Biophysics and Physicobiology

Regular Article

Potential energy landscape and thermodynamic transitions of coarse-grained protein models revealed by the multicanonical generalized hybrid Monte Carlo method

Natsuki Mukuta[1] and Shinichi Miura[2]

[1] *Division of Mathematical and Physical Sciences, Graduate School of Natural Science and Technology, Kanazawa University, Kakuma, Kanazawa, Ishikawa 920-1192, Japan*
[2] *Faculty of Mathematics and Physics, Kanazawa University, Kakuma, Kanazawa, Ishikawa 920-1192, Japan*

In the present study, thermodynamic properties of coarse-grained protein models have been studied by an extended ensemble method. Two types of protein model were analyzed; one is categorized into a fast folder and the other into a slow folder. Both models exhibit the following thermodynamic transitions: the collapse transition between random coil states and spatially compact, but non-native states and the folding transition between the collapsed states and the folded native states. Caloric curve for the fast folder shows strong statistical ensemble dependence, while almost no ensemble dependence is found for the slow folder. Microcanonical caloric curve for the fast folder exhibits S-shaped temperature dependence on the internal energy around the collapse transition which is reminiscent of the van der Waals loop observed for the first order transition; at the transition temperature, the collapsed and random coil states coexist dynamically. The corresponding microcanonical heat capacity is found to have negative region around the transition. This kind of exotic behaviors could be utilized to distinguish fast folding proteins.

Introduction

Proteins are heteropolymers consisting of amino acid residues linearly linked by peptide bonds. Protein molecules are known to spontaneously fold into compact three-dimensional structures and to biologically function in the native states. Since the pioneering experiments by Anfinsen and coworkers [1], it has been known that the amino acid sequence uniquely determines the native states. To understand how the information coded in the amino acid sequence is translated into the corresponding three-dimensional structure is the core of the protein folding problem. Understanding the problem would allow us to predict three-dimensional structures in the native states only by the knowledge of the amino acid sequence. Despite many remarkable advances in the last decades [2–7], the

Corresponding author: Shinichi Miura, Faculty of Mathematics and Physics, Kanazawa University, Kakuma, Kanazawa, Ishikawa 920-1192, Japan.
e-mail: smiura@mail.kanazawa-u.ac.jp

◄ *Significance* ►

Coarse-grained protein models studied exhibit two thermodynamic transitions: the collapse transition between random coil and spatially compact, but non-native states and the folding transition between the collapse and native states. Caloric curve for a fast folding model shows strong statistical ensemble dependence especially around the collapse transition, while almost no ensemble dependence is found for a slow folder. Microcanonical caloric curve for the fast folder exhibits S-shaped temperature dependence on the internal energy around the collapse transition; the corresponding microcanonical heat capacity is found to have negative region around the transition. This kind of exotic behaviors could be utilized to distinguish fast folding proteins.

protein folding problem is still far from a solution.

It is important to point out that not all polypeptides are proteins; only a very small subset of all the possible sequences of the 20 naturally occurring amino acids has been selected by evolution. While natural proteins fold into uniquely determined native states quickly, a generic polypeptide does not. Much effort has been made to understand this issue; one important perspective is provided by the consistency principle [2] which states that various energy terms in proteins, for example, short- and long-range interactions, are consistent with each other. This idea has been extended as the principle of minimum frustration and applied to the random energy model to shed light on the phase behavior of proteins [8]. These theoretical efforts have been unified as a perspective assuming that protein has a funnel-shaped energy landscape with a bias toward native structures [3,9].

In the present study, we have adopted a coarse-grained protein model developed by Honeycutt and Thirumalai [10] to investigate thermodynamic behavior associated with protein folding. This HT model has originally been introduced to examine the possibility that metastable states are relevant for protein folding. In this model, conformational frustration prevents efficient relaxation to the global potential energy minimum. To model a fast folding protein, the original model has been modified with the help of the consistency principle [11–14], the resulting coarse-grained model is referred to be a Go-like HT model. The original and Go-like HT models have been studied theoretically [10–18]. Potential energy landscape (PEL) of the HT models has been explored by the eigenvector following and basin-hopping methods [12,14]; while the original HT model is characterized by the frustrated PEL around the global potential energy minimum, the Go-like HT model where the frustration in the original model is removed exhibits a funnel type PEL. Thermodynamic properties have been studied by the histogram method and several enhanced sampling methods [11,13,15–18]. In the present study, we have applied our enhanced sampling method, the multicanonical generalized hybrid Monte Carlo method [19] to the original and Go-like HT models to explore thermodynamic transitions of the protein models. Special attention has been paid to the statistical ensemble dependence of caloric curves to distinguish fast folders from slow folders. Convexity of the microcanonical entropy is demonstrated to play a key role to characterize the fast and slow folders.

Method

We first briefly review the multicanonical ensemble method. The multicanonical ensemble [20–23] is an artificial statistical ensemble whose probability density in the potential energy space $\rho_{mc}(U)$ is given using the density of the potential energy states $\widetilde{\Omega}(U)$ by

$$\rho_{mc}(U) \propto \widetilde{\Omega}(U)e^{-W(U)} = \text{constant} \tag{1}$$

where $W(U)$ is a weight function to realize the constant flat distribution of the potential energy. The obvious choice of the weight function is provided by $W(U) = \ln \widetilde{\Omega}(U)$, however, $\widetilde{\Omega}(U)$ is not known a priori. Thus, it must be determined by other techniques; we have used the replica exchange method [24,25] with the Weighted Histogram Analysis Method (WHAM) [26,27]. In the present study, the multicanonical distribution is generated by the generalized hybrid Monte Carlo that is originally a method to generate the canonical ensemble [28].

The generalized hybrid Monte Carlo (GHMC) method is a method to combine MD and MC methods. The configuration of the system is generated by an equation of motion and the trial configuration is accepted according to the Metropolis criterion as in MC method. Here, we briefly describe our GHMC method to generate the multicanonical ensemble [19]. We first regard the multicanonical ensemble as a fictitious canonical ensemble by introducing the following multicanonical potential: $U_{mc}(U)=k_B T_0 W(U(\{r_i\})$ where T_0 is an arbitrary temperature. Then, the multicanonical distribution can be interpreted as a fictitious canonical distribution at the temperature T_0 by

$$\rho_{mc}(U) \propto \widetilde{\Omega}(U)e^{-U_{mc}(U)/k_B T_0} \tag{2}$$

where k_B is the Boltzmann constant. Here, we introduce the following Hamiltonian H_{mc}:

$$H_{mc} = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + U_{mc}(U) \tag{3}$$

where p_i is the fictitious momentum and m_i is the associated fictitious mass of an ith particle. Using the Hamiltonian H_{mc}, we obtain the following equations of motion [22,23]:

$$\frac{dr_i}{dt} = \frac{\partial H_{mc}}{\partial p_i} = \frac{p_i}{m_i},$$

$$\frac{dp_i}{dt} = -\frac{\partial H_{mc}}{\partial r_i} = -\frac{\partial U_{mc}}{\partial U}\frac{\partial U}{\partial r_i}. \tag{4}$$

Given an initial state, the system state is evolved by numerically integrating the above equations over n_{MD} steps with the time increment Δt using a symplectic integrator [29]. The map from the initial state to the trial state is represented as $U_{\Delta\tau}$: $(\{r_i\},\{p_i\}) \rightarrow (\{r_i'\},\{p_i'\})$ where $\Delta\tau = n_{MD} \times \Delta t$. The trial state is accepted by the following Metropolis criterion:

$$(\{r_i'\}, \{p_i'\}) = \begin{cases} \mathscr{F} \cdot U_{\Delta\tau}(\{r_i\}, \{p_i\}) \text{ with a probability} \\ \qquad\qquad \min\left\{1, e^{-\Delta H_{mc}/k_B T_0}\right\} \\ (\{r_i\}, \{p_i\}) \qquad \text{otherwise} \end{cases}$$
$$\tag{5}$$

where
$$\Delta H_{mc} = H_{mc}(\mathscr{F} \cdot U_{\Delta \tau}(\{r_i\}, \{p_i\})) - H_{mc}((\{r_i\}, \{p_i\}))$$ and $\mathscr{F}$ is an operator that negates the momentum

$$\mathscr{F} : (\{r_i\}, \{p_i\}) \rightarrow (\{r_i\}, \{-p_i\}). \tag{6}$$

The momentum flip is needed to satisfy the detail balance condition in the phase space. It is noted that the Hamiltonian is invariant under the momentum flip. Then, we finally apply the momentum flip and define the next accepted state vector: $\mathscr{F}^t(\{r'_i\}, \{p'_i\})$. The extra momentum flip is included so that the trajectory is reversed on an MC rejection instead of on an acceptance.

In the GHMC method, the momentum is partially refreshed at each GHMC step. We mix the momentum p with a Gaussian noise vector u drawn from the Maxwell distribution at the temperature T_0 that is carried out by the following equation:

$$\begin{pmatrix} p'_i \\ u'_i \end{pmatrix} = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \cdot \begin{pmatrix} p_i \\ u_i \end{pmatrix}, \quad \text{for} \quad i = 1, \cdots, N. \tag{7}$$

Here, u_i is generated by $u_i = (m_i k_B T_0)^{1/2} \xi_i$ where each component of ξ_i is given by the Gaussian random number with zero mean and unit variance. The angle ϕ is introduced in the range of $0 < \phi \leq \pi/2$ that controls the ratio of the momentum mixing. It is worthwhile to note that the standard hybrid Monte Carlo [30,31] is recovered by $\phi = \pi/2$ where the momenta are fully refreshed at each GHMC step.

Here, we summarize the algorithm of the multicanonical GHMC method. We start with an initial state of the system $(\{r_i\}, \{p_i\})$ and mix Gaussian noise vector u_i drawn from the Maxwell distribution at the temperature T_0: $p_i \leftarrow p_i \cos \phi + u_i \sin \phi$ for $i = 1, ..., N$. Then, molecular dynamics calculation based on Eq. (4) is performed for the time increment $\Delta \tau = n_{MD} \times \Delta t$ to generate a trial state that is accepted by the probability $\min\{1, e^{-\Delta H_{mc}/k_B T_0}\}$. If the trial state is rejected, the momenta in the initial state are negated by the operator $\mathscr{F}$:

$$(\{r'_i\}, \{p'_i\}) \leftarrow (\{r_i\}, \{-p_i\}). \tag{8}$$

Models

In the present study, we have adopted a coarse-grained model of a protein molecule developed by Honeycutt and Thirumalai [10]; hereafter, we refer it HT model. The model consists of N beads. Each bead corresponds to a coarse-grained amino acid residue. Three types of beads are introduced in this model; hydrophobic (B), hydrophilic (L), and neutral (N). The primary sequence of the HT model for $N=46$ is as follows: $B_9N_3(LB)_4N_3B_9N_3(LB)_5L$. The potential energy for the HT model consists of the following terms: bond-length, bond-angle, torsion-dihedral, and nonbonded potential terms [32]:

$$U_{HT} = U_{bond} + U_{angle} + U_{torsion} + U_{nonbonded}. \tag{9}$$

The bond-length energy U_{bond} is given by

$$U_{bond} = \sum_{i=1}^{N-1} \frac{k_r}{2}(|r_{i+1} - r_i| - a)^2 \tag{10}$$

where $k_r = 400\varepsilon/a^2$, ε is the average strength of the hydrophobic interaction, and a is the equilibrium bond length; $a = \varepsilon = 1$ in the present study. The bond-angle energy is given by

$$U_{angle} = \sum_{i=1}^{N-2} \frac{k_\theta}{2}(\theta_i - \theta_0)^2 \tag{11}$$

where $k_\theta = 20\varepsilon/\text{rad}^2$, θ_i is the bond angle defined by the successive three beads along the chain i, $i+1$, $i+2$, and $\theta_0 = 1.8326$ rad is the equilibrium bond angle. The torsion dihedral energy is represented as

$$U_{torsion} = \sum_{i=1}^{N-3} [A_i(1 + \cos \phi_i) + B_i(1 + \cos 3\phi_i)] \tag{12}$$

where ϕ_i is the dihedral angle defined by the successive four beads i, $i+1$, $i+2$, $i+3$; $A_i = 0$ and $B_i = 0.2\varepsilon$ if two or more neutral (N) beads are included, otherwise $A_i = B_i = 1.2\varepsilon$. The nonbonded energy is given by

$$U_{nonbonded} = 4\varepsilon \sum_{i=1}^{N-3} \sum_{j-i+3}^{N} C_{ij} \left[\left(\frac{a}{r_{ij}}\right)^{12} - D_{ij} \left(\frac{a}{r_{ij}}\right)^6 \right] \tag{13}$$

where $r_{ij} = |r_i - r_j|$. If both i and j beads are the hydrophobic (B), $C_{ij} = D_{ij} = 1$; $C_{ij} = 2/3$ and $D_{ij} = -1$ for LL and LB pairs. In the case of a pair including at least one neutral (N) bead, $C_{ij} = 1$ and $D_{ij} = 0$. The global minimum energy structure is known to be a β-barrel structure with an energy of $U_0 = -49.2635$ for the HT model; the minimum energy structure is defined to be the native structure of the HT model. The potential energy landscape around the native structure is found to be highly frustrated. Here, we introduce a Go-like HT model guided by the consistency principle [11–14]; the original HT model has been modified by removing the attractive interactions between non-contacting hydrophobic beads in the native structure of the HT model. By this treatment, the native structure is specifically stabilized and the frustration in the original model is expected to be greatly reduced. In the present study, the Go-like model is constructed by setting $D_{ij} = 0$ for the pair whose interparticle distance r_{ij} is longer than $1.3a$ in the native structure of the original model [11]. The resulting minimum potential energy is found to be $U_0 = -41.0228$ that is higher than that of the original model due to switching off the attractive interaction between noncontact pairs in the global minimum structure, while the native structure itself is intact.
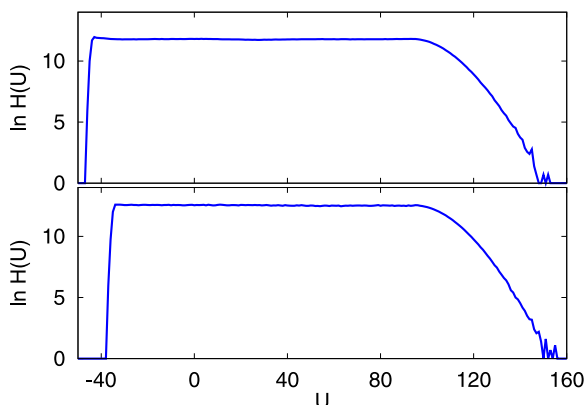
Figure 1 Potential energy distributions generated by the multicanonical GHMC calculations for the HT model (upper panel) and the Go-like HT model (lower panel).

Computational details

In this section, we describe details on the calculations of the original and Go-like HT models. Algorithmic parameters of the multicanonical GHMC method are the following: the fictitious temperature T_0, mixing angle ϕ, MD time step Δt and MD steps n_{MD} in single GHMC step. The fictitious temperature was set to be 0.5. As guided by our previous results [19], we adopted $\phi = \pi/8$ and $n_{MD} = 20$ for both original and Go-like HT models in the present calculations. We set $\Delta t = 0.015$ for the original HT model and $\Delta t = 0.020$ for the Go-like HT model. Total number of the GHMC steps amounted to 2.0×10^7 for the original HT model and 4.0×10^7 for the Go-like HT model. The potential energy distribution was designed to be flat in the energy range [–43, 95] for the HT model and [–34, 95] for the Go-like HT model. The potential energy distributions calculated by our multicanonical GHMC are presented in Figure 1. We find that the distribution in the range specified above is flat enough, demonstrating efficient sampling in configuration space. To perform the multicanonical

calculation, we first numerically determined the multicanonical weight $W(U)$. For the original HT model, we performed the replica exchange HMC calculations. 30 replicas were used to cover the energy range to be flatten; the exchange probabilities were in the range from 40 to 60%. 1.0×10^6 HMC steps were carried out for each replica. Then, in order to obtain $W(U)$, the density of potential energy states $\widetilde{\Omega}(U)$ was evaluated using the replica exchange results with the help of the WHAM method. For the Go-like HT model, the density of potential energy states $\widetilde{\Omega}(U)$ was evaluated similarly.

Results and Discussion

We first show the time series of the potential energy of the protein models along the multicanonical GHMC trajectory. The potential energies are presented in Figure 2. It is found that the broad range of the potential energy is covered by the multicanonical GHMC method, as compared to the standard canonical simulation, see Figure 7. The multicanonical calculations sample various structures from spatially extended random coils to compact structures around the global potential energy minimum. While the potential energy continuously changes in the case of the original HT model, intermittent fluctuation of the potential energy is observed for the Go-like HT model. To reveal the underlying potential energy landscape, we have applied the inherent structure analysis [33–39] to the multicanonical GHMC results. In this analysis, the structures along the trajectory are instantaneously quenched down to the absolute zero; each structure is mapped to the potential energy minimum nearest in the configuration space. In the present study, this mapping was performed by the conjugate gradient method. As mentioned in Sec. 3, the minimum potential energy is given by $U_0 = -49.2635$ for the original HT model and by $U_0 = -41.0228$ for the Go-like HT model. For the original HT model, almost energetically degenerate inherent structures are found around the global minimum,
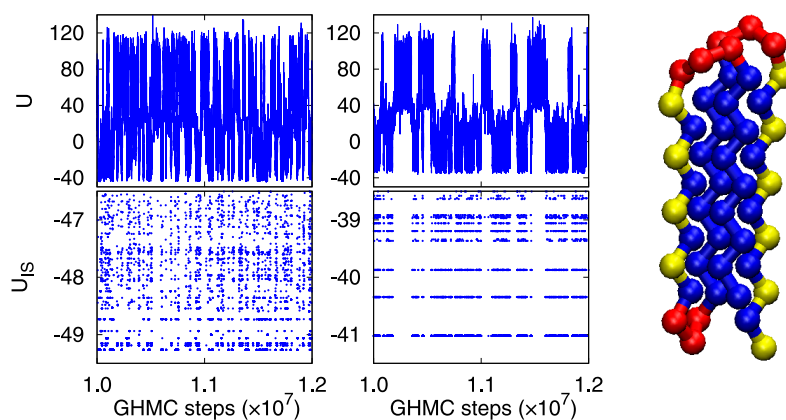
Figure 2 The potential energy U and the potential energy of the associated inherent structure U_{IS} as a function of GHMC steps for the HT model (left panels) and the Go-like HT model (right panels). The snapshot gives the global minimum energy structure. Each bead corresponds to a coarse-grained amino acid residue. Three types of beads are introduced in this model; hydrophobic (blue), hydrophilic (yellow), and neutral (red).

Table 1   Number of inherent structures

HT model			Go-like HT model	
Energy range	Present work	STMD	Energy range	Present work
$U_0 < U \leq -49$	5	5	$U_0 < U \leq -41$	1
$-49 < U \leq -48$	35	35	$-41 < U \leq -40$	1
$-48 < U \leq -47$	159	149	$-40 < U \leq -39$	5
$-47 < U \leq -46$	383	309	$-39 < U \leq -38$	15
$-46 < U \leq -45$	742	547	$-38 < U \leq -37$	31
$-45 < U \leq -44$	1222	760	$-37 < U \leq -36$	80
$-44 < U \leq -43$	1886	918	$-36 < U \leq -35$	457

STMD indicates the statistical temperature molecular dynamics results reported in Ref. [17].

demonstrating frustrated energy landscape. On the other hand, the inherent structures around the global minimum are sparsely distributed for the Go-like HT model; the energy of the global minimum structure is found to be well separated from those of metastable structures. In Table 1, the number of the inherent structures found is listed for both the original and Go-like HT models. For the original HT model, the number is in good agreement with a previous work [17] in low energy ranges; in higher energy ranges, larger number of inherent structures are found in comparison with the previous work. This demonstrates that broader sampling in the configuration space has been realized by our multicanonical GHMC calculation. In the case of the Go-like HT model, the number of the inherent structures is reduced especially in the low energy ranges, corresponding to the sparse potential energy distribution of the inherent structures seen in Figure 2.

Using the configurations distributed according to the multicanonical density, we can evaluate the canonical ensemble average of physical quantities. The canonically averaged value of a physical quantity $A(\boldsymbol{r}_1, ..., \boldsymbol{r}_N)$ is given by

$$\langle A \rangle = \frac{\left\langle A e^{W(U) - \beta U} \right\rangle_{\mathrm{mc}}}{\left\langle e^{W(U) - \beta U} \right\rangle_{\mathrm{mc}}} \tag{14}$$

where $\langle ... \rangle$ indicates the canonical ensemble average while $\langle ... \rangle_{\mathrm{mc}}$ indicates the multicanonical ensemble average. We also analyze the results on the basis of the microcanonical ensemble. The thermodynamic potential of the microcanonical ensemble is given by the entropy $S(E)$ calculated using the density of states $\Omega(E)$ by

$$S(E) = k_B \ln \Omega(E) \tag{15}$$

where E is the internal energy. In the present multicanonical calculations, the density of potential energy states $\widetilde{\Omega}(U)$ can straightforwardly be evaluated. The density of states $\Omega(E)$ is linked with the density of the potential energy states $\widetilde{\Omega}(U)$ by the following relation [40]:

$$\Omega(E) = C_N \int_{U_0}^{E} (E - U)^{\frac{3N}{2} - 1} \widetilde{\Omega}(U) dU \tag{16}$$

where $C_N = (2m\pi)^{3N/2} Z / \Gamma(3N/2)$; here Z is the partition function and Γ is the gamma function. Using the entropy, for example, the temperature $T(E)$ is obtained by

$$\frac{1}{T(E)} = \frac{\partial S(E)}{\partial E}. \tag{17}$$

We present the caloric curves for the original and Go-like HT models in Figure 3. The internal energy E by the canonical ensemble is given as follows:

$$E = \frac{3}{2} N k_B T + \langle U \rangle \tag{18}$$

The associated heat capacity is calculated using the variance of the potential energy

$$C = \frac{3}{2} N k_B + \frac{\langle U^2 \rangle - \langle U \rangle^2}{k_B T^2} \tag{19}$$

For both models, the internal energy decreases on cooling. For the Go-like HT model, steep decrease is found around $T = 0.61$. The corresponding heat capacity has a clear peak around the temperature. The peak signals the transition from a random coil state at high temperature to a spatially compact collapsed, but non-native state at low temperature, which has been verified by examining the order parameter introduced below. Regarding the spatial extent, we find steep decrease of the radius of gyration R_g around the transition on cooling, as shown in Figure 3. According to the previous studies [16,41–43], we call it the collapse transition. For the original HT model, similar but much milder temperature dependence is found for the caloric curve and the heat capacity; a peak in C also signals the collapse transition. In Figure 3, the microcanonical caloric curve, E vs $T = (\partial S/\partial E)^{-1}$ is also presented. For the Go-like HT model, remarkable ensemble dependence is observed; the caloric curve is found to exhibit S-shaped dependence around the collapse transition temperature T_θ, which is reminiscent of the van der Waals loops regarding the first order transition
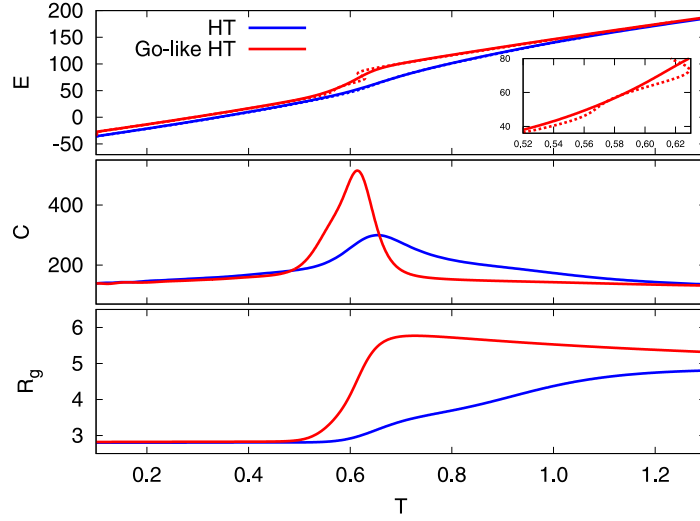
Figure 3 The internal energy E (top panel), the heat capacity C (middle panel) and the radius of gyration R_g (bottom panel) as a function of temperature T. Blue curves for the original HT model and red curves for the Go-like HT model; solid lines indicate the canonical ensemble results and dashed lines in the upper panel indicate the microcanonical ensemble results. The internal energy for the Go-like HT model around $T = 0.57$ is given in an inset.

[44]. This point will further be discussed later. In the microcanonical caloric curve, we find another inflection point around $T = 0.57$ corresponding to the folding transition temperature characterized by the order parameter below. This observation suggests that the microcanonical caloric curve contains richer information compared with the canonical counterpart. For the original HT model, almost no ensemble dependence is found for the caloric curve. This type of strong ensemble dependence of the caloric curve has been observed for a fast folding protein model characterized by two states, folded and random coil states [45]. In Ref. [45], the model has been demonstrated to show a first-order type transition; the above statistical ensemble dependence is known to be observed for finite systems exhibiting first-order type transitions. In the present study, the HT models are characterized by three states, folded, collapsed and random coil states.

To quantitatively analyze the folding transition to the native state, the following order parameter is introduced [16,42]:

$$Q = \frac{1}{M} \sum_{i,\, j > i+4}^{N} \theta\big(\epsilon - |r_{ij} - r_{i,j}^0|\big), \qquad (20)$$

where r_{ij}^0 is the distance between beads i and j in the native state, $\epsilon = 0.2$ in the present study, M is the normalization constant and $\theta(x)$ is the Heaviside step function. This order parameter is designed to give unity for the native structure and small values for other structures. We also examine the order parameter fluctuation χ_{QQ} defined by

$$\chi_{QQ} = \langle Q^2 \rangle - \langle Q \rangle^2. \qquad (21)$$
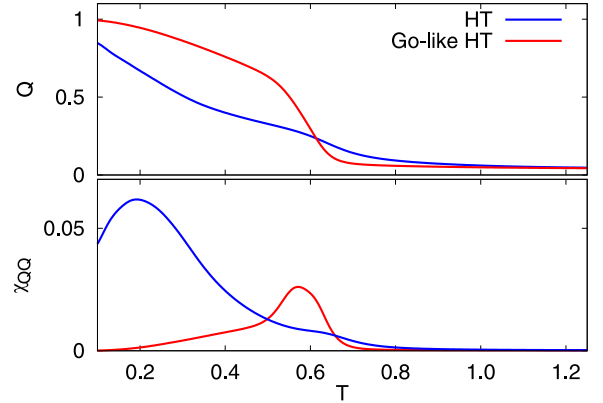
Figure 4 The order parameter Q (upper panel) and the order parameter fluctuation χ_{QQ} (lower panel) as a function of temperature T; blue curves for the original HT model and red curves for the Go-like HT model.

The calculated results are presented in Figure 4. At high temperature, the averaged order parameter is vanishingly small for the both models, since the proteins are in random coil states. For the Go-like HT model, the order parameter rapidly increases around $T = 0.6$ with lowering the temperature; the associated fluctuation χ_{QQ} has a peak at $T_f = 0.57$ that indicates the folding transition from the collapsed state to the native state. For the original HT model, on the other hand, the averaged order parameter gradually increases; a peak in the fluctuation function is found at much lower temperature $T_f = 0.19$ in comparison with the Go-like HT model. The transition temperatures are collected in Table 2; the transition temperatures are in good agreement with previously reported values [11]. Using the transition

Table 2   The folding transition temperature T_f and the collapse transition temperature T_θ for the HT and Go-like HT models

	T_f	T_θ	σ
HT model	0.19 (0.19)	0.65 (0.66)	0.71 (0.71)
Go-like HT model	0.57 (0.59)	0.61 (0.61)	0.07 (0.03)

The parameter σ is defined by $\sigma=(T_\theta-T_f)/T_\theta$. Values in parentheses are given in Ref. [11].

temperatures, the following parameter σ can be introduced and is known to be correlated with the folding time of proteins [43]:

$$\sigma = \frac{T_\theta - T_f}{T_\theta} \tag{22}$$

where T_θ is the collapse transition temperature. Protein molecules for $\sigma<0.1$ are empirically known to show fast folding and those for $\sigma>0.6$ slow folding. As seen in the Table, the Go-like HT model falls in the range of the fast folder and the original model in the slow folder.

To further gain the insight on the nature of the transitions, we present the entropy $S(E)$ as a function of the internal energy E around the collapse transition in Figure 5. While the entropy is found to be a concave function of E for the original HT model, the entropy has a convex E dependence in some energy region for the Go-like HT model. In the convex region, we can draw a tangent common to two energy values; the slope is found to be indeed the inverse T_θ. The internal energy distribution at the temperature T_θ is also presented in Figure 5. The internal energy indicates bimodal distribution for the Go-like HT model. The lower energy peak corresponds to the collapsed state and the higher energy peak the random coil state; the

peak energy difference gives the latent heat as seen in the first order transition. Thus, the collapse transition for the Go-like HT model could be categorized into the first order type transition. For the original HT model, the energy distribution is unimodal, suggesting the collapse transition could be categorized into a continuous type transition. As shown in Figure 6, the energy distribution for the both models at the folding transition temperature is found to be unimodal; again, the folding transition for the both models could be a continuous type. In Figure 7, the instantaneous potential energy and the order parameter are plotted along the canonical GHMC trajectory controlled at the temperature T_θ. For the Go-like HT model, low and high potential energy states are clearly shown to coexist dynamically; the former is the collapsed state and the latter the random coil state. Typical configurations of the collapsed and the random coil states are presented in Figure 7. It is worthwhile to mention that the collapsed state has relatively large value of the order parameter Q. This is due to the fact that the collapsed state for the Go-like HT model has a barrel region consisting of three β-strands common to the native structure. To further clarify this point, we introduce a partial order parameter Q_p that is defined by Eq. (20) only using the above β-barrel region. This parameter is normalized to be unity if the region is similar enough with the β-barrel found in the native state. In Figure 8, the calculated Q_p is presented. The large Q_p values are observed for the collapsed state. The trend is clearer for the partial order parameters calculated using the inherent structures which have larger values close to unity. This demonstrates that the collapsed state could be regarded to be a partially folded state. On the other hand, the collapsed state for the original HT model has no such partial structural similarity with the native state.

In the microcanonical ensemble, it might seem that a
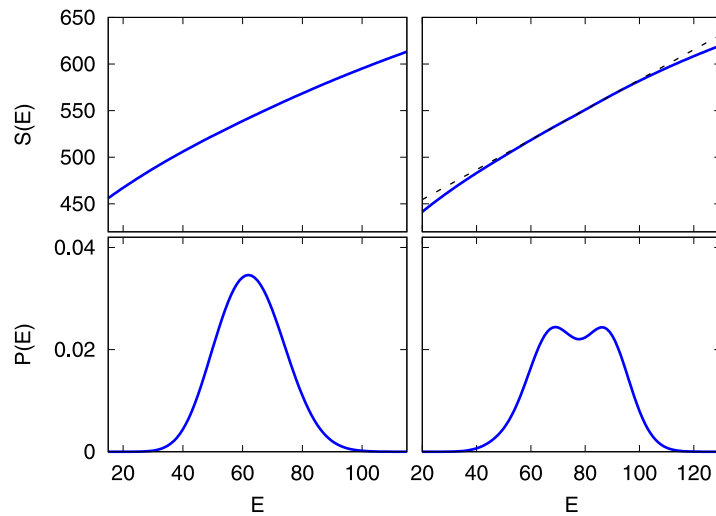
Figure 5   Microcanonical entropy $S(E)$ and the potential energy distribution $P(E)$ at the collapse transition temperature T_θ for the HT model (left panels) and the Go-like HT model (right panels). Dashed line in the right upper panel indicates T_θ.
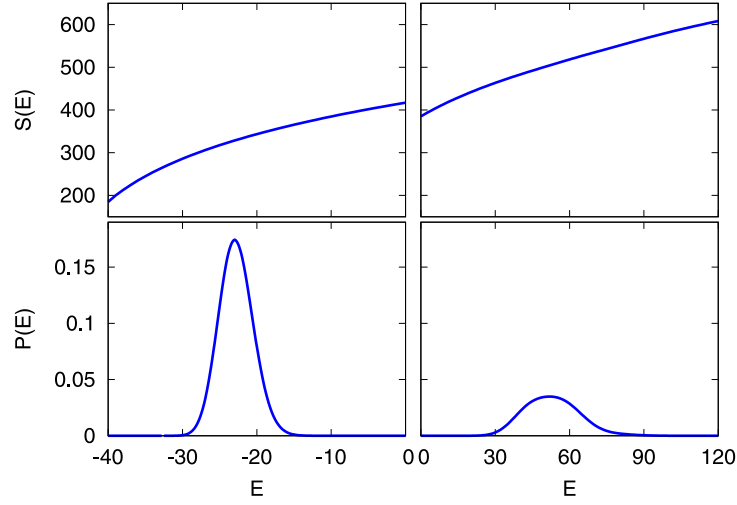
Figure 6   Microcanonical entropy $S(E)$ and the potential energy distribution $P(E)$ at the folding transition temperature T_f for the HT model (left panels) and the Go-like HT model (right panels).
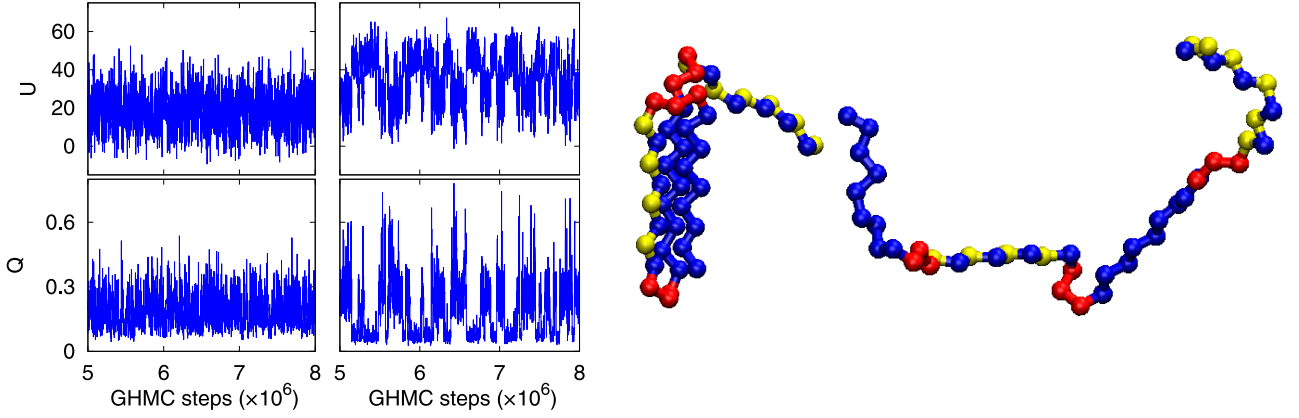
Figure 7   The potential energy U and the order parameter Q at the collapse transition temperature T_θ as a function of the canonical GHMC steps for the HT model (left panels) and the Go-like HT model (right panels). The left and right figures are typical configurations of the collapsed and the random coil states for the Go-like HT model, respectively.
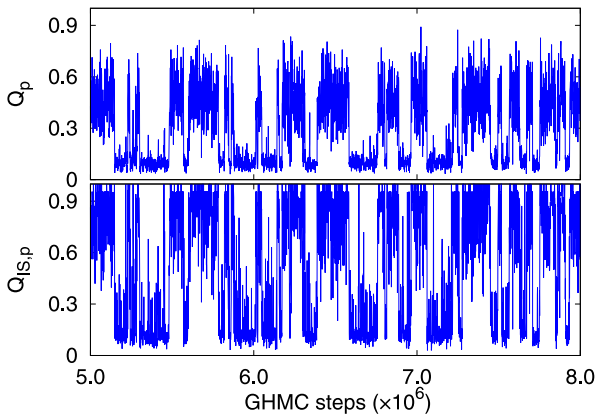

Figure 8   The partial order parameter Q_p (upper panel) and the corresponding partial order parameter calculated by the inherent structure $Q_{IS,p}$ (lower panel) for the Go-like HT model at the collapse transition temperature T_θ as a function of the canonical GHMC steps.

stable internal energy intermediate between the collapsed and random coil states could be given by phase separation; however, the system is too small to phase-separate. The van der Waals loop type behavior observed for the caloric curve can be understood by the kinetic energy distribution at fixed internal energy E. The kinetic energy distribution $P(K; E)$ for a given internal energy E can be written by [46]

$$P(K; E) = \widetilde{\Omega}(E - K)K^{(s-1)/2} \qquad (23)$$

where K is the kinetic energy and $s=3N–6$. The calculated result is presented in Figure 9. For the Go-like HT model, in the vicinity of the collapse transition, the kinetic energy distribution broadens toward lower energy with increasing the internal energy. This arises from the system climbing up the potential energy surface, accompanying the β-barrel deformation in the collapsed state. The resulting averaged
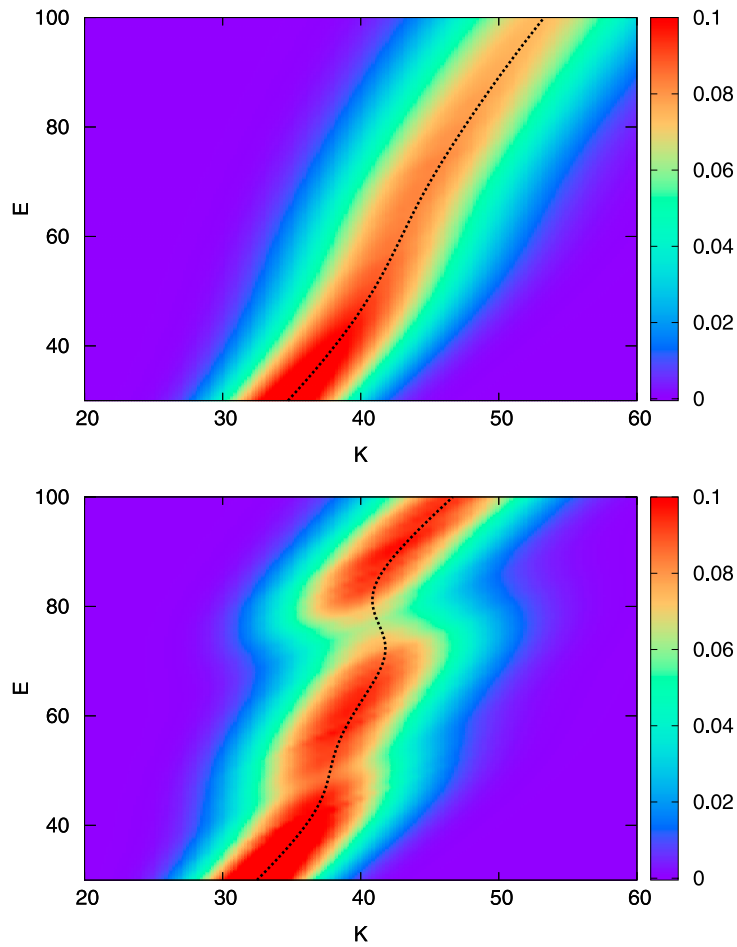
Figure 9   The kinetic energy distribution as a function of the internal energy *E* for the HT model (upper panel) and the Go-like HT model (lower panel). Dashed lines indicate the averaged kinetic energy as a function of the internal energy.

kinetic energy shows the S-shaped dependence as seen in the microcanonical caloric curve. For the original model, monotonic dependence of the averaged kinetic energy is observed, although the kinetic energy distribution broadens at the collapse transition. The S-shaped dependence seen for the Go-like HT model yields the exotic energy dependence of the heat capacity. The microcanonical heat capacity is evaluated by

$$C(E) = \frac{\partial E}{\partial T} = -\frac{1}{T(E)^2}\left(\frac{\partial^2 S(E)}{\partial E^2}\right)^{-1} \qquad (24)$$

The calculated heat capacity is presented in Figure 10. For the Go-like HT model, the negative heat capacity is found around the collapse transition, reflecting the S-shaped loop in the microcanonical caloric curve. Indeed, this is due to the fact that an increase in the internal energy causes a temperature reduction. A peak is found around the folding transition arising from the inflection point in the micro-canonical caloric curve. For the original HT model, a broad peak is found at the collapse transition.
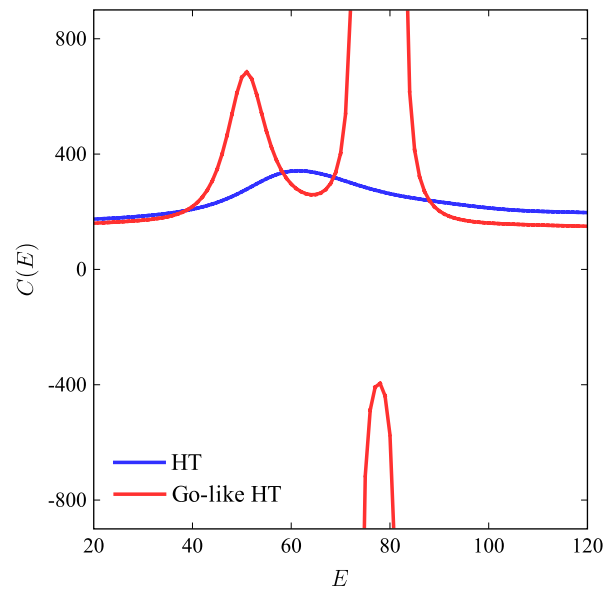
Figure 10   The microcanonical heat capacity as a function of the internal energy *E* for the HT (blue curve) and the Go-like HT (red curve) models.

Conclusion

In the present study, we have performed the multicanonical generalized hybrid Monte Carlo (GHMC) calculations for coarse-grained protein models. One is a model proposed by Honeycutt and Thirumalai (HT) and the other is a modified version called a Go-like HT model that the global minimum structure is specifically stabilized. While the original HT model is characterized by highly frustrated energy landscape near the global potential energy minimum, the Go-like HT model has a deep global potential energy minimum of the native state. The protein models exhibit two types of thermodynamic transitions; the collapse transition between random coil states and spatially compact, but non-native states and the folding transition between the collapsed states and the folded native states. The collapsed state for the Go-like HT model actually is found to be a partially folded state, while the collapsed state for the original HT model has no structural similarity with the native folded state. An empirical parameter regarding foldability defined using the transition temperatures indicates that the original HT model is a slow folder and the Go-like HT model is a fast folder. Caloric curve for the Go-like HT model is found to show strong statistical ensemble dependence, while that for the original model does not. Microcanonical caloric curve for the Go-like HT model shows S-shaped temperature dependence of the internal energy around the collapse transition that is reminiscent of the van der Waals loop characteristic of the first order transition; at the transition temperature, the collapsed and random coil states coexist dynamically. The associated microcanonical heat capacity has negative region around the transition. On the other hand, the original HT model does not show such exotic behaviors. This ensemble dependence and the associated negative heat capacity could be utilized to distinguish fast folding proteins. An important future direction of the present study includes extension of our treatment to more realistic all atom models to explore the energy landscape of hydrated proteins; such a study with the help of integral equation theories of liquids [47,48] is in progress in our research group.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Author Contributions

N. M. performed all the simulations and analyses. N. M and S. M. directed the entire project and co-wrote the manuscript.

References

[1] Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **181**, 223–230 (1973).

[2] Go, N. Theoretical Studies of Protein Folding. *Annu. Rev. Biophys. Bioeng.* **12**, 183–210 (1983).

[3] Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).

[4] Pande, V. S., Grosberg, A. Y. & Tanaka, T. Heteropolymer freezing and design: Towards physical models of protein folding. *Rev. Mod. Phys.* **72**, 259–314 (2000).

[5] Shakhnovich, E. Protein Folding Thermodynamics and Dynamics: Where Physics, Chemistry, and Biology Meet. *Chem. Rev.* **106**, 1559–1588 (2006).

[6] Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. The Protein Folding Problem. *Annu. Rev. Biophys.* **37**, 289–316 (2008).

[7] Dill, K. A. & MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **338**, 1042–1046 (2012).

[8] Bryngelson, J. D. & Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc. Nati. Acad. Sci. USA* **84**, 7524–7528 (1987).

[9] Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* **21**, 167–195 (1995).

[10] Honeycutt, J. D. & Thirumalai, D. Metastability of the folded states of globular proteins. *Proc. Natl. Acad. Sci. USA* **87**, 3526–3529 (1990).

[11] Kim, J. & Keyes, T. Influence of go-like interactions on global shapes of energy landscapes in β-barrel forming model proteins: Inherent structure analysis and statistical temperature molecular dynamics simulation. *J. Phys. Chem. B* **112**, 954–966 (2008).

[12] Miller, M. A. & Wales, D. J. Energy landscape of a model protein. *J. Chem. Phys.* **111**, 6610–6616 (1999).

[13] Kim, J., Keyes, T. & Straub, J. E. Relationship between protein folding thermodynamics and the energy landscape. *Phys. Rev. E* **79**, 030902 (2009).

[14] Oakley, M. T., Wales, D. J. & Johnston, R. L. Energy landscape and global optimization for a frustrated model protein. *J. Phys. Chem. B* **115**, 11525–11529 (2011).

[15] Nymeyer, H., Garcia, A. E. & Onuchic, J. N. Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl. Acad. Sci. USA* **95**, 5921–5928 (1998).

[16] Guo, Z. & Brooks, C. L. Thermodynamics of protein folding: a statistical mechanical study of a small all-β protein. *Biopolymers* **42**, 745–757 (1997).

[17] Kim, J., Straub, J. E. & Keyes, T. Statistical temperature molecular dynamics: Application to coarse-grained β-barrel-forming protein models. *J. Chem. Phys.* **126**, 135101 (2007).

[18] Kim, J. & Keyes, T. Inherent structure analysis of protein folding. *J. Phys. Chem. B* **111**, 2647–2657 (2007).

[19] Mukuta, N. & Miura, S. Development of a generalized hybrid Monte Carlo algorithm to generate the multicanonical ensemble with applications to molecular systems. *J. Chem. Phys.* **149**, 072322 (2018).

[20] Berg, B. A. & Neuhaus, T. Multicanonical algorithms for first order phase transitions. *Phys. Lett. B* **267**, 249–253 (1991).

[21] Berg, B. A. & Neuhaus, T. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys. Rev. Lett.* **68**, 9–12 (1992).

[22] Hansmann, U. H. E., Okamoto, Y. & Eisenmenger, F. Molecular dynamics, langevin and hybrid Monte Carlo simulations in a multicanonical ensemble. *Chem. Phys. Lett.* **259**, 321–330 (1996).

[23] 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).

[24] Hukushima, K. & Nemoto, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Physical Soc. Japan* **65**, 1604–1608 (1996).

[25] Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).

[26] Ferrenberg, A. M. & Swendsen, R. H. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* **63**, 1195–1198 (1989).

[27] Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).

[28] Horowitz, A. M. A generalized guided Monte Carlo algorithm. *Phys. Lett. B* **268**, 247–252 (1991).

[29] Hairer, E., Lubich, C. & Wanner, G. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations* (Springer, 2002).

[30] Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **195**, 216–222 (1987).

[31] Mehlig, B., Heermann, D. W. & Forrest, B. M. Hybrid Monte Carlo method for condensed-matter systems. *Phys. Rev. B* **45**, 679–685 (1992).

[32] Lee, Y. H. & Berne, B. J. Global Optimization: Quantum Thermal Annealing with Path Integral Monte Carlo. *J. Phys. Chem. A* **104**, 86–95 (2000).

[33] Stillinger, F. H. & Weber, T. A. Hidden structure in liquids. *Phys. Rev. A* **25**, 978–989 (1982).

[34] Noguti, T. & Gō, N. Structural basis of hierarchical multiple substates of a protein. I: Introduction. *Proteins* **5**, 97–103 (1989).

[35] Noguti, T. & Gō, N. Structural basis of hierarchical multiple substates of a protein. II: Monte carlo simulation of native thermal fluctuations and energy minimization. *Proteins* **5**, 104–112 (1989).

[36] Noguti, T. & Gō, N. Structural basis of hierarchical multiple substates of a protein. III: Side chain and main chain local conformations. *Proteins* **5**, 113–124 (1989).

[37] Noguti, T. & Gō, N. Structural basis of hierarchical multiple substates of a protein. IV: Rearrangements in atom packing and local deformations. *Proteins* **5**, 125–131 (1989).

[38] Noguti, T. & Gō, N. Structural basis of hierarchical multiple substates of a protein. V: Nonlocal deformations. *Proteins* **5**, 132–138 (1989).

[39] Go, N. & Noguti, T. Structural basis of hierarchical multiple substates of a protein. *Chem. Scr.* **29A**, 151–164 (1989).

[40] Shimizu, H. Estimation of the density of states by multi-canonical molecular dynamics simulation. *Phys. Rev. E* **70**, 056704 (2004).

[41] Guo, Z. & Thirumalai, D. Kinetics of protein folding: Nucleation mechanism, time scales, and pathways. *Biopolymers* **36**, 83–102 (1995).

[42] Camacho, C. J. & Thirumalai, D. Kinetics and thermo-dynamics of folding in model proteins. *Proc. Natl. Acad. Sci. USA* **90**, 6369–6372 (1993).

[43] Klimov, D. K. & Thirumalai, D. Criterion that determines the foldability of proteins. *Phys. Rev. Lett.* **76**, 4070–4073 (1996).

[44] Reichl, L. E. *A Modern Course in Statistical Physics* (Wiley, 2009).

[45] Hernández-Rojas, J. & Gomez Llorente, J. M. Micro-canonical versus canonical analysis of protein folding. *Phys. Rev. Lett.* **100**, 258104 (2008).

[46] Labastie, P. & Whetten, R. L. Statistical thermodynamics of the cluster solid-liquid transition. *Phys. Rev. Lett.* **65**, 1567–1570 (1990).

[47] Hansen, J. P. & McDonald, I. R. *Theory of Simple Liquids: with Applications to Soft Matter* (Academic Press, 2013).

[48] Gray, C. G. & Gubbins, K. E. *Theory of Molecular Fluids: I: Fundamentals* (OUP Oxford, 1984).