Effect of hydrogen bond interaction on protein phase transition

Hon-Wai Leong[1,2] and Lock Yue Chew[2]

[1]*High Performance Computing Centre, Nanyang Technological University, Block NS4-04-25, 50 Nanyang Avenue, 639798 Singapore*

[2]*Division of Physics & Applied Physics, School of Physical & Mathematical Sciences, Nanyang Technological University, SPMS-04-01, 21 Nanyang Link, 637371 Singapore*

(Received 6 November 2011; revised manuscript received 16 July 2012; published 4 September 2012)

We derive the grand partition function of protein chain by restricting dihedral angles to exist only in five distinct states and assume that the dominant noncovalent potential is the hydrogen bond interaction. We investigate the phase transition of protein secondary structures and the order of the transition through analyzing its heat capacity. Our theory demonstrates the presence of α-β-coil structural phase transition in the protein polyalanine.

DOI: 10.1103/PhysRevE.86.031902 PACS number(s): 87.15.bd, 87.15.Zg, 87.14.E−, 87.15.Cc

I. INTRODUCTION

Various aspects of protein have been studied since many decades ago. This includes analysis of biological functions, computation of normal modes, study of protein-protein interaction, and most importantly the folding mechanism of protein from a given sequence of amino acid residues. Protein is known to consist of a chain of amino acid residues. Given a sequence of amino acids, which is known as the primary structure, the chain typically folds into one unique structure [1]. However, protein misfolding has been observed in some cases which causes diseases such as the prion disease (also known as the mad cow disease) [2]. Protein misfolding has also been observed in the β amyloid whose aggregation into neurotoxic oligomers is believed to be the source of Alzheimer's disease. A detailed study of protein misfolding in terms of all atomic models is currently limited by the available computational resources. At the same time, models such as the Zimm-Bragg [3] and Lifson-Roig [4] may be too simplistic to capture certain details of the protein phase transition that arises during protein misfolding. Recently, more advanced models, such as Yakubovich *et al.* [5–7], Ding *et al.* [8], Hong-Lei *et al.* [9], Yasar-Demir [10], Gibbs-DiMarzio [11], Leong *et al.* [12], etc., have been developed that are able to predict additional features of protein phase transition. In this paper, we shall further develop our model in Ref. [12], which is based on the Hamiltonian formulation. Our aim is to include additional structural and configurational details into the model and see how it will affect protein secondary structure phase transition as a function of temperature.

The organization of our paper is as follows. We first introduce our model and derive the canonical formalism of the protein chain in Sec. II, where we describe the use of the dihedral angles (ϕ and ψ) as generalized coordinates instead of the Cartesian coordinates of the protein molecules. Based on the Ramachandran plot [13], we consider dihedral angles that correspond to five distinct protein states. A combination of these sets of angles can lead to the following protein structures: α helix, β sheet, β turns, and coil. In our model, we assume that the hydrogen bond is the dominant noncovalent interactive potential that stabilizes the formation of these protein structures. This is a simplifying assumption as we ignore many energy terms such as the stretching, bending, torsion, as well as the van der Waals and electrostatic interactions which are employed in more sophisticated models. The advantage of this simplification is that it allows us to derive the grand partition function of a protein chain in an analytical way, which is shown in Sec. III of this paper. Although our model is crude, it has allowed us to gain insights into the statistical physics of protein phase transition, which will be covered in Secs. IV and V. Finally, Sec. VI concludes our paper.

II. CANONICAL FORMALISM OF PROTEIN CHAIN

In our mathematical model, we define the independent unit of a protein chain as "crank,' which connects the center of one residue to the next. Each of these cranks consists of C_α, C, N, O, H, and S (hard-sphere side chain) molecules (Fig. 1). We assume the crank to be perfectly rigid, and the only degrees of freedom among the cranks are given by pairs of dihedral angles (ϕ and ψ). Such a description has been used in the CSAW (conditioned self-avolding walk) model of protein folding [14,15]. For a given n number of cranks, we denote the position vector of each atom by R_j, with $j = 1,\ldots,6n$. Furthermore, we let the dihedral angles $\phi_i = q_{i1}$ and $\psi_i = q_{i2}$. This allows us to define the notation q_α, where $\alpha = \{i,k\}$, with $i = 1,\ldots,n-1$ and $k = 1,2$. We are to treat R_j as a function of q_α.

Let p and q be the canonical position and momentum. The Hamiltonian of our protein chain can be written as

$$H(p,q) = K(p,q) + U(q) + U^e, \tag{1}$$

where $K(p,q)$ is the kinetic energy of the molecules, $U(q)$ is the effective potential of the hydrogen bonds, and U^e is the chemical potential of the hydrogen bonds in the protein chain.

According to Ref. [12], the total kinetic energy is given as follows:

$$K(q,\dot{q}) = \frac{1}{2}\sum_{j=1}^{6n} m_j \dot{\mathbf{R}}_j^2 = \frac{1}{2}\sum_{\alpha,\beta}\dot{q}_\alpha\left(\sum_{j=1}^{6n} m_j \frac{\partial \mathbf{R}_j}{\partial q_\alpha}\cdot\frac{\partial \mathbf{R}_j}{\partial q_\beta}\right)\dot{q}_\beta = \frac{1}{2}\dot{q}^T M \dot{q}, \tag{2}$$

where the mass matrix is given by

$$M_{\alpha\beta} = \sum_{j=1}^{6n} m_j \frac{\partial \mathbf{R}_j}{\partial q_\alpha}\cdot\frac{\partial \mathbf{R}_j}{\partial q_\beta}. \tag{3}$$

Note that M is a symmetric matrix, i.e., $M^T = M$.

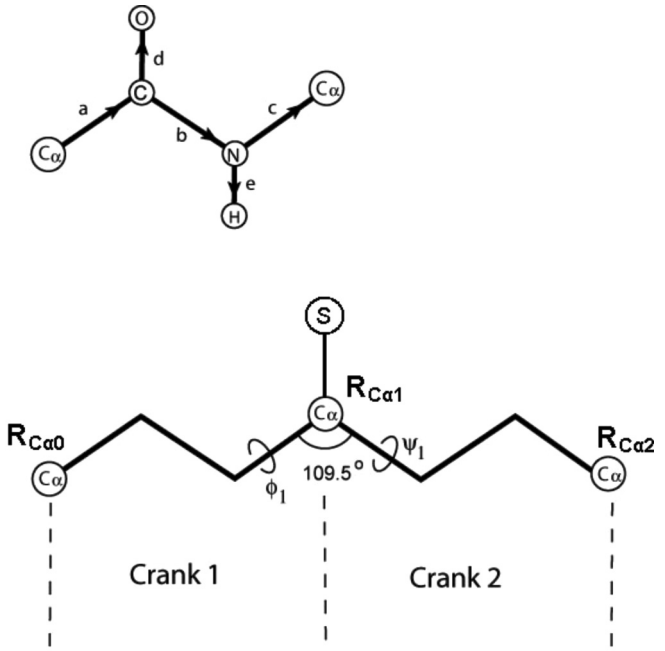

FIG. 1. Upper panel illustrates one crank unit. Lower panel shows the manner in which the cranks are connected to form the backbone of the protein. "S" denotes the side chain molecule. It attaches to the C_α atom which serves as the point of connection between the two cranks.

In our model, we assume that the hydrogen bond is the dominant effective potential that stabilizes the secondary structure within a protein chain. In the protein chain, hydrogen bonds are formed between the N-H and C=O groups of different residues [16,17]. We assume that a hydrogen bond is formed when the distance between the H and O atoms is 2.0 ± 1.0 Å, and the bond angle between the N-H and C=O is $180 \pm 45^\circ$ [16].

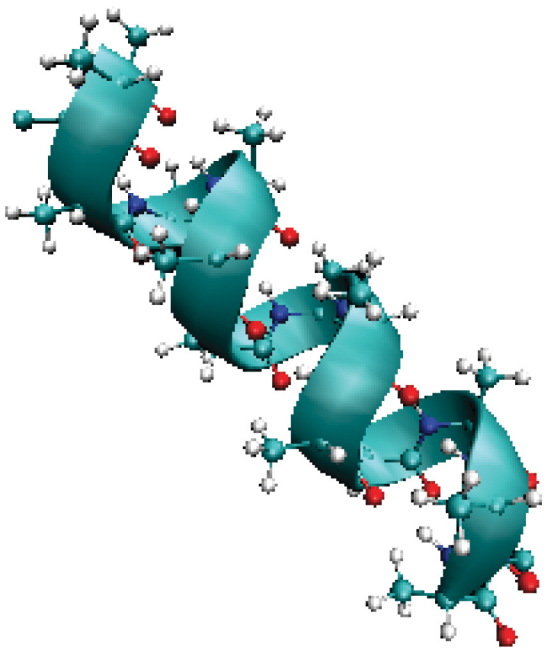

FIG. 2. (Color online) A seven crank polyalanine in α helix conformation.

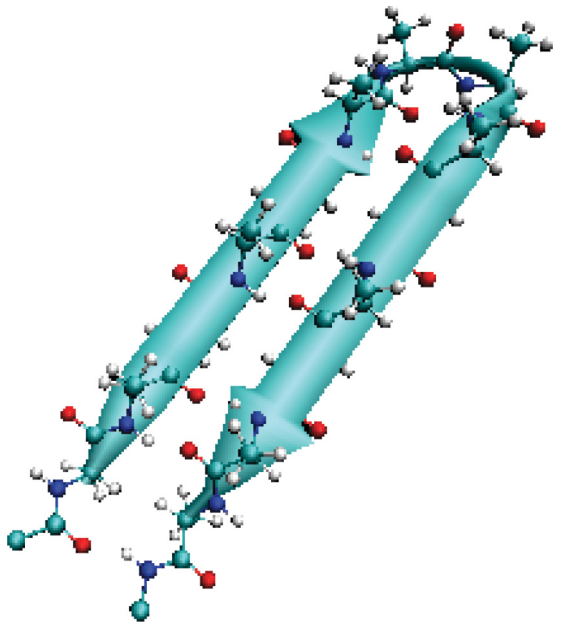

FIG. 3. (Color online) A seven crank polyalanine in β sheet conformation.

Let $\mathbf{b}_k$ be the kth hydrogen bond vector between the O atom and the H atom in the equilibrium situation. On the other hand, $\mathbf{b}'_k$ is the same vector when the configuration is displaced from equilibrium. The effective potential of the hydrogen bonds can be modeled as follows:

$$U = \frac{1}{2} q^T (\kappa_1 D + \kappa_2 C) q, \tag{4}$$

$$D_{\alpha\beta} = \sum_k \left| \hat{\mathbf{b}}_k \cdot \frac{\partial \mathbf{b}'_k}{\partial q_\alpha} \right|_0 \cdot \left| \hat{\mathbf{b}}_k \cdot \frac{\partial \mathbf{b}'_k}{\partial q_\beta} \right|_0, \tag{5}$$

$$C_{\alpha\beta} = \sum_k \left| \hat{\mathbf{b}}_k \times \frac{\partial \mathbf{b}'_k}{\partial q_\alpha} \right|_0 \cdot \left| \hat{\mathbf{b}}_k \times \frac{\partial \mathbf{b}'_k}{\partial q_\beta} \right|_0, \tag{6}$$

where $\hat{\mathbf{b}}_k = \mathbf{b}_k / |\mathbf{b}_k|$ is the unit vector of hydrogen bond at equilibrium. The subscript 0 indicates that the evaluation is to be performed at the equilibrium configuration. κ_1 and κ_2 are the force constants associated with the stretching and bending of the hydrogen bonds. We have let $\kappa_1 = 13$ N/m and $\kappa_2 = 3$ N/m [12].

The formation of a hydrogen bond carries a "chemical" potential. Furthermore, the stability of each secondary structure is determined by its configuration of hydrogen bonds. We assume that the chemical potential of hydrogen bond associated with the α helix and β sheet has different statistical weight. We

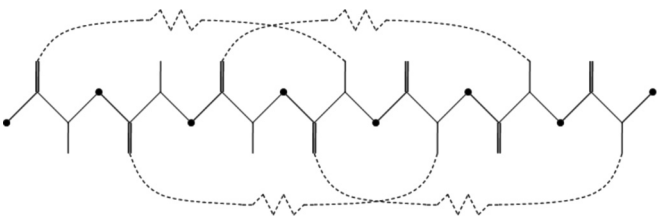

FIG. 4. The configuration of hydrogen bonds in the α helix.

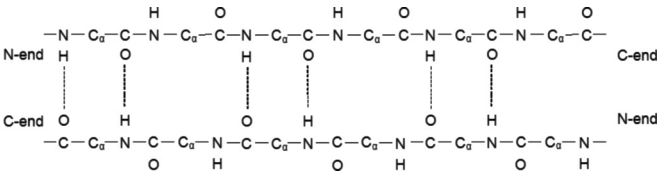

FIG. 5. The configuration of hydrogen bonds in the β sheet.

write the equation of chemical potential as follows:

$$U^e = NU_h^e + N_a U_a^e + N_b U_b^e, \tag{7}$$

where N gives the total number of hydrogen bonds, while N_a and N_b give the number of hydrogen bonds associated with the α helix and β sheet, respectively. Note that U_h^e is the chemical potential of one hydrogen bond. U_a^e and U_b^e denotes the statistical weights of the chemical potential of the hydrogen bond associated with the α helix and β sheet structures, respectively.

In our coarse-grained model, we shall restrict the dihedral angles to five distinct sets of angles. The selection is made according to the Ramachandran plot, which indicates regions in terms of dihedral angles where stable secondary structures can form Ref. [13]. The selected angles or states are $\{-57.4^\circ, -47.5^\circ\}$, $\{-139^\circ, 135^\circ\}$, $\{315^\circ, 110^\circ\}$, $\{105^\circ, 330^\circ\}$, and $\{180^\circ, 180^\circ\}$. Note that the last set of angles form the coil structure. A combination of these angles leads to the formation of α helix (Fig. 2), β sheet (Fig. 3), and β turn. A schematic on the configuration of hydrogen bonds that give rise to the α helix and β sheet is shown in Figs. 4 and 5, respectively. The random coil does not have any hydrogen bond in its structure.

III. PARTITION FUNCTION OF PROTEIN CHAIN

If a protein has n cranks, the restriction of the sets of dihedral angles for each crank to five distinct states implies that the chain can form into a total of 5^n possible structures. Let each structure be i, then the partition function of the ith structure is given as follows:

$$Z_i = \int \exp\left\{-\beta \cdot \left[K_i(p) + U_i(q) + U_i^e\right]\right\} d\vec{p}\, d\vec{q}, \tag{8}$$

where $\beta = 1/k_B T$, with k_B the Boltzmann constant and T the temperature. Note that from now on the subscript i means that the value of the quantity it attaches to is to be determined via the set of n dihedral angles or states which corresponds to the ith structure. In Eq. (8), K_i is written to depend only on p because the effective potential is found to be essentially independent of the chain conformations [12]. Since $\dot{q} = M^{-1} p$, Eq. (2) can also take the following form:

$$K_i(p) = \tfrac{1}{2} p^T M_i^{-1} p, \tag{9}$$

with the mass matrix M_i independent of q.

Substituting Eqs. (4), (7), and (9) into Eq. (8), we obtain

$$Z_i = \int \exp\left\{-\beta \cdot \left[\frac{1}{2} p^T M_i^{-1} p + \frac{1}{2} q^T (\kappa_1 D_i + \kappa_2 C_i) q + \left(N_i U_h^e + N_{\text{ài}} U_a^e + N_{bi} U_b^e\right)\right]\right\} d\vec{p}\, d\vec{q}. \tag{10}$$

After evaluating the integral in Eq. (10), we arrive at the following expression for the partition function:

$$\begin{aligned} Z_i &= \frac{(2\pi k_B T)^{n-1}}{\sqrt{\det M_i^{-1}}} \frac{(2\pi k_B T)^{N_i}}{\sqrt{\det_p (\kappa_1 D_i + \kappa_2 C_i)}} (2\pi)^{2(n-1)-2N_i} e^{-\beta U_i^e} \\ &= \frac{(2\pi k_B T)^{n-1}}{\left(\bar{\lambda}_i^M\right)^{2(n-1)}} \frac{(2\pi k_B T)^{N_i}}{\left(\bar{\lambda}_i^K\right)^{2N_i}} (2\pi)^{2(n-1)-2N_i} \\ &\quad \times e^{-(1/k_B T)(N_i U_h^e + N_{ai} U_a^e + N_{bi} U_b^e)}, \end{aligned} \tag{11}$$

where

$$\begin{aligned} \left(\bar{\lambda}_i^M\right)^{2(n-1)} &= \left(\omega_i^{M1} \omega_i^{M2} \omega_i^{M3} \ldots \omega_i^{M2(n-1)}\right)^{1/2}, \\ \left(\bar{\lambda}_i^K\right)^{2N_i} &= \left(\omega_i^{K1} \omega_i^{K2} \omega_i^{K3} \ldots \omega_i^{K2N_i}\right)^{1/2}. \end{aligned} \tag{12}$$

Note that ω_i^{Mj} and ω_i^{kj} are the jth eigenvalues of the matrices M_i^{-1} and $(\kappa_1 D_i + \kappa_2 C_i)$, respectively. Since N_i is always less than n, zero eigenvalues are to be expected in the potential matrix $(\kappa_1 D_i + \kappa_2 C_i)$. Thus, we have a "pseudodeterminant" term $\det_p()$ in Eq. (11) which takes in only the nonzero eigenvalues.

By summing over the partition function Z_i of each structure i, we form the grand partition function of a protein chain as

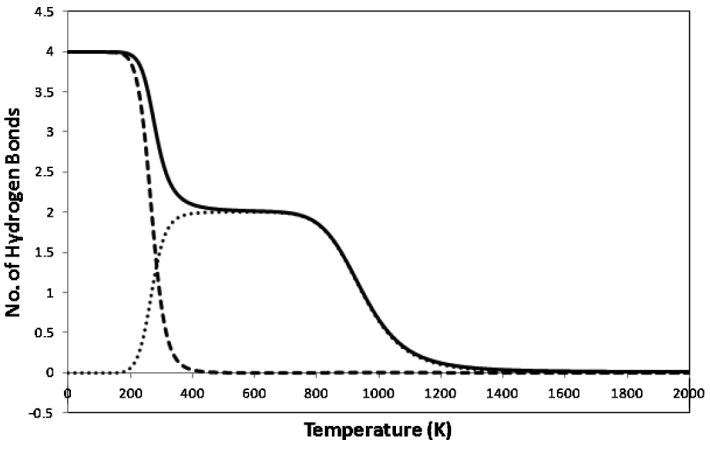

FIG. 6. The average number of hydrogen bonds in a seven crank polyalanine versus temperature obtained from an ensemble of 5^7 configurations. The three lines are the total number of hydrogen bonds (solid line), number of hydrogen bonds associated with α helix (dashed line), and number of hydrogen bonds associated with β sheet (dotted line).

follows:

$$\mathbb{Z} = \sum_{i=1}^{5^n} Z_i = \sum_{i=1}^{5^n} \left\{ \frac{(2\pi k_B T)^{n-1}}{\left(\bar{\lambda}_i^M\right)^{2(n-1)}} \frac{(2\pi k_B T)^{N_i}}{\left(\bar{\lambda}_i^K\right)^{2N_i}} (2\pi)^{2(n-1)-2N_i} \times e^{-(1/k_B T)(N_i U_h^e + N_{ai} U_a^e + N_{bi} U_b^e)} \right\}. \quad (13)$$

In lieu of the fact that the term $(\bar{\lambda}_i^M)^{2(n-1)}$ is large during computation, each Z_i is very small. It turns out that the value of the grand partition function is approximately zero even after performing the 5^n summation. In order to avoid numerical inaccuracies, we have rescaled Eq. (13) by $\bar{\lambda}_M^{2(n-1)}$, with $\bar{\lambda}_M = \sum_i \bar{\lambda}_i^M / 2^n$, as follows:

$$\mathbb{Z} = \frac{(2\pi k_B T)^{n-1}}{\bar{\lambda}_M^{2(n-1)}} \sum_{i=1}^{5^n} \left\{ \frac{\bar{\lambda}_M^{2(n-1)}}{\left(\bar{\lambda}_i^M\right)^{2(n-1)}} \frac{(2\pi k_B T)^{N_i}}{\left(\bar{\lambda}_i^K\right)^{2N_i}} (2\pi)^{2(n-1)-2N_i} \times e^{-(1/k_B T)(N_i U_h^e + N_{ai} U_a^e + N_{bi} U_b^e)} \right\}. \quad (14)$$

Note that the rescaling has no consequence on our investigation of the statistical properties of the phase transition in the next section since the term $\bar{\lambda}_M^{2(n-1)}$ will be canceled out during the ensemble averaging operation.

IV. STATISTICAL PHYSICS OF PROTEIN PHASE TRANSITION

We shall employ the grand partition function to calculate the ensemble average of different statistical properties of a protein chain. The first property that we investigate is the average number of hydrogen bonds in the protein as a function of temperature. This average number $\langle N \rangle$ is obtained from the partition function in the following manner:

$$\langle N \rangle = \frac{\sum_{i=1}^{5^n} (N_i Z_i)}{\sum_{i=1}^{5^n} Z_i}. \quad (15)$$

Note that Z_i is given by Eq. (11). In the case of a small protein, $\langle N \rangle$ gives a good estimate on the type of its secondary structure.

In order to proceed, we require the relative weight of U_a^e and U_b^e. Let us assume that $U_a^e = U_h^e$ and $U_b^e = 2U_h^e$, which is reasonable according to Refs. [18,19]. With this assumption, we have computed the number of hydrogen bonds against temperature for a seven crank polyalanine, which is plotted in Fig. 6. From Fig. 6, we observe that the number of hydrogen bonds decreases as temperature increases. On further analysis, we found that $\langle N \rangle$ in Fig. 6 is the sum of the hydrogen bonds of the α helical and β sheet structures. At low temperature, we observe four hydrogen bonds and the protein is an α helix. When the temperature is increased to a critical value, a transition occurs with the protein containing a mixture of α helical and β sheet structures, before turning into a full β sheet. As this happens, the number of hydrogen bonds reduces. A further increase in temperature eventually leads to the next critical value at which the β sheet denatures into a random coil.

The secondary structure transition that we have described possesses the characteristic of a first-order phase transition. In first-order phase transition, a sharp peak in the heat capacity is to be expected against a change in temperature. This results from a sudden change in the internal energy of the system [17]. Thus, in order to establish the nature of the transitions observed in Fig. 6, we compute the ensemble average of our second property of interest: the heat capacity of the protein polyalanine as a function of temperature and we then observe the possible presence of sharp peaks.

The definition of the heat capacity is given as follows:

$$C(T) = k_B T \frac{\partial^2 T \ln \mathbb{Z}}{\partial T^2}. \quad (16)$$

By using Eq. (14), the heat capacity of our model can be expressed in the following manner:

$$C(T) = k_B T \left[\frac{2}{\mathbb{Z}} \frac{\partial \mathbb{Z}}{\partial T} - \frac{T}{\mathbb{Z}^2} \left(\frac{\partial \mathbb{Z}}{\partial T} \right)^2 + \frac{T}{\mathbb{Z}} \frac{\partial^2 \mathbb{Z}}{\partial T^2} \right] = k_B T \left\{ \frac{2}{(2\pi k T)^{n-1} \mathbb{A}} [(n-1)(2\pi k_B T)^{n-2}(2\pi k_B)\mathbb{A} + (2\pi k_B T)^{n-1}\mathbb{B}] - \frac{T}{(2\pi k_B T)^{2(n-1)} \mathbb{A}^2} [(n-1)(2\pi k_B T)^{n-2}(2\pi k_B)\mathbb{A} + (2\pi k_B T)^{n-1}\mathbb{B}]^2 + \frac{T}{(2\pi k_B T)^{n-1}\mathbb{A}} [(n-1)(n-2)(2\pi k_B T)^{n-3}(2\pi k_B)^2 \mathbb{A} + 2(n-1)(2\pi k_B T)^{n-2}(2\pi k_B)\mathbb{B} + (2\pi k_B T)^{n-1}\mathbb{C}] \right\}, \quad (17)$$

where

$$\mathbb{A} = \sum_{i=1}^{5^n} \left\{ \frac{\bar{\lambda}_M^{2(n-1)}}{\left(\bar{\lambda}_i^M\right)^{2(n-1)}} \frac{(2\pi k_B T)^{N_i}}{\left(\bar{\lambda}_i^K\right)^{2N_i}} (2\pi)^{2(n-1)-2N_i} e^{-U_i^e / k_B T} \right\}, \quad (18)$$

$$\mathbb{B} = \sum_{i=1}^{5^n} \left\{ \frac{\bar{\lambda}_M^{2(n-1)} (2\pi)^{2(n-1)-2N_i}}{\left(\bar{\lambda}_i^M\right)^{2(n-1)} \left(\bar{\lambda}_i^K\right)^{2N_i}} \left[N_i (2\pi k_B T)^{N_i - 1} (2\pi k_B) e^{-U_i^e / k_B T} + (2\pi k_B T)^{N_i} e^{-U_i^e k_B T} \left(-\frac{U_i^e}{k_B T^2} \right) \right] \right\}, \quad (19)$$

$$\mathbb{C} = \sum_{i=1}^{5^n} \left\{ \frac{\bar{\lambda}_M^{2(n-1)} (2\pi)^{2(n-1)-2N_i}}{\left(\bar{\lambda}_i^M\right)^{2(n-1)} \left(\bar{\lambda}_i^K\right)^{2N_i}} e^{-U_i^e k_B T} \left[N_i (N_i - 1)(2\pi k_B T)^{N_i-2} (2\pi k_B)^2 + 2N_i (2\pi k_B T)^{N_i-1} (2\pi k_B) \left(-\frac{U_i^e}{k_B T^2} \right) \right.\right.$$
$$\left.\left. + (2\pi k_B T)^{N_i} \left(\left[-\frac{U_i^e}{k_B T^2} \right]^2 + \frac{2U_i^e}{k_B T^3} \right) \right] \right\}. \quad (20)$$

We plot heat capacity against temperature in Fig. 7, where we observe the occurrence of two peaks which confirm the presence of two first-order phase transitions. The first peak represents the transition from α helix to β sheet at a temperature of around 300 K. The second peak arises from the transition from β sheet to random coil at a temperature of around 950 K.

In summary, our results as displayed in Figs. 6 and 7 illustrate that the physics of protein phase transitions is captured by the grand partition function $\mathbb{Z}$. The results in these figures imply two important facts: (1) the network of hydrogen bonds stabilizes the protein secondary structure, and (2) the hydrogen bond loses its hold on the structure when the temperature reaches a critical value, and as this happens a new secondary structure appears.

With $\mathbb{Z}$, we can explore various aspects of the protein dynamics through its statistical physics, and this makes our theory more comprehensive than the classical two-state models [3,4]. Our theory differs from others that also take a partition function approach, such as theories due to Zimm-Bragg, Lifson-Roig, and Lei *et al*. In these theories, the residue of a protein is assigned a state from a discrete set, and each state is given a statistical weight. In our model, we assign five states to each crank unit, but in the form of five distinct sets of dihedral angles. Furthermore, instead of assigning a different statistical weight to each state, we assign the statistical weight based on the type of the hydrogen bond, i.e., depending on whether the hydrogen bond is associated with the α helix or the β sheet. Our approach has the advantage of being less restrictive because a residue can adopt more than just the pairs of dihedral angles that form a helix, sheet, or coil. However, it requires greater computational power and memory due to the need to sum through a larger number of configurations during the calculation of the grand partition function of the protein chain. This limitation has essentially restricted our study to at most seven cranks with a maximum of five distinct states for each crank.

Comparing against our earlier results in Ref. [12] where the α-β and the β-coil transition temperatures are 475 and 600 K, respectively, we observe that the α-β transition temperature has lowered to 300 K while the β-coil transition temperature has increased to 950 K. This results from our assumption in this paper that the statistical weighting of the hydrogen bonds in the β sheet is greater than that in the α helix. With the hydrogen bonds being relatively more stable in the β sheet than in the α helix, we anticipate the β sheet to form at a lower temperature than before. Furthermore, the increased stability of the hydrogen bonds in the β sheet implies that a higher temperature is needed to break all its hydrogen bonds and converts it into a random coil. In fact, the α-β transition temperature of 300 K that we have obtained is rather close to the 315 K predicted by Ding *et al*. [8] through discrete molecular dynamics simulation, although our β-coil transition temperature of 950 K is vastly different from the 327 K obtained by them. Figure 6 also indicates that the helix-coil transition temperature occurs at about 300 K when the number of hydrogen bonds from the α helix drops to zero. A comparison of this temperature against the predictions made by the models of Yasar and Demir [10] and Lee *et al*. [20] has further validated our approach to be a reliable concept.

We have plotted in Fig. 8 the Helmholtz free energy of the seven crank polyalanine in the α helix, the β sheet, and the random coil configuration as a function of temperature (see also Fig. 8 of Ref. [12]). In the calculation of the Helmholtz free energy (which is based on the approach in Ref. [12]), we have considered the different statistical weights for the hydrogen bonds in the α helix and in the β sheet structure as introduced in this section. From Fig. 8, we observe an α-β transition temperature of about 400 K and a β-coil transition temperature of 1900 K. This is to be expected based on the

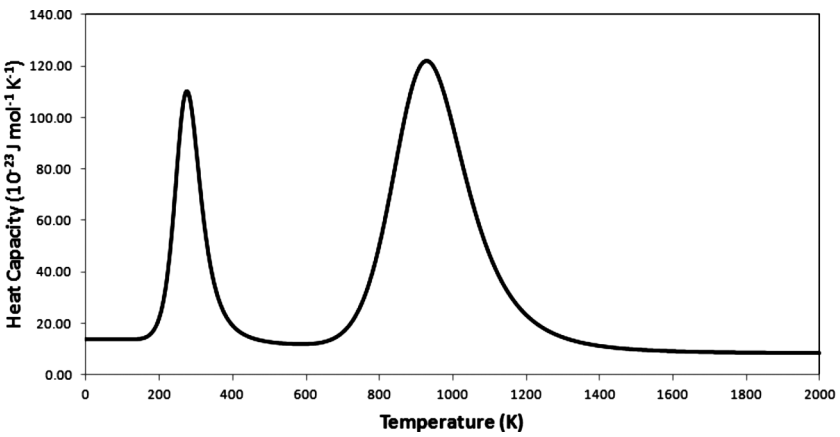

FIG. 7. The heat capacity of a seven crank polyalanine versus temperature obtained from an ensemble of 5^7 configurations. The first phase transition occurs at a temperature around 300 K with the α helix transforming into a β sheet. The second phase transition occurs at around 950 K with the β sheet turning into a random coil.

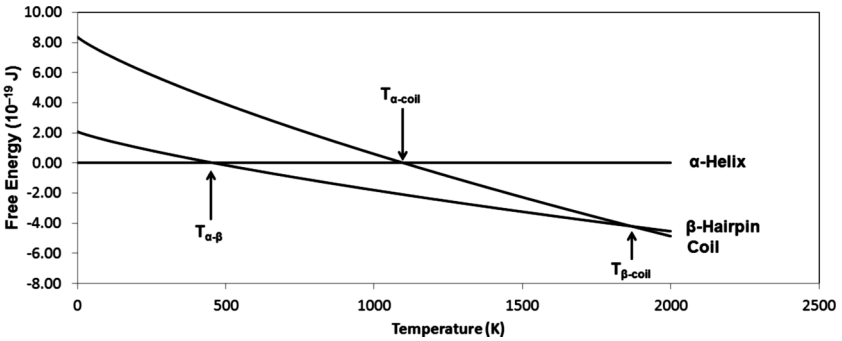

FIG. 8. Free energy curves of α helix, β sheet, and random coil for a seven crank polyalanine. Note that the intersection of the curves indicate phase transition. The transition temperatures are $T_{\alpha\text{-}\beta} = 400$ K, $T_{\alpha\text{-coil}} = 1100$ K, and $T_{\beta\text{-coil}} = 1900$ K.

argument above with the β sheet being relatively more stable than the α helix.

V. FURTHER ANALYSIS AND DISCUSSION ON PROTEIN PHASE TRANSITION

By assigning five states to each crank such that each of the 5^n configurations is associated to the partition function Z_i as given by Eq. (8), the grand partition function $\mathbb{Z}$ describes a free energy landscape consisting of each of these 5^n configurations as local minimum. In other words, we have assumed that the 5^n structures are relatively more stable compared to their neighboring structures in the free energy space in lieu of our approach in evaluating the Gaussian integral given by Eq. (10) based on the stationary phase approximation. While we expect our assumption to remain true as n increases under the above scenario (which is validated for a two-state polypeptide below), the situation would be different if we were to increase the number of states per crank or to consider a sequence dependent protein structure, since a configuration under these circumstances may not constitute a local minimum. Nonetheless, the inclusion of additional states or information per crank may lead to new stable structural phases that emerge out of the extra statistical information hidden in the grand partition function.

Let us now employ the above argument to explain the results obtained in Fig. 8, where we have come to understand that a single state of either α helix, β sheet, or random coil has been allocated to each crank. This implies the presence of a single minimum for the free energy landscape, and the curves in Fig. 8 plot the Helmholtz free energy of the minimum for the corresponding conformation. In order to determine the stable configuration at a particular temperature, the three separate minimum free energies are then compared at the given temperature with the configuration having the lowest free energy being chosen as the stable conformation. This approach differs from that of the grand partition function where the statistical information of all the possible configurations is contained within the function itself. The presence of more information in the grand partition function approach leads to a more comprehensive treatment of phase transition. Nevertheless, both approaches reach the same conclusion of α helix being formed at low temperature, followed by β sheet being the stable structure as temperature increases, and the polypeptide becomes denatured at high temperature.

At the moment, our ability to develop further analysis on our theory using a longer polypeptide with more than five states per crank is severely limited by the available memory resources, even though our computation is performed on two compute nodes with 8 core and 16 GB RAM per node. As a result, in order to investigate the impact of a larger number of cranks, we have to reduce the number of states per crank. In this respect, we have considered a two-state polypeptide, i.e., a crank with dihedral angles $\{-57.4^\circ, -47.5^\circ\}$ and $\{180^\circ, 180^\circ\}$, which correspond to the α helical and the random coil configurations, respectively. For n cranks, we thus anticipate a free energy landscape with 2^n local minima. Our results show a persistent α helix to random coil phase transition at the same critical temperature for different n. This confirms our expectation that the phase transition in a polypeptide is independent of its size. We observe that the energies of the local minimum of the stable phases remain significant as n increases. Figure 9 illustrates the α-coil transition of the two-state polypeptide when the length ranges from $n = 7$ to $n = 15$. The figure clearly shows that the first order phase transition becomes more prominent as n becomes larger. To summarize, we observe that the reduction

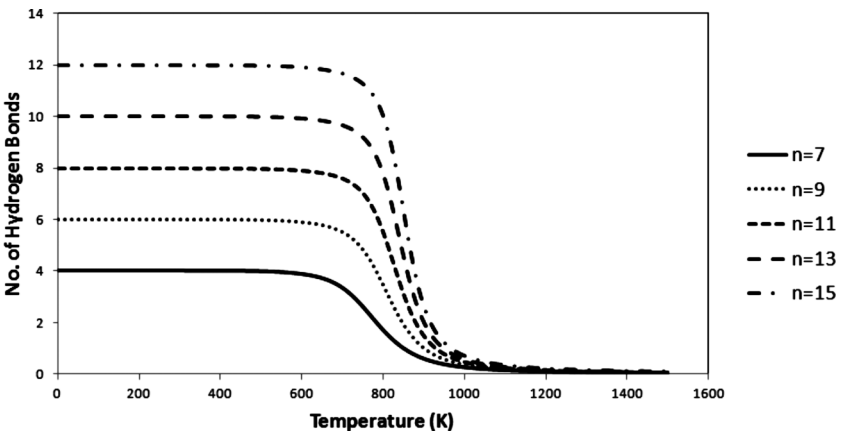

FIG. 9. The average number of hydrogen bonds versus temperature for different lengths of a two-state polypeptide.

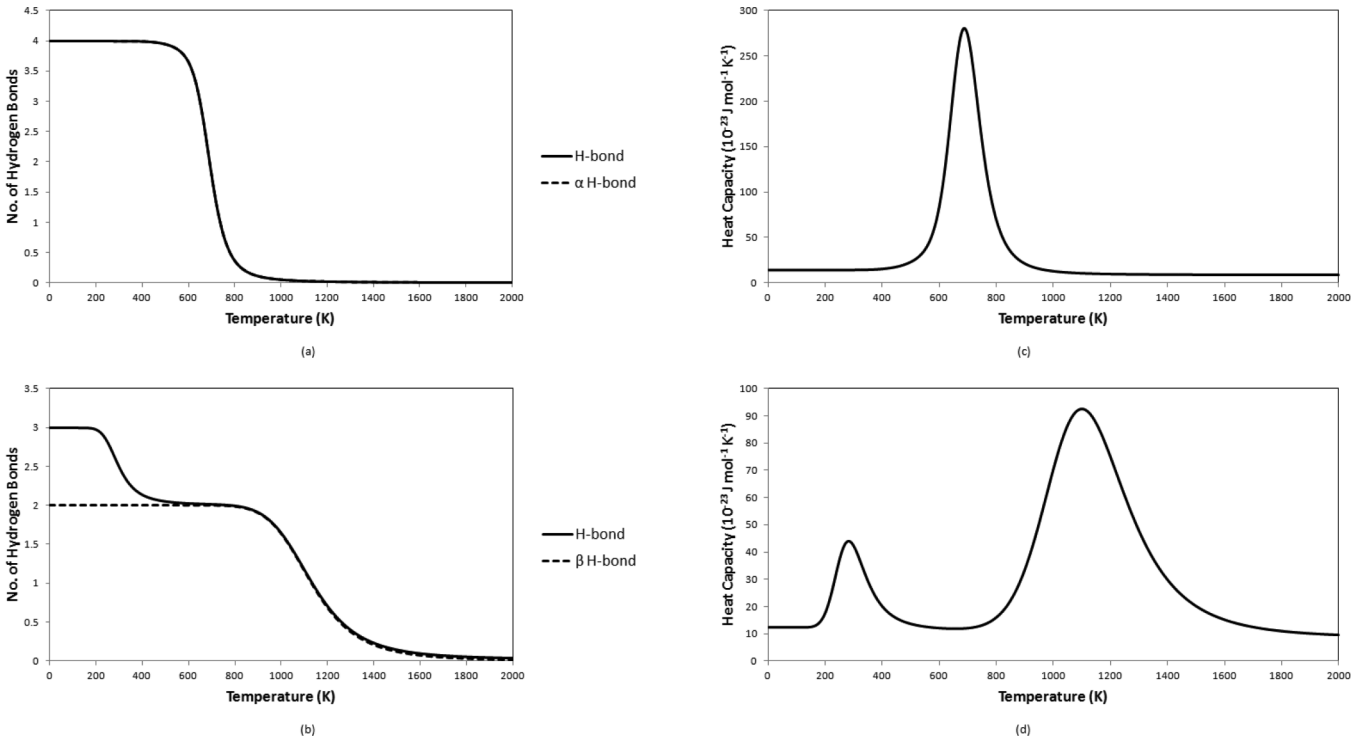

FIG. 10. Results for the three-state polypeptide obtained from an ensemble of 3^7 configurations: (a) and (b) show the plots of average number of hydrogen bonds versus temperature; (c) and (d) show the corresponding plots of heat capacity versus temperature. In (b), we observe the presence of an additional hydrogen bond at lower temperature, which is associated with neither the α helical nor β sheet structures. It corresponds to the first peak in the heat capacity plot, while the second peak represents the β-coil phase transition.

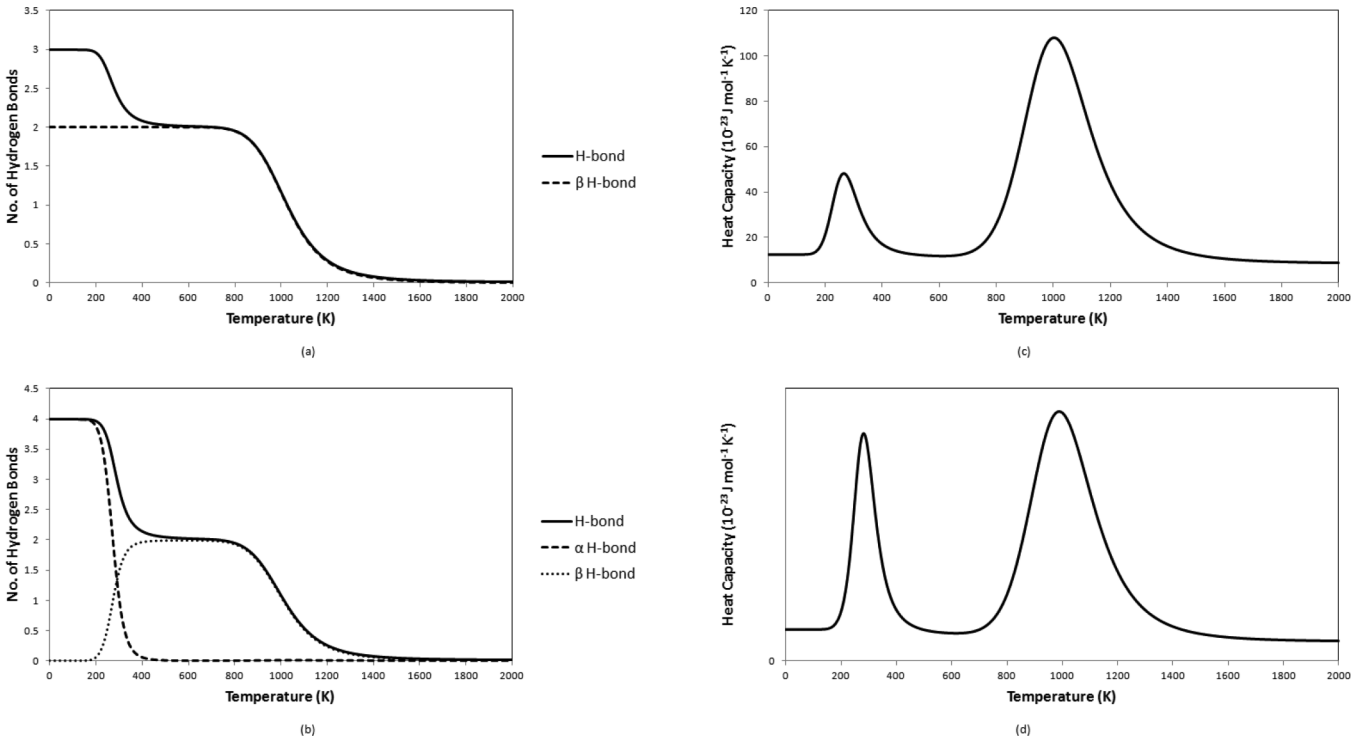

FIG. 11. Results for the four-state polypeptide obtained from an ensemble of 4^7 configurations: (a) and (b) show the plots of average number of hydrogen bonds versus temperature; (c) and (d) show the corresponding plots of heat capacity versus temperature. While (a) is analogous to Fig. 10(b) depicting a β-coil phase transition, (b) illustrates the occurrence of the α-β-coil phase transition similar to Fig. 6.

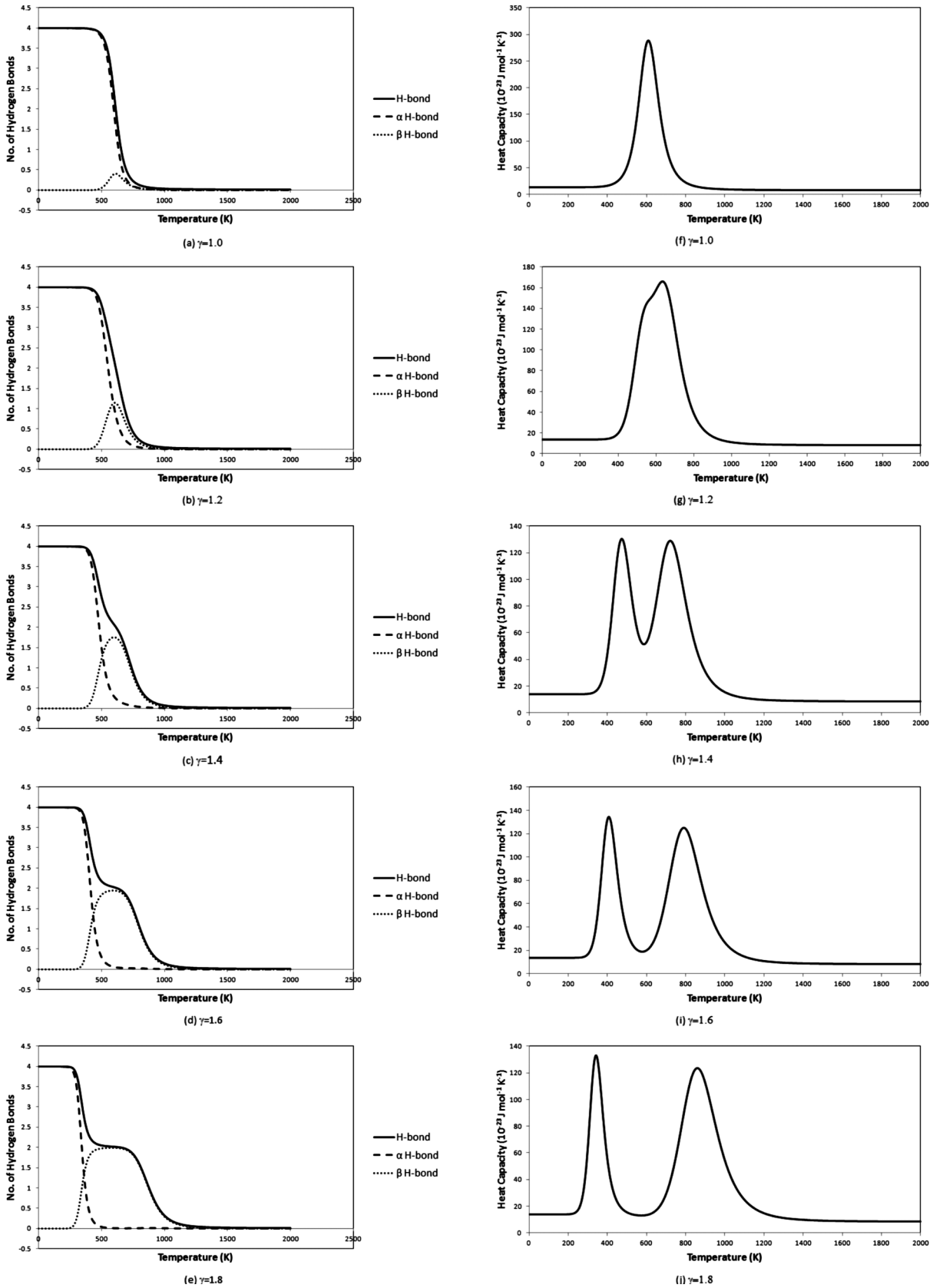

FIG. 12. (a)–(e) Phase transition of a seven-crank polyalanine at different U_b^e to U_a^e ratios, from $\gamma = 1.0$ to $\gamma = 1.8$; their corresponding plots of heat capacity are displayed in (f)–(j).

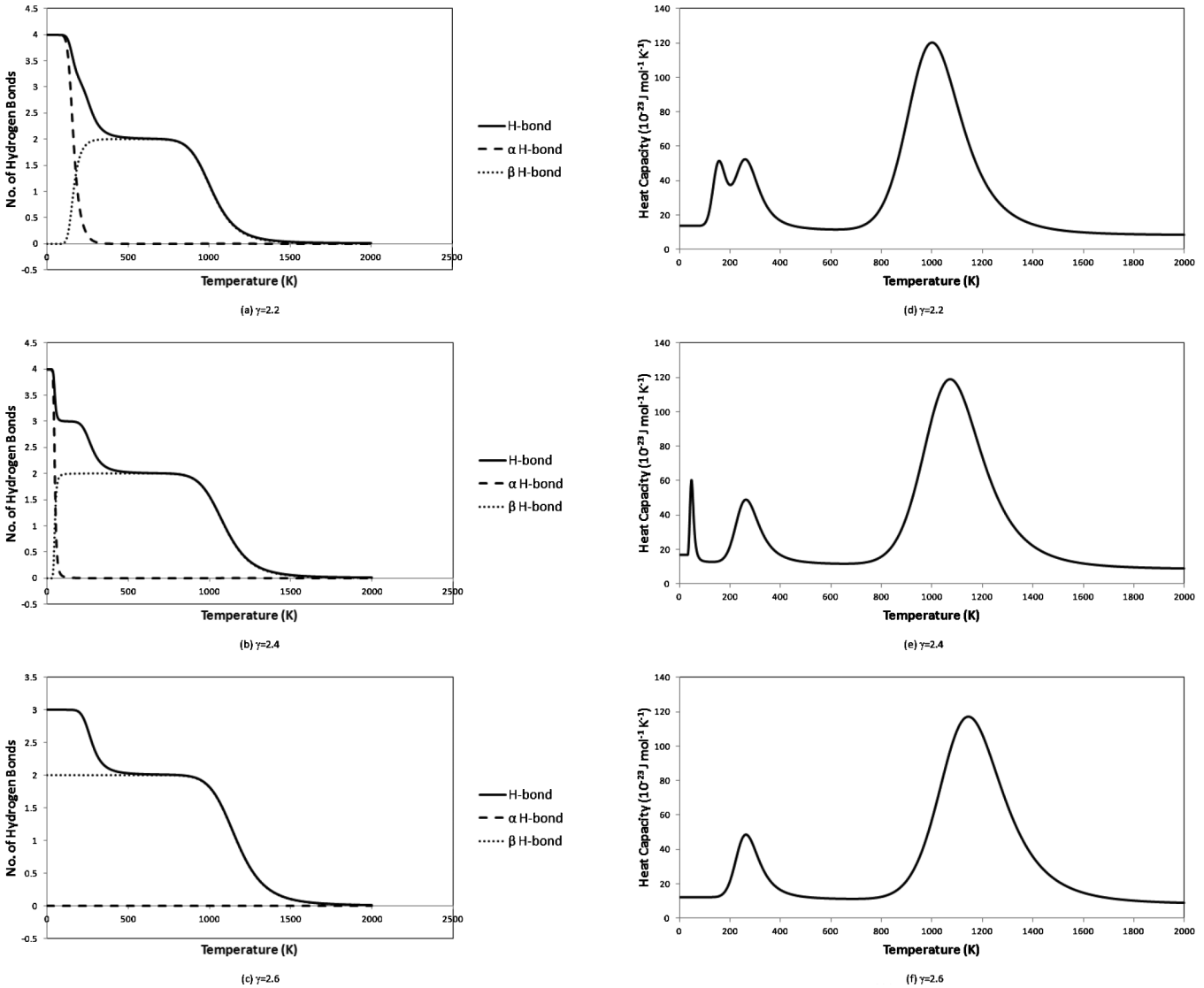

FIG. 13. (a)–(c) Phase transition of a seven-crank polyalanine at different U_b^e to U_a^e ratios, from $\gamma = 2.2$ to $\gamma = 2.6$; their corresponding plots of heat capacity are displayed in (d)–(f).

of the number of states per crank from five to two has led to the disappearance of the phase of stable β sheet. In other words, the phase transition from α helix to β sheet, and from β sheet to random coil becomes untenable.

We can continue to investigate the role of the total number of distinct states on protein phase transition by considering the case with three and four states per crank. By adding the dihedral angle $\{315^\circ,110^\circ\}$ to the set $\{-57.4^\circ, -47.5^\circ\}$ and $\{180^\circ,180^\circ\}$ to form a three-state seven-crank polypeptide, we notice the occurrence of an α-coil phase transition as the two-state polypeptide that we have discussed above, albeit with a reduced critical temperature of about 687 K [see Figs. 10(a) and 10(c)]. However, if we were to use the set $\{-139^\circ,135^\circ\}$, $\{315^\circ,110^\circ\}$, and $\{105^\circ,330^\circ\}$ instead, we observe two transitions. The first transition at 282 K is not a secondary structure phase transition since it involves the denaturing of a single hydrogen bond not associated with any secondary structure. On the other hand, the second transition at 1100 K is a β-coil phase transition. The results are shown in Figs. 10(b) and 10(d). Similar transition scenarios occur when we add the dihedral angle $\{180^\circ,180^\circ\}$ to the latter set of angles to form a four-state polypeptide. We again detect a first-order phase transition from the β sheet to the random coil, although it now arises at a lower critical temperature of 1002 K [see Figs. 11(a) and 11(c)]. If we were to use the set of angles $\{-57.4^\circ, -47.5^\circ\}$, $\{-139^\circ,135^\circ\}$, $\{315^\circ,110^\circ\}$, and $\{105^\circ,330^\circ\}$ instead to form a four-state polypeptide, we observe both the α-β and β-coil phase transition as our original five-state polypeptide, with the transition temperatures now happening at 283 and 990 K, respectively [see Figs. 11(b) and 11(d)]. These results illustrate that while the number of states per crank is small, the selection of dihedral angles can have important consequences on the emergence of the type of stable secondary structures, which is being revealed as the temperature changes. Conversely, we believe that new stable structural phases and new phase transition scenarios should arise if we were to increase the number of states per crank beyond 5.

While we have let $U_b^e = \gamma U_a^e$ with $\gamma = 2.0$ in the evaluation of our results in Figs. 6 and 7 it is interesting to explore the effect of varying the relative statistical weight between U_b^e

and U_a^e as represented by γ on the structural phase transition (see Figs. 12 and 13). When $\gamma = 1.0$ and 1.2, we observe a transition from α helix to random coil as temperature increases. A stable phase of β sheet starts to appear only at $\gamma = 1.4$, and becomes apparent after $\gamma = 1.6$. More specifically, we observe two phase transitions from α helix to β sheet, and then from β sheet to random coil, at $\gamma = 1.6,\ 1.8,\ 2.0$, and 2.2. Subsequently, at $\gamma = 2.4$, we observe the unexpected presence of three stable phases: a phase with four hydrogen bonds (α helix); a phase with three hydrogen bonds (a mixture of α helix and β sheet); a phase with two hydrogen bonds (β sheet); and a phase with no hydrogen bond (coil). In fact, the new phase of a mixture of α helix and β sheet begins to emerge at $\gamma = 2.2$ according to our heat capacity calculation. When γ is further increased to 2.6, the phase of α helix disappears, and phase transition now occurs between the three phases of the mixture of α helix and β sheet, the β sheet, and the random coil. These outcomes can be understood to result from the progressive enhancement in stability of the β sheet structure as the weight of hydrogen bond energy associated to the β sheet becomes relatively larger than that associated to the α helix. Finally, it is important to note that all the transitions observed are first-order phase transitions based on the results of heat capacity (see Figs. 12 and 13).

VI. CONCLUSION

In this paper, we have investigated into the mechanism of protein phase transition by means of the Hamiltonian formalism, which allows us to construct the grand partition function of the protein chain. In our model, we have assumed that the dominant noncovalent potential is due to the hydrogen bond interaction. In addition, we have set different statistical weights for the strength of the hydrogen bonds within the α helix and the β sheet structure. By considering five distinct states for the dihedral angles, we have computed the average number of hydrogen bonds and the heat capacity of the protein polyalanine using the grand partition function at different temperatures. We observe a sudden change in the average number of hydrogen bonds at two critical temperatures, 300 and 950 K, signifying an α-β and a β-coil phase transition. Our results based on the heat capacity confirm that these two transitions are first-order phase transitions.

ACKNOWLEDGMENTS

We would like to thank NTU High Performance Computing Centre for providing the computational resources. This work was partially supported by MOE AcRF Tier 1 Grant No. RG52/08.

[1] S. B. Prusiner, Proc. Natl. Acad. Sci. USA **95**, 13363 (1998).
[2] T. E. Creighton, *Protein Folding* (W. H. Freeman & Co., New York, 1992).
[3] B. H. Zimm and J. K. Bragg, J. Chem. Phys. **31**, 526 (1959).
[4] S. Lifson and A. Roig, J. Chem. Phys. **34**, 1963 (1961).
[5] A. V. Yakubovich, I. A. Solov'yov, A. V. Solov'yov, and W. Greiner, Eur. Phys. J. D **40**, 363 (2006).
[6] A. V. Yakubovich, I. A. Solov'yov, A. V. Solov'yov, and W. Greiner, Eur. Phys. J. D **46**, 215 (2008).
[7] I. A. Solov'yov, A. V. Yakubovich, A. V. Solov'yov, and W. Greiner, Eur. Phys. J. D **46**, 227 (2008).
[8] F. Ding, J. M. Borreguero, S. V. Buldyrey, H. E. Stanley, and N. V. Dokholyan, Proteins **53**, 220 (2003).
[9] L. Hong and J. Lei, Phys. Rev. E **78**, 051904 (2008).
[10] F. Yasar and K. Demir, Comput. Phys. Commun. **175**, 604 (2006).
[11] J. H. Gibbs and E. A. DiMarzio, J. Chem. Phys. **30**, 271 (1959).
[12] H. W. Leong, L. Y. Chew, and K. Huang, Phys. Rev. E **82**, 011915 (2010).
[13] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, J. Mol. Biol. **7**, 95 (1963).
[14] K. Huang, Biophys. Rev. Lett. **3**, 1 (2008).
[15] K. Huang, Biophys. Rev. Lett. **2**, 139 (2007).
[16] A. Karshiko, *Non-Covalent Interactions in Proteins* (Imperial College Press, London, 2006).
[17] A. V. Finkelstein and O. B. Ptitsyn, *Protein Physics: A Course of Lectures* (Academic, London, 2002).
[18] N. Y. Chen, C. Y. Mou, and Z. Y. Su, Phys. Rev. Lett. **96**, 078103 (2006).
[19] B. C. Goh, H. W. Leong, X. Qu, and L. Y. Chew, Eur. Phys. J. E **35**, 27 (2012).
[20] M. S. Lee, G. G. Wood, and D. J. Jacobs, J. Phys.: Condens. Matter **16**, S5035 (2004).