Title	Effects of side-chain packing on the formation of secondary structures in protein folding.
Author(s)	Yasuda, Satoshi; Yoshidome, Takashi; Oshima, Hiraku; Kodama, Ryota; Harano, Yuichi; Kinoshita, Masahiro
Citation	The Journal of chemical physics (2010), 132(6)
Issue Date	2010-02-14
URL	http://hdl.handle.net/2433/108226
Right	© 2010 American Institute of Physics
Type	Journal Article
Textversion	publisher

Effects of side-chain packing on the formation of secondary structures in protein folding

Satoshi Yasuda,[1] Takashi Yoshidome,[1] Hiraku Oshima,[1] Ryota Kodama,[1] Yuichi Harano,[2] and Masahiro Kinoshita[1,a)]

[1]*Institute of Advanced Energy, Kyoto University, Uji, Kyoto 611-0011, Japan*
[2]*Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan*

We have recently shown that protein folding is driven by the water-entropy gain. When the α-helix or β-sheet is formed, the excluded volumes generated by the backbone and side chains overlap, leading to an increase in the total volume available to the translational displacement of water molecules. Primarily by this effect, the water entropy becomes higher. At the same time, the dehydration penalty (i.e., the break of hydrogen bonds with water molecules) is compensated by the formation of intramolecular hydrogen bonds. Hence, these secondary structures are very advantageous units, which are to be formed as much as possible in protein folding. The packing of side chains, which leads to a large increase in the water entropy, is also crucially important. Here we investigate the roles of the side-chain packing in the second structural preference in protein folding. For some proteins we calculate the hydration entropies of a number of structures including the native structure with or without side chains. A hybrid of the angle-dependent integral equation theory combined with the multipolar water model and the morphometric approach is employed in the calculation. Our major findings are as follows. For the structures without side chains, there is an apparent tendency that the water entropy becomes higher as the α-helix or β-sheet content increases. For the structures with side chains, however, a higher content of α-helices or β-sheets does not necessarily lead to larger entropy of water due to the effect of the side-chain packing. The thorough, overall packing of side chains, which gives little space in the interior, is unique to the native structure. To accomplish such specific packing, the α-helix and β-sheet contents are prudently adjusted in protein folding. © *2010 American Institute of Physics*. [doi:10.1063/1.3319509]

I. INTRODUCTION

Protein folding causes a substantially large loss of the conformational entropy. Following the anticipation by Pauling *et al.*,[1] many researchers have regarded the formation of intramolecular hydrogen bonds as a factor exceeding the conformational-entropy loss. However, it accompanies serious dehydration penalty, that is, the break of hydrogen bonds with water molecules. Terazima *et al.*[2] developed a novel experimental technique which enables us to directly measure the enthalpic change upon protein folding at a prescribed temperature. They showed that apoplastocyanin (apoPC) folding at 298 K, which occurs with both of the pressure and the system volume almost unchanged, accompanies a large enthalpic loss. Their result indicates that the folding leads to a great gain in the water entropy, which surpasses the enthalpic loss plus the conformational-entropy loss.

According to the usual view,[3] the water adjacent to a nonpolar group is entropically unstable, and protein folding is driven by the release of such unfavorable water to the bulk through the burial of nonpolar groups. We have recently shown that the entropic gain originating from this view is too small to elucidate the water-entropy gain manifesting the apoPC-folding data.[4] Upon protein folding, the water crowd-

ing is largely reduced primarily by the decrease in the excluded volume (EV) (i.e., the volume of the space which the centers of water molecules cannot enter) followed by the increase in the total volume available to the translational displacement of water molecules.[4–9] (In a strict sense, the water-entropy change is affected not only by the EV but also by three more geometric measures as described in Sec. II B.) Namely, the folding leads to the increases in the number of accessible translational configurations of water and in the water entropy. The usual view considers only the water in the close vicinity of the protein surface, while ours does the water within a far larger length scale.[4–9] The great gain in the water entropy upon apoPC folding experimentally observed can quantitatively be reproduced only by the theoretical method based on our view.[4]

It has been argued for a solute with tubular shape that the formation of the helical structure by a tube or the lateral contact of tubes leads to a reduction of the EV.[10–12] This argument can be applied to the backbone of a protein. The formation of the helical structure by a portion of the backbone gives the α-helix [Fig. 1(a)]. The lateral contact of portions of the backbone occurs in the formation of the β-sheet [Fig. 1(b)]. Thus, the formation of these secondary structures leads to a reduction of the EV followed by a corresponding gain in the water entropy. From this standpoint, some theoretical studies[13–16] with tube models have been car-

a)Author to whom correspondence should be addressed. Electronic mail: kinoshit@iae.kyoto-u.ac.jp.
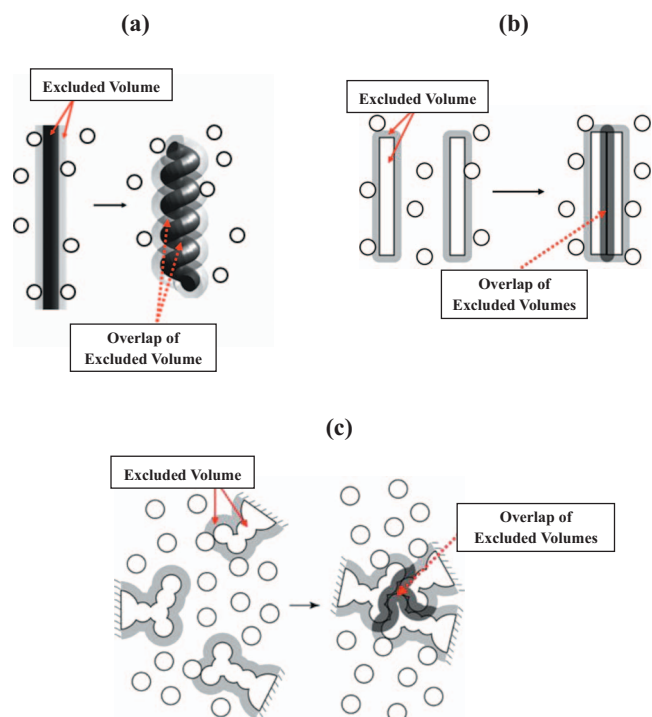
(a)

(b)

(c)

FIG. 1. (a) Formation of the helical structure by a portion of the backbone. (b) Lateral contact of portions of the backbone. (c) Close packing of side chains. EVs overlap [in (a) the overlap of an EV occurs], and the total volume available to the translational displacement of water molecules increases by this amount.

ried out to examine which of the α-helix and the β-sheet is more stabilized when the density and molecular diameter of the solvent are varied as important parameters. However, the effect of the side-chain packing specific to their geometric features cannot be taken into consideration in the tube models. The presence of a side chain generates an EV. When side chains are closely packed, a remarkable overlap of the EVs generated by them occurs [Fig. 1(c)], leading to a large gain in the water entropy.[4–9]

The driving force in protein folding is not the formation of intramolecular hydrogen bonds but the water-entropy gain. However, we never claim that the former is unimportant. It is imperative in compensating the dehydration penalty during the folding process.[4–9] The α-helix or β-sheet formation accompanies the overlap of the EVs generated by the side chains as well as by the backbone, and it leads to a large gain in the water entropy. At the same time it compensates the dehydration penalty by forming the intramolecular hydrogen bonds. Hence, these secondary structures are very advantageous units which are to be formed as much as possible, and it is no wonder that they frequently appear in the native structures.

In the present article, we investigate the roles of the side-chain packing in the second structural preference in protein folding. For some proteins we calculate the hydration entropies of a number of structures including the native structure with or without side chains. A hybrid of the angle-dependent integral equation theory[17–28] combined with the multipolar water model[29,30] and the morphometric approach[31,32] is employed in the calculation. All residues in each structure are replaced by glycine (Gly) (the side chain

of Gly is only a small hydrogen atom) to mimic the structure without side chains. We find that for the structures without side chains, there is an apparent tendency that the water entropy becomes higher as the α-helix or β-sheet content increases. For the structures with side chains, however, a higher content of α-helices or β-sheets does not necessarily lead to larger entropy of water due to the effect of the side-chain packing. Even when the content of the secondary structures is higher, the stability in terms of the water entropy can be lower if the packing of side chains is less effective. In the native structure, there is little space in its interior due to the thorough, overall packing of side chains, which is not shared by any other structure. To accomplish such specific packing, the α-helix and β-sheet contents are prudently adjusted in protein folding.

On the basis of the results of our investigation, we discuss a relevance to the following experimental results observed for some β-sheet proteins such as β-lactoglobulin.[33,34] A highly α-helical intermediate is stabilized in the earlier stage of the folding process, and then it is reconstructed toward the predominantly β-sheet structure with the thorough, overall packing of side chains in the later stage.

II. MODEL AND THEORY

A. Models of water and protein

We employ a multipolar model for water.[29,30] A water molecule is modeled as a hard sphere with diameter $d_S = 0.28$ nm in which a point dipole and a point quadrupole of tetrahedral symmetry are embedded. The effect of the molecular polarizability is taken into account using the self-consistent mean field (SCMF) theory.[29,30] At the SCMF level the many-body induced interactions are reduced to pairwise additive potentials involving an effective dipole moment.

The hydration entropy S is determined principally by the EV effect originating from the translational displacement of water molecules that coexist with a protein and not significantly dependent on the protein-water interaction potentials. Imai *et al.*[35] considered the native structures of a total of eight peptides and proteins and calculated S using the three-dimensional reference interaction site model (3D-RISM) theory combined with the all-atom [Coulomb plus Lennard-Jones (LJ)] potentials and the SPC/E water model. Even when the protein-water electrostatic potentials, which are quite strong, are shut off and only the LJ potentials are retained, $|S|$ decreases merely by less than 5%. Therefore, a protein can be modeled as a set of fused hard spheres.

The hydration free energy μ, entropy S, and energy U under the isochoric condition are calculated for a spherical solute with diameter 0.28 nm using the angle-dependent integral equation theory[17–28] combined with the multipolar water model.[29,30] For the hard-sphere solute with zero charge, the result is $\mu = 5.95 k_B T$, $S = -9.22 k_B$, and $U = -3.27 k_B T$. When the point charge $-0.5e$ (e is the electronic charge) is embedded at its center, the result changes as the following: $\mu = -32.32 k_B T$, $S = -10.11 k_B$, and $U = -42.43 k_B T$. Thus, S is fairly insensitive to the solute-water interaction potential
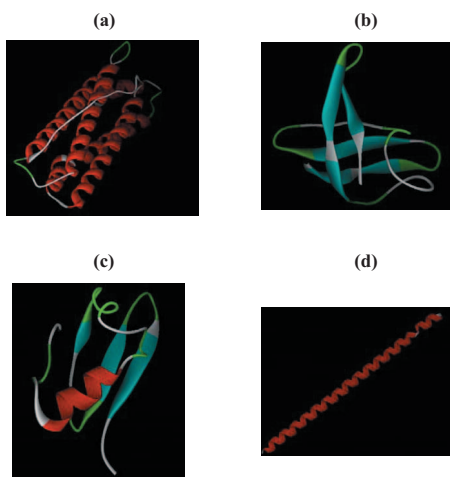
(a)

(b)

(c)

(d)

FIG. 2. Ribbon representation for the native structures of human H chain ferritin (a) and CspB (b), and that for a non-native structure (c) and "all α" (d) of CspB. They are drawn using the DS visualizer 2.0.

while μ and U are largely influenced by it. This gives another justification of the protein model, a set of fused hard spheres.

We test some proteins including human H chain ferritin (PDB Code: 2fha, 172 residues) and the major cold shock protein of bacillus subtilis, CspB (PDB Code: 1csp, 67 residues). The native structure of human H chain ferritin possesses only the α-helices while that of CspB does only the β-sheets. We consider a number of non-native structures taken from the Rosetta decoy set.[36] In order to conform the number of residues of the native structure to that of the non-native structures, we removed two small end portions of the native structure (12 and 3 residues for human H chain ferritin and for CspB, respectively). The portions are not buried, and they do not participate in the α-helix or β-sheet formations. Therefore, the removal has no significant effect on the structural characteristics.

The α-helix and β-sheet contents of a structure are important parameters in the present investigation. The DSSP program[37] is employed for calculating the contents. For the native structures, the α-helix content of human H chain ferritin is ~76% and the β-sheet content of CspB is ~47%. For each protein, the complete α-helix structure is generated using the TINKER program package[38] maintained at "http://dasher.wustl.edu/tinker/" with the dihedral angles (φ, ψ, and ω) set at ($-60°$, $-45°$, and $180°$) (hereafter, this structure is referred to as "all α"). The ribbon representations for the native structures of the two proteins, a non-native structure of CspB, and "all α" of CspB are shown in Fig. 2. The LJ potential energy for the non-native structures and "all α" is positive and quite large due to the unrealistic overlaps of protein atoms. Such overlaps are removed by local minimization of the energy function using the CHARMM biomolecular simulation program[39] through the multiscale modeling tools in structural biology (MMTSB) program.[40] The minimization is performed so that the original structures can be retained as much as possible. We employ the CHARMM22 (Ref. 41) with the CMAP correction[42] as the force-field parameters. Electrostatic and nonbonded interactions are all evaluated without any cutoff, and the Generalized-Born

(GBMV/SA) approximation[43–45] is employed for the electrostatic part of the hydration energy. After the minimization, each structure is switched to a set of fused hard spheres for calculating the hydration entropy. Now there are no serious overlaps of protein atoms.

It is physically insightful to separate the effect of side chains from those of the backbone. To this end, in each structure all residues are replaced by Gly using the CHARMM and MMTSB programs.[39,40] The replacement is carried out after the minimization mentioned above. The structure thus made has essentially no side chains (hereafter, these are referred to as "the structures without side chains"). For each protein we now have the native structure, a number of non-native structures, and "all α" with or without side chains.

In the calculation of the hydration entropy S, the (x, y, z)-coordinates of protein atoms (N, C, O, H, and S) are used as part of the input data to characterize each structure. The diameter of each atom is set at the sigma value of the LJ potential parameters, which are taken from the CHARMM22.[41] $-S$ represents the loss of the water entropy upon the insertion of a protein with a prescribed structure. The information on the effect of side chains is contained in "$-S$ of a structure with side chains" $-$ "$-S$ of the corresponding structure without side chains (i.e., with the backbone alone)." The latter is denoted by $-S_b$. Our major concern is to investigate the relation between $-S$, $-S_b$, or $-S-(-S_b)$ and the α-helix or β-sheet content.

B. Theoretical method

Since a molecular model is employed for water, the angle-dependent version[17–28] must be used for the integral equation theory, an elaborate statistical-mechanical theory. However, its extension to complex solute molecules such as proteins is rather difficult. This problem can be overcome by combing it with the morphometric approach.[31,32] The approach is applicable only when the solute-solvent interaction potential includes no electrostatic part. There is no problem in the present case where a protein is modeled as a set of fused hard spheres. The idea of the approach is to express a hydration quantity such as S by the linear combination of only four geometric measures of a solute molecule,

$$S/k_B = C_1 V_{ex} + C_2 A + C_3 X + C_4 Y. \tag{1}$$

Here, k_B is Boltzmann's constant, V_{ex} is the EV, A is the water-accessible surface area, and X and Y are the integrated mean and Gaussian curvatures of the accessible surface, respectively. Though $C_1 V_{ex}$ is the principal term, the other three terms also influence S. In the approach, the solute shape enters S only via the four geometric measures. Therefore, the four coefficients (C_1–C_4) can be determined in a simple geometry. They are calculated from the values of S for hard-sphere solutes with various diameters immersed in our model water. The angle-dependent integral equation theory is employed in the calculation. Refer to our earlier publications[4,7–9,32,46–53] for more details.

The dielectric constant of bulk water calculated using the angle-dependent integral equation theory combined with the multipolar water model is ~83 that is in good agreement

with the experimental value ~ 78. As proved in our earlier work,[26] the angle-dependent integral equation theory is far superior to the RISM and related theories[17,54,55] in analyses on the hydrophobic hydration. The former gives a quantitatively accurate value of the hydration free energy of a nonpolar solute. The high reliability of the morphometric approach has already been demonstrated.[4,7–9,32,47] For example, the results from the 3D integral equation theory[56,57] applied to the same model protein immersed in a simple solvent (the solvent particles interact through strongly attractive potential such as water molecules) can be reproduced with sufficient accuracy by the morphometric approach where the four coefficients are determined in the manner explained above. By a hybrid of the angle-dependent integral equation theory combined with the multipolar water model and the morphometric approach, the experimentally measured changes in thermodynamic quantities upon apoPC folding are quantitatively reproduced.[4] Moreover, great progresses have been made in elucidating the microscopic mechanisms of pressure,[46–48] cold,[49,50] and heat[51] denaturating of proteins, and in the prediction of the native structure[9,52,53] by our theoretical methods in which the morphometric approach is combined with the integral equation theory or its angle-dependent version.

III. RESULTS AND DISCUSSION

A. Characteristics of the native structure

We have tested several proteins, but we present the results for human H chain ferritin and CspB. In Fig. 3, $-S/k_B$ is plotted against the radius of gyration R_g for the native structure, "all α," and non-native structures of human H chain ferritin (a) or CspB (b). The loss of the water entropy upon the insertion $(-S)$ for the native structure is much smaller than that for any other structure. This is ascribed primarily to the exceptionally small EV of the native structure in which the backbone and side chains are closely packed with little space in the interior.[58]

In Fig. 4, $-S_b/k_B$ is plotted against R_g for the native structure, "all α," and non-native structures of human H chain ferritin (a) or CspB (b). $-S_b$ denotes the water-entropy loss upon the insertion of a protein in the structure without side chains (i.e., with the backbone alone). There are significantly many structures $-S_b$ of which is smaller than that of the native structure. It is interesting that $-S_b$ for "all α" is the smallest.

In Fig. 5, $-S/k_B-(-S_b/k_B)$ is plotted against R_g for the native structure, "all α," and non-native structures of human H chain ferritin (a) or CspB (b). $-S-(-S_b)$ represents the increase in the water-entropy loss caused by the incorporation of side chains. $-S-(-S_b)$ for the native structure is far smaller than any other structure. In the native structure the incorporation of side chains gives rise to the smallest increase in the EV and the side chains are most closely packed. In particular, the thorough, overall packing of side chains is a feature of the native structure, which is not shared by the other structures considered.
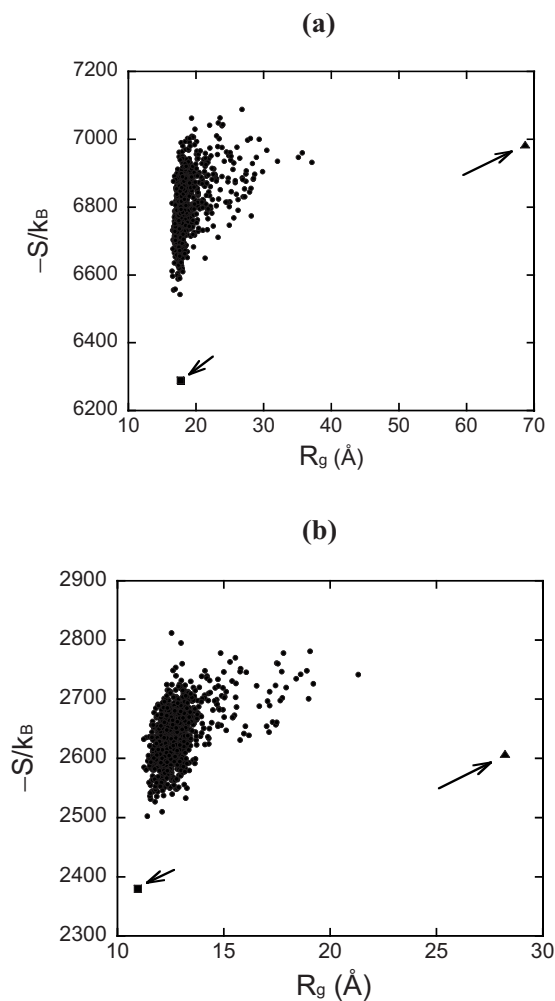

FIG. 3. Relation between the water-entropy loss upon the insertion of a protein with a prescribed structure, $-S$, and the radius of gyration R_g for human H chain ferritin (a) or CspB (b). In the ordinate $-S$ is scaled by k_B. Closed circles: data points for the nonnative structures. Closed square: data point for the native structure (this is indicated by the short arrow). Closed triangle: data point for "all α" (this is indicated by the long arrow).

B. Roles of side-chain packing: α-helix protein

We first consider human H chain ferritin whose native structure possesses only α-helices. In Fig. 6, $-S/k_B$ (a) or $-S_b/k_B$ (b) is plotted against the α-helix content for the native structure, "all α," and non-native structures. The α-helix contents of the non-native structures are roughly in the range 50%–90%, and the β-sheets are not present in most of them. As observed in Fig. 6(a), there is only weak correlation between $-S$ and the α-helix content. By contrast, there is a strong tendency that $-S_b$ becomes smaller as the α-helix content increases, which is obvious from Fig. 6(b). $-S_b$ for "all α" is the smallest.

There are many non-native structures whose details of the side-chain packing are considerably different even when they share almost the same α-helix content. This leads to no appreciable correlation between $-S$ and the α-helix content. An example is illustrated in Fig. 7. The values of $-S/k_B$ for the two non-native structures shown are quite different (6656 and 7089, respectively) despite the same α-helix content, $\sim 70\%$, and only a slight difference in the values of $-S_b/k_B$. The local and nonlocal packing of side chains is significantly
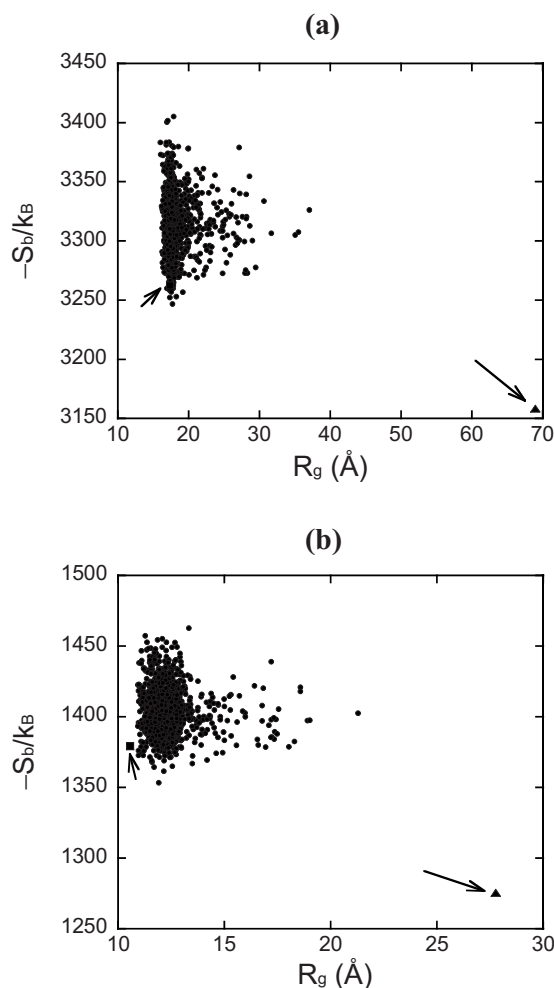
(a)

(b)

FIG. 4. Relation between the water-entropy loss upon the insertion of a protein with the structure all residues of which are replaced by Gly (i.e., the structure with the backbone alone), $-S_b$, and the radius of gyration R_g for human H chain ferritin (a) or CspB (b). In the ordinate $-S_b$ is scaled by k_B. Closed circles: data points for the non-native structures. Closed square: data point for the native structure (this is indicated by the short arrow). Closed triangle: data point for "all α" (this is indicated by the long arrow).
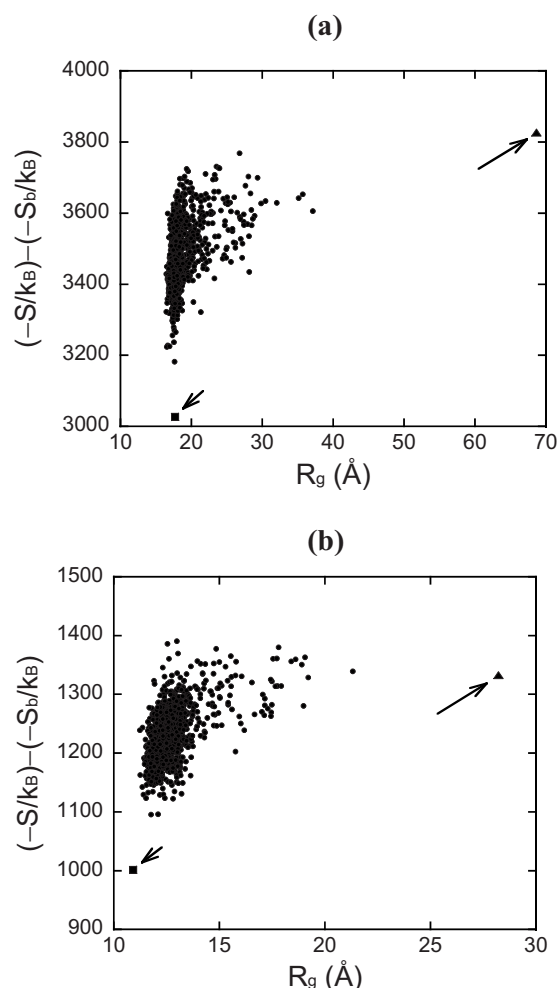

(a)

(b)

FIG. 5. Relation between $-S/k_B - (-S_b/k_B)$ and the radius of gyration R_g for human H chain ferritin (a) or CspB (b). $-S - (-S_b)$ represents the increase in the water-entropy loss caused by the incorporation of side chains. Closed circles: data points for the non-native structures. Closed square: data point for the native structure (this is indicated by the short arrow). Closed triangle: data point for "all α" (this is indicated by the long arrow).

more effective in (a), leading to smaller $-S$. Here, "local" means the packing or contact of nearby side chains while "nonlocal" does the packing or contact of side chains distant in the amino-acid sequence.

An important result is that to form the α-helix as much as possible is the most favorable as far as the backbone is concerned. However, due to the effect of the side-chain packing, a structure with a higher α-helix content is not always more favorable. For the native structure, $-S$ is the smallest [Fig. 6(a)] while $-S_b$ is not [Fig. 6(b)]. The thorough, overall packing of side chains is achieved by the α-helix content $\sim$76% in the case of human H chain ferritin.

C. Roles of side-chain packing: β-sheet protein

We then consider CspB whose native structure possesses only the β-sheets. In Fig. 8, $-S/k_B$ (a) or $-S_b/k_B$ (b) is plotted against the α-helix content for the native structure, "all α," and non-native structures. The α-helix and β-sheet contents of the non-native structures are roughly in the ranges 0%–50% and 0%–70%, respectively, but the number of the

structures with zero β-sheet content is quite small. Most of them have both of α-helices and β-sheets, and a structure with higher α-helix content tends to have lower β-sheet content (a structure with higher β-sheet content tends to have lower α-helix content). Although there is little correlation between $-S$ and the α-helix content [Fig. 8(a)], there is a tendency that $-S_b$ becomes smaller as the α-helix content increases [Fig. 8(b)].

In Fig. 9, $-S/k_B$ (a) or $-S_b/k_B$ (b) is plotted against the β-sheet content for the native structure, "all α," and non-native structures. There is almost no correlation between $-S$ and the β-sheet content, and $-S_b$ and the β-sheet content seem to be only weakly correlated with each other. The reason for this should be the following. The correlation between $-S_b$ and the β-sheet content is disturbed by that between $-S_b$ and the α-helix content. In order to extract the effect of the β-sheet content, only the non-native structures with zero α-helix content are chosen, and $-S/k_B$ (a) or $-S_b/k_B$ (b) is plotted against the β-sheet content in Fig. 10. There is little correlation between $-S$ and the β-sheet content. However, there is a rather strong tendency that $-S_b$ becomes smaller as
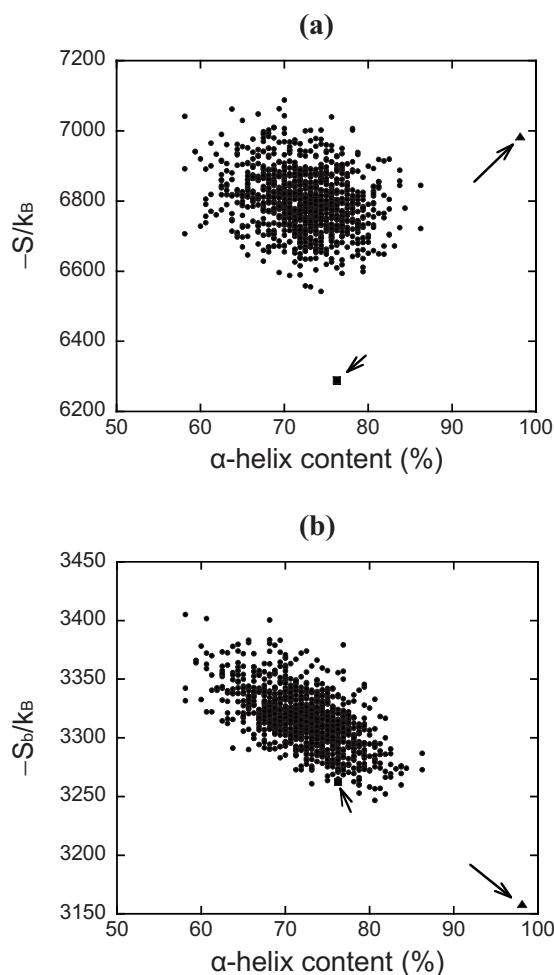
(a)

(b)

FIG. 6. Relation between $-S/k_B$ (a) or $-S_b/k_B$ (b) and the α-helix content for human H chain ferritin. Closed circles: data points for the non-native structures. Closed square: data point for the native structure (this is indicated by the short arrow). Closed triangle: data point for "all α" (this is indicated by the long arrow).
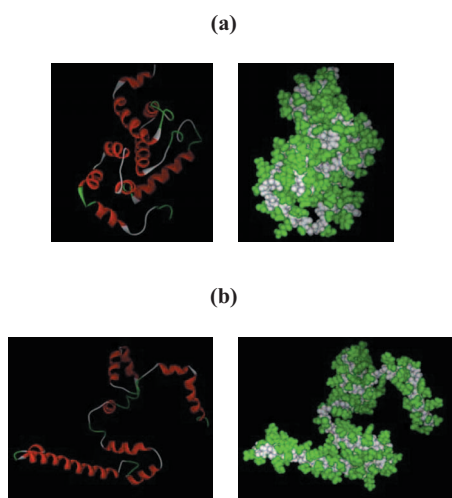
(a)

(b)

FIG. 7. Ribbon and space-filled representations for two non-native structures of human H chain ferritin. They are drawn using the DS visualizer 2.0. In the space-filled representation, the side chains are marked in green. The two structures share the same α-helix content $\sim$70%. The values of $(-S_b/k_B,$ $-S/k_B)$ are (3320, 6656) for (a) and (3294, 7089) for (b).
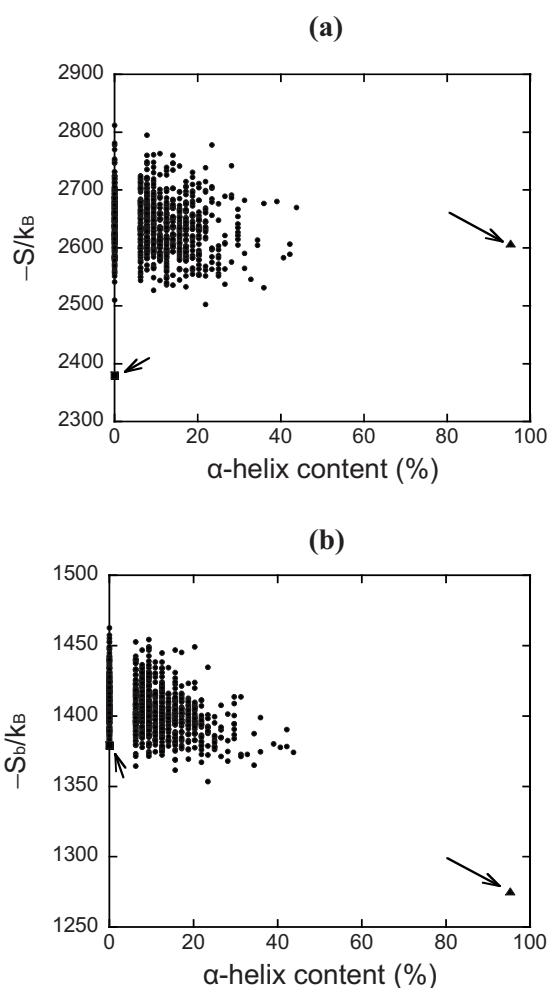
(a)

(b)

FIG. 8. Relation between $-S/k_B$ (a) or $-S_b/k_B$ (b) and the α-helix content for CspB. Closed circles: data points for the non-native structures. Closed square: data point for the native structure (this is indicated by the short arrow). Closed triangle: data point for "all α" (this is indicated by the long arrow).

the β-sheet content increases. The β-sheet formation certainly leads to a reduction of the EV generated by the backbone followed by a corresponding increase in the water entropy.

The values of $-S/k_B$, $-S_b/k_B$ and $(-S/k_B)-(-S_b/k_B)$ of some representative structures are given in Table I. Non-natives 1 and 2 have high α-helix content ($\sim$40%) with no β-sheet and non-natives 3–6 have almost equally high β-sheet content ($\sim$40%) with no α-helix. For "all α," $-S_b$ is the smallest but $(-S)-(-S_b)$ is the largest. $(-S)-(-S_b)$ as well as $-S$ is the smallest for the native structure. We now compare the values for the six non-native structures. In terms of $-S_b$, the values for non-natives 1 and 2 are smaller than those for non-natives 3–6. However, in terms of $(-S)-(-S_b)$ and $-S$, the values for non-natives 1 and 2 are larger than those for non-natives 3 and 4 but smaller than those for non-natives 5 and 6. Thus, the second structural preference is largely influenced by the side-chain packing. In particular, the thorough, overall packing of side chains is crucially important, and it is achieved by the β-sheet content $\sim$47% in the case of CspB.
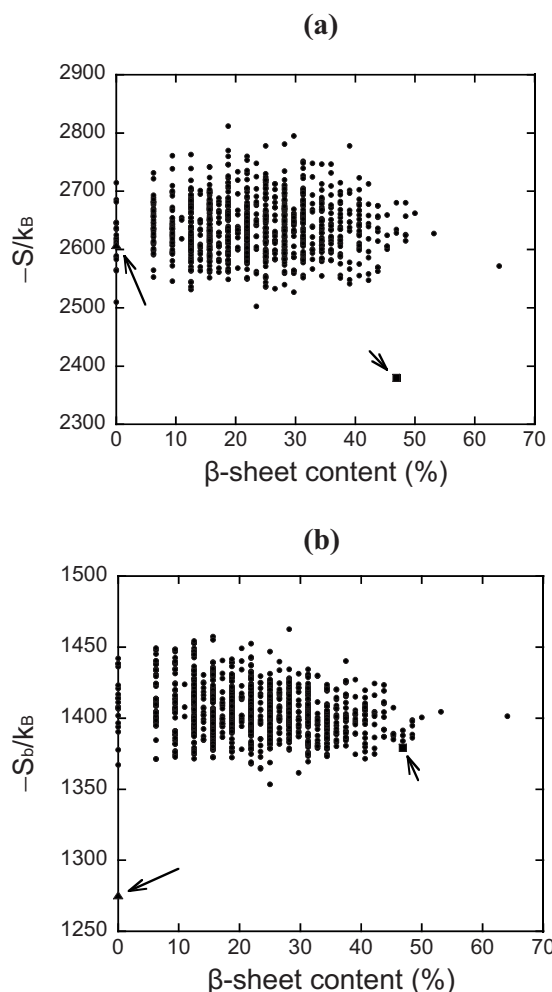
(a)

(b)

(a)

(b)

FIG. 9. Relation between $-S/k_B$ (a) or $-S_b/k_B$ (b) and the β-sheet content for CspB. Closed circles: data points for the non-native structures. Closed square: data point for the native structure (this is indicated by the short arrow). Closed triangle: data point for "all α" (this is indicated by the long arrow).

FIG. 10. Relation between $-S/k_B$ (a) or $-S_b/k_B$ (b) and the β-sheet content for CspB. Only the non-native structures with zero α-helix content are chosen. Closed circles: data points for the non-native structures. Closed square: data point for the native structure (this is indicated by the short arrow).

D. Further discussion

It should be noted that the correlation between the α-helix or β-sheet content and $-S_b$ is independent of the protein species. In Fig. 6(b), most of the non-native structures have only the α-helices. In Figs. 8(b) and 9(b), most of the non-native structures have both of α-helices and β-sheets, and a structure with higher α-helix content tends to have lower β-sheet content (a structure with higher β-sheet content tends to have lower α-helix content). In Fig. 10(b), all the non-native structures chosen have only β-sheets. Among the four figures, the correlation is the strongest in Fig. 6(b). Although in Fig. 8(b) the correlation between the α-helix content and $-S_b$ is somewhat weakened by that between β-sheet content and $-S_b$, the former is still dominant. Thus, for the non-native structures tested in the present study, the α-helices seem to be preferable to the β-sheets as far as the backbone is concerned. To draw a definite conclusion, however, further examination is needed for the non-native structures with wider ranges of the α-helix and β-sheet contents.

The thorough, overall packing of side chains in the protein interior is necessary to ensure that the native structure be
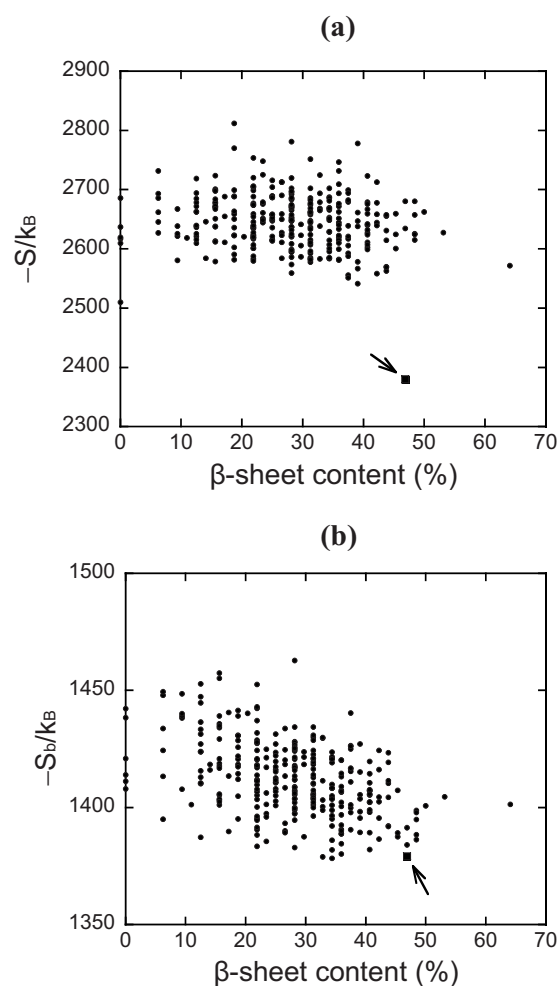
stable and that partially denatured, inactive structures have negligible probability at ambient temperatures.[59] It is also necessary to increase the water entropy as much as possible. Only the amino-acid sequences, with which the specific packing can be realized, have been selected by nature. It is reached by the judicious adjustment of the α-helix and β-sheet contents as follows: by the α-helices alone (e.g., for human H chain ferritin), by the β-sheets alone (e.g., for CspB), or by a mixture of α-helices and β-sheets. The resultant content of the secondary structures is not always high, giving rise to a significantly large total dehydration penalty and an enthalpic loss upon protein folding.[2,4]

As in the usual theoretical methods for proteins, the hydration properties of many different structures fixed are compared to analyze the structural stability in aqueous solution. The TINKER program[38] is used simply to generate the complete α-helix structure, and the CHARMM and MMTSB programs[39,40] are just for removing the unrealistic overlaps of the protein atoms and for replacing all residues by Gly. The DSSP program[37] is necessitated in the calculation of the α-helix and β-sheet contents. The employment of other computer programs can lead to slightly different results in a quantitative sense, but the qualitative aspects of our conclusions are not likely to be altered.

TABLE I. Values of $-S/k_B$, $-S_b/k_B$, and $(-S/k_B)-(-S_b/k_B)$ calculated for the native structure, six non-native structures, and "all α" of CspB. $-S_b$ denotes the water-entropy loss upon the insertion of a protein with the structure all residues of which are replaced by Gly (i.e., the structure with the backbone alone). $-S-(-S_b)$ represents the increase in the water-entropy loss caused by the incorporation of side chains.

	α-helix content (%)	β-sheet content (%)	$-S/k_B$	$-S_b/k_B$	$(-S/k_B)-(-S_b/k_B)$
Native	0	46.9	2380	1379	1000
Non-native 1	40.6	0	2583	1378	1205
Non-native 2	42.2	0	2589	1390	1199
Non-native 3	0	37.5	2551	1417	1135
Non-native 4	0	39.1	2541	1396	1146
Non-native 5	0	39.1	2647	1415	1232
Non-native 6	0	39.1	2661	1404	1258
"All α"	95.3	0	2608	1275	1333

E. Relevance to nonhierarchical properties of protein folding

The folding mechanisms can be categorized into hierarchical and nonhierarchical models.[33,34] Experimental studies have indicated the presence of the intermediate states of protein folding. In the hierarchical model, the intermediate is a compact denatured state with largely flexible side chains but with *nativelike* secondary structures.[33,34,60–62] The folding proceeds sequentially toward the native structure. On the other hand, in the nonhierarchical model, the secondary structures of the intermediate are quite different from those of the native structure.[63] Although the hierarchical model has been shown to be true for many proteins by experiments,[60–62] it is not applicable to some β-sheet proteins. An experimental evidence is that β-lactoglobulin, which comprises antiparallel β-sheets and one α-helix, forms an intermediate with the α-helices alone.[34,64] The intermediate is devoid of the thorough, overall packing of side chains.[33,34] It is attained by complete reconstruction of the secondary structures, resulting in the native β-sheet structure.

The native structure is completed in the final stage where the side chains with a variety of geometrical features are specifically locked.[65] It is completed only by the thorough, overall packing of side chains. In the earlier stage of protein folding, such specific packing does not come into play. Only a decrease in the EV generated by the backbone and the local and nonlocal packing of side chains are promoted. When such promotion is made more effectively by the α-helix formation, the α-helices are preferable to the β-sheets. A highly helical intermediate is formed. This should be the case for β-lactoglobulin. For this protein, however, the ultimate packing of side chains is attained not by the α-helices but by the β-sheets. Therefore, the intermediate must be reconstructed all over again toward the native β-sheet structure. Thus, the nonhierarchical properties of protein folding are relevant to the large effects of the thorough, overall packing of side chains on the formation of secondary structures.

IV. CONCLUDING REMARKS

We have investigated how the secondary structural preference is influenced by the side-chain packing. For some proteins we calculate the hydration entropies of a number of structures including the native structure, and the results for human H chain ferritin and CspB are presented. A hybrid of the angle-dependent integral equation theory[17–28] combined with the multipolar water model[29,30] and the morphometric approach[31,32] is employed in the calculation. For each structure, all residues are replaced by Gly to mimic the structure without side chains. We can thus extract the effect of side chains.

In the significance, the present work is clearly different from our earlier works[5–8] showing that the water-entropy gain is a major driving force in protein folding. Upon the folding, a large gain in the water entropy occurs, but the gain originates from various factors such as the α-helix and β-sheet formations by portions of the backbone, the local and nonlocal packing of side chains, and the thorough, overall packing of side chains taking place in the final stage of the folding process. In the present work we analyze how large the contributions from these factors to the water-entropy gain are, how they are correlated, and how the side-chain packing has influences on the second structural preference.

When the α-helix or β-sheet is formed, the EVs generated by the backbone and side chains overlap, leading to an increase in the total volume available to the translational displacement of water molecules and to a corresponding gain in the water entropy. At the same time, the formation assures the compensation of the dehydration penalty by the intramolecular hydrogen bonds during protein folding. Hence, these secondary structures are very advantageous units and to be formed as much as possible. The packing of side chains (e.g., those in different secondary structures), which results in a large gain in the water entropy, is also crucially important.

For the backbone alone, there is a strong tendency that the water entropy becomes higher as the α-helix or β-sheet content increases. Due to the effect of the side-chain packing, however, a higher content of α-helices, β-sheets, or the secondary structures does not necessarily lead to larger entropy of water. The native structure is characterized by the side chains with a variety of geometrical features specifically locked with little space in its interior. To accomplish such thorough, overall packing of side chains, the α-helix and β-sheet contents are prudently adjusted in the folding process. The content of the secondary structures in the native

structure thus reached is not always high. Moreover, some native structures possess the α-helices or β-sheets alone, and others consist of mixtures of α-helices and β-sheets. The side-chain packing has a substantially large effect on the second structural preference.

We never claim that polyglycine favors the folding to the α-helix or β-sheet. The hydration entropy of polyglycine is calculated just to extract the contribution from the α-helix and β-sheet formation by portions of the backbone to the water-entropy gain upon a transition of the protein structure to a more compact one. For a protein, the water-entropy gain becomes large enough to surpass the conformational entropy loss and the dehydration penalty only after the close packing of the backbone and side chains discussed throughout the present article. For polyglycine with essentially no side chains, the conformational entropy loss and the dehydration penalty predominate over the water-entropy gain. Polyglycine is not foldable.

The nonhierarchical aspect of β-lactoglobulin folding[33,34,64] can be understood as follows. The thorough, overall packing of side chains is the most important for increasing the water entropy as much as possible. Such specific packing, which is achievable by the β-sheets for β-lactoglobulin, occurs in the final stage of protein folding.[65] In the earlier stage, only a decrease in the EV generated by the backbone and the local and nonlocal packing of side chains are promoted. When such promotion is made more effectively by the formation of the α-helices than by that of the β-sheets, a highly helical intermediate can precede the formation of the native β-sheet structure. This should be the case for β-lactoglobulin. The intermediate, which is devoid of the thorough, overall packing of side chains, must be reconstructed all over again toward the native structure. A more detailed study which is specific to β-lactoglobulin is to be carried out to substantiate our claim.

ACKNOWLEDGMENTS

[1] L. Pauling, R. B. Corey, and H. R. Branson, Proc. Natl. Acad. Sci. U.S.A. **37**, 205 (1951).
[2] N. Baden, S. Hirota, T. Takabe, N. Funasaki, and M. Terazima, J. Chem. Phys. **127**, 175103 (2007).
[3] K. A. Dill, Biochemistry **29**, 7133 (1990).
[4] T. Yoshidome, M. Kinoshita, S. Hirota, N. Baden, and M. Terazima, J. Chem. Phys. **128**, 225104 (2008).
[5] Y. Harano and M. Kinoshita, Chem. Phys. Lett. **399**, 342 (2004).
[6] Y. Harano and M. Kinoshita, Biophys. J. **89**, 2701 (2005).
[7] M. Kinoshita, Front. Biosci. **14**, 3419 (2009).
[8] M. Kinoshita, Int. J. Mol. Sci. **10**, 1064 (2009).
[9] T. Yoshidome, K. Oda, Y. Harano, R. Roth, Y. Sugita, M. Ikeguchi, and M. Kinoshita, Proteins **77**, 950 (2009).
[10] M. Kinoshita, Chem. Phys. Lett. **387**, 47 (2004).
[11] M. Kinoshita, Chem. Phys. Lett. **387**, 54 (2004).
[12] M. Kinoshita, Chem. Eng. Sci. **61**, 2150 (2006).
[13] Y. Snir and R. D. Kamien, Science **307**, 1067 (2005).
[14] Y. Snir and R. D. Kamien, Phys. Rev. E **75**, 051114 (2007).
[15] H. Hansen-Goos, R. Roth, K. Mecke, and S. Dietrich, Phys. Rev. Lett. **99**, 128101 (2007).
[16] C. Poletto, A. Giacometti, A. Trovato, J. R. Banavar, and A. Maritan, Phys. Rev. E **77**, 061804 (2008).
[17] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic, London, 2005).
[18] M. Kinoshita and M. Harada, Mol. Phys. **81**, 1473 (1994).
[19] M. Kinoshita, S. Iba, and M. Harada, J. Chem. Phys. **105**, 2487 (1996).
[20] M. Kinoshita and D. R. Bérard, J. Comput. Phys. **124**, 230 (1996).
[21] N. M. Cann and G. N. Patey, J. Chem. Phys. **106**, 8165 (1997).
[22] D. R. Bérard, M. Kinoshita, N. M. Cann, and G. N. Patey, J. Chem. Phys. **107**, 4719 (1997).
[23] M. Kinoshita, J. Solution Chem. **33**, 661 (2004).
[24] M. Kinoshita, J. Mol. Liq. **119**, 47 (2005).
[25] M. Kinoshita, N. Matubayasi, Y. Harano, and M. Nakahara, J. Chem. Phys. **124**, 024512 (2006).
[26] M. Kinoshita, J. Chem. Phys. **128**, 024507 (2008).
[27] M. Kinoshita and M. Suzuki, J. Chem. Phys. **130**, 014707 (2009).
[28] M. Kinoshita and T. Yoshidome, J. Chem. Phys. **130**, 144705 (2009).
[29] P. G. Kusalik and G. N. Patey, J. Chem. Phys. **88**, 7715 (1988).
[30] P. G. Kusalik and G. N. Patey, Mol. Phys. **65**, 1105 (1988).
[31] P. M. König, R. Roth, and K. R. Mecke, Phys. Rev. Lett. **93**, 160601 (2004).
[32] R. Roth, Y. Harano, and M. Kinoshita, Phys. Rev. Lett. **97**, 078101 (2006).
[33] K. Shiraki, K. Nishikawa, and Y. Goto, J. Mol. Biol. **245**, 180 (1995).
[34] D. Hamada, S. Segawa, and Y. Goto, Nat. Struct. Biol. **3**, 868 (1996).
[35] T. Imai, Y. Harano, M. Kinoshita, A. Kovalenko, and F. Hirata, J. Chem. Phys. **125**, 024911 (2006).
[36] K. T. Simons, C. Kooperberg, E. Huang, and D. Baker, J. Mol. Biol. **268**, 209 (1997).
[37] W. Kabsch and C. Sander, Biopolymers **22**, 2577 (1983).
[38] J. W. Ponder and F. M. Richards, J. Comput. Chem. **8**, 1016 (1987).
[39] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[40] M. Feig, J. Karanicolas, and C. L. Brooks III, J. Mol. Graphics **22**, 377 (2004).
[41] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
[42] A. D. Mackerell, Jr., M. Feig, and C. L. Brooks III, J. Comput. Chem. **25**, 1400 (2004).
[43] M. S. Lee, F. R. Salsbury, Jr., and C. L. Brooks III, J. Chem. Phys. **116**, 10606 (2002).
[44] M. S. Lee, M. Feig, F. R. Salsbury, Jr., and C. L. Brooks III, J. Comput. Chem. **24**, 1348 (2003).
[45] J. Chocholoušová and M. Feig, J. Comput. Chem. **27**, 719 (2006).
[46] Y. Harano, T. Yoshidome, and M. Kinoshita, J. Chem. Phys. **129**, 145103 (2008).
[47] T. Yoshidome, Y. Harano, and M. Kinoshita, Phys. Rev. E **79**, 011912 (2009).
[48] T. Yoshidome and M. Kinoshita, Chem. Phys. Lett. **477**, 211 (2009).
[49] T. Yoshidome and M. Kinoshita, Phys. Rev. E **79**, 030905(R) (2009).
[50] H. Oshima, T. Yoshidome, K. Amano, and M. Kinoshita, J. Chem. Phys. **131**, 205102 (2009).
[51] K. Amano, T. Yoshidome, Y. Harano, K. Oda, and M. Kinoshita, Chem. Phys. Lett. **474**, 190 (2009).
[52] Y. Harano, R. Roth, and M. Kinoshita, Chem. Phys. Lett. **432**, 275 (2006).
[53] Y. Harano, R. Roth, Y. Sugita, M. Ikeguchi, and M. Kinoshita, Chem. Phys. Lett. **437**, 112 (2007).
[54] J. S. Perkyns and B. M. Pettitt, Chem. Phys. Lett. **190**, 626 (1992).

[55] J. S. Perkyns and B. M. Pettitt, J. Chem. Phys. **97**, 7656 (1992).

[56] M. Ikeguchi and J. Doi, J. Chem. Phys. **103**, 5011 (1995).

[57] M. Kinoshita, J. Chem. Phys. **116**, 3493 (2002).

[58] C. N. Pace, Biochemistry **40**, 310 (2001).

[59] M. Karplus and E. Shakhnovich, in *Protein Folding*, edited by T. E. Creighton (Freeman, New York, 1992), pp. 127–195.

[60] M. Ohgushi and A. Wada, FEBS Lett. **164**, 21 (1983).

[61] K. Kuwajima, Proteins **6**, 87 (1989).

[62] O. B. Ptitsyn, in *Protein Folding*, edited by T. E. Creighton (Freeman, New York, 1992), pp. 243–300.

[63] V. I. Lim, FEBS Lett. **89**, 10 (1978).

[64] K. Kuwajima, H. Yamaya, S. Miwa, S. Sugai, and T. Nagamura, FEBS Lett. **221**, 115 (1987).

[65] C. M. Dobson, Nature (London) **426**, 884 (2003).