

The helix-coil transition revisited

Yantao Chen,¹ Yaoqi Zhou,^{2,3} and Jiandong Ding^{1*}

¹ Department of Macromolecular Science, Key Laboratory of Molecular Engineering of Polymers of Ministry of Education, Advanced Materials Laboratory, Fudan University, Shanghai 200433, China

² Department of Physiology and Biophysics, Howard Hughes Medical Institute Center for Single Molecule Biophysics, State University of New York at Buffalo, Buffalo, New York 14214

³ Indiana University School of Informatics and Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Walker Plaza, Indianapolis, Indiana 46202

ABSTRACT

In this article, we perform a dynamic Monte Carlo simulation study of the helix-coil transition by using a bond-fluctuation lattice model. The results of the simulations are compared with those predicted by the Zimm-Bragg statistical thermodynamic theory with propagation and nucleation parameters determined from simulation data. The Zimm-Bragg theory provides a satisfactory description of the helix-coil transition of a homopolymer chain of 32 residues ($N = 32$). For such a medium-length chain, however, the analytical equation based on a widely-used large- N approximation to the Zimm-Bragg theory is not suitable to predict the average length of helical blocks at low temperatures when helicity is high. We propose an analytical large-eigenvalue (λ) approximation. The new equation yields a significantly improved agreement on the average helix-block length with the original Zimm-Bragg theory for both medium and long chain lengths in the entire temperature range. Nevertheless, even the original Zimm-Bragg theory does not provide an accurate description of helix-coil transition for longer chains. We assume that the single-residue nucleation of helix formation as suggested in the original Zimm-Bragg model might be responsible for this deviation. A mechanism of nucleation by a short helical block is proposed by us and provides a significantly improved agreement with our simulation data.

Proteins 2007; 69:58–68.
© 2007 Wiley-Liss, Inc.

Key words: Zimm–Bragg theory; helix–coil transition; dynamic Monte Carlo simulation; lattice model; nucleation; protein folding.

INTRODUCTION

α -helix, first proposed by Pauling and coworkers in 1950,^{1,2} is one of the basic structural elements in proteins.³ So far, it has been used as a model system for studying the formation of secondary structures in proteins by experimental,^{4–13} theoretical,^{14–25} and simulation methods.^{26–51} Theories of helix-coil transitions have been used for interpreting experimental and simulation results. One of the best-known theories for the helix-coil transition was established by Zimm and Bragg¹⁵ since five decades ago. In this classical theory, each residue is assumed to be either in a coiled or a helical state. The formation of a native hydrogen bond (H-bond) distinguishes a “helical” residue from a “coil” residue, as schematically shown in Figure 1. The theory has two parameters. A nucleation parameter (σ) reflects the degree of difficulty in nucleating a helical residue among coil blocks and a propagation parameter (s) measures the ability of helix propagation at the helix-coil interface.

These two Zimm–Bragg parameters (σ and s) have been considered as the fundamental parameters for describing the helix-coil transition and the tendency of different residues in helix formation.^{5–9} For example, Scheraga and coworkers first put forward a host-guest method, in which the amino acid of interest is incorporated into long random copolymers to measure Zimm–Bragg parameters.⁵ Later, an alternative method was proposed by Baldwin and coworkers, in which the parameters are determined via substitution into short polyalanine helices.⁹ These experimental results revealed different tendencies of helical formations for different residues. The two parameters have also been determined by computer simulations.^{27–32,38,40,46,47,49,51} Recently, the importance of these parameters was reassessed by Baldwin.⁵²

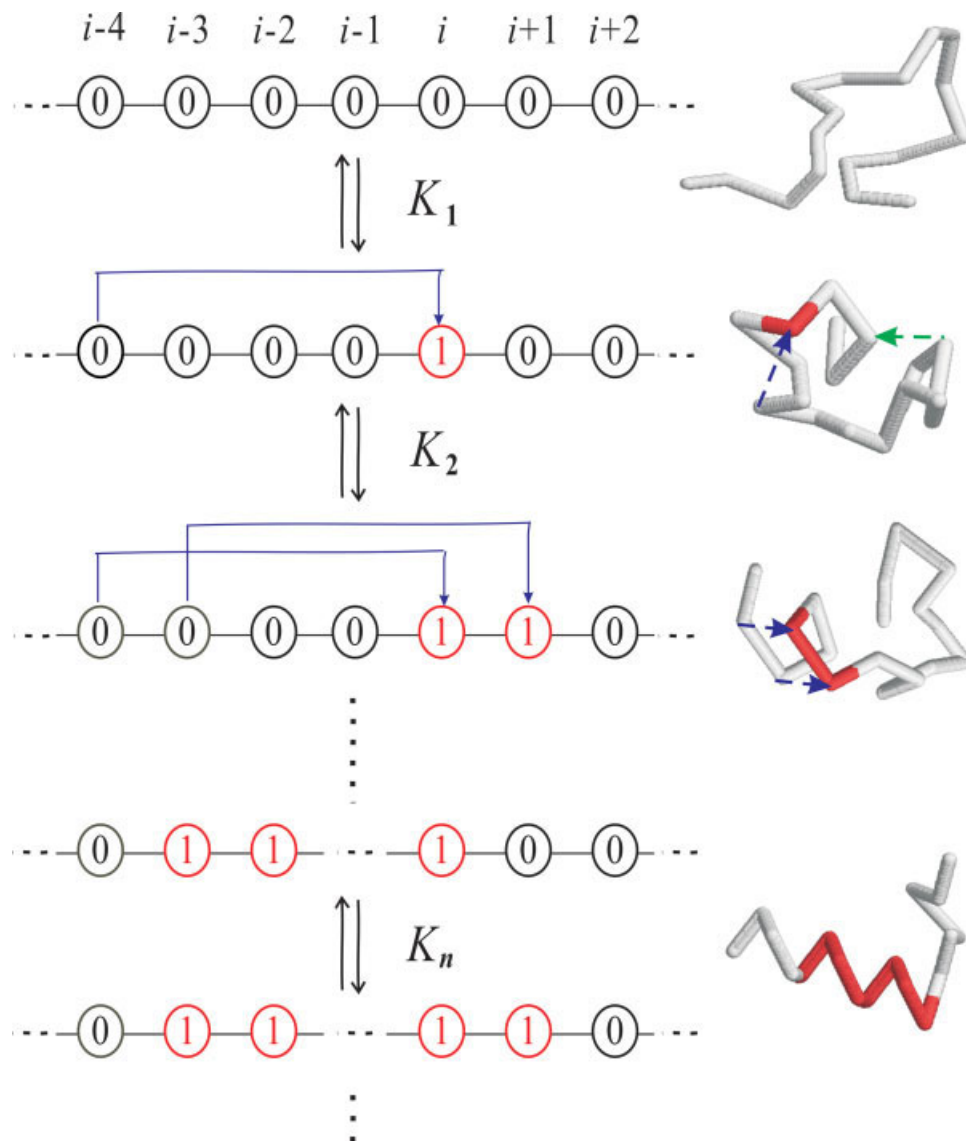
Grant sponsor: NSF of China; Grant numbers: 50533010, 20574013, Two-Base Grant; Grant sponsor: Key Grant of Chinese Ministry of Education; Grant number: 305004; Grant sponsor: 973 Project; Grant number: 2005CB522700; Grant sponsor: Science and Technology Developing Foundation of Shanghai; Grant number: 055207082; Grant sponsor: 863 Project from Chinese Ministry of Science and Technology; Grant sponsor: Innovation Foundation for Graduate Students from Fudan University; Grant number: CQH1717007; Grant sponsor: NIH; Grant numbers: R01 GM 966049; R01 GM 068530

*Correspondence to: Jiandong Ding, Department of Macromolecular Science, Key Laboratory of Molecular Engineering of Polymers of Ministry of Education, Advanced Materials Laboratory, Fudan University, Shanghai 200433, China. E-mail: jdding1@fudan.edu.cn

Received 23 November 2006; Revised 9 February 2007; Accepted 5 March 2007

Published online 27 June 2007 in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/prot.21492

**Figure 1**

Schematic representation of the formation of an α -helix block. Symbols "0" and "1" denote the coiled residue and helical residue, respectively. K_1 , K_2 , and K_n denote the equilibrium constants of the conversion between the coiled state and the helical state of the interface residue between a helical block and a coil block. The associated representative snapshots from our simulations of a homopolyptide are shown on the right side. The helical and coiled residues are colored as red and gray, respectively. A blue arrow denotes a native H-bonding between the two residues that are four residues from each other in the sequence. A green arrow denotes a nonnative H-bonding. Here, residue i is denoted as state "1" (a helical residue) only when a native H-bonding with residue $i - 4$ is formed.

In this work, a dynamic Monte Carlo (MC) simulation has been performed to investigate the helix-coil transition of a self-avoiding lattice chain in the three-dimensional space. Statistical analysis of equilibrium constants for the dynamic processes of helical block formation and disruption allows us to directly calculate σ and s parameters in the simulation^{27,51} in addition to evaluation of helicity θ and average helical block length L . The availability of σ , s , θ , and L from our simulations permits a

detailed comparison between the simulated data and the predictions given by the Zimm-Bragg theory. We show that it is necessary to introduce a short helical block (rather than a single helical residue) for nucleation in order to bridge the gap between theoretical predictions and simulation results for chains longer than 32 residues. Moreover, we find a large-eigenvalue (λ) approximation that leads to a more accurate simple analytic equation of the average length of helix blocks.

METHODS

The Zimm–Bragg theory

The partition function Q in the Zimm–Bragg theory¹⁵ is expressed as⁵³

$$Q = \sum_{k=1}^{N-4} \sum_{j=1}^{\min(k, N-k-4)} \Omega_{j,k} \sigma^j s^k, \quad (1)$$

where $\Omega_{j,k}$ is the total number of ways to have k helical hydrogen bonds in j distinct α -helical blocks. Note that a helical block is made of consecutive unbroken helical residues and $N - 4$ is the maximum number of α -helical hydrogen bonds in a polypeptide chain of length N . In this theory, the nucleation process is the formation of the first α -helical residue (the first native hydrogen bond) among coil blocks and its equilibrium constant K_1 is the product of the nucleation parameter and the propagation parameter. The propagation process, on the other hand, is the formation of the helical residues that are the neighbors of existing helical residues. The equilibrium constants for adding and removing a helical residue from an existing helical block of length n (K_n , $n > 1$) are assumed to be the same regardless the length of existing helical block. So,

$$K_1 = \sigma s, \quad (2)$$

$$K_2 = K_3 = \cdots = K_n = \cdots = s. \quad (3)$$

For this simple model, Q has an analytical form given by

$$Q = \frac{\lambda_0^{N-3}(\lambda_0 - s) - \lambda_1^{N-3}(\lambda_1 - s)}{\lambda_0 - \lambda_1}, \quad (4)$$

where $\lambda_{0,1} = \left[s + 1 \pm \sqrt{(1-s)^2 + 4\sigma s} \right] / 2$, $\lambda_0 > \lambda_1$, λ_0 , and λ_1 are the large and the small eigenvalues of the characteristic Zimm–Bragg matrix ($2^\mu \times 2^\mu$), respectively. Here, $\mu = 1$ for local interactions between two adjacent residues only.¹⁵

Once Q is known, the helicity or the ratio of the number of helical residues in $N - 4$ residues, θ , the average number of α -helical blocks, v , (helical segments separated by coiled residues), and the number average length of α -helical blocks, L , can be obtained from

$$\theta = \frac{1}{N-4} \frac{\partial \ln Q}{\partial \ln s}, \quad (5)$$

$$v = \frac{\partial \ln Q}{\partial \ln \sigma}, \quad (6)$$

and

$$L = (N-4) \frac{\theta}{v} = \frac{s}{\sigma} \frac{\partial \ln Q / \partial s}{\partial \ln Q / \partial \sigma}. \quad (7)$$

Here, Eqs. (5) and (6) correspond to Eqs. (2) and (44), respectively, in the original paper of Zimm and Bragg.¹⁵ Using Eq. (4), one can obtain

$$\theta = \frac{s}{N-4} \left[\frac{\left(\frac{N-3}{\lambda_0} + \frac{1}{\lambda_0-s} \right) \frac{\partial \lambda_0}{\partial s} - \frac{1}{\lambda_0-s}}{1 - \frac{\lambda_1-s}{\lambda_0-s} \frac{\lambda_1^{N-3}}{\lambda_0^{N-3}}} + \frac{\left(\frac{N-3}{\lambda_1} + \frac{1}{\lambda_1-s} \right) \frac{\partial \lambda_1}{\partial s} - \frac{1}{\lambda_1-s} \frac{\partial \lambda_0}{\partial s} - \frac{\partial \lambda_1}{\partial s}}{1 - \frac{\lambda_0-s}{\lambda_1-s} \frac{\lambda_0^{N-3}}{\lambda_1^{N-3}}} - \frac{\partial \lambda_0}{\partial s} - \frac{\partial \lambda_1}{\partial s} \right], \quad (8)$$

$$v = \sigma \left(\frac{\frac{N-3}{\lambda_0} + \frac{1}{\lambda_0-s}}{1 - \frac{\lambda_1-s}{\lambda_0-s} \frac{\lambda_1^{N-3}}{\lambda_0^{N-3}}} \frac{\partial \lambda_0}{\partial \sigma} + \frac{\frac{N-3}{\lambda_1} + \frac{1}{\lambda_1-s}}{1 - \frac{\lambda_0-s}{\lambda_1-s} \frac{\lambda_0^{N-3}}{\lambda_1^{N-3}}} \frac{\partial \lambda_1}{\partial \sigma} - \frac{\partial \lambda_0}{\partial \sigma} - \frac{\partial \lambda_1}{\partial \sigma} \right), \quad (9)$$

$$L = \frac{s}{\sigma} \frac{\frac{\left(\frac{N-3}{\lambda_0} + \frac{1}{\lambda_0-s} \right) \frac{\partial \lambda_0}{\partial s} - \frac{1}{\lambda_0-s}}{1 - \frac{\lambda_1-s}{\lambda_0-s} \frac{\lambda_1^{N-3}}{\lambda_0^{N-3}}} + \frac{\left(\frac{N-3}{\lambda_1} + \frac{1}{\lambda_1-s} \right) \frac{\partial \lambda_1}{\partial s} - \frac{1}{\lambda_1-s} \frac{\partial \lambda_0}{\partial s} - \frac{\partial \lambda_1}{\partial s}}{1 - \frac{\lambda_0-s}{\lambda_1-s} \frac{\lambda_0^{N-3}}{\lambda_1^{N-3}}} - \frac{\partial \lambda_0}{\partial s} - \frac{\partial \lambda_1}{\partial s}}{\frac{\frac{N-3}{\lambda_0} + \frac{1}{\lambda_0-s}}{1 - \frac{\lambda_1-s}{\lambda_0-s} \frac{\lambda_1^{N-3}}{\lambda_0^{N-3}}} \frac{\partial \lambda_0}{\partial \sigma} + \frac{\frac{N-3}{\lambda_1} + \frac{1}{\lambda_1-s}}{1 - \frac{\lambda_0-s}{\lambda_1-s} \frac{\lambda_0^{N-3}}{\lambda_1^{N-3}}} \frac{\partial \lambda_1}{\partial \sigma} - \frac{\partial \lambda_0}{\partial \sigma} - \frac{\partial \lambda_1}{\partial \sigma}}. \quad (10)$$

The equations above are bit complicated. The large- N approximation¹⁵ (i.e., $\ln Q \approx N \ln \lambda_0$) was proposed to simplify the equations and leads to^{15,30,53}

$$\theta = \frac{1}{2} - \frac{1-s}{2\sqrt{(1-s)^2 + 4\sigma s}}, \quad (11)$$

$$L = 1 + \frac{2s}{1-s + \sqrt{(1-s)^2 + 4\sigma s}}. \quad (12)$$

It is of interest to check if the above equations satisfy the high-helicity limit at low temperatures. At this limit, it is very easy to propagate once the nucleation process is completed. That is, $s \gg 1$ and, of course, also $s \gg \sigma$ since $\sigma < 1$ in all cases. According to Eqs. (11) and (12), $\theta \rightarrow 1$ and $L \rightarrow 1 + (s-1)/\sigma$ at this limit. While θ approaches the correct limit, the correct limit for L should be $N - 4$. Thus, the large- N approximation is just appropriate for θ but not for L because $1 + (s-1)/\sigma$ can be significantly greater than N at low temperatures. In the next subsection, we shall make another approximation for simplifying those equations.

The large- λ approximation

To achieve the correct limit for L at $s \gg 1$ and $\gg \sigma$. We introduce a large-eigenvalue (λ) approximation. A naïve way is to neglect all of terms about λ_1 in the

expression of Q . Then, Eq. (4) becomes

$$\ln Q \approx (N - 4) \ln \lambda_0 + \ln(\lambda_0 - s). \quad (13)$$

This leads to [cf. Eq. (6)].

$$\frac{v}{\frac{2\sigma s}{\sqrt{\Delta}}} = \frac{N - 4}{\sqrt{\Delta} + s + 1} + \frac{1}{\sqrt{\Delta} - s + 1}, \quad (14)$$

where

$$\Delta \equiv (1 - s)^2 + 4\sigma s. \quad (15)$$

Combining Eqs. (11) and (14), we obtain

$$L = \frac{N - 4}{N - 3} \left[1 + \frac{2s - 2s/(N - 3)}{\sqrt{(1 - s)^2 + 4\sigma s + 1 - s + 2s/(N - 3)}} \right]. \quad (16)$$

In this equation, $L \rightarrow N - 4$ when $s \gg 1$ and $\gg N\sigma$. The equation can be further reduced to Eq. (12) if $N \gg 3 + (s - 1)/\sigma$ or simply if $N \gg s/\sigma$.

Equations (11) and (16) could also be obtained by a direct simplification of Eqs. (8) and (10) under the approximation $\lambda_0^{N-3} \gg \lambda_1^{N-3}$ with a bit tedious procedure. In fact, the so-called large- N approximation in the literature and the so-called large- λ approximation in this article both contain $\lambda_0 \gg \lambda_1$ and $N \gg 1$. But the large- N approximation is stronger than the large- λ approximation. At low temperatures, $s \gg 1$ and $\sigma \ll 1$. Then, $N \gg s/\sigma$ is only justified for an extremely long chain. As a result, the large- λ approximation is useful for both a medium chain and a long chain, whereas the large- N approximation is just for a very long chain at the high helicity limit.

Nucleation by short helical block

In the Zimm-Bragg theory, nucleation is completed after the formation of the first helical residue and the rate of propagation is the same regardless the length of the initial helical block. We think that it is possible that nucleation requires more than one helical residue. In that case, K_n will depend on the length of the existing helical block (n) at least for the first few helical residues (short helical block). To take into account this possibility, we define

$$K_1 = \sigma_1 s, \quad (17)$$

$$K_2 = \sigma_2 s, \quad (18)$$

$$K_n = \sigma_n s, \quad (19)$$

$$K_{n+1} = K_{n+2} = \cdots = s, \quad (20)$$

where n is the size of the nucleation helix block. The nucleation parameter σ^{block} for helical block is given by

$$\sigma^{\text{block}} = \sigma_1 \sigma_2 \cdots \sigma_n. \quad (21)$$

Here, we extend the Zimm-Bragg theory by using the block nucleation parameter σ^{block} and the propagation parameter s obtained from the equilibrium constant K_i , $i > n$. Equations from the original Zimm-Bragg theory, the large- N approximation, and large- λ approximation can be used to calculate helical ratio and the average length of helical blocks.

Simulation model

In this work, a coarse-grained lattice model is employed to investigate the helix-coil transition. We use a suitable lattice model because it allows efficient calculations while retaining essential physics of the transition.^{36, 54–58} This model is an improved version of a lattice model developed previously by us.⁵¹

We consider a single polypeptide chain of length (N) of 32 amino-acid residues, unless otherwise indicated. The lattice space is composed of $36 \times 36 \times 36$ cubic lattice sites and the periodic boundary condition is set along each dimension. In our model, each residue is represented by a cube made of eight lattice sites surrounding its center (Fig. 2). The bond length between two consecutive residues may fluctuate within a defined range.⁵⁹ There are 87 bond orientations (relative orientations between cubes) in three dimensions that permit a large number of possible bond angles.⁵⁹ This bond-fluctuation lattice model^{59–61} allows a better account for the excluded volume effect and a more accurate description of helical geometry without significantly changing the efficiency associated with simple lattice models.^{62,63} An alternative approach to improve the accuracy of lattice models was the 210 lattice model developed by Kolinski and Skolnick.^{64–66} A large number of bond orientations reduce the artificial anisotropy of a lattice model.

The total residue-residue interaction energy E is the sum of the internal energies for bond length (E_l), bond angle (E_θ), chirality (E_{ch}) and nonbonded interaction for hydrogen bond (E_{HB}).

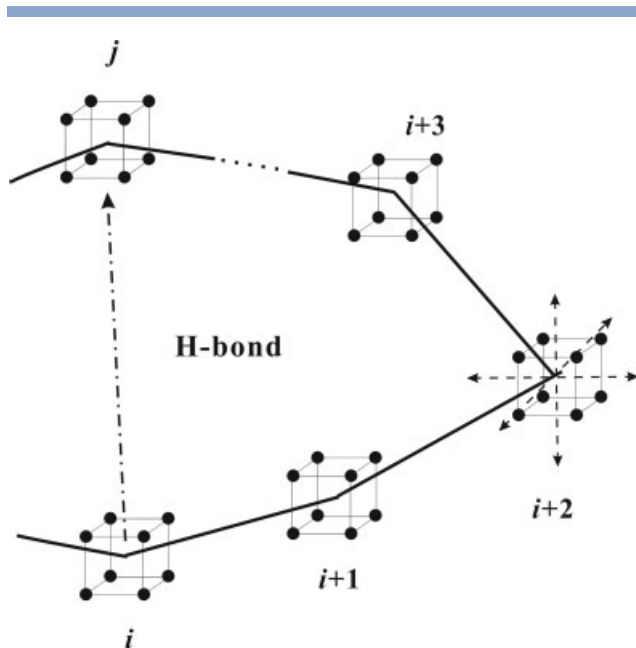
$$E = E_l + E_\theta + E_{\text{ch}} + E_{\text{HB}}, \quad (22)$$

where

$$E_l = \sum_{i=1}^{N-1} u_{l,i} = \sum_{i=1}^{N-1} \frac{1}{2} k_l (l_i - l_0)^2, \quad (23)$$

and

$$E_\theta = \sum_{i=2}^{N-1} u_{\theta,i} = \sum_{i=2}^{N-1} \frac{1}{2} k_\theta (\cos \theta_i - \cos \theta_0)^2. \quad (24)$$

**Figure 2**

Schematic representation of the coarse-grained lattice model. Each residue occupies eight lattice sites. Each lattice site can only be occupied by one residue at one time. The elementary movement is set to allow “half residue” to jump to the four nearest-neighbor vacant sites along one of the six marked directions. A hydrogen bond can be formed between two residues with sequence separation of more than three residues within a cutoff distance.

Here k_l and k_θ are the corresponding force constants for the associated harmonic potentials; l_i and θ_i are the i th bond length and bond angle, respectively; θ_0 and l_0 are their respective expected values. θ_0 is set to be 88.9° as the actual value in α -helix. All lengths are scaled by the size of cube. The bond length l_0 is set to be $\sqrt{6}$, the distance between (0,0,0) and (2,1,1). This distance allows the bonded chain to move efficiently. In this coarse-grained model, the side chains of amino-acid residues are not treated explicitly, thus the specific packing effect cannot be reproduced. As an alternative, we include a term E_{ch} to reproduce the fact that most helices in proteins are right-handed helices.

$$E_{ch} = \sum_{i=1}^{N-3} u_{ch,i}, \quad (25)$$

where

$$u_{ch,i} = \begin{cases} \varepsilon_{ch} & (\mathbf{l}_i \times \mathbf{l}_{i+1} \cdot \mathbf{l}_{i+2} \leq 0) \\ 0 & (\mathbf{l}_i \times \mathbf{l}_{i+1} \cdot \mathbf{l}_{i+2} > 0) \end{cases}. \quad (26)$$

That is, a left-handed chirality for three continuous bond vectors will be penalized with a positive energy ε_{ch} .

The interaction energy of H-bonding is expressed as

$$E_{HB} = \sum_{i=1}^{N-4} u_{HB,i}, \quad (27)$$

where

$$u_{HB,i} = \begin{cases} 0 & (h_{i,j} < h_1) \\ -\varepsilon_{HB} & (h_1 \leq h_{i,j} \leq h_2), \\ 0 & (h_{i,j} < h_2) \end{cases}, \quad (28)$$

and

$$j \geq i + 4. \quad (29)$$

Here, ε_{HB} is the energy of each H-bond, and $h_{i,j}$ is the distance between residue i to residue j (The position of a residue is defined by the residue center). The H-bonding interaction occurs when the inter-residue distance is within the distance range from h_1 to h_2 and their sequence separation is four residues or more. This model allows nonnative H-bonds for sequence separation of five residues or more. The model is an improvement over the previous model which limits the H-bonds for $|i-j| = 4$ (i.e. native H-bonds only).⁵¹ Values for h_1 and h_2 as well as ε_{ch} , ε_{HB} , k_b and k_θ are determined by trials and errors. The parameters are set so that helix conformation is stable at low temperatures. $h_1 = \sqrt{11.5}$, $h_2 = \sqrt{12.5}$, $\varepsilon_{ch} = \varepsilon_{HB}/2$, $k_l = \varepsilon_{HB}$, and $k_\theta = 2\varepsilon_{HB}$.

Simulation method

The bond-fluctuation model is simulated with the same dynamic MC simulation technique employed in our previous work.⁵¹ Every MC step contains N attempts. In each attempt, a residue and four sites that are the nearest neighbors to one of its six cube surfaces are selected randomly (Fig. 2).^{51,59,60} The residue moves when the selected four sites become the new cube surface of the residue. This move is accepted if all the four sites are not occupied and the standard Metropolis criterion⁶⁷ is satisfied.

Equilibrium constants of helix-coil transitions from simulations

We use “0” and “1” to denote the coiled and helical residues, respectively. As shown in Figure 1, the i -th residue is labeled as “1” if an α -helical H-bond is formed between the i -th residue and the $(i-4)$ -th residue and “0” otherwise. (To avoid redundancy, we use only one direction to define helical residues). We also use “100 \rightarrow 110” to denote the propagation of a helix regardless the length of α -helical block by converting a coiled residue located between a helical residue and a coiled residue to a helical one. $n_{100 \rightarrow 110, \text{try}}$ and $n_{100 \rightarrow 110, \text{suc}}$ are the number of tried and the number of successful attempts

for making a conversion. The probability of propagating $p_{100 \rightarrow 110}$ is thus calculated from

$$p_{100 \rightarrow 110} = \frac{\langle n_{100 \rightarrow 110, \text{suc}} \rangle}{\langle n_{100 \rightarrow 110, \text{try}} \rangle}, \quad (30)$$

here ‘ $\langle \rangle$ ’ denotes the average over all residues and associated attempts. Similarly, the possibility of successful attempts for the reverse process (helical to coiled residue), $p_{110 \rightarrow 100}$, can also be calculated. The equilibrium constant $K_{100 \leftrightarrow 110}$ between helix propagating and its reverse process can be calculated from $p_{100 \rightarrow 110}/p_{110 \rightarrow 100}$. Similarly, the equilibrium constant $K_{000 \leftrightarrow 010}$ between nucleation and its reverse process and $K_{101 \leftrightarrow 111}$ between formation and breakage of helical block can be respectively determined by $p_{000 \rightarrow 010}/p_{010 \rightarrow 000}$ and $p_{101 \rightarrow 111}/p_{111 \rightarrow 101}$. Here, $p_{000 \rightarrow 010}$, $p_{010 \rightarrow 000}$, $p_{101 \rightarrow 111}$ and, $p_{111 \rightarrow 101}$ denote the probabilities for the nucleation process, its reverse process, formation, and breakage of an α -helical block, respectively. In the Zimm–Bragg theory, $K_{100 \leftrightarrow 110} = s$, $K_{000 \leftrightarrow 010} = \sigma s$, $K_{101 \leftrightarrow 111} = s/\sigma$. Thus, the knowledge of any two equilibrium constants will allow us to calculate the σ and s parameters from simulation data.

To apply the extended Zimm–Bragg theory based on the helical-block nucleation, we obtain the number of tries and successes for adding or removing a helical residue from a helical block of length m ($n_{m,100 \rightarrow 110, \text{try}}$, $n_{m,100 \rightarrow 110, \text{suc}}$, $n_{m,110 \rightarrow 100, \text{try}}$ and $n_{m,110 \rightarrow 100, \text{suc}}$). If the nucleation block size is n , then

$$\begin{aligned} \sigma^{\text{block}} &= \frac{1}{s^n} \prod_{m=1}^n K_{m,100 \leftrightarrow 110} \\ &= \frac{1}{s^n} \prod_{m=1}^n \frac{\langle n_{m,100 \rightarrow 110, \text{suc}} \rangle / \langle n_{m,100 \rightarrow 110, \text{try}} \rangle}{\langle n_{m,110 \rightarrow 100, \text{suc}} \rangle / \langle n_{m,110 \rightarrow 100, \text{try}} \rangle}. \end{aligned} \quad (31)$$

For s , we simply collect the statistics for $m \geq n$, regardless the length of helix block, in order to increase the statistical certainty of the result.

Reduced specific heat

The specific heat (C_V) or reduced specific heat⁵¹ (c_V^*) is calculated from the energy fluctuation theorem

$$c_V^* = \frac{C_V}{Nk_B} = \frac{\langle E^2 \rangle - \langle E \rangle^2}{N(k_B T)^2}, \quad (32)$$

where k_B , T , and E denote the Boltzmann constant, absolute temperature, and the total energy of the system, respectively, and “ $\langle \rangle$ ” means the averaging over sampled conformations. A reduced, dimensionless temperature $T^*(\equiv k_B T/\epsilon_{\text{HB}})$ will be used in this paper.

Simulation procedure

The model is simulated at 19 different temperatures ($1/T^* = 0, 0.415, 0.830, 1.245, 1.660, 2.075, 2.490, 2.905, 3.320, 3.735, 4.150, 4.565, 4.980, 5.395, 5.810, 6.225, 6.640, 7.055$, and 7.470). Each simulation starts from an equilibrated conformation from a previous simulation at a higher temperature and the temperature is gradually decreased from the highest temperate to the desired temperature. Each temperature is independently simulated 20 times for error estimation. For each simulation, the first half of the simulation is for equilibration and the second half is for data collection. The total number of MC steps for each simulation is between 2 millions for the highest temperature and 38 millions for the lowest temperature. The program is written in FORTRAN. For $N = 32$, it takes about 90 h for all of the 19 simulations on a PC with 2.4 GHz Xeon CPU.

While the majority of our simulations are on the chain length of 32 residues, we also simulate chain length of 12, 16, 24, 50, 65, 80, and 100. The sizes of simulation boxes are adjusted according to chain size to avoid self-interaction due to the period boundary condition. For example, the size of the simulation box is $70 \times 70 \times 70$ for a chain of 100 residues. Simulating chains with different lengths allows us to examine how chain length affects the performance of the Zimm–Bragg theory.

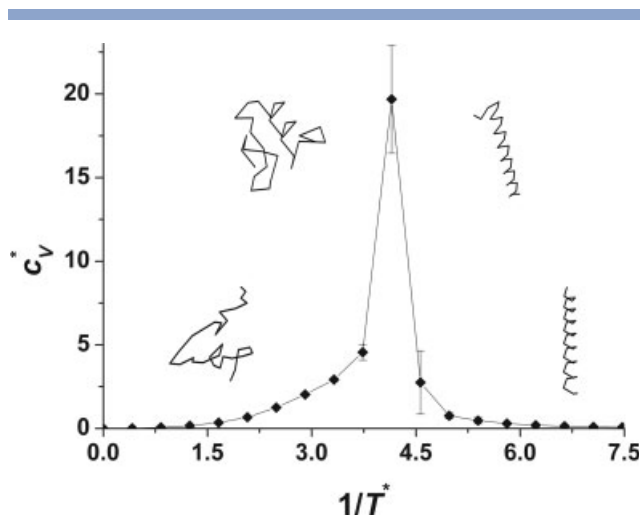
RESULTS

Specific heat

The reduced specific heat as a function of inverse temperature is shown in Figure 3. There is a peak at $1/T^* = 4.15$ indicating a thermodynamic transition for the simulated single homopolyptide chain. Examination of structures reveals that this process is a transition from coil to helix.

Nucleation and propagation parameters for the lattice model

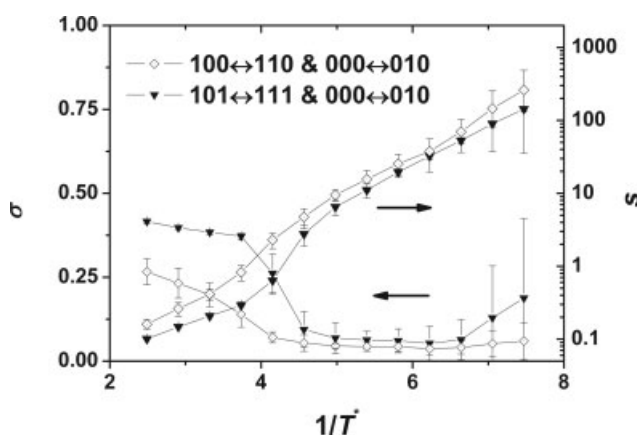
The nucleation and propagation parameters (σ , s) are calculated from equilibrium constants from the simulations. Figure 4 shows the values of σ and s that are calculated either from $K_{100 \leftrightarrow 110}$ and $K_{000 \leftrightarrow 010}$ or from $K_{101 \leftrightarrow 111}$ and $K_{000 \leftrightarrow 010}$ at different temperatures. The figure shows that two parameters, σ and s , have very different temperature dependence. The nucleation parameter varies between 0 and 0.4 over the entire temperature range simulated while the propagation parameter changes three orders of magnitude in the same temperature range. This means that the initiation of the first helical residue is only slightly easier (high σ values) at high temperatures than at low temperature. Once the first helical residue is formed, helical propagation is much easier at lower temperatures when helix structures are more

**Figure 3**

Reduced specific heat c_v^* as a function of reduced inverse temperature $1/T^*$. Insets are the typical snapshots of the configurations at associated reduced inverse temperature. The error bars indicate the standard deviations of all examined trajectories, and are shown unless smaller than the sizes of underlying data points.

stable. Similar temperature dependences of σ and s are also observed by other theoretical studies.^{32,51}

Figure 4 also shows that the differences between the two sets of σ and s are small, especially at low temperatures. This validates the approaches we used to calculate σ and s . Moreover, the results further suggest that helical nucleation and propagation parameters provide a reasonable description for the formation and breakage of helical blocks ($K_{101 \leftrightarrow 111}$) as well. We used σ and s calculated

**Figure 4**

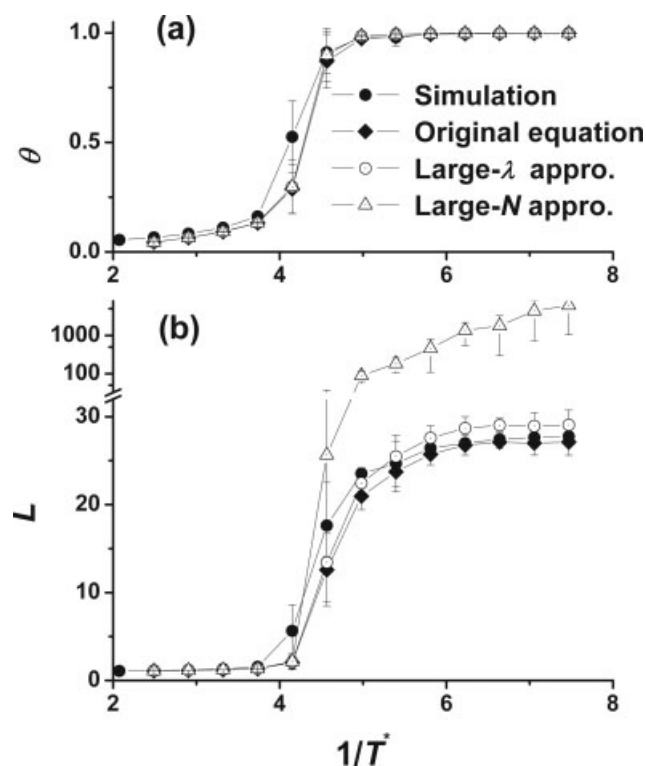
Nucleation parameter σ and propagation parameter s obtained from the equilibrium simulations. They are calculated from the combination of two equilibrium constants (either from $K_{100 \leftrightarrow 110}$ and $K_{000 \leftrightarrow 010}$ or from $K_{101 \leftrightarrow 111}$ and $K_{000 \leftrightarrow 010}$).

from $K_{100 \leftrightarrow 110}$ and $K_{000 \leftrightarrow 010}$ in the remaining sections of this article, because this calculation is easier and conceptually more direct.

Helical ratio and helical block length

The average helical ratio (helicity) and the average length of helical blocks can be obtained directly from simulations. They can also be calculated from the original Zimm–Bragg theory, the large- N approximation, and the large- λ approximation by using σ and s derived from the simulation data as mentioned above.

As shown in Figure 5(a), the helical ratios determined from Eq. (8) (the original Zimm–Bragg theory) and Eq. (11) (the large- N approximation) are in a good agreement with the direct simulation results in the entire temperature range simulated. There are also good agreements between the simulated mean lengths of α -helical blocks and calculated values based on the original Zimm–Bragg theory [Fig. 5(b)]. The large- N approximation, however, produces unphysically high lengths for helical blocks at low temperatures (much greater than the maximum possible length of 28 residues). The results from the large- λ

**Figure 5**

The helical ratio θ (a) and the averaged length of α -helical blocks, L (b) obtained directly from simulation (solid circles) or calculated from the Zimm–Bragg theory using the original equation [Eqs. (8) and (10), diamonds], large- N approximation [Eqs. (11) and (12), open triangles], and large- λ approximation [Eqs. (11) and (16), open circles].

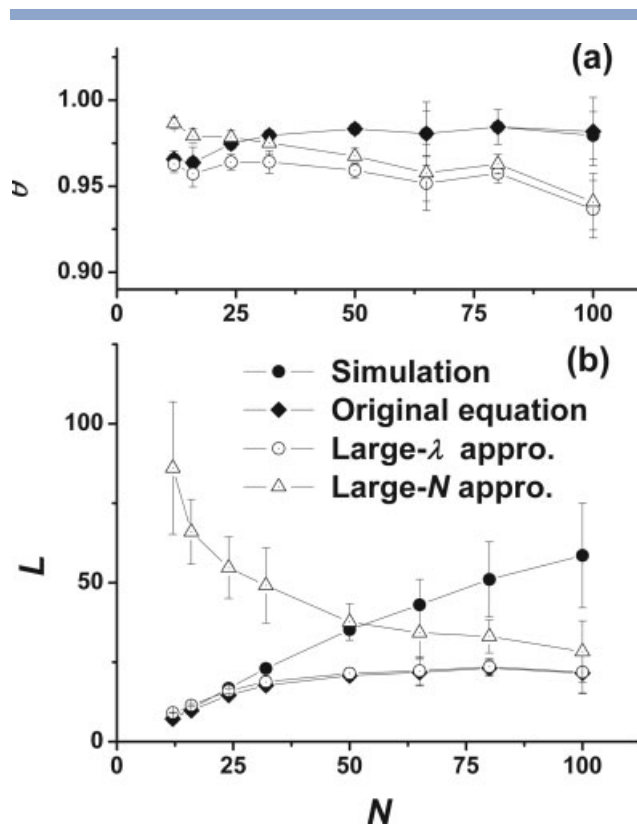


Figure 6

As in Figure 5 but as a function of chain length. The simulation temperature is at $T_c^*/1.3$. T_c^* refers to the temperature at which specific heat is at its maximum. Simulated chain lengths are 12, 16, 24, 32, 50, 65, 80, and 100.

approximation make a significant improvement over those from the large- N approximation by satisfying the low temperature high-helicity limit.

One interesting question is whether or not the agreement between the Zimm–Bragg theory and the large- N approximation will improve as the chain length increases. Figure 6 compares the average lengths of helical blocks for different chain lengths at a temperature that is lower than the transition temperature. Indeed, the results from the large- N approximation become in close agreement with those from the original Zimm–Bragg theory at $N = 100$. However, we surprisingly found that the agreement between simulation results and the results from the Zimm–Bragg theory (no matter from the original formulae, large- N approximation, or large- λ approximation) was not satisfactory even when chain length was increased.

Nucleation by short helical block

The significant difference between what predicted from the Zimm–Bragg theory and what obtained from simula-

tions in Figure 6 implies that there is room for improvement of the original Zimm–Bragg model.

In the Zimm–Bragg theory, it is assumed that the propagation parameter (s) is independent of the length of α -helical block. That is, the equilibrium constant $K_{m,100 \leftrightarrow 110}$ between adding a helical residue to a helical block of length m and its reverse process is a constant (equal to s), independent of m .

In Figure 7, $K_{m,100 \leftrightarrow 110}$ is plotted as a function of m for the homopolyptide chain of 32 residues. The results show that $K_{m,100 \leftrightarrow 110}$ could be regarded as a constant only after helical block is longer than six residues for the model studied here. $K_{1,100 \leftrightarrow 110}$ is the smallest among $K_{m,100 \leftrightarrow 110}$. This means that the formation of the first helical residue is the most difficult. $K_{2,100 \leftrightarrow 110}$ is the second smallest among $K_{m,100 \leftrightarrow 110}$. So, the second helical residue is still difficult to form. Thus, while it is mostly true that the propagation parameter is the same after nucleation it takes more than one residue for the completion of nucleation process.

To improve the Zimm–Bragg model, we assume that helix propagation is initiated by a helical block of length n rather than a single helical residue. The block nucleation parameter can be calculated from $K_{m,100 \leftrightarrow 110}$, $m \leq n$ (see methods). The new block nucleation parameter and propagation parameter are then used to calculate helical ratio and average helical-block length. Figure 8 shows that when helical blocks are used for nucleation, the agreement between predicted and simulated values for average helix length and helical ratio are significantly improved, at long chain length, in particular. A block of two residues makes the largest improvement in agreement between theory and simulation results.

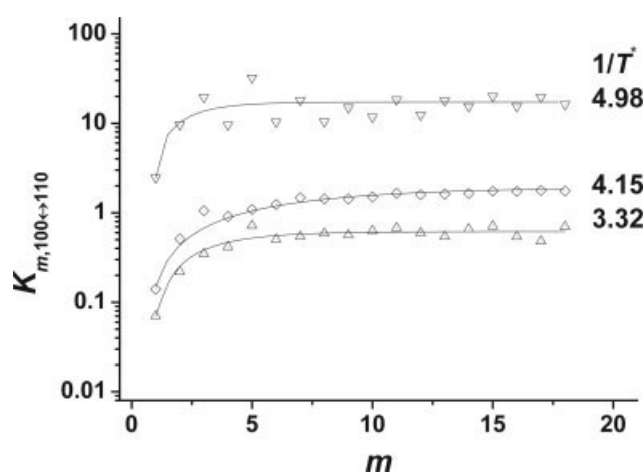
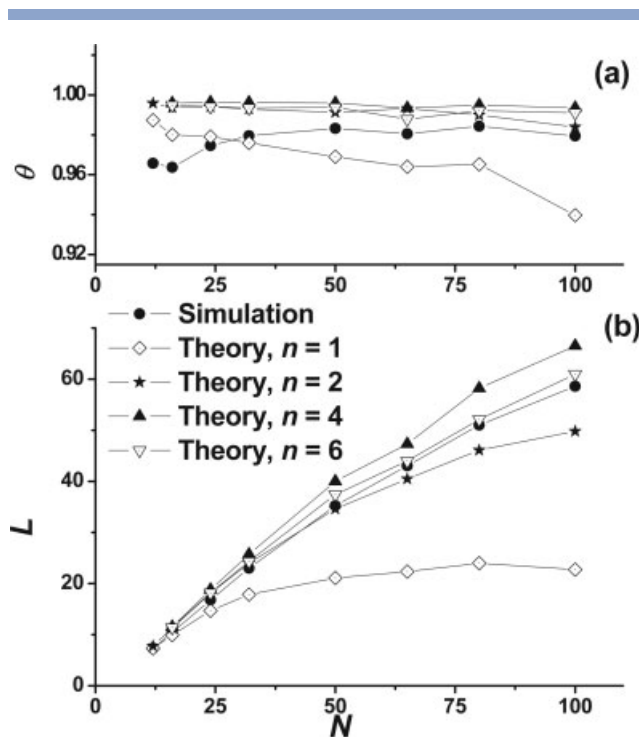


Figure 7

Equilibrium constants $K_{m,100 \leftrightarrow 110}$ as a function of the existing helical block length m . The lines are for guiding eyes.

**Figure 8**

As in Figure 6 but to compare with theoretical predictions based on the mechanism of helical-block nucleation. $n = 1, 2, 4, 6$ indicates the size (number of residue) of nucleation helical block. The original Zimm–Bragg theory corresponds to $n = 1$.

DISCUSSION

In this article, we investigate the helix–coil transition of a bond-fluctuation lattice model by using MC simulation techniques. For a chain length of 32 residues, the simulation result can be described reasonably well by the Zimm–Bragg theory. This is significant because our lattice model allows the formation of nonnative H-bonds. The result is consistent with other simulation studies.^{32,38} For example, Takano et al.³⁸ showed a qualitative agreement between the Zimm–Bragg theory and the all-atom molecular dynamics simulation of a short polyaniline chain ($N = 15$).

The commonly-used large- N approximation for the Zimm–Bragg theory is shown not accurate for predicting the average helix length based on nucleation and propagation parameters. This is because a chain length of 32 residues is not long enough to satisfy the large- N approximation. Similar result has been observed by Ohkubo and Brooks.³² In fact, our detailed analysis indicates that the large- N approximation would be satisfied only if $N \gg s/\sigma$ rather than $N \gg 1$. Because the helical lengths in proteins are often short, it is necessary to have a simple equation that is applicable at the medium chain length as

well. We proposed a large- λ approximation for the average helix length. The result of the new equation is in excellent agreement with the result from the more sophisticated equation in the original Zimm–Bragg theory. The large- λ approximation allows a simple way to relate the helical ratio and the average helix length to the nucleation and propagation parameters in accuracy similar to the original Zimm–Bragg theory.

We further show that the propagating ability is nearly independent of the length of α -helical block if the block is longer than a certain length. This validates the basic assumption in the Zimm–Bragg theory that the propagation parameter is constant after the nucleation process.¹⁵ The phenomena is reminiscent of the radical polymerization in polymer chemistry, in which the polymerizing ability of each step is nearly independent of the chain length after initiation.⁶⁸ However, the length of nucleation block is greater than 1. The formations of first several helical residues (not just the first residue) are all difficult although the first helical residue is the hardest to form. This is understandable because the α -helical H-bond happens not between two neighboring residues but between two residues that has a sequence separation of four residues.

Our simulation discovers that the single-residue nucleation used in the Zimm–Bragg theory leads to a systematic deviation from the simulation results as the chain length becomes longer. We demonstrate that this deviation can be removed by using a mechanism of helix-block nucleation.

ACKNOWLEDGMENT

We thank Professor Jian Zi for a critical reading of our manuscript.

REFERENCES

- Pauling L, Corey RB. Two hydrogen-bonded spiral configuration of the polypeptide chain. *J Am Chem Soc* 1950;72:5349–5349.
- Pauling L, Corey RB, Barnson HR. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci USA* 1951;37:205–211.
- Barlow DJ, Thornton JM. Helix geometry in proteins. *J Mol Biol* 1988;201:601–619.
- Doty P, Yang JT. Polypeptide. VI. Poly-benzyl-L-glutamate: the helix-coil transition in solution. *J Am Chem Soc* 1956;78:498–499.
- Wojcik J, Altmann KH, Scheraga HA. Helix-coil stability constants for the naturally occurring amino acids in water. XXIV. Half-cystine parameters from random poly(hydroxybutylglutamine-co-S-methylthio-L-cysteine) *Biopolymers* 1990;30:121–134.
- Oneil KT, Degrado WF. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* 1990;250:646–651.
- Lyu PC, Liff MI, Marky LA, Kallenbach NR. Side-chain contributions to the stability of alpha-helical structure in peptides. *Science* 1990;250:669–673.
- Scholtz JM, Qian H, York EJ, Stewart JM, Baldwin RL. Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water. *Biopolymers* 1991;31:1463–1470.

9. Chakrabartty A, Kortemme T, Baldwin RL. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci* 1994;3:843–852.
10. Thompson PA, Eaton WA, Hofrichter J. Laser temperature jump study of the helix reversible arrow coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model. *Biochemistry* 1997; 36:9200–9210.
11. Huang CY, Getahun Z, Zhu YJ, Klemke JW, DeGrado WF, Gai F. Helix formation via conformation diffusion search. *Proc Natl Acad Sci USA* 2002;99:2788–2793.
12. Loladze VV, Makhatadze GI. Both helical propensity and side-chain hydrophobicity at a partially exposed site in alpha-helix contribute to the thermodynamic stability of ubiquitin. *Proteins* 2005;58:1–6.
13. Barrett DG, Minder CM, Mian MU, Whittington SJ, Cooper WJ, Fuchs KM, Tripathy A, Waters ML, Creamer TP, Pielak GJ. Pressure perturbation calorimetry of helical peptides. *Proteins* 2006;63:322–326.
14. Schellman JA. The stability of hydrogen-bonded peptide structures in aqueous solution. *C R Trav Lab Carlsberg [Chim]* 1955;29:230–259.
15. Zimm BH, Bragg JK. Theory of the phase transition between helix and random coil in polypeptide chains. *J Chem Phys* 1959;31:526–535.
16. Lifson S, Roig A. On the theory of helix-coil transition in polypeptides. *J Chem Phys* 1961;34:1963–1974.
17. Flory PJ, Miller WG. A general treatment of helix-coil equilibria in macromolecular systems. *J Mol Biol* 1966;15:284.
18. Fixman M, Zeroka D. Helix-coil transition in heterogeneous chains. I. Protein model. *J Chem Phys* 1968;48:5223–5230.
19. Poland D, Scheraga HA. Theory of helix-coil transitions in biopolymers. New York: Academic Press; 1970.
20. Qian H, Schellman JA. Helix-coil theories: a comparative study for finite length polypeptides. *J Phys Chem* 1992;96:3987–3994.
21. Doig AJ, Chakrabartty A, Klingler TM, Baldwin RL. Determination of free-energies of N-capping in alpha-helices by modification of the Lifson-Roig helix-coil theory to include N-capping and C-capping. *Biochemistry* 1994;33:3396–3403.
22. Munoz V, Serrano L. Elucidating the folding problem of helical peptides using empirical parameters. *Nat Struct Biol* 1994;1:399–409.
23. Brooks CL, III. Helix-coil kinetics: folding time scales for helical peptides from a sequential kinetic model. *J Phys Chem* 1996; 100:2546–2549.
24. Avbelj F, Luo PZ, Baldwin RL. Energetics of the interaction between water and the helical peptide group and its role in determining helix propensities. *Proc Natl Acad Sci USA* 2000;97:10786–10791.
25. Buchete NV, Straub JE. Mean first-passage time calculations for the coil-to-helix transition: the active helix Ising model. *J Phys Chem B* 2001;105:6684–6697.
26. McCammon JA, Northrup SH, Karplus M, Levy RM. Helix-coil transitions in a simple polypeptide model. *Biopolymers* 1980;19: 2033–2045.
27. Daggett V, Kollman PA, Kuntz ID. A molecular dynamics simulation of polyaniline: an analysis of equilibrium motions and helix-coil transitions. *Biopolymers* 1991;31:1115–1134.
28. Daggett V, Levitt M. Molecular dynamics simulations of helix denaturation. *J Mol Biol* 1992;223:1121–1138.
29. Yang AS, Honig B. Free-energy determinants of secondary structure formation. 1. Alpha-helices. *J Mol Biol* 1995;252:351–365.
30. Okamoto Y, Hansmann UHE. Thermodynamics of helix-coil transitions studied by multicanonical algorithms. *J Phys Chem* 1995;99: 11276–11287.
31. Young WS, Brooks CL, III. A microscopic view of helix propagation: N and C-terminal helix growth in alanine helices. *J Mol Biol* 1996;259:560–572.
32. Ohkubo YZ, Brooks CL, III. Exploring Flory's isolated-pair hypothesis: statistical mechanics of helix-coil transitions in polyaniline and the C-peptide from RNase A. *Proc Natl Acad Sci USA* 2003; 100:13916–13921.
33. Zhou YQ, Karplus M. Folding thermodynamics of a model three-helix-bundle protein. *Proc Natl Acad Sci USA* 1997;94:14429–14432.
34. Zhou YQ, Linhananta A. Thermodynamics of an all-atom off-lattice model of the fragment B of Staphylococcal protein A: implication for the origin of the cooperativity of protein folding. *J Phys Chem B* 2002;106:1481–1485.
35. Kemp J, Chen Z. Formation of helical states in wormlike polymer chains. *Phys Rev Lett* 1998;81:3880–3883.
36. Sikorski A, Kolinski A, Skolnick J. Computer simulations of de novo designed helical proteins. *Biophys J* 1998;75:92–105.
37. Takano M, Yamato T, Higo J, Suyama A, Nagayama K. Molecular dynamics of a 15-residue poly(L-alanine) in water: helix formation and energetics. *J Am Chem Soc* 1999;121:605–612.
38. Takano M, Nagayama K, Suyama A. Investigating a link between all-atom model simulation and the Ising-based theory on the helix-coil transition: equilibrium statistical mechanics. *J Chem Phys* 2002; 116:2219–2228.
39. Vila JA, Ripoll DR, Scheraga HA. Physical reasons for the unusual alpha-helix stabilization afforded by charged or neutral polar residues in alanine-rich peptides. *Proc Natl Acad Sci USA* 2000; 97:13075–13079.
40. Smith AV, Hall CK. Alpha-helix formation: discontinuous molecular dynamics on an intermediate-resolution protein model. *Proteins* 2001;44:344–360.
41. Wu YD, Zhao YL. A theoretical study on the origin of cooperativity in the formation of 3(10)- and alpha-helices. *J Am Chem Soc* 2001;123:5313–5319.
42. Chowdhury S, Zhang W, Wu C, Xiong GM, Duan Y. Breaking non-native hydrophobic clusters is the rate-limiting step in the folding of an alanine-based peptide. *Biopolymers* 2003;68:63–75.
43. Zhang W, Lei HX, Chowdhury S, Duan Y. Fs-21 peptides can form both single helix and helix-turn-helix. *J Phys Chem B* 2004; 108:7479–7489.
44. Knott M, Chan HS. Exploring the effects of hydrogen bonding and hydrophobic interactions on the foldability and cooperativity of helical proteins using a simplified atomic model. *Chem Phys* 2004; 307:187–199.
45. Sorin EJ, Pande VS. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys J* 2005;88:2472–2493.
46. Shental-Bechor DS, Kirca S, Ben-Tal N, Haliloglu T. Monte Carlo studies of folding, dynamics, and stability in alpha-helices. *Biophys J* 2005;88:2391–2402.
47. van Giessen AE, Straub JE. Monte Carlo simulations of polyaniline using a reduced model and statistics-based interaction potentials. *J Chem Phys* 2005;122:024904.
48. Salvatella X, Dobson CM, Fersht AR, Vendruscolo M. Determination of the folding transition states of barnase by using Phi(I)-value-restrained simulations validated by double mutant Phi(II)-values. *Proc Natl Acad Sci USA* 2005;102:12389–12394.
49. Gnanakaran S, Garcia AE. Helix-coil transition of alanine peptides in water: force field dependence on the folded and unfolded structures. *Proteins* 2005;59:773–782.
50. Podtelezhnikov AA, Wild DL. Exhaustive metropolis Monte Carlo sampling and analysis of polyaniline conformations adopted under the influence of hydrogen bonds. *Proteins* 2005;61:94–104.
51. Chen YT, Zhang Q, Ding JD. A coarse-grained model and associated lattice Monte Carlo simulation of the coil-helix transition of a homopolypeptide. *J Chem Phys* 2004;120:3467–3474.
52. Baldwin RL. In search of the energetic role of peptide hydrogen bonds. *J Biol Chem* 2003;278:17581–17588.
53. Cantor CR, Schimmel PR. *Biophysical chemistry*, Vol.3. New York: Freeman; 1980. pp 1041–1073.
54. Taketomi H, Ueda Y, Gö N. Studies on protein folding, unfolding and fluctuations by computer simulations. *Int J Pept Protein Res* 1975;7:445–459.

55. Lau KF, Dill KA. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules* 1989;22:3986–3997.
56. Kaya H, Chan HS. Simple two-state protein folding kinetics requires near-levinthal thermodynamic cooperativity. *Proteins* 2003;52:510–523.
57. Sali A, Shakhnovich E, Karplus M. Kinetics of protein folding—a lattice model study of the requirements for folding to the native state. *J Mol Biol* 1994;235:1614–1636.
58. Li L, Mirny LA, Shakhnovich EI. Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus. *Nat Struct Biol* 2000;7:336–342.
59. Deutsch HP, Binder K. Interdiffusion and self-diffusion in polymer mixtures: a Monte Carlo study. *J Chem Phys* 1991;94:2294–2304.
60. Carmesin I, Kremer K. The bond fluctuation method: a new effective algorithm for the dynamics of polymers in all spatial dimensions. *Macromolecules* 1988;21:2819–2823.
61. Xu GQ, Ding JD, Yang YL. Monte Carlo simulation of self-avoiding lattice chains subject to simple shear flow. I. Model and simulation algorithm. *J Chem Phys* 1997;107:4070–4084.
62. Verdier PH, Stockmayer WH. Monte Carlo calculations on the dynamics of polymers in dilute solution. *J Chem Phys* 1962;36:227–235.
63. Hilhorst HJ, Deutch JM. Analysis of Monte Carlo results on the kinetics of lattice polymer chains with excluded volume. *J Chem Phys* 1975;63:5153–5161.
64. Kolinski A, Skolnick J. Discretized model of proteins. I. Monte Carlo study of cooperativity in homopolypeptides. *J Chem Phys* 1992;97:9412–9426.
65. Kolinski A, Skolnick J. Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. *Proteins* 1994;18:338–352.
66. Kolinski A, Skolnick J. Monte Carlo simulations of protein folding. II. Application to protein A, ROP, and crambin. *Proteins* 1994;18:353–366.
67. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys* 1953;21:1087–1092.
68. Flory PJ. Principles of polymer chemistry. Ithaca: Cornell University Press; 1953.