



SCIENTIFIC REPORTS

OPEN

Theoretical analysis on thermodynamic stability of chignolin

Tomonari Sumi ^{1,2} & Kenichiro Koga ^{1,2}

Understanding the dominant factor in thermodynamic stability of proteins remains an open challenge. Kauzmann's hydrophobic interaction hypothesis, which considers hydrophobic interactions between nonpolar groups as the dominant factor, has been widely accepted for about sixty years and attracted many scientists. The hypothesis, however, has not been verified or disproved because it is difficult, both theoretically and experimentally, to quantify the solvent effects on the free energy change in protein folding. Here, we developed a computational method for extracting the dominant factor behind thermodynamic stability of proteins and applied it to a small, designed protein, chignolin. The resulting free energy profile quantitatively agreed with the molecular dynamics simulations. Decomposition of the free energy profile indicated that intramolecular interactions predominantly stabilized collapsed conformations, whereas solvent-induced interactions, including hydrophobic ones, destabilized them. These results obtained for chignolin were consistent with the site-directed mutagenesis and calorimetry experiments for globular proteins with hydrophobic interior cores.

Understanding the dominant factor behind thermodynamic stability of proteins remains a challenging issue in biochemistry, biophysics, and molecular biology¹⁻³. Several theories explaining protein stability have been proposed. In 1936, Pauling and Mirsky suggested that a protein achieved a uniquely defined conformation held in place by N-H...O hydrogen bonds between the nitrogen and oxygen atoms in the peptide chain, the interaction energy of each bond being approximately 5 kcal/mol⁴. Three years later, Bernal suggested that the hydrophilic residues of a protein were exposed to the aqueous solution, whereas the hydrophobic parts were in contact with each other in the interior of the protein⁵. In 1951, Pauling's group discovered the most important structural elements in globular proteins: alpha helices⁶ and beta sheets⁷. They, furthermore, pointed out that the backbone N-H and O forming intramolecular hydrogen bonds were approximately 2 kcal/mol more stable than those forming intermolecular hydrogen bonds with surrounding water molecules⁷. In 1959, Kauzmann concluded in his seminal review⁸ that hydrophobic attraction was a dominant factor in the thermodynamic stability of the folded conformation for many globular proteins. This has been supported by the following experimental observations: (i) the change in Gibbs energy for transferring a small nonpolar molecule from an aqueous solution to an organic solvent is large and negative⁸; (ii) the net effect of electrostatic interactions on protein stability is negligibly small⁹; and (iii) numerous nonpolar residues are indeed located in the interior of globular proteins^{10,11}.

In the late 1980s, it became possible to examine the dominant factor behind protein stability by applying site-directed mutagenesis. It was shown that (1) both hydrophobic interactions¹² and intramolecular hydrogen bonding¹³ contributed substantially to protein stability; (2) the enhancement of van der Waals interactions due to tight packing in the protein interior caused by the replacement of small hydrophobic residues with larger ones resulted in increased protein stability¹⁴; and (3) the effect of hydrogen bonding of peptide groups on protein stability was comparable to that of hydrogen bonding of side chains^{13,15}. These observations suggest the importance of intramolecular interactions in protein stability. Nevertheless, the thermodynamic stability of proteins has been basically modeled according to Kauzmann's hydrophobic interaction hypothesis (e.g., refs¹⁶⁻¹⁸).

Molecular dynamics (MD) simulations are a powerful tool that enables us to investigate large conformational changes in proteins¹⁹⁻²⁷. Generalized ensemble MD simulations have been applied to calculate temperature and pressure dependence of free energy profile of proteins and peptides^{24,25}. Long equilibrium MD simulations have

¹Research Institute for Interdisciplinary Science, Okayama University, 3-1-1 Tsushima-Naka, Kita-ku, Okayama, 700-8530, Japan. ²Department of Chemistry, Faculty of Science, Okayama University, 3-1-1 Tsushima-Naka, Kita-ku, Okayama, 700-8530, Japan. Correspondence and requests for materials should be addressed to T.S. (email: sumi@okayama-u.ac.jp)

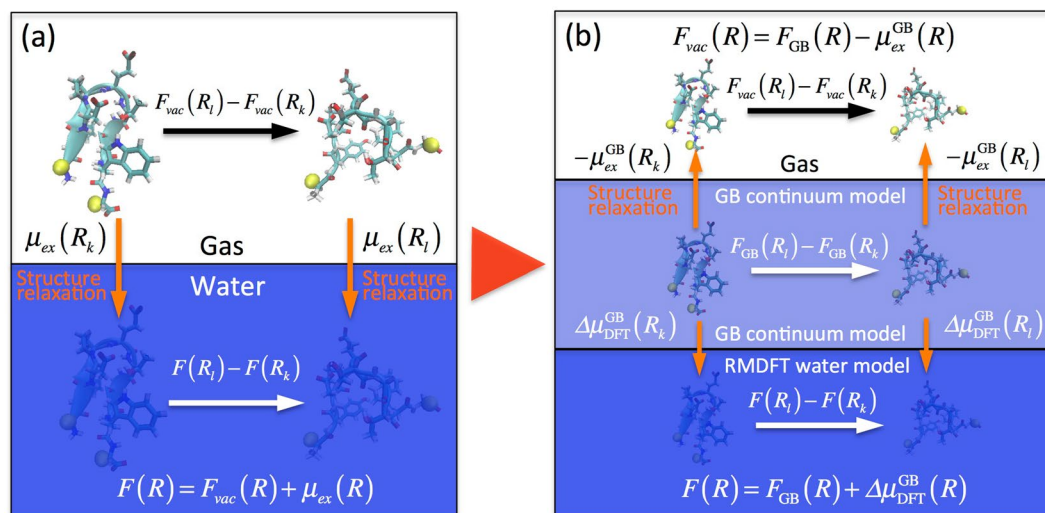


Figure 1. Calculation scheme of free energy profile of a protein. The distance between alpha carbon atoms at the C-terminus and N-terminus (shown as yellow spheres) is introduced as the coordinate R specifying the dimensions of the protein. **(a)** Thermodynamic cycle depicting the relationship in Eq. 1 among the free energy profile of the protein in water, $F(R)$, the free energy profile in vacuum, $F_{vac}(R)$, and the excess chemical potential profile, $\mu_{ex}(R)$, at a distance R . **(b)** Two different thermodynamic cycles, which allow for the calculation of $F(R)$ and $F_{vac}(R)$, if the following are obtained: free energy profile determined by the generalized Born (GB) model umbrella sampling MD simulations with a dielectric constant $\epsilon_r = 80$, $F_{GB}(R)$, excess chemical potential of the protein in the GB model with $\epsilon_r = 80$ (Eq. 4), $\mu_{ex}^{GB}(R)$, and free energy difference between the protein in water described by the GB model and that described by the RMDFT model (Eq. 7), $\Delta\mu_{DFT}^{GB}(R)$.

been performed to characterize folding pathways and free energy changes^{26,27}. In general, a free energy profile of a protein can be expressed as a function of a coordinate R :

$$F(R) = F_{vac}(R) + \mu_{ex}(R), \quad (1)$$

where $F_{vac}(R)$ is the free energy profile in vacuum and $\mu_{ex}(R)$ is the excess chemical potential profile of the protein, i.e., the free energy change for hydration, at the coordinate R (Fig. 1a). $F_{vac}(R)$ consists of intramolecular energy and entropy:

$$F_{vac}(R) = E_{vac}^{intra}(R) - TS_{vac}^{intra}(R). \quad (2)$$

$\mu_{ex}(R)$ in Eq. 1, which acts as a solvent-induced interaction on the protein, can be expressed as

$$\mu_{ex}(R) = \mu_{nonpol}(R) + \mu_{pol}(R), \quad (3)$$

where $\mu_{nonpol}(R)$ is the nonpolar part of $\mu_{ex}(R)$, which is calculated by omitting all partial charges on the protein and $\mu_{pol}(R)$ is the remaining polar part of $\mu_{ex}(R)$. $\mu_{nonpol}(R)$ is expected to provide the upper limit of the contribution of the solvent-induced hydrophobic interaction to the hydrophobic collapse of the protein since $\mu_{nonpol}(R)$ includes the nonpolar contributions to $\mu_{ex}(R)$ arising from all polar residues as well. Such a decomposition of $F(R)$ should provide insights into understanding the dominant factor in the thermodynamic stability of the protein. Here, it can be seen that $\mu_{ex}(R)$ is essentially different from the simple ensemble-average value of solvation free energy calculated using conformations generated under the coordinate R in the solvent. This is because $\mu_{ex}(R)$ includes the effect of conformation relaxation during the gradual annihilation of either the protein or all the solvent molecules²⁸ (see Fig. 1a). Although the importance of the conformation entropy in $F_{vac}(R)$ has widely been realized, the effect of conformation relaxation on $\mu_{ex}(R)$ has been overlooked or simply ignored (see an example in the Supplementary Information); however, we have explicitly evaluated this as described below. Furthermore, through liquid-state DFT, it was easy to determine $\mu_{ex}(R)$ and $\mu_{nonpol}(R)$ (or $\mu_{pol}(R)$) as well as the intramolecular conformation entropy $TS_{vac}^{intra}(R)$, which would have been hard if we had employed explicit-solvent all-atom MD simulations.

In this study, we present an efficient computational method to evaluate the free energy profile and the components as a function of a coordinate R using a combination of continuum solvent MD simulations and a recently developed reference-modified density functional theory (RMDFT) for calculation of solvation free energy^{29–32}. The reliability of RMDFT in calculating the solvation free energy has been demonstrated by comparing experiments on organic solute molecules^{29,32}. In contrast to continuum solvent models, the present method allows for hydration effects to be taken into consideration at a molecular level. We applied this method to a small, designed protein, chignolin, consisting of ten amino acids with the sequence GYDPETGTWG³³, and revealed a dominant factor responsible for the thermodynamic stability as well as temperature- and pressure-induced unfolding of chignolin.

Computational Method

We calculated the free energy profile $F(R)$ and its components $F_{vac}(R)$ and $\mu_{ex}(R)$ for chignolin as a function of a coordinate R . The distance between the alpha carbon atoms at the C-terminus and the N-terminus was chosen as the coordinate R . It is not easy to determine the most suitable single reaction coordinate to characterize folding kinetics. However, we aimed to obtain the free energy profile $F(R)$ as a function of the measure of the dimensions of chignolin, and we chose the end-to-end distance as R as it is one of the common measures of the dimensions of polymers. It should be noted that the free energy *difference* and its decomposition between folded and extended conformations do not depend on the pathway and depend only on the initial and final state. Thus, the conclusion based on the decomposed free energy differences is not affected by the choice of the coordinate parameter.

There are steep energy barriers associated with structural changes involving cleavage of intramolecular hydrogen bonding, which make it difficult to directly calculate the free energy profile in vacuum, $F_{vac}(R)$, using umbrella sampling. In fact, exchange of the intramolecular hydrogen bonding in vacuum is very rarer than in water due to the higher energy barriers since no water molecules are surrounding chignolin. We thus chose an alternative approach: first, we calculated the free energy profile $F_{GB}(R)$ in a continuum solvent with a dielectric constant of water ($\epsilon_r = 80$), $F_{GB}(R)$, using implicit-solvent generalized Born (GB) umbrella sampling MD simulations (Fig. 1b), which are much faster than explicit-solvent all-atom umbrella sampling MD simulations; second, as shown later, we obtained $F_{vac}(R)$ from $F_{GB}(R)$.

The excess chemical potential profile for the GB model, $\mu_{ex}^{GB}(R)$, was calculated by the free energy perturbation method as follows:

$$\mu_{ex}^{GB}(R) = \sum_{a=0}^n [\Delta\mu_{\epsilon_{a+1}}^{\epsilon_a}(R) - \Delta\mu_{\epsilon_a}^{\epsilon_{a+1}}(R)]/2, \quad (4)$$

with

$$\Delta\mu_{\epsilon_j}^{\epsilon_i}(R) = -k_B T \ln \left\langle \exp \left[-(\Delta G_{solv}^{GB}(\epsilon_j) - \Delta G_{solv}^{GB}(\epsilon_i))/k_B T \right] \right\rangle_R^{\epsilon_i}, \quad (5)$$

where n is the number of intermediate states between $\epsilon_r = 1$ (vacuum) and 80 (liquid water), $\langle \dots \rangle_R^{\epsilon_i}$ represents the ensemble average of conformations generated by GB MD simulations with a dielectric constant of ϵ_i , in which the coordinate is fixed at R , and $\Delta G_{solv}^{GB}(\epsilon_i)$ is the solvation free energy of chignolin calculated from the dielectric constant $\epsilon_r = \epsilon_i$. It should be noted that the solvation free energy given for each conformation serves as a potential energy term in the effective Hamiltonian of GB MD simulations. Thus, the solvation free energy appears in Eq. 5 instead of the protein-water interaction energy that usually appears in explicit-solvent all-atom MD free energy perturbation calculations (Appendix A in the Supplementary Information). As shown in Fig. 1b, the free energy profile in vacuum, $F_{vac}(R)$, can be obtained from

$$F_{vac}(R) = F_{GB}(R) - \mu_{ex}^{GB}(R). \quad (6)$$

An interval of $\Delta R = 0.1$ nm was employed, and thus, steep changes in $F_{vac}(R)$, included within an interval less than ΔR , were fully or partially omitted. However, this was not problematic because we focused on the overall profile of $F_{vac}(R)$.

The GB model is computationally favorable, but it is less accurate and tends to underestimate solvation of polar groups^{34–36}. To obtain a more reliable free energy profile, we calculated the free energy difference between the solute in water described by the GB model and that described by the RMDFT model,

$$\Delta\mu_{DFT}^{GB}(R) = -k_B T \ln \left\langle \exp \left[-(\Delta G_{hyd}^{DFT} - \Delta G_{solv}^{GB}(\epsilon_r = 80))/k_B T \right] \right\rangle_R^{\epsilon_r=80}, \quad (7)$$

where ΔG_{hyd}^{DFT} is the hydration free energy according to the RMDFT model and includes the effects of temperature and pressure³². Finally, the free energy profile in water, $F(R)$, and the excess chemical potential profile, $\mu_{ex}(R)$, were obtained from

$$F(R) = F_{GB}(R) + \Delta\mu_{DFT}^{GB}(R) \quad (8)$$

and

$$\mu_{ex}(R) = \mu_{ex}^{GB}(R) + \Delta\mu_{DFT}^{GB}(R), \quad (9)$$

respectively (Fig. 1b). It is noted that we can use the free energy perturbation by Eq. 7 to calculate $F(R)$ at high pressures, if the ensemble average $\langle \dots \rangle_R^{\epsilon_r=80}$ includes enough conformations that becomes important at the high pressures. The validity is all the free energy perturbation calculations used above should be assessed by the standard error. We were now able to evaluate separately the two components of the free energy profile, namely the purely intramolecular part, $F_{vac}(R)$, and the solvent-induced part, $\mu_{ex}(R)$. These were further decomposed as discussed below.

Results

Decomposition of the free energy profile. The curves in Fig. 2a represent the free energy profiles $F(R)$ and $F_{GB}(R)$ at 298 K. The distance $R = 0.5$ nm corresponds to the native state, and these free energy profiles are plotted, so that the value becomes zero at this distance. The misfolded state^{37–41} is found at distances around $R = 0.6$ nm (see Fig. S1 in the Supplementary Information) due to inaccurate force field parameters for glycine

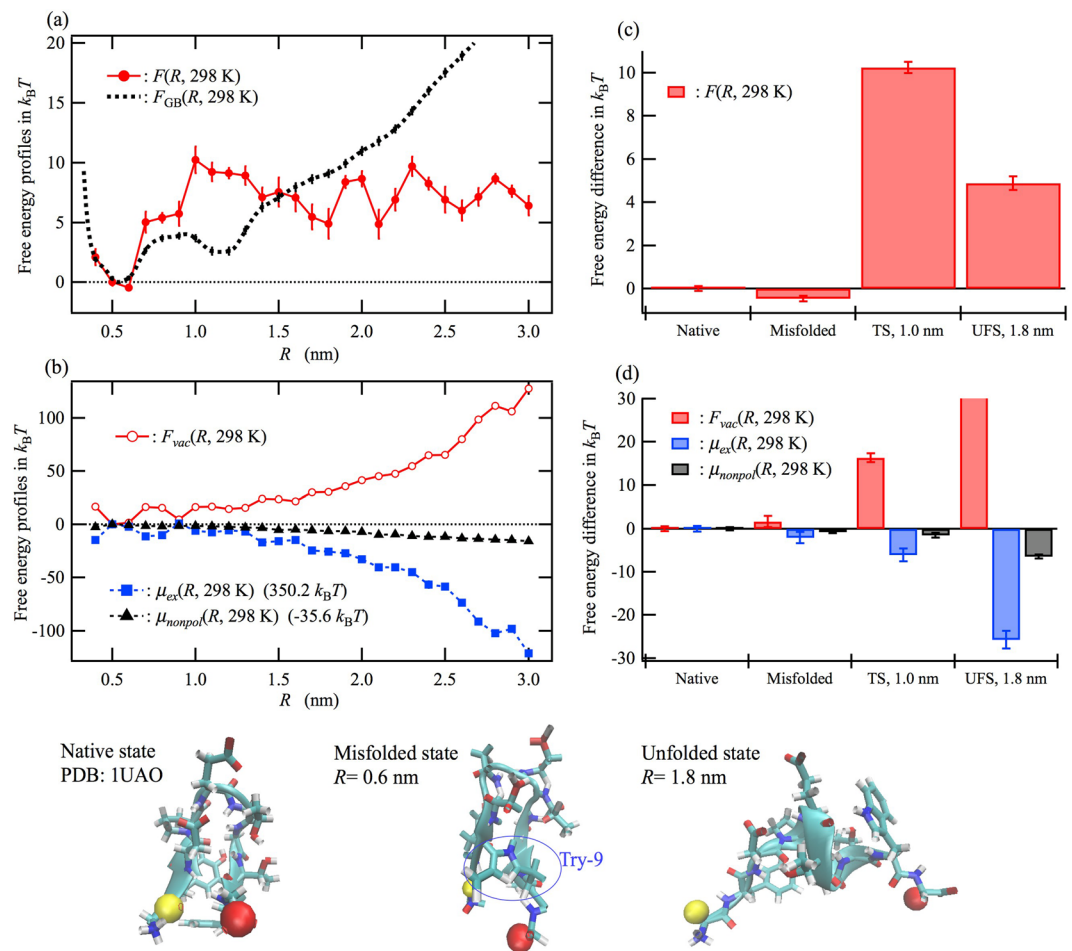


Figure 2. Free energy profiles of chignolin in water at 298 K. All profiles are shown in $k_B T$ and are shifted vertically, so that the value becomes zero at a distance of $R = 0.5$ nm, which corresponds to the native state. **(a)** Free energy profile calculated according to the GB model, $F_{GB}(R)$, and that corrected by the RMDFT method, $F(R)$. **(b)** Free energy profile in vacuum, $F_{vac}(R)$, excess chemical potential profile, $\mu_{ex}(R)$, and nonpolar part of $\mu_{ex}(R)$, $\mu_{nonpol}(R)$, calculated by removing all partial charges on chignolin (see Computational Details). **(c,d)** These free energy differences from the native state ($R = 0.5$ nm) are shown for the native state from the Protein Data Bank database (PDB: 1UAO), misfolded state obtained at $R = 0.6$ nm, and unfolded state obtained at $R = 1.8$ nm. The numbers beside the legends in **(b)** indicate the vertically shifted value for these profiles. Here and hereafter, the error bars for all profiles indicate the standard error. The shown ternary structures are for the native state from the Protein Data Bank database (PDB: 1UAO), misfolded state obtained at $R = 0.6$ nm, and unfolded state obtained at $R = 1.8$ nm. The yellow and red spheres depict the alpha carbon atoms at the C-terminus and N-terminus, respectively. In the misfolded state, the relative position of Try-9 compared with Try-2 is different from that of the native state because of rotation in backbone torsion angle ψ for Gly-7³⁹.

backbone³⁹. The free energy profile corrected by the RMDFT method, $F(R)$, sharply increases and then reaches a plateau as the distance R increases from 0.5 nm, whereas $F_{GB}(R)$ gradually increases. There are some small minima in the plateau region of $F(R)$ (e.g., $R = 1.8, 2.1,$ and 2.6 nm). A similar plateau in $F(R)$ has been reported by a generalized-ensemble all-atom MD simulation with explicit solvent by Okumura²⁵. The free energy difference between the native and denatured states was determined by Okumura to be $4.6 k_B T$. This value is quantitatively consistent with the $4.9 k_B T$ observed at the minimum on the plateau region of $F(R)$. This quantitative agreement shows the validity of the present method, at least, for chignolin. In contrast, the free energy profile of the GB model, $F_{GB}(R)$, increased with increasing R for $R > 1.3$ nm. This substantial difference indicates the necessity of using the RMDFT method to describe the solvation of chignolin in water.

The curves in Fig. 2b show the free energy profile in vacuum, $F_{vac}(R)$, the excess chemical potential profile, $\mu_{ex}(R)$, and the nonpolar part of $\mu_{ex}(R)$, $\mu_{nonpol}(R)$, at 298 K. $\mu_{ex}(R)$ decreases with increasing R , whereas $F_{vac}(R)$ increases. Thus, $\mu_{ex}(R)$, namely, the solvent-induced part of $F(R)$, appears to stabilize the unfolded state. Indeed, earlier theoretical studies on small peptides^{42,43} and several proteins^{44–46} predicted a lower solvation free energy for unfolded conformations than for folded ones. In addition, the nonpolar part $\mu_{nonpol}(R)$ also decreases with increasing R , indicating that the solvent-induced hydrophobic interaction $\mu_{nonpol}(R)$, which gives the upper limit of the hydrophobic contribution of $\mu_{ex}(R)$, also stabilizes the unfolded state rather than the folded state. This

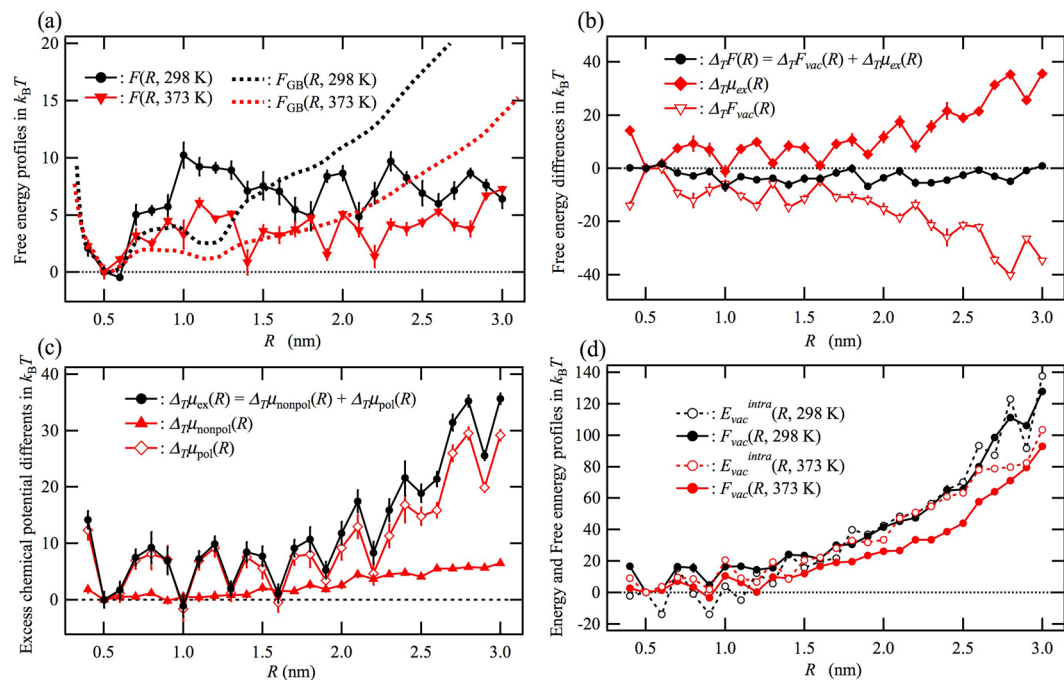


Figure 3. Comparison of free energy profiles for chignolin in water at 298 K and 373 K. **(a)** Free energy profile calculated according to the GB model, $F_{GB}(R)$, and that corrected by the RMDFT method, $F(R)$. **(b)** Difference between the free energy profiles at 373 K and 298 K, $\Delta_T F(R) = F(R, 373 \text{ K})/k_B T - F(R, 298 \text{ K})/k_B T$ and the corresponding differences for the two components, $\Delta_T F_{vac}(R)$ and $\Delta_T \mu_{ex}(R)$, where $\Delta_T F(R) = \Delta_T F_{vac}(R) + \Delta_T \mu_{ex}(R)$. **(c)** Nonpolar part and polar part for $\Delta_T \mu_{ex}(R)$, $\Delta_T \mu_{nonpol}(R)$ and $\Delta_T \mu_{pol}(R)$, respectively, where $\Delta_T \mu_{ex}(R) = \Delta_T \mu_{nonpol}(R) + \Delta_T \mu_{pol}(R)$. **(d)** Free energy profile in vacuum, $F_{vac}(R)$, and its energy part, $E_{vac}^{intra}(R)$, at 298 K and 373 K. The difference between $F_{vac}(R)$ and $E_{vac}^{intra}(R)$ corresponds to the entropic term of $F_{vac}(R)$, $-T\Delta S_{vac}^{intra}(R) = F_{vac}(R) - E_{vac}^{intra}(R)$ (see Eq. 2).

result qualitatively agrees with MD simulations of peptides by Kokubo *et al.*⁴³ A similar solvent-induced repulsive force has also been shown for large hydrophobic molecules (e.g., fullerene C_{60}) in water^{44,45}. The obtained results for chignolin indicate that the *intramolecular* interactions including van der Waals and electrostatic forces are a dominant factor in the collapse of chignolin because the intramolecular-force-driven collapse slightly conquers the solvent-induced expansion. This observation is consistent with large-scale molecular dynamics simulations for multi-peptide aggregations that are predominantly caused by both van der Waals and Coulomb interactions between hydrophobic amino acids⁴⁶. Furthermore, it is implied that a subtle balance between the competitive factors, i.e., the intramolecular interactions and the solvent-induced interactions determines the conformation of the native state. These results are inconsistent with Ben-Naim's theory where hydrophilic interactions play more important role on protein stability rather than hydrophobic ones^{2,47,48}; however, we agree with him in terms of the importance of electrostatic interactions in protein stability.

Figure 2c,d plot the free energy differences from the native state ($R = 0.5 \text{ nm}$) for the misfolded state ($R = 0.6 \text{ nm}$), transition state ($R = 1.0 \text{ nm}$), and unfolded state ($R = 1.8 \text{ nm}$). $F_{vac}(R)$ and $\mu_{ex}(R)$ are respectively increased and decreased at the misfolded state, implying that the relative stability between the native and misfolded state is determined as a result of the competition between the intramolecular interactions and the solvent-induced interactions. From the comparison with the unfolded state at $R = 1.8 \text{ nm}$, the elevation of $F(R)$ at $R = 1.0 \text{ nm}$ is attributable to the reduced stabilization in the solvent-induced interactions for the transition state.

Effect of temperature. The red solid curve in Fig. 3a shows the RMDFT free energy profile $F(R)$ at 373 K. The relative stability of unfolded conformations compared with that of the native state is lower at 373 K than at 298 K. The free energy change from the folded to unfolded state, ΔF_u , was experimentally determined by Honda *et al.*³³ to be $0.5 k_B T$ at 298 K and $-2.2 k_B T$ at 373 K. Not only the RMDFT method but also the generalized-ensemble all-atom MD simulation by Okumura²⁵ estimated the ΔF_u higher than the experimental values at both temperatures. On the other hand, the relative stabilization of the unfolded state by heating, $\Delta_T F_u = \Delta F_u(373 \text{ K})/k_B T - \Delta F_u(298 \text{ K})/k_B T$, where the former and latter were obtained as -0.1 and 2.7 , respectively, was, thus, estimated by RMDFT to be -2.8 , if we assume the sum of the state probabilities at $R = 0.5$ and 0.6 nm as the probability of the folded state. This value is comparable with the experimental value, -2.7 .

Figure 3b shows the difference in the free energy profile between 373 K and 298 K, $\Delta_T F(R) = F(R, 373 \text{ K})/k_B T - F(R, 298 \text{ K})/k_B T$, and the corresponding differences for the two components, $\Delta_T F_{vac}(R)$ and $\Delta_T \mu_{ex}(R)$, where $\Delta_T F(R) = \Delta_T F_{vac}(R) + \Delta_T \mu_{ex}(R)$. $\Delta_T \mu_{ex}(R)$ increases with increasing R mainly because of the increase in the electrostatic part of $\Delta_T \mu_{ex}(R)$, $\Delta_T \mu_{pol}(R)$, as shown in Fig. 3c. This result shows that the electrostatic part of the

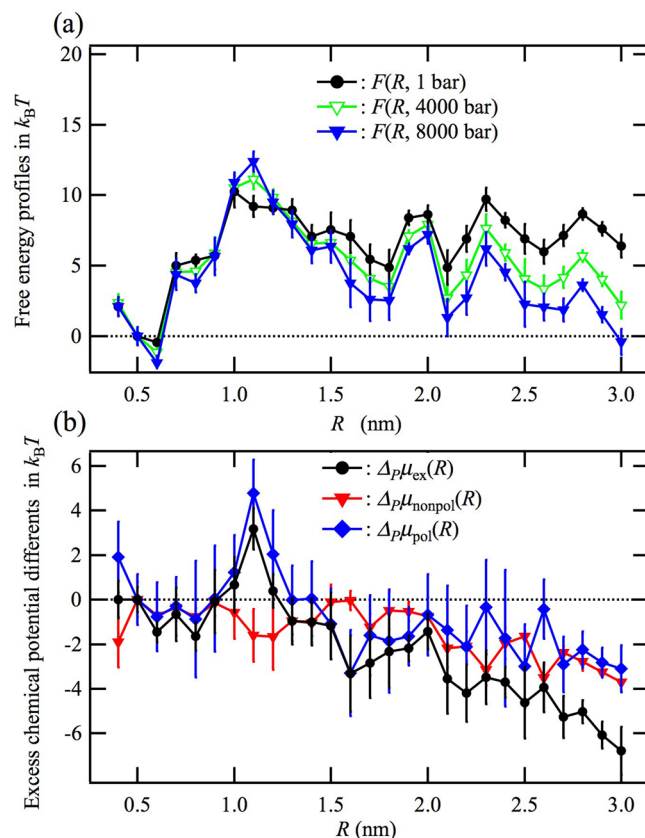


Figure 4. Effect of pressure on the free energy profile $F(R)$. (a) Pressure dependence of the free energy profile at 298 K. (b) Difference in the free energy between 8000 bar and 1 bar, $\Delta_p F(R) = F(R, 8000 \text{ bar})/k_B T - F(R, 1 \text{ bar})/k_B T$. Shown are the excess chemical potential difference, $\Delta_p \mu_{ex}(R)$, the nonpolar part $\Delta_p \mu_{nonpol}(R)$, and the electrostatic part, $\Delta_p \mu_{pol}(R) = \Delta_p \mu_{ex}(R) - \Delta_p \mu_{nonpol}(R)$, resulting from the independence of $F_{vac}(R)$ on pressure.

solvent-induced interaction significantly suppresses the high-temperature unfolding. In contrast, $\Delta_T F_{vac}(R)$ decreases with increasing R , as seen in Fig. 3b, indicating that the stabilization of the unfolded state at high temperature is attributable to the intramolecular free energy $F_{vac}(R)$. In Fig. 3d, we show a comparison between $F_{vac}(R)$ and the energy part of $F_{vac}(R)$, $E_{vac}^{intra}(R)$, at 298 K and 373 K. $E_{vac}^{intra}(R)$ is similar at 373 K and 298 K, whereas the intramolecular free energy $F_{vac}(R)$ for the unfolded conformations is lower at 373 K than at 298 K. We, therefore, conclude that the dominant factor in high-temperature unfolding is intramolecular conformation entropy, $-T S_{vac}^{intra}(R)$ (see Eq. 2), because the intramolecular-conformation-entropy-driven unfolding slightly overcomes the solvent-induced collapse.

Effect of pressure. Figure 4a shows the pressure dependence of $F(R)$ at 298 K. The unfolded state becomes more stable than the folded state with increasing pressure, although the barrier at $R = 1.1 \text{ nm}$ is somewhat raised by the pressurization. High-pressure unfolding of chignolin has previously been observed by FT-IR and FRET experiments⁴⁹ and a similar behavior for $F(R)$ has also been obtained by the generalized-ensemble all-atom MD simulations by Okumura²⁵. The partial molar volume change from the folded to unfolded state, $\Delta V = (\partial \Delta F_u / \partial P)_T$, which is a crucial thermodynamic quantity that characterizes the pressure-induced unfolding, was obtained by RMDFT as $-5.3 \text{ cm}^3/\text{mol}$ in the same manner as the heat denaturation. This value is comparable with the all-atom MD simulation value obtained by Okumura, $-5.6 \text{ cm}^3/\text{mol}$ ²⁵. Both the values are slightly larger than the experimental value, $-8.8 \text{ cm}^3/\text{mol}$ ⁴⁹.

Figure 4b shows the difference in the free energy between 8000 bar and 1 bar, $\Delta_p F(R) = F(R, 8000 \text{ bar})/k_B T - F(R, 1 \text{ bar})/k_B T$, which is identical to the difference in the excess chemical potential profile $\Delta_p \mu_{ex}(R)$ because $F_{vac}(R)$ is independent of pressure. The maximum of $\Delta_p \mu_{ex}(R)$ at a distance of $R = 1.1 \text{ nm}$ indicates that the conformations near the transition state have a larger excess partial molar volume than the other conformations including the native ones. This is because there would exist many narrow spaces that prevent water molecules from accessing the protein surface. It is remarkable that both the polar part of $\Delta_p \mu_{ex}(R)$, $\Delta_p \mu_{pol}(R)$, and the nonpolar part, $\Delta_p \mu_{nonpol}(R)$, are lower by 3–4 at a distance of $R = 3.0 \text{ nm}$. The unfolded state, therefore, is stabilized at high pressure by both the nonpolar part⁵⁰ and electrostatic part of the solvent-induced interaction. The decrease in the polar part, $\Delta_p \mu_{pol}(R)$, with increasing R under high pressure is consistent with the mechanism of high-pressure unfolding of proteins proposed by Chalikian and Macgregor⁵¹.

Concluding Remarks

Kauzmann's hydrophobic interaction hypothesis rests on the following assumption: either there is a precise compensation in the interaction energy between dehydration of polar groups and intramolecular hydrogen-bond formation or the effect of electrostatic interactions on protein stability is negligibly small. For chignolin, which is thought to possess both intramolecular hydrogen bonds and potential hydrophobic bonds in the native state³³, it was shown that neither the nonpolar part nor the polar part of the solvent-induced interaction plays a predominant role in the collapse of chignolin. In fact, the solvent-induced interaction rather stabilizes the unfolded conformations mainly due to the electrostatic part of $\mu_{ex}(R)$. Therefore, the dominant factor in the collapse of chignolin is the intramolecular interactions including van der Waals and electrostatic forces. Furthermore, the conformation of the native state is determined by the subtle balance between competitive factors, i.e., the intramolecular interactions and the solvent-induced interactions. It is remarkable that these results obtained for chignolin are consistent with the conclusions of previous studies on site-directed mutagenesis of globular proteins with interior hydrophobic cores^{13,14}.

It has been demonstrated by calorimetry experiments that thermal unfolding of globular proteins is an endothermic and entropy-driven process, where $T\Delta S_u = \Delta H_u > 0$, if $\Delta G_u = 0$, and where ΔG_u , ΔH_u , and ΔS_u are change in Gibbs energy, enthalpy, and entropy due to unfolding, respectively^{52,53}. In contrast, low-temperature unfolding is, therefore, an exothermic and entropy-reduction process, while such an observation is limited due to freezing of protein solutions. Therefore, widely observed protein folding from high-temperature unfolded states is an exothermic process with a decrease in entropy. The intramolecular-force-driven collapse hypothesis extracted from chignolin is consistent with the exothermic behavior upon protein folding, even though chignolin has no hydrophobic interior cores. This hypothesis can also provide a proper explanation for the exothermic aggregation generally observed for heat-denatured proteins^{53,54}. However, these arguments as well as the results obtained for chignolin do not necessarily guarantee the validity of this hypothesis on the thermodynamic stability of general proteins, thus we need further investigations for several larger proteins to assess the general validity of this conclusion.

The thermodynamic stability mechanism presented on the basis of the competition between the intramolecular interactions and the solvent-induced interactions may underlay the remarkable successes of the protein tube-like model^{55,56} and the Go-like model^{57,58}. Furthermore, the competition between these opposing factors would also provide a new insight into self-assembly of bio/soft materials; especially, the physical origin of their softness in aqueous solutions may be attributed to the solvent-induced interactions because amphiphilic polymers such as poly(N-isopropylacrylamide) that are solid in vacuum due to direct intramolecular and intermolecular interactions become soft and swollen when immersed in water.

In our previous study³², based on an effective energy defined as the sum of intramolecular interaction energy and solvation free energy of each conformation generated by MD simulations in water, we had obtained qualitatively equivalent results for the thermodynamic stability of chignolin, which, though, had significantly overestimated the stability of folded conformations [see the argument at the beginning of the Supplementary Information as well as ref.³²]. Thus, in the present work, we developed the computational method on the direct free energy decomposition for flexible protein molecule and applied it to chignolin again. The direct free energy decomposition demonstrates that we can apply the effective energy analysis to determine the predominant factor in the thermodynamic stability of proteins instead of using the time-consuming direct free energy decomposition. Investigating whether the competition mechanism holds in the thermodynamic stability of proteins with hydrophobic interior cores as well, using these methods or more improved ones with respect to conformation sampling, is our future important project.

Computational Details

Molecular simulations. Isothermal MD simulations⁵⁹ were performed using the Gromacs 5.0.7 suite⁶⁰ with the generalized Born (GB)/surface area (SA) continuum solvent model⁶¹ and the Amber99SB force field⁶². The time step in the MD simulations was 1.0 fs. All intramolecular bonds were constrained using the LINCS algorithm⁶³. Non-bonded interactions were not truncated. In all the GB MD simulations for the coordinate R except for umbrella sampling, the distance between the alpha carbon atoms at the C- and N-terminus was fixed by a constraint.

Umbrella sampling. The free energy profile for the unfolding of chignolin in a continuum solvent described by the GB model, $F_{GB}(R)$, was calculated from a set of umbrella sampling MD simulations using the weighted histogram analysis method (WHAM)⁶⁴. A harmonic potential with a force constant of $1875 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ was applied for the distance between the alpha carbon atoms at the C-terminus and N-terminus. The histogram of the force on these atoms was obtained from a 50-ns simulation for every window. The spacing of the windows along the coordinate R was 0.0125 nm and the number of windows was 208. The standard deviation of $F_{GB}(R)$ was estimated by a bootstrap analysis⁶⁴.

RMDFT calculations. The hydration free energy of the RMDFT model, ΔG_{hyd}^{DFT} , was calculated for 5000 conformations generated by a 50-ns GB MD production run after a 10-ns equilibration at each R . The details of the RMDFT model are given in Appendix C of the Supplementary Information and our previous study³². The standard deviation of $\Delta \mu_{DFT}^{GB}(R)$ was evaluated by decomposing the 5000 conformations into five equal blocks.

Excess chemical potential of the GB model, $\mu_{ex}^{GB}(R)$. Twelve intermediate states, $\varepsilon_r = 40, 20, 10, 5, 4.2, 3.5, 2.9, 2.4, 2.0, 1.7, 1.4$, and 1.2 were considered when calculating $\mu_{ex}^{GB}(R)$. The GB MD simulations of the intermediate states were performed sequentially. The conformation obtained at the end of the simulation with one step higher ε_r value was used as the initial conformation of the simulation at one step lower ε_r value. An equilibration

run was performed for 10 ns, and the data of 5000 conformations generated by the following 50-ns production run were used to calculate $\Delta\mu_{\epsilon_r}^{\epsilon_i}(R)$ for every intermediate state. The standard deviation of $\Delta\mu_{\epsilon_j}^{\epsilon_i}(R)$ was evaluated in the same manner as for $\Delta\mu_{\text{DFT}}^{\text{GB}}(R)$.

Nonpolar part of $\mu_{\text{ex}}(R)$, $\mu_{\text{nonpol}}(R)$. $\Delta\mu_{80}^1(R)$ and $\Delta\mu_1^{80}(R)$ were calculated from 5000 conformations generated by 50-ns production run without all partial charges on chignolin using the dielectric constants $\epsilon_r = 1$ (in vacuum) and $\epsilon_r = 80$ (with only the nonpolar SA term), respectively. $\mu_{\text{nonpol}}(R)$ is given as the sum of $[\Delta\mu_{80}^1(R) - \Delta\mu_1^{80}(R)]/2$ and $\Delta\mu_{\text{DFT}}^{\text{GB}}(R)$, as given by Eq. 7.

References

- Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29**, 7133–7155 (1990).
- Ben-Naim, A. The Rise and Fall of the Hydrophobic Effect in Protein Folding and Protein-Protein Association, and Molecular Recognition. *Open Journal of Biophysics* **1**, 1–7 (2011).
- Pace, C. N., Scholtz, J. M. & Grimsley, G. R. Forces stabilizing proteins. *FEBS Lett* **588**, 2177–2184 (2014).
- Mirsky, A. E. & Pauling, L. On the Structure of Native, Denatured, and Coagulated Proteins. *Proc. Natl. Acad. Sci. USA* **22**, 439–447 (1936).
- Bernal, J. D. Structure of Proteins. *Nature* **143**, 663–667 (1939).
- Pauling, L. & Corey, R. B. Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds: Two New Pleated Sheets. *Proc. Natl. Acad. Sci. USA* **37**, 729–740 (1951).
- Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* **37**, 205–211 (1951).
- Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1–63 (1959).
- Jacobsen, C. F. & Linderström-Lang, K. Salt Linkages in Proteins. *Nature* **164**, 411–412 (1949).
- Chothia, C. Structural invariants in protein folding. *Nature* **254**, 304–308 (1975).
- Guy, H. R. Amino-Acid Side-Chain Partition Energies and Distribution of Residues in Soluble-Proteins. *Biophys. J.* **47**, 61–70 (1985).
- Pace, C. N. *et al.* Contribution of hydrophobic interactions to protein stability. *J. Mol. Biol.* **408**, 514–528 (2011).
- Pace, C. N. Polar Group Burial Contributes More to Protein Stability than Nonpolar Group Burial. *Biochemistry* **40**, 310–313 (2001).
- Eriksson, A. E. *et al.* Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**, 178–183 (1992).
- Bunagan, M. R., Gao, J., Kelly, J. W. & Gai, F. Probing the Folding Transition State Structure of the Villin Headpiece Subdomain via Side Chain and Backbone Mutagenesis. *J. Am. Chem. Soc.* **131**, 7470–7476 (2009).
- Dill, K. A. Theory for the Folding and Stability of Globular-Proteins. *Biochemistry* **24**, 1501–1509 (1985).
- Yasuda, S., Oshima, H. & Kinoshita, M. Structural stability of proteins in aqueous and nonpolar environments. *J Chem Phys* **137**, 135103–135103 (2012).
- Graziano, G. On the molecular origin of cold denaturation of globular proteins. *Phys Chem Chem Phys* **12**, 14245–14252 (2010).
- Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **314**, 141–151 (1999).
- Itoh, S. G. & Okumura, H. Replica-Permutation Method with the Suwa-Todo Algorithm beyond the Replica-Exchange Method. *J. Chem. Theory Comput.* **9**, 570–581 (2013).
- Nakajima, N., Nakamura, H. & Kidera, A. Multicanonical Ensemble Generated by Molecular Dynamics Simulation for Enhanced Conformational Sampling of Peptides. *J Phys Chem B* **101**, 817–824 (1997).
- Paschek, D. & García, A. E. Reversible Temperature and Pressure Denaturation of a Protein Fragment: A Replica Exchange Molecular Dynamics Simulation Study. *Phys. Rev. Lett.* **93**, 238105 (2004).
- Okumura, H. & Okamoto, Y. Multibaric-multithermal ensemble molecular dynamics simulations. *J. Comput. Chem.* **27**, 379–395 (2006).
- Paschek, D., Gnanakaran, S. & García, A. E. Simulations of the pressure and temperature unfolding of an alpha-helical peptide. *Proc. Natl. Acad. Sci. USA* **102**, 6765–6770 (2005).
- Okumura, H. Temperature and pressure denaturation of chignolin: Folding and unfolding simulation by multibaric-multithermal molecular dynamics method. *Proteins* **80**, 2397–2416 (2012).
- Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **334**, 517 (2011).
- Piana, S., Lindorff-Larsen, K. & Shaw, D. E. Atomic-level description of ubiquitin folding. *PNAS* **110**, 5915–5920 (2013).
- Yu, H. A. & Karplus, M. A thermodynamic analysis of solvation. *J Chem Phys* **89**, 2366–2379 (1988).
- Sumi, T., Mitsutake, A. & Maruyama, Y. A solvation-free-energy functional: A reference-modified density functional formulation. *J. Comput. Chem.* **36**, 1359–1369 (2015).
- Sumi, T., Mitsutake, A. & Maruyama, Y. Erratum: 'A solvation-free-energy functional: A reference-modified density functional formulation' [J. Comput. Chem. 2015, 36, 1359–1369]. *J. Comput. Chem.* **36**, 2009–2011 (2015).
- Sumi, T., Maruyama, Y., Mitsutake, A. & Koga, K. A reference-modified density functional theory: An application to solvation free-energy calculations for a Lennard-Jones solution. *The Journal of Chemical Physics* **144**, 224104–224104 (2016).
- Sumi, T., Maruyama, Y., Mitsutake, A., Mochizuki, K. & Koga, K. Application of reference-modified density functional theory: Temperature and pressure dependences of solvation free energy. *J. Comput. Chem.* **39**, 202–217 (2018).
- Honda, S., Yamasaki, K., Sawada, Y. & Morii, H. 10 residue folded peptide designed by segment statistics. *Structure* **12**, 1507–1518 (2004).
- Calimet, N., Schaefer, M. & Simonson, T. Protein molecular dynamics with the generalized Born/ACE solvent model. *Proteins* **45**, 144–158 (2001).
- Khandogin, J. & Brooks, C. L. Toward the accurate first-principles prediction of ionization equilibria in proteins. *Biochemistry* **45**, 9363–9373 (2006).
- Harris, R. C. & Pettitt, B. M. Examining the assumptions underlying continuum-solvent models. *J. Chem. Theory Comput.* **11**, 4593–4600 (2015).
- Satoh, D., Shimizu, K., Nakamura, S. & Terada, T. Folding free-energy landscape of a 10-residue mini-protein, chignolin. *FEBS Lett* **580**, 3422–3426 (2006).
- Harada, R. & Kitao, A. Exploring the folding free energy landscape of a β -hairpin miniprotein, chignolin, using multiscale free energy landscape calculation method. *J Phys Chem B* **115**, 8806–8812 (2011).
- Kührová, P., De Simone, A., Otyepka, M. & Best, R. B. Force-field dependence of chignolin folding and misfolding: comparison with experiment and redesign. *Biophys. J.* **102**, 1897–1906 (2012).
- Mitsutake, A. & Takano, H. Relaxation mode analysis and Markov state relaxation mode analysis for chignolin in aqueous solution near a transition temperature. *The Journal of Chemical Physics* **143**, 124111–124111 (2015).
- Maruyama, Y. & Mitsutake, A. Analysis of Structural Stability of Chignolin. *J Phys Chem B* **122**, 3801–3814 (2018).

42. Kokubo, H., Hu, C. Y. & Pettitt, B. M. Peptide conformational preferences in osmolyte solutions: transfer free energies of decaalanine. *J. Am. Chem. Soc.* **133**, 1849–1858 (2011).
43. Kokubo, H., Harris, R. C., Asthagiri, D. & Pettitt, B. M. Solvation free energies of alanine peptides: the effect of flexibility. *J Phys Chem B* **117**, 16428–16435 (2013).
44. Li, L., Bedrov, D. & Smith, G. D. Repulsive solvent-induced interaction between C60 fullerenes in water. *Phys. Rev. E* **71**, 011520 (2005).
45. Makowski, M., Czaplewski, C., Liwo, A. & Scheraga, H. A. Potential of mean force of association of large hydrophobic particles: toward the nanoscale limit. *J Phys Chem B* **114**, 993–1003 (2010).
46. Kuroda, Y., Suenaga, A., Sato, Y., Kosuda, S. & Taiji, M. All-atom molecular dynamics analysis of multi-peptide systems reproduces peptide solubility in line with experimental observations. *Sci. Rep.* **6**, 19479 (2016).
47. Ben-Naim, A. Inversion of the hydrophobic/hydrophilic paradigm demystifies the protein folding and self-assembly of problems. *International Journal of Physics*, <https://doi.org/10.12691/ijp> (2013).
48. Durell, S. R. & Ben-Naim, A. Hydrophobic-hydrophilic forces in protein folding. *Biopolymers* **107** (2017).
49. Imamura, H. & Kato, M. 3P066 Unfolding of β -hairpin peptides by pressure: FT-IR and FRET studies (Protein: Property, The 48th Annual Meeting of the Biophysical Society of Japan). *Annu Rev Biophys* **50**, S156 (2010).
50. Sumi, T. & Sekino, H. Possible mechanism underlying high-pressure unfolding of proteins: formation of a short-period high-density hydration shell. *Phys Chem Chem Phys* **13**, 15829–15832 (2011).
51. Chalikian, T. V. & Macgregor, R. B. Origins of pressure-induced protein transitions. *J. Mol. Biol.* **394**, 834–842 (2009).
52. Salvetti, G., Tombari, E., Mikheeva, L. & Johari, G. P. The Endothermic Effects during Denaturation of Lysozyme by Temperature Modulated Calorimetry and an Intermediate Reaction Equilibrium. *J Phys Chem B* **106**, 6081–6087 (2002).
53. Schön, A., Clarkson, B. R., Jaime, M. & Freire, E. Temperature stability of proteins: Analysis of irreversible denaturation using isothermal calorimetry. *Proteins* **85**, 2009–2016 (2017).
54. Makhatazde, G. I. & Privalov, P. L. Energetics of Protein Structure. *Advances in Protein Chemistry* **47**, 307–425 (Elsevier, 1995).
55. Maritan, A., Micheletti, C., Trovato, A. & Banavar, J. R. Optimal shapes of compact strings. *Nature* **406**, 287–290 (2000).
56. Hoang, T. X. *et al.* Common attributes of native-state structures of proteins, disordered proteins, and amyloid. *Proc. Natl. Acad. Sci. USA* **103**, 6883–6888 (2006).
57. Go, N. & Taketomi, H. Respective Roles of Short-Range and Long-Range Interactions in Protein Folding. *Proc. Natl. Acad. Sci. USA* **75**, 559–563 (1978).
58. Wu, L., Zhang, J., Qin, M., Liu, F. & Wang, W. Folding of proteins with an all-atom Go-model. *J Chem Phys* **128**, 235103–235103 (2008).
59. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101 (2007).
60. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).
61. Qiu, D., Shenkin, P. S., Hollinger, F. P. & Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *The Journal of Physical Chemistry A* **101**, 3005–3014 (1997).
62. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725 (2006).
63. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS - A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
64. Hub, J. S., de Groot, B. L. & van der Spoel, D. g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **6**, 3713–3720 (2010).

Acknowledgements

This work was supported in part by JSPS KAKENHI Grant No. JP16K05657, No. JP26287099, and No. JP18KK0151. We would like to thank Dr. Yutaka Maruyama for the implementation of the RMDFT model in the graphics processing unit (GPU) version of the 3D-RISM integral equation. We would also like to thank Dr. Takuma Yagasaki for the stimulating discussion and useful suggestions, and Dr. Hiroshi Imamura, Dr. Ryuichi Okamoto, and Dr. Masayuki Irisa for the useful comments.

Author Contributions

T.S. developed the method and performed the calculations. Both the authors, T.S. and K.K. discussed the results and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-41518-1>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019