Molecular Theory of Solvation Thermodynamics and Kinetics

Kazuya Okita

MARCH 2025

Molecular Theory of Solvation Thermodynamics and Kinetics

A dissertation submitted to
THE GRADUATE SCHOOL OF ENGINEERING SCIENCE
OSAKA UNIVERSITY
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY IN SCIENCE

by

Kazuya Okita

MARCH 2025

Abstract

Various dynamic phenomena occur in solution systems, which play an important role in chemistry. Solution systems have both a degree of ordered structure and thermal fluctuations, and these properties give rise to a wide variety of functions. To understand the dynamics of the solution, we have developed a new formulation based on the theory of statistical mechanics and molecular simulation.

Chapter 1 gives a general introduction to the theoretical backgrounds of the dynamics in the solution system.

Chapter 2 presents the new diffusion equation theory in the energy representation (ER) solution theory. First, the exact framework based on the generalized Langevin equation (GLE) formula is formulated, incorporating the ER theory. The derived equation (ER-GLE) describes the time evolution of the distribution functions on the energy coordinate. By introducing systematic approximations, such as the overdamped limit, the new diffusion equation for the molecular liquids in the ER theory (ERSV equation) is obtained. The present theory is applied to the solvation dynamics of water induced by the photoexcitation of benzonitrile. The new theory was able to predict the long-time behavior of the relaxation of the solvation structure, which is in agreement with that obtained by the molecular dynamics simulation. The analysis based on the new theory indicates that the collective motion of water is important during the relaxation process.

In Chapter 3, we applied the ERSV equation to the solvation dynamics of 6-propionyl2-dimethylamino naphthalene (Prodan) in water and alcohol solvents. For all solvents, the ERSV equation successfully predicts the relaxation timescale on the long timescale. We found that the relaxation timescale is linearly correlated with the inverse of the translational diffusion coefficient for the alcohol solvents. For water, this linear relationship is broken, indicating the difference in the importance of collective motion between water and alcohol solvents.

Chapter 4 provides the new computational scheme for calculating the free energy change of the molecular association based on the ER theory. The rigorous equation of the standard free energy change caused by the molecular association is derived. This equation consists of three terms, the conditional solvation free energy in the bound state, the solvation free energy in the bulk state, and the standard state correction term. The first two of these terms can be calculated based on the framework of the ER theory. The comparison of the results obtained by our new method and the conventional method showed the validity of our new approach.

Finally, the general conclusion is shown in Chapter 5.

Contents

Chapter 1

General Introduction

1.1 Integral equation theory for molecular liquids: Static theory

Solution systems are the common field where various chemical dynamics occur. [1–3] Since the solution systems are always under thermal fluctuation, the statistical mechanics treatments are essential. As such a treatment, the integral equation theory has long been used to understand the properties of solutions. The Ornstein-Zernike (OZ) relation is one of the most fundamental equations, which determines the relationship between the total correlation function and the direct correlation function. Because the OZ equation contains two unknown functions, the additional equation to close the OZ equation, known as the closure equation is required. The hyper-netted chain (HNC) closure and the Percus-Yevick (PY) closure are commonly used as the closure equation, and they can be derived from the functional expansions. Based on the OZ equation and the closure equation, the radial distribution function (RDF) between monoatomic molecules or its Fourier transform, the static structure factor can be obtained. Thanks to its analytical nature, the OZ equation is free from statistical errors, and its computational cost is much lower than molecular simulations, such molecular dynamics simulations and Monte Carlo simulations.

The extensions of the OZ equation for the molecular liquids are also often used in the solution chemistry. One of the commonly used theories for molecular liquids is the reference interaction site model (RISM) theory. [4–7] In the RISM theory, molecules are considered to be composed of atomic sites, and the site-site Ornstein-Zernike (SSOZ) equation can be derived by averaging the orientation of the molecules. The SSOZ equation and the closure equation give the RDFs between atomic sites, and the properties of the solution is discussed through the set of the RDFs. The extension of the RISM theory for the three-dimensional distribution functions (SDFs) is known as the 3D-RISM theory. [8,9] Because the SDFs are more intuitive and contain more information than RDFs, the 3D-RISM theory has been used to analyze the solvation structure of the biomolecules. [10, 11] The solvation free energy, which is one of the most important thermodynamic variables and very computationally expensive in MD simulations, can also be expressed analytically as a function of the correlation functions. Hence, once the correlation functions are calculated through the RISM theory, the solvation free energy is also easily obtained without any statistical error.

The hybrid approach of RISM theory and quantum chemistry, such as the RISM-SCF method and similar methods, has also been developed to incorporate the solvent effect

into the quantum chemical calculation. [12–16] It should be noted that these methods were originally derived with the intuition and reformulated on the basis of the variational principles. [14] The RISM-based methods incorporating the polarization of the solvent [17–20] have been also proposed based on the charge response kernel (CRK), which is the response function of the point charge on the atomic site against the electrostatic potential and determined by the quantum chemistry calculation. [21]

1.2 Integral equation theory for molecular liquids: Dynamics theory

This section introduces the theoretical approaches to describe the dynamical phenomena in terms of the integral equation theory. [1, 6, 22, 23] One of the representative approaches is the transport equation, such as the Smoluchowski equation. [24] The Smoluchowski equation is the extension of the diffusion equation to include the effect of the free energy surface. In the Smoluchowski equation, the time evolution of the distribution function is characterized by the diffusive motion and the drift motion due to the free energy profile. The Smoluchowski-Vlasov (SV) equation is the extension of the Smoluchowski equation to include the collective motion, using the direct correlation function in the OZ equation as the effective interaction between particles. [25, 26] The logarithms of the stationary solutions of the Smoluchowski equation and the SV equation must be proportional to the free energy profile. This is an important feature of these two equations, and the diffusion equation has no such feature.

The generalized Langevin equation (GLE) plays a significant role in this field. It is rigorously derived using the Zwanzig-Mori projection operator method and describes the time evolution of the dynamical variable such as particle velocity, distribution and the corresponding current. [27–29] The time evolution of the dynamical variable is characterized by three terms, the collective frequency term, the memory term and the fluctuating force term. The second term corresponds to the convolution integral of the dynamical variable and the memory function, which is considered to be the generalized friction coefficient. As a feature of the convolution integral, the second term is history-dependent, in other words, all information about the time evolution of the dynamical variable is required. This is the decisive difference between the GLE and the (ordinal) Langevin equation, which is written only by the instantaneous variables. The GLE for the density and the current fields, which is applicable to inhomogeneous systems with external fields was also formulated. [30] Although it requires the two-point density-density correlation function, which is computationally expensive, the introduction of the approximation based on the HNC closure makes it possible to use the one-body correlation function instead of the many-body function.

Although the derivation of the GLE itself is rigorous, it is impossible to know the exact form of the memory function. Therefore, some approximations for the memory function are necessary. The simplest approximation is the overdamped limit, which assumes that the memory function decays immediately. This approximation is valid when the timescale of the dynamical variable and other variables is largely different. It should be noted that the Smoluchowski equation and the SV equation are derived from the GLE using this approximation. When the timescale gap is not large enough, other approximations, such as the viscoelastic model and the mode-coupling theory (MCT) are used. In the viscoelastic model, the memory function is approximated by the single exponential function with the

wavenumber-dependent time constant. This wavenumber-dependent time constant can be determined to satisfy the asymptotic behavior of the dynamics structure factor. [31, 32] The MCT introduces the new projection operator to extract the slower motion, and the memory function is divided into two components, the fast and the slow part. Employing the decoupling approximation, the MCT gives the closed formula for the slower part of the memory function. [33]

Treatments of the dynamics of molecular liquids were achieved by the site-site Smoluchowski-Vlasov (SSSV) equation and the site-site generalized Langevin equation (SSGLE). [34–37] These equations give the time evolution of the site-site van Hove correlation function or its Fourier transform, the site-site intermediate scattering function. The SSGLE is derived through the Zwanzig-Mori projection operator method and the RISM theory, and the overdamped limit approximation yields the SSSV equation. As in the case of monoatomic molecular liquids, the viscoelastic model and the mode-coupling theory for the memory function are proposed. [36, 38–42] The three-dimensional version of the SSSV equation (3D-SSSV) was developed and applied to the diffusion process in the electrolyte solutions. [43] The site-site description of the Smoluchowski equation under the external field was also formulated and used to analyze the diffusion-controlled reaction of the diatomic molecule. [44] The surrogate Hamiltonian theory enabled them to treat the solvation dynamics, which is the response of the solvation structure after the photoexcitation of the solute. [45–55] This theory can express the solvation time correlation function (STCF) through the intermediate scattering function and the site-site direct correlation functions between the solute and the solvent. It is noteworthy that the site-site intermediate scattering function in the pure solvent system can be used instead of that in the solution system and the site-site direct correlation function is regarded as the effective solute-solvent interaction. Such a treatment is justified by the linear response theory, in other words, it is reasonable when the solute-solvent interaction is sufficiently weak and the dynamics of the solvent is not changed significantly by the solute.

Another promising approach is the time-dependent density functional theory (TDDFT) [56, 57] in terms of the site-site description. [58] It describes the dynamics of the solvation structure by the functional derivative of the free energy with the distribution. Its major advantage is that it can treat the nonlinear response, whereas the surrogate theory is a linear response theory. The combination of the TDDFT and the 3D-RISM theory with the polarizable solvent has also been developed and applied to the solvation dynamics. [59]

1.3 Energy representation solution theory

This section gives a brief introduction to the Energy Representation (ER) solution theory. [60–65] Whereas many other IE theories use geometric quantities, such as positions and angles, the ER theory employs the pair interaction energy between the solute and the solvent as a argument of the distribution function. In this theory, the full set of the coordinates of position and orientation with intramolecular degrees of freedom is projected onto the solute-solvent pair interaction energy, which is used as a coordinate, namely the energy coordinate. This process can be regarded as a kind of dimensional reduction. In contrast to the RISM theory, where the molecules are considered to be composed of atomic sites, the ER theory takes each of the solute and solvent molecules as a single unit, making it straightforward to handle the molecules with the intramolecular degree of freedom. It should be noted that the potential energy function used for the projec-

tion does not necessarily correspond to the true solute-solvent interaction that appears in the Hamiltonian of the system. To clarify this difference, the former is referred to as the defining potential. The Kirkwood charging formula, the well-known expression for the solvation free energy, can also be written in the energy representation. In this formula, the solvation free energy is expressed as the functional of the solute-solvent interaction potential function $u_\lambda^{\mathrm{e}}(\varepsilon)$ and the energy distribution function $\rho_\lambda^{\mathrm{e}}(\varepsilon)$, where λ means the coupling parameter. The one-to-one correspondence between the solute-solvent interaction potential and the energy distribution function is proved under the condition that the solute-solvent interaction potential can be written as a function of the defining potential. [60] This fact means that the functional of the solute-solvent interaction can be treated as a functional of the energy distribution. Thanks to this fact, the solvation free energy can be seen as a functional of $\rho_\lambda^{\mathrm{e}}(\varepsilon)$ only. Although this formula is exact, it requires $\rho_\lambda^{\mathrm{e}}(\varepsilon)$ for all λ, which is almost impossible to calculate. To avoid this problem, the HNC-type and PY-type approximations are introduced, and the approximated the solvation free energy functional is derived. [61] As a result, only the functions at the end-points of the charging process ($\lambda = 0,\ 1$) are needed to calculate the solvation free energy, leading to the drastic reduction of the computational costs.

One of the significant points of the ER theory is its applicability to functional molecular systems, such as the micelles, [66,67] the lipid membranes, [68,69] the polymers, [70–75] the biomolecules, [76–80] the ionic liquids, [81] and the solid-liquid interface. [82,83] This broad applicability is achieved by generalizing the concept of solvation. In this concept, the molecule which comes into the system is referred to as the solute, and all other molecules already present in the system are regarded as the mixed solvent. For example, in the case of the dissolution process of the peptide to the lipid membrane, the lipid is considered to be one of the components of the mixed solvents. This treatment makes it straightforward to apply the ER theory to systems with nano- or macroscopic inhomogeneity. The combination with the quantum-mechanical/molecular-mechanics (QM/MM) calculation is also realized, [84–87] and the reduction process is viewed as the solvation of the electron in the mixed solvent composed of the oxidant and surrounding molecules.

1.4 Objective and focus of this dissertation

As shown in section 1.3, the ER solution theory is a powerful framework for studying the thermodynamics of functional molecular systems. Recently, the diffusion-influenced bimolecular reaction theory, which uses the energy coordinate as the reaction coordinate, has been proposed and found to be useful in describing the host-guest binding processes. [88,89] In this approach, the intermediate and bound states are defined by the host-guest interaction energy, and the binding process is described as a motion on the energy coordinate. It is shown that this method yields the rate constants in agreement with the experimental results. This suggests that the ER solution theory is effective not only in free energy analyses but also in kinetic theories. For many chemical processes, both thermodynamic and kinetic properties play an important role. For example, in the elucidation of host-guest binding processes, both the thermodynamic stability of the bound state and the rate constants of association and dissociation are important. Therefore, the development of theoretical frameworks to describe the thermodynamic and kinetic properties of chemical processes occurring in functional molecular systems is crucial for microscopic insights into these phenomena.

In this dissertation, we aim to understand the thermodynamics and kinetics of the

solvation based on the theory of solutions and the nonequilibrium statistical mechanics. The diffusion equation theory plays central role in elucidating the molecular motions in solutions. We formulate the new diffusion equation theory to analyze the dynamic phenomena of the functional molecules in terms of the energy representation solution theory and achieve the systematic analysis based on the new theory. We also develop the new theoretical framework for the calculation of the binding free energy based on the energy representation solution theory using the generalized concept of the solvation.

1.5 Organization of this dissertation

In Chapter 2, the rigorous theoretical framework of the energy-represented dynamics is shown. Introducing the systematic approximations, such as the coarse graining of the time, leads to the new diffusion equation theory in the energy representation. Application to the relaxation process of the hydration structure around the benzonitrile shows the validity of the new theory in the long-time region, and the systematic analyses based on the new theory are conducted. In Chapter 3, we applied the new theory to the solvation dynamics of Prodan in various solvents. For all solution systems, our theory predict the long-timescale dynamics successfully. The comparison among the different solvent species indicates that the relaxation process of water has a different feature from that of the alcohol solvents. Chapter 4 provides the new framework to compute the binding free energy based on the theory of solutions. It gives the rigorous relationship between the binding free energy and the conditional solvation free energy. The new framework is shown to be more robust and computationally efficient than the existing rigorous one.

Chapter 2

Diffusion theory of molecular liquids in the energy representation and application to solvation dynamics

2.1 Introduction

Molecular motions in a condensed phase are essential in various dynamic processes. Such dynamics are often characterized by sophisticated spectroscopy techniques. [90, 91] For instance, the time-resolved fluorescence spectroscopy can be useful to probe the solvation dynamics with a certain dye (solute), i.e., a solvent relaxation around a solute induced by the photoexcitation. [3,92,93] Recently, this spectroscopy has been extensively utilized to investigate the dynamics in heterogeneous environments such as the surface of biological lipid bilayers. [94,95] In its measurement, however, since the information on the dynamics is contracted to the time development of the fluorescence wavenumber, it is challenging to get the atomistic insights directly. Thus, developing a theoretical fundament to analyze the dynamics in the heterogeneous environments from the atomistic points of view would be useful for further understanding.

A theoretical approach to describe the dynamics in solution is based on transport equations such as Smoluchowski equation, in which the dynamics is characterized by the diffusive motion and drift motion governed by the free energy gradient. [96–98] Smoluchowski-Vlasov (SV) equation is an extension of the original Smoluchowski equation to describe the collective motion of solvents, which is a crucial part of solvation dynamics. [25,26,99,100] The molecular mode-coupling theories were also proposed to describe the dynamics of supercooled liquids. [101–103] However, these equations are useful only for simple liquids such as monoatomic and diatomic liquids, because the difficulties in handling both the translational and rotational motions (six-dimensional in total) arise in the case of molecular liquids.

A route to describe such motions while avoiding the difficulty is to introduce the interaction site representation of statistical mechanics of solution. [6, 7] In its representation, the static liquid structure is expressed with a set of radial distribution functions (RDFs) between the atomic sites in molecules, realizing the effective treatment of a molecular shape and orientation. The theoretical framework for the RDFs is provided by reference interaction site model (RISM) theory, an integral equation (IE) theory of molecular liquids, [1, 4, 104, 105] and its extension towards the dynamic processes is achieved with Zwanzig-Mori projection operator theory. [23, 27, 28, 34, 106] According to this approach,

site-site Smoluchowski-Vlasov (SSSV) equation, [34] and several types of site-site generalized Langevin equations (SSGLE) have been derived [35,40,42,107] and applied to various dynamic processes such as collective excitation, [38], pressure dependence of solvent diffusivity, [108–110], electrical conduction, [111, 112] ultrasound adsorption, [113, 114] and solvation dynamics. [45, 47–49, 52, 53, 55] Upgrading the description of the dynamics from the radial coordinate to the spatial coordinate using three-dimensional RISM (3D-RISM) theory [8, 9, 115] and the time-dependent density functional theory (TDDFT) [57, 93] is a recent progress in the IE-based dynamics theories. [43, 58, 59] Thanks to the analytical nature of the theories, the developed theories are free from the sampling problems which appear in the molecular dynamics (MD) simulations, and the computational cost is low. However, the applicability of the theories is still limited to simple polyatomic solvents due to the orientational averaging imposed on the solvent molecules.

Another progress in the field of the statistical mechanics of solution is the development of the energy representation (ER) theory. [60–62, 116] Instead of the spatial coordinates, ER theory employs the interaction energy between a solute and solvent molecule as a coordinate, namely energy coordinate, for treating the relative position and orientation for molecules. It corresponds to the reduction of the dimensions required for the atomistic description from six-dimensions (6D) to one-dimension (1D). As well as in the case of the spatial coordinate, the Ornstein-Zernike (OZ) equation and the relationship between the density correlations and the solvation free energy can be constructed on the energy coordinate. In ER theory, once we evaluate such density correlations through the MD simulations, the accurate calculation of the solvation free energy of a solute is realized. [61, 64, 117] Thanks to the reduction of the dimensions, the calculation of the density correlations on the energy coordinate is much easier than those on the 6D-space. A remarkable point of ER theory is the applicability to heterogeneous systems. [64, 118] In the case of the binding of small solutes in a lipid membrane, for example, the lipid can be regarded as part of the mixed solvents. [68, 69] Very recently, a diffusion-influenced reaction theory incorporating the energy coordinate was proposed to quantify the rate constants of host-guest binding processes, and its application to the binding processes of aspirin and 1-butanol to β-cyclodextrin yielded the rate constants consistent with the experimental results. [88] Accordingly, the energy coordinate would be also useful for describing dynamic processes at the atomistic scale. Furthermore, since the mathematical form of the equations derived in the framework of ER theory is almost parallel to that for the IE theory on the spatial coordinate of full 6D, position and orientation, such as the OZ equation, the ideas of the existing IE-based dynamics theories mentioned in the last paragraph could be imported into ER framework.

In the present study, we develop a theoretical framework for elucidating the dynamics in condensed phases by utilizing the energy coordinate. According to Zwanzig-Mori projection operator theory and ER theory, an energy representation of GLE is derived. Introducing the systematic approximations such as the overdamped limit yields the energy representation of SV (ERSV) equation by generalizing the scheme of deriving the Smoluchowski equation for molecular liquids. [44] The ERSV equation provides the solute-solvent dynamics as outputs from the static solute-solvent distribution and the diffusion coefficient as inputs. Those inputs can be readily obtained in molecular simulations, and ERSV offers a scheme to approaching the (long-time) dynamics by combining MD and GLE. We show that the solvation dynamics can be described by ERSV equation in conjunction with the linear response theory. The present theory is applied to the solvation dynamics of solvent water induced by the photoexcitation of benzonitrile.

2.2 Theory

2.2.1 Ornstein-Zernike equation in the energy representation

In this subsection, we describe a brief summary of the Ornstein-Zernike equation formalism in the energy representation. [60–62] Let us consider a system containing a solute molecule u immersed in solvent. We regard the solute u as the source of an external field and the position is defined as the origin of the system. We denote the full-phase space coordinate of ith molecule of the solvent v with respect to u as $\boldsymbol{x}_{v,i}$, which refers collectively to the coordinates of the center of mass (CoM) and the orientation of the ith solvent molecule. If the solvent molecule is flexible, the intramolecular degrees of freedom are also incorporated in $\boldsymbol{x}_{v,i}$. The instantaneous distribution of solvent v in the energy representation is defined as

$$\rho_v(\varepsilon) \equiv \sum_{i \in v} \delta(\varepsilon - \varepsilon_v(\boldsymbol{x}_{v,i})), \tag{2.1}$$

and its fluctuation around the equilibrium average is

$$\delta\rho_v(\varepsilon) \equiv \rho_v(\varepsilon) - \langle \rho_v(\varepsilon) \rangle \,, \tag{2.2}$$

where $\langle \cdots \rangle$ stands for the ensemble average, and ε_v is the defining potential between solute u and solvent v, which defines the energy coordinate, ε. Note that the defining potential, ε_v, is not limited to the solute-solvent interaction, and can be defined arbitrarily depending on the phenomena of interest. It should be also noted that either the solute or solvent species are not assumed to be rigid. Moreover, the solute molecule is actually not necessary to be located at fixed position and/or orientation. The distribution of Eq. (2.1) can be constructed at any snapshot configuration of the solute and solvent by referring only to the value of the defining potential $\varepsilon_v(\boldsymbol{x}_{v,i})$. The density-density correlation function between solvent species v and w on the energy coordinate $\chi_{vw}(\varepsilon, \eta)$ is written as

$$\begin{aligned} \chi_{vw}(\varepsilon, \eta) &\equiv \langle \delta\rho_v(\varepsilon) \delta\rho_w(\eta) \rangle \\ &= \langle \rho_v(\varepsilon) \rangle \langle \rho_w(\eta) \rangle h_{vw}(\varepsilon, \eta) + \delta_{vw} \delta(\varepsilon - \eta) \langle \rho_v(\varepsilon) \rangle \,, \end{aligned} \tag{2.3}$$

where $h_{vw}(\varepsilon, \eta)$ is the total correlation function defined as

$$h_{vw}(\varepsilon, \eta) \equiv \frac{\langle \rho_v(\varepsilon) \rho_w(\eta) \rangle - \delta_{vw} \delta(\varepsilon - \eta) \langle \rho_v(\varepsilon) \rangle}{\langle \rho_v(\varepsilon) \rangle \langle \rho_w(\eta) \rangle} - 1. \tag{2.4}$$

The Ornstein-Zernike equation in the energy representation is given as follows:

$$h_{vw}(\varepsilon, \eta) = c_{vw}(\varepsilon, \eta) + \sum_{v'} \int \mathrm{d}\zeta \; c_{vv'}(\varepsilon, \zeta) \langle \rho_{v'}(\zeta) \rangle h_{v'w}(\zeta, \eta). \tag{2.5}$$

Here, $c_{vw}(\varepsilon, \eta)$ is the direct correlation function, which can be related with the inverse of $\chi_{vw}(\varepsilon, \eta)$ as

$$\chi^{-1}_{vw}(\varepsilon, \eta) = \frac{\delta_{vw}}{\langle \rho_v(\varepsilon) \rangle} \delta(\varepsilon - \eta) - c_{vw}(\varepsilon, \eta). \tag{2.6}$$

Note that $\chi^{-1}_{vw}(\varepsilon, \eta)$ is a key quantity for constructing the dynamics theory based on the generalized Langevin equation (GLE) formalism. The numerical computation of $\chi^{-1}_{vw}(\varepsilon, \eta)$

can be performed from $\langle\rho(\varepsilon)\rangle_v$ and $\chi_{vw}(\varepsilon,\eta)$ with the aid of Moore-Penrose pseudo-inverse method when the ensemble is NVT or NPT and the number of particles in the system is invariant. The inverse matrix obtained from this method contains an additive constant. This constant is determined so as to assure the intensive property of solvation free energy in the ER theory. [62,116] As shown in the next subsection, since the derivative of $\chi^{-1}_{vw}(\varepsilon,\eta)$ affects the dynamics, the additive constant does not contribute to the time development of the system. [62, 116]

2.2.2 Diffusion equation in the energy representation

We derive a diffusion equation in the energy representation. The starting point of the derivation is the GLE formalism based on Zwanzig-Mori projection operator method. We choose a set of the time-dependent density fluctuation, $\{\delta\rho_v(\varepsilon,t)\}$ defined below, as the dynamical variable $\boldsymbol{A}(\varepsilon,t)$

$$\boldsymbol{A}(\varepsilon,t) \equiv \begin{pmatrix} \delta\rho_1(\varepsilon,t) \\ \delta\rho_2(\varepsilon,t) \\ \vdots \end{pmatrix}, \tag{2.7}$$

$$\delta\rho_v(\varepsilon,t) \equiv \sum_i \delta(\varepsilon - \varepsilon_v(\boldsymbol{x}_{v,i}(t))) - \langle\rho_v(\varepsilon)\rangle\,. \tag{2.8}$$

Hereafter, the functions which do not contain time t as their arguments represent the values at $t=0$. Performing the time derivative of $\delta\rho_v(\varepsilon,t)$ yields the continuity equation as

$$\frac{\partial\delta\rho_v(\varepsilon,t)}{\partial t} = -\frac{\partial j_v(\varepsilon,t)}{\partial\varepsilon}. \tag{2.9}$$

where $j_v(\varepsilon,t)$ is the current field on the energy coordinate given by

$$j_v(\varepsilon,t) = \sum_i \frac{\mathrm{d}\varepsilon_v(\boldsymbol{x}_{v,i}(t))}{\mathrm{d}t}\delta(\varepsilon - \varepsilon_v(\boldsymbol{x}_{v,i}(t))). \tag{2.10}$$

Let us introduce an operator $\mathcal{P}$ projecting dynamical variables onto the subspace spanned by $\boldsymbol{A}(\varepsilon)$ defined as

$$\mathcal{P}\boldsymbol{X}(\varepsilon,t) \equiv \int \mathrm{d}\eta\mathrm{d}\zeta\ \left\langle \boldsymbol{X}(\varepsilon,t)\boldsymbol{A}^\dagger(\eta)\right\rangle \left\langle \boldsymbol{A}(\eta)\boldsymbol{A}^\dagger(\zeta)\right\rangle^{-1} \boldsymbol{A}(\zeta), \tag{2.11}$$

where $\dagger$ means the adjoint. According to the standard procedure of the Zwanzig-Mori projection operator method, the following GLE in the energy coordinate is derived.

$$\frac{\partial\delta\rho_v(\varepsilon,t)}{\partial t} = -\sum_w \int_0^t \mathrm{d}\tau \int \mathrm{d}\eta\ K_{vw}(\varepsilon,\eta,\tau)\delta\rho_w(\eta,t-\tau) + F_v(\varepsilon,t). \tag{2.12}$$

Here, $K_{vw}(\varepsilon,\eta,t)$ and $F_v(\varepsilon,t)$ are the memory function and the fluctuating force, respectively. $F_v(\varepsilon,t)$ is defined as

$$F_v(\varepsilon,t) \equiv \exp\left[\mathcal{Q}i\mathcal{L}t\right]\mathcal{Q}i\mathcal{L}\delta\rho_v(\varepsilon) = -\exp\left[\mathcal{Q}i\mathcal{L}t\right]\frac{\partial j_v(\varepsilon)}{\partial\varepsilon}. \tag{2.13}$$

where $\mathcal{L}$ is the Liouville operator of the system and $\mathcal{Q}$ is the orthogonal operator defined as $\mathcal{Q} \equiv 1 - \mathcal{P}$. The memory function, $K_{vw}(\varepsilon, \eta, t)$, is represented with $F_v(\varepsilon, t)$ as follows

$$\begin{aligned} K_{vw}(\varepsilon, \eta, t) &\equiv \sum_{v'} \int \mathrm{d}\zeta \ \langle F_v(\varepsilon, t) F_{v'}(\zeta) \rangle \, \chi^{-1}_{v'w}(\zeta, \eta) \\ &= \sum_{v'} \int \mathrm{d}\zeta \ \left\langle \left\{ \exp\left[\mathcal{Q} i \mathcal{L} t\right] \frac{\partial j_v(\varepsilon)}{\partial \varepsilon} \right\} \frac{\partial j_{v'}(\zeta)}{\partial \zeta} \right\rangle \chi^{-1}_{v'w}(\zeta, \eta). \end{aligned} \tag{2.14}$$

A route for deriving the diffusion equation from the projection operator method is to replace the orthogonal time evolution operator $\exp\left[\mathcal{Q} i \mathcal{L} t\right]$ involved in Eq. (2.14) by the usual time evolution operator $\exp\left[i \mathcal{L} t\right]$, [44, 119, 120] corresponding to

$$F_v(\varepsilon, t) \simeq -\frac{\partial j_v(\varepsilon, t)}{\partial \varepsilon}, \tag{2.15}$$

where we have used Eq. (2.13) with the relationship given by $\exp\left[i \mathcal{L} t\right] A = A(t)$ that holds for arbitrary functions over the phase space. This approximation leads to

$$K_{vw}(\varepsilon, \eta, t) = \sum_{v'} \int \mathrm{d}\zeta \ \frac{\partial^2 \left\langle j_v(\varepsilon, t) j_{v'}(\zeta) \right\rangle}{\partial \varepsilon \, \partial \zeta} \chi^{-1}_{v'w}(\zeta, \eta). \tag{2.16}$$

Similar to the conventional diffusion theory, we introduce the overdamped approximation expressed in the energy representation as

$$\left\langle j_v(\varepsilon, t) j_w(\eta) \right\rangle \simeq 2 \delta_{vw} \delta(\varepsilon - \eta) \delta(t) D^{\mathrm{e}}_v(\varepsilon) \left\langle \rho_v(\varepsilon) \right\rangle. \tag{2.17}$$

where $D^{\mathrm{e}}_v(\varepsilon)$ is the diffusion coefficient in the energy coordinate defined as

$$\begin{aligned} D^{\mathrm{e}}_v(\varepsilon) &\equiv \frac{1}{\langle \rho_v(\varepsilon) \rangle} \int_0^\infty \mathrm{d}t \int \mathrm{d}\eta \ \langle j_v(\varepsilon, t) j_v(\eta) \rangle \\ &= \frac{1}{\langle \rho_v(\varepsilon) \rangle} \int_0^\infty \mathrm{d}t \ \left\langle j_v(\varepsilon, t) \dot{E}_v \right\rangle \end{aligned} \tag{2.18}$$

where $E_v(t)$ is the total interaction energy between the solute and the solvent v defined as

$$E_v(t) = \sum_i \varepsilon_v(\boldsymbol{x}_{v,i}(t)). \tag{2.19}$$

Using the above approximation, Eq. (2.16) can be rewritten as

$$K_{vw}(\varepsilon, \eta, t) = -2\delta(t) \frac{\partial}{\partial \varepsilon} \left[D^{\mathrm{e}}_v \left\langle \rho_v(\varepsilon) \right\rangle \frac{\partial \chi^{-1}_{vw}(\varepsilon, \eta)}{\partial \varepsilon} \right]. \tag{2.20}$$

Substituting Eqs. (2.6) and (2.20) into Eq. (2.12) yields the following diffusion equation as

$$\begin{aligned} \frac{\partial \delta\rho_v(\varepsilon, t)}{\partial t} = \frac{\partial}{\partial \varepsilon} \Bigg[& D^{\mathrm{e}}_v(\varepsilon) \frac{\partial \delta\rho_v(\varepsilon, t)}{\partial \varepsilon} - D^{\mathrm{e}}_v(\varepsilon) \frac{\mathrm{d} \ln \left\langle \rho_v(\varepsilon) \right\rangle}{\mathrm{d}\varepsilon} \delta\rho_v(\varepsilon, t) \\ & - D^{\mathrm{e}}_v(\varepsilon) \left\langle \rho_v(\varepsilon) \right\rangle \sum_w \int \mathrm{d}\eta \ \frac{\partial c_{vw}(\varepsilon, \eta)}{\partial \varepsilon} \delta\rho_w(\eta, t) \Bigg] + F_v(\varepsilon, t). \end{aligned} \tag{2.21}$$

The first and second terms respectively indicate the diffusive motions caused by the energy distribution gradient and drift motions by the free energy gradient over the energy coordinate, respectively. The third term describes the collective motions of the solvents. Since the mathematical form of Eq. (2.21) is parallel to that of the Smoluchowski-Vlasov (SV) equation on the spatial coordinate proposed by Calef *et al.*, [25, 26] this equation can be regarded as the energy-represented SV (ERSV) equation. The diffusive and drift terms in Eq. (2.21) could be combined as follows

$$\frac{\partial \delta\rho_v(\varepsilon,t)}{\partial t} = \frac{\partial}{\partial \varepsilon}\left[D_v^{\mathrm{e}}(\varepsilon)\left\langle \rho_v(\varepsilon)\right\rangle \frac{\partial}{\partial \varepsilon}\left(\frac{\delta\rho_v(\varepsilon,t)}{\left\langle \rho_v(\varepsilon)\right\rangle}\right) \right. \\ \left. - D_v^{\mathrm{e}}(\varepsilon)\left\langle \rho_v(\varepsilon)\right\rangle \sum_w \int \mathrm{d}\eta \, \frac{\partial c_{vw}(\varepsilon,\eta)}{\partial \varepsilon}\delta\rho_w(\eta,t)\right] + F_v(\varepsilon,t), \quad (2.22)$$

indicating that the time evolution of the energy distribution is governed by $D_v^{\mathrm{e}}(\varepsilon)\left\langle \rho_v(\varepsilon)\right\rangle$ instead of $D_v^{\mathrm{e}}(\varepsilon)$. If we neglect the third term, Eq. (2.21) is reduced to the energy-represented Smoluchowski (ERS) equation written as

$$\frac{\partial \delta\rho_v(\varepsilon,t)}{\partial t} = \frac{\partial}{\partial \varepsilon}\left[D_v^{\mathrm{e}}(\varepsilon)\frac{\partial \delta\rho_v(\varepsilon,t)}{\partial \varepsilon} - D_v^{\mathrm{e}}(\varepsilon)\frac{\mathrm{d}\ln\left\langle \rho_v(\varepsilon)\right\rangle}{\mathrm{d}\varepsilon}\delta\rho_v(\varepsilon,t)\right] + F_v(\varepsilon,t). \quad (2.23)$$

Similar to the case of the spatial coordinate, [23] the van Hove correlation function in the energy representation can be defined as follows:

$$G_{vw}(\varepsilon,\eta,t) \equiv \left\langle \delta\rho_v(\varepsilon,t)\delta\rho_w(\eta)\right\rangle . \quad (2.24)$$

From Eq. (2.21) and the orthogonality between $F_v(\varepsilon,t)$ and $\delta\rho(\varepsilon)$, we can obtain the differential equation for $G_{vw}(\varepsilon,\eta,t)$ as

$$\frac{\partial G_{vw}(\varepsilon,\eta,t)}{\partial t} = \frac{\partial}{\partial \varepsilon}\left[D_v^{\mathrm{e}}(\varepsilon)\frac{\partial G_{vw}(\varepsilon,\eta,t)}{\partial \varepsilon} - D_v^{\mathrm{e}}(\varepsilon)\frac{\mathrm{d}\ln\left\langle \rho_v(\varepsilon)\right\rangle}{\mathrm{d}\varepsilon}G_{vw}(\varepsilon,\eta,t) \right. \\ \left. - D_v^{\mathrm{e}}(\varepsilon)\left\langle \rho_v(\varepsilon)\right\rangle \sum_{v'} \int \mathrm{d}\zeta \, \frac{\partial c_{vv'}(\varepsilon,\zeta)}{\partial \varepsilon}G_{v'w}(\zeta,\eta,t)\right] . \quad (2.25)$$

2.2.3 Approximate expression of the diffusion coefficient $D_v^{\mathrm{e}}(\varepsilon)$

The explicit formula of the diffusion coefficient in the energy coordinate is already available (Eq. (2.18)), which requires the energy-time correlation function, $\langle j_v(\varepsilon,t)j_v(\eta)\rangle$. In this subsection, we derive a tractable expression of $D_v^{\mathrm{e}}(\varepsilon)$. Substituting Eq. (2.10) into Eq. (2.18) yields

$$D_v^{\mathrm{e}}(\varepsilon) = \frac{1}{\left\langle \rho_v(\varepsilon)\right\rangle}\sum_{i,j}\int_0^\infty \mathrm{d}t \ \left\langle \dot{\varepsilon}_v(\boldsymbol{x}_{v,i}(t))\dot{\varepsilon}_v(\boldsymbol{x}_{v,j})\delta(\varepsilon - \varepsilon_v(\boldsymbol{x}_{v,i}(t)))\right\rangle . \quad (2.26)$$

We assume that the solute molecule is located at fixed position and orientation. Then, by expressing the full coordinate $\boldsymbol{x}_{v,i}(t)$ as a set of the positions of atoms μ in the molecule, $\boldsymbol{r}_{i,\mu}$, $\dot{\varepsilon}_v(\boldsymbol{x}_{v,i}(t))$ can be written as

$$\dot{\varepsilon}_v(\boldsymbol{x}_{v,i}(t)) = -\sum_{\mu\in i}\boldsymbol{v}_{i,\mu}(t)\cdot\boldsymbol{f}_{i,\mu}(t) = -\sum_{\mu\in i}\sum_{\alpha=x,y,z} v_{i,\mu}^{\alpha}(t)f_{i,\mu}^{\alpha}(t). \quad (2.27)$$

Here, $v_{i,\mu}^{\alpha}(t)$ is the velocity of atom μ in the ith molecule, and $f_{i,\mu}^{\alpha}(t)$ is the "force" corresponding to the defining potential, that is introduced as

$$\boldsymbol{f}_{i,\mu}(t) = -\frac{\partial \varepsilon_v(\boldsymbol{x}_{v,i}(t))}{\partial \boldsymbol{r}_{i,\mu}}. \tag{2.28}$$

If the defining potential ε_v is the same as the solute-solvent interaction, $\boldsymbol{f}_{i,\mu}(t)$ corresponds to the force acting on atom μ due to the solute-solvent interaction. In the next subsection, we treat the solvation dynamics and set ε_v to the difference in the solute-solvent interaction potential between the ground and excited states of the solute. Using Eq. (2.27), Eq. (2.26) is written as

$$D_v^{\mathrm{e}}(\varepsilon) = \sum_{i,j} \sum_{\mu\in i,\lambda\in j} \sum_{\alpha,\beta=x,y,z} \frac{1}{\langle \rho_v(\varepsilon)\rangle} \int_0^\infty \mathrm{d}t \ \left\langle v_{i,\mu}^{\alpha}(t) v_{j,\lambda}^{\beta} f_{i,\mu}^{\alpha}(t) f_{j,\lambda}^{\beta} \delta(\varepsilon - \varepsilon_v(\boldsymbol{x}_{v,i}(t)))\right\rangle . \tag{2.29}$$

By decoupling the velocity from the force and energy, Eq. (2.29) is written as

$$D_v^{\mathrm{e}}(\varepsilon) = \sum_{i,j} \sum_{\mu\in i,\lambda\in j} \sum_{\alpha,\beta=x,y,z} \frac{1}{\langle \rho_v(\varepsilon)\rangle} \int_0^\infty \mathrm{d}t \ \left\langle v_{i,\mu}^{\alpha}(t) v_{j,\lambda}^{\beta}\right\rangle \left\langle f_{i,\mu}^{\alpha}(t) f_{j,\lambda}^{\beta} \delta(\varepsilon - \varepsilon_v(\boldsymbol{x}_{v,i}(t)))\right\rangle , \tag{2.30}$$

Tu *et al.* revealed that the time integral of velocity correlation function between different atomic sites is equal to the integral for the CoM, [121] i.e.,

$$\int_0^\infty \mathrm{d}t \ \left\langle v_{i,\mu}^{\alpha}(t) v_{j,\lambda}^{\beta}\right\rangle = \int_0^\infty \mathrm{d}t \ \left\langle V_i^{\alpha}(t) V_j^{\beta}\right\rangle , \tag{2.31}$$

where $\boldsymbol{V}(t)$ is the CoM velocity of ith molecule. It is well known that this integral gives the translational diffusion coefficient of species v, D_v, if $i = j$ and $\alpha = \beta$. Then, we introduce the following overdamped approximation.

$$\left\langle v_{i,\mu}^{\alpha}(t) v_{j,\lambda}^{\beta}\right\rangle \simeq 2D_v \delta_{\alpha\beta}\delta_{ij}\delta(t). \tag{2.32}$$

It should be noted that a similar type of approximation is utilized for deriving SSSV equation [34]. Substitution of Eq. (2.32) into Eq. (2.30) gives

$$D_v^{\mathrm{e}}(\varepsilon) = \frac{1}{\langle \rho_v(\varepsilon)\rangle} D_v \sum_i \left\langle |\boldsymbol{f}_i^{\mathrm{g}}|^2 \delta(\varepsilon - \varepsilon_v(\boldsymbol{x}_{v,i}))\right\rangle , \tag{2.33}$$

where $\boldsymbol{f}_i^{\mathrm{g}}$ is the force acting on the CoM of ith molecule defined as $\boldsymbol{f}_i^{\mathrm{g}} \equiv \sum_\mu \boldsymbol{f}_{i,\mu}$. The scheme of computing $\boldsymbol{f}_i^{\mathrm{g}}$ is provided in Appendix A. As compared to Eq. (2.18), this formula does not require the computation of the energy-time correlation function, leading to the reduction of the computational cost for evaluating $D_v^{\mathrm{e}}(\varepsilon)$. On the other hand, since the spatial diffusivity is represented with bulk diffusion coefficient of solvents, Eq. (2.33) is valid only when the solute-solvent interactions are so weak that the diffusivity of the solvent is not changed drastically around the solute. In Eq. (2.33), the rotational diffusion does not appear explicitly due to the decoupling of velocity from the force and energy, and the overdamped approximation. It could cause an insufficient description about the rotational motion of the solvent which is important on the short timescale.

2.2.4 Application to solvation dynamics

In this subsection, we present a scheme for describing the solvation dynamics, a response of the solvation structure to a sudden change of the electronic structure of a solute due to the photoexcitation. Let us introduce the Hamiltonians of the system at the ground and excited states as $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. The difference of the Hamiltonians is defined as

$$\begin{aligned}\Delta\mathcal{H} &\equiv \mathcal{H}_1 - \mathcal{H}_0 \\ &= \sum_v \sum_{i\in v} \left\{ u_v^{\mathrm{e}}(\boldsymbol{x}_{v,i}) - u_v^{\mathrm{g}}(\boldsymbol{x}_{v,i}) \right\} + \Delta E_{\mathrm{solute}},\end{aligned} \tag{2.34}$$

where $u_v^{\mathrm{g}}(\boldsymbol{x}_{v,i})$ and $u_v^{\mathrm{e}}(\boldsymbol{x}_{v,i})$ are respectively the interaction potential between a solute and solvent v at the ground and excited states. $\Delta E_{\mathrm{solute}}$ is the change of the solute energy due to the excitation. Only the pairwise additivity of the difference of the solute-solvent interactions at the ground and excited states are assumed in Eq. (2.34), and the interactions themselves and solvent-solvent interactions may not be necessarily pairwise additive. We consider the following nonequilibrium process: at $t < 0$, the system is in equilibrium at the ground state and then the system is changed to the excited state from $t = 0$. The Hamiltonian corresponding to this process is represented as

$$\mathcal{H}(t) = \mathcal{H}_0 + \Theta(t)\Delta\mathcal{H}, \tag{2.35}$$

where $\Theta(t)$ is the Heaviside step function. The solvation dynamics is characterized with the solvation time correlation function (STCF), $S(t)$, defined as

$$S(t) \equiv \frac{\langle \Delta\mathcal{H}(t) \rangle_{\mathrm{ne}} - \langle \Delta\mathcal{H}(\infty) \rangle_{\mathrm{ne}}}{\langle \Delta\mathcal{H}(0) \rangle_{\mathrm{ne}} - \langle \Delta\mathcal{H}(\infty) \rangle_{\mathrm{ne}}}. \tag{2.36}$$

Here, $\langle \cdots \rangle_{\mathrm{ne}}$ indicates the nonequilibrium average governed by Eq. (2.35). In the present study, we assume that the solute molecule is rigid and its electronic structure is not modulated by the solvents, i.e., $\Delta E_{\mathrm{solute}}$ is a constant during the relaxation process. Thus, introducing the difference of the total solute-solvent interaction between the excited and ground states at time t as

$$\Delta E(t) \equiv \sum_v \sum_{i\in v} \left\{ u_v^{\mathrm{e}}(\boldsymbol{x}_{v,i}(t)) - u_v^{\mathrm{g}}(\boldsymbol{x}_{v,i}(t)) \right\}, \tag{2.37}$$

Eq. (2.36) is rewritten as

$$S(t) = \frac{\langle \Delta E(t) \rangle_{\mathrm{ne}} - \langle \Delta E(\infty) \rangle_{\mathrm{ne}}}{\langle \Delta E(0) \rangle_{\mathrm{ne}} - \langle \Delta E(\infty) \rangle_{\mathrm{ne}}}. \tag{2.38}$$

The fluctuation dissipation theorem provides the following tractable expression of $S(t)$ in the linear-response regime.

$$S(t) = \frac{\langle \delta\Delta E(t) \delta\Delta E \rangle}{\langle \delta\Delta E \delta\Delta E \rangle}. \tag{2.39}$$

Here, $\langle \cdots \rangle$ denotes an ensemble average at equilibrium with the Hamiltonian $\mathcal{H}_0$ and $\delta\Delta E(t)$ is the fluctuation of $\Delta E(t)$ in the equilibrium ensemble with $\mathcal{H}_0$, $\delta\Delta E(t) =$

$\Delta E(t) - \langle \Delta E \rangle$. Since the system is described by $\mathcal{H}_0$ till $t = 0$, $\langle \cdots \rangle$ corresponds to the average at $t = 0$. If the potential to define the energy coordinate is set to

$$\varepsilon_v(\boldsymbol{x}_{v,i}) = u_v^{\mathrm{e}}(\boldsymbol{x}_{v,i}) - u_v^{\mathrm{g}}(\boldsymbol{x}_{v,i}), \tag{2.40}$$

$\langle \Delta E(t) \rangle_{\mathrm{ne}}$ can be simply written as

$$\Delta E(t) = \sum_v \int \mathrm{d}\varepsilon \; \varepsilon \rho_v(\varepsilon, t), \tag{2.41}$$

$$\langle \Delta E(t) \rangle_{\mathrm{ne}} = \sum_v \int \mathrm{d}\varepsilon \; \varepsilon \left\langle \rho_v(\varepsilon, t) \right\rangle_{\mathrm{ne}}. \tag{2.42}$$

In the original ER theory, the energy coordinate is defined as the solute-solvent interaction for free energy calculation, and its definition is almost parallel to that in the present study: the initial state contains a phantom solute that does not interact with the solvent, and the full solute-solvent interactions are present at the final state.

It indicates that the difference between these two states becomes the full solute-solvent interaction potential. Therefore, the difference of the solute-solvent interaction between the ground and excited states (Eq. (2.40)) is a natural choice as the energy coordinate for solvation dynamics. In empirical models, solvation dynamics has been regarded as the dynamics along the "solvation coordinate", which is a collective coordinate of solvent molecules usually defined as the difference of the total potential energy between the two different states. [122, 123] Unlike such a collective coordinate, the energy coordinate (Eq. (2.40)) refers the difference of pairwise solute-solvent interaction.

According to the linear response theory, [1, 3, 30, 124] the approximate relationships are given by

$$\langle \rho_v(\varepsilon, t) \rangle_{\mathrm{ne}} - \langle \rho_v(\varepsilon, \infty) \rangle_{\mathrm{ne}} = \beta \langle \delta\rho_v(\varepsilon, t) \Delta\mathcal{H} \rangle = \beta \langle \delta\rho_v(\varepsilon, t) \Delta E \rangle \tag{2.43}$$

and

$$\langle \delta\rho_v(\varepsilon) \, \delta\rho_w(\eta) \rangle = \langle \delta\rho_v(\varepsilon) \, \delta\rho_w(\eta) \rangle_{\mathrm{ES}}. \tag{2.44}$$

Here, β is inverse temperature and $\langle \cdots \rangle_{\mathrm{ES}}$ represents the equilibrium average at the excited state. Eq. (2.43) is rewritten as

$$\langle \rho_v(\varepsilon, t) \rangle_{\mathrm{ne}} - \langle \rho_v(\varepsilon, \infty) \rangle_{\mathrm{ne}} = \beta \sum_w \int \mathrm{d}\eta \; \eta G_{vw}(\varepsilon, \eta, t) = \beta Q_v(\varepsilon, t), \tag{2.45}$$

where $G_{vw}(\varepsilon, \eta, t)$ is introduced by Eq. (2.24) and $Q_v(\varepsilon, t)$ is defined as

$$Q_v(\varepsilon, t) \equiv \sum_w \int \mathrm{d}\eta \; \eta G_{vw}(\varepsilon, \eta, t) = \langle \delta\rho_v(\varepsilon, t) \delta\Delta E \rangle. \tag{2.46}$$

From the conservation of the number of molecules in the system, the integral of $Q_v(\varepsilon, t)$ over the energy leads to

$$\int_{-\infty}^{\infty} \mathrm{d}\varepsilon \, Q_v(\varepsilon, t) = 0. \tag{2.47}$$

Since $\langle\rho_v(\varepsilon,\infty)\rangle_{\rm ne}$ is the equilibrium distribution at the excited state, $\langle\rho_v(\varepsilon)\rangle_{\rm ES}$, Eq. (2.45) is expressed as

$$\langle\rho_v(\varepsilon,t)\rangle_{\rm ne} = \langle\rho_v(\varepsilon)\rangle_{\rm ES} + \beta Q_v(\varepsilon,t)\,. \tag{2.48}$$

An alternative expression is obtained from the equivalence of $\langle\rho_v(\varepsilon,0)\rangle_{\rm ne}$ to $\langle\rho_v(\varepsilon)\rangle$ as

$$\langle\rho_v(\varepsilon,t)\rangle_{\rm ne} = \langle\rho_v(\varepsilon)\rangle + \beta\{Q_v(\varepsilon,t) - Q_v(\varepsilon,0)\}. \tag{2.49}$$

Substituting Eqs. (2.42) and (2.45) into Eq. (2.36) gives

$$S(t) = \frac{\sum_v \int \mathrm{d}\varepsilon\ \varepsilon Q_v(\varepsilon,t)}{\sum_v \int \mathrm{d}\varepsilon\ \varepsilon Q_v(\varepsilon)}. \tag{2.50}$$

$G_{vw}(\varepsilon,\eta,t)$ in Eq. (2.45) evolves with time through with Eq. (2.25). Since Eq. (2.25) does not contain the differential operator acting on the energy coordinate η, the differential equation for $Q_v(\varepsilon,t)$ is readily obtained as

$$\begin{aligned}\frac{\partial Q_v(\varepsilon,t)}{\partial t} = \frac{\partial}{\partial\varepsilon}\Bigg[&D_v^{\rm e}(\varepsilon)\frac{\partial Q_v(\varepsilon,t)}{\partial\varepsilon} - D_v^{\rm e}(\varepsilon)\frac{\mathrm{d}\ln\langle\rho_v(\varepsilon)\rangle}{\mathrm{d}\varepsilon}Q_v(\varepsilon,t)\\ &-D_v^{\rm e}(\varepsilon)\langle\rho_v(\varepsilon)\rangle\sum_w\int\mathrm{d}\eta\ \frac{\partial c_{vw}(\varepsilon,\eta)}{\partial\varepsilon}Q_w(\eta,t)\Bigg].\end{aligned} \tag{2.51}$$

It indicates that the solvation dynamics can be described on the energy coordinate by numerically solving Eq. (2.51). Eq. (2.51) reduces to

$$\frac{\partial Q_v(\varepsilon,t)}{\partial t} = \frac{\partial}{\partial\varepsilon}\left[D_v^{\rm e}(\varepsilon)\frac{\partial Q_v(\varepsilon,t)}{\partial\varepsilon} - D_v^{\rm e}(\varepsilon)\frac{\mathrm{d}\ln\langle\rho_v(\varepsilon)\rangle}{\mathrm{d}\varepsilon}Q_v(\varepsilon,t)\right]. \tag{2.52}$$

In Sec. 2.2.3, an approximation for $D_v^e(\varepsilon)$ has been described. Eqs. (2.51) and (2.52) have been derived, on the other hand, only by combining the formulation in Sec. 2.2.2 and the linear-response treatment for the solvation dynamics without resorting to any approximations to $D_v^e(\varepsilon)$. The overdamped condition of Eq. (2.17) is a step to Eq. (2.51), and an approximation may be independently adopted for $D_v^e(\varepsilon)$. $\langle\rho_v(\varepsilon)\rangle$, $D_v^e(\varepsilon)$, and $c_{vw}(\varepsilon,\eta)$ are involved in Eqs. (2.51) and (2.52), and to solve these equations, $Q_v(\varepsilon,t=0)$ is also needed as the initial condition. It should be noted that $\langle\rho_v(\varepsilon)\rangle$, $D_v^e(\varepsilon)$, $c_{vw}(\varepsilon,\eta)$, and $Q_v(\varepsilon,t=0)$ can be computed from the molecular simulation of the static ($t=0$) solute-solvent system at equilibrium with the solute being at its ground state, as well as the diffusion coefficient of the bulk solvent. The direct correlation function, $c_{vw}(\varepsilon,\eta)$, can be computed from Eq. (2.6) using the simulation at the ground state. Hence, Eq. (2.51) can be a scheme to evaluate the (long-time) dynamics on the basis of readily accessible properties in molecular simulation. Note that there are three expressions of $Q_v(\varepsilon,t=0)$ in the linear response limit. From Eqs. (2.44) and (2.46), one can obtain

$$Q_v^{(1)}(\varepsilon,t=0) = \sum_w\int d\eta\,\eta\,\langle\delta\rho_v(\epsilon)\delta\rho_w(\eta)\rangle\,, \tag{2.53}$$

and

$$Q_v^{(2)}(\varepsilon,t=0) = \sum_w\int d\eta\,\eta\,\langle\delta\rho_v(\epsilon)\delta\rho_w(\eta)\rangle_{\rm ES}\,. \tag{2.54}$$

Since the value of $\langle \rho_v(\varepsilon, t=0) \rangle$ is equivalent to $\langle \rho_v(\varepsilon) \rangle$, Eq. (2.48) is rewritten as

$$Q_v^{(3)}(\varepsilon, t=0) = \frac{1}{\beta} \Big\{ \langle \rho_v(\varepsilon) \rangle - \langle \rho_v(\varepsilon) \rangle_{\mathrm{ES}} \Big\}. \tag{2.55}$$

While Eqs. (2.53) and (2.54) respectively require the molecular simulations at the ground state and at the excited states, the simulations both at the ground and excited states are required for Eq. (2.55). In the present study, we utilize Eq. (2.53) for computing the initial value.

2.3 Computational methods

2.3.1 System modeling

We investigated the aqueous solution of benzonitrile in the infinite dilution limit. The structure of benzonitrile was optimized at MP2/6-31G(d) level calculation with Gaussian16. [125] The atomic point charges of benzonitrile at the ground and excited states were taken from the previous study by Ishida *et al.*, [49] and these values are shown in Fig. 2.1. The force fields for benzonitrile and water molecules are the generalized Amber force field (GAFF) [126] and TIP3P, respectively. In order to check the box size dependency of the present theory, we prepared two systems with different box sizes, $60 \times 60 \times 60$ Å^3 (7200 water molecules) and $80 \times 80 \times 80$ Å^3 (17067 water molecules). The schemes of the MD simulations described below were adopted for both the systems. The initial configurations were built using Packmol. [127]

2.3.2 Simulation setups

We performed three types of MD simulations of benzonitrile in water at 298.15 K at constant volume and numbers of particles. One is at equilibrium with benzonitrile at the ground state, and the others are with the excited-state benzonitrile. The second type is in the equilibrium condition, and the last one is the nonequilibrium MD (NEMD) simulations corresponding to the solvent relaxation process due to the photoexcitation of benzonitrile. The results from the NEMD simulations were used for comparison with those from the present theory.

To perform a series of MD simulations at equilibrium, we prepared the system configurations separately at the ground and excited states for benzonitrile. At each state, an MD simulation was performed for 1 ns. Then, we carried out the simulations for 5 ns at the ground state and for 1 ns at the excited state to extract the configurations every 1 ps. The total numbers of sampled configurations were 5000 and 1000 at the ground and excited states, respectively. We performed the simulations for 0.5 ns starting from the sampled configurations for equilibration, where the random seed for the thermostat was different among the distinct runs. The production MD simulations (1 ns for each) were performed at the ground and excited states, and the number of trajectories was 1000 at each state. We also conducted the 5000 NEMD simulations (2 ps for each) starting from the configurations at the ground state while using the atomic charges of benzonitrile at the excited state.

For all the simulations, we used the velocity Verlet (VVER) integrator [128] with a time interval of 2 fs and the Bussi thermostat. [129] The benzonitrile molecule was treated as rigid by setting the velocities of its atoms to zero, corresponding to the neglect

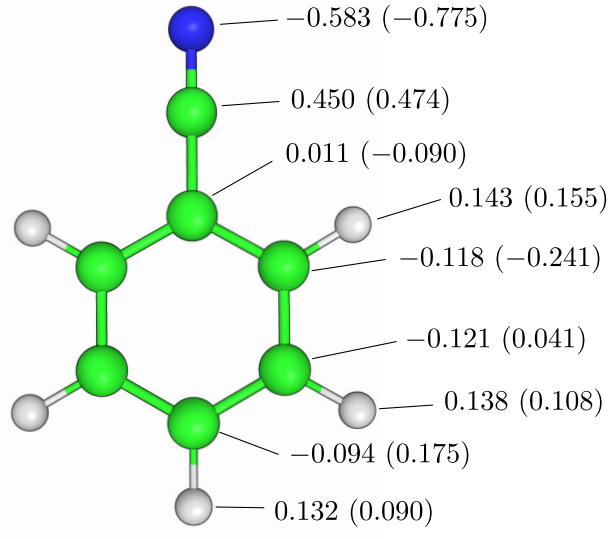


Fig. 2.1: Atomic point charges for benzonitrile at the ground and excited states. The values in parentheses mean the charges at the excited state. The carbon, nitrogen, and hydrogen atoms are depicted in green, blue, and gray, respectively. The absolute value of the dipole moment is changed from 5.88 D to 12.6 D due to the excitation.

of both the translational and rotational motions of benzonitrile. The cutoff distance of Lennard-Jones potentials was 9.0 Å, and the particle-mesh Ewald (PME) method was used to calculate the electrostatic potential. Water molecules were kept rigid using the SETTLE algorithm. [130] All the MD simulations were performed with GENESIS 2.0 beta [131–133] in which the freeze scheme of a molecule was implemented by us. All the analyses were performed using in-house Fortran90/95 programs combined with the visual molecular dynamics (VMD) package (ver. 1.9.4 alpha) [134] and ERmod 0.3.7. [116]

2.3.3 Solver for ERS and ERSV equations

In order to numerically solve ERS and ERSV equations, we constructed a scheme of treating these equations in the discretized forms. It is known that the solvent distribution on the energy coordinate has a very sharp peak around $\varepsilon = 0$. For numerical accuracy, we introduced a function for generating non-uniform grids on the energy coordinate, which are fine around $\varepsilon = 0$ (see Sec. S1 and Fig. S1 of the supplementary material). The discretization was performed with the finite volume method (FVM) to assure the conservation law of particle number, as shown in Fig. S2 of the supplementary material. [135] As well as in the case of a solver of Smoluchowski equation, [136] the drift term was discretized by the 1st-order upwind difference scheme, and ERS and ERSV equations were integrated by the full implicit algorithm to obtain the numerical stability. The time grid Δt is set to be 1 fs. Further details of the solver is shown in Sec. S2 of the supplementary material

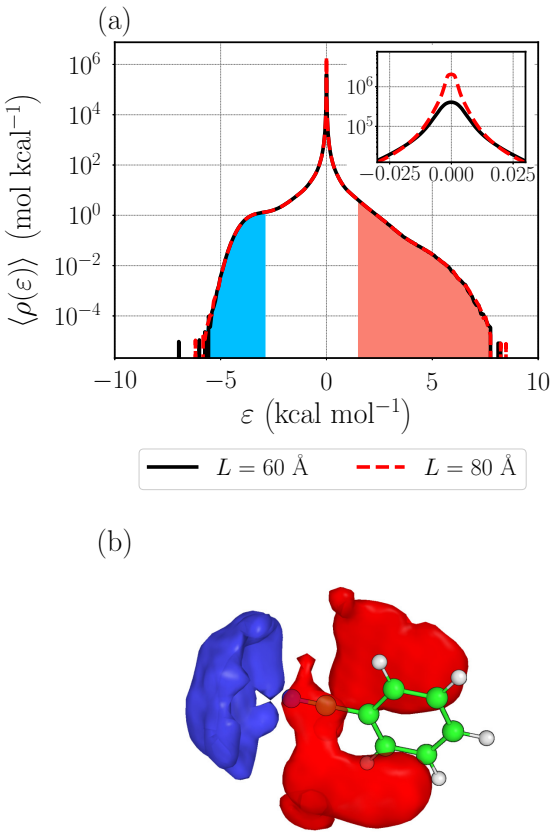


Fig. 2.2: Solvent distributions at the ground state. (a) The distributions on the energy coordinate obtained from the simulations with different box lengths, L = 60 Å and 80 Å. The energy coordinate ε is defined as the difference between the solute-solvent interaction potential at the ground state and that at the excited state. (b) The spatial distribution function (SDFs), $g_S(\boldsymbol{r})$, corresponding to the destabilized region ($\varepsilon \geq 1.5$ kcal mol^{-1}, red) and the stabilized region ($\varepsilon \leq -2.8$ kcal mol^{-1}, blue). The $g_S(\boldsymbol{r})$ at the isovalue of 0.4 are visualized with PyMOL. [137] In (a), those regions are shaded in cyan and orange, respectively.

2.4 Results and discussion

2.4.1 Solvent distribution on the energy coordinate

First, we examine the solvent distributions on the energy coordinate at the ground state of benzonitrile, $\langle\rho(\varepsilon)\rangle$ (Fig. 2.2(a)). The energy coordinate, ε, is defined as the difference of the solute-solvent interaction energy between the excited and the ground states. Thus, the destabilized and stabilized water molecules due to the excitation contribute to the distribution at $\varepsilon > 0$ and $\varepsilon < 0$, respectively.

It is seen for the benzonitrile solute at the ground state that the distribution monotonically increases with ε at $\varepsilon < 0$. When the MD simulations are done with the excited-state benzonitrile, a shallow well of the distribution exists at $\varepsilon = -2.8$ kcal mol^{-1}, and the population corresponding to the stabilized water molecules ($\varepsilon < 0$) is increased (see Fig. S3 of the supplementary material). The shapes of the distributions obtained from the MD simulations with the different box lengths are almost the same as each other except for $\varepsilon \simeq 0$. It is well known that the energy distribution function has the system size dependency around $\varepsilon = 0$ and shows divergent behavior as the box size increases. As will

be discussed in Sec. 2.4.2, on the other hand, the system size dependency of $\langle\rho(\varepsilon)\rangle$ is cancelled with that of $D^{\mathrm{e}}(\varepsilon)$ for the dynamics.

The locations of stabilized and destabilized water molecules can be illustrated by computing the water CoM densities in the spatial regions corresponding to $\varepsilon \leq -2.8\,\mathrm{kcal\,mol^{-1}}$ and $\varepsilon \geq 1.5\ \mathrm{kcal\ mol^{-1}}$, respectively, and determining the spatial distribution functions (SDFs), $g_S(\boldsymbol{r})$, as their ratios to the bulk density of water. SDFs are shown in Fig. 2.2(b), and they indicate that the most stabilized water molecules are localized around the nitrogen atom of benzonitrile, while the destabilized ones are distributed around the benzene ring. The same trend is also observed in the SDFs with different isovalues (see Fig. S4(a) of the supplementary material). The stabilized region at the excited state becomes wider than that at the ground state (see Fig. S4(b) of the supplementary material). It can be confirmed from the radial distribution function (RDF) for the stabilized region that the stabilized water molecules dominantly contribute to the first solvation shell around the nitrogen atom (see Fig. S5 of the supplementary material). Considering the fact that the excitation makes the nitrogen atom more negative, the distribution of the stabilized water molecules is reasonable.

2.4.2 Diffusion coefficient on the energy coordinate

The diffusion coefficients on the energy coordinate, $D^{\mathrm{e}}(\varepsilon)$, calculated from Eq. (2.33) are shown in Fig. 2.3(a). The minimum is located around $\varepsilon = 0$, indicating that the difference of the solute-solvent interaction between the excited and ground states varies slowly in the region. This is because the bulk water molecules are dominant around $\varepsilon = 0$ and their wandering has little contribution to the change on the energy coordinate. On the other hand, high diffusivity is observed in both the stabilized ($\varepsilon < 0$) and destabilized ($\varepsilon > 0$) regions. From the definition (Eq. (2.33)), $D^{\mathrm{e}}(\varepsilon)$ is large where the large force acts on the solvent molecules. Therefore, the high diffusivity in the stabilized and destabilized regions can be interpreted that the difference of the solute-solvent interactions in these regions varies significantly depending on their configurations. The system size dependency is slightly discernible around $\varepsilon = 0$. As shown in Eq. (2.22), however, $D^{\mathrm{e}}(\varepsilon)$ appears with $\langle\rho(\varepsilon)\rangle$, i.e.,

$$D^{\mathrm{e}}(\varepsilon)\,\langle\rho(\varepsilon)\rangle = D\sum_i \left\langle \left|\boldsymbol{f}_i^{\mathrm{G}}\right|^2 \delta(\varepsilon - \varepsilon(\boldsymbol{x}_i))\right\rangle, \tag{2.56}$$

and hence the system size dependency should be discussed with $D^{\mathrm{e}}(\varepsilon)\,\langle\rho(\varepsilon)\rangle$ instead of $D^{\mathrm{e}}(\varepsilon)$. Fig. 2.3(b) reveals that the dependency of $D^{\mathrm{e}}(\varepsilon)\,\langle\rho(\varepsilon)\rangle$ on the system size is negligibly small. The number of bulk molecules corresponding to region $\varepsilon \sim 0$ depends on the system size, but the force $|\boldsymbol{f}_i^{\mathrm{g}}|$ for such molecules are small. Thus, they do not contribute to $D^{\mathrm{e}}(\varepsilon)\,\langle\rho(\varepsilon)\rangle$ (Eq. (2.56)), leading to the cancellation of the system size dependencies appearing in $\langle\rho(\varepsilon)\rangle$ and $D^{e}(\varepsilon)$. The water molecules around benzonitrile which are neither stabilized nor destabilized by the excitation dominantly contribute to region $\varepsilon \sim 0$. According to Fig. 2.2, most of water molecules which belong to the high peak of $D^{\mathrm{e}}(\varepsilon)\,\langle\rho(\varepsilon)\rangle$ located at $\sim\!-3.5\,\mathrm{kcal\,mol^{-1}}$ in Fig. 2.3(b) coordinate to the nitrogen atom of benzonitrile. Hence, the well existing at $\sim\!-2\,\mathrm{kcal\,mol^{-1}}$ can be interpreted as a kinetic trap for changing the states of water molecules around benzonitrile.

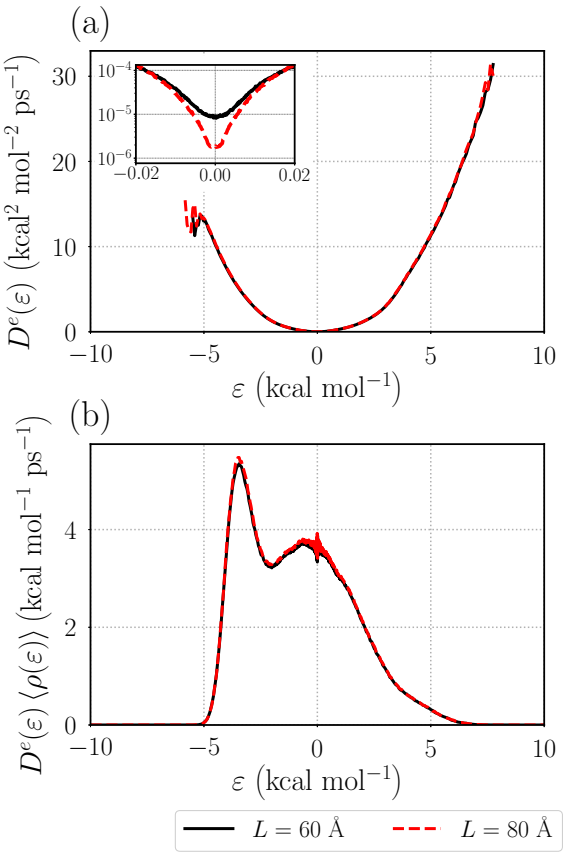


Fig. 2.3: Solvent diffusivity on the energy coordinate. (a) Diffusion coefficients, $D^{\mathrm{e}}(\varepsilon)$, and (b) $D^{\mathrm{e}}(\varepsilon)\,\langle\rho(\varepsilon)\rangle$.

2.4.3 Relaxation of the solvent distribution

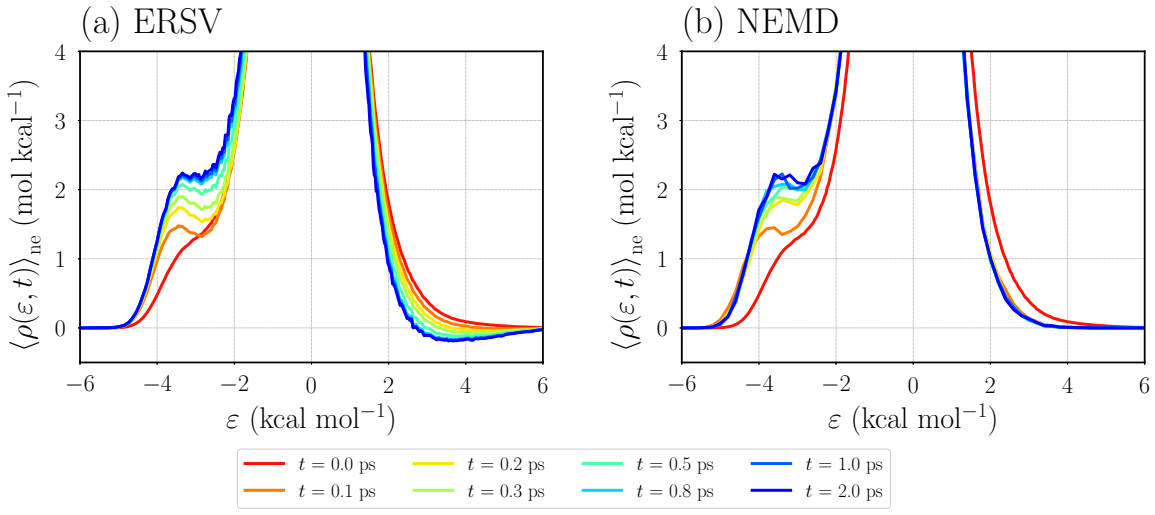


Fig. 2.4: Time developments of the distributions on the energy coordinate after the excitation of benzonitrile, obtained from (a) ERSV equation and (b) NEMD simulations. In the case of ERSV equation, Eq. (2.49) is used for computing $\langle\rho(\varepsilon,t)\rangle$. The box length of the simulations is 60 Å for both the cases.

We discuss the time development of $\langle \rho(\varepsilon, t) \rangle_{\mathrm{ne}}$ obtained by numerically solving ERSV equation for $Q(\varepsilon, t)$. In the computation, the initial value of $Q(\varepsilon, t)$ is evaluated through Eq. (2.53), i.e., $Q(\varepsilon, t = 0) = Q^{(1)}(\varepsilon, t = 0)$. The comparison of $Q^{(1)}(\varepsilon, t = 0)$ with $Q^{(2)}(\varepsilon, t = 0)$ and $Q^{(3)}(\varepsilon, t = 0)$ (Fig. S6 of the supplementary material) reveals that the difference between them is sufficiently small. Note that the computed $Q_v(\varepsilon, t)$ satisfactory reproduces the conservation law (Eq. (2.47)), revealing the validity of the present numerical scheme (see Fig. S7 of the supplementary material). In Fig. 2.4, the time development of $\langle \rho(\varepsilon, t) \rangle_{\mathrm{ne}}$ using Eq. (2.49) is illustrated together with that from the NEMD simulations. It is seen from $\langle \rho(\varepsilon, t) \rangle_{\mathrm{ne}}$ with ERSV equation (Fig. 2.4(a)) that a peak arises at $\sim$−3.5 kcal mol^{-1} immediately after the excitation and its peak height becomes higher as time proceeds. A similar trend is also observed in the distribution from the NEMD simulations (Fig. 2.4(b)). Although the differences of the peak heights and shapes at $0.1 \leq t/\mathrm{ps} \leq 0.8$ are discernible between ERSV equation and the NEMD simulations, a good agreement is realized at $t \geq 1.0$ ps. The rate of decreasing the population in $\varepsilon > 0$ predicted from ERSV equation is much slower compared with the NEMD simulations, suggesting the importance of the inertial and memory effects, which are neglected in ERSV equation, for reproducing the fast relaxation process. The time development of $\langle \rho(\varepsilon, t) \rangle_{\mathrm{ne}}$ using Eq. (2.48) is similar to that using Eq. (2.49) (see Fig. S8 of the supplementary material). In Fig. 2.4(a), an unphysical negative distribution is slightly discernible around $\varepsilon \sim 3$ kcal mol^{-1}, as also reported in the case of the spatial site distribution functions obtained by the surrogate theory. [45] Such a behavior can be interpreted as a nonlinearlity of the solvation dynamics. Further discussion is found in Sec. S3 and Fig. S9 of the supplementary material. On the other hand, it should be emphasized that the negative distribution in Fig. 2.4(a) is negligibly small, and hence the linear response treatment should be valid for the present system.

For more quantitative comparison, we analyze the change in the number of water molecules in the stabilized region ($\varepsilon \leq -2.8$ kcal mol^{-1}) and in the destabilized region ($\varepsilon \geq 1.5$ kcal mol^{-1}), as shown in Fig. S10 of the supplementary material. It is found that the difference of the number of water molecules between ERSV equation and the NEMD simulations is smaller than unity even in the destabilized region. Hence, the prediction of the time development of $\langle \rho(\varepsilon, t) \rangle_{\mathrm{ne}}$ using ERSV equation is useful for understanding the dynamics on the long timescale in detail.

2.4.4 Solvation time correlation function

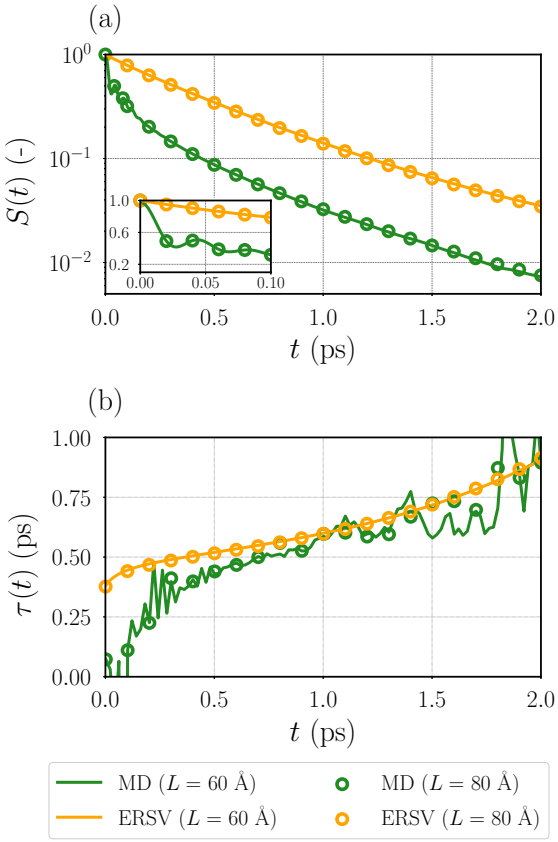


Fig. 2.5: Time correlation functions for the solvent relaxation process obtained from ERSV equation and the MD simulation at the equilibrium ground state. (a) Solvation time correlation functions (STCFs), $S(t)$, and (b) time-dependent relaxation time coefficients, $\tau(t)$.

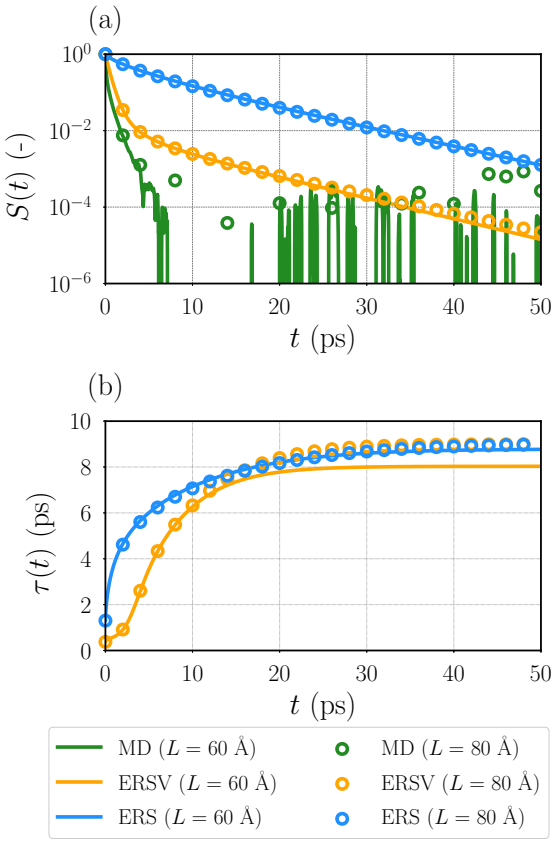


Fig. 2.6: Time correlation functions for the solvent relaxation process obtained from ERSV and ERS equations. (a) Solvation time correlation functions (STCFs), $S(t)$, and (b) time-dependent relaxation time coefficients. For comparison, $S(t)$ obtained from the MD simulation at the equilibrium ground state is also shown.

In this subsection, we address the accuracy of the solvation time correlation functions (STCFs), $S(t)$, obtained from ERSV equation. For comparison, we also compute the function from the MD simulations at the equilibrium ground state by employing an approximate expression derived from the linear response theory, Eq. (2.39). Note that the approximated STCF from the MD simulations matches the exact function from the NEMD simulations with Eq. (2.36), proving the validity of the linear response theory for the present system (see Fig. S11 of the supplementary material). Fig. 2.5 shows the STCFs and the time-dependent relaxation time coefficients defined as

$$\tau(t) = -\left(\frac{\mathrm{d}\ln S(t)}{\mathrm{d}t}\right)^{-1}, \tag{2.57}$$

obtained from ERSV equation and from the MD simulations at the equilibrium ground state (Eq. (2.39)). Note that $\tau(\infty)$ coincides with the relaxation time constant in the diffusion regime. The linear plots of the STCFs are available in Fig. S12 of the supplementary material. In order to reduce the statistical noise in $S(t)$ from the MD simulations, the moving average scheme was employed. The numbers of points for the average were 5 for $0 < t/\mathrm{ps} < 0.5$, 21 for $0.5 < t/\mathrm{ps} < 1.0$ and 51 for $1.0 < t/\mathrm{ps} < 2.0$, where the time interval used for computing $S(t)$ is 0.002 ps. It is confirmed that this operation hardly changes the curves of $S(t)$. While $S(t)$ obtained from ERSV equation decays monotonically, the initial Gaussian decay followed by the damped oscillation appears on

a short timescale in $S(t)$ from the MD simulations (see Fig. S13 of the supplementary material). As Nishiyama *et al.* mentioned in their study using SSSV and RISM/mode-coupling theories, [55] such oscillations reflect the memory effects of solvent motions. In order to elucidate the origin of the oscillations, we compute the time correlation functions of the rotational motions of the water molecules coordinating to the nitrogen atom of benzonitrile with the hydrogen bonding from the MD simulations. The definitions of the functions are described in Sec. S4 and Fig. S14 of the supplementary material. The time correlation functions are shown in Fig. S15 of the supplementary material. The phase and frequency of the damped oscillations appearing in these functions are found to be the almost same as that in $S(t)$. Hence, the rotational motion of water molecules restricted by the hydrogen bonding has a vital role in the dynamics on the short timescale. The slower decay of $S(t)$ from ERSV equation compared with the MD simulation might stem from the insufficient description of the rotational motion in $D^e(\varepsilon)$ (Eq. (2.33)) due to the decoupling approximation (Eq. (2.30)). An upgrade of the method so as to include the memory effect and rotational motion is important for accurately describing the fast relaxation dynamics. On the other hand, although the discrepancy between ERSV equation and the MD simulations is shown due to the difference of $S(t)$ appearing at $t \leq 0.5$ ps, the slopes of $S(t)$ on the logarithmic scale for ERSV equation and the MD simulations are similar to each other at $t > 1.0$ ps. As also shown in Fig. 2.5(b), $\tau(t)$ from ERSV equation is in accord with that from the MD simulations at $t > 0.5$ ps. It indicates that the dynamics on the long timescale is well captured by ERSV equation. Furthermore, it should be emphasized that the present theory can be used for robustly discussing the dynamics on the long timescale while avoiding the sampling error. In comparison, $\tau(t)$ obtained from the MD simulations shows the noise due to a poor sampling, and this noise could be problematic to determine the relaxation time constant on the long timescale.

In addition to the usefulness of the present theory for describing the dynamics on the long timescale, a systematic analysis about the effect of the collective solvent diffusion is available using both ERS and ERSV equations. Since ERS equation is derived by neglecting the collective diffusion part from ERSV equation, only the single-diffusion process is considered in ERS equation. Thus, it can be rigorously stated that the difference of $S(t)$ between ERS and ERSV equations reflects the significance of the collective diffusion on the relaxation process. The time developments of $S(t)$ and $\tau(t)$ computed from these equations are shown in Fig. 2.6. It is seen that the system size dependency of $S(t)$ is slightly appeared for ERSV equation at $t > 20$ ps, while such dependency is not present in the function from ERS equation. In the case of ERS equation, the deviation of $S(t)$ from the MD simulations is larger as compared to ERSV equation. It is further observed that $S(t)$ decays more slowly with ERS than with ERSV, showing that the collective motion of solvent molecules facilitates the relaxation. The difference is evident on the timescale of $t < 10.0$ ps, and still alive on the longer timescale. Interestingly, the value of $\tau(t)$ from ERS equation becomes close to that from ERSV equation as time proceeds, as described in Fig. 2.6(b). At $t > 15.0$ ps, the match of $\tau(t)$ with ERSV equation is observed when the box length of the system, L, is 80 Å. Accordingly, the collective diffusion has an impact on the relaxation process until 15.0 ps.

2.5 Concluding remarks

We formulated the energy-represented generalized Langevin equation (ERGLE), an exact differential equation, based on Zwanzig-Mori projection operator theory. In the equation, the molecular motions are described over one-dimensional space without any approximations using the solute-solvent interaction energy as the coordinate. The quantities involved in ERGLE were introduced on low-dimensional coordinates compared with those employing the conventional coordinates of the positions and orientations of molecules, and hence the their evaluation using the molecular dynamics (MD) simulation can be much easier. Thanks to the exact treatment of the dynamics, ERGLE would be useful for describing the dynamic processes occurring in the molecular liquids composed of the arbitrary complicated molecules and in the heterogeneous environments such as lipid membranes. The energy-represented Smoluchowski-Vlasov (ERSV) and Smoluchowski (ERS) equations were formulated by systematically neglecting the inertial and memory effects, which can be important for the short-time dynamics. They are thus a scheme for describing the long-time dynamics and adopts the (static) solvent distribution and diffusion coefficient on the energy coordinate as inputs, which can be easily evaluated through MD simulations. It should be noted that the present theory has no assumptions about the chemical properties of solvents and hence offers a theoretical fundament to analyze the dynamics in various solvents systems such as protic/aprotic solvents and ionic liquids.

ERSV and ERS equations were applied to the water relaxation process triggered by the photoexcitation of benzonitrile in conjunction with the linear response theory. The solvent distribution and diffusion coefficient on the energy coordinate were examined. The energy coordinate, ε, was defined as the difference of the solute-solvent interaction energy between the excited and ground states of benzonitrile. The distribution showed the populations of the stabilized and destabilized water molecules due to the excitation of benzonitrile, and these molecules were found to be localized around the nitrogen atom and benzene ring of benzonitrile, respectively. The diffusivity of water on the energy coordinate was high in the stabilized and destabilized regions on the energy coordinate, while that for the bulk water was quite small. Both the distribution and diffusion coefficient showed the system-size dependencies around $\varepsilon \sim 0$, but we revealed that these dependencies bring no system-size dependency to the dynamical behavior obtained from ERSV and ERS equations. The time development of the nonequilibrium distribution obtained from ERSV equation showed that the growth of the peak in the stabilized region, which is also observed in that from the nonequilibrium MD (NEMD) simulations. Although the difference of the distribution between ERSV and the NEMD simulation was discernible especially in the destabilized region on the short timescale, the good agreement was realized on the long timescale. The solvation time correlation function (STCF) obtained from ERSV equation deviated from that from the MD simulations due to the difference appearing on the short timescale. On the other hand, ERSV equation reproduced the relaxation time coefficient from the MD simulations on the long timescale, indicating the usefulness of ERSV equation for describing the long-time dynamics. The comparison of the STCF obtained from ERS equation with that from ERSV equation clarified the impact of the collective solvent diffusion on the relaxation time. For further investigation, on the other hand, developing the methodology to realize the decomposition of $S(t)$ into the contributions of the optical and acoustic modes in ER framework would be important for increasing the usefulness of the present study.

We shall comment on the possibility for improving the description of the dynamics

based on the present energy-represented theory. In the present study, an approximated expression of the diffusion coefficient on the energy coordinate uses the translational diffusion coefficient of water in the bulk, and this treatment should be valid only for homogeneous systems with weak solute-solvent interactions. Since ERSV and ERS have been derived without referring to an explicit expression for the diffusion coefficient, any upgrade for the approximation for the diffusivity will lead to improvement of the theory. The inertial and memory effects are vital for describing the short-time dynamics. As well as in the case of the theories using the spatial coordinate, [30] a formulation of the GLEs for the solvent distribution and its current might be important for treating the inertial effect. As for the memory effect, importing the approximations utilized in the viscoelastic [31, 32] and mode-coupling [138–140] theories into ER framework might be promising. We believe that further improvement would lead to deeper understanding of the dynamical behaviors in complex molecular liquids and in heterogeneous environments.

Appendix A

Appendix of chapter 1

A.1 Method to build non-uniform grid for energy coordinate

In order to numerically solve ERSV and ERS equations, we constructed a scheme of treating these equations in the discretized forms. The solvent distribution on the energy coordinate has a very sharp peak around $\varepsilon = 0$. Therefore, to perform discretization with numerical accuracy and efficiency, we have used following non-uniform grids for the energy coordinate,

$$\varepsilon_i = \frac{\varepsilon_{\max}}{\sinh\left[a\frac{1}{N}\frac{N-1}{2}\right]} \sinh\left[a\frac{1}{N}\left(i - \frac{N+1}{2}\right)\right] \qquad (i = 1, \cdots, N). \tag{A.1}$$

Here, N is the number of grids. $\varepsilon_{\max}$ determines the upper and lower bounds of energy coordinate ($\varepsilon_1 = -\varepsilon_{\max}$, $\varepsilon_N = \varepsilon_{\max}$). a corresponds to the change ratio of the grid width.

In this study, we have adopted $N = 1000$, $\varepsilon_{\max} = 10$ kcal mol^{-1} and $a = 15$. The relationship between the discretized energy coordinate ε_i and the grid width $\Delta\varepsilon_i \equiv \varepsilon_{i+1} - \varepsilon_i$ is shown in Fig. A.1.

A.2 Numerical scheme to solve ERSV equation

ERSV equation is given by Eq. (51) and is written as

$$\frac{\partial Q(\varepsilon,t)}{\partial t} = \frac{\partial}{\partial \varepsilon}\left[D^{\mathrm{e}}(\varepsilon)\frac{\partial Q(\varepsilon,t)}{\partial \varepsilon} - D^{\mathrm{e}}(\varepsilon)\frac{\mathrm{d}W(\varepsilon)}{\mathrm{d}\varepsilon}Q(\varepsilon,t) - \psi(\varepsilon)\int \mathrm{d}\eta\, \frac{\partial c(\varepsilon,\eta)}{\partial \varepsilon}Q(\eta,t)\right]. \tag{A.2}$$

For simplicity, we consider the single-component solvent and define

$$W(\varepsilon) \equiv \ln\langle\rho(\varepsilon)\rangle\,, \tag{A.3}$$

$$\psi(\varepsilon) \equiv \langle\rho(\varepsilon)\rangle\, D^{\mathrm{e}}(\varepsilon). \tag{A.4}$$

The subscript v is also dropped. We discretize Eq. (2.51) based on finite volume method (FVM) and 1st-order upwind difference. [135, 136] The schematic view of FVM is shown in Fig. A.2. Here, the subscript i and the superscript n mean ith grid on the energy

coordinate and nth time-step, respectively. By using the full implicit method, the time development of Q_i^n is written as

$$\left(\frac{\Delta\varepsilon_{i-1}}{2}+\frac{\Delta\varepsilon_i}{2}\right)\frac{Q_i^n-Q_i^{n-1}}{\Delta t}=j_{i,w}^n-j_{i,e}^n. \tag{A.5}$$

Then, we consider the form of currents. The contribution from diffusion term of Eq. (2.51) to the net current $j_{i,w}^n-j_{i,e}^n$ is

$$-D_{i,w}^{\mathrm{e}}\frac{Q_i^n-Q_{i-1}^n}{\Delta\varepsilon_{i-1}}+D_{i,e}^{\mathrm{e}}\frac{Q_{i+1}^n-Q_i^n}{\Delta\varepsilon_i}, \tag{A.6}$$

where $D_{i,w}^{\mathrm{e}}$ and $D_{i,e}^{\mathrm{e}}$ are the harmonic means of the diffusion coefficients in adjacent cells.

$$D_{i,w}^{\mathrm{e}}\equiv\frac{2D_i^{\mathrm{e}}D_{i-1}^{\mathrm{e}}}{D_i^{\mathrm{e}}+D_{i-1}^{\mathrm{e}}}, \tag{A.7}$$

$$D_{i,e}^{\mathrm{e}}\equiv\frac{2D_{i+1}^{\mathrm{e}}D_i^{\mathrm{e}}}{D_{i+1}^{\mathrm{e}}+D_i^{\mathrm{e}}}. \tag{A.8}$$

Note that $W(\varepsilon)$ is very sharp near the origin, causing numerical instability for computing the drift term. Then, we used 1st-order upwind difference to obtain the stability of the solution. Since the current occurs from the small $W(\varepsilon)$ to the large $W(\varepsilon)$, we adopt the following discretized equations

$$\left(D^{\mathrm{e}}(\varepsilon)Q(\varepsilon,t)\frac{\mathrm{d}W(\varepsilon)}{\mathrm{d}\varepsilon}\right)_{i,w}^n=\begin{cases}D_{i-1}^{\mathrm{e}}Q_{i-1}^n\dfrac{W_i-W_{i-1}}{\Delta\varepsilon_{i-1}} & \text{for } (W_i-W_{i-1}>0)\\ D_i^{\mathrm{e}}Q_i^n\dfrac{W_i-W_{i-1}}{\Delta\varepsilon_{i-1}} & \text{for } (W_i-W_{i-1}<0)\end{cases}, \tag{A.9}$$

$$\left(D^{\mathrm{e}}(\varepsilon)Q(\varepsilon,t)\frac{\mathrm{d}W(\varepsilon)}{\mathrm{d}\varepsilon}\right)_{i,e}^n=\begin{cases}D_i^{\mathrm{e}}Q_i^n\dfrac{W_{i+1}-W_i}{\Delta\varepsilon_i} & \text{for } (W_{i+1}-W_i>0)\\ D_{i+1}^{\mathrm{e}}Q_{i+1}^n\dfrac{W_{i+1}-W_i}{\Delta\varepsilon_i} & \text{for } (W_{i+1}-W_i<0)\end{cases}. \tag{A.10}$$

For simplicity, we define the following notations

$$\begin{aligned}&\delta W_{i+\frac{1}{2}}\equiv W_{i+1}-W_i,\quad \delta W_{i-\frac{1}{2}}\equiv W_i-W_{i-1},\\ &\delta W_{i\pm\frac{1}{2}}^{(+)}\equiv\frac{\delta W_{i\pm\frac{1}{2}}+\left|\delta W_{i\pm\frac{1}{2}}\right|}{2},\quad \delta W_{i\pm\frac{1}{2}}^{(-)}\equiv\frac{\delta W_{i\pm\frac{1}{2}}-\left|\delta W_{i\pm\frac{1}{2}}\right|}{2}.\end{aligned} \tag{A.11}$$

By using these notations, the net current due to the drift terms is

$$\frac{1}{\Delta\varepsilon_{i-1}}D_{i-1}^{\mathrm{e}}Q_{i-1}^n\delta W_{i-\frac{1}{2}}^{(+)}+\frac{1}{\Delta\varepsilon_{i-1}}D_i^{\mathrm{e}}Q_i^n\delta W_{i-\frac{1}{2}}^{(-)}-\frac{1}{\Delta\varepsilon_i}D_i^{\mathrm{e}}Q_i^n\delta W_{i+\frac{1}{2}}^{(+)}-\frac{1}{\Delta\varepsilon_i}D_{i+1}^{\mathrm{e}}Q_{i+1}^n\delta W_{i+\frac{1}{2}}^{(-)}. \tag{A.12}$$

The net current due to the direct correlation term is

$$\begin{aligned}&\psi_{i,w}\sum_j\Delta\varepsilon_j\frac{c_{i,j}-c_{i-1,j}}{\Delta\varepsilon_{i-1}}Q_j^n-\psi_{i,e}\sum_j\Delta\varepsilon_j\frac{c_{i+1,j}-c_{i,j}}{\Delta\varepsilon_i}Q_j^n\\ &=\sum_j\left\{\frac{\psi_{i,w}}{\Delta\varepsilon_{i-1}}\Delta\varepsilon_j(c_{i,j}-c_{i-1,j})-\frac{\psi_{i,e}}{\Delta\varepsilon_i}\Delta\varepsilon_j(c_{i+1,j}-c_{i,j})\right\}Q_j^n.\end{aligned} \tag{A.13}$$

$\psi_{i,w}$ and $\psi_{i,e}$ are the harmonic means of ψ in adjacent cells.

$$\psi_{i,w} \equiv \frac{2\psi_i\psi_{i-1}}{\psi_i + \psi_{i-1}}, \tag{A.14}$$

$$\psi_{i,e} \equiv \frac{2\psi_{i+1}\psi_i}{\psi_{i+1} + \psi_i}. \tag{A.15}$$

Substituting Eqs. (A.6), (A.12) and (A.13) into Eq. (A.5) yields

$$\begin{aligned}
&\left(\frac{\Delta\varepsilon_{i-1}}{2} + \frac{\Delta\varepsilon_i}{2}\right)\frac{Q_i^n - Q_i^{n-1}}{\Delta t} \\
&= -D_{i,w}^{\mathrm{e}}\frac{Q_i^n - Q_{i-1}^n}{\Delta\varepsilon_{i-1}} + D_{i,e}^{\mathrm{e}}\frac{Q_{i+1}^n - Q_i^n}{\Delta\varepsilon_i} \\
&\quad + \frac{1}{\Delta\varepsilon_{i-1}}D_{i-1}^{\mathrm{e}}Q_{i-1}^n\delta W_{i-\frac{1}{2}}^{(+)} + \frac{1}{\Delta\varepsilon_{i-1}}D_i^{\mathrm{e}}Q_i^n\delta W_{i-\frac{1}{2}}^{(-)} \\
&\quad - \frac{1}{\Delta\varepsilon_i}D_i^{\mathrm{e}}Q_i^n\delta W_{i+\frac{1}{2}}^{(+)} - \frac{1}{\Delta\varepsilon_i}D_{i+1}^{\mathrm{e}}Q_{i+1}^n\delta W_{i+\frac{1}{2}}^{(-)} \\
&\quad + \sum_j\left\{\frac{\psi_{i,w}}{\Delta\varepsilon_{i-1}}\Delta\varepsilon_j(c_{i,j} - c_{i-1,j}) - \frac{\psi_{i,e}}{\Delta\varepsilon_i}\Delta\varepsilon_j(c_{i+1,j} - c_{i,j})\right\}Q_j^n.
\end{aligned} \tag{A.16}$$

From the simple manipulation, the following simultaneous equation for Q_j^n is derived.

$$\sum_j A_{ij}Q_j^n = b_i, \tag{A.17}$$

$$\begin{aligned}
A_{ij} \equiv{}& \left\{-D_{i,e}^{\mathrm{e}}\Delta\varepsilon_{i-1} + D_{i+1}^{\mathrm{e}}\Delta\varepsilon_{i-1}\delta W_{i+\frac{1}{2}}^{(-)}\right\}\delta_{i+1,j} \\
&+ \left\{\alpha_i + D_{i,w}^{w}\Delta\varepsilon_i + D_{i,e}^{\mathrm{e}}\Delta\varepsilon_{i-1} - D_i^{\mathrm{e}}\Delta\varepsilon_i\delta W_{i-\frac{1}{2}}^{(-)} + D_i^{\mathrm{e}}\Delta\varepsilon_{i-1}\delta W_{i+\frac{1}{2}}^{(+)}\right\}\delta_{i,j} \\
&+ \left\{-D_{i,w}^{\mathrm{e}}\Delta\varepsilon_i - D_{i-1}^{\mathrm{e}}\Delta\varepsilon_i\delta W_{i-\frac{1}{2}}^{(+)}\right\}\delta_{i-1,j} \\
&- \left\{\psi_{i,w}\Delta\varepsilon_i\Delta\varepsilon_j(c_{i,j} - c_{i-1,j}) - \psi_{i,e}\Delta\varepsilon_{i-1}\Delta\varepsilon_j(c_{i+1,j} - c_{i,j})\right\},
\end{aligned} \tag{A.18}$$

$$b_i \equiv \alpha_i Q_i^{n-1}, \tag{A.19}$$

$$\alpha_i \equiv \frac{\Delta\varepsilon_{i-1}\Delta\varepsilon_i(\Delta\varepsilon_{i-1} + \Delta\varepsilon_i)}{2\Delta t}. \tag{A.20}$$

Here, δ_{ij} is Kronecker's delta. By conducting LU decomposition of A_{ij}, one can obtain time development of Q_i^n. The same scheme can be used for ERS equation by setting the values of $\{c_{ij}\}$ to be zero.

We impose Neumann boundary condition, in which the current at the boundary is zero. In addition to the boundary condition, we omit the direct correlation term because $\langle\rho(\varepsilon)\rangle$ is zero at the boundary. If the current between $(i-1)$th point and ith point is zero, the following equation is satisfied.

$$-D_{i,w}^{\mathrm{e}}\frac{Q_i^n - Q_{i-1}^n}{\Delta\varepsilon_{i-1}} + D_{i-1}^{\mathrm{e}}Q_{i-1}^n\frac{1}{\Delta\varepsilon_{i-1}}\delta W_{i-\frac{1}{2}}^{(+)} + D_i^{\mathrm{e}}Q_i^n\frac{1}{\Delta\varepsilon_{i-1}}\delta W_{i-\frac{1}{2}}^{(-)} = 0. \tag{A.21}$$

Accordingly, the boundary condition is expressed as

$$Q_1^n = \frac{D^{\mathrm{e}}_{2,w} - D^{\mathrm{e}}_2 \delta W^{(-)}_{\frac{3}{2}}}{D^{\mathrm{e}}_{2,w} + D^{\mathrm{e}}_1 \delta W^{(+)}_{\frac{3}{2}}} Q_2^n, \tag{A.22}$$

$$Q_N^n = \frac{D^{\mathrm{e}}_{N,w} + D^{\mathrm{e}}_{N-1} \delta W^{(+)}_{N-\frac{1}{2}}}{D^{\mathrm{e}}_{N,w} - D^{\mathrm{e}}_N \delta W^{(-)}_{N-\frac{1}{2}}} Q_{N-1}^n. \tag{A.23}$$

A.3 Nonlinearlity of the solvation dynamics

In the present study, we utilize the linear response theory to describe the solvation dynamics. The solvent distribution functions on the energy coordinate obtained from ERSV equation and the linear response theory, $\langle \rho(\varepsilon, t) \rangle_{\mathrm{ne}}$, exhibit an unphysical distribution around $\varepsilon \sim 3$ kcal mol^{-1} (Fig. 4(a)). In this section, we show that such a negative distribution stems from a nonlinearity of the solvation dynamics. As shown in Sec. II D with Eqs. (2.53), (2.54) and (2.55), there are three expressions of $Q_v(\varepsilon, t=0)$ in the linear response limit as

$$Q_v^{(1)}(\varepsilon, t=0) = \sum_w \int \mathrm{d}\eta\, \eta \langle \delta\rho_v(\varepsilon)\, \delta\rho_w(\eta) \rangle, \tag{A.24}$$

$$Q_v^{(2)}(\varepsilon, t=0) = \sum_w \int \mathrm{d}\eta\, \eta \langle \delta\rho_v(\varepsilon)\, \delta\rho_w(\eta) \rangle_{\mathrm{ES}}, \tag{A.25}$$

$$Q_v^{(3)}(\varepsilon, t=0) = \frac{1}{\beta} \left\{ \langle \rho_v(\varepsilon) \rangle - \langle \rho_v(\varepsilon) \rangle_{\mathrm{ES}} \right\}. \tag{A.26}$$

In the main text, we employ Eq. (2.53) for evaluating $Q_v(\varepsilon, t=0)$. These three functions agree with one another in the linear response limit. Thus, the difference between them can be interpreted as a nonlinearity of the solvation dynamics. The time development of the solvent distribution, $\langle \rho_v(\varepsilon, t) \rangle_{\mathrm{ne}}$, is given by Eq. (49) as

$$\langle \rho_v(\varepsilon, t) \rangle_{\mathrm{ne}} = \langle \rho_v(\varepsilon) \rangle + \beta \{ Q_v(\varepsilon, t) - Q_v(\varepsilon, t=0) \}, \tag{A.27}$$

and in the limit of $t \to \infty$ one can obtain

$$\langle \rho_v(\varepsilon) \rangle_{\mathrm{ES}} = \langle \rho_v(\varepsilon) \rangle - \beta Q_v(\varepsilon, t=0), \tag{A.28}$$

where we have used the relationships given by

$$\langle \rho_v(\varepsilon, t \to \infty) \rangle = \langle \rho_v(\varepsilon) \rangle_{\mathrm{ES}}, \tag{A.29}$$

$$Q_v(\varepsilon, t \to \infty) = 0. \tag{A.30}$$

While $Q_v^{(3)}(\varepsilon, t=0)$ (Eq. (2.55)) always satisfies Eq. (A.28), $Q_v^{(1)}(\varepsilon, t=0)$ and $Q_v^{(2)}(\varepsilon, t=0)$ satisfy this relationship only if the linear response limit exactly describes the solvation dynamics. The plots of $\langle \rho(\varepsilon) \rangle - \beta Q^{(i)}(\varepsilon, t=0)$ (Fig. A.9) show that the deviations from $\langle \rho(\varepsilon) \rangle_{\mathrm{ES}}$ appear in the cases of $Q^{(1)}(\varepsilon, t=0)$ and $Q^{(2)}(\varepsilon, t=0)$. Furthermore, $\langle \rho(\varepsilon) \rangle - \beta Q^{(1)}(\varepsilon, t=0)$ has a negative distribution around $\varepsilon \sim 3$ kcal mol^{-1} as well as $\langle \rho(\varepsilon, t) \rangle_{\mathrm{ne}}$ obtained from $Q^{(1)}(\varepsilon, t=0)$ (Fig. 4(a)). On the other hand, it should be emphasized that the difference between these expressions is sufficiently small, and hence the linear response limit is essentially valid for the aqueous solution of benzonitrile.

A.4 Analysis of the rotational motions of water around a solute

In order to elucidate the rotational motions of water around a solute, we analyzed the time correlation functions of the vectors associated with the geometry of a water molecule. Let us define the positions of the oxygen and two hydrogen atoms in ith water molecule at time t as $\boldsymbol{r}_{\mathrm{O}}^{(i)}(t)$, $\boldsymbol{r}_{\mathrm{H}_1}^{(i)}(t)$, and $\boldsymbol{r}_{\mathrm{H}_2}^{(i)}(t)$, respectively. Then, the dipole moment of ith water molecule, $\hat{\boldsymbol{\mu}}_i\left(t\right)$, is defined as

$$\hat{\boldsymbol{\mu}}_i\left(t\right) = \sum_{\lambda \in i\text{th water}} q_\lambda \boldsymbol{r}_\lambda^{(i)}\left(t\right), \tag{A.31}$$

where q_λ is the charges on λ atom. The direction of $\hat{\boldsymbol{\mu}}_i(t)$ is on the plane spanned with the three atoms in the water molecule, and a vector normal to the plane, $\hat{\boldsymbol{n}}_i(t)$, is given by

$$\hat{\boldsymbol{n}}_i\left(t\right) = \frac{\boldsymbol{r}_{\mathrm{OH}_1}^{(i)}\left(t\right) \times \boldsymbol{r}_{\mathrm{OH}_2}^{(i)}\left(t\right)}{\left|\boldsymbol{r}_{\mathrm{OH}_1}^{(i)}\left(t\right) \times \boldsymbol{r}_{\mathrm{OH}_2}^{(i)}\left(t\right)\right|}, \tag{A.32}$$

where $\boldsymbol{r}_{\lambda\nu}^{(i)}\left(t\right) = \boldsymbol{r}_{\nu}^{(i)}\left(t\right) - \boldsymbol{r}_{\lambda}^{(i)}\left(t\right)$. The directions of $\hat{\boldsymbol{\mu}}_i$ and $\hat{\boldsymbol{n}}_i$ with respect to a water molecule are illustrated in Fig. A.14. By introducing the distance between benzonitrile (solute) and ith water molecule, r_i, and cutoff distance, r_{cut}, the following time correlation functions for the rotational motions of water molecules in the vicinity of benzonitrile are defined.

$$C_\mu\left(t, r_{\mathrm{cut}}\right) = \frac{\sum_i \left\langle \hat{\boldsymbol{\mu}}_i\left(t\right) \cdot \hat{\boldsymbol{\mu}}_i \Theta\left(r_{\mathrm{cut}} - r_i\left(t\right)\right) \Theta\left(r_{\mathrm{cut}} - r_i\right)\right\rangle}{\sum_i \left\langle \hat{\boldsymbol{\mu}}_i \cdot \hat{\boldsymbol{\mu}}_i \Theta\left(r_{\mathrm{cut}} - r_i\right)\right\rangle}, \tag{A.33}$$

$$C_n\left(t, r_{\mathrm{cut}}\right) = \frac{\sum_i \left\langle \hat{\boldsymbol{n}}_i\left(t\right) \cdot \hat{\boldsymbol{n}}_i \Theta\left(r_{\mathrm{cut}} - r_i\left(t\right)\right) \Theta\left(r_{\mathrm{cut}} - r_i\right)\right\rangle}{\sum_i \left\langle \hat{\boldsymbol{n}}_i \cdot \hat{\boldsymbol{n}}_i \Theta\left(r_{\mathrm{cut}} - r_i\right)\right\rangle}. \tag{A.34}$$

Here, $\Theta(\cdot)$ is Heaviside step function. In the present study, benzonitrile-water distance is defined as the distance between the nitrogen atom of benzonitrile and oxygen atom of water molecule. Note that $C_\mu(t, r_{\mathrm{cut}})$ and $C_n(t, r_{\mathrm{cut}})$ respectively reflect the pitch and roll motions. The time developments of $C_\mu(t, r_{\mathrm{cut}})$ and $C_n(t, r_{\mathrm{cut}})$ at different values of r_{cut} are shown in Fig. A.15 together with the two-dimensional potential of mean force (PMF) along the benzonitrile-water distance, r, and the energy coordinate, ε. It is seen that the damped oscillation is observed in $C_\mu(t, r_{\mathrm{cut}})$ and $C_n(t, r_{\mathrm{cut}})$ at small values of r_{cut}, and its phase and frequency are almost the same as that in $S\left(t\right)$. Hence, the rotational motion of water molecules restricted by the hydrogen bonding to benzonitrile has a vital role in the dynamics on a short timescale.

A.5 Supplementary figures

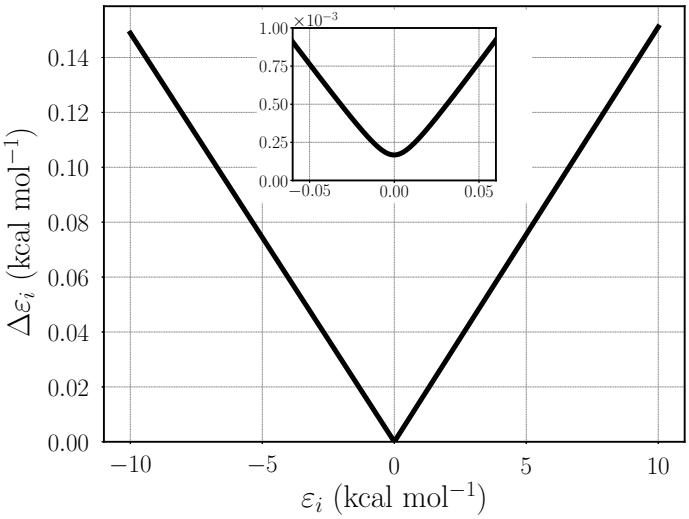


Fig. A.1: Relationship between the discretized energy coordinate ε_i and the grid width $\Delta\varepsilon_i$ in our non-uniform grid.

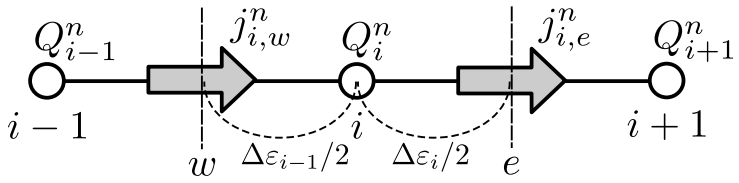


Fig. A.2: Schematic diagram of finite volume method (FVM)

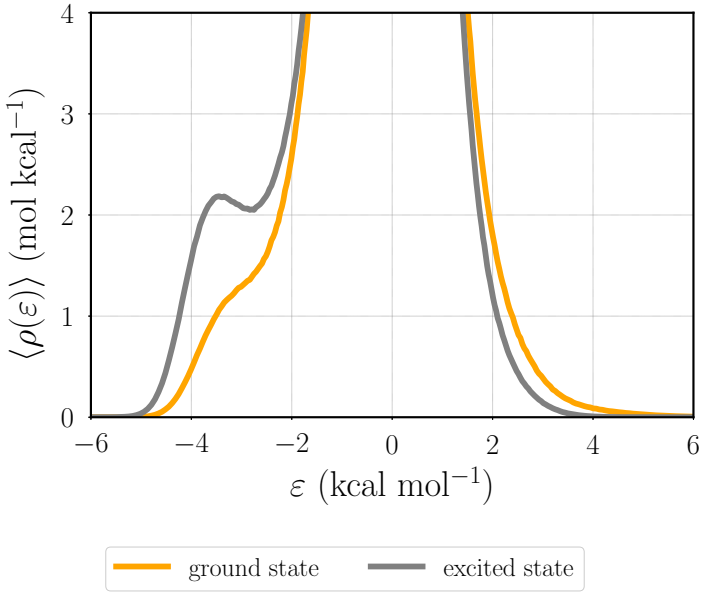
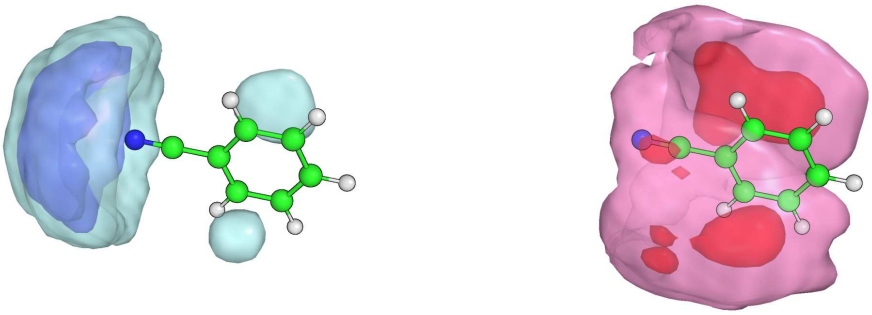


Fig. A.3: Solvent distributions on the energy coordinate at the ground and excited states. The energy coordinate, ε, is defined as the difference of the interaction energy with the solute at the excited state from that at the ground state.

(a) ground state



(b) excited state

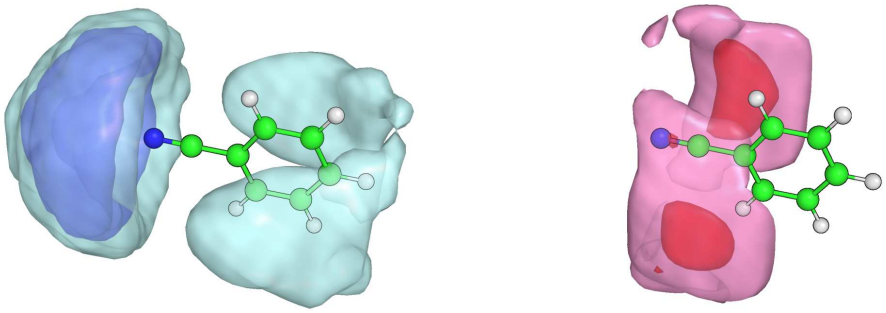


Fig. A.4: Spatial distribution functions (SDFs), $g_S(\boldsymbol{r})$, conditioned by the energy coordinate at (a) the ground state and (b) the excited state. The stabilized region ($\varepsilon \leq -2.8$ kcal mol^{-1}) corresponds to the blue ($g_S(\boldsymbol{r}) = 0.4$) and light blue ($g_S(\boldsymbol{r}) = 0.08$) regions. The destabilized region ($\varepsilon \geq 1.5$ kcal mol^{-1}) corresponds to the red ($g_S(\boldsymbol{r}) = 0.4$) and light red ($g_S(\boldsymbol{r}) = 0.2$) regions.

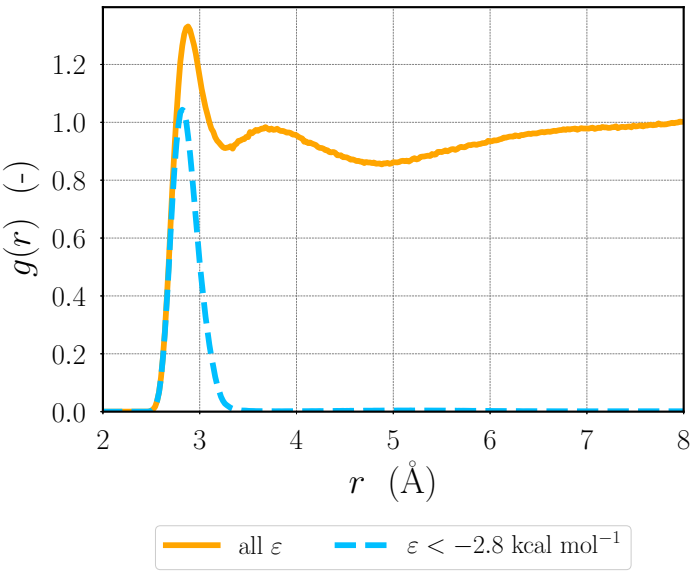


Fig. A.5: Radial distribution function (RDF) conditioned by $\varepsilon < -2.8$ kcal mol^{-1}. In the function, the distance between the nitrogen atom and the center-of-mass (CoM) of water is used. For comparison, the standard RDF is also shown.

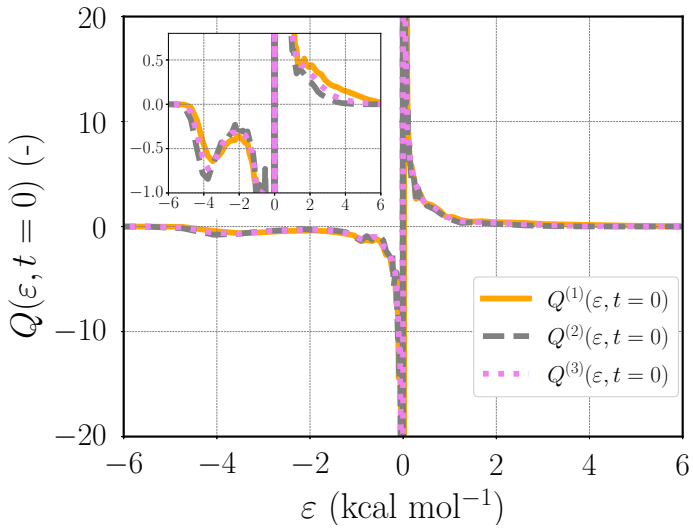


Fig. A.6: Initial values of $Q(\varepsilon, t)$ using different expressions, $Q^{(1)}(\varepsilon, t = 0)$ (Eq. (53)), $Q^{(2)}(\varepsilon, t = 0)$ (Eq. (54)), and $Q^{(3)}(\varepsilon, t = 0)$ (Eq. (55)).

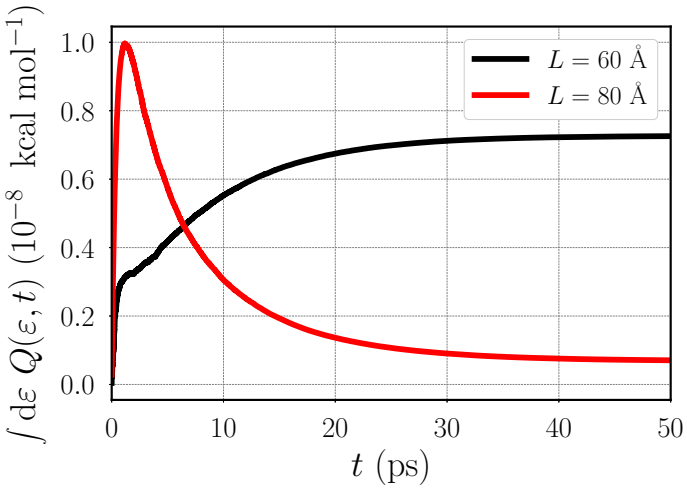


Fig. A.7: Time development of the integral of $Q(\varepsilon, t)$ for ERSV equation over the energy coordinate obtained from our numerical scheme.

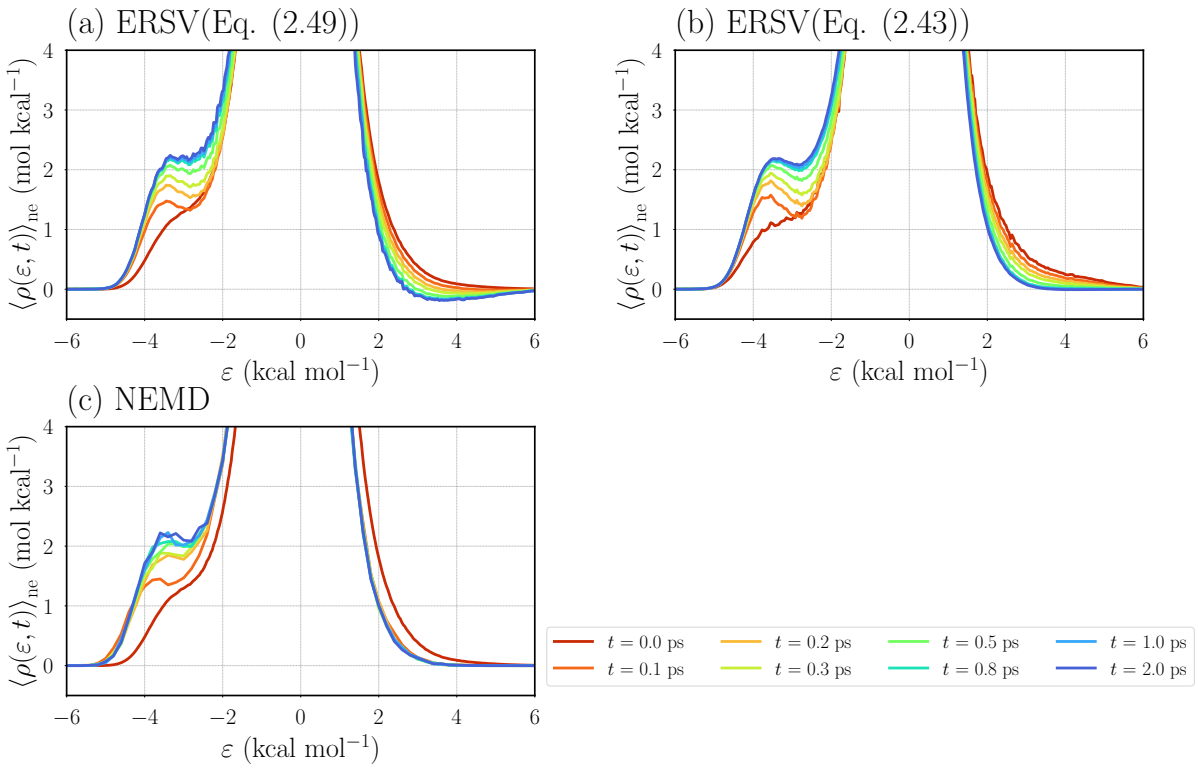


Fig. A.8: Time developments of the solvent distributions on the energy coordinate after the excitation of benzonitrile obtained from (a) ERSV equation with Eq. (2.49), (b) ERSV equation with Eq. (2.43), and (c) nonequilibrium MD (NEMD) simulation. (a) and (c) are identical to Figs. 4(a) and 4(b), respectively.

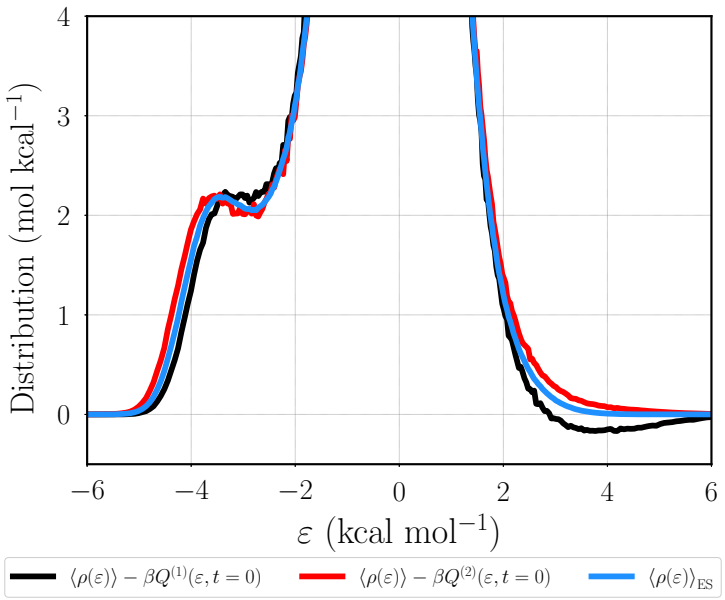


Fig. A.9: Equilibrium solvent distributions at the excited state using different expressions, $\langle\rho(\varepsilon)\rangle - \beta Q^{(1)}(\varepsilon, t=0)$, $\langle\rho(\varepsilon)\rangle - \beta Q^{(2)}(\varepsilon, t=0)$, and $\langle\rho(\varepsilon)\rangle_{\mathrm{ES}}$ (exact). Note that $\langle\rho(\varepsilon)\rangle_{\mathrm{ES}}$ is equivalent to $\langle\rho(\varepsilon)\rangle - \beta Q^{(3)}(\varepsilon, t=0)$.

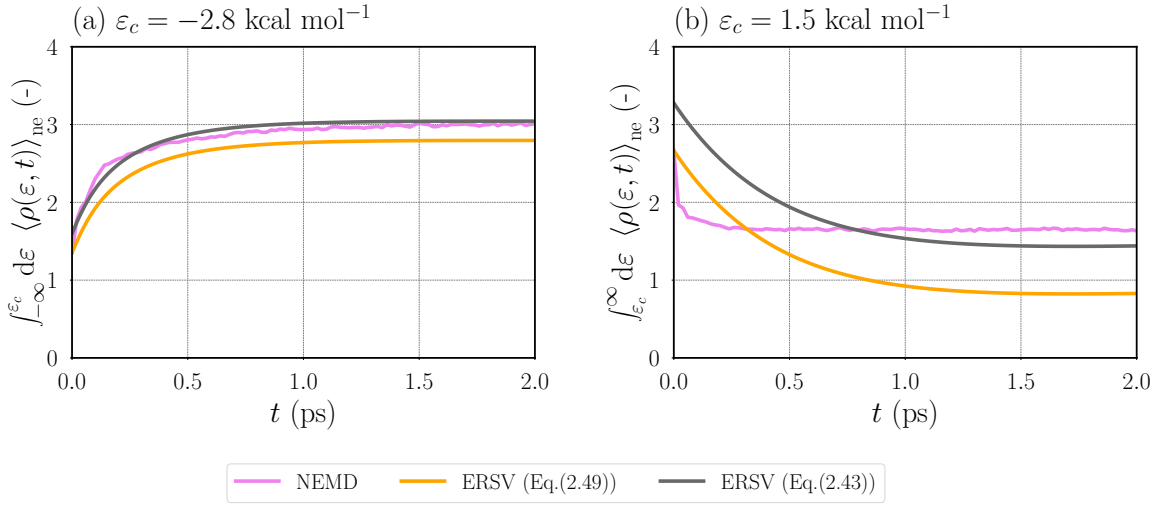


Fig. A.10: Time developments of (a) the number of stabilized water molecules ($\varepsilon \leq -2.8$ kcal mol^{-1}) and (b) the number of destabilized water molecules ($\varepsilon \geq 1.5$ kcal mol^{-1}).

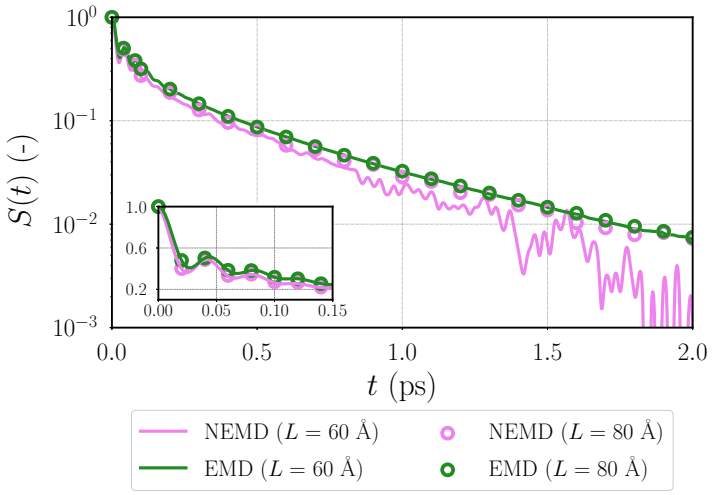


Fig. A.11: Solvation time correlation functions (STCFs) obtained from NEMD simulation with Eq. (36) and equilibrium MD (EMD) simulation at the ground state with Eq. (39).

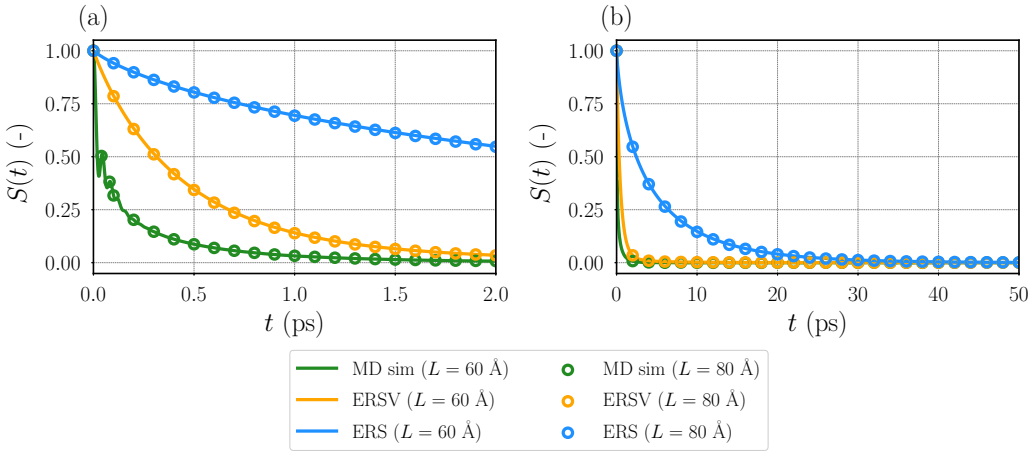


Fig. A.12: Solvation time correlation functions in the time-range of (a) 0 - 2.0 ps and (b) 0 - 50 ps.

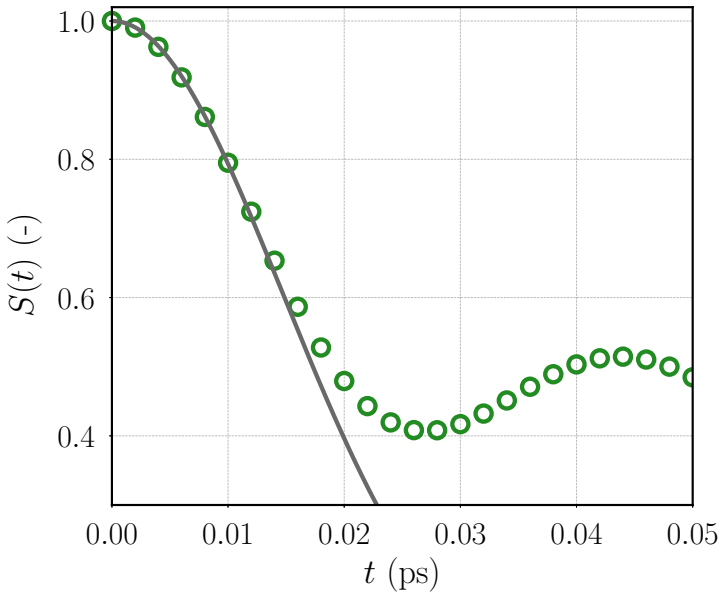


Fig. A.13: Solvation time correlation function in the short-time range (L = 60 Å) in the equilibrium MD with the ground-state benzonitrile. The solid line represents the curve fitted to Gaussian in the time range of $0 \leq t/\mathrm{ps} \leq 0.01$.

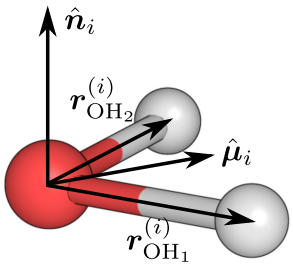


Fig. A.14: Illustrations of dipole vector $\boldsymbol{\mu}$ and normal vector $\boldsymbol{n}$ of water.

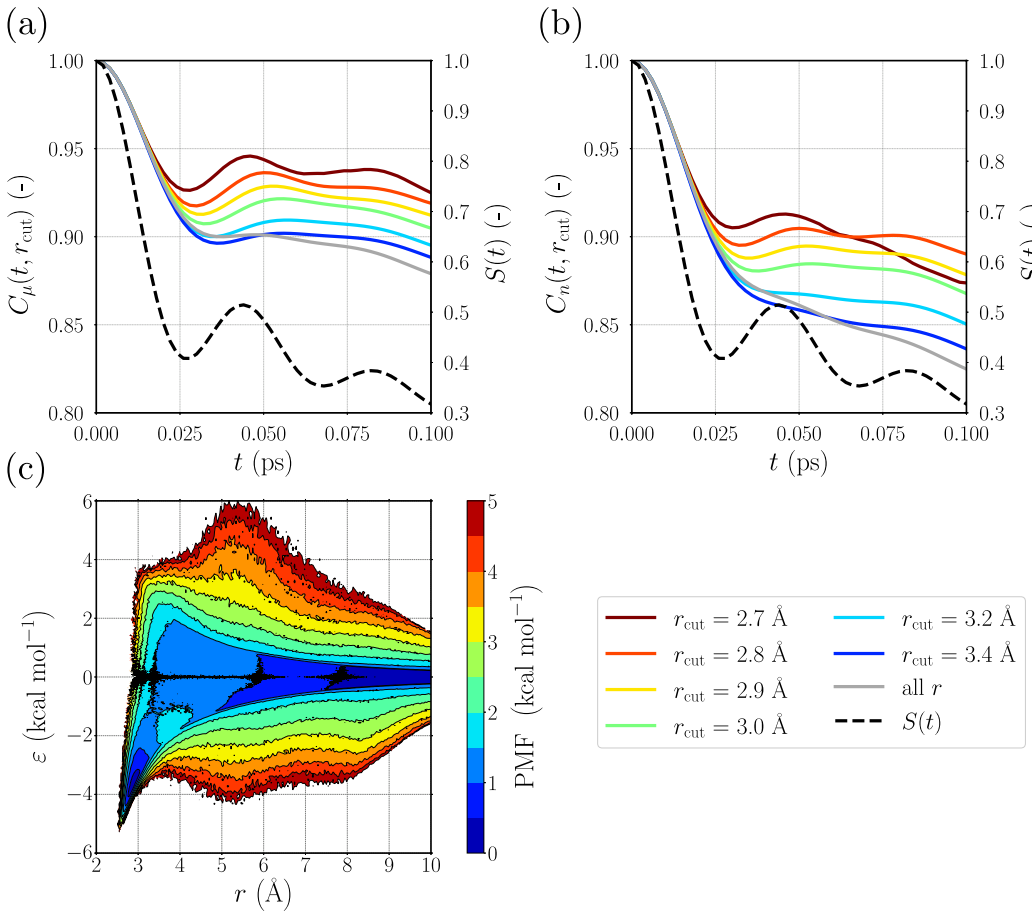


Fig. A.15: Time correlation functions of (a) dipole vector, (b) normal vector of water conditioned by the benzonitrile-water distance (r) and (c) potential of mean force (PMF) of the energy coordinate (ε). The definition of r is the distance between the nitrogen atom of benzonitrile and oxygen atom of water.

Chapter 3

Solvation dynamics on the diffusion timescale elucidated using energy-represented dynamics theory

3.1 Introduction

Solute-solvent interaction is vital for the mass transport as well as the chemical reaction of a solute. The friction associated with the solute motion is affected by such an interaction in addition to the solvent diffusivity. [3,28,141] The motion of the solvent molecules around a solute is also different from the simple diffusion in the bulk phase due to the inhomogeneous environment produced by a solute. [93] Photoexcitation phenomena give us useful information on the solute-solvent interaction through the spectroscopy techniques. The local solvent environments can be elucidated by the the fluorescence spectra of solvent-sensitive probes. The spectra are often characterized with the physicochemical concepts such Stokes shifts and polarity. [90, 91] Furthermore, the solvent response to the change in the electronic structure of a solute by the photoexcitation, referred to as solvation dynamics or dynamic Stokes shift, reflects the dynamic properties of solvents. [93, 142, 143] The solvation dynamics can be utilized for investigating the heterogeneous environments such as the ionic liquids [144–146], biomolecular solutions [147, 148] and biological membranes [94, 149, 150].

The theoretical and computational methods are useful for obtaining the atomistic information on the photoexcitation dynamics in solutions which is difficult to be accessed from the experiments. Molecular dynamics (MD) simulation is a representative method to obtain such information. [151, 152] The solvation dynamics of small solute molecules such as coumarin has been extensively studied by means of MD simulations so far. [143,153–155] The recent advance in computers enables us to investigate the solvation dynamics of macromolecules such as deoxyribonucleic acids (DNA) [156] and proteins [157] in aqueous solutions. Ab initio MD (AIMD) simulation [158] is also a powerful tool for realistically describing solvation dynamics. [159, 160] For instance, the application of this method to the solvation dynamics of N-methyl-6-oxyquinolinium betaine in water revealed that the dominant mode for the solvent motion varies depending on the timescale and that the collective solvent rearrangement plays an important role on a picosecond timescale. [160] As well as in the case of the spectroscopy experiments, the solvation dynamics in these studies is characterized in terms of the solvation time correlation functions (STCF). However, the accurate computation of the time correlation functions on the long timescale

often suffers from the insufficient sampling.

The dynamics theories based on the integral equation theory of molecular liquids with the interaction-site representation [1, 5, 6] have been also widely used to investigate the solvation dynamics of the various solutes in molecular liquids. [7, 45, 46, 48, 51–53, 53–55, 161] The governing equations have a mathematically closed form for both the static and dynamic properties of solutions, and hence the information on the dynamics can be obtained by solving the equations in an algebraic manner, meaning that the obtained results are free from the sampling problem. [34, 40, 44, 59, 107] The application of these theories to the solvation dynamics is possible with the aid of a modified linear response theory [45, 46] or time-dependent density functional theory (TD-DFT) of liquids. [57–59, 93] Due to the orientational average of solvent molecules introduced for avoiding the explicit treatment of the orientational degrees of freedom, however, the theories with the interaction-site representation are applicable only to the simple polyatomic solvents. In addition, the rigidity of the molecules is assumed in these theories. Note that the flexibility of molecules is known to have an impact on the dynamic behavior of solutions. [162, 163]

Recently, we have derived an alternative expression of the molecular diffusion based on the energy representation (ER) which is amenable to MD simulations. [164] In this theory, the configuration of a solvent molecule is projected onto the solute-solvent interaction energy, namely energy coordinate. [60, 61, 64] It should be noted that the intramolecular degrees of freedom is naturally taken into account using the energy coordinate. This dimensionality reduction also enables us to construct the one-dimensional free-energy functional for describing the solvation thermodynamics in complex solutions including polymer solutions [73] and lipid membrane systems [68, 69] by means of MD simulations. By applying the Zwanzig-Mori projection operator method to the solvent distribution on the energy coordinate, the energy-represented generalized Langevin equation (ERGLE) can be derived in an exact way. [164] The systematic approximations about the dynamics such as the overdamped limit give the energy-represented Smoluchowski-Vlasov (ERSV) equation, a diffusion equation of solution that describes the self-diffusion, drift motion, and collective motion of solvents on the energy coordinate. Once we calculate the static correlation functions and diffusion coefficient of solvents involved in the ERSV equation, the time development of the solvent distribution functions can be obtained by solving this equation. Since the sampling of the information on the dynamics is not required in this treatment, the analysis of the long-timescale dynamics is realized without the sampling problem. It is confirmed that the ERSV equation reproduces the relaxation time coefficient on the diffusion timescale for the solvation dynamics of benzonitrile in water. Furthermore, we revealed the importance of the collective diffusion on the solvent relaxation on the intermediate timescale.

In the present study, we apply the ERSV equation to the solvation dynamics of 6-propionyl-2-dimethylamino naphthalene (Prodan) in water and alcohol solvents (methanol, ethanol, and 1-propanol) for clarifying the differences of the relaxation processes among these solvents. Prodan is a solvent-sensitive fluorescence probe that exhibits the significant Stokes shifts, making it possible to analyze the local environments of the systems of interests. [165–170] The excited states of Prodan are well characterized with the quantum chemical calculations [171–174] Prodan is widely used for analyzing the membrane properties such as local polarity and gel phase transition. [175–178] The time-resolved infrared (IR) spectroscopy analysis of the solvent-dependent feature of the excited state for Prodan probed that the S_1 ($\pi - \pi^*$) state undergoing a solvent-driven charge redistribution from dimethylamine (DMA) to carbonyl (C=O) group dynamically alters the solvation

structure such as the patterns in hydrogen-bonding of Prodan with the surrounding solvent molecules, suggesting the importance of the atomistic description of the solvation dynamics. [170] Very recently, the systematic analysis of the membrane properties at the interfacial region was proposed based on the fluorescence decays of Prodan in a series of solvent mixtures. [179] The microscopic information on the solvation dynamics in various solvents could thus be useful to deepen our understanding of heterogeneous environments.

We focus on the solvation dynamics of Prodan triggered by S_1 ($\pi - \pi^*$) excitation described by means of the TD-DFT of electronic structures. The solvation time correlation correlation function (STCF) for each solvent is calculated with the ERSV equation based on the MD simulations with Prodan in its ground state. Furthermore, we introduce a scheme of decomposing the diffusion coefficients in the energy representation into the contributions of the moieties in Prodan to unveil the difference of the relaxation processes depending on the solvent species.

3.2 Methods

3.2.1 Energy-represented Smoluchowski-Vlasov (ERSV) equation

In this section, we briefly summarize the energy-represented dynamics theory and its application to the solvation dynamics. [164] The energy representation means that the solvent configuration around a solute is projected onto the solute-solvent pair interaction energy, energy coordinate. Then, the dynamic processes of solvents are represented as the time development of the solvent distribution on the energy coordinate in the theory. This treatment enables us to effectively treat the solvent position and orientation on one-dimensional space.

Let us consider a dilute solution containing a solute molecule in a single-component solvent. The formulation for multi-component solvents is available in Ref. [164]. Furthermore, we assume that the solute molecule is fixed in space. We define the full coordinate (position and orientation with the intramolecular degrees of freedom) of the ith solvent molecule as $\boldsymbol{x}_i$. The instantaneous solvent distribution on the energy coordinate (energy distribution), $\rho(\varepsilon, t)$, is defined as,

$$\rho(\varepsilon, t) = \sum_i \delta(u(\boldsymbol{x}_i(t)) - \varepsilon), \tag{3.1}$$

where $u(\cdot)$ is an energy function between the solute and solvent (defining potential). The fluctuation of $\rho(\varepsilon, t)$ is also defined as

$$\delta\rho(\varepsilon, t) = \rho(\varepsilon, t) - \langle\rho(\varepsilon)\rangle\,, \tag{3.2}$$

where $\langle\cdots\rangle$ represents the ensemble average at the equilibrium state. The Zwanzig-Mori projection operator method gives the energy-represented generalized Langevin equation (ERGLE), an exact partial differential equation of $\delta\rho(\varepsilon, t)$. Imposing the overdamped limit on the ERGLE, one can derive the energy-represented Smoluchowski-Vlasov (ERSV)

equation given by

$$\frac{\partial \delta\rho(\varepsilon,t)}{\partial t} = \frac{\partial}{\partial \varepsilon}\left[D^{\mathrm{e}}(\varepsilon)\frac{\partial \delta\rho(\varepsilon,t)}{\partial \varepsilon} - D^{\mathrm{e}}(\varepsilon)\frac{\mathrm{d}\ln\langle\rho(\varepsilon)\rangle}{\mathrm{d}\varepsilon}\delta\rho(\varepsilon,t) - D^{\mathrm{e}}(\varepsilon)\langle\rho(\varepsilon)\rangle\int \mathrm{d}\eta\ \frac{\partial c(\varepsilon,\eta)}{\partial \varepsilon}\delta\rho(\eta,t)\right] + F(\varepsilon,t). \tag{3.3}$$

Here, $D^{\mathrm{e}}(\varepsilon)$ and $F(\varepsilon,t)$ represent the energy-represented diffusion coefficient and fluctuating force, respectively. $c\left(\varepsilon,\eta\right)$ is the direct correlation function that describes the relationship between the solvent molecules whose values of the defining potential are ε and η. The approximate expression of $D^{\mathrm{e}}(\varepsilon)$ is given by

$$D^{\mathrm{e}}(\varepsilon) = D\left\langle \left|\, \boldsymbol{f}_i^{\mathrm{G}} \,\right|^2 \right\rangle_{\varepsilon}, \tag{3.4}$$

where D and $\boldsymbol{f}_i^{\mathrm{G}}$ are the translational diffusion coefficient of the solvent and the force acting on the center of mass (CoM) of the ith solvent molecule corresponding to the defining potential. $\langle\cdots\rangle_{\varepsilon}$ means the ensemble average conditioned by the energy coordinate ε defined as

$$\langle\cdots\rangle_{\varepsilon} = \frac{1}{\langle\rho(\varepsilon)\rangle}\sum_i \left\langle(\cdots)\,\delta(u(\boldsymbol{x}_i)-\varepsilon)\right\rangle. \tag{3.5}$$

Thanks to the additivity of the solute-solvent interaction, $\boldsymbol{f}_i^{\mathrm{G}}$ can be decomposed into the forces acting on the moieties (m) of the solute molecule, $\boldsymbol{f}_i^m$, as

$$\boldsymbol{f}_i^{\mathrm{G}} = \sum_m^{\text{all moieties}} \boldsymbol{f}_i^m. \tag{3.6}$$

Thus, $D^{\mathrm{e}}\left(\varepsilon\right)$ can be rewritten as

$$D^{\mathrm{e}}\left(\varepsilon\right) = \sum_m^{\text{all moieties}} D^{\mathrm{e},m}\left(\varepsilon\right), \tag{3.7}$$

$$D^{\mathrm{e},m}(\varepsilon) = D\left\langle \boldsymbol{f}_i^m \cdot \boldsymbol{f}_i^{\mathrm{G}}\right\rangle_{\varepsilon}. \tag{3.8}$$

Note that $D^{\mathrm{e},m}\left(\varepsilon\right)$ can be regarded as the contribution of moiety m to $D^{\mathrm{e}}\left(\varepsilon\right)$.

Each term in the ERSV equation (Eq. (3.3)) has a clear physical meaning. The first term in the square bracket of Eq. (3.3) describes the simple diffusion of solvents due to the gradient of the solvent distribution. Since $-\ln\langle\rho\left(\varepsilon\right)\rangle$ multiplied by the inverse temperature is the free energy profile on the energy coordinate, the drift motion caused by the free energy gradient is expressed with the second term. The third term describes the collective diffusion through the direct correlation function, $c\left(\varepsilon,\eta\right)$. If the collective term is neglected in the ERSV equation (Eq. (3.3)), one can obtain the energy represented Smoluchowski equation (ERS) describing the single-particle diffusion process as

$$\frac{\partial \delta\rho(\varepsilon,t)}{\partial t} = \frac{\partial}{\partial \varepsilon}\left[D^{\mathrm{e}}(\varepsilon)\frac{\partial \delta\rho(\varepsilon,t)}{\partial \varepsilon} - D^{\mathrm{e}}(\varepsilon)\frac{\mathrm{d}\ln\langle\rho(\varepsilon)\rangle}{\mathrm{d}\varepsilon}\delta\rho(\varepsilon,t)\right] + F(\varepsilon,t). \tag{3.9}$$

The solvation dynamics triggered by the sudden change of a solute molecule can be described based on the ERSV equation and linear response theory. [30,124] Let us consider an nonequilibrium process that the solute-solvent pair interaction energy is changed from $u^{\mathrm{g}}(\boldsymbol{x}_i)$ to $u^{\mathrm{ex}}(\boldsymbol{x}_i)$ due to the photoexcitation of a solute at $t = 0$. If we approximate that the intramolecular energy of the solute is unchanged during the relaxation process, the solvation time correlation function (STCF) is expressed as

$$S(t) = \frac{\langle \Delta E(t) \rangle_{\mathrm{ne}} - \langle \Delta E(\infty) \rangle_{\mathrm{ne}}}{\langle \Delta E(0) \rangle_{\mathrm{ne}} - \langle \Delta E(\infty) \rangle_{\mathrm{ne}}}, \tag{3.10}$$

where $\langle \cdots \rangle_{\mathrm{ne}}$ is the nonequilibrium ensemble average and

$$\Delta E(t) = \sum_i \left\{ u^{\mathrm{ex}}(\boldsymbol{x}_i) - u^{\mathrm{g}}(\boldsymbol{x}_i) \right\}. \tag{3.11}$$

Next, we introduce a scheme for describing the solvation dynamics. The solvation dynamics is characterized with the solvation time correlation function (STCF), $S(t)$. Based on the linear response theory, $S(t)$ can be expressed as

$$S(t) = \frac{\langle \delta\Delta E(t) \delta\Delta E \rangle}{\langle \delta\Delta E \delta\Delta E \rangle}. \tag{3.12}$$

Here, $\langle \cdots \rangle$ is the ensemble average at the ground state and $\delta\Delta E(t)$ is the fluctuation of $\Delta E(t)$ defined as $\delta\Delta E(t) = \Delta E(t) - \langle \Delta E \rangle$. If the defining potential is the difference of the solute-solvent pair interaction energies between the excited and ground states as $u(\boldsymbol{x}_i) = u^{\mathrm{ex}}(\boldsymbol{x}_i) - u^{\mathrm{g}}(\boldsymbol{x}_i)$, $\delta\Delta E(t)$ and its autocorrelation function can be written as

$$\delta\Delta E(t) = \int \mathrm{d}\varepsilon \; \varepsilon \delta\rho(\varepsilon, t), \tag{3.13}$$

$$\begin{aligned} \langle \delta\Delta E(t) \delta\Delta E \rangle &= \iint \mathrm{d}\varepsilon \mathrm{d}\eta \; \varepsilon\eta \langle \delta\rho(\varepsilon, t) \delta\rho(\eta) \rangle \\ &= \int \mathrm{d}\varepsilon \; \varepsilon Q(\varepsilon, t), \end{aligned} \tag{3.14}$$

where we have introduced a new function $Q(\varepsilon, t)$ defined as

$$Q(\varepsilon, t) = \int \eta \langle \delta\rho(\varepsilon, t) \delta\rho(\eta) \rangle \, \mathrm{d}\eta = \langle \delta\rho(\varepsilon, t) \delta\Delta E \rangle. \tag{3.15}$$

Substituting Eq. (3.14) into Eq. (3.12) yields

$$S(t) = \frac{\displaystyle\int \mathrm{d}\varepsilon \; \varepsilon Q(\varepsilon, t)}{\displaystyle\int \mathrm{d}\varepsilon \; \varepsilon Q(\varepsilon, t = 0)}. \tag{3.16}$$

By using Eq. (3.3), one can obtain the ERSV equation for $Q(\varepsilon, t)$ as

$$\begin{aligned} \frac{\partial Q(\varepsilon, t)}{\partial t} = \frac{\partial}{\partial \varepsilon} \Bigg[& D^{\mathrm{e}}(\varepsilon) \frac{\partial Q(\varepsilon, t)}{\partial \varepsilon} - D^{\mathrm{e}}(\varepsilon) \frac{\mathrm{d} \ln \langle \rho(\varepsilon) \rangle}{\mathrm{d}\varepsilon} Q(\varepsilon, t) \\ & - D^{\mathrm{e}}(\varepsilon) \langle \rho(\varepsilon) \rangle \int \mathrm{d}\eta \; \frac{\partial c(\varepsilon, \eta)}{\partial \varepsilon} Q(\eta, t) \Bigg]. \end{aligned} \tag{3.17}$$

The ERS equation for $Q(\varepsilon, t)$ can be also derived from Eq. (3.9) as

$$\frac{\partial Q(\varepsilon,t)}{\partial t} = \frac{\partial}{\partial \varepsilon}\left[D^{\mathrm{e}}(\varepsilon)\frac{\partial Q(\varepsilon,t)}{\partial \varepsilon} - D^{\mathrm{e}}(\varepsilon)\frac{\mathrm{d}\ln\langle\rho(\varepsilon)\rangle}{\mathrm{d}\varepsilon}Q(\varepsilon,t)\right]. \tag{3.18}$$

The STCF can be computed by solving the ERSV or ERS equation under the following initial condition.

$$Q(\varepsilon, t=0) = \int \langle \delta\rho(\varepsilon)\,\delta\rho(\eta)\rangle\, d\eta. \tag{3.19}$$

3.2.2 Computational Details

System modeling

The structure of Prodan (Fig. 3.1) was obtained by performing the geometry optimization with CAM-B3LYP/cc-pVDZ level calculation [180] at the ground state. Then, the S_1 ($\pi \rightarrow \pi^*$) excited state was computed using TD-DFT (CAM-B3LYP)/cc-pVDZ calculation. Charges from electrostatic potentials using a grid based method (CHelpG) [181] was used for calculating the atomic point charges for both the ground and excited states (Table S1, Fig. S1, ESI). All the quantum chemical calculations were performed with Gaussian16. [125]

We prepared four different solution systems consisting of one Prodan molecule and solvent molecules, water, methanol (MeOH), ethanol (EtOH), and 1-propanol (PrOH). The force field for Prodan, MeOH, EtOH, and PrOH was the CHARMM generalized force field (CGenFF) [182] and the parameters were obtained using CHARMM-GUI server, [183] while the atomic charges on Prodan for the ground and excited states were evaluated using the quantum chemical calculations mentioned above. CHARMM-compatible TIP3P model was used for water. [184] The numbers of solvent molecules were 7200, 3100, 2210, and 1690 for water, methanol, ethanol, and 1-propanol systems, respectively. For all the systems, the initial configurations were prepared using Packmol [127] with the cubic box whose volume is 60^3 Å^3. We also prepared the pure solvent systems for calculating the diffusion coefficients of the solvents. The numbers of solvent molecules and the volume are the same as those of corresponding solution systems.

Simulation setups

For each solution system with Prodan in its ground state, we performed three types of NVT simulations: (i) equilibration, (ii) sampling of the system configurations and (iii) production simulations started from the configurations obtained from (ii). For equilibration (i), the MD simulations were performed for 1 ns. Then, we conducted the simulation (ii) for 1 ns to extract the configurations every 1 ps (The total number of the samples was 1000 for each system). After re-distributing the velocities of the sampled configurations so as to generate the Maxwell-Boltzmann distribution, we performed the simulations (iii) for 0.1 ns equilibration, followed by 1 ns production simulation for each trajectory. As for each pure-solvent system, we performed 0.4 ns production simulation after 1 ns simulation for equilibration.

For all the simulations, the equation of motion was integrated using the velocity Verlet algorithm [128] with a time interval of 2 fs. The temperature was set at 300 K with the Bussi thermostat. [129] The Prodan molecule was fixed in space by making the velocities of its atoms zero. The Lennard-Jones (LJ) interaction was truncated by applying

the switching function, with the switching range of 10–12 Å. The smoothed particle mesh Ewald (SPME) method [185,186] was used to calculate the electrostatic potential. Water molecules were kept rigid using the SETTLE algorithm [130] and the bonds involving the hydrogen atoms were fixed with SHAKE/RATTLE algorithm. [187, 188] All the simulations were performed with GENESIS 2.0. [131–133] All the analyses were performed using in-house Fortran90/95 programs combined with the visual molecular dynamics (VMD) package (ver. 1.9.3), [134] PyMOL, [137] and ERmod 0.3.7. [116]

3.2.3 Solver for ERSV and ERS equations

To solve the ERSV and ERS equation, we used the scheme developed in our previous study. [164] We used the finite volume method (FVM) to discretize the energy coordinate. For numerical efficiency and accuracy, we used non-uniform grids on the energy coordinate which are fine around $\varepsilon = 0$. The drift term of these equations was discretized by the 1st-order upwind difference scheme. The full implicit algorithm was employed to integrate these equation for the numerical stability. The time grid Δt was set to be 1 fs. The translational diffusion coefficients used as the inputs of the ERSV equations (Table 3.1) were calculated from the mean square displacements (MSD) of the solvent molecules in the pure solvent systems.

Table 3.1: Translational diffusion coefficients of the solvents (D) obtained from the MD simulations for the pure solvent systems. These values are used as the inputs of the ERSV equation.

	Water	Methanol	Ethanol	1-propanol
D ($10^{-5}\,\mathrm{cm^2\,s^{-1}}$)	5.91	3.28	1.51	0.94

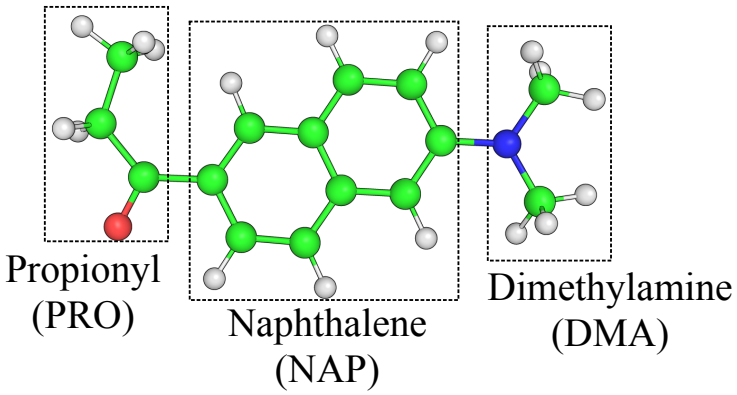


Fig. 3.1: Structure of Prodan. The hydrogen, carbon, nitrogen and oxygen atoms are depicted in gray, green, blue and red, respectively.

3.3 Results and Discussion

3.3.1 Distribution functions on the energy coordinate

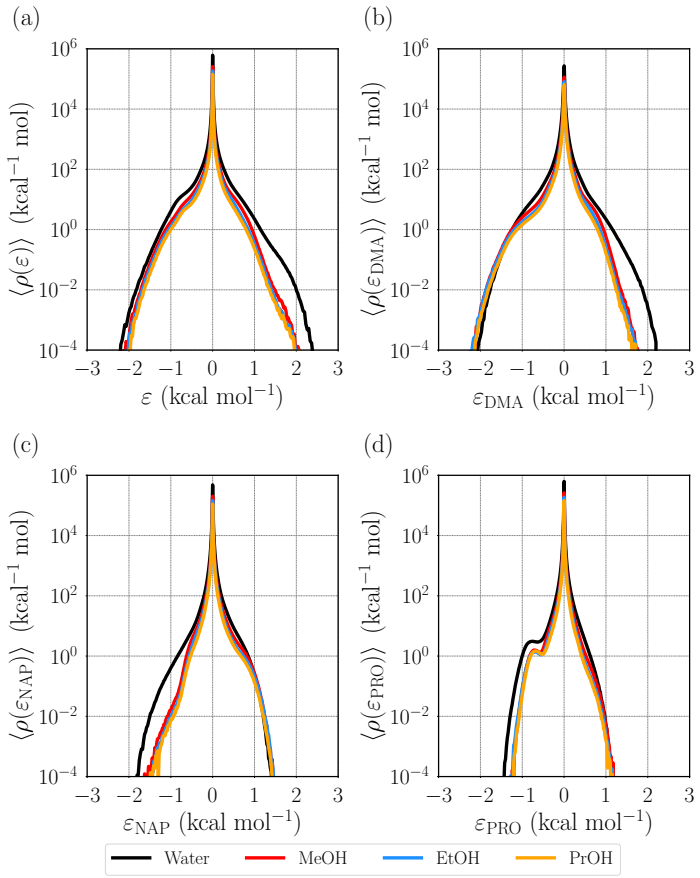


Fig. 3.2: Solvent distribution functions on the energy coordinate (energy distribution distribution function), $\langle\rho(\varepsilon)\rangle$. The distribution functions on the decomposed defining potentials based on the moieties (m) of Prodan, $\langle\rho(\varepsilon_m)\rangle$, are also shown. (a) $\langle\rho(\varepsilon)\rangle$, (b) $\langle\rho(\varepsilon_{\mathrm{DMA}})\rangle$, (c) $\langle\rho(\varepsilon_{\mathrm{NAP}})\rangle$, and (d) $\langle\rho(\varepsilon_{\mathrm{PRO}})\rangle$. DMA, NAP, and PRO respectively denote N,N-dimethylamine, naphtalene, and propionyl moieties of Prodan (Fig. 3.1).

We first examine the energy distribution function, $\langle\rho(\varepsilon)\rangle$, for different solvent systems, water, methanol (MeOH), ethanol (EtOH), and 1-propanol (PrOH) (Fig. 3.2 (a)). The defining potential, u, is defined as the difference in the solute-solvent pair interaction energy between the excited and ground states. Hence, the distributions at $\varepsilon < 0$ and $\varepsilon > 0$ respectively correspond to the stabilized and destabilized solvent molecules due to the excitation of Prodan. The sharp peak at $\varepsilon \sim 0$ observed for all the systems comes from the bulk solvent molecules that are not interacted with Prodan. It is seen that the shapes of the energy distributions in the different solvents are similar, although the populations of both the stabilized and destabilized molecules for water are slightly higher than those for the alcohol solvents. The distribution becomes broader in the order of water > MeOH > EtOH > PrOH. This ordering coincides with ascending order of the solvent polarity. For further analysis, we decompose the defining potential (u) into the contribution from the moieties of Prodan as N,N-dimethylamine (u_{DMA}), naphthalene

(u_{NAP}), and propionyl (u_{PRO}) moieties (Fig. 3.1).

$$u = u_{\mathrm{DMA}} + u_{\mathrm{NAP}} + u_{\mathrm{PRO}}. \tag{3.20}$$

The distribution functions on the decomposed defining potentials defined as

$$\langle \rho\left(\varepsilon_m\right)\rangle = \sum_i \langle \delta\left(u_m\left(\boldsymbol{x}_i\right) - \varepsilon_m\right)\rangle, \quad m = \mathrm{DMA}, \mathrm{NAP}, \mathrm{PRO}, \tag{3.21}$$

are shown in Figs. 3.2(b)-(d). The profile of $\langle \rho\left(\varepsilon_{\mathrm{DMA}}\right)\rangle$ reveals that the water molecules are more destabilized by DMA than other solvent molecules (Fig. 3.2(b)). From the spatial distribution functions (SDFs) corresponding to the destabilized region (Fig. S2, ESI), it is confirmed that the destabilized water molecules are distributed around DMA. On the other hand, the water molecules are more stabilized by NAP and PRO than the other solvent molecules (Figs. 3.2(c) and (d), and Fig. S3, ESI). $\langle \rho\left(\varepsilon_{\mathrm{PRO}}\right)\rangle$ has a small peak around -0.8 kcal mol^{-1}, while the other distribution functions change monotonically at $\varepsilon_m < 0$ for all the solvents. The radial densities around the carbonyl oxygen of Prodan indicate that the carbonyl group forms the hydrogen bonds with the hydroxyl group of the solvent molecules (Fig. S4, ESI). The oxygen atom of Prodan becomes more negative upon excitation ($-0.441e \rightarrow -0.456e$). Thus, the peak in $\langle \rho(\varepsilon_{\mathrm{PRO}})\rangle$ around -0.8 kcal mol^{-1} stems from the strengthened hydrogen bonding by the excitation.

3.3.2 Diffusivity on the energy coordinate

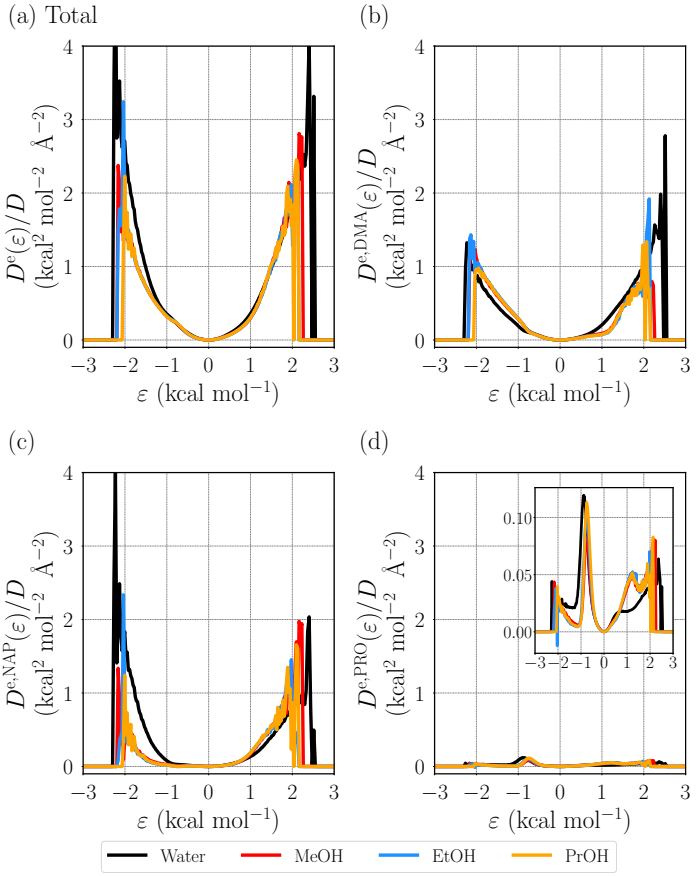


Fig. 3.3: (a) Energy-represented diffusion coefficients scaled with the translational diffusion coefficients, $D^{\mathrm{e}}(\varepsilon)/D$, and their decomposition into the contributions from the moieties (m) of Prodan, $D^{\mathrm{e},m}(\varepsilon)/D$. (a) $D^{\mathrm{e}}(\varepsilon)/D$, (b) $D^{\mathrm{e,DMA}}(\varepsilon)/D$, (c) $D^{\mathrm{e,NAP}}(\varepsilon)/D$, and (d) $D^{\mathrm{e,PRO}}(\varepsilon)/D$. DMA, NAP, and PRO respectively denote N,N-dimethylamine, naphthalene, and propionyl moieties of Prodan (Fig. 3.1).

The energy-represented diffusion coefficients, $D^{\mathrm{e}}(\varepsilon)$, are calculated using Eq. (3.4). $D^{\mathrm{e}}(\varepsilon)$ can be expressed in terms of the translational diffusion coefficient of the solvent, D, and the force associated with the defining potential acting on the solvent molecule, $\boldsymbol{f}_i^{\mathrm{G}}$. Since $D^{\mathrm{e}}(\varepsilon)/D = \left\langle \left| \boldsymbol{f}_i^{\mathrm{G}} \right|^2 \right\rangle_{\varepsilon}$ is a static correlation function, Eq. (3.4) realizes the decomposition of the diffusivity on the energy coordinate into the dynamic contribution (D) and static contribution ($\left\langle \left| \boldsymbol{f}_i^{\mathrm{G}} \right|^2 \right\rangle_{\varepsilon}$). As shown in Table 3.1, water shows the highest diffusivity. In the case of the alcohol solvents, the translational diffusivity becomes low as the molecular size increases. For all the systems, $D^{\mathrm{e}}(\varepsilon)/D$ has a minimum at $\varepsilon \sim 0$ (Fig. 3.3(a)). Since $\boldsymbol{f}_i^{\mathrm{G}}$ is negligibly small for the solvent molecules in the bulk that has a dominant population at $\varepsilon \sim 0$, the appearance of such a minimum is a typical behavior of $D^{\mathrm{e}}(\varepsilon)/D$. Interestingly, all the examined alcohol solvents show almost identical profiles of $D^{\mathrm{e}}(\varepsilon)/D$, indicating that the difference in the diffusivity on the energy coordinate among the alcohol solvents dominantly originates from the translational diffusion coefficients. While the profile of $D^{\mathrm{e}}(\varepsilon)/D$ for water is close to those for the alcohol solvents at $\varepsilon > -1$ kcal mol^{-1}, it exhibits a higher diffusivity of water at $\varepsilon < -1$ kcal mol^{-1}.

The decomposition of $D^{\mathrm{e}}(\varepsilon)/D$ into the contributions of the moieties of Prodan is performed based on Eqs. (3.7) and (3.8). As well as in the case of the solvent distributions, we decompose Prodan into the three moieties, DMA, NAP, and PRO (Fig. 3.1). The decomposed profiles of $D^{\mathrm{e},m}(\varepsilon)/D$ are shown in Figs. (3.3)(b)-(d). The contributions from the DMA (Fig. 3.3(b)) and PRO (Fig. 3.3(d)) are almost the same for all the solvents. Furthermore, the contribution from PRO is found to be negligibly small compared with the other contributions. As for NAP, water shows a higher diffusivity at $\varepsilon < -1$ kcal mol^{-1} than the alcohol solvents. It indicates that the difference of $D^{\mathrm{e}}(\varepsilon)/D$ between the water and alcohol solvents originates from NAP. This behavior corresponds to the fact that the stabilized water molecules are highly populated around NAP (Fig. 3.2 and Fig. S3, ESI).

3.3.3 Solvation time correlation functions (STCFs)

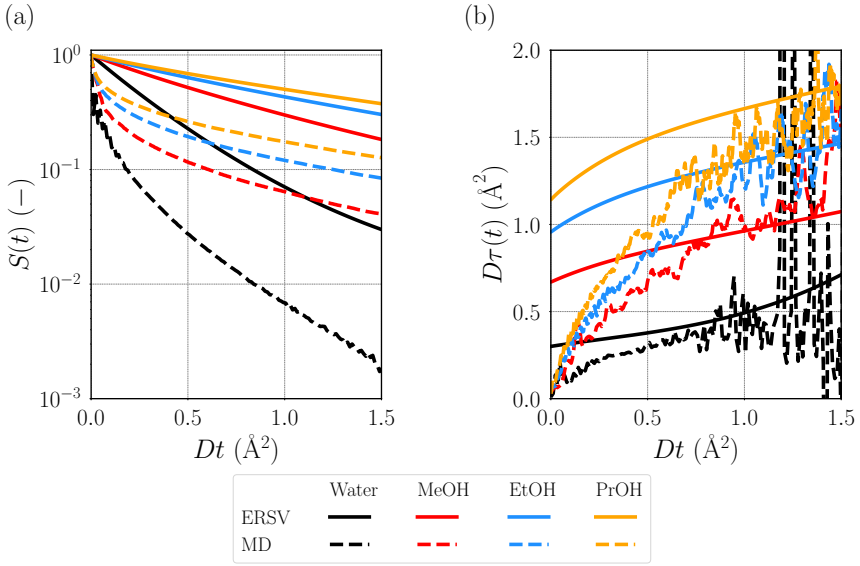


Fig. 3.4: (a) Solvation time correlation functions (STCFs), $S(t)$, and (b) relaxation time coefficients, $\tau(t)$, defined as the time derivative of $\ln S(t)$. These functions obtained from the ERSV equation and MD simulations are shown. The timescale is scaled with the diffusion coefficient for each solvent (D).

The ERSV equation enables us to compute the solvation time correlation functions (STCFs), $S(t)$, describing the nonequilibrium solvent relaxation process triggered by the photoexcitation of Prodan with the aid of the linear response theory (Eq. (3.16)). $S(t)$ can be also computed using the MD simulations at the ground state using the linear response theory (Eq. (3.12)).The comparison of $S(t)$ obtained from the two approaches is useful to understand how the approximations introduced in the ERSV equation affect the description of the dynamics. The time derivative of $S(t)$ gives the relaxation time coefficient, $\tau(t)$.

$$\tau(t) = -\left(\frac{\mathrm{d}\ln S(t)}{\mathrm{d}t}\right)^{-1}. \tag{3.22}$$

The time developments of $S(t)$ and $D\tau(t)$ are plotted in Fig. 3.4(a) and (b), respectively. In these plots, we use the timescale scaled with the diffusion coefficients. For all

the solvents, $S(t)$ calculated with the MD simulation shows the fast decay of $S(t)$ on the short timescale. In addition, $S(t)$ in water shows the damped oscillation at $Dt < 0.2$ Å^2. In our previous study, [164] a similar oscillation is observed for the solvation dynamics of benzonitrile in water due to the rotational motion of the water molecules hydrogen bonded with benzonitrile. The rotational motion of the water molecules would also bring the oscillation observed in the present water system. As for $S(t)$ obtained from the ERSV equations, the fast decay on the short timescale is not observed for all the solvents. Since the ERSV equation is derived with the overdamped limit that causes the neglects of the memory and inertial effects of solvent motions, the discrepancy between the ERSV equation and MD simulations clearly reveals the importance of these effects on the short timescale. In the case of the long timescale ($Dt > 1$ Å^2), the slope of $S(t)$ on the logarithmic scale obtained from the ERSV equation are similar to those from the MD simulations, as also shown in the plots of $D\tau(t)$ (Fig. 3.4). Petrone *et al.* revealed that the collective solvent rearrangement dominates the solvation dynamics on a picosecond timescale in the case of N-methyl-6-oxyquinolinium betaine in water. [160] Since $Dt = 1$ Å^2 corresponds to $t \sim 1.7$ ps in water, the good agreement between the ERSV equation and MD simulations on this timescale for the Prodan system suggests that such a solvent motion can be described using the ERSV equation through the collective term expressed with the direct correlation function, $c(\varepsilon, \eta)$. Note that the ERSV equation can compute the time developments of $\tau(t)$ without the statistical noise observed in those from the MD simulations. Thus, the rigorous estimation of the time coefficient on the diffusion timescale is possible using the ERSV equation.

3.3.4 Importance of collective motion of solvents

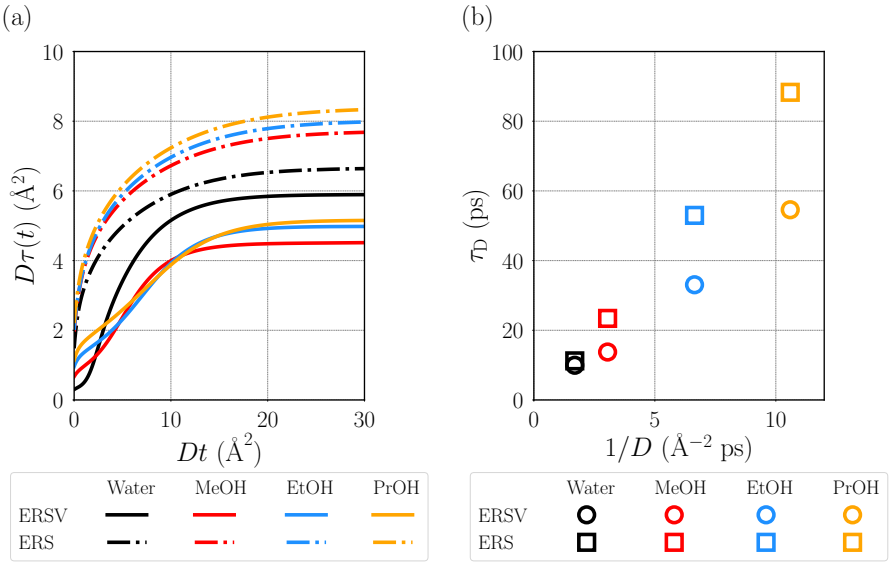


Fig. 3.5: Comparison of $\tau(t)$ obtained from the ERSV and ERS equations. (a) $D\tau(t)$, and (b) correlation plots of the relaxation time coefficient on the diffusion timescale, τ_D, against $1/D$. The values of $\tau(t)$ at $Dt = 30$ Å^2 are defined as τ_D.

In this subsection, we focus on analyzing the collective motion of solvents based on the energy representation. The neglect of the term for the collective motion of solvents in the ERSV equation (Eq. (3.17)) leads to the energy-represented Smoluchowski (ERS)

equation for $Q(\varepsilon, t)$ (Eq. (3.18)). Therefore, the comparison of $\tau(t)$ (Eq. (3.22)) obtained from the ERSV and ERS equations enables us to elucidate the importance of the collective motion on the solvation dynamics.

Fig. 3.5(a) shows the time development of $D\tau(t)$ obtained from the ERS equation, together with those from the ERSV equation. It is seen that the ERS equation overestimates the values of $\tau(t)$ on the short timescale, indicating that the collective motion of solvents promotes the relaxation of the solvation structure. For both the equations, $\tau(t)$ for all the solvents converge at $Dt > 25$ Å^2. The convergence for water is faster on the scaled timescale (Dt) than for the other solvents in the case of the ERS equation. It reflects the high diffusivity of water observed in $D^{\mathrm{e}}(\varepsilon)/D$ at $\varepsilon < -1$ kcal mol^{-1} (Fig. 3.3(a)). A small difference in the converged values of $\tau(t)$ for water is discernible between the ERSV and ERS equations. As for the alcohol solvents, on the other hand, the converged values for the ERS equation are larger than those for the ERSV equation. Hence, treating the collective motion is necessary for the quantitative estimation of $\tau(t)$ on the long timescale for the alcohol solvents.

We examine how the relaxation time coefficient on the diffusion timescale, τ_{D}, depends on the solvent species for both the ERSV and ERS equations. The values of $\tau(t)$ at $Dt = 30$ Å^2 are used as τ_{D}. Fig. 3.5(b) shows the correlation plots of τ_{D} against $1/D$. For both the ERSV and ERS equations, the value of τ_{D} is larger in the order of PrOH > EtOH > MeOH > water. It indicates that the contribution of the collective motion does not alter the ordering, while the absolute values of τ_{D} are affected by that motion for the alcohol solvents. Note that the ordering is apparently changed when $\tau(t)$ obtained from the ERSV equation is multiplied by D (Fig. 3.5(a)). In the case of the ERS equation, the correlation plot falls into a single line. Since the ERS equation expresses time development by the terms involving D as a product (see Eqs. (3.4) and (3.18)), the linear relationship between $\tau(t)$ and $1/D$ holds when $\langle\rho(\varepsilon)\rangle$ and $D^{\mathrm{e}}(\varepsilon)/D$ are similar among the different solvents. Furthermore, on the diffusion timescale, the dynamic behaviors are dominated by $\langle\rho(\varepsilon)\rangle$ and $D^{\mathrm{e}}(\varepsilon)$ around $\varepsilon \sim 0$, because the energy distribution functions decay slowly in this region (almost corresponding to the bulk). Thus, it can be concluded that the similarities of $\langle\rho(\varepsilon)\rangle$ and $D^{\mathrm{e}}(\varepsilon)/D$ around $\varepsilon \sim 0$ among different solvents (Figs. 3.2(a) and 3.3(a)) give the linear relationship between τ_{D} and $1/D$. As for the ERSV equation, the alcohol solvents show the same linear relationship, suggesting that the collective motions (Eq. (3.17)) in different alcohol solvents are similar except for the contribution from the translational diffusion coefficient. On the other hand, τ_{D} for water deviates from the linear relationship observed in the case of the alcohol solvents. Since only the water system shows the small difference of τ_{D} between the ERSV and ERS equations, this deviation reflects the difference in the importance of the collective motion on the diffusion timescale between the water and alcohol solvents.

3.4 Conclusion

We investigated the solvation dynamics of Prodan triggered by the photoexcitation ($S_1(\pi - \pi^*)$) using the energy-represented Smoluchowski-Vlasov (ERSV) equation. The difference in the dynamics for different solvents (water, methanol (MeOH), ethanol (EtOH), and 1-propanol (PrOH)) was elucidated. The ERSV equation enables us to calculate the time development of the systems from the several quantities on the energy coordinate, solvent static distribution (energy distribution), $\langle\rho(\varepsilon)\rangle$, direct correlation function, $c(\varepsilon, \eta)$, and diffusion coefficient, $D^{\mathrm{e}}(\varepsilon)/D$, computed using the molecular dynamics (MD) sim-

ulations at the ground state. The defining potential was set to the difference in the solute-solvent pair interaction energy between the excited and ground states.

We found that the all the solvents had the similar energy distributions, although the populations of both the stabilized and destabilized molecules are higher for water. The detailed analysis of the distributions was realized with the decomposition of the defining potential into the contributions from the moieties of Prodan, *N*,*N*-dimethyl (DMA), naphthalene (NAP), and propionyl (PRO) moieties. The profiles of $D^{\mathrm{e}}(\varepsilon)/D$, where D is the translational diffusion coefficient of a solvent, showed no significant difference among the alcohol solvents, indicating that the difference in the diffusivity on the energy coordinate is brought by D. $D^{\mathrm{e}}(\varepsilon)/D$ for water revealed the higher diffusivity than those for the alcohol solvents at $\varepsilon < -1$ kcal mol^{-1} that stems from the water molecules coordinating to NAP. Using the ERSV equation, we computed the time development of the relaxation time coefficient, $\tau(t)$, defined as the time derivative of the logarithm of solvation time correlation function (STCF). On the short timescale, $\tau(t)$ calculated with the ERSV equation largely deviates from those with the MD simulations for all the solvents. This deviation clearly reveals the importance of the memory and inertial effects ignored in the ERSV equation on that timescale. $\tau(t)$ on the long timescale from the MD simulations were well reproduced with the ERSV equation. We also computed $\tau(t)$ using the energy-represented Smoluchowski (ERS) equation derived by neglecting the term for the collective motion of solvents in the ERSV equation. As for water, the time coefficient on the diffusion timescale, τ_{D}, obtained from the ERS equation was similar to that from the ERSV equation. On the other hand, τ_{D} for the alcohol solvents from the ERS equation were larger than those from the ERSV equation, indicating that the collective motion tends to promote the solvent relaxation for these solvents. We found that the set of τ_{D} from the ERS equation linearly correlates with $1/D$ because of the similarities of $\langle\rho(\varepsilon)\rangle$ and $D^{\mathrm{e}}(\varepsilon)/D$ among the different solvents. In the case of the ERSV equation, τ_{D} for the alcohol solvents also fell into a single line, but τ_{D} for water deviated from that line. It reveals that the importance of the collective motion on the diffusion timescale is different between the water and alcohol solvents.

An advantage of employing the energy representation is that molecular motion can be effectively described in one-dimensional space without explicitly treating the orientational degree of freedom. On the other hand, the dielectric models, such as the Debye model, have focused upon the reorientational relaxations within the framework of the continuum treatment of the solvent. [93,142,189–191] Then, to compare the dielectric and the ERSV methods, the collective reorientation modes need to be extracted in the energy representation. It will be an interesting subject to express a variety of solvent motions effectively over the energy coordinate.

To realize the more realistic description of the solvation dynamics on the diffusion timescale, sophisticating the theoretical treatment of the electronic structure of a fluorescent probe (solute) is necessary. In this study, we approximated that the electronic structure of the solute was unchanged during the solvation dynamics. However, it is well known that solute molecules are polarized depending on the surrounding environments, affecting the solvation dynamics. In the case of the integral equation theory with the interaction site representation, the formulations for treating the polarization effects have been developed so far. Naka *et al.* proposed the methodologies of incorporating the polarization effects [17, 18] into the reference interaction site model self-consistent field (RISM-SCF) method [13, 14] using the charge-response kernel (CRK) model. [21] The recently developed theory by Yamaguchi and Yoshida can describe the solvent polarization effects on

the solvation dynamics [59] based on the solvent-polarizable 3D-RISM theory [19] and time-dependent density functional theory (TD-DFT). [58] Since the energy coordinate is suitable for treating the flexibility of solvents and heterogenous environments compared with the spatial coordinate employed in the above theories, the energy-represented dynamics theory incorporating the polarization effects based on the CRK model could be promising for a realistic description of the solvation dynamics in complex systems such as polymer solutions and lipid membrane systems. The importance of solute motion should also be noted. Recent experimental and simulation studies revealed that the vibrational solute motion has a non-negligible influence on the dynamics of the solvents inside the first solvation shell for small probes in water. [192, 193] Thus, including the solute motion in the framework of the energy-represented dynamics theory is also crucial for a more realistic description of the solvation dynamics. We believe that the ERSV equation and its extension will deepen our understanding of the nonequilibrium processes at the excited states.

Appendix B

Appendix of chapter 2

B.1 Supplementary figures

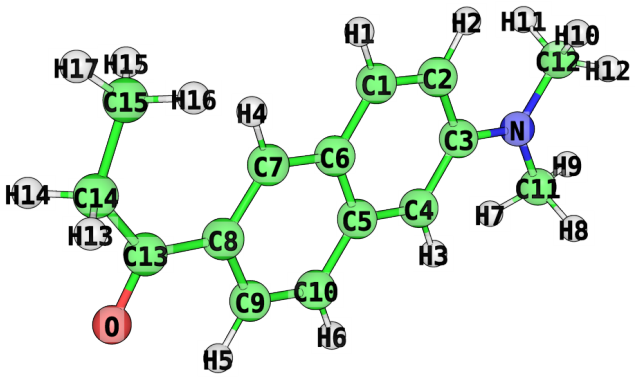


Fig. B.1: Definitions of the atom labels for Prodan.

Table B.1: Atomic charges on Prodan at the ground and excited states. The label of each atom is shown in Fig. B.1.

site name	atomic charge at GS (e)	atomic charge at ES (e)	difference
C1	-0.16703	-0.20922	-0.04219
C2	-0.21334	-0.30112	-0.08778
C3	0.30188	0.30558	0.00370
C4	-0.40712	-0.30964	0.09748
C5	0.20250	0.16139	-0.04111
C6	0.11120	0.20825	0.09705
C7	-0.21616	-0.33364	-0.11748
C8	-0.07123	0.02150	0.09273
C9	-0.01159	-0.05330	-0.04171
C10	-0.23902	-0.25537	-0.01635
C11	0.09300	0.06970	-0.02330
C12	0.09210	0.08956	-0.00254
N	-0.26505	-0.19167	0.07338
C13	0.38845	0.31197	-0.07648
O	-0.44136	-0.45568	-0.01432
C14	0.15965	0.24546	0.08581
C15	-0.14782	-0.19386	-0.04604
H1	0.11247	0.11805	0.00558
H2	0.12682	0.13465	0.00783
H3	0.16547	0.13830	-0.02717
H4	0.11577	0.10156	-0.01421
H5	0.08113	0.08488	0.00375
H6	0.11117	0.10350	-0.00767
H7, H8, H9	0.01343	0.03329	0.01986
H10, H11, H12	0.01296	0.02878	0.01582
H13, H14	-0.02970	-0.04745	-0.01775
H15, H16, H17	0.03278	0.03928	0.00650

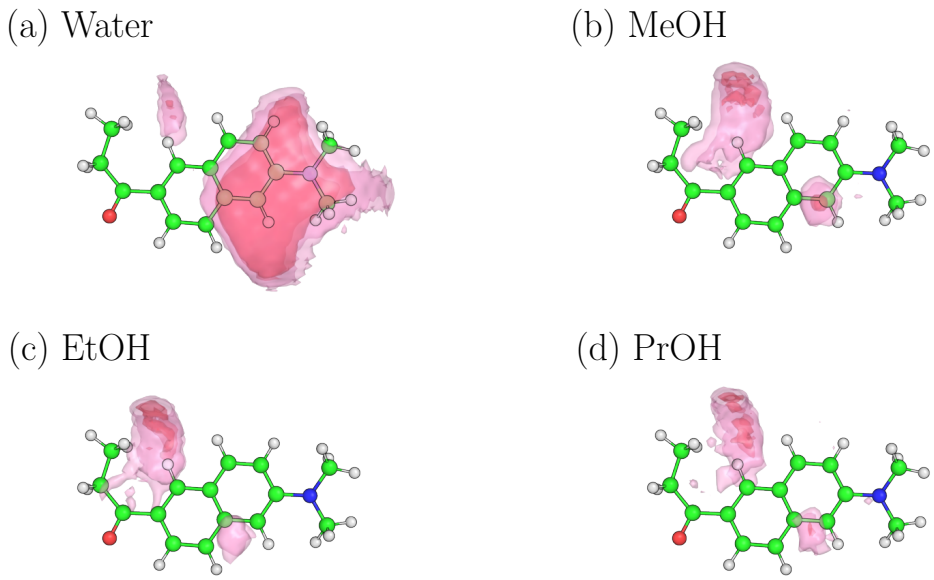


Fig. B.2: Spatial distribution functions (SDFs) of the hydroxyl hydrogens corresponding to the destabilized region ($\varepsilon \geq 1.0$ kcal mol^{-1}) for (a) water, (b) MeOH, (c) EtOH, and (d) PrOH. The isovalues of SDFs for solid and transparent surfaces are 0.1 and 0.05, respectively.

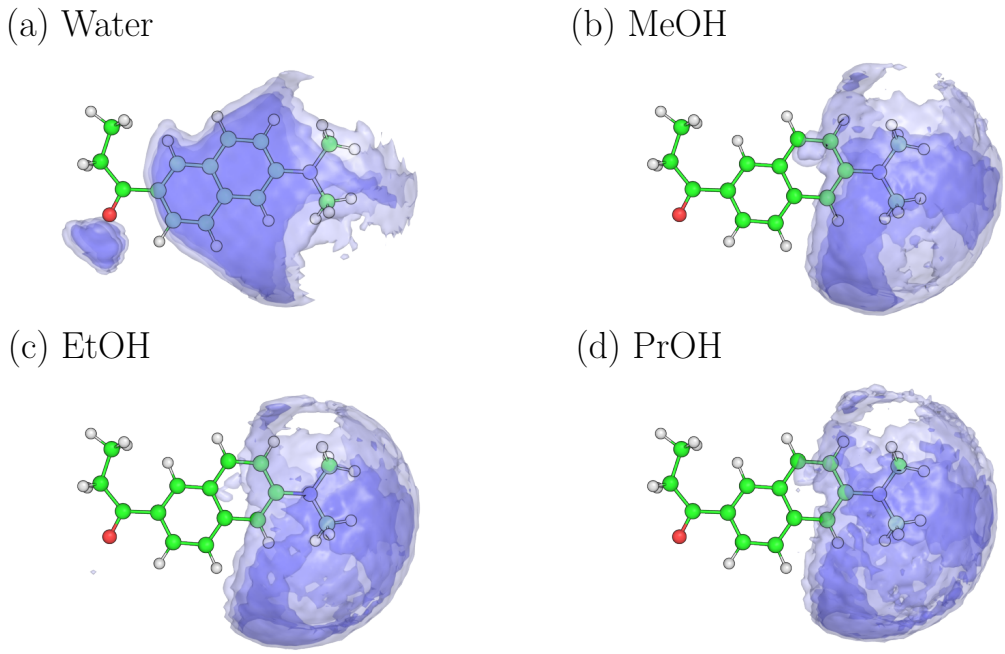


Fig. B.3: Spatial distribution functions (SDFs) of the hydroxyl hydrogens corresponding to the stabilized region ($\varepsilon \leq -1.0$ kcal mol^{-1}) for (a) water, (b) MeOH, (c) EtOH, and (d) PrOH, respectively. The isovalues of SDFs for solid and transparent surfaces are 0.1 and 0.05, respectively.

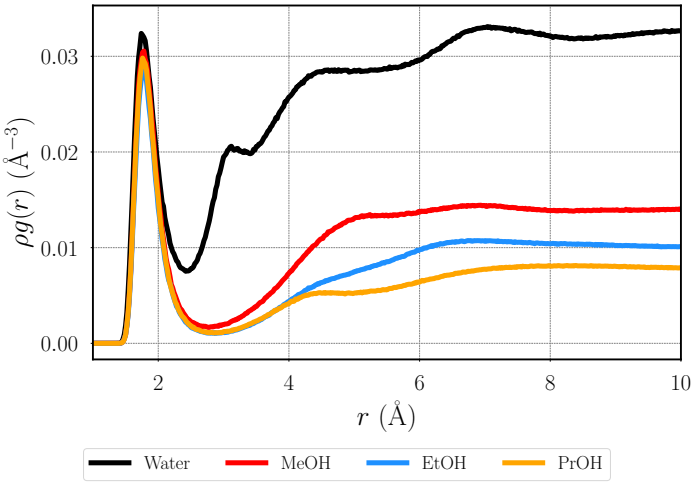


Fig. B.4: Radial densities of the hydroxyl hydrogens around the oxygen atom of Prodan, $\rho g\,(r)$.

Chapter 4

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

4.1 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. [194] The binding process is also crucial for drug molecules, which regulate (promote or inhibit) cellular functions such as cell proliferation mediated by signal transduction. [195, 196] Molecular dynamics (MD) simulation has played an important role in drug discovery and design [197, 198] 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. [199] 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. [200–203] 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), [204] free-energy perturbation (FEP), [205] and Bennett acceptance ratio (BAR) [206] 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 alchemical pathway. The double-annihilation scheme (DAS) [207] and double-decoupling scheme (DDS) [208] 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 state and bound state. An effective setting of the intermediate states on the pathways is proposed for the DAS. [209, 210] 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. [211, 212] 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. [213, 214] The automation of the free-energy calculation based on the methodologies mentioned above has expanded the versatility. [215–220] 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 free-energy estimation of binding can be made efficient by adopting approximate methods. [197, 221, 222] 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 unbound states. [223] The effect of entropy was incorporated by mining local energy minima and quantifying the extents of local fluctuations. [224] 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. [65] 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. [60, 64] 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, [64] polymer solutions, [70, 73] and crystal-surface systems. [83] 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. [89] 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 has been reported, [225–227] 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 intermediate state. This intermediate state was proposed in a previous study on the dissolution of water into polymer membranes. [70]

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, [228] 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. [227], 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.

4.2 Theory

4.2.1 Theoretical expression of binding free energy

In this subsection, 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

$$\mathrm{H} + \mathrm{G} \rightleftharpoons \mathrm{B}. \tag{4.1}$$

Here, H, G, and B signify host, guest and bound complex, respectively. The equilibrium constant of the above reaction, K_a, is related with the binding free energy, $\Delta G^\circ_\mathrm{B}$, as

$$\begin{aligned} \Delta G^\circ &= -\frac{1}{\beta} \log c^\circ K_\mathrm{a} \\ &= -\frac{1}{\beta} \log c^\circ \left(\frac{[\mathrm{B}]}{[\mathrm{H}]\,[\mathrm{G}]} \right), \end{aligned} \tag{4.2}$$

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. (4.1) is expressed as

$$\mu_\mathrm{B} - (\mu_\mathrm{H} + \mu_\mathrm{G}) = 0, \tag{4.3}$$

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 or G) is given by [89]

$$\begin{aligned} \mu_\mathrm{S} = &\frac{1}{\beta} \log \left([\mathrm{S}]\, \lambda_\mathrm{S} \right) \\ &- \frac{1}{\beta} \log \frac{\displaystyle\int d\mathbf{x}_\mathrm{S} \int d\mathbf{X}_\mathrm{V}\, e^{-\beta(U_\mathrm{S} + U_\mathrm{SV} + U_\mathrm{V})}}{\displaystyle V \int d\mathbf{X}_\mathrm{V}\, e^{-\beta U_\mathrm{V}}}, \end{aligned} \tag{4.4}$$

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. $\mathbf{x}_\mathrm{S}$ is the full-coordinate of species S, and $\mathbf{X}_\mathrm{V}$ 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. (4.4), but the configurational integral over the full-coordinates of H and G, $\mathbf{x}_\mathrm{HG} = \{\mathbf{x}_\mathrm{H}, \mathbf{x}_\mathrm{G}\}$, needs to be restricted to the region corresponding to the bound state. Let $\Theta_\mathrm{B}\left(\mathbf{x}_\mathrm{HB}\right)$ be the characteristic function whose value is unity when the bound complex is formed and zero otherwise. Then, μ_B is expressed as

$$\begin{aligned} \mu_\mathrm{B} = &\frac{1}{\beta} \log \left([\mathrm{B}]\, \lambda_\mathrm{H} \lambda_\mathrm{G} \right) \\ &- \frac{1}{\beta} \log \frac{\displaystyle\int d\mathbf{x}_\mathrm{HG} \int d\mathbf{X}_\mathrm{V}\, \Theta_\mathrm{B}\left(\mathbf{x}_\mathrm{HG}\right) e^{-\beta(U_\mathrm{B} + U_\mathrm{BV} + U_\mathrm{V})}}{\displaystyle V \int d\mathbf{X}_\mathrm{V}\, e^{-\beta U_\mathrm{V}}}, \end{aligned} \tag{4.5}$$

where $d\mathbf{x}_{\mathrm{HG}} = d\mathbf{x}_{\mathrm{H}} d\mathbf{x}_{\mathrm{G}}$. U_{B} is the potential of H and G that is composed of the intramolecular energies of H, U_{H}, and of G, U_{G}, and the interaction energy between H and G, U_{HG}, as

$$U_{\mathrm{B}} = U_{\mathrm{H}} + U_{\mathrm{G}} + U_{\mathrm{HG}}. \tag{4.6}$$

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}.

$$U_{\mathrm{BV}} = U_{\mathrm{HV}} + U_{\mathrm{GV}}. \tag{4.7}$$

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 $\Delta G^{\circ}_{\mathrm{B}}$ from Eq. (4.2), as will be discussed later. As for the dissociate state, let us introduce the solution and reference systems whose total potentials are respectively defined as

$$\mathcal{V}^{\mathrm{D}}_{\mathrm{sol}} = U_{\mathrm{G}} + U_{\mathrm{GV}} + U_{\mathrm{V}}, \tag{4.8}$$
$$\mathcal{V}^{\mathrm{D}}_{\mathrm{ref}} = U_{\mathrm{G}} + U_{\mathrm{V}}. \tag{4.9}$$

The solvation free energy, $\Delta\mu^{\mathrm{D}}_{\mathrm{G}}$, can be described as

$$\Delta\mu^{\mathrm{D}}_{\mathrm{G}} = -\frac{1}{\beta}\log\frac{\displaystyle\int d\mathbf{x}_{\mathrm{G}}\int d\mathbf{X}_{\mathrm{V}}\, e^{-\beta\mathcal{V}^{\mathrm{D}}_{\mathrm{sol}}}}{\displaystyle\int d\mathbf{x}_{\mathrm{G}}\int d\mathbf{X}_{\mathrm{V}}\, e^{-\beta\mathcal{V}^{\mathrm{D}}_{\mathrm{ref}}}}. \tag{4.10}$$

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

$$\Delta\mu^{\mathrm{B}}_{\mathrm{G}} = -\frac{1}{\beta}\log\frac{\displaystyle\int d\mathbf{x}_{\mathrm{HG}}\int d\mathbf{X}_{\mathrm{V}}\,\Theta_{\mathrm{B}}\left(\mathbf{x}_{\mathrm{HG}}\right) e^{-\beta\mathcal{V}^{\mathrm{B}}_{\mathrm{sol}}}}{\displaystyle\int d\mathbf{x}_{\mathrm{HG}}\int d\mathbf{X}_{\mathrm{V}}\,\Theta_{\mathrm{B}}\left(\mathbf{x}_{\mathrm{HG}}\right) e^{-\beta\mathcal{V}^{\mathrm{B}}_{\mathrm{ref}}}}, \tag{4.11}$$

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

$$\mathcal{V}^{\mathrm{B}}_{\mathrm{sol}} = U_{\mathrm{G}} + U_{\mathrm{H}} + U_{\mathrm{HG}} + U_{\mathrm{GV}} + U_{\mathrm{HV}} + U_{\mathrm{V}}, \tag{4.12}$$
$$\mathcal{V}^{\mathrm{B}}_{\mathrm{ref}} = U_{\mathrm{G}} + U_{\mathrm{H}} + U_{\mathrm{HV}} + U_{\mathrm{V}}. \tag{4.13}$$

In Eq. (4.12), all the interactions among G, H, and V are operative, and in Eq. (4.13), the interactions between G and H and between G and V are turned off. $\Delta\mu^{\mathrm{B}}_{G}$ 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. (4.11) means that the solvation process of species G, which forms the bound complex in both the solution and reference systems, is represented by $\Delta\mu^{\mathrm{B}}_{\mathrm{G}}$.

Substituting Eqs. (4.4) and (4.5) into Eq. (4.2) yields

$$\Delta G^{\circ} = \Delta\mu^{\mathrm{B}}_{\mathrm{G}} - \Delta\mu^{\mathrm{D}}_{\mathrm{G}} + \Delta G^{\circ}_{\mathrm{corr}}, \tag{4.14}$$

where we have used Eqs. (4.10) and (4.11), and $\Delta G^\circ_{\rm corr}$ is the standard-state correction term, ensuring the concentration of G in the dissociate state is c°, expressed as

$$\begin{aligned}
&\Delta G^\circ_{\rm corr} \\
&= -\frac{1}{\beta}\log\left(c^\circ V \frac{\displaystyle\int d\mathbf{x}_{\rm HG}\int d\mathbf{X}_{\rm V}\,\Theta_{\rm B}\left(\mathbf{x}_{\rm HG}\right)e^{-\beta\mathcal{V}^{\rm B}_{\rm ref}}}{\displaystyle\int d\mathbf{x}_{\rm HG}\int d\mathbf{X}_{\rm V}\,e^{-\beta\mathcal{V}^{\rm B}_{\rm ref}}}\right).
\end{aligned} \tag{4.15}$$

The configurations of G and those of H and the solvent molecules are independently generated by $U_{\rm G}$ and $U_{\rm H}+U_{\rm HV}+U_{\rm V}$, respectively, in the reference system, and thus the logarithm in Eq. (4.15) can be computed by the test-particle insertion of G into the configurations of H and the solvent molecules. Furthermore, $\Delta G^\circ_{\rm corr}$ is intensive and the spatial region for insertion can be made smaller than the simulation cell. [89]

To utilize Eq. (4.14), the definition of $\Theta_{\rm B}$ is needed for $\Delta\mu^{\rm B}_{\rm G}$ (Eq. (4.14)) and $\Delta G^\circ_{\rm corr}$ (Eq. (4.15)). 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, $\Theta_{\rm B}$ can be set to accept all the sampled configurations. In this case, on the other hand, the explicit form of $\Theta_{\rm B}$ is needed in the reference system to distinguish between the bound complex and others. The unique determination of $\Theta_{\rm B}$ is generally impossible except for simple host and guest molecules, such as monoatomic molecules, and ΔG° appears to be dependent on the choice of $\Theta_{\rm B}$ through the sampling in the reference system. Actually, it can be proved that ΔG° is not affected by the choice of $\Theta_{\rm B}$ for the reference solvent as described below. Let us introduce the characteristic function that is different from $\Theta_{\rm B}$, $\Theta'_{\rm B}$, and the following quantities.

$$\begin{aligned}
\Delta\mu^{\rm B'}_{\rm G} &= \Delta\mu^{\rm B}_{\rm G} - \frac{1}{\beta}\log\frac{\displaystyle\int d\mathbf{x}_{\rm HG}\int d\mathbf{X}_{\rm V}\,\Theta_{\rm B}\left(\mathbf{x}_{\rm HG}\right)e^{-\beta\mathcal{V}^{\rm B}_{\rm ref}}}{\displaystyle\int d\mathbf{x}_{\rm HG}\int d\mathbf{X}_{\rm V}\,\Theta'_{\rm B}\left(\mathbf{x}_{\rm HG}\right)e^{-\beta\mathcal{V}^{\rm B}_{\rm ref}}} \\
&= -\frac{1}{\beta}\log\frac{\displaystyle\int d\mathbf{x}_{\rm HG}\int d\mathbf{X}_{\rm V}\,\Theta_{\rm B}\left(\mathbf{x}_{\rm HG}\right)e^{-\beta\mathcal{V}^{\rm B}_{\rm sol}}}{\displaystyle\int d\mathbf{x}_{\rm HG}\int d\mathbf{X}_{\rm V}\,\Theta'_{\rm B}\left(\mathbf{x}_{\rm HG}\right)e^{-\beta\mathcal{V}^{\rm B}_{\rm ref}}},
\end{aligned} \tag{4.16}$$

$$\begin{aligned}
\Delta G^{\circ\prime}_{\rm corr} &= \Delta G^\circ_{\rm corr} \\
&\quad + \frac{1}{\beta}\log\frac{\displaystyle\int d\mathbf{x}_{\rm HG}\int d\mathbf{X}_{\rm V}\,\Theta_{\rm B}\left(\mathbf{x}_{\rm HG}\right)e^{-\beta\mathcal{V}^{\rm B}_{\rm ref}}}{\displaystyle\int d\mathbf{x}_{\rm HG}\int d\mathbf{X}_{\rm V}\,\Theta'_{\rm B}\left(\mathbf{x}_{\rm HG}\right)e^{-\beta\mathcal{V}^{\rm B}_{\rm ref}}} \\
&= -\frac{1}{\beta}\log\left(c^\circ V \frac{\displaystyle\int d\mathbf{x}_{\rm HG}\int d\mathbf{X}_{\rm V}\,\Theta'_{\rm B}\left(\mathbf{x}_{\rm HG}\right)e^{-\beta\mathcal{V}^{\rm B}_{\rm ref}}}{\displaystyle\int d\mathbf{x}_{\rm HG}\int d\mathbf{X}_{\rm V}e^{-\beta\mathcal{V}^{\rm B}_{\rm ref}}}\right),
\end{aligned} \tag{4.17}$$

By substituting Eqs. (4.16) and (4.17) into Eq. (4.14), one can rewrite Eq. (4.14) without any approximations as

$$\Delta G^\circ_{\rm B} = \Delta\mu^{\rm B'}_{\rm G} - \Delta\mu^{\rm D}_{\rm G} + \Delta G^{\circ\prime}_{\rm corr}, \tag{4.18}$$

indicating that $\Delta G^{\circ}_{\mathrm{B}}$ does not depend on the choice of Θ'_{B} when all the terms in the above equation are computed in an exact manner.

4.2.2 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 $\Delta\mu^{\mathrm{B}}_{\mathrm{G}}$ (Eq. (4.11)), as the theoretical developments of the ER theory for $\Delta\mu^{\mathrm{D}}_{\mathrm{G}}$ (Eq. (4.10)) have been already reported elsewhere. [61, 64, 116]

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

$$\hat{\rho}_{\alpha}(\varepsilon) = \sum_{i\in\alpha} \delta\left(u_{\alpha}\left(\mathbf{x}_{\mathrm{G}}, \mathbf{x}_{\alpha,i}\right) - \varepsilon\right). \tag{4.19}$$

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

$$\left\langle \cdots \right\rangle_{\mathrm{sol},\Theta_{\mathrm{B}}} = \frac{\displaystyle\int d\mathbf{x}_{\mathrm{HG}} \int d\mathbf{X}_{\mathrm{V}}\ (\cdots)\,\Theta_{\mathrm{B}}(\mathbf{x}_{\mathrm{HG}})\, e^{-\beta\mathcal{V}^{\mathrm{B}}_{\mathrm{sol}}}}{\displaystyle\int d\mathbf{x}_{\mathrm{HG}} \int d\mathbf{X}_{\mathrm{V}}\,\Theta_{\mathrm{B}}(\mathbf{x}_{\mathrm{HG}})\, e^{-\beta\mathcal{V}^{\mathrm{B}}_{\mathrm{sol}}}}, \tag{4.20}$$

$$\left\langle \cdots \right\rangle_{\mathrm{ref},\Theta_{\mathrm{B}}} = \frac{\displaystyle\int d\mathbf{x}_{\mathrm{HG}} \int d\mathbf{X}_{\mathrm{V}}\ (\cdots)\,\Theta_{\mathrm{B}}(\mathbf{x}_{\mathrm{HG}})\, e^{-\beta\mathcal{V}^{\mathrm{B}}_{\mathrm{ref}}}}{\displaystyle\int d\mathbf{x}_{\mathrm{HG}} \int d\mathbf{X}_{\mathrm{V}}\,\Theta_{\mathrm{B}}(\mathbf{x}_{\mathrm{HG}})\, e^{-\beta\mathcal{V}^{\mathrm{B}}_{\mathrm{ref}}}}, \tag{4.21}$$

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

$$\rho^{\mathrm{B}}_{\mathrm{sol},}(\varepsilon) = \left\langle \hat{\rho}_{\alpha}(\varepsilon) \right\rangle_{\mathrm{sol},\Theta_{\mathrm{B}}}, \tag{4.22}$$

$$\rho^{\mathrm{B}}_{\mathrm{ref},\alpha}(\varepsilon) = \left\langle \hat{\rho}_{\alpha}(\varepsilon) \right\rangle_{\mathrm{ref},\Theta_{\mathrm{B}}}, \tag{4.23}$$

respectively. According to the Kirkwood's charging formula for the alchemical pathway connecting the solution and reference systems through the coupling parameter, $\Delta\mu^{\mathrm{B}}_{\mathrm{G}}$ 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 [64]

$$\begin{aligned}
\Delta\mu^{\mathrm{B}}_{\mathrm{G}} = &\sum_{\alpha}\int d\varepsilon\,\varepsilon\rho^{\mathrm{B}}_{\mathrm{sol},\alpha}(\varepsilon) \\
&- \frac{1}{\beta}\sum_{\alpha}\int d\varepsilon\,\left(\rho^{\mathrm{B}}_{\mathrm{sol},\alpha}(\varepsilon) - \rho^{\mathrm{B}}_{\mathrm{ref},\alpha}(\varepsilon)\right) \\
&+ \frac{1}{\beta}\sum_{\alpha}\int d\varepsilon\,\rho^{\mathrm{B}}_{\mathrm{sol},\alpha}(\varepsilon)\log\frac{\rho^{\mathrm{B}}_{\mathrm{sol},\alpha}(\varepsilon)}{\rho^{\mathrm{B}}_{\mathrm{ref},\alpha}(\varepsilon)} \\
&+ \mathcal{F}\left[\rho^{\mathrm{B}}_{\mathrm{sol},\alpha}(\varepsilon), \rho^{\mathrm{B}}_{\mathrm{ref},\alpha}(\varepsilon), \chi^{\mathrm{B}}_{\alpha\beta}(\varepsilon,\eta)\right],
\end{aligned} \tag{4.24}$$

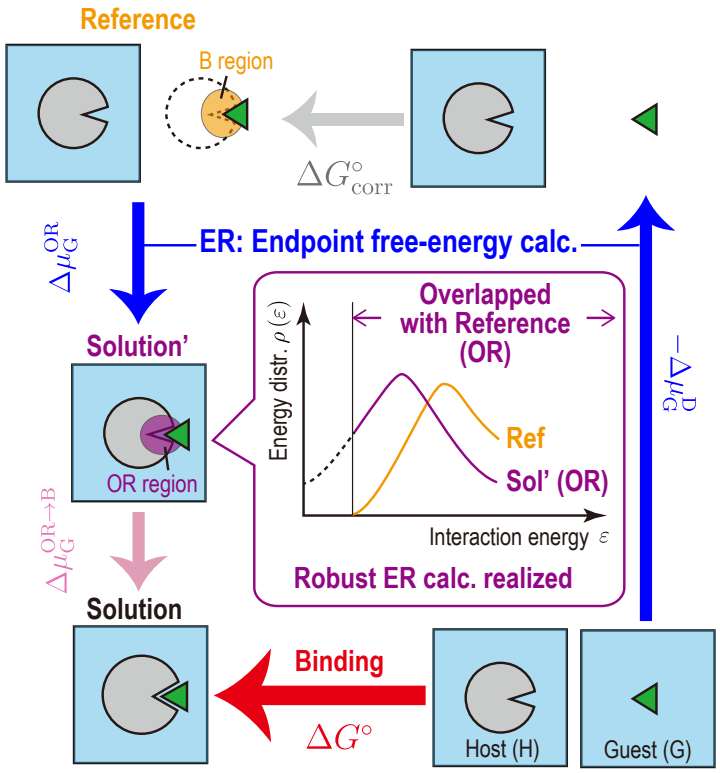


Fig. 4.1: Thermodynamic cycle employed in the energy representation (ER) method incorporating a solution state with overlapped distributions with reference (ER-OR).

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

$$\chi^{\mathrm{B}}_{\alpha\beta}(\varepsilon,\eta) = \langle\hat{\rho}_{\alpha}(\varepsilon)\,\hat{\rho}_{\beta}(\eta)\rangle_{\mathrm{ref},\Theta_{\mathrm{B}}} - \langle\hat{\rho}_{\alpha}(\varepsilon)\rangle_{\mathrm{ref},\Theta_{\mathrm{B}}}\langle\hat{\rho}_{\beta}(\eta)\rangle_{\mathrm{ref},\Theta_{\mathrm{B}}}. \tag{4.25}$$

The first three terms in Eq. (4.24) 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. 64.

Evaluating the free energy using Eq. (4.24) is effective when the distributions in the solution ($\rho^{\mathrm{B}}_{\mathrm{sol},}(\varepsilon)$) and reference ($\rho^{\mathrm{B}}_{\mathrm{ref},}(\varepsilon)$) 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 $\rho^{\mathrm{B}}_{\mathrm{sol},\alpha}(\varepsilon)$ and $\rho^{\mathrm{B}}_{\mathrm{ref},\alpha}(\varepsilon)$ may not overlap well. In such host-guest systems, the ε-region with $\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon) \neq 0$ and $\rho_{\mathrm{ref},\alpha}(\varepsilon) = 0$, which is problematic due to the integrand of the third term in Eq. (4.24)

$$\rho^{\mathrm{B}}_{\mathrm{sol},\alpha}(\varepsilon)\log\frac{\rho^{\mathrm{B}}_{\mathrm{sol},\alpha}(\varepsilon)}{\rho^{\mathrm{B}}_{\mathrm{ref},\alpha}(\varepsilon)}, \tag{4.26}$$

may be too broad, especially in the energy distribution for H ($\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon)$ and $\rho^{\mathrm{B}}_{\mathrm{ref,H}}(\varepsilon)$). Then, we introduce a solution state with overlapped distributions with the reference (OR state), in which the above problematic ε-region is absent (Fig. 4.1). Such a state can be defined using the following characteristic function.

$$\Theta_{\mathrm{OR}}(\mathbf{x}_{\mathrm{HG}},\mathbf{X}_{\mathrm{V}}) = \Theta_{\mathrm{B}}(\mathbf{x}_{\mathrm{HG}})\prod_{\alpha}\prod_{i\in\alpha}\theta\left(\rho^{\mathrm{B}}_{\mathrm{ref},\alpha}\left(u_{\alpha}\left(\mathbf{x}_{\mathrm{H}},\mathbf{x}_{\alpha,i}\right)\right)\right). \tag{4.27}$$

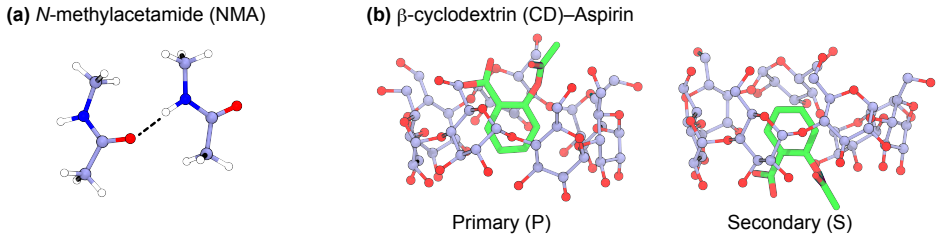


Fig. 4.2: Target binding systems. (a) Self-association of *N*-methylacetamide (NMA). (b) *β*-cyclodextrin (CD)-aspirin binding. The CD-aspirin complexes in which the hydroxyl group of aspirin points towards the primary and secondary faces are labeled as P and S, respectively.

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

$$\theta(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}. \tag{4.28}$$

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

$$\Delta\mu_{\mathrm{G}}^{\mathrm{OR}} = -\frac{1}{\beta}\log\frac{\int d\mathbf{x}_{\mathrm{HG}}\int d\mathbf{X}_{\mathrm{V}}\Theta_{\mathrm{OR}}(\mathbf{x}_{\mathrm{HG}},\mathbf{X}_{\mathrm{V}})\,e^{-\beta\mathcal{V}_{\mathrm{sol}}^{\mathrm{B}}}}{\int d\mathbf{x}_{\mathrm{HG}}\int d\mathbf{X}_{\mathrm{V}}\Theta_{\mathrm{B}}(\mathbf{x}_{\mathrm{HG}})\,e^{-\beta\mathcal{V}_{\mathrm{ref}}^{\mathrm{B}}}}, \tag{4.29}$$

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

$$\Delta\mu_{\mathrm{G}}^{\mathrm{B}} = \Delta\mu_{\mathrm{G}}^{\mathrm{OR}} + \Delta\mu_{\mathrm{G}}^{\mathrm{OR}\to\mathrm{B}}, \tag{4.30}$$

where $\Delta\mu_{\mathrm{G}}^{\mathrm{OR}\to\mathrm{B}}$ is described as

$$\Delta\mu_{\mathrm{G}}^{\mathrm{OR}\to\mathrm{B}} = \frac{1}{\beta}\log P_{\mathrm{OR}}, \tag{4.31}$$

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

$$P_{\mathrm{OR}} = \frac{\int d\mathbf{x}_{\mathrm{HG}}\int d\mathbf{X}_{\mathrm{V}}\Theta_{\mathrm{OR}}(\mathbf{x}_{\mathrm{HG}},\mathbf{X}_{\mathrm{V}})\,e^{-\beta\mathcal{V}_{\mathrm{sol}}^{\mathrm{B}}}}{\int d\mathbf{x}_{\mathrm{HG}}\int d\mathbf{X}_{\mathrm{V}}\Theta_{\mathrm{B}}(\mathbf{x}_{\mathrm{HG}})\,e^{-\beta\mathcal{V}_{\mathrm{sol}}^{\mathrm{B}}}}. \tag{4.32}$$

Since $\rho_{\mathrm{sol},\alpha}^{\mathrm{B}}(\varepsilon)$ is the product of $\rho_{\mathrm{ref},\alpha}^{\mathrm{B}}(\varepsilon)$ and a term referring to the solvent-mediated contribution to the potential of mean force, in principle $\rho_{\mathrm{ref},\alpha}^{\mathrm{B}}(\varepsilon) \neq 0$ when $\rho_{\mathrm{sol},\alpha}^{\mathrm{B}}(\varepsilon) \neq 0$. The difficulty related to Eq. (4.26) is a practical problem due to finite sampling. $\rho_{\mathrm{ref},\alpha}^{\mathrm{B}}(\varepsilon)$ appearing in Eq. (4.27) 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. (4.27), the OR state is a subset of the 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 $\rho_{\mathrm{ref},\alpha}^{\mathrm{B}}(\varepsilon)$. If a sampled configuration in the B state contains a pair energy which corresponds to zero $\rho_{\mathrm{ref},\alpha}^{\mathrm{B}}(\varepsilon)$, that configuration is excluded

from OR. The OR state was called intermediate state in Refs. 70 and 116, however, to avoid possible confusion with an intermediate state in BAR, it is denoted with 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. C.1 of the supplementary material.

4.3 Computational methods

4.3.1 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. 4.2).

TIP3P model was used for water, and the general Amber force field (GAFF) [126,229] was used for the other species. Following modeling scheme was adopted to all the species except for water. We employed the restrained electrostatic potential (RESP) method [230] 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 water and β-cyclodextrin (CD). According to our previous study, [88] we optimized the CD structure at HF/6-31G(d) level calculations. The quantum chemical calculations mentioned above were performed with Gaussian 16 [125] and Antechamber program was used for the RESP method. [231] The initial configurations of the systems of interest were built using Packmol. [127]

All the simulations were performed with GENESIS 2.0. [131–133] The Bussi method was used for generating the NVT and NPT ensembles. [129, 232] The velocity Verlet (VVER) [128] and reversible reference system propagator algorithm (r-RESPA) [233] integrators were employed for the equilibration and production runs, respectively. The time intervals for VVER and r-RESPA were 2 fs and 2.5 fs, respectively. The cutoff distance for the Lennard-Jones (LJ) interactions was 9 Å, and smooth particle-mesh Ewald (SPME) [186] 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, [187,188] and water molecules were treated as rigid molecules using SETTLE method. [130]

4.3.2 *N*-methylacetamide (NMA) systems

Simulation setups

For the computation of $\Delta\mu_{\mathrm{G}}^{\mathrm{D}}$ (Eq. (4.10)), we prepared the trajectories for the system containing an NMA molecule in solvents and for the pure solvent systems, respectively corresponding to the solution and reference systems for the D state. For both systems, the box size was 60^3 Å^3, and the numbers of solvent molecules were set to 1809, 1605, and 1545 in acetone, 1,4-dioxane, and chloroform, respectively. The numbers of solvent molecules were 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 2 ns NVT simulation for equilibration, followed by 10 ns NVT simulation for production.

Regarding $\Delta\mu_{\mathrm{G}}^{\mathrm{B}}$ (Eq. (4.11)), 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 2 ns NVT simulation for equilibration, followed by 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).

$$U_{\mathrm{HFB}}(d) = \begin{cases} 0 & d < d_0 \\ k(d-d_0)^2 & d \geq d_0 \end{cases}. \tag{4.33}$$

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 1 ns NVT simulation for production.

To calculate ΔG° using the PMF-based method (exact), [88, 234] we also conducted 100 ns NVT simulations in the solution system from the 2 ns equilibration mentioned above while applying $U_{\mathrm{HFB}}(d)$ (Eq. (4.33)) with $d_0 = 15$ Å and $k = 10$ kcal mol^{-1} Å^{-2}.

Binding free-energy calculations

The binding free energies, ΔG°, were evaluated through the computation of $\Delta\mu_{\mathrm{G}}^{\mathrm{D}}$, $\Delta\mu_{\mathrm{G}}^{\mathrm{B}}$, and $\Delta G_{\mathrm{corr}}^{\circ}$. In the case of $\Delta\mu_{\mathrm{G}}^{\mathrm{D}}$, 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 $\Delta\mu_{\mathrm{G}}^{\mathrm{D}}$ was done by dividing the solution trajectories into 10 blocks for averaging.

Regarding $\Delta\mu_{\mathrm{G}}^{\mathrm{B}}$, the energy distributions in the solution system, $\rho_{\mathrm{sol,H}}^{\mathrm{B}}(\varepsilon)$, was 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. (4.33). $\rho_{\mathrm{sol,H}}^{\mathrm{B}}(\varepsilon)$ 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, $\rho_{\mathrm{ref,}}^{\mathrm{B}}(\varepsilon)$, the characteristic function for the B state, $\Theta_{\mathrm{B}}(\mathbf{x}_{\mathrm{HB}})$, was constructed using the spatial distribution function for the guest NMA, $g(\mathbf{r})$, and Weeks-Chandler-Andersen (WCA) potential, [235] $u_{\mathrm{WCA}}(\mathbf{x}_{\mathrm{HB}})$, 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. $\rho_{\mathrm{ref},\alpha}^{\mathrm{B}}(\varepsilon)$ was constructed using the configurations obtained from the test-particle insertion that satisfy $d_{\mathrm{min}} \leq 3.5$ Å, $g(\mathbf{r}) > 0$ and $u_{\mathrm{WCA}}(\mathbf{x}_{\mathrm{HB}}) \leq 15$ kcal mol^{-1}. The same characteristic function was used to perform the test-particle insertion for $\Delta G_{\mathrm{corr}}^{\circ}$. The number of insertion was 1000 for each configuration of the reference system. We estimated the statistical error in $\Delta\mu_{\mathrm{G}}^{\mathrm{B}}$ by dividing the solution trajectories into 10 blocks for averaging.

Table 4.1: Information on the trajectories used for the free-energy calculations in the CD-aspirin systems. The values in parentheses indicate the numbers of replicas for the BAR simulations. For the BAR simulations, the last 40 ns and 30 ns were used for the D and B states, respectively.

	States					
	D			B		
	# of Traj.	Simul. length	Total	# of Traj.	Simul. length	Total
BAR	1 (24)	100 ns	2400 ns	10 (57)	150 ns	85500 ns
ER, ER-OR (Solution)	1	20 ns	20 ns	25	20 ns	500 ns
ER, ER-OR (Reference)	1	10 ns	10 ns	25	20 ns	500 ns

For comparison, we also computed ΔG° using the PMF-based approach. [88] In this calculation, only $d_{\min} < 3.5$ Å was used for the bound-complex criteria as well as in the calculation of $\rho^{\mathrm{B}}_{\mathrm{sol},\alpha}(\varepsilon)$. 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.

4.3.3 β-cyclodextrin (CD)-aspirin system

Simulation setups

We prepared the trajectories required for the computation of $\Delta\mu^{\mathrm{D}}_{\mathrm{G}}$ and $\Delta\mu^{\mathrm{B}}_{\mathrm{G}}$ (Table 4.1). The simulation scheme was constructed according to our previous study. [88] 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 K to 298 K during 0.1 ns NVT simulation, we performed 1 ns NVT simulation for equilibration. Then, we decided the system size by 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 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 K to 298 K during 0.1 ns NVT simulation, followed by 0.1 ns NVT simulation for equilibration. Then, we conducted 10 ns NVT simulation for production.

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. [236] In the P and S complexes, the hydroxyl group of aspirin points towards the primary and secondary faces of CD, respectively. We selected 25 different conformations from the trajectories in our previous study for each complex. [88] 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 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, 0.1 ns NVT equilibration was performed. Following this, we conducted 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 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) [206] 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) [237] implemented in GENESIS [238, 239] 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. [240] The last 40 ns trajectory for each state was used for the analysis. Regarding the B state, we selected the 10 configurations of the solution system obtained after the 5 ns production simulations for each bound complex. Then, we conducted the 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 correspond to the fully coupled and decoupled states for each interaction-energy component, respectively. The total number of states is 57. For all the states, we imposed the HFB potential (Eq. (4.33)) 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 [88] between CD and aspirin, $u_{\mathrm{attr}}(\mathbf{x}_{\mathrm{HG}})$, defined as

$$U_{\mathrm{FB}}^{\mathrm{uattr}}(u_{\mathrm{attr}}) = \begin{cases} k\left(u_{\mathrm{attr}} - u_{\mathrm{lower}}\right)^2 & u_{\mathrm{attr}} \leq u_{\mathrm{lower}} \\ 0 & u_{\mathrm{lower}} < u_{\mathrm{attr}} \leq u_{\mathrm{upper}} \\ k\left(u_{\mathrm{attr}} - u_{\mathrm{upper}}\right)^2 & u_{\mathrm{attr}} > u_{\mathrm{upper}} \end{cases}, \tag{4.34}$$

The force constant, k, was set to 10 kcal^{-1} mol, and $\left(u_{\mathrm{lower}}, u_{\mathrm{upper}}\right)$ 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 $\left(u_{\mathrm{lower}}, u_{\mathrm{upper}}\right)$ were determined from the lower and upper limits of u_{attr} observed in the solution systems.

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. 4.3.2), and the same approach was applied for $\Delta\mu_{\mathrm{G}}^{\mathrm{D}}$. Therefore, only the settings specific to the computation of $\Delta\mu_{\mathrm{G}}^{\mathrm{B}}$ 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 $\rho_{\mathrm{ref},\alpha}^{\mathrm{B}}(\varepsilon)$ for each bound pose. For the computation of $\rho_{\mathrm{ref},\alpha}^{\mathrm{B}}(\varepsilon)$ and $\Delta G_{\mathrm{corr}}^\circ$, we constructed the characteristic function for the B state, $\Theta(\mathbf{x}_{\mathrm{HG}})$, in terms of the spatial distribution function for the CoM of aspirin, $g(\mathbf{r})$, and u_{attr}. We computed $\rho_{\mathrm{ref},\alpha}^{\mathrm{B}}(\varepsilon)$ using the configurations obtained from the test-particle insertion of aspirin into the reference trajectories that satisfy $g(\mathbf{r}) > 0$ and $u_{\mathrm{lower}} \leq u_{\mathrm{attr}}(\mathbf{x}_{\mathrm{HG}}) \leq u_{\mathrm{upper}}$. $\left(u_{\mathrm{lower}}, u_{\mathrm{upper}}\right)$ 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 insertion for each configuration was 10000 for both $\rho_{\mathrm{ref},\alpha}^{\mathrm{B}}(\varepsilon)$ and $\Delta G_{\mathrm{corr}}^\circ$. As noted in the last paragraph of Sec. 4.2.2, $\rho_{\mathrm{ref},\alpha}^{\mathrm{B}}(\varepsilon)$ was constructed by discretizing the energy coordinate ε. The bin width of discretization was 0.05 kcal mol^{-1} in the relevant

energy range. In the case of $\Delta G^{\circ}_{\mathrm{corr}}$, the structure of an isolated aspirin was inserted to the spatial region containing CD, with the volume of 20^3 Å. The error in $\Delta\mu^{\mathrm{B}}_{\mathrm{G}}$ was estimated from the different trajectories for the solution system.

We also computed ΔG° using the BAR simulations. Let $\Delta G_{\mathrm{BAR,D}}$ and $\Delta G_{\mathrm{BAR,B}}$ 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

$$\Delta G^{\circ} = \Delta G_{\mathrm{BAR,B}} - \Delta G_{\mathrm{BAR,D}} + \Delta G^{\circ}_{\mathrm{BAR,corr}}, \tag{4.35}$$

where $\Delta G^{\circ}_{\mathrm{BAR,corr}}$ is the standard-state correction. For the computation of $\Delta G_{\mathrm{BAR,B}}$, the snapshots satisfying $u_{\mathrm{lower}} \leq u_{\mathrm{attr}} \leq u_{\mathrm{upper}}$, $d \leq 6$ Å, and the primary/secondary poses criteria for the aspirin's orientation (Fig. S3 of the supplementary material) were used for the fully coupled ($\lambda_{\mathrm{elec}} = \lambda_{\mathrm{vdW}} = 1$) and intermediate states. As for the fully decoupled state ($\lambda_{\mathrm{elec}} = \lambda_{\mathrm{vdW}} = 1$), the aspirin's orientation was not used for the selecting the snapshots. According to Eq. (4.15), the standard-state correction, $\Delta G^{\circ}_{\mathrm{BAR,corr}}$, is expressed as

$$\Delta G^{\circ}_{\mathrm{BAR,corr}} = -\frac{1}{\beta}\log\left(c^{\circ}V\frac{\displaystyle\int d\mathbf{x}_{\mathrm{G}}\int d\mathbf{X}_{\mathrm{V}}\,\Theta^{\mathrm{BAR}}_{\mathrm{B}}\left(\mathbf{x}_{\mathrm{HG}}\right)e^{-\beta\mathcal{V}^{\mathrm{B}}_{\mathrm{ref}}}}{\displaystyle\int d\mathbf{x}_{\mathrm{HG}}\int d\mathbf{X}_{\mathrm{G}}\,e^{-\beta\mathcal{V}^{\mathrm{B}}_{\mathrm{ref}}}}\right), \tag{4.36}$$

where

$$\Theta^{\mathrm{BAR}}_{\mathrm{B}}\left(\mathbf{x}_{\mathrm{HG}}\right) = \begin{cases} 1 & u_{\mathrm{lower}} \leq u_{\mathrm{attr}} \leq u_{\mathrm{upper}} \text{ and } d \leq 6 \text{ Å} \\ 0 & \text{otherwise} \end{cases}. \tag{4.37}$$

$\Delta G^{\circ}_{\mathrm{BAR,corr}}$ was computed using the test-particle insertion of aspirin to the reference trajectories. The number of insertion was 10000. The error in $\Delta G_{\mathrm{BAR,D}}$ was estimated by dividing the trajectory of each state into 8 blocks for averaging, and that in $\Delta G_{\mathrm{BAR,B}}$ was estimated from the different BAR simulation runs.

4.4 Results and discussion

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

Energy distribution

We examine the energy distributions of the host NMA molecule in the solution ($\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon)$) and reference ($\rho^{\mathrm{B}}_{\mathrm{ref,H}}(\varepsilon)$) systems for the bound complex (Fig. 4.3). Note that one of the NMA molecules is regarded as the host, while the other is considered guest. In acetone (Fig. 4.3(a)), $\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon)$ exhibits a broad peak at $\varepsilon \sim -8.5$ kcal mol^{-1}. As illustrated in Fig. 4.2(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 $\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon)$. 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

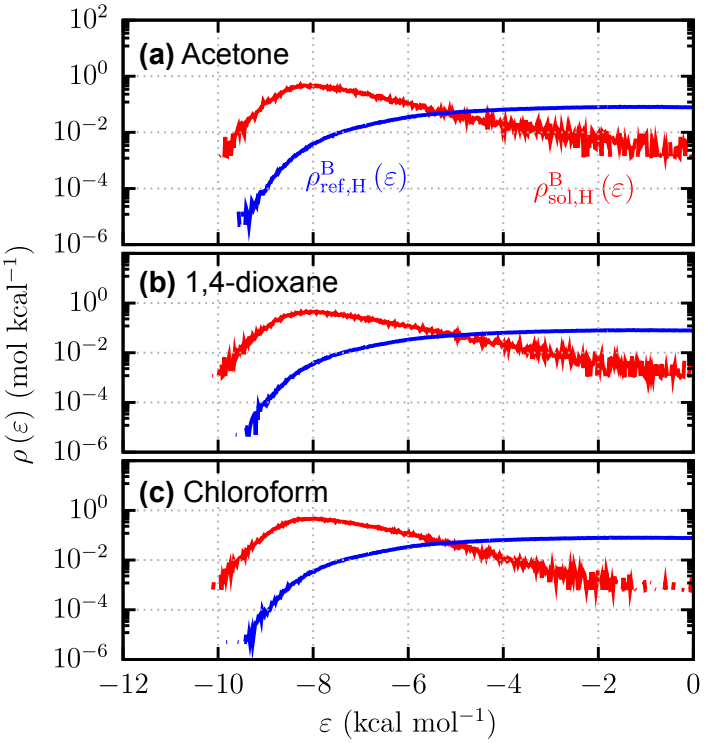


Fig. 4.3: Energy distributions of the host NMA molecule in (a) acetone, (b) 1,4-dioxane, and (c) chloroform.

Table 4.2: Binding free energies for the NMA systems obtained through PMF, ER, and ER-OR methods. The errors are provided at the standard error.

	ΔG° (kcal mol^{-1})		
Solvent	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

der Waals component, -1.049 ± 0.006 kcal mol^{-1}. It is found that the profile of $\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon)$ is largely independent on the solvent species, meaning that the distribution of the bound-complex structures is insensitive to the surrounding environments. Similar to $\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon)$, $\rho^{\mathrm{B}}_{\mathrm{ref,H}}(\varepsilon)$ hardly changes its profile across the solvent species. Since the non-overlapping ε-region ($\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon) \neq 0$ and $\rho^{\mathrm{B}}_{\mathrm{ref,H}}(\varepsilon) = 0$) is sufficiently narrow in all the solvent systems, it is expected that the "solvation free energy" in the bound complex, $\Delta\mu^{\mathrm{B}}_{\mathrm{G}}$, can be computed using the ER method without introducing the OR state (Eq. (4.27)), as will be discussed in the next subsection.

Binding free energy

In this subsection, 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 that the energy distributions in the solution and reference systems ($\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon)$ and $\rho^{\mathrm{B}}_{\mathrm{ref,H}}(\varepsilon)$) overlap well,

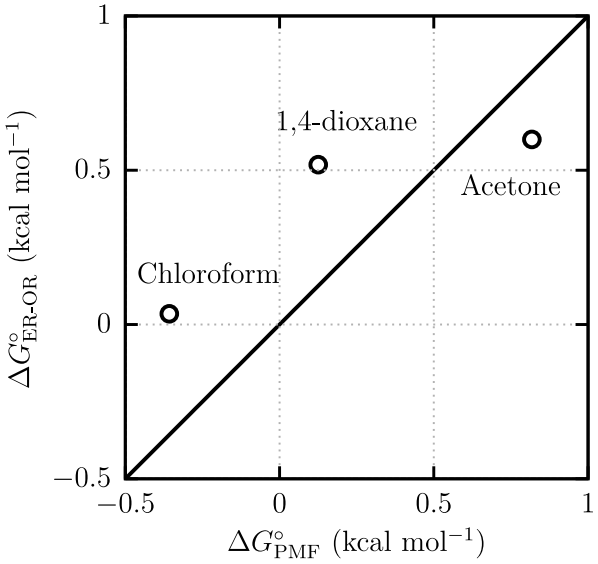


Fig. 4.4: Correlation plots of the binding free energies obtained from the ER-OR method, $\Delta G^\circ_{\text{ER-OR}}$, against those from the PMF, $\Delta G^\circ_{\text{PMF}}$. The error bars are not shown in the figure because the standard errors are smaller than 0.1 kcal mol^{-1} in all the solvent systems.

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 4.2, and the correlation plots of the binding free energies obtained from the ER-OR method, $\Delta G^\circ_{\text{ER-OR}}$, against those from the PMF, $\Delta G^\circ_{\text{PMF}}$, are shown in Fig. 4.4. 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. [228] 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.

4.4.2 β-cyclodextrin (CD)-aspirin system

Structure and energy distribution of CD

We first assess the impact of binding on the structural population of CD. Fig. 4.5(a) illustrates the distribution of the minimum 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 con-

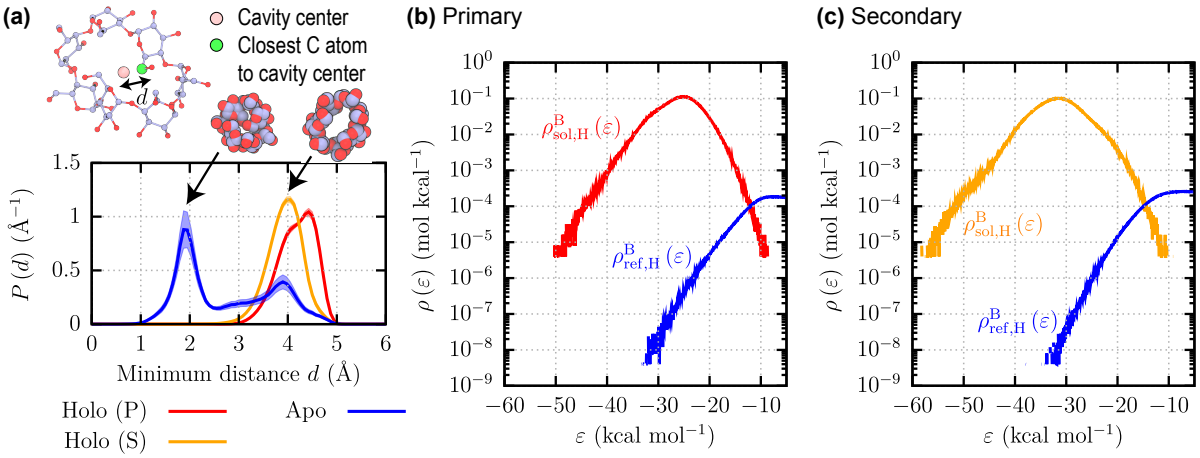


Fig. 4.5: Structural difference of CD between apo- and holo-forms, and energy distributions of CD. (a) Minimum-distance distribution between the cavity center and carbon atoms in CD. (b) and (c) Energy distributions of CD in the solution, $\rho^{\mathrm{B}}_{\mathrm{sol,H}}(\varepsilon)$, and in the reference system, $\rho^{\mathrm{B}}_{\mathrm{ref,H}}(\varepsilon)$, for complex P (b) and those for complex S (c). The cavity center of CD is defined as the center of mass for the ether oxygen atoms.

Table 4.3: 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° (kcal mol^{-1})		
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

formations in the apo-form are also reported by Tang *et al.* [227] as well as by Harris *et al.* [241] 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 those for S are shown in Fig. 4.5(b) and (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. 4.5(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. 4.5(a)).

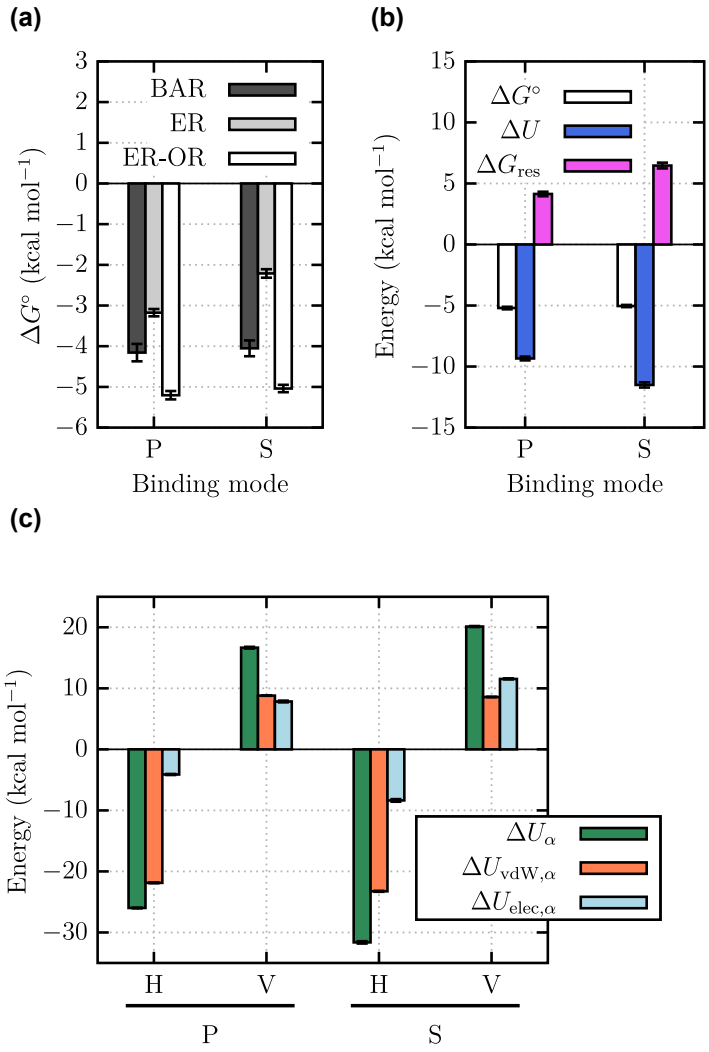


Fig. 4.6: Binding free energy, ΔG, and the decomposition analysis based on the interaction energy. (a) ΔG° evaluated through the BAR, ER, and ER-OR methods. (b) Decomposition of ΔG° into the interaction-energy (ΔU) and other ($\Delta G_{\rm res}$) contributions using Eq. (4.38). In this analysis, the values of ΔG° obtained from the ER-OR methods are used. (c) Decomposition of ΔU into the interaction-energy components for each species α ($\alpha =$ H (CD) or V (water)). The errors are provided at the standard error.

Binding free energy

We summarize the binding free energies, ΔG°, evaluated using the BAR, ER, and ER-OR methods in Fig. 4.6(a) and Table 4.3. 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. On the other hand, 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 ($\rho^{\rm B}_{{\rm sol,H}}(\varepsilon) \neq 0$ and $\rho^{\rm B}_{{\rm ref,H}}(\varepsilon) = 0$) 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 ($\geq$ 120 ns in this system) is required to obtain the reliable estimates of ΔG° using the BAR method. According to Table 4.1, 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 $\sim$2. Introducing a sophisticated scheme of applying the restraint potentials in the BAR simulation, such as the virtual bond algorithm (VBA), [211,212] 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, $\Delta\mu_{\rm G}^{\rm X}$ (X = B or D) (Eqs. (4.10) and (4.11)), into the ensemble average of the interaction energy between aspirin and its surrounding environments in the solution system at state X, $U^{\rm X}$, and the residual part, $\Delta\mu_{\rm res}^{\rm X}$. Thus, ΔG° can be expressed from Eq. (4.14) as

$$\Delta G^\circ = \Delta U + \Delta G_{\rm res}, \tag{4.38}$$

where

$$\Delta U = U^{\rm B} - U^{\rm D}, \tag{4.39}$$

$$\Delta G_{\rm res} = \Delta\mu_{\rm res}^{\rm B} - \Delta\mu_{\rm res}^{\rm D} + \Delta G_{\rm corr}^\circ. \tag{4.40}$$

Note that $\Delta G_{\rm res}$ consists of the contribution from the pair free-energy components, the many-body entropic contributions, and the standard-state correction. Fig. 4.6(b) shows the decomposition of ΔG° using Eq. (4.38). 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 $\Delta G_{\rm res}$, and that the trend of $|\Delta U| > \Delta G_{\rm res}$ 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 $\Delta G_{\rm res}$, resulting in comparable stability of P and S. Since the distribution of $P(d)$ for S is sharper than that for P (Fig. 4.6(a)), the entropic penalty due to the restriction of the CD structure in S may account for the larger value of $\Delta G_{\rm res}$ 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 $\Delta U_{{\rm vdW},\alpha}$ and $\Delta U_{{\rm elec},\alpha}$, respectively. The analysis based on this decomposition is presented in Fig. 4.6(c). Regardless of the complex types, the attractive interaction between aspirin and CD, $\Delta U_{\rm H}$, primarily contributes to ΔU through the van der Waals component, $\Delta U_{{\rm vdW},\alpha}$. 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. [242] The contribution of the interaction energy between aspirin and water, $\Delta U_{\rm V}$, tends to inhibit binding, reflecting the dehydration penalty. It is observed that both $\Delta U_{\rm vdW,V}$ and $\Delta U_{\rm elec,V}$ contribute almost equally to this penalty.

4.5 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 were 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 ΔG° 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 unbound structures are provided. Still, it can be employed with any schemes of structure prediction. For instance, AlphaFold 2/3 [243, 244] is a choice for preparing a structure which 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. [245] 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 cost. 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. [246] 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 ΔG° 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.

Appendix C

Appendix of chapter 5

C.1 supplementary figures

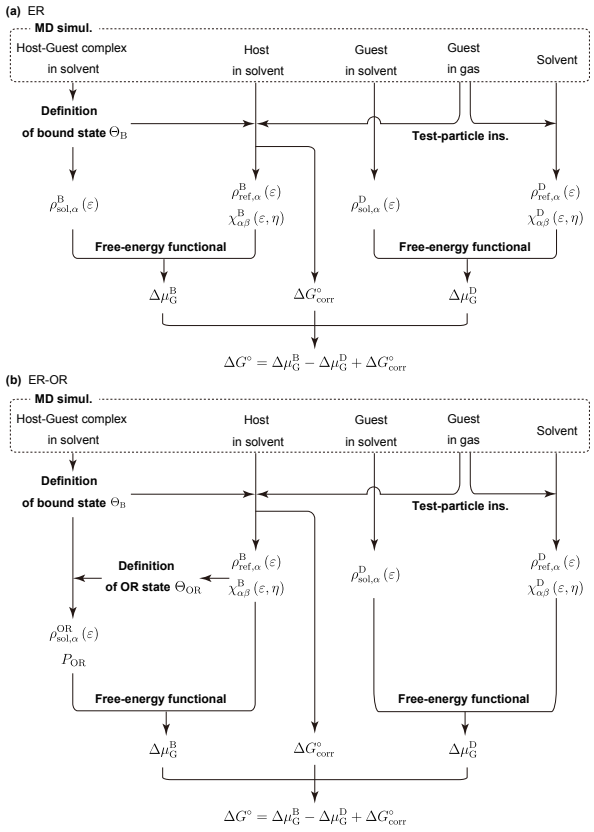


Fig. C.1: Protocols for computing binding free energy, ΔG°, through (a) energy-representation (ER) and (b) ER incorporating a solution state with overlapped distributions with reference (ER-OR). $\rho^{B}_{sol,\alpha}(\varepsilon)$, $\rho^{D}_{sol,\alpha}(\varepsilon)$, and $\rho^{OR}_{sol,\alpha}(\varepsilon)$ are the energy distributions for the bound (B), dissociate (D), and OR states, respectively. $\chi^{X}_{\alpha\beta}(\varepsilon,\eta)$ (X = B or D) is the two-body density-correlation function, P_{OR} is the probability of finding the OR state in the B state, $\Delta\mu^{X}_{G}$ (X = B or D) is the solvation free energy of guest in the X state, and ΔG°_{corr} is the standard-state correction.

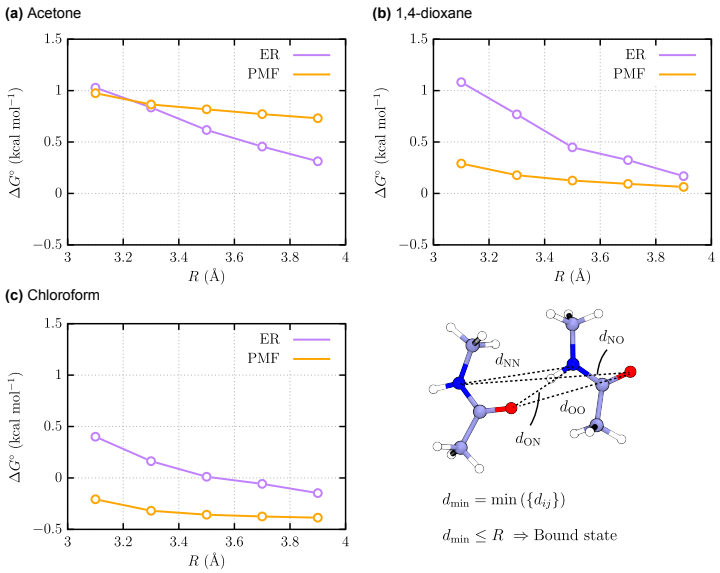


Fig. C.2: Dependency of ΔG° on the definition of the bound state for the NMA systems. (a) acetone, (b) 1,4-dioxane, and (c) chloroform.

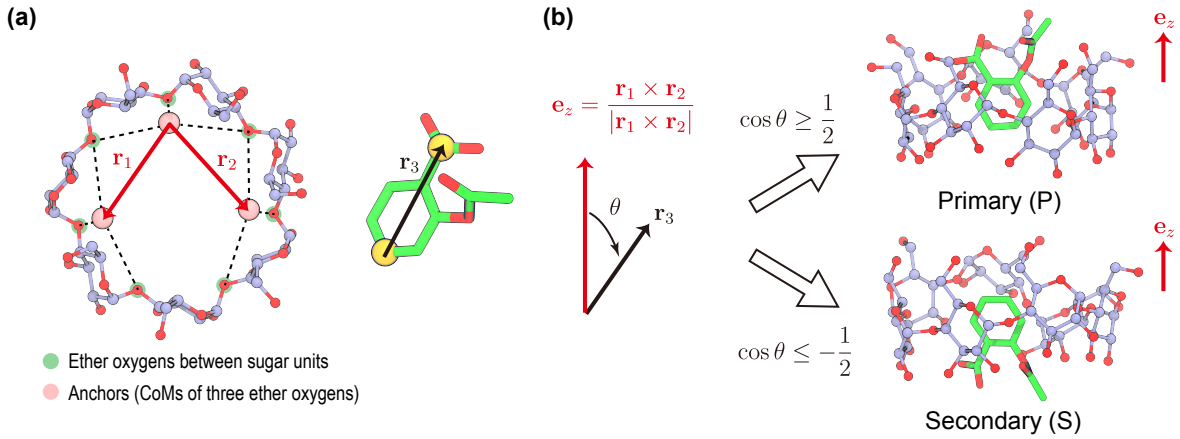


Fig. C.3: Definition of the primary (P) and secondary (S) poses in the β-cyclodextrin (CD)-aspirin system. (a) The vectors on the cross section of CD, $\mathbf{r}_1$ and $\mathbf{r}_2$, and the molecular axis of aspirin, $\mathbf{r}_3$. $\mathbf{r}_1$ and $\mathbf{r}_2$ are the vectors connecting the three anchor particles defined based on the centers of mass (CoMs) of the ether oxygens between the sugar units. The ether oxygens used for defining each anchor particle are indicated as the dash lines in the figure. $\mathbf{r}_3$ is defined using the two carbon atoms in aspirin illustrated as the yellow particles. (b) z-direction ($\mathbf{e}_z$) is defined as the vector product of $\mathbf{r}_1$ and $\mathbf{r}_2$. The angle between $\mathbf{e}_z$ and $\mathbf{r}_3$ is defined as θ, and the P and S poses are defined as the complexes that satisfy $\cos\theta \geq 1/2$ and $\cos\theta \leq -1/2$, respectively.

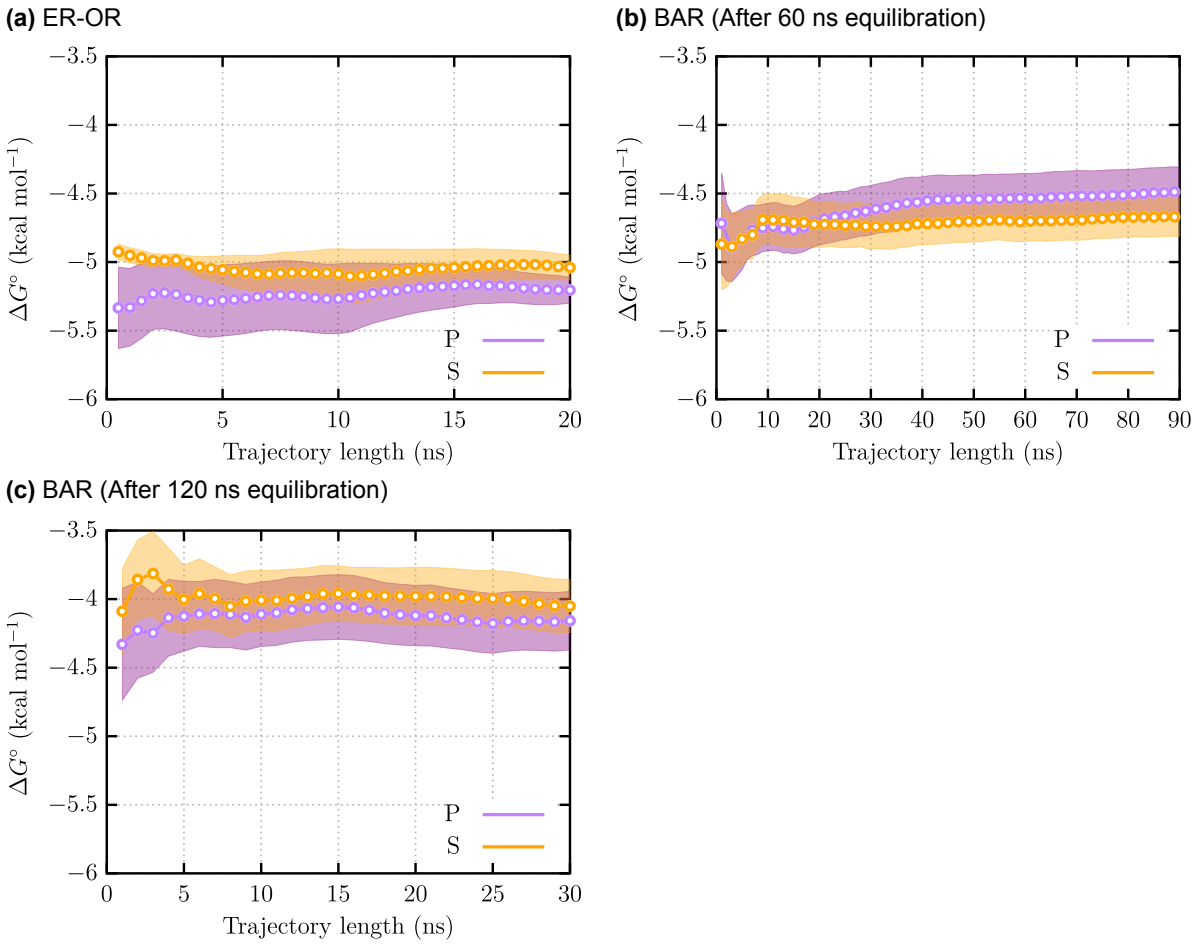


Fig. C.4: Convergence of ΔG° with the trajectory length. (a) ER-OR, (b) BAR (after 60 ns equilibration), and (c) BAR (after 120 ns equilibration).

Chapter 5

General Conclusion

The new molecular theory of solvation thermodynamics and kinetics are formulated based on the theory of solutions and the nonequilibrium statistical mechanics. The following findings have been obtained.

In Chapter 2, the new dynamics theory in the energy representation (ER) solution theory is derived. The Zwanzig-Mori projection operator method and the ER solution theory give the energy-represented generalized Langevin equation (ERGLE). In this equation, the molecular motions are expressed as the time-dependent distribution function on the solute-solvent interaction energy (energy coordinate). Since its derivation does not depend on any approximations, it will be the theoretical basis of the dynamic process in functional molecular systems. Neglecting the inertial and memory effects yields the energy-represented Smoluchowski-Vlasov (ERSV) equation and Smoluchowski (ERS) equation. Although these equations need input functions, such as the energy-represented diffusion coefficient, the free energy profile and the direct correlation function on the energy coordinate, they are easily evaluated by MD simulations. The ERSV and ERS equation have were applied to the solvation dynamics of water triggered by the photoexcitation of benzonitrile. The time constant of the relaxation of the solvation structure showed the validity of the new theory in the long-time region. The comparison of the results obtained by the ERSV equation and the ERS equation respectively indicated the importance of the collective motion during the initial stage of the relaxation.

Chapter 3 has applied the new theory introduced in Chapter 2 to the solvation dynamics of Prodan in water and alcohol solvents (methanol, ethanol, and 1-propanol). It is found that the populations of both the stabilized and destabilized molecules are higher for water. Although the profiles of the energy-represented diffusion coefficient divided by the translational diffusion coefficient of the solvent are almost the same for the alcohol solvents, that of water is higher than that of alcohol solvents in the stabilized region. The decomposition analysis has shown that the high diffusivity of water is due to the presence of water molecules that are highly stabilized by the naphthalene moiety of Prodan. The solvation time correlation functions (STCFs) and their time coefficients are calculated by the ERSV equation. The time coefficients are well reproduced with the ERSV equation in the long-time region, indicating the validity of the ERSV equation in this time region. The time coefficients calculated by the ERSV equation are always smaller than those by the ERS equation for all solvents. This result is consistent with that obtained in Chapter 2 and show that the collective motion tends to promote the relaxation of the solvation structure.

In Chapter 4, the new framework of computing the binding free energy is formu-

lated. First, the rigorous thermodynamic cycle for the binding free energy calculation was introduced. In this cycle, the binding free energy is decomposed into three terms, the conditional solvation free energy in the bound state, the solvation free energy in the bulk, and the standard state correction term, and the former two terms can be calculated based on the ER theory. When the structure of the host molecule is changed by the guest molecule, the overlap between the energy distribution functions in the solution and reference systems is small, making it difficult to calculate the conditional solvation free energy in the bound state. To avoid this difficulty, the intermediate solution state which has the overlapped distribution with the reference (OR state) is introduced. This intermediate state allows the more robust protocol (ER-OR) for the conditional solvation free energy in the bound state. The new method was applied to the self-association, or dimerization of *N*-methylacetamide (NMA) in different solvents (acetone, 1,4-dioxane, and chloroform). Since the energy distribution of the solution system has enough overlapped region with that of the reference systems, the values of the binding free energy were almost identical with and without the OR method. Their deviations from those by the rigorous method were within 0.5 kcal mol^{-1} for all solvents. The binding free energy of aspirin to β-cyclodextrin (CD) in water was the next target. In this system the structure of CD is affected by aspirin molecule, leading to the poor overlap between the energy distributions of the solution and the reference systems. The values of the binding free energy differed by more than 1 kcal mol^{-1}, and those obtained from the ER-OR method were much closer to those from the rigorous method, showing the robustness of the new method.

Finally, we will show the summary of this dissertation. The new theoretical frameworks based on the concept of solvation were formulated through the theory of solutions and the nonequilibrium statistical mechanics. Chapter 2 and 3 introduced the new diffusion equation theory on the energy coordinate. It can describe the long-timescale dynamics, and the systematic analyses based on it were shown. In Chapter 4, the new scheme of calculating the binding free energy using the endpoints density-functional theory was derived. The new scheme was shown to be robust and to produce results consistent with the rigorous method. We believe that these theoretical frameworks developed in this work will deepen our understanding of the thermodynamics and the kinetics in solution systems.

References

[1] J.-P. Hansen, I. R. McDonald, *Theory of Simple Liquids: With Applications to Soft Matter*, Academic Press, 2013.

[2] A. Ben-Naim, *Molecular Theory of Solutions*, OUP Oxford, 2006.

[3] A. Nitzan, *Chemical Dynamics in Condensed Phases: Relaxation, Transfer, and Reactions in Condensed Molecular Systems*, of *Oxford Graduate Texts*, Oxford University Press, Oxford, New York, 2013.

[4] D. Chandler, H. C. Andersen, *J. Chem. Phys.*, 1972, **57**, 1930–1937.

[5] H. Sato, *Phys. Chem. Chem. Phys.*, 2013, **15**, 7450.

[6] F. Hirata, *Molecular Theory of Solvation*, Springer Science & Business Media, 2003.

[7] K. Kasahara, H. Sato, *Phys. Chem. Chem. Phys.*, 2017, **19**, 27917–27929.

[8] D. Beglov, B. Roux, *J. Phys. Chem. B*, 1997, **101**, 7821–7826.

[9] A. Kovalenko, F. Hirata, *Chem. Phys. Lett.*, 1998, **290**, 237–244.

[10] N. Yoshida, T. Imai, S. Phongphanphanee, A. Kovalenko, F. Hirata, *J. Phys. Chem. B*, 2009, **113**, 873–886.

[11] Y. Kiyota, N. Yoshida, F. Hirata, *J. Chem. Theory Comput.*, 2011, **7**, 3803–3815.

[12] S. Ten-no, F. Hirata, S. Kato, *Chem. Phys. Lett.*, 1993, **214**, 391–396.

[13] S. Ten-no, F. Hirata, S. Kato, *J. Chem. Phys.*, 1994, **100**, 7443–7453.

[14] H. Sato, F. Hirata, S. Kato, *J. Chem. Phys.*, 1996, **105**, 1546–1551.

[15] D. Yokogawa, *Bull. Chem. Soc. Jpn.*, 2018, **91**, 1540–1545.

[16] K. Imamura, D. Yokogawa, H. Sato, *J. Chem. Phys.*, 2024, **160**, 050901.

[17] K. Naka, A. Morita, S. Kato, *J. Chem. Phys.*, 1999, **110**, 3484–3492.

[18] K. Naka, A. Morita, S. Kato, *J. Chem. Phys.*, 1999, **111**, 481–491.

[19] N. Yoshida, T. Yamaguchi, *J. Chem. Phys.*, 2020, **152**, 114108.

[20] N. Yoshida, T. Yamaguchi, H. Nakano, *Chem. Phys. Lett.*, 2022, **797**, 139579.

[21] A. Morita, S. Kato, *J. Am. Chem. Soc.*, 1997, **119**, 4021–4032.

[22] J. P. Boon, S. Yip, *Molecular Hydrodynamics*, McGraw-Hill Inc.,US, New York usw., Hamburg, 1980.

[23] U. Balucani, M. Zoppi, *Dynamics of the Liquid State*, Clarendon Press, 1995.

[24] S. W. Haan, *Phys. Rev. A*, 1979, **20**, 2516–2520.

[25] D. F. Calef, P. G. Wolynes, *J. Chem. Phys.*, 1983, **78**, 4145–4153.

[26] S. A. Egorov, *Phys. Rev. Lett.*, 2004, **93**, 023004.

[27] H. Mori, *Prot. Theor. Phys.*, 1965, **33**, 423–455.

[28] R. Zwanzig, *Nonequilibrium Statistical Mechanics*, Oxford University Press, 2001.

[29] M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, Oxford University Press, 2023.

[30] T. Yamaguchi, T. Matsuoka, S. Koda, *J. Chem. Phys.*, 2005, **123**, 034504.

[31] S. W. Lovesey, *J. Phys. C: Solid State Phys.*, 1971, **4**, 3057.

[32] J. R. D. Copley, S. W. Lovesey, *Rep. Prog. Phys.*, 1975, **38**, 461–563.

[33] D. R. Reichman, P. Charbonneau, *J. Stat. Mech.: Theory Exp.*, 2005, **2005**, P05013.

[34] F. Hirata, *J. Chem. Phys.*, 1992, **96**, 4619–4624.

[35] S.-H. Chong, F. Hirata, *Phys. Rev. E*, 1998, **57**, 1691–1701.

[36] S.-H. Chong, F. Hirata, *J. Mol. Liq.*, 1998, **77**, 105–120.

[37] K. Iida, H. Sato, *J. Chem. Phys.*, 2012, **137**, 034506.

[38] S.-H. Chong, F. Hirata, *J. Chem. Phys.*, 1999, **111**, 3083–3094.

[39] S.-H. Chong, F. Hirata, *J. Chem. Phys.*, 1999, **111**, 3095–3104.

[40] S.-H. Chong, F. Hirata, *Phys. Rev. E*, 1998, **58**, 6188–6198.

[41] S.-H. Chong, F. Hirata, *Phys. Rev. E*, 1998, **58**, 7296–7308.

[42] S.-H. Chong, W. Götze, *Phys. Rev. E*, 2002, **65**, 051201.

[43] K. Kasahara, H. Sato, *J. Chem. Phys.*, 2014, **140**, 244110.

[44] K. Kasahara, H. Sato, *J. Chem. Phys.*, 2016, **145**, 194502.

[45] F. O. Raineri, B.-C. Perng, H. L. Friedman, *Chem. Phys.*, 1994, **183**, 187–205.

[46] F. O. Raineri, H. Resat, B.-C. Perng, F. Hirata, H. L. Friedman, *J. Chem. Phys.*, 1994, **100**, 1477–1491.

[47] H. L. Friedman, F. O. Raineri, F. Hirata, B.-C. Perng, *J. Stat. Phys.*, 1995, **78**, 239–266.

[48] F. Hirata, T. Munakata, F. Raineri, H. L. Friedman, *J. Mol. Liq.*, 1995, **65–66**, 15–22.

[49] T. Ishida, F. Hirata, S. Kato, *J. Chem. Phys.*, 1999, **110**, 11423–11432.

[50] K. Nishiyama, F. Hirata, T. Okada, *Chem. Phys. Lett.*, 2000, **330**, 125–131.

[51] K. Nishiyama, F. Hirata, T. Okada, *J. Chin. Chem. Soc.*, 2000, **47**, 837–842.

[52] K. Nishiyama, F. Hirata, T. Okada, *J. Mol. Struct.*, 2001, **565–566**, 31–34.

[53] K. Nishiyama, F. Hirata, T. Okada, *J. Chem. Phys.*, 2003, **118**, 2279–2285.

[54] K. Nishiyama, T. Yamaguchi, F. Hirata, T. Okada, *Pure Appl. Chem.*, 2004, **76**, 71–77.

[55] K. Nishiyama, T. Yamaguchi, F. Hirata, *J. Phys. Chem. B*, 2009, **113**, 2800–2804.

[56] A. Chandra, B. Bagchi, *J. Chem. Phys.*, 1989, **91**, 1829–1842.

[57] A. Yoshimori, *J. Theor. Comput. Chem.*, 2004, **03**, 117–144.

[58] A. Yoshimori, *J. Phys. Soc. Japan*, 2011, **80**, 034801.

[59] T. Yamaguchi, N. Yoshida, *J. Chem. Phys.*, 2021, **154**, 044504.

[60] N. Matubayasi, M. Nakahara, *J. Chem. Phys.*, 2000, **113**, 6070–6081.

[61] N. Matubayasi, M. Nakahara, *J. Chem. Phys.*, 2002, **117**, 3605–3616.

[62] N. Matubayasi, M. Nakahara, *J. Chem. Phys.*, 2003, **119**, 9686–9702.

[63] A. I. Frolov, *arXiv preprint arXiv:1502.04355*, 2015.

[64] N. Matubayasi, *Bull. Chem. Soc. Jpn.*, 2019, **92**, 1910–1927.

[65] R. M. Levy, N. Matubayasi, B. W. Zhang, *J. Phys. Chem. B*, 2020, **124**, 11771–11782.

[66] N. Matubayasi, K. K. Liang, M. Nakahara, *J. Chem. Phys.*, 2006, **124**, 154908.

[67] A. Date, R. Ishizuka, N. Matubayasi, *Phys. Chem. Chem. Phys.*, 2016, **18**, 13223–13231.

[68] N. Matubayasi, W. Shinoda, M. Nakahara, *J. Chem. Phys.*, 2008, **128**, 195107.

[69] T. Mizuguchi, N. Matubayasi, *J. Phys. Chem. B*, 2018, **122**, 3219–3229.

[70] T. Kawakami, I. Shigemoto, N. Matubayasi, *J. Chem. Phys.*, 2012, **137**, 234903.

[71] T. Kawakami, I. Shigemoto, N. Matubayasi, *J. Chem. Phys.*, 2018, **148**, 214903.

[72] K. Yamada, N. Matubayasi, *Macromolecules*, 2020, **53**, 775–788.

[73] H. Kojima, K. Handa, K. Yamada, N. Matubayasi, *J. Phys. Chem. B*, 2021, **125**, 9357–9371.

[74] N. Yasoshima, T. Ishiyama, N. Matubayasi, *J. Phys. Chem. B*, 2022, **126**, 4389–4400.

[75] K. Yamada, N. Matubayasi, *Macromolecules*, 2023, **56**, 3857–3872.

[76] Y. Yamamori, R. Ishizuka, Y. Karino, S. Sakuraba, N. Matubayasi, *J. Chem. Phys.*, 2016, **144**, 085102.

[77] Y. Yamamori, N. Matubayasi, *J. Chem. Phys.*, 2017, **146**, 225103.

[78] K. Masutani, Y. Yamamori, K. Kim, N. Matubayasi, *J. Chem. Phys.*, 2019, **150**, 145101.

[79] N. Matubayasi, K. Masutani, *Biophy. physicobiology*, 2019, **16**, 185–195.

[80] T. M. Do, D. Horinek, N. Matubayasi, *Phys. Chem. Chem. Phys.*, 2024, **26**, 11880–11892.

[81] R. Ishizuka, N. Matubayasi, K.-M. Tu, Y. Umebayashi, *J. Phys. Chem. B*, 2015, **119**, 1579–1587.

[82] N. Yamamoto, I. Nakakuki, N. Matubayasi, *J. Chem. Phys.*, 2018, **149**, 014504.

[83] S. Tanaka, N. Yamamoto, K. Kasahara, Y. Ishii, N. Matubayasi, *J. Phys. Chem. B*, 2022, **126**, 5274–5290.

[84] H. Takahashi, N. Matubayasi, M. Nakahara, T. Nitta, *J. Chem. Phys.*, 2004, **121**, 3989–3999.

[85] H. Takahashi, H. Ohno, R. Kishi, M. Nakano, N. Matubayasi, *J. Chem. Phys.*, 2008, **129**, 205103.

[86] H. Takahashi, K. Maruyama, Y. Karino, A. Morita, M. Nakano, P. Jungwirth, N. Matubayasi, *J. Phys. Chem. B*, 2011, **115**, 4745–4751.

[87] N. Matubayasi, H. Takahashi, *J. Chem. Phys.*, 2012, **136**, 044505.

[88] K. Kasahara, R. Masayama, K. Okita, N. Matubayasi, *J. Chem. Phys.*, 2021, **155**, 204503.

[89] K. Kasahara, R. Masayama, K. Okita, N. Matubayasi, *J. Chem. Phys.*, 2023, **159**, 134103.

[90] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Springer Science & Business Media, 2007.

[91] I. L. Medintz, N. Hildebrandt, *FRET - Förster Resonance Energy Transfer: From Theory to Applications*, John Wiley & Sons, 2013.

[92] M. Maroncelli, *J. Mol. Liq.*, 1993, **57**, 1–37.

[93] B. Bagchi, *Molecular Relaxation in Liquids*, Oxford University Press, USA, 2012.

[94] G. Gunther, L. Malacrida, D. M. Jameson, E. Gratton, S. A. Sánchez, *Acc. Chem. Res.*, 2021, **54**, 976–987.

[95] S. Osella, S. Knippenberg, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2021, **1863**, 183494.

[96] S. A. Rice, *Diffusion-Limited Reactions*, Elsevier, 1985.

[97] N. Pottier, *Nonequilibrium Statistical Physics: Linear Irreversible Processes*, Oxford University Press, 2009.

[98] K. Lindenberg, R. Metzler, G. Oshanin, *Chemical Kinetics: Beyond The Textbook*, World Scientific, 2019.

[99] A. Chandra, B. Bagchi, *Chem. Phys. Lett.*, 1988, **151**, 47–53.

[100] A. Chandra, B. Bagchi, *J. Chem. Phys.*, 1991, **94**, 8367–8377.

[101] R. Schilling, T. Scheidsteger, *Phys. Rev. E*, 1997, **56**, 2932–2949.

[102] L. Fabbian, A. Latz, R. Schilling, F. Sciortino, P. Tartaglia, C. Theis, *Phys. Rev. E*, 1999, **60**, 5768–5777.

[103] T. Theenhaus, R. Schilling, A. Latz, M. Letz, *Phys. Rev. E*, 2001, **64**, 051505.

[104] F. Hirata, P. J. Rossky, *Chem. Phys. Lett.*, 1981, **83**, 329–334.

[105] F. Hirata, P. J. Rossky, B. M. Pettitt, *J. Chem. Phys.*, 1983, **78**, 4133–4144.

[106] F. O. Raineri, Y. Zhou, H. L. Friedman, G. Stell, *Chem. Phys.*, 1991, **152**, 201–220.

[107] T. Yamaguchi, F. Hirata, *J. Chem. Phys.*, 2002, **117**, 2216–2224.

[108] T. Yamaguchi, S.-H. Chong, F. Hirata, *J. Chem. Phys.*, 2003, **119**, 1021–1034.

[109] T. Yamaguchi, S.-H. Chong, F. Hirata, *J. Mol. Liq.*, 2004, **112**, 117–124.

[110] A. E. Kobryn, T. Yamaguchi, F. Hirata, *J. Mol. Liq.*, 2005, **119**, 7–13.

[111] S.-H. Chong, F. Hirata, *J. Chem. Phys.*, 1998, **108**, 7339–7349.

[112] T. Yamaguchi, T. Matsuoka, S. Koda, *J. Chem. Phys.*, 2007, **127**, 234501.

[113] T. Yamaguchi, T. Matsuoka, S. Koda, *J. Chem. Phys.*, 2007, **126**, 144505.

[114] A. E. Kobryn, F. Hirata, *J. Chem. Phys.*, 2007, **126**, 044504.

[115] A. Kovalenko, F. Hirata, *J. Chem. Phys.*, 1999, **110**, 10095–10112.

[116] S. Sakuraba, N. Matubayasi, *J. Comput. Chem.*, 2014, **35**, 1592–1608.

[117] A. I. Frolov, *J. Chem. Theory Comput.*, 2015, **11**, 2245–2256.

[118] R. Urano, G. A. Pantelopulos, J. E. Straub, *J. Phys. Chem. B*, 2019, **123**, 2546–2557.

[119] E. Fick, G. Sauermann, *The Quantum Statistics of Dynamic Processes*, Springer Berlin Heidelberg, 1990.

[120] T. Koide, *Phys. Rev. E*, 2005, **72**, 026135.

[121] K.-M. Tu, R. Ishizuka, N. Matubayasi, *J. Chem. Phys.*, 2014, **141**, 244507.

[122] R. A. Marcus, *J. Chem. Phys.*, 1965, **43**, 3477–3489.

[123] E. A. Carter, J. T. Hynes, *J. Phys. Chem.*, 1989, **93**, 2184–2187.

[124] B. B. Laird, W. H. Thompson, *J. Chem. Phys.*, 2007, **126**, 211104.

[125] 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, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, *Gaussian 16 Revision C.01*, 2016.

[126] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

[127] L. Martínez, R. Andrade, E. G. Birgin, J. M. Martínez, *J. Comput. Chem.*, 2009, **30**, 2157–2164.

[128] W. C. Swope, H. C. Andersen, P. H. Berens, K. R. Wilson, *J. Chem. Phys.*, 1982, **76**, 637–649.

[129] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

[130] S. Miyamoto, P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962.

[131] J. Jung, T. Mori, C. Kobayashi, Y. Matsunaga, T. Yoda, M. Feig, Y. Sugita, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2015, **5**, 310–323.

[132] C. Kobayashi, J. Jung, Y. Matsunaga, T. Mori, T. Ando, K. Tamura, M. Kamiya, Y. Sugita, *J. Comput. Chem.*, 2017, **38**, 2193–2206.

[133] J. Jung, C. Kobayashi, K. Kasahara, C. Tan, A. Kuroda, K. Minami, S. Ishiduki, T. Nishiki, H. Inoue, Y. Ishikawa, M. Feig, Y. Sugita, *J. Comput. Chem.*, 2021, **42**, 231–241.

[134] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.

[135] J. H. Ferziger, M. Peric, *Computational Methods for Fluid Dynamics*, Springer Science & Business Media, 2012.

[136] K. Kasahara, H. Sato, *J. Comput. Chem.*, 2018, **39**, 1491–1497.

[137] LLC. Schrödinger, W. DeLano, *PyMOL*, 2020.

[138] L. Sjögren, *Phys. Rev. A*, 1980, **22**, 2883.

[139] L. Sjögren, *Phys. Rev. A*, 1980, **22**, 2866.

[140] L. Sjogren, *J. Phys. C: Solid State Phys.*, 1980, **13**, 705.

[141] G. B. Dutt, *ChemPhysChem*, 2005, **6**, 413–418.

[142] B. Bagchi, D. W. Oxtoby, G. R. Fleming, *Chem. Phys.*, 1984, **86**, 257–267.

[143] M. Maroncelli, J. MacInnis, G. R. Fleming, *Science*, 1989, **243**, 1674–1681.

[144] A. Samanta, *J. Phys. Chem. B*, 2006, **110**, 13704–13716.

[145] Y. Nagasawa, H. Miyasaka, *Phys. Chem. Chem. Phys.*, 2014, **16**, 13008–13026.

[146] A. Samanta, *J. Phys. Chem. Lett.*, 2010, **1**, 1557–1562.

[147] G. U. Nienhaus, J. Wiedenmann, *ChemPhysChem*, 2009, **10**, 1369–1379.

[148] H. Shweta, N. Pal, M. K. Singh, S. D. Verma, S. Sen in *Reviews in Fluorescence 2017*, C. D. Geddes (Ed.), Springer International Publishing, Cham, 2018, pp. 231–279.

[149] J. R. Lakowicz, *J. Biochem. Biophys. Methods*, 1980, **2**, 91–119.

[150] M. Amaro, R. Šachl, P. Jurkiewicz, A. Coutinho, M. Prieto, M. Hof, *Biophys. J.*, 2014, **107**, 2751–2760.

[151] D. Frenkel, B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Elsevier, 2001.

[152] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, 2017.

[153] M. L. Horng, J. A. Gardecki, A. Papazyan, M. Maroncelli, *J. Phys. Chem.*, 1995, **99**, 17311–17337.

[154] M. S. Skaf, B. M. Ladanyi, *J. Phys. Chem.*, 1996, **100**, 18258–18268.

[155] M. N. Kobrak, *J. Chem. Phys.*, 2006, **125**, 064502.

[156] S. Mukherjee, S. Mondal, S. Acharya, B. Bagchi, *J. Phys. Chem. B*, 2018, **122**, 11743–11761.

[157] L. Nilsson, B. Halle, *Proc. Natl. Acad. Sci. U.S.A.*, 2005, **102**, 13867–13872.

[158] D. Marx, J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*, Cambridge University Press, 2009.

[159] C. Allolio, M. Sajadi, N. P. Ernsting, D. Sebastiani, *Angew. Chem. Int. Ed.*, 2013, **52**, 1813–1816.

[160] A. Petrone, G. Donati, P. Caruso, N. Rega, *J. Am. Chem. Soc.*, 2014, **136**, 14866–14874.

[161] B. Bagchi, *Annu. Rev. Phys. Chem.*, 1989, **40**, 115–141.

[162] H. Feng, W. Gao, J. Nie, J. Wang, X. Chen, L. Chen, X. Liu, H.-D. Lüdemann, Z. Sun, *J. Mol. Model.*, 2013, **19**, 73–82.

[163] D. Stadelmaier, W. Köhler, *Macromolecules*, 2009, **42**, 9147–9152.

[164] K. Okita, K. Kasahara, N. Matubayasi, *J. Chem. Phys.*, 2022, **157**, 244505.

[165] A. Pandey, R. Rai, M. Pal, S. Pandey, *Phys. Chem. Chem. Phys.*, 2014, **16**, 1559–1568.

[166] R. Karmakar, A. Samanta, *J. Phys. Chem. A*, 2002, **106**, 6670–6675.

[167] B. A. Rowe, C. A. Roach, J. Lin, V. Asiago, O. Dmitrenko, S. L. Neal, *J. Phys. Chem. A*, 2008, **112**, 13402–13412.

[168] M. Raguž, J. Brnjas-Kraljević, *J. Chem. Inf. Model.*, 2005, **45**, 1636–1640.

[169] C. C. Vequi-Suplicy, K. Coutinho, M. T. Lamy, *J. Fluoresc.*, 2015, **25**, 621–629.

[170] P. Pospíšil, L. Cwiklik, J. Sýkora, M. Hof, G. M. Greetham, M. Towrie, A. Vlček, *J. Phys. Chem. B*, 2021, **125**, 13858–13867.

[171] A. Marini, A. Muñoz-Losa, A. Biancardi, B. Mennucci, *J. Phys. Chem. B*, 2010, **114**, 17128–17135.

[172] G. Parisio, A. Marini, A. Biancardi, A. Ferrarini, B. Mennucci, *J. Phys. Chem. B*, 2011, **115**, 9980–9989.

[173] C. C. Vequi-Suplicy, Y. Orozco-Gonzalez, M. T. Lamy, S. Canuto, K. Coutinho, *J. Chem. Phys.*, 2020, **153**, 244104.

[174] B. Mennucci, M. Caricato, F. Ingrosso, C. Cappelli, R. Cammi, J. Tomasi, G. Scalmani, M. J. Frisch, *J. Phys. Chem. B*, 2008, **112**, 414–423.

[175] E. K. Krasnowska, E. Gratton, T. Parasassi, *Biophys. J.*, 1998, **74**, 1984–1993.

[176] F. Moyano, M. A. Biasutti, J. J. Silber, N. M. Correa, *J. Phys. Chem. B*, 2006, **110**, 11838–11846.

[177] B. A. Rowe, S. L. Neal, *J. Phys. Chem. B*, 2006, **110**, 15021–15028.

[178] R. Adhikary, C. A. Barnes, J. W. Petrich, *J. Phys. Chem. B*, 2009, **113**, 11999–12004.

[179] N. Ito, N. M. Watanabe, Y. Okamoto, H. Umakoshi, *Biophys. J.*, 2023, S0006349523006896.

[180] T. Yanai, D. P. Tew, N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.

[181] C. M. Breneman, K. B. Wiberg, *J. Comput. Chem.*, 1990, **11**, 361–373.

[182] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. Mackerell Jr., *J. Comput. Chem.*, 2010, **31**, 671–690.

[183] S. Jo, T. Kim, V. G. Iyer, W. Im, *J. Comput. Chem.*, 2008, **29**, 1859–1865.

[184] J. Huang, A. D. MacKerell Jr, *J. Comput. Chem.*, 2013, **34**, 2135–2145.

[185] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

[186] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

[187] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

[188] H. C. Andersen, *J. Comput. Phys.*, 1983, **52**, 24–34.

[189] C.-P. Hsu, X. Song, R. A. Marcus, *J. Phys. Chem. B*, 1997, **101**, 2546–2551.

[190] F. Ingrosso, B. Mennucci, J. Tomasi, *J. Mol. Liq.*, 2003, **108**, 21–46.

[191] B. Bagchi, B. Jana, *Chem. Soc. Rev.*, 2010, **39**, 1936.

[192] E. Heid, C. Schröder, *J. Phys. Chem. B*, 2017, **121**, 9639–9646.

[193] M. Gerecke, C. Richter, M. Quick, I. N. Ioffe, R. Mahrwald, S. A. Kovalenko, N. P. Ernsting, *J. Phys. Chem. B*, 2017, **121**, 9631–9638.

[194] R. A. Weinberg, R. A. Weinberg, *The Biology of Cancer*, Garland Science, 2013.

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

[196] T. Daviter, C. M. Johnson, S. H. McLaughlin, M. A. Williams, *Protein-Ligand Interactions: Methods and Applications*, Springer US, 2021.

[197] M. K. Gilson, H.-X. Zhou, *Annu. Rev. Biochem.*, 2007, **36**, 21–42.

[198] R. Baron, *Computational Drug Discovery and Design*, Humana Press, 2011.

[199] J. R. Schames, R. H. Henchman, J. S. Siegel, C. A. Sotriffer, H. Ni, J. A. McCammon, *J. Med. Chem.*, 2004, **47**, 1879–1881.

[200] W. L. Jorgensen, *Science*, 2004, **303**, 1813–1818.

[201] J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson, V. S. Pande, *Curr. Opin. Struct. Biol.*, 2011, **21**, 150–160.

[202] M. De Vivo, M. Masetti, G. Bottegoni, A. Cavalli, *J. Med. Chem.*, 2016, **59**, 4035–4061.

[203] E. King, E. Aitchison, H. Li, R. Luo, *Front. Mol. Biosci.*, 2021, **8**, 712085.

[204] J. G. Kirkwood, *J. Chem. Phys.*, 1935, **3**, 300–313.

[205] R. W. Zwanzig, *J. Chem. Phys.*, 1954, **22**, 1420–1426.

[206] C. H. Bennett, *J. Comput. Phys.*, 1976, **22**, 245–268.

[207] W. L. Jorgensen, J. K. Buckner, S. Boudon, J. Tirado-Rives, *J. Chem. Phys.*, 1988, **89**, 3742–3746.

[208] M. Gilson, J. Given, B. Bush, J. McCammon, *Biophys. J.*, 1997, **72**, 1047–1069.

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

[210] H. Fujitani, Y. Tanida, A. Matsuura, *Phys. Rev. E*, 2009, **79**, 021914.

[211] S. Boresch, F. Tettinger, M. Leitgeb, M. Karplus, *J. Phys. Chem. B*, 2003, **107**, 9535–9551.

[212] S. Boresch, *J. Chem. Inf. Model.*, 2024, **64**, 3605–3609.

[213] H.-J. Woo, B. Roux, *Proc. Natl. Acad. Sci. U.S.A.*, 2005, **102**, 6825–6830.

[214] J. C. Gumbart, B. Roux, C. Chipot, *J. Chem. Theory Comput.*, 2013, **9**, 794–802.

[215] S. Jo, W. Jiang, H. S. Lee, B. Roux, W. Im, *J. Chem. Inf. Model.*, 2013, **53**, 267–277.

[216] G. Heinzelmann, M. K. Gilson, *Sci. Rep.*, 2021, **11**, 1116.

[217] H. Fu, C. Chipot, X. Shao, W. Cai, *J. Phys. Chem. B*, 2023, **127**, 10459–10468.

[218] H. Fu, H. Chen, W. Cai, X. Shao, C. Chipot, *J. Chem. Inf. Model.*, 2021, **61**, 2116–2123.

[219] S. Kim, H. Oshima, H. Zhang, N. R. Kern, S. Re, J. Lee, B. Roux, Y. Sugita, W. Jiang, W. Im, *J. Chem. Theory Comput.*, 2020, **16**, 7207–7218.

[220] R. Liu, W. Li, Y. Yao, Y. Wu, H.-B. Luo, Z. Li, *J. Chem. Inf. Model.*, 2023, **63**, 7755–7767.

[221] E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J. Z. H. Zhang, T. Hou, *Chemical Reviews*, 2019, **119**, 9478–9508.

[222] A. De Ruiter, C. Oostenbrink, *Curr. Opin. Struct. Biol.*, 2020, **61**, 207–212.

[223] T. Hansson, J. Marelius, J. Åqvist, *J. Comput. Aided Mol. Des.*, 1998, **12**, 27–35.

[224] M. K. Gilson, L. E. Stewart, M. J. Potter, S. P. Webb, *J. Chem. Theory Comput.*, 2024, **20**, 6328–6340.

[225] D. L. Mobley, A. P. Graves, J. D. Chodera, A. C. McReynolds, B. K. Shoichet, K. A. Dill, *J. Mol. Biol.*, 2007, **371**, 1118–1134.

[226] D. L. Mobley, J. D. Chodera, K. A. Dill, *J. Chem. Theory Comput.*, 2007, **3**, 1231–1235.

[227] Z. Tang, C.-e. A. Chang, *J. Chem. Theory Comput.*, 2017, **13**, 2230–2244.

[228] H. L. Schenck, K. Hui, *J. Chem. Educ.*, 2011, **88**, 1158–1161.

[229] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. and Modelling*, 2006, **25**, 247–260.

[230] P. Cieplak, W. D. Cornell, C. Bayly, P. A. Kollman, *J. Comput. Chem.*, 1995, **16**, 1357–1377.

[231] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.

[232] G. Bussi, T. Zykova-Timan, M. Parrinello, *J. Chem. Phys.*, 2009, **130**, 074101.

[233] M. Tuckerman, B. J. Berne, G. J. Martyna, *J. Chem. Phys.*, 1992, **97**, 1990–2001.

[234] S. Doudou, N. A. Burton, R. H. Henchman, *J. Chem. Theory Comput.*, 2009, **5**, 909–918.

[235] J. D. Weeks, D. Chandler, H. C. Andersen, *J. Chem. Phys.*, 1971, **54**, 5237–5247.

[236] S. Khuttan, S. Azimi, J. Z. Wu, S. Dick, C. Wu, H. Xu, E. Gallicchio, *Phys. Chem. Chem. Phys.*, 2023, **25**, 24364–24376.

[237] W. Jiang, B. Roux, *J. Chem. Theory Comput.*, 2010, **6**, 2559–2565.

[238] H. Oshima, Y. Sugita, *J. Chem. Inf. Model.*, 2022, **62**, 2846–2856.

[239] Y. Matsunaga, M. Kamiya, H. Oshima, J. Jung, S. Ito, Y. Sugita, *Biophys. Rev.*, 2022, **14**, 1503–1512.

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

[241] R. C. Harris, N. Deng, R. M. Levy, R. Ishizuka, N. Matubayasi, *J. Comput. Chem.*, 2017, **38**, 1198–1208.

[242] E. Schneiderman, A. M. Stalcup, *J. Chromatogr. B Biomed. Appl.*, 2000, **745**, 83–102.

[243] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature*, 2021, **596**, 583–589.

[244] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J. M. Jumper, *Nature*, 2024, **630**, 493–500.

[245] M. Muttenthaler, G. F. King, D. J. Adams, P. F. Alewood, *Nat. Rev. Drug Discov*, 2021, **20**, 309–325.

[246] P. Csermely, R. Palotai, R. Nussinov, *Trends Biochem. Sci.*, 2010, **35**, 539–546.

List of Publications

Main Papers

1. K. Okita, K. Kasahara and N. Matubayasi, "Diffusion theory of molecular liquids in the energy representation and application to solvation dynamics", *J. Chem. Phys.*, **157**, 244505 (2022).

2. K. Okita, N. Ito, N. Morishita-Watanabe, H. Umakoshi, K. Kasahara and N. Matubayasi, "Solvation dynamics on the diffusion timescale elucidated using energy-represented dynamics theory", *Phys. Chem. Chem. Phys.*, **26**, 12852 (2024).

3. K. Okita, Y. Maruyama, K. Kasahara and N. Matubayasi, "Flexible framework of computing binding free energy using the energy representation theory of solution", *J. Chem. Phys.*, in press.

Related papers

1. K. Kasahara, R. Masayama, K. Okita and N. Matubayasi, "Atomistic description of molecular binding processes based on returning probability theory", *J. Chem. Phys.*, **155**, 204503 (2021).

2. K. Kasahara, R. Masayama, K. Okita and N. Matubayasi, "Elucidating protein–ligand binding kinetics based on returning probability theory" *J. Chem. Phys.*, **159**, 134103 (2023).

Acknowledgements

The author is indebted to Prof. Nobuyuki Matubayasi and Asst. Prof. Kento Kasahara for their constant guidance and helpful advice throughout this work. The author is grateful to Assoc. Prof. Kang Kim for valuable discussions.

The author thanks all the members of the Matubayasi laboratory for their friendship. Special thanks to my friends for their support and valuable discussions: Shota Goto, Senri Tanaka and Kengo Takemoto.

This work was supported by the Grant-in-Aid for Scientific Research (Grant Nos. JP21K14589, JP22J21080, JP22KJ2210, JP21H05249, JP23H01924, JP23H02622, JP23K27313, and JP23K26617) from the Japan Society for the Promotion of Science, the Fugaku Supercomputer Project (Grant No. JPMXP1020200308, JPMXP1020230325, JPMXP1020230327, and JPMXP1122714694), the Data-Driven Material Research Project (Grant No. JPMXP1122714694) and the Elements Strategy Initiative for Catalysts and Batteries (Grant No. JPMXP0112101003) from the Ministry of Education, Culture, Sports, Science, and Technology, Maruho Collaborative Project for Theoretical Pharmaceutics, the Core Research for Evolutional Science and Technology (CREST), and Japan Science and Technology Agency (JST) (No. JPMJCR22E3). The simulations were conducted using TSUBAME3.0 at Tokyo Institute of Technology and Fugaku at RIKEN Advanced Institute for Computational Science through the HPCI System Research Project (Project No. hp220254, hp230101, hp230205, hp230212, hp230158, hp240223, hp240224, hp240195, and hp240111).

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