

The Origin and Extent of Coarse-Grained Regularities in Protein Internal Packing

Zerrin Bagci,¹ Andrzej Kloczkowski,² Robert L. Jernigan,² and Ivet Bahar^{1*}

¹Center for Computational Biology & Bioinformatics, and Department of Molecular Genetics & Biochemistry, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania

²Laurence H. Baker Center for Bioinformatics and Biological Statistics, Iowa State University, Ames, Iowa

ABSTRACT Despite the suitability of various lattice geometries for coarse-grained modeling of proteins, the actual packing geometry of residues in folded structures has remained largely unexplored. A strong tendency to assume a regular packing geometry is shown here by optimally reorienting and superimposing clusters of neighboring residues from databank structures examined on a coarse-grained (single-site-per-residue) scale. The orientation function (or order parameter) of the examined coordination clusters with respect to fcc lattice directions is found to be 0.82. The observed geometry, which may be termed an incomplete distorted face-centered cubic (fcc) packing, is apparently favored by the drive to maximize packing density, in a fashion analogous to the way identical spheres pack densely and follow fcc geometry. About 2/3 of all residues obey this packing geometry, while the remainder occupy other context-dependent positions. The preferred coordination directions show relatively small variations over the various amino acid types, consistent with uniform residue viewpoint. Both the extremes of solvent-exposed and completely buried residue neighborhoods approximate the same generic packing, the only difference being in the numbers (and not the orientations) of coordination sites that are occupied (or left void for solvent occupancy). We observe the prevalence of a rather uniform (tight) residue packing density throughout the structure, including even the residues packed near solvent-exposed regions. The observed orientation distribution reveals an underlying, intrinsic orientation lattice for proteins. *Proteins* 2003;52:56–67. © 2003 Wiley-Liss, Inc.

Key words: residue coordination geometry; face-centered-cubic lattice; packing density; order parameters; closest cubic packing; quaternion superimposition method

INTRODUCTION

Residue packing has been suggested to play a selective role in determining protein structure.^{1–3} Yet, the way in which residues pack in folded proteins, and the extent of a preferential geometry constraining a structure are still open issues, as is the question even of the existence of such inherent geometric packing preferences.

Several controversial models have been advanced for the packing of sidechains in globular proteins; these differ in their regularities and degrees of freedom. The extreme cases are ideal packing conforming with the closest packed cubic geometry of identical spheres,⁴ perfect complementarity similar to a jigsaw puzzle,^{1,5} or completely random arrangement⁶ (Fig. 1). There have been other studies of directional specificities for interacting pairs of residues, indicating some preferred directions, but also some inherent plasticity consistent with the tolerance to mutations.⁷ The aim of the present work is to explore the occurrence of generic packing characteristics for residue clusters. Understanding how amino acids are packed, assessing the extent of randomness/regularity in their spatial arrangements, and defining their coordination patterns, are issues of crucial importance for the design of proteins and their complexes.

Proteins exhibit structural motifs or secondary/supersecondary structures that recur regardless of amino acid type. Such generic structural properties provide an explanation for the commonly observed insensitivity of structures to single site mutations.^{8,9} These patterns might also provide insights into the responses of proteins to external agents, and could have important therapeutic implications. Insensitivity to mutations could be suggestive of either an absence of amino acid *specificity* in their packing, or a lack of *regularity* or *order*. Non-specific behavior could indeed be associated with either of these two opposite views: (1) an *ordered* coarse-grained packing, say of identical spheres, that can tolerate substitutions over some range of sizes and shapes, or (2) a *disordered* packing having sufficient free volume to accommodate substitutions sometimes with concomitant local conformational changes. Tolerance to mutations cannot give unambiguous information about packing characteristics.

Proteins may have some evolutionarily selected and conserved *regular* architectures for optimizing residue coordination, required for biological function or stability, which could be stabilized by many alternative residues.

*Correspondence to: Ivet Bahar, Center for Computational Biology & Bioinformatics, and Department of Molecular Genetics & Biochemistry, School of Medicine, University of Pittsburgh, W1043 Biomedical Science Tower, 200 Lothrop St, Pittsburgh, PA 15213. E-mail: bahar@pitt.edu.

Received 15 May 2002; Accepted 3 February 2003

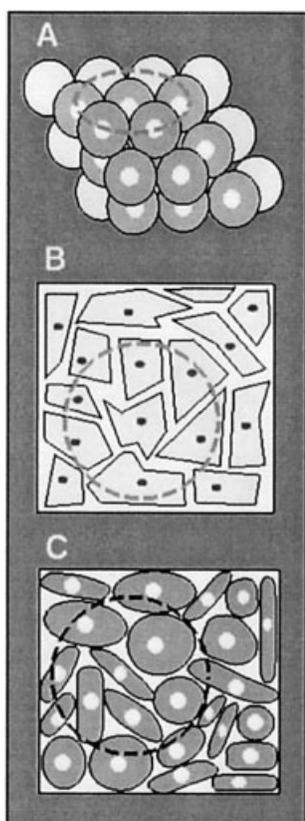


Fig. 1. Different coarse-grained models for residue packing in proteins: (A) face-centered-cubic lattice packing, the completely ordered and regular geometry proposed by Raghunathan and Jernigan,⁴ (B) jigsaw puzzle model emphasizing shape complementarity of neighboring residues,^{1,5} and (C) nuts-and-bolts model where sidechains mimic random association of nuts and bolts in a jar.⁶ The dashed circle shows the first coordination shell around a central residue in each case. Parts (B) and (C) may be viewed as a cross-section of the 3D-representation. Notably, despite their differences, the three coarse-grained models have common characteristics, such as the average number (6 in this schema) of residues within a first shell, the most probable angular positions of the neighbors, and the overall packing density. Note that the inter-residue distances as determined from the centers of interaction (shown by dots) can vary depending on the residue sizes, whereas the coordination angles themselves exhibit greater regularities.

Helical motifs, for example, are suggested by Maritan et al. to have emerged due to such generic packing preferences.¹⁰ Several studies support the concept of compactness inducing significant secondary structure.^{11–16} Statistical analyses based on Delaunay tessellation also indicate a correlation between coordination architecture and secondary structure.¹⁷

While several studies indicate a relationship between regular backbone conformations and packing efficiency, no direct evidence of regular (non-bonded) coordination has been reported to date. Studies aimed at characterizing amino acid sidechain coordination geometry^{7,17–20} indicate some nonrandomness, but this would not prevent the association of sidechains from being ductile similar to those in the nuts-and-bolts model.⁶ Our recent study²¹ suggests that the regularities observed on a coarse-grained scale can be attributed to the generic tendency of residues to be filling uniformly the protein interiors in a

randomly distributed spacing, rather than selecting particular coordination directions. This study did not distinguish among different types of amino acids. Selectivities imparted by residue specificities, therefore, were not accounted for. The questions to answer in a more informative analysis of residue packing are simple and direct: is there any intrinsic orientation regularity in residue packing, and if so, is it residue-specific?

In the present study, contrasting views of packing are reconciled. We recently showed in a detailed examination of databank structures that the twelve coordination sites of the fcc geometry represent directions that are “likely” to be occupied by the first neighbors (about 6–7, on the average^{22,23}). But usually not all of these sites are filled, except for the very most densely packed core regions.²¹ One can equally predefine other coordination directions (usually conforming to particular lattice geometries), and the residue clusters can be fit with good fidelity to such predefined sites. The quality of the fit naturally improves with the increasing coordination number of the predefined lattice.^{24–27} On the other hand, in the absence of a *a priori* choice of a lattice geometry, a generic fcc-like architecture emerges, and as will be shown this architecture shows little dependence on residue type.

Importantly, fcc packing is the *closest* packing geometry of identical spheres.^{28,29} Thus, amino acids, when observed at coarse-grained scale, tend to assume this *universal closest packing geometry*. This observation suggests that the drive for maximizing packing efficiency not only stabilizes secondary structures (helices), as pointed out in other studies,^{10–16} but also induces regularities in tertiary packing. The requirement for achieving a high inter-residue packing geometry, and the ensuing tendency for closest cubic packing, could be exploited to reduce the conformational space in the quest for, and engineering of, tertiary structures. A recent study supports the view that packing may indeed be a key factor in selecting or stabilizing the hydrophobic cores of proteins.³⁰

MATERIALS AND METHODS

A total of 28,730 residue clusters were extracted from 150 nonhomologous Protein Data Bank³¹ (PDB) structures. The database of analyzed protein coordination clusters is deposited as a supplementary material, and the reasons for selecting this database were explained in our earlier articles.^{21,24} Figure 2 illustrates a coordination cluster formed by a central residue (Gly65) in myoglobin³² and the surrounding $m = 10$ residues. A bundle of directional unit vectors pointing from the central residue toward the m coordinating residues in a close neighborhood characterizes each cluster. Residues are represented by their C^β atoms, except glycine, for which C^α atoms are used. The use of C^β atoms for representing residues has been common practice in single-site-per-residue models of proteins. An alternative approach is the choice of sidechain centroids. We have adopted C^β atoms because the different sizes and rotameric states of sidechains could make it more difficult to discern the possible packing regularities of residues.

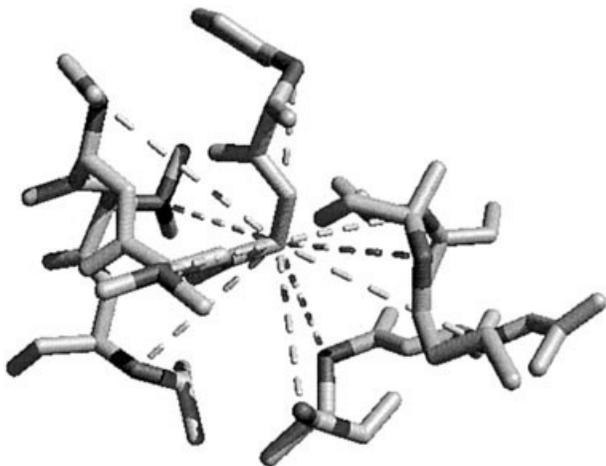


Fig. 2. An example of a cluster of residues from a databank structure (myoglobin, PDB code: 1mbn³²). Here, residue Gly65 is selected as the central residue. The dashed lines represent the coordination vectors that relate the central residue to the C β atoms of 10 other residues ($m = 10$), located within 6.8 Å distance. The cluster of coordinating residues contains both bonded and non-bonded neighbors (residue numbers: 22, 25, 27, 29, 62, 64, 66, 67, 68, 69).

The number and geometry of the residues in the clusters can vary with the cut-off distance, hence the use of the Delaunay tessellation method in previous examinations of residue clusters.³³ The clusters examined in the present study consist of the neighbors located within a first coordination shell near a central residue. A cutoff distance of 6.8 Å has been found in previous knowledge-based studies of PDB structures to be a realistic distance for including first neighbors in the adopted single-site-per-residue model.^{22,23} Bonded and non-bonded neighbors are not distinguished. A justification for this approximation also comes from the Gaussian network model that successfully describes the fluctuation dynamics of proteins with a single parameter harmonic potential for all (bonded and non-bonded) contacts.³⁴ Additional support is provided by our recent study comparing coordination angles for bonded and non-bonded neighbors.^{21,35}

Two types of computations are performed with the objectives of (1) assessing how well the directional unit vectors fit a predefined lattice geometry (*constrained fit method*), and (2) identifying the common intrinsic coordination directions assumed by the clusters (*optimal superimposition*) as a function of the type of the central residue of the cluster. In each case, the central residue defines the origin of a spherical reference system and the tips of the directional vectors end at the surface of a unit sphere where the most highly populated regions indicate the preferred coordination angles. These regions are characterized by two spherical angles (θ : polar and ϕ : azimuthal). In the first version of *optimal superimposition*,²¹ clusters were selected and rotated randomly at each Monte Carlo (MC) step. The resulting mean deviation between all (N) clusters taken pairwise was computed as

$$\langle \epsilon \rangle = \sum_i \sum_j \epsilon_{ij} / [(N(N-1)/2)] \quad (1)$$

where $\epsilon_{ij} = \sum s_k / \min(m_i, m_j)$ is the deviation between clusters i and j , s_k is the distance between the tips of their closest unit vectors, m_i is the coordination number of cluster i , and $\min(m_i, m_j)$ designates the minimum of the two coordination numbers. If $\langle \epsilon \rangle$ decreased compared to its previous value, the rotation was accepted and vice versa. *Constrained fit*, on the other hand, requires the superimposition of the (k) directional unit vectors of the cluster with those of the target lattice, with the mean deviation being found in this case from $\langle \epsilon \rangle_{\text{cons}} = \sum_k \epsilon_k / N$. A total of 3×10^6 MC steps for each of the 20 types of amino acids positioned at the center of the examined clusters was found to yield reproducible distributions. Sets of 1,000 coordination clusters were found to be large enough to obtain statistically accurate results, and small enough to be computationally tractable. The convergence of the results took ~ 50 h (real time) for 1,000 clusters on a R10,000 SGI workstation, and grew exponentially with increasing numbers of clusters, and with increasing coordination number. Because of the serious limitations imposed by this “brute force” computational method, we have devised here a different theoretical approach that substantially reduces the required computer time for calculations.

First we use the symmetry of the fcc lattice to limit the number of possible sets of directional unit vectors on the lattice. If we choose the coordinate system such that four perpendicular directions of the lattice are oriented along the xy -plane, the lattice twelve directional unit vectors are:

$$\begin{aligned} \mathbf{e}_1 &= (1, 0, 0) \\ \mathbf{e}_2 &= (-1, 0, 0) \\ \mathbf{e}_3 &= (0, 1, 0) \\ \mathbf{e}_4 &= (0, -1, 0) \\ \mathbf{e}_5 &= \left(\frac{1}{2}, \frac{1}{2}, \frac{\sqrt{2}}{2} \right) \\ \mathbf{e}_6 &= \left(-\frac{1}{2}, \frac{1}{2}, \frac{\sqrt{2}}{2} \right) \\ \mathbf{e}_7 &= \left(\frac{1}{2}, -\frac{1}{2}, \frac{\sqrt{2}}{2} \right) \\ \mathbf{e}_8 &= \left(-\frac{1}{2}, -\frac{1}{2}, \frac{\sqrt{2}}{2} \right) \\ \mathbf{e}_9 &= \left(\frac{1}{2}, \frac{1}{2}, -\frac{\sqrt{2}}{2} \right) \\ \mathbf{e}_{10} &= \left(-\frac{1}{2}, \frac{1}{2}, -\frac{\sqrt{2}}{2} \right) \\ \mathbf{e}_{11} &= \left(\frac{1}{2}, -\frac{1}{2}, -\frac{\sqrt{2}}{2} \right) \\ \mathbf{e}_{12} &= \left(-\frac{1}{2}, -\frac{1}{2}, -\frac{\sqrt{2}}{2} \right) \end{aligned} \quad (2)$$

The problem of using the symmetry of the system to reduce the total number of combinatorial possibilities was studied earlier by us in the generation of Hamiltonian paths within rectangles (in 2D) or parallelepipeds (in 3D).³⁶ Various symmetries are represented by allowed permuta-

tions and/or changes of signs of coordinates of vertices of rectangle or parallelepiped. In the case of a cube, there are six possible permutations of the coordinates x, y, z , and eight possible combinations of signs, which leads to a total number of 48 symmetries. The symmetry of the set of 12 directional vectors of the fcc lattice differs from that of the cube since the directional vectors $(0, 0, 1)$ and $(0, 0, -1)$ along the z -axis do not exist. Because of this, the permutations involving the z -coordinate are not allowed and the total number of symmetries is reduced by a factor of 3 (to 16), similar to the case of a parallelepiped with a square base.³⁶ A more even distribution of directional vectors might be obtained by using directional vectors of regular polyhedra, especially those of Platonic solids, such as icosahedron or dodecahedron. An icosahedron has the same coordination number 12 as the fcc lattice and is built of 20 regular triangles, with five triangles meeting at each of 12 vertices. However, the icosahedra cannot completely fill the space, and because of this, the fcc packing is more efficient. Icosahedral ordering is frequently observed in amorphous materials such as glasses and quasicrystals.³⁷ A dodecahedron is built of 12 regular pentagons, with three pentagons meeting at each of its 20 vertices, and its coordination number (20) is too high to be selected by amino acids in protein interiors.

The sixteen symmetric transformations for the directional vectors of the fcc lattice can be found from the combinatorial possibilities $(\pm x, \pm y, \pm z) \rightarrow (\pm y, \pm x, \pm z)$ for superimposing pairs of structures. Each of these symmetries transforms a given directional vector into another directional vector (or itself) from the set of twelve vectors. For example, the symmetry operation $(x, y, z) \rightarrow (-y, x, -z)$ transforms the set of vectors $(e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9, e_{10}, e_{11}, e_{12})$ defined in eq. 2 into $(e_4, e_3, e_2, e_1, e_{10}, e_9, e_{12}, e_{11}, e_6, e_8, e_5, e_7)$. The symmetry of the system was then used to reduce the number of possible sets of directional vectors on the fcc lattice. For example, the total number of $\binom{12}{2} = 66$ combinations of pairs of directional vectors can be reduced to 7, namely (e_1, e_2) , (e_1, e_3) , (e_1, e_5) , (e_1, e_6) , (e_5, e_6) , (e_5, e_8) and (e_5, e_{12}) since all other combinations can be reduced to those irreducible cases by the symmetry operations. We wrote an algorithm that identifies and stores all irreducible n -tuplets of directional vectors for a given number n ($1 < n < 12$) of fcc directional vectors. Those n -tuplets were then used in the orientational alignment of the clusters of residues. Instead of using the MC method for finding the best fit of clusters to the directional vectors of the fcc lattice, we have used a method for finding the best rotation to relate the two sets of vectors. The problem was studied by Kabsch in the late 1970s.^{38,39} If \mathbf{a}_n and \mathbf{b}_n ($n = 1, 2, \dots, N$) denote two sets of vectors and w_n denotes weights corresponding to each pair, then we find a rotation \mathbf{U} that aligns the set \mathbf{a}_n with the set \mathbf{b}_n and minimizes the error defined as

$$E = \sum_{i=1}^n w_i (\mathbf{U}\mathbf{a}_i - \mathbf{b}_i)^2 \quad (3)$$

Kabsch's original solution of the problem in terms of Lagrange multipliers,³⁸ led, however, sometimes to improper rotations, and was later modified to avoid ambiguities. The ambiguities in this method are intrinsically associated with the effect of the order of applying Euler angle rotations. A more elegant and powerful approach to this problem is the use of quaternions,⁴⁰ which yield unique parameterization of rotations and allows additionally for reflections and inversions, operations not realizable by pure rotations. A quaternion is a quartet (4-tuple) of real numbers, which might be viewed as a generalization of complex numbers

$$q = (\alpha, x\mathbf{i}, y\mathbf{j}, z\mathbf{k}) \quad (4)$$

with the unit vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$ obeying the following multiplication rules:

$$\begin{aligned} \mathbf{i} \cdot \mathbf{i} &= \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = -1 \\ \mathbf{i} \cdot \mathbf{j} &= -\mathbf{j} \cdot \mathbf{i} = \mathbf{k} \end{aligned} \quad (5)$$

and with cyclic permutations defining the rest of the algebra. The conjugate of a quaternion is $q^\dagger = (\alpha, -x\mathbf{i}, -y\mathbf{j}, -z\mathbf{k})$ and the norm $|q|^2 = q^\dagger q$. Any point (x, y, z) in the Cartesian coordinate system can be represented by a quaternion as $q = (0, x\mathbf{i}, y\mathbf{j}, z\mathbf{k})$. Quaternions with unit norm represent rotation operations in three dimensions. The quaternion

$$q = \left(\cos \frac{\phi}{2}, \sin \frac{\phi}{2} (n_x \mathbf{i}, n_y \mathbf{j}, n_z \mathbf{k}) \right) \quad (6)$$

can be viewed as a rotation by an angle of ϕ about the normalized axis (n_x, n_y, n_z) . This rotation changes the coordinates of a point a represented by the quaternion $(0, a_x \mathbf{i}, a_y \mathbf{j}, a_z \mathbf{k})$ to

$$a' = q^\dagger a q \quad (7)$$

If \mathbf{A} and \mathbf{B} are matrices built from two sets of vectors \mathbf{a}_n and \mathbf{b}_n ($n = 1, 2, \dots, n$)

$$\mathbf{A} = \begin{bmatrix} a_{1x} & a_{1y} & a_{1z} \\ a_{2x} & a_{2y} & a_{2z} \\ \dots & \dots & \dots \\ a_{nx} & a_{ny} & a_{nz} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_{1x} & b_{1y} & b_{1z} \\ b_{2x} & b_{2y} & b_{2z} \\ \dots & \dots & \dots \\ b_{nx} & b_{ny} & b_{nz} \end{bmatrix} \quad (8)$$

then the minimization of the square difference $E = |q^\dagger \mathbf{B} q - \mathbf{A}|^2$ between \mathbf{A} and \mathbf{B} subject to the constraint $q^\dagger q = 1$ corresponds to finding a quaternion (and the associated rotation) that maximizes the overlap $q^\dagger \mathbf{B} q \mathbf{A} + \mathbf{A}^\dagger q^\dagger \mathbf{B} q$. The problem is thus reduced to a matrix diagonalization and finding the eigenvector associated with the largest eigenvalue. The left rotation that superimposes (in the least mean square meaning) the set \mathbf{B} over \mathbf{A} is given in terms of the quaternion components as⁴¹

$$\mathbf{U} = \begin{bmatrix} \alpha^2 + x^2 - y^2 - z^2 & 2(xy + \alpha z) & 2(zx - \alpha y) \\ 2(xy - \alpha z) & \alpha^2 - x^2 + y^2 - z^2 & 2(yz + \alpha x) \\ 2(zx + \alpha y) & 2(yz - \alpha x) & \alpha^2 - x^2 - y^2 + z^2 \end{bmatrix} \quad (9)$$

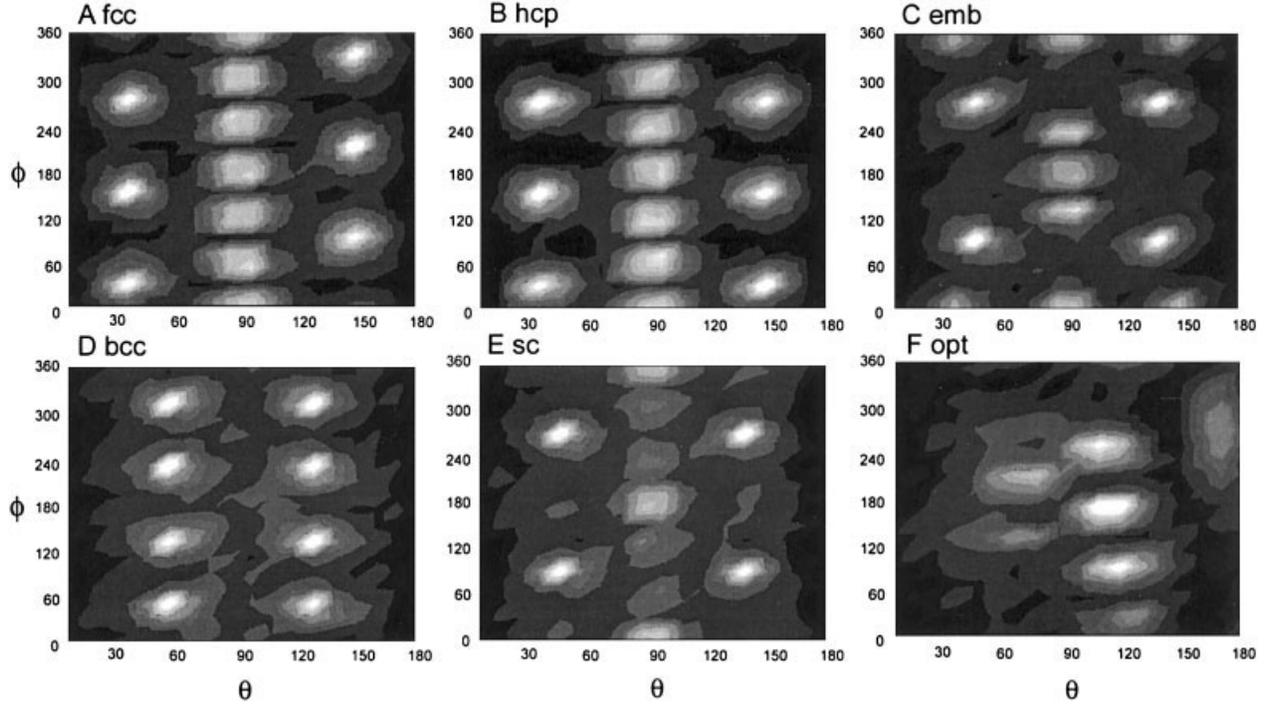


Fig. 3. Distributions obtained from constrained calculations: (A) face-centered cubic (fcc) lattice, (B) hexagonal closest packed (hcp) lattice, (C) tetrahedral embedded in simple cubic (emb) lattice, (D) body-centered cubic (bcc) lattice, and (E) simple cubic (sc) lattice. The coordination orientations are $[(\theta^\circ, \phi^\circ)] = (35, 30), (35, 150), (35, 270), (90, 0), (90, 60), (90, 120), (90, 180), (90, 240), (90, 300), (145, 90), (145, 210), (145, 330)$ in fcc; same except for the last three replaced by $(145, 30), (145, 150), (145, 270)$ in hcp; $(35, 0), (45, 90), (45, 270), (90, 0), (90, 125), (90, 180), (90, 235), (145, 0), (135, 90), (135, 270)$ in emb; $(55, 45), (55, 135), (55, 225), (55, 315), (125, 45), (125, 135), (125, 225), (125, 315)$ in bcc, and $(45, 90), (45, 270), (90, 0), (90, 180), (135, 90), (135, 270)$ in sc. (F) The results from the optimal (opt) superimposition of the clusters.

We have used the algorithm QTRFIT developed by David J. Heisterberg⁴¹ and based on the quaternion method to find the best superimposition of two sets of unit direction vectors. Each coordination cluster of n directional unit vectors was superimposed with all (unrelated by symmetries) n -tuplets of directional unit vectors of the fcc lattice. In this way, the best possible superimpositions in the constrained fit method were found. The QTRFIT algorithm was also used in the optimal superimposition method to find the best pairwise rotational superimpositions of coordination clusters. This new computational scheme significantly reduced the computer time required for calculations.

Order parameters were found from the squared cosine of the angle $\Delta\alpha$ between the fcc lattice directions and the closest directional vectors (of clusters) after rigid body reorientation of the optimally superimposed clusters to yield the closest match to the fcc lattice geometry. An average value $\langle \cos^2 \Delta\alpha \rangle_{\text{cluster}} = \sum_k \cos^2 \Delta\alpha_k / m$ was found for each cluster of m residues, where the summation is over all directional unit vectors in that cluster $1 \leq k \leq m$, and $\langle \cos^2 \Delta\alpha \rangle_{\text{cluster}}$ was further averaged over all clusters in a given subset to find the order parameter $S = \langle 3/2 \langle \cos^2 \Delta\alpha \rangle_{\text{cluster}} - 1/2 \rangle$ corresponding to the examined subset. S gives a measure of the extent of order with respect to the fcc lattice. In general S varies in the range $0 \leq S \leq 1$; the upper limit of $S = 1$ refers to the perfect alignment along the lattice directions, i.e., $\langle \cos^2 \Delta\alpha \rangle = 1$, and $S = 0$ if the

coordination vectors are *uncorrelated* with the lattice directional vectors, i.e., $\langle \cos^2 \Delta\alpha \rangle = \frac{1}{3}$ (random orientation value).

RESULTS AND DISCUSSION

Results From Constrained Fit Method

In Figure 3(A–E), the probability distributions of coordination angles obtained from constrained fit calculations for the respective target lattice directions—fcc, hexagonal closest packed (hcp), simple cubic embedded to tetrahedral (emb), body-centered cubic (bcc), and simple cubic (sc)—are depicted. The clear sites are the most densely populated regions. The excellent agreement between these peaks and the coordination sites of the target lattices shows that the packing architecture can be well represented by various lattice geometries, as elaborated in a detailed study.²⁴ This does not, however, imply that the clusters, themselves, are optimally superimposed onto one another. The point is that the coordination number of the target lattice (z) is usually larger than that (m) of the cluster so that the best fitting subset of m clusters out of the total set of $z!/[m!(z-m)!]$ combinations can be adopted for each superposition. Different clusters, therefore, select different subsets of directional vectors amongst the z accessible choices, and this freedom results in a relatively poor superimposition ($\langle \epsilon \rangle = 0.56–0.60$) among the clusters, themselves. Optimal superimposition of clusters is pre-

sented next. A lower $\langle \epsilon \rangle$ value (0.39) among clusters is found in this case.

Optimal Superimposition of Clusters: Generic Characterization of All Amino Acids

Figure 3(F) shows the result from the *optimal superimposition* of orientation clusters, one to another, irrespective of amino acid type or coordination number. This represents the intrinsic orientation lattice directly observable for proteins. The number of peaks in Figure 3(F) is significantly lower than the coordination numbers of the lattices targeted in the constrained fit calculations, except for the sc lattice. The number of peaks reflects the average coordination number (~ 6.5) in the databank clusters. The most populated coordination angles are:

	1	2	3	4	5	6	7
$\theta(^{\circ})$	110	105	70	65	115	165	120
$\phi(^{\circ})$	170	250	210	130	90	270	20
$P(\theta, \phi)$	0.15	0.10	0.08	0.02	0.10	0.13	0.05

The last row designates the fraction of residues located within a 20° solid angle deviation from each direction. The sum of these probabilities is 0.63, i.e., almost 2/3 of residues occupy these states, while the remainder samples any other suitable positions in space.

Interestingly, the directional vectors are not uniformly distributed in space, but closely clustered to cover only a portion of the coordination sphere. The remaining empty (or sparsely occupied) regions can be anticipated to be those allocated to surrounding solvent.

Optimal Superimposition for Specific Amino Acids at Centers of Clusters

Figure 4 displays the same type of results [as Fig. 3(F)], but for each type of individual amino acid occupying the central position of the examined clusters. The clusters have been grouped into 20 subsets according to the identity of the central residue, and the clusters in each subset have been optimally superimposed using the optimal superimposition algorithm for each subset, independently.

Recurrent orientation coordination patterns with slight variations are detected for different amino acids in Figure 4. An additional eighth coordination state emerges in some cases. Table I lists these residue-specific coordination states. It is observed that (1) the residue-specificity is relatively weak, the coordination directions being preserved with small deviations in coordination angles, and (2) not all coordination states are occupied in the neighborhood of all amino acids. Apparently, residues near a central amino acid usually select sites from amongst these eight most probable directions, depending on the type of amino acid.

The two rows preceding the last one in Table I list the mean values for the directional unit vectors characterizing the most frequently occupied coordination sites. The first of these simply reproduces the seven sites already identified for all clusters [Figure 3(F)] independent of residue type, along with the eighth site preferentially occupied for several specific residues, mostly hydrophobic and glycine.

And the second is another representation of the *same* set of directional vectors obtained after their rigid-body rotation so as to clarify their correspondence to the actual fcc lattice directional vectors listed in the last row.

Coordination of Core Residues

Our analysis shows that a substantial portion of the space near the central examined residue is either unoccupied, or irregularly and weakly populated, and it was anticipated that this feature is indicative of the solvent-exposed regions. To verify this conjecture, subsets of clusters composed of $m = 10$ or more residues have been considered. These are evidently densely packed clusters, and could be viewed as reflecting the behavior of *core* residues. Their optimal superimposition yields the distribution of coordination angles displayed in Figure 5(A). There are now more peaks, and these are more or less uniformly distributed in space. The directional vectors characterizing the core coordination clusters (after rigid-body comparison to facilitate the comparison with fcc directional vectors) are:

	1	2	3	4	5	6	7	8	9	10
$\theta(^{\circ})$	45	45	45	95	105	60	100	85	105	140
$\phi(^{\circ})$	40	180	280	360	60	100	140	240	300	220

The distributions of coordination directions for three special cases, Ala, Cys, and Gly, are presented in Figure 5(B–D). See also Table II for the corresponding coordination states. Comparison of Figure 5(A) and (B) shows that the geometry for “all” clusters is well represented by the subset “alanine.” Figure 5(C) and (D), on the other hand, exhibit distinctive features. Glycine samples an additional eleventh state, consistent with its smaller size and presumably due to the fact that it lacks a sidechain. Cys residues are special because of their tendency to form disulfide bridges. In the last row of Table II are listed the corresponding directional vectors for the fcc lattice for comparison, which confirms that the optimal geometry in the core closely resembles that of fcc packing with two empty sites.

The observed *incomplete, distorted fcc packing* can be rationalized after a closer examination. First, there are two unoccupied sites in core clusters, because this particular subset of clusters is dominated by clusters of $m = 10$ coordinating residues. Calculations repeated for clusters of even higher density ($m \geq 12$) indeed verified that the remaining two unoccupied sites also become filled upon optimal superimposition of such clusters. Table III summarizes the results. Results are reported as the angular difference between the observed coordination angles and the nearest fcc lattice directional vectors for different subsets of residue clusters. The preference for site 10 over the unoccupied sites 11 and 12 could be associated with the relatively large (100° instead of 60°) azimuthal angle difference between the two nearest hexagonal sites 7 and 8. The relatively small (40°) azimuthal angle difference between site 6 and its nearest neighbors (5 and 7) is apparently accommodated by a polar angle distortion (60° instead of 90°). Interestingly, even the eighth site observed

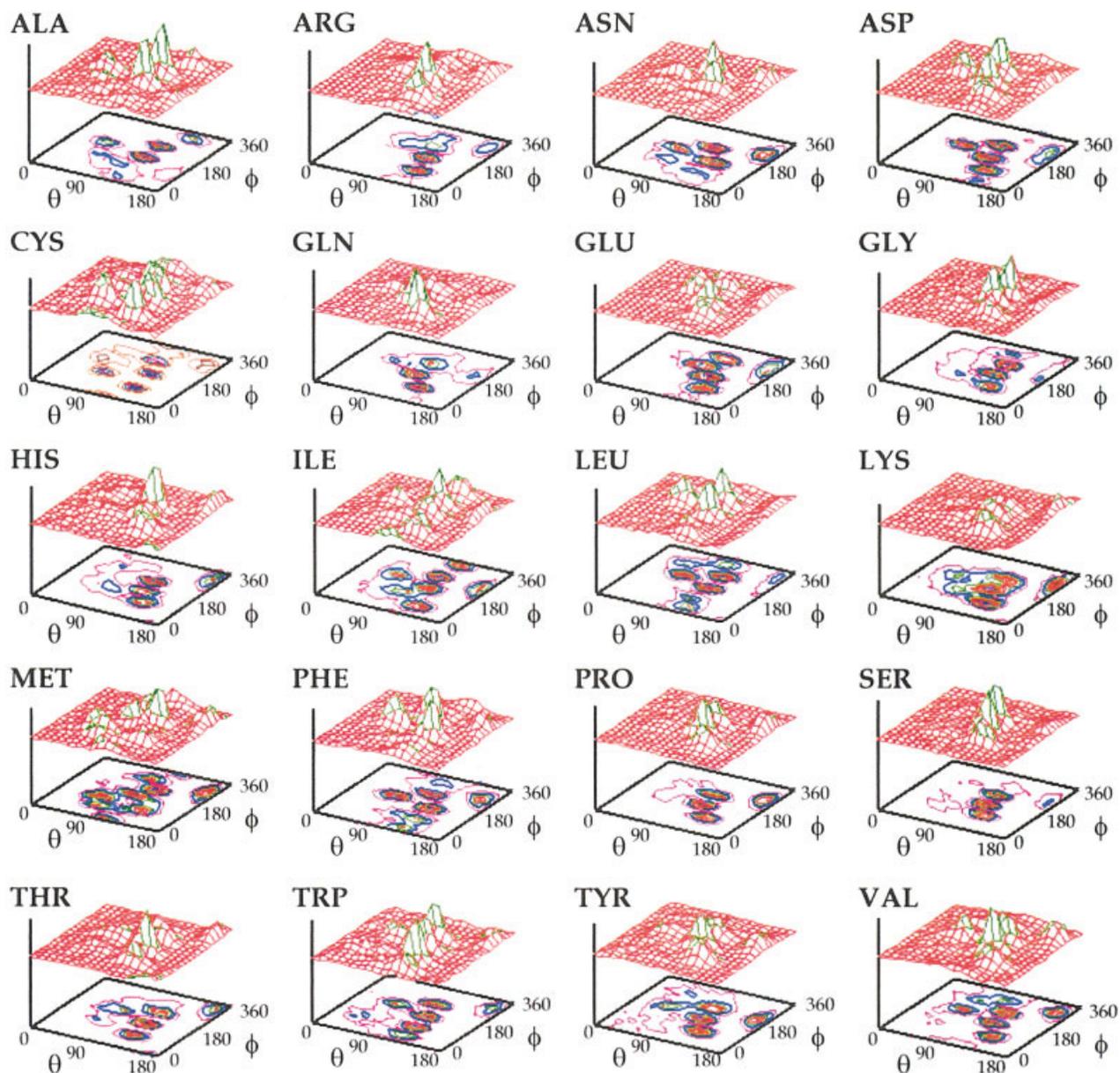


Fig. 4. Coordination orientation geometry around specific amino acids, shown as probability distribution surfaces and projected countours as a function of the spherical angles θ and ϕ . See Table I for the list of coordination loci (peaks in the distributions).

for a number of specific residues (Table I) conforms to one of the directional vectors (7) of the fcc geometry.

Finally, surface residues (subset of clusters having coordination numbers of 4 or less) have been examined. As expected, fewer peaks are observed in this case, but these can be easily attributed to four directions of the fcc geometry. See the first row in Table III. The fraction of residues occupying these four coordination directions near solvent-exposed residues is calculated to be 0.40, again allowing for 20° deviations with respect to the central directional vectors. The deviations from standard fcc directions are found to be even smaller in this case.

The twelve directional unit vectors identified for the most densely packed core residues ($m \geq 12$) exhibit an

occupancy ratio of 0.76. Therefore, approximately $\frac{3}{4}$ of residues occupy these “regular” coordination directions, while the remainder is “disordered.” We see that a larger fraction of residues occupy the “regular” coordination directions as one examines increasingly denser clusters. See the last column in Table III. This provides evidence that the origin of much of the observed regularity originates in an excluded volume effect.

Figure 6 provides a summary of the results obtained by optimal superimposition of directional clusters having different coordination numbers. Figure 6(A) displays the most probable coordination orientations for surface residues ($m \leq 4$). The sites are assigned identifying numbers consistent with those in Table III. Figure 6(B) shows the

TABLE I. Coordination States for Different Types of Residues in Proteins[†]

		1	2	3	4	5	6	7	8
ALA	θ, ϕ	115, 170	115, 250	50, 210	—	90, 110	135, 330	115, 30	
ARG	θ, ϕ	115, 170	105, 240	75, 230	—	125, 90	160, 300	115, 20	
ASN	θ, ϕ	120, 170	95, 250	65, 195	65, 130	120, 90	155, 260	—	
ASP	θ, ϕ	105, 170	100, 250	65, 210	—	120, 90	165, 260	130, 20	
CYS	θ, ϕ	120, 170	100, 255	45, 210	70, 130	120, 85	155, 300	95, 0	
GLN	θ, ϕ	95, 170	95, 240	55, 210	—	120, 90	160, 230	135, 20	
GLU	θ, ϕ	115, 170	105, 280	85, 225	—	125, 90	165, 270	—	
GLY	θ, ϕ	115, 170	115, 245	—	75, 130	125, 105	165, 220	—	100, 310
HIS	θ, ϕ	110, 160	110, 230	—	70, 140	125, 95	165, 310	—	
ILE	θ, ϕ	115, 170	105, 250	60, 210	95, 130	130, 70	165, 240	—	110, 330
LEU	θ, ϕ	115, 170	105, 250	65, 210	—	105, 90	165, 300	120, 40	95, 320
LYS	θ, ϕ	105, 170	105, 250	55, 210	—	120, 90	165, 290	—	
MET	θ, ϕ	95, 190	110, 245	—	60, 130	125, 110	165, 280	95, 10	85, 290
PHE	θ, ϕ	120, 170	100, 250	65, 190	—	120, 90	155, 270	120, 25	95, 320
PRO	θ, ϕ	105, 170	105, 230	—	—	125, 90	165, 270	—	
SER	θ, ϕ	105, 170	100, 230	—	60, 110	115, 110	165, 260	—	
THR	θ, ϕ	115, 170	110, 250	75, 210	60, 130	120, 90	165, 320	—	
TRP	θ, ϕ	115, 170	100, 250	60, 210	—	115, 110	165, 310	115, 45	
TYR	θ, ϕ	105, 170	100, 240	50, 195	—	120, 90	165, 230	—	
VAL	θ, ϕ	110, 170	110, 235	60, 210	—	125, 90	155, 290	—	90, 300
AVG ^a	θ, ϕ	110, 170	105, 250	70, 210	65, 130	115, 90	165, 270	120, 20	100, 310
AVG ^b	θ, ϕ	40, 10	35, 200	45, 285	95, 350	105, 50	55, 115	120, 115	90, 180
FCC ^c	θ, ϕ	35, 30	35, 150	35, 270	90, 360	90, 60	90, 120	145, 90	90, 180

[†]The coordination angles (in degrees) refer to the centers of the peaks observed in Figure 3. The absolute values of the angles can change depending on the reference frame, but the relative values of the individual amino acids remain unchanged.

^aIncludes the seven peaks identified for all amino acids [Fig. 3 (F)], and the site 8 visited by specific residues.

^bSame results as in the preceding row, except for a rigid body rotation of the cluster to match the fcc lattice geometry.

^cRemaining unoccupied fcc sites are (90, 240), (90, 300), (145, 210), and (145, 330).

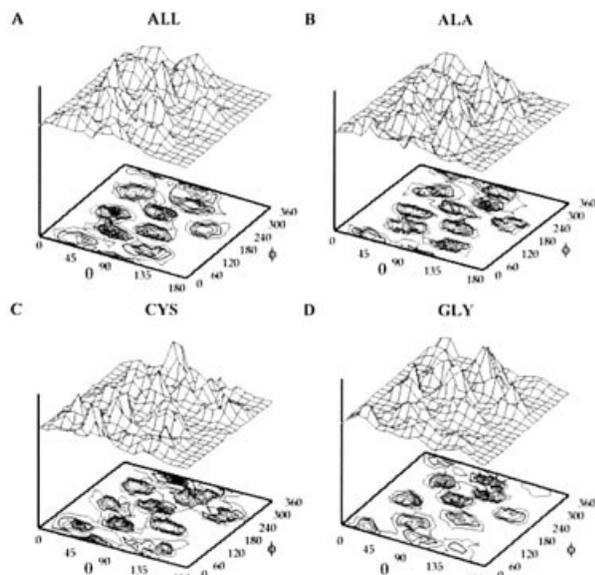


Fig. 5. Coordination orientation geometry around a central residue for high-density clusters (A) around all types of central residues, (B) around Ala, (C) around Cys, and (D) around Gly.

results for all residues, obtained by a rigid body rotation of the map in Figure 3(F). Figure 6(C) and (D) describes the coordination for core residues having $m \geq 10$ and $m \geq 12$, respectively. As can be seen in Figure 6, sites are gradually filled as the coordination number increases. It is important

to note that the occupancy is not sparse or staggered, but instead relatively close sites are filled first so that the density, excluding the solvent space, is approximately constant. However, there is some increasing sharpness of the peaks as m increases, again reflecting an excluded volume effect.

The Extent of Order in the Coordination of Residues

The degree of ordering is assessed by calculating the orientation functions (S), also called order parameters, for the four cases. The nearest fcc directional unit vectors after optimal superimposition of the clusters are considered for each coordination unit vector in order to evaluate S . S is found to be 0.82 for the surface clusters ($3 \leq m \leq 4$), 0.82 for the clusters all over the protein ($3 \leq m \leq 14$), 0.83 for the clusters at the core with coordination number $m \geq 10$, 0.81 for the clusters at the core with coordination number $m \geq 12$. The order parameters provide a quantitative measure of the intrinsic order in folded structures when examined on a coarse-grained scale. While proteins can enjoy a higher disorder at the atomic scale, the tendency to uniformly fill the protein interior in conformity with a regular high-density packing geometry leads to these relatively high order parameter values on a coarse-grained scale. Interestingly, the same level of order relative to an fcc packing is observed in the protein core and near the surface. This similar behavior can, however, result from two different causes: The high coordination

TABLE II. Coordination Directions About Specific Residues in the Protein Core

		1	2	3	4	5	6	7	8	9	10	11
ALL	θ ($^\circ$)	45	45	45	95	105	60	100	85	105	140	
	ϕ ($^\circ$)	40	180	280	360	60	100	140	240	300	220	
ALA	θ ($^\circ$)	40	50	50	95	115	65	105	85	105	145	
	ϕ ($^\circ$)	20	170	280	340	60	100	160	220	280	220	
CYS	θ ($^\circ$)	35	40	45	70	130	70	115	85	100	150	
	ϕ ($^\circ$)	60	180	290	360	60	120	150	220	320	250	
GLY	θ ($^\circ$)	30	40	50	90	85	65	105	80	105	145	130
	ϕ ($^\circ$)	30	180	280	340	40	90	160	220	280	230	45
FCC	θ ($^\circ$)	35	35	35	90	90	90	90	90	90	145	145
	ϕ ($^\circ$)	30	150	270	360	60	120	180	240	300	210	90

TABLE III. Deviations of Coordination Angles of Residues From the Face-Centered-Cubic Lattice Directions

Coordination number		Coordination states ($^\circ$)												P_{tot}^a	
		1	2	3	4	5	6	7	8	9	10	11	12		
$3 \leq m \leq 4$	$\Delta\theta$	5	+10			+5	0								0.40
	surface $\Delta\phi$	0	+20			-10	-10								
$3 \leq m \leq 14$	$\Delta\theta$	+5	0	+10	+5	+15	-35	0 ^b					-25	0.63	
	all $\Delta\phi$	-20	+50	+15	-10	-10	-5	0 ^b					+25		
$m \geq 10$ core	$\Delta\theta$	+10	+10	+10	+5	+15	-30	+10	-5	+15	-5			0.65	
	$\Delta\phi$	+10	+30	+10	0	0	-20	-40	0	0	+10				
$m \geq 12$	$\Delta\theta$	+10	-10	+15	-20	+10	-15	-10	-15	+15	-5	0	-15	0.76	
	$\Delta\phi$	+30	+20	+10	-20	+20	0	-20	-20	-40	-10	0	+30		

^aProbability of the complete set of coordination states, found from the fraction of residues occupying the listed set of directions within 20° deviation about the central directional vectors.

^bObserved for a subset of specific amino acids (see Table 1).

clusters ought to be optimally packed because of competition for space, and should approximate as expected a high coordination lattice geometry, which explains the selection of fcc lattice on a coarse-grained scale. As the latter case of solvent-exposed residues, these residues have fewer intramolecular contacts, and the fewer coordination unit vectors can be matched to a good approximation with one of the 12 directions accessible in the fcc lattice.

CONCLUSIONS

Packing Architecture of Residues in Folded Proteins Can Be Modeled With a Variety of Regular Geometries, Although Not All Coordination Directions Are Filled

Our analysis demonstrates that the coordination geometry can be fit to a variety of regular geometries, with approximately equal fidelity. See Figure 3(A–E). The point is that clusters containing $m < 7$ residues, which may be viewed as a low-to-intermediate density state for folded proteins, dominate the observed behavior. And such clusters can be suitably allocated to 6–7 sites out of the z choices available in the target geometry. This leaves about 5–6 empty sites compared to the packing in solid-like high coordination lattice. The variation in the number of coordination numbers ($3 \leq m \leq 14$) or unfilled sites in the close neighborhood of residues, along with the tendency to populate close coordination sites to achieve a tight packing, is consistent with the results from a recent rigorous Delaunay triangulation, which shows that proteins look more like liquids and glasses by the criterion of their free

volume distributions, although they resemble crystals on the basis of their average density.⁴²

Optimal Superimposition of Clusters Conforms to Partially Filled and Distorted Face-Centered Cubic Packing

The optimal superimposition of the clusters irrespective of any predefined target lattice [Figs. 3(F) and 5(A) for *all* and *core* residues, respectively] reveal a tendency to occupy fcc-like orientations, which is also supported by maps in Figure 6. The optimal architecture is a distorted, incomplete fcc packing that is gradually filled as the coordination number increases. See also Table III. A cubic close-packed geometry may, thus, be viewed as a generic packing architecture in protein interiors. It is important to note that in the clusters having relatively low coordination numbers, the coordination directions are closely clustered in space, i.e., the coordinating residues do not fill sparsely the coordination sphere in the neighborhood of the central residues, but are closely grouped to occupy directions approximating those of fcc packing. See, for example, Figure 6(A). Therefore, the uniform (high) densities of residues are still maintained even in solvent-exposed regions, with the only difference being that not all orientations are occupied, and there are slight distortions in the unit direction vectors.

Alternatively stated, the coordination numbers are different for surface and core residues. However, if one considers the subspace allocated to residues, only, the density is uniform. The same feature has been pointed out by Tsai et al.⁴³ based on a different knowledge-based approach.

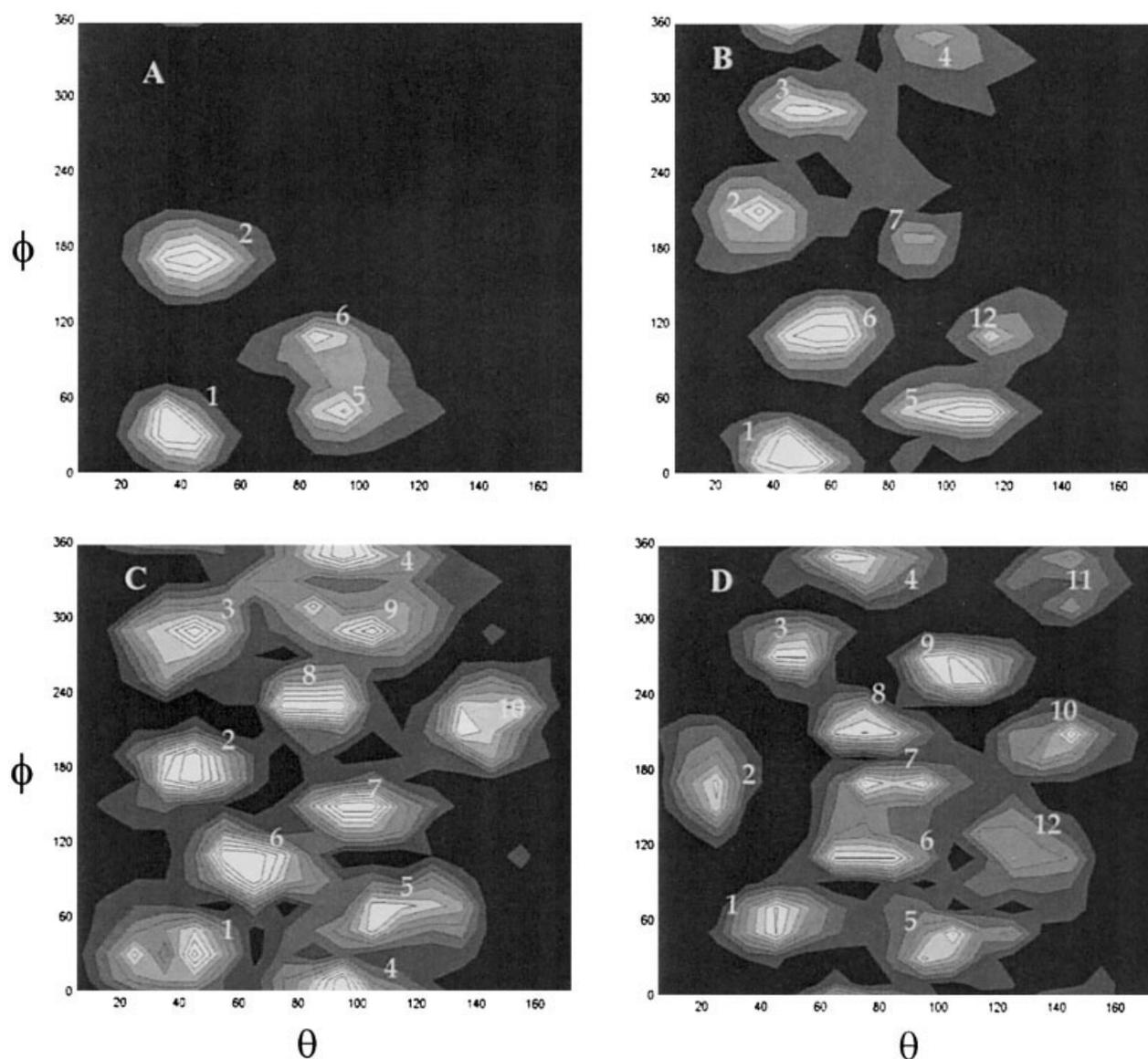


Fig. 6. The coordination orientation geometries for different packing densities: (A) $3 \leq m \leq 4$, (B) $3 \leq m \leq 14$, (C) $m \geq 10$, and (D) $m \geq 12$, where m is the coordination number. The labels in the contours refer to the angular positions given in Table III.

Fcc Orientational Packing Could Result From the Drive for Maximizing Packing Density Rather Than Specific Directional Preferences

The fcc geometry has been recently shown^{28,29} to be the closest packing geometry for identical spheres. The intrinsic tendency of residues (when examined at the coarse-grained level of a single-site-per-residue) to assume such a packing architecture, can equally originate in the drive for maximizing packing density. The fcc directions are, therefore, populated not because of a specific preference, but due to the fact that residues ought to be closely packed, and fcc geometry achieves a close packing. The peaks emerging at the fcc sites can, in fact, be viewed as the optimal *discrete* representation of a *uniform* sampling of the coordination space, as suggested by our recent calculations performed for a hypothetical 2-dimensional (square)

lattice coordination geometry.²¹ Therein, we examined clusters of four coordination vectors each sampling a distinct quadrant on a square lattice. While the vectors were allowed to select any orientation within the allocated quadrant (such that the complete coordination space was uniformly sampled), their optimal superimposition yield four peaks separated by 90° angles. The appearance of the peaks in the angular distributions after optimal superimposition does not, therefore, reflect the preference for particular directions, but the uniform sampling of different space grids. The present observation of 12 peaks at approximately fcc directions could equally be a manifestation of the uniform sampling (packing) of the coordination space on a coarse-grained (single-site-per-residue) scale.

The packing density in proteins can even exceed that (0.7405) of fcc geometry.⁴⁴ In fact, it is possible to exceed

this upper limit for packing of identical spheres by considering particles differing in size. For example, an equimolar mixture of two particles with respective radii in the ratio of 1.00:0.414 has a packing density of 0.7931 when the bigger spheres are placed at the fcc lattice vertices, and the smaller spheres are fitted exactly into the empty spaces left between the big spheres. This suggests that the heterogeneous sizes and shapes of amino acids are well-suited for maximizing packing density. Kussell and coworkers studied packing of sidechains in proteins and obtained a set of incorrectly packed decoys using excluded volume interactions.⁴⁵ It would be interesting to examine these decoys with regard to the packing tendencies indicated in the present study. Interestingly, mutational studies of the hydrophobic core of Fyn SH3 domain indicate that efficient packing not only stabilizes structures but accelerates folding.⁴⁶

Residue Packing Architecture Is Highly Versatile

The above analysis demonstrates that almost 2/3 of residues in folded proteins are packed in conformity with a regular (fcc-like) architecture while the remaining 1/3 occupy any suitable position. The fraction of regularly placed residues decreases near solvent-exposed regions, and increases in the core regions. The apparent random positioning of the 1/3 (on average) could be selected to optimize the bonded and non-bonded interactions in a given irregular context, hence the adaptability of tertiary structures to single-site mutations and the ability to bind multiple ligands,⁴⁷ another observation that lends support to the nonspecific uniform packing of residues. We note that Soyer et al.³³ also suggested that residue packing in proteins is like random packings of hard spheres.

The Observed Packings Relate to the Solid-Like Versus Liquid-Like Nature of the Protein Interior and Exterior

Klapper⁴⁸ proposed about three decades ago that molecules may be described as those containing a hard core along with soft interacting surfaces, and that the protein interior is closer to a solid than a liquid. The same feature, that of a liquid-like exterior and an aperiodic solid core, was also suggested by Fraunfelder et al.⁴⁹ and shown by Zhou and coworkers⁵⁰ to be consistent with the Lindemann criterion, which compares the root mean square atomic fluctuation amplitude to the lattice constant a of a crystal. When this ratio is above a certain value, the fluctuations start to damage and destroy the crystal lattice. The present findings suggest that the liquid-like surface is not necessarily associated with a lower packing density of surface residues, but rather with the flexibility (or ductility) induced by the higher fraction (0.60) of residues occupying random, or disordered, positions near solvent-exposed residues. The solid-like core, on the other hand, can be understood both in view of the staggered close packed distribution of residues, and the fact that more than 2/3 of residues occupy well-defined coordination directions.

Packing Architecture Exhibits Weak Residue-Specificity, Consistent With Many Sequences Being Compatible With a Given Structure

The similarities between amino acids illustrated in Figure 4 (see also Table I) reveal a nonspecificity in packing on a coarse-grained scale. This result conforms to the fact that sequences diverge faster than structures, or several sequences map to the same structure (or fold). Designable protein structures are those mapped by a large number of sequences,⁵¹ and the observed generic packing reproduced for all types of amino acids (Fig. 4) is consistent with the designability requirement of folded structures,^{51,52} inasmuch as the regular (or symmetric) coordination architecture can accommodate different types of residues. Behe et al.⁵³ pointed out that packing does not determine the native fold. The weak specificity observed here supports this view. Not surprisingly, inclusion of knowledge-based data about packing preferences is of limited utility in sequence-recognizes-structure protocols.²⁰ On the other hand, knowledge of the generic packing architecture of residue clusters in folded structures, and its correlation with secondary structure, might provide some guidance in reducing the space for conformational search and for the computational prediction of 3D structures.

REFERENCES

- Richards FM, Lim WA. An analysis of packing in the protein folding problem. *Q Rev Biophys* 1993;26:423–498.
- Ponder JW, Richards FM. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol* 1987;193:775–791.
- Munson M, Balasubramanian S, Fleming KG, Nagi AD, O'Brien R, Sturtevant JM, Regan L. What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Protein Sci* 1996;5:1584–1593.
- Raghuathan G, Jernigan RL. Ideal architecture of residue packing and its observation in protein structures. *Protein Sci* 1997;6:2072–2083.
- Richards FM. Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng* 1977;6:151–176.
- Bromberg S, Dill KA. Side-chain entropy and packing in proteins. *Protein Sci* 1994;3:997–1009.
- Bahar I, Jernigan RL. Coordination geometry of nonbonded residues in globular proteins. *Fold Des* 1996;1:357–370.
- Lim WA, Sauer RT. The role of internal packing interactions in determining the structure and stability of a protein. *J Mol Biol* 1991;219:359–376.
- Matthews BW. Studies on protein stability with T4 lysozyme. *Adv Protein Chem* 1995;46:249–278.
- Maritan A, Micheletti C, Trovato A, Banavar JR. Optimal shapes of compact strings. *Nature* 2000;406:287–290.
- Chan HS, Dill KA. Origins of structure in globular proteins. *Proc Natl Acad Sci USA* 1990;87:6388–6392.
- Gregoret LM, Cohen FE. Protein folding. Effect of packing density on chain conformation. *J Mol Biol* 1991;219:109–122.
- Hao MH, Rackovsky S, Liwo A, Pincus MR, Scheraga HA. Effects of compact volume and chain stiffness on the conformations of native proteins. *Proc Natl Acad Sci USA* 1992;89:6614–6618.
- Hunt NG, Gregoret LM, Cohen FE. The origins of protein secondary structure. Effects of packing density and hydrogen bonding studied by a fast conformational search. *J Mol Biol* 1994;241:214–225.
- Yee DP, Chan HS, Havel TF, Dill KA. Does compactness induce secondary structure in proteins? A study of poly-alanine chains computed by distance geometry. *J Mol Biol* 1994;241:557–573.
- Stasiak A, Maddocks JH. Mathematics. Best packing in proteins and DNA. *Nature* 2000;406:251–253.
- Singh RK, Tropsha A, Vaisman II. Delaunay tessellation of

- proteins: four body nearest-neighbor propensities of amino acid residues. *J Comput Biol* 1996;3:213–221.
18. Singh J, Thornton JM. SIRIUS. An automated method for the analysis of the preferred packing arrangements between protein groups. *J Mol Biol* 1990;211:595–615.
 19. Vriend G, Sander C. Quality control of protein models: Directional atomic contact analysis. *J Appl Crystallogr* 1993;26:47–60.
 20. Keskin O, Bahar I. Packing of sidechains in low-resolution models for proteins. *Fold Des* 1998;3:469–479.
 21. Bagci Z, Jernigan RL, Bahar I. Residue packing in proteins: Uniform distribution on a coarse-grained scale. *J Chem Phys* 2002;117:2269–2276.
 22. Jernigan RL, Bahar I. Structure-derived potentials and protein simulations. *Curr Opin Struct Biol* 1996;6:195–209.
 23. Miyazawa S, Jernigan RL. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J Mol Biol* 1996;256:623–644.
 24. Bagci Z, Jernigan RL, Bahar I. Residue coordination in proteins conforms to the closest packing of spheres. *Polymer* 2002;43:451–459.
 25. Godzik A., Kolinski A, Skolnick J, Lattice representations of globular proteins: How good are they? *J Comput Chem* 1993;14:1194–1202.
 26. Kolinski A, Skolnick J, Lattice models of protein folding, dynamics and thermodynamics. R.G. Landes: Austin, TX (Distributed by Chapman & Hall: New York) 1996.
 27. Sikorski A, Kolinski A, Skolnick J, Computer simulations of the properties of the α_2 , α_2C , and α_2D de novo designed helical proteins, *Proteins* 2000;38:17–28.
 28. Cipra B. Mathematics: packing challenge mastered at last. *Science* 1998;281:1267–1267.
 29. Sloane NJA. Kepler's conjecture confirmed. *Nature* 1998;395:435–436.
 30. Chen J, Stites WE. Packing is a key selection factor in the evolution of protein hydrophobic cores. *Biochemistry* 2001;40:15280–15289.
 31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shinyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–242.
 32. Watson HC. The stereochemistry of the protein myoglobin. *Progr Stereochem* 1969;4:299–305.
 33. Soyer A, Chomilier J, Mornon JP, Jullien R, Sadoc JF. Voronoi tessellation reveals the condensed matter character of folded proteins. *Phys Rev Lett* 2000;85:3532–3535.
 34. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 1997;2:173–181.
 35. Jernigan RL, Bahar I. Geometric regularities among bonded and non-bonded residues in proteins. In: Vijayan M, Yathindra N, Kolaskar AS, editors. *Perspectives in structural biology: a volume in honor of G. N. Ramachandran*. Hyderabad: Indian Academy of Sciences, Universities Press; 1999. p 209–225.
 36. Kloczkowski A, Jernigan RL. Computer generation and enumeration of compact self-avoiding walks within simple geometries on lattices. *Comput Theoret Polym Sci* 1997;7:163–173.
 37. Janot C. Quasicrystals. A primer. Oxford University Press. 1992.
 38. Kabsch W. A solution for best rotation to relate two sets of vectors. *Acta Cryst* 1976;A32:922–923.
 39. Kabsch W. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Cryst* 1978;A34:827–828.
 40. Kuipers JB. Quaternions and rotation sequences: a primer with applications to orbits, aerospace, and virtual reality. Princeton: Princeton University Press; 2002.
 41. Heisterberg DJ. QTRFIT algorithm for superimposing two similar rigid molecules, The Ohio Supercomputer Center Ohio State University, Columbus, OH.
 42. Liang J, Dill KA. Are proteins well-packed? *Biophys J* 2001;81:751–766.
 43. Tsai J, Taylor R, Chothia C, Gerstein M. The packing density in proteins: standard radii and volumes. *J Mol Biol* 1999;290:253–266.
 44. Harpaz Y, Gerstein M, Chothia C. Volume changes on protein folding. *Structure* 1994;2:641–649.
 45. Kussell E, Shimada J, Shakhnovich EI. Excluded volume in protein side-chain packing. *J Mol Biol* 2001;311:183–193.
 46. Northey JG, Di Nardo AA, Davidson AR. Hydrophobic core packing in the SH3 domain folding transition state. *Nat Struct Biol* 2002;9:126–130.
 47. Ma B, Shatsky M, Wolfson HJ, Nussinov R. Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein Sci* 2002;11:184–197.
 48. Klapper MH. On the nature of the protein interior. *Biochim Biophys Acta* 1971;229:557–566.
 49. Frauenfelder H, Petsko GA, Tsernoglou D. Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature* 1979;280:558–563.
 50. Zhou Y, Vitkup D, Karplus M. Native proteins are surface-molten solids: application of the Lindemann criterion for the solid versus liquid state. *J Mol Biol* 1999;285:1371–1375.
 51. Li H, Helling R, Tang C, Wingreen N. Emergence of preferred structures in a simple model of protein folding. *Science* 1996;273:666–669.
 52. Wang TR, Miller J, Wingreen NS, Tang C, Dill KA. Symmetry and designability for lattice protein models. *J Chem Phys* 2000;113:8329–8336.
 53. Behe MJ, Lattman EE, Rose GD. The protein-folding problem: the native fold determines packing, but does packing determine the native fold? *Proc Natl Acad Sci USA* 1991;88:4195–4199.