

The Three-Dimensional Structure of the Fab' Fragment of a Human Myeloma Immunoglobulin at 2.0-Å Resolution

(β -sheets/sequence alignments/hypervariable regions/active site)

R. J. POLJAK, L. M. AMZEL, B. L. CHEN, R. P. PHIZACKERLEY, AND F. SAUL

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

Communicated by Max F. Perutz, June 14, 1974

ABSTRACT The structural analysis of the Fab' fragment of human myeloma immunoglobulin IgG1(λ) New has been extended to a nominal resolution of 2.0 Å. Each of the structural subunits corresponding to the variable and to the constant homology regions of the light and heavy chains contains two irregular β -sheets which are roughly parallel to each other and surround a tightly packed interior of hydrophobic side chains. About 50-60% of the amino-acid residues are included in β -pleated sheets. Sequence alignments between the homology regions of Fab' New obtained by comparison of their three-dimensional structures are given. Some of the sequence variations observed in light and heavy chains and the role of the regions of hypervariable sequence in defining the size and shape of the active site of different immunoglobulin molecules are discussed on the basis of the three-dimensional model of Fab' New.

In a previous paper (2) we described the three-dimensional structure of the Fab' fragment of human myeloma immunoglobulin IgG New based on the interpretation of an electron density map at 2.8-Å resolution. The molecule was found to consist of four globular subunits which correspond to the variable (V_L , V_H) and constant (C_L and C_{H1}) homology regions of the light (L) and heavy (H) polypeptide chains, arranged in tetrahedral configuration. The homology subunits*, which closely resemble each other, share a basic pattern of polypeptide chain folding. This basic pattern (immunoglobulin fold) and an additional loop of polypeptide chain describe the more complex folding of the variable subunits. The additional loop in the V_L subunit of IgG New is shortened by a deletion of seven amino acids which is unique to this L chain. The regions of hypervariable sequence in the L and H chains were found to occur in close spatial proximity, at one end of the molecule. Two reports dealing with the three-dimensional structure of human immunoglobulin L chains have been published (3, 4). In this paper we extend the structural analysis of Fab' New to a nominal resolution of 2.0 Å and continue the description and discussion of its three-dimensional structure.

METHODS

Preliminary measurement of 13,000 x-ray reflections with spacings ranging from 2.8 to 2.0 Å was carried out on native Fab' crystals. About 2000 of these reflections had intensities

Abbreviations: The nomenclature of immunoglobulins, their chains and fragments is as recommended by the World Health Organization (1); Dnp, 2,4-dinitrophenyl.

* A homology subunit of an immunoglobulin is defined as the globular unit of three-dimensional structure containing the amino-acid sequence of a homology region.

significantly above background scattering and were selected for intensity measurements with techniques previously outlined (5). Procedures for the preparation of heavy atom derivatives and phase determination have been previously given (5, 6). Three heavy atom compounds, phenylmercuric acetate, diacetoxymethyl dipropylene dioxide and 1,4-diacetoxymethyl-2,3-dimethoxybutane, were used to obtain isomorphous replacements to extend the phase information from 2.8- to 2.0-Å resolution. The average figure of merit obtained for the 2000 reflections with spacings between 2.8 and 2.0 Å that were phased was 0.69. A 2.0-Å resolution Fourier map was calculated including a total of 12,000 reflections with an average figure of merit of 0.76. Based on this map a model was constructed using an optical comparator (7) and Kendrew skeletal models.

Purified Fab New was digested with CNBr and the resulting fragments were separated by gel filtration on Sephadex G-100 columns. The amino-acid sequences of some of the tryptic and chymotryptic peptides of the CNBr fragments belonging to the V_H region were determined by procedures previously outlined (8). These peptides were aligned by homology with other human H chains, by sequence overlaps, and by information derived from the Fourier map.

RESULTS AND DISCUSSION

The 2.0-Å map shows increased resolution with respect to the previous 2.8-Å Fourier map even though only 2000 additional reflections were included. The amino-acid sequences of the C_L , C_{H1} , and V_L regions and the tentative sequence of V_H (see Fig. 1) correspond very closely to the features of the 2.0-Å Fourier map. The positions of most of the carbonyl groups of the polypeptide chain can be easily identified in the electron density map. Careful placing of the carbonyl and α -imino groups invariably maximized features of secondary structure such as hydrogen bonding between adjacent stretches of polypeptide chain. The V_L , V_H , C_L , and C_{H1} homology subunits are predominantly folded in β -pleated sheet conformation. The C_L homology subunit, for example, consists of a β -pleated sheet made up by four hydrogen-bonded antiparallel chains (residues 116-120, 132-140, 160-169, and 173-182) and another β -pleated sheet containing three antiparallel chains (residues 147-151, 193-199, and 202-208, see Fig. 2). These two twisted and roughly parallel sheets surround a tightly packed interior of hydrophobic side chains including the intrachain disulfide bond which links the two sheets in a direction approximately perpendicular to their planes. About 60% of the C_L residues are included in these

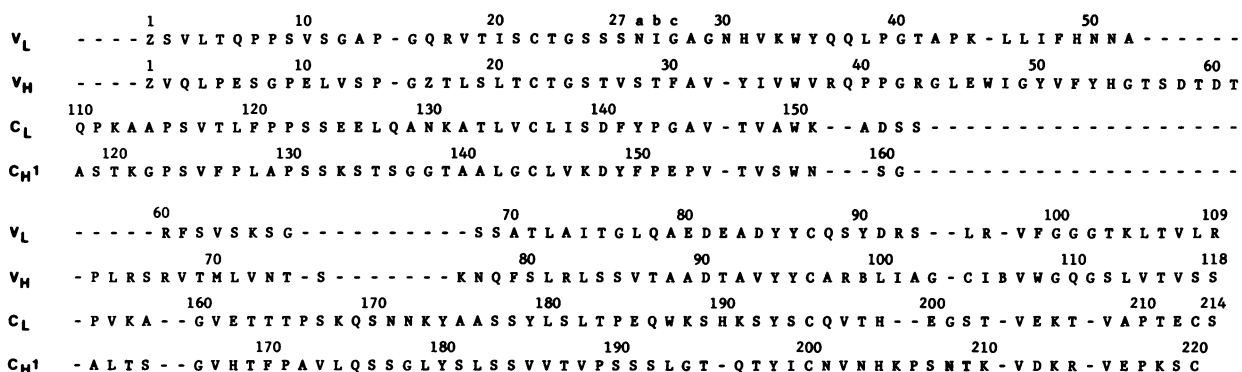


FIG. 1. Amino-acid sequences of the V_L, C_H, V_H, and C_H1 homology regions of Fab New, aligned by comparison of their three-dimensional structures (see *text*). Parts of the tentative V_H sequence given here were obtained by Drs. Y. Nakashima and W. Konigsberg (personal communication). The C_H1 sequence is taken from ref. 9. Abbreviations are from ref. 10.

two β -pleated sheets. The homologous subunits C_H1, V_L, and V_H (see Figs. 2 and 3) can be described in similar terms. In addition to the features of secondary structure just discussed, V_L residues 26–27c appear in a helical conformation close to that of a π -helix (11). Also, V_L residues 78–82, and the homologous V_H residues 87–91, fold with the conformation of a distorted π or α -helix. In general, the precise conformation of amino-acid residues that do not belong to the regions of secondary structure discussed above is more difficult to define, especially when they correspond to regions of lower electron density on the surface of the molecule.

Amino-acid sequence comparisons have been extensively used in several laboratories to align the primary structures of

different immunoglobulin molecules and their homology regions. The alignment of V_L and V_H sequences with those of the C_L and C_H homology regions is less straightforward than alignments between V_L and V_H or between C_L and C_H regions. This problem can now be approached by matching their similar three dimensional structures and aligning residues that occupy an identical or similar position in the constant pattern of the tertiary structure of the homology subunits. Alignments obtained by this procedure, which is independent of amino-acid sequence homologies, are shown in Fig. 1. Where homologies have been detected by sequence comparison, it can be safely assumed that the three-dimensional structure of the homologous polypeptide chains will be very

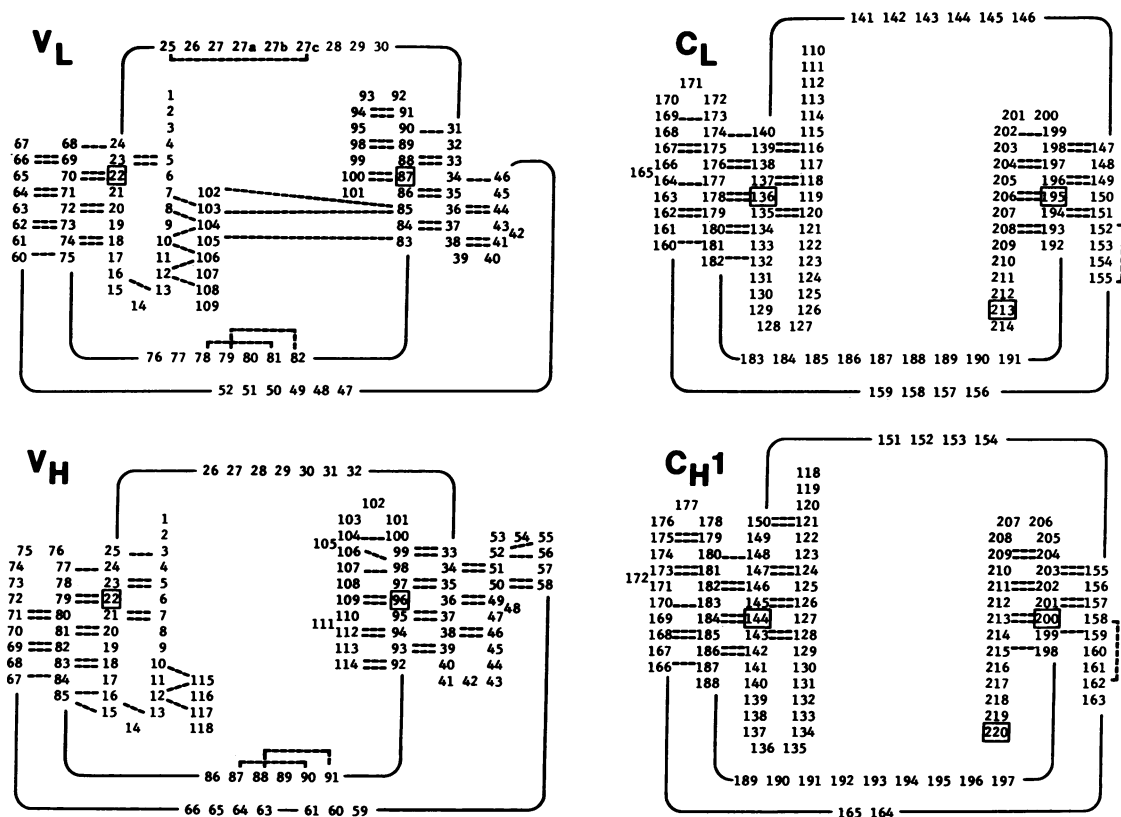


FIG. 2. Diagram of the hydrogen bonds (broken lines) between main chain atoms for V_L, C_L, V_H, and C_H1. The hydrogen-bonded clusters correspond to the two β -sheets of each subunit (see *text*). Cysteine residues that participate in the intrachain disulfide bonds linking the two β -sheets are enclosed in rectangles; C_L residue 213 and C_H1 residue 220 form the interchain disulfide bond and are also enclosed in rectangles. Numbers refer to residues as given in Fig. 1.

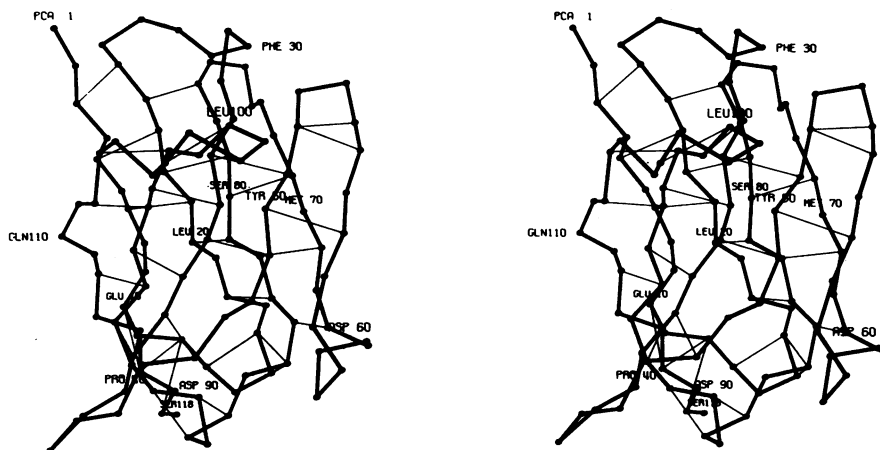


FIG. 3. Stereo pair drawing of the α -carbon backbone of the V_H structure. Thin lines connect the α -carbon atoms of residues that are hydrogen-bonded; see Fig. 2. PCA is pyrrolidone carboxylic acid.

similar. Since the amino-acid sequence of β_2 -microglobulin is highly homologous to that of C_L and C_H1 (12), it is reasonable to conclude that it shares the basic three dimensional structure or immunoglobulin fold of C_L and C_H1 .

As discussed before (2), the Kern and Oz serological markers of human λ chains, which correspond to positions 154 and 191 in the L chain sequence, occur on the surface of the molecule, in close spatial proximity. The *I_H* allotypic markers of human κ chains have recently been shown (13) to involve Ala/Val and Val/Leu substitutions at positions 153 and 191, respectively, which closely correspond to the positions of the Kern and Oz markers in human λ chains. Replacements at positions 153 and 191 in κ chains will also affect antigenic determinants of the molecule that are recognized by anti-allotypic antisera. Since the distance between the α -carbon atoms of the homologous residues is about 8–10 Å, replacements involving both positions can be simultaneously recognized by a single anti-allotypic antibody molecule.

Some of the variable and constant features of V_L sequences can be discussed in terms of the three-dimensional structure of V_λ New. Hairpin bends in the polypeptide chain of V_λ New occur around positions 14–15, 27–30, 39–40, 67–68, and 92–93 and an approximate 90° bend around residues 75–76. Except for the bend at positions 92–93 (a hypervariable region) all others involve a Gly residue that is constant in human λ chain sequences. Most of these bends also involve a constant or nearly constant Pro-Gly or Ser- or Thr-Gly sequence. A similar conclusion has been obtained from the study of a crystalline V_κ fragment (4). Glycine residues also contribute to a constant sequence (Phe-Gly-Gly-Gly, positions 99–102) that is not part of a bend. The constant character of this sequence in all λ chains can be explained by the following observations: (a) Phe 99 is located in an internal, interchain hydrophobic pocket that includes the homologous constant Trp 108 in V_H , related to Phe 99 by a local pseudo 2-fold axis of symmetry; it can be assumed that Phe 99 (and Trp 108) make an interchain contact that is important for V_L - V_H

assembly; (b) Gly 100 in V_λ (or Gly 109 in V_H) is tightly packed between the intrachain disulfide bond and Leu 4 (a constant residue in V_L and V_H); (c) Gly 101 is relatively close to the constant Gln 6, although here there is room for a side chain as observed in V_κ (Gln 101) or in V_H (position 110); (d) Gly 102 (111 in V_H) is very close to a constant aromatic residue (Tyr/Phe, 86 in V_λ or 95 in V_H) so that the limited space available requires the presence of a Gly residue at this position. Some other residues that are constant in V_λ or that show only a limited degree of variability, such as Tyr 35, Gln 37, Ala 42, Pro 43, and Asp 84 are involved in close contacts with the V_H subunit. Other constant residues such as Gln 37 and Tyr 85, Glu 82 and Tyr 142 (C_L) make internal hydrogen bonds. In addition to the residues just discussed, most of the nonpolar, hydrophobic amino acids that occur in the interior of the structure between the two β -sheets are invariant or are replaced by other hydrophobic residues. They are Leu 4, Gln 6, Val 10, Val 18, Ile 20, Cys 22, Val 32, Trp 34, Leu 46, Phe 61, Val 63, Ala 70, Leu 72, Ile 74, Leu 77, Ala 83, Tyr 85, Cys 87, Ser 89, Val 98, Thr 103, Leu 105, and Val 107. All the constant or nearly constant residues that appear at bends or contribute to intra- and inter-subunit bonds seem to be important for the preservation of structure. Mutations that alter any of these residues, which constitute more than 50% of the total number of residues of the V_λ sequence, cannot be considered to be "neutral or modulating" (14). Different combinations of these invariant and semi-invariant residues that are compatible with the requirements of the constant three-dimensional structure of the homology subunits are, in our view, better explained by a process of evolutionary germ line gene divergence than by random somatic mutations. By contrast, the nature of the residues that occur in the regions of hypervariable sequence, at one end of the molecule and fully exposed to the solvent, is not limited by any visible structural constraints.

It has recently been pointed out (15) that the sequence Arg/Lys-Phe-Ser-Gly-Ser-Lys (positions 60–65) is constant in λ chains from several animal species that have been ana-

lyzed and that this sequence could, therefore, fulfill a special function. The side chains of Phe 61 and Gly 63 (Val 63 in λ Newm) are part of the internal hydrophobic core surrounded by the β -pleated sheets; consequently, they could be replaced only by other nonpolar residues. Arg 60 appears to be involved in an internal hydrogen bond to Glu 80 and/or Gln 78. However, the side chains of the other residues of that constant sequence are external with no apparent structural constraint. Residues that occur in the constant or nearly constant N-terminal sequences of V_L , V_K , and V_H can be analyzed in similar terms: some are internal (residues 4, 6, 10, 12 in V_L) and would not be expected to show much change, but others have external side chains with no visible constraints (positions 5, 9, 11, 13, 16 in V_L).

As discussed before (2), the V_H subunit exposes a larger area in the region of the active site. Comparison of the three-dimensional model of Fab' New with that of a λ chain dimer (3) and a V_K dimer (4) suggests that the V_H subunit plays an essential role in defining the conformation and the specificity of the antigen binding site. The V_H hypervariable sequences extending from positions 50–60 and 100–110 are longer than the homologous regions of L chains. In particular, the third hypervariable region of V_H has been found to consist of a variable number of amino-acid residues, ranging from 13 to 20 when counted from Cys 96 to Trp 108, whereas the homologous loop of V_L and V_K (Cys 87 to Phe 99) has been found to contain only 11 to 13 residues. This hypervariable loop of V_H does not conform to the approximate local 2-fold axis of symmetry relating V_L to V_H . Instead, this loop bends towards the L chain (see Fig. 4), making the structural pocket at the active site (2, 16) narrower than it is in L chain dimers, where it appears as a large cavity (3, 4). The width and depth of the active site pocket of different immunoglobulins can, therefore, be altered by variations in the length of this hypervariable V_H loop. Thus different H chains that pair with the same L chain in induced antibodies (17) could modulate affinity not only by changes in the amino acids present in this sequence but also by alterations in the length of the polypeptide chain in this region.

Although the length of the L chain hypervariable regions is relatively constant, some human and mouse (18, 19) κ chains have been found to include additional residues in the first hypervariable region (positions 25–32). This additional length of polypeptide chain will also have an important effect in determining the dimensions of the active site pocket and the antigen specificity of the immunoglobulin molecules to which they belong.

With the model of Fab' New as a basic structural framework, a striking correlation between the structure and function of the well-studied MOPC 315 anti-Dnp mouse myeloma protein (20, 21) can be obtained. The L(λ) chains of IgG New and MOPC 315 are highly homologous and contain an equal number of residues in the third hypervariable region between the constant amino acid residues Cys 89 and Phe 99. Also, a comparison of the tentative sequence of V_H New (Fig. 1) and that of V_H MOPC 315 indicates that the third hypervariable regions of V_H in both chains between the constant Cys 96 and Trp 108 include the same number of amino acid residues. It is, therefore, feasible to fit the MOPC 315 sequence (18) to the basic V_L and V_H structures obtained in the 2.0-Å Fourier map of Fab' New. The model of the MOPC 315 binding site that emerges from this comparison includes a

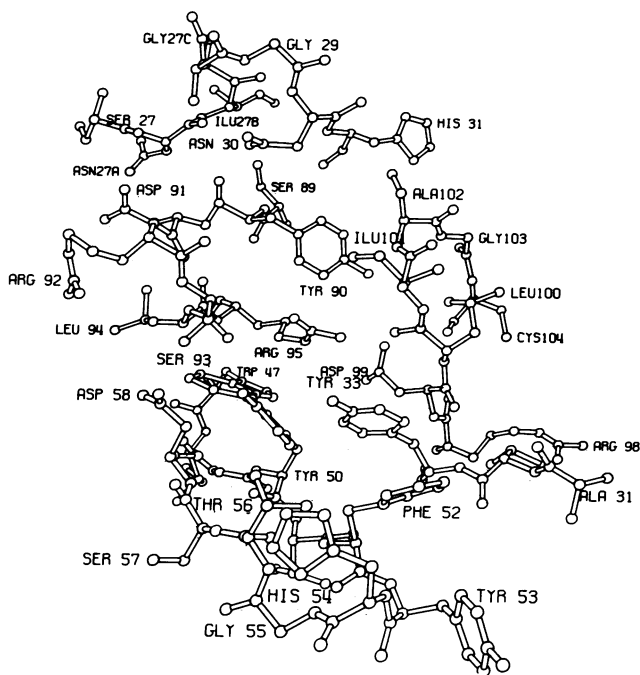


FIG. 4. View of some of the amino-acid residues at the active site of IgG New. Residue numbers for V_L (27 to 31 and 89 to 95) and V_H correspond to those of Fig. 1.

crevice similar to that shown in Fig. 4, in which L chain Tyr 34 forms the "upper" limit, H chain Trp 47 and Phe 50 form the "lower" limit, L chain Phe 99, H chain Tyr 104 and Phe 34 contribute to the "sides" and L chain Trp 98 and Phe 103 and H chain Phe 105 form the 6- to 10-Å deep "floor." The high density of adjacent aromatic side chains that line this crevice, at the center of the active region, is striking and correlates with the observed specificity of MOPC IgA immunoglobulin for Dnp and other haptens that include benzene and naphthalene aromatic rings. The relatively shallow depth of the active center is in agreement with the electron microscopy study of a complex between MOPC 315 IgA and the bifunctional hapten bis(Dnp- β -alanyl)-diaminosuccinate (20) in which the Dnp groups that join the Fab arms of different IgA molecules end to end are only 15 Å apart.

The possibility of a conformational change as a biological signal triggered by antigen binding is an important question to be considered in discussing immunoglobulin structure. In this context, the Fab structure can be described as a tetrahedral arrangement of homologous subunits, covalently linked in pairs (V_L to C_L and V_H to C_{H1}) by linear stretches of polypeptide chain ("switch" regions) bent to a larger extent in the F_d' chain than in the L chain (2), suggesting flexibility. Furthermore, two identical L chains of a dimer assume different conformations (3), such that one of them appears similar to the L chain of Fab', making an angle greater than 90° between the major axes of the C_L and V_L subunits, whereas the other L chain of the dimer makes an angle smaller than 90°, as observed between the V_H and C_{H1} subunits of Fab' New. These observations suggest that a conformational change could take place by a hinge-like movement at one or both switch regions. Since a disulfide bond linking V_L to C_L has been found in some rabbit IgG molecules (22, 23), the flexibility of the more open L chain may be more limited than that of the H chain. An "opening" of the F_d' chain, as illustrated in Fig. 5,

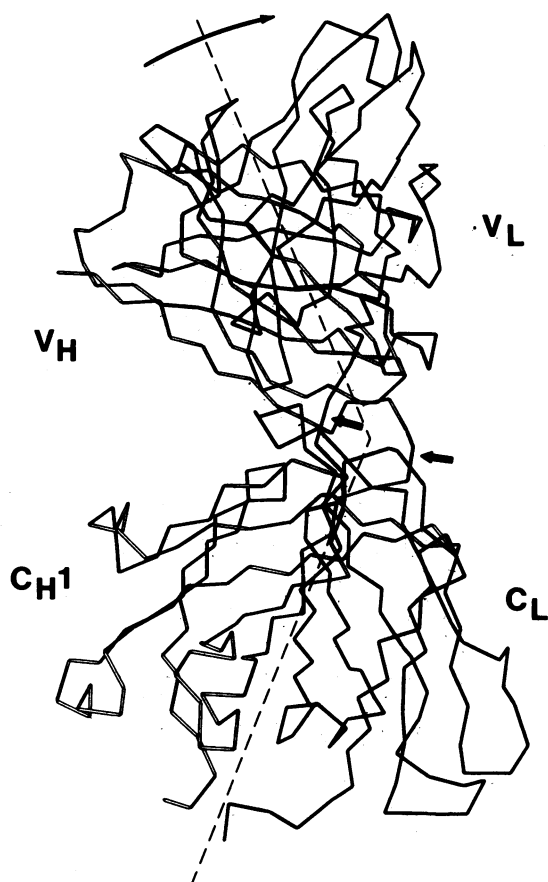


FIG. 5. A view of the α -carbon backbone of Fab' New. The V and C₁ domains, the L chain (open line), the Fd' chain (solid line) and the local, approximate 2-fold axes (broken lines) relating the V_L to the V_H subunit and the C_L to the C_{H1} subunit are shown. The two short arrows indicate the switch region of both chains. The longer arrow indicates a possible relative motion of the V and C₁ domains (see text).

would lead to a relative movement of the structural subunits and to the exposure of some of the V_H and C_{H1} side chains that were not previously exposed. A relative movement of structural subunits leading to changes in quaternary structure has been demonstrated by crystallographic analyses of reduced and oxygenated hemoglobin molecules (24). Although no major conformational changes have been observed after binding of ligands to Fab' fragments (16, 25) the occurrence of such changes cannot be ruled out on the basis of these experiments carried out in the crystalline state. Furthermore, some of the interactions necessary to trigger the postulated conformational change may not be present in the binding of relatively small haptenic groups such as phosphoryl choline (25)

and vitamin K₂OH (16) which can only interact with some of the side chains of the active site.

This research was supported by Grants AI 08202 from the National Institutes of Health, NP-141A from the American Cancer Society, and N.I.H. Research Career Development Award AI 70091 to R.J.P.

1. World Health Organization (1964) *Bull. W.H.O.* **30**, 447-450.
2. Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerley, R. P. & Saul, F. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3305-3310.
3. Schiffer, M., Girling, R. L., Ely, K. R. & Edmundson, A. B. (1973) *Biochemistry* **12**, 4620-4631.
4. Epp, O., Colman, P., Fehlhammer, H., Bode, W., Schiffer, M. & Huber, R. (1974) *Eur. J. Biochem.*, **45**, 513-524.
5. 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**, 421-425.
6. Poljak, R. J., Amzel, L. M., Avey, H. P., Becka, L. N. & Nisonoff, A. (1972) *Nature New Biol.* **235**, 137-140.
7. Richards, F. M. (1968) *J. Mol. Biol.* **37**, 225-230.
8. Chen, B. L. & Poljak, R. J. (1974) *Biochemistry* **13**, 1295-1302.
9. Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. D., Rutishauser, U. & Waxdal, M. J. (1969) *Proc. Nat. Acad. Sci. USA* **63**, 78-85.
10. Dayhoff, M. O., ed. (1972) *Atlas of Amino Acid Sequence and Structure*, Vol. 5 (National BioMedical Research Foundation, Silver Spring, Md.).
11. Low, B. W. & Grenville-Wells, H. J. (1953) *Proc. Nat. Acad. Sci. USA* **39**, 785-801.
12. Peterson, P. A., Cunningham, B. A., Berggard, I. & Edelman, G. M. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 1697-1701.
13. Milstein, C. P., Steinberg, A. G., McLaughlin, C. L. & Solomon, A. (1974) *Nature* **248**, 160-161.
14. Weigert, M. G., Cesari, I. M., Yonkovich, S. J. & Cohn, M. (1970) *Nature* **228**, 1045-1047.
15. Stanton, T., Sledge, C., Capra, J. D., Woods, R., Clem, W. & Hood, L. (1974) *J. Immunol.* **112**, 633-640.
16. Amzel, L. M., Poljak, R. J., Saul, F., Varga, J. M. & Richards, F. F. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 1427-1430.
17. Friedenson, B., Appella, E., Takeda, Y., Roholt, O. A. & Pressman, D. (1973) *J. Biol. Chem.* **248**, 7073-7079.
18. McKean, D., Pötter, M. & Hood, L. (1973) *Biochemistry* **12**, 760-771.
19. Barstad, P., Rudikoff, S., Potter, M., Cohn, M., Konigsberg, W. & Hood, L. (1974) *Science* **183**, 962-964.
20. Green, N. M., Dourmashkin, R. R. & Parkshouse, R. M. E. (1971) *J. Mol. Biol.* **56**, 203-206.
21. Francis, S. H., Leslie, R. G. Q., Hood, L. & Eisen, H. N. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 1123-1127.
22. Poulsen, K., Fraser, K. J. & Haber, E. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 2495-2499.
23. Appella, E. (1973) *Biochem. Biophys. Res. Commun.* **53**, 1122-1129.
24. Perutz, M. F., Muirhead, H., Cox, J. M. & Goaman, L. C. G. (1968) *Nature* **219**, 131-139.
25. Padlan, E. A., Segal, D. M., Spande, T. F., Davies, D. R., Rudikoff, S. & Potter, M. (1973) *Nature New Biol.* **245**, 165-167.