THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 253, No. 2, Issue of January 25, pp. 585-597, 1978
Printed in U.S.A.

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. POLJAK‡

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 0.62 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 ϕ, ψ angular values which corresponds to the predominant β-pleated sheet conformation present in the structure.

The structures of the homology subunits V_H, V_L, C_H1, and C_L were superimposed by pairs and quantitatively compared. The closest similarities were observed between V_H and V_L and between C_H1 and C_L. Amino acid sequence alignments obtained from this structural superposition are given. The closest sequence homology in Fab New is observed between C_H1 (γ heavy chain) and C_L (λ 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 ≤ 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 V_H and V_L and between C_H1 and 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.[1] 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, ϕ and ψ 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·hapten 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

* This work was supported by Research Grant AI 08202 from the National Institutes of Health and by Research Grants NP-141B and IM-105C from the American Cancer Society. 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.

‡ Recipient of United States Public Health Service Research Career Development Award AI-70091.

[1] 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 $\tau_\alpha(\tau_{N\text{-}C\alpha\text{-}C})$ 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 input coordinates or which had an abnormal τ_α value were remeasured and checked for correspondence with the Fourier map. These discrepancies could always be traced to errors in the measurement of coordinates or to distortions of the brass model. When necessary the coordinates were measured again after rebuilding distorted regions and the remeasured coordinates were submitted to the model building procedure as guide points. The final average value of $\Delta\tau_\alpha(\Delta\tau_\alpha = |\tau_\alpha - 109.3°|)$ for the 440 residues in the model-built structure was 5.43°. Pro 151 in C_H1 was built and refined in the *cis* conformation. The root mean square shift from the initial coordinates for all non-hydrogen atoms in the structure was 0.2 Å.

Real Space 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, (a = 111.43 Å), 1/80 along y (b = 56.68 Å), and 1/130 along z (c = 90.30 Å) in sections of constant y. 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. Values of atomic radii for carbon, nitrogen, oxygen, and sulfur that gave fastest convergence in trials using a small part of the structure were adopted and kept constant during refinement. Progress in the refinement process was followed by inspection of the root mean square shifts in coordinates, shifts in the coordinates of individual atoms, and adjustments of amino acid scale factors. Unusually large shifts in coordinates were checked using the optical comparator and the Fourier map. All refinement calculations were carried out using computer programs implemented at the Brookhaven National Laboratories.

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. 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 C_H1 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,[2] based on F_c obtained with the coordinates in "Appendix" and an overall temperature factor[3] (B = 18.0), 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_H, 2.00 Å; V_L, 2.46 Å; C_H1, 2.30 Å; C_L, 2.30 Å; C_H1-C_L, 2.43 Å.

Ramachandran Plots

The Ramachandran plots of the V_L and C_L homology sub-

TABLE I

Conditions of model building refinement[a]

Parameters varied were: ϕ, ψ, χ, τ(N Cα C). Flexible proline residues were used. In addition, τ(Cα Cβ Cγ) was allowed to vary in Cys, His, Phe, Trp, and Tyr residues. A list of the sources of the amino acid groups used in model building is given in Table I of Diamond (7).

Probe	Length (residues)	Filter constants	
		C_1	C_2
1	1	0.1	10^{-4}
2	2	0.1	10^{-4}
3	3	0.1	10^{-4}

[a] See Diamond (5) for definition of terms.

TABLE II

Conditions of real space refinement[a]

Zone length: 5
Margin width: 6
Fixed atomic radii: 1.4 Å
Relative atomic weights; C:6, O:7, N:8, S:16
Relative softness of angular parameters that were allowed to vary:
ϕ, ψ: 4.0
χ: 3.2
Filter levels:

	λ_{min}	$\lambda_{min}/\lambda_{max}$
Scale factor and background	0.0001	0.01
Translational refinement	0.001	0.01
Rotational refinement[b]	0.001	0.001

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.

[2] $R = \Sigma|F_o - F_c|/\Sigma F_o$, where F_o is observed structure amplitude and F_c is calculated structure amplitude.
[3] The isotropic temperature factor (B) used in the expression $exp\langle -B \sin^2\theta/\lambda^2\rangle$.

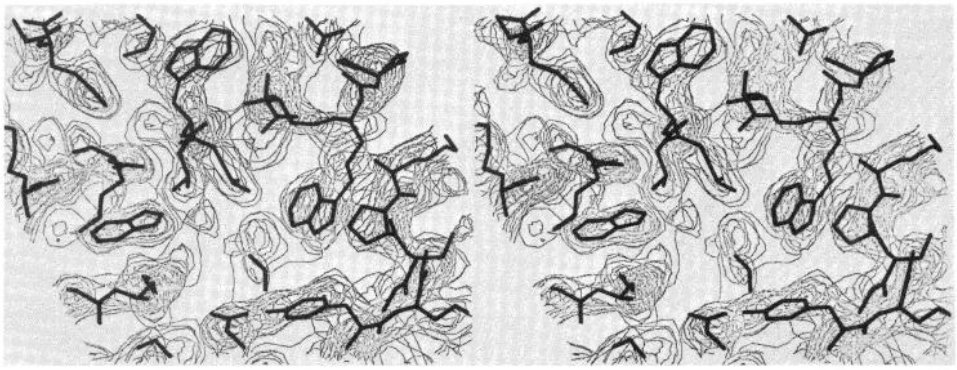

FIG. 1 (*upper*). Stereo view of some main chain and side chain groups of Fab New after five cycles of real space refinement superimposed on the corresponding electron density of the multiple isomorphous replacement 2 Å resolution map.

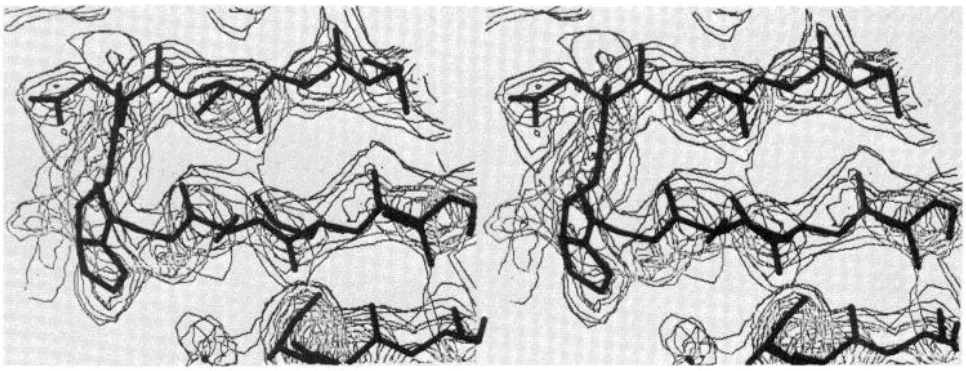

FIG. 2 (*lower*). Stereo view of some amino acid residues of Fab New superimposed on the corresponding density of the multiple isomorphous replacement map. Carbonyl groups of the main polypeptide chain are clearly seen.

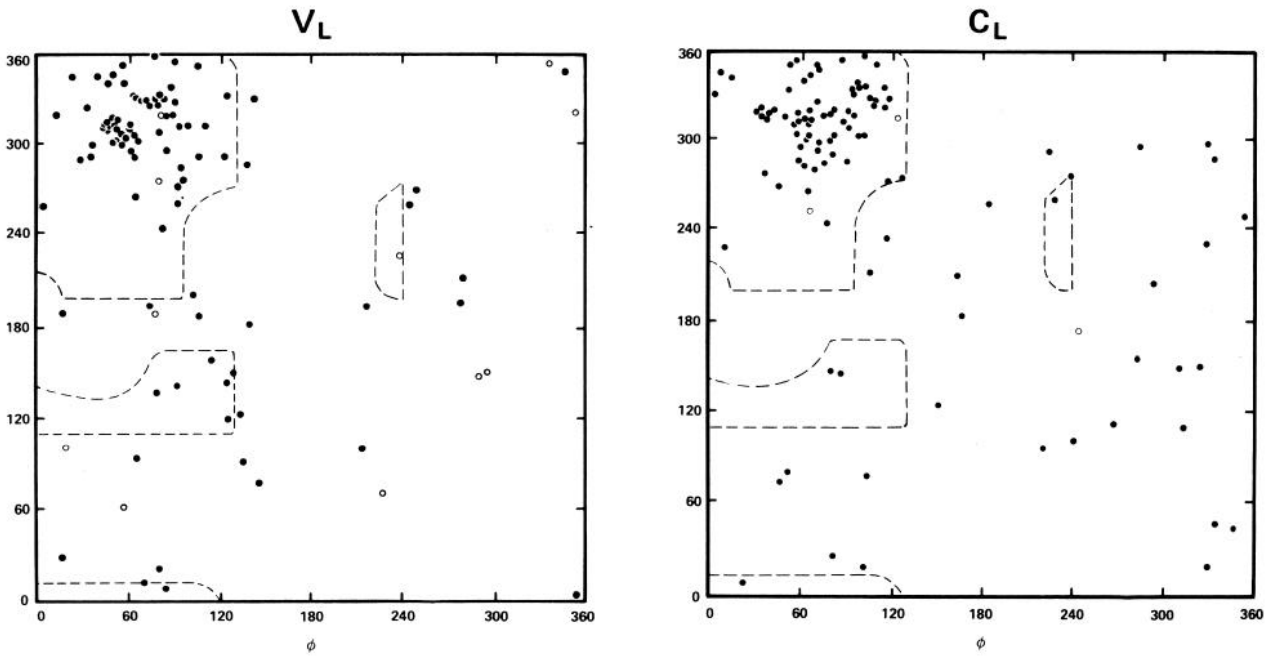

FIG. 3. Ramachandran plots of the V_L and C_L homology subunits of Fab New. The distribution of ϕ, ψ angles indicates the predominant antiparallel β-pleated sheet structure of the subunits. Glycine residues are indicated by ○.

units (see Fig. 3) show a distribution of ϕ, ψ angles which corresponds with the predominant antiparallel β-pleated sheet structure present in V_H, V_L, C_H1, and C_L. As observed in several other protein structures, the ϕ, ψ 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 observed 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 Cα coordinates of the homology subunits were obtained from a small number of structurally equivalent amino acids. The number of equivalences was then extended by an automatic search for stretches of polypeptide chain for which the distances between putatively equivalent Cαs was smaller than 3.8 Å. 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 Cαs occurring at distances of less than 1.5 Å and less than 3.0 Å, 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 Cα distances given in Table III could become smaller with further crystallographic refinement. The number of Cαs which superimpose with distances shorter than 1.5 Å and 3.0 Å 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 Cαs that are structurally equivalent and the average minimum base change per codon. Furthermore, when a restrictive condition for structural equivalence is imposed ($d_{C\alpha-C\alpha} \leq 1.5$ Å) 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_H 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.

TABLE III

Alignment of α-carbon coordinates of four homology subunits of Fab (New) using method of Rao and Rossmann (9)

Subunits	Number of C_α pairs equivalenced with $d_{C\alpha-C\alpha} \leq 1.5$ Å	Average minimum base change per codon for $d_{C\alpha-C\alpha} \leq 1.5$ Å	Number of C_α pairs equivalenced with $d_{C\alpha-C\alpha} \leq 3.0$ Å	Average minimum base change per codon for $d_{C\alpha-C\alpha} \leq 3.0$ Å
V_H-V_L	56	0.98	81	0.97
C_H1-C_L	60	0.71	82	0.80
C_L-V_L	40	1.03	66	1.23
C_L-V_H	29	1.04	59	1.28
C_H1-V_L	27	1.04	58	1.24
C_H1-V_H	25	1.29	49	1.40

The Cαs of the homologous sequences, -Phe-Gly-Gly-Gly- (99 to 102) in V_L and -Trp-Gly-Gln-Gly- (107 to 110) in V_H, can be closely superimposed as can Cα 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 γ and μ 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_H1. The sequence -Thr-Val-Leu- (106 to 108) corresponds to the COOH terminus of V_L and the residues -Gln-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 Cα atoms separated by a distance of 8 Å 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

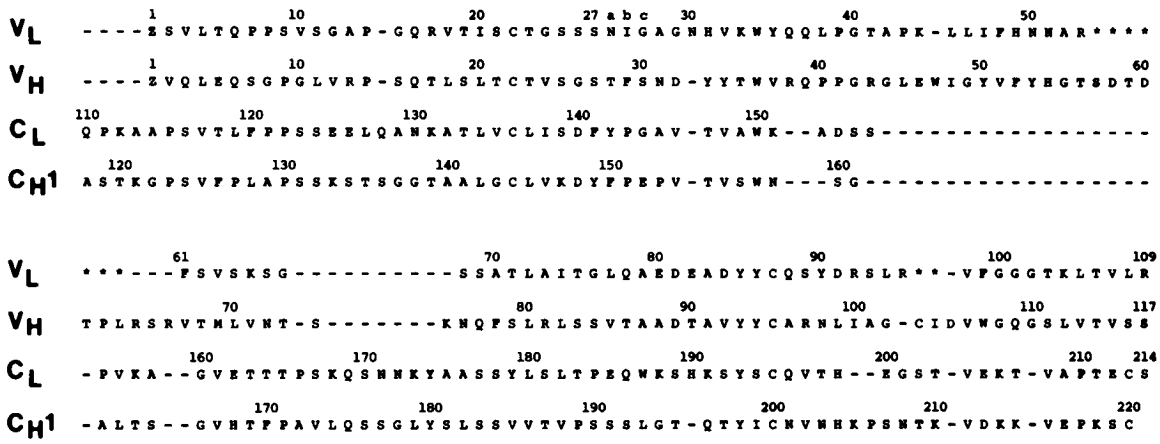

FIG. 4. Amino acid sequences of the V_L, C_L, V_H, and C_H1 homology regions of Fab New aligned by comparison of their three-dimensional structures. – – – indicate gaps introduced to maximize alignment of the three-dimensional structures. * indicate deletions in the V_L sequence. See Ref. 11 for the V_L and C_L sequences, Ref. 12 for V_H, and Ref. 13 for C_H1. Abbreviations for amino acids are as given in Ref. 14.

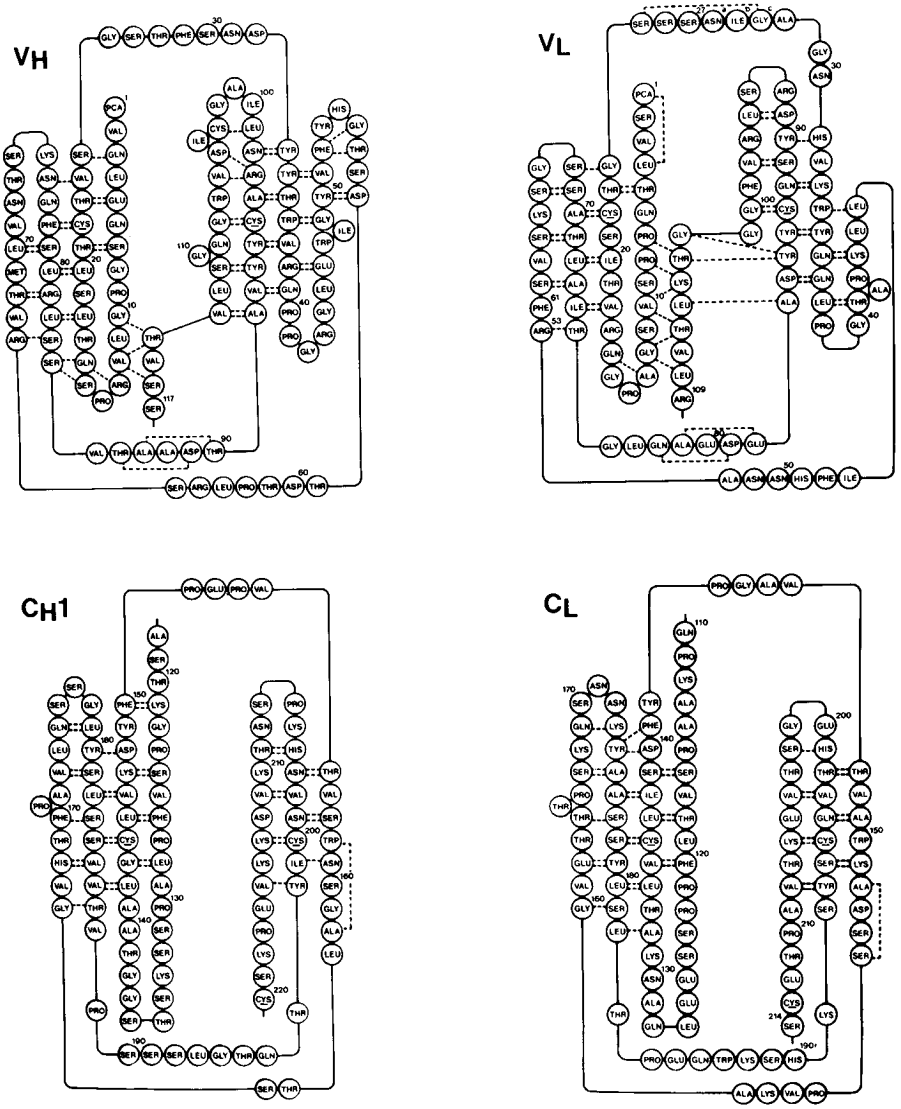

FIG. 5. Diagram of hydrogen bonding (*broken lines*) between main chain atoms for the V_L, C_L, V_H, and C_H1 homology regions of Fab New. The hydrogen-bonded clusters correspond to the two β-sheet structures of each subunit. Cysteine residues that participate in intrachain and interchain disulfide bonds are *underlined*.

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-

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.