PHYSICAL REVIEW E, VOLUME 64, 041907

Modeling two-state cooperativity in protein folding

Ke Fan, Jun Wang, and Wei Wang*
National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing 210093, China

(Received 19 February 2001; revised manuscript received 4 June 2001; published 21 September 2001)

A protein model with the pairwise interaction energies varying as the local environment changes, i.e., including some kind of collective effect between the contacts, is proposed. Lattice Monte Carlo simulations on the thermodynamical characteristics and free energy profile show a well-defined two-state behavior and cooperativity of folding for such a model. As a comparison, related simulations for the usual Gō model, where the interaction energies are independent of the local conformation, are also made. Our results indicate that the evolution of interactions during the folding process plays an important role in the two-state cooperativity in protein folding.

DOI: 10.1103/PhysRevE.64.041907 PACS number(s): 87.15.Aa, 87.15.Cc, 87.15.He

Due to the development of experimental methods and theoretical models, many achievements in the field of protein folding have been made recently [1]. A protein can fold itself to its uniquely well-defined native structure in a biologically short time, regardless of the huge number of possible conformations, showing a highly cooperatively kinetic behavior. It is now clear that the cooperativity of folding may result from backbone hydrogen bonding, sidechain packing, and hydrophobic interactions; among these the hydrophobic interactions are believed to be the dominant driving force for folding [2]. For many small single-domain proteins or lattice proteinlike models, there is a two-state behavior between the unfolded states and the folded native one [3,4]. Recently, Chan and Kaya [5] indicated that according to the calorimetric criterion, which is widely used in experiments as a condition for two-state folding, popular lattice models, e.g., the two-letter HP [6] and 20-letter [7] models, are far from two-state models. This may be due to some flawed assumptions in the potential functions used in these models. Lattice models usually use statistical potential functions extracted from the pairing frequencies of 20 kinds of amino acid in databases of protein structures [8]. Although these knowledge-based potentials may be a good approximation to the relative strength of interactions between the residues in the native state, they provide no information about how the interactions evolve during folding. For computational convenience, a common assumption in lattice models is that the interactions are additive, and they are the same during folding as in the native state. This means that the interaction energies are conformation independent. Clearly this is not relevant to the experimental situation [9]. In fact, as Dill pointed out [10], the thermodynamic additivity principle which is widely used in chemistry may be unsuitable in biochemistry. Some recent experiments also indicated that the transition state is an expanded version of the native state, where the majority of interactions are partially formed [11], and their strengths are different from those in the native state (with $\Phi<1$). That is, these interactions depend on the conformation [9], especially the local structures around the contacts, as emphasized recently in Ref. [12]. Previously, the nonadditivity was built into a lattice model for packing effects [13]; and the hydrophobic force depending on the local density of peptide atoms was also taken into account in an off-lattice model [14]. Studies on these models show that the introduction of the non additivity is significiant, but the two-state cooperativity of these models was not checked, and the effects of the nonadditivity on the thermodynamics and kinetics of folding need to be further studied.

In this paper, we develop a refined Gō model in which the pairwise interaction energies vary as the local environment changes, i.e., some kinds of collective effects between contacts are introduced. Our purpose here is to study the two-state cooperativity of protein folding and its physical origin with such a model. Our results give a general picture of how the conformation-dependent interactions affect the folding kinetics, which is consistent with the phenomenological explanation based on experimental results.

We model a polypeptide chain as a self-avoiding chain on a cubic lattice. A contact is formed if two residues are space adjacent but not sequence adjacent. If two residues form a contact that is the same as in the native state, we call this contact a native contact; otherwise it is a non-native contact. Following the Gō model [15], only native contacts are considered to contribute to the total energy. In contrast to the Gō model, we assume that the interaction energies between residues are conformation dependent and vary with changes of the local environment. To achieve this, we introduce a parameter S to describe the degree to which a residue is ordered relative to the native state. For the ith residue in a certain conformation, its degree of order S_i is defined as

$$S_i=z_i/z_i^{nat}, \tag{1}$$

where z_i is the number of native contacts in this conformation, and z_i^{nat} is the number of contacts formed in the native state. Obviously, S_i varies between 0 (the ith residue being fully disordered) and 1 (being fully ordered). Thus, the interaction energy between residues i and j, $B_{ij}=-(S_i+S_j)\varepsilon/2$ is defined, where S_i and S_j are the degrees of order for residues i and j, respectively. ε is the unit of energy and is set to be 1 in this work. The total energy of the confor-

---

*Email address: wangwei@nju.edu.cn

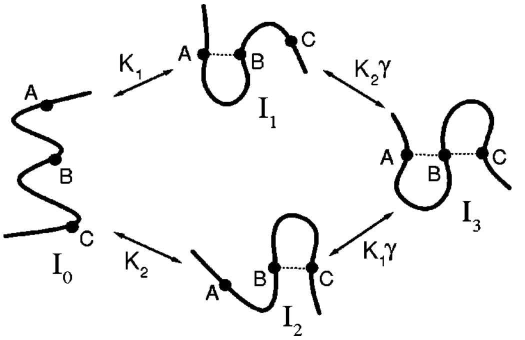

FIG. 1. Schematic illustration of collective effect between two interactions. From a state I_0 with three unstructured residues, the chain can settle in a state I_1 (or I_2) with a contact *A*-*B* (or *B*-*C*) and an equilibrium constant K_1 (or K_2). A state I_3 with two contacts *A*-*B* and *B*-*C* can be reached from state I_1 or I_2, but with different equilibrium constants $K_2\gamma$ or $K_1\gamma$. In state I_3, each interaction is stronger by a factor γ due to the existence of the other contact.

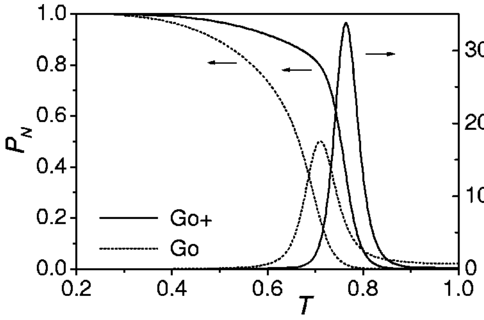

FIG. 2. Population P_N and specific heat C_v varying with the temperature *T* for a 36-mer chain.

mation is then $E=\Sigma_{i<j}\Delta_{ij}B_{ij}$, where Δ is unity when residues *i* and *j* form a native contact, and zero otherwise. Here, 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 (for the Gō model, one always has $B_{ij}=-\varepsilon$). In general, a contact formed between residues *i* and *j* will stabilize, to some extent, other contacts that residue *i* or *j* forms with other residues, and its breakage may destabilize those contacts as well. Therefore, the introduction of the degree of order for a residue into the potential function reflects the cooperativity between the residues. Although the correlation distance is small, only one lattice unit, the many-body effects are obviously included in our model. Figure 1 shows such a collective effect. The interaction energies of contact *A*-*B* (or *B*-*C*) are different when the other contact is present or not present. Clearly, the energy of state I_3 is lower than the sum of the energies of states I_1 and I_2, indicating the interaction nonadditivity. Each contact is stabilized by the other contact due to the collective effect. This is somewhat similar to the solvent accessible surface area model in which the hydrophobic force is decidedly nonadditive [16]. Note that in this paper our model is called the Gō+ model to distinguish it from the Gō model [15].

Now let us present the Monte Carlo simulations of the thermodynamic and kinetic features for both models. The mean first passage time (MFPT), as a common measure of folding rate, is calculated by an average of the first passage times (FPT's) over 1000 runs. Each run begins with a random conformation, and ends when the native state is reached for the first time. The FPT is the number of Monte Carlo steps (MCS's) consumed in a run.

Generally, as the temperature *T* decreases, the population of the native state, P_N, increases from zero to about unity. The degree of sharpness of changes in P_N, similar to the "rapidity" in Ref. [17], is a measure of the cooperativity of the folding reaction. Figure 2 shows the population P_N and the specific heat C_v versus temperature *T* for a 36-mer chain for both models. P_N is defined as $P_N=e^{-E_N/T}/\Sigma_E\Omega(E)e^{-E/T}$, where $\Omega(E)$ is the density of states for energy *E* and E_N is the energy of the native state. $\Omega(E)$ is calculated with the Monte Carlo histogram method [18]. From Fig. 2 we can see that the folding transition for our Gō+ model is much sharper than that of the Gō model, i.e., there is a sharper change in P_N. There is also a single peak in the C_v curve, but it is narrower than that of the Gō model. For our Gō+ model, the maximum of C_v occurs at a temperature that is nearly the midpoint temperature of the transition with $P_N=1/2$, i.e., the difference between these two temperatures is quite small. This is consistent with recent studies on naturally occurring proteins [17,19], implying a good cooperativity of folding in this model. In contrast, this temperature difference is large for the Gō model (see Fig. 2), indicating that the folding of the Gō model is much less cooperative than that of the Gō+ model. Since the sharpness is only a qualitative description for the transition, we further calculate the equilibrium energy distribution at the folding transition temperature T_f. Figure 3 shows such distributions for both models. Clearly our Gō+ model shows a good bimodal behavior, and the denatured-state energy is distributed in a narrow region [see Fig. 3(a)]. This means a two-state folding and there are essentially no intermediate states at equilibrium. In contrast, for the Gō model as shown in Fig. 3(b), there are many intermediate states and the bimodal behavior is not so significant as that in Fig. 3(a). Thus for the Gō model the folding is not of a two-state nature. This is in agreement with Chan and Kaya's argument [5].

In experiments, a well-established criterion for two-state folding is that the van't Hoff enthalpy ΔH_{vH} around the transition midpoint is equal, or very close, to the calorimetric

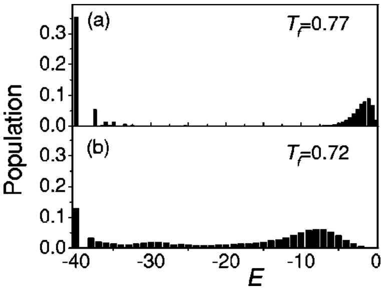

FIG. 3. The energy distribution for the same 36-mer used in Fig. 2, using (a) Gō+ potential and (b) Gō potential at the appropriate folding transition temperature T_f.

TABLE I. The ratios of $\Delta H_{vH}/\Delta H_{cal}$ for the Gō model and for our Gō+ model. Ten sequences are calculated for each chain size.

	$\Delta H_{vH}/\Delta H_{cal}$	
Chain size	Gō model	Gō+ model
27-mer	0.62 ± 0.01	0.91 ± 0.01
36-mer	0.60 ± 0.01	0.91 ± 0.01
48-mer	0.74 ± 0.01	0.95 ± 0.01

enthalpy ΔH_{cal} of the entire transition. In this work, we calculate the ratio $\Delta H_{vH}/\Delta H_{cal}$ as suggested in Ref. [5] [here, the definition of $\Delta H_{vH}/\Delta H_{cal}$ is equal to $(k_2)^2$ in Ref. [5]], and list the results in Table I. From Table I, we can clearly see the difference between the Gō model and our Gō+ model. The Gō model, which is considered as a model with minimal energetic frustrations, does not meet the calorimetric two-state criterion and gives a value of $\Delta H_{vH}/\Delta H_{cal}$ far from 1. Nevertheless, our model satisfies the criterion quite well (for real proteins, the value of $\Delta H_{vH}/\Delta H_{cal}$ is 0.96 ± 0.03 [20]). This, again, implies the two-state folding and the good cooperativity of our Gō+ model.

Physically, the high cooperativity of our model may result from the narrow distribution of denatured states and the high population of the native state at the folding temperature (see also Figs. 2 and 3). In our model, the energy spectrum relating to various conformations is redistributed, comparing with that of the Gō model, due to the collective effect between interactions. As a result, the energies of non-native conformations are moved to higher levels and a larger energy gap is left between the non-native conformations and the native one (for the two models, the energies of the native state are the same). The large energy gap makes the native state particularly stable, which is believed to be a necessary condition for cooperative folding [21]. This may be the physical origin of the two-state cooperativity. It can be further explained from the viewpoint of the free energy profile. For our Gō+ model, as shown in Fig. 4, the free energy profiles have broad activation barriers. The broad activation barriers can account for the large movement of the transition state caused by mutation or temperature changes, and are considered as a common feature of two-state folding [22]. Our numerical results are surprisingly consistent with a phenomenological speculation for the existence of such a free energy profile in Ref. [22]. It should be noted that the broad activation barriers are consistent with the narrow distribution of denatured states.

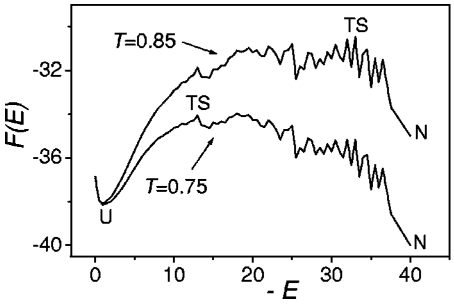

FIG. 4. The free energy profile $F(E)=E-TS(E)$ of our Gō+ model at different temperatures, where entropy $S(E)$ is calculated by the using entropy sampling Monte Carlo method [24]. Here *U*, *N*, and TS denote the unfolded state, native state, and transition state, respectively. Note that the free energy profile at high temperature is shifted overall so that the unfolded states are overlapped.

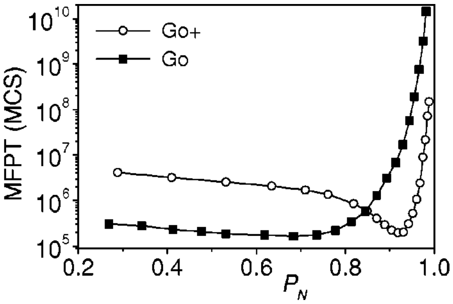

FIG. 5. MFPT versus P_N for a 36-mer chain.

Now let us make a comparison of the foldability based on the plots of the MFPT versus P_N for both models. Note that we use P_N instead of the commonly used temperature T on the horizontal axis on Fig. 5. This is because identical conditions should be taken for the comparison. In lattice simulations, the temperature has arbitrary units and also has no direct relationship with the real temperature. The comparison between two different models at the same temperature may make no sense. Nevertheless, at identical conditions with the same P_N, the differences in the foldability can be well defined. This is similar to other conditions used previously [23]. From Fig. 5, we can see that the MFPT for our Gō+ model shows a slow decrease as P_N increases, it reaches a minimum at $P_N \approx 0.93$, and then it increases. For the Gō model, there is also a minimum but at $P_N \approx 0.71$. It is clear that when the native state is stable (say, $P_N \geqslant 0.9$) our Gō+ model folds significantly faster, i.e., the MFPT is smaller by one or two orders of magnitude than that of the Gō model. Physically, this can be explained as follows. From Eq. (1) we can easily see that the energy gain of forming a contact is usually smaller for our Gō+ model than that for the Gō model. At high temperatures, the entropic contribution is dominant in the free energy barrier, and the loss of entropy is always undercompensated by the energy gain; thus the Gō+ model folds more slowly 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 kinetic traps, and the folding is faster. 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 Gō+ model, a good core the assembly of nonpolar residues, is formed much earlier at low temperatures than that in the Gō model. Detailed kinetic results will be reported elsewhere. We also note that similar results are obtained for different chain sizes.

In conclusion, our Gō+ model, with many-body interactions depending on the local structures included, exhibitsgood two-state folding behavior. Our results suggest that a detailed treatment of the pairwise potential is likely to be more relevant than the consideration of other forces. This provides another view of the origin of two-state cooperativity and also has special significance for protein structure prediction, in which the potential functions are generally similar to those used in protein folding.

We thank H. S. Chan, A. Maritan, and D. Thirumalai for useful suggestions. This work was supported by the Foundation of NNSF (Grant Nos. 19625409 and 10074030).

[1] K.A. Dill and H.S. Chan, Nat. Struct. Biol. **4**, 10 (1997); D. Baker, Nature (London) **405**, 39 (2000); W.A. Eaton *et al.*, Annu. Rev. Biophys. Biomol. Struct. **29**, 327 (2000); J.N. Onuchic, L.-S. Zaida, and P.G. Wolynes, Annu. Rev. Phys. Chem. **48**, 525 (1997); C.M. Dobson and M. Karplus, Curr. Opin. Struct. Biol. **9**, 92 (1999); J. Wang and W. Wang, Nat. Struct. Biol. **6**, 1033 (1999).

[2] J.M. Sorenson and T. Head-Gordon, Folding Des. **3**, 523 (1998).

[3] S.E. Jackson, Folding Des. **3**, R81 (1998).

[4] H.S. Chan, S. Bromberg, and K.A. Dill, Philos. Trans. R. Soc. London, Ser. B **348**, 61 (1995); D.K. Klimov and D. Thirumalai, Folding Des. **3**, 127 (1998).

[5] H.S. Chan, Proteins **40**, 543 (2000); H. Kaya and H.S. Chan, *ibid.* **40**, 637 (2000); Phys. Rev. Lett. **85**, 4823 (2000).

[6] H.S. Chan and K.A. Dill, Proteins **30**, 2 (1998).

[7] A.M. Gutin, V.I. Abkevich, and E.I. Shakhnovich, Folding Des. **3**, 183 (1998).

[8] S. Miyazawa and R.L. Jernigan, Macromolecules **18**, 534 (1985); A. Koliski, A. Godzik, and J. Skolnick, J. Chem. Phys. **98**, 7420 (1993).

[9] A.R. Fersht *et al.*, Proc. Natl. Acad. Sci. U.S.A. **91**, 10 426 (1994); R.L. Baldwin, Nature (London) **369**, 183 (1994).

[10] K.A. Dill, J. Biol. Chem. **272**, 701 (1997).

[11] D.E. Otzen *et al.*, Proc. Natl. Acad. Sci. U.S.A. **91**, 10 422 (1994); F. Chiti *et al.*, Nat. Struct. Biol. **6**, 1005 (1999); J.C. Martinez and L. Serrano, *ibid.* **6**, 1010 (1999); D.S. Riddle *et al.*, *ibid.* **6**, 1016 (1999).

[12] J.R. Banavar and A. Maritan, Proteins **42**, 433 (2001).

[13] C.J. Camacho and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **90**, 6369 (1993).

[14] S. Takada, Z. Luthey-Schulten, and P.G. Wolynes, J. Chem. Phys. **110**, 11616 (1999).

[15] N. Gō, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).

[16] D. Eisenberg and A.D. McLachlan, Nature (London) **319**, 199 (1986).

[17] F. Cecconi *et al.*, Proteins **43**, 365 (2001).

[18] A.M. Ferrenberg and R.H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989); N.D. Socci and J.N. Onuchic, J. Chem. Phys. **103**, 4732 (1995).

[19] C. Micheletti *et al.*, Phys. Rev. Lett. **82**, 3372 (1999).

[20] P.L. Privalov, Adv. Protein Chem. **33**, 167 (1979).

[21] A. Sali, E.I. Shakhnovich, and M. Karplus, Nature (London) **369**, 248 (1994).

[22] M. Oliveberg *et al.*, J. Mol. Biol. **277**, 933 (1998); D.E. Otzen *et al.*, Biochemistry **38**, 6499 (1999); M. Oliveberg, Acc. Chem. Res. **31**, 765 (1998).

[23] V.I. Abkevich, A.V. Gutin, and E.I. Shakhnovich, J. Mol. Biol. **252**, 460 (1995); D.K. Klimov and D. Thirumalai, Proteins **26**, 411 (1996).

[24] J. Lee, Phys. Rev. Lett. **71**, 211 (1993); M.H. Hao and H.A. Scheraga, J. Phys. Chem. **98**, 4940 (1994).