

Structural stability of proteins in aqueous and nonpolar environments

Satoshi Yasuda, Hiraku Oshima, and Masahiro Kinoshita

Citation: *J. Chem. Phys.* **137**, 135103 (2012); doi: 10.1063/1.4755755

View online: <http://dx.doi.org/10.1063/1.4755755>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v137/i13>

Published by the [American Institute of Physics](#).

Additional information on J. Chem. Phys.

Journal Homepage: <http://jcp.aip.org/>

Journal Information: http://jcp.aip.org/about/about_the_journal

Top downloads: http://jcp.aip.org/features/most_downloaded

Information for Authors: <http://jcp.aip.org/authors>

ADVERTISEMENT



Goodfellow
metals • ceramics • polymers • composites
70,000 products
450 different materials
small quantities fast
www.goodfellowusa.com

Structural stability of proteins in aqueous and nonpolar environments

Satoshi Yasuda,¹ Hiraku Oshima,² and Masahiro Kinoshita^{2,a)}¹Graduate School of Energy Science, Kyoto University, Uji, Kyoto 611-0011, Japan²Institute of Advanced Energy, Kyoto University, Uji, Kyoto 611-0011, Japan

(Received 5 August 2012; accepted 13 September 2012; published online 3 October 2012)

A protein folds into its native structure with the α -helix and/or β -sheet in aqueous solution under the physiological condition. The relative content of these secondary structures largely varies from protein to protein. However, such structural variability is not exhibited in nonaqueous environment. For example, there is a strong trend that alcohol induces a protein to form α -helices, and many of the membrane proteins within the lipid bilayer consists of α -helices. Here we investigate the structural stability of proteins in aqueous and nonpolar environments using our recently developed free-energy function $F = (\Lambda - TS)/(k_B T_0) = \Lambda/(k_B T_0) - S/k_B$ ($T_0 = 298$ K and the absolute temperature T is set at T_0) which is based on statistical thermodynamics. $\Lambda/(k_B T_0)$ and S/k_B are the energetic and entropic components, respectively, and k_B is Boltzmann's constant. A smaller value of the positive quantity, $-S$, represents higher efficiency of the backbone and side-chain packing promoted by the entropic effect arising from the translational displacement of solvent molecules or the CH_2 , CH_3 , and CH groups which constitute nonpolar chains of lipid molecules. As for Λ , in aqueous solution, a transition to a more compact structure of a protein accompanies the break of protein-solvent hydrogen bonds: As the number of donors and acceptors buried without protein intramolecular hydrogen bonding increases, Λ becomes higher. In nonpolar solvent, lower Λ simply implies more intramolecular hydrogen bonds formed. We find the following. The α -helix and β -sheet are advantageous with respect to $-S$ as well as Λ and to be formed as much as possible. In aqueous solution, the solvent-entropy effect on the structural stability is so strong that the close packing of side chains is dominantly important, and the α -helix and β -sheet contents are judiciously adjusted to accomplish it. In nonpolar solvent, the solvent-entropy effect is substantially weaker than in aqueous solution. Λ is crucial and the α -helix is more stable than the β -sheet in terms of Λ , which develops a tendency that α -helices are exclusively chosen. For a membrane protein, α -helices are stabilized as fundamental structural units for the same reason, but their arrangement is performed through the entropic effect mentioned above. © 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.4755755>]

I. INTRODUCTION

We have been arguing that protein folding in aqueous solution under the physiological condition is driven by a gain in the configurational entropy of water.^{1–6} The folding process accompanies the break of hydrogen bonds with water molecules. There is no problem if the resulting energetic penalty is compensated by the formation of protein intramolecular hydrogen bonds, but such compensation is not always achievable. Hence, the enthalpic change upon protein folding takes a large, positive value. Moreover, protein folding undergoes a large loss of the protein conformational entropy. It follows that the water-entropy gain is powerful enough to surpass the positive enthalpic change plus the conformational-entropy loss. This concept has been proved by Terazima and his co-workers^{3,7} who developed a novel experimental technique enabling us to directly measure the enthalpic change upon apoplastocyanin (apoPC) folding at 298 K.

Upon protein folding, the excluded volume (EV) (i.e., the volume of the space which the centers of water molecules cannot enter) decreases to a large extent, which is followed

by a corresponding increase in the total volume available to the translational displacement of water molecules and a reduction in the water crowding.^{1–6} Namely, the folding leads to gains in the number of accessible translational configurations of water and in the water entropy. The formation of α -helix or β -sheet leads to a significant decrease in the EV (see Figs. 1(a) and 1(b)).^{5,6} At the same time, this formation compensates the loss of hydrogen bonds with water molecules by ensuring intramolecular hydrogen bonds. Hence, these secondary structures are very advantageous units which are to be formed as much as possible. Close packing of side chains (see Fig. 1(c)), which reduces the EV to a considerable extent, is also crucial in protein folding.^{5,6}

We have recently investigated how the secondary structural preference is influenced by the side-chain packing.⁶ The findings are as follows. As far as the backbone is concerned, 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 (i.e., the sum of α -helices and β -sheets) does not always lead to higher entropy of water. The native structure (NS) is characterized by the side chains with a variety of geometrical features specifically locked with little space in its interior. To accomplish

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

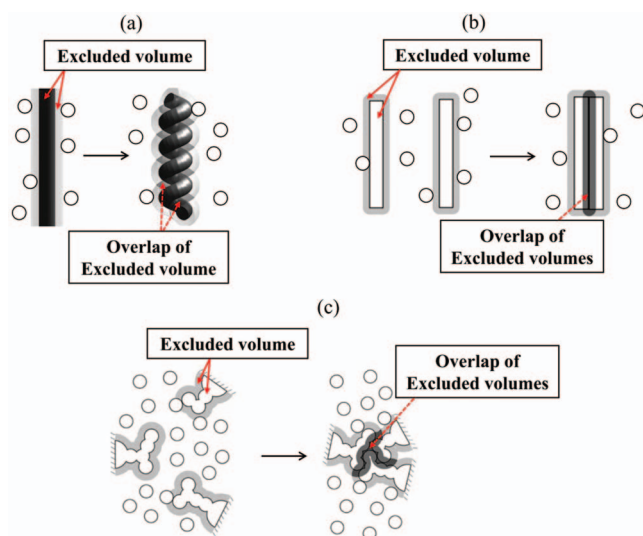


FIG. 1. (a) Formation of helical structure by a portion of the backbone. (b) Lateral contact of portions of the backbone. (c) Close packing of side chains. The overlap of excluded volumes or that in an excluded volume occurs, and the total volume available to the translational displacement of solvent molecules increases by the overlapped volume.

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 NS thus constructed is not necessarily high. Moreover, some native structures possess α -helices or β -sheets alone, and others consist of mixtures of α -helices and β -sheets. The side-chain packing has substantially large effects on the secondary structural preference.

For the above reason, the relative content of the α -helix and β -sheet largely varies from protein to protein in aqueous environment. Such structural variability, however, is not exhibited in nonaqueous environment. For example, it has been found in experiments^{8,9} that β -lactoglobulin, which comprises antiparallel β -sheets and only one α -helix, undergoes a structural transition to an almost completely helical structure when a sufficiently large amount of alcohol is added to aqueous solution: The α -helix content shows a rather abrupt increase beyond a threshold value of the alcohol concentration. The α -helix content of the resulting structure is estimated to reach $\sim 90\%$.⁸ Cytochrome *c* is a protein with α -helices and its α -helix content manifests a further increase as the alcohol concentration becomes higher.¹⁰ Thus, there is a strong trend that alcohol induces a protein to form α -helices. It has been demonstrated that though the helical structure induced by alcohol is independent of the alcohol species, the degree of inducing α -helices is strengthened as the hydrocarbon group of an alcohol molecule becomes more bulky.¹¹ As the alcohol concentration and the bulkiness of the hydrocarbon group increase, the solvent environment becomes *more nonaqueous*, promoting the formation of α -helices to a larger extent. As an extreme example of nonaqueous environment, a membrane protein is within the lipid bilayer. Many of the membrane proteins consist of α -helices.^{12,13} The high propensity to form α -helices in nonaqueous environment is in marked contrast to the structural variability in aqueous environment.

In the present study, we investigate the structural stability of proteins in aqueous and nonpolar environments using our recently developed free-energy function^{14–16} $F = (\Lambda - TS)/(k_B T_0) = \Lambda/(k_B T_0) - S/k_B$ ($T_0 = 298$ K and the absolute temperature T is set at T_0) and its energetic and entropic components, $\Lambda/(k_B T_0)$ and S/k_B , respectively. The high performance of F was demonstrated in our earlier works (see Sec. II E). Two solvents, water and nonpolar solvent, are considered for aqueous and nonaqueous environments, respectively. The nonpolar solvent provides an extreme example of nonaqueous environment. A water molecule is modeled as a hard sphere in which a point dipole and a point quadrupole of tetrahedral symmetry are embedded.^{17,18} The nonpolar solvent is formed by neutral hard spheres. The two solvents share the same molecular diameter and number density. F is strongly dependent on the protein structure. S is the solvation entropy and is calculated using a hybrid of the angle-dependent^{17–29} or radial-symmetric³⁰ integral equation theory and the morphometric approach.^{31–33} A smaller value of the positive quantity, $-S$, represents higher efficiency of the backbone and side-chain packing promoted by the entropic effect arising from the translational displacement of solvent molecules. For a membrane protein, the entropic effect is attributable to the translational displacement of the CH_2 , CH_3 , and CH groups (the CH_2 groups form the major component) which constitute nonpolar chains of lipid molecules. When the solvent is water, Λ is calculated in accordance with a simple manner which still accounts for physically the most important factors: intramolecular and protein-water hydrogen bonds. A procedure suited to the calculation of Λ for a nonpolar solvent is newly proposed. In aqueous solution, a transition to a more compact structure of a protein accompanies the break of protein-solvent hydrogen bonds: As the number of donors and acceptors buried without protein intramolecular hydrogen bonding increases, Λ becomes higher. In nonpolar solvent, lower Λ simply implies more intramolecular hydrogen bonds formed.

Our major findings are as follows. The solvent-entropy effect by water is substantially stronger than that by the nonpolar solvent. In aqueous solution, the α -helix and β -sheet contents are judiciously adjusted to accomplish the close packing of side chains which is dominantly important. As a consequence, a variety of structures are stabilized: For instance, the total content of the secondary structures is not necessarily high, some proteins possess only α -helices or β -sheets while others consist of mixtures of α -helices and β -sheets, and the relative content of the α -helix and β -sheet largely varies from protein to protein. This structural variability of proteins in aqueous environment, which is presented by the water-entropy effect, can be relevant to the diversity of their functions. In the nonpolar solvent, the importance of forming as many intramolecular hydrogen bonds as possible predominates over the solvent-entropy effect. Λ tends to become lower as the α -helix or β -sheet content increases, but this tendency is much stronger for the α -helix content: α -helices are exclusively chosen and a higher content of the α -helix is more favorable. For a membrane protein, α -helices are stabilized as fundamental structural units for the same reason, but their arrangement

is performed through the entropic effect originating from the translational displacement of the CH₂, CH₃, and CH groups of lipid molecules. We are successful in showing that for significantly many proteins in aqueous solution, the NS is more stable in terms of our free-energy function F than any of the non-native structures artificially constructed. In nonpolar solvent, by contrast, only highly helical structures are stabilized. Further, we demonstrate for a membrane protein that its NS is more stable in terms of F than any of the non-native structures generated using a computer simulation.

II. MODEL AND THEORY

A. Solvent and protein models

We consider two solvents, water and nonpolar solvent. A multipolar model is adopted for water: 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.^{17,18} In the angle-dependent integral equation theory^{17–29} (ADIET), the theoretical tool we employ for the solvation thermodynamics for water, the effect of the molecular polarizability is taken into account using the self-consistent mean field (SCMF) theory.^{17,18} At the SCMF level the many-body induced interactions are reduced to pairwise additive potentials involving an effective dipole moment. The value of $\rho_S d_S^3$ (ρ_S is the bulk number density) is set at that of water at 298 K and 1 atm: $\rho_S d_S^3 = 0.7317$. We remark that the solvation free energies of nonpolar solutes calculated by the ADIET with the multipolar model are in quantitatively excellent agreement with those from Monte Carlo computer simulations.²⁷ Further, the dielectric constant of bulk water calculated, which is a good measure of the validity of a theory, is ~ 83 that is in good accord with the experimental value ~ 78 . The nonpolar solvent is formed by neutral hard spheres with diameter $d_S = 0.28$ nm and $\rho_S d_S^3$ is set at 0.7317. We note that our model water and nonpolar solvent share the same molecular diameter and number density. The radial-symmetric integral equation theory³⁰ (RSIET) is employed for the nonpolar solvent.

The solvation entropy (in particular, that under the isochoric condition) is fairly insensitive to the solute-solvent interaction potential as proved in our earlier works. For example, the solvation free energy, entropy, and energy under the isochoric condition denoted by μ , S , and E , respectively, are calculated at $T = 298$ K (T is the absolute temperature) for a spherical solute with diameter 0.28 nm using the ADIET^{17–29} combined with the multipolar water model.^{17,18} For the hard-sphere solute with zero charge, the calculated values are $\mu = 5.95 k_B T$, $S = -9.22 k_B$, and $E = -3.27 k_B T$ (k_B is Boltzmann's constant). When the point charge $-0.5 e$ (e is the elementary electric charge) is embedded at its center, the calculated values are $\mu = -32.32 k_B T$, $S = -10.11 k_B$, and $E = -42.43 k_B T$. Thus, S is fairly insensitive to the solute-solvent interaction potential while μ and E are largely influenced by it. Imai *et al.*³⁴ have considered the native structures of a total of eight peptides and proteins and calculated S using the three-dimensional reference interaction site model 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%.

For the above reasons, a protein in aqueous environment as well as that in nonaqueous environment can be modeled as a set of fused hard spheres just for calculating its solvation entropy. The solvation energy in aqueous environment, which is influenced by the protein-solvent interaction potential, is separately treated: It is estimated in a different manner (see Sec. II D). The x - y - z coordinates of the protein atoms (H, C, N, O, and S) are used as part of the input data to account for the polyatomic structure at the atomic level. The diameter of each atom is set at the sigma value of the LJ potential parameters which are taken from the CHARMM22.³⁵

B. Free-energy function

Our free-energy function F (F is dimensionless due to the normalization by $k_B T_0$; see Eq. (1)) is expressed for a prescribed structure of a protein by

$$F = (E_I + \mu)/(k_B T_0), T_0 = 298 \text{ K}, \quad (1)$$

where E_I is the protein intramolecular energy and μ is the solvation free energy (i.e., excess chemical potential) that is the most important thermodynamic quantity of protein solvation. We note that μ is the same irrespective of the protein insertion condition: isobaric or isochoric,²² and we consider the isochoric condition that is much more convenient in a theoretical treatment. Using the relation

$$\mu = E - TS, \quad (2)$$

where E and S denote the solvation energy and entropy, respectively, and defining Λ by

$$\Lambda = E_I + E, \quad (3)$$

we obtain

$$F = (\Lambda - TS)/(k_B T_0) = \Lambda/(k_B T_0) - S/k_B, T = T_0. \quad (4)$$

We emphasize that T is set at T_0 in the present study. For the hard-sphere solvent, E in Eq. (3) is zero. Λ is calculated by choosing a fully extended structure as the standard structure. S represents the solvent-entropy loss upon the protein insertion. $-S$ is positive, and Λ is positive for water and negative for the nonpolar solvent. Λ , S , and F are strongly dependent on the protein structure. The procedures of calculating the entropic component S/k_B and the energetic component $\Lambda/(k_B T_0)$ are briefly described below (more details are given in our earlier publications^{3,4,15,36}).

C. Entropic component

The calculation of S is finished quite rapidly by combining the ADIET^{17–29} or RSIET³⁰ with the morphometric approach (MA).^{31–33} We refer to this combination as the hybrid method. The idea of the MA is to express a hydration quantity such as S by the linear combination of only four geometric

measures of a solute molecule:

$$S/k_B = C_1 V_{\text{ex}} + C_2 A + C_3 X_M + C_4 Y_M. \quad (5)$$

In this equation, which is referred to as the morphometric form, V_{ex} is the EV, A is the solvent-accessible surface area, and X_M and Y_M are the integrated mean and Gaussian curvatures of the accessible surface, respectively. The solvent-accessible surface is the surface that is accessible to the centers of solvent molecules. The volume that is enclosed by this surface is the EV. In the MA, the solute shape enters S only via the four geometric measures. Therefore, the four coefficients (C_1 – C_4) can be determined in simple geometries. They are calculated from the values of S for hard-sphere solutes with various diameters immersed in one of the two solvents considered. In the calculation, the ADIET and RSIET are employed for water and for the nonpolar solvent, respectively.

The procedure of calculating S of a protein with a prescribed structure comprises the following four steps.

- (1) S of a hard-sphere solute with diameter d_U is calculated using the ADIET or RSIET. The values of S are prepared for sufficiently many different values of d_U ($0 \leq d_U \leq 30d_S$; d_S is the solvent diameter).
- (2) The four coefficients are determined by the least squares fitting applied to the following equation for hard-sphere solutes (i.e., the morphometric form applied to hard-sphere solutes):

$$\begin{aligned} S/k_B = & C_1(4\pi R^3/3) + C_2(4\pi R^2) + C_3(4\pi R) \\ & + C_4(4\pi), \end{aligned} \quad (6)$$

$$R = (d_U + d_S)/2.$$

The most recent method of the least square fitting is described in one of our publications.³³

- (3) The four geometric measures of a protein (V_{ex} , A , X_M , and Y_M) with a prescribed structure are calculated by means of an extension³² of Connolly's algorithm.^{37,38} The x - y - z coordinates of the protein atoms are used as part of the input data to account for the polyatomic structure at the atomic level. The diameter of each atom is set at the sigma value of the LJ potential parameters which are taken from the CHARMM22.³⁵
- (4) S of a protein with a prescribed structure is obtained from Eq. (5) in which the four coefficients determined in step (2) are used. It should be emphasized that the computation time required for step (4) is only ~ 0.1 s on our workstation.

The high reliability of the hybrid method in calculating S has been demonstrated in the following examples: quantitative reproduction of the experimentally measured changes in thermodynamic quantities upon apoPC folding;³ elucidation of the molecular mechanisms of pressure^{39,40} and cold^{41,42} denaturing of proteins; proposal of a reliable measure of the thermal stability of proteins,^{43,44} and prediction of the so-called hot spots (i.e., residues accounting for the majority of the protein-protein binding free energy despite that they comprise only a small fraction of the protein-protein interface) in protein-protein complexes.⁴⁵

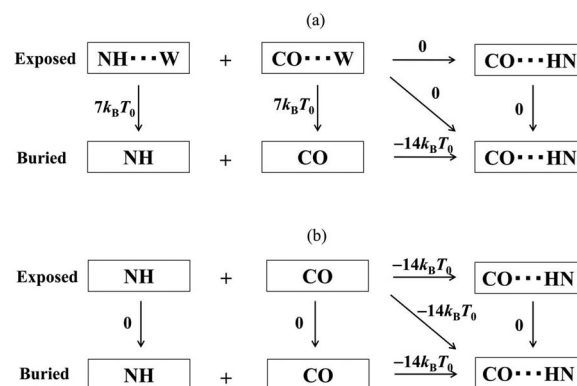


FIG. 2. Our basic strategy for calculating Λ . “W” and “...” represent a water molecule and a hydrogen bond, respectively. $T_0 = 298$ K. (a) is for aqueous environment and (b) is for nonpolar environment.

D. Energetic component

When the solvent is water, Λ defined by Eq. (3) is calculated in accordance with a simple manner which still accounts for physically the most important factors: protein intramolecular and protein-water hydrogen bonds. The calculation of Λ is performed by choosing a fully extended structure, which possesses the maximum number of hydrogen bonds with water molecules but no intramolecular hydrogen bonds, as the standard structure. Compared to the fully extended structure with $\Lambda = 0$, in a more compact structure some donors and acceptors (e.g., N and O, respectively) are buried in the interior after the break of hydrogen bonds with water molecules ($\text{CO} \cdots \text{W}$, $\text{NH} \cdots \text{W}$, etc.). There is no problem if intramolecular hydrogen bonds ($\text{CO} \cdots \text{HN}$, etc.) are formed. However, such formation is not always achieved, giving rise to an energetic penalty.

Our procedure of calculating Λ can be summarized as follows. As explained in Fig. 2(a), when a donor and an acceptor are buried in the interior after the break of hydrogen bonds with water molecules, if they form an intramolecular hydrogen bond, we impose no penalty. On the other hand, when a donor or an acceptor is buried with no intramolecular hydrogen bond formed, we impose the penalty of $7 k_B T_0$ ($T_0 = 298$ K). The value, $7 k_B T_0$, is based on the energy-decrease of $-14 k_B T_0$ arising from hydrogen-bond formation between two formamide molecules in a nonpolar liquid.⁴⁶ The water-accessible surface areas of all the donors and acceptors in the backbone and side chains are calculated by means of Connolly's algorithm^{37,38} (the TINKER program package⁴⁷ is used). When a donor or an acceptor has the area which is smaller than $0.001 \text{ \AA}^2$, it is considered buried. To determine if an intramolecular hydrogen bond is formed or not, we use the criteria proposed by McDonald and Thornton.⁴⁸

The above procedure must be modified when the solvent is nonpolar as explained in Fig. 2(b). The formation of an intramolecular hydrogen bond, regardless that it is buried or not, surely leads to the energy lowering of $-14 k_B T_0$. Even when a donor or an acceptor is buried with no intramolecular hydrogen bond formed, there is no penalty to be imposed. In both of aqueous and nonpolar environments, it is assumed that any gain in the protein intramolecular van der Waals attractive

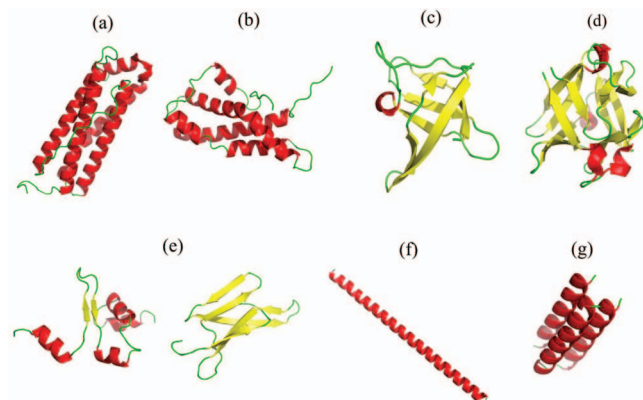


FIG. 3. Ribbon representations for the native structures of human H chain ferritin (a), fertilization protein (b), CspB (c), and HB/FGF (d). Ribbon representations for two representative non-native structures (e), “all α ” (f), and “all α -3” (g) considered for CspB. They are drawn using the PyMOL.

interaction is cancelled out by the loss of the protein-solvent van der Waals attractive interaction. This is justifiable because the effects of the break of hydrogen bonds with water molecules and the formation of intramolecular hydrogen bonds should form a principal constituent of the energetic

component. We note that Λ is negative in nonpolar solvent, whereas it is positive in aqueous solution. The time required in the calculation of Λ is only ~ 0.1 s per protein structure on our workstation.

E. Performance of free-energy function

We have examined the performance of our free-energy function F in discriminating the native fold from a number of misfolded decoys.^{15,16} The examination is carried out for a total of 133 proteins in 8 decoy sets. The solvent is water. F is shown to be far superior to any of the previously reported functions. When the NS model is determined by the X-ray crystallography, the discrimination is always successful. In the case of NMR models, as long as a sufficiently good NS model is included in the candidate models, the discrimination is accomplished with 100% accuracy. Further, using F and its energetic and entropic components ($\Lambda/(k_B T_0)$ and S/k_B , respectively) we have recently developed a reliable method of characterizing the NS models of a protein determined through the X-ray crystallography and NMR experiments combined with structure calculations.³⁶ The approximations employed in calculating Λ , S , and F can thus be justified by these successful results.

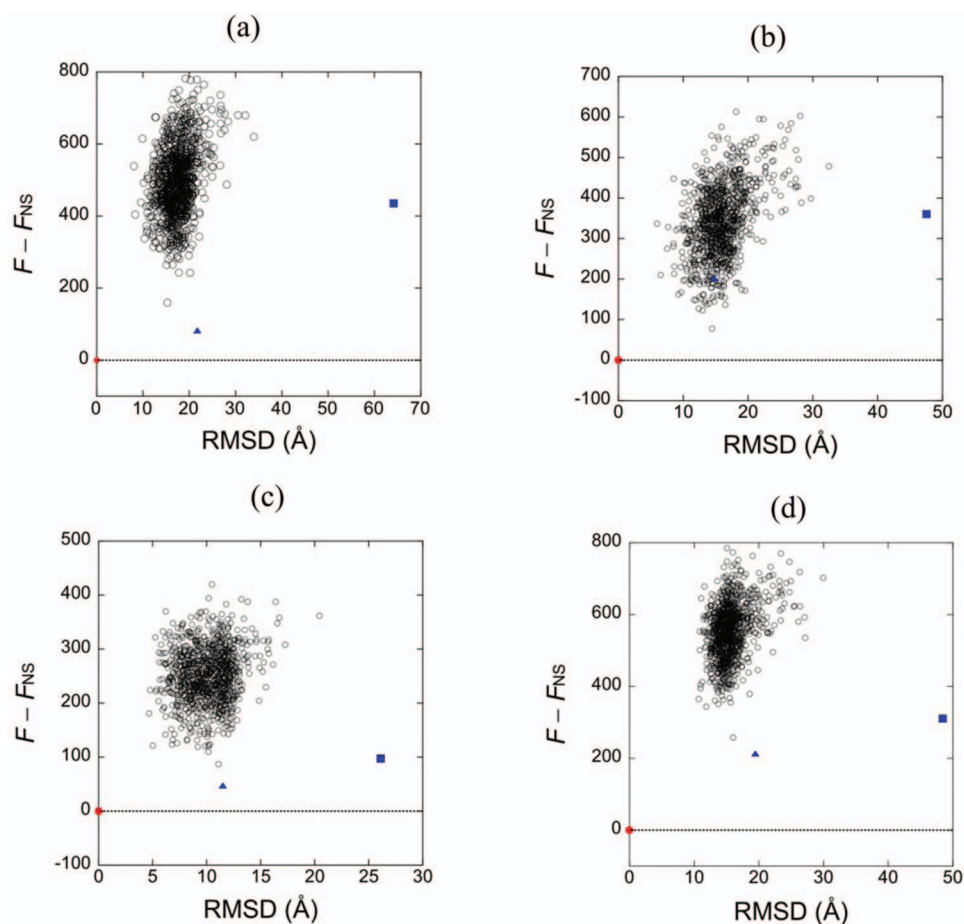


FIG. 4. Relation between F and RMSD from the native structure in aqueous environment for human H chain ferritin (a), fertilization protein (b), CspB (c), and HB/FGF (d). Black circles: data points for the non-native structures. Red circle: data point for the native structure. Blue square: data point for “all α .” Blue triangle: data point for “all α -3.”

F. Proteins considered

We have tested significantly many proteins including human H chain ferritin (PDB Code: 2fha, $N_r = 172$, where N_r is the number of residues), fertilization protein (PDB Code: 1lis, $N_r = 131$), major cold shock protein of bacillus subtilis, CspB (PDB Code: 1csp, $N_r = 67$), and basic fibroblast growth factor, HB/FGF (PDB code: 4fgf, $N_r = 124$). We present only the results of these four proteins because the qualitative aspects of our conclusions are not altered by the results of the other proteins. The native structures (i.e., the native folds in aqueous solution) of human H chain ferritin and fertilization protein possess only α -helices while those of CspB and HB/FGF possess only β -sheets. The ribbon representations for the native structures of the four proteins are shown in Figs. 3(a)–3(d). For each protein, we consider a number of non-native structures taken from the Rosetta decoy set.⁴⁹ In order to conform the number of residues of the NS to that of the non-native structures, we removed two small end portions of the NS (12, 21, 3, 4 residues for human H chain ferritin, fertilization protein, CspB, and HB/FGF, respectively). The portions are not buried, and they do not participate in the α -helix or β -sheet formation. Therefore, the removal has no significant

effects on the structural characteristics. The ribbon representations for two representative non-native structures considered for CspB are shown in Fig. 3(e).

For each protein, the complete α -helix structure is generated using the TINKER program package⁴⁷ with the dihedral angles (φ , ψ , and ω) set at (60° , 45° , and 180°). Hereafter, this structure is referred to as “all α ” (see Fig. 3(f)). In order to construct a more compact structure whose α -helix content is almost as high as that of “all α ”, we divide “all α ” into three portions and arrange them. The arrangement is made using Hyperchem 8 so that they can form triangle-like geometry when they are viewed from the top side (see Fig. 3(g)). The resulting structure is referred to as “all α -3.” Unlike “all α ,” “all α -3” possesses the side-chain packing among the three α -helices.

The LJ potential energy for the native structures, non-native structures, “all α ”, and “all α -3” can be 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 program⁵⁰ through the multiscale modeling tools in structural biology (MMTSB) program.⁵¹ The minimization is performed so that the original structures can be retained as much

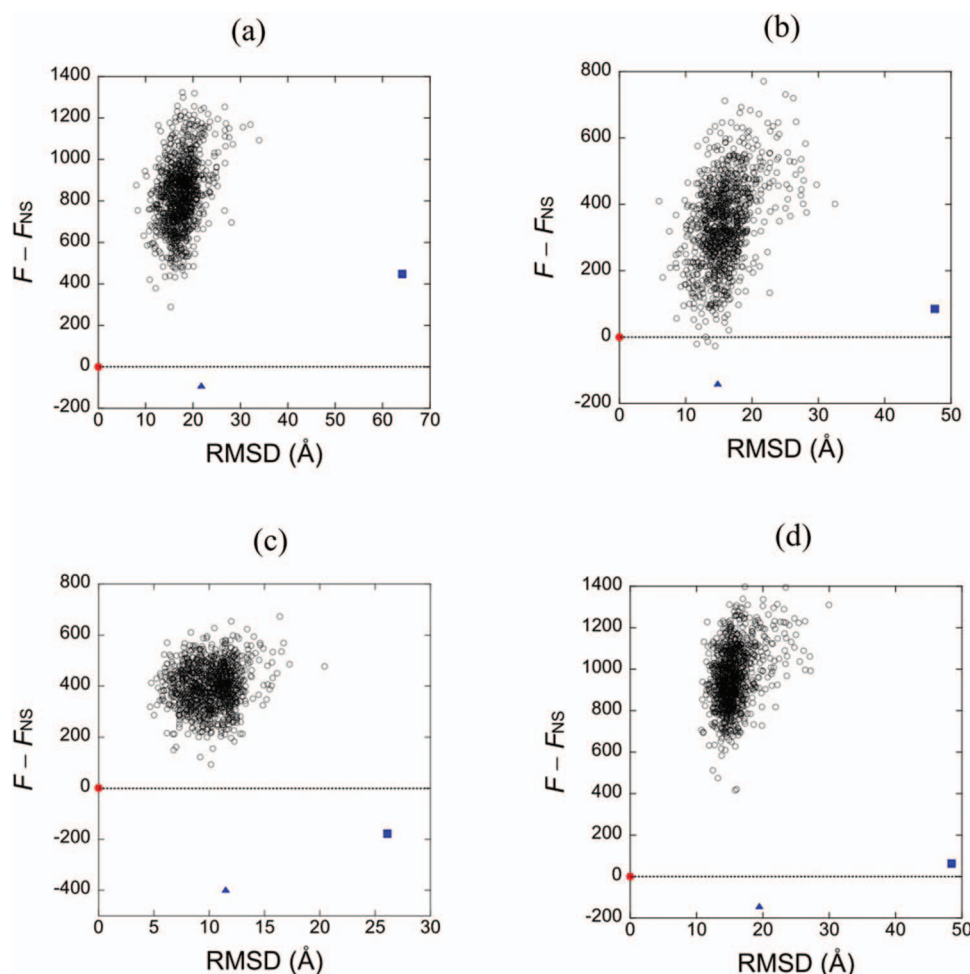


FIG. 5. Relation between F and RMSD from the native structure in nonpolar environment for human H chain ferritin (a), fertilization protein (b), CspB (c), and HB/FGF (d). Black circles: data points for the non-native structures. Red circle: data point for the native structure. Blue square: data point for “all α .” Blue triangle: data point for “all α -3.”

as possible. We employ the CHARMM22 as the force-field parameters.³⁵ Electrostatic and nonbonded interactions are all evaluated without any cutoff. After the minimization, each structure is switched to a set of fused hard spheres for calculating the solvation entropy. Now there are no serious overlaps of protein atoms.

The α -helix and β -sheet contents of a structure are important parameters in the present study. The DSSP program⁵² is employed for calculating these contents. The α -helix contents of the native structures of human H chain ferritin and fertilization protein are $\sim 76\%$ and $\sim 77\%$, respectively. The β -sheet contents of the native structures of CspB and HB/FGF are $\sim 47\%$ and $\sim 37\%$, respectively.

III. RESULTS AND DISCUSSION

A. Comparison between structural stabilities in aqueous and nonpolar environments

For the four proteins, we compare the stabilities of the NS, non-native structures, “all α ,” and “all α -3” in aqueous solution with those in nonpolar solvent. Figures 4 and 5 show the plots of $(F - F_{\text{NS}})$ (the subscript “NS” denotes the value of the NS) against the root mean square deviation (RMSD) from the NS. In Fig. 4 where aqueous solution is considered, the NS is the most stable for all the four proteins in the sense that its F

is the lowest. This gives a demonstration of the high reliability of F . In nonpolar solvent, on the other hand, “all α -3” is the most stable as observed in Fig. 5. Moreover, for fertilization protein, three of the non-native structures are more stable than the NS. For CspB, “all α ” is also more stable than the NS.

In what follows, we investigate the physical origins of these differences between structural stabilities in aqueous and nonpolar environments. For the investigation, we define X and Y as

$$X = \Lambda/(k_B T_0) - \Lambda_{\text{NS}}/(k_B T_0) \quad (7)$$

and

$$Y = -S/k_B - (-S_{\text{NS}}/k_B), \quad (8)$$

respectively. X and Y , respectively, denote $\Lambda/(k_B T_0)$ and $-S/k_B$ of a structure *relative to* those of the NS. Close packing of the backbone and side chains leads to the reduction of the EV generated for solvent molecules by the protein followed by the relaxation of solvent crowding. $-S$ is made as small as possible primarily by this effect. $-S$ or Y represents the efficiency of backbone and side-chain packing, and lower efficiency results in a higher value of $-S$ or Y . In nonpolar solvent, a lower value of Λ or X implies more intramolecular hydrogen bonds formed. In aqueous solution, whether donors and acceptors are buried or not is also an important

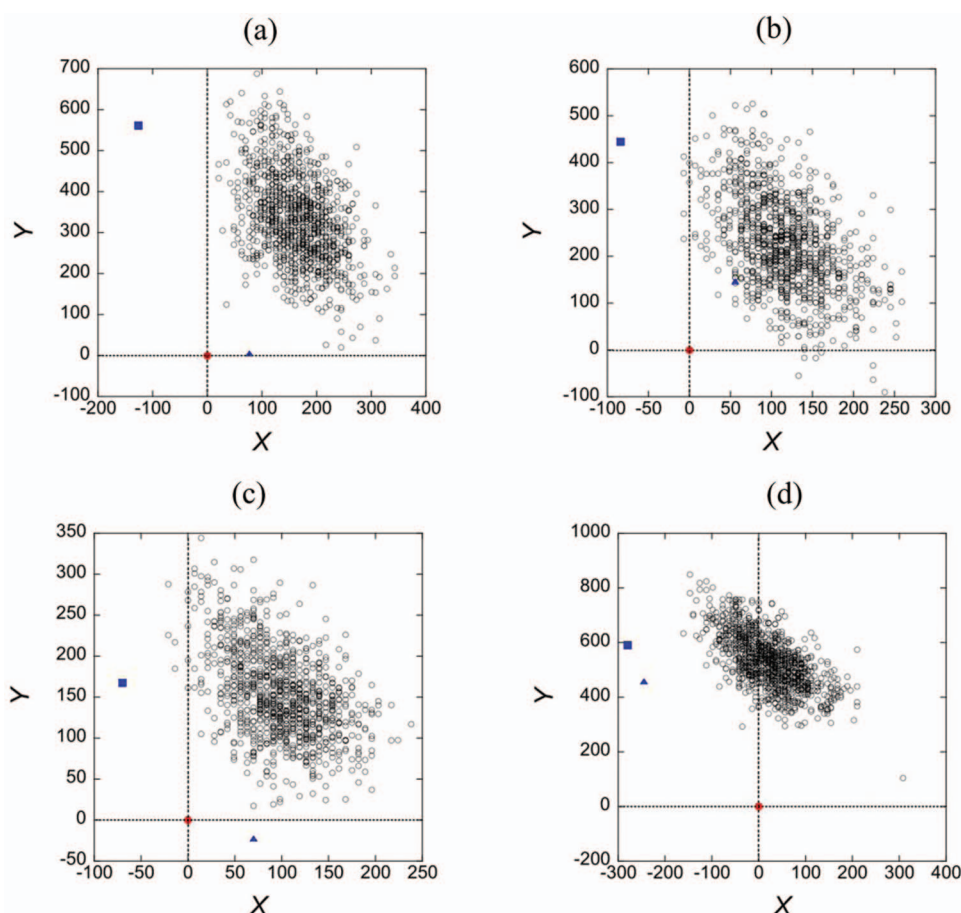


FIG. 6. Relation between X and Y in aqueous environment for human H chain ferritin (a), fertilization protein (b), CspB (c), and HB/FGF (d). Black circles: data points for the non-native structures. Red circle: data point for the native structure. Blue square: data point for “all α .” Blue triangle: data point for “all α -3.”

factor. As the number of donors and acceptors buried without intramolecular hydrogen bonding increases, Δ or X becomes higher.

B. Physics of structural stability in aqueous environment

Figure 6 shows X plotted against Y for the NS, non-native structures, “all α ,” and “all α -3” in aqueous solution. For all the four proteins, “all α ” merits the lowest value of X . Several non-native structures possess lower X than the NS for fertilization protein and CspB, but this is not the case for human H chain ferritin. Strikingly, the NS of HB/FGF suffers considerably high X : It is much less stable than many of the non-native structures as well as “all α ” and “all α -3” in terms of X . The reason for this result is that the NS of HB/FGF possesses significantly many donors and acceptors without the formation of intramolecular hydrogen bonds. For all the four proteins, the solvent entropy has substantially large effects on the stability of the NS. In particular, the NS of HB/FGF becomes the most stable due to the solvent-entropy effect represented by Y . The considerably high stability of “all α -3” is due to Y for human H chain ferritin and CspB, whereas it is due to X for HB/FGF.

Overall, there can be structures which are more stable than the NS with respect to X , but such structures are problematic in terms of Y . Though structures which are more favorable than the NS with regard to Y can be constructed, they suffer unreasonably high X . Thus, the NS is optimized for the sum of Δ and $-TS$.

It is physically insightful to separate the effect of side chains from that of the backbone for the entropic component. To perform this separation, we replace all residues in each structure by Gly using the CHARMM and MMTSB programs.^{50,51} The replacement is carried out after the slight modification of the structure described in Sec. II F. The structure thus made has essentially no side chains (hereafter, these are referred to as “the structures without side chains”). $-S$ 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 $-S_b$: $-S = -S_b + (-S_{sc})$ where $-S_b$ and $-S_{sc}$ denote the contributions from the backbone and side chains to $-S$, respectively. We then define Y_b and Y_{sc} as

$$Y_b = -S_b/k_B - (-S_{b,NS}/k_B) \quad (9)$$

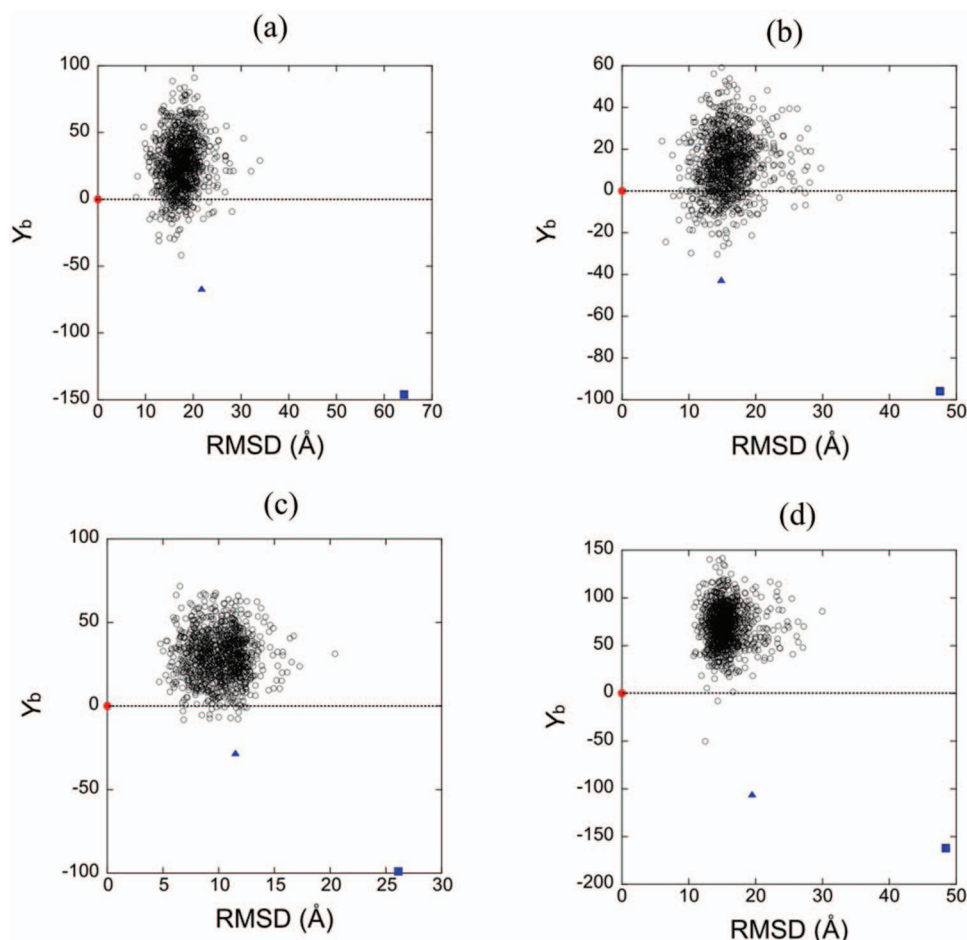


FIG. 7. Relation between Y_b and RMSD from the native structure in aqueous environment for human H chain ferritin (a), fertilization protein (b), CspB (c), and HB/FGF (d). Black circles: data points for the non-native structures. Red circle: data point for the native structure. Blue square: data point for “all α .” Blue triangle: data point for “all α -3.”

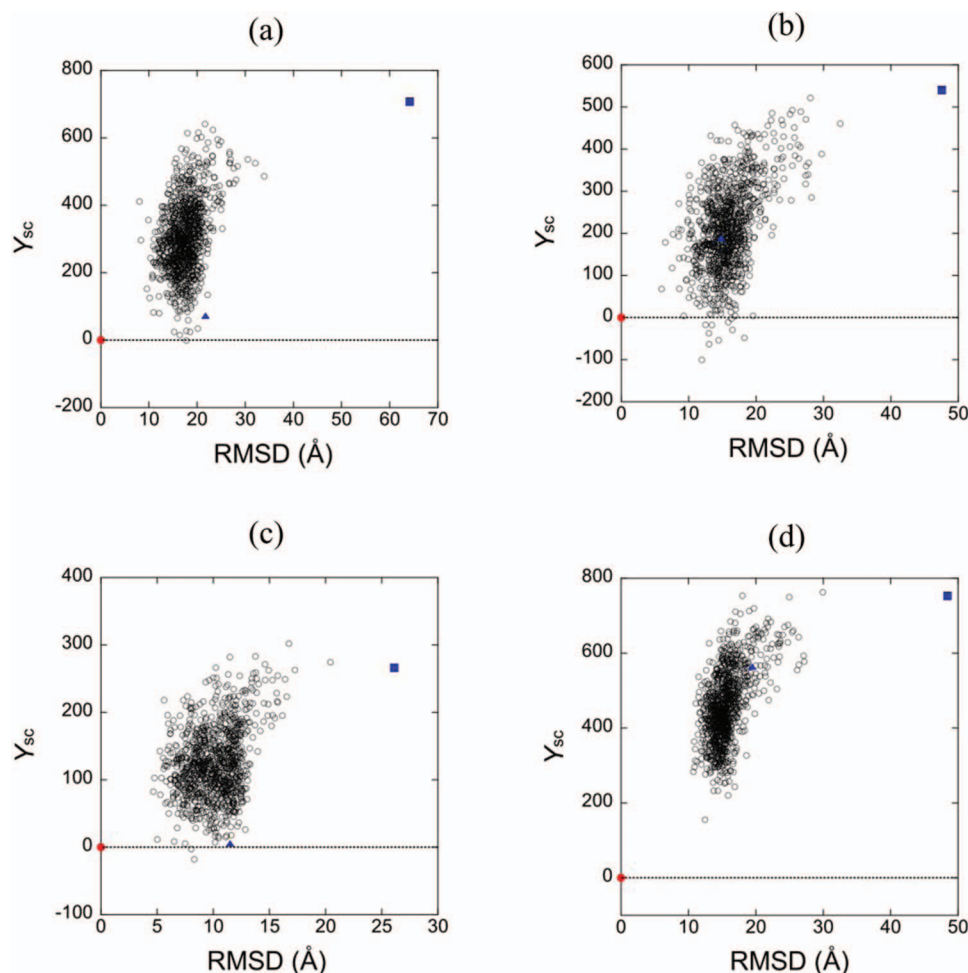


FIG. 8. Relation between Y_{sc} and RMSD from the native structure in aqueous environment for human H chain ferritin (a), fertilization protein (b), CspB (c), and HB/FGF (d). Black circles: data points for the non-native structures. Red circle: data point for the native structure. Blue square: data point for “all α .” Blue triangle: data point for “all α -3.”

and

$$Y_{sc} = -S_{sc}/k_B - (-S_{sc,NS}/k_B), \quad (10)$$

respectively.

We plot Y_b in Fig. 7 and Y_{sc} in Fig. 8 against the RMSD from the NS. In terms of Y_b , as expected, “all α ” and “all α -3” (especially the former) are quite stable (see Fig. 1(a)). There can be significantly many non-native structures Y_b of which is lower than that of the NS: Such structures feature high α -helix content. We note that the magnitude of the ordinate in Fig. 8 is much larger than that in Fig. 7. The efficiency of the side-chain packing has dominant effects on Y . As a conspicuous example, the NS of HB/FGF is stabilized by the close packing of side chains which predominates over the unfavorable properties of intramolecular hydrogen bonding observed in Fig. 6(d).

The α -helix and β -sheet are very advantageous with respect to both of Δ and $-S$, and these structural units should be formed as much as possible. In aqueous solution, however, the solvent-entropy effect plays critical roles in the structural stability of a protein. In particular, the closer packing of side chains often takes priority to the better properties of intramolecular hydrogen bonding. As a result, the content of

the secondary structures in the NS is not necessarily high (a good example is HB/FGF with β -sheets alone, whose content of the secondary structures is only $\sim 37\%$; see Fig. 3(d)). The thorough, overall packing of side chains, which gives little space in the protein interior, is unique to the NS. To accomplish such specific packing, the α -helix and β -sheet contents are prudently adjusted in protein folding in aqueous solution.

C. Physics of structural stability in nonpolar environment

Figure 9, which should be compared with Fig. 6, shows the plot of X against Y for the NS, non-native structures, “all α ,” and “all α -3” in nonpolar solvent. We note that the magnitude of the abscissa in Fig. 9 is much larger than that in Fig. 6. This can readily be understood in the following way: In nonpolar solvent, the formation of intramolecular hydrogen bonds leads to the energy lowering regardless that they are exposed to the solvent or buried in the protein interior, with the result that the differences among the structures in X are magnified. By comparing Fig. 9(d) with Fig. 6(d), we notice that the NS of HB/FGF is less stable than many of the non-native structures with respect to X in aqueous solution, whereas this

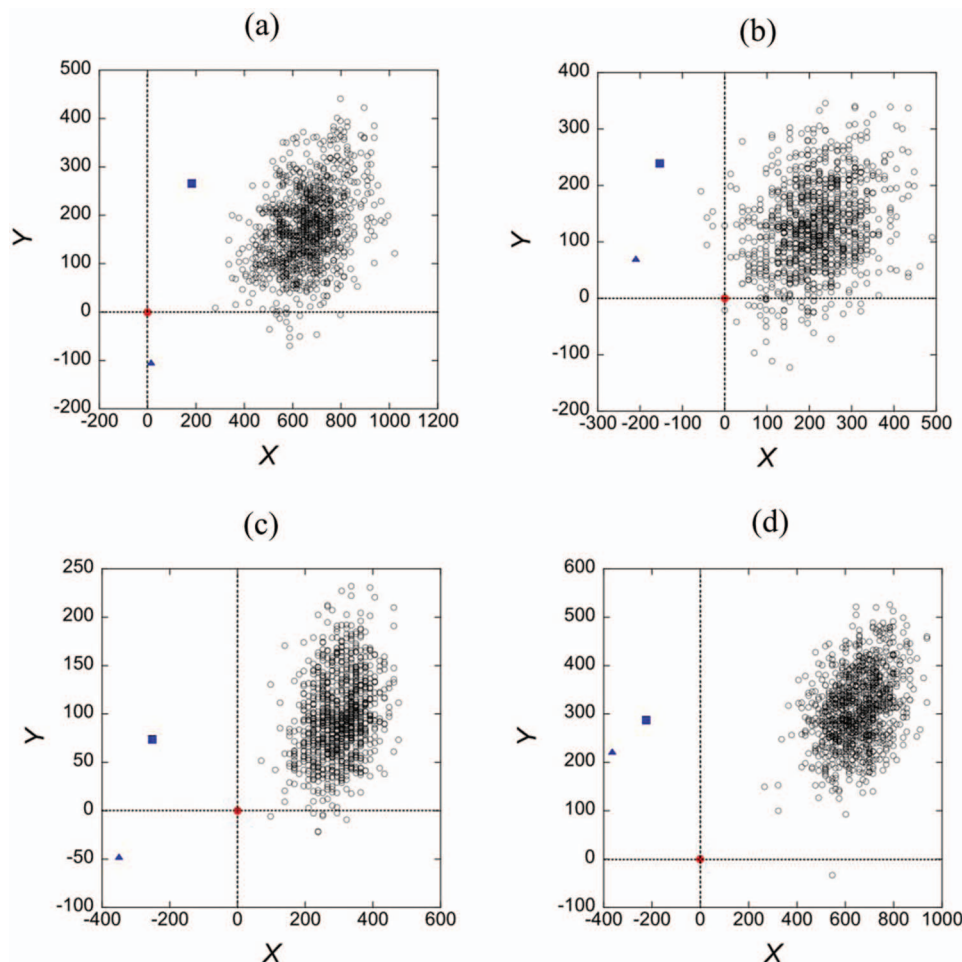


FIG. 9. Relation between X and Y in nonpolar environment for human H chain ferritin (a), fertilization protein (b), CspB (c), and HB/FGF (d). Black circles: data points for the non-native structures. Red circle: data point for the native structure. Blue square: data point for “all α .” Blue triangle: data point for “all α -3.”

is not true in nonpolar solvent. This is because the NS of HB/FGF possesses significantly many donors and acceptors without the formation of intramolecular hydrogen bonds, as described in Sec. III B. In aqueous environment, the closer packing of side chains takes priority to the intramolecular hydrogen bonding in the protein interior. Turning our attention to the ordinates of Figs. 9 and 6, we find that the magnitude of the ordinate in Fig. 9 is much smaller than that in Fig. 6. In nonpolar solvent, the effects of the solvent entropy on the structural stability of a protein are considerably reduced.

In Fig. 10(a), Δ is plotted against the α -helix content for fertilization protein. There is a strong tendency that Δ becomes lower as the α -helix content increases. The plot of Δ against the α -helix content for CspB is shown in Fig. 10(b). It is observed that on the whole Δ becomes lower as the α -helix content increases. In Fig. 10(c), Δ is plotted against the β -sheet content for CspB. There is no apparent correlation between Δ and the β -sheet content. Most of the non-native structures of CspB possess both α -helices and β -sheets. The α -helix and β -sheet contents are roughly related to each other in the way that as one of them increases, the other decreases. To extract the pure relation between Δ and the β -sheet content, the plot is made in Fig. 10(d) by choosing the structures with no α -helices. It appears that Δ tends to become lower

as the β -sheet content increases, but the tendency is significantly weaker than in cases of the α -helix content (we remark that the magnitude of the ordinate in Fig. 10(d) is smaller than that in Fig. 10(b)). This result is reasonable, because the donors and acceptors form intramolecular hydrogen bonds only on the inside upon the β -sheet formation (see Fig. 1(b)), whereas most of the donors and acceptors participate in forming intramolecular hydrogen bonds upon the α -helix formation. Further, the manner of constructing β -sheets is complex and diverse: A higher β -sheet content does not always lead to the formation of more intramolecular hydrogen bonds, because the formation is influenced by the specific details of the configuration of β -sheets.

An important result is that the solvent-entropy effect in nonpolar solvent is substantially weaker than that in aqueous solution. The formation of as many intramolecular hydrogen bonds as possible is imperative in nonpolar solvent. As discussed above, α -helices are more favorable than β -sheets from this point of view, and a structure with exceedingly high α -helix content (e.g., “all α -3”; see Fig. 3(g)) becomes especially stable. It is interesting to note that a variety of structures such as those in Figs. 3(a)–3(d) are stabilized in aqueous environment, whereas such variability is not exhibited in nonpolar solvent.

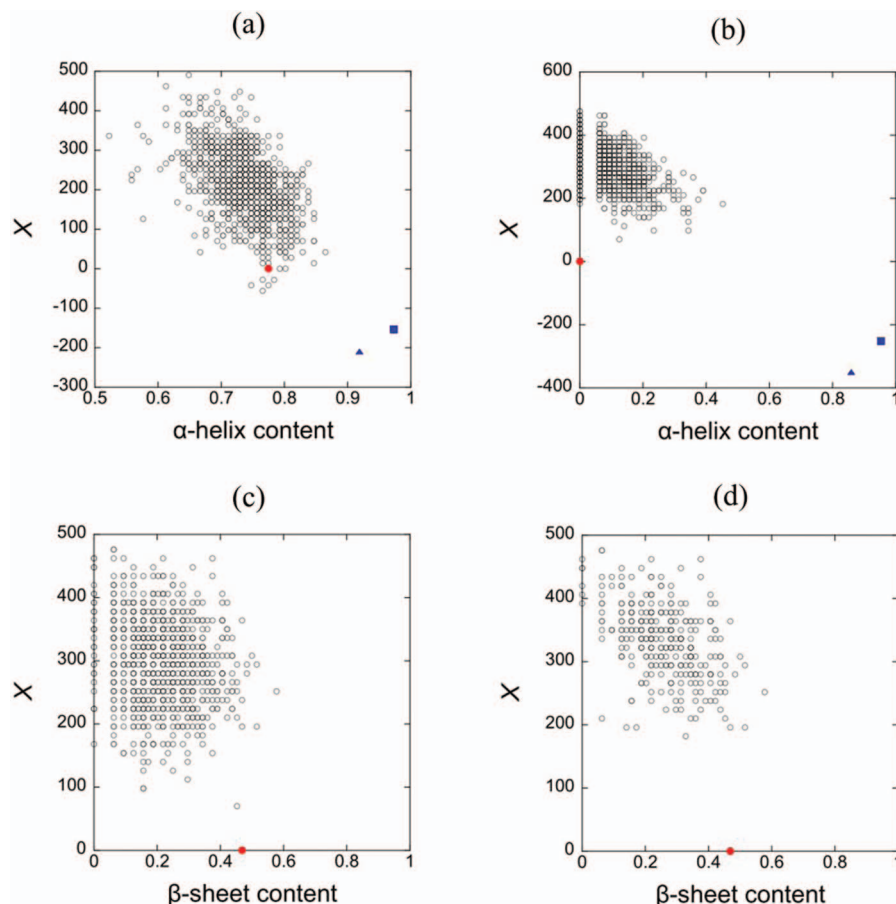


FIG. 10. Relation between Δ and α -helix content in nonpolar environment for fertilization protein (a) and CspB (b). Relation between Δ and β -sheet content in nonpolar environment for CspB (c, d). In (d), only the non-native structures with zero α -helix content are chosen. Black circles: data points for the non-native structures. Red circle: data point for the native structure. Blue square: data point for “all α .” Blue triangle: data point for “all α -3.” The data points for “all α ” and “all α -3” are not included in (c) and (d).

D. Structural stability of a membrane protein

Since it is difficult to experimentally determine the NS of a membrane protein, the structure prediction by a theoretical approach or by a computer simulation plays crucially important roles.^{53–58} The membrane protein is a typical example of proteins in nonpolar environment. In this case, “solvent molecules” correspond to the CH_2 , CH_3 , and CH groups, which are usually represented by united atoms, constituting nonpolar chains of lipid molecules. The entropic effect arising from the translational displacement of these groups can be taken into consideration by immersing the protein in our model nonpolar solvent. The particle diameter and packing fraction of the model solvent are set at those of water. In general, the solvent-entropy effect becomes larger as the solvent diameter decreases or the packing fraction increases. The diameters of the CH_2 , CH_3 , and CH groups are larger than the molecular diameter of water but their packing fraction is higher than the water value. (Water can be characterized by its relatively low packing fraction due to hydrogen bonds.) These two properties are rather compensating, and the model solvent is expected to provide a good model of the environment.

The two-stage model⁵⁹ has been proposed for the structure formation process. In this model, it is assumed that in-

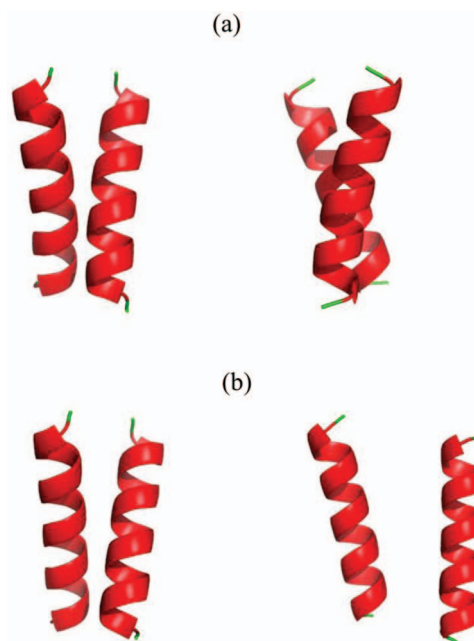


FIG. 11. Ribbon representations for transmembrane α -helical segments of GpA in the native structure viewed from two different angles (a) and for those in two representative non-native structures (b). They are drawn using the PyMOL.

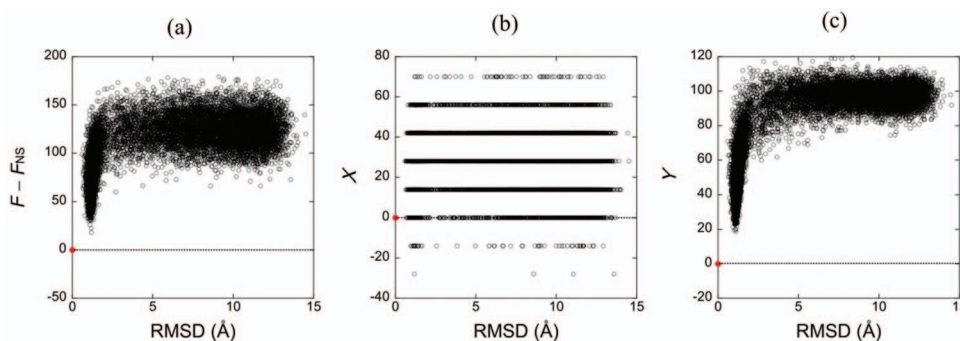


FIG. 12. Relation between F (a), X (b), and Y (c) and RMSD from the native structure in nonpolar environment for GpA. Black circles: data points for the non-native structures. Red circle: data point for the native structure.

dividual α -helices of the protein are separately stabilized as constituent domains within a lipid bilayer in the first stage and then the NS is completed by the side-to-side association of these helices in the second stage. The high propensity to form α -helices in the first stage can reasonably be understood from our result described above. In this section, we are concerned with the second stage.

Kokubo and Okamoto⁶⁰ studied transmembrane dimer of glycoprotein A, GpA (PDB code: 1afo), one of the membrane proteins whose native structures are experimentally available.⁶¹ This protein has two transmembrane α -helical segments and is a paradigm for the exploration of the helix-helix interaction within a lipid bilayer.^{55,56} Figure 11(a) shows ribbon representations for the arrangement of the two segments in the NS. The number of residues in each segment is 18 and the amino-acid sequence is TLIIFGV-MAGVIGTILLI. In the study of Kokubo and Okamoto,⁶⁰ the structure of GpA stabilized in the second stage was predicted using a Monte Carlo (MC) simulation based on the replica-exchange method (REM) with the CHARMM param19 parameter set for the force field. The interaction between the segments and nonpolar chains of lipid molecules was neglected as in usual studies.^{55,56} The backbone structures were fixed during the simulation and the MC-move type was taken to be rigid translation and rotation of each segment and torsion-angle rotations of side chains. When the dielectric constant was set at unity, the energy took the lowest value for a structure whose RMSD from the NS was 0.54 Å. The structure was claimed to be sufficiently close to the NS. However, it was not definite if the structure could be identified as the most stable one or not. The entropic effect arising from the translational displacement of the CH₂, CH₃, and CH groups was not taken into account.

Here we calculate F for the NS and a number of non-native structures (the total number is $\sim 15\,000$) generated by Kokubo and Okamoto⁶⁰ in the MC simulation based on the REM. The unrealistic overlaps of protein atoms in the NS and non-native structures are removed by local minimization of the energy function in the manner described in Sec. II F. The arrangements of the two segments in the NS and in two representative non-native structures are illustrated in Fig. 11. ($F - F_{\text{NS}}$), X , and Y are plotted against the RMSD from the NS in Fig. 12. It is found that the NS is the most stable (see Fig. 12(a)). In the NS, no hydrogen bonds are formed

between the two α -helical segments.⁶¹ For this reason, as observed in Fig. 12(b), the NS is less stable than significantly many non-native structures in terms of Λ or X . Interestingly, Fig. 12(c) indicates that the NS is the most stable in terms of $-S$ or Y . Thus, α -helices are chosen as fundamental structural units, but their arrangement is performed through the entropic effect originating from the translational displacement of the CH₂, CH₃, and CH groups. This result could be applicable to general membrane proteins for which the structures looking like those in Fig. 3(g) and Fig. 11(a) are stabilized.

IV. CONCLUDING REMARKS

We have investigated the structural stability of proteins in aqueous and nonpolar environments using our free-energy function¹⁴⁻¹⁶ $F = (\Lambda - TS)/(k_B T_0) = \Lambda/(k_B T_0) - S/k_B$ ($T_0 = 298$ K and the absolute temperature T is set at T_0) and its energetic and entropic components, $\Lambda/(k_B T_0)$ and S/k_B , respectively. Two solvents, water and nonpolar solvent, are considered. The nonpolar solvent provides an extreme example of nonaqueous environment. S is the solvation entropy and calculated using a hybrid of the angle-dependent¹⁷⁻²⁹ or radial-symmetric³⁰ integral equation theory and the morphometric approach.³¹⁻³³ As the backbone and side chains are more efficiently packed, the positive quantity, $-S$, becomes smaller. This is ascribed primarily to a smaller excluded volume, a larger total volume available to the translational displacement of solvent molecules, less solvent crowding, a larger number of accessible translational configurations of the solvent, and higher solvent entropy. When the solvent is water, Λ is calculated in accordance with a simple manner which still accounts for physically the most important factors: protein intramolecular and protein-water hydrogen bonds. A procedure suited to the calculation of Λ for a nonpolar solvent is newly proposed. In nonpolar environment, lower Λ simply implies more intramolecular hydrogen bonds formed. In aqueous environment, a transition to a more compact structure of a protein accompanies the break of protein-solvent hydrogen bonds: As the number of donors and acceptors buried without intramolecular hydrogen bonding increases, Λ becomes higher. Λ is positive and negative in aqueous and nonpolar environments, respectively. A number of different structures of significantly many proteins are considered as described in Sec. II F.

Though the α -helix and β -sheet are very advantageous with respect to both of Δ and $-S$ and these structural units should be formed as much as possible in this respect, the structural stability of a protein in aqueous solution is largely influenced by the solvent-entropy effect. The closer packing of side chains is particularly important, and to accomplish such packing, the α -helix and β -sheet contents are prudently adjusted in protein folding. As a result, the total content of the secondary structures in the NS is not necessarily high, some proteins possess only α -helices or β -sheets while others consist of mixtures of α -helices and β -sheets, and the relative content of the α -helix and β -sheet largely varies from protein to protein: A variety of structures are stabilized in aqueous solution (see Figs. 3(a)–3(d)). We thus find that the structural variability of proteins in aqueous environment is presented by the water-entropy effect.

In nonpolar solvent, on the other hand, the solvent-entropy effect becomes substantially weaker. The formation of as many intramolecular hydrogen bonds as possible is highly important. It is shown that α -helices are more favorable than β -sheets from this viewpoint, and a structure with exceedingly high α -helix content (e.g., that in Fig. 3(g)) is the most stable: α -helices are exclusively chosen as the secondary structures. The variability of the protein structures is no more exhibited in nonpolar solvent. It is experimentally known that alcohol induces a protein to form α -helices.^{8–11} The power of this induction is strengthened when the alcohol concentration is higher and the hydrocarbon group of an alcohol molecule is more bulky. The addition of a larger amount of alcohol and a more bulky hydrocarbon group make the solvent *more non-aqueous* (i.e., closer to nonpolar environment), and these alcohol effects can be understood on the basis of our theoretical results.

For a membrane protein,⁶⁰ we have succeeded in showing that the NS is more stable in terms of our free-energy function than any of $\sim 15\,000$ non-native structures generated by a computer simulation. In the two-stage model for the structure formation process,⁵⁹ it is assumed that individual helices of the protein are separately stabilized as constituent domains within a lipid bilayer in the first stage and then the NS is completed by the side-to-side association of these helices in the second stage. The high propensity to form α -helices in the first stage can reasonably be understood from our result for nonpolar environment. The arrangement of the fundamental structural units, α -helices, in the second stage is performed through the entropic effect originating from the translational displacement of the CH_2 , CH_3 , and CH groups which constitute nonpolar chains of lipid molecules. To the best of our knowledge, the importance of this entropic effect has not yet been pointed out in the literature.

ACKNOWLEDGMENTS

We thank Yuko Okamoto for sending us the non-native structures of transmembrane dimer of glycophorin A. The computer program for the morphometric approach was developed with Roland Roth and Yuichi Harano. This work was supported by Grants-in-Aid for Scientific Research on Innovative Areas (Grant No. 20118004) and that on (B) (Grant No.

22300100) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and by Grant-in-Aid for JSPS (Japan Society for the Promotion of Science) fellows.

- ¹Y. Harano and M. Kinoshita, *Chem. Phys. Lett.* **399**, 342 (2004).
- ²Y. Harano and M. Kinoshita, *Biophys. J.* **89**, 2701 (2005).
- ³T. Yoshidome, M. Kinoshita, S. Hirota, N. Baden, and M. Terazima, *J. Chem. Phys.* **128**, 225104 (2008).
- ⁴M. Kinoshita, *Front. Biosci.* **14**, 3419 (2009).
- ⁵M. Kinoshita, *Int. J. Mol. Sci.* **10**, 1064 (2009).
- ⁶S. Yasuda, T. Yoshidome, H. Oshima, R. Kodama, Y. Harano, and M. Kinoshita, *J. Chem. Phys.* **132**, 065105 (2010).
- ⁷N. Baden, S. Hirota, T. Takabe, N. Funasaki, and M. Terazima, *J. Chem. Phys.* **127**, 175103 (2007).
- ⁸N. Hirota-Nakaoka and Y. Goto, *Bioorg. Med. Chem.* **7**, 67 (1999).
- ⁹V. N. Uversky, N. V. Narizhneva, S. O. Kirschstein, S. Winter, and G. Löber, *Folding Des.* **2**, 163 (1997).
- ¹⁰Y. O. Kamatari, T. Konno, M. Kataoka, and K. Akasaka, *J. Mol. Biol.* **259**, 512 (1996).
- ¹¹N. Hirota, K. Mizuno, and Y. Goto, *J. Mol. Biol.* **275**, 365 (1998).
- ¹²J. U. Bowie, *Nature (London)* **438**, 581 (2005).
- ¹³L. E. Hedin, K. Illergård, and A. Elofsson, *J. Proteome Res.* **10**, 3324 (2011).
- ¹⁴Y. Harano, R. Roth, Y. Sugita, M. Ikeguchi, and M. Kinoshita, *Chem. Phys. Lett.* **437**, 112 (2007).
- ¹⁵T. Yoshidome, K. Oda, Y. Harano, R. Roth, Y. Sugita, M. Ikeguchi, and M. Kinoshita, *Proteins* **77**, 950 (2009).
- ¹⁶S. Yasuda, T. Yoshidome, Y. Harano, R. Roth, H. Oshima, K. Oda, Y. Sugita, M. Ikeguchi, and M. Kinoshita, *Proteins* **79**, 2161 (2011).
- ¹⁷P. G. Kusalik and G. N. Patey, *J. Chem. Phys.* **88**, 7715 (1988).
- ¹⁸P. G. Kusalik and G. N. Patey, *Mol. Phys.* **65**, 1105 (1988).
- ¹⁹M. Kinoshita and M. Harada, *Mol. Phys.* **81**, 1473 (1994).
- ²⁰M. Kinoshita, S. Iba, and M. Harada, *J. Chem. Phys.* **105**, 2487 (1996).
- ²¹M. Kinoshita and D. R. Bérard, *J. Comput. Phys.* **124**, 230 (1996).
- ²²N. M. Cann and G. N. Patey, *J. Chem. Phys.* **106**, 8165 (1997).
- ²³D. R. Bérard, M. Kinoshita, N. M. Cann, and G. N. Patey, *J. Chem. Phys.* **107**, 4719 (1997).
- ²⁴M. Kinoshita, *J. Solution Chem.* **33**, 661 (2004).
- ²⁵M. Kinoshita, *J. Mol. Liq.* **119**, 47 (2005).
- ²⁶M. Kinoshita, N. Matubayasi, Y. Harano, and M. Nakahara, *J. Chem. Phys.* **124**, 024512 (2006).
- ²⁷M. Kinoshita, *J. Chem. Phys.* **128**, 024507 (2008).
- ²⁸M. Kinoshita and M. Suzuki, *J. Chem. Phys.* **130**, 014707 (2009).
- ²⁹M. Kinoshita and T. Yoshidome, *J. Chem. Phys.* **130**, 144705 (2009).
- ³⁰J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic, London, 2006).
- ³¹P. M. König, R. Roth, and K. R. Mecke, *Phys. Rev. Lett.* **93**, 160601 (2004).
- ³²R. Roth, Y. Harano, and M. Kinoshita, *Phys. Rev. Lett.* **97**, 078101 (2006).
- ³³R. Kodama, R. Roth, Y. Harano, and M. Kinoshita, *J. Chem. Phys.* **135**, 045103 (2011).
- ³⁴T. Imai, Y. Harano, M. Kinoshita, A. Kovalenko, and F. Hirata, *J. Chem. Phys.* **125**, 024911 (2006).
- ³⁵A. 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. Kucsera, 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. Wiórkiewicz-Kucsera, D. Yin, and M. Karplus, *J. Phys. Chem. B* **102**, 3586 (1998).
- ³⁶H. Mishima, S. Yasuda, T. Yoshidome, H. Oshima, Y. Harano, M. Ikeguchi, and M. Kinoshita, *J. Phys. Chem. B* **116**, 7776 (2012).
- ³⁷M. L. Connolly, *J. Appl. Crystallogr.* **16**, 548 (1983).
- ³⁸M. L. Connolly, *J. Am. Chem. Soc.* **107**, 1118 (1985).
- ³⁹Y. Harano, T. Yoshidome, and M. Kinoshita, *J. Chem. Phys.* **129**, 145103 (2008).
- ⁴⁰T. Yoshidome, Y. Harano, and M. Kinoshita, *Phys. Rev. E* **79**, 011912 (2009).
- ⁴¹T. Yoshidome and M. Kinoshita, *Phys. Rev. E* **79**, 030905(R) (2009).
- ⁴²H. Oshima, T. Yoshidome, K. Amano, and M. Kinoshita, *J. Chem. Phys.* **131**, 205102 (2009).
- ⁴³K. Amano, T. Yoshidome, Y. Harano, K. Oda, and M. Kinoshita, *Chem. Phys. Lett.* **474**, 190 (2009).
- ⁴⁴K. Oda, R. Kodama, T. Yoshidome, M. Yamanaka, Y. Sambongi, and M. Kinoshita, *J. Chem. Phys.* **134**, 025101 (2011).

- ⁴⁵H. Oshima, S. Yasuda, T. Yoshidome, M. Ikeguchi, and M. Kinoshita, *Phys. Chem. Chem. Phys.* **13**, 16236 (2011).
- ⁴⁶S. F. Sneddon, D. J. Tobias, and C. L. Brooks III, *J. Mol. Biol.* **209**, 817 (1989).
- ⁴⁷J. W. Ponder and F. M. Richards, *J. Comput. Chem.* **8**, 1016 (1987).
- ⁴⁸I. K. McDonald and J. M. Thornton, *J. Mol. Biol.* **238**, 777 (1994).
- ⁴⁹K. T. Simons, R. Bonneau, I. Ruczinski, and D. Baker, *Proteins* **37** (Suppl. 3), 171 (1999).
- ⁵⁰B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
- ⁵¹M. Feig, J. Karanicolas, and C. L. Brooks III, *J. Mol. Graphics Modell.* **22**, 377 (2004).
- ⁵²W. Kabsch and C. Sander, *Biopolymers* **22**, 2577 (1983).
- ⁵³W. R. Taylor, D. T. Jones, and N. M. Green, *Proteins* **18**, 281 (1994).
- ⁵⁴M. Suwa, T. Hirokawa, and S. Mitaku, *Proteins* **22**, 363 (1995).
- ⁵⁵P. D. Adams, D. M. Engelman, and A. T. Brünger, *Proteins* **26**, 257 (1996).
- ⁵⁶R. V. Pappu, G. R. Marshall, and J. W. Ponder, *Nat. Struct. Biol.* **6**, 50 (1999).
- ⁵⁷T. Hirokawa, J. Uechi, H. Sasamoto, M. Suwa, and S. Mitaku, *Protein Eng.* **13**, 771 (2000).
- ⁵⁸N. Vaidehi, W. B. Floriano, R. Trabanino, S. E. Hall, P. Freddolino, E. J. Choi, G. Zamanakos, and W. A. Goddard III, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12622 (2002).
- ⁵⁹J. L. Popot and D. M. Engelman, *Annu. Rev. Biochem.* **69**, 881 (2000).
- ⁶⁰H. Kokubo and Y. Okamoto, *Chem. Phys. Lett.* **383**, 397 (2004).
- ⁶¹K. R. MacKenzie, J. H. Prestegard, and D. M. Engelman, *Science* **276**, 131 (1997).