Construction of an intermediate-resolution lattice model and re-examination of the helix-coil transition: a dynamic Monte Carlo simulation

Yantao Chen* and Jiandong Ding*

*State Key Laboratory of Molecular Engineering of Polymers, Shenzhen Key Laboratory of Functional Polymer, College of Chemistry and Chemical Engineering, Shenzhen University, Shenzhen 518060, China; †Department of Macromolecular Science, State Key Laboratory of Molecular Engineering of Polymers, Laboratory of Advanced Materials, Fudan University, Shanghai 200433, China

Communicated by Ramaswamy H. Sarma

(Received 18 February 2013; final version received 27 March 2013)

In protein modeling, spatial resolution and computational efficiency are always incompatible. As a compromise, an intermediate-resolution lattice model has been constructed in the present work. Each residue is decomposed into four basic units, i.e. the α-carbon group, the carboxyl group, the imino group, and the side-chain group, and each basic coarse-grained unit is represented by a minimum cubic box with eight lattice sites. The spacing of the lattice is about 0.56 Å, holding the highest spatial resolution for the present lattice protein models. As the first report of this new model, the helix-coil transition of a polyalanine chain was examined via dynamic Monte Carlo simulation. The period of formed α-helix was about 3.68 residues, close to that of a natural α-helix. The resultant backbone motion was found to be in the realistic regions of the conformational space in the Ramachandran plot. Helix propagation constant and nucleation constant were further determined through the dynamic hydrogen bonding process and torsional angle variation, and the results were used to make comparison between classical Zimm-Bragg theory and Lifson-Roig theory based on the Qian-Schellman relationship. The simulation results confirmed that our lattice model can reproduce the helix-coil transition of polypeptide and construct a moderately fine α-helix conformation without significantly weakening the priority in efficiency for a lattice model.

Keywords: helix-coil transition; intermediate-resolution; lattice chain model; dynamic Monte Carlo simulation; polypeptide

1. Introduction

Computer simulation is referred to as an important complement to the experimental and theoretical methods, via which wealthy information at various levels of spatial and temporal resolution can be obtained. In the field of protein folding, many interesting simulations have been done about the folding process of a protein or related polypeptide (Chan, Zhang, Wallin, & Liu, 2011; Dill, Ozkan, Weikl, Chodera, & Voelz, 2007; Zhang et al., 2007; Zhang et al., 2009). Among them, all-atom molecular dynamics simulation is widely employed, via which many important insights have been provided (Lindorff-Larsen, Piana, Dror, & Shaw, 2011; Luo, Ding, & Zhou, 2008; Shaw et al., 2010), and even straightforward comparisons (De Sancho & Best, 2011; Jani, Sonavane, & Joshi, 2011; Tao, Rao, & Liu, 2010) with experimental results have been made. However, the lack of reliable potentials and accessible computer times has limited the application of this method. As an alternative, Levitt and Warshel firstly employed a simple representation for BPTI to explore the movement of non-local structures and refolded it in allowed computing time (Levitt & Warshel, 1975). At the same year, another simple Gō model also emerged (Taketa, Ueda, & Gō, 1975). The field of coarse-grained modeling of proteins has grown tremendously since that time (Bereau & Deserno, 2009; Chen, Wang, Zhang, & Liu, 2010; Montecelli et al., 2008; Noid et al., 2008; Pandey & Farmer, 2010; Takada, 2012; Tozzini, 2005).

The minimalist low-resolution model is on the other end of atomic high-resolution model, in which one residue is usually considered as the basic unit and represented by a sphere. It holds the priority of rapid collection of meaningful statistics with ensemble average and precisely formulated questions that can be quantitatively answered via converged computations. Thus, in many cases it can serve as a good choice for molecular simulation (Dill et al., 1995). However, those highly coarse-grained models clearly lack the details necessary for interplay between
different driving forces. For example, hydrogen bonding can’t come into being reasonably without explicitly including amino and carboxyl groups. The packing effect is not easy to reflect either if self-avoiding side-chain groups were not fully considered. Therefore, simple modeling cannot fully reproduce the complexity of energy landscape and sufficiently address the so-called Levinthal paradox.

As a good compromise, an intermediate-resolution model may provide a balance between low- and high-resolution models for protein modeling. The simplest intermediate-resolution model is a two-unit model, in which each residue is divided into the backbone group and the side-chain group (Klimov & Thirumalai, 1998; Kolinski & Skolnick, 1994; Levitt & Warshel, 1975; Li, Mirny, & Shakhnovich, 2000; Mukherjee & Bagchi, 2004; Skolnick & Kolinski, 1990). To further mimic the structure of side chain, this two-unit scheme was extended to the three-unit scheme, in which the backbone group is coarse-grained into one unit and the side-chain group is represented by one or two units according to the ratio between α-carbon group and side-chain group for the residue examined (Wallqvist & Ullner, 1994; Zhang, Kolinski, & Skolnick, 2003). To improve details of the backbone structure, another three-unit model is developed by Scheraga and his co-workers (Liwo et al., 1997), which is represented by an α-carbon group, an ellipsoidal side-chain group and a united peptide group located in the center of peptide bond (He, Chen, & Xiao, 2009). In many cases, the backbone structure can be decomposed into three groups, i.e. the α-carbon group, the carboxyl group, and the amino group. The four-unit model was thus put forward, in which each residue is represented by three backbone units and one side-chain unit (Ding, Borreguero, Buldyrev, Stanley, & Dokholyanothers, 2003; Smith & Hall, 2001; Takada, Luthey-Schulten, & Wolynes, 1999). Alternatively, virtual-amino group and virtual-carboxyl group have been introduced into the single-unit model to represent the reacting sites for hydrogen bonding, (Chen & Imamura 2003; Chen, Zhang, & Ding, 2006; Klimov, Betancourt, & Thirumalai, 1998; Yang, Li, Zhang, Ur Rehman, & Liang, 2010) which can be an interesting complement to the explicit representation. If the oxygen in carboxyl group and hydrogen in amino group are explicitly represented by two additional units, it can be called as the six-unit models (Irbyack, Sjunnesson, & Wallin, 2000). Further efforts have also been made. Recently, Ding et al. (Ding, Buldyrev, & Dokholyan, 2005) added one or two more units on the side-chain group for some residues with large side chain (Ding et al., 2003). Shakhnovich and co-workers (Shimada, Kussell, & Shakhnovich, 2001) developed an all-(heavy) atom model without consideration of hydrogen atoms, and Zhou and co-worker (Zhou & Linhananta, 2002) have include all heavy atoms and polar hydrogen atoms into their model, which can be considered as the finest intermediate-resolution model.

The idea of using a low-resolution lattice model for computer studies of protein folding dates back to Gō and co-workers (Taketomi et al., 1975), then broadly employed by other researchers (Dill et al., 1995), and later expanded on the intermediate and high-coordinate lattice spaces (Kolinski & Skolnick, 2004; Li et al., 2000). We have constructed a one-unit lattice model to reproduce the helix-coil transition of homo-polypeptide (Chen, Zhang, & Ding, 2004). It remains two limitations: (1) the resultant helical period was an integer 4 not 3.6; (2) the side-chain group has not been separately reflected. As a further step, a four-unit lattice model has been built in this work and one residue is decomposed into four basic units, including the α-carbon group, the carboxyl group, the imino group, and the side-chain group. The new lattice model holds very high spatial resolution for the representation of proteins, hence able to construct a very fine α-helix in lattice space.

As an elementary process, helix-coil transition has been extensively explored for understanding of protein folding (Chen & Ding, 2010a, 2010b; Chen, Zhou, & Ding, 2007; De Sancho & Best, 2011; Lifson & Roig, 1961; Makhatadze, 2006; Qian & Schellman, 1992; Scholtz & Baldwin, 1992; van Giessen & Straub, 2006; Vitalis & Caflisch, 2012; Xu, Ren, & Li, 2013; Zimm & Bragg, 1959). In this work, polyalanine's helix-coil transition has been reproduced via the four-unit lattice model and dynamic Monte Carlo simulation, and several order parameters have been employed to describe this transition process and the shape of formed α-helix, then a comparison has been made between the classical Zimm-Bragg (ZB) theory and Lifson-Roig (LR) theory based upon the determined helix nucleation parameter and propagation parameter.

2. Simulation model and approach
2.1. Four-unit lattice model
In this intermediate-resolution model, each residue is composed of four basic units, i.e. α-carbon group, imino group, carboxyl group, and side-chain group. For alanine residue, each group is assumed to occupy equivalent volume and represented by a minimum cubic composed of 8 lattice sites at its apexes, as shown in Figure 1; as for other residues, the side-chain group can be represented by none or more than one minimum cubic according to the side-chain volume. In this work, a polyalanine chain with fifteen alanines (N=15) is embedded into a cubic box composed of 48 × 48 × 48 lattice sites. The periodic boundary condition is set along each dimension.

In this work, we aim to construct an α-helix, thus hydrophobicity has not been included just for simplicity. Hydrogen bonding and chirality are of key importance
for $\alpha$-helix formation and thus have been taken into consideration. The energy function for our helix model is written as

$$ E = E_l + E_d + E_\theta + E_\omega + E_{\phi} + E_{ch} + E_{HB}. $$  \hspace{1cm} (1)

Here, the potentials for bond lengths and bond angles between consecutive units are given by

$$ E_l = \sum_{i} u_{l,i} = \sum_{i} \frac{1}{2} k_i (l_i - l_0)^2, $$ \hspace{1cm} (2)

and

$$ E_\theta = \sum_{i} u_{\theta,i} = \sum_{i} \frac{1}{2} k_\theta (\theta_i - \theta_0)^2, $$ \hspace{1cm} (3)

To describe the peptide plane, the following potential is used for the dihedral angle $\omega$,

$$ E_\omega = \sum_{i} u_{\omega,i}, $$ \hspace{1cm} (4)

where

$$ u_{\omega,i} = \begin{cases} \frac{1}{2} k_\omega (\omega_i - \omega_0)^2, & |\omega_i - \omega_0| \leq \sigma, \\ \infty, & \text{otherwise} \end{cases} $$ \hspace{1cm} (5)

Here $k_l$, $k_\theta$, and $k_\omega$ are the corresponding elastic constants for the associated harmonic potentials; $l_i$ is the bond length between unit $i$ to unit $i+1$; $\sigma$ is a dihedral-angle-flexibility parameter; $l_\omega$, $\theta_\omega$, and $\omega_0$ are expectation values as presented in Table 1, which are determined according to the spatial coordinate of a natural $\alpha$-helical polyalanine and based on some degree of coarse-graining. In this lattice model, $l_o$ for the bond length between $\text{C}_\alpha$ and CO groups and that between $\text{C}_\alpha$ and NH groups were set as the same, and $l_o$ for the bond length between $\text{C}_\alpha$ and R groups was referred to that between $\text{C}_\alpha$ and $\text{C}_B$ atoms of the natural polyalanine.

Compared with real $\alpha$-helix and off-lattice models, volume of the unit is limited by the representing minimal cubic in the lattice model. To strengthen the volume-packing effect and confine the unit’s movement into a reasonable range, some “artificial” potential functions have been employed, which are

$$ E_d = \sum_{i} u_{d,i} = \sum_{i} \frac{1}{2} k_d (d_i - d_0)^2, $$ \hspace{1cm} (6)

$$ E_{\phi} = \sum_{i} u_{\phi,i} = \sum_{i} \frac{1}{2} k_{\phi} (1 + \cos 3\phi_i), $$ \hspace{1cm} (7)

$$ E_{\psi} = \sum_{i} u_{\psi,i} = \sum_{i} \frac{1}{2} k_{\psi} (1 + \cos 3\psi_i). $$ \hspace{1cm} (8)

Here, Equation (6) is used to confine the pseudo-bond’s fluctuations (Smith & Hall, 2001). In our model,

<table>
<thead>
<tr>
<th>Angle</th>
<th>( \theta_0 ) (°)</th>
<th>( k_\theta \times 10,000 )</th>
<th>Bond</th>
<th>( l_0 ) (Å)</th>
<th>( k_l )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \angle (\text{N}<em>\alpha-\text{C}</em>\alpha-R_\text{R}) )</td>
<td>109.5</td>
<td>2( 1_{\text{HB}} )</td>
<td>( \text{N}<em>i-\text{C}</em>\alpha )</td>
<td>1.5</td>
<td>( 1_{\text{HB}} )</td>
</tr>
<tr>
<td>( \angle (\text{C}<em>\alpha-\text{C}</em>\alpha-R_\text{R}) )</td>
<td>109.5</td>
<td>2( 1_{\text{HB}} )</td>
<td>( \text{C}<em>\alpha-\text{C}</em>\alpha )</td>
<td>1.5</td>
<td>( 1_{\text{HB}} )</td>
</tr>
<tr>
<td>( \angle (\text{N}<em>\alpha-\text{N}</em>\alpha-\text{C}_\alpha) )</td>
<td>109.5</td>
<td>2( 1_{\text{HB}} )</td>
<td>( \text{N}<em>\alpha-\text{C}</em>\alpha )</td>
<td>1.3</td>
<td>( 2_{\text{HB}} )</td>
</tr>
<tr>
<td>( \angle (\text{C}<em>\alpha-\text{C}</em>\alpha-\text{N}_\alpha) )</td>
<td>120</td>
<td>2( 1_{\text{HB}} )</td>
<td>( \text{C}<em>\alpha-\text{R}</em>\text{i} )</td>
<td>1.6</td>
<td>( 1_{\text{HB}} )</td>
</tr>
<tr>
<td>( \angle (\text{C}<em>\alpha-\text{N}</em>\alpha-\text{C}_\alpha) )</td>
<td>120</td>
<td>2( 1_{\text{HB}} )</td>
<td>( \text{C}<em>\alpha-\text{C}</em>\alpha )</td>
<td>1.6</td>
<td>( 1_{\text{HB}} )</td>
</tr>
<tr>
<td>Dihedral angle</td>
<td>( \omega_\theta ) (°)</td>
<td>( k_\omega \times 10,000 )</td>
<td>Pseudo-bond</td>
<td>( d_\delta ) (Å)</td>
<td>( k_d )</td>
</tr>
<tr>
<td>( \angle (\text{C}<em>\alpha-\text{C}</em>\alpha-\text{N}<em>\alpha-\text{C}</em>\alpha) )</td>
<td>180</td>
<td>3( 1_{\text{HB}} )</td>
<td>( \text{C}<em>\alpha-\text{C}</em>\alpha )</td>
<td>3.8</td>
<td>0.4( 1_{\text{HB}} )</td>
</tr>
</tbody>
</table>
there is only one pseudo-bond, i.e. between two continuous α-carbon groups. Equations (7) & (8) were used to
confine the Ramachandran torsional angles (ϕ, ψ) (Ramachandran & Sasisekharan, 1968), where ϕ and ψ
denote the rotation around the backbone bond (Figure 2).

Generally, the introduction of side-chain group and L-
configuration special restriction can result in a right-hand chirality. However, the relative small volume of side-chain
group weakens this effect to a large extent. Actually, there
is no way to reflect the chiral effect completely even in a
Gō-like model unless the modeling on side-chain groups
has sufficient resolution (Kwiecinska & Cieplak, 2005).

As an alternative, chirality effect is strengthened via the
following expression (Chen & Ding, 2010b)

\[
E_{ch} = \sum_{i=1}^{N-3} u_{ch,i}
\]

where

\[
u_{ch,i} = \begin{cases} 
\epsilon_{ch}, & l_i \times l_{i+1} \cdot l_{i+2} < 0 \\
0, & l_i \times l_{i+1} \cdot l_{i+2} \geq 0.
\end{cases}
\]

Here, \(l_i\) is the bond vector connecting residue \(i\) to residue
\(i+1\). The left-handed chirality for three continuous
bonds will be penalized with a positive energy \(\epsilon_{ch}\); the
right-handed chirality and the planar conformation can
both come into being for three continuous bond vectors.

This simple definition is just for the formation of α-helix
in lattice space, and the construction of other complex
protein structures in a coarse-grained model can consult
a delicate representation for side chain’s orientation
(Buchete, Straub, & Thirumalai, 2004).

For hydrogen bond (H-bond) interactions, four
criteria should be satisfied (Figure 3a): (1) the distance
between amide hydrogen atom and carbonyl oxygen
atom is in an appropriate range, i.e. [1.6 Å, 2.4 Å]; (2)

Figure 3. Schematic representation for (a) hydrogen bonding and
(b) location of the pseudo-hydrogen atom in the \(i\)th imino
group and the pseudo-oxygen atom in the \(j\)th carboxyl group
\((i\) and \(j\) are the serial numbers of residue instead of peptide
bond). \(\gamma_{NHO}\) and \(\gamma_{COH}\) denote the nitrogen-hydrogen-oxygen
and carbon-oxygen-hydrogen angles, respectively; \(h_{i,j}\) denotes
the distance between the \(i\)th pseudo-hydrogen \(H_i\) and the \(j\)th
pseudo-oxygen \(O_j\). In our model, \(H_i\) and \(O_j\) were not explicitly
included. The location of \(H_i\) can be determined from the
pseudo-vector linking \(N_i\) to \(H_i\). The orientation of this pseudo-
vector is determined along the anti-direction of the bisector of
the angle \(C_{i-1}-N_i-C_{i+1}\). In analogy to it, the location of \(O_j\) can
be determined by the combination of \(C_{j-1}, C_{j}\) and \(N_{j+1}\). A
hydrogen bond is formed with suitable \(h_{i,j}\), \(\gamma_{NHO}\) and \(\gamma_{COH}\)
where \(j \geq i+4\).

the nitrogen-to-hydrogen vector in the amino group and
carbon-to-oxygen vector in the carboxyl group are rea-
onably coaxial; (3) neither amino group nor carboxyl
group has already been involved in a formed H-bond;
(4) the involved amino group and downstream carboxyl
group is apart of at least four amino acids in sequence.

Thus, the associated potential for hydrogen bonding
interaction can be expressed as

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

where

\[
u_{HB,i,j} = \begin{cases} 
\epsilon_{HB}, & |h_{i,j} - h_0| \leq \delta h_0 \land \gamma_{COH} \geq \gamma_0 \land \gamma_{NHO} \geq \gamma_0 \\
0, & \text{otherwise}
\end{cases}
\]

Figure 2. Schematic representation of dihedral angles and
pseudo-bonds. \(\alpha\) denotes the dihedral angle for peptide plane;
ϕ, ψ denote the Ramachandran torsional angles. The dashed
lines represent pseudo-bonds.
and Oj were not explicitly included. As shown in Figure 3b, the location of Hj can be determined from the nitrogen-to-hydrogen vector (Ni→Hj), which is the combination of carbon-to-nitrogen vector (Ci→Nj), and nitrogen-to-oxygen vector (Ni→Co); similarly, Oj can be located by the combination of vector from Coj to Cj and that from Nj−1 to Cj. Hydrogen bonding interaction occurs only when sequence interval is equal to or larger than 4 in the direction from the N terminus to the C terminus, i.e. j \geq i + 4.

In this model, we set the length of hydrogen bond (from hydrogen atom to oxygen atom) h0 as 2 Å, the distance from associated nitrogen atom to hydrogen atom as 1 Å, and the distance from associated carbon atom to oxygen atom as 1.2 Å. The special parameters are presented in Table 1, and other parameters in our potentials were determined mainly by trial and error, which are listed as following: εHb = 0.412, kR = kϕ = 0.412, σ = 35°, δ = 0.2, and γ = 145°. In this work, the H-bonding energy εHb is used as a basic unit and other energy-related parameters are scaled with εHb.

### 2.2. Simulation approach

The bond-fluctuation model (Carmesin & Kremer, 1988; Cui, Ding, & Chen, 2006; Deutsch & Binder, 1991; Xu, Ding, & Yang, 1997) was employed, which was originally proposed to deal with commodity polymers. Bond length between two neighboring units can fluctuate within a defined range. One advantage of this model over the conventional single-site cubic lattice model (Hilhorst & Deutch, 1975; Verdier & Stockmayer, 1962) is that the number of the permitted bond orientation in three dimensions is as large as 87 and thus can be considered as quasi-continuous or quasi-off-lattice one, which can be an evident superiority when constructing a “realistic” helix via lattice models (Chen et al., 2004, 2006). Another advantage is that the branching point that can be easily incorporated, such as the disulfide bond and the side chain. In the four-unit model, the α-carbon group is a branching point, as it should simultaneously bond to the amino group, the carboxyl group, and the side-chain group.

For this lattice model in three dimensional space, bond length can take five discrete values: 2, √3, √8, √10, and the equilibrium value is about 2.68 for a random walking chain. This corresponds to the expectation value of bond length between α-carbon group and amino group or carboxyl group of the same residue, i.e. 1.50 Å (Table 1). Other bond lengths in Table 1 have been set via the same proportion. Thus, the spacing of the lattice in this four-unit lattice model is about 0.56 Å, which is better than that of the high-coordinate 310 hybrid lattice model, i.e. 1.22 Å (Kolinski & Skolnick, 2004). This high spatial resolution can not only increase the ability for the construction of more “realistic” α-helix but also obviously reduce the artificial anisotropic effect which is inevitably resulted from the limited available orientation of resultant helixes in lattice space (Kolinski & Skolnick, 2004).

The dynamic Monte Carlo simulation (Carmesin & Kremer, 1988; Deutsch & Binder, 1991) was chosen. One Monte Carlo step (MCS) contains N attempts. In each attempt, a residue is randomly chosen, and then one of the nearest neighbor “site groups” in six principal directions is randomly selected, as shown in Figure 1. As for a volume-excluded chain, if any of the four sites in the selected “site group” is occupied, the residue’s movement can not be performed. Metropolis importance sampling (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) is employed as the last criterion for each attempt, and also thermal annealing for jumping out local minima. Each helix-coil transition trajectory starts from a different random-coil transition and one trajectory, about 30 annealing steps are needed from athermal state to the lowest simulated temperature. For each temperature, w MCSs are spent to reach to its thermodynamic equilibrium, and then another w MCSs for the collection of statistics. The result of each collection is averaged over 1 000 data points. More MCSs are required to reach its thermodynamic equilibrium state at lower temperature. In the athermal state, w = w0, and during annealing process, w = INT (0.8/T + 1) × w0. Here, a dimensionless temperature T* (≡ kBT/εHb) is defined, kB is Boltzmann constant, and T the absolute temperature. The program was coded via FORTRAN by us. For the polypeptide of 15 alanines, w0 = 4 × 10^6 MCSs and about 800 h was spent to run all of the 30 trajectories on a PC with 2.4 GHz CPU.

### 2.3. Measured quantities

As a very good order parameter for describing the thermodynamic transition, specific heat has been broadly used in protein folding research. Many significant changes in the simulated system can be indicated via the appearance of peaks of specific heat. The reduced specific heat (Chen et al., 2004, 2006) is defined as,

\[
C_v = \frac{C_v}{Nk_B} = \frac{\langle E^2 \rangle - \langle E \rangle^2}{N(k_B T)^2}.
\]

Here, \(\langle \cdot \rangle\) denotes the conformation average.

To characterize the size of chain conformation and associated variation, the mean square end-to-end distance <R^2> and mean square radius of gyration <S^2> were employed:

\[
R^2 = \frac{1}{N} \sum_{i=1}^{N} r_i^2
\]

(14)
and

\[ S^2 = \frac{1}{N} \sum_{i=1}^{N} (\vec{r}'_i - \vec{r}'_{cm})^2. \]  

Here, \( \vec{r}'_i \) is the position vector for the \( i \)th residue's \( \alpha \)-carbon group in the laboratory frame and \( \vec{r}'_{cm} \) for the center of mass of the whole polypeptide chain as determined by just the \( \alpha \)-carbon groups. For simplicity, the conformation related order parameters were mainly calculated based on the coordinates of \( \alpha \)-carbon group.

The flexibility of a polypeptide chain and associated functional groups is of interest. Generally, the susceptibility of conformational order parameters is a good measure of structural flexibility. Thus, three order parameters are defined: \( D_{Ca} \) denotes the distance between the \( i \)th \( \alpha \)-carbon group and the \((i+4)\)th one; \( D_{SC} \) denotes that between the \( i \)th side-chain group and the \((i+4)\)th one; \( D_{CN} \) denotes that between the \( i \)th carboxylic group and the \((i+4)\)th imino group. Thus, the susceptibility of \( D_{Ca} \) is defined as

\[ \chi_{Ca} = \frac{\langle D_{Ca}^2 \rangle - \langle D_{Ca} \rangle^2}{\langle D_{Ca}^2 \rangle^2}, \]  

and similarly for \( \chi_{SC} \) and \( \chi_{CN} \).

In the simulation of helix formation, it is important to judge the formation of a helix block and the corresponding length. Generally, the judgment of a helix block is based on the dihedral angles \((\phi, \psi)\) or the hydrogen bond. In this work, the H-bond related pattern recognition method (Kabsch & Sander, 1983) was employed as the criterion to determine the so-called \( \alpha \)-helix block. Two consecutive \( \alpha \)-helical H-bonds (from \( i \) to \( i + 4 \)) can shape a minimal \( \alpha \)-helix block, in which all sequential residues can be called as \( \alpha \)-helical residues. Two minimal \( \alpha \)-helix blocks can build a long \( \alpha \)-helix block if at least two consecutive residues are shared. Thus, the maximum number of \( \alpha \)-helical residues is \( N \), when the whole chain forms one helix block.

At a finite temperature the stable helical structure may not be perfect. In order to well describe a non-perfect helix statistically, a spatial correlation function has been suggested in our previous paper (Chen et al., 2006) to describe the periodicity and regularity of a helix as

\[ G(m) = \frac{1}{N - 3} \sum_{i=1}^{N-2} g(m, i). \]  

where

\[ g(m, i) = \frac{1}{N - m - 1} \sum_{j=i}^{N-m-1} (\cos \theta_{ij} - \cos \theta_{ij}) (\cos \theta_{ij+m} - \cos \theta_{ij}), \]

and

\[ \cos \theta_{ij} = \frac{l_i \cdot l_j}{|l_i| |l_j|}. \]

In this function, \( m \) means sequence interval, which ranges from 0 to \( N/2 - 1 \) in our calculation. \( \cos \theta_{ij} \) denotes the average of \( \cos \theta_{ij} \) over \( j \) from 1 to \( N-1 \). The values of \( G(m) \) for a series of sequence interval can be fitted by the following equation,

\[ G(m) = \exp(-m/\xi) \cos(2\pi m/P). \]

Here, two parameters, \( \xi \) and \( P \) are interpreted as the orientation correlation length and period of helix, respectively. \( \xi \) describes the correlation of residues in the direction of helix's principal axis, thus reflecting the regularity of a secondary structure. \( P \) is the mean number of residues per turn of the analyzed polypeptide chain.

The ZB theory (Zimm & Bragg, 1959) and LR theory (Lifson & Roig, 1961) are the two most classical theories for the helix-coil transition, but focus on different sides of helix formation, i.e. hydrogen bonding and torsional angle variations, respectively. Two sets of parameters are defined, i.e. \((\sigma, s)\) in ZB theory and \((v, w)\) in LR theory. Here, nucleation constants, \( \sigma \) and \( v \), reflect the probability of nucleating a helical block among coil blocks; propagation constants, \( s \) and \( w \), measure the propagating ability at the helix-coil interface for an existed helical block. After reconsidering the difference in the reference state between ZB theory and LR theory, (Qian & Schellman, 1992) have established a relationship between those two sets of parameters 20 years ago. They are

\[ \sigma \leftrightarrow v^2/(1 + v)^4, \]

and

\[ s \leftrightarrow w/(1 + v). \]

In this four-unit model, both set of parameters can be determined independently from the dynamic process of inter-conversion between helical state and coiled state for a specific residue, hence a direct test of the Qian-Schellman relationships can be made. In our preceding work
(Chen et al., 2004, 2006), a useful method has been suggested to determine \( \sigma \) and \( s \) based on the dynamic process of H-bond formation and breakage in the thermodynamic equilibrium. Propagation parameter \( s \) is identical to the equilibrium constant \( K_{100 \rightarrow 110} \) between helix propagation and shortening. (Zimm & Bragg, 1959) which is determined by

\[
K_{100 \rightarrow 110} \equiv \frac{P_{100 \rightarrow 110}}{P_{110 \rightarrow 100}}.
\]

Here, “0” and “1” denote the coiled and helical residues, respectively; “100 \( \rightarrow \) 110” denotes the helix propagation. According to ZB theory, the \( i \)th residue is marked as “1” if an \( \alpha \)-helical H-bond \( (i \rightarrow i+4) \) is formed, otherwise as “0”. For the judgment of \( \alpha \)-helical hydrogen bond \( (i \rightarrow i+4) \), an additional requirement should also be satisfied, that is, the three in-between residues \( (i+1, i+2, i+3) \) should be in the \((\phi, \psi)\) space of an \( \alpha \)-helix, i.e. \((-60^\circ \pm 30^\circ, -60^\circ \pm 30^\circ)\) (Qian & Schellman, 1992). Here, the range \(\pm 30^\circ\) is selected by trial and error, which is the same as Sorin and Pande (Sorin & Pande, 2005).

The quantity \( P_{100 \rightarrow 110} \) denotes the helix propagating probability, which is calculated from \( \langle n_{100 \rightarrow 110, \text{tr}} \rangle / \langle n_{100 \rightarrow 110, \text{suc}} \rangle \). Here \( n_{100 \rightarrow 110, \text{tr}} \) and \( n_{100 \rightarrow 110, \text{suc}} \) denote the number of tried and successful attempts of converting the specific coiled residue to the helical state; \( \langle \cdot \rangle \) denotes the average over all residues and attempts. Similarly, the helix shortening possibility, \( P_{111 \rightarrow 110} \), can also be calculated. As for the nucleation constant \( \sigma \), it can be determined by the combination of \( K_{100 \rightarrow 110} \) and \( K_{000 \rightarrow 010} \) (\( \equiv \sigma s \)). In a similar way, \( v \) and \( w \) can be determined by the following equilibrium constants: \( K_{\text{ccc}c-\text{chcc}} \), \( K_{\text{chcc}-\text{chhc}} \), and \( K_{\text{hhcc}-\text{thhc}} \); according to LR theory (Lifson & Roig, 1961), \( K_{\text{ccc}c-\text{chcc}} = K_{\text{chcc}-\text{chhc}} \equiv v \); \( K_{\text{hhcc}-\text{thhc}} \equiv w \). Here, “c” and “h” denote the coiled and helical residues, respectively, and the \( i \)th residue is marked as “h” if its conformation is in \( \alpha \)-helix related \((\phi, \psi)\) space, i.e. \((-60^\circ \pm 30^\circ, -60^\circ \pm 30^\circ)\), otherwise as “c”.

Based upon ZB theory, helical ratio or helicity \( \theta \) can be calculated via the following formulas (Okamoto & Hansmann, 1995),

\[
\theta = \frac{<H >}{N-4} = \frac{1 - s}{2\sqrt{(1 - s)^2 + 4s\sigma}}.
\]

where \( n \) refers to the whole number of residues in the \( \alpha \)-helix blocks for a polypeptide chain.

3. Results and discussion

3.1. Specific heat peak and thermodynamic transition

The variations of both reduced specific heat and chain energy during annealing are displayed in Figure 4. Here, the reduced inverse temperature \( 1/T^* (\equiv e_{\text{HB}}/(k_B T)) \) is employed as an indicator for the temperature annealing progress. At about \( 1/T^* = 6.4 \), the reduced specific heat exhibits a peak with significant fluctuations, and simultaneously the chain energy \( E \) decreases more quickly. Such a large specific heat peak is often interpreted as an indicator of a thermodynamic transition. Figure 4b also illustrates that the hydrogen bonding interaction is the dominant driving force for the transition in this model, and the torsional angle energy takes the second place.

3.2. Chain conformation and number of helix blocks

The measured quantities for polymer conformation, i.e. mean squared radius of gyration \(<S^2>\) and mean squared end-to-end distance \(<H^2>\), were employed to describe the variations of polypeptide chain’s conformation, as shown in Figure 5a. With decreasing temperature, the chain expands to a certain degree, leading to peaks around \( 1/T^* = 5 \) for both \(<H^2> \) and \(<S^2> \). For a random coil, \(<H^2> \) is proportional to \(<S^2> \) with the ratio of 6. (The left and right normal coordinates in Figure 5a are set via this ratio). However, this proportion is clearly destroyed after the transition point, which can be understood in light of the formation of rod-like conformation.

The average length and the number of helix blocks per chain, i.e. \( L \) and \( v \), have been determined (Figure 5b) based on the pattern recognition method (Kabsch & Sander, 1983). In the transition process, \( L \) increases rapidly, and then reaches its maximum, i.e. \( N \). In
contrast, \( \nu \) decreases to the number of one after passing its peak around the transition point, and the longer the chain, the higher the peak (data not shown). Hence, helix formation is probably initiated from different residues simultaneously, multiple nucleuses can co-exist in a polypeptide chain, leading to a local minimum for chain conformation around \( 1/T^* = 7.5 \) in Figure 5a, and then short helix blocks merged together to form longer rod-like helix at low temperature. A similar phenomenon was also found in simulations based upon the low-resolution single-unit model (Chen & Ding, 2010b) and the all-atom model (Ohkubo & Brooks, 2003).

3.3. Ramachandran plot for the formed helix
To further explore the formed helix and test our potentials, the Ramachandran plot of \( \phi \) versus \( \psi \) has been generated (Figure 6). In the athermal state, there is no clear dominant region for \( \phi \) and \( \psi \) in this plot (Figure 6), and the polyalanine chain can approach almost all the possible conformations. During the thermodynamic transition process, more and more conformations are excluded, and \( \phi \), \( \psi \) show a preference for the third quadrant of the conformational space, such as around the transition point as displayed in Figure 6b, which can be an important hint for the partial formation of \( \alpha \)-helix. At low temperatures, only a few conformations can be approached, and \( \phi \), \( \psi \) focuses on a very small region about \((-60^\circ, -60^\circ)\) (Figure 6c), which is satisfactorily consistent to the torsional angle distribution of \( \alpha \)-helix in real proteins and other simulation results\(^{25-26}\). What’s more, the distribution of \( \phi \), \( \psi \) are also consistent to the helical ratio as judged via the H-bond pattern recognition method (Figure 5b). Thus, all of above results gave us confidence that our four-unit lattice model can finely simulate the polypeptide chain not only in random-coil state but also in \( \alpha \)-helical state.

3.4. Helical period and persistent length
The orientation correlation length \( \xi \) and period \( P \) were also determined via fitting to equation (20). During annealing, \( \xi \) first decreases slowly, and then increases most rapidly after the transition point (Figure 7). Hence, a much regular \( \alpha \)-helix can be formed via our four-unit lattice model. On the other hand, the period \( P \) presents a complex behavior before the transition point, and after transition \( P \) decreased rapidly to about 3.68, which was close to 3.6 in a real \( \alpha \)-helix. As a comparison, the period of helix formed based on our previous single-unit model lattice model was an integrator of 4 (Chen et al., 2010b).
Thus, more realistic \( \alpha \)-helix can be constructed via the present four-unit lattice model, which is non-trivial for modeling helix bundles and coiled coils. As each residue has its specific position in the amphiphilic helix blocks, and period is critical when embedding amphiphilic helix bundle into lattice models.

Compared to our previous simulation of helix formation (Chen et al., 2004, 2006), the variation for \( P \) is more complex. Other than the difference on spatial resolution between previous single-unit model and the present four-unit model, the chirality in potential function has been renewed. As shown in equation (10), not only right-hand conformation but also planar conformation is allowed to form in the four-unit model. Hence, the complex period variation may be resulted from some planer conformation. To get more insight for this, more work will be done based on all-atom simulation as a validation.

### 3.5. Comparison between ZB theory and LR theory

The nucleation constant \( \sigma \) and propagation constant \( s \) have been determined according to the variations of H-bonding and Ramachandran torsional angles. As displayed in Figure 8, the propagation constant increased exponentially during the simulated temperature range and was nearly 1 at the helix-coil transition point. Different from the temperature-independence assumption for the nucleation constant in ZB theory, our results show a clear dependence among the studied range of temperature, and more complex variations have also been detected in other researches (Ohkubo & Brooks, 2003; van Giessen & Straub 2006). Additionally, Figure 8 demonstrates that the direct simulation outputs basically agree with the theoretical predictions based upon the calculated \( \sigma \) and \( s \) with equation (24), which further verifies that our four-unit model is self-consistent.

In this work, both propagation constant and nucleation constant for ZB theory (\( \sigma \) & \( s \)) and LR theory (\( v \) & \( w \)) have been determined based on the dynamic process of helix formation. Thanks to Qian-Schellman relationship (Qian & Schellman, 1992), \( v \) & \( w \) have been converted to \( \sigma \) & \( s \) according to equations (21) & (22).

As shown in Figure 8, the propagation constant and nucleation constant were determined based upon the hydrogen bonding and Ramachandran torsional angles and are close to each other. Hence, though ZB theory and LR theory described the helix-coil transition in different points of view, they both catch the essential of helix formation. As shown in Figure 8c, there is still notably difference for helicity especially around the transition point. This may be resulted from LR theory, which restricted to the assumption of only a single helical segment being allowed to form. Around the transition temperature, there are many cases that multishort helical segments co-exist and the mean number of helix can be larger than one as shown in Figure 5, hence leading to a different nucleation constant for LR theory as compared with ZB theory. Similar phenomena for LR theory has also been detected in a recent work by Vitalis and Caflisch (Vitalis & Caflisch, 2012).
3.6. Flexibility of the chain and associated functional groups

Susceptibility of the distance between two groups is presented in Figure 9. As shown in equation (16), the susceptibility is scaled by the distance itself, and identical to the thermal fluctuation, thus can be a suitable measure of chain flexibility. At high temperatures, all the building units for polypeptide chain are in the state with large fluctuation. Any interaction with finite energy potential can be suppressed, thus the flexibility mainly resulted from the thermal fluctuation and identical to each other. The case is different near the transition region, and peaks appear for $\chi_{CN}$, $\chi_{C\alpha}$, and $\chi_{SC}$. $D_{CN}$ exhibits the maximum fluctuation in this region, which can be understood in light of the hydrogen bonding between the $i$th carboxyl group and the $(i+4)$th amino group. What's more, the temperature for the peak of $\chi_{CN}$ is very close to the transition point, which is also consistent with the major-driving-force role for the hydrogen bonding (Figure 4). The third case appeared at low temperatures when helix has been basically formed. For a solid-like helix, the $i$th carboxyl group and the $(i+4)$th imino group are fixed by the hydrogen bonding, which sharply decreases the corresponding fluctuation $\chi_{CN}$. Different behaviors were found for the side-chain groups. With less structural limitation, the side-chain groups move more freely, thus exhibiting larger flexibility $\chi_{SC}$.

4. Concluding remarks

In this work, a four-unit lattice model for the polypeptide chain has been constructed in lattice space, in which one residue is decomposed into four basic units, i.e. the $\alpha$-carbon group, the carboxyl group, the amino group, and the side-chain group. This intermediate-resolution model is a good compromise between low-resolution models (few details, computationally fast) and high-resolution models (many details, computationally slow) for protein modeling. The lattice space is represented via an eight-site lattice model, which can not only produce a comparatively realistic secondary structure but also provide branching point for residue’s side chain and disulfide bond. Through dynamic Monte Carlo simulation, helix-coil transition of polyalanine has been reproduced and associated thermodynamic properties were described by several measured quantities. Helix propagation and nucleation constants have been determined through the dynamic hydrogen bonding process and torsional angle variation. Based on the Qian-Schellman relationship, we found that the classical ZB theory and LR theory can both give a good prediction for helicity for the helix-coil transition.

This new intermediate-resolution model exhibit several advantages. (1) The spacing of lattice in this four-unit lattice model is about 0.56 Å, holding maybe the highest spatial resolution for lattice protein model. (2) As hydrogen bonding interaction occurred between the explicit carboxyl and amino group in the model, a fine $\alpha$-helix with period 3.68 can be formed, close to natural value and much better than that with integral period (=4) formed via sing-unit model in our previous work (Chen et al., 2004). The formation of fine $\alpha$-helix was also attested by the region of dihedral angle ($\phi, \psi$) in the Ramachandran plot. (3) In this four-unit model, the side group is not limited to occupy just one basic eight-site lattice group and the hydrophobicity can be assigned here. Hence, the most important improvement of the present four-unit lattice model as compared to our one-unit lattice of polypeptide is that sequence is much easy to explicitly dealt with, although merely a homopolypeptide chain was examined in this paper. On the other hand, this four-unit lattice model is not so much time-consuming as the simulation is performed in a discontinuous space. (4) Not only $\alpha$-helical conformation but also planar $\beta$-sheet is allowed in our model, thus it can simulate the conformational transformation process for polypeptide with alternated side-chain group’s surroundings.

In conclusion, the present model and associated dynamic MC simulation are prospective in some challenging fields such as folding kinetics, protein dynamics, and multi-peptide aggregation process.
Acknowledgments

The authors are grateful for the financial support from NSF of China (Nos. 21034002, 21174088), Chinese Ministry of Science and Technology (973 program No. 2009CB930000), the Basic Research Program of Shenzhen Municipal Scientific and Technological R & D funding (No. JCYJ201105130434A), Opening Project of State Key Laboratory of Molecular Engineering of Polymers (No. Z2012-03), for the computing resources provided by the High Performance Computing Center (HPCC) of Nanjing University and the National Supercomputing Centre in Shenzhen of China.

References


