Exactly solvable model for helix-coil-sheet transitions in protein systems

John S. Schreck[*] and Jian-Min Yuan[†]

Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104, USA
(Received 27 January 2010; revised manuscript received 29 April 2010; published 17 June 2010)

In view of the important role helix-sheet transitions play in protein aggregation, we introduce a simple model to study secondary structural transitions of helix-coil-sheet systems using a Potts model starting with an effective Hamiltonian. This energy function depends on four parameters that approximately describe entropic and enthalpic contributions to the stability of a polypeptide in helical and sheet conformations. The sheet structures involve long-range interactions between residues which are far in sequence, but are in contact in real space. Such contacts are included in the Hamiltonian. Using standard statistical mechanical techniques, the partition function is solved exactly using transfer matrices. Based on this model, we study thermodynamic properties of polypeptides, including phase transitions between helix, sheet, and coil structures.

PACS number(s): 87.15.Cc, 87.15.A−, 64.60.De

In the late 1950s and early 1960s, Zimm and Bragg (ZB) and Lifson-Roig (LR) studied helix-coil transitions of simple models of homopolypeptides by employing rigorous statistical methods based on partition functions and transfer matrices [1]. In the 1970s and 1980s, these models were extended to include copolymers and medium-ranged interactions, and were used to characterize the experimental results of all amino acids and many proteins [2]. Because of the close coupling between the theoretical and experimental studies, ZB, LR, and related models have stimulated much interest in helix-coil transitions [3], which is still an active field of research up to the present time [4,5]. For reviews, see Ref. [2]. However, conformation changes of polypeptides involving sheet structures, such as helix-sheet transitions, are not as well characterized as for helix-coil transitions. In the late 1970s, using a multistate model, Tanaka and Scheraga [6] considered extended and chain-reversal states in addition to helix-coil transitions. In Ref. [7], medium-range interactions were taken into account to study helices, extended structures, and coils. More recently, Mattice and Scheraga [8], Sun and Doig [9], Hong and Lei [10], and others have included sheet structures in statistical models for homopolypeptides. The difficulties in constructing models for sheets lie primarily in the interactions between residues that are long-range in sequence but are close in physical space, and in the rich variety of structures associated with sheets, turns, and loops, thus a large number of parameters required for their description. In this paper, we introduce a simple statistical mechanical model for helix-coil-sheet transitions of homopolypeptides, starting with an effective Hamiltonian. Instead of an Ising-like model, the treatment is built on a multistate Potts model, which is capable of explicitly describing some of the long-range interactions exhibited by sheet structures. One of the objectives is that this simple model extends the helix-coil treatments to protein systems with three or more secondary structures.

An important step in a statistical mechanical approach such as ZB, LR, Ising, and Potts models is to construct the partition function for the system, based on which all thermodynamic properties are obtainable. As in ZB and LR models, partition functions factorize in terms of transfer matrices. However, ZB or LR theories start with a combinatorial partition function without defining an effective Hamiltonian. More generally, if an energy function $H(i)$ is defined, where $i=(i_1, \ldots, i_n)$ and i_n is the microstate of the nth residue which could occupy one of q possible states (conformations) labeled as $\{1, 2, \ldots, q\}$, the partition function for a system of N residues with periodic boundary conditions reduces to

$$Z_N = \sum_{i_1=1}^{q} \sum_{i_2=1}^{q} \cdots \sum_{i_N=1}^{q} e^{-\beta H(i)} = Tr(T^N), \qquad (1)$$

where $\beta=(k_B T)^{-1}$, k_B is Boltzmann's constant, and Tr is the matrix trace operation. The dimension of a transfer matrix in a one-dimensional (1D) Ising model is 2×2 and for a q-state Potts model, the dimension of a transfer matrix is $q \times q$. For Potts models with long-range interactions of range L along a 1D chain, as Glumac and Uzelac [11] showed in their formulation, the dimension of a transfer matrix becomes $q^L \times q^L$. Equation (1) may be further simplified by diagonalizing the transfer matrix T.

More recently, Hamiltonians of polypeptide chains have been described using a variety of Ising-like models [4,12,13] and Potts models [14,15], and also using an *ab initio* model [5]. In particular, the WSME model [12,13] uses two terms to construct an effective Hamiltonian and partition function: (1) the free energy term associated with the entropic cost of forming a pair of native residue conformations with restricted dihedral angles and (2) an enthalpic term associated with solvent-mediated contact energies between residues. Thus, residues may be either native or denatured, but not specific enough to distinguish sheets from helices. Our approach to polypeptides is based on a Potts model, where residues could assume many conformations including sheet, helix, coil, and turn. Before discussing the full helix-coil-sheet system, let us consider the simpler case of helix-coil transitions where an effective ($q=2$) Potts Hamiltonian (free energy in reality) can be written for a protein consisting of N residues as
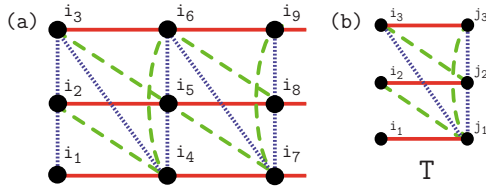
*jss74@drexel.edu
†yuanjm@drexel.edu

FIG. 1. (Color online) (a) Graphical representation of the partition function for the case $L=3$. The black dots mark the locations of particles along the chain. The dotted (blue) lines, K_1, are nearest-neighbor interactions. The dashed (green) lines, K_2, are next nearest-neighbor interactions. Solid (red) lines, K_3, are the $L=3$ interactions. (b) Graphical representation of the transfer matrix T.

$$-\beta H_{hc} = h_1 \sum_{n=1}^{N} \delta(1,i_n) + \beta J_1 \sum_{n=k+1}^{N} \delta(i_{n-k},i_n)\prod_{j=0}^{k-1} \delta(1,i_{n-j}),$$
(2)

where we assign $i_n=1$ to a residue in helix conformation and $i_n=2$ to a residue in coil conformation. The subscript "hc" in $-\beta H_{hc}$ means "helix-coil" and "1" in h_1 and J_1 refers to helix. The meanings of these parameters are similar to those described in the WSME model, where $h_1<0$ refers to an entropic cost from converting a coil to a helical residue, and $J_1>0$ refers to a contact energy between residues. In the present article, contact energies J_i are free-energies associated with solvent-mediated interactions, including hydrogen bonds, van der Waals, polar interactions, etc. The Kronecker delta $\delta(1,i_n)$ yields one if the nth residue is helical, and zero otherwise. In the second term of Eq. (2), $\delta(i_{n-k},i_n)$ yields one when residues at positions $n-k$ and k have the same conformation and are in contact, and zero otherwise. Contacts only contribute to the energy function when the participating residues are helical. Additionally, residues between two contacting residues must also all be helical. Thus, the range k determines the range of interaction. In α-helices, where k equals 4, residues at positions $n-4,n-3,\ldots,n$ are all helical when an H-bond forms between the $(n-4)$th and nth residues. When $k=1$, the effective Hamiltonian becomes $-\beta H_{hc} = h_1\Sigma_{n=1}^{N}\delta(1,i_n)+\beta J_1\Sigma_{n=2}^{N}\delta(1,i_n)\delta(i_{n-1},i_n)$. The second term in Eq. (2) is also similar to the Hamiltonian of the GMPC

model, which is a microscopic theory for helix-coil transitions based on a q-state Potts model [14,16].

To write down an effective Hamiltonian suitable for β-sheets, we need to include in it interactions up to length L along the polypeptide chain. Such a Hamiltonian can be constructed by adapting the long-range spin model of Glumac and Uzelac [11]. For a chain of N spins, their Hamiltonian can be written as

$$-\beta H = \sum_{l=1}^{L} \sum_{n=1}^{N} \beta K_l \delta(i_n,i_{n+l})$$
(3)

where K_l is distance-dependent. Figure 1(a) illustrates a graphical representation of the $L=3$ case and facilitates the construction of transfer matrices for long-range Potts systems. For Potts systems on a 1D lattice, Glumac and Uzelac grouped the spins along a chain into columns of height L, the longest interaction length, transforming a long-range problem of spin interactions into a short range one relating nearest-neighbor columns of height L [11,17], illustrated in Fig. 1(b). Each column of spins represents a vector that can take on one of q^L possible states. The transfer matrix thus has dimension $q^L \times q^L$. The various lines in Fig. 1 represent interactions $K_1,K_2,\ldots,K_L$ in Eq. (3), and contribute to the partition function when the arguments in the Kronecker delta's are equal.

Two modifications are made to apply the Glumac-Uzelac method of constructing transfer matrices to a protein system. Figure 2(a) illustrates a segment of an antiparallel β-sheet, where interactions can occur between residues which are remote in relative chain position, but are nearby in space. This is what is meant by 'long-range' in protein systems. Thus, the long-range nature of a protein system comes from labeling the residues according to the sequence order and does not come from the spatial distance between two residues. Even with the difference in the definition of long-range-ness, the Glumac-Uzelac method can be used in solving the protein problem. The strengths of interactions between each residue-residue pair are similar and not dependent on the relative chain position l. This is a main difference between our Hamiltonian [see Eq. (5) below] and Eq. (3). For simplicity, in this article we shall consider all contacts between
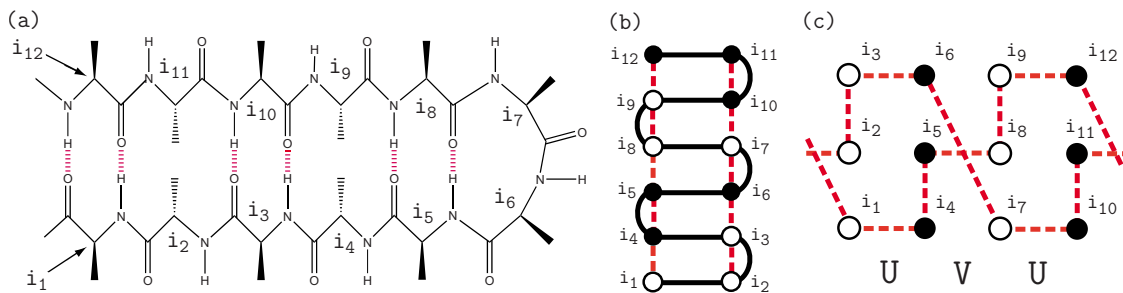

FIG. 2. (Color online) (a) A segment of an $L=11$ antiparallel β-sheet chain. The sequence position of a residue is labeled and H-bonds are referenced by the dashed lines. (b) A simple pattern illustrating repeating $L=3$ and nearest-neighbor contact interactions, denoted by dashed (red) lines. The solid (black) lines represent peptide bonds. In (c), a diagram representing the partition function for the structure in (b). The first column in (c) are residues i_1,i_2,i_3, the second column are residues i_4,i_5,i_6, etc. Contacts are represented by dashed lines. The color of residues comprising the columns alternate in color from white to black, which corresponds to the residue pattern in (b). Repeated multiplication of matrices U and V generates the partition function for the whole chain.

β-strands are of the same strength. In making this modification, Eq. (3) is recast as $-\beta H = \beta K \Sigma_{l=1}^{L} \Sigma_{n=1}^{N} \delta(i_n, i_{n+l})$, which drops the l-dependence of K, but maintains the long-range nature of the Kronecker interactions.

Second, according to Fig. 2(a), two hydrogen bonds form between residue-residue pairs, which occur for every other residue along a strand terminating at the turn. On the other hand, the residues along the β-strand that are not involved in hydrogen bonds with the opposite β-strand, could be involved in hydrophobic interactions with the opposite strand. To simplify the model, we assume that every residue-residue pair along neighboring strands forms contacts of the same strength, as stated above. The following pattern then represents H-bonding or hydrophobic interactions between two residues along neighboring strands, which we identify as contacts: $i_1 \rightarrow i_{1+L}, i_2 \rightarrow i_{1+L-1}, \cdots, i_{(L+1)/2} \rightarrow i_{(L+1)/2+1}$. In the present work, the turn conformation is also counted as a sheet conformation, but, in principle, the model can be extended to include specifically turn conformations if $q > 3$. The Kronecker delta's given in Eq. (3) are then modified to represent the aforementioned sheet-pattern. Additionally, for protein systems where the neighboring strands have the same interaction length L, the number of strands M, and the total number of residues N are related by

$$N = MR, \quad R = (L+1)/2. \tag{4}$$

We can write the two-state effective Hamiltonian for a pattern such as the one in Fig. 2(a) extended for any L, while taking into account the two modifications made to Eq. (3), as

$$-\beta H_{sc} = h_3 \sum_{n=1}^{N} \delta(3, i_n)$$
$$+ \beta J_3 \sum_{k=1}^{R} \sum_{m=1}^{M-1} b(i_{k,m}) \delta(i_{k+R(m-1)}, i_{1-k+R(m+1)}), \tag{5}$$

where we denote $i_n = 2$ (coil), or 3 (sheet), $b(i_{k,m}) \equiv \delta(3, i_{1-k+R(m+1)})$ and only allows J_3 terms to accumulate when the residues at position $k + R(m-1)$ and $1 - k + R(m+1)$ are locked in a sheet conformation and are in contact. The term $J_3 > 0$ now represents contacts between sheet residues, $h_3 < 0$ is the reduced entropic cost for coil to sheet conversions. The subscript "sc" in $-\beta H_{sc}$ refers to "sheet-coil" and subscript "3" in h_3 and J_3 refers to sheet. Unlike in Eq. (2), we do not require all residues between two residues in contact to be locked into the sheet state.

To see the general pattern described by the second term in Eq. (5), we start by considering the simplest $L = 3$ case as shown in Fig. 2(b). In reality, the minimal structure in Fig. 2(b) may not even be considered as a sheet structure, but nevertheless illustrates the general behavior that the transfer matrix can be decomposed into a product of subtransfer matrices. For $L = 3$ case, the transfer matrix decomposes into a product of two matrices U and V, as illustrated by Figs. 2(b) and 2(c). U and V are required to write out a general sequence of M strands and are explicitly written with the help of Fig. 2(c) as

$$\langle i|U|j \rangle = x^{\delta(i_1, j_1) + \delta(i_3, j_3) + \delta(j_1, j_2)},$$

$$\langle i|V|j \rangle = x^{\delta(i_2, j_2) + \delta(i_3, j_1) + \delta(j_2, j_3)}, \tag{6}$$

where $|i\rangle$ and $|j\rangle$ are neighboring column vectors of length L, where, for example, in Fig. 2(c), they can be $\langle i| = \langle i_1 i_2 i_3|$ and $|j\rangle = |i_4 i_5 i_6\rangle$, and $x = \exp\{\beta J_3\}$. Each transfer matrix U and V has dimension $q^L \times q^L$. This methodology works for any finite L, where the total number of transfer matrices needed to generate a periodic pattern for general L is found to be equal to the total number of interactions over the distance $L + 1$, which happens to equal the number R [18]. For example, for the $L = 3$ case illustrated in Fig. 2(c), there are two interactions, a nearest-neighbor [for example, i_2, i_3, in Fig. 2(c)] and one over the longest range of interaction [for example, i_1, i_4, in Fig. 2(c)] thus two matrices are sufficient. For illustrating purposes, we explicitly consider a simple model of antiparallel sheet-helix-coil systems, which starts with a three-state ($q = 3$) effective Hamiltonian with four parameters that can describe transitions between sheet, helix, and coil structures. Helical conformations are assumed to form contacts between nearest neighbors only, that is, the $k = 1$ case of Eq. (2). The total effective Hamiltonian can be written as

$$-\beta H_{\text{hcs}} = -\beta H_{hc} - \beta H_{sc}, \tag{7}$$

where now $i_n = 1$, 2, or 3, refers to helix, coil, and sheet, respectively, and the subscript "hcs" in $-\beta H_{\text{hcs}}$ refers to "helix-coil-sheet." The partition function can be written in the form of Eq. (1), when periodic boundary conditions are imposed, and calculated using transfer matrices, similar to the $L = 3$ case as illustrated in Figs. 2(b) and 2(c).

The parameters h_i and J_i are chosen so that the helix state is the most stable conformation at the lowest temperature in the interested temperature range. The coil dominates at high temperatures, where contact energies become relatively weak compared to thermal fluctuations. The sheet is thus an intermediary state [19]. For some proteins, the sheet is seen as the most stable conformation at low temperature, where the helix conformation becomes an intermediary state [20]. Our model can accommodate this case as well as a variety of others with proper choices of parameters.

For systems with fixed numbers of residues, the partition function facilitates calculation of numerous thermodynamical quantities, such as the average energy, $\langle E \rangle$, the heat capacity, C, and the order parameters, Θ_i, which are the average fractional content of ith state among q conformations at a particular temperature. To calculate the partition function, we choose a multistranded β-barrel system, which serves as an example of a protein system satisfying periodic boundary conditions. Inserting Eq. (7) into Eq. (1) and differentiating, we have for such a system

$$C = \frac{\partial \langle E \rangle}{\partial T} = \frac{\partial}{\partial T} \left(k_B T^2 \frac{\partial \ln Z_N}{\partial T} \right) \quad \text{and} \quad \Theta_i = \frac{\partial \ln(Z_N)}{\partial J_i}, \tag{8}$$

respectively. In Fig. 3(a), the order parameters for helix, coil, and sheet are presented for the case $L = 11$, $M = 100$, and in Figs. 3(b) and 3(c), we plot the temperature dependence of
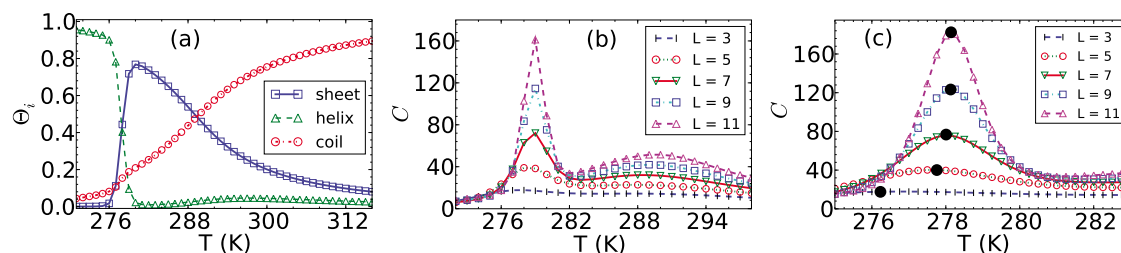
FIG. 3. (Color online) All calculated quantities using $J_1 = 2.85$ kcal/mol, $J_3 = 2.45$ kcal/mol, $h_1 = -4.91$, and $h_3 = -4.20$. a) Order parameters for the case $L = 11$ with $M = 100$. (b) Heat capacity (kcal/mol/K) vs T for various strand lengths L with $M = 100$. (c) The same plot as in (b) with more details of the helix-sheet transition given. Black dots denote transition temperatures which increase with range parameter L.

the heat capacity for various L cases. The heat capacity curve show two peaks: the sharp, low-temperature peak signifies the helix-sheet transition, and the broad, high-temperature peak signifies the sheet-coil transition. These peak positions are approximately given by the crossing points of Θ_i, shown in Fig. 3(a), between the helix and sheet and between the sheet and coil curves.

In conclusion, we have shown that, for a simple pattern associated with antiparallel β-sheet structures, an effective Hamiltonian using a minimal number of parameters and its corresponding partition function can be constructed to study its helix-coil-sheet transitions. The partition function can be exactly computed by means of transfer matrices, which are used to calculate thermodynamical properties of the system, including the order parameters for helices and sheets and the heat capacity, which show that increasing strand length, L, plays a stabilizing role in the protein.

[1] B. H. Zimm and J. K. Bragg, J. Chem. Phys. **31**, 526 (1959); S. Lifson and A. Roig, *ibid.* **34**, 1963 (1961).

[2] H. A. Scheraga, J. A. Vila, and D. R. Ripoll, Biophys. Chem. **101-102**, 255 (2002); A. J. Doig, *ibid.* **101-102**, 281 (2002).

[3] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic, New York, 1970).

[4] M. Takano, K. Nagayama, and A. Suyama, J. Chem. Phys. **116**, 2219 (2002).

[5] A. V. Yakubovich, I. A. Solov'yov, A. V. Solov'yov, and W. Greiner, Eur. Phys. J. D **46**, 215 (2008).

[6] S. Tanaka and H. Scheraga, Macromolecules **9**, 812 (1976); **10**, 9 (1977); **10**, 305 (1977).

[7] H. Wako, N. Saito, and H. A. Scheraga, J. Protein Chem. **2**, 221 (1983).

[8] W. L. Mattice and H. A. Scheraga, Biopolymers **23**, 1701 (1984).

[9] J. K. Sun and A. J. Doig, J. Phys. Chem. B **104**, 1826 (2000).

[10] L. Hong and J. Lei, Phys. Rev. E **78**, 051904 (2008); L. Hong, J. Chem. Phys. **129**, 225101 (2008).

[11] K. Uzelac and Z. Glumac, J. Phys. A **21**, L421 (1988); Z. Glumac and K. Uzelac, *ibid.* **22**, 4439 (1989).

[12] V. Muñoz, P. A. Thompson, J. Hofrichter, and W. A. Eaton, Nature (London) **390**, 196 (1997).

[13] V. Muñoz, E. R. Henry, J. Hofrichter, and W. A. Eaton, Proc. Natl. Acad. Sci. U.S.A. **95**, 5872 (1998); V. Muñoz and W. A. Eaton, *ibid.* **96**, 11311 (1999); P. Bruscolini and A. Pelizzola, Phys. Rev. Lett. **88**, 258101 (2002).

[14] N. S. Ananikyan, Sh. A. Hajryan, E. Sh. Mamasakhlisov, and V. F. Morozov, Biopolymers **30**, 357 (1990).

[15] R. E. Goldstein, Phys. Lett. **104A**, 285 (1984).

[16] A. V. Badasyan, A. Giacometti, Y. Sh. Mamasakhlisov, V. F. Morozov, and A. S. Benight, Phys. Rev. E **81**, 021921 (2010).

[17] Z. Glumac and K. Uzelac, J. Phys. A **26**, 5267 (1993).

[18] J. Schreck and J. M. Yuan (unpublished).

[19] F. Ding, J. M. Borreguero, S. V. Buldyrev, H. E. Stanley, and N. V. Dokholyan, Proteins **53**, 220 (2003).

[20] M. Andrec, A. K. Felts, E. Gallicchio, and R. M. Levy, Proc. Natl. Acad. Sci. U.S.A. **102**, 6801 (2005).