

Title	Elucidating protein-ligand binding kinetics based on returning probability theory
Author(s)	Kasahara, Kento; Masayama, Ren; Okita, Kazuya et al.
Citation	Journal of Chemical Physics. 2023, 159(13), p. 134103
Version Type	VoR
URL	<a href="https://hdl.handle.net/11094/93213">https://hdl.handle.net/11094/93213</a>
rights	This article may be downloaded for personal use only. Any other use requires prior permission of the author and AIP Publishing. This article appeared in The Journal of Chemical Physics 159, 134103 (2023) and may be found at <a href="https://doi.org/10.1063/5.0165692">https://doi.org/10.1063/5.0165692</a> .
Note	

***Osaka University Knowledge Archive : OUKA***

<https://ir.library.osaka-u.ac.jp/>

Osaka University

RESEARCH ARTICLE | OCTOBER 03 2023

## Elucidating protein–ligand binding kinetics based on returning probability theory

Kento Kasahara  ; Ren Masayama ; Kazuya Okita ; Nobuyuki Matubayasi 



*J. Chem. Phys.* 159, 134103 (2023)

<https://doi.org/10.1063/5.0165692>



CrossMark

### Articles You May Be Interested In

Completing the dark matter solutions in degenerate Kaluza-Klein theory

*J. Math. Phys.* (April 2019)

Gibbs measures based on 1d (an)harmonic oscillators as mean-field limits

*J. Math. Phys.* (April 2018)

An upper diameter bound for compact Ricci solitons with application to the Hitchin–Thorpe inequality. II

*J. Math. Phys.* (April 2018)

500 kHz or 8.5 GHz?  
And all the ranges in between.

Lock-in Amplifiers for your periodic signal measurements



Find out more



# Elucidating protein–ligand binding kinetics based on returning probability theory

Cite as: *J. Chem. Phys.* **159**, 134103 (2023); doi: [10.1063/5.0165692](https://doi.org/10.1063/5.0165692)

Submitted: 29 June 2023 • Accepted: 14 September 2023 •

Published Online: 3 October 2023



View Online



Export Citation



CrossMark

Kento Kasahara,<sup>a)</sup>  Ren Masayama,<sup>b)</sup>  Kazuya Okita,<sup>b)</sup>  and Nobuyuki Matubayasi<sup>b)</sup> 

## AFFILIATIONS

Division of Chemical Engineering, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan

<sup>a)</sup> Author to whom correspondence should be addressed: [kasahara@cheng.es.osaka-u.ac.jp](mailto:kasahara@cheng.es.osaka-u.ac.jp)

<sup>b)</sup> Electronic mail: [nobuyuki@cheng.es.osaka-u.ac.jp](mailto:nobuyuki@cheng.es.osaka-u.ac.jp)

## ABSTRACT

The returning probability (RP) theory, a rigorous diffusion-influenced reaction theory, enables us to analyze the binding process systematically in terms of thermodynamics and kinetics using molecular dynamics (MD) simulations. Recently, the theory was extended to atomistically describe binding processes by adopting the host–guest interaction energy as the reaction coordinate. The binding rate constants can be estimated by computing the thermodynamic and kinetic properties of the reactive state existing in the binding processes. Here, we propose a methodology based on the RP theory in conjunction with the energy representation theory of solution, applicable to complex binding phenomena, such as protein–ligand binding. The derived scheme of calculating the equilibrium constant between the reactive and dissociate states, required in the RP theory, can be used for arbitrary types of reactive states. We apply the present method to the bindings of small fragment molecules [4-hydroxy-2-butanone (BUT) and methyl methylthiomethyl sulphoxide (DSS)] to FK506 binding protein (FKBP) in an aqueous solution. Estimated binding rate constants are consistent with those obtained from long-timescale MD simulations. Furthermore, by decomposing the rate constants to the thermodynamic and kinetic contributions, we clarify that the higher thermodynamic stability of the reactive state for DSS causes the faster binding kinetics compared with BUT.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0165692>

## I. INTRODUCTION

Molecular binding processes are ubiquitous in various fields of science. In biological systems, for instance, the ligand (substrate or drug) binding to its target protein induces or inhibits the expression of a biological function such as cell proliferation. Significant efforts have been expended to investigate thermodynamic properties, such as binding free energy, for screening and lead optimization of the drug candidates in the field of drug discovery.<sup>1,2</sup> In addition to the thermodynamics, the kinetic properties, such as the binding/unbinding, have been also utilized as an indicator of drug efficacy.<sup>3–6</sup> Thus, analyzing the detailed binding mechanism in terms of thermodynamics and kinetics is essential for rational drug design. Since molecular dynamics (MD) simulation provides atomistic information on the system of interest, it can be useful for realizing such analysis.<sup>7–9</sup>

A binding process is thermodynamically quantified by the standard free energy of binding. The free energy perturbation (FEP)<sup>10</sup> and thermodynamic integration (TI)<sup>11</sup> methods are widely used for computing the free energy difference between two different states

of interest with MD simulations in an exact way. In these methods, the free energy difference can be evaluated through the MD simulations for a set of intermediate states connecting the two states of interest.<sup>12</sup> The thermodynamic cycles practically suitable for the binding free energy based on these methods are proposed.<sup>13–16</sup> The endpoint classical density functional theory (DFT), a statistical mechanics theory of solution, is also useful for the free energy calculation.<sup>17</sup> Unlike the FEP and TI methods, since only the information on the two states of interest is needed in the framework of endpoint DFT, the computational cost is reduced. The energy representation (ER) theory of solution<sup>18–21</sup> is one of the endpoint DFTs that realizes the accurate estimation of the solvation free energy of a solute from the MD simulations by employing the solute–solvent pair interaction energy as a coordinate, namely, the energy coordinate. The dimensionality reduction with the energy coordinate enables us to effectively treat the position and orientation of a solvent molecule with intramolecular degrees of freedom. The ER theory can be used also for evaluating the equilibrium binding constants associated with the binding of a solute into lipid membranes.<sup>22</sup>

MD-based methodologies for elucidating the binding kinetics have been extensively developed.<sup>23–25</sup> For example, the kinetics theories of treating the state-to-state transition probabilities, such as Markov state model (MSM)<sup>26–28</sup> and milestoning theory,<sup>29–31</sup> can be used to estimate the binding/unbinding rate constants. Enhanced sampling methods, such as weighted ensemble (WE),<sup>32,33</sup> parallel-cascade MD (PaCS-MD),<sup>34</sup> infrequent metadynamics,<sup>35</sup> scaled MD,<sup>36,37</sup> Gaussian accelerated MD (GaMD),<sup>38,39</sup> simulation-enabled estimation of kinetic rates (SEEKR),<sup>40,41</sup> and resampling of ensembles by variation optimization (REVO),<sup>42</sup> are available for the efficient calculations of the rate constants.

Diffusion-influenced reaction (DIR) theories<sup>43–47</sup> provide a useful framework for describing the binding kinetics. In the DIR theories, the theoretical expressions of the rate constants are derived based on the transport equations treating the reaction (binding) processes. Recently, we proposed a methodology of quantifying the binding rate constants based on returning probability (RP) theory and MD simulations.<sup>48,49</sup> The RP theory is a rigorous DIR theory based on the Liouville equation of the phase space densities with the reaction sink term.<sup>50</sup> The reaction sink term is introduced for describing the reaction (binding) probability on the reactive state defined on the reaction coordinates. The RP theory provides a tractable expression of the rate constant characterized in terms of the thermodynamic and kinetic properties of the reactive state, and hence, the systematic analysis is possible for relative importance of the thermodynamic and kinetic contributions. Lee *et al.* applied the RP theory to the approaching process of the super oxide anion radical ( $O_2^-$ ) to Cu/Zn superoxide dismutase (SOD).<sup>51</sup> They used the radial coordinate as a reaction coordinate since  $O_2^-$  is diatomic. Employing the energy coordinate as a reaction coordinate enables us to apply the RP theory to complex binding systems. Application of the RP theory to the inclusion systems that consist of  $\beta$ -cyclodextrin and small compounds yields the binding rate constants consistent with the experimental observations.<sup>48</sup>

Here, we present a methodology based on the RP theory applicable to elucidating protein–ligand binding kinetics with an improved treatment of free energy calculation. The free energy difference between the reactive and dissociate states is required for utilizing the RP theory. In the previous application,<sup>48</sup> the free energy calculation is performed with the potential of mean force (PMF).<sup>52</sup> For complex binding systems, however, the reliable estimate of the PMF involving the reactive and dissociate states requires high computational costs with MD simulations. Then, we construct a scheme of calculating the free energy difference based on the ER theory.<sup>21</sup> The derived expression of the free energy difference is applicable to arbitrary types of reactive states defined on reaction coordinates, and hence, combining the RP theory with ER theory is expected to be useful for various binding systems.

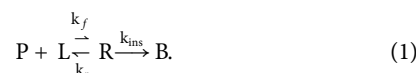
We apply the present method to the bindings of small fragment molecules to FK506 binding protein (FKBP) in an aqueous solution. FKBP is a receptor for immunosuppressive drugs, such as cyclosporin and FK506, and the FKBP–drug complexes inhibit the immune rejection reaction.<sup>53</sup> Thus, FKBP is recognized as an important drug target.<sup>54</sup> The binding rate constants for small fragments to FKBP are reported by Pan *et al.* from long timescale MD simulations performed on Anton2, an MD-specialized purpose machine, without any enhanced sampling methods.<sup>9</sup> Thus, these systems are suitable for testing the present method. We also show a

systematic analysis by decomposing the obtained rate constants into the kinetic and thermodynamic contributions for understanding binding mechanisms.

## II. THEORY

### A. Returning probability (RP) theory

Returning probability (RP) theory is a rigorous diffusion-influenced reaction (DIR) theory for elucidating a bimolecular reaction, originally proposed by Kim and Lee.<sup>50</sup> Recently, this theory was extended to atomistically describe host–guest binding processes by us.<sup>48</sup> We apply the theory to the protein (P)–ligand (L) binding kinetics. In the RP theory, the following reaction scheme is assumed (Fig. 1):



Here, R and B denote the reactive and bound states, respectively.  $k_f$  is the rate constant for forming state R from the dissociate state, and  $k_r$  is the rate constant for the dissociation process from state R.  $k_{\text{ins}}$  is the rate constant for the ligand insertion into the binding pocket of the protein. We define state R as the region that is close to the free energy barrier. The intermediate region covering the local minimum on the free energy profile can be also used as state R. The RP theory is based on the Liouville equation of the phase space densities with the reaction sink term that describes the insertion process. Hence, the RP theory could be useful for describing the binding kinetics in the heterogeneous environments, such as macromolecular crowded solutions.<sup>55</sup> The binding process refers to the conversion from the dissociate state to state B and passes through state R. The RP theory focuses on state R and facilitates the atomistic description of the thermodynamics and kinetics of binding.

Let us introduce the reaction coordinate,  $\Lambda$ , which distinguishes state B, state R, and dissociate state. Then, state R is defined on the  $\Lambda$  coordinate as  $Y$ . By introducing the characteristic function

$$\Theta(\Lambda) = \begin{cases} 1, & \Lambda \in Y, \\ 0, & \Lambda \notin Y, \end{cases} \quad (2)$$

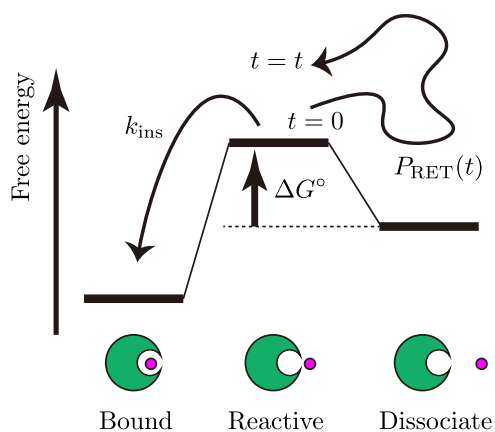


FIG. 1. Schematic illustration of returning probability (RP) theory.

the reaction sink is defined as

$$S(\Lambda) = k_{\text{ins}}\Theta(\Lambda), \quad (3)$$

which represents the frequency of the insertion events when the ligand is in state R. In the RP theory, the rate constant for the overall binding process at the steady state,  $k_{\text{on}}$ , is described as

$$k_{\text{on}} = K^* \left( \frac{1}{k_{\text{ins}}} + \int_0^\infty d\tau P_{\text{RET}}(\tau) \right)^{-1}. \quad (4)$$

Here,  $K^*$  is the equilibrium constant between the dissociate state and state R, represented with the standard free energy change for forming the latter,  $\Delta G^\circ$ , as

$$K^* = \frac{[R]}{[P][L]} = \frac{1}{c^\circ} e^{-\beta\Delta G^\circ}, \quad (5)$$

where  $c^\circ$  is the standard state concentration (1 mol/l) and  $\beta$  is the inverse temperature.  $[P]$ ,  $[L]$ , and  $[R]$  are the concentrations of P, L, and R, respectively.  $P_{\text{RET}}(t)$  is the returning probability, a conditional probability that the reactants form state R at time  $t = t$ , given that they formed state R at time  $t = 0$ . Using Eq. (2), the probability is defined as

$$P_{\text{RET}}(t) = \frac{\langle \Theta(\Lambda(t))\Theta(\Lambda(0)) \rangle}{\langle \Theta(\Lambda(0)) \rangle}, \quad (6)$$

where  $\langle \dots \rangle$  is the ensemble average in the system at  $t = 0$ . In the framework of the RP theory, the rate constant for the dissociation from state R to the dissociate state,  $k_r$ , can be expressed as

$$k_r = \left( \int_0^\infty d\tau P_{\text{RET}}(\tau) \right)^{-1}. \quad (7)$$

Equation (4) can be derived from the Liouville equation with the reaction sink term by assuming that the repeated returning to state R is a Markovian process.<sup>50</sup> Thus, state R should be narrow to assure the Markovianity on the state.

The previous study for the host–guest binding systems reveals that the host–guest interaction energy can effectively describe the reactants' relative position and orientation on one-dimensional space.<sup>48</sup> Furthermore, the structural distinction is realized more clearly with the attractive part of the Lennard-Jones (LJ) interaction defined as

$$U_{\text{attr}} = \sum_{i \in \text{protein}} \sum_{j \in \text{ligand}} u_{\text{attr},ij}, \quad (8)$$

$$u_{\text{attr},ij} = \begin{cases} 4\epsilon_{ij} \left\{ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right\}, & r_{ij} \geq 2^{1/6} \sigma_{ij}, \\ -\epsilon_{ij}, & r_{ij} < 2^{1/6} \sigma_{ij}, \end{cases} \quad (9)$$

where  $r_{ij}$  is the distance between atoms  $i$  and  $j$  and  $\sigma_{ij}$  and  $\epsilon_{ij}$  are the LJ parameters. Accordingly, we employ  $U_{\text{attr}}$  as the reaction coordinate for the protein–ligand binding system,  $\Lambda = U_{\text{attr}}$ . The energy range for state R,  $\mathbf{Y}_R = \{U_0 \leq U_{\text{attr}} \leq U_1\}$ , is determined based on the positions of the peak tops or shoulders appearing in the potential of mean force (PMF) on the  $U_{\text{attr}}$  coordinate.

## B. Theoretical expression of the free energy of forming state R $\Delta G^\circ$

In this section, we derive the theoretical expression of the free energy difference between state R and the dissociate state,  $\Delta G^\circ$ , in terms of the solvation free energies. Let us define the full coordinates of the protein and ligand as  $\mathbf{x}_P$  and  $\mathbf{x}_L$ , respectively, and define the set of full coordinates of the solvent molecules as  $\mathbf{X}_V$ . The theoretical expression of the chemical potential of species S ( $S = P$  or  $L$ ) under the NPT condition,  $\mu_S$ , was derived in Appendix A of Ref. 56 up to its Eq. (29). The derived expression is exact when the intramolecular energy of species S and the total potential of the solvent depend only on  $\mathbf{x}_S$  and  $\mathbf{X}_V$ , respectively. Note that this assumption is valid when the classical force fields with the pairwise additivity of non-bonded interactions are used. The same expression of  $\mu_S$  can be also derived under the NVT condition as described in Appendix A. The resultant expression of  $\mu_S$  is given by

$$\mu_S = -\frac{1}{\beta} \log \frac{Z_S}{\lambda_S V [S]} + \Delta\mu_S^{\text{bulk}}, \quad (10)$$

where  $V$  is the volume of the system,  $[S]$  is the concentration of species S whose dimension is the inverse of volume, and  $\lambda_S$  is the kinetic factor that comes from the integration of the partition function about the kinetic energy of species S. The dimension of  $\lambda_S$  is the same as that of  $\mathbf{x}_S$ .  $Z_S$  and  $\Delta\mu_S^{\text{bulk}}$  are the configurational integral of the isolated species S and solvation free energy, respectively. The definition of  $Z_S$  is

$$Z_S = \int d\mathbf{x}_S \exp[-\beta U_S(\mathbf{x}_S)], \quad (11)$$

where  $U_S(\mathbf{x}_S)$  is the intramolecular energy of species S. The dimension of  $Z_S$  is the same as that of  $\mathbf{x}_S$ . Since  $[S]$  has the dimension of the inverse of the volume,  $Z_S/\lambda_S V [S]$  is dimensionless. When the dilute condition of species S is imposed,  $\Delta\mu_S$  is given by

$$\Delta\mu_S^{\text{bulk}} = -\frac{1}{\beta} \log \frac{\int d\mathbf{x}_S \int d\mathbf{X}_V e^{-\beta \mathcal{V}_S^{\text{sol}}(\mathbf{x}_S, \mathbf{X}_V)}}{\int d\mathbf{x}_S \int d\mathbf{X}_V e^{-\beta \mathcal{V}_S^{\text{ref}}(\mathbf{x}_S, \mathbf{X}_V)}}, \quad (12)$$

where  $\mathcal{V}_S^{\text{sol}}(\mathbf{x}_S, \mathbf{X}_V)$  is the potential of the solution system of interest consisting of species S and solvents and  $\mathcal{V}_S^{\text{ref}}(\mathbf{x}_S, \mathbf{X}_V)$  is the potential of the reference solvent system with species S in which the interactions between species S and solvents are absent. By defining the total potential of the solvents as  $U_V(\mathbf{X}_V)$  and the interaction between species S and solvents as  $U_{SV}(\mathbf{x}_S, \mathbf{X}_V)$ ,  $\mathcal{V}_S^{\text{sol}}(\mathbf{x}_S, \mathbf{X}_V)$  and  $\mathcal{V}_S^{\text{ref}}(\mathbf{x}_S, \mathbf{X}_V)$  are, respectively, expressed as

$$\mathcal{V}_S^{\text{sol}}(\mathbf{x}_S, \mathbf{X}_V) = U_S(\mathbf{x}_S) + U_{SV}(\mathbf{x}_S, \mathbf{X}_V) + U_V(\mathbf{X}_V), \quad (13)$$

$$\mathcal{V}_S^{\text{ref}}(\mathbf{x}_S, \mathbf{X}_V) = U_S(\mathbf{x}_S) + U_V(\mathbf{X}_V). \quad (14)$$

Substitution of Eqs. (11), (12), and (14) into Eq. (10) leads to

$$\mu_S = \frac{1}{\beta} \log([S]\lambda_S) - \frac{1}{\beta} \log \frac{\int d\mathbf{x}_S \int d\mathbf{X}_V e^{-\beta \mathcal{V}_S^{\text{sol}}(\mathbf{x}_S, \mathbf{X}_V)}}{V \int d\mathbf{X}_V e^{-\beta U_V(\mathbf{X}_V)}}. \quad (15)$$

While the chemical potential for state R,  $\mu_R$ , is also expressed like Eq. (10), the definitions of the configurational integral of the isolated

$R$ ,  $Z_R$ , and solvation free energy,  $\Delta\mu_R^{\text{bulk}}$ , should be slightly modified. This is because the configurations of protein and ligand are restricted to the region corresponding to state  $R$ ,  $\Omega_R$ , defined in terms of  $\mathbf{x}_P$  and  $\mathbf{x}_L$ . By introducing the following characteristic function

$$\theta_R(\mathbf{x}_P, \mathbf{x}_L) = \begin{cases} 1, & (\mathbf{x}_P, \mathbf{x}_L) \in \Omega_R, \\ 0, & (\mathbf{x}_P, \mathbf{x}_L) \notin \Omega_R, \end{cases} \quad (16)$$

$Z_R$  and  $\Delta\mu_R^{\text{bulk}}$  are respectively written as

$$Z_R = \int d\mathbf{x}_P \int d\mathbf{x}_L \theta_R(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta(U_P(\mathbf{x}_P) + U_L(\mathbf{x}_L) + U_{PL}(\mathbf{x}_P, \mathbf{x}_L))}, \quad (17)$$

$$\Delta\mu_R^{\text{bulk}} = -\frac{1}{\beta} \log \frac{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V \theta_R(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta\mathcal{V}_R^{\text{sol}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}}{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V \theta_R(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta\mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}}. \quad (18)$$

Here,  $U_P(\mathbf{x}_P)$  and  $U_L(\mathbf{x}_L)$  are the intramolecular energies of protein and of ligand, respectively, and  $U_{PL}(\mathbf{x}_P, \mathbf{x}_L)$  is the protein–ligand interaction.  $\mathcal{V}_R^{\text{sol}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)$  and  $\mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)$  are, respectively, defined as

$$\mathcal{V}_R^{\text{sol}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V) = U_P(\mathbf{x}_P) + U_L(\mathbf{x}_L) + U_{PL}(\mathbf{x}_P, \mathbf{x}_L) + U_{PV}(\mathbf{x}_P, \mathbf{X}_V) + U_{LV}(\mathbf{x}_L, \mathbf{X}_V) + U_V(\mathbf{X}_V), \quad (19)$$

$$\mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V) = U_P(\mathbf{x}_P) + U_L(\mathbf{x}_L) + U_{PL}(\mathbf{x}_P, \mathbf{x}_L) + U_V(\mathbf{X}_V), \quad (20)$$

where  $U_{PV}(\mathbf{x}_P, \mathbf{X}_V)$  and  $U_{LV}(\mathbf{x}_L, \mathbf{X}_V)$  are protein–solvent and ligand–solvent interactions, respectively. Similar to Eq. (15),  $\mu_R$  can be described as

$$\mu_R = \frac{1}{\beta} \log ([R]\lambda_R) - \frac{1}{\beta} \log \frac{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V \theta_R(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta\mathcal{V}_R^{\text{sol}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}}{V \int d\mathbf{X}_V e^{-\beta U_V(\mathbf{X}_V)}}. \quad (21)$$

Note that  $\lambda_R$  can be decomposed as  $\lambda_R = \lambda_P \lambda_L$ . The dimension of the quantity inside the logarithm of the first term is canceled by that of the second term.

From the equilibrium condition,  $\mu_R - \mu_P - \mu_L = 0$ ,  $\Delta G^\circ$  is represented using Eqs. (5), (15), and (21) as

$$\Delta G^\circ = -\frac{1}{\beta} \log c^\circ K^* = \Delta\mu_L^R - \Delta\mu_L^{\text{bulk}} + \Delta G_{\text{corr}}^\circ, \quad (22)$$

where  $\Delta\mu_L^R$  is the solvation free energy of the ligand conditioned by  $\theta_R(\mathbf{x}_P, \mathbf{x}_L)$  as

$$\Delta\mu_L^R = -\frac{1}{\beta} \log \frac{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V \theta_R(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta\mathcal{V}_R^{\text{sol}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}}{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V \theta_R(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta\mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}} \quad (23)$$

and  $\Delta G_{\text{corr}}^\circ$  is the correction term defined as

$$\Delta G_{\text{corr}}^\circ = -\frac{1}{\beta} \log \left( c^\circ V \frac{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V \theta_R(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta\mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}}{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V e^{-\beta\mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}} \right). \quad (24)$$

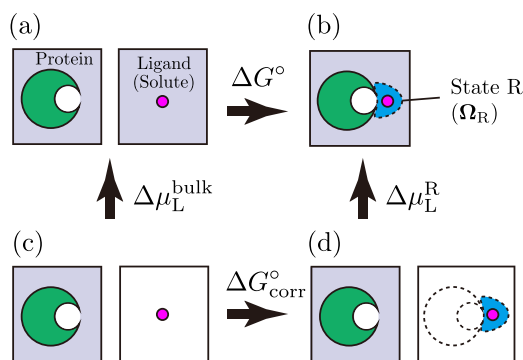
Here,  $\mathcal{V}_R^{\text{sol}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)$  and  $\mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)$  are, respectively, defined as

$$\mathcal{V}_R^{\text{sol}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V) = \mathcal{V}_R^{\text{sol}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V), \quad (25)$$

$$\mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V) = U_P(\mathbf{x}_P) + U_L(\mathbf{x}_L) + U_{PV}(\mathbf{x}_P, \mathbf{X}_V) + U_V(\mathbf{X}_V). \quad (26)$$

Note that the interaction between the protein and ligand [ $U_{PL}(\mathbf{x}_P, \mathbf{x}_L)$ ] and that between ligand and solvents [ $U_{LV}(\mathbf{x}_L, \mathbf{X}_V)$ ] are absent in  $\mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)$ . Accordingly, Eq. (23) is the free-energy change for turning on the intermolecular interactions of ligand  $L$  with protein  $P$  and solvent  $V$  under the condition imposed by  $\theta_R(\mathbf{x}_P, \mathbf{x}_L)$ .  $\Delta\mu_L^R$  is called the solvation free energy at Eq. (23) by viewing the ligand as the solute and the protein and solvent together as mixed solvents.  $\Delta\mu_L^{\text{bulk}}$  is the “usual” solvation free energy of the ligand in the bulk solvent as provided by Eq. (12). The logarithm in Eq. (24) can be computed by randomly inserting the ligand into the configurations of the protein and solvent molecules. This is a procedure of test-particle insertion, and the ligand is only placed in the protein–solvent system without affecting their configurations. Furthermore, only the calculation of the protein–ligand interaction employed as the reaction coordinate is required after the insertion, and the solvent configurations are not relevant with the computation. Thus, sampling the configurations corresponding to state  $R$  is possible by means of random insertion. Since the insertion process can be performed with low computational cost, the accurate calculation of  $\Delta G_{\text{corr}}^\circ$  is realized. Since  $\Delta G_{\text{corr}}^\circ$  involves the volume of the system,  $\Delta G_{\text{corr}}^\circ$  appears to be dependent on the system size. Actually, it can be proved that  $\Delta G_{\text{corr}}^\circ$  has no such a dependency (see Appendix B). A scheme of efficiently calculating  $\Delta G_{\text{corr}}^\circ$  is found in Appendix C. It should be further noted that the choice of the standard state affects only  $\Delta G_{\text{corr}}^\circ$  through  $c^\circ$ .

Equation (22) can be interpreted with a simple thermodynamic cycle involving the solvation processes of a ligand shown in Fig. 2. In this scheme, we regard the ligand as a solute and the protein as a part of the mixed solvent together with water and salts (if contained in the system). The state change (c)  $\rightarrow$  (a) indicates the solvation process of



**FIG. 2.** Thermodynamic cycle for forming the reactive ( $R$ ) state from the dissociate state. (a) Dissociate state. (b) State  $R$ . (c) Protein and ligand are in solution and in the gas phase, respectively. (d) Protein and ligand are in solution and in the gas phase, respectively, while the configuration of ligand is restricted to the region corresponding to state  $R$ .

a ligand from the gas phase to the bulk, and its free energy difference is  $\Delta\mu_L^{\text{bulk}}$ . The change (d)  $\rightarrow$  (b) associated with the free energy difference,  $\Delta\mu_L^{\text{R}}$ , is also the solvation of the ligand, but the configuration of the ligand is restricted to state R both in the gas and solution phases. In order to bridge the above two processes, it is necessary to consider an additional process [(c)  $\rightarrow$  (d)] that the ligand is brought to state R in the gas phase. The free energy change during this process is  $\Delta G_{\text{corr}}^{\circ}$ . Equation (24) is a generalization of the standard correction term derived by Gilson *et al.*, which is useful when the position and orientation of the ligand are used as the reaction coordinates.<sup>13</sup> It should be noted that Eq. (24) is applicable to arbitrary types of reaction coordinates. We describe the ER theory for computing  $\Delta\mu_L^{\text{bulk}}$  and  $\Delta\mu_L^{\text{R}}$  in Secs. II C and II D, respectively.

The above formulation in this section was performed in the NVT ensemble. Our developments are also valid in NPT when  $U_V(\mathbf{X}_V)$  of Eqs. (13)–(15), (19)–(21), and (26) is replaced to  $U_V(\mathbf{X}_V) + pV$  and the configurational integral is written with the integration over  $V$ , where  $p$  and  $V$  refer to the pressure and volume of the system, respectively. Furthermore, the solvation free energies and the free-energy changes obtained from them have the same values in the NPT and NVT ensembles as far as the system is large enough (thermodynamic limit) and the pressure for the NPT ensemble and the volume for NVT are connected through the right equation of state.

### C. Solvation free energy of ligand in the bulk $\Delta\mu_L^{\text{bulk}}$

The theoretical expression of  $\Delta\mu_L^{\text{bulk}}$  associated with the solvation process from the gas phase is obtained by considering  $S = L$  in Eq. (12) as

$$\Delta\mu_L^{\text{bulk}} = -\frac{1}{\beta} \log \frac{\int d\mathbf{x}_L \int d\mathbf{X}_V e^{-\beta \mathcal{V}_L^{\text{sol}}(\mathbf{x}_L, \mathbf{X}_V)}}{\int d\mathbf{x}_L \int d\mathbf{X}_V e^{-\beta \mathcal{V}_L^{\text{ref}}(\mathbf{x}_L, \mathbf{X}_V)}}. \quad (27)$$

The ER theory provides a theoretical expression for  $\Delta\mu_L^{\text{bulk}}$  with the approximated functional  $\mathcal{F}_{\text{ER}}$  as<sup>21</sup>

$$\Delta\mu_L^{\text{bulk}} = \sum_{\alpha} \int_{-\infty}^{\infty} d\epsilon \epsilon \rho_{\alpha}^e(\epsilon) + \mathcal{F}_{\text{ER}}[\{\rho_{\alpha}^e(\epsilon)\}, \{\rho_{\alpha,0}^e(\epsilon)\}, \{\chi_{\alpha\beta,0}^e(\epsilon, \eta)\}], \quad (28)$$

where  $\rho_{\alpha}^e(\epsilon)$  and  $\rho_{\alpha,0}^e(\epsilon)$  are the energy distribution functions for  $\alpha$ th solvent species in the solution and in the reference solvent, respectively, and  $\chi_{\alpha\beta,0}^e(\epsilon, \eta)$  is the two-body energy correlation function between the  $\alpha$ th and  $\beta$ th solvent species in the reference solvent. The definitions of these functions are

$$\rho_{\alpha}^e(\epsilon) = \left\langle \sum_i \delta(u_{\alpha}(\mathbf{x}_L, \mathbf{x}_i) - \epsilon) \right\rangle, \quad (29)$$

$$\rho_{\alpha,0}^e(\epsilon) = \left\langle \sum_i \delta(u_{\alpha}(\mathbf{x}_L, \mathbf{x}_i) - \epsilon) \right\rangle_0, \quad (30)$$

$$\chi_{\alpha\beta,0}^e(\epsilon, \eta) = \left\langle \sum_{ij} \delta(u_{\alpha}(\mathbf{x}_L, \mathbf{x}_i) - \epsilon) \delta(u_{\beta}(\mathbf{x}_L, \mathbf{x}_j) - \eta) \right\rangle_0 - \rho_{\alpha,0}^e(\epsilon) \rho_{\beta,0}^e(\eta). \quad (31)$$

Here,  $\langle \dots \rangle$  and  $\langle \dots \rangle_0$ , respectively, indicate the ensemble averages in the solution and in the reference solvent, where the solution refers to the system sampled with  $\mathcal{V}_L^{\text{sol}}(\mathbf{x}_L, \mathbf{X}_V)$  and the reference solvent is generated by  $\mathcal{V}_L^{\text{ref}}(\mathbf{x}_L, \mathbf{X}_V)$ .  $u_{\alpha}(\mathbf{x}_L, \mathbf{x}_i)$  is the interaction between the ligand and the  $i$ th solvent molecule of species  $\alpha$  whose full coordinate is  $\mathbf{x}_i$ .

### D. Solvation free energy of ligand for state R $\Delta\mu_L^{\text{R}}$

In this subsection, after describing the solvation free energy of the ligand for arbitrary state A defined in terms of  $\mathbf{x}_L$  and  $\mathbf{x}_P$ ,  $\Delta\mu_L^{\text{A}}$ , we show the scheme of computing  $\Delta\mu_L^{\text{R}}$ . In state A, the configuration of the ligand is restricted, and hence,  $\Delta\mu_L^{\text{A}}$  can be expressed by replacing  $\theta_{\text{R}}(\mathbf{x}_P, \mathbf{x}_L)$  in Eq. (23) with  $\theta_{\text{A}}(\mathbf{x}_P, \mathbf{x}_L)$ . We first define the following characteristic function associated with state A as

$$\theta_{\text{A}}(\mathbf{x}_P, \mathbf{x}_L) = \begin{cases} 1, & (\mathbf{x}_P, \mathbf{x}_L) \in \Omega_{\text{A}}, \\ 0, & (\mathbf{x}_P, \mathbf{x}_L) \notin \Omega_{\text{A}}, \end{cases} \quad (32)$$

where  $\Omega_{\text{A}}$  is the region corresponding to state A. Then,  $\Delta\mu_L^{\text{A}}$  is expressed as

$$\Delta\mu_L^{\text{A}} = -\frac{1}{\beta} \log \frac{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V \theta_{\text{A}}(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta \mathcal{V}_R^{\text{sol}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}}{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V \theta_{\text{A}}(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta \mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}}. \quad (33)$$

Let us introduce the conditional ensemble average as

$$\langle \dots \rangle_{\text{A}} = \frac{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V (\dots) \theta_{\text{A}}(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta \mathcal{V}_R^{\text{sol}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}}{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V \theta_{\text{A}}(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta \mathcal{V}_R^{\text{sol}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}}, \quad (34)$$

$$\langle \dots \rangle_{0,\text{A}} = \frac{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V (\dots) \theta_{\text{A}}(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta \mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}}{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V \theta_{\text{A}}(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta \mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}}. \quad (35)$$

In parallel to the note for Eqs. (29)–(31), the systems described by Eqs. (34) and (35) are called the solution and reference solvent, respectively. The conditional energy distribution and correlation functions are written as

$$\rho_{\alpha}^{e(\text{A})}(\epsilon) = \left\langle \sum_i \delta(u_{\alpha}(\mathbf{x}_L, \mathbf{x}_i) - \epsilon) \right\rangle_{\text{A}}, \quad (36)$$

$$\rho_{\alpha,0}^{e(\text{A})}(\epsilon) = \left\langle \sum_i \delta(u_{\alpha}(\mathbf{x}_L, \mathbf{x}_i) - \epsilon) \right\rangle_{0,\text{A}}, \quad (37)$$

$$\chi_{\alpha\beta,0}^{e(\text{A})}(\epsilon, \eta) = \left\langle \sum_{ij} \delta(u_{\alpha}(\mathbf{x}_L, \mathbf{x}_i) - \epsilon) \delta(u_{\beta}(\mathbf{x}_L, \mathbf{x}_j) - \eta) \right\rangle_{0,\text{A}} - \rho_{\alpha,0}^{e(\text{A})}(\epsilon) \rho_{\beta,0}^{e(\text{A})}(\eta), \quad (38)$$

where  $\alpha$  and  $\beta$  refer to the protein or solvents.  $\Delta\mu_L^{\text{A}}$  is expressed using the ER theory as

$$\Delta\mu_L^{\text{A}} = \sum_{\alpha} \int_{-\infty}^{\infty} d\epsilon \epsilon \rho_{\alpha}^{e(\text{A})}(\epsilon) + \mathcal{F}_{\text{ER}}[\{\rho_{\alpha}^{e(\text{A})}(\epsilon)\}, \{\rho_{\alpha,0}^{e(\text{A})}(\epsilon)\}, \{\chi_{\alpha\beta,0}^{e(\text{A})}(\epsilon, \eta)\}]. \quad (39)$$

The conditional ensemble averages in the reference solvent can be computed by considering the solvent configurations with the inserted ligand whose configurations are restricted to  $\Omega_A$ . On the other hand, if state A is defined only with  $U_{\text{attr}}$  [Eq. (8)] as  $\mathbf{Y}_A = \{U_{0,A} \leq U_{\text{attr}} \leq U_{1,A}\}$ , the locations of the inserted ligand sometimes can differ significantly from those observed in the solution system; the  $U_{\text{attr}}$  value alone may not fully capture the binding configurations in the solution and can be very “coarse” to specify insertion configurations in the reference solvent. In order to supplement the information for the binding configurations, we introduce the center of mass (CoM) of the ligand with respect to the CoM of the protein,  $\mathbf{r}_{\text{PL}}$ , and the minimum of the interatomic distances between the protein and ligand,  $r_{\text{min}}$ , as the auxiliary reaction coordinates. Then, we define  $\Omega_A$  as

$$\Omega_A = \{U_{\text{attr}} \in \mathbf{Y}_A \wedge \rho_A(\mathbf{r}) > 0 \wedge r_{\text{min}} \geq R_{\text{sol}}\}. \quad (40)$$

Here,  $\rho_A(\mathbf{r})$  is the spatial density of the ligand with respect to the protein in the solution system defined as

$$\rho_A(\mathbf{r}) = \frac{\langle \delta(\mathbf{r} - \mathbf{r}_{\text{PL}}) \rangle_{\mathbf{Y}_A}}{\int d\mathbf{r} \langle \delta(\mathbf{r} - \mathbf{r}_{\text{PL}}) \rangle_{\mathbf{Y}_A}}, \quad (41)$$

where  $\langle \dots \rangle_{\mathbf{Y}_A}$  stands for the ensemble average in the solution system conditioned by  $U_{\text{attr}} \in \mathbf{Y}_A$  and  $R_{\text{sol}}$  is the minimum of  $r_{\text{min}}$  observed in the solution system. Note that  $\rho_A(\mathbf{r}) > 0$  and  $r_{\text{min}} \geq R_{\text{sol}}$  are satisfied in the solution system when the system is in state R with  $U_{\text{attr}} \in \mathbf{Y}_A$ . Since  $\Omega_A$  [Eq. (40)] involves  $\rho_A(\mathbf{r})$  and  $R_{\text{sol}}$ , we need to run the MD simulations for the solution system before computing Eqs. (37) and (38). A scheme of the test-particle insertion that satisfies Eq. (40) is found in Sec. S2 of the supplementary material.

The theoretical expression of  $\Delta\mu_L^R$  can be described by considering  $A = R$  in Eq. (33), and the application of the ER theory to state R is straightforward. However, since state R is defined as the narrow region with high free energy on the reaction coordinate, it is difficult to obtain the adequate sampling for computing the energy distribution and correlation functions needed in Eq. (39). Then, we consider the intermediate (IM) state located between state R and the dissociate state in which the robust computation of these functions can be achieved. Adopting Eqs. (32) and (33) to states R and IM gives

$$\Delta\mu_L^R - \Delta\mu_L^{\text{IM}} = -\frac{1}{\beta} \log \frac{\langle \theta_R \rangle}{\langle \theta_R \rangle_0} + \frac{1}{\beta} \log \frac{\langle \theta_{\text{IM}} \rangle}{\langle \theta_{\text{IM}} \rangle_0}, \quad (42)$$

where  $\langle \dots \rangle$  and  $\langle \dots \rangle_0$  stand for the ensemble averages in the solution system governed by  $\mathcal{V}_R^{\text{sol}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)$  [Eq. (25)] and in the reference solvent system governed by  $\mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)$  [Eq. (26)], respectively. Equation (42) can be rewritten as

$$\Delta\mu_L^R = \Delta\mu_L^{\text{IM}} - \frac{1}{\beta} \log \frac{\langle \theta_R \rangle}{\langle \theta_{\text{IM}} \rangle} + \frac{1}{\beta} \log \frac{\langle \theta_R \rangle_0}{\langle \theta_{\text{IM}} \rangle_0}. \quad (43)$$

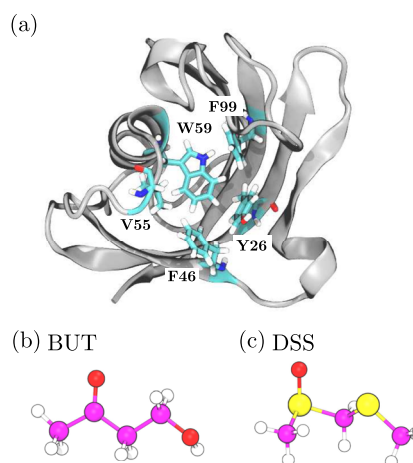
Since  $\langle \theta_R \rangle / \langle \theta_{\text{IM}} \rangle$  and  $\langle \theta_R \rangle_0 / \langle \theta_{\text{IM}} \rangle_0$ , respectively, represent the population ratios in the solution system and in the reference solvent system, these quantities can be computed through molecular simulations in the solution system and with random insertion of the ligand into the solvent system.

### III. COMPUTATIONAL METHODS

We investigated two different protein–ligand binding systems composed of FK506 binding protein (FKBP) and fragment molecules, 4-hydroxy-2-butanone (BUT) and methyl methylthiomethyl sulphoxide (DSS) in an aqueous solution at 300 K (Fig. 3). The modeling for each molecule is described in Sec. III A, and simulation procedures are in Secs. III C and III D. The details of the equilibration schemes in the MD simulations are found in Tables S1–S5 in the supplementary material. The initial configurations of the systems of interest were built using Packmol.<sup>57</sup> All the MD simulations were performed with GENESIS2.0<sup>58–60</sup> The Bussi thermostat was used for temperature control in the NVT and NPT simulations, and the Bussi barostat was used for the NPT simulations.<sup>61</sup> The velocity Verlet integrator<sup>62</sup> and reversible reference system propagator algorithm (r-RESPA)<sup>63</sup> were employed for the equilibration and production runs, respectively. The cutoff distance for the Lennard-Jones interactions was 9.0 Å. We employed smooth particle-mesh Ewald (SPME)<sup>64</sup> for computing the electrostatic interactions, and the number of grids was automatically determined in GENESIS2.0 so that the grid spacing was shorter than 1.4 Å. All bonds involving hydrogen atoms were constrained using the SHAKE/RATTLE method,<sup>65,66</sup> and water molecules were kept rigid using the SETTLE method.<sup>67</sup>

#### A. Molecular models

The structure of FKBP was taken from the crystal structure provided in RCSB PDB (PDB-ID: 1D7H).<sup>68</sup> The used force field for the protein and ions was the ff99SB\*-ILDN force field.<sup>69,70</sup> As for the ligands, we used the generalized Amber force field (GAFF).<sup>71,72</sup> The TIP3P model was used for water molecules. The optimized structures of ligand molecules were obtained with the quantum chemical calculation at the MP2/6-31G(d) level, and then, the restrained electrostatic potential (RESP) charges<sup>73</sup> were evaluated using the Antechamber<sup>74</sup> program based on HF/6-31G(d) level



**FIG. 3.** Molecular structures of (a) FK506 binding protein (FKBP), (b) 4-hydroxy-2-butanone (BUT), and (c) methyl methylthiomethyl sulphoxide (DSS). The residues composing the binding pocket of FKBP are highlighted in (a).



calculations. All the quantum chemical calculations were performed with Gaussian 16.<sup>75</sup>

## B. Equilibration of FKBP structure

We first performed the MD simulations of FKBP immersed in the 150 mM NaCl aqueous solution at 300 K to obtain the equilibrated structure of FKBP. The numbers of water, Na<sup>+</sup>, and Cl<sup>-</sup> were 32 600, 88, and 89, respectively. Since FKBP has a net charge of +1|e|, the number of Cl<sup>-</sup> added to the system is one more than that of Na<sup>+</sup> for charge neutrality. The initial box size was 100<sup>3</sup> Å<sup>3</sup>. After the equilibration with NVT and NPT simulations, we conducted 100 ns MD (NVT) simulations. The final structure of FKBP was used as the initial structure in the succeeding MD simulations described in the following.

## C. MD simulations for protein-ligand systems

We prepared the protein-ligand systems containing FKBP and ligand (BUT or DSS) immersed in 150 mM NaCl aqueous solutions whose initial box sizes were 100<sup>3</sup> Å<sup>3</sup>. The numbers of water, Na<sup>+</sup>, and Cl<sup>-</sup> were 32 600, 88, and 89, respectively. We equilibrated the systems with the protein and ligand being separated by introducing the flat-bottom (FB) potential defined as

$$U_{\text{FB}}^d(d) = \begin{cases} k(d - d_1)^2, & d \leq d_1, \\ 0, & d_1 < d \leq d_2, \\ k(d - d_2)^2, & d > d_2, \end{cases} \quad (44)$$

where  $d$  is the distance between the CoM of Trp59 and CoM of a ligand and hydrogen atoms were omitted in the calculations of CoMs. For both the ligands, the values of  $k$ ,  $d_1$ , and  $d_2$  were set to 1 kcal mol<sup>-1</sup> Å<sup>-2</sup>, 17 Å, and 18 Å, respectively. The box sizes obtained after the equilibration (NPT) were 100.329<sup>3</sup> and 100.386<sup>3</sup> Å<sup>3</sup> for the BUT and DSS systems, respectively. We sampled 20 configurations during the above simulations. Then, 60 ns MD (NVT) simulations were conducted from each of the sampled configurations with different random seeds of thermostat while imposing the half flat-bottom (HFB) potential defined as

$$U_{\text{HFB}}^d(d) = \begin{cases} 0, & d \leq d_2, \\ k(d - d_2)^2, & d > d_2. \end{cases} \quad (45)$$

For the accurate estimation of the PMFs, we performed the replica-exchange umbrella sampling (REUS)<sup>76</sup> on the energy coordinate,  $U_{\text{attr}}$  [Eqs. (8) and (9)]. The initial configurations for the REUS simulations were sampled from the above 20 runs with the HFB potential. The atoms in the sidechains of Tyr26, Phe46, Val55, Trp59, and Phe99 (Fig. 3) were considered for the calculations of  $U_{\text{attr}}$ . The configurations that satisfy  $-5.0 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -1.0$  were sampled as the initial ones with an interval of  $1 \pm 0.1$  kcal mol<sup>-1</sup> for BUT. As for DSS, the configurations that satisfy  $-7.0 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -1.0$  were taken with an interval of  $1 \pm 0.1$  kcal mol<sup>-1</sup>. The following harmonic potential was employed for the  $i$ th window in the REUS simulations:

$$U_{\text{harm}}(U_{\text{attr}}) = k_i(U_{\text{attr}} - U_i)^2. \quad (46)$$

For both the ligands, the force constants,  $k_i$ , were fixed to 0.8 kcal<sup>-1</sup> mol, and the initial values of reference energies,  $U_i$ , were set to -5, -4, -3, -2, and -1 kcal mol<sup>-1</sup> for BUT and to -7, -6, -5, -4, -3, -2, and -1 kcal mol<sup>-1</sup> for DSS. Performing the parameter tuning simulations (NVT)<sup>77</sup> implemented in Genesis2.0 yielded  $U_i = -4.60, -3.75, -2.93, -1.98, \text{ and } -1.00$  kcal mol<sup>-1</sup> for BUT and  $U_i = -6.82, -6.05, -5.04, -3.93, -2.88, -2.11, \text{ and } -1.00$  kcal mol<sup>-1</sup> for DSS. Then, we conducted the production REUS (NVT) simulations with the exchange period of 1 ps while imposing the following HFB potential for avoiding the sampling of the configurations in the dissociate state that are far separated from R:

$$U_{\text{HFB}}^{\text{attr}}(U_{\text{attr}}) = \begin{cases} 0, & U_{\text{attr}} \leq U_{\text{dissoc}}, \\ k(U_{\text{attr}} - U_{\text{dissoc}})^2, & U_{\text{attr}} > U_{\text{dissoc}}. \end{cases} \quad (47)$$

Here,  $k = 10$  kcal<sup>-1</sup> mol and  $U_{\text{dissoc}} = -0.5$  kcal mol<sup>-1</sup> for both the ligands. The trajectory lengths for production were 250 and 350 ns for BUT and DSS, respectively.

The configurations obtained from the REUS simulations were used as the initial configurations for the simulations to calculate  $P_{\text{RET}}(t)$  and  $k_{\text{ins}}$ . The number of sampled configurations at  $-4.3 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -3.0$  was 300 for BUT, and that at  $-2.8 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -1.5$  was 500 for DSS. For the calculations of  $P_{\text{RET}}(t)$ , after the initialization of velocities and equilibrations, we performed 20 and 30 ns MD (NVT) simulations for BUT and DSS, respectively, with the following HFB potential for each trajectory:

$$U_{\text{HFB}}^{\text{attr}}(U_{\text{attr}}) = \begin{cases} k(U_{\text{attr}} - U_{\text{ins}})^2, & U_{\text{attr}} \leq U_{\text{ins}}, \\ 0, & U_{\text{attr}} > U_{\text{ins}}. \end{cases} \quad (48)$$

Here,  $k = 10.0$  kcal<sup>-1</sup> mol and  $U_{\text{ins}} = -4.3$  kcal mol<sup>-1</sup> for BUT and  $k = 10.0$  kcal<sup>-1</sup> mol and  $U_{\text{ins}} = -2.8$  kcal mol<sup>-1</sup> for DSS. As for  $k_{\text{ins}}$ , we conducted 30 ns MD (NVT) simulations for both the ligands while imposing the HFB potential defined in Eq. (47) with  $k = 10.0$  kcal<sup>-1</sup> mol and  $U_{\text{dissoc}} = -3.0$  kcal mol<sup>-1</sup> for BUT and with  $k = 10.0$  kcal<sup>-1</sup> mol and  $U_{\text{dissoc}} = -1.5$  kcal mol<sup>-1</sup> for DSS.

## D. Additional MD simulations for evaluating solvation free energies

In order to evaluate the solvation free energies in the bulk,  $\Delta\mu_{\text{L}}^{\text{bulk}}$ , by means of the ER theory, we need to perform the MD simulations for the systems with and without a ligand in 150 mM aqueous solutions. As for the system with BUT or DSS (solution system), the numbers of water, Na<sup>+</sup>, and Cl<sup>-</sup> were 32 600, 88, and 88, respectively. After the equilibrations under the NVT and NPT conditions, the box sizes for BUT and DSS were obtained as 99.767<sup>3</sup> and 99.788<sup>3</sup> Å<sup>3</sup>, respectively. Then, we conducted 80 and 500 ns MD (NVT) simulations for production. In the cases of the systems without ligands (reference solvent systems), the composition of the solvents and box sizes were set to be the same as those for the systems with ligands. We performed the MD (NVT) simulations for equilibration, followed by 10 ns MD (NVT) simulations for production.

For the computation of  $\Delta\mu_{\text{L}}^{\text{R}}$ , the simulations for protein-ligand systems (solution systems) and protein systems (reference solvent systems) in 150 mM aqueous solutions are needed. In the cases of

the protein–ligand systems, the trajectories obtained from the REUS simulations described in Sec. III C can be also used for this purpose. We prepared the protein systems whose solvent compositions and box sizes were the same as those for the protein–ligand systems. Then, we performed 50 ns MD (NVT) simulations for production after the equilibration.

We prepared the configurations of the isolated ligands by performing the MD simulations in the gas phase. After the equilibration using the 2 ns MD (NVT) simulations, we conducted 1  $\mu$ s MD (NVT) simulations for production. The number of obtained configurations was 1 000 000, and these configurations were used for the insertion.

### E. Computation of thermodynamic and kinetic quantities

The PMFs on the energy coordinate were evaluated using the trajectories obtained from the REUS simulations. The weight of each snapshot was estimated with the multistate Bennet acceptance ratio (MBAR) method<sup>78</sup> implemented in GENESIS2.0.<sup>79</sup> The solvation free energies were computed by means of ERmod 0.3.7.<sup>20</sup> For the computation of  $\Delta\mu_L^{\text{bulk}}$ , the ligand was randomly inserted into the corresponding reference solvent. The trajectory of the solution system was splitted into 20 blocks, and then, the average of  $\Delta\mu_L^{\text{bulk}}$  and its standard error were computed. As for  $\Delta\mu_L^{\text{IM}}$ , we used the trajectories obtained from the REUS simulations for the solution system. As well as in the case of  $\Delta\mu_L^{\text{bulk}}$ , we splitted the REUS trajectories into 20 blocks for the error estimation. In order to calculate  $\Delta\mu_L^{\text{R}}$  through Eq. (43), we computed  $\Delta G_{\text{IM}\rightarrow\text{R}} = (-1/\beta)\log\langle\theta_{\text{R}}\rangle/\langle\theta_{\text{IM}}\rangle$  and  $\Delta G_{0,\text{IM}\rightarrow\text{R}} = (-1/\beta)\log\langle\theta_{\text{R}}\rangle_0/\langle\theta_{\text{IM}}\rangle_0$  from the MD simulations for the solution system (REUS) and for the reference solvent system, respectively. The average and standard error of  $\Delta G_{\text{IM}\rightarrow\text{R}}$  were computed through the Monte-Carlo (MC) bootstrap method.<sup>80</sup> In this method, the number of bootstrap samples generated by selecting the frames was set to 100. For the computation of  $\Delta G_{0,\text{IM}\rightarrow\text{R}}$ , the test particle insertion was performed, and the reference solvent configurations with the inserted solute were splitted into 20 blocks for the error estimation. The numbers of insertions for calculating  $\Delta\mu_L^{\text{bulk}}$ ,  $\Delta\mu_L^{\text{IM}}$ ,  $\Delta G_{\text{corr}}^{\circ}$ , and  $\Delta G_{0,\text{IM}\rightarrow\text{R}}$  were described in Table S6 of the supplementary material.

The rate constant of insertion,  $k_{\text{ins}}$ , is given by

$$k_{\text{ins}} = \frac{\sum_{i=1}^{N_{\text{traj}}} \delta_{\text{ins}}^{(i)}}{\sum_{i=1}^{N_{\text{traj}}} \sum_{j=1}^{N_{\text{step}}} \Theta(U_{\text{attr}}^{(i)}(t_j)) \Delta t}. \quad (49)$$

Here,  $N_{\text{traj}}$  is the number of trajectories and  $N_{\text{step}}$  is the number of time steps until the insertion event is observed for the first time.  $\Delta t$  is the time interval between adjacent frames. The value of  $\Delta t$  was set to 1 ps.  $U_{\text{attr}}^{(i)}(t)$  is the time series of  $U_{\text{attr}}$  obtained from the  $i$ th trajectory.  $\Theta$  is a characteristic function for state R that is the same as Eq. (2) when  $U_{\text{attr}}$  is used as the reaction coordinate  $\Lambda$ .  $\delta_{\text{ins}}^{(i)}$  is a characteristic function for insertion, which is unity when the insertion event is observed in the  $i$ th trajectory and vanishes otherwise. Each MD is terminated when the insertion event is observed. Accordingly,  $\delta_{\text{ins}}^{(i)}$  can be only one or zero. When  $\delta_{\text{ins}}^{(i)} = 0$ , there are no insertions in the  $i$ th trajectory and  $N_{\text{step}}$  is the same as the total number of steps in the trajectory. The entry of the ligand into the region defined as

$U_{\text{attr}} \leq -12.0$  kcal mol<sup>-1</sup> for BUT and as  $U_{\text{attr}} \leq -11.5$  kcal mol<sup>-1</sup> for DSS during the simulation was considered as insertion. The MC bootstrap method was used for the error estimations of  $k_{\text{ins}}$  and  $P_{\text{RET}}(t)$ . The number of bootstrap samples generated by selecting the trajectories was set to 1000 for both the quantities.

## IV. RESULTS AND DISCUSSION

### A. Potentials of mean force (PMFs)

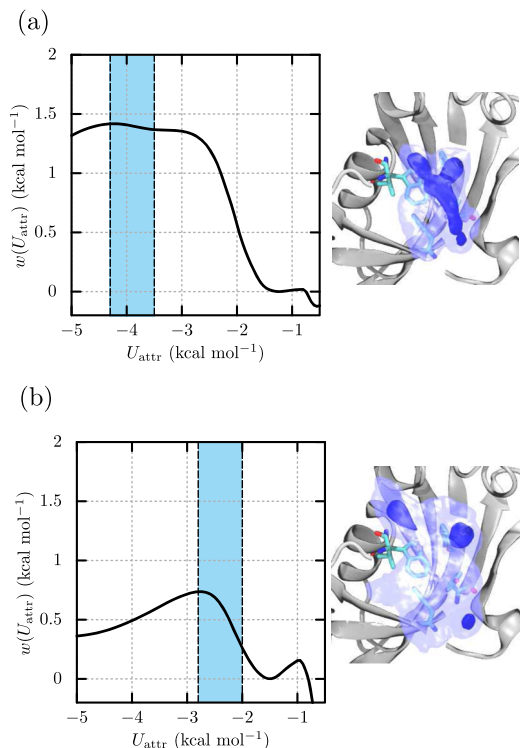
We first examine the potentials of mean force (PMFs) on the energy coordinate,  $w(U_{\text{attr}})$ , from the REUS simulations reweighted with the MBAR method (Fig. 4). The spatial densities of ligands for state R are also shown in Fig. 4. We define the energy ranges for state R as  $-4.3 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -3.5$  for BUT and  $-2.8 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -2.0$  for DSS. The lower bounds of state R correspond to the free energy barriers in the PMFs for both the ligands. The choice of the energy range to define state R will be addressed in Sec. IV D. We set the energy ranges so that the binding rate constants are not sensitive to the variations in the upper bounds. In the case of BUT [Fig. 4(a)], the profile in the PMF is almost flat around the free energy barrier [ $-4.3 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -2.8$ ]. The spatial density for state R is distributed around the entrance of the binding pocket. The region corresponding to the high population is found near the hydrophobic residues (Tyr26, Phe46, and Phe99). From the definition of  $U_{\text{attr}}$  [Eqs. (8) and (9)], the decrease in  $U_{\text{attr}}$  corresponds to the increase in the number of contacts between the binding pocket of the protein and ligand. Hence, the flat region on the PMF indicates that the ligand configurations with different contact patterns with the protein show similar thermodynamic stability in this region. A local-minimum region is discernible around  $-1.5 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -0.75$ , and no barrier exists between the local minimum and dissociate ( $U_{\text{attr}} \sim 0$  kcal mol<sup>-1</sup>) states. In comparison with the case of BUT, DSS has a lower free energy barrier located at  $U_{\text{attr}} = -2.8$  kcal mol<sup>-1</sup>, and no plateau exists around the barrier. The spatial density for state R is delocalized around the binding pocket compared with BUT, and the high populations exist separately around Tyr26, Val55, and Phe99. Unlike in the case of BUT, in addition, the local minimum found at  $U_{\text{attr}} \sim -1.5$  kcal mol<sup>-1</sup> and dissociate state ( $U_{\text{attr}} \sim 0$  kcal mol<sup>-1</sup>) are separated by a free energy barrier at  $U_{\text{attr}} \sim -1.0$  kcal mol<sup>-1</sup>.

### B. Solvation free energies

We discuss the solvation free energies of the ligands for the bulk ( $\Delta\mu_L^{\text{bulk}}$ ) and for state R ( $\Delta\mu_L^{\text{R}}$ ) obtained through the ER theory. In order to calculate the solvation free energy for state R,  $\Delta\mu_L^{\text{R}}$ , using Eq. (43), we define state IM as  $-2.0 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -1.0$  in which the local minima are present for both the ligands (Fig. 4). In the framework of endpoint DFT, the solvation free energy for state A,  $\Delta\mu_L^{\text{A}}$ , can be exactly described as

$$\Delta\mu_L^{\text{A}} = \langle U_{\text{LE}} \rangle_{\text{A}} + \Delta\mu_{\text{L,res}}^{\text{A}}, \quad (50)$$

where  $U_{\text{LE}}$  is the average interaction energy of a ligand with the surrounding environments, such as solvents and protein, and  $\Delta\mu_{\text{L,res}}^{\text{A}}$  is the residual part of  $\Delta\mu_L^{\text{A}}$  composed of the pair entropy and many-body terms.  $\Delta\mu_{\text{L,res}}^{\text{A}}$  corresponds to the free-energy penalty due to structural changes of the solvents and protein caused by



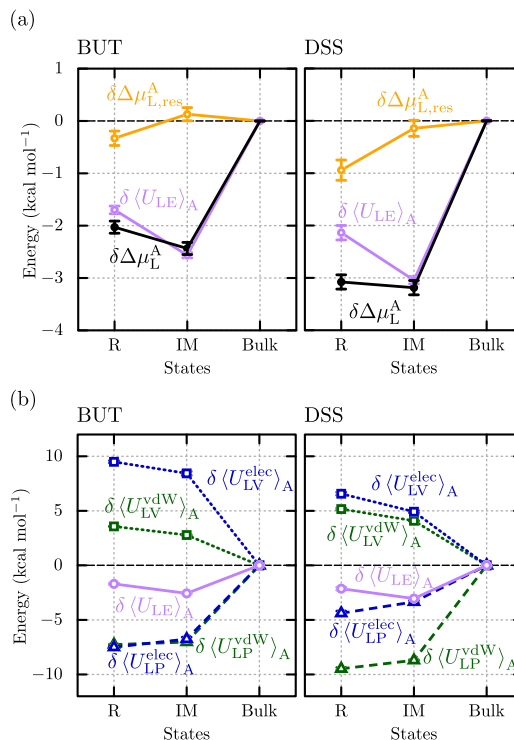
**FIG. 4.** Potentials of mean force (PMFs) on the energy coordinate,  $w(U_{\text{attr}})$ , for (a) BUT and (b) DSS obtained from the REUS simulations. The spatial densities of ligands [Eq. (41)] corresponding to state R are also shown in the right. The definitions of state R are  $-4.3 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -3.5$  for BUT and  $-2.8 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -2.0$  for DSS, respectively, and these regions are highlighted in the PMFs. The isovalues of spatial densities for transparent and solid surfaces are  $10^{-5}$  and  $10^{-3} \text{ \AA}^{-3}$ , respectively. The spatial densities were visualized with the Visual Molecular Dynamics (VMD) package.<sup>81</sup>  $w(U_{\text{attr}})$  is set to zero at the local minimum around  $-1.5 \text{ kcal mol}^{-1}$ .

the binding of the ligand. Note that the ER theory introduces the approximations to the many-body term. In the case of the bulk,  $U_{\text{LE}}$  is ligand–solvent (water and ions) interaction,  $U_{\text{LV}}$ . For states IM and R,  $U_{\text{LE}}$  is the sum of ligand–solvent interaction (LV) ( $U_{\text{LV}}$ ) and ligand–protein (LP) interaction [ $U_{\text{LP}}$ , equivalent to  $U_{\text{PL}}$  introduced in Eq. (19)]. Further decomposition of  $U_{\text{LE}}$  into the interaction energy components of LV and LP interactions can be achieved as

$$U_{\text{LE}} = U_{\text{LV}}^{\text{vdW}} + U_{\text{LV}}^{\text{elec}} + U_{\text{LP}}^{\text{vdW}} + U_{\text{LP}}^{\text{elec}}, \quad (51)$$

where  $U_{\text{LV}}^{\text{vdW}}$  and  $U_{\text{LV}}^{\text{elec}}$  ( $X = \text{V}$  and  $\text{P}$ ) are, respectively, the van der Waals and electrostatic interaction energies in LX interaction.  $U_{\text{LP}}^{\text{vdW}}$  and  $U_{\text{LP}}^{\text{elec}}$  are zero for the bulk state. The values of  $\Delta\mu_{\text{L, res}}^{\text{A}}$  are evaluated directly through the ER theory for the bulk and state IM. As for state R,  $\Delta\mu_{\text{L, res}}^{\text{R}}$  is computed with Eq. (43), and then, we obtain  $\Delta\mu_{\text{L, res}}^{\text{R}}$  by subtracting  $\langle U_{\text{LE}} \rangle_{\text{R}}$  from  $\Delta\mu_{\text{L}}^{\text{R}}$ .

Figure 5(a) shows the changes in  $\Delta\mu_{\text{L}}^{\text{A}}$ ,  $\langle U_{\text{LE}} \rangle_{\text{A}}$ , and  $\Delta\mu_{\text{L, res}}^{\text{A}}$  ( $\text{A} = \text{bulk, IM, or R}$ ) from the bulk state. In the figure,  $\delta X$  ( $X = \Delta\mu_{\text{L}}^{\text{A}}$ ,  $\langle U_{\text{LE}} \rangle_{\text{A}}$ , and  $\Delta\mu_{\text{L, res}}^{\text{A}}$ ) is defined as the difference between the values of  $X$  for state A and for the bulk state. The values of  $\Delta\mu_{\text{L}}^{\text{A}}$



**FIG. 5.** Changes in the solvation free energies and their decomposition from the bulk state. (a)  $\delta\Delta\mu_{\text{L}}^{\text{A}}$ ,  $\delta\langle U_{\text{LE}} \rangle_{\text{A}}$ , and  $\delta\Delta\mu_{\text{L, res}}^{\text{A}}$ . (b) Decomposition of  $\delta\langle U_{\text{LE}} \rangle_{\text{A}}$  into the interaction energy components of ligand–solvent (LV) and ligand–protein (LP) interactions. The errors are provided at the standard error.

and their decomposition are summarized in Table I.  $\delta\Delta\mu_{\text{L}}^{\text{A}}$  for BUT decreases by  $2.4 \text{ kcal mol}^{-1}$  from the bulk to state IM, corresponding to the stabilization of BUT. At state R, the value of  $\delta\Delta\mu_{\text{L}}^{\text{A}}$  is slightly increased. It is seen that the profile of  $\delta\langle U_{\text{LE}} \rangle_{\text{A}}$  is similar to that of  $\delta\Delta\mu_{\text{L}}^{\text{A}}$ , and the absolute value of  $\delta\Delta\mu_{\text{L, res}}^{\text{A}}$  is small. Hence,  $\langle U_{\text{LE}} \rangle_{\text{A}}$  is responsible for the stabilization of BUT in the vicinity of the protein. State IM for DSS has a negatively large value of  $\delta\Delta\mu_{\text{L}}^{\text{A}}$ ,  $-3.2 \text{ kcal mol}^{-1}$ , because of  $\delta\langle U_{\text{LE}} \rangle_{\text{A}}$ . Unlike in the case of BUT,  $\delta\Delta\mu_{\text{L}}^{\text{A}}$  hardly changes from state IM to R in spite of the increase in  $\delta\langle U_{\text{LE}} \rangle_{\text{A}}$ . This is because the decrease in  $\delta\Delta\mu_{\text{L, res}}^{\text{A}}$  is observed from state IM to R, suggesting the importance of the entropy and many-body contribution to state R.

The decomposition of  $\delta\langle U_{\text{LE}} \rangle_{\text{A}}$  into the interaction energy components based on Eq. (51) is shown in Fig. 5(b). For both the ligands, the LV and LP interactions have the positive and negative contributions to  $\delta\langle U_{\text{LE}} \rangle_{\text{A}}$ , respectively. It is found that the desolvation penalty for BUT stems mainly from the electrostatic interaction,  $\delta\langle U_{\text{LV}}^{\text{elec}} \rangle_{\text{A}}$ . Since BUT has two hydrophilic groups, hydroxyl and carbonyl groups [Fig. 3(b)], the breaking of hydrogen bonding should bring the penalty. The van der Waals and electrostatic components in the LP interaction,  $\delta\langle U_{\text{LP}}^{\text{vdW}} \rangle_{\text{A}}$  and  $\delta\langle U_{\text{LP}}^{\text{elec}} \rangle_{\text{A}}$ , contribute almost equally to the stabilization, making  $\delta\langle U_{\text{LE}} \rangle_{\text{A}}$  negative at states IM and R. As for DSS, since a hydroxyl group is absent in the structure [Fig. 3(c)], the desolvation penalty from  $\delta\langle U_{\text{LE}} \rangle_{\text{A}}$  at states IM and R is smaller than that for BUT, while that from  $\delta\langle U_{\text{LV}}^{\text{vdW}} \rangle_{\text{A}}$  is increased. On the other hand, the stabilization effect originating

**TABLE I.** Solvation free energies for different states and their decomposition based on Eq. (50). All values are in kcal mol<sup>-1</sup>. The energy ranges of state R for BUT and DSS are, respectively, defined as  $-4.3 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -3.5$  and  $-2.8 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -2.0$ . As for state IM, the energy range is defined as  $-2.0 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -1.0$  for both the ligands. The errors are provided at the standard error.

	State A	$\Delta\mu_L^A$	$\langle U_{LE} \rangle_A$	$\Delta\mu_{L,\text{res}}^A$	$\langle U_{LV} \rangle$	$\langle U_{LP} \rangle$
BUT	Bulk	$-8.69 \pm 0.05$	$-34.32 \pm 0.02$	$25.6 \pm 0.05$	$-34.32 \pm 0.02$	...
	IM	$-11.1 \pm 0.1$	$-36.88 \pm 0.04$	$25.8 \pm 0.1$	$-23.1 \pm 0.2$	$-13.8 \pm 0.2$
	R	$-10.7 \pm 0.1$	$-36.02 \pm 0.07$	$25.3 \pm 0.1$	$-21.3 \pm 0.1$	$-14.8 \pm 0.1$
DSS	Bulk	$-7.8 \pm 0.1$	$-29.94 \pm 0.03$	$22.2 \pm 0.1$	$-29.94 \pm 0.03$	...
	IM	$-10.96 \pm 0.09$	$-32.99 \pm 0.06$	$22.0 \pm 0.1$	$-20.9 \pm 0.1$	$-12.1 \pm 0.1$
	R	$-10.85 \pm 0.09$	$-32.1 \pm 0.1$	$21.2 \pm 0.2$	$-18.2 \pm 0.2$	$-13.9 \pm 0.2$

from  $\delta(U_{LP}^{\text{vdW}})$  at states IM and R is enhanced as compared to BUT. It indicates the importance of the hydrophobic nature of DSS for the thermodynamic stability of states IM and R. The geometries of BUT and DSS [Figs. 3(b) and 3(c)] are similar to each other, while the hydrophobicity of DSS is higher. Hence, the difference of the profiles between BUT and DSS shown in Fig. 5 may explain the general effect of the hydrophobicity of ligands on the formation of state R for hydrophobic binding pockets.

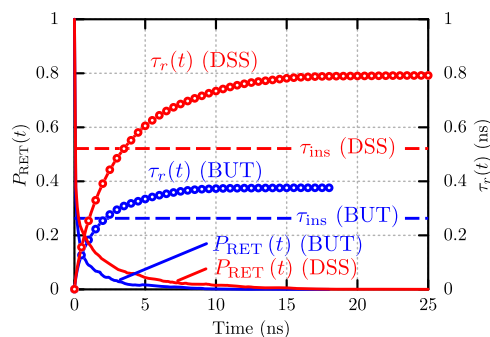
In order to evaluate the equilibrium constant between the dissociate state and state R through Eq. (22),  $K^*$ , we also need to compute the correction term,  $\Delta G_{\text{corr}}^{\circ}$  [Eq. (24)], by means of the test-particle insertion. For both the systems, we confirm that the convergence of  $\Delta G_{\text{corr}}^{\circ}$  is fast with the number of frames used for random insertion (Fig. S3 of the supplementary material), and thus, the statistically reliable estimate of  $\Delta G_{\text{corr}}^{\circ}$  is possible with the test-particle insertion.

### C. Kinetics of returning and insertion processes

The returning probabilities,  $P_{\text{RET}}(t)$ , are plotted in Fig. 6 together with the running time integrals defined as

$$\tau_r(t) = \int_0^t d\tau P_{\text{RET}}(\tau). \quad (52)$$

The energy ranges for state R are, respectively, defined as  $-4.3 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -3.5$  for BUT and  $-2.8 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -2.0$ . The time constants of insertion defined as

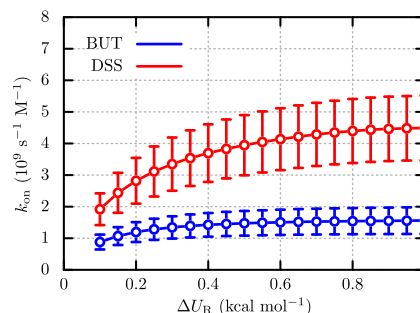


**FIG. 6.** Returning probability,  $P_{\text{RET}}(t)$ , and its running time integral,  $\tau_r(t)$ . The time constants of insertion defined as  $\tau_{\text{ins}} = 1/k_{\text{ins}}$  are also shown for comparison.

$\tau_{\text{ins}} = 1/k_{\text{ins}}$  (Fig. 1) are also shown for comparison. Since the HFB potential [Eq. (48)] is imposed to prevent the ligands from inserting into the binding pocket during the MD simulations for  $P_{\text{RET}}(t)$ , the lower bound for state R is not considered in the computation of  $P_{\text{RET}}(t)$ . Note that the inverse of the integration up to  $t \rightarrow \infty$  gives the dissociation rate constant from state R,  $k_r$  [Eq. (7)].  $\tau_r(t)$  converge at  $t \sim 10$  ns for BUT and  $t \sim 20$  ns for DSS, and the converged values are  $0.38 \pm 0.09$  and  $0.79 \pm 0.07$  ns for BUT and DSS, respectively. Hence, the dissociation kinetics of DSS is around twice slower than that of BUT. As shown in the PMF (Fig. 4), the free energy barrier around state R for DSS is lower than that for BUT. Furthermore, DSS has a barrier between state IM [ $-2.0 \leq U_{\text{attr}}/(\text{kcal mol}^{-1}) \leq -1.0$ ] and dissociate state ( $U_{\text{attr}} \sim 0$ ), while such a barrier does not exist for BUT. Thus, DSS tends to return from state IM to state R more often than BUT, resulting in a slow-down of the dissociation kinetics.

For a hard-sphere system<sup>82</sup> and host-guest systems composed of  $\beta$ -cyclodextrin and small compounds,<sup>48</sup>  $P_{\text{RET}}(t)$  has an asymptotic decay as  $t^{-3/2}$ . Even for the present protein-ligand binding systems,  $P_{\text{RET}}(t)$  shows the same asymptotic behavior (see Fig. S1 of the supplementary material). Hence, this asymptoticity may hold for different types of binding systems.

The time constants of insertion,  $\tau_{\text{ins}}$ , are  $0.26 \pm 0.03$  ns for BUT and  $0.52 \pm 0.03$  ns for DSS, respectively. For both the ligands,  $\tau_{\text{ins}}$  are smaller than  $\tau_r(\infty)$ . Thus, the ligands in state R prefer to move from state R to the bound state than to the dissociate state.  $\tau_{\text{ins}}$  of DSS is larger than that of BUT. The difference of  $\tau_{\text{ins}}$  could be interpreted



**FIG. 7.** Dependency of binding rate constants,  $k_{\text{on}}$ , on the choice of the energy range of state R,  $U_{\text{low}} \leq U_{\text{attr}} \leq U_{\text{low}} + \Delta U_R$ . The values of  $U_{\text{low}}$  for BUT and DSS are fixed to  $-4.3$  and  $-2.8$  kcal mol<sup>-1</sup>, respectively.

**TABLE II.** Thermodynamic and kinetic quantities associated with the protein–ligand binding processes. The values of  $k_{\text{on}}$  obtained through the MD simulations are taken from Ref. 9. The errors are provided at the standard error.

	$k_{\text{on}} (10^9 \text{ s}^{-1} \text{ M}^{-1})$		$\chi (10^9 \text{ s}^{-1})$	$K^* (\text{M}^{-1})$	$k_{\text{ins}} (10^9 \text{ s}^{-1})$	$k_r (10^9 \text{ s}^{-1})$
	This study	MD				
BUT	$1.5 \pm 0.4$	$1.23 \pm 0.03$	$1.6 \pm 0.3$	$1.0 \pm 0.2$	$3.8 \pm 0.4$	$2.7 \pm 0.6$
DSS	$4 \pm 1$	$2.1 \pm 0.2$	$0.76 \pm 0.07$	$6 \pm 1$	$1.3 \pm 0.1$	$1.9 \pm 0.1$

with the profiles of the PMFs between state R and bound state (Fig. S2 of the supplementary material). The PMF of DSS has a local minimum at  $U_{\text{attr}} \sim -5 \text{ kcal mol}^{-1}$ , which is close to state R, while such a minimum is not present in the case of BUT. Thus, DSS coming from state R might be trapped at this minimum, causing the larger value of  $\tau_{\text{ins}}$  compared with BUT.

#### D. Binding rate constant $k_{\text{on}}$

In this subsection, we address the dependency of binding rate constants,  $k_{\text{on}}$ , on the choice of the energy range of state R. We represent the energy range for state R as

$$U_{\text{low}} \leq U_{\text{attr}} \leq U_{\text{low}} + \Delta U_{\text{R}}. \quad (53)$$

The values of  $U_{\text{low}}$  for BUT and DSS are, respectively, fixed to  $-4.3$  and  $-2.8 \text{ kcal mol}^{-1}$  that are the same as the peak positions in the PMFs (Fig. 4). Figure 7 shows the values of  $k_{\text{on}}$  as a function of  $\Delta U_{\text{R}}$ . When  $\Delta U_{\text{R}}$  is small,  $k_{\text{on}}$  for both the ligands are found to be dependent on  $\Delta U_{\text{R}}$  especially for DSS. On the other hand,  $k_{\text{on}}$  converge to certain values for both the cases with the increase in  $\Delta U_{\text{R}}$ . This monotonic behavior enables us to determine the energy range suitable for state R by gradually changing the upper bound. For both the ligands, we set the value of  $\Delta U_{\text{R}}$  to  $0.8 \text{ kcal mol}^{-1}$ , i.e., the energy ranges for BUT and DSS are set to  $-4.3 \leq U_{\text{attr}} / (\text{kcal mol}^{-1}) \leq -3.5$  and  $-2.8 \leq U_{\text{attr}} / (\text{kcal mol}^{-1}) \leq -2.0$ , respectively. The values of  $k_{\text{on}}$  under the above condition are  $(1.5 \pm 0.4) \times 10^9$  and  $(4 \pm 1) \times 10^9 \text{ s}^{-1} \text{ M}^{-1}$  for BUT and DSS, respectively. The observed trend that  $k_{\text{on}}$  of BUT is larger than that of DSS is consistent with the previous long timescale MD simulations<sup>9</sup> [ $(1.23 \pm 0.08) \times 10^9$  and  $(2.1 \pm 0.2) \times 10^9 \text{ s}^{-1} \text{ M}^{-1}$  for BUT and DSS, respectively] although  $k_{\text{on}}$  of DSS predicted from the present method is around twice larger.

We analyze the binding kinetics using the theoretical expression of  $k_{\text{on}}$  provided by the RP theory [Eq. (4)]. Let us define

$$\chi = (k_{\text{ins}}^{-1} + k_r^{-1})^{-1} = (\tau_{\text{ins}} + \tau_r(\infty))^{-1}, \quad (54)$$

which means the frequency at which either the insertion or dissociation occurs. From Eqs. (7) and (54), Eq. (4) is rewritten as

$$k_{\text{on}} = \chi K^*. \quad (55)$$

Since  $K^*$  reflects the thermodynamic stability of state R, Eq. (55) is a decomposition of  $k_{\text{on}}$  into the thermodynamic ( $K^*$ ) and kinetic ( $\chi$ ) contributions. We summarize the thermodynamic and kinetic quantities associated with the protein–ligand binding processes in Table II. The value of  $\chi$  is  $0.76 \pm 0.07 \text{ ns}$  for DSS, which is smaller than for BUT ( $1.6 \pm 0.3 \text{ ns}$ ). It indicates the slower kinetics of

DSS around state R as compared to BUT. On the other hand,  $K^*$  for DSS is  $6 \pm 1 \text{ M}^{-1}$ , which is  $\sim 6$  times larger than that for BUT ( $1.0 \pm 0.2 \text{ M}^{-1}$ ). Thus, the high stability of state R for DSS is responsible for the larger value of  $k_{\text{on}}$ .

#### V. CONCLUSION

In this study, we proposed a new method to quantify the binding rate constants of protein–ligand binding,  $k_{\text{on}}$ , by means of molecular dynamics (MD) simulations. The method is based on returning probability (RP) and energy representation (ER) theories of solution. The RP theory provides not only a tractable expression of the binding rate constant but also enables us to systematically analyze the binding processes in terms of the thermodynamics and kinetics on the reactive state existing in the binding processes. By means of the ER theory, the reliable estimate of the solvation free energy of a solute is realized in complex solution systems. We constructed a scheme of computing the free energy difference between the reactive and dissociate state, required in the RP theory, based on the ER theory. Note that this scheme is applicable to arbitrary types of reactive states on reaction coordinates. Thus, the incorporation of the ER theory expanded the versatility of the RP theory.

We applied the present method to the protein–ligand binding systems that consist of FK506 binding protein (FKBP) and small fragments [4-hydroxy-2-butanone (BUT) and methyl methylthiomethyl sulphoxide (DSS)] in a 150 mM NaCl aqueous solution. The reactive and intermediate states were characterized with the potentials of mean force (PMFs) on the attractive part of Lennard-Jones interaction between the protein and ligand,  $U_{\text{attr}}$ . From the analysis of the interaction energy components, we quantified the stabilizing and destabilizing effects on the reactive state coming from the ligand–solvent and ligand–protein interactions, respectively. The computed values of  $k_{\text{on}}$  were found to be hardly dependent on the choice of the reactive state when the energy range of  $U_{\text{attr}}$  is sufficiently wide. The present method reproduced the trends reported in the previous study using the long-timescale MD simulations that the value of  $k_{\text{on}}$  for DSS is larger than that for BUT.<sup>9</sup> Furthermore, the systematic analysis based on the RP theory clarified that the higher thermodynamic stability of the reactive state for DSS causes the faster binding kinetics compared with BUT.

Since both the RP and ER theories are applicable to the heterogeneous systems, the present method could be utilized to elucidate the binding kinetics in complex solution systems from an atomistic point of view. It is known that macromolecular crowded environments significantly affect the binding efficiency.<sup>83</sup> The application of the present method to the binding phenomena occurring in such environments could provide a physicochemical insight into the crowder effects.

In the present method, a challenge lies in the determination of state R. Currently, we have to check the validity of the definition of state R by systematically changing the parameter associated with the region of state R on the reaction coordinate. This is due to the lack of the scheme to theoretically determine state R. As mentioned by Kim and Lee,<sup>50</sup> the Markovianity on state R is important for quantitatively evaluating the rate constant. Recently, a methodology of constructing optimal Markovian models from time-series data has been proposed based on the Koopman operator theory.<sup>84</sup> Thus, incorporating this method might be useful for overcoming the above challenge. We expect that the present method and its extension will allow us to elucidate a variety of binding systems, including permeation through membranes.

## SUPPLEMENTARY MATERIAL

The supplementary material contains the simulation protocols, the scheme of conditional test-particle insertion, the setups of test-particle insertion, the asymptotic behavior of returning probability, the potentials of mean force between the bound and reactive states, and the convergence of  $\Delta G_{\text{corr}}^{\circ}$ .

## ACKNOWLEDGMENTS

This work is supported by the Grants-in-Aid for Scientific Research (Grant Nos. JP21K14589, JP22J21080, JP21H05249, and JP23H01924) from the Japan Society for the Promotion of Science and by the Fugaku Supercomputer Project (Grant Nos. JPMXP1020230325 and JPMXP1020230327) and the Data-Driven Material Research Project (Grant No. JPMXP1122714694) from the Ministry of Education, Culture, Sports, Science and Technology. The simulations were conducted using TSUBAME3.0 at Tokyo Institute of Technology and Fugaku at RIKEN Advanced Institute for Computational Science through the HPCC System Research Project (Project Nos. hp220254, hp230101, hp230205, hp230212, and hp230158).

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

## Author Contributions

**Kento Kasahara:** Conceptualization (lead); Data curation (lead); Formal analysis (lead); Funding acquisition (lead); Investigation (lead); Methodology (lead); Project administration (lead); Resources (lead); Software (lead); Supervision (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). **Ren Masayama:** Investigation (supporting); Software (supporting). **Kazuya Okita:** Investigation (supporting); Methodology (supporting); Writing – review & editing (supporting). **Nobuyuki Matubayasi:** Conceptualization (supporting); Funding acquisition (supporting); Methodology (lead); Resources (supporting); Software (supporting); Writing – original draft (supporting); Writing – review & editing (equal).

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## APPENDIX A: THEORETICAL EXPRESSION OF CHEMICAL POTENTIAL UNDER NVT CONDITION

In this appendix, we derive a theoretical expression of the chemical potential under the NVT condition. The derivation is almost parallel to that under the NPT condition described in Ref. 56. Let us consider a system consisting of the solute species S and solvents V. The number of molecules for species S and solvents is  $N_S$  and  $N_V$ , respectively. We define the full coordinates of the  $i$ th molecules of S as  $\mathbf{x}_{S,i}$  and the set of  $\mathbf{x}_{S,i}$  as  $\mathbf{X}_S^{(N_S)}$ . The set of the full coordinates of the solvents is also defined as  $\mathbf{X}_V$ . Then, we introduce the total potential of the system as

$$\mathcal{V}_S(\mathbf{X}_S^{(N_S)}, \mathbf{X}_V; N_S) = \sum_{i=1}^{N_S} U_S(\mathbf{x}_{S,i}) + \sum_{i=1}^{N_S} \sum_{j>i}^{N_S} U_{SS}(\mathbf{x}_{S,i}, \mathbf{x}_{S,j}) + \sum_{i=1}^{N_S} U_{SV}(\mathbf{x}_{S,i}, \mathbf{X}_V) + U_V(\mathbf{X}_V), \quad (\text{A1})$$

where  $U_S(\mathbf{x}_{S,i})$  is the intramolecular energy of a solute molecule,  $U_{SS}(\mathbf{x}_{S,i}, \mathbf{x}_{S,j})$  is the solute–solute pair interaction energy,  $U_{SV}(\mathbf{x}_{S,i}, \mathbf{X}_V)$  is the interaction energy of a solute with the solvents, and  $U_V(\mathbf{X}_V)$  is the total potential of the solvents. The partition function of the system after performing the integration of the kinetic energy,  $Q(N_S)$ , is defined as

$$Q(N_S) = \frac{1}{N_S! N_V! \lambda_S^{N_S} \lambda_V^{N_V}} \int d\mathbf{X}_S^{(N_S)} \times \int d\mathbf{X}_V \exp[-\beta \mathcal{V}_S(\mathbf{X}_S^{(N_S)}, \mathbf{X}_V; N_S)], \quad (\text{A2})$$

where  $\beta$  is the inverse temperature and  $\lambda_S$  and  $\lambda_V$  are the kinetic contributions that come from the integration of the partition function about the kinetic energy of species S and solvent species, respectively. The dimensions of  $\lambda_S^{N_S}$  and  $\lambda_V^{N_V}$  are the same as those of  $\mathbf{X}_S^{(N_S)}$  and  $\mathbf{X}_V$ , respectively, and thus,  $Q(N_S)$  is dimensionless. If both the solute and solvent species are fully flexible,  $\lambda_l$  ( $l = S$  or  $V$ ) can be represented as the product of the thermal de Broglie wavelengths,

$$\lambda_l = \prod_{k=1}^{n_l} \left( \frac{\beta h^2}{2\pi m_{l,k}} \right)^{3/2}. \quad (\text{A3})$$

Here,  $h$  is Planck constant and  $m_{l,k}$  is the mass of the  $k$ th atom in species  $l$ . The quantum correction is usually multiplied to Eq. (A3), in fact, and when some bond lengths are fixed, the correction is further necessary to Eq. (A3) and the integration variables  $\mathbf{X}_S^{(N_S)}$  and  $\mathbf{X}_V$  in Eq. (A2). In any case, Eq. (A2) becomes dimensionless due to the contributions from  $\lambda_S$  and  $\lambda_V$ . The chemical potential of species S is defined as

$$\mu_S = -\frac{1}{\beta} \log \frac{Q(N_S + 1)}{Q(N_S)}. \quad (\text{A4})$$

Substituting Eq. (A2) into Eq. (A4) yields

$$\mu_S = -\frac{1}{\beta} \log \left( \frac{1}{\lambda_S(N_S + 1)} \frac{\int d\mathbf{X}_S^{(N_S+1)} \int d\mathbf{X}_V \exp[-\beta \mathcal{V}_S(\mathbf{X}_S^{(N_S+1)}, \mathbf{X}_V; N_S + 1)]}{\int d\mathbf{X}_S^{(N_S)} \int d\mathbf{X}_V \exp[-\beta \mathcal{V}_S(\mathbf{X}_S^{(N_S)}, \mathbf{X}_V; N_S)]} \right). \quad (\text{A5})$$

By introducing the solvation free energy of species S,

$$\Delta\mu_S = -\frac{1}{\beta} \log \left( \frac{\int d\mathbf{X}_S^{(N_S+1)} \int d\mathbf{X}_V \exp[-\beta \mathcal{V}_S(\mathbf{X}_S^{(N_S+1)}, \mathbf{X}_V; N_S + 1)]}{\int d\mathbf{X}_S^{(N_S+1)} \int d\mathbf{X}_V \exp[-\beta (U_S(\mathbf{x}_{S,N_S+1}) + \mathcal{V}_S(\mathbf{X}_S^{(N_S)}, \mathbf{X}_V; N_S))]} \right). \quad (\text{A6})$$

Equation (A5) can be rewritten as

$$\mu_S = -\frac{1}{\beta} \log \frac{Z_S}{\lambda_S N_S} + \Delta\mu_S, \quad (\text{A7})$$

where  $N_S + 1 \approx N_S$  and  $Z_S$  is the configurational integral of species S in an isolated state defined as

$$Z_S = \int d\mathbf{x}_S \exp[-\beta U_S(\mathbf{x}_S)]. \quad (\text{A8})$$

In addition, by defining the concentration of species S as  $[S] = N_S/V$ , Eq. (A7) can be rewritten as

$$\mu_S = -\frac{1}{\beta} \log \frac{Z_S}{\lambda_S V [S]} + \Delta\mu_S. \quad (\text{A9})$$

## APPENDIX B: INDEPENDENCE OF $\Delta G_{\text{corr}}^\circ$ FROM THE SYSTEM SIZE

In this appendix, we prove the independence of  $\Delta G_{\text{corr}}^\circ$  [Eq. (24)] from the system size,

$$e^{-\beta \Delta G_{\text{corr}}^\circ} = c^\circ V \frac{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V \theta_R(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta \mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}}{\int d\mathbf{x}_P \int d\mathbf{x}_L \int d\mathbf{X}_V e^{-\beta \mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}}. \quad (\text{B1})$$

Note that  $\mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)$  is defined in Eq. (26) and the intermolecular interactions are absent for the ligand with the protein and solvent. We first define the integration of  $e^{-\beta \mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}$  over the solvent coordinates as  $e^{-\beta X(\mathbf{x}_P, \mathbf{x}_L)}$ ,

$$e^{-\beta X(\mathbf{x}_P, \mathbf{x}_L)} = \int d\mathbf{X}_V e^{-\beta \mathcal{V}_R^{\text{ref}}(\mathbf{x}_P, \mathbf{x}_L, \mathbf{X}_V)}. \quad (\text{B2})$$

Then, Eq. (B1) is rewritten as

$$e^{-\beta \Delta G_{\text{corr}}^\circ} = c^\circ V \frac{\int d\mathbf{x}_P \int d\mathbf{x}_L \theta_R(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta X(\mathbf{x}_P, \mathbf{x}_L)}}{\int d\mathbf{x}_P \int d\mathbf{x}_L e^{-\beta X(\mathbf{x}_P, \mathbf{x}_L)}}. \quad (\text{B3})$$

Since  $\mathbf{x}_P$  and  $\mathbf{x}_L$  are decoupled with each other in  $X(\mathbf{x}_P, \mathbf{x}_L)$ , the integration over the center of masses (CoMs) of the protein and ligand can be performed in the denominator of Eq. (B3) as

$$\int d\mathbf{x}_P \int d\mathbf{x}_L e^{-\beta X(\mathbf{x}_P, \mathbf{x}_L)} = V^2 \int^* d\mathbf{x}_P \int^* d\mathbf{x}_L e^{-\beta X(\mathbf{x}_P, \mathbf{x}_L)}, \quad (\text{B4})$$

where  $\int^* d\mathbf{x}_P$  and  $\int^* d\mathbf{x}_L$  indicate the integration over the orientational and internal degrees of freedom of the protein and of the

ligand, respectively, with the CoM fixed. In the numerator,  $\mathbf{x}_P$  and  $\mathbf{x}_L$  are coupled due to the presence of  $\theta_R(\mathbf{x}_P, \mathbf{x}_L)$ . On the other hand, by defining the relative coordinate of the ligand with respect to the protein as  $\mathbf{x}_{\text{PL}}$ , the integration over the CoM of the protein can be performed as

$$\begin{aligned} \int d\mathbf{x}_P \int d\mathbf{x}_L \theta_R(\mathbf{x}_P, \mathbf{x}_L) e^{-\beta X(\mathbf{x}_P, \mathbf{x}_L)} \\ = V \int^* d\mathbf{x}_P \int d\mathbf{x}_{\text{PL}} \theta_R(\mathbf{x}_P, \mathbf{x}_{\text{PL}}) e^{-\beta X(\mathbf{x}_P, \mathbf{x}_{\text{PL}})}. \end{aligned} \quad (\text{B5})$$

Substitution of Eqs. (B4) and (B5) into Eq. (B3) yields

$$e^{-\beta \Delta G_{\text{corr}}^\circ} = c^\circ \frac{\int^* d\mathbf{x}_P \int d\mathbf{x}_{\text{PL}} \theta_R(\mathbf{x}_P, \mathbf{x}_{\text{PL}}) e^{-\beta X(\mathbf{x}_P, \mathbf{x}_{\text{PL}})}}{\int^* d\mathbf{x}_P \int^* d\mathbf{x}_L e^{-\beta X(\mathbf{x}_P, \mathbf{x}_L)}}. \quad (\text{B6})$$

Evidently, the integrations labeled with  $\int^*$  do not depend on the system size (the system-size dependence from the solvent degrees of freedom is canceled between the denominator and numerator). The integrand in the numerator shows non-zero values only if the protein and ligand form the complex of state R, and hence, the integration over  $\mathbf{x}_{\text{PL}}$  is hardly dependent on the system size. As a result, it can be concluded that the value of  $\Delta G_{\text{corr}}^\circ$  is not dependent on the system size.

## APPENDIX C: SCHEME OF COMPUTING $\Delta G_{\text{corr}}^\circ$

The free energy correction,  $\Delta G_{\text{corr}}^\circ$  [Eq. (24)], can be computed by randomly inserting the ligand into the configurations of the “mixed solvent” consisting of the protein together with water and salts (if contained in the system). If we define the total number of the solvent configurations with the inserted ligand as  $N_{\text{tot}}$  and the number of the configurations in which the protein and inserted ligand form a complex of state R as  $N_{\Omega_R}$ , Eq. (24) can be described as

$$\Delta G_{\text{corr}}^\circ = -\frac{1}{\beta} \log \left( c^\circ V \frac{N_{\Omega_R}}{N_{\text{tot}}} \right). \quad (\text{C1})$$

Since the formation of the complex of state R is hardly observed when the system size is large and the ligand is inserted randomly into the system, the convergence of  $N_{\Omega_R}/N_{\text{tot}}$  becomes slow with the increase in the system volume. As discussed in Appendix B, on the other hand,  $\Delta G_{\text{corr}}^\circ$  does not depend on the system size, indicating that  $N_{\Omega_R}/N_{\text{tot}}$  is proportional to the system volume. Thus, if we consider the insertion of the ligand into a spatial region whose volume

is  $V'$  ( $V' < V$ ) and that contains state R, the following relationship holds:

$$\frac{N_{\Omega_R}}{N_{\text{tot}}} = \frac{V'}{V} \left( \frac{N_{\Omega_R}}{N_{\text{tot}}} \right)_{V'} \quad (\text{C2})$$

Here,  $(N_{\Omega_R}/N_{\text{tot}})_{V'}$  means the population of the configurations corresponding to state R obtained from the insertions of ligand into the spatial region mentioned above. Substitution of Eq. (C2) into Eq. (C1) yields

$$\Delta G_{\text{corr}}^{\circ} = -\frac{1}{\beta} \log \left[ c^{\circ} V' \left( \frac{N_{\Omega_R}}{N_{\text{tot}}} \right)_{V'} \right] \quad (\text{C3})$$

Therefore, the efficient computation of  $\Delta G_{\text{corr}}^{\circ}$  can be performed by considering the molecular-size region  $V'$  for insertion in Eq. (C3).

## REFERENCES

- 1 K. M. Merz, Jr., D. Ringe, and C. H. Reynolds, *Drug Design: Structure-and Ligand-Based Approaches* (Cambridge University Press, 2010).
- 2 J.-P. Renaud and M.-A. Delsuc, *Curr. Opin. Pharmacol.* **9**, 622 (2009).
- 3 D. C. Swinney, *Nat. Rev. Drug Discovery* **3**, 801 (2004).
- 4 R. A. Copeland, D. L. Pompliano, and T. D. Meek, *Nat. Rev. Drug Discovery* **5**, 730 (2006).
- 5 R. A. Copeland, *Nat. Rev. Drug Discovery* **15**, 87 (2016).
- 6 D. A. Schuetz, W. E. A. de Witte, Y. C. Wong, B. Knasmueller, L. Richter, D. B. Kokh, S. K. Sadiq, R. Bosma, I. Nederpelt, L. H. Heitman, E. Segala, M. Amaral, D. Guo, D. Andres, V. Georgi, L. A. Stoddart, S. Hill, R. M. Cooke, C. De Graaf, R. Leurs, M. Frech, R. C. Wade, E. C. M. de Lange, A. P. IJzerman, A. Müller-Fahrnow, and G. F. Ecker, *Drug Discovery Today* **22**, 896 (2017).
- 7 M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli, *J. Med. Chem.* **59**, 4035 (2016).
- 8 Z. Tang and C.-e. A. Chang, *J. Chem. Theory Comput.* **14**, 303 (2017).
- 9 A. C. Pan, H. Xu, T. Palpant, and D. E. Shaw, *J. Chem. Theory Comput.* **13**, 3372 (2017).
- 10 R. W. Zwanzig, *J. Chem. Phys.* **22**, 1420 (1954).
- 11 J. G. Kirkwood, *J. Chem. Phys.* **3**, 300 (1935).
- 12 C. Chipot and A. Pohorille, *Free Energy Calculations* (Springer, 2007), Vol. 86.
- 13 M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, *Biophys. J.* **72**, 1047 (1997).
- 14 S. Boresch, F. Tettinger, M. Leitgeb, and M. Karplus, *J. Phys. Chem. B* **107**, 9535 (2003).
- 15 Y. Deng and B. Roux, *J. Phys. Chem. B* **113**, 2234 (2009).
- 16 H. Fujitani, Y. Tanida, and A. Matsuura, *Phys. Rev. E* **79**, 021914 (2009).
- 17 R. M. Levy, N. Matubayasi, and B. W. Zhang, *J. Phys. Chem. B* **124**, 11771 (2020).
- 18 N. Matubayasi and M. Nakahara, *J. Chem. Phys.* **113**, 6070 (2000).
- 19 N. Matubayasi and M. Nakahara, *J. Chem. Phys.* **117**, 3605 (2002).
- 20 S. Sakuraba and N. Matubayasi, *J. Comput. Chem.* **35**, 1592 (2014).
- 21 N. Matubayasi, *Bull. Chem. Soc. Jpn.* **92**, 1910 (2019).
- 22 N. Matubayasi, W. Shinoda, and M. Nakahara, *J. Chem. Phys.* **128**, 195107 (2008).
- 23 N. J. Bruce, G. K. Ganotra, D. B. Kokh, S. Kashan Sadiq, and R. C. Wade, *Curr. Opin. Struct. Biol.* **49**, 1 (2018).
- 24 J. Wang, H. N. Do, K. Koirala, and Y. Miao, *J. Chem. Theory Comput.* **19**, 2135 (2023).
- 25 F. Sohraby and A. Nunes-Alves, *Trends Biochem. Sci.* **48**, 437 (2022).
- 26 J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, *J. Chem. Phys.* **126**, 155101 (2007).
- 27 J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, *J. Chem. Phys.* **134**, 174105 (2011).
- 28 G. R. Bowman, V. S. Pande, and F. Noé, *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation* (Springer Science & Business Media, 2013), Vol. 797.
- 29 A. K. Faradjian and R. Elber, *J. Chem. Phys.* **120**, 10880 (2004).
- 30 E. Vanden-Eijnden, M. Venturoli, G. Ciccotti, and R. Elber, *J. Chem. Phys.* **129**, 174102 (2008).
- 31 E. Vanden-Eijnden and M. Venturoli, *J. Chem. Phys.* **130**, 194101 (2009).
- 32 G. A. Huber and S. Kim, *Biophys. J.* **70**, 97 (1996).
- 33 D. M. Zuckerman and L. T. Chong, *Annu. Rev. Biophys.* **46**, 43 (2017).
- 34 R. Harada and A. Kitao, *J. Chem. Phys.* **139**, 035103 (2013).
- 35 P. Tiwary and M. Parrinello, *Phys. Rev. Lett.* **111**, 230602 (2013).
- 36 W. Sinko, Y. Miao, C. A. F. de Oliveira, and J. A. McCammon, *J. Phys. Chem. B* **117**, 12759 (2013).
- 37 D. A. Schuetz, M. Bernetti, M. Bertazzo, D. Musil, H.-M. Eggenweiler, M. Recanatini, M. Masetti, G. F. Ecker, and A. Cavalli, *J. Chem. Inf. Model.* **59**, 535 (2018).
- 38 Y. Miao, V. A. Feher, and J. Andrew McCammon, *J. Chem. Theory Comput.* **11**, 3584 (2015).
- 39 Y. Miao, A. Bhattarai, and J. Wang, *J. Chem. Theory Comput.* **16**, 5526 (2020).
- 40 L. W. Votapka, B. R. Jagger, A. L. Heyneman, and R. E. Amaro, *J. Phys. Chem. B* **121**, 3597 (2017).
- 41 B. R. Jagger, A. A. Ojha, and R. E. Amaro, *J. Chem. Theory Comput.* **16**, 5348 (2020).
- 42 N. Donyapour, N. Roussey, and A. Dickson, *J. Chem. Phys.* **150**, 244112 (2019).
- 43 S. A. Rice, *Diffusion-Limited Reactions* (Elsevier, 1985), Vol. 25.
- 44 K. Lindenberg, R. Metzler, and G. Oshanin, *Chemical Kinetics: Beyond the Textbook* (World Scientific, 2019).
- 45 S. H. Northrup, S. A. Allison, and J. A. McCammon, *J. Chem. Phys.* **80**, 1517 (1984).
- 46 B. A. Luty, J. Andrew McCammon, and H.-X. Zhou, *J. Chem. Phys.* **97**, 5682 (1992).
- 47 S. Lee and M. Karplus, *J. Chem. Phys.* **86**, 1883 (1987).
- 48 K. Kasahara, R. Masayama, K. Okita, and N. Matubayasi, *J. Chem. Phys.* **155**, 204503 (2021).
- 49 K. Kasahara, R. Masayama, Y. Matsubara, and N. Matubayasi, *Chem. Lett.* **51**, 823 (2022).
- 50 J.-H. Kim and S. Lee, *J. Chem. Phys.* **131**, 014503 (2009).
- 51 J.-U. Lee, W.-J. Lee, H.-S. Park, and S. Lee, *Bull. Korean Chem. Soc.* **33**, 862 (2012).
- 52 S. Doudou, N. A. Burton, and R. H. Henchman, *J. Chem. Theory Comput.* **5**, 909 (2009).
- 53 J. M. Kolos, A. M. Voll, M. Bauder, and F. Hausch, *Front. Pharmacol.* **9**, 1425 (2018).
- 54 D. A. Holt, J. I. Luengo, D. S. Yamashita, H.-J. Oh, A. L. Konialian, H.-K. Yen, L. W. Rozamus, M. Brandt, M. J. Bossard, M. A. Levy, D. S. Eggleston, T. J. Stout, J. Liang, L. W. Schultz, and J. Clardy, *J. Am. Chem. Soc.* **115**, 9925 (2002).
- 55 K. Kasahara, S. Re, G. Nawrocki, H. Oshima, C. Mishima-Tsumagari, Y. Miyata-Yabuki, M. Kukimoto-Niino, I. Yu, M. Shirouzu, M. Feig, and Y. Sugita, *Nat. Commun.* **12**, 4099 (2021).
- 56 K. Yamada and N. Matubayasi, *Macromolecules* **53**, 775 (2020).
- 57 L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, *J. Comput. Chem.* **30**, 2157 (2009).
- 58 J. Jung, T. Mori, C. Kobayashi, Y. Matsunaga, T. Yoda, M. Feig, and Y. Sugita, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **5**, 310 (2015).
- 59 C. Kobayashi, J. Jung, Y. Matsunaga, T. Mori, T. Ando, K. Tamura, M. Kamiya, and Y. Sugita, *J. Comput. Chem.* **38**, 2193 (2017).
- 60 J. Jung, C. Kobayashi, K. Kasahara, C. Tan, A. Kuroda, K. Minami, S. Ishiduki, T. Nishiki, H. Inoue, Y. Ishikawa, M. Feig, and Y. Sugita, *J. Comput. Chem.* **42**, 231 (2021).
- 61 G. Bussi, D. Donadio, and M. Parrinello, *J. Chem. Phys.* **126**, 014101 (2007).
- 62 W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, *J. Chem. Phys.* **76**, 637 (1982).



- <sup>63</sup>M. Tuckerman, B. J. Berne, and G. J. Martyna, *J. Chem. Phys.* **97**, 1990 (1992).
- <sup>64</sup>U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).
- <sup>65</sup>J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).
- <sup>66</sup>H. C. Andersen, *J. Comput. Phys.* **52**, 24 (1983).
- <sup>67</sup>S. Miyamoto and P. A. Kollman, *J. Comput. Chem.* **13**, 952 (1992).
- <sup>68</sup>P. Burkhard, P. Taylor, and M. D. Walkinshaw, *J. Mol. Biol.* **295**, 953 (2000).
- <sup>69</sup>R. B. Best and G. Hummer, *J. Phys. Chem. B* **113**, 9004 (2009).
- <sup>70</sup>Parameter files for ff99SB\*-ILDN is provided by Robert Best (National Institutes of Health) at GitHub. [https://github.com/bestlab/force\\_fields](https://github.com/bestlab/force_fields).
- <sup>71</sup>J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, *J. Comput. Chem.* **25**, 1157 (2004).
- <sup>72</sup>J. Wang, W. Wang, P. A. Kollman, and D. A. Case, *J. Mol. Graphics Modell.* **25**, 247 (2006).
- <sup>73</sup>P. Cieplak, W. D. Cornell, C. Bayly, and P. A. Kollman, *J. Comput. Chem.* **16**, 1357 (1995).
- <sup>74</sup>D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, *J. Comput. Chem.* **26**, 1668 (2005).
- <sup>75</sup>M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji *et al.*, *Gaussian* **16**, 2016.
- <sup>76</sup>Y. Sugita, A. Kitao, and Y. Okamoto, *J. Chem. Phys.* **113**, 6042 (2000).
- <sup>77</sup>M. Bonomi and C. Camilloni, *Biomolecular Simulations* (Springer, 2019), Vol. 2022.
- <sup>78</sup>M. R. Shirts and J. D. Chodera, *J. Chem. Phys.* **129**, 124105 (2008).
- <sup>79</sup>Y. Matsunaga, M. Kamiya, H. Oshima, J. Jung, S. Ito, and Y. Sugita, *Biophys. Rev.* **14**, 1503 (2022).
- <sup>80</sup>B. Efron, *Bootstrap Methods: Another Look at the Jackknife* (Springer, 1992).
- <sup>81</sup>W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graphics* **14**, 33 (1996).
- <sup>82</sup>J. Lee, S. Yang, J. Kim, and S. Lee, *J. Chem. Phys.* **120**, 7564 (2004).
- <sup>83</sup>J. C. M. Uitdehaag, J. de Man, N. Willemsen-Seegers, M. B. W. Prinsen, M. A. A. Libouban, J. G. Sterrenburg, J. J. de Wit, J. R. F. de Vetter, J. A. D. M. de Roos, R. C. Buijsman, and G. J. R. Zaman, *J. Mol. Biol.* **429**, 2211 (2017).
- <sup>84</sup>H. Wu and F. Noé, *J. Nonlinear Sci.* **30**, 23 (2020).