Effectiveness of Φ-value analysis in the binding process of Arc repressor dimer

Xinlu Guo, Jian Zhang, Le Chang, Jun Wang, and Wei Wang
National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, 210093 China
(Received 14 December 2010; revised manuscript received 12 May 2011; published 14 July 2011)

Φ-value analysis has proved to be a powerful technique in protein folding studies. It has been used to study the transition state structures, to infer microscopic folding pathways, and to identify key residues in the protein folding process. However, its effectiveness in protein binding reaction has not been tested, especially when a fly-casting mechanism is involved. In this article we attempt to answer this question through a coarse-grained study of the binding reaction of Arc repressor dimer. Our simulations show that its binding process proceeds through a fly-casting mechanism, consistent with previous results. We then estimate the importance of the residues for the fly-casting binding by computationally mutating each of them into glycine and calculating their respective effects on the capture radius. It is found that (1) the residues with high Φ values may not be responsible for the large capture radius in the fly-casting binding; (2) mutation of residues with low Φ values may destabilize the denatured states and then increase the capture radius, presumably further increasing the binding rate. Based on our simulations, we conclude that Φ values do not reflect well the importance of residues in the fly-casting binding reaction and need to be combined with other techniques to provide a complete picture.

DOI: 10.1103/PhysRevE.84.011909 PACS number(s): 87.15.A—, 87.15.Cc

I. INTRODUCTION

Understanding the recognition and binding process between macromolecules is important for understanding their functions, designing drugs, and helping people find ways of inhibiting pathogenic protein aggregations. Several scenarios have been proposed for describing the recognition process between molecules. Early theory dates back to the “lock and key” mechanism [1], which depicted the enzyme reactions as a fit of the key (substrate) into the lock (enzyme). In this picture both molecules were described as rigid bodies. Over the years, it was gradually recognized that the flexibility of molecules plays an important role in many reactions; therefore, the concept “induced fit” was introduced into the recognition process by Koshland [2]. The induced fit picture developed the lock and key mechanism by stating that the fit between lock and key “occurs only after the changes induced by the substrate itself” [2]. In recent years, growing evidence, primarily by NMR and computer simulations, suggested that proteins in the unbound state populate different conformations and select one of them on binding to its partners [3]. This is the “conformational selection” or “population shift” view of the recognition and binding process.

Although the role of flexibility in the recognition process has been recognized by the “induced fit” model, the range of motions allowed in the folded state is still more restricted than those allowed in the unfolded state. Therefore, it came as a surprise that the intrinsically disordered proteins were common in the cell and the potential advantage of being unfolded before binding was revealed [4,5]. Possibly inspired by these observations, a new scenario of “fly casting” was proposed by Wolynes and colleagues [6].

The essence of this binding mechanism is an increased capture radius due to the larger fluctuation of the (partially) unfolded protein, compared with that of the folded protein [6–8]. The large capture radius facilitates the early recognition of the target; once weak binding contacts are formed between partners, the (partially) unfolded protein reels in the target and folds at the same time. In a theoretical study of the free energy profile of four homodimers performed by Levy and colleagues [7], a long-range and attractive force between monomers was observed. Remarkably, this attraction occurred at a separation distance 30 Å larger than the native separation distance. Later, the same group found that a fly-casting mechanism can be facilitated further by nonspecific electrostatic interactions in the case of the Ets domain of SAP-1 binding to DNA [9]. The attraction between protein and DNA appeared at roughly the same separation distance as in four homodimers, dependent on the strength of the coulomb interactions in the model. Recently, in a 15N relaxation-dispersion experiment, the coupled binding and folding of pKID, an intrinsically disordered protein, to the KIX domain of the CREB binding protein was studied [10]. It was found that pKID binds KIX primarily through nonspecific hydrophobic contacts and evolves by way of an intermediate to the fully folded state without dissociation from KIX, showing a fly-casting scenario. A later computational study of the same protein complex found that increasing the amount of structure in the unbound PKID reduces the binding rate, also suggesting a fly-casting process [11].

In short, the fly-casting scenario may provide a possible explanation for the efficient binding processes in cell and is attracting more and more interests. To fully understand this scenario, it is necessary to investigate how each residue contributes to this scenario and to identify key residues in this process. In protein folding studies, Φ-value analysis has proved to be a powerful tool for understanding protein folding pathways on a microscopic basis and identifying key residues in the folding process. Recently, it has been used to study the binding reaction of protein oligomers to localize the crucial interfacial contacts that are necessary for association [8,12]. However, the effectiveness of Φ values in identifying key residues in the binding reaction, especially when the fly-casting mechanism is involved, requires investigation. Presumably, Φ-value analysis may not work well for this purpose, as important attractions between partners may have
been occurring much earlier than their drifting close and forming transition states. Besides, according to previous studies, the free energy surface as a function of interpartner distance is rather flat and broad at the intermediate and large distance when the fly-casting mechanism is involved [7,9], indicating the failure of the transition state theory itself.

Therefore, in this paper we investigate the relation between \( \Phi \) values and fly-casting binding reaction, that is, how effective the \( \Phi \)-value analysis is in identifying key residues in the binding process in which the fly-casting mechanism is involved. Since the essence of the fly-casting binding is the increased capture radius due to early formation of contacts between interfacial residues, whose mutation will significantly decrease the capture radius and reaction rate, we tentatively use the change of capture radius after mutations to estimate the importance of residues in the fly-casting mechanism. That is, if the mutation of a residue results in a large decrease of capture radius with respect to that of the wild type, the residue is assumed to be important for the binding reaction; otherwise it is not. We then compare the key residues identified in this way with that by \( \Phi \)-value analysis.

The paper is organized as follows. In the methods section we describe our model system, the Arc repressor homodimer, and the methods that are used in this work. In the results section we first present the calculated \( \Phi \) values and discuss the binding kinetics of Arc repressor. We then show the free energy curves and discuss their relationship with the fly-casting mechanism. This part is followed by the central section of the paper, where we try to answer the question raised above by comparing the \( \Phi \) values of the interfacial residues and the change of capture radii after their mutations. Last we discuss the lessons we may learn from this study.

II. MODELS AND METHODS

A. Unlinked and linked models

The Arc repressor dimer (PDB ID: 1ARR) was selected as our model system. The native structure of this protein consists of two antiparallel \( \beta \) sheets and two \( \alpha \) helices, as shown in Fig. 1(a).

We built two models for this protein, the linked model and the unlinked model. In the linked model, as in most previous works, a linker was added to join the C terminus of one subunit to the N terminus of the other subunit to restrict the diffusive motion of the monomers, facilitating the sampling of the binding events. The linker is passive and has no interaction with all the other residues. Experimentally, it has been shown that the transition state of the linked protein has a similar solvent exposure compared with that of the wild type [13], indicating that the transition state is not changed much by the linker. This argument is also supported by our calculation, as is shown later. The length of the linker can be translated into an effective concentration of the monomer. The Arc protein shown in Fig. 1(b) has a linker 15 residues long, corresponding to an effective concentration of 2.7 mM.

Although the linker is passive, it may change the dynamics of the system at large intermonomer distance, which is important for studying the fly-casting mechanism. Therefore we also constructed an unlinked model in which the linker was eliminated and the two monomers were able to diffuse freely in space [Fig. 1(a)].

B. Molecular dynamics simulations

In both the linked and the unlinked models, residues were modeled as beads interacting through a Go-type force field used previously by Clementi et al. [14]:

\[
V(\Gamma, \Gamma_0) = \sum_{\text{bonds}} K_r(r_i - r_0)^2 + \sum_{\text{angles}} K_\phi(\theta_i - \theta_0)^2 + \sum_{\text{dihedral contact}} [K_r^{(1)}(1 - \cos(\phi_i - \phi_0^{(1)}))] + K_\phi^{(3)}[1 - \cos 3(\phi_i - \phi_0^{(3)})]] + \sum_{i>j-3} \varepsilon_1 \left[ 5 \left( \frac{r_{0ij}}{r_{ij}} \right)^{12} - 6 \left( \frac{r_{0ij}}{r_{ij}} \right)^{10} \right] + \sum_{i>j-3} \varepsilon_2 \left( \frac{C}{r_{ij}} \right)^{12},
\]

FIG. 1. (Color online) Molecular structure of the Arc repressor. (a) Ribbon representation of the native structure of the unlinked model. The structure was taken from the PDB database (PDB entry: 1ARR). (b) Ribbon representation of the native structure of the linked model. The linker connects the C terminus of one subunit with the N terminus of the other subunit.
where \( r_i, \theta_i, \) and \( \phi_i \) stand for the \( i \)th virtual bond length between the \( i \)th and \((i + 1)\)th residues, the virtual bond angle centered at the \( i \)th residue, and the virtual dihedral angle around the \( i \)th bond, respectively. \( r_{ij} \) is the distance between the \( i \)th and \( j \)th amino acid residues. The parameters with subscript 0 denote the corresponding values in the native structure. We set \( K_r = 100.0 \varepsilon, K_{\theta} = 20.0 \varepsilon, K_{\phi}^{(4)} = 10.0 \varepsilon, K_{\phi}^{(3)} = 0.5 \varepsilon; \)
\( \varepsilon_1 = 1.0 \varepsilon, \varepsilon_2 = 1.0 \varepsilon, \) and \( C = 4.0 \text{ Å}, \) where \( \varepsilon \) is the unit energy. The interactions for the native and non-native contacts were treated differently, as shown by the last two terms of Eq. (1).

Molecular dynamics (MD) was employed to evolve the system. The equation of motion is described by a Langevin equation as follows:

\[
m\ddot{x}(t) = F_c - \gamma \dot{x}(t) + \Gamma(t),
\]

where \( m \) is the mass of a bead, \( F_c = -\partial V/\partial x \) is the conformational force, which is the negative gradient of potential energy \( V \) with respect to the coordinate \( x, \gamma \) is the friction coefficient, and \( \Gamma \) is the Gaussian random force, which is used to balance the energy dissipation caused by friction [15]. The random force satisfies the autocorrelation function

\[
\langle \Gamma(t) \Gamma(t') \rangle = 2m\gamma k_B T \delta(t - t'),
\]

where \( k_B \) is the Boltzmann constant and \( T \) is the absolute temperature. Readers are referred to previous literature for more technical details [15].

The native contacts were calculated from the PDB structure of Arc repressor (PDB ID: 1ARR). A native contact is deemed to be the observation of rebinding events very hard. Therefore, we employed umbrella sampling to enhance the conformational sampling and to calculate the potential of mean force (PMF) as a function of the intermonomer distance. The intermonomer distance was defined as the distance between the centers of mass (c.m.) of the two monomers [21,22]. Harmonic potentials were used to constrain simulations at a serial of equally spaced windows, ranging from 1 to 100 Å; the separation between neighboring windows was 3.8 Å, leading to a total of 25 windows. The strength of the harmonic potential was set to 5 \( \varepsilon, \) 1/20 times of the strength of the bond interactions. These values guaranteed a sufficient overlap of populations between the neighboring windows, as is shown later. The MD simulation in each umbrella window was run for \( 10^8 \) steps. WHAM algorithm was employed to combine data from multiple windows to get PMF as a function of intermonomer distance [18–20].

**C. Umbrella sampling**

For the unlinked model, the monomers are likely to diffuse away from each other once they unbind, making the observation of rebinding events very hard. Therefore, we employed umbrella sampling to enhance the conformational sampling and to calculate the potential of mean force (PMF) as a function of the intermonomer distance. The intermonomer distance was defined as the distance between the centers of mass (c.m.) of the two monomers [21,22]. Harmonic potentials were used to constrain simulations at a serial of equally spaced windows, ranging from 1 to 100 Å; the separation between neighboring windows was 3.8 Å, leading to a total of 25 windows. The strength of the harmonic potential was set to 5 \( \varepsilon, \) 1/20 times of the strength of the bond interactions. These values guaranteed a sufficient overlap of populations between the neighboring windows, as is shown later. The MD simulation in each umbrella window was run for \( 10^8 \) steps. WHAM algorithm was employed to combine data from multiple windows to get PMF as a function of intermonomer distance [18–20].

**D. \( \Phi \)-value analysis**

\( \Phi \)-value analysis was used to characterize the transition states of the binding process [8,14]. The \( \Phi \) value of a contact pair \( i, j \) was calculated as follows:

\[
\Phi_{ij} = \frac{\Delta AG^{TS - U}}{\Delta AG^{F - U}} \approx \frac{P_{ij}^{TS} - P_{ij}^{U}}{P_{ij}^{F} - P_{ij}^{U}},
\]

where \( \Delta AG \) is the free energy difference between wild type and the mutant and \( P_{ij} \) is the formation probability of the contact between residue \( i \) and \( j \) [23,24]. The subscripts \( F, U, \) and \( TS \) denote the folded, unfolded, and transition state ensembles, respectively. The \( \Phi_i \) of a single residue \( i \) is defined as the average of the \( \Phi_{ij} \) over all residues \( j \) that has contacts with \( i \). For a specific residue, a \( \Phi_i \) close to 1 suggests that the local structure around this residue in the transition state ensemble is similar to that in the folded ensemble, and that this residue plays an important role in the binding process. On the contrary, a \( \Phi_i \) close to 0 means that the local structure is similar to that in the unfolded ensemble and this residue may not be a key one.

The \( F, U, \) and \( TS \) ensembles were determined from the histogram plots of the \( Q \) values. For the linked model, 100 independent direct MD simulations were performed with different seeds of random numbers and started from extended chains. Each trajectory lasts for \( 6 \times 10^7 \) MD steps and at least ten binding/unbinding events were observed in each
Metropolis Monte Carlo (MC) simulations to determine the commitment probability. At each MC step, a trial move toward either smaller or larger intermonomer distance was attempted; the acceptance probability was determined according to the PMF potential using the Metropolis criterion. The simulation was stopped when the virtual ball either commits itself to the binding or unbinding basin of attraction or the simulation length exceeded a threshold. We run 100 independent MC simulations to get the commitment probability for each initial position; the statistical error of the probability was estimated to be 0.02. By repeating this calculation at different initial positions, the commitment probability as a function of the initial position was obtained. The $R_{\text{eff}}$ was designated as the intermonomer distance where the commitment probability was 0.5. The error bar of the $R_{\text{eff}}$ was estimated from the variation of distance when the corresponding commitment probability changes from 0.48 to 0.52. It should be noted that the $R_{\text{eff}}$ was only calculated for the unlinked model. For the linked model, the two monomers were constrained by a linker and they sooner or later bind each other at appropriate temperature. Therefore, the definition of $R_{\text{eff}}$ in this situation makes no sense.

III. RESULTS AND DISCUSSION

A. Kinetics of the linked and unlinked models

Figure 2 shows several typical trajectories at the folding temperature for both linked and unlinked models. The folding temperature was determined from the peak of the heat capacity and found to be 1.055 for this dimer. The accuracy of this value can be seen from the equal population of the native and denatured states (Fig. 2), as the relative population is very sensitive to the temperature near the folding temperature. Figure 2(a) shows that the trajectory for the linked model manifests a cooperative two-state behavior; that is, the monomers would not fold until they bind to each other. The unbound monomers are largely unstructured and no intermediate with unbound folded monomer is observed, in agreement with previous studies [7,8,25–27]. Figure 2(b) presents five independent trajectories from direct MD simulations performed for the unlinked model. It can be seen that these trajectories also show a cooperative two-state behavior, almost indistinguishable from that of the linked model. This similarity between two models suggests that the passive linker does not change the nature of the binding kinetics.

B. $\Phi$-value analysis and the transition state ensemble

The $\Phi_{ij}$ values calculated for both the linked and the unlinked models and for two Go-type force fields are shown in Fig. 3. The distributions of $\Phi_{ij}$ values of four cases are similar, with the maximal correlation coefficient between them to be 0.97, and the minimal to be 0.81. The standard errors of $\Phi_{ij}$ values were estimated to be around 0.02 for the linked model or 0.03 for the unlinked model, which indicates a good statistical quality of the calculated $\Phi$ values.
both studies suggested that the most structured element in the transition state was that between two antiparallel \( \beta \) sheets.

Figure 3 shows that the \( \Phi_{ij} \) values of the intramonomer contacts is asymmetric for the two subunits, indicating that one subunit is more structured than the other in the transition state. This asymmetry is in contrast to previous findings [8]. In that work, although the transition states were also found to be asymmetric in single trajectories, the overall \( \Phi_{ij} \) values were symmetric, attributed to the symmetric native structures by the authors [8]. The above differences between two studies may be attributed to the different definitions of native contacts. We checked our calculated native contacts and found that the native contacts are not exactly symmetric due to the slight asymmetry in the native tertiary structure. For example, the residue E28 in chain B forms 4 contacts with chain A, while E28 in chain A only forms 2 contacts with chain B. Other 8 residues (R16, V25, N29, R31, V33, N34, E36, R40) in the most structured region in chain B were also found forming more interfacial contacts than their corresponding residues in chain A. This asymmetry in the number of native contacts resulted in an asymmetric kinetics; the transition states with a more structured chain B was observed more frequently than the opposite. Therefore, the overall \( \Phi_{ij} \) values is asymmetric.

Furthermore, we calculated \( \Phi_{ij} \) values using the Go model with non-native interactions described in the method section. We found that the asymmetry in the transition states was reduced slightly [Figs. 3(c) and 3(d)], especially for the unlinked model. Therefore, we concluded that the difference in the \( \Phi_{ij} \) values between our work and the previous one [8] is due to the different parameters employed in two studies, such as the definition of native contacts, the threshold used to determine the F, U, and TS ensembles, and the force field employed.

In spite of the difference in the absolute value of \( \Phi_{ij} \) values, both our study and the previous one [8] observed an asymmetric kinetics; that is, one monomer is more structured than its partner in a single binding event. The advantage of this feature for the binding reaction can be understood as follows. If the two monomers were both folded before reaching the transition state, the capture radius will be small and therefore the collision probability is effectively reduced. If two monomers were both unfolded, there is no template for each to bind on, leading to an increased searching phase space and thereby an increased entropic barrier. Therefore, an asymmetric transition state is advantageous for increasing the collision probability as well as decreasing the free-energy barrier.
C. Potential of mean force and the fly-casting mechanism

Although the transition states and $\Phi$ values provide rich information for the binding process, the picture may still be incomplete, as the fly-casting recognition may have occurred much earlier than two monomers drifting close and forming transition state. To investigate the possible binding behaviors especially when the two monomers are separated by a large distance, we calculated the potential of mean force as a function of the inter-monomer distance. For the linked model, direct MD simulations were employed, and for the unlinked model, umbrella sampling technique was used to restrain the simulations at a series of equally spaced windows. The results for the two models at three temperatures are shown in Fig. 4. We also present the distributions of the number of conformations along the separation distance for all umbrella windows in Fig. 5, which indicates that the statistical quality of the calculation is good.

According to Fig. 4, the two models show similar features at short inter-monomer distance (<7 Å); both have a basin of attraction at 7 Å, corresponding to the bound state of the dimer. With increasing the temperature, the bound state is destabilized and this basin gradually disappears.

The PMF at large values of the intermonomer distance (>45 Å) is different for two models. For the linked model, the PMF gradually increases as the distance increases, simply due to the energy penalty from the linker. In contrast, for the unlinked model, the PMF decreases as the distance increases (when $d > 45$ Å), manifesting a repelling force between two monomers within this region. This repelling force is clearly because of the gain of the translational entropy when two monomers drift away from each other.

For the unlinked model [Fig. 4(b)], one remarkable feature is that the PMF curve is rather flat and broad, as has also been seen in previous computational studies [7,9]. It should be noted that the slope of the PMF at the folding temperature ($T = 1.055$) is positive at the intermediate distance (10–45 Å), suggesting that the interaction between partners becomes attractive when they drift close than a critical distance (45 Å). Moreover, the existence of such an attractive force at the intermediate distance is a robust feature, as is seen later in our simulations for many mutants and with a different force field. This critical distance (45 Å) is rather large compared with the radius of the monomer (∼15 Å) and the intermonomer distance at the native state (∼25 Å). It suggests that the monomers have to be partially unstructured to interact with each other at such a large distance, indicating a fly-casting mechanism.

D. The effectiveness of the $\Phi$ values

To test how well the $\Phi$ values can characterize the relative importance of residues in the fly-casting binding process, we did the following calculations. For a specific residue, we mutated it to glycine by deleting all the involved contacts of this residue. Then for this mutant we calculated the PMF as a function of the intermonomer distance and characterized the overall shape of the curve by the effective capture radius $R_{\text{eff}}$. The difference between the $R_{\text{eff}}$ of the mutant and wild type is assumed to be a reflection of the importance of the mutated residue in the fly-casting binding, based on the reasoning discussed in the Introduction. A large decrease of $R_{\text{eff}}$ indicates that the mutated residue plays a key role while a minor change denotes it is not deeply involved.

All the interfacial residues between two monomers were mutated. The $\Phi$ values of these residues versus the new $R_{\text{eff}}$ after mutations are shown in Fig. 6. To facilitate the following discussions, we roughly divide the data points in Fig. 6 into four clusters, represented by V25G, M42G, N11G, and P8G, respectively. Presumably, the mutations of the residues with high $\Phi$ values may lead to a significant decrease of the capture radius. This argument is only roughly correct according to Fig. 6, suggesting that $\Phi$ values are able to characterize the importance of residues in fly-casting binding process, but only qualitatively.
respectively. The wild type and its error. The data points can be roughly divided into four clusters, represented by V25G, M42G, N11G, and P8G, respectively.

Quantitatively, the correlation between two quantities is not good. For example, the cluster represented by P8G and W14G shows an exceptionally reduced capture radii, although the residues in this cluster have similar $\Phi$ values as that in the cluster represented by N11G. This result may be explained as follows. The interaction between P8 and its partners occurs at a very early stage when the intermonomer distance is still large (fly-casting); therefore, the mutation of P8 to glycine greatly reduces the capture radius. In contrast, although the local structure around N11 is similarly structured as that around P8 in the transition state, the binding of residue N11 with its partner occurs late and hence its mutation to glycine does not affect the capture radius very much. The above hypothesis is further supported by Fig. 7, which shows that the binding of P8 with its partner indeed happens much earlier than that of N11.

Another surprising aspect concerns the cluster represented by V25G. The residues in this cluster have smaller $\Phi$ values but their mutations lead to increased capture radii (compared with that of the wild type), manifesting an even pronounced effect of mutations and the associated change of contact networks.

To further visualize the above picture, we present in Fig. 8 the PMF for the representative mutants of the four clusters. The curves are arranged in a decreasing order of $\text{R}_{\text{eff}}$. It can be seen clearly that the mutant P8G decreases the slope of the PMF (from positive to a negative value) at the intermediate and large intermonomer regions, resulting in a significant decrease of $\text{R}_{\text{eff}}$. The mutant V25G destabilizes both the native and unfolded states, leading to an elevation of the curve at the intermediate and large distance region and thus an increased capture radius. In Fig. 8, we also show the representative structures at various binding stages. These structures visualize the following features. (1) The protein will not fold until the partners bind and form an interface. (2) The interactions (fly-casting) between partners appear at a large distance (>45 Å). (3) The fly-casting events usually occur between two $\beta$ sheets. (4) The PMF curves are rather flat and delicately influenced by mutations and the associated change of contact networks.

In short summary, our results demonstrate that the $\Phi$ values cannot solely characterize well the delicate rules of residues in the fly-casting binding. By combining $\Phi$-value analysis and the mutation study of capture radii, the residues can be divided into several groups, for example, the residues mainly taking part in the transition state (e.g., N11), those not getting much involved in the transition state but having influence on the capture radius (e.g., V25), and those playing key roles in both cases (e.g., P8). The combination of two methods is able to give a more complete picture for the binding reaction.

E. Results from the Go model with non-native interactions

To test how the above results depend on the employed model, we set up a modified Go model by adding attractive interactions.
non-native and nonspecific interactions into the standard Go model. We then calculated the \( \Phi \) values, PMFs, and the effective capture radii for the wild type dimer and mutants and compared these results with that obtained by the standard Go model.

Figure 9 gives the PMFs calculated by the modified Go model. It can be seen that the slopes of all curves at the intermediate distance (10–45 \( \text{Å} \)) are significantly increased compared with that in Fig. 8. These larger positive slopes suggest stronger attractive forces, apparently due to the nonspecific interactions introduced in the modified model.

Figure 10 shows the \( \Phi \) values of several representative residues versus and change of capture radii after their mutations, calculated with the modified model. Since the calculation of PMF with umbrella sampling is expensive, we did not make computations for all interfacial residues but only selected these representative ones. According to Fig. 10, roughly speaking, the relative positions of the four cluster representatives are similar to that in Fig. 6, suggesting that our previous discussions of the relationship between \( \Phi \) values and capture radii still hold. Therefore, our results are robust against the variation of force fields.

An interesting difference of Fig. 10 from Fig. 6 is that the capture radii are systematically larger, indicating a more pronounced fly-casting mechanism. The physical picture may be that the probability of the encountering events between monomers is increased by the nonspecific attractive interactions between non-native contacts. This result is in agreement with previous computational studies [11,28].
FIG. 10. (Color online) \( \Phi \) values for several representative residues in the interface versus the capture radii \( R_{\text{eff}} \) of the corresponding mutants, calculated by the Go model with non-native interactions. The horizontal short bars denote the error of the calculated \( R_{\text{eff}} \) for the mutants. The long vertical solid and dashed lines give the \( R_{\text{eff}} \) of the wild type and its error.

**IV. CONCLUSION**

By using the linked and unlinked models and two coarse-grained force fields, we studied the recognition and binding process of the Arc repressor dimer. All models show cooperative two-state behaviors and similar transition state structures. The calculated PMFs as a function of intermonomer distance demonstrate that the binding process proceeds through a fly-casting mechanism.

We investigated the correlation between \( \Phi \) values of the interfacial residues and the change of capture radii after their mutations. It was found that the two quantities roughly correlate with each other, however, with important deviations. First, the mutation of the residues with similarly high \( \Phi \) values leads to significantly different capture radii, for example, the P8G versus N11G. This is attributed to the difference in time stages when these residues get involved in the binding reaction. The residue P8 begins to interact with its partner at a very early stage, much earlier than the dimer’s reaching the transition state; while the residue N11 binds to its partner much later. Second, the mutation of several residues with small \( \Phi \) values (e.g., V25G) results in an increased capture radius and thereby a more pronounced fly-casting mechanism. Detailed free energy analysis revealed that the mutation destabilizes the unfolded state and effectively increases the attraction between monomers, consistent with the fact that many proteins in a cell are intrinsically unstructured. These delicate roles of residues in the fly-casting kinetics cannot be well reflected by their \( \Phi \) values. We also performed the same computations with a modified Go model with non-native interactions and found that the above results are robust against the change of models.

The above results can be understood as follows. Physically, \( \Phi \) values describe the structure of the transition states, which conceptually locate at the top of the free energy barrier. However, to facilitate the recognition and binding process between biomolecules, nature has evolved to take advantage of the (partial) unfolded structures through fly-casting mechanism; therefore, the interactions between partners may have been occurring much earlier than the partners drift close and encounter the major barrier. Therefore, the residues that are important for binding may not be necessarily reflected by their \( \Phi \) values. Due to the same reason, the PMF as a function of partner’s distance is rather flat and broad, as seen in the previous and our studies. This feature of the PMF also shakes the ground of characterizing this type of binding reaction with the transition state theory, which requires a sufficient high free energy barrier.

The fly-casting mechanism emphasizes the importance of flexibility, unfolded structures, and nonspecific interactions in the recognition and binding process. It provides not only a new scenario for the recognition and binding process but also a new challenge for how to characterize such a process. This study suggests that the \( \Phi \)-value analysis is not sufficient for this end and brings forward the need to develop new tools to study this process, especially in this age of protein interaction networks and system biology.

**ACKNOWLEDGMENTS**

This work was supported by the NNSF (Grants No. 10834002, No. 10974088, and No. 10704033), JiangSu Province (Grant No. BK2009008), and the NBRPC (Grant No. 2007CB814800). The authors acknowledge Shanghai Supercomputer Center and HPCC of Nanjing University for the computational supports.