Article No. jmbi.1998.2439 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (1999) 285, 2177–2198

JMB

The Atomic Structure of Protein-Protein Recognition Sites

Loredana Lo Conte[1], Cyrus Chothia[1] and Joël Janin[1,2]*

[1]*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB1 1JX, England*

[2]*Laboratoire d'Enzymologie et de Biochimie Structurales CNRS UPR9063 91198 Gif-sur-Yvette, France*

**Corresponding author*

The non-covalent assembly of proteins that fold separately is central to many biological processes, and differs from the permanent macromolecular assembly of protein subunits in oligomeric proteins. We performed an analysis of the atomic structure of the recognition sites seen in 75 protein-protein complexes of known three-dimensional structure: 24 protease-inhibitor, 19 antibody-antigen and 32 other complexes, including nine enzyme-inhibitor and 11 that are involved in signal transduction.

The size of the recognition site is related to the conformational changes that occur upon association. Of the 75 complexes, 52 have ''standard-size'' interfaces in which the total area buried by the components in the recognition site is 1600 (±400) Å^2. In these complexes, association involves only small changes of conformation. Twenty complexes have ''large'' interfaces burying 2000 to 4660 Å^2, and large conformational changes are seen to occur in those cases where we can compare the structure of complexed and free components. The average interface has approximately the same non-polar character as the protein surface as a whole, and carries somewhat fewer charged groups. However, some interfaces are significantly more polar and others more non-polar than the average.

Of the atoms that lose accessibility upon association, half make contacts across the interface and one-third become fully inaccessible to the solvent. In the latter case, the Voronoi volume was calculated and compared with that of atoms buried inside proteins. The ratio of the two volumes was 1.01 (±0.03) in all but 11 complexes, which shows that atoms buried at protein-protein interfaces are close-packed like the protein interior. This conclusion could be extended to the majority of interface atoms by including solvent positions determined in high-resolution X-ray structures in the calculation of Voronoi volumes. Thus, water molecules contribute to the close-packing of atoms that insure complementarity between the two protein surfaces, as well as providing polar interactions between the two proteins.

Keywords: protein-protein complexes; interface area; polar interactions; packing density; conformation changes

Introduction

Many biological processes are carried out, or regulated, through the interactions between preformed proteins. The importance of such interactions in biology has made the protein recognition process an area of considerable interest. Here, we describe aspects of the atomic structure of the recognition sites seen in protein-protein complexes of known structure. The protein-protein complexes whose structures we analyse here are non-covalent assemblies of proteins that fold separately to carry out independent functions before they associate, as opposed to permanent macromolecular assemblies, such as oligomeric proteins, virus capsids or muscle fibres. Though the two types of complexes have features in common, they also have many that differ, and are best treated separately.

General analyses of structural aspects of protein-protein interaction have been carried out (Chothia & Janin, 1975; Argos, 1988; Janin & Chothia, 1990; Janin, 1995, 1996; Jones & Thornton, 1995, 1996,

Abbreviations used: ASA, solvent-accessible surface area; PDB, Protein Data Bank; PTI, pancreatic trypsin inhibitor.

E-mail address of the corresponding author: janin@lebs.cnrs-gif.fr

0022-2836/99/052177–22 $30.00/0

Table 1. Protein-protein complexes

Code	Complex	Res (Å)	Interface area B (Å^2)	H-bonds N_{hb}	Water molecules N_{wat}[a]	Reference
A. *Protease-inhibitor (19)*						
2ptc	Trypsin-PTI	1.9	1430	12	18	Huber *et al.* (1974)
1avw	Trypsin-soybean inhibitor	1.8	1740	11	12	Song & Suh (1998)
1mct	Trypsin-bitter gourd inhibitor	1.6	1520	10	13	Huang *et al.* (1993)
3tpi	Trypsinogen-PTI	1.9	1420	11	17	Marquart *et al.* (1983)
1tgs	Trypsinogen-PSTI	1.8	1730	10	18	Bolognesi *et al.* (1982)
1cho	Chymotrypsin-ovomucoid	1.8	1470	10	17	Fujinaga *et al.* (1987)
1acb	Chymotrypsin-eglin C	2.0	1540	9	7	Frigerio *et al.* (1992)
1cbw	Chymotrypsin-PTI	2.6	1460	9		Scheidig *et al.* (1997)
1ppf	Elastase-ovomucoid	1.8	1330	6	18	Bode *et al.* (1986)
1fle	Elastase-elafin	1.9	1780	9	11	Tsunemi *et al.* (1996)
2kai	Kallikrein-PTI	2.5	1440	10		Chen & Bode (1983)
1hia	Kallikrein-hirustatin	2.4	1740	10	14	Mittl *et al.* (1997)
3sgb	*S. griseus* protease B-ovomucoid	1.8	1280	8	16	Read *et al.* (1983)
1mkw	Thrombin-prethrombin	2.3	1280	14	6	Malkowski *et al.* (1997)
1cse	Subtilisin-eglin C	1.2	1490	12	23	Bode *et al.* (1987)
2sic	Subtilisin-SSI	1.8	1620	10	20	Takeuchi *et al.* (1991)
2sni	Subtilisin-CI2	2.1	1630	10	10	McPhalen & James (1987)
1stf	Papain-stefin	2.4	1790	5	13	Stubbs *et al.* (1990)
4cpa	Carboxypeptidase A-inhibitor	2.5	1360	5		Rees & Lipscomb (1982)
B. *Large protease complexes (5)*						
1bth	Thrombin E192Q-PTI	2.3	2380	13	5	van de Locht *et al.* (1997)
4htc	Thrombin-hirudin	2.3	3350	16	38	Rydel *et al.* (1991)
1tbq	Thrombin-rhodniin	3.1	3510	12		van de Locht *et al.* (1995)
1toc	Thrombin-ornithodorin	3.1	3510	15		van de Locht *et al.* (1996)
1dan	Factor VIIA-soluble tissue factor	2.0	3770	23	28	Banner *et al.* (1996)
C. *Antibody-antigen (19)*						
1vfb	Fv D1.3-lysozyme	1.8	1400	9	28	Bhat *et al.* (1994)
1mlc	Fab D44.1- lysozyme	2.1	1410	5	7	Braden *et al.* (1994)
1jhl	Fv D11.15-lysozyme	2.4	1260	5		Chitarra *et al.* (1993)
3hfl	Fab Hy-HEL5-lysozyme	2.7	1730	11		Sheriff *et al.* (1987)
3hfm	Fab Hy-HEL10-lysozyme	3.0	1610	11		Padlan *et al.* (1989)
1fbi	Fab 9.13.7-lysozyme	3.0	1720	12		Lescar *et al.* (1995)
1mel	Camel H chain-lysozyme	2.5	1710	8		Desmyter *et al.* (1996)
1jel	Fab Jel42-HPR	2.8	1360	5		Prasad *et al.* (1993)
1nsn	Fab N10-Staph. nuclease	2.9	1800	4		Bossart-Whitaker *et al.* (1995)
1osp	Fab-*Borrelia* OSP-A	2.0	1500	7	27	Li *et al.* (1997)
1nca	Fab NC41-flu neuraminidase	2.5	1960	12		Tulip *et al.* (1992)
1nmb	Fab NC10-flu neuraminidase	2.5	1500	9		Malby *et al.* (1994)
(*1)[b]	Fab BH151-flu hemagglutinin X31	2.8	1550	5		Fleury *et al.* (personal communication)
(*2)[b]	Fab HC45-flu hemagglutinin X31	2.8	1850	4		Fleury *et al.* (personal communication)
1dvf	Fv D1.3-Fv E5.2	1.9	1680	10	15	Braden *et al.* (1996)
1iai	Fab 730.1.4-Fab 409.5.3	2.9	1900	6		Ban *et al.* (1994)
1nfd	Fab H57-N15 T cell receptor	2.8	1710	8		Wang *et al.* (1998)
1kb5	Fab Désiré-1-TCR Fv domain	2.5	2340	15		Housset *et al.* (1997)
1ao7	T cell receptor-HLA-A2/peptide	2.6	1990	15		Garboczi *et al.* (1996)
D. *Enzyme complexes (9)*						
(*3)[b]	β-Lactamase-BLIP	1.7	2560	11	31	Strynadka *et al.* (1996)
1brs	Barnase-barstar	2.0	1570	13	32	Buckle *et al.* (1994)
1dfj	RNase A-RNase inhibitor	2.5	2600	7		Kobe & Deisenhofer (1995)
1dhk	α-Amylase-bean inhibitor	1.9	3080	12	37	Bompard-Gilles *et al.* (1996)
1fss	Acetylcholinesterase-fasciculin	3.0	1970	7		Harel *et al.* (1995)
1gla	Glycerol kinase-Factor IIIGlc	2.6	1300	4		Hurley *et al.* (1993)
1udi	Uracil-DNA glycosylase-inhibitor	2.7	2020	11		Savva & Pearl (1995)
1ydr	Protein kinase A-inhibitor	2.2	2000	14	6	Engh *et al.* (1996)
2pcc	Cytochrome peroxydase-cytochrome c	2.3	1170	1	15	Pelletier & Kraut (1992)
E. *G-proteins, cell cycle, signal transduction (11)*						
1tx4	Rho-Rho GAP	1.6	2280	14	50	Rittinger *et al.* (1997)
1gua	Rap1A-cRaf1	2.0	1310	12	7	Nassar *et al.* (1996)
1a2k	Ran-NFT2	2.5	1590	10		Stewart *et al.* (1998)
1efu	EFtu-EFts. *E. coli*	2.5	3660	11		Kawashima *et al.* (1996)
1aip	EFtu-EFts. *T. thermophilus*	3.0	2940	9		Wang *et al.* (1997)
1gg2	$G_{i\alpha 1}$-$G_{i\beta 1\gamma 2}$	2.4	2360	13	10	Wall *et al.* (1995)
1got	Transducin $G_{t\alpha}$-$G_{t\beta\gamma}$	2.0	2500	18	36	Lambright *et al.* (1996)
2trc	$G_{t\beta\gamma}$-phosducin	2.4	4660	34	26	Gaudet *et al.* (1996)

1agr	$G_{i\alpha}$-RGS4	2.8	1650	13		Tesmer *et al.* (1997)
1fin	CDK2-cyclin A	2.3	3400	17	18	Jeffrey *et al.* (1995)
1a0b	CheA-Che Y	2.9	1140	5		Welch *et al.* (1997)
F. *Miscellaneous (12)*						
1fc2	Protein A-Fc fragment	2.8	1300	2		Deisenhofer (1981)
1igc	Protein G-Fab MOPC21	2.6	1350	14		Derrick & Wigley (1994)
1ak4	Cyclophilin-HIV capsid	2.4	1170	8	3	Gamble *et al.* (1996)
1efn	Fyn SH3 domain-HIV Nef	2.5	1260	7		Lee *et al.* (1996)
1atn	Actin-DNase I	2.8	1780	8		Kabsch *et al.* (1990)
2btf	Actin-profilin	2.5	2090	11		Schutt *et al.* (1993)
1dkg	Grep E-DNA K	2.8	1980	4		Harrison *et al.* (1997)
1ebp	Erythropoietin receptor-peptide	2.8	1940	13		Livnah *et al.* (1996)
1hwg	HGH receptor-human growth hormone	2.5	4200	11		Sundström *et al.* (1996)
1seb	HLA DR1-enterotoxin B	2.7	1340	8		Jardetsky *et al.* (1994)
1tco	FKBP12-Calcineurin	2.5	2470	8		Griffith *et al.* (1995)
1ycs	p53-53BP2	2.2	1500	7	8	Gorina & Pavletich (1996)
Protease-inhibitor		Mean	1530	9.5	14.6	
		s.d.	170	2.3	4.7	
Large protease complexes		Mean	3300	15.8		
		s.d.	540	4.3		
Antibody-antigen		Mean	1680	8.5		
		s.d.	260	3.5		
Enzyme complexes		Mean	2030	8.9		
		s.d.	630	4.4		
G-proteins, signal transduction		Mean	2500	14.2		
		s.d.	1090	7.5		
Miscellaneous		Mean	1870	8.4		
		s.d.	840	3.4		
All complexes		Mean	1940	10.1	18.3	
		s.d.	760	4.8	11.1	

[a] Number of interface solvent molecules in 36 X-ray structures with resolution 2.4 Å or better. Due to the small number of occurences in other categories, the mean and standard deviation are quoted only for protease-inhibitor and for all complexes.
[b] Co-ordinates are gift from authors. The files are refered to as *1 to *3.

1997; Tsai *et al.*, 1996; Chothia, 1997). Recently, however, there has been a large increase in the number of known three-dimensional structures that contain protein-protein recognition sites. These structures cover a much broader range of activities than earlier ones, which were almost exclusively protease-inhibitor and antibody-antigen complexes. They allow us to determine the extent to which rules based on the few structures that were first available, can be generalised. We also find features that were missed in previous work.

The aspects of structure with which we are concerned here are those related to the stabilization of protein association: the size and chemical character of the protein surface that is buried at interfaces; the packing density of atoms that make contacts across the interface, which expresses complementarity; and polar interactions through hydrogen bonds and interface water molecules. Each of these aspects can be described at the level of the individual atom.

Protein-Protein Complexes of Known Structure

From the Protein Databank (PDB; Bernstein *et al.*, 1977), we selected 72 non-redundant files that contain X-ray structures of protein-protein complexes determined at a resolution of 3.1 Å or better. When more than one complex was present in the asymmetric unit, only one copy was retained. The atomic co-ordinates of another three complexes were given to us by their authors.

Table 1 lists the 75 structures: 24 are complexes of proteases with protein inhibitors, 19 are complexes of antibody Fab or Fv fragments with cognate protein antigens, and 32 complexes of other kinds. Among the latter, there are nine complexes between enzymes and protein inhibitors or substrates, and 11 that involve GTP/GDP-binding proteins (G-proteins) or other components of the cellular cycle or signal transduction pathways. Although nearly all these complexes are binary in the sense that they have only two protein components, a few also contain cofactors or other non-protein components.

The Interface Areas of Protein-Protein Recognition Sites

Our basic tool for measuring the extent of a protein-protein recognition site is the interface area: the area of the accessible surface on both partners that becomes inaccessible to solvent due to protein-protein contacts. This area is the sum of the solvent-accessible surface areas (ASA) of the isolated components less that of the complex.

Solvent accessible surface areas were determined here using the Lee & Richards (1971) algorithm. The computer program is based on that originally produced by Lee & Richards (1971), with subsequent modifications and extensions by Harpaz *et al.* (1994) and by Gerstein *et al.* (1995). Cofactors,

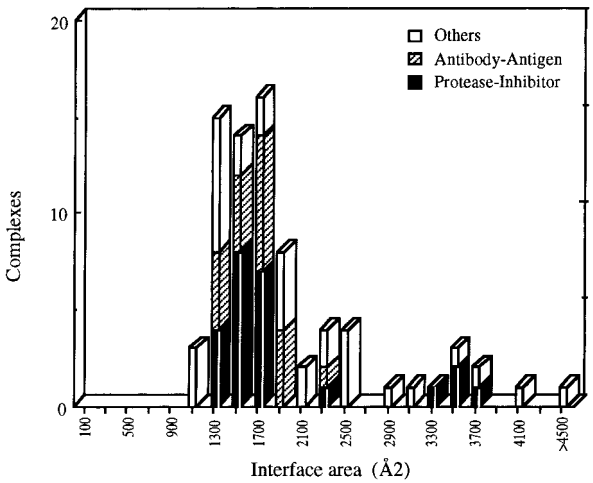

Figure 1. Histogram of interface areas. The distribution of the values of B which are given in Table 2.

but not solvent molecules, were included in the calculations. Group radii were taken from Chothia (1975). The radius of the water probe was 1.4 Å.

We give in Table 1 the interface area B, and the number of hydrogen bonds and water molecules observed at each interface (see below for their definition). The B value measures the interaction between components. The smallest interfaces have $B \approx 1150$ Å^2 in the yeast cytochrome peroxydase-cytochrome c complex (2pcc), the Che A-Che Y complex (1a0b) and between cyclophilin and the HIV capsid protein (1ak4). The largest interface has $B = 4660$ Å^2, and is between phosducin and the $G_{\beta\gamma}$ subunit of transducin (2trc). Thus, the overall range of interface areas B is wide. However, the histogram of Figure 1 shows that most values are clustered at the lower end of this range. All but one of the antibody-antigen complexes, and 19 out of 24 protease-inhibitor complexes have interface areas between 1250 and 2000 Å^2. In antibody-antigen complexes, the mean value of B is 1680 Å^2, and the standard deviation is 260 Å^2. Only the interface between Fab Désiré-1 and the T-cell receptor (1kb5) has B above 2000 Å^2. The distribution of interface areas in protease-inhibitor complexes is bimodal. The major peak (19 complexes) has a mean B of 1530 Å^2, and a standard deviation of 170 Å^2. These protease-inhibitor complexes are comparable with the antibody-antigen complexes. The four complexes in the minor peak centred near $B = 3300$ Å^2 involve the blood proteases thrombin and factor VIIA. They have interfaces twice as large and are listed as ''large'' for that reason.

The histogram of the interface areas for other types of complexes shows a wider distribution. Yet nearly half (15 out of 32) are in the range 1600 (±400) Å^2, which we shall call ''standard-size'' interfaces. Below this range, we find the three ''small'' interfaces with $B \approx 1150$ Å^2; there are 14 large interfaces. In total, we have 3 small, 52 standard-size (70 %) and 20 large interfaces (27 %).

Whereas the value of B quoted in Table 1 applies to the two component proteins of a complex, interface areas were evaluated on each one (data not shown). The components contribute almost equally (50 (±3)%) to B, with the exception of the standard-size protease-inhibitor complexes. There, B is distributed 46:54 on average between the protease and the inhibitor. This asymmetry is the result of a convex inhibitor surface fitting into a concave active site. Because the ASA is measured one water probe away from the molecular surface, convexity tends to make it relatively larger, and concavity, smaller. The curvature effect is also seen on the peptide inhibitor bound at the active site of protein kinase A (1ydr) and in the cyclophilin-HIV capsid protein complex (1ak4). It is negligible in other complexes.

The descriptions small, standard and large refer to the size of the interface, not the complex, as there is little correlation between B and the size of the complex measured by its molecular weight or its ASA. The correlation coefficient between B and the ASA is only 0.38, and the ratio of B to the ASA varies between 4 % and 18 % in different complexes. The complex between thrombin and hirudin (4htc) has a lower molecular weight (40 kDa) than a Fab-lysozyme complex, yet its interface is more than twice as large. Conversely, Fab-lysozyme complexes have the same size interface as the much smaller (20 kDa) complex of the bacterial ribonuclease barnase with its inhibitor barstar (1brs; Guillet *et al.*, 1993; Buckle *et al.*, 1994).

Conformation Changes

For many of the complexes listed in Table 1, atomic structures are known not just for the complex but also for one or both of the free components. Comparison of the free and complexed structures can show the changes that take place upon association. With one exception, the proteases, inhibitors, antigens and antibodies that form complexes with standard size interfaces undergo only small changes in conformation. These include shifts in surface loops or movements of short segments of polypeptide chain by up to 1.5 Å, and the rotation of surface side-chains (Janin & Chothia, 1990; Davies & Cohen, 1996). The trypsinogen-pancreatic trypsin inhibitor (PTI) complex is the exception: part of the polypeptide chain is disordered in the isolated enzyme, but ordered in the complex (Bode *et al.*, 1978). The disorder-to-order transition is reflected in the association constant of PTI, which is several orders of magnitude smaller for trypsinogen than for trypsin which has a fully ordered structure, though the two complexes are otherwise extremely similar.

Conversely, the formation of complexes that have large interfaces is seen to involve large changes in conformation in those cases where native structures are available. These changes are generally described in publications reporting the

Table 2. Conformation change upon formation of large interfaces

Code	Complex	B (Å^2)	Type of change
A. *Large protease complexes*			
1bth	Thrombin E192Q-PTI	2380	Enzyme: large loops movements. PTI: no change.
4htc	Thrombin-hirudin	3350	Enzyme: no change. Hirudin: C-terminal tail becomes ordered.
1tbq	Thrombin-rhodniin	3510	Enzyme: no change. Rhodniin has two domains with a flexible linker.
1toc	Thrombin-ornithodorin	3510	Enzyme: no change. Ornithodorin has two domains with a flexible linker.
1dan	Factor VIIA-soluble tissue factor	3770	Factor VIIA is an extended molecule made of five domains that wrap around Tissue Factor. Tissue factor: no change.
B. *Antibody-antigen complexes*			
1kb5	Fab Désiré-1-TCR Fv domain	2340	No structure of free components.
C. *Others*			
(*3)	β-Lactamase-BLIP	2560	Enzyme: no change. BLIP: bending of saddle-shaped inhibitor.
1dfj	RNase A-RNase inhibitor	2600	Enzyme: no change. Horseshoe shaped inhibitor opens up to fit the enzyme in.
1dhk	α-Amylase-bean inhibitor	3080	Enzyme: large loop movements. No free inhibitor structure.
1udi	U-DNA glycosylase-inhibitor	2020	Enzyme: no change. No free inhibitor structure.
1tx4	Rho-Rho GAP	2280	Rho: loop becomes ordered. Rho GAP: a 10 residue loop becomes ordered.
1efu	EFtu-EFts, *E. coli*	3660	EFtu: domains and loops move, a α-helix refolds. No free EFts structure.
1aip	EFtu-EFts, *T. thermophilus*	2940	EFtu: domains move, a α-helix and loops shift. EFts: N-terminal residues become ordered.
1gg2	$G_{i\alpha 1}$-$G_{i\beta 1\gamma 2}$	2360	G_α: a α-helix rotates, loops move; N- and C-terminal residues become ordered or refold. No free $G_{\beta\gamma}$ structure.
1got	Transducin $G_{t\alpha}$-$G_{t\beta\gamma}$	2500	See the text.
2trc	$G_{t\beta\gamma}$-phosducin	4660	$G_{\beta\gamma}$: loops move. Phosducin is an extended molecule that wraps around $G_{\beta\gamma}$.
1fin	CDK2-cyclin A	3400	Kinase: domains move, a α-helix rotates, a loop moves. No free cyclin structure.
2btf	Actin-profilin	2090	Actin: domain rotation relative to actin-DNase complex. Profilin: no change.
1hwg	HGH receptor-HGH	4200	Receptor dimerizes. Large helix movements in HGH.
1tco	FKBP12-Calcineurin	2470	No change in FKBP12. No structure of free calcineurin.

X-ray structure of the complexes. They are summarised in Table 2 for the 20 complexes that have $B > 2000$ Å^2. In one (1kb5), no free component structure is known. In 14 others, at least one of the components is seen to undergo large changes in conformation upon complex formation. There are three more blood protease-inhibitor complexes (1tbq, 1toc, 1dan) where the protease undergoes little change. Whilst the structure of the free inhibitors is not known, they are elongated molecules made of well-defined domains separated by flexible linkers, and unlikely to have the same structure outside the complexes. The remaining two cases are the FKBP12-calcineurin complex (1tco) which has a large non-protein component at the interface and no known structure for the free calcineurin, and the uracil-DNA-glycosylase-inhibitor complex (1udi). There, the enzyme shows little change and the inhibitor is a small compact molecule, but their interface is just above the 2000 Å^2 cut-off.

Three major types of structural changes can be deduced from the information available at present: disorder-to-order transitions as regions of the polypeptide chain that are disordered in the free protein become ordered in the complex; large movements of the main-chain as loops switch to quite different conformations; and in multi-domain proteins, changes in the relative position of the domains. The first type is illustrated by the thrombin-hirudin complex (4htc). In addition to a N-terminal globular domain that binds at the active site, the hirudin inhibitor has a C-terminal tail that becomes ordered as it binds to a secondary site of the protease (Rydel *et al.*, 1991). The second type is illustrated by transducin as discussed below, and the third by the EFtu component of the *Escherichia coli* EFtu-EFts complex (1efu): two domains swing by 18°, undergoing movements of 10 Å on average (Kawashima *et al.*, 1996). These three types of structural changes are the most common, but others are observed and in many of the cases, combinations of changes are seen. For instance, the CDK2 kinase undergoes domain and loop movements upon binding cyclin A, but it also contains a α-helix that rotates about its axis while becoming part of the interface (Jeffrey *et al.*, 1995).

Classes of Atoms at Protein-Protein Recognition Sites

In describing the atomic structure of recognition sites it is useful to distinguish between three classes of atoms (Figure 2). These are (i) interface atoms (types A, B, C): all atoms that lose solvent-accessible surface on the formation of the complex. Interface atoms on different components are within the sum of the Van der Waals radii plus the diameter of the water probe (2.8 Å). Not all interface atoms make actual Van der Waals contact across the interface. Those which do will be called: (ii) contact

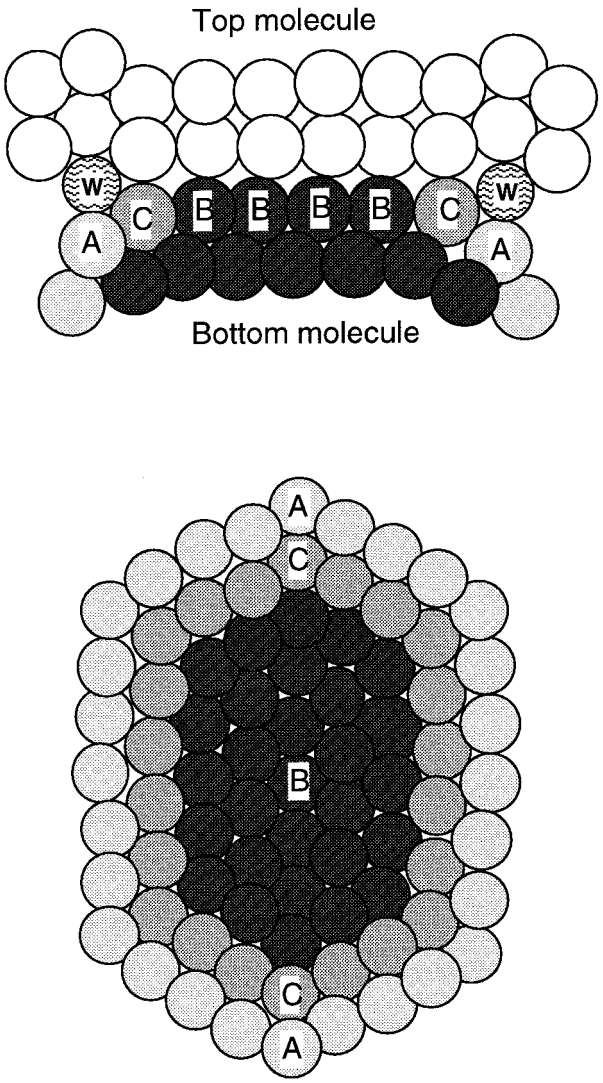

Figure 2. Classes of interface atoms. (a) Side view: all the atoms of the bottom molecule that lose accessibility to solvent in the presence of the top molecule are interface atoms. Type B is fully buried, types A and C retain partial accessibility. Types B and C are contact atoms; they make Van der Waals contact with the top molecule, while type A does not. W represents the water probe. (b) Top view: 29 atoms of type B are surrounded by a layer of 22 atoms of type C and another of 28 atoms of type A. All 79 atoms are interface atoms, their respective numbers approximate those of buried (B) and contact (B and C) atoms in one side of a standard-size interface.

atoms (types B and C): all atoms within a distance of atoms on the partner protein equal to the sum of the Van der Waals radii plus 0.5 Å. Contact atoms are a subset of interface atoms. Though contact atoms make Van der Waals contacts across the interface, they may retain some accessibility to the solvent. (iii) Buried interface atoms (type B): atoms that are accessible in the free component but have zero accessible surface area in the complex.

The average interface contains 211 atoms belonging to 52 amino acid residues. Of these atoms, approximately one-half are contact atoms and one-third are buried (Table 3). For a standard-size interface, each side has an average of 87 atoms from 22 residues, of which 44 are contact and 30 buried atoms. Figure 2(b) shows how these numbers can be rationalised in a simple way. Buried atoms forming an irregular hexagonal array are surrounded by a ring of contact atoms and another of non-contact atoms that lose accessible surface. As they are all interface atoms, their numbers approximately correspond with one side of a standard-size interface. The ratio between the three classes depends on the shape of the array, recalling that real interfaces involve patches of the protein surface that are not planar and have irregularly shaped edges. The shape in Figure 2(b) reproduces an average of these effects.

The number of interface atoms scales linearly with the interface area. On average, there is one interface atom per 9.2 Å^2 of interface area. As one might expect, contact atoms bury more surface than interface atoms not in direct contact. Whereas contact atoms are only one-half of all the interface atoms, they contribute three-quarters of the interface area. The remaining one-quarter belongs to peripheral interface atoms of type A in Figure 2, whose closest approach to atoms on the other component is less than the sum of the Van der Waals radii, plus the diameter of a water molecule. Buried atoms contribute less interface area than other contact atoms, because most are already partly buried in the free proteins.

The Chemical Character of the Interfaces

Chemical groups at the protein surface may be allotted to one of three types: non-polar (all groups containing aliphatic and aromatic carbons); neutral polar groups (all groups containing non-carbon atoms, except those carrying a net electric charge), and charged groups. Their relative contribution to the interface area is listed in Table 6 (see page 2186). The average composition of the solvent-accessible surface of small globular proteins is 57 % non-polar, 24 % neutral polar and 19 % charged (Miller *et al.*, 1987a). The average surface of the 75 protein-protein interfaces is 53 % non-polar, the average interface, 56 %. Thus, the fraction contributed by non-polar groups is similar and, on average, patches of protein surface that form interfaces are no more hydrophobic than the accessible surface of small globular proteins. The fraction contributed by neutral polar groups is somewhat higher, whereas charged groups are somewhat less abundant.

However, the spread around the average values is large (Figure 3). We do observe interfaces that are significantly less polar than the average protein surface, and others that are significantly more polar. Standard-size protease-inhibitor interfaces tend to be of the first type with a mean 61 % non-polar. Antibody-antigen interfaces are of the

Table 3. Categories of interface atoms

	All interfaces	Protease-inhibitor	Antigen-antibody	Other interfaces	Accessible surface
Average number	211	207	185	229	-
A. *Number fraction*					
Contact	0.51	0.56	0.48	0.49	
Buried[a]	0.35	0.40	0.31	0.32	-
Buried with water[b]	0.59	0.60	0.62	0.55	-
B. *Area fraction*[c]					
Contact	0.76	0.79	0.73	0.75	-
Buried[a]	0.26	0.32	0.22	0.24	-
Buried with water[b]	0.59	0.59	0.63	0.57	-
Main-chain	0.19	0.23	0.19	0.17	0.22
Side-chain	0.80	0.76	0.81	0.82	0.77
Non-protein groups	0.004	0	0	0.01	0.006

[a] Interface atoms with zero accessible-surface area.
[b] Interface atoms with zero accessible-surface area when solvent is taken into account; the data set is restricted to 36 X-ray structures with resolution 2.4 Å or better.
[c] The average interface area per interface atom is close to 9.2 Å^2 in all complexes.

second type with a mean 51 % non-polar. The interface between the two antibody Fv fragments D1.3 and E5.2 (1dvf) is the most hydrophilic in our sample (42 % non-polar). The most hydrophobic (≈70 % non-polar) are in two protease-inhibitor complexes (1ppf,1acb) and in the erythropoietin receptor-peptide complex (1ebp). The latter complex contains a peptide that has been selected *in vitro* for high affinity, but is unrelated to the natural ligand hormone. The three most hydrophobic interfaces are of standard-size and bury ≈1100 Å^2 of non-polar surface. No large interface is as non-polar as these three. Yet, the total amount of non-polar protein surface that is buried at a large interface can be huge, over 2000 Å^2 in the *E. coli* EFtu-EFts complex (1efu) for instance.

The non-polar fraction is very similar for contact atoms (58 %) and for interface atoms in general (56 %). This might be expected because the fraction is calculated on interface areas, of which contact atoms contribute three-quarters. Buried atoms are definitely more non-polar, 63 % on average, than other interface atoms, which remain partly accessible to the solvent.

The polar surface includes neutral and charged groups. On average, neutral polar groups contribute 29 % to the interface area. Charged side-chain groups contribute 15 %, with a slight excess of the positively charged surface over the negative one, whereas the accessible surface of the complexes is 18 % charged and equally distributed between positive and negative charges. We find that interfaces are somewhat depleted in negatively charged groups relative to the rest of the protein surface (6.7 % *versus* 9.1 %), but not in positively charged groups. Again, these numbers are averages and the contribution of charged groups varies widely from one complex to another, from zero at the erythropoietin receptor-peptide interface (1ebp), to 27 % at the Rap1A-Raf1 interface (1gua).

The polypeptide main-chain atoms contribute to 19 % of the interface area of the average protein-protein complex (Table 3). Side-chain atoms contribute 80 %, and non-protein groups 0.4 %. Again, these fractions are close to the average surface. They show that the main-chain contribution to interfaces cannot be ignored, especially that of the carbonyl oxygen which accounts for 11 % of the interface area, more than any other type of atom. Non-protein groups (HETATM in PDB files) are unimportant, with the exception of the FK506 immunosuppressant drug in the FKBP12 immunophilin-calcineurin complex (1tco), which is a major part of the interface and accounts for 30 % of the accessible surface area lost by the immunophilin when it binds to calcineurin. In a few other complexes, metal ions occur at the interface, but their buried surface is small.

The Character of Amino Acids at Interfaces

Table 4 describes the contribution of each of the 20 amino acid types to the interfaces and to the protein surface that remains accessible to solvent in the complexes. The two surfaces have amino acid compositions that are significantly different. Interfaces are much richer in aromatic residues His, Tyr, Phe and Trp than the average protein surface (21 % *versus* 8 %), and somewhat richer in aliphatic residues Leu, Ile, Val and Met (17 % *versus* 11 %). They are depleted in the charged residues Asp, Glu and Lys, but not Arg, which is the residue type that makes the largest overall contribution to interfaces (10 %). Some variations are seen between different types of complexes, but the large contribution of Arg and depletion in Lys are general. Protease-inhibitor interfaces are particularly rich in Cys (mostly from disulphide bridges), antigen-anti-

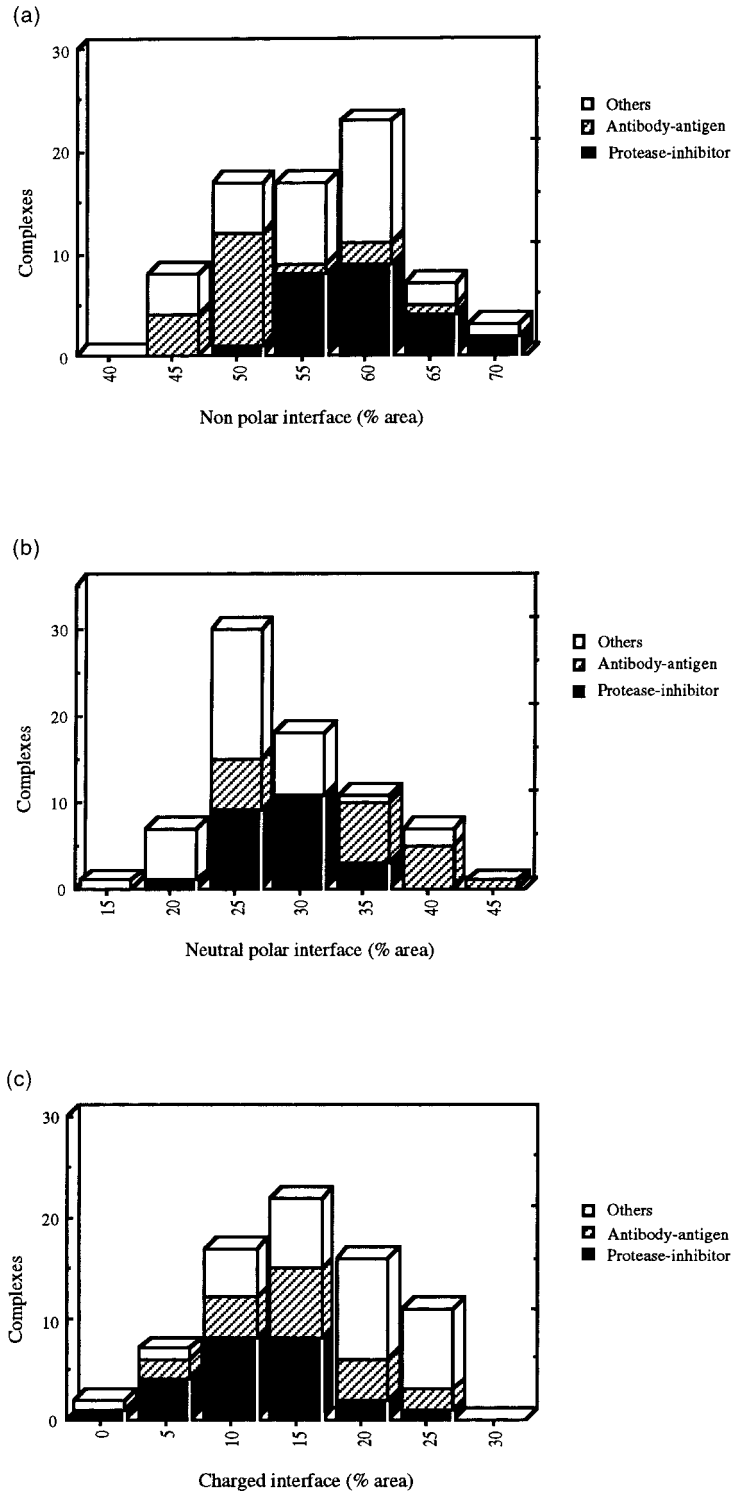

Figure 3. Histograms of the chemical character of protein-protein interfaces. Distributions of the values for the non-polar (top), neutral polar (middle) and charged (bottom) fractions of the interface areas given in Table 6.

body interfaces in Tyr, which contributes one-sixth of their area.

The amino acid composition of protein-protein interfaces is compared in Table 4 and Figure 4 to that of interfaces between subunits in 23 oligomeric proteins, and also to their accessible surface and to the protein surface buried upon subunit folding (Janin *et al.*, 1988). These compositions refer to the contribution of the residue types to the surface area, not to the number of residues. For pairwise comparison, a Euclidian distance Δf can be computed in the 19-dimension space of amino acid composition, from:

$$(\Delta f)^2 = 1/19\Sigma_i(f_i - f_1')^2$$

where f_i and f_i' are the percentage areas contributed by residue type i to the two surfaces. Distances

Table 4. Amino acid composition of protein-protein interfaces

Type	All complexes		Protease-inhibitor		Antibody-antigen		Others		Oligomeric proteins[a]		
	A	*B*	*A*	*B*	*A*	*B*	*A*	*B*	Surface	Interior	Interface
Ala	4.0	2.6	4.6	2.3	3.1	1.8	4.5	3.2	5.9	6.3	4.1
Arg	8.9	10.1	8.8	11.9	7.8	9.2	9.8	9.3	8.4	6.3	9.9
Asn	6.2	5.5	7.7	3.6	6.1	9.2	5.5	5.1	5.2	3.8	4.6
Asp	7.1	5.2	6.0	2.3	6.6	7.3	8.0	6.1	7.8	4.4	4.8
Cys	0.7	1.5	1.1	3.2	0.6	0.0	0.6	1.1	0.4	1.6	0.8
Gln	6.0	4.2	5.9	4.4	6.2	3.8	5.9	4.3	5.4	3.3	3.5
Glu	9.8	6.1	7.5	6.1	7.1	4.0	12.9	7.2	10.3	4.9	4.1
Gly	4.5	4.6	5.0	5.7	5.2	5.9	3.8	3.4	4.8	4.0	4.2
His	1.9	3.6	1.6	3.9	1.4	1.4	2.4	4.4	3.5	3.4	4.5
Ile	2.4	4.2	2.6	5.2	2.1	3.1	2.5	4.0	2.2	6.8	4.6
Leu	4.1	5.5	4.3	7.0	3.5	3.0	4.4	5.6	3.8	11.4	10.5
Lys	11.8	6.7	11.8	5.5	11.6	6.8	12.0	7.6	14.9	5.6	5.4
Met	1.2	3.2	0.7	3.2	0.6	0.8	1.8	4.2	1.5	2.9	3.9
Phe	2.0	4.4	2.4	4.4	1.4	3.0	2.2	5.0	1.9	5.9	6.0
Pro	5.1	4.0	5.3	4.8	5.2	2.6	4.8	4.2	5.6	3.6	5.3
Ser	8.4	5.5	8.5	5.7	12.6	7.4	5.1	4.6	6.3	4.5	4.1
Thr	7.3	5.1	6.6	4.7	9.8	6.4	5.7	4.8	5.5	4.8	4.7
Trp	1.3	4.5	1.2	5.2	1.6	5.7	1.1	3.4	0.8	2.5	2.4
Tyr	3.2	9.1	4.4	6.8	3.0	16.6	2.7	7.2	2.7	5.1	5.4
Val	3.5	3.8	3.6	3.9	3.5	1.5	3.5	4.7	3.2	9.1	7.3
Others[b]	0.7	0.6	0.3	0.2	1.0	0.5	0.7	0.9			
Ratio[c]	2.9	1.3	2.5	1.1	3.0	2.4	3.0	1.3	3.3	0.6	0.7

The amino acid compositions are given as per cent area contributions to the solvent-accessible surface area, *A*, and the interface area, *B*, in protein-protein complexes.

[a] Composition of the solvent accessible surface area, of the area of protein surface buried inside subunits (interior) and of the area of the interfaces between subunits in a sample of 23 oligomeric proteins (Janin *et al.*, 1988).

[b] Non-protein groups.

[c] Charged/hydrophobic ratio of the sums (Arg,Lys,Asp,Glu) and (Ile,Leu,Val,Phe,Met).

between some of the compositions in Table 4 are shown in Figure 4. The distance of 3.9 % between the accessible surface of oligomers and the surface buried upon folding, represents the difference in amino acid composition between the protein surface and interior. The distance between the average solvent-accessible surfaces of the complexes and the oligomeric proteins is 1.2 %, which may be seen as a background value.

Not surprisingly, the two types of interfaces are in between the protein surface and its interior. However, the composition of interfaces in protein-protein complexes is just half-way, whereas the composition of subunit interfaces in oligomers is much closer to the interior than to the surface. The average subunit interface in oligomers is also more hydrophobic (Janin *et al.*, 1988; Jones & Thornton, 1996) than the average protein-protein interface, and strongly depleted in charged groups. This can be seen from the amino acid composition by calculating the ratio of the fraction of charged amino acids Arg, Lys, Glu, Asp to that of hydrophobic

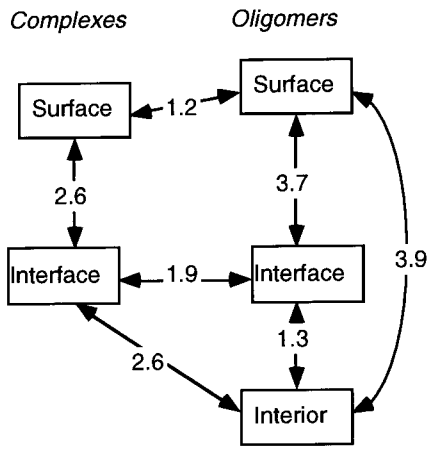

Figure 4. Distance between amino acid compositions. The numbers, expressed as per cent surface area, are distances between the amino acid compositions and are listed in Table 4.

Table 5. Polar interactions at protein-protein interfaces

Type	Fraction
A. *H-bond*[a]	
Main-chain-main-chain	0.24
Main-chain-side-chain	0.40
Side-chain-side-chain	0.36
Neutral	0.70
Charged[b]	0.30
Salt bridge	0.13
B. *Donor/acceptor group*[c]	
Main-chain N	0.18
O	0.26
Side-chain N^+ (Arg, Lys)	0.17
O^- (Asp, Glu)	0.14
Amide (Asn, Gln)	0.09
OH (Ser, Thr, Tyr)	0.13
Others (His, Trp, Cys)	0.03

[a] The average number of hydrogen bonds in 75 interfaces is 10.1.

[b] Charged H-bonds involve either one or two charged groups and include salt bridges.

[c] Fraction of the number of polar groups involved in all hydrogen bonds as donor or acceptor.

Table 6. Chemical character and atomic volumes

File	Chemical character (% interface area)			Interface atoms	Volume ratios[a]			
					Buried atoms		Buried with water	
	Non-polar	Polar	Charged		V/V_o	%B	V'/V_o	%B
A. *Protease-inhibitor*								
2ptc	55	31	14	163	0.99	38	1.00	75
1avw	56	31	13	195	0.99	35	1.00	63
1mct	58	31	10	173	0.99	33	0.98	61
3tpi	55	30	15	161	0.98	37	1.00	78
1tgs	61	34	4	195	1.00	31	1.00	68
1cho	64	27	9	163	0.99	26	1.00	66
1acb	68	26	5	176	0.99	29	0.99	42
1cbw	59	30	11	153	1.00	37		
1ppf	71	26	4	154	0.98	38	1.01	80
1fle	61	24	14	184	1.03	32	1.03	52
2kai	59	28	13	164	1.00	43		
1hia	58	27	15	169	1.03	32	1.00	53
3sgb	64	27	8	153	0.99	27	1.00	59
1mkw	65	25	10	159	0.98	7	0.98	25
1cse	61	31	8	156	0.99	39	0.97	82
2sic	60	33	7	182	0.99	40	1.00	87
2sni	64	28	8	187	1.01	44	1.00	62
1stf	62	35	1	198	1.00	27	1.00	51
4cpa	57	32	9	148	1.01	45		
B. *Large protease complexes*								
1bth	55	29	16	262	0.98	34	0.98	44
4htc	54	26	20	351	1.01	24	0.99	58
1tbq	57	24	19	374	1.04	26		
1toc	54	29	17	373	1.00	25		
1dan	52	21	26	380	1.05	14	1.02	41
C. *Antibody-antigen*								
1vfb	45	37	18	157	1.00	14	1.00	80
1mlc	52	35	13	151	1.06	35	1.04	50
1jhl	47	38	15	155	1.06	19		
3hfl	49	36	15	185	1.05	13		
3hfm	45	39	16	177	1.01	34		
1fbi	48	33	19	189	1.02	22		
1mel	61	27	12	183	1.01	35		
1jel	52	37	11	149	1.00	25		
1nsn	52	33	15	194	1.03	9		
1osp	51	25	25	160	1.01	14	1.01	65
1nca	50	40	10	223	1.01	32		
1nmb	48	40	7	153	1.03	13		
(*1)	61	23	16	178	0.97	23		
(*2)	64	24	12	198	0.98	18		
1dvf	42	38	18	186	1.00	23	1.04	59
1iai	51	45	4	217	1.04	16		
1nfd	48	24	25	187	0.97	22		
1kb5	48	35	17	249	1.04	16		
1ao7	55	27	18	230	1.02	25		
D. *Enzymes complexes*								
(*3)	59	23	18	281	1.02	25	1.00	64
1brs	52	23	25	184	1.00	25	1.00	98
1dfj	46	30	24	293	1.07	14		
1dhk	57	29	13	344	1.02	22	1.02	70
1fss	56	28	16	221	1.01	26		
1gla	62	16	23	136	1.00	14		
1udi	61	27	12	210	1.03	20		
1ydr	57	17	27	209	1.03	27	1.03	38
2pcc	50	25	23	111	1.00	5	1.07	53
E. *G-proteins, cell cycle, signal transduction*								
1tx4	55	25	18	234	1.02	31	1.00	98
1gua	44	29	27	135	1.02	23	1.01	43
1a2k	60	26	14	158	1.04	19		
1aip	61	24	15	306	1.02	26		
1efu	63	21	16	380	0.99	26		
1gg2	55	24	21	254	1.00	28	1.02	46
1got	57	25	18	266	1.02	28	1.00	72
2trc	49	25	22	502	1.01	27	1.00	42
1agr	52	26	23	184	1.05	22		
1fin	61	26	13	363	1.02	31	1.01	49
1a0b	59	21	20	122	1.01	21		

F. *Miscellaneous*								
1fc2	59	38	3	146	1.01	39		
1igc	52	40	8	138	1.07	35		
1ak4	60	28	12	132	0.95	22	0.95	23
1efn	61	22	18	134	0.98	33		
1atn	65	23	12	193	1.01	26		
2btf	55	22	23	224	1.04	28		
1dkg	61	19	20	212	1.00	17		
1ebp	69	31	0	196	1.01	31		
1hwg	57	23	19	476	1.05	29		
1seb	59	30	11	144	1.04	24		
1tco	47	25	13	259	0.98	21		
1ycs	44	33	23	168	0.97	26	0.99	41
Protease-inhibitor								
Mean	61	29	9	170	1.00	34	1.00	63
s.d.	4	3	4	16	0.01	9	0.01	16
Large protease complexes								
Mean	54	26	20	348	1.02	25		
s.d.	4	3	4	49	0.03	7		
Antibody-antigen								
Mean	51	33	15	185	1.01	21		
s.d.	4	3	4	29	0.03	8		
Enzymes complexes								
Mean	56	24	20	221	1.02	20		
s.d.	4	3	4	75	0.02	7		
G-proteins,cell cycle, signal transduction								
Mean	56	25	19	264	1.02	26		
s.d.	4	3	4	117	0.02	4		
Miscellaneous								
Mean	57	28	14	202	1.01	28		
s.d.	4	3	4	95	0.04	6		
All complexes								
Mean	56	29	15	211	1.01	26	1.00	59
s.d.	6	6	6	81	0.02	9	0.02	18

[a] V is the Voronoi volume of buried atoms; V_o, is the reference volume. The $\%B$ is the fraction of the interface area contributed by buried atoms. For V', solvent molecules reported in 36 X-ray structures with resolution 2.4 Å or better were included in the volume calculation. Due to the small number of these high-resolution structures in other categories, the mean and standard deviation are quoted only for protease-inhibitor and for all complexes.

amino acids Ile, Leu, Val, Phe, Met (Table 4). The ratio is close to 3 for the protein surface, and 0.7 for the surface buried either inside proteins or between subunits. It is more than 1 in all categories of protein-protein interfaces, particularly in antibody-antigen interfaces.

Polar Interactions: Hydrogen Bonds and Interface Water Molecules

Polar interactions at interfaces were determined using the program HBPLUS (McDonald *et al.*, 1994) and standard geometrical parameters. The number of such bonds, N_{hb} , in each complex is listed in Table 1. The average interface contains ten hydrogen bonds, but N_{hb} can be as low as 1 between cytochrome *c* and cytochrome peroxidase (2pcc) or as high as 34 between $G_{t\beta\gamma}$ and phosducin (2trc). Standard-size interfaces in protease-inhibitor and antibody-antigen complexes interfaces contain 9 (±5) hydrogen bonds. The N_{hb} value generally increases with the interface area, but the correlation is mediocre and sensitive to the quality of the X-ray structures. For the 36 structures with resolution of 2.4 Å or better, there is on average one hydrogen bond per 170 Å^2 of interface area B, and the correlation coefficient is 0.84. For structures at lower resolution, there are fewer hydrogen bonds and the correlation with interface area vanishes, which suggests that errors in atomic co-ordinates mask existing hydrogen bonds.

Table 5 shows that almost one hydrogen bond in three involves a charged side-chain, and that salt bridges are not rare. In bonds with only one charge, positive charges from Arg/Lys side-chains are somewhat more frequent than negative ones from Asp/Glu side-chains. The polypeptide main-chain is a major actor in polar interactions, being part of almost two out of three through main-chain to main-chain and main-chain to side-chain hydrogen-bonds. The carbonyl oxygen is the acceptor group most frequently used in hydrogen bonding at protein interfaces.

We analysed the role of water molecules found at interfaces. Because the positioning of solvent atoms is generally less accurate than for protein atoms, we restricted our analysis to structures determined at a resolution of 2.4 Å or better. Water molecules were assumed to be at the interface if they were within the sum of the Van der Waals radii, plus 1 Å of atoms of both proteins. The great majority form hydrogen bonds with both sides. The number (N_{wat}) of interface water molecules in the 36 structures is listed in Table 1. It is 18 on average, but the range is wide, from 3 to 50. There is about one interface water per 100 Å^2 of

interface area B, suggesting that water-mediated polar interactions are more numerous than direct hydrogen bonds, but the correlation between N_{wat} and B is poor (0.5). The different practises of crystallographers in locating and listing solvent molecules are responsible for at least some of this variation. Still, it is evident that there are interfaces from which water is excluded and where solvent molecules are distributed on the periphery of the contact, and others where they occur in central regions of the interface. This can be seen by examining the complexes individually and will be discussed below where it is illustrated in Figure 6.

Atomic packing at interfaces

The volumes occupied by atoms inside proteins or buried in their interfaces can be determined by constructing Voronoi polyhedra around their atoms, and calculating their volume. The packing density at interfaces was determined by calculating the volume of the Voronoi polyhedron of each atom buried in the interface, summing the values to give a total volume V, and comparing V to a reference value V_o. To derive V_o, we used the mean volumes that the atomic groups of residues occupy in protein interiors. A V/V_o ratio larger than unity means the packing density at interfaces is lower than that in protein interiors, and a ratio less than unity means a higher packing density.

Volumes of atomic groups buried in the interfaces were calculated using an implementation of the Voronoi procedure that is based on an original program by Richards (1974), with its subsequent modifications and extensions by Harpaz *et al.* (1994) and by Gerstein *et al.* (1995). The same program has been used to determine the reference values for the average volume of atoms buried inside globular proteins (Harpaz *et al.*, 1994; Tsai *et al.*, unpublished results). These volumes are, on average, 5% smaller than the volumes of equivalent groups in small molecule crystals, of amino acids, indicating that the protein inside is better packed on average than small molecule crystals in spite of the occasional presence of internal cavities which increase the average atomic volume. As small molecule crystals are usually considered to be close-packed, volumes observed inside proteins are a good reference for atomic packing.

For a Voronoi polyhedron to be built around an atom, it must be fully surrounded by other atoms. This implies that the calculation can only be carried out on buried atoms unless the structure of the solvent is known. Only about one-third of all interface atoms are buried. We calculated their volumes in each of the 75 complexes; Table 6 lists the values of the V/V_o volume ratio observed at each interface. A histogram of these values is shown in Figure 5(a). The mean is very close to unity and the standard deviation small. The range of V/V_o is 1.01(±0.06) for all complexes, and 1.01(±0.03) for all but 11. Thus, the packing density within protein-protein interfaces analysed here is very similar to that within the protein interior.

However, the volume measurement applies only to about one-third of the interface atoms, and says nothing of the two-thirds that have residual accessibility in the complex. In the smaller interfaces, buried interface atoms are few. The cytochrome *c* oxydase-cytochrome *c* interface (2pcc) fully buries only 14 atoms; the cyclophilin-HIV capsid complex (1ak4), which has the lowest of all volume ratios (0.95), buries 38 atoms. On so few atoms, the volume ratio is poorly defined and the finding that they are closely packed may be thought to be insignificant. On the other hand, it is certainly meaningful for a larger interface like the transducin $G_{t\beta\gamma}$-phosducin complex (2trc) that buries 173 atoms, which occupy a volume equivalent to the interior volume of a small protein and are close-packed with a volume ratio of 1.01.

Two complexes have high volume ratios near 1.07: the RNase A-inhibitor (1dfj) and the Fab MOPC21-protein G (1igc) complex. The RNase inhibitor is shaped like a horseshoe and forms a large interface while sequestering the enzyme (Kobe & Deisenhofer, 1995). Protein G binds to the constant domain (not to the antigen-combining site of the antibody), and forms a standard size interface. In spite of the different interface size, both complexes bury about 60 atoms. Errors in atomic positions due to the limited resolution (2.5-2.6 Å) of the two X-ray structures may contribute to the poor packing suggested by the volume ratio. Our data indicate that the less accurate structures tend to have a high volume ratio at the interface. However, the bias is small at less than 2%. The mean value of the volume ratio is 1.004 for 22 X-ray structures of complexes with a resolution of 2.0 Å or better, 1.016 for 38 structures with a resolution of 2.5 Å or worse. Thus, the high value of V/V_o in the RNase A-inhibitor and Fab-protein G complexes may be real, and indicate that these two interfaces are poorly packed compared with others.

Because two-thirds of the interface atoms have non-zero solvent accessibility, their volume cannot be calculated in the absence of information of the structure of the solvent molecules with which they are in contact. However, many of the data sets report at least some of the solvent positions which, if included in the calculation, increases the proportion of interface atoms with zero accessibility; 36 of them have resolution 2.4 Å or better. For these structures, we could calculate volumes for some protein atoms that are partly solvent accessible by taking reported solvent positions into account when drawing the Voronoi polyhedra. No volume was calculated for the solvent molecules.

The new volume ratio noted V'/V_o in the 36 interfaces is given in Table 6, and a histogram of its values is shown in Figure 5(b). The range is the same as for V/V_o, but 16 values are within 0.5% of 1.00. For 17 of the complexes, over 60% of the interface atoms are included; the mean V'/V_o ratio

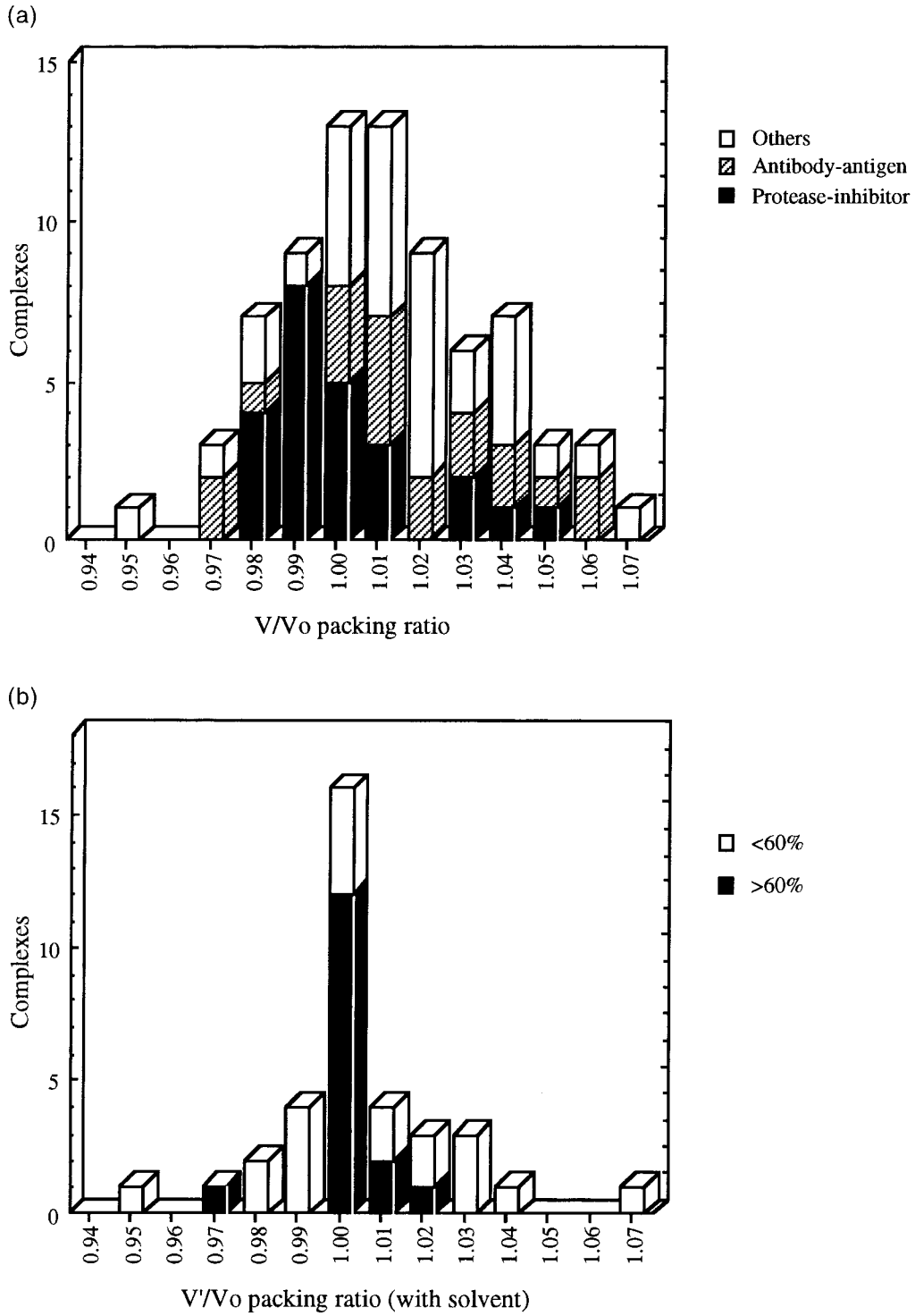

Figure 5. Histogram of packing densities for buried atoms. (a) Distribution of the values of the V/V_o ratio cited in Table 6 for atoms fully buried at the 75 protein-protein interfaces. V is the sum of their Voronoi volumes, V_o, is a reference volume observed for atoms buried inside proteins. (b) Distribution of the values of the V'/V_o ratio in 36 structures with resolution of 2.4 Å or better. The Voronoi volume V' is calculated for interface atoms in the presence of the solvent molecules reported in PDB files. This increases the fraction of interface atoms with zero accessibility. In 17 complexes (black columns), these atoms account for more than 60% of the interface area.

is 1.00, and the standard deviation 0.01. In the barnase-barstar (1brs) and the Rho-Rho GAP (1tx4) complexes, the calculation involved almost all the interface atoms. Both have V'/V_o ratios of 1.00 and, therefore, these two interfaces are close-packed over their entire extent on the condition that solvent molecules are taken into account. As only 33-35% of their interface atoms are actually buried, this shows how important water molecules are in the packing of these two interfaces.

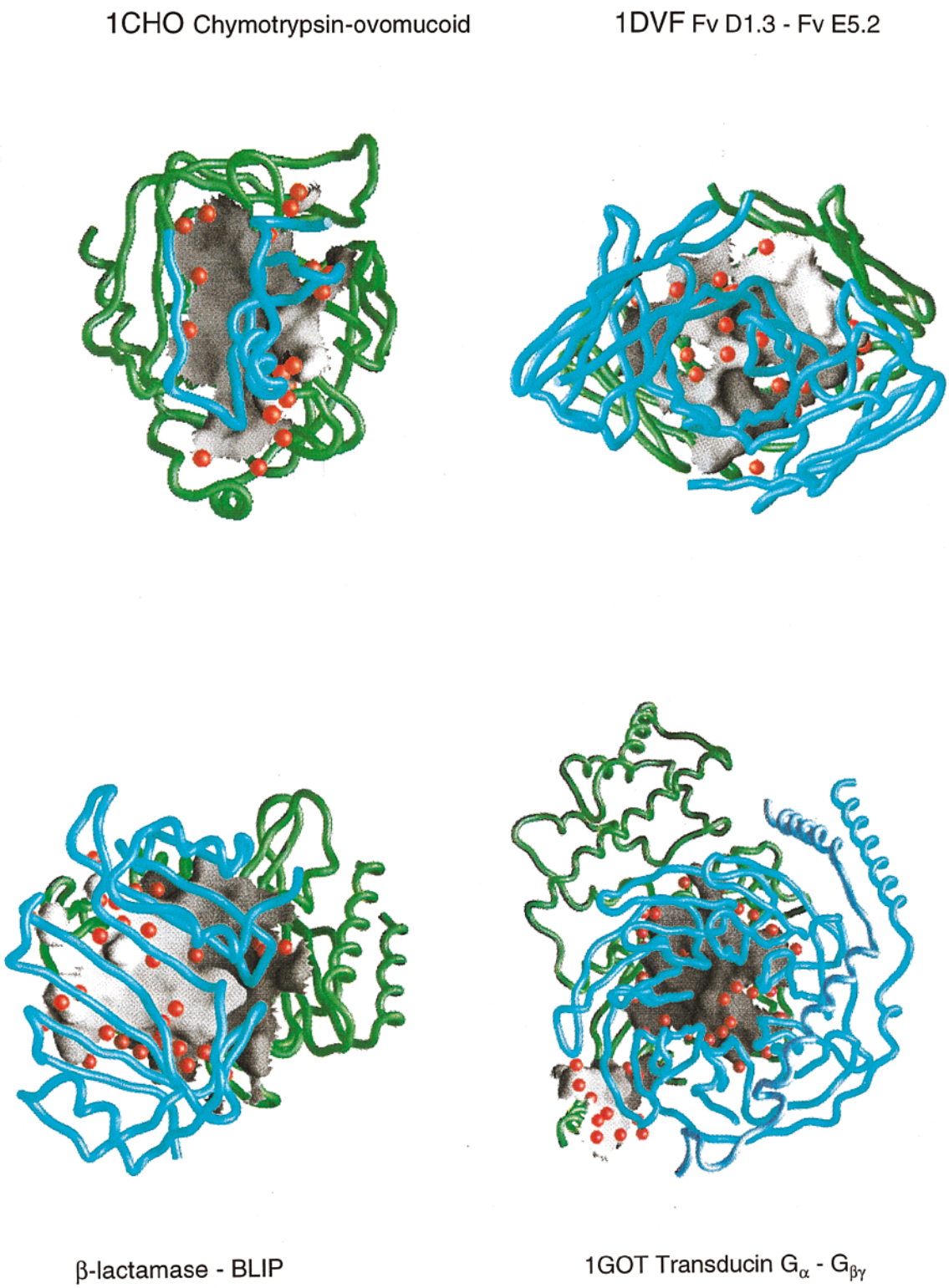

Figure 6. Interfaces in four protein-protein complexes. The backbone of the two components are coloured blue and green. The shaded surface belongs to the green component at the back, and is in contact with the blue component in front. Red spheres are interface water molecules. From left to right and top to bottom: 1cho, complex of chymotrypsin on the back with the turkey ovomucoid third domain inhibitor (Fujinaga *et al.*, 1987); 1dvf, complex between two Fv fragments of antibody D1.3 and the anti-idiotypic antibody E5.2 (Braden *et al.*, 1996); *3, complex of RTEM β-lactamase on the back with the BLIP inhibitor (Strynadka *et al.*, 1996); 1got, the transducin $G_{t\alpha}$-$G_{t\beta\gamma}$ heterotrimer (Lambright *et al.*, 1996); $G_{t\alpha}$ is at the back. The Figure was drawn with GRASP (Nicholls *et al.*, 1992).

Diversity of interfaces

The analysis of the atomic structure presented in the previous sections shows that recognition sites have some features that all, or almost all, share and other features that vary. To illustrate these general points at a more detailed level we describe and discuss here the recognition sites in two complexes that have standard size interfaces and two complexes with large interfaces. All four structures were determined at high resolution (2 Å or better). The recognition sites in these complexes are illustrated in Figure 6.

The two standard size interfaces are between chymotrypsin and the ovomucoid inhibitor (1cho; Fujinaga *et al.*, 1987) and between Fv D1.3 and the anti-idiotypic Fv E5.2 immunoglobulin fragment (1dvf; Braden *et al.*, 1996). Their general features are representative of other protease-inhibitor and antigen-antibody complexes. They have pico- to nanomolar affinity ($K_d = 5$ pM for chymotrypsin-turkey ovomucoid; Bigler *et al.*, 1993; $K_d = 11$ nM for D1.3-E5.2; Goldman *et al.*, 1997). Association takes place with little change in the structure of the components. The two interfaces have areas near 1600 Å^2 and use a single patch on the surface of each component; buried atoms are well packed, with volume ratios very close to 1. They contain the same number (ten) of hydrogen bonds, and approximately the same number (15-17) of water molecules.

The two complexes also have features that are not the same. They differ in the hydrophobic/hydrophilic character of their interfaces. The chymotrypsin-inhibitor interface is largely non-polar (64%), the Fv-Fv interface much less so (42%). This may be why water molecules form a ring around the more hydrophobic chymotrypsin-inhibitor interface, whereas they are distributed over most of the more polar Fv-Fv interface. Chymotrypsin-inhibitor contacts mostly involve extended segments of the two polypeptide chains, and the peptide group contributes more hydrogen bonds than side-chains. Contacts between the two Fv fragments involve loops from the complementarity-determining regions of the two antibodies, and side-chains contribute most of the hydrogen bonds. The chymotrypsin-inhibitor interface is depleted in charged groups and has only one charged hydrogen bond, whilst the Fv-Fv interface has six (none of them a salt bridge, however).

As examples of large interfaces, we describe those between β-lactamase and the BLIP inhibitor (*3; Strynadka *et al.*, 1996) and between the $G_{t\alpha}$ and $G_{t\beta\gamma}$ subunits of transducin (1got), a trimeric G-protein (Lambright *et al.*, 1996). Both have interface areas near 2500 Å^2 and polar characters that are close to the average in our sample. Buried atoms have a volume ratio near 1.02 ignoring water, and 1.00 taking water into account. Thus, they close-packed over large areas.

The two large complexes are otherwise quite different. The β-lactamase-inhibitor complex has picomolar affinity (Strynadka *et al.*, 1996). The contact region is a single large patch covering both the active site of the enzyme and an adjacent convex surface made of a protruding loop and helix. On the BLIP inhibitor, the contact region is the concave face of a large saddle-shaped β-sheet. The enzyme changes little upon association, but the inhibitor bends like a clamp over the protruding region of the enzyme to which it binds. The interface has 11 hydrogen-bonds, all involving side-chains, and all but one charged, three being salt bridges. It also contains 31 water molecules which are distributed around and within the interface, and provide additional polar bonds bridging the two components of the complex.

The interaction between the $G_{t\alpha}$ and $G_{t\beta\gamma}$ components of transducin involves two regions of $G_{t\alpha}$ forming two discrete patches. The $G_{i\alpha}$-$G_{i\beta\gamma}$ complex (1gg2) has very similar features. One patch comprises the N-terminal helix and buries approximately 900 Å^2. The helix is missing in the free $G_{t\alpha}$ structure (Lambright *et al.*, 1994) and largely disordered in $G_{i\alpha}$. The affinity of $G_{t\alpha}$ for $G_{t\beta\gamma}$ is weak (micromolar or worse) when the N-terminal helix is deleted. In the natural form of transducin, $G_{t\alpha}$ and $G_{t\gamma}$ are acylated and their attachment to the membrane profoundly affects the stability of the complex (Bigay *et al.*, 1994). The second patch of the interface buries 1700 Å^2, equivalent to a standard-size interface. It involves regions called switch I and switch II of $G_{t\alpha}$ (Noel *et al.*, 1993). These regions are close to the phosphates of the bound GTP/GDP nucleotide and undergo major conformation changes upon GTP hydrolysis. Upon $G_{t\beta\gamma}$ binding, a different but even larger conformation change is observed in $G_{t\alpha}$ (Lambright *et al.*, 1994, 1996). A turn of 3_{10}-helix unfolds, and large amplitude main-chain and side-chain movements remodel the protein surface in this region. On the $G_{t\beta\gamma}$ side of the contact, the $G_{t\beta}$ subunit, a protein with a characteristic β-propeller fold, has loops connecting β-strands in the β-propeller to form the $G_{t\alpha}$ binding site. Conformation changes are less dramatic than in $G_{t\alpha}$, yet a comparison with the free structure (Sondek *et al.*, 1996) shows a loop in $G_{t\beta}$ moving by up to 6 Å. The $G_{t\alpha}$-$G_{t\beta\gamma}$ interface contains 18 hydrogen bonds, of which all but one involve side-chains; nine are charged, including five salt bridges. It also contains 36 solvent molecules, located mostly at the periphery of the two patches.

Discussion

Early analyses of protein-protein recognition sites were based on a small number of structures, nearly exclusively protease-inhibitor and antigen-lysozyme complexes (Chothia & Janin, 1975; Janin & Chothia, 1990). They revealed a remarkable homogeneity in the size of the interfaces measured by the interface area, and suggested that regions of the protein surface that form recognition sites do

not differ greatly in their chemical character from the rest of the solvent-accessible surface. The interfaces have a large non-polar component which provides hydrophobic free energy in favour of association as it becomes buried in the complexes, and a polar component providing hydrogen bonds. Both are essential to recognition. Recognition was shown to occur between protein surfaces of similar character having complementary shapes and excluding water over at least 1200 Å^2 as the two partners come into contact. It was hypothesised that this is the minimum size of an interface yielding a stable structure. In most cases, the shapes were preformed and conformation changes limited to local movements of small amplitude: complexes formed by rigid-body association.

Standard-size, small and large interfaces

The present study uses a much larger sample. Protease-inhibitor and antibody-lysozyme complexes are still in the majority, but there is a number of examples of recognition sites involved in very different physiological processes. Their analysis confirms earlier conclusions and introduces new ones, because the diversity of interfaces is much greater. Standard-size interfaces with B =1600 ($\pm$400) Å^2 correspond to the description of protein-protein recognition sites given by Janin & Chothia (1990). Here, we see that the range of structures in which they occur is extended. In addition to protease-inhibitor and antibody-antigen complexes, they occur in about half of the other complexes. Below and above the limits of standard-size interfaces, novel features are observed. The smallest interfaces have $B = 1150$ Å^2. Though they are few at present, we suggest that they represent a type of interaction in which proteins form short-lived, low-stability complexes. The complex between cytochrome peroxydase and cytochrome c (2pcc) is an example. Cytochrome *c* is a substrate of the enzyme, and must not bind too tightly for turnover to take place. This type of interaction is essential in many cellular processes, but still poorly represented in the PDB.

Large interfaces are better represented, as they form 27% of the present sample. They were essentially absent in earlier studies. Interfaces with B in the range 2000-4660 Å^2 occur mostly between proteases and a particular class of inhibitors and between G-proteins and other components of the signal transduction system. By their size, they resemble subunit interfaces in oligomeric proteins analysed by Miller *et al.* (1987b), Janin *et al.* (1988), Argos (1988), Jones & Thornton (1995, 1996) and Tsai & Nussinov (1997). However, they are less hydrophobic and contain a larger proportion of charged groups. This reflects the fact that the protein components of a complex must be stable and soluble in water, whereas isolated subunits of oligomeric proteins are generally not stable and water soluble. Yet, we find another point of similarity between large interfaces in complexes and in oligomeric proteins: their formation is generally accompanied by major conformation changes. In most oligomeric proteins, association is coupled with subunit folding. The disorder-to-order transition illustrated by the C-terminal tail of hirudin (4htc) and the N-terminal helix of transducin G_α belongs to that category. At many large interfaces, the polypeptide chain partly refolds, loops move and the shape of the protein surface changes completely upon association. Thus, the complementary regions that form recognition sites do not pre-exist association as they do in standard-size interfaces.

Close-packing of atoms at recognition sites

We reported here that the atoms buried in the interfaces of trypsin-PTI and the Fab HyHEL-lysozyme complexes have packing densities like those in organic solids (Janin & Chothia, 1976, 1990). Here, we have extended this work in two directions. First, we have carried out calculations of Voronoi volumes on a large number of complexes. Second, in structures where the solvent structure at the interface is known, we can calculate densities for atoms that are not fully buried. For interface atoms that are buried in 75 complexes, which represents one-third of all interface atoms, the packing density derived from the Voronoi volume is within 7% of that of the protein interior. In most of the complexes, it is within 4%. Atoms in protein interiors have mean volumes that are 5% smaller than those found in crystals of amino acids (Harpaz *et al.*, 1994; Tsai *et al.*, unpublished results). Thus, close-packing of atoms buried at protein-protein interfaces is equal to or better than in organic solids. By taking into account solvent molecules observed in high-resolution X-ray structures, we can extend this conclusion to a larger fraction of interface atoms which make up more than 60% of the interface in 17 complexes and almost 100% in two complexes.

Comparison with other studies

Most surveys of protein-protein interaction (Miller *et al.*, 1987b; Janin *et al.*, 1988; Argos, 1988; Jones & Thornton, 1995, 1996, 1997; Tsai *et al.*, 1996; Xu *et al.*, 1997) deal principally with oligomeric proteins, which we do not consider here. However, the study by Jones & Thornton (1996) has 19 structures in common with our study, mostly protease-inhibitor and antigen-antibody complexes. The statistics on interface areas†, hydrophobicity and hydrogen bonds agree with our study for these types of complexes. For other

† ΔASA values in Jones & Thornton (1996) are given per subunit and differ by a factor of 2 from interface areas quoted here.

types, the overlap of the structural samples is too small and heterogeneous in composition for a comparison to be meaningful. A common conclusion is that, except perhaps in protease-inhibitor complexes, interfaces are no more hydrophobic than the rest of the protein surface. This finding is at variance with the general description of recognition sites as hydrophobic clusters given by Young *et al.* (1994), a description that applies only to protease-inhibitor complexes. The amino acid composition of interfaces reported here agrees with residue propensities cited by Jones & Thornton (1996) for interfaces in heterocomplexes.

Geometric complementarity

Several methods have been proposed to estimate geometric complementarity. Lawrence & Colman (1993) calculated a shape correlation index based on distance and the angle of the normal vectors to the molecular surface. Jones & Thornton (1996) proposed a gap index which is the volume of cavities between the components of complexes divided by the interface area. Values of the shape correlation and of the gap indices were taken to indicate that antigen-antibody interfaces are less complementary than interfaces in oligomeric proteins and in protease-inhibitor complexes. Whereas neither quantity can be directly compared with the ratio of Voronoi volumes we use, our data suggest that the difference, if real, must be marginal. We find the average value of the volume ratio is 1.00 in standard size protease-inhibitor interfaces (Table 6) and 1.01 at antigen-antibody interfaces. In both cases, buried atoms are better packed than small molecule crystals, and cavities are no more abundant than inside monomeric proteins. When water is included in the calculation, this conclusion applies to nearly two-thirds of the interface area in both types of complexes.

The apparent contradiction may originate from the presence of water at interfaces and its role in packing. Solvent was present in only one of the antigen-antibody complexes analysed by Lawrence & Colman (1993), and not considered in the gap index of Jones & Thornton (1996). We find that crystallographic water molecules surround the contact region and may also fill cavities within it. The number of water molecules is about the same in standard-size protease-inhibitor and antibody-antigen interfaces (Table 1) and their importance in recognition has been emphasised in both types of complexes (Huang *et al.*, 1995; Bhat *et al.*, 1994). Water molecules that surround an interface are illustrated in Figure 6 by the chymotrypsin-ovomucoid protease-inhibitor complex; an example of water molecules within the interface is the D1.3-E5.2 antibody-antigen complex. Both have well packed interfaces, but water plays a greater part in the packing of the second category. Poorly packed interfaces may exist (Ysern *et al.*, 1998), and if there is an example in our sample, it is in the RNase A-inhibitor or the protein G complexes; there are no examples in antigen-antibody complexes.

Structure of interfaces and the effects of mutations

Atoms that form protein-protein recognition sites can be put into one of three classes: atoms that lose accessibility but do not make direct contacts across the interface; atoms that make direct contacts but remain partly accessible; and atoms that become buried. This description of interfaces, sketched in Figure 2, should help in interpreting data on the effect of site-directed mutagenesis on the stability of protein-protein complexes. Systematic studies of this type have been performed on several of the complexes considered here: trypsin and chymotrypsin with PTI (Castro & Anderson, 1996), antibody D1.3 with lysozyme or antibody E5.2 (Dall'Acqua *et al.*, 1996, 1998; Goldman *et al.*, 1997), barnase and barstar (Schreiber & Fersht, 1993, 1995), the human growth hormone and its receptor (Cunningham & Wells, 1993; de Vos *et al.*, 1992; Clackson & Wells, 1995; Wells, 1996), factor VIIA and soluble tissue factor (Kelley *et al.*, 1995).

Bogan & Thorn (1998) recently compiled data on affinity changes, expressed as changes in dissociation free enthalpies ($\Delta\Delta G_d$), that occur when alanine is substituted for individual contact residues. They proposed a O-ring model of interfaces similar to that illustrated in Figure 2(b), yet based on a very different approach. They found that almost all residues that yield large $\Delta\Delta G_d$ values have very little accessibility in the complex, and that they tend to cluster at the centre of the interfaces. This finding may help in solving an apparent contradiction between the structural data and a major conclusion of the mutagenesis studies: that only a small number of residues matter in terms of affinity. In barnase-barstar, 43 residues lose accessible surface, a typical value for a standard-size interface. Yet, only 11 yield $\Delta\Delta G_d > 1$ kcal/mol when mutated to alanine. In the hormone receptor-human growth hormone complex, the numbers are 115 and 25. Thus, in these two systems, about three-quarters of the interface residues can be substituted with alanine with no or little effect on affinity. The difference is less (48 *versus* 22) in the antibody D1.3-antibody E5.2 system.

To understand why the ''functional epitope'' defined by mutagenesis is much smaller than the ''structural epitope'' seen in the three-dimensional structures (Cunningham & Wells, 1993), several features of interfaces should be taken into account. First, the polypeptide main-chain is inaccessible to site-directed mutagenesis. However, it is an important contributor to interfaces as it represents about one-fifth of the interface area and contributes nearly two-thirds of the hydrogen bonds. Second, three-quarters of the interface area comes from atoms that remain partly accessible to solvent. When these atoms are deleted in an alanine mutant, they can be replaced by water molecules

at much less cost than fully buried atoms. In contrast, deleting buried atoms which we find to be close-packed optimising Van der Waals interactions, cannot be done without a cost.

Conclusion

The examples of antibody-antigen, protease-inhibitors and many other complexes show that a standard-size interface is sufficient for both stability and specificity. These standard recognition sites form stable protein-protein complexes irrespective of the size of the proteins, and no large conformational change is involved in their formation. On average, these interfaces bury 1600 Å^2, have a chemical character that is close to the average protein surface, and involve nine hydrogen bonds and the same packing density as the protein interior. Individual complexes have variations around these mean values, which are large in the case of their chemical character but small in the case of packing densities. These variations are not apparently correlated with each other, with the size of the components or with affinity constants. They balance each other in a manner not open to simple structural analysis.

Large interfaces are not designed to improve stability. In transducin or in the bacterial elongation factors, stable association of the subunits could presumably be achieved by forming a standard interface. However, it would serve no function: in the G-proteins, subunit assembly is tightly coupled with the hydrolysis or the release of the bound nucleotide through large conformation changes. In transducin, the changes mediate the visual signal; in elongation factors, they allow aminoacyl-tRNA to bind. Compared with those between trypsin and PTI or barnase and barstar, the interfaces between the $G_{t\alpha}$ and $G_{t\beta\gamma}$ subunits of transducin or the Ts and Tu subunits of the elongation factors, have a similar atomic composition and they are equally well-packed, but they are much larger. This must be the source of additional free energy for the conformation changes that enable these proteins to carry their function.

Acknowledgements

L.L. thanks IBM for an IBM Cooperative Fellowship and The Fondazione Cassa di Risparmio and the Compagnia di San Paolo of Torino for support. J.J. acknowledges financial support from EMBL-European Bioinformatics Institute (Hinxton, Cambridge, UK) and Université Paris-Sud (Orsay, France) during a sabbatical leave. We thank Dr S. Wodak for discussion, Dr N. Strynadka, S. Smerdon, M. Knossow, D. Fleury, N. Chinardet and H.K. Song for providing co-ordinates prior to deposition.

References

Argos, P. (1988). An investigation of protein subunit and domain interfaces. *Protein Eng.* **2**, 101-113.

Ban, N., Escobar, C., Garcia, R., Hasel, K., Day, J., Greenwood, A. & McPherson, A. (1994). Crystal structure of an idiotype-anti-idiotype Fab complex. *Proc. Natl Acad. Sci. USA,* **91**, 1604-1608.

Banner, D. W., d'Arcy, A., Chene, C., Winkler, F., Guha, A., Konigsberg, W. H., Nemerson, Y. & Kirschhofer, D. (1996). The crystal structure of the complex of blood coagulation factor VIIA with soluble tissue factor. *Nature,* **380**, 41-46.

Bernstein, F. C., Koetzle, T. F., Williams, J. B., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank. A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I., Tello, D., Dall'Acqua, W., Souchon, H., Schwarz, F. P., Mariuzza, R. A. & Poljak, R. J. (1994). Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl Acad. Sci. USA,* **91**, 1089-1093.

Bigay, J., Faurobert, E., Franco, M. & Chabre, M. (1994). Roles of lipid modifications of transducin subunits in their GDP-dependent association and membrane binding. *Biochemistry,* **33**, 14081-14090.

Bigler, T. L., Lu, W., Park, S. J., Tashiro, M., Wieczorek, M., Wynn, R. & Laskowski, M., Jr (1993). Binding of aminoacid side chains to preformed cavities: interaction of serine proteases with turkey ovomucoid third domains with coded and noncoded P1 residues. *Protein Sci.* **2**, 786-799.

Bode, W., Schwager, P. & Huber, R. (1978). The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. *J. Mol. Biol.* **118**, 99-112.

Bode, W., Wei, A. Z., Huber, R., Meyer, E., Travis, J. & Neumann, S. (1986). X-ray crystal structure of the complex of human leukocyte (Pmn elastase) and the third domain of the turkey ovomucoid inhibitor. *EMBO J.* **5**, 2453-2458.

Bode, W., Papamokas, E. & Musil, D. (1987). The high resolution X-ray crystal structure of the complex formed between subtilsin Carlsberg and eglin C, an elastase inhibitor from the leech *Hirudo medicinalis*. *Eur. J. Biochem.* **166**, 673-692.

Bogan, A. A. & Thorn, K. S. (1998). An analysis of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1-9.

Bolognesi, M., Gatti, G., Menegatti, E., Guarneri, M., Marquart, M., Papamokos, E. & Huber, R. (1982). Three-dimensional structure of the complex between pancreatic secretory inhibitor (Kazal type) and trypsinogen at 1.8 Å resolution. *J. Mol. Biol.* **162**, 839-868.

Bompard-Gilles, C., Rousseau, P., Rougé, P. & Payan, P. (1996). Substrate mimicry in the active center of mammalian α-amylase: structural analysis of an enzyme-inhibitor complex. *Structure,* **4**, 1441-1452.

Bossart-Whitaker, P., Chang, C. Y., Novotny, J., Benjamin, D. C. & Sheriff, S. (1995). The crystal structure of the antibody N10–Staphylococcal nuclease complex at 2.9 Å resolution. *J. Mol. Biol.* **253**, 559-575.

Braden, B. C., Souchon, H., Eiselé, J.-L., Bentley, G. A., Bhat, T. N., Navaza, J. N. & Poljak, R. J. (1994). Three-dimensional structures of the free and anti-

gen-complexed Fab from monoclonal anti-lysozyme antibody D44.1. *J. Mol. Biol.* **242**, 767-781.

Braden, B. C., Fields, B. A., Ysern, X., Dall'Acqua, W., Goldbaum, F. A., Poljak, R. J. & Mariuzza, R. A. (1996). Crystal structure of an Fv-Fv idiotope-anti-idiotope complex at 1.9 Å resolution. *J. Mol. Biol.* **264**, 137-151.

Buckle, A. M., Schreiber, G. & Fersht, A. R. (1994). Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0 Å resolution. *Biochemistry,* **33**, 8878-8889.

Castro, M. J. & Anderson, S. (1996). Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: effects on the kinetics and thermodynamics of binding to β-trypsin and a-chymotrypsin. *Biochemistry,* **35**, 11435-11446.

Chen, Z. & Bode, W. (1983). Refined 2.5 Å X-ray crystal structure of the complex formed by porcine kallikrein A and the bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **164**, 283-311.

Chitarra, V., Alzari, P. M., Bentley, G. A., Bhat, T. N., Eisele, J. L., Houdusse, A., Lescar, J., Souchon, H. & Poljak, R. J. (1993). Three-dimensional structure of a heteroclitic antigen-antibody cross-reaction complex. *Proc. Natl Acad. Sci. USA,* **90**, 7711-7715.

Chothia, C. (1975). Structural invariants in protein folding. *Nature,* **254**, 304-308.

Chothia, C. (1997). Protein-protein and protein-carbohydrate recognition. In *Molecular Aspects of Host-Pathogen Interaction* (McCrae, M. A., Saunders, J. R., Smyth, C. J. & Stow, N. D., eds), Cambridge University Press.

Chothia, C. & Janin, J. (1975). Principles of protein-protein recognition. *Nature,* **256**, 705-708.

Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science,* **267**, 383-386.

Cunningham, B. C. & Wells, J. A. (1993). Comparison of a structural and functional epitope. *J. Mol. Biol.* **234**, 554-563.

Dall'Acqua, W., Goldman, E. R., Eisenstein, E. & Mariuzza, R. A. (1996). A mutational analysis of bthe inding of two different proteins to the same antibody. *Biochemistry,* **35**, 9667-9676.

Dall'Acqua, W., Goldman, E. R., Lin, W., Teng, C., Tsuchiya, D., Li, H., Ysern, X., Braden, B. C., Li, Y., Smith-Gill, S. J. & Mariuzza, R. A. (1998). A mutational analysis of binding interactions in an antigen-antibody protein-protein complex. *Biochemistry,* **37**, 7981-7991.

Davies, D. R. & Cohen, G. H. (1996). Interactions of protein antigens with antibodies. *Proc. Natl Acad. Sci. USA,* **93**, 7-12.

Deisenhofer, J. (1981). Crystallographic refinement and atonic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9 and 2.8 Å resolution. *Biochemistry,* **20**, 2361-2370.

Derrick, J. P. & Wigley, D. B. (1994). The third IgG-binding domain from streptococcal protein G. *J. Mol. Biol.* **243**, 906-18.

Desmyter, A., Transue, T. R., Arbabi-Ghahroudi, M., Dao-Thi, M., Poortmans, F., Hamers, R., Muyldermans, S. & Wyns, L. (1996). Crystal structure of a single-domain Vh antibody fragment in complex with lysozyme. *Nature Struct. Biol.* **3**, 803-811.

Engh, R. A., Girod, A., Kinzel, V., Huber, R. & Bossemeyer, D. (1996). Crystal structures of catalytic subunit of cAMP-dependent protein kinase in complex with isoquinolinesulfonyl protein kinase inhibitor H7, H8 and H89. Structural implications for selectivity. *J. Biol. Chem.* **271**, 26157-26164.

Frigerio, F., Coda, A., Puglise, L., Lionetti, C., Menegatti, E., Amiconi, G., Schnebli, H. P., Ascenzi, P. & Bolognesi, M. (1992). Crystal and molecular structure of the bovine a-chymotrypsin-eglin C complex at 2.0 Å resolution. *J. Mol. Biol.* **225**, 107-123.

Fujinaga, M., Sielecki, R., Read, R. J., Ardelt, W., Laskowski, M. & James, M. N. J. (1987). Crystal and molecular structures of α-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution. *J. Mol. Biol.* **195**, 397-418.

Gamble, T. R., Vajdos, F. F., Yoo, S., Worthylake, D. K., Houseweart, M., Sundquist, W. I. & Hill, C. P. (1996). Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV1 capsid. *Cell,* **87**, 1285-1294.

Garboczi, D. N., Ghosh, P., Utz, U., Fan, Q. R., Biddison, W. E. & Wiley, D. C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature,* **384**, 134-141.

Gaudet, R., Bohm, A. & Sigler, P. B. (1996). Crystal structure at 2.4 Å resolution of the complex of transducin βγ and its regulator, phosducin. *Cell,* **87**, 577-588.

Gerstein, M., Tsai, J. & Levitt, M. (1995). The volume of atoms on the protein surface calculated from simulation using *Voronoi polyhedra*. *J. Mol. Biol.* **249**, 955-966.

Goldman, E. R., Dall'Acqua, W., Braden, B. C. & Mariuzza, R. A. (1997). Analysis of binding interactions in an idiotope-antiidiotope protein-protein complex by double mutant cycles. *Biochemistry,* **36**, 49-56.

Gorina, S. & Pavletich, N. P. (1996). Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science,* **274**, 1001-1005.

Griffith, J. P., Kim, J. L., Kim, E. E., Sintchak, M. D., Thomson, J. A., Fitzgibbon, M. J., Fleming, M. A., Caron, P. R., Hsiao, K. & Navia, M. A. (1995). X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex. *Cell,* **82**, 507-522.

Guillet, V., Lapthorn, A., Hartley, R. W. & Mauguen, Y. (1993). Recognition between a bacterial ribonuclease, barnase, and its natural inhibitor, barstar. *Structure,* **1**, 165-177.

Harel, M., Kleywegt, G. J., Ravelli, R. B. G., Silman, I. & Sussman, J. L. (1995). Crystal structure of an acetylcholinesterase-fasciclin complex: interaction of a three-fingered toxin from snake venom with its target. *Structure,* **3**, 1355-1366.

Harrison, C. J., Hayer-Hartl, M., Di Liberto, M., Hartl, F. U. & Kuriyan, J. (1997). Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. *Science,* **276**, 431-435.

Harpaz, Y., Gerstein, M. & Chothia, C. (1994). Volume changes on protein folding. *Structure,* **2**, 641-649.

Housset, D., Mazza, G., Grégoire, C., Piras, C., Malissen, B. & Fontecilla-Camps, J. C. (1997). The three-dimensional structure of a T-cell antigen receptor Vα-Vβ heterodimer reveals a novel rearrangement of the Vβ domain. *EMBO J.* **16**, 4205-4216.

Huang, Q., Liu, S. & Tang, Y. (1993). The refined 1.6 Å resolution crystal structure of the complex formed

between porcine β-trypsin and MCTI-A, a trypsin inhibitor of the squash family. *J. Mol. Biol.* **229**, 1022-1036.

Huang, K., Lu, W., Anderson, S., Laskowski, M., Jr & James, M. N. G. (1995). Water molecules participate in protease-inhibitor interactions: crystal structures of Leu-18, Ala-18 and Gly-18 variants of turkey ovomucoid inhibitor third domain in complex with *S. griseus* proteinase B. *Protein Sci,* **4**, 1985-1997.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. & Steigemann, W. (1974). Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II Crystallographic refinement at 1.9 Å resolution. *J. Mol. Biol.* **89**, 73-101.

Hurley, J. H., Faber, H. R., Worthylake, D., Meadow, N. D., Roseman, S., Pettigrew, D. W. & Remington, S. J. (1993). Structure of the regulatory complex of *E. coli* III^{Gel} with glycerol kinase. *Science,* **259**, 673-677.

Janin, J. (1995). Principles of protein-protein recognition from structure to thermodynamics. *Biochimie,* **77**, 497-505.

Janin, J. (1996). Protein-protein recognition. *Prog. Biophys. Mol. Biol.* **64**, 145-165.

Janin, J. & Chothia, C. (1976). Stability and specificity of protein-protein interactions: the case of the trypsin-trypsin inhibitor complexes. *J. Mol. Biol.* **100**, 197-211.

Janin, J. & Chothia, C. (1990). The structure of protein-protein recognition sites. *J. Biol. Chem.* **265**, 16027-16030.

Janin, J., Miller, S. & Chothia, C. (1988). Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.* **204**, 155-164.

Jardetzky, T. S., Brown, J. H., Gorga, J. C., Stern, L., J., Urban, R. G., Chi, Y.-I., Stauffacher, C., Strominger, J., L. & Wiley, D. C. (1994). Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen. *Nature,* **368**, 711-718.

Jeffrey, P. D., Russo, A. A., Polyak, K., Gibbs, E., Hurwitz, J., Massagué, J. & Pavletich, (1995). Mechanism of CDK activation revealed by the structure of a cyclin A-CDK2 complex. *Nature,* **376**, 313-320.

Jones, S. & Thornton, J. M. (1995). Protein-protein interaction: a review of protein dimer structures. *Prog. Biophys. Mol. Biol.* **63**, 131-165.

Jones, S. & Thornton, J. M. (1996). Principles of protein-protein interactions. *Proc. Natl Acad. Sci. USA,* **93**, 13-20.

Jones, S. & Thornton, J. M. (1997). Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol.* **272**, 121-132.

Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. (1990). Atomic structure of the actin:DNAse I complex. *Nature,* **347**, 37-44.

Kawashima, T., Berthet-Colominas, C., Wulff, M., Cusack, S. & Leberman, R. (1996). The structure of the *E. coli* EF-Tu. EF-Ts complex at 2.5 Å resolution. *Nature,* **379**, 511-518.

Kelley, R. F., Costas, K. E., O'Connell, M. P. & Lazarus, R. A. (1995). Analysis of the factor VIIa binding site on human tissue factor: effects of tissue factor mutations on the kinetics and thermodynamics of binding. *Biochemistry,* **34**, 10383-10392.

Kobe, B. & Deisenhofer, J. (1995). A structural basis of the interactions between leucine rich repeats and protein ligands. *Nature,* **374**, 183-186.

Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. (1994). Structural determinants for activation of the α-subunit of a heterotrimeric G protein. *Nature,* **369**, 621-628.

Lambright, D. G., Sondek, J., Bohm, A., Skiba, N. P., Hamm, H. E. & Sigler, P. B. (1996). The 2.0 Å crystal structure of a heterotrimeric G-protein. *Nature,* **379**, 311-319.

Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946-950.

Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.

Lee, C.-H., Saksela, K., Mirza, U. A., Chait, B. T. & Kuriyan, J. (1996). Crystal structure of the conserved coreof HIV-1 nef complexed with a Src family SH3 domain. *Cell,* **85**, 831-942.

Lescar, J., Pellegrini, M., Souchon, H., Tello, D., Poljak, R. J., Peterson, N., Greene, M. & Alzari, P. M. (1995). Crystal Structure of a cross reaction complex between Fab F9.13.7 and guinea fowl lysozyme. *J. Biol. Chem.* **270**, 18067-18076.

Li, H., Dunn, J. J., Luft, B. J. & Lawson, C. L. (1997). Crystal structure of Lyme disease antigen OspA complexed with a Fab. *Proc. Natl Acad. Sci. USA,* **94**, 3584-3589.

Livnah, O., Stura, E. A., Johnson, D. L., Middleton, S. A., Mulcahy, L. S., Wrighton, N. C., Dower, W. J., Jolliffe, L. K. & Wilson, I. A. (1996). Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor at 2.8 Å resolution. *Science,* **273**, 464-471.

Malby, R. L., Tulip, W. R., Harley, V. R., McKimm-Breschkin, J. L., Laver, W. G., Webster, R. G. & Colman, P. M. (1994). The structure of a complex between the NC10 antibody and influenza virus neuraminidase and comparison with the overlapping site of the NC41 antibody. *Structure,* **2**, 733-746.

Malkowski, M. G., Martin, P. D., Guzik, G. C. & Edwards, B. F. (1997). The co-crystal structure of unliganded bovine α-thrombin and prethrombin-2:movement of the Tyr-Pro-Pro-Trp segment upon ligand binding. *Protein Sci.* **6**, 1438-1448.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallog. sect. B,* **39**, 480-490.

McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777-793.

McPhalen, C. A. & James, M. N. G. (1987). Structural comparison of two serine proteinase-protein inhibitor complexes. Eglin C-subtilisin Carlsberg and CI 2−subtilisin novo. *Biochemistry,* **27**, 6582-6598.

Miller, S., Janin, J., Lesk, A. M. & Chothia, C. (1987a). Interior and surface of monomeric proteins. *J. Mol. Biol.* **196**, 641-656.

Miller, S., Lesk, A. M., Janin, J. & Chothia, C. (1987b). The accessible surface area and stability of oligomeric proteins. *Nature,* **328**, 834-836.

Mittl, P. R., Di Marco, S., Fendrich, G., Pohlig, G., Heim, J., Sommerhoff, C., Fritz, H., Priestle, J. P. & Grutter, M. G. (1997). A new structural class of serine protease inhibitors revealed by the structure of the hirustatin-kallikrein complex. *Structure,* **5**, 253-264.

Nassar, N., Horn, G., Herrmann, C., Scherer, A., McCormick, F. & Wittinghofer, A. (1996). The 2.2 Å crystal structure of the ras-binding domain of the serine/threonine kinase c-Raf1 in a complex with Rap1A and a GTP analogue. *Nature,* **375**, 554-660.

Nicholls, A., Sharp, K. & Honig, B. (1992). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-296.

Noel, J. P., Hamm, H. E. & Sigler, P. B. (1993). The 2.2 Å crystal structure of transducin complexes with GTPgS. *Nature,* **366**, 654-662.

Padlan, E. A., Silverton, E. W., Sheriff, S., Cohen, G. H., Smith-Gill, S. J. & Davis, D. R. (1989). Structure of an antibody-antigen complex: crystal structure of the HyHEL-10 Fab-lysozyme Complex. *Proc. Natl Acad. Sci. USA,* **86**, 5938-5952.

Pelletier, H. & Kraut, J. (1992). Crystal structure of a complex between electron transfer partners, cytochrome c peroxidase and cytochrome *c*. *Science,* **258**, 1748-1755.

Prasad, L., Sharma, S., Vandonselaar, M., Quail, J. W., Lee, J. S., Waygood, E. B., Wilson, K. S., Dauter, Z. & Delbaere, L. T. J. (1993). Evaluation of mutagenesis for epitope mapping: structure of an antibody-protein antigen complex. *J. Biol. Chem.* **268**, 10705-10708.

Read, R. J., Fujinaga, M., Sielecki, R. & James, M. N. G. (1983). Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. *Biochemistry,* **22**, 4420-4433.

Rees, D. C. & Lipscomb, W. N. (1982). Refined crystal structure of potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *J. Mol. Biol.* **160**, 475-498.

Richards, F. M. (1974). The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* **82**, 1-14.

Rittinger, K., Walker, P. A., Eccleston, J. F., Smerdon, S. J. & Gamblin, S. J. (1997). Structure at 1.65 Å of RhoA and its GTPase -activating protein in complex with a transition-state analogue. *Nature,* **389**, 758-762.

Rydel, T. J., Tulinsky, A., Bode, W. & Huber, R. (1991). The refined crystal structure of the hirudin-thrombin complex. *J. Mol. Biol.* **221**, 583-601.

Savva, R. & Pearl, L. H. (1995). Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase-uracil glycoslase inhibitor protein complex. *Nature Struct. Biol.* **2**, 752-757.

Scheidig, A. J., Hynes, T. R., Pelleyier, L. A., Wells, J. A. & Kossiakoff, A. A. (1997). Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid b-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI). *Protein Sci.* **6**, 1806-1824.

Schreiber, G. & Fersht, A. R. (1993). Interaction of barnase with its polypeptide inhibitor barstat studied by protein engineering. *Biochemistry,* **32**, 5145-5150.

Schreiber, G. & Fersht, A. R. (1995). Energetics of protein-protein interactions: analysis of the Barnase-Barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* **248**, 478-486.

Schutt, C. E., Myslik, J. C., Rozycki, M. D., Goonesekere, N. C. W. & Lindberg, U. (1993). The structure of crystalline profilin-β-actin. *Nature,* **365**, 810-816.

Sheriff, S., Silverton, 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. Natl Acad. Sci. USA,* **84**, 8075-8079.

Sondek, J., Bohm, A., Lambright, D. G., Hamm, H. E. & Sigler, P. B. (1996). Crystal structure of a G_A protein $\beta\gamma$ dimer at 2. 1 Å resolution. *Nature,* **379**, 369-374.

Song, H. K. & Suh, S. W. (1998). Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight of the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *J. Mol. Biol.* **275**, 347-363.

Stewart, M., Kent, H. M. & McCoy, A. J. (1998). Structural basis for molecular recognition between nuclear transport factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase Ran. *J. Mol. Biol.* **277**, 635-646.

Strynadka, N. C. J., Jensen, S. E., Alzari, P. M. & James, M. N. G. (1996). A potent new mode of β-lactamase inhibition revealed by the 1.7 Å X-ray crystallographic structute of the TEM-1–BLIP complex. *Nature Struct. Biol.* **3**, 290-297.

Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. & Turk, V. (1990). The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinease papain: a novel type of proteinase inhibitor interaction. *EMBO J.* **9**, 1939-1947.

Sundström, M., Lundqvist, T., Rödin, J., Giebel, L. B., Milligan, D. & Norstedt, G. (1996). Crystal structure of an antagonist mutant of human growth hormon G120R in complex with its receptor at 2.9 Å resolution. *J. Biol. Chem.* **271**, 32197-32203.

Takeuchi, Y., Satow, Y., Nakamura, K. T. & Mitsui, Y. (1991). Refined crystal structure of the complex subtilisin BPN and *Streptomyces* subtilisin inhibitor at 1.8 Å resolution. *J. Mol. Biol.* **221**, 309-325.

Tesmer, J. J. G., Berman, D. M., Gilman, A. G. & Sprang, S. R. (1997). Structure of RGS4 bound to $AlF4^-$ activated $G_{i\alpha1}$: stabilisation of the transition state of GTP hydrolysis. *Cell,* **89**, 251-261.

Tsai, C. J. & Nussinov, R. (1997). Hydrophobic folding units at protein-protein interfaces: implications to protein folding and to protein-protein association. *Protein Sci.* **6**, 1426-1437.

Tsai, J., Lin, S. L., Wolfson, H. & Nussinov, R. (1996). A dataset of protein-protein interfaces generated with a sequence order-independent comparison technique. *J. Mol. Biol.,* **260**, 604-620.

Tsunemi, M., Matsuura, Y., Sakakibara, S. & Katsube, Y. (1996). Crystal structure of an elastase-specific inhibitor elafin complexed with porcine pancreatic elastase. *Biochemistry,* **35**, 11570-11576.

Tulip, W. R., Varghese, J. N., Laver, W. G., Webster, R. G. & Colman, P. M. (1992). Refined crystal structure of the influenza virus N9 neuraminidase-NC41 Fab complex. *J. Mol. Biol.* **227**, 122-148.

van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kröger, B., Höffken, W. & Bode, W. (1995). Two heads are better than one: crystal structure of the insect derived double domain Kazal inhibitor rhodnin in complex with thrombin. *EMBO J.* **14**, 5149-5157.

van de Locht, A., Stubbs, M. T., Bode, W., Friedrich, T., Bollschweiler, C., Hoffken, W. & Huber, R. (1996). The ornithodorin-thrombin crystal structure, a key to the TAP enigma? *EMBO J.* **15**, 6011-6017.

van de Locht, A., Bode, W., Huber, R., Le, Bonniec B. F., Stone, S. R., Esmon, C. T. & Stubbs, M. T. (1997). The thrombin E192Q-BPTI complex reveals gross

structural rearrangements: implications for the interaction with antithrombin and thrombomodulin. *EMBO J.* **16**, 2977-2984.

Vos, A. M., de Ultsch, M. & Kossiakoff, A. A. (1992). Human growth hormone and extracellular domain of its receptor: crystal structure of thr complex. *Science,* **255**, 306-312.

Wall, M. A., Coleman, D. E., Lee, E., Iniguez-Lluhi, J. A., Posner, B. A., Gilman, A. G. & Sprang, S. R. (1995). The structure of the G protein heterotrimer $G_{i\alpha1\beta1\gamma2}$. *Cell,* **83**, 1047-1058.

Wang, Y., Jiang, Y., Meyering-Voss, M., Sprinzl, M. & Sigler, P. B. (1997). Crystal structure of the EF-Tu. EF-Ts complex from *Thermus thermophilus*. *Nature Struct. Biol.* **4**, 650-656.

Wang, J. H., Lim, K., Smolyar, A., Teng, M. K., Liu, J. H., Tse, A. G. D., Liu, J., Hussey, R. E., Chishti, Y., Thomson, C. T., Sweet, R. M., Nathenson, S. G., Chang, H. C., Sacchenetti, J. C. & Reinherz, E. L. (1998). Atomic structure of an αβ T-cell receptor (TCR) heterodimer in complex with an anti-TCR Fab fragment derived from a mitogenic antibody. *EMBO J.* **17**, 10-26.

Welch, M., Chinardet, N., Mourey, L., Birck, C. & Samama, J. P. (1997). Structure of the CheY-binding domain of histidine kinase CheA in complex with CheY. *Nature Struct. Biol,* **5**, 25-29.

Wells, J. A. (1996). Binding in the growth hormone receptor complex. *Proc. Natl Acad. Sci. USA,* **93**, 1-6.

Xu, D., Tsai, C. J. & Nussinov, R. (1997). Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng.* **10**, 999-1012.

Young, L., Jernigan, R. L. & Covell, D. G. (1994). A role for surface hydrophobicity in protein-protein recognition. *Protein Sci.* **3**, 717-729.

Ysern, X., Li, H. & Mariuzza, R. A. (1998). Imperfect interfaces. *Nature Struct. Biol.* **5**, 412-414.

Edited by A. R. Fersht

(Received 1 October 1998; received in revised form 30 November 1998; accepted 30 November 1998)