Protein Science (1998), 7:897–905. Cambridge University Press. Printed in the USA.

Heterogeneity of packing: Structural approach

NATALYA KUROCHKINA AND GEORGE PRIVALOV

Applied Thermodynamics, 222 Schilling Circle, Suite 130, Hunt Valley, Maryland 21031-0157

(RECEIVED July 31, 1997; ACCEPTED December 30, 1997)

Abstract

Analysis of the heterogeneity of packing in proteins showed that different groups of the protein preferentially contribute to low- or high-density regions. Statistical distribution reveals the two preferable values for packing density in the form of two peaks. One peak occurs in the range of densities 0.55–0.65, the other occurs in the range 0.75–0.8. The high-density peak is originated primarily by high packing inside the hydrogen bonded backbone and to some extent by side chains. Polar/charged and apolar side chains both contribute to the low-density peak. The average packing density values of individual atomic groups significantly vary for backbone atoms as well as for side chain atoms. The carbonyl oxygen atoms of protein backbone and the end groups of side chains show lower packing density than the rest of the protein. The side-chain atomic groups of a secondary structure element when packed against the neighboring secondary structure element form stronger contacts with the side chains of this element than with its backbone. Analysis of the low-density regions around each buried peptide group was done for the set of proteins with different types of packing, including α–α, α–β, and β–β packing. It was shown that cavities are regularly situated in the groove of secondary structure element packed against neighboring elements for all types of packing. Low density in the regions surrounding the peptide groups and the end groups of side chains can be explained by their positioning next to a cavity formed upon the association of secondary structure elements. The model proposed can be applied to the analysis of protein internal motions, mechanisms of cellular signal transduction, diffusion through protein matrix, and other events.

Keywords: cellular signal transduction; diffusion; internal motions; packing density of proteins; protein architecture; ridges into grooves packing model

Molecular packing is an important characteristic of protein structure. Several models describing the protein interior were suggested. The model proposed by Kauzmann (1959) suggests that the interior of a protein is similar to an organic liquid "oil-drop." While studying the packing density of lysozyme and ribonuclease, Richards (1974) determined the average value of packing density to be close to 0.75. That makes the protein interior more similar to an organic crystal. For the buried residues of nine proteins analyzed by Chothia (1975), the volume occupied by a residue in the protein interior is constant for all the residues of the same kind; it corresponds to the volume calculated from the crystal structure. This confirmed that the most accurate description of a protein is as "molecular crystal." In the model suggested by Klapper (1971), it is assumed that molecules may be described as those containing "hard" core and "soft" interacting surfaces, and that the protein interior is more close to solid than liquid.

Although on average protein interior is highly packed, there are variations in the packing efficiency. For ribonuclease, it varies in the range from 0.66 to 0.84 (Richards, 1974). Regions with less efficient packing were found close to the active site groove surrounded by more densely packed areas. Considerable variability was found for five globular proteins with predominant beta-structure. Analysis of the local densities orthogonal to beta-sheet amide planes predominated by van der Waals interactions demonstrated that the regions with high density occur mostly at the ends of beta-structure strands (Beardsley & Kauzmann, 1996). The regions of high packing density were found to correlate with the location of folding intermediates (Privalov, 1995). The packing efficiency of atoms in locally concave, grooved regions of protein surfaces is generally looser than that around atoms in locally convex, ridge regions. It demonstrates surface curvature-dependent hydration (Gerstein & Chothia, 1996).

Packing of different atomic groups also varies. Standard deviation for most of the groups was found to be 10–15% of their mean values (Finney, 1975) or even 20–25% (Richards, 1974). Analysis of the statistical distribution of atomic groups with low and high density made by Privalov (1995) shows that there are two preferable values for packing density—one around 0.55–0.65 and the other close to 0.75.

Packing deficiency is directly related to movement properties of protein molecules. Experimental data provides the evidence that protein molecule is able to "breathe," and that the motions of atoms play very important role in their function (McCammon & Harvey, 1987). Analysis of the diffraction patterns from insulin

Reprint requests to: Natalya Kurochkina, Applied Thermodynamics, 222 Schilling Circle, Suite 130, Hunt Valley, Maryland 21031-0157; e-mail: Kurochkina@aol.com.

crystals (Caspar et al., 1988), lysozyme (Faure et al., 1994), and tropomyosin alpha-helical coiled coil shows very diffuse liquid-like diffraction. The movements of atoms within the molecules with RMS amplitudes of 0.4–0.45 Å are coupled over a range of ~6 Å as in a liquid. These locally coupled movements account for the most of the disorder (Boylan & Phillips, 1986).

If the regions with lower density are required in protein interior to make motions needed for the functioning of the molecule, then the heterogeneity of packing is a necessary feature that is characteristic for each protein.

Here we analyze what is the structural basis for the regions with high and low packing density, i.e., how these regions are related to the architecture of the protein molecule. The fact that the packing efficiency around individual atomic groups varies may result from some common properties. It is known that approximately 80% of a protein is organized in the elements of a secondary structure—hydrogen-bonded backbone of the molecule. α-Helices and β-sheets are attached to each other by the protruding side chains interpenetrating into each other. The fact that predominantly hydrophobic regions are packed with lower density than the rest of the protein (Kuntz, 1972) may indicate that the regions of the secondary structure that are packed with better efficiency are connected to each other by lower density, softer surfaces of interpenetrating side chains, and this is common for each protein. To summarize, we tried to address the question of whether the heterogeneity of packing is the general feature, which characterizes the association of secondary structure elements in all proteins.

It was shown that (1) different groups of protein preferentially contribute to low- or high-density regions. Statistical distribution reveals the two preferable values for packing density in the form of two peaks. One occurs in the range of densities 0.55–0.65, the other occurs in the range 0.75–0.8. The high-density peak is originated primarily by high packing inside the hydrogen-bonded backbone and to some extent by side chains. Polar-charged and apolar side chains both contribute to the low-density peak (Privalov, 1995). (2) The average packing density values of individual atomic groups significantly vary for backbone atoms as well as for side-chain atoms. The carbonyl oxygen atoms of protein backbone and the end groups of side chains show lower packing density than the rest of the protein. (3) Cavities are regularly situated in the groove of secondary structure element packed against neighboring elements for all types of packing, including α–α, α–β, and β–β packing. (4) Low density in the regions surrounding the peptide groups and the end groups of side chains can be explained by their positioning next to a cavity formed upon the association of secondary structure elements. Side chains are packed in the groove of a neighboring secondary structure element in such a way that they do not reach its backbone and leave a cavity around the peptide group.

Results

We analyzed how different atomic groups of a protein are packed, and whether they preferentially contribute to low- or high-density regions.

Statistical distribution of packing

The statistical distribution of the packing density function for three proteins (hen egg white lysozyme, barnase, and cytochrome *c*) was analyzed. The statistical distribution is the function of frequencies of occurrence of certain density values or histograms of packing density. Packing density values and their frequency of occurrence were calculated (see Statistical distribution of packing in Materials and methods). The histogram for all the analyzed proteins has two peaks. One peak occurs in the range of densities between 0.55 and 0.65. The other, higher peak, occurs in the range 0.75–0.8.

To determine what particular groups in proteins are contributing to the given ranges of the packing density values, the frequencies were recalculated within the volumes of the particular structures separately. The structures being analyzed include apolar side chains, polar side chains, and backbone atoms. We did not take into account surface residues because it is difficult to analyze the arrangement of water molecules surrounding the protein, thus making the estimation of packing arbitrary. Only buried atomic groups were considered. Figure 1 shows the frequencies of occurrence of packing density values for atoms belonging to different categories of side chains as well as for backbone atoms only. For all these categories the lower density peak is relatively smaller. Shown on the same figure are distributions of density for different categories of atoms, which do give some insight on the origin of two peak distribution.

It appeared that the right peak is originated primarily by high packing inside the hydrogen bonded backbone and to some extent by side chains (polar and apolar). The left peak is dominated by relatively loose side chains and, to some extent, by segments of backbone that are loose.

It follows from the above that it is the packing density that predominantly falls in two categories—high and low packing. These categories can be further split into subcategories.

The heterogeneity of packing was analyzed at the level of individual amino acid residue. Packing density was estimated for different atomic groups of hydrophobic residues that are fully buried in the protein interior and form the hydrophobic core. The procedure is described in Packing density of atomic groups in Materials and methods. The average packing density of each atomic group is shown in Table 1. As we can see, both main-chain and side-chain atoms show heterogeneity. For main-chain atoms there is tendency of the carbonyl oxygen to be in low-density regions. Its packing density is always lower than the average packing density of main

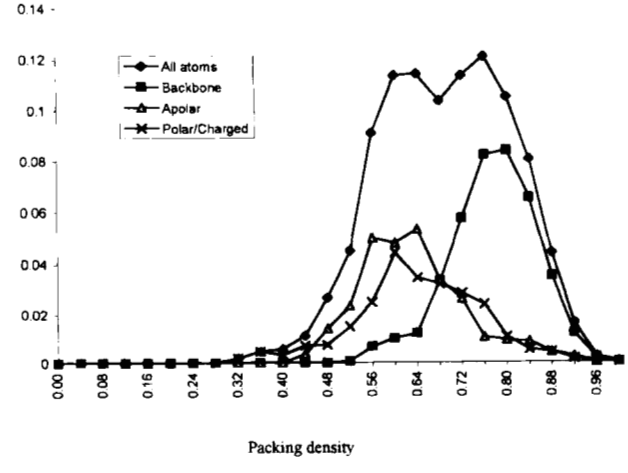

Fig. 1. Distribution of packing density values for extended set of atoms, containing buried atoms from HEW lysozyme, barnase, and cytochrome *c*. Total distribution (all atoms) is broken into different types of atoms: backbone, apolar side chains, and polar/charged side chains.

Table 1. *The average packing density of atomic groups in the protein interior for buried amino acid residues*[a]

	ALA	VAL	ILE	LEU	PHE	TYR	TRP
No.	21	25	18	41	26	17	8
N	0.73	0.74	0.7	0.75	0.75	0.77	0.74
Cα	0.67	0.84	0.84	0.81	0.8	0.87	0.69
C	0.78	0.92	0.87	0.86	0.87	0.9	0.82
O	0.5	0.52	0.39	0.47	0.49	0.5	0.49
Cβ	0.52	0.72	0.81	0.68	0.64	0.68	0.7
Cγ		0.54	0.72	0.73	0.87	0.91	0.87
		0.58	0.58				
Cδ			0.56	0.52	0.66	0.7	0.59
				0.54	0.64	0.69	0.83
Xε					0.57	0.68	0.54
					0.58	0.64	0.84
							0.58
Xζ					0.56	0.81	0.53
							0.54
Xη						0.47	0.5
m.ch.	0.67	0.76	0.7	0.72	0.73	0.76	0.69
s.ch.	0.52	0.61	0.67	0.62	0.65	0.7	0.65
Total	0.64	0.69	0.68	0.67	0.68	0.72	0.66

[a]Residues were taken from the proteins D-xylose isomerase 9xia, lysozyme 3lzm, rubredoxin 5rxn, myohemerythrin 2mhr, FAB fragment of antibody 4fab, apoferritin 1hrs, Bence-Jones protein 2rhe, carboxypeptidase A 5cpa, alpha-lactalbumin 1alc, phospholipase A 1bp2, trypsin inhibitor 5pti, alpha-chymotrypsin 6cha, and granulocyte colony-stimulating factor 1bgc.

chain. In aliphatic side chains, $-CH_3$ groups (C_β of alanine, C_γ atoms of valine, $C_{\gamma 2}$ and $C_{\delta 1}$ of isoleucine, and C_δ atoms of leucine) have density less than average. Aromatic residues have lower than average density for –CH– groups.

Contacts

To analyze whether atomic groups with high and low density differ in types of interactions with surrounding groups, the distances were calculated for each buried side-chain atomic group of all secondary structure elements (see Contacts of atomic groups in Materials and methods). Distance D_m is the minimum distance between the given group and the main-chain atoms of neighboring secondary structure elements. Distance D_s is the minimum distance between the given group and the side-chain atoms of neighboring secondary structure elements (Fig. 2). These distances are shown in Figure 3 for 100 atomic groups. Distance D_s, which characterizes contacts of the side chain–side chain type fills the range 3.2–5.2 Å. This corresponds approximately to the sum of van der Waals radii of the interacting atomic groups. It indicates that van der Waals contacts between the side chains of secondary structure elements play important role in stability of their interfaces. Distance D_m is in the range 3.6–7.6 Å. This shows that van der Waals contacts between the side chain of one secondary structure element and the main chain of another secondary structure element are weaker compared to the side chain–side chain contacts. The value $(D_m - D_s)$ estimated for each atomic group is from -0.2 to 3 Å. We have noticed that distance D_m is almost always more than distance D_s for the given atomic group. The side chains of the approaching element of secondary structure do not reach the backbone of the considered element. They interact mainly with the protruding side chains of this element on some distance from its backbone.

Search for cavities in low-density regions

Analysis was done by Beardsley and Kauzmann (1996) for the buried peptide groups of several proteins with beta-structure dominantly. It revealed that there is a tendency for the lowest density regions to concentrate in zones that are on a distance of 2–4 Å from the peptide group plane in the direction perpendicular to the plane, and residing in the radius of 1–2 Å around the C–N bond. This region of the peptide group is close to the carbonyl oxygen and the amide nitrogen, which are hydrogen bonded to the neighboring strand. The hydrogen bonded atoms lie in the groove where the side chains from a neighboring element of a secondary structure are packed. Because it is in average common for all buried peptide groups, a sequence of these groups should form a channel in the groove. Our analysis of packing density for the atomic groups have shown that the carbonyl oxygens, $-CH_3$ groups of aliphatic side chains, and –CH– groups of aromatic side chains have lower than average packing density values (Table 1).

To answer the question of whether the groove formed by two hydrogen bonded beta-strands or two residues in an α-helix has empty space when the side chains of the neighboring strand or helix are packed against it, we analyzed this region on the presence of cavities. Each buried peptide group was surrounded by a square filled by points. The search of cavities was done by an algorithm described in Algorithm for the finding of cavities in the protein

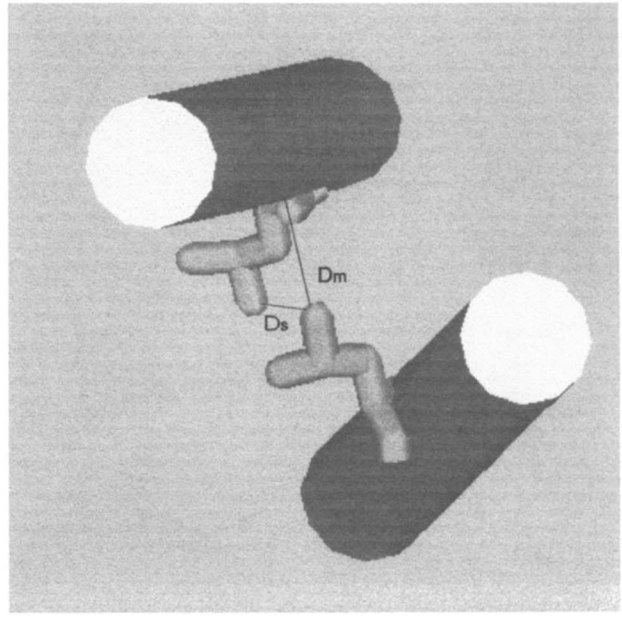

Fig. 2. Distances D_m and D_s for a pair of secondary structure elements. Secondary structure element are shown by cylinders; side chains are in thick stick representation. D_m is the minimum distance between the side-chain atomic group of one secondary structure element and the main chain of neighboring secondary structure element. D_s is the minimum distance between the side-chain atomic group of one secondary structure element and the side chain of neighboring secondary structure element.

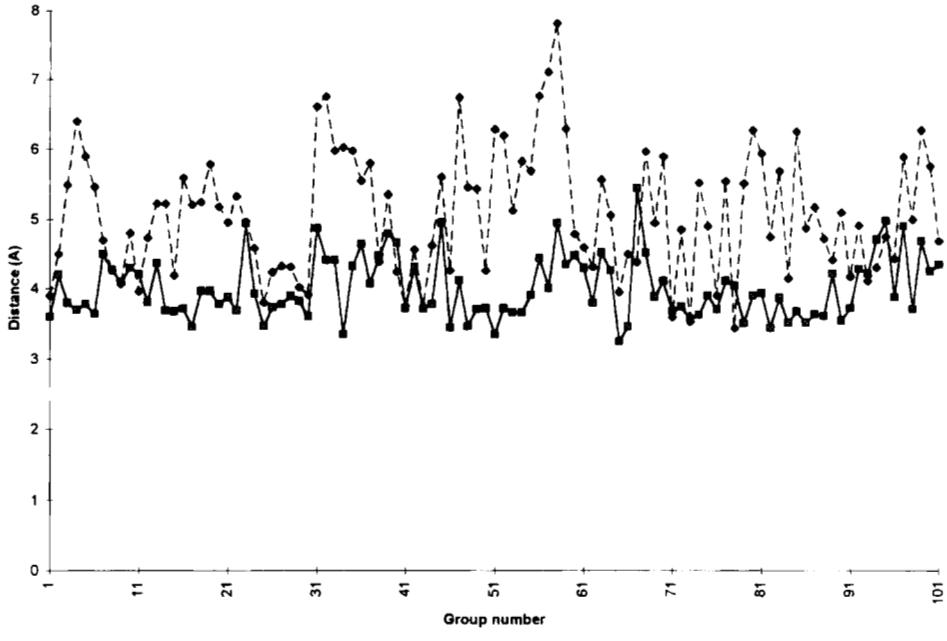

Fig. 3. D_m (broken line) and D_s (solid line) for the side-chain atomic groups of secondary structure elements in 14 proteins. Protein set is given in Packing density of atomic groups in Materials and methods.

interior in Materials and methods. This procedure was applied to proteins that have α–α, α–β, and β–β types of packing. Proteins selected were myoglobin (α–α) 1mbn, carboxypeptidase A (α–β, α–α) 5cpa, α-chymotrypsin (β–β, α–β) 6cha, lysozyme (α–α, α–β) 7lyz, and ribonuclease (α–α, α–β) 1rns. Thirty-eight out of 58 buried peptide groups of myoglobin contained series of interconnected spherical cavities with a radius of more than 1 Å. Among the rest of the 20 groups, 15 groups had cavities with a radius of 0.8–0.9 Å. Similar result was observed for other proteins.

In Figure 4a cavities around peptide groups are shown for the E helix of myoglobin. The molecular surface of E helix is shown in blue using hard sphere representation. The cavities in the grooves $i \div i + 4$ and $i \div i + 3$ are shown in red. The helices surrounding the E helix are shown in magenta; they are given in ribbon representation to reveal the interior of packing. Helix E interacts also with heme; the cavities are formed on the interface of E-helix with heme as well. Figure 4b gives an example of cavities on the α-β interfaces for the ribonuclease S. Helix 2 is shown by its molecular surface, strand 1 is packed against it. In Figure 4C, α-chymotrypsin β–β packing is shown for α subunit.

Superposition of cavities found by Lee & Richards algorithm on packing diagrams

To analyze whether cavities in a groove are formed as a result of helix-helix packing, we built the diagrams for the packing of alpha-helices in a myoglobin molecule on the basis of the diagrams for the hemoglobin molecule (Chothia et al., 1981). Sequence alignment of myoglobin with hemoglobin was used. The location of cavities, which was found for myoglobin by Lee and Richards (1971), was analyzed in combination with the packing diagrams. Superposition of these data is shown in Figure 5. Most of the amino acid residues that form helix–helix interfaces participate in the cavity formation. The residues that are found on the borders of the cavities are shown in double circles.

Discussion

Heterogeneity of packing occurs in all protein structures analyzed. Different atomic groups of the protein preferentially contribute to low- or high-density regions. We analyzed the heterogeneity of packing from the point of view of protein architecture. Statistical distribution of the packing density values reveal the two preferable values for packing density in the form of two peaks. One occurs in the range of densities 0.55–0.65, the other occurs in the range 0.75–0.8. When backbone and side-chain atoms are analyzed separately, one can see that the high-density peak is originated primarily by high packing inside the hydrogen-bonded backbone and only to some extent by side chains. Polar-charged and apolar side chains both contribute to the low-density peak.

Packed side chains, even though they constitute only a small fraction of the volume, may play an important role in stabilizing the tertiary structure. The low-packing category primarily contains the side chains and, to some extent, backbone atoms. Therefore, the protein structure can be represented as a rigid frame that consists of α-helices spot welded in different places by complementary side chains. This frame is filled and surrounded by flexible and loose padding. This arrangement perhaps provides proteins for the necessary balance between stability and flexibility needed for their functioning.

The average packing density values of individual atomic groups significantly vary for backbone atoms as well as for side-chain atoms. Analysis of the average packing density for different atomic groups shows that the oxygen atoms of backbone, $-CH_3$ groups of aliphatic side chains, and –CH– groups of aromatic side chains have lower density than the rest of the protein. We can suggest that these facts may be related to each other. Structural analysis of the regions with low density shows that the latter usually occur in the grooves of secondary structure elements. These grooves contain hydrogen-bonded oxygen and amide nitrogen atoms. The grooves are usually filled by the side chains of neighboring secondary

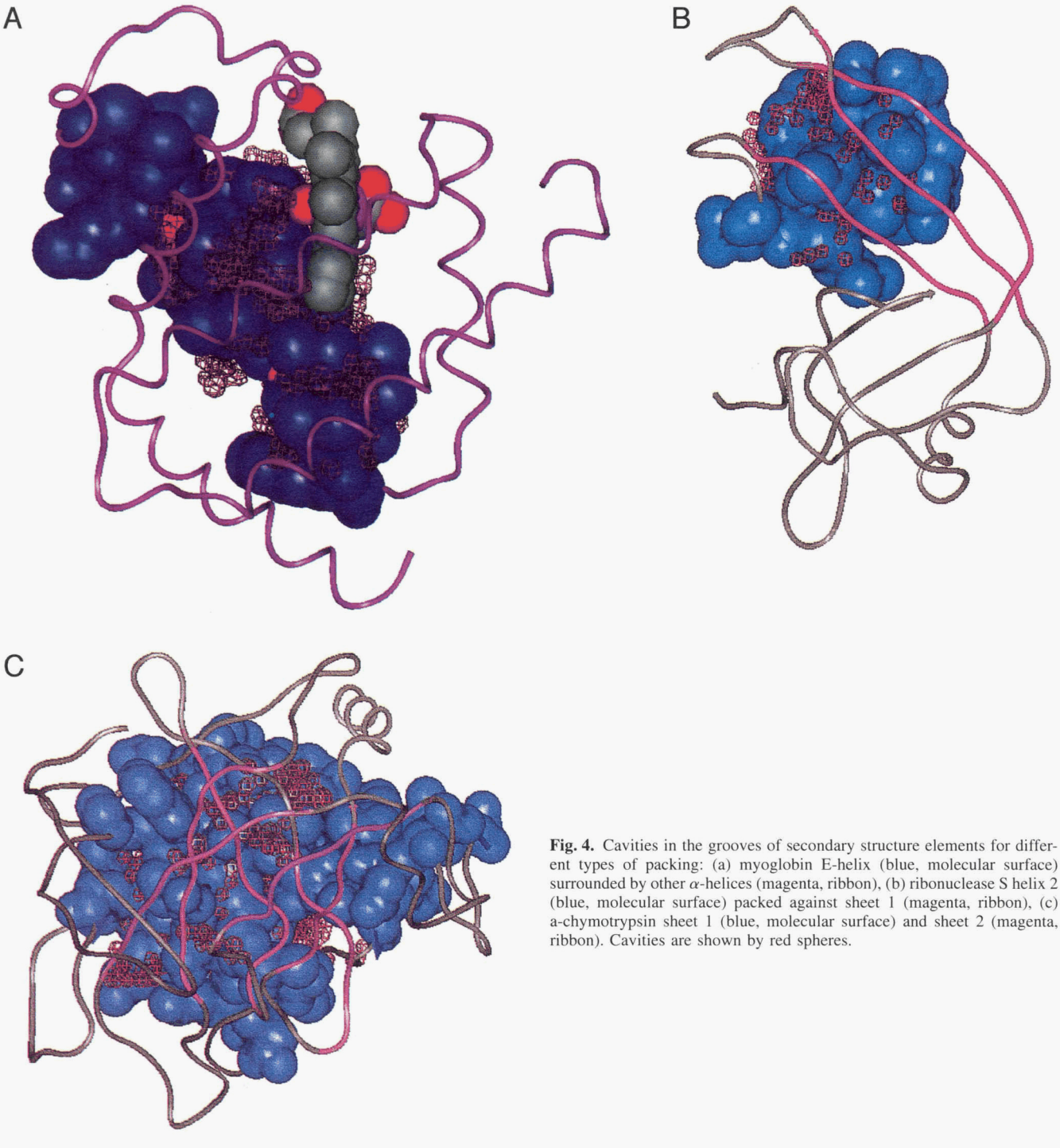

Fig. 4. Cavities in the grooves of secondary structure elements for different types of packing: (a) myoglobin E-helix (blue, molecular surface) surrounded by other α-helices (magenta, ribbon), (b) ribonuclease S helix 2 (blue, molecular surface) packed against sheet 1 (magenta, ribbon), (c) a-chymotrypsin sheet 1 (blue, molecular surface) and sheet 2 (magenta, ribbon). Cavities are shown by red spheres.

structure elements. The low-density regions found around the peptide plane by Beardsley and Kauzmann, low density at the ends of side chains, large fluctuations in average density for atomic groups noticed by Finney (1975) and Richards (1974), two peaks distribution found by Privalov (1995), and our analysis confirm the model we suggest.

The region with lowest packing density is located at the distance of approximately 4 Å from the peptide group plane. It is distinct by very limited long-range contacts of the backbone with the side chains of neighboring secondary structure elements. This confirms our observation that the hydrogen-bonded backbone and the side chains of approaching secondary structure elements do not pack tightly with each other. On the contrary, they tend to be separated by the buffer zone of lower density or "empty space."

This regular pattern of heterogeneous packing is a common feature for all proteins with different function because all of them consist of packed secondary structure elements. We included in our analysis proteins diverse in their function and three-dimensional

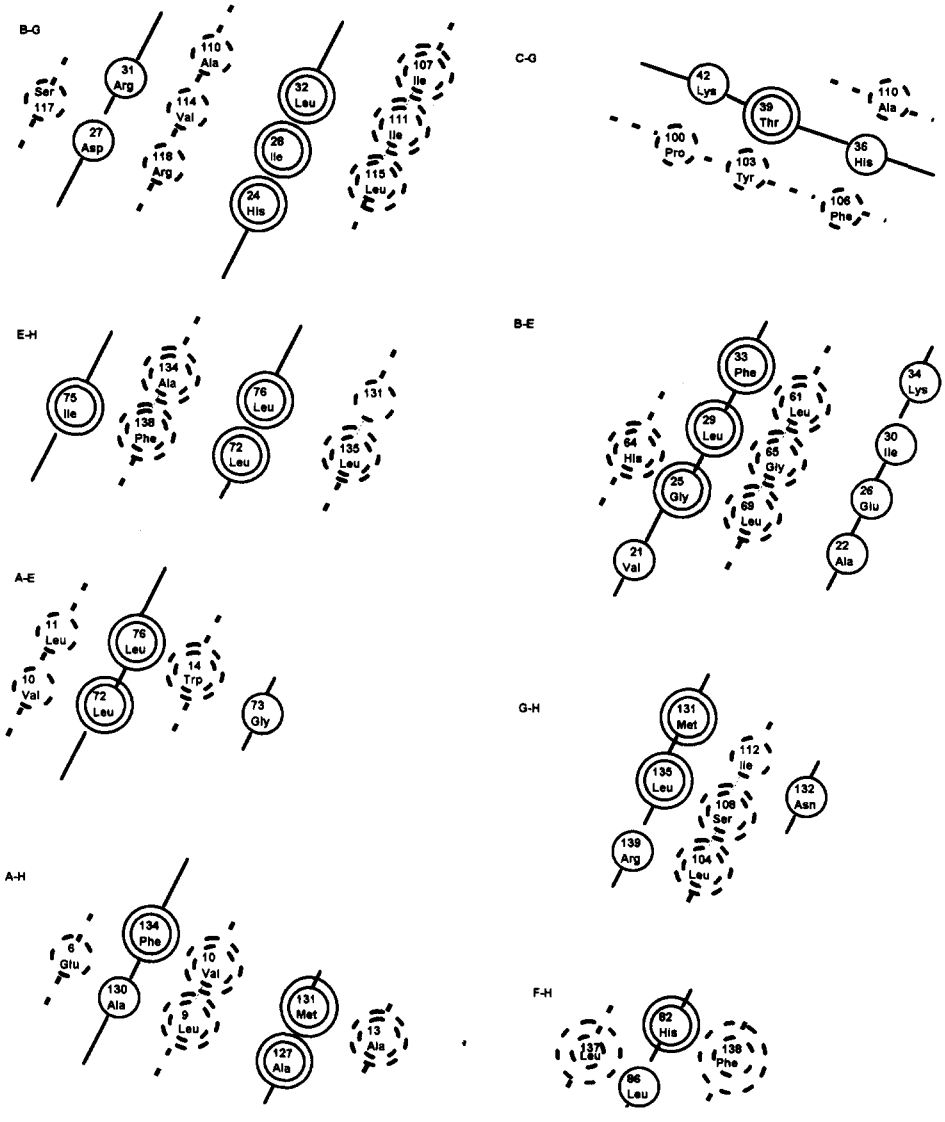

Fig. 5. Helix–helix interfaces for myoglobin. In a pair of α-helices, the amino acid residues of one helix are shown as a solid line, and those of the contacting helix by a dotted line. α-Helices are given by a letter A–H. Double circles are for those amino acid residues that participate in the cavity formation found by Lee and Richards (1971).

structure—cytokines, antibodies, enzymes, oxygen carriers with α–α, α–β, and β–β types of packing. The pattern repeats for all of them.

The secondary structure is an ordered backbone of molecule stabilized by hydrogen bonds. Protein backbone can be organized into several types of hydrogen bonding pattern (α-, π-, 3_{10}-helix, polyproline helix, β-sheet). It is expected that some properties of the molecules also show regular patterns. We think as a result of our analysis that pattern of cavities in a molecule is regular; it is the result of the packing of secondary structure elements against each other. If it is regular, it should be the result of perfect "non-packing," rather than imperfect packing. It is known that proteins move while performing their function. The presence of "empty spots" in the protein interior makes these movements possible.

Our model is supported by experimental data on diffusion through the protein matrix. For example, buried tryptophan in ribonuclease T1 is being quenched by acrylamide (Eftink & Ghiron, 1975). This indicates that O_2 is able to penetrate into the protein interior, and the holes appear in the protein interior due to nanosecond fluctuations. Channels formed in the grooves of the elements of secondary structure could be the routes leading from the solvent inside the protein.

Myoglobin molecule binds and releases oxygen to perform its physiological function. The way oxygen penetrates from the solvent into the protein interior to reach the heme iron is not clear. It was suggested (Perutz & Matthews, 1966) that histidine E7, or possibly one of the other side chains blocking the entrance to the ligand site, might act as a gate that has to swing out to open the access to ligand. Mutant proteins can be produced with altered ligand binding parameters. Mutation in the position E7 His $\rightarrow$ Gly causes dramatic increase in the rates of ligand binding. Val in position E11 if mutated to Ala, Val, or Ile change the rate of the ligand association; the rate decreases with the increase in the size of side chain (Springer et al., 1988). Both of these side chains are

situated on the E helix, which interacts with A, B, H helices, and heme itself. As we can see, packing against the other helices and heme leaves cavities in the grooves of the helix. According to our model, oxygen could move along these channels from the exposed beginning of the helix–helix interface inside the molecule. Mutations to smaller side chains at the sites of the helix distant from the heme but located along the channel leading to heme probably make the entrance to the channel more open. They also can modify the width of the channel. It may become easier for the ligand to penetrate inside as a result of these changes.

This model also can be applied to the analysis of mechanisms of electron transfer through the membrane protein. We know from the three-dimensional structure of photosynthetic reaction center from *Rhodopseudomonas viridis*, which converts the energy of sunlight into electrical and chemical energy, that no connections were found between pigments involved in the electron transfer process (Deisenhofer & Michel, 1989). They are separated by relatively large distances and fixed in nonpolar regions of the protein matrix. Low density regions around the hydrogen-bonded backbone surrounded by nonpolar side chains could be carriers of the electron flow.

Cellular signal transduction is regulated often via ligand binding to the extracellular portion of a receptor molecule that causes structural changes on the cytosolic portion of the receptor. Because the elements of secondary structure cross the membrane and are packed against each other all the way from the extracellular space inside the cell, interconnected channels situated along the grooves are crossing the membrane. These channels are specific because the side chains of different amino acids in different conformation face the interior of the channel, therefore forming the basis for selectivity. Ligand binding can regulate the width of the channel and close or open the entrance into the channel that is located at the beginning of the secondary structure interface outside the membrane.

Pore-forming peptides as alamethicin produce voltage-gated ion channels in membranes and exhibit multiple conductance levels (Eisenberg et al., 1973). The conductance substates can be explained by a change in the channel cross section due to a fluctuation in the number of monomer units forming oligomer (Boheim, 1974). This is consistent with our model of channel formation between the two elements of secondary structure packed against each other. Crystallographic structure for alamethicin peptide revealed that it is largely an α-helix (Butters, 1981).

Therefore, pattern of regular cavities situated inside the protein shown on the basis of the analysis of low-density regions and their relation to the protein architecture, we think, is a general feature of all protein molecules. A model of the channels formed as a result of the packing of secondary structure elements against each other is confirmed by experimental data. This model can be used in the analysis of protein internal motions, diffusion through protein matrix, mechanisms of electron transfer, cellular signal transduction, and other events.

Materials and methods

Statistical distribution of packing density values

The relative frequency of the occurrence of packing density values was calculated for hen egg white lysozyme, barnase, and cytochrome *c*. Atoms were divided into the following categories: backbone, polar and apolar side chains, and all atoms. Atomic groups of buried amino acids were considered. Amino acid residue was considered as buried if its solvent accessible area was less than 5% of the total area. Packing density was estimated with porcupine algorithm based on molecular volume partitioning using the Voronoi polyhedra (Privalov, 1995) described below in Porcupine algorithm.

Packing density of atomic groups

Packing density of individual atomic groups of buried amino acids was calculated for the set of proteins using a porcupine algorithm. Amino acid residue was considered buried if its solvent accessible area was less than 5% of the total area.

Proteins we chose for the analysis are diverse in their function and three-dimensional structure. To show that heterogeneous packing is the general feature of protein structure, we included in our set antibodies, electron carriers, cytokines, and other proteins. Some proteins in this set are enzymes; those that do not have enzymatic activity were also included. These proteins represent different fold types (α, β, and α/β). The set includes D-xylose isomerase 9xia, lysozyme 3lzm, rubredoxin 5rxn, myohemerythrin 2mhr, FAB fragment of antibody 4fab, apoferritin 1hrs, Bence-Jones protein 2rhe, carboxypeptidase A 5cpa, alpha-lactalbumin 1alc, phospholipase A 1bp2, trypsin inhibitor 5pti, alpha-chymotrypsin 6cha, granulocyte colony-stimulating factor 1bgc, and cytochrome *c* 5cyt.

Porcupine algorithm

This is a numerical method that we call radical dissection. The quantitative measure of the complementarity of groups packed in the interior of protein is the packing density. Macroscopically, packing density is defined as the ratio between the van der Waals envelope of the molecule and the total volume that the molecule occupies in context of surrounding molecules (occupancy volume). However, at the level of individual atoms the definition of the volume that any given atom occupies becomes less obvious (Richards, 1974). The most widely used definition of occupancy volume (Chothia, 1975; Finney, 1975; Richards, 1977, 1979, 1985; Janin & Chothia, 1990; Gerstein & Lynden-Bell, 1993; Gerstein & Chothia, 1996) is based on a geometric composition known as the Voronoi Diagram (Voronoi, 1908). The same approach has also been used in the analysis of pure water simulations (Shih et al., 1994).

Given an arbitrary set of points in space, the classical Voronoi diagram defines the areas of proximity to each point from that set. An example of such a composition is given in Figure 6. In the case of the atoms within a protein, the area of proximity for any given atom can be constructed by connecting this atom with every other atom (that is reasonably close) then bisecting these interatomic vectors by planes normal to connecting vectors. The smallest polyhedra resulting from the intersection of these planes around this atom will define the proximity area or Voronoi polyhedron that determines the occupancy volume. One can now define the packing density on the atomic level as a ratio of the volume of the van der Waals envelope for this atom and the volume of the Voronoi polyhedra. For covalently bonded atoms the volume of the van der Waals envelope is determined by the fraction of the van der Waals sphere that is cut by the radical planes.

Compared to classic Voronoi dissection, our algorithm uses radical planes. Given two neighboring atoms A and B (Fig. 7) the radical plane is constructed at the distance d from the center of the atom.

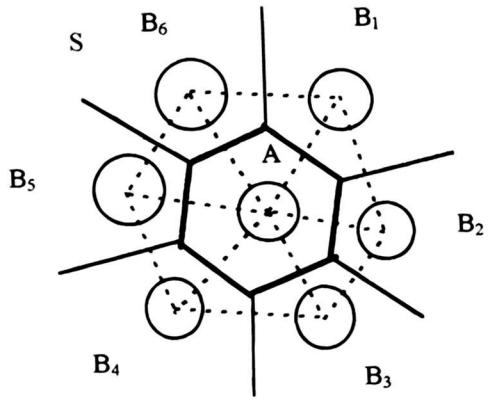

Fig. 6. An example of the Voronoi diagram in two dimensions around the point produced by the set of neighboring points $B_1, B_2, \ldots, B_6$ for the set of points S shown as circles. Bold lines outline the areas of proximity. Any point within this area is closer to the point from the set that is within this area than to any other point from the set S. To construct a Voronoi diagram one should triangulate the set of points (connect each point to its nearest neighbors). Bisection of the triangulation lines yields the classical Voronoi diagram.

The occupancy volume of atom A is determined as smallest polyhedron each facet of which is determined by radical planes constructed between atom A and its heaviest neighbors.

Contacts of atomic groups

To analyze how atomic groups with high and low density differ in the types of interactions with surrounding groups, two distances were calculated for each buried side-chain atomic group of all secondary structure elements. Amino acid residues were chosen from the set of proteins described in the Packing density of atomic groups section. Totally, 100 atomic groups were selected from the amino acid residues if the residue was found in a secondary structure element, and it was buried (its solvent-accessible area was less than 5% of the total area).

The distances were calculated for each buried side-chain atomic group of all secondary structure elements. Distance D_m is the minimum distance between the given group and the main-chain atoms of neighboring secondary structure elements. Distance D_s is the minimum distance between the given group and the side-chain atoms of neighboring secondary structure elements (Fig. 2).

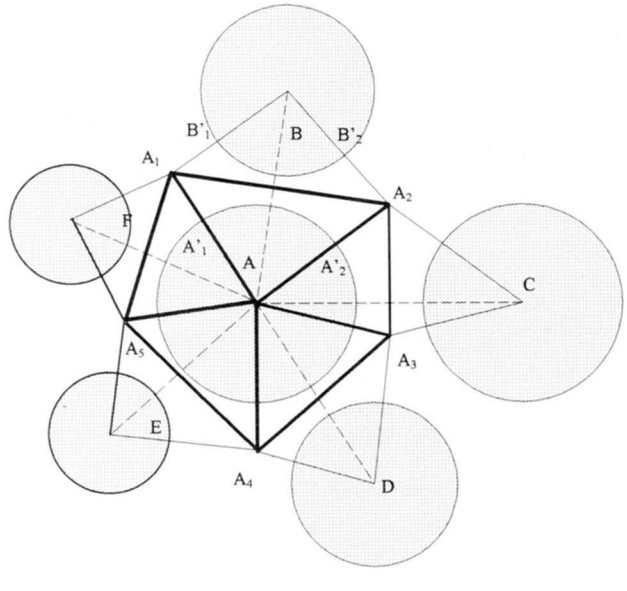

Fig. 7. Two-dimensional illustration of mutual occupancy of the atom A with each of its neighbors B, C, D, E, and F. The occupancy volume for atom A is determined by polyhedron $A_1A_2A_3A_4A_5$. This polyhedron consists of five pyramids shown here as triangles (A,A_1,A_2), $(A,A_2,A_3,)$, (A,A_3,A_4), (A,A_4,A_5), and (A,A_5,A_1), which determine mutual occupancy with each neighbor. Mutual packing of two atoms A and B can be calculated as the ratio of volumes defined by the two sectors of the spheres (A,A_1',A_2') and (B,B_1',B_2') over the volume of polyhedron (A,A_1,B,A_2).

Algorithm for the finding of cavities in the protein interior

To analyze low-density regions, buried peptide groups from the proteins with different types of packing (α–α, α–β, and β–β packing) were considered. The orientation of each peptide group was chosen similarly to the method proposed by Beardsley and Kauzmann (1996). It is shown in Figure 8. Cavities were searched around each peptide group by the following algorithm. The peptide bond was surrounded by a set of points in a cube with the dimensions $4 \times 4 \times 6$ Å along the axis x, y, and z correspondingly. Each point was considered as filled if it was found inside the van der Waals envelope of the molecule (radii were chosen the same as in Chothia, 1975). Other points were considered as empty. For each empty point the contacts with surrounding atomic groups were calculated. If minimum contact distance (d_m) exceeded 0.5 Å, the cavity was defined with the center in this point and with the radius equal to d_m. All the cavities were listed. The next cavity was added to the list if it was overlapping with another cavity not less than half of its radius. Cavities were selected if they were at the distance of 2–4 Å from the peptide group plane and within the radius of 1–2 Å around the C–N bond.

Packing diagrams for ridges into grooves packing model

To analyze whether cavities in a groove are formed as a result of helix–helix packing, the superposition of data was made for the hemoglobin ridges into grooves packing suggested by Chothia et al. (1981) and the cavity volume calculations made for myoglobin by Lee and Richards (1971). The sequence of myoglobin was aligned with the sequence of hemoglobin. The packing diagrams

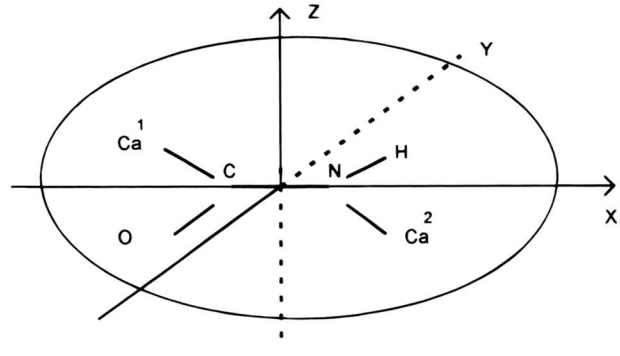

Fig. 8. The orientation of a peptide group in XY plane with the middle of the C–N bond as an origin.

were built for myoglobin similarly to those for hemoglobin. The amino acid residues of myoglobin located on the borders of the cavities (Lee & Richards, 1971) were shown on the packing diagrams.

Algorithms and computer graphics

All algorithms for protein structure analysis (packing density, solvent-accessible area, packing clusters, cavity search) and computer graphics were used from MOLE program (Applied Thermodynamics, Hunt Valley, MD) and performed on Gateway 2000/90 computer.

Acknowledgments

We would like to thank Prof. Peter L. Privalov of the Department of Biology, John Hopkins University for valuable discussion of this paper.

References

Beardsley DS, Kauzmann W. 1996. Local densities orthogonal to beta-sheet amide planes: Patterns of packing in globular proteins. *Proc Natl Acad Sci USA 93*:4448–4453.

Boheim G. 1974. Statistical analysis of alamethicin channels in black lipid membranes. *J Membr Biol 19*:277–303.

Boylan D, Phillips GN. 1986. Motions of tropomyosin. *Biophys J 49*:76–78.

Butters T. 1981. *Angew Chem 93*:904–905.

Caspar DL, Clarage J, Salunke DM, Clarage M. 1988. Liquid-like movements in crystalline insulin. *Nature 332*:659–662.

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

Chothia C, Levitt M, Richardson D. 1981. Helix to helix packing in proteins. *J Mol Biol 145*:215–250.

Deisenhofer J, Michel H. 1989. The photosynthetic reaction centre from the purple bacterium *Rhodopseudomonas viridis*. *EMBO J 8*:2149–2170.

Eftink MR, Ghiron CA. 1975. Dynamics of a protein matrix revealed by fluorescence quenching. *Proc Natl Acad Sci USA 72*:3290–3294.

Eisenberg M, Hall JE, Mead CA. 1973. The nature of the voltage-dependent conductance induced by alamethicin in black lipid membranes. *J Membr Biol 14*:143–147.

Faure P, Micu A, Perahia D, Smith JC, Benoit JP. 1994. Correlated intramolecular motions and diffuse X-ray scattering in lysozyme. *Nat Struct Biol 1*:124–128.

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

Fisher W, Koch EZ. 1979. Geometrical packing analysis of molecular compounds *Z Kristallogr 150*:245–260.

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

Gerstein M, Lynden-Bell RM. 1993. What is the natural boundary of the protein in solution? *J Mol Biol 230*:641–650.

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

Karplus M, McCammon JA. 1981. The internal dynamics of globular proteins. *CRC Crit Rev Biochem 9*:293–349.

Kauzmann W. 1959. Some factors in the interpretation of protein denaturation. *Adv Protein Chem 14*:1–64.

Klapper MH. 1971. On the nature of the protein interior. *BBA 229*:557–566.

Kuntz ID. 1972. Tertiary structure in carboxypeptidase. *J Am Chem Soc 94*:8568–8572.

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

McCammon JA, Harvey SC. 1987. *Dynamics of proteins and nucleic acids*. Cambridge, Massachusetts: Cambridge University Press.

Perutz MF, Matthews FS. 1966. An X-ray study of azide methaemoglobin. *J Mol Biol 21*:199–202.

Privalov G. 1995. Packing of protein interior: Structure and distribution of packing density [Dissertation]. The John Hopkins University.

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

Richards FM. 1977. Areas, volumes, packing, and protein structure. *Annu Rev Biophys Bioeng 6*:151–176.

Richards FM. 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 FM. 1985. Calculation of molecular volumes and areas for structures of known geometry. *Methods Enzymol 115*:440–464.

Shih JP, Sheu SY, Mou CY. 1994. A Voronoi polyhedra analysis of structures of liquid water. *J Chem Phys 100*:2202–2212.

Springer BA, Edeberg KD, Sligar SG, Rohlfs RJ, Matthews A, Olson JS. 1988. Site-directed mutagenesis of sperm whale myoglobin: Role of His E7 and Val E11 in ligand binding. Program and Abstracts of Symposium on Oxygen Binding Heme Proteins, Azilomar.

Voronoi GF. 1908. Nouveles applications des parametres continus à la theorie des formes quadratiques. *J Reine Angew Math 134*:198–297.