Folding transition of model protein chains characterized by partition function zeros

Jun Wang and Wei Wang[a)]
National Laboratory of Solid Microstructure, Department of Physics, Nanjing University, Nanjing 210093, China

(Received 9 September 2002; accepted 14 November 2002)

The folding transition of model protein chains with various kinds of Go̅-type interactions are investigated by partition function zeros on complex temperature plane. Using multicanonical sampling procedures, the density of states and thus the partition function zeros are precisely obtained. Several factors related to the local distribution of the partition function zeros near the real axis are extracted and used to characterize the features of folding transition. The results show that the folding transition is of first-order-like, and is weakly dependent on the native structures. The efficiency of the method of partition function zero is also illustrated. The correlation between some conventional thermodynamic factors for characterizing the stability and foldability and those obtained from the zeros are also studied. Finally, a mapping between the various models and the Go̅-type models is proposed based on the factors related to zeros, which suggests the wide applicability of the method based on the partition function zeros. © *2003 American Institute of Physics.* [DOI: 10.1063/1.1536162]

I. INTRODUCTION

Protein folding, as a challenging problem in molecular biology, has attracted great attention during the last decades.[1–8] The folding of protein molecules, namely, the process from less-structured denatured states to the well-shaped native conformation, is depicted as a directed diffusion on a rugged funnel-like landscape. Nevertheless, proteins can also be unfolded under certain external conditions.[9] Following the change of such environmental factors, the protein systems undergo a transition between two macrostates, namely, the denatured and the native one.[10] This kind of transition is a fundamental phenomenon in protein systems, which shows the competition of various ingredients of protein systems.[11,12] To understand the characteristics of the transition and its physical origin is an essential point for the research of protein folding[13–21] and modeling.[22–24]

Different from the second-order-like transition of the flexible polymers,[25] the folding generally shows a highly first-order cooperativity during the transition,[13–15] and the kinetics is similar to a Kramer process between two free energy minima.[26] The energy distinctiveness of the native structure[27] ensures the existence of two phases.[13] The large change in symmetry of structures during the folding also implies that such a kind of transition is generally of first-order type, such as ice melting.[28] Yet, the characterization on the phase transition in protein systems still has some difficulties due to the choice of the order parameter. To find a suitable description of folding transition not only provides some insights on the mechanism of folding processes but also extends our knowledge on the description of complex macromolecular systems.

A procedure based on the Fisher zeros or Lee-Yang zeros of the partition function[29,30] seems to have become an important scheme for investigating the thermodynamic transition in a finite-size system.[17–21,31–34] It has been successfully used to study the cluster melting problem,[31] the helix-coil transition of the polypeptide system,[17] Bose–Einstein condensation,[32] and the lattice Gross–Neveu model with Wilson fermions.[34] Especially, using the partition function zero for the analysis on the phase transitions in biomolecules, such as the helix-coil transition[17,18] and the folding transition[19–21] of protein chains, has been successfully implemented by Hansmann and co-workers. This kind of method shows its consistence with the Ehrenfest criterion at the thermodynamics limit, and provides more flexibility for finite-size systems. For protein-like model systems, it is a powerful way to extract information of the transition via the study on the partition function zeros (PFZs).

In this work, we report a study on the features of folding transitions of lattice model proteins based on the partition function zeros. Comparing with the studies by Hansmann and co-workers in Refs. 20 and 21, we use different models and perform a more detailed investigation on the characterization of the folding properties of model protein systems. These give us a more comprehensive understanding on the folding transition. The transition of the model protein system is characterized with multiple parameters extracted from the distribution of the partition function zeros. The order of the transition is discussed on the basis of the partition function zeros. The folding transition is identified as a first-order-like transition. The transition temperature and the strength of the transition directly relate to the intercept and the slope of the local distribution of the zeros near the real axis. The conventional characterization of proteins is assessed from their correlations with the related parameters to the zeros. With these

a)Author to whom correspondence should be addressed; Electronic address: wangwei@nju.edu.cn

understandings, a mapping of various protein models to the Gō-type models[35,36] provides a possible comparison between different models. The whole scheme with partition function zeros shows a thorough analysis for protein systems. The main contents are arranged as follows: The basic scheme of partition function zero analysis and the related implementations are presented in Sec. II. The phase transition of Gō-model protein chains and its characteristics are discussed in Sec. III. Different native structures and the chain size effect are also studied. In Sec. IV, the evolution of the transitions for different non-native strengths is discussed. Various factors for the foldability are also discussed by comparing with the characteristics from the partition function zeros. These investigations enable us have a more comprehensive understanding on the protein folding problem. As an extrapolation for our conclusions, a mapping of different types of model proteins to the Gō-type models is made, which shows a semiquantitative thermodynamic assessment of different models. Finally, we give a summary in the last section.

II. MODELS AND METHODS

A. Details of modeling

In the present work, we investigate several three-dimensional lattice models with Gō-type interactions. The model proteins are modeled with self-avoided random walks on a three-dimensional cubic lattice with fixed length. One walk corresponds to a conformation of the model protein. The spatially neighboring sites interact with each other and form a contact when they are not sequential neighbors. The Gō-type interaction assumes that the contacts in the preselected native conformation are most energetically favored, and their strengths $\epsilon_N<0$ are all equal. For the sake of simplicity, the quantity ϵ_N is chosen as the unit of the energy and the temperature. Here the Boltzmann constant k_B is set as 1. To include some effects of energetic frustration, the nonnative contacts are considered to have a nonzero attractive strength as $\epsilon_{NN}=\Delta\epsilon_N$ for comparison. Therefore, the energy for any conformation C of the model chain is

$$E=\sum_{i<j}\left[\epsilon_N\Lambda_{ij}^{(C)}\Lambda_{ij}^{(N)}+\epsilon_{NN}\Lambda_{ij}^{(C)}(1-\Lambda_{ij}^{(N)})\right]. \tag{1}$$

Here, $\Lambda_{ij}^{(C)}$ and $\Lambda_{ij}^{(N)}$ indicate the contact maps for conformation C and the native conformation N, respectively. The nonnative strength adjusts the foldability of proteins. The original Gō model corresponds to the case $\Delta=0$.[35,36] Since it is believed that there is some simplicity embedded in protein systems, as suggested from the connection between the native topology and folding kinetics[37] and from the simplification of the amino acid alphabets,[38–41] simplified modeling for proteins is generally feasible. The Gō-type models maximally emphasize the topological ingredients of protein systems and simplify the sequence dependence, and are believed to contain the main feature of the protein systems.[42] For these models, simulations are carried out with the regular Monte Carlo (MC) methods.[43] The density of states is extracted with multicanonical sampling method (see the following).

TABLE I. The simulation details for various models. The maximal Monte Carlo (MC) steps in one sweep, the number of sweeps of iteration, and the number of independent runs for error evaluations are listed.

	MC steps in one sweep (maximal)	Iteration sweeps	Independent runs
$\Delta=0$	4×10^9	20	40
$\Delta\leqslant0.9$	5×10^{10}	50	16

B. Sampling schemes

The most important point for obtaining an accurate partition function, thus partition function zeros, is to have a precise estimation on the density of states. Thus, some sophisticated skills, such as multicanonical sampling, are necessary to be used.[17,44–47] A well-designed sampling over different energy regions can ensure the similar errors for the density of states for the whole energy spectrum. Therefore, it is possible to give the correct distribution of the PFZs, which enables us to have a clear understanding on the phase structure of the proteinlike model system. In our work, multicanonical sampling is used to obtain a suitable estimation on the whole landscape along the energy axis.

In multicanonical sampling, the energy landscape are flattened by realizing an equal probability for each energy level as

$$P_{\rm ms}(E)\propto n(E)\rho_{\rm ms}(E)=\text{const}, \tag{2}$$

where $\rho_{\rm ms}(E)$ means the statistical weight of a microstate with energy E, $P_{\rm ms}(E)$ means the probability on an energy level E, and $n(E)$ represents the corresponding density of the states. Here the subscript is for the multicanonical sampling. Therefore, the statistical weight has the form

$$\rho_{\rm ms}(E)\propto\frac{1}{n(E)}=e^{-S(E)/k_B}, \tag{3}$$

where $S(E)=k_B\ln[n(E)]$ is the microcanonical entropy. The task to find the statistical weight is implemented by an iterative procedure with a prior guess for the weight $\rho_{\rm ms}(E)$. The detailed implementation can see the references, e.g., Ref. 46.

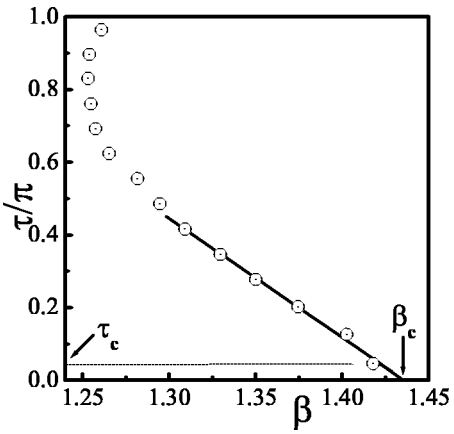

FIG. 1. The PFZ distribution in primary region $[0,\pi]$. The zeros near the real axis lie on a straight line. The solid line shows the fitting, and it cuts the real axis at β_c. The minimal value of image of zeros is τ_c as shown in the figure.

In this work, we study the cases with the non-native strength ϵ_{nn} up to $0.9\epsilon_n$, i.e., $0 \leqslant \epsilon_{nn} \leqslant 0.9\epsilon_n$. The detailed parameters for the simulations of various models are listed in Table I. It is noted that, for systems with more frustrations, to obtain the convergence a large number of iterations is needed due to the deep and distant traps on the free-energy landscape, and the simulation time for each running of the iterations is also necessary to be increased to realize a sufficient statistical significance. The errors of results are estimated from some independent runs.

Generally, various native structures have different designability and different kinetic partition properties. Different native topologies may have various kinetics, but their thermodynamics is qualitatively the same. In our work, we use the 27-site structure (Fig. 1 in Ref. 48) as the native one for our Gō model if there is no further declaration. We will discuss the structural difference briefly in the following section.

C. Partition function zeros

The partition function zeros in the complex temperature plane can be determined by solving the equation $Z(B) = \Sigma_E n(E)\exp(-BE) = 0$ for their complex roots. To describe the distribution of these zeros, we adopt the parameters proposed and studied in Refs. 19 and 31. The detailed forms of the evaluation and characterization of PFZs are given in the Appendix.

III. Gō MODELS WITHOUT NON-NATIVE INTERACTION

A. An example of analysis of PFZs for the Gō model

With the methods described in preceding section, the Gō model without non-native interactions can be analyzed thoroughly. The PFZs of this system are extracted and shown in Fig. 1. The corresponding factors are $\beta_c = 1.4153(19)$, $\gamma = -0.0621(12)$, $\tau_c = 0.1430(5)$, and $\alpha = 0.0817(64)$, respectively. The result shows that there is a transition at the temperature $T_c = 1/\beta_c = 0.707(1)$. For the order of the transition, let us examine the density exponent α, which is slightly larger than zero. Strictly, the transition should be of second order, which is the same as that in Ref. 21. However, from a coarse-grained view, the value of the factor α is small and is approximately equal to zero, that is, the zeros almost have a uniform space between the neighbors. This resembles a first-order transition. Our small positive value of the factor α may be attributed to the contact assumption of the interaction in the model. In fact, we have seen a smaller value of the factor α for a model including cooperative many-body interactions[23] than that of the present work (results are not shown here). This deduction is consistent with some other theoretical and computational studies.[13,49] Besides, the finite population of the intermediate states makes the transition slightly second-order-like since for these states the gap between the native state and the denatured states is not large enough. Considering other factors, the system has a small γ, i.e., has an almost vertical line of zeros, which reflects a strong transition. These features provide a comprehensive description for the transition of model protein systems.

To obtain a more precise picture of the transition, let us view the complex temperature plane. This plane is divided into two parts by the line of the zeros. Each part corresponds to a thermodynamic phase related to the suggested transition by the zeros. The line of zeros outlines the boundary between different phases. We calculated the magnitude of the average energy E, namely $|\langle E\rangle|$, for the system in the complex temperature plane, and superpose it with the distributions of the zeros (see Fig. 2). It is clear that the zero line separates two areas with obviously different energies. Now at these zeros, the quantity $|\langle E\rangle|$ shows a discontinuity, which is related to the feature of the first-order transitions.

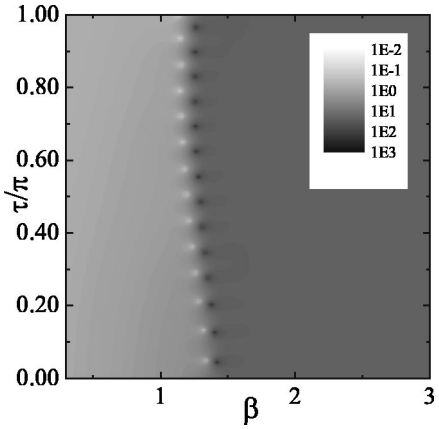

FIG. 2. The average energy $\langle E\rangle$ (in absolute value) on the complex-temperature plane. The line consisted of PFZs clearly shows an abrupt transition in energy, which suggests the first-order-transition feature.

Other physical quantities, such as the average of the total contacts N, $\langle N\rangle$, all follow the same behavior (see Fig. 3) since the distribution of the zeros is independent of the investigated quantities. There is only a single transition here for the Gō-model chains without non-native interactions. The collapse and the folding (namely the recognization of the specific structure) occur simultaneously for this small model protein with its size of 27.

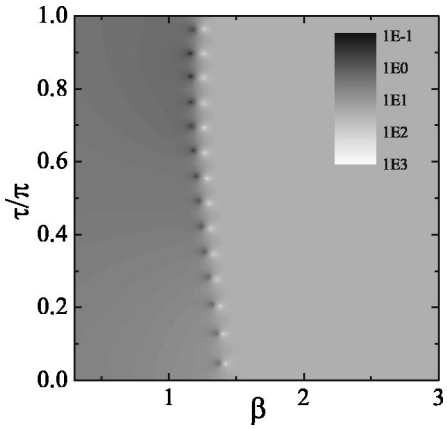

FIG. 3. The average of the total contacts N, $\langle N\rangle$ (in absolute value) on the complex-temperature plane. There is also an abrupt transition in number of bonds, and the transition occurs at the same temperature as Fig. 2.

TABLE II. The factors related to the transition of Gō chains with different native structures.

Structure index	β_c	γ	τ_c	α
1	1.4153(19)	−0.0621(12)	0.1430(5)	0.0816(64)
2	1.4037(23)	−0.0397(17)	0.1398(7)	0.1037(82)
3	1.4241(15)	−0.0475(12)	0.1501(4)	0.2101(51)
4	1.4804(12)	−0.0838(11)	0.1674(4)	0.0952(47)
5	1.4409(17)	−0.0757(14)	0.1563(5)	0.0969(60)
6	1.4492(17)	−0.0719(12)	0.1527(4)	0.1692(54)
7	1.4518(18)	−0.0442(15)	0.1546(6)	0.1405(55)
8	1.4375(14)	−0.0466(14)	0.1531(4)	0.1176(52)
9	1.4060(21)	−0.0479(17)	0.1397(6)	0.1123(74)

B. Structure dependence of PFZs for the Gō model

Do these features of the transition depend on the choice of the native structure? We check this for several other native structures. The distributions of their zeros are shown in Fig. 4. The corresponding factors are listed in Table II, and the correlations between these factors are calculated as listed in Table III. There are a strong correlation (with the correlation coefficient $R=0.951$) between the transition temperature (the factor β_c) and finite size effect (the factor τ_c). It seems that the Gō models with different native structures have different ''distances'' from the thermodynamic limit, and the variation of the transition temperatures comes from the size dependence of the energy gap, which results in the correlations with the finite-size effect. This phenomenon is attributed to the fact that the distinct topologies of native conformations have different effective free-energy barriers, which produces different degrees of resemblance to the barrier for a system with infinite size in the thermodynamic limit. This also provides some interpretations for the correlations between the factors γ and τ_c (or β_c). The other correlations are rather weak, showing the irrelevance between the factor α and the others. This suggests that the order of the transition seems to be decoupled with other features of the transition. This is qualitatively the same as the examples shown in Ref. 31. For different choices of the native structures, there are some variations in their transition temperatures. However, the properties of the transition are basically the same. Thus, the above properties of the phase transition are the basic ingredients of proteinlike systems, which does not depend on the native topology qualitatively. In other words, the native structure will not change the thermodynamics of the protein system seriously. This conclusion matches with the general observations that different proteins are similar in their first-order feature of the folding transitions.[13–15] On the other hand, the detailed transition properties depend on the features of the native structures. The detailed transition properties vary from protein to protein, which implies that the extrapolation to the thermodynamic limit for heterogeneous protein systems is correct in a qualitative sense, rather than quantitatively. This kind of qualitative consistence coincides with the self-average property of the protein systems,[50] which validates the regular mean-field analysis.

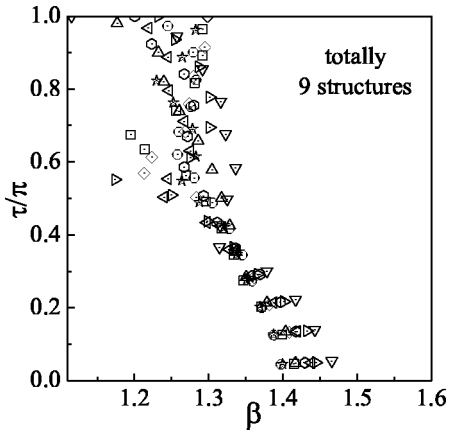

FIG. 4. The PFZs for nine Gō model chains with different native structures. They basically show the similar behaviors, and do have some differences in various factors from structure to structure.

C. Size dependence of PFZs for Gō model

As a finite ''cluster'' of residues, different sizes of proteins exist in nature. Do they experience different transitions? It is a question concerning both biology and physics. We study several Gō-type model proteins with different numbers of beads (marked as the length $L=18,36,48,64$). The corresponding factors are listed is Table IV. Their native structures are picked up randomly from the set of the most compact structures, generally in cuboid forms. As pointed out above, different choices of the native structures still keep the behavior of the folding transition qualitatively. We find that the various folding properties change almost monotonically as the lengths of model proteins increase. The folding temperature increases as the chain length increases due to the slower increase of the configurational entropy than that of the energetic factor (see Fig. 5). Especially, the factor α related to the local density of zeros decreases. This implies that with sufficient numbers of degrees of freedom, the folding transition in protein systems acts more similarly to a first-order phase transition. The apparent decrease of the value of the factor α also suggests that the size effect is significant for

TABLE III. The correlation coefficients between different factors related to the transition of the Gō chains with different native structures.

β_c	γ	τ_c	α
β_c	−0.717	0.951	−0.019
γ		−0.693	0.232
τ_c			0.0313

TABLE IV. The factors related to the transition of the Gō chains with different lengths.

Length	β_c	γ	τ_c	α
18	1.6255(21)	−0.1431(21)	0.7493(9)	0.2877(83)
27	1.4153(19)	−0.0621(12)	0.1430(5)	0.0816(64)
36	1.3556(16)	−0.0248(8)	0.1042(5)	0.1192(55)
48	1.2841(20)	−0.00677(5)	0.0644(4)	0.0427(48)
64	1.2337(23)	−0.00472(6)	0.0459(5)	0.0355(51)

the lattice protein models with a moderate length (such as $L=27$; see Fig. 6). Moreover, the slope of the line of zeros increases as chain length L increases. This implies that the first-order feature of the transition is consolidated for longer chains, which relates to the general statistical arguments. The variation of the factor τ_c also shows the weakening of the finite size effect for longer model proteins. These results suggest that the first-order character of the folding transition is an intrinsic feature from the proteinlike properties. The sizes of systems modulate the strength of the transitions, but will not change the basic trend of the transition. As a note, here we do not check the exact scaling for various properties, unlike the studies for the general phase transition in statistical physics where a finite-size scaling is often visited. In fact, for proteins, the thermodynamic limit is meaningless due to their moderate sizes. The generally interesting single-domain proteins have a size near several hundred residues, which obviously is far from the thermodynamic limit. The thermodynamic limit for proteins cannot be realized or approached in nature. It is noted that the precise scaling behavior for heterogeneous protein systems is not easy to be worked out.

IV. Gō MODELS WITH NON-NATIVE INTERACTION

A. The PFZs for various Gō models

For naturally occurring proteins, the various types of residues make the non-native contacts relevant. Some correlations or attractions exist between the residues, also for the pairs not presented in the native form. This phenomenon implies that there are some energetic frustrations in natural protein systems, which is absent in the Gō-type model without non-native interaction. The non-native ingredient changes the landscape and thus the population of states and the kinetics.[51] How does the energetic frustration change the picture and the properties of the folding transition? It is of importance for the understanding on the naturally occurring proteins. The PFZs do provide a power tool to study this problem.

To efficiently find the density of state with different non-zero non-native strengths, we sample the conformations in a N vs Q space with the entropic sampling algorithm. Here N and Q represent the total contact number and the native contact number for the sampled conformations, respectively. It is obvious that the conformations with same N and Q share a same energy $E=Q\epsilon_N+(N-Q)\epsilon_{NN}$. The energy unit ϵ_0 can be easily obtained from ϵ_N and ϵ_{NN}. The zeros can be obtained by solving the corresponding polynomial equations.

Studying the distributions of the PFZs, there are several kinds of patterns for different values of frustration Δ. The simplest one is the case with a tiny non-native interaction, namely $\Delta\simeq 0$. There is simply a single-phase boundary consisted of zeros, which separates two phases (the denatured states and the native state). When the energetic frustration increases, there are more states being stable in the model systems, besides the native state favored by the energetic factor and the coil states with their entropic preference. Therefore, more phases appear due to the tradeoff between different interactions. For example, when the frustration factor $\Delta=0.6$, there are three phases near the real axis as marked in Fig. 7(a), and a new phase with intermediate size

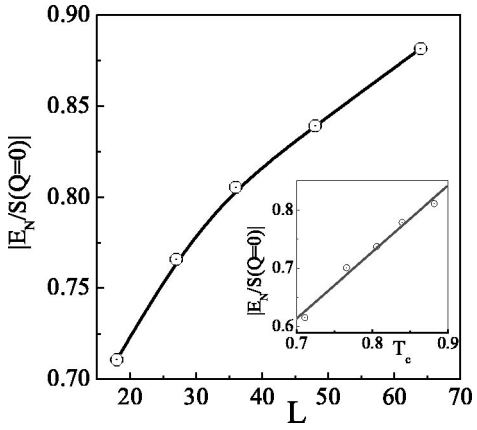

FIG. 5. The competition between the energy and the entropy for different sizes of chains. The ratio $E_N/S(Q=0)$ serves as a raw estimation on this competition. It shows a monotonical increase with the chain length. The inset shows the correlation between the ratio $E_N/S(Q=0)$ and the transition temperature $T_c=1/\beta_c$. The solid line is the fit.

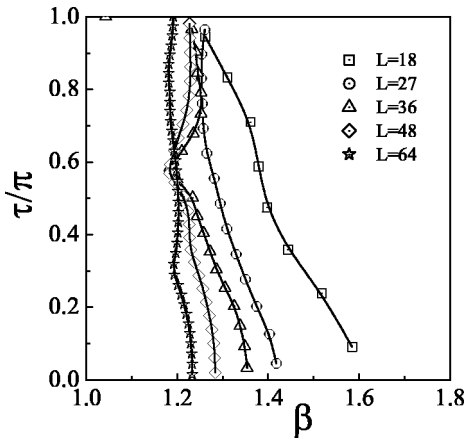

FIG. 6. The PFZs for the Gō model chains with different lengths of $L=18$, 27, 36, 48, and 64, respectively. The solid lines are the guide line for eyes. They basically have similar forms, and do show systematic change in their forms.

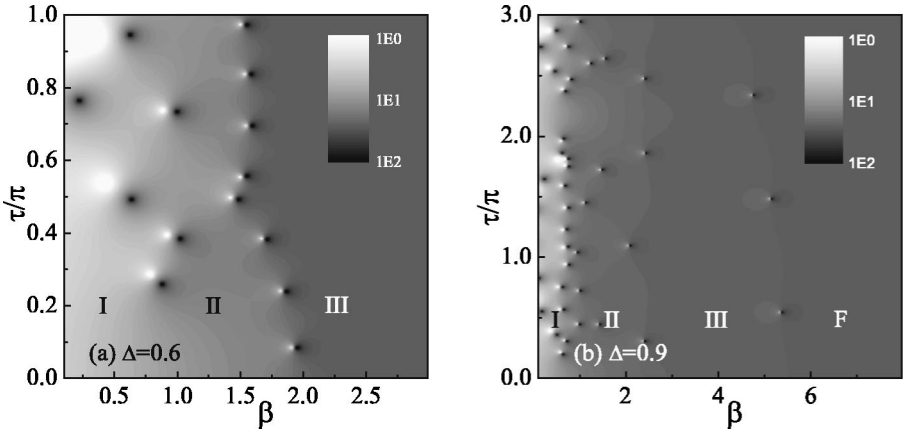

FIG. 7. The average energy $\langle E \rangle$ (in absolute value) on the complex-temperature plane for the cases (a) $\Delta = 0.6$ and (b) $\Delta = 0.9$. There are three phases with different energetics for case (a), and there are four phases in case (b), in which F characterizes the folded (native) state.

of energies appears. The zeros roughly make the boundary. This corresponds to the ''molten-globule'' state as argued in other works.[52] The proper strength of the non-native interactions (i.e., the value Δ) induces a separation in the energies of an ensemble of states from the coil states. In more detail, a new phase occurs for $\Delta = 0.2-0.3$. Noted that the ability to identify the phases is enhanced with the aid of the PFZ schemes in the complex-temperature plane. Nevertheless, the change between the two-phase region and three-phase region cannot be identified from the regular heat capacity for real temperature since there is only one peak in that case. The reason is because of the roundoff of the finite size and the screening or overlapping between the peaks. For larger frustrations, such as $\Delta = 0.9$, the phase space becomes more complex [see Fig. 7(b)]. The detailed differences between various ensembles of structures are magnified due to the non-native interaction, and these differences may produce some new phases at low temperatures. Due to the assumption on the energetics of the model proteins, the native state generally has the lowest energy. Therefore, the phase occurring in the low-temperature area corresponds to the folded state. As a result, the transition from unfolded states to the folded one still exists as the one with the lowest real temperature. This argument is consistent with the former understanding on various model protein systems.[53,54] In this work, we mainly concentrate in the folding transitions only.

The factors related to the folding transitions of the chains with different frustrations are listed is Table V. From Table V, we can see that when the frustration Δ increases, the folding transition shifts towards the region with larger β, namely at lower temperature T. The increase of the strength of the non-native contacts enhances the stability of the denatured states, which makes the energy gap between the native state and the denatured states decrease. The native state generally exhibits a larger preference at lower temperature. The slope of the line of the PFZs decreases as the frustration increases, which indicates that the transition becomes weaker and weaker. The finite-size effect changes slightly as indication by the parameter τ_0. It is worth noting that the properties of the folding transition with $\Delta = 0.9$ is a little different from the others. The transition shows better first-order features with respect to the cases with intermediate non-native strength. It is ascribed to the difference of the states relevant to these transitions. For the intermediate non-native strengths, the transition relates to a large ensemble of states, broad in energetics, while for the very large energetic frustration, $\Delta = 0.9$, the concerned states is rather restricted with a narrow energy distribution. This ''anomaly'' provides us some physical insights of the factors from the distribution of PFZs.

B. The comparison of various characteristic temperatures

From the systematic variance of the PFZ-related factors with the non-native strengths, it seems that these factors provide a description for the transitions of the chains with different foldability. Do these settlements coincide with the former investigations? First, we investigate the correlations of the transition temperatures from the PFZ scheme and from some other definitions of folding temperatures. Conventionally, the following characteristic temperatures are used for the location the folding transition: (1) the melting temperature, T_m, at which the native population equals to 1/2; (2) the temperature of the peak of heat capacity, T_{C_p}, at which the heat capacity of the system takes its maximum; (3) the temperature of the peak of native-similarity fluctuation $\langle Q^2 \rangle - \langle Q \rangle^2$, T_Q, which intuitively determines the structural change; and (4) the phenomenological transition temperature T_B based on the Q coordinate, at which there is a balance between native states and the unfolded free-energy minimum. With the entropies for different states, these tempera-

TABLE V. The factors related to the transitions of the Go-type chains with different energetic frustrations. The abnormality for the parameter α of the case $\Delta = 0.8$ is from the spatial diversity of the zeros, which makes the determination of the factor α imprecise.

Δ	β_c	γ	τ_c	α
0	1.4153(19)	−0.0621(12)	0.1430(5)	0.0816(64)
0.1	1.3995(16)	−0.0482(8)	0.1391(11)	0.0576(55)
0.2	1.487(22)	−0.1016(5)	0.159(2)	0.101(32)
0.3	1.564(24)	−0.1265(7)	0.172(3)	0.123(73)
0.4	1.675(17)	−0.145(1)	0.192(2)	0.084(31)
0.5	1.796(11)	−0.156(3)	0.221(2)	0.059(51)
0.6	1.975(35)	−0.176(2)	0.264(1)	0.142(75)
0.7	2.31(8)	−0.212(13)	0.338(7)	−0.017(71)
0.8	2.78(15)	−0.295(27)	0.589(14)	−0.43(12)
0.9	5.51(27)	−0.076(24)	1.712(16)	0.112(87)

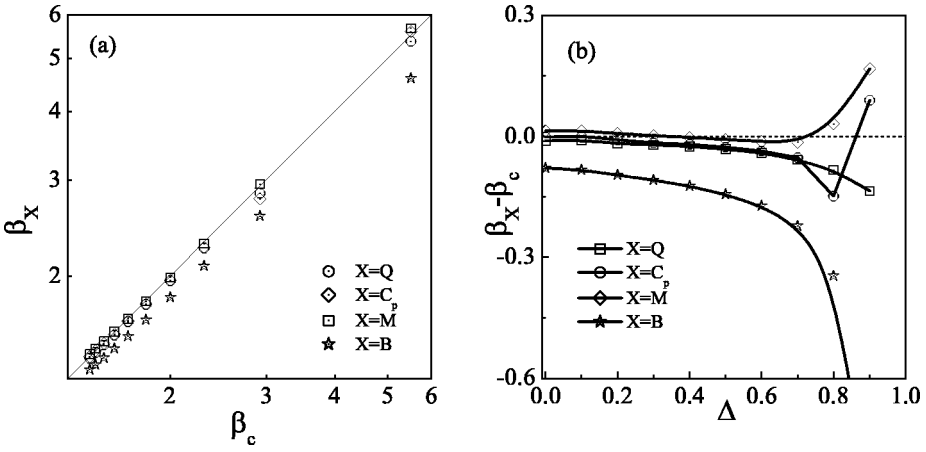

FIG. 8. (a) The relation between different β_c and β_X, with $X=Q$, C_p, M, and B, respectively. For the detailed meaning of X, see the text. The solid diagonal corresponds to $\beta=\beta_c$. The double logarithmic plot is used just for a better view. (b) The difference $\beta_X-\beta_c$ vs the frustration Δ. The dashed horizontal line represents the case with null differences.

tures can be deduced easily. Their correlations are shown in Fig. 8. It is clear that these temperatures all have good correlation with the transition temperature β_c, namely, they have good linear relations between each other. In this sense, these temperatures all can be used qualitatively to show the variation of the transition. However, the temperature obtained from the equality of the free energy between native states and unfolded state along the Q coordinate, β_B (or T_B), is generally lower than β_c (or higher than T_c). This results from the underestimation of the unfolded state ensemble. Especially, for large frustrations, the deviation becomes more significant so that the difference $\beta_B-\beta_c$ becomes large (see Fig. 8). It is interesting that β_Q and β_{C_p} vary similarly under a certain frustration, suggesting that the energetic and structural formation generally are concurrent, especially for the folders with small frustrations, namely with optimal foldability. This is consistent with the arguments related to broad transition states[55] and with some experimental results that found the concurrent folding and collapse for some small proteins.[56] From our investigation, it seems that the temperature β_M has good coincidence with the β_c in the widest region of the frustrations, which may result from the consideration of all the configurations. The temperatures β_{C_p} and β_Q have systematic deviations. It is ascribed to the overestimation on the entropy of the native phase. There are some configurations similar to the native state that contribute to the average at low temperatures. An enlarged native basin to include these configurations can provide a clear picture for the transition characterized by these temperatures.[57] Interestingly, the temperature T_m defining the native dominance shows a coincidence in a large range of frustrations. It comes not only from the proper estimations for different ensembles, but also from the dominance of the folding transition. For a very large frustration, such as $\Delta=0.9$, different temperature schemes all show a large deviation of β_c, which is attributed to the weak strength of the transition. Especially, for the cases with large frustrations, the rule of population variation changes. Consequently, the temperature T_m deviates from the folding transition considerably. Due to the large value of β_c in those cases, this deviation will not introduce large errors in extracting the transition temperature.

C. Why the PFZ scheme is better

It is worth noting that different physical quantities, such as the heat capacity C_p and the fluctuation of similarity σ_Q, show various thermodynamic behaviors with the same set of the PFZs, especially for the cases with large energetic frustrations. This results in the phenomenon that the regular analysis of the real temperature axis depends on the choice of the investigated physical quantity. This kind of difference does not exist in the analysis of PFZs, showing the power of the analysis based on the scheme of PFZs. On the other hand, what makes the difference for various quantities? As we discussed above, the PFZs correspond to the poles of the quantities on the complex temperature plane. How do these poles affect the thermodynamic behavior of the investigated quantities? At the poles, the physical quantities generally are divergent. Going away from these poles, these quantities decay with a form proportional to $1/|B-B_0|$, which can be seen from the definition of the PFZs. For the physically interested real temperature area, the values of the physical quantities are the summation of the contributions of these poles. Therefore, the rate of decaying away from the poles determines the importance of the corresponding zeros. To have an estimation of the decay rate of the PFZs, the zeros of the physical quantities provide us some information. If there are many zeros near the poles (PFZs), the decay generally is fast to fit the zeros. Meanwhile, a systematic large deviation of zeros from poles makes the poles contribute more to the physically meaningful area.

For a quantity A, these zeros satisfy an equation

$$\sum_{k,l} A_{kl} n(k,l) u^k = \sum_k \left[\sum_l A_{kl} n(k,l) \right] u^k = 0, \tag{4}$$

where k and l are integers. Here, the index k relates to the energy, while l is defined relating to other degrees of freedom. If the quantity A is irrelevant to energy E, namely $A_{kl}=A_l$, the above equation can be rearranged as

$$\left(\sum_l A_l \right) \sum_k n(k) u^k = 0, \tag{5}$$

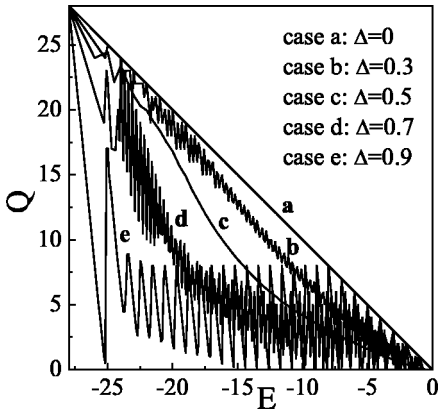

FIG. 9. The similarity Q vs the energy E for different cases of the frustrations. The zigzag of the curves result from the discreetness of the density of states.

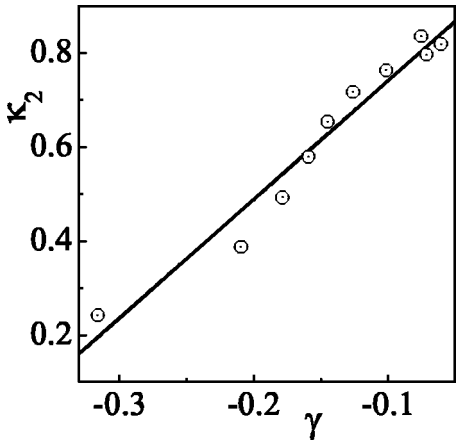

FIG. 10. The correlation between the cooperativity κ_2 and the transition strength factor γ for the systems with different frustrations Δ from $\Delta=0$ to $\Delta=0.9$.

with $n(k)=\Sigma_l n(k,l)$. This is equivalent to the equation for PFZs. That is to say, these poles have no contribution to the quantity A. On the other hand, if the quantity A tightly correlates with the energy E, such as $A_{kl}\propto k$, the zeros are systematically away from the poles. Then, all poles affect the behavior of the quantity A for different temperatures. As we check the correlation between A and E, we find qualitatively which parts of zeros take their effects on the real temperature axis. For the quantity energy E, only the latter case takes place. As a result, the zeros near the real axis generally all take their roles to the heat capacity, which results in a form of multipeaks. However, for the quantity Q, we have different results. We analyze the average of the similarity Q for different energy levels (see Fig. 9). We find that there is a decorrelation between Q and E as the frustration Δ increases except for the cases near the low-energy part. For the case of $\Delta=0.9$, in a broad range of energies (from 0 to -25), the average of Q oscillates around a value, which is similar to the former case mentioned above. Therefore, there is a weak correlation between Q and E. Nevertheless, in the energy range from -25 to -28, there is a good linear relationship between Q and E. Therefore, the PFZs relating to the high-energy part (namely the zeros corresponding to small β) have little effects for behaviors of the fluctuation of Q, $\sigma_Q(\beta)$, and the leftmost zero line dominates the real temperature behavior of the fluctuation σ_Q. As a result, for this model protein system, the feature of quantity Q for real temperature can be used to identify the folding transition, but reflect little information of the change of other thermodynamical features of the system, such as those found from the variation of heat capacity. Therefore, the set of zeros shows much information, independent of the physical quantities that are used. In other words, one should be very careful in using any quantity to analysis the phase-diagram of the system. This also shows the necessity and effectivity of using the PFZs.

D. Cooperativity and foldability of various Gō models

The cooperativity is another regularly used term for characterizing the folding transition. The calorimetric cooperativity has been carefully investigated in the literatures.[13,14] This quantity characterizes the sharpness of the transition and has been considered as an important factor for describing the folding transition. What is the relation between the cooperativity and the phase-transition-related factors as studied above? Here, we use the quantity κ_2 in Ref. 14 as the definition of the cooperativity. That is, we have

$$\kappa_2=2\sqrt{k_B C_p(\beta_{\max})}/\beta_{\max}\Delta H_{\mathrm{cal}}. \tag{6}$$

Here, $\beta_{\max}=1/T_{\max}$ corresponds to the peak of the heat capacity C_p. Note that for the cases with large frustration, there are many peaks. We use the sigmoidal fit for the local change of the energy related to the folding transition (this figure is not presented here), and then obtain the calorimetric enthalpy ΔH_{cal} related to this transition and the corresponding cooperativity. This definition coincides with the definition used in Ref. 14 for the cases with a single transition. We investigate the correlations between the cooperativity κ_2 with the factors β_c, τ_c, γ, and α. It is found that there is good correlation only between κ_2 and γ in the whole range of frustration Δ with $\Delta<1$, even for a very large frustration (such as $\Delta=0.9$) (see Fig. 10). This implies the intrinsic connection between the strength of the transition and the cooperativity, and is similar to a suggested definition of cooperativity from the relative shapeness of the peak of heat capacity by Klimov and Thirumali.[58] The high cooperativity means a strong transition between the denatured state and native state, which is intuitively correct. This relationship provides a bridge to understand the importance of the transition properties for the folding thermodynamics and kinetics. It is worth pointing out that the folding transition for large Δ shows a good cooperativity because of the restricted energy distribution of the related unfolded state ensemble. The relations with other factors are not linear, especially for the cases with large frustrations. The correlation of κ_2 with the density variance is very bad, implying that the cooperativity has little relation with the order of the transition. When we limit our interests in the region with a single peak of the heat capacity, the monotonic relationship between the cooperativity and the factor β_c or τ_c is obtained. Considering the correlation of cooperativity with the relative shapeness mea-

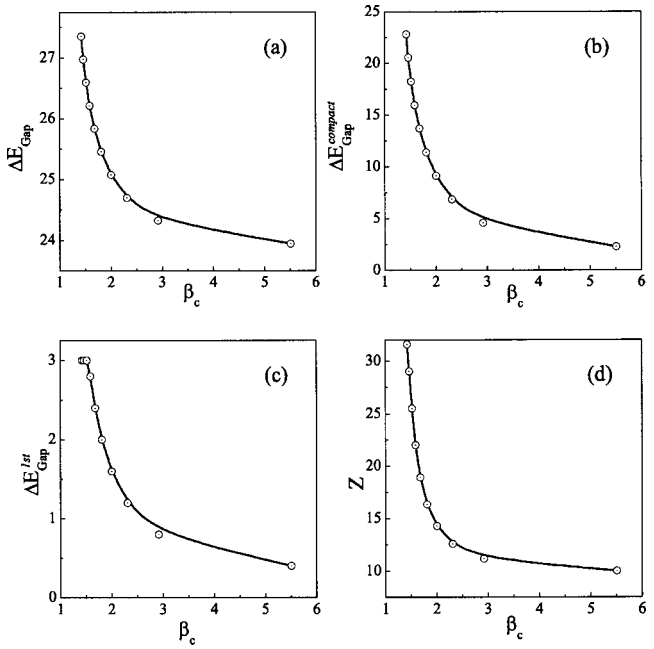

FIG. 11. The relation between the various gaps vs folding temperature β_c, (a) ΔE_{gap}, (b) $\Delta E_{\mathrm{gap}}^{\mathrm{compact}}$, (c) $\Delta E_{\mathrm{gap}}^{1\mathrm{st}}$, and (d) Z score.

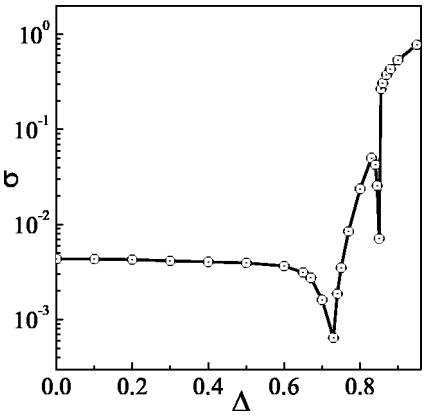

FIG. 12. The relation between the foldability σ and the frustration.

sured with folding temperature, the increase of cooperativity versus β_c and the decrease of κ_2 versus τ_c can be understood straightforwardly. In this sense, the correlation between κ_2 and β_c (τ_c) for the cases with proper frustrations may result from the correlation between κ_2 and γ.

As discussed above, the properties of the folding transition determined from the scheme of the PFZs have good correlation with regular definition. Based on the energy landscape theory, the folding properties of model proteins have a close relationship with the thermodynamic features. What can we understand the foldability from the folding transition? It is an interesting problem receiving much attention. What can we know from the properties extracted from the analysis of PFZs? It is an important aspect for understanding the folding from the PFZs. We study the correlation of the transition factors with some foldability parameters, including the energy gap between denatured states and the native state $\Delta E_{\mathrm{gap}}=\langle E\rangle-E_N$, the first excited energy gap $\Delta E_{\mathrm{gap}}^{1\mathrm{st}}=E_{\mathrm{excited}}^{1\mathrm{st}}-E_N$, the energy gap with the average of the most compact structures $\Delta E_{\mathrm{gap}}^{\mathrm{compact}}=\langle E\rangle_{\mathrm{compact}}-E_N$, the Z score $Z=\Delta E_{\mathrm{gap}}/(\langle E^2\rangle-\langle E\rangle^2)$ and the σ factor defined by Klimov and Thirumalai.[59] Here, the $\langle\cdots\rangle$ or $\langle\cdots\rangle_{\mathrm{compact}}$ means the ensemble average over all conformations or over all the compact conformations, and E_N and $E_{\mathrm{excited}}^{1\mathrm{st}}$ represent the energy of the ground state (the native state) and that of the first excited state. The practical Z score defined based on the set of the most compact structures is not discussed here,[60] since this quantity shows a trend different from the theoretical arguments for the Go̅-type models. This is attributed to a bad estimation on the width of the denatured ensemble from the set of the compact structures. Due to the linear variation of the various gaps, monotonic changes of the gaps with the transition factors are expected as shown in the Fig. 11. There is a small horizontal section for the case $\Delta E_{\mathrm{gap}}^{1\mathrm{st}}$ [see Fig. 11(c)], which suggests that the first excited gap is not suitable for describing the foldability in this range. These all qualitatively show that the proteins with lower transition temperatures have a worse foldability. This is consistent with other arguments. Clearly, for large values of β_c (i.e., low temperature), there is a saturation of various gaps. Therefore, the deterioration of the transition temperature becomes sensitive to the gap size. For the scaled gap, Z score, similar conclusions can be obtained. Thus, for the largely frustrated systems, the simple gap is not a good parameter for understanding the folding transition and furthermore the foldability. This is also coincident with the conventional observations.[61,62] While for the foldability parameter σ, the relation is a little different. For simplicity, we choose the position of the first peak of the heat capacity to define the collapse temperature T_θ since generally it relates to the largest energy jump. Roughly, the total tendency of σ increases as the frustration increases (see Fig. 12). In more detail, for the case of small values of frustration, the σ varies little, even decreases as β_c increases. After a threshold frustration, the parameter σ increases with an increase of β_c. When a new phase is approached from the heat capacity curve, there is generally a drop of σ. Therefore, for the systems with a common phase structure, the parameter σ can approximately identify the foldability. The validity of σ for foldability is a qualitative conclusion, which needs a careful speculation. However, the parameter σ is more sensitive than the regular gap-type quantities. The change of the phase structure can been found from the variance of σ. It is also interesting to point out that $\sigma=0.1$ was set as the threshold between the fast folder and the folders with an intermediate rate.[63] This parameter can be understood easily from our results, since the value 0.1 clearly separates the cases with small frustrations ($\Delta<0.85$) and those with large frustrations ($\Delta>0.85$) from Fig. 12.

E. Mapping between different protein models

Finally, it is interesting that the results obtained from our simulations on the Go̅ models can be extrapolated to other models with different interactions and sequence characters. Based on the patterns of the PFZ distributions, we can map the various models into one of Go̅ cases with some certain frustrations. This can be understood as follows. We define a

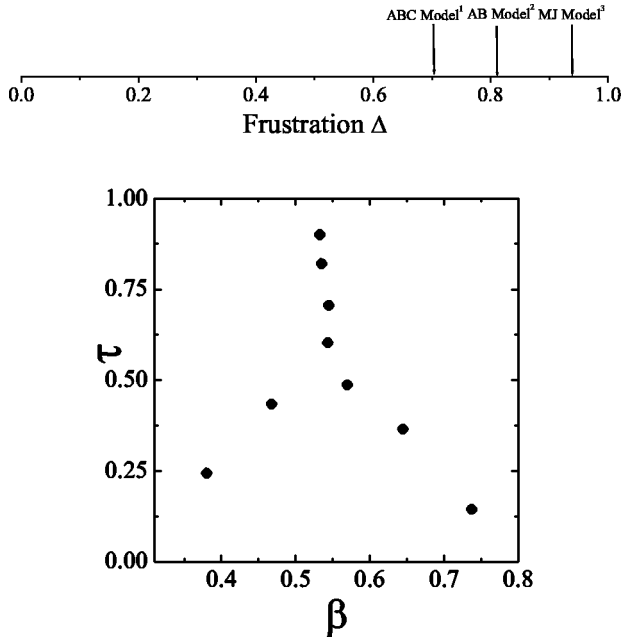

FIG. 13. Mapping of different models according to the frustrations: (1) the model in Ref. 64; (2) the model in Ref. 43; (3) the potential of Table 3 in Ref. 65.

''distance'' between two models M_1 and M_2 with the mean of the squares of the various parameters, that is,

$$D_{M_1M_2}=\sqrt{\left(\frac{\beta_{c,M_1}}{\tau_{c,M_1}}-\frac{\beta_{c,M_2}}{\tau_{c,M_2}}\right)^2+(\gamma_{M_1}-\gamma_{M_2})^2}. \qquad (7)$$

Here it is worth noting that the factor α varies little for folding transitions so that we omit it in Eq. (7). The combination of the two factors β_c and τ_c is used to eliminate the correlation between these two factors. Comparing the selected model M with the Go̅ models with various non-native interactions, the frustration Δ_M, which results in the minimal distance between the models, sets up a mapping of the model M to the Go̅-like model. We study some models or potentials employed in other studies (see the caption of Fig. 13). They are mapped into the space coordinated with frustration. Some examples of their PFZ distributions are shown in Fig. 13(b). This kind of ''linearization'' provides a way to compare various folding models and potentials. For example, the distribution of the PFZs of a model protein with two kinds of monomers is shown in Fig. 13. The corresponding parameters related to the folding transition are $\beta_c=0.798$, $\tau_c=0.145$, $\gamma=-0.419$, and $\alpha=1.118$, respectively. The related frustration for the Go̅ model is about 0.71. From these mappings, we can understand the thermodynamics of various models and their folding transition by employing the mentioned above natures obtained from the Go̅-type models. From this aspect, the PFZ scheme also provides us with a regular method to evaluation various folding models from their thermodynamics.

V. CONCLUSION

In this work, we apply a method to analyze the folding transition based on the distribution of the partition function zeros. Several factors related to the transition are extracted from the distribution of the zeros. These factors characterize the position, the strength, the order, and the finite-size effect of the folding transition. These factors for the Go̅-type models with and without non-native interactions are studied. It is found that transition of folding is approximately of first order, which does not depend on the chosen native structure. The longer chains increase the strength of the transition. These results qualitatively coincide with the previous understandings.[17–21] The PFZ scheme provides a straightforward method without any further assumptions for the reaction coordinates in analysis. The factors related to the transition effectively identify the properties of folding transition. Different regular transition-related quantities are evaluated. Their validity and physical background are discussed through these factors. The thermodynamic foldability factors are also compared with these factors. The fitness of these factors for the description of foldability are discussed. Finally we discuss briefly the mapping between different models and the Go̅-type models using the PFZ method, which possibly validates the conclusions of our studies for other protein models.

Conclusively, the method based on PFZ distributions provides a systematic way to analyze the thermodynamic properties of the model protein systems. Analysis of these zeros was recently introduced as a tool for investigating structural transitions in proteins[17–21] and is used here to study Go-type models with and without non-native interactions. Comparing with the regular schemes, the distribution of zeros includes much information rather than a single parameter.[59,61] This is consistent with the present understanding on protein folding.[66] This knowledge from the distribution of partition function zeros enables us to extract the thermodynamic information without artificial assumptions on the related processes. Furthermore, this scheme of the PFZs is helpful for the analysis of some other first-order phase transitions.[67] With a suitable form of free energy, not only the regular thermodynamic transitions but also some other phase-transition-like processes such as the random first-order transition can be characterized with the related partition function zeros. Finally, it is worthy to note that the error-bars of our results are fairly small for the cases with weak frustrations, while there are some large fluctuations for the cases with strong frustrations due to the inefficiency on sampling. This means that the method based on the PFZ distributions is useful for analyzing the folding transitions of finite-size systems, such as the proteins, and our conclusions are rather robust for the optimal protein systems.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (Nos. 90103031, 10074030, 10021001, and 10204013), and the Nonlinear Project (973) of NSM. J.W. is a Fellow of the Ke-Li Research Foundation. The authors acknowledge valuable comments by Professor T. K. Lee, during our visit to the CTS, Taiwan, where part of simulations were done.

APPENDIX: PFZ: THEIR EVALUATION AND CHARACTERIZATION

As for a system with a density of state (DOS), namely $n(E)$, the partition function for a canonical ensemble has the form,

$$Z=\int n(E)\exp(-\beta E)dE, \tag{A1}$$

where $\beta=1/T$ is the inversion of the temperature T, and E is the energy of the system. Consequently, the average of a physical quantity A for the canonical ensemble can have the form

$$\langle A\rangle=\frac{\int A(E)n(E)\exp(-\beta E)dE}{Z}. \tag{A2}$$

It is easy to find that the divergence of quantity A, namely the singularities of the function $A(\beta)$, can be determined directly from the zeros of the partition function $Z(\beta)$, that is, the solutions of the equation $Z(\beta)=0$. Therefore, the problems for studying the thermodynamic transition between different phases can be translated into the one to study the characteristics of the distribution of the PFZs. The consideration from the PFZs is consistent with the regular Erenfest criterion[31] for the systems under thermodynamic limits and is more flexible in analyzing the finite-size systems.[17,18,21,31,32]

Generally, in the complex temperature plane, the partition function Z of a lattice protein can be formulated in terms of the function of the inversion of the complex temperature B as

$$Z(B)=\sum_E n(E)e^{-BE}=\sum_k n'(k)z(\beta)^k$$

$$=\Re(B)\prod_i [z(B)-z_i]. \tag{A3}$$

Here $\Re(B)$ is the regular part of the partition function, which is analytic in the whole complex plane. The function $z(B)$ is defined as $z(B)=e^{-B\epsilon}$ and ϵ is the unit of the energy to ensure that $k=E/\epsilon$ is an integer. The coefficients of the z-related polynomial $n'(k)$ has the form $n'(k)=n(k\epsilon)$. The parameters $\{z_i\}$ are a series of constants related to the PFZs. Therefore, the solutions $\{B_i=\beta_i+\omega\tau_i\}$ of the equation $Z(B)=0$, namely $Z(B_i)=0$, have the form $B_i=-\ln(z_i)/\epsilon$ with the integer i being the solution index. Here β_i and τ_i are the real and the imaginary part of the complex solution B_i, and ω is the unit of the imaginary number. It is clear that the number of zeros depends on the number of the energy levels and thus the size of the system. Since the partition function is a real function for real temperatures (real β), B_i and its conjugate B_i^* are all the solutions of the equation $Z=0$. There is a mirror symmetry with respect to the real axis for the distribution of the PFZs in the complex temperature plane. Due to the periodicity of the exponential function along the image-axis direction, we have $\exp(-B_i\epsilon+\omega 2n\pi)=\exp(-B_i\epsilon)$, where n is an integer. Therefore, along the image-axis direction, B_i has a period $2\pi/\epsilon$. Considering these two symmetries, we generally need to investigate the region with imaginary part $\tau\in[0,\pi/\epsilon]$.

Practically, one way to extract the PFZs is to solve explicitly the polynomial equation $Z(z)=0$ [see Eq. (A3)]. With the solutions $z=z_i$, it is easy to deduce the corresponding β_i from the relation $\beta_i=-\ln z_i/\epsilon$. This method is rather efficient when applying to systems with a few degrees of freedom. Another usual estimation for the zeros is to scan the variation of physical quantities (such as the heat capacity) around the whole complex-temperature plane to monitor their divergence. This idea is commonly used for the real-temperature cases. The complex temperatures corresponding to the divergence are generally the zeros of the partition function. The choice of the thermodynamic quantities to monitor are not required to be unique since the singularity is an inherent characteristics of the partition function and generally appears in various thermodynamic functions. With this method, not only can the zeros be extracted easily, but also we can find out the variation of some physical quantities that include more information about the systems. At any rate, for the zeros, these two methods produce the same results. It is worth noting that, for the same set of PFZs, the behavior of various quantities in the real-temperature range may be different in some cases since not all PFZs contribute equally to the thermodynamic quantities. Therefore, sometimes the transitions cannot be identified from the behavior at the real temperature based on some certain quantities in finite-size systems, but the analysis based on the PFZs does not suffer from this feature. This shows the power of the analysis based on the PFZs (more detailed discussions are presented in Sec. IV).

To describe the distribution of the obtained zeros, we adopt the parameters proposed and studied in Refs. 19 and 31. Since the thermodynamic properties existing at real temperatures are largely affected by the zeros near the real axis, the distribution of zeros near the real temperature axis is important for physical studies. For example, we consider the Go-model case without energetic frustration, that is $\Delta=\epsilon_{nn}/\epsilon_n=0$. For such an optimized proteinlike system, the zeros are all aligned on a line (as seen in Fig. 1). Near the real axis, the zeros are almost in a straight line, which intersects with the real axis at a positive intercept, β_c. This intercept β_c generally is considered as the inversion of the predicted transition temperature. Due to the importance of the zero nearest to the real axis, the image part of this zero, τ_c, is another key parameter for the system. By considering the thermodynamic limit, the value of parameter τ_c indicates the nearness of the system to that with an infinite size, that is, with τ_c being a measure on the finite-size effect of the system. Besides this zero, other zeros near to the axis also have their important contributions to the thermodynamic behavior of the system. To characterize the span of the transition, the slope of the cutting line composed of zeros is proposed as

$$\gamma=\frac{\beta_2-\beta_1}{\tau_2-\tau_1}, \tag{A4}$$

which have some connections with the strength of the transition. Here the index 1 and 2 indicate the first zero and the second zero from the real axis. A small value of γ means a vertical line to the real axis, that is, almost all the zeros have the same real part. There is a small span due to the finite-size

effect. For a large value of γ, however, the line of zeros affects a larger range of temperatures, which corresponds to a wide span of the transition region. Another parameter, the local density of zeros ϕ near the real axis, is employed to characterize the contribution of the zeros besides the nearest one to the real axis, which is generally considered to have some relation with the order of the transition. For the discrete distribution of zeros, the density ϕ is defined based on the interspace between the neighboring zeros. A regular form of the zero density is

$$\phi_i = \frac{1}{2}\left(\frac{1}{|B_i - B_{i+1}|} + \frac{1}{|B_i - B_{i-1}|}\right), \tag{A5}$$

where B_i is the complex solution corresponding to the ith zero along the transition line. Sometimes, when the line of zeros is rather short (due to the competition between different phases), another discrete form of the density of zeros is used as follows:

$$\phi_{i+0.5} = \frac{1}{|B_i - B_{i+1}|}. \tag{A6}$$

It is argued that the density of the zeros would vary with a power-law form, that is, $\phi_i \sim \tau_i^{\alpha}$. The exponent α shows the variation of the density of zeros. Practically, the exponent α can be fitted from the relation of ϕ_i versus τ_i. Here we adopt a discrete differential as the form of the exponent,

$$\alpha = \frac{\ln(\phi_2) - \ln(\phi_1)}{\ln(\tau_2) - \ln(\tau_1)} \tag{A7}$$

or

$$\alpha = \frac{\ln(\phi_{2.5}) - \ln(\phi_{1.5})}{\ln[(\tau_2 + \tau_3)/2] - \ln[(\tau_1 + \tau_2)/2]}. \tag{A8}$$

The order of the transition then can be identified from this exponent α. The relation $\alpha \leqslant 0$ corresponds to the first-order transition, while a positive α suggests a high-order transition. Obviously, this coincides with the Ehrenfest criterion for the infinite systems clearly. The illustration for these quantities can be clearly seen from Fig. 1. With these quantities, the basic features of the distribution of the zeros near the real axis are outlined and the properties of transition can be determined.[19,31]

[1] T. E. Creighton, *Protein Folding* (Freeman, New York, 1992).
[2] A. R. Fersht, *A Guide to Enzyme Catalysis and Protein Folding* (W. H. Freeman, New York, 1999).
[3] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[4] K. A. Dill and H. S. Chan, Nat. Struct. Biol. **4**, 10 (1997).
[5] A. R. Dinner, A. Sali, L. J. Smith, C. M. Dobson, and M. Karplus, Trends Biochem. Sci. **25**, 331 (2000).
[6] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[7] A. R. Fersht, Proc. Natl. Acad. Sci. U.S.A. **92**, 10869 (1995).
[8] V. S. Pande, A. Yu Grosberg, and T. Tanaka, Rev. Mod. Phys. **72**, 259 (2000).
[9] H. Wu, Chinese J. Physiol. **5**, 321 (1931); Am. J. Physiol. **90**, 562 (1929).
[10] C. B. Anfinsen, Science **181**, 223 (1973).
[11] C. J. Camacho, Phys. Rev. Lett. **77**, 2324 (1996).
[12] Z. W. Xing, J. Wang, and W. Wang, Phys. Rev. E **58**, 3552 (1998).
[13] H. S. Chan, Proteins: Struct., Funct., Genet. **40**, 534 (2000).
[14] H. Kaya and H. S. Chan, Proteins: Struct., Funct., Genet. **40**, 637 (2000).
[15] E. I. Shakhnovich, Curr. Opin. Struct. Biol. **7**, 29 (1997).
[16] J. Wang, K. Fan, and W. Wang, Phys. Rev. E **65**, 041925 (2002).
[17] N. A. Alves and U. H. E. Hansmann, Phys. Rev. Lett. **84**, 1836 (2000).
[18] N. A. Alves and U. H. E. Hansmann, Physica A **84**, 509 (2001).
[19] N. A. Alves, J. P. N. Ferrite, and U. H. E. Hansmann, Phys. Rev. E **65**, 036110 (2002).
[20] N. A. Alves, U. H. E. Hansmann, and Y. Peng, Int. J. Mol. Sci. **3**, 17 (2002).
[21] N. A. Alves and U. H. E. Hansmann, J. Chem. Phys. **117**, 2337 (2002).
[22] H. Kaya and H. S. Chan, Phys. Rev. Lett. **85**, 4823 (2000).
[23] K. Fan, J. Wang, and W. Wang, Phys. Rev. E **64**, 041907 (2001).
[24] H. Chen, X. Zhou, and Z. C. Ou-Yang, Phys. Rev. E **64**, 041905 (2001).
[25] I. M. Lifshitz, A. Yu Grosberg, and A. R. Khokhlov, Rev. Mod. Phys. **50**, 683 (1978).
[26] D. K. Klimov and D. Thirumalai, Phys. Rev. Lett. **79**, 317 (1997).
[27] H. Chen, X. Zhou, and Z. C. Ou-Yang, Phys. Rev. E **63**, 031913 (2001).
[28] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Clarendon, Oxford, 1983).
[29] C. N. Yang and T. D. Lee, Phys. Rev. **97**, 404 (1952); **87**, 410 (1952).
[30] M. E. Fisher, in *Lectures in Theoretical Physics* (University of Colorado Press, Boulder, 1965), Vol. 7c, p. 1.
[31] P. Borrmann, O. Mulken, and J. Harting, Phys. Rev. Lett. **84**, 3511 (2000).
[32] O. Mülken, P. Borrmann, J. Harting, and H. Stamerjohanns, Phys. Rev. A Phys. Rev. A **64**, 013611 (2001).
[33] W. Janke and R. Kenna, Nucl. Phys. B **106**, 905 (2002); J. Stat. Phys. **102**, 1211 (2001).
[34] R. Kenna and J. C. Sexton, Phys. Rev. D **65**, 014507 (2002).
[35] N. Go, J. Stat. Phys. **30**, 413 (1983).
[36] N. Go, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).
[37] D. Baker, Nature (London) **405**, 39 (2000).
[38] J. Wang and W. Wang, Nat. Struct. Biol. **6**, 1033 (1999).
[39] H. S. Chan, Nat. Struct. Biol. **6**, 994 (1999).
[40] J. Wang and W. Wang, Phys. Rev. E **61**, 6981 (2000).
[41] J. Wang and W. Wang, Phys. Rev. E **65**, 041911 (2002).
[42] S. Takada, Proc. Natl. Acad. Sci. U.S.A. **96**, 11698 (1999).
[43] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **101**, 1519 (1994).
[44] J. J. de Pablo, Q. L. Yan, and F. A. Escobedo, Annu. Rev. Phys. Chem. **50**, 377 (1999).
[45] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
[46] B. A. Berg, J. Stat. Phys. **82**, 323 (1996).
[47] U. H. E. Hansmann and Y. Okamoto, Curr. Opin. Struct. Biol. **9**, 177 (1999).
[48] H. Li, R. Helling, C. Tang, and N. S. Wingreen, Science **273**, 666 (1996).
[49] B. A. Shoemaker and P. G. Wolynes, J. Mol. Biol. **287**, 657 (1999); B. A. Shoemaker, J. Wang, and P. G. Wolynes, **287**, 675 (1999).
[50] J. Chuang, A. Yu Grosberg, and M. Kardar, Phys. Rev. Lett. **87**, 078104 (2001).
[51] Y. Q. Zhou and M. Karplus, Nature (London) **401**, 400 (1999).
[52] V. S. Pande and D. S. Rokhsar, Proc. Natl. Acad. Sci. U.S.A. **95**, 1490 (1998).
[53] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).
[54] A. Dinner, A. Sali, M. Karpus, and E Shakhnovich, J. Chem. Phys. **101**, 1444 (1994).
[55] M. Oliveberg, Curr. Opin. Struct. Biol. **11**, 94 (2001).
[56] K. W. Plaxco, I. S. Millet, D. J. Segel, S. Doniach, and D. Baker, Nat. Struct. Biol. **6**, 554 (1999).
[57] T. Veitshans, D. K. Klimov, and D. Thirumalai, Fold Des **2**, 1 (1997).
[58] D. K. Klimov and D. Thirumalai, Fold Des **3**, 127 (1998).
[59] D. K. Klimov and D. Thirumalai, Phys. Rev. Lett. **76**, 4070 (1996).
[60] J. Wang and W. Wang (unpublished).
[61] A. R. Dinner, V. Abkevich, E. Shakhnovich, and M. Karplus, Proteins **35**, 34 (1999).
[62] N. E. G. Buchler and R. A. Goldstein, J. Chem. Phys. **111**, 6599 (1999).
[63] D. K. Klimov and D. Thirumalai, Curr. Opin. Struct. Biol. **9**, 197 (1999).
[64] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, J. Chem. Phys. **104**, 5860 (1996).
[65] S. Miyazawa and R. L. Jernigan, J. Mol. Biol. **256**, 632 (1996).
[66] A. R. Dinner and M. Karplus, Nat. Struct. Biol. **8**, 21 (2001).
[67] M. P. Eastwood and P. G. Wolynes, cond-mat/0206006 (2002).