Quantifying the Kinetic Paths of Flexible Biomolecular Recognition

Jin Wang,*† Kun Zhang,* Hongyang Lu,* and Erkang Wang*
*State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin, China; and †Department of Chemistry and Department of Physics, State University of New York at Stony Brook, Stony Brook, New York

ABSTRACT Biomolecular recognition often involves large conformational changes, sometimes even local unfolding. The identification of kinetic pathways has become a central issue in understanding the nature of binding. A new approach is proposed here to study the dynamics of this binding-folding process through the establishment of a path-integral framework on the underlying energy landscape. The dominant kinetic paths of binding and folding can be determined and quantified. The significant coupling between the binding and folding of biomolecules often exists in many important cellular processes. In this case, the corresponding kinetic paths of binding are shown to be intimately correlated with those of folding and the dynamics becomes quite cooperative. This implies that binding and folding happen concurrently. When the coupling between binding and folding is weak (strong), the kinetic process usually starts with significant folding (binding) first, with the binding (folding) later proceeding to the end. The kinetic rate can be obtained through the contributions from the dominant paths. The rate is shown to have a bell-shaped dependence on temperature in the concentration-saturated regime consistent with experiment. The changes of the kinetics that occur upon changing the parameters of the underlying binding-folding energy landscape are studied.

INTRODUCTION

Biomolecular recognition is an important issue in modern molecular biology. Both affinity and specificity are crucial in understanding the nature of the binding process. The affinity is measured by the free energy difference between an unbinding complex and a native binding complex mimicking the natural attraction or tendency to form a stable complex. The specificity, on the other hand, measures the discrimination between the native state and others (1–6). In practice, good drug design needs to quantify both affinity and specificity accurately (7–12). The general docking approach assumes rigid or semirigid conformations of the binding complex for the purpose of simplifying the computation of searching for conformational degrees of freedom. In other words, the binding proceeds with two already (or almost) folded proteins. In nature, binding often involves conformational changes. Therefore, the binding mechanism should be more like induced-fit rather than lock-and-key in nature (13,14). In cells, it was estimated from the current available data that ~15% of the proteins when isolated are in their unfolded form. The actual percentage might be even higher. This indicates that some natural proteins prefer to be unfolded in the cell environment in order to function (15,16). It implies that significant binding occurs as the binding and folding cooperatively couple together. That is, binding proceeds from the two unfolded or partially folded proteins instead of two already-formed ones. Here, flexibility rather than rigidity is crucial for binding as well as for biological function.

To understand the interplay between binding and folding dynamically, one needs first to have a good description of the binding and folding degrees of freedom. One way to do that is to employ atomic detailed calculations. This way of doing it often runs into the trouble of not being able to sample enough of the configurational space. However, it is possible to use a phenomenological approach by identifying the quasi-reaction coordinate or order parameter mimicking the binding and folding process. This is in analogy to the liquid-gas phase transition. Molecular-level calculations give more detail, whereas the phenomenological use of density as an order parameter reveals the universal features of the liquid-to-gas phase transition (17–19). In a similar spirit, the approach used here attempts to study the binding-folding phenomena common in nature with at least two order parameters, Q_b and Q_f. Here, Q_b is the fraction of native binding spatial contacts and Q_f is the fraction of the native folding spatial contacts. $Q_b = 1$ when the binding is completed; that is, all the native interface contacts are formed. On the other hand, $Q_f = 1$ represents the situation where all the native contacts of folding are formed (see Fig. 1) from binding of two proteins where one protein is rigid but the interface and the other protein is flexible. This minimal representation is used to study the thermodynamics of binding-folding process (15,16). It is found that the folding and binding processes are often intimately coupled in nature. The crucial question one needs to address is how the dynamics actually occur. This is not only relevant for uncovering the fundamental mechanism but important also in guiding more accurate rational drug design.

Submitted September 19, 2005, and accepted for publication February 6, 2006.

Address reprint requests to Erkang Wang, E-mail: ekwang@ciac.jl.cn.

0006-3495/06/08/866/07 $2.00 doi: 10.1529/biophysj.105.074716

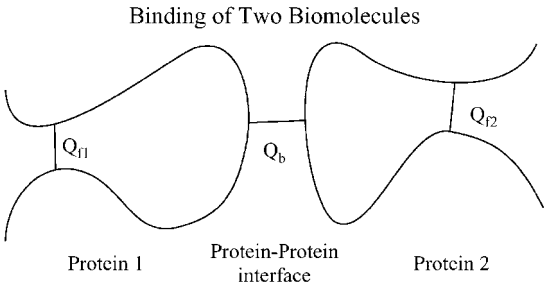

FIGURE 1 Flexible binding-folding.

Identifying the important dynamic flow of paths that the binding complex takes to reach the native state is crucial in uncovering the fundamental kinetic mechanisms of the binding-folding process and has been a central issue in the experimental community (20,21). So far, very limited efforts have been put on the actual kinetic binding intermediate process or the identification of kinetic paths connecting the initial and final states (22–29). We will study the dynamics of binding-folding coupling by developing a path-integral formulation. Path-integral formulations have been developed successfully in studying many different areas in physics, chemistry, and even finance (30–34). The advantage of this approach is that it addresses the fundamental issues of kinetic pathways directly. The paths can be identified and quantitatively determined (See Fig. 2).

Another important question is related to how the many possible degrees of configuration could fall to the unique native-state basin. The most natural and simple way of resolving this so-called Levinthal paradox (35) is that the underlying energy landscape should be funneled to guarantee both the thermodynamic stability and specificity (1–6,36–39). This should also lead to faster kinetics (40–42).

Under this funneled energy landscape, in general there are multiple paths or a dominant flow of paths toward the native-state basin. The discrete paths emerge when the landscape is rough and local bumps or traps play important roles. Thus, the kinetics can be obtained through studying the behavior of the paths. By approximating the path integral using a dominant-path approach, we will describe with realistic parameters estimated from the current available data the dynamics and the degrees of cooperativity in the binding-folding process. The kinetic rate or timescale can be obtained by summing over the appropriate weighted contributions from the dominant paths. The rate is shown to have a bell-shaped dependence on the temperature in the concentration-saturated regime. This is consistent with the kinetic experiments studies on reaction with large conformational changes (see Chevron plots, 43–45).

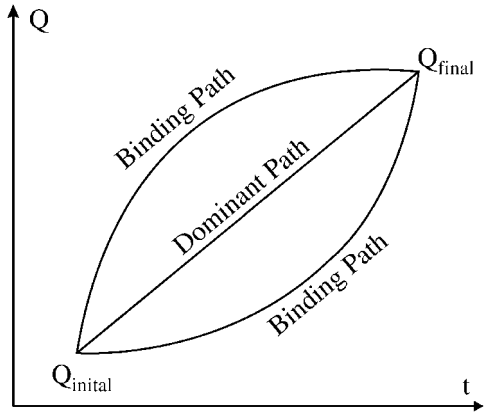

FIGURE 2 Binding paths.

METHODS AND MATHEMATICAL DETAILS

To proceed, let us briefly review the formulation of the thermodynamic energy landscape of the binding-folding process (6,15,16). We first start with the energy function of a polypeptide chain with an interface,

$$H^{\mathrm{f}} = \sum \epsilon_{\mathrm{ij}}^{\mathrm{f}} \sigma_{\mathrm{ij}}^{\mathrm{f}} \tag{1}$$

and

$$H^{\mathrm{b}} = \sum \epsilon_{\mathrm{ij}}^{\mathrm{b}} \sigma_{\mathrm{ij}}^{\mathrm{b}} \tag{2}$$

and total energy is

$$H = H^{\mathrm{f}} + H^{\mathrm{b}}, \tag{3}$$

where H^{f} and H^{b} are the energy functions of polypeptide chain energy (folding) and interface contacts (binding). The ϵ_{ij} values are the contact-energy strengths, while σ_{ij} is the contact variable equal to 1 when there is a spatial contact, and zero when there is no spatial contact (a spatial contact is controlled by the cutoff distance within which a contact is defined).

If we take a reference state **a**, and split up the energy as a point E_{f} associated with the folding degree of freedom, a point E_{b} associated with the binding degree of freedom and a reference energy E_{a} is

$$E_{\mathrm{f}} = E_{\mathrm{af}} + E'_{\mathrm{f}}, \tag{4}$$

$$E_{\mathrm{b}} = E_{\mathrm{ab}} + E'_{\mathrm{b}}, \tag{5}$$

$$E_{\mathrm{a}} = E_{\mathrm{af}} + E_{\mathrm{ab}} + E'_{\mathrm{a}}, \tag{6}$$

where E_{af} is part of the overlap energy of folding relative to the reference state a and E_{bf} is part of the overlap energy of binding relative to the reference state **a**. The conditional probability in finding the folding energy and binding energy in reference to state **a** is given by

$$P(E_{\mathrm{a}}|E_{\mathrm{f}}, E_{\mathrm{b}}) = \frac{\langle \delta(E_{\mathrm{a}} - H(Q_{\mathrm{f}}, Q_{\mathrm{b}}))\delta(E_{\mathrm{f}} - H(Q_{\mathrm{f}}))\delta(E_{\mathrm{b}} - H(Q_{\mathrm{b}})\rangle}{\delta(E_{\mathrm{f}} - H(Q_{\mathrm{f}}))\delta(E_{\mathrm{b}} - H(Q_{\mathrm{b}}))}. \tag{7}$$

If reference state **a** is the native state, then the Q_{f} and Q_{b} become the fraction of native folding and binding contacts, respectively. The entropy can be obtained as

$$S(E_{\mathrm{a}}, Q_{\mathrm{f}}, Q_{\mathrm{b}}) = ln(\Omega(Q_{\mathrm{f}}, Q_{\mathrm{b}})P(E_{\mathrm{a}}|E_{\mathrm{f}}, E_{\mathrm{b}})), \tag{8}$$

where $\Omega(Q_{\mathrm{f}}, Q_{\mathrm{b}})$ is the total number of configurational degrees of freedom at a given Q_{f} and Q_{b}. Therefore, by employing the microcanonical ensemble and thermodynamic relationships, we can easily obtain the free energy of the system,

$$F(Q_f, Q_b) = N\delta E_f Q_f + CN\delta E_b Q_b - NS_0(Q_f, Q_b)T - N\frac{\Delta E_f^2(1-Q_f)(1+\gamma_f Q_f)}{2\,k_B T} - CN\frac{\Delta E_b^2(1-Q_b)(1+\gamma_b Q_b)}{2\,k_B T}, \quad (9)$$

where N is the number of the amino-acid residues, δE_f is the energy gap or bias toward the native folded state, δE_b is the energy gap or bias toward the native binding state, ΔE_f is the roughness or spread of the folding energy, and ΔE_b is the roughness or spread of the binding energy. The values γ_f and γ_b are the inhomogeneity coefficients for folding and binding. The value S_0 is the entropy of the configurations $S_0 = ln\Omega$. The value C is a constant scale factor of the binding relative to folding.

The entropy function can be fitted with a simple function by noticing that the entropy of the completely native folding and binding state $S(1,1)$ is zero and the entropy of the completely unfolded and unbinding state is $S(0,0)$, the entropy of native folded but completely unbinding state is $S(1,0)$, and the entropy for the completely unfolded and native binding state is $S(0,1)$. These quantities can all be estimated. So the entropy has a functional form given by

$$S(Q_f, Q_b) = (1-Q_f)(1-Q_b)S(0,0) + Q_f(1-Q_b)S(1,0) + Q_b(1-Q_f)S(0,1). \quad (10)$$

The binding-folding energy landscape typically has several phases: the total native phase, native binding but unfolded phase, folded but unbinding phase, and completely unbinding and unfolded phase. In addition, there might exist a possible trapping phase for the whole complex, as well as a partial trapping phase for folded states alone, and a partial trapping phase for the binding states alone. To guarantee thermodynamic stability and discriminate between the local bumps or traps (the specificity), the temperature of the transition to a stable native state should be significantly larger than the trapping temperature. Since the ratio between native phase transition temperature and glassy trapping temperature, $T_{native}/T_g = \sqrt{\Lambda}+\sqrt{\Lambda-1}$, monotonically depends on the ratio of the gap/roughness of the underlying binding-folding energy landscape $((\Lambda = \delta E_f + \delta E_b/\sqrt{\Delta E_f^2 + \Delta E_b^2}\sqrt{2S0}))$, the gaps should be significantly larger than the spread of the energy spectrum. In other words, $(\delta E_f + \delta E_b/\sqrt{\Delta E_f^2 + \Delta E_b^2})$ should be significantly larger than 1. This implies the landscape in the two dimensions of folding and binding should be funneled toward the native state. Let us turn to the discussion now to the kinetics.

Under the free-energy profiles, the equation of motion for native contact vector $\mathbf{Q} = (Q_f, Q_b)$ formation can be formulated as

$$d\mathbf{Q}/dt = -\frac{\partial\beta\mathbf{F}(\mathbf{Q})}{\partial\mathbf{Q}} + \eta. \quad (11)$$

Due to the long timescale, the binding-folding motions are overdamped. Therefore, the second derivatives of $\mathbf{Q}$ with respect to time t may be ignored. Here, $-\partial\beta\mathbf{F}(\mathbf{Q})/\partial\mathbf{Q}$ is the gradient force that the motion of $\mathbf{Q}$ vector would follow and η is the noise term assumed to be Gaussian and white. The correlation of the noise is given by $\langle\eta(\mathbf{Q}, t)\eta(\mathbf{Q}, 0)\rangle = 2\mathbf{D}(\mathbf{Q})\delta(t)$. The $\mathbf{D}(\mathbf{Q})$ is the $\mathbf{Q}$-dependent diffusion coefficient tensor (or matrix). The binding-folding process has many degrees of freedom; therefore, when looking at the motion along the reduced two-dimensional order parameter or reaction coordinate $\mathbf{Q}$, there is an effective noise or friction force from the rest of the other dimensions.

We can now formulate the dynamics for the probability of starting from initial configuration $Q_{initial}$ at $t = 0$ and end at the final configuration of Q_{final} at time t, with the Onsager-Machlup functional (32,33) as

$$P(\mathbf{Q}_{final}, t, \mathbf{Q}_{initial}, 0) = \int D\mathbf{Q}\mathrm{Exp}\left[-\int dt\left(\frac{1}{4}\frac{\left(\frac{d\mathbf{Q}}{dt} + \frac{\mathbf{D}(\mathbf{Q})\partial\beta F(\mathbf{Q})}{\partial\mathbf{Q}}\right)^2}{\mathbf{D}(\mathbf{Q})} - \frac{1}{2}\frac{\partial\mathbf{D}(\mathbf{Q})}{\partial\mathbf{Q}}\frac{\beta F(\mathbf{Q})}{\partial\mathbf{Q}}\right)\right] = \int D\mathbf{Q}\mathrm{Exp}\left[-\int L(\mathbf{Q}(t))dt\right]. \quad (12)$$

The integral over $D\mathbf{Q}$ represents the sum over all possible paths connecting $\mathbf{Q}_{initial}$ at time $t = 0$ to $\mathbf{Q}_{final}$ at time t. The exponential factor gives the weight of each path. So the probability of binding-folding dynamics from nonnative configurations $\mathbf{Q}_{initial}$ to native configuration $\mathbf{Q}_{final}$ is equal to the sum of all possible paths with different weights. The $L(\mathbf{Q}(t))$ is the Lagrangian or the weight for each path (Fig. 2).

Notice that not all the paths give the same contribution. We can approximate the path integrals with a set of dominant paths. Since each path is exponentially weighted, the other subleading path contributions are often small and can be ignored. One can easily use this observation to find the paths with the optimal weights. The dominant paths should satisfy the Euler-Lagrangian equation (see Fig. 2),

$$\frac{d}{dt}\frac{\partial L}{\partial\dot{\mathbf{Q}}} - \frac{\partial L}{\partial\mathbf{Q}} = 0, \quad (13)$$

and the resulting equation becomes

$$\ddot{\mathbf{Q}} - \frac{1}{2}\frac{\frac{\partial\mathbf{D}(\mathbf{Q})}{\partial\mathbf{Q}}}{\mathbf{D}(\mathbf{Q})}\dot{\mathbf{Q}}^2 - 2\mathbf{D}(\mathbf{Q})\frac{\partial V(\mathbf{Q})}{\partial\mathbf{Q}} = 0, \quad (14)$$

where

$$V(\mathbf{Q}) = \frac{\partial\beta F(\mathbf{Q})}{\partial\mathbf{Q}}\frac{\mathbf{D}(\mathbf{Q})}{4}\frac{\partial\beta F(\mathbf{Q})}{\partial\mathbf{Q}} - \frac{\mathbf{D}(\mathbf{Q})}{2}\frac{\partial^2\beta F(\mathbf{Q})}{\partial\mathbf{Q}^2} - \frac{1}{2}\frac{\partial\mathbf{D}(\mathbf{Q})}{\partial\mathbf{Q}}\frac{\partial\beta F(\mathbf{Q})}{\partial\mathbf{Q}}. \quad (15)$$

The equation of motion of Q has the acceleration term $\ddot{Q}$, the frictional (positive and negative) term $\frac{1}{2}((\partial\mathbf{D}(\mathbf{Q})/\partial\mathbf{Q})/\mathbf{D}(\mathbf{Q}))\mathbf{Q}^2$, and the force term $2\mathbf{D}(\mathbf{Q})(\partial\mathrm{V}(\mathbf{Q})/\partial\mathbf{Q})$. Define $-\partial U(\mathbf{Q})/\partial\mathbf{Q} = 2\mathbf{D}(\mathbf{Q})\partial V(\mathbf{Q})/\partial\mathbf{Q}$. Then the problem becomes one of a two-dimensional particle moving in a potential well U with friction.

When $\mathbf{D}(\mathbf{Q})$ is a constant, the friction term is zero. The diffusion coefficient tensor matrix is diagonal, with only two elements (D_{ff} and D_{bb}) present, while the nondiagonal elements are zero ($D_{fb} = D_{bf} = 0$).

We can also write out explicitly the equation of motion in the scalar form as

$$\ddot{Q}_f - 2D_{ff}\frac{\partial V}{\partial Q_f} = 0, \quad (16)$$

and

$$\ddot{Q}_b - 2D_{bb}\frac{\partial V}{\partial Q_b} = 0. \quad (17)$$

Notice that frictional term becomes zero under the current assumption of a Q-independent diffusion coefficient.

By solving these two equations with initial points of $Q_f = Q_b = 0$ and end points at $Q_f = Q_b = 1$, we can obtain the dominant path contribution to the weight of the paths. Substituting the dominant path solution back into the path-integral formulation, we can obtain the expression for the rate of the kinetic process from nonnative states to native state.

RESULTS

Through data mining of the protein interaction database, one can classify protein-protein binding complex into two classes

(15,16). One type is related to the more stable proteins while isolated, and the other type is related to the less stable proteins, or often unfolded when isolated. It is estimated that 15% of the proteins are floppy and unstable when isolated at room temperature (15,16). Our choices of the parameters (especially for the stability parameter: *energy gap*) are as follows: Set-I parameters are for the average proteins that are often more stable; and Set-II parameters are for the more stable binding interactions with less stable and more floppy proteins. That is,

$$\delta E_{\rm f} = -11.8\frac{\rm kJ}{\rm mol}, \quad \delta E_{\rm b} = -9.3\frac{\rm kJ}{\rm mol} \qquad \text{(Set I)}$$

and

$$\delta E_{\rm f} = -10.3\frac{\rm kJ}{\rm mol}, \quad \delta E_{\rm b} = -12.9\frac{\rm kJ}{\rm mol}. \qquad \text{(Set II)}$$

The energy scale factor of binding relative to folding is also given as $C = 0.2$.

The other related parameters are the same for both Set I and Set II:

$$\gamma_{\rm f} = 1.0, \gamma_{\rm b} = 0, S(0,0) = 3 + \frac{9.7}{200}k_{\rm B}, S(1,0) = \frac{9.7}{200}k_{\rm B},$$
$$S(0,1)/S(0,0) = 0.75, \sqrt{\Delta E_{\rm f}} = 3.4\frac{\rm kJ}{\rm mol}, \sqrt{\Delta E_{\rm b}} = 3.4\frac{\rm kJ}{\rm mol}. \qquad (18)$$

The free energy F as a function of $Q_{\rm f}$ and $Q_{\rm b}$ as well as the dominant kinetic paths are shown in Fig. 3 for the parameter Set I (Fig. 3, *left panel*) and Set II (Fig. 3, *right panel*). We can see that the underlying landscape is downhill and funneled toward the native state.

We can see clearly that for the parameter Set I (the more stable proteins, or the ones with a larger folding gap compared with the binding gap—i.e., more stability on folding and less stability on binding), the kinetic process proceeds with a significant fraction of folding initially and then proceeds with the completion of the binding process (Fig. 3, *left panel*). The folding and binding are not very strongly coupled. But with the parameter Set II (the floppy proteins, or the ones of smaller folding gap compared with the binding gapless stability on folding and more stability on binding), we see that significant binding occurs first and then proceeds with folding and binding together toward the native state (Fig. 3, *right panel*). So in this case, the folding and binding process are more cooperatively coupled (15,16,46,47).

The effect of temperature on kinetics can be seen from the change of rate. The rate is plotted in Fig. 4 versus temperature. The rate is shown to have a bell-shape. At high temperatures, as the temperature increases, the native state is unstable so the kinetic rate decreases. On the other hand, at low temperatures, when the temperature decreases, there exists the possibility of local trapping, so the rate decreases again. This explains why the rate has a bell-like shape. There exists an optimal rate at a certain temperature between high and low temperatures where the kinetic process is the fastest. When the gap of either folding or binding increases, the kinetic rate increases, as we can see. This is due to the greater bias toward the native-state basin. We can also see that the kinetic process is faster for flexible binding (with the more-stable binding (gap) rather than the folding (gap)). The more-stable folding (gap) implies that the binding process starts first with significant folding-complex forming and then proceeds with binding; it is basically more of a rigid-binding process. In contrast, the more-stable binding (gap) implies that significant binding starts first and induces the folding. In other words, binding and folding are intimately coupled. So we have shown here that the flexible binding (binding-folding

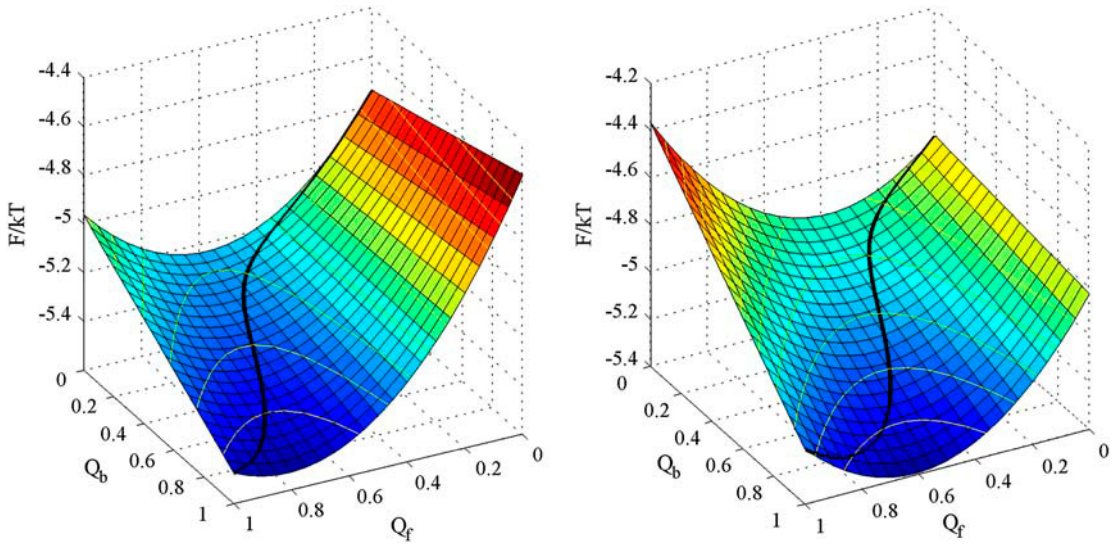

FIGURE 3 (*Left*) Free-energy profile and dominant kinetic paths with respect to $Q_{\rm f}$ and $Q_{\rm b}$ for more stable proteins. (*Right*) Free-energy profile and dominant kinetic paths with respect to $Q_{\rm f}$ and $Q_{\rm b}$ for more floppy proteins.

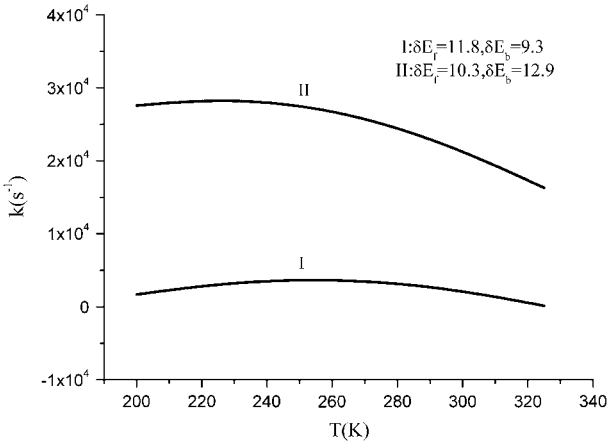

FIGURE 4 The kinetic rate of binding-folding versus temperature.

coupled) has a kinetic advantage (faster) over rigid binding (folding first and then binding). Binding with large conformational changes helps to reach the kinetic specificity rather than the rigid one. This is due to the larger capture radius for the flexible binding. It is analogous to fly-casting in fishing (15,16). The recent simulations and experiments seem to imply that this mechanism might be quite general for the flexible binding (48–52).

It is worth pointing out the binding involves two molecules and the reaction is bimolecular. Thus, the kinetics, in general, is concentration-dependent, and directly proportional to the concentration in the low concentration limit. Here we study the kinetic from nonspecific unbinding unfolded states to the native state. It is likely we are looking at the high concentration regime, where the concentrations already have become saturated, and thus are constants—so the kinetic rate we obtained here should be considered valid only at the concentration-saturation limit. To obtain the kinetics in the full range of concentrations, one has to carefully take into account the concentration-dependence of the temperatures. This would be an interesting topic worth further investigation.

We also have studied the influence of diffusion on the kinetic paths (*left panel*, Fig. 5) and kinetic rates (*right panel*, Fig. 5) with parameter Set II of floppy binding proteins. We can see that the dominant kinetic paths will be shifted more toward initial binding (folding) first and then proceeds with folding (binding) when the diffusion for binding is faster (slower), i.e., the diffusion coefficients of binding increases (decreases). The corresponding binding kinetics is faster for a faster, or a larger, binding-diffusion coefficient. We also studied the case of varying the folding diffusion coefficients, and found that dominant kinetic paths will be shifted more toward initial folding (binding) first and then proceed with binding (folding) when the diffusion for folding is faster (slower). The corresponding binding kinetics is faster for faster diffusion.

DISCUSSION AND CONCLUSIONS

Through the establishment of a two-dimensional path-integral formulation, we studied the kinetics on a binding-folding energy landscape. By quantitatively determining the kinetic paths, we find that a binding-folding mechanism can be broken into several categories, with one in which folding proceeds first and binding-folding proceeds later; and another, in which the binding and folding processes are dynamically coupled.

It is worthwhile to point out that the two-dimensional binding-folding coupling we considered here assumes one of the binding partners is rigid so that we only need to consider essentially two degrees of freedom: the binding interface and folding or conformational changes of the other partner (Q_{f1}, Q_b), as illustrated in Fig. 1. In general, one should consider the situation of both partners being flexible and binding to

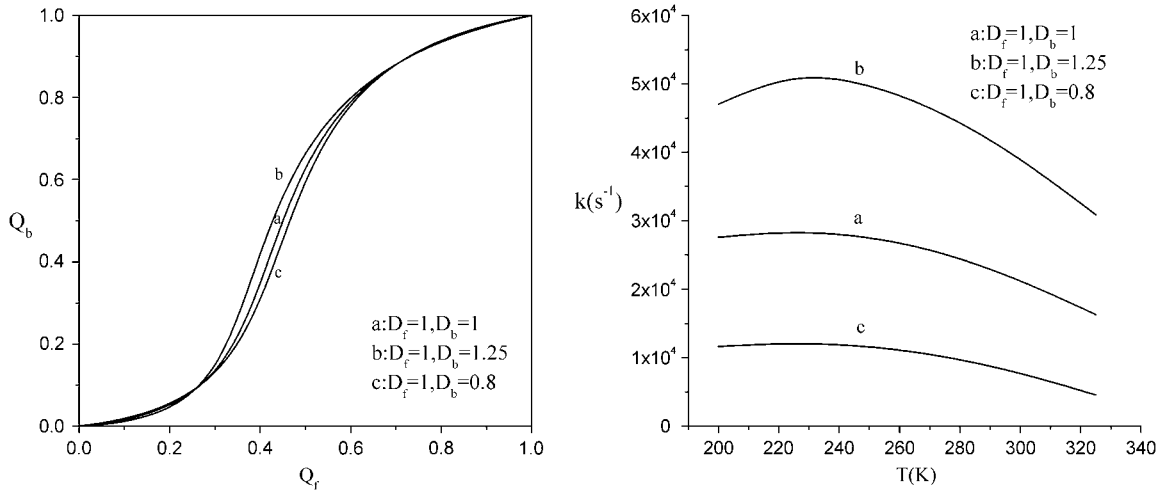

FIGURE 5 (*Left*) Dominant kinetic paths. (*Right*) Kinetic rates for different binding diffusion coefficients.

each other, so that three degrees of freedom need to be considered (Q_{f1}, Q_{f2}, Q_b) (Fig. 1). We will first have to establish the thermodynamic free energy profiles either from simulations or by developing an analytical theory. Then we will construct the corresponding path-integral formulation for the case. This study is in progress. It will be an interesting extension of the current framework, which will be discussed in a future publication.

For simplicity, we have ignored the position dependence of the diffusion coefficient in obtaining the kinetic rate. The Q-dependence of diffusion coefficient D due to the size of the configurational space to explore at a particular position could have a significant influence on the quantitative values of the rate. It will lead to an effective free energy so that the actual barrier for kinetics will deviate from the thermodynamic barrier. This is worth further study.

On the other hand, we did not take into account the possibility of the time-dependence or memory of the diffusion coefficient due to the correlations in energies among different states. In principle, one can put that in. Then the transport properties become time-correlated. This correlation in time of the diffusion coefficient would also contribute to the effective free energy making the corner, cutting without passing through the actual thermodynamic barrier.

It is worthwhile to explore further the influence of the nature of the hydrophobic multibody force on kinetics. This can be implemented with the formalism established in this article. In that situation, the resulting free-energy profile will most likely have free-energy valleys and barriers. It is expected that the dominant paths at long times will be, most likely, the instantons traversing back and forth between native and nonnative states.

In addition, at low temperatures, one expects that the current continuous description of paths breaks down. Instead, a discrete version of the path integral needs to be developed for treating this regime.

J.W. thanks Prof. Peter Wolynes, Prof. Jose Onuchic, Prof. Andy McCammon, Prof. George Stell, and Dr. Koby Levy for helpful discussions.

K.Z., H.L., and E.K.W. are supported by the National Science Foundation of China. J.W. is supported by the National Science Foundation Career Award (USA), the Petroleum Research Fund, and the K.C. Wong Foundation Research Award.

REFERENCES

1. Rejto, P. A., and G. M. Verkhivker. 1996. A mean field model of ligand-protein interactions: implications for the structural assessment of human immunodeficiency virus type 1 protease complexes and receptor-specific binding; unraveling principles of lead discovery: from unfrustrated energy landscapes to novel molecular anchors. *Proc. Natl. Acad. Sci. USA.* 93:60–64, 8945–8950.
2. Janin, J. 1996. Quantifying biological specificity: the statistical mechanics of molecular recognition. *Proteins.* 25:438–445.
3. Tsai, C. J., S. Kumar, B. Ma, and R. Nussinov. 1999. Folding funnels, binding funnels and protein function. *Protein Sci.* 8:1181–1190.
4. Tovchigrechko, A., and I. A. Vakser. 2001. How common is the funnel-like energy landscape in protein-protein interactions? *Protein Sci.* 10:1572–1583.
5. Gerland, U., J. D. Moroz, and T. Hwa. 2002. Physical constraints and functional characteristics of transcription factor-DNA interaction. *Proc. Natl. Acad. Sci. USA.* 99:12015–12020.
6. Wang, J., and G. M. Verkhivker. 2003. Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. *Phys. Rev. Lett.* 90:188101–188104.
7. Cohen, N. C. 1996. Molecular Modeling in Design. Academic Press, New York.
8. Wlodawer, A., and J. W. Erickson. 1993. Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* 62:543–585.
9. Clackson, T., and J. A. Wells. 1995. A hot-spot of binding energy in a hormone-receptor interface. *Science.* 267:383–386.
10. Cherfils, J., and J. Janin. 1993. Protein docking algorithms: simulating molecular recognition. *Curr. Opin. Struct. Biol.* 3:265–269.
11. Oshiro, C. M., I. D. Kuntz, and J. S. Dixon. 1995. Flexible ligand docking using a genetic algorithm. *J. Comput. Aided Mol. Des.* 9:113–130.
12. Gallop, M. A., R. W. Barrett, W. J. Dower, S. P. A. Fodor, and E. M. Gordon. 1994. Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries. *J. Med. Chem.* 37:1233–1251, 1385–1401.
13. Koshland, D. E., Jr. 1958. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA.* 44:98–104.
14. McCammon, J. A. 1998. Theory of biomolecular recognition. *Curr. Opin. Struct. Biol.* 8:245–249.
15. Shoemaker, B. A., J. J. Portman, and P. G. Wolynes. 2000. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. USA.* 97:8868–8873.
16. Papoian, G. A., and P. G. Wolynes.2003. The physics and bioinformatics of binding and folding—an energy landscape perspective. *Biopolymers.* 68:333–349.
17. Weeks, J. D., D. Chandler, and H. C. Andersen. 1971. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* 54:5237.
18. Chandler, D., and H. C. Andersen. 1972. Optimized cluster expansions for classical fluids. *J. Chem. Phys.* 57:1930.
19. Lowden, L. J., and D. Chandler. 1974. Theory of intermolecular pair correlations for molecular liquids. *J. Chem. Phys.* 61:5228.
20. Frauenfelder, H., F. Parak, and R. D. Young. 1988. Conformational substates in proteins. *Annu. Rev. Biophys. Biophys. Chem.* 17:451–479.
21. Frauenfelder, H., S. G. Sligar, and P. G. Wolynes. 1991. The energy landscapes and motions of proteins. *Science.* 254:1598–1603.
22. Berkowitz, M., and J. A. McCammon. 1981. Brownian motion of a system of coupled harmonic oscillators. *J. Chem. Phys.* 75:957–961.
23. Berkowitz, M., and J. A. McCammon. 1982. Molecular dynamics with stochastic boundary conditions. *Chem. Phys. Lett.* 90:215–217.
24. Berkowitz, M., J. D. Morgan, and J. A. McCammon. 1981. Memory kernels from molecular dynamics. *J. Chem. Phys.* 75:2462.
25. Berkowitz, M., J. D. Morgan, and J. A. McCammon. 1983. Generalized Langevin dynamics simulations with arbitrary time-dependent memory kernels. *J. Chem. Phys.* 78:3256.
26. Berkowitz, M., J. D. Morgan, J. A. McCammon, and S. H. Northrup. 1983. Diffusion-controlled reactions. A variational formula for the optimum reaction coordinate. *J. Chem. Phys.* 79:5563–5565.
27. McCammon, J. A., and S. H. Northrup. 1983. Saddle-point avoidance in diffusional reactions. *J. Chem. Phys.* 78:987.
28. Olender, R., and R. Elber. 1996. Calculation of classical trajectories with a very large time step: formalism and numerical examples. *J. Chem. Phys.* 105:9299–9315.
29. Elber, R., J. Meller, and R. Olender. 1999. Stochastic path approach to compute atomically detailed trajectories: application to the folding of C peptide. *J. Phys. Chem. B.* 103:899.

30. Wiener, N. 1964. Generalized Harmonic Analysis and Tauberian Theorems. MIT Press, Boston, MA.

31. Feynman, R. P., and A. R. Hibbs. 1965. Quantum Mechanics and Path Integrals. McGraw-Hill, New York.

32. Onsager, L., and S. Machlup. 1953. Fluctuations and irreversible processes. *Phys. Rev.* 91:1505–1512.

33. Hanggi, P. 1989. Path integral solutions for non-Markovian processes. *Z. Phys. B.* 75:275–281.

34. Hunt, K. L. C., and J. Ross. 1981. Path integral solutions of stochastic equations for nonlinear irreversible processes: the uniqueness of the thermodynamic Lagrangian. *J. Chem. Phys.* 75:976.

35. Levinthal, C. 1969. Proceedings in Mossbauer spectroscopy in biological systems. University of Illinois Press, Urbana. IL:22.

36. Goldstein, R. A., Z. A. Luthey-Schulten, and P. G. Wolynes. 1992. Optimal protein-folding codes from spin-glass theory. *Proc. Natl. Acad. Sci. USA.* 89:4918–4922.

37. Abkevich, V. I., A. M. Gutin, and E. I. Shakhnovich. 1994. Free energy landscape for protein folding kinetics: intermediates, traps, and multiple pathways in theory and lattice model simulations. *J. Chem. Phys.* 101:6052–6062.

38. Klimov, D. K., and D. Thirumalai. 1998. Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics. *J. Chem. Phys.* 109:4119–4125.

39. Tsai, C. J., D. Xu, and R. Nussinov. 1998. Protein folding via binding, and vice versa. *Folding Des.* 3:R71–R80.

40. Lee, C. L., G. Stell, and J. Wang. 2003. First passage time distribution and non-Markovian dynamics of protein folding. *J. Chem. Phys.* 118: 959–968.

41. Lee, C. L., G. Stell, and J. Wang. 2003. Diffusion dynamics, moments and distributions of first passage time on protein folding energy landscapes, with applications to single molecules *Phys. Rev. E.* 67: 41905.

42. Zhou, Y., C. Zhang, G. Stell, and J. Wang. 2003. Temperature dependence of the distribution of the first passage time, results from discontinuous molecular dynamics simulations of an all-atom model of the second β-hairpin fragment of protein G. *J. Am. Chem. Soc.* 125: 6300–6305.

43. Kaya, H., and H. S. Chan. 2000. Energetic components of cooperative protein folding. *Phys. Rev. Lett.* 85:4823–4826.

44. Kaya, H., and H. S. Chan. 2002. Towards a consistent modeling of protein thermodynamic and kinetic cooperativity: how applicable is the transition state picture to folding and unfolding? *J. Mol. Biol.* 315: 899–908.

45. Kuhlman, B., D. L. Luisi, P. A. Evans, and D. P. Raleigh. 1998. Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9. *J. Mol. Biol.* 284:1661–1670.

46. Plotkin, S. S., J. Wang, and P. G. Wolynes. 1996. A correlated energy landscape model for finite, random heteropolymers. *Phys. Rev. E.* 53: 6271–6296.

47. Plotkin, S. S., J. Wang, and P. G. Wolynes. 1997. Statistical mechanics of a correlated energy landscape model for protein folding funnels. *J. Chem. Phys.* 106:2932–2948.

48. Levy, Y., J. N. Onuchic, and P. G. Wolynes. 2004. Protein topology determines binding mechanism *Proc. Natl. Acad. Sci. USA.* 101: 511–516.

49. Levy, Y., S. S. Cho, J. N. Onuchic, and P. G. Wolynes. 2005. A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J. Mol. Biol.* 346:1121–1145.

50. Jana, R., T. R. Hazbun, A. K. M. M. Mollah, and M. C. Mossing. 1997. A folded monomeric intermediate in the formation of λ-cro dimer-DNA complex. *J. Mol. Biol.* 273:402–416.

51. Gloss, L. M., and C. R. Matthews. 1998. The barriers in the bimolecular and unimolecular folding reactions of the dimeric core domain of *Escherichia coli* Trp repressor are dominated by enthalpic contributions. *Biochemistry.* 37:16000–16010.

52. Rousseau, F., J. W. H. Schymkowitz, and L. S. Itzhaki. 2003. The unfolding story of three-dimensional domain swapping. *Structure* (*Lond.*). 11:243–251.