2007

Exploring the Mechanism of Flexible Biomolecular Recognition with Single Molecule Dynamics

H. Peter Lu
Bowling Green State University, hplu@bgsu.edu

Qiang Lu

Jin Wang

Exploring the Mechanism of Flexible Biomolecular Recognition with Single Molecule Dynamics

Qiang Lu,[1,2] H. Peter Lu,[3,*] and Jin Wang[1,2,†]

[1]*State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin 130022, People's Republic of China*

[2]*Department of Chemistry and Department of Physics, State University of New York at Stony Brook, Stony Brook, New York 11794-3400, USA*

[3]*Department of Chemistry, 141 Overman Hall, Bowling Green State University, Bowling Green, Ohio 43403, USA*

(Received 30 January 2006; published 21 March 2007)

Combining a single-molecule study of protein binding with a coarse grained molecular dynamics model including solvent (water molecules) effects, we find that biomolecular recognition is determined by flexibilities in addition to structures. Our single-molecule study shows that binding of CBD (a fragment of Wiskott-Aldrich syndrome protein) to Cdc42 involves bound and loosely bound states, which can be quantitatively explained in our model as a result of binding with large conformational changes. Our model identified certain key residues for binding consistent with mutational experiments. Our study reveals the role of flexibility and a new scenario of dimeric binding between the monomers: first bind and then fold.

DOI: 10.1103/PhysRevLett.98.128105 PACS numbers: 87.15.−v

The study of associations between two biomolecules (e.g., proteins, RNA, or DNA) is the key to understanding molecular recognition and function. A standard paradigm is that molecular function (realized by binding) is determined by the structure of the molecule [1]. In nature, however, the binding between two biomolecules is often accompanied by large conformational changes which are essential for cell function [2–5]. The flexible binding processes or disordered forms of the proteins in the cells can be targeted for rapid enzymatic turnover, thus providing an additional lever of control, and can sometimes be required for regulation. However, the flexible binding processes are not yet completely understood [2–5]. Addressing this issue will answer the critical questions of how molecular function, in addition to structure, is determined by conformational flexibility and dynamics [6].

The high temporal and spatial resolution obtainable in single-molecule fluorescence spectroscopy [7–11] makes it ideal for studying conformational dynamics and localization of proteins under cellular physiological conditions. A Wiskott-Aldrich syndrome protein (WASP) fragment, CBD, that binds only the activated intracellular signaling protein Cdc42 [12–18] was labeled with a novel solvatochromic dye and used to probe hydrophobic interactions at the protein-protein interface significant to Cdc42/CBD recognition. The single-molecule experimental study showed static and dynamic inhomogeneous conformational fluctuations of the protein complex that involved bound and loosely bound states (Fig. 1) [16–18].

Because of the large conformational degrees of freedom involved in flexible binding, an important question is how the huge number of configurations could fall to the unique native state basin. The most natural solution to this so-called Levinthal paradox [19], originally proposed for folding, is the idea that the underlying energy landscape should be funneled against roughness or traps to guarantee both the thermodynamic stability and specificity [4,5,20]. Go models with native structure information and uniform energy biased towards the native state (i.e., an absence of energetic heterogeneity) lead to a funneled landscape and prove to be consistent with many experimental findings in protein folding [21,22]. It is expected that this type of model will also provide useful information about the global topology of the underlying binding-energy landscape [4,5].

In our theoretical investigations, the native structure information and uniform stabilization binding energy were used. We performed the residue-level Go-model simulations of Cdc42-protein/WASP binding dynamics with solvent on a supercomputer at the Pacific Northwest National Lab. Forty long-time trajectories of 80×10^6 steps each were collected for a reliable statistical analysis. We used the experimentally determined native structure of the Cdc42/CBD binding complex (PDB:1cee) [12–15].

The Go model [21,22] takes into account only interactions that exist in the native structure and, therefore, does not include energetic frustration (i.e., includes only topo-

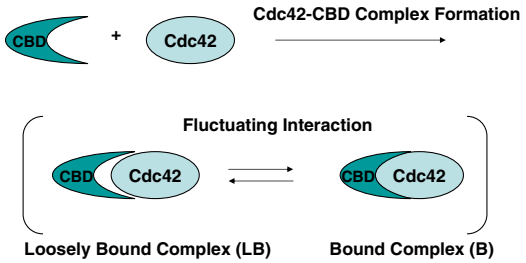

FIG. 1 (color online). A two-state interaction fluctuation model of the Cdc42/CBD complex. The loosely bound state was presumably a subset of conformations that deviated from the bound equilibrium states, probed by the dye, without disrupting the subnanometer long-range interactions, so the overall protein complex was still associated.

logical frustration). In this work, we used an off-lattice Go model, where each residue was represented by a single bead centered on its α-carbon (C α) position. Adjacent beads were pieced together into a polymer chain by means of a potential encoding bond length and angle constraints. The secondary structure was encoded in the dihedral angle potential and the nonbonded (native contact) potential. The interaction energy U at a given protein conformation Γ is given by:

$$U(\Gamma,\Gamma_0) = \sum_{\text{bonds}}^{N-1} K_b(b_i - b_{0i})^2 + \sum_{\text{angles}}^{N-2} K_\theta(\theta_i - \theta_{0i})^2 + \sum_{\text{dihedrals}}^{N-3} \{K_\phi^1(1 - \cos[(\phi_i - \phi_{0i})]) + K_\phi^3(1 - \cos[3(\phi_i - \phi_{0i})])\} + \sum_{|i-j|>3}^{\text{native}} \epsilon\left[5\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{0ij}}{r_{ij}}\right)^{10}\right] + \sum_{|i-j|>3}^{\text{non-native}} \epsilon\left(\frac{C}{r_{ij}}\right)^{12} + U_{\text{water}}.$$

The first three terms in the above equation represent the energies from the backbone chemical bond vibration and dihedral rotations. The fourth term represents the native interaction energy contribution to binding between two residues, and the fifth term represents the non-native interaction energy contribution to binding between two residues. In the equation, b_i, θ_i, and ϕ_i stand for the ith virtual bond length between ith and $(i + 1)$th residues, the virtual bond angle between $(i - 1)$th and ith bonds, and the virtual dihedral angle around the ith bond in conformation Γ, respectively. The parameters b_{0i}, θ_{0i}, and ϕ_{0i} stand for the corresponding variables at the native structure Γ_0. These three terms control the local conformations within four residues. In the framework of the model, all native contacts in the fourth term are represented by the 10–12 Lennard-Jones potential form without any discrimination between the various chemical types of interaction. We used a particular version of contacts of structural units software to count the number of contacts. The total number of native contacts for Cdc42 folding is 468, for the folding of the CBD portion of WASP is 62, and for the binding interface of Cdc42/CBD is 98. The r_{ij} and r_{0ij} are the $C_\alpha - C_\alpha$ distances between the contacting residues i and j in conformation Γ and Γ_0 (the protein data bank structure), respectively. In the summation over non-native contacts in the fifth term, C (=4.0 Å) parametrizes the excluded volume repulsion between residue pairs that do not belong to the given native contact set. The last two terms control the nonlocal interactions. All temperatures and energies are reported in units of 4.0 Å. For other parameters, similar values were used in several folding studies [4,5,21,22], namely, $K_b = 100.0$, $K_\theta = 20.0$, $K_\phi^1 = 1.0$, $K_\phi^3 = 0.5$, and $\epsilon = 1.0$. U_{water} is the potential energy due to the presence of solvent (water molecules), which has been used recently [23] for exploring the cooperative effects of squeezing out water molecules from the partially hydrated protein core.

To enhance the sampling of binding events, we tried a method of linking the two monomers of the dimer by setting the cutoff distance between the two centers of the mass to be within 20 Å. Doing so can significantly reduce the computational time for the two monomers to find each other. This center of mass constraint is realized by implementing a steep parabola with a flat bottom so that the centers of the masses of the two monomers are constrained to be within 20 Å. For the studied dimers, multiple constant temperature molecular dynamics simulations were performed (using the simulation package AMBER6 as an integrator) starting from either the dimeric conformation or the unfolded monomers.

In order to describe the flexible binding, we need at least three reaction coordinates. We define Q_I as the fraction of native spatial tertiary contacts that represent the degrees of freedom of interface binding, Q_{f1} as the fraction of native spatial tertiary contacts that represent the degrees of freedom of folding or the flexible conformational change of Cdc42, and Q_{f2} as the fraction of native spatial tertiary contacts that represent the degrees of freedom of folding or the flexible conformational change of the CBD complex (see Fig. 2).

We investigated the two-dimensional free-energy contour plot that provides the more complete picture of the binding process. Since in our simulation temperature range Cdc42 was stable in its native state, we could treat it as a rigid folded state and consider only the folding and binding of CBD with the interface between CBD and Cdc42. Figure 3(a) shows the free energy (three-dimensional inset) and free-energy contour as a function of the fraction of native contacts of the binding interface and the fraction of native contacts of folding of CBD. The underlying free-energy landscape of the binding clearly shows distinct conformational states: a partially folded state ($Q_{f2} = 0.15$, where $0.15 = 15\%$ native) with no binding

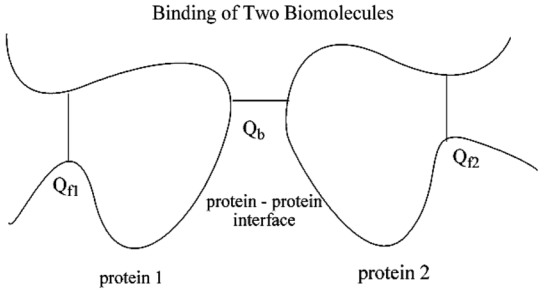

FIG. 2. Flexible binding-folding with a fraction of native contacts for folding of protein 1 Q_{f1} and protein 2 Q_{f2} and the binding interface between 1 and 2Q_b.

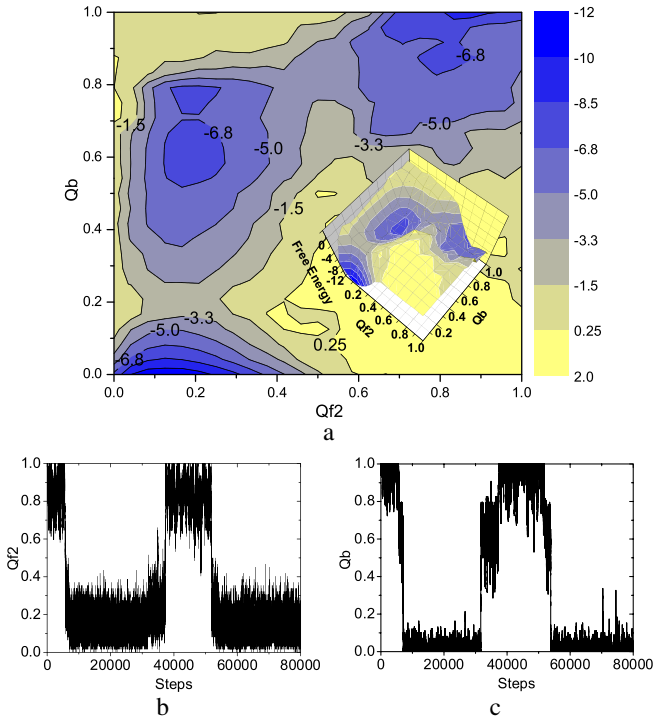

FIG. 3 (color online). (a) Free-energy contour profile, (b) time trajectories of the conformation/folding degree of freedom for CBD, and (c) the interface degree of freedom between the CBD portion of WASP and Cdc42.

($Q_b = 0$), a partially folded ($Q_{f2} = 0.2$) with partial native binding state ($Q_b = 0.6$), and a mostly folded ($Q_{f2} = 0.8$) with almost native binding state ($Q_b = 0.9$). This shows that folding and binding do not proceed independently (individually): The binding and folding are intimately coupled. It is more likely that the whole binding process proceeds first with a partial folding of CBD to a very limited extent (only 15% mostly through local folding), then to significant interface binding (60% without much further folding), and finally to binding and folding proceeding cooperatively to the native state. The experimental finding [16–18] of the dynamic fluctuations between the loosely bound and closely bound conformational states, when identified with our theoretical calculated free-energy minimum, corresponds to the cooperative binding-folding process or disorder-to-order transition of CBD upon binding. The cooperative binding-folding coupling and the inherent hydrophobic interactions lead to the formation of loosely bound and bound dual states and provide the basis and micro-origin of the conformational changes seen in the experiments [16–18]. The free-energy barrier separating these two basins is 2.7 kT. From the trajectories in time [Figs. 3(b) and 3(c)], we see the binding remains (Q_b and Q_{f2}) in the non-native partially folded and bound states, jumps to the native state, stays in the native state, and fluctuates back and forth together. This clearly shows the cooperativity between binding and folding.

Studying the transition state properties, especially the inhomogeneous distribution of contacts between residues, will help us better understand the mechanism of binding by locating the sites of nucleation for binding. The local parts of the free energy are perturbed by mutations, causing changes in both the equilibrium constant and the kinetics of binding. The ratio Φ is defined as the change of the kinetic rate and the equilibrium constant upon mutation at residue i. The ϕ values quantify the importance of certain residues in the transition state ensemble [24]. From experimental data, some ''hot residues'' with high Φ values can be identified in a transition state as they cluster together, indicating that the nucleus for binding can be identified.

Based on the landscape theory and the Go-model simulation with solvent effects taken into account, we have located the transition state from the free-energy profile as a function of the reaction coordinate ($Q_b = 0.8$, $Q_{f2} = 0.5$). We performed analysis on the ϕ values of Cdc42/CBD and its associated distribution to explore the nature of the binding transition state ensemble. For Cdc42/CBD, the binding seems to proceed like a nucleation process. The distribution of ϕ values in Fig. 4(a) shows prominent peaks near 1, indicating that there are certain residues with large ϕ values. These hot spots are crucial for the kinetic process and act as the nucleus or nucleation seeds of binding [see Fig. 4(b) for the theoretical ϕ values along the sequence positions of CBD portions of WASP and Cdc42]. Residues Phe37 and Phe56 for Cdc42 and Leu263 as well as Leu 267 for CBD have relatively high ϕ values. These findings are consistent with the limited mutational experiments done so far [12–14]. The theoretical predictions about ϕ values can be used to guide further experiments into which other hot residues could be chosen for study. Further mutational experiments on Cdc42/CBD are crucial in determining all of the hot residues crucial for flexible binding. The free-energy profiles and phi value predictions are influenced by the solvent. This seems to indicate that a small number of changes in the contact map may result in significant deviations in the free-energy profile, i.e., sensitivity to the contact map [25].

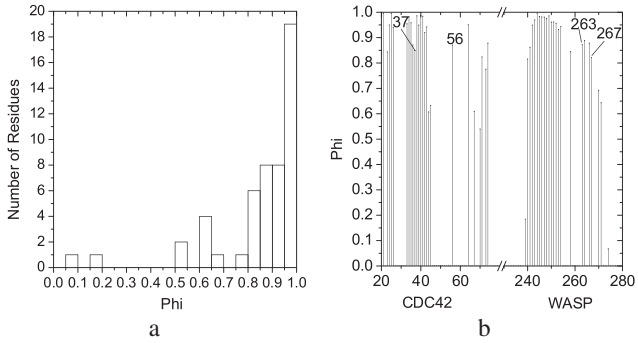

FIG. 4. Distribution of phi values (a) and phi values versus protein primary sequences (b) phi values along the primary sequences of the CBD portion of WASP and Cdc42.

In principle, there exist three possible scenarios for binding: (1) First folding and then binding, (2) binding and folding simultaneously, and (3) binding first and then folding. Scenario 1 is common and corresponds to the conventional thinking. There are some flexible binding complexes involved in signaling and gene regulations following scenario 2 [5]. Scenario 3 has never before been seen, but here we have shown the system to bind significantly before folding. Thus, this is the first study to realize scenario 3, combining both theory and experiment. Flexibility here is necessary in order to realize the biological function and specificity [4,20].

The simulations carried here are at higher concentrations (constrained center of mass distance) and higher temperatures than the biological ones to enhance sampling ($\Delta G \sim 0$, $K_d \sim 1$ M). The experimental measured K_d is in the range of μM [26]. Thus, our conclusion relies on the assumption that changing the effective temperature and confining the proteins does not qualitatively change the behavior or the underlying mechanism of the binding. This might be due to the possible existence of a universal phase boundary between monomers and oligomers in different ranges of conditions [27].

In this work, we explored flexible binding dynamics. The complex conformational fluctuations upon binding found in the single-molecule experiments directly reflects the topography of the binding landscape and reveals the cooperative nature of the binding-folding dynamics with large conformational changes that result in loosely bound and bound states [16–18]. Limited mutational experiments [12–14] show some hot spots crucial for flexible binding. The theoretical approach here can help to identify the hot residues and guide further experiments, and experiments can provide grounds for the theoretically constructed binding-energy landscape [28].

We thank Peter Wolynes, Jose Onuchic, Andy McCammon, Koby Levy, Sam Cho, Klaus Hahn, and Erich Vorpagel for helpful discussions. We acknowledge support from NSF, DOE, NSFC (China), and a computational pilot grant from PNNL.

*Electronic address: hplu@bgsu.edu
†Electronic address: jin.wang.1@stonybrook.edu

[1] E. Fischer, Ber. Dtsch. Chem. Ges. B **23**, 2611 (1890).
[2] D. E. Koshland, Jr., Proc. Natl. Acad. Sci. U.S.A. **44**, 98 (1958).
[3] A. K. Dunker *et al.*, Biochemistry **41**, 6573 (2002).
[4] B. A. Shoemaker, J. J. Portman, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **97**, 8868 (2000).
[5] Y. Levy, P. G. Wolynes, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **101**, 511 (2004).
[6] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).
[7] J. Wang and P. G. Wolynes, Phys. Rev. Lett. **74**, 4317 (1995).
[8] W. E. Moerner, Acc. Chem. Res. **29**, 563 (1996).
[9] H. P. Lu, L. Y. Xun, and X. S. Xie, Science **282**, 1877 (1998).
[10] L. Edman and R. Rigler, Proc. Natl. Acad. Sci. U.S.A. **97**, 8266 (2000).
[11] X. W. Zhuang, L. E. Bartley, H. P. Babcock, R. Russell, T. J. Ha, D. Herschlag, and S. Chu, Science **288**, 2048 (2000).
[12] N. Abdul-Manan, B. Aghazadeh, G. A. Liu, A. Majumdar, O. Ouerfelli, K. A. Siminovitch, and M. K. Rosen, Nature (London) **399**, 379 (1999).
[13] M. G. Rudolph, P. Bayer, A. Abo, J. Kuhlmann, I. R. Vetter, and A. J. Wittinghofer, Biol. Chem. **273**, 18 067 (1998).
[14] J. W. Erickson and R. A. Cerione, Curr. Opin. Cell Biol. **13**, 153 (2001).
[15] P. Nalbant, L. Hodgson, V. Kraynov, A. Toutchkine, and K. M. Hahn, Science **305**, 1615 (2004).
[16] X. Tan, P. Nalbant, A. Toutchkine, D. Hu, E. R. Vorpagel, K. M. Hahn, and H. P. Lu, J. Phys. Chem. B **108**, 737 (2004).
[17] H. Peter Lu, Acc. Chem. Res. **38**, 557 (2005).
[18] H. P. Lu, L. M. Iakoucheva, and E. J. Ackerman, J. Am. Chem. Soc. **123**, 9184 (2001).
[19] C. Levinthal, *Proceedings in Mossbauer Spectroscopy in Biological Systems* (University of Illinois, Urbana, 1969), p. 22.
[20] J. Wang and G. M. Verkhivker, Phys. Rev. Lett. **90**, 188101 (2003).
[21] N. Go, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).
[22] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).
[23] M. S. Cheung, A. E. Garcia, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **99**, 685 (2002).
[24] L. S. Itzhaki, D. E. Otzen, and A. R. Fersht, J. Mol. Biol. **254**, 260 (1995).
[25] H. Kaya and H. S. Chan, J. Mol. Biol. **315**, 899 (2002).
[26] L. Hemsath, R. Dvorsky, D. Fiegen, M. Carlier, and M. R. Ahmadian, Mol. Cell **20**, 313 (2005).
[27] S. Yang, H. Levine, and J. N. Onuchic, J. Mol. Biol. **352**, 202 (2005).
[28] H. J. Dyson and P. E. Wright, Curr. Opin. Struct. Biol. **12**, 54 (2002).