J. Mol. Biol. (1982) **160**, 325–342

Evolution of Proteins Formed by β-Sheets

II. The Core of the Immunoglobulin Domains

ARTHUR M. LESK[1,†] AND CYRUS CHOTHIA[1,2]

[1] *Medical Research Council*
Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH, England

[2] *William Ramsay, Ralph Foster and Christopher Ingold Laboratories*
University College London, Gower Street
London WC1E 6BT, England

(Received 5 April 1982)

We have analysed and compared the structures of 11 immunoglobulin domains available from high-resolution crystal-structure determinations. These include examples of the domain classes V_κ, V_λ, V_γ, C_λ, $C_\gamma1$, $C_\gamma2$ and $C_\gamma3$. The core of the structure of each of the 11 domains is formed by a set of 35 or 36 homologous residues that form two β-sheets packed face-to-face. Only three residues are common to all domains: two cysteines that form a disulphide bridge between the β-sheets, and a tryptophan that packs against them. The other interior residues tend to retain hydrophobic character but vary greatly in size (the mean variation is 60 Å^3) and shape.

We examined how the interface between the β-sheets accommodates these mutations. We describe examples of the following mechanisms.

(1) Displacements and rotations of the β-sheets relative to each other by up to 2 Å and 20°.

(2) Lateral insertion of side-chains from external loops into the interface region, to compensate for reductions in the volume of β-sheet residues.

(3) Insertion of a residue into a strand to form a β-bulge.

(4) Local changes in conformation.

(5) Only rarely, complementarity in adjacent mutations.

Although certain of these mechanisms are specific to the double β-sheet structure, the immunoglobulin domains, like other families of proteins, accommodate mutations of interior residues through substantial structural changes consistent with the preservation of their function.

1. Introduction

The immunoglobulins are a family of proteins found in vertebrates, which mediate the immune response to antigens. The structures of several whole molecules and separate domains have been determined by X-ray crystallography (Edmundson

† Permanent address: Fairleigh Dickinson University, Teaneck-Hackensack Campus, Teaneck, NJ 07666, U.S.A.

0022-2836/82/260325-18 $03.00/0

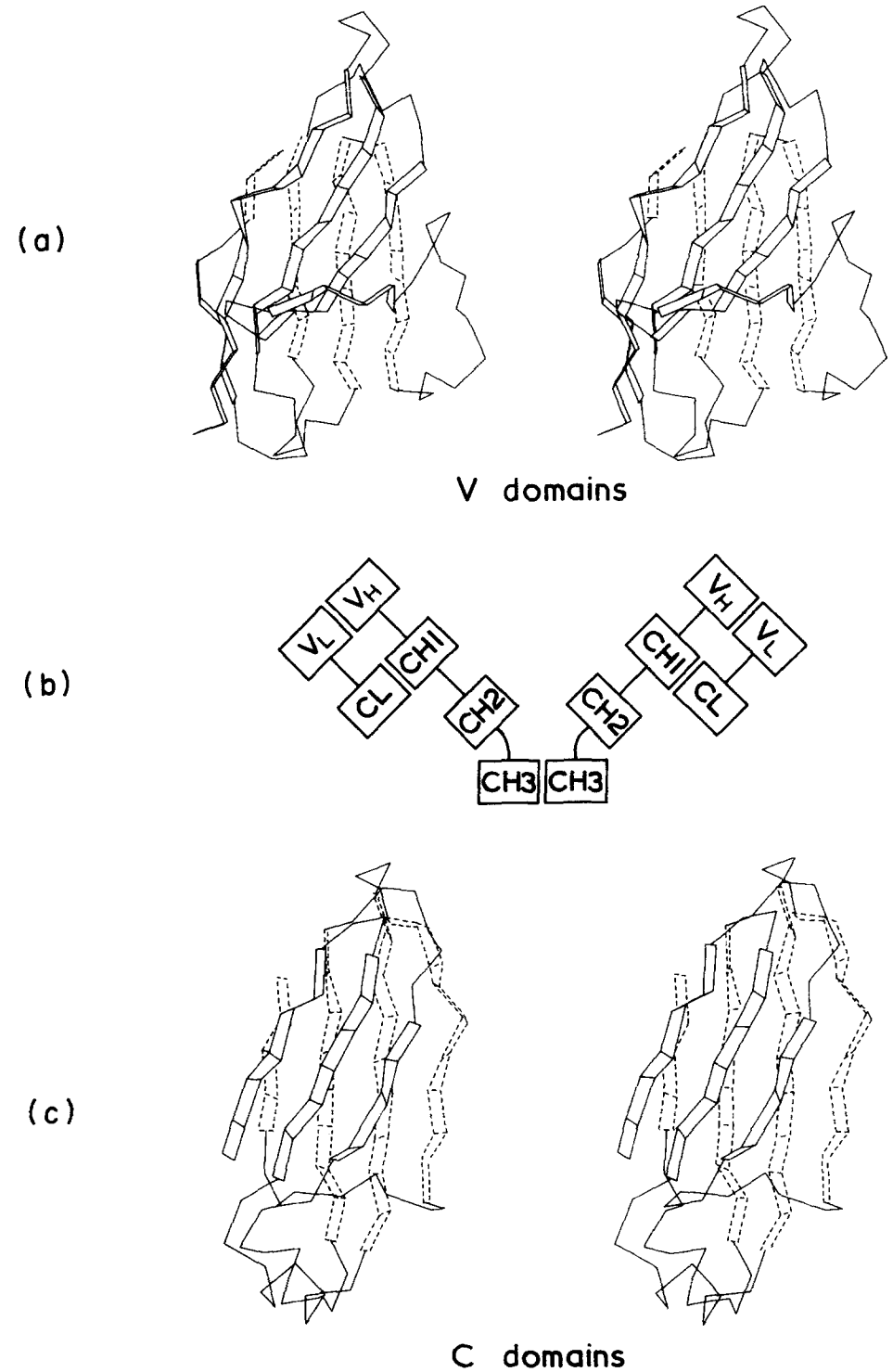


FIG. 1.

et al., 1975; Davies *et al.*, 1975; Huber *et al.*, 1976; Poljak, 1978). All immunoglobulin molecules consist of several domains which are believed to be derived from a common ancestor (Edelman & Gall, 1969). They have similar secondary and tertiary structures: each domain contains two β-sheets packed face-to-face and pinned together by a disulphide bridge (Fig. 1). However, the domains vary greatly in amino acid sequence.

Two significant aspects to the variability in sequence are: first, the antibody combining site, which involves the end of the β-sheets and the loops between their strands, is subject to extensive insertion and deletion of amino acids as well as substitution. This is related to the versatility of the structure in creating binding sites to a wide variety of antigens. Here different amino acid sequences produce *different* three-dimensional structures.

Second, the residues packed in the region between the β-sheets show extensive substitution. Only three residues are invariant in all domains: the cysteines of a disulphide bridge between the sheets and a tryptophan that packs against it. Here, different amino acid sequences are producing *similar* secondary and tertiary structures. It is the purpose of this study to examine the ways in which this second class of substitutions is compatible with a conserved structure.

We used 11 immunoglobulin domains, the structures of which have been determined by X-ray crystallography (see Table 1).

TABLE 1

Immunoglobulin structures used in this work

	Classification of the domains			
Protein fragment	Class	Subgroup or subclass	Allotypes	Reference for the structure determination
Bence–Jones protein REI	V_κ	I		Epp *et al.* (1974)
Fab′ (NEW)	V_λ	I		Saul *et al.* (1978)
	C_λ			
	V_γ	II		
	$C_\gamma 1$	I	Gm (3)	
Fab (KOL)	V_λ	I		Marquart *et al.* (1980)
	C_λ			
	V_γ	III		
	$C_\gamma 1$	I	Gm (17)	
Fc	$C_\gamma 2$			Deisenhofer (1981)
	$C_\gamma 3$			

FIG. 1. The assembly of the domains of an immunoglobulin G molecule and the structures of typical V and C domains. (b) A schematic diagram of the immunoglobulin showing the two heavy chains and 2 light chains. Each light chain contains 2 domains, a variable domain V_L and a constant domain C_L. Each heavy chain contains 4 domains, a variable domain V_H and 3 constant domains CH1, CH2 and CH3. The chains are held together by disulphide bonds and by hydrophobic and van der Waals' interactions. (a) The structure of variable domains. Each contains 2 β-sheets packed face-to-face. In V domains, one sheet contains 5 strands and the other contains 4 strands. (c) The structure of constant domains. Each contains 2 β-sheets packed face-to-face. In C domains one sheet contains 3 strands and the other contains 4 strands. A hairpin loop in the 5-strand sheet of V domains is deleted in the C domains. (a) and (c) were drawn from the co-ordinates of V_L (KOL) and C_L (KOL), respectively (see Table 1).

2. Co-ordinates and Calculations

(a) *Sources of atomic co-ordinates*

The structures used in this work are listed in Table 1. Co-ordinates of V_κ (REI) and Fab′ (NEW) were taken from the Brookhaven National Laboratory Protein Data Bank (Bernstein *et al.*, 1977). Co-ordinates for Fab (KOL) and Fc were given to us by Drs J. Deisenhofer and R. Huber.

(b) *Computing techniques*

The visual study and analysis of the structures were facilitated by the computer graphics system in the Laboratory of Dr P. Pauling at University College London, using a program system designed and written by Mr D. Richardson.

Determination and analysis of the pattern of contacts were accomplished using programs written by Dr M. Levitt.

Accessible surface areas (Lee & Richards, 1971) were computed in a manner described previously (Chothia, 1976) using a program of Dr M. Levitt.

Superpositions of structures were carried out by minimizing the mean-square deviation of selected sets of corresponding atoms by the method of Kabsch (1976).

3. Structure of Immunoglobulin Molecules and Their Domains

(a) *The overall geometry*

A complete immunoglobulin G molecule (IgG) contains four polypeptide chains. The two light chains contain two domains each; the two heavy chains contain four domains each (see Fig. 1). (Each domain contains approx. 100 residues.) The light and heavy chains are held together by disulphide bridges, and by non-covalent interactions.

The domains at the amino-terminal ends of the light and heavy chains form the antigen binding site. Because of the large number of sequences found in any species, these are called the *variable* domains (abbreviated V_L and V_H). The other domains, for which only a few different sequences occur, are called the *constant* domains: abbreviated CL in the light chain and CH1, CH2 and CH3 in the heavy chain.

Similarities among amino acid sequences distinguish two classes of light chains, κ and λ, and five classes of heavy chains, μ, α, γ, δ and ϵ. V domains vary in length, as a result of insertion and deletion of residues, but all known V domains have similar secondary and tertiary structure: a five-strand β-sheet and a four-strand β-sheet packed face-to-face (Fig. 1). The secondary and the tertiary structures of C domains are similar to those of V domains, except that a hairpin loop is absent from the edge of one of the sheets, reducing it from five strands in the V domains to three strands in the C domains. Topologically, the pattern of linkages between strands is the same in all domains (Fig. 1).

The structures used in this analysis contain examples of V_κ, V_λ, V_γ, C_λ, $C_\gamma1$, $C_\gamma2$ and $C_\gamma3$ domains (see Table 1).

(b) *Analysis and alignment of the sequences*

Three residues are common to all immunoglobulin domains: two cysteines and a tryptophan at the centre of the β-sheets. The cysteines form a disulphide bridge linking the sheets; and the tryptophan, also a β-sheet residue, packs its side-chain against the cysteines. These three invariant residues provide a marker useful in superimposing different domains to determine which residues occupy geometrically similar positions.

By pairwise superpositions of the 11 domains listed in Table 1, we found that the structure of a domain can be divided into three parts (Fig. 2).

(1) The *pin*. This is a group of atoms present in all domains, at the centre of the region between the two β-sheets. It includes the disulphide bridge, the invariant Trp, a residue near one of the cysteines that is Val, Ile or Leu, and portions of five other residues (see Fig. 3). These form a compact contiguous parcel of 56 atoms, the structure of which is very similar but not identical in all the domains we have analysed. Superpositions of the pin atoms from different pairs of domains gave root-mean-square deviations of 0·4 to 0·9 Å.

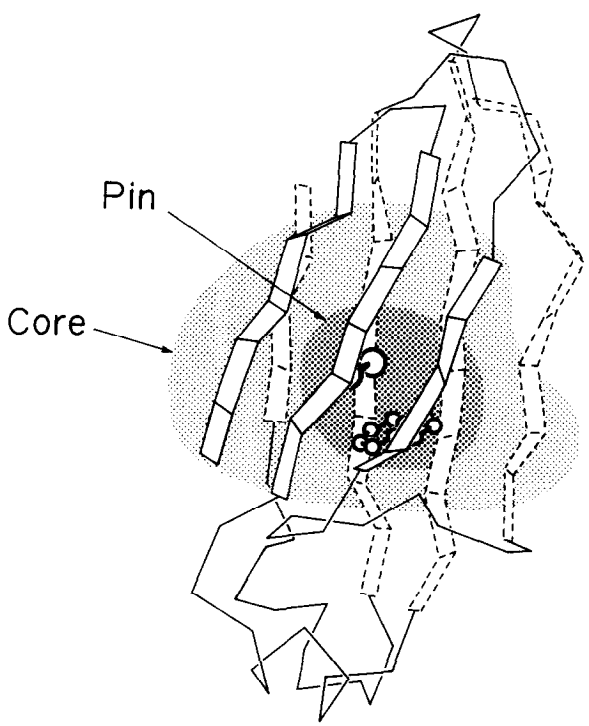


FIG. 2. The pin, the core and the periphery of the immunoglobulin domains. The pin is a compact parcel of 56 atoms present at the centre of all immunoglobulin domains and which retains a nearly constant structure. The pin includes the constant disulphide bridge between the sheets, the constant tryptophan that packs against it, and portions of 6 other residues (see Fig. 3). The core contains the 35 residues that are in strands of β-sheet in all domains (including the pin residues). The periphery contains the remainder of the residues, i.e. those outside the core.

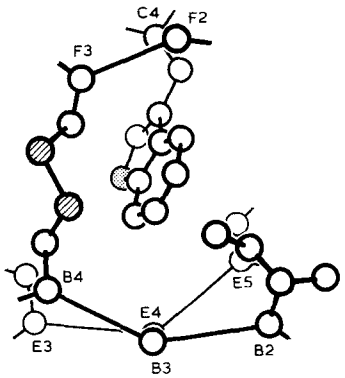


FIG. 3. The structure of the pin in V_K (REI). The pin contains all or part of 9 residues from 4 strands of sheet: B2 (Ile, Val or Leu), B3 (main-chain atoms only), B5 (Cys), E3 (main-chain atoms+C^β), E4 (main-chain atoms only), E5 (main-chain atoms+$C\beta$), C4 (Trp), F2 (main-chain atoms only) and F3 (Cys). In this illustration the backbone is represented by C^α atoms only.

(2) The β-sheet *core* structure. This comprises the residues in positions that are in β-sheet strands in all domains. It includes the "pin" residues. The core contains 35 residues in C and V domains, plus an additional residue in V domains that forms a β-bulge (Richardson *et al.*, 1978) in a strand at the edge of one of the sheets. (This definition is similar but not identical to the use of the term "core" by Novotný & Franěk (1975).)

We shall discuss the superpositions of the core structures in section 5(a), below.

(3) The *periphery* includes all residues outside the core: about 70 residues in V domains and 60 residues in C domains. Approximately one-third of these are in strands of β-sheet and represent dissimilarities of secondary structure between V and C domains or between individual domains in either class. The other residues in the periphery are in connecting loops between strands of sheet; these vary greatly in length, sequence and conformation.

4. Residues that Form the β-Sheet Core Structure

Table 2 displays a sequence alignment of the β-sheet core residues of the 11 domains studied. The alignment of the Fab' (NEW) domains is the same as that given by Poljak *et al.* (1974) and by Beale & Feinstein (1976). Strands are designated by letters, A, B, C, etc. in order of appearance in the sequence. Thus, the core residues of one sheet are in strands A, B, D and E, and the core residues of the other are in strands C, F and G. We denote positions of residues in the strands by A1, A2, etc. (see Table 2). The core of the V domains contains an additional residue in a β-bulge in the G strand that does not occur in the C domains.

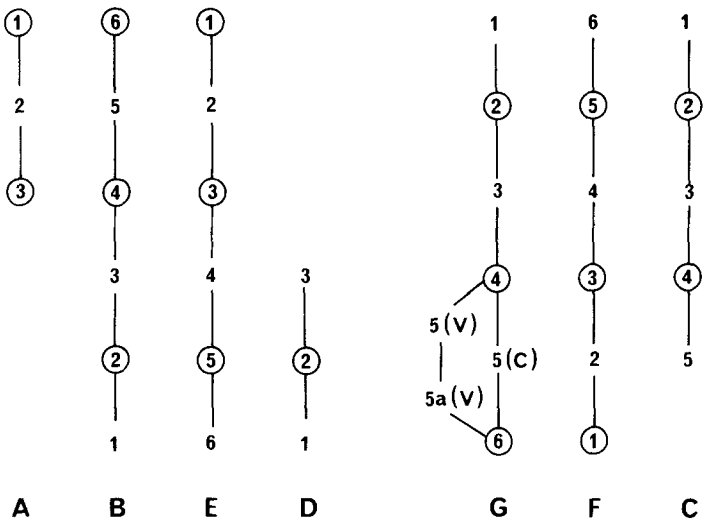
(a) *Conservation and chemical nature of the core residues*

When the sequences of residues in the core of domains of different classes are compared pairwise, 10 to 21 residues are identical between two V domains. 12 to 18

TABLE 2

Alignment of the residues in the β-sheet core structures

A. *Plan of the β-sheets*



Circled residues form the interface between the β-sheets

B. *Residues in the β-sheets*

Strand:	A	B	C	D
Position:	1 2 3	1 2 3 4 5 6	1 2 3 4 5	1 2 3
V_κ (REI)	^{4}MTQ	^{20}TITCQA	^{32}YLNWY	^{63}SGS
V_λ (NEW)	^{4}LTQ	^{19}TISCTG	^{31}HVKWY	^{62}SVS
V_λ (KOL)	^{4}LTQ	^{19}TISCSG	^{31}SVNWY	^{62}SGS
V_γ (NEW)	^{4}LEQ	^{19}SLTCTV	^{33}YYTWV	^{68}TML
V_γ (KOL)	^{304}LVQ	^{319}RLSCSS	^{333}AMYWV	^{369}TIS
C_λ (NEW)	^{117}VTL	^{133}TLVCLI	^{147}TVAWK	^{160}GVE
C_λ (KOL)				
$C_\gamma1$ (NEW)	$^{125}_{426}$VFP	$^{141}_{442}$ALGCLV	$^{155}_{456}$TVSWN	$^{166}_{467}$GVH
$C_\gamma1$ (KOL)				
$C_\gamma2$	^{240}VFL	^{258}EVTCVV	^{274}KFNWY	^{288}KTK
$C_\gamma3$	^{348}VYT	^{364}SLTCLV	^{378}AVEWE	^{390}NYK

Strand:	E	F	G
Position:	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 5a 6
V_κ (REI)	^{69}TDYTFT	^{86}YYCQQY	^{96}YTFGQGT
V_λ (NEW)	^{68}SSATLA	^{85}YYCQSY	^{95}RVFGGGT
V_λ (KOL)	^{68}TSASLG	^{85}YYCASW	^{97}YVFGTGT
V_γ (NEW)	^{76}NQFSLR	^{93}YYCARN	^{105}DVWGQGS
V_γ (KOL)	^{377}NTLFLQ	^{394}YFCARD	^{406}DYWGQGT
C_λ (NEW)	^{176}ASSYLS	^{193}YSCQVT	^{203}TVEK T V
C_λ (KOL)			
$C_\gamma1$ (NEW)	$^{182}_{483}$LSSVVT	^{198}YICNVN	$^{210}_{511}$KVDK R V K
$C_\gamma1$ (KOL)			
$C_\gamma2$	^{302}VVSVLT	^{319}YKCKVS	^{331}PIEK T I
$C_\gamma3$	^{406}LYSKLT	^{423}FSCSVM	^{436}YTQK S L

residues are identical between two C domains, and 5 to 11 residues are identical between a V and a C domain. Of these, only the two cysteines (B4 and F3) and Trp (C4) are common to all structures. At position F1 the 11 domains studied here have only Tyr or Phe, but the more extensive alignments of Beale & Feinstein (1976), Barker *et al.* (1978) and Kabat *et al.* (1979) show that Val, Ile, Thr, Ala, Ser and His can occur at this position.

Because of the alternating directions of side-chains in β-sheet regions, residues at 17 of the 35 or 36 core positions point their side-chains into the region between the sheets. These occupy positions:

A1, A3; B2, B4, B6; C2, C4; D2;
E1, E3, E5; F1, F3, F5; G2, G4, G6.

At 13 of these positions residues are non-polar or Ser, Thr or Tyr. At the other four positions (A3, E1, F5 and G4), Gln, Asn, Arg and Lys can occur. Residues at three positions (A3, E1 and F5) are at the edge of a sheet and the ends of the side-chains are exposed to solvent in all cases. The Gln side-chains at position G4 are buried and form internal hydrogen bonds. We shall discuss this residue in more detail in section 5(c), below.

(b) *Variation of side-chain volumes*

There is substantial variation in side-chain volume among homologous residues occupying inward-pointing positions in the core. Figure 4 shows the distribution of ranges of volume at these 17 positions. (For any position this range is the difference between the volumes of the largest and smallest residues found at that position in the 11 domains studied (see Table 2).) The mean variation is 60 Å^3 with a standard deviation of 45 Å^3.

Inspection of the individual structures shows that in general there is no local compensation of these volume changes by complementary mutations in adjacent residues. For example, residues at positions A1, B6 and E1 in one β-sheet pack against residues at G2, F5 and C2 in the other β-sheet. A comparison of V_λ (NEW) and $C_\gamma 2$ shows:

V_λ (NEW)			$C_\gamma 2$		
Leu4	Gly24	Ser68	Val240	Val263	Val302
A1	B6	E1	A1	B6	E1
Val98	Ser89	Val32	Ile332	Val323	Phe275
G2	F5	C2	G2	F5	C2

On going from V_λ (NEW) to $C_\gamma 2$ the side-chain volume increases at each of these positions except A1. In the next section we examine the mechanisms by which such volume changes are accommodated.

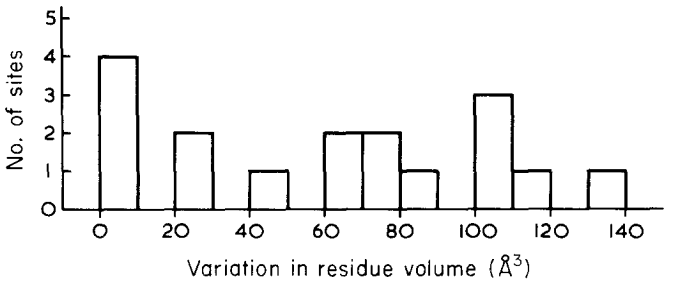


FIG. 4. The variation in volume of residues in the 17 interface positions of the core of the immunoglobulin domains. The volume variation is the difference in volume between the smallest and largest residues that occur at that position (Table 2). Residue volumes used here are listed by Lesk & Chothia (1980).

5. Structural Accommodations of Mutations

In the previous section we have shown that the residues forming the core of the immunoglobulin domain can vary in chemical nature and in size. How does this family of molecules retain the dense packing of side-chains in the region between the sheets? The mechanisms by which the immunoglobulin structure accommodates these mutations include movements of the sheets relative to one another, the insertion of residues from the periphery into the region between the sheets, local conformational changes and, rarely, local complementarity of mutations. We shall illustrate these changes in this section.

(a) *Shifts in the relative position of the packed β-sheets*

The core structure consists of one β-sheet formed of strands A, B, D and E, and another β-sheet formed of strands C, F and G. For each pair of domains we superimposed the homologous backbone atoms of the individual sheets and of both sheets together (see Table 3). The root-mean-square (r.m.s.) deviations lie in the range 0·4 to 1·0 Å for the pairs of ABDE sheets, 0·4 to 1·8 Å for the CFG sheets, and 0·5 to 1·9 Å for both sheets. In several cases the r.m.s. deviation after superposition of both sheets is larger than the corresponding deviations of the individual sheets (e.g. $C_\gamma 2$ *versus* $C_\gamma 3$; see Table 3). This indicates that there has been a change in the relative position or orientation of one sheet with respect to the other†.

We determined the changes in relative geometry of the two sheets by a double

† The reasoning is as follows. Suppose that structure X_1 fits structure Y_1 with an r.m.s. deviation 0·5 Å and structure X_2 fits structure Y_2 with an r.m.s. deviation 0·5 Å. If X_1 and X_2 have the same relative disposition in space as Y_1 and Y_2, then the r.m.s. deviation of the fit of X_1+X_2 to Y_1+Y_2 will also be 0·5 Å. But if there are shifts or rotations in the substructures relative to each other, the result will be to increase the r.m.s. deviation of the fit of X_1+X_2 to Y_1+Y_2. More formally, if Δ_1 is the r.m.s. deviation of the superposition of the N_1 atoms of X_1 and Y_1, and Δ_2 is the r.m.s. deviation of the superposition of the N_2 atoms of X_2 and Y_2, and if the rigid body motions (translation + rotations) for the optimal superpositions of X_1 onto X_2 and Y_1 onto Y_2 are the *same*, then Δ_{12}, the r.m.s. deviation of the optimal superposition of X_1+X_2 and Y_1+Y_2 will be equal to $[(N_1\Delta_1^2+N_2\Delta_2^2)/(N_1+N_2)]^{\frac{1}{2}}$. In this formula, if $\Delta_1=\Delta_2$, then $\Delta_{12}=\Delta_1=\Delta_2$. If *different* rigid-body motions are required for optimal superpositions of X_1 and Y_1 and of X_2 and Y_2, then $\Delta_{12} > [(N_1\Delta_1^2+N_2\Delta_2^2)/(N_1+N_2)]^{\frac{1}{2}}$.

TABLE 3

Fit of the immunoglobulin core structures

Class of comparison	Domains compared	r.m.s. difference in position of main-chain atoms after a least-squares fit (Å) ABED β-sheet	CFG β-sheet	Total core
Homologous domains†	V_λ (NEW) V_λ (KOL)	0·53	0·83	0·72
	V_γ (NEW) V_γ (KOL)	0·68	1·05	1·01
	C_λ (NEW) C_λ (KOL)	0·39	0·48	0·46
	$C_\gamma 1$ (NEW) $C_\gamma 1$ (KOL)	0·51	0·41	0·51
V domains	V_λ (NEW) V_γ (NEW)	0·62	1·48	1·49
	V_κ (REI) V_γ (NEW)	0·72	1·19	1·38
	V_λ (KOL) V_γ (NEW)	0·64	1·48	1·49
	V_κ (REI) V_γ (KOL)	0·67	0·73	0·94
	V_λ (NEW) V_γ (KOL)	0·66	1·22	1·14
	⋮ ⋮	⋮	⋮	⋮
	V_κ (REI) V_λ (KOL)	0·49	0·85	0·75
C domains	$C_\gamma 1$ (KOL) $C_\gamma 2$	0·80	1·06	1·06
	$C_\gamma 1$ (KOL) $C_\gamma 3$	0·66	0·87	0·89
	$C_\gamma 2$ $C_\gamma 3$	0·70	0·70	0·87
	$C_\gamma 1$ (NEW) C_λ (KOL)	0·74	1·07	1·02
	$C_\gamma 1$ (NEW) $C_\gamma 2$	0·84	1·13	1·09
	⋮ ⋮	⋮	⋮	⋮
	C_λ (KOL) $C_\gamma 2$	0·54	0·49	0·55
V–C domains	V_γ (NEW) $C_\gamma 1$ (NEW)	0·96	1·73	1·78
	V_γ (NEW) $C_\gamma 1$ (KOL)	0·85	1·76	1·80
	V_γ (NEW) C_λ (NEW)	0·78	1·49	1·57
	V_γ (NEW) C_λ (KOL)	0·72	1·34	1·40
	V_γ (KOL) $C_\gamma 1$ (NEW)	0·89	1·49	1·48
	V_γ (KOL) $C_\gamma 1$ (KOL)	0·87	1·49	1·51
	V_λ (NEW) $C_\gamma 3$	0·79	1·05	1·14
	V_λ (KOL) $C_\gamma 3$	0·75	0·86	1·01

We list here the values for: (1) homologous domains; (2) V–V, C–C and V–C pairs that show the largest difference in the relative position of their β-sheets (see Table 4); (3) the V–V and C–C pairs that show the smallest differences.

† The residue identity is 69% for the 2 V_λ domains, 47% for the V_γ, 96% for the C_λ and 98% for the C_γ.

superposition calculation. For each pair of domains, first we superimposed the ABDE sheets and then computed the additional shift and rotation required to superimpose the CFG sheets starting from the position and orientation produced by the ABDE superposition (see Table 4).

In discussing these results, it is important to distinguish the intrinsic effects of mutations on a structure, from the effects of different environments in the crystal, and of experimental errors in the co-ordinates. A "control experiment" that provides an estimate of the magnitude of extrinsic factors is provided by the independent determination of the very similar constant domains in Fab (KOL) and Fab′ (NEW). The sequences of the C_λ and $C_\gamma 1$ domains of these proteins differ by

TABLE 4

Differences in the relative position of the β-sheets in the core structure

Class of comparison	Domains compared	Shift required to superimpose main-chain atoms of the CFG β-sheets *after* the superposition of the ABED β-sheets	
		Translation (Å)	Rotation (deg.)
Homologous domains	V_λ (NEW) V_λ (KOL)	0·34	2·8
	V_γ (NEW) V_γ (KOL)†	1·14	10·1
	C_λ (NEW) C_λ (KOL)	0·15	2·8
	$C_\gamma 1$ (NEW) $C_\gamma 1$ (KOL)	0·37	5·0
V domains	V_λ (NEW) V_γ (NEW)	1·29	18·0
	V_κ (REI) V_γ (NEW)	1·20	17·2
	V_λ (KOL) V_γ (NEW)	1·60	16·0
	V_κ (REI) V_γ (KOL)	0·45	10·2
	V_λ (NEW) V_γ (KOL)	0·46	10·0
	⋮ ⋮	⋮	⋮
	V_κ (REI) V_λ (KOL)	0·41	3·9
C domains	$C_\gamma 1$ (KOL) $C_\gamma 2$	1·24	4·4
	$C_\gamma 1$ (KOL) $C_\gamma 3$	0·36	10·3
	$C_\gamma 2$ $C_\gamma 3$	1·17	7·9
	$C_\gamma 1$ (NEW) C_λ (KOL)	1·22	5·0
	$C_\gamma 1$ (NEW) $C_\gamma 2$	1·12	4·8
	⋮ ⋮	⋮	⋮
	C_λ (KOL) $C_\gamma 2$	0·36	3·2
V–C domains	V_γ (NEW) $C_\gamma 1$ (NEW)	1·97	21·7
	V_γ (NEW) $C_\gamma 1$ (KOL)	1·63	21·7
	V_γ (NEW) C_λ (NEW)	0·99	20·1
	V_γ (NEW) C_λ (KOL)	0·83	17·4
	V_γ (KOL) $C_\gamma 1$ (NEW)	0·75	16·1
	V_γ (KOL) $C_\gamma 1$ (KOL)	0·41	16·1

We list here values for (1) homologous domains (see also Table 5); (2) the V–V, C–C and V–C pairs that show the largest difference; (3) the V–V and C–C pairs with the smallest differences.

† Note that the subgroup of V_γ (NEW) is II and that of V_γ (KOL) is III.

four and two residues, respectively. These few changes are all in surface residues and are therefore not expected to affect the structure of the β-sheet core.

The results of the superpositions of the backbones of the core structures of the C_λ and $C_\gamma 1$ domains of KOL and NEW are reported in Table 5. In comparing C_λ (KOL) with C_λ (NEW) and $C_\gamma 1$ (KOL) with $C_\gamma 1$ (NEW) we are matching domains almost identical in sequence and expect to find very similar structures. If, in these cases, the ABDE β-sheets are superimposed, the CFG β-sheets in the pair C_λ (KOL)–C_λ (NEW) differ in position by 0·15 Å and in orientation by 2·8°, and the CFG β-sheets in the $C_\gamma 1$ domains differ in position by 0·37 Å and in orientation by 5·0°. We took these figures as an estimate of the variability in the core structure that all effects other than mutations could produce.

Further, the comparison of the C_λ (KOL) domain with the $C_\gamma 1$ (KOL) domain (which have different sequences) shows that if the ABDE sheets are superimposed, the CFG domains differ in position by 0·89 Å and in orientation by 7·2°. The

TABLE 5

Differences in the relative positions and orientations of the β-sheets in the domains C_λ (KOL), C_λ (NEW), $C_\gamma 1$ (KOL) and $C_\gamma 1$ (NEW)

Pair of domains	Relative geometry of CFG sheet after superposition of ABED sheet				
	Rotation angle (deg.)	Translation components			Total displacement (Å)
		x	*y*	*z*	
C_λ (KOL)–C_λ (NEW)	2·8				0·15
$C_\gamma 1$ (KOL)–$C_\gamma 1$ (NEW)	5·0				0·37
C_λ (NEW)–$C_\gamma 1$ (NEW)	2·2	(−0·58,	0·77	−0·09)	0·97
C_λ (KOL)–$C_\gamma 1$ (KOL)	5·7	(−0·53,	0·60	−0·22)	0·83

Each pair of structures is compared by first superimposing the ABED sheets and then calculating the additional translation and rotation required to superpose the CFG sheets. The ABED sheets are oriented (as in Figs 1 and 2) so that the mean plane of this sheet is the *x*–*y* plane and the strands are approximately parallel to the *y*-axis.

corresponding results for the comparison of C_λ (NEW) and $C_\gamma 1$ (NEW) are 0·97 Å and 2·2°. The observation of similar values and, in particular, the result that the shifts are in the same direction (see Table 5) gives us confidence that we have not underestimated the threshold of significance for interpreting these results.

Indeed, comparisons of other pairs of domains show changes in the relative geometry of the two β-sheets that are substantially larger (see Table 4), with residues at the end of β-sheets differing by 2 to 3 Å. For example, in Figure 5 we show the core structures of V_κ (REI) and $C_\gamma 1$ (KOL), with the ABDE sheets superimposed. The CFG sheets differ in position by 0·41 Å and in orientation by 16·1°. Primarily as a result of the rotation, homologous residues at the ends of the CFG strands differ in position by 2 Å. In this and in other cases, displacements of homologous atoms in the region of the "pin" are not larger than ~1 Å. These displacements are accommodated locally, by small changes in torsional angles.

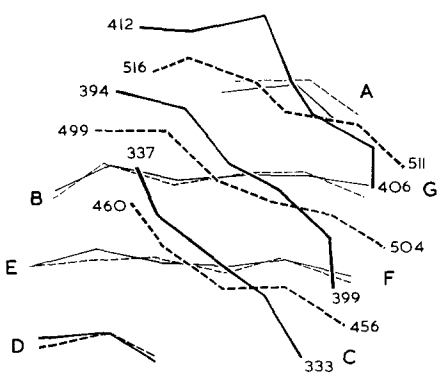


FIG. 5. The differences in the relative geometry of the sheets in V_κ (REI) and $C_\gamma 1$ (KOL). The ABDE sheets of these 2 domains are superimposed (bold unbroken and broken lines). The CFG sheets differ in position by 0·41 Å and in orientation by 16·1° (light unbroken and broken lines). Homologous residues at the ends of the CFG strands differ in position by ~2 Å.

(b) *Lateral insertion of loop residues into the region between the β-sheets*

Mutations that reduce the volumes of side-chains in the region between the β-sheets threaten to create spaces that must be filled if good packing is to be maintained. One mechanism for accommodating such mutations is the insertion of side-chains from the loops between strands, into the region between the β-sheets. A comparison of V_κ (REI) and V_λ (KOL) illustrates the lateral insertion of the loop between strands B and C in V_λ (KOL) into the region where side-chain volumes have been reduced (see Fig. 6).

Residues at position C2 in one sheet and E3 in the other point into the same region of the interface. In V_κ (REI) the large side-chains Leu33 (C2) and Tyr71 (E3) pack against each other. In V_λ (KOL), smaller residues occupy these positions: Val32 (C2) and Ser69 (E3). The C^α–C^α distance between residues C2 and E3 is 8·1 Å in V_κ (REI) and 9·5 Å in V_λ (KOL). In V_λ (KOL) these residues do not pack against each other: the space between them is filled by the Ile side-chain of residue 27B, in the BC loop (see Fig. 6).

This insertion also helps to accommodate another mutation, at position F5, Gln90 in V_κ (REI) and Ser89 in V_λ (KOL). This involves a repacking of the residues that fill the space around position B6, which is small in both of these domains (Ala25 in V_κ (REI) and Gly24 in V_λ (KOL)). In V_κ (REI) the space over B6 is filled by the side-chain of A1 (Met4), from the adjacent strand of the sheet (see Fig. 6). In V_λ (KOL) the inserted residue Ile27 fills this space. This frees the residue at A1 (Leu4 in V_λ (KOL)) to adopt an extended conformation and thereby occupy some of the space created by the reduction in size of the residue at F5 (Fig. 6).

(c) *The buried glutamine residue and the β-bulge*

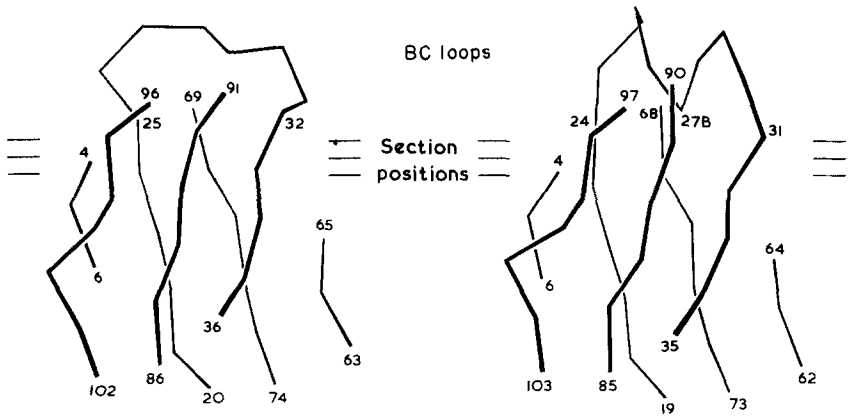
Two interdependent differences between V and C domains occur in the A and G strands: V domains have a buried Gln residue at position A3, associated with a β-bulge (Richardson *et al.*, 1978) in the other sheet in strand G. The G strands of C domains are regular and position A3 of C domains contains Leu, Thr or Pro.

The buried Gln at A3 in V domains forms hydrogen bonds to residues of the opposite β-sheet (Fig. 7). These involve main-chain atoms of residues F1 and F3, and the side-chain of G6 (Thr or Ser). In V_κ (REI) the hydrogen bonds are:

Gln6 (A3) Oε1 . . . N Cys88 (F3)
Nε2 . . . O Tyr86 (F1)
Nε2 . . . Oγ Thr102 (G6)

These interactions of Gln (A3) with residues of the opposite sheet interrupt the normal β-sheet association of the G and F strands. The insertion of an extra residue in the G strand permits formation of a β-bulge which detours around the amide group of Gln (A3).

In V domains residues at positions G4–G5–G5a–G6 have sequences of the form Gly–X–Gly–Thr/Ser. Position G6 retains a hydroxyl group for the hydrogen bond to Gln (A3). χ at G5 has an χ_R conformation and forms the β-bulge. Gly at G5a has $(\phi, \psi) = (+100, +150)$, outside the allowed range for other amino acids: this



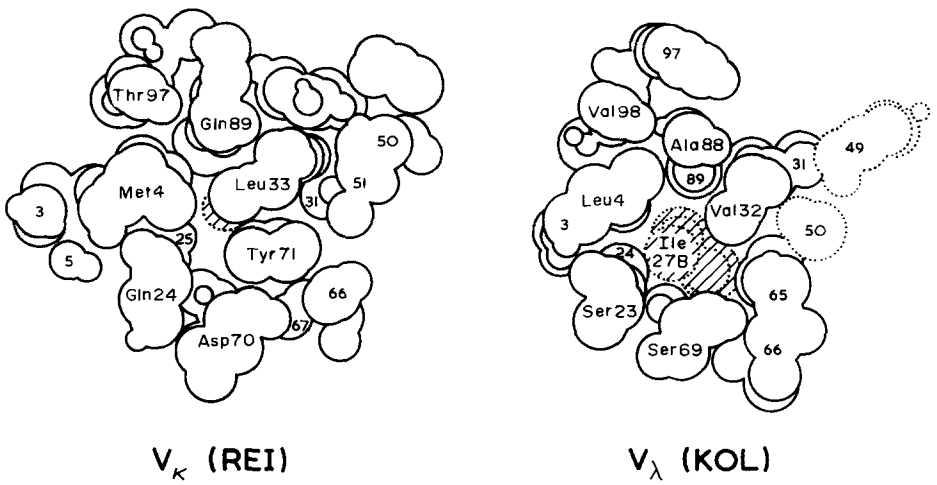


FIG. 6. A residue from a loop can be inserted laterally into the region between the sheets. In V_λ (KOL) residue Ile27B occupies a region of space between the core β-sheets. In V_κ (REI) residues from the β-sheets themselves pack together to fill this region. (a) Shows the position of C^α atoms of the β-sheet core end of the BC loop in V_κ (REI) and V_λ (KOL). (b) Shows 3 superimposed serial sections cut through computer-generated space-filling models of these proteins. The sections are separated by 1 Å and their position is indicated on the C^α drawings in (a). The main-chain of the BC loop is at the back of the stack of sections; the insertion of residue Ile27B of V_λ (KOL) is taking place *towards* the viewer. The sections are cut perpendicular to the β-sheet interface, with the strands running into the page. The residues in the β-sheets are indicated by unbroken lines. Residues not in the β-sheets are hatched or dotted.

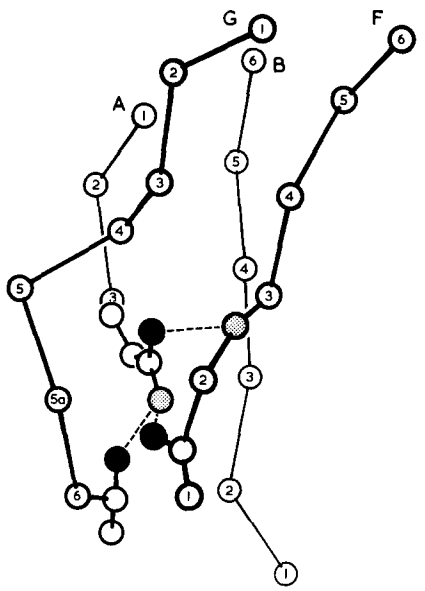


FIG. 7. The buried glutamine at position A3. This Figure shows, for V_κ (REI), the C^α atoms of strands A, B, F and G and other selected atoms for residues A3, F1, F3 and G6. The buried amide of the Gln at A3 hydrogen bonds to the main-chain of the F strand at positions 1 and 3 and to the side-chain of the Thr at G6. The β-bulge in the G strand (see the text) allows it to fold round the buried amide.

conformation permits the bulge to form without changing the direction of the strand. The Gly at position G4 is retained for a different reason: a side-chain at this position would occupy the same region of space as the amide group of Gln (A3).

(d) *Change in conformation*

One example of the use of conformational change to accommodate mutation was mentioned in section 5(b), above. Figure 8 illustrates another. This involves position E5, occupied by Phe73 in V_κ (REI) and by Val487 in $C_\gamma 1$ (KOL). These residues point into the interface between the sheets, and pack against adjacent inward-pointing residues along the E strand: Ile75 in V_κ and Val489 in $C_\gamma 1$. In both molecules the two E-strand residues pack against the Tyr at position F1 in the opposite β-sheet (Fig. 8). The replacement of Phe by Val at E5 would create a large cavity in $C_\gamma 1$ if no other change took place. Such a cavity does not form in $C_\gamma 1$ because these three residues are repacked by conformational changes (Fig. 8). The torsion angles of the main-chain and side-chains differ, between V_κ and $C_\gamma 1$, so as to retain the contacts of the E strand residues with each other and with the Tyr in the opposite sheet (Fig. 8).

(e) *Local complementarity in mutations*

Inspection of homologous proteins shows that local complementarity in mutations is rare; see the discussion of this point by Lesk & Chothia (1980). In the

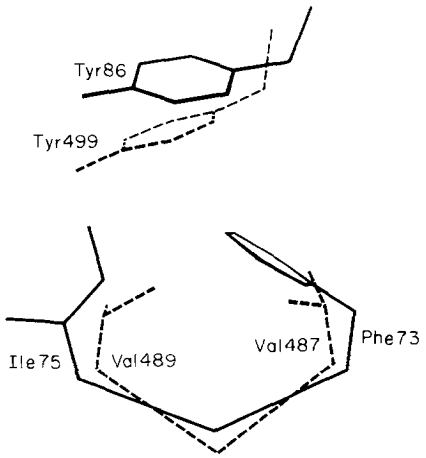


FIG. 8. Repacking by conformational change (see the text). The residues Phe73, Ile75 and Tyr86 in V_κ (REI) (unbroken line) are homologous to Val487, Val489 and Tyr499 in $C_\gamma1$ (KOL) (broken line). Differences in main-chain and side-chain torsion angles allow both sets of residues to be in contact.

immunoglobulins an example occurs in C_λ (KOL) and $C_\gamma2$ at two positions that are in contact across the β-sheet interface: C2 and B6. The residues at these positions are:

	C_λ	$C_\gamma2$
C2	Val148	Phe275
B6	Ile138	Val263

Although the Val → Ile change at B6 cannot entirely compensate (in volume) for the Phe → Val change at C2, small conformational changes bring the side-chains closer in C_λ: the C^β–C^β distance is 7·0 Å in C_λ, compared with 8·5 Å in $C_\gamma2$.

The preceding paragraphs have described and illustrated five mechanisms by which immunoglobulin domains accommodate mutations that occur in their interiors. The separation of these mechanisms for expository purposes is of course artificial: the actual response to a mutation will involve several occurring simultaneously and in a co-ordinated manner.

The principal driving force behind the changes we have described is the maintenance of a close-packed interior. It is unlikely, *a priori*, that all the immunoglobulin domains are close-packed equally well. At present there is no way to compare their total energies quantitatively. What does nevertheless emerge quite clearly is that changes in the size of interior residues produce quite extensive structural adjustments to maintain the overall packing.

6. Conclusions

The 17 homologous residues of immunoglobin domains that pack between the double β-sheet structure are primarily responsible for creating the basic "immunoglobulin fold". These cores provide the building blocks for entire

immunoglobulin molecules, and support the versatile creation of antigen binding sites from loops between the strands of β-sheets. Our results concern the effects of mutations in these residues on the geometry of the β-sheets and their interface. Other constraints on the evolution of immunoglobulin domains arising from interdomain interactions (and, in particular, the creation of the binding site) are beyond the scope of this study.

Although the residues between the β-sheets of any domain tend to remain hydrophobic in character, they show extensive substitution within this class, leading to large volume changes. The mean volume change at any position in the interface is 60 $Å^3$; the standard deviation is 45 $Å^3$. This mean value is comparable to that observed for the analogous set of residues in the globins (57 $Å^3$), and in plastocyanin and azurin (54 $Å^3$) (Lesk & Chothia, 1980; accompanying paper by Chothia & Lesk, 1982). These changes in size and shape require a rearrangement of the interface in order to maintain good packing.

The interface between the β-sheets is dominated by the "pin"; the disulphide bridge and contiguous residues that link the two β-sheets. The retention of this pin limits the range of excursion of the β-sheets, relative to one another. The atoms of the pin itself form very similar structures in different domains (r.m.s. deviations are all approximately 0·4 to 0·9 Å) and the relative position and orientation of the sheets varies, from domain to domain, by shifts of up to 2 Å and rotation around the pin of up to 20° in the structures studied.

In the globins, and in plastocyanin and azurin, the repacking of interfaces between secondary structures is often associated with large shifts and rotations of individual secondary structural elements with respect to each other (Lesk & Chothia, 1980; Chothia & Lesk, 1982). In the globins, helices in contact may be displaced by up to 7 Å, and the β-sheets in plastocyanin and azurin have shifted relative to one another by 3·75 Å. Such large relative motions are not possible for the immunoglobulin domains, because of the restraints imposed by the pin.

The mechanisms of accommodating mutations in the immunoglobulin domains present some features not yet observed to the same extent in other families of proteins; perhaps as a result of the restraints imposed by the pin. In common with other protein families, there are displacements of elements of secondary structure (though small), occasional locally complementary mutations, and conformational changes in the interface residues. But in addition, certain immunoglobulin domains fill the space created by mutations from large to small residues by the deep lateral insertion of side-chains into the interface region between the sheets.

In the other protein families we have studied (the globins, plastocyanin and azurin) the most stringent constraint on evolution was the conservation of the geometry of the binding site. This requirement limited the *net* effect of all other changes in secondary and tertiary structure. In both of these families, the major binding sites are contained in a monomer unit, and associated directly with one or more units of secondary structure. In these cases the maintenance of the binding site constrained the relative geometry of the secondary structural units.

In the immunoglobulins, the antigen binding site is created from adjacent V domains by the end of the strands and loops. Whereas in other families of proteins

the binding site is conserved or modified only slightly, in immunoglobulins the binding site is the *most* variable portion of the molecule. The formation of the binding site from loops, rather than from elements of secondary structure, makes it simpler to create different binding sites by changes in amino acid sequences, because the backbone torsion angles are relatively free to vary.

Thus, the core structure provides a scaffolding on which new binding sites may be built. But its role in this process is not entirely passive. The BC loop is part of the binding site and we have seen, in section 5(b) and Figure 5, how mutations in the core structure affect its conformation.

We thank Drs Deisenhofer and Huber for sending us atomic co-ordinates, John Cresswell for the Figures, A. Feinstein and M. Perutz for comments on the manuscript, and the Royal Society, National Science Foundation (PCM80–12007), the National Institute of General Medical Sciences (GM 25435), the North Atlantic Treaty Organization Division of Scientific Affairs (RG 275.80) and the European Molecular Biology Organization (SF4–1980) for support.

REFERENCES

Barker, W. C., Ketcham, L. K. & Dayhoff, M. O. (1978). In *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed.), vol. 5, suppl. 3, pp. 197–227, National Biomedical Research Foundation, Washington.

Beale, D. & Feinstein, A. (1976). *Quart. Rev. Biophys.* **9**, 135–180.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Chothia, C. (1976). *J. Mol. Biol.* **105**, 1–12.

Chothia, C. & Lesk, A. M. (1982). *J. Mol. Biol.* **160**, 309–323.

Davies, D. R., Padlan, E. A. & Segal, D. M. (1975). *Annu. Rev. Biochem.* **44**, 639–667.

Deisenhofer, J. (1981). *Biochemistry*, **20**, 2361–2370.

Edelman, G. M. & Gall, W. E. (1969). *Annu. Rev. Biochem.* **38**, 415–466.

Edmundson, A. B., Ely, K. R., Abola, E. E., Schiffer, M. & Panagiotopoulos, N. (1975). *Biochemistry*, **14**, 3953–3961.

Epp, O., Colman, P., Fehlhammer, H., Bode, W., Schiffer, M. & Huber, R. (1974). *Eur. J. Biochem.* **45**, 513–524.

Huber, R., Deisenhofer, J., Colman, P. M., Matsushima, I. & Palm, W. (1976). In *27 Mosbacher Colloquium*, Springer-Verlag, Berlin.

Kabat, E., Wu, T. T. & Bilofsky, H. (1979). *Sequences of Immunoglobulin Chains*, National Institutes of Health, Bethesda.

Kabsch, W. (1976). *Acta Crystallogr. sect. A*, **32**, 922–923.

Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **82**, 393–420.

Lesk, A. M. & Chothia, C. (1980). *J. Mol. Biol.* **136**, 225–270.

Marquart, M., Deisenhofer, J. & Huber, R. (1980). *J. Mol. Biol.* **141**, 369–391.

Novotný, J. & Franěk, F. (1975). *Nature (London)*, **258**, 641–643.

Poljak, R. J. (1978). *CRC Crit. Rev. Biochem.* **5**, 45–84.

Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P. & Saul, F. (1974). *Proc. Nat. Acad. Sci., U.S.A.* **71**, 3440–3444.

Richardson, J. S., Getzoff, E. D. & Richardson, D. C. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 2574–2578.

Saul, F. A., Amzel, L. M. & Poljak, R. J. (1978). *J. Biol. Chem.* **253**, 585–597.

Edited by R. Huber