Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 10167–10172, September 1996
Biophysics

Packing at the protein–water interface

(x-ray crystal structure/protein surface/Voronoi polyhedra/solvent structure)

MARK GERSTEIN*† AND CYRUS CHOTHIA†‡

*Department of Structural Biology, Stanford University, Stanford, CA 94305; and ‡Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom

Communicated by Robert L. Baldwin, Stanford University Medical Center, Stanford, CA, May 24, 1996 (received for review March 6, 1996)

ABSTRACT **We have determined the packing efficiency at the protein–water interface by calculating the volumes of atoms on the protein surface and nearby water molecules in 22 crystal structures. We find that an atom on the protein surface occupies, on average, a volume ≈7% larger than an atom of equivalent chemical type in the protein core. In these calculations, larger volumes result from voids between atoms and thus imply a looser or less efficient packing. We further find that the volumes of individual atoms are not related to their chemical type but rather to their structural location. More exposed atoms have larger volumes. Moreover, the packing around atoms in locally concave, grooved regions of protein surfaces is looser than that around atoms in locally convex, ridge regions. This as a direct manifestation of surface curvature-dependent hydration. The net volume increase for atoms on the protein surface is compensated by volume decreases in water molecules near the surface. These waters occupy volumes smaller than those in the bulk solvent by up to 20%; the precise amount of this decrease is directly related to the extent of contact with the protein.**

The central role of water in the stability, structure, and function of proteins has made protein–water interactions a subject of general interest (1, 2). One problem often discussed is the nature of the packing at the protein–water interface. This is important, because of its implications for the way proteins interact with other molecules (3). The topic was widely discussed in the 1970s (4–7). Since then, structures determined by x-ray crystallography have provided detailed descriptions of the arrangements of water molecules around specific regions of protein surfaces. To our knowledge, however, the surface packing problem has yet to be resolved.

Proteins have high packing efficiencies: the ratio of the volume enclosed within the van der Waals envelope of residues to the actual space these residues occupy in protein structures is close to 0.8 (4, 8). About half the atoms on protein surfaces are polar and half are nonpolar. Water molecules, on the other hand, are only polar, and the tetrahedral arrangement of their hydrogen bonds means that they form an open structure with a packing efficiency of <0.4. The incommensurate nature of the chemical character and packing properties of proteins and water make it difficult to determine, *a priori*, how they pack together. On one hand, one might think that the tetrahedral hydrogen-bonded geometry of water would have difficulty accommodating all the protrusions and indentations on the protein surface, leading to the creation of defects and a very loose packing. Alternatively, one could equally well expect that protein atoms would interpenetrate into the very open structure of water, leading to a tight packing at the protein–water interface. This second scenario is supported by liquid-transfer experiments with simple apolar molecules. It is found, for instance, that the mixing water and alcohol leads to a net decrease in volume (9). To find an answer to the packing problem, we have calculated directly the packing efficiencies at the protein–water interface in 22 crystal structures (10–31).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Ultra-High-Resolution Protein Structures

For the analysis described here, we used the structures listed in Table 1. These structures, which were determined to resolutions between 1.0 and 1.5 Å and refined to give R factors between 12 and 20%, contain 54–306 residues and 63–350 resolved water sites. In some structures, a small proportion of the water sites represent alternate locations for single water molecules (or sets of adjacent, interlocking water molecules) and have partial occupancy. It is difficult to disentangle these sites, so they were not included in our calculations.

In all, we used 3255 water molecules from the 22 structures for our calculations. Of these water molecules, 160 were buried in the protein interior, and the other 3095 are part of the first or second layer around the protein surface. Tables 1 and 2 describe the general features of the contacts made by the water molecules. They cover between 28 and 75% of the surface of the different proteins. Three-quarters of them make good hydrogen bonds to polar atoms on protein surfaces: about half make one bond, one-third make two bonds, and the rest make three or more. They also make van der Waals contacts to apolar protein atoms. If n is the number of hydrogen bonds made to the protein, they have, on average, close to $1 + 1.2n$ contacts (Table 2). The large majority of the water molecules that are not directly hydrogen bonded to the protein surface forms one or more good hydrogen bonds to other water molecules.

Determining Packing Efficiency with Voronoi Polyhedra

To determined the packing efficiency at protein–water interfaces, we used a geometrical procedure originally developed by Voronoi (32) and first applied to proteins by Richards (4). This procedure allocates all the space within a structure, including cavities or defects, to its constituent atoms by constructing around each atom a minimally sized polyhedron (called a Voronoi polyhedron). As shown in Fig. 1, the faces of a Voronoi polyhedron are formed by planes perpendicular to vectors between an atom and its neighbors, and the edges of a polyhedron result from the intersection of these planes. The volume inside of a polyhedron is inversely proportional to the packing efficiency of its central atom—i.e., a big volume is indicative of a poorly packed atom.

To determine the volume effectively occupied by an atom, the Voronoi procedure requires the location of all of its neighbors. As shown in Fig. 1, missing neighbors can lead to highly distorted, misshapen polyhedra. This is particularly important for atoms on the protein surface, because many of their neighbors are water molecules that may not be completely defined even in very high resolution crystal structures. To deal with the incomplete hydration at the protein surface, we carried out packing calculations only on those surface atoms that are fully surrounded by solvent. We located these atoms by calculating the accessible surface area

†To whom reprint requests should be addressed.

Table 1. Protein structures used in the calculation and the volume increase of their surfaces

Protein	Protein Data Bank ID	Protein Data Bank Ref.	Resolution, Å	R factor, %	Surface atoms Total, no.	Surface atoms Hydrated, %	Volume increase, %
Trypsin inhibitor	5PTI	10	1.00	20	230	40	2.5
Protein G	1IGD	11	1.10	19	232	66	3.0
Rubredoxin	5RXN	12	1.20	12	182	67	4.3
Scorpion neurotoxin	2SN3	13	1.20	19	232	62	8.1
Achromobacter protease I	1ARB	14	1.20	15	712	31	3.5
Trypsin inhibitor	9PTI	15	1.22	17	230	47	4.2
Cutinase	1CUS	16	1.25	16	579	54	4.0
Ribonuclease A	7RSA	17	1.26	15	450	57	5.8
Scorpion toxin II	1PTX	18	1.30	15	238	75	5.4
Ribonuclease F1	1FUS	19	1.30	19	347	49	4.4
Oncomodulin	1RRO	20	1.30	18	386	47	6.2
Hen lysozyme	135L	21	1.30	19	431	48	5.5
Plastocyanin	1PLC	22	1.33	15	312	56	6.2
Thermitase	1THM	23	1.37	17	723	40	5.5
Repressor of primer	1RPO	24	1.40	19	229	58	6.8
Ribonuclease A	1RPG	25	1.40	17	423	53	7.1
Lysozyme	2IHL	26	1.40	17	441	50	4.7
Haemoglobin	1ECO	27	1.40	18	430	34	4.0
Flavodoxin	1RCF	28	1.40	14	499	58	3.8
Subtilisin	1ST3	29	1.40	19	687	28	4.3
Carboxypeptidase A	2CTC	30	1.40	16	911	31	2.0
Insulin	4INS	31	1.50	15	337	71	6.3
Average values			1.30	17	420	47	4.9

The last three columns show the following: the number of atoms on the surface of each protein; the percentage of these surface atoms that are hydrated sufficiently enough so that they are completely buried by crystallographically resolved waters and thus can be used for Voronoi volume calculations; and the percentage that the observed, total volume of these sufficiently hydrated surface atoms is larger than that of an equivalent set of atoms buried in the interior of the protein, respectively. That is, the last column is $(S - R)/R$, where S is the total Voronoi volume we calculate for the sufficiently hydrated surface atoms and R is the total volume for these same atoms if we use standard core volumes derived from ref. 8 for each atom.

(33) of the protein, first in the absence of and then in presence of solvent. We consider any atom that, in the absence of solvent, has a measurable accessible surface area to be a surface atom and any surface atom whose accessible surface area is reduced to zero by the presence of the solvent to be sufficiently hydrated for our calculations.

The Voronoi calculations were carried out using a C-language program that was written by Y. Harpaz (Medical Research Council Laboratory of Molecular Biology) and M.G. and is directly based on the original program of Richards (version 6.0, written March 1983 by M. D. Handschumacher and F. M. Richards). The most complete description of the procedure and program is given by Richards (34). To position the dividing plane between atoms of unequal size, we used method B, which makes the division according the radii of the contacting atoms, and, for atomic radii, we used the values listed in Chothia (35). In Table 3, we show that the results of the calculations are relatively insensitive to the choice of radii set. All calculations were carried out using the full unit cell with all symmetry-related water molecules and protein atoms.

Table 2. Contacts made to the protein surface by water molecules

No. of H-bonds made to the protein surface, n	Proportion of water molecules making n H-bonds,†* %	Average no. of van der Waals contacts to the protein by water molecules making n H-bonds†
0	26	0.9
1	41	2.4
2	24	3.3
3 or more	9	4.4

*Hydrogen bonds listed here are those for which the H_2O. . . . {N, O} distance is 3.3 Å or less. The use of a longer distance for the cutoff would give somewhat more hydrogen bonds but would not alter the general result. Of the water molecules making no hydrogen bonds to the protein, 85% are hydrogen bonded to other water molecules.

†Water molecules and nonpolar atoms were taken to be in contact if the distance between them was equal to, or less than, the sum of their Van der Waals radii plus 0.4 Å.

In all, we found 4332 protein surface atoms that were sufficiently hydrated for our calculations. As discussed in the caption to Fig. 1, none of these have distorted Voronoi polyhedra, and all were used in the packing calculations described here. As will become apparent, it is convenient to express the volumes of these surface atoms as the percentage by which they are larger than those of chemically equivalent atoms buried in the interior of the protein; that is, $(\bar{V}_s - \bar{V}_r)/\bar{V}_r$, where $\bar{V}_s$ is the volume of the surface atom and $\bar{V}_r$ is the mean volume for the same atom type when buried inside the protein. Values for the volumes of buried atoms were taken from the work of Harpaz *et al.* (8). On average, the surface atoms used in these calculations occupy volumes 5% larger than those of interior atoms (Table 1). For individual atoms, the extent of the increase in volume varies, and, in the following sections, we describe the extent to which this variation is related to structural environment and chemical type.

Volumes of Protein Atoms in Relation to their Exposure to Solvent

For each surface atom, we determined the fraction of its surface covered by water. This quantity was used to sort the atoms in ascending order and place them in 13 bins. For the atoms in each bin, we calculated their average exposure to solvent and their volume increase (as compared with chemically equivalent buried atoms). A plot of these numbers is shown in Fig. 2*A*. Though the volumes of individual surface atoms vary, it is clear from the figure that, on average, volume increase is nearly proportional to exposure; the correlation coefficient is 0.98. A simple line fit gives

Table 3. Effect of different atomic radii sets on the results

van der Waals radii set*	Values in Å of van der Waals radii	Surface volume increase for 5PTI,† %
Standard set (6 radii)	sp^3-C 1.87, S 1.85, sp^2-C 1.76, N 1.65, charged-N 1.5, O 1.4	2.5
Standard set with larger O (6 radii)	Above with O of 1.6	0.6
Set derived from ENCAD LJ parameters, with hydrogens (9 radii)	sp^3-C or S 1.82, sp^2-C 1.74, N 1.72, charged-N 1.68, water-O 1.54, sp^3-O 1.48, sp^2-O 1.35, H 1.17, polar-H 0.35	2.8
Simple Bondi set, with hydrogens (5 radii)	S 1.8, C 1.7, N 1.55, O 1.52, H 1.2	2.8

*The first radii set is derived from ref. 35. It reflects the closest contact or minimal distance between atoms in small-molecule crystal structures and has a 1.4 Å value for the water radius. It was used for all the calculations in the paper, both volume and surface area, and in the following discussion it will be referred to as the "standard set." It is of concern that this set does not explicitly model hydrogens. It also has been pointed out that the 1.4-Å value of the water radius reflects only hydrogen-bonding distances and not Lennard–Jones interaction parameters (36). In the other radii sets in the table, we attempt to assess whether these points lead to significant differences in the volume calculation. Initially, we increased the radius of the oxygen atom by ≈15% to give a second radii set. This models a larger water and gives an appreciable but smaller volume increase. However, it produces, in a sense, an unphysical radii set since the radii for the other atoms were not adjusted to give overall consistency. The third set remedies this problem. It is derived directly from the Lennard–Jones parameters in the Energy Calculations and Dynamics (ENCAD) potential (37). For each atom type, we found the radius where the Lennard–Jones potential just started to become repulsive [i.e., we solved for r in $U(r) = (A/r^{12}) - (B/r^6) = 0$]. The resulting radii set includes a larger value for the radius of a water molecule (about the same as in the second set), but it has the other radii scaled to compensate to some degree. This set also includes hydrogens and, in total, is rather complex, containing nine different radii. Nevertheless, despite the great differences in the values of the individual radii from the standard set, it results in a very similar volume increase. The fourth set is derived from Bondi (38). It also contains hydrogens and a large water radius but is much simpler with only five radii in total. It is rather notable that it gives essentially the same volume increase as either the ENCAD set or the standard set. This shows how insensitive our results are to choice of radii set.

†The last column contains the percentage increase in volume of all the sufficiently hydrated surface atoms in the 5PTI crystal structure. 5PTI was chosen because it is a neutron structure and has most hydrogen atoms positioned. This column shows the exact same quantity as the last column in Table 1—i.e., $(S - R)/R$, where S is the total Voronoi volume we calculate for the sufficiently hydrated surface atoms and R is the total volume for these same atoms if we use standard core volumes derived from ref. 8 for each atom.

an atom's percentage increase in volume as about one-fifth of its percentage increase in exposure.

The Packing in Grooves and Ridges on Protein Surfaces

Examination of the first protein structures for which the positions of a large number of water molecules were determined showed that water molecules are more clearly resolved in regions of the protein surface that are grooved or concave in shape than in the regions that have a ridged or convex shape (39, 40). Recent calculations by Kuhn *et al.* (41) showed that this is generally true. We extended our analysis of the relation between structural environment and volume of surface atoms by determining the packing differences that occur in groove and ridge regions of protein surfaces.

To distinguish between grooves and ridges, we calculated for each atom the ratio of its surface area in the context of a whole protein to its surface area in a Gly-Xaa-Gly tripeptide. This quantity, called fractional accessibility, is the relative surface area of an atom in the presence or absence of its immediate nonbonded neighbors. As such, it measures the "local," small-scale curvature in a given region. It ranges from 0 for atoms in deep grooves to 1 for atoms protruding straight out into solution. Our definition of fractional accessibility is the same as that used in the original paper on accessible surface area by Lee and Richards (33). It gives results consistent with other measures for curvature of the protein surface (41, 42).

We found a fractional accessibility of 0.35 to be a convenient threshold to separate atoms on ridges from those in grooves. Using this threshold, the relative percentage of the atoms buried, in surface grooves, or on surface ridges in the 22 structures is 47:30:23. Alternative, reasonable values for the threshold would, of course, give slightly different proportions but would not effect our conclusions. Our calculations also show that for the 22 structures, the proportion of completely hydrated protein atoms is 43–96% in groove regions but only 10–56% in ridge regions. Thus, we found water molecules to be more clearly resolved around grooves than ridges, as expected from previous work (39–41).

Overall, the 3260 hydrated atoms in groove regions occupy a volume 4.2% larger than equivalent atoms in protein interiors (Table 4), and the 1072 hydrated atoms in ridge regions occupy a volume 7.6% larger.

The larger volumes for atoms on ridges in comparison to those in grooves reflect, to a large degree, the fact that they are more exposed to solvent. What is not as obvious, however, is that the volume increases for ridge atoms are actually smaller in proportion to their exposed surface area. For instance, based on the numbers in Table 4, a carbonyl oxygen in a groove has, on average, a volume 0.7 $Å^3$ larger than it does in the protein interior and a surface area of 3.6 $Å^2$; on a ridge, the corresponding numbers are 2.1 $Å^3$ for volume increase and 15.0 $Å^2$ for surface area. This means that the volume increase per unit area ($\Delta V/A$) is 0.20 Å in grooves but only 0.14 Å on ridges. Very similar numbers are found for the other common atom types individually and collectively. For instance, methylene groups have a $\Delta V/A$ in grooves of 0.2 Å but only 0.15 Å on ridges, and averaged over all atoms, $\Delta V/A$ is 0.21 Å for grooves and 0.15 Å for ridges. These numbers show that water molecules do not pack as well around grooves as ridges.

In connection with this, it is interesting to note that the ratio of hydrogen-bond acceptors to donors is larger in the grooves than on the ridges. (Specifically, the ratio is 1.8 for the population of groove atoms versus 1.1 for the population of ridge atoms). One might imagine that the greater orientational flexibility that acceptors (O atoms) have in forming hydrogen bonds to water than donors (-NH-, $-NH_2$, etc.) might allow them to do so easily and thus produce tighter packing. However, the relatively large increases in their volumes show, in fact, that this flexibility is not great enough to overcome the restrictions imposed by their being in grooves.

Volume and Chemical Type of Surface Atoms

To determine whether the volume increase in the surface atoms is related to their chemical type, we grouped the protein atoms into 20 distinct chemical types (e.g., main chain carbonyl oxygen, side chain methylene carbon, etc.), using a scheme based on that of Richards (4). For both the ridge and groove regions a mean volume increase was determined for each atom type (Table 4).

Examination of the data in Table 4 shows little or no correlation between the volume increase of the surface atoms and their chemical character. On the groove surfaces, the magnitude of the volume increase is similar for all atom types. On the ridge surfaces, the extent of the volume increase for a given atom type is roughly related to how accessible it is to

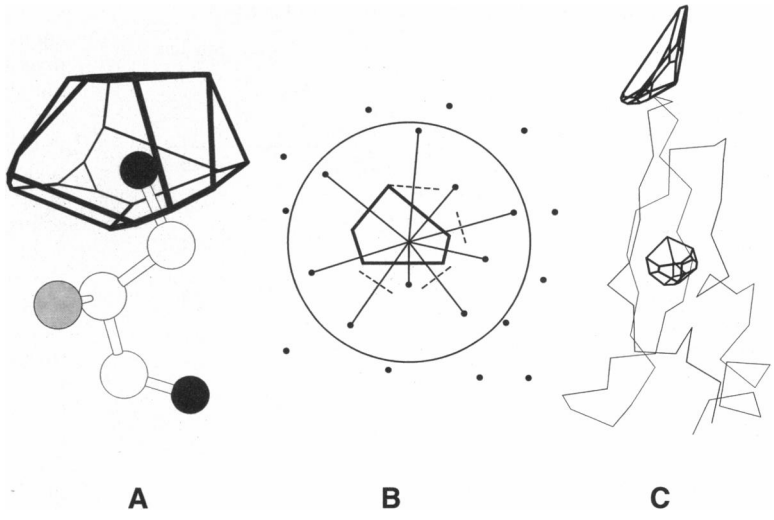

FIG. 1. Voronoi polyhedra. (*A*) A typical Voronoi polyhedron around a protein atom (the Oγ in a Ser). (*B*) To construct a Voronoi polyhedron, one draws lines connecting a central atom to all of its neighbors within a certain "cutoff" distance (indicated by the large circle in the figure). Then one constructs planes perpendicular to these lines, positioning them according to the ratio of atomic radii. The smallest polyhedron formed by the intersection of the dividing planes is unique and is the Voronoi polyhedron associated with the central atom. Those dividing planes far from the central atom (indicated by dashed lines) are excluded. If Voronoi polyhedra are constructed around atoms in a periodic system, such as in a crystal, all the volume in the unit cell will be apportioned to the atoms. However, if some of the neighbors around an atom are missing, the constructed polyhedron will be too large to be physically "reasonable" and will allocate "too much" space to the atom. This is the problem at the protein surface, where one often does not have enough neighboring water atoms. Furthermore, when some neighbors are missing, even if an apparently reasonable polyhedron can be constructed, it will often have a very "pointy" or distended shape. (*C*) A distorted polyhedron on the surface of trypsin inhibitor in contrast to a well-shaped one in the core. The shape of a polyhedron can be measured in terms of an asymmetry parameter, the ratio of maximum to minimum distances to the vertices. The asymmetry parameters for the polyhedra in *C* are 1.3 for the well-shaped one and 5.8 for the distorted one. Overall, for atoms in the protein core, the asymmetry parameter ranges from 1 to ≈2.7. However, for atoms on the surface with "pointy" polyhedra the asymmetry parameter can range up to 5 or more. We find that 3 is a useful threshold separating noticeably distorted from undistorted polyhedra, and all polyhedra used in the calculations had asymmetry parameters below this.

solvent. In particular, the main chain N, Cα, and C atoms and side chain $\rangle$C = atoms have volume increases (3–7%) and surface areas (1.5–5.0 Å^2) that are only slightly greater on the ridges than in the groove regions. On the other hand, the main chain oxygen atom and nearly all the other side chain atoms have volume increases and surface areas that are much larger on the ridges than in the grooves (on average, a 12% volume increase and 15.0 Å^2 of accessible surface).

The Total Volume of Surface Atoms

In all 22 structures, the aggregate volume of the surface atoms is larger than that of equivalent interior atoms (Table 1). Individual structures have volume increases that range from 2 to 8%, and 16 of the 22 have increases between 3.5 and 6.5%.

However, the water molecules that surround the most exposed atoms, those in the ridge regions, tend not to be well-resolved in the structure determinations. This means that the atoms with the largest volume changes are under represented in our calculations and that the overall surface volume increase is underestimated. We can compensate for this bias by using the data for the frequencies of different atom types, their average volumes on the surface, and the accessible surface areas of atoms included in our calculations in comparison to those left out because of incomplete hydration. Calculations with these numbers give volume increases for atoms in grooves that are very similar to those calculated directly (≈4.5%), but for the atoms on ridges they give a slightly larger number (≈10%). This implies that the surface atoms overall have ≈7% larger volumes than those of interior atoms.

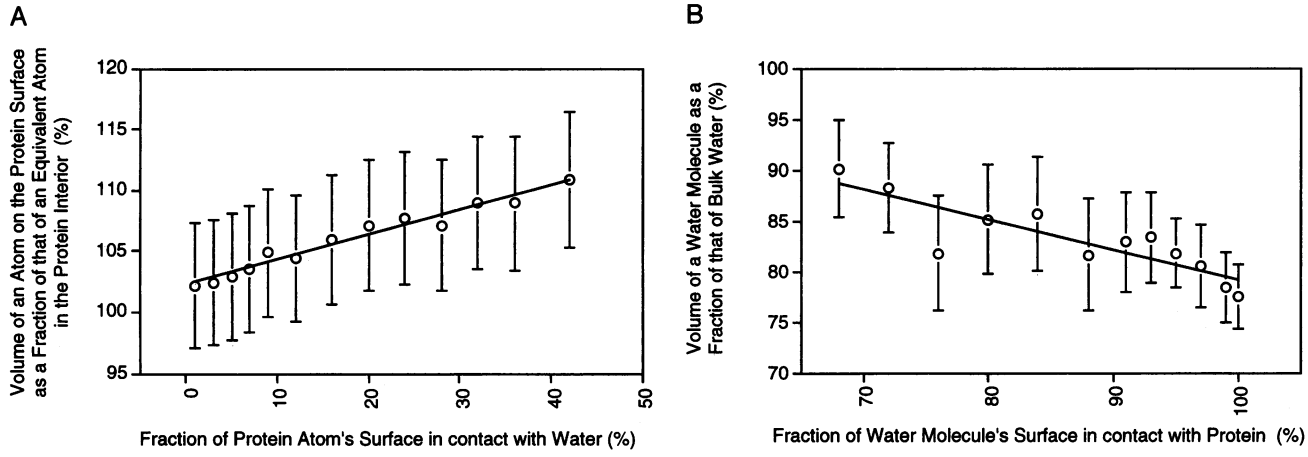

FIG. 2. The relationship between the degree of protein–water contact and change in volume. (*A*) The increase in volume of protein atoms as they make more contact with water, and (*B*) the decrease in volume of water molecules as they make more contact with the protein. The increase in protein atom volume, which is calculated over all atoms, is expressed as the measured volume of the protein atom on the surface divided by the reference volume of a chemically equivalent buried atom (and multiplied by 100%). Similarly, the decrease in water volume is shown as the ratio of an observed water molecule volume to that of a molecule in bulk water (29.7 Å^3). The degree of contact is expressed as a fraction of either the protein's atom surface covered by water or the water molecule's surface covered by protein. For both *A* and *B*, the error bars represent the standard deviation of values at each amount of contact. Simple line fits give proportionality constants between the percentage change in volume and the degree of contact: +0.20 for the *A* and −0.30 for *B*. Linear correlation coefficients, which give a measure of the spread of the points, can indicate how useful the line-fits are: they are 0.98 for the *A* and 0.87 for *B*.

Table 4. How the volume increase depends on atom type and location

	Total atoms, no.	N_s No.	N_s %	$\overline{V}_s$, Å^3	$\overline{V}_r$, Å^3	$\Delta V/\overline{V}_r$, %	$\overline{A}$, Å^2
In grooves (locally concave regions)							
All atoms	5243	3260	62	20.2	19.4	4.2	3.8
Main-chain atoms							
N	316	264	84	14.8	14.0	5.1	1.9
Cα	366	256	70	14.2	13.5	5.1	1.8
C	76	50	66	10.0	9.3	5.7	1.1
O	905	690	76	16.6	15.9	4.4	3.6
Cα(G)	122	64	52	24.2	23.4	2.7	5.7
Nonpolar side-chain atoms							
>CH—	90	70	78	15.2	14.7	2.2	2.0
—CH$_2$—	1140	655	57	24.5	23.7	3.3	4.0
—CH$_3$	547	262	48	37.6	36.6	2.9	6.1
>C=	190	97	51	10.6	10.1	4.7	2.0
—CH=	364	222	61	21.9	21.0	4.6	3.8
Polar side-chain atoms							
Other atoms	208	115	55	18.0	16.7	7.4	4.0
—NH$_2$	254	124	49	24.8	23.4	6.1	6.5
—OH	275	166	60	17.8	17.3	3.3	4.7
=O	124	67	54	18.3	16.8	6.8	4.8
—O (−)	266	158	59	16.7	16.0	3.2	5.1
On ridges (locally convex regions)							
All atoms	3998	1072	27	15.9	14.7	7.6	8.0
Main-chain atoms							
N	293	197	67	14.9	14.0	7.0	4.7
Cα	542	262	48	14.1	13.5	4.9	5.0
C	313	161	51	9.6	9.3	3.1	1.5
O	424	80	19	18.0	15.9	12.6	15.0
Side-chain atoms (polar and apolar)							
—CH$_2$—	780	111	14	26.0	23.7	10.1	15.3
>CH—	303	69	23	10.9	10.2	5.9	4.5
—CH=	151	33	22	23.4	21.0	6.6	14.4
—O (−)	198	32	16	18.0	16.1	8.0	20.8
Other atoms	994	127	13	32.5	29.6	11.8	25.5

Starting with the one labeled "Total atoms," from left to right the columns show: the number of atoms on the surface of each protein; the absolute number N_s and proportion of these surface atoms that are hydrated sufficiently enough so that they are completely buried by crystallographically resolved waters and thus can be used for volume calculations; the average observed volume $\overline{V}_s$ for the sufficiently hydrated surface atoms in the second column; the average volume of these same atoms $\overline{V}_r$ if the standard reference volumes for buried atoms were substituted for the observed Voronoi volume; the percentage difference between columns 4 and 5, $\Delta V/\overline{V}_r$, where $\Delta V = \overline{V}_s - \overline{V}_r$; and the average surface area $\overline{A}$ of the sufficiently hydrated atoms in column 2. Note that the statistical error in the volumetric quantities in the table is very small. Essentially, we are assessing whether $\overline{V}_s$ is significantly different from $\overline{V}_r$, compared to the null hypothesis that they are the same. For the large number of observations considered here, we would expect the difference ΔV to be distributed approximately normally with an estimated standard error $SE(\Delta V) = \sqrt{S_s^2/N_s + s_r^2/N_r}$, where s_s is the standard deviation in our sample of surface volumes, and s_r and N_r are the number of buried volumes we measured and their fluctuation in computing the reference volumes in ref. 8, respectively. Since we measured a much larger number of reference volumes than surface volumes and they had smaller amounts of fluctuation, the standard error reduces to $SE(\Delta V) \approx S_s/\sqrt{N_s}$. Evaluating this quantity, for instance, for the Cα atom in groove regions (which has an observed mean volume of 14.2 Å^3 and a standard deviation of 1.12 Å^3), we find that $SE(\Delta V) = 1.12/\sqrt{366} \approx 0.059$ Å^3, which is ~0.4%. The error in mean for the associated buried reference volume (13.5 Å^3) is much less. Consequently, the volume increase (5.1%) is quite large compared to the overall statistical error in the calculation (0.4%), about 13σ above the mean (Z score). This has a virtually nil chance (i.e., P value) of occurring under the null hypothesis.

The Volume Occupied by Water Near the Protein

The surface area calculations used here show that 160 water molecules are buried in protein interiors and that 393 water molecules that contact the protein surface are completely covered by a second layer of water molecules. This means we can use the Voronoi procedure to calculate the volumes of water molecules associated with the protein. The buried water molecules have a mean volume of 22.9 Å^3 (with a relatively narrow standard deviation of 7%). The 393 water molecules near the protein surface have a mean volume of 24.5 Å^3 with >90% between 20 and 30Å^3. This value is ≈18% less than the volume of a molecule in bulk water (29.7 Å^3) but still considerably larger than the van der Waals volume of an isolated water molecule (11.5 Å^3, assuming a 1.4 Å radius).

We examined the relationship between the volume of water molecules and the extent of their contact with the protein using a procedure similar to that for the protein surface atoms. Although we have data for only a relatively small number of water molecules, they do show a clear relationship between contact area and volume change. For a given contact area, individual water volumes vary, but the mean values are inversely proportional to the extent of the contact (Fig. 2*B*). The proportionality constant (−0.3) is a little larger in magnitude than that for protein atoms.

Discussion

The surface water around a given protein has, in different crystal forms, quite different configurations. Zegers *et al.* (25) recently examined the water seen around ribonuclease A in crystals from

five different space groups and found that only 17 water sites are conserved in all structures. Very similar results have been obtained from analyzing different crystal forms of pancreatic trypsin inhibitor (43) and trypsin (44). Thus, the structural data analyzed here represents not only different configurations of water molecules around different proteins but also just one (or two) of the many different configurations that are to be found in solution around each of them—configurations that are stabilized, in part at least, by particular crystal forms.

Overall, we have shown that, although the water around proteins has considerable variations in its local structure, its average properties have clear regularities. We estimate that surface atoms on average occupy volumes ≈7% greater than those of atoms buried within proteins. These overall results are consistent with those obtained recently in a molecular dynamics simulation of a small protein (45), where it was found that the volumes of atoms on the protein surface are, on average, 6% larger than equivalent atoms inside the protein core. We also obtained consistent results when we carried out our packing calculations on the protein–water interface in a very recent structure determination, which used direct experimental phase information for the protein and solvent (46).

For individual atoms, we find that the volume increase has little or no relation to chemical character but is roughly proportional to the extent of exposure to solvent. In detail, we find that atoms in grooves expand more per unit of exposed surface than those on ridges. This effect must be due, at least in large part, to the difficulty water molecules have in fitting into grooved regions of the protein surface while still maintaining their tetrahedral hydrogen-bonding. It may, thus, be a manifestation of "structural hydrophobicity"—i.e., a hydrophobic effect due to the shape and not the chemical character of the groups on the protein surface. This phenomena, which is also called curvature-dependent hydration, has been previously been argued for on the basis of macroscopic solution-transfer models (42) and microscopic simulations (47). Thus, our results are a direct observation in crystal structures of an effect that, up to now, has only been indirectly substantiated.

Turning attention to the water, we find that water molecules near the protein surface decrease in volume by an amount that is proportional to the extent of their contact with the protein. This means the volume changes in the atoms at the protein–water interface must cancel, at least in part. Previously, we demonstrated that protein volumes calculated just from the standard volumes of buried residues (after making allowance for electrostriction around charged groups) gave values that on average were only 0.4% larger than those determined experimentally (8). This suggested that either residues on the surface have the same volumes as those in the interior or that the volume changes that occur at the surface tend to cancel. The results reported here demonstrate that it is the latter explanation that is correct.

The picture of the protein–water interface that emerges from our results is the following: the close packing of the protein core produces a structure more tightly packed than even organic crystals (8), which is incommensurate with the hydrogen-bonded open structure of water. At the protein–water interface, these two structures interpenetrate with water molecules, making mainly one or two hydrogen bonds to polar groups on the protein and some van der Waals contacts to nonpolar protein atoms. And as we move across the interface, we gradually move between the different packing environments, through a region of intermediate density with more loosely packed protein and more tightly packed water.

We thank M. Levitt, Y. Harpaz, J. Tsai, A. Franke, R. Baldwin, and D. Kharakoz for useful discussions and comments on the manuscript. We also thank Fred Richards for conversations in 1974 when these calculations were first attempted in his laboratory by one of us. M.G. is supported by a Damon Runyon–Walter Winchell fellowship (DRG-1272). [Supplementary material relevant to this paper (source code for volume calculating program, tables of standard volumes, and a hyper text tutorial on the Voronoi construction) is available via the World Wide Web at the following locator: http://hyper.stanford.edu/~mbg/Surface Volumes.]

1. Kuntz, I. D. & Kauzmann, W. (1974) *Adv. Protein Chem.* **28,** 239–345.
2. Edsall, J. T. & McKenzie, H. A. (1983) *Adv. Biophys.* **16,** 53–183.
3. Kauzmann, W. (1987) *Nature (London)* **325,** 763–764.
4. Richards, F. M. (1974) *J. Mol. Biol.* **82,** 1–14.
5. Kauzmann, W., Moore K. & Schultz, D. (1974) *Nature (London)* **248,** 447–449.
6. Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng.* **6,** 151–176.
7. Liquori, A. M. & Sadun, C. (1981) *Int. J. Biol. Macromol.* **3,** 56–59.
8. Harpaz, Y., Gerstein, M. & Chothia, C. (1994) *Structure* **2,** 641–649.
9. Franks, F. (1983) *Water* (The Royal Society of Chemistry, London).
10. Wlodawer, A., Walter, J., Huber, R. & Sjölin, L. (1984) *J. Mol. Biol.* **180,** 301–329.
11. Derrick, J. P. & Wigley, D. B. (1994) *J. Mol. Biol.* **243,** 906–918.
12. Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1980) *J. Mol. Biol.* **138,** 615–633.
13. Zhao, B., Carson, M., Ealick, S. E. & Bugg, C. E. (1992) *J. Mol. Biol.* **227,** 239–252.
14. Tsunasawa, S., Masaki, T., Hirose, M., Soejima, M. & Sakiyama, F. (1989) *J. Biol. Chem.* **264,** 3832–3839.
15. Eigenbrot, C., Randal, M. & Kossiakoff, A. (1991) Protein Data Bank, File 9PTI (Brookhaven National Laboratory, Upton, NY).
16. Martinez, C., De Geus, P. D., Lauwereys, M., Matthyssens, G. & Cambillau, C. (1992) *Nature (London)* **356,** 615–618.
17. Wlodawer, A., Svensson, L. A., Sjölin, L. & Gilliland, G. L. (1988) *Biochemistry* **27,** 2705–2717.
18. Housset, D., Habersetzer-Rochat, C., Astier, J. P. & Fontecilla-Camps, J. C. (1994) *J. Mol. Biol.* **238,** 88–103.
19. Vassylyev, D. G., Katayanagi, K., Ishikawa, K., Tsujimotohirano, K., Danno, M., Pahler, A., Matsumoto, O., Matsushima, M., Yoshida, H. & Morikawa, K. (1993) *J. Mol. Biol.* **230,** 979–966.
20. Ahmed, F. R., Przybylska, M., Rose, D. R., Birnbaum, G. I., Pippy, M. E. & Macmanus, J. P. (1990) *J. Mol. Biol.* **216,** 127–140.
21. Harata, K. (1993) *Acta Crystallogr. D* **49,** 497–504.
22. Guss, J. M., Bartunik, H. D. & Freeman, H. C. (1992) *Acta Crystallogr. B* **48,** 790–811.
23. Teplyakov, A. V., Kuranova, I. P., Harutyunyan, E. H., Vainshtein, B. K., Frommel, C., Hohne, W. E. & Wilson, K. S. (1990) *J. Mol. Biol.* **214,** 261–279.
24. Vlassi, M., Steif, C., Weber, P. Tsernoglou, D., Wilson, K. S., Hinz, H. J. & Kokkinidis, M. (1994) *Nat. Struct. Biol.* **1,** 706–716.
25. Zegers, I., Maes, D., Daothi, M. H., Poortmans, F., Palmer, R. & Wyns, I. (1994) *Protein Sci.* **3,** 2322–2339.
26. Houdusse, A., Bentley, G. A., Poljak, R. J., Souchon, H. & Zhang, Z. (1991) Protein Data Bank, File 2IHL (Brookhaven National Laboratory, Upton, NY).
27. Steigemann, W. & Weber, E. (1981) *J. Mol. Biol.* **127,** 309–338.
28. Rao, S. T., Shaffie F., Yu, C., Satyshur, K. A., Stockman, B. J., Markley, J. L. & Sundaralingam, M. (1992) *Protein Sci.* **1,** 1413–1427.
29. Goddette, D. W., Paech, C., Yang, S. S., Mielenz, J. R., Bystroff, C., Wilke, M. E. & Fletterick, R. J. (1992) *J. Mol. Biol.* **228,** 580–595.
30. Teplyakov, A., Wilson, K. S., Orioli, P. & Mangani, S. (1993) *Acta Crystallogr. D* **49,** 534–540.
31. Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M. C., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, K., Sakabe, N. & Vijayan, N. M. (1988) *Phil. Trans. R. Soc. (London)* **319,** 369–456.
32. Voronoi, G. F. (1908) *J. Reine Angew. Math.* **134,** 198–287.
33. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400.
34. Richards, F. M. (1985) *Methods Enzymol.* **115,** 440–464.
35. Chothia, C. (1975) *Nature (London)* **254,** 304–308.
36. Madan, B. & Lee, B. K. (1994) *Biophys. Chem.* **51,** 279–289.
37. Levitt, M., Hirschberg, M., Sharon, R. & Daggett, V. (1995) *Comput. Phys. Comm.* **91,** 215–231.
38. Bondi, A. (1964) *J. Phys. Chem.* **68,** 441–451.
39. Watenpaugh, K. D., Margulis, T. N., Sieker, L. C. & Jensen, L. H. (1978) *J. Mol. Biol.* **122,** 175–190.
40. Blake, C. C. F., Pulford, W. C. A. & Artymiuk, P. J. (1983) *J. Mol. Biol.* **167,** 693–723.
41. Kuhn, L. A., Siani, M. A., Pique, M. E., Fisher, C. I., Getzoff, E. D. & Tainer, J. A. (1992) *J. Mol. Biol.* **228,** 13–22.
42. Sharp, K., Nicholls, A., Fine, R. & Honig, B. (1991) *Science* **252,** 106–109.
43. Wlodawer, A., Nachman, J., Gilliland, G. L., Gallagher, W. & Woodward, C. (1987) *J. Mol. Biol.* **198,** 469–480.
44. Finer-Moore, J. S., Kossiakoff, A. A., Hurley, J. H., Earnest, T. & Stroud, R. M. (1992) *Proteins* **12,** 203–222.
45. Gerstein, M., Tsai, J. & Levitt, M. (1995) *J. Mol. Biol.* **249,** 955–966.
46. Burling, F. T., Weis, W. I., Flaherty, K. M. & Brunger, A. T. (1996) *Science* **271,** 72–77.
47. Gerstein, M. & Lynden-Bell, R. M. (1993) *J. Mol. Biol.* **230,** 641–650.