Statistical mechanical model for helix-sheet-coil transitions in homopolypeptides

Liu Hong* and Jinzhi Lei†
Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University, Beijing, People's Republic of China, 100084
(Received 5 June 2008; published 6 November 2008)

In this paper, we propose a simple statistical mechanical model to study the conformation transition between the α helix, β sheet, and random coil in homopolypeptides. In our model, five parameters are introduced to obtain the partition function. There are two factors for helical propagation and initiation, which are the same as those used in the Zimm-Bragg model, and three newly introduced parameters for β structures: the strand propagation factor for residues in β strands and two correction factors for the initiation effect of the β strand and β sheet. Our model shows that the variation of these parameters may induce conformation transition from α helix or random coil to β sheet. The sharpness of the transition depends on the initiation factors.

DOI: 10.1103/PhysRevE.78.051904 PACS number(s): 87.15.bd, 87.15.Cc, 87.15.hp

I. INTRODUCTION

Formation of regular secondary structures in proteins has attracted great interest in the last few decades [1,2]. Many well-known diseases, such as Alzheimer's, Mad Cow, and Parkinson's disease, are caused by protein structure transition. For example, the interconversion between α-helix and β-sheet conformation of the protein prion is closely related to Mad Cow disease [3]. However, despite numerous efforts, the underlying mechanisms of how such a transition happens are still controversial [1,2]. Many interactions are involved in the transition process and determine the stability of final conformations. According to their distinct properties, the interactions can be grouped into two classes: short-range interactions between amino acids which are near neighbors in sequence; and long-range interactions that involve amino acids remote in sequence but spatially close [4].

The α-helix formation involves only short-range interactions, and can be studied under the framework of phase transition theory [1,2,5,6]. The first such kind of theory for the helix-coil transition was proposed by Zimm and Bragg (ZB) in the late 1950s [5], and then reformulated by Lifson and Roig in the early 1960s [6]. In the ZB model, there are two parameters that are essential for helix formation: the propagation parameter s to represent the successive hydrogen bonding inside an α helix, and the factor σ for the initiation of a helix. Accordingly, the partition function can be constructed by enumerating all possible conformations of a chain. Then its statistical properties, including the helical content and the average number of α helices, are easily to be obtained from the partition function.

After the pioneer work of Zimm and Bragg, helix-coil transition theory has been greatly developed in the past fifty years [7–11]. In these studies, the parameters in the model have been carefully compared with experiments [12–14]. Additional interactions that may affect the stability of the α helix, such as terminal effects, capping motifs, helix dipoles, and side chain interactions, were widely studied by many authors [15–33]. Extensions to different kinds of helices, including the π and 3_{10} helix, have also been discussed [34].

When there are β structures in a polypeptide, the connecting patterns among the β strands can be very complicated because of the presence of long-range interactions. As a result, the developing of transition theories of β structures is still at the very beginning. And most existing works are limited to the β hairpin or antiparallel β sheet [35–48]. To our knowledge, no general consideration about the helix-sheet-coil transition is available, especially for the general case with parallel β sheets.

In this paper, we seek to develop a simple statistical mechanical model for the helix-sheet-coil transition in a generic homopolypeptide chain. Instead of providing a comprehensive understanding for the secondary structure formation, we will mainly focus on the role of long-range interactions in the β structure formation and helix-sheet transition.

The rest of the paper is organized as follows. The basic model is introduced in Sec. II, with detailed mathematical treatment of the partition function in Sec. III. Section IV presents the computational results for the transition from random coil and α helix to β sheet. The paper is concluded in Sec. V.

II. MODEL

For a statistical mechanical model, its central problem is to calculate the partition function. To this end, we introduce a systematical way to describe the possible conformations of a chain, which is done in two steps.

At first, we define the state of each amino acid residue, in other words, which one is in the α helix and which one is in the β strand. The definition of a residue in the α helix is straightforward, i.e., those residues whose NH group is bonded to the CO group of the fourth preceding residue. Here we assume that the hydrogen bonding of a residue, if it occurs in the α helix, is always to the fourth preceding one, and disregard other helical structures such as the π helix or 3_{10} helix [1,2,34]. To determine whether a residue is in the β strand is tricky, and depends on the state of its neighboring residues. Here we adopt a simple classification that a residue is regarded to be in the β strand, if the torsion angle pairs (ϕ,ψ) of itself and its two neighboring residues all take values in the upper left quadrant of the Ramachandran plot.

*hong-l04@mails.tsinghua.edu.cn
†jzlei@mail.tsinghua.edu.cn

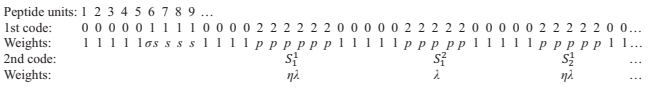

FIG. 1. A statistical mechanical model of polypeptide conformation.

Now, a chain of n residues can be described by a sequence of n symbols, each of which has one of three following values: digit 0 represents a random coil residue, 1 for a residue in the α helix, and 2 for a residue in the β strand. An example is shown by the first code in Fig. 1.

Knowing the state of each residue is not sufficient. We need to know how the β strands connect to each other to form β sheets. This is done by the second code shown in Fig. 1. We assign a symbol S_i^j for the jth strand in the ith β sheet. Here the number j is assigned not according to the sequence along the chain, but by the spatial connecting order in the β sheet. Now, the conformation of a chain can be described by two sequences of codes as shown in Fig. 1. It is easy to see that the second code is a consequence of the long-range interactions in the β sheet, which will be essential to helix-sheet transition.

Now, we introduce the statistical weights for the residues and β strands according to their codes shown in Table I. The physical meanings of every terms are given as follows. Then the relatively occurring probability for a given conformation is the product of these statistical weights along the chain.

(1) The factor 1 is arbitrarily assigned to random coil residues because only the relative ratio is effective.

(2) The helical propagation factor s measures the contribution of an α-helical residue to the partition function. It contains a decrease due to the restriction of torsion angle, and an increase due to the local hydrogen bonding. The value of s for a particular amino acid depends on its tendency to form a helix. In general, if the amino acid has a small side chain and is easy to form local hydrogen bonds, s will be slightly larger than unity; otherwise, s is smaller than unity [12–14].

(3) The helical initiation factor σ represents the decrease in the helical propagation factor s for the first unit in an α helix. This is because the formation of the first hydrogen bond needs extra restriction of torsion angles of residues between the bonded ones. According to Lifson-Roig model [6], this factor measures the boundary effect, and can be applied to both ends of an α helix. Additional capping parameters may also be considered [16–20]. However, here we only introduce the initiation factor at one end, and omit the capping parameters for simplicity.

TABLE I. Statistical weights in the model.

Code	Weight	State
0	1	coil residue
(0)1	σs	initiation of an α helix
(1)1	s	residue in the α helix
2	p	residue in the β sheet
S_i^1	$\eta\lambda$	initiation of a β sheet
$S_i^j (j \geqslant 2)$	λ	initiation of a β strand

(4) The strand propagation factor p measures the contribution of a residue in the β strand to the partition function. This factor contains a decrease due to the restriction of torsion angle, and an increase due to hydrogen bonding and hydrophobic effect. These interactions may occur between residues that are remote along the chain but close in space. Experimentally, the value of p for a particular amino acid can be obtained from its relative tendencies to form β sheets [49–51].

(5) The strand initiation factor λ measures the decrease in the strand propagation factor p for the first unit in a β strand. This is because the torsion angles of residues between the β strands are greatly restricted in order to connect two consecutive β strands to form the first pair of hydrogen bonds. Here, we introduce one initiation factor for each β strand, and also neglect the capping parameters for simplicity.

(6) The β-sheet initiation factor η measures the decrease in bonding effect for the first pair of β strands in a β sheet. Since the residues in the outer strands can only form one hydrogen bond, rather than two as those in the inner strands. Similar to the case of helix, one can also introduce an extra parameter for the last β strand in each β sheet. However, this will not affect the results and will be omitted here for simplicity.

In summary, we have introduced five tuneable parameters. The factors s and p measure the interactions on single residue in the α helix and β strand; the factors σ, λ, and η stand for boundary effects of the α helix, β strand, and β sheet, respectively. These effective model parameters are closely related to the chemical properties of the residues. Even small changes in their values may cause a large structure transition from one conformation to the other, just as what have been discovered in many proteins and biopolymers [1,2,52,53]. In general, s and p are sightly larger than unity, while σ, λ and η are less than unity.

In addition to the above five parameters, we introduce a parameter μ to represent the minimum number of coil residues to separate two consecutive secondary structure elements.

III. MATHEMATICAL TREATMENT

Based on the previous coding method for the conformations of a chain, we can construct the partition function of a homopolypeptide by enumerating all possible ways of arranging the digits 0,1,2 and symbols S_i^j. For each arrangement, its corresponding macroscopic state can be represented as $X=(n_a,n_b,l_a,l_b,k_b)$, where n_a, n_b, l_a, l_b, k_b are the number of α residues, β residues, α helices, β strands, and β sheets, respectively. The statistical weight of X is given by

$$Q(X) = s^{n_a} p^{n_b} \sigma^{l_a} \lambda^{l_b} \eta^{k_b}. \tag{1}$$

Let $S(X)$ be the number of conformations with the same macroscopic state X, and Ω be the set of all permissible conformations, then the partition function Z is formulated as

$$Z = \sum_{X \in \Omega} Q(X) S(X). \tag{2}$$

However, the explicit forms of $S(X)$ and Ω are very difficult to obtain, especially in the presence of β structures. In the

current study, all of our considerations are limited to homopolypeptides. We also neglect the stereochemical restrictions in the real polypeptides, such as the limitation on the length of loop segment between consecutive parallel β strands, and the constraint for the length of β strands in the same β sheets, etc. Furthermore, we ignore the possible energy difference between the parallel and antiparallel β sheets too. Despite the possible loss of exactness, these simplifications can provide some valuable insight into the role of long-range interactions in the β structure formation and the helix-sheet transition, and would also be helpful towards a complete model in the future.

After some tedious computations based on above assumptions, the explicit form of $S(X)$ is given by

$$S(X) = C_{n_a-2l_a-1}^{l_a-1} C_{n_b-l_b-1}^{l_b-1} C_{l_b-k_b-1}^{k_b-1} l_b!/2 \times C_{n-n_a-n_b-(\mu-1)(l_a+l_b)+\mu}^{l_a+l_b} C_{l_a+l_b}^{l_a} 2^{l_b-k_b}. \quad (3)$$

The meaning of each term in Eq. (3) is listed below.

(1) The first factor $C_{n_a-2l_a-1}^{l_a-1}$ gives the number of ways to partition n_a α residues into l_a helices, with the restriction that there are at least three residues in each helix. This term can be obtained by the classical combinatorics formula [54]. First of all, we take out $3l_a$ residues from the n_a α-residues in order to remove the constraint. Then partition the remaining (n_a-3l_a) identical residues into l_a groups freely (there may be empty groups). This process is equivalent to take (l_a-1) elements out of $(n_a-3l_a)+(l_a-1)$ units, which gives the combination $C_{n_a-2l_a-1}^{l_a-1}$. Finally, we divide the $3l_a$ residues equally into l_a helices to make sure that each helix has at least three residues.

(2) The second term $C_{n_b-l_b-1}^{l_b-1}$ is the number of ways to partition n_b β residues into l_b β strands, providing that each strand has at least two residues. Its derivation is similar to the one above.

(3) The third term $C_{l_b-k_b-1}^{k_b-1} l_b!/2$ represents the number of ways to arrange l_b β strands into k_b β sheets, with each β sheet having at least two strands. The factor $l_b!/2$ counts all possible permutations of the spatial connecting orders of β strands. Here, two arrangements with completely opposite ordering should be considered the same, as they represent the same spatial conformation.

(4) The factor $C_{n-n_a-n_b-(\mu-1)(l_a+l_b)+\mu}^{l_a+l_b}$ gives the number of ways to insert $(n-n_a-n_b)$ coil residues between the (l_a+l_b) secondary structure elements, so that consecutive elements are separated by at least μ coil residues. This term can be obtained in the same way as the first one. As we can see, there are $(n-n_a-n_b)$ coil residues to be assigned to either the (l_a+l_b-1) inner gaps between the secondary structure elements or the two ends of the chain. Each inner gap requires at least μ coil residues; while the chain ends can be empty. Similar to the previous method, we firstly take out $\mu(l_a+l_b-1)$ coil residues to remove the constraint. Then assign the remaining $[n-n_a-n_b-\mu(l_a+l_b-1)]$ residues to the (l_a+l_b+1) positions, which gives $C_{n-n_a-n_b-(\mu-1)(l_a+l_b)+\mu}^{l_a+l_b}$. Finally, we reinsert the $\mu(l_a+l_b-1)$ residues equally into (l_a+l_b-1) inner gaps to complete the whole process.

(5) The factor $C_{l_a+l_b}^{l_b}$ represents the number of ways to arrange the order of l_a α helices and l_b β strands, which is given by choosing l_a α helices from the total (l_a+l_b) secondary structure elements.

(6) The last term $2^{l_b-k_b}$ accounts the different connecting patterns of parallel and antiparallel β sheet. In each β sheet, when the initial β strand is fixed, the remaining strands can either parallel or antiparallel the previous one. And this gives the total $2^{l_b-k_b}$ combinations.

According to the chain structure, the partition function in Eq. (2) can be rewritten as

$$Z = 1 + Z_\alpha + Z_\beta + Z_{\alpha/\beta}, \quad (4)$$

where

$$Z_\alpha = \sum_{l_a=1}^{n} \sigma^{l_a} \sum_{n_a=3l_a}^{n} s^{n_a} C_{n_a-2l_a-1}^{l_a-1} C_{n-n_a-(\mu-1)l_a+\mu}^{l_a}, \quad (5)$$

$$Z_\beta = \sum_{k_b=1}^{n} \eta^{k_b} \sum_{l_b=2k_b}^{n} \lambda^{l_b} C_{l_b-k_b-1}^{k_b-1} 2^{l_b-k_b} l_b!/2 \times \sum_{n_b=2l_b}^{n} p^{n_b} C_{n_b-l_b-1}^{l_b-1} C_{n-n_b-(\mu-1)l_b+\mu}^{l_b}, \quad (6)$$

$$Z_{\alpha/\beta} = \sum_{k_b=1}^{n} \eta^{k_b} \sum_{l_b=2k_b}^{n} \lambda^{l_b} C_{l_b-k_b-1}^{k_b-1} 2^{l_b-k_b} l_b!/2 \times \sum_{n_b=2l_b}^{n} p^{n_b} C_{n_b-l_b-1}^{l_b-1} \sum_{l_a=1}^{n} \sigma^{l_a} C_{l_a+l_b}^{l_a} \times \sum_{n_a=3l_a}^{n} s^{n_a} C_{n_a-2l_a-1}^{l_a-1} C_{n-n_b-n_a-(\mu-1)(l_a+l_b)+\mu}^{l_a+l_b}. \quad (7)$$

Here we set $C_n^m = \frac{n!}{m!(n-m)!}$, if $n \geqslant 0$, $m \geqslant 0$ and $n \geqslant m$; otherwise, we set $C_n^m = 0$.

In Eq. (4), Z_α represents the partition function for all α structures $(n_b=0)$, Z_β for all β structures $(n_a=0)$, and $Z_{\alpha/\beta}$ for α/β mixed structures $(n_a, n_b \geqslant 1)$. If no β structure is presented, we have $Z=1+Z_\alpha$, which is consistent with the results in the ZB model [5].

Comparing Eqs. (5) and (6), we can see that the factor $C_{l_b-k_b-1}^{k_b-1} 2^{l_b-k_b} l_b!/2$ is introduced particularly for β structures. It reveals the entropy from different connective patterns of arranging β strands into β sheets. This effect is originated from the long-range interactions and can induce interesting transitions from the α helix to β sheet (to be detailed later).

According to statistical mechanics, the average number of α and β residues are given, respectively, as [5–11]

$$\langle n_a \rangle = \frac{\partial \ln Z}{\partial \ln s} = \frac{1}{Z} \sum_{X \in \Omega} n_a Q(X) S(X), \quad (8)$$

$$\langle n_b \rangle = \frac{\partial \ln Z}{\partial \ln p} = \frac{1}{Z} \sum_{X \in \Omega} n_b Q(X) S(X). \quad (9)$$

The average number of α helices and β strands are

$$\langle l_a \rangle = \frac{\partial \ln Z}{\partial \ln \sigma} = \frac{1}{Z} \sum_{X \in \Omega} l_a Q(X) S(X), \tag{10}$$

$$\langle l_b \rangle = \frac{\partial \ln Z}{\partial \ln \lambda} = \frac{1}{Z} \sum_{X \in \Omega} l_b Q(X) S(X). \tag{11}$$

Further formulas for n_a, n_b, l_a, and l_b can be obtained directly from the partial derivatives of Eqs. (4)–(7). As the detailed expressions are tedious, they will be omitted here.

IV. RESULTS

From previous discussions, in the absence of the β sheet, we reobtain the partition function in the ZB model for helix-coil transtion, which have been extensively studied [5–11]. In this paper, we will focus on the case with the presence of β structures, i.e., coil-sheet and helix-sheet transitions. Such transitions have been observed in many proteins and other biopolymers [1,2,51–53].

As we know, there are five parameters in our model. Despite the well studied helical parameters, no experimental values for the β structure parameters are available so far [55–61]. In this study, we refer to the helical parameters of the amino acid Leu, which is found to have nearly equal tendency to form the α helix and β sheet [51,62–64], with $s=1.14$ and $\sigma=0.0033$. Thus, we will roughly assign $p=s=1.14$ and $\lambda=\sigma=0.0033$, since their physical meanings are similar: s and p are for interactions on single residue; σ and λ are for initiative effect. Here, as the results are insensitive to η (see Fig. 3 below) and μ (data not shown), we will take $\eta=1$, $\mu=3$ in later computations for simplicity.

A. Coil-sheet transition

In this section, we omit the α helix and study the transition from random coil to the β sheet, thus $Z=1+Z_\beta$. Now, we have three tunable parameters p, λ, and η. Typical dependence of $\langle n_b \rangle$ and $\langle l_b \rangle$ on these parameters are shown in Fig. 2.

From Figs. 2(a) and 2(c), we see that the number of β residues increases from zero to the maximum value of 197 monotonously as p increases. This indicates that as the bonding effect becomes stronger, more coil residues transit to β residues. When p is large enough, the chain tends to adopt a conformation of two long β strands, which are connected by a β turn with $\mu(=3)$ residues. The sharpness of transition depends on the parameters λ and η. The smaller λ and η are, the sharper the transition will be.

Unlike the monotonously increasing of the average number of β residues, the average number of β strands shows a different dependence on p. It increases first to reach the maximum at around the transition point $(p=1)$, and then decreases to $\langle l_b \rangle=2$ as p keeps increasing [Figs. 2(b) and 2(d)].

A notable fact in Fig. 2(a) is that when λ is small ($\lambda<0.01$), there is only one transition region at around $p=1$, which consists with the results in the ZB model for helix-coil transition. When λ is increasing, however, the transition region will separate into two parts with $p<1$ and $p>1$, respectively. The underlying mechanisms of these two transitions are different. In the first transition region $(p<1)$, the average number of β strands increases along with $\langle n_b \rangle$ [Fig. 2(b)]. It indicates that the chain prefers to adopt a conformation with many short β stands in order to gain larger entropy. The entropy is given by the term $C_{l_b-k_b-1}^{k_b-1} 2^{l_b-k_b} l_b!/2$ in the partition function [Eq. (6)]. It comes from the multiple ways of arranging the β strands into β sheets, which is a consequence of the long-range interactions. While in the other transition region $(p>1)$, the average number of β strands decreases quickly to 2. This suggests that when p is large enough, the bonding effect on a single residue takes over the entropy contribution. And then, the chain favors the conformation of one β sheet with two long β strands to maximize the number of β residues. So the presence of two transition regions is a consequence of the competition between enthalpy and entropy effect in protein structure formation, and is different from the traditional helix-coil transition.

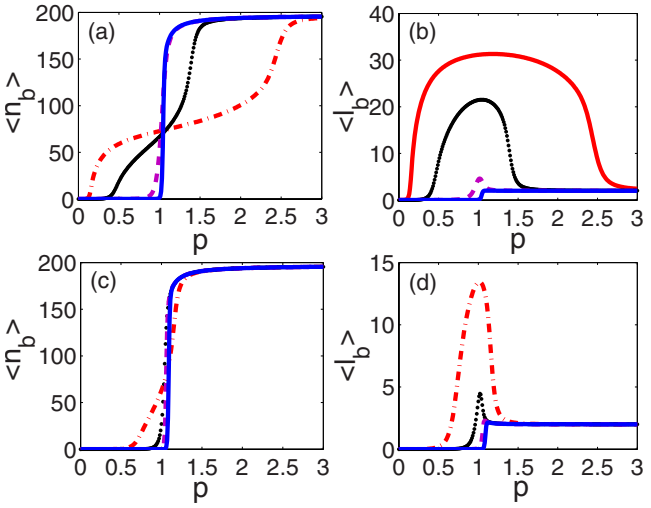

FIG. 2. (Color online) Coil-sheet transition. The dependence of the average number of β residues $\langle n_b \rangle$ and the average length of β strands $\langle l_b \rangle$ on the parameter p, with different values of λ or η. In the (a) and (b), $\eta=1$ and λ is changed ($\lambda=10^{-1}$: red dashed and dotted line; $\lambda=10^{-2}$: black dotted line; $\lambda=10^{-3}$: pink dashed line; $\lambda=10^{-4}$: blue solid line). In (c) and (d), $\lambda=0.0033$ and η is changed ($\eta=1$: red dashed and dotted line; $\eta=10^{-2}$: black dotted line; $\eta=10^{-4}$: pink dashed line; $\eta=10^{-6}$: blue solid line). The length of the chain is $n=200$.

B. Helix-sheet transition

Figure 3 shows the dependences of $\langle n_a \rangle$, $\langle n_b \rangle$, $\langle l_a \rangle$, $\langle l_b \rangle$ with respect to the model parameters p, λ, and η. We can see that there is evident transition of the chain conformation from the α helix to β sheet, when p or λ is increasing.

In Fig. 3(a), the chain transits from the α helix to β sheet, when the bonding effect p is increasing. The all α structure and all β structure are well separated. Around the transition point $(p \approx 1.15)$, small changes of the parameter p is able to force a peptide to change from a mostly helical state to a β-sheet dominated one. This hints that even single mutation of the amino acid is possible to induce a helix-sheet transition in natural protein, just as we have seen in $A\beta(1-42)$

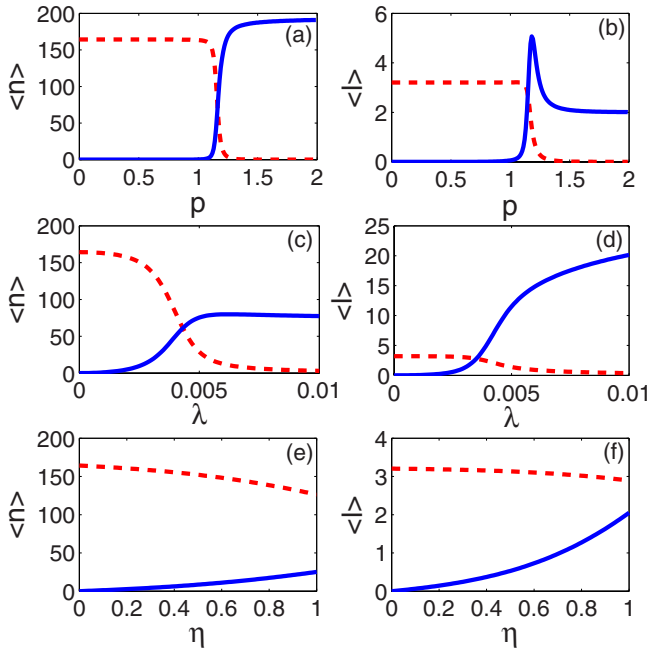

FIG. 3. (Color online) Helix-sheet transition. The dependence of $\langle n_a\rangle$, $\langle l_a\rangle$ (red dashed line) and $\langle n_b\rangle$, $\langle l_b\rangle$ (blue solid line) on the model parameters. The default parameters (except the one under studied) in the computation are $p=s=1.14$, $\lambda=\sigma=0.0033$, $\eta=1$. The chain length is $n=200$.

[53]. Further computations show that the sharpness of the transition depends on the values of λ and η (data not shown). The smaller λ and η are, the sharper the transition will be. This is similar to the case in sheet-coil transitions.

From Fig. 3(b), the average number of the α helix decreases with respect to p. The average number of β strands increases in the small p region. When reaching the maximum value of $\langle l_b\rangle \approx 5$ at $p \approx 1.2$, it then decreases to $\langle l_b\rangle = 2$ as p is increasing further. In particular, when p is small, the chain favors many short β strands; while p becomes larger, the chain prefers long β strands to maximize the number of β residues. Thus the average number of β strands decreases rapidly to 2. As we have pointed out in the case of coil-sheet transition, the turning point at $p \approx 1.2$ reveals the critical value where the enthalpy and entropy effects are exquisitely balanced by each other.

Figures 3(c) and 3(d) show the transition from α structure to β structure when the β strand initiation factor λ is increasing. The transition is smoother than that caused by the changes of p. Unlike Fig. 3(b) that shows a turning point at $p \approx 1.2$, $\langle l_b\rangle$ in Fig. 3(d) increases monotonously with parameter λ. This is because the larger β-strand initiation factor yields larger tendency to form β strands.

In Figs. 3(e) and 3(f), the transition from α structure to β structure is slow when the β-sheet initiation factor η is increasing. This result shows that the β sheet initiation factor η has minor effect on the helix-sheet transition.

V. DISCUSSIONS

In this paper, we have developed a simple statistical mechanical model for the helix-sheet-coil transition in generic homopolypeptide. We introduce a double coding method to describe the possible conformations of a chain, from which the partition function of a homopolypeptide is obtained when neglecting some stereochemical constraints. Though far from complete, our model has provided some physical insights about the helix-sheet-coil transition, especially the significant role of long-range interaction in β structure formation. Our computations show that the final equilibrium state of a homopolypeptide chain depends on the exquisite balance between short-range bonding interactions, long-range interactions, and entropy effects.

We must point out that the results in this study are only valid for homopolypeptides, and cannot be applied to heteropolypeptides directly. For example, Figs. 2 and 3 suggests that the chain may adopt long α helix or β strands when the bonding effect is strong enough. However this is not often seen in natural proteins. Instead the average length of the α helix in natural proteins is 11, and the length of the β strand is about 5–6 [52,65]. This is because on the one hand an α helix (or a β strand) can easily be interrupted by the residues with low tendency to adopt a helical (or strand) conformation in nature proteins. On the other hand, the tertiary structure, that we have neglected, may bring about strong geometrical constraint on the secondary structure. The long α helix or β strand are not stable alone for many other interactions, that are not considered in our model, such as the electronic interaction and hydrophobic interaction, etc. [1,2,51].

As we have pointed out that a comprehensive statistical mechanical description for the β structure formation is very difficult. Our current study is an attempt towards such an aim, although still at the very beginning. Further improvements need to be done before applying to real proteins.

ACKNOWLEDGMENTS

We thank Professor Kerson Huang and Professor C.C. Lin for their guidance and many helpful discussions.

[1] C. Branden and J. Tooze, *Introduction to Protein Structure* (Garland, London, 1998).
[2] R. H. Pain, *Mechanisms of Protein Folding* (Oxford University Press, Oxford, 2000).
[3] S. B. Prusiner, Science **252**, 1515 (1991).
[4] P. J. Flory, *Statistical Mechanics of Chain Molecules* (Hanser, Munich, 1988).
[5] B. H. Zimm and J. K. Bragg, J. Chem. Phys. **31**, 526 (1959).
[6] S. Lifson and A. Roig, J. Chem. Phys. **34**, 1963 (1961).
[7] D. Poland and H. A. Scheraga, *Theory of Helix-coil Transitions in Biopolymers* (Academic Press, New York, 1970).
[8] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of*

Macromolecules (AIP, New York, 1994).
[9] T. M. Birstein and O. B. Ptitsyn, *Conformations of Macromolecules* (Interscience, London, 1966).
[10] H. Qian and J. A. Schellman, J. Phys. Chem. **96**, 3987 (1992).
[11] A. J. Doig, Biophys. Chem. **101**, 281 (2002).
[12] C. A. Rohl and R. L. Baldwin, Biochemistry **36**, 8435 (1997).
[13] W. Shalongo, L. Dugad, and E. Stellwagen, J. Am. Chem. Soc. **116**, 2500 (1994).
[14] A. V. Finkelstein, Protein Eng. **8**, 207 (1995).
[15] A. J. Doig, A. Chakrabartty, T. M. Klingler, and R. L. Baldwin, Biochemistry **33**, 3396 (1994).
[16] A. Chakrabartty, A. J. Doig, and R. L. Baldwin, Proc. Natl. Acad. Sci. U.S.A. **90**, 11332 (1993).
[17] A. J. Doig and R. L. Baldwin, Protein Sci. **4**, 1325 (1995).
[18] N. H. Andersen and H. Tong, Protein Sci. **6**, 1920 (1997).
[19] D. A. E. Cochran, S. Penel, and A. J. Doig, Protein Sci. **10**, 463 (2001).
[20] D. A. E. Cochran and A. J. Doig, Protein Sci. **10**, 1305 (2001).
[21] M. Petukhov, V. Munoz, N. Yumoto, S. Yoshikawa, and L. Serrano, J. Mol. Biol. **278**, 279289 (1998).
[22] M. Petukhov, K. Uegaki, N. Yumoto, S. Yoshikawa, and L. Serrano, Protein Sci. **8**, 2144 (1999).
[23] E. T. Harper and G. D. Rose, Biochemistry **32**, 7605 (1993).
[24] J. M. Scholtz, H. Qian, V. H. Robbins, and R. L. Baldwin, Biochemistry **32**, 9668 (1993).
[25] W. Shalongo and E. Stellwagen, Protein Sci. **4**, 1161 (1995).
[26] B. J. Stapley, C. A. Rohl, and A. J. Doig, Protein Sci. **4**, 2383 (1995).
[27] M. Vasquez and H. A. Scheraga, Biopolymers **27**, 41 (1988).
[28] C. H. Roberts, Biopolymers **30**, 335 (1990).
[29] P. J. Gans, P. C. Lyu, P. C. Manning, R. W. Woody, and N. R. Kallenbach, Biopolymers **31**, 1605 (1991).
[30] H. Qian, Biophys. J. **67**, 349 (1994).
[31] A. Mitsutake and Y. Okamoto, J. Chem. Phys. **112**, 10638 (2000).
[32] Y. Peng, U. H. E. Hansmann, and N. A. Alves, J. Chem. Phys. **118**, 2374 (2003).
[33] H. Wako, N. Saito, and H. A. Scheraga, J. Protein Chem. **2**, 221 (1983).
[34] C. A. Rohl and A. J. Doig, Protein Sci. **5**, 1687 (1996).
[35] V. Munoz, E. R. Henry, J. Hofrichter, and W. A. Eaton, Nature (London) **390**, 196 (1998).
[36] D. K. Klimov and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **97**, 2544 (2000).
[37] A. R. Dinner, T. Lazaridis, and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **96**, 9068 (1999).
[38] W. L. Mattice, Annu. Rev. Biophys. Biophys. Chem. **18**, 93 (1989).
[39] C. W. David and R. Schor, J. Chem. Phys. **42**, 2156 (1965).
[40] C. W. David, H. B. Haukaas, J. G. Kalnins, and R. Schor, Biophys. J. **7**, 505 (1967).
[41] R. Schor, H. B. Haukaas, and C. W. David, J. Chem. Phys. **49**, 4726 (1968).
[42] V. Munoz, E. R. Henry, J. Hofrichter, and W. A. Eaton, Proc. Natl. Acad. Sci. U.S.A. **95**, 5872 (1998).
[43] W. L. Mattice and H. A. Scheraga, Biopolymers **23**, 1701 (1984).
[44] W. L. Mattice and H. A. Scheraga, Biopolymers **23**, 2879 (1984).
[45] W. L. Mattice and H. A. Scheraga, Biopolymers **24**, 565 (1985).
[46] S. J. Chen and K. A. Dill, J. Chem. Phys. **103**, 5802 (1995).
[47] S. J. Chen and K. A. Dill, J. Chem. Phys. **109**, 4602 (1998).
[48] J. K. Sun and A. J. Doig, J. Phys. Chem. B **104**, 1826 (2000).
[49] D. L. Minor Jr and P. S. Kim, Nature (London) **367**, 660 (1994).
[50] D. L. Minor Jr and P. S. Kim, Nature (London) **371**, 264 (2002).
[51] G. D. Fasman, *Prediction of Protein Structure and the Principles of Protein Conformation* (Plenum, New York, 1989).
[52] A. M. Lesk, *Introduction to Protein Architecture* (Oxford University Press, Oxford, 2003).
[53] K. Janek, S. Rothemund, K. Gast, M. Beyermann, J. Zipper, H. Fabian, M. Bienert, E. Krause, and E. Krause, Biochemistry **40**, 5457 (2001).
[54] R. A. Brualdi, *Introductory Combinatorics* (Pearson/Prentice Hall, New York, 2004).
[55] A. Chakrabartty, T. Kortemme, and R. L. Baldwin, Protein Sci. **3**, 843 (1994).
[56] L. Wang, T. O. Connell, A. Tropsha, and J. Hermans, Proc. Natl. Acad. Sci. U.S.A. **92**, 10924 (1995).
[57] S. Patel, P. V. Balaji, and Y. U. Sasidhar, J. Pept. Res. **13**, 314 (2007).
[58] J. A. Vila, R. L. Williams, J. A. Grant, J. Wojcik, and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **89**, 7821 (1992).
[59] H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **97**, 13075 (2000).
[60] J. A. Vila, D. R. Ripoll, and H. A. Scheraga, Biopolymers **58**, 235 (2001).
[61] J. Wojcik, K. H. Altman, and H. A. Scheraga, Biopolymers **30**, 121 (1990).
[62] H. A. Scheraga, J. A. Vila, and D. R. Ripoll, Biophys. Chem. **101-102**, 255 (2002).
[63] H. A. Scheraga, in *Perspectives in Structure Biology*, edited by M. Vijayan, N. Y. athindra, and A. S. Kolaskar (Indian Academy of Sciences, Bangalore, 1999), pp. 275–292.
[64] P. Y. Chou and G. D. Fasman, Biochemistry **13**, 211 (1974).
[65] Srinivas Aluru, *Handbook of Computational Molecular Biology* (CRC Press, Boca Raton, FL, 2006).