Cavities and packing at protein interfaces

SIMON J. HUBBARD AND PATRICK ARGOS

European Molecular Biology Laboratory, Postfach 10.2209, Meyerhofstrasse 1, 69012 Heidelberg, Germany

(RECEIVED July 25, 1994; ACCEPTED September 16, 1994)

Abstract

An analysis of internal packing defects or “cavities” (both empty and water-containing) within protein structures has been undertaken and includes 3 cavity classes: within domains, between domains, and between protein subunits. We confirm several basic features common to all cavity types but also find a number of new characteristics, including those that distinguish the classes. The total cavity volume remains only a small fraction of the total protein volume and yet increases with protein size. Water-filled “cavities” possess a more polar surface and are typically larger. Their constituent waters are necessary to satisfy the local hydrogen bonding potential. Cavity-surrounding atoms are observed to be, on average, less flexible than their environments. Intersubunit and interdomain cavities are on average larger than the intradomain cavities, occupy a larger fraction of their resident surfaces, and are more frequently water-filled. We observe increased cavity volume at domain–domain interfaces involved with shear type domain motions. The significance of interfacial cavities upon subunit and domain shape complementarity and the protein docking problem, as well as in their structural and functional role in oligomeric proteins, will be discussed. The results concerning cavity size, polarity, solvation, general abundance, and residue type constituency should provide useful guidelines for protein modeling and design.

Keywords: cavities; complementarity; docking; interfaces; packing

The interfaces formed between protein domains and subunits, as detailed by their tertiary and quaternary structures, have been investigated by many workers and considerable detail is currently known about their general properties (Chothia & Janin, 1975; Argos, 1988; Janin et al., 1988). Subunit interfaces are complementary in shape and well packed (Chothia & Janin, 1975), albeit with some variation in quality (Tulip et al., 1992; Walls & Sternberg, 1992; Lawrence & Colman, 1993), and also show a high degree of hydrophobic complementarity (Korn & Burnett, 1991; Young et al., 1994). Nonetheless, the packing quality at domain–domain interfaces has not been rigorously studied. An excellent and recent review of packing and its significance stresses its importance for the protein folding problem (Richards & Lim, 1994). The dense atomic packing observed within proteins, comparable with that found in small organic crystals (Richards, 1974, 1977), has also been used in many forms for structure prediction and modeling, including the inverse folding problem (e.g., Ponder & Richards, 1987), the prediction of side-chain conformation (e.g., Holm & Sander, 1991; Lee & Subbiah, 1991; Eisenmenger et al., 1993), and as a criterion to test models (e.g., Novotny et al., 1988).

Interfacial shape complementarity between protein subunits is similarly important to the protein docking problem, which

Connolly (1986) proposed to solve on a geometrical basis alone. Considerable further efforts have relied on a high match of shape to obtain good fitting solutions (Zielenkiewicz & Rabcenko, 1984; Connolly, 1986; Wodak et al., 1987; Jiang & Kim, 1991; Shiochet & Kuntz, 1991, 1993; Bacon & Moulton, 1992; Walls & Sternberg, 1992; Helmer-Citterich & Tramontano, 1994; Norel et al., 1994). Despite this profusion of methods, shape complementarity has only been quantified for a limited number of protein interfaces, using atomic packing density (Tulip et al., 1992; Walls & Sternberg, 1992) and a statistical measure (Lawrence & Colman, 1993). A comprehensive survey over a larger data set is required for a more complete picture. We provide an alternative method to estimate interfacial fit by calculating the total internal cavity surface found at interfaces and express it as a fraction of the total interface surface, resulting in a quantitative estimate of surface complementarity. By “cavities,” we refer to a “hole” or packing defect in a protein tertiary structure entirely surrounded by protein atoms, such that a probe of given size can be accommodated but cannot find a path to the external surface. The “cavity” may therefore contain water, which we refer to as a water-filled or solvated cavity, or be void of any other atoms, which we refer to as an “empty” cavity. Atoms contributing to the surface of cavities are often referred to as “cavity atoms.”

Packing defects or cavities within monomeric protein subunits have been well characterized (Connolly, 1985a; Rashin et al., 1986; Hubbard et al., 1994) and a number of general proper-

Reprint requests to: Simon J. Hubbard, European Molecular Biology Laboratory, Postfach 10.2209, Meyerhofstrasse 1, 69012 Heidelberg, Germany; e-mail: hubbard@embl-heidelberg.de.

ties have been elucidated. Though our previous work (Hubbard et al., 1994) considered cavities within single protein subunits, no distinction was made between those lying within and between protein domains. It has already been postulated that interdomain cavities may play a functional role by assisting domain movements (Rashin et al., 1986); however, the general properties of such cavities in a large data set have yet to be investigated.

In the present work, we have studied all cavities within 2 data sets of protein tertiary structures taken from the Brookhaven Protein Data Bank (Bernstein et al., 1977): multidomain and multisubunit proteins. Cavities were calculated using the algorithm of Connolly (1983, 1985b, 1993) and were divided into 3 distinct classes: intradomain, interdomain, and intersubunit. Their general properties have been compared and contrasted. We conclude that the 3 cavity classes exhibit many different properties and yet often follow the rules already detailed for cavities found within monomeric units (Hubbard et al., 1994). Intersubunit and interdomain cavities had constituted a significant fraction of the interfacial surfaces and were more prevalent than expected, especially given their weaker presence within domains and the generally accepted tenet that interfaces are tightly packed (Chothia & Janin, 1975). Cavities were further classified, where possible, as either solvated or "empty" and their properties compared. The implications of these results are discussed with reference to protein folding, packing, and docking, as well as to the functional roles these interfacial cavities may have.

Results and discussion

General packing properties of surfaces

Table 1A and B lists the total cavity volume, interface surface area, and fractional cavity interfacial surface area for each protein in the multidomain and multisubunit data sets, respectively. Only 10% of the multidomain proteins studied were without detectable cavities at their domain interfaces, and these proteins

were all less than 175 residues in size. Furthermore, 106 of the 356 cavities delineated in multidomain proteins were interdomain, which represents 38.1% of the total cavity volume in the data set (Table 1A). Hence, most interdomain protein interfaces are imperfectly matched, as is also the case for the subunit interfaces (Table 1B), where 77% of the oligomers have at least 1 cavity at their subunit interfaces and 30% of all cavity volume is intersubunit.

The fractional cavity interface surface for both data sets is plotted in Figure 1. As a comparison, a similar value is calculated for the intradomain cavities (those that do not lie at domain–domain boundaries in the domain data set) using the entire surface buried upon folding of the single domains (thus excluding surface buried between domains) as the reference surface. Cavities found exclusively within domains constitute a much smaller fraction of their reference surface (mean 2.8%) than interdomain and intersubunit cavities, which form on average 10% of the interdomain and intersubunit surfaces, respectively. Seven of 52 intersubunit surfaces have more than 20% of their interface surface contributed by cavity surface, with the largest value being 36.4% for southern bean mosaic virus (SBMV) and the next largest 24.6%. The SBMV result may be in part due to poorly defined metal ion density at the quasi-3-fold axis (Silva & Rossmann, 1985). Nevertheless, these examples represent a significant deviation from optimal shape complementarity, given that the interfacial cavity regions preclude good van der Waals contacts. Figure 2 shows such an example, the Trp repressor protein dimer (2wrp), with an intermolecular cavity of 143 Å³.

Most of the interdomain and intersubunit cavity volume is solvent filled. Within the high-resolution protein subsets reporting at least 2 waters per residue, 77% of the interdomain cavity volume is contributed by solvent-filled cavities; the corresponding number for the intersubunit cavities is 63%. Empty cavities constitute on average only 2.3% and 3.0% of the domain and subunit interface surfaces, respectively. This demonstrates the important role water plays at protein–protein boundaries.

Fig. 1. Distribution histograms of the total cavity surface of each protein expressed as a percentage of the resident surface for the relevant data set. All surfaces used were molecular surfaces. **A:** Intradomain cavity surface as a percentage of the buried surface. **B:** Interdomain cavity surface as a percentage of the interdomain interfacial surface. **C:** Intersubunit cavity surface as a percentage of the intersubunit interfacial surface.

Table 1. Multidomain (A) and multisubunit (B) data sets, interface and interface cavity surface

Protein	PDB code ^a	Number of domains/ subunits ^b	Total number of residues	Interface area (Å ²)	Interfacial cavity area (Å ²)	Cavity area as percentage of interfacial area
A. Multidomain data set						
Ferredoxin	1fdx	2	54	1,346	0.0	0.0
Calbindin	4icb ^c	2	76	1,094	43.4	4.0
Parvalbumin	4cpv ^c	3	108	2,359	198.0	8.4
Lysozyme	6lyz ^c	2	129	1,055	154.6	14.6
Aspartate carbamoyltransferase	2atc	2	139	318	0.0	0.0
Lysozyme	2lzm ^c	2	162	991	88.2	8.9
Glutathione peroxidase	1gpl ^c	2	168	3,684	214.4	5.8
Wheat germ agglutinin	9wga ^c	4	171	871	0.0	0.0
γ-B crystallin	4gcr ^c	2	174	790	24.7	3.1
Serine proteinase	2sga ^c	2	181	2,165	140.4	6.5
α-Lytic protease	2alp ^c	2	198	2,616	345.9	13.2
Papain	9pap ^c	2	212	1,570	344.0	21.9
α Chymotrypsin	6cha ^c	2	228	2,983	451.9	15.2
Elastase	3est ^c (u)	2	240	2,736	230.9	8.4
Rhodanese	1rhd	2	266	2,723	237.4	8.7
Subtilisin	1sbt	3	275	4,467	400.9	9.0
Tomato bushy stunt virus	2tbv (h)	2	287	210	22.0	10.5
Cytochrome c peroxidase	2cyp ^c	2	292	1,896	257.5	13.6
Arabinose binding protein	5abp ^c (h)	2	305	1,080	66.3	6.1
Thermolysin	3tln ^c (u)	2	316	2,045	118.8	5.8
Penicillopepsin	3app ^c (s)	2	323	1,773	176.8	10.0
Rhizopuspepsin	2apr ^c (s)	2	325	2,024	344.1	17.0
Pepsin	4pep ^c (s)	2	326	1,843	238.9	13.0
Lactate dehydrogenase	6ldh ^c	2	330	2,428	161.9	6.7
Glyceraldehyde 3P dehydrogenase	3gpd (s)	2	334	2,233	419.4	18.8
Alcohol dehydrogenase	8adh (s)	2	374	1,899	147.8	7.8
Phosphoglycerate kinase	3pgk (s)	2	415	562	72.7	12.9
Citrate synthase	5cts (s)	2	429	2,117	309.0	14.6
Glutathione reductase	3grs ^c	3	433	4,584	168.8	3.7
Catalase	8cat	4	498	4,633	591.6	12.8
B. Multisubunit data set						
Avian pancreatic polypeptide	1ppt	2	72	812	157.3	19.4
Insulin	3ins ^c	4	102	2,579	202.4	7.8
Mellitin	2mlt ^c	4	104	4,213	0.0	0.0
Uteroglobin	2utg ^c	2	140	2,220	80.5	3.6
Gene 5 binding protein	2gn5	2	174	507	0.0	0.0
HIV proteinase	5hvp ^c	2	198	2,409	0.0	0.0
Trp repressor	2wrp ^c	2	208	2,759	330.4	12.0
Immunoglobulin	1rei ^c	2	214	794	191.2	24.1
Subtilisin inhibitor	2ssi	2	214	905	195.2	21.6
Immunoglobulin	2rhe ^c	2	228	996	93.7	9.4
Rous sarcoma virus protease	2rsp ^c	2	228	1,966	0.0	0.0
Phospholipase	1pp2	2	244	1,844	0.0	0.0
Cytochrome c'	2ccy ^c	2	254	983	0.0	0.0
Hemoglobin	2sdh	2	292	940	0.0	0.0
Superoxide dismutase	2sod	2	304	1,054	0.0	0.0
Wheat germ agglutinin	9wga ^c	2	342	586	0.0	0.0
Glutathione peroxidase	1gpl ^c	2	370	1,388	284.4	20.5
Gene activator protein	3gap	2	413	1,722	81.1	4.7
Immunoglobulin	3mcg ^c	2	432	2,515	203.6	8.1
Immunoglobulin	2fbj ^c	2	433	2,092	23.1	1.1
Immunoglobulin	4fab	2	435	1,870	274.3	14.7
Southern bean mosaic virus	4sbv	3	620	4,804	1,749.8	36.4
Tumor necrosis factor	1tnf	3	456	4,133	1,138.1	24.6
Prealbumin	2pab	4	456	4,133	781.6	18.9
Concanavalin A	1cn1	2	474	1,140	78.1	6.9
α Chymotrypsin	5cha ^c	2	474	732	0.0	0.0
Triose phosphate isomerase	6tim	2	498	2,198	356.4	16.2
Thymidylate synthase	2tsc ^c	2	528	3,058	383.0	12.5

(continued)

Table 1. Continued

Protein	PDB code ^a	Number of domains/subunits ^b	Total number of residues	Interface area (Å ²)	Interfacial cavity area (Å ²)	Cavity area as percentage of interfacial area
B. Multisubunit data set (continued)						
Inorganic pyrophosphatase	1pyp	2	562	783	0.0	0.0
Hemoglobin	4hhb ^c	4	574	3,553	300.3	8.5
Tyrosyl transfer RNA synthetase	4ts1	2	628	2,263	168.3	7.4
Acetyltransferase	3cla ^c	3	639	3,750	301.9	8.1
Malate dehydrogenase	4mdh	2	668	3,379	104.7	3.1
Alcohol dehydrogenase	6adh	2	748	1,873	65.3	3.5
Rhinovirus	4rhv	3	764	9,004	845.7	9.4
Aspartate aminotransferase	3aat	2	792	4,474	746.7	16.7
Isocitrate dehydrogenase	9icd	2	828	4,273	675.6	15.8
Enolase	7enl	2	872	3,792	571.8	15.1
Citrate synthase	4cts	2	874	6,922	1,311.1	18.9
Rubisco	2rus	2	881	5,310	1,038.6	19.6
Tomato bushy stunt virus	2tbv	3	895	2,047	347.3	17.0
Aspartate carbamoyltransferase	8atc	4	912	2,523	284.4	11.3
Phosphoglycerate mutase	3pgm	4	920	2,855	0.0	0.0
Glutathione reductase	3grs ^c	2	922	4,382	493.4	11.3
Bean pod mottle virus	1bmV	3	1,122	3,490	649.1	18.6
Phosphofructokinase	1pfk	4	1,280	7,082	1,577.6	22.3
Lactate dehydrogenase	1llc	4	1,284	6,965	973.4	14.0
Lactate dehydrogenase	8ldh	4	1,320	8,156	1,793.6	22.0
Glyceraldehyde 3P dehydrogenase	4gpd	4	1,332	8,922	1,686.9	18.9
Glyceraldehyde 3P dehydrogenase	2gd1	4	1,336	11,492	2,061.3	17.9
Hemagglutinin	2hmg	3	1,509	6,854	601.9	8.8
Xylose isomerase	5xia	4	1,572	16,893	863.5	5.1

^a Identification code of protein in the Brookhaven Protein Data Bank (Bernstein et al., 1977) of deposited coordinates. Proteins followed by letters in parentheses are in common with the domain-motion classified proteins of Gerstein et al. (1994a), where (s) is a shear motion, (h) a hinge motion, and (u) unclassifiable, but not predominantly shear or hinge.

^b For definitions of domains, see Argos (1988). Number of subunits for the multisubunit data set.

^c Proteins included in the high-resolution, water-dense subsets, with 2.0 Å or better resolution and at least 1 water per 2 residues.

Fig. 2. Example intersubunit cavity. A slice through the Trp repressor dimer (2wrp) is shown, with atoms depicted as filled circles with appropriate van der Waals radius, and dark and light shading used to distinguish the 2 subunits. The accessible surface traced by the probe center is shown as a dotted line bordering a light gray surface around the atoms. The intersubunit cavity, as indicated, is unshaded and includes the path traced by the center of the defining probe (dotted line).

Interdomain cavities have been linked with function by facilitating domain-domain motions (Rashin et al., 1986). The motion between domains has recently been reviewed and classified into 2 predominant types: shear and hinge motions (Gerstein et al., 1994a). The former involves a translation motion of the 2 interfaces parallel to each other and consequently is constrained by the close packing of the interface residues. Hinge or rotational motions would not be expected to be so constrained, and, consequently, cavities would be expected in greater numbers at domain interfaces involved in shear type motions. Eleven of our multidomain proteins are found in the domain-motion proteins listed by Gerstein et al. (1994a): 7 of these exhibit shear type motions and have on average 13.5% of their interfacial surfaces occupied by cavity surface, which is in excess of the 7.7% mean for the remaining 4 proteins. The corresponding mean value for all 30 multidomain proteins, excluding the 7, is 8.6%. This result suggests that cavities play a role in such shear domain movements by allowing the relaxation of packing constraints.

To test if proteins make compensatory interactions for the loss of van der Waals contact at their interfaces due to cavities, we considered the buried area, hydrophobicity, and hydrogen bond-

ing at the interfaces. Interface area per subunit or domain varies considerably from protein to protein and there is no simple correlation with domain or monomeric molecular weight, or the total accessible surface (Argos, 1988; Janin et al., 1988). Similarly, we find no correlation with interfacial complementarity. Previous results (Janin et al., 1988) have found there to be approximately 1 hydrogen bond per 200 Å² of accessible interface surface within oligomeric proteins, although we find closer to 1 per 100 Å² using the molecular surface. This latter value is quite well maintained over domain and subunit interfaces and shows no compensatory increase for loosely packed interfaces. However, the hydrophobic surface fraction of the interfaces and the number of hydrogen bonds per Å² show an inverse correlation of -0.76 , suggesting that when one is poor the other is strong, but do not correlate with shape complementarity.

Interfacial cavity size

The total cavity volume within monomeric proteins was found to be a crude function of the protein size (Hubbard et al., 1994). It is plotted here for the different cavity types in the 2 current data sets against protein size (number of residues) in Figure 3. The intradomain and intrasubunit cavity volume forms the bulk

of the total protein cavity volume. The total cavity volume summed over all cavity types for all the proteins in both data sets is also plotted against protein size in Figure 3D. For each plot a least-squares linear regression line is shown. Though the total cavity volume roughly increases with protein size in all cases, the best fit to a straight line is obtained in the all cavity case. This suggests that the total cavity volume in the complete and functional protein is best maintained, constituting on average about 2 Å³/residue, whereas the cavity distribution in various regions is more flexible. The ratio of solvated to "empty" cavity volume increases from intradomain to interdomain to intersubunit, the ratios being 1.25, 1.83, and 2.23, respectively.

Histograms of cavity volume distributions are shown in Figure 4 for solvated and "empty" cavities over the 3 cavity types: intradomain, interdomain, and intersubunit. We present here data derived from the full protein data sets in Table 1 because the high-resolution subsets displayed similar trends. Consistent with previous results (Rashin et al., 1986; Hubbard et al., 1994), cavity solvation supports larger cavities. Additionally, the mean cavity size increases from intradomain to interdomain to intersubunit surfaces: mean volumes for solvated cavities are, respectively, 39.8, 43.8, and 86.5 Å³ (39.4, 43.9, and 60.4 Å³ for the high-resolution subsets) and for "empty" cavities 18.5, 27.1, and

Fig. 3. Total cavity volume plotted against the number of residues in the multidomain or multisubunit proteins for (A) intradomain cavities in the multidomain data set, (B) interdomain cavities in the multidomain data set, (C) intersubunit cavities in the oligomeric data set, and (D) all cavities from all proteins.

Fig. 4. Distributions of cavity volumes for (A) intradomain solvated, (B) interdomain solvated, (C) intersubunit solvated, (D) intradomain empty, (E) interdomain empty, and (F) intersubunit empty cavities.

28.1 Å³ (17.8, 21.1, and 37.3 Å³ for the high-resolution subsets), respectively. Very few “empty” cavities in monomeric proteins exceed 50 Å³ in volume (Hubbard et al., 1994).

Cavity surface constitution

Histograms of the polarities of the different cavity types are shown in Figure 5. Solvated cavity surface is consistently more polar than the empty cavity surface, where the mean polar fraction of the respective surfaces are 0.47, 0.48, and 0.41 for intradomain, interdomain, and intersubunit cavities (0.48, 0.47, 0.44 for high-resolution subsets). The equivalent mean polarities for the empty cavities are 0.14, 0.20, and 0.26 (0.15, 0.15, 0.22 for high-resolution subsets). The reason for this increasing trend in polarity for the “empty” cavities is unclear, although they remain largely hydrophobic. These surfaces may be required to be more polar in character because they are more often exposed to solvent or may be involved in ligand binding and even catalysis.

Table 2 lists the percentage contributions to the various surfaces from main- and side-chain atoms, polar and nonpolar atoms, and the 3 classical secondary structural types, over the high-resolution, water-dense subsets. A number of clear trends are evident. As noted by Chothia (1976), the accessible surface is dominated by side chains, and the buried surface more so by main chain. The surface of solvated cavities within domains and at domain/subunit interfaces has a high main-chain component, as does the buried surface, and is consistently more polar than the surface of “empty” cavities, which are dominated by non polar side-chain groups. Because the 2 interface types for domains and subunits possess similar distributions of main/side-chain

and polar/nonpolar surfaces, it is likely that the cavity type is the major determinant of its properties rather than the interface in which it lies. The secondary structural preferences were similar to those previously described (Hubbard et al., 1994), where solvated cavities used less surface from regular secondary structures and more from coil residues, whereas “empty” cavities had a higher secondary structural contribution, despite the broad variation in the secondary structural constitution at protein interfaces generally (Argos, 1988; Miller, 1989).

Amino acid preferences for cavity surface

The amino acid side-chain surface preferences were calculated as described in the Materials and methods section and are plotted in Figure 6 for 10 different surface types. The classic propensities for accessible and buried surfaces are observed (Chothia, 1976) with hydrophilic residues prevalent at the accessible surface and hydrophobic ones in the buried core. A number of common residue types are preferred at “empty” cavity surfaces: Phe, Leu, Met, and Ala. In general, the bulky aromatic and aliphatic residues are favored at “empty” cavity surfaces, with intra- and interdomain “empty” cavities sharing the same top 7 preferred residues. Proline is the only residue type favored positively for all 3 types of solvated cavity surfaces, although several polar residues are positively favored for two (e.g., Thr, Glu, and Ser). The relative inflexibility of the proline ϕ torsion angle restricting the local main chain may lead to poorer local packing and therefore cavities.

In accordance with Argos (1988), the aromatic residues (Phe, His, Trp, and Tyr) exhibit strong preferences for the domain and subunit interfaces as well as Met. The interfacial cavities are

Fig. 5. Distributions of the polarity or fractional polar surface for (A) intradomain solvated, (B) interdomain solvated, (C) intersubunit solvated, (D) intradomain empty, (E) interdomain empty, and (F) intersubunit empty cavities.

broadly adapted from their respective interface types. For example, the "empty" interdomain cavity surface demotes the polar His, which is positively favored at domain interfaces, in favor of the strictly hydrophobic residue types. Many more polar residues become more favored at the interdomain solvated cavity surface at the expense of residues such as Leu and Met.

A number of residue types are universally disfavored at cavity surfaces, in particular the large charged side groups Arg and Lys. Indeed, despite being favored at intersubunit interface surface, Arg is disfavored at both solvated and "empty" intersub-

unit cavity surfaces. Although common at the buried surface, cysteines are disfavored at nearly all cavity surfaces and, in the majority of cavity cases, do not make disulfide bonds, presumably due to the destabilizing effect of the cavities on the bond.

Properties of cavity solvent

Statistics concerning the solvent in all 3 cavity classes (intradomain, interdomain, and intersubunit) are given in Table 3, calculated from the high-resolution, water-dense subsets. Similar

Table 2. Percent composition of protein surfaces

	Percent surface composition from						
	Main chain	Side chain	Polar	Nonpolar	Helix ^a	Strand ^a	Coil ^a
Accessible surface	29.6	70.4	38.5	61.5	23.9	15.1	61.0
Buried surface	42.0	58.0	36.9	63.1	29.6	30.5	39.9
Domain interface	28.0	72.0	25.4	74.6	29.6	23.2	47.2
Intradomain							
(solvated)	43.9	56.1	42.6	57.4	29.6	22.2	48.2
Intradomain (empty)	19.5	80.5	14.2	85.8	38.2	37.0	24.8
Interdomain							
(solvated)	42.0	58.0	43.5	56.5	22.5	34.2	43.3
Interdomain (empty)	21.3	76.7	19.5	80.5	42.6	23.6	33.8
Subunit interface	23.1	76.9	27.6	72.4	29.9	19.8	50.4
Intersubunit (solvated)	31.4	68.6	38.3	61.7	32.9	16.6	50.5
Intersubunit (empty)	24.9	75.1	28.2	71.8	35.1	26.9	38.0

^a Residue secondary structural assignments made using the method of Kabsch and Sander (1983).

Fig. 6. Amino acid propensities at various protein surfaces. Sorted propensities are plotted with one subtracted from each score so that a neutral propensity becomes zero. All solvated and empty cavity propensities were derived from the high-resolution, water-dense subsets. The full multidomain data set was used in calculations involving buried (within single domains only) and accessible (multidomain units) surfaces. A list of the actual values can be obtained from the authors upon application.

properties are exhibited by the water molecules for all 3 cavity types. Each water is confined on average to a volume of $30 \text{ \AA}^3$, as found in bulk water. However, the single-water cavities possess less average volume per water than multiple-water cavities,

found more often in subunit interfaces, which are generally the least well packed. Because solvated intersubunit cavities possess a somewhat more hydrophobic surface (Table 2) than the other 2 classes, fewer polar atoms would be available for hydrogen

Table 3. *Solvated cavity statistics*

Number of waters in cavity	Frequency	Cavity volume per water ($\text{\AA}^3$)	Cavity surface per water ($\text{\AA}^2$)	Polar surface per water ($\text{\AA}^2$)	Minimum polar surface per water ($\text{\AA}^2$)
Intradomain solvated cavities					
1	39	26.1	43.9	19.8	9.5
2	4	35.7	51.0	18.8	9.2
3	4	38.9	52.1	22.1	16.4
4	1	30.9	41.1	21.9	21.9
All	48	28.0	45.1	19.9	9.2
Interdomain solvated cavities					
1	27	23.8	41.6	19.9	11.3
2	11	29.5	42.1	17.2	9.2
3	2	24.4	35.4	14.6	13.5
4	1	34.8	44.3	17.0	17.0
5	1	26.7	32.2	11.1	11.1
6	1	29.7	33.3	13.2	13.2
All	43	25.8	41.1	18.6	9.2
Intersubunit solvated cavities					
1	13	22.3	38.9	18.1	8.7
2	5	45.5	60.8	28.4	12.6
3	1	44.7	53.3	5.1	5.1
4	1	41.4	48.8	24.4	24.4
5	1	44.0	56.9	10.9	10.9
All	21	30.8	46.1	19.9	5.1

bonding and thus, fewer waters would be expected to be comfortably accommodated. Nevertheless, all solvated cavity-type surfaces are mostly polar, possessing a minimum of $5.1 \text{ \AA}^2$ of polar surface per water, but on average $20 \text{ \AA}^2/\text{water}$.

Hydrogen bonding of cavity atoms and solvent

The hydrogen bonding statistics for cavity polar atoms are given in Table 4 and calculated with only the high-resolution, water-dense subsets. The polar atoms contributing to the surface of solvated cavities make a much more significant fraction of their hydrogen bonds to water molecules, without which they would average less than 1 hydrogen bond per atom. The "empty" cavity polar atoms are generally less favorably hydrogen bonded but make more than 1 hydrogen bond on average. It should be noted that the "waters" to which empty cavity polar atoms make hydrogen bonds lie either at the protein surface or in other cavities because the cavity is itself, by definition, empty. Ignoring the contribution from solvent to cavity hydrogen bonding, the "empty" cavity polar atoms are better hydrogen bonded, although this distinction is less clear for the intersubunit cavity atoms. It is not clear why this should be so. Similar statistics for buried atoms (taken from Hubbard et al., 1994) are listed for comparison. Solvated cavity polar atoms make more hydrogen bonds on average than buried atoms only when internal waters are considered, whereas "empty" cavities always make less.

The cavity waters in all 3 cavity classes are highly coordinated and bind preferentially to main-chain carbonyl oxygens, as observed previously (Baker & Hubbard, 1984; Hubbard et al., 1994). Each cavity water makes on average 3.4 hydrogen bonds.

Mobility of cavity atoms and solvent

Distributions of the average temperature factors of each cavity's atoms and the cavity's solvent as a number of standard deviations from the mean (NSDM) of buried and cavity-defining atoms, respectively, are plotted in Figure 7. Cavity-defining atoms are consistently less mobile than buried atoms. This contrasts with the results of our previous study (Hubbard et al., 1994) due to the exclusive definition of buried atoms, which are not considered to be accessible or interface atoms. These findings are also consistent with the recent proposal of Lee (1993), who suggests that the maximum destabilization of a protein due to mutation of large core residues to smaller ones occurs when the local environment is rigid, hence preventing the protein from rearranging its structure to eliminate the cavity.

The amino acid propensities calculated here for the buried surface exhibit a strong inverse correlation of -0.84 with the mean B_{norm} flexibility parameters used to predict mobility (Karplus & Schulz, 1985). Similarly, the amino acid propensities for "empty" cavities at intradomain, interdomain, and intersubunit surfaces show correlations of -0.67 , -0.74 , and -0.41 , respectively, with the Karplus and Schulz parameters. These results also suggest that cavities may be a product of the intrinsic inflexibility of the surrounding residue types, and thus the cavity itself is a result of the "unpackedness" of the local environment. This is especially the case within monomer (Hubbard et al., 1994) and domain cores (data not shown), where the cavities are often concentrated.

The atoms contributing to the surface of "empty" cavities are slightly more flexible than for solvated cavities, with the for-

Table 4. Potential hydrogen bonding of cavity polar atoms and solvent

Cavity type	Hydrogen bonds made by atomic subclass	Number of atoms	Mean potential hydrogen bonds per atom (including waters)	Mean potential hydrogen bonds per atom (excluding waters)
Intradomain solvated	All	239	1.71	0.84
	Main-chain oxygen	98	1.67	0.81
	Main-chain nitrogen	68	1.40	0.65
	Side-chain polar	73	2.04	1.05
Intradomain empty	All	96	1.20	0.98
	Main-chain oxygen	53	1.25	1.00
	Main-chain nitrogen	24	1.04	0.92
	Side-chain polar	19	1.26	1.00
Interdomain solvated	All	213	1.66	0.72
	Main-chain oxygen	96	1.73	0.75
	Main-chain nitrogen	52	1.46	0.75
	Side-chain polar	65	1.72	0.66
Interdomain empty	All	30	1.37	1.17
	Main-chain oxygen	15	1.27	1.20
	Main-chain nitrogen	9	1.33	1.22
	Side-chain polar	6	1.67	1.00
Intersubunit solvated	All	111	1.57	0.78
	Main-chain oxygen	39	1.69	0.85
	Main-chain nitrogen	28	0.93	0.36
	Side-chain polar	44	1.86	1.00
Intersubunit empty	All	48	1.00	0.81
	Main-chain oxygen	17	1.06	0.88
	Main-chain nitrogen	13	0.62	0.62
	Side-chain polar	18	1.22	0.89
Buried atoms ^a	All	10,025	1.40	1.31
	Main-chain oxygen	3,875	1.48	1.37
	Main-chain nitrogen	5,252	1.28	1.23
	Side-chain polar	898	1.74	1.52

^a Taken from Hubbard et al. (1994).

mer possessing mean NSDMs from buried atoms of -0.48 , -0.33 , and -0.31 for the intradomain, interdomain, and intersubunit cases, respectively, compared respectively to -0.56 , -0.35 , and -0.52 for solvated cavity atoms. The cavity waters may restrict local motions even more, both sterically and through hydrogen bonding with the protein. The cavity solvent itself exhibits a broad range of relative mobilities, although on average it is more flexible than cavity-forming atoms. Cavity waters in all 3 classes were consistently less mobile than bulk water, as expected, with mean NSDM values of -1.57 , -1.33 , and -1.17 for intradomain, interdomain, and intersubunit cavity solvent, respectively.

Conclusions

The total cavity volume in a folded protein remains a small fraction of the total protein volume but increases in absolute amount with protein size. Cavities are almost always found in proteins greater than 100 residues in size. Solvated and "empty" cavities share the same basic requirements, i.e., solvation supports larger cavities and a more polar cavity surface, and the hydrogen bonding and packing of internal solvent is not compromised by the

cavity packing constraints. These properties are maintained in cavities that are formed within domains, between domains, and between subunits. Interdomain and intersubunit cavities are on average larger and more often solvated than intradomain cavities and exhibit different amino acid preferences. The results presented here suggest broad guidelines regarding total cavity volume, cavity size, solvation, and amino acid composition for all cavity classes, which should be of use in protein engineering and design.

The results of our studies show very clearly that many domain-domain and subunit-subunit interfaces are far from perfectly packed although, in general, reasonably good shape complementarity is maintained. Lawrence and Colman (1993) considered the quality of surface match between unlike protein interfaces, such as antibody/antigen complexes, and concluded that the latter were not as complementary as oligomeric or proteinase/inhibitor complexes. Our results over a much larger data set suggest that oligomeric interfaces themselves are nonoptimally matched. These findings have impact on the protein docking problem because the majority of the associated methods rely heavily on shape complementarity to eliminate poor complexes from further consideration. Furthermore, many

Fig. 7. Distributions of the number of standard deviations from a given mean of the average temperature factors of cavity-forming atoms and cavity waters. Distributions are shown with respect to the mean temperature factor of buried atoms in each protein for (A) intradomain, (B) interdomain, and (C) intersubunit cavity-forming atoms, and with respect to the mean temperature factor of cavity-forming atoms for (D) intradomain, (E) interdomain, and (F) intersubunit cavity waters.

of the techniques are unable to select unambiguously the correct "docked" conformation from a set of "top-scoring" surface-complementary solutions. It appears unlikely that the native fit of 2 interacting surfaces is geometrically the most optimal (cf. Shiochet & Kuntz, 1991, 1993), a hypothesis originally proposed by Connolly (1986). We suggest that the findings presented here provide consistency checks with the cavity characteristics shown by real complexes when testing final solutions from docking algorithms. Because a significant proportion of interface surface is constituted by solvated cavities, inclusion of solvent in such simulations is also strongly recommended.

The creation of cavities in proteins has been shown to be destabilizing to the free energy of folding (Kellis et al., 1988, 1989; Karpusas et al., 1989; Eriksson et al., 1992, 1993). However, it is difficult to quantify the destabilization directly due to cavity formation and reported values differ (Rashin et al., 1986; Eriksson et al., 1993). Using the estimate from Eriksson et al. (1993) of 24 cal per mol $\text{\AA}^3$ of cavity formed, we obtain a value of 48 cal per mol residue as an average for proteins. A 100-residue protein must accommodate about 5 kcal mol⁻¹. However, this value makes no distinction between "empty" and solvated cavities. Repeating our calculation using only empty cavities in the high-resolution data set, we obtain a value closer to 12 cal per mol residue, which may be a more realistic value because Eriksson and coworkers' estimate is derived from purely empty cavities. Nevertheless, these values should be considered highly approximate.

How do proteins accommodate such packing defects? We were unable to find any simple correlation of shape complementarity with interfacial area, hydrophobicity, or hydrogen bonding, and conclude that proteins accommodate cavities in many

different ways involving many or all of the above and probably other features. Furthermore, different multimeric proteins may require different levels of stability for their function and regulation; for example, a multimeric protein involved in an allosteric response such as hemoglobin-binding oxygen would not function properly with overly stable subunit interactions.

The increased size and polarity of some intersubunit cavities may be due to their possible ligand-binding role. However, this hypothesis is not proven. The size of the cavities detected may not be large enough to comfortably bind ligands larger than a large side-chain group such as Phe, which is about 190 $\text{\AA}^3$ (Gerstein et al., 1994b). We find only 15 intersubunit cavities above this size, of which only 4 are empty or unknown. It is possible they may be required to rearrange their structure to accommodate ligands or act as intermediate points for ligands traversing the protein to some other site (Elber & Karplus, 1990; Lambright et al., 1994) or to bind smaller ligands.

The observation that intradomains are better packed than interdomain or subunit regions can be explained by a number of hypotheses concerning cavities and their roles in protein structure and function.

1. The mutational pressure at protein interfaces leads to a greater cavity density. Domains are independent folding units and can tolerate many mutationally prone loops at their surfaces, yielding varying interfaces and more cavities. Within domains themselves, the core cannot tolerate so many mutations and cavities are correspondingly fewer.

2. Domains evolved and developed before multidomain and multimeric proteins and have subsequently optimized packing with few cavities. Because domains are used as building blocks

for higher-order protein structures coming later in evolution, there has been insufficient time to achieve the same level of packing between them. Thus, the looser packing between domains and subunits is simply a result of the "packing history" of the protein in question.

3. Interface cavities are structurally and/or functionally important, for the like of conformational changes when interacting with substrates or flexibility between units, and hence are found in a higher proportion. The intradomain cavities would be fewer because such motions are generally undesirable for stability of the overall fold; nevertheless, a smaller proportion of intradomain cavities could also serve functional purposes such as aiding catalysis through atomic rearrangements.

It is possible that the higher percentage of interface cavities is a result of all three. The functional role of cavities in general in proteins is currently under investigation (S.J. Hubbard & P. Argos, manuscript in prep.), although it seems certain that some cavities are highly functional, involving channels to active sites (Finer-Moore et al., 1992; Kossiakoff et al., 1992) and domain-domain movements (Gerstein et al., 1994a). As yet, the fraction of cavities that are nonfunctional, and are therefore simply tolerated, is still unknown.

Materials and methods

Data

The 2 protein data sets (multidomain and multisubunit) chosen for analysis are shown in Table 1A and B, along with the Brookhaven (Bernstein et al., 1977) 4-letter code identifier given for each protein. The selected sequences within each data set are less than 40% identical to every other member to eliminate bias from homologous proteins. The classification of protein domains was taken primarily from Argos (1988), along with the multisubunit protein data set and, in a number of instances, the domain definitions were extended for recent Brookhaven entries. For the comparison of solvated and "empty" cavity properties, subsets of both data sets were chosen where the resolution of the structure was 2.0 Å or better and a minimum of 1 water per 2 residues was given in the Brookhaven file, to reduce misclassification of cavity solvation state.

Methods

Surface and volume calculations

All surfaces and volumes were calculated analytically using the suite of programs by Connolly (1983, 1985b, 1993). As found to be optimal in our previous work (Hubbard et al., 1994), a probe of radius 1.25 Å was rolled over the van der Waals surface of the protein to delineate unique surface components describing the atomic (external) accessible surface and successive internal cavities. The molecular surface (the sum of the contact and reentrant surfaces) was used for all calculations in this paper because it converges to $4\pi r^2$ (where r is the probe radius) rather than 0 for the smallest detectable cavity. The van der Waals radii used for the protein atoms are listed in Hubbard et al. (1994) and were originally used by Rashin et al. (1986). Heteroatoms as defined in the Brookhaven Data Bank for cofactors, inhibitors, ligands, and the like were ignored for the cal-

culation of surfaces; nonetheless, all cavities were subsequently checked for their presence. Only two were found for the proteins in our data sets and were ignored for the purposes of this study. Three extremely large cavities greater than 500 Å³ were found lying between subunits. On consultation of the literature, these were seen to be either putative ligand binding sites or formed active sites and were thus not considered in this study.

Classification of cavities

Cavities were divided into 3 classes: intradomain, interdomain, and intersubunit. Cavities with atomic surface components arising from more than 1 domain were deemed to be interdomain, and from more than 1 subunit, intersubunit. The intradomain cavities were taken only from the single domains of multidomain proteins (Table 1A). Cavities were further classified according to solvation state by assigning all waters and heteroatoms to that cavity if they were within a small distance g of a set of grid points spaced at g Å and filling the cavity (see Hubbard et al., 1994). Only when waters were present in the Brookhaven entry, but none assigned to the cavity, were the cavities deemed "empty."

Cavity conformational parameters

All nitrogen and oxygen atoms were considered polar and the remainder nonpolar. The polarity of each cavity was calculated as the polar surface fraction of the total cavity surface.

Individual amino acid side-chain propensities for the different protein surfaces were calculated by:

$$P_{ts} = \frac{A_{ts}/A_s}{C_t/C_{tot}},$$

where A_{ts} is the total surface of type s contributed by the residue type t , A_s is the total surface of type s , C_t the total "possible" surface contribution of that residue type, and C_{tot} the total "possible" surface over the whole data set. The "possible" surface is the sum of the molecular surface of each residue type t in an extended tripeptide Ala- t -Ala. Preferences were calculated using only side-chain atoms of the 20 amino acid types for the following surface types: accessible, buried, interface (domain), interface (subunit), intradomain cavity, interdomain cavity, and intersubunit cavity. The cavity propensities calculated separately for solvated and "empty" cavities relied only on proteins from the high-resolution subsets. Accessible surface propensities were calculated from the accessible surface of the multidomain proteins, excluding the interdomain and all cavity surfaces. The buried surface, taken from the same data set, was defined as the surface lost in going from the extended to folded state of each individual domain, excluding all cavity surface.

The secondary structural preferences of all residues contributing to various surfaces were also estimated, where the method of Kabsch and Sander (1983) was used to assign each protein residue as helix (H), strand (E), or coil (not H or E). The percentage contributions from these 3 states for the different surface types were then compared, along with the contributions from main- and side-chain, and polar and nonpolar atoms. Hydrogen bonding of cavity atoms and solvent was calculated as previously described (Hubbard et al., 1994), using an acceptor-donor cutoff distance of 3.6 Å and a hydrogen bond angle of 90–180°.

The crystallographically defined atomic temperature factors (B -values) were utilized to measure the relative mobility of the cavity-defining atoms. The average temperature factor for such atoms in each cavity was expressed as NSDM of totally buried atoms in a given protein. The mean was then determined over all cavities and subsequently over all proteins considered. We define the buried atom subset to be exclusive, i.e., atoms with zero accessibility to the solvent and lacking involvement with interfaces or cavity surfaces. Similarly, NSDM values were calculated for the cavity waters in solvated cavities by expressing their mean B -values relative to the mean and standard deviation of those for the cavity-forming atoms and also the bulk solvent (all waters not contained in cavities).

Acknowledgments

This work was in part supported by an Institutional EC Human Capital and Mobility grant ERBCHBGCT930241 to P.A. and S.J.H. and in part from a grant from the German Bundesministerium für Forschung und Technologie (FG5-1075 to P.A.).

References

- Argos P. 1988. An investigation of protein subunit and domain interfaces. *Protein Eng* 2:101–113.
- Bacon DJ, Moulton J. 1992. Docking by least-squares fitting of molecular surface patterns. *J Mol Biol* 225:849–858.
- Baker EN, Hubbard RE. 1984. Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol* 44:97–179.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:532–542.
- Chothia C. 1976. The nature of the accessible and buried surface in proteins. *J Mol Biol* 105:1–14.
- Chothia C, Janin J. 1975. Principles of protein–protein recognition. *Nature* 256:705–708.
- Connolly ML. 1983. Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221:709–713.
- Connolly ML. 1985a. Atomic size packing defects in proteins. *Int J Pept Protein Res* 28:360–363.
- Connolly ML. 1985b. Computation of molecular volume. *J Am Chem Soc* 107:1118–1124.
- Connolly ML. 1986. Shape complementarity at the hemoglobin $\alpha 1 \beta 1$ subunit interface. *Biopolymers* 25:1229–1247.
- Connolly ML. 1993. The molecular surface package. *J Mol Graphics* 11:139–141.
- Eisenmenger F, Argos P, Abagyan R. 1993. A method to configure protein sidechains from the mainchain trace in homology modelling. *J Mol Biol* 231:849–860.
- Elber R, Karplus M. 1990. Enhanced sampling in molecular dynamics: Use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J Am Chem Soc* 112:9161–9175.
- Eriksson AE, Baase WA, Matthews BW. 1993. Similar hydrophobic replacements of Leu 99 and Phe 153 within the core of T4 lysozyme have different structural and thermodynamic consequences. *J Mol Biol* 229:747–769.
- Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW. 1992. Response of a protein structure to a cavity creating mutations and its relation to the hydrophobic effect. *Science* 255:178–183.
- Finer-Moore JS, Kossiakoff AA, Hurley JH, Earnest T, Stroud RM. 1992. Solvent structure in crystals of trypsin determined by X-ray and neutron diffraction. *Proteins Struct Funct Genet* 12:203–222.
- Gerstein M, Lesk AM, Chothia C. 1994a. Structural mechanisms for domain movement in proteins. *Biochemistry* 33:6739–6749.
- Gerstein M, Sonnhammer ELL, Chothia C. 1994b. Volume changes in protein evolution. *J Mol Biol* 236:1067–1078.
- Helmer-Citterich M, Tramontano A. 1994. PUZZLE: A new method for automated protein docking based on surface complementarity. *J Mol Biol* 235:1021–1031.
- Holm L, Sander C. 1991. Database algorithm for generating protein backbone and sidechain co-ordinates from a C^α trace. Application to model building and detection of co-ordinate errors. *J Mol Biol* 218:183–194.
- Hubbard SJ, Gross KH, Argos P. 1994. Intramolecular cavities in globular proteins. *Protein Eng* 7:613–626.
- Janin J, Miller S, Chothia C. 1988. Surface, subunit interfaces and interior of oligomeric proteins. *J Mol Biol* 204:155–164.
- Jiang F, Kim SH. 1991. "Soft docking" matching of molecular surface cubes. *J Mol Biol* 219:79–102.
- Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* 22:2577–2637.
- Karplus PA, Schulz GE. 1985. Prediction of chain flexibility in proteins. *Naturwissenschaften* 72:212–213.
- Karpusas M, Baase WA, Matsumura M, Matthews BW. 1989. Hydrophobic packing in T4 lysozyme probed by cavity-filling mutants. *Proc Natl Acad Sci USA* 86:8237–8241.
- Kellis JT Jr, Nyberg K, Fersht AR. 1989. Energetics of complementary side chain packing in a hydrophobic core. *Biochemistry* 28:4914–4922.
- Kellis JT Jr, Nyberg K, Sali D, Fersht AR. 1988. Contribution of hydrophobic interactions to protein stability. *Nature* 333:784–786.
- Korn AP, Burnett RM. 1991. Distribution and complementarity of hydrophobicity in multisubunit proteins. *Proteins Struct Funct Genet* 9:37–55.
- Kossiakoff AA, Sintchak MD, Shpungin J, Presta LG. 1992. Analysis of solvent structure in proteins using D_2O - H_2O solvent maps: Pattern of primary and secondary hydration of trypsin. *Proteins Struct Funct Genet* 12:223–236.
- Lambright DG, Balasubramanian S, Decatur SM, Boxer SG. 1994. Anatomy and dynamics of a ligand binding pathway in myoglobin. *Biochemistry* 33:5518–5525.
- Lawrence MC, Colman PM. 1993. Shape complementarity at protein/protein interfaces. *J Mol Biol* 234:946–950.
- Lee B. 1993. Estimation of the maximum change in stability of globular proteins upon mutation of a hydrophobic residue to another of a smaller size. *Protein Sci* 2:733–738.
- Lee C, Subbiah S. 1991. Prediction of protein sidechain conformation by packing optimisation. *J Mol Biol* 217:373–388.
- Miller S. 1989. The structure of interfaces between subunits of dimeric and tetrameric proteins. *Protein Eng* 3:77–83.
- Norel R, Lin SL, Wolfson HJ, Nussinov R. 1994. Shape complementarity and protein–protein interfaces. *Biopolymers* 34:933–940.
- Novotny J, Rashin AA, Brucoleri RE. 1988. Criteria that discriminate between native proteins and incorrectly folded models. *Proteins Struct Funct Genet* 4:19–30.
- Ponder JW, Richards FM. 1987. Tertiary templates for proteins: Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol* 193:775–791.
- Rashin AA, Iofin M, Honig B. 1986. Internal cavities and buried waters in globular proteins. *Biochemistry* 25:3619–3625.
- Richards FM. 1974. The interpretation of protein structures: Total volume, group volume distributions and packing density. *J Mol Biol* 82:1–14.
- Richards FM. 1977. Areas, volumes, packing and protein structures. *Annu Rev Biophys Bioeng* 6:151–176.
- Richards FM, Lim WA. 1994. An analysis of packing in the protein folding problem. *Q Rev Biophys* 26:423–498.
- Shiochet BK, Kuntz ID. 1991. Protein docking and complementarity. *J Mol Biol* 221:327–346.
- Shiochet BK, Kuntz ID. 1993. Matching chemistry and shape in molecular docking. *Protein Eng* 6:723–732.
- Silva AM, Rossmann MG. 1985. The refinement of southern bean mosaic virus in reciprocal space. *Acta Crystallogr B* 41:147–157.
- Tulip WR, Varghese JN, Laver WG, Webster RG, Colman PM. 1992. Refined crystal structure of the influenza virus N9 neuraminidase–NC41 Fab complex. *J Mol Biol* 227:122–148.
- Walls PH, Sternberg MJE. 1992. New algorithm to model protein–protein recognition based on surface complementarity. *J Mol Biol* 228:277–297.
- Wodak SJ, De Crombrughe M, Janin J. 1987. Computer studies of interactions between macromolecules. *Prog Biophys Mol Biol* 49:29–63.
- Young L, Jernigan RL, Covell DG. 1994. A role for surface hydrophobicity in protein–protein recognition. *Protein Sci* 3:717–729.
- Zielenkiewicz O, Rabcenko A. 1984. Protein–protein recognition: Method for finding complementary surfaces of interacting proteins. *J Theor Biol* 111:17–30.