PROTEINS: Structure, Function, and Genetics 53:708–719 (2003)

Dissecting Subunit Interfaces in Homodimeric Proteins

Ranjit Prasad Bahadur,[1] Pinak Chakrabarti,[1] Francis Rodier,[2] and Joël Janin[2]*
[1]*Department of Biochemistry, Bose Institute, Calcutta, India*
[2]*Laboratoire d'Enzymologie et de Biochimie Structurales, CNRS UPR 9063, Gif-sur-Yvette, France*

ABSTRACT **The subunit interfaces of 122 homodimers of known three-dimensional structure are analyzed and dissected into sets of surface patches by clustering atoms at the interface; 70 interfaces are single-patch, the others have up to six patches, often contributed by different structural domains. The average interface buries 1,940 Å^2 of the surface of each monomer, contains one or two patches burying 600–1,600 Å^2, is 65% nonpolar and includes 18 hydrogen bonds. However, the range of size and of hydrophobicity is wide among the 122 interfaces. Each interface has a core made of residues with atoms buried in the dimer, surrounded by a rim of residues with atoms that remain accessible to solvent. The core, which constitutes 77% of the interface on average, has an amino acid composition that resembles the protein interior except for the presence of arginine residues, whereas the rim is more like the protein surface. These properties of the interfaces in homodimers, which are permanent assemblies, are compared to those of protein–protein complexes where the components associate after they have independently folded. On average, subunit interfaces in homodimers are twice larger than in complexes, and much less polar due to the large fraction belonging to the core, although the amino acid compositions of the cores are similar in the two types of interfaces. Proteins 2003;53:708–719.**

Key words: interface area; protein–protein recognition; protein–protein interaction; interfaces; hydrophobicity; polar interactions; oligomeric proteins

INTRODUCTION

Homodimers are the simplest example of the noncovalent self-assembly of proteins. This class of proteins is abundantly represented in the Protein Data Bank (PDB).[1] Their subunit interfaces, defined as the regions of the protein surface that are involved in subunit contacts, display geometric and chemical properties that give the assembly its stability and specificity. The interfaces formed by the components of protein–protein complexes could be expected to share the same properties. However, most homodimeric proteins are permanent assemblies, and their polypeptide chains assemble at the time they fold. In contrast, complexes involve proteins that fold separately and remain monomers until they meet and associate. Several studies have been devoted to subunit interfaces in oligomeric proteins[2,3] and in protein–protein complexes.[3,4] As the number of new X-ray structures increases rapidly, these studies must be updated regularly. Here, we examine a sample of 122 homodimer interfaces with tools that we already applied to interfaces in complexes.[5] We show that subunit interfaces in homodimers are very diverse both in their size and in their chemical composition. On average, they are more hydrophobic and bury twice as much protein surfaces as in complexes. The larger interfaces involve several patches of the subunit surface, often contributed by separate structural domains. Like in complexes, the subunit interfaces have a core of residues containing atoms that are fully buried in the dimer. The core has an amino acid composition similar to that of the protein interior, and is surrounded by a rim that remains solvent accessible and has an amino acid composition similar to that of the protein surface.

METHODS

One hundred twenty-two entries representing X-ray structures of homodimeric proteins were taken from the Protein Data Bank[1] at the Research Collaboratory for Structural Bioinformatics or the Protein Quaternary Structure Server (PQS)[6] at the European Bioinformatics Institute. When more than one entry was available for a particular protein, the structure with the highest resolution was retained; 84 out of 122 structures have resolutions better than 2.0 Å; 32 have resolution in the range 2–2.5 Å, and 6 in the range 2.5–3.0 Å. The sequence identity between any two proteins in the set was less than 25%. In cases where the dimer did not constitute the crystal asymmetric unit, it was generated with the information of record BIOMT if it was present. Alternatively, we used the program CRISPACK[7] to generate neighbors, and selected the pair of twofold related subunits that had the largest interface area. The procedure was implemented in program ASPIC.

The solvent accessible surface area (ASA) was computed using the program ACCESS,[8] which is an implementation of the algorithm of Lee and Richards.[9] The subunit interface area was estimated as $B/2$ where B was:

Grant sponsor: Indo-French Centre for the Promotion of Advanced Research (CEFIPRA); Grant number: 2203-2.

*Correspondence to: J. Janin, LEBS-CNRS, 91198-Gif-sur-Yvette, France. E-mail: janin@lebs.cnrs-gif.fr

Received 10 December 2002; Accepted 20 February 2003

TABLE I. PDB Entries for Homodimers

12as	1a3c	1a4i	1a4u	1aa7	1ad3	1ade	1af5	1afw	1ajs
1alo	1amk	1aor	1aq6	1auo	1b3a	1b5e	1b67	1b8a	1b8j
1bam	1bbh	1bd0	1bif	1biq	1bis	1bjw	1bkp	1bmd	1brw
1bsl	1bsr	1buo	1bxg	1bxk	1cdc	1cg2	1chm	1cmb	1cnz
1coz	1csh	1ctt	1cvu	1czj	1daa	1dor	1dpg	1dqs	1dxg
1e98	1ebh	1f13	1fip	1fro	1gvp	1hhp	1hjr	1hss	1hxp
1icw	1imb	1isa	1ivy	1jhg	1jsg	1kba	1kpf	1lyn	1m6p
1mkb	1mor	1nox	1nse	1nsy	1oac	1opy	1pgt	1pre	1qfh
1qhi	1qr2	1r2f	1reg	1rfb	1rpo	1ses	1slt	1smn	1smt
1sox	1tc1	1tox	1trk	1uby	1utg	1vfr	1vok	1wtl	1xso
2arc	2ccy	2hdh	2ilk	2lig	2mcg	2nac	2ohx	2spc	2sqc
2tct	2tgi	3dap	3grs	3sdh	3ssi	4cha	4kbp	5csm	5rub
8prk	9wga								

$$B = \mathrm{ASA}_{\mathrm{monomer1}} + \mathrm{ASA}_{\mathrm{monomer2}} - \mathrm{ASA}_{\mathrm{dimer}} \quad (1)$$

All atoms or amino acid residues in the monomer that lose more than 0.1 Å^2 ASA in the dimer were counted as interface atoms or residues. Residues with one or more interface atoms completely buried (zero ASA) in the dimer, were considered as belonging to the core. Residues where all interface atoms have residual accessibility formed the rim of the interface.

Clusters of atoms forming patches at interfaces were identified by the algorithm based on the average linkage method reported in Chakrabarti and Janin.[5] The clustering algorithm was run on only one subunit of each dimer. The algorithm requires setting a threshold distance for atoms in a cluster. A threshold value of 22 Å was selected as representing half the average value of the maximum distance D_{max} between any two interface atoms within a subunit. In our sample, the range of D_{max} was 24–100 Å. All interfaces were visually checked for the occurrence of the patches. In eight small interfaces, the clusters defined by the program did not seem to be justified on visual inspection, and the threshold value was reduced accordingly. In four cases, it was increased to 25 or 28 Å.

RESULTS

The Sample

The selection of PDB entries representing homodimers and of the relevant interfaces is not straightforward for several reasons: (1) the quaternary structure of a protein in solution is often unknown or uncertain, whether it is reported in the PDB entry or not; (2) atomic coordinates derived from X-ray structures relate to the crystal asymmetric unit rather than the biological unit; (3) in a crystal, the dimer interface is only one of the many interfaces between polypeptide chains that hold the crystal together.[6,10,11] The procedure we followed in this work takes these points into account.

We selected the PDB entries of proteins that were reported as being dimers in solution based on biochemical or biophysical data such as analytical centrifugation. We applied crystal symmetries and translations to generate all neighbors of the polypeptide chain(s) present in the asymmetric unit, and selected pairs on the ground that they had either exact (crystallographic) or approximate (local) twofold symmetry, and a large interface area. The selected interface was generally the largest of all the pairwise interfaces between neighboring chains. This selection procedure is in line with the work of Ponstingl et al.[11] who assembled a set of 76 homodimers on the same basis. As neither the symmetry nor the interface area criterion is absolute, we rejected entries where the structure in solution was uncertain, or where two interfaces of similar size were observed in the crystal. In the latter case, site-directed mutagenesis may help identifying the biologically relevant interface, but this information is rarely available. Because the present study deals with the protein component of the interfaces, we also rejected dimers where more than 5% of the interface area was contributed by ligands, prosthetic groups, or other nonprotein elements.

In a second step, we checked the selected pairs against the information listed in headers of the PDB entry, and against the PQS Server. A few discrepancies could be traced to errors in either the PDB, the PQS data base, or our selection. Others pointed to undetected ambiguities, and led us to reject the entry. We finally assembled the set of 122 homodimers listed in Table I. Of the 76 entries in Ponstingl et al.,[11] we have retained 62, excluded 3 as containing a covalently bound cofactor or a large prosthetic group at the interface, and updated 11 to include higher resolution structures. The list has 49 novel entries. In 30 of the homodimers, the twofold symmetry was a crystal symmetry; the remainder had noncrystalline symmetry.

Size of the Subunit Interfaces in Homodimers

We estimated the size of interfaces in dimers by measuring the area of the protein surface buried in subunit contacts. This area is quoted per subunit in Table II and noted *B/2* for consistency with our previous work on protein–protein complexes,[4,5] where *B* represented the total surface area lost by the two components of the complex. In a homodimer, twofold symmetry implies that each component contributes equally to *B*, which is not necessarily the case in a heterodimer or a complex.

Figure 1(A) is a plot of the *B/2* values observed in the 122 homodimers against the protein size, estimated by the subunit accessible surface area ($\mathrm{ASA}_{\mathrm{monomer1}}$ in Eq. 1).

TABLE II. Properties of Homodimer Interfaces

			Interface[a]				Area fraction	
PDB entry	Protein	Patches	$B/2$ (Å^2)	Residues	Atoms	H-bonds	In core[b]	Polar[c]
12as	Asparagine synthetase	1	1,989	52	182	5	0.63	0.28
1a3c	Pyrimidine operon regulator PyrR	1	853	27	95	5	0.87	0.47
1a4i	Methylene-tetrahydrofolate Dehydrogenase/cyclohydrolase	1	1,353	38	134	12	0.73	0.43
1a4u	Alcohol dehydrogenase	1	2,547	63	238	10	0.78	0.28
1aa7	Influenza virus matrix protein	1	1125	26	106	4	0.89	0.45
1ad3	Aldehyde dehydrogenase class 3	4	3,936	110	388	20	0.81	0.40
1ade	Adenylosuccinate synthetase	1	2,708	78	283	11	0.68	0.41
1af5	I-CREI	1	856	24	97	1	0.74	0.27
1afw	3-Ketoacyl-CoA thiolase	2	2,400	70	255	16	0.77	0.37
1ajs	Aspartate aminotransferase	3	3,401	99	335	19	0.75	0.38
1alo	Aldehyde oxidoreductase	3	1,215	43	137	2	0.49	0.42
1amk	Triose-phosphate Isomerase	1	1,477	39	153	11	0.94	0.38
1aor	Aldehyde-ferredoxin oxydoreductase	1	1,180	33	125	8	0.76	0.45
1aq6	L-2-haloacid dehalogenase	1	2,232	54	225	11	0.93	0.35
1auo	Carboxylesterase	2	662	20	76	4	0.73	0.35
1b3a	RANTES	1	763	23	86	5	0.79	0.39
1b5e	Deoxycytidylate hydroxymethylase	1	2,581	65	253	15	0.79	0.40
1b67	Histone HMFA	2	1,607	40	158	4	0.77	0.22
1b8a	Aspartyl-tRNA synthetase	3	4,391	114	448	26	0.90	0.34
1b8j	Alkaline phosphatase	3	3,794	103	392	25	0.83	0.39
1bam	Endonuclease BAM-HI	1	745	18	76	3	0.75	0.37
1bbh	Cytochrome C'	1	771	24	86	1	0.84	0.20
1bd0	Alanine racemase	2	3,091	85	323	20	0.63	0.44
1bif	6-P-fructo-2-kinase/bisphosphatase	2	858	28	97	5	0.64	0.50
1biq	Ribonucleotide reductase R2	3	3,004	74	301	19	0.84	0.37
1bis	HIV-1 integrase core domain	1	1,495	40	161	4	0.81	0.30
1bjw	Aspartate aminotransferase	3	2,938	76	323	23	0.94	0.45
1bkp	Thymidylate synthase A	1	2,206	58	223	22	0.79	0.45
1bmd	Malate dehydrogenase	2	1,564	41	159	8	0.73	0.40
1brw	Pyrimidine nucleoside phosphorylase	1	1,083	27	112	5	0.88	0.33
1bsl	Bacterial luciferase beta	1	1,918	55	212	4	0.93	0.34
1bsr	Ribonuclease (bovine, seminal)	2	1,888	48	197	12	0.79	0.43
1buo	Zinc finger protein PLZF	2	1,972	49	189	10	0.85	0.33
1bxg	Phenylalanine dehydrogenase	1	1,041	25	103	3	0.70	0.44
1bxk	dTDP-glucose 4,6-dehydratase	1	1,286	38	139	3	0.72	0.34
1cdc	CD2 N-terminal domain	2	3,918	85	383	31	0.89	0.34
1cg2	Carboxypeptidase G2	1	1,298	40	130	4	0.78	0.31
1chm	Creatine amidinohydrolase	3	3,171	85	338	14	0.81	0.39
1cmb	Met apo-repressor (MetJ)	1	1,797	39	168	7	0.82	0.24
1cnz	3-Isopropylmalate dehydrogenase	2	2,447	63	247	18	0.84	0.31
1coz	Glycerol-3-P cytidyltransferase	1	1,050	24	104	5	0.90	0.30
1csh	Citrate synthase	4	5,057	117	497	23	0.80	0.35
1ctt	Cytidine deaminase	1	1,990	53	205	14	0.85	0.33
1cvu	Prostaglandin H2 synthase-2	3	2,436	69	260	18	0.79	0.41
1czj	Cytochrome C3	1	829	20	80	8	0.79	0.39
1daa	D-aminoacid aminotransferase	1	2,193	61	227	8	0.84	0.31
1dor	Dihydroorotate dehydrogenase A	2	2,189	60	219	10	0.72	0.27
1dpg	Glucose 6-P dehydrogenase	2	2,293	62	224	9	0.72	0.30
1dqs	Phosphate-lyase (cyclizing)	2	1,640	49	174	7	0.68	0.39
1dxg	Desulforedoxin	1	729	21	71	7	0.83	0.39
1e98	Thymidylate kinase	1	770	21	83	1	0.75	0.19
1ebh	Enolase	1	1,784	55	197	11	0.74	0.48
1f13	Coagulation factor XIII zymogen	5	2,556	83	290	12	0.64	0.43
1fip	Factor for inversion stimulation	1	1,836	43	187	5	0.91	0.25
1fro	Lactoylglutathione lyase	1	3,505	97	356	18	0.70	0.33
1gvp	Gene V protein	1	908	24	88	1	0.71	0.20

TABLE II. (Continued)

PDB entry	Protein	Patches	Interface[a]				Area fraction	
			B/2 (Å^2)	Residues	Atoms	H-bonds	In core[b]	Polar[c]
1hhp	HIV-1 protease	2	1,599	36	164	10	0.81	0.39
1hjr	RuvC holliday junction resolvase	1	962	26	99	1	0.75	0.23
1hss	Alpha-amylase inhibitor	1	1,101	31	106	5	0.56	0.36
1hxp	Hexose-1-P uridyltransferase	2	3,402	87	353	15	0.84	0.35
1icw	Interleukin-8	1	954	28	103	7	0.87	0.45
1imb	Inositol monophosphatase	1	1,623	44	159	7	0.73	0.32
1isa	Superoxide dismutase	1	920	23	96	7	0.87	0.40
1ivy	Carboxypeptidase L	1	1,601	49	178	5	0.62	0.40
1jhg	Trp operon repressor	1	2,207	53	218	3	0.70	0.26
1jsg	Oncogene product P14TCL1	1	794	21	80	2	0.71	0.32
1kba	Kappa-bungarotoxin	1	498	14	50	4	0.77	0.38
1kpf	Protein kinase C interacting protein	1	1,867	46	194	14	0.93	0.41
1lyn	Sperm lysin	1	948	24	96	2	0.84	0.27
1m6p	Mannose-6-P receptor	1	1,025	34	108	4	0.57	0.35
1mkb	β-OH-decanoyl-thioester dehydrase	1	1,605	43	167	12	0.82	0.41
1mor	Glucosamine 6-P synthase	2	2,540	69	269	17	0.76	0.42
1nox	NADH oxidase	2	3,033	75	304	18	0.75	0.33
1nse	Nitric oxide synthase	2	2,736	75	283	11	0.83	0.34
1nsy	NAD synthetase	2	2,592	64	267	5	0.82	0.30
1oac	Copper amine oxidase	6	7,149	192	724	48	0.77	0.38
1opy	Δ5-3-ketosteroid isomerase	1	1,048	30	114	2	0.82	0.34
1pgt	Glutathione S-transferase	1	1,238	32	117	4	0.77	0.40
1pre	Proaerolysin	5	2,300	80	252	10	0.41	0.48
1qfh	Gelation factor ROD domains 5–6	2	2,264	58	242	16	0.81	0.43
1qhi	Thymidine kinase	2	1,714	48	165	5	0.75	0.21
1qr2	Quinone reductase type2	2	1,947	54	215	10	0.79	0.34
1r2f	Ribonucleotide reductase R2	1	1,746	50	187	8	0.68	0.32
1reg	T4 Reg A	1	659	18	72	3	0.76	0.40
1rfb	Interferon gamma	2	2,650	73	288	4	0.37	0.28
1rpo	ROP (repressor of primer)	1	1,405	35	128	5	0.71	0.28
1ses	Seryl-tRNA synthetase	1	2,211	59	218	15	0.91	0.35
1slt	S-Lectin	1	536	16	58	4	0.67	0.44
1smn	Extracellular endonuclease	1	866	27	86	6	0.83	0.38
1smt	Transcriptional repressor SmtB	2	1,970	49	195	4	0.82	0.24
1sox	Sulfite oxidase	1	1,404	47	146	7	0.64	0.39
1tc1	Hypoxanthine P-ribosyltransferase	1	1,540	39	150	7	0.68	0.38
1tox	Diphtheria toxin	2	3,721	119	374	11	0.65	0.40
1trk	Transketolase	4	4,476	118	460	25	0.86	0.36
1uby	Farnesyl diphosphate synthase	1	2,168	53	215	6	0.73	0.31
1utg	Uteroglobin	1	1,485	40	135	0	0.82	0.23
1vfr	NAD(P)\:FMN oxidoreductase	2	3,431	91	337	25	0.78	0.34
1vok	TATA-box-binding protein	2	1,577	47	158	5	0.69	0.32
1wtl	Bence-Jones protein	1	698	21	80	2	0.82	0.26
1xso	Cu, Zn superoxide dismutase	1	662	21	74	2	0.87	0.24
2arc	Ara operon regulatory protein	1	765	18	86	1	0.80	0.43
2ccy	Cytochrome C’	1	792	20	80	0	0.85	0.28
2hdh	Hydroxyacyl-CoA dehydrogenase	1	1,524	40	154	3	0.73	0.27
2ilk	Interleukin-10	2	4,542	96	442	15	0.81	0.24
2lig	Aspartate receptor	2	1,686	47	183	6	0.73	0.39
2mcg	IgG lambda light chain dimer (Mcg)	2	1,646	57	177	9	0.55	0.39
2nac	NAD-dept formate dehydrogenase	2	3,789	98	387	19	0.87	0.35
2ohx	Alcohol dehydrogenase	1	1,718	49	178	8	0.84	0.36
2spc	Spectrin (one repeat unit)	3	2,508	59	237	11	0.76	0.28
2sqc	Squalene-hopene cyclase	1	809	23	77	2	0.64	0.36
2tct	Tetracycline repressor	2	2,675	63	273	9	0.75	0.27
2tgi	Transforming growth factor-beta2	2	1,262	37	124	2	0.73	0.27
3dap	Diaminopimelic acid dehydrogenase	2	2,661	71	271	13	0.81	0.27
3grs	Glutathione reductase	2	3,302	87	331	14	0.82	0.32
3sdh	Hemoglobin	1	873	25	87	3	0.72	0.45

TABLE II. (Continued)

PDB entry	Protein	Patches	Interface[a] *B*/2 (Å^2)	Residues	Atoms	H-bonds	Area fraction In core[b]	Polar[c]
3ssi	Streptomyces subtilisin inhibitor	1	866	31	89	2	0.80	0.32
4cha	Alpha-chymotrypsin	1	1,026	33	107	5	0.51	0.39
4kbp	Purple acid phosphatase	1	1,478	42	143	2	0.64	0.30
5csm	Chorismate mutase	3	2,007	57	203	5	0.62	0.33
5rub	RUBISCO	3	2,859	82	291	15	0.74	0.37
8prk	Inorganic pyrophosphatase	1	969	23	110	3	0.75	0.25
9wga	Wheat germ agglutinin	1	2,293	67	238	14	0.74	0.41
All		Average	1,940	52	198	9.0	0.77	0.35
	122	s.d.	1,100	29	111	6.7	0.10	0.07
1 patch		Average	1,370	37	140	6.0	0.77	0.35
	70	s.d.	620	16	62	4.6	0.09	0.07
2 patch		Average	2,380	63	243	11.3	0.76	0.34
	35	s.d.	890	22	88	6.5	0.10	0.07
3–6 patches		Average	3,365	92	346	18.5	0.78	0.38
	17	s.d.	1,380	34	136	10.3	0.14	0.05

[a]Due to twofold symmetry, each interface contains twice the interface area *B*/2, and twice (or very close to twice if the symmetry is noncrystallographic), the number of patches, interface atoms, residues, and H-bonds cited here per subunit.
[b]Fraction of the interface area contributed by core residues; rim residues contribute the complement to 1.
[c]Fraction of the interface area contributed by polar (N, O, and S containing) groups; nonpolar (carbon-containing) group residues contribute the complement to 1.

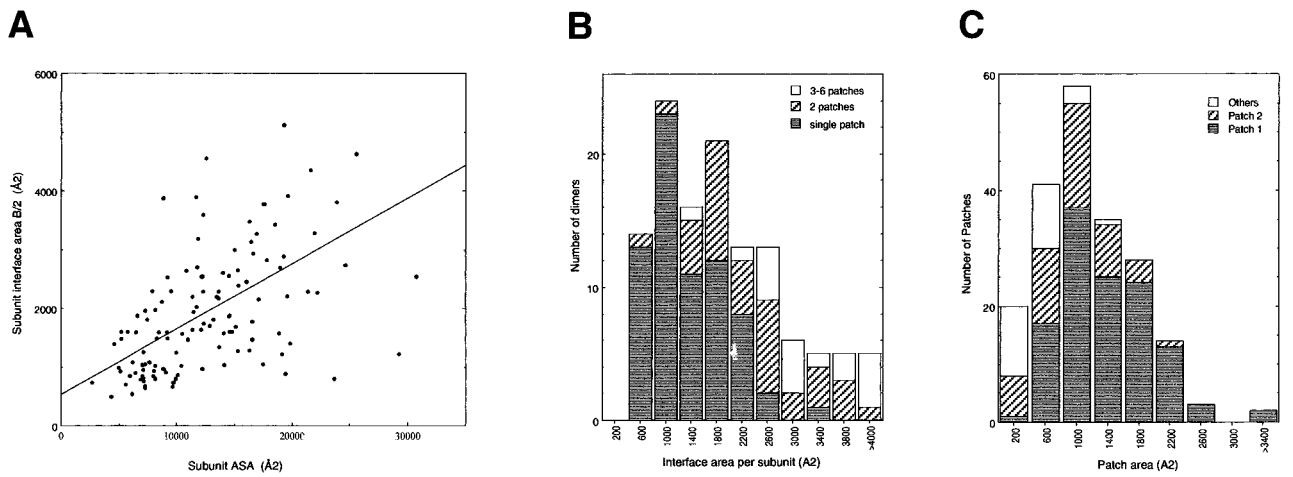

Fig. 1. Subunit surface, interface, and patch areas. **A:** The area of the 122 subunit interfaces in homodimers is plotted against the solvent accessible area of the free subunit. The regression line has a slope 0.11 and $R^2 = 0.34$. **B:** Histogram of the subunit interface areas; 70 of the 122 interfaces are single-patch; 35 have two patches, 17 have 3 or more. All areas are duplicated by symmetry. The labels mark the middle of the range in each column. **C:** Patch areas. Patch 1 includes the patches of single-patch interfaces and the largest patch in multipatch interfaces. Patch 2 is the second patch by size in a multipatch interface.

The range of *B*/2 extends from 500 Å^2 to above 7,000 Å^2 . Although the smaller proteins obviously cannot form very large interfaces, the correlation with size is mediocre ($R^2 = 0.34$). Interfaces bury 16% of the subunit surface on average, but this fraction varies from 3 to 44% in our sample. The histogram of *B*/2 values [Fig. 1(B)] is skewed, so that the median (1,700 Å^2) is less than the mean (1,940 Å^2, Table II). In protein–protein complexes,[4] *B*/2 has a mean of 970 Å^2 and a range of 570–2,330 Å^2. Thus, the average subunit interface in homodimers buries twice more protein surface than in complexes, and it can reach sizes that are never observed in nonpermanent assemblies. Copper amine oxidase[12] (1oac) has the largest interface in our sample. It buries 14,300 Å^2 (7,150 Å^2 per subunit) representing 22% of solvent accessible surface area of this large (160 kDa) protein. At the low end of the distribution, two small homodimers have an interface burying about 500 Å^2 per subunit, less than any interface found in complexes: S-lectin (1slt) and kappa-bungarotoxin (1kba). Kappa-bungarotoxin is listed as a dodecamer in PQS, and Ponstingl et al.[11] note that the dimer interface described by the authors of the X-ray structure[13] is smaller than other pairwise interfaces generated by the crystal symmetry. However, NMR and ultracentrifugation data[14] favor the existence in solution of the proposed dimer, which we retained in spite of its anomalously small interface.

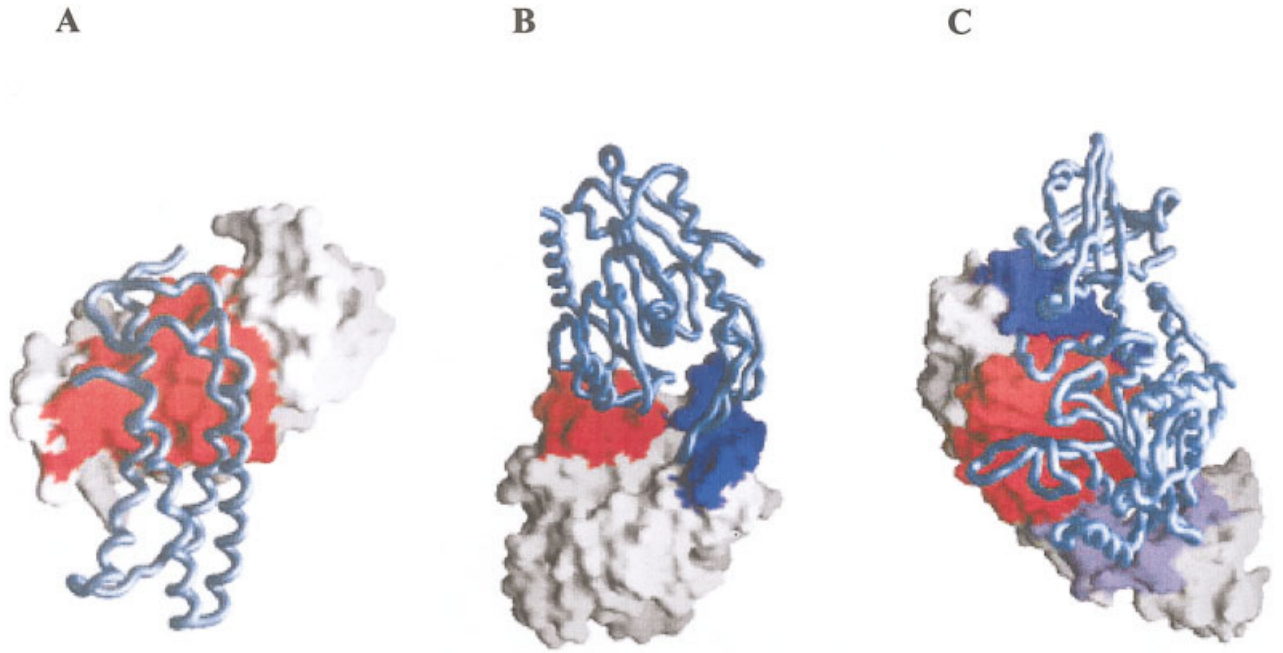

Fig. 2. Recognition patches within subunit interfaces. **A:** The single-patch interface of the cytochrome *c′* dimer (2ccy).[15] **B:** The interface of the carboxylesterase dimer[16] (1auo) contains two recognition patches formed by two separate domains. **C:** That of aspartyl-tRNA synthetase[17] (1b8a) has three patches; the magenta and red patches involve the larger catalytic domain, and the blue patch, the smaller anticodon-binding domain. Patches are colored on the molecular surface of one subunit; the other subunit being imaged as a backbone tube. Drawn with GRASP.[40]

Recognition Patches

Within each of the subunit interfaces, patches were identified by the geometric clustering algorithm, which we have already applied to protein–protein complexes.[5] Of the 122 interfaces, 70 formed a single patch; 35 had two patches, 17 had 3 to 6 patches [Fig. 1(B,C)]. Figure 2 compares the single patch interface of cytochrome *c′* (2ccy)[15] to the two-patch interface of carboxylesterase (1auo)[16] and the three-patch interface of aspartyl-tRNA synthetase (1b8a).[17] The carboxylesterase subunit interface is smaller than most other two-patch interfaces, or than the cytochrome *c′* interface. Nevertheless, it involves two structural domains within the protein subunits, and each domain contributes a patch. The aspartyl-tRNA synthetase interface, which is very extensive, contains a patch contributed by the anticodon-binding domain, and two others, by the catalytic domain.

The number of patches is correlated to the size of the interface. This can be seen in the histogram of Figure 1(B), where all but one of the 70 single-patch interfaces have $B/2 < 2{,}800$ Å^2 and all but two of the 52 multipatch interfaces have $B/2 > 1{,}500$ Å^2. For comparison, Figure 1(C) is a histogram of the area of the patches. Patch areas have the same distribution in single-patch interfaces and for the largest patch of multipatch interface. Consequently, the two types are lumped together as Patch 1 in Figure 1(C). The area histogram of Patch 1 has a peak near 1,000 Å^2, but several large patches cover more than 2,000 Å^2, and the mean value of the Patch 1 area is about 1,400 Å^2 (Table II). The next largest patch in multipatch interfaces is noted Patch 2 in Figure 1(C). The mean and peak values of the Patch 2 area distribution are both near 1,000 Å^2. In these two categories, 65% of the patches have areas in the range 1,100 ± 500 Å^2. Other patches in multipatch interfaces are generally small, with $B/2 < 600$ Å^2.

Polar Interactions and the Hydrophobic Character of the Interfaces

On average, 18% of the protein surface that is buried at homodimer interfaces belongs to main chain atoms, 82% to side chain atoms. These interfaces are 35% polar and 65% nonpolar (Table II), counting all carbon-containing groups as nonpolar, nitrogen, oxygen, and sulphur as polar. The nonpolar/polar ratio becomes 66/34 instead of 65/35 if we count the interface atoms instead of their contribution to *B/2*. In comparison, the nonpolar/polar area ratio is 56/44 for the average interface in complexes,[4] and 57/43 for the average solvent accessible surface of small globular proteins.[18] A histogram of the contribution of polar groups to the subunit interface area in the 122 dimers is shown in Figure 3. About one-quarter of the subunit interfaces are 40–50% polar, which makes their chemical composition similar to that of the interfaces in complexes, and also to the protein accessible surface. Another quarter is only 19–30% polar, and therefore hydrophobic. The chemical composition of these subunit interfaces resembles that of the surface buried between secondary structure elements inside proteins, which is very different from the protein surface.[18] The remainder, about half of the homodimer interfaces, is in between these two extremes. There is no systematic tendency for the nonpolar/polar area ratio to change with the size of the interface or with the number of

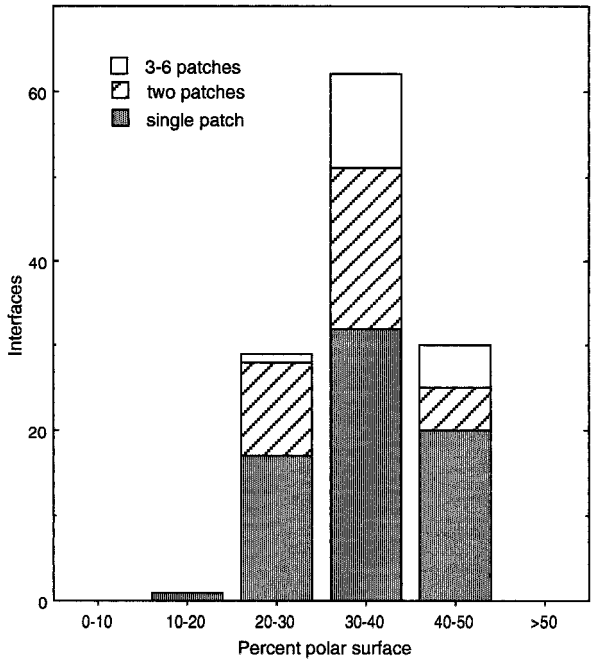

Fig. 3. The polar/non-polar character of the interfaces. Histogram of the contribution of polar (N, O, and S containing) groups to the interfaces in the 122 homodimers, expressed as a fraction of the interface area.

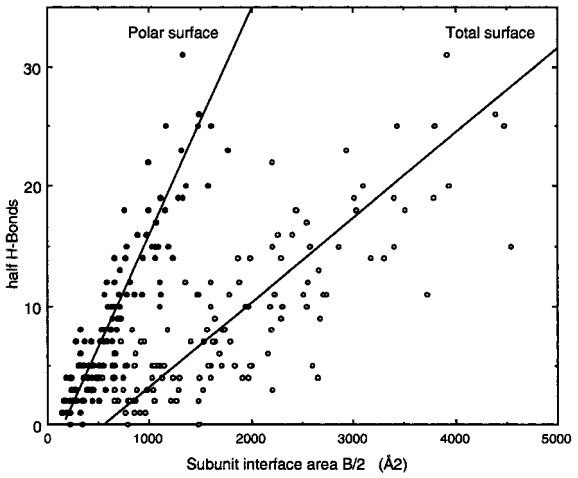

Fig. 4. Polar interactions. The number of interface H-bonds per subunit (half the total number of H-bonds present at the interface) is plotted against *B/2*, the subunit interface area, or against the polar surface component of *B/2*. The correlation coefficient R^2 is 0.75 with the subunit interface area, 0.83 with the polar surface area.

patches, but the spread of individual values around the mean is less in the larger interfaces, presumably due to the better sampling. The most polar interface (50% polar in 6-phosphofructose-2-kinase/bisphosphatase,[19] 1bif) and the least polar one (19% polar in thymidylate kinase,[20] 1e98) are both relatively small.

Hydrogen bond interactions between dimer subunits were detected using the program HBPLUS[21] with default parameters. On average, there are 9 such unique interactions per subunit (Table II), and 18 per interface due to the twofold symmetry. Although main chain atoms contribute only 18% of the interface area, they are implicated in a majority (61%) of the polar interactions, either in main chain/main chain (21%) or in side chain/main chain (40%) H-bonds; only 38% are side chain/side chain H-bonds. The number of H-bonds varies widely from one interface to another depending on its size and its hydrophobicity. Two interfaces contain no H-bond at all: in uteroglobin[22] (1utg) and cytochrome c' (2ccy).[15] These are among the most hydrophobic interfaces in our set, with only 23 and 28% polar area, respectively. At the other end of the spectrum, the huge interface of the copper amine oxidase dimer (1oac)[12] contains 96 H-bonds. Figure 4 shows that the number of H-bonds correlates weakly with the interface area. There is an average of one bond per 210 Å^2 with a correlation coefficient $R^2 = 0.75$. The correlation is better ($R^2 = 0.83$) with the polar surface area, which accounts for the different hydrophobic character of individual interfaces as well as their size. On average, there is one bond per 75 Å^2 of polar surface. Protein-protein complexes have an average of one H-bond per 170 Å^2 of interface area[4] equivalent to 75 Å^2 of polar surface, these interfaces being 44% polar on average. Thus, the density of H-bonds per unit interface area is lower in homodimers than in complexes, but the density per unit *polar* area is the same.

Amino Acid Composition

The average dimer interface contains 198 atoms from 52 residues in each subunit (Table II). The number of interface atoms and residues in each homodimer is linearly correlated to the interface area, with a slope of 37 Å^2 per residue ($R^2 = 0.98$), and 9.8 Å^2 per atom ($R^2 = 0.997$). Very similar values of the interface area per atom and per residue are observed in protein–protein complexes.[4,5] Table III gives the amino acid composition of the interfaces, evaluated either as the fractional number of residues of each type that lose accessible surface area, or as their fractional contribution to the interface area. The latter values take into account the different sizes of the side chains as well as their relative tendency to be at an interface. The average composition of the subunit interfaces is compared to that of the solvent accessible surface of the 122 homodimers, and that of the interfaces in protein–protein complexes.[3–5]

The most abundant residue at homodimer interfaces is leucine, which contributes about 10% of the buried surface area. The interfaces are enriched almost twofold in Leu relative to either the solvent accessible protein surface, or the interfaces in complexes. This being true of other aliphatic residues, Leu, Ile, Val, and Met together contribute 25% of the dimer interface area, whereas they form only 12% of the protein surface, and 17% of the interfaces in complexes. The two types of interfaces are also enriched in the aromatic residues: Phe, Tyr, and Trp are more abundant by a factor of 2.5 to 3 than on the protein surface. Neutral polar residues distribute equally between the surface and the interfaces. In contrast, the latter are depleted by a factor of about 2 in charged residues Asp, Glu, and Lys. Together, these three residues contribute

TABLE III. Amino Acid Composition of Homodimer Interfaces[†]

	% Number				% Area				Propensities[a]			
Type	Surface	Interface	Core	Rim	Surface	Interface	Core	Rim	Core	Rim	Lo Conte et al.[4]	Jones and Thomton[25]
Ala	7.6	7.4	7.6	7.1	5.2	5.2	5.2	5.0	0.00	−0.04	−0.43	−0.17
Arg	6.5	6.4	6.8	5.8	9.8	9.7	9.8	9.4	0.00	−0.04	0.13	0.27
Asn	5.0	4.7	3.9	5.8	5.1	3.8	3.5	4.8	−0.38	−0.06	−0.12	0.12
Asp	7.5	5.5	4.4	7.1	8.2	4.3	3.8	6.1	−0.77	−0.30	−0.31	−0.38
Cys	1.0	1.4	1.6	1.3	0.5	1.2	1.2	1.5	0.88	1.10	0.76	0.43
Gln	4.7	4.0	3.7	4.5	6.1	4.1	3.9	4.7	−0.45	−0.26	−0.36	−0.11
Glu	8.8	6.2	5.2	7.7	11.9	5.6	5.2	6.7	−0.83	−0.57	−0.47	−0.13
Gly	7.5	6.1	5.3	7.3	4.4	3.5	3.0	5.4	−0.38	0.20	0.02	−0.07
His	2.6	3.2	3.4	2.9	2.6	3.3	3.5	2.5	0.30	−0.04	0.64	0.41
Ile	3.5	4.9	6.0	3.3	2.4	5.5	5.9	3.9	0.90	0.49	0.56	0.44
Leu	6.1	9.0	10.6	6.5	4.4	10.4	11.2	7.9	0.93	0.59	0.29	0.40
Lys	8.1	5.4	3.7	8.1	13.2	5.4	4.1	9.7	−1.17	−0.31	−0.57	−0.36
Met	1.6	2.9	3.4	2.3	1.5	4.0	4.1	3.8	1.00	0.93	0.98	0.66
Phe	2.8	4.6	6.0	2.4	2.0	6.3	7.5	2.4	1.32	0.18	0.79	0.82
Pro	5.7	5.2	4.3	6.5	5.9	4.9	4.4	6.7	−0.29	0.13	−0.25	−0.25
Ser	5.9	5.6	5.3	6.1	4.7	4.0	3.7	4.9	−0.24	0.04	−0.42	−0.33
Thr	5.7	5.4	5.0	5.8	4.8	4.3	4.1	4.8	−0.16	0.00	−0.35	−0.18
Trp	1.2	1.7	2.0	1.2	1.1	2.6	3.0	1.3	1.00	0.17	1.25	0.83
Tyr	3.3	4.6	5.5	3.2	2.9	6.3	7.3	3.1	0.92	0.07	1.04	0.66
Val	4.8	6.0	6.4	5.3	3.3	5.6	5.6	5.5	0.53	0.51	0.09	0.27

[†]The compositions are calculated on the basis of either the number of residues or the area contributed to the solvent accessible protein surface and to the subunit interface in the 122 homodimers. Core residues contain at least one fully buried atom, rim residues have only accessible atoms. The sample contains 3,844 core and 2,518 rim residues.

[a]The propensity of a residue to be part of the core or the rim of a dimer interface is p = In (f/f°), where f is the area-based composition of the core or rim, f°, the area-based composition of the accessible surface of the homodimers cited in column "surface". Data for the interfaces in complexes are from table 4 of Lo Conte et al.[4] and table 2 of Jones and Thornton[25] the latter also including oligomeric proteins.

33% of the accessible surface area of the homodimers, but only 15% of their interface area, and 18% in complexes.

Remarkably, arginine is not excluded from interfaces in spite of its charge. Arg contributes about 10% of the accessible surface area, of the surface buried at dimer interfaces and the surface buried in complexes. Arg is the largest single contributor to the interfaces in complexes, the second largest in homodimers after Leu, but it ranks after Lys and Glu on the protein surface. Its guanidinium group is the prime donor of interface H-bonds: on average, 2.5 H-bonds per subunit (5 per homodimer), implicate Arg side chains. The H-bonds acceptors are main chain carbonyl groups (43% of cases), Asp/Glu carboxylates (43%), or other side chains (14%). Thus, Arg-Asp/Glu salt bridges are relatively numerous (2.2 per homodimer), but not more frequent than bonds to the main chain. The abundance of arginine at interfaces was also seen in protein–protein complexes.[4] It constitutes a major difference between the surfaces that are buried upon association or dimerization, and those that are buried inside proteins upon folding, as arginine is even more severely excluded from the protein interior than the other charged residues.

Core and Rim of Subunit Interfaces

As we did for protein–protein complexes,[5] we split the interfaces of the homodimers into a core and a rim. The core contains residues with one or more interface atoms that are fully buried in the dimer, the rim, residues in which all interface atoms have residual accessibility to the solvent. Figure 5 illustrates the core and rim of the subunit interfaces in enolase (1ebh)[23] and hypoxanthine phosphoribosyltransferase (1tc1),[24] two homodimers that have rather large single-patch interfaces. Typically, rim residues surround the core residues buried at the center. Rim residues contribute one-quarter of the area of the enolase interface, core residues, three-quarters. This is close to the average in our sample, where the interface areas are distributed for 23% in the rim, 77% in the core (Table II). However, the distribution of the area between rim and core depends on the size and the shape of the interface, and it varies widely from one dimer to another.

On average, an interface involves some 52 residues per subunit; the core contains about 32, the rim, about 20, but the core residues contribute most of the buried surface area. Thus, the area-based amino acid composition of the interfaces is essentially that of their core (Table III). As already observed for oligomeric proteins in general,[2,3] the core of subunit interfaces is enriched in aliphatic and aromatic residues, depleted in charged residues (except for Arg), and very much resembles the protein interior. We quantify the similarity between the amino acid compositions of two surfaces or interfaces by calculating their Euclidean distance Δf:

$$(\Delta f)^2 = 1/19 \sum_{i=1 \text{ to } 20} (f_i - f'_i)^2 \qquad (2)$$

where f_i and f'_i are the percent contribution of amino acid type i to the surface or the interface area. The distance

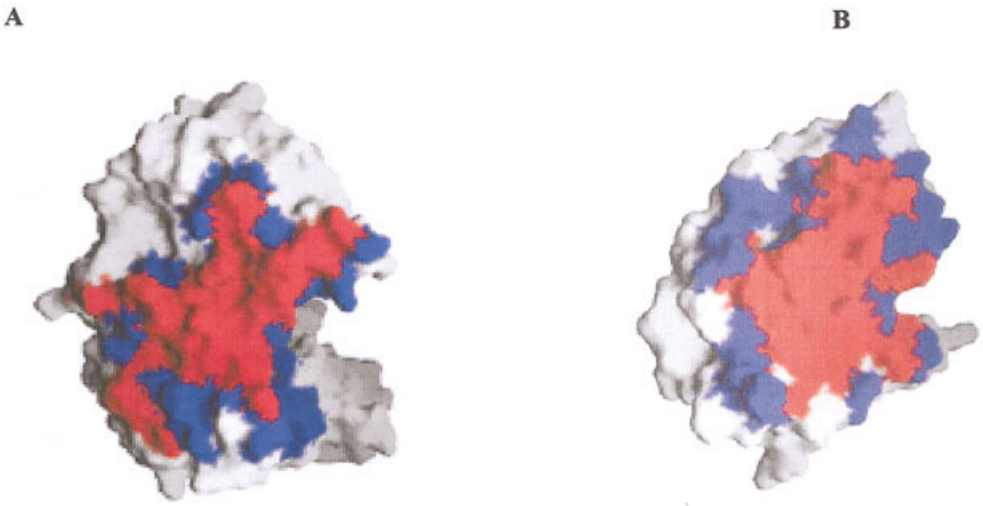

Fig. 5. The core and rim in a subunit interface. **A:** The surface of the enolase subunit[23] (1ebh) involved in the dimer contact is imaged. Core residues in red have one or more atoms with zero ASA in the dimer, whereas rim residues in blue remain solvent accessible. **B:** Similar features in the interface of hypoxanthine phosphoribosyltransferase[24] (1tc1). Drawn with GRASP.[38]

between the average compositions of the dimer interfaces and their core is small, only 0.6%. It is greater, 2.5%, between the core and the rim, and reaches 3.9% between the core and the protein surface [Fig. 6(A)]. In contrast, the distance between the core and the protein interior compositions is only 1.5%. These values confirm that, in homodimers, the surfaces that are buried when the polypeptide chains fold and assemble are similar in composition, and very different from those that remain solvent accessible.

DISCUSSION

The proteins analyzed here were selected as being homodimers in solution on the basis of the best available information. Most are permanent assemblies, but some may dissociate into monomers or form larger oligomers depending on concentration, pH, ligand binding, and other parameters that are known to modulate protein–protein interaction. When this happens, the dissociation constant of the dimer is very rarely known. Thus, we made no attempt to introduce thermodynamic stability as a criterion for choosing the proteins. Moreover, we identified subunit interfaces primarily on the basis of geometric criteria derived from the atomic coordinates. Ambiguous cases were encountered, and in one case at least, kappa-bungarotoxin, the geometric criteria conflicted with the literature. Then, we cannot exclude that site-directed mutagenesis or other biochemical methods will prove the selected interface to be incorrect. However, these cases were very few and in most of the PDB entries of Table I, the extensive contacts observed between polypeptide chains make it very likely that they assemble as dimers at the time of folding, and remain permanently so.

We observe obvious differences between interfaces in homodimers and in protein–protein complexes. They concern the size of the interfaces, which on average is twice larger in homodimers than in complexes, and their hydrophobic character, which is more pronounced in homodimers. Whilst a majority of the complexes form "standard size" interfaces[4] with *B/2* in the range 600–1,000 Å^2, 77% of the homodimers have *B/2* greater than 1,000 Å^2. Each component of a complex with a standard-size interface generally carries a single recognition patch, whilst larger interfaces are multipatch, and each patch typically covers 800 Å^2 of the protein surface.[5] The subunit interfaces of homodimers also contain patches, which tend to be more numerous than in complexes, and also larger: two-thirds of the Patches 1 and 2 in homodimers bury 1,100 ± 500 Å^2. These differences must relate to the modes of assembly. The protein surface that is buried at the interface of a protein–protein complex is in contact with water until the complex assembles, and it cannot carry large hydrophobic patches, which would make the protein insoluble. In contrast, burying hydrophobic patches favors the formation of a permanent assembly during both folding and dimerization. The core of the interfaces is their buried and most hydrophobic region. On average, it includes 32 residues per monomer in a dimer, a number to be compared with the 12 core residues per component of a complex with a standard-size interface.[5]

Nevertheless, the two types of interfaces have several features in common. For instance, they both contain about one H-bond per 75 Å^2 of polar interface area. Thus, dimer interfaces may be depleted of polar groups relative to interfaces in complexes, but the same fraction of these polar groups form polar interactions. More significantly, the cores of the interfaces in complexes and homodimers have similar amino acid compositions in spite of their different sizes. This can be seen in Table III, which lists the propensities of each residue type for being part of the core or the rim of an interface, rather than of the solvent accessible protein surface. The correlation between the sets of propensities is poor ($R^2 = 0.55$) between the core and the rim of dimer interfaces. It is much better (R^2 =

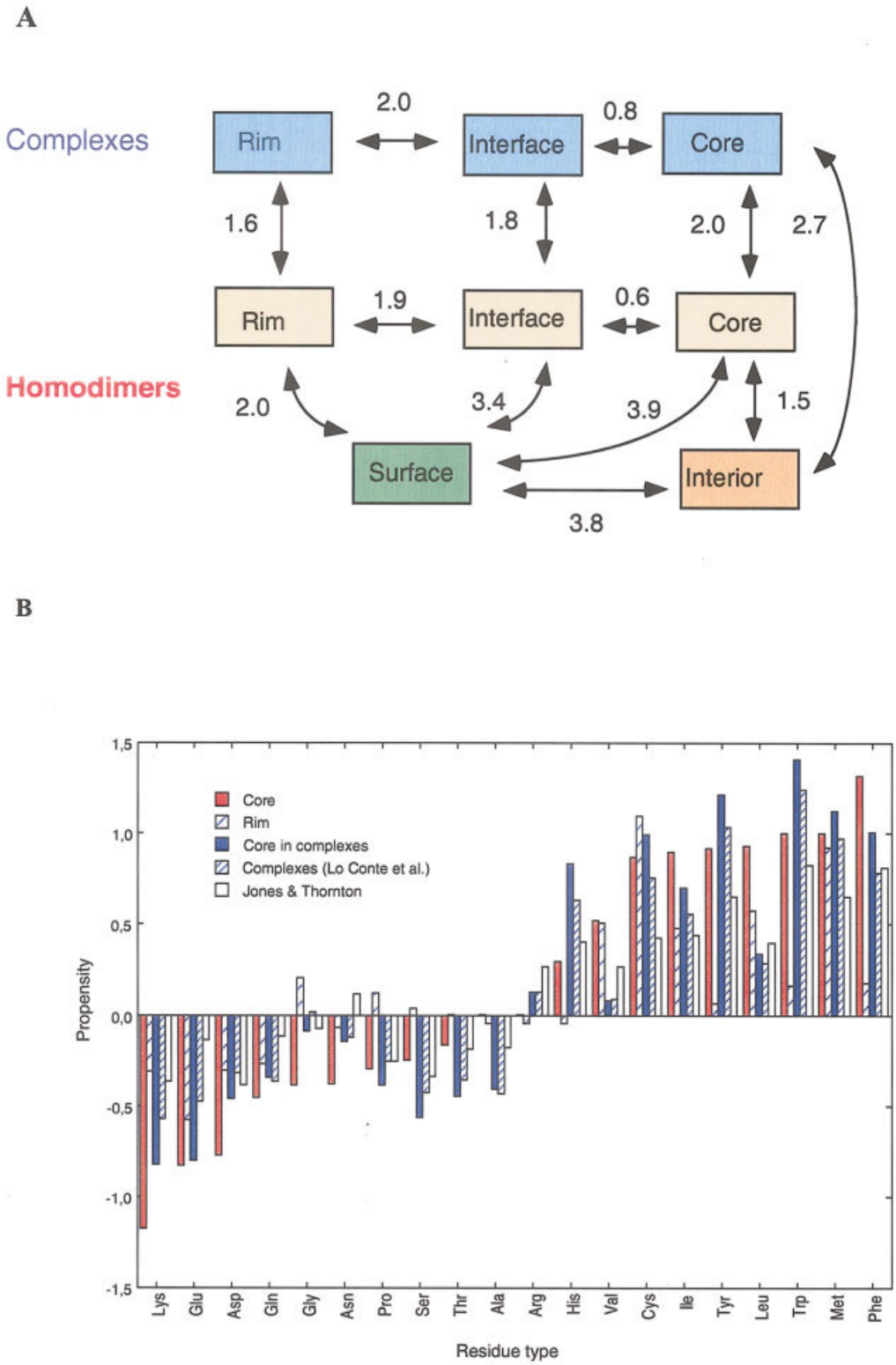

Fig. 6. Amino acid compositions and residue propensities for the interfaces. **A:** The Euclidean distances between area-based compositions are defined by Eq. 2 in the text. The average compositions of the interfaces and of the solvent accessible protein surface ('Surface') in homodimers are taken from Table III; those of the interfaces in complexes from Chakrabarti and Janin[5]; that of the protein surface buried upon folding ('Interior') is taken from Miller et al.[18] **B:** The propensity of a residue type to be part of an interface rather than of the protein surface is ln (f /f °) where f and f ° are the average area contributions of the residue to, respectively, the interface and the solvent accessible surface. The propensity for the core of interfaces in homodimers and in complexes[5] is compared to those for the whole interface in 75 complexes[4] and in a sample containing both complexes and oligomers.[25,26] The residue types are ranked in order of increasing propensities for the core of homodimer interfaces.

0.81) between the cores of the two types of interfaces. Figure 6(B) is a plot of different sets of propensities for interfaces vs. the protein surface. It shows the propensities to be more pronounced after the cores are distinguished from the bulk of the interfaces, and suggests that these particular regions of the subunit surface could be recognized on the basis of their amino acid composition.

Our results generally agree with those of Miller et al.[2] and Jones and Thornton[3,25,26] who studied smaller samples of homodimers with, respectively, 11 and 32 proteins. The remark that interfaces are generally much larger and more hydrophobic in homodimers than in protein–protein complexes, was made in all these studies, and their great heterogeneity was noted. The average number of H-bonds per unit interface area observed here (one per 210 $Å^2$), is the same as in Miller et al.[2] The amino acid residue preferences, expressed as propensities in Table III, are linearly correlated, but they are more strongly marked in our sample, which is larger and contains only homodimers. Relative to the propensities of Jones and Thornton,[25] which were derived from a sample that also contained protein–protein complexes, ours are twice greater, although the correlation is high ($R^2 = 0.83$).

Our analysis of the patches and the distinction between the rim and core of the interfaces, is novel for homodimers. Jones and Thornton[3,25] defined the "segmentation" of an interface in terms of the number of polypeptide chain segments that contribute to it, and they designed "surface patches" to cover the complete protein surface. Segments and surface patches differ in definition and purpose from the patches that our geometric clustering algorithm generates. All the larger interfaces have several of those, implicating distant parts of the polypeptide chains, often in separate domains. The distinction that we also made between the solvent accessible rim of an interface and its buried core, is justified by their different compositions. It gives a quantitative basis to a description of interfaces with a "defined hydrophobic core," which has been made by many authors.[27–30] This description applies to homodimers much better than to complexes or other nonpermanent assemblies.[28] Moreover, homodimer interfaces have a higher degree of residue conservation than elsewhere on the protein surface,[31,32] but the difference is marginal on whole interfaces, and the detection of conservation is also likely to benefit from the distinction between the rim and the core.

Several other studies[33–36] of protein–protein interfaces have used much larger samples of proteins, generated automatically from the PDB by assuming the content of the asymmetric unit to represent the protein molecule. These samples lack oligomers with crystal symmetry, and they include dimers that are crystal packing artifacts along with genuine dimers or protein–protein complexes. Crystal packing generally involves interfaces that are much smaller and less hydrophobic than in complexes or oligomers,[10,32,37–39] but there are exceptions, and, in the absence of biochemical information, the distinction between the two cannot be made with more than 85–90% confidence.[6,11,32] Nevertheless, studies based on automatically generated interfaces lead to conclusions concerning the preference of amino acids for interfaces,[34,36] the role of hydrophobicity,[30,34] and of polar interactions[35] that are qualitatively similar to ours. The better defined sample that we have assembled for this study enables us to describe properties of homodimers and dissect their interfaces, with confidence that our conclusions apply to biologically relevant protein–protein interactions.

REFERENCES

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
2. Miller S, Lesk AM, Janin J, Chothia C. The accessible surface area and stability of oligomeric proteins. Nature 1987;328:834–836.
3. Jones S, Thornton JM. Principles of protein–protein interactions. Proc Natl Acad Sci USA 1996;93:13–20.
4. Lo Conte L, Chothia C, Janin J. The atomic structure of protein–protein recognition sites. J Mol Biol 1999;285:2177–2198.
5. Chakrabarti P, Janin J. Dissecting protein–protein recognition sites. Proteins 2002;47:334–343.
6. Henrick K, Thornton JM. PQS: a protein quaternary structure file server. Trends Biochem Sci.1998;23:358–361.
7. Rodier F, Chiadmi M, Crosio MP. An array processor program for computing the neighbours in molecular packing. Acta Cryst. 1990;A46:37
8. Hubbard S. ACCESS: A program for calculating accessibilities. Department of Biochemistry and Molecular Biology. University College of London, 1992.
9. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–400.
10. Janin J, Rodier F. Protein-protein interaction at crystal contacts. Proteins 1995;23:580–587.
11. Ponstingl H, Henrick K, Thornton JM. Discriminating between homodimeric and monomeric proteins in the crystalline state. Proteins 2000;41:47–57.
12. Parsons MR, Convery MA, Wilmot CM, Yadav KD, Blakeley V, Corner AS, Phillips SE, McPherson MJ, Knowles PF. Crystal structure of a quinoenzyme: copper amine oxidase of *Escherichia coli* at 2 Å resolution. Structure 1995;3:1171–1184.
13. Dewan JC, Grant GA, Sacchettini JC. Crystal structure of κ-bungarotoxin at 2.3 Å resolution. Biochemistry 1994;33:13147–13154.
14. Oswald RE, Sutcliffe MJ, Bamberger M, Loring RH, Braswell E, Dobson CM. Solution structure of neuronal bungarotoxin determined by two-dimensional NMR spectroscopy: sequence-specific assignments, secondary structure, and dimer formation. Biochemistry 1991;30:4901–4909.
15. Finzel BC, Weber PC, Hardman KD, Salemme FR. Structure of ferricytochrome *c′* from *Rhodospirillum molischianum* at 1.67 Å resolution. J Mol Biol 1985;186:627–643.
16. Kim KK, Song HK, Shin DH, Hwang KY, Choe S, Yoo OJ, Suh SW. Crystal structure of carboxylesterase from *Pseudomonas fluorescens*, an alpha/beta hydrolase with broad substrate specificity. Structure 1997;5:1571–1584.
17. Schmitt E, Moulinier L, Fujiwara S, Imanaka T, Thierry J-C, Moras D. Crystal structure of aspartyl-tRNA synthetase from *Pyrococcus kodakaraensis* KOD: archaeon specificity and catalytic mechanism of adenylate formation. EMBO J 1998;17:5227–5237.
18. Miller S, Janin J, Lesk AM, Chothia C. Interior and surface of monomeric proteins. J Mol Biol 1987;196:641–656.
19. Hasemann CA, Istvan ES, Uyeda K, Deisenhofer J. The crystal structure of the bifunctional enzyme 6-phosphofructo-2-kinase/fructose-2-6-bisphosphatase reveals distinct domain homologies. Structure 1996;4:1017–1029.
20. Ostermann N, Lavie A, Padiyar S, Brundiers R, Veit T, Reinstein J, Goody RS, Konrad M, Schlichting I. Potentiating AZT activation: Structures of wild-type and mutant human thymidylate kinase suggest reasons for the mutants' improved kinetics with the HIV prodrug metabolite AZTMP. J Mol Biol 2000;304:43–53.
21. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J Mol Biol 1994;238:777–793.
22. Morize I, Surcouf E, Vaney MC, Epelboin Y, Buehner M, Fridlansky F, Milgrom E, Mornon JP. Refinement of the $C222_1$ crystal

form of oxidized uteroglobin at 1.34 Å resolution. J Mol Biol 1987;194:725–739.
23. Wedekind JE, Reed GH, Rayment I. Octahedral coordination at the high affinity metal site in enolase; crystallographic analysis of the Mg^{II}-enzyme complex from yeast at 1.9 Å resolution. Biochemistry 1995;34:4325–4330.
24. Focia PJ, Craig III SP, Nieves-Alicea R, Fletterick RJ, Eakin AE. A 1.4 Å crystal structure for the hypoxanthine phosphoribosyltransferase of *Trypanosoma cruzi*. Biochemistry 1998;37:15066–15075.
25. Jones S, Thornton JM. Analysis of protein–protein interaction sites using surface patches. J Mol Biol 1997;272:121–132.
26. Jones S, Thornton JM. Protein-protein interactions: a survey of protein dimer structures. Prog Biophys Mol Biol 1995;63:31–65.
27. Miller S. The structure of interfaces between subunits of dimeric and tetrameric proteins. Prot Eng 1989;3:77–83.
28. Larsen TA, Olson AJ, Goodsell DS. Morphology of protein–protein interfaces. Structure 1998; 6:421–427.
29. Goodsell DS, Olson AJ. Structural symmetry and protein function. Annu Rev Biophys Biomol Struct 2000;29:105–153.
30. Young L, Jernigan RL, Covell DG. A role for surface hydrophobicity in protein–protein recognition. Protein Sci 1994;20:320–329.
31. Valdar SJ, Thornton JM. Protein-protein interfaces: analysis of aminoacid conservation in homodimers. Proteins 2001;42:108–124.
32. Valdar SJ, Thornton JM. Conservation helps to identify biologically relevant crystal contacts. J Mol Biol 2001;399–416.
33. Tsai C-J, Lin SL, Wolfson HJ, Nussinov R. A dataset of protein–protein interfaces generated with a sequence-order independent comparison technique. J Mol Biol 1996;260:604–620.
34. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. Study of protein–protein interfaces: a statistical analysis of the hydrophobic effect. Protein Sci 1997;6:53–64.
35. Xu D, Tsai CJ, Nussinov R. Hydrogen bonds and salt bridges across protein–protein interfaces. Protein Eng 1997;10:999–1012.
36. Glaser F, Steinberg DM, Vakser IA, Ben-Tal N. Residue frequencies and pairing preferences at protein–protein interfaces. Proteins 2001;43:89–102.
37. Janin J. Principles of protein–protein recognition from structure to thermodynamics. Biochimie 1995;:497–505.
38. Janin J. Specific *vs* non-specific contacts in protein crystals. Nature Struct Biol 1997;4:973–974.
39. Dasgupta S, Iyer GH, Bryant SH, Lawrence CE, Bell JA. Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. Proteins 1997; 28:494–514.
40. Nicholls A, Sharp K, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–296.