Protein Geometry: Volumes, Areas, and Distances

M Gerstein & F M Richards

Department of Molecular Biophysics & Biochemistry
266 Whitney Avenue, Yale University
PO Box 208114, New Haven, CT 06520

A Manuscript for inclusion in:
The International Tables for Crystallography

Volume F: Macromolecular Crystallography
Chapter 22: Molecular Geometry and Features
Editors: M G Rossmann and E Arnold

Manuscript is 28 Pages in Length (including this one)
Version: fr318 (file ...0\fmr\final\geom-inttab.doc)

Introduction

For geometrical analysis, a protein consists of a set of points in three dimensions. This information corresponds to the actual data provided by experiment, which is fundamentally of a geometrical rather than chemical nature. That is, crystallography primarily tells one about the positions of atoms and perhaps an approximate atomic number, but not their charge or number of hydrogen bonds.

For the purposes of geometrical calculation, each point has an assigned identification number and a position defined by three coordinates in a right-handed Cartesian system. (These coordinates will be based on the electron density for X-ray derived structures and on nuclear positions for those derived from neutron scattering. Each coordinate is usually assumed to have accuracy between 0.5 and 1.0 Å.) Normally, only one additional characteristic is associated with each point, its size, usually measured by a van der Waals (VDW) radius. Furthermore, characteristics such as chemical nature and covalent connectivity, if needed, can be obtained from lookup tables keyed on the ID number.

Our model of a protein, thus, is the van der Waals envelope, the set of interlocking spheres drawn around each atomic center. In brief, the geometrical quantities of the model of particular concern in this section are its total surface area, total volume, the division of these totals among the amino acid residues and individual atoms, and the description of the empty space (cavities) outside of the van der Waals envelope. These values are then used in the analysis of protein structure and properties.

All the geometric properties of a protein (e.g., surfaces, volumes, distances, etc.) are obviously interrelated. So the definition of one quantity, e.g., area, obviously impacts on how another, e.g., volume, can be consistently defined. Here, we will endeavor to present definitions for measuring protein volume, showing how they are related to various definitions of linear distance (VDW parameters) and surface. Further information related to macromolecular geometry, focussing on volumes, is available from:

http://bioinfo.mbb.yale.edu/geometry.

Definitions of Protein Volume

Volume in terms of Voronoi Polyhedra: Overview

Protein volume can be defined in a straightforward sense through a particular geometric construction called Voronoi polyhedra. In essence, this construction provides a useful way of partitioning space amongst a collection of atoms. Each atom is surrounded by a single convex polyhedron and allocated the space within it (figure 1). The faces of Voronoi polyhedra are formed by constructing dividing planes perpendicular to vectors connecting atoms, and the edges of the polyhedra result from the intersection of these planes.

Voronoi polyhedra were originally developed by Voronoi (1908) nearly a century ago. Bernal & Finney (1967) used them to study the structure of liquids in the 1960s. However, despite the general utility of these polyhedra, their application to proteins was limited by a serious methodological difficulty. While the Voronoi construction is based on partitioning space amongst a collection of "equal" points, all protein atoms are not equal. Some are clearly larger than others. In 1974 a solution was found to this problem (Richards, 1974), and since then Voronoi polyhedra have been applied to proteins.

The Basic Voronoi Construction

(a) Integrating on a Grid

The simplest method for calculating volumes with Voronoi polyhedra is to put all atoms in the system on a fine grid. Then go to each grid-point (i.e., voxel) and add its infinitesimal volume to the atom center closest to it. This is prohibitively slow for a real protein structure, but it can be made somewhat faster by randomly sampling grid-points. It is, furthermore, a useful approach for high-dimensional integration (Sibbald & Argos, 1990).

More realistic approaches to calculating Voronoi volumes have two parts: (1) for each atom find the vertices of the polyhedron around it and (2) systematically collect these vertices to draw the polyhedron and calculate its volume.

(b) Finding Polyhedron Vertices

In the basic Voronoi construction (figure 1), each atom is surrounded by a unique limiting polyhedron such that all points within an atom's polyhedron are closer to this atom than all other atoms. Consequently, points equidistant from 2 atoms lie on a dividing plane; those equidistant from 3 atoms are on a line, and those equidistant from 4 centers form a vertex. One can use this last fact to easily find all the vertices associated with an atom. With the coordinates of four atoms, it is straightforward to solve for possible vertex coordinates using the equation of a sphere. (That is, one uses four sets of coordinates (x,y,z) and the equation $(x-a)^2 + (y-b)^2 + (z-c)^2 = r^2$ to solve for the center (a,b,c) and radius (r) of the sphere.) One then checks whether this putative vertex is closer to these four atoms than any other atom; if so, it is a real vertex.

Note that this procedure can fail for certain pathological arrangements of atoms that would not normally be encountered in a real protein structure. These occur if there is a center of symmetry, as in a regular cubic lattice or in a perfect hexagonal ring in a protein (see Procacci & Scateni, 1992). Centers of symmetry can be handled (in a limited way) by randomly perturbing the atoms a small amount and breaking the symmetry. Alternatively, the "chopping down" method described below is not affected by symmetry centers -- an important advantage to this method of calculation.

(c) Collecting Vertices and Calculating Volumes

To systematically collect the vertices associated with an atom, label each one by the indices of the four atoms with which it is associated (figure 2). To traverse the vertices on one face of a polyhedron, find all vertices that share two indices and thus have two atoms in common — e.g., a central atom (atom 0) and another atom (atom 1). Arbitrarily pick a vertex to start and walk around the perimeter of the face. One can tell which vertices are connected by edges because they will have a third atom in common (in addition to atom 0 and atom 1). This sequential walking procedure also provides a way to draw polyhedra on a graphics device. More importantly, with reference to the starting vertex, the face can be divided into triangles, for which it is trivial to calculate areas and volumes (see figure caption for specifics).

Adapting Voronoi Polyhedra to Proteins

In the procedure outlined above, all atoms are considered equal, and the dividing planes are positioned midway between atoms (figure 3). This method of partition, called bisection, is not physically reasonable for proteins, which have atoms of obviously different size (such as oxygen and sulfur). It chemically misallocates volume, giving excess to the smaller atom.

Two principal methods of re-positioning the dividing plane have been proposed to make the partition more physically reasonable: method B (Richards, 1974) and the radical-plane method (Gellatly & Finney, 1982). Both methods depend on the radii of the atoms in contact (R for the larger atom and r for the smaller one) and the distance between the atoms (D). As shown in figure 3, they position the plane at a distance d from the larger atom. This distance is always set such that the plane is closer to the smaller atom.

(a) Method B and a Simplification of it: The Ratio Method

Method B is the more chemically reasonable of the two and will be emphasized here. For atoms that are covalently bonded, it divides the distance between the atoms proportionately according to their covalent-bond radii:

$$d = D\, R/(R+r). \qquad [1]$$

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

$$d = R + (D-R-r)/2. \qquad [2]$$

For separations that are not much different from the sum of the radii, the two formulas for method B give essentially the same result. Consequently, it is worthwhile to try a slight simplification of method B, which we call the "ratio method." Instead of using the first formula for bonded atoms and the second for non-bonded ones, one can just use formula 2 in both cases with either VDW or covalent radii (Tsai & Gerstein, 1999). Doing this gives more consistent reference volumes (manifest in terms of smaller standard deviations about the mean).

(b) Vertex Error

If bisection is not used to position the dividing plane, it is much more complicated to find the vertices of the polyhedron, since a vertex is no longer equidistant from 4 atoms. Moreover, it is also necessary to have a reasonable scheme for “typing” atoms and assigning them radii.

More subtly, when using the plane-positioning determined by method B, the allocation of space is no longer mathematically perfect, since the volume in a tiny tetrahedron near each polyhedron vertex is not allocated to any atom (figure 3). This is called vertex error. However, calculations on periodic systems have shown that, in practice, vertex error does not amount to more than 1 part in 500 (Gerstein, Tsai & Levitt, 1995).

(c) "Chopping-down" Method of Finding Vertices

Because of vertex error and the complexities in locating vertices, a different algorithm has to be used for volume calculation with method B. (It can also be used with bisection.) First, surround the central atom (for which a volume is being calculated) by a very large, arbitrarily positioned tetrahedron. This is initially the "current polyhedron." Next, sort all neighboring atoms by distance from the central atom and go through them from nearest to farthest. For each neighbor, position a plane perpendicular to the vector connecting it to the central atom according to the predefined proportion (i.e., from the Method B formulas or bisection). Since a Voronoi polyhedron is always convex, if any vertices of the current polyhedron are on a different side of this plane than the central atom, they cannot be part of the final polyhedra and should be discarded. After this has been done, the current polyhedron is recomputed using the plane to "chop it down." This process is shown schematically in figure 4. When it is finished, one has a list of vertices, which can be traversed to calculate volumes, as in the basic Voronoi procedure.

(d) Radical-Plane Method

The radical-plane method does not suffer from vertex error. In this method, the plane is positioned according to the following formula:

$$d = (D^2+R^2-r^2)/2D \qquad [3]$$

The Delaunay Triangulation

Voronoi polyhedra are closely related (i.e., dual) to another useful geometric construction called the Delaunay triangulation. This consists of lines, perpendicular to Voronoi faces, connecting each pair of atoms that share a face (Figure 5).

The Delaunay triangulation is described here as a derivative of the Voronoi construction. However, it can be constructed directly from the atom coordinates. In 2D, one connects with a triangle any triplet of atoms if a circle through them does not enclose any additional atoms. Likewise, in 3D one connects 4 atoms in a tetrahedron if the sphere through them does not contain any further atoms. Notice how this construction is equivalent to the specification for Voronoi polyhedra and, in a sense, is simpler. One can immediately see the relationship between the triangulation and the Voronoi volume by noting that the volume is the distance between neighbors (as determined by the triangulation) weighted by the area of each polyhedral face. In practice, it is often easier in drawing to construct the triangles first and then build the Voronoi polyhedra from them.

The Delaunay triangulation is useful in many "nearest-neighbor" problems in computational geometry -- e.g., trying to find the neighbor of a query point and finding the largest empty circle in a collection of points (O'Rourke, 1994). Since this triangulation has the "fattest" possible triangles, it is the choice for procedures such as finite element analysis.

In terms of protein structure, the Delaunay triangulation is the natural way to determine packing neighbors, either in protein structure or molecular simulation (Singh, Tropsha & Vaisman, 1996; Tsai, Gerstein & Levitt, 1996, 1997). Its advantage is that the definition of "a neighbor" does not depend on distance. The alpha shape is a further generalization of the Delaunay triangulation that has proven useful in identifying ligand-binding sites (Edelsbrunner, Facello & Liang, 1996; Edelsbrunner et al., 1995; Edelsbrunner & Mucke, 1994; Peters, Fauck & Frommel, 1996).

Definitions of Protein Surface

The Problem of the Protein Surface

When one is carrying out the Voronoi procedure, if a particular atom does not have enough neighbors, the "polyhedron" formed around it will not be closed, but rather will have an open, concave shape. As it is not often possible to place enough water molecules in an X-ray crystal structure to cover all the surface atoms, these "open polyhedra" occur frequently on the protein surface (Figure 6). Furthermore, even when it is possible to define a closed polyhedron on the surface, it will often be distended and too large. This is the problem of the protein surface in relation to the Voronoi construction.

There are a number of practical techniques for dealing with this problem. First, one can use very high-resolution protein crystal structures, which have many solvent atoms positioned (Gerstein & Chothia, 1996). Alternatively, one can make up the positions of missing solvent molecules. These can be placed either according to a regular grid-like arrangement or, more realistically, according to the results of molecular simulation (Finney et al., 1980; Gerstein, Tsai & Levitt, 1995; Richards, 1974).

Definitions of Surface in terms of Voronoi Polyhedra (the Convex Hull)

More fundamentally, however, the "problem of the protein surface" indicates how closely linked the definitions of surface and volume are and how the definition of one, in a sense, defines the other. That is, the 2D surface of an object can be defined as the boundary between two 3D volumes. More specifically, the polyhedral faces defining the Voronoi volume of a collection of atoms also define their surface. The surface of a protein consists of the union of (connected) polyhedra faces. Each face in this surface is shared by one solvent atom and one protein atom (Figure 7).

Another somewhat related definition is the convex hull, the smallest convex polyhedron that encloses all the atom centers (Figure 7). This is important in computer graphics applications and as an intermediary in many geometric constructions related to proteins (Connolly, 1991; O'Rourke, 1994). The convex hull is a subset of the Delaunay triangulation of the surface atoms. It is quickly located by the following procedure (Connolly, 1991): Find the atom farthest from the molecular center. Then choose two of its neighbors (as determined by the Delaunay triangulation) such that a plane through these three atoms has all the remaining atoms of the molecule on one side of it (the "plane test"). This is the first triangle in the convex hull. Then one can choose a fourth atom connected to at least two of the three in the triangle and repeat the plane test, and by iteratively repeating this procedure, one can "sweep" across the surface of the molecule and define the whole convex hull.

Other parts of the Delaunay triangulation can define additional surfaces. The part of the triangulation connecting the first layer of water molecules defines a surface, as does the part joining the second layer. The second layer of water molecules, in fact, has been suggested on physical grounds to be the natural boundary for a protein in solution (Gerstein & Lynden-Bell, 1993c). Protein surfaces defined in terms of the convex hull or water layers tend to be "smoother" than those based on Voronoi faces, omitting deep grooves and clefts (see Figure 7).

Definitions of Surface in terms of a Probe Sphere

In the absence of solvent molecules to define Voronoi polyhedra, one can define the protein surface in terms of the position of a hypothetical solvent, often called the probe sphere, that "rolls" around the surface (Richards, 1977) (Figure 7). The surface of the probe is imagined to be maintained tangent to the van der Waals surface of the model.

Various algorithms are used to cause the probe to visit all possible points of contact with the model. The locus of either the center of the probe or the tangent point to the model is recorded. Either through exact analytical functions or numerical approximations of adjustable accuracy, the algorithms provide an estimate of the area of the resulting surface. (See Section 22A in this series for a more extensive discussion of the definition, calculation, and use of areas.)

Depending on the probe size and whether its center or point of tangency is used to define the surface, one arrives at a number of commonly used definitions, summarized in table 2 and Figure 7.

(a) <u>van der Waals Surface (VDWS)</u>

The area of the van der Waals surface will be calculated by the various area algorithms (see section 22a) when the probe radius is set to zero. This is a mathematical calculation only. There is no physical procedure that will measure van der Waals surface area directly. From a mathematical point of view, it is just the first of a set of solvent-accessible surfaces calculated with differing probe radii.

(b) <u>Solvent Accessible Surface (SAS)</u>

The solvent accessible surface is convex and closed, with defined areas assignable to each individual atom (Lee & Richards, 1971). However, the individual calculated values vary in a complex fashion with variations in the radii of the probe and protein atoms. This radius is frequently, but not always, set at a value considered to represent a water molecule (1.4 Å). The total SAS area increases without bound as the size of the probe increases.

(c) Molecular Surface as the sum of the Contact and Reentrant Surfaces (MS = CS + RS)

Like the solvent accessible surface, the molecular surface is also closed, but it contains a mixture of convex and concave patches, the sum of the contact and reentrant surfaces. The ratio of these two surfaces varies with probe radius. In the limit of infinite probe radius, the molecular surface becomes convex and attains a limiting minimum value (i.e., it becomes a convex hull, similar to the one described above). The molecular surface cannot be divided up and assigned unambiguously to individual atoms.

The contact surface is not closed. Instead, it is a series of convex patches on individual atoms, simply related to the solvent accessible surface of the same atoms. In complementary fashion, the reentrant surface is also not closed but is a series of concave patches that is part of the probe surface where it contacts 2 or 3 atoms simultaneously. At infinite probe radius, the reentrant areas are plane surfaces at which point the molecular surface becomes a convex surface. The reentrant surface cannot be divided up and assigned unambiguously to individual atoms. Note, the molecular surface is simply the union of the contact and reentrant surfaces, so in terms of area, MS = CS + RS.

(d) Further Points

The detail provided by these surfaces will depend on the radius of the probe used for their construction.

One may argue that the behavior of the rolling probe sphere does not accurately model real, hydrogen-bonded water. Instead, its "rolling" more closely mimics the behavior of a non-polar solvent. An attempt has been made to incorporate more realistic hydrogen-bonding behavior into the probe sphere, allowing for the definition of a hydration surface more closely linked to the behavior of real water (Gerstein & Lynden-Bell, 1993c).

The definitions of accessible surface and molecular surface can be related back to the Voronoi construction. The molecular surface is similar to "time-averaging" the surface formed from the faces of Voronoi polyhedra (the Voronoi surface) over many water configurations, and the accessible surface is similar to averaging the Delaunay triangulation of the first layer of water molecules over many configurations.

There are a number of other definitions of protein surfaces that are unrelated to either probe sphere or Voronoi polyhedra and provide complementary information (Kuhn et al., 1992; Leicester, Finney & Bywater, 1988; Pattabiraman, Ward & Fleming, 1995).

Definitions of Atomic Radii

The definition of protein surfaces and volumes depends greatly on the values chosen for various parameters of linear dimension -- in particular, van der Waals and probe-sphere radii.

van der Waals radii

For all the calculations outlined above, the hard sphere approximation is used for the atoms. (One must remember that in reality atoms are neither hard nor spherical, but this approximation has a long history of demonstrated utility.) There are many lists prepared in different laboratories for the radii of such spheres, both for single atoms and for unified atoms, where the radii are adjusted to approximate the joint size of the heavy atom and its bonded hydrogen atoms (clearly not an actual spherical unit).

Some of these lists are reproduced in Table 1. They are derived from a variety of approaches -- e.g., looking for the distances of closest approach between atoms (the Bondi set) and energy calculations (the CHARMM set). The differences between the sets often boil down to how one decides to truncate the Lennard-Jones potential function. Further differences arise from the parameterization of water and other hydrogen bonding molecules, as these substances really should be represented with two radii, one for their hydrogen-bonding interactions and one for their VDW interactions.

Perhaps because of the complexities in defining VDW parameters, there are some great differences in Table 1. For instance, the radius for an aliphatic CH (>CH-) ranges from 1.7 to 2.38 Å, and the radius for carboxyl oxygen ranges from 1.34 to 1.89 Å. Both of these represent at least a 40% variation. Moreover, such differences are practically quite significant, since many geometrical and energetic calculations are very sensitive to the choice of VDW parameters, particularly the relative values within a single list. (Repulsive core interactions, in fact, vary almost exponentially.) Consequently, proper volume and surface comparisons can only be based on numbers derived through use of the same list of radii.

In the last column of the table we give a recent set of VDW radii that has been carefully optimized for use in volume and packing calculations. It is derived from analysis of the most common distances between atoms in small-molecule crystal structures in the Cambridge Structural Database (Rowland & Taylor, 1996; Tsai et al., 1999).

The Probe radius

A series of surfaces can be described by using a probe sphere with a specified radius. Since this is to be a convenient mathematical construct in calculation, any numerical value may be chosen with no necessary relation to physical reality. Some commonly used examples are listed in table 2.

The solvent accessible surface is intended to be a close approximation to what a water molecule as a probe might "see" (Lee & Richards, 1971). However, there is no uniform agreement on what the proper water radius should be. Usually it is chosen to be about 1.4 Å.

Application of Geometry Calculations: The Measurement of Packing

Using Volume to Measure Packing Efficiency

Volume calculations are principally applied in measuring packing. This is because the packing efficiency of a given atom is simply the ratio of the space it could minimally occupy to the space that it actually does occupy. As shown in Figure 8, this ratio can be expressed as the VDW volume of an atom divided by its Voronoi volume (Richards, 1974; Richards, 1985; Richards & Lim, 1994). (Packing efficiency also sometimes goes by the equivalent terms “packing density” or “packing coefficient.”) This simple definition masks considerable complexities -- in particular, how does one determine the volume of the VDW envelope (Petitjean, 1994)? This requires knowledge of what the VDW radii of atoms are, a subject on which there is not universal agreement (see above), especially for water molecules and polar atoms (Gerstein, Tsai & Levitt., 1995; Madan & Lee, 1994).

Knowing that the absolute packing efficiency of an atom is a certain value is most useful in a comparative sense -- i.e., when comparing equivalent atoms in different parts of a protein structure. In taking a ratio of two packing efficiencies, the VDW envelope volume remains the same and cancels. One is left with just the ratio of space that an atom occupies in one environment to what it occupies in another. Thus, for the measurement of packing, standard reference volumes are particularly useful. Recently calculated values of these standard volumes are shown in Tables 3 and 4 for atoms and residues (Tsai et al., 1999).

In analyzing molecular systems, one usually finds that close packing is the default (Chandler, Weeks & Andersen, 1983) -- i.e., atoms pack like billiard balls. Unless there are highly directional interactions (such as hydrogen bonds) that have to be satisfied, one usually achieves close packing to optimize the attractive tail of the VDW interaction. Close-packed spheres of the same size have a packing efficiency of ~0.74. Close-packed spheres of different size are expected to have a somewhat higher packing efficiency. In contrast, water is not close-packed because it has to satisfy the additional constraints of hydrogen bonding. It has an open, tetrahedral structure with a packing efficiency of ~0.35. (This difference in packing efficiency is illustrated in figure 8B.)

The Tight Packing of the Protein Core

The protein core is usually considered to be the atoms inaccessible to solvent -- i.e., with an accessible surface area of zero or a very small number, such as 0.1 $Å^2$. Packing calculations on the protein core are usually done by calculating the average volumes of the buried atoms and residues in a database of crystal structures. These calculations were first done more than two decades ago (Chothia & Janin, 1975; Finney, 1975; Richards, 1974). The initial calculations revealed some important facts about protein structure. Atoms and residues of a given type inside of proteins have a roughly constant (or invariant) volume. This is because the atoms inside proteins are packed together fairly tightly, with the protein interior better resembling a close-packed solid than a liquid or gas. In fact, the packing efficiency of atoms inside of proteins is roughly what is expected for the close packing of hard spheres (.74).

More recent calculations measuring the packing in proteins (Harpaz, Gerstein & Chothia, 1994; Tsai et al., 1999) have shown that the packing inside of proteins is somewhat tighter (~4%) than that observed initially and that the overall packing efficiency of atoms in the protein core is greater than that in crystals of organic molecules. When molecules are packed this tightly, small changes in packing efficiency are quite significant. In this regime, the limitation on close packing is hard-core repulsion, which is expected to have a twelfth power or exponential dependence, so even a small change is quite substantial energetically. Furthermore, the number of allowable configurations that a collection of atoms can assume without core overlap drops off very quickly as these atoms approach the close-packed limit (Richards & Lim, 1994).

The exceptionally tight packing in the protein core seems to require a precise jigsaw puzzle-like fit of the residues. This appears to be the case for the majority of atoms inside of proteins (Connolly, 1986). The tight packing in proteins has, in fact, been proposed as a quality measure in protein crystal structures (Pontius, Richelle & Wodak, 1996). It is also believed to be a strong constraint on protein flexibility and motions (Gerstein et al., 1993; Gerstein, Lesk & Chothia, 1994a). However, there are exceptions, and some studies have focussed on these, showing how the packing inside proteins is punctuated by defects, or cavities (Hubbard & Argos, 1994; Hubbard & Argos, 1995; Kleywegt & Jones, 1994; Kocher et al., 1996; Rashin et al., 1986; Richards, 1979; Williams et al., 1994). If these defects are large enough, they can contain buried water molecules (Baker & Hubbard, 1984; Matthews et al., 1995; Sreenivasan & Axelsen, 1992).

Surprisingly, despite the intricacies of the observed jigsaw-puzzle-like packing in the protein core, it has been shown that one can simply achieve the "first-order" aspect of this, getting the overall volume of the core right rather easily (Gerstein, Sonnhammer & Chothia, 1994a; Kapp et al., 1995; Lim & Ptitsyn, 1970). This has to do with simple statistics for summing random numbers and the fact that the distribution of sizes for amino acids usually found inside proteins is rather narrow (Table 3). In fact, the similarly sized residues Val, Ile, Leu and Ala (with volumes 138, 163, 163 and 89) make up about half of the residues buried in the protein core. Furthermore, aliphatic residues, in particular, have a relatively large number of adjustable degrees of freedom per cubic Angstrom, allowing them to accommodate a wide range of packing geometries. All of this suggests that many of the features of protein sequences may only require random-like qualities for them to fold (Finkelstein, 1994).

Looser Packing on the Surface

Measuring the packing efficiency inside of the protein core provides a good reference point for comparison, and a number of other studies have looked at this in comparison to other parts of the protein. The most obvious thing to compare with the protein inside is the protein outside, or surface. This is particularly interesting from a packing perspective, since the protein surface is covered by water, and water is packed much less tightly than protein and in a distinctly different fashion. (The tetrahedral packing geometry of water molecules gives a packing efficiency of less than half that of hexagonal close-packed solids.)

Calculations based on crystal structures and simulations have shown that the protein surface has intermediate packing, being packed less tightly than the core but not as loosely as liquid water (Gerstein & Chothia, 1996; Gerstein, Tsai & Levitt, 1995). One can understand the looser packing at the surface than in the core in terms of a simple trade-off between hydrogen bonding and close-packing, and this can be explicitly visualized in simulations of the packing in simple toy systems (Gerstein & Lynden-Bell, 1993a, 1993b).

Figure Captions and Tables

Figure 1. The Voronoi Construction in 2D and 3D

Representative Voronoi polyhedra from 1CSE (subtilisin). TOP-LEFT, six polyhedra around the atoms in a Phe ring. TOP-RIGHT, a single polyhedron around the sidechain hydroxyl oxygen (OG) of a serine. BOTTOM, a schematic showing the construction of a Voronoi polyhedron in 2-dimensions. The broken lines indicate planes that were initially included in the polyhedron but then removed by the "chopping down" procedure (see figure 4).

Figure 2. Labeling Parts of Voronoi Polyhedra

The figure illustrates a labeling scheme for parts of Voronoi polyhedra. The central atom is atom 0, and each neighboring atom has a sequential index number (1,2,3...). Consequently, in 3D, planes are denoted by the indices of the 2 atoms that form them (e.g., 01); lines are denoted by the indices of 3 atoms (e.g., 012); and vertices are denoted by 4 indices (e.g., 0123). In the 2D representation shown here, lines are denoted by 2 indices, and vertices, by 3. From a collection of points, a volume can be calculated by a variety of approaches: First of all, the volume of a tetrahedron determined by four points can be calculated by placing one vertex at the origin and evaluating the determinant formed from the remaining three vertices. (The tetrahedron volume is one-sixth of the determinant value.) The determinant can be quickly calculated by a vector triple product, $\mathbf{w} \bullet (\mathbf{u} \times \mathbf{v})$, where u, v, and w are vectors between the vertex selected to be the origin and the other three vertices of the tetrahedron. Alternately, the volume of the pyramid from a central atom to a face can be calculated from the usual formula Ad/3, where A is the area of the face and d is the distance to the face.

Figure 3. Positioning of the Dividing Plane

PART A illustrates how the dividing plane is positioned at a distance d from the larger atom with respect to radii of the larger atom (R) and the smaller atom (r) and the total separation between the atoms (D).

PART B illustrates Vertex Error. One problem with using Method B is that 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 the figure. The radical plane method does not suffer from vertex error, but it is not as chemically reasonable as method B.

Figure 4. The "Chopping Down" Method of Polyhedra Construction

The figure illustrates the "Chopping Down" method of calculation. This is necessary when using method B for plane positioning, since one can no longer solve for the position of vertices. One starts with a large tetrahedron around the central atom and then "chops it down" by removing vertices that are outside of the plane formed by each neighbor. For instance, say vertex 0214 of the current polyhedron is outside of the plane formed by neighbor 6. One needs to delete 0214 from the list of vertices and recompute the polyhedron using the new vertices formed from the intersection of the plane formed by neighbor 6 and the current polyhedron. Using the labeling conventions in figure 2, one finds that these new vertices are formed by the intersection of 3 lines (021, 024, and 014) with plane 06. So one adds the new vertices 0216, 0246, and 0146 to the polyhedron. However, there is a snag: it is necessary to check whether any of the 3 lines are not also outside of the plane. To do this, when a vertex is deleted, all the lines forming it (e.g., 021, 024, 014) are pushed onto a secondary list. Then when another vertex is deleted, one checks whether any of its lines have already been deleted. If so, this line is not used to intersect with the new plane. This process is shown schematically in 2D in the figure.

For the purposes of the calculations, it is useful to define a plane created by a vector **v** from the central atom to the neighboring atom by a constant K so that for any point **u** on the plane: $\mathbf{u} \bullet \mathbf{v} = K$. If $\mathbf{u} \bullet \mathbf{v} > K$, **u** is on the wrong side of the plane, otherwise it is on the right side. A vertex point **w** satisfies the equations of three planes: $\mathbf{w} \bullet \mathbf{v}_1 = K_1$, $\mathbf{w} \bullet \mathbf{v}_2 = K_2$, and $\mathbf{w} \bullet \mathbf{v}_3 = K_3$. These three equations can be solved to give the components of **w**. For example, the x component is:

$$w_x = \begin{vmatrix} K_1 & v_{1y} & v_{1z} \\ K_2 & v_{2y} & v_{2z} \\ K_3 & v_{3y} & v_{3z} \end{vmatrix} \Bigg/ \begin{vmatrix} v_{1x} & v_{1y} & v_{1z} \\ v_{2x} & v_{2y} & v_{2z} \\ v_{3x} & v_{3y} & v_{3z} \end{vmatrix} .$$

Figure 5. The Delaunay Triangulation Defines Packing Neighbors

The figure illustrates Delaunay Triangulation and its relation to the Voronoi construction. LEFT shows a standard schematic of the Voronoi construction. The atoms used to define the Voronoi planes around the central atom are highlighted. Lines connecting these atoms to the central one are part of the Delaunay Triangulation, which is shown at RIGHT. Note that atoms included in the triangulation cannot be selected strictly on the basis of a simple distance criterion relative to the central atom. The two circles about the central atoms illustrate this. Some atoms within the outer circle but outside of the inner circle are included in the triangulation, but others are not. In the context of protein structure, the Delaunay triangulation is useful in identifying true "packing contacts," in contrast to those contacts found purely by distance threshold. The broken lines in the LEFT subfigure indicate planes that were initially included in the polyhedron but then removed by the "chopping down" procedure (see figure 4).

Figure 6. The Problem of the Protein Surface

The figure shows the difficulty in constructing Voronoi Polyhedra for atoms on the protein surface. If all the water molecules near the surface are not resolved in a crystal structure, one often does not have enough neighbors to define a closed polyhedron. This figure is to be compared to figure 1, illustrating the basic Voronoi construction. Both figures are exactly the same except that in this figure some of the atoms on the left are missing, giving the central atom an open polyhedron. The broken lines indicate planes that were initially included in the polyhedron but then removed by the "chopping down" procedure (see figure 4).

Figure 7. Definitions of the Protein Surface

<u>PART A</u> shows the classic definitions of protein surface in terms of the probe sphere, the accessible surface and the molecular surface. (The figure is adapted from Richards, 1977).
<u>PART B</u> shows how Voronoi polyhedra and Delaunay Triangulation can also be used to define a protein surface. In this schematic, the large spheres represent closely packed protein atoms, and the smaller spheres represent the small loosely packed water molecules. The Delaunay Triangulation is shown by dotted lines. Some parts of the triangulation can be used to define surfaces. The outermost part of the triangulation of *just* the protein atoms forms the convex hull.

This is indicated by the thick line around the protein atoms. For the convex hull construction, one imagines that the water is not present. This is highlighted by the difference between the thick dotted line, which shows how Delaunay triangulation of the surface atoms in the presence of the water diverges from the convex hull near a deep cleft. Another part of the triangulation, also indicated by thick black lines, connects the first layer of water molecules (those that touch protein atoms). A time-averaged version of this line approximates the accessible surface. Finally, the light thick lines show the Voronoi faces separating the protein surface atoms from the first layer of water molecules. Note how this corresponds approximately to the molecular surface (considering the water positions to be time-averaged). These correspondences between the accessible and molecular surfaces and time-averaged parts of the Voronoi construction are understandable in terms of which part of the probe sphere, center or point of tangency, is used for the surface definition. The accessible surface is based on the position of the center of the probe sphere, while the molecular surface is based on the points of tangency between the probe sphere and the protein atoms, and these tangent points are similarly positioned to Voronoi faces, which bisect inter-atomic vectors between solvent and protein atoms.

Figure 8. Packing Efficiency

PART A illustrates the relationship between Voronoi polyhedra and packing efficiency. Packing efficiency is defined as the volume of an object as a fraction of the space that it occupies. (It is also known as “packing coefficient” or “packing density”.) In the context of molecular structure, it is measured by the ratio of the VDW volume (V_{VDW}, shown by a light gray line) and Voronoi volume (V_{Vor}, shown by a dotted line). This calculation gives absolute packing efficiencies. In practice, one usually measures a relative efficiency, relative to the atom in a reference state: $(V_{VDW}/V_{Vor})/(V_{VDW}/V_{Vor}(ref))$. Note that in this ratio the unchanging VDW volume of an atom cancels out, leaving one with just a ratio of two Voronoi volumes. Perhaps more usefully, when one is trying to evaluate the packing efficiency P at an interface, one computes $P=p\Sigma V_i/\Sigma v_i$, where p is packing efficiency of the reference dataset (usually .74), V_i is the actual measured volume of each atom i at interface, and v_i is the reference volume corresponding to the type of atom i.

PART B graphically illustrates the difference between tight packing and loose packing. Frames from a simulation are shown for liquid water (bottom) and for liquid argon, a simple liquid (top). Due to its hydrogen-bonds, water is much less tightly packed than argon (packing efficiency of 0.35 vs. ~0.7). Each water molecule has only four to five nearest neighbors while each argon has about ten.

Table 1. Standard Atomic Radii

Atom Type & Symbol		Bondi 1968	Lee & Richards 1971	Shrake & Rupley 1973	Richards 1974	Chothia 1975	Richmond & Richards 1978	Gelin & Karplus 1979	Dunfield et al. 1979	ENCAD derived 1995	CHARMM derived 1995	Tsai et al. 1999
$-CH_3$	Aliphatic, methyl	2.00	1.80	2.00	2.00	1.87	1.90	1.95	2.13	1.82	1.88	1.88
$-CH_2-$	Aliphatic, methyl	2.00	1.80	2.00	2.00	1.87	1.90	1.90	2.23	1.82	1.88	1.88
>CH-	Aliphatic, CH	-	1.70	2.00	2.00	1.87	1.90	1.85	2.38	1.82	1.88	1.88
≥CH	Aromatic, CH	-	1.80	1.85	*	1.76	1.70	1.90	2.10	1.74	1.80	1.76
>C=	Trigonal, aromatic	1.74	1.80	*	1.70	1.76	1.70	1.80	1.85	1.74	1.80	1.61
$-NH_3+$	Amino, protonated	-	1.80	1.50	2.00	1.50	0.70	1.75		1.68	1.40	1.64
$-NH_2$	Amino or amide	1.75	1.80	1.50	-	1.65	1.70	1.70		1.68	1.40	1.64
>NH	Peptide, NH or N	1.65	1.52	1.40	1.70	1.65	1.70	1.65	1.75	1.68	1.40	1.64
=O	Carbonyl Oxygen	1.50	1.80	1.40	1.40	1.40	1.40	1.60	1.56	1.34	1.38	1.42
-OH	Alcoholic hydroxyl	-	1.80	1.40	1.60	1.40	1.40	1.70		1.54	1.53	1.46
-OM	Carboxyl Oxygen	-	1.80	1.89	1.50	1.40	1.40	1.60	1.62	1.34	1.41	1.42
-SH	Sulfhydryl	-	1.80	1.85	-	1.85	1.80	1.90		1.82	1.56	1.77
-S-	Thioether or –S-S-	1.80	-	-	1.80	1.85	1.80	1.90	2.08	1.82	1.56	1.77

All values in Angstroms. Comments below. "*" means to see note below on a specific value.

Bondi: Values assigned on the basis of observed packing in condensed phases (Bondi, 1968).

Lee & Richards: Values adapted from Bondi (1964) and used in Lee & Richards (1971).

Shrake & Rupley: Values taken from Pauling (1960) and used in Shrake & Rupley (1973). >C= value can be either 1.5 or 1.85.

Richards: Minor modification of the original Bondi set in Richards (1974). (Rationale not given.) See original paper for discussion of aromatic carbon value.

Chothia: From packing in amino acid crystal structures. Used in Chothia (1975).

Richmond & Richards: No rationale given for values used in Richmond & Richards (1978).

Gelin & Karplus: Origin of values not specified. Used in Gelin & Karplus (1979).

Dunfield et al: Detailed description of deconvolution of molecular crystal energies. Values represent one-half of the heavy-atom separation at the minimum of the Lennard-Jones 6-12 potential functions for symmetrical interactions. Used in Nemethy et al. (1983) and Dunfield et al. (1979).

ENCAD: A set of radii, derived in Gerstein et al. (1995), based solely on the ENCAD molecular dynamics potential function in Levitt et al. (1995). To determine these radii, the separation at which the 6-12 Lennard-Jones interaction energy between equivalent atoms was 0.25 k_BT was determined (0.15 kcal/mole).

CHARMM: Determined in the same way as the ENCAD set, but now for the CHARMM potential (Brooks et al., 1983) (parameter set 19).

Tsai et al.: Values derived from a new analysis (Tsai et al., 1999) of the most common distances of approach of atoms in the Cambridge Structural Database.

Table 2. Probe Radii and their Relation to Surface Definition

Probe Radius	Part of Probe Sphere	Type of Surface
0	Center (or Tangent)	Van der Waals Surface (VDWS)
1.4 Å	Center	Solvent Accessible Surface (SAS)
""	Tangent (1 atom)	Contact Surface (CS, from parts of atoms)
""	Tangent (2 or 3 atoms)	Reentrant Surface (RS, from parts of probe)
""	Tangent (1,2, or 3 atoms)	Molecular Surface (MS = CS + RS)
10 Å	Center	A Ligand or Reagent Accessible Surface
∞	Tangent	Minimum limit of MS (related to convex hull)
""	Center	Undefined

The 1.4 and, especially, 10 Å are only approximate figures. One could, of course, use 1.5 Å for a water radius or 15 Å for a ligand radius, depending on the specific application.

Table 3. Standard Residue Volumes

Residue	Volume	SD	Freq.
Ala	89.3	3.5	13%
Val	138.2	4.8	13%
Leu	163.1	5.8	12%
Gly	63.8	2.7	11%
Ile	163.0	5.3	9%
Phe	190.8	4.8	6%
Ser	93.5	3.9	6%
Thr	119.6	4.2	5%
Tyr	194.6	4.9	3%
Asp	114.4	3.9	3%
Cys	102.5	3.5	3%
Pro	121.3	3.7	3%
Met	165.8	5.4	2%
Trp	226.4	5.3	2%
Gln	146.9	4.3	2%
His	157.5	4.3	2%
Asn	122.4	4.6	1%
Glu	138.8	4.3	1%
Cyh	112.8	5.5	1%
Arg	190.3	4.7	1%
Lys	165.1	6.9	1%

The table shows for each residue its standard volume and its frequency of occurrence in the protein core. Considering cysteine (Cyh, reduced) to be chemically different from cystine (Cys, involved in a disulfide and hence oxidized) gives 21 different types of residues. For each residue a mean volume and the standard deviation about this mean are shown in the two left columns in cubic Angstroms. These residue volumes are adapted from the ProtOr parameter set (also known as the BL+ set) in Tsai et al. (1999) and Tsai & Gerstein (1999). For this set, the averaging is done over 87 representative high-resolution crystal structures, only buried atoms not in contact with ligand are selected, the radii set shown in the last column of Table 1 is used, and the volumes are computed in the presence of the crystal water. The frequencies for buried residues are from Gerstein et al. (1994b).

Table 4. Standard Atomic Volumes

atom type	cluster	Description	mean	SD	num	Symbol
>C=	bigger	Trigonal (unbranched), aromatics	9.7	0.7	4184	C3H0b
>C=	smaller	Trigonal (branched)	8.7	0.6	11876	C3H0s
≥CH	bigger	Aromatic, CH (facing away from mainchain)	21.3	1.9	2063	C3H1b
≥CH	smaller	Aromatic, CH (facing towards mainchain)	20.4	1.7	1742	C3H1s
>CH-	bigger	Aliphatic, CH (unbranched)	14.4	1.3	3642	C4H1b
>CH-	smaller	Aliphatic, CH (branched)	13.2	1.0	7028	C4H1s
-CH2-	bigger	Aliphatic, methyl	24.3	2.1	1065	C4H2b
-CH2-	smaller	Aliphatic, methyl	23.2	2.3	4228	C4H2s
-CH3		Aliphatic, methyl	36.7	3.2	3497	C4H3u
>N-		Pro N	8.7	0.6	581	N3H0u
>NH	bigger	sidechain NH	15.7	1.5	446	N3H1b
>NH	smaller	Peptide	13.6	1.0	10016	N3H1s
-NH2		Amino or amide	22.7	2.1	250	N3H2u
-NH3+		Amino, protonated	21.4	1.2	8	N4H3u
=O		Carbonyl Oxygen	15.9	1.3	7872	O1H0u
-OH		Alcoholic hydroxyl	18.0	1.7	559	O2H1u
-S-		Thioether or –S-S-	29.2	2.6	263	S2H0u
-SH		Sulfhydryl	36.7	4.2	48	S2H1u

Standard atomic volumes for each of the 18 main types of atoms. Tsai et al. (1999) and Tsai & Gerstein (1999) clustered all the atoms in proteins into 18 basic types. These are shown in this table. Most of these have a simple chemical definition – e.g., "=O" are carbonyl carbons. However, some of the basic chemical types, such as the aromatic CH group ("≥CH"), need to be split into two subclusters (bigger and smaller), as is indicated by the column labeled "cluster". Volume statistics were accumulated for each of the 18 types based on averaging over 87 high-resolution crystal structures (in the same fashion as described for the residue volumes in Table 3). These statistics are shown in the mean, SD, and num columns, which give the average volume in cubic Angstroms, the standard deviation about this (also in cubic Angstroms) and the number of atoms averaged over. The final column ("symbol") gives the standardized symbol used to describe the atom in Tsai et al. (1999). These atom volumes shown here are part of the ProtOr parameter set (also known as the BL+ set) in Tsai et al. (1999).

References

Baker, E N & Hubbard, R E (1984). Hydrogen Bonding in Globular Proteins. *Prog. Biophys. Mol. Biol.* **44**: 97-179.

Bernal, J D & Finney, J L (1967). Random close-packed hard-sphere model II. Geometry of random packing of hard spheres. *Disc. Faraday Soc.* **43**: 62-69.

Bondi, A (1964). van der Waals Volumes and Radii. *J. Phys. Chem.* **68**: 441-451.

Bondi, A (1968). *Molecular Crystals, Liquids and Glasses* (Wiley, New York).

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.

Chandler, D, Weeks, J D & Andersen, H C (1983). Van der Waals Picture of Liquids, Solids, and Phase-transformations. *Science* **220**: 787-794.

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

Chothia, C & Janin, J (1975). Principles of protein-protein recognition. *Nature* **256**: 705-708.

Connolly, M (1986). Measurement of protein surface shape by solid angles. *J. Mol. Graph.* **4**: 3-6.

Connolly, M L (1991). Molecular Interstitial Skeleton. *Computers Chem.* **15**: 37-45.

Dunfield, L G, Burgess, A W & Scheraga, H A (1979). *J. Phys. Chem.* **82**: 2609.

Edelsbrunner, H, Facello, M & Liang, J. *On the definition and construction of pockets in macromolecules* 1-272-287 (World Scientific, Singapore, 1996).

Edelsbrunner, H, Facello, M, Ping, F & Jie, L (1995). Measuring proteins and voids in proteins. *Proc. 28th Hawaii Int. Conf. Sys. Sci.* 256-264.

Edelsbrunner, H & Mucke, E (1994). Three-dimensional alpha shapes. *ACM Transactions on Graphics* **13**: 43-72.

Finkelstein, A (1994). Implications of the random characteristics of protein sequences for their three-dimensional structure. *Curr. Opin. Str. Biol.* **4**: 422-428.

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, 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.

Gelin, B R & Karplus, M (1979). Side-chain torsional potentials: effect of dipeptide, protein, and solvent environment. *Biochemistry* **18**: 1256-1268.

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 & Chothia, C (1996). Packing at the Protein-Water Interface. *Proc. Natl. Acad. Sci. USA* **93**: 10167-10172.

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.

Gerstein, M, Lesk, A M & Chothia, C (1994). Structural Mechanisms for Domain Movements. *Biochemistry* **33**: 6739-6749.

Gerstein, M & Lynden-Bell, R M (1993). Simulation of Water around a Model Protein Helix. 1. Two-dimensional Projections of Solvent Structure. *J. Phys. Chem.* **97**: 2982-2991.

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

Gerstein, M, Sonnhammer, E & Chothia, C (1994). Volume Changes on Protein Evolution. *J. Mol. Biol.* **236**: 1067-1078.

Gerstein, M, Tsai, J & Levitt, M (1995). The volume of atoms on the protein surface: Calculated from simulation, using Voronoi polyhedra. *J. Mol. Biol.* **249**: 955-966.

Harpaz, Y, Gerstein, M & Chothia, C (1994). Volume Changes on Protein Folding. *Structure* **2**: 641-649.

Hubbard, S J & Argos, P (1994). Cavities and packing at protein interfaces. *Protein Science* **3**: 2194-2206.

Hubbard, S J & Argos, P (1995). Evidence on close packing and cavities in proteins. *Current Opinion In Biotechnology* **6**: 375-381.

Kapp, O H, Moens, L, Vanfleteren, J, Trotman, C N A, Suzuki, T & Vinogradov, S N (1995). Alignment of 700 globin sequences: Extent of amino acid substitution and its correlation with variation in volume. *Prot. Sci.* **4**: 2179-2190.

Kleywegt, G J & Jones, T A (1994). Detection, Delineation, Measurement and Display of Cavities in Macromolecular Structures. *Acta Cryst.* **D50**: 178-185.

Kocher, J P, Prevost, M, Wodak, S J & Lee, B (1996). Properties of the protein matrix revealed by the free energy of cavity formation. *Structure* **4**: 1517-1529.

Kuhn, L A, Siani, M A, Pique, M E, Fisher, C L, Getzoff, E D & Tainer, J A (1992). The Interdependence of Protein Surface Topography and Bound Water Molecules Revealed by Surface Accessibility and Fractal Density Measures. *J. Mol. Biol.* **228**: 13-22.

Lee, B & Richards, F M (1971). The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.* **55**: 379-400.

Leicester, S E, Finney, J L & Bywater, R P (1988). Description of molecular surface shape using Fourier descriptors. *J. Mol. Graphics* **6**: 104-108.

Levitt, M, Hirschberg, M, Sharon, R & Daggett, V (1995). Potential Energy Function and Parameters for Simulations of the Molecular Dynamics of Proteins and Nucleic Acids in Solution. *Computer Phys. Comm.* **91**: 215-231.

Lim, V I & Ptitsyn, O B (1970). On the Constancy of the Hydrophobic Nucleus Volume in Molecules of Myoglobins and Hemoglobins. *Mol. Biol. (USSR)* **4**: 372-382.

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. *Biophysical Chemistry* **51**: 279-289.

Matthews, B W, Morton, A G & Dahlquist, F W (1995). Use of NMR to detect water within nonpolar protein cavities [letter]. *Science* **270**: 1847-1849.

Nemethy, G, Pottle, M S & Scheraga, H A (1983). *J. Phys. Chem.* **87**: 1883.

O'Rourke, J (1994). *Computational Geometry in C* (Cambridge UP, Cambridge).

Pattabiraman, N, Ward, K B & Fleming, P J (1995). Occluded molecular surface: analysis of protein packing. *J Mol Recognit* **8**: 334-344.

Pauling, L C (1960). *The nature of the chemical bond* (Cornell Univ. Press, Ithaca, NY).

Peters, K P, Fauck, J & Frommel, C (1996). The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. *J Mol Biol* **256**: 201-213.

Petitjean, M (1994). On the analytical calculation of van der Waals surfaces and volumes: some numerical aspects. *Journal of Computational Chemistry* **15**: p.

Pontius, J, Richelle, J & Wodak, S J (1996). Deviations from Standard Atomic Volumes as a Quality Measure for Protein Crystal Structures. *J Mol Biol.* **264**: 121-136.

Procacci, P & Scateni, R (1992). A General Algorithm for Computing Voronoi Volumes: Application to the Hydrated Crystal of Myoglobin. *Int. J. Quant. Chem.* **42**: 151-1528.

Rashin, A A, Iofin, M & Honig, B (1986). Internal cavities and buried waters in globular proteins. *Biochemistry* **25**: 3619-3625.

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. *Ann. 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 in Enzymology* **115**: 440-464.

Richards, F M & Lim, W A (1994). An analysis of packing in the protein folding problem. *Quart. Rev. Biophys.* **26**: 423-498.

Richmond, T J & Richards, F M (1978). Packing of alpha-helices: Geometrical constraints and contact areas. *J. Mol. Biol.* **119**: 537-555.

Rowland, R S & Taylor, R (1996). Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii. *J. Phys. Chem.* **100**: 7384-7391.

Shrake, A & Rupley, J A (1973). *J. Mol. Biol.* **79**: 351.

Sibbald, P R & Argos, P (1990). Weighting Aligned Protein or Nucleic Acid Sequences to Correct for Unequal Representation. *J. Mol. Biol.* **216**: 813-818.

Singh, R K, Tropsha, A & Vaisman, I I (1996). Delaunay Tessellation of Proteins: Four Body Nearest-Neighbor Propensities of Amino Acid Residues. *J. Comp. Biol.* **3**: 213-222.

Sreenivasan, U & Axelsen, P H (1992). Buried Water in Homologous Serine Proteases. *Biochemistry* **31**: 12785-12791.

Tsai, J & Gerstein, M (1999). Volume Calculations of Protein Atomic Groups: Factors Affecting the Calculation and Derivation of a Minimal yet Optimal Set of Volumes. *Proteins* (submitted).

Tsai, J, Gerstein, M & Levitt, M (1996). Keeping the shape but changing the charges: A simulation study of urea and its iso-steric analogues. *J. Chem. Phys.* **104**: 9417-9430.

Tsai, J, Gerstein, M & Levitt, M (1997). Estimating the size of the minimal hydrophobic core. *Protein Science* **6**: 2606-2616).

Tsai, J, Taylor, R, Chothia, C & Gerstein, M (1998). The Packing Density in Proteins: Standard Radii and Volumes. *J. Mol. Biol.* (in press).

Voronoi, G F (1908). Nouvelles 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 Science* **3**: 1224-1235.

Figure Graphics (Follow Sequentially)

Figure 1. The Voronoi Construction in 2D and 3D

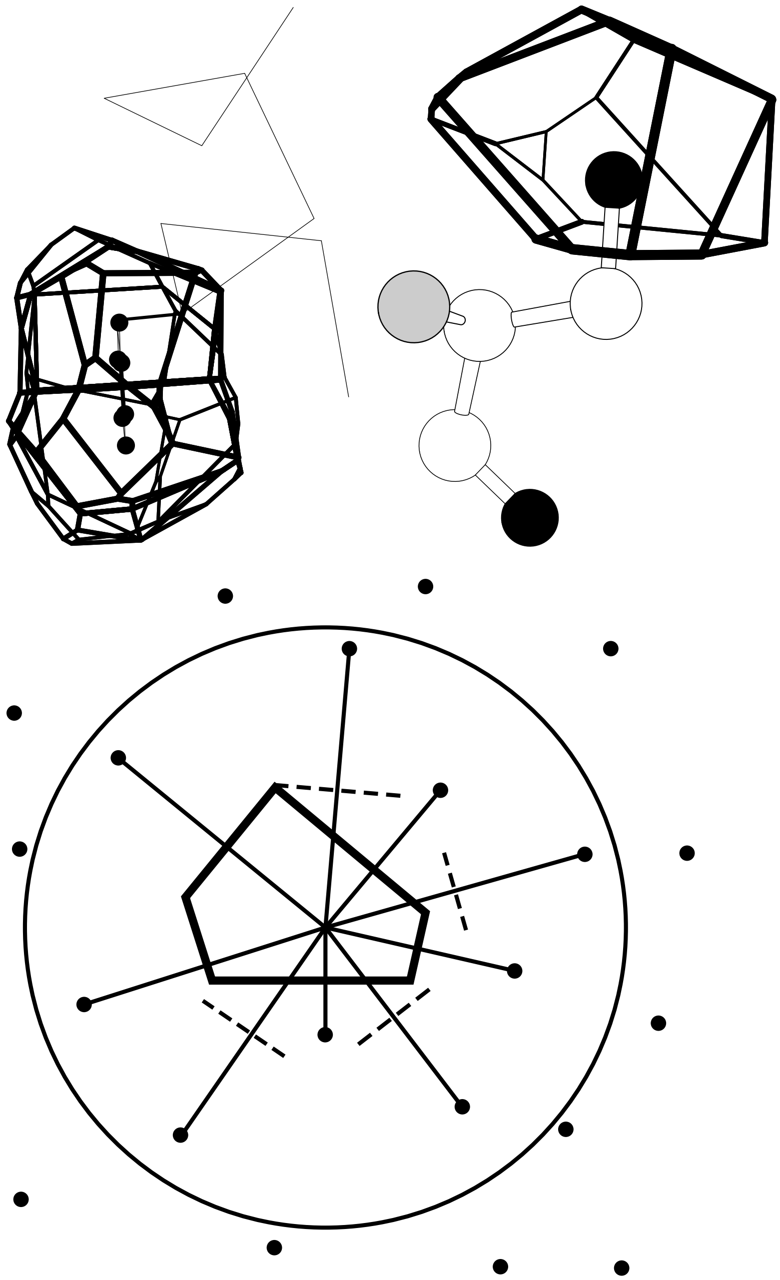

Figure 2. Labeling Parts of Voronoi Polyhedra

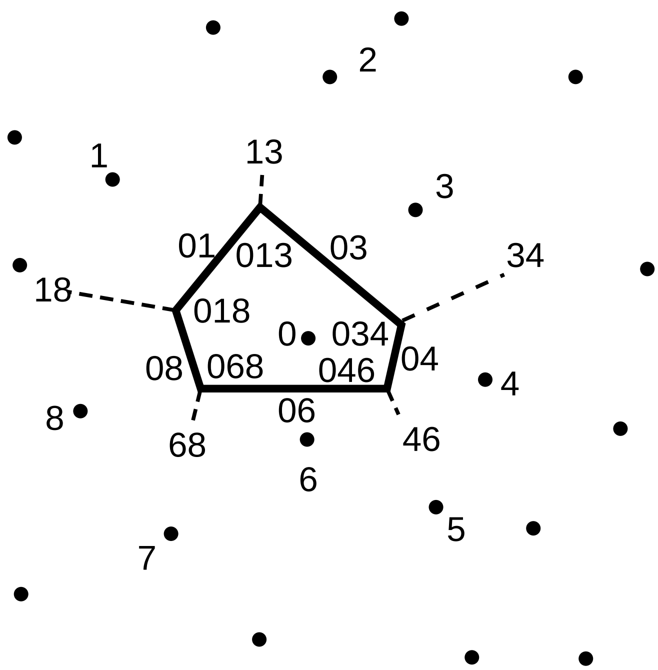

Figure 3. Positioning of the Dividing Plane

Figure 3A. Definitions

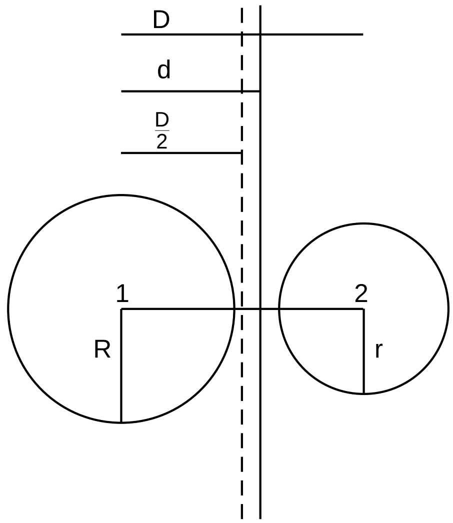

Figure 3B. Vertex Error

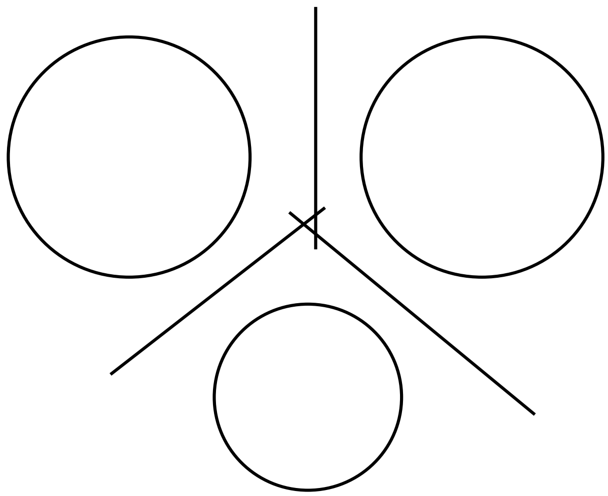

Figure 4. The "Chopping Down" Method of Polyhedra Construction

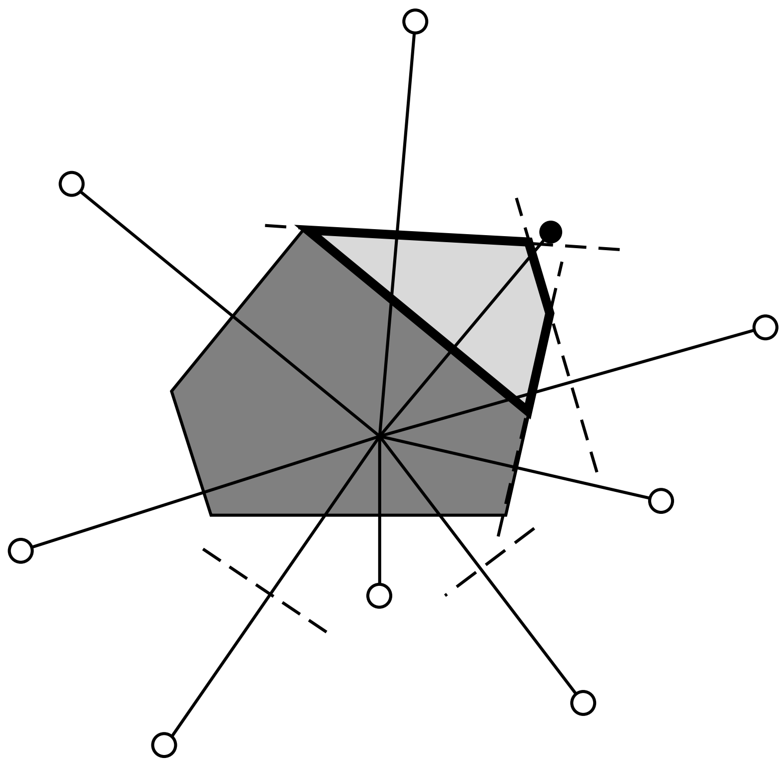

Figure 5. The Delaunay Triangulation Defines Packing Neighbors

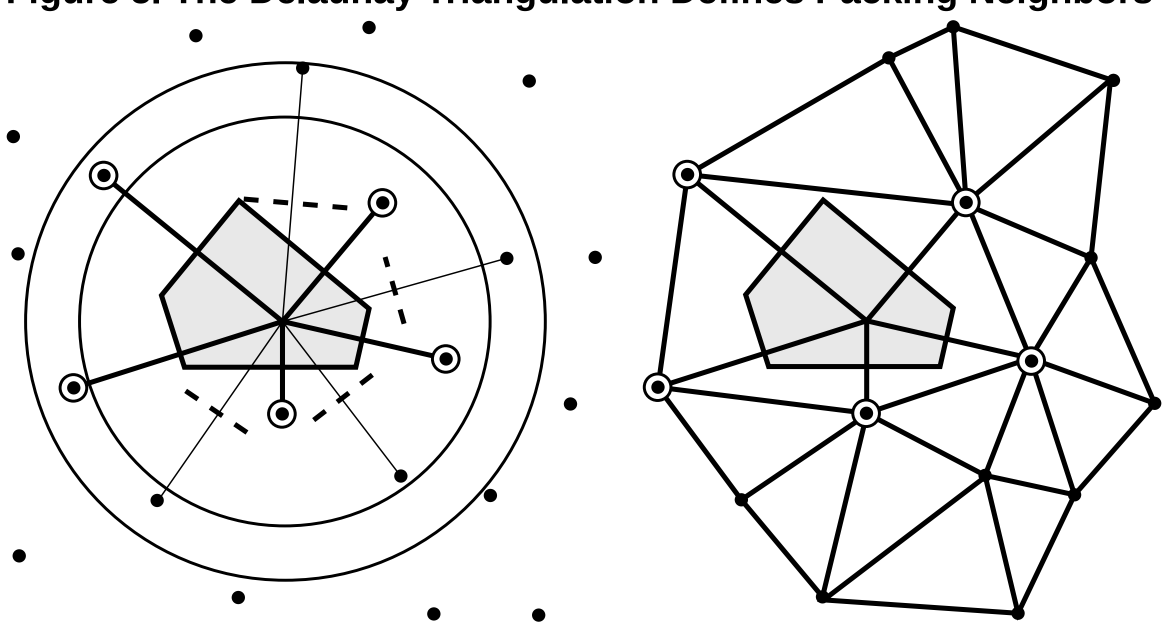

Figure 6. The Problem of the Protein Surface

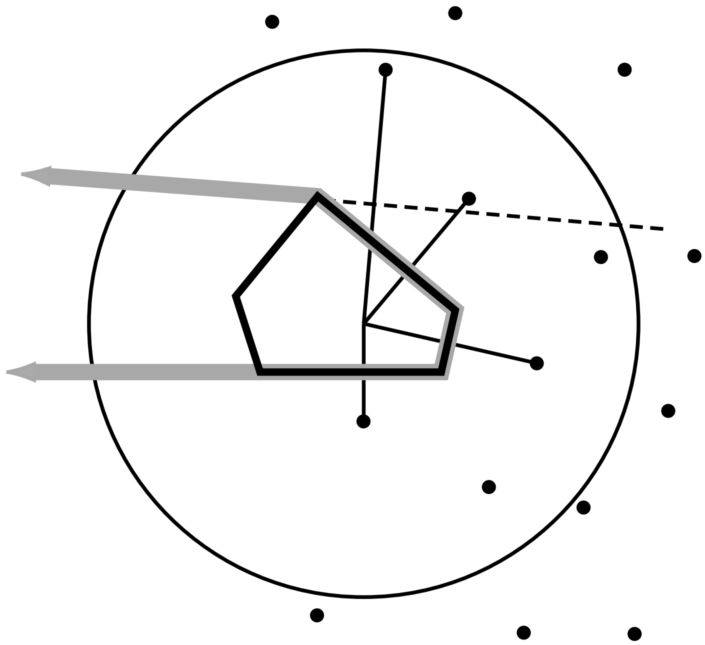

Figure 7. Definitions of the Protein Surface

Figure 7A. Surface Definitions Based on the Probe Sphere

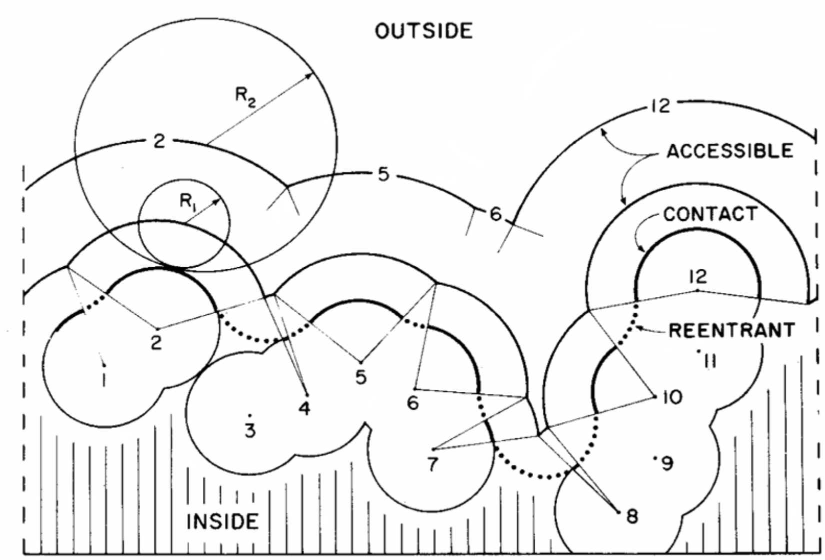

Figure 7. Surface Definitions Based on the Voronoi Construction

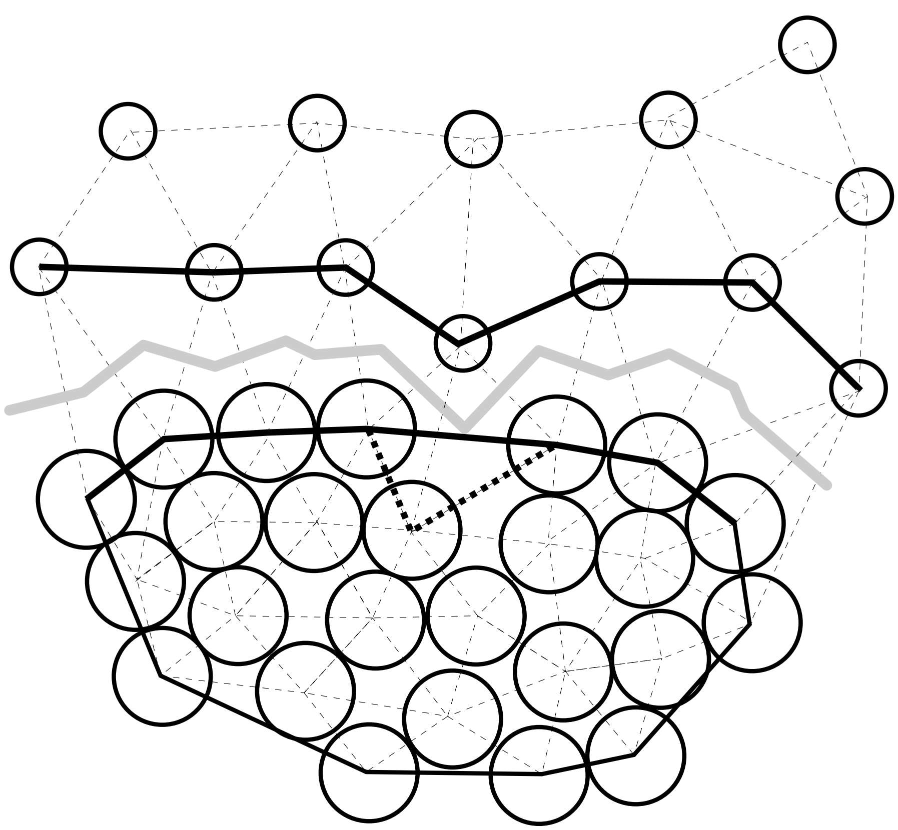

Figure 8. Packing Efficiency

Figure 8A. Measurement of Packing Efficiency in terms of Voronoi Volumes

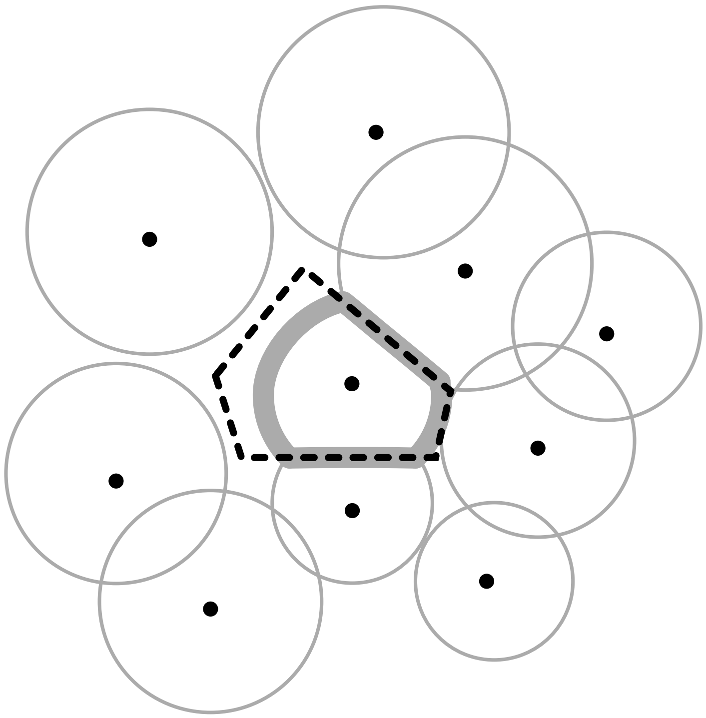

Figure 8B. Illustration of Tight Packing vs. Loose Packing

(This figure should appear in color. See

bioinfo.mbb.yale.edu/geometry/pic/hoh-v-ar.rot.jpg

and

http://bioinfo.mbb.yale.edu/geometry/pic/hoh-v-ar.jpg

)