Reconstitution of the apoenzyme of cytochrome oxidase from Pseudomonas aeruginosa with heme a, and other heme groups.
Kerriane E. Hill and David C. Wharton

A uniting mechanism for stimulation of mammalian glycogen debranching, kinase by reduced nicotinamide adenine dinucleotide, dibasic phosphatase, acetyl coenzyme A, or pyruvate.
Richard L. Cole and Thomas E. Bohn

Purification and characterization of human erythrocyte pyruvate dehydrogenase by reduced nicotinamide adenine dinucleotide phosphate. Yuvali Waxin, Deborah Doyle, and David W. Martin, Jr.

Interaction between DNA and Escherichia coli Protein α. Formation of a complex between single-stranded DNA and α protein.
Richard E. Depue, Leroy F. Liu, and James C. Wang

Preparation and properties of the major copper-binding component in human fetal liver. Its identification as metallothionein.
Laura Ryden and Harold F. Deitch

Molecular weights of aggregation states of Staphylococcus homoyalamin.
Sharon Quitter, Laura A. Watts, Carol Crosby, and Robert Basby

Radioimmunoassay and characterization of enokphalins in rat tissues.
Richard J. Miller, Kwan-Has Chung, Barrett Cooper, and Peldy Cautreusos

Inhibition of mammalian S-adenosylmethionine decarboxylase activity by 1,1-(ethane-1,2-diyl)bis[2-methyl-1H-benzimidazole].
Anthony E. Pegg

3,3',5-Tetrahydropyridine acid. Conformation of the tetrahydropyrimidine ring.
Martin Pfe and Karol Haage

Interactions of α-antipain with trypsin and chymotrypsin.
James W. Blom and Margaret J. Hunter

Purification and properties of a heat-stable protein inhibiting polynucleotide phosphorylase from rabbit liver.
Rhul A. Khondeal and Sunit M. Ewman

Gene 4 protein of bacteriophage T7. Purification, physical properties, and stimulation of T7 DNA polymerase during the elongation of polynucleotide chains.
Richard Kolodner, Yakov Manasune, J. Eugene LeClerc, and Charles C. Richardson

Gene 4 protein of bacteriophage T7. Characterization of the product synthesized by the T7 DNA polymerase and gene 4 protein in the absence of ribonucleoside 5'-triphosphate.
Richard Kolodner and Charles C. Richardson

Preliminary refinement and structural analysis of the Fab fragment from human immunoglobulin New at 2.8 A resolution.
Frederick A. Saul, L. Maria Ansal, and Robert J. Poljak

Purification of interferon from mouse Ehrlich ascites tumor cells.
Masaru Itagaki, Barulome Cabe, Haberabeb Taiwe, Mathey Rebello, Elizabeth Slateter, Hengshui Wondol, and Peter Lengel

Cellular RNA synthesis in normal and meng-support-infected L-929 cells.
James W. Agrell and Edward E. Penhoet

Nucleotide sequence of the DNA encoding the 5'-terminal sequences of simian virus 40 late mRNA.
Basi Dhar, V. Bhaskara Reddy, and Sherman M. Weissman

Nucleotide sequence of the genes for the simian virus 40 protein VP2 and VP1.
V. Bhaskara Reddy, Basi Dhar, and Sherman M. Weissman

Production of a non-immunoglobulin thyroid stimulator by human lymphocytes during mixed culture with human thyroid cells.
Paul Rapaport, Rao J. Pillarjetty, Elizabeth A. Herman, Oris F. Clark, and Evangelos G. Conger

Corticosteroid suppression of lymphocytic thyroid stimulator production.
Elizabeth A. Herman, Rao J. Pillarjetty, and Paul Rapaport

ADDITIONS AND CORRECTIONS

Charles McHenry and Karth Kornberg

NOTE: EFFECTIVE JANUARY 1, 1978, THE MANUSCRIPT HANDLING CHARGE IS NO LONGER REQUIRED.

THE JOURNAL OF BIOLOGICAL CHEMISTRY

GENERAL INFORMATION

MANUSCRIPT SUBMISSION AND PAGE CHARGES

Submit manuscripts of full papers in duplicate and communications in triplicate to:
Editor, The Journal of Biological Chemistry
920 Rockville Pike
Bethesda, Maryland 20011, U.S.A.

Accepted manuscripts will be published with the implicit understanding that the authors will pay a charge of $50 per published page (subsequent supplements at cost). Under exceptional circumstances, when no source of grant or other support exists, the author(s) of accepted manuscripts may apply for a grant-in-aid to Chairman, Publications Committee, American Society of Biological Chemists, Inc., 920 Rockville Pike, Bethesda, Maryland 20014.

Manuscripts on matters of general editorial policy, requests for reprints of the "Instructions to Authors," or of the "Editorial Policy and Practice," or for permission to reproduce any part of a previously published article should be directed to the Journal Editorial Office in Bethesda. (Telephone 301-520-7150.)

Address all correspondence relative to subscriptions, subscription fulfillment, and orders for back copies to: The Journal of Biological Chemistry, 412 East Pratt Street, Baltimore, Maryland 21202, U.S.A. Requests for termination of paid and duly entered subscriptions will be honored with proportional refunds less a $10 cancellation charge.

Published semi-monthly by the American Society of Biological Chemists, Inc. Annual volume: $200, United States; $210, other countries. Single copy $8.50. Special subscription rates are available to members, graduate students, and postdoctoral fellows for personal use only. Qualifying forms are available from the Bethesda Office. Second class postage paid at Baltimore, Maryland 21202, U.S.A., and at additional mailing offices.

PFIZER EX. 1583
Page 3
Preliminary Refinement and Structural Analysis of the Fab Fragment from Human Immunoglobulin New at 2.0 Å Resolution*

(Received for publication, July 7, 1977)

FREDERICK A. SAUL, L. MARIO AMZEL, AND ROBERTO J. POLIAK†

From the Department of Biophysics, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

The three-dimensional structure of the Fab fragment from human myeloma IgG New has been refined using "model building" and "real space" procedures. By these techniques, the correlation between the amino acid sequences and the 2.0 Å resolution multiple isomorphous replacement Fourier map has been optimized. The average shift of all atoms during real space refinement was 0.82 Å. A list of the refined atomic coordinates for the 440 amino acid residues in the structure is given. Ramachandran plots prepared using the refined coordinates show a distribution of $\phi$, $\psi$ angular values which corresponds to the predominant $\beta$-pleated sheet conformation present in the structure.

The structures of the homology subunits $\text{V}_H$, $\text{V}_L$, $\text{C}_H1$, and $\text{C}_L$ were superimposed by pairs and quantitatively compared. The closest similarities were observed between $\text{V}_H$ and $\text{V}_L$ and between $\text{C}_H1$ and $\text{C}_L$. Amino acid sequence alignments obtained from this structural superposition are given. The closest sequence homology in Fab New is observed between $\text{C}_H1$ (gamma heavy chain) and $\text{C}_L$ (lambda light chain). In addition, there is considerable homology between the variable and constant regions.

The distances of close contacts between the homology subunits of Fab New have been determined. The closer contacts, those between atoms at a distance $\leq 1.2$ times their van der Waals radii, are analyzed in relation to the constant, variable, and hypervariable nature of the immunoglobulin sequence positions at which they occur. Most of the residues which determine the closer contacts between $\text{V}_H$ and $\text{V}_L$ and between $\text{C}_H1$ and $\text{C}_L$ are structurally homologous and highly conserved or conservatively replaced in immunoglobulin sequences.

The relation between idiotypic determinants, antigen combining site and hypervariable regions, is discussed in terms of the refined model.

In this paper we present the results of a preliminary crystallographic refinement and a list of atomic coordinates of Fab New. Some features of the refined structure are discussed in relation to the genetic control and physiological function of immunoglobulins.

It is generally accepted (see review in Ref. 1) that electron density maps calculated by multiple isomorphous replacement techniques contain significant errors which may lead to imprecise determination of structural details such as the location of amino acid side chain atoms, bond angles, $\phi$ and $\psi$ values, cis or trans character of proline residues, etc. This refinement project was undertaken with the aim of obtaining more accurate coordinates which can be applied to structural studies of other immunoglobulins and Fab·haptenc complexes (2). The starting atomic coordinates were those of the structure previously described (3, 4) obtained using multiple isomorphous heavy atom replacements. Since the complete refinement of the structure of Fab New is a complex undertaking, we present here initial results obtained after application of two consecutive refinement techniques. In the first step we have applied a "model building" procedure (5) in which the measured atomic coordinates were adjusted to impose standard bond lengths and bond angles. In a second step a "real space" procedure (6, 7) was used to optimize the correlation between the Fab New model and the multiple isomorphous replacement, electron density Fourier map.

The coordinates obtained by these procedures have provided an improved model which has been used to compare the tertiary structure of the homology subunits, to calculate interatomic contacts that define the quaternary structure of Fab, and to re-examine the conformation of the combining site.

METHODS

Measurement of Model Coordinates—Atomic coordinates were measured on the 2 Å (nominal) resolution model previously described (4). A two-pointer device was used for this purpose: a horizontal pointer (50 inches long) was brought to touch atom centers by displacing the device on the base of the model, adjusting the height of the pointer along a graduated scale (z coordinate) while a second, parallel, fixed pointer of equal length gave the $x$, $y$ coordinates on a grid at the base of the model. The use of a level and leveling screws at the base of the two-pointer device was essential for obtaining reproducible coordinates. While these measurements were made, the image of the model and the atom centers were

† The abbreviations used for immunoglobulins, their polypeptide chains, and fragments are as recommended in (1964) Bull. W H O 30, 447.
projected on the corresponding sections of the Fourier map using an optical comparator (8) to verify that their location and the coordinate values corresponded with the Fourier map.

**Model Building Procedure** —The set of measured coordinates for the 3185 non-hydrogen atoms in the structure provided the starting point for this procedure. In general, these coordinates are subject to errors due to measurement uncertainty and to mechanical deformations of the skeletal brass model. Consequently, the mathematical model building procedure of Diamond (5) programmed for a digital computer was used in order to impose standard bond lengths and bond angles in the model. The measured coordinates were used to provide a guide point for each (non-hydrogen) atom in the structure. Some conditions used in this part of the refinement process are given in Table I. All the varied angles are dihedral; the \( \tau = \tau_{\text{C3-C-C-C}} \) angle was allowed to vary since this condition gave a much closer correlation with input coordinates without introducing large distortions in the idealized, model-built geometry of the molecule. Residues for which the model-built coordinates differed considerably from the structure was removed. (11) The calculated, model-built geometry of the molecule.

**Real Spatial Refinement** — The model-built coordinates were used as a starting set for real space refinement (6). The 2.0 Å electron density map used for the automatic fitting of the atomic coordinates was calculated using multiple isomorphous replacement phases as described before (3, 4). The electron density function was calculated at intervals of 1/160 along \( x \) \((x = 111.43, 1/80\) along \( y \) \((y = 56.68\) Å), and 1/130 along \( z \) \((z = 90.30\) Å) in sections of constant \( \gamma \). A computer program incorporating a fast-Fourier transform algorithm was used for this calculation. Five cycles of real space refinement were carried out using the conditions defined in Table II.

The root mean square shift of the initial coordinates for all non-hydrogen atoms in the structure was 0.2 Å.

**RESULTS AND DISCUSSION**

The root mean square shifts of atomic coordinates after five cycles of real space refinement converged to an average value of 0.09 Å. The values after each cycle were: 0.46 Å, 0.20 Å, 0.13 Å, 0.10 Å, and 0.09 Å. The average shift of all atoms during real space refinement was 0.62 Å. In most parts of the molecule, the coordinates after refinement are in very good agreement with the features of the electron density map (Fig.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi ), ( \psi ), ( \chi )</td>
<td>2.0 Å</td>
<td>For comparison with other structures.</td>
</tr>
<tr>
<td>( \tau_{\text{C3-C-C-C}} )</td>
<td>2.0 Å</td>
<td>For comparison with other structures.</td>
</tr>
</tbody>
</table>

**TABLE I**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi ), ( \psi ), ( \chi )</td>
<td>2.0 Å</td>
<td>For comparison with other structures.</td>
</tr>
<tr>
<td>( \tau_{\text{C3-C-C-C}} )</td>
<td>2.0 Å</td>
<td>For comparison with other structures.</td>
</tr>
</tbody>
</table>

**TABLE II**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Conditions of real space refinement**a

| Zone length | 5 | |
| Margin width | 6 | |
| Fixed atomic radii | 1.4 Å | |
| Relative atomic weights | C, O, N, H: 12, 16, 14, 1 | |
| Relative softness of angular parameters that were allowed to vary: \( \phi \), \( \psi \), \( \chi \) | 4.0 | |
| Filler levels: | 3.2 | |

**Electron density map grid:** 111.43/160, 56.68/80, 90.30/130 along cell edges

**a** See Diamond (6, 7) for definition of terms.

**b** Nonlinear constraints were used to preserve chain continuity.

1. Many carbonyl groups of the main polypeptide chain can be reliably positioned (Fig. 2). In general, coordinates of atoms in regions of strong, well defined electron density converged rapidly in the first cycle of refinement and moved very little in subsequent cycles. Atoms in regions of lower density converged more slowly. A few residues in poorly defined regions of low density showed little convergence, although their total shift from the starting coordinates was small. The movement of main chain atoms tended to be smaller than those of side chains, presumably due to their better defined electron density and to greater constraints on their positions.

The progress of refinement was checked after each cycle by inspecting the fit of atoms whose shifts were substantially greater than the average. Some side chain groups which had been shifted by the Diamond model building procedure were moved back to their original positions by real space refinement. In the fifth cycle of refinement, an average shift greater than 1 Å occurred for three consecutive amino acid residues (Gly-166, Val-167, and His-168) in the C11 region. Inspection of the position of these atoms showed that they had moved to a conformation that appears to be in better agreement with the electron density map than the original model. The coordinates for all atoms of Fab New after refinement, given in the "Appendix," are filed with the Protein Data Bank at Brookhaven National Laboratory. No major features of the map remain unexplained, although a number of possible solvent molecules are found on the surface of the molecule. The conventional \( R \) factor, based on \( F_{o} \), obtained with the coordinates in "Appendix" and an overall temperature factor (\( B = 18.0 \) Å), is \( R = 0.46 \). This value is reasonable given the refinement approximations outlined in Table II and the fact that no solvent atoms were included in the model. Further refinement using observed structure factors and calculated phases is currently underway.

The \( S-S \) distances in the five Fab disulfide bonds were allowed to vary without constraints. At the end of the refinement these distances were found to be: \( V_{L} \), 2.0 Å; \( V_{L} \), 2.46 Å; \( C_{11} \), 2.30 Å; \( C_{L} \), 2.30 Å; \( C_{L} \), 2.45 Å.

**Ramachandran Plots**

The Ramachandran plots of the \( V_{L} \) and \( C_{L} \) homology sub-
Fig. 1 (upper). Stereo view of some main chain and side chain groups of Fab New after five cycles of real space refinement superimposed on the corresponding electron density of the multiple isomorphous replacement 2 Å resolution map.

Fig. 2 (lower). Stereo view of some amino acid residues of Fab New superimposed on the corresponding density of the multiple isomorphous replacement map. Carbonyl groups of the main polypeptide chain are clearly seen.

Fig. 3. Ramachandran plots of the VL and CL homology subunits of Fab New. The distribution of $\phi$, $\psi$ angles indicates the predominant antiparallel $\beta$-pleated sheet structure of the subunits. Glycine residues are indicated by $\circ$.

units (see Fig. 3) show a distribution of $\phi$, $\psi$ angles which corresponds with the predominant antiparallel $\beta$-pleated sheet structure present in $V_H$, $V_L$, $C_H$, and $C_L$. As observed in several other protein structures, the $\phi$, $\psi$ angles for glycine residues are scattered in the plot, frequently appearing in nonallowed regions for L-amino acids. Several other residues which occur outside areas of allowed conformation are found in bends of the polypeptide chains; it is expected that further
refinement will improve the angular values for these residues.

**Structural Comparison of Homology Subunits**

The structural homology of $V_H$, $V_L$, $C_{H1}$, and $C_L$ has been described before (3, 4). A quantitative analysis of this homology using the method of Rao and Rossmann (9) is presented here. A similar analysis has been made by Richardson et al. (10) comparing the structures of superoxide dismutase and the murine Fab McPC 603 fragment.

Initial matrices relating the $C_\alpha$ coordinates of the homology subunits were obtained from a small number of structurally equivalent amino acids. The number of equivalents was then extended by an automatic search for stretches of polypeptide chain for which the distances between putatively equivalent $C_\alpha$ was smaller than 3.8 A. Based on the extended equivalences new matrices were calculated and the process was iterated until no changes in equivalences were observed.

A summary of the results is presented in Table III which lists the number of $C_\alpha$ occurring at distances of less than 1.5 A and less than 3.0 A, for all the six possible pairings of subunits which were superimposed and compared by this process. The average value of the minimum base change necessary to exchange the codons of the structurally equivalent amino acids is also given in Table III.

As can be seen in Table III, there is an even closer structural homology between $V_H$, $V_L$, $C_{H1}$, and $C_L$ in Fab New than that observed for McPC 603 Fab (10), probably reflecting the higher resolution of the Fab New model. Presumably the $C_\alpha$ distances given in Table III could become smaller with further crystallographic refinement. The number of $C_\alpha$ which super-impose with distances shorter than 1.5 A and 3.0 A is larger when comparing $V_H$ to $V_L$ and $C_{H1}$ to $C_L$. Also, there is good (inverse) correlation between the number of $C_\alpha$ that are structurally equivalent and the average minimum base change per codon. Furthermore, when a restrictive condition for structural equivalence is imposed ($d_{Ca-Ca} \leq 1.5$ A) the average base change per codon becomes smaller, reflecting a higher degree of conservation of amino acid sequences.

It should be emphasized here that amino acid sequence information is not used in the quantitative three-dimensional alignment procedure described above. However, this procedure leads to amino acid sequence alignments that clearly reflect the well established homologies between the $V_H$ and $V_L$, and between the $C_{H1}$ and $C_L$ regions of immunoglobulins (see Figs. 4 and 5). The closest sequence similarity in Fab New occurs between $C_{H1}(\gamma)$ and $C_L(\lambda)$, although, as shown in Table III, the structural similarity between $V_H$ and $V_L$ is close to that between $C_{H1}$ and $C_L$ (see Figs. 6, 7, and 8). In addition, Table III shows that there is considerable homology between the V and C regions. These results can be interpreted to indicate that all the homology regions contain a basic core of amino acid residues with highly preserved three-dimensional structure. The chemical nature of these residues is also preserved as indicated by the correspondingly lower values of the average base change per codon. As stated before (3) these findings strongly support the postulate (15) of a gene duplication mechanism which gave rise to the different homology regions of immunoglobulins.

The $C_\alpha$ of the homologous sequences, -Phe-Gly-Gly-Gly- (99 to 102) in $V_H$ and -Trp-Gly-Gln-Gly- (107 to 110) in $V_H$, can be closely superimposed as can $C_\alpha$ atoms immediately preceding and following those residues. This conserved conformation gives no evidence supporting the postulate (16) that the glycine residues could serve as a pivot to allow for optimal contacts between an antibody and its ligands. An alternative explanation for these constant, homologous $V_H$ and $V_L$ sequences has been proposed (4) in terms of intersubunit ($V_H$ to $V_L$, see below) and intrasubunit contacts.

The limits between the V and C homology regions can be defined from the model. The sequence -Val-Ser-Ser- (115 to 117, Fig. 4) which is shared by $\gamma$ and $\mu$ human H chains marks the COOH terminus of $V_H$, and following a sharp bend in the polypeptide chain the sequence -Ala-Ser-Thr- (118 to 120) marks the NH$_2$ terminus of $C_L$. The sequence -Thr-Val-Leu- (106 to 108) corresponds to the COOH terminus of $V_L$ and the residues -Gly-Pro-Lys- (110 to 112) constitute the NH$_2$ terminus of $C_L$. Thus, in the three-dimensional model Arg 109 (usually assigned to $V_L$) could be considered either as the end of $V_L$ or as the beginning residue of $C_L$. By the structural alignment described here however, Arg 109 can be properly considered as the COOH terminus of $V_L$.

In agreement with the Gm(4-) serological specificity of IgG New (17), a lysine residue is placed at position 214 in $C_{H1}$. This residue, corresponding to the Gm(17) allotype provides a better fit with the Fourier map than an arginine residue (which correlates with Gm(4)) at that position. This interpretation is reinforced by the Gm(1+) serological specificity of IgG New (17) which has been verified by amino acid sequence analysis (12).

**Quaternary Structure**

Contacts between Homology Subunits — The closer contacts between the homology subunits of Fab New are diagrammatically represented in Fig. 9 by lines joining $C_\alpha$ atoms separated by a distance of 8 A or less. This figure provides a description of regions of $V_H$, $V_L$, $C_{H1}$, and $C_L$ in which there are higher density of contacts. Inspection of Figs. 8 and 9 indicates that the interactions between $V_H$ and $V_L$ and between $C_{H1}$ and $C_L$ are more extensive than those between $V_H$ and $C_{H1}$ and those between $V_L$ and $C_L$. The fact that the $V_H$ and $C_{H1}$ subunits (whose major axes make an angle smaller than 90°) interact more extensively than $V_L$ and $C_L$ (whose major axes makes an angle larger than 90°) is also reflected in Figs. 8 and 9.

Intersubunit contacts between side chain and main chain atoms situated at a distance not larger than 1.2 times their van der Waals radii are given in Table IV. This table lists contacting residues and the number of close contacts that atoms from a given residue make with atoms of other residues. Evidently, amino acids with larger side chains have a potential to make more contacts with other amino acids, thus for example, $V_H$ Trp 107 makes 29 intersubunit contacts, Trp 47...
FIG. 4. Amino acid sequences of the \( V_L \), \( C_L \), \( V_H \), and \( C_H \) homology regions of Fab New aligned by comparison of their three-dimensional structures. \(-\-\-\) indicate gaps introduced to maximize alignment of the three-dimensional structures. \(*\) indicate deletions in the \( V_L \) sequence. See Ref. 11 for the \( V_L \) and \( C_L \) sequences, Ref. 12 for \( V_H \), and Ref. 13 for \( C_H \). Abbreviations for amino acids are as given in Ref. 14.

makes 28 contacts and Arg 43 makes 24 contacts.

The contacts between \( V_H \) and \( V_L \) are of particular interest in view of the fact that different H and L immunoglobulin chains can form structurally viable pairs. Three types of \( V_H - V_L \) contacts will be considered in this discussion: first, the contacts which are at the core of the contacting region, made by residues which are invariant or semi-invariant in \( V_H \) and \( V_L \) sequences; second, the contacts made by invariant or semi-
Structural Refinement of Fab New

Fig. 6. Stereo pair drawings of the α carbon backbones of the $V_L$ (top) and $V_H$ (bottom) subunits. The subunits are viewed here in similar orientations.

Fig. 7. Stereo pair drawing of the α carbon backbones of the $C_\alpha$ (top) and $C_{\beta1}$ (bottom) subunits viewed in similar orientations.
invariant residues with hypervariable residues; and finally, those made between hypervariable residues.

The core of the V_{H}-V_{L} contacting region can be described as determined by residues Val 37, Gln 39, Leu 45, Tyr 94, and Trp 107 in V_{H} and by residues Tyr 35, Gln 37, Ala 42, Pro 43, Tyr 86, and Phe 99 in V_{L}. These residues are structurally homologous with the exception that V_{L} Ala 42 has no clear correspondence in V_{H} due to a structural "insertion" (see Fig. 4). These homologous V_{H} and V_{L} residues make numerous contacts with each other (about 50% of those listed in Table IV) or with other, nonhypervariable residues. The rings of Trp 107 (V_{H} and Pro 43 (V_{L}), at the center of the V_{H}-V_{L} contacting region, are nearly parallel and stacked on each other. The contact residues listed above are invariant or are replaced by homologous residues in V_{H} (κ and λ) and V_{L} sequences from different animal species. For example, Tyr 35, Gln 37, Pro 43, and Phe 99 appear constant in human L chains (κ or λ), and Gln 39, Tyr 94 (replaced by Phe in a very few cases), and Trp 107 (replaced by Phe or Tyr in a very few cases) appear nearly constant in human H chains. Ala 42, Tyr 86 in V_{H} and Val 37 in V_{L} are more frequently replaced by homologous residues: Ser 42, Phe 86, and Ile 37. The invariant or nearly invariant nature of these residues of the main V_{H} V_{L} contacting area provides a structural basis (together with interactions between C_{H} and C_{L}, see below) for the property of different H and L chains to recombine into new immunoglobulin molecules (see References 18, 19, and in particular 20, for a recent review and experimental data on this topic).

A second type of contact listed in Table IV is made between constant or nonhypervariable residues and hypervariable residues. For example, the side chain atoms of V_{H} Trp 47, a constant residue in human, mouse, guinea pig, and in most rabbit immunoglobulin sequences, make close contacts with Ser 93, Leu 94, and Arg 95 in the third hypervariable region of V_{L}. However, a large number of these contacts involve the peptide chain atoms of the V_{L} residues. Replacements in the V_{L} side chains will not necessarily alter the nature of these contacts. Similar contacts appear to be made by V_{L} Leu 45 (invariant or semi-invariant in human L chains) with the peptide chain at V_{H} hypervariable position 104. Contacts of this type could also be made from V_{L} Tyr 35 to the peptide chain atoms of the fourth hypervariable region of V_{H} in chains of different length than V_{H} New.

The third type of contact to be discussed here is that made between hypervariable residues, such as those made between V_{H} Asn 98 and V_{L} Arg 95. These contacts are more difficult to evaluate in general terms (a) because the location of some of the residues involved might be changed by further refinement to a larger extent than those of most other residues in the sequence, and (b) because it is possible that in other immunoglobulins, replacements by different amino acid side chains at these positions could be accommodated by small displacements of the hypervariable peptide loops. These "idiotypic" interactions are consequently more difficult to assess. However, they could perhaps explain the preferred reassociation observed between complementary H and L chains derived from a single immunoglobulin molecule (20). Most of the contacts discussed above consist of van der Waals interactions between hydrophobic side chains. However, a few hydrogen bonds can be indicated: V_{H} Gln 39 to V_{L} Gln 37, and V_{H} Asn 98 to V_{L} Tyr 90 and/or V_{L} Arg 95. Also, an ion pair is formed between V_{H} Arg 43 and V_{L} Asp 84.

In the Fab New model the contacts between V_{H} and V_{L} are very close (Table IV), giving rise to a compact dimer. No haptons or even solvent molecules can be accommodated between V_{H} and V_{L} beyond the combining site, a situation which is different from that described for an L-chain dimer (22).

As shown in Table IV the interactions between C_{H} and C_{L} are extensive. The core of the contact area between C_{H} and C_{L} is defined by C_{H} residues Leu 128, Ala 129, Gln 143, and Leu 145 and the structurally homologous C_{L} residues Phe 120, Pro 121, Val 135, and Leu 137. These residues appear to be invariant or nearly invariant in the H and L sequences from different animal species. Most of the other contact residues such as C_{H}: Phe 126, Pro 127, Thr 139, Lys 147, Phe 170, Pro 171, Val 173, Gln 175, Ser 181, Val 185, Lys 218, and C_{L}: Thr 118, Ser 123, Gln 125, Glu 126, Lys 131, Thr 133, Thr 164, Ser 177, Tyr 179, Lys 206, are also invariant or replaced by homologous residues in the immunoglobulin chains from different animal species. In the contact area the central location of C_{H} Leu 128 and C_{L} Phe 120 is reflected in the large number of contacts (20 contacts) they make with each other and with many other residues (see Table IV). As pointed out by Novotny and Franek (22) the amino acid sequence of the four-stranded β-sheet is more conserved than the rest of the α-helices (or genealogic tree) of distorted evolutionary distances. This observation can be analyzed in terms of the structural model presented here as follows. The four-stranded β-sheets of C_{H} and C_{L} contain side chains which make intrasubunit contacts and in particular, they contain all or nearly all of the contact residues between C_{H} and C_{L} (discussed above). Evidently, mutational events leading to amino acid replacements at these positions would have to occur in a complementary pattern in both C_{H} and C_{L} in order to preserve tertiary and quaternary immunoglobulin structure, and consequently they would be expected to occur at a slower rate than mutations in other regions of C_{H} and C_{L}.

As can be seen in Table IV the region immediately preceding the interchain disulfide bond does not provide close contacts between C_{H} and C_{L}. In addition, the two strands of
polyypeptide chain that come together at the interchain disulfide bond do not closely interact with the rest of CH1 or CL. This region can be described as having a loose conformation, with a lower electron density in the Fourier map. These structural features are in agreement with the notion of segmental flexibility residing around this part of the immunoglobulin structure and in the immediately adjacent hinge region of the H chain.

**Hyervariable Regions, Idiotypes, and Combining Site**

The results of several experimental approaches (see Chapter II...
in Ref. 23 for a comprehensive review) strongly suggested that hypervariable residues in the amino acid sequences of H and L chains, idiotypic determinants, and combining site residues of an antibody molecule partially overlap. X-ray crystallographic analyses (2-4, 21, 24, 25) provided an unequivocal confirmation of these conclusions and three-dimensional models of different immunoglobulin molecules in which the structural bases of these operational concepts could be further defined.

The hypervariable regions of the H and L chains are located on exposed bends of the polypeptide chains, in regions which can be folded without a defined secondary structure (such as a regular β-bend) and which consequently appear capable of accommodating sequence variations without major structural constraints. From analyses of the structural bases of the Inv allotypic antigenic determinants of human κ chains (26) and of the Oz, Kern isotypic antigenic determinants of human λ chains (3, 27, 28) it is known that a variation in a single exposed immunoglobulin amino acid side chain is sufficient to generate a specifically recognizable antigenic determinant. There are regions of immunoglobulin molecules such as those around residues 85 to 90 in human H chains (29) or around residues 1 to 3 in rabbit L chains (30) where such specific unique antigenic determinants (idiotypes) could arise. However, since most of the sequence variation occurs at an exposed end, where VH and VL hypervariable regions join in determining the conformation of the combining site, most of the idiotypic determinants would be expected to occur at this site or immediately adjacent to it (see Fig. 10). Two questions arise: (a) are all antigen contact (complementarity-determining) residues idiotypic antigenic determinants?; (b) do all idiotypic determinants contribute directly to antigen binding? These questions have been explored before in some ligand-antibody systems (see review in Ref. 23). For example, idiotypic determinants of anti-phosphorylcholine antibodies from inbred mice can be divided into site-associated determinants and other idiotypic determinants which are not site-associated (31, 32). In order to provide answers to these questions on the basis of the three-dimensional model of an immunoglobulin presented in this paper, it is necessary to make some assumptions which will be briefly reviewed in the following paragraph.

The first assumption is, necessarily, that the conformation of an Fab fragment in the crystalline state is the same as in the parent immunoglobulin in solution. Also, it is assumed that the combining site is limited to the crevice or cavity delineated by the VH and VL hypervariable regions as previously described (3). In addition, it will be assumed that antigen binding will not induce major conformational changes at the combining site although by analogy with enzyme-ligand systems (see for example Ref. 33) smaller movements of amino acid side chains and polypeptide backbone chains can be expected and have to be allowed for. Since the definition of idiotypic determinants relies on serological procedures, the mode of interaction between the recognized immunoglobulin and the recognizing anti-idiotypic antibody is important. Although no suitable structural model is available to describe them they can be expected to be complex. A minimal assumption that can be made is that all amino acid side chains which are accessible to solvent are potential antigenic determinants.

![Fig. 10. View of some of the amino acid residues at the combining site of IgG New.](image-url)
that can be recognized by an anti-idiotypic antibody.

Within the assumptions outlined above we conclude that most (not necessarily all) complementarity-determining residues are idiotypic determinants. The set of possible idiotypic determinants (exposed amino acid side chains) is larger than the set of complementarity-determining residues. The number of hypervariable region residues is larger than the number of idiotypic determinants since some of the hypervariable residues are idiotypic determinants. The set of possible idiotypic determinants since some of the hypervariable region residues is larger than the number of determining residues in the combining site, the following amino acid residues which could also make contacts include Thr 56 and Asp 98. Atoms along the main peptide chain of the residues listed above and of several other residues will also add to contacts with antigens or haptens. VH, Tyr 90 and VL, Trp 47 are fairly conserved residues in VH and VL sequences, however, they line up the bottom of the combining site and should be expected to contribute to antigen contacts. In other immunoglobulin molecules having different lengths of polypeptide chains at the hypervariable regions, the complementarity-determining residues will not be expected to occur at the same positions as those listed above for IgG New. All the residues listed above, except VH, Trp 47 and VL, Tyr 90, are possible idiotypic determinants.

It should be emphasized here that the list of complementarity-determining residues discussed above is based on the characterization of the combining site as a cavity or pocket between VH and VL hypervariable regions. This characterization has been made on the basis of its conformation described as a "shallow groove" for IgG New (3) and on the basis of the binding of ligands that can be considered as haptens (or cross-reacting haptens) with defined stoichiometry and association constant (2, 24). Thus, there is no doubt that a central region of the combining site has been identified and characterized. However, no suitable structural models have yet been obtained to exclude the possibility that the combining site extends to regions around this central cavity. For example, there is a "side" cavity or pocket in IgG New, surrounded by VH residues 2, 3, 25, 27, 27a, 91, 92, 93, 94, and 98, and VL residues 57 through 61. It could be assumed that this is also part of the combining site or a compartment of it. In support of this view (a) hypervariable region residues modulate the conformation of this region to make it unique for every immunoglobulin molecule, (b) affinity labeling experiments with anti-2,4-dinitrophenyl and anti-arsenate guinea pig antibodies (34) have tagged VL residues Tyr 60 and Lys 59, respectively, which occur in this side site. A symmetrical side site is bound by residues of the second hypervariable region of VH, and by hypervariable regions around positions 30 and 95 in VH. As in the symmetrical side site described above, constant or less variable residues of VH and VL also contribute to the conformation of this subsite. If this view of an extended combining site is correct, it will have implications such as (a) more amino acid side chains in the hypervariable regions contribute to the definition of antigen binding specificity; thus, sequence hypervariability would have a bigger influence in the physiological process of antigen recognition and binding; (b) antibodies are highly reactive proteins utilizing a great contact surface for antigen recognition; (c) because of the sequence diversity and the varied conformational components of such an extended site it appears inescapable that a given antibody molecule should be able to react with different ligands; substances such as 2,4-dinitrophenyl which are highly reactive toward proteins would be bound by many different antibody molecules, thus an anti-2,4-dinitrophenyl response would have the potential to be extremely heterogeneous.

It should also be mentioned here that if the antigen combining site is extended as described above, amino acid variations in the BALB/c murine λ chain system (35) will contribute to antigen binding, at variance with our previous analysis (36). Positions 25, 91, and 94 (numbered following the Vh sequence given in Fig. 4) in which variations have been detected in the murine λ chain sequences contribute to the conformation of the side cavity or pocket described above. Irrespective of this putative antigen binding role, these chain sequence variations, which have been interpreted as products of somatic mutation mechanism (35), are likely to give rise to new

---

**TABLE IV**

**Intersubunit contacts**

<table>
<thead>
<tr>
<th>Vh</th>
<th>Vl</th>
<th>No. of contacts</th>
<th>Cαl</th>
<th>Cβ</th>
<th>No. of contacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val 37</td>
<td>Phe 99</td>
<td>1</td>
<td>Phe 126</td>
<td>Glu 126</td>
<td>14</td>
</tr>
<tr>
<td>Gin 39</td>
<td>Gin 37</td>
<td>5</td>
<td>Phe 126</td>
<td>Glu 125</td>
<td>1</td>
</tr>
<tr>
<td>Arg 43</td>
<td>Asp 84</td>
<td>14</td>
<td>Phe 126</td>
<td>Ser 123</td>
<td>3</td>
</tr>
<tr>
<td>Gin 43</td>
<td>Tyr 86</td>
<td>10</td>
<td>Leu 128</td>
<td>Phe 120</td>
<td>20</td>
</tr>
<tr>
<td>Gin 43</td>
<td>Gin 37</td>
<td>3</td>
<td>Leu 128</td>
<td>Val 135</td>
<td>2</td>
</tr>
<tr>
<td>Leu 45</td>
<td>Tyr 86</td>
<td>4</td>
<td>Leu 128</td>
<td>Pro 121</td>
<td>1</td>
</tr>
<tr>
<td>Leu 45</td>
<td>Phe 99</td>
<td>2</td>
<td>Ala 129</td>
<td>Phe 120</td>
<td>8</td>
</tr>
<tr>
<td>Glu 46</td>
<td>Phe 99</td>
<td>3</td>
<td>Ala 129</td>
<td>Pro 121</td>
<td>2</td>
</tr>
<tr>
<td>Trp 47</td>
<td>Arg 95</td>
<td>19</td>
<td>Lys 133</td>
<td>Glu 212</td>
<td>1</td>
</tr>
<tr>
<td>Trp 47</td>
<td>Leu 94</td>
<td>8</td>
<td>Thr 139</td>
<td>Thr 118</td>
<td>3</td>
</tr>
<tr>
<td>Trp 47</td>
<td>Ser 93</td>
<td>1</td>
<td>Thr 139</td>
<td>Lys 206</td>
<td>2</td>
</tr>
<tr>
<td>Asp 58</td>
<td>Ser 93</td>
<td>2</td>
<td>Ala 141</td>
<td>Phe 120</td>
<td>6</td>
</tr>
<tr>
<td>Asp 60</td>
<td>Leu 94</td>
<td>3</td>
<td>Leu 142</td>
<td>Phe 120</td>
<td>4</td>
</tr>
<tr>
<td>Thr 61</td>
<td>Leu 94</td>
<td>1</td>
<td>Gly 143</td>
<td>Phe 120</td>
<td>5</td>
</tr>
<tr>
<td>Tyr 94</td>
<td>Ala 42</td>
<td>5</td>
<td>Leu 145</td>
<td>Tyr 179</td>
<td>1</td>
</tr>
<tr>
<td>Asn 98</td>
<td>Arg 95</td>
<td>8</td>
<td>Leu 145</td>
<td>Val 135</td>
<td>1</td>
</tr>
<tr>
<td>Leu 99</td>
<td>Arg 95</td>
<td>3</td>
<td>Lys 147</td>
<td>Glu 126</td>
<td>1</td>
</tr>
<tr>
<td>Ala 101</td>
<td>Tyr 90</td>
<td>6</td>
<td>Lys 147</td>
<td>Lys 131</td>
<td>2</td>
</tr>
<tr>
<td>Ala 101</td>
<td>His 31</td>
<td>7</td>
<td>Lys 147</td>
<td>Thr 133</td>
<td>2</td>
</tr>
<tr>
<td>Ala 101</td>
<td>Lys 33</td>
<td>2</td>
<td>Phe 170</td>
<td>Leu 137</td>
<td>10</td>
</tr>
<tr>
<td>Gly 102</td>
<td>Lys 33</td>
<td>6</td>
<td>Phe 170</td>
<td>Ile 138</td>
<td>4</td>
</tr>
<tr>
<td>Ile 104</td>
<td>Tyr 35</td>
<td>3</td>
<td>Phe 170</td>
<td>Ser 177</td>
<td>4</td>
</tr>
<tr>
<td>Ile 104</td>
<td>Gin 88</td>
<td>3</td>
<td>Pro 171</td>
<td>Ser 167</td>
<td>2</td>
</tr>
<tr>
<td>Ile 104</td>
<td>Leu 45</td>
<td>2</td>
<td>Pro 171</td>
<td>Ala 175</td>
<td>1</td>
</tr>
<tr>
<td>Trp 107</td>
<td>Pro 43</td>
<td>21</td>
<td>Val 173</td>
<td>Tyr 179</td>
<td>6</td>
</tr>
<tr>
<td>Trp 107</td>
<td>Ala 42</td>
<td>2</td>
<td>Glu 175</td>
<td>Glu 162</td>
<td>7</td>
</tr>
<tr>
<td>Trp 107</td>
<td>Phe 99</td>
<td>4</td>
<td>Ser 176</td>
<td>Glu 162</td>
<td>8</td>
</tr>
<tr>
<td>Trp 107</td>
<td>Tyr 35</td>
<td>2</td>
<td>Leu 182</td>
<td>Tyr 179</td>
<td>2</td>
</tr>
<tr>
<td>Cαl</td>
<td>Vh</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala 118</td>
<td>Leu 11</td>
<td>2</td>
<td>Val 185</td>
<td>Leu 137</td>
<td>3</td>
</tr>
<tr>
<td>Ser 119</td>
<td>Leu 11</td>
<td>2</td>
<td>Val 185</td>
<td>Phe 120</td>
<td>3</td>
</tr>
<tr>
<td>Thr 120</td>
<td>Leu 11</td>
<td>3</td>
<td>Lys 218</td>
<td>Cys 213</td>
<td>2</td>
</tr>
<tr>
<td>Phe 150</td>
<td>Leu 11</td>
<td>3</td>
<td>Ser 219</td>
<td>Glu 212</td>
<td>2</td>
</tr>
<tr>
<td>Phe 150</td>
<td>Thr 114</td>
<td>1</td>
<td>Ser 219</td>
<td>Cys 213</td>
<td>6</td>
</tr>
<tr>
<td>Pro 151</td>
<td>Leu 11</td>
<td>2</td>
<td>Cys 220</td>
<td>Cys 213</td>
<td>7</td>
</tr>
<tr>
<td>Pro 151</td>
<td>Thr 114</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu 152</td>
<td>Leu 112</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro 153</td>
<td>Leu 112</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cβ</td>
<td>Vl</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gin 110</td>
<td>Glu 82</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys 168</td>
<td>Pro 39</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn 172</td>
<td>Glu 82</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
idiotypic specificities and to an altered regulatory idotype network (37).

In order to further map and characterize the combining site it might be necessary to investigate chemical interactions in other hapten-antibody or antigen-antibody systems. At present this appears to be a task of considerable magnitude.

Acknowledgments—We thank Dr. H. Bernstein and the staff of the CRYSNET project for their assistance in implementing programs at Brookhaven National Laboratories.

REFERENCES


Structural Refinement of Fab New

595

Page 14