Supplementary Information for

Variational Implicit-Solvent Predictions of the Dry-Wet Transition Pathways for Ligand-Receptor Binding and Unbinding Kinetics

Shenggao Zhou, R. Gregor Weiß, Li-Tien Cheng, Joachim Dzubiella, J. Andrew McCammon, and Bo Li

J. Andrew McCammon (Email: jmccammon@ucsd.edu) and Bo Li (Email: bli@math.ucsd.edu)

This PDF file includes:

Supplementary text
Figs. S1 to S5
Table S1
References for SI reference citations

Supporting Information Text

Table of contents.

Abbreviations. BD: Brownian dynamics. CTMC: continuous-time Markov chain. FPE: Fokker–Planck equation. LJ: Lennard-Jones. MD: molecular dynamics. MEP: minimum energy path. MFPT: mean first-passage time. PMF: potential of mean force. vdW: van der Waals. VISM: variational implicit-solvent model.

1. The Level-Set Method for Minimizing the VISM Solvation Free-Energy Functional

We consider the solvation of solute molecules with all the solute atoms located at $\mathbf{r}_1, \dots, \mathbf{r}_N$ in an aqueous solvent. A solute-solvent interface Γ is a closed surface that encloses all the solute atoms but no solvent molecules. The interior and exterior of such a surface Γ, denoted by Ω_{m} and Ω_{w}, are termed the solute and solvent regions, respectively. In the variational implicit-solvent model (VISM), we minimize the solvation free-energy functional (cf. Eq. [2] in the main text) (1, 2)

$$G[\Gamma] = \Delta P \, \mathrm{vol}\,(\Omega_{\mathrm{m}}) + \gamma_0 \int_\Gamma (1 - 2\tau H) \, dS + \rho_0 \sum_{i=1}^{N} \int_{\Omega_{\mathrm{w}}} U_i(|\mathbf{r} - \mathbf{r}_i|) \, dV + G_{\mathrm{e}}[\Gamma] \qquad [1]$$

among all the solute-solvent interfaces Γ. The parameters ΔP, γ_0, τ, and ρ_0 are the difference of pressures across Γ, the surface tension constant for a planar solute-solvent interface, the curvature correction coefficient (i.e., the Tolman length), and the bulk solvent density, respectively. In Eq. [1], H is the local mean curvature and each U_i is a 12-6 Lennard-Jones (LJ) potential with parameters σ_i and ε_i. We shall set the electrostatic part $G_{\mathrm{e}}[\Gamma] = 0$ in this study. But we will make a remark at the end of this section on the full VISM with the electrostatics. We call a solute-solvent interface a VISM surface if it minimizes (locally) the VISM functional Eq. [1], i.e., if it is a stable equilibrium. A VISM surface is dry, representing a dry hydration state, if it loosely wraps up all the solute atoms with enough space for a few solvent molecules, or wet, representing a wet hydration state, if it tightly wraps up all the solute atoms without extra space for a solvent molecule.

We have designed and implemented a robust level-set method to numerically minimize the VISM solvation free-energy functional Eq. [1] in the three-dimensional setting (3–8). Beginning with an initial solute-solvent interface that may have a large value of solvation free energy, our level-set method moves the interface in the direction of steepest descent of the VISM solvation free energy step by step until a VISM surface is reached. The (normal component of the) boundary force that moves the interface is given by the negative first variation, $F_n = -\delta_\Gamma G[\Gamma]$, of the VISM solvation free-energy functional Eq. [1] (with $G_{\mathrm{e}}[\Gamma] = 0$) (3, 7):

$$F_n(\mathbf{r}) = -\Delta P - 2\gamma_0 [H(\mathbf{r}) - \tau K(\mathbf{r})] + \rho_0 \sum_{i=1}^{N} U_i(|\mathbf{r} - \mathbf{r}_i|) \qquad \forall \mathbf{r} \in \Gamma, \qquad [2]$$

where $K(\mathbf{r})$ is the Gaussian curvature at $\mathbf{r}$. As our level-set method is an optimization method of the steepest descent type, different initial interfaces are relaxed to different VISM surfaces, often representing different hydrations states. We often use the following two types of initial interfaces: a tight wrap that is a surface of the union of van der Waals (vdW) spheres centered at solute atoms with reduced radii; and a loose wrap that is a large surface loosely enclosing all the solute atoms.

To apply the level-set method (9–11) to minimizing the functional Eq. [1], we represent a solute-solvent interface Γ as the zero level set (i.e., level surface) of a function $\phi = \phi(\mathbf{r})$ (called a level-set function), i.e., $\Gamma = \{\mathbf{r} : \phi(\mathbf{r}) = 0\}$. We keep a level-set function to be negative and positive inside and outside the interface Γ, respectively. The unit normal $\mathbf{n}$ pointing from the solute to solvent region, the mean curvature H, and the Gaussian curvature K at a point

$\mathbf{r}$ on the interface can be readily expressed as $\mathbf{n} = \nabla\phi/|\nabla\phi|$, $H = (1/2)\nabla \cdot \mathbf{n}$, and $K = \mathbf{n} \cdot \mathrm{adj}\,(\nabla^2\phi)\mathbf{n}$, respectively. Here, $\nabla^2\phi$ is the Hessian matrix of the function ϕ with entries being the second order partial derivatives $\partial^2_{ij}\phi$ of the level-set function ϕ, and $\mathrm{adj}\,(\nabla^2\phi)$ is the adjoint matrix of the Hessian $\nabla^2\phi$. The motion of the interface $\Gamma = \Gamma(t)$, where t denotes the relaxation time, is then tracked by locating the level set of the corresponding level-set function $\phi = \phi(\mathbf{r}, t)$ that solves the so-called level-set equation

$$\partial_t\phi + F_n|\nabla\phi| = 0, \qquad [3]$$

where the boundary force F_n, given in Eq. [2], is extended to the entire computational box or a band centered around the interface Γ. We start from an initial level-set function ϕ_0 at $t = 0$ and solve the equation by iteration in time until a steady-state solution is reached. To avoid the gradient $\nabla\phi$ being too small which can lead to numerical instability in locating the interface, we reinitialize the level-set function ϕ every few time steps in iteration. The reinitialization is done by solving

$$\partial_t\phi + \mathrm{sign}\,(\phi_0)(|\nabla\phi| - 1) = 0, \qquad [4]$$

where ϕ_0 is the level-set function before reinitialization, $\mathrm{sign}\,(\phi_0)$ is the sign of ϕ_0, and the time t can be different from that in the original level-set equation [3]. See (3, 5–7) for more details.

We remark that the electrostatic part of the solvation free energy, $G_{\mathrm{e}}[\Gamma]$, can be included as the Coulomb-field approximation (CFA) (12, 13) or the dielectric-boundary Poisson–Boltzmann (PB) electrostatic free energy (14–16). The CFA does not include the ionic effect but is efficient as it requires no numerical solution of partial differential equations. The PB free energy is determined by the electrostatic potential that is the unique solution to a boundary-value problem of the dielectric-boundary PB equation. Explicit formula of the (normal component of the) dielectric-boundary force, defined as the negative variation $-\delta_\Gamma G_{\mathrm{e}}[\Gamma]$, has been obtained (17–19). We have implemented both CFA and PB electrostatics; cf. (6–8).

2. The Level-Set Implementation of the VISM-String Method

Let us fix all the solute atoms $\mathbf{r}_i$ $(i = 1, \ldots, N)$ and consider two different VISM surfaces Γ_0 and Γ_1, represented by two level-set functions ϕ_0 and ϕ_1, respectively. We use the string method (20–22) to find minimum energy paths (MEPs) that connect these two states. A string or path here is a family of solute-solvent interfaces $\{\Gamma_\alpha\}_{\alpha\in[0,1]}$, or their corresponding level-set functions $\{\phi_\alpha\}_{\alpha\in[0,1]}$, that connect the two states Γ_0 and Γ_1, or their level-set functions ϕ_0 and ϕ_1. A MEP here is a string that is orthogonal to the level surfaces of the VISM solvation free-energy functional. In the level-set formulation, a MEP can be obtained by solving for a steady-state solution of the equation for the level-set function $\phi_\alpha = \phi_\alpha(x, t)$

$$\partial_t\phi_\alpha = -F_n(\phi_\alpha)|\nabla\phi_\alpha| + \lambda_\alpha \frac{\partial_\alpha\phi_\alpha}{\|\partial_\alpha\phi_\alpha\|} \qquad \text{for each } \alpha \in (0, 1),$$

together with a given initial string $\{\phi^{(0)}_\alpha\}_{\alpha\in[0,1]}$ that connects ϕ_0 and ϕ_1, Here, the normal component of the boundary force $F_n(\phi_\alpha) = -\delta_\Gamma G[\Gamma_\alpha]$ (with ϕ_α being a zero level-set of Γ_α) is given in Eq. [2], $\partial_\alpha\phi_\alpha/\|\partial_\alpha\phi_\alpha\|$ is the unit vector tangential to the string, the constant λ_α is a Lagrange multiplier for enforcing particular parameterization (e.g., equal arc-length or energy weighted arc-length parameterization) of the string, and $\|\cdot\|$ denotes the $L^2(\Omega)$-norm.

Let us focus now on the model ligand-pocket system (cf. Fig. 1 in the main text) with a fixed reaction coordinate z. We implement a simplified version of the string method (21) to numerically find a MEP connecting two hydration states Γ_0 and Γ_1, with their level-set functions ϕ_0 and ϕ_1, respectively. To do so, we select some integer $M \geq 2$ and discretize the parameter $\alpha \in [0, 1]$ by $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_M < \alpha_{M+1} = 1$, and consider the corresponding level-set functions ϕ_{α_j} $(j = 0, 1, \ldots, M + 1)$ that represent some solute-solvent interfaces. Each ϕ_{α_j} is called an image. These images are discrete points of a string or path connecting ϕ_0 and ϕ_1. They are updated iteratively to reach a stable steady state, representing a MEP. We set the initial images for the iteration to be

$$\phi^{(0)}_{\alpha_j} = \phi_0 + \alpha_j(\phi_1 - \phi_0) \qquad (j = 1, \ldots, M). \qquad [5]$$

Each iteration is a two-step process: relaxation and redistribution. Suppose we know all the interior images $\phi^{(k)}_{\alpha_j}$ $(j = 1, \ldots, M)$ after the kth iteration. In the first step, we solve the level-set equation [3] for each j $(1 \leq j \leq M)$ with the initial function $\phi^{(k)}_{\alpha_j}$ but only for one time step, followed by the reinitialization (cf. Eq. [4]), and obtain a solution $\phi^*_{\alpha_j}$. These images $\phi^*_{\alpha_j}$ $(j = 1, \ldots, M)$ should make the new string "closer" to being normal to the free-energy level surfaces, but may also cluster around the two states ϕ_0 and ϕ_1, as they are local minimizers of the VISM solvation free-energy functional. In the second step, we redistribute these intermediate images by linear interpolation to generate new and well-separated images $\phi_{\alpha_j}^{(k+1)}$. More precisely, we set $s_0 = 0$ and $s_j = s_{j-1} + \|\phi^*_{\alpha_j} - \phi^*_{\alpha_{j-1}}\|$ $(j = 1, \dots, M+1)$, where $\phi^*_{\alpha_0} = \phi_0$ and $\phi^*_{\alpha_{M+1}} = \phi_1$. We also set $\alpha^*_j = s_j/s_M$ $(j = 0, 1, \dots, M+1)$. For each j $(1 \le j \le M)$, we find the unique i $(1 \le i \le M+1)$ that depends on j such that $\alpha^*_{i-1} \le \alpha_j < \alpha^*_i$. We then calculate $\phi_{\alpha_j}^{(k+1)}$ by the linear interpolation

$$\phi_{\alpha_j}^{(k+1)} = \phi^*_{\alpha_{i-1}} + \frac{\alpha_j - \alpha^*_{i-1}}{\alpha^*_i - \alpha^*_{i-1}}(\phi^*_{\alpha_i} - \phi^*_{\alpha_{i-1}}). \qquad [6]$$

Once the iteration converges to a MEP, we find an interior image that has the largest VISM solvation free energy among all the images, and identify it as a saddle point. Note that different initial images may lead to different MEPs; cf. Fig. 3 in the main text.

Algorithm of a Simplified String Method.

Step 1. Input all the parameters ΔP, γ_0, τ, ρ_0, and $\mathbf{r}_i$, σ_i, and ε_i for all $i = 1, \dots, N$. Input the level-set functions ϕ_0 and ϕ_1 for the two states. Input M, the number of (interior) images in the string, the parameters α_j $(j = 0, 1, \dots, M+1)$ for the string images, and the initial (interior) image level-set functions $\phi_j^{(0)}$ $(j = 1, \dots, M)$; cf. Eq. [5]. Input the time step Δt. Set the iteration counter $k = 0$.

Step 2. Given the interior images $\phi_{\alpha_j}^{(k)}$ $(j = 1, \dots, M)$. For each j $(1 \le j \le M)$, solve the level-set equation [3] using the initial solution $\phi_{\alpha_j}^{(k)}$ for one time step to obtain the image $\bar{\phi}_{\alpha_j}$. Compute the image $\phi^*_{\alpha_j}$ by solving the reinitialization equation [4] with $\bar{\phi}_{\alpha_j}$ as the initial solution.

Step 3. Compute the arc lengths $s_0 = 0$ and $s_j = s_{j-1} + \|\phi^*_{\alpha_j} - \phi^*_{\alpha_{j-1}}\|$ $(j = 1, \dots, M+1)$ and the parameters $\alpha^*_j = s_j/s_M$ $(j = 0, 1, \dots, M+1)$. Generate the images $\phi_{\alpha_j}^{(k+1)}$ $(j = 1, \dots, M)$ by Eq. [6].

Step 4. Check the stopping criteria. If failed, set $k := k+1$ and go to Step 2.

To find possible multiple MEPs connecting the two states ϕ_0 and ϕ_1, we can alternatively apply the climbing string method (23) to first find saddle points near ϕ_0. In implementation, we fix the first image ϕ_0 but allow the last image to climb uphill in the direction tangental to the string. The string converges when the last image approaches a saddle point close to the starting state ϕ_0. Usually, we use more images close to the last one to more efficiently find a saddle point. Once a saddle point is found, we then relax it to a level-set function representing a VISM surface. If this function is ϕ_1, then we can use the simplified string method described above, in which we keep the saddle point as an image during the iteration, to find an MEP that connects these two states ϕ_0 and ϕ_1, and that passes through the found saddle point. Otherwise, we start over with different initial images. Since we usually have at most three significant hydration states for each reaction coordinate, we can efficiently find multiple MEPs (if exist) connecting these states.

Algorithm of a Climbing String Method.

Step 1. Input all the parameters ΔP, γ_0, τ, ρ_0, and $\mathbf{r}_i$, σ_i, and ε_i for all $i = 1, \dots, N$. Input a level-set function ϕ_0 for a VISM surface. Input M, with $M+2$ the number of images in the string, the parameters $\{\alpha_j\}_{j=0}^{M+1}$ for the string images with $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_{M+1} < 1$, and the initial image level-set functions $\{\phi_{\alpha_j}^{(0)}\}_{j=1}^{M+1}$. Input the time step Δt. Set the iteration counter $k = 0$.

Step 2. Given the images $\phi_{\alpha_j}^{(k)}$ $(j = 1, \dots, M+1)$. For each j $(1 \le j \le M+1)$, solve the level-set equation [3] using the initial solution $\phi_{\alpha_j}^{(k)}$ for one time step to obtain an image $\bar{\phi}_j$. Solve the reinitialization equation [4] using the initial solution $\bar{\phi}_j$ for one time step to obtain an image ϕ^*_j.

Step 3. Update the last image

$$\phi_{\alpha_{M+1}}^{(k+1)} = \phi^*_{M+1} - 2\langle \phi^*_{M+1} - \phi_{\alpha_{M+1}}^{(k)}, \hat{\tau}_{M+1}\rangle \hat{\tau}_{M+1} \quad \text{with} \quad \hat{\tau}_{M+1} = \frac{\phi_{\alpha_{M+1}}^{(k)} - \phi_{\alpha_M}^{(k)}}{\|\phi_{\alpha_{M+1}}^{(k)} - \phi_{\alpha_M}^{(k)}\|},$$

where $\langle \cdot, \cdot \rangle$ denotes the $L^2(\Omega)$-inner product.

Step 4. Compute the arc lengths $s_0 = 0$ and $s_j = s_{j-1} + \|\phi^*_{\alpha_j} - \phi^*_{\alpha_{j-1}}\|$ $(j = 1, \dots, M+1)$, and set $\alpha^*_j = s_j/s_M$ $(j = 0, 1, \dots, M+1)$. Update the other images to obtain $\phi_{\alpha_j}^{(k+1)}$ $(j = 1, \dots, M)$ by Eq. [6].

Step 5. Check the stopping criteria. If failed, set $k := k+1$ and go to Step 2.

3. Algorithms for Brownian Dynamics Simulations of the Ligand Stochastic Motion

In the absence of the pocket dry-wet fluctuations, the random position $z = z(t)$ (also denoted z_t) can be determined by the standard Brownian dynamics (BD) simulations that solve numerically the stochastic differential equation

$$dz_t = \left[-\frac{1}{k_{\mathrm{B}}T} D(z_t) V'(z_t) + D'(z_t) \right] dt + \sqrt{2D(z_t)}\, d\xi_t, \qquad [7]$$

together with a given initial position $z(0)$, where $V(z)$ is the equilibrium potential of mean force (PMF) (defined in Eq. [1] and plotted in Fig. 2 (B), both in the main text), ξ_t is the standard Brownian motion, and a prime stands for derivative. The effective and position-dependent diffusion coefficient $D = D(z)$ is a smooth interpolation of the diffusion constants D_{in} and D_{out} for the ligand inside and outside the pocket, respectively. It is given by

$$D(z) = \frac{D_{\mathrm{in}} + D_{\mathrm{out}}}{2} - \frac{D_{\mathrm{in}} - D_{\mathrm{out}}}{2} \tanh\left[\nu\,(z - z_{\mathrm{c}})\right], \qquad [8]$$

where $\nu > 0$ is a parameter that controls the width of the transition from D_{in} to D_{out} and z_{c} is a threshold reaction coordinate distinguishing the ligand being inside or outside the pocket. Solutions to Eq. [7] are constrained by $z(t) \in [z_{\mathrm{L}}, z_{\mathrm{R}}]$ for all t for some boundaries z_{L} and z_{R}, with z_{L} close to the pocket and z_{R} far away from the pocket, respectively. For the binding simulation (i.e., the simulation of a binding process), we reset the value of $z(t)$ to be $2z_{\mathrm{R}} - z(t)$ if $z(t) \geq z_{\mathrm{R}}$, and we stop the simulation if $z(t) \leq z_{\mathrm{L}}$. For the unbinding simulation (i.e., the simulation of an unbinding process), we reset the value of $z(t)$ to be z_{L} if $z(t) \leq z_{\mathrm{L}}$, and we stop the simulation if $z(t) \geq z_{\mathrm{R}}$.

Algorithm for BD Simulations without the Dry-Wet Fluctuations.

Step 1. Input the diffusion constants D_{in} and D_{out}, the controlling parameter ν, the threshold position z_{c}, the total PMF $V(z)$, an initial ligand position z_{init}, and the simulation time step δt. Set Time = 0, $z^{(0)} = z_{\mathrm{init}}$, and $k = 0$.

Step 2. Given a ligand position $z^{(k)}$. Calculate $z^{(k+1)}$ by

$$z^{(k+1)} - z^{(k)} = -\left[\frac{1}{k_{\mathrm{B}}T} D(z^{(k)}) V'(z^{(k)}) + D'(z^{(k)}) \right] \delta t + \sqrt{2D(z^{(k)})\delta t}\, \xi,$$

where ξ is a random number with the standard normal distribution.

Step 3. Set Time := Time $+\delta t$.

(a) For binding simulations: If $z^{(k+1)} \geq z_R$, set $z^{(k+1)} := 2z_R - z^{(k+1)}$; If $z^{(k+1)} \leq z_L$, then stop.

(b) For unbinding simulations: If $z^{(k+1)} \leq z_L$, set $z^{(k+1)} := z_L$; If $z^{(k+1)} \geq z_R$, then stop.

Step 4. Set $k := k + 1$ and go to Step 2.

To study the effect of dry-wet fluctuations on the kinetics of ligand-pocket binding/unbinding, let us define a position-dependent, three-state, random variable $\eta = \eta(z) \in \{0, 1, 2\}$ by $\eta(z) = 0$, 1, or 2, if the hydration state of the system at a given reaction coordinate z is 1s-dry, 2s-dry, or 2s-wet, respectively. The (discrete) probability density of $\eta(z)$ is defined by the equilibrium probabilities $P_i^{\mathrm{eq}}(z)$ $(i = 0, 1, 2)$:

$$\mathrm{Prob}\left(\{\eta(z) = i\}\right) = P_i^{\mathrm{eq}}(z) = \frac{e^{-G[\Gamma_i(z)]/k_{\mathrm{B}}T}}{\sum_{j=0}^{2} e^{-G[\Gamma_j(z)]/k_{\mathrm{B}}T}}, \qquad i = 0, 1, 2,$$

where $G[\Gamma_i(z)]$ is the VISM solvation free energy at the ith hydration state represented by the VISM surface $\Gamma_i(z)$, and the sum runs over all the hydration states, at the given reaction coordinate z. To account for the fluctuations among the three states at each reaction coordinate, we further define a potential $V_{\mathrm{fluc}} = V_{\mathrm{fluc}}(\eta, z)$ by $V_{\mathrm{fluc}}(\eta, z) = V_i(z)$ if $\eta = i$ for $i \in \{0, 1, 2\}$, where the potential functional $V_i(z)$, defined in Eq. [3] in the main text, is the sum of the solvation free energy of the ith hydration state and the ligand-pocket vdW interaction energy at the reaction coordinate z. If at a given coordinate z, there is only one or two hydration states, then we set $V_i(z) = 0$ for the other states i.

We perform our continuous-time Markov chain (CTMC) BD simulations, i.e., numerically solve the following stochastic differential equation for the ligand position $z = z(t) = z_t$ (same as that in the CTMC BD simulations part of section Theory and Methods in the main text):

$$\begin{cases} dz_t = \left[-\dfrac{1}{k_{\rm B}T}D(z_t)\dfrac{\partial V_{\rm fluc}(\eta(z_t), z_t)}{\partial z} + D'(z_t)\right] dt + \sqrt{2D(z_t)}\, d\xi_t, \\ \eta(z_t) \in \{0,1,2\} \text{ is a CTMC with the transition rate matrix} \\ \begin{pmatrix} -[R_{01}(z_t)+R_{02}(z_t)] & R_{01}(z_t) & R_{02}(z_t) \\ R_{10}(z_t) & -[R_{10}(z_t)+R_{12}(z_t)] & R_{12}(z_t) \\ R_{20}(z_t) & R_{21}(z_t) & -[R_{20}(z_t)+R_{21}(z_t)] \end{pmatrix}, \end{cases} \tag{9}$$

together with a given initial position $z_0 = z_{\rm init}$. Here, the partial derivative of $V_{\rm fluc}$ is with respect to its second variable, ξ_t is the standard Brownian motion, and the rates of transitions $R_{ij}(z)$ from the ith state to the jth state for all $i, j = 0, 1, 2$ are defined in Theory and Methods in the main text. Solutions to Eq. [9] are constrained by $z(t) \in [z_{\rm L}, z_{\rm R}]$ for all t for some boundaries $z_{\rm L}$ and $z_{\rm R}$. Again, for a binding simulation, we reset the value of $z(t)$ to be $2z_{\rm R} - z(t)$ if $z(t) \ge z_{\rm R}$, and we stop the simulation if $z(t) \le z_{\rm L}$. For an unbinding simulation, we reset the value of $z(t)$ to be $z_{\rm L}$ if $z(t) \le z_{\rm L}$, and we stop the simulation if $z(t) \ge z_{\rm R}$.

Algorithm for CTMC BD Simulations.

Step 1. Input the diffusion constants $D_{\rm in}$ and $D_{\rm out}$, the controlling parameter ν, the threshold position $z_{\rm c}$, the potential functions $V_0(z)$, $V_1(z)$, and $V_2(z)$, an initial position $z_{\rm init}$, and the simulation time step δt. Initialize the hydration state $\eta(z_{\rm init})$ according to the probabilities $P_i^{\rm eq}(z_{\rm init})$ $(i = 0, 1, 2)$. Set Time $= 0$, $z^{(0)} = z_{\rm init}$, and $k = 0$.

Step 2. Given a ligand position $z^{(k)}$. Calculate $z^{(k+1)}$ by

$$z^{(k+1)} - z^{(k)} = -\left[\frac{1}{k_{\rm B}T}D(z^{(k)})V_i'(z^{(k)}) + D'(z^{(k)})\right]\delta t + \sqrt{2D(z^{(k)})\delta t}\,\xi \qquad \text{if } \eta(z^{(k)}) = i,$$

where ξ is a random number with the standard normal distribution.

Step 3. Update the hydration state η. If $z^{(k+1)} \le z_{\rm c}$, set $\eta = 0$; else, determine η as follows:

For $\eta = i$, if $e^{-\delta t \sum_{j\neq i} R_{ij}(z^{(k+1)})} \ge \zeta$, keep $\eta = i$,; otherwise, determine the transition from state i to state j according to the probability $R_{ij}/\sum_{k\neq i} R_{ik}$ $(i \neq j)$, where ζ is a random number uniformly distributed between 0 and 1.

Step 4. Set Time := Time $+\delta t$.

(a) For binding simulations: If $z^{(k+1)} \ge z_R$, set $z^{(k+1)} := 2z_R - z^{(k+1)}$; If $z^{(k+1)} \le z_L$, then stop.

(b) For unbinding simulations: If $z^{(k+1)} \le z_L$, set $z^{(k+1)} := z_L$; If $z^{(k+1)} \ge z_R$, then stop.

Step 5. Set $k := k + 1$ and go to Step 2.

4. Generalized Fokker–Planck Equations and the Mean First-Passage Time

Let us denote by $\bar P(z,t)$ the probability density of the ligand random position $z = z(t) \in [z_{\rm L}, z_{\rm R}]$ in the absence of pocket dry-wet fluctuations. It is determined by the following Fokker–Planck equation (FPE) that is associated with the stochastic differential equation [7]:

$$\frac{\partial \bar P}{\partial t} = \frac{\partial}{\partial z}\left\{D(z)\left[\frac{\partial \bar P}{\partial z} + \frac{1}{k_{\rm B}T}V'(z)\bar P\right]\right\}, \tag{10}$$

where $V = V(z)$ is the equilibrium PMF defined in Eq. [1] in the main text. The initial condition for this equation is $\bar P(z, 0) = \bar P^{(0)}(z)$ for some $\bar P^{(0)}(z)$ and the boundary conditions are designed separately for the simulation of binding and that of unbinding:

$$\begin{aligned} &\bar P(z_{\rm L}, t) = 0 \quad \text{and} \quad \frac{\partial \bar P(z_{\rm R}, t)}{\partial z} = 0 && \text{for binding,} \\ &\frac{\partial \bar P(z_{\rm L}, t)}{\partial z} + \frac{1}{k_{\rm B}T}V'(z_{\rm L})\bar P(z_{\rm L}, t) = 0 \quad \text{and} \quad \bar P(z_{\rm R}, t) = 0 && \text{for unbinding.} \end{aligned} \tag{11}$$

The mean first-passage time (MFPT) of binding/unbinding is given by

$$\tau_{\rm MFPT}(z_{\rm init}) = \int_0^\infty \int_{z_{\rm L}}^{z_{\rm R}} \bar P(z,t)\, dz\, dt,$$

where z_{init} is the initial ligand position, or equivalently, the initial value of $\bar{P}$ is given by $\bar{P}(z,0)=\delta(z-z_{\text{init}})$, the Dirac mass concentrated at z_{init}. Integrating both sides of Eq. [10] with respect to time, we arrive at

$$-\bar{P}^{\text{init}}(z,z_{\text{init}})=\frac{d}{dz}\left\{D(z)\left[\frac{d\bar{P}^{I}(z)}{dz}+\frac{1}{k_{\text{B}}T}\bar{P}^{I}(z)V'(z)\right]\right\}, \qquad [12]$$

where $\bar{P}^{\text{init}}(z,z_{\text{init}})=\delta(z-z_{\text{init}})$ is the initial probability density, and

$$\bar{P}^{I}(z)=\int_0^\infty \bar{P}(z,t)dt.$$

The solution to Eq. [12] can be obtained by integrating the equation twice with the boundary conditions Eq. [11]. For instance, the unbinding MFPT of a ligand starting at z_{init} without solvent fluctuations is given by

$$\begin{aligned}\tau_{\text{MFPT}}(z_{\text{init}})&=\int_{z_{\text{L}}}^{z_{\text{R}}}\bar{P}^{I}(z)dz\\&=\int_{z_{\text{init}}}^{z_{\text{R}}}\frac{e^{\beta V(z)}}{D(z)}dz\int_{z_{\text{L}}}^{z_{\text{init}}}e^{-\beta V(z)}dz+\int_{z_{\text{init}}}^{z_{\text{R}}}e^{-\beta V(z)}\left[\int_{z}^{z_{\text{R}}}\frac{e^{\beta V(z')}}{D(z')}dz'\right]dz,\end{aligned}$$

where $\beta=1/(k_{\text{B}}T)$. To get an explicit analytical solution for the MFPT of the binding, we make an assumption that $V'(z_R)=0$, which is often true when z_R is far from the pocket. Under such an assumption, the binding MFPT of a ligand starting at z_{init} without solvent fluctuations is obtained analogously:

$$\begin{aligned}\tau_{\text{MFPT}}(z_{\text{init}})&=\int_{z_{\text{L}}}^{z_{\text{R}}}\bar{P}^{I}(z)dz\\&=\int_{z_{\text{L}}}^{z_{\text{init}}}\frac{e^{\beta V(z)}}{D(z)}dz\int_{z_{\text{init}}}^{z_{\text{R}}}e^{-\beta V(z)}dz+\int_{z_{\text{L}}}^{z_{\text{init}}}e^{-\beta V(z)}\left[\int_{z_{\text{L}}}^{z}\frac{e^{\beta V(z')}}{D(z')}dz'\right]dz.\end{aligned}$$

We now consider the MFPT with dry-wet fluctuations (or the solvent fluctuations). We solve the following system of generalized FPEs for the probability densities, $P_0(z,t)$, $P_1(z,t)$, and $P_2(z,t)$, for the probabilities of finding the ligand at location z at time t with the system being in the states of 1s-dry, 2s-dry, and 2s-wet, respectively (24):

$$\frac{\partial P_i}{\partial t}=\frac{\partial}{\partial z}\left\{D(z)\left[\frac{\partial P_i}{\partial z}+\frac{1}{k_{\text{B}}T}V_i'(z)P_i\right]\right\}+\sum_{0\le j\le 2,\, j\ne i}R_{ji}(z)P_j-\left(\sum_{0\le j\le 2,\, j\ne i}R_{ij}(z)\right)P_i \qquad \text{for } i=0,1,2. \qquad [13]$$

This is the same equation as in section Theory and Methods in the main text.

These equations correspond to the stochastic differential equation [9] for our CTMC BD simulations. They are solved with some initial values and also for $z_{\text{L}}<z<z_{\text{R}}$, with the boundary conditions

$$\begin{aligned}&P_i(z_{\text{L}},t)=0 \quad \text{and} \quad \frac{\partial P_i(z_{\text{R}},t)}{\partial z}=0 && \text{for binding,}\\&\frac{\partial P_i(z_{\text{L}},t)}{\partial z}+\frac{1}{k_{\text{B}}T}V'(z_{\text{L}})P_i(z_{\text{L}},t)=0 \quad \text{and} \quad P_i(z_{\text{R}},t)=0 && \text{for unbinding,}\end{aligned}$$

where $i=0,1,2$.

To calculate the MFPT for the ligand-pocket binding/unbinding starting from z_{init}, we let

$$P_i^{\text{init}}(z,z_{\text{init}})=P_i^{\text{eq}}(z_{\text{init}})\delta(z-z_{\text{init}})$$

be the initial probability densities for P_i with $i=0,1,2$. Integrating both sides of the Eq. [13] with respect to time, we have

$$-P_i^{\text{init}}(z,z_{\text{init}})=\frac{d}{dz}\left\{D(z)\left[\frac{dP_i^I(z)}{dz}+\frac{1}{k_{\text{B}}T}P_i^I(z)V_i'(z)\right]\right\}+\sum_{j\ne i}R_{ji}(z)P_j^I-\left(\sum_{j\ne i}R_{ij}(z)\right)P_i^I,$$

where

$$P_i^I(z)=\int_0^\infty P_i(z,t)\,dt. \quad \text{for } i=0,1,2.$$

With certain boundary conditions, the boundary-value problem can be solved with the finite difference method. The MFPT is then given by

$$\tau_{\mathrm{MFPT}}(z_{\mathrm{init}}) = \sum_{i=0}^{2} \int_{z_{\mathrm{L}}}^{z_{\mathrm{R}}} P_i^I(z)dz.$$

This can be calculated with numerical integration.

5. Parameters

We list the values and units of all the parameters in our computations. These are the same as those described in the main text.

Symbol	Description	Units	Value
T	Temperature	Kelvin	298
ΔP	Pressure difference (cf. Eq. [1]) ^{a}	bar	0
γ_0	surface tension for a planar interface (cf. Eq. [1])	$k_{\mathrm{B}}T$/Å^2	0.143
τ	Tolman length (cf. Eq. [1])	Å	0.8
ρ_0	bulk solvent (i.e., water) density (cf. Eq. [1])	Å^{-3}	0.033
σ_{water}	LJ length parameter for a solvent molecule ^{b}	Å	3.154
σ_{wall}	LJ length parameter for a wall particle ^{b}	Å	4.152
σ_{ligand}	LJ length parameter for the ligand ^{b}	Å	3.73
$\varepsilon_{\mathrm{water}}$	LJ energy parameter for a solvent molecule ^{b}	$k_{\mathrm{B}}T$	0.26
$\varepsilon_{\mathrm{wall}}$	LJ energy parameter for a wall particle ^{b}	$k_{\mathrm{B}}T$	9.67E-4
$\varepsilon_{\mathrm{ligand}}$	LJ energy parameter for the ligand ^{b}	$k_{\mathrm{B}}T$	0.5
M	Number of (interior) images of a string	No units	10
z_{c}	The coordinate of the pocket entrance	Å	-0.5
z_{L}	Smallest value of the reaction coordinate ^{c}	Å	-4
z_{R}	Largest value of the reaction coordinate ^{c}	Å	15.5
R_0	Prefactor of transition rates ^{d}	ps^{-1}	0.13
D_{in}	Diffusion constant inside the pocket (cf. Eq. [8]) ^{e}	Å^2/ps	1
D_{out}	Diffusion constant outside the pocket (cf. Eq. [8]) ^{f}	Å^2/ps	0.26
ν	The control parameter in D (cf. Eq. [8])	1/Å	5

Table S1. Parameters.

^{a} The term $\Delta P\,\mathrm{vol}\,(\Omega_{\mathrm{m}})$ is very small compared with the other terms in Eq. [1].
^{b} The values are taken from (25, 26). We use the Lorentz–Berthelot mixing rules to determine the LJ parameters for the interaction of two particles.
^{c} These values can vary.
^{d} R_0 is estimated from the relaxation timescale $(R_{\mathrm{dw}} + R_{\mathrm{wd}})^{-1} \approx 10$ ps of water fluctuations in the pocket when the ligand is far away (27), where $R_{\mathrm{dw}} = R_0 e^{-B_{\mathrm{dw}}/k_{\mathrm{B}}T}$ and $R_{\mathrm{wz}} = R_0 e^{-B_{\mathrm{wd}}/k_{\mathrm{B}}T}$ with B_{dw} and B_{wd} the barriers in the pocket dry-wet and wet-dry transitions when the ligand is far away; cf. section Theory and Methods in the main text.
^{e} This is a trial value. See subsection B in section 6. Additional Simulation Results.
^{f} The value is taken from (27).

6. Additional Simulation Results

A. Minimum Energy Paths for $z = 2$ Å and $z = 10$ Å. At the reaction coordinate $z = 2$ Å, there are two hydration states: 2s-wet and 1s-dry, and only one MEP is found to connect these two states. Fig. S1 shows this MEP, together with the solute-solvent interfacial structures of the two hydration states (marked (I) and (III), respectively) and the only transition state (marked (II)). Note that the 1s-dry has a lower solvation free energy.

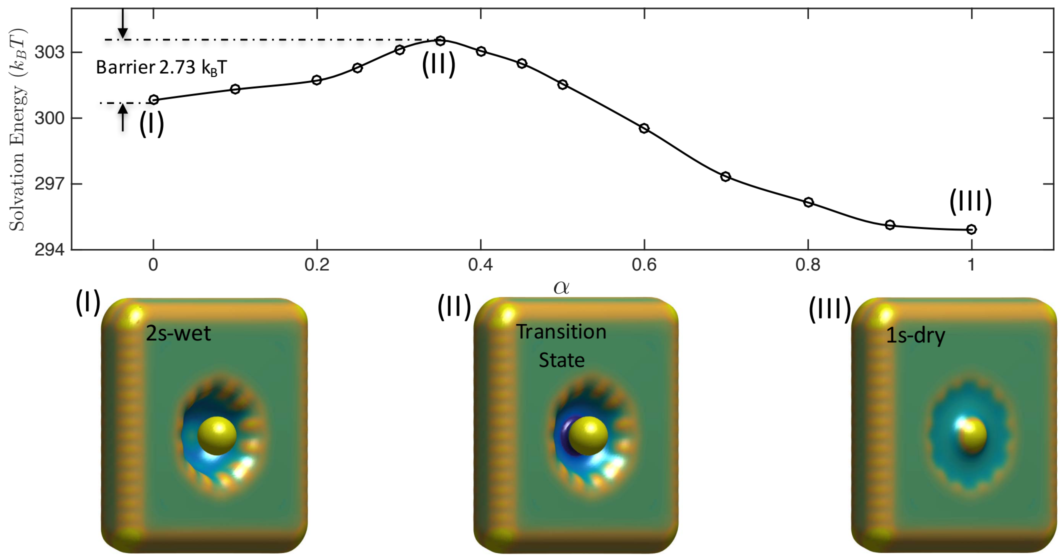

Fig. S1. The MEP connecting the only hydration states of 2s-wet (marked (I)) and 1s-dry (marked (III)) when the ligand is placed at $z = 2$ Å. The solute-solvent interfaces of these hydration states, and the transition state (marked (II)) are also shown. The energy barrier in the dewetting transition from 2s-wet to 1s-dry is $2.73\ k_\mathrm{B}T$.

Fig. S2 shows the MEP connecting the only hydration states 2s-wet and 2s-dry for the reaction coordinate $z = 10$ Å. The calculated activation energy barrier is about 0.68 $k_\mathrm{B}T$. In contrast to the dewetting energy barrier (0.70 $k_\mathrm{B}T$) for $z = 6$ Å (cf. Fig. 3 in the main text), one finds that the presence of the ligand with a smaller ligand-pocket distance increases the dewetting energy barrier of the hydrophobic pocket. This is because that, when the ligand is close, part of the solvent region with the attractive solute-solvent vdW interaction is lost in such a dewetting transition. From an explicit-solvent point of view, the water molecules in the hydration shell of the methane particle hinders the evaporation of water molecules from the pocket.

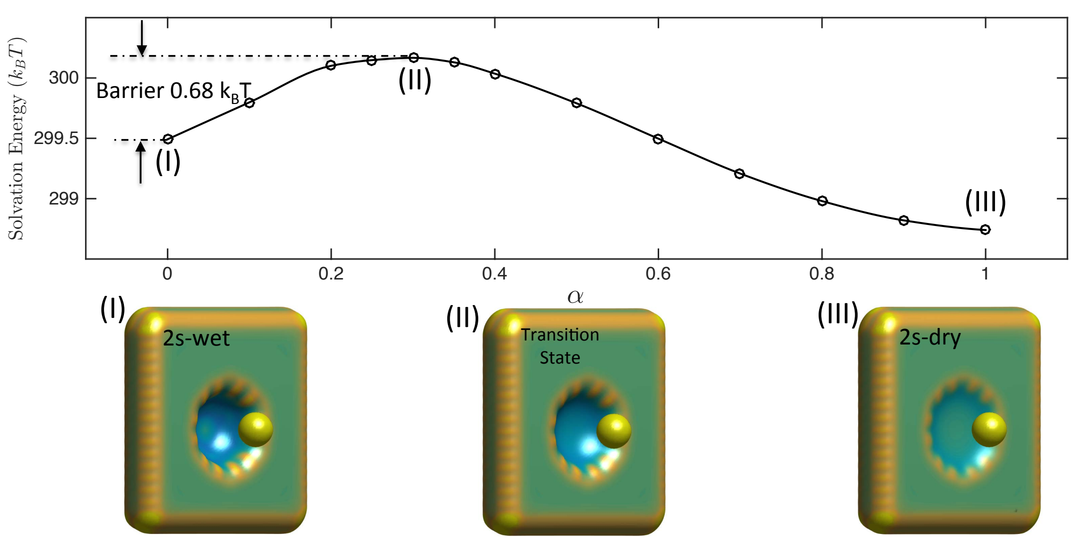

Fig. S2. The MEP connecting the hydration states 2s-wet and 2s-dry with the ligand is placed at $z = 10$ Å. The energy barrier in the dewetting transition from 2s-wet to 2s-dry is $0.68\ k_\mathrm{B}T$. The solute-solvent interfaces of the hydration states 2s-wet (marked (I)) and 2s-dry (marked (III)), and that of the transition state (marked (II)) are also shown.

B. Effect of D_in. We choose two very different values of the diffusion constant $D_\mathrm{in} = 1\ \mathrm{Å}^2/\mathrm{ps}$ and $D_\mathrm{in} = 1{,}000\ \mathrm{Å}^2/\mathrm{ps}$, and hence determine two, effective and position-dependent diffusion coefficient $D(z)$ by Eq. [8]. With these diffusion coefficients, we solve numerically Eq. [10], and Eqs [13], and then calculate the MFPT for the binding and unbinding process. Fig. S3 shows that the large difference in the diffusion constant D_in does not affect the MFPT with or without the dry-wet fluctuations.

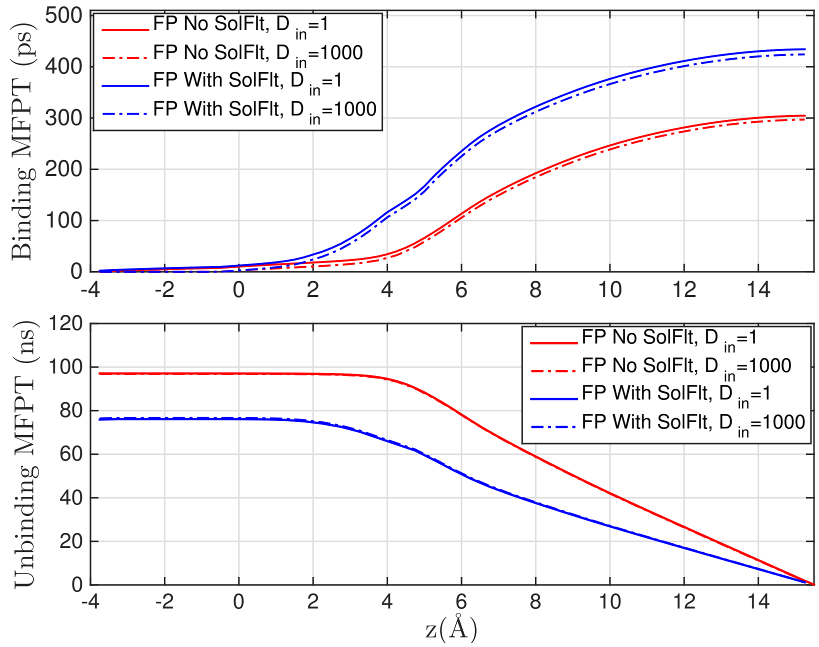

Fig. S3. The FPE calculations of the binding and unbinding MFPT of the ligand starting at z with two different values of the diffusion coefficient D_{in} for the ligand inside the pocket. SolFlt stands for the solvent fluctuations, i.e., the pocket dry-wet fluctuations.

C. Evolution of Probability Density of Ligand Position. To further understand the effect of solvent fluctuations, we investigate the decay rate of the probability densities $\bar{P}(z,t)$ and $P_{\mathrm{tot}}(z,t) = \sum_{i=0}^{2} P_i(z,t)$ in binding and unbinding processes Here, $\bar{P}(z,t)$ is the probability density for the ligand random position $z(t)$ in the absence of dry-wet fluctuations (cf. Eq. [10]), and each $P_i(z,t)$ $(i = 0, 1, \text{or } 2)$ is the probability density for the ligand random position $z(t)$ with the system being at the ith hydration state (cf. Eq. [13]). Fig. S4 displays the evolution of the probability densities normalized by the initial value at the positions $z = 6$ and $z = -2$ in binding and unbinding simulations, respectively. In the binding processes, the normalized probability density decays slower when solvent fluctuations are included, because the pocket fluctuates between dry and wet states and the PMF of the wet branch is repulsive. On the contrary, the normalized probability density decays faster in unbinding processes, and hence a shorter residence time when solvent fluctuations are included. This is again due to the repulsive PMF of the wet branch. The pocket fluctuates to the wet state when the unbinding ligand approaches the entrance of the pocket.

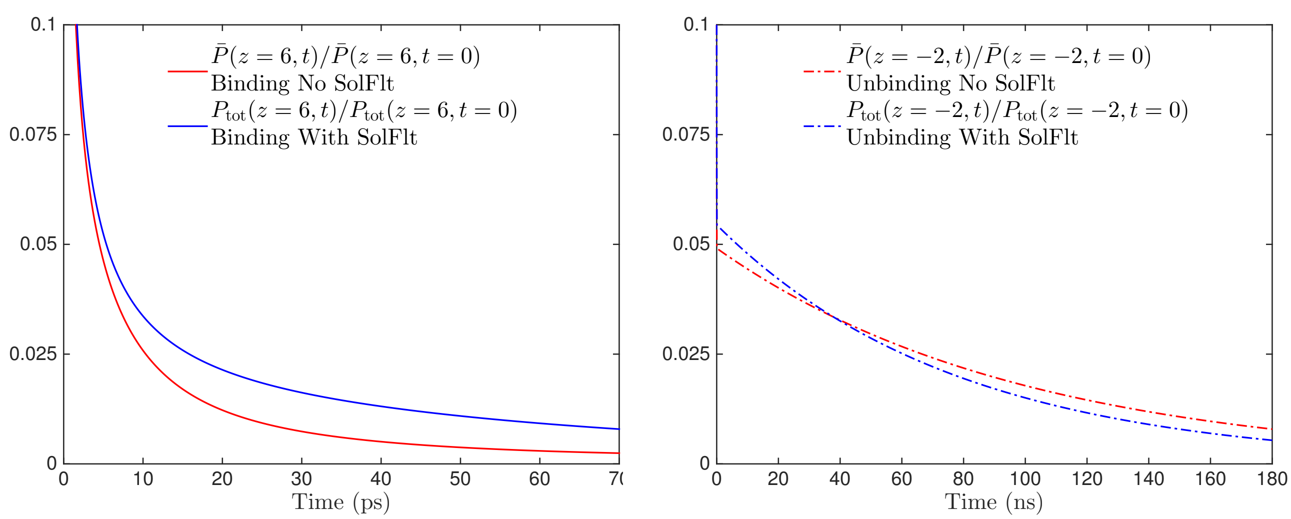

Fig. S4. Evolution of probability densities, $\bar{P}(z,t)$ (cf. Eq. [10]) and $P_{\mathrm{tot}}(z,t) = P_0(z,t) + P_1(z,t) + P_2(z,t)$ (cf. Eq. [13]) normalized by the initial values at $z = 6$ (left) and $z = -2$ (right) in the binding and unbinding simulations with and without solvent fluctuations. SolFlt stands for the solvent fluctuations, i.e., the pocket dry-wet fluctuations.

D. Sensitivity of R_0. We now discuss the effect of R_0 on the binding and unbinding kinetics. Fig. S5 presents the MFPT of the binding and unbinding of ligand against the starting position $z_{\rm init} = z$ with different values of R_0. We see that the results predicted by the CTMC BD simulations and FPE calculations agree with each other perfectly. As R_0 decreases, both binding and unbinding MFPTs increase. With a smaller R_0, the dewetting transition rate decreases and the ligand stays in the branch of 2s-wet for longer time in binding processes. This explains the longer binding MFPT with a smaller R_0. For unbinding, a smaller R_0 leads to a smaller wetting transition rate, restraining the transition starting from the 1s-dry state whose PMF is attractive. This explains the increasing unbinding MFPT with a decreasing value of R_0.

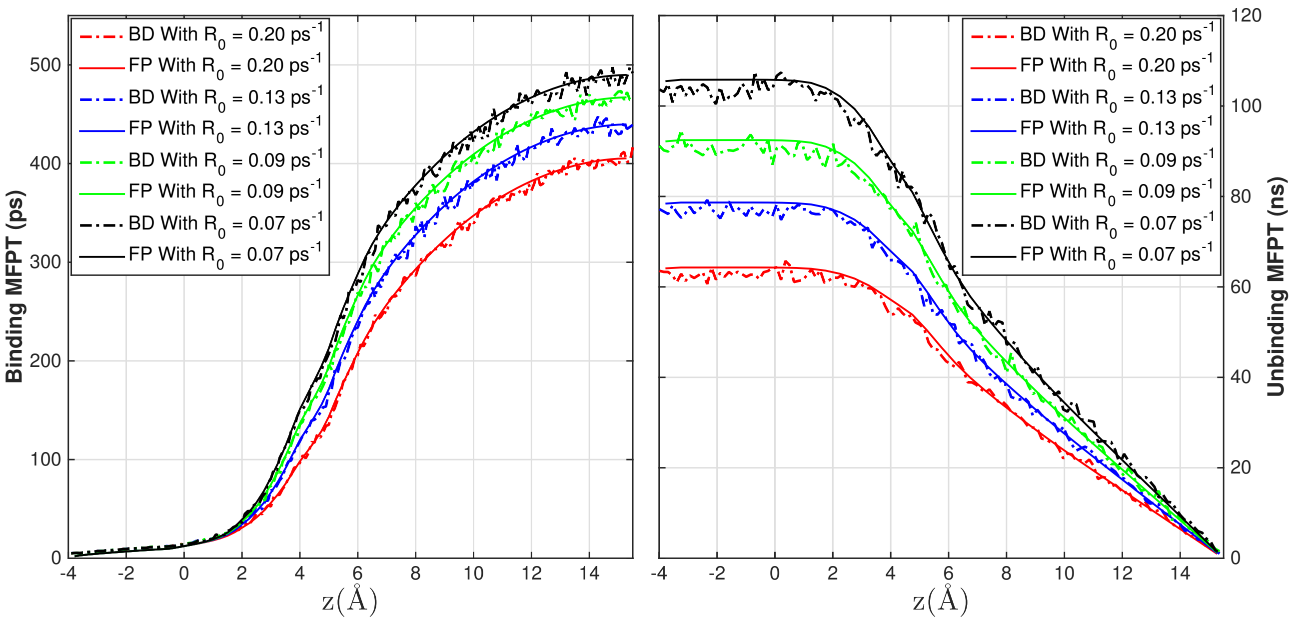

Fig. S5. The MFPT for the binding (left) and the unbinding of ligand that starts from $z_{\rm init} = z$, predicted by the CTMC BD simulations and FPE calculations with different values of R_0.

References

1. Dzubiella J, Swanson JMJ, McCammon JA (2006) Coupling hydrophobicity, dispersion, and electrostatics in continuum solvent models. *Phys. Rev. Lett.* 96:087802.
2. Dzubiella J, Swanson JMJ, McCammon JA (2006) Coupling nonpolar and polar solvation free energies in implicit solvent models. *J. Chem. Phys.* 124:084905.
3. Cheng LT, Dzubiella J, McCammon JA, Li B (2007) Application of the level-set method to the implicit solvation of nonpolar molecules. *J. Chem. Phys.* 127:084503.
4. Cheng LT, et al. (2009) Coupling the level-set method with molecular mechanics for variational implicit solvation of nonpolar molecules. *J. Chem. Theory Comput.* 5:257–266.
5. Cheng LT, Li B, Wang Z (2010) Level-set minimization of potential controlled Hadwiger valuations for molecular solvation. *J. Comput. Phys.* 229:8497–8510.
6. Wang Z, et al. (2012) Level-set variational implicit solvation with the Coulomb-field approximation. *J. Chem. Theory Comput.* 8:386–397.
7. Zhou S, Cheng LT, Dzubiella J, Li B, McCammon JA (2014) Variational implicit solvation with Poisson–Boltzmann theory. *J. Chem. Theory Comput.* 10(4):1454–1467.
8. Zhou S, et al. (2015) LS-VISM: A software package for analysis of biomolecular solvation. *J. Comput. Chem.* 36:1047–1059.
9. Osher S, Sethian JA (1988) Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations. *J. Comput. Phys.* 79:12–49.
10. Sethian JA (1999) *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision, and Materials Science.* (Cambridge University Press), 2nd edition.
11. Osher S, Fedkiw R (2002) *Level Set Methods and Dynamic Implicit Surfaces.* (Springer, New York).
12. Bashford D, Case DA (2000) Generalized Born models of macromolecular solvation effects. *Ann. Rev. Phys. Chem.* 51:129–152.

13. Cheng HB, Cheng LT, Li B (2011) Yukawa-field approximation of electrostatic free energy and dielectric boundary force. *Nonlinearity* 24:3215–3236.
14. Davis ME, McCammon JA (1990) Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* 90:509–521.
15. Sharp KA, Honig B (1990) Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Chem.* 19:301–332.
16. Che J, Dzubiella J, Li B, McCammon JA (2008) Electrostatic free energy and its variations in implicit solvent models. *J. Phys. Chem. B* 112:3058–3069.
17. Cai Q, Ye X, Wang J, Luo R (2011) Dielectric boundary forces in numerical Poisson–Boltzmann methods: Theory and numerical strategies. *Chem. Phys. Lett.* 514:368–373.
18. Li. B, Cheng XL, Zhang ZF (2011) Dielectric boundary force in molecular solvation with the Poisson–Boltzmann free energy: A shape derivative approach. *SIAM J. Applied Math.* 71:2093–2111.
19. Cai Q, Ye X, Luo R (2012) Dielectric pressure in continuum electrostatic solvation of biomolecules. *Phys. Chem. Chem. Phys.* 14:15917–15925.
20. E W, Ren W, Vanden-Eijnden E (2002) String method for the study of rare events. *Phys. Rev. B* 66:052301.
21. E W, Ren W, Vanden-Eijnden E (2007) Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.* 126:164103.
22. E. W, Vanden-Eijnden E (2010) Transition-path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem.* 61:391–420.
23. Ren W, Vanden-Eijnden E (2013) A climbing string method for saddle point search. *J. Chem. Phys.* 138:134105.
24. Mondal J, Morrone J, Berne BJ (2013) How hydrophobic drying forces impact the kinetics of molecular recognition. *Proc. Natl. Acad. Sci., USA* 110(33):13277–13282.
25. Setny P, et al. (2009) Dewetting-controlled binding of ligands to hydrophobic pockets. *Phys. Rev. Lett.* 103:187801.
26. Cheng LT, et al. (2009) Interfaces and hydrophobic interactions in receptor-ligand systems: A level-set variational implicit solvent approach. *J. Chem. Phys.* 131:144102.
27. Setny P, Baron R, Kekenes-Huskey P, McCammon JA, Dzubiella J (2013) Solvent fluctuations in hydrophobic cavity-ligand binding kinetics. *Proc. Natl. Acad. Sci., USA* 110:1197–1202.