Simplified lattice model for polypeptide fibrillar transitions

Xuhui Xiao[1] and Ming-Chya Wu[2,3,*]

[1]*Department of Physics, National Taiwan University, Taipei 10617, Taiwan*

[2]*Research Center for Adaptive Data Analysis, National Central University, Chungli 32001, Taiwan*

[3]*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*

(Received 31 March 2014; revised manuscript received 21 May 2014; published 2 October 2014)

Polypeptide fibrillar transitions are studied using a simplified lattice model, modified from the three-state Potts model, where uniform residues as spins, placed on a cubic lattice, can interact with neighbors to form coil, helical, sheet, or fibrillar structure. Using the transfer matrix method and numerical calculations, we analyzed the partition function and construct phase diagrams. The model manifests phase transitions among coil, helix, sheet, and fibril through parameterizing bond coupling energy ε_h,ε_s,ε_f, structural entropies s_h,s_s,s_f of helical, sheet, and fibrillar states, and number density ρ. The phase diagrams show the transition sequence is basically governed by ε_h, ε_s, and ε_f, while the transition temperature is determined by the competition among ε_h, ε_s, and ε_f, as well as s_h, s_s, s_f, and ρ. Furthermore, the fibrillation is accompanied with an abrupt phase transition from coil, helix, or sheet to fibril even for short polypeptide length, resembling the feature of nucleation-growth process. The finite-size effect in specific heat at transitions for the nonfibrillation case can be described by the scaling form of lattice model. With rich phase-transition properties, our model provides a useful reference for protein aggregation experiments and modeling.

DOI: 10.1103/PhysRevE.90.042701 PACS number(s): 87.15.Cc, 87.15.A−, 64.60.De

I. INTRODUCTION

Amyloidosis is a general name for a group of amyloid-related diseases, including Alzheimer's disease, Creutzfeldt-Jakob disease, Parkinson's disease, Huntington's disease, diabetes mellitus type 2, etc. [1], which are known for their incurability and invariably fatality. Though the clinical symptoms of these diseases differ, they share similar characteristics that there are generally abnormal aggregated proteins that deposit in extracellular spaces of tissues and organs in the patient's body, which are usually toxic [2]. The depositions of amyloidosis are mainly composed of insolvable protein fibrillar structures, which are generally long, unbranched, and about 10 nm in diameter [3,4]. They appear as extensive β-sheet structures in the analysis of x-ray diffraction [4,5] and possibly aggregate to hairpinlike structures [6–10]. Since the aggregated proteins of various diseases differ markedly in the primary structures of native states [1], there must exist general rules for the aggregation processes and inherent structure formation.

Recently, there have been extensive studies on protein aggregation, both from experimental observations and theoretical modeling. Remarkably, analysis on the native state of several amyloid-like proteins shows that helices observed in experiments are predicted as β sheets by secondary structure predictors [11], implying helix-to-sheet transitions before fibrillation. Further, numerical simulations [12] on aggregation of separated non-native proteins to fibril suggest that aggregation may be the result of a nucleation-growth (NG) mechanism [13,14] and a template-assembly (TA) process [15–17]. It is proposed that nucleation speeds up aggregation, while the fibrillar surface forms templates to induce fibrillation of nonfibrillar proteins.

This study is aimed at constructing a model for protein aggregation capable of reproducing the properties mentioned above. The model assumes the effective interactions at the molecular level, incorporating features of multiple chain interactions, homogeneous residues, and helix-to-sheet, helix-to-fibril, and sheet-to-fibril transformations. We have been aware of the recent studies based on lattice models which offer physical pictures to understand the transition from "normal" structures to fibrils in a relatively simple but heuristic way. Among others, the Wako-Saitô-Muñoz-Eaton (WSME) model [18–22] is a course-grained but solvable model which assumes residues are basic units in native and non-native states and that interactions among residues exist only for native states. Following similar concepts, Schreck and Yuan [23] proposed an alternative based on the three-state Potts model [24–26], in which phase transition is possible among secondary structures and coil. However, they did not explicitly define the fibrillar structure in the model due to the tendency of fibrillation from a sheet structure and hence higher transition probability for the cases with β-sheet-rich structures. On the other hand, Zamparo *et al.* [27] extended the WSME model and introduced a toy (while exactly solvable) model for the transitions from the helical structure to the fibrillar structure. This model does not take into account the existence of sheet structures in an individual polypeptide, hence there is no transition from the "normal" sheet structure to the fibrillar structure. Here we propose to take advantage of the two models to introduce a new one to account explicitly for the coexistence of sheet and fibrillar structures. The model is essentially a three-state Potts model [24–26], constituted by multiple polypeptides, each of which consists of uniform residues occupying a finite number of lattice sites in a layer, and multiple layers form a three-dimensional lattice.

We analyzed the partition function of the model with the help of the transfer matrix method and from which we numerically calculated the average fractional contents of coil,

*mcwu@ncu.edu.tw

helix, sheet, and fibril and the specific heat of phase transitions. The phase diagrams were then constructed to explore the effects of bond coupling energy, structural entropy, and system concentration. For the case without a fibrillar phase, numerical calculations can be implemented on relatively larger system sizes (under the constraint of our computational power), and the finite-size effect in specific heat was also analyzed.

The paper is organized as follows. In Sec. II, a simplified model for polypeptide fibrillar transition is introduced. The exact expression of solution of the model, which is amenable to detailed numerical analysis, is derived in Sec. III. The parameter choices and procedures of numerical calculations are explained in Sec. IV, and the numerical results including phase diagrams of various typical realizations are presented and discussed in Sec. V. Finally, we conclude briefly in Sec. VI.

II. THE MODEL

The system under consideration is a plurality of polypeptides in solvent, and each polypeptide is a single chain consisting of a finite number of residues. Before fibrillation, a polypeptide may be composed of coils, or fold to form helices and sheets, while after the fibrillar transition, most polypeptides are collectively in a state composed of coils, helices, sheets, and fibrillar structure. To simulate the transition, we adapt the three-state Potts model [24–26], in which N polypetides are placed on a three-dimensional square lattice; each polypeptide consists of L uniform residues as spins occupying L lattice sites in a plane (see Figs. 1 and 2). A residue can be in a state σ, corresponding to the coil, helical, sheet, or hairpinlike fibrillar structure, specified by interactions to be introduced below. The system is open and properly described by the grand-canonical ensemble. The energy function is composed of the energy contribution from individual polypeptides, contact energy of fibrillar structure, entropic cost of fibrillation, and the chemical potential accounting for the polypeptides involved in the system. Detailed interactions including the H-bonds, van der Waals, polar interactions, and hydrophobic interactions are represented by effective structural bond coupling energies [23]. Then the partition

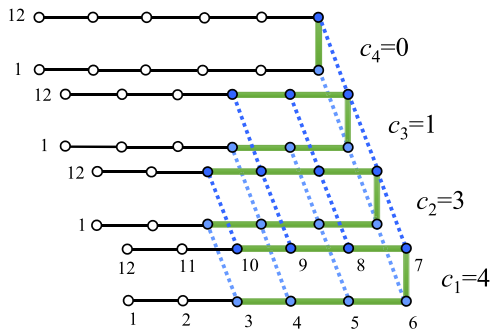

FIG. 1. (Color online) Schematic illustration of a typical hairpinlike fibril ($N = 4$, $L = 12$) on a cubic lattice of 4×2×6. Each polypeptide is placed on a distinct 2×6 square lattice. Fibrillar structure is highlighted by a bold line (green). Dotted lines (blue) represent fibrillar bonds.

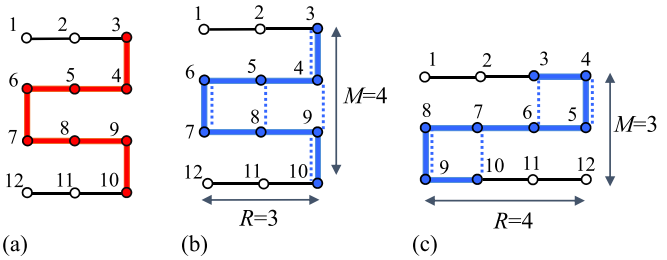

FIG. 2. (Color online) A typical polypeptide ($L = 12$) with secondary structures on a 3×4 square lattice. (a) A helix with length of eight residues is highlighted in red. Two examples of sheet structure are represented with blue, (b) the width $R = 3$ and maximum strand $M = 4$, and (c) $R = 4, M = 3$.

function of the system reads as

$$Z = \sum_{\{\sigma\}} \exp[-\beta H(\sigma)], \tag{1}$$

with

$$H = \sum_{n=1}^{N} \gamma_n H_n, \tag{2}$$

$$H_n = Q(m_n, c_n, c_{n-1}) - \left[2c_n\varepsilon_f + \frac{1}{\beta}S_f(c_n)\right] - \mu, \tag{3}$$

where $\beta = 1/k_BT$, $k_B = 1.38\times10^{-23}$ (J/K) is the Boltzmann's constant, H is the energy function of the system, and H_n is the energy function of the nth polypeptide. For the lth residue of the nth polypeptide, the state variable σ is conveniently characterized by the parameters γ_n, $m_n(l)$, and c_n. γ_n is a binary; $\gamma_n = 1$ if the nth polypeptide is involved in the system and 0 otherwise. Hence, $N_\gamma = \sum_{n=1}^{N}\gamma_n$, a nonfixed quantity, is the number of polypeptides involved in the system. Q is the energy function for structures within a polypeptide. $m_n(l)$ specifies the structure where the lth residue of the nth polypeptide locates, and $m_n(l) = 0$ for coil, 1 for helix, and 2 for sheet or fibril. $2c_n\varepsilon_f$ represents that there are $2c_n$ fibrillar bonds between two adjacent, nth and $(n+1)$th polypeptides, each has a bond coupling energy ε_f, and here the free boundary condition $c_0 = c_N = 0$ is considered. For a hairpinlike structure, c_n satisfies $0 \leqslant c_n \leqslant L/2$. S_f is the entropic cost for structure conversion of a polypeptide in fibrillation, and $S_f(c_n = 0) = 0$, $S_f(c_n \neq 0) = s_f L$, under the assumption of most of the entropic loss at the formation of just two fibrillar bonds (i.e., $c_n = 1$) [27] and the simplification of length dependence of S_f. μ is the chemical potential of the system.

The explicit form of Q is specified by the model. Here we assume that a coil does not have any bond coupling among residues, a helix is constituted by at least four consecutive residues, and sheets and fibrils have at least two bonds. Figure 2 illustrates typical helical and sheet structures for a polypeptide of length $L = 12$ in a plane. The bonds of helix, sheet, and fibril are specified by distinct bond coupling energies. Further specifications include the width of sheet which is to be fixed throughout the calculation of a realization, and the arrangement of fibrillar bonds, which is assumed to distribute

sequentially and symmetrically from the middle [$(L/2)$th and $(L/2+1)$th residues] to the two ends (1st and Lth residues) (see Fig. 1). Then Q is written as

$$Q = -\varepsilon_h \sum_{k=1}^{L-3}\prod_{l=k}^{k+3}\delta(m_n(l),1) - \frac{1}{\beta}s_h\sum_{r=1}^{L}\delta(m_n(r),1) - \varepsilon_s\sum_{r=1}^{R}\sum_{p=1}^{M-1}\delta(m_n(u_{r,p}),2)\delta(m_n(u_{r,p}),m_n(v_{r,p})) - \frac{1}{\beta}s_s\sum_{r=1}^{L}\delta(m_n(r),2) - E(c), \quad (4)$$

where ε_h and ε_s are bond coupling energies for helical and sheet structures, respectively, and s_h and s_s stand as the entropies for helical and sheet structures, respectively. R is the width of the sheet in the range from 1 to $L/2$ (L even), and M is the maximum number of strands for a polypeptide of length L, defined via the ceiling function $M = \lceil L/R \rceil$ that takes the smallest integer $M \geqslant L/R$. u and v are residue sequence numbers of sheets in the given polypeptide and are defined as $u_{r,p} = 1 - r + R(p+1)$ and $v_{r,p} = r + R(p-1)$. The last term, $E(c)$, where $c = \max(c_{n-1}, c_n) = 0, 1, \ldots, L/2$, is the energy of the parts that have formed fibrillar structures. If $c = 0$, then $E = 0$, indicating that no resides in the given polypeptide are a part of the fibril. If $c \neq 0$, then $E \neq 0$ and it must be subtracted from Q. The function E reads as

$$E = -\varepsilon_h \sum_{l=\max(\frac{L}{2}-2-c,1)}^{\min(\frac{L}{2}+c,L-3)}\prod_{k=l}^{l+3}\delta(m_n(k),1) - \varepsilon_s\sum_{r=1}^{R}\sum_{p=1}^{M-1}\delta(m_n(u),2)\delta(m_n(u),m_n(v)) + \varepsilon_s\sum_{r=1}^{R}\sum_{p=1}^{M'-1}\delta(m_n(u),2)\delta(m_n(u),m_l(v)) + \varepsilon_s\sum_{r=1}^{R}\sum_{p=1}^{M'-1}\delta(m_n(u'),2)\delta(m_n(u'),m_n(v')), \quad (5)$$

where $M' = \lceil L'/R \rceil$, $L' = (L-2c)/2$, $u' = u + L/2 + c$, and $v' = v + L/2 + c$. In other words, when a polypeptide is partially involved in fibrillation, part ($2c$) of its sheet structures form fibrillar bonds with neighboring polypeptides, while the remaining parts are symmetric segments that can be regarded as two shorter (L') polypeptides with helical and/or sheet structures. The two shorter segments are to be considered as they were independent polypeptides. Further, the fibrillar structures are essentially in sheet state, such that the function E does not involve entropic contribution.

III. CALCULATION OF THE PARTITION FUNCTION

In this section, we calculate the partition function of our model, following the scheme in Refs. [23,27], which exploits the transfer matrix method. The transfer matrix method, illustrated in Fig. 3, recombines the configuration sum of the partition function into boundary and inner parts. The

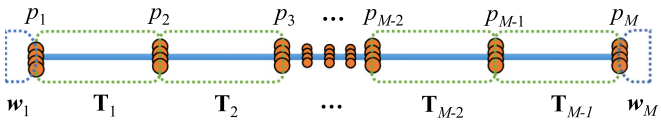

FIG. 3. (Color online) Illustration of the recombination scheme of the transfer matrix method. The partition function of the system is to be represented as a product of boundary parts $\mathbf{w}_1(p_1)$ and $\mathbf{w}_M(p_M)$, and the transfer matrix $\mathbf{T}_i(p_i, p_{i+1})$. p_i is a parameter used to label the sequence.

boundary parts $w_1(p_1)$ and $w_M(p_M)$ can be represented as a $1\times\aleph$ matrix, while the inner part is the transfer matrix $\mathbf{T}_i(p_i, p_{i+1})$ with dimension $\aleph\times\aleph$. Such recombination can be achieved on the basis of the property of short-range interaction and translational symmetry. The calculation of the partition function then reduces to the calculation of the transfer matrix.

Let us first observe our model for the case of $\gamma_n = 0$; there is no need to consider $m_n(l)$, and it is equivalently set to $m_n(l) = 0$ for all l, denoted by $\delta_{m_n 0}$ for the nth polypeptide. This is taken into account to avoid overestimating the configuration sum of the partition function by rewriting Eq. (1) as

$$Z = \sum_{\{m_n(l)\}}\sum_{\{c_n\}} G(\{m_n(l)\},\{c_n\}) \quad (6)$$

and

$$G = \sum_{\{\gamma_n\}}\prod_{n=1}^{N}\left[\gamma_n + (1-\gamma_n)\delta_{m_n 0}\right]\exp(-\beta H_n) = \prod_{n=1}^{N}\left[\delta_{m_n 0} + \exp\beta(\mu - Q)\right]\exp[2\beta c_n\varepsilon_f + S_f(c_n)]. \quad (7)$$

Furthermore, fibrillar bonds c_n will only form in sheet state when $c_n \leqslant B_{n,n+1}$,

$$B_{n,n+1} = \min(\varphi_n, \varphi_{n+1}), \quad (8)$$

with

$$\varphi_n = \sum_{i=1}^{\frac{L}{2}}\prod_{l=i}^{L+1-i}\delta(m_n(l),2). \quad (9)$$

Then Eq. (7) is written as

$$G = \prod_{n=1}^{N}\left[\delta_{m_n 0} + \exp\beta(\mu - Q)\right]\exp(2\beta c_n\varepsilon_f + S_f) \times\theta(B_{n,n+1} - c_n), \quad (10)$$

where $\theta(x)$ is the Heaviside step function, satisfying the condition $\theta(x \geqslant 0) = 1$ and $\theta(x < 0) = 0$. The function B implies that the fibrillar structure grow symmetrically to a width c_n from the turning loop to the two ends of the polypeptide. θ can be expressed alternatively as

$$\prod_{n=1}^{N}\theta(B_{n,n+1} - c_n) = \prod_{n=1}^{N}\theta(\varphi_n - c_n)\prod_{n=1}^{N}\theta(\varphi_{n+1} - c_n) = \prod_{n=1}^{N}\theta(\varphi_n - c), \quad (11)$$

and we have

$$G = \prod_{n=1}^{N} \theta(\varphi_n - c)\left[\delta_{m_n 0} + \exp\beta(\mu - Q)\right] \times \exp(2\beta c_n \varepsilon_f + S_f). \quad (12)$$

It is straightforward to sum over $m_n(l)$ in Eq. (6) and to follow the recombination scheme of Fig. 3 to express the partition function as

$$Z = \sum_{\{c_n\}} w_1(c_1) T(c_1,c_2)\cdots T(c_{N-2},c_{N-1}) w_{N-1}(c_{N-1}) = (\mathbf{w}_1, \mathbf{T}^{N-2}\mathbf{w}_{N-1}), \quad (13)$$

where $\mathbf{w}_n$ is a $1\times(L/2+1)$ matrix with the entry

$$w_n = \Psi(c_n)\exp\left[\beta c_n \varepsilon_f + \tfrac{1}{2} S_f(c_n)\right], \quad (14)$$

taking the value $c_n = 0,1,\ldots,L/2$. $\mathbf{T}$ is the transfer matrix with entries $T(c_{n-1},c_n)$ defined as

$$T = \Psi(c)\exp\left\{\beta\varepsilon_f(c_{n-1}+c_n) + \tfrac{1}{2}[S_f(c_{n-1}) + S_f(c_n)]\right\}, \quad (15)$$

with

$$\Psi(c) = \begin{cases} 1 + e^{\beta\mu}\tilde{Z}_L, & \text{if } c = 0, \\ e^{\beta\mu + 2cs_s}\tilde{Z}^2_{\frac{L}{2}-c}, & \text{if } 1 \leqslant c \leqslant L/2, \end{cases} \quad (16)$$

where $\tilde{Z}_{\tilde{L}}$ is the partition function of the homogeneous and noninteractive polypeptide with length $\tilde{L}$. The first case of $c = 0$ in Eq. (16) corresponds to the situation of no fibrils in the system, and the second case refers to the coexistence of fibril and secondary structures and thus $\tilde{L}$ is the length of one of the two equal segments of the nonfibrillar parts and is not necessarily an even number. For convenience, either the full-length polypeptides or remaining segments of nonfibrillar parts are regarded as independent polypeptides. Using p_i consisting of 3^R elements to indicate the strand number of sheets, the expression of $\tilde{Z}_{\tilde{L}}$ can be obtained by modifying directly Eq. (1), and, again in the recombination form of Fig. 3,

$$\tilde{Z}_{\tilde{L}} = \sum_{\{m\}} \exp[-\beta Q(m,0)] = \sum_{\{m\}} \tilde{w}_1(p_1)\tilde{T}(p_1,p_2)\cdots\tilde{T}(p_{\tilde{M}},p_{\tilde{M}+1})\tilde{w}_{\tilde{M}+1}(p_{\tilde{M}+1}) = (\tilde{\mathbf{w}}_1, \tilde{\mathbf{T}}^{\tilde{M}}\tilde{\mathbf{w}}_{\tilde{M}+1}), \quad (17)$$

where $\tilde{\mathbf{w}}$ is a 1×3^R matrix with the entry $\tilde{w}_{p+1}(p+1)$,

$$\tilde{w}_{p+1} = \exp\left\{\frac{1}{2}\sum_{r=pR+1}^{(p+1)R}[s_h\delta(m(r),1) + s_s\delta(m(r),2)]\right\}, \quad (18)$$

taking the value $m = 0,1,2$, $\tilde{M} = \lceil \tilde{L}/R \rceil$, and $\tilde{\mathbf{T}}$ is the transfer matrix of dimension $3^R\times 3^R$, with entries $\tilde{T}_{p+1}(p+1,p+2)$ and $p = 0,\ldots,\tilde{M}-1$,

$$\tilde{T}_{p+1} = \exp\left\{\sum_{r=pR+1}^{(p+1)R}\left[\beta\varepsilon_h\prod_{n=0}^{3}\delta(m(r+n),1)\right]\right\}$$

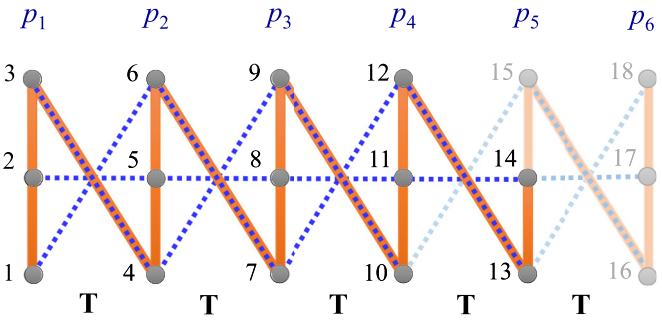

FIG. 4. (Color online) Illustration for the treatment of the boundary with incomplete residue sets for the case of $\tilde{L} = 14$ and $R = 3$. The blue dotted line indicates interactions between residues in the sheet state, while the orange solid line indicates the helical state. Residues of number 15, 16, 17, and 18 do not exist in the system. The transfer matrices $\tilde{\mathbf{T}}$ are defined by the recombination scheme.

$$\exp\left\{\sum_{r=pR+1}^{(p+1)R}\beta\varepsilon_s\delta(m(r),2)\delta(m(2(p+1)R-r+1),2)\right\} \exp\left\{\frac{1}{2}\sum_{r=pR+1}^{(p+2)R}[s_h\delta(m(r),1) + s_s\delta(m(r),2)]\right\}, \quad (19)$$

where $m(r) = 0$ if $r > \tilde{L}$, to account for the overestimation of the incomplete residue sets at the boundary. For example, in the calculation of polypeptide chains with length $\tilde{L} = 14$ and sheet length $R = 3$ (see Fig. 4), we have

$$\tilde{Z}_{14} = \sum_{\{m\}} \tilde{w}_1(p_1)\tilde{T}(p_1,p_2)\cdots\tilde{T}(p_5,p_6)\tilde{w}_6(p_6) \quad (20)$$

and $m(15) = m(16) = m(17) = m(18) = 0$.

A. Average fractional contents of structures

Having the partition function of the system ready, one can now calculate thermodynamic variables by using Eq. (13). The grand potential density ϕ_N defined as

$$\phi_N = -\frac{1}{LN\beta}\ln Z, \quad (21)$$

is useful in our framework. At the thermodynamic limit of $N\to\infty$, ϕ_∞ (abbreviated as ϕ hereafter for simplicity) is calculated through the largest eigenvalue τ of the transfer matrix $\mathbf{T}$, i.e.,

$$\phi = -\frac{1}{L\beta}\ln\tau = -\frac{1}{L\beta}\ln(\bar{\mathbf{w}},\mathbf{T}\bar{\mathbf{w}}), \quad (22)$$

where $\bar{\mathbf{w}}(c)$ is the corresponding eigenvector. The number density ρ, defined as the average number of polypeptides $\langle N_\gamma\rangle$ involved in the system divided by the system size $N\to\infty$, is the derivative of the grand potential density ϕ with respect to the chemical potential μ,

$$\rho \equiv \lim_{N\to\infty}\frac{\langle N_\gamma\rangle}{N} = -L\frac{\partial}{\partial\mu}\phi(Q,S_f,\mu), \quad (23)$$

which can be calculated through the derivative of the transfer matrix in Eq. (15),

$$\rho = \frac{1}{\beta\tau}\left(\bar{\mathbf{w}}, \frac{\partial \mathbf{T}}{\partial \mu}\bar{\mathbf{w}}\right) = 1 - \frac{\bar{w}^2(0)}{\tau}, \tag{24}$$

where $\bar{w}(0)$ takes the value of the first entry of the eigenvector $\bar{\mathbf{w}}$. Further, the average fractional content of the fibril P_f, coil P_c, helix P_h, and sheet P_s of the system, defined as the probabilities of the corresponding structures to show up in the system, can be calculated according to

$$\begin{aligned} P_x &= -\frac{1}{\rho}\frac{\partial}{\partial \lambda}\phi(Q - \lambda\zeta_x, S_f, \mu)|_{\lambda=0} \\ &= \frac{1}{L\beta\rho\tau}\left(\bar{\mathbf{w}}, \frac{\partial \mathbf{T}(\lambda,\zeta_x)}{\partial \lambda}\bar{\mathbf{w}}\right)_{\lambda=0}, \end{aligned} \tag{25}$$

where the subscript x denotes one of fibril (f), coil (c), helix (h), and sheet (s). The function ζ_x is defined as

$$\zeta_f(c) = 2c, \tag{26}$$

for fibril and

$$\zeta_q(m,c) = \sum_{k=1}^{\frac{L}{2}-c} \delta(m(k),q) + \sum_{k=\frac{L}{2}+c+1}^{L} \delta(m(k),q), \tag{27}$$

with $q = 0,1,2$ for coil, helix, and sheet, respectively. For the case of P_f, only $c \neq 0$ in Eq. (16) is considered. Then the relation

$$\frac{\partial}{\partial \lambda} T(\lambda, c_{n-1}, c_n)|_{\lambda=0} = \beta\zeta_f(c) T(c_{n-1}, c_n) \tag{28}$$

holds, and we have

$$P_f = \frac{1}{L\rho\tau}(\bar{\mathbf{w}}, \zeta_f \mathbf{T}\bar{\mathbf{w}}). \tag{29}$$

For the case of P_q, we have to consider secondary structures in polypeptides with and without segments in fibrillar structures. Accordingly, we have

$$P_q = \frac{e^{\beta\mu}}{L\beta\rho\tau}\left(\bar{\mathbf{w}}, \frac{\mathbf{T}}{\Psi}\frac{\partial \tilde{Z}_L(\lambda,\zeta_q)}{\partial \lambda}\bar{\mathbf{w}}\right)_{\lambda=0} \tag{30}$$

if $c = 0$ and

$$P_q = \frac{2e^{\beta\mu + 2cs_s}}{L\beta\rho\tau}\left(\bar{\mathbf{w}}, \frac{\mathbf{T}}{\Psi}\tilde{Z}_{\frac{L}{2}-c}\frac{\partial \tilde{Z}_{\frac{L}{2}-c}(\lambda,\zeta_q)}{\partial \lambda}\bar{\mathbf{w}}\right)_{\lambda=0} \tag{31}$$

if $c = 1 \sim L/2$, where

$$\left.\frac{\partial \tilde{Z}_{\tilde{L}}}{\partial \lambda}\right|_{\lambda=0} = \sum_{\{m\}} \beta \sum_{k=1}^{\tilde{L}} \delta(m(k),q) \exp\left(-\beta Q(m(k),0)\right), \tag{32}$$

which can be calculated by plugging $\beta\sum_{k=1}^{\tilde{L}}\delta(m(k),q)$ into Eq. (17). Subsequent calculations will be carried out by numerical approach in the next section.

B. Specific heat

Another interesting quantity, the specific heat of the system at thermodynamic limit, is calculated through the Helmholtz free energy, represented here in terms of the grand potential ϕ and the number density ρ,

$$\frac{c_V}{k_B} = \beta^2 \frac{\partial^2}{\partial \beta^2}\left(-\beta\phi - \beta\mu\frac{\rho}{L}\right). \tag{33}$$

Substituting the results of ϕ in Eq. (22) and ρ in Eq. (24), we have

$$\frac{c_V}{k_B} = \frac{\beta^2}{L}\frac{\partial^2}{\partial \beta^2}\left(\ln\tau + \beta\mu\frac{\bar{w}^2(0) - \tau}{\tau}\right). \tag{34}$$

This quantity is to be analyzed numerically.

IV. PARAMETER CHOICES AND PROCEDURES OF NUMERICAL CALCULATIONS

In our numerical analysis, the values (or ranges of the values) of some parameters are determined in advance. These include two structural entropies (s_h, s_s), three bond coupling energies (ε_h, ε_s, ε_f), and sheet width (R). Before preceding to the results section, here we explain how the parameters are chosen and the procedures of numerical calculations.

The structural analysis involves secondary structure formations and corruptions, which are associated with entropic change. When a polypeptide folds to a secondary structure, it reduces configuration entropy due to the lost of spacial degree of freedom. For a nonspecified sequence, we determine the entropic change as a result of a structural transition by calculating the size difference of the configuration space of the structure with respect to the entire allowed domain in a Ramachandran plot [28]. In a typical statistics on 121 870 residues from 463 structures [29] in the protein data bank [30], sheet structures have a larger configuration space than helices, such that $s_h < s_s$ is assumed for homogeneous sequence, while this is not necessarily true for a particular sequence. Specifically, under the assumption of homogeneous sequence, the Ramachandran plot of Fig. 5 in Ref. [29] is divided into a matrix of 36×36 cells; each cell represents a structure state. Then there are totally 910 cells for the allowed region. It follows that the entire allowed region (covering 910 cells) is assumed for coil, among which helices distribute in 58 cells and sheets occupy 84 cells. Thus, we have structural entropy $s_c = \ln(910/910) = 0$ for coils, $s_h = \ln(58/910) = -2.7530$ for helices, and $s_s = \ln(84/910) = -2.3826$ for sheets. There is no preliminary knowledge to determine the fibrillar structural entropy, thus different values of structural entropy of fibril s_f will be used in this work.

Furthermore, the bond coupling energy $\varepsilon_h, \varepsilon_s, \varepsilon_f$ parameterizing competitive interactions in structure formation are properly chosen to have the transitions located around room temperature. This reduces the parameter space to a significantly smaller one in which phase-transition behaviors are usually typical and one of them will be presented for demonstration. To further simplify parameter choices and for comparison purpose, the same values of parameters will be used in different cases if changing parameters are not necessary. Meanwhile, as mentioned in Sec. II, a particular value of R is selected and then fixed in the calculation of a realization.

Within the parameter space, we calculate the transfer matrix $\mathbf{T}$ of Eq. (15), from which the largest eigenvalue τ and

the corresponding eigenvector $\bar{\mathbf{w}}$ are obtained. Then, using Eqs. (23) and (24), we derive the polypeptide number density ρ and four average fractional contents (P_c, P_h, P_s, P_f), expressed by Eqs. (29)–(31). Finally, we calculate the specific heat c_V numerically according to Eq. (34), using the central difference method.

V. RESULTS AND DISCUSSIONS

Now we present the results of numerical calculations of the average fractional contents and the specific heat of the transitions among secondary structures and fibril. To demonstrate the average fractional contents of structures of polypeptides consisting of residues which can be in various state configurations, a system is called an x phase if the average fractional content of a structure P_x exceeds 50%.

A. The case without fibrillation ($s_f = 0, \varepsilon_f = 0$)

First, we consider the case without fibrillation, in which there is no fibrillar structural entropy ($s_f = 0$) and fibrillar bond coupling energy is zero ($\varepsilon_f = 0$), and the number density $\rho = 1.0$ is set. When $\varepsilon_f = 0$ is considered, only coil, helical, and sheet structures are involved in the system, and the system reduces to the model proposed by Schreck and Yuan in Ref. [23]. We chose $\varepsilon_h = 1.827$ kcal/mol and $\varepsilon_h = 1.566$ kcal/mol such that there are two transitions around room temperature: one at around 280 K corresponds to the helix-to-sheet transition, while the other at around 295 K is for the sheet-to-coil transition. This generally represents that our model can reproduce the results of Ref. [23]. Specifically, Fig. 5 shows heat capacity per polypeptide $L \times c_V / k_B$ and the average fractional contents P_x of the system as functions of temperature T for sheet width $R = 3,4,5,6$ and polypeptide length $L = 300,400,500,600$. Here R and L have been chosen to have a product of 100 transfer matrices. The general features are, with the decrease of temperature, that the system from high to low temperatures undergoes a first transition from coil to sheet and then a second transition from sheet to helix. With the increase of system size, the height of the transition curve becomes higher and higher, and the sheet-to-helix transition temperature shifts from 278 K to 280 K, and the coil-to-sheet transition temperature shifts from 294 K to 295 K [see the insets in Fig. 5(a)]. The much smaller change in height and shift in transition temperature for the coil-to-sheet transition compared to the sheet-to-helix transition is a consequence of only a relatively smaller amount of structure changes and a slower process with respect to temperature change of the former.

It is interesting that the structure change from sheet to helix is mostly around the transition temperature [see Fig. 5(b)], such that there is a sharp peak in specific curves [Fig. 5(a)]. For comparison, Fig. 6 shows the specific heat as a function of temperature T for different sheet widths $R = 3,4,5,6$ but a fixed polypeptide length $L = 256$ and number density $\rho = 1.0$. The longer the sheet width R, the shaper the first transition peak (at around 280 K) and the more flat and smooth the second transition (at around 295 K). This is a result of the corruption of more sheet bonds for larger R in a transition from sheet to helix such that more energy (heat) is needed for such a transition.

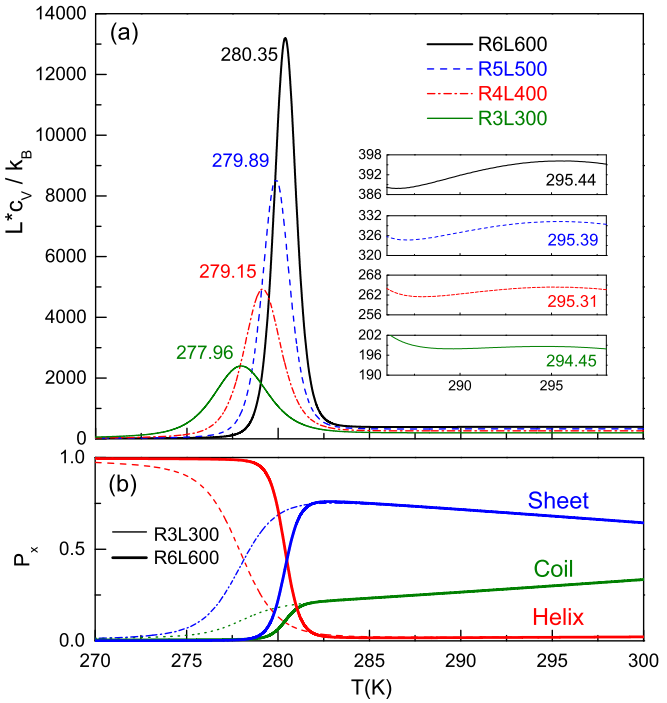

FIG. 5. (Color online) (a) Heat capacity per polypeptide $L \times c_V / k_B$ and (b) the average fractional contents P_x of the system as functions of temperature T for sheet width $R = 3,4,5,6$ and polypeptide length $L = 300,400,500,600$. Here the structural entropy $s_h = -2.7530, s_s = -2.3826, s_f = 0$, bond coupling energy $\varepsilon_h = 1.827$ kcal/mol, $\varepsilon_s = 1.566$ kcal/mol, $\varepsilon_f = 0$, and number density $\rho = 1.0$. Each realization has two phase transitions: one at around 280 K is for the helix-to-sheet transition, and the other at around 295 K is for the sheet-to-coil transition. The insets in (a) are enlarged view for the second transition. The transition temperatures are indicated near curves.

Figure 7 shows the phase diagram of the case without fibrillation at absolute temperature $T = 290$ K, with fixed polypeptide length $L = 256$, structural entropy $s_h = -2.7530$,

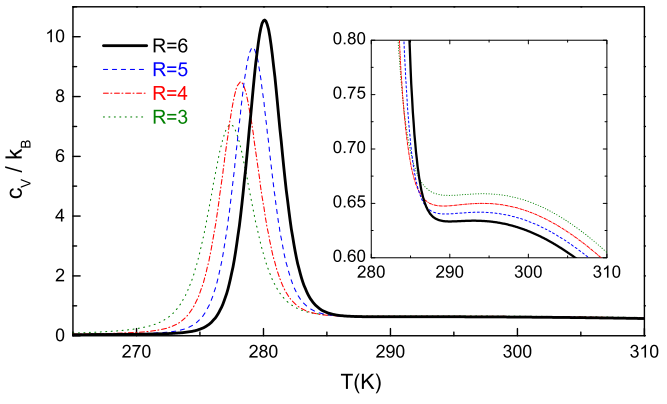

FIG. 6. (Color online) Specific heat of the case without fibrillation, as a function of temperature T, for sheet width $R = 3,4,5,6$ and fixed polypeptide length $L = 256$. Here the structural entropy $s_h = -2.7530$, $s_s = -2.3826, s_f = 0$, and bond coupling energy $\varepsilon_h = 1.827$ kcal/mol, $\varepsilon_s = 1.566$ kcal/mol, $\varepsilon_f = 0$, and number density $\rho = 1.0$. The inset is an enlarged view showing the second transition.

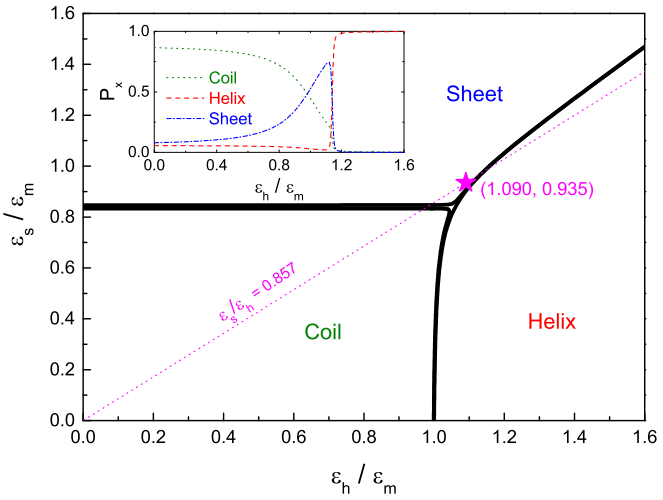

FIG. 7. (Color online) The phase diagram of the system without fibrillation ($s_f = 0$ and $\varepsilon_f = 0$) at absolute temperature $T = 290$ K and fixed polypeptide length $L = 256$. Here structural entropy $s_h = -2.7530, s_s = -2.3826$, and number density $\rho = 1.0$. The bold black lines denote phase boundaries. A small region enclosed by the boundaries corresponds to undefined phase with the average factional content $P_x < 50\%$ for coil, helix, and fibril. $\varepsilon_m = 1.675$ kcal/mol. The slope of the phase boundary between the helix and sheet at the large-ε_h regime is 1. The inset shows P_x as function of $\varepsilon_h/\varepsilon_m$, taken along the magenta dashed line with slope $\varepsilon_s/\varepsilon_h = 0.857$. The magenta star (1.090,0.935) on the dashed line represents the case with bond coupling energy $\varepsilon_h = 1.827$ kcal/mol and $\varepsilon_s = 1.566$ kcal/mol.

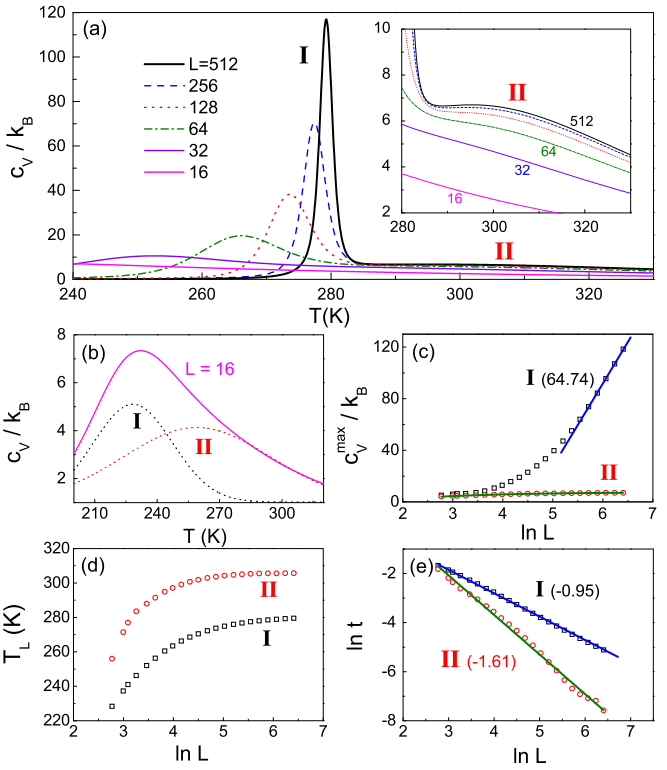

FIG. 8. (Color online) Finite-size scaling analysis of the transition temperatures and specific heat for the system without fibrillation. (a) Specific heat c_V/k_B as a function of temperature T for different polypeptide length L. (b) Decomposition of the specific-heat curve into curve I and curve II, corresponding to transitions I and II, respectively. (c) Maxima of specific heat $c_V^{\max}/k_B$ and fittings. (d) Transition temperatures T_L as functions of L. (e) Scaling of the transition temperatures. t is the reduced temperature defined as $t = (T_\infty - T_L)/T_\infty$. Here structural entropy $s_h = -2.7530$, $s_s = -2.3826, s_f = 0$, bond coupling energy $\varepsilon_h = 1.827$ kcal/mol, $\varepsilon_s = 1.566$ kcal/mol, $\varepsilon_f = 0$, fixed sheet width $R = 3$, and number density $\rho = 1.0$. The inset in (a) is an enlarged view of the second (II) transition at around 296 K. The numbers in (c) and (e) denote the slops of the fittings.

$s_s = -2.3826$, and number density $\rho = 1.0$. The phase boundary is determined by the average fractional content $P_x \leqslant 50\%$. The region surrounded by the three phases is an undefined phase where $P_x < 50\%$ for helix, sheet, and coil. The slope of the phase boundary between helix and sheet at large ε_h is 1, which is a result of the same maximum number of residues in helix and sheet for a given polypeptide length. For convenience, we rescaled the axes by ε_m determined through normalizing the phase boundary between the helical and coil phases at $\varepsilon_s = 0$ to 1. The boundaries of helix and sheet merge at $\varepsilon_h/\varepsilon_m = 1.051$ and $\varepsilon_s/\varepsilon_m = 0.852$. A typical composition of the system in a chemical may correspond to the dashed line in the diagram. Specifically, this line represents the ratio of bond coupling energies $\varepsilon_s/\varepsilon_h = 0.857$. By adjusting the chemical condition, the bond coupling energy changes and structure phase transitions take place. For example, along the dashed line in the phase diagram, the P_x of coil becomes lower and lower as the ratio $\varepsilon_h/\varepsilon_m$ increases. With the decrease of the fractional content of coil, more sheet structure forms before $\varepsilon_h/\varepsilon_m = 1.15$, corresponding to a first transition from coil to sheet. Interestingly, there is an abrupt change of P_x for a decrease of coil and sheet, while P_h increases from very small (at around 0) to very large (at around 1.0) when $\varepsilon_h/\varepsilon_m$ approaches 1.15. The abrupt changes in P_s and in P_h represent sharp structure phase transitions (a second transition corresponding to a sheet-to-helix transition) at this ratio, which manifests a thermodynamical-limit-like behavior of a finite-size system for a particular range of $\varepsilon_h/\varepsilon_m$. It is clear that from the phase diagram that only the system with particular parameters has two phase transitions. Such a parameter choice must have two cross sections with two phase boundaries, such as the situation represented by the dashed line in Fig. 7. The magenta star on the magenta dashed line represents the case with $\varepsilon_h = 1.827$ kcal/mol, $\varepsilon_s = 1.566$ kcal/mol, used in the plots of Figs. 5, 6, and 8, where there are two phase transitions for each specific-heat curve.

Furthermore, the properties of the transition peaks are also associated with the polypeptide length. Figure 8(a) shows the specific heat c_V/k_B as a function of temperature T for different polypeptide length L. As previously mentioned, the specific-heat curve involves two phase transitions, which are visually distinguishable when L is large enough. Using two-peak curve (Gaussian function) fitting, each specific-heat curve can be decomposed into two curves: one corresponds to the first (I) transition at lower temperature around 280 K and the other to the second (II) transition at higher temperature about 295 K. Figure 8(b) shows a typical decomposition of the specific-heat curve of $L = 16$. The transition temperatures T_L of two transitions are shown in Fig. 8(d), and the heights of the

maxima of specific-heat curves $c_V^{\max}/k_B$ as functions of L are shown in Fig. 8(c). Note that though the difference between $c_V^{\max}/k_B$ of the first (I) and second (II) transitions increases with L due to the fact that the growth of $c_V^{\max}/k_B$ of the former is much faster, the variation of the difference between the two transition temperatures is generally small.

To see the finite-size effects of the system, we analyzed the specific-heat curves with different polypeptide lengthes, and the results are shown in Fig. 8(c) and Fig. 8(e). The transition temperatures of $L=\infty$ are determined through extrapolation and in this case $T_\infty^I = 281.22 \pm 0.07$ K and $T_\infty^{\rm II} = 305.87 \pm 0.25$ K for the two transitions, respectively. Employing the concept of the scaling function of lattice model [31–34], we assumed a scaling form involving $\ln L$ and $1/L$ terms for the maxima of specific-heat curves of the first (I) transition. We used a linear fitting for the curve at large L regime to determine the coefficient of $\ln L$ and a curve fitting for the curve at small L regime to determine the coefficient of $1/L$, and the result is

$$\frac{c_V^{\max}}{k_B} \approx (64.74 \pm 0.68)\ln L - (296.8 \pm 4.1) + (2248 \pm 65)\frac{1}{L}. \tag{35}$$

In contrast, the maxima of the specific-heat curves of the second (II) transition saturates at large L. We found that $c_V^{\max}/k_B$ as a function of temperature is well described by

$$\frac{c_V^{\max}}{k_B} \approx (7.27 \pm 0.02) - (30.0 \pm 1.8)L^{-0.75\pm0.03}. \tag{36}$$

Likewise, the scaling form of the reduced temperature is determined through the assumption of $t = (T_\infty - T_L)/T_\infty \sim L^{-\nu}$. Using the linear fitting in the form of

$$\ln t \approx \text{const} - \nu \ln L, \tag{37}$$

for both transitions, we determined $\nu = 0.95 \pm 0.01$ for transition I and $\nu = 1.61 \pm 0.03$ for transition II [see Fig. 8(e)]. Since our model is not a primary three-state Potts model, there is no reference for the values of ν.

The correlation length of the helical and sheet structures were further analyzed using the definition of the correlation function $g(r) = \langle\delta(m(i),q)\delta(m(i+r),q)\rangle - \langle\delta(m(i),q)\rangle\langle\delta(m(i+r),q)\rangle$ on a polypeptide. $g(r)$ has a maximum g_0 at $r=1$, which will be used here as a self-normalization factor. Figures 9(a) and 9(b) show the normalized correlation function $g(r)/g_0$ of the system with polypeptide length $L=128$ at different temperatures for helical and sheet structures. $g(r)/g_0$ roughly decays in an exponential form for a realization at temperature far from the transition temperature $T_{128} \approx 273.5$ K. When T approaches T_L, $g(r)/g_0$ decreases more quickly before $r=L$. This scenario is clear in Fig. 9(c), where $g(r)/g_0$ for the helical structure at T_L is plotted for different L. $g(r)/g_0$ at T_L for $L=64$ to 512 share the same function form (have partial overlap). Consequently, the correlation length $\xi(T)$ is calculated through fitting a preassumed function

$$\frac{g(r)}{g_0} \propto f(r)\exp\left[-\frac{r}{\xi(T)}\right], \tag{38}$$

where $f(r)$ is a function characterizing the interactions in the system. The correlation length $\xi(T)$ as a function of T is shown in Fig. 9(d). Both $\xi(T)$ of helix and sheet reach maxima at the transition temperature $T_{128} \approx 273.5$ K. These results indicate that the correlation length at the transition temperature exceeds the polypeptide length even for a small system. Due to the limit of our computation power, our numerical data are insufficient to determine the explicit function form of $f(r)$.

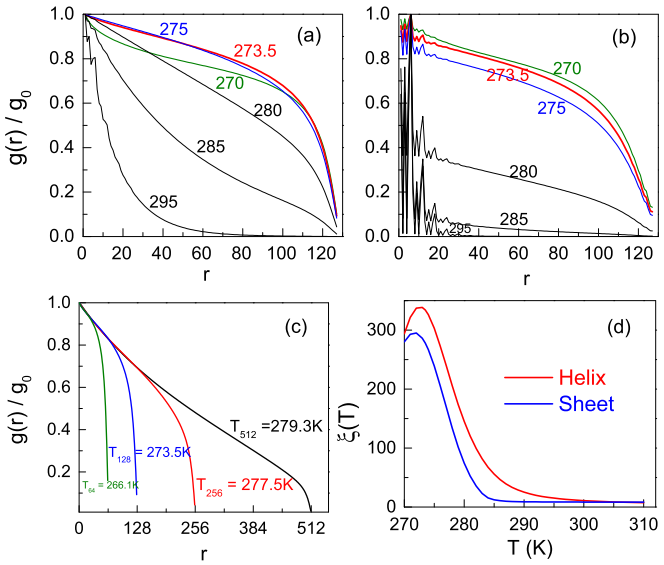

FIG. 9. (Color online) The normalized correlation function $g(r)/g_0$ as a function of separation r between two residues for the system with polypeptide length $L=128$ at different temperatures T for (a) helix and (b) sheet. The transition temperature is $T_{128} = 273.5$ K. (c) $g(r)/g_0$ as a function of r at the transition temperature T_L for different L. (d) Correlation length $\xi(T)$ as a function of T for the system with $L=128$.

B. The case with fibrillation and $s_s \neq 0, \varepsilon_s = 0$

To simulate the situation that fibrillation usually takes place for sheet-rich molecules, we consider the case where forming sheet structures is a signature of fibrillation. This is similar to the situation considered by Zamparo *et al.* in Ref. [27]. Again, our model can reproduce the features observed in Ref. [27]. Here we assumed $s_h = -2.7530, s_s = -2.3826, s_f = -0.75$, and $\varepsilon_s = 0$ and studied the effects of the competing bond coupling energy of ε_h and ε_f. Figure 10(a) shows the phase diagram of the system as a function of helical and fibrillar bond coupling energy $\varepsilon_h, \varepsilon_f$ at absolute temperature $T = 290$ K, with polypeptide length $L=16$ and number density $\rho = 0.6$. Note that though $\varepsilon_s = 0$ is assumed, the average fractional content of sheet P_s, by definition, is not necessarily zero but very small [see insets in Fig. 10(a)]. There is therefore no sheet phase in the phase diagram. The phase diagram generally consists of three regions of coil, helical, and fibrillar phases, separated by phase boundaries. There is an undefined region between the coil and helical phases, in which $P_x < 50\%$ for both coil and helix for the existence of minor sheet structures. Due to the limit of computing power, we present only the calculation of the case with $L=16$, and the axes have been rescaled by $\varepsilon_m = 1.995$ kcal/mol, determined through normalizing the phase boundary between the coil and helical phases at $\varepsilon_f = 0$ to 1. Even for such a

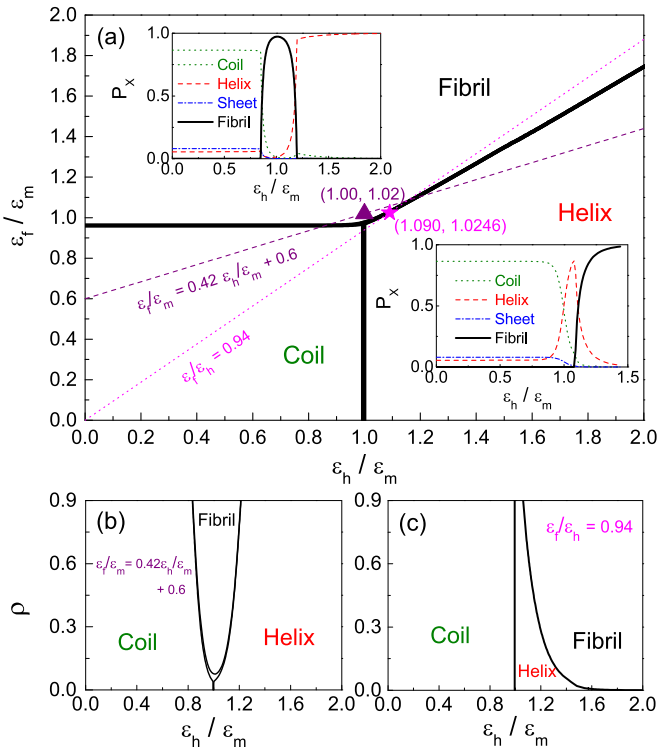

FIG. 10. (Color online) The phase diagram as functions of structure bond coupling energy and number density. (a) The phase diagram of the system as a function of normalized helical and fibrillar bond coupling energy $\varepsilon_h/\varepsilon_m, \varepsilon_f/\varepsilon_m$ at absolute temperature $T = 290$ K. Here the structural entropy $s_h = -2.7530, s_s = -2.3826$, $s_f = -0.75$, sheet width $R = 3$, and number density $\rho = 0.6$, polypeptide length $L = 16$, sheet width $R = 3$. $\varepsilon_m = 1.995$ kcal/mol. The slope of the phase boundary between helix and fibril at the large-ε_h regime is 13/16. The left upper inset in (a) shows the average fractional content P_x as a function of $\varepsilon_h/\varepsilon_m$, taken along the purple dashed line. The purple triangle on the dashed line represents the case with $\varepsilon_h/\varepsilon_m = 1.00$, $\varepsilon_f/\varepsilon_m = 1.02$. The right bottom inset is the same plot taken along the magenta dotted line. The magenta star on the dotted line locates at $\varepsilon_h/\varepsilon_m = 1.090$, $\varepsilon_f/\varepsilon_m = 1.0246$. (b) The phase diagram as function of ρ and $\varepsilon_h/\varepsilon_m$ along the purple dashed line in (a), and (c) is that along the magenta dotted line.

small system, rich phase-transition behaviors can be observed, including a single-phase transition and successive two-phase transitions. Depending on parameters of ε_h and ε_f chosen, the phase transition of the system from high to low temperatures can be the coil-to-fibrillar transition, a coil-to-fibril-to-helix transition, or a coil-to-helix transition. Note that there is no transition from coil to fibril to helix, because the slope of the phase boundary between helix and fibril is 13/16. More specifically, a parameter configuration of the system described by a line in the phase diagram, with a slope greater than 13/16 and passing through origin, does not have an intersection with the phase boundary between the fibrillar and helical phases. The number 13/16 is associated with the ratio between the maximum bond numbers of helix and fibril. For the system with $L = 16$, the maximum number of helical bond is 13, while it is 16 for the fibrillar bond.

One of the interesting features observed in this case is successive two-phase transitions from coil to fibril to helix with respect to the increase of $\varepsilon_h/\varepsilon_m$. This is shown in the left upper inset of Fig. 10(a), which is taken along the purple dashed line corresponding to $\varepsilon_f/\varepsilon_m = 0.42\varepsilon_h/\varepsilon_m + 0.6$. By controlling ε_h while keeping the relationship between ε_f and ε_h, the transition from the helical phase ($P_h \geqslant 50\%$) to the fibrillar phase ($P_f \geqslant 50\%$) is forbidden. More specifically, the transition sequence is determined by the slope (i.e., the ratio $\varepsilon_f/\varepsilon_h$) of the line passing the origin through the point of interest. The appearance of the fibrillar phase in this case also depends on the number density ρ. Figure 10(b) shows the phase diagram as function of ρ and $\varepsilon_h/\varepsilon_m$, taken along the purple dashed line in Fig. 10(a). As is expected, the fibrillar phase shrinks as ρ decreases and, finally, vanishes when ρ is below a threshold where there exist insufficient polypeptides for fibrillation. Thus, for a fixed $\varepsilon_f/\varepsilon_h = 0.42$, fibril is more abundant for larger ρ. Below the threshold of ρ, the system will undergo the transition from coil to helix directly, without the fibrillar phase [see Fig. 10(b)]. Note in the figure that, surrounded by coil, fibrillar, and helical phases, there is a small region with $P_x < 50\%$ for all structures.

Another successive two-phase transition can be seen by considering the magenta dotted line in Fig. 10(a), which corresponds to $\varepsilon_f/\varepsilon_h = 0.94$. The right bottom inset shows the average fractional content P_x as a function of $\varepsilon_h/\varepsilon_m$. The abrupt changes of P_x for helix and fibril can be found at around $\varepsilon_h/\varepsilon_m = 1.2$. Remarkably, the two insets show there is an interconversion between native intrachain and fibril interchain sheet content, which likely implies that sheet structure is a transitional state with respect to the variation of $\varepsilon_h/\varepsilon_m$. Furthermore, when the ratio $\varepsilon_f/\varepsilon_h = 0.94$ is kept fixed, fibrillation takes place even for small ρ, as seen in Fig. 10(c). Only when ρ is extremely low does the system become too "dilute" to form fibrillar bonds. In this condition, helix is the last structure before structure corruption. More interestingly, Fig. 10(c) shows that for a fixed parameter configuration except ρ, a system primarily in the helical phase tends to form fibrillar structure when the concentration of the system is increased.

The specific heat c_V/k_B and the average fractional content P_x as a function of temperature T, with parameters $\varepsilon_h/\varepsilon_m = 1.00$, $\varepsilon_f/\varepsilon_m = 1.02$, corresponding to the purple triangle in Fig. 10(a), is shown in Fig. 11(a). From high to low temperatures, the system undergoes a transition mostly from the coil phase to the fibrillar phase at $T = 307$ K, and some helical and sheet content to the fibrillar phase during fibrillation [see Fig. 11(a)], but such a small amount of residues are involved in them that these transitions are unobservable in the specific-heat curves. The transition from coil phase to fibrillar phase is rather sharp, indicating that a significant amount of polypeptides collectively participate in this energy consumptive process. This is due to the fact that the fibrillation from coils involves formation of both sheet and fibrillar bonds, and more than one polypeptides must present in the formation of fibrillar bonds. The scenario has been found in all typical realizations with fibrillation and this is considered as a feature of NG process observed in some studies of aggregation [13,14].

As an example for comparison, Fig. 11(b) shows the specific heat c_V/k_B and the average fractional content P_x of the system as functions of temperature T, with parameters $\varepsilon_h/\varepsilon_m = 1.090$, $\varepsilon_f/\varepsilon_m = 1.0246$, corresponding to the magenta star in Fig. 10(a). From high to low temperatures, the system undergos

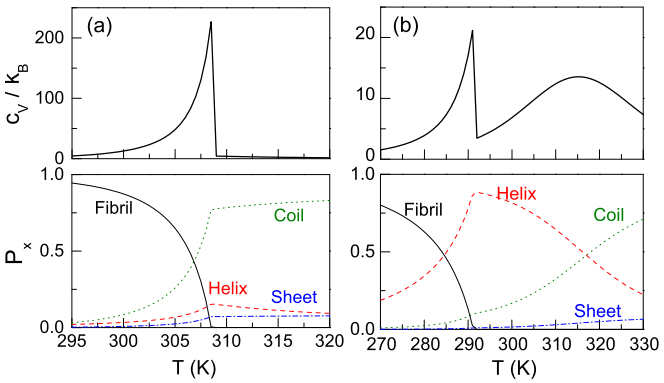

FIG. 11. (Color online) The specific heat c_V/k_B and the average fractional contents P_x of the system as functions of temperature T for sheet width $R = 3$ and polypeptide length $L = 16$. Here, structural entropy $s_h = -2.7530, s_s = -2.3826, s_f = -0.75$, bond coupling energy $\varepsilon_h = 1.995$ kcal/mol, $\varepsilon_s = 0$, $\varepsilon_f = 2.035$ kcal/mol, and number density $\rho = 0.6$. The plots of (a) and (b) correspond to parameters of the purple triangle and the magenta star in Fig. 10(a), respectively.

a transition from the coil phase to the helical phase at around 315 K and then another transition from the helical phase to the fibrillar phase at around 290 K. The former transition is quite smooth at this system size, while the later is sharp, resembling the case in Fig. 11(a). Nevertheless, in contrast to the case shown in Fig. 11(a), while the transition proceeds from the coil phase to the helical phase, the sheet structure is corrupted during the formation of the helical structure before fibrillation. Thus, in this condition the structures transiting to the fibrillar structure are mostly helices.

C. The case with fibrillation and $s_s \neq 0, \varepsilon_s \neq 0$

Now we consider a more general case, in which $s_s \neq 0$ and $\varepsilon_s \neq 0$. This corresponds to a system of the coexistence of sheet and fibrillar structures, which is the main design of our model. In contrast to the previous cases, here the sheet and fibrillar states are distinct. This explicitly indicates that not all sheet structures are involved in fibrillation.

Figure 12(a) shows the phase diagram of the system as a function of normalized helical and fibrillar bond coupling energy $\varepsilon_h/\varepsilon_m$ and $\varepsilon_s/\varepsilon_m$ at absolute temperature $T = 290$ K. The value of $\varepsilon_m = 1.995$ kcal/mol is determined through rescaling the phase boundary of the coil phase and the helical phase at $\varepsilon_s = 0$ to 1. Here, for convenience, but without losing generality, the fibrillar bond coupling energy ε_f is assumed to have a fixed relationship with the sheet bond coupling energy ε_s and is determined through $\varepsilon_f = 1.06\varepsilon_s$. The other parameters are structural entropy $s_h = -2.7530$, $s_s = -2.3826$, sheet width $R = 3$, number density $\rho = 0.6$, polypeptide length $L = 16$, and sheet width $R = 3$. The fibrillar structural entropy $s_f = -0.75$ is used to plot the phase boundaries in a bold solid line, while other choices of s_f have been shown for comparison. The phase boundaries corresponding to $s_f = 0$, $s_f = -0.75$, $s_f = -1.00$, and $s_f = -1.50$ are parallel to one another in their common linear regimes according to the linear relationship between ε_f and s_f in Eq. (3). The slope of the phase boundary between the fibrillar and helical phases at the

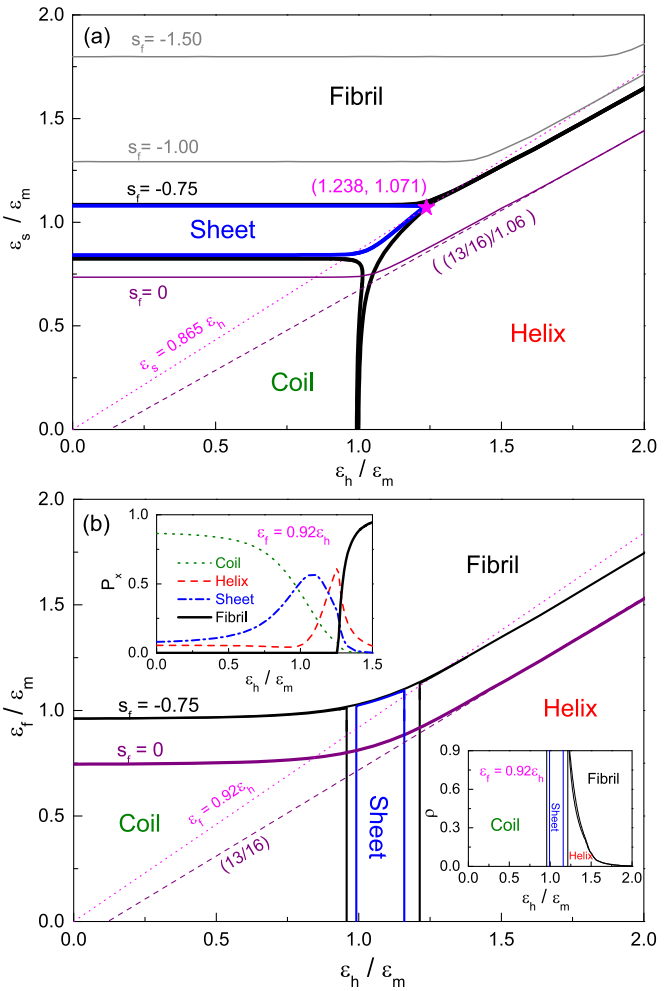

FIG. 12. (Color online) The phase diagram of the system (a) as a function of normalized helical and sheet bond coupling energy $\varepsilon_h/\varepsilon_m, \varepsilon_s/\varepsilon_m$, and (b) as a function of normalized helical and fibrillar bond coupling energy $\varepsilon_h/\varepsilon_m, \varepsilon_f/\varepsilon_m$, at absolute temperature $T = 290$ K. $\varepsilon_m = 1.995$ kcal/mol. Here, structural entropy $s_h = -2.7530, s_s = -2.3826$, sheet width $R = 3$, and number density $\rho = 0.6$, polypeptide length $L = 16$, sheet width $R = 3$. In (a), ε_f is determined through $\varepsilon_f = 1.06\varepsilon_s$. The purple dashed line represents the slope ((13/16)/1.06) of the straight line fitting the boundary curve of $s_f = 0$ at $\varepsilon_h/\varepsilon_m > 1.5$ regime. In (b), $\varepsilon_s = 0.857\varepsilon_h$ is assumed. The left upper inset and right bottom inset in (b) show respectively the average fractional content P_x and phase diagram as a function of ρ and $\varepsilon_h/\varepsilon_m$, taken along the magenta dotted line. The slope of the straight line fitting the boundary curve of zero fibrillar structural entropy $s_f = 0$ is 13/16.

large-ε_h regime is (13/16)/1.06. This number is associated with the ratio between the maximum bond numbers of fibril and helix for $L = 16$, as discussed in the previous subsection. Remarkably, for the case with $s_f = 0$, the region of fibrillar phase expands to cover the whole region primarily occupied by sheet phase for the case of $s_f = -0.75$. This shows that the system tends to form fibril instead of sheet structure if there is no entropic difference between fibril and sheet. This is similar to the case of nonexistent sheet structures in the previous subsection.

For comparison, in Fig. 12(b), the sheet bond coupling energy ε_s is assumed to have a fixed relationship with helical bond coupling energy ε_h, in which ε_s is determined through $\varepsilon_s = (1.566/1.827)\times\varepsilon_h = 0.857\varepsilon_h$. Again, the two phase boundaries corresponding to $s_f = 0$ and $s_f = -0.75$ are parallel to each other. Thus, a lower fibrillar structural entropy leads to a larger region for sheet phase. The slope of the boundary separating fibrillar and helical phases is 13/16. For a realization with the parameter configuration along the magenta dotted line of $\varepsilon_f = 0.92\varepsilon_h$ in Fig. 12(b), the average fractional content P_x as a function of $\varepsilon_h/\varepsilon_m$ is shown in the left upper inset of Fig. 12(b). Varying ε_h from small to large, there are successive three-phase transitions from the coil phase to the sheet phase, then from the sheet phase to the helical phase, and, finally, from the helical phase to the fibrillar phase. The slope of the line passing the origin through the point of interest determines the phase transitions. Thus, in this case, it is impossible to have a transition from coil to helix without visiting the sheet phase. This is shown more clearly in the right lower inset in Fig. 12(b), in which the phase diagram as a function of ρ and $\varepsilon_h/\varepsilon_m$ is shown. Comparing with Fig. 10(c), the sheet phase in this case partially occupies the region between coil and helical phases, defining the phase transition sequence with respect to $\varepsilon_h/\varepsilon_m$. Similar to the scenario discussed regarding Fig. 10(c), a system primarily in helical phase tends to form fibrillar structure as the concentration of polypeptide is increased, while the system remains unchanged if it is in the coil or sheet phase.

To explore the effects of the fibrillar structural entropy s_f, the fibrillar coupling energy ε_f, and the number density ρ on fibrillation, we chose the parameter settings of $s_f = -0.75$, $\varepsilon_h = 2.47$ cal/mol, $\varepsilon_s = 2.1366$ cal/mol, $\varepsilon_f = 2.2647$ cal/mol, $\rho = 0.6$ corresponding to configuration represented by the magenta star on the magenta dotted line of $\varepsilon_s = 0.865\varepsilon_h$ in Fig. 12(a), as a reference, and varied s_f, ε_f, and ρ one time per parameter to see changes in the specific-heat curve c_V/k_B and the average fractional contents P_x. Figure 13 shows c_V/k_B and P_x of the system as functions of temperature T for s_f ranging from -0.75 to -0.6 [Fig. 13(a)], ε_f ranging from 2.26 to 2.4 [Fig. 13(b)], and ρ ranging from 0.3 to 0.9 [Figs. 13(c) and 13(d)]. It turns out that the increases of the fibrillar structural entropy s_f, the fibrillar coupling energy ε_f, and the number density ρ all lead to increases of the transition temperature T_L and the height of the peak of the specific-heat curve $c_V^{\max}/k_B$. Typically, changing one of the parameters, for example, s_f, can lead to a change from successive two-phase transitions to a single-phase transition, as a result of shifting one transition from the fibrillar phase to the helical phase toward the other transition from the helical phase to the sheet phase and becoming a single transition from the fibrillar phase to the sheet phase. Remarkably, according to the relative variations of P_x for coil, helix, sheet, and fibril shown in Fig. 13, fibrillation is similar to a rolling snowball that destroys all the other phases to grow itself. The parameters control rolling of the snowball, resulting in corruptions of coil, helix, and sheet and formation of fibrillar structures.

We further analyzed the average length of fibrils and the average fraction of residues in the fibrillar state. The average length of fibrils is temperature dependent. For a given temperature, it increases abruptly from 2 to 10^{11} during

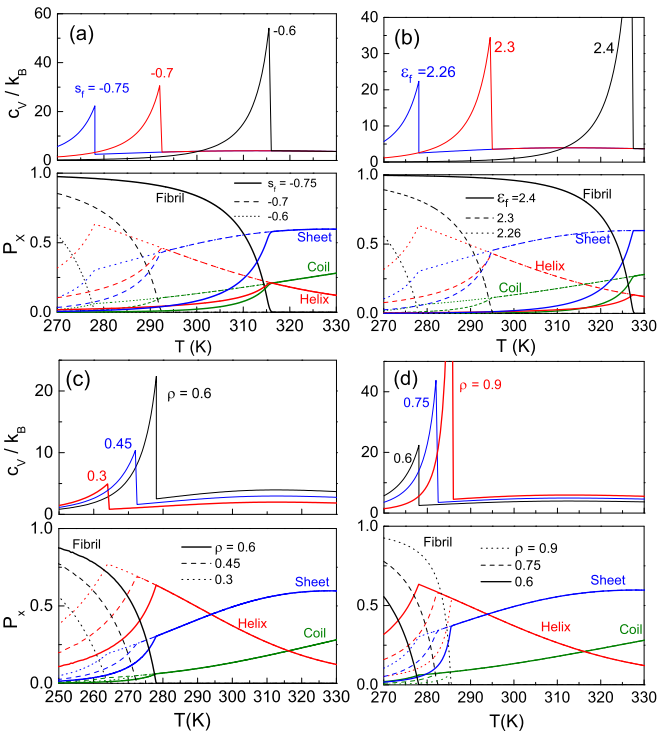

FIG. 13. (Color online) Specific heat c_V/k_B and the average fractional contents P_x of the system with as a function of temperature T for (a) structural entropy $s_f = -0.75, -0.7, -0.6$, (b) for fibrillar bond coupling energy $\varepsilon_f = 2.26, 2.3, 2.4$, (c) for number density $\rho = 0.3, 0.45, 0.6$, and (d) for $\rho = 0.6, 0.75, 0.9$. Here, structural entropy $s_h = -2.7530, s_s = -2.3826, s_f = -0.75$, bond coupling energy $\varepsilon_h = 2.47$ kcal/mol, $\varepsilon_s = 2.1366$ kcal/mol, $\varepsilon_f = 2.2647$ kcal/mol, sheet width $R = 3$ and polypeptide length $L = 16$, taken along the magenta dotted line in Fig. 12(a).

fibrillation (for infinite N) and then grows exponentially. The average fraction of residues in fibrillar state n_f during fibrillation is analyzed by calculating the average number of residues in fibrillar state in a polypeptide for different polypeptide length L and normalizing the outcome by L, and the result is shown in Fig. 14. It is interesting that at a transition temperature T_L, a longer polypeptide has a larger

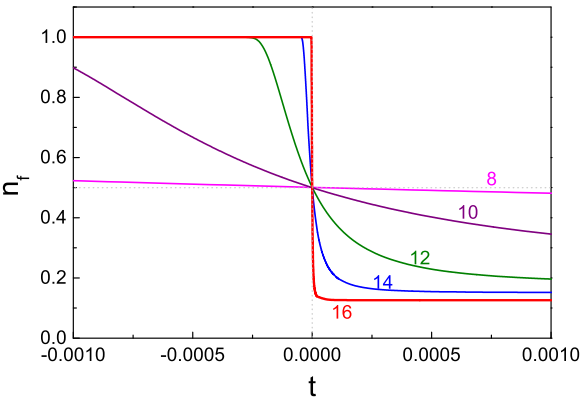

FIG. 14. (Color online) Average fraction of fibrillar residues in a polypeptide n_f as a function of the reduced temperature $t = (T - T_L)/T_L$.

fraction of residues transiting from or to the fibrillar state with respect to temperature change, which in turn results in a more abrupt (sharper) transition. This is a consequence of a longer polypeptide has a fewer residues at boundary that cannot be arranged in fibrillar state. Thus, under the framework of a hairpinlike fibrillar structure, it is easier to have a system with longer polypeptides transition to the fibrillar phase.

VI. CONCLUSIONS

We have proposed a polypeptide fibrillar transition model based on the three-state Potts model on cubic lattice and analyzed the model using the transfer matrix method and numerical calculations. Through parameterizing the competing bond coupling energy and structural entropies of helical, sheet, and fibrillar states, and concentration of the system, the model manifests phase transitions among coil, secondary structures, and fibril. We systematically analyzed the average fractional content of structures at phase transitions and constructed the phase diagrams to explore the transitions. The phase diagrams show the transition sequences are governed by the bond coupling energy, while the transition temperature is determined by the competition among the bond coupling energy of helix, sheet, and fibril, as well as the structural entropies, and concentration of the system. Depending on the parameters, there can be a single-phase transition, successive two-phase transitions, and successive three-phase transitions. No matter what the transition sequence is, the fibrillation is accompanied with an abrupt transition from the coil or secondary structure phase to the fibrillar phase, even for short polypeptide length. For the case without fibrillation, numerical data suggest that the finite-size effects in specific heat can be roughly described by the scaling form of the lattice Ising model.

Specifically, our model is able to mimic the following features: (i) Nucleation-growth-like behavior, which is revealed from the abrupt transition in fibrillation; (ii) relatively easier fibrillation for sheet-rich polypeptides (Fig. 12); (iii) successive two- and three-phase transitions, governed by parameters adjustable in experiments; (iv) forbidden transitions from high to low temperatures, such as the fibrillar-phase-to-sheet-phase transition in the phase diagram of Fig. 12. In addition, our model shows the possibility of the transition from the fibrillar to the helical phase for particular bond coupling energy (Fig. 12), suggesting a route of fibrillation control. Meanwhile, our model is limited by the simplification of homopolypeptides and assumption of hairpinlike fibrillar structure. A more realistic model without these limits will be a topic of our further study.

In conclusion, our studies on polypeptide phase transitions using a simplified lattice model based on a three-state Potts model provide useful information to understand fibrillation. Here we remark that our model reduces to the model proposed by Schreck and Yuan [23] at the limit of zero fibrillar bond coupling energy and zero fibrillar structural entropy and can reproduce the features observed in Ref. [23]. On the other hand, our model can also simulate the features of polypeptide fibrillation observed by Zamparo *et al.* in Ref. [27] when the sheet bond coupling energy is set to zero. The features revealed in the generalized version between the two limit cases show that our model offers richer phase-transition and physical properties of polypeptide fibrillation. Our model may be useful to raise the possibility of related experiments in protein aggregation, providing a way to design chemical conditions to make fibrillation control become possible.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Science and Technology of the Republic of China (Taiwan), under Grants No. NSC 100-2112-M-008-003-MY3 and No. MOST 103-2112-M-008-008-MY3, and the NCTS of Taiwan.

[1] F. Chiti and C. M. Dobson, Annu. Rev. Biochem. **75**, 333 (2006).
[2] C. Soto, Nat. Rev. Neurosci. **4**, 49 (2003).
[3] L. A. Morozova-Roche, J. Zurdo, A. Spencer, W. Noppe, V. Receveur, D. B. Archer, M. Joniau, and C. M. Dobson, J. Struct. Biol. **130**, 339 (2000).
[4] M. Fändrich, Cell. Mol. Life Sci. **64**, 2066 (2007).
[5] M. R. Nilsson, Methods **34**, 151 (2004).
[6] A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, and R. Tycko, Proc. Natl. Acad. Sci. USA **99**, 16742 (2002).
[7] N. Ferguson, J. Becker, H. Tidow, S. Tremmel, T. D. Sharpe, G. Krause, J. Flinders, M. Petrovich, J. Berriman, H. Oschkinat, and A. R. Fersht, Proc. Natl. Acad. Sci. USA **103**, 16248 (2006).
[8] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. Ø. Madsen, C. Riekel, and D. Eisenberg, Nature (London) **447**, 453 (2007).
[9] A. Trovato, F. Chiti, A. Maritan, and F. Seno, PLoS Comput. Biol. **2**, e170 (2006).
[10] T. Lürs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Döbeli, D. Schubert, and R. Riek, Proc. Natl. Acad. Sci. USA **102**, 17342 (2005).
[11] Y. Kallberg, M. Gustafsson, B. Persson, J. Thyberg, and J. Johansson, J. Biol. Chem. **276**, 12945 (2001).
[12] C. Wu and J.-E. Shea, Curr. Opin. Struct. Biol. **21**, 209 (2011).
[13] D. B. Teplow, N. D. Lazo, G. Bitan, S. Bernstein, T. Wyttenbach, M. T. Bowers, A. Baumketner, J.-E. Shea, B. Urbanc, L. Cruz, J. Borreguero, and H. E. Stanley, Acc. Chem. Res. **39**, 635 (2006).
[14] J. T. Jarrett and P. T. Lansbury, Cell **73**, 1055 (1993).
[15] W. P. Esler, E. R. Stimson, J. M. Jennings, H. V. Vinters, J. R. Ghilardi, J. P. Lee, P. W. Mantyh, and J. E. Maggio, Biochem. **39**, 6288 (2000).
[16] M. J. Cannon, A. D. Williams, R. Wetzel, and D. G. Myszka, Anal. Biochem. **328**, 67 (2004).
[17] P. Nguyen, M. S. Li, J. E. Staub, and D. Thirumalai, Proc. Natl. Acad. Sci. USA **104**, 111 (2007).
[18] H. Wako and N. Saitô, J. Phys. Soc. Jpn. **44**, 1931 (1978).
[19] H. Wako and N. Saitô, J. Phys. Soc. Jpn. **44**, 1939 (1978).
[20] V. Muñoz, P. A. Thompson, J. Hofrichter, and W. A. Eaton, Nature (London) **390**, 196 (1997).

[21] V. Muñoz, E. R. Henry, J. Hofrichter, and W. A. Eaton, Proc. Natl. Acad. Sci. USA **95**, 5872 (1998).
[22] V. Muñoz and W. A. Eaton, Proc. Natl. Acad. Sci. USA **96**, 11311 (1999).
[23] J. S. Schreck and J.-M. Yuan, Phys. Rev. E **81**, 061919 (2010).
[24] R. B. Potts, Ph. D. thesis, University of Oxford, 1951.
[25] R. B. Potts, Proc. Camb. Phil. Soc. **48**, 106 (1952).
[26] F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).
[27] M. Zamparo, A. Trovato, and A. Maritan, Phys. Rev. Lett. **105**, 108102 (2010).
[28] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, J. Mol. Biol. **7**, 95 (1963).
[29] A. L. Morris, M. W. Macarthur, E. G. Hutchinson, and J. M. Thorntonm, Proteins **12**, 345 (1992).
[30] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucl. Acids Res. **28**, 235 (2000).
[31] M.-C. Wu and C.-K. Hu, J. Phys. A **35**, 5189 (2002).
[32] M.-C. Wu, C.-K. Hu, and N. Sh. Izmailian, Phys. Rev. E **67**, 065103(R) (2003).
[33] M.-C. Wu, Phys. Rev. E **73**, 046135 (2006).
[34] M.-C. Wu, C.-L. Li, C.-K. Hu, Y.-C. Chang, Y.-H. Liaw, L.-W. Huang, C.-S. Chang, T.-T. Tsong, and T. Hsu, Phys. Rev. B **74**, 125424 (2006).