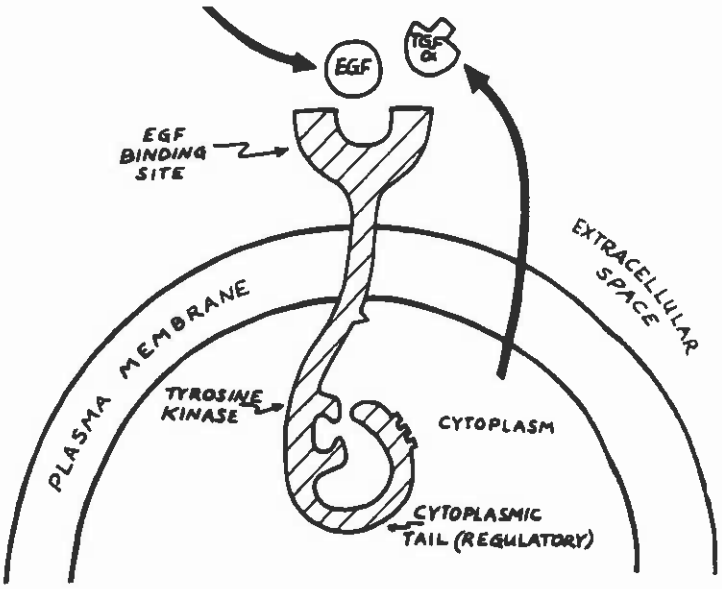



FIG. 1. Diagram of an A431 cell. The EGF receptor (cross hatched) consists of three portions: an extracellular domain with the capacity to bind EGF or TGF alpha with high affinity; a short hydrophobic transmembrane portion; and an intracellular domain which includes a tyrosine-specific protein kinase. There are three tyrosine residues on the cytoplasmic tail which are autophosphorylated when the receptor is activated by ligand, and this further enhances the intrinsic tyrosine kinase activity. The cytoplasmic portion of the receptor is homologous with the *v-erb* B oncogene product.

EGF receptor with affinity comparable to the natural ligand ($K_D$ = 3 nM), compete with EGF binding, precipitate the receptor, and block EGF-induced tyrosine kinase activation (11–14). Figure 2 shows the results of an experiment with intact A431 cells which demonstrates the inhibition of EGF-induced autophosphorylation of receptor tyrosine residues by 225 IgG1, present in 15-fold excess over the concentration of EGF (15, 16). The left panel demonstrates that exposure to EGF for 15 minutes stimulates receptor tyrosine kinase, while 225 IgG1 has no direct activity on the receptor. The right panel shows data demonstrating that 225 IgG1 inhibits EGF-induced phosphorylation of receptor tyrosine residues. The mechanism by which threonine and serine phosphorylation remains stimulated in the presence of excess antibody is under exploration.

The mAbs do not have intrinsic inhibitory activity against normal human cells in culture. However, they can block EGF-induced stimulation of growth rate (12). For cultured tumor cells bearing high numbers of EGF receptors, the presence of nanomolar quantities of EGF in the

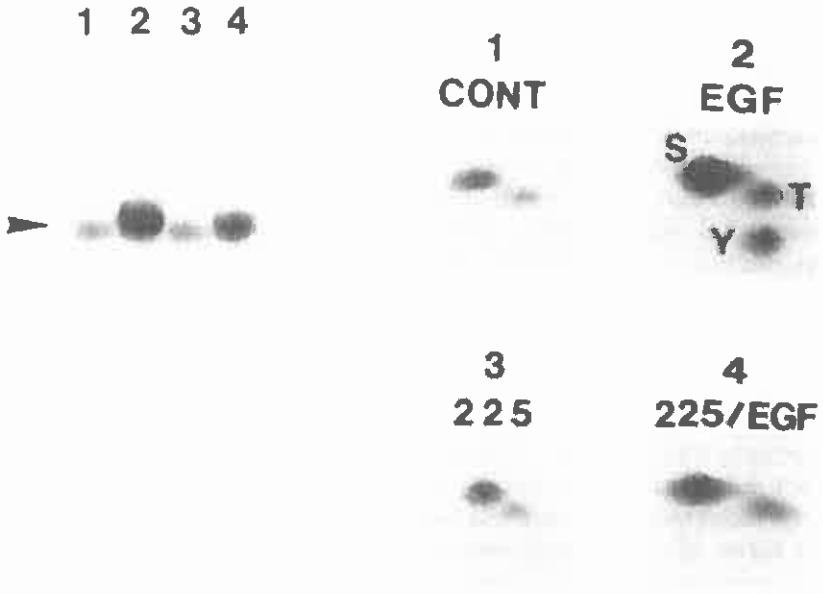



FIG. 2. Modulation of the EGF receptor phosphorylation by mAb 225. Near confluent A431 cells prelabeled with $^{32}$P-orthophosphate at 37°C for 15 h, were incubated with 1) no addition; 2) 20 nM EGF (37°C, 30 min); 3) 300 nM mAb 225 (0°C, 30 min followed by 37°C, 30 min); and 4) 300 nM mAb (0°C, 30 min), then together with 20 nM EGF (37°C, 30 min). In order to maintain consistency in the conditions, every culture was placed on ice for 30 min either in the presence or absence of mAb 225, prior to the addition of EGF. *Left*, autoradiography of $^{32}$P-EGF receptor (arrow head) immunoprecipitated with 528 IgG2a and separated in a 7% polyacrylamide gel in the presence of SDS. *Right*, two-dimensional thin layer electrophoresis patterns of $^{32}$P-phosphoamino acids associated with the EGF receptor. S, phosphoserine; T, phosphothreonine; Y, phosphotyrosine. With longer autoradiography exposure time, baseline tyrosine phosphorylation in control cells becomes evident.

culture medium inhibits proliferation (17, 18), although 10–100 picomolar EGF stimulates growth (11). Monoclonal antibodies 225 IgG1 and 528 IgG2a, at a concentration of 2–20 nM, also reduce the rate of proliferation of A431 human epidermoid carcinoma cells, cultured in the absence of exogenous EGE (11, 12). A comparable direct inhibitory effect has been observed against some, but not all, epidermal tumor cell lines tested and against human mammary adenocarcinoma cell line MDA-468 (19, 20). The cell lines inhibited by antibody are those that are also inhibited by 5–20 nM EGF, and they generally express increased levels of EGF receptors.

*In vivo* effects of treatment with anti-EGF receptor mAbs were assayed against xenografts of human A431 tumors in athymic mice. A treatment schedule of intraperitoneal injections twice weekly was selected, based on a measured mAb half-life in serum of 3 days. Administration of either competitive mAb intraperitoneally, concurrent with tumor cell implantation subcutaneously, resulted in a dose-dependent inhibition of tumor growth (21). Injection of 2 mg twice weekly for 3 weeks resulted in total suppression of tumor growth, which persisted for 3 months after completion of a 3-week course of therapy. Comparable inhibition of xenograft tumor growth was observed with two additional squamous cell carcinomas, which were also inhibited in culture, but was not observed against five other tumor cell lines, which were not inhibited in culture (21, 22, 19). MDA-468 breast adenocarcinoma xenografts were also inhibited by treatment with anti-EGF receptor mAb (20). Evidence that the antitumor effect is mediated by a physiologic effect upon the receptor, rather than by an immune mechanism, has been presented (23).

A series of experiments was carried out to determine whether anti-EGF receptor mAbs could selectively image A431 tumors bearing elevated levels of EGF receptors (20). Tumor cells ($10^7$) were injected into athymic mice subcutaneously, in the flank, and anti-EGF receptor mAb 225, conjugated with diethylenetriaminepentaacetic acid (DTPA) to $^{111}$In, was administered intraperitoneally when tumor size reached 50–100 mg. As a control for these studies, $^{111}$In was conjugated to mAb KS1/4S-1, which is an IgG1 that does not bind to human cells. Antibody concentrations ranged from 100–200 mcg, with radioactivity in the range of 150 $\mu$Ci. Animals were imaged with a gamma camera on successive days. Mice injected with labeled 225 showed excellent localization of the subcutaneous A431 xenografts, while animals injected with labeled control mAb KS1/4-S1 showed poor tumor images. After 3 days, animals were sacrificed in groups of 6, and samples of various tissues were removed, washed in buffered saline, weighed, and counted in a gamma counter. The tumor uptake of 225 represented 28% I.D./gm, while the tumor uptake of KS1/4-S1 was 7.3% I.D./gm. There was greater uptake of labeled 225 in tumor tissue than in all normal tissues tested (range of Tissue Distribution Ratio = 4.5–41.0). In contrast, there was less localization of labeled control mAb in tumor compared with normal tissues. The results demonstrate that there was a far higher concentration of radioactivity in tumor tissue, when mAb specific for the EGF receptor was administered to the tumor-bearing animals.

## CLINICAL TRIAL

Monoclonal antibody 225 IgG1 was the first agent selected for our clinical trials, because it would be less likely to activate complement and

inflammatory mechanisms than the 528 IgG2a. By using 225 IgG1, we were able to design a more pure test of the hypothesis that an antibody can exert antitumor effects *in vivo* by directly affecting a physiologic function of the antigen, in this case, the EGF receptor.

A Phase I study of $^{111}$In-labeled 225 IgG1 was initiated in a collaboration between the Departments of Medicine and Nuclear Medicine at Memorial Sloan-Kettering Cancer Center and Hybritech, Inc. (24). The goals of this trial are two-fold: 1) to define the toxicity and pharmacokinetics of $^{111}$In-labeled mAb 225 IgG1 in patients with squamous cell carcinoma of the lung, and 2) to determine if $^{111}$In-labeled 225 IgG1 localizes to sites of epidermoid lung cancer in patients with these tumors. Squamous carcinoma of the lung was selected because all reports in the literature (cited above) have demonstrated increased EGF receptors in these tumors.

Eligible patients had Stage III or IV epidermoid carcinoma of the lung, with a performance status >60% (Karnofsky), and they were first offered treatment with courses of radiation or chemotherapy before being offered the opportunity to participate in the mAb trial. Patients in five groups of 3 have received one-hour infusions of mAb. The first group received 1 mg of unlabeled mAb. The subsequent groups received 4 mg mAb labeled with 5 mCi $^{111}$In, with groups 3, 4 and 5 additionally receiving 16, 36 and 116 mg unlabeled mAb, respectively. The clinical trial is not completed, but certain preliminary results can be reported. First, and most importantly, we have not observed any toxicity in patients treated with up to 120 mg of anti-EGF receptor mAb. Second, imaging studies on patients who received more than 20 mg of mAb revealed lung tumor visualization optimally, on days 3 and 5 after injection. As expected from previous studies utilizing indium-labeled mAbs, there was extensive uptake in the liver, as well as the large intestine. In addition, hepatocytes are known to bear EGF receptors. Finally, the serum half-life of $^{111}$In-225 IgG1 increased from less than 5 minutes to 21 hours, when the administered dose was escalated from 4 to 120 mg. The results of detailed pharmacokinetic studies and evaluation for human anti-mouse antibodies are pending, and further dose escalation is ongoing.

These preliminary observations suggest that 225 IgG1 may be useful for imaging tumors which bear increased numbers of EGF receptors, and they raise the possibility that therapy with this mAb, or with immunoconjugates of the antibody, may be worthwhile areas for clinical investigation in the future.

## SUMMARY

We have produced two monoclonal antibodies (mAbs), 528 IgG2a and 225 IgG1, which bind to the human EGF receptor with an affinity

comparable to EGF, compete with EGF for binding to the receptor, and prevent EGF-induced activation of receptor tyrosine kinase. The mAbs inhibit the proliferation of a number of human tumor cells bearing high numbers of EGF receptors in culture and in nude mouse xenographs. A431 xenographs bearing high numbers of EGF receptors can be imaged with $^{111}$In-labeled mAb. Phase I trials were initiated in patients with advanced squamous cell carcinoma of the lung, which consistently expresses high numbers of EGF receptors. $^{111}$In-labeled 225 IgG1 was given by single intravenous infusion. Preliminary results show no toxicity from doses up to 120 mg. Tumor visualization was seen to be dose-related. However, as with other imaging studies with indium-labeled antibodies, significant liver and bowel isotope uptake was observed. Further dose escalation is ongoing to assess the relationship between dose and tumor uptake.

## ACKNOWLEDGEMENTS

These studies have been carried out in collaboration with investigators at the University of California, San Diego, Memorial Sloan-Kettering Cancer Center and Hybritech, Inc. They include (chronologically) Hideo Masui, Gordon Sato, Tomoyuki Kawamoto, Denry Sato, Jan Castagnola, Gordon Gill, Carol MacLeod, Hironabu Sunada, Alec Goldenberg, Peter Yu, Bruce Merchant, Michael Unger, Sally Schweighardt, Sydney Welt, Francisco Real, Chaitanya Divgi, Samuel Yeh, Richard Gralla, Mark Kris, and Steven Larson.

## REFERENCES

1. Derynck R, Roberts AB, Winkler ME, et al. Human transforming growth factor-alpha: precursor structure and expression in E. coli. *Cell* 1984; 38: 287.
2. Ozanne BW, Richards CS, Hendler FJ, et al. Over-expression of the EGF receptor is a hallmark of squamous cell carcinoma. *J Path* 1986; 149: 9.
3. Hendler FJ, Ozanne BW. Human squamous cell lung cancers express increased epidermal growth factor receptors. *J Clin Invest* 1984; 74: 647.
4. Hwang D, Tay YC, Lin SS, Lev-Ran A. Expression of epidermal growth factor receptors in human lung tumors. *Cancer* 1986; 58: 2260.
5. Gullick WJ, Marsden JJ, Whittle N, et al. Expression of epidermal growth factor receptors on human cervical, ovarian, and vulvar carcinomas. *Cancer Res* 1986; 46: 285.
6. Cowley GP, Smith JA, Gusterson BA. Increased EGF receptors on human squamous carcinoma cell lines. *Br J Cancer* 1986; 53: 223.
7. Libermann TA, Razon N, Bartal AD, et al. Expression of EGF receptors in human brain tumors. *Cancer Res* 1984; 44: 753.
8. Filmus J, Pollack MN, Cailleau R, et al. MDA-468, a human breast cancer cell line with a high number of EGF receptors, has an amplified EGF receptor gene and is growth inhibited by EGF. *Biochem Biophys Res Com* 1985; 128: 898.
9. Fabricant RN, DeLarco JE, Todaro GJ. Nerve growth factor receptors on human melanoma cells in culture. *Proc Natl Acad Sci USA* 1977; 74: 565.

10. Derynck R, Goeddel DV, Ulrich A, et al. Synthesis of messenger RNAs for transforming growth factors alpha and beta and the epidermal growth factor receptor by human tumors. *Cancer Res* 1987; 47: 707.
11. Kawamoto T, Sato JD, Le A, et al. Growth stimulation of A431 cells by EGF: Identification of high affinity receptors for epidermal growth factor by anti-receptor monoclonal antibody. *Proc Natl Acad Sci USA* 1983; 80: 1337.
12. Sato JD, Kawamoto T, Le AD, et al. Biological effects in vitro of monoclonal antibodies to human EGF receptor. *Mol Biol Med* 1983; 1: 511.
13. Kawamoto T, Mendelsohn J, Le A, et al. Relation of epidermal growth factor receptor concentration to growth of human epidermoid carcinoma A431 cells. *J Biol Chem* 1984; 259: 7761.
14. Gill GN, Kawamoto T, Cochet C, et al. Monoclonal anti-epidermal growth factor receptor antibodies which are inhibitors of epidermal growth factor binding and antagonists of epidermal growth factor-stimulated tyrosine protein kinase activity. *J Biol Chem.* 1984; 259: 7755.
15. Sunada H, Magun B, Mendelsohn J, MacLeod CL. Monoclonal antibody against EGF receptor is internalized without stimulating receptor phosphorylation. *Proc Natl Acad Sci USA* 1986; 83: 3825.
16. Sunada H, Yu P, Peacock JS, Mendelsohn J. Modulation of tyrosine, serine and threonine phosphorylation and intracellular processing of the epidermal growth factor receptor by anti-receptor monoclonal antibody. *Submitted* 1988.
17. Gill GN, Lazar CS. Increased phosphotyrosine content and inhibitions of proliferations in EGF-treated A431 cells. *Nature* 1981; 293: 305.
18. Barnes DW. Epidermal growth factor inhibits growth of A431 epidermoid carcinoma in serum-free cell culture. *J Cell Biol* 1982; 93: 1.
19. Mendelsohn J, Masui H, Sunada H, MacLeod CL. Monoclonal antibodies against the receptor for EGF as potential anticancer agents. In: Minna J, Kuehl NM, eds. *Cellular and Biology of Tumor and Potential Clinical Applications.* New York: Alan R Liss; 1988: 307.
20. Goldenberg A, Masui H, Kamrath H, et al. EGF receptor overexpression and localization of nude mouse xenografts using a $^{111}$In-labeled anti-EGF recpetor monoclonal antibody. Submitted for publication.
21. Masui H, Kawamoto T, Sato JD, et al. Growth inhibition of human tumor cells in athymic mice by anti-EGF receptor monoclonal antibodies. *Cancer Res* 1984; 44: 1002.
22. Mendelsohn J, Masui H, Goldenberg A. Anti-epidermal growth factor receptor monoclonal antibodies may inhibit A431 tumor cell proliferation by blocking on autocrine pathway. *Trans Assoc Amer Phys* 1987; 100: 173.
23. Masui H, Morayama T, Mendelsohn J. Mechanism of anti-tumor activity in mice for anti-EGF receptor monoclonal antibodies with different isotypes. *Cancer Res* 1986; 46: 5592.
24. Divgi C, Welt S, Real F, et al. Unpublished observations.

## DISCUSSION

**Middleton** (Buffalo): Could you tell us a bit about the normal regulation mechanisms of EGF receptors? What is known about the expression of those receptors?

**Mendelsohn:** The receptor is expressed normally in numbers of ten to thirty thousand per cell in contrast to some tumor cells which have up to two million. When the receptor is activated by the binding of EGF, both are internalized and both traffic to the lysomal compartment and are destroyed or catabolized within a matter of two to four hours. Our antibody follows a similar pathway so that it mimics the normal ligand. The cell generates

an incredible amount of energy and protein because it has to continually repopulate the surface with receptor. There must be a reason why the cell wants to do this. It seems to be an important molecule for many of these tumor cells.

**Middleton** (Buffalo): Was the antibody of heterologous or human origin?

**Mendelsohn:** The antibody is a murine monoclonal antibody. Therefore if these experiments do show a potential for clinical efficacy, which is still very much a dream rather than a reality, it will be important to convert these antibodies to human epitopes so that an immune reaction won't occur in the patients.

**Field** (Houston): I'd like to congratulate you, John, on this beautiful presentation and the results. I have two somewhat unrelated questions. The first: Among your monoclonal antibodies did you observe any that reproduced the effects of EGF rather than block them on the receptor? Studies of antibodies in patients with Graves' Disease have demonstrated both stimulating and blocking actions.

**Mendelsohn:** We haven't. One IgM antibody was produced by Dr. Joseph Schlessinger. Unfortunately that clone was lost but that antibody was able to mimic the stimulation of EGF. This was the first formal proof that the activation of the EGF receptor was not dependent upon the ligand but was totally within the capacity of the receptor molecule itself.

**Field** (Houston): The second question: In your experiments in nude mice you used the antibody which blocked totally the growth of the tumor. One might then hypothesize that the antibody itself was responsible for the destruction of the tumor and that one wouldn't need to couple poisons or radioactive molecules to the antibody to destroy the tumor. Do you have any preliminary evidence, that administration of antibody to patients with squamous cell carcinoma had any effect on the size or growth of the tumor?

**Mendelsohn:** The answer is no, we have no indications of this as yet. We only know that the drug was not toxic in this phase one trial. A couple of points: First of all the nude mouse experiment involved repeated injections; secondly, as you noticed the antibody administration was begun on the day of tumor innoculation. We have studies in which we let the tumors grow to substantial size, say after a couple of weeks when the tumors were 1-2 cubic centimeters in size, then initiate therapy. The antibody in this situation is cytostatic. The tumor growth curve flattens and when we stop therapy with the antibody, the tumor takes off again. We don't know the reason for this but if you remember in the *in vitro* studies in culture the antibody alone is cytostatic. The cell number increased but the growth curve was attenuated. Thus we believe that as with any other form of antitumor therapy, one agent is not going to achieve our total goal and we are thinking about experiments combining chemotherapy with the antibodies. The first step, of course, is to prove that it isn't toxic in patients and so far it hasn't been.

**Crofford** (Nashville): Will you tell us what kind of toxic effects you have observed in experimental animals?

**Mendelsohn:** We have to date observed no toxic effects in the animals and that raises another interesting point which we didn't have time to go into. The antibody is specific for the human EGF receptor binding site. It is not cross-reacting with the murine receptor so that it is an unfair test of toxicity to ask what happened to the animals. However, after waiting for five years for a commercial organization to synthesize anti-mouse EGF receptor antibodies we gave up and made our own. We are in the process now of studying the very point you raised. So far I can tell you we have raised ascites with rat anti-mouse EGF receptor antibodies in mice and we know that the antibody gets into the circulation. The mice have done fine. We're going to study that more formally. It looks like an antibody which can block the EGF receptor is not going to produce major toxicity, which surprised us. I hope we will be able to give you an update on this as we learn more about it.

**Ultimate goal of therapeutic antibodies**

Most of the therapeutic antibodies are used for the treatment of cancers, autoimmune diseases, transplantation rejections, etc. Therefore, prolonged use will be essential. Unfortunately, as discussed before, there are two important factors related to this: (1) half-life in human patients, and (2) possible length of therapy. Both are related to the antigenicity of these therapeutic antibodies in humans. In general, humanized mouse and fully human therapeutic antibodies have longer half-lives. They can also be used for longer periods of treatment. In most cases, very few of these therapeutic antibodies can be used indefinitely.

As a result, in clinical treatment of cancer for example, therapeutic antibodies are currently used in conjunction with chemotherapy and radiation therapy. However, such combinations may involve contradictory ideas involving human patients. Therapeutic antibodies require a robust human immune system. On the other hand, both chemotherapy and radiation therapy try to selectively kill cancer cells. They also provide an adverse environment to rapidly dividing cells. One of such normal systems in humans is the immune system. As a result, such combined therapy may somewhat improve the survival rate of human cancer patients slightly, but with absolutely no cure. This is illustrated by the recently FDA approved therapeutic antibody ipilimumab (Sondak VK, Smalley KSM, Kudchadkar R, Frippon S, Kirkpatrick P (2011) Ipilimumab. Nature Review, 10, 411-412). In a clinical trial of metastatic melanoma using chemotherapy with or without ipilimumab, the survival rates were as follows (Robert C et al. (2011) Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. N Eng J Med, 364, 2517-2526):

| Treatment | 2 -year survival | 3-year survival |
|---|---|---|
| Ipilimumab + dacarbazine | 28.5 % | 20.8 % |
| Placebo + dacarbazine | 17.9 % | 12.2 % |

Therefore, in principle, the ultimate goal of therapeutic antibodies should aim at the complete cure of cancer or other diseases mentioned above. The idea of immune surveillance in the prevention of cancer basically suggests that our immune system is capable of preventing the development of cancers. Somehow, the ideal therapeutic antibodies must be able to activate such surveillance.

Some of the humanized mouse therapeutic antibodies are thus fine-tuned with the modification of their surface amino acid residues to reduce immunogenicity in human patients. Some of their CDR's can be made more similar to human CDR's. In the case of fully human therapeutic antibodies, they should be non-immunogenic in humans. However, if a few of these molecules are not folded properly, they may also become foreign antigens over very long periods of usage.

FRESH FROM THE PIPELINE

# Ipilimumab

*Vernon K. Sondak, Keiran S. M. Smalley, Ragini Kudchadkar, Seden Grippon and Peter Kirkpatrick*

In March 2011, ipilimumab (Yervoy; Bristol-Myers Squibb), a human monoclonal antibody specific for human cytotoxic T lymphocyte-associated antigen 4 (CTLA4), was approved by the US Food and Drug Administration (FDA) for the treatment of unresectable or metastatic melanoma.

Metastatic melanoma currently has a 10-year survival rate of less than 10%, and is estimated to be responsible for more than 8,000 deaths annually in the United States alone[1]. Outside a clinical trial setting, the standard treatment for metastatic melanoma is cytotoxic chemotherapy, typically with the alkylating agent dacarbazine, which was approved by the FDA more than three decades ago. Other chemotherapies have also been used, particularly the orally administered dacarbazine analogue temozolomide, platinum-based agents such as carboplatin and nitrosoureas such as fotemustine. However, like dacarbazine, these drugs only have modest antitumour efficacy, with response rates typically ranging from ~5–20% and no documented improvement in overall survival[2].

Treatment with high doses of interleukin-2 (IL-2), a cytokine that stimulates T cell proliferation and function, can result in durable complete responses in a small proportion of patients with metastatic melanoma[2], and a recombinant IL-2 therapy, aldesleukin (Proleukin; Prometheus), was approved by the FDA in 1998. Various approaches involving combinations of chemotherapy and biologics have also been investigated in the past decade, but although it seems that response rates can be improved compared to single-agent chemotherapy, the combinations are associated with greater toxicity but no definitive survival benefit[2]. Furthermore, there has been no accepted second-line therapy for patients with metastatic melanoma.

## Basis of discovery

Tumours express tumour-associated antigens that could theoretically provide the basis for an effective immune response. Co-stimulatory signals are needed to promote effective T cell activation following presentation of antigens that are bound to major histocompatibility complex (MHC) molecules on the surface of antigen-presenting cells (APCs)[3,4] (FIG. 1). Most melanomas, however, are poorly immunogenic[3], in part owing to insufficient T cell activation resulting from lack of co-stimulation, but also potentially owing to immunosuppression induced by the tumour, possibly through regulatory T ($T_{Reg}$) cells.

A key co-stimulatory signal is the binding of CD80 or CD86 on APCs to CD28 on T cells[3]. CTLA4 is related to CD28 but has greater binding affinity for CD80 and CD86. Binding of CTLA4 to CD80 or CD86 prevents co-stimulation and thus acts as a negative regulator of T cell activation. CTLA4 is also involved in the functioning of $T_{Reg}$ cells, in which it is likely to have an important role in limiting the immune response to self antigens. Based on these discoveries, it was proposed that blockade of CTLA4 interactions could enhance T cell-mediated antitumour responses[3]. Initial experiments in mice showed that *in vivo* administration of CTLA4-specific antibodies resulted in the rejection of established tumours[4]. This led to the development of CTLA4-specific human monoclonal antibodies as potential cancer immunotherapeutics[5].

## Drug properties

Ipilimumab is a recombinant human immunoglobulin G1κ monoclonal antibody that binds to CTLA4 and blocks the interaction of CTLA4 with CD80 or CD86 (REF. 6). Blockade of CTLA4 has been shown to augment T cell activation and proliferation[2,3].

## Clinical data

The safety and efficacy of ipilimumab (administered at 3 mg per kg via intravenous infusion every 3 weeks for four doses) were assessed in a randomized, double-blind trial involving 676 patients with unresectable or metastatic melanoma whose disease had progressed after treatment with one or more 'standard therapies': dacarbazine, aldesleukin, temozolomide, fotemustine or carboplatin[6,7]. Patients were assigned in a 3:1:1 ratio to receive ipilimumab in combination with an investigational gp100 peptide vaccine (403 patients), ipilimumab alone (137 patients) or gp100 alone (136 patients)[6,7]. The trial only enrolled HLA-A*0201-positive patients, as the gp100 peptide vaccine is presented to the immune system in the context of this specific MHC molecule[6,7]. The primary end point was overall survival in the ipilimumab plus gp100 arm compared to the gp100 arm[6,7]. Secondary end points included overall survival in the ipilimumab plus gp100 arm compared to the ipilimumab arm, overall survival in the ipilimumab arm compared to the gp100 arm, and objective response rates and duration of response in each of the study arms[6]. Assessment of tumour response was conducted at 12 and 24 weeks, and every 3 months thereafter[7]. Patients with evidence of objective tumour response at 12 or 24 weeks were assessed for confirmation of durability of response at 16 or 28 weeks, respectively[6,7].

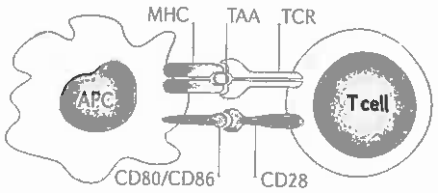



**Figure 1 | T cell activation.** As well as the presentation of tumour-associated antigens (TAAs) bound to major histocompatibility complex (MHC) molecules on the surface of antigen-presenting cells (APCs) to T cell receptors (TCR), a second co-stimulatory signal involving interaction between CD80 or CD86 on APCs and CD28 on T cells is needed to complete T cell activation and expansion. Adapted from REF. 3.

Ipilimumab, with or without the gp100 vaccine, improved overall survival; median overall survival was 10.0 months in the ipilimumab plus gp100 arm, 10.1 months in the ipilimumab arm and 6.4 months in the gp100 arm[6,7]. The investigator-assessed objective response rate was 5.7% in the ipilimumab plus gp100 arm, 10.9% in the ipilimumab arm and 1.5% in the gp100 arm[6]. The median duration of response was 11.5 months in the ipilimumab plus gp100 arm and has not been reached in the ipilimumab alone or gp100 alone arms[6].

## Indications

Ipilimumab is approved by the FDA for the treatment of unresectable or metastatic melanoma[6], with a risk evaluation and mitigation strategy to ensure that treatment benefits outweigh risks.



© 2011 Macmillan Publishers Limited. All rights reserved

Analysing issues in the treatment of metastatic melanoma are Vernon K. Sondak, M.D., Chief, Department of Cutaneous Oncology, and Keiran S. M. Smalley, Ph.D., Assistant Professor, H. Lee Moffitt Cancer Center and Research Institute, Florida, USA.

Metastatic melanoma has long been associated with a dismal prognosis and a low level of response to standard treatments. However, with the recent FDA approval of ipilimumab — based on a Phase III trial in patients with refractory melanoma, which showed that ipilimumab improved survival compared to a vaccine[7] — the treatment landscape is beginning to change substantially. We are eagerly awaiting the results of another Phase III trial of ipilimumab that was conducted in previously untreated patients with metastatic melanoma and compared dacarbazine chemotherapy alone to dacarbazine and ipilimumab (at a higher dose than used in the previous study). A preliminary report in a press release issued by Bristol-Myers Squibb on 21 March 2011 noted that the primary end point of improving overall survival was met in the ipilimumab arm. So, clinicians are likely to be faced with questions about what dose of ipilimumab to use and whether to give ipilimumab alone or with dacarbazine to patients who have not previously received chemotherapy.

Impressive results of Phase II and Phase III trials of inhibitors of the product of the mutant *BRAF* gene — which is found in 40–50% of melanomas — will also raise questions about when and how to use ipilimumab. The BRAF inhibitor vemurafenib has shown dramatic responses in a high percentage of *BRAF* mutant melanoma cases, although it appears that resistance is a major factor limiting its long-term use[8]. A press release issued by Genentech on 18 January 2011 reported that a recently completed Phase III trial showed a survival benefit for vemurafenib compared to dacarbazine, thus suggesting that FDA approval of this agent may also be forthcoming.

Attempts are now underway to explore the efficacy of ipilimumab in combination with vemurafenib or other BRAF inhibitors. This combination approach makes sense clinically, given the different kinetics of response and the nonoverlapping toxicities of the two classes of drugs, and is supported by preclinical studies suggesting that BRAF inhibition improves immune recognition of melanoma cells through increased antigen presentation[9]. A therapeutic regimen can be envisioned in which a BRAF inhibitor could be used first to 'debulk' tumours in patients with $BRAF^{V600E}$ mutant melanoma, followed by an immunomodulatory therapy to 'mop up' any remaining cancer cells. As these combinations are developed further, the choice of the targeted inhibitor to partner with immunotherapy is likely to be critical, because different inhibitors may have subtly different 'off-target' immunosuppressive effects.

Options for patients with wild-type *BRAF* melanomas are more limited, and ipilimumab — alone or with chemotherapy, vaccines or other immunomodulators — will have the predominant role in the management of these melanomas for the foreseeable future. The relatively low response rates with ipilimumab, coupled with the toxicity profile, make patient selection and particularly the development of biomarkers that are predictive of response a key future issue. Our growing understanding of the negative regulatory checkpoints in the immune system has also resulted in the development of new immunomodulators, such as antibodies that are directed against programmed death receptor 1 and its ligand, which ultimately may augment or even replace ipilimumab. Overall, capitalizing on the unprecedented opportunities to develop more efficacious treatments for patients with metastatic melanoma — and perhaps even to prevent high-risk patients from developing metastasis — could result in metastatic melanoma becoming perceived not as an invariably fatal disease but as one that could be amenable to long-term disease control.

*Vernon K. Sondak, Keiran S. M. Smalley and Ragini Kudchadkar are at the H. Lee Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, Tampa, Florida 33612, USA.*

*Seden Grippon is at IMS Health, 7 Harewood Avenue, London NW1 6JB, UK.*

*Peter Kirkpatrick is at* Nature Reviews Drug Discovery.

*e-mails: vernon.sondak@moffitt.org; SGrippon@uk.imshealth.com; p.kirkpatrick@nature.com*

doi:10.1038/nrd3463

1. Jemal, A. *et al.* Cancer statistics, 2010. *CA Cancer J. Clin.* **60**, 277–300 (2010).
2. Garbe, C. *et al.* Systematic review of medical treatment in melanoma: current status and future prospects. *Oncologist* **16**, 5–24 (2011).
3. Salama, A. K. & Hodi, F. S. Cytotoxic T-lymphocyte-associated antigen-4. *Clin. Cancer Res.* 5 Apr 2011 (doi: clincanres.2232.2011).
4. Leach, D. R. *et al.* Enhancement of antitumor immunity by CTLA-4 blockade. *Science* **271**, 1734–1736 (1996).
5. Phan, G. Q. *et al.* Cancer regression and autoimmunity induced by cytotoxic T lymphocyte-associated antigen 4 blockade in patients with metastatic melanoma. *Proc. Natl Acad. Sci. USA* **100**, 8372–8377 (2003).
6. US FDA. FDA labeling information — Yervoy. *FDA website* [online], http://www.accessdata.fda.gov/drugsatfda_docs/label/2011/125377s000lbl.pdf (2011).
7. Hodi, F. S. *et al.* Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* **363**, 711–723 (2010).
8. Flaherty, K. T. *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.* **363**, 809–819 (2010).
9. Boni, A. *et al.* Selective $BRAF^{V600E}$ inhibition enhances T-cell recognition of melanoma without affecting lymphocyte function. *Cancer Res.* **70**, 5213–5219 (2010).
10. Tong, M. *et al.* Wells Fargo Securities, equity research. (28 March 2011).
11. Holfrod, J. *et al.* Jefferies company note, Bristol-Myers Squibb. (31 March 2011).

**Box 1 | Market for metastatic melanoma**

**Analysing the market for metastatic melanoma is Seden Grippon, IMS Health, London, UK.**

**On 25 March 2011, ipilimumab (Yervoy; Bristol-Myers Squibb) became the first drug in more than a decade to be approved in the United States for unresectable or metastatic melanoma. It is currently awaiting approval in the European Union for the same indication, and it is anticipated that approvals in Australia, Canada and other countries will follow. Given the limitations of the few established treatment options for metastatic melanoma (see main text), there is a strong physician interest in ipilimumab. In the United States, ~68,000 patients are diagnosed with melanoma each year, and an estimated 15% of these are expected to initially receive ipilimumab[10]. A course of ipilimumab treatment in the United States is four doses (at US$30,000 per dose), but Bristol-Myers Squibb's patient assistance programme is expected to reduce net pricing to $80,000 per course.**

**In melanoma patients alone, the drug is predicted to reach worldwide sales of $2 billion in 2015 (REF. 10). Ipilimumab is also in clinical development for renal cell carcinoma, ovarian cancer, non-small cell lung cancer, prostate cancer, pancreatic cancer, breast cancer, superficial bladder cancer and lymphoma. Analysts' peak sales estimates in prostate cancer (launch estimated in 2013) and lung cancer (launch estimated in 2015) are $2 billion each[11]. There are also several agents in Phase III trials for metastatic melanoma, including the BRAF inhibitor vemurafenib (developed by Plexxikon/Roche) and the oncolytic virus OncoVEX (developed by Biovex, which was recently acquired by Amgen). Depending on the efficacy and safety data from these trials, some of these agents could provide competition for ipilimumab, or be used with it.**

**Competing financial interests**
The authors declare no competing financial interests.

287



© 2011 Macmillan Publishers Limited. All rights reserved

ORIGINAL ARTICLE

# Ipilimumab plus Dacarbazine for Previously Untreated Metastatic Melanoma

Caroline Robert, M.D., Ph.D., Luc Thomas, M.D., Ph.D., Igor Bondarenko, M.D., Ph.D., Steven O'Day, M D., Jeffrey Weber M.D., Ph.D., Claus Garbe, M.D., Celeste Lebbe, M.D., Ph.D., Jean-François Baurain, M.D., Ph.D., Alessandro Testori, M.D., Jean-Jacques Grob, M.D., Neville Davidson, M.D., Jon Richards, M.D., Ph.D., Michele Maio, M.D., Ph.D., Axel Hauschild, M.D., Wilson H. Miller, Jr., M.D., Ph.D., Pere Gascon, M.D., Ph.D., Michal Lotem, M.D., Kaan Harmankaya, M.D., Ramy Ibrahim, M.D., Stephen Francis, M.Sc., Tai-Tsang Chen, Ph.D., Rachel Humphrey, M.D., Axel Hoos, M.D., Ph.D., and Jedd D. Wolchok, M.D., Ph.D.

## ABSTRACT

### BACKGROUND

Ipilimumab monotherapy (at a dose of 3 mg per kilogram of body weight), as compared with glycoprotein 100, improved overall survival in a phase 3 study involving patients with previously treated metastatic melanoma. We conducted a phase 3 study of ipilimumab (10 mg per kilogram) plus dacarbazine in patients with previously untreated metastatic melanoma.

### METHODS

We randomly assigned 502 patients with previously untreated metastatic melanoma, in a 1:1 ratio, to ipilimumab (10 mg per kilogram) plus dacarbazine (850 mg per square meter of body-surface area) or dacarbazine (850 mg per square meter) plus placebo, given at weeks 1, 4, 7, and 10, followed by dacarbazine alone every 3 weeks through week 22. Patients with stable disease or an objective response and no dose-limiting toxic effects received ipilimumab or placebo every 12 weeks thereafter as maintenance therapy. The primary end point was overall survival.

### RESULTS

Overall survival was significantly longer in the group receiving ipilimumab plus dacarbazine than in the group receiving dacarbazine plus placebo (11.2 months vs. 9.1 months, with higher survival rates in the ipilimumab–dacarbazine group at 1 year (47.3% vs. 36.3%), 2 years (28.5% vs. 17.9%), and 3 years (20.8% vs. 12.2%) (hazard ratio for death, 0.72; P<0.001). Grade 3 or 4 adverse events occurred in 56.3% of patients treated with ipilimumab plus dacarbazine, as compared with 27.5% treated with dacarbazine and placebo (P<0.001). No drug-related deaths or gastrointestinal perforations occurred in the ipilimumab–dacarbazine group.

### CONCLUSIONS

Ipilimumab (at a dose of 10 mg per kilogram) in combination with dacarbazine, as compared with dacarbazine plus placebo, improved overall survival in patients with previously untreated metastatic melanoma. The types of adverse events were consistent with those seen in prior studies of ipilimumab; however, the rates of elevated liver-function values were higher and the rates of gastrointestinal events were lower than expected on the basis of prior studies. (Funded by Bristol-Myers Squibb; ClinicalTrials.gov number, NCT00324155.)

From the Institute Gustave, Roussy, Villejuif (C.R.); Centre Hospitalier, Lyon Sud, Pierre Bénite (L.T.); Hôpital Saint Louis University, Paris (C.L.); and Hôpital de la Timone, Marseille (J.-J.G.) — all in France; Dnepropetrovsk State Medical Center, Dnepropetrovsk, Ukraine (I.B.); The Angeles Clinic and Research Institute, Santa Monica, CA (S.O.); H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL (J.W.); University Medical Center, Tuebingen, Germany (C.G.); Cliniques Universitaires Saint-Luc, Brussels (J.-F.B.) and Bristol-Myers Squibb, Braine-l'Alleud (S.F.) — both in Belgium; Istituto Europeo di Oncologia, Milan (A.T.) and University Hospital of Siena, Siena (M.M.) — both in Italy; Mid Essex Hospitals National Health Service Trust, Essex, United Kingdom (N.D.); Oncology Specialists, Park Ridge, IL (J.R.); University of Kiel, Kiel, Germany (A.H.); Lady Davis Institute and Segal Cancer Center, Jewish General Hospital, McGill University, Montreal (W.M.); Hospital Clinic, Barcelona, Spain (P.G.); Hadassah Hebrew University Hospital, Sharett Institute of Oncology, Jerusalem, Israel (M.L.); Medical University of Vienna, Vienna (K.H.); Medimmune, Gaithersburg, MD (R.I.); Bristol-Myers Squibb, Wallingford, CT (T.-T.C., A.H.); Bristol-Myers Squibb, Lawrenceville, NJ (R.H.); and Memorial Sloan-Kettering Cancer Center, New York (J.D.W.). Address reprint requests to Dr. Wolchok at Memorial Sloan-Kettering Cancer Center, 1275 York Ave., New York, NY 10021, or at wolchokj@mskcc.org.

This article (10.1056/NEJMoa1104621) was published on June 5, 2011, and updated on June 6, 2011, at NEJM.org.

N Engl J Med 2011;364:2517-26.
Copyright © 2011 Massachusetts Medical Society.



The New England Journal of Medicine
Downloaded from nejm.org at GALTER HEALTH SCIENCES LIBRARY on July 18, 2011. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

THE SURVIVAL RATE FOR PATIENTS WITH metastatic melanoma is low, with an expected 2-year survival rate of 10 to 20%.[1-3] Although dacarbazine has never been shown to improve survival in randomized, controlled studies, it has been the drug most frequently compared with new agents or combination therapies in randomized trials involving patients with melanoma.[4,5] High-dose interleukin-2 is associated with durable, complete responses, with a survival benefit, in a small subgroup of patients with metastatic melanoma.[6,7] Ipilimumab, a fully human, IgG1 monoclonal antibody, blocks cytotoxic T-lymphocyte–associated antigen 4 (CTLA-4), a negative regulator of T cells, and thereby augments T-cell activation and proliferation.[8-12]

Two separate phase 2 studies support the rationale for the current study. The first phase 2 study, involving 217 patients with previously treated metastatic melanoma, showed a dose-dependent response rate, with the highest rate of response seen in the group receiving ipilimumab at a dose of 10 mg per kilogram of body weight (11.1% in the group receiving 10 mg per kilogram vs. 4.2% and 0% in the groups receiving 3 mg per kilogram and 0.3 mg per kilogram, respectively; P=0.002).[13] In the second, small, randomized, phase 2 study, involving 72 patients, treatment with a combination of dacarbazine (250 mg per square meter of body-surface area per day for 5 days every 3 weeks) and ipilimumab (3 mg per kilogram every 4 weeks for 4 doses) was associated with durable objective responses, with no new adverse events; the rationale for such combinations has been reviewed previously.[14-16] We conducted a phase 3 study to determine whether ipilimumab (at a dose of 10 mg per kilogram) plus dacarbazine, as compared with dacarbazine and placebo, improves overall survival in patients with previously untreated metastatic melanoma.

## METHODS

### PATIENTS

Eligible patients were at least 18 years of age and had previously untreated stage III (unresectable) or stage IV melanoma with measurable lesions, an Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1 (with 0 indicating that the patient is fully active and able to carry on all predisease activities without restriction, and 1 indicating that the patient is restricted in physically strenuous activity but is ambulatory and able to carry out work of a light or sedentary nature, such as light housework or office work),[17] and a life expectancy of 16 weeks or more. The baseline serum lactate dehydrogenase level did not affect eligibility. Patients were ineligible if they had received any prior treatment for metastatic disease. Patients who received prior adjuvant therapy were not excluded. Concomitant treatment with immunosuppressive agents or long-term use of systemic glucocorticoids (except for the management of adverse events during the course of the study) was not allowed. Patients were also ineligible if they had evidence of brain metastasis (as confirmed on imaging), primary ocular or mucosal melanoma, or autoimmune disease.

### STUDY OVERSIGHT

The protocol was approved by the appropriate institutional review boards or independent ethics committees. The study was conducted in accordance with the ethical principles originating from the Declaration of Helsinki and with Good Clinical Practice as defined by the International Conference on Harmonization. All participating patients gave written informed consent.

The trial was designed jointly by the senior academic authors and the sponsor, Bristol-Myers Squibb. Data were collected by the sponsor and analyzed in collaboration with the senior academic authors, who vouch for the completeness and accuracy of the data and analyses. An initial draft of the manuscript was prepared jointly by a writing committee that included five academic authors, the sponsor, and a professional medical writer employed by the sponsor. All the authors contributed to subsequent drafts and made the decision to submit the manuscript for publication. All the authors signed a confidentiality disclosure agreement with the sponsor. The protocol, including the statistical analysis plan, is available with the full text of this article at NEJM.org.

### STUDY DESIGN AND TREATMENT

In this multinational, randomized, double-blind, phase 3 study, patients (stratified according to metastasis stage, study site, and ECOG performance status) were randomly assigned, in a 1:1 ratio, to receive either ipilimumab (at a dose of 10 mg per kilogram) plus dacarbazine (850 mg per square meter) or dacarbazine (850 mg per square meter) plus placebo at weeks 1, 4, 7, and 10, followed by dacarbazine alone every 3 weeks through week 22 (induction phase). Treatment was discontinued if



289

The New England Journal of Medicine
Downloaded from nejm.org at GALTER HEALTH SCIENCES LIBRARY on July 18, 2011. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

toxic effects associated with the drug or progressive disease was noted during weeks 12 to 24. At week 24, patients with stable disease or an objective response during the induction phase who did not have a dose-limiting adverse event were eligible to enter a maintenance phase in which they received placebo or ipilimumab every 12 weeks until progression of the disease, development of toxic effects, or the end of the study.

In the initial design of the study, 500 patients were to undergo randomization, and progression-free survival was to be the primary end point. Emerging data from other ipilimumab trials suggested that conventional definitions of disease progression and response incompletely reflect overall survival among patients who appear to have a long-term benefit,[18,19] and in an amendment approved by the Food and Drug Administration on October 9, 2008, the primary end point was changed from progression-free survival to overall survival before the treatment assignments were revealed. No change in the size of the study was required, since it was already fully powered to assess overall survival. No interim analysis was conducted. Secondary end points included progression-free survival, the rate of best overall response (defined as the proportion of all randomly assigned patients who had a complete or partial response), the rate of disease control (defined as a complete response, a partial response, or stable disease), the time to a response, the duration of the response, and safety. Responses were defined according to the modified World Health Organization criteria as the sum of the products of bidimensional measurements of target lesions; a complete response is defined by the disappearance of all known lesions, a partial response by a decrease of at least 50% from baseline in the sum of the products of the diameters of index lesions, stable disease by failure to meet the criteria for either partial response or progressive disease, and progressive disease by a 25% increase in an existing lesion or the development of a new lesion. Tumor assessments were performed by the local investigator and by a central independent review committee.

### ASSESSMENTS

Radiologic and photographic tumor assessments were performed in all patients at baseline and at week 12, and in patients in whom the disease had not progressed, at weeks 16, 20, and 24 and every 6 weeks through week 48. For patients in whom the disease had not progressed and who remained in the study beyond week 48, tumor assessments were performed every 12 weeks. Patients with progressive disease who entered the follow-up phase were recommended to receive two additional tumor assessments as part of the standard of care. All efficacy end points (except survival) were based on assessments performed by the independent review committee, whose members were not aware of the treatment assignments.

Adverse events were graded with the use of the National Cancer Institute Common Terminology Criteria for Adverse Events, version 3.0 (http://ctep.info.nih.gov/protocolDevelopment/electronic_applications/docs/ctcaev3.pdf). Adverse events were reported from the date of the first dose up to and including 70 days after the final dose. The study included two additional safety analyses. One was an analysis of immune-related adverse events (described previously[12]), which were defined as adverse events that were, according to the judgment of the investigator, associated with inflammation. The list of potential inflammation-associated adverse events was defined prospectively, in the protocol, before the database was locked. The other additional safety analysis was an exploratory analysis performed to determine whether specific adverse events were likely to be immune-mediated and associated with ipilimumab treatment. Immune-mediated adverse reactions, identified through retrospective review of the data, were adjudicated by a physician employed by the sponsor, in order to exclude noninflammatory causes of the event, such as infection or tumor progression. The immune-mediated adverse reactions included specific events of clinical interest, such as enterocolitis, hepatitis, dermatitis, neuropathies, or endocrinopathies, including hypophysitis.

### STATISTICAL ANALYSIS

All analyses of efficacy were performed as prespecified in the protocol except for the 3-year survival rate, which was analyzed post hoc. The target number of events for the primary analysis was 416 deaths, which we estimated would give the study approximately 90% power to detect a 37% increase in median overall survival to 11 months with ipilimumab plus dacarbazine, with a corresponding hazard ratio for death of 0.727, assuming a total sample of 500 patients (250 randomly assigned to each group) and a median survival of 8 months for the patients receiving dacarbazine plus placebo. For the analysis of overall survival, a log-



The New England Journal of Medicine
Downloaded from nejm.org at GALTER HEALTH SCIENCES LIBRARY on July 18, 2011. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

rank test was performed, at a two-sided alpha level of 0.05, stratified according to metastasis stage (M0, M1a, M1b, or M1c) and ECOG performance status (0 or 1), as classified at the time of randomization. The final analysis of progression-free survival was conducted on a database that was locked after 416 events had been documented, as originally planned. The hazard ratios for overall survival and progression-free survival with ipilimumab plus dacarbazine as compared with dacarbazine plus placebo, and the associated two-sided 95% confidence intervals, were estimated with the use of a stratified Cox proportional-hazards model. Survival functions were estimated with the use of the Kaplan–Meier method. A hierarchical testing procedure was implemented for a comparison of the treatment groups in the order of the following end points: overall survival, progression-free survival, rate of disease control, and rate of best overall response. Additional details of the methods are provided in the Supplementary Appendix, available at NEJM.org.

## RESULTS

### PATIENTS AND TREATMENT

Between August 8, 2006, and January 22, 2008, a total of 250 patients were randomly assigned to receive ipilimumab plus dacarbazine (ipilimumab–dacarbazine group) and 252 were randomly assigned to receive dacarbazine plus placebo (dacarbazine group). The baseline characteristics of the patients were balanced between the two groups (Table 1). A total of 92 patients in the ipilimumab–dacarbazine group (36.8%) and 165 patients in the dacarbazine group (65.5%) received all four doses of ipilimumab or placebo. At least one maintenance dose was administered in 43 patients in the ipilimumab–dacarbazine group (17.2%) and 53 in the dacarbazine group (21.0%). The follow-up time between the start of the study (the first visit of the first enrolled patient) and the end of the study (the last visit of the last enrolled patient) was 54 months, and the follow-up time between the time the last patient underwent randomization (the first visit of the last patient) and the end of the study was 36.6 months.

The most frequent reason for discontinuation of the study drug across the entire study was disease progression (in 46.2% of patients in the ipilimumab–dacarbazine group and 77.3% in the dacarbazine group). Discontinuation due to a drug-related adverse event was reported in 89 of the 247 patients in the ipilimumab–dacarbazine group who received at least one dose of the study drug (36.0%) and 10 of the 251 patients in the dacarbazine group who received at least one dose of the study drug (4.0%): 85 of the 247 patients (34.4%) and 10 of the 251 patients (4.0%) in the two groups, respectively, discontinued the study drug because of a drug-related adverse event after receiving treatment in the induction phase; 4 of the 43 patients in the ipilimumab–dacarbazine group who received at least one maintenance dose (9.3%) and none of the 53 patients in the dacarbazine group who received at least one maintenance dose discontinued the therapy after receiving treatment in the maintenance phase.

Therapy after disease progression was balanced between the two groups; 54.7% of the patients in the ipilimumab–dacarbazine group and 59.0% in the dacarbazine group received subsequent therapy. In the ipilimumab–dacarbazine group, 37.7% of the patients received chemotherapy, and 2% immunotherapy; in the dacarbazine group, 34.7% received chemotherapy, and 2% immunotherapy. One patient in the dacarbazine group received the BRAF inhibitor, vemurafenib, at the time of disease progression, but tumors were not routinely assessed for the presence of the BRAF V600E mutation.

### EFFICACY

Efficacy analyses were performed on the intention-to-treat population. A survival analysis was performed after 414 deaths occurred, 37 months after the last patient was enrolled. The median overall survival in the ipilimumab–dacarbazine group was 11.2 months (95% confidence interval [CI], 9.4 to 13.6), as compared with 9.1 months (95% CI, 7.8 to 10.5) in the dacarbazine group, with estimated survival rates in the two groups, respectively, of 47.3% and 36.3% at 1 year, 28.5% and 17.9% at 2 years, and 20.8% and 12.2% at 3 years (hazard ratio for death with ipilimumab–dacarbazine, 0.72; $P<0.001$) (Fig. 1A). Ipilimumab was associated with improved overall survival across patient subgroups, including those defined according to age, sex, ECOG performance status, baseline serum lactate dehydrogenase level, and substage of metastatic disease (Fig. 2).

There was a 24% reduction in the risk of progression in the ipilimumab–dacarbazine group as compared with the dacarbazine group (hazard ratio for progression, 0.76; $P=0.006$). The median values for progression-free survival were similar



291

The New England Journal of Medicine
Downloaded from nejm.org at GALTER HEALTH SCIENCES LIBRARY on July 18, 2011. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

in the two groups because the first assessment of progression occurred at week 12 after the true median. After the first tumor assessment, the Kaplan–Meier curves separated (Fig. 1B).

The rate of disease control (i.e., a complete or partial response or stable disease) did not differ significantly between the two groups: 33.2% in the ipilimumab–dacarbazine group and 30.2% in the dacarbazine group (P=0.41). The rate of best overall response (i.e., a complete or partial response) was 15.2% in the ipilimumab–dacarbazine group and 10.3% in the dacarbazine group (P=0.09) (Table 2). The median duration of response among all randomly assigned patients with a complete or partial response was 19.3 months (95% CI, 12.1 to 26.1) in the ipilimumab–dacarbazine group and 8.1 months (95% CI, 5.19 to 19.8) in the dacarbazine group (P=0.03) (Fig. 1C). Some patients in the study who were receiving ipilimumab had an improvement from partial response to complete response after 6 months (data not shown). Responses in the presence of new lesions were also observed, although these were not captured as part of the best overall response.

**Table 1. Demographic and Baseline Clinical Characteristics of the Patients.***

| Characteristic | Ipilimumab plus Dacarbazine (N=250) | Placebo plus Dacarbazine (N=252) |
|---|---|---|
| Mean age — yr | 57.5 | 56.4 |
| Sex — no. (%) | | |
| Male | 152 (60.8) | 149 (59.1) |
| Female | 98 (39.2) | 103 (40.9) |
| ECOG performance status — no. (%)† | | |
| 0 | 177 (70.8) | 179 (71.0) |
| 1 | 73 (29.2) | 73 (29.0) |
| Metastasis stage — no. (%)‡ | | |
| M0 | 6 (2.4) | 8 (3.2) |
| M1a | 37 (14.8) | 43 (17.1) |
| M1b | 64 (25.6) | 62 (24.6) |
| M1c | 143 (57.2) | 139 (55.2) |
| Lactate dehydrogenase — no. (%) | | |
| ≤ULN | 157 (62.8) | 140 (55.6) |
| >ULN | 93 (37.2) | 110 (43.7) |
| Unknown | 0 | 2 (0.8) |
| Prior adjuvant therapy — no. (%) | 66 (26.4) | 67 (26.6) |

* ULN denotes upper limit of the normal range.
† An Eastern Cooperative Oncology Group (ECOG) performance status of 0 indicates that the patient is fully active and able to carry on all predisease activities without restriction, and an ECOG performance status of 1 indicates that the patient is restricted in physically strenuous activity but is ambulatory and able to carry out work of a light or sedentary nature, such as light housework or office work.[17]
‡ The M1a stage is characterized by metastasis of the tumor to the skin, subcutaneous tissues, or distant lymph nodes, with normal lactate dehydrogenase level; M1b by metastasis to the lung, with normal lactate dehydrogenase level; and M1c by metastasis to any other visceral site or to any site with an elevated lactate dehydrogenase level.

### SAFETY

The safety analysis included all patients who underwent randomization and received at least one dose of the assigned study drug (498 patients). The adverse events reported in the safety population are listed in Table 3. Adverse events (all grades) for which there was a higher incidence in the ipilimumab–dacarbazine group than in the dacarbazine group included elevation of alanine aminotransferase levels (in 33.2% of patients vs. 5.6%), elevation of aspartate aminotransferase levels (29.1% vs. 5.6%), diarrhea (36.4% vs. 24.7%), pruritus (29.6% vs. 8.8%), and rash (24.7% vs. 6.8%). Grade 3 or 4 adverse events occurred in 56.3% of patients receiving ipilimumab plus dacarbazine and in 27.5% of patients receiving placebo plus dacarbazine (P<0.001).

No gastrointestinal perforations were reported. No cases of hypophysitis were noted in the ipilimumab–dacarbazine group except for a single case in a patient receiving maintenance therapy that was reported on day 364 (which was outside the protocol-specified reporting window of <70 days after the last dose — a period representing 5 times the half-life of ipilimumab — and was therefore not categorized as an "on-study" event). No drug-related deaths were reported in the ipilimumab–dacarbazine group; one fatal gastrointestinal hemorrhage was reported in the dacarbazine group. Nausea and vomiting (see the table in the Supplementary Appendix) as well as myelosuppression (data not shown), which are common side effects of dacarbazine, were not increased in the ipilimumab–dacarbazine group as compared with the dacarbazine group.

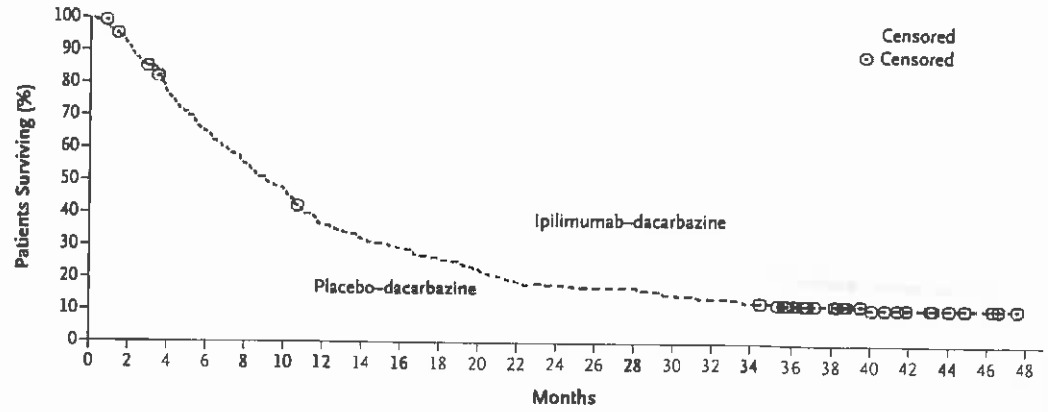
The most common study-drug–related adverse events were those classified as immune-related adverse events, which were seen in 77.7% of the patients treated with ipilimumab plus dacarbazine and 38.2% of the patients treated with dacarbazine and placebo (Table 3). The most common immune-related adverse events in the ipilimumab–dacarbazine group were elevated liver-function values, with grade 3 or 4 elevations in liver-function values noted in 17.4 to 20.7% of the patients.

An alternative characterization of inflammatory events was performed in which noninflammatory causes of the event (e.g., tumor progression or in-



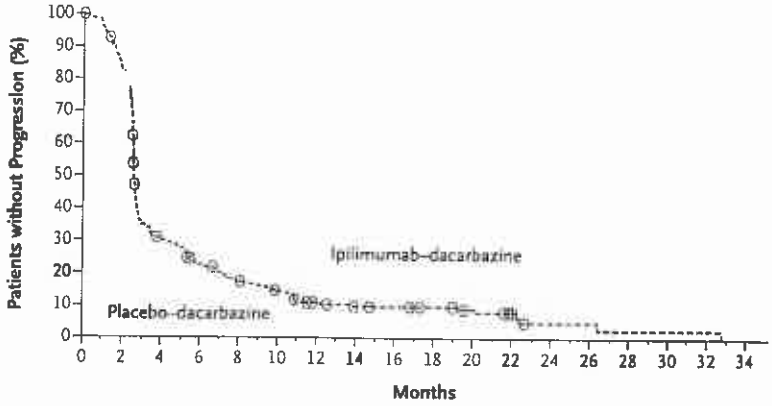
The New England Journal of Medicine
Downloaded from nejm.org at GALTER HEALTH SCIENCES LIBRARY on July 18, 2011. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

A



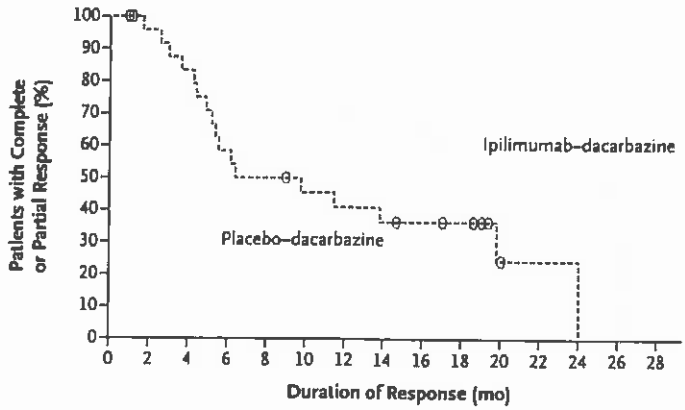

| No. at Risk | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ipilimumab–dacarbazine | 250 | 230 | 199 | 181 | 157 | 131 | 114 | 104 | 91 | 85 | 79 | 74 | 68 | 61 | 59 | 56 | 56 | 52 | 41 | 31 | 17 | 10 | 4 | 2 | 0 |
| Placebo–dacarbazine | 252 | 229 | 190 | 160 | 136 | 116 | 89 | 78 | 72 | 64 | 56 | 47 | 44 | 42 | 42 | 37 | 34 | 31 | 26 | 19 | 11 | 7 | 5 | 3 | 0 |

B




| No. at Risk | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ipilimumab–dacarbazine | 250 | 199 | 85 | 70 | 57 | 45 | 40 | 35 | 30 | 25 | 16 | 10 | 6 | 4 | 3 | 2 | 1 0 |
| Placebo–dacarbazine | 252 | 205 | 72 | 52 | 39 | 30 | 20 | 16 | 15 | 13 | 10 | 7 | 2 | 2 | 1 | 1 | 1 0 |

C




| No. at Risk | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ipilimumab–dacarbazine | 38 | 38 | 33 | 30 | 27 | 23 | 22 | 20 | 17 | 8 | 4 | 1 | 1 | 1 | 0 |
| Placebo–dacarbazine | 26 | 23 | 20 | 14 | 12 | 10 | 9 | 8 | 7 | 6 | 2 | 1 | 1 | 0 | 0 |



293

The New England Journal of Medicine
Downloaded from nejm.org at GALTER HEALTH SCIENCES LIBRARY on July 18, 2011. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

**Figure 1 (facing page). Kaplan–Meier Curves for Overall Survival, Progression-free Survival, and Duration of Response.**

Overall survival and progression-free survival in the intention-to treat population are shown in Panels A and B, respectively, and the duration of complete or partial response is shown in Panel C.

fection) were ruled out (immune-mediated adverse reactions). Severe (grade 3 or higher) immune-mediated adverse reactions were seen in 38.1% of the patients in the ipilimumab–dacarbazine group and 4.4% of the patients in the dacarbazine group. The most common grade 3 or 4 immune-mediated adverse reaction was immune-mediated hepatitis, which was seen in 78 patients in the ipilimumab–dacarbazine group (31.6%) and in 6 patients in the dacarbazine group (2.4%). Grade 3 or 4 immune-mediated enterocolitis was seen in 12 patients in the ipilimumab–dacarbazine group (4.9%) and no patients in the dacarbazine group. Hepatic immune-mediated adverse reactions were generally reversible with treatment according to established guidelines specified in the research protocol. The proportion of patients who received glucocorticoids or other immunosuppressant agents after the emergence of high-grade immune-mediated hepatitis was 80.8% (63 of 78 patients, including 5 patients who received mycophenolate mofetil) in the ipilimumab–dacarbazine group and 33.3% (2 of 6 patients) in the dacarbazine group. High-grade elevations in liver-function values were normalized in 67.9% of patients in the ipilimumab–dacarbazine group, with a median time to normalization of 9.9 weeks (range, 1.0 to 56.1; interquartile range, 7.0 to 15.0). No patient died

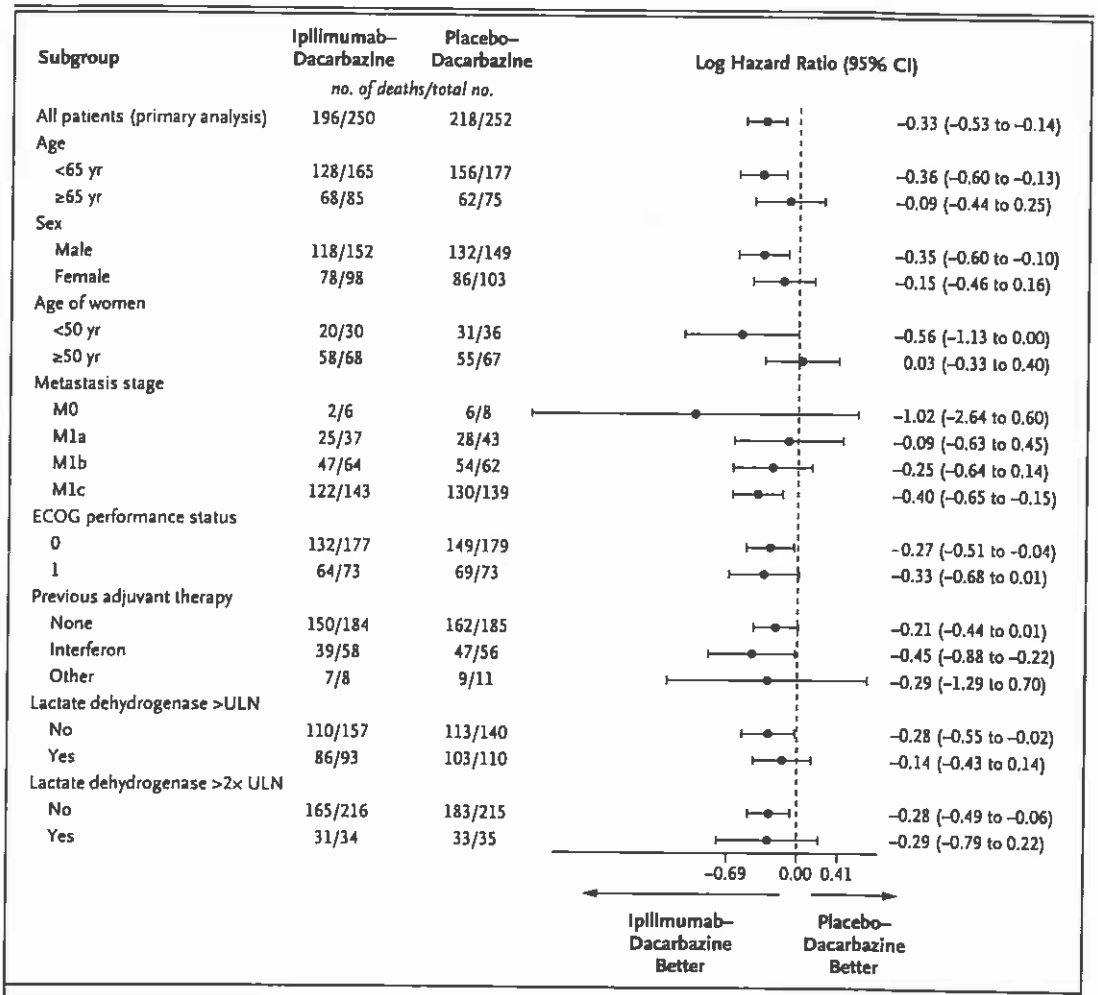



**Figure 2. Overall Survival According to Subgroup.**

ECOG denotes Eastern Cooperative Oncology Group, and ULN the upper limit of the normal range.



The New England Journal of Medicine
Downloaded from nejm.org at GALTER HEALTH SCIENCES LIBRARY on July 18, 2011. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

**Table 2. Efficacy Results.**

| End Point | Ipilimumab plus Dacarbazine (N=250) | Placebo plus Dacarbazine (N=252) | Hazard Ratio with Ipilimumab plus Dacarbazine (95% CI) | P Value |
|---|---|---|---|---|
| **Primary end point: overall survival** | | | | |
| No. of deaths | 196 | 218 | 0.72 (0.59–0.87) | <0.001 |
| Survival — % (95% CI) | | | | |
| 1 yr | 47.3 (41.0–53.6) | 36.3 (30.4–42.4) | | |
| 2 yr | 28.5 (22.9–34.2) | 17.9 (13.3–22.8) | | |
| 3 yr | 20.8 (15.7–26.1) | 12.2 (8.2–16.5) | | |
| **Secondary end points** | | | | |
| Disease progression — no. of events | 203 | 223 | 0.76 (0.63–0.93) | 0.006 |
| Best overall response — no. (%)* | 38 (15.2) | 26 (10.3) | | |
| Complete response | 4 (1.6) | 2 (0.8) | | |
| Partial response | 34 (13.6) | 24 (9.5) | | |
| Stable disease — no. (%)* | 45 (18.0) | 50 (19.8) | | |
| Progressive disease — no. (%) | 111 (44.4) | 131 (52.0) | | |
| Response not evaluated — no. (%)† | 56 (22.4) | 45 (17.9) | | |

* The rate of disease control (defined as a complete response, a partial response, or stable disease) was 33.2% in the ipilimumab–dacarbazine group and 30.2% in the dacarbazine group (P=0.41).

† To be evaluated for a response, patients had to undergo baseline and follow-up scanning. Owing to early progression, no follow-up scans were available for some patients.

as a result of complications of immune-mediated hepatitis or enterocolitis during the course of the study.

Among the 43 patients who received ipilimumab and the 53 patients who received placebo during the maintenance phase, the most common adverse events (all grades, occurring in >5% of patients in the ipilimumab group) were rash (25.6% vs. 5.7%), pruritus (16.3% vs. 3.8%), diarrhea (14.0% vs. 5.7%), nausea (7.0% vs. 5.7%), and fatigue (9.3% vs. 3.8%). Low-grade increases in liver-function values were noted in 2 patients receiving ipilimumab during the maintenance phase. No high-grade elevations in liver-function values were noted among the patients receiving ipilimumab during the maintenance phase. High-grade adverse events during the maintenance phase were infrequent; among the grade 3 or 4 adverse events noted were colitis and diarrhea (in 1 patient) and rash or pruritus (in 3 patients). No grade 3 or 4 adverse events occurred in patients receiving placebo during the maintenance phase.

## DISCUSSION

In this phase 3 study, ipilimumab (at a dose of 10 mg per kilogram), in combination with dacarbazine (at a dose of 850 mg per square meter), as compared with dacarbazine plus placebo, was associated with a significant improvement in overall survival among patients with previously untreated metastatic melanoma. More than 50% of the patients had M1c disease (indicating the presence of visceral metastases, elevated lactate dehydrogenase levels, or both), and more than 35% had baseline elevations in lactate dehydrogenase levels — both of which are associated with poor survival.[20,21] Durable objective responses were observed (median duration of best overall response, 19.3 months in the ipilimumab–dacarbazine group, vs. 8.1 months in the dacarbazine group). Prolonged survival was noted among some patients who were followed for up to 4 years. In the ipilimumab–dacarbazine group, an estimated 28.5% of the patients were alive at 2 years, and an estimated 20.8% at 3 years, as compared with an estimated 17.9% and 12.2%, respectively, in the dacarbazine group.

The rates of some high-grade adverse events in the current study differ from those in previous phase 2 studies of ipilimumab monotherapy (at a dose of 10 mg per kilogram).[13,22-24] The absence of gastrointestinal perforations and the low rates of other grade 3 or 4 gastrointestinal



295

The New England Journal of Medicine
Downloaded from nejm.org at GALTER HEALTH SCIENCES LIBRARY on July 18, 2011. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

**Table 3. Adverse Events and Immune-Related Adverse Events.***

| Adverse Event | Ipilimumab plus Dacarbazine (N=247) | | | Placebo plus Dacarbazine (N=251) | | |
|---|---|---|---|---|---|---|
| | Total | Grade 3 | Grade 4 | Total | Grade 3 | Grade 4 |
| | *number of patients (percent)* | | | | | |
| **All adverse events, regardless of cause†** | | | | | | |
| Any event | 244 (98.8) | 99 (40.1) | 40 (16.2) | 236 (94.0) | 45 (17.9) | 24 (9.6) |
| Gastrointestinal: diarrhea | 90 (36.4) | 10 (4.0) | 0 | 62 (24.7) | 0 | 0 |
| Dermatologic | | | | | | |
| Pruritus | 73 (29.6) | 5 (2.0) | 0 | 22 (8.8) | 0 | 0 |
| Rash | 61 (24.7) | 3 (1.2) | 0 | 17 (6.8) | 0 | 0 |
| Hepatic | | | | | | |
| Increase in alanine aminotransferase | 82 (33.2) | 40 (16.2) | 14 (5.7) | 14 (5.6) | 2 (0.8) | 0 |
| Increase in aspartate aminotransferase | 72 (29.1) | 36 (14.6) | 9 (3.6) | 14 (5.6) | 3 (1.2) | 0 |
| Other | | | | | | |
| Pyrexia | 91 (36.8) | 0 | 0 | 23 (9.2) | 0 | 0 |
| Chills | 28 (11.3) | 0 | 0 | 10 (4.0) | 0 | 0 |
| Weight loss | 27 (10.9) | 1 (0.4) | 0 | 13 (5.2) | 1 (0.4) | 0 |
| **Immune-related adverse events** | | | | | | |
| Any event | 192 (77.7) | 78 (31.6) | 25 (10.1) | 96 (38.2) | 8 (3.2) | 7 (2.8) |
| Dermatologic | | | | | | |
| Pruritus | 66 (26.7) | 5 (2.0) | 0 | 15 (6.0) | 0 | 0 |
| Rash | 55 (22.3) | 3 (1.2) | 0 | 12 (4.8) | 0 | 0 |
| Gastrointestinal | | | | | | |
| Diarrhea | 81 (32.8) | 10 (4.0) | 0 | 40 (15.9) | 0 | 0 |
| Colitis | 11 (4.5) | 4 (1.6) | 1 (0.4) | 0 | 0 | 0 |
| Hepatic‡ | | | | | | |
| Increase in alanine aminotransferase | 72 (29.1) | 37 (15.0) | 14 (5.7) | 11 (4.4) | 2 (0.8) | 0 |
| Increase in aspartate aminotransferase | 66 (26.7) | 34 (13.8) | 9 (3.6) | 8 (3.2) | 1 (0.4) | 0 |
| Hepatitis | 4 (1.6) | 3 (1.2) | 0 | 0 | 0 | 0 |

* The safety analysis included all patients who underwent randomization and received at least one dose of study drug (498 patients). Adverse events and immune-related adverse events were prospectively defined: the *Medical Dictionary for Regulatory Activities* (MedDRA), version 13.0, was used for the reporting of adverse events, and a list of events prespecified in the protocol was used to capture immune-related adverse events, which were a subgroup of the reported adverse events. The categories are not mutually exclusive (i.e., one patient could have events in multiple categories).

† A complete list of adverse events that occurred in at least 10% of patients is available in the Supplementary Appendix.

‡ Terms used in the category of hepatic immune-related adverse events are MedDRA preferred terms, as listed by the investigator in case-report forms.

events are in contrast to the results of phase 2 studies involving 325 patients (rate of diarrhea, 4% in this study vs. 11% in the previous studies; rate of colitis, 2% vs. 5%). No cases of hypophysitis were observed during the course of this study, whereas 2% of the patients in the previous studies had high-grade endocrinopathy. The rate of grade 3 or 4 rash was similar (1% in the current study and 2% in the previous studies). However, the rates of high-grade hepatic adverse events in the current study were higher than those in the previous studies (elevated alanine aminotransferase level, 21% vs. 8%; elevated aspartate aminotransferase level, 17% vs. 7%). The apparent shift in the rates of adverse events associated with ipilimumab may be due to its combination with dacarbazine, which is known to cause hepatotoxic effects when it is used as monotherapy.[25-27] The role of systemic glucocorticoids (which were administered to manage hepatic events) in the appar-



The New England Journal of Medicine
Downloaded from nejm.org at GALTER HEALTH SCIENCES LIBRARY on July 18, 2011. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

ent reduction in gastrointestinal and endocrine events is undefined.

In summary, this trial showed that there was a significant improvement in overall survival among patients with previously untreated metastatic melanoma who received ipilimumab plus dacarbazine as compared with dacarbazine plus placebo. Adverse events other than those typically seen with dacarbazine or ipilimumab therapy were not identified. An increase in liver-function values is an important side effect that was observed more frequently than expected with the combination therapy. Other ipilimumab-associated adverse events (enterocolitis and endocrinopathy) were observed, albeit at a rate that was lower than expected. The key side effects of ipilimumab were managed through adherence to treatment according to well-established guidelines, including the administration of systemic glucocorticoids or other immunosuppressant agents.

Supported by Bristol-Myers Squibb.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank the patients and all the investigators who participated in the study and the representatives of the sponsor, Haolan Lu, Ph.D., Mamatha Kodidela, and Veerle de Pril, for their support in statistical analysis. Professional medical writing and editorial assistance were provided by Sudha Vemuri, Ph.D., and Ananya Bhattacharya, Ph.D., of Bristol-Myers Squibb.

## REFERENCES

1. Middleton MR, Grob JJ, Aaronson N, et al. Randomized phase III study of temozolomide versus dacarbazine in the treatment of patients with advanced metastatic malignant melanoma. J Clin Oncol 2000;18:158-66.
2. Falkson CI, Ibrahim J, Kirkwood JM, Coates AS, Atkins MB, Blum RH. Phase III trial of dacarbazine versus dacarbazine with interferon alpha-2b versus dacarbazine with tamoxifen versus dacarbazine with interferon alpha-2b and tamoxifen in patients with metastatic malignant melanoma: an Eastern Cooperative Oncology Group Study. J Clin Oncol 1998;16:1743-51.
3. Rusthoven JJ, Quirt IC, Iscoe NA, et al. Randomized, double-blind, placebo-controlled trial comparing the response rates of carmustine, dacarbazine, and cisplatin with and without tamoxifen in patients with metastatic melanoma. J Clin Oncol 1996;14:2083-90.
4. Julia F, Thomas L, Dumontet C, Dalle S. Targeted therapies in metastatic melanoma: toward a clinical breakthrough? Anticancer Agents Med Chem 2010;10:661-5.
5. Agarwala SS. Current systemic therapy for metastatic melanoma. Expert Rev Anticancer Ther 2009;9:587-95.
6. Atkins MB, Lotze MT, Dutcher JP, et al. High-dose recombinant interleukin 2 therapy for patients with metastatic melanoma: analysis of 270 patients treated between 1985 and 1993. J Clin Oncol 1999;17:2105-16.
7. Atkins MB, Kunkel L, Sznol M, Rosenberg SA. High-dose recombinant interleukin-2 therapy in patients with metastatic melanoma: long-term survival update Cancer J Sci Am 2000;6:Suppl 1:S11-S14.
8. Melero I, Hervas-Stubbs S, Glennie M, Pardoll DM, Chen L. Immunostimulatory monoclonal antibodies for cancer therapy. Nat Rev Cancer 2007;7:95-106.
9. O'Day SJ, Hamid O, Urba WJ. Targeting cytotoxic T-lymphocyte antigen-4 (CTLA-4): a novel strategy for the treatment of melanoma and other malignancies. Cancer 2007;110:2614-27.
10. Fong L, Small EJ. Anti-cytotoxic T-lymphocyte antigen-4 antibody: the first in an emerging class of immunomodulatory antibodies for cancer treatment. J Clin Oncol 2008;26:5275-83.
11. Robert C, Ghiringhelli F. What is the role of cytotoxic T lymphocyte-associated antigen 4 blockade in patients with metastatic melanoma? Oncologist 2009;14:848-61.
12. Hodi FS, O'Day SJ, McDermott DF, et al. Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med 2010;363:711-23. [Erratum, N Engl J Med 2010;363:1290.]
13. Wolchok JD, Neyns B, Linette G, et al. Ipilimumab monotherapy in patients with pretreated advanced melanoma: a randomised, double-blind, multicentre, phase 2, dose-ranging study. Lancet Oncol 2010;11:155-64.
14. Hersh EM, O'Day SJ, Powderly J, et al. A phase II multicenter study of ipilimumab with or without dacarbazine in chemotherapy-naive patients with advanced melanoma. Invest New Drugs 2011;29:489-98.
15. Mokyr MB, Kalinichenko T, Gorelik L, Bluestone JA. Realization of the therapeutic potential of CTLA-4 blockade in low-dose chemotherapy-treated tumor-bearing mice. Cancer Res 1998;58:5301-4.
16. Zitvogel L, Apetoh L, Ghiringhelli F, Kroemer G. Immunological aspects of cancer chemotherapy. Nat Rev Immunol 2008;8:59-73.
17. Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982;5:649-55.
18. Wolchok JD, Hoos A, O'Day S, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. Clin Cancer Res 2009;15:7412-20.
19. Hoos A, Eggermont AM, Janetzki S, et al. Improved endpoints for cancer immunotherapy trials. J Natl Cancer Inst. 2010;102:1388-97.
20. Balch CM, Gershenwald JE, Soong SJ, et al. Final version of 2009 AJCC melanoma staging and classification. J Clin Oncol 2009;27:6199-206.
21. Bedikian AY, Johnson MM, Warneke CL, et al. Prognostic factors that determine the long-term survival of patients with unresectable metastatic melanoma. Cancer Invest 2008;26:624-33.
22. Weber J, Thompson JA, Hamid O, et al. A randomized, double-blind, placebo controlled, phase II study comparing the tolerability and efficacy of ipilimumab administered with or without prophylactic budesonide in patients with unresectable stage III or IV melanoma. Clin Cancer Res 2009;15:5591-8.
23. O'Day SJ, Maio M, Chiarion-Sileni V, et al. Efficacy and safety of ipilimumab monotherapy in patients with previously treated, advanced melanoma: a multicenter single-arm phase II study. Ann Oncol 2010;21:1712-7.
24. Schmidt H, Hamid O, Nissan A, et al. Identification of tumor biopsy markers as potential predictors of ipilimumab clinical activity in patients with advanced melanoma. Eur J Cancer Suppl 2009;7:577.
25. Dancygier H, Runne U, Leuschner U, Milbradt R, Classen M. Dacarbazine (DTIC)-induced human liver damage light and electron-microscopic findings. Hepatogastroenterology 1983;30:93-5.
26. Greenstone MA, Dowd PM, Mikhailidis DP, Scheuer PJ. Hepatic vascular lesions associated with dacarbazine treatment. Br Med J (Clin Res Ed) 1981;282:1744-5.
27. Frosch PJ, Czarnetzki BM, Macher E, Grundmann E, Gottschalk I. Hepatic failure in a patient treated with dacarbazine (DTIC) for malignant melanoma. J Cancer Res Clin Oncol 1979;95:281-6.

Copyright © 2011 Massachusetts Medical Society.



297

The New England Journal of Medicine
Downloaded from nejm.org at GALTER HEALTH SCIENCES LIBRARY on July 18, 2011. For personal use only. No other uses without permission.
Copyright © 2011 Massachusetts Medical Society. All rights reserved.

After the initial administration of therapeutic antibodies, one may then ask whether the human patient can use the killed cancer cells as vaccines to induce active immunity. As in the case of bacterial infection, when the bacteria is killed by immune serum or antibiotics, the dead bacteria can stimulate the human immune system to generate more anti-bacterial antibodies thus protecting the patient from re-infection. However, the monoclonal therapeutic antibodies used in the treatment of cancer, autoimmunity, transplantation rejection, etc. are directed against normal human proteins. Our immune system has evolved to prevent attacking our own proteins, a mechanism still not fully understood and known as tolerance. As a result, in general, no active immunity results after therapeutic antibody treatments.

**Limitations of therapeutic antibodies**

Just like any other medication, monoclonal therapeutic antibodies decay in the human system even for the best designed ones. Their half-lives are about three to four weeks at the best. To understand the situation in detail, we need to analyze it using standard engineering approaches. Let V denote the amount of cancer cells, for example, and let A denote the amount of therapeutic antibodies injected into the human patient. Then we have the following two differential equations for the rates of changes of V and A:

$$dV / dt = \alpha V - \beta V A$$

$$dA / dt = - \varepsilon A$$

The initial conditions are at $t = 0$, $V = V_0$ and $A = A_0$. The constants $\alpha$, $\beta$ and $\varepsilon$ are all positive. The second equation can be solved and the solution substituted into the first equation.

$$A = A_0 e^{-\varepsilon t}$$

The first equation becomes:

$$dV / V = [\alpha - \beta A_0 e^{-\varepsilon t}] dt$$

The solution of this equation is:

$$\ln V = \alpha t + (\beta A_0 / \varepsilon) e^{-\varepsilon t} + \text{constant}$$

The constant can be determined from the initial condition. Therefore we have:

$$\ln V = \ln V_0 + \alpha t - (\beta A_0 / \varepsilon) [1 - e^{-\varepsilon t}]$$

This result can also be written as:

$$V = V_0 \exp \{\alpha t - (\beta A_0 / \varepsilon) [1 - e^{-\varepsilon t}]\}$$

In order for V to decrease initially, i.e. for small t, the term inside { } must be negative. It can be approximated as:

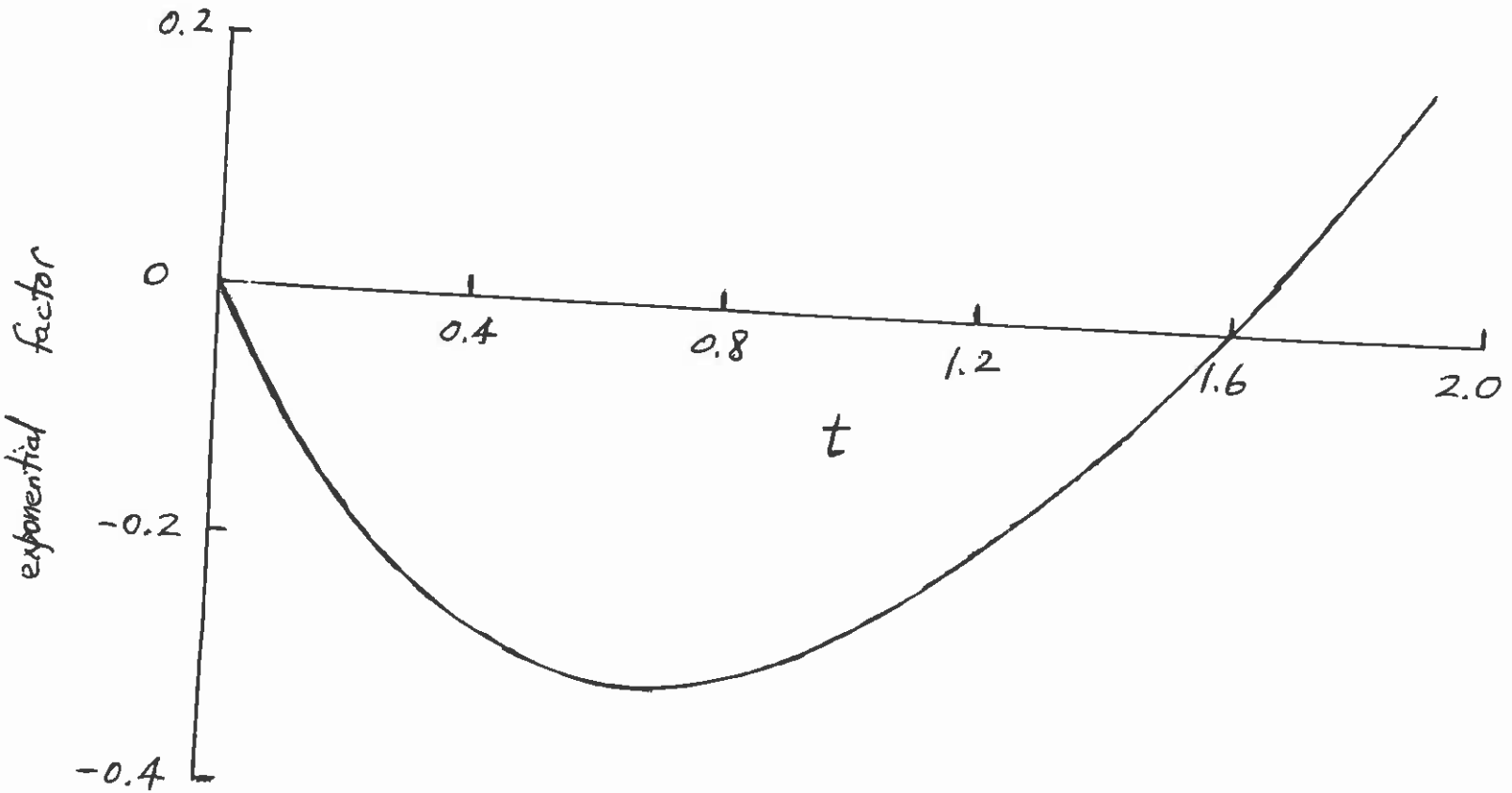
$$\alpha t - (\beta A_0 / \varepsilon) [1 - (1 - \varepsilon t)] = (\alpha - \beta A_0) t$$

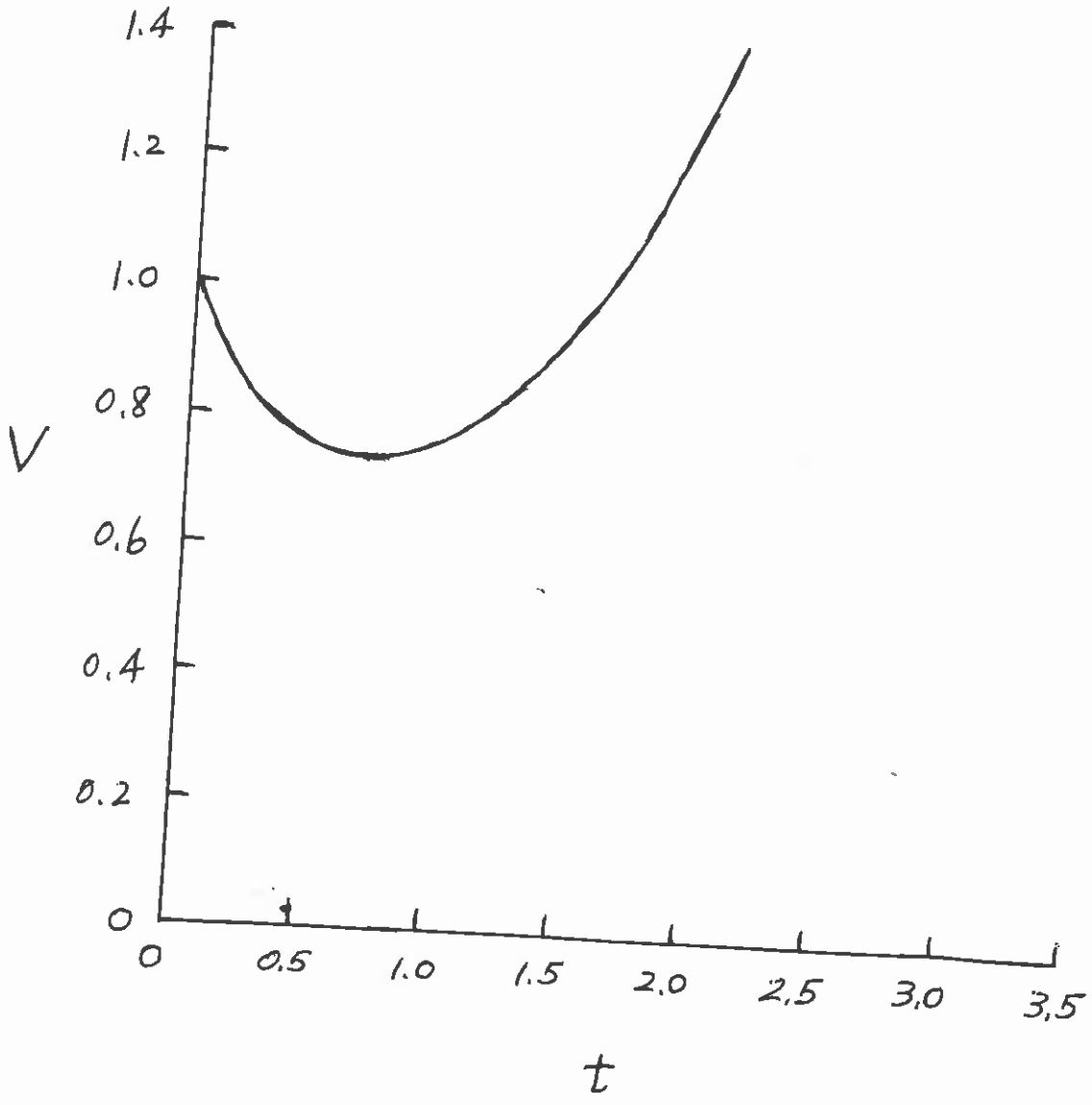
Therefore the amount of monoclonal therapeutic antibody to be given initially, i.e. $A_0$, has to be:

$$A_0 > \alpha / \beta$$

The constant $\alpha$ is the rate of growth of cancer cells, and $\beta$ is related to the efficiency of the elimination of cancer cells by the monoclonal therapeutic antibody. In other words, depending on this efficiency, a substantial amount of the therapeutic antibody has to be given initially to reduce the number of cancer cells.

As an explicit example, consider a simple example with $V_0 = 1$, $A_0 = 1$, $\alpha = 1$, $\beta = 2$, and $\varepsilon = 1$. We can plot the terms inside { }, i.e. the exponential term, as a function of t, as well as V as a function of t. These curves are shown on the next two pages.


0.2
0
-0.2
-0.4
exponential factor
0.4
0.8
1.2
1.6
2.0
t


1.4
1.2
1.0
0.8
V
0.6
0.4
0.2
0
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
t

As illustrated here, passive immunity using monoclonal therapeutic antibodies can only reduce the number of cancer cells, but can never eliminate them completely even with numerous repeated injections. Thus, to "cure" cancer with therapeutic antibodies, one needs to use these therapeutic antibodies indefinitely, similar to the indefinite injections of insulin to treat type I diabetes mellitus.

What is a possible solution to this situation? Somehow, active immunity has to be induced. Then the patient can start to manufacture antibodies against the specific cancer cells. In this case, one needs to overcome the mechanism of tolerance.

In an interesting study of mouse immunized with pigeon cytochrome c, the initial antibodies produced by the mouse has higher binding affinity for the mouse cytochrome c than the pigeon cytochrome c (Minnerath JM, Wakem LP, 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, 92, 12379-12383). These authors suggested that their result supported the longstanding but unproven hypothesis that B lymphocytes specific for self antigens can be used in the immune response to foreign antigens.

If this hypothesis is correct, then one may consider the following approach for the treatment of cancer. Let us consider the use of rituximab against the B lymphocyte surface antigen CD20. After four injections of rituximab, this medication cannot be continued for the treatment of certain types of B cell lymphomas due to the antigenicity of the mouse portion of the therapeutic antibody. Then at that stage, one may inject mouse CD20 into the patient. This should stimulate the patient's B lymphocytes which produce anti-CD20 antibodies into action, as in the case of pigeon cytochrome c in mouse. If this active immunity can be sustained, then the patient should produce enough human anti-CD20 antibodies to attack the B cell lymphoma. This may be a possible approach to overcome tolerance to generate autoimmunity to fight cancer cells.

*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 $V_\kappa$ gene combinations and, possibly, somatic mutation of the $V_H$ and $V_\kappa$ 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; $V_\kappa$, 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).

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 $V_\kappa$ 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

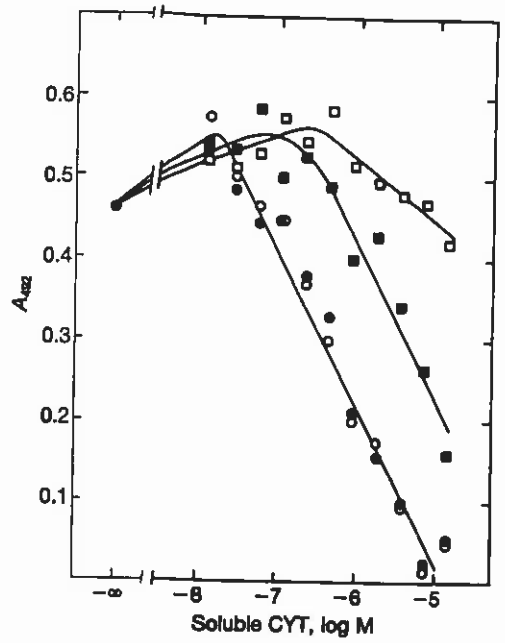



FIG. 1. Competitive ELISA of a mAb obtained 8 days after injecting BALB/c mice with pigeon CYT-OVA. mAbs, such as the one tested here, were generated in splenic fragment cultures. For the competitive ELISA, pigeon CYT was adsorbed to microtiter plates, and the ability of several CYTs to inhibit mAb binding 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 $V_\kappa$ 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 $V_\kappa$ gene (*H3*; ref. 27). Although this $V_\kappa$ segment has not been observed among anti-mouse CYT-specific mAbs, it could be used infrequently. In fact, several different $V_\kappa$ 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

305

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-

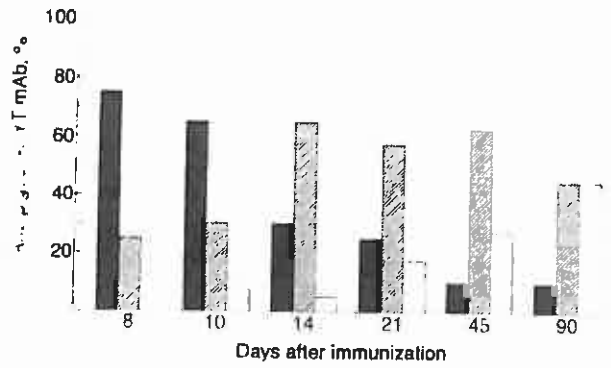



FIG. 3. Relative binding of anti-pigeon CYT mAbs to pigeon and mouse CYTs at various times during the B-lymphocyte response to pigeon CYT-OVA. ■, Binding to mouse CYT > pigeon CYT; ■, binding to mouse CYT = pigeon CYT; □, binding to mouse CYT < pigeon CYT. Binding was assessed by indirect ELISA. Results 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 $V_\kappa$ gene encoding the L chain of mAb 5H11 differs from the $V_\kappa$ gene *R9*, it appears to derive from the $V_\kappa$ 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 $V_\kappa$ 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 $V_\kappa$ 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 $V_\kappa$ gene *R9*. The third mAb derives from a $V_\kappa$ 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

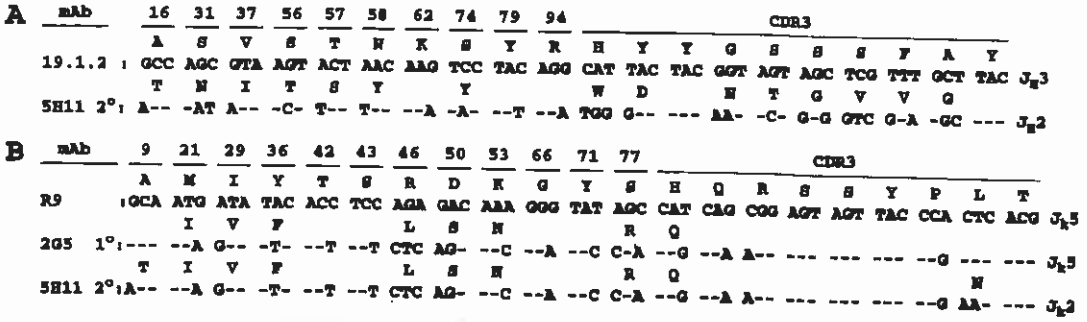



FIG. 4. cDNA and deduced amino acid sequences of the pigeon CYT-specific mAb 5H11 H chain (*A*) and L chain (*B*). The 5H11 hybridoma was prepared 3 months after injecting BALB/c mice with pigeon CYT-OVA. See the legend to Fig. 2 for details of the sequence comparisons shown here. The L chain cDNA sequence of the mouse CYT-specific mAb 2G5 is also given. Like *R9*, the $V_\kappa$ 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.