

Refined Crystal Structure of a Recombinant Immunoglobulin Domain and a Complementarity-determining Region 1-grafted Mutant

Boris Steipe^{1†}, Andreas Plückthun² and Robert Huber¹

¹*Abteilung Strukturforschung, Max-Planck-Institut für Biochemie, Am Klopferspitz
8033 Martinsried, Germany*

²*Genzentrum der Ludwig-Maximilians-Universität, c/o Max-Planck-Institute für Biochemie
Am Klopferspitz, 8033 Martinsried, Germany*

(Received 13 August 1991; accepted 15 January 1992)

We report the solution of the crystal structure of a mutant of the immunoglobulin V_L domain of the antibody McPC603, in which the complementarity-determining region 1 segment is replaced with that of a different antibody. The wild-type and mutant crystal structures have been refined to a crystallographic R -factor of 14.9% at a nominal resolution of 1.97 Å. A detailed description of the structures is given. Crystal packing results in a dimeric association of domains, in a fashion closely resembling that of an Fv fragment. The comparison of this V_L domain with the same domain in the Fab fragment of McPC603 shows that the structure of an immunoglobulin V_L domain is largely independent of its mode of association, even in places where the inter-subunit contacts are not conserved between V_L and V_H . In all three complementarity-determining regions we observe conformations that would not have been predicted by the canonical structure hypothesis. Significant differences between the V_L domain dimer and the Fab fragment in the third complementarity-determining region show that knowledge of the structure of the dimerization partner and its exact mode of association may be needed to predict the precise conformation of antigen-binding loops.

Keywords: V_L domain; domain interaction; CDR; canonical structures; antibodies

1. Introduction

With immunoglobulins, nature has perfected a protein engineering system permitting the generation of a seemingly unlimited repertoire of complementary molecular surfaces. The variable, antigen-binding immunoglobulin domain consists of a conserved core structure formed by a β -barrel, and topped by three loops of highly variable length and sequence, the so-called complementarity-determining regions (CDR \ddagger : for recent reviews, see Alzari *et al.*, 1988; Davies *et al.*, 1990). These loops, grafted on a structurally conserved platform, confer antigen affinity and binding specificity to the antibody molecule. A variable domain of the light chain

(V_L) combines with a variable domain of the heavy chain (V_H) in the antibody molecule to form the heterodimeric Fv fragment, the smallest immunoglobulin substructure that is fully competent to bind the antigen. A number of structures of immunoglobulin variable domains have been determined since the early 1970s. These proteins were either light-chain dimers secreted by myelomas, or the proteolytically accessible antigen-binding antibody fragment (Fab).

Research into the structures of specific CDR sequences is of interest beyond basic science, where they are regarded as a paradigm for local effects on protein structure. The construction of recombinant "humanized" antibodies, in which all the CDRs of a specific murine antibody have been grafted onto a human framework (e.g. see Riechmann *et al.*, 1988), has recently become of interest for new therapeutic approaches in medicine. As sequence information for specific antibodies is nowadays readily obtainable, understanding of the sequence-structure relationships of the CDRs has become the focus of

[†] Author to whom correspondence should be addressed.

[‡] Abbreviations used: CDR, complementarity-determining region; V_L -D, homodimer of V_L domains; V_L -Fab, V_L domain in the context of the Fab fragment; RMSD, root-mean-square difference; r.m.s., root-mean-square.

Table 1
Aligned sequences for the V_L CDR1 of McPC603, MOPC167 and the mutant M3

Position:	28	29	30	31	31a	31b	31c	31d	31e	31f	32
McPC603 V _L	Ser	Leu	Leu	Asn	Ser	Gly	Asn	Glu	Lys	Asn	Phe
M3:	Ser	Leu	Leu	Tyr	Lys	—	Asp	Gly	Lys	Asn	Phe
MOPC167 V _L	Ser	Leu	Leu	Tyr	Lys	—	Asp	Gly	Lys	Thr	Tyr

Mutated residues relative to McPC603 are indicated by bold type.

theoretical (representative examples are: Stanford & Wu, 1981; Snow & Amzel, 1986; Fine *et al.*, 1986; Shenkin *et al.*, 1987; Bruccoleri *et al.*, 1988) and comparative efforts (representative examples are: Mainhart *et al.* 1984; de la Paz *et al.*, 1986; Chothia & Lesk, 1987; Martin *et al.*, 1989; Chothia *et al.*, 1989; Holm *et al.*, 1990). A particularly attractive approach is the "canonical-structure" hypothesis that has been forwarded by Chothia & Lesk (1987). Their comparison of published immunoglobulin structures and sequences seems to indicate that the fold of the CDRs critically depends only on the nature of a small number of conserved residues and the length of the loop. Unfortunately, the relative scarcity of high-resolution structural information slows progress in this field, and it is not yet clear whether this canonical-structure approach will be generally applicable for accurate and reliable structural predictions.

With the successful functional expression of immunoglobulin domains in *Escherichia coli* (Skerra & Plückthun, 1988; reviewed by Plückthun, 1991), it has become possible to easily obtain sufficient quantities of specific immunoglobulin domains for detailed study. We have previously reported the crystallization and the solution of the crystal structure of the recombinant κ -V_L domain of the phosphorylcholine-binding murine IgA antibody McPC603 (Glockshuber *et al.*, 1990). This structure shows a dimeric crystal packing arrangement corresponding to V_L dimers previously reported in the literature (Fehlhammer *et al.*, 1975; Epp *et al.*, 1975; Colman *et al.*, 1977), resembling the arrangement of V_L and V_H in the antigen-binding Fv fragment. As the structure of the whole Fab fragment of McPC603 has been published previously (Satow *et al.*, 1986), we can now compare the structure of a V_L domain in the context of the homodimer (V_L-D) with the structure of the same domain in the heterodimeric association (V_L-Fab).

Interestingly, none of the three CDRs of this domain participates in interdomain contacts in the homodimer structure. We thus predicted that this system would be particularly useful for the structural study of different hypervariable loops. Indeed, we were able to crystallize a mutant of the McPC603 V_L, in which the first CDR sequence (CDR1) was replaced by that of a different phosphorylcholine-binding antibody, MOPC167 (Potter, 1977; Perlmutter *et al.*, 1984). The mutant domain, named M3, crystallized isomorphously under the same conditions as the wild-type protein. We have now solved the structure using the same model as

described previously (Glockshuber *et al.*, 1990), refined the structural model to a crystallographic *R*-factor of 14.9% and used the refined model for a final refinement of the wild-type domain.

2. Materials and Methods

(a) Preparation and crystallization

The preparation, crystallization and initial structural determination of the McPC603 V_L (subsequently called M603) has been described in detail (Glockshuber *et al.*, 1990). A mutant of this V_L fragment, called M3, replaces the CDR1 of McPC603 with that of MOPC167 (Table 1).

The plasmid encoding the Fv fragment of the mutant M3, constructed by site-directed mutagenesis of the synthetic M603 gene, was obtained from J. Stadlmüller. The first 114 amino acid residues of the light chain, comprising the V_L domain, are encoded. The M3 Fv fragment binds phosphorylcholine with a binding constant of $3.5 \times 10^5 \text{ M}^{-1}$ (Stadlmüller, 1991). Since this is similar to that observed for the Fv fragment of McPC603 (Skerra & Plückthun, 1988), the experimental protocol described previously for the purification of M603 could be followed for the preparation of M3. The purified protein crystallized from 1.7 M-ammonium sulfate, 0.1 M-acetate, (pH 4.0). Large, stable, macroscopically hexagonal crystals grew within 2 days, and were morphologically indistinguishable from the wild-type. They were completely isomorphous to the M603 crystals with space group *P*6₁22 and cell dimensions of $a = b = 86.5 \text{ \AA}$, $c = 74.6 \text{ \AA}$ ($1 \text{ \AA} = 0.1 \text{ nm}$). The asymmetric unit contains 1 V_L domain. Thus 12 V_L domains, with a combined molecular weight of 148.8 kDa, occupy a unit cell of $4.8 \times 10^5 \text{ \AA}^3$. The ratio volume/molecular weight is $3.25 \text{ \AA}^3/\text{Da}$ and the solvent content is estimated to be 62% (Matthews, 1968). This value is at the upper end of those observed for protein crystals and it is remarkable that, in spite of the high solvent content, the crystals are very well ordered and diffract to beyond 1.9 Å resolution.

(b) Data collection

A complete, native data set consisting of 96,963 measurements ($R_{\text{merge}} = (\sum |I|)/\sum I = 0.086$) for 11,262 unique reflections was collected from a single crystal on a FAST diffractometer (Enraf-Nonius, Delft) using a Rigaku Rotaflex rotating-anode generator. Measurements were made using the MADNES software (Messerschmidt & Pflugrath, 1987). Measurements were corrected for relative scale, temperature factors and absorption (Huber & Kopfmann, 1969; Steigemann, 1974; Messerschmidt *et al.*, 1990). The R_{sym} of averaged Friedel pairs was 0.029. Of the possible reflections to 1.97 Å, 91.3% were measured (more than 2σ above background). The last shell of resolution from 1.99 Å to 1.97 Å was complete to 50%.

Table 2

Correspondence of residue numbers following Kabat *et al.* (1987) and Chothia & Lesk (1987) and a sequential numbering for M3 and M603 V_L domains

Kabat <i>et al.</i> (1987)	Sequential M603	Sequential M3
1-31	1-31	1-31
31a	32	32
31b	33	-
31c-31f	34-37	33-36
32-109	38-115	37-114

M603 has an insertion of 6 residues relative to the Kabat alignment (31a to 31f) and M3 has an insertion of 5 residues (deleting position 31b).

(c) Structure solution

As the crystals were isomorphous to those described previously for the wild-type domain, the structure was readily solved by difference-Fourier techniques using the model phases for the wild-type V_L domain.

(d) Nomenclature and definitions

The numbering of residues throughout this paper is that proposed by Kabat *et al.* (1987). The correspondence between this numbering and a sequential numbering is given in Table 2. Amino acids from symmetry-related domains will be designated by a #. The numbers denoting position of amino acid residues from the heavy chain are preceded by the letter H. The β -strands will be designated as described by Marquardt & Deisenhofer (1982), starting with strand A at the N terminus and ending with the C terminal strand H. To describe directions and orientations within a domain, we will consider the dimerization interface of a domain to be the "front", the CDRs to lie on "top" and the N and C terminus to lie on the "left"-hand side. Structural data for comparisons has been taken from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977) using the following entries: IREI (V_L dimer REI, Epp *et al.*, 1975), 2FB4 (Fab KOL, Marquardt *et al.*, 1980), 2FBJ (Fab J539, Suh *et al.*, 1986), 2HFL (Fab HyHel 5, Sheriff *et al.*, 1987), 2MCG (λ chain dimer MCG, Ely *et al.*, 1989), 2MCP (Fab McPC603, Satow *et al.*, 1986), 2RHE (V_L dimer RHE, Furey *et al.*, 1983), 3HFM (Fab HyHel 10, Padlan *et al.*, 1989), 3MCG (λ chain dimer MCG, Ely *et al.*, 1989) and 4FAB (Fab 4-4-20, Herron *et al.*, 1989). All superpositions for structural comparison were done using co-ordinates for the residues

that we designate the core β -strand region: 3 to 7 and 9 to 14 (not for λ chains), 16 to 24, 33 to 38, 45 to 50, 52 to 55, 61 to 66, 70 to 76, 84 to 90 and 97 to 107. Superposition matrices were calculated to minimize the root-mean-square co-ordinate difference (RMSD) for the N, C α , C and O atoms of these residues. Hydrogen bonds were considered if the donor-H...acceptor distance was less than 2.7 Å and the donor-H...acceptor angle was greater than 145°. The co-ordinates of the wild-type V_L -D and the M3 mutant have been deposited with the Brookhaven Protein Data Bank and are available directly from the authors on request until they have been processed and released.

(e) Refinement of M3

An initial structural model of the M3 protein was based on the co-ordinates of the V_L domain from the McPC603 Fab fragment. These were rotated into the orientation found for the wild-type V_L domain (Glockshuber *et al.*, 1990) and the mutated loop was built into the difference electron density map calculated from the wild-type and M3 reflection data and the preliminary wild-type V_L phases.

Refinement was started for this initial model employing the program EREF (energy restrained crystallographic refinement: Jack & Levitt, 1978). Refinement cycles were run until the structure's *R*-factor had converged, followed by interactive model building on an Evans and Sutherland graphics terminal using the program FRODO (Jones, 1978).

As the refinement progressed to an *R*-factor below 25%, solvent molecules were placed into regions of the $F_o - F_c$ difference electron density map which were above 4.1 σ of the mean and less than 4.0 Å distance from the protein atoms. Only such solvent molecules that were consistently observed, did not significantly increase their *B*-values during further refinement and were in stereochemically plausible positions were retained. A strong peak of residual electron density located between 2 symmetry-related arginine side-chains was interpreted as a sulfate ion. A consistent residual electron density peak at the surface of the protein, containing 4 maxima, was interpreted as an acetate ion. A summary of the refinement process is given in Table 3.

Whereas most regions of the protein refined without major rebuilding, the site of the mutation itself, the CDR1, had to be remodelled several times. The electron density of this loop is weaker than that of the rest of the molecule and it was difficult to interpret unambiguously the structure of the main chain. Atoms without defined electron density were not used for calculation of structure factors.

Table 3

Summary of M3 crystallographic refinement

Cycles	Procedure	<i>R</i> -factor
	Start refinement, data from 8.0 to 2.5 Å	34.0
1-4	Manual adjustments, 5 solvent molecules added	25.4
5-6	54 solvent molecules and 1 sulfate ion added	21.5
7-12	Data from 8.0 to 1.97 Å, manual adjustments	19.7
13	48 solvent molecules added	18.1
14	Use unconstrained individual atomic <i>B</i> values and add phase correction for disordered solvent space	16.5
15-19	Manual adjustments, 41 solvent molecules added	15.0
20	Add 1 acetate molecule, orient carboxamide groups	14.9

Manual adjustments were followed by EREF cycles. The final model possesses a standard deviation of bond lengths of 0.015 Å, and a standard deviation of bond angles of 2.23° from ideal values.

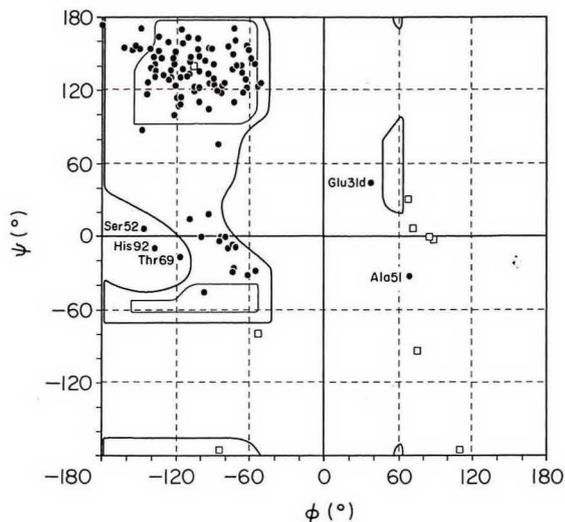


Figure 1. Ramachandran plot for the (ϕ, ψ) torsional angles of the M3 V_L domain. Glycine residues are plotted with a square, all other residues are plotted with a circle. The residues lying outside the predicted conformational boundaries are labelled.

The final model shows clear electron density for all residues from Asp1 to Lys107. The side-chain of Arg108 and the last residue, Ala109, are disordered. The solvent-exposed loop of CDR1 gives only poor electron density. Here, Lys31a is almost undefined. The main-chain atoms from N Asp31c to C Gly31d also show weak electron density. The whole loop seems to be rather flexible, with main chain B -values above 50 \AA^2 . Thus the structural model has to be interpreted with some caution in this region.

(f) Refinement of the M603 V_L domain

A model for the V_L domain was constructed by building the CDR1 loop between 31 and 31e from the co-ordinates of the V_L domain from the McPC603 Fab fragment superimposed onto the refined M3 co-ordinates. This model immediately refined to an R -factor of 15.8%. The R -factor dropped to 14.9% after a final round of minor modifications. Most solvent molecules refined into identical positions as in the M3 mutant with only minor changes in their B -values.

3. Results

(a) The final models

The final models for M3 and M603 contains 864 and 866 protein atoms as well as 121 and 123 solvent atoms, respectively. The two structures are practically identical. The root-mean-square (r.m.s.) co-ordinate difference for all atoms (excluding solvent and the mutated loop) between M3 and M603 is 0.06 \AA . Thus, except where explicitly stated, the following description is valid for both structures. The final R -factor for both models is 14.9% with a standard deviation of bond lengths from ideal values of 0.015 \AA and of bond angles of 2.23° . The maximum average co-ordinate error as calculated from a Luzzati plot (Luzzatti, 1952) is

0.16 \AA . No residual peaks greater than $|0.25 e/\text{\AA}^3|$ were found in the $F_o - F_c$ maps.

The Ramachandran plot for M3 is shown in Figure 1. The majority of residues are found in the antiparallel β -strand region ($\phi = -60^\circ$ to 150° , $\psi = 100^\circ$ to 170°) (Richardson, 1981) and all residues but five are found within the energy boundaries postulated by Ramachandran *et al.* (1966). Three exceptions are Ser52 ($\phi = -141^\circ$, $\psi = 7^\circ$), His92 ($\phi = -135^\circ$, $\psi = 14^\circ$) and Thr69 ($\phi = -115^\circ$, $\psi = 18^\circ$), which are located in well-defined electron density and deviate only slightly from the low energy contours. The most conspicuous deviation is seen at Ala51 ($\phi = 67^\circ$, $\psi = -34^\circ$), which will be discussed below. Gly31d ($\phi = -36^\circ$, $\psi = -42^\circ$) is poorly defined in electron density (see above).

The refined V_L domain shows the typical immunoglobulin variable domain structure (Fig. 2). It consists of two β -sheets with four and five antiparallel strands, forming a typical β -barrel. The barrel is closed on the left-hand side by a compact subdomain containing two strands forming the second CDR (Fig. 3). It is closed on the right-hand side by the first β -strand, which is divided by the *cis*-proline Pro8 into a part A, hydrogen bonding in antiparallel fashion to the B β -strand and a part A' bonding in parallel fashion to the H strand.

(b) Crystal packing

A view of a C^α trace of domains in the unit cell, along the 6-fold symmetry axis, clearly shows the symmetry elements of the $P6_122$ space group (Fig. 4). Each monomer contacts four domains in the crystal lattice. One of these domains forms the V_L dimer, which is discussed in more detail below. Corresponding to the relative orientation of domains to each other, these symmetry relationships will be designated *syn*, *anti* and *meta*. These are generated through the following symmetry operations: *syn*: $(X - Y, Y, -Z)$; *anti*: $(X, X - Y, \frac{1}{6} - Z)$; and *meta*: $(-Y, X - Y, \frac{1}{3} + Z)$ or $(-X + Y, X - Y, \frac{2}{3} + Z)$.

The *syn*-interaction corresponds to the V_L dimer (Glockshuber *et al.*, 1990). The amino acids Tyr36, Gln38, Pro43, Pro44, Leu46, Tyr49, Glu55, Tyr87, Gln89, Asp91, Tyr94, Pro95, Leu96, Phe98, Gly99 and Ala100 lose more than 10% of their solvent-accessible surface upon association and thus form the dimer interface.

The *anti*-symmetry partner binds to the V_L domain with an antiparallel β -strand from the amino acids Ser7 to Ser14 and Ser#14 to Ser#7. As this part of the first β -strand (A') contributes to the front β -sheet of the domain, this β -sheet is extended in the crystal lattice to twice its size.

The *meta*-association is non-symmetric. In this mode of association, the amino acids Lys24, Thr69 and Gly70 at the top of the domain contact Ala#15, Gly#16 and Gln#79 at the bottom of the *meta*-partner $(-Y, X - Y, \frac{1}{3} + Z)$. Equivalently, at the bottom of the domain Ala15, Gly16 and Gln79

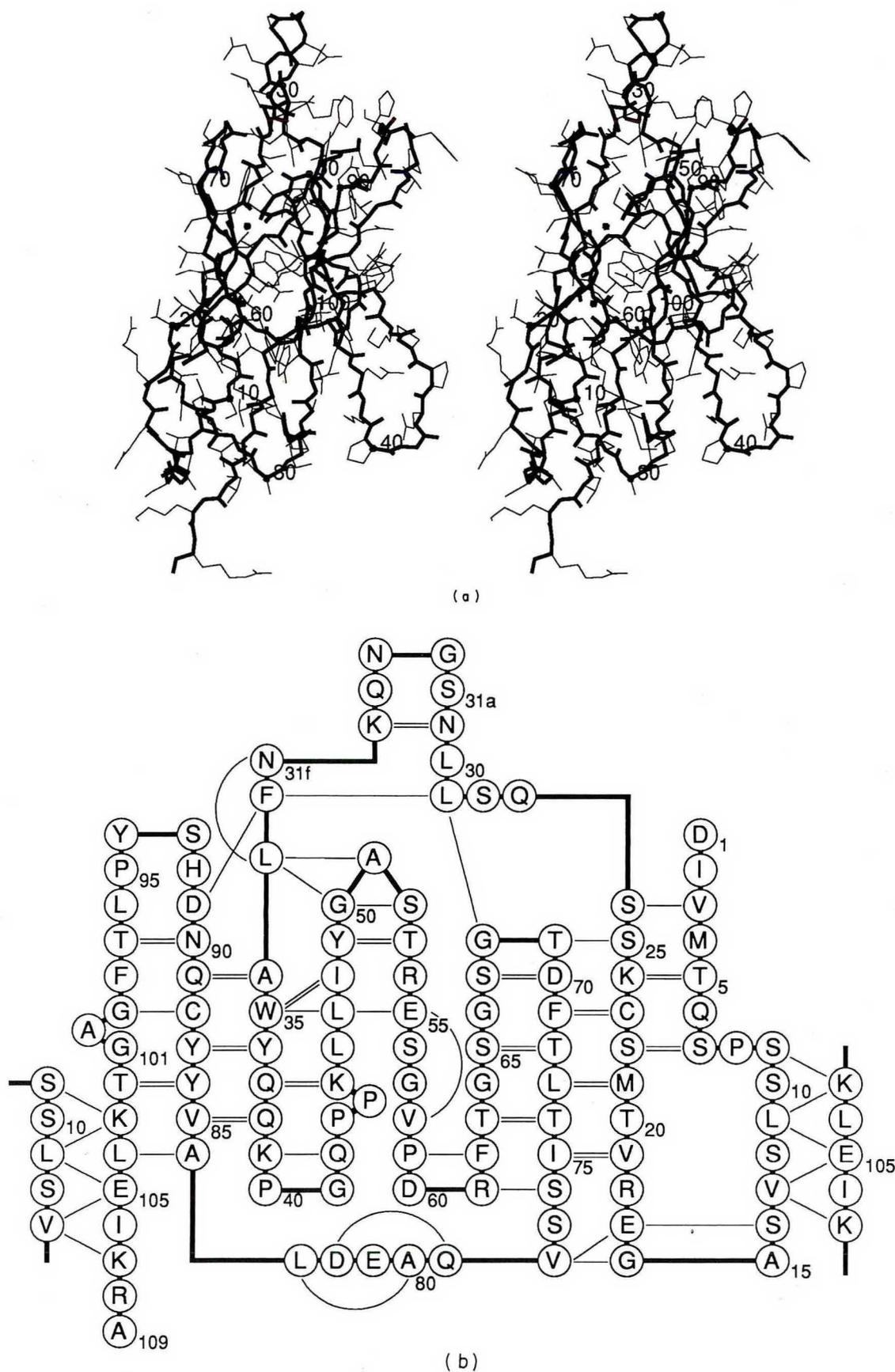


Figure 2. The refined structure of the McPC603 V_L domain. (a) The 3-dimensional structure. Solvent molecules are omitted, except for the 2 internal water molecules. The view is from the right hand side. The main chain is plotted with a bold line, the side-chains are plotted with thin lines. Every 10th C^α atom is labelled. (b) Secondary structure of the V_L domain. Main-chain hydrogen bonds are indicated with a thin line.



Figure 3. The CDR2 subdomain of V_L . Residues from Leu46 to Ser65 are shown together with the main stabilizing residues of the protein core, Leu33, Trp35 and Asp82. The view is from the front of the domain. The framework C^α atom co-ordinates are connected with a thin line, the residues mentioned in the text are drawn with a bold line. Hydrogen bonds are drawn with a broken line.

contact Lys# 24, Thr# 69 and Gly # 70 at the top of the next *meta*-partner ($-X + Y$, $-X$, $\frac{2}{3} + Z$). This top-to-bottom arrangement can be extended in space. The projection of the long axes of *meta*-associated monomers onto the "a b" plane of the crystal lattice gives an angle of 60° . *Meta*-associated monomers turn around the 3-fold screw axis.

Thus the crystal lattice can be easily described. A helix of *meta*-associated monomers turns around a 3-fold screw axis. A second helix, with a phase difference of exactly one half turn, associates through the *anti*-interaction. This double helix now has a period length of the unit cell height. All monomers within this helix point their *syn*-interfaces radially outward. Finally, these columns aggregate in a hexagonal packing *via* the *syn*-interaction to form the crystal lattice. A protein-free tunnel with a radius of about 30 Å, parallel to the protein columns, is found in the crystal lattice at sites corresponding to every third position of a hexagonal close packing. All three CDRs point into this tunnel.

Additionally, a domain is linked to the *meta*-partner of the *anti*-domain, ($-Y$, $-X$, $\frac{5}{6} - Z$), through a double-salt link: two symmetry-equivalent Arg60 residues bind a sulfate ion from the solvent.

(c) Hydrogen bonding

The secondary structure elements, along with the hydrogen bonds between main-chain atoms are shown in Figure 2(b). The structure consists predominantly of antiparallel β -strands, and one turn of a 3_{10} -helix. The boundaries of the β -strands and

the classification of the turns are summarized in Table 4. The four β -II and β -II' turns observed here, contain glycine in the third position (Venkatachalam, 1968). We find that 97% of main-chain polar atoms form hydrogen bonds, 54% (118) are involved in main chain to main-chain bonds, 12% (26) are to side-chain atoms, the rest are to solvent molecules.

All buried polar side-chain atoms make good hydrogen bonds. This implies that the structure is

Table 4
 β -Strand boundaries and turns

Element	Type	Boundaries
β -Strand A		Val3-Ser7
β -Strand A'		Ser9-Ser14
Turn	β -II turn	Ala14-Gly15
β -Strand B		Gly16-Ser26
β -Strand C		Lys31e-Gln38
Turn	β -II turn	Pro40-Gly41
β -Strand D		Lys45-Gly 50
Turn	γ -turn	Gly50-Ser52
β -Strand X		Ser52-Glu 55
Turn	β -II turn	Ser54-Gly 57
Turn	β -I turn	Asp60-Arg61
β -Strand E		Arg61-Ser67
Turn	β -II' turn	Gly68-Thr69
β -Strand F		Thr69-Ser74
Helix	3_{10} -Helix	Gln79-Leu82
β -Strand G		Ala84-Asn90
β -Strand H		Thr97-Lys107

Boundaries of β -Strands are defined by the hydrogen-bonding patterns of the structural model and classification of turns was performed according to the ϕ - and ψ -angles of consecutive amino-acids as given by Richardson (1981).

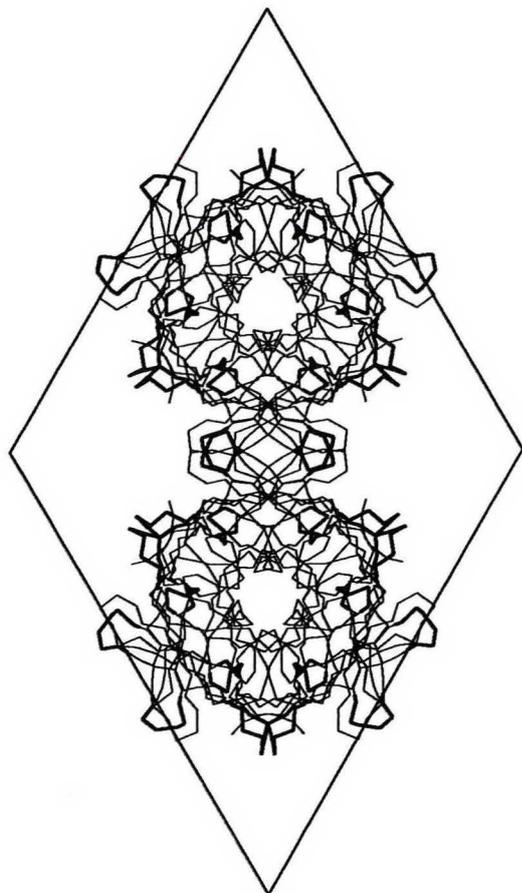


Figure 4. V_L domains in the crystallographic unit cell viewed along the 6-fold axis of symmetry. A C^α atom trace is shown, with the CDRs plotted with a bold line. The symmetry elements of the $P6_122$ space group are clearly visible.

at a conformational optimum, in the sense that a mutation at any of these sites would increase the conformational energy or necessitate some (local) refolding. The buried polar side-chain atoms are all associated with some distinct conformational feature of the domain, and all are among the most highly conserved amino acid residues in the immunoglobulins. (1) Gln6 hydrogen bonds to the structural water molecule W2 (see below) and to amide nitrogen atoms of Gly101 and Cys88. Thus, it may be important in the formation or stabilization of the β -bulge of Ala100 in strand H. This β -bulge generates the twist in the front β -sheet, which characterizes the immunoglobulin variable domain dimerization (Chothia *et al.*, 1985). Gln6 is almost completely conserved in the V_L domains and found to be mutated to Glu in only half of the V_H domains. (2) O^γ Ser25 bonds to O Gln27 (2.73 Å), stabilizing the first turn from β -strand B into the CDR1. It is not part of the core structure and, indeed, an alanine is found in this position in the structures of REI, J539, HyHel5 and HyHel10 without significant structural consequences. (3) $N^{\epsilon 1}$ Trp35 bonds to the buried solvent W1 (see below). (4) O^γ Tyr86 hydrogen bonds to O Asp82 (2.61 Å), thus affixing

the 3_{10} -helix, which crosses under the domain, to the core of the structure. Again, this residue is almost completely conserved, and it is significant that a mutation to Phe seems to be strongly selected against in the immunoglobulin sequence. (There are only 2 Phe residues in this position in all the V_L sequences, and all the rest are Tyr.) This hydrogen bond is conserved in all the published immunoglobulin V_L structures. (5) Thr97 is the first conserved residue after the CDR3. This position is also frequently taken by a valine and thus the requirements seem to favor a small, β -branched amino acid. The $O^{\gamma 1}$ Thr97 to O Ile2 hydrogen bond (2.72 Å) may stabilize the N terminus of the protein. As can be seen from Figure 2(a) and (b), the first two residues cannot participate in the regular anti-parallel β -strand, since strand B turns at the height of these residues to cross over the top of the domain and form the CDR1. In the structures of KOL, MCG and RHE, this hydrogen bond is absent, since Val, Val and Gly are found, respectively, in position 97. This correlates with an N terminus that is somewhat displaced from the framework in these structures. On the other hand, wherever Thr is found, the N terminus is integrated into the rest of the domain in a manner comparable to the V_L -D. (6) Thr102 forms a hydrogen bond to O Pro8 (2.78 Å) to provide the crucial interaction that breaks the regular main-chain hydrogen-bonding pattern, as the first β -strand makes its transition from A' (anti-parallel β -structure) to A' (parallel β -structure). This conformation is conserved in all the published immunoglobulin V_L structures.

Additionally, the stereochemistry of the polar side-chain atoms buried in the domain core conforms well to the spatial preference regions found by Ippolito *et al.* (1990) in their recent analysis of side-chain hydrogen bonding in well-refined high resolution protein structures. The hydrogen bond O^γ Tyr86 to O Asp82 (2.61 Å) is found almost perfectly in plane and at a 120° angle, consistent with an sp^2 -hybridization of the tyrosine hydroxyl group. $N^{\epsilon 1}$ Trp35 bonds W1 only 4° out of plane and 11.2° out of line. $O^{\gamma 1}$ Thr97 bonds O Ile2 13° out of *trans* and $O^{\gamma 1}$ Thr102 bonds O Pro8 Ser only 13° out of *gauche*⁻. Finally, the salt link between Arg61 and Asp82 also displays ideal stereochemistry (type 1: Singh *et al.*, 1987), 9° out of the guanidinium group plane and 3° out of the carboxylate group plane, with the oxygen atoms approaching the nitrogen atoms N^ϵ and $N^{\gamma 1}$ at 2.78 and 2.77 Å. Again this important interaction is conserved in the other immunoglobulin structures examined. Thus, the conformation of the whole domain seems highly optimized towards optimal stereochemistry of the conserved buried polar side-chain atoms.

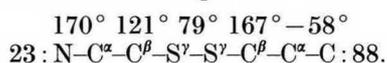
(d) *Orientation of N and O atoms in carboxamide groups of Asn and Gln*

Assuming, that the side-chain carboxamide atom B -values should be approximately equal, those that showed large differences in their B -values, with

values for O atoms larger than those for N atoms, were rotated by 180° from their initial orientation taken from the V_L -Fab. The solvent-exposed side-chain of Gln27 was modelled better this way, indicating a conformational preference, while Gln79 and Gln89 seem statistically oriented. More significant differences are the hydrogen bonding carboxamide groups of (1) Gln37, which hydrogen bonds with its $N^{\epsilon 2}$ to the lone pair of O $^\gamma$ Tyr86, which in turn hydrogen bonds to the carbonyl group of Asp82, and (2) Asn90, which has a stereochemically implausible conformation in the V_L -Fab.

(e) *The disulfide bridge*

Epp *et al.* (1975) have described two alternative conformations of the intramolecular disulfide bond between Cys23 and Cys88 in the REI V_L dimer. Here, clearly only a single conformation is seen; it possesses the torsional angles:



Thus it would be considered a right-handed disulfide spiral (Richardson, 1981). The disulfide bond bridges the large $C^\alpha-C^\alpha$ distance of 6.6 Å, typical for immunoglobulin domains.

(f) *The structural water molecules W1 and W2*

Two water molecules are found within the V_L domains that are not part of the hydration shell, but are integral parts of the domain architecture. These are designated W1 and W2. W1 is found in a cavity between the CDR2 subdomain and the core of the protein. It makes three hydrogen bonds of near-ideal geometry with the indole nitrogen atom of the conserved core residue Trp35, and the backbone carbonyl oxygen atoms of Ala51 and Ser65. This water molecule is found in all κ -domain structures solved at sufficient resolution, REI and J539. Thus, we postulate that this water molecule is a universally conserved structural feature of κ - V_L domains. W2 is found in a cavity between the β -strands A, B and H, some 5 Å from the protein surface. Again, we find three hydrogen bonds with near-ideal geometry. They connect the water molecule with the buried glutamine carboxamide nitrogen atom, $N^{\epsilon 2}$ Gln6, with the carbonyl O atom of Ser22 and with the buried hydroxyl group of Thr102. Interestingly, this water molecule is not found in the structure of REI even though all three potential ligands are present. The only local structural difference is a change from methionine (Met21) to isoleucine at the bottom of the cavity that contains this water molecule. The water molecule is present in the structure of J539, despite an isoleucine in position 21.

(g) *The V_L dimer compared to the McPC603 Fab structure*

As has been noted previously, the V_L monomer is seen to associate as a homodimer in the crystal

lattice. The second domain can be generated through the symmetry operator ($X - Y, -Y, -Z$) and a translation vector of ($A, 2B, C$) in the non-orthogonal co-ordinate system of the unit cell. This association is very similar to the V_L dimer REI and places the two V_L domains in a relative spatial arrangement corresponding closely to the structure of the heterodimeric Fv fragment. A structural superposition of the V_L domains of this study and the published McPC603 Fab shows only small differences. The core β -strand region C^α atoms of the two structures can be superimposed with an r.m.s. deviation of 0.32 Å, which is within the limits seen for other proteins crystallized under non-identical conditions (Chothia & Lesk, 1986). M603 and REI have only 74% amino-acid identity in this region (63% overall). Still, the RMSD of superposition of these two different proteins is the same as that between the V_L -D and the V_L -Fab. As the resolution of the Fab fragment is only 3.1 Å, the significance of conformational differences between the V_L -D and the V_L -Fab is difficult to judge in some cases. The more conspicuous differences are: (1) the side-chain of Leu11 packs differently in the V_L -D; this allows the main chain to be modelled into more favorable stereochemistry. (2) The loop between Ser14 and Glu17 was rebuilt in the V_L -D, the orientation of the Ser14-Ala15 peptide plane was changed by 180° and the rest of the residues moved accordingly. The V_L -D conformation corresponds more closely to the conformation seen in other high-resolution V_L structures. (3) Pro43 and Pro44 move back in the V_L -Fab, possibly due to different interface packing interactions. (4) The peptide plane between Gly50 and Ala51 was turned by 180° . (5) The conformation of the CDR3 is significantly changed, as discussed in more detail below.

In accordance with the observation that residues at the interface between the two domains are highly conserved and their spatial location is preserved among V_L and V_H domains (Chothia *et al.*, 1985), the structure of the V_L -D and the V_L -Fab superimposes particularly well in this region. There are no major rotational or translational movements seen. Evolutionary pressure seems to have allowed only very conservative changes from the time V_H and V_L diverged from a common precursor. It is only in one place, where TrpH103 of V_H is replaced by Phe#98 of the V_L dimer, that a small cavity would be created. But the interface region in this three-layer packing is sufficiently flexible to fill this space through a slight shift in the side-chain positions of Phe98 and Tyr36 (Fig. 5). A Gln-Gln hydrogen bond within the interface, a structural feature found in all immunoglobulin variable domain dimers, is also seen in V_L -D. Gln38 lies at the 2-fold symmetry axis of the dimer and forms two hydrogen bonds of good geometry with Gln#38. We conclude that crystal packing and domain association play only a minor role in the generation of this structure.

A comparison of the mode of association of the V_L -D and the Fv fragment shows that the V_L domain loses 30% less of its solvent-accessible

surface through association in the V_L dimer than in the complex with the Fv fragment. The solvent-accessible surface of the V_L domain of 5620 \AA^2 is reduced to 5050 \AA^2 (a difference of 570 \AA^2) in the V_L dimer, but to 4720 \AA^2 (a difference of 900 \AA^2) in the Fv fragment. Carbon atoms contribute 65% of the interface surface. The absence of the extensive interactions that the V_H CDR3 makes with V_L in the Fab fragment does not cause significant structural changes in the V_L dimer. But a number of specific interactions between the V_L domains can be discerned, each of which would be predicted to cause a decreased association energy of the V_L dimer as compared to the Fv fragment. (1) Asp91, which hydrogen bonds to AsnH95 of the V_H domain in the Fab fragment, is close to the 2-fold symmetry axis in the V_L dimer. Here the side-chain of Asp91 is turned, so that the carboxyl groups of Asp91 and Asp#91 are 12 \AA apart. Even so, repulsive interactions of this paired, negatively charged side-chain, close to the 2-fold axis of symmetry, might serve to explain the fact that V_L crystallizes only at acidic pH. (2) The carboxyl group of Glu55 is seen to make a close contact to the hydrophobic ring of Pro#95 of the CDR3. This might come in conflict with the hydration shell of the carboxyl group. In the Fab fragment, Glu55 is modelled pointing into the solvent. (3) The hydroxyl group of Tyr94 is hydrogen bonded to the carboxyl group of GluH35 in the Fab fragment. This hydrogen bond is absent from the V_L dimer and a well-defined solvent molecule is found in place of the carboxyl group. (4) The most conspicuous difference in the association is the absence of the long CDR3 of the V_H domain, which packs against the descending part of the V_L CDR1, the C strand in the V_L domain. Residue PheH100c is absent from the V_L dimer. This residue plugs deeply into the interface, in the heterodimer. The top of the interface, which is the bottom of the hydrophobic hapten-binding pocket of the Fv fragment, is formed by a hydrogen bond between Asp91 of V_L and AsnH95 of V_H . The absence of this bond and the PheH100c side-chain leave a deep cleft in the interface that extends the solvent-accessible surface almost 10 \AA deep into the interior of the dimer (Fig. 5). Additionally, the hydrogen bond from O^{γ} Tyr36 to N PheH100c has no counterpart in the V_L -D.

(h) Structure of the CDRs

(i) CDR1

The long, solvent-exposed loop of the M603 CDR1 between Asn31 and Lys31e makes few interactions with the main body of the domain. It is thus not surprising that it is found to be partially disordered or flexible with main chain B -values above 50 \AA^2 . Although the M3 sequence is one residue shorter, the loop is found to occupy approximately the same region of space as the M603 loop. The M3 loop is found similarly disordered, with the weakest electron density at the first "corner" of the loop (Ser31a in M603 and Lys31a in M3).

(ii) CDR2

The CDR2 is very well defined and the electron density permits an unambiguous definition of the position of all involved atoms (Fig. 6(a)). It is bridged by four hydrogen bonds. A "classical" γ -turn (Rose *et al.*, 1985; class 3 γ -turn, Milner-White *et al.*, 1988) is found at the apex of the two β -strands D and X, and Ala51, bracketed by this turn, displays the unusual (ϕ, ψ) combination of $(\phi = 66^\circ, \psi = -35^\circ)$. This region of the (ϕ, ψ) map was originally "forbidden" by Ramachandran *et al.* (1966) but was found to be favorable in more refined calculations which allow some flexibility of atomic bonds (e.g. see Weiner *et al.*, 1984). The conformation around Gly50 and Ala51 is different from the one modelled by Segal *et al.* (1974) in the V_L -Fab (Fig. 6(b)). But the clearly defined electron density and more favorable stereochemistry in this region prompt us to postulate that the conformation observed here is the correct one. This makes the conformation of the CDR2 uniformly conserved among immunoglobulin V_L domains.

The main chain continues with two β -turns (Ser56-Gly57, Asp60-Arg61) after the CDR2. They fold the main chain back onto itself to form one of the most compact subdomains observed so far in proteins (Zehfus & Rose, 1986). This compact subdomain, running from Leu46 to Ser65 wraps around the side-chain of Ile48 and forms a plug, sitting against the hydrophobic core of the protein and closing it off to the side (Fig. 3). Only the N- and C-terminal segments of this subdomain are anchored to the main body of the protein with: (1) a peculiar sequence of double-single-double hydrogen bonds between β -strands C and D (these are: N Leu47 (2.8 \AA) and N Ile48 (3.2 \AA) with O Trp35; O Ile48 with N Trp35 (2.9 \AA); N Gly50 and N Ala51 with O Leu33 (both 3.0 \AA)); (2) the hydrogen bond of O Ala51 to the core tryptophan Trp35 *via* the internal water molecule W1; (3) a salt link between Arg61 and Asp82, two residues that are almost universally conserved in the V_L domains; and (4), regular anti-parallel β -structure of the strands E and F, beginning at Arg61 (Fig. 2(b)). Thus, the stabilization of this subdomain largely relies on non-covalent interactions with only three residues, Leu33, Trp35 and Asp82, besides van der Waals' interactions.

(iii) CDR3

The CDR3 conformation is significantly different from the V_L -Fab conformation. Again, the electron density is very well defined and the main chain and side-chains of this region can be unambiguously placed (Fig. 7(a)). The main chain is moved approximately 2 \AA towards the front and the side-chain of His92 is tilted in the same direction, so that the C^α His92 is at a distance of 3.0 \AA from the V_L -Fab conformation. The imidazole group moves by 6.8 \AA (Fig. 7(b)). The RMSD of the 44 atoms from Asn90 to Tyr94 is 2.76 \AA (2.05 \AA for the 19 backbone atoms). Additionally, the hydrogen bonding topology of this loop changes: Asn90 is found to make

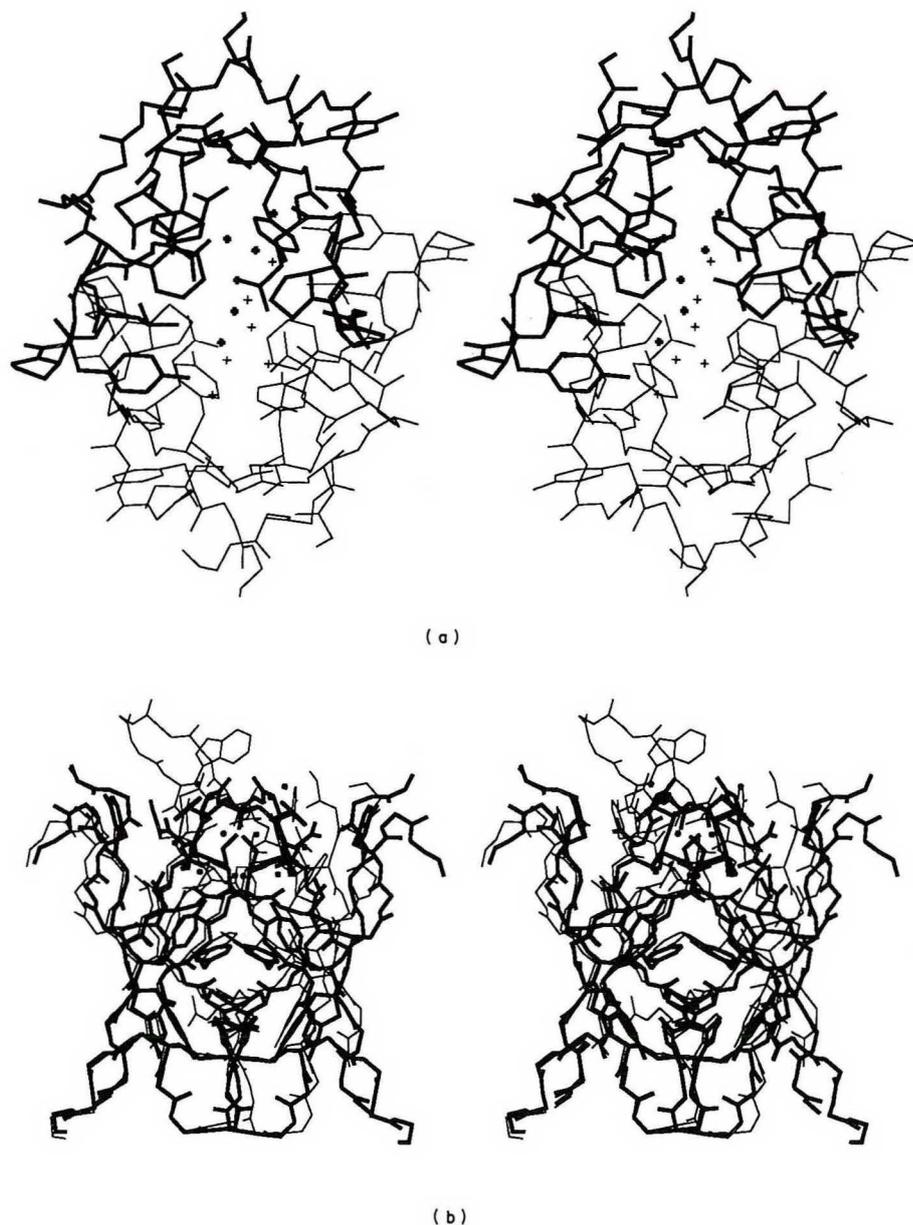


Figure 5. Stereoplot of the dimerization interface of the V_L dimer and the Fv fragment. (a) The view is from the top along the 2-fold axis of symmetry, only the front β -sheets and interacting side-chains are shown. One domain of the V_L -D dimer is drawn with bold lines, the other is drawn with thin lines. Solvent molecules of this region are included. The hydrogen bonding Gln38 and Gln#38 are clearly seen at the bottom of the interface. (b) View from the right-hand side, only the front β -sheets and interacting side-chains are shown. Both domains of the V_L -D dimer are drawn with bold lines, the Fv fragment is drawn with thin lines. Solvent molecules for V_L -D are included. The overall geometry of the interface, which forms a β -barrel, can be well appreciated. Note that this β -barrel is not closed through secondary structure. The V_H CDR3 extends over the top of the domains. Note, by comparison, how deep the solvent molecules penetrate into the dimer interface in the V_L -D. The side-chains of Phe98 and Trp103H are clearly seen in the center of the interface.

two weak hydrogen bonds with N Ser99 (3.06 Å) and O Ser99 (3.06 Å), instead of the strong bond to N His98 (2.54 Å) seen in V_L -Fab.

(i) *Comparison of M603 and M3*

The structures of M603 and M3, excluding the mutated loop from Leu30 to Asn31f, can be superimposed with an RMSD of 0.06 Å. They are virtually identical, down to the B -values of most

solvent molecules. This shows that the conformation of the mutated loop has no significance for the remaining domain structure.

4. Discussion and Conclusions

Chothia & Lesk (1987) have put forward the interesting hypothesis that for the accurate prediction of CDR conformations, knowledge of the nature of a limited number of key residues and a

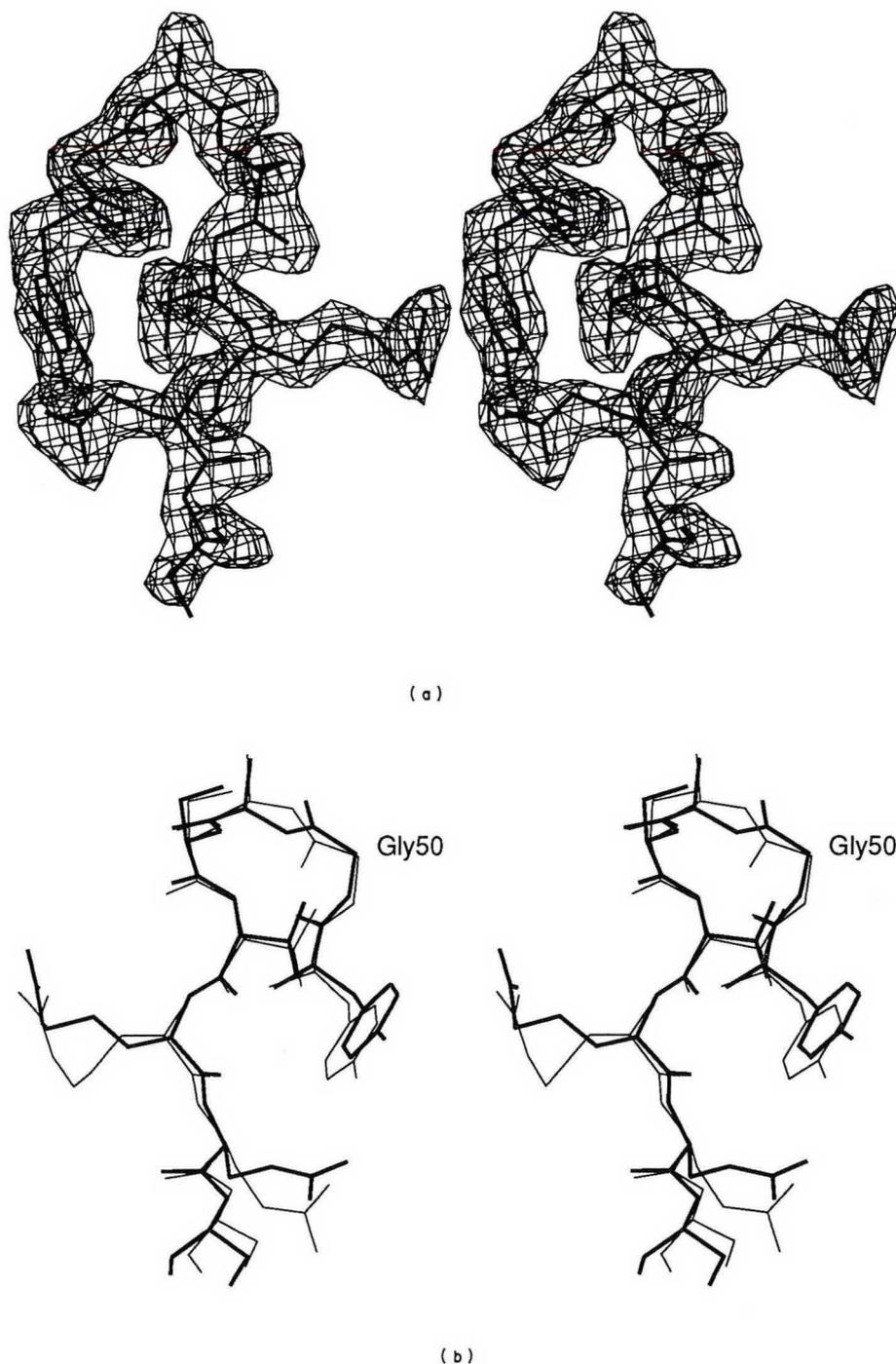


Figure 6. Stereoplots of the CDR2 region of V_L . (a) The residues from Tyr49 to Ser62 are shown. Ala51 is at the top of the loop and the view is from the side of the domain. An electron density map is displayed, contoured at a level of $0.4 \text{ e}/\text{\AA}^3$ (1.25σ). Density for the carbonyl oxygen atoms is clearly seen. (b) Comparison of the V_L -D and V_L -Fab CDR2. The view is from the front as in Figure 3. V_L -D co-ordinates are drawn with bold lines; V_L -Fab co-ordinates are drawn with thin lines.

representative structure might be sufficient. These key residues would contain the relevant information to generate a typical "canonical" fold, regardless of the rest of the sequence. By this comparative modelling strategy, it has been possible to predict the conformation of some CDR loops prior to the publication of the experimental structure (Chothia *et al.*, 1989). If the implied causal relationship between the observed key residues and the three-

dimensional loop structure was of general significance, this would provide us with an instance of being able to localize essential elements of folding information among the primary structure. Besides, the knowledge of precisely which residues are necessary and sufficient to generate a certain canonical fold, would prove extremely helpful in efforts aimed at predicting antibody structure from sequence data alone, with a view to elucidating molecular inter-

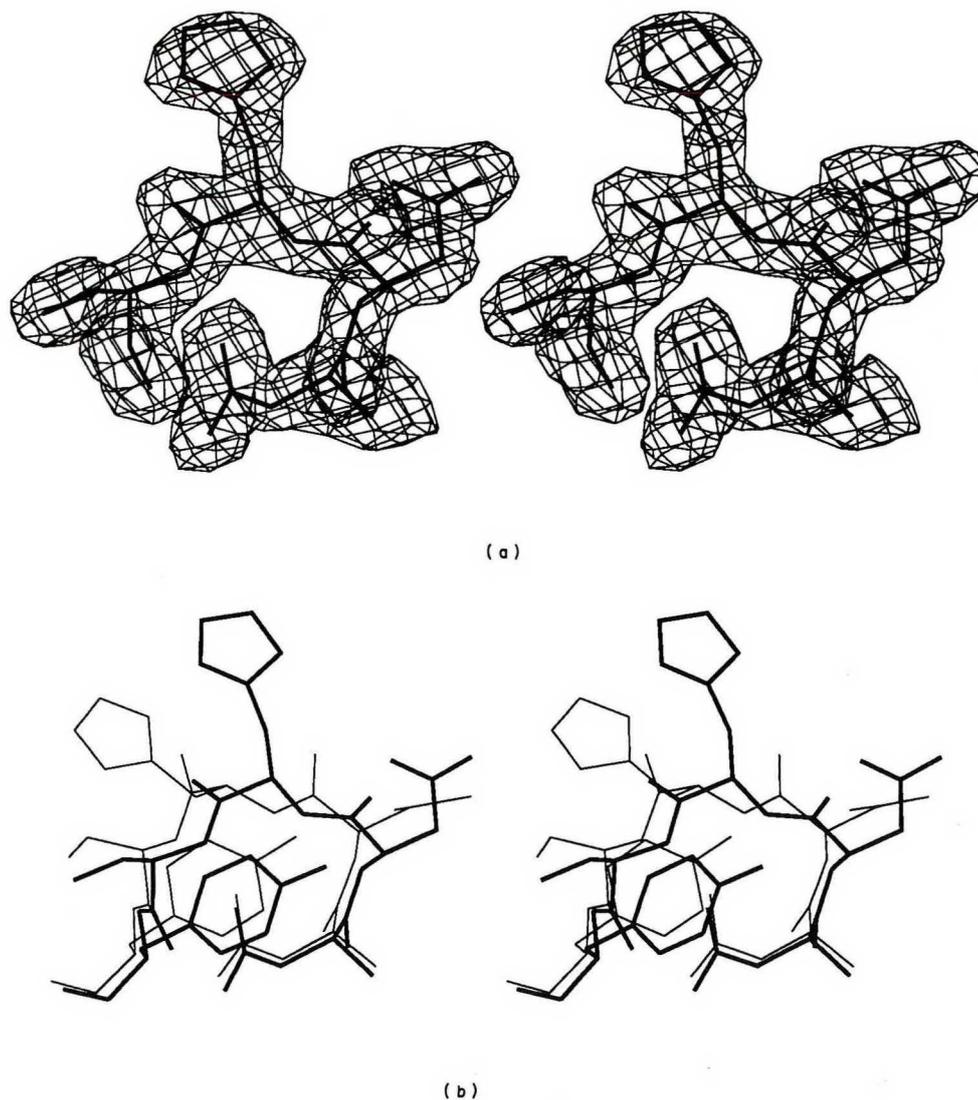


Figure 7. Comparison of the V_L -D and V_L -Fab CDR3. (a) The residues from Asn90 to Ser93 are shown. His92 is at the top and the view is from the front of the domain. An electron density map is displayed, contoured at a level of $0.4 \text{ e}/\text{\AA}^3$ (1.25σ). (b) The residues from Asn90 to Tyr94 are shown. V_L -D co-ordinates are drawn with bold lines; V_L -Fab co-ordinates are drawn with thin lines.

actions with a given hapten or antigen. Indeed, once an antibody structure is precisely known, a docking procedure may provide the correct mode of association with the antigen in favorable cases (Goodsell & Olson, 1990).

How well is the canonical structure hypothesis supported by the findings of this X-ray crystallographic study?

(1) The CDR1, arching across the top of the β -barrel, is found to be pinned into the core of the structure, in a fashion typical for CDR1s previously observed, by a conserved hydrophobic amino acid (Leu29 in this case). Leu29 is tightly packed into the core with low B -values and the CDR1 loop up to Leu30 is well stabilized. On the other hand, the long solvent-exposed loop, continuing from Asn31 is hardly stabilized at all. It had been necessary to postulate a separate canonical structure for 4-4-20 (Chothia *et al.*, 1989), which is one amino acid

residue shorter than McPC603 and has a conformation very different from that of McPC603 (Herron *et al.*, 1989). M3 has the same length of the CDR1 as 4-4-20, but the conformation is found again to be quite different, so that yet another canonical structure should have to be introduced. It is interesting to note that the CDR1 B -values of 4-4-20 are much lower than those observed for either M603 or M3. Since the 4-4-20 loop does not seem to have significantly more stabilizing interactions with the domain scaffold than M603 or M3, other than the approach to 3.5 \AA from an arginine residue of a symmetry-related domain in the crystal and a possible participation of His31 in hapten binding, this argues for an important role for internal stabilization of the loop itself. Three hydrogen bonds in 4-4-20 could be important in this context: the first bonds O Gln31b to Oⁿ Tyr32 (which is Phe32 in M3). The other two bridge the back of the

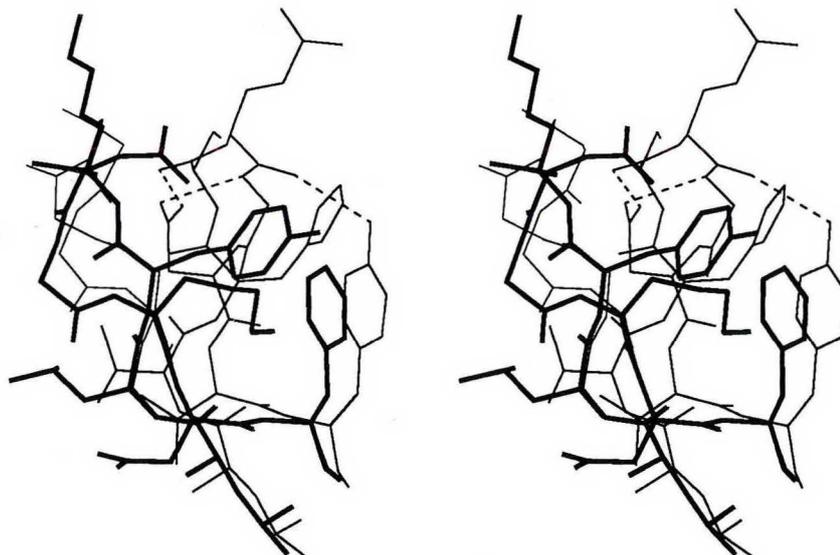


Figure 8. Comparison of the M3 and 4-4-20 V_L -Fab CDR1 conformations. Residues from 29 to 32 are shown. The view is from the right-hand side as in Fig. 2. M3 co-ordinates are drawn with thin lines, 4-4-20 co-ordinates are drawn with bold lines. The RMSD for the 36 backbone atoms from Leu29 to Phe32 is 3.75 Å. The V_L -D CDR1 conformation is drawn with thin lines for comparison. It closely corresponds to the M3 L1 conformation. Hydrogen bonds that possibly stabilize the loop conformation of 4-4-20 are also shown.

loop with $N^{\delta 2}$ Asn31e (which is Lys31e in M3) to O Ser31a and $O^{\delta 1}$ Asn31e to N Gly31d, in a conformation reminiscent of the stabilization of some CDR3 loops (Fig. 8). In the face of some residual uncertainty about the accuracy of the 4-4-20 model in this region and until experimental data on the relative importance of Tyr32, Asn31e or other residues within the loop for stabilizing this structure are available, we would suggest that the structure of loops inserted between position 30 and 32 should be considered to be unpredictable.

(2) In the structure of the CDR2, Chothia & Lesk (1987) had postulated that the variant conformation observed in McPC603 was linked to the presence of a glycine residue (Gly50). Our data on the V_L -D structure indicate, however, that the CDR2 conformation is highly conserved among all V_L domains, thus unifying the observed canonical structures, and a glycine residue at position 50 does not generate an exception.

(3) The canonical structure of this CDR3 has been postulated to be determined by the conserved *cis*-proline Pro95 and two hydrogen bonds from the carboxamide group of Asn90 to the back of the loop (Tramontano *et al.*, 1989). We observe a different topology of hydrogen bonds (which are, in addition, all longer than 3.0 Å) from Asn90 in this structure. Thus, we would suggest that the conformational importance of the hydrogen-bond stabilization of this solvent-exposed loop is not yet completely clear. Furthermore, the significant structural differences between the V_L -D and the V_L -Fab CDR3 indicate that the basic premise of the canonical-structure hypothesis has to be viewed with some caution. A strong influence of V_H (or the second V_L) on the conformation of a CDR (possibly mediated in this case through electrostatic effects involving

Asp91) would mean that this conformation would depend both on the precise mode of association between the two domains in the homo- or heterodimer. This would be especially problematic if a CDR conformation would depend on the precise conformation of the V_H CDR3. Both cannot be ruled out and both may be impossible to predict with sufficient accuracy (Stevens *et al.*, 1988; Colman *et al.*, 1987).

In conclusion, the potential of the canonical-structure hypothesis to predict antigen-binding loops from sequence data alone with sufficient accuracy to permit modelling of binding interactions, must still be viewed with some caution, and it becomes clear that much further structural work will be necessary. An attempt to rationalize the structural differences between the M3 and 4-4-20 CDR1 illustrates the difficulty of drawing conclusions from the observation of mere sequence-structure correlations, as too little is known about true causal relationships. The conformation of the CDR2 and its lack of correlation with a glycine residue at position 50 illustrates the limitation of such knowledge-based conformational predictions that rest in the accuracy and reliability of the underlying experimental structures. Finally, the difference in CDR3 conformation between the V_L -D and the V_L -Fab illustrates the difficulties for conformational predictions in cases where the local structure may be significantly influenced by long-range interactions or quaternary structure.

The finding that a CDR mutation such as in M3 can have negligible structural consequences for the rest of the domain is encouraging for further engineering: experiments to transplant and combine loops of known structure to generate new binding properties and efforts to solve the structure of new

sequences may one day lead to an improved understanding of sequence-structure relationships in immunoglobulin domains.

References

- Alzari, P. M., Lascombe, M. B. & Poljak, R. J. (1988). 3-dimensional structure of antibodies. *Annu. Rev. Immunol.* **6**, 555-580.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein databank: a computer based archival file for macromolecular structure. *J. Mol. Biol.* **112**, 535-542.
- Bruccoleri, R. E., Haber, E. & Novotny, J. (1988). Structure of antibody hypervariable loops reproduced by a conformational search algorithm. *Nature (London)*, **335**, 564-568.
- Chothia, C. & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823-826.
- Chothia, C. & Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196**, 901-917.
- Chothia, C., Novotny, J., Bruccoleri, R. & Karplus, M. (1985). Domain association in immunoglobulin molecules. *J. Mol. Biol.* **186**, 651-633.
- Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E., Davies, D. R., Tulip, W. R., Colman, P., Spinelli, S., Alzari, P. M. & Poljak, R. J. (1989). Conformations of immunoglobulin hypervariable regions. *Nature (London)*, **342**, 877-883.
- Colman, P. M., Schramm, H. J. & Guss, J. M. (1977). Crystal and molecular structure of the dimer of variable domains of the Bence-Jones Protein ROY. *J. Mol. Biol.* **116**, 73-79.
- Colman, P. M., Laver, W. G., Varghese, J. N., Baker, A. T., Tulloch, P. A., Air, G. M. & Webster, R. G. (1987). Three-dimensional structure of a complex of antibody with influenza-virus neuraminidase. *Nature (London)*, **326**, 358-363.
- Davies, D. R., Padlan, E. A. & Sheriff, S. (1990). Antibody-antigen complexes. *Annu. Rev. Biochem.* **59**, 439-473.
- de la Paz, P., Sutton, B. J., Darsley, M. J. & Rees, A. R. (1986). Modelling the combining sites of three antilysozyme monoclonal antibodies and of the complex between one of the antibodies and its epitope. *EMBO J.* **5**, 415-425.
- Ely, K. R., Herron, J. N., Harker, M. & Edmundson, A. B. (1989). Three-dimensional structure of a light chain dimer crystallized in water. *J. Mol. Biol.* **210**, 601-615.
- Epp, O., Lattman, E. E., Schiffer, M., Huber, R. & Palm, W. (1975). The molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2.0 Å resolution. *Biochemistry*, **14**, 4943-4952.
- Fehlhammer, H., Schiffer, M., Epp, O., Colman, P. M., Lattman, E. E., Schwager, P. & Steigemann, W. (1975). The structure determination of the variable portion of the Bence-Jones protein AU. *Biophys. Struct. Mechanism*, **1**, 139-146.
- Fine, R. M., Wang, H., Shenkin, P. S., Yarmush, D. L. & Levinthal, C. (1986). Predicting antibody hypervariable loop conformations. II. Minimization and molecular dynamics studies of McPC603 from many randomly generated loop conformations. *Proteins*, **1**, 342-462.
- Furey, W., Wang, B. C., Yoo, C. S. & Sax, M. (1983). Structure of a novel Bence-Jones protein (RHE) fragment at 1.6 Å resolution. *J. Mol. Biol.* **167**, 661-692.
- Glockshuber, R., Steipe, B., Huber, R. & Plückthun, A. (1990). Crystallization and preliminary X-ray studies of the V_L domain of the antibody McPC603 produced in *E. coli*. *J. Mol. Biol.* **213**, 613-615.
- Goodsell, D. S. & Olson, A. J. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins*, **8**, 195-202.
- Herron, J. N., He, X.-H., Mason, M. L., Voss, E. W. & Edmundson, A. B. (1989). Three-dimensional structure of a fluorescein-Fab complex crystallized in 2-methyl-2,4-pentanediol. *Proteins*, **5**, 271-280.
- Holm, L., Laaksonen, L., Kaartinen, M., Teeri, T. T. & Knowles, J. C. (1990). Molecular modelling study of antigen binding to oxazolone-specific antibodies: the Ox1 idiotype IgG and its mature variant with increased affinity to 2-phenyloxazolone. *Protein Eng.* **3**, 403-409.
- Huber, R. & Kopfmann, G. (1969). A method of absorption correction by X-ray-intensity measurements. *Acta Crystallogr. sect. A*, **25**, 143-152.
- Ippolito, J. A., Alexander, R. S. & Christianson, D. W. (1990). Hydrogen bond stereochemistry in protein structure and function. *J. Mol. Biol.* **215**, 457-471.
- Jack, T. & Levitt, M. (1978). Refinement of structures by simultaneous minimization of energy and R-factor. *Acta Crystallogr. sect. A*, **34**, 931-935.
- Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268-272.
- Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesmann, K. (1987). Sequences of proteins of immunological interest, Public Health Service, Natl. Inst. of Health, Bethesda, MD, U.S.A.
- Luzatti, V. (1952). Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallogr. Sect. A*, **5**, 802-810.
- Mainhart, C. R., Potter, M. & Feldman, R. J. (1984). A refined model for the variable domains (F_v) of the J539 β(1, 6)-D-galactan-binding immunoglobulin. *Mol. Immunol.* **21**, 469-478.
- Marquart, M. & Deisenhofer, J. (1982). The three-dimensional structure of antibodies. *Immunol. Today*, **3**, 160-166.
- Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. (1980). Crystallographic refinement and atomic models of the intact immunoglobulin molecule KOL and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. *J. Mol. Biol.* **141**, 369-391.
- Martin, A. C. R., Cheetham, J. C. & Rees, A. R. (1989). Modelling antibody hypervariable loops: a combined algorithm. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 9268-9272.
- Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **244**, 491-497.
- Messerschmitt, A. & Pflugrath, J. W. (1987). Crystal orientation and X-ray pattern prediction routines for area-detector diffractometer systems in macromolecular crystallography. *J. Appl. Crystallogr.* **20**, 306-315.
- Messerschmidt, A., Schneider, M. & Huber, R. (1990). ABSCOR: a scaling and absorption correction program for the FAST area detector diffractometer. *J. Appl. Crystallogr.* **23**, 436-439.

- Milner-White, J. E., Ross, B. M., Ismail, R., Khaled, B.-M. & Poet, R. (1988). One type of gamma-turn, rather than the other gives rise to chain-reversal in proteins. *J. Mol. Biol.* **204**, 777–782.
- Padlan, E. A., Silvertown, E. W., Sheriff, S., Cohen, G. H., Smith-Gill, S. J. & Davies, D. R. (1989). Structure of an antibody-antigen complex: crystal structure of the HyHEL-10 Fab-lysozyme complex. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 5938–5942.
- Perlmutter, R. M., Crews, S. T., Douglas, R., Sorensen, G., Johnson, N., Nivera, N., Gearhart, P. J. & Hood, L. (1984). The generation of diversity in phosphorylcholine-binding antibodies. *Advan. Immunol.* **35**, 1–37.
- Plückthun, A. (1991). Antibody engineering: advances from the use of *Escherichia coli* expression systems. *Bio/technology*, **9**, 545–551.
- Potter, M. (1977). Antigen-binding myeloma proteins of mice. *Advan. Immunol.* **25**, 141–211.
- Ramachandran, G. N., Venkatachalam, C. M. & Krimm, S. (1966). Stereochemical criteria for polypeptide and protein conformations. *Biophys. J.* **6**, 849–872.
- Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–339.
- Riechmann, L., Clark, M., Waldmann, H. & Winter, G. (1988). Reshaping human antibodies for therapy. *Nature (London)*, **332**, 323–327.
- Rose, G. D., Gierasch, L. M. & Smith, J. A. (1985). Turns in peptides and proteins. *Advan. Protein Chem.* **37**, 1–109.
- Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). Phosphocholine binding immunoglobulin Fab McPC603. *J. Mol. Biol.* **190**, 593–604.
- Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. R. (1974). The three-dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen binding site. *Proc. Nat. Acad. Sci., U.S.A.* **71**, 4298–4302.
- Shenkin, P. S., Yarmush, D. L., Fine, R. M., Wang, H. & Levinthal, C. (1987). Predicting antibody hyper-variable loop conformations. I. Ensembles of random conformations for ringlike structures. *Biopolymers*, **26**, 2053–2085.
- Sheriff, S., Silvertown, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. C. & Davies, D. R. (1987). Three-dimensional structure of an antibody/antigen complex. *Proc. Nat. Acad. Sci., U.S.A.* **84**, 8075–8079.
- Singh, J., Thornton, J. M., Snarey, M. & Campbell, S. F. (1987). The geometries of interacting arginine-carboxyls in proteins. *FEBS Letters*, **224**, 161–171.
- Skerra, A. & Plückthun, A. (1988). Assembly of a functional Fv fragment in *Escherichia coli*. *Science*, **240**, 1038–1041.
- Snow, M. E. & Amzel, L. M. (1986). Calculating three-dimensional changes in protein structure due to amino acid substitutions: the variable region of immunoglobulins. *Proteins*, **1**, 267–279.
- Stadlmüller, J. (1991). Das Fv-Fragment des Phosphorylcholin-bindenden Antikörpers McPC603: Mutanten, Bindungseigenschaften und katalytische Aktivität. Ph.D thesis, Ludwig-Maximilians-University, Munich, F.R.G.
- Stanford, J. M. & Wu, T. T. (1981). A predictive method for determining possible three-dimensional foldings of immunoglobulin backbones around antibody combining sites. *J. Theoret. Biol.* **88**, 421–439.
- Steigemann, W. (1974). Die Entwicklung und Anwendung von Rechenverfahren und Rechenprogrammen zur Strukturanalyse von Proteinen am Beispiel des Trypsin-Trypsininhibitor Komplexes, des freien Inhibitors und der L-Asparaginase. Ph.D thesis, Technical University, Munich, F.R.G.
- Stevens, F. J., Chang, C.-H. & Schiffer, M. (1988). Dual conformations of an immunoglobulin light-chain dimer: heterogeneity of antigen specificity and idiotype profile may result from multiple variable-domain interaction mechanisms. *Immunology*, **85**, 6895–6899.
- Suh, S. W., Bhat, T. N., Navia, M. A., Cohen, G. H., Rao, D. N., Rudikoff, S. & Davies, D. R. (1986). The galactan binding immunoglobulin Fab J539: an X-ray diffraction study at 2.6 Å resolution. *Proteins*, **1**, 74–80.
- Tramontano, A., Chothia, C. & Lesk, A. M. (1989). Structural determinants of the conformations of medium-sized loops in proteins. *Proteins*, **6**, 382–394.
- Venkatachalam, C. M. (1968). Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers*, **6**, 1425–1436.
- Weiner, S. J., Singh, U. C., O'Donnell, T. J. & Kollman, P. A. (1984). Quantum and molecular mechanical studies on alanyl dipeptide. *J. Amer. Chem. Soc.* **106**, 6243–6245.
- Zehfus, M. H. & Rose, G. D. (1986). Compact units in proteins. *Biochemistry*, **25**, 5759–5765.

