The helix–coil transition revisited

Yantao Chen,1 Yaoqi Zhou,2,3 and Jiandong Ding1*

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 homopeptide 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.

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: CQF1717007; Grant sponsor: NIH; Grant numbers: R01 GM 066049; 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
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 \( r \) and \( s \) parameters in the simulation\(^{27,51} \) in addition to evaluation of helicity \( y \) and average helical block length \( L \). The availability of \( r, s, y, \) 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 \( (\lambda) \) approximation that leads to a more accurate simple analytic equation of the average length of helix blocks.

---

**Figure 1**
Schematic representation of the formation of an \( \alpha \)-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 homopolypeptide 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 non-native 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.
METHODS

The Zimm—Bragg theory

The partition function \(Q\) in the Zimm—Bragg theory is expressed as\(^\text{53}\)

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

where \(\Omega_{j,k}\) is the total number of ways to have \(k\) helical hydrogen bonds in \(j\) distinct \(\alpha\)-helical blocks. Note that a helical block is made of consecutive uninterrupted helical residues and \(N - 4\) is the maximum number of \(\alpha\)-helical hydrogen bonds in a polypeptide chain of length \(N\). In this theory, the nucleation process is the formation of the first \(\alpha\)-helical residue (the first native hydrogen bond) among coiled 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,

\[
K_1 = \sigma s, \tag{2}
\]

\[
K_2 = K_3 = \ldots = K_n = \ldots = s. \tag{3}
\]

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

\[
Q = \frac{\lambda_0^{N-3}(\lambda_0 - s) - \lambda_1^{N-3}(\lambda_1 - s)}{\lambda_0 - \lambda_1}, \tag{4}
\]

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

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

\[
\theta = \frac{1}{N - 4} \frac{\partial \ln Q}{\partial \ln s}, \tag{5}
\]

\[
\nu = \frac{\partial \ln Q}{\partial \ln \sigma}, \tag{6}
\]

and

\[
L = (N - 4) \frac{\theta}{\nu} = \frac{s}{\sigma} \frac{\partial \ln Q/\partial s}{\partial \ln Q/\partial \sigma}. \tag{7}
\]

Here, Eqs. (5) and (6) correspond to Eqs. (2) and (44), respectively, in the original paper of Zimm and Bragg.\(^\text{15}\)

Using Eq. (4), one can obtain

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

\[
\nu = \sigma \left[ \frac{(N-3)\lambda_0 + 1}{\lambda_0 - s} \frac{\partial \lambda_0}{\partial \sigma} + \frac{(N-3)\lambda_1 + 1}{\lambda_1 - s} \frac{\partial \lambda_1}{\partial \sigma} \frac{\partial \lambda_0}{\lambda_0 - \lambda_1} \right], \tag{9}
\]

\[
L = \frac{s}{\sigma} \left[ \frac{1}{\lambda_0 - s} \frac{\partial \lambda_0}{\partial \sigma} + \frac{1}{1 - \lambda_0 - s} \frac{\partial \lambda_0}{\partial \sigma} \frac{\partial \lambda_1}{\lambda_0 - \lambda_1} \right]. \tag{10}
\]

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

\[
\theta = \frac{s}{2} - \frac{1}{2 \sqrt{(1-s)^2 + 4s\sigma}}, \tag{11}
\]

\[
L = 1 + \frac{2s}{1 - s + \sqrt{(1-s)^2 + 4s\sigma}}. \tag{12}
\]

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), \(\theta \to 1\) and \(L \to 1 + (s - 1)/\sigma\) at this limit. While \(\theta\) approaches the correct limit, the correct limit for \(L\) should be \(N - 4\). Thus, the large-\(N\) approximation is just appropriate for \(\theta\) 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-\(\lambda_1\) approximation

To achieve the correct limit for \(L\) at \(s \gg 1\) and \(\gg \sigma\). We introduce a large-eigenvalue (\(\lambda_1\)) approximation. A naive way is to neglect all of terms about \(\lambda_1\) in the
expression of Q. Then, Eq. (4) becomes

\[
\ln Q \approx (N - 4) \ln \lambda_0 + \ln(\lambda_0 - s).
\]

(13)

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

\[
\frac{v}{S_0} = \frac{N - 4}{\sqrt{\Delta + s + 1}} + \frac{1}{\sqrt{\Delta - s + 1}.}
\]

(14)

where

\[
\Delta \equiv (1 - s)^2 + 4\sigma s.
\]

(15)

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

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

(16)

In this equation, \(L \rightarrow N - 4\) when \(s \gg 1\) and \(\gg N\). 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-\(\lambda\) 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-\(\lambda\) 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-\(\lambda\) 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 of 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

\[
K_1 = \sigma_1 s,
\]

(17)

\[
K_2 = \sigma_2 s,
\]

(18)

\[
K_n = \sigma_n s,
\]

(19)

\[
K_{n+1} = K_{n+2} = \cdots = s,
\]

(20)

where \(n\) is the size of the nucleation helix block. The nucleation parameter \(\sigma^\text{block}\) for helical block is given by

\[
\sigma^\text{block} = \sigma_1 \sigma_2 \cdots \sigma_n.
\]

(21)

Here, we extend the Zimm–Bragg theory by using the block nucleation parameter \(\sigma^\text{block}\) and the propagation parameter \(s\) obtained from the equilibrium constant \(K_n\).

In this equation, \(\lambda_0^{N-3} \gg \lambda_1^{N-3}\) with a bit tedious procedure. In fact, the so-called large-\(N\) approximation in the literature and the so-called large-\(\lambda\) 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 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 \((E_0)\), bond angle \((E_0)\), chirality \((E_0)\) and nonbonded interaction for hydrogen bond \((E_0)\).

\[
E = E_1 + E_0 + E_0 + E_0.
\]

(22)

where

\[
E_i = \sum_{i=1}^{N-1} i u_{i,j} = \sum_{i=1}^{N-1} \frac{1}{2} k_i (l_i - l_0)^2,
\]

(23)

and

\[
E_0 = \sum_{i=2}^{N-1} i u_{i,j} = \sum_{i=2}^{N-1} \frac{1}{2} k_0 (\cos \theta_i - \cos \theta_0)^2.
\]

(24)
The interaction energy of H-bonding is expressed as

\[ E_{HB} = \sum_{i=1}^{N-4} u_{HB,i}, \tag{27} \]

where

\[ u_{HB,i} = \begin{cases} 0 & (h_{ij} < h_1) \\ -e_{HB} & (h_1 \leq h_{ij} \leq h_2) \\ 0 & (h_{ij} < h_2) \end{cases}, \tag{28} \]

and

\[ j \geq i + 4. \tag{29} \]

Here, \( e_{HB} \) is the energy of each H-bond, and \( h_{ij} \) 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 \( e_{ch} \), \( e_{HB} \), \( k_1 \) and \( k_0 \) are determined by trials and errors. The parameters are set so that helix conformation is stable at low temperatures. \( h_1 = \sqrt{11.5}, h_2 = \sqrt{12.5}, e_{ch} = e_{HB}/2, k_1 = e_{HB}, \) and \( k_0 = 2e_{HB} \).

**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 criterion\(^67\) 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 \( \alpha \)-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 \( \alpha \)-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 \( n_{100 \rightarrow 110, \text{succ}} \) 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

$$p_{100 \rightarrow 110} = \frac{<n_{100 \rightarrow 110, suc}>}{<n_{100 \rightarrow 110, try}>},$$  \hspace{1cm} (30)$$

here ‘<>’ 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 $K_{100 \rightarrow 110}$ between helix propagating and its reverse process can be calculated from $p_{100 \rightarrow 110}/p_{110 \rightarrow 100}$. Similarly, the equilibrium constant $K_{000 \rightarrow 010}$ between nucleation and its reverse process and $K_{011 \rightarrow 111}$ between formation and breakage of helical block can be respectively determined by $p_{000 \rightarrow 010}/p_{010 \rightarrow 000}$ and $p_{011 \rightarrow 111}/p_{111 \rightarrow 101}$. Here, $p_{000 \rightarrow 010}$, $p_{010 \rightarrow 000}$, $p_{011 \rightarrow 111}$ 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, $K_{100 \rightarrow 110} = s$, $K_{000 \rightarrow 010} = \sigma s$, $K_{011 \rightarrow 111} = s/\sigma$. 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 $m$ ($n_{m,100 \rightarrow 110,try}^n$, $n_{m,100 \rightarrow 110,suc}^n$, $n_{m,110 \rightarrow 100,try}^n$ and $n_{m,110 \rightarrow 100,suc}^n$). If the nucleation block size is $n$, then

$$\sigma^{\text{block}} = \frac{1}{s} \prod_{m=1}^{n} K_{m,100 \rightarrow 110}$$

$$= \frac{1}{s} \prod_{m=1}^{n} \frac{<n_{m,100 \rightarrow 110, suc}>}{<n_{m,100 \rightarrow 110, try}>},$$  \hspace{1cm} (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 heat ($c_V^*$) is calculated from the energy fluctuation theorem

$$c_V^* = \frac{C_V}{Nk_B} = \frac{\langle E^2 \rangle - \langle E \rangle^2}{N(k_B T)^2},$$  \hspace{1cm} (32)$$

where $k_B$, $T$, and $E$ denote the Boltzmann constant, absolute temperature, and the total energy of the system, respectively, and “<>” means the averaging over sampled conformations. A reduced, dimensionless temperature $T^* (\equiv k_B T/61k_B)$ 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 temperature 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 homopolypeptide 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 ($\sigma$, $s$) are calculated from equilibrium constants from the simulations. Figure 4 shows the values of $\sigma$ and $s$ that are calculated either from $K_{100 \rightarrow 110}$ and $K_{000 \rightarrow 010}$ or from $K_{011 \rightarrow 111}$ and $K_{000 \rightarrow 010}$ at different temperatures. The figure shows that two parameters, $\sigma$ 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 $\sigma$ 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
stable. Similar temperature dependences of \( \sigma \) and \( s \) are also observed by other theoretical studies.\(^{32,51}\)

Figure 4 also shows that the differences between the two sets of \( \sigma \) and \( s \) are small, especially at low temperatures. This validates the approaches we used to calculate \( \sigma \) 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 (\( K_{101\rightarrow111} \)) as well. We used \( \sigma \) and \( s \) calculated from \( K_{100\rightarrow110} \) and \( K_{000\rightarrow010} \) 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-\( \lambda \) approximation by using \( \sigma \) 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 \( \alpha \)-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-\( \lambda \)
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-\(\lambda\) approximation) was not satisfactory even when chain length was increased.

**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.

Nucleation by short helical block

The significant difference between what predicted from the Zimm–Bragg theory and what obtained from simulations 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 \(\alpha\)-helical block. That is, the equilibrium constant \(K_{m,100}^{110}\) 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, \(K_{m,100}^{110}\) is plotted as a function of \(m\) for the homopolypeptide chain of 32 residues. The results show that \(K_{m,100}^{110}\) could be regarded as a constant only after helical block is longer than six residues for the model studied here. \(K_{1,100}^{110}\) is the smallest among \(K_{m,100}^{110}\). This means that the formation of the first helical residue is the most difficult. \(K_{2,100}^{110}\) is the second smallest among \(K_{m,100}^{110}\). 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 \(K_{m,100}^{110}\), \(m < n\) (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 \(K_{m,100}^{110}\) as a function of the existing helical block length \(m\). The lines are for guiding eyes.
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. For example, Takano et al. 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. In fact, our detailed analysis indicates that the large-N approximation would be satisfied only if \( N \gg \sigma^2 \) 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-\( \lambda \) approximation for the average helix length. The result of the new equation is in excellent agreement with the result from the more sophisticate equation in the original Zimm–Bragg theory. The large-\( \lambda \) 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 \( \alpha \)-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. 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. However, the length of nucleation block is greater than \( \lambda \). 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 \( \alpha \)-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

The Helix–Coil Transition Revisited


