PROTEINS: Structure, Function, and Genetics 23:580–587 (1995)

Protein–Protein Interaction at Crystal Contacts

Joel Janin and Francis Rodier
Laboratoire de Biologie Structurale, UMR 9920 CNRS-Université Paris-Sud, 91198 Gif-sur-Yvette, France

ABSTRACT **Packing contacts are crystal artifacts, yet they make use of the same forces that govern specific recognition in protein–protein complexes and oligomeric proteins. They provide examples of a nonspecific protein–protein interaction which can be compared to biologically relevant ones. We evaluate the number and size of pairwise interfaces in 152 crystal forms where the asymmetric unit contains a monomeric protein. In those crystal forms that have no element of 2-fold symmetry, we find that molecules form 8 to 10 pairwise interfaces. The total area of the surface buried on each molecule is large, up to 4400 Å^2. Pairwise interfaces bury 200–1200 Å^2, like interfaces generated at random in a computer simulation, and less than interfaces in protease–inhibitor or antigen–antibody complexes, which bury 1500 Å^2 or more. Thus, specific contacts occurring in such complexes extend over a larger surface than nonspecific ones. In crystal forms with 2-fold symmetry, pairwise interfaces are fewer and larger on average than in the absence of 2-fold symmetry. Some bury 1500–2500 Å^2, like interfaces in oligomeric proteins, and create "crystal oligomers" which may have formed in the solution before crystallizing.**

Key words: biological specificity, interface area, symmetry

INTRODUCTION

Protein–protein recognition has been given a structural basis by analyzing the crystal structure of protein–protein complexes such as protease–inhibitor or antigen–antibody complexes,[1–5] and of oligomeric proteins.[6] All are examples of highly stable and specific interactions. Crystal packing interactions, which have been carefully reviewed for organic molecules,[7] have been given much less attention with proteins, except as a possible source of perturbation of the molecular structure. Indeed, packing interactions are crystal artifacts: they play no part in function and are not subject to natural selection. Does their study make biological sense? We think it does for the following reasons. First, the forces that stabilize a protein crystal are the same that govern protein folding and protein–protein recognition. Then, analyzing specific contacts in protease–inhibitor or antigen–antibody complexes is not enough to give specificity a structural basis. We must have counterexamples, which crystal packing interactions provide in abundance, and make a comparison.

Crystal packings are diverse and complicated. Here, we take a first step by considering only the simplest case: crystal forms where the molecule is also the asymmetric unit. Ignoring for the moment electrostatic interactions and the chemical nature of the interfaces formed between molecules, we focus on their number and on their size, estimated from the protein surface area excluded from contact with the solvent.[8] We find systematic differences between crystal forms which do and do not have elements of 2-fold symmetry. When they do not, pairwise interfaces created by lattice translations or screw rotations are systematically smaller and more numerous than when they do. Pairwise interfaces are significantly less extensive than in specific complexes, and more like interfaces generated in computer simulations by randomly tossing protein molecules against each other. In contrast, crystal forms with 2-fold symmetry show examples of large pairwise interfaces resembling those of oligomeric proteins. They contain "crystal oligomers" which may have assembled in the crystallization solution. Conclusions of this study are therefore relevant to the mechanism of crystallization[9] as well as to biological specificity.

METHODS

Atomic coordinates were obtained from the Brookhaven Protein Data Bank.[10] Data sets at better than 2.5 Å resolution and with one molecule in the asymmetric unit were selected by excluding all macromolecular complexes and proteins known to be oligomers in solution. Several nonisomorphous crystal forms of the same protein may satisfy this criterion. They yield independent information and were included as separate entries. When isomorphous or closely related crystal forms of the same protein were present, only one example was kept, usually the one determined at the highest resolu-

Received April 4, 1995; revision accepted June 23, 1995.
Address reprint requests to Joel Janin, Laboratoire de Biologie Structurale, UMR 9920 CNRS-Université Paris-Sud, 91198 Gif-sur-Yvette, France.

tion. Solvent molecules were deleted. The following 152 files were retained:

1AAJ	1AAK	1ABE	1ACX	1ALC	1APH	1APW	1ARB	1ARP	1AST
1AYH	1BBC	1BFG	1BGC	1BLC	1BNH	1BP2	1CA2	1CAA	1CBN
1CCR	1CGT	1CLL	1CLM	1CMS	1CTF	1CUS	1CY3	1DMB	1DRI
1ECO	1END	1ESA	1EZM	1FDX	1FGA	1FLP	1FNR	1FUS	1FX1
1GCR	1GCT	1GOF	1GPR	1HBQ	1HEL	1HIP	1HPI	1HYP	1LIB
1LZ1	1LZT	1MBO	1NAR	1NOA	1NTP	1OFV	1OMP	1PAZ	1PCY
1PDA	1PGX	1PHP	1PK4	1POA	1PPG	1PPO	1PPT	1PTF	1RAR
1RAY	1RCB	1RDG	1REC	1RHD	1RNH	1RNS	1RNT	1RRO	1RSM
1RTC	1SBP	1SGT	1SHA	1SHG	1SIM	1ST3	1STN	1TCA	1TEN
1TGT	1THM	1TLD	1TML	1TON	1TOP	1TPP	1UBQ	2ACH	2ACT
2ALP	2CCP	2CTX	2CYP	2FGF	2GN5	2IHL	2LHB	2LZM	2MCM
2PRK	2RN2	2SGA	2SN3	351C	3ADK	3B5C	3C2C	3CHY	3COX
3EBX	3PSG	3TLN	4APE	4BP2	4CLN	4FD1	4FGF	4FXN	4I1B
4ICB	4MT2	4OVO	4PEP	4PTI	5ACN	5CPV	5CYT	5PEP	5PTI
6APR	6CPA	6PTI	6RXN	7RSA	821P	8DFR	8EST	8GCH	8PTI
8RXN	9PAP								

Crystal packings were produced with an improved version of program SYMVOI[11] running under UNIX and renamed CRISPACK. The program can be obtained by applying at the e-mail address rodier@cygne.lbs.cnrs-gif.fr. It generates copies of the reference molecule by applying space group symmetry operations and lattice translations, then tests the copies for contacts shorter than 4 Å with atoms of the reference molecule. Steric clashes were flagged. Some packings had too many, indicating that either the lattice parameters or the orthogonalization matrices taken from file headers were in error. These were checked with authors of the X-ray structures and corrected if needed.

Accessible surface areas (ASA) were calculated for individual molecules and for all pairs of molecules in contact. We used program ASA from Pr. A. Lesk (Cambridge) which implements the Shrake-Rupley dot algorithm,[12] with a probe radius of 1.4 Å. The molecule's ASA being A_1 and that of the pair A_2, the pairwise interface area B_{int} was taken to be to $2A_1-A_2$. B_{int} is distributed over the two molecules. B_{tot}, the total area of the buried surface of one molecule in the crystal, was derived as one-half of the sum of its pairwise interface areas. If two interfaces overlap, this overestimates the area of the surface actually buried. We checked for overlaps in test cases and found their effect to be negligible.

RESULTS

Protein–Protein Interfaces in Crystals

In a crystal, each protein molecule makes several contacts, which are different for different proteins or, for the same protein, in different crystal forms. The sample of protein–protein interaction at crystal contacts is therefore very large. For simplicity, we consider here only monomeric proteins and crystal forms having one molecule in the asymmetric unit. With more than one molecule in the asymmetric unit, noncrystallographic symmetries combine with crystallographic ones, and the packing can be very complex. Here, neighboring molecules are related by either a lattice translation, a point group symmetry, or a screw rotation. Each has N_{int} neighbors and forms N_{int} different interfaces. In the two-dimensional crystal of Figure 1, N_{int} is 5. Interface *A:a* is created by 2-fold symmetry about axis *v* and it involves the same set of atoms on both sides. Two interfaces *B:c* and *C:b* are created by the screw rotation along axis *v'*. They are identical, yet they involve two distinct regions *B* and *C* on the reference molecule. Similarly, the lattice translation along *v* creates a pair of interfaces *D:e* and *E:d*. Contacts like *A:a* have been called *isologous* by Monod et al.,[13] all others being *heterologous*.

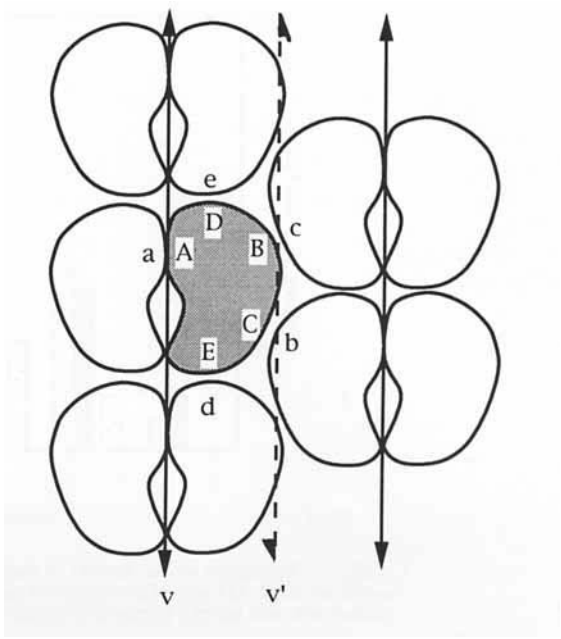

Fig. 1. Two-dimensional crystal packing. *v* is a 2-fold axis of symmetry, *v'*, a screw 2-fold. The reference molecule (shaded) makes one isologous contact with the 2-fold-related molecule on the left, creating interface *A:a;* and four heterologous contacts: two on the right with molecules related by the screw 2-fold, forming interfaces *B:c* and *C:b;* two on top and bottom with molecules related by the lattice translation along *v,* forming interfaces *D:e* and *E:d.*

Heterologous contacts due to lattice translations or screw symmetry are ubiquitous, whereas those resulting from 3-, 4-, and 6-fold point group symmetries are rare. Isologous contacts are also uncommon, even though most space groups available to proteins have elements of 2-fold symmetry. Such space groups are a minority in our sample: 58 crystal forms out of a total of 152, or 38%. In keeping with previous observations,[14,15] space groups lacking point-group symmetry are overrepresented; 87 crystal forms (57% of the sample) belong to three space groups only: triclinic $P1$, monoclinic $P2_1$, and orthorhombic $P2_12_12_1$. In these space groups and others that lack 2-fold symmetry, the number N_{int} of interfaces must be even. In general, N_{int} is broadly distributed between 4 and 14 (Fig. 2). The mean is

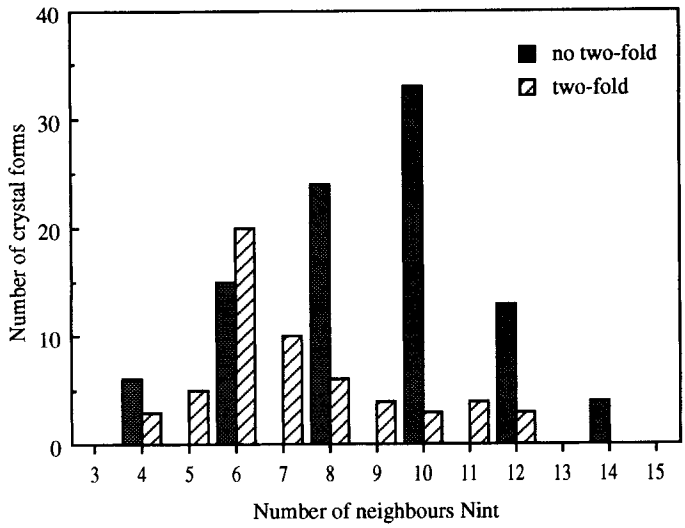

Fig. 2. Histogram of the number of neighbors. N_{int} is the number of neighbors the reference molecule has in 152 crystal forms containing a monomeric protein in their asymmetric units. Space groups with and without elements of 2-fold symmetry are shown separately.

8.3 in all crystal forms, 9.0 in crystal forms with no 2-fold, 7.2 in those with. N_{int} is either 8 or 10 in 60% of the crystal forms with no 2-fold, and either 6 or 7 in 50% of those with. Thus, each molecule has about two neighbors less on average in the second case.

The size of the pairwise interface between two neighboring molecules can be evaluated as the number of van der Waals contacts, or, more conveniently, as the area B_{int} of the protein surface that is buried on both partners when the pair assembles. B_{int} is approximately proportional to the number of atom pairs in contact and also to the number of amino acid residues involved (approximately one residue per 30 Å^2). There are 1320 pairwise interfaces in our sample. The mean value of their area is 570 Å^2, which implies that they cover about 280 Å^2 on the surface of each partner molecule. A typical pairwise interface contains 30 van der Waals contacts and involves about 10 amino acid residues of each partner. However, these numbers are only averages and pairwise interfaces can vary in size by at least an order of magnitude (Fig. 3). The bulk is in the range 200–1200 Å^2, yet some involve a single amino acid residue on each partner molecule and bury 50 Å^2, whereas others involve 100 residues and bury over 2000 Å^2.

Figure 3 compares pairwise interfaces to those of specific protein–protein complexes,[16] and to interfaces created at random by tossing two protein molecules against each other in a computer experiment.[17] Interfaces in complexes bury 1250–1900 Å^2, and they are much more homogeneous in size than the other two kinds. Random interfaces cover 200–1200 Å^2, the same range as for pairwise interfaces in crystals. This concordance of range fits with the idea that crystal contacts make use of randomly selected regions of the protein surface, a conclusion already reached in the case of pancreatic ribonuclease. This protein crystallizes in many different crystal forms,[18,19] and, taken together, the pairwise interfaces found in these crystals cover essentially the whole protein.[20]

In a crystal, there are always several pairwise interfaces, and they obey space group symmetry. The area B_{tot} of the surface buried on the reference molecule is half the sum of the pairwise interface areas. The other half is distributed among its neighbors, thus, the area of the protein surface buried when one molecule is incorporated in a crystal is actually $2B_{tot}$. In our sample, B_{tot} varies from 1100 to 4400 Å^2. The smallest value is for orthorhombic plasminogen kringle crystals (1PK4), where each molecule forms four small interfaces. The largest is for trigonal troponin C (1TOP), which forms 12 interfaces, including two large ones along a screw 3-fold. Unlike the accessible surface area A of monomeric proteins, which is a simple function of their molecular weight M,[21,22] the surface area buried in crystal contacts does not correlate with the protein size. A correlation has been described[21] between M and the area of the surface remaining accessible, which is $A - B_{tot}$. We attribute it to the correlation between M and A, since B_{tot} can change for the same protein in different crystal forms: by almost a factor of two in the case of pancreatic ribonuclease.[20] In our sample, the fraction B_{tot}/A of the protein surface that becomes buried in a crystal is as low as 8% in orthorhombic cyclodextrin glycosyltransferase (1CGT), and as high as 78% in the avian pancreatic peptide (1PPT). In the majority of crystal forms, B_{tot}/A is in

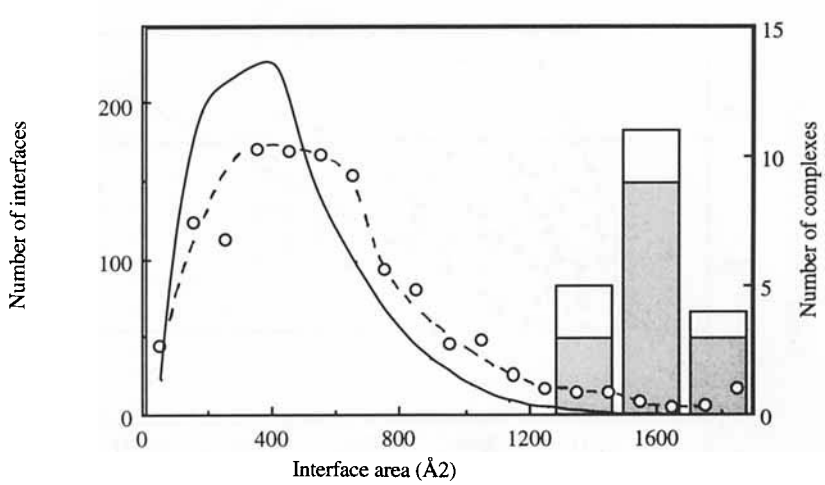

Fig. 3. Distribution of interface areas. (○) Area B_{int} of 1320 pairwise interfaces in 152 crystal forms measured with the Shrake–Rupley algorithm.[12] The dashed line is drawn to guide the eye and help in the comparison with the histogram of interface areas found in 15 protease–inhibitor (stripes) and 5 antigen–antibody (empty box) complexes, and also with random interfaces (full line). Data for complexes are taken from Janin.[16] Random interfaces were formed in a computer experiment by docking hen lysozyme onto the combining site of antilysozyme antibody HyHEL-5 in all possible orientations. The number of dockings yielding an interface of given area B has been scaled vertically to yield the same integrated total as for pairwise interfaces in crystals, and horizontally to correct for a systematic error of about 20% which affects interface areas calculated by an analytical approximation used for efficiency in docking, rather than with the Shrake–Rupley algorithm. Very similar results were obtained with other pairs of proteins.[17]

the range 15–60% (Fig. 4). The mean value is 32% for all proteins of our sample. It decreases from 40% for the smaller proteins (M less than 15,000) to 20% for the larger ones (M over 25,000). This strongly suggests that there is a minimum, but not necessarily a maximum, for the area of the protein surface that must be buried to make the crystal stable, and that the minimum does not depend on the protein size.

Crystal Oligomers

The buried surface per molecule is large in all protein crystals, but, as it is spread among 4 to 14 pairwise interfaces of quite different sizes, some interfaces may be marginal whereas others make a dominant contribution to the stability of the crystal. We will call "dominant" any pairwise interface that contributes more than 30% to B_{tot}, an arbitrary choice which is sufficient to reveal a clear difference between crystal forms. Figure 5 is a histogram of the area B_{max} of the largest pairwise interface in each crystal form. In the absence of 2-fold symmetry, the average value of B_{max} is 900 Å^2, only 50% more than for the average interface. Fewer than 5% of these crystal forms have a dominant interface as defined above, and in many, all interfaces bury less than 700 Å^2. They are therefore much smaller than in protein–protein complexes. In crystal forms with 2-fold symmetry, the average B_{max} is 1350 Å^2, which approaches the size of interfaces in protein–protein complexes. Moreover, half of the crystals have a dominant interface. Its large size compensates for the smaller number of neighbors each molecule has relative to crystals with no 2-fold symmetry.

Our sample contains 14 pairwise interfaces, 1% of the total number, that bury more than 1600 Å^2. All but two are isologous, and so are 23 out of 27 interfaces that bury more than 1400 Å^2. Almost all occur in high-symmetry crystal forms. The largest is an isologous interface that buries 2600 Å^2 in trigonal crystals of adenylate kinase (3ADK). Interfaces in the range 1600–2200 Å^2 are found in tetragonal crystal forms of α_1-antichymotrypsin (2ACH), recoverin (1REC), γ-chymotrypsin (1GCT), and metallothionein (4MT2), in two different trigonal forms of pancreatic ribonuclease (1RNS, 1RAR), and in trigonal forms of troponin C (1TOP), papain (1PPO), and the $p21^{ras}$ protein (821P). Only a few occur in crystal forms of lower symmetry: orthorhombic aconitase (5ACN), porphobilinogen deaminase (1PDA), and monoclinic pepsinogen (3PSG). All point to isologous contacts between 2-fold symmetry-related protein molecules, except for tetragonal recoverin where the symmetry is 4-fold,[23] and trigonal troponin C where it is a screw 3-fold.[24,25]

The very large isologous interface in trigonal crystals of porcine adenylate kinase has already been described.[26] It defines a dimer where the active site cleft of each molecule interlocks with its counterpart on a molecule related by a crystallographic 2-point to fold (Fig. 6). It covers 12% of the protein surface and involves nearly one-fourth of the 194 residues. Though charged residues are abundant and several hydrogen bonds or salt bridges are formed, nonpolar

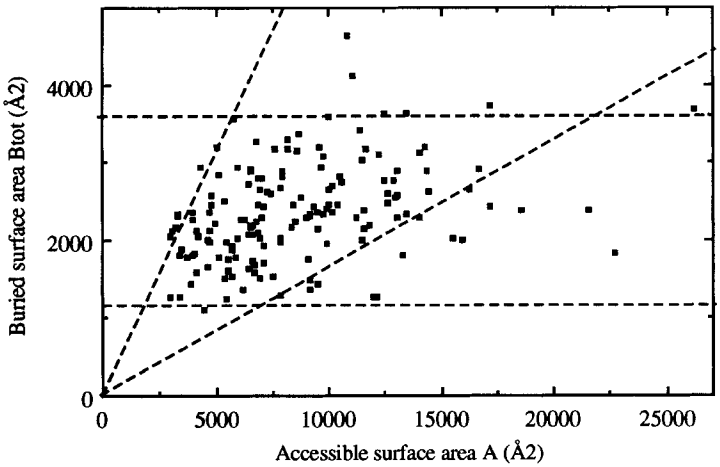

Fig. 4. Buried surface area in crystal packing. B_{tot} is the surface that becomes buried on the reference molecule as a result of contacts with all its neighbors in the 152 crystal forms. It is plotted against the solvent accessible surface area A of the reference molecule; A is related to the protein molecular weight M_r by: $A \approx 6.3\ M_r^{0.73}$;[22] 85% of the data points are within the region limited by the four dotted lines. They satisfy the conditions 1200 Å^2 < B_{tot} < 3600 Å^2 (horizontal lines) and 0.15 < B_{tot}/A < 0.6 (oblique lines).

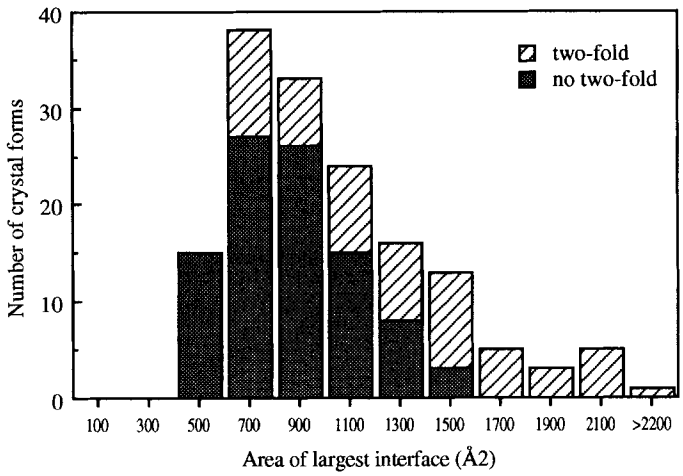

Fig. 5. Large interfaces in crystals. The interface area measured here is that of the largest among the N_{int} pairwise interfaces present in each of 94 crystal forms with no 2-fold symmetry or 58 with. All but two of the interfaces with B_{int}>1600 Å^2 are isologous.

atoms are a majority (55%) at the interface.[26] Similar features are observed in the pancreatic ribonuclease dimer, which is observed in several different crystals forms.[18–20] There, the contact hinders access to the active site and it would inhibit the enzyme if it existed in solution. The adenylate kinase dimer found in the trigonal crystal form certainly lacks enzymic activity, since in this case substrate binding and catalysis require significant changes in the conformation of those parts of the protein that make the contact.[27]

Point symmetries of order higher than two may also form oligomers in crystals. Crystal contacts involving proper (as opposed to screw) rotational symmetry are fairly scarce in our sample: we have only two examples of 3-fold symmetry, four of 4-fold, and none of 6-fold. The 3-folds occur in cubic insulin (1APH) and hexagonal leukocyte elastase (1PPG). In cubic insulin crystals,[28] the hormone forms hexamers and the largest interface is in the dimer, not the trimer. In leukocyte elastase crystals, interfaces between 3-fold related molecules are of medium size. In recoverin crystals (1REC), the 4-fold symmetry forms a putative tetramer with interfaces burying 2100 Å^2 each. In addition to a number of polar interactions, the association between 4-fold related molecules involves a ridge of nonpolar residues fitting into a patch of aromatic side chains.[23] We sug-

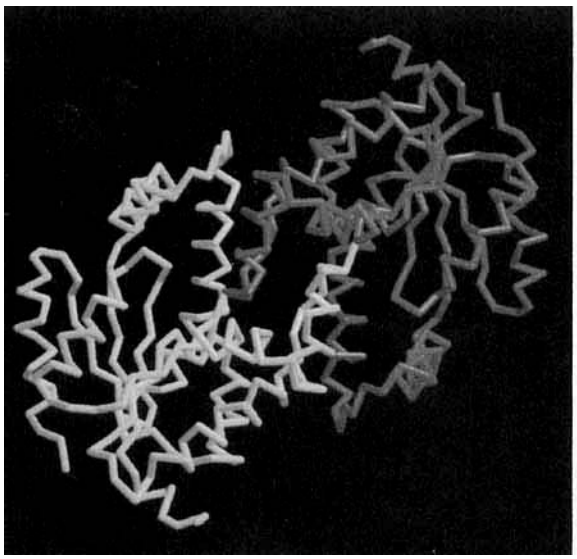

Fig. 6. The adenylate kinase dimer. The crystal packing in trigonal porcine cytosolic adenylate kinase generates a dimer where two molecules interact through their active site cleft.[26] In spite of the very large interface (2600 Å^2) and of several polar bonds, adenylate kinase appears to be a monomer in solution. The dimer, where the active site is blocked, might be induced under high-salt conditions (3 M ammonium sulfate) used for crystallization. Drawn with MolScript[34] with Raster3D[35] rendering.

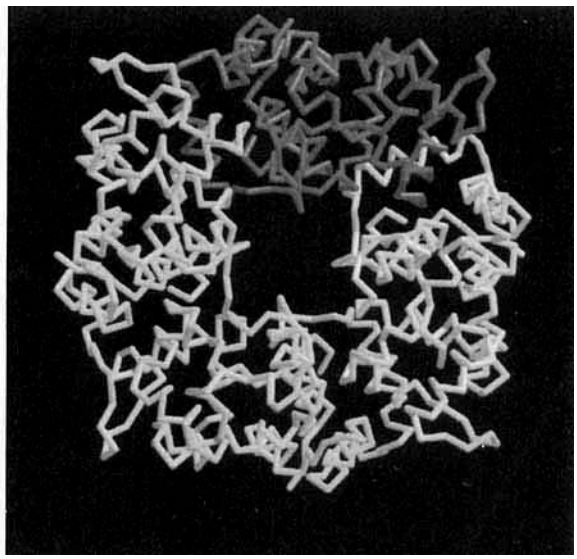

Fig. 7. The recoverin tetramer. Recoverin is a calcium binding protein involved in vision.[23] Tetragonal crystals grown in 70–90% saturated ammonium sulfate contain tetrads of molecules associated by a large pairwise interface (2100 Å^2) and related by the 4-fold axis of symmetry. Drawn with MolScript[34] with Raster3D[35] rendering

gest that the high-salt conditions used to produce these crystals stabilize a tetrameric form illustrated in Figure 7. Contacts between 4-fold related molecules also occur in tetragonal crystals of ribonuclease inhibitor (1BNH) and ovomucoid domain 3 (4OVO), but these interfaces are much less extensive (1000–1300 Å^2) than for recoverin. In tetragonal crystals of cobra phospholipase A2 (1POA), the largest interfaces are created by lattice translations, not rotations.

DISCUSSION

The survey of crystal packings presented here is purposely incomplete and limited to the case where the chemical unit is the crystal asymmetric unit. It nevertheless reveals properties that are likely to belong to all protein crystals. The most obvious is the ubiquitous existence of interfaces burying less than 1200 Å^2 on the two molecules in contact. These interfaces are similar in size to those created at random in computer experiments. These experiments simulate diffusional collision in solution, and interfaces created in this way should not be expected to create stable complexes. Their overwhelming abundance in protein crystals is consistent with models of crystal growth where monomeric units are added onto faces where they can form several pairwise interfaces simultaneously. The protein surface area becoming buried is only a fraction of the quantity $2B_{tot}$ which we measure for a molecule in the bulk of the crystal, but it still is well over 1200 Å^2, which we consider to be a lower limit for stable association: the newly incorporated molecule makes enough van der Waals or electrostatic interactions to become immobilized.

Pairwise interfaces in crystals may also be compared to those of protein–protein complexes and oligomeric proteins. In complexes, specific recognition relies on forming a well-packed interface between two surfaces that display shape and chemical complementarity. A typical protease-inhibitor or antibody-antigen complex buries 1500±250 Å^2 and includes 10±3 hydrogen bonds.[2–5] In oligomeric proteins, subunit contacts are even more extensive.[6] They bury up to 10,000 Å^2 and include an average of one hydrogen bond per 200 Å^2. We should not forget that these data have been derived from the study of crystals in which specific interactions are mingled with others that are crystal packing artefacts. In delineating the multimeric structure, we make the implicit assumption that *the biologically relevant interaction creates the most extensive interface.* This assumption has now been validated in so many cases that it is hardly a matter of debate, but it raises the question of whether unspecific interactions can create interfaces of comparable size.

The general answer being no, we must now consider the exceptions. A few unexpectedly large interfaces occur in crystals of monomeric proteins. They are nearly always associated with 2-fold symmetry, or, at least, with point group symmetry. Such symmetries, characteristic of oligomeric proteins, are

also remarkably frequent in crystals having more than one molecule per asymmetric unit, even when the molecule is monomeric.[15] Indeed, "crystal oligomers" may be a much more common feature in such crystals. We suggest that the proteins that form large pairwise interfaces in crystals are already oligomers in the solution from which they crystallize. Even though we have excluded from our sample the proteins known to form stable oligomers under standard conditions, the precipitants used for crystallization are also potential agents for dimerization. Moreover, the oligomeric observed in crystals may occasionally have biological relevance. Insulin is an example of a protein that exists both as a monomer and an oligomer in nature, and the insulin hexamer seen in cubic crystals corresponds to a natural storage form of the hormone.[29] In general, regions of the protein surface that form large interfaces in the crystal are candidates for specific interactions with partner molecules, absent from the crystal. Thus, it has been suggested that contacts seen in crystals of troponin C mimic its interaction with other proteins of the muscle fiber,[24] and that the hydrophobic region involved in the recoverin tetramer has a functional role in the retina.[23]

Oligomerization occurring under conditions leading to crystallization has been demonstrated in several cases. Spectroscopic data indicate that pancreatic ribonuclease dimerizes in high-salt solutions,[30] and that canavalin forms trimers well before it crystallizes.[31] The dissociation constant of the ribonuclease dimer must be less than the protein concentration in these experiments, and therefore in the 10^{-3}–10^{-4} M range. At lower ionic strength, it may be an order or two of magnitude larger, perhaps 10^{-2} M. Yet, the 1800 Å^2 interface and 10 hydrogen bonds observed in this dimer are equivalent to those of specific complexes with dissociation constants in the 10^{-7}–10^{-14} M range.[2] Most likely, repulsive forces prevent dimerization: for instance, electrostatic interactions between molecules bearing the same electric charge (positive in the case of ribonuclease). This reasoning also applies to adenylate kinase, where the dimer seen in trigonal crystals has an even larger interface and is devoid of enzymic activity. Whereas in specific complexes repulsive interactions have been eliminated by natural selection, we suggest that the converse prevails in general: *repulsive interactions are selected for.* They prevent association which would occur spontaneously in cells where protein concentration is high and the potentiality for nefarious interactions tremendous. A good example is hemoglobin: its concentration is almost the same in erythrocytes and in the crystal, and solubility places severe constraints on the composition of its surface. In sickle cell hemoglobin, the substitution of a single residue suffices to promote association of the molecules into fibers with dramatic consequences.[32]

CONCLUSION

Our analysis of the number and size of packing interfaces in protein crystals bears implications on the mechanism of crystallization. It also brings strong support to a conclusion that was reached from computer simulations of protein-protein complexes (reviewed in ref. 33): specific recognition relies first on pairing large regions of the partner proteins surface that have complementary shapes, as can be judged from the exclusion of water molecules at the interface. Only a small minority of the complexes formed at random in simulations achieve interfaces burying 1500 Å^2 as in the native complexes. Those which do are "false positive," incorrect structures that can have the same number of electrostatic interactions, hydrogen bonds, and salt bridges, as the correct one. Interactions in false positives may contribute less free energy to the association, but this calculation is exceedingly unreliable,[16] whereas the buried surface area is easy to measure. We now find that most crystal packing contacts are like contacts formed at random in terms of the protein area removed from contact with the solvent, and not like contacts involved in recognition. The exceptions are the large interfaces present in a few crystal forms. In several cases, they mimic functional interactions that the protein makes outside the crystal and therefore represent a degree of specific recognition.

ACKNOWLEDGMENTS

We are grateful to all authors of the many protein X-ray structures which were used here and could not be included in the list of references. We acknowledge the competent help of F. Zimmermann and discussions with Dr. A. Ducruix in Gif-sur-Yvette.

REFERENCES

1. Chothia, C., Janin, J. Principles of protein recognition. Nature (London) 256:705–708, 1975.
2. Janin, J., Chothia, C. The structure of protein-protein recognition sites. J. Biol. Chem. 265:16027–16030, 1990.
3. Bode, W., Huber, R. Ligand binding: proteinase-protein inhibitor interactions. Curr. Opinion Struct. Biol. 1:45–52, 1991.
4. Davies, D.R., Padlan, E.A., Sheriff, S. Antibody-antigen complexes. Annu. Rev. Biochem. 59:439–473, 1990.
5. Wilson, I.A., Stanfield, R.L. Antibody-antigen interaction. Curr. Opinion Struct. Biol. 3:113–118, 1993.
6. Janin, J. Miller, S., Chothia, C. Surface, subunit interfaces and interior of oligomeric proteins. J. Mol. Biol. 204:155–164, 1988.
7. Kitaigordskii, A.I. "Organic Chemical Crystallography." New York: Consultants Bureau, 1961.
8. Lee, B.K., Richards, F.M. Solvent accessibility of groups in proteins. J. Mol. Biol. 55:379–400, 1971.
9. Salemme, F.R., Genieser, L., Finzel, B.C., Hilmer, R.M., Wendoloski, J.J. Molecular factors stabilizing protein crystals. J. Crystal Growth 90:273–282, 1988.
10. Bernstein, F.C., Koetzle, T.F., Williams, J.B., Meyer, E.F., Jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Data Bank. A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
11. Rodier, F., Crosio, M.P., Chiadmi, M. An array processor program for computing the list of neighbours in molecular packing. Acta Crystallogr. A46:37, 1990.

12. Shrake, A., Rupley, A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. J. Mol. Biol. 79:351–371, 1973.
13. Monod, J., Wyman, J., Changeux, J.P. On the nature of allosteric transitions: A plausible model. J. Mol. Biol. 12: 88–118, 1965.
14. Padmaja, N., Ramakumar, S., Viswamitra, M.A. Space-group frequencies of proteins and of organic compounds with more than one formula unit in the asymmetric unit. Acta Crystallogr. A46:725–730, 1990.
15. Wang, X., Janin, J. Orientation of non-crystallographic symmetry axes in protein crystals. Acta Crystallogr. D49: 505–513, 1993.
16. Janin, J. Elusive affinities. Proteins 21:30–39, 1995.
17. Cherfils, J., Duquerroy, S., Janin, J. Protein-protein recognition analyzed by docking simulation. Proteins 11:271–280, 1991.
18. Svensson, L.A., Dill, J. Sjölin, L. Wlodawer, A., Toner, M. Bacon, D., Moult, J., Veerapandian, B., Gilliland, G.L. The crystal packing interactions of two different forms of bovine ribonuclease A.J. Crystal Growth 110:119–130, 1991.
19. Kim, E.E., Varadarajan, R., Wyckoff, H.W., Richards, F.M. Refinement of crystal structure of ribonuclease S. Comparison with and between the various ribonuclease A structures. Biochemistry 31:12304–12314, 1992.
20. Crosio, M.P., Janin, J., Jullien, M. Crystal packing in six crystal forms of pancreatic ribonuclease. J. Mol. Biol. 228: 243–251, 1992.
21. Islam, S.A., Weaver, D.L. Molecular interactions in protein crystals: solvent accessible surface and stability. Proteins 8:1–5, 1990.
22. Miller, S., Janin, J., Lesk, A.M., Chothia, C. Interior and surface area of monomeric proteins. J. Mol. Biol. 196:641–656, 1987.
23. Flaherty, K.M., Zozulya, S., Stryer, L., McKay, D.B. Three-dimensional structure of recoverin, a calcium sensor in vision. Cell 75:709–716, 1993.
24. Satyshur, K.A., Pyzalska, D., Graeser, M., Rao, S.T., Sundaralingam, M. Structure of chicken muscle troponin C at 1.78 Å resolution. Acta Crystallogr. D50:40–49, 1994.
25. Herzberg, O., James, M.N.G. Refined crystal structure of troponin C from turkey skeletal muscle at 2.0 Å resolution. J. Mol. Biol. 203:76–779, 1988.
26. Dreusicke, D., Karplus, P.A., Schulz, G.E. Refined structure of porcine cytosolic adenylate kinase at 2.1 Å resolution. J. Mol. Biol. 199:359–371, 1988.
27. Schulz, G.E., Müller C.W., Diederichs, K. Induced-fit movements in adenylate kinases. J. Mol. Biol. 213:627–630, 1990.
28. Gursky, O., Li, Y., Badger, J., Caspar, D.L.D. Monovalent cation binding to cubic insulin crystals. Biophys. J. 61: 604–611, 1992.
29. Hodgkin, D.C. Varieties of insulin. J. Endocrinol. 63:1–15, 1974.
30. Jullien, M., Crosio, M.P., Baudet-Nessler, B., Mérola, F., Brochon, J.C. Evidence for a dimeric intermediate on the crystallization pathway of ribonuclease A. Acta Crystallogr. D50:398–403, 1994.
31. Kadima, W., McPherson, A., Dunn, M.F., Jurnak, F.A. Characterization of precrystallization aggregation of canavalin by dynamic light scattering. Biophys. J. 57:125–132, 1990.
32. Perutz, M. "Protein Structure: New Approaches to Disease and Therapy." New York: W.H. Freeman, 1992.
33. Cherfils, J., Janin, J. Protein docking algorithms: Simulating molecular recognition. Curr. Opinion Struct. Biol. 3:265–269, 1993.
34. Kraulis, P.J. MolScript: A program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24:946–950, 1991.
35. Merritt, E.A., Murphy, E.P. Raster3D Version 2.0. A program for photorealistic molecular graphics. Acta Crystallogr. D50:869–873, 1994.