Folding of proteins with an all-atom Gō-model

L. Wu, J. Zhang, M. Qin, F. Liu, and W. Wang[a)]
National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing 210093, China

(Received 22 January 2008; accepted 19 May 2008; published online 19 June 2008)

The Gō-like potential at a residual level has been successfully applied to the folding of proteins in many previous works. However, taking into consideration more detailed structural information in the atomic level, the definition of contacts used in these traditional Gō-models may not be suitable for all-atom simulations. Here, in this work, we develop a rational definition of contacts considering the screening effect in the crowded intramolecular environment. In such a scheme, a large amount of screened atom pairs are excluded and the number of contacts is decreased compared to the case of the traditional definition. These contacts defined by such a new definition are compatible with the all-atom representation of protein structures. To verify the rationality of the new definition of contacts, the folding of proteins CI2 and SH3 is simulated by all-atom molecular dynamics simulations. A high folding cooperativity and good correlation of the simulated Φ-values with those obtained experimentally, especially for CI2, are found. This suggests that the all-atom Gō-model is improved compared to the traditional Gō-model. Based on the comparison of the Φ-values, the roles of side chains in the folding are discussed, and it is concluded that the side-chain structures are more important for local contacts in determining the transition state structures. Moreover, the relations between side chain and backbone orderings are also discussed. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2943202]

I. INTRODUCTION

Gō-model has been widely used to study the folding of small globular proteins and has shown remarkable successes.[1–7] In the traditional Gō-model, each residue is modeled by a single bead centered at the C_α atom, and the native contacts are assumed to have the same contact energies. Hereafter, we denote the traditional Gō-model by C_α Gō-model. The successes of such a simplistic model can be understood by the notion that the protein folding is mainly determined by the native topology and that non-native interactions only play minor roles for most proteins. This is believed to be due to the fact that the energetic frustrations of most proteins have been minimized by evolution, i.e., "the principle of minimal frustration."[8] However, from a physical point of view, the C_α Gō-model has several obvious limitations. First, the one-bead representation in the C_α Gō-model ignores the side-chain structures, thus reducing the conformational entropy.[9,10] Second, the specific packing patterns of side chains that determine the secondary structure propensities of residues cannot be delineated by the C_α Gō-model.[11] Third, in the C_α Gō-model, the native contacts are defined when a pair of atoms or residues have their native distances less than a cutoff value. This is a simplification since different physical interactions have different characteristic distances. Moreover, it is recently shown that the sequence-dependent energetic effects can also play important roles in the folding processes of some proteins.[12,13] The topology-only C_α Gō-model is unable to correctly predict the folding behaviors of such proteins with heterogeneous distribution of contact energies.

The most accurate simulation method to study protein folding is the all-atom model combined with a modern force field, such as AMBER, CHARMM, OPLS, GROMOS, etc.[14–16] However, it is very computational expensive to carry out such simulations and the lengths of trajectories are usually limited to a time scale of nanoseconds that is too short for describing the protein folding. In order to satisfy the requirement of time scale to study the protein folding problem, there are many efforts in the community toward compromising the precision and computational accessibility. One trend is to reduce the structural resolution, using several beads to represent a residue, and develop parameters based on various physical considerations.[17,18] Another trend prefers to keep native topology-based potentials and seeks to use high structural resolution models. The most representative one is the all-atom Gō-model, which explicitly models all the heavy atoms except hydrogens and uses Gō-like potentials to describe the interactions between atoms.[19–21]

In this work, in order to examine the improvements of including side-chain topologies in Gō-like models, we preserve the all-atom structures of proteins and use Gō-like potential to carry out the dynamic simulations. Since the contact network plays a central role and determines the folding behaviors in Gō-like models, the simplified traditional definition of the native contact is reconsidered. The free energy profiles and reversible folding/unfolding trajectories for CI2 and SH3 are obtained using the all-atom Gō-model with our new definition of contact. Our results show that proteins have improved folding cooperativity in the all-atom Gō-

a)Electronic mail: wangwei@nju.edu.cn.

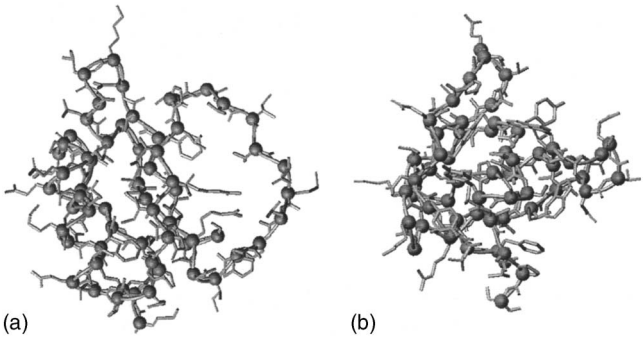

FIG. 1. The native structures of CI2 (a) and SH3 (b). The C_α atoms are represented by balls and side chains are represented by sticks.

model compared to the C_α Gō-model. Moreover, Φ-values are also calculated using both C_α and all-atom Gō-models and compared with those from experiments. It is found that the Φ-values obtained from the all-atom Gō-model for residues in the α-helix in CI2 show improved agreement with experimental values, and that the two models have similar performance in predicting Φ-values for residues in β-structures in CI2 and SH3. Based on the improved performance of the all-atom Gō-model, the roles of side-chain topology in folding processes are discussed, and there are indications that the side-chain structures play more important roles in forming local contacts. Furthermore, the relations between side-chain and backbone orderings are also discussed. The side-chain ordering is found to be slower than the backbone folding, especially at lower temperatures. Even so, a coupling of side-chain ordering and backbone folding is still required for proteins to fold into their native states.

II. MODEL

The model proteins used here are chymotrypsin inhibitor 2 (CI2, PDB: 1coa) and src SH3 domain (SH3, PDB: 1nlo), whose folding behaviors are extensively studied by both experiments and simulations.[22–24] The protein CI2 has an α/β structure and the src SH3 domain has an all-β native structure. The all-atom representations of CI2 and SH3 are shown in Fig. 1. All the C_α atoms are represented by balls and the side chains are shown in sticks. Both CI2 and SH3 have been well established to display two-state folding behaviors, but the patterns of the transition states are different. The SH3 domain folds with a polarized transition state, whereas the protein CI2 folds with a diffused transition state.

In the C_α Gō-model, native contacts are traditionally defined by a cutoff distance. With such a definition, the total number of native contacts will increase unlimitedly with the increase in the cutoff value. However, physically, the interactions between two residues should fade out and become zero at sufficiently long distance due to the screening effect from other atoms between them. Therefore, it is clear that in the traditional definition of contacts the screening effect is not considered, and such a definition is not satisfactory from a physical point of view. This problem looms large in the all-atom Gō-models due to its crowded intramolecular environment. In previous all-atom Gō-model, very low cutoff values (4 Å) are selected.[21] Although such low cutoff distance will effectively exclude screened atom pairs, it may not be sufficient for characterizing all the native interactions between atoms. The detailed physical interatomic contacts in a protein can be provided by the contacts of structural units (CSU) program, which analyzes the contact interactions based on the method of surface complementarity incorporating information about the shape and chemical nature of the atoms.[25] In this method, the solvent molecule is used as a probe to detect the contact area between two atoms, and the maximum distance between two contacting atoms is determined by the van der Waals radii of them and the solvent molecule. In this way, many atom pairs with very long distance (>6 Å) are defined as native atomic contacts. However, in the all-atom Gō-model, a realistic all-atom force field is used for the bonded interactions, which have conformational frustrations in the native conformation. Therefore, the native distance of atoms cannot be taken as the equilibrium distance of the Lennard–Jones potential for nonbond contacts. In this work, we adopt a similar method to the work of Clementi *et al.*,[21] and simply use uniform equilibrium distance for atomic contacts specified for their (hydrophobic-polar property. Consequently, the CSU method is not suitable in this model, because that the long native contacts defined by the CSU method should be significantly biased from native conformation due to the much shorter equilibrium distances. To develop a new definition of contacts that is compatible with the all-atom Gō-model, a general method of considering screening effect based on the traditional definition of native contacts is needed, and the native contacts calculated by the CSU program can be used as the reference to refine the parameters of our new definition of contact. Considering the simplistic nature of Gō-like potentials, we believe that the more realistic all-atom topology contributes more to the improvement of the minimalist all-atom Gō-model than the refinement of definition of atomic contacts according to the chemicophysical properties. Furthermore, the contacts formed between distant atoms should have weak interactions in the native state and have little effect on the folding behaviors.

To incorporate the screening effect into the model, the angle formed by any two contacts involving a common atom is checked, and a cutoff angle is set to determine whether one of them is screened. As shown in Fig. 2(a), according to the traditional definition of contacts, atoms i and j form native contacts with a common atom c, respectively. Here, if the angle α between the two contacts is less than the cutoff angle, the contact with a longer distance between atoms c and j should be excluded, which is referred to as "screened contact" hereafter. Then the contact list defined by the new method is compared with the CSU results, and the distance cutoff and angle cutoff are accordingly set to 4.7 Å and 35°, respectively. By excluding the screened contacts, a plot [Fig. 2(b)] reflecting the saturation of the total number of native contacts is obtained. Compared with the definition of contacts using 4.0 Å as the cutoff value in other all-atom Gō-like models, our definition increases the percentage of true-positive residue contacts (contacts identified by both our criteria and the CSU method) from 73% to 89% for CI2, and

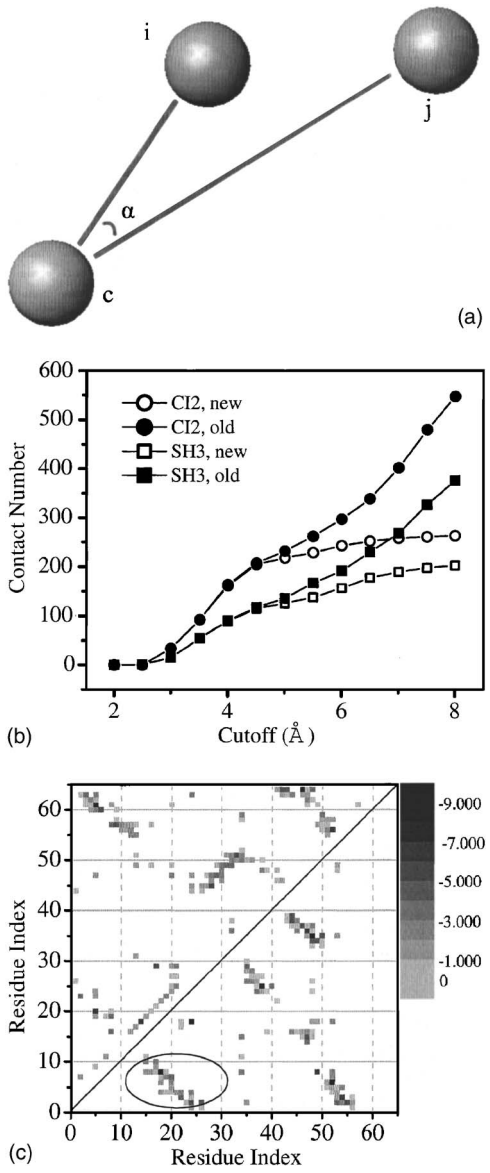

FIG. 2. (a) The diagram illustrates the modification of the definition of native contact. (b) The number of contacts as a function of the cutoff values. The solid symbols correspond to that calculated by cutoff-only definition and the hollow symbols correspond to our new definition of contacts. (c) The contact maps of CI2 (upper corner) and SH3 (lower corner). The gray scale illustrates the integrated interaction energy of atomic contacts between corresponding residues, and the darker gray scales represent stronger interaction energies. The native contacts within the RT loop of SH3 are enclosed by an ellipse.

45% of atom pairs within 4.7 Å cutoff are screened. The CSU contact list for the SH3 domain contains much more hydrophobic interactions over long distances that cannot be identified in our method, so the true-positive percentage is relatively low for SH3, which increased from 69% to 84%, and 49% of atom pairs are screened for the SH3 domain. Thus the screened atoms are effectively excluded by using the new definition of contact. This makes proper cutoff values available in the all-atom Gō-model, and the residual contacts integrated from the atomic contacts identified by our method agree well with that from the CSU method. It is noticeable that in the all-atom Gō-model, the native contacts are between heavy atoms, instead of residues represented by virtual beads located at the C_α atoms.

We assign a 12-6 Lennard–Jones potential to each native contact, which is

$$4\varepsilon_0[(\sigma_i/r)^{12} - (\sigma_i/r)^6], \tag{1}$$

with $i=1,2$, and a repulsive one,

$$4\varepsilon_0[(\sigma_3/r)^{12} + (\sigma_3/r)^6] \tag{2}$$

to each non-native contact. The parameter $\sigma_1=2.75$ Å is used for native contacts between two polar atoms, $\sigma_2=3.25$ Å for the other native contacts, and $\sigma_3=3.25$ Å. The first two values are slightly lower than the average distance between the corresponding kinds of native contacts, in order to balance the frustrations. The native contact map is shown in Fig. 2(c). The squares denote the contacts between residues, and the color is scaled by the total energy of all native atom pairs between corresponding residues. According to Fig. 2(c), the heterogeneous distribution of contact energies is naturally introduced into the all-atom Gō-model since the side-chain atoms are explicitly modeled. Physically, this feature should distinguish the all-atom Gō-model from the C_α one and improve its performance. The simulations are carried out using AMBER8 simulation package with OPLS force field at constant temperatures. The integration step is set to 0.002 ps. The transition temperatures T_f determined by the all-atom Gō-model correspond to 450 and 490 K for CI2 and SH3, respectively. Free energy profiles are calculated using the WHAM method.[26]

III. RESULTS

A. Folding behaviors

The dynamic trajectories of CI2 and SH3 domains at each transition temperature are calculated and shown as the function of order parameter Q, i.e., the fraction of native contacts [Figs. 3(a) and 3(c)]. The repeated folding and unfolding events are clearly shown in the trajectory profiles. Such reversible folding processes accord with the basic characteristics of protein kinetics, and guarantee the validity of the simulations. The bimodal distributions of Q values show clearly the two-state folding behaviors. Sharp transitions connecting the unfolded states ($Q<0.3$) and folded states ($Q>0.6$) indicate high folding cooperativity. The one dimensional free energy profiles for CI2 and SH3 are calculated from both the all-atom Gō-model and the C_α Gō-model, and are shown in Figs. 3(b) and 3(d). The free energies calculated by both models exhibit two basins of attraction separated by a large barrier. Comparing with the C_α model, however, the free energy barrier obtained from the all-atom Gō-model is relatively higher for both proteins, suggesting higher folding cooperativity. Moreover, a slight movement of the free energy barrier away from the unfolded state is observed for both proteins.

In addition, comparing with the C_α Gō-model, a less structured unfolded state is observed for CI2, manifested by the narrowing and shifting of unfolded basin to the low Q value. This is consistent with the experimental observations that CI2 has extremely denatured unfolded states, which look

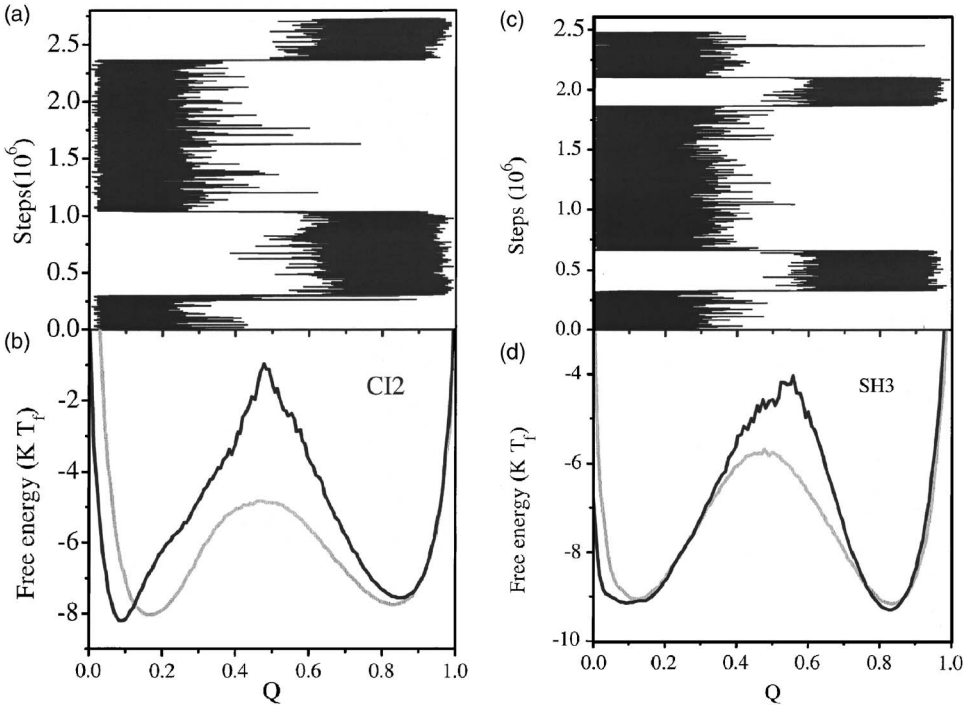

FIG. 3. The folding trajectories for CI2 (a) and SH3 (c) sampled from dynamic simulations with the all-atom Gō-model, and the comparison of free energy profiles between the all-atom Gō-model and the C_α Gō-model for CI2 (b) and SH3 (d).

like random coils.[27] This can be understood as follows. Unlike in the C_α Gō-models, the side chains have been explicitly modeled in the all-atom Gō-model, therefore, residues have to adjust their side-chain conformations to achieve specific packing patterns and form native contacts. This makes the formation of local structures relatively difficult; thus, the all-atom Gō-model produces more denatured unfolded states. The free energy profile for the SH3 domain is shown in Fig. 3(d). The unfolded basin shifts to the low Q value and becomes broader with the introduction of side-chain structures, indicating larger structural heterogeneity in the unfolded states of SH3.

To quantitatively characterize the change in folding cooperativities, the κ_2 is calculated for both the C_α and the all-atom Gō-models. The quantity κ_2 is proposed by Chan and co-workers to assess the folding cooperativity of two-state proteins,[28,29] which is defined as

$$\kappa_2 = 2T_f\sqrt{k_B C_v(T_f)}/\Delta E_{\mathrm{cal}}. \tag{3}$$

A κ_2 value close to 1 indicates high folding cooperativity, whereas a value close to 0 indicates noncooperative folding. Comparing with the C_α model, the calculated κ_2 increases from 0.86 ± 0.06 to 0.95 ± 0.05, and from 0.87 ± 0.06 to 0.97 ± 0.05 for CI2 and SH3, respectively. The increase of κ_2 demonstrates the improved folding cooperativity in the all-atom Gō-model. Such increase of folding cooperativity all-atom Gō-model is also observed in former studies.[21] The increased cooperativity may contribute to the coupling between side-chain packing and the formation of tertiary structure. The packing of side chain is both distance and orientation dependent, and these factors should need support from tertiary structure. When the protein reaches the transition state, the nativelike structure facilitates the packing of side chains and folds to the native state rapidly. Therefore, the conformational entropy introduced by side chains is quickly lost when passing the transition state and thus increases the free energy barrier and moves the transition state away from the unfolded state slightly. The different methods of characterizing the local interactions in the all-atom and C_α Gō-model may be another factor influencing the folding cooperativity. In the C_α Gō-model, the 1-3 and 1-4 residue pairs (two residues that are separated by one and two amino acid residues) are restrained by virtual bond angles and dihedral angles. The interaction energy of these bond angles is one order of magnitude stronger than that of the interactions of nonlocal contacts, and the potential energy of dihedral angles is also stronger than that of the nonbond interactions. In comparison, in the all-atom Gō-model, the local native contacts formed by 1-3 and 1-4 residue pairs are all considered as nonbond interactions. Therefore, compared with local interactions, the nonlocal interactions in the all-atom Gō-model are *stronger* than those in the C_α Gō-model. Since stronger nonlocal interactions usually lead to higher folding cooperativity, the change of balance between the local and nonlocal interactions should improve the folding cooperativity. Such a change in the all-atom Gō-model results from the more realistic representation of protein structures. Therefore, the increased folding cooperativity induced by the change of balance of the local and nonlocal interactions in the all-atom Gō-model is rational, and the interactions of the 1-3 and 1-4 residue pairs are overestimated in C_α Gō-model.

The foldings of both CI2 and SH3 are also simulated using the all-atom Gō-model with traditional definition of native contacts in order to make comparison with our new definition of native contacts. Considering the ignorance of screening effect, the cutoff value should be *smaller* than that in our definition to keep the ratio of nonbond interaction energy to covalent bond interaction energy within a reasonable range. Here, we test a series of cutoff values from 4.0 to 4.7 Å in step of 0.1 Å. It is found that when the cutoff

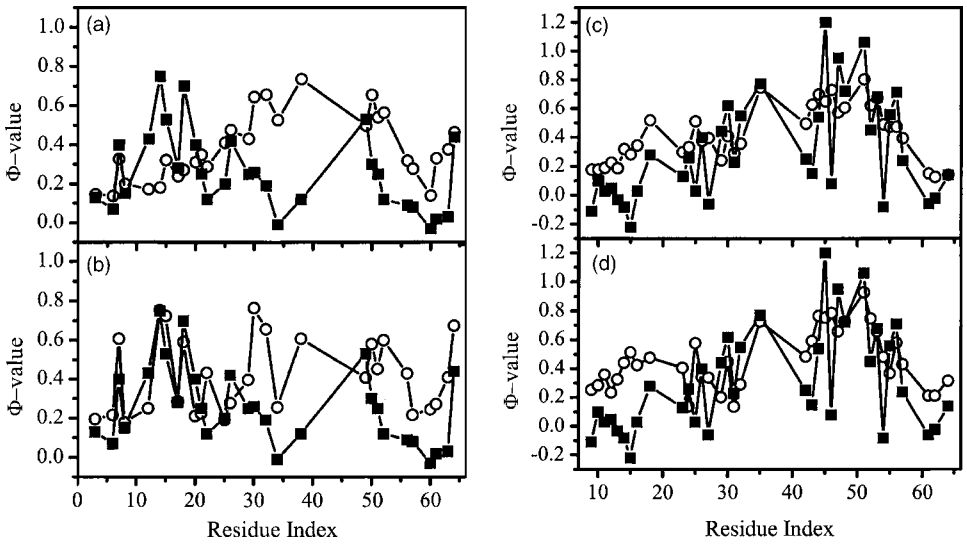

FIG. 4. The comparison of calculated Φ-values (hollow circles) and experimental values (solid squares). (a) Φ-values of CI2 calculated from the C_α Gō-model. (b) Φ-values of CI2 calculated from the all-atom Gō-model. (c) Φ-values of SH3 calculated from the C_α model. (b) Φ-values of SH3 calculated from the all-atom Gō-model.

value is smaller than 4.2 Å, folding of CI2 will be trapped in a partially folded state because of the absence of the interactions between the N-terminal loop and the α-helix. These interactions allow the N-terminal loop docking to the α-helix and decrease the distance between N- and C-termini. This process may be important to the formation of native contacts between N- and C-termini of CI2. When the cutoff value is increased above 4.2 Å, more native contacts between the N-terminal loop and the α-helix are defined, and CI2 folds with a stable folding intermediate in which the N–C terminal interactions are not formed. This is inconsistent with the experimental findings that CI2 is a two-state protein. Similarly, two different stable intermediates are also observed in the folding of SH3 with traditional definition of native contacts. The difference between folding intermediates results from the competition between the RT loop [Fig. 2(c)] and N–C terminal interactions. For smaller cutoff distances, there is a lack of native contacts within the RT loop. Therefore, an intermediate, in which the N–C terminal contacts are not formed, is found. However, the number of native contacts within the RT loop quickly increases with cutoff distance and the RT loop forms much faster due to the small loop entropy. In the traditional definition of the native contacts, the interaction energy between residues varies significantly with the distance between the contacted residues. Therefore, the folding intermediates of CI2 and SH3 may result from energetic traps caused by very inhomogeneous interaction energies. With such patterns of native interaction networks, both CI2 and SH3 exhibit improper folding behaviors that are much different from experimental results. This emphasizes the importance of considering screening effect in the definition of native contacts in the all-atom Gō-model.

B. Φ-value analysis

The Φ-values for CI2 and SH3 are calculated to make quantitative comparison on the ability of predicting the structure of transition states between the two models. Φ-value is originally proposed by Fersht to investigate the structure of transition state ensemble from mutation data.[30] A Φ-value close to 1 means that the structure around the mutated residue has already formed in the transition state, whereas a Φ-value close to 0 means the related structure is denatured in the transition state. Here, the Φ-values for residues are evaluated as the fraction of formed native contacts involving the corresponding residues in the transition state.

For the protein CI2, the experimental Φ-values of the residues within the hydrophobic core, the minicore, and the secondary structures are selected to make comparisons.[23] Abnormal experimental Φ-values, such as those much larger than 1 or with negative values, are excluded. The residues located in the active site (the loop, GLY35-ILE44) are also excluded due to their high exposure to solvents, thus the minor change of stabilities after mutations and the concomitant large experimental errors.[31] The Φ-values of CI2 are shown in Figs. 4(a) and 4(b), calculated from the C_α and the all-atom model, respectively. The agreement of Φ-values between experiments and simulations is improved in the all-atom Gō-model, especially for the N-half of the sequence (from the N-termini to GLU26) where the calculated values overlap the experimental ones perfectly. However, for the C-half of the sequence, both models overestimate the Φ-values and the all-atom Gō-model only improves the prediction slightly. The correlation coefficient of Φ-values of the entire protein obtained from experiments and simulations is improved from −0.06 to 0.46.

Structure analysis shows that the N-half sequence of CI2 corresponds mostly to the α-helix and the C-half forms β-structures in the native states. The different performances of the all-atom Gō-model on these two parts indicate different importance of the side chains in the folding of α-helices and β-strands. Comparing with the β-structures, the packing density of α-helices is larger,[32] thus, the specific packing and the interaction energies between the side chains should contribute greatly to the formation of α-helix. This can be illustrated from the detailed analysis on the improved Φ-values. For example, as shown in Fig. 4, the Φ-values for GLU14 and LYS18 located in the α-helix increase in the all-atom Gō-model. The side chains of these two residues extend parallelly, forming a large number of close atom pairs. This feature can be explicitly characterized by the all-atom Gō-model [see Fig. 2(c)], but in the C_α Gō-model, the interaction

of the close contact between GLU14 and LYS18 is underestimated due to the nonspecific contact energy. Here, the heterogeneity of contact energies only results from the introduction of the side chains, without considering the chemicophysical properties of the side-chain atoms.

For the β-structures, since most native contacts have long loop lengths, the topology-based diffusion process plays a more important role in the formation of the β-sheets. This is supported by the fact that most of the residues with lower Φ-values in the C-half have relatively longer average loop lengths of native contacts. Consequently, the transition state structure of the β-sheets in CI2 is determined by the loop entropy, and therefore the C_α model is able to predict the transition state ensemble of the C-half of CI2 as good as the all-atom Gō-model does. However, for the α-helices, the diffusion process is not so important in forming native structures since it involves only local structure arrangement. This result leads us to the conclusion that the side chains play less important roles in determining the transition state ensemble of structures stabilized mostly by long range interactions such as the β-structures. Moreover, both models overestimate the Φ-values in C-half, suggesting that the formation rate of the β-structures is much accelerated in the two models. Physically, the acceleration is highly possible due to the incorporated information of the turn position in the Gō-like potential, which reduces the torsional energetic frustrations, thus greatly facilitating the initial formation of turns. In fact, the folding rate of β-hairpin is supposed to correlate with the formation of the turn structure.[33]

The Φ-values for the SH3 domain are calculated using both models and are shown in Figs. 4(c) and 4(d). SH3 is an all-β protein and has a polarized transition state, manifested by the high Φ-values in the midpart of sequence and the low Φ-values in the two termini. As expected, according to Fig. 4, both the C_α and the all-atom Gō-model correctly reproduce the Φ-values and the polarized characteristic of the transition state of SH3. This supports the conclusion that the side-chain topology has a little effect on predicting transition states of the β-structures. In contrast with the overestimation of the Φ-values corresponding to the β-structures in CI2, The correlation coefficients of the Φ-values for the SH3 domain are larger than 0.6 for both models. The better results in the calculation of SH3 domain may be due to that since the SH3 domain is an all-β protein, and the formation speed of all the β-strands within it are equally accelerated, leading to correct structures of transition state ensemble in the simulation.

C. Side-chain and backbone ordering

The relation between the side-chain ordering and backbone folding is of particular interest to the study of protein folding. Understanding how the side-chain ordering coupled with backbone folding should provide an insight into the dynamics and thermodynamics of protein folding. Benefited from the atomic resolution of protein structures and easily accessible time scale of folding in the all-atom Gō-model, the kinetic behavior of the side-chain and backbone atoms during folding can be discussed. Here, the fraction of native contacts Q is used as the folding coordinate. We decompose the Q value into two components, i.e., Q_s and Q_m, representing the fraction of atomic contacts between the side-chain atoms and backbone atoms, respectively. The Q_s curves as a function of Q_m are plotted both for CI2 and SH3 in Figs. 5(a) and 5(b). These Q_s curves are obtained from arithmetic average of multiple folding simulations under different temperatures. The black lines correspond to temperatures 440 and 470 K that are near the transition temperatures of CI2 and SH3, respectively, and the gray lines correspond to much lower temperatures (i.e., 350 K for CI2 and 390 K for SH3). It is noticeable that the temperatures do *not* correspond to the actual value because the Gō-type potential function is simplified and *not* very physically realistic. By average, backbone contacts have a higher formation probability than side-chain contacts. Q_s has the same value as Q_m in the unfolded state region. However, Q_s grows slower than Q_m during formation of the tertiary structures, and stabilizes at a lower value than Q_m in the folded state. The Q_s curves for both CI2 and SH3 near transition temperatures have a transition at Q_m values near 0.5, suggesting a nearly simultaneous ordering of the side-chain and backbone atoms. This provides a support to our earlier conclusion that the coupling of side-chain entropy reduction and tertiary structure formation may contribute to the increase of folding cooperativity. At lower temperatures, the transition part of Q_s curves moves toward the folded state and the transition occurs when Q_m values are around 0.7, indicating that the side-chain ordering is much slower than the backbone folding. This is in accordance with other studies and reports.[34,35] In addition, there is a difference in Q_s curves between CI2 and SH3 at lower temperatures. The Q_s values for CI2 separate from Q_m at relatively higher Q_m values compared with SH3. This difference results from the existence of the α-helix in CI2. The side-chain and backbone orderings within the α-helix are highly coupled.

Figures 5(c) and 5(d) show typical folding trajectories for CI2 and SH3 near the transition temperatures. The black lines represent Q_m trajectories and the gray lines represent Q_s trajectories. The side-chain and backbone contacts form almost coincidently near the transition temperatures for both CI2 and SH3. The Q_m values are a little higher than Q_s in a range of Q_m values from 0.2 to 0.4, indicating that the backbone folding slightly precedes the side-chain ordering before the transition state. This is in accordance with the results from average Q_s curves. The foldings of CI2 and SH3 are much slower in lower temperatures. The separations of Q_s and Q_m trajectories under 350 and 390 K for CI2 and SH3 are clearly shown in Figs. 5(e) and 5(f), respectively. There is a sharp transition of Q_m around 750 000 simulation steps for CI2, with only a relatively mild increase of Q_s values. This indicates that the side-chain ordering is far behind the folding transition of backbone structure at lower temperature. When the side-chain contacts are mostly formed after 1 500 000 simulation steps, the Q_m reaches a slightly higher value, indicating a more compact native structure compared with that when side chains are mostly disordered. In the folding trajectory of SH3, the sharp transition occurs after the separation region. Although the backbone contacts have sub-

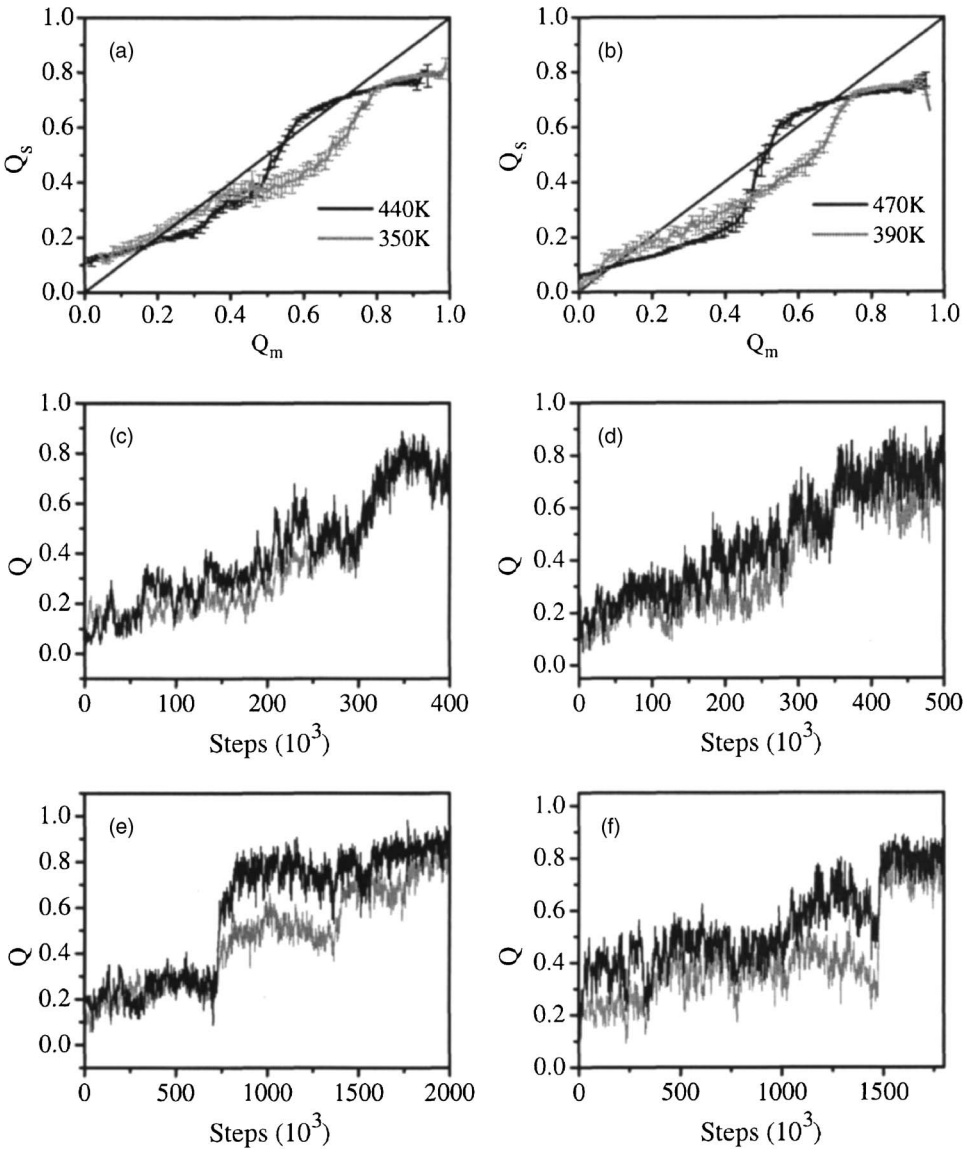

FIG. 5. (a) The average Q_s values plotted against Q_m for CI2. The black line corresponds to dynamic simulations under 440 K, and the gray line corresponds to simulations under 350 K. (b) The average Q_s values plotted against Q_m for SH3. The black line and gray line correspond to dynamic simulations under 470 and 390 K, respectively. Four typical folding trajectories are selected to be shown in (c)–(f), and the black lines represent Q_m trajectories and gray lines represent Q_s trajectories. (c) Folding trajectory of CI2 at 440 K. (d) Folding trajectory of SH3 at 470 K. (e) Folding trajectory of CI2 at 350 K. (d) Folding trajectory of SH3 at 390 K.

stantially higher formation probability before 1 480 000 simulation steps, the Q_s and Q_m trajectories jump to the compact native state simultaneously.

Combining the observations from average Q values and detailed folding trajectories, we obtain a qualitative understanding of relations between the side-chain ordering and backbone folding. First, the backbone folding is *faster* than the side-chain ordering, especially at lower temperatures. The higher temperature results in more rapid thermal motion of the backbone structures that may provide more opportunity for the side-chain structures to adopt their optimal rotamers, and the efficiency of conformational search of the side-chain rotamers may also be enhanced by thermal motion of the side-chain structures. As a result, the side-chain ordering is almost synchronized with the backbone folding near transition temperatures. However, at lower temperatures, the side-chain ordering is *slower*, leading to a separation from the backbone folding. Second, the side-chain ordering is essential for proteins to pass the transition state and fold to the compact native state. From the folding trajectories of SH3 at lower temperature, we observe kinetic coupled backbone folding and side-chain ordering. The backbone needs support from well-organized side-chain structures to achieve the native state. Generally, the all-atom Gō-model is able to *successfully* reproduce qualitatively correct side-chain behaviors.

IV. CONCLUSIONS

In the present work, we have successfully developed a new definition of interatomic contacts for the all-atom Gō-model. By considering the screening effect, all atom pairs, which are screened by other atoms, are not considered as native contacts even if they are within the cutoff distance. Very simple geometrical parameters are used to discern the native atom pairs that are directly interacting with each other without being screened. To verify the rationality of such a modification, we perform simulations using the C_α Gō-model and the all-atom Gō-model with both traditional and our definition of native contacts. The two-state folding behavior of CI2 and SH3 in the all-atom model can be observed *only* when the screening effect is incorporated into the definition of native contacts. With the newly defined inter-

atomic contacts, the all-atom Gō-model shows better performance compared to the C_α Gō-model, demonstrating the rationality of our new definition of contacts.

As in other applications of all-atom Gō-like models, our simulation results again show an *improvement* of the folding cooperativity. The higher folding cooperativity possibly results from the precise representation of the side-chain structures and the more realistic interaction energies. The inclusion of side-chain topologies increases the degree of freedom for each residue, thus resulting in a large increase of conformational entropy in the unfolded states of the all-atom Gō-models. The additional conformational entropy contributed from the side chains rapidly *loses* when the protein passes the transition state, thus increased the free energy barrier. On the other hand, compared with the C_α Gō-model, the interaction energies between 1-3 and 1-4 residue pairs are much smaller. This also results from the high-resolution structure in the all-atom model, and is more consistent with the real molecular system. Since the 1-3 and 1-4 residues form local contacts, the balance between the nonlocal and local interactions is changed in the all-atom Gō-model. In this way, the all-atom Gō-model exhibits a *higher* folding cooperativity compared with the C_α Gō-model.

The Φ-values calculated from the all-atom Gō-model have better agreement with experimental results, especially for the residues corresponding to the α-helical structures in CI2, because of the inclusion of the side-chain packing in folding. The side-chain topology is believed to play more important roles in determining the folding of structures that are mostly stabilized by local interactions, such as α-helices. In contrast, the folding of β-sheets can be well characterized by the traditional C_α Gō-model because the transition state ensemble of β-structures is mostly determined by the loop entropy. Although the side-chain topology plays a less important role in the formation of β-sheets, there will be other factors that influence the folding behavior, for example, the chemicophysical properties of residues such as hydrophobic clusters between the β-strands. In addition, the inherent limitation of the Gō-like potentials is also reflected in the all-atom Gō-model. The preassumed turn positions in the Gō-like potential greatly accelerate the formation of the β-hairpins, thus resulting in overestimation of the Φ-values for β-sheets.

Moreover, the all-atom Gō-model is able to qualitatively reproduce the side-chain ordering process. The side-chain ordering proceeds slower than the backbone folding for both CI2 and SH3, especially at lower temperatures. From the detailed trajectory analysis, we found that the side-chain ordering and backbone folding are *not* independent. The folding of proteins needs coupling from side chains.

In conclusion, the all-atom Gō-like model successfully compromises the precision and the computational accessibility. The improvements of such a model indicate the importance of incorporating the side-chain topology in simulating protein folding kinetics, especially for the proteins that have a large content of local interactions. The generality of our definition of contacts in the all-atom Gō-model facilitates the applicability of this model to study other proteins, and further study using this model can be expected to yield more insights to the protein folding problem.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 90403120, 10504012, and 10474041), and the National Basic Research Program of China (2006CB910302 and 2007CB814806). The simulation was done at the Shanghai Supercomputer Center.

[1] J. E. Shea, J. N. Onuchic, and C. L. Brooks III, Proc. Natl. Acad. Sci. U.S.A. **96**, 12512 (1999).
[2] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).
[3] C. Clementi, P. A. Jennings, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **97**, 5871 (2000).
[4] C. Clementi, P. A. Jennings, and J. N. Onuchic, J. Mol. Biol. **311**, 879 (2001).
[5] N. Koga and S. J. Takada, J. Mol. Biol. **313**, 171 (2001).
[6] W. X. Xu, J. Wang, and W. Wang, Proteins **61**, 777 (2005).
[7] G. H. Zuo, J. Wang, and W. Wang, Proteins **63**, 165 (2006).
[8] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).
[9] A. J. Doig and J. E. Sternberg, Protein Sci. **4**, 2247 (1995).
[10] S. D. Pickett and M. J. E. Sternberg, J. Mol. Biol. **231**, 825 (1993).
[11] M. Blaber, X. J. Zhang, and B. W. Matthews, Science **260**, 1637 (1993).
[12] S. S. Plotkin and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **97**, 6509 (2000).
[13] M. Lingberg, J. Tångrot, and M. Oliveberg, Nat. Struct. Biol. **9**, 818 (2002).
[14] T. Lazaridis and M. Karplus, Science **278**, 1928 (1997).
[15] A. Li and V. Daggett, Proc. Natl. Acad. Sci. U.S.A. **91**, 10430 (1994).
[16] A. Li and V. Daggett, J. Mol. Biol. **257**, 412 (1996).
[17] J. Karanicolas and C. L. Brooks III, J. Mol. Biol. **334**, 309 (2003).
[18] S. Y. Lee, Y. Fujitsuka, D. H. Kim, and S. Takada, Proteins **55**, 128 (2004).
[19] L. Li and E. I. Shakhnovich, Proc. Natl. Acad. Sci. U.S.A. **98**, 13014 (2001).
[20] J. Shimada and E. I. Shakhnovich, Proc. Natl. Acad. Sci. U.S.A. **99**, 11175 (2002).
[21] C. Clementi, A. E. Garcia, and J. N. Onuchic, J. Mol. Biol. **326**, 933 (2003).
[22] S. E. Jackson, N. elMasry, and A. R. Fersht, Biochemistry **32**, 11270 (1993).
[23] L. S. Itzhaki, D. E. Otzen, and A. R. Fersht, J. Mol. Biol. **254**, 260 (1995).
[24] V. Grantcharova, D. Riddle, J. Santiago, and D. Baker, Nat. Struct. Biol. **5**, 714 (1998).
[25] V. Sobolev, R. Wade, G. Vriend, and M. Edelman, Proteins: Struct., Funct., Genet. **25**, 120 (1996).
[26] R. H. Swendsen, Physica A **194**, 53 (1993).
[27] S. L. Kazmirski, K. Wong, S. M. V. Freund, Y. Tan, A. R. Fersht, and V. Daggett, Proc. Natl. Acad. Sci. U.S.A. **98**, 4349 (2001).
[28] H. Kaya and H. S. Chan, Proteins **40**, 637 (2000).
[29] H. S. Chan, S. Shimizu, and H. Kaya, Methods Enzymol. **380**, 350 (2004).
[30] A. R. Fersht, Curr. Opin. Struct. Biol. **5**, 79 (1994).
[31] A. R. Fersht and S. Sato, Proc. Natl. Acad. Sci. U.S.A. **101**, 7976 (2004).
[32] P. J. Fleminga and F. M. Richardsa, J. Mol. Biol. **299**, 487 (2000).
[33] D. Du, Y. J. Zhu, C. Y. Huang, and F. Gai, Proc. Natl. Acad. Sci. U.S.A. **101**, 15915 (2004).
[34] E. Kussell and E. I. Shakhnovich, Phys. Rev. Lett. **89**, 168101 (2002).
[35] Y. Wei and W. Nadler, J. Chem. Phys. **125**, 164902 (2006).