JMB

Protein Packing: Dependence on Protein Size, Secondary Structure and Amino Acid Composition

Patrick J. Fleming* and Frederic M. Richards

Department of Molecular Biophysics and Biochemistry Yale University, New Haven CT 06520-8114, USA

We have used the occluded surface algorithm to estimate the packing of both buried and exposed amino acid residues in protein structures. This method works equally well for buried residues and solvent-exposed residues in contrast to the commonly used Voronoi method that works directly only on buried residues. The atomic packing of individual globular proteins may vary significantly from the average packing of a large data set of globular proteins. Here, we demonstrate that these variations in protein packing are due to a complex combination of protein size, secondary structure composition and amino acid composition. Differences in protein packing are conserved in protein families of similar structure despite significant sequence differences. This conclusion indicates that quality assessments of packing in protein structures should include a consideration of various parameters including the packing of known homologous proteins. Also, modeling of protein structures based on homologous templates should take into account the packing of the template protein structure.

© 2000 Academic Press

Keywords: occluded surface; protein packing; homologous proteins; Voronoi volume; atom density

*Corresponding author

Introduction

Atomic packing has been recognized as an important metric for characterizing protein structures since it was observed in 1974 that the average packing density for the interior of proteins is approximately the same as that for crystals of small organic molecules (Richards, 1974). This observation was extended to the general statement that the volume occupied by a particular residue in protein interiors is the same as that occupied by the residue in its crystalline form (Chothia, 1975). In these studies the volume occupied by an amino acid residue was calculated using the Voronoi procedure (Finney, 1970) and packing density was defined as the ratio of van der Waals volume to Voronoi volume. Such residue Voronoi volumes have been shown to be a sensitive measure of the quality of a protein structure (Pontius *et al.*, 1996).

Variation in the packing within and between protein structural families has been indirectly studied by means of Voronoi volume analysis (Lesk & Chothia, 1980; Ptitsyn & Volkenstein, 1986; Gerstein *et al.*, 1994; Tsai *et al.*, 1999). The questions addressed in these studies concerned conservation of overall buried residue core volume and vari-

ations of residue volumes at individual sites. The results of these studies suggested that variations of buried residue core volumes within each family were due to normal statistical variation. In addition, it was found that the variation of residue volumes at specific buried residue sites was limited by ''local size constraints imposed by the protein structure'' at these sites (Gerstein *et al.* 1994). It was also suggested that the packing density is uniform in different buried regions of a protein interior when crystallographic water is included in the calculation (Tsai *et al.*, 1999). However, the question of different average packing densities between protein structural families was not specifically addressed in these previous studies.

Other methods, related to atomic packing but involving the frequency and quality of atom pair contacts, have also been devised for the evaluation of protein structures (Tanaka & Scheraga, 1976; Manavalan & Ponnuswamy, 1977; Warme & Morgan, 1978; Narayana & Argos, 1984; Gregoret & Cohen, 1990; Maiorov & Crippen, 1992; Singh & Thornton, 1992; Colovos & Yeates; 1993; Vriend & Sander, 1993; Bauer & Beyer, 1994; Miyazawa & Jernigan, 1996; Abagyan & Totrov; 1997; Melo & Feytmans, 1997). A recent study of the packing of 100 structures in the PDB, using a small-probe contact dot method, showed an inverse correlation between efficient packing and

E-mail address of the corresponding author: fleming@csb.yale.edu

crystallographic resolution (Word *et al.* 1999). In the highest resolution structures, when hydrogen atoms were properly added, the interdigitation and atomic fitting were highly compact. In addition to evaluation of structural models, packing analyses have proven useful in the prediction of allowed sequences for specific structures (Ponder & Richards, 1987) and the design of novel proteins (Lee & Subbiah, 1991; Dahiyat & Mayo, 1997; Kono *et al.*, 1998).

An alternative to the above methods for calculation of protein packing is the occluded surface algorithm (Pattabiraman *et al.*, 1995; DeDecker *et al.*, 1996). The occluded surface method calculates two initial parameters; the molecular surface area of each residue occluded by neighboring atoms, and the distance from every occluded molecular surface point to the occluding atom surface. Figure 1 graphically compares the occluded surface method (Figure 1(b)), the small- probe contact dot method (Figure 1(c)) and the Voronoi method (Figure 1(d)) for estimating the packing environment of the CB atom of an alanine residue in a protein structure.

In the occluded surface method (Figure 1(b)) a dot surface is calculated. Each molecular surface dot which is occluded by a neighboring atom has a distance normal associated with it and each dot represents a defined patch of surface area (circles). The distance normal extends to the van der Waals surface of the occluding atom (only selected occluding atoms are shown in the Figure). Surface dots with normals longer than 2.8 Å are considered non-occluded and are not shown in Figure 1. One obtains information on both the overall extent of packing interactions (total occluded surface area) and also a measure of the complementarity of packing (distribution of extended normals). In the small-probe contact method (Figure 1(c)) dots are placed on the molecular surface at each point that is 0.25 Å or less from a neighboring atom surface. Distance normals may also be calculated from these dots to the contacting atom surface in a manner similar to the occluded surface method. As seen in the example in Figure 1(b) the occluded surface method provides packing information on all atoms defining the packing environment
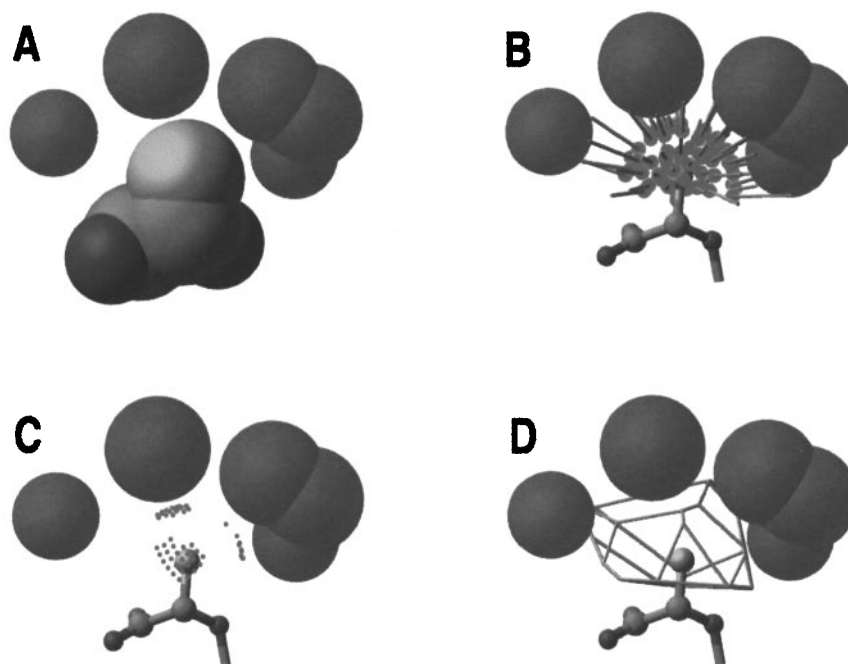

Figure 1. Methods to evaluate packing. Three different methods to evaluate the packing of the CB atom of Ala20 from ribonuclease A (7rsa) are shown. (a) the Ala20 residue is shown as intersecting van der Waals spheres and in (b), (c) and (d) as ball and stick figures. (a), (b), (c) and (d) Five selected neighboring atoms are also shown as van der Waals spheres; for clarity not all neighboring atoms are shown. (b) Occluded surface method. Surface normals are extended from the molecular surface of the CB atom to the intersection of neighboring van der Waals surfaces. Only those extended normals that intersect another van der Waals surface within a surface to surface distance of 2.8 Å are counted. The CB atom surface area associated with each extended normal is shown as a small circle. The occluded surface (*OS*) of the CB atom is the sum of the surface areas associated with intersecting normals and the occluded surface packing value (*OSP*) is a function of both this occluded surface area and the distribution of extended normal lengths as described in Methods. (c) The small probe contact dot method. A dot is placed on the molecular surface of the CB atom wherever this surface is within 0.25 Å of a neighboring van der Waals surface. Packing is estimated as a function of the gap distance between each dot and the neighboring atom's surface (Word *et al.*, 1999). (d) The Voronoi method. A plane is drawn at the midpoint between the CB van der Waals surface and all neighboring van der Waals surfaces. The intersections of these planes define the Voronoi volume occupied by the CB atom. The packing density is defined as the ratio of van der Waals volume to this Voronoi volume (Richards, 1974). The Figure was made with the program RIBBONS (Carson, 1987).

around an atom rather than, as in Figure 1(c), just the contacting atoms as in the small-probe contact dot method. The Voronoi method (Figure 1(d)) constructs a plane between the atom in question and each neighboring atom; the intersections of these planes define a convex polyhedron around the specific atom. The smallest polyhedron so constructed defines the Voronoi volume of the atom in question. The Voronoi method has a limitation in that it is most useful only for buried residues. This procedure requires an atom to be surrounded by neighboring atoms but on the protein surface these neighboring atoms frequently belong to water molecules. The exact location of surface water molecules is not known in most cases.

An important aspect of the occluded surface method is that, since packing is estimated only for the residue surface which is occluded by other atoms, the method works equally well for both buried residues and surface residues. Although the packing values of a limited number of X-ray crystal structures have been analyzed and compared using the occluded surface method (DeDecker *et al.*, 1996; Ratnaparkhi *et al.*, 1998) no extensive comparison of a large data set has been published.

Here, we addressed the question of whether all soluble globular proteins have the same overall atomic packing or have identifiable packing differences. We report the results of an analysis of the atomic packing in a data set of 152 non-homologous, high-resolution protein structures determined by X-ray crystallography and in four different families of proteins. We find that the packing of proteins is dependent on the size of the protein, the content of secondary structure elements and the amino acid composition. Differences in protein packing are conserved in protein families of similar structure.

Results and Discussion

Distribution of occluded surface packing values for a data set of protein structures

To determine the distributions of protein packing parameters, we used the occluded surface algorithm to estimate the atomic packing in protein structures. The occluded surface method takes into account two parameters: the surface area of each atom occluded by other atoms and the distance from the occluded surface to the van der Waals surface of occluding atoms (see Figure 1(b)). An occluded surface packing value, *OSP*, for each residue is calculated from these two parameters as described in Methods. Figure 2 shows the *OSP* for each residue in the ribonuclease A structure, 7rsa, as an example. The buried residues (defined as residues with zero solvent-accessible surface area) have uniformly large *OSP* values (filled circles; average, 0.557) compared to the solvent-exposed residues (open diamonds; average, 0.331). The values of *OSP* for the solvent-exposed residues show more variation than the buried residues,
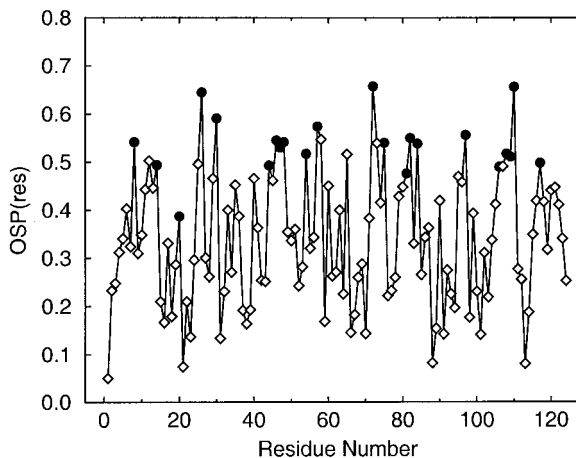

Figure 2. Residue packing of ribonuclease A. The individual residue occluded surface packing values (*OSP*) for the ribonuclease structure 7rsa are plotted *versus* residue number. *OSP* is an estimate of packing obtained by combining, for each residue, the occluded surface area and distance to occluding atom surfaces and was calculated using the occluded surface algorithm as described in Methods. The filled circles represent the individual *OSP* values for buried residues (defined as those residues with zero accessible surface area). The open diamonds represent the individual *OSP* values for solvent-exposed residues (all other residues). Some residues categorized as exposed may have only minor amounts of solvent-accessible surface area and therefore may have large *OSP* values. An example is Phe8 which has only 3.6 Å^2 of accessible surface area.

reflecting the wider range of occluded surface areas associated with exposed residues. Inspection of the data in Figure 2 readily provides a method to identify those residues most efficiently packed in a protein, including both buried and solvent-exposed residues. This information may be useful in the design of protein site mutations.

Since buried residues tend to have larger *OSP* values than surface residues, the average *OSP* of all residues in a protein should increase with increasing protein size as the number of buried residues relative to surface residues increases. The average *OSP* for each protein in a data set of 152 non-homologous proteins that range in size from 50 to 753 residues is plotted *versus* protein size in Figure 3. As seen in Figure 3, the average *OSP* increases markedly with number of residues per protein up to approximately 200 residues and then increases only slightly with larger proteins. This change in relationship that appears at approximately 200 residues reflects the previously documented observation that the maximum size of protein domains approaches 200 residues (Privalov, 1989). Larger proteins tend to be composed of a number of such domains. Thus the relative number of buried residues, as compared to exposed residues, increases very little for proteins larger than 200 residues and correspondingly the average packing value increases very little for
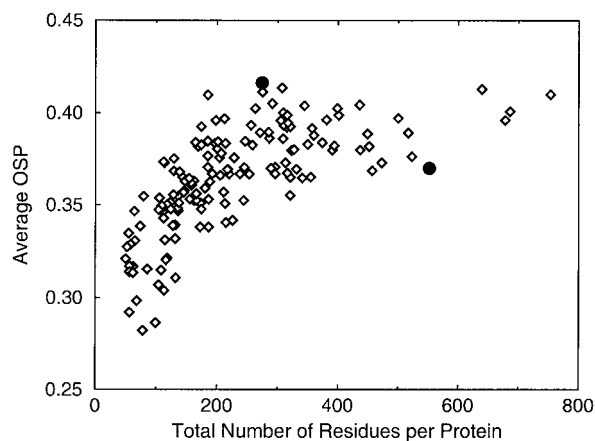
Figure 3. Average protein packing as a function of the total number of residues per protein chain. The average occluded surface packing values (*OSP*) for each of the 152 non-homologous protein chains in the Supplementary Material are shown. Average packing values are calculated as the average *OSP* for all residues in each respective protein chain as described in Methods. The two filled circles represent subtilisin Carlsberg (274 residues) and plant ascorbate oxidase (552 residues) which are discussed in the text.

these proteins also. This change in volume to surface area effect is not the only reason for the observed relationship, but it is a contributor. As will be discussed below, the *OSP* of solvent-exposed residues also changes with protein size. The filled circles in Figure 3 represent subtilisin Carlsberg (average *OSP*, 0.416) and plant ascorbate oxidase (average *OSP*, 0.370). These proteins are representatives of very high and very low *OSP* values, respectively and are also discussed below.

Although local variations in the packing of buried residues have been observed (Richards, 1974; Gerstein *et al.*, 1994) it has been generally assumed that the average packing of proteins is similar (Chothia, 1975; Pontius *et al.*, 1996). The results in Figure 3 show, even for proteins larger than 200 residues where the ratio of buried to solvent-exposed residues does not change a large amount, that the average *OSP* may vary by between 0.35 and 0.42. In view of the 20% variation in this measure of packing we have reinvestigated the question of packing differences between proteins. First, we determined if the differences in *OSP* were reflected in other measures of protein packing. A standard method for estimating protein packing is the Voronoi procedure. The Voronoi method is well known and commonly used but the occluded surface method has an advantage in that it directly estimates the packing of solvent-exposed residues as well as buried residues. Therefore we compared the average *OSP* and Voronoi packing density for buried residues only in the data set of protein structures. Figure 4(a) shows the average *OSP* of buried residues for the data set of 152 non-homologous proteins. The average *OSP* of all buried resi-

dues in the data set is 0.56. Using the current version of the occluded surface algorithm the *OSP* value obtained for the PHE residue in GLY-PHE-GLY crystals is 0.53. This is in agreement with earlier similar comparisons using the Voronoi method which showed that buried residue packing density in proteins was similar to small molecule crystal packing density (Richards, 1974). The van der Waals volume to Voronoi volume ratios, or packing densities, for buried residues for the same data set of 152 proteins are shown in Figure 4(b). The mean packing density in this data set of proteins is 0.74, a value within the range of 0.71-0.78 first reported by Richards using the two proteins, lysozyme and ribonuclease S (Richards, 1974). The range of packing densities (Figure 4(b)) is similar to the range of *OSP* values (Figure 4(a)). However, for many of the proteins in this data set the number of buried residues is small, especially for proteins with less than 200 total residues (see Figure 4(c)) and therefore meaningful comparisons of the average packing estimates between proteins is difficult when considering only the buried residues.

Of those proteins with more than 200 total residues subtilisin Carlsberg (1cse, 274 total residues, 50 buried residues) has the highest *OSP* of buried residues (0.590) and plant ascorbate oxidase (1aoz, 552 total residues, 49 buried residues) has the lowest *OSP* for buried residues (0.523). These are indicated as filled circles in Figure 4. The Voronoi packing densities for these two proteins are also unusually high (0.783) and low (0.708), respectively (see Figure 4(b)). A consideration of the overall *OSP* for all residues in these two proteins also shows the same trend, although such a comparison
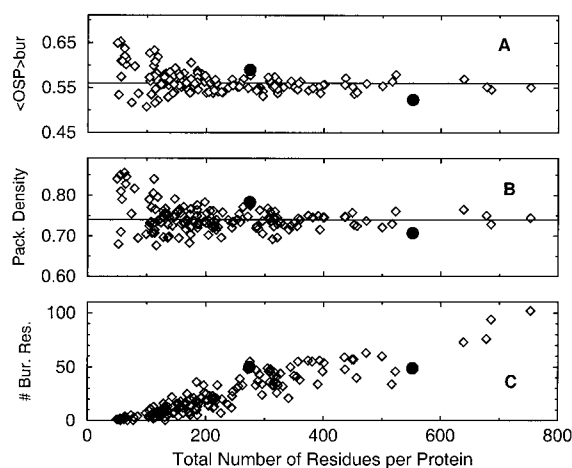

Figure 4. Packing of buried residues. (a) The average occluded surface packing, $\langle OSP \rangle_{\text{bur}}$; (b) average Voronoi packing density, for buried residues only are plotted *versus* the total number of residues in each protein chain listed in the Supplementary Material. (c) The number of buried residues in each protein is plotted. Buried residues are defined as those with zero accessible surface area. The two filled circles represent subtilisin Carlsberg (274 residues) and plant ascorbate oxidase (552 residues) which are discussed in the text.

must take into account differences in the number of buried residues *versus* solvent-exposed residues. Below we will explore whether or not the variation of packing seen in Figures 3 and 4 and exemplified by subtilisin Carslberg and plant ascorbate oxidase is due to normal statistical variation or due to real differences between proteins. But first we examine the packing of only the solvent-exposed residues for these proteins.

An advantage of the occluded surface method is that it allows one to directly obtain a packing value for surface residues as well as buried residues. The calculation of *OSP* combines two parameters: the extended normal distances (ray lengths) to all occluding atoms and the fractional occluded surface area. This combination provides a good estimate of the overall interaction with neighboring protein atoms but means that the *OSP* depends on the extent of surface occlusion by these neighboring atoms as well as the complementary nature of the interactions. Figure 5(a) shows the average *OSP* of solvent-exposed residues for the proteins in the data set. Again, the *OSP* increases rapidly with protein size up to approximately 200 residues and then only slightly with size for larger proteins. Of the two parameters that are combined in the calculation of *OSP* only the occluded surface areas of exposed residues increases with protein size, the average ray lengths have no significant correlation with protein size (data not shown).

The occluded surface of an atom in a protein structure is directly related to the atom density surrounding that atom. Therefore the observed change in occluded surface with size of the protein seen in Figure 5(a) may be explained by a change in atom

density surrounding exposed atoms in protein structures. As shown in Figure 5(b), the atom density (expressed as atoms per $Å^3$ within a radius of 6.5 Å) surrounding solvent-exposed atoms increases with protein size. This dependence is similar to the size-dependence of exposed residue *OSP* (Figure 5(a)) and explains the change in occluded surface area observed for exposed residues. The relationships of both overall average *OSP* and exposed residue average *OSP* with protein size shown in Figures 3 and 5 point out the fact that total van der Waals interactions change in a complex way with increasing protein size. Recently, specific atom densities about different amino acid residues have been extensively categorized (Karlin *et al.*, 1999). But the dependence of atom densities on protein size was not reported.

However it is important to note that the differences in atom density do not completely account for variations in the observed *OSP* of exposed residues. For example, the *OSP* of exposed residues for subtilisin Carslberg and plant ascorbate oxidase follows the same trend seen in buried residue average *OSP* values; subtilisin has a higher *OSP* than plant ascorbate oxidase (Figure 5(a), filled circles). Yet the atom density for exposed residues is not different for these two proteins (Figure 5(b), filled circles). This latter result indicates that the difference in *OSP* of exposed residues between these two very differently packed proteins is due to differences in the ray lengths as calculated by the occluded surface method and not to differences in occluded surface areas.

The distributions of ray lengths are a measure of the goodness of fit between the complementary van der Waals surfaces of the interacting atoms. This parameter gives an indication of the atom to atom packing for all atoms in a protein. We calculated the average of ray lengths for each protein in the data set of 152 proteins. The mean for all proteins in the data set is 0.906 (SD, 0.030) Å. Variation in average ray lengths between proteins is similar to the variation seen in *OSP* shown in Figures 3, 4 and 5 (data not shown). In this measure of packing efficiency the protein subtilisin Carlsberg (274 residues) has shorter than average ray lengths (0.883), indicating high packing efficiency, and plant ascorbate oxidase (552 residues) has longer than average ray lengths (0.961), indicating low packing efficiency. Thus the different packing efficiencies for these two proteins is seen with all calculated packing parameters.
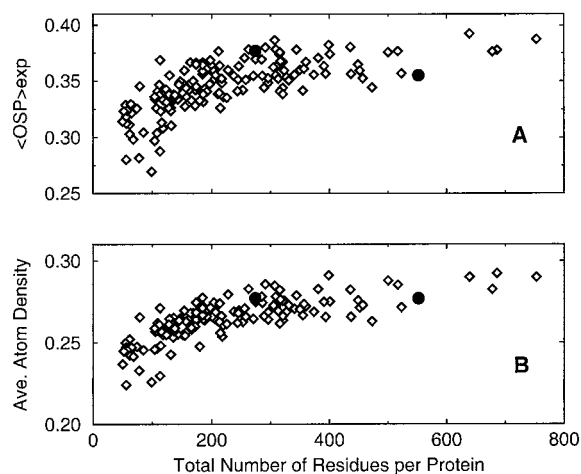
Figure 5. Packing of solvent exposed residues. (a) The average occluded surface packing, $\langle OSP \rangle_{exp}$, for solvent-exposed residues only is plotted *versus* the total number of residues in each protein chain listed in the Supplementary Material. (b) the average atom density (expressed as atoms per $Å^3$ within a radius of 6.5 Å) surrounding solvent-exposed atoms is plotted *versus* the total number of residues in each protein chain. Subtilisin Carlsberg (274 residues) and plant ascorbate oxidase (552 residues) (two filled circles) are discussed in the text.

Variation in average packing is not due to experimental conditions

The results in Figures 3, 4 and 5 show that individual proteins may have average packing parameters that vary as much as 20 %. Is this variation due to experimental differences, to different specific characteristic average packing for individual proteins, or to the expected statistical variation

described by Gerstein *et al.* (1994) for protein buried residue volume changes?

One experimental parameter that may result in different packing of X-ray crystal protein structures is the temperature at which data were collected. Many X-ray crystal structures are determined at cryogenic temperatures and others at or near 25 °C. Tilton *et al.* (1992) have determined the crystal structure of ribonuclease A at different temperatures between 50 and −172 °C. They report a 2.7 % decrease of all inter-atomic distances less than 3.25 Å as the temperature is decreased over this range. Using the same coordinate data set we observe a 4.5 % increase in the average *OSP* of buried residues and a 3.8 % increase in overall average *OSP* in ribonuclease A over this temperature range (data not shown). Similar small changes of average *OSP* are seen in other proteins whose structures have been determined at both room temperature and cryogenic temperature. For example: trypsinogen (1tgc and 2tgt), crambin (1crn and 1cbn); and elastase (1esa and 3est) all have average *OSP* differences of less than 2 % at the two temperatures. Thus, temperature alone cannot account for the 20 % range of packing parameters seen above.

Another experimental factor that may introduce variation in the packing of crystal structures is the crystal lattice. Several arguments may be made that different crystal lattice contacts should not be a significant energetic force in determining the overall packing of a protein. The total inter-protein crystal contact area in most protein crystals is less than the contact area in specific protein-protein complexes observed in solution (Janin & Rodier, 1995). Crystal lattice contacts may alter the packing of a few surface residues in a crystalline protein, but the binding energy of these contacts represents a small fraction of the overall conformational energy of a macromolecule. This energy is not large enough to force the macromolecule into conformations that are not readily available to the molecule in solution (Dickerson *et al.*, 1994). However, to investigate this question further we calculated the average *OSP* for hen lysozyme in three different crystal space groups (1hel, *P* 43 21 2; 2lzt, *P* 1; 4lyt, *P* 21). The average *OSP* for these three structures (0.377, 0.371, 0.376) varies by only 1.4 %. This small difference in overall *OSP* is despite the fact that the number of residues involved in crystal contacts in the 1hel structure is 33 per protein and in 4lyt the number of residues involved in contacts is only four per protein. Therefore, differences in crystal packing cannot account for the variation in packing values seen in the data set of non-homologous proteins.

Finally, we investigated the variation in occluded surface packing that may be introduced by differences in refinement protocols of the X-ray structures. For this purpose we calculated the average *OSP* in seven ribonuclease A structures with a resolution of 2.0 Å or better (1afk, 1aqp, 1bel, 1rnx, 1rpg, 1xps, 3rn3) from different laboratories. Four different programs (X-PLOR, Brünger, 1992; TNT, Tronrud *et al.*, 1987; PROFFT, Finzel, 1987;

RESTRAIN, Driessen *et al.*, 1989) were used in the refinement of these seven structures. The average *OSP* for these seven structures is 0.355 (SD, 0.0044). This represents a variation of 1.2 % and is consistent with the variation seen in different crystal forms for lysozyme described above. From all of these comparisons we expect the variation from experimental sources to be no more than 2 % for the data set of proteins examined here.

Dependence of average protein packing on secondary structure composition

Since the observed variation from the average seen in the packing parameters of individual proteins is much greater than expected from experimental conditions it is reasonable to consider that the packing of a particular protein is specific for that protein. If, in fact, packing is a differentiating characteristic of a protein it should be conserved in closely similar proteins. Of the two proteins with unusual packing discussed above (subtilisin Carlsberg and plant ascorbate oxidase) only subtilisin Carlsberg has closely similar proteins represented in the PDB. Subtilisin BL from Bacillus (1st3) has 73 % sequence homology to subtilisin Carlsberg and proteinase K from Tritirachium (1pek) has 51 % sequence homology. Both of these similar proteins have unusually high packing efficiency, as seen in the data presented in Table 1. All three proteins have average occluded surface packing greater than 0.41, a value which puts them among the proteins with the highest average *OSP* (cf. Figure 3). They also have average occluded surface ray lengths less than 0.89, a value shorter than the data set mean of 0.91, and their Voronoi packing densities are all greater than 0.77, well above the mean for proteins (cf. Figure 4). These results suggest that the unusual packing is conserved in this family of proteins. One major difference between subtilisin Carlsberg and plant ascorbate oxidase is that the subtilisin has 30 % of its residues in helical and 19 % in extended strand secondary structure while ascorbate oxidase has only 10 % in helical and 38 % in extended strand. We therefore investigated the dependence of protein packing on the relative composition of secondary structure as an explanation for the differences in observed protein packing. To eliminate effects of differing relative solvent exposures we compared only the average occluded surface ray lengths of whole proteins.

Figure 6 displays the average of occluded surface ray lengths for each protein in the data set *versus* the percentage of helix or strand in each structure. It can be seen from the results in this Figure that proteins with a higher percentage of helix tend to have shorter occluded surface ray lengths, and proteins with a higher percentage of strand tend to have longer occluded surface ray lengths. In other words, helices appear to be more efficiently packed compared to strands. This trend is also seen in the packing of buried amino acid residues only. All helical buried residues in the

Table 1. Comparison of packing for subtilisin-like protein

PDB code	Description	Number of residues	Number of buried residues[a]	Seq. homol.[b] (%)	Average *OSP*[c]	Average ray length (Å)[d]	Voronoi packing density[e]
1cse	Subtilisin Carlsberg	274	50	100	0.416	0.883	0.783
1st3	Subtilisin BL	269	45	73	0.415	0.886	0.775
1pek	Proteinase K	279	56	51	0.437	0.880	0.820

[a] Buried residues are defined as those residues with zero solvent-accessible surface area as calculated by the program ACCESS using a probe size of 1.4 Å.
[b] Pairwise sequence homology compared to subtilisin Carlsberg using the program XALIGN.
[c] Average residue packing value (*OSP*) calculated with the occluded surface method as described in Methods.
[d] The occluded surface extended normal distance (ray length) was calculated as described in Methods.
[e] Voronoi packing density was calculated using the program VOLUME and the atom radii listed in Methods.

data set have an average *OSP* of 0.577 compared to buried residues in extended strand, which have an average *OSP* of 0.551. The same result is seen in Voronoi packing densities: helical residue average packing density is 0.744 and strand residue packing density is 0.723. Residues in defined turns have a high packing efficiency: *OSP* of 0.577 and packing density of 0.794. However, the residues in turns seldom contribute to more than 25 % of a protein and therefore do not influence the overall average packing as much as either helix or strand is able to. From these results it is clear that secondary structure composition may have a significant effect on overall packing.

An interesting extension of the question relating to secondary structure packing is: what are packing interactions between different secondary structural elements? Table 2 shows the average partial *OSP* for all residues in the data set broken down into

Figure 6. Dependence of protein packing on composition of secondary structure. The averages of occluded surface ray lengths were calculated for each protein listed in the Supplementary Material. Ray length is the length (in Å) of the extended normal from an atom surface to its intersection with an occluding atom surface. See Figure 1 for a description of the extended normals. The average ray length is plotted *versus* the percentage of (a) helix or (b) strand in each protein. A linear regression is shown for the data in (a) and (b) by a continuous line. The mean ray length for all proteins is 0.91 and is indicated by a broken line.

categories by the secondary structure classification of the occluding residue. As expected from the close interaction of strands in beta-sheets, strand-strand interaction shows the highest inter-element packing. Other categories of interactions between specific secondary structure elements show decreased *OSP* values. Second in packing efficiency is the interaction between residues within the same helical segment. This is primarily due to cross-helix interactions within the backbone atoms. It is this latter class of interactions that cause the high packing associated with helical segments.

Is overall packing conserved in protein families of similar secondary structure other than the subtilisin-like family discussed above? In order to investigate this question we have calculated occluded surface packing within three families of proteins which range in size from approximately 130 residues to approximately 375 residues. The data in Table 3 compare the average *OSP* and occluded surface ray lengths for proteins within the alcohol dehydrogenase, dihydrofolate reductase and lysozyme families from various species. One would expect the *OSP* to vary between families because of differences in overall geometry giving rise to differences in the ratio of buried residues to solvent-exposed residues and to differences in atom density surrounding individual residues. However, the average *OSP* data are interesting in two respects. First, the variation in average *OSP* within each family is small, as indicated by the small standard deviation of *OSP* values listed in Table 3. This standard deviation represents a variation in *OSP* of less than 5 % from the family mean value. This small variation is evidence for conservation of packing across related proteins with sequence homology of as little as 35 %. Second, as the family with the smallest proteins, one would expect the lysozymes to have the lowest average *OSP* value, yet the average *OSP* for the family is 0.379, intermediate between the other two families of larger proteins and high for proteins of only approximately 130 residues (cf. Figure 3).

To estimate the conservation of packing within these protein families independent of protein size and volume to surface ratio, we also calculated the average occluded surface ray lengths for each protein. The average ray length value for the alcohol
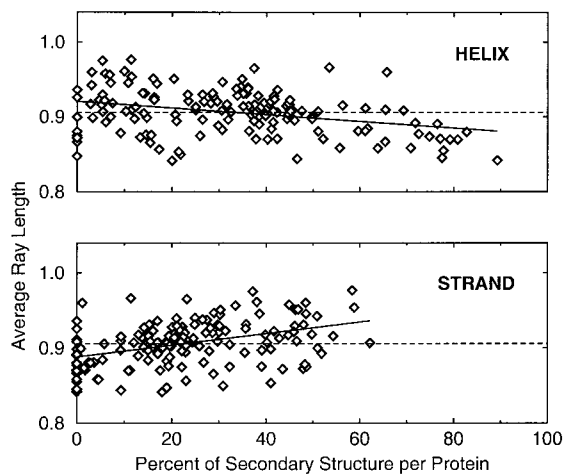
Table 2. Dependence of occluded surface packing on secondary structure interactions

	Intra	>Helix	>Strand	>Turn	>Other
Helix	0.189 (0.085)	0.122 (0.089)	0.096 (0.077)	0.075 (0.061)	0.079 (0.064)
Strand	0.077 (0.040)	0.097 (0.076)	0.210 (0.103)	0.066 (0.061)	0.085 (0.066)

Each residue in the data set of proteins listed in the Supplementary Material was categorized into one of four secondary structures using the program DSSP. The occluded surface algorithm was used to calculate the *OSP*, and the identity of occluding atom, for each occluded surface patch for all residues in the data set. The secondary structure-specific *OSP* for each residue is calculated as follows:

$$OSP_{ss} = \frac{[OS_{ss} * \langle 1 - RL \rangle_{ss}]}{MS_{res}}$$

where OS_{ss} is the occluded surface area occluded by atoms belonging to residues in helix, strand, turn or other secondary structure, MS_{res} is total molecular surface of the residue (sum of occluded and non-occluded areas) and RL is the length of the extended normal divided by 2.8 Å. Values in the Table are the average *OSP* values with standard deviations in parentheses. The secondary structure assignment of the occluded residue is indicated on the left of the Table and the secondary structure assignment of the occluding residue is indicated across the top. Interactions within the same secondary structural element are labeled Intra.

dehydrogenase protein family (0.927; SD, 0.016) is not significantly different from the average ray length value for the dihydrofolate reductase family (0.940; SD, 0.019). However, the lysozyme family average ray length value is significantly lower (0.867; SD, 0.014) than either of these protein families. Here, shorter ray lengths indicate more efficient packing. This difference indicates that the lysozymes are more efficiently packed than the other two families of proteins.

From the data in Table 3 and the data presented in Table 1 for subtilisin-like proteins we conclude that differences in packing between proteins appears to be reasonably well conserved in protein families. A partial explanation for the unusually high occluded surface packing of the lysozymes is the high helical content and low extended strand content compared to the other families shown in Table 3. Lysozyme has approximately 40% helix and 11% strand, while

Table 3. Packing of protein families

Family and PDB code	Species	Number of residues	Number of buried residues[a]	Seq. homol.[b] (%)	Average *OSP*[c]	Average ray length (Å)[d]	% Helix[e]	% Strand[e]
Alcohol dehydrogenase					**0.391 (0.007)**	**0.927 (0.016)**		
1deh	Human	374	61	100	0.392	0.934	28	25
2ohx	Horse	374	56	93	0.384	0.931	29	25
1cdo	Cod	374	50	72	0.385	0.944	31	25
1qor	*E. coli*	326	35	39	0.390	0.901	35	25
1uok	*B. cereus*	558	68	35	0.402	0.925	34	19
Dihydrofolate reductase					**0.350 (0.005)**	**0.940 (0.019)**		
3dfr	*L. casei*	162	11	100	0.352	0.941	25	33
4dfr	*E. coli*	159	13	51	0.347	0.936	26	32
8dfr	Chicken	186	16	48	0.355	0.907	23	34
1drf	Human	186	14	47	0.354	0.948	22	32
1ai9	*Candida*	192	18	49	0.352	0.943	23	35
1dyr	*Pneumocystis*	205	14	43	0.341	0.967	25	30
Lysozymes					**0.379 (0.004)**	**0.867 (0.014)**		
4lyt	Chicken	129	13	100	0.376	0.870	41	11
2ihl	Quail	129	12	98	0.375	0.870	40	11
135l	Turkey	129	14	97	0.379	0.888	41	11
1ghl	Pheasant	129	13	96	0.384	0.880	41	11
1hhl	Guinea fowl	129	11	96	0.375	0.870	42	11
1lmn	Trout	129	11	77	0.379	0.854	38	11
1lz1	Human	130	12	76	0.385	0.863	37	12
2eql	Horse	129	14	70	0.379	0.845	40	11

[a] Buried residues are defined as those residues with zero solvent-accessible surface area as calculated by the program ACCESS using a probe size of 1.4 Å.

[b] Pairwise sequence homology compared to the first sequence in each family using the program XALIGN.

[c] Average residue packing value (*OSP*) calculated with the occluded surface method as described in Methods. The bold numbers indicate the mean *OSP* and standard deviation for each protein family.

[d] Ray length is the length of the extended normal from an atom surface to its intersection with an occluding atom surface. See Figure 1 for a description of the extended normals. The numbers listed are the average ray lengths for all residues in each protein; the bold numbers indicate the mean and standard deviation for each family.

[e] Each residue was assigned to a secondary structure category using the program DSSP as described in Methods.

the other two families have more equal distributions. However, it is clear from the results shown in Figure 6 that the composition of helix and strand cannot completely explain the overall packing values of a protein.

Dependence of average protein packing on amino acid composition

Finally, we address the question of whether or not the differences in packing shown here are partially due to amino acid composition differences. To answer this question we calculated the average of occluded surface ray lengths for each amino acid type in the data set of 152 proteins. These results were further divided to indicate the influence of secondary structure on the ray lengths of the amino acid residue in question. The averages of occluded surface ray lengths of each type of amino acid residue, for each general type of secondary structure, are plotted in Figure 7. The residues with shortest average ray lengths are CYS, the aromatic residues Trp and Tyr, and the small residue Gly. The residues with longest ray lengths are the aliphatic residues Val, Ile and Leu. Residues of all types of amino acid residues in helical segments tend to have shorter ray lengths than residues in strand segments in agreement with the results shown in Figure 6 and discussed above. Since cysteine residues have the shortest ray lengths and would be expected to increase packing density in the vicinity of disulfide bonds we asked whether or not the overall packing values of proteins in the data set were correlated with percentage of cysteine composition. The cysteine composition ranges from zero to 13% in proteins within this data set and there was no correlation with *OSP* values (correlation coefficient −0.36). Although the different average ray length values are not significantly different between amino acid types because of the large standard deviations (typically 0.12 Å) an amino acid composition favoring the aromatic and small residues would tend to decrease the average ray length of a protein and *vice versa*.

In fact the amino acid compositions of the alcohol dehydrogenases, dihydrofolate reductases and lysozymes follow this prediction. Figure 8 shows the percentage amino acid composition for each of these families of proteins. The order of amino acid types in Figure 8 is the same as in Figure 7 for comparison. Lysozyme (Figure 8(c)) has relatively more of the closely packed Cys, Trp and Gly residues and relatively fewer of the loosely packed residues Val, Ile and Leu compared to either the alcohol dehydrogenases (Figure 8(a)) or dihydrofolate reductases (Figure 8(b)). This difference in amino acid composition also contributes to the unusually high packing efficiency of the lysozymes.

Conclusions

We have investigated the question of whether or not the atomic packing in globular proteins shows specific differences between proteins. The occluded surface algorithm was used to estimate the packing
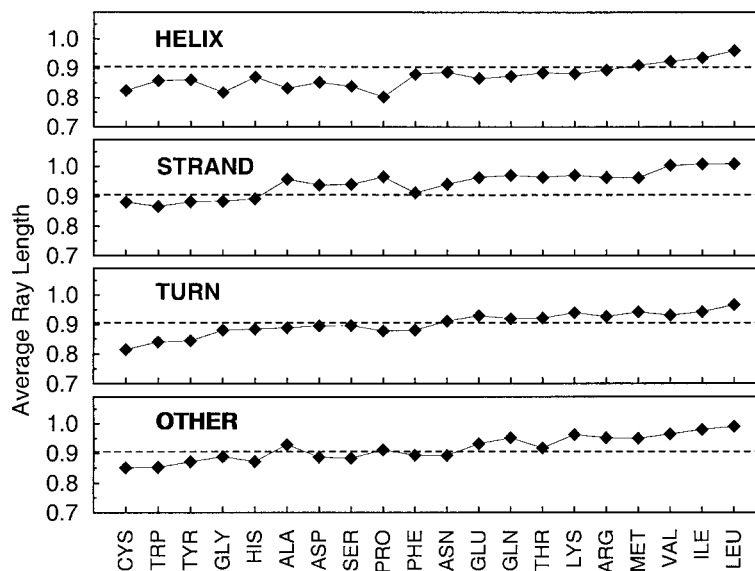

Figure 7. Dependence of amino acid residue packing on secondary structure assignment. The averages of occluded surface ray lengths were calculated for all amino acid residues in the proteins listed in the Supplementary Material. Ray length is the length in Å of the extended normal from an atom surface to its intersection with an occluding atom surface. See Figure 1 for a description of the extended normals. Each residue was also categorized according to secondary structure assignment. The average of ray lengths for each residue in four different secondary structure categories is plotted. The order of amino acids is according to their rank order of overall average ray lengths. The mean ray length for all residues is 0.91 (broken line).

of both buried residues and solvent-exposed residues. Occluded surface packing results for buried residues were consistent with Voronoi volume packing densities. We find that some proteins may have packing efficiencies that vary significantly from the average. The 20 % variation of packing efficiencies seen in a large data set of proteins is not completely due to statistical variation. Rather, specific protein types have specific packing that is determined by the protein size, composition of secondary structure and composition of amino acid residues. The functional significance of this differentiated packing remains to be elucidated. Specific packing efficiencies may correlate with the functions of different proteins. Calculation of packing values using a sensitive method such as described here, both for whole proteins and for individual segments of proteins, provides a new parameter with which to investigate these possible correlations. Future results in this direction may be useful for the prediction of protein function in structural genomic studies. More immediately, this result has ramifications for evaluation of protein models and modeling of protein structures. Quality assessments of packing in protein structures should include a consideration of various parameters including the packing of known homologous proteins. One would expect increased overall packing in large proteins, proteins with a high percentage of helix and proteins with a high percentage of aromatic and small residues and a low percentage of aliphatic residues. Furthermore, since protein packing is well conserved in protein families of similar structure, a protein structural model based on a homologous template structure should have the same overall packing as the template structure.
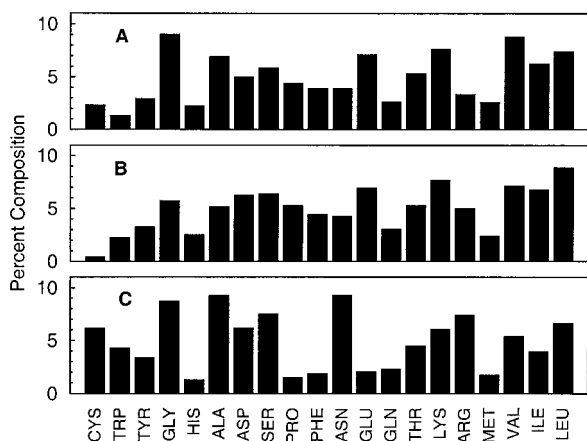

Figure 8. Amino acid composition of protein families. The average percent compositions of each amino acid type for (a) the alcohol dehydrogenases, (b) the dihydrofolate reductases and (c) the lysozymes listed in Table 3 are plotted. The order of amino acid residues is the same as in Figure 7 for comparison.

Methods

Data sets

A data set of 152 non-homologous X-ray crystal protein structures of single chains with a resolution of $< = 2.0$ Å and an R factor of $<=0.20$ and containing all non-hydrogen atoms was used for the analysis of packing (Supplementary Material). The original list was obtained as described by Hobohm & Sander (1994) from the PDB (Berstein *et al.*, 1977) and culled to meet the all non-hydrogen atom criterion and to remove ASX or GLX amino acid residues or other non-standard amino acid residues and to remove membrane proteins. In the case of alternative conformations of residues only the first conformer (A) was used. HETATOM records and hydrogen atoms, if present, were not used in the packing analysis.

Packing analysis

An extension of the occluded surface algorithm (Pattabiraman *et al.*, 1995) as described by DeDecker *et al.* (1996) was used for packing analysis. In brief, a molecular dot surface of each residue is calculated with a 1.4 Å probe using the MS program (Connolly, 1985). Dot density was chosen such that each dot represents approximately 0.215 Å^2 of surface area. Higher dot densities did not significantly change the results of the occluded surface calculation but did significantly extend the computation time. A normal is extended radially from each dot until it either intersects the van der Waals surface of a neighboring atom or reaches a length of 2.8 Å (the diameter of a water molecule). Occluded surface, *OS*, is that molecular surface area on the originating atom associated with normals that intersect with another atom surface as opposed to reaching the 2.8 Å limit. All other molecular surface area is considered non-occluded. The packing value, *OSP* for each residue is defined as:

$$OSP = \frac{\sum_{\text{atom}}^{\text{res}} \left[OS_{\text{atom}} * \langle 1 - RL \rangle_{\text{atom}} \right]}{MS_{\text{res}}}$$

where MS_{res} is total molecular surface of the residue (sum of occluded and non-occluded areas) and *RL* is the length of the extended normal divided by 2.8 Å. Division by the total molecular surface area normalizes the *OSP* value to account for various sizes of amino acid residues. The average occluded surface packing value (*OSP*) for a protein is simply the average of all residue *OSP* values for that protein.

As a baseline example, the maximum theoretical *OSP* value for spherical object packing is that obtained for a hexagonal array of closely packed spheres. For such an array of radius 1.4 Å spheres, the *OSP* is 0.79. This value represents an occluded surface area of 23.9 Å^2, a total surface area of 24.6 Å^2 and an average ray length (extended normal) of 0.36 Å. The occluded surface area is slightly less than the total surface area because some extended normals are able to extend into void space for a distance greater than 2.8 Å.

Small probe contact dots were calculated using the program PROBE (Word *et al.*, 1999). The program ACCESS (Lee & Richards, 1971) was used with a probe size of 1.4 Å to calculate buried residues. Those

residues with zero solvent-accessible surface area were considered buried. Voronoi volumes were calculated using the program VOLUME (Richards, 1974) using the Richards method B for placing the inter-atomic planes. The van der Waals volumes were calculated using the MSROLL program (Connolly,1993) with a probe size of 1.4 Å and atom radii given below. Voronoi volume vertices for Figure 1 were obtained using the program DUMP-POLYHEDRA (Gerstein et al., 1995). Secondary structure was calculated using the program DSSP (Kabsch & Sanders, 1983). The program XALIGN (Wishart et al., 1994) was used to calculate pair-wise sequence homology.

Atom radii

The atom radii used in the packing analysis are: C, 1.90 Å; N, 1.50 Å; O, 1.40 Å; S, 1.85 Å. The choice of van der Waals radii for analysis of protein structures is a difficult one. Lists of unified atomic radii for C, N, O and S have been derived from inspection of small molecule crystal structures (Pauling, 1967; Kitaigorodsky, 1973; Rowland & Taylor, 1996; Tsai et al., 1999) and protein structures (Li & Nussinov, 1998), from a consideration of liquid state properties (Bondi, 1964; Jorgensen et al., 1984), and from calculation of electron density contours at specific charge densities (Arteca, 1991). Many radii used in the analysis of protein structures have been derived from the original lists of Pauling (1967) or Bondi (1964) and vary by ~0.2 Å (Lee & Richards, 1964; Rashin, 1984; Word, 1999) although more recent protein simulation studies suggest that the effective radii may be larger (Jorgensen & Tirado-Rives, 1988). The radii used here are essentially the same as the original values published by Pauling (1967) for C, N, O and S. These radii are at the low end of the published range for unified atom radii. However, given the recent demonstration by Word et al. (1999) that most hydrogen atoms in high-resolution crystal structures are ''interdigitated'' it is more appropriate to use these atom radii for the packing analysis described here. Most important, all calculations done here used the same atom radii list and therefore the conclusions made concerning comparisons within the data set of proteins are valid.

References

Abagyan, R. A. & Totrov, M. M. (1997). Contact area difference (CAD): a robust measure to evaluate accuracy of protein models. J. Mol. Biol. 268, 678-685.

Arteca, G. A., Grant, N. D. & Mezey, P. G. (1991). Variable atomic radii based on some approximate configurational invariance and transferability properties of the electron density. J. Comput. Chem. 12, 1198-1210.

Bauer, A. & Beyer, A. (1994). An improved pair potential to recognize native protein folds. Proteins: Struct. Funct. Genet. 18, 254-261.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. 112, 535-542.

Bondi, A. (1964). van der Waals volumes and radii. J. Phys. Chem. 68, 441-451.

Brünger, A. T. (1992). X-PLOR, Version 3.1 Manual, Yale University, New Haven, CT.

Carson, M. (1987). Ribbon models of macromolecules. J. Mol. Graph. 5, 103-106.

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

Colovos, C. & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 2, 1511-1519.

Connolly, M. L. (1985). Computation of molecular volume. J. Am. Chem. Soc. 107, 1118-1124.

Connolly, M. L. (1993). The molecular surface package. J. Mol. Graph. 11, 139-141.

Dahiyat, B. I. & Mayo, S. L. (1997). Probing the role of packing specificity in protein design. Proc. Natl Acad. Sci. USA, 94, 10172-10177.

DeDecker, B. S., O'Brien, R., Fleming, P. J., Geiger, J. H., Jackson, S. P. & Sigler, P. B. (1996). The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. J. Mol. Biol. 264, 1072-1084.

Dickerson, R. E., Goodsell, D. S. & Neidle, S. (1994). ''. . . the tyranny of the lattice . . .''. Proc. Natl Acad. Sci. USA, 91, 3579-3583.

Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Khan, G. & Moss, D. S. (1989). RESTRAIN: restrained structure-factor least-squares refinement program for macromolecular structures. J. Appl. Crystallog. 22, 510-516.

Finney, J. L. (1970). Random packings and the structure of simple liquids. I. The geometry of random close packing. Proc. Roy. Soc. ser. A, 319, 479-493.

Finzel, B. C. (1987). Incorporation of fast Fourier transforms to speed restrained least-squares refinement of protein structures. Acta Crystallog. 20, 5355.

Gerstein, M., Sonnhammer, E. L. & Chothia, C. (1994). Volume changes in protein evolution. J. Mol. Biol. 236, 1067-1078.

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

Gregoret, L. M. & Cohen, F. E. (1990). Novel method for the rapid evaluation of packing in protein structures. J. Mol. Biol. 211, 959-974.

Hobohm, U. & Sander, C. (1994). Enlarged representative set of protein structures. Protein Sci. 3, 522-524.

Janin, J. & Rodier, F. (1995). Protein-protein interaction at crystal contacts. Proteins: Struct. Funct. Genet. 23, 580-587.

Jorgensen, W. L. & Tirado-Rives, J. (1988). The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J. Am. Chem. Soc. 110, 1657-1666.

Jorgensen, W. L., Madura, J. D. & Swenson, C. J. (1984). Optimized intermolecular potential functions for liquid hydrocarbons. J. Am. Chem. Soc. 106, 6638-6646.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers, 22, 2577-2637.

Karlin, S., Zhu, Z.-Y. & Baud, F. (1999). Atom density in protein structures. *Proc. Natl Acad. Sci. USA,* **96**, 12500-12505.

Kitaigorodsky, A. I. (1973). Molecular crystals and molecules. *Phys. Chem.* **29**, 1-223.

Kleywegt, G. J. & Jones, T. (1997). Model-building and refinement practice. *Methods Enzymol.* **277**, 208-230.

Kono, H., Nishiyama, M., Tanokura, M. & Doi, J. (1998). Designing the hydrophobic core of Thermus flavus malate dehdrogenase based on side-chain packing. *Protein Eng.* **11**, 47-52.

Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.

Lee, C. & Subbiah, S. (1991). Prediction of protein side-chain conformation by packing optimization. *J. Mol. Biol.* **217**, 373-388.

Lesk, A. M. & Chothia, C. (1980). How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins. *J. Mol. Biol.* **136**, 225-270.

Li, A. J. & Nussinov, R. (1998). A set of VDW and columbic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins: Struct. Funct. Genet.* **32**, 111-127.

Maiorov, V. N. & Crippen, G. M. (1992). Contact potential that recognizes the correct folding of globular proteins. *J. Mol. Biol.* **227**, 876-888.

Manavalan, P. & Ponnuswamy, P. K. (1977). A study of the preferred environment of amino acid residues in globular proteins. *Arch. Biochem. Biophys.* **184**, 476-487.

Melo, F. & Feytmans, E. (1997). Novel knowledge-based mean force potential at atomic level. *J. Mol. Biol.* **267**, 207-222.

Miyazawa, S. & Jernigan, R. L. (1996). Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **256**, 623-644.

Narayana, S. V. L. & Argos, P. (1984). Residue contacts in protein structures and implications for protein folding. *Int. J. Pept. Protein Res.* **24**, 25-39.

Pauling, L. (1967). *The Chemical Bond. A Brief Introduction to Modern Structural Chemistry*, pp. 135-153, Cornell University Press, Ithaca, NY.

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

Ponder, J. W. & Richards, F. M. (1987). Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **193**, 775-791.

Pontius, J., Richelle, J. & Wodak, S. J. (1996). Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J. Mol. Biol.* **264**, 121-136.

Privalov, P. L. (1989). Thermodynamic problems of protein structure. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 47-69.

Ptitsyn, O. B. & Volkenstein, M. V. (1986). Protein structures and the neutral theory of evolution. *J. Biomol. Struct. Dynam.* **4**, 134-156.

Rashin, A. A. (1984). Buried surface area, conformational entropy, and protein stability. *Biopolymers,* **23**, 1605-1620.

Ratnaparkhi, G. S., Ramachandran, S., Udgaonkar, J. B. & Varadarajan, R. (1998). Discrepancies between the NMR and X-ray structures of uncomplexed barstar: analysis suggests that packing densities of protein structures determined by NMR ar unreliable. *Biochemistry,* **37**, 6958-6966.

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

Rowland, R. S. & Taylor, R. (1996). Intermolecular non-bonded contact distances in organic crystal structures: comparison with distances expected from van der Waals radii. *J. Phys. Chem.* **100**, 7384-7391.

Singh, J. & Thornton, J. M. (1992). *Atlas of Protein Side-Chain Interactions*, vols 1 and 2, pp. 1-846, IRL Press, Oxford.

Tanaka, S. & Scheraga, H. A. (1976). Medium- and long-range interaction parameters between amino acids for predicitng three-dimensional structures of proteins. *Macromolecules,* **9**, 945-950.

Tilton, R. F., Dewan, J. C. & Petsko, G. A. (1992). Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different tempera- tures from 98 to 320 K. *Biochemistry,* **31**, 2469-2481.

Tronrud, D. E., Ten, Eyck L. F. & Matthews, B. W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallog. sect. A,* **43**, 489-501.

Tsai, J., Taylor, T., Chothia, C. & Gerstein, M. (1999). The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* **290**, 253-266.

Vriend, G. & Sander, C. (1993). Quality control of protein models: directional atomic contact analysis. *J. Appl. Crystallog.* **26**, 47-60.

Warme, P. K. & Morgan, R. S. (1978). A survey of atomic interactions in 21 proteins. *J. Mol. Biol.* **118**, 273-287.

Wishart, D., Boyko, R. & Sykes, B. (1994). Constrained multiple sequence alignment using XALIGN. *CABIOS,* **10**, 678-688.

Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S. & Richardson, D. C. (1999). Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* **285**, 1711-1733.

JMB Online

http://www.academicpress.com/jmb

Supplementary material comprising one Table is available from JMB Online.