J. Mol. Biol. (1995) **249**, 955–966

JMB

The Volume of Atoms on the Protein Surface: Calculated from Simulation, using Voronoi Polyhedra

Mark Gerstein*, Jerry Tsai and Michael Levitt

Department of Structural Biology, Fairchild Building D109, Standford University Standford, CA 94305, USA

**Corresponding author*

We analyze the volume of atoms on the protein surface during a molecular-dynamics simulation of a small protein (pancreatic trypsin inhibitor). To calculate volumes, we use a particular geometric construction, called Voronoi polyhedra, that divides the total volume of the simulation box amongst the atoms, rendering them relatively larger or smaller depending on how tightly they are packed. We find that most of the atoms on the protein surface are larger than those buried in the core (by $\sim$6%), except for the charged atoms, which decrease in size, presumably due to electroconstriction. We also find that water molecules are larger near apolar atoms on the protein surface and smaller near charged atoms, in comparison to ''bulk'' water molecules far from the protein. Taken together, these findings necessarily imply that apolar atoms on the protein surface and their associated water molecules are less tightly packed (than corresponding atoms in the protein core and bulk water) and the opposite is the case for charged atoms. This looser apolar packing and tighter charged packing fundamentally reflects protein–water distances that are larger or smaller than those expected from van der Waals radii. In addition to the calculation of mean volumes, simulations allow us to investigate the volume fluctuations and hence compressibilities of the protein and solvent atoms. The relatively large volume fluctuations of atoms at the protein–water interface indicates that they have a more variable packing than corresponding atoms in the protein core or in bulk water. We try to adhere to traditional conventions throughout our calculations. Nevertheless, we are aware of and discuss three complexities that significantly qualify our calculations: the positioning of the dividing plane between atoms, the problem of vertex error, and the choice of atom radii. In particular, our results highlight how poor a ''compromise'' the commonly accepted value of 1.4 Å is for the radius of a water molecule.

Keywords: Voronoi volumes; pancreatic trypsin inhibitor; molecular dynamics; accessible surface; high-pressure simulation

Introduction

The protein surface is of great interest since proteins recognize other molecules and perform their functions through their surfaces. Central to understanding the protein surface is understanding its interaction with water. Much work has been done on protein–water interactions and this subject has recently been reviewed (Levitt & Park, 1993; Saenger, 1987; Teeter, 1991). One of the more interesting questions relating to protein–water interactions concerns how tightly water is packed against the protein surface. It is difficult to simply guess an answer to this question since water and protein have such radically different structures. Water structure is determined by highly directional hydrogen bonding while protein structure is greatly influenced by the tight packing of apolar atoms.

Voronoi volume calculations are one of the best methods available to study packing. These calculations were originally developed by Voronoi (1908) and were first applied to proteins by Richards (1974). Since then they have been used successfully in the calculation of standard volumes of protein residues, in characterizing protein–protein interactions, in understanding protein motions, and in analyzing cavities in protein structure (Chothia, 1975; Harpaz *et al*., 1994; Janin & Chothia, 1990; Richards, 1977, 1979, 1985; Gerstein *et al*., 1993; Finney, 1975; Finney *et al*., 1980). They have also been used in the analysis of pure water simulations (Shih *et al*., 1994), and the

Abbreviations used: PTI, pancreatic trypsin inhibitor; VDW, van der Waals.

0022–2836/95/250955–12 $08.00/0

faces of Voronoi polyhedra have been used to characterize protein accessibility and assess the fit of docked substrates in enzymes (Finney, 1978; David, 1988).

The Voronoi procedure allocates all space amongst a collection of atoms. Each atom is surrounded by a polyhedron and allocated the space within it (Figure 1). The faces of Voronoi polyhedra are formed by constructing dividing planes perpendicular to the interatomic vectors between atoms, and the edges of the polyhedra result from the intersection of these planes.

In applying the Voronoi procedure to the protein surface, there are a number of difficulties which have to be addressed. First of all, to determine accurately an atom's volume, the Voronoi procedure requires the location of all of its neighbors. This is possible in the protein core, but on the protein surface many of the neighbors of a protein atom are water molecules, which are often not located in crystal structures. To address this difficulty, we have used molecular simulation to realistically position water molecules around a protein and to simulate their movement. Using molecular simulation, moreover, provides us with many instances to average over, so we can get much better statistics from simulation than from crystal structures.

The second complexity in applying the Voronoi procedure to the protein surface is that protein atoms have different sizes. In the original Voronoi procedure, all atoms are considered equal, and each dividing plane is positioned midway between two atoms (Figure 2A). However, doing this tends to chemically misallocate volume between atoms of *a priori* different size (i.e. a carbon and an oxygen). A number of investigators, therefore, have modified the procedure so that the dividing plane is positioned according to atom radii (Richards, 1974; Gellatly & Finney, 1982). Making this modification introduces further complexities into the Voronoi procedure since it is necessary to have a reasonable scheme for ''typing'' atoms and assigning them radii. Furthermore, in Richards' modification, which is perhaps the most chemically reasonable, the allocation of space is no longer mathematically perfect because the volume in tiny tetrahedrons near each polyhedron vertex is not allocated to any atom (''vertex error'' as shown in Figure 2C). The problems of assigning atom radii and vertex error are most acute on the protein surface. For here, while we do not really have an *a priori* idea of the exact size relationships between atoms, we do expect the differences in atom size to be large (i.e. between a methyl carbon and a water oxygen).

We are conscious of these problems and discuss their impact on our results. However, we have not been able to find an alternative better than Voronoi polyhedra for studying packing on the protein surface. Moreover, we have not found it worthwhile to make substantial modifications in the usual dividing-plane positioning scheme (known as Richard's method B) and usual set of atom radii (from Chothia, 1975) used to adapt the Voronoi procedure to proteins. Rather, we have done our calculations in a traditional fashion to make them fully comparable with previous work (Richards, 1974; Chothia, 1975; Harpaz *et al*., 1994).

Overall, we find that most atoms on the protein surface are larger than those in the core, the only exception being charged atoms. Furthermore, we find that the waters around the protein are marginally larger around hydrophobic atoms than charged atoms. We have also taken advantage of the molecular simulation technique and in addition

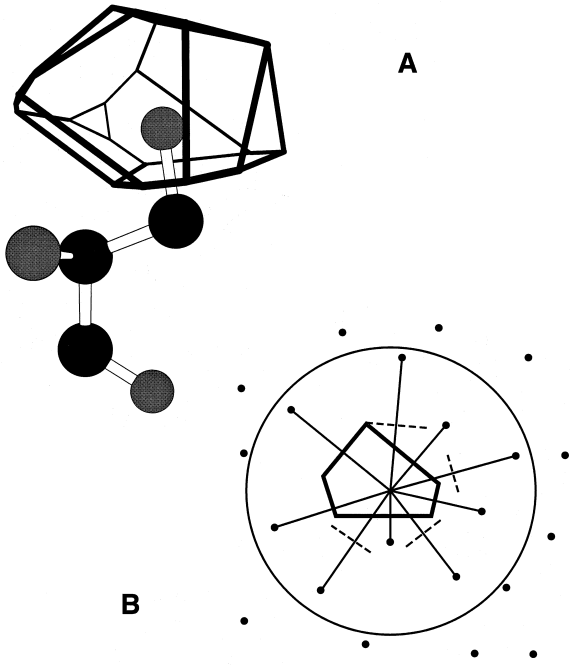

Figure 1. Using Voronoi polyhedra to calculate volumes. Voronoi (1908) found a way of partitioning all space amongst a collection of points using specially constructed polyhedra. In A, a typical Voronoi polyhedron around a protein atom (the O^{γ} in a Ser). As shown in B, to construct a Voronoi polyhedron, one draws lines connecting a central point (i.e. atom) to all of its neighbors within a certain ''cutoff'' distance (indicated by the large circle in the Figure). Then one constructs all planes perpendicular to these lines. As discussed in the legend to Figure 2, these dividing planes are positioned midway between the atoms in Voronoi's original construction and according to the ratio of the atoms' VDW radii in Richards' (1974) modification to the procedure (method B). The smallest polyhedron formed by the intersection of the dividing planes is unique and is the Voronoi polyhedron associated with the central atom. Only the innermost planes will be part of the final polyhedron; those dividing planes far from the central atom (indicated by broken lines in the Figure) are not part of the final polyhedron. The resulting Voronoi polyhedron contains all points in space closer to the central atom than to any others. In a sense the polyhedron represents the weighted sum of the distances from the central atom to its neighbors, where the weighting factor is the contact area between the atoms. 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. There will be no gaps or cavities as there would be if one, for instance, simply drew spheres around the atoms.

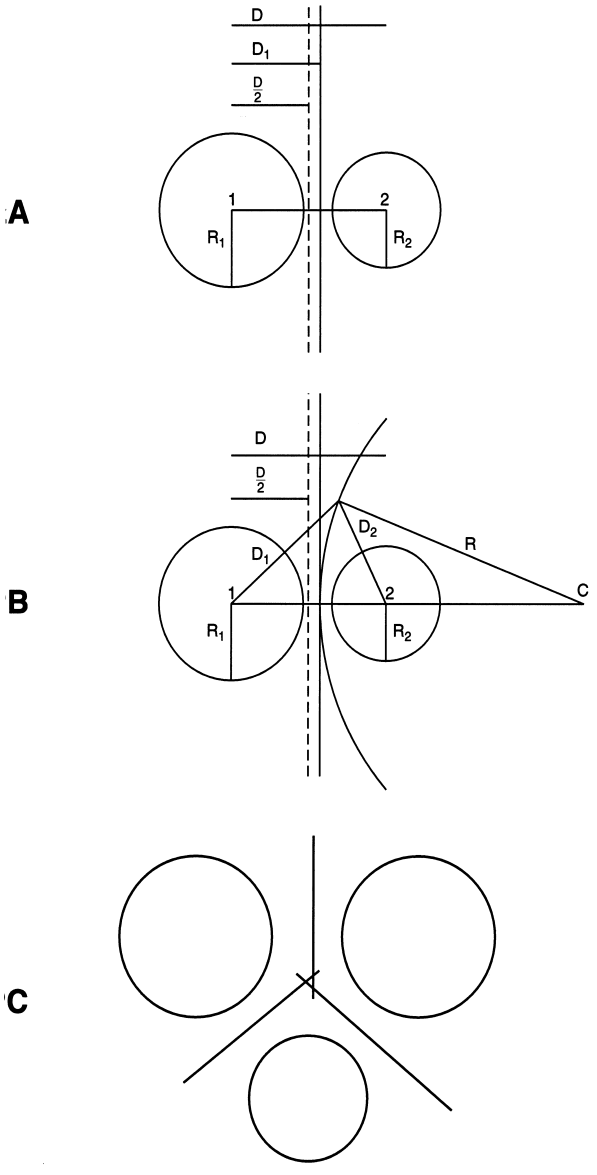

Figure 2. The partitioning of volume between different atoms. One problem with applying the Voronoi procedure to proteins is that protein atoms are of intrinsically different size. In the standard Voronoi procedure the dividing plane is positioned midway between the two atoms, as indicated in A by the dotted line a distance $D/2$ from each atom. This is not physically reasonable for atoms of obviously different size (such as oxygen and sulfur), so a variety of different methods have been proposed to reposition the plane to make the partition more physically reasonable. These methods depend on the radii of the atoms in contact (R_1 and R_2) and the distance between the atoms (D). As shown in A, they position the plane at a distance D_1 from the first atom. This distance is always set such that the plane is closer to the smaller atom. The two principal methods for determining D_1 are method B (Richards, 1974) and the radical plane method (Gellatly & Finney, 1982). For atoms that are covalently connected, method B divides the distance between the atoms proportionality according to their covalent-bond radii.

$$D_1 = D\,\frac{R_1}{R_1 + R_2} \tag{4}$$

to mean volumes, we have looked at volume fluctuations. In the same way that the vibrational amplitude of a spring can be related to its spring constant, these volume fluctuations can be directly related to compressibilities. Calculations of compressibility and volume fluctuations provide valuable new information relevant to the packing at the protein surface since they measure how variable this packing is over time.

Our statements regarding the volume changes of atoms reflect a particular allocation of volume that is contingent upon the radii chosen and the partitioning method used. It is important to realize that these volume changes could arise from two conceptually

---

For atoms that are not covalently connected, method B splits the remaining distance between them after subtracting away their VDW radii.

$$D_1 = R_1 + \frac{D - (R_1 + R_2)}{2} \tag{5}$$

For separations that are not much different from the sum of the radii, the above two formulas ((4) and (5)) give essentially the same result.

Method B suffers from vertex error. That is, the calculation does not account for all space, and tiny tetrahedrons of unallocated volume are created near the vertices of each polyhedron. Such an error tetrahedron is shown in C. The radical plane method does not suffer from vertex error, but it is not as chemically reasonable as method B. In this method, the plane is positioned according to the following formula:

$$D_1 = \frac{D^2 + R_1^2 - R_2^2}{2D} \tag{6}$$

In our calculations, we use method B. We do this mostly so that our work is comparable to previous calculations, in particular that of Harpaz *et al*. (1994) and Chothia (1975).

Method B and all the other partitioning schemes derived from Voronoi polyhedra are limited to describing the boundaries between atoms with planes. It is impossible for all points on a planar surface to have the same ratio of distances to two atoms (assuming the ratio is not 1). This can be construed as a limitation since the essential proportioning criteria of all the methods is that space be divided amongst the atoms in accordance with their size. To understand how using planes may understate or overstate the volumes of atoms, we generalized method B to use non-planar boundaries. The method B formula for positioning the dividing plane (equation (4)) can be rewritten as $D_1 = rD_2$, where D_1 is the distance from the plane to atom 1, D_2 is the distance to atom 2, and

$$r = \frac{R_1}{R_2}.$$

Now, as shown in B, to generalize this formula, one imagines that D_1 and D_2 describe distances to atom 1 and 2 from any point in space, not just from a point on the interatomic vector between the atoms. One can then adapt the poor man's Voronoi procedure described in Methods to allocate a volume ''voxel'' to atom 1 if $D_1 < rD_2$ and to atom 2 if otherwise. The resulting boundaries between atom volumes are now curved arcs. The center C and radius R of the boundary arc shown in B are:

$$C = \frac{d}{r^2 - 1} \quad \text{and} \quad R = rC. \tag{7}$$

Table 1. Volumes of particular types of protein atoms in the simulation

Atom type	Charge	Buried[a] Num.	Buried[a] Volume	Surface[a] Num.	Surface[a] Volume	Difference[b] Volume (%)
C		33	9.7	19	9.9	+2.8
CA		20	14.7	32	15.1	+2.7
O[c]		18	16.8	39	18.1	+8.0
N		24	14.2	24	15.3	+7.3
>C═		11	10.7	15	10.8	+0.6
—CH═		16	22.8	20	24.0	+5.6
>CH—		3	14.9	5	15.8	+6.5
—CH2—		11	25.5	67	27.4	+7.5
—CH3		0	—	20	42.7	
═O[c]		2	17.5	2	18.1	+3.5
—O[c]	–0.9	1	13.7	9	13.2	–3.6
OH		2	17.3	7	23.5	+35.8
—NH2	+0.4	2	29.7	14	28.0	–5.7
—NH3	+1.0	0	—	4	20.6	
—S—		4	32.2	2	34.5	+7.1

[a] Average volumes (in Å^3), averaged over the simulation, are shown for selected atom types.

[b] The percentage difference between surface and buried volumes is shown.

[c] The smaller volume of the carboxyl oxygen in comparison to the carbonyl oxygen most probably reflects the effect of electroconstriction. That is, since these two atoms differ only in charge and not in hybridization, number of attached hydrogen atoms, and (in the ENCAD parameter set) Lennard-Jones parameters, the tighter packing of water molecules around the carboxyl oxygen results from its greater charge. It is worth mentioning, however, that another effect could also account for the smaller volume for the carboxyl oxygen: the different steric environments of the carboxyl and carbonyl oxygen atoms. The carboxyl oxygen usually projects directly out into solution, at the end of a flexible side-chain, while the packing geometry around the carbonyl oxygen is more constrained by its direct connection to the protein main-chain and its close proximity to the β-carbon. Thus, the smaller volume for the carboxyl oxygen in comparison to the carbonyl oxygen could also be explained by the ability of water molecules to more easily solvate it.

distinct physical processes. On one hand, the larger volume of, say, a typical surface atom (in comparison to a corresponding core atom) could arise because this atom is physically bigger on the surface. On the other hand, this larger volume could result from water molecules not packing tightly against the surface atom, giving rise to cavity volumes, which then, in turn, could be allocated to the surface atom. This distinction is to some degree a matter of semantics since it is contingent upon how we choose to define a cavity and allocate its volume.

In doing our calculations and interpreting our results, we have tried to take as simple as possible an approach. In our calculations, we have not tried to define (arbitrarily) the positions and sizes of cavities. Rather, we have let the Voronoi procedure naturally allocate the volume of a cavity to atoms neighboring it. Thus, in reporting our results, we make no mention of cavities but rather indirectly describe their presence or absence in terms of larger or smaller atomic volumes. However, in interpreting our results, we assume that there is no physical process changing the size of atoms on the protein surface (in the way, for instance, that the loss of an electron decreases the size of a sodium ion in comparison to a neutral sodium atom), so we explain all the volume changes we observe in terms of tighter or looser packing.

Consequently, we interpret our principal result that surface hydrophobic atoms and their associated water molecules are larger to imply that hydrophobic atoms on the protein surface and their associated water molecules are less tightly packed (than corresponding atoms in the protein core and bulk water). Conversely, we interpret the smaller volumes of surface charged atoms and their neighboring water molecules to imply tighter packing.

Results

Results from simulation

In performing the volume calculations on the simulations, we first compared the volumes of representative types of atoms on the protein surface with those in the protein core (Table 1). Except for charged oxygen and nitrogen atoms, all the atoms on the surface are larger than in the core by $\sim$6%. The volume of charged atoms decreases, but it is difficult to accurately make an assessment of the magnitude of this decrease due to the small number of buried or partially buried charged atoms. The decrease in size of an atom as it becomes more highly charged is also borne out by the smaller volume of the carboxyl oxygen in comparison to the carbonyl

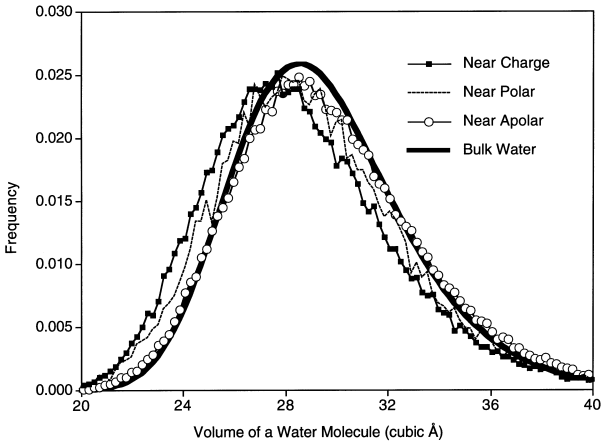

Figure 3. Distribution of water volumes in bulk and near the protein. The distribution of water volumes in the four different classes is shown: in the bulk and near protein atoms that are polar, apolar, or charged. The centerpoint and width of these distributions are directly related to the volume and compressibility of these four classes of water molecules. The mean volumes of the four classes of water molecules are: bulk, 29.68 Å^3; near charged protein atoms, 28.75 Å^3; near polar protein atoms, 29.16 Å^3; and near apolar protein atoms, 29.86 Å^3. The compressibilities for these four classes are shown in Table 4.

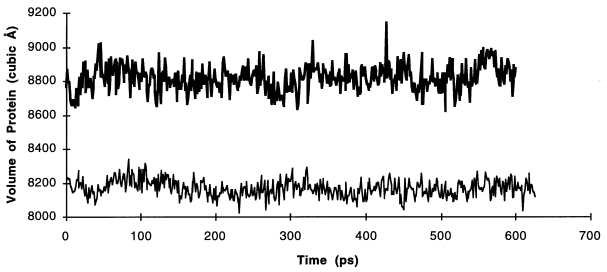

Figure 4. Volume of trypsin inhibitor over time in the normal-pressure (top) and high-pressure (bottom) simulations

oxygen, since these atoms differ only in charge, and not in hybridization or number of attached hydrogen atoms. (As shown in Table 1, for carbonyl and carboxyl oxygen atoms on the surface, the difference is 18.1 Å^3 *versus* 13.2 Å^3, respectively.)

Compensating for the larger size of most of the atoms on the protein surface, water molecules around the protein are, overall, slightly smaller than those in the bulk (Figure 3). In particular, water molecules near polar groups are smaller in volume by $\sim$0.5 Å^3 and those around charged groups are even smaller (by $\sim$1 Å^3). However, water molecules near hydrophobic groups are slightly larger in volume (by $\sim$0.2 Å^3, as compared to those in the bulk).

The way the overall protein volume fluctuates over the trajectory is shown in Figure 4. Neither the mean volume nor the magnitude of the fluctuations changes much, indicating that the simulations were stable. Values for the fluctuations of particular classes of protein and water atoms are shown in Table 4. Atoms on the protein surface have greater fluctuations in volume than atoms buried in the protein core, giving them on average greater compressibilities by a factor of 2.1. Similarly, water molecules near the protein surface have larger volume fluctuations (by $\sim$20%) than those in bulk solvent. Figure 5 shows how the atomic volume fluctuations vary over the protein surface. Atoms with smaller fluctuations preferentially occur on protruding ''ridges'', reflecting to some degree the fact that the charged atoms mostly reside on ridges.

In the normal pressure simulation, the volume fluctuations of water molecules are about twice as large as those of protein atoms. This translates into a difference in compressibility that is strikingly manifest in the high-pressure simulation. As shown in Table 4, at high pressure, the volume of the protein core decreases by only 5% while the bulk water shrinks by 13%. The atoms on the protein surface decrease in volume by an amount between these extremes (8%), and in accord with their large volume fluctuations, the water molecules near the protein surface shrink by 14%, on average.

The decreases in volume evident in the high pressure simulation make it possible to compare the average compressibilities calculated from the definition ($dV/V\,dP$, see Methods) with specific compressibilities calculated at normal pressure from volume fluctuations. As shown in Table 4, we find that the compressibilities calculated in both ways are consistent. In particular, the pattern of higher and lower compressibilities for different atoms calculated from the definition is almost perfectly correlated (with a correlation coefficient of 0.91) with that calculated from fluctuations.

Results relating to the Voronoi procedure

Our results, both from crystal structures and from simulation, depend on the particular way we chose to partition the volume between neighboring atoms. We have used method B in conjunction with the radii of Chothia (1975) to position the dividing plane between different atoms. We feel these are reasonable choices. Nevertheless, here we will discuss some aspects of these choices and explore possible alternatives.

One problem with method B is vertex error. We would, moreover, expect the problem of vertex error to be greatest at the protein surface, where water molecules pack onto hydrophobic atoms, since the water radius of 1.4 Å is one of the smallest radii in comparison to that of aliphatic carbons (1.87 Å). Since the volume of the simulation box is necessarily fixed, we can directly calculate the vertex error from comparing its volume to the total Voronoi volume of its constituents. We find that it is 46 Å^3 out of 86,152 Å^3. As there is no vertex error in calculating polyhedra for the bulk water (since the dividing planes will always be positioned midway between the water molecules), all the vertex error is concentrated in the protein and in water near the protein ($\sim$22,000 Å^3). Thus, we estimate the vertex

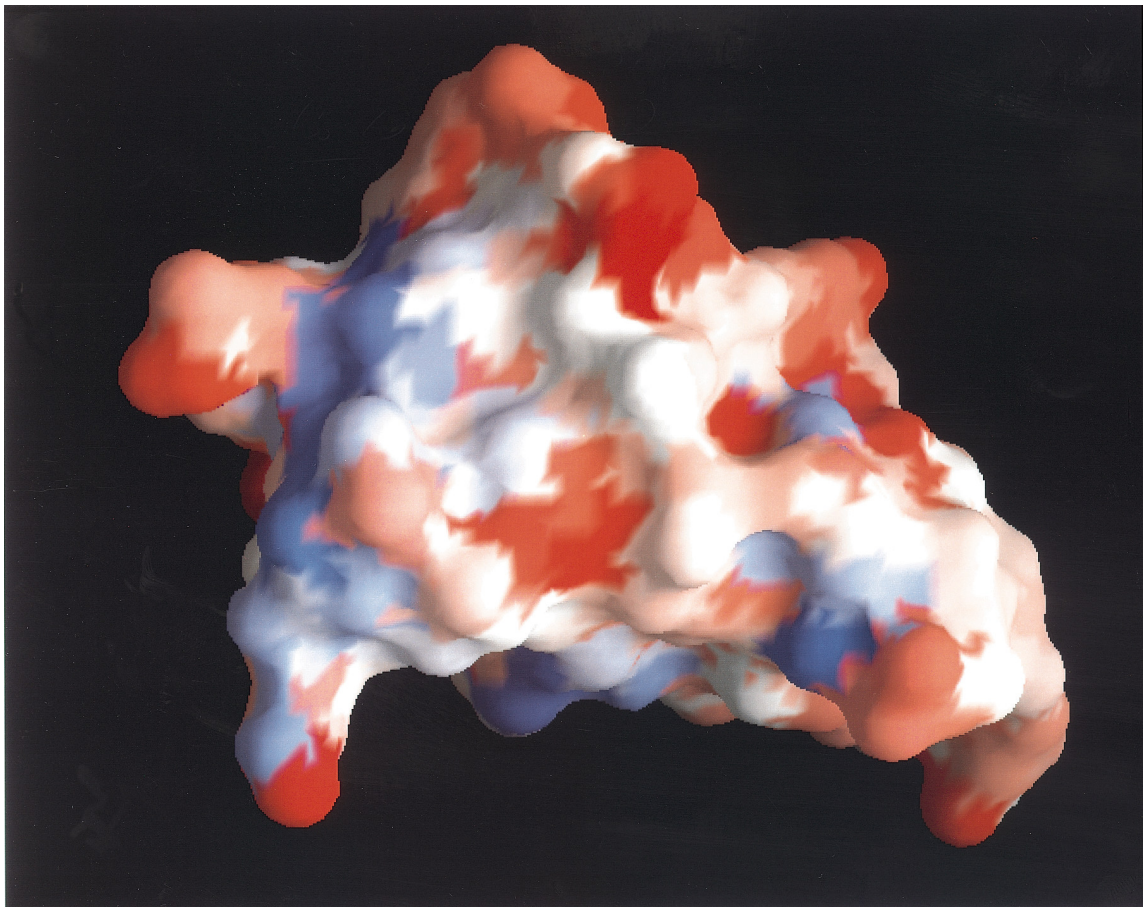

Figure 5. Volume fluctuations of surface atoms over the simulation. The surface of trypsin inhibitor is colored according to the fluctuations in atom volumes over the simulation. As discussed in Methods, the amplitude of an atom's volume fluctuation is directly related to the compressibility of that atom. Atoms with large volume fluctuations are shown in blue and those with small fluctuations are shown in red. To some degree the atoms on protruding ''ridges'' have smaller volume fluctuations than other atoms on the protein surface. This suggests that water molecules may be able to pack in a more consistent and efficient manner around these protruding atoms. The Figure was drawn with GRASP (Nicholls *et al*., 1991).

error to be only about one part in 500 and not a significant factor in our calculations.

One obvious alternative to using method B is to use the original Voronoi method (bisection) to position the dividing plane. We found, as others before us (Richards, 1974; Gellatly & Finney, 1982), that bisection is not suitable for calculating volumes in the protein core. Bisection systematically misallocates volume inside of the protein, producing a larger variance in the volume for any particular atom type. However, we found that a hybrid approach, using method B to position the plane between protein atoms and bisection in all other cases (i.e. in water–water and protein–water contacts), worked quite well and may be useful for future investigations. Our reasoning for advocating this hybrid approach will become clear later after our discussion of atom radii. Overall we found that the hybrid approach transfers ~10% of the protein volume to the water.

Another alternative, perhaps not so obvious, would be to keep the spirit of the chemically reasonable partitioning implied by method B but not to use Voronoi polyhedra. That is, we tried generalizing method B to use more elaborate shapes than polyhedra for dividing the space between atoms. Because method B seeks to partition space between two neighboring atoms in proportion to the ratio of their radii and a constant ratio of distances to two different atoms cannot be maintained on a dividing plane, using a dividing boundary other than a plane may actually be more in keeping with the spirit of method B. As described in Figure 2B, we have replaced the dividing plane by ''a curved dividing surface,'' constructed so as to keep the distance ratio to the two neighboring atoms constant. Using our new construction, we estimated that method B, in a sense, overstates the volume of a small atom, such as a water oxygen, by ~7% when it is in contact with a larger atom, such as a methyl carbon. This volume differential is in the opposite direction to that observed for atoms on the protein surface (where large atoms such as aliphatic carbon atoms get bigger and small atoms such as the water oxygen get smaller). Thus, a more elaborate partitioning than Voronoi polyhedra would, if anything, tend to accentuate the volume increase we found for surface atoms.

Method B requires the assignment of atom radii to various atom types. The radii are essentially *a priori*

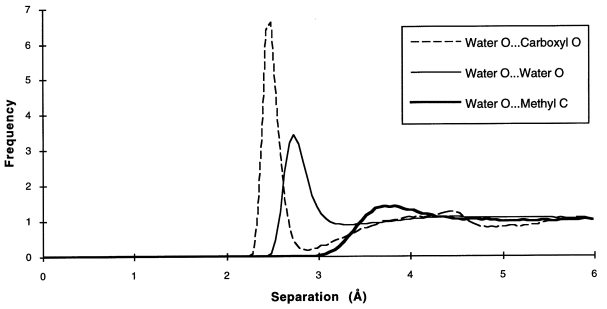

Figure 6. Representative distributions of separations involving water in the simulation. The distribution of interatomic distances, collected over the simulation, between a water oxygen and three types of atoms (a methyl carbon, another water oxygen, and a carboxyl oxygen) are shown. The most likely separation between water oxygens is 2.75 Å, and half of this separation is usually used for the water VDW radius. Since this VDW radius is essentially being determined by water–water hydrogen bonding, there is a potential for inconsistency when it is used to describe atoms interacting in a different fashion. Such inconsistencies are evident for water interacting with atoms on the protein surface. The most likely separation between a water oxygen and methyl carbon (—CH3) is more than the sum of their VDW radii (3.75 Å *versus* 1.4 + 1.87 Å), and the most likely separation between water and a carboxyl oxygen (—O) is less than the sum of their VDW radii (2.45 Å *versus* 1.4 + 1.4 Å).

sizes for atoms. In particular, the value used for the water radius is most critical in determining the overall partitioning between protein and solvent. Increasing its value from the usual 1.4 Å has the effect of ''transferring volume'' from the protein to the surrounding water molecules. In fact, we found as the water radius (R) is increased from 1 Å to 2.5 Å, the protein volume (in Å^3) decreases in perfectly linear fashion, according to the relation: $V = 8753 - 1965(R - 1.4)$.

We have investigated the appropriateness of the 1.4 Å radius for water as well as of the radii for other atoms through the calculation of distributions of interatomic distances in our PTI simulation. We have calculated the distribution of non-bonded distances between any two types of atoms i and j. We show three representative distance distributions involving water molecules in Figure 6. The first peak in these distributions represents the preferred separation between an atom of type i and another of type j. It should occur at the sum of the van der Waals (VDW) radii.

We find that the usual 1.4 Å value for the water radius is a tenuous compromise at best. As shown in Figure 6 and Table 2, around hydrophobic atoms water molecules are farther from the protein than the sum of VDW radii would suggest. For instance, for water around a methyl carbon the preferred separation is ~0.6 Å greater than the sum of the VDW radii, 1.4 Å + 1.87 Å. This larger than ''expected'' distance is, of course, the origin of the volume increase for both hydrophobic atoms on the protein surface and water molecules contacting them.

Table 2. Comparison of ''expected'' and observed interatomic spacings involving water

A. *Separation between water and all atom types*

Separation from water oxygen (OW) to . . .		Most likely distance in simulation (Å)	Difference from sum of VDW radii (Å)
CA,	>CH—	3.93	+0.66
C,	>C═	3.78	+0.62
—S—		3.80	+0.53
—CH═		3.68	+0.52
—CH2—	—CH3	3.73	+0.46
N,	—NH2	2.98	–0.07
—NH3		2.83	–0.07
OW		2.73	–0.07
—OH		2.68	–0.13
O,	═O	2.63	–0.18
—O		2.48	–0.33

B. *Least-squares fits to observed separations*

For interactions between water and . . .	The best water radius is . . . (Å)
Apolar atoms	1.96
Polar atoms, including water	1.29
The carboxyl oxygen	1.08

In A, we tabulate the most likely separations between water and all atom types observed in the simulation and compare them to the expected distance from the sum of their VDW radii. The differences with respect to the VDW radii fall into three categories. All the apolar atoms (carbon and sulfur atoms) are further away than expected, all the polar atoms (nitrogen and oxygen) beside the carboxyl oxygen are approximately at the distance expected, and the carboxyl oxygen is closer than expected. These three categories are the basis for our suggestion that three different radii be used for the water molecule depending on whether it is interacting with an apolar, polar, or carboxyl atom. As shown in B, we do a least-squares fit to the observed separations to find three suitable radii. To illustrate the improvement in the fit with three as opposed to one water radii we calculated the following r.m.s. ''residual'':

$$\sqrt{\frac{\sum_{i=1}^{N} (D_i - (w_i + r_i))^2}{N}}$$

where D_i is an observed separation from the above Table, w_i is the appropriate water radius (either 1.4 Å or one of the three above), and r_i is the VDW radius (from Table 4) for the atom contacting the water. We find the r.m.s. residual decreases from 0.42 to 0.11 Å when we switch from using one to three water radii.

The opposite situation is observed around charged atoms, where water molecules are closer than expected.

To avoid this compromise between extremes, we find that we can get a much better fit to the distributions (decreasing the ''residual'' by a factor of 4) if we use three different radii for water: one for interactions with apolar atoms, a second for interactions with polar atoms, and a third for interactions with the highly charged carboxyl oxygen (Table 2).

We have also tried to use our distributions of interatomic distances to determine a set of effective VDW radii from simulation. Unfortunately, we could not get good enough statistics between many of the protein atoms to determine a complete set. We also found that this calculation was complicated by the way the geometry of the polypeptide

Table 3. Effective VDW radii derived from potential function parameters

Atom types	Radii (Å) from Table 5 [a]	Effective radii (Å) derived from potential ENCAD [b]	CHARMM [c]
C, >C═, —CH═	1.76	1.74	1.80
>CH—, —CH2—, —CH3	1.87	1.82	1.88
N	1.65	1.68	1.40
—NH3	1.50	1.68	1.40
O, ═O	1.40	1.34	1.38
OW	1.40	1.54	1.53
—O	1.40	1.34	1.41
—S—, —SH	1.85	1.82	1.56

[a] As discussed in Table 5, for all our Voronoi calculations we used the radii of Chothia (1975). These radii were derived from analysis of the packing and geometry of small-molecule crystal structures.

[b] For comparison with the radii of Chothia (1975), we have determined a set of radii based solely on the ENCAD molecular dynamics potential function (Levitt *et al*., 1994) used for the calculations. To determine these radii, we calculated the separation at which the Lennard-Jones interaction energy U between equivalent atoms was 0.25 k_BT (0.15 kcal/mol).

$$U = \frac{A}{r^{12}} - \frac{B}{r^6} = \frac{k_BT}{4}$$

Then after substituting the values for A and B from the ENCAD potential, we calculated a value for r for each atom type, which we report in the Table. (Note, in the ENCAD potential $A = \epsilon r_0^{12}$ and $B = 2\epsilon r_0^6$.) The r.m.s. residual between the effective radii in this column and the radii from Chothia (1975) (in column [a]) is 0.09 Å.

[c] For further comparison, we have also determined radii based on a second molecular dynamics potential function, the CHARMM potential (Brooks *et al*., 1983). The values for A and B for this comparison were taken from CHARMM parameter set 19, as used in X-PLOR (Brunger *et al*., 1987). (The CHARMM parameters have a slightly different form from those in ENCAD, so $A = 4\epsilon\sigma^{12}$ and $B = 4\epsilon\sigma^6$.) The r.m.s. residual between the effective radii in this column and the radii from Chothia (1975) (in column [a]) is 0.15 Å, and the correlation coefficient with the radii derived from the ENCAD potential is 0.62.

chain fixes atoms at preferred distances irrespective of their VDW radii (e.g. the relative location of the carbonyl oxygen and the β-carbon has more to do with secondary structure than VDW radii). Consequently, we have found it better to calculate a set of effective VDW radii directly from the Lennard-Jones parameters used in the ENCAD potential. This is shown in Table 3. These effective radii are clearly similar to the radii set of Chothia (1975) that we used for our Voronoi calculations. For comparison, we have also calculated a set of effective radii based on another molecular-dynamics potential, the CHARMM potential. This radii set tends to give smaller radii to the uncharged nitrogen and sulfur.

Discussion

Volume changes

Our Voronoi volume calculations showed that most of the atoms on the protein surface are larger than those in the protein core by ~6%. The simulations provided further detail on the volume of the water molecules themselves, which expanded around hydrophobic atoms on the protein surface and contracted around charged atoms. Fundamentally, the expansion of hydrophobic atoms on the protein surface and the water around them is a result of the greater than expected distance of water molecules from these atoms. This larger separation has been pointed out before (Levitt & Sharon, 1988; Madan & Lee, 1994; Williams *et al*., 1994) and is clearly evident in the distributions of interatomic distances calculated from the simulation. Whether the volume increase is allocated preferentially to the water or the protein is a consequence of the method of positioning the dividing plane and the particular VDW radii used.

Two competing volumetric effects influence the size of hydrophobic atoms on the protein surface. On one hand, when one unit of an apolar liquid, such as hexane, is added to one unit of water, the total volume of the resulting mixture is less than the sum of its parts, i.e. it is less than two units (Franks, 1983). This volume decrease is usually explained in terms of the apolar molecule fitting into the cage-like interstices of the much more open structure of liquid water. On the other hand, as has been pointed out before (Finney, 1975), the packing at the protein surface is expected to be less than optimal since this is where two chemically different species meet. Furthermore, the convoluted and irregular shape of the protein surface is expected to make it all the more difficult for water to pack tightly around it. Simulation studies (Gerstein & Lynden-Bell, 1993a,b), in fact, have directly shown how water–water hydrogen bonding can prevent water molecules from packing well in crevices on the protein surface.

Clearly, our finding that the protein surface is not packed as tightly as the protein core is more in consonance with the second effect. Furthermore, the large volume fluctuations we have observed for both the water and protein atoms at this interface (i.e. the atoms on the protein surface and the water molecules in contact with them) provide further evidence for loose packing at the protein surface. These large fluctuations suggest that water molecules do not easily form stable, well-packed structures around the protein surface. They are also consistent with the experimental observation that water is more compressible around apolar solutes than in the bulk (Gekko & Noguchi, 1974).

The only atoms that do not increase in size on the protein surface are the charged atoms, which actually shrink. The reduction in volume of charged atoms is matched by a volume contraction of the water around them. These coupled volume reductions, in turn, are a direct consequence of the short average separations between water and charged atoms. The volume decrease of charged atoms in solution is a well-established consequence of the Drude-Nernst theory of electroconstriction (Drude & Nernst, 1894; Kauzmann *et al*., 1962),

which predicts that the degree of contraction around an ion should be roughly proportional to the square of its charge.

Our two findings that atoms on the surface are not packed as tightly as those in the core and that there is electroconstriction around charged atoms are consistent with the results of Harpaz *et al*. (1994). Harpaz *et al*. compared the volumes of core residues in proteins with those for amino acids in solution and found that aliphatic groups occupy more volume in solution than in the protein interior and that the converse was true for charged and amide groups.

They also showed that the protein interior is packed exceptionally tightly, with a greater packing density than that in organic crystals. They argued that this tight packing can occur because the core structure of a protein does not have to satisfy ''external'' constraints, such as the geometry of a crystal lattice. That is, the protein may sacrifice packing at its surface for optimal packing in its core. Thus, considering the exceptionally tight core packing found by Harpaz *et al*., it appears doubtful whether the protein surface could even possibly be packed as well as the core. Furthermore, comparison of the tight core packing, with the looser packing in apolar solids and, obviously, liquids, is in itself enough to suggest that when a protein unfolds, it will not exhibit a volume decrease similar to that found when an apolar liquid, such as hexane, is mixed with water (as discussed above).

Methodological issues

We have shown how our results are qualified by issues related to the construction of Voronoi polyhedra. That is, we show how volume is re-apportioned (in method B) by the choice of atom radii, in particular that of the water radius. This is especially important because the usual 1.4 Å water radius is not a proper VDW radius; rather, it reflects the hydrogen bond spacing in pure water. We show that the observed distributions of interatomic distances in the simulation can be much better accounted for if we use three radii for water: one for polar interactions, one for apolar interactions, and one for interactions with the highly charged carboxyl oxygen. Using such multiple radii for water has been suggested before (Williams *et al*., 1994; Savage & Finney, 1986).

Conceptually, the Voronoi calculations using Richards' method B could be adapted to use these extra water radii since method B already makes a distinction between two different atom radii (i.e. covalent and VDW radii). This adaptation would tend to transfer volume to the water from the protein and would thus cause the water around hydrophobic groups to appear even larger. However, on a practical level, this adaptation would add further complexity and ambiguity into an already involved calculation. A simpler alternative may be to use our hybrid method (method B to position the dividing plane between protein atoms and bisection to position the dividing plane involving water molecules). This would achieve virtually the same effect as adding extra water radii since the extra radii are scaled to be roughly the same as that of the atom the water is interacting with. On balance, we feel this is the best way to perform volume calculations involving the protein surface. If one is not constrained to be consistent with past calculations, we would recommend this modification of method B for future calculations on the volume of the protein surface.

Methods

Simulation setup and volume calculations

For the simulation studies, pancreatic trypsin inhibitor (the 4PTI crystal structure, Marquart *et al*., 1983) was immersed in a box of 2600 F3C waters (Levitt *et al*., unpublished results) and six chlorides at 298 K. Two simulations at constant volume were done: a normal pressure simulation, lasting 609 ps, in a box of dimensions 48.336 Å × 42.290 Å × 42.041 Å; and a high-pressure simulation, lasting 634 ps, in box of dimensions 46.324 Å × 40.530 Å × 40.291 Å. The dimensions of the simulation boxes were chosen so that the overall densities of the system were consistent with pressures of 1 atm and 5000 atm, respectively (Vedam & Holton, 1967; Grindley & Lind, 1971). The ENCAD program and potentials were employed (Levitt *et al*., 1994), and these simulations used the same methodology and conditions as a number of previous PTI simulations (Daggett & Levitt, 1992; Levitt & Sharon, 1988). The protein took up only a tenth of the volume of the simulation box (Table 4). Periodic boundary conditions were used throughout, both for the simulation and for the Voronoi calculations.

All the Voronoi calculations were done with a newly written C-language program that was directly based on a program of Richards (version of 6 March 1983 by M. D. Handschumacher & F. M. Richards; the Voronoi algorithm is most clearly described by Richards, 1985). As discussed in Table 5, mostly following the conventions of Richards (1974), we grouped the protein atoms into 20 different types, and we used atom radii taken from Chothia (1975). As shown in Figure 2A, we used Richards' method B to position the plane between atoms of different size (Richards, 1974).

We also used Voronoi polyhedra to determine whether atoms were in contact. In particular, we considered a water molecule to be ''near'' a protein atom if it shared a Voronoi polyhedron face with that atom. When a water molecule shared faces with multiple types of protein atoms, it was considered near the type it shared the most faces with. A bulk water molecule only shared faces with other water molecules.

We checked our calculations with a ''poor man's'' Voronoi procedure. We sampled the space in the simulation box according to a three-dimensional grid. The volume of each grid ''voxel'' was assigned to the atom it was closest to. To get accurate numbers, we needed to use a very fine grid, but this calculation could be made significantly faster by using Monte-Carlo integration and sampling the grid randomly. We also used this poor man's procedure for calculating volumes bounded by non-planar surfaces (as described in Figure 2B).

Table 4. Volumes of parts of the simulation box at two different pressures

Components[a]	Average volume 1 atm[b]	Average volume 5000 atm[c]	Change[d] (%)	Volume fluctuation S.D.[e] (%)	β[f]
Whole Box	86,107	75,600	−12	11.9	0.40
Protein	8810	8167	−7	10.9	0.24
Core	2527	2409	−5	9.3	0.14
Surface	6283	5759	−8	11.7	0.29
Charged	590	546	−7	9.6	0.19
Polar	1447	1337	−8	12.4	0.27
Apolar	4245	3875	−9	11.7	0.32
Chloride	281	260	−7	6.6	0.20
Water	77,017	67,173	−13	12.1	0.42
Bulk	60,775	53,236	−12	11.9	0.41
Cl Shell	3001	2538	−15	11.8	0.40
Near Protein	13,241	11,398	−14	13.2	0.50
Near Charged	4484	3739	−17	13.2	0.50
Near Polar	1367	1200	−12	12.9	0.49
Near Apolar	7390	6459	−13	13.3	0.51

[a] The first column shows how the simulation box can be divided into various components. The volume of the whole box is the sum of the protein, chloride, and water volumes, the protein volume is the sum of the core and surface volumes, and the water volume is the sum of the bulk, chloride shell, and ''near protein'' volumes. The volume of the near protein water and that of the protein surface can be further divided as shown.

[b, c] The second and third columns show average volumes (in Å^3) calculated from the normal and high pressure simulations. These are total volumes of the protein, water, and so forth.

[d] The percentage change in total volume between the normal and high-pressure simulations. As described in Methods, this quantity is closely related to the definition of compressibility, i.e.:

$$P\beta = \left(\frac{V_{5000} - V_1}{\frac{1}{2}(V_{5000} + V_1)} \right).$$

[e] The second to last column shows the average atomic volume fluctuations in the 1 atm simulation in terms of the standard deviation over the simulation divided by the mean:

$$\frac{\sqrt{\langle \Delta V^2 \rangle}}{\langle V \rangle}.$$

This quantity reflects the average fluctuation in atomic volume for atoms in a particular group and not the fluctuation in their total volume (that is, it shows the average fluctuation for a protein atom and not for the protein as a whole). Consequently, it reflects two separate stages of averaging. For each distinct atom j, the mean volume is first determined by ''time-averaging'' over all frames i in the simulation. Then an ''ensemble-average'' is determined for all atoms in the same group (i.e. all protein atoms):

$$\langle V \rangle = \left\langle \langle V_{ij} \rangle_i \right\rangle_{j \text{ in a group}}.$$

This two-step averaging procedure is likewise used to calculate the variance:

$$\langle \Delta V^2 \rangle = \left\langle \langle V_{ij}^2 \rangle_i - \langle V_{ij} \rangle_i \right\rangle_{j \text{ in a group}}.$$

[f] The last column shows the average atomic volume fluctuations in the 1 atm simulation expressed in slightly different way from that in column 5, in terms of the variance over the simulation divided by the mean:

$$\frac{\langle \Delta V^2 \rangle}{\langle V \rangle}.$$

This quantity, which is expressed in cubic Å, is averaged in a similar two-step fashion to the S.D. in column 5. While it (obviously) exhibits similar behavior to the S. D., it is shown here because it is directly related to the compressibility (i.e. it is $\beta k_B T$; see Methods). To convert the values in this column into compressibilities in usual units, one needs to divide them by the appropriate value of $k_B T$, 40,900 Å^3 atm. This will give the compressibilities in units of 1/atm. After performing this conversion, we find that the compressibilities calculated from the fluctuations are consistent with those calculated from the definition (i.e. those in column 4, marked change). That is, the correlation coefficient between the values in this column and those in column 4 is 0.91. Furthermore, for bulk water, we find that the actual value of the compressibility calculated from fluctuations is about 40% of that calculated from the definition. This is in quite reasonable agreement considering that we are comparing an ''instantaneous'' value to an ''average'' value.

Volume fluctuations and compressibility

The adiabatic compressibility β is defined as the fractional change in volume for a given change in pressure at constant temperature:

$$\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T$$

(where V is the volume, P is the pressure, and T is the temperature). The fluctuation of volume over a

Table 5. The atom types and radii used for the calculation

Atom types[a]		Radii[b]
Main-chain atoms		
Oxygen	O	1.40
Carbonyl carbon	C	1.76
Nitrogen (not P)	N	1.65
Alpha carbon (not G or P)	CA	1.87
Pro N	NPRO	1.65
Gly CA	CGLY	1.87
Pro CA	CPRO	1.87
Side-chain atoms		
Trigonal or aromatic carbon	>C═	1.76
Aromatic CH (H, F, W, Y)	—CH═	1.76
Aliphatic CH	>CH	1.87
Methylene group	—CH2—	1.87
Methyl group (A, V, L, I)	—CH3	1.87
2° amine (R, H, W)	—NH—	1.65
Amino or amide (R, N, Q)	—NH2	1.65
Tetrahedral nitrogen (K)	—NH3	1.50
Carboxyl oxygen (D, E)	—O	1.40
Carbonyl oxygen (N, Q)	═O	1.40
Hydroxyl (S, T)	—OH	1.40
Thioether or disulfide (C)	—S—	1.85
Sulfhydryl (C)	—SH	1.85
Water oxygen		
Overall radius	OW	1.40

[a] Considering each atom in each type of residue distinct (and considering cysteine chemically different from cystine) gives 176 different types of atoms. Basically following the atom typing of Richards (1985), we have grouped these into the 20 classes shown above. The only modification we have made is keeping main-chain and side-chain atoms distinct throughout the calculation.

[b] The VDW radius (in Å) assigned to each atom type. For all our volume calculations, we use a radii set derived from Chothia (1975). This set has six distinct types of radii: tetrahedral carbon atoms, 1.87 Å; trigonal and aromatic carbon atoms atoms, 1.76 Å; oxygen atoms (including water), 1.4 Å; sulfur atoms, 1.85 Å; tetrahedral nitrogen atoms (as in lysine), 1.5 Å; and trigonal nitrogen atoms, 1.65 Å.

constant-pressure simulation can be related to this compressibility:

$$\beta = \frac{\langle \delta V^2 \rangle}{k_B T \langle V \rangle}, \qquad (1)$$

(where k_B is Boltzmann's constant).

This relationship does not hold rigorously for the constant volume conditions under which we carried out our simulations. However, we show that it holds approximately for small volumes. Consider a single atom of volume V_1 and compressibility β_1 fluctuating in volume in the simulation box. A volume change of δV for it has to be matched by an equal but opposing volume change in the rest of the system. The energy associated with this volume change can be found by integrating the pressure-volume work $P\,dV$ over the volume change and using the definition of compressibility to express P in terms of V:

$$U = \int_{V_1}^{V_1+\delta V} P\,dV = \int_{V_1}^{V_1+\delta V} \frac{V - V_1}{\beta_1 V_1}\,dV = \frac{\delta V^2}{2\beta_1 V_1} \qquad (2)$$

Analogously the energy associated with the volume change for the rest of the system is

$$\frac{\delta V^2}{2\beta_r V_r}$$

(where V_r and β_r are the relevant volume and compressibility). This is a situation similar to that of two coupled springs (or pistons), where the spring constant is $1/\beta V$. Since these two energies are associated with only a single degree of freedom, equipartition requires that their sum, averaged over the simulation, be $\frac{1}{2}k_B T$. Performing this sum and rearranging, we find that

$$\frac{k_B T}{\langle \delta V^2 \rangle} = \frac{1}{\beta_1 \langle V_1 \rangle} + \frac{1}{\beta_r \langle V_r \rangle} \qquad (3)$$

This equation applies to any partition of the volume V_1 and V_r. However, when V_r is much larger than V_1 (as when comparing one atom to the rest of the simulation box), this formula reduces to the constant-pressure formula for β (equation (1), above).

Availability of results on the Internet

We make available on the Internet, code for calculating Voronoi polyhedra (source and executables), data associated with this paper (i.e. standard volumes of buried atoms in proteins), and some explanations of Voronoi polyhedra in hypertext form. These items can be retrieved by sending e-mail to mbg@hyper.stanford.edu or levitt@hyper.stanford.edu or through using anonymous ftp or the World Wide Web with the following URLs:

ftp://hyper.stanford.edu/pub/mbg/SurfaceVolumes/
http://hyper.stanford.edu/~mbg/SurfaceVolumes/

Acknowledgements

M.G. acknowledges support from the Damon Runyon-Walter Winchell foundation (grant DRG-1272). M.L. and J.T. acknowledge support from the NIH (grant GM 41455). We thank all of our reviewers for helpful suggestions and A. Franke for a careful reading of the manuscript.

References

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187–217.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science*, **235**, 458–460.

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

Daggett, V. & Levitt, M. (1992). A model of the molten globule state from molecular dynamics simulations. *Proc. Nat. Acad. Sci. USA*, **89**, 5142–5146.

David, C. W. (1988). Voronoi polyhedra as structure probes in large molecular systems. *Biopolymers*, **27**, 339–344.

Drude, P. & Nernst, W. (1894). Über elektrostriktion durch freie ionen. *Z. physik. Chem.* **15**, 80–85.

Finney, J. L. (1975). Volume occupation, environment and accessibility in proteins. The problem of the protein surface. *J. Mol. Biol.* **96**, 721–732.

Finney, J. L. (1978). Volume occupation, environment, and accessibility in proteins. Environment and molecular area of RNase-S. *J. Mol. Biol.* **119**, 415–441.

Finney, J. L., Gellatly, B. J., Golton, I. C. & Goodfellow, J. (1980). Solvent effects and polar interactions in the structural stability and dynamics of globular proteins. *Biophys. J.* **32**, 17–33.

Franks, F. (1983). *Water*. The Royal Society of Chemistry, London.

Gekko, K. & Noguchi, H. (1974). Hydration behavior of ionic dextran derivatives. *Macromolecules*, **7**, 224–228.

Gellatly, B. J. & Finney, J. L. (1982). Calculation of protein volumes. An alternative to the Voronoi procedure. *J. Mol. Biol.* **161**, 305–322.

Gerstein, M. & Lynden-Bell, R. M. (1993a). Simulation of water around a model protein helix. 2. The relative contributions of packing, hydrophobicity, and hydrogen bonding. *J. Phys. Chem.* **97**, 2991–2999.

Gerstein, M. & Lynden-Bell, R. M. (1993b). What is the natural boundary for a protein in solution? *J. Mol. Biol.* **230**, 641–650.

Gerstein, M., Lesk, A. M., Baker, E. N., Anderson, B., Norris, G. & Chothia, C. (1993). Domain closure in lactoferrin: two hinges produce a see-saw motion between alternative close-packed interfaces. *J. Mol. Biol.* **234**, 357–372.

Grindley, T. & J. E. Lind, J. (1971). PVT properties of water and mercury. *J. Chem. Phys.* **54**, 3983–3989.

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

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

Kauzmann, W., Bodanszky, A. & Rasper, J. (1962). Volume changes in protein reactions. II. Comparison of ionization reactions in proteins and small molecules. *J. Amer. Chem. Soc.* **84**, 1777–1788.

Levitt, M. & Park, B. H. (1993). Water: now you see it now you don't. *Structure*, **1**, 223–226.

Levitt, M. & Sharon, R. (1988). Accurate simulation of protein dynamics in solution. *Proc. Natl. Acad. Sci. USA*, **85**, 7557–7561.

Levitt, M., Hirschberg, M., Sharon, R. & Daggett, V. (1994). Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Comp. Phys. Comm*. In the press.

Madan, B. & Lee, B. (1994). Role of hydrogen bonds in hydrophobicity: the free energy of cavity formation in water models with and without the hydrogen bonds. *Biophys. Chem.* **51**, 279–289.

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

Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.

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

Richards, F. M. (1977). Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.

Richards, F. M. (1979). Packing defects, cavities, volume fluctuations, and access to the interior of proteins. Including some general comments on surface area and protein structure *Carlsberg. Res. Commun.* **44**, 47–63.

Richards, F. M. (1985). Calculation of molecular volumes and areas for structures of known geometry. *Methods Enzymol.* **115**, 440–464.

Saenger, W. (1987). Structure and dynamics of water surrounding biomolecules. *Annu. Rev. Biophys. Biophys. Chem.* **16**, 93–114.

Savage, H. & Finney, J. L. (1986). Repulsive regularities of water structure in ices and crystalline hydrates. *Nature*, **322**, 717–720.

Shih, J. P., Sheu, S. Y. & Mou, C. Y. (1994). A Voronoi polyhedra analysis of structures of liquid water. *J. Chem. Phys.* **100**, 2202–2212.

Teeter, M. M. (1991). Water-protein interactions: theory and experiment. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 577–600.

Vedam, R. & Holton, G. (1968). Specific volumes of water at high pressures, obtained from ultrasonic-propagation measurements. *J. Acous. Soc. Am.* **43**, 108–116.

Voronoi, G. F. (1908). Nouveles applications des paramétres continus à la théorie des formes quadratiques. *J. Reine Angew. Math.* **134**, 198–287.

Williams, M. A., Goodfellow, J. M. & Thornton, J. M. (1994). Buried waters and internal cavities in monomeric proteins. *Protein Sci.* **3**, 1224–1235.

Edited by F. Cohen

(Received 30 January 1995; accepted in revised form 3 April 1995)