Nucleation process in the folding of a domain-swapped dimer

Zhiqiang Yan, Jun Wang,[*,†] Yujie Zhang, Meng Qin, and Wei Wang[*,‡]
National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing 210093, China
(Received 12 July 2009; revised manuscript received 7 December 2009; published 8 February 2010)

Nucleation processes are important for the understanding in protein dynamics. To evaluate the effect of nucleation mechanism in dimerization process, a domain-swapped dimer (Esp8) is simulated with the symmetrized Gō model and the classical Gō model. The pathways of the dimerization are analyzed with computational ϕ-analysis method. It is found out that some nuclei are observed in the kinetic steps of the dimeric association though the whole pathway is a process with multiple intermediate states. The key residues in the nuclei are rather similar to those observed in the monomeric folding. The differences with the monomeric cases are also discussed. These differences illustrate the effects of dimeric feature on the nucleation process. Besides, manual mutations are carried out to illustrate the importance of the interactions related to the nuclei. It is observed that the mutations in the nuclei-related interactions apparently change the dynamics while other mutations have little effect on the kinetics. All of these results outline a picture that the nucleation processes act as the fundamental steps of high-order organization of protein systems.

DOI: 10.1103/PhysRevE.81.021910 PACS number(s): 87.15.Cc, 87.15.hg, 87.15.km

I. INTRODUCTION

Nucleation is a typical process for protein molecules to form ordered native structures. This kind of process is a signature of monomeric proteins different from random polymers [1–5]. A lot of researches are proposed to characterize the cooperativity and the pathways related to the nucleation during the folding processes of small single-domain proteins [6–13]. The identification of the nuclei in protein molecules (namely, the key residues and the related interactions which contribute essentially to folding dynamics) greatly promotes the understanding on the folding of monomeric proteins [13–18]. The nucleation-condensation mechanism has become one of the paradigms to understand the physics of protein folding processes [5,19–21].

Recently, the binding, association, and aggregation between multiple protein molecules have attracted much more interest in searching for key interactions governing the formation of the complexes and in observing the mechanisms emerging in the self-organizations of multiple proteins (see the reviews [22–24] and the references therein). Though there are no comprehensive theories for these complicated processes, many observations imply that these processes are often related to the behavior of nucleation. For example, during the formation of amyloid fibrils, there are some compact intermediate structures between the disordered oligomers and the prefibrils. This kind of intermediates act as the boundary between two phases with different structural orders and toxicity [25–27]. A "nucleated conformational conversion" has been proposed to understand such kind of observations, and has been supported with some experiments and theories [28–31]. The nucleation is still believed to be an important mechanism for the dynamic processes with multiple protein chains.

Meanwhile, due to the complex interactions and the great flexibility of the proteins, the realistic dynamics of the oligomerization and aggregation are much more complicated than the two-state folding behaviors of small monomeric proteins. When would the nucleation happen during the association and aggregation processes? Would the nucleation be also related to the interactions between some specific residues? What would be the differences between the nucleation in the association and aggregation processes and that in the monomeric folding? The answers to these questions could bridge the understanding between the monomeric folding behaviors and the dynamics of multiple molecules, and would help to build up a physical picture for the complex processes of protein association and aggregation.

In this work, the nucleation in the association processes of proteins is systematically investigated, since the binding and association between proteins act as the basic steps to form higher-order structures and complex aggregates. Based on a symmetrized Gō model, the dimerization between two SH3 domains through domain-swapping behavior is simulated. The whole dimerization progresses step by step with multiple intermediate states, and cannot be regarded as a nucleation process. Meanwhile, the transitions between the intermediates are cooperative, which suggests the possible existence of nucleation in each step of the binding processes. These steps are further analyzed in detail with a classical Gō model. For each step, a cluster of contacts are observed to be formed earlier than the others in the rate-limiting barrier. The concerned interactions and residues establish a nucleus related to the formation of a certain subunit. The residues in the nucleus are similar as those appeared in the nucleus for monomeric folding. This is consistent with the picture of minimal frustration for the domain-swapping-based dimerization. The differences between the nuclei for dimerization and for monomeric folding are also discussed. The differences reflect the effects of association behaviors. To evaluate and illustrate the importance of the nucleus-related interactions, manual mutations in a series of contacts are carried out. There are prominent variations in the association kinetics when just a few of the contacts related to nucleus are

*Corresponding author.
†FAX: 86-25-83595535; wangj@nju.edu.cn
‡FAX: 86-25-83595535; wangwei@nju.edu.cn

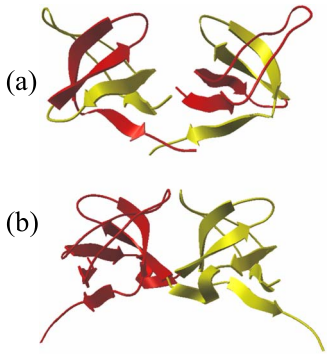

FIG. 1. (Color online) Cartoon view of (a) two MSH3 and (b) DSH3. Two subunits of DSH3 are rendered with different colors.

mutated. This further demonstrates that the nucleation is one of the important factors during the complex organization of proteins.

II. MODELS AND METHODS

A. Protein model

In this work, an intertwined, domain-swapped dimer of epidermal growth factor receptor pathway substrate 8 (Eps8) src homology 3 (SH3) domain is adopted as the model system. The monomeric structure (MSH3, PDB entry 1I0C) and dimeric structure (DSH3, PDB entry 1I07) are shown in Fig. 1 [32,33]. In the Eps8 SH3 dimer, two monomers are defined as chain *A* and chain *B*, respectively, according to the naming convention in the PDB file. Each monomer has about half of its residues exchanged with the corresponding part of the other monomer. Two structural subunits are formed through this domain-swapping behavior. Each subunit could be highly superimposable with the monomeric structure except the hinge region. These two structural subunits are named as SU1 (with the residues 6–39 in chain *A* and the residues 40–64 in chain *B*) and SU2 (with the rest of the residues), respectively. These two subunits are rendered with different colors in Fig. 1(b). The reason to choose Eps8 SH3 dimer is that the thermodynamics and dynamics of the monomeric SH3 domain have been widely studied in simulations and experiments [14,34–37]. The information on the monomer enables us to carry out the comparisons between nucleation processes of the monomer and the dimer.

In our work, the protein system is modeled with an off-lattice model. The amino acids are represented as the beads located at their corresponding C_β positions (C_α in case of amino acid Gly) [38]. Interactions for the contacts in native structure (namely, the native contacts) are modeled with the attractive potential,

$$H_{i,j}^{\text{contact}} = \begin{cases} 0 & r/d_{ij} > \lambda_2 \\ -\epsilon & \lambda_1 \le r/d_{ij} \le \lambda_2 \\ \infty & r/d_{ij} < \lambda_1, \end{cases} \tag{1}$$

where i and j are indices of the concerned residues, the parameter d_{ij} represents the distance between the residues i and j in the native conformation, and the parameters $\lambda_{1(2)}=0.9(1.3)$ determine the width of the attractive region. Here, the contacts are defined when the distance between the corresponding C_β atoms is smaller than 7.5 Å [34]. The interactions for the other contacts generally take the form of the purely repulsive potential,

$$H_{i,j}^{\text{core}} = \begin{cases} 0 & r/d_{\text{core}} > 1 \\ \infty & \text{otherwise}, \end{cases} \tag{2}$$

in which the parameter $d_{\text{core}}=4.5$ Å measures the size of repulsive core of the residues. Besides, a bond function is applied to neighboring residues,

$$H_{i,i+1}^{\text{bond}} = \begin{cases} 0, & 1-\delta \le r/d_{i,i+1} \le 1+\delta \\ \infty, & \text{otherwise}, \end{cases} \tag{3}$$

where the parameter $\delta=0.02$ gives out the tolerance for length fluctuation of the bonds. This kind of potential has been widely applied in many protein systems [14,39,40].

B. Symmetrized Gō model and classical Gō model

To have a thorough study on the dynamics of SH3 dimer, two kinds of Gō-like models are employed to simulate the dimerization processes, the symmetrized Gō model and the classical Gō model. These models are all originated from the idea of the minimal frustration principle and have been successfully used in dimeric modeling [41–43]. The symmetrized Gō model is specifically used to simulate domain-swapping processes of proteins. It considers the similarity between the monomer and the subunits of domain-swapped dimer, and uses the interactions in the monomer structure as the basic input. The intermonomer interactions are assigned based on the interaction in the monomer. That is, for each native intrachain contact between residue i and j in the monomer, there is a corresponding interchain contact between residue i in the chain *A* (or *B*) and j in the chain *B* (or *A*). This kind of interaction is useful to describe the conversion between the monomers and the domain-swapped dimer, and outlines the whole picture for the domain-swapping process. However, this kind of model may meet large barriers related to the unfolding of the monomers. This feature makes the thermodynamic equilibrium difficult in simulations. To enhance the sampling efficiency (especially around the transition states of the association processes), a classical Gō model is also used. The classical Gō model uses the native dimeric conformation as its input. This model can also describe the association of the dimers, since it has almost the same interchain interactions as the symmetrized Gō model due to the similarity between the native structures of the monomer and the subunits of dimer. Clearly, the classical Gō model may not precisely depict the monomeric unfolding, but it could precisely describe the binding of two molecules, which has been successfully applied in many literatures such as [41,43].

In our work, the symmetrized Gō model is used to give a global view on the whole free-energy landscape and the classical Gō model focuses on the detailed analysis for binding processes of the dimer. The interactions of the symmetrized Gō model are derived based on the monomeric native struc-

ture, and the native conformation of the classical Gō model is defined based on the native structure obtained from the simulations with symmetrized Gō model. This kind of assignment ensures the consistence of the simulations with two models. It is worth noting that the dimeric native conformation we used is rather similar to the crystal structure of SH3 dimer. The dRMSD [44] between these two structures is 3.79 Å, which mainly comes from the hinge region. The contacts in our model are also similar to those from the crystal structure, with 95% contacts being the same. These validate both models to simulate such a dimer.

C. Spatial confinement for efficient simulations

For the dimeric system, the translational entropy is an important factor for the dimerization [43,45,46]. Variation in the concentration of the substrate proteins would modulate the translational entropy and apparently affect the balance of dimerization [43]. It is also found that the overall shape of the free-energy landscape may change little following the variation in concentration though the heights of various barriers would be clearly altered. Considering the fact that proper crowded environment could optimize the dimerization[43], a proper concentration of proteins may be chosen in the simulation. In our simulation, a spatial confinement is used to mimic the crowded cellular environment which favors the dimerization. Similar as previous implementation [47,48], an inert spherical shell with its radius as 40 Å is used to encapsulate the Eps8 dimer. This size corresponds to the case with protein concentration of 12 mM. This concentration is on the same order as that in experiments [32,33] to observe coexisting of monomeric and dimeric SH3 domains in a dynamic equilibrium. This kind of modeling would speed up the conformational search and may not change the basic picture for the dimerization. The spherical wall is set as an impenetrable wall for residues, which induces elastic collisions when beads hit on the spherical surface. Quantitatively, the confinement interaction for the residue i is described as

$$H_i^{\text{confine}} = \begin{cases} 0 & r > d_{\text{core}}/2 \\ \infty & \text{otherwise.} \end{cases} \tag{4}$$

This kind of "optimization" for simulations has been used in some previous literatures [49].

D. Discrete molecular dynamics

Compatible with the above discontinuous potentials, the event-driven discrete molecular dynamics (DMD) [43,50–53] are used in our simulations. The collisions and propagations of the beads are proceeded sequentially, and all these events build up the kinetics of the systems. With the discontinuous potentials, the processing of the events are all algebra calculations rather than the integrations of dynamic equations, which greatly reduces computational demands. It is worth noting that the event-driven feature makes the time of the evolution not a counting of steps, but a summation of the time related to all propagations. The time unit (τ_0) is defined as $\sigma_0\sqrt{m_0/\epsilon}$, where σ_0 is the unit length, which takes the value of one angstrom (1 Å) and m_0 is the unit of mass. In the DMD simulations, a constant-temperature ensemble is realized with the Andersen's method [54–56]. The collisions with ghost particles are controlled as about 4% of all collisions, which ensures thermodynamic equilibrium.

E. Weighted histogram analysis method

Weighted histogram analysis method (WHAM) is employed to calculate thermodynamic properties [56,57]. In this method, a series of histograms obtained from sampling could be utilized together to find out the density of states and the partition function based on a self-consistent method. With this kind of method, the statistical errors could be minimized for a large range of temperatures. For our discontinuous potential, the minimal interval ($\epsilon=1$) of energy spectrum is generally taken as the bin size of histograms.

III. RESULTS AND DISCUSSIONS

A. Dimerization of the domain-swapped dimer

In the association of multiple protein molecules, there are generally some specific organizations between the molecules. This kind of specificity would produce the oligomers with specific structures. The Eps8 dimer is an example of such kind of molecular organization. This structural feature introduces a generalized rigidity [58] to the systems around the oligomeric conformation. As suggested by the previous studies, the landscape for the oligomers with specific structures are believed to have minimal frustrations [41]. To check whether there are nucleation events during the oligomerization processes becomes a task to extend the landscape theory to oligomeric systems. In this work, we use dimerization process as an instance.

With the symmetrized Gō model, the dimerizations of Esp8 dimer are simulated at various temperatures from $0.9T_F$ to $1.3T_F$. Here, T_F is the temperature corresponding to folding transition of monomeric SH3 domain, which is determined from the peak of heat capacity of the monomeric system. The free-energy landscape at the temperature T_F is built up with WHAM method as a function of the numbers of native contacts between two chains (N_{inter}) and inside one chain (N_{intra}), as shown in Fig. 2(a). Here, N_{intra} measures the similarity of the monomeric structure to the native structure of SH3 domain, and N_{inter} describes the progresses of dimerization via swapping. It is observed that there are five free-energy minima on the landscape. These minima (basins) represent the thermodynamic states of the dimeric system. The basin A has the largest N_{intra}, representing the condition with two isolated folded monomers. Meanwhile, the state E with the largest value of N_{inter}, and is a collection of the structures similar to the native domain-swapped dimeric conformation. The other basins are the intermediates during dimerization. They correspond to the conformations with unfolded monomers (the states B and C) or with partial folded dimer (the state D), respectively. Combined with kinetic trajectories, the dimerization processes could be described as the pathway along these states, as illustrated by the thick arrows on the landscape [as shown in Fig. 2(a)]. It is worth noting that the

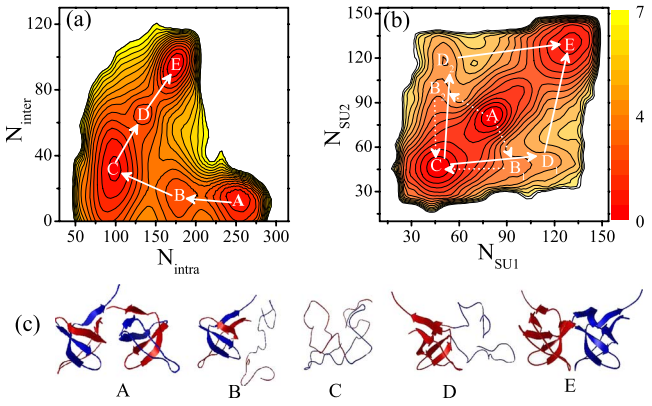

FIG. 2. (Color online) Free-energy landscape of dimer with the coordinates (a) N_{intra} and N_{inter}, (b) N_{SU1} and N_{SU2}. The basins and the pathways related to the dimerization are marked. (c) Cartoon view of typical structures for states *A* to *E*.

dimeric structure could be established from the state *A* to the state *E* step by step through the intermediates (namely, the states *B*, *C*, and *D*). Not only does the unfolding of monomers (from *A* to *C*) have the local minimum with partial unfolded structures, but also the binding between two monomers (from *C* to *E*) experiences the intermediate with partially formed dimeric structures. Both the destruction of monomeric interaction and the establishment of dimeric structure are the processes with the structural order broken or formed progressively. There are not sudden variations in structural order as a whole. The typical conformations for states *A* to *E* are shown in Fig. 2(c). These intermediates are generally related to the partially unfolded structures, and the conformational entropy contributes essentially to the stability of these intermediates. The existence of these kinds of intermediates is also proposed in experiments [59,60].

This kind of behavior with multiple intermediates is apparently different from the cooperative folding for the monomers. It is not easy to judge whether the nucleation happens for such a process. For the binding process, which is largely governed by the native dimeric structure, even multiple pathways are observed in the dynamics, as shown in Fig. 2(b). In Fig. 2(b), the free-energy landscape is reprojected with the coordinates N_{SU1} and N_{SU2}, which record the numbers of formed native contacts related to corresponding subunits. Based on the dynamic trajectories, these intermediates are obligatory for the binding processes from the states *C* to *E*, that is, there are no direct transitions between the states *C* and *E*. This outlines two pathways for the binding process. Two pathways are marked on the landscape with solid arrows. The picture of dimerization process is also consistent with previous simulations with different kinds of confinements [41,42]. This consistence indicates that the extent of translational entropy loss due to dimerization [45,46] wouldn't affect the dimerization pathway although the rate of dimerization is highly affected [43]. During the dimerization process, fully unfolding of two monomers is obligatory along the pathway, it agrees with the experimental observations that binding of the domain-swapped dimer initiates from the unfolded states for proteins such as human prion [61] and p13suc1 [59,60]. It could be concluded that two chains of

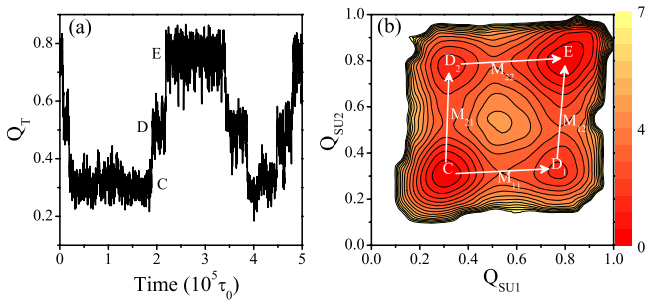

FIG. 3. (Color online) The association process between two chains. (a) The temporal variations of Q_T showing the cooperativity of transitions $C \leftrightarrow D$ and $D \leftrightarrow E$; (b) the free-energy landscape of association processes. The basins and barriers are marked in this figure.

dimer associate in the unfolded state and subsequently fold to the dimeric state. This kind of behavior makes it probable to investigate the nucleation events during the dimerization with a similar way used in protein folding.

B. Contacts and residues affecting dimerization kinetics

As shown above, the unfolding of SH3 domains are generally monomeric events, while the binding processes during the dimerization (namely, from *C* to *E*) mainly relates to the formation of the interchain contacts. This reflects that the unfolding and binding of the monomers are controlled by the intra- and interchain interactions, respectively. With this fact, the classical Gō model with precise interchain interactions could be used to properly describe the binding processes of the dimer. Therefore, for the domain-swapped protein systems, the classical Gō model is not a duplication of symmetrized Gō model, but a refinement for the description of the binding processes.

Different from the global multistate feature, the transitions between the sequential states (such as from *C* to *D*, or from *D* to *E*) are rather cooperative [as shown in Fig. 3(a)]. These transitions are all related to the large structural variations in certain parts of proteins. These steps exhibit some nucleation signals. This kind of phenomenon is also observed in many experiments for dimers [59,60]. It is valuable to investigate these steps in detail to disclose whether and how the nucleation happens for the association processes of proteins. The landscape for the binding processes at the temperature $T=1.03T_F$ is constructed with the coordinates Q_{SU1} and Q_{SU2} based on the similar method as in the last section [as shown in Fig. 3(b)]. Here, $Q_{SU1}=N_{SU1}/N_{SU1}^{\text{Total}}$ and $Q_{SU2}=N_{SU2}/N_{SU2}^{\text{Total}}$ are the ratios of native contacts in the concerned subunits, where the normalization factors N_{SU1}^{Total} and N_{SU2}^{Total} are the numbers of all native contacts in the subunits SU1 and SU2, respectively. On this landscape, there are two intermediates, one with a large value of Q_{SU1} and a small value of Q_{SU2}, and the other with a smaller Q_{SU1} and a larger Q_{SU2}. They are marked as D_1 and D_2, respectively. The coordinates $Q_{SU1,2}$ for these states indicate that the intermediates are generally have one subunit well formed and the other with random structures. With these intermediates, the kinetics for the dimerization could be described as a series of

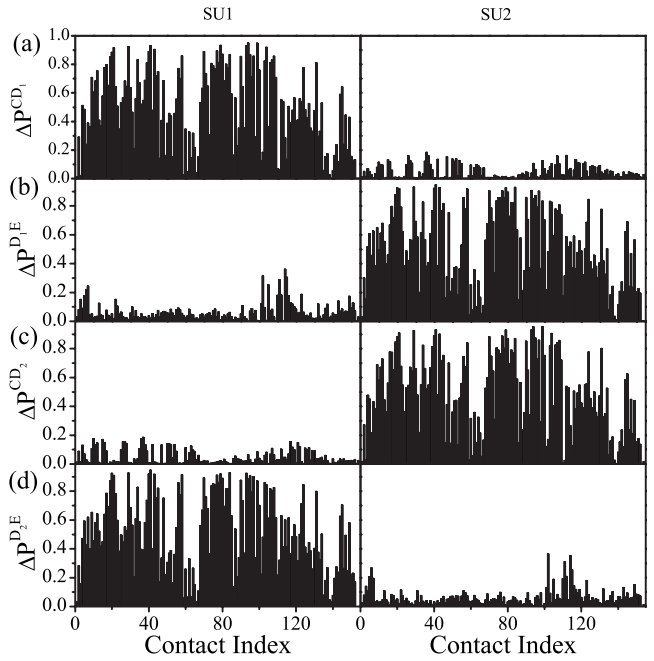

FIG. 4. Variations in formation probabilities of native contacts between basins (a) ΔP^{CD_1}, (b) ΔP^{D_1E}, (c) ΔP^{CD_2}, (d) ΔP^{D_2E}. Here, the contacts are aligned based on the concerned residue indices.

cooperative steps, $C \leftrightarrow D_1(D_2) \leftrightarrow E$. These observations are consistent with the results from the symmetrized Gō model and some related simulation [41] and experiments [59,60]. It implies that the simplified interaction captures the basic essence of the binding process in domain-swapped dimerizations.

For each step $X \rightarrow Y$ of the dimerization (such as $C \rightarrow D_1$), the protein system behaves somehow like a two-state folder. There is a rate-limiting free-energy barrier M between the state X and Y [as the regions $M_{11,12,21,22}$ shown in Fig. 3(b)]. Similar as the ϕ value widely used in analysis of two-state folding [7], a parameter $\phi=(P_\mu^M-P_\mu^X)/(P_\mu^Y-P_\mu^X)$ is defined to indicate the kinetic progress of a contact μ in the rate-limiting state M related to the step $X \rightarrow Y$, where P_μ^S measures the formation probability of the contact μ in the state S. A contact with a large ϕ value indicates that this contact has been well formed in the rate-limiting step and would be an essential part of the interactions inside the nucleus of the concerned process. ϕ value provides a way to identify the contacts related to nucleation processes. Practically, for the steps of dimerization, the denominator $\Delta P^{XY}=P_\mu^Y-P_\mu^X$ may take a small value for the contacts irrelevant to the concerned step, while, for the other contacts, ΔP^{XY} would have the values as large as 1 (as shown in Fig. 4). With this feature, the ϕ value could be simplified as $\varphi=P_\mu^M-P_\mu^X$, which avoids ill definition for ϕ values in some cases.

For four steps in the dimerization, the parameter φ are evaluated for all contacts (as shown in Fig. 5). The practical definitions for the concerned states are given in Table I. It is found that the φ values are not uniformly distributed for contacts in all kinetic steps. Based on the phrases for two-state folders, they generally have highly polarized "transition state." For example, for the step $C \rightarrow D_1$ [as shown in Fig. 5(a)] the interchain contacts with large φ values ($\varphi > 0.55$) are congregated together and form in three clusters, while the other contacts have small values of φ. This difference between these contacts is clearly demonstrated from the color rendering in Fig. 5(a). Mapping these high-φ clusters of contacts to the protein structure, it is found that these contacts are related to the interactions between distal hairpin and some β strands. It is reasonable to point out that a contact network centered around a few residues $L23_A$, $L31_A$, $V43_B$, and $G51_B$ (where the subscripts represent the concerned chains) is well formed in the rate-limiting barrier M_{11}. This gives out a nucleation picture for this step. This kind of phenomenon could also be observed in other kinetic steps [as shown in Figs. 5(b)–5(d)]. All the rate-limiting barriers have polarized structures, which is consistent with our speculation based on cooperative behaviors of the kinetic steps. It is interesting to find out that the key residues related to the nucleus are almost the same for four steps. This is due to the similarity of two concerned subunits. That is, the similarity of structure of subunits would make all the four steps experience similar driving forces since each step corresponds to the formation of a certain subunit. These results outline a picture that the dimerization is implemented with a series of nucleation steps. The nucleation would be fundamental steps of high-order organization of protein systems.

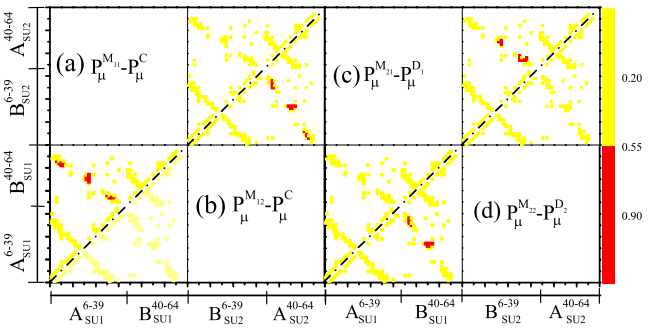

FIG. 5. (Color online) The simplified φ factor for various transitions, (a) $P^{M11}-P^C$, (b) $P^{M12}-P^C$, (c) $P^{M21}-P^{D1}$, (d) $P^{M22}-P^{D2}$. These subfigures are shown in triangular areas, separated with thick dash-dot lines.

It is also interesting to compare the structure and interaction of nucleus for these kinetic steps with those for monomeric SH3 domain. It is observed that the concerned residues in nucleus for both cases are rather similar except that they are located in different chains for the dimeric cases. This is attributed to the structural similarity of the subunits and the monomer. This supports the idea that the dimerization is also related to a landscape with minimal frustration as that for monomer proteins [13–18]. With this observation, it could be concluded that the nucleation in the steps of dimeric association would largely affected by the monomeric features. This proposes some physical understandings on the question of how the nucleation enters into the dynamics of high-order organization of proteins. Though this observation may be closely related to the domain-swapping feature, we believe that the nucleation feature in monomeric dynamics would greatly affect the behaviors of oligomers and further aggregates, as suggested by the previous work [62]. Surely, the

TABLE I. The parameters N_T and N_I to determine the unfolded state (C), the intermediate states (D_1 and D_2), the dimeric native state (E) and the barriers during the pathways (M_{11}, M_{12}, M_{21}, and M_{22}).

State	C	M_{11}	M_{21}	D_1	D_2	M_{12}	M_{21}	E
N_{SU1}	<46	82–87	46–51	116–121	46–51	121–126	79–84	>123
N_{SU2}	<46	46–51	82–87	46–51	116–121	79–84	121–126	>123

dimeric feature also introduces some properties for the nucleation in the kinetic steps. For the first step of the binding processes (namely, the step $C \rightarrow D_1$ or $C \rightarrow D_2$), there are more interactions (contacts) in the nucleus comparing with that for the monomeric SH3 domain. This is because there is larger translational entropy (rather than the configurational entropy only for monomeric case) which is necessary to be balanced with interactions in the rate-limiting barrier. This kind of effect may be more important for larger system with more chains. The translational motion of the chains provides some modulations for the nucleation in the initial recognition processes. Meanwhile, for the second step of association (namely, the step $D_1 \rightarrow E$ or $D_2 \rightarrow E$), there are apparently fewer well formed contacts in the rate-limiting barrier than those in the first step. The difference of the φ between the contacts in nucleus and out of nucleus is also smaller. This implies weaker cooperativity for the second step. Besides the decrease in translational entropy due to the constraint by the formed subunit, the interactions between the hinge region of dimers induce the formation of the second subunit, which weakens the effect of nucleation as observed in Figs. 5(c) and 5(d). These observations demonstrate the modulation of the interchain interaction on the intrinsic nucleation processes. As a conclusion, these comparisons between nucleation processes in dimeric and monomeric dynamics further enrich our picture about the nucleation-coupled high-order dynamics.

C. Mutations of contacts related to nucleus

As discussed above, the dimerization of SH3 domains is composed of a series of steps with nucleation features. In this sense, it is possible to speculate that the modification for just a few interactions related to nucleus may alter the dynamic features of whole dimerization. This idea is an extrapolation of the observation on nucleation processes for monomeric proteins [63]. The evaluation for such an idea is another check for the effect of nucleation in dimerization processes.

To check the effect of the interactions (contacts) related to nucleus on the dynamics, a series of computational mutations are carried out. The schemes of mutations include (I) the deletion of two interchain contacts related to the nucleus; (II) the deletion of two interchain contacts irrelevant to the nucleus; (III) the deletion of two intrachain contacts. Considering the situation that two subunits are the same in their contacts, the mutated contacts are randomly selected in the subunit SU1. The results are not sensitive to the selection of contacts for all cases. In our simulations, the mutated contacts are given in Table II. After the mutation, it is found out that there is an apparent variation for the free-energy landscape of case I [Fig. 6(a)] while tiny changes are observed for the cases II and III [Figs. 6(b) and 6(c)]. For the case I, the minima D_1 becomes much more shallow comparing with that for wildtype case. The cease of the intermediate results in a high barrier for the dynamics from C to E through the state D_1. Therefore, in this case, the dimerization would largely proceed through the pathway along the state D_2. That is, the formation of the dimer would often be initiated through a certain part of protein chains. Just two contacts (about 1.3% of all contacts) produce such an apparent change for the dynamics. On the other hand, the mutations for other two contacts irrelevant to nucleus have little effect on the landscape and the dynamics. The comparison between these cases clearly illustrates the importance of interactions related to nucleus. These observations further cements the understanding on the relationship between the nucleation and high-order dynamics.

IV. CONCLUSION

In this work, an off-lattice model with the symmetrized and the classical Gō-type potential is used to study the thermodynamics and kinetics of a domain-swapped dimer of Eps8 SH3 domain. Based on our simulations, the dimerization is progressed step by step without clear nucleation signals. Meanwhile, in each step of the binding process, a few residues and interactions take an important role in the corresponding rate-limiting barrier. The key residues and interactions are similar between subunits and monomer, and resemble the nucleation process. These illustrate that the

TABLE II. The contacts to be deleted in our computational mutations. They are (I) the interchain contacts related to nucleus, (II) the interchain contacts irrelevant to nucleus, (III) the intrachain contacts irrelevant to nucleus.

Categories	I. Interchain contacts related to nucleus	II. Interchain contacts irrelevant to nucleus	III. Intrachain contacts irrelevant to nucleus
Contacts	$L23_A$-$S50_B$	$L10_A$-$D59_B$	$W10_A$-$L31_A$
	$E22_A$-$G51_B$	$D35_A$-$W41_B$	$R43_B$-$V53_B$

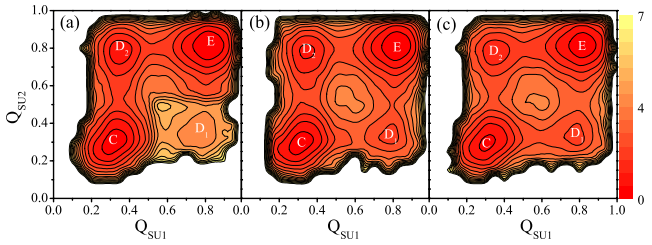

FIG. 6. (Color online) Free-energy landscapes with coordinates Q_{SU1} and Q_{SU2} for mutants (a) deleting two interchain contacts in nucleus (b) deleting two interchain contacts irrelevant to nucleus (c) two intrachain contacts.

nucleation acts as the fundamental step of high-order organization of proteins. Computational mutations for the dimeric interaction are also carried out. The results illustrate that the interactions related to nucleus residues are important to adjust the kinetic pathways and that the other interactions have weaker effect. This further supports the nucleation-condensation picture. It also demonstrates the way to change the association dynamics efficiently. As a conclusion, the dimerization processes of this dimer via domain swapping are tightly related to the nucleation processes. This is also a result of the minimal frustration feature for natural proteins. This kind of characteristics may be helpful to understand the domain-swapped dimerization processes and high-order aggregations.

ACKNOWLEDGMENTS

This work was supported by the Foundation of NNSF (Grants No. 10774069 and No. 10834002), National Basic Research Program of China (Grants No. 2006CB910302 and No. 2007CB814806), and the support from Jiangsu Province (Grant No. BK2009008). J.W. thanks the support from FANEDD, and Z.Q.Y. thanks the support from Scientific Research Foundation of Graduate School of Nanjing University (Grant No. 2007CL08). We thank Shanghai Supercomputer Center and High Performance Computing Center of Nanjing University for the computational supports.

[1] D. Thirumalai and Z. Guo, Biopolymers **35**, 137 (1995).
[2] V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich, Biochemistry **33**, 10026 (1994).
[3] A. R. Fersht, Curr. Opin. Struct. Biol. **7**, 3 (1997).
[4] H. Frauenfelder, S. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).
[5] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[6] D. K. Klimov and D. Thirumalai, Proteins **43**, 465 (2001).
[7] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).
[8] I. A. Hubner, K. A. Edmonds, and E. I. Shakhnovich, J. Mol. Biol. **349**, 424 (2005).
[9] I. A. Hubner, J. Shimada, and E. I. Shakhnovich, J. Mol. Biol. **336**, 745 (2004).
[10] V. S. Pande, A. Yu Grosberg, T. Tanaka, and D. S. Rokhsar, Curr. Opin. Struct. Biol. **8**, 68 (1998).
[11] H. S. Chan, S. Shimizu, and H. Kaya, Methods Enzymol. **380**, 350 (2004).
[12] J. E. Shea, J. N. Onuchic, and C. L. Brooks, Proc. Natl. Acad. Sci. U.S.A. **99**, 16064 (2002).
[13] V. S. Pande and D. S. Rokhsar, Proc. Natl. Acad. Sci. U.S.A. **96**, 1273 (1999).
[14] F. Ding, N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich, Biophys. J. **83**, 3525 (2002).
[15] G. Settanni, J. Gsponer, and A. Caflisch, Biophys. J. **86**, 1691 (2004).
[16] N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich, J. Mol. Biol. **296**, 1183 (2000).
[17] I. A. Hubner, M. Oliveberg, and E. I. Shakhnovich, Proc. Natl. Acad. Sci. U.S.A. **101**, 8354 (2004).
[18] M. Vendruscolo, E. Paci, C. M. Dobson, and M. Karplus, Nature (London) **409**, 641 (2001).
[19] B. Nölting and D. A. Agard, Proteins: Struct., Funct., Bioinf. **73**, 754 (2008).
[20] V. Daggett and A. R. Fersht, Trends Biochem. Sci. **28**, 18 (2003).
[21] L. Lindorff-Larsen, P. Røgen, E. Paci, M. Vendruscolo, and C. M. Dobson, Trends Biochem. Sci. **30**, 13 (2005).
[22] A. M. Morris, M. A. Watzky, and R. G. Finke, Biochim. Biophys. Acta **1794**, 375 (2009).
[23] G. Morra, M. Meli, and G. Colombo, Curr. Protein Pept. Sci. **9**, 181 (2008).
[24] R. V. Pappu, X. Wang, A. Vitalis, and S. L. Crick, Arch. Biochem. Biophys. **469**, 132 (2008).
[25] J. R. Silveira, G. J. Raymond, A. G. Hughson, R. E. Race, V. L. Sim, and B. Caughey, Nature (London) **437**, 257 (2005).
[26] V. N. Uversky, J. Li, and A. L. Fink, J. Biol. Chem. **276**, 10737 (2001).
[27] J. M. Gruschus, Amyloid **15**, 160 (2008).
[28] T. R. Serio, A. G. Cashikar, A. S. Kowal, G. J. Sawicki, J. J. Moslehi, L. Serpell, M. E. Arnsdorf, and S. L. Lindquist, Science **289**, 1317 (2000).
[29] T. Scheibel, J. Bloom, and S. L. Lindquist, Proc. Natl. Acad. Sci. U.S.A. **101**, 2287 (2004).
[30] H. D. Nguyen and C. K. Hall, Proc. Natl. Acad. Sci. U.S.A. **101**, 16180 (2004).
[31] S. Auer, C. M. Dobson, M. Vendruscolo, and A. Maritan, Phys. Rev. Lett. **101**, 258101 (2008).
[32] K. V. Radha Kishan, G. Scita, W. T. Wong, P. P. Di Fiore, and M. E. Newcomer, Nat. Struct. Biol. **4**, 739 (1997).
[33] K. V. Radha Kishan, M. E. Newcomer, T. H. Rhodes, and S. D. Guilliot, Protein Sci. **10**, 1046 (2001).
[34] J. M. Borreguero, N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich, J. Mol. Biol. **318**, 863 (2002).
[35] A. R. Viguera, L. Serrano, and M. Wilmanns, Nat. Struct. Biol. **3**, 874 (1996).
[36] V. P. Grantcharova, D. S. Riddle, J. V. Santiago, and D. Baker,

Nat. Struct. Biol. **5**, 714 (1998).
[37] D. S. Riddle, V. P. Grantcharova, J. V. Santiago, E. Alm, A. E. Ruczinski, and D. Baker, Nat. Struct. Biol. **6**, 1016 (1999).
[38] B. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Res. **28**, 235 (2000).
[39] Y. Q. Zhou and M. Karplus, Nature (London) **401**, 400 (1999).
[40] Z. Q. Yan, J. Wang, and W. Wang, Proteins: Struct., Funct., Bioinf. **72**, 150 (2008).
[41] S. Yang, S. S. Cho, Y. Levy, M. S. Cheung, H. Levine, P. G. Wolynes, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **101**, 13786 (2004).
[42] F. Ding, K. C. Prutzman, S. L. Campbell, and N. V. Dokholyan, Structure (London) **14**, 5 (2006).
[43] W. Wang, W. X. Xu, Y. Levy, E. Trizac, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **106**, 5517 (2009).
[44] L. Holm and C. Sander, J. Mol. Biol. **233**, 123 (1993).
[45] L. M. Amzel, Proteins **28**, 144 (1997).
[46] X. Siebert and L. M. Amzel, Proteins **54**, 104 (2004).
[47] W. X. Xu, J. Wang, and W. Wang, Proteins **61**, 777 (2005).
[48] A. I. Jewett, A. Baumketner, and J. E. Shea, Proc. Natl. Acad. Sci. U.S.A. **101**, 13192 (2004).
[49] T. Furuta, Y. Fujitsuka, G. Chikenji, and S. Takada, Biophys. J. **94**, 2558 (2008).
[50] B. J. Alder and T. E. Wainwright, J. Chem. Phys. **31**, 459 (1959).
[51] D. C. Rapaport, J. Chem. Phys. **71**, 3299 (1979).
[52] Y. Q. Zhou and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **94**, 14429 (1997).
[53] S. W. Smith, C. K. Hall, and B. D. Freeman, J. Comput. Phys. **134**, 16 (1997).
[54] H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).
[55] A. Voegler Smith and C. K. Hall, Proteins **44**, 344 (2001).
[56] Y. Q. Zhou and M. Karplus, J. Chem. Phys. **107**, 10691 (1997).
[57] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[58] P. W. Anderson, *Basic Notions of Condensed Matter Physics* (Benjamin/Cummings, New York, 1997).
[59] F. Rousseau, J. W. H. Schymkowitz, H. R. Wilkinson, and L. S. Itzhaki, Structure **10**, 649 (2002).
[60] F. Rousseau, J. W. H. Schymkowitz, H. R. Wilkinson, and L. S. Itzhaki, J. Biol. Chem. **279**, 8368 (2004).
[61] K. J. Knaus, M. Morillas, W. Swietnicki, M. Malone, W. K. Surewicz, and V. C. Yee, Nat. Struct. Biol. **8**, 770 (2001).
[62] M. Silow, Y. J. Tan, A. R. Fersht, and M. Oliveberg, Biochemistry **38**, 13006 (1999).
[63] J. Li, J. Wang, and W. Wang, Proteins: Struct., Funct., Bioinf. **71**, 1899 (2008).