Reprinted from *J. Mol. Biol.* (1985) **186**, 651–663

Domain Association in Immunoglobulin Molecules

The Packing of Variable Domains

Cyrus Chothia, Jiří Novotný, Robert Bruccoleri
and Martin Karplus

J. Mol. Biol. (1985) 186, 651–663

Domain Association in Immunoglobulin Molecules

The Packing of Variable Domains

Cyrus Chothia

MRC Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH

and

Christopher Ingold Laboratories
Department of Chemistry, University College London
20 Gordon Street, London WC1H 0AJ, England

Jiří Novotný, Robert Bruccoleri

Molecular & Cellular Laboratory
Massachusetts General Hospital

and

Harvard Medical School, Boston MA 02114, U.S.A.

and

Martin Karplus

Department of Chemistry
Harvard University, Cambridge MA 02138, U.S.A.

(Received 17 July 1984, and in revised form 19 July 1985)

We have analyzed the structure of the interface between VL and VH domains in three immunoglobulin fragments: Fab KOL, Fab NEW and Fab MCPC 603. About 1800 Å^2 of protein surface is buried between the domains. Approximately three quarters of this interface is formed by the packing of the VL and VH β-sheets in the conserved "framework" and one quarter from contacts between the hypervariable regions. The β-sheets that form the interface have edge strands that are strongly twisted (coiled) by β-bulges. As a result, the edge strands fold back over their own β-sheet at two diagonally opposite corners. When the VL and VH domains pack together, residues from these edge strands form the central part of the interface and give what we call a three-layer packing; i.e. there is a third layer composed of side-chains inserted between the two backbone side-chain layers that are usually in contact. This three-layer packing is different from previously described β-sheet packings. The 12 residues that form the central part of the three observed VL–VH packings are absolutely or very strongly conserved in all immunoglobulin sequences. This strongly suggests that the structure described here is a general model for the association of VL and VH domains and that the three-layer packing plays a central role in forming the antibody combining site.

1. Introduction

Immunoglobulins are the best-studied examples of a large and ancient family of proteins, which also includes β-microglobulins, *Thy-1* antigens, major (i.e. class I) and minor (i.e. class II) histocompatibility antigens and cell surface receptors. Functionally, all these structures are involved in cell recognition processes (Jensenius & Williams, 1982), either actively as vehicles endowed with

recognition specificity (antigen-combing antibodies) or passively as surface structures that are being recognized (histocompatibility antigens). Only the immunoglobulin tertiary structures are known to date (Schiffer *et al.*, 1983; Epp *et al.*, 1974; Saul *et al.*, 1978; Segal *et al.*, 1974; Marquart *et al.*, 1980; Deisenhofer, 1981; Phizackerley *et al.*, 1979). However, the homology among primary structures of immunoglobulin, β-microglobulin, *Thy-1* antigen, some of the histocompatibility antigen domains, T-cell receptor β chain and the transepithelial "secretory component" has been interpreted as evidence for a common fold (Cunningham *et al.*, 1973; Orr *et al.*, 1979; Feinstein, 1979; Cohen *et al.*, 1980, 1981*a*; Novotný & Auffray, 1984; Yangai *et al.*, 1984; Hedrick *et al.*, 1984; Mostov *et al.*, 1984).

A typical antibody molecule (IgG1) consists of two pairs of light chains (M_r 25,000) and two pairs of heavy chains (M_r 50,000), each of the chains being composed of domains made up of approximately 100 amino acid residues. The domains are autonomous folding units; it has been demonstrated experimentally (Hochman *et al.*, 1973; Goto & Hamaguchi, 1982) that a polypeptide chain segment corresponding to a single domain can be refolded independently of the rest of the polypeptide chain. All the immunoglobulin domains are formed by two β-sheets packed face-to-face and covalently connected together by a disulfide bridge. The topology of the N-terminal, variable domains in both the light and heavy chains differs from that of the C-proximal constant domains. While the two variable domain sheets consist of five and four strands, respectively, the constant domain sheets are three- and four-stranded (Fig. 1). The four-stranded β-sheets of the two domain types are homologous; the five- or four- stranded β-sheet of the variable domains derives from the three-strand sheet of the constant domains by the addition, at one side, of a two-stranded β-hairpin or a single β-strand, respectively.

In a complete immunoglobulin molecule, domains that correspond to different polypeptide chains associate to form domain dimers VL–VH, CL–CH1 and CH3–CH3. Edmundson *et al.* (1975) were the first to note the phenomenon of rotational allomerism between the variable and constant domain dimers, that is, whereas the C–C dimers interact *via* a close packing of their four-strand sheets, the V–V dimers pack "inside out", with the five-stranded sheets oriented face-to-face. The reversal of domain–domain interaction is reflected in the amino acid sequence homology between, and among, the constant and variable domains (Novotný & Franěk, 1975; Beale & Feinstein, 1976; Novotný *et al.*, 1977).

Different antibody molecules in the same organism bind different antigenic structures. The variation in specificity is produced by several mechanisms: mutations, deletions and insertions in the binding regions of the VL and VH domains; and the association of different light and heavy chains. Aspects of the second mechanism are analyzed in this paper. In particular, the nature of the interface between VL and VH domains is examined by comparing the Fab fragments of KOL, NEW and MCPC 603 myeloma proteins whose X-ray structures are known. The relative contributions to the buried surface between the domains from the conserved framework residues and the hypervariable regions are determined. Attention is focused on the unique packing of the interfaces and the reasons for this packing are examined.

2. Materials and Methods

(a) *Fab fragment co-ordinates*

Cartesian co-ordinates for Fab fragments KOL, NEW and MCPC 603 were obtained from the Brookhaven Data Bank (Bernstein *et al.*, 1977). Table 1 lists the domain classification, the nominal resolutions and the crystallographic residuals (*R* factors) for the 3 Fab fragments. To facilitate comparisons of the 3 structures, their residue numbering was changed from that used in the original descriptions to that used by Kabat *et al.* (1983). Thus, in this paper residues that are structurally homologous have the same sequence number.

To obtain consistent sets of atomic co-ordinates, the original co-ordinates were dissected into individual VL–

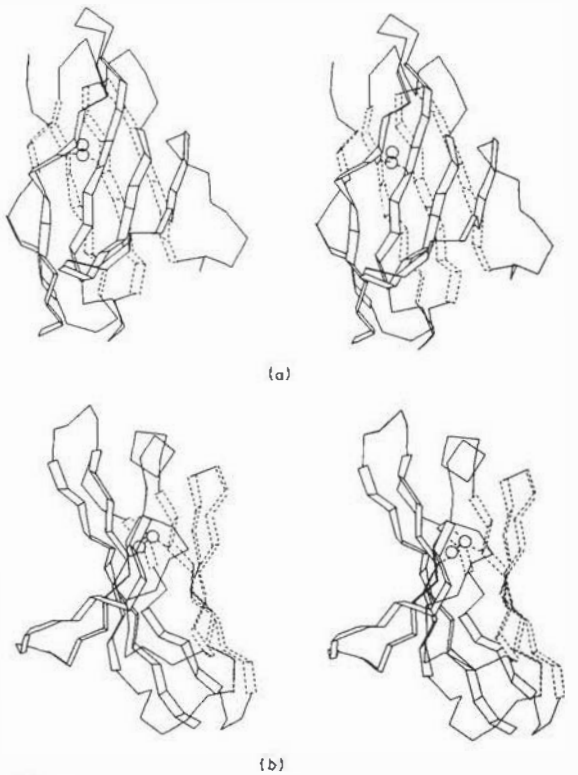

Figure 1. The β-sheets in typical immunoglobulin domains. Vertices represent the position of Cα atoms: those in β sheets are linked by ribbons; and those between strands by lines. (a) The VL domain of KOL: the β-sheet involved in VL–VH contacts is closer to the viewer (unbroken line). (b) The same VL domain rotated by approximately 90°. Note that the interface-forming β-sheet is strongly twisted at diagonally opposite corners (drawing by A. M. Lesk).

Table 1
Summary of X-ray crystallographic data

Protein	L and H chain types	X-ray data Resolution (Å)	X-ray data *R* factor (%)	Minimized Energy (kJ)	Minimized r.m.s. shift (Å)	Reference
Fab KOL human	λI, γIII	1·9	26	−3010	—	Marquart *et al.* (1980)
Fab NEW human	λI, γII	2·0	19	−2592	0·21	Saul *et al.* (1978)
Fab MCPC 603 mouse	κ, γI	2·7	24	−3703	0·26	Segal *et al.* (1974)

The energy given for Fab KOL is that of the unminimized crystallographic data.

VH domain dimers. The structures were subjected to 100 cycles of constrained energy minimization with the program CHARMM version 16 using the adopted-basis Newton-Raphson procedure (Brooks *et al.*, 1983) with constraints of 41·8 kJ (10 kcal) present on all the atoms (Bruccoleri & Karplus, unpublished results). Typically, the constrained minimization converged from original positive values of potential energy to values of about −2·1 kJ/atom (−0·50 kcal/atom) with an average root-mean-square co-ordinate different from the original X-ray structure of 0·3 Å (see Table 1). The results indicate that the crystallographic structures were satisfactory and that acceptable values of potential energy can be achieved by small adjustments of the co-ordinates. Thus, both energy minized structures and the crystallographic co-ordinates were used in the present study: essentially identical results were obtained from the 2 types of co-ordinates sets.

(b) *Computation of solvent-accessible surfaces and contact areas*

Solvent-accessible surfaces (Lee & Richards, 1971) were computed with programs written by A. M. Lesk using the method of Shrake & Rupley (1973) and by T. Richmond using the methods of Lee & Richards (1971) and Richmond & Richards (1978). The latter program was obtained from Yale University. The water probe radius used was 1·4 Å and the section interval along the *Z* axis was 0·05 Å; the atom van der Waals' radii used were 2 Å for all the (extended) tetrahedral carbon atoms, 1·85 Å for all the planar (*sp2* hybridized) carbons, 1·4 Å and 1·6 Å for carbonyl and hydroxyl oxygens, respectively, 1·5 Å for a carbonyl OH group, 2·0 Å for all the (extended) tetrahedral nitrogen atoms, 1·5 Å, 1·7 Å and 1·8 Å for *sp2*-hybridized nitrogen atoms carrying no hydrogen, 1 and 2 hydrogen atoms, respectively, 2·0 Å for a sulfhydryl group and 1·85 Å for a divalent sulfur atom with no hydrogens.

(c) *β-Strands and β-sheets*

Protein structures were analyzed using the CHARMM program (Brooks *et al.*, 1983) in the so-called explicit hydrogen atom representation: aliphatic hydrogens were combined together with their heavy atoms into "extended atoms" whereas hydrogens bound to polar atoms and possibly involved in hydrogen bonds were explicitly present. The β-strands and β-sheets were defined by their inter-strand backbone (C = O . . . H-N) hydrogen-bonding pattern. A hydrogen bond list was generated in CHARMM for all the polypeptide chain segments under consideration and amino acids with hydrogen bonds of nearly optimal geometry (energy of −4·18 kJ/bond or less) were taken to be parts of the β-sheets (cf. Fig. 3 of Novotný *et al.*, 1983). This method of defining β-strand boundaries gives results essentially identical to those obtained by visual inspection of crystallographic models, although it tends to be somewhat more restrictive (the 2 methods sometimes differ in inclusion of the N- or C-terminal β-strand residues). Ambiguities arise in cases of edge β-strands that start and end with irregular conformations (β-bulges); such cases are discussed in more detail below.

(d) *β-Strand conformation*

In a typical extended polypeptide chain segment the dihedral angle between the 2 consecutive side-chains is not 180° as in the ideal β-sheet (Pauling *et al.*, 1951) but closer to −160°; that is, the β-strands are twisted (Chothia, 1973). The out-of-planarity angle (180° − 160°) = 20° can be obtained explicitly from the values of the principal backbone torsion angles φ, ψ and ω (see, e.g. Chou *et al.*, 1982). We define the local backbone twist for 2 consecutive residues as:

$$\vartheta = \left(-\frac{\tau}{|\tau|}\right)(180 - |\tau|).$$

where τ is the torsion angle Cβ–Cα–C′α–C′β and |τ| denotes its magnitude. When glycine residues that lack Cβ atoms are encountered, the torsion angle ϑ is measured with respect to the C′β atom following the glycine. Thus, glycine residues contribute to the local backbone twist indirectly, by being included in the virtual bond Cα–Cα that spans from the residue preceding the glycine to that which follows it.

Backbone twist profiles (plots of ϑ as a function of the amino acid residue) serve to characterize polypeptide chain conformations. Certain conformational characteristics of polypeptides are more clearly seen using ϑ values instead of the $\varphi\psi$ values for individual residues. In our plots, the value of the torsion angle Cα–Cβ–C′α–C′β is assigned to the second (C′) residue. The angle ϑ is related to "the amount of twist per 2 residues", defined as δ by Chou *et al.* (1982); in fact, $\vartheta = \frac{1}{2}\delta$. It thus follows that ϑ can be obtained from the helical parameters u (number of residues per turn), h (the rise per residue) and T ($T = 360°/u$) in a corresponding way to that described for δ by Chou *et al.* (1982).

3. Results

(a) *Domain–domain contact surfaces*

We identified the residues that form the interface between VL and VH by calculation of the solvent-accessible surface of the domains, first in isolation and second when associated. Any residue that lost surface on the association of VL and VH was taken as part of the interface between them. We also determined which residues form van der Waals' contacts across the interface (distance cutoff 4·1 Å). The lists of residues obtained by the two methods were very similar. Thus, except for a few marginal cases, the residues that lose surface in domain–domain contacts also have van der Waals' interactions between the domains, indicating that the VL–VH interface is tightly packed.

The total surface areas of the separated VL and VH domains and that buried on the association is shown in Table 2. The values for the buried surface area (between 1700 and 1900 Å^2) and the fraction of the buried surface that is composed of polar atoms are similar to those found in other cases (Chothia & Janin, 1975). For the bovine pancreatic trypsin inhibitor and trypsin it is known that the structure of the isolated proteins does not change significantly on association. In most cases, as for the VL and VH domains considered here, there are no data concerning the structure of the unassociated domains.

Of the total area buried between the VL–VH dimers about one quarter comes from residues in the hypervariable regions and about three quarters from residues in β-sheets. Figure 2 shows the residues that form the interfaces and the areas that are buried for the three VH–VL packings. Two important features are evident in this Figure. First, homologous residues form the interface in the three structures. Second, the pattern formed by the contact residues is most unusual. The contacts of residues on the edge strands of the β-sheets are more extensive than those of residues on the inner strands. This is the opposite of the behavior found in previously described β-sheet packings, where it is the central strands that have the largest contact. For example, for packing of β-sheets in the same domain, the region of maximal contact generally runs diagonally across the sheets at 45° with respect to the β-strands (Cohen *et al.*, 1981*b*; Chothia & Janin, 1981). The point is clearly illustrated in the Cα backbone plot in Figure 2(c); here, for each of the Cα atoms a circle is displayed, the area of which is proportional to the total contact area made by the residue with the other sheet. As we describe below, the unusual packing is a direct consequence of the distortions present in this type of β-sheet.

(b) *Conformation of interface β-sheets*

The deviation of the conformations of the β-sheets that form the interface between VL and VH from the idealized flat structure (i.e. twisting, coiling and bending) can be characterized by the variations in the twist angle ϑ (see Materials and Methods). On such twist profiles, regular twisted β-sheets correspond to horizontal lines with an average $\vartheta = +20°$, right-handed α helices to lines of $\vartheta = -110°$ and tight reverse turns as triplets of points of approximately the same magnitude and alternating sign. The insertion of an additional residue in an edge strand of a β-sheet, so that two edge residues face one another on an inner strand, forms what has been called a β-bulge (Richardson *et al.*, 1978). Such insertions can have a variety of conformational effects depending upon the exact $\varphi\psi$ values of the inserted residue and those of its neighbors. Usually a sharp bend or local coiling is produced in the edge strand; this gives rise to a single- or double-point peak or trough in the ϑ values.

In Figure 3 we show the ϑ values for the VL–VH interface segments (β-strands with the adjacent hypervariable loops) in KOL, NEW and MCPC 603. Two important features of these β-sheets are evident from the Figure. First, most of the individual values of ϑ, and the patterns formed by the variations in ϑ angles, are very similar in the different sheets, particularly in the inner β-strands (β1, β3, β5 and β8 of Fig. 3) and in the β-bulges; the edge β-strands (β2, β4, β6 and β9 of Fig. 3) have

Table 2

Accessible surfaces and those lost on VL–VH association (Å^2)

	Isolated surface			Contact surface		
Domain pair	Hydrophobic	Polar	Total	Hydrophobic	Polar	Total
KOL VL domain	1121	658	1779	580	311	891
KOL VH domain	1216	700	1926	615	250	865
VL–VH in KOL	2337	1358	3705	1195	561	1756
NEW VL domain	1233	744	1977	529	387	916
NEW VH domain	1186	801	1985	506	386	892
VL–VH in NEW	2419	1545	3962	1035	773	1808
MCPC 603 VL domain	1082	689	1771	676	299	975
MCPC 603 VH domain	1156	760	1916	619	324	943
VL–VH in MCPC 603	2238	1449	3687	1295	623	1918
VL–VH average	2331	1714	3785	1195	652	1827

DOCKET ALARM

Explore Litigation Insights

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

Real-Time Litigation Alerts

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

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

Advanced Docket Research

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

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

Analytics At Your Fingertips

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

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

API

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

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

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

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

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