

*Proteins*. Author manuscript; available in PMC 2011 January 11.

Published in final edited form as:

Proteins. 2008 November 15; 73(3): 730–741. doi:10.1002/prot.22092.

# Orientational distributions of contact clusters in proteins closely resemble those of an icosahedron

Yaping Feng<sup>1,2</sup>, Robert L. Jernigan<sup>1,2</sup>, and Andrzej Kloczkowski<sup>1,2,\*</sup>

<sup>1</sup>Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University Ames, IA 50011-0320, USA

<sup>2</sup>L.H.Baker Center for Bioinformatics and Biological Statistics, Iowa State University, Ames IA 50011-3020, USA

#### Abstract

The orientational geometry of residue packing in proteins was studied in the past by superimposing clusters of neighboring residues with several simple lattices.1<sup>,2</sup> In this work, instead of a lattice we use the regular polyhedron, the icosahedron, as the model to describe the orientational distribution of contacts in clusters derived from a high-resolution protein dataset (522 protein structures with high resolution < 1.5Å). We find that the order parameter (orientation function) measuring the angular overlap of directions in coordination clusters with directions of the icosahedron is 0.91, which is a significant improvement in comparison with the value 0.82 for the order parameter with the face-centered cubic (fcc) lattice. Close packing tendencies and patterns of residue packing in proteins is considered in detail and a theoretical description of these packing regularities is proposed.

#### **Keywords**

residue packing in proteins; icosahedron; packing pattern

### INTRODUCTION

Protein packing is an important aspect of structural biology related to many other problems, such as: protein structure design<sup>3,4</sup>, quality evaluation of protein structures <sup>5</sup>, prediction of protein-ligand binding<sup>6,7</sup>, and calculation of the intrinsic compressibility of proteins<sup>8,9</sup>. Many previous studies of packing at the atomic level show that proteins have an exceptionally high packing density in their interior regions8<sup>10</sup> and that side-chains in the protein cores are neatly interlocked<sup>11</sup>Word et al., 1999). The tight packing of the hydrophobic core mainly caused by the tendency for nonpolar residues to aggregate in water has been considered to play a key role in the stability of proteins. <sup>12</sup>Close packing of the hydrophobic core has been indicated to be a key selection factor in evolution from investigations of stabilities and interaction energies of a series of mutants in the major hydrophobic core of staphylococcal nuclease and 42 homologous proteins. <sup>13</sup> The surface parts of proteins are considered to be less tightly packed than the core parts. <sup>14</sup> The protein size also affects the packing: larger proteins are usually packed more loosely than smaller proteins. <sup>15</sup>

<sup>\*</sup>Corresponding author: Phone: (515) 294-7261; Fax:(515) 294-3841; kloczkow@iastate.edu.

Small ranges of torsion angles are allowed for the backbone conformations of proteins because of the restriction imposed by peptide bonds. Ramachandran plots show that dihedral angles in proteins are mainly localized within a few regions of the psi-phi angles corresponding to different secondary structures, which is indicative of the packing regularities of protein backbones. The side chain packing problem is more complicated and the existence of regular and ordered packing of side chains is usually unclear when studied at the atomic level. Conflicting experimental observations and theoretical analysis about random or ordered side chain packing patterns <sup>16</sup>, 17 make the side chain packing problem particularly interesting for a more thorough exploration. Several models have been put forward to study this problem. Richards firstly proposed in 1977 the jigsaw puzzle model to elucidate the side chain packing problem.18 Another completely different packing model of the nuts and bolts in a jar that was described by Bromberg and Dill.16 Raghunathan and Jernigan utilized a lattice model of sphere packing and found that almost all residues conform perfectly to this lattice model when 6.5 Å is used as the cutoff to define nonbonded interacting residues.<sup>2</sup> The face-centered-cubic (fcc) lattice model and several other lattice models have been used to find the side chain packing regularities when proteins are studied at the coarse-grained level.<sup>1</sup>,19

In the present paper, we use same quaternion-based superimposition algorithm (QTRFIT) employed earlier by us for the fcc lattice model<sup>1</sup> to superimpose the unit vector clusters collected from real protein structures with the directional vectors of the icosahedron model to investigate packing patterns, packing regularities, and their relations to the packing density. Several recent studies on packing density motivated us to investigate the icosahedron as a new model for the distribution of directions among closely packed residues. It has been proved that the fcc lattice is the closest packing geometry of equalsized spheres.20.21 However, if ellipsoids are used instead of spherical particles, the random packing density will increase because achieving a higher density relates to having a larger number of degrees of freedom; and ellipsoids have more degrees of freedom than spheres. 22,<sup>23</sup> The irregular shapes of protein side chains imply that each residue resembles more closely an ellipsoid than a sphere. Because of this, we hypothesize that the packing density of proteins may be higher than that in the fcc lattice used in our previous study<sup>1</sup>, and therefore a new model having the possibility of slightly higher packing density is proposed. In this study, we choose the icosahedron as a new model to investigate the protein packing problem on the coarse-grained level. The central sphere of an icosahedron has a higher local packing fraction 0.76 than that of the fcc lattice, which has the same local packing fraction 0.74 for all spheres.<sup>24</sup> The icosahedron is the Platonic solid P<sub>3</sub> with 12 vertices, 30 edges, and 20 equivalent equilateral triangle faces. The regular property of the icosahedron has other advantages in its regularity in angles and even reduces computational complexity. There are a total of 12 directional vectors from the center of icosahedron to its 12 vertices. Each of the vector clusters obtained from the protein dataset 1.5Å522 represents the cluster of unit vectors between the central residues and its neighbors. We use the quaternion-based QTRFIT algorithm to superimpose the set of directional vectors of coordination clusters with the set of directional vectors of the icosahedron model. We observe that the icosahedron model can represent coordination clusters derived from protein structures much better than the fcc lattice model. The superimposition results provide us with extremely valuable information about residue packing patterns and regularities, packing density, etc.

#### **MATERIALS AND METHODS**

#### Selection of dataset

A dataset of 522 protein structures, named here as 1.5Å522 was randomly selected from our larger dataset of 774 structures  $1.5\text{Å}774^{25}$  which we extracted from the Protein Data Bank using the online server PISECES<sup>26</sup> by imposing the following criteria: percentage sequence

identity: 30%, resolution: 1.5 Å or better, R-factor: 0.3, with only X-ray-determined structures included. A total of 110,255 coordination clusters were extracted from the 1.5Å522 dataset, which is nearly 4 times more than the total number of coordination clusters used in our previous study  $^1$ . Protein packing is a complex problem and many experimental data and theoretical analyses are mutually conflicting.  $^{16,17}$  Here we use coarse-grained models to reduce the complexity of the problem while investigating packing regularities in proteins. All residues are represented by their  $C^\beta$  atoms except glycines, which are represented by the  $C^\alpha$  atoms. Figure 2 in our previous paper  $^1$  shows an example of the coordination cluster formed by the central residue (GLY65) and all it spatial neighbors within 6.8Å in myoglobin. Each of the 110,255 coordination clusters studied here is represented by a set of unit vectors pointing from the central residue to its neighbors lying within 6.8Å. We do not differentiate here between bonded and non-bonded neighbors. The reasons for choosing a cutoff distance 6.8Å and for including both bonded and non-bonded neighbors have been discussed in detail in our previous paper  $^1$ .

# Construction of directional vectors for the icosahedron model and the generation of irreducible combinations of m (m≤12) directional unit vectors

The icosahedron is one of the most interesting regular polyhedra and has been widely used in physics, material science, and biological sciences. <sup>24,27-29</sup> It has 12 vertices, 30 edges and 20 equilateral triangle faces with five of them meeting at each of the 12 vertices. If we choose the icosahedron center as the center of the coordinate system and specify the vectors from the center of the icosahedron to each of its 12 vertices to be the unit vector, and then compute the Cartesian coordinates for the 12 directional unit vectors, we obtain the following 12 directional unit vectors:

```
\begin{array}{lll} e_1 &=& (0.894,0,0.447) \\ e_2 &=& (0.276,0.851,0.447) \\ e_3 &=& (-0.724,0.526,0.447) \\ e_4 &=& (-0.724,-0.526,0.447) \\ e_5 &=& (0.276,-0.851,0.447) \\ e_6 &=& (0.724,0.526,-0.447) \\ e_7 &=& (-0.276,0.851,-0.447) \\ e_8 &=& (-0.894,0,-0.447) \\ e_9 &=& (-0.276,-0.851,-0.447) \\ e_{10} &=& (0.724,-0.526,-0.447) \\ e_{11} &=& (0,0,1) \\ e_{12} &=& (0,0,-1) \end{array}
```

The coordinate system of the icosahedron model that we choose has two opposite vertices located along the z-axis, five vertices constructing an equilateral pentagon are parallel to the xy-plane at the distance 0.447 above the xy-plane, and the other five vertices forming also an equilateral pentagon are located at almost symmetrical positions opposite to the previous pentagon along the xy-plane except  $36^{\circ}$  ( $\pi/5$ ) rotation along the z-axis. We show the icosahedron model in Figure 1. The number beside each node labels the order of assignment of 12 unit vectors used in our work.

The first five unit vectors located at the vertices of the upper pentagon have coordinates:

$$e_{i} = \frac{\sqrt{4\sin^{2}\frac{\pi}{5} - 1}}{2\sin^{2}\frac{\pi}{5}} \left[ \cos\frac{2\pi (i - 1)}{5}, \sin\frac{2\pi (i - 1)}{5}, \frac{1 - 2\sin^{2}\frac{\pi}{5}}{\sqrt{4\sin^{2}\frac{\pi}{5} - 1}} \right]; 1 \le i \le 5$$
(2)

The next five unit vectors located at the vertices of the lower pentagon that is rotated by the angle  $\pi/5$  with respect to the upper pentangle have coordinates:

$$e_{i} = \frac{\sqrt{4\sin^{2}\frac{\pi}{5} - 1}}{2\sin^{2}\frac{\pi}{5}} \left[ \cos\frac{\pi + 2\pi(i - 6)}{5}, \sin\frac{\pi + 2\pi(i - 6)}{5}, \frac{-1 + 2\sin^{2}\frac{\pi}{5}}{\sqrt{4\sin^{2}\frac{\pi}{5} - 1}} \right]; 6 \le i \le 10$$
(3)

We use these 12 directional unit vectors from the icosahedron model to fit our coordination clusters from the  $1.5\text{\AA}522$  dataset. If a given coordination cluster contains m neighbors,

represented by m unit vectors; then there are  $\binom{12}{m}$  different ways of choosing m ( $1 \le m \le 12$ ) directional unit vectors in the icosahedron model to fit this coordination cluster. However, we can significantly reduce this number by removing sets of directional vectors related by

symmetry. For the simplest case  $\binom{12}{1}$ , theoretically there are 12 combinations given by the binomial coefficient formula. However, since all vertices of the icosahedron are geometrically equivalent we can choose a single one to represent all others. We have shown previously that the number of possible compact lattice conformations can be reduced by removing conformations related by symmetries of the shape. For example, the cube has the total number of symmetries 48, and the number of compact self-voiding walks on the cubic lattice within a cubic shape can be reduced by the factor  $\sigma = 48.30$  Similarly, we construct irreducible sets of m ( $1 \le m \le 12$ ) directional vectors of icosahedron. We first enumerate all possible combinations of choosing m directional vectors from 12. If two of them are symmetric, they will overlap after applying proper rotation using QTRFIT algorithm and we can eliminate one of them. By considering all combinations of directional vectors and rotations superimposing these sets we obtain irreducible combinations of the m ( $1 \le m \le 12$ ) directional vectors of icosahedron.

The probabilities of various irreducible combinations of m directional unit vectors are different. If we assume that all combinations are equally probable, then the probabilities of irreducible combinations can be computed from the following formula:

$$P_{\rm in}$$
=the total number of reducible combinations having the same pattern/ $\begin{pmatrix} 12 \\ m \end{pmatrix}$  (4)

In the case of m=2, we have 3 irreducible combinations  $(\mathbf{e}_1, \mathbf{e}_2)$ ,  $(\mathbf{e}_1, \mathbf{e}_3)$ , and  $(\mathbf{e}_1, \mathbf{e}_8)$  that we call patterns. The pattern  $(\mathbf{e}_1, \mathbf{e}_2)$  corresponds to the case when two vertices of the icosahedron are the nearest neighbors; the pattern  $(\mathbf{e}_1, \mathbf{e}_3)$  represents the case when the two vertices are second nearest neighbors; and  $(\mathbf{e}_1, \mathbf{e}_8)$  corresponds to the situation when the two vertices are opposite points, the most distant nodes of the icosahedron. Patterns  $(\mathbf{e}_1, \mathbf{e}_2)$  and  $(\mathbf{e}_1, \mathbf{e}_3)$  have the same probability  $P_{irr}$  0.455, while  $P_{irr}$  of the pattern  $(\mathbf{e}_1, \mathbf{e}_8)$ , is five times less frequent, is only 0.091.

Obviously the two vectors in the pattern  $(\mathbf{e}_1, \mathbf{e}_8)$  are less densely distributed than the vectors in patterns  $(\mathbf{e}_1, \mathbf{e}_2)$  and  $(\mathbf{e}_1, \mathbf{e}_3)$ . We use the following equation to compute the density of different patterns, denoted as  $D_{\text{patterm}}$ : For m vectors, we will totally have

 $\binom{12}{2} = m(m-1)/2$  pairs. And we compute the summation of the differences of all these m(m-1)/2 pairs and the summation is then scaled by dividing the number of pairs, that is m(m-1)/2)

$$D_{\text{pattern}} = \sqrt{\frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} (\mathbf{e}_i - \mathbf{e}_j)^2}$$
 (5)

where m is the number of vectors in each irreducible combination.

#### QTRFIT algorithm for superimposing two clusters of unit vectors

The QTRFIT algorithm was developed in 1990 by David J. Heisterberg<sup>31</sup> to superimpose atoms of two molecules by quaternion-based approach. In this algorithm, two matrices  $\mathbf{A} = [a_{ij}]$  and  $\mathbf{B} = [b_{ij}]$  of size  $n \times n$  built from vectors  $\mathbf{a}_i$  and  $\mathbf{b}_i$ ;  $(1 \le i \le n)$  represent conformations of each of two molecules, or as in our study, two sets of unit vectors. The goal is to find a rotation represented by the matrix  $\mathbf{U}$  which minimizes the error of superimposition defined as:

$$E=Tr\left[\mathbf{W}(UA-B)^{2}\right]$$
 (6)

where **W** is the weight matrix, which in our case is simply the identity matrix **I**, and Tr denotes the trace of the matrix. A detailed description of the quaternion representation of three-dimensional rotations and the QTRFIT algorithm was given in our previous study<sup>1</sup>. In Eq. 6, **B** is the target matrix composed of directional vectors, and the matrix **A** (also composed of directional vectors), tries to fit it by optimized rotation. The order of directional vectors (columns) in matrices **A** and **B** is fixed. However this order may not be optimal to minimize the error E in Eq. 6. To globally minimize this error we should consider all permutations of directional vectors in matrix **A**. If **A** is composed of m non-zero vectors, we can rearrange them in m! different ways. After superimposing **A** and all its rearrangements with **B**, we can find the order of directional vectors in the matrix **A** having the smallest root mean square distance (RMSD)

$$RMSD(\mathbf{A}', \mathbf{B}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{3} (a_{ij}' - b_{ij})^{2}}$$
(7)

upon superimposition between A' = UA and B.

#### **Evaluation criteria for superimposition**

We use the same criterion as in our previous study<sup>1</sup> to evaluate the quality of superimposition between the directional vectors of coordination clusters derived from our 1.5Å522 dataset and the directions of the icosahedron model. The quality of the superimposition of two sets of vectors is measured by the order parameter (*OP*), defined as the average square of the cosine of the angle  $\Delta\alpha$  between two superimposed directional vectors:

$$OP = \frac{\sum_{i=1}^{m} \cos^2 \Delta \alpha_i}{m} \tag{8}$$

where m is the number of superimposed directional vectors.

For two sets of vectors that are perfectly superimposed the order parameter OP = 1. We also use RMSD defined by Eq. 7 to measure the difference between two superimposed sets of vectors. If two sets of vectors perfectly overlap then RMSD equals zero. Since all vectors studied here are unit vectors, RMSD is an alternative measure of the quality of a superposition to OP.

#### Linear regression

We use the R programming language to perform linear regression analysis in our studies. R is a language and environment for statistical computing and graphics freely available as a part of the GNU Project. R currently developed by the R Development Core Team (http://www.r-project.org) is similar to the S language and environment developed in the mid-70s at Bell Laboratories by John Chambers and colleagues. The R function, 'lm', is used to fit linear models. We also use the built-in function step() to choose a best formula-based linear regression model.

#### **RESULTS**

#### Results from protein dataset

The distribution of unit vector clusters from the 1.5Å522 dataset has a nice bell shape with its peak around 6-7 vectors (Fig. 2). In our dataset, there are few coordination clusters having more than 12 neighbors (directional vectors). The clusters having 13 and 14 directional vectors account for 0.2% and 0.04% of the total number of clusters, respectively. Because the icosahedron model has only 12 vertices, we ignore those cases with more than 12 directional vectors.

#### Results from analyzing the icosahedron model

We choose m ( $1 \le m \le 12$ ) vectors from the 12 directional vectors of the icosahedron pointing from its center to 12 vertices. Table I lists all irreducible combinations of m vectors. The irreducible m-tuplets of directional vectors greatly reduce the combinatorial size of the problem. Without this combinatorial reduction in choosing sets of m ( $1 \le m \le 12$ ) vectors from 12, we would have  $2^{12} - 1 = 4095$  of all possible combinations. This can be reduced to 63 irreducible m-tuplets by eliminating sets of vectors related by symmetry of the icosahedron, using the method described by us earlier. It should be noted that there is a symmetry between the distribution of the m-tuplets and that of the (12-m)-tuplets of vectors since the later ones correspond to the removal of m vectors from the single 12-tuplet. Because of this the maximum number of irreducible m-tuplets (12) is observed for m = 6 (see Table I).

For each irreducible combination of m orientational vectors (m-tuplet), we consider two properties: pattern density  $D_{pattern}$  defined by Eq. 5 and the probability of such irreducible m-tuplet.  $D_{pattern}$  is a measure of how directional vectors in a given coordination cluster are close to each other. The smaller  $D_{pattern}$  value is, the closer are the vectors in this cluster. For the 1-tuplet  $D_{pattern}$  is zero by definition. It is rather obvious that combinations containing neighboring vertices of the icosahedron should have low values of  $D_{pattern}$ , while those including opposite, most distant vertices should have large values of  $D_{pattern}$ . The

results shown in Table I clearly demonstrate the correctness of this supposition. For all irreducible m-tuplets shown in Table I the triplet C3\_1 that has the lowest  $D_{pattern}=1.051$  is the combination of three vectors:  $\mathbf{e}_1$ ,  $\mathbf{e}_2$ , and  $\mathbf{e}_6$  joining the center of the icosahedron with three vertices located on the same equilateral triangular face. On the other hand the doublet C2\_3 that has the largest value of  $D_{pattern}=2.000$  represents the combination of two oppositely directed vectors  $\mathbf{e}_1$  and  $\mathbf{e}_8$ .

Probability of a given irreducible m-tuplet informs us if a particular irreducible combination is more likely than other m-tuplets. For choosing 1 or 11 vectors from 12, there are 12 different combinations; however all of them are reducible and we have only a single irreducible combination (packing pattern). For choosing 12 vectors from 12, obviously only one combination is available. There are several different irreducible combinations (shown in Table I) when we choose m ( $2 \le m \le 10$ ) out of 12 unit vectors of the icosahedron. Each irreducible combination (packing pattern) represents a certain number of reducible combinations, which defines the probability of this packing pattern. We separately consider m-tuplets with a different number m ( $1 \le m \le 12$ ) of directional vectors, so the probabilities of different packing patterns with the same number m of vectors by definition sum to 1. There are no obvious correlations between the density of a given pattern  $D_{pattern}$  and its probability.

# Results from superimposition of coordination clusters derived from the 1.5Å522 dataset with directional vectors in icosahedron model

We have used the QTRFIT algorithm to superimpose coordination clusters derived from the 1.5Å522 dataset with directional vectors of the icosahedron model. If a given coordination cluster contains m ( $1 \le m \le 12$ ) vectors, we superimpose it with the irreducible m-tuplets of directional vectors (packing patterns) having the same number m of vectors. The packing pattern having the lowest RMSD or the highest OP among all packing patterns with same number m of vectors is chosen as the best fit of a given coordination cluster. The results of superimposition are shown in Table II. In Figure 3, we show the distribution of the polar ( $\phi$ ) and azimuthal ( $\theta$ ) angles of each vector in the coordination clusters after superimposition with the icosahedron model. These polar and azimuthal angles are unevenly distributed among 12 peaks corresponding to 12 nodes of the icosahedron model, because we don't choose irreducible vectors randomly.

We use two parameters: RMSD and OP, to measure how well coordination clusters derived from the 1.5Å522 dataset fit the directional vectors of the icosahedron. If two sets of vectors are completely overlapping, the the RMSD is 0 and the OP is 1. With increasing number m of vectors, the mean values of RMSD increase and the mean values of OP decrease. The standard deviations of both RMSD and OP are small and practically don't depend on m. If two sets of superimposed clusters are uncorrelated, the OP value is 1/3.\(^1\) Our superimposition results show clearly that the OP values are much better than 1/3 (see Table II). In our previous study, superimposition of coordination clusters derived from protein structures with directional vectors of the fcc lattice gave the order parameter value OP = 0.82.\(^1\) We have also observed in those studies that the OP values decrease with growing number m of directional vectors in the cluster.

Here we study a different model, with directional vectors of the icosahedron instead of the fcc lattice. We use the same QTRFIT algorithm for superimposition of clusters, and we obtain much better overlap (higher *OP* values) between directions in coordination clusters derived from the protein dataset and the model. This shows that the icosahedron is a better model to represent the residue packing problem than the fcc lattice.

We have performed superimposition computations for all 110,255 coordination clusters derived from the 1.5Å522 dataset. Table II shows the results obtained for best packing pattern for each coordination cluster averaged over all clusters. We were able also to compute distributions of coordination clusters among different packing patterns for a given number m of directional vectors in the cluster. Another interesting problem is what factors affect these distributions. Figure 4. shows the dependence between frequencies of various patterns and their densities  $D_{pattern}$  (defined by Eq. 5 and listed for various patterns in Table I). We observe a general trend that the frequency of a pattern decreases for lower densities (the increase of D<sub>pattern</sub> value), which implies that coordination clusters derived from the 1.5Å522 dataset tend to pack as closely as possible. Since each vector actually represents a residue, coordination clusters represent not only directions of contacts, but also locations of neighboring residues. Our observation (Fig. 4) that residues try to pack as closely as possible is consistent with results of many earlier studies.1'8'10', 12 Regardless of whether coordination clusters include only two vectors (for surface residues) or ten or more vectors (for buried residues inside protein core), they always tend to be packed as closely as possible.

Packing patterns (irreducible combinations of directional vectors) have different probabilities since each irreducible set of vectors represents a different number of reducible vector combinations. We are interested in knowing whether the varying probabilities of packing patterns affect the distribution of coordination clusters among different patterns. In Figure 5, we plot fractions of various patterns vs. their probabilities for vector clusters containing from 2 to 10 vectors, and cannot find any correlations between them. However, if we check these vector clusters individually, we find that the pattern fraction is affected by its probability, although the dependence on pattern density  $D_{\text{pattern}}$  is dominant. Figure 6 shows the relations between pattern fraction and pattern density  $D_{\text{pattern}}$  (Fig. 6a) and pattern probability (Fig. 6b) for vector clusters containing 6 directional vectors. Two points in Figs 6a and 6b pointed out in two circles illustrate how pattern probability influences pattern fraction. These two points have lower  $D_{\text{pattern}}$  (1.305 and 1.350) than others, so their fractions should be higher. However they obviously do not follow the overall trend, which may be explained by the low probabilities (0.013 and 0.22) of these patterns.

#### Multiple regression analysis of packing patterns

In this next section we will apply regression analysis to study the relationships among pattern density, pattern probability and pattern fraction. There are at least two independent variables: pattern density  $(x_1)$  and pattern probability  $(x_2)$  related to our dependent variable: pattern fraction (y), and therefore this is a multiple linear regression problem. By checking the correlation of our independent and dependent variables, we found that there is no simple linear correlation between them. Because of this we include both  $x_1$ ,  $x_2$  and  $x_1^2$ ,  $x_2^2$  terms in the regression analysis. We also included the cross-term  $x_1x_2$  corresponding to the interaction between  $x_1$  and  $x_2$  jointly affecting y.

Before performing a regression, we transformed y to  $\sqrt{y}$  so that it appears more like a normal distribution. In order to check the quality of the transformation and to diagnose if the normality assumption behind the regression model is reasonable for our data fitting problem, we draw a normal probability plot of the residuals shown in Figure 7. The normal probability plot shows the actual percentiles of the residuals vs. the theoretical percentiles of a normal distribution with the same mean and the variance. Ideally, this plot should be a diagonal straight line. Figure 7 shows only a slight departure from normality in the tail of the distribution. This small extent of non-normality suggests that it may be appropriate to use a linear regression model for our data fitting. (http://www.duke.edu/~rnau/regnotes.htm)

The choice of the best possible model is important for enabling good predictions. First, we use all independent variables:  $x_1$ ,  $x_2$ ,  $x_1^2$ ,  $x_2^2$ , and  $x_1x_2$ , for building a full linear regression analysis model. Since not all independent variables are significant, we use the build-in function step() in R to choose a best reduced model. 32,33 The model we finally choose is  $\sqrt{y}$ =a+b<sub>1</sub>x<sub>1</sub>+b<sub>2</sub>x<sub>1</sub><sup>2</sup>+b<sub>3</sub>x<sub>1</sub>x<sub>2</sub>. In this model, all three terms with independent variables depend on x<sub>1</sub> (density of pattern D<sub>pattern</sub>), but only one cross-term corresponding to the interaction of x<sub>1</sub> and x<sub>2</sub> is related to the pattern probability. This implies that D<sub>pattern</sub> plays a more dominant role in determination of the pattern fraction than pattern probability. This result is consistent with our previous observations reported in Figures 4 and 6.

The coefficients obtained by fitting our data to the model (equation) are 3.10, - 2.86, 0.61, and 0.51 for a,  $b_1$ ,  $b_2$ , and  $b_3$  respectively (Table III), which explains why y has a positive correlation with  $x_2$ , and the interactions of  $x_1$  and  $x_2$ , but a negative correlation with  $x_1$ . Figure 8 shows the scatter plot of predicted vs. observed square root fractions. The dotted lines denote 95% confidence intervals for predicted values computed by using the maximal standard error of predicted values. Most of predicted  $\sqrt{y}$  are located in the region of 95% confidence intervals.

#### Does the central residue affect protein packing?

We want to know if the central residue type influences the orientational packing of its neighbors. All 110,255 coordinate clusters from the 1.5Å522 dataset were divided into 20 subsets according to the type of the central residue. For each subset, there is no significant difference in RMSD and OP values, which implies that the type of central residue is not very important for the superimposition of coordination clusters with the icosahedron model. In order to learn about possible central residue-type effect on protein packing pattern, we compared pairwise all pattern fraction distributions for the 20 different types of central residues and the overall distribution (ALL) and computed Pearson's correlation coefficients among all distributions (Table IV). Most of correlation coefficients are above 0.9 except for Gly and Cys, which have a clearly different behavior from the other amino acids. Glycine does not have a side chain and the  $C^{\alpha}$  atom represents the whole residue, and many cysteines form disulfide bonds, both of which appear to influence the packing patterns of their neighboring residues.

Although all other central residues have better correlation coefficients than Gly and Cys, they show different orders of these correlations upon a careful analysis. For example, Trp has correlations with other residues varying from 0.890 to 0.978 in the following order: Glu < Lys < Arg < Gln, < Asp < Ala < Thr < Asn < Ser < ALL < His < Pro < Tyr < Met < Val < Phe < Leu < Ile. This order suggests that Trp has the most similar distribution to the other hydrophobic residues, such as Ile, Leu, Phe, Val, Met and Tyr, and the least similar distributions to hydrophilic residues such as Glu, Lys, Arg, Gln and Asp in the center of the coordination cluster. We check also the clusters with central residue Asp, which shows almost reverse order of correlations: Phe < Tyr < Trp < Ile < Val < Lys < Leu < Arg,Met,Pro < His < Thr < Gln < Glu < Ala < Ser < Asn < ALL with correlation coefficients ranging from 0.919 to 0.991. Basically, ASP has the most similar distribution to other hydrophilic residues and the least similar distribution to hydrophobic residues in the center of the coordination cluster. A similar order of correlations can be obtained for all other central residues except for Gly and Cys. This observation implies that central residues with similar hydrophobicity should exhibit similar packing behavior.

#### DISCUSSION

In the present study, we use a larger and higher quality protein dataset than in our earlier work to study the residue packing problem. We derived 110,255 coordination clusters from our new 1.5Å522 dataset that is thus almost 4 times larger than the number of clusters used previously<sup>1</sup>.

Protein packing is an important problem in structural biology and relates to protein structure design and prediction, binding site prediction etc. It is also a difficult and somewhat controversial problem because of its complexity, existence of conflicting experimental data, and the absence of deep and rational analysis. A reasonable and simple model is a critical step in the study of the protein packing problem at the residue level. In previous studies, lattice models were used to fit coordination clusters derived from protein structure datasets. In the present study we use the icosahedron as a model of residue packing. The icosahedron has a higher local packing density than any other lattice model, which may account for the reason why we obtain better superimpositions of coordination clusters with directional vectors pointing from the centers of the icosahedra to their 12 vertices (OP = 0.91) than with the 12 directions of the fcc lattice (OP = 0.82). This is somewhat surprising since all angles of the icosahedron are identical while the fcc directions have several angles. It is extremely important to have good superimpositions to simplify the residue packing problem. We can explain residue packing in proteins using simple theoretical models only if coordination clusters derived from experimental protein structures match the theoretical model well. Improvement of the superimposition results using the icosahedron model enables us to explain the residue packing problem in a simpler and clearer way.

We carefully analyzed properties of different irreducible combinations of unit directional vectors of the icosahedron (packing patterns) by introducing two novel parameters: pattern density  $D_{pattern}$  and pattern probability. We found that coordination clusters from the 1.5Å522 dataset fit these packing patterns in a non-random way. The preference is given to packing patterns having higher pattern density (lower value of  $D_{pattern}$ ) and higher pattern probability. Such packing behavior suggests that protein packing is driven mostly to maximize the packing density because the preference of low values of pattern density  $D_{pattern}$  indicates that proteins tend to be packed at higher density. The probability of packing patterns is a novel concept not previously studied. Although the probability of packing patterns does not have a dominant effect on the distribution of residue clusters, it does affect them as seen by the analysis of a specific example (Fig. 6) and by developing a linear regression model.

The residue clusters with the different central residues (except Gly and Cys) have similar preferences to packing patterns as seen from examination of pairwise correlation coefficients between them. This observation is consistent with our previous studies<sup>1</sup>. Additionally we have found that correlation coefficients are related to the hydrophobicity of the central residue in the coordination cluster.

One of the most interesting parts of our study is the prediction of pattern fractions using a linear regression model. The predicted square roots of fractions obtained from a multiple linear regression model exhibit a good correlation with the observed ones as seen in Figure 8. The application of this model in the future might significantly aid in predicting protein structures and in protein design. We hope that it may be possible to convert this model into a set of effective energy functions. Another interesting question is whether this regression model is sufficiently robust for selecting among different fitting models.

### **Acknowledgments**

RLJ and AK acknowledge financial support provided by the NIH grants 1R01GM072014-01, 1R01GM073095-02, and 1R01GM081680-01.

#### REFERENCES

1. Bagci Z, Kloczkowski A, Jernigan RL, Bahar I. The origin and extent of coarse-grained regularities in protein internal packing. Proteins: Structure, Function, and Genetics 2003;53(1):56–67.

- 2. Raghunathan G, Jernigan RL. Ideal architecture of residue packing and its observation in protein structures. Protein Science 1997;6(10):2072–2083. [PubMed: 9336831]
- Dahiyat BI, Mayo SL. Probing the role of packing specificity in protein design. Proc Natl Acad Sci U S A FIELD Full Journal Title:Proceedings of the National Academy of Sciences of the United States of America 1997;94(19):10172–10177.
- Kono, H.; Nishiyama, M.; Tanokura, M.; Doi, J. Design of hydrophobic core of E. coli malate dehydrogenase based on the side-chain packing; Pacific Symposium on Biocomputing '97; Maui, Hawaii. Jan 6-9, 1997; 1997. p. 210-221.
- Pontius J, Richelle J, Wodak SJ. Deviations from standard atomic volumes as a quality measure for protein crystal structures. J Mol Biol FIELD Full Journal Title: Journal of molecular biology 1996;264(1):121–136.
- 6. Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci FIELD Full Journal Title:Protein science: a publication of the Protein Society 1998;7(9):1884–1897.
- Kuhn LA, Siani MA, Pique ME, Fisher CL, Getzoff ED, Tainer JA. The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. J Mol Biol FIELD Full Journal Title: Journal of molecular biology 1992;228(1):13–22.
- 8. Harpaz Y, Gerstein M, Chothia C. Volume changes on protein folding. Structure FIELD Full Journal Title:Structure (London, England: 1993) 1994;2(7):641–649.
- Paci E, Marchi M. Intrinsic compressibility and volume compression in solvated proteins by
  molecular dynamics simulation at high pressure. Proc Natl Acad Sci U S A FIELD Full Journal
  Title:Proceedings of the National Academy of Sciences of the United States of America
  1996;93(21):11609–11614.
- 10. Hubbard SJ, Gross KH, Argos P. Intramolecular cavities in globular proteins. Protein Eng FIELD Full Journal Title:Protein engineering 1994;7(5):613–626.
- Word JM, Lovell SC, LaBean TH, Taylor HC, Zalis ME, Presley BK, Richardson JS, Richardson DC. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. J Mol Biol FIELD Full Journal Title: Journal of molecular biology 1999;285(4): 1711–1733.
- 12. Dill KA. Dominant forces in protein folding. Biochemistry FIELD Full Journal Title:Biochemistry 1990;29(31):7133–7155.
- 13. Chen J, Stites WE. Packing is a key selection factor in the evolution of protein hydrophobic cores. Biochemistry FIELD Full Journal Title:Biochemistry 2001;40(50):15280–15289.
- 14. Gerstein M, Chothia C. Packing at the protein-water interface. Proc Natl Acad Sci U S A FIELD Full Journal Title:Proceedings of the National Academy of Sciences of the United States of America 1996;93(19):10167–10172.
- Liang J, Dill KA. Are proteins well-packed? Biophys J FIELD Full Journal Title: Biophysical journal 2001;81(2):751–766.
- 16. Bromberg S, Dill KA. Side-chain entropy and packing in proteins. Protein Sci FIELD Full Journal Title:Protein science: a publication of the Protein Society 1994;3(7):997–1009.
- 17. 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 U S A FIELD Full Journal Title:Proceedings of the National Academy of Sciences of the United States of America 1991;88(10):4195–4199.

18. Richards FM. Areas, volumes, packing and protein structure. Annu Rev Biophys Bioeng FIELD Full Journal Title: Annual review of biophysics and bioengineering 1977;6:151–176.

- 19. Bagci Z, Jernigan RL, Bahar I. Residue packing in proteins: Uniform distribution on a coarse-grained scale. Journal of Chemical Physics 2002;116(5):2269–2276.
- Cipra B. Mathematics: Packing challenge mastered at last. Science (Washington, D C) 1998;281(5381):1267.
- 21. Sloane NJA. Kepler's conjecture confirmed. Nature 1998;395(6701):435-436.
- 22. Donev A, Cisse I, Sachs D, Variano E, Stillinger FH, Connelly R, Torquato S, Chaikin PM. Improving the density of jammed disordered packings using ellipsoids. Science 2004;303(5660): 990–993. [PubMed: 14963324]
- 23. Weitz DA. Packing in the spheres. Science 2004;303(5660):968–969. [PubMed: 14963318]
- 24. Hermann H, Elsner A, Gemming T. Influence of the packing effect on stability and transformation of nanoparticles embedded in random matrices. Materials Science-Poland 2005;23(2):541–549.
- 25. Feng Y, Kloczkowski A, Jernigan RL. Four-body contact potentials derived from two protein datasets to discriminate native structures from decoys. Proteins: Structure, Function, and Bioinformatics 2007;68(1):57–66.
- 26. Wang GL, Dunbrack RL. PISCES: a protein sequence culling server. Bioinformatics 2003;19(12): 1589–1591. [PubMed: 12912846]
- Chushak Y, Travesset A. Coarse-grained molecular-dynamics simulations of the self-assembly of pentablock copolymers into micelles. Journal of Chemical Physics 2005;123(23):234905/234901– 234905/234907. [PubMed: 16392947]
- Sathaliyawala T, Rao M, Maclean DM, Birx DL, Alving CR, Rao VB. Assembly of human immunodeficiency virus (HIV) antigens on bacteriophage T4: a novel in vitro approach to construct multicomponent HIV vaccines. Journal of Virology 2006;80(15):7688–7698. [PubMed: 16840347]
- Crick FHC, Watson JD. Structure of Small Viruses. Nature 1956;177(4506):473–475. [PubMed: 13309339]
- 30. Kloczkowski A, Jernigan RL. Computer generation and enumeration of compact self-avoiding walks within simple geometries on lattices. Computational and Theoretical Polymer Science 1997;7(3-4):163–173.
- 31. Heisterberg, DJ. QTRFIT algorithm for superimposing two similar rigid molecules. The Ohio Supercomputer Center Ohio State University; Columbus, OH: 1990.
- 32. Hastie, TJ.; Pregibon, D. Generalized linear models. Wadsworth & Brooks/Cole; 1992.
- 33. Venables, WN.; Ripley, BD. Modern applied statistics with S. Springer; New York: 2002.

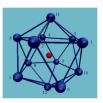


Fig. 1. The icosahedron model

The numbers beside nodes are in the same order as the vectors defined in Eq. 1 connecting the center of the icosahedron (red point) with each of the nodes.

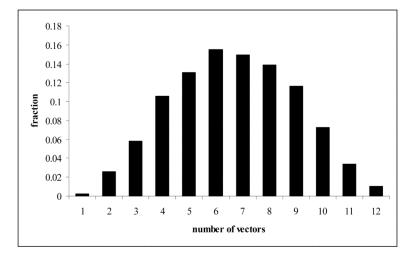


Fig. 2.

The distribution of clusters of directional vectors derived from the 1.5Å522 dataset.

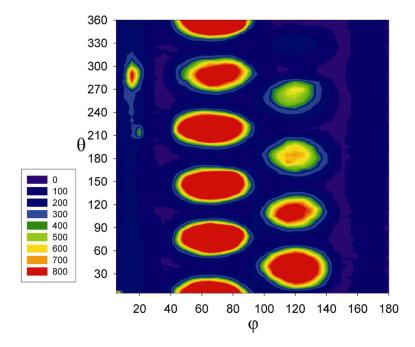


Fig. 3. The distribution of polar and azimuthal angles of vectors in coordination clusters after superimposition with the icosahedron model  ${\bf r}$ 

The 12 peaks correspond to orientations  $[(\phi,\theta)]=(63,0)$ , (63,72), (63,144), (63,216), (63,288), (117,36), (117,108), (117,180), (117,252), (117,324), (0,0), (180,0) in the icosahedron, and are listed in the same order as the vectors defined in Eq. 1 and shown in Figure 1.

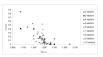
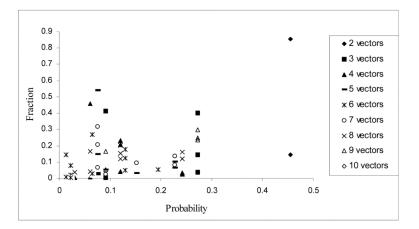
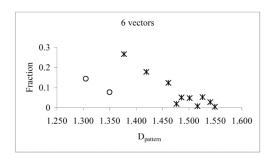


Fig. 4. Pattern fraction vs. pattern density  $D_{pattern}$ Distribution of coordination clusters among different patterns is negatively proportional to densities of these patterns, indicating that the closely positioned clusters are more frequent.



**Fig. 5. Pattern fraction vs. pattern probability**There is no clear relationship between the two measures.

(a)



(b)

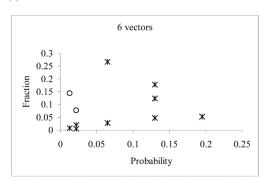
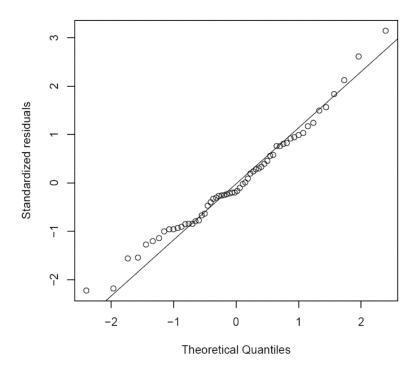


Fig. 6. Pattern fraction vs. pattern density  $D_{pattern}$  (a) and pattern fraction vs. pattern probability (b) for the coordination clusters containing 6 vectors

There is an overall trend for diminishing fractions at lower packing density (higher value of  $D_{pattern}$ ). However, the two circles clearly fall outside the overall trends.

## **Normal Probability Plot**



**Fig. 7.** Normal probability plot for the square root of pattern fraction.

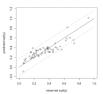


Figure 8. Observed vs. predicted values of  $\sqrt{y}$ 

The dotted lines show the 95% confidence intervals for the predicted values. The black diagonal line shows the ideal case when observed and predicted values are identical.

Table I

NIH-PA Author Manuscript

Densities of patterns D<sub>pattern</sub> and their probabilities for all irreducible combinations of m vectors chosen from 12 directional vectors of the icosahedron.

Feng et al.

	vectors	- pattern	•	defi	ned ir	ı Eq. 1 s	defined in Eq. 1 and shown in Figure 1)		
C1	1	0	1.000	1	—	_		—	_
C2_1	2	1.052	0.455		2				
C2_2	2	1.701	0.455	_	ж				
C2_3	2	2.000	0.091						
C3_1	3	1.051	0.091		2 6	9			
C3_2	3	1.305	0.273	_	7				
C3_3	3	1.516	0.273	-	7	4			
C3_4	3	1.632	0.273		7				
C3_5	к	1.701	0.091	_	<u>د</u>				
C4_1	4	1.185	0.061		2	3   11			
C4_2	4	1.305	0.121	_	7	3 6			
C4_3	4	1.414	0.121	-	7	4			
C4_4	4	1.478	0.242	_	67	8			
C4_5	4	1.516	0.121	-	7	3 12			
C4_6	4	1.576	0.242		7	3 9			
C4_7	4	1.611	0.061	_	7	4   12			
C4_8	4	1.633	0.03	_	7	6 8			
C5_1	5	1.282	0.076	1	2	3 4			
C5_2	5	1.39	0.076	-	7	3 6	10		
C5_3	5	1.414	0.091	_	7	4	8		
C5_4	5	1.453	0.227	-	67	4	9		
C5_5	5	1.513	0.227	-	67	4	6		
C5_6	5	1.536	0.076	-	7	4	12		
C5_7	5	1.549	0.152	-	67	8	6		
C5_8	5	1.571	0.076	-	7	3	12		

Page 21

Feng et al.

names	# of vectors	Dpattern	Probability	Irr	jmed jmed	ible in E	vector	r comb	inatic wn in	Irreducible vector combination (number i represents vector e, defined in Eq. 1 and shown in Figure 1)	nber 1	i repr	esents	vecto	ı e;
C6_1	9	1.305	0.013	-	2	3	4	5	11						
C6_2	9	1.35	0.022	_	2	3	4	7	11						
C6_3	9	1.377	0.065	_	2	3	4	9	11						
C6_4	9	1.419	0.13	_	2	3	4	9	7						
C6_5	9	1.461	0.13	_	2	3	4	5	9						
0_6_6	9	1.476	0.022	_	2	3	4	7	12						
C6_7	9	1.486	0.195	_	2	3	4	9	8						
8_62	9	1.501	0.13	_	2	3	4	9	12						
6 <sup>-</sup> 92	9	1.516	0.013	_	2	3	4	5	12						
C6_10	9	1.525	0.195	_	2	3	4	9	6						
C6_11	9	1.54	0.065	_	2	3	4	6	12						
C6_12	9	1.549	0.022		2	3	∞	6	10						
C7_1	7	1.388	0.076		2	3	4	5	9	11					
C7_2	7	1.436	0.076	_	2	3	4	9	7	∞					
C7_3	7	1.447	0.091	_	2	3	4	5	=	12					
C7_4	7	1.466	0.227	_	2	3	4	5	9	7					
C7_5	7	1.494	0.227	_	2	3	4	5	9	∞					
C7_6	7	1.505	0.076	_	2	3	4	5	9	12					
C7_7	7	1.512	0.152		2	3	4	9	8	6					
C7_8	7	1.522	0.076		2	ж	4	9	6	12					
C8_1	~	1.42	0.061		2	8	4	5	9	7	11				
C8_2	∞	1.442	0.121	_	2	3	4	5	9	∞	Ξ				
C8_3	∞	1.464	0.121	_	2	3	4	5	9	11	12				
C8_4	∞	1.477	0.242	_	2	3	4	5	9	7	∞				
C8_5	∞	1.485	0.121	_	2	3	4	5	9	7	12				
0_8Z	∞	1.499	0.242		2	3	4	5	9	7	6				
C8_7	∞	1.507	0.061	=	2	3	4	5	9	∞	12				
	∞	1.512	0.03		2	ю	4	9	∞	6	10				
G <sub>-1</sub>	6	1.447	0.091		2	3	4	5	9	7		=			

Page 22

names	# of vectors	D <sub>pattern</sub>	Probability	Irr	educ	ible in E	vecton	r comb	own ir	Irreducible vector combination (numbe defined in Eq. 1 and shown in Figure 1)	ımber re 1)	Irreducible vector combination (number i represents vector equefined in Eq. 1 and shown in Figure 1)	esents	vecto	ır e <sub>i</sub>
C9_2	6	1.464	0.273	-	2	3	4	5	9	7	6	111			
C9_3	6	1.481	0.273	-	2	3	4	5	9	7	∞	12			
C9_4	6	1.491	0.273	-	2	3	4	5	9	7	∞	6			
C9_5	6	1.497	0.091	-	7	ж	4	S	9	7	6	12			
C10_1	10	1.469	0.455		2	3	4	5	9	7	∞	6	11		
C10_2	10	1.483	0.455	-	2	3	4	2	9	7	∞	6	12		
C10_3	10	1.491	0.091	-	2	3	4	S	9	7	∞	6	10		
C111	11	1.477	1.000		2	3	4	5	9	7		6	10	Ξ	
C12	12	1.477	1.000		2	3	4	5	9	7		6	10	11	12

Feng et al.

Page 23

#### Table II

The mean and standard deviation (std) values of *RMSD* and *OP* from superimposition of coordination clusters containing m directional vectors ( $1 \le m \le 12$ ) with the icosahedron model

# of vectors	RM	SD	0	P
	Mean	Std	Mean	Std
1	0	0	1	0
2	0.107	0.015	0.984	0.101
3	0.184	0.051	0.964	0.017
4	0.223	0.045	0.949	0.022
5	0.251	0.050	0.936	0.024
6	0.271	0.052	0.927	0.027
7	0.281	0.044	0.922	0.023
8	0.286	0.039	0.920	0.022
9	0.289	0.039	0.918	0.021
10	0.301	0.042	0.911	0.023
11	0.309	0.047	0.901	0.027
12	0.353	0.051	0.881	0.031

Table III

Summary of the linear regression model:  $\sqrt{y}=a+b_1x_1+b_2x_1^2+b_3x_1x_2$ 

	Estimate	Std. Error	T statistics	p-value
A	3.10	0.58	5.36	1.60e-06
b <sub>1</sub>	-2.86	0.78	-3.68	5.3e-04
b <sub>2</sub>	0.61	0.26	2.32	0.024
b <sub>3</sub>	0.51	0.08	6.71	1.03e-08
Res	idual standard	d error: 0.0976	on 56 degrees	of freedom
Mul	ltiple R-squar	ed: 0.7499, Ad	justed R-square	ed: 0.7365
F-st	atistic: 55.98	on 3 and 56 D	F, p-value: < 2.	2e-16

A p-value less than 0.05 means high significance.

Table IV

NIH-PA Author Manuscript

CYS         1.000           MET         0.895         1.000           PHE         0.916         0.966         1.000           ILE         0.886         0.977         0.976           LEU         0.801         0.984         0.976           VAL         0.886         0.970         0.968           TYR         0.915         0.968         0.968           ALA         0.852         0.976         0.968           GLY         0.889         0.980         0.945           ASN         0.881         0.982         0.936           GLN         0.881         0.962         0.913           GLN         0.885         0.966         0.913           GLN         0.881         0.966         0.913           GLU         0.881         0.966         0.913           GLU         0.815         0.946         0.880           HIS         0.912         0.946         0.865           ARG         0.884         0.956         0.913           LYS         0.892         0.947         0.896           LYS         0.954         0.958	PHE ILE	LEU	VAL	TRP	TYR	ALA	GLY	THR	SER	ASN	GLN	ASP	GLU	HIS	ARG	LYS	PRO	ALL
0.895         1.000           0.916         0.966         1.000           0.886         0.977         0.976           0.887         0.970         0.968           0.839         0.960         0.968           0.852         0.976         0.927           0.887         0.976         0.927           0.889         0.980         0.945           0.871         0.982         0.934           0.875         0.965         0.913           0.876         0.965         0.913           0.875         0.966         0.913           0.876         0.966         0.913           0.871         0.966         0.913           0.880         0.946         0.880           0.912         0.984         0.965           0.881         0.986         0.913           0.882         0.947         0.896           0.882         0.974         0.896           0.882         0.974         0.896																		
0.916         0.966         1.000           0.886         0.977         0.976           0.901         0.984         0.976           0.885         0.960         0.963           0.815         0.968         0.986           0.852         0.976         0.927           0.880         0.986         0.945           0.871         0.982         0.936           0.875         0.965         0.913           0.850         0.966         0.913           0.815         0.966         0.913           0.816         0.986         0.913           0.817         0.984         0.965           0.818         0.966         0.913           0.828         0.947         0.896           0.828         0.947         0.896           0.828         0.974         0.896																		
0.886         0.977         0.976           0.901         0.984         0.976           0.886         0.970         0.963           0.839         0.968         0.986           0.852         0.976         0.927           0.883         0.846         0.804           0.889         0.982         0.934           0.871         0.982         0.934           0.887         0.962         0.913           0.850         0.969         0.919           0.815         0.946         0.880           0.815         0.946         0.865           0.816         0.984         0.965           0.818         0.946         0.913           0.828         0.946         0.880           0.828         0.947         0.896           0.828         0.947         0.896           0.892         0.974         0.958	_																	
0.901         0.984         0.976           0.886         0.970         0.963           0.839         0.968         0.986           0.852         0.976         0.927           0.883         0.846         0.804           0.899         0.980         0.945           0.871         0.982         0.934           0.871         0.982         0.913           0.850         0.965         0.913           0.851         0.966         0.913           0.815         0.946         0.880           0.912         0.984         0.965           0.848         0.956         0.913           0.880         0.947         0.896           0.828         0.947         0.896           0.828         0.947         0.896	5 1.000																	
0.886         0.970         0.963           0.839         0.960         0.968           0.915         0.968         0.986           0.882         0.846         0.804           0.882         0.982         0.945           0.887         0.982         0.936           0.871         0.982         0.934           0.850         0.965         0.913           0.859         0.966         0.913           0.912         0.984         0.965           0.912         0.984         0.965           0.848         0.966         0.913           0.828         0.947         0.896           0.828         0.947         0.896           0.829         0.974         0.896	5 0.991	1.000																
0.839         0.960         0.968           0.915         0.968         0.986           0.852         0.976         0.927           0.868         0.846         0.804           0.882         0.982         0.935           0.871         0.982         0.934           0.850         0.965         0.913           0.850         0.969         0.919           0.815         0.946         0.880           0.912         0.981         0.965           0.848         0.956         0.913           0.828         0.947         0.896           0.828         0.947         0.896           0.829         0.974         0.958	3 0.985	0.983	1.000															
0.915         0.968         0.986           0.852         0.976         0.927           0.889         0.846         0.804           0.882         0.982         0.936           0.871         0.982         0.934           0.850         0.965         0.913           0.859         0.966         0.913           0.815         0.946         0.880           0.912         0.984         0.965           0.848         0.956         0.913           0.828         0.947         0.896           0.829         0.974         0.896	8 0.978	0.973	0.961	1.000														
0.852         0.976         0.927           0.868         0.846         0.804           0.899         0.980         0.945           0.871         0.982         0.934           0.850         0.965         0.913           0.859         0.969         0.913           0.815         0.966         0.913           0.912         0.984         0.965           0.848         0.956         0.913           0.828         0.947         0.896           0.829         0.974         0.896	5 0.981	0.980	0.970	0.959	1.000													
0.868         0.846         0.804           0.899         0.980         0.945           0.881         0.982         0.934           0.871         0.982         0.934           0.850         0.965         0.913           0.859         0.946         0.880           0.912         0.984         0.965           0.848         0.956         0.913           0.828         0.947         0.896           0.828         0.974         0.896           0.828         0.974         0.896	7 0.964	0.972	0.955	0.931	0.937	1.000												
0.899         0.980         0.945           0.882         0.982         0.936           0.871         0.982         0.934           0.850         0.965         0.913           0.859         0.969         0.913           0.912         0.984         0.965           0.912         0.984         0.965           0.828         0.947         0.896           0.828         0.974         0.896           0.828         0.974         0.896	t 0.778	0.805	0.787	0.717	0.813	0.819	1.000											
0.882         0.982         0.936           0.871         0.982         0.934           0.850         0.965         0.913           0.859         0.969         0.919           0.815         0.966         0.880           0.912         0.984         0.965           0.848         0.956         0.913           0.828         0.947         0.896           0.892         0.974         0.958	996.0	0.978	0.975	0.937	0.956	0.975	0.835	1.000										
0.871     0.982     0.934       0.850     0.965     0.913       0.859     0.969     0.919       0.815     0.946     0.880       0.912     0.984     0.965       0.848     0.956     0.913       0.828     0.947     0.896       0.829     0.974     0.958	5 0.967	0.977	0.968	0.942	0.944	0.987	0.818	986.0	1.000									
0.850     0.965     0.913       0.859     0.969     0.919       0.815     0.946     0.880       0.912     0.984     0.965       0.848     0.956     0.913       0.828     0.947     0.896       0.892     0.974     0.958	1 0.964	0.974	0.958	0.938	0.944	0.988	0.825	0.981	0.989	1.000								
0.859 0.969 0.919 0.815 0.946 0.880 0.912 0.984 0.965 0.848 0.956 0.913 0.828 0.947 0.896	3 0.946	0.962	0.949	0.916	0.927	0.981	0.834	0.980	0.973	0.978	1.000							
0.815 0.946 0.880 0.912 0.984 0.965 0.848 0.956 0.913 0.828 0.947 0.896 0.892 0.974 0.958	0.958	0.968	0.962	0.928	0.927	986.0	0.813	0.982	0.987	0.990	0.984	1.000						
0.912 0.984 0.965 0.848 0.956 0.913 0.828 0.947 0.896 0.892 0.974 0.958	0.924	0.942	0.923	0.890	968.0	0.979	0.810	0.961	0.965	0.978	0.988	0.984	1.000					
0.848     0.956     0.913       0.828     0.947     0.896       0.892     0.974     0.958	5 0.975	0.985	0.979	0.954	0.971	996.0	0.846	0.985	726.0	0.979	996.0	0.973	0.946	1.000				
0.828     0.947     0.896       0.892     0.974     0.958	3 0.937	0.955	0.932	0.912	0.928	0.968	0.833	0.972	0.959	0.970	0.991	696.0	0.979	0.960	1.000			
0.892 0.974 0.958	5 0.917	0.941	0.918	0.900	0.909	0.956	0.821	0.967	0.952	0.964	0.987	0.964	0.979	0.952	0.995	1.000		
	3 0.980	0.980	926.0	0.958	0.963	0.972	0.827	0.974	0.979	0.970	0.955	0.969	0.937	926.0	0.947	0.931	1.000	
ALL 0.890 0.990 0.953	3 0.978	0.987	0.977	0.951	0.961	0.992	0.841	0.991	0.993	0.993	0.985	0.991	0.974	0.988	0.974	0.965	0.985	1.000