Author(s)	Okita, Kazuya; Maruyama, Yusei; Kasahara, Kento et al.

Flexible framework of computing binding free energy using the energy representation theory of solution ⊘

Kazuya Okita (iD) ; Yusei Maruyama (iD) ; Kento Kasahara ✉ (iD) ; Nobuyuki Matubayasi (iD)

Check for updates

CHORUS

View Online   Export Citation

Flexible framework of computing binding free energy using the energy representation theory of solution

View Online   Export Citation   CrossMark

Kazuya Okita,   Yusei Maruyama,   Kento Kasahara,[a]   and Nobuyuki Matubayasi[b]

AFFILIATIONS

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

[a]Author to whom correspondence should be addressed: kasahara@cheng.es.osaka-u.ac.jp
[b]Electronic mail: nobuyuki@cheng.es.osaka-u.ac.jp

ABSTRACT

Host–guest binding plays a crucial role in the functionality of various systems, and its efficiency is often quantified using the binding free energy, which represents the free-energy difference between the bound and dissociated states. Here, we propose a methodology to compute the binding free energy based on the energy representation (ER) theory of solution, which enables us to evaluate the free-energy difference between the systems of interest with the molecular dynamics (MD) simulations. Unlike the other free-energy methods, such as the Bennett acceptance ratio (BAR), the ER theory does not require the MD simulations for hypothetical intermediate states connecting the systems of interest, leading to reduced computational costs. By constructing the thermodynamic cycle of the binding process that is suitable for the ER theory, a robust calculation of the binding free energy is realized. We apply the present method to the self-association of N-methylacetamide in different solvents and the binding of aspirin to β-cyclodextrin (CD) in water. In the former case, the present method estimates that the binding free energy decreases as the solvent polarity decreases. This trend is consistent with the experimental finding. For the latter system, the binding free energies for the two representative CD–aspirin bound complexes, primary (P) and secondary (S) complexes, are estimated to be -5.2 ± 0.1 and -5.03 ± 0.09 kcal mol^{-1}, respectively. These values are satisfactorily close to those from the BAR method [-4.2 ± 0.2 and -4.1 ± 0.2 kcal mol^{-1} for P and S, respectively]. Furthermore, the interaction-energy component analysis reveals that the van der Waals interaction between aspirin and CD dominantly contributes to the stabilization of the bound complexes, which is in harmony with the well-known binding mechanism in the CD systems.

I. INTRODUCTION

Host–guest binding has been recognized as one of the most fundamental processes in various fields of science. For instance, substrate binding to its target protein is a central issue in biology because most proteins exert their biological functions upon binding.[1] The binding process is also crucial for drug molecules, which regulate (promote or inhibit) cellular functions such as cell proliferation mediated by signal transduction.[2,3] The molecular dynamics (MD) simulation has played an important role in drug discovery and design[4,5] thanks to its capability of elucidating the binding mechanisms at the atomistic detail based on classical mechanics. For instance, an inhibitor of HIV integrase was successfully identified through the MD simulations combined with molecular docking

techniques.[6] Binding free energy, the free-energy difference between the bound and dissociated states, is regarded as a useful indicator for the efficiency of binding processes and has been extensively evaluated through the MD-based approaches.[7–10] Therefore, developing methodologies to efficiently compute the binding free energy while enabling systematic analysis would be beneficial for *in silico* screening of drug candidates.

The thermodynamic integration (TI),[11] free-energy perturbation (FEP),[12] and Bennett acceptance ratio (BAR)[13] methods offer a theoretical foundation for estimating the free-energy difference between the two states of interest (endpoints) in an exact manner using MD simulations. In these methods, the free-energy difference can be evaluated by considering a set of intermediate states that connect the endpoint states, which is often referred to as the

16 January 2025 01:52:31

alchemical pathway. The double-annihilation scheme (DAS)[14] and double-decoupling scheme (DDS)[15] are representative approaches for computing the binding free energy using the alchemical pathways. The DAS describes the binding process using the alchemical pathways associated with the gradual vanishing of the guest in both the dissociated and bound states. An effective setting of the intermediate states on the pathways is proposed for the DAS.[16,17] The alchemical pathways employed in the DDS are similar to those in the DAS, but the restraint is imposed on the guest in the bound state to keep it within the binding pocket of the host for all the intermediate states on the pathway.[18,19] Note that the effect of the restraints on the free energy can be removed analytically. The potential of mean force (PMF) approach coupled with the FEP realizes the free-energy calculation for large and flexible guest molecules.[20,21] The automation of the free-energy calculation based on the methodologies mentioned above has expanded the versatility.[22–27] However, these methods require conducting the MD simulations for all the intermediate states along the alchemical pathway. Therefore, the reduction in the computational cost is one of the important subjects.

The estimation of binding free energy can be made efficient by adopting approximate methods.[4,28,29] The method of linear interaction energy belongs to this category and performs an approximate evaluation of free energy from the energetics of the bound and dissociated states.[30] The effect of entropy was incorporated by mining local energy minima and quantifying the extents of local fluctuations.[31] The balance between the accuracy and speed of free-energy estimation is pursed when an approximate method is developed.

The classical density functional theory (DFT) of liquids allows for the analytical treatment of alchemical pathways in an approximate manner.[32] The energy representation (ER) theory of solution is a DFT theory that employs the solute–solvent pair interaction energy as a reaction coordinate for effectively describing the relative position and orientation of the solvent molecules around a solute on one-dimensional space.[33,34] This treatment enables us to construct an approximate free-energy functional represented in terms of only the information on the endpoint states obtained through the MD simulations, leading to the reduction in the computational cost as compared to the other alchemical free-energy methods. The ER theory was formulated to estimate the solvation free energy of a solute, which is a free-energy difference associated with the transfer of a solute from the gas phase to the solution phase. It has proven useful to analyze the solvation energetics for various systems, including lipid bilayers,[34] polymer solutions,[35,36] and crystal-surface systems.[37] Recently, the ER theory has been extended to compute the binding free energy for host–guest systems using an alchemical pathway similar to that of the DDS.[38] However, this method is applicable only to host molecules whose holo-form structures resemble their apo-form structures. Since a number of host molecules exhibiting the significant structural changes in the structure due to the binding have been reported,[39–41] further development could enhance the applicability and versatility of the ER theory.

Here, we present an ER-based methodology of computing the binding free energy applicable to the host–guest systems where the binding event induces a structural change in the host molecule. In this approach, the difference in the host structure between the holo- and apo-forms is characterized using the distribution on the host–guest interaction energy. The problematic energy domain in these distributions, which affects the free-energy calculation using the ER theory, is theoretically addressed by introducing a suitable additional state. This state was proposed in a previous study on the dissolution of water into polymer membranes.[35]

We apply the developed method to two systems: the self-association of N-methylacetamide (NMA) in different solvents and the binding of aspirin to β-cyclodextrin in water. In the first system, NMA molecules are known to weakly bind to each other,[42] allowing for the accurate evaluation of the binding free energy through the brute-force MD simulations. This makes the system suitable for verifying the accuracy of the present method. In the second system, β-cyclodextrin exhibits different structural populations between the apo- and holo-forms,[41] and thus, it can be used for testing the applicability of the present method. We also discuss the contributions of the interaction energies between the guest and surrounding environments to the binding thermodynamics, aiming to clarify the driving force of the binding processes.

II. THEORY

A. Theoretical expression of binding free energy

Here, we describe the theoretical expression of the binding free energy in terms of solvation free energy. The reaction scheme for host–guest binding is given by

$$H + G \rightleftharpoons B. \tag{1}$$

Here, H, G, and B signify host, guest, and bound complex, respectively. The equilibrium constant of the above reaction, K_a, is related to the binding free energy, ΔG°, as

$$\Delta G^\circ = -\frac{1}{\beta} \log c^\circ K_a = -\frac{1}{\beta} \log c^\circ \left(\frac{[B]}{[H][G]} \right), \tag{2}$$

where c° is the standard concentration ($c^\circ = 1$ M, typically) and $[H]$, $[G]$, and $[B]$ are the concentrations of H, G, and B, respectively. The equilibrium condition of Eq. (1) is expressed as

$$\mu_B - (\mu_H + \mu_G) = 0, \tag{3}$$

where μ_H, μ_G, and μ_B are the chemical potentials of H, G, and B, respectively. The chemical potential of species $S(S = H \text{ or } G)$ is given by[38]

$$\mu_S = \frac{1}{\beta} \log \left([S]\lambda_S\right) - \frac{1}{\beta} \log \frac{\int d\mathbf{x}_S \int d\mathbf{X}_V \; e^{-\beta(U_S + U_{SV} + U_V)}}{V \int d\mathbf{X}_V \; e^{-\beta U_V}}, \tag{4}$$

where β is the inverse temperature, V is the system volume, and λ_S is the kinetic contribution for species S obtained by the integration of the Maxwell–Boltzmann velocity distribution. $\mathbf{x}_S$ is the full-coordinate of species S, and $\mathbf{X}_V$ is the set of the full-coordinates of solvents. U_S, U_{SV}, and U_V are the intramolecular energy of S, the interaction energy between S and the solvents, and the total potential of the solvents, respectively. Regarding species B, the mathematical form of μ_B is similar to Eq. (4), but the configurational

integral over the full-coordinates of H and G, $\mathbf{x}_{HG} = \{\mathbf{x}_H, \mathbf{x}_G\}$, needs to be restricted to the region corresponding to the bound state. Let $\Theta_B(\mathbf{x}_{HB})$ be the characteristic function whose value is unity when the bound complex is formed and zero otherwise. Then, μ_B is expressed as

$$\mu_B = \frac{1}{\beta} \log\left([B]\lambda_H\lambda_G\right)$$
$$- \frac{1}{\beta} \log \frac{\int d\mathbf{x}_{HG} \int d\mathbf{X}_V \; \Theta_B(\mathbf{x}_{HG}) e^{-\beta(U_B + U_{BV} + U_V)}}{V \int d\mathbf{X}_V \; e^{-\beta U_V}}, \qquad (5)$$

where $d\mathbf{x}_{HG} = d\mathbf{x}_H d\mathbf{x}_G$. U_B is the potential of H and G, which is composed of the intramolecular energies of H, U_H, and G, U_G, and the interaction energy between H and G, U_{HG}, as

$$U_B = U_H + U_G + U_{HG}. \qquad (6)$$

U_{BV} is defined as the sum of the interaction energy between H and the solvents, U_{HV}, and that between G and the solvents, U_{GV},

$$U_{BV} = U_{HV} + U_{GV}. \qquad (7)$$

The solvation free energy of species G represents the change in free energy associated with the solvation process. This quantity is useful to derive the tractable expression of ΔG° from Eq. (2), as will be discussed later. As for the dissociated state, let us introduce the solution and reference systems, whose total potentials are, respectively, defined as

$$\mathcal{V}_{sol}^D = U_G + U_{GV} + U_V, \qquad (8)$$

$$\mathcal{V}_{ref}^D = U_G + U_V. \qquad (9)$$

The solvation free energy, $\Delta\mu_G^D$, can be described as

$$\Delta\mu_G^D = -\frac{1}{\beta} \log \frac{\int d\mathbf{x}_G \int d\mathbf{X}_V \; e^{-\beta \mathcal{V}_{sol}^D}}{\int d\mathbf{x}_G \int d\mathbf{X}_V \; e^{-\beta \mathcal{V}_{ref}^D}}. \qquad (10)$$

Since the interaction between G and the solvents, U_{GV}, is present in $\mathcal{V}_{sol}^D$ and absent in $\mathcal{V}_{ref}^D$, $\Delta\mu_G^D$ can be interpreted as the free-energy change resulting from the appearance of U_{GV} for the dissociated state. Similarly, we define the "solvation free energy" of G in the bound complex, $\Delta\mu_G^B$, as

$$\Delta\mu_G^B = -\frac{1}{\beta} \log \frac{\int d\mathbf{x}_{HG} \int d\mathbf{X}_V \; \Theta_B(\mathbf{x}_{HG}) e^{-\beta \mathcal{V}_{sol}^B}}{\int d\mathbf{x}_{HG} \int d\mathbf{X}_V \; \Theta_B(\mathbf{x}_{HG}) e^{-\beta \mathcal{V}_{ref}^B}}, \qquad (11)$$

where $\mathcal{V}_{sol}^B$ and $\mathcal{V}_{ref}^B$ are the potentials for the solution and reference systems corresponding to the bound state, respectively, defined as

$$\mathcal{V}_{sol}^B = U_G + U_H + U_{HG} + U_{GV} + U_{HV} + U_V, \qquad (12)$$

$$\mathcal{V}_{ref}^B = U_G + U_H + U_{HV} + U_V. \qquad (13)$$

In Eq. (12), all the interactions among G, H, and V are operative, and in Eq. (13), the interactions between G and H and between G and V are turned off. $\Delta\mu_G^B$ is thus the free-energy change for introducing

the interactions of G with H and V. It is called solvation free energy by viewing G as the solute and H and V as the solvent. The presence of Θ_B in both the numerator and denominator of Eq. (11) means that the solvation process of species G, which forms the bound complex in both the solution and reference systems, is represented by $\Delta\mu_G^B$.

Substituting Eqs. (4) and (5) into Eq. (2) yields

$$\Delta G^\circ = \Delta\mu_G^B - \Delta\mu_G^D + \Delta G_{corr}^\circ, \qquad (14)$$

where we have used Eqs. (10) and (11), and ΔG_{corr}° is the standard-state correction term, ensuring that the concentration of G in the dissociated state is c°, expressed as

$$\Delta G_{corr}^\circ = -\frac{1}{\beta} \log\left(c^\circ V \frac{\int d\mathbf{x}_{HG} \int d\mathbf{X}_V \; \Theta_B(\mathbf{x}_{HG}) e^{-\beta \mathcal{V}_{ref}^B}}{\int d\mathbf{x}_{HG} \int d\mathbf{X}_V \; e^{-\beta \mathcal{V}_{ref}^B}}\right). \qquad (15)$$

The configurations of G and those of H and the solvent molecules are independently generated by U_G and $U_H + U_{HV} + U_V$, respectively, in the reference system, and thus, the logarithm in Eq. (15) can be computed by the test-particle insertion of G into the configurations of H and the solvent molecules. Furthermore, ΔG_{corr}° is intensive and the spatial region for insertion can be made smaller than the simulation cell.[38]

To utilize Eq. (14), the definition of Θ_B is needed for $\Delta\mu_G^B$ [Eq. (14)] and ΔG_{corr}° [Eq. (15)]. The determination from the shape of the free-energy profile on certain reaction coordinates is a straightforward approach. If species G remains inside the binding site of species H during the simulations starting from the bound complex in the solution system, Θ_B can be set to accept all the sampled configurations. In this case, on the other hand, the explicit form of Θ_B is needed in the reference system to distinguish between the bound complex and others. The unique determination of Θ_B is generally impossible except for simple host and guest molecules, such as monoatomic molecules, and ΔG° appears to be dependent on the choice of Θ_B through the sampling in the reference system. Actually, it can be proved that ΔG° is not affected by the choice of Θ_B for the reference solvent as described below. Let us introduce the characteristic function that is different from Θ_B, Θ_B', and the following quantities:

$$\Delta\mu_G^{B'} = \Delta\mu_G^B - \frac{1}{\beta} \log \frac{\int d\mathbf{x}_{HG} \int d\mathbf{X}_V \; \Theta_B(\mathbf{x}_{HG}) e^{-\beta \mathcal{V}_{ref}^B}}{\int d\mathbf{x}_{HG} \int d\mathbf{X}_V \; \Theta_B'(\mathbf{x}_{HG}) e^{-\beta \mathcal{V}_{ref}^B}}$$
$$= -\frac{1}{\beta} \log \frac{\int d\mathbf{x}_{HG} \int d\mathbf{X}_V \; \Theta_B(\mathbf{x}_{HG}) e^{-\beta \mathcal{V}_{sol}^B}}{\int d\mathbf{x}_{HG} \int d\mathbf{X}_V \; \Theta_B'(\mathbf{x}_{HG}) e^{-\beta \mathcal{V}_{ref}^B}}, \qquad (16)$$

$$\Delta G_{corr}^{\circ'} = \Delta G_{corr}^\circ + \frac{1}{\beta} \log \frac{\int d\mathbf{x}_{HG} \int d\mathbf{X}_V \; \Theta_B(\mathbf{x}_{HG}) e^{-\beta \mathcal{V}_{ref}^B}}{\int d\mathbf{x}_{HG} \int d\mathbf{X}_V \; \Theta_B'(\mathbf{x}_{HG}) e^{-\beta \mathcal{V}_{ref}^B}}$$
$$= -\frac{1}{\beta} \log\left(c^\circ V \frac{\int d\mathbf{x}_{HG} \int d\mathbf{X}_V \; \Theta_B'(\mathbf{x}_{HG}) e^{-\beta \mathcal{V}_{ref}^B}}{\int d\mathbf{x}_{HG} \int d\mathbf{X}_V e^{-\beta \mathcal{V}_{ref}^B}}\right). \qquad (17)$$

By substituting Eqs. (11) and (17) into Eq. (14), one can rewrite Eq. (14) without any approximations as

$$\Delta G^\circ = \Delta\mu_G^{B'} - \Delta\mu_G^D + \Delta G_{corr}^{\circ'}, \qquad (18)$$

indicating that ΔG° does not depend on the choice of Θ_B' when all the terms in the above equation are computed in an exact manner.

B. Energy representation (ER) theory of solution

The energy representation (ER) theory offers an efficient method for computing solvation free energies using information about the endpoint states. In this approach, the full coordinates of the solvents are projected onto the solute–solvent pair interaction energy, and the free-energy functional is constructed based on the solvent distribution on the interaction energy, referred to as the energy distribution. In this subsection, it is our intent here to describe the ER theory only for $\Delta\mu_G^B$ [Eq. (11)], as the theoretical developments of the ER theory for $\Delta\mu_G^D$ [Eq. (10)] have been already reported elsewhere.[34,43,44]

Let $\hat{\rho}_\alpha(\varepsilon)$ denote the instantaneous distribution for the αth species, defined as follows:

$$\hat{\rho}_\alpha(\varepsilon) = \sum_{i\in\alpha}\delta(u_\alpha(\mathbf{x}_G,\mathbf{x}_{\alpha,i})-\varepsilon). \tag{19}$$

Here, u_α is the pair interaction-energy function between G and the αth species and $\mathbf{x}_{\alpha,i}$ is the full-coordinate of the ith molecule of the αth species. α refers to the solvent species (such as water) or H. By defining the ensemble average in the solution system conditioned by Θ_B and that in the reference system conditioned by Θ_B, respectively, as

$$\langle\cdots\rangle_{\text{sol},\Theta_B} = \frac{\int d\mathbf{x}_{HG}\int d\mathbf{X}_V\ (\cdots)\Theta_B(\mathbf{x}_{HG})e^{-\beta\mathcal{V}_{\text{sol}}^B}}{\int d\mathbf{x}_{HG}\int d\mathbf{X}_V\ \Theta_B(\mathbf{x}_{HG})e^{-\beta\mathcal{V}_{\text{sol}}^B}}, \tag{20}$$

$$\langle\cdots\rangle_{\text{ref},\Theta_B} = \frac{\int d\mathbf{x}_{HG}\int d\mathbf{X}_V\ (\cdots)\Theta_B(\mathbf{x}_{HG})e^{-\beta\mathcal{V}_{\text{ref}}^B}}{\int d\mathbf{x}_{HG}\int d\mathbf{X}_V\ \Theta_B(\mathbf{x}_{HG})e^{-\beta\mathcal{V}_{\text{ref}}^B}}, \tag{21}$$

the αth solvent distributions in the solution and reference systems when the bound complex is formed can be expressed as

$$\rho_{\text{sol},\alpha}^B(\varepsilon) = \langle\hat{\rho}_\alpha(\varepsilon)\rangle_{\text{sol},\Theta_B}, \tag{22}$$

$$\rho_{\text{ref},\alpha}^B(\varepsilon) = \langle\hat{\rho}_\alpha(\varepsilon)\rangle_{\text{ref},\Theta_B}, \tag{23}$$

respectively. According to Kirkwood's charging formula for the alchemical pathway connecting the solution and reference systems through the coupling parameter, $\Delta\mu_G^B$ is expressed using the integral over the coupling parameter. Introducing the Percus–Yevick (PY)-type and hypernetted-chain (HNC)-type approximations against the distributions for the non-endpoint systems on the alchemical pathway yields[34]

$$\Delta\mu_G^B = \sum_\alpha\ \int d\varepsilon\ \varepsilon\rho_{\text{sol},\alpha}^B(\varepsilon) - \frac{1}{\beta}\sum_\alpha\ \int d\varepsilon\ (\rho_{\text{sol},\alpha}^B(\varepsilon)-\rho_{\text{ref},\alpha}^B(\varepsilon))$$
$$+ \frac{1}{\beta}\sum_\alpha\ \int d\varepsilon\ \rho_{\text{sol},\alpha}^B(\varepsilon)\log\frac{\rho_{\text{sol},\alpha}^B(\varepsilon)}{\rho_{\text{ref},\alpha}^B(\varepsilon)}$$
$$+ \mathcal{F}[\rho_{\text{sol},\alpha}^B(\varepsilon),\rho_{\text{ref},\alpha}^B(\varepsilon),\chi_{\alpha\beta}^B(\varepsilon,\eta)], \tag{24}$$

where $\chi_{\alpha\beta}^B(\varepsilon,\eta)$ is the two-body density-correlation function defined as

$$\chi_{\alpha\beta}^B(\varepsilon,\eta) = \langle\hat{\rho}_\alpha(\varepsilon)\hat{\rho}_\beta(\eta)\rangle_{\text{ref},\Theta_B} - \langle\hat{\rho}_\alpha(\varepsilon)\rangle_{\text{ref},\Theta_B}\langle\hat{\rho}_\beta(\eta)\rangle_{\text{ref},\Theta_B}. \tag{25}$$

The first three terms in Eq. (24) are the pair free-energy components without approximations, and $\mathcal{F}$ is the approximate free-energy functional for the many-body entropic contributions. The explicit form of $\mathcal{F}$ is available in Ref. 34.

Evaluating the free energy using Eq. (24) is effective when the distributions in the solution ($\rho_{\text{sol},\alpha}^B(\varepsilon)$) and reference ($\rho_{\text{ref},\alpha}^B(\varepsilon)$) systems overlap well with each other. However, if the holo-form structures of host molecules observed in the solution system differ from their apo-form structures in the reference system, the distributions $\rho_{\text{sol},\alpha}^B(\varepsilon)$ and $\rho_{\text{ref},\alpha}^B(\varepsilon)$ may not overlap well. In such host–guest systems, the ε-region with $\rho_{\text{sol},H}^B(\varepsilon)\neq0$ and $\rho_{\text{ref},\alpha}^B(\varepsilon)=0$, which is problematic due to the integrand of the third term in Eq. (24),

$$\rho_{\text{sol},\alpha}^B(\varepsilon)\log\frac{\rho_{\text{sol},\alpha}^B(\varepsilon)}{\rho_{\text{ref},\alpha}^B(\varepsilon)}, \tag{26}$$

may be too broad, especially in the energy distribution for H [$\rho_{\text{sol},H}^B(\varepsilon)$ and $\rho_{\text{ref},H}^B(\varepsilon)$]. Then, we introduce a solution state involving the overlapped distributions with the reference (OR state), in which the above problematic ε-region is absent (Fig. 1). Such a state can be defined using the following characteristic function:

$$\Theta_{OR}(\mathbf{x}_{HG},\mathbf{X}_V) = \Theta_B(\mathbf{x}_{HG})\prod_\alpha\prod_{i\in\alpha}\theta(\rho_{\text{ref},\alpha}^B(u_\alpha(\mathbf{x}_H,\mathbf{x}_{\alpha,i}))). \tag{27}$$

Here, $\theta(x)$ is the Heaviside step function given by

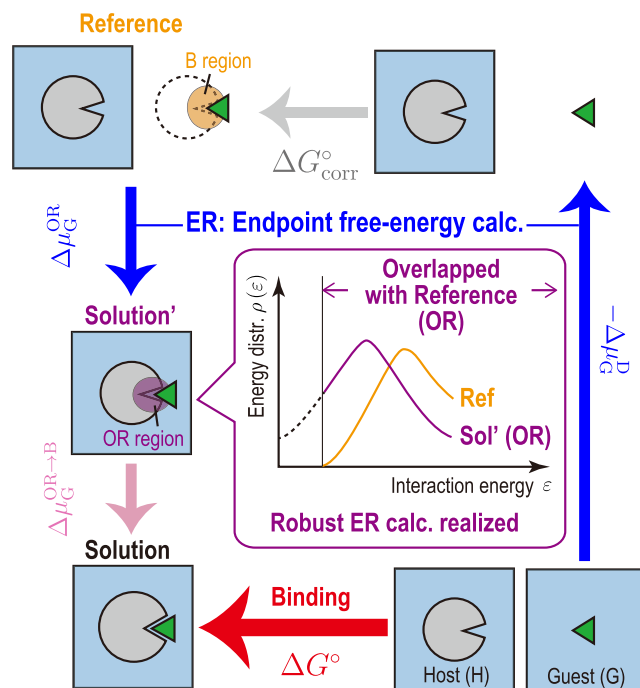$$\theta(x) = \begin{cases} 0, & x\leq0, \\ 1, & x>0. \end{cases} \tag{28}$$

FIG. 1. Thermodynamic cycle employed in the energy representation (ER) method incorporating a solution state involving the overlapped distributions with reference (ER-OR).

By defining the free-energy change associated with the transition from B in the reference system to the OR state in the solution system expressed as

$$\Delta\mu_G^{OR} = -\frac{1}{\beta}\log\frac{\int d\mathbf{x}_{HG}\int d\mathbf{X}_V \Theta_{OR}(\mathbf{x}_{HG},\mathbf{X}_V)e^{-\beta\mathcal{V}_{sol}^B}}{\int d\mathbf{x}_{HG}\int d\mathbf{X}_V \Theta_B(\mathbf{x}_{HG})e^{-\beta\mathcal{V}_{ref}^B}}, \quad (29)$$

$\Delta\mu_G^B$ can be decomposed into $\Delta\mu_G^{OR}$ and the free-energy change due to the transition from OR to B in the solution system, $\Delta\mu_G^{OR\to B}$, as

$$\Delta\mu_G^B = \Delta\mu_G^{OR} + \Delta\mu_G^{OR\to B}, \quad (30)$$

where $\Delta\mu_G^{OR\to B}$ is described as

$$\Delta\mu_G^{OR\to B} = \frac{1}{\beta}\log P_{OR}, \quad (31)$$

and P_{OR} is the probability of finding the OR state among the bound-complex configurations, given by

$$P_{OR} = \frac{\int d\mathbf{x}_{HG}\int d\mathbf{X}_V \Theta_{OR}(\mathbf{x}_{HG},\mathbf{X}_V)e^{-\beta\mathcal{V}_{sol}^B}}{\int d\mathbf{x}_{HG}\int d\mathbf{X}_V \Theta_B(\mathbf{x}_{HG})e^{-\beta\mathcal{V}_{sol}^B}}. \quad (32)$$

Since $\rho_{sol,\alpha}^B(\varepsilon)$ is the product of $\rho_{ref,\alpha}^B(\varepsilon)$ and a term referring to the solvent-mediated contribution to the potential of mean force, in principle $\rho_{ref,\alpha}^B(\varepsilon) \neq 0$ when $\rho_{sol,\alpha}^B(\varepsilon) \neq 0$. The difficulty related to Eq. (26) is a practical problem due to finite sampling. $\rho_{ref,\alpha}^B(\varepsilon)$ appearing in Eq. (27) should thus be understood as a numerically computed one, and in actual simulations, its argument (energy coordinate ε) is discretized to a set of bins with finite widths. According to Eq. (27), the OR state is a subset of the bound (B) state consisting only of the configurations for which all the pair-interaction energies of G with each species α fall into energy bins with non-zero $\rho_{ref,\alpha}^B(\varepsilon)$. If a sampled configuration in the B state contains a pair energy, which corresponds to zero $\rho_{ref,\alpha}^B(\varepsilon)$, that configuration is excluded from OR. The OR state was called intermediate state in Refs. 35 and 44. However, to avoid possible confusion with an intermediate state in BAR, it is denoted as OR in this work. The interaction of the solute with the surroundings is partially turned on in intermediate states of BAR, while the solute's interactions are fully turned on in the OR state. Hereafter, the ER theory incorporating the OR state is referred to as ER-OR. The ER-OR procedures are schematically depicted in Fig. S1 of the supplementary material.

III. COMPUTATIONAL METHODS

A. System setups

We investigated the self-association of *N*-methylacetamide (NMA) in different solvents (acetone, 1,4-dioxane, and chloroform) and the binding of aspirin to *β*-cyclodextrin (CD) in water (Fig. 2).

The TIP3P model was used for water, and the general Amber force field (GAFF)[45,46] was used for the other species. The following modeling scheme was adopted for all the species except water. We employed the restrained electrostatic potential (RESP) method[47] to determine the point charges on the atoms at HF/6-31G(d) level calculations. The optimized structures used for the RESP method were prepared at MP2/6-31G(d) level calculations except for *β*-cyclodextrin (CD). According to our previous study,[48] we optimized the CD structure at HF/6-31G(d) level calculations. The quantum chemical calculations mentioned above were performed with Gaussian 16,[49] and the Antechamber program was used for the RESP method.[50] The initial configurations of the systems of interest were built using Packmol.[51]

All the simulations were performed with GENESIS 2.0.[52–54] The Bussi method was used for generating the NVT and NPT ensembles.[55,56] The velocity Verlet (VVER)[57] and reversible reference system propagator algorithm (r-RESPA)[58] integrators were employed for the equilibration and production runs, respectively. The time intervals for VVER and r-RESPA were 2 and 2.5 fs, respectively. The cutoff distance for the Lennard-Jones (LJ) interactions was 9 Å, and smooth particle-mesh Ewald (SPME)[59] was used for computing the electrostatic interactions. The number of grids for SPME was automatically determined in GENESIS so that the grid spacing was shorter than 1.4 Å. All the bonds that involve hydrogen atoms were constrained with the SHAKE/RATTLE method,[60,61] and water molecules were treated as rigid molecules using the SETTLE method.[62]

B. *N*-methylacetamide (NMA) systems

1. Simulation setups

For the computation of $\Delta\mu_G^D$ [Eq. (10)], we prepared the trajectories for the system containing an NMA molecule in solvents and for the pure solvent systems, corresponding to the solution and reference systems for the dissociated (D) state. For both systems, the box size was 60^3 Å^3, and the number of solvent molecules was set to 1809, 1605, and 1545 for acetone, 1,4-dioxane, and chloroform, respectively. The number of solvent molecules was determined using the NPT simulations to ensure that the system volume fluctuated around 60^3 Å^3 at 300 K and 1 atm. For each system, we conducted

(a) *N*-methylacetamide (NMA)

(b) β-cyclodextrin (CD)–Aspirin
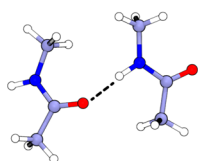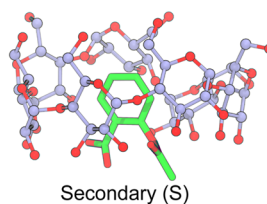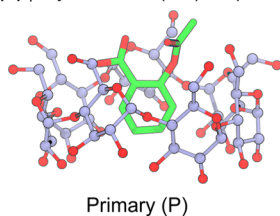
Primary (P)

Secondary (S)

FIG. 2. Target binding systems. (a) Self-association of *N*-methylacetamide (NMA). (b) *β*-cyclodextrin (CD)–aspirin binding. The CD–aspirin complexes in which the hydroxyl group of aspirin points toward the primary and secondary faces are labeled as P and S, respectively.

a 2 ns NVT simulation for equilibration, followed by a 10 ns NVT simulation for production.

Regarding $\Delta\mu_G^B$ [Eq. (11)], the trajectories for the solution and reference systems for the B state, respectively containing 2 and 1 NMA molecules, are needed. As for the latter system (reference), the trajectory of the solution system for the D state can be used. In the case of the former systems (solution), we conducted a 2 ns NVT simulation for equilibration, followed by a 100 ns NVT simulation for production. These simulations were conducted while applying a following half-flat bottom (HFB) potential on the distance between the centers of mass (CoM) of the two NMA molecules (d),

$$U_{HFB}(d) = \begin{cases} 0, & d < d_0, \\ k(d - d_0)^2, & d \geq d_0. \end{cases} \quad (33)$$

Here, $d_0 = 7$ Å and $k = 10$ kcal mol^{-1} Å^{-2}.

We performed the NVT simulations for an isolated NMA molecule that is required for the test-particle insertion in the ER-based methods. After 1 ns NVT simulation for equilibration, we performed a 1 ns NVT simulation for production.

To calculate ΔG° using the PMF-based method (exact),[48,63] we also conducted 100 ns NVT simulations in the solution system from the 2 ns equilibration mentioned above while applying $U_{HFB}(d)$ [Eq. (33)] with $d_0 = 15$ Å and $k = 10$ kcal mol^{-1} Å^{-2}.

2. Binding free-energy calculations

The binding free energies, ΔG°, were evaluated through the computation of $\Delta\mu_G^D$, $\Delta\mu_G^B$, and ΔG_{corr}°. In the case of $\Delta\mu_G^D$, the energy distributions for the solution and reference systems for the D state were computed. For the reference system, the test-particle insertion was performed for computing the distribution, with 1000 insertions for each configuration of the reference system. The error estimation of $\Delta\mu_G^D$ was done by dividing the solution trajectories into ten blocks for averaging.

Regarding $\Delta\mu_G^B$, the energy distributions in the solution system, $\rho_{sol,H}^B(\varepsilon)$, were computed using the configurations that satisfy the bound-complex criteria. In this work, the criteria were defined using the interatomic distances involving the oxygen (O) atoms of the carbonyl group and nitrogen (N) atoms of the secondary amine. If the minimum distance among the O–O, N–N, and O–N interatomic distances, d_{min}, was shorter than 3.5 Å, the NMA dimer was considered a bound complex. The configurations of the system that satisfy this criterion are part of the configurations generated

with the restraining potential of Eq. (33). $\rho_{sol,H}^B(\varepsilon)$ was constructed by using only those configurations within the distance threshold of 3.5 Å, and the other configurations were discarded. See Fig. S2 of the supplementary material for how the choice of the threshold affects the binding free energy. For the computation of the energy distribution in the reference system, $\rho_{ref,\alpha}^B(\varepsilon)$, the characteristic function for the B state, $\Theta_B(\mathbf{x}_{HB})$, was constructed using the spatial distribution function for the guest NMA, $g(\mathbf{r})$, and the Weeks–Chandler–Andersen (WCA) potential,[64] $u_{WCA}(\mathbf{x}_{HB})$, in addition to d_{HB}. Here, $\mathbf{r}$ is the CoM of the guest NMA and $g(\mathbf{r})$ was computed using the solution trajectories. $\rho_{ref,\alpha}^B(\varepsilon)$ was constructed using the configurations obtained from the test-particle insertion, which satisfy $d_{min} \leq 3.5$ Å, $g(\mathbf{r}) > 0$, and $u_{WCA}(\mathbf{x}_{HB}) \leq 15$ kcal mol^{-1}. The same characteristic function was used to perform the test-particle insertion for ΔG_{corr}°. The number of insertions was 1000 for each configuration of the reference system. We estimated the statistical error in $\Delta\mu_G^B$ by dividing the solution trajectories into ten blocks for averaging.

For comparison, we also computed ΔG° using the PMF-based approach.[48] In this calculation, only $d_{min} < 3.5$ Å was used for the bound-complex criteria as well as in the calculation of $\rho_{sol,\alpha}^B(\varepsilon)$. Note that the standard-state concentration was properly treated in this method, allowing for a valid comparison of the ΔG° values obtained from this method with those from the ER-based methods.

C. β-cyclodextrin (CD)–aspirin system
1. Simulation setups

We prepared the trajectories required for the computation of $\Delta\mu_G^D$ and $\Delta\mu_G^B$ (Table I). The simulation scheme was constructed according to our previous study.[48] The pure water system composed of 7200 water molecules with the box size of 60^3 Å^3 was built as the reference system for the D state. After the annealing of the system from 548 to 298 K during a 0.1 ns NVT simulation, we performed a 1 ns NVT simulation for equilibration. Then, we determined the system size by a 1 ns NPT simulation at 1 atm. The system size at the final step was 60.20^3 Å^3, and this size was used for the other systems described below. After further equilibration (0.1 ns NVT), we conducted a 10 ns NVT simulation for production. The solution system for the D state contains an aspirin and 7200 water molecules. The system was annealed from 548 to 298 K during a 0.1 ns NVT simulation, followed by a 0.1 ns NVT simulation for equilibration. Then, we conducted a 10 ns NVT simulation for production.

TABLE I. Information on the trajectories used for the free-energy calculations in the CD–aspirin systems. The values in parentheses indicate the number of replicas for the BAR simulations. For the BAR simulations, the last 40 and 30 ns were used for the D and B states, respectively.

	States					
	D			B		
	No. of Traj.	Simul. length (ns)	Total (ns)	No. of Traj.	Simul. length (ns)	Total (ns)
BAR	1 (24)	100	2400	10 (57)	150	85 500
ER, ER-OR (solution)	1	20	20	25	20	500
ER, ER-OR (reference)	1	10	10	25	20	500

In the CD–guest systems, it is well known that there are two distinct bound complexes, referred to as primary (P) and secondary (S) complexes.[65] In the P and S complexes, the hydroxyl group of aspirin points toward the primary and secondary faces of CD, respectively. We selected 25 different conformations from the trajectories in our previous study for each complex.[48] Using these conformations, 25 initial configurations of the solution system for the B state, each containing an aspirin, a CD, and 7200 water molecules were constructed for each complex. For each initial configuration, we performed a 0.1 ns NVT simulation for equilibration while imposing the positional restraints on the heavy atoms of the CD and aspirin with the force constant of 1 kcal mol^{-1} Å^{-2}. Then, a 0.1 ns NVT equilibration was performed. Following this, we conducted a 25 ns NVT production run, and the final 20 ns trajectory was used for analysis. In the case of the reference system, a CD and 7200 water molecules are involved. We prepared 25 initial configurations for this system. Then, we equilibrated the system using a 0.1 ns NVT simulation for each configuration, followed by 25 ns NVT simulations for production. The last 20 ns trajectory was used for analysis.

We also performed the Bennett acceptance ratio (BAR)[13] simulations for the D and B states to compute ΔG° based on the double-annihilation scheme (DAS). In the case of the D state, the initial configuration was taken from the final snapshot after the equilibration for the solution system. The BAR method combined with Hamiltonian replica-exchange MD (BAR/H-REMD)[66] implemented in GENESIS[67,68] was performed with a simulation time of 100 ns. The setup of the intermediate states (24 states) was the same as that in our previous study on the membrane permeation.[69] The last 40 ns trajectory for each state was used for the analysis. Regarding the B state, we selected the ten configurations of the solution system obtained after the 5 ns production simulations for each bound complex. Then, we conducted a 150 ns BAR/H-REMD simulation for each configuration. The last 30 ns trajectory for each state was used for the analysis. The intermediate states were defined using the soft-core electrostatic (elec) and van der Waals (vdW) interactions with the coupling parameters λ_{elec} (1.000, 0.950, 0.900, 0.850, 0.800, 0.750, 0.700, 0.650, 0.600, 0.550, 0.500, 0.450, 0.400, 0.350, 0.300, 0.250, 0.200, 0.150, 0.100, 0.050, and 0.000) and λ_{vdW} (1.000, 0.950, 0.900, 0.850, 0.800, 0.750, 0.700, 0.650, 0.600, 0.550, 0.500, 0.450, 0.400, 0.350, 0.325, 0.300, 0.275, 0.250, 0.225, 0.200, 0.175, 0.150, 0.140, 0.130, 0.120, 0.110, 0.100, 0.090, 0.080, 0.070, 0.060, 0.050, 0.040, 0.030, 0.020, 0.010, and 0.000). $\lambda_i = 0$ and 1, respectively, correspond to the fully decoupled and coupled states for interaction energy component i (i = elec or vdW). The total number of states is 57. For all the states, we imposed the HFB potential [Eq. (33)] on the distance between the CoMs of CD and aspirin, d, with $d_0 = 6$ Å and $k = 10$ kcal mol^{-1} Å^{-2}, and on the attractive part of the LJ interaction[48] between CD and aspirin, $u_{attr}(\mathbf{x}_{HG})$, defined as

$$U_{FB}^{uattr}(u_{attr}) = \begin{cases} k(u_{attr} - u_{lower})^2, & u_{attr} \leq u_{lower}, \\ 0, & u_{lower} < u_{attr} \leq u_{upper}, \\ k(u_{attr} - u_{upper})^2, & u_{attr} > u_{upper}. \end{cases} \quad (34)$$

The force constant, k, was set to 10 kcal^{-1} mol, and (u_{lower}, u_{upper}) was set to ($-38.86, -9.35$) for P and to ($-37.71, -11.95$) for S in units of kcal mol^{-1}. Note that the values of (u_{lower}, u_{upper}) were determined from the lower and upper limits of u_{attr} observed in the solution systems.

2. Binding free-energy calculations

The scheme for computing ΔG° using the ER-based methods was almost parallel to that used for the NMA systems (Sec. III B 2), and the same approach was applied for $\Delta\mu_G^D$. Therefore, only the settings specific to the computation of $\Delta\mu_G^B$ in the CD–aspirin system are described here. In the simulations for the B state in the solution system, we confirmed that aspirin maintained its initial bound pose throughout the simulations. Thus, all the configurations generated in the solution system for the B state were used to compute $\rho_{sol,\alpha}^B(\varepsilon)$ for each bound pose. For the computation of $\rho_{ref,\alpha}^B(\varepsilon)$ and ΔG_{corr}°, we constructed the characteristic function for the B state, $\Theta(\mathbf{x}_{HG})$, in terms of the spatial distribution function for the CoM of aspirin, $g(\mathbf{r})$, and u_{attr}. We computed $\rho_{ref,\alpha}^B(\varepsilon)$ using the configurations obtained from the test-particle insertion of aspirin into the reference trajectories that satisfy $g(\mathbf{r}) > 0$ and $u_{lower} \leq u_{attr}(\mathbf{x}_{HG}) \leq u_{upper}$. ($u_{lower}, u_{upper}$) was set to ($-38.86, -9.35$) for P and to ($-37.71, -11.95$) for S in units of kcal mol^{-1}. The number of insertions for each configuration was 10 000 for both $\rho_{ref,\alpha}^B(\varepsilon)$ and ΔG_{corr}°. As noted in the last paragraph of Sec. II B, $\rho_{ref,\alpha}^B(\varepsilon)$ was constructed by discretizing the energy coordinate ε. The bin width of discretization was 0.05 kcal mol^{-1} in the energy range where Eq. (26) becomes problematic. In the case of ΔG_{corr}°, the structure of an isolated aspirin was inserted to the spatial region containing CD, with the volume of 20^3 Å. The error in $\Delta\mu_G^B$ was estimated from the different trajectories for the solution system.

We also computed ΔG° using the BAR simulations. Let $\Delta G_{BAR,D}$ and $\Delta G_{BAR,B}$ represent the free-energy changes along the alchemical pathways in the BAR simulations associated with the appearance of the interactions between aspirin and its surrounding environments for the D and B states, respectively. Then, ΔG° can be expressed as

$$\Delta G^\circ = \Delta G_{BAR,B} - \Delta G_{BAR,D} + \Delta G_{BAR,corr}^\circ, \quad (35)$$

where $\Delta G_{BAR,corr}^\circ$ is the standard-state correction. For the computation of $\Delta G_{BAR,B}$, the snapshots satisfying $u_{lower} \leq u_{attr} \leq u_{upper}$, $d \leq 6$ Å, and the primary/secondary poses criteria for the aspirin's orientation (Fig. S3 of the supplementary material) were used for the fully coupled ($\lambda_{elec} = \lambda_{vdW} = 1$) and intermediate states. As for the fully decoupled state ($\lambda_{elec} = \lambda_{vdW} = 0$), the aspirin's orientation was not used for selecting the snapshots. According to Eq. (15), the standard-state correction, $\Delta G_{BAR,corr}^\circ$, is expressed as

$$\Delta G_{BAR,corr}^\circ = -\frac{1}{\beta}\log\left(c^\circ V \frac{\int d\mathbf{x}_G \int d\mathbf{X}_V\, \Theta_B^{BAR}(\mathbf{x}_{HG})e^{-\beta v_{ref}^B}}{\int d\mathbf{x}_{HG} \int d\mathbf{X}_G\, e^{-\beta v_{ref}^B}}\right), \quad (36)$$

where

$$\Theta_B^{BAR}(\mathbf{x}_{HG}) = \begin{cases} 1, & u_{lower} \leq u_{attr} \leq u_{upper} \text{ and } d \leq 6 \text{ Å}, \\ 0, & \text{otherwise.} \end{cases} \quad (37)$$

$\Delta G_{BAR,corr}^\circ$ was computed using the test-particle insertion of aspirin to the reference trajectories. The number of insertions was 10 000. The error in $\Delta G_{BAR,D}$ was estimated by dividing the trajectory of each state into eight blocks for averaging, and that in $\Delta G_{BAR,B}$ was estimated from the different BAR simulation runs.

IV. RESULTS AND DISCUSSION

A. Self-association of *N*-methylacetamide (NMA) in different solvents

1. Energy distribution

We examine the energy distributions of the host NMA molecule in the solution ($\rho_{\text{sol,H}}^{\text{B}}(\varepsilon)$) and reference ($\rho_{\text{ref,H}}^{\text{B}}(\varepsilon)$) systems for the bound complex (Fig. 3). Note that one of the NMA molecules is regarded as the host, while the other is considered guest. In acetone [Fig. 3(a)], $\rho_{\text{sol,H}}^{\text{B}}(\varepsilon)$ exhibits a broad peak at $\varepsilon \sim -8.5$ kcal mol^{-1}. As illustrated in Fig. 2(a), the bound complex is stabilized by the hydrogen bond between the carbonyl oxygen atom and the hydrogen atom in the secondary amine, and thus, the electrostatic interaction has a dominant contribution to $\rho_{\text{sol,H}}^{\text{B}}(\varepsilon)$. We confirm that the average value of the interaction energies between the two NMA molecules for the electrostatic component is -6.28 ± 0.01 kcal mol^{-1}, which is significantly larger in magnitude than that for the van der Waals component, -1.049 ± 0.006 kcal mol^{-1}. It is found that the profile of $\rho_{\text{sol,H}}^{\text{B}}(\varepsilon)$ is largely independent on the solvent species, meaning that the distribution of the bound-complex structures is insensitive to the surrounding environments. Similar to $\rho_{\text{sol,H}}^{\text{B}}(\varepsilon)$, $\rho_{\text{ref,H}}^{\text{B}}(\varepsilon)$ hardly changes its profile across the solvent species. Since the non-overlapping ε-region $[\rho_{\text{sol,H}}^{\text{B}}(\varepsilon) \neq 0$ and $\rho_{\text{ref,H}}^{\text{B}}(\varepsilon) = 0]$ is sufficiently narrow in all the solvent systems, it is expected that the "solvation free energy" in the bound complex, $\Delta\mu_{\text{G}}^{\text{B}}$, can be computed using the ER method without introducing the OR state [Eq. (27)], as will be discussed in Subsection IV A 2.

2. Binding free energy

Here, we compare the values of the binding free energies, ΔG°, obtained from the potentials of mean force (PMF) in an exact way with those from the ER and ER-OR methods to verify the accuracy of the ER-based methods. The ER and ER-OR methods yield virtually identical ΔG° values in all the solvents examined. Given
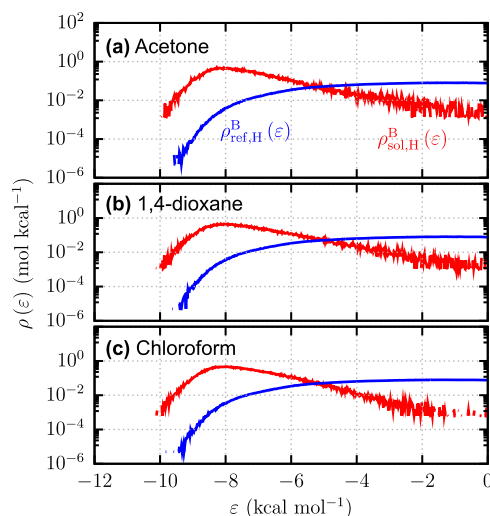
TABLE II. Binding free energies for the NMA systems obtained through PMF, ER, and ER-OR methods. The errors are provided at the standard error.

Solvent	ΔG° (kcal mol^{-1})		
	PMF	ER	ER-OR
Acetone	0.8 ± 0.1	0.61 ± 0.02	0.60 ± 0.02
1,4-dioxane	0.1 ± 0.1	0.52 ± 0.03	0.52 ± 0.03
Chloroform	-0.36 ± 0.07	0.04 ± 0.03	0.03 ± 0.04

that the energy distributions in the solution and reference systems $[\rho_{\text{sol,H}}^{\text{B}}(\varepsilon)$ and $\rho_{\text{ref,H}}^{\text{B}}(\varepsilon)]$ overlap well, it is reasonable for the ER-based methods to yield the identical ΔG° values, regardless of the introduction of the overlapped state. The values of ΔG° evaluated in this work are listed in Table II, and the correlation plots of the binding free energies obtained from the ER-OR method, $\Delta G_{\text{ER-OR}}^{\circ}$, against those from the PMF, $\Delta G_{\text{PMF}}^{\circ}$, are shown in Fig. 4. The PMF method estimates that ΔG° decreases in the order of acetone >1,4-dioxane >chloroform. This ordering is consistent with the experimental findings that the binding constant for the self-association of NMA increases as solvent polarity decreases.[42] It is seen that the ER and ER-OR methods reproduce the ΔG° ordering predicted from the PMF method. Furthermore, the deviation from the PMF method is within 0.5 kcal mol^{-1} in all the solvents. Since the NMA dimer has the shallow free-energy minimum in the PMF, ΔG° is sensitive to the variation in the bound-complex criteria (Fig. S2 of the supplementary material). However, we confirm that both the ER and PMF methods exhibit the similar behaviors against the variation, and the ΔG° ordering is not altered.

B. *β*-cyclodextrin (CD)–aspirin system

1. Structure and energy distribution of CD

We first assess the impact of binding on the structural population of CD. Figure 5(a) illustrates the distribution of the minimum

FIG. 3. Energy distributions of the host NMA molecule in (a) acetone, (b) 1,4-dioxane, and (c) chloroform.
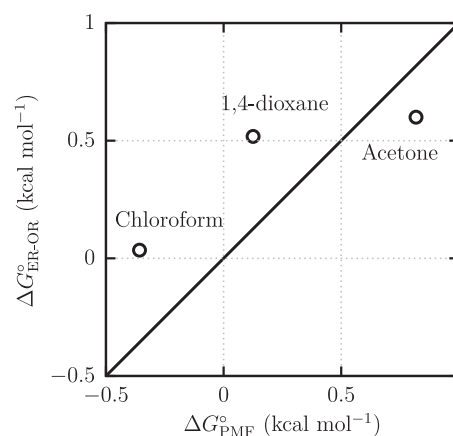

FIG. 4. Correlation plots of the binding free energies obtained from the ER-OR method, $\Delta G_{\text{ER-OR}}^{\circ}$, against those from the PMF, $\Delta G_{\text{PMF}}^{\circ}$. The error bars are not shown in the figure because the standard errors are smaller than 0.1 kcal mol^{-1} in all the solvent systems.
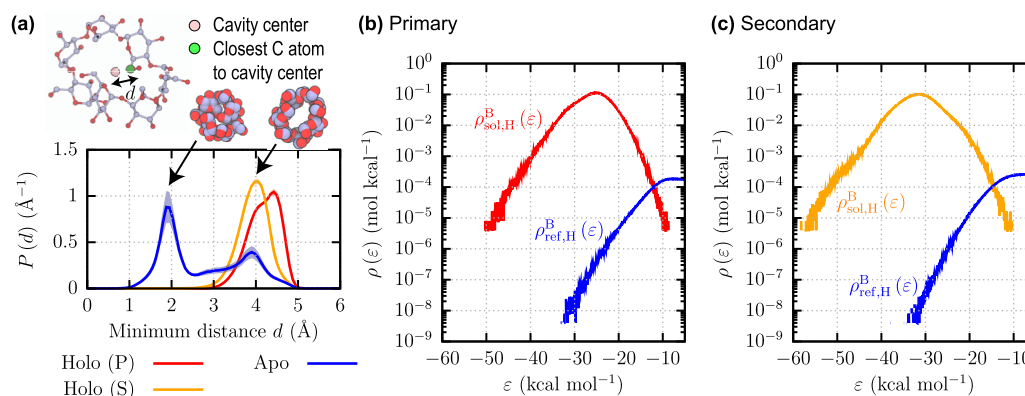
FIG. 5. Structural difference of CD between apo- and holo-forms, and energy distributions of CD. (a) Minimum-distance distribution between the cavity center and carbon atoms in CD. (b) and (c) Energy distributions of CD in the solution system, $\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon)$, and the reference system, $\rho^{\mathrm{B}}_{\mathrm{ref,H}}(\varepsilon)$, for complex P (b) and complex S (c). The cavity center of CD is defined as the center of mass for the ether oxygen atoms.

distance between the cavity center of CD and its carbon atoms, denoted as $P(d)$. The cavity center is defined as the center of mass for the ether oxygen atoms in CD. For the holo-forms of CD in complexes P and S, peaks in $P(d)$ at $d = 4.5$ Å and 4 Å, respectively, indicate open conformations where the internal cavity of CD is accessible to the guest molecule. In addition to these open conformations, the apo-form exhibits a sharp peak at $d = 2$ Å, reflecting closed conformations where the center of CD is occupied by its own atoms. In such conformations, one of the sugar rings in CD rotates so that its plane is closer to the cavity center. The open and closed conformations in the apo-form are also reported by Tang and Chang[41] as well as by Harris et al.[70] The localization of $P(d)$ to the distribution corresponding to the open conformations upon binding indicates that the structural fluctuations of CD are suppressed by aspirin.

The energy distributions of CD in the solution $(\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon))$ and reference systems $(\rho^{\mathrm{B}}_{\mathrm{ref,H}}(\varepsilon))$ for P and S are shown in Figs. 5(b) and 5(c), respectively. The peak of $\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon)$ is located at $\varepsilon \sim -25$ kcal mol^{-1} for P and $\varepsilon \sim -31.5$ kcal mol^{-1} for S. Furthermore, the tail of $\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon)$ extends further into the negative region for S than for P. This indicates that the direct interaction between aspirin and CD is stronger for S than for P. In the case of the reference system, the difference in $\rho^{\mathrm{B}}_{\mathrm{ref,H}}(\varepsilon)$ between P and S is found to be negligibly small. The peak position of $\rho^{\mathrm{B}}_{\mathrm{ref,H}}(\varepsilon)$ is $\varepsilon \sim -10$ kcal mol^{-1}, and it is shifted in the positive direction from that of $\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon)$ (-25 and -31.5 kcal mol^{-1} for P and S, respectively). As shown in the profiles of $P(d)$ [Fig. 5(a)], the accessible d-region in the holo-form (solution system) is fully covered by that in the apo-form (reference system). However, a wide non-overlapping region between $\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon)$ and $\rho^{\mathrm{B}}_{\mathrm{ref,H}}(\varepsilon)$ is present. This suggests the presence of a difference in the CD structure between the apo- and holo-forms that is not captured by $P(d)$ [Fig. 5(a)].

2. Binding free energy

We summarize the binding free energies, ΔG°, evaluated using the BAR, ER, and ER-OR methods in Fig. 6(a) and Table III.

According to the results from the BAR method, the thermodynamic stabilities of complexes P and S are nearly comparable to each other. In the case of the ER method, the stability of P is predicted to be higher than that of S. The values of ΔG° for S obtained from the ER method differ by more than 2 kcal mol^{-1} from those obtained using the BAR method. Meanwhile, the ER-OR method reproduces the result revealed by the BAR method that the values of ΔG° for P and S are similar to each other. The improvement achieved by the ER-OR method indicates that the introduction of the OR state into the ER method is beneficial for robust free-energy calculations, when the non-overlapping ε-region [$\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon) \neq 0$ and $\rho^{\mathrm{B}}_{\mathrm{ref,H}}(\varepsilon) = 0$] is too broad and the interpolation/extrapolation scheme employed in the ER method does not work properly. The most time-consuming part of the ΔG° calculation for both the ER-OR and BAR methods is the MD simulations for the B state. The convergence of the computed ΔG° with respect to the simulation timescale is much faster with the ER-OR method than with the BAR method (Fig. S4 of the supplementary material). It is also found that sufficiently long equilibration (≥ 120 ns in this system) is required to obtain the reliable estimates of ΔG° using the BAR method. According to Table I, the computational cost of these simulations required in the ER-OR method is orders-of-magnitude lower than in BAR given that the error in ER-OR is smaller by a factor of ~2. Introducing a sophisticated scheme of applying the restraint potentials in the BAR simulation, such as the virtual bond algorithm (VBA),[18,19] could accelerate the convergence while maintaining the robustness.

To clarify the driving force of the binding, we elucidate the importance of the interaction energy between aspirin and the surrounding environments on ΔG°. According to the endpoint DFT theory, one can decompose the solvation free energy of aspirin, $\Delta\mu^{\mathrm{X}}_{\mathrm{G}}$ (X = B or D) [Eqs. (10) and (11)], into the ensemble average of the interaction energy between aspirin and its surrounding environments in the solution system at state X, U^{X}, and the residual part, $\Delta\mu^{\mathrm{X}}_{\mathrm{res}}$. Thus, ΔG° can be expressed from Eq. (14) as

$$\Delta G^{\circ} = \Delta U + \Delta G_{\mathrm{res}}, \tag{38}$$
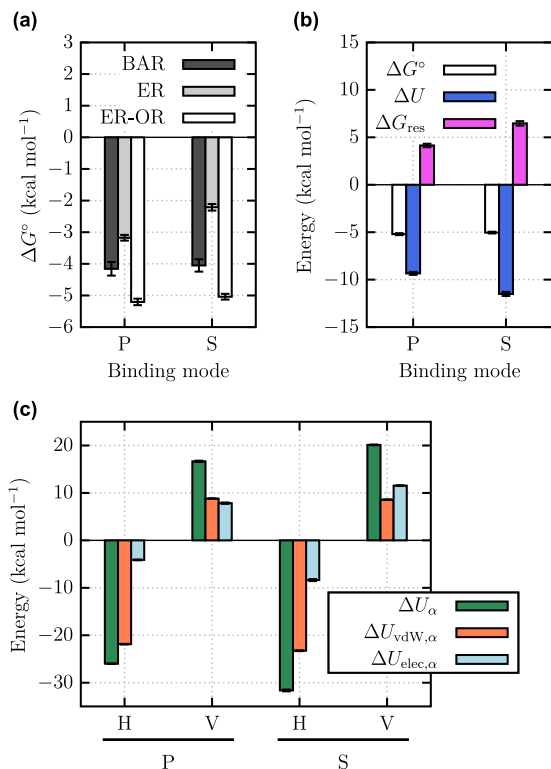
FIG. 6. Binding free energy, ΔG°, and the decomposition analysis based on the interaction energy. (a) ΔG° evaluated through the BAR, ER, and ER-OR methods. (b) Decomposition of ΔG° into the interaction-energy (ΔU) and other (ΔG_{res}) contributions using Eq. (38). In this analysis, the values of ΔG° obtained from the ER-OR methods are used. (c) Decomposition of ΔU into the interaction-energy components for each species α [α = H(CD) or V(water)]. The errors are provided at the standard error.

TABLE III. Binding free energies for the CD–aspirin system obtained through BAR, ER, and ER-OR methods. The errors are provided at the standard error.

	ΔG° $\left(\text{kcal mol}^{-1}\right)$		
Bound complex	BAR	ER	ER-OR
P	-4.2 ± 0.2	-3.14 ± 0.09	-5.2 ± 0.1
S	-4.1 ± 0.2	-2.2 ± 0.1	-5.03 ± 0.09

where

$$\Delta U = U^{\mathrm{B}} - U^{\mathrm{D}}, \tag{39}$$

$$\Delta G_{\mathrm{res}} = \Delta\mu_{\mathrm{res}}^{\mathrm{B}} - \Delta\mu_{\mathrm{res}}^{\mathrm{D}} + \Delta G_{\mathrm{corr}}^\circ. \tag{40}$$

Note that ΔG_{res} consists of the contribution from the pair free-energy components, the many-body entropic contributions, and the standard-state correction. Figure 6(b) shows the decomposition of ΔG° using Eq. (38). In this analysis, the values of ΔG° obtained from the ER-OR methods are used. In both bound complexes, it is found

that the binding is facilitated by ΔU and suppressed by ΔG_{res} and that the trend of $|\Delta U| > \Delta G_{\mathrm{res}}$ leads to a negative ΔG°. The value of ΔU for S is decreased from that for P, but this decrease is almost canceled out by the increase of ΔG_{res}, resulting in comparable stability of P and S. Since the distribution of $P(d)$ for S is sharper than that for P [Fig. 6(a)], the entropic penalty due to the restriction of the CD structure in S may account for the larger value of ΔG_{res} for S.

ΔU is decomposed into the van der Waals and electrostatic interaction-energy components of aspirin with species α [α = H(CD) or V(water)], denoted as $\Delta U_{\mathrm{vdW},\alpha}$ and $\Delta U_{\mathrm{elec},\alpha}$, respectively. The analysis based on this decomposition is presented in Fig. 6(c). Regardless of the complex types, the attractive interaction between aspirin and CD, ΔU_{H}, primarily contributes to ΔU through the van der Waals component, $\Delta U_{\mathrm{vdW},\alpha}$. This observation is consistent with the well-known binding mechanism in which the CD cavity provides a hydrophobic environment, enabling guest molecules to be captured through the hydrophobic interactions with CD.[71] The contribution of the interaction energy between aspirin and water, ΔU_{V}, tends to inhibit binding, reflecting the dehydration penalty. It is observed that both $\Delta U_{\mathrm{vdW},\mathrm{V}}$ and $\Delta U_{\mathrm{elec},\mathrm{V}}$ contribute almost equally to this penalty.

V. CONCLUSION

In this study, we developed a methodology to compute the binding free energies based on the energy representation (ER) theory. The ER theory enables us to calculate the free-energy difference between the two systems of interest, referred to as the solution and reference systems. Unlike other free-energy methods, there is no need to conduct the MD simulations for the intermediate states connecting the solution and reference systems, leading to the reduction in the computational cost. In applications to the calculation of the binding free energy for the host–guest systems (ΔG°), however, the applicability of the ER theory was limited to the host molecules whose structures in the holo-form resemble those in the apo-form. In the present method, this problematic structural difference was identified through the distributions on the host–guest interaction energy (energy distributions) for the solution and reference systems. By introducing a solution state involving the overlapped distributions with the reference (OR state), we achieved a robust binding free-energy calculation for such host molecules. The original method is referred to as the ER method, while the present method is referred to as the ER-OR method. It is noteworthy that, since this state is a subset of the target solution state, introducing the additional state into the ER method brings no extra computational costs compared to the ER method.

The present method (ER-OR) was first applied to the self-association of N-methylacetamide (NMA) in different solvents (acetone, 1,4-dioxane, and chloroform). It was found that ΔG° decreases in the order of acetone >1,4-dioxane >chloroform, which aligns with the experimental observations. Since the energy distribution for the guest NMA in the solution and in the reference system overlapped well, the ΔG° values evaluated through the ER and ER-OR methods were virtually identical. The comparison of the obtained ΔG° values with those from the exact method revealed that the differences in ΔG° between the two methods are within 0.5 kcal mol^{-1} in all the solvents.

The binding of aspirin to β-cyclodextrin (CD) in water was selected as the second target. In this system, there are two distinct bound complexes, primary (P) and secondary (S) complexes, and the CD structure in the holo-form is significantly different from that in the apo-form. For the bound state, the energy distribution of CD for the solution system was found to be not overlapped well with those for the reference system due to the difference in the CD structures between the holo- and apo-forms. As a result, the differences in $\Delta G°$ between the ER and ER-OR methods were larger than 1 kcal mol^{-1} for both P and S. The ER-OR method reproduced the result revealed by the BAR method that the thermodynamic stabilities of P and S are similar to each other, indicating an increase in reliability with the introduction of the OR state.

The present method works when both of the bound and dissociated structures are provided. Still, it can be employed with any schemes of structure prediction. For instance, AlphaFold 2/3[72,73] is a choice for preparing a structure that is not known in advance. The combination of the machine learning (ML)-based schemes of structure prediction and an all-atom scheme for free-energy evaluation, such as the present method, will be a promising direction.

Since the computational cost is lower in the present method compared to the other free-energy calculation methods, its application to the complex host–guest binding systems appears promising. For instance, peptide compounds that bind to their target exhibit the high flexibility.[74] In the present method, the simulations are required only for the endpoint (solution and reference) states, allowing for the incorporation of advanced sampling techniques to treat such high flexibility, despite their high computational costs. On the other hand, challenges still remain in the theoretical treatment of the host conformations. In the present method, we assumed the existence of an overlapped region in the energy distributions between the solution and reference systems. However, such a region may be absent in proteins (hosts) that exhibit global conformational changes through the induced-fit mechanism.[75] Introducing an additional state, in which the host molecule has structures close to those at the bound state but does not bind the guest, into the thermodynamic cycle for $\Delta G°$ might be useful for overcoming this challenge. We believe that the present method and its extensions would be beneficial for unveiling binding mechanisms in various host–guest binding systems.

SUPPLEMENTARY MATERIAL

The supplementary material contains the protocols for computing binding free energy through the ER-based methods, the dependency of $\Delta G°$ on the definition of the bound state for the NMA systems, the definition of the primary (P) and secondary (P) poses in the CD–aspirin system, and the convergence of $\Delta G°$ with the trajectory length.

ACKNOWLEDGMENTS

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

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

DATA AVAILABILITY

The source code of ER-OR program and sample data are available at https://github.com/kenkasa/eror-bfe. The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] R. A. Weinberg and R. A. Weinberg, *The Biology of Cancer* (WW Norton & Company, 2006).

[2] X. Du, Y. Li, Y.-L. Xia, S.-M. Ai, J. Liang, P. Sang, X.-L. Ji, and S.-Q. Liu, Int. J. Mol. Sci. **17**, 144 (2016).

[3] T. Daviter, C. M. Johnson, S. H. McLaughlin, and M. A. Williams, *Protein-ligand Interactions* (Springer, 2021).

[4] M. K. Gilson and H.-X. Zhou, Annu. Rev. Biophys. Biomol. Struct. **36**, 21 (2007).

[5] R. Baron, *Computational Drug Discovery and Design* (Springer, 2012).

[6] J. R. Schames, R. H. Henchman, J. S. Siegel, C. A. Sotriffer, H. Ni, and J. A. McCammon, J. Med. Chem. **47**, 1879 (2004).

[7] W. L. Jorgensen, Science **303**, 1813 (2004).

[8] J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson, and V. S. Pande, Curr. Opin. Struct. Biol. **21**, 150 (2011).

[9] M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli, J. Med. Chem. **59**, 4035 (2016).

[10] E. King, E. Aitchison, H. Li, and R. Luo, Front. Mol. Biosci. **8**, 712085 (2021).

[11] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).

[12] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).

[13] C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).

[14] W. L. Jorgensen, J. K. Buckner, S. Boudon, and J. Tirado-Rives, J. Chem. Phys. **89**, 3742 (1988).

[15] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, Biophys. J. **72**, 1047 (1997).

[16] H. Fujitani, Y. Tanida, M. Ito, G. Jayachandran, C. D. Snow, M. R. Shirts, E. J. Sorin, and V. S. Pande, J. Chem. Phys. **123**, 084108 (2005).

[17] H. Fujitani, Y. Tanida, and A. Matsuura, Phys. Rev. E **79**, 021914 (2009).

[18] S. Boresch, F. Tettinger, M. Leitgeb, and M. Karplus, J. Phys. Chem. B **107**, 9535 (2003).

[19] S. Boresch, J. Chem. Inf. Model. **64**, 3605 (2024).

[20] H.-J. Woo and B. Roux, Proc. Natl. Acad. Sci. U. S. A. **102**, 6825 (2005).

[21] J. C. Gumbart, B. Roux, and C. Chipot, J. Chem. Theory Comput. **9**, 794 (2013).

[22] S. Jo, W. Jiang, H. S. Lee, B. Roux, and W. Im, J. Chem. Inf. Model. **53**, 267 (2013).

[23] G. Heinzelmann and M. K. Gilson, Sci. Rep. **11**, 1116 (2021).

[24] H. Fu, C. Chipot, X. Shao, and W. Cai, J. Phys. Chem. B **127**, 10459 (2023).

[25] H. Fu, H. Chen, W. Cai, X. Shao, and C. Chipot, J. Chem. Inf. Model. **61**, 2116 (2021).

[26] S. Kim, H. Oshima, H. Zhang, N. R. Kern, S. Re, J. Lee, B. Roux, Y. Sugita, W. Jiang, and W. Im, J. Chem. Theory Comput. **16**, 7207 (2020).

[27] R. Liu, W. Li, Y. Yao, Y. Wu, H.-B. Luo, and Z. Li, J. Chem. Inf. Model. **63**, 7755 (2023).

[28] E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J. Z. Zhang, and T. Hou, Chem. Rev. **119**, 9478 (2019).

[29] A. de Ruiter and C. Oostenbrink, Curr. Opin. Struct. Biol. **61**, 207 (2020).

[30] T. Hansson, J. Marelius, and J. Åqvist, J. Comput.-Aided Mol. Des. **12**, 27 (1998).

[31] M. K. Gilson, L. E. Stewart, M. J. Potter, and S. P. Webb, J. Chem. Theory Comput. **20**, 6328 (2024).

[32] R. M. Levy, N. Matubayasi, and B. W. Zhang, J. Phys. Chem. B **124**, 11771 (2020).

[33] N. Matubayasi and M. Nakahara, J. Chem. Phys. **113**, 6070 (2000).

[34] N. Matubayasi, Bull. Chem. Soc. Jpn. **92**, 1910 (2019).

[35] T. Kawakami, I. Shigemoto, and N. Matubayasi, J. Chem. Phys. **137**, 234903 (2012).

[36] H. Kojima, K. Handa, K. Yamada, and N. Matubayasi, J. Phys. Chem. B **125**, 9357 (2021).

[37] S. Tanaka, N. Yamamoto, K. Kasahara, Y. Ishii, and N. Matubayasi, J. Phys. Chem. B **126**, 5274 (2022).

[38] K. Kasahara, R. Masayama, K. Okita, and N. Matubayasi, J. Chem. Phys. **159**, 134103 (2023).

[39] D. L. Mobley, A. P. Graves, J. D. Chodera, A. C. McReynolds, B. K. Shoichet, and K. A. Dill, J. Mol. Biol. **371**, 1118 (2007).

[40] D. L. Mobley, J. D. Chodera, and K. A. Dill, J. Chem. Theory Comput. **3**, 1231 (2007).

[41] Z. Tang and C.-e. A. Chang, J. Chem. Theory Comput. **14**, 303 (2017).

[42] H. L. Schenck and K. Hui, J. Chem. Educ. **88**, 1158 (2011).

[43] N. Matubayasi and M. Nakahara, J. Chem. Phys. **117**, 3605 (2002).

[44] S. Sakuraba and N. Matubayasi, J. Comput. Chem. **35**, 1592 (2014).

[45] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).

[46] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, J. Mol. Graphics Modell. **25**, 247 (2006).

[47] P. Cieplak, W. D. Cornell, C. Bayly, and P. A. Kollman, J. Comput. Chem. **16**, 1357 (1995).

[48] K. Kasahara, R. Masayama, K. Okita, and N. Matubayasi, J. Chem. Phys. **155**, 204503 (2021).

[49] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji et al., GAUSSIAN 16, Revision B.01, Gaussian, Inc., Wallingford, CT, 2016.

[50] 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).

[51] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, J. Comput. Chem. **30**, 2157 (2009).

[52] J. Jung, T. Mori, C. Kobayashi, Y. Matsunaga, T. Yoda, M. Feig, and Y. Sugita, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **5**, 310 (2015).

[53] C. Kobayashi, J. Jung, Y. Matsunaga, T. Mori, T. Ando, K. Tamura, M. Kamiya, and Y. Sugita, J. Comput. Chem. **38**, 2193 (2017).

[54] J. Jung, C. Kobayashi, K. Kasahara, C. Tan, A. Kuroda, K. Minami, S. Ishiduki, T. Nishiki, H. Inoue, Y. Ishikawa et al., J. Comput. Chem. **42**, 231 (2021).

[55] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).

[56] G. Bussi, T. Zykova-Timan, and M. Parrinello, J. Chem. Phys. **130**, 074101 (2009).

[57] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, J. Chem. Phys. **76**, 637 (1982).

[58] M. Tuckerman, B. J. Berne, and G. J. Martyna, J. Chem. Phys. **97**, 1990 (1992).

[59] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).

[60] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, J. Comput. Phys. **23**, 327 (1977).

[61] H. C. Andersen, J. Comput. Phys. **52**, 24 (1983).

[62] S. Miyamoto and P. A. Kollman, J. Comput. Chem. **13**, 952 (1992).

[63] S. Doudou, N. A. Burton, and R. H. Henchman, J. Chem. Theory Comput. **5**, 909 (2009).

[64] J. D. Weeks, D. Chandler, and H. C. Andersen, J. Chem. Phys. **54**, 5237 (1971).

[65] S. Khuttan, S. Azimi, J. Z. Wu, S. Dick, C. Wu, H. Xu, and E. Gallicchio, Phys. Chem. Chem. Phys. **25**, 24364 (2023).

[66] W. Jiang and B. Roux, J. Chem. Theory Comput. **6**, 2559 (2010).

[67] H. Oshima and Y. Sugita, J. Chem. Inf. Model. **62**, 2846 (2022).

[68] Y. Matsunaga, M. Kamiya, H. Oshima, J. Jung, S. Ito, and Y. Sugita, Biophys. Rev. **14**, 1503 (2022).

[69] Y. Matsubara, R. Okabe, R. Masayama, N. M. Watanabe, H. Umakoshi, K. Kasahara, and N. Matubayasi, J. Chem. Phys. **161**, 024108 (2024).

[70] R. C. Harris, N. Deng, R. M. Levy, R. Ishizuka, and N. Matubayasi, J. Comput. Chem. **38**, 1198 (2017).

[71] E. Schneiderman and A. M. Stalcup, J. Chromatogr. B: Biomed. Sci. Appl. **745**, 83 (2000).

[72] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko et al., Nature **596**, 583 (2021).

[73] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick et al., Nature **630**, 493 (2024).

[74] M. Muttenthaler, G. F. King, D. J. Adams, and P. F. Alewood, Nat. Rev. Drug Discovery **20**, 309 (2021).

[75] P. Csermely, R. Palotai, and R. Nussinov, Trends Biochem. Sci. **35**, 539 (2010).