Chinese Science Bulletin 2006 Vol. 51 No. 21 2559—2565

DOI: 10.1007/s11434-006-2172-3

The orientational preferences of backbones of proteins

CHEN Ying, WANG Jun, XUE Bin & WANG Wei

National Laboratory of Solid State Microstructure, Department of Physics, and Institute of Biophysics, Nanjing University, Nanjing 210093, China

Correspondence should be addressed to Wang Wei (email: wangwei@nju.edu.cn)

Received November 9, 2005; accepted January 25, 2006

Abstract The orientation between the backbone residues of proteins is defined based on the local configurations and the corresponding preferences are analyzed by statistics. It is found that all the residue pairs have some specific preferences of orientations. The statistical analysis is mainly concentrated in the orientational distributions for two kinds of groupings of residues based on the hydrophobicity and secondary structural features. The statistics for such two types of groupings shows different orientational preferences. It is found that for the former grouping the orientational preference is rather weak, while for the later a kind of strong orientational preferences. This suggests that the formation of local structures and of secondary structures are highly related to the orientational preferences.

Keywords: residues, native structures of proteins, orientational preferences of backbones.

Tertiary structures of proteins are essential for their physical properties and biological functions[1,2]. One of the popular characterizations for protein structures is the contact map[3−6] which records the neighborhood of every residue and could be used to rebuild the main topology of protein structures. The relationship between residues is reduced as a map of distances between the residues. In the studies on various features of proteins, distance-dependent interactions are widely used based on the description of contact map. In fact, due to the anisotropy of residues and the interactions between the atoms, the orientations of residues are not randomly distributed[7]. The realistic interaction between a pair of residues would be a mixture of distance- and orientation-dependent interactions. In most previous models, the orientational features are seldom explicitly considered. Such problems have attracted some interests and a kind of orientation-dependent statistical potentials is proposed[7,8]. Indeed, proteins experience a lot of kinetic stages during their folding, such as the hydrophobic collapse, the formation of secondary structures, and further packing of secondary structures. What kind of role do the orientation-dependent interactions play in protein folding process? Are the orientation-dependent interactions important for every stage of folding? These are not well clarified.

Here we report a statistical study on the orientations between residues in proteins, which may shed some lights on the problems mentioned above. To consider various interactions, two types of groupings on 20 types of amino acid residues based on the hydrophobicity and secondary structural features are made. The statistics for these two groupings shows different preferences of orientations. The secondary structural features are related to strong orientation alignment. This suggests that the formation of local structures and of secondary structures are highly related to the orientation-dependent interactions.

1 Model and methods

As a simplified description for protein molecules, the backbone trace of protein chain outlines the basic geometry of the concerned molecules. The side chains of proteins can be partly rebuilt from the backbone information. In other words, the configuration of backbone atoms is a results of the intra-molecular interactions. As well known, the contact map is often adopted as a description of the interaction in protein molecules. Thus, in our statistics, only the residue pairs in contact are considered, that is, only the interacted residues are considered. Here a contact is defined when two residues *i* and *j* have at least a pair of neighboring non-hydrogen atoms whose distance is smaller than 5.0 Å[6,9]. The orientation between any a pair of contacted residues is defined based on the configurations of the backbone atoms of the two residues since the backbone outlines the basic topology of a protein. Quantitatively, the orientation is measured by projecting one local reference frame (LRF) to the other (see Fig. 1). Here the LRF is an orthogonal frame constructed by the local configurations of atoms, with C_α at its origin, the direction $\overrightarrow{C_\alpha - N}$ as the *X* axis, and the direction of normal vector of plane $C-C_\alpha-N$ (namely $\overrightarrow{C_\alpha - N}\times\overrightarrow{C_\alpha - C}$) as *Z* axis (as shown in Fig. 1(a)). The inter-residue rela-

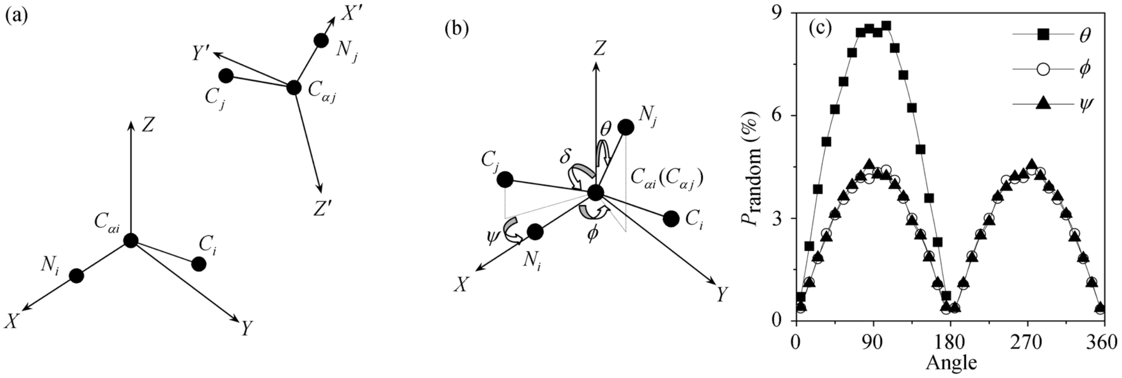

Fig. 1. (a) The definition of local reference frame (LRF) based on the backbone atoms C_α, C, and N. The X axis is along the direction of the vector $\overrightarrow{C_\alpha - N}$. The Z axis is defined as the direction of the normal vector of the plane by three atoms C_α, C, and N. The Y axis is the cross product of unit vectors of Z and X axes. (b) The definition of the orientational angles by projection one LRF to the other. The orientation of residue j related to residue i is defined as the triplet (θ, φ, ψ) based on the projection. (c) The distributions of the angles θ, φ, and ψ corresponding to random orientation.

tionship could be characterized by the LRF located at each residue. Similar to the relationship between rigid bodies, there are 3 degrees of freedom for the relative orientation between two residues. Here we take three angles θ, φ, and ψ (as shown in Fig. 1(b)) to describe the relationship between two residues with

$$\cos\theta = (\vec{X})_j \cdot (\vec{Z})_i, \quad \cos\varphi = (\vec{X})_j \cdot (\vec{X})_i / \sin\theta,$$

$$\cos\psi = (\overrightarrow{C_\alpha - C})_j \cdot (\vec{X})_i / \sin\delta,$$

where $\cos\delta = (\overrightarrow{C_\alpha - C})_j \cdot (\vec{Z})_i$.

For a certain pair of residues with the sequence indices i and j ($j > i$ in the sequence), only the projection from j to i is considered to avoid the double counting. Besides, two kinds of contacts with different sequential distances $|j-i|$ are considered. One is the case with the condition $|j-i| \geqslant 4$ which corresponds to the definition of regular contacts (termed as RC), and the other considers the case of long-range contacts with $|j-i| \geqslant 10$ (termed as LRC). The results of statistics for these cases are compared in our work to illustrate the effects of the local connectivity.

Based on the above definition, 500 non-homologous proteins which are randomly selected from ASTRAL set[10] are evaluated. It is checked that all the statistics is not sensitive to the selection of the proteins. The orientational angles θ, φ, and ψ are discretized into a number of bins each with the width of 10°. Generally, the distribution of relative orientations between residues with types I and J can be described as the probability $p_{I,J}(\theta,\varphi,\psi)$, which is calculated by the formula $n_{I,J}(\theta,\varphi,\psi)/N_{I,J}$, where $n_{I,J}$ and $N_{I,J}$ represent the counts of residue pairs with a certain orientation (θ,φ,ψ) and of all possible pairs between residues of type I and J. It is worth noting that the probability distribution for two independent residues is not uniform as shown in Fig. 1(c). More detailedly, for a uniform distribution of the vector X_j in the corresponding LRF (as shown in Fig. 1(a)), the corresponding function $\cos\theta$ is also uniformly distributed. Therefore, the distribution density of the angle θ has the form $|\sin\theta|$. Similarly, the distribution densities of other two angles have forms as $|\sin\varphi|$ and $|\sin\psi|$, respectively. To illustrate the orientational preferences of residues, this kind of background distribution should be removed from the distribution $p_{I,J}$ as the quantity $\Delta p_{I,J}$ in our work.

2 Results and discussion

Based on our statistics, the pairs of various types of residues all have preferential orientations, that is, between any two types of residues, they do have some orientations (θ,φ,ψ) with high occurrence frequencies. This means that the orientation-dependent interactions actually exist in the packing of proteins. As an example, the distribution $p_{I,J}$ of the orientations (θ,φ,ψ) for residue pair Glu-Glu is shown in Fig. 2. To clearly show the distributions, some projections to the planes of $\theta-\varphi$, of $\theta-\psi$, and of $\varphi-\psi$ are made as shown in Fig. 2(a), (b), and (c), respectively. In these

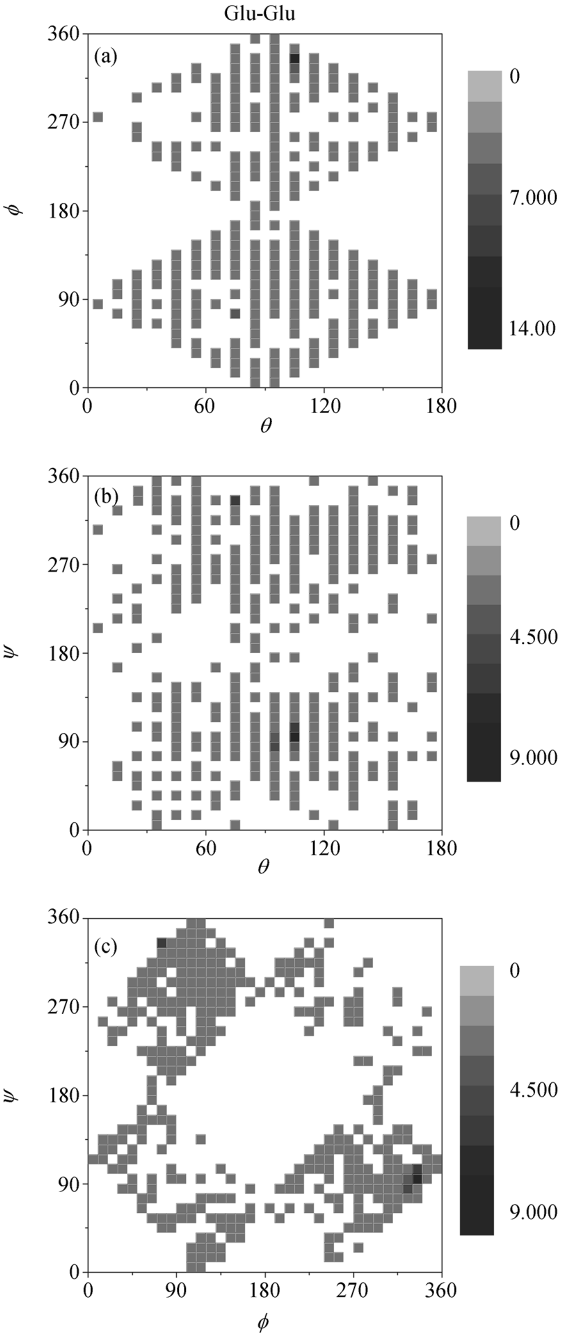

Fig. 2. The projection of orientational distribution to (a) $\theta-\varphi$ plane, (b) $\theta-\psi$ plane, (c) $\varphi-\psi$ plane. The region with darker color has more population. Due to the correlation of the angles, not all the combinations of the angles are possible, as the blank regions in the figure.

plots, there are some black zones which correspond to high occurrence frequencies of orientations, with $\theta\approx 100°$, $\varphi\approx 330°$, and $\psi\approx 100°$. The difference between the black regions and the gray regions of the background indicates that these orientations are not fluctuations, showing clearly the existence of the preferential orientations. Besides, there are some zones with light color around the angles of $\theta\approx 80°$, $\varphi\approx 70°$, and 330°. These values of angles are approximately complementary or equal to the values related the above black zones, which suggests that the zones with light color may be related to the inversion of the backbone directions of the black zones. The difference of the populations of these zones may reflect the preferences of different chiralities. For other kinds of residue pairs, the distributions have the similar features, though the preferential orientations are quantitatively different (unpublished data). Therefore, the orientational preferences between residues are not trivial features in protein systems.

To further understand the physics related to the orientational preferences, it is necessary to utilize the detailed information of various residue pairs. Here, coarse-grained analysis based on the groupings of the amino acid residues is made to capture some universal features of residues in proteins. It is worth pointing out that the grouping of residues is often related to a certain kind of interactions. Therefore, the extent of the orientation-related component of a kind of interactions may be detected from the statistics based on the groupings related to the interactions. In other words, the statistics based on a proper kind of grouping of residues may help us to understand the role that orientational factor plays in protein systems. In this paper, we mainly concentrate on the groupings based on the hydrophobicity of the residues and on the secondary structures of proteins since the hydrophobic interaction and secondary-structure-related interactions contribute essentially to the folding process. Here the grouping to consider the hydrophobicity refers to HP grouping in the reference[11]. The residues are grouped into H or P group based on the types of residues, i.e., residues (F, W, Y, C, M, V, I, L) and (H, A, G, T, P, S, Q, N, R, D, E, K) are classified into the H and P groups, respectively. The grouping of secondary structures involves the type α-, β-, and coil structures, based on the definition of the database DSSP[12]. Different from the HP grouping, the amino acid residues are allocated into the α, β, or coil group based on their structural features in native state. This also reflects the difference between two kinds of interactions. After grouping, the effective number of residues is 2 (for HP grouping) and 3 (for secondary-structure-related grouping).

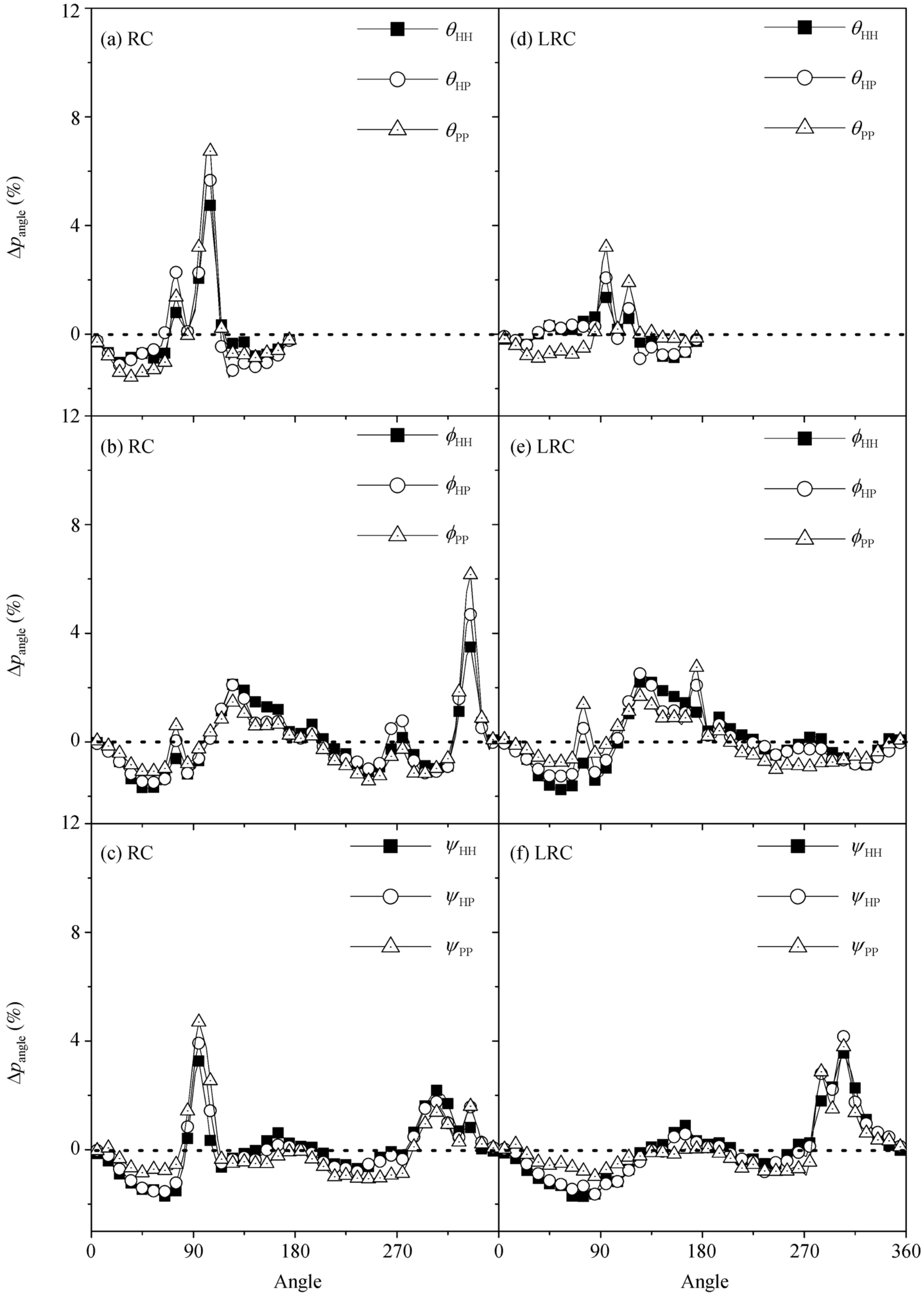

Fig. 3. The 1-d projection of orientational distribution of HP grouping for the cases (a) projected to θ for the RCs with , (b) projected to φ for the RCs with , (c) projected to ψ for the RCs with , (d) projected to θ for the LRCs with , (e) projected to φ for the LRCs with , (f) projected to ψ for the LRCs with . Here the background distribution of random orientation is subtracted.

Considering the HP grouping, the statistics similar to that for individual residue pair are made. The distributions are presented in Fig. 3. To have a better view, all the distributions are projected to one dimension of θ, of φ, or of ψ. Interestingly, all distributions are almost the same with a little fluctuation for three possible pairs (namely HH, HP, and PP), both for LRCs and RCs. This suggests that the orientation-dependent interactions are insensitive to the hydrophobic type. The independence of preference on the residue type also implies that this kind of distribution reflects the basic orientational preferences of any pair of residues. For the LRCs, there are some moderate preferences (regularly not greater than 4%) for all three angles. The preferential angles are 90°, 120°, and 330°. The bonds of two residues seem to prefer the parallel configurations. For the RCs, the preferential angles locate at 100°, 330°, and 90°, different from that of the LRCs. The related peaks are obviously larger than those of LRCs. This kind of enhancement could be attributed to the reason that the local connectivity contributes more importantly to the local orientations rather than that by the hydrophobic interaction. These observations suggest that there are some minor contents in the hydrophobic interaction, but this kind of component is type-independent. These features are consistent with the entropy feature of the hydrophobicity.

For the grouping based on the secondary structural features (Fig. 4), the distributions of orientations are totally different from those for HP grouping. This reflects the difference between the hydrophobicity and the secondary structural feature. For the LRCs, there are no orientational preferences for pairs of two residues both with α configurations. Those small peaks could be attributed to the fluctuations. For the LRCs, only the residues with large sequential separations are considered in statistics. Therefore, the statistics for the LRCs here refers mainly to the packing of helices. The absence of the preferences between α-type residues implies that the packing of α -helices has nothing to do with the orientation-dependent interactions. This is consistent with the regular idea for the packing of helices. Differently, the distribution for β-β pairs has large

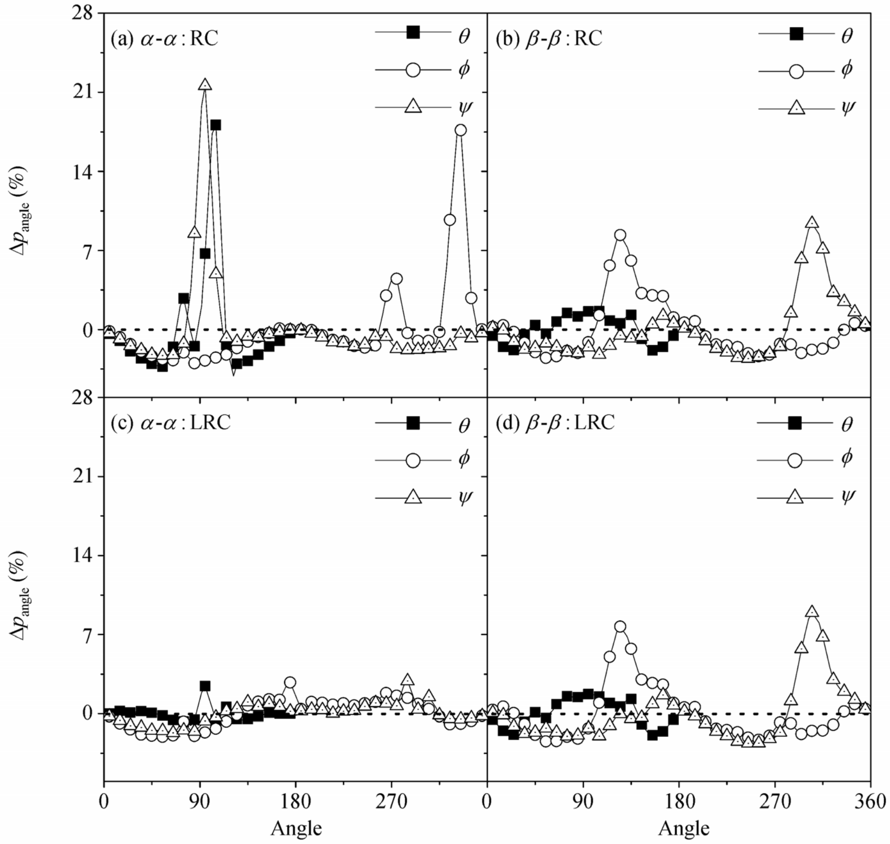

Fig. 4. The 1-d projection of orientational distribution of secondary-structure-related grouping for the cases (a) α-α pairs for the RCs with , (b) β-β pairs for the RCs with , (c) α-α pairs for the LRCs with , (d) β-β pairs for the LRCs with . Here the background distribution of random orientation is subtracted.

peaks at $\varphi \approx 120°$ and $\psi \approx 300°$. The distribution of angle θ is flat with little clear preferences. There is a clear orientational correlation between β-strands. Especially, the preferences of angle φ and ψ reflect the correlation of the trend of strands. This may be related to the the orientational requirement of the interaction of hydrogen bonds between the neighboring strands. This is different from the nonspecific packing of the helices. For the case of RCs, some more residue pairs with short sequential separations are considered. The architecture of secondary structures could be reflected from the statistics. Actually, the strong preference appears for pairs of α-type residues, while the distribution for the pairs of β-type residues changes a little comparing with the LRCs. The emergent preference for cases of α-type residues is a result of the intra-helix hydrogen bonds. Comparing with the LRCs for α-type residues, the orientation-dependent interaction (here the hydrogen bonds) is indeed a source of orientational preferences of the statistics. This is a valuable instance for our assumption on the relation between interactions and orientations. For the cases of β-type residues, the similarity between the distribution for the LRCs and that of RCs suggests that the local properties may contribute similarly to the orientational preferences as those by long-range contacts. In other words, the formation of β-sheet structure may be due to a cooperativity of both the local rigidity and the long-range hydrogen bonds. As a result, the statistics shows a weak dependence on the sequential separation of residues. Based on all these analyses, we conclude that the existence of orientation-dependent interactions could produce a lot of apparent orientational preferences. The importance of hydrogen bonds for orientational preferences in secondary structures is further supported by the investigations on some typical pairs of residues (unpublished data).

It is worth noting that the chemical features of protein chains may present some preferences independent of the type of amino acids, such as the large populations of (i, $i + 4$) in α-class proteins due to local hydrogen bonds. This produces the high peaks in Fig. 4(a). While for the pairs with large sequence separations, these peaks disappear. Therefore, the comparative analysis for RC and LRC cases demonstrates the interplay between directional interaction and the connectivity of protein chains.

Based on our statistics, different interactions may induce different orientational preferences. Therefore, the orientational order would not be formed uniformly in various stages during protein folding. During the process of hydrophobic collapse, there would be little structures with specific orientations. The orientational order is not so important during this stage. Nevertheless, during the grown up of secondary structures, some specific orientations between residues emerges due to the related directional interactions. The secondary structures are built up following the accumulation of this kind of local orientational bindings. Thus, the formation of secondary structures is a process of establishment of the orientational order. Here, the orientational order includes both local factor (the formation of α-helices) and long-range ingredient (the formation of β-sheets), which are all the result of the orientation-dependent interaction (hydrogen bonds). In this sense, the observed orientational preferences of proteins probably form accompanied with the formation of secondary structures.

3 Conclusions

In conclusion, we construct a local reference frame based on the local geometry of the residues. The relative orientation between a pair of residues is defined by projecting their related frames together. A series of statistics on the orientations between some certain kinds of residue pairs are made. It is found that the preferences between the hydrophobic and hydrophilic residues are weak and the preferences between some residues with some specific secondary structures are strong. These illustrate that the occurrence of the orientational preferences is mainly related to the formation of secondary structures.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 90403120, 10474041, 10204013 and 10021001) and FANEDD to Wang Jun.

References

1 Branden C I, Tooze J. Introduction to Protein Structure. 2nd ed. New York: Garland Publishing Inc., 1998. 3—12

2 Fersht A. Structure and Mechanism in Protein Science. New York: WH Freeman and Company, 1998

3 Levitt M. A simplified representation of protein conformations for rapid simulation of protein folding. J Mol Biol, 1976, 104: 59—107

4 Skolnick J, Kolinski A. Monte Carlo approaches to the protein

folding protein. Adv Chem Phys, 1999, 105: 203－242

5 Zhou Y Q, Karplus M. Interpreting the folding kinetics of helical proteins. Nature, 1999, 401: 400－403

6 Chen J, Wang J, Wang W. Transition states for folding of circular-permuted proteins. Proteins-Struct Funct Bioinf, 2004, 57: 153－171[DOI]

7 Buchete N V, Straub J E, Thirumalai D. Development of novel statistical potentials for protein fold recognition. Curr Opin Struct Biol, 2004, 14: 225－232[DOI]

8 Buchete N V, Straub J E, Thirumalai D. Anisotropic coarse-grained statistical potentials improve the ability to identify nativelike protein structures. J Chem Phys, 2003, 118: 7658－7671[DOI]

9 Chavez L L, Onuchic J N, Clementi C. Quantifying the roughness on the free energy landscape: Entropic bottlenecks and protein folding rates. J Am Chem Soc, 2004, 126: 8426－8432[DOI]

10 Chandonia J M, Hon G, Walker N S, et al. The ASTRAL compendium in 2004. Nucl Acids Res, 2004, 32: D189－D192[DOI]

11 Wang J, Wang W. A computational approach to simplifying the protein folding problem. Nature Struct Biol, 1999, 6: 1033－1038[DOI]

12 Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers, 1983, 22: 2577－2637[DOI]