Eur. Phys. J. B **30**, 381–391 (2002)
DOI: 10.1140/epjb/e2002-00393-4

THE EUROPEAN PHYSICAL JOURNAL B

# Folding of lattice protein chains with modified Gō potential

K. Fan, J. Wang, and W. Wang[a]

National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing 210093, PR China

Received 10 April 2002 / Received in final form 20 August 2002
Published online 19 December 2002 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2002

**Abstract.** We propose a modified Gō model in which the pairwise interaction energies vary as local environment changes. The stability difference between the surface and the core is also well considered in this model. Thermodynamic and kinetic studies suggest that this model has improved folding cooperativity and foldability in contrast with the Gō model. The free energy landscape of this model has broad barriers and narrow denatured states, which is consistent with that of the two-state folding proteins and is lacked for the Gō model. The role of non-native interactions in protein folding is also studied. We find that appropriate consideration of the contribution of the non-native interactions may increase the folding rate around the transition temperature. Our results show that conformation-dependent interaction between the residues is a realistic representation of potential functions in protein folding.

**PACS.** 87.15.Aa Theory and modeling; computer simulation – 87.15.Cc Folding and sequence analysis – 87.15.He Dynamics and conformational changes

## 1 Introduction

Protein folding remains one of the most challenging problems in structural biology [1–8]. Although it is generally accepted that the structural information of a protein was encoded in its amino acid sequence [9], how this information is encoded is still not well understood. Even there is an astronomically large number of possible conformations, a protein can find its unique native conformation within a physiologically short time. This indicates that protein folding is not a randomly native state searching process but a highly cooperative one. An "old view" interprets the reason of fast folding as that folding is progressed upon a definitive pathway [10]. Differently, the "new view" describes the energy landscape of folding as funnel-like, in which the native state places at the bottom of the funnel [11]. In the new view, there are many parallel pathways downwards to the native state. Both in the old view and in the new view, the native conformation locates at the global free energy minima, and folding is biased towards the native state [11,12].

In the theoretical studies of protein folding, a major issue is the choice of the potential functions [13,14]. An appropriate potential function should describe the real situation of the energetics of protein as much as possible and at the same time be computationally feasible for the folding simulation. In the pioneering work by Gō and his collaborators [15], they proposed a potential function to interpret the protein structures in which only interactions presented in the native state (the so-called native interactions) are considered, and interactions not presented in the native state (the so-called non-native interactions) have no contribution to the energy of the system. Because there is no energetic frustration in the Gō model, the energy landscape is of almost perfect funnel shape and the energy of the native conformation is absolutely the lowest. However, during the folding process, it seems to be unrealistic that only native interactions are considered and there are no strength difference between the various native contacts. Therefore, some researchers developed more "realistic" potentials, such as the hydrophobic and polar (HP) model [16] and the Miyazawa-Jernigan matrix (the so called MJ matrix [17]) or similar interaction matrices. Especially, those statistical potentials, *e.g.* the MJ matrix, extracted from the pairing frequencies of 20 kinds of amino acids in databases of protein structures [14,17,18], were expected to represent the energetics of protein system more faithfully than the Gō potential. Nonetheless, there is no stringent evidence that those knowledge-based potentials are superior to the Gō potential.

Recently, a number of researchers utilized the Gō-like potential in their simplified theories to predict the folding rate, and obtained fair predictability for many fast folding proteins [19–21]. Besides, another exciting finding was made by Baker's group. They found that there is a significant correlation between the contact order (a quantity defined as the average sequence separation of contacts in the native structure) and folding rate [22], suggesting that the

[a] e-mail: wangwei@nju.edu.cn

physical mechanism underlying protein folding would be surprisingly simple [23]. Many experiments also indicated that the role of topology is more important than that of the sequence details in protein folding [24–26], which is in consistent with the point of view that topology is a major determinant of folding kinetics [27]. This relates to the Gō model, in which the role of the native topology of a protein is emphasized, while the sequence details are reduced, and explains to some extent why the Gō-like potentials, although not "realistic", make great success in predicting the folding rate for small globular proteins.

Is the Gō model a perfect one for protein folding? The answer may be negative. There is a number of evidences indicating that there are some limits to the description of the perfect funnel model. First, it is found experimentally that there is small, but non-negligible, number of residues which have abnormal $\phi$-values (larger than unity or smaller than zero). These have been interpreted as the effects of interactions that are not presented in the native state but are important in the transition ensemble [28,29]. Second, it is also found recently that there are broad barriers and multiple transition ensembles as well as narrow denatured state in the free energy landscape for the two-state folding proteins [30–32]. But this is lacked for the Gō model. Recent works made by Chan *et al.* also indicated that based on the calorimetric criterion, the Gō model is far from a two-state model [33–35]. In fact, most of the questions are directly related to some flawed assumptions in the Gō potential. For example, the interactions are assumed to be additive, and they are the same during the folding as in the native state, *i.e.*, interactions are environment-independent. This is in confliction with the recent experimental results, which clearly show that interactions in the native state are different from those in the intermediate state and transition state, and specially, interactions are weaker in the folding process than in the folded state [36–38]. Thus, the viewpoint that interactions are environment-dependent is more and more widely accepted [39,40]. Another example of the flawed assumption is that the interactions are assumed to be the same between the residues at the surface as those in the interior. Proteins usually have a closely packed core but a rather mobile shell, which indicates clearly an existence of differences between interactions at the surface and in the core [41]. Besides, interactions are assumed to be temperature-independent. This assumption makes present model could not explain the cold denaturation [42]. Taken together, we can say that the Gō model is a beautiful model but far from a perfect model.

Fortunately, with the development of experimental methods and theoretical studies, our understanding of protein and protein folding has been greatly improved in the last decade. Based on these understanding, one can make some reasonable modification on the Gō model to enable it to work more realistically. There have been a number of works done in this direction, using both lattice [43–45] and off-lattice models [46]. The non-additive effects were incorporated into the models in manner of solvation forces [44], local atom density dependent hydrophobic interaction [46], or some predefined manner [43], etc. Some introduced hydrogen bond strength at the same time [46]. In addition, some other important issues on non-additive effects of hydrophobic forces are also discussed in references [47–52]. The folding cooperativity was indeed improved with these modification. However, these models are not necessarily two-state, since they may not satisfy the calorimetric two-state criterion [33,34]. In addition, there are also a number of works related to the studies on the cooperativity of protein folding [53–56]. In this work, we only consider the non-additive hydrophobic force without introducing other forces. Such a modified Gō model is referred to as the Gō++ model. (The Gō-like model with different modification in reference [57] is referred to as the Gō+ model.) It is found that the present Gō++ model could reproduce the two-state folding behavior comparable with that of real proteins.

This paper is organized as follows. Our model and method are described in Section 2. In Section 3 we present our results on thermodynamics and kinetics of folding for both the Gō model and the Gō++ model. The role of non-native interactions is studied in Section 4, and a comparison with the Gō+ model is made in Section 5. Finally we give a summary and outlook in Section 6.

## 2 Model and method

We use the most widely used lattice model. A protein chain is represented by a self-avoiding walk on a cubic lattice. An amino acid residue is reduced to a hard sphere at the lattice site, for which the side chain and atomic details are ignored. If two nonbonded residues are spatially neighboring, we say that they are in contact. If a contact is the same as it presents in the native structure, it is called a native contact, otherwise a non-native contact. In the Gō model, all the native interactions are attractive, and the non-native interactions have no contribution to the energy. The energy function of the chain can then be written as

$$E = \sum_{i<j} \Delta_{ij}^{N} B_{ij}, \tag{1}$$

where $\Delta_{ij}^{N}$ is unity when residues $i$ and $j$ form a native contact, and zero otherwise. Because in the Gō model each interaction contributes an equal energy, one always has $B_{ij} = -\varepsilon$. Here $\varepsilon$ is the unit of energy.

Considering a real protein system, the interaction between the residues, though local in space, certainly depends on the surroundings, which may affect the polarization, steric arrangement, and so on. Therefore, a realistic modification of the Gō model should consider the local cooperation of interactions. Generally, the clustering of native contacts are more preferable than isolated formation of native contacts. Thus a contact formed between residues $i$ and $j$ may have different energies in different conformations, *i.e.*, $B_{ij}$ may change from one conformation to another. Specifically, we define $B_{ij}$ as

$$B_{ij} = -\varepsilon(n_i + n_j)/2, \tag{2}$$

where $n_i$ (or $n_j$) is the number of native contacts for residue $i$ (or residue $j$) formed in a given conformation. The variability of $B_{ij}$ is considered to reflect the cooperativity between residues. In addition, equation (2) also reflects the surface/core difference for a protein in solution. Let us make an argument as follows.

It is well known that there is a significant difference in the number of native contacts between the surface and interior residues of a protein in solution [41]. The residues at surface are in general more mobile than those in the interior and some surface side-chains even have no unique conformations. At the native conformation, residues in the core generally form more contacts than those at the surface, that is, the value of $n_i$ and thus $|B_{ij}|$ for residues in the core are larger than those at the surface. Therefore, in the present model, residues in the core contribute more to the energy than those at the surface, which incorporates the surface/core difference and the cooperative interactions between residues. In addition, residues with more solvent exposure area show weak preference and thus weak interactions, which is qualitatively consistent with the analysis on the solvent effect by many people previously [58,59]. Thus, the surface/core difference is considered in the present model, which is, in some sense, similar to the solvent accessible surface area model of protein folding [59]. Here a contact formed between residues $i$ and $j$ will stabilize, to some extent, other contacts that residue $i$ or $j$ formed with other residues. On the contrary, its breakaway may destabilize those contacts as well. In short, we treat the hydrophobic interactions not be additive but as many-body interactions. This is similar to our previously Gō+ model [57]. Differently, in the Gō+ model, all the native contacts contribute equally to the energy, which could not account for the differences between the surface and the core.

As described above, our modified model has large difference comparing with the Gō model. We introduce a variability of the strength to characterize the interactions during the folding process, which results in a difference in the stability between the core and the surface. We expect these differences could be reflected on the thermodynamics and the kinetics of folding.

To study the thermodynamics and the kinetics of protein folding, we use the standard Monte Carlo method [60,61]. The move set includes corner, crankshaft, end, and null moves, which is believed to have the similar time scale as that of polymer relaxation and may faithfully simulate the process of protein folding [62–64]. The rejection or acceptance of a new conformation is judged with the Metropolis criterion [65].

With respect to the calculation of thermodynamic quantities, such as the specific heat $C_v$ and the population of the native state $P_N$, we use the well-known Monte Carlo histogram method [61,66,67]. The $P_N$ is calculated as follows

$$P_N = \mathrm{e}^{-E_N/T} / \sum_E \Omega(E) \mathrm{e}^{-E/T}, \tag{3}$$

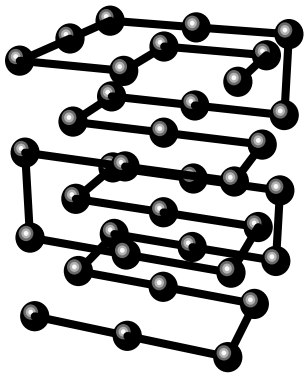

**Fig. 1.** Target native conformation for a 36-mer chain used in this work.

where $\Omega(E)$ is the density of states with energy $E$, and $E_N$ is the energy of the native state. $\Omega(E)$ is calculated with the entropic sampling Monte Carlo method [68–70].

In the theoretical study of protein folding, one usually needs to determine the folding transition temperature, $T_f$. The folding transition temperature is a temperature at which the free energy of the native state is equal to that of the denatured state. There are many methods to determine $T_f$. In this work we use the one proposed by Thirumalai and coworkers [71]. First, we calculate a structural overlap factor $\chi$,

$$\chi = 1 - \frac{1}{N^2 - 3N + 2} \sum_{i \neq j, j \pm 1} \delta\left(r_{ij} - r_{ij}^N\right), \tag{4}$$

where $N$ is the number of residues, $r_{ij}$ is the distance between residues $i$ and $j$ in an arbitrary conformation and $r_{ij}^N$ is the corresponding distance in the native state. The fluctuation in $\chi$ is

$$\Delta\chi = \langle\chi^2\rangle - \langle\chi\rangle^2 \cdot \tag{5}$$

The folding temperature is then determined from the peak of $\Delta\chi$ [71].

## 3 Thermodynamics and kinetics of folding

The results presented below are mainly obtained based on Monte Carlo simulations for a 36-mer chain whose native structure is shown in Figure 1. It is noted that even for different native structures, we have basically similar results as presented in the following.

Figure 2 shows the average $\langle\chi\rangle$ and its fluctuation *versus* temperature $T$ for a 36-mer chain. Note all the temperatures are scaled with the folding temperature $T_f$ for the convenience of comparison. Because there are many differences between the Gō++ model and the Gō model, which results in a large difference on the absolute values of $T_f$ between two models, a direct comparison between the results without the scaling with $T_f$ for the temperature is not meaningful. It is shown that for both models the values of $\langle\chi\rangle$ decrease as $T$ decreases and the degree of decrease

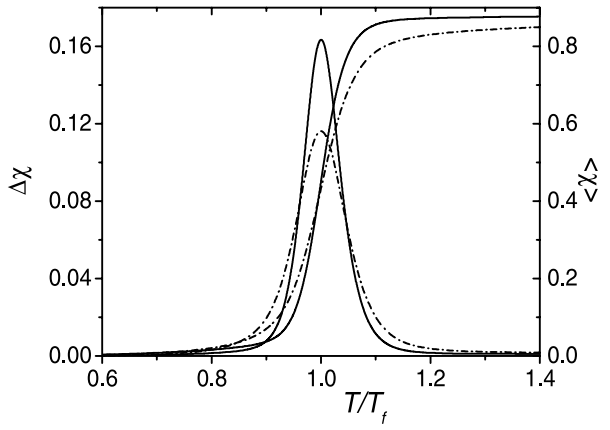


**Fig. 2.** Temperature dependence of the structural overlap factor $\langle\chi\rangle$ and its fluctuation $\Delta\chi$ for the 36-mer chain. Solid lines are for the Gō++ model and broken lines are for the Gō model.

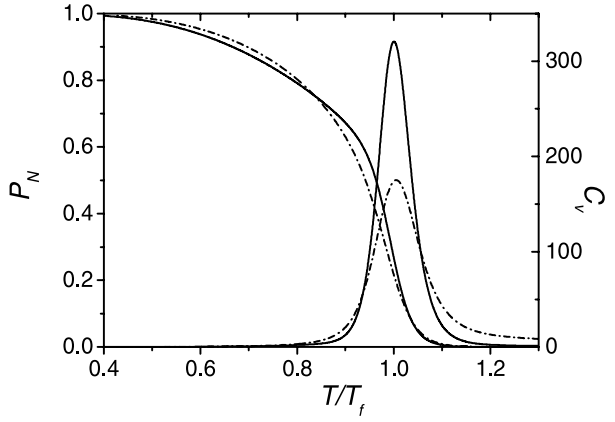


**Fig. 3.** Population of the native state $P_N$ and specific heat $C_v$ varying with the temperature $T$ for the 36-mer chain. Solid lines are for the Gō++ model and broken lines are for the Gō model. Note that the specific heat for the Gō++ model are divided by 25 so as to be comparable with that of the Gō model.

reaches a maximum at the folding temperature, indicating that a sharp folding transition occurs at this temperature. The degree of sharpness of changes in $\langle\chi\rangle$ is a measure of the cooperativity of the folding reaction. From Figure 2 we can see that there is a much shaper transition for the Gō++ model than that of the Gō model, *i.e.*, a sharper changes in $\langle\chi\rangle$. The fluctuation $\Delta\chi$ also gives consistent results which can be seen from the high peak and narrow distribution for the Gō++ model.

Similar phenomenon occurs for the case of the population probability $P_N$ and the specific heat $C_v$. From Figure 3 we can see that near the folding transition temperature $T_f$ there is a sharper change in $P_N$ for the Gō++ model than that for the Gō model, indicating a more cooperatively folding transition. As for the specific heat, there is also a single peak in the $C_v$ curve, and it is narrower than that of the Gō model. Specially, we define the width of the heat capacity peak as the difference in the two temperatures at which the value of heat capacity is half of it maximal value. For the Gō++ model, the width of the heat capacity peak is 0.08 $T_f$, while for the Gō model the width is 0.11 $T_f$. Moreover, for the Gō++ model, the collapse transition temperature $T_\theta$ as determined from the maximum of $C_v$ is the same as the folding temperature $T_f$, indicating that the folding and collapse occur almost simultaneously. Furthermore, the difference between $T_\theta$ and the midpoint temperature of transition with $P_N$=1/2 is also smaller than that of the Gō model, *i.e.*, $\Delta T = T_\theta - T_{P_N=\frac{1}{2}} = 0.03$, providing an alternative evidence for the good cooperativity of folding for the Gō++ model. Similar results are obtained for other chain sizes. Note that for the Gō model, one has $\Delta T = 0.07$.

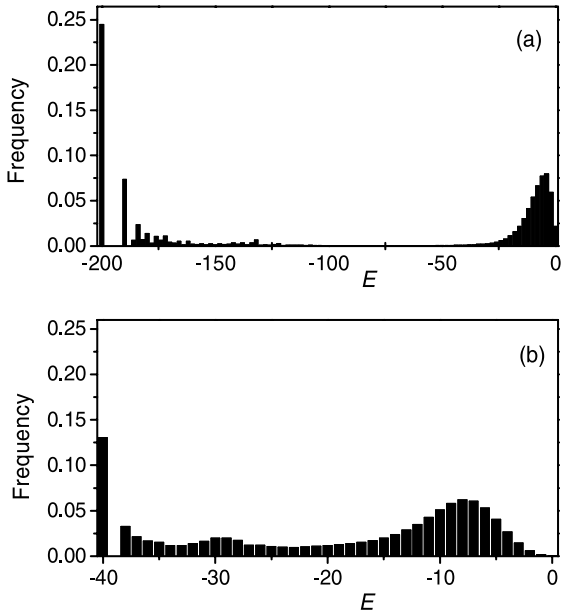


**Fig. 4.** The energy distribution of the 36-mer chain, for (a) the Gō++ model and (b) the Gō model at respective folding transition temperature, $T_f$.

Figures 2 and 3 provide evidences that the Gō++ model improves the folding cooperativity compared with the Gō model. Now we show a more direct evidence that can be found from the equilibrium energy distribution at the folding transition temperature. Figure 4 shows such distributions for both models. Clearly for the Gō++ model there is a good bimodal distribution, and the energies of the denatured states are distributed in a rather narrow region (see Fig. 4a). Furthermore, the native and denatured states are well separated, and there is basically no intermediate states at equilibrium, indicating that the folding is of undoubtedly two-state for the Gō++ model. Differently, for the Gō model as shown in Figure 4b, the bimodal behavior is not so significant as that in Figure 4a and there are many intermediate states. The variation of $Q$ (the number of the native contacts) and $C$ (the number of the total contacts including both native and non-native ones) during the folding also gives consistent results. As shown in Figure 5, for the Gō++ model the folding is clearly a two-state process, the native and denatured states are in rapid equilibrium at $T_f$. While for the

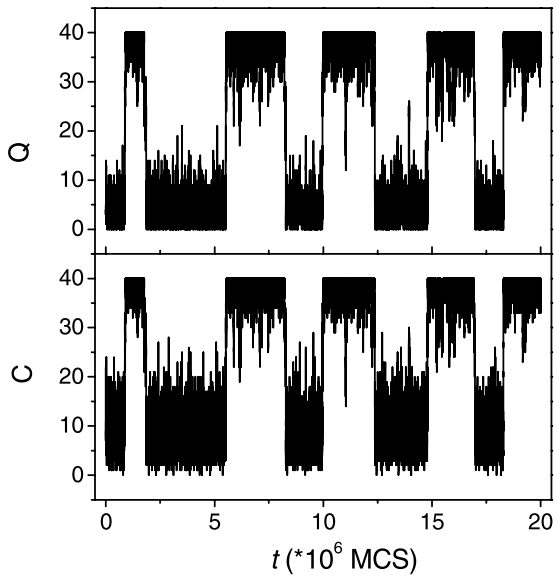


**Fig. 5.** The number of native contacts $Q$ and the number of contacts $C$ *versus* time $t$ in one of the typical trajectories at $T_f$ for the Gō++ model.

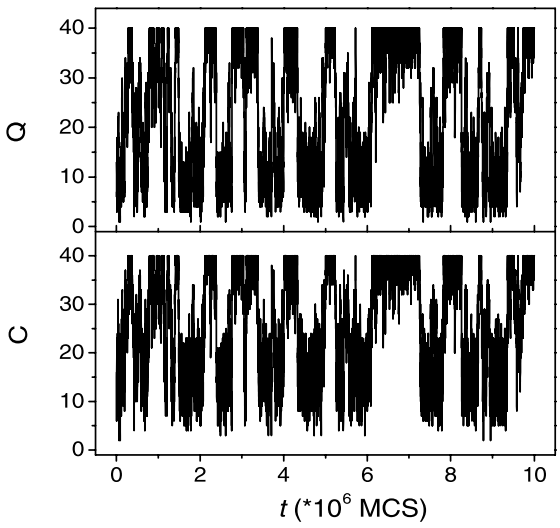


**Fig. 6.** The number of native contacts $Q$ and the number of contacts $C$ *versus* time $t$ in one of the typical trajectories at $T_f$ for the Gō model. The two-state behavior is less significant than that for the Gō++ model.

Gō model, there are obviously intermediates in the folding process (see also Fig. 6). As a result, we conclude that for the Gō model the folding is not of a two-state, or the two-state behavior is not so good. This is in agreement with Chan and Kaya's argument [33,34].

In experiments, to determine whether the folding of a protein is a two-state process or not, a general method is to use the calorimetric criterion. It is well-established that for the two-state folding the van't Hoff enthalpy $\Delta H_{vH}$ around the transition midpoint is equal, or very close, to the calorimetric enthalpy $\Delta H_{cal}$ of the entire transition. In this work, we calculate the ratio of $\Delta H_{vH}/\Delta H_{cal}$ as suggested in reference [33] (here, the definition of $\Delta H_{vH}/\Delta H_{cal}$ is equal to $(k_2)^2$ in reference [33]). It is found that the Gō model, which is considered as a model with minimal energetic frustrations, does not meet the calorimetric two-state criterion and gives the value of $\Delta H_{vH}/\Delta H_{cal}$=0.60. Nevertheless, our Gō++ model gives this value to be 0.80 which is much closer to that of real proteins (for real proteins, the value of $\Delta H_{vH}/\Delta H_{cal}$ is 0.96±0.03 [72]). This provides another evidence for the two-state folding and the good cooperativity of the Gō++ model.

What is the physical origin of the high cooperativity of our modified model, *i.e.*, the Gō++ model? Physically, the high cooperativity of our model may result from the narrow distribution of the denatured states and the high population of the native state at the folding temperature (see also Figs. 3 and 4). In the Gō++ model, due to the collective effect between the interactions, the energy spectrum relating to various conformations is redistributed comparing with that of the Gō model. Specifically, the energies of non-native conformations are moved to higher energy levels. As a result, the number of conformations with intermediate energies decreases and a large energy gap between the non-native conformations and the native one is left. The decrease in the number of conformations with intermediate energies results in a concave segment in the microcanonical entropy curve (data not shown), which is a character of two-state folding in a protein model. At the same time, the large energy gap makes the native state paticularly stable, which is believed to be a necessary condition for cooperative folding [73]. Considering the process microscopically, residues with more native contacts are energetically more favorable. Therefore, the macrostates with lower energy are stabilized respect to those with high energy. Thus the difference between original two states is magnified. This enhances the gap of two-state separation, which may be the physical origin of the two-state cooperativity. It can be further explained from the viewpoint of the free energy profile. Figure 7 shows the free energy profiles for both models, clearly there is a significant difference between the two models. For the Gō++ model, as shown in Figure 7a, the free energy profile has broad activation barriers and the region of the denatured states is rather narrow. This is very different from previous populated free energy profiles for a two-state folding protein which is similar to the case of the Gō model (see also Fig. 7b). However, views on the energy profile have been changed in recent years. The broad activation barriers in free energy profile are considered as a common feature of the two-state folding [30–32], because it gives an alternative interpretation for the curved chevron plots and reasonably account for the large movement of transition state caused by mutation or temperature changes. Our numerical results are strikingly consistent with a phenomenological speculation for the existence of such a free energy profile in references [30] and [31], indicating a two-state folding for the Gō++ model. Thus, the above

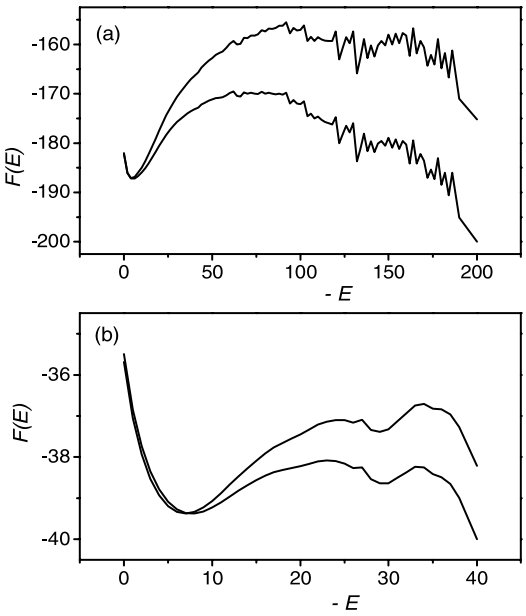


**Fig. 7.** The free energy profile $F(E) = E - TS(E)$ of (a) the Gō++ model and (b) the Gō model at two temperatures. One is higher than $T_f$, the other is lower than $T_f$. The entropy $S(E)$ is calculated using the entropy sampling Monte Carlo method [68]. Note that the free energy profiles at high temperature are shifted overall so that the unfolded states are overlapped.

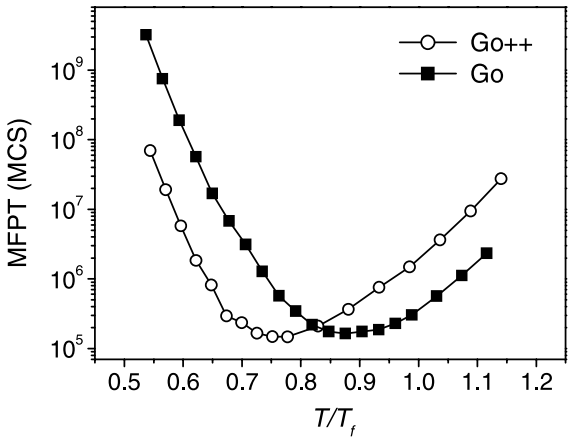


**Fig. 8.** Plot of MFPT of the 36-mer chain at different temperatures for the Gō++ and the Gō model.

comparison provides distinct evidences that the Gō++ model improves the cooperativity and exhibits a good two-state behavior.

Now, we turn to the comparison on the foldability between the two models. In the lattice simulations of protein folding, a common measure of the folding rate is the mean first passage time (MFPT) to the native state. In this work, MFPT is obtained by an average of first passage time (FPT) over 800 runs, and the FPT is the number of Monte Carlo steps (MCS's) consumed in a run. The plots of the MFPT *versus* temperature for the two models are shown in Figure 8. Note that the temperature is scaled with a factor $T_f$. This is because that an identical condition should be taken for the comparison. In the lattice simulations, the temperature has an arbitrary unit and also has no direct relationship with the real temperature. Therefore, the comparison between two different models at the same temperature unscaled may make no sense. Nevertheless, at an identical condition the differences in the foldability can be well-defined. This is similar to other conditions used previously [74,75]. From Figure 8, we can see that the MFPT for the Gō++ model shows a decrease as temperature increases, it reaches a minimum at $T/T_f \approx 0.75$, and then it increases. For the Gō model, there is also a minimum but the location of this minimum is at $T/T_f \approx 0.87$. It is clearly that at a low temperature when the native state is stable (say, $T/T_f \leq 0.7$), the Gō++ model folds significantly faster, *i.e.*, the MFPT is smaller with one or two orders of magnitude than that of the Gō model. This is also consistent with our previous results that larger difference between $T_f$ and $T_{min}$ means better foldability [76]. Physically, this can be explained as follows. From the definition of the Gō++ model we can easily see that the energy gain of forming a contact is usually smaller than that for the Gō model. At high temperatures, entropic contribution is dominant to the free energy barrier, and the loss of entropy is always undercompensated by the energy gain, thus the Gō++ model folds slower for its smaller energy gain. Whereas at low temperatures, folding is nearly a downhill process, and the loss of entropy is always overcompensated by the energy gain. Therefore, for the Gō++ model, it is easier to escape from the kinetic traps, and the folding is faster. It should be noted that here the kinetic traps come from the topological frustration, because there is no energetic frustrations for the Gō and Gō++ models. Because of steric constraint, some native contacts have to be broken before all correct contacts formed. This leads to the topological frustration, which is inevitable even for the Gō model. Finally, we note that for the two models the pathways of reaching the transition state from the denatured state are different. Due to the high cooperativity in our modified model, a good core, the assembly of non-polar residues, is formed much earlier at low temperatures than that in the Gō model. We also note that similar results are obtained for different chain sizes. From Figure 9, we can see that for three chain sizes, the minimal MFPT is almost the same for the two models, but the population of the native state $P_N$ differs largely at $T_{min}$ (see Fig. 10). Obviously, $P_N$ at $T_{min}$ for the Gō++ model is much larger than that for the Gō model, indicating that the foldability is greatly improved for the Gō++ model.

## 4 The role of non-native interactions in protein folding

Recently, the role of non-native interactions in protein folding becomes an interesting issue [28,77–83]. Since the non-native contacts are not presented in the native state,

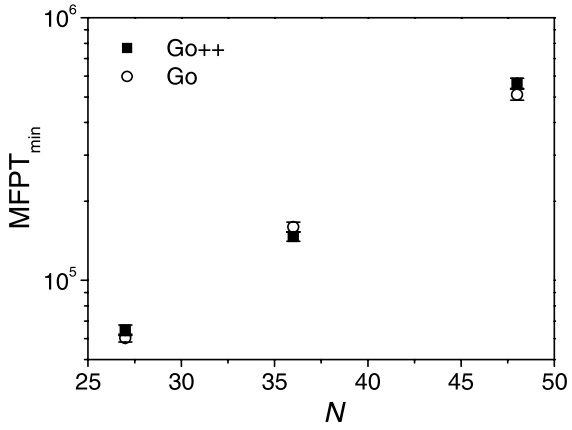


**Fig. 9.** The minimal MFPT ($MFPT_{min}$) *versus* the chain length $N$ for the two models. Note that $MFPT_{min}$ may vary with the target structures.

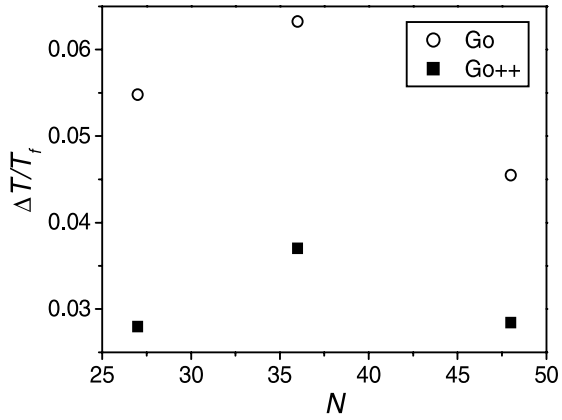


**Fig. 10.** $P_{min}$ *versus* the chain length $N$. Here $P_{min}$ is the population of the native state at the temperature $T_{min}$ at which the folding is fastest. Note that $P_{min}$ may also vary with the target structures.

*i.e.*, they have no contribution to the energy of the native state, one intuitive assumption is that they play a minor role in protein folding with respect to the native ones. Such an extreme example is made for the Gō model in which the role of the non-native interactions is totally ignored. However, recent experiments suggested that the non-native interactions might play significant role in protein folding and should not be ignored in the modeling of proteins. For example, a kinetic important intermediate with non-native helical secondary structures occurs in the folding of beta-lactoglobulin [79,80]. It has also been found that there are many non-native structures in the denatured states of spectrin SH3 domain [81]. Although the role of the non-native interactions in protein folding is not fully clarified, their importance is undoubted. Therefore, a model with proper consideration on the non-native interactions would be more realistic. Then a question arises naturally. What effect do the non-native interactions result in? Surely the formation of non-native contacts may increase the compactness of the structure, which could decrease the solvent exposure area of the formed native contacts and thus stabilize the partially formed native structure. Following a similar consideration as that for equation (2), we can assume that the residue binding preference has the form

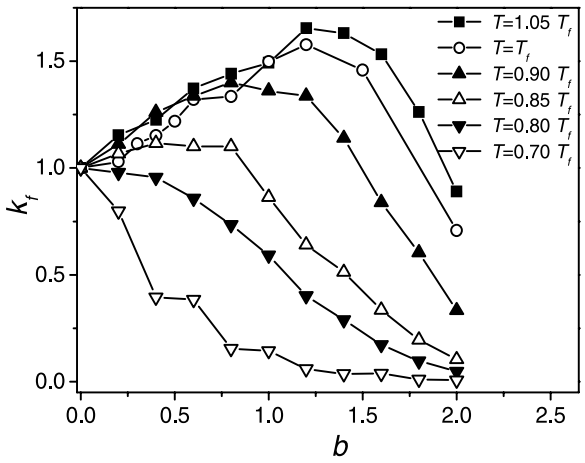


**Fig. 11.** The folding rate $k_f$ *versus* the frustration factor $b$ at various temperatures. Here we assume that the folding rate $k_f = 1/MFPT$, which is true for a single exponential relaxation process [29]. For the convenience of comparison, $k_f$ is scaled to be 1 at $b$=0.

$$B_{ij} = -\varepsilon[(n_i + n_j) + b(n'_i + n'_j)]/2. \quad (6)$$

Here $n'$ is the number of non-native contacts and $b$ is a variant to control how much frustration is added. At this stage, formation of native contacts is not only affected by the local folding progress but also by the local compactness. The contribution of non-native contacts to the energy are thus included in the formula of potential function. Note that some non-native contacts are energetically favorable if their formation could decrease the solvent exposure area of neighboring native contacts. With respect to the case of non-native interaction not included explicitly (see Eq. (1)), equation (6) is a further modification of the Gō model. Thus some of the Gō-like properties should be still kept.

Now we show how this modification affects the folding kinetics. In Figure 11, we can see that at temperatures around the folding transition temperature $T_f$, the folding rate, $k_f = 1/MFPT$, increases first with the increase of the frustration $b$, and reaches a maximum at $b$=1.2. Then the folding rate decreases rapidly with the increase of the frustration. This demonstrates that appropriate energetic frustration could increase the folding rate around the folding temperature, which remind us the role of trifluoroethanol (TFE) in protein folding [84–87]. TFE is a cosolvent which may increase the tendency to form the local contacts (both native and non-native). It has been found that low concentration TFE could increase the folding rate of some small proteins, while high concentration TFE may decrease the folding rate. The mechanism of the role of the TFE in protein folding is still under research. Although the role of the TFE may not be the same as the role of frustration in our study, we believe that our results may provide some helpful clues for interpreting the

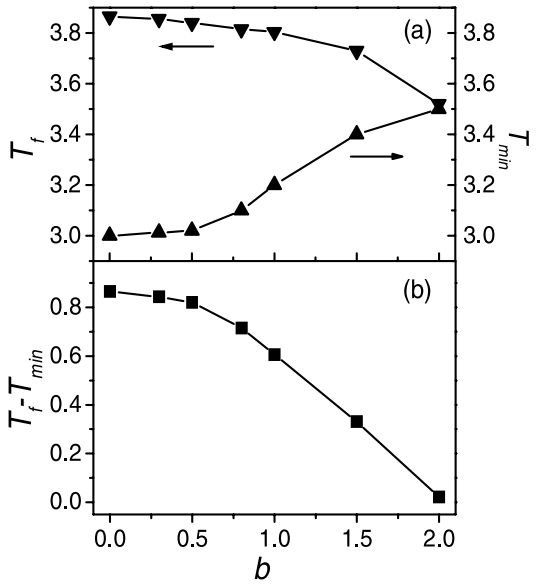


**Fig. 12.** (a) The folding transition temperature $T_f$ and the fast folding temperature $T_{min}$ *versus* the frustration factor $b$, respectively; (b) $T_f - T_{min}$ *versus* the frustration factor $b$.

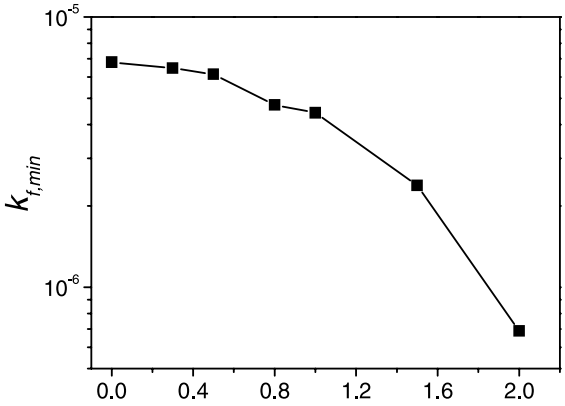


**Fig. 13.** The fastest folding rate $k_{f,min}$ *versus* the frustration factor $b$.

role of the TFE in protein folding. In fact, the view that appropriate frustration may help folding has been theoretically predicted by Plotkin [83]. He used a general idea from the energy landscape theory and perturbed a Gō model towards a realistic protein Hamiltonian by adding non-native interactions, and found that the folding rate was enhanced at a modest level of non-native interactions. Here our simulation results provide a substantial support to such a theoretical prediction. In the following, we give a reasonable interpretation for the role of frustration in protein folding based on the energy landscape theory. The contribution to the free energy comes from two components, namely, entropy and energy. Above and at the folding transition temperature, the effect of entropy dominates the folding behavior of the system. An intermediate level of non-native attraction ($b = 0.5 - 1.5$), *i.e.*, appropriate energy compensation for the entropy loss resulting from the formation of non-native contacts, may lead to the higher compactness. As the result, the conformational search is performed among more compact conformations, leading to a lower entropic cost of contact formation. This lowers the free energy barrier to the native state, and thus make the folding fast. However, too much energy compensation may lead to some trapped non-native conformations, and the energy landscape becomes even more rugged. Different case occurs at low temperature. As the temperature lowers, the effect of energy becomes more and more important relative to the entropy. At an appropriate temperature ($T = 0.85T_f$ at the present study), the entropy loss resulting from the formation of non-native contacts is balanced right by the energy compensation, and the folding rate keeps unchanged at a rather broad range of the frustration $b$ (see also Fig. 11). Below this temperature, the effect of entropy becomes rather minor, thus compactness is high even without the non-native attraction, and the non-native attraction leads to formation of the low-energy traps, that slow down folding. Thus at low temperature, as shown in Figure 11, the folding rate decreases monotonically with the increase of the frustration.

Next we study how the energetic frustration affects the stability and other characteristics of model proteins. In Figure 12 we plot the folding transition temperature $T_f$ and the fastest folding temperature $T_{min}$ *versus* $b$. Here $T_f$ can be viewed as a criterion of stability. In general, high $T_f$ means high stability. From Figure 12a we can see that $T_f$ decreases monotonically with the increase of the value of the frustration $b$, indicating that the stability of model proteins becomes worse with the introduction of the frustration. Moreover, the temperature $T_{min}$, at which folding is fastest, increases with the increase of the energetic frustration $b$ (see Fig. 12a), thus make the difference between $T_f$ and $T_{min}$ become small (see Fig. 12b), indicating at low temperature the foldability becomes worse with the increase of the frustration. From Figure 13, we see clearly when $b \leq 0.5$, the folding behavior is basically kept unchanged as that in the absence of the frustration. However, when $b > 0.5$, the folding behavior becomes bad, and as the value of $b$ approaches to 2.0, the folding rate is quite low. In Figure 13, we see that the fastest folding rate decrease with the increase of the energetic frustration, which is consistent with the above results.

## 5 Comparison with the Gō+ model

In the previous paper [57], we proposed a so-called Gō+ model in which the non-additivity of the interactions is similar as the present Gō++ model. The difference between the two models is that the energy difference between residues at the surface and in the core is considered in the present Gō++ model, while in the Gō+ model all the native contacts contribute equally to the energy. Now, we compare the correlation property for the two models. For the convenience, the curve of the Gō model is also plotted in the same figure. For a specific residue $i$, the

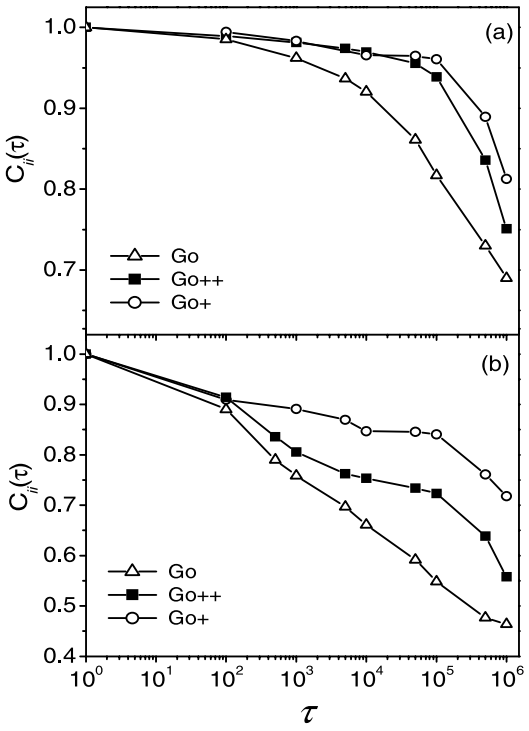


**Fig. 14.** The auto-correlation function $C_{ii}(\tau)$ for (a) a core residue and (b) a surface one *versus* $\tau$ for three models at respective folding transition temperature $T_f$.

auto-correlation function $C_{ii}(\tau)$ can be defined as

$$C_{ii}(\tau) = \sum_{t=0}^{M} n(i,t)n(i,t+\tau) / \sum_{t=0}^{M} n(i,t)^2, \qquad (7)$$

where $M$ is the number of samples, $n(i,t)$ is the number of native contacts of residue $i$ at a specific moment $t$, and $\tau$ is the time lag between two samplings.

Figure 14a shows the auto-correlation function $C_{ii}(\tau)$ of a core residue *versus* the time lag $\tau$ for three models. From Figure 14, We can see that there is a small difference in the values of $C_{ii}(\tau)$ between the Gō++ model and the Gō+ model, while the value of $C_{ii}(\tau)$ for the Gō model decreases faster than that for both the Gō++ and the Gō+ models. There exists a correlation time $\tau_c \approx 100,000$ MCS for both the Gō++ and the Gō+ models, and when the time lag $\tau > \tau_c$ the correlation decrease rapidly. This clearly indicates the formation of the contacts among the core residues during the folding. For a surface residue, the difference in the auto-correlation between the Gō++ model and the Gō+ model becomes significant (see Fig. 14b). However, the correlation is still kept at high values for the Gō++ and the Gō+ model. This implies that the correlation between the number of native contacts for residues at the surface shows strong characteristic of mobility, and the formation or breakdown of the contacts correlate strongly.

Physically, for both the Gō++ and the Gō+ model, the strong correlation between the number of native contacts is due to the introduction of the non-additivity in the interaction. Such a many-body effect makes the formation of contacts promoting further binding reaction of other contacts, thus the neighboring contacts correlate with each other with a high correlation. However, for the Gō model, there is no such strong correlation since the interaction of contacted residues (or the formation of a contact) has no direct relation with the status of contact of other residues. Therefore, the correlation between the contact number of residues really reflects the microscopical features of the folding kinetics.

Now let us discuss the difference between the Gō++ and Gō+ models. From the results presented above, one may find that the Gō+ model behaves better than the Gō++ model in some sense. As we have mentioned in the introduction, the interaction (or the contact) degree described as the ratio of $n_i/n_i^N$ in the Gō+ model has been overestimated since obviously the degrees of all residues are equal in the native state. This is a somewhat artifical assumption. It is well known that the contact number for a residue in the core of a protein is larger than that of the one at the surface. Thus, the overestimation on the interaction of the residues at the surface certainly results in a strong correlation of the folding. This is obviously not the real case of natural proteins. While for the Gō++ model the results are more close to the real situation, and the difference between the core and the surface residues is emphasized. Although its folding behavior is not so good as that of the Gō+ model, it still exhibits good foldability and proteinlike two-state behavior. Here a critical question that should be considered in modeling protein folding is what characteristics of proteins are subject to evolutionary pressures. Some researchers suggested that nature selected proteins stable against mutations [88,89]. Other researchers argued that evolution selected for rapid folding [90–92]. While Baker's group found that the sequences of small proteins are not extensively optimized for fast folding by natural selection [93]. Different effects of natural selection have also been discussed [94–97]. Actually, the biological requirements for successful folding of most proteins are more complex than one or a few criteria. Cooperativity, stability and foldability might be several important criteria to judge whether a model is more realistic, but not the whole of requirements. Our present model incorporated to some extent the surface/core difference and the environment-dependent interactions, which reproduces the known thermodynamics and kinetics of protein folding and exhibits the two-state behavior comparable with that of real proteins. Therefore, we believe that our modified model might be an interesting attempt to model the protein folding.

## 6 Summary and outlook

In this work, we made some reasonable modification on the Gō model based on the contact number of each residue of the model chain. The cooperative interactions between residues and the surface/core difference are incorporated in the present Gō++ model. A detailed examination on the thermodynamics and the kinetics of folding has been

made for both the Gō++ and the Gō models. Our results suggested that the Gō++ model, although simple, exhibits excellent two-state behavior, and has improved cooperativity and foldability compared to the original Gō model. The study on the role of non-native interactions shows that appropriate consideration of the non-native interactions could make the model protein fold faster near the folding transition temperature. In summary, we emphasize that the hydrophobic interactions are non-additive, and may vary with the local environment during the folding process. Our results suggest that the many-body interactions may be a major source of two-state cooperativity, although we do not rule out other possibility which might make contribution to the two-state cooperativity. From a viewpoint of physics, proteins are a rather complex system, and the folding cooperativity may originate from an integration of many aspects, *e.g.*, sidechain packing, hydrogen bonding strength, electrostatic force, and so on.

What kinds of potential function is more close to the realistic potential for protein folding? This is still an open question. Much effort has been done to extract interaction potentials from the pairing frequencies of 20 kinds of amino acids in databases of various protein structures [13,14,17,18]. However, the accuracy of these knowledge-based potentials is not known [98,99]. Even though one assumes that these potentials could really represent the energetics of the native proteins, they provide no more informations about how the interactions vary during the folding. Previous point of view that assumed interactions are invariable during folding is proved to be inappropriate now. Therefore, understanding how the interactions vary during folding is an essential step in understanding the mechanisms of protein folding. It seems that the Gō model with some modification may be a good starting point to interpret the folding of proteins, especially for the states not far away from the native state. Our present work provides one trying in this direction, and we expect further studies could provide deeper insight into this question. Finally, we note that the model studied in this work is easily extended to the off-lattice case.

This work was supported by the NNSF of China (Grant No.10204013, No.90103031, No.10074030, and No.10021001), and the Nonlinear Project (973) of the NSM.

## References

1. S.L. Radford, Trends Biochem. Sci. **25**, 611 (2000)
2. W.A. Eaton *et al.*, Annu. Rev. Biophys. Biomol. Struct. **29**, 327 (2000)
3. J.N. Onuchic, Z. Luthey-Schulten, P.G. Wolynes, Ann. Rev. Phys. Chem. **48**, 545 (1997)
4. C.M. Dobson, M. Karplus, Curr. Opin. Struct. Biol. **9**, 92 (1999)
5. V.S. Pande, A.Y. Grosberg, T. Tanaka, Rev. Mod. Phys. **72**, 259 (2000)
6. J. Wang, W. Wang, Nat. Struct. Biol. **6**, 1033 (1999)
7. H.S. Chan, Nat. Struct. Biol. **6**, 994 (1999)
8. C.M. Dobson, A. Sali, M. Karplus, Angew. Chem. Int. Ed. **37**, 868 (1998)
9. C.B. Anfinsen, Science **181**, 223 (1973)
10. C. Levinthal, J. Chim. Phys. **65**, 44 (1968)
11. K.A. Dill, H.S. Chan, Nat. Struct. Biol. **4**, 10 (1997)
12. P.G. Wolynes, J.N. Onuchic, D. Thirumalai, Science **267**, 1619(1995)
13. M.-H. Hao, H.A. Scheraga, Curr. Opin. Struct. Biol. **9**, 184 (1999)
14. L.A. Mirny, E.I. Shakhnovich, J. Mol. Biol. **264**, 1164 (1996)
15. N. Gō, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983)
16. H.S. Chan, K.A. Dill, Proteins **30**, 2 (1998)
17. S. Miyazawa, R.L. Jernigan, Macromolecules **18**, 534 (1985); J. Mol. Biol. **256**, 623 (1996)
18. A. Kolinski, A. Godzik, J. Skolnick, J. Chem. Phys. **98**, 7420 (1993)
19. O.V. Galzitskaya, A.V. Finkelstein, Proc. Natl. Acad. Sci. USA **96**, 11299 (1999)
20. E. Alm, D. Baker, Proc. Natl. Acad. Sci. USA **96**, 11305 (1999)
21. V. Munoz, W.A. Eaton, Proc. Natl. Acad. Sci. USA **96**, 11311 (1999)
22. K.W. Plaxco, K.T. Simons, D. Baker, J. Mol. Biol. **277**, 985 (1998)
23. D. Baker, Nature (London) **405**, 39 (2000)
24. F. Chiti *et al.*, Nat. Struct. Biol. **6**, 1005 (1999)
25. J.C. Martinez, L. Serrano, Nat. Struct. Biol. **6**, 1010 (1999)
26. D.S. Riddle *et al.*, Nat. Struct. Biol. **6**, 1016 (1999)
27. K.W. Plaxco, K.T. Simons, I. Ruczinski, D. Baker, Biochemistry **39**, 11177 (2000)
28. L. Li, L.A. Mirny, E.I. Shakhnovich, Nat. Struct. Biol. **7**, 336 (2000)
29. A.M. Gutin, V.I. Abkevich, E.I. Shakhnovich, Fold. Des. **3**, 183 (1998)
30. M. Oliveberg *et al.*, J. Mol. Biol. **277**, 933 (1998)
31. D.E. Otzen *et al.*, Biochemistry **38**, 6499 (1999)
32. M. Oliveberg, Acc. Chem. Res. **31**, 765 (1998)
33. H.S. Chan, Proteins **40**, 543 (2000)
34. H. Kaya, H.S. Chan, Proteins **40**, 637 (2000)
35. H. Kaya, H.S. Chan, Phys. Rev. Lett. **85**, 4823 (2000)
36. A.R. Fersht *et al.*, Proc. Natl. Acad. Sci. USA **91**, 10426 (1994)
37. R.L. Baldwin, Nature (London) **369**, 183 (1994)
38. D.E. Otzen *et al.*, Proc. Natl. Acad. Sci. USA **91**, 10422 (1994)
39. K.A. Dill, J. Biol. Chem. **272**, 701 (1997)
40. J.R. Banavar, A. Maritan, Proteins **42**, 433 (2001)
41. A.R. Fersht, *Structure and mechanism in protein science* (Freeman, New York, 1999)
42. A. Hansen, M.H. Jansen, K. Sneppen, G. Zocchi, Eur. Phys. J. B **6**, 157 (1998)
43. C.J. Camacho, D. Thirumalai, Proc. Natl. Acad. Sci. USA **90**, 6369 (1993)
44. J.M. Sorenson, T. Head-Gordon, Fold. Des. **3**, 523 (1998)
45. A. Fernandez, J. Chem. Phys. **115**, 7293 (2001)
46. S. Takada, Z. Luthey-Schulten, P.G. Wolynes, J. Chem. Phys. **110**, 11616 (1999)
47. D.M. Huang, D. Chandler, Proc. Natl. Acad. Sci. USA **97**, 8324 (2000)
48. A. Van Der Vaart, B.D. Bursulaya, C.L. Brooks, K.M. Merz, J. Phys. Chem. B **104**, 9554 (2000)

49. G. Hummer, S. Garde, A.E. Garcia, L.R. Pratt, Chem. Phys. **258**, 349 (2000)
50. A. Kolinski, W. Galazka, J. Skolnick, Proteins **25**, 286 (1996)
51. W.P. Eastwood, P.G. Wolynes, J. Chem. Phys. **114**, 4702 (2001)
52. S.S. Plotkin, J. Wang, P.G. Wolynes, J. Chem. Phys. **106**, 2932 (1997)
53. A. Horovitz, A.R. Fersht, J. Mol. Biol. **214**, 613 (1990)
54. K.P. Murphy, V. Bhakuni, D. Xie, E. Freire, J. Mol. Biol. **227**, 293 (1992)
55. K.A. Dill, K.M. Fiebig, H.S. Chan, Proc. Natl. Acad. Sci. USA **90**, 1942 (1993)
56. D.K. Klimov, D. Thirumalai, Fold. Des. **3**, 127 (1998)
57. K. Fan, J. Wang, W. Wang, Phys. Rev. E **64**, 041907 (2001)
58. C. Czaplewski *et al.*, Protein Sci. **9**, 1235 (2000)
59. D. Eisenberg, A.D. McLachlan, Nature (London) **319**, 199 (1986)
60. N.D. Socci, J.N. Onuchic, J. Chem. Phys. **101**, 1519 (1994)
61. N.D. Socci, J.N. Onuchic, J. Chem. Phys. **103**, 4732 (1995)
62. P.H. Verdier, W.H. Stockmayer, J. Chem. Phys. **36**, 227 (1962)
63. H.J. Hilhorst, J.M. Deutch, J. Chem. Phys. **63**, 5153 (1975)
64. M.T. Gurler, C.C. Crabb, D.M. Dahlin, J. Kovac, Macromolecules **16**, 398 (1983)
65. N. Metropolis *et al.*, J. Chem. Phys. **21**, 1087 (1953)
66. A.M. Ferrenberg, R.H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988)
67. A.M. Ferrenberg, R.H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989)
68. J. Lee, Phys. Rev. Lett. **71**, 211 (1993)
69. M.-H. Hao, H.A. Scheraga, J. Phys. Chem. **98**, 4940 (1994)
70. M.-H. Hao, H.A. Scheraga, J. Phys. Chem. **98**, 9882 (1994)
71. D.K. Klimov, D. Thirumalai, J. Mol. Biol. **282**, 471 (1998)
72. P.L. Privalov, Adv. Protein Chem. **33**, 167 (1979)
73. A. Sali, E.I. Shakhnovich, M. Karplus, Nature (London) **369**, 248 (1994)
74. V.I. Abkevich, A.V. Gutin, E.I. Shakhnovich, J. Mol. Biol. **252**, 460 (1995)
75. D.K. Klimov, D. Thirumalai, Proteins **26**, 411 (1996)
76. J. Wang, K. Fan, W. Wang, Phys. Rev. E **65**, 041925 (2002)
77. S.L. Kazmirski, V. Daggett, J. Mol. Biol. **284**, 793 (1998)
78. D.M. Rothwarf, H.A. Scheraga, Biochemistry **35**, 13797 (1996)
79. D. Hamada, Y. Goto, J. Mol. Biol. **269**, 479 (1997)
80. D. Hamada, S. Segawa, Y. Goto, Nat. Struct. Biol. **3**, 868 (1996)
81. F.J. Blanco, L. Serrano, J.D. Forman-Kay, J. Mol. Biol. **284**, 1153 (1998)
82. J. Prieto *et al.*, J. Mol. Biol. **268**, 760 (1997)
83. S.S. Plotkin, Proteins **45**, 337 (2001)
84. D. Hamada *et al.*, Nat. Struct. Biol. **7**, 58 (2000)
85. E.R.G. Main, S.E. Jackson, Nat. Struct. Biol. **6**, 831 (1999)
86. R.M. Lonescu, C.R. Matthews, Nat. Struct. Biol. **6**, 304 (1999)
87. F. Chiti *et al.*, Nat. Struct. Biol. **6**, 380 (1999)
88. M.R. Ejtehadi, *et al.*, Phys. Rev. E **57**, 3298 (1998)
89. H. Li, R. Hellings, C. Tang, N.S. Wingreen, Science **273**, 666 (1996)
90. A.M. Gutin, V.I. Abkevich, E.I. Shakhnovich, Proc. Natl. Acad. Sci. USA **92**, 1282 (1995)
91. L.A. Mirny, V.I. Abkevich, E.I. Shakhnovich, Proc. Natl. Acad. Sci. USA **95**, 4976 (1998)
92. A.R. Ortiz, J. Skolnick, Biophys. J **79**, 1787 (2000)
93. D.E. Kim, H.D. Gu, D. Baker, Proc. Natl. Acad. Sci. USA **95**, 4982 (1998)
94. D.W. Bolen, I.V. Baskakov, J. Mol. Biol. **310**, 955 (2001)
95. S. Saito, M. Sasai, T. Yomo, Proc. Natl. Acad. Sci. USA **94**, 11324 (1997)
96. V.I. Abkevich, A.M. Gutin, E.I. Shakhnovich, Proc. Natl. Acad. Sci. USA **93**, 839 (1996)
97. C. Clementi, P.A. Jennings, J.N. Onuchic, Proc. Natl. Acad. Sci. USA **97**, 5871 (2000)
98. P.D. Thomas, K.A. Dill, J. Mol. Biol. **257**, 457 (1996)
99. M. Rooman, D. Gilis, Eur. J. Biochem. **254**, 135 (1998)