Theory and simulation on the kinetics of protein–ligand binding coupled to conformational change

Lu Cai[1] and Huan-Xiang Zhou[2,a)]
[1]*Department of Polymer Science and Engineering, CAS Key Laboratory of Soft Matter Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China*
[2]*Department of Physics and Institute of Molecular Biophysics, Tallahassee, Florida 32306, USA*

(Received 21 November 2010; accepted 12 February 2011; published online 8 March 2011)

Conformational change during protein–ligand binding may significantly affect both the binding mechanism and the rate constant. Most earlier theories and simulations treated conformational change as stochastic gating with transition rates between reactive and nonreactive conformations uncoupled to ligand binding. Recently, we introduced a dual-transition-rates model in which the transition rates between reactive and nonreactive conformations depend on the protein–ligand distance [H.-X. Zhou, Biophys. J. **98**, L15 (2010)]. Analytical results of that model showed that the apparent binding mechanism switches from conformational selection to induced fit, when the rates of conformational transitions increase from being much slower than the diffusional approach of the protein–ligand pair to being much faster. The conformational-selection limit (k_{CS}) and the induced-fit limit (k_{IF}) provide lower and upper bounds, respectively, for the binding rate constant. Here we introduce a general model in which the energy surface of the protein in conformational space is coupled to ligand binding, and present a method for calculating the binding rate constant from Brownian dynamics simulations. Analytical and simulation results show that, for an energy surface that switches from favoring the nonreactive conformation while the ligand is away to favoring the reactive conformation while the ligand is near, k_{CS} and k_{IF} become close and, thus, provide tight bounds to the binding rate constant. This finding has significant mechanistic implications and presents routes for obtaining good estimates of the rate constant at low cost. © *2011 American Institute of Physics*. [doi:10.1063/1.3561694]

I. INTRODUCTION

Protein–ligand binding is generally accompanied by conformational change. Both the diffusional approach of the protein–ligand pair and the conformational change can be rate-limiting steps.[1] When binding is rate-limited by diffusion, only the relative translation and the overall rotation of the binding molecules need to be considered; theories and simulations have had great successes in this regime.[2,3] Outside this regime, one has to explicitly treat conformational degrees of freedom. Theories have been presented where the conformational degrees of freedom are uncoupled to the translational–rotational degrees of freedom.[4–7] However, one expects the energy surface in conformational space to change significantly when the protein–ligand pair moves from far apart to near contact. Indeed, a far more realistic energy surface is one that switches from favoring a "nonreactive" conformation while the protein–ligand pair is far apart to favoring a "reactive" conformation while the protein–ligand pair is near contact (Fig. 1). In this paper we deal with such a binding-coupled energy surface.

Szabo *et al.*[4] introduced a model in which conformational change of the protein is treated as stochastic gating,

$$\text{nonreactive} \underset{\omega_-}{\overset{\omega_+}{\rightleftarrows}} \text{reactive}, \tag{1}$$

with the transition rates $\omega_\pm$ between reactive and nonreactive conformations independent of the ligand position. Binding to the protein can occur only when the ligand is in the "reactive" region while the protein is in the reactive conformation. Such a model predicts the following relation for the binding rate constant k_{on}:[4,7]

$$\frac{1}{k_{on}} = \frac{1}{k_{on0}} + \frac{\omega_-}{\omega_+}\frac{1}{(\omega_+ + \omega_-)\hat{k}_0(\omega_+ + \omega_-)}, \tag{2}$$

where $\hat{k}_0(s)$ is the Laplace transform of the time-dependent rate coefficient when the protein always stays in the reactive conformation and k_{on0} is the long-time limit of that rate coefficient. The limits of $k_{on,}$ when the gating rates are infinitely slow and infinitely fast, are of particular interest. We denote these two limits as k_{slow} and k_{fast}. When the reactivity of the protein–ligand pair in the reaction region is small, k_{slow} and k_{fast} are identical:

$$k_{slow} = k_{fast} = p_+ k_{on0}, \tag{3}$$

where $p_+ = \omega_+/(\omega_+ + \omega_-)$ is the equilibrium probability for the protein to adopt the reactive conformation. However, when the reactivity is large the two limits are very different:

$$k_{slow} = p_+ k_{on0}, \tag{4a}$$

$$k_{fast} = k_{on0}. \tag{4b}$$

So the slow-gating limit can be significantly less than the fast-gating limit.

a)Electronic mail: hzhou4@fsu.edu.

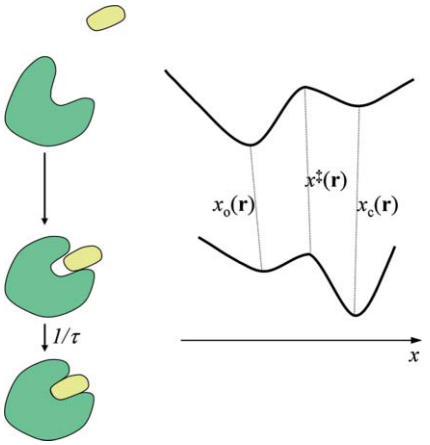

FIG. 1. A model in which conformational change and ligand binding are coupled. The energy surface of the protein depends on the protein–ligand separation **r**.

We[6] developed a related model, for the binding of a ligand to a buried site in a protein, which is accessible only through a gate that undergoes open–closed transitions. The open conformation allows for ligand entrance but the closed conformation blocks it. The open–closed transitions are assumed to be independent of the ligand position and described by the scheme of Eq. (1). Instead of stochastic gating, Agmon[5] treated the conformational change of an enzyme as diffusion along a continuous coordinate, but here again the energy surface of the enzyme was assumed to be unaffected by substrate binding.

Wade, McCammon, and co-workers have carried out Brownian dynamics and molecular dynamics simulations to explore the conformational changes of a number of proteins in the absence of ligands.[8–12] Equation (4) was used to estimate the ligand-binding rate constant. Similarly, the gating dynamics of an enzyme with a buried site, obtained from molecular dynamics simulations in the absence of the substrate, was used to estimate the substrate-binding rate constant.[13] In another interesting study,[14] an ensemble of conformations of an enzyme was generated by molecular dynamics simulations and multiple "static" snapshots were used to calculate the substrate-binding rate constant.

The most expensive and of most relevance to the present paper is a study by Wade *et al.*,[15] in which conformational fluctuation and substrate binding to an enzyme were coupled. The diffusional motion of the substrate and conformational fluctuation of the enzyme were both modeled by Brownian dynamics simulations. To reduce the computational cost of simulating the conformational fluctuation, only a critical loop over the active site was allowed to be flexible during the simulations. We note that Kim and Lee[16] have recently derived an formal expression for the binding rate coefficient for reactant molecules with internal degrees of freedom.

To further investigate how conformational change that is coupled to ligand binding affects binding mechanism and rate constant, we[17] extended the model of Szabo *et al.*[4] to include two sets of transition rates, one when the ligand is far away from the protein and the other when the ligand is near. The transition rates were chosen to favor the nonreactive conformation when the ligand is away but the reactive conformation when the ligand is near. In order to obtain analytical results, the geometry of the model considered was very simple. The protein was modeled as a sphere with a uniformly reactive surface. It was found that, when the rates of conformational transitions are low, binding proceeds via an apparently conformational-selection (CS) mechanism; in contrast, when the rates of conformational transitions are high, binding proceeds via an apparently induced-fit (IF) mechanism. Moreover, the rate constants in these limiting situations, k_{CS} and k_{IF}, provide lower and upper bounds, respectively.

Here we increase the realism of this line of development by restricting the reactive region to a localized site on the protein surface. We describe a method for calculating the binding rate coefficient from Brownian dynamics simulations. Our simulation results on a model system showed that the CS and IF limits become closer and closer as the reactive region becomes more and more localized. This finding prompted us to derive general relations for k_{CS} and k_{IF}. They confirm that, for an energy surface that switches from favoring the nonreactive conformation while the ligand is away to favoring the reactive conformation while the ligand is near, k_{CS} and k_{IF} are close. Therefore, these two limiting values provide tight bounds on the binding rate constant. No simulation of conformational change is necessary in the calculation of either k_{CS} or k_{IF}; hence these limits provide good estimates of k_{on} at low computational cost. The closeness of k_{CS} and k_{IF} under conditions that realistically model protein–ligand binding has significant mechanistic implications.

The rest of the paper is organized as follows. In Sec. II, we present a general formulation of the time-dependent binding rate coefficient when the energy surface of the protein is coupled to ligand binding. There we also present a method for calculating the rate coefficient from Brownian dynamics simulations. In Sec. III, we present analytical results for the binding rate constant and its CS and IF limits. In Sec. IV, we illustrate our Brownian dynamics method for calculating the rate constant on a model system. Details of the implementation are described. In Sec. V, we present simulation results and compare them against available analytical results. Finally, in Sec. VI we discuss the implications and further extensions of the present study.

II. FORMULATION OF THE GENERAL PROBLEM

Here, we present a general formulation of the time-dependent binding rate coefficient when the energy surface of the protein is coupled to ligand binding. A method to calculate the rate coefficient from Brownian dynamics simulations is described. For simplicity, we treat a pointlike ligand and describe the conformational space by a single coordinate x. Our basic formulation can be easily extended to a more general case involving both translational and rotational diffusion of the protein and the ligand as well as a higher-dimensional conformational space. The motion along x is first treated as continuous diffusion and then as transitions between two discrete conformations.

A. Continuous diffusion in conformational space

As illustrated in Fig. 1, in the system that we consider, the protein with a conformation space specified by a one-dimensional variable x binds with a pointlike ligand. For example, x may represent the opening size of the ligand-binding pocket on the protein surface. We note that the mathematics is identical if the protein is modeled as rigid and the ligand, such as a peptide, is modeled as flexible; in that case x represents the conformational degree of freedom of the ligand. Let the protein–ligand separation be a three-dimensional vector $\mathbf{r}$. We treat the motions in $\mathbf{r}$ and along x both as diffusion, with diffusion constants D and D_1, respectively. Let the potential of mean force in $\mathbf{r}$ and x be $U(\mathbf{r}, x)$. We envision that, for a given $\mathbf{r}$, $U(\mathbf{r}, x)$ has a double-well shape along x (Fig. 1); for example, one well may correspond to a closed conformation and the other well an open conformation. When the ligand is far away, the Boltzmann weight of the open conformation is much greater than that of the closed form, so that the protein mostly adopts the open conformation. We denote the potential as $U_\infty(x)$ when the protein and ligand are far apart. When the ligand is near or inside the binding pocket, due to protein–ligand interactions, the Boltzmann weights of the open and closed conformations are reversed to favor the latter. Within this loosely-bound state, the protein–ligand pair can form the native complex. We treat the last step as a rate process with a rate constant $1/\tau$.

In the Smoluchowski theory for diffusion-influenced bimolecular reactions, the rate coefficient, $k(t)$, for protein–ligand binding is calculated from the pair distribution function, $P(\mathbf{r}, x, t)$, where t denotes time.[7,16,18] The pair distribution function satisfies the Smoluchowski equation augmented by a "sink" term to account for the formation of the native complex:

$$\begin{aligned}\frac{\partial P(\mathbf{r},x,t)}{\partial t} &= \nabla \cdot D e^{-\beta U(\mathbf{r},x)} \nabla e^{\beta U(\mathbf{r},x)} P(\mathbf{r},x,t) \\ &\quad + \frac{\partial}{\partial x} D_1 e^{-\beta U(\mathbf{r},x)} \frac{\partial}{\partial x} e^{\beta U(\mathbf{r},x)} P(\mathbf{r},x,t) \\ &\quad - \tau^{-1} H(\mathbf{r},x) P(\mathbf{r},x,t), \end{aligned} \tag{5}$$

where $\beta = (k_B T)^{-1}$ and $H(\mathbf{r}, x)$ is either 1 if the ligand is in the "reactive" region (forming the loosely-bound state with the protein in the closed conformation) or 0 otherwise. At infinite protein–ligand separation the pair distribution function is

$$P(\mathbf{r},x,t) = \frac{e^{-\beta U_\infty(x)}}{\int_{-\infty}^{\infty} dx e^{-\beta U_\infty(x)}} \equiv P_{\text{eq}}(x), \quad r = |\mathbf{r}| \to \infty. \tag{6}$$

The inner boundary is reflecting, since the protein and ligand cannot interpenetrate. The initial distribution is assumed to be an equilibrium distribution:

$$P(\mathbf{r},x,0) = P_{\text{eq}}(\mathbf{r},x) \equiv \frac{e^{-\beta U(\mathbf{r},x)}}{\int_{-\infty}^{\infty} dx e^{-\beta U_\infty(x)}}. \tag{7}$$

The rate coefficient is then

$$k(t) = \int d\mathbf{r} dx \tau^{-1} H(\mathbf{r},x) P(\mathbf{r},x,t). \tag{8}$$

This result generalizes the usual formulation of $k(t)$ by including the coordinate x for modeling conformational change. Using Eq. (7), the initial value $k(0)$ is found as

$$k(0) = \int d\mathbf{r} dx \tau^{-1} H(\mathbf{r},x) P_{\text{eq}}(\mathbf{r},x). \tag{9}$$

If diffusions (both in $\mathbf{r}$ and along x) were infinitely fast (or equivalently, reactions were infinitely slow, i.e., $\tau \to \infty$), the protein–ligand pair would stay in the equilibrium distribution and the rate coefficient would stay at $k(0)$. In general, $k(t)$ would decay over time.

We have previously developed an algorithm for calculating $k(t)$ from Brownian dynamics simulations.[19] The algorithm is based on the relationship between the pair distribution function $P(\mathbf{r}, x, t)$ and the survival probability $S(\mathrm{t}|\mathbf{r}, x)$:

$$P(\mathbf{r},x,t) = P_{\text{eq}}(\mathbf{r},x) S(t|\mathbf{r},x). \tag{10}$$

The survival probability is the probability that a protein–ligand pair, started at time $t = 0$ from separation $\mathbf{r}$ and conformation x, has not formed the native complex by time t. Using Eq. (10) in Eq. (8), we have

$$k(t) = \int d\mathbf{r} dx \tau^{-1} H(\mathbf{r},x) P_{\text{eq}}(\mathbf{r},x) S(t|\mathbf{r},x). \tag{11}$$

Using Eq. (9), we find the ratio $k(t)/k(0)$ as

$$\frac{k(t)}{k(0)} = \frac{\int d\mathbf{r} dx H(\mathbf{r},x) P_{\text{eq}}(\mathbf{r},x) S(t|\mathbf{r},x)}{\int d\mathbf{r} dx H(\mathbf{r},x) P_{\text{eq}}(\mathbf{r},x)}. \tag{12}$$

The last result is the basis of our Brownian dynamics algorithm for calculating the binding rate coefficient. When Brownian trajectories are started at $t = 0$ from the reactive region with a distribution proportional to $P_{\text{eq}}(\mathbf{r}, x)$, the average survival probability at time t is $k(t)/k(0)$. Multiplied by $k(0)$ calculated according to Eq. (9), we obtain $k(t)$.

Let us briefly outline how our formulation of $k(t)$ can be extended to the more general case involving both translational and rotational diffusion of the protein and the ligand as well as a higher-dimensional conformational space. Now $\mathbf{r}$ represents the protein–ligand relative translational and overall rotational degrees of freedom and x represents coordinates in the higher-dimensional conformational space. The function $H(\mathbf{r}, x)$ still defines the reaction criterion: it has a value of 1 only if the protein–ligand pair is separated and orientated such that it is in the loosely-bound state with the two molecules correctly facing each other and in their respective reactive conformations. The protein–ligand pair is started from the region in configurational–conformaitonal space where $H(\mathbf{r}, x) = 1$ and then propagated. Whenever it is found in the region where $H(\mathbf{r}, x) = 1$, the protein–ligand pair is allowed to be transformed, with a mean lifetime τ, into the native complex. Again, the survival probability at time t gives $k(t)/k(0)$.

B. Transitions between two discrete conformations

For the model shown in Fig. 1, at each protein–ligand separation $\mathbf{r}$, the conformation of the protein is either open or

closed. Let the location of the energy barrier separating the two conformations be $x = x^{\ddagger}(\mathbf{r})$. The open and closed conformations are specified by $-\infty < x < x^{\ddagger}(\mathbf{r})$ and $x^{\ddagger}(\mathbf{r}) < x < \infty$, respectively. If the energy barrier is high, then equilibration within each well is fast and transitions between the two wells can be treated as rate processes:

$$\mathrm{o} \underset{\omega_-(\mathbf{r})}{\overset{\omega_+(\mathbf{r})}{\rightleftharpoons}} \mathrm{c}. \tag{13}$$

According to Kramers,[20] the transition rate constants are given by

$$\omega_+(\mathbf{r}) = \frac{1}{\int_{-\infty}^{x^{\ddagger}(\mathbf{r})} dx e^{-\beta U(\mathbf{r},x)} \int_{x_{\mathrm{o}}(\mathbf{r})}^{x_{\mathrm{c}}(\mathbf{r})} dx [D_1 e^{-\beta U(\mathbf{r},x)}]^{-1}}, \tag{14a}$$

$$\omega_-(\mathbf{r}) = \frac{1}{\int_{x^{\ddagger}(\mathbf{r})}^{\infty} dx e^{-\beta U(\mathbf{r},x)} \int_{x_{\mathrm{o}}(\mathbf{r})}^{x_{\mathrm{c}}(\mathbf{r})} dx [D_1 e^{-\beta U(\mathbf{r},x)}]^{-1}}, \tag{14b}$$

where $x_{\mathrm{o}}(\mathbf{r})$ and $x_{\mathrm{c}}(\mathbf{r})$ are the values of x at the bottoms of the open and closed energy wells, respectively. In line with fast equilibration within both energy wells, we define conformation-specific potential energies, $U_{\mathrm{o}}(\mathbf{r})$ and $U_{\mathrm{c}}(\mathbf{r})$, for protein–ligand interaction by

$$e^{-\beta U_{\mathrm{o}}(\mathbf{r})} = \frac{\int_{-\infty}^{x^{\ddagger}(\mathbf{r})} dx e^{-\beta U(\mathbf{r},x)}}{\int_{-\infty}^{x_{\infty}^{\ddagger}} dx e^{-\beta U_{\infty}(x)}}, \tag{15a}$$

$$e^{-\beta U_{\mathrm{c}}(\mathbf{r})} = \frac{\int_{x^{\ddagger}(\mathbf{r})}^{\infty} dx e^{-\beta U(\mathbf{r},x)}}{\int_{x_{\infty}^{\ddagger}}^{\infty} dx e^{-\beta U_{\infty}(x)}}, \tag{15b}$$

where $x_{\infty}^{\ddagger}$ is the location of the energy barrier separating the open and closed conformations when the ligand is far away. The denominators in Eq. (15) ensure that $U_{\mathrm{o}}(\mathbf{r})$ and $U_{\mathrm{c}}(\mathbf{r})$ go to zero at infinite protein–ligand separation. Correspondingly we denote the pair distribution as $P_{\mathrm{o}}(\mathbf{r}, t)$ and $P_{\mathrm{c}}(\mathbf{r}, t)$ when the protein is in the open and closed conformations, respectively,

$$P_{\mathrm{o}}(\mathbf{r}, t) = \int_{-\infty}^{x^{\ddagger}(\mathbf{r})} dx P(\mathbf{r}, x, t), \tag{16a}$$

$$P_{\mathrm{c}}(\mathbf{r}, t) = \int_{x^{\ddagger}(\mathbf{r})}^{\infty} dx P(\mathbf{r}, x, t). \tag{16b}$$

The initial values of the pair distribution function are

$$P_{\mathrm{g}}(\mathbf{r}, 0) = P_{\mathrm{geq}}(\mathbf{r}) \equiv p_{\infty \mathrm{g}} e^{-\beta U_{\mathrm{g}}(\mathbf{r})}, \tag{17a}$$

where g = o or c, and the equilibrium probability, $p_{\infty \mathrm{g}}$, of conformation g at infinite protein–ligand separation is

$$p_{\infty \mathrm{o}} = \frac{\int_{-\infty}^{x_{\infty}^{\ddagger}} dx e^{-\beta U_{\infty}(x)}}{\int_{-\infty}^{\infty} dx e^{-\beta U_{\infty}(x)}}; \quad p_{\infty \mathrm{c}} = \frac{\int_{x_{\infty}^{\ddagger}}^{\infty} dx e^{-\beta U_{\infty}(x)}}{\int_{-\infty}^{\infty} dx e^{-\beta U_{\infty}(x)}}. \tag{17b}$$

With the transitions between the two conformations treated as rate processes, the pair distribution now satisfies (see Appendix A)

$$\frac{\partial P_{\mathrm{o}}(\mathbf{r}, t)}{\partial t} = \nabla \cdot D e^{-\beta U_{\mathrm{o}}(\mathbf{r})} \nabla e^{\beta U_{\mathrm{o}}(\mathbf{r})} P_{\mathrm{o}}(\mathbf{r}, t) - \omega_+(\mathbf{r}) P_{\mathrm{o}}(\mathbf{r}, t) + \omega_-(\mathbf{r}) P_{\mathrm{c}}(\mathbf{r}, t), \tag{18a}$$

$$\frac{\partial P_{\mathrm{c}}(\mathbf{r}, t)}{\partial t} = \nabla \cdot D e^{-\beta U_{\mathrm{c}}(\mathbf{r})} \nabla e^{\beta U_{\mathrm{c}}(\mathbf{r})} P_{\mathrm{c}}(\mathbf{r}, t) + \omega_+(\mathbf{r}) P_{\mathrm{o}}(\mathbf{r}, t) - \omega_-(\mathbf{r}) P_{\mathrm{c}}(\mathbf{r}, t) - \tau^{-1} H_{\mathrm{c}}(\mathbf{r}) P_{\mathrm{c}}(\mathbf{r}, t), \tag{18b}$$

where $H_{\mathrm{c}}(\mathbf{r})$ is given by Eq. (A8). Note that the ratio of the transition rates between the two conformations is

$$\frac{\omega_+(\mathbf{r})}{\omega_-(\mathbf{r})} = \frac{\int_{x^{\ddagger}(\mathbf{r})}^{\infty} dx e^{-\beta U(\mathbf{r},x)}}{\int_{-\infty}^{x^{\ddagger}(\mathbf{r})} dx e^{-\beta U(\mathbf{r},x)}} = \frac{\int_{x_{\infty}^{\ddagger}}^{\infty} dx e^{-\beta U_{\infty}(x)}}{\int_{-\infty}^{x_{\infty}^{\ddagger}} dx e^{-\beta U_{\infty}(x)}} e^{-\beta [U_{\mathrm{c}}(\mathbf{r}) - U_{\mathrm{o}}(\mathbf{r})]}. \tag{19a}$$

Denoting the transition rate constants at infinite separation as $\omega_{\infty \pm}$, we have

$$\frac{\omega_{\infty +}}{\omega_{\infty -}} = \frac{\int_{x_{\infty}^{\ddagger}}^{\infty} dx e^{-\beta U_{\infty}(x)}}{\int_{-\infty}^{x_{\infty}^{\ddagger}} dx e^{-\beta U_{\infty}(x)}} = \frac{p_{\infty \mathrm{c}}}{p_{\infty \mathrm{o}}} \tag{19b}$$

and Eq. (19a) can be expressed as

$$\frac{\omega_+(\mathbf{r})}{\omega_-(\mathbf{r})} = \frac{\omega_{\infty +}}{\omega_{\infty -}} e^{-\beta [U_{\mathrm{c}}(\mathbf{r}) - U_{\mathrm{o}}(\mathbf{r})]}. \tag{19c}$$

This result has been written down based on detailed balance.[17] The gating model of Shoup *et al.*[4] has $\omega_+(\mathbf{r}) = \omega_{\infty +}$ and $\omega_-(\mathbf{r}) = \omega_{\infty -}$; correspondingly $U_{\mathrm{c}}(\mathbf{r}) = U_{\mathrm{o}}(\mathbf{r})$ in their model.

The binding rate coefficient is now given by

$$k(t) = \int d\mathbf{r} \tau^{-1} H_{\mathrm{c}}(\mathbf{r}) P_{\mathrm{c}}(\mathbf{r}, t). \tag{20}$$

In the present case, the pair distribution function $P_{\mathrm{g}}(\mathbf{r}, t)$ relates to the survival probability $S(t|\mathbf{r}, \mathrm{g})$ via

$$P_{\mathrm{g}}(\mathbf{r}, t) = p_{\infty \mathrm{g}} e^{-\beta U_{\mathrm{g}}(\mathbf{r})} S(t|\mathbf{r}|, \mathrm{g}). \tag{21}$$

Using the last result, we find

$$k(0) = \int dr \tau^{-1} H_{\mathrm{c}}(\mathbf{r}) p_{\infty \mathrm{c}} e^{-\beta U_{\mathrm{c}}(\mathbf{r})}, \tag{22}$$

$$\frac{k(t)}{k(0)} = \frac{\int d\mathbf{r} H_{\mathrm{c}}(\mathbf{r}) e^{-\beta U_{\mathrm{c}}(\mathbf{r})} S(t|\mathbf{r}, \mathrm{c})}{\int d\mathbf{r} H_{\mathrm{c}}(\mathbf{r}) e^{-\beta U_{\mathrm{c}}(\mathbf{r})}}. \tag{23}$$

The last result shows that $k(t)/k(0)$ is given by the average survival probability at time t when Brownian trajectories are started with the ligand in the reactive region specified by the condition $H_{\mathrm{c}}(\mathbf{r}) = 1$, with position following the Boltzmann distribution $\exp[-\beta U_{\mathrm{c}}(\mathbf{r})]$, and with the protein in the closed, i.e., reactive conformation.

C. Asymptotic behavior of $k(t)$

While the Brownian dynamics algorithm outlined above gives the full time dependence of the binding rate coefficient,

the limit of $k(t)$ at $t \to \infty$, i.e., the steady-state value, is of particular interest. We refer to this steady-state value the rate constant and denote it as k_{on}. In previous work,[7] we have shown that $k(t)$ has the following asymptotic behavior at long times:

$$k(t) = k_{on}\left[1 + \frac{k_{on}}{4\pi D}(\pi Dt)^{-1/2} + \cdots\right]. \tag{24}$$

Even though Brownian dynamics simulations must have a finite cutoff time (t_{cut}), yielding $k(t)$ up to $t = t_{cut}$, by fitting the long-time portion of $k(t)$ to Eq. (24) we can obtain the steady-state value k_{on}.[19] The fitting function is linear when the independent variable is chosen as $(\pi Dt)^{-1/2}$. The intercept of the linear fit is k_{on} and the slope is ${k_{on}}^2/4\pi D$ which can be calculated from the intercept. That the slope is not a free parameter but can be calculated from the intercept provides a convenient way to validate the simulation results and determine whether t_{cut} is sufficiently long to ensure asymptotic behavior.

III. ANALYTICAL RESULTS FOR RATE CONSTANTS

The preceding section outlines a Brownian dynamics algorithm for calculating the binding rate coefficient when protein–ligand binding is coupled to conformational change. The conformational change is either treated as continuous diffusion along a one-dimensional coordinate or as rate processes between two discrete conformations. The latter treatment will be the focus for the remainder of this paper. We now present some analytical results for the rate constant k_{on}, i.e., the steady-state value of $k(t)$. Both explicit k_{on} results for model systems and general relations for the CS and IF limits of k_{on} will be given. The analytical results obtained on model systems can serve an important role in validating simulation algorithms (see Sec. V).

A. Dual-transition-rates model: Spherical symmetry

In earlier work we introduced the dual-transition-rates model.[17] The model has spherical symmetry in $\mathbf{r}$. The transition rates between the open and closed conformations have two sets of values, $\omega_\pm$ in the loosely-bound state, where $R < r < R + \Delta \equiv R_1$, and $\omega_{\infty\pm}$, when $r > R_1$. The conformation-specific interaction potentials are 0 when $r > R_1$ and constants U_o and U_c, respectively, for the open and closed conformations when $R < r < R_1$. According to Eq. (19c), the two sets of transition rates are constrained by

$$\frac{\omega_+}{\omega_-} = \frac{\omega_{\infty+}}{\omega_{\infty-}} e^{-\beta(U_c - U_o)}. \tag{25}$$

The reactive region is a very thin shell in $R < r < R + \varepsilon$. When $\varepsilon \to 0$ and $\tau \to 0$ but ε/τ has a finite value κ, the sink term is equivalent to a radiation boundary condition[19]

$$De^{-\beta U_c(\mathbf{r})}\frac{\partial}{\partial r}e^{\beta U_c(\mathbf{r})}P_c(\mathbf{r},t) = \kappa P_c(\mathbf{r},t), \quad r = R. \tag{26}$$

The rate constant k_{on} for this model is given by[17]

$$\begin{aligned}4\pi DR_1 p_{\infty c} e^{-\beta U_c}(B_+C_- + B_-C_+)/k_{on} \\ = p_c(A_+C_- + A_-C_+) + p_c(e^{-\beta U_{eff}} - 1)(B_+C_- + B_-C_+) \\ -(p_c p_{\infty o} - p_o p_{\infty c})[4 + (p_c p_{\infty o}/p_o - p_{\infty c}) \\ \times (B_+e^{\lambda\Delta} + B_-e^{-\lambda\Delta})],\end{aligned} \tag{27}$$

where

$$A_\pm = 1 \pm (1 + \kappa R/D \pm \lambda R)/p_c\kappa\lambda R^2/D, \tag{28a}$$

$$B_\pm = \pm(1 \pm \lambda R)/\lambda R_1, \tag{28b}$$

$$C_\pm = [1 \pm \lambda R_1 - (1 + \lambda_\infty R_1)p_{\infty o}/p_o e^{-\beta U_c}]e^{\mp\lambda\Delta}, \tag{28c}$$

$$e^{-\beta U_{eff}} = p_{\infty o}e^{-\beta U_o} + p_{\infty c}e^{-\beta U_c}, \tag{28d}$$

$$\lambda = [(\omega_+ + \omega_-)/D]^{1/2}, \tag{28e}$$

$$\lambda_\infty = [(\omega_{\infty+} + \omega_{\infty-})/D]^{1/2}, \tag{28f}$$

$$p_c = \omega_+/(\omega_+ + \omega_-); \quad p_o = \omega_-/(\omega_+ + \omega_-). \tag{28g}$$

In the dual-transition-rates model, conformational selection and induced fit are manifest at two extremes of the transition rates. When the transition rates are extremely small, conformational selection appears as the binding mechanism and the rate constant is

$$k_{CS} = p_{\infty c}k_{on0}, \tag{29}$$

where k_{on0} is the rate constant when the protein stays in the closed (i.e., reactive) conformation, given by

$$\frac{1}{k_{on0}} = \frac{1}{4\pi\kappa R^2 e^{-\beta U_c}} + \frac{1 + (e^{-\beta U_c} - 1)R/R_1}{4\pi DRe^{-\beta U_c}}. \tag{30}$$

In the opposite extreme, where the transition rates are large, induced fit appears as the binding mechanism, and the rate constant is given by

$$\frac{1}{k_{IF}} = \frac{1}{4\pi p_c\kappa R^2 e^{-\beta U_{eff}}} + \frac{1 + (e^{-\beta U_{eff}} - 1)R/R_1}{4\pi DRe^{-\beta U_{eff}}}. \tag{31}$$

Note that the last result is equivalent to the rate constant when the protein stays in the reactive conformation but with the reactivity κ replaced by $p_c\kappa$ and the potential U_c replaced by U_{eff}. The CS and IF limits are lower and upper bounds of k_{on}, respectively.

B. Dual-transition-rates model: Reactive patch

The spherically symmetric dual-transition-rates model can be made more realistic when the reactive region is restricted to a patch, with the polar angle θ limited to the range from 0 to θ_0 (Fig. 2). We have not been able to find a general expression for the rate constant of this model. However, the CS and IF limits can be obtained. These results are based

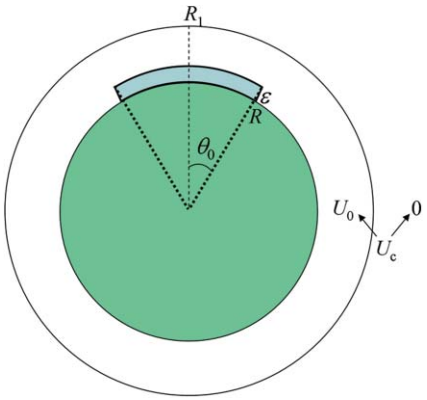

FIG. 2. The model system for which analytical and simulation results are obtained in the present study.

on the rate constant for the situation where the protein always stays in the reactive conformation. The latter, denoted as $k_{on0}(\theta_0, \kappa, U)$ to signify the dependence on the patch angle θ_0, the reactivity κ at $r = R$, and the potential U in $R < r < R_1$, is given by[21]

$$\frac{1}{k_{on0}(\theta_0, \kappa, U)} = \frac{1}{2\pi(1-\cos\theta_0)\kappa R^2 e^{-\beta U}} + \frac{1}{k_D(\theta_0, U)}, \tag{32}$$

where

$$\frac{1}{k_D(\theta_0, U)} = \frac{\sum_{l=0}^{\infty} \frac{c_l}{(l+1)(2l+1)} \frac{1+(l+1)A_l}{1-lA_l}}{4\pi D R e^{-\beta U} c_0}, \tag{33a}$$

$$c_l = [P_{l-1}(\cos\theta_0) - P_{l+1}(\cos\theta_0)]^2, \tag{33b}$$

$$A_l = \frac{(e^{-\beta U}-1)(R/R_1)^{2l+1}}{l+1+le^{-\beta U}}, \tag{33c}$$

with $P_l(x)$ denoting Legendre polynomials. The first term on the right hand side of Eq. (32) gives the rate constant when the reactivity is small, i.e., $\kappa R/D \to 0$, whereas the second term gives the rate constant when the reactivity is large, i.e., $\kappa R/D \to \infty$. A similar interpretation applies to the two terms of Eq. (30) or (31).

When $\theta_0 = 180^\circ$, Eq. (32) reduces to Eq. (30) with $U = U_c$. However, for other values of θ_0, the derivation of Eq. (32) relies on the so-called constant-flux approximation.[22] As $\theta_0 \to 0$, Eq. (33a) is known to underestimate $k_D(\theta_0, U)$ by a factor $32/3\pi^2 \approx 1.08$. Following our earlier work,[21] we scale up this $k_D(\theta_0, U)$ by $32/3\pi^2$ to calculate $k_{on0}(\theta_0, \kappa, U)$ when comparison is made against simulation results for small reactive patches (i.e., $\theta_0 \leq 15^\circ$).

In the limit that conformational transitions are extremely slow, the protein cannot change its conformation during the binding process. If the protein started out in the closed (i.e., reactive) conformation, the binding rate constant would be $k_{on0}(\theta_0, \kappa, U_c)$. If the protein started out in the open (i.e., nonreactive) conformation, it would not be able to bind with the ligand at all and the binding rate constant would be 0. Now the probability for the protein to start in the closed conformation while the ligand is far away is $p_{\infty c}$. Therefore the binding rate constant is the average over the two starting conformations:

$$k_{CS} = p_{\infty c} k_{on0}(\theta_0, \kappa, U_c). \tag{34}$$

In the limit that conformational transitions are extremely fast, the ligand always experiences an average effective potential, given by Eq. (28d). In the reactive region, the ligand likewise experiences an average reactivity, given by $p_c\kappa$. The binding rate constant is, thus,

$$k_{IF} = k_{on0}(\theta_0, p_c\kappa, U_{eff}). \tag{35}$$

When $\kappa R/D \to 0$, k_{CS} becomes

$$k_{CS} \to 2\pi(1-\cos\theta_0)\kappa R^2 p_{\infty c} e^{-\beta U_c}, \tag{36}$$

which is the initial value $k(0)$ of the time-dependent rate coefficient [see Eq. (22)]. In fact, when $\kappa R/D \to 0$, the rate constant should always be $k(0)$ irrespective of the timescale of the conformational transitions. It can be easily verified that that k_{IF} approaches $k(0)$ when $\kappa R/D \to 0$.

How do k_{CS} and k_{IF} compare when $\kappa R/D \to \infty$? For notational simplicity we denote $k_{on0}(\theta_0, \kappa, U)$ as $k_D(\theta_0, U)$ when $\kappa R/D \to \infty$. For a small reactive patch (i.e., $\theta_0 \to 0$), which models stereospecific ligand binding, an approximate dependence of $k_D(\theta_0, U)$ on U,[21,23]

$$k_D(\theta_0, U) \approx k_D(\theta_0, 0) e^{-\beta U}, \tag{37}$$

is very useful here. The accuracy of this approximation is also affected by the range (Δ in the present case) of the interaction potential U, improving as Δ increases. Assuming that Eq. (37) holds, then when $\kappa R/D \to \infty$, k_{CS} becomes

$$k_{CS} \to k_D(\theta_0, 0) p_{\infty c} e^{-\beta U_c}, \tag{38}$$

and k_{IF} becomes

$$\begin{aligned} k_{IF} \to k_D(\theta_0, 0) e^{-\beta U_{eff}} &= k_D(\theta_0, 0)(p_{\infty o} e^{-\beta U_o} + p_{\infty c} e^{-\beta U_c}) \\ &= k_D(\theta_0, 0) p_{\infty o} e^{-\beta U_o} \\ &\quad + k_D(\theta_0, 0) p_{\infty c} e^{-\beta U_c}. \end{aligned} \tag{39}$$

Note that the second term of Eq. (39) is the same as k_{CS} of Eq. (38). According to Eq. (25), the ratio of the second term to the first term on the right hand side of Eq. (39) is p_c/p_o. For systems of biological interest, this ratio should be much greater than 1. Therefore, Eq. (39) is dominated by the second term. We can thus conclude that k_{CS} and k_{IF} are close when $\kappa R/D \to \infty$. Since k_{CS} and k_{IF} are actually identical when $\kappa R/D \to 0$, we can further conclude that k_{CS} and k_{IF} are close irrespective of the value of $\kappa R/D$, as long as the reactive patch is small.

C. Stereospecific binding in general

The closeness of k_{CS} and k_{IF} derived for the model with a small reactive patch actually holds in general, as long as

the reactive region is small relative to the range of the interaction potential. This is a highly significant result. Because k_{CS} and k_{IF} provide lower and upper bounds of k_{on}, if they are close in value then either of them can provide a good estimate for k_{on}. Below we outline a derivation for the closeness of k_{CS} and k_{IF}, making clear the approximation involved. The derivation given here is for the case where conformational change is treated as rate processes between two discrete conformations. The case where conformational change is treated as continuous diffusion along a one-dimensional coordinate is presented in Appendix B.

When the reactivity is small, both k_{CS} and k_{IF} are identical to $k(0)$. The difference between k_{CS} and k_{IF} is the greatest when the reactivity is big. In that situation, when conformational change is treated as rate processes between two discrete conformations, we have

$$k_{CS} \to p_{\infty c} k_D[U_c(\mathbf{r})], \tag{40}$$

where $k_D[U_c(\mathbf{r})]$ is the rate constant when the protein does not undergo conformational change, the interaction potential is $U_c(\mathbf{r})$, and the reactivity is infinite. Correspondingly k_{IF} approaches

$$k_{IF} \to k_D[U_{eff}(\mathbf{r})], \tag{41}$$

where $U_{eff}(\mathbf{r})$ is the conformation-averaged effective potential

$$e^{-\beta U_{eff}(\mathbf{r})} = \frac{\int_{-\infty}^{\infty} dx e^{-\beta U(\mathbf{r},x)}}{\int_{-\infty}^{\infty} dx e^{-\beta U_\infty(x)}}, \tag{42a}$$

$$= p_{\infty o} e^{-\beta U_o(\mathbf{r})} + p_{\infty c} e^{-\beta U_c(\mathbf{r})}. \tag{42b}$$

When the reactive region is small relative to the range of the interaction potential $U(\mathbf{r})$, our previous work[21,23] suggests an approximate dependence of $k_D[U(\mathbf{r})]$ on $U(\mathbf{r})$:

$$k_D[U(\mathbf{r})] \approx k_{D0} e^{-\beta U(\mathbf{r}_{RR})}, \tag{43}$$

where k_{D0} denotes the counterpart of $k_D[U(\mathbf{r})]$ when the interaction potential $U(\mathbf{r})$ is turned off, and $\mathbf{r}_{RR}$ is a representative point in the reactive region. This approximation holds when the range of the potential $U(\mathbf{r})$ is significantly greater than the size of the reactive region. Applying this approximation, we have

$$k_{CS} \to k_{D0} p_{\infty c} e^{-\beta U_c(\mathbf{r}_{RR})}, \tag{44}$$

$$\begin{aligned} k_{CS} &\to k_{D0} e^{-\beta U_{eff}(\mathbf{r}_{RR})} \\ &= k_{D0}[p_{\infty o} e^{-\beta U_o(\mathbf{r}_{RR})} + p_{\infty c} e^{-\beta U_c(\mathbf{r}_{RR})}] \\ &= k_{D0} p_{\infty o} e^{-\beta U_o(\mathbf{r}_{RR})} + k_{D0} p_{\infty c} e^{-\beta U_c(\mathbf{r}_{RR})}. \end{aligned} \tag{45}$$

The second term of Eq. (45) is the same as Eq. (44). In addition, we also expect the second term of Eq. (45) to dominate over its first term. The preceding derivation thus shows that, even when the reactivity increases to infinity, where the difference between k_{CS} and k_{IF} is the greatest, these two bounds of k_{on} are close as long as $U_o(\mathbf{r})$ and $U_c(\mathbf{r})$ vary slowly over the reactive region and beyond.

IV. ILLUSTRATION OF THE BROWNIAN DYNAMICS ALGORITHM

In Sec. II we described an algorithm for calculating $k(t)$ from Brownian dynamics simulations. This algorithm is adapted from our earlier work[19] but now accounts for protein conformational change that is coupled to ligand binding. It consists of the following steps. (i) At the start of each Brownian trajectory, the protein is in the closed, i.e., reactive conformation and the ligand is in the reactive region. The starting position of the ligand follows the Boltzmann distribution $\exp[-\beta U_c(\mathbf{r})]$. (ii) The trajectory is then propagated (conformational transition for the protein and translational diffusion for the ligand). (iii) Whenever the ligand moves into the reactive region while the protein is in the closed conformation, they can react with the rate constant $1/\tau$ to form the native complex. If the reaction occurs, the trajectory is terminated; otherwise it is propagated to a preset cutoff time t_{cut}. (iv) Finally, after all the trajectories (totaling N_{traj}) are finished, the fraction of surviving trajectories at time $t < t_{cut}$ is calculated. That fraction is equal to $k(t)/k(0)$. Multiplying by $k(0)$ [calculated according to Eq. (22)], we obtain $k(t)$. Below we further explain steps (ii) and (iii) of our algorithm and provide details of its implementation on a model system.

A. Trajectory propagation

Our simulations consist of simultaneous transition of the protein in conformational space and translational diffusion of the ligand in the three-dimensional space exterior to the protein. The translational diffusion follows the Ermak–McCammon algorithm:[24]

$$\mathbf{r} = \mathbf{r}_0 + \beta \mathbf{F}_0 D \Delta t + (2D\Delta t)^{1/2} \mathbf{R}, \tag{46}$$

where $\mathbf{r}_0$ is the current position, $\mathbf{r}$ is the position after timestep Δt, $\mathbf{F}_0 = -\nabla U_{g_0}(\mathbf{r}_0)$ is the force calculated for the current ligand position $\mathbf{r}_0$ and the current protein conformation g_0, and $\mathbf{R}$ is a random vector with components generated from a normal distribution.

The conformational transition of the protein is implemented as follows. Suppose that the current conformation of the protein is open. The probability that, within the timestep Δt, the protein switches to the closed conformation is $\exp[-(\omega_+(\mathbf{r}_0) + \omega_+(\mathbf{r}))\Delta t/2]$. This probability is compared with a random number uniformly distributed between 0 and 1. If the random number is smaller, then the conformational switch takes place; otherwise the protein stays in the open conformation. The case where the current conformation is closed can be similarly treated.

The reflecting protein surface has to be dealt with. We follow a very simple recipe.[19] If the new position of the ligand penetrates inside the protein surface, the ligand is "reflected" back to its current position and, at the same time, the protein is restored to its current conformation. The lifetime is incremented by the timestep, just like in a regular move.

B. Determination of whether reaction occurs

The probability that, within the timestep Δt, the protein and ligand react to form the native complex is $\exp[-(i_0 + i)\Delta t/2\tau]$, where $i_0 = 1$ if the current ligand position is inside the reactive region and the current protein conformation is closed and $= 0$ otherwise; and i is the counterpart of i_0 after the timestep. This probability is compared with a random number uniformly distributed between 0 and 1. If the random number is smaller, then the reaction takes place and the trajectory is terminated; otherwise the trajectory is continued.

C. Implementation on a model system

As illustration, we implemented the Brownian dynamics algorithm for calculating $k(t)$ on the dual-transition-rates model with a reaction patch, introduced in Subsection III.B (see Fig. 2). The interaction potential U_o for the open conformation is 0 for any $r > R$. The interaction potential U_c for the closed conformation has the form

$$U_c(r) = -\frac{U_0}{2}\left(\tanh\frac{r - R_1}{L} - 1\right), \tag{47}$$

which, when $L/\Delta \to 0$, approaches the step-function potential, having value U_0 for $R < r < R_1$ and 0 for $r > R_1$, considered in the analytical solutions of Subsections III A and III B. This smooth potential function is much easier to deal with in the Brownian dynamics simulations than the step-function potential. We chose $R = 20$ Å, $R_1 = 22$ Å, and $L = 0.1$ Å. When $r - R_1 \gg L$, $U_c(r)$ is practically 0. To minimize force calculations, we turned off $U_c(r)$ when $r > R_2 = 23$ Å.

We set $\omega_- = \omega_{\infty-}$. There was no particular reason for this choice except to reduce the number of parameters. Then Eq. (19c) gives

$$\omega_+ = \omega_{\infty+} e^{-\beta U_c(r)}. \tag{48}$$

We set $\omega_{\infty+}/\omega_{\infty-}$ to 0.1, reflecting the fact that the closed, i.e., reactive conformation is disfavored when the ligand is away. The value of $\omega_{\infty-}$ was varied to explore the full timescale range of the conformational transitions, approaching the CS limit at small $\omega_{\infty-}$ and the IF limit at large $\omega_{\infty-}$. The CS and IF limits were also studied directly. In these simulations, the protein was kept in the closed (i.e., reactive) conformation. The interaction potential was $U_c(r)$ for the CS limit and the effective potential $U_{\text{eff}}(r)$ for the IF limit. For the present model system, $\exp[-\beta U_{\text{eff}}(r)] = p_{\infty c}\exp[-\beta U_c(r)] + p_{\infty o}$. The reactivity was κ and $p_c\kappa$, respectively, in the CS and IF simulations. The rate constant from the former simulations, after multiplication by $p_{\infty c}$, yields k_{CS}, while the rate constant from the latter simulations directly gives k_{IF}.

The ligand diffusion constant D was chosen to be 10 Å^2/ns. The timestep was $\Delta t_i = 10^{-6}$ ns for $r < R_2$. For $r > R_2$, the timestep was $\Delta t_i + 10^{-3}(r - R_2)^2/2D$. The cutoff time for all the simulations was 10^4 ns.

The reactive region was defined by $R < r < R + \varepsilon$ and $\theta < \theta_0$. Corresponding to a reactivity κ, the rate constant for forming the native constant is $1/\tau = \kappa/\varepsilon$. We fixed ε at 2 Å and studied a range of κ values. When the reactive region covered the full surface ($\theta_0 = 180°$), κ was 0.5 Å/ns; 2×10^4 trajectories were run. For small reactive patches, κ was 5, 50, and 500 Å/ns; 2×10^4, 4×10^4, and 10^5 trajectories, respectively, were run.

V. SIMULATION RESULTS

Here we present simulation results for the rate coefficient of the model system shown in Fig. 2 and compare them with available analytical results. We pay particular attention to how close the CS and IF limits of the rate constant are.

A. Asymptotic behavior of $k(t)$

At long times, $k(t)$ is predicted to follow the asymptotic behavior given by Eq. (24) Specifically, $k(t)$ is a linear function of $(\pi Dt)^{-1/2}$. Furthermore, the slope is determined by the intercept. All our simulation results conform to this asymptotic behavior. For illustration, in Fig. 3 we show two sets of results, one for the case where the reactive region covers the whole surface of the protein with $\kappa = 0.5$ Å/ns and

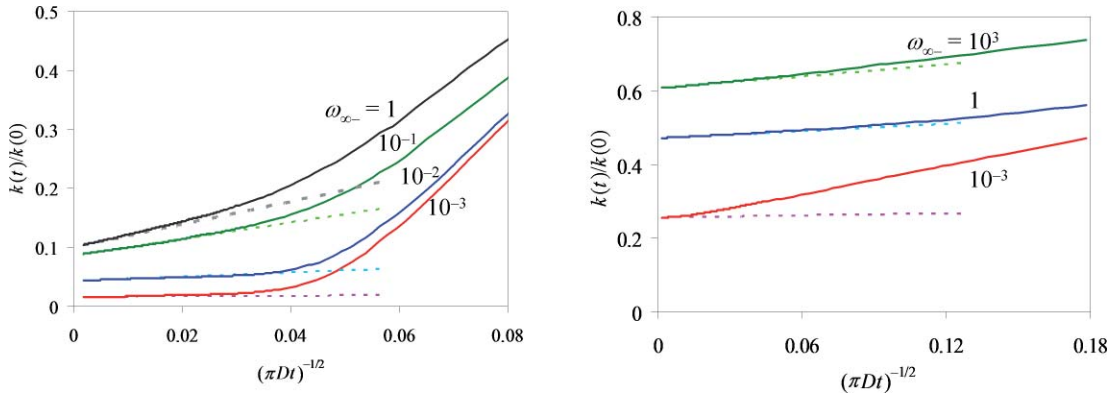

FIG. 3. The asymptotic behavior of the rate coefficient. The solid curves show simulation results for $k(t)/k(0)$; the dashed line show fits to a linear function of $(\pi Dt)^{-1/2}$. The slope of the linear function is determined from the intercept according to Eq. (24). The values of $\omega_{\infty-}$, in units of ns^{-1}, are shown in the figure. Results for two sets of model parameters are given: (a) $\theta_0 = 180°$, $\kappa = 0.5$ Å/ns, and $\exp(-\beta U_0) = 100$; (b) $\theta_0 = 10°$, $\kappa = 5$ Å/ns, and $\exp(-\beta U_0) = 10$.

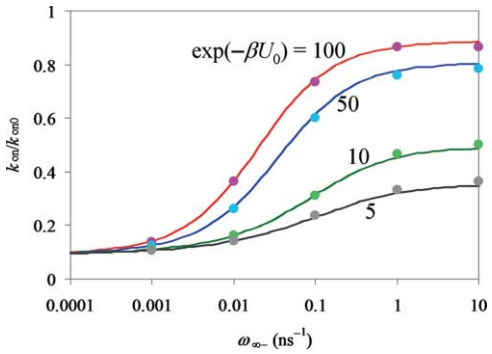

FIG. 4. Comparison of simulation (symbols) and analytical (curves) results for k_{on}/k_{on0}. The model parameters are: $\theta_0 = 180^\circ$, $\kappa = 0.5$ Å/ns, with values of $\exp(-\beta U_0)$ shown in the figure.

TABLE I. Comparison of simulation and theoretical results on k_{CS} and k_{IF} for a range of reactive patch angles.

			$k_{CS}/k(0)$		$k_{IF}/k(0)$	
θ_0 °	$\exp(-\beta U_0)$	κ	Simulation	Theory	Simulation	Theory
3	5	5	0.73	0.665	0.93	0.881
3	10	5	0.69	0.634	0.89	0.826
5	5	5	0.55	0.516	0.87	0.817
5	10	5	0.49	0.468	0.79	0.735
8	5	5	0.38	0.370	0.78	0.738
8	10	5	0.31	0.314	0.67	0.629
10	5	5	0.31	0.306	0.72	0.693
10	10	5	0.24	0.250	0.61	0.573
15	5	5	0.20	0.208	0.62	0.605
15	10	5	0.15	0.158	0.49	0.469

$\exp(-\beta U_0) = 100$; the other for the case where the reactive region spans up to $\theta = 10^\circ$ with $\kappa = 5$ Å/ns and $\exp(-\beta U_0) = 10$. The long-time portion of each $k(t)/k(0)$ curve can be fitted well to a linear function in $(\pi Dt)^{-1/2}$, with the slope m related to the intercept b via $m = b^2 k(0)/4\pi D$, exactly as Eq. (24) predicts. Form here on we focus on the steady-state rate constant k_{on}, which is given by $bk(0)$.

B. Dual-transition-rates model: Spherical symmetry

For the case where the reactive region covers the whole surface of the protein, the analytical result for k_{on} is given by Eq. (27). In Fig. 4 we compare the simulation results for k_{on} against the analytical theory. In the simulations, the reactive region has a finite thickness ε, but in the theory the thickness approaches 0, while keeping the same reactivity κ. To reduce the effect of this difference for treating the reactive region, we scale both the simulation and analytical results by their respective results for k_{on0}. Figure 4 shows that the simulation results for k_{on}/k_{on0} over the full range of $\omega_{\infty-}$ and a range of U_0 are in excellent agreement with the analytical theory.

At a given U_0, as $\omega_{\infty-}$ increases from 0 to ∞, k_{on} increases from the CS limit to the IF limit. Note that $k_{CS}/k_{on0} = p_{\infty c}$. That is why the k_{on}/k_{on0} curves for different U_0 values all start from the same $\omega_{\infty-} \to 0$ limit. However, the $\omega_{\infty-} \to \infty$ limit, k_{IF}/k_{on0}, increases with increasing $|U_0|$.

C. Dual-transition-rates model: Reactive patch

We now turn attention to the model with a small reactive patch. In Table I we compare our simulation and analytical results on k_{CS} and k_{IF} for patch angles at 3°, 5°, 8°, 10°, and 15°. It can be seen that there is close agreement between simulation and theory. The simulation results do seem to be slightly larger than the analytical results. We attribute these small discrepancies to the different treatments of the reaction region in the simulations and in the theory. The effect of the different treatments is magnified when the reactive patch becomes very small.

To show the transition of k_{on} from the CS limit to the IF limit and the difference in magnitude between k_{IF} and k_{CS}, in Figs. 5 and 6 we display k_{on}/k_{CS} as a function of $\omega_{\infty-}$ for a variety of model parameters. Figure 5(a) presents the results for $\exp(-\beta U_0) = 5$ and $\kappa = 5$ Å/ns. As the patch angle decreases from 15° to 3°, k_{IF}/k_{CS} reduces from 3.1 to 1.3. This is the main finding of the present study: k_{IF} and k_{CS} becomes close (i.e., $k_{IF}/k_{CS} \approx 1$) as the reactive patch becomes smaller. Figure 5(b) shows that the closeness of k_{IF} and k_{CS} is affected by the magnitude of the interaction potential. At $\exp(-\beta U_0)$

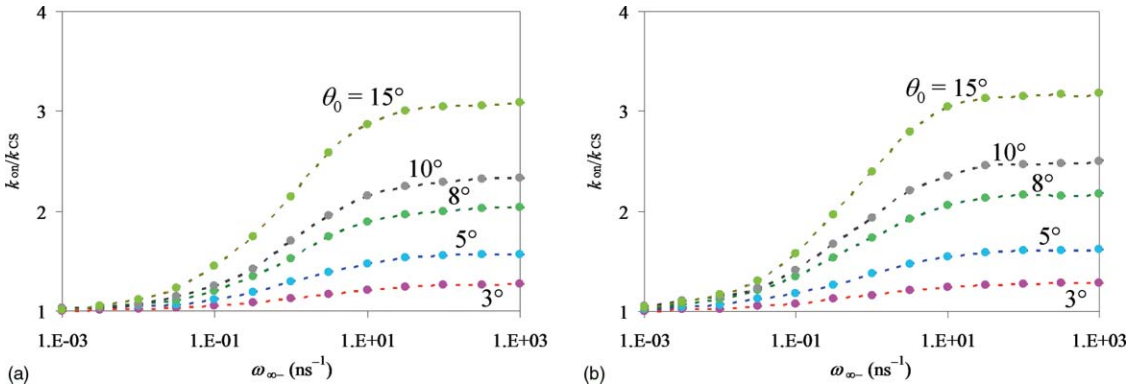

FIG. 5. Dependence of k_{on}/k_{CS} on $\omega_{\infty-}$ for a range of reactive patch angles (shown in the figure). Results for two values of $\exp(-\beta U_0)$: (a) 5 and (b) 10 are shown to illustrate that the difference between the CS and IF limits of k_{on} increases as the magnitude of the interaction potential increases. $\kappa = 5$ Å/ns. Data for the same patch angle are connected by dash.

$= 10$, k_{IF}/k_{CS} increases slightly to 3.2 for $\theta_0 = 15^\circ$ and remains at 1.3 for $\theta_0 = 3^\circ$.

Figure 6 further shows that, with $\exp(-\beta U_0)$ fixed at 5, k_{IF}/k_{CS} increases as the reactivity κ increases. At $\kappa = 50$, k_{IF}/k_{CS} changes from 5.8 to 2.6 as the patch angle decreases from 15° to 3°. In comparison, at $\kappa = 500$, k_{IF}/k_{CS} changes from 6.3 to 4.3 as the patch angle decreases from 15° to 3°.

The simulation results presented here are for the extreme case where the range of the interaction potential is the same as the thickness of the reactive region. The analytical results of Subsections III B and III C show that the approximation $k_{IF}/k_{CS} \approx 1$ improves when the range of the interaction potential widens beyond the reactive region.

VI. DISCUSSION

In the present study we have introduced a general model for protein–ligand binding in which the energy surface of the protein in conformational space is coupled to the binding of its ligand. Two common features of most protein–ligand systems are that the energy surface switches from favoring nonreactive conformations while the ligand is away to favoring reactive conformations while the ligand is near and the reactive region is highly localized.[25,26] We have also presented a method for calculating the binding rate constant from Brownian dynamics simulations.

For a given energy surface, the timescale of the conformational transitions may affect the binding mechanism and the binding rate constant. The binding mechanism appears as conformational selection when conformational transitions are slow relative to the time required for the diffusional approach of the protein–ligand pair, and gradually shifts to manifestly induced fit when the rate of conformational transition increases. The CS and IF limits provide lower and upper bounds, respectively, for the binding rate constant. The difference between k_{IF} and k_{CS} increases as the reactivity increases. At infinite reactivity, we have derived the general result,

$$\frac{k_{IF}}{k_{CS}} \to \frac{p_{\infty o} e^{-\beta U_o(\mathbf{r}_{RR})} + p_{\infty c} e^{-\beta U_c(\mathbf{r}_{RR})}}{p_{\infty c} e^{-\beta U_c(\mathbf{r}_{RR})}} = \frac{1}{p_c(\mathbf{r}_{RR})}, \tag{49}$$

where the reactive region is assumed to be small relative to the range of the interaction potential, and $p_c(\mathbf{r}_{RR})$ is the equilibrium probability of the closed (i.e., reactive) conformation in the reactive region. For systems of actual interest, when the ligand is in the reactive region, the equilibrium probability for the reactive conformation is much higher than that for the nonreactive conformation, i.e., $p_c(\mathbf{r}_{RR}) \to 1$. Hence k_{IF}/k_{CS} is close to 1. Our simulation results on a model with a small reactive patch confirm (in fact, motivated the analytical derivation of) the closeness of k_{IF} and k_{CS}.

The closeness of k_{IF} and k_{CS} in our model is in contrast to the relation between the limits of the rate constant under fast and slow gating in the model of Szabo *et al.*[4] In their model the energy surface of the protein is unaffected by ligand binding. At infinite reactivity, their model predicts $k_{fast}/k_{slow} \to 1/p_{\infty c}$, where $p_{\infty c}$ is the equilibrium probability for the reactive conformation when the ligand is away. For systems of actual interest, one expects that $p_{\infty c} \to 0$; hence $k_{fast} \gg k_{slow}$ in the model of Szabo *et al.* Our model in fact assumes a low equilibrium probability (i.e., $p_{\infty c} \to 0$) for the protein being in the reactive conformation when the ligand is away. However, when near the protein, intermolecular interactions place the ligand on the $U_c(\mathbf{r})$ potential, which facilitates the diffusion of the ligand toward the binding site and therefore enhances the binding rate constant. Consequently, the difference between the limits (k_{CS} and k_{IF}) of the rate constant under slow and fast conformational transitions is significantly reduced.

Because k_{CS} and k_{IF} are lower and upper bounds of the binding rate constant k_{on}, when they are close they provide good estimates of k_{on}. This is significant because the calculation of k_{on} involves expensive simulations of protein conformational fluctuation while running the Brownian dynamics simulations of the ligand translational motion. In contrast, for the calculations of both k_{CS} and k_{IF}, the ligand experiences a fixed interaction potential and hence no simulation of protein conformational fluctuation is necessary. For k_{CS}, the interaction potential is $U_c(\mathbf{r})$ as the protein is fixed in the reactive conformation. For k_{IF}, the interaction potential is $U_{eff}(\mathbf{r})$ as it is assumed that, for each give ligand position $\mathbf{r}$, the protein instantaneously adopts an equilibrium distribution in conformational space. The closeness of k_{CS} and k_{IF} and how good estimates they provide for k_{on} will be investigated in the future with realistic representations of protein–ligand systems.

If conformational selection and induced fit can achieve similar rate constants, is it still significant to distinguish the two binding mechanisms? If so, what kinetic experiments can ascertain the binding mechanism to be one but not the other? These mechanistic questions remain to be addressed.

Finally, we note that the formulation of $k(t)$ depends on a single protein–ligand pair. As we pointed out in Subsection II A, $k(t)$ is unaffected whether the conformational coordinate x refers to the protein or the ligand. That is, the same $k(t)$ is obtained whether it is the protein or the ligand that undergoes conformational change. However, in experiments, rate coefficients can be measured only when many protein–ligand pairs are present. Consider the case where one species is in excess (with concentration C) so that the binding kinetics is pseudo-first-order. If the ligand is the species that is both in excess and undergoing conformational change, then the many ligand molecules around a single protein molecule are independent. The probability that the protein escapes binding with all the ligand molecules is[27]

$$S(t) = e^{-C\int_0^t dt_1 k(t_1)}. \tag{50}$$

However, when the ligand is in excess but the protein undergoes conformational change (or *vice versa*), then Eq. (50) is no longer valid, except in the limit of fast conformational change. That is because now the ligand molecules are no longer independent: they experience the same change in reaction condition at the same time whenever the protein goes through a conformational transition. Therefore one has to deal with a many-body problem. For the situation where conformational change is uncoupled to ligand binding, Zhou and Szabo[27] have derived an approximate approach for calculating the many-body $S(t)$, with $k(t)$ from the pair problem as

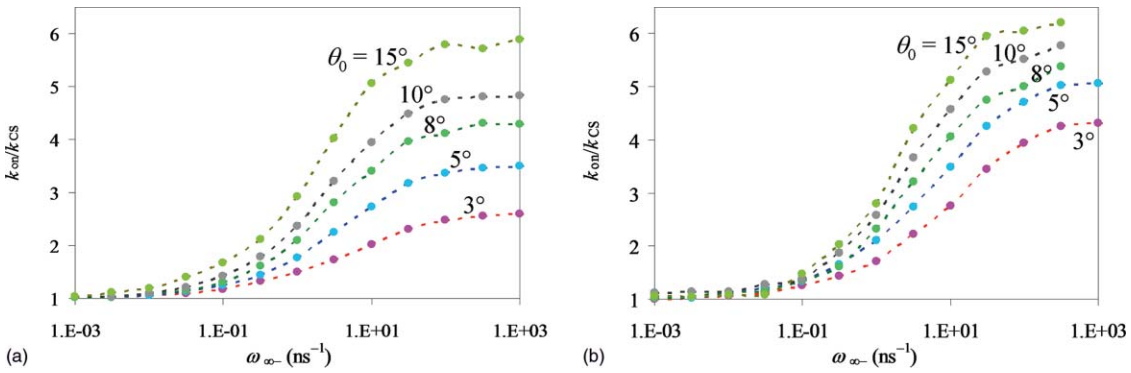

FIG. 6. Dependence of k_{on}/k_{CS} on $\omega_{\infty-}$ for a range of reactive patch angles. Results for two values of κ: (a) 50 and (b) 500 Å/ns are shown to illustrate that the difference between the CS and IF limits of k_{on} increases as the reactivity increases. $\exp(-\beta U_0) = 5$.

input. When conformational change is coupled to ligand binding we anticipate that the many-body binding kinetics will have much richer features. In particular, using analyses based on ordinary chemical kinetics, Hammes *et al.*[28] proposed that increasing ligand concentration can shift the binding mechanism at the many-body level from conformational selection to induced fit. We plan to investigate the many-body problem in the future.

ACKNOWLEDGMENTS

This work was supported in part by Grant No. GM58187 from the National Institutes of Health (NIH).

APPENDIX A: GOVERNING EQUATION FOR DISCRETE CONFORMATIONS

Here we derive Eqs. (18a) and (18b), the governing equation for the pair distribution function when the transitions between the two conformations of the protein are treated as rate processes. To find $P_o(\mathbf{r}, t)$, we integrate both sides of Eq. (5) over x from $-\infty$ to $x^\ddagger(\mathbf{r})$, obtaining

$$\frac{\partial P_o(\mathbf{r},t)}{\partial t} = \int_{-\infty}^{x^\ddagger(\mathbf{r})} dx \nabla \cdot D e^{-\beta U(\mathbf{r},x)} \nabla e^{\beta U(\mathbf{r},x)} P(\mathbf{r},x,t) + \int_{-\infty}^{x^\ddagger(\mathbf{r})} dx \frac{\partial}{\partial x} D_1 e^{-\beta U(\mathbf{r},x)} \frac{\partial}{\partial x} e^{\beta U(\mathbf{r},x)} P(\mathbf{r},x,t). \tag{A1}$$

The third term on the right hand side is absent because $H(\mathbf{r}, x)$ is zero when $x < x^\ddagger(\mathbf{r})$ (formation of the native complex requires that the protein be in the closed, i.e., reactive conformation). Since we assume fast equilibration within the open energy well, we have

$$P(\mathbf{r},x,t) \approx P_{oeq}(x|\mathbf{r}) P_o(\mathbf{r},t), \quad x < x^\ddagger(\mathbf{r}), \tag{A2}$$

where $P_{oeq}(x|\mathbf{r})$ is the equilibrium distribution for x in $x < x^\ddagger(\mathbf{r})$ with a given $\mathbf{r}$:

$$P_{oeq}(x|\mathbf{r}) = \frac{e^{-\beta U(\mathbf{r},x)}}{\int_{-\infty}^{x^\ddagger(\mathbf{r})} dx e^{-\beta U(\mathbf{r},x)}} = \frac{e^{-\beta U(\mathbf{r},x)}}{e^{-\beta U_o(\mathbf{r})} \int_{-\infty}^{x^\ddagger_\infty} dx e^{-\beta U_\infty(x)}}. \tag{A3}$$

Combining the last two equations, we find

$$e^{\beta U(\mathbf{r},x)} P(\mathbf{r},x,t) \approx \frac{e^{\beta U_o(\mathbf{r})} P_o(\mathbf{r},t)}{\int_{-\infty}^{x^\ddagger_\infty} dx e^{-\beta U_\infty(x)}}, \quad x < x^\ddagger(\mathbf{r}). \tag{A4}$$

The first term on the right hand side of Eq. (A1) now becomes

$$\int_{-\infty}^{x^\ddagger(\mathbf{r})} dx \nabla \cdot D e^{-\beta U(\mathbf{r},x)} \nabla e^{\beta U(\mathbf{r},x)} P(\mathbf{r},x,t) \approx \int_{-\infty}^{x^\ddagger(\mathbf{r})} dx \nabla \cdot D \frac{e^{-\beta U(\mathbf{r},x)}}{\int_{-\infty}^{x^\ddagger_\infty} dx e^{-\beta U_\infty(x)}} \nabla e^{\beta U_o(\mathbf{r})} P_o(\mathbf{r},t).$$

Now the integration over x is dominated by the region around $x = x_o(\mathbf{r})$, and hence the precise value of the upper bound of the integration is unimportant; therefore we may switch the order of the integration over x and the gradient operator on $\mathbf{r}$, leading to

$$\begin{aligned}\int_{-\infty}^{x^\ddagger(\mathbf{r})} dx \nabla \cdot D e^{-\beta U(\mathbf{r},x)} \nabla e^{\beta U(\mathbf{r},x)} P(\mathbf{r},x,t) &\approx \nabla \cdot D \frac{\int_{-\infty}^{x^\ddagger(\mathbf{r})} dx e^{-\beta U(\mathbf{r},x)}}{\int_{-\infty}^{x^\ddagger_\infty} dx e^{-\beta U_\infty(x)}} \nabla e^{\beta U_o(\mathbf{r})} P_o(\mathbf{r},t) \\ &= \nabla \cdot D e^{-\beta U_o(\mathbf{r})} \nabla e^{\beta U_o(\mathbf{r})} P_o(\mathbf{r},t).\end{aligned}$$

To treat the second term on the right hand side of Eq. (A1), we follow the well-known reduction of the diffusive barrier crossing problem to a rate-equation description. The second term is essentially the rate of change in the probability of the inactive conformation. In the rate-equation description, this is

$$-\omega_+(\mathbf{r}) P_o(\mathbf{r},t) + \omega_-(\mathbf{r}) P_c(\mathbf{r},t),$$

where $\omega_\pm(\mathbf{r})$ are the rate constants introduced in Eq. (14a) and (14b). Combining the results for the two terms, we find

$$\frac{\partial P_o(\mathbf{r},t)}{\partial t} = \nabla \cdot D e^{-\beta U_o(\mathbf{r})} \nabla e^{\beta U_o(\mathbf{r})} P_o(\mathbf{r},t) - \omega_+(\mathbf{r}) P_o(\mathbf{r},t) + \omega_-(\mathbf{r}) P_c(\mathbf{r},t). \tag{A5}$$

The equation for the distribution function $P_c(\mathbf{r}, t)$ when the protein is in the closed conformation can be similarly derived. Now the integration over x is from $x^\ddagger(\mathbf{r})$ to ∞ and we need to assume

$$P(\mathbf{r}, x, t) \approx P_{\text{ceq}}(x|\mathbf{r})P_c(\mathbf{r}, t), \quad x > x^\ddagger(\mathbf{r}), \tag{A6}$$

where

$$P_{\text{ceq}}(x|\mathbf{r}) = \frac{e^{-\beta U(\mathbf{r},x)}}{\int_{x^\ddagger(\mathbf{r})}^{\infty} dx e^{-\beta U(\mathbf{r},x)}}. \tag{A7}$$

In the closed conformation, formation of the native complex can occur when the ligand is in the reactive region. Correspondingly the third term on the right hand side of Eq. (5) yields

$$\int_{x^\ddagger(\mathbf{r})}^{\infty} dx \tau^{-1} H(\mathbf{r}, x) P(\mathbf{r}, x, t) \approx \int_{x^\ddagger(\mathbf{r})}^{\infty} dx \tau^{-1} H(\mathbf{r}, x) P_{\text{ceq}}(x|\mathbf{r}) \times P_c(\mathbf{r}, t) \equiv \tau^{-1} H_c(\mathbf{r}) P_c(\mathbf{r}, t),$$

where

$$H_c(\mathbf{r}) = \frac{\int_{x^\ddagger(\mathbf{r})}^{\infty} dx H(\mathbf{r}, x) e^{-\beta U(\mathbf{r},x)}}{\int_{x^\ddagger(\mathbf{r})}^{\infty} dx e^{-\beta U(\mathbf{r},x)}}. \tag{A8}$$

Finally, the equation for $P_c(\mathbf{r}, t)$ is

$$\begin{aligned} \frac{\partial P_c(\mathbf{r}, t)}{\partial t} &= \nabla \cdot D e^{-\beta U_c(\mathbf{r})} \nabla e^{\beta U_c(\mathbf{r})} P_c(\mathbf{r}, t) \\ &\quad + \omega_+(\mathbf{r}) P_o(\mathbf{r}, t) - \omega_-(\mathbf{r}) P_c(\mathbf{r}, t) \\ &\quad - \tau^{-1} H_c(\mathbf{r}) P_c(\mathbf{r}, t). \end{aligned} \tag{A9}$$

APPENDIX B: CLOSENESS OF k_{CS} AND k_{IF} FOR CONTINUOUS CONFORMATIONS

Here we show that k_{CS} and k_{IF} are close when conformational change is treated as continuous diffusion along a one-dimensional coordinate, provided that the reactive region is small relative to the range of the interaction potential. As in the main text, we use $\mathbf{r}_{RR}$ to denote a representative point in the reactive region. When the ligand is in the reactive region, the closed conformation of the protein is specified by $h_c(x) = 1$. We focus the calculation of k_{CS} and k_{IF} on an infinite reactivity, since this is when the difference between them is the greatest.

In calculating k_{CS}, one first assumes that the protein is fixed in conformation x; the resulting rate constant is then averaged over the equilibrium distribution, $P_{eq}(x)$, of x. For a fixed protein conformation x, the interaction potential that the ligand experiences is $\Delta U(\mathbf{r}|x) \equiv U(\mathbf{r}, x) - U_\infty(x)$, where the subtraction by $U_\infty(x)$ is made so that the interaction potential approaches 0 at infinite protein–ligand separation. When the reactivity is infinite, we have

$$k_{CS} \to \int_{-\infty}^{\infty} dx P_{eq}(x) h_c(x) k_D[\Delta U(\mathbf{r}|x)], \tag{B1}$$

where $h_c(x)$ is inserted to indicate the fact that binding occurs only if the protein starts from a closed conformation [as specified by $h_c(x) = 1$]; and $k_D[\Delta U(\mathbf{r}|x)]$ is the rate constant when the protein conformation is fixed at x, the interaction potential is $\Delta U(\mathbf{r}|x)$, and the reactivity is infinite. For calculating k_{IF}, the ligand experiences the conformation-averaged effective potential $U_{\text{eff}}(\mathbf{r})$ given by Eq. (42a). When the reactivity is infinite,

$$k_{IF} \to k_D[U_{\text{eff}}(\mathbf{r})], \tag{B2}$$

where $k_D[U_{\text{eff}}(\mathbf{r})]$ is the rate constant when the protein does not under conformational change, the interaction potential is $U_{\text{eff}}(\mathbf{r})$, and the reactivity is infinite.

To proceed further, we now use the approximate dependence of the $k_D[U(\mathbf{r})]$ on the interaction potential $U(\mathbf{r})$, given by Eq. (43). Using this approximation, Eq. (B1) becomes

$$\begin{aligned} k_{CS} &\to k_{D0} \int_{-\infty}^{\infty} dx P_{eq}(x) h_c(x) e^{-\beta[U(\mathbf{r}_{RR},x) - U_\infty(x)]} \\ &= k_{D0} \frac{\int_{-\infty}^{\infty} dx h_c(x) e^{-\beta U(\mathbf{r}_{RR},x)}}{\int_{-\infty}^{\infty} dx e^{-\beta U_\infty(x)}}, \end{aligned} \tag{B3}$$

where we have used the expression for $P_{eq}(x)$ given by Eq. (6). On the other hand, Eq. (B2) becomes

$$\begin{aligned} k_{IF} &\to k_{D0} e^{-\beta U_{\text{eff}}(\mathbf{r}_{RR})} \\ &= k_{D0} \frac{\int_{-\infty}^{\infty} dx e^{-\beta U(\mathbf{r}_{RR},x)}}{\int_{-\infty}^{\infty} dx e^{-\beta U_\infty(x)}} \\ &= k_{D0} \frac{\int_{-\infty}^{\infty} dx h_c(x) e^{-\beta U(\mathbf{r}_{RR},x)}}{\int_{-\infty}^{\infty} dx e^{-\beta U_\infty(x)}} \\ &\quad + k_{D0} \frac{\int_{-\infty}^{\infty} dx [1 - h_c(x)] e^{-\beta U(\mathbf{r}_{RR},x)}}{\int_{-\infty}^{\infty} dx e^{-\beta U_\infty(x)}}. \end{aligned} \tag{B4}$$

Note that the first term is the same as Eq. (B3). In addition, we also expect the first term of Eq. (B4) to dominate over its second term, since the closed (i.e., reactive) conformation should dominate over the open (i.e., nonreactive) conformation when the ligand is in the reactive region. Therefore, k_{CS} and k_{IF} are close even for an infinite reactivity as long as the interaction potential varies slowly over the reactive region and beyond.

[1] R. Alsallaq and H.-X. Zhou, Proteins **71**, 320 (2008).
[2] S. H. Northrup, S. A. Allison, and J. A. McCammon, J. Chem. Phys. **80**, 1517 (1984).
[3] H. X. Zhou, Q. Rev. Biophys. **43**, 219 (2010).
[4] A. Szabo, D. Shoup, S. H. Northrup, and J. A. McCammon, J. Chem. Phys. **77**, 4484 (1982).
[5] N. Agmon, J. Theor. Biol. **113**, 711 (1985).
[6] H.-X. Zhou, J. Chem. Phys. **108**, 8146 (1998).
[7] H.-X. Zhou and A. Szabo, Biophys. J. **71**, 2440 (1996).
[8] R. C. Wade, M. E. Davis, B. A. Luty, J. D. Madura, and J. A. McCammon, Biophys. J. **64**, 9 (1993).
[9] G. H. Peters, O. H. Olsen, A. Svendsen, and R. C. Wade, Biophys. J. **71**, 119 (1996).
[10] C. E. Chang, T. Shen, J. Trylska, V. Tozzini, and J. A. McCammon, Biophys. J. **90**, 3880 (2006).
[11] A. A. Gorfe, C. E. Chang, I. Ivanov, and J. A. McCammon, Biophys. J. **94**, 1144 (2008).
[12] R. V. Swift and J. A. McCammon, J. Am. Chem. Soc. **131**, 5126 (2009).
[13] H. X. Zhou, S. T. Wlodek, and J. A. McCammon, Proc. Natl. Acad. Sci. U.S.A **95**, 9280 (1998).
[14] A. A. Gorfe, B. Lu, Z. Yu, and J. A. McCammon, Biophys. J. **97**, 897 (2009).

[15]R. C. Wade, B. A. Luty, E. Demchuk, J. D. Madura, M. E. Davis, J. M. Briggs, and J. A. McCammon, Nat. Struct. Biol. **1**, 65 (1994).
[16]J. H. Kim and S. Lee, J. Chem. Phys. **131**, 014503 (2009).
[17]H.-X. Zhou, Biophys. J. **98**, L15 (2010).
[18]M. V. Smoluchowski, Z. Phys. Chem. **92**, 129 (1917).
[19]H.-X. Zhou, J. Phys. Chem. **94**, 8794 (1990).
[20]H. A. Kramers, Physica **7**, 284 (1940).
[21]H. X. Zhou, Biophys. J. **73**, 2441 (1997).
[22]D. Shoup, G. Lipari, and A. Szabo, Biophys. J. **36**, 697 (1981).
[23]H.-X. Zhou, J. Chem. Phys. **105**, 7235 (1996).
[24]D. L. Ermak and J. A. McCammon, J. Chem. Phys. **69**, 1352 (1978).
[25]A. Eletsky, A. Kienhofer, D. Hilvert, and K. Pervushin, Biochemistry **44**, 6788 (2005).
[26]S. M. Sullivan and T. Holyoak, Proc. Natl. Acad .Sci. U.S.A. **105**, 13829 (2008).
[27]H.-X. Zhou and A. Szabo, J. Phys. Chem. **100**, 2597 (1996).
[28]G. G. Hammes, Y. C. Chang, and T. G. Oas, Proc. Natl. Acad. Sci. U.S.A. **106**, 13737 (2009).