Are Proteins Well-Packed?

Jie Liang* and Ken A. Dill†
*Department of Bioengineering, University of Illinois at Chicago, Chicago, Illinois 60607-7052 and †Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143-1204

ABSTRACT The average packing density inside proteins is as high as in crystalline solids. Does this mean proteins are well-packed? We go beyond average densities, and look at the full distribution functions of free volumes inside proteins. Using a new and rigorous Delaunay triangulation method for parsing space into empty and filled regions, we introduce formal definitions of interior and surface packing densities. Although proteins look like organic crystals by the criterion of average density, they look more like liquids and glasses by the criterion of their free volume distributions. The distributions are broad, and the scalings of volume-to-surface, volume-to-cluster-radius, and numbers of void versus volume show that the interiors of proteins are more like randomly packed spheres near their percolation threshold than like jigsaw puzzles. We find that larger proteins are packed more loosely than smaller proteins. And we find that the enthalpies of folding (per amino acid) are independent of the packing density of a protein, indicating that van der Waals interactions are not a dominant component of the folding forces.

INTRODUCTION

What is a good model for the packing inside a protein? Is a protein packed more like a liquid or a solid? Based on the observations of high packing densities (Richards, 1977) and low compressibilities (Gavish et al., 1983), protein cores are often considered to be more like solids than liquids (Chothia, 1984; Murphy and Gill, 1991). Packing in proteins was first analyzed quantitatively by Richards (Richards, 1974) and Finney (Finney, 1975). They used a Voronoi analysis for proteins in a space-filling model, where each atom is taken to be a sphere with a fixed radius, given by the van der Waals radius. These and other classic papers showed that the average packing density in a protein is as high as that inside crystalline solids (Chothia, 1975; Harpaz et al., 1994; Gerstein and Chothia, 1996).

In contrast, proteins are tolerant to mutations (Lim and Sauer, 1989; Shortle et al., 1990; Richards and Lim, 1994; Axe et al., 1996), suggesting that proteins can be regarded as plastic or liquid-like. It is not unreasonable to believe that different properties reflect different aspects of packing. A recent review discussed the implications of packing to protein folding (Honig, 1999).

To analyze the volume of a protein molecule, it is not sufficient to take a simple sum of the atomic volumes of its atoms. There are two reasons. First, each atom in a molecule is not an isolated sphere; the geometric description of each atom depends on spatial and connected neighbors. For example, covalently linked atoms are not simply adjacent spheres of the appropriate van der Waals radii, because the bonding between them shortens their separation to the point that a van der Waals sphere representation would have volume overlaps. There are examples of overlaps involving up to seven atoms (Petitjean, 1994). It is important to find better ways to measure and represent overlaps because they can be responsible for errors in resolution and in fitting models to nuclear magnetic resonance and crystal structures of proteins (Word et al., 1999). Second, inside proteins, there are unfilled spaces, not occupied by any atomic spheres. Such free volume is often described by terms such as cavities, voids, pockets, etc. The overlaps and unfilled empty spaces are sometimes correlated with each other. For example, overlapping atoms can sometimes fully enclose an unfilled empty space, causing a void. Any treatment of protein packing must treat overlaps and empty spaces.

Here we describe a way to model the packing of proteins based on developments in computational geometry (Edelsbrunner and Mücke, 1994; Edelsbrunner, 1995; Liang et al., 1998a; Edelsbrunner et al., 1998). The "alpha shape method" we describe treats volume overlaps fully and accurately. We use this method to study how free space is distributed in proteins. We also formalize the definitions of the interior packing density P_i, the surface packing density P_s, and the total packing density P_t. We calculate these parameters for a set of 636 proteins for which atomic structures are known.

At the same time, we address a traditional problem with Voronoi methods. Voronoi methods are challenged to determine where interior empty space ends and the outside bulk begins. Standard methods often require the creation of fictitious surface solvent molecules, but such choices can be arbitrary. This problem has been referred to as the "can-of-worms problem of molecular speleologists" (Kleywegt and Jones, 1994). Here we solve this problem by using the concept of the "convex hull" to define the boundaries of surface pockets and depressions of a molecule (Preparata and Shamos, 1985; Edelsbrunner, 1987; O'Rourke, 1994). To illustrate the convex hull, consider a two-dimensional

Received for publication 13 February 2001 and in final form 2 May 2001.
Address reprint requests to Jie Liang, Dept. of Bioengineering, 851 S. Morgan St., Room 218, SEO, MC063, University of Illinois at Chicago, Chicago, IL 60607-7052. Tel.: 312-355-1789; Fax: 312-996-5921; E-mail: jliang@uic.edu.

0006-3495/01/08/751/16 $2.00

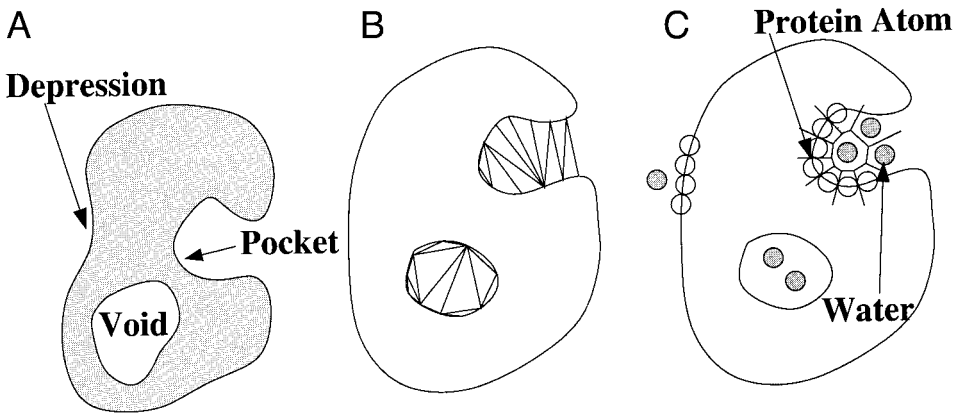

FIGURE 1 The concave regions and surface packing of proteins. (*a*) There are three types of concave regions on protein surfaces: Fully enclosed voids with no outlet, pockets accessible from the outside but with constriction at mouths, and shallow depressions. (*b*) In the analysis of protein self packing, the volume of interior voids and surface pockets are calculated based on Delaunay triangulation and alpha shape described later. All water molecules need to be removed first. (*c*) Protein–water packing has been studied by measuring the volume of Voronoi polyhedra of water molecules and neighboring protein atoms (Harpaz et al., 1994; Gerstein et al., 1995; Gerstein and Chothia, 1996). Here water molecules define the Voronoi volume of protein surface atoms.

(2D) example of nails hammered into a board. Stretching a rubber band around the entire collection of nails forms the convex hull. In three dimensions, a convex hull would be a foil tightly wrapped around a collection of points. Each planar triangular piece of such a foil passes through atom center points. Collectively, the triangular pieces on the convex hull form the boundaries between the empty spaces of the molecule and the external surrounding medium. The empty spaces that we want to identify and measure are the spaces contained inside the convex hull that are unoccupied by protein atoms. All empty spaces can be classified as either voids, pockets, or depressions, as described in the next section.

COMPUTATIONAL MODELS AND METHODS

In this section, we review the computational methods that are used for studying packing in proteins. Details of these methods have been described previously.

Empty spaces in proteins: voids, pockets, and depressions

The external surface of a protein can be represented in various ways. On the one hand, the outermost atoms can be represented as atomic spheres having the appropriate van der Waals radii. On the other hand, a molecule can be defined by the solvent-accessible surface model of Lee and Richards (Lee and Richards, 1971). The solvent-accessible molecule is enclosed by the surface swept out by the center of the spheres of the probe balls. In this case, the surface of the protein is defined by the sum of the van der Waals radii of the outermost atoms plus the solvent probe spheres. We use the term "inflated atoms" to refer to the atomic spheres expanded to include the radius of the probe atom.

To avoid ambiguity, we define some terms. There are three types of empty spaces in proteins. Voids are unfilled spaces inside the protein that are fully enclosed by inflated atoms. A void is sufficiently enclosed that a probe ball is too large to escape. Voids are traps for probe balls. Pockets are caverns that open to the outside of the protein through mouths that are small relative to cavern dimensions but big enough that the probe ball has access to the outside of the molecule. The mouth of a pocket is narrower than at least one cross section of the interior of the pocket. Depressions are concave regions on protein surfaces that have no constriction at the mouth. From the deepest part toward the outside, a depression widens monotonically (see Fig. 1) (Edelsbrunner, 1995; Edelsbrunner et al., 1998; Liang et al., 1998b,c).

Our interest here is in "internal" packing, so we focus on voids and pockets, and do not consider depressions to be a component of packing. Whether a region is a pocket or a void can depend on the size of the probe. If a probe is small enough to escape to the outside of the protein, it is in a pocket. If the probe is too big to escape to the outside, it is in a void (Edelsbrunner et al., 1998). Once a probe radius is defined, there is no ambiguity.

Several important studies of packing in proteins have been carried out (Chothia, 1975; Rashin et al., 1986; Hubbard et al., 1994; Hubbard and Argos, 1994; Gerstein and Chothia, 1996; Liang et al., 1998b,c). Usually voids are not distinguished from pockets, and both are often referred to as cavities. For void detection and calculation, there are many numerical methods (Lee and Richards, 1971; Shrake and Rupley, 1973; Voorintholt et al., 1989; Pascual-Ahuir and Silla, 1990; Ho and Marshall, 1990; Delaney, 1992; Kleywegt and Jones, 1994), and several analytical methods (Connolly, 1983; Richmond, 1984; Gibson and Scheraga, 1987; Perrot et al., 1992; Edelsbrunner, 1995; Sastry et al., 1997; Liang et al., 1998a). It is more challenging to identify pockets and calculate their sizes. Most studies use a variation of a void-detection algorithm, by inflating the solvent probe radius to a size that traps it.

In this study, we apply methods based on the alpha shape theory from computational geometry to identify and measure both voids and pockets (Edelsbrunner and Mücke, 1994; Edelsbrunner, 1995; Edelsbrunner et al., 1995, 1998; Liang et al., 1998a,b,c). This method has been shown to be useful in a number of applications (Akkiraju and Edelsbrunner, 1996; Peters et al., 1996; Kim et al., 1997; McGee et al., 1998; Liang and Subranmaniam, 1997; Liang and McGee, 1998).

The alpha shape method is based on the Delaunay triangulation, which is mathematically equivalent to Voronoi diagram (Richards, 1974; Finney, 1975; Chothia, 1975). To illustrate, Fig. 2 *a* shows a 2D molecule formed by a collection of disks of uniform size. The Voronoi diagram is shown in Fig. 2 *a*. Each Voronoi cell is defined by its boundaries, shown as dashed lines in the figure. Every Voronoi edge is a perpendicular bisector of the line between two atom centers. Each Voronoi cell contains one atom, and

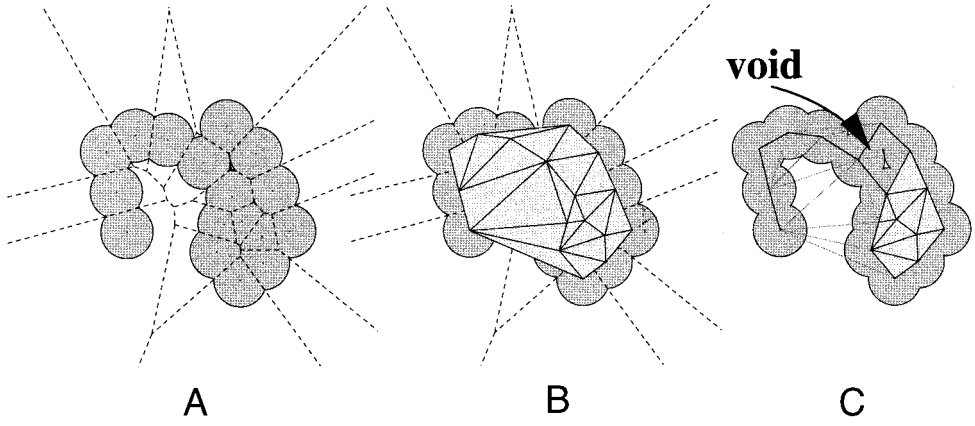

FIGURE 2 Geometry of a simplified 2D model molecule to illustrate the procedure mapping the Voronoi diagram to the Delaunay triangulation. (*a*) The molecule formed by the union of atom disks of uniform size. Voronoi diagram is in dashed lines. (*b*) The shape enclosed by the boundary polygon is the convex hull. It is tessellated by the Delaunay triangulation. (*c*) The alpha shape of the molecule is formed by removing those Delaunay edges and triangles that are not completely contained within the molecule. A molecular void is represented in the alpha shape by two empty triangles.

every point inside a Voronoi cell is closer to this atom than to any other atom. Three connected Voronoi edges meet at a Voronoi vertex.

Using this construction, the Delaunay triangulation can be explained by the following simple exercise. For each Voronoi edge, connect the corresponding two atom centers with a line segment, and for each Voronoi vertex, place a triangle spanning the three atom centers of the three Voronoi cells. Completing this for all Voronoi edges and Voronoi vertices gives a collection of line segments and triangles. Together with the vertices representing atom centers, they form the Delaunay complex, which is the underlying structure of Delaunay triangulation. The Delaunay triangulation for the 2D molecule is shown in Fig. 2 *b*. This unique triangulation has the property that the interior of the circumscribed circle of any triangle does not contain a fourth atom center. Delaunay triangulation can be computed efficiently (Lawson, 1977; Joe, 1991; Guibas et al., 1992; Edelsbrunner and Shah, 1996).

Now remove all Delaunay edges (or line segments) where the two atoms have no two-body volume overlaps. And remove all Delaunay triangles where the three atoms do not have three-body volume overlaps (see Fig. 2 *c*). These triangles are the empty Delaunay triangles. The subset of the Delaunay complex formed by the remaining triangles, edges and vertices (atom centers) is called the "dual complex." It is also often called the "alpha complex" or the "alpha shape."

The alpha shape contains much geometric information about the molecule, including its volume and area (Liang et al., 1998a). The empty spaces, i.e., the voids, pockets, and depressions, can be identified and mapped from the empty Delaunay triangles in the alpha shape. Furthermore, an empty triangle is merged with its neighboring empty triangle(s), if the interface Delaunay edge is also removed by virtue of no two-body volume overlaps between two atoms. In the end, we have a discrete collection of sets. Each set contains a number of merged empty triangles and reflects one of the discrete empty spaces. Figure 2 *c* shows a void in the alpha complex formed by two empty triangles. The size of a void, pocket, or depression can be determined from its corresponding set of empty triangles: the sum of the areas of all the triangles is calculated, from which we subtract the area of the fraction (or sector) of each of the three atom disks contained within each triangle. The result after all the subtractions is the actual size of the void, pocket, or depression.

The discrete flow (Edelsbrunner et al., 1998) explains the distinction between a depression and a pocket. It is a way of collecting together connected packets of empty space. The empty triangles in a 2D Delaunay triangulation that are not part of the alpha shape can be classified into obtuse triangles and acute triangles. The largest angle of an obtuse triangle is more than 90 degrees, and the largest angle of an acute triangle is less than 90 degrees. An empty obtuse triangle can be regarded as a "source" of empty space that "flows" to its neighbor, and an empty acute triangle a "sink" that collects flow from its obtuse empty neighboring triangle(s). In Fig. 3 *a*, obtuse triangles 1, 3, 4, and 5 flow to the acute triangle 2, which is a sink. Each of the discrete empty spaces on the surface of protein can be organized by the flow systems of the corresponding empty triangles: those that flow together belong to

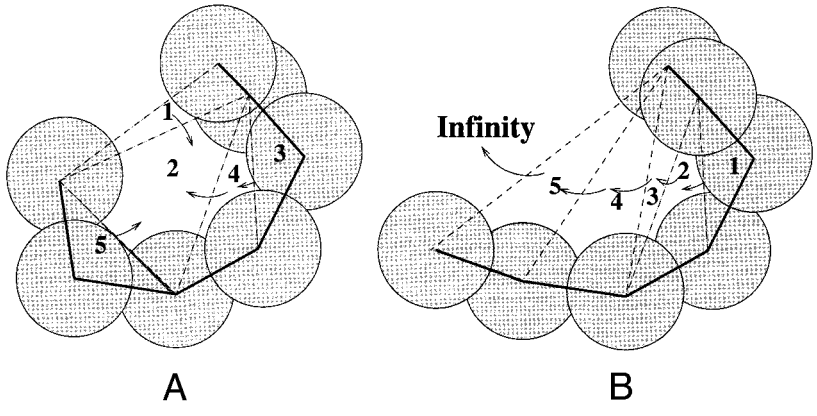

FIGURE 3 Discrete flow of empty space illustrated for 2D disks. (*a*) Discrete flow of a pocket. Triangles 1, 3, 4, and 5 are obtuse. The free volume flows to the "sink" triangle 2, which is acute. (*b*) In a depression, the flow is from obtuse triangles to the outside.

the same discrete empty space. For a pocket, there is at least one sink among the empty triangles. For a depression, all triangles are obtuse, and the discrete flow goes from one obtuse triangle to another, from the innermost region to outside the convex hull. The discrete flow of a depression therefore goes to infinity. Figure 3 *b* gives an example of a depression formed by a set of obtuse triangles.

All of the concepts above have counterparts in three-dimensional (3D) space. For example, the convex hull of a 3D molecule is a convex polytope rather than a polygon, and it is tessellated by tetrahedra rather than triangles. Furthermore, everything we have discussed can be generalized to real molecules, where atoms have different radii. Now we no longer choose the Voronoi planes to be bisectors that would be equidistant between two atom centers (Richards, 1977). Instead, we use the radical plane, defined by the property that any point on it will have equal lengths of tangent line segments to the two atoms (Gellatly and Finney, 1982). The result is that the Voronoi diagram is now correctly weighted by atomic radius.

The corresponding Delaunay triangulation weighted by atomic radius can be found efficiently (Edelsbrunner and Shah, 1996; Facello, 1995). In the weighted Delaunay triangulation, an atom is represented by the coordinates of its center $q \in \mathbb{R}^3$ in 3D space, and its weight $W_q = r_{q2}$, where r_q is the radius of the atom. The square of the length of the tangent line segment between a point $p \in \mathbb{R}^3$ and an atom with atomic center q is defined as the "power distance" $\pi_p(q)$, where $\pi_p(q) = |pq|^2 - W_q$, and $|pq|$ is the Euclidean distance between p and q. This definition of weight W_q and the use of power distance $\pi_p(q)$ ensures that the corresponding weighted Voronoi cells are polyhedra with planar boundaries rather than curved surfaces. For a tetrahedron *abcd*, there is a unique point *z*, the orthogonal center, that has the same power distance W_z to the four atom centers at points *a*, *b*, *c*, and *d*. Similar to the example in 2D space, voids and pockets of a molecule are represented by discrete sets of empty Delaunay tetrahedra that are not contained within the alpha shape. Again, pockets are organized by discrete flow. The discrete flow in 3D space is defined as the following. If two empty Delaunay tetrahedra τ and σ share an interfacial triangle ϕ, and if the orthogonal center z_τ of τ is not contained in the interior of τ (but instead resides on the other side of the plane where ϕ is embedded), then we say that tetrahedron τ flows to tetrahedron σ. This is a generalization of the idea that an obtuse triangle flows to its neighbor to the case of 3D space. It allows us to treat nonuniform atomic radii (Edelsbrunner et al., 1998).

After tetrahedra representing a pocket are organized by the discrete flow, we can calculate the volume of the pocket. We first calculate the volume of each empty tetrahedron. Because this tetrahedron is empty, the four atoms at its four vertices do not completely fill the volume of the tetrahedron. The fractions of the four atoms contained within the tetrahedron are then subtracted from the volume of this tetrahedron. We repeat this process for all empty tetrahedra organized by the same discrete flow. In the end, the sum of the remaining volumes of each tetrahedron gives the volume of the pocket, through an analytical expression. Details of such calculations can be found in (Edelsbrunner et al., 1995; Liang et al., 1998a).

Computing surface and interior packing densities

"Packing density" describes how effectively atoms fill space. It is defined as the amount of the space that is occupied within the van der Waals envelope of the molecule divided by the total volume of space that contains the molecule (Richards, 1974; Richards and Lim, 1994). But what space is occupied by the molecule? This is addressed by the Voronoi polyhedra method (Richards, 1974; Finney, 1975; Chothia, 1975; Harpaz et al., 1994; Gerstein et al., 1995; Gerstein and Chothia, 1996; Gerstein and Richards, 1999). Because each Voronoi cell contains one atom, and everywhere inside the cell is closer to that atom than to any other atoms, Voronoi cells provide a natural definition for the space that can be assigned to an individual atom. Figure 2 *a* shows that a Voronoi region can be either completely filled by atomic spheres, or can contain some empty space that is part of a void, a pocket, or a depression.

The Voronoi method works well for nonsurface atoms, where each Voronoi cell is closed and fully bounded by faces. But since Voronoi cells are defined by neighboring atoms, the Voronoi approach does not work for atoms on the surface. There are no neighboring protein atoms beyond the surface, so the Voronoi cells of convex hull atoms extend to infinity, and cells of surface atoms that are not on the convex hull can have artificially large sizes. This problem is usually addressed by empirical placements of auxiliary heteroparticles such as water molecules around the protein surface atoms, for example, by a periodic grid (Richards, 1974), solvent shell (Finney, 1975) or by molecular dynamics simulations (Gerstein et al., 1995). Where such approaches are problematic is in comparing protein self packing at surfaces versus interiors (see Fig. 1, *b* and *c*), and in treating protein–protein interfaces. Here we are interested in protein self packing at the surface, and water molecules are not required. We avoid the difficulty of the Voronoi method by using the convex hull to bound the molecule. The boundary of the surface pockets and depressions are uniquely defined by the convex hull and the Delaunay triangles of the molecule. Concave regions on the protein surface can be uniquely determined, including their volumes and areas.

We divide the space occupied by a molecule into three parts: the space occupied by the union of atomic spheres, the interior voids, and the surface empty spaces (pockets and depressions). To compute packing densities, we first determine the van der Waals volume of the molecule, where each atom has a fixed van der Waals radius, and no solvent probe is used. We then expand the atom spheres by the radius of the solvent probe (1.4 Å). The alpha shape of the union of the inflated atom balls is then used to compute the volume of the molecule as defined by the molecular surface (or the Connolly surface). Next the volumes of the voids, and the volumes of the pockets are also found, as defined by the molecular surface. These volume measurements are then used to calculate packing densities below.

The van der Waals volume of a molecule is the space taken up by the molecule when all atoms are idealized as balls with fixed van der Waals radii. Solvent is modeled as a spherical probe with a radius of 1.4 Å. When a solvent probe rolls around the van der Waals molecule from the outside, its front sweeps out an envelope surface. A probe sphere that is trapped inside the molecule sweeps out a void or cavity surface. The volume contained within the envelope surface, the envelope volume, minus the cavity volume is the molecular surface volume of the molecule: $V_{ms} = V_{env} - V_{cav}$. It is usually slightly larger than the van der Waals volume, because it includes crevices and other dead spaces at the intersections of atoms, which are not part of the empty space accessible to the probe.

Following Richards' (1974) original definition, we define the interior packing density as the ratio of the van der Waals volume divided by the envelope volume:

$$P_i = \frac{V_{vdw}}{V_{env}} = \frac{V_{vdw}}{V_{void} + V_{ms}}.$$

This parameter takes into consideration both the inaccessible cavities and the small crevice dead space. It gives a good description of the interior packing of a protein, and is also a good approximation of the original definition of Richards. Surface packing density is the ratio of the van der Waals volume divided by the sum of molecular surface volume and the pocket volume. It excludes the interior cavities and therefore describes the packing on the surface between different regions of the protein:

$$P_s = \frac{V_{vdw}}{V_{pocket} + V_{ms}}.$$

Note that P_s does not reflect protein–water packing on surface regions. Total packing density is the ratio of the van der Waals volume divided by

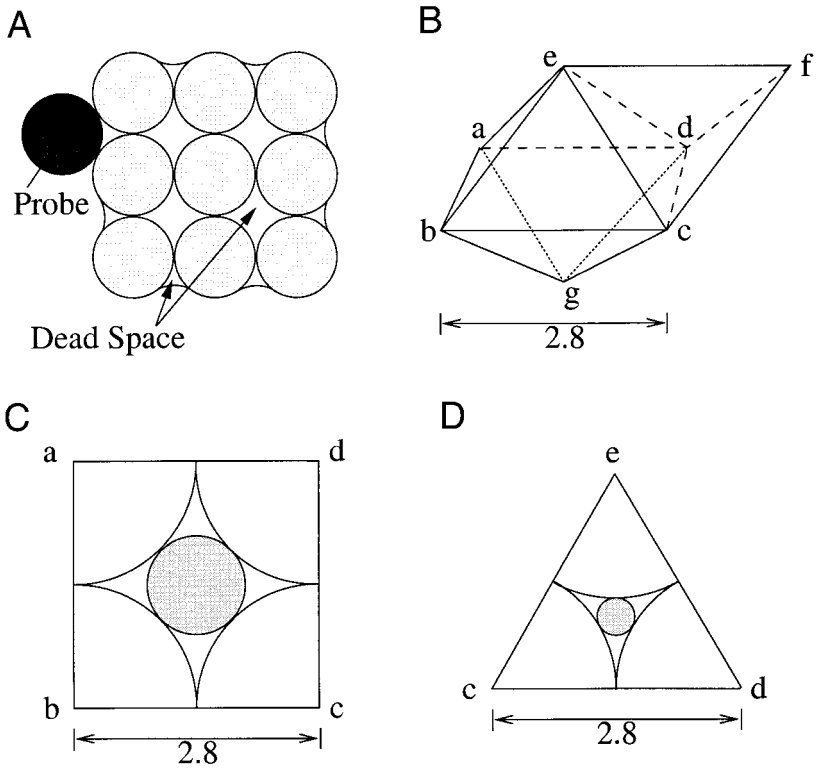

FIGURE 4 Close-packed face-centered solid. (*a*) The dead spaces are the interstitial space inside the Connolly surface envelope but unoccupied by the van der Waals atoms. They are too small to contain the probe. Although such spaces contribute to the packing density, they are not part of the empty space that is accessible to the probe. (*b*) The unoccupied empty spaces are of two types: octahedral space is enclosed by the two pyramids (*a*, *b*, *c*, *d*, *e*) and (*d*, *c*, *b*, *a*, *g*), and tetrahedral space is enclosed by (*c*, *d*, *e*, *f*). (*c*) The planar cross-section of square (*a*, *b*, *c*, *d*) of the octahedral space. The radius of the largest probe that can pass is 0.580 Å. (*d*) The planar cross section of triangle (*c*, *d*, *e*) interfacing the octahedral and tetrahedral spaces. The radius of the largest probe that can pass is 0.217 Å.

the sum of the envelope volume and the pocket volume. It takes both surface pocket and interior voids into consideration, and therefore gives an overall packing description of the protein:

$$P_t = \frac{V_{vdw}}{V_{void} + V_{pocket} + V_{ms}}.$$

RESULTS

We first compare the packing inside proteins to the packing of spheres in crystalline solids and in computer simulations of liquids. We then compare protein packing with that of random spheres.

A model solid: close-packed spheres

We consider two different model close-packed solids, each containing 250 atom balls (simulation results are graciously provided by F. Richards, Yale University). One is a face-centered periodic cubic array; the other is a hexagonal array. The balls have a radius of 1.4 Å. No voids or pockets are found when the solid is sampled with a probe of radius 1.4 Å, indicating that it is close-packed because the empty spaces are too small for access by the probe. Because our results are essentially the same for both data sets (except for small boundary effects), we only describe the face-centered close-packed spheres in detail.

We compute the packing density P for each model solid based on the sum of van der Waals volumes of the atoms, V_{vdw}, and on the volume enclosed within the molecular surface, V_{ms}, found by a 1.4-Å probe. Because a 1.4-Å probe detects neither voids nor pockets, we have $V_{void} = V_{pockets} = 0$ and $V_{env} = V_{ms}$. This gives

$$P = \frac{V_{vdw}}{V_{ms}}.$$

Using the VOLBL program (Edelsbrunner et al., 1995; Liang et al., 1998a), we find $P = 0.74$ for face-centered spheres. For hexagonally close packed spheres, $P = 0.76$. Hence, this simply confirms the tight packing that is well known for close-packed solids (Richards and Lim, 1994). The empty space arises from the unavoidable free volume, as illustrated in two dimensions in Fig. 4 *a*. Here the free volume or dead space is the fraction of space contained within the boundary curve rolled out by the probe (*dark*) but unoccupied by the van der Waals spheres (*gray*). These dead spaces are not accessible to the probe.

A smaller probe can occupy empty space that is not accessible to the larger 1.4-Å probe. We systematically increase the probe radius from 0.1 to 2.4 Å in steps of 0.1 Å. Figure 5, *a* and *b*, show the number of voids detected and the total void volume measured at various probe sizes. Voids in the model solids can only be detected with a very narrow range of probe radii (0.3–0.5 Å). At a radius of 0.1 or 0.2 Å, a probe can meander freely between the atoms, so it detects no isolated voids. It sees only a very large pocket (539 and 455 $Å^3$ at 0.1 and 0.2 Å, respectively). This large

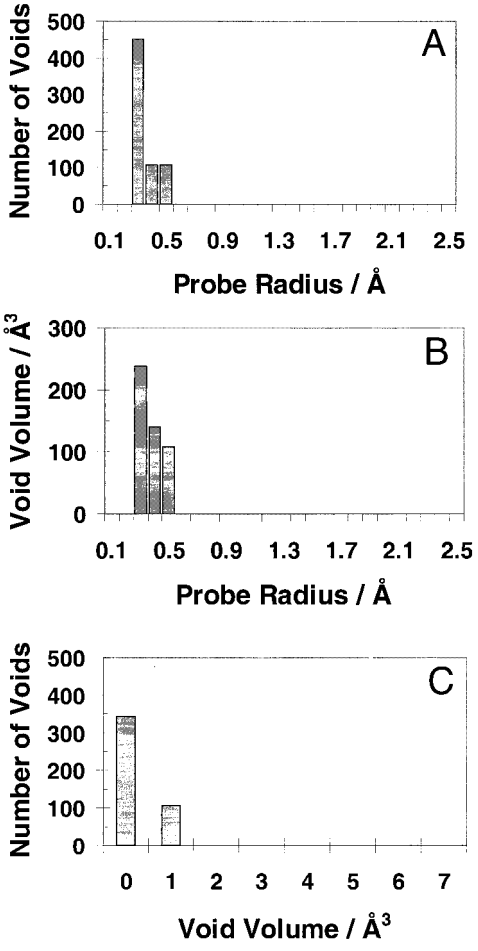

FIGURE 5 Voids and pockets in the model solid. (*a*) The number of voids in the model solid detected by probes of various sizes. Voids are detectable only over a narrow range of probe radii (0.3–0.5 Å). (*b*) Total molecular surface volume of voids of different probe radii. (*c*) A probe of 0.3-Å radius detects 451 voids in the model solid. These voids are of only two sizes: 108 voids are 1.72 $Å^3$ and 343 voids are 0.16 $Å^3$.

pocket represents the interstitial space of the atoms. However, at radius of 0.3 Å, the probe no longer has such freedom of passage. It is big enough to become trapped between atoms, so it now sees space divided up into numerous small isolated voids. Altogether there are 451 such voids. These voids are of only two sizes: 1.72 $Å^3$ for 108 of them, and 0.16 $Å^3$ for 343 of them (Fig. 5 *c*). A probe of radius 0.3 Å can be trapped within voids of either of the two sizes. A probe of 0.4- or 0.5-Å radius is too large for the small voids seen at 0.3 Å; therefore, it sees only the 108 larger voids. A probe larger than 0.6 Å is too big to be trapped by any of the voids.

These observations can be explained by the geometry of face-centered close-packed spheres. Figure 4 *b* shows seven such spheres. Each vertex represents the position of the center of an atom. The distance between neighboring atom centers is 2.8 Å. There are two types of voids: One type with larger size is enclosed within the double pyramids: (*a*, *b*, *c*, *d*, *e*) and (*d*, *c*, *b*, *a*, *g*), which form an octahedron. The other type with smaller size is enclosed within the tetrahedron (*c*, *d*, *e*, *f*). This is why a probe of 0.3 Å detects voids of two different sizes. Figure 4 *c* shows the planar cross-section of square (*a*, *b*, *c*, *d*) in the middle of the octahedron. There are four spheres placed at each of the four corner vertices. The area unoccupied by the four 90°-angled pie-shaped sectors at the corners is the empty space for this 2D cross section. The largest solvent probe that can pass through this space has a radius $(\sqrt{2} - 1) \times 1.4 = 0.580$. No probe of radius >0.580 Å can fit in. This is why neither voids nor pockets are detected when we use probes of 0.6 Å or larger.

Figure 4 *d* shows the planar cross-section of triangle (*c*, *d*, *e*). Again, the atom sectors do not fill the full area of the triangle. The largest probe that can fit in its empty space is $1.4/\cos(\pi/6) - 1.4 = 0.217$ Å. Any probe smaller than 0.217 Å can cross through the triangular interface between the octahedral empty space and the tetrahedral empty space. These probes detect a large pocket that represents the fully connected interstitial space. This is why we detect a large interstitial pocket using probes of 0.1 and 0.2 Å. Probes with radii >0.217 Å detect no pockets. Probes between 0.217 and 0.580 Å cannot cross triangular interfaces, but they still detect voids embedded in octahedral spaces (108 for this data set), and the smaller voids embedded in tetrahedral spaces (343 for this data set).

The packets of free volume inside solids of close-packed spheres are quite uniform. The range of probe radii that can detect pockets or voids is very narrow (0.1–0.2, 0.3–0.5 Å, respectively). Both the total number (451 to 108) and the volumes of the voids monotonically decrease as the probe radius increases. In addition, at probe of 0.3 Å, where the largest number of voids are detected, there are only two different sizes for the 451 voids we find.

Model liquid: random spheres

We also studied the packing of 631 spheres that are distributed randomly and loosely as in a low-density liquid, obtained from computer simulations (Finney, 1970). These spheres have radius 1.4 Å and are uniformly distributed in a roughly spherical space. A probe of 1.4 Å detects a large pocket (39,831 $Å^3$), with 416 mouth openings on the boundary of the collection of spheres. There are no voids, because all the empty space can be reached by the 1.4-Å probe, and the interstitial space of this liquid of random spheres is fully connected.

All the empty space reachable by a 1.4-Å probe is represented by a big pocket. For a 1.4-Å probe,

$$\frac{V_{\mathrm{vdw}}}{V_{\mathrm{ms}}} = 0.82.$$

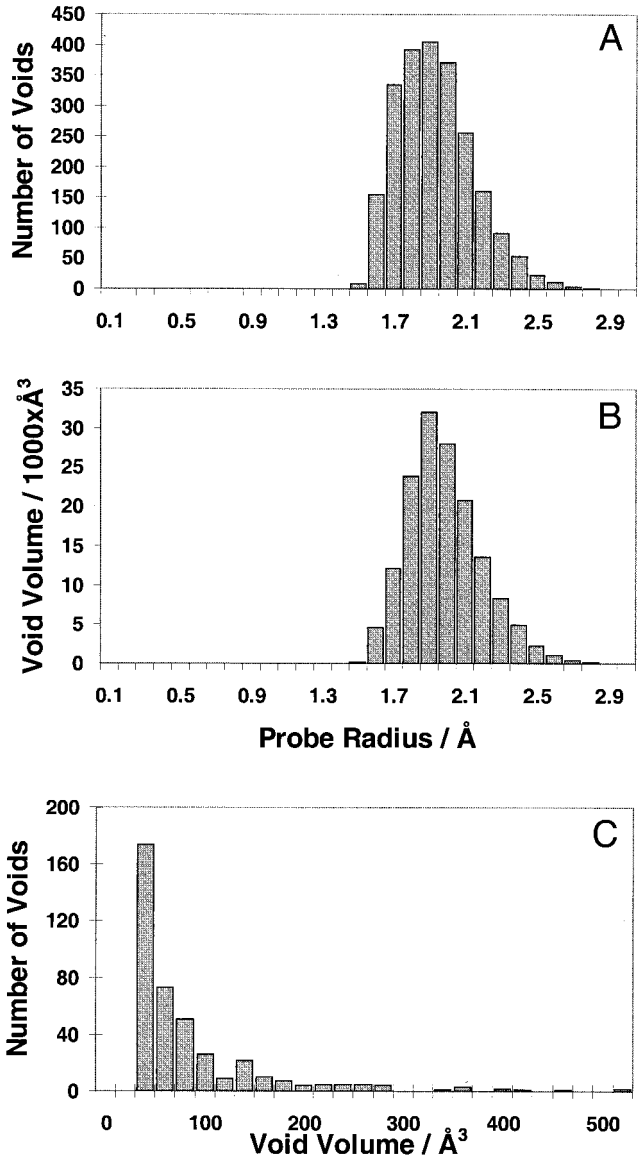

FIGURE 6 Voids in the model liquid. (*a*) The number of voids in the model liquid detected by probes of various sizes. Voids are detectable with a wide range of probe radii (1.5–2.7 Å). (*b*) Total molecular surface volume of voids at different probe radius. (*c*) A probe of 1.9 Å detects the largest number of voids in the model liquid. The probe detects a wide distribution of void sizes.

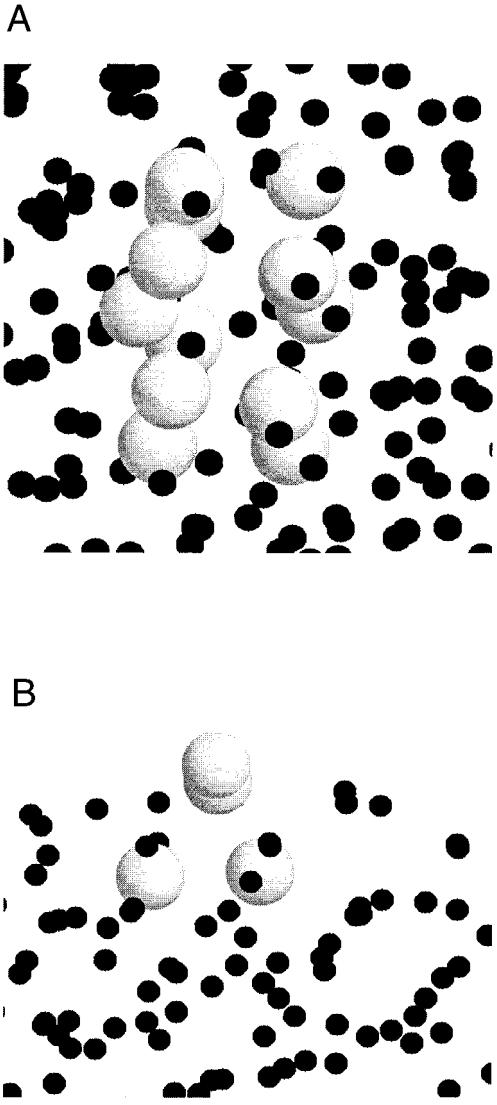

FIGURE 7 Voids in the model liquid. A large void (*a*) and a small void (*b*) in a liquid of random spheres.

Because $V_{\text{void}} = 0$, the total packing density P_t of the model liquid is

$$P_t = \frac{V_{\text{vdw}}}{V_{\text{ms}} + V_{\text{pocket}}} = 0.15.$$

This liquid is very loosely packed, and the interstitial space is fully connected.

We explored the distribution of free space in the liquid by systematically increasing the probe radius from 0.1 to 3.6 Å (Fig. 6). The total numbers and volumes of voids are shown in Fig. 6, *a* and *b*. No voids can be detected with any probe smaller than 1.5 Å because such probes can move freely throughout the space. Beginning at 1.5 Å, the big pocket seen at smaller probe sizes starts to break up into smaller unconnected voids. A probe of 1.9 Å detects the largest number of voids (405). This probe also detects 92 pockets. The distribution of the voids by volume is shown in Fig. 6 *c*. Unlike the model solid (Fig. 5 *c*) where voids are of only two sizes, voids in the model liquid have a broad distribution of sizes. There are a few large voids, and many smaller ones in the liquid. A visualization of a large void and a small void for the 1.9-Å probe is shown in Fig. 7. The model liquid has some voids that are much larger than in the model solid, and has a much broader distribution of void sizes. Probes ranging from 1.5 to 2.7 Å can detect them.

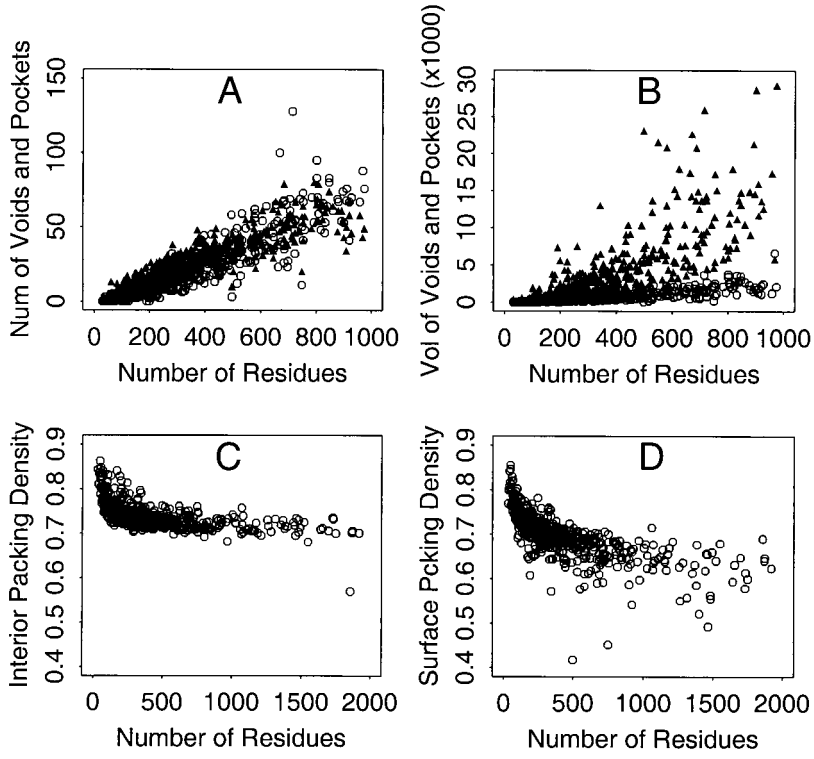

FIGURE 8 Voids, pockets, and packing densities for a set of 636 proteins representing most of the known protein folds. (*a*) The number of voids and pockets detected with a 1.4-Å probe is linearly correlated with the number of residues in a protein. Only proteins with less than 1000 residues are shown. Solid triangles and empty circles represent the pockets and the voids, respectively. (*b*) Total molecular surface volumes of the voids and the pockets does not correlate strongly with the number of residues. (*c*) The interior packing density (P_i) versus the number of residues for a representative set of proteins. The mean value is 0.74, with a standard deviation of 0.03. (*d*) The surface packing density (P_s) for the same set of proteins. The mean value for P_s is 0.70, with a standard deviation of 0.05. There are more outliers from the mean P_s.

The packing in the model liquid is loose and irregular. Both the total number and total volume of the voids follow skewed Gaussian-like distributions, unlike the solids, which have monotonic decreases in voids as a function of size and number. In addition, at a probe size of 1.9 Å, where the largest number of voids is detected, the distribution of voids is very broad.

Surface pockets, interior voids, and protein packing densities

Now we focus on voids and pockets in proteins. We studied a subset of 636 proteins from the November 1999 release of PDBSELECT, which represents a good sampling of all known protein folds (Hobohm and Sander, 1994). For these proteins, sequence identity between any two structures is less than 25%. The volumes, areas of pockets and void distributions for each protein are computed using a 1.4-Å probe. The van der Waals radii of protein atoms are from Table 2 of Tsai et al. (1999). Computation using OPLS radii ($0.5 * \sigma$) (Jorgensen and Tirado-Rives, 1988) qualitatively gives the same conclusions, although the specific values are different (data not shown).

Figure 8 *a* shows the numbers of pockets (*solid triangles*) and the numbers of voids (*empty circles*) plotted against the number of residues in each protein. For proteins with less than 1,000 residues, both are approximately linear, although the deviation increases with the size of the protein. For every additional 100 residues, a protein has about an additional 7–8 voids and 7–8 pockets. These spaces are found by a 1.4-Å probe, so they are large enough to contain at least one water molecule. The linear relationship between the numbers of voids and pockets and the number of residues has been observed in a small set of enzymes (Liang et al., 1998c). Figure 8 *b* shows that, although the number of voids is correlated with chain length, the total pocket volume does not correlate well.

For proteins with less than 2000 residues, Fig. 8 *c* shows that the packing densities are negatively correlated with the chain length, especially for proteins with chain length less than 200 residues. The mean interior packing density is ~0.74, similar to model solids, with a standard deviation of 0.03. The surface-packing density P_s has a mean value of 0.70, with a standard deviation of 0.05 (Fig. 8 *d*). Surface packing is looser in larger proteins. We find that the surface of a protein is never packed more tightly than its interior. This confirms results from the study by Gerstein and Chothia (1996).

Many proteins that are loosely packed (small P_t) are formed by multiple subunits and contain tunnels or holes of large size (bacterioferritin, 1bfr, 0.44; chaperone GP31, 1g31, 0.45; importin beta subunit, 1qgk, 0.53; synthase/riboflavin synthase complex of bacillus subtilis, 1rvv, 0.46; ribulose 1,5-bisphosphate carboxylase/oxygenase, 8ruc, 0.46; 20S proteasome from yeast, 1ryp, 0.50; matoprin, 1mpq, 0.54). These tunnels or holes may be important for function. The protein with the smallest P_t (0.42) is methylamine dehydrogenase (1mae). In contrast proteins that are

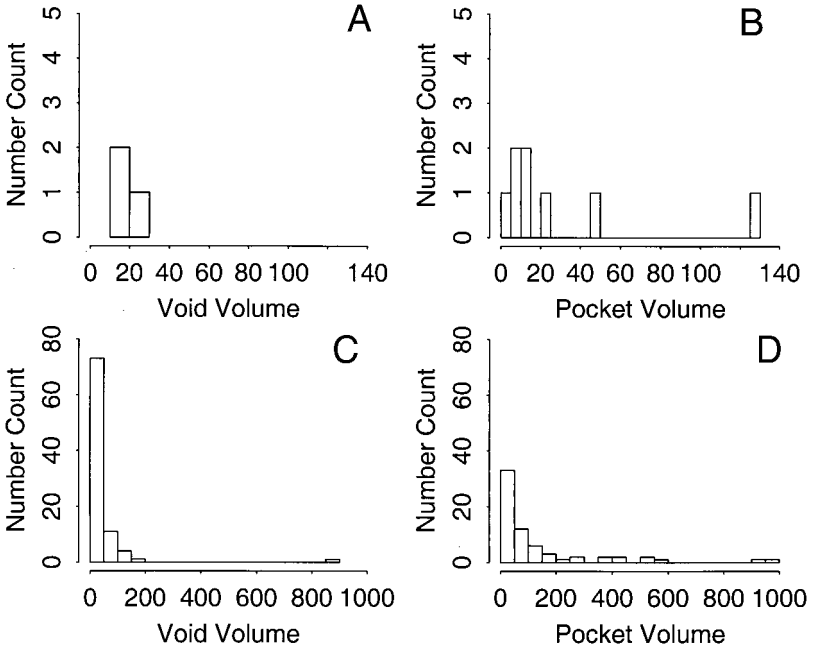

FIGURE 9 The size distributions of pockets and voids for a small and a large protein. The histograms are not to the same scale. The number of the voids (*a*) and pockets (*b*) in ribonuclease F1 (1fut), a small protein, and the number of the voids (*c*) and pockets (*d*) in glycogen phosphorylase (1gpb), a larger protein. The larger protein has many more pockets on the surface, and the range of void and pocket sizes is also broader.

packed well (high P_t) are often small proteins or polypeptides of simple shape that do not have enough residues to form complex structures (e.g., crambin, 1cbn, 0.84; ROP protein, 1nkd, 0.83; scorpion toxin variant-3, 2sn3, 0.82).

Figure 9 compares the distribution of empty space in a small well-packed protein, ribonuclease F1 (106 residues, pdb code 1fut, interior packing density $P_i = 0.76$, surface packing density $P_s = 0.75$, and total packing density $P_t = 0.74$), with a large poorly packed protein, glycogen phosphorylase B (766 residues, 1gpb; interior, surface and total packing densities are 0.71, 0.68, and 0.65, respectively). In the latter, packing is more heterogeneous, and there are some large empty spaces, whereas the well-packed protein has few large packing defects.

Packing in proteins is heterogeneous

Figure 10 *a* shows the distribution of voids and pocket volumes in a set of 12 proteins that were matched in their packing densities (P_t ranges only between 0.688 and 0.723), and sizes (ranging only between 185 and 190 residues). The distributions of free volumes are all rather similar. They show that most free volumes in proteins are in small packets, but that there are also usually a few large defects. These distributions look remarkably similar to those of the model liquid, Fig. 6 *c*. Hence, although the average packing density in these proteins resembles that of a solid, the distribution function resembles that of a liquid. The issue of whether a protein is more like a liquid or like a solid can depend on what is the property of interest.

Figure 10 *b* shows the free-volume distributions of 34 different proteins that cover a broad range of average packing densities (P_t of 0.42–0.86 for the subset from PDBSELECT). This figure shows that poorly packed proteins have more small cavities and more large cavities than better packed proteins.

Figure 10 *c* shows the averages of normalized free-volume distributions of proteins with P_t between 0.60 and 0.86. Proteins are grouped together into bins, each with P_t ranging within 0.02 units. This figure shows that worse packing in proteins is not attributable to voids of any particular size; there are just more voids of all sizes.

Proteins packing: jig-saw puzzles or random spheres?

Another way to characterize internal packing is through surface/volume relationships. For example, for a perfectly solid 3D sphere of radius r, the relationship between volume $V = 4\pi r^3/3$ and surface area $A = 4\pi r^2$ is $V \propto A^{3/2}$. In contrast, Fig. 11 *a* shows that the van der Waals volume scales linearly with the van der Waals surface areas of proteins. The same linear relationship holds irrespective of whether we relate molecular surface volume and molecular surface area, or solvent-accessible volume and solvent-accessible surface area (data not shown).

A model for disordered materials is clusters of random uncorrelated spheres, which has a characteristic scaling behavior (Lorenz et al., 1993). Monte Carlo studies show that the volume V of clusters of random spheres, of either uniform radius or of mixtures of different radii, scales linearly with the surface area A of the cluster: $V \propto A$ (Lorenz et al., 1993; Stauffer, 1985). The same scaling is found in lattice models of simple clusters (Stauffer, 1985). This

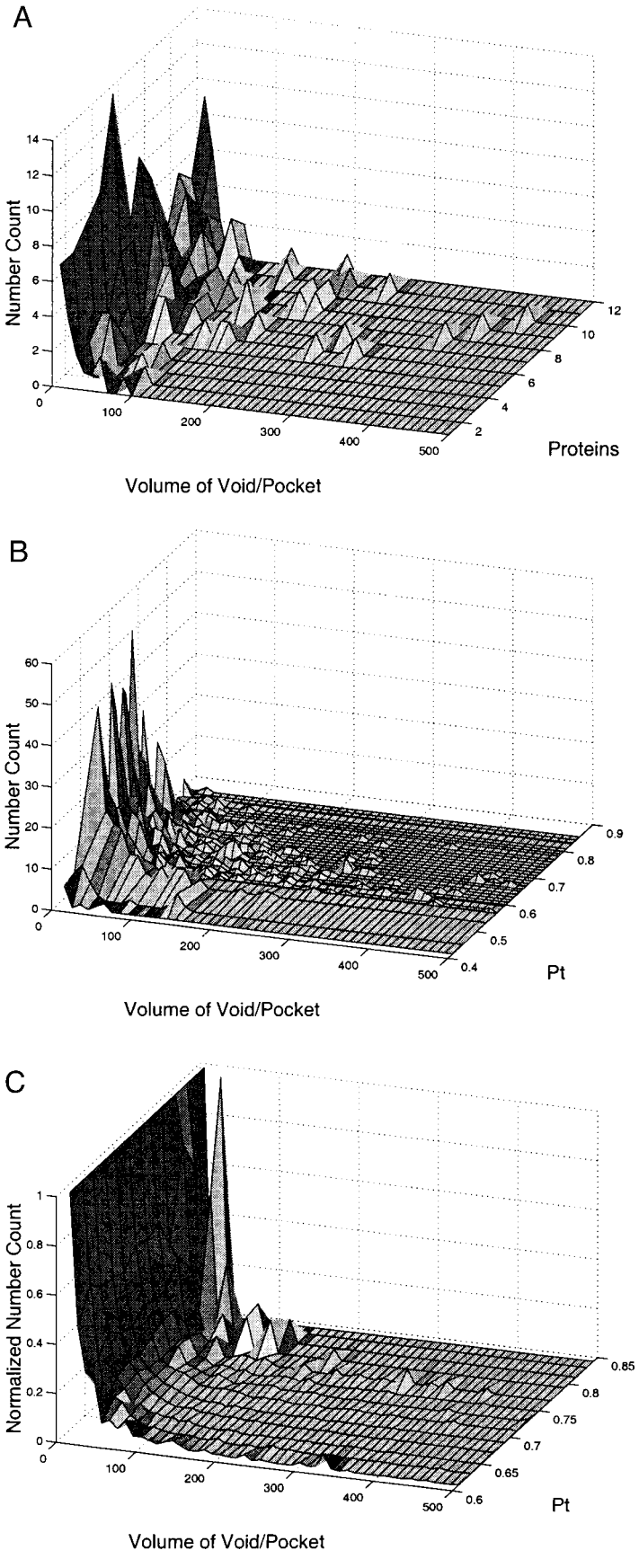

FIGURE 10 The free-volume distributions of the pockets and the voids of a sample of many proteins. Here only voids/pockets smaller than 500 Å are counted. (*a*) The free volume distributions of a set of 12 proteins of similar packing (P_t between 0.688 and 0.723) and similar size (the number of residue between 185 and 190). The distributions vary across different proteins, but all look rather similar. The pdb names and P_t of the 14 proteins are: 1gky 0.688, 153l 0.692, 1xnb 0.703, 2sas 0.703, 1mol 0.705, 1rec 0.708, 1a2x 0.711, 1ay7 0.711, 3tss 0.712, 1ygs 0.717, 1knb 0.723, and 2gar 0.723. The proteins are ranked on the *Proteins* axis in ascending order of P_t. (*b*) The free volume distributions of a set of 34 proteins with different P_t, indicating the heterogeneous nature of protein packing. The pdb names and P_ts of these proteins are: 1mae 0.417, 1g31 0.452, 1qgk 0.534, 9wga 0.569, 2cas 0.571, 1pre 0.580, 1eai 0.596, 1ps1 0.601, 1p32 0.610, 1p35 0.620, 1cfm 0.630, 1ak4 0.640, 1cyd 0.652, 1lbe 0.660, 1a28 0.670, 1cmk 0.671, 1adw 0.680, 1ac5 0.690, 1aun 0.700, 1aep 0.710, 1aw8 0.720, 1rie 0.730, 1bxo 0.740, 2spc 0.750, 1gvp 0.761, 1ben 0.770, 1utg 0.783, 1hta 0.791, 2cbp 0.800, 1mof 0.810, 1nkd 0.829, 2erl 0.832, 1cbn 0.843, and 2ifo 0.855. (*c*) The averages of normalized free volume distributions of proteins with P_t between 0.60 and 0.86. Proteins within 0.02 unit of P_t are grouped together, and their average distributions are plotted. The distribution of free volumes is about the same in well-packed as in poorly packed proteins. Worse packing in proteins is not attributable to voids of any particular size.

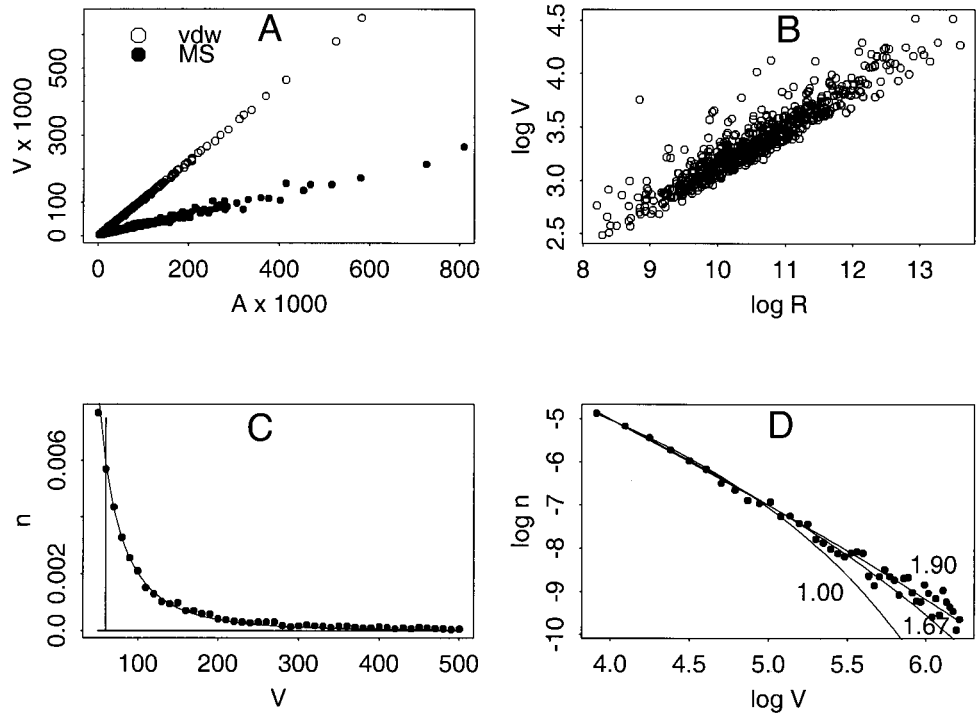

FIGURE 11 The scaling behavior of the geometric properties of proteins. (*a*) The van der Waals (*vdw*) volume and van der Waals area of proteins scale linearly with each other. Here, the van der Waals volume is the volume of the union of overlapping atom balls adopting van der Waals radii. Similarly, molecular surface (*MS*) volume also scales linearly with molecular surface area using a probe radius of 1.4 Å. The proteins are the same as in Fig. 8. (*b*) The logarithm of protein molecular surface volume log V scales with the logarithm of the length of the protein log R with the characteristic slope d of 2.47. For van der Waals and solvent-accessible volumes, $d = 2.42 \pm 0.03$ and $d = 2.34 \pm 0.04$. (*c*) The size distribution n_V of voids and pockets of volume V. n_V (*solid circles*) is the number of voids and pockets of volume V, divided by the chain length of the protein. Because voids and pockets are sparse in proteins, n_V is the average of all proteins with chain length ≥200. n_V follows the curve $n_V \propto V^{-\theta} c_0^V$, where $\theta = 1.67 \pm 0.07$, $c_0 = 0.9966 \pm 0.0008$. Note that the size distribution of voids and pockets of regular packing as in a model of solid would be a delta function, as also shown figuratively here. (*d*) The log–log plot of n_V and V indicates that $n_V \propto V^{-\theta} c_0^V$ fits the data well with $\theta = 1.67$ or $\theta = 1.90$, but not when $\theta = 1.0$.

linear relation of V with A is also what we observe in proteins (Fig. 11 *a*). Similarly, there is a linear correlation between the volume and the number of atoms N (or the number of residues) of proteins (data not shown), the same as observed in random-packed spheres (Lorenz et al., 1993).

A key property of randomly packed spheres is the so-called percolation threshold. For randomly packed spheres, when the packing density p is greater than a threshold density p_c, clusters become connected to each other, and the size of the largest cluster approaches the size of the whole system (Meester et al., 1994; Lorenz et al., 1993; van der Marck, 1996). At this threshold p_c, the volume V of a cluster of random spheres is known to scale with the length R of the cluster as $V \propto R^D$, with a characteristic exponent $D = 2.5$ in 3D space (Stauffer, 1985; Lorenz et al., 1993). The same exponent is also found in lattice models of clusters (Adler et al., 1990). The size R of a cluster of spheres can be calculated as the maximum extent of the cluster along the coordinate axes

$$R = \frac{1}{2d} \sum_{j=1}^{d} (x_{j,max} - x_{j,min}),$$

where $d = 3$ in $\mathbb{R}^3$ (Lorenz et al., 1993). For random-packed spheres, $D = 2.5$ at $p = p_c$, but no scaling behavior is known for $p > p_c$. Based on 3D lattice studies, it is expected that D will cross over from $D = 2.5$ if $p \approx p_c$ to $D = 3$ if $p \gg p_c$ (Stauffer, 1985).

Figure 11 *b* shows that, in proteins, $\ln V \propto D \ln R$, with a fractal dimension $D = 2.47 \pm 0.04$ (a nonlinear fit to the model $V = aR^D$ gives a similar value for the exponent $D = 2.35$). This suggests that packing in proteins behaves like random spheres near their percolation threshold.

The surface areas A of proteins also scale with R with a fractal exponent $D = 2.26$, obtained from nonlinear fit of $A = aR^D$. Therefore, both V and A scale with R with a similar fractal dimensionality. This is consistent with the direct linear correlation we observed between V and A.

Another way to examine the random spheres model is through the size distribution n_V of voids and pockets of different volume V. Here we obtain n_V by normalizing the number of voids and pockets N_V of size V by the chain length N_{aa} of the protein: $n_V = N_V/N_{aa}$. In random-packed spheres and in lattice models where $p < p_c$, it is observed

TABLE 1 Summary of protein unfolding experimental data

Protein*	pdb† Name	Number of Residues	P_t	Temperature Range T_{low}	Temperature Range T_{high}	$\Delta^U_N C_p$‡ (kJ/mol°C)	Intercept kJ/mol)	Ref.§
Pepsinogen	2psg	346	0.675	51	66	25.4	−551.5	[1]
IL 1β	avg (4)	150	0.698	37	53	8	−73.9	[2]
Chymotrypsin	5cha	212	0.704	37	57.2	10.25	82	[3]
T4 lysozyme	3lzm	156	0.712	33	51	9.75	24.98	[4]
Papain	9pap	200	0.700	56	83.8	13.23	−228.2	[5]
Myoglobin	1mbo	147	0.716	50	70.8	11.31	−290.2	[3]
Rnase A	avg (4)	114	0.725	41	62.2	4.52	186.2	[3]
Lysozyme	1lzl	130	0.725	47.1	77.6	6.36	82.8	[3]
Ubiquitin	1ubq	73	0.733	57	90	3.33	8.29	[6]
Barnase	1rnb	106	0.734	22	54.3	5.54	192	[7]
Rnase T1	8rnt(K25)	93	0.733	44	61.2	5.18	154.1	[8]
Cytochrome C	5cyt	98	0.736	51.6	80	7.53	−154	[3]
Eglin c	avg (7)	41	0.743	41	85.7	3.05	34.16	[9]
Tendamistat	1hoe	70	0.753	68.3	81.6	2.89	38.23	[10]
CI 2	2ci2	62	0.761	41.4	70.9	3.31	46.52	[11]
SH3	1shg	52	0.762	34	66	3.25	−15.62	[12]
BPTI	5pti	58	0.767	86	104.5	1.97	112	[13]
Protein G	1pgx(B2)	66	0.778	58.4	79.4	2.67	15.24	[14]

For BPTI, we use 5pti. ROP and Met-J proteins are excluded because they are dimers. Their thermodynamics cannot easily be compared with that of the rest of the proteins. Water molecules are removed before calculation of packing densities. The linear regression parameters are directly cited from the original publication, are fitted from listed data, or are estimated from published figures.

*Exceptions are: eglin C, where the P_t is the average of 7 structures, separated from complexes with thermitase (1tec, 2tec, 3tec, residues 1–6 cleaved), with subtilisin (2sec (8–70), 1mee (7–70), 1cse (8–70)), and with alpha chymotrypsin (1acb (8–70)). The P_t for ribonuclease A is the average from 1rat, 1xps (two chains), and 3rn3. The structures for human interleukin 1β are 1ilb (res 3–153), 2ilb (1–153), 4ilb (3–153), and 5ilb (3–153). 2mib and 8ilb used in Makhatadze and Privalov (1995) are mouse interleukin 1β, and are therefore not included.

†The pdb structures for the P_t calculations are from Table 1 of Makhatadze and Privalov (1995).

‡Parameters of linear regression model ($\Delta^U_N H(T) = \Delta^U_N C_p \times T +$ Intercept) of calorimetric enthalpic changes of unfolding are listed.

§The sources of the experimental data are: [1] Mateo and Privalov (1981), [2] Makhatadze et al. (1994), [3] Privalov and Khechinashvili (1974), [4] Kitamura and Sturtevant (1989), [5] Tiktopulo and Privalov (1978), [6] Wintrode (1994), [7] Griko et al. (1994), [8] Yu et al. (1994), [9] Bae and Sturtevant (1995), [10] Renner et al. (1992), [11] Jackson and Fersht (1991), [12] Viguera et al. (1994), [13] Makhatadze et al. (1993), and [14] Alexander et al. (1992).

that $n_V \propto V^{-\theta} c_0^V$, where $\theta = 1.5$ in 3D space (Stauffer, 1985; Lorenz et al., 1993). Here we regard voids and pockets as clusters. Because we have so little data for any one protein, we calculate the average of n_V for all proteins longer than 200 amino acids from the data set used in Fig. 8. We also exclude all voids and pockets smaller than 50 Å^3, because the scaling law applies only to clusters of larger size. The voids and pockets are binned in sizes from 50 to 500 Å^3 at a step size of 10 Å^3, and the average number of voids and pockets of all proteins in each bin is calculated. Figure 11, *c* and *d*, shows that the normalized size distribution of voids and pockets follows the scaling law, $n_V \propto V^{-\theta} c_0^V$, where θ and c_0 are estimated by nonlinear curve fitting to be 1.67 ± 0.07 and 0.9966 ± 0.0008, respectively.

One simple alternative model, for comparison, would be perfect packing, as in a jigsaw puzzle, with identical small packets of free volume everywhere, as in a solid. This would give a delta function for n_V versus V, with a plateau at $n_V = 0$, as illustrated in Fig. 11 *c*. In lattice models, the exponent θ does not depends on the details of the lattice structures but only on the dimension d of the space: $\theta = 1$ for $d = 2$, $\theta = 1.5$ for $d = 3$, and $\theta = 1.9$ for $d = 4$. The estimated θ value of 1.67 falls between 1.5 and 1.9, the values for three- and four-dimensional systems. With the sampling errors we have, Fig. 11 *d* shows that protein void and pocket size distributions could fit models with dimensionality 3 or 4, but not 2.

What is the physical meaning of the parameters θ and c_0? For a mostly occupied lattice, if the probability for a lattice site to be unoccupied is p, and if p is small (as in proteins, where voids and pockets are only a small fraction of the total volume), the number n_s of clusters of unoccupied sites containing s sites is

$$n_s(p \to 0) = \sum g_{st} p^s (1 - p)^t.$$

Here s is the size (volume) of the cluster, and t is the peripheral (area) of the cluster. For a cluster of size s, we need to have s connected and unoccupied sites (with probability p^s), and t sites that are immediate neighbors of the cluster all occupied (with probability $(1 - p)^t$). Because a cluster of size s may have many different shapes, g_{st} represents the number of clusters of volume s and area t, and $g_s = \sum_t g_{st}$ represents the number of configurations of clusters of size s, namely, the number of ways s pieces can be put

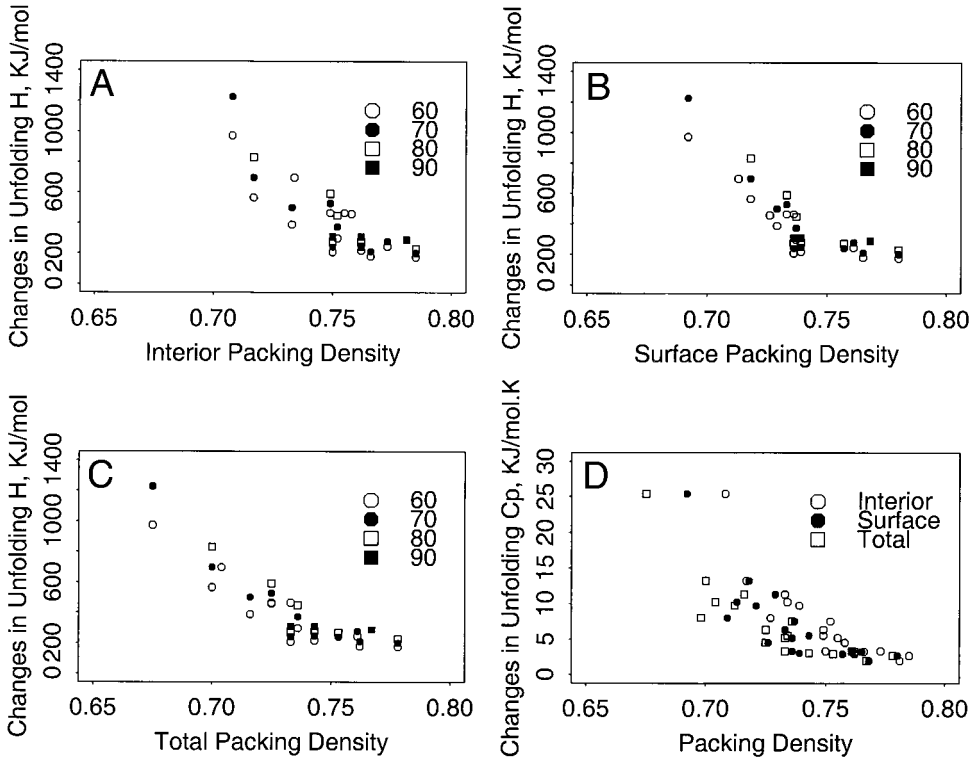

FIGURE 12 Protein packing density and protein unfolding thermodynamics. The thermodynamic parameters are from linear regressions of experimentally measured enthalpies (see Table 1). Correlations of enthalpy of unfolding $\Delta^{U}_{N}H$ (kJ/mol) at 60, 70, 80, and 90°C with (*a*) interior packing density (P_i), (*b*) surface packing density (P_s), (*c*) total packing density (P_t), and (*d*) $\Delta^{U}_{N}C_p$ values, as determined from the slope of the $\Delta^{U}_{N}H \sim T$ regressions, as plotted against P_i, P_s, and P_t.

together in a domino game. For convex polyominoes, it is known that (Delest and Viennot, 1983),

$$g_{st} = (t + 11) \cdot 4^{t/2} - 4 \cdot (t + 1)\binom{t}{t/2}$$

but there is no general analytical solution for either g_{st} or g_s, and enumeration is intractable if $n_s > 30$ (Redelmeier, 1981). However, it is known that g_s scales asymptotically as $g_s \propto s^{-\theta}\lambda^s$ (Stauffer, 1979). Here θ is a universal exponent, which depends only on the dimensionality of the space.

The parameter λ, however, depends on the details of the lattice. Let $a_\infty(p) = \lim_{s\to\infty}(t_s/s)$ be the average perimeter-to-size ratio of large clusters. It is a function of the filling fraction p. Then, it is known (Stauffer, 1979) that

$$\ln \lambda = \ln \frac{1}{p_c} + \int_0^{p_c} \frac{a_\infty(p)}{1 - p}\, dp,$$

where p_c is the threshold probability when a percolating cluster of unoccupied sites appears. In general, λ is a function of the perimeter-to-size ratio a (Stauffer, 1979). For small p, $(1 - p)^t \approx 1$, we have

$$n_s(p \to 0) \approx \sum g_{st} p^s \propto s^{-\theta}(\lambda p)^s = s^{-\theta} c_1^s,$$

where $c_1 = \lambda p$. Therefore, c_1 is a property of the surface-to-volume ratio of the voids and the pockets.

We conclude that the scaling behavior of the volume, area, and size distribution of voids and pockets, suggests that proteins are packed more like random spheres than like jig-saw puzzles.

Limitations of these models

Is there a best model for packing inside a protein? It may be that neither solids nor liquids is the appropriate model for proteins. A heterogeneous environment, such as a protein core, is different from a homogeneous and regular environment. For example, here is a third criterion, according to which the interior of a protein resembles neither a solid nor a liquid. This criterion is the connectivity of free space as seen by a 1.4-Å sphere. In glycogen phosphorylase, a 1.4-Å probe detects many voids, and they have a wide size range (0–1000 Å^3), with the majority of voids between 10 and 100 Å^3 (Fig. 9 *c*). But the same probe sphere sees the model solid, as being completely inaccessible, because the voids are too small. And this probe sees the model liquid, as a single pocket with no voids, because all the free space is interconnected and fully accessible.

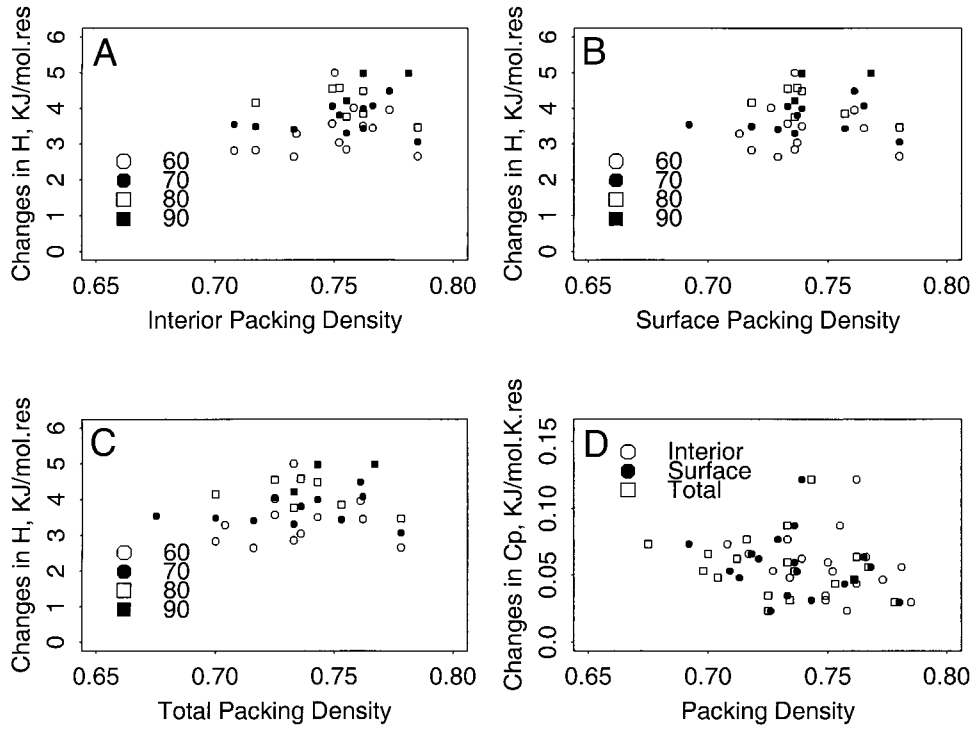

FIGURE 13 Correlations of unfolding enthalpy per monomer $\Delta_N^U H/N$ (kJ/mol) at 60, 70, 80, and 90°C with (*a*) interior packing density (P_i), (*b*) surface packing density (P_s), (*c*) total packing density (P_t), and (*d*) heat capacity changes per monomer $\Delta_N^U C_p/N$ plotted against packing densities P_i, P_s, and P_t. The proteins are the same as in Fig. 12.

Of course, the resemblance of the protein packing to the model systems can be made closer by rescaling the sizes of particles. If the model solid had much smaller spheres, a 1.4-Å probe could enter it. And if the model liquid had higher density, a 1.4-Å probe would see some connected voids. In this regard, the liquid may be the better overall model of protein cores, because choosing the right particle sizes and liquid density could probably cause the liquid model to reproduce the protein-void distribution function, whereas there is no rescaling that could cause the model solid to have the correct shape of the distribution function of free volumes.

However, there might be better models altogether. Maybe polymer liquids are better models. Or perhaps mixtures of spheres of different sizes. In summary, crystalline solids of spheres is a model for packing inside proteins only in the respect that it reproduces the mean packing density. Better models should also account for the substantial numbers and volumes of packing defects that we find in proteins.

Packing and protein unfolding thermodynamics

In this section, we compare the native packing density to the unfolding thermodynamics, $\Delta_N^U C_p$ and $\Delta_N^U H$, for 18 small proteins reported by (Makhatadze and Privalov, 1995). In a recent study, a correlation between packing efficiency and protein stability has been noted for the A–B helix regions of α-lactalbumin and better packed lysozyme (Demarest et al., 2001). It is also observed that lysozyme has more buried polar surfaces (Demarest et al., 2001). It is not clear whether either or both observations generalize for other proteins. In this study, we focus on packing efficiency as reflected by the values of packing densities. The other effect of packing, namely, the volume overlaps between nonbonded atoms (polar–polar, polar–nonpolar, and nonpolar–nonpolar interactions), which reflects more directly the burial effects of polar and nonpolar surfaces, is not the focus of this study. Table 1 lists the proteins, pdb names, numbers of residues, P_t values, and temperature ranges (T_{low} and T_{high}) of the experiments. Because different temperature ranges were reported in different experiments, we use such linear regressions so $\Delta_N^U C_p$ can be interpolated at standard temperature points of 60, 70, 80, and 90°C. We extrapolate no more than 5°C from either T_{low} or T_{high}. Our aim is to compare results across different proteins at these standard temperatures. Figure 12, *a–c*, shows that unfolding enthalpy $\Delta_N^U H$ becomes smaller as packing density increases. The unfolding heat capacity changes $\Delta_N^U C_p$ also decreases with increasing packing densities P_t (Fig. 12 *d*).

These results indicate that larger proteins, which are more loosely packed than smaller proteins, also have more favorable total folding enthalpies. Because protein size is corre-

lated with looseness of packing, it raises the question of what is the most appropriate independent variable: compactness or chain length. Figure 13 shows that, when unfolding enthalpies and unfolding heat capacities are divided by chain length, $\Delta_N^U H$ and $\Delta_N^U C_p$ are independent of the packing densities inside proteins. Because the enthalpies per monomer are independent of the volume per monomer, it indicates that the dominant thermodynamic quantities for folding are not the van der Waals interactions. This view is consistent with an earlier modeling effort (Dill and Bromberg, 1994), suggesting that average protein-core packing is more like nuts and bolts in a jar than like a jigsaw puzzle, in the sense that packing has not been evolutionarily optimized and designed. Internal packing may be a consequence of folding, rather than a cause of it.

We thank Drs. Fred Richards and Pat Fleming for providing the data sets of the model solid and liquid, and Dr. John Finney for making the data of computer simulated model liquid available. J.L. thanks Drs. Ralf Bundschuh, Charles DeLisi, Herbert Edelsbrunner, Simon Kasif, Renhao Li, and Clare Woodward for helpful discussions.

J.L. thanks the National Science Foundation for support through grants MCB9980088 and DBI0078270, the American Chemical Society-Petroleum Research Fund through grant 35616-G7, and K.A.D. thanks the National Institutes of Health for support through grant GM34993.

REFERENCES

Adler, J., Y. Meir, A. Aharony, and A. Harris. 1990. Series study of percolation moments in general dimension. *Physic. Rev. B.* 41: 9183–9206.

Akkiraju, N., and H. Edelsbrunner. 1996. Triangulating the surface of a molecule. *Discrete Appl. Math.* 71:5–22.

Alexander, P., S. Fahnestock, T. Lee, J. Orban, and P. Bryan. 1992. Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains B1 and B2: why small proteins tend to have high denaturation temperatures. *Biochemistry.* 31:3597–3603.

Axe, D., N. Foster, and A. Fersht. 1996. Active barnase variants with completely random hydrophobic cores. *Proc. Natl. Acad. Sci. U.S.A.* 93:5590–5594.

Bae, S.-J., and J. Sturtevant. 1995. Thermodynamics of the thermal unfolding of eglin c in the presence and absence of guanidinium chloride. *Biophys. Chem.* 55:247–252.

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

Chothia, C. 1984. Principles that determine the structure of proteins. *Ann. Rev. Biochem.* 53:537–572.

Connolly, M. 1983. Analytical molecular surface calculation. *J. Appl. Cryst.* 16:548–558.

Delaney, J. 1992. Finding and filling protein cavities using cellular logic operations. *J. Mol. Graph.* 10:174–177.

Delest, M., and G. Viennot. 1983. Algebraic languages and polyominoes enumeration. *In* Automata, Languages and Programming. Proc. 10th Colloquium, Barcelona, Spain, July 18–22. J. Diaz, editor. Vol. 154 of Lecture Notes in Computer Science. Springer-Verlag, Berlin.

Demarest, S., S.-Q. Zhou, J. Robblee, R. Fairman, B. Chu, and D. Raleigh. 2001. A comparative study of peptide models of the α-domain of α-lactalbumin, lysozyme, and α-lactablbumin/lysozyme chimeras allows the elucidation of critical factors that contribute to the ability to form stable partially folded states. *Biochemistry.* 40:2138–2147.

Dill, K., and S. Bromberg. 1994. Side-chain entropy and packing in proteins. *Protein Sci.* 3:997–1009.

Edelsbrunner, H. 1987. Algorithms in Combinatorial Geometry. Springer-Verlag, Berlin.

Edelsbrunner, H. 1995. The union of balls and its dual shape. *Discrete Comput. Geom.* 13:415–440.

Edelsbrunner, H., M. Facello, P. Fu, and J. Liang. 1995. Measuring proteins and voids in proteins. *In* Proc. 28th Ann. Hawaii Int'l Conf. System Sciences. Vol. 5, IEEE Computer Society Press, Los Alamitos, CA. 256–264.

Edelsbrunner, H., M. Facello, and J. Liang. 1998. On the definition and the construction of pockets in macromolecules. *Discrete Appl. Math.* 88: 83–102.

Edelsbrunner, H., and E. Mücke. 1994. Three-dimensional alpha shapes. *ACM Trans. Graphics.* 13:43–72.

Edelsbrunner, H., and N. Shah. 1996. Incremental topological flipping works for regular triangulations. *Algorithmica.* 15:223–241.

Edelsbrunner, H., M. Facello, and J. Liang. 1998. On the definition and the construction of pockets in macromolecules. *Discrete Appl. Math.* 88: 18–29.

Facello, M. 1995. Implementation of a randomized algorithm for Delaunay and regular triangulations in three dimensions. *Comp. Aided Geom. Design.* 12:349–370.

Finney, J. 1970. Random packings and the structures of simple liquids. I. The geometry of random close packing. *Proc. Roy. Soc. Lond. A.* 319:479–493.

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

Gavish, B., E. Gratton, and C. Hardy. 1983. Adiabatic compressibility of globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* 80:750–754.

Gellatly, B., and J. Finney. 1982. Calculation of protein volume: an alternative to the Voronoi procedure. *J. Mol. Biol.* 161:305–322.

Gerstein, M., and C. Chothia. 1996. Packing at the protein–water interface. *Proc. Natl. Acad. Sci. U.S.A.* 93:10167–10172.

Gerstein, M., and F. Richards. 1999. Protein Geometry: Distances, Areas, and Volumes, Vol. F, chap. 22. International Union of Crystallography.

Gerstein, M., J. Tsai, and M. Levitt. 1995. The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra. *J. Mol. Biol.* 249:955–966.

Gibson, K., and H. Scheraga. 1987. Exact calculation of the volume and surface area of fused hard sphere molecules with unequal atomic radii. *Mol. Phys.* 62:1247–1265.

Griko, Y., G. Makhatadze, and P. Privalov. 1994. Thermodynamics of barnase unfolding. *Protein Sci.* 3:669–676.

Guibas, L., D. Knuth, and M. Sharir. 1992. Randomized incremental construction of Delaunay and Voronoi diagrams. *Algorithmica.* 7:381–413.

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

Ho, C., and G. Marshall. 1990. Cavity search: an algorithm for the isolation and display cavity-like binding regions. *J. Comput. Aided Mol. Des.* 4:337–354.

Hobohm, U., and C. Sander. 1994. Enlarged representative set of protein structures. *Protein Sci.* 3:522–524.

Honig, B. 1999. Protein folding: from the Levinthal paradox to structure prediction. *J. Mol. Biol.* 293:283–293.

Hubbard, S. J., and P. Argos. 1994. Cavities and packing at protein interfaces. *Protein Sci.* 3:2194–2206.

Hubbard, S., K. Gross, and P. Argos. 1994. Intramolecular cavities in globular-proteins. *Protein Eng.* 7:613–626.

Jackson, S., and A. Fersht. 1991. Folding of chymotrypsin inhibitor 2. 1. evidence for a two-state transition. *Biochemistry.* 30:10428–10435.

Joe, B. 1991. Construction of three-dimensional Delaunay triangulation using local transformations. *Comp. Aided Geom. Design.* 8:123–142.

Jorgensen, W., and J. Tirado-Rives. 1988. The OPLS potential function for proteins: energy minimizations for crystals of cyclic peptides and crambin. *J. Amer. Chem. Soc.* 110:1657–1666.

Kim, S., J. Liang, and B. Barry. 1997. Chemical complementation identifies a proton acceptor for redox-active tyrosin D in photosystem II. *Proc. Natl. Acad. Sci. U.S.A.* 94:14406–14411.

Kitamura, S., and J. Sturtevant. 1989. A scanning calorimetric study of the thermal denaturation of the lysozyme of phate T4 and the Arg 96→His mutant form thereof. *Biochemistry.* 28:3788–3792.

Kleywegt, G., and T. Jones. 1994. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Cryst. D.* 50: 178–185.

Lawson, C. 1977. Software for C^1 surface interpolation. *In* Mathematical Software III, Academic Press, New York. 161–194.

Lee, B., and F. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.

Liang, J., H. Edelsbrunner, P. Fu, P. Sudhakar, and S. Subramaniam. 1998a. Analytical shape computing of macromolecules. I. Molecular area and volume through alpha-shape. *Proteins.* 33:1–17.

Liang, J., H. Edelsbrunner, P. Fu, P. Sudhakar, and S. Subramaniam. 1998b. Analytical shape computing of macromolecules. II. Identification and computation of inaccessible cavities inside proteins. *Proteins.* 33: 18–29.

Liang, J., H. Edelsbrunner, and C. Woodward. 1998c. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci.* 7:1884–1897.

Liang, J., and M. McGee. 1998. Mechanisms of coagulation factor Xa inhibition by antithrombin. Correlation between hydration structure and water transfer during reactive loop insertion. *Biophys. J.* 75:573–582.

Liang, J., and S. Subramaniam. 1997. Computation of molecular electrostatics with boundary element methods. *Biophys. J.* 73:1830–1841.

Lim, W., and R. Sauer. 1989. Alternative packing arrangements in the hydrophobic core of λ repressor. *Nature.* 339:31–36.

Lorenz, B., I. Orgzall, and H.-O. Heuer. 1993. Universality and cluster structures in continuum models of percolation with two different radius distributions. *J. Phys. A. Math. Gen.* 26:4711–4722.

Makhatadze, G., G. Clore, A. Gronenborn, and P. Privalov. 1994. Thermodynamics of unfolding of the all β-sheet protein interleukin-1β. *Biochemistry.* 33:9327–9332.

Makhatadze, G., K.-S. Kim, C. Woodward, and P. Privalov. 1993. Thermodynamics of BPTI folding. *Protein Sci.* 2:2028–2036.

Makhatadze, G., and P. Privalov. 1995. Energetics of protein structure. *Adv. Protein Chem.* 47:307–425.

Mateo, P., and P. Privalov. 1981. Pepsinogen denaturation is not a two-state transition. *FEBS Lett.* 123:189–192.

McGee, M., H. Teuschler, and J. Liang. 1998. Effective electrostatic charge of coagulation factor X in solution and on phospholipid membranes: implications for activation mechanisms and structure–function relationships of the Gla domain. *Biochem. J.* 330: 533–539.

Meester, R., R. Roy, and A. Sarkar. 1994. Nonuniversality and continuity of the critical covered volume fraction in continuum persolation. *J. Stat. Phys.* 75:123–134.

Murphy, K., and S. Gill. 1991. Solid model compounds and the thermodynamics of protein unfolding. *J. Mol. Biol.* 222:699–709.

O'Rourke, J. 1994. Computational Geometry in C. Cambridge University Press, New York.

Pascual-Ahuir, J., and E. Silla. 1990. GEPOL: an improved description of molecular surfaces. I. Building the spherical surface set. *J. Comput. Chem.* 11:1047–1060.

Perrot, G., B. Cheng, K. Gilson, K. Palmer, A. Nayeem, B. Maigret, and H. Scheraga. 1992. MSEED: a program for the rapid analytical determination of accessible surface areas and their derivatives. *J. Comp. Chem.* 13:1–11.

Peters, K., J. Fauck, and C. Frömmel. 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. *J. Comput. Chem.* 15:507–523.

Preparata, F., and M. Shamos. 1985. Computational Geometry: An Introduction. Springer-Verlag, New York.

Privalov, P., and N. Khechinashvili. 1974. A thermodynamic approach to the problem of stabilization of globular protein structure. *J. Mol. Biol.* 86:665–684.

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

Redelmeier, D. 1981. Counting polyominoes: yet another attack. *Discrete Math.* 36:191–203.

Renner, M., H. Hans-Jurgen, M. Scharf, and J. Engels. 1992. Thermodynamics of unfolding of the α-amylase inhibitor tendamistat: correlations between accessible surface area and heat capacity. *J. Mol. Biol.* 223: 769–779.

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

Richards, F. 1977. Areas, volumes, packing, and protein structures. *Ann. Rev. Biophys. Bioeng.* 6:151–176.

Richards, F., and W. Lim. 1994. An analysis of packing in the protein folding problem. *Q. Rev. Biophys.* 26:423–498.

Richmond, T. 1984. Solvent accessible surface area and excluded volume in proteins: analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.* 178:63–89.

Sastry, S., D. Corti, P. Debenedetti, and F. Stillinger. 1997. Statistical geometry of particle packing. I. Algorithm for exact determination of connectivity, volume, and surface areas of void space in monodisperse and polydisperse sphere packings. *Phys. Rev. E.* 56:5524–5532.

Shortle, D., W. Stites, and A. Meeker. 1990. Contributions of the large hydrophobic amino acids to the stability of staphyloccocal nuclease. *Biochemistry.* 29:8033–8041.

Shrake, A., and J. Rupley. 1973. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* 79:351–371.

Stauffer, D. 1979. Scaling theory of percolation clusters. *Physics Rep.* 54:1–74.

Stauffer, D. 1985. Introduction to Percolation Theory. Taylor & Francis, London.

Tiktopulo, E., and P. Privalov. 1978. Papain denaturation is not a two-state transition. *FEBS Lett.* 91:57–58.

Tsai, J., R. Taylor, C. Chothia, and M. Gerstein. 1999. The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* 290:253–266.

van der Marck, S. 1996. Network approach to void percolation in a pack of unequal spheres. *Phys. Rev. Lett.* 77:1785–1788.

Viguera, A., J. Martinez, V. Filimonov, P. Mateo, and L. Serrano. 1994. Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. *Biochemistry.* 33:2142–2150.

Voorintholt, R., M. Kosters, G. Vegter, G. Vriend, and W. Hol. 1989. A very fast program for visualizing protein surfaces, channels and cavities. *J. Mol. Graphics.* 7:243–245.

Wintrode, P. 1994. Thermodynamics of uniquitin unfolding. *Proteins.* 18:246–253.

Word, J., S. Lovell, T. LaBean, H. Taylor, M. Zalis, B. Presley, J. Richardson, and D. Richardson. 1999. Visualizing and quantitating molecular goodness-of-fit: small-probe contact dots with explicit hydrogens. *J. Mol. Biol.* 285:1711–1733.

Yu, Y., G. Makhatadze, C. Pace, and P. Privalov. 1994. Energetics of ribonuclease T1 structure. *Biochemistry.* 33:3312–3319.