Normal modes and phase transition of the protein chain based on the Hamiltonian formalism

Hon-Wai Leong,[1] Lock Yue Chew,[1] and Kerson Huang[2,3]

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

[2]*Institute of Advanced Studies, Nanyang Technological University, Nanyang Executive Centre No. 02-18, 60 Nanyang View,*
Singapore 639673, Singapore

[3]*Physics Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

We use the torsional angles of the protein chain as generalized coordinates in the canonical formalism, derive canonical equations of motion, and investigate the coordinate dependence of the kinetic energy expressed in terms of the canonical momenta. We use the formalism to compute the normal-frequency distributions of the α helix and the β sheet, under the assumption that they are stabilized purely through hydrogen bonding. In addition, we obtain the free-energy relations of the α helix, the β sheet, and the random coil of a 15-residue polyalanine. Interestingly, our results predict a phase transition from an α helix to a β sheet at a critical temperature.

PACS number(s): 87.15.bd, 87.15.Zg, 87.14.E−, 87.15.Cc

I. INTRODUCTION AND SUMMARY

At present, protein folding is one of the outstanding problems in protein science. Extensive research has been carried out to understand how specific native structure can arise from a given sequence of amino acids. This structure, which determines the function of a protein, is known to exhibit a certain range of slow vibration modes. It is these slow vibration modes that play an important part in the catalytic function of proteins [1]. Studies on these slow modes have been performed both experimentally and theoretically [2]. Among these, the normal-mode analysis (NMA) is a technique that explores the flexibility and the slow range of motion of diverse protein configurations [3–8]. In the classical NMA approach, the degrees of freedom are described by a set of generalized coordinates, with the force field being given semiempirically, consisting of energy terms due to stretching, bending, torsion, as well as van der Waals and electrostatic interactions [4]. Although classical NMA has reduced the computational expense on the investigation of low-frequency collective modes of proteins in comparison to molecular dynamics, it is still limited to small proteins. This has motivated the formulation of coarse-grained models such as the Gō model [9,10], the elastic network model [11–13], and the Gaussian network model [14–16], which leads to greater computational efficiency. Interestingly, the low-frequency normal modes determined from these simplistic models correspond closely not only to those obtained from the more sophisticated force fields, but they also correspond to the conformational transitions observed experimentally [17,18]. In particular, the application of NMA to these simplistic models has led to theoretical studies on model refinement of crystallographic or diffraction data [19,20], cooperative and hinge-bending motion in enzyme [21], mechanism of allosteric communication [22], NMR order parameters [23,24], and conformational changes in large complexes such as viruses [25,26]. Remarkably, the results from these studies are found to correlate well with those obtained experimentally.

It is known that the very low-frequency modes of a protein can be associated with those of secondary structures, namely, α helices and β sheets. It is of particular interest to study possible transitions between these structures, which may cause protein misfolding, resulting in a loss of normal biological functions. Protein misfolding is known to be a source of debilitating diseases, such as the "mad cow" disease caused by prion misfolding [27]. A number of theories for such transitions have been proposed [28–32]. Our investigation of the slow vibration modes involves the consideration of protein as being made up of a chain of amino acids connected by covalent bonds that are unbreakable. From the point of view of the mechanics of protein folding, the independent variables are the dihedral angles between successive units of the chain. Other degrees of freedom, such as the vibrations of bond lengths and angles, do not affect the general shape of the protein and will be neglected.

We begin our work with a development of the canonical formalism, using the dihedral angles as generalized coordinates. An important consequence is that the kinetic energy, when expressed in terms of the canonical momenta, becomes a function of the coordinates. Specifically, masses are replaced with a mass matrix which depends on the coordinates, and this gives rise to an effective potential. By studying this mass matrix numerically, we find that the effective potential is approximately constant for almost all conformations of the chain. This result is significant for practical applications, particularly for the conditioned self-avoiding walk (CSAW) model [33,34], in which these generalized coordinates were explicitly used.

The canonical formulation has a wide range of applications, including the formulation of kinetic equations for protein folding and the calculation of physical quantities from statistical mechanics. In this work, we studied the phase transition of protein. We considered only hydrogen bonding as the potential in our protein model. Thus, the low-frequency collective modes considered in this paper arise solely from this simple potential field. Although our model is crude, it is proven to be sufficient. This was verified from the computation of the distribution functions of normal frequencies for a pure α helix and a pure β sheet. We assume that the α helix and β sheet are stabilized purely through hydrogen bonding,

and the restoring potential about the equilibrium conformation arises from the deformation of hydrogen bonds.

We find that all normal frequencies are positive definite,[1] and this shows that the α helix and the β sheet are mechanically stable under hydrogen bonding alone. This leads one to expect that, in the unfolded protein chain, which is subject to random bombardments from water molecules, these secondary structures should have transient existence. We further examine our theory on an all-α myoglobin protein. Our computed results agree with those calculated using the classical NMA.

We calculate the partition functions of a pure α helix and a pure β sheet, which are treated as collections of harmonic oscillators with given frequency distributions. For comparison, we also calculate the partition function of a random coil. We compare the free energies of these conformations as functions of temperature. By identifying the one with least free energy as the equilibrium conformation, we find the critical temperatures at which phase transitions occur. With choices restricted to pure α helix, pure β sheet, or random coil, we find that the α helix has the lowest free energy at low temperatures. It makes a transition to the β sheet at 475 K, and the latter makes a transition to the random coil at 600 K. These results agree with those from independent calculations. In a real protein, the critical temperatures may be modified by interactions with the rest of the protein and by the effects of dissipation.

II. MODELING THE PROTEIN CHAIN

The protein chain consists of a sequence of amino acids chosen from a pool of 20. These amino acids all center about a carbon atom called C_α and differ from one another only in the side chains connected to C_α. The side chains are typically represented by hard spheres in our model. When the amino acids are joined into a chain, they become interlocked "residues." From a dynamical point of view, the independent units of the chain are "cranks" made up of coplanar chemical bonds, which connect one C_α to the next, as shown in Fig. 1. The bond lengths and bond angles in a crank are given in Table I [36].

The backbone of the protein chain is thus a sequence of cranks. The angle between two adjacent cranks is fixed at the tetrahedral angle $\cos^{-1}(-1/3) \approx 109.5°$. Thus, the orientation of one crank with respect to its predecessor is specified by two dihedral angles $\{\phi, \psi\}$, as illustrated in Fig. 1.

The conformation of the backbone of the protein is completely specified by a set of dihedral angles $\{\phi_1, \psi_1; \phi_2, \psi_2; \ldots\}$. In this study, we only consider these dihedral degrees of freedom, ignoring the small high-frequency vibrations within the cranks. Such a description has been used in the CSAW model of protein folding [33,34].

III. CANONICAL FORMALISM

Consider a crank consisting of C_α, C, N, O, H, and S (hard-sphere side chain) molecules. For a chain of n cranks,

---

[1]Note that modes with zero eigenvalues have been ignored since they do not relate to the normal frequencies and are found to correspond to the entropic freedom of the protein chain.
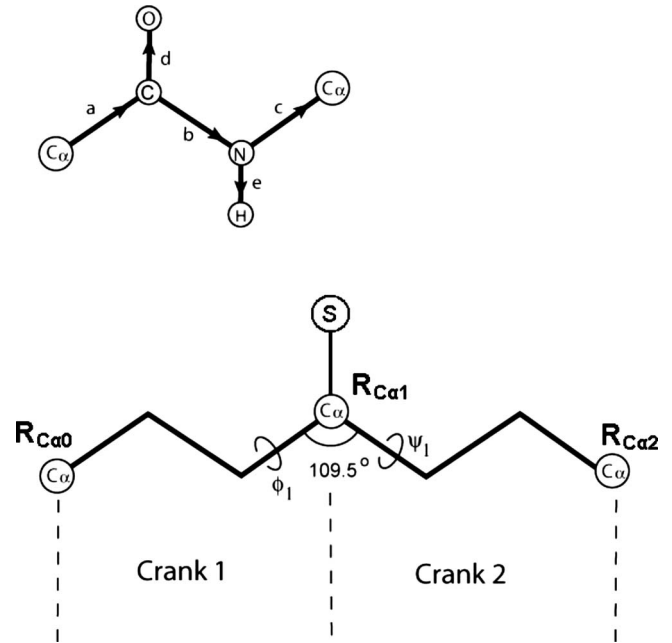

FIG. 1. Upper panel shows the crank that connects the center of one residue to the next. The vectors **a**, **b**, **c**, **d**, and **e** represent chemical bonds, which all lie in the same plane. (Side chains have been omitted for clarity.) Lower panel shows how the cranks are connected to form the backbone of the protein. The vector position of C_α is denoted by $\mathbf{R}_{C\alpha i}$. S denotes the side chain molecule being modeled as a hard sphere. The bond length of hard-sphere side chain can be obtained from [35]. The angle between cranks is fixed; the relative orientation of successive cranks is specified by two dihedral angles ϕ and ψ. The conformation of the backbone chain is completely specified by a set of dihedral angles. Data for bond lengths and angles are given in Table I.

this gives a total of $6n$ atoms. We label each atom by an index j. Let

$$\mathbf{R}_j = \text{position vector of atom } j,$$

$$j = 1, \ldots, 6n. \tag{1}$$

Let the set of dihedral angles be $\{\phi_i, \psi_i\}$ $(i = 1, \ldots, n-1)$ which makes up a total of $2(n-1)$ degrees of freedom. Assuming the cranks to be perfectly rigid, it places constraints on the vector positions. The constraints are solved by using

TABLE I. Bond lengths and bond angles. Data are obtained from Protein Data Bank website [36].

	Bond length (Å)		Bond angle (deg)
C_α—C	1.525	$\angle C_\alpha CN$	116.2
C—N	1.329	$\angle CNC_\alpha$	121.7
N—C_α	1.458	$\angle NC_\alpha C$	109.5
C—O	1.231	$\angle C_\alpha CO$	120.8
N—H	1.000	$\angle C_\alpha NH$	114.0

the dihedral angles as generalized coordinates, which we denote by the notation

$$q_{ik} \quad (i = 1, \ldots, n-1; \quad k = 1, 2). \tag{2}$$

Thus, for example,

$$q_{11} = \phi_1, \quad q_{12} = \psi_1,$$

$$q_{21} = \phi_2, \quad q_{22} = \psi_2, \text{ etc.} \tag{3}$$

We also use the notation q_α, where $\alpha = \{i, k\}$. We are to regard $\mathbf{R}_j$ as functions of $\{q_\alpha\}$.

The velocity is given by

$$\dot{\mathbf{R}}_j = \sum_{i=1}^{n-1} \sum_{k=1}^{2} \frac{\partial \mathbf{R}_j}{\partial q_{ik}} \dot{q}_{ik} = \sum_{\alpha} \frac{\partial \mathbf{R}_j}{\partial q_\alpha} \dot{q}_\alpha. \tag{4}$$

The total kinetic energy is

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

$$= \frac{1}{2} \dot{q}^T M \dot{q}, \tag{5}$$

where the mass matrix is given by

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

This is a symmetric matrix, with $M^T = M$.

The Lagrangian of the backbone chain is given by

$$L(q, \dot{q}) = K(q, \dot{q}) - U(q) = \frac{1}{2} \dot{q}^T M \dot{q} - U(q). \tag{7}$$

Hence, the canonical momentum is

$$p = \frac{\partial L}{\partial \dot{q}} = M \dot{q} \tag{8}$$

and the generalized force is

$$\frac{\partial L}{\partial q} = \frac{1}{2} \dot{q}^T \frac{\partial M}{\partial q} \dot{q} - \frac{\partial U}{\partial q}. \tag{9}$$

The Lagrange equation of motion,

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \right) = \frac{\partial L}{\partial q}, \tag{10}$$

leads to

$$M \ddot{q} = -\frac{1}{2} \dot{q}^T \frac{\partial M}{\partial q} \dot{q} - \frac{\partial U}{\partial q}. \tag{11}$$

The Hamiltonian is

$$H(p, q) = K(p, q) + U(q) = \frac{1}{2} \dot{q}^T M \dot{q} + U(q). \tag{12}$$

From Eq. (8) we have $\dot{q} = M^{-1} p$ and $\dot{q}^T = p^T (M^{-1})^T$. Therefore,

$$H(p, q) = \frac{1}{2} p^T M^{-1} p + U(q). \tag{13}$$

The canonical equations of motion,

$$\dot{p} = -\frac{\partial H}{\partial q}, \quad \dot{q} = \frac{\partial H}{\partial p}, \tag{14}$$

take the forms

$$\dot{p} = -\frac{1}{2} p^T \frac{\partial M^{-1}}{\partial q} p - \frac{\partial U}{\partial q},$$

$$\dot{q} = M^{-1} p. \tag{15}$$

These are, of course, the same as the Lagrangian equation of motion as given by Eq. (11).

IV. EFFECTIVE POTENTIAL

The partition function of the system is, up to a constant scale factor, given by

$$Z = \int \int e^{-\beta[K(p,q)+U(q)]} d\vec{p} \, d\vec{q}, \tag{16}$$

where $\beta = (k_B T)^{-1}$ is the inverse temperature. The p integration is Gaussian and can be immediately carried out, and the result generally depends on q,

$$Z = \int \left[\int e^{-\beta K(p,q)} d\vec{p} \right] e^{-\beta U(q)} d\vec{q} \tag{17}$$

$$= \int \left[\int e^{-\beta(1/2)p^T M^{-1} p} dp^{2(n-1)} \right] e^{-\beta U(q)} d\vec{q}$$

$$= \left(\frac{2\pi}{\beta} \right)^{n-1} \int \sqrt{\det M} \, e^{-\beta U(q)} d\vec{q}. \tag{18}$$

This gives rise to an effective potential $V_{\text{eff}}(q)$, which is defined through the relation

$$e^{-\beta V_{\text{eff}}(q)} \equiv \left(\frac{\beta}{2\pi} \right)^{n-1} \int e^{-\beta K(p,q)} dp^{2(n-1)}. \tag{19}$$

Thus,

$$Z = \int \rho_{\text{con}}(q) d\vec{q},$$

$$\rho_{\text{con}}(q) \equiv \left(\frac{2\pi}{\beta} \right)^{n-1} e^{-\beta(U+V_{\text{eff}})}, \tag{20}$$

where ρ_{con} is the configurational probability density. That is, $\rho_{\text{con}} dq$ is the relative probability of finding the system in dq, regardless of the momentum p. If the kinetic energy is independent of q, the effective potential is a constant.

In a canonical ensemble, the relative probability of finding the state in the element $d\vec{p} d\vec{q}$ in phase space is given by $d\vec{p} d\vec{q} \exp(-\beta H)$. If we are only interested in the probability of finding the state in dq, we integrate the above over p and obtain
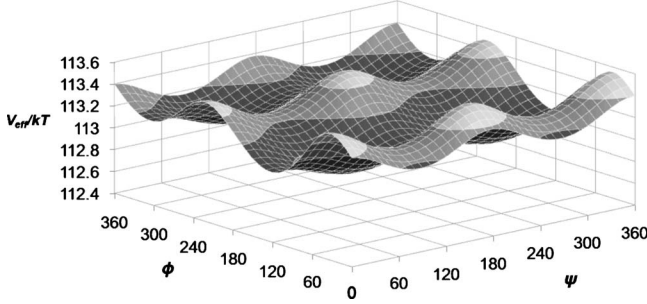
FIG. 2. Effective potential of a three-crank chain at different $\{\phi, \psi\}$ angles between the second and third cranks. The percentage change of the effective potential is less than 0.2%.

$$d\vec{q} \int d\vec{p} \exp(-\beta H) = d\vec{q} \exp(-\beta U) \int d\vec{p} \exp(-\beta K)$$
$$= d\vec{q} \left(\frac{2\pi}{\beta} \right)^{n-1} \exp[-\beta(U + V_{\mathrm{eff}})]. \tag{21}$$

This is the probability to be used, for example, in the Monte Carlo algorithm in the CSAW model [33,34]. From Eqs. (18) and (19), we have

$$\beta V_{\mathrm{eff}}(q) = -\tfrac{1}{2}\ln(\det M) = -\tfrac{1}{2}\mathrm{Tr}(\ln M). \tag{22}$$

Since $V_{\mathrm{eff}}(q)$ depends on all the dihedral angles, it is a function of the chain conformation. Our calculations show that it is sensibly constant for almost all different conformations. The percentage change of the effective potential for a three-crank model is less than 0.2%. Hence, we shall assume that it is a constant in our subsequent derivation. Representative results are shown in Fig. 2.

V. POTENTIAL ENERGY OF HYDROGEN BONDING

In our model, the main stabilizing agents for protein secondary structures such as α helix and β sheet are hydrogen bonds, which exist between N—H and C=O groups from different residues [37,38]. 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 N—H and C=O is $180° \pm 45°$ [37].

The α helix, also known as the 4_{13} helix, is the most abundant secondary structure due to its tight conformation [38]. In this configuration, a hydrogen bond connects the C=O group of the ith crank to the N—H group of the $(i+3)$th crank.

The β sheet is a two-dimensional mat made up of backbone strands stitched together by hydrogen bonds [38]. The participating strands may be parallel or antiparallel. In this paper, we will only study the latter case.

We wish to study the normal modes of small vibrations about an equilibrium configuration. The potential energy U is assumed to be minimum, and taken to be zero, at this configuration. The equilibrium is assumed to be maintained by hydrogen bonds. Deviations from equilibrium arise from the stretching and bending of these bonds. Let $\mathbf{b}_i$ be the bond

vector of the ith hydrogen bond, i.e., the vector between O and the bonded H, in the equilibrium situation. Let $\mathbf{b}_i'$ be the same vector when the configuration is displaced from equilibrium. The displacement vector is given by

$$\mathbf{u}_i = \mathbf{b}_i' - \mathbf{b}_i. \tag{23}$$

For small displacements, we take the potential energy to be

$$U = \frac{1}{2}\kappa_1 \sum_i (|\hat{\mathbf{b}}_i \cdot \mathbf{u}_i|)^2 + \frac{1}{2}\kappa_2 \sum_i (|\hat{\mathbf{b}}_i \times \mathbf{u}_i|)^2, \tag{24}$$

where $\hat{\mathbf{b}}_i = \mathbf{b}_i / |\mathbf{b}_i|$, and κ_1 and κ_2 are the force constants associated with the stretching and bending of hydrogen bonds, respectively [39],

$$\kappa_1 = 13 \ \mathrm{N/m},$$

$$\kappa_2 = 3 \ \mathrm{N/m}. \tag{25}$$

Let the generalized coordinates be denoted by

$$q = q_0 + \lambda, \tag{26}$$

where q_0 corresponds to equilibrium and λ represents a small deviation. We can write

$$\mathbf{u}_i = \sum_\alpha \left(\frac{\partial \mathbf{b}_i'}{\partial q_\alpha} \right)_0 \lambda_\alpha + O(\lambda^2), \tag{27}$$

where the subscript 0 indicates evaluation at equilibrium. This leads to the quadratic form

$$U = \frac{1}{2}\lambda^T(\kappa_1 D + \kappa_2 C)\lambda, \tag{28}$$

$$D_{\alpha\beta} = \sum_i \left| \hat{\mathbf{b}}_i \cdot \frac{\partial \mathbf{b}_i'}{\partial q_\alpha} \right|_0 \cdot \left| \hat{\mathbf{b}}_i \cdot \frac{\partial \mathbf{b}_i'}{\partial q_\beta} \right|_0, \tag{29}$$

$$C_{\alpha\beta} = \sum_i \left| \hat{\mathbf{b}}_i \times \frac{\partial \mathbf{b}_i'}{\partial q_\alpha} \right|_0 \cdot \left| \hat{\mathbf{b}}_i \times \frac{\partial \mathbf{b}_i'}{\partial q_\beta} \right|_0. \tag{30}$$

VI. NORMAL MODES

For small oscillations about equilibrium, the linearized equation of motion is

$$M\ddot{\lambda} = -\frac{\partial U}{\partial q}. \tag{31}$$

From Eq. (28) we have

$$\frac{\partial U}{\partial q} = (\kappa_1 D + \kappa_2 C)\lambda. \tag{32}$$

Thus,

$$M\ddot{\lambda} + (\kappa_1 D + \kappa_2 C)\lambda = 0. \tag{33}$$

The normal frequencies ω and normal modes λ are eigenvalues and eigenvectors of the equation
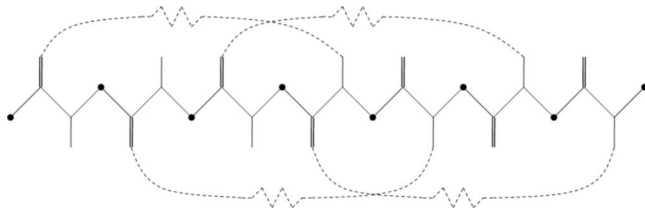
FIG. 3. The mechanical system corresponding to small oscillations of the α helix (solid lines). Springs are hydrogen bonds (dashed lines).

$$M^{-1}(\kappa_1 D + \kappa_2 C)\lambda = \omega^2 \lambda. \tag{34}$$

Our model's validity is subject to the following conditions:

(1) We treat small oscillation about a presumed equilibrium configuration q_0. Whether q_0 indeed corresponds to equilibrium can be verified through the requirement that all normal frequencies are nonzero and positive.

(2) We ignore electrostatic and other interactions. Our results can serve as a test on whether the structures investigated can maintain equilibrium purely through hydrogen bonding. Inclusion of other interactions will introduce corrections.

(3) Actual α and β structures are embedded inside a protein molecule in solution and are subject to other forces not considered here, particularly those arising from Brownian motion in the solution, the hydrophobic effect, and interaction with other atoms in the protein. These forces will give rise to corrections and may even destroy the stability of the structure.

In view of the limitations of the model, we only examine normal modes in a frequency range corresponding to wave numbers $10^{-1} - 10^3$ cm^{-1}. This is because, in a real protein, the very low-frequency end will be dominated by binding effects to the rest of the protein, while the very high-frequency region will be dominated by bond oscillations.

VII. α HELIX

In this study, we have taken polyalanine as our model for a generic α helix. Since the CH$_3$ side chain group of alanine residues is not a big molecule, it is easily expressed in terms of a fully atomic model instead of using the implicit hard-sphere model. Formally, it is straightforward to generalize this all-atom side chain model.

The dihedral angles ϕ, ψ of polyalanine [40] are given by

$$\{\phi, \psi\} = \{-57.4°, -47.5°\}. \tag{35}$$

The equivalent spring system is illustrated in Fig. 3. In this example, there are seven cranks, but only four hydrogen bonds. In general, for n cranks, the number of hydrogen bonds is $n-3$. The number of degrees of freedom from stretching and bending of the hydrogen bonds is thus $2(n-3)$. The total number of degrees of freedom of the system, however, is $2(n-1)$. Thus, we expect to have four zero modes, apart from rigid translations and rotations. These will not be included in our results.

Figure 4 shows the distributions of normal modes as a function of wave number, for different crank numbers n. All calculated frequencies are positive. The distributions exhibit four peaks associated with various types of deformation, which can be ascertained by examining the corresponding eigenvectors. The results are listed in Table II.

VIII. β SHEET

We model a generic antiparallel β sheet [41] by setting the dihedral angles in each strand to

$$\{\phi, \psi\} = \{-139°, 135°\}. \tag{36}$$

The connectivity of hydrogen bonds for the antiparallel β sheet is shown in Fig. 5. An extra crank is included to join two adjacent strands.

Compared to the α helix, the β sheet has fewer hydrogen bonds formed within the structure. Thus, we expect that in our model there will be more zero modes compared to the α
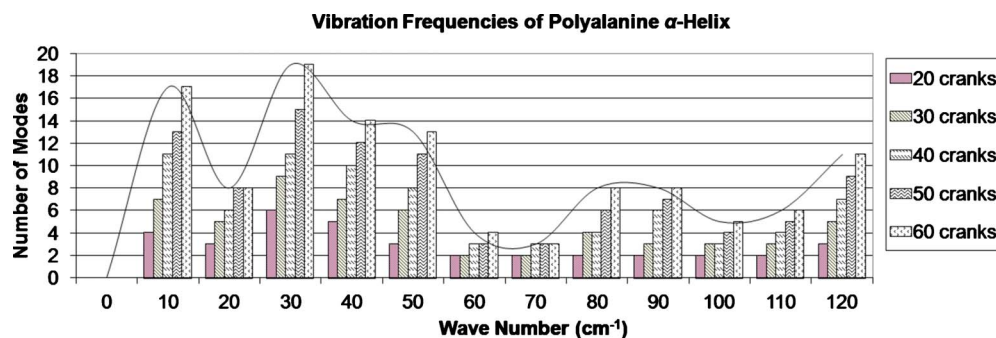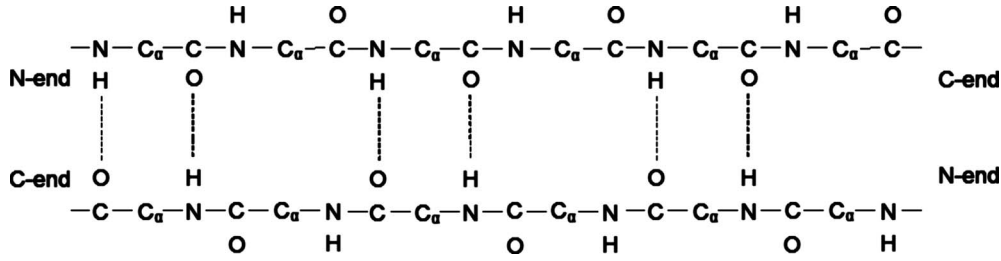
TABLE II. Normal modes of α helix.

Frequency (cm^{-1})	Mode
0–10	Twisting
30–40	Stretching and bending
90–100	Bending
120–130	Bending

FIG. 4. (Color online) Normal-frequency distributions for the α helix, with different numbers of cranks n. Types of distortion corresponding to the peaks are listed in Table II.

FIG. 5. Schematic diagram of antiparallel β sheet, illustrating the connectivity of hydrogen bonds.

helix; but we ignore them for reasons stated previously. Otherwise, all calculated frequencies are positive.

Normal frequencies are computed for varying numbers of strands, and cranks per strand. We display representative distributions in Fig. 6. We see that the frequencies are dominating the far-infrared frequency region. This is consistent with calculations on real protein with β-sheet structure [2,41–43]. In general, the peak positions of the distributions depend only on the number of cranks per strand and are independent of the number of strands. The peaks tend to widen with increasing crank number.

IX. COMPARISON WITH OTHER RESEARCH WORK

Our calculation of normal modes for both α helix and β sheet is in agreement with results calculated by ben-Avraham [1]. ben-Avraham found that the density of states of globular proteins follows a characteristic universal curve (c.f. Fig.1 of [1]). We found similar behavior for both the α helix and β sheet. The number of modes for these two secondary structures converges to their respective characteristic curves that are independent of the number of cranks for the α helix and the number of strands for the β sheet. In addition, our results show that the main contribution to the normal modes of these secondary structures comes from physical effects due to the hydrogen bonds.

In order to further validate our results, we obtain the normal-mode distribution of myoglobin (1MBD) [44] (see Fig. 7), which is made up of eight α helices. We found that our result is in good agreement with those obtained by Krimm and Reisdorf, Jr. (c.f. Fig. 7(b) of [40]), who derived their normal-mode distribution from a different approach.

X. α-β TRANSITION

The transition between α helix and β sheet is an important subject in view of the existence of proteins with ambiva-

lent structures [45]. From our results, we now proceed to making the following calculation. Based on Eqs. (13) and (28), we can write the partition function as follows:

$$
\begin{aligned}
Z_i &= \int \exp\{-\beta[K_i(p) + U_i(q) + U_i^e]\}d\vec{p}d\vec{q} \\
&= \int \exp\left\{-\beta\left[\frac{1}{2}p^T M_i^{-1} p + \frac{1}{2}q^T(\kappa_1 D_i + \kappa_2 C_i)q \right.\right. \\
&\qquad\qquad \left.\left. + U_i^e\right]\right\}d\vec{p}d\vec{q},
\end{aligned}
\tag{37}
$$

with the index i being α, β, or c depending on whether the configuration is a α helix, a β sheet, or a random coil, respectively. Note that U_i^e is the total chemical potential of the hydrogen bonds formed within a particular polypeptide configuration. U_i^e can also be viewed as the energy minima of the potential well and, since it is independent of the dihedral angles, it does not affect the earlier normal-mode analysis.
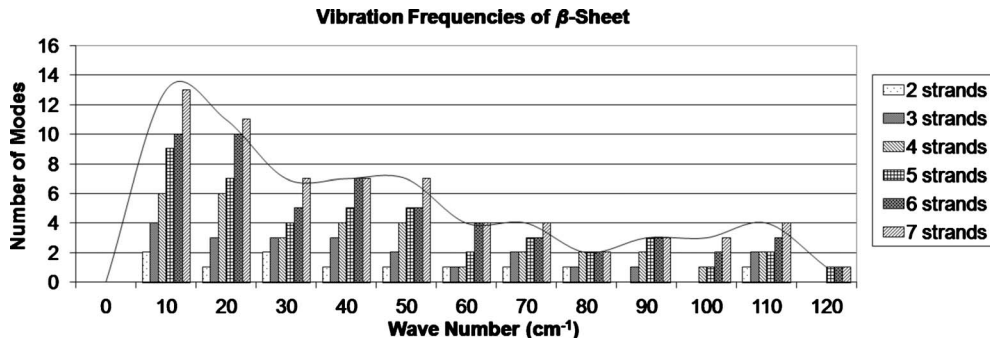
After evaluating the integral in Eq. (37), the partition function takes the following form:

$$
\begin{aligned}
Z_i &= \frac{(2\pi k_B T)^{n-1}}{\sqrt{\det M_i^{-1}}}\frac{(2\pi k_B T)^{N_i}}{\sqrt{\det_p(\kappa_1 D_i + \kappa_2 C_i)}}(2\pi)^{2(n-1)-2N_i}e^{-\beta U_i^e} \\
&= \frac{(2\pi k_B T)^{n-1}}{(\bar{\lambda}_i^M)^{2(n-1)}}\frac{(2\pi k_B T)^{N_i}}{(\bar{\lambda}_i^K)^{2N_i}}(2\pi)^{2(n-1)-2N_i}e^{-\beta U_i^e},
\end{aligned}
\tag{38}
$$

where

$$
(\bar{\lambda}_i^M)^{2(n-1)} = (\omega_i^{M1}\omega_i^{M2}\omega_i^{M3}\cdots\omega_i^{M2(n-1)})^{1/2},
$$

$$
(\bar{\lambda}_i^K)^{2N_i} = (\omega_i^{K1}\omega_i^{K2}\omega_i^{K3}\cdots\omega_i^{K2N_i})^{1/2}.
\tag{39}
$$

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. In Eq. (38), n is the

FIG. 6. Normal-frequency distributions of β sheets, for the same number of cranks per strand, but different number of strands.
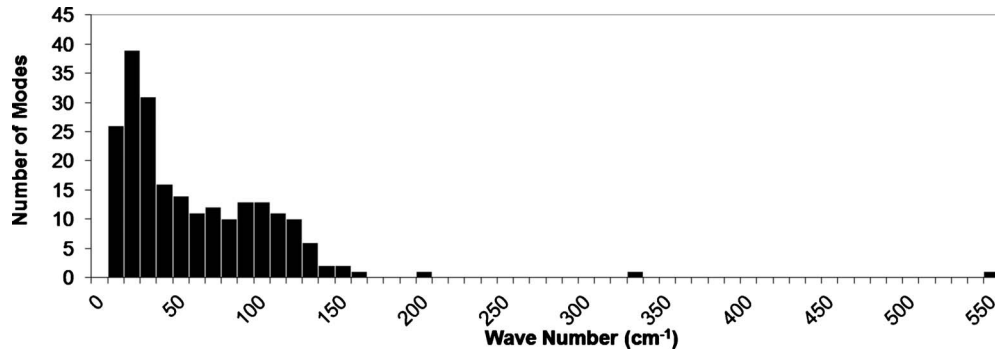
FIG. 7. Normal-frequency distribution for myoglobin, consisting of eight α helices.

number of cranks, while N_i is the number of hydrogen bonds in the configuration. Since N_i is less than n, zero eigenvalues are to be expected in the potential matrix $(\kappa_1 D_i + \kappa_2 C_i)$. Hence, we have a "pseudo" determinant term $\det_p(\ )$ which takes in only the nonzero eigenvalues.

Next, we evaluate the Helmholtz free energy of a particular protein configuration from Eq. (38),

$$F_i = -k_B T \ln Z_i = U_i - k_B T S_i, \tag{40}$$

where

$$S_i = (n-1)\ln(2\pi k_B T) - 2(n-1)\ln \bar{\lambda}_i^M + N_i \ln(2\pi k_B T)$$
$$- 2N_i \ln \bar{\lambda}_i^K + N_i^z \ln(2\pi). \tag{41}$$

We observe that the zero modes now play a significant role in the determination of the free energy of the protein configuration. In other words, the zero modes are directly related to the entropy of the protein configuration with $N_i^z = 2(n-1) - 2N_i$ zero modes giving rise to N_i^z multiples of 2π entropy. Furthermore, the third and fourth terms on the right-hand side of Eq. (41) give the competition between the entropic effects of the environment and the internal stretching and bending energy of the hydrogen bonds, respectively. These results imply that the smaller the number of hydrogen bonds in the protein configuration, the higher the entropy. Since a protein in a random coil configuration has no hydrogen bond (i.e., $N_i = 0$), it has the highest entropy.

Figure 8 shows the Helmholtz free energy of a 15-crank polyalanine in the α helix, in the two-strand β hairpin, and the random coil configuration as a function of temperature. The α helix has been taken as a reference state in the figure in the following way:

$$F_i = (U_i - U_\alpha) - k_B T(S_i - S_\alpha). \tag{42}$$

Note that the values of the quantities used in the plot are given in Table III, with $U_i^e = -N_i \epsilon_{hb}$, where $\epsilon_{hb} = 5$ kcal/mol is the potential energy of a hydrogen bond. The plot shows that the random coil has the steepest curve, followed by the β hairpin and then the α helix. This is to be expected since without the constraint of hydrogen bonds, the random coil has the highest entropy. At the other extreme, the structure of the α helix is stabilized by 12 hydrogen bonds, six more than the β hairpin. Hence, it has the least configurational freedom and the flattest curve of the three.

The critical temperature at which protein phase transition occurs can be determined through the free-energy curves in Fig. 8. For example, by examining the intersection between the free-energy curves of the α helix and the random coil, the critical temperature of 535 K is observed for the α-coil transition. Similarly, the critical temperatures for the β-coil and α-β transitions are found to occur at 600 and 475 K, respectively. As a protein adopts its stable configuration by minimizing its free energy, we expect the 15-crank polyalanine to form an α-helix structure at low temperature. As temperature increases, entropy begins to gain importance against the dominance of the internal energy of the protein at low temperature. At the α-β transition temperature, the more en-
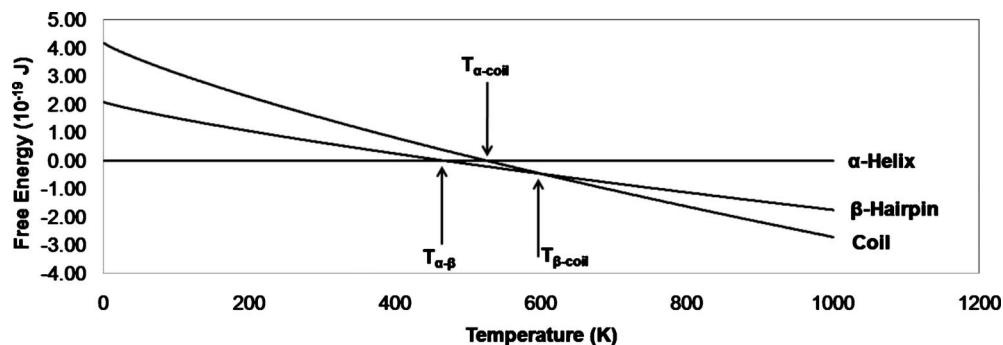

FIG. 8. Free-energy curves of α helix, β sheet, and random coil. The intersection of the curves indicates phase transition. The transition temperatures are found to be $T_{\alpha\beta} = 475$ K, $T_{\alpha c} = 535$ K, and $T_{\beta c} = 600$ K.

TABLE III. Quantities employed in the calculation of the free energy versus temperature curve for the α helix, β sheet, and random coil configurations.

i	$U_i^e \times 10^{-19}$ (J)	N_i	$\bar{\lambda}_i^M \times 10^{22}$ (kg$^{-1/2}$ m^{-1})	$\bar{\lambda}_i^K \times 10^{-10}$ (kg$^{1/2}$ m s^{-1})
α helix	−4.17	12	1.165	5.878
β hairpin	−2.09	6	1.190	6.960
Random coil	0.00	0	1.714	

tropic β hairpin emerges as the more stable structure with a lower free energy. A further increase in temperature will eventually lead to the critical state of β-coil transition, upon which the β hairpin denatures into a random coil.

Interestingly, our results correspond to those obtained by Ding *et al.* [46] through molecular-dynamics simulations. However, we notice that they have overestimated their degrees of freedom in their analytical estimates. Since the degree of freedom in their model is also based on the ϕ and ψ dihedral angles, the total degree of freedom is $2N$ for a pair of dihedral angles per residue, with N being the number of residues. Thus, a more reasonable entropy calculation should be $S_i = A \ln\langle$rms $d_r\rangle_i + S_0$ with $A = 2N$ instead of $A = 3N$. Note that $\langle$rms $d_r\rangle_i$ is the expected root-mean-square deviation for configuration i from a chosen reference structure [46]. Remarkably, this correction leads to results which corroborate with ours in Fig. 8 (see Fig. 9), with α-β, α-coil, and β-coil transition temperatures of 427, 504, and 616 K, respectively, by employing the data given in [46]. Further support of our approach is given by Yasar and Demir for a hydrophobic homogeneous polypeptide chain with a helix-coil transition temperature of 550 K [47]. Our result is also in agreement with computational work performed by Lee *et al.* [48], whose analysis has found a helix-coil transition temperature of about 475 K for a 15-residue polyalanine (refer to Fig. 2(a) of [48]).

Finally, it should be noted that by assuming a simple hydrogen bond potential, we have considered the polyalanine to be in the gas phase in this paper. However, if we were to include the effects of solvent, we would expect a lowering of the Helmholtz free energies as the polyalanine folds toward a new optimal conformation [47]. Indeed, if the solvent is water, the new conformation is non-α helical since polyalanine is a purely hydrophobic polypeptide.

XI. CONCLUSION

In this paper, we have formulated the mechanics of protein chains in terms of the Hamiltonian formalism. We have applied this formalism to determine the normal modes of the α helix and the β sheet, and the phase transition of these secondary structures. By modeling the protein as a sequence of cranks and making small deviations from the equilibrium potential of the hydrogen bonds, we have obtained normal-mode distributions that correspond to those computed based on classical NMA [3]. This has served to validate our formalism. Unlike these prior models that employed a complete range of force fields, results from our simplified model reveal that the slow modes observed in the α helices and β sheets are mainly attributed to the vibrational effects of the hydrogen bonds in these structures.

Finally, we have pursued the subject of protein phase transition from a different perspective by formulating the partition function and the Helmholtz free energy in terms of results obtained from our canonical formalism and normal-mode analysis. Remarkably, our analytical results on the free energy for a 15-crank polyalanine have reaffirmed previous numerical prediction [46] of an α-β followed by a β-coil transition, as temperature is increased. While we had performed our analysis in the gas phase, which leads to a high critical temperature of 475 K for the α-β transition, we expect such a transition to occur under normal physiological conditions if there are present enzymatic influences in a hydrated environment. We plan to investigate deeper on such enzymatic effects in our future research. Our model could add on to the existing studies of structural transition in protein secondary structures [31,32,45–48], which we believe could lead us to another step closer to an understanding on the mechanism of both helix-coil and α-β transitions in protein.
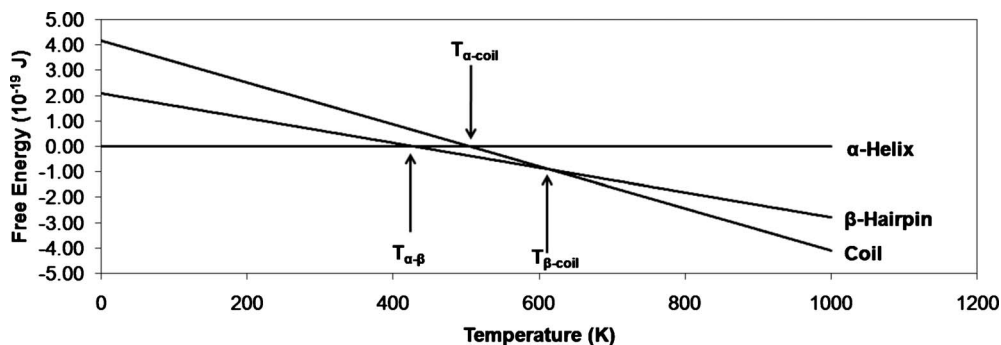

FIG. 9. Modification of the free-energy curves of Ding *et al.* [46] based on $S_i = 2N \ln\langle$rms $d_r\rangle_i + S_0$. The transition temperatures are found to be $T_{\alpha\beta} = 427$ K, $T_{\alpha c} = 504$ K, and $T_{\beta c} = 616$ K.

[1] D. ben-Avraham, Phys. Rev. B **47**, 14559 (1993).
[2] Y. Abe and S. Krimm, Biopolymers **11**, 1817 (1972).
[3] D. ben-Avraham and M. M. Tirion, Physica A **249**, 415 (1998).
[4] B. R. Brooks and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **80**, 6571 (1983).
[5] M. Levitt, C. Sander, and P. S. Stern, J. Mol. Biol. **181**, 423 (1985).
[6] D. A. Case, Curr. Opin. Struct. Biol. **4**, 285 (1994).
[7] A. Kitao and N. Gō, Curr. Opin. Struct. Biol. **9**, 164 (1999).
[8] F. Tama, F. X. Gadea, O. Marques, and Y. H. Sanejouand, Proteins **41**, 1 (2000).
[9] N. Gō, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).
[10] N. Gō, T. Noguti, and T. Nishikawa, Proc. Natl. Acad. Sci. U.S.A. **80**, 3696 (1983).
[11] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[12] W. J. Zheng and B. R. Brooks, Biophys. J. **88**, 3109 (2005).
[13] A. B. Besya, H. Mobasheri, and M. R. Ejtehadi, Phys. Rev. E **81**, 051911 (2010).
[14] I. Bahar, A. R. Atilgan, and B. Erman, Folding Des. **2**, 173 (1997).
[15] T. Haliloglu, I. Bahar, and B. Erman, Phys. Rev. Lett. **79**, 3090 (1997).
[16] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, Biophys. J. **80**, 505 (2001).
[17] S. Takada, Proc. Natl. Acad. Sci. U.S.A. **96**, 11698 (1999).
[18] F. Tama, Protein and Peptide Letters **10**, 119 (2003).
[19] A. Kidera and N. Gō, Proc. Natl. Acad. Sci. U.S.A. **87**, 3718 (1990).
[20] M. M. Tirion, D. Benavraham, M. Lorenz, and K. C. Holmes, Biophys. J. **68**, 5 (1995).
[21] I. Bahar, B. Erman, R. L. Jernigan, A. R. Atilgan, and D. G. Covell, J. Mol. Biol. **285**, 1023 (1999).
[22] I. Bahar and R. L. Jernigan, Biochemistry **38**, 3478 (1999).
[23] R. Brüschweiler and D. A. Case, Phys. Rev. Lett. **72**, 940 (1994).
[24] S. Sunada, N. Gō, and P. Koehl, J. Chem. Phys. **104**, 4768 (1996).
[25] P. Doruker, R. L. Jernigan, and I. Bahar, J. Comput. Chem. **23**, 119 (2002).
[26] F. Englert, K. Peeters, and A. Taormina, Phys. Rev. E **78**, 031908 (2008).
[27] S. B. Prusiner, Proc. Natl. Acad. Sci. U.S.A. **95**, 13363 (1998).
[28] Y. Peng and U. H. E. Hansmann, Phys. Rev. E **68**, 041911 (2003).
[29] A. V. Yakubovich, I. A. Solov'yov, A. V. Solov'yov, and W. Greiner, Eur. Phys. J. D **40**, 363 (2006).
[30] A. V. Yakubovich, I. A. Solov'yov, A. V. Solov'yov, and W. Greiner, Eur. Phys. J. D **46**, 215 (2008).
[31] L. Hong and J. Lei, Phys. Rev. E **78**, 051904 (2008).
[32] B. H. Zimm and J. K. Bragg, J. Chem. Phys. **31**, 526 (1959).
[33] K. Huang, Biophys. Rev. Lett. **3**, 1 (2008).
[34] K. Huang, Biophys. Rev. Lett. **2**, 139 (2007)..
[35] J. S. Anderson and H. A. Scheraga, J. Protein Chem. **1**, 281 (1982).
[36] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Res. **28**, 235 (2000); http://www.pdb.org
[37] A. Karshiko, *Non-Covalent Interactions in Proteins* (Imperial College Press, London, 2006).
[38] A. V. Finkelstein and O. B. Ptitsyn, *Protein Physics: A Course of Lectures* (Academic, London, 2002).
[39] K. Itoh and T. Shimanouchi, Biopolymers **9**, 383 (1970).
[40] K. Krimm and W. C. Reisdorf, Jr., Faraday Discuss. **99**, 181 (1994).
[41] W. H. Moore and S. Krimm, Biopolymers **15**, 2465 (1976).
[42] J. Bandekar and S. Krimm, Biopolymers **27**, 909 (1988).
[43] A. M. Dwivedi and S. Krimm, Macromolecules **15**, 186 (1982).
[44] S. E. Phillips and B. P. Schoenborn, Nature (London) **292**, 81 (1981), PDB ID: 1MBD.
[45] S. Patel, P. V. Baleji, and Y. U. Sasidhar, J. Pept. Sci. **13**, 314 (2007).
[46] F. Ding, J. M. Borreguero, S. V. Buldyrev, H. E. Stanley, and N. V. Dokholyan, Proteins **53**, 220 (2003).
[47] F. Yasar and K. Demir, Comput. Phys. Commun. **175**, 604 (2006).
[48] M. S. Lee, G. G. Wood, and D. J. Jacobs, J. Phys.: Condens. Matter **16**, S5035 (2004).