Development of Analysis Method for Solvation Effects on Brownian Motion of Macromolecule Using Molecular Dynamics Studies

岩下, 智哉

Development of Analysis Method for Solvation Effects on Brownian Motion of Macromolecule Using Molecular Dynamics Studies

分子動力学法を用いた

巨大分子のブラウン運動への溶媒効果の解析手法の開発

Tomoya Iwashita

岩下　智哉

Department of Chemistry, Kyushu University, Fukuoka, 812-0395, Japan

February 21, 2025

Abstract

Molecular dynamics studies were carried out to reveal the solvation effects on the macromolecule diffusion. This thesis was motivated by the experimental results obtained by Terazima and coworkers. They have shown that minor changes in the protein structure can change the diffusion coefficient significantly. Their results indicate the diffusion coefficients contain information about the solvation, such as hydrogen bonding. Molecular dynamics simulation would be a powerful tool to reveal the relation between the diffusion coefficient and the solvation. In this thesis, I attempted to develop a method for decomposing the friction coefficients with MD simulations. The decomposed components of the friction coefficient can be used as the indices of the diffusion coefficient since the friction coefficient is inversely proportional to the diffusion coefficient. This approach could be useful in explaining the experimentally observed significant differences in the diffusion coefficients of proteins.

First, I assessed the system-size dependence of the diffusion coefficients. The diffusion coefficient calculated from MD simulation depends on a simulation cell length, i.e., cell size. The dependence is significant when discussing protein diffusions in this subject. Yeh and Hummer have proposed correction methods. In the conventional usage of Yeh and Hummer method, the higher-order correction terms have been almost ignored. I examined the higher-order correction terms for the Yeh–Hummer method. The results show the usefulness of the higher-order correction terms. The correction method with the higher-order correction terms corrects the diffusion coefficient accurately while reducing the computational cost.

Second, I developed the method to decompose the friction coefficients into the contribution of individual components on the solute surface using MD simulations. The method was constructed in the framework of the generalized Langevin equation. To verify the method, MD simulations of $C_{60}(OH)_n$ (fullerenol) in water were carried out. The diffusion coefficients of fullerenol decrease as the number of OH groups increases. The proposed method reveals that the decrease in diffusivities is mainly driven by the roughness of the solute surface and hydrogen bonding. The results support the experimental results found by Terazima et al. that minor changes in protein structure can significantly alter the diffusion coefficient.

Contents

Chapter 1

General introduction

The self-diffusion coefficient shows how quickly Brownian particles diffuse the medium. The coefficient has been obtained using various experimental methods such as Dynamic Light Scattering, Nuclear Magnetic Resonance, Fluorescence Correlation Spectroscopy, and Transient Grating. The diffusion coefficient is well known to depend strongly on the hydrodynamic radius of Brownian particles through the Stokes-Einstein law. Recently, the Terazima group has shown that the hydrodynamic radius not only depends on the particle radius but also on the solvation of the particle experimentally [1, 2, 3, 4, 5, 6, 7, 8, 9]. The solvation effect on the diffusion coefficient (or the hydrodynamic radius) is often significant. For example, in the case of the LOV2-linker protein, the change in hydrodynamic radius was comparable to the volume change associated with dimerization. However, the volume change in the protein, nor dimerization didn't explain the phenomenon. This study aims to clarify the relation between diffusion coefficient and solvation using molecular dynamics (MD) simulation and propose an analysis method for the diffusion phenomena.

First, this chapter describes the historical background and basic knowledge of diffusion coefficient. The experimental results obtained by the Terazima group are emphasized because they show the significant solvation effects on the diffusion coefficient. Second, I will explain what is necessary to solve the problems using the MD simulations. Finally, the purpose of this study is specified, and the approaches are explained.

1.1 Self-diffusion coefficient and Stokes-Einstein (SE) law

If a drop of ink falls into a liquid, the ink molecules spontaneously diffuse through the liquid from areas of high concentration to areas of low concentration. This diffusion phenomenon is described by the Fick Law [10, 11] (see Appendix A):

$$J = -D\frac{dc(x)}{dx}, \tag{1.1}$$

where J is the amount of substance passing through a unit area per unit time, $c(x)$ is a molecular concentration at position x, and D is the diffusion coefficient. Here, J is known as the diffusion flux. The Fick law indicates that J is proportional to the gradient of $c(x)$ and the constant of proportionality D. The diffusion processes are not always the macroscopic non-equilibrium phenomena. In 1905, Einstein attributed the Brownian motion to the diffusion process [12, 13, 14, 15]. He thought that the motion-producing force in Brownian motions originates from the osmotic pressure due to the local concentration gradient.

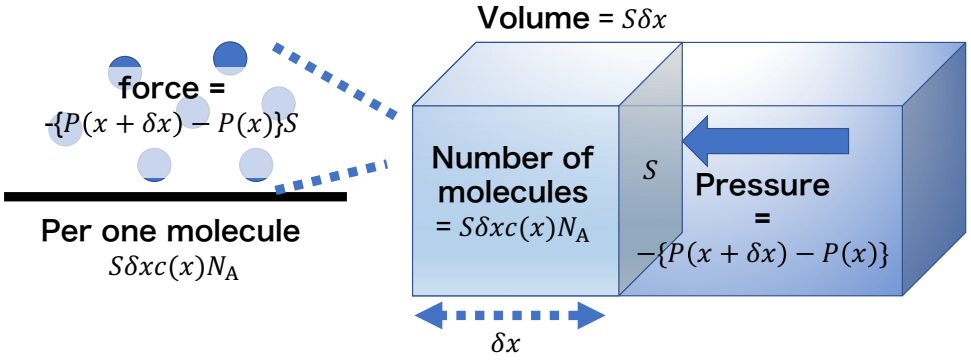

Figure 1.1: Osmotic force acts on a particle. The thermodynamic force is caused by the local concentration gradient.

Here, I describe the summary of Einstein's theory [12]. Even in the system under thermodynamic equilibrium, local concentration differences exist because of the fluctuation. Thus, the local osmotic pressure fluctuates spontaneously. If the solution is sufficiently dilute, the osmotic pressure p obeys the van't Hoff's law: $p = cR_{\rm gas}T$, where $R_{\rm gas}$ is gas constant and T is temperature. Then, as shown in Fig. 1.1, the osmotic force $K_{\rm osmo}$ per solute molecule acted by the concentration gradient is:

$$K_{\rm osmo} = -\frac{\{p(x+\delta x) - p(x)\}\cdot S}{S\delta x c(x) N_{\rm A}} = -\frac{k_{\rm B}T}{c(x)}\frac{c(x+\delta x) - c(x)}{\delta x} = -\frac{k_{\rm B}T}{c(x)}\frac{dc(x)}{dx}, \tag{1.2}$$

where $N_{\rm A}$ is Avogadro constant, S is unit cross section on which the pressure act, $k_{\rm B}$ is Boltzmann constant, and T is temperature. Here, a relation: $k_{\rm B} = R_{\rm gas}/N_{\rm A}$ was used. Einstein also

derived Eq. (1.2) from the perspective of the extreme value condition of free energy [13]. Then, the solute velocity $v(t)$ is expressed by the following equation:

$$m\frac{dv(t)}{dt} = -\gamma v(t) + K_{\mathrm{osmo}}, \tag{1.3}$$

where γ is the friction coefficient, and $-\gamma v(t)$ is the drag force. Assuming high viscosity and zero inertial term, i.e., $m\frac{dv(t)}{dt} = 0$ (see Appendix B), we obtain the following relation about the solute velocity:

$$v(t) = K_{\mathrm{osmo}}/\gamma \tag{1.4}$$

Substituting Eq. (1.2), this equation can be rewritten as:

$$vc(x) = -\frac{k_{\mathrm{B}}T}{\gamma}\frac{dc(x)}{dx}, \tag{1.5}$$

which is identical to the Fick law since $J = vc(x)$. This equation indicates the Brownian motions originate from the diffusion process, and gives the relation about the diffusion coefficient:

$$D = \frac{k_{\mathrm{B}}T}{\gamma}. \tag{1.6}$$

This is one of the fluctuation-dispersion theorems. In 1906, Smoluchowski also derived this equation from another perspective in which the motion-producing force originates from the multiple collisions of molecules [16] (see Fig. 1.2). Thus, this equation is called Einstein–Smoluchowski equation.

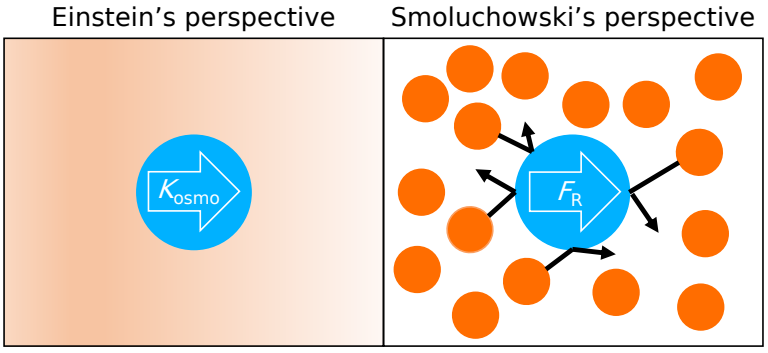

Figure 1.2: The difference in perspective between Einstein and Smoluchowski is shown. The former regards the solvent as a continuum, while the latter regards it as composed of particles.

Moreover, Einstein adopted the Stokes law: $\gamma = 6\pi\eta R_{\mathrm{h}}$, which predicts the friction coeffi-

cient of a sphere in fluid, and derived the Stokes–Einstein (SE) law:

$$D = \frac{k_{\mathrm{B}}T}{6\pi\eta R_{\mathrm{h}}} \tag{1.7}$$

where η is solvent viscosity, and R_{h} is the hydrodynamic radius of the Brown particle. The SE law predicts that the diffusion coefficient depends on the hydrodynamic radius of the Brownian particle.

Einstein also pointed out the mean-squared displacements $\langle \Delta x(t)^2 \rangle = \langle (x(t) - x(0))^2 \rangle$ of the particles in the diffusion process are proportional to the time t as follows:

$$\langle \Delta x(t)^2 \rangle = 2Dt, \tag{1.8}$$

where $\langle \cdots \rangle$ denotes taking an ensemble average [13]. In 1908, based on the Smoluchowski's perspective (see Fig. 1.2), Langevin introduced a mathematical model for Brownian motion:

$$m\frac{dv(t)}{dt} = -\gamma v(t) + F_{\mathrm{R}}, \tag{1.9}$$

where F_{R} is random force caused by the multiple collisions of molecules [17, 18]. By solving this equation, Langevin showed that the mathematical model is consistent with Eqs. (1.6)–(1.8). In this case, the constant of proportionality is also 2D. These arguments show that Brownian motion is a consequence of diffusion processes. The diffusion coefficient in Brownian motion is specially called 'the self-diffusion coefficient'. Therefore, the self-diffusion coefficient is a proportionality constant that indicates how quickly the Brownian motion, which is a kind of diffusion phenomena, proceeds. In this thesis, the term 'diffusion coefficient' primarily refers to the self-diffusion coefficient. Next section describes the experimental aspects of the diffusion coefficient and the SE law.

1.2 Applications of the SE law

Perrin experimentally verified the theory of Brownian motion proposed by Einstein. In 1909, he also determined the value of the Avogadro number using a colloidal dispersion system and the SE law [19, 20]. If one knows the particle radius R in advance and measures the particle displacements, the SE law gives N_{A} as follows:

$$N_{\mathrm{A}} \approx \frac{R_{\mathrm{gas}}T}{6\pi\eta RD}. \tag{1.10}$$

Thus, Perrin and his co-workers prepared colloidal particles of uniform size, and measured the displacements of those particles to determine the Avogadro number. They continued the measurements with various experimental conditions such as particle mass (75000–4.8×10^{-14} g), particle radius (0.212–5.5 μm), and solvent viscosity (1–125 $\times10^{-2}$ dyn·s/cm^2) over four years. They found that the N_A remained although there was a great diversity of the experimental conditions. The values for N_A were 5.5–8.0 $\times$ 10^{23} mol^{-1}, which is approximately the same as 6.0×10^{23} mol^{-1}: the value currently known. This error is smaller than that of a similar experiment using a CCD camera (15–30 %) [21]. Therefore, Perrin's experiments were very precise for his era, and we could recognize that the SE law was reasonable.

After Perrin's experiments, the SE law was widely accepted and applied by many researchers. One of its key applications is particle size analysis. Here, I introduce dynamic light scattering (DLS) as an example of the experimental application[22]. DLS does not track the trajectories of Brownian particles but instead obtains the diffusion coefficients by analyzing light scattered from the particles. When a flash of light hits a particle with a radius much smaller than the wavelength of light, the light is scattered in all directions. The intensity of the scattered light fluctuates due to interference with other scattered light and Brownian motion. The time constant of the intensity's autocorrelation function is inversely proportional to the particle's diffusion coefficient as follows:

$$\frac{\langle I(t)\cdot I(t+\tau)\rangle}{\langle I(t)\rangle^2} = 1 + e^{-2Dq^2t}, \tag{1.11}$$

where $I(t)$ is the intensity of light and q is a scattering vector. The scattering vector is determined by the scattering angle θ and the wavelength of light λ through $4\pi\sin(\theta/2)/\lambda$. Therefore, by constructing a setup shown in Fig. 1.3 and observing the fluctuations in scattered light intensity with a detector, we can obtain the diffusion coefficient of the particles through Eq. (1.11). Once we have a diffusion coefficient of a Brownian particle, we can estimate the radius as the hydrodynamic radius through SE law as follows:

$$R_\mathrm{h} = \frac{k_\mathrm{B}T}{6\pi\eta D}. \tag{1.12}$$

Therefore, we can estimate the "radius" of invisible particles through the SE law and scattering light from the particle.

In 1970, Pike and co-workers developed the first digital autocorrelator and demonstrated that the diffusion coefficient of hemocyanin can be measured accurately using DLS [23]. After that, DLS has evolved through various methodological advancements to become a widely used experimental technique for particle size analysis. In fact, DLS is now introduced by Interna-

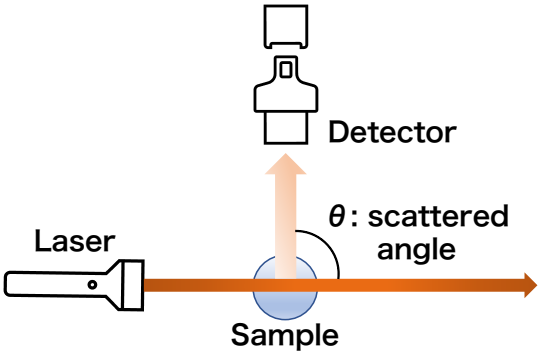

Figure 1.3: An example of a setup for DLS.

tional Standards Organization (ISO) as a method of particle size analysis [24].

The information from diffusion coefficients obtained by experimental techniques, such as DLS, is not always the particle radius. We can estimate the particle shape through the hydrodynamic radius and the gyration radius obtained from static light scattering, SAXS or SANS methods. The gyration radius R_{g} is also a kind of particle radius defined as

$$R_{\mathrm{g}} = \sqrt{\frac{\int_V |\mathbf{r} - \mathbf{r}_{\mathrm{cm}}|^2 \rho dV}{\int_V \rho dV}}, \tag{1.13}$$

where $\mathbf{r}_{\mathrm{cm}}$ is the center of mass, ρ is the mass density, and $\int_V$ denotes volume integral. Here, we assume a sphere with homogeneous mass density and the radius R, the gyration radius is expressed as follows:

$$R_{\mathrm{g}} = \sqrt{\frac{\int_0^R \rho r^2 4\pi r^2 dr}{4/3\pi\rho R^3}} = \sqrt{\frac{3}{5}} R. \tag{1.14}$$

Then, assuming that the hydrodynamic radius is equal to the sphere radius, the following relationship holds for the ratio of the gyration radius to the hydrodynamic radius for spherical molecules:

$$\frac{R_{\mathrm{g}}}{R_{\mathrm{h}}} \approx \sqrt{\frac{3}{5}} \tag{1.15}$$

Predicting the molecular shape of diffuse species using the ratio $R_{\mathrm{g}}/R_{\mathrm{h}}$ was initiated by W. Burchard et al. to help distinguish polymer shapes [25]. In the review by J. Stetefeld et al. [22], several results for the $R_{\mathrm{g}}/R_{\mathrm{h}}$ values of proteins and RNAs are tabulated, showing that $R_{\mathrm{g}}/R_{\mathrm{h}}$ features molecular shape. In the table, the ratio for globular molecules approximates $\sqrt{3/5}$, while the ratio for elongated molecules is larger than $\sqrt{3/5}$. These results could be evidence that the SE law captures the particle radius precisely in those DLS experiments.

1.3 Problems in the application

In the previous section, I explained the SE law as the prediction method for the Avogadro number, the size, and the shape of particles using diffusion coefficients obtained experimentally. Thus, the SE law has been a helpful tool for analyzing the physical properties of microscopic particles for over 100 years. However, some problems remain in the usage of the SE law. One of these is the "radius": the hydrodynamic radius is not always equal to the particle radius. The contribution of solvation to diffusion is not always small. For example, we don't know how the solute-solvent interactions, such as hydrogen bonds, affect the diffusion coefficients. Especially, the importance of solvation on the diffusion coefficient was emphasized by the experimental results conducted by Terazima and co-workers.

Terazima and co-workers used a transient grating (TG) method to obtain the diffusion coefficients. I describe a brief introduction to the TG method[26]. When two laser lights (wave 1 and wave 2 as shown in Fig. 1.4 (a) and (b)) in different directions are crossed, the interference light (wave 1 + wave 2 as shown in Fig. 1.4 (c)) makes an interference pattern as shown in Fig.1.4 (d). The TG method uses this phenomenon with two pulse lasers to make an interference pattern in the sample solution. In a bright spot of the interference pattern, chemical reactions are initiated, and refractive index change occurs with concentration modulation of chemical species. This grating pattern of refractive indices functions as a diffraction grating. Therefore, when another probe light is illuminated to the sample solution, we obtain a diffraction light called a TG signal. A setup is shown in Fig. 1.5. Assuming no chemical reaction in the observation time window and no time dependence of molecular diffusion coefficients, the change of the TG signal is expressed as:

$$I_{\mathrm{TG}} = \alpha\{\delta n_{\mathrm{th}} \exp(-D_{\mathrm{th}} q^2 t) + \delta n_{\mathrm{P}} \exp(-D_{\mathrm{P}} q^2 t) - \delta n_{\mathrm{R}} \exp(-D_{\mathrm{R}} q^2 t)\}^2, \tag{1.16}$$

where α, δn_{th}, D_{th}, δn_{P}, D_{P}, δn_{R}, and D_{R} are constant, the initial refractive index change by the temperature change, the thermal diffusivity of the solution, the initial refractive index change by the change of the product concentration, the diffusion coefficient of the product, the initial refractive index change by the reactant concentration, and the diffusion coefficient of the reactant, respectively. Therefore, we can obtain the diffusion coefficients of the reactant and products by analyzing the TG signal. Because the TG method simultaneously gives those diffusion coefficients in the same condition, we can evaluate the difference of diffusion coefficients between the reactant and product precisely. It is an advantage of the TG method over other methods, such as DLS.

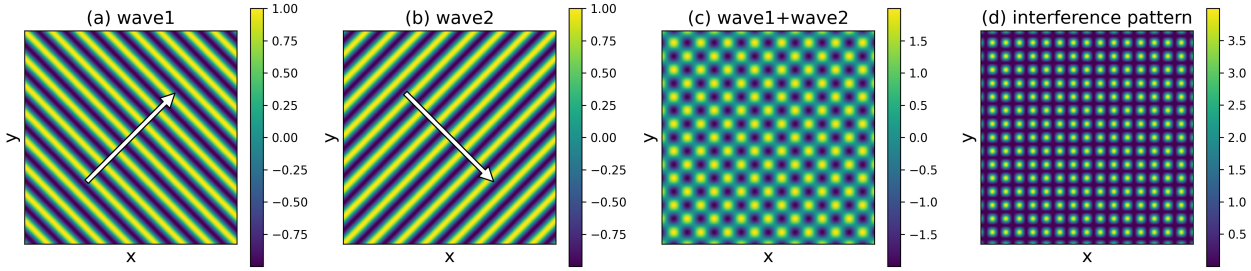

Figure 1.4: Interference pattern of two laser light. (a) and (d) show two plane waves from different directions. (b) shows the superposition of two waves. (d) shows intensity, i.e., squared amplitude, of the wave shown in (b). In (a), (b), and (d), color-bars show the amplitude of the wave, while in (d), the color-bar shows the squared amplitude of the wave.

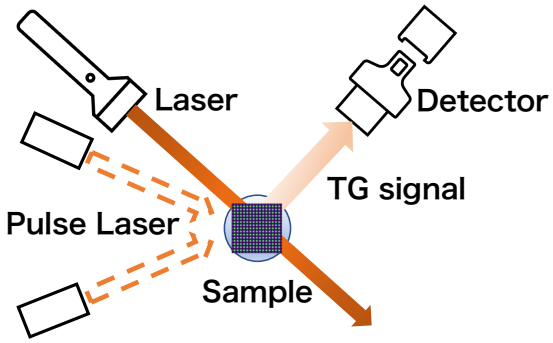

Figure 1.5: An example of a setup for the TG method.

Using the TG method, Terazima and co-workers investigated the thermodynamic and transport properties of intermediate states of photoactive yellow protein (PYP) during the photocyclic reaction [1]. In the study, they found that the diffusion coefficient of the ground state is about 1.2 times larger than that of the second intermediate state. If we assume this difference has been caused only by the volume difference, the volume ratio would be 1.7 times. Such a significant volume change would lead to a significant change in the refractive index. However, according to their articles [1], no such change in the refractive index was observed in the TG profile. Therefore, they supposed that the difference is caused by the changes in the protein–water interaction. In fact, before that, Oishi et al. have suggested that the surface area of the second intermediate state becomes larger than that of the ground state in a denaturation experiment of PYP [27], which is consistent with the interpretation of Terazima and co-workers. Thus, Terazima and co-workers concluded that the surface of PYP in the second intermediate state is rougher than that in the ground state.

Terazima and co-workers have observed large changes in the diffusion coefficient due to structural changes in a protein. They have examined various proteins, such as carboxymyo-

globin (MbCO) [2] and cytochrome c (Cyt-c) [3, 4]. To study the origin of the diffusivity changes of Cyt-c, they investigated the diffusion coefficient of poly-L-glutamic acid (PLG) at various pHs [5]. PLG forms the random-coil conformation at $6 < \mathrm{pH} < 7$, while PLG begins to form α-helix formation at pH lower than 5.3. Using the nature of PLG, they aimed to explore the relationship between conformation and diffusion coefficient D, while also attempting to confirm the hypothesis that the change in the diffusion coefficient of Cyt-c originates from the collapse of the α-helix. The results show that the diffusion coefficient of α-helix is about 1.3 times larger than that of random coil. Although the structure of random coil is unknown, assuming the structure of the short α-helix composed of about 10 amino acids is a spherical shape, they pointed out the structure of random coil should also be a spherical shape from the discussion with the theoretical equation proposed by Perrin:

$$\gamma/\gamma_0 = \frac{(a/b)^{2/3}(1-b^2/a^2)^{1/2}}{\ln[\{1+(1-b^2/a^2)\}^{1/2}/(b/a)]}, \tag{1.17}$$

where γ and γ_0 are a friction coefficients of a prolate molecule with a semimajor axis a and a minor axis b and a spherical molecule with the same volume as the prolate molecule. Here, we expected that the ratio (a/b) should be 7.45 to achieve the observed ratio $\gamma/\gamma_0 = 1.3$. However, such an extremely elongated shape is obviously unrealistic because a freely moving chain should form a near-globular shape with high entropy. Thus, they attributed the diffusivity changes to the change in the hydrogen-bond network from the intramolecular to the intermolecular one. This suggestion also indicates that the change in the diffusion coefficient of Cyt-c is caused by the hydrogen-bond network due to the unfolding of the α-helix. In fact, subsequent studies revealed that the origin of the diffusion coefficient of PYP is also conformation changes at N-terminal helices [28] and that other proteins also show large diffusivity changes for a similar reason. I summarized the large diffusivity changes observed by Terazima and co-workers in Table 1.1. Those diffusivity changes are not accompanied by dimerization and large volume change. Especially, for example of the LOV2-linker, although the gyration radius of the ground state (2.43nm) is almost the same as that of the intermediate state (2.50nm), the diffusion coefficient in the ground state is 1.2 times larger than that in the intermediate state [29, 6]. Those results cannot be explained by SE law with the assumption "hydrodynamic radius = particle radius", and it indicates that the diffusion coefficient can strongly depend on the solvation.

Table 1.1: Summary of large changes in the diffusion coefficient due to structural changes of protein in aqueous solution. The diffusion coefficients D of two states (D1 and D2) are listed for each protein. The unit of D is 10^{-10} m^2/s.

protein	D1	D2	reference
PYP	1.21 (grand state)	1.00 (second intermediate state)	[1]
MbCO	1.29 (native)	0.32 (unfolded)	[2]
Cyt-c	1.18 (native)	0.65 (unfolded)	[3, 4]
PLG	3.00 (α-helix)	2.39 (random-coil)	[5]
Photo2-LOV2-linker	0.88 (grand state)	0.68 (intermediate state)	[6]
Photo1-LOV2-linker	0.92 (grand state)	0.50 (intermediate state)	[7]
Rhodopsin	0.6 (grand state)	0.28 (final product)	[8]
Heliorhodopsin	0.54 (grand state)	0.45 (intermediate state)	[9]

1.4 Solvation effects in the diffusion phenomena

Terazima and co-workers show the importance of solvation in the diffusion phenomena. Then, where is room for solvation effects in the SE law? One might think "How far is SE law correct in the first place?". First, we discuss the Einstein–Smoluchowski (ES) equation since the SE law is derived from it. Second, I show the necessity of the friction coefficient to discuss the solvation effects on the SE law.

In 1965, Mori showed that the generalized form of Langevin equation (Eq. (1.9)), so-called generalized Langevin equation (GLE), can be derived directly from Newton's equations using a projection operator method [30]. The one-dimensional GLE for the Brownian particle (solute) is

$$F(t) = -\int_0^t d\tau \Gamma(\tau) v(\tau - t) + F_{\mathrm{R}}(t), \tag{1.18}$$

where F, Γ, v, and F_{R} are total force acting on solute, friction kernel of solute, velocity of the mass center relative to the solvent, and a random force acting on solute, respectively. Here, the convolution integral represents the drag force, which takes into account the memory effect. I prepared the re-derivation of GLE for the Brownian motion in Appendix C [30, 31]. From Appendix C, we can see the followings:

1. GLE for Brownian motion can be derived from Newton's equations with no approximations.
2. A term $i\omega A$, which appeared in Mori's paper, is not necessary for Brownian motion.
3. Langevin equation (eq. (1.9)) can be derived from GLE with Markov approximations.
4. In GLE, the friction coefficient is defined as follows:

$$\gamma = \int_0^\infty d\tau \Gamma(\tau). \tag{1.19}$$

5. The random force is uncorrelated with the initial velocity, i.e.,

$$\langle v(0) \cdot R(\tau) \rangle = 0. \tag{1.20}$$

6. The random force is related to the friction kernel by the fluctuation-dissipation theorem [32]:

$$\langle F_{\mathrm{R}}(0) \cdot F_{\mathrm{R}}(t) \rangle = k_{\mathrm{B}} T \Gamma(t). \tag{1.21}$$

Now, let's rederive the ES equation from GLE. Multiplying both sides of GLE by the initial velocity $v(0)$ and taking an ensemble average, we obtain the following equation:

$$m \frac{dC^{vv}(t)}{dt} = -\int_0^t d\tau \Gamma(\tau) C^{vv}(\tau - t), \tag{1.22}$$

where $C^{vv}(t) = \langle v(0) \cdot v(t) \rangle$. We define the Laplace transforms of $C^{vv}(t)$ and $\Gamma(t)$ as $\mathcal{L}[C^{vv}] = \int_0^\infty C^{vv}(t) e^{-st} dt$ and $\mathcal{L}(\Gamma) = \int_0^\infty \Gamma(t) e^{-st} dt$, respectively. Then, the Laplace transform of eq. (1.22) can be expressed as follows:

$$m(s\mathcal{L}(C^{vv}) - C^{vv}(0)) = -\mathcal{L}(C^{vv})\mathcal{L}(\Gamma), \tag{1.23}$$

Hence,

$$\mathcal{L}(C^{vv}) = \frac{mC^{vv}(0)}{m \cdot s + \mathcal{L}(\Gamma)}. \tag{1.24}$$

The derivation of this equation can also be seen in a book of Evans and Morriss [See Eq. (4.8) of p. 80 in Ref. [31]]. Moreover, when $s \to 0$, this equation goes to:

$$\int_0^\infty C^{vv}(t) dt = \frac{mC^{vv}(0)}{\int_0^\infty \Gamma(t) dt} = \frac{k_{\mathrm{B}} T}{\gamma}. \tag{1.25}$$

Then, using the Green–Kubo relation:

$$\frac{d\langle \Delta x(t)^2 \rangle}{dt} = 2 \int_0^\infty C^{vv}(t) dt, \tag{1.26}$$

and eqs. (1.8) and (1.25), we obtain the ES equation, i.e., $D = k_{\mathrm{B}} T / \gamma$. Therefore, the ES equation holds even in the GLE, which means that the ES equation is valid in classical

mechanics. The remaining step for derivation of SE law is the application of Stokes law, i.e.,

$$\gamma = \int_0^\infty d\tau \Gamma(\tau) \approx 6\pi\eta R_{\mathrm{h}}. \tag{1.27}$$

In other words, there is room for solvation effects only here. We have to consider the friction coefficient more deeply to reveal the relationship of the diffusion phenomena and solvation.

First, we focus on the Stokes law, i.e., $\gamma = 6\pi\eta R$. Stokes law assumes the spherical shape of the particle and stick boundary conditions, which impose the fluid velocity on the particle surface to zero. The boundary condition might correspond to the particle–fluid interaction. In fact, assuming slip condition, which allows the fluid velocity on the particle surface to be non-zero, gives Stokes law of $\gamma = 4\pi\eta R$. Considering the difference of the boundary condition, the SE law is expressed as follows:

$$D = \frac{k_{\mathrm{B}}T}{c_{\mathrm{surf}}\pi\eta R_{\mathrm{h}}}, \tag{1.28}$$

where c_{surf} is constant determined by the boundary condition.

Using a hard sphere mixture model where a solute with radius R dissolved in a mixture of solvent 1 with radius a and co-solvent 2 with radius b, Nakamura et al. studied theoretically the relationship between the constant c_{surf} and the radial distribution functions between the solute particle and solvent particles [33, 34, 35, 36, 37]. They showed c_{surf} is given by the following equation:

$$c_{\mathrm{surf}} = 6\left\{\frac{2-\epsilon\alpha+\epsilon\beta}{3+(\epsilon\alpha)^2+\epsilon\beta}\right\}, \tag{1.29}$$

with

$$\epsilon\alpha = \frac{2}{R+a}\int_{R+a}^\infty \left\{\frac{W_{\mathrm{T}}(r)}{W_{\mathrm{T}}(\infty)} - 1\right\} dr, \tag{1.30}$$

and

$$\begin{aligned}\epsilon\beta = \frac{1}{R+a}\Bigg\{ &\int_{R+a}^{R+b}\Delta\nu_1(r)dr + \int_{R+b}^\infty \Delta\nu_{\mathrm{T}}(r)dr \\ &- \left(1+\frac{\xi+\eta/3}{\eta}\right)\left[\int_{R+a}^{R+b} W_1(r)\int_r^{R+b}\frac{\omega_1(r')}{W_1(r')}\Delta\nu_1(r')dr'dr + \int_{R+b}^\infty W_{\mathrm{T}}(r)\int_r^\infty \frac{\omega_{\mathrm{T}}}{W_{\mathrm{T}}}\Delta\nu_{\mathrm{T}}(r')dr'dr\right] \\ &- \left(1+\frac{\xi+\eta/3}{\eta}\right)\int_{R+a}^{R+b} W_1(r)dr\left[\int_{R+b}^\infty \frac{\omega_{\mathrm{T}}(r)}{W_{\mathrm{T}}(r)}\Delta\nu_{\mathrm{T}}(r)dr + \Delta\nu_1(R+b)\left(\frac{1}{W_1(R+b)} - \frac{1}{W_{\mathrm{T}}(R+b)}\right)\right]\Bigg\},\end{aligned} \tag{1.31}$$

where,

$$\Delta\nu_j(r) = \frac{2\omega_j(r)}{W_j(r)}\int_{R+a}^r W_j(r')dr', \tag{1.32}$$

$$\omega_j(r) = \frac{1}{W_j(r)}\frac{dW_j(r)}{dr}, \tag{1.33}$$

for j = 1, 2, or T(1 + 2). Here, ξ is bulk viscosity and $W_j(r)$ is an equilibrium mass density, which can be calculated from the radial distribution function $g_i(r)$ between solute and i-th solvent through $W_j(r) = \rho_i g_i(r)$.

Nakamura et al. calculated the $g_i(r)$ using the integral equation theory and calculated the diffusion coefficient through Eqs. (1.28) and (1.29). They found the difference of the mole fraction and packing fraction of co-solvent changes the diffusion coefficients as well as the boundary conditions. Their results indicate the boundary condition takes into account the difference in solute–solvent interaction.

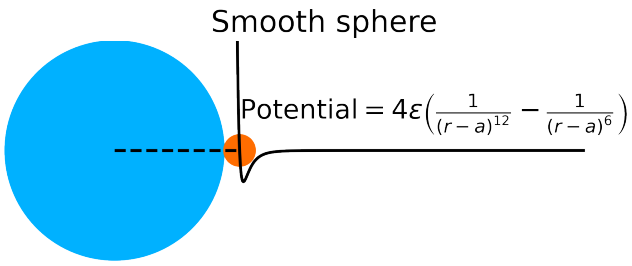

Figure 1.6: Image of smooth sphere model.

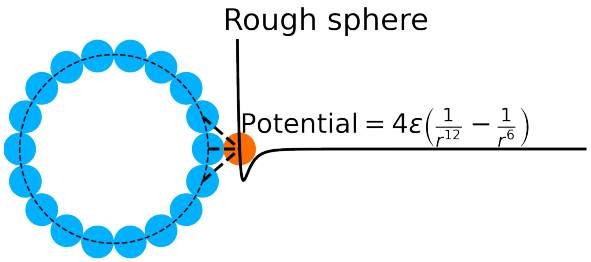

Figure 1.7: Image of rough sphere model.

However, the changes in experiments of Terazima and co-workers would not be simply explained by the transition of slip to stick boundary conditions. Schmidt and Skinner conducted MD simulations to study the problem of boundary conditions [38, 39]. They investigated mainly two different solute models in LJ liquid where the solvent–solvent interaction $P_{\mathrm{solv-solv}}(r)$ is

given by Lennard-Jones (LJ) potential:

$$P_{\mathrm{solv-solv}}(r) = 4\epsilon\left\{\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right\}. \tag{1.34}$$

One solute type is a smooth sphere model, in which the solute–solvent interaction $P_{\mathrm{sol-solv}}(r)$ is given by modified LJ potential:

$$P_{\mathrm{sol-solv}}(r) = 4\epsilon\left\{\left(\frac{\sigma}{r-a}\right)^{12} - \left(\frac{\sigma}{r-a}\right)^{6}\right\}, \tag{1.35}$$

where r is the distance between centers of solute and solvent, a is a parameter to decide the solute size. This solute type can vary the attraction between the solute and solvent by changing the parameter ϵ, in which the solute–solvent interaction $P_{\mathrm{sol-solv}}(r)$ is corrected by a switch function:

$$P_{\mathrm{sol-solv}}(r) = 4\epsilon\left\{\left(\frac{\sigma}{r-a}\right)^{12} - \left(\frac{\sigma}{r-a}\right)^{6}\right\}S(r-a), \tag{1.36}$$

where $S(r-a)$ is 1 for $r < r_1$, $(r_2^2 - r^2)^2(r_2^2 + 2r^2 - 3r_1^2)/(r_2^2 - r_1^2)^2$ for $r_1 < r < r_2$, and zero for $r > r_2$. They take $r_1 = 2^{1/6}\sigma$ and $r_2 = 3\sigma/2$. The second solute type is a rough sphere model, in which the solute surface is composed by LJ particle, which is the same as the solvent. Therefore, in the rough sphere model, the interaction between the solute surface and solvent is the same as LJ potential. For the same parameter ϵ, the two solute models exhibit almost the same attraction between the solute and solvent. The schematic diagrams of the smooth sphere model and the rough sphere model are shown in Figs. 1.6 and 1.7, respectively.

If one has a diffusion coefficient D and the radius R, we can estimate the boundary condition as follows:

$$c_{\mathrm{surf}} = \frac{k_{\mathrm{B}}T}{D\pi\eta R_{\mathrm{h}}}, \tag{1.37}$$

Using the equation and estimated radii, Schmidt and Skinner showed the smooth sphere model exhibits the slip boundary condition at first, and as the attraction becomes larger, the boundary condition changes to the stick boundary condition. On the other hand, from the start, the rough sphere model exhibited the stick boundary condition. This means that the stick boundary condition is achieved with basic Van der Waals interaction and rough surface. Therefore, in a system of protein and water, the stick boundary condition has been achieved by the Van der Waals interaction, and other solvent effects, such as hydrogen bonding, would be additive, i.e., $6\pi\eta R_{\mathrm{h}} + \alpha$.

1.5 Next study of the solvation effect

In the previous section, we saw that the friction coefficient is important in revealing the relationship between the diffusion phenomena and solvation. Nakamura et al. conducted the theoretical study using the hard sphere mixture model. They showed explicitly that the diffusion coefficient of a solute particle depends on the radial distribution function between solute and solvent particles. Their results could be evidence of the relation between the diffusion coefficient and solvation. However, their studies are not enough to discuss the experiments in detail because protein surfaces are more complicated. In fact, as shown by Schmidt and Skinner, the boundary condition of the solute surface could be changed drastically by the addition of a simple rough surface.

Moreover, protein surfaces are a mixture of hydrophobic and hydrophilic surfaces. A decomposition method, which divides the friction or diffusion coefficients into the contributions from these surfaces, must help to understand the solvation effects on protein diffusion. We can expect the following: MD simulation technique gives adequate data for the decomposition method. Unfortunately, there is no specific method to decompose the friction coefficient of a poly-atomic molecule into several contributions. Thus, I attempt to develop such a method with MD simulations in this thesis.

1.6 Artifact in MD simulations

Before the summary of this chapter, I have to introduce a non-negligible artifact that arises when dealing with diffusion phenomena in MD simulations.

MD simulation cannot handle the "real" bulk system, such as a system with 6.0×10^{23} molecules, due to the large computational costs. Thus, MD simulation usually uses a periodic boundary condition (PBC) to solve the problem. In PBC, particles that exit the system re-enter through the corresponding opposite surfaces, which assume a system shown in Fig. 1.8. While in the infinite system, a flow around the swimming sphere spreads far and wide (See Fig. 1.9(a)), the flow is artificially re-entered due to the PBC. This causes a reduction in the diffusivity of particles. Therefore, we have to correct the diffusion coefficient obtained from MD simulation to estimate the diffusion coefficient of the infinite system which can be applied by the SE law.

The correction method was proposed by Fushiki [40] and developed by Yeh and Hummer [41]. Fushiki showed the diffusion coefficient depends on the system cell size as follows:

$$D_{\mathrm{MD}} = D_{\infty} - \alpha \frac{1}{L}, \tag{1.38}$$

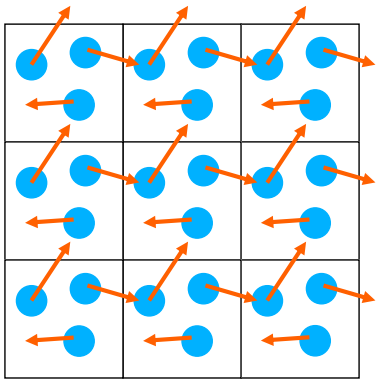

Figure 1.8: An example of periodic boundary condition (PBC).

where D_{MD}, D_{∞}, α, and L are the diffusion coefficient obtained from MD simulation with PBC, the diffusion coefficient of the infinite system, a constant that is independent of the system size, and the cell size of the system. Later, Yeh and Hummer explicitly showed that $\alpha = k_{\mathrm{B}}T\xi/(6\pi\eta)$, where ξ is 2.837297 for a cubic system, based on the Kirkwood–Riseman theory of polymer diffusion. Then, we get the correction equation as follows:

$$D_{\mathrm{MD}} = D_{\infty} - \frac{k_{\mathrm{B}}T\xi}{6\pi\eta L}. \tag{1.39}$$

Yeh and Hummer also demonstrated that this equation functions properly even in a small system composed of 27 LJ particles. After that, the correction method has been used by many researchers. In this thesis, the method will be called Yeh–Hummer method.

However, based on Hashimoto's solution for the Stokes equation under periodic boundary conditions [42], Yeh and Hummer mentioned the higher-order correction as expressed as follows:

$$D_{\mathrm{MD}} = D_{\infty} - \frac{k_{\mathrm{B}}\,T\xi}{6\pi\eta L} + \frac{2k_{\mathrm{B}}\,TR_{\mathrm{h}}^2}{9\eta L^3}. \tag{1.40}$$

The third term was relatively small for $R_{\mathrm{h}} < L/2$, and has been ignored. Here, let's estimate the error of the third term. Eq. (1.40) can be written as follows:

$$D_{\mathrm{MD}} = D_{\infty} - \frac{k_{\mathrm{B}}\,T\xi}{6\pi\eta L}\left(1 - \frac{4\pi R_{\mathrm{h}}^2}{3\xi L^2}\right), \tag{1.41}$$

where the term $\frac{4\pi R^2}{3\xi L^2}$ is the percentage of the third term in the correction term. Moreover, assuming the SE law, eq. (1.41) can be written as follows:

$$D_{\mathrm{MD}} = D_{\infty}\left\{1 - \frac{R_{\mathrm{h}}}{L}\left(\xi - \frac{4\pi R_{\mathrm{h}}^2}{3L^2}\right)\right\}, \tag{1.42}$$

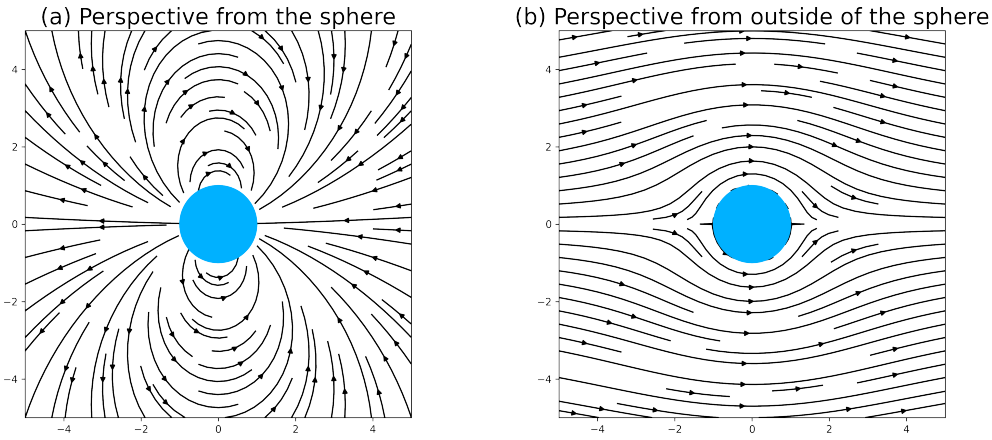

Figure 1.9: The flow around the sphere where the circle indicates the sphere and the arrows indicate the velocity field around the sphere. One might imagine the flow shown in (b), but it assumes a finite flow U_∞ at infinity. Thus, we need to take off U_∞ from the velocity field shown in (b) to consider the flow around the swimming sphere with velocity U_∞. Then, we get the velocity field shown in (a).

where the term $\frac{R_\mathrm{h}}{L}\left(\xi - \frac{4\pi R_\mathrm{h}^2}{3L^2}\right)$ is the percentage of the third term in D_∞. Therefore, if we want to let the error of the third term, i.e., $\frac{R_\mathrm{h}}{L}\left(\xi - \frac{4\pi R_\mathrm{h}^2}{3L^2}\right)$, be smaller than 1%, we have to prepare the cell size of larger than 7.4 R_h. This condition can be satisfied easily with a system of small particles, such as LJ fluid and water. In the case of large solute molecules, conducting an MD simulation that satisfies the condition becomes very computationally expensive. For example, assuming the LOV2-linker protein, the R_h is about 3nm. Then, we must prepare a basic cell for which it is at least 22nm. If we assume that the basic cell has one LOV2-linker protein, the number of water molecules is expected to be more than 3×10^5. The calculation cost for D_∞ can be more expensive as the size of the solute molecule increases. Then, we need to use Eq. (1.42). Unfortunately, Eq. (1.42) has been ignored in other studies and has not been adequately validated. Therefore, we need to verify Eq. (1.42) before using it.

1.7 Summary and purpose of this Study

The SE law shows the particle size dependence on the diffusion coefficient. It has been used as a prediction tool for particle size for over 100 years because other parameters can be obtained using macroscopic measurements. However, Terazima and co-workers showed experimental results that cannot be explained by the SE law with the assumption 'hydrodynamic radius = particle radius,' and emphasized the importance of solvation effects on the diffusion coeffi-

cients [1, 2, 3, 4, 5, 6, 7, 8, 9].

The friction coefficient, which is proportional to the inverse of the diffusion coefficient, must account for the solvation effects. Studies of the friction coefficient for various hard-sphere systems have been conducted by Nakamura et al. [33, 34, 35, 36, 37]. They examined various binary hard-sphere fluids. Their theoretical studies suggest that the solvation effects drastically change the diffusion coefficient of macromolecules immersed in hard-sphere fluids. However, it is not enough to discuss the experimental results in detail because they only examined spherical macroparticles. On the other hand, proteins generally have more complex surfaces, which are mosaic patterns consisting of hydrophobic and hydrophilic surfaces. Therefore, to understand the solvation effects on protein diffusion, a method to decompose the contribution of each surface to the friction or diffusion coefficient is required.

The main purpose of this study is to construct a method for the decomposition of the friction coefficients to analyze the hydration effect on a protein. The analysis needs a detailed model (or all-atom model) and extensive calculations obtained from the MD simulations. However, I also pointed out the problem in the MD simulation in the previous section [40, 41]. Thus, this thesis comprises two main chapters. In the first part, I assess the higher-order correction term of the Yeh–Hummer method. Here, the usefulness of the higher-order term when handling protein diffusion coefficients in the solvent will be shown. In the second part, I introduce the method to decompose the contribution of each surface to the friction or diffusion coefficient. In this part, I define the contribution of each site in a molecule to the friction coefficient. I apply the theoretical method to systems that have a fullerenol immersed in ambient water. The fullerenol molecule is adopted as a model molecule to mimic a protein because it has hydrophobic and hydrophilic surfaces. The OH groups also contribute to the surface roughness. Therefore, fullerenol molecules are adequate as a model protein.

Chapter 2

Usefulness of higher-order system-size correction for calculations of macromolecule diffusion coefficients [43]

2.1 Introduction

Usually, textbooks for molecular simulation describe calculation methods for the diffusion coefficient using MD trajectories [see p60 in Ref. [44]]. These methods are adequate for obtaining the coefficient in a finite system under periodic boundary conditions (PBC). However, the coefficient for an infinite system cannot be obtained because the system size dependence is significant. In other words, the PBC causes an artifact when attempting to obtain the diffusion coefficient in an infinite system. Researchers did not initially recognize this artifact, and as a result, many diffusion coefficients were published under the PBC without any correction. Recognition of this artifact has gradually increased over the last twenty years.

Fushiki pointed out the artifact and proposed a method to obtain the diffusion coefficient for a finite system in 2003 [40], and Yeh and Hummer later explicitly derived the correction equation in 2004 [41]. The method proposed by Yeh et al. is now widely used by many researchers. Moreover, it should be noted that the method ignores the higher-order correction term. As mentioned above, the correction becomes increasingly important as the solute molecule size increases, particularly in the calculation of the diffusion coefficient of a protein (see Section 1.6).

In this chapter, I present a method to obtain an infinite-dilution diffusion coefficient for a

macromolecule, D_∞, using data from MD simulations of small systems. Notably, the ratio of protein molecules to solvent molecules in these small systems is significantly higher than that in an infinitely dilute system. In other words, the concentration in the simulated systems is much higher than that in an infinitely dilute system. However, a new strategy for determining the diffusion coefficient D_∞ can be proposed based on the higher-order correction term.

First, I will explain the derivation of the correction method proposed by Yeh and Hummer and discuss the origin of the higher-order correction term. Second, I will apply the correction methods to the results of MD simulations. Then, I will demonstrate the significant deviations that arise from ignoring the higher-order correction term when using small simulation cells with a protein as the solute. Next, I will introduce a small scheme for the estimation method based on Eq. (1.40), which does not require the hydrodynamic radius. Although Eq. (1.40) was applied in a previous study [45], the hydrodynamic radius R_h was determined through different methods. The differences between the present and previous studies will be highlighted. The current approach reduces the computational cost from 'enormous' to 'acceptable.'

2.2 Theory for system size dependence of the diffusion coefficient

MD simulations provide time-dependent data on particle positions, velocities, and forces (I prepared the description of MD simulation in Appendix D). Using this data, we can calculate the diffusion properties of particles through specific relationships, such as Eq. (1.8). However, MD simulations for bulk fluids use periodic boundary conditions (PBC), where the target cell is surrounded by its replica cells. When the basic cell has one target particles, the concentration is one over the basic cell volume. Then, the concentration of the target particle is high if the basic cell size is usual. In these cases, we cannot adopt the SE law because the SE law requires a dilution limit. Fortunately, the effect can be estimated because of the simple periodicity. Thus, we can obtain the diffusion coefficient for the infinite system as a corrected value. Correction methods have been proposed by Fushiki [40], as well as by Yeh and Hummer [41]. In this section, I will first describe the derivation of the correction method proposed by Yeh and Hummer, and later discuss the higher-order correction terms for the Yeh–Hummer method. Then, I also discuss the origin of the higher-order term.

2.2.1 Yeh–Hummer method [41]

Assuming a finite periodic system in which a perturbation force $\mathbf{F}$ acts on a target particle positioned at the center of a cubic cell filled with an incompressible fluid, the velocity field

caused by the perturbation force $\mathbf{v}(\mathbf{r})$ is described by the Stokes equation as follows:

$$\eta\nabla^2\mathbf{v}(\mathbf{r}) = \nabla p(\mathbf{r}) - \left[\delta(\mathbf{r}) - \frac{1}{V}\right]\mathbf{F}, \tag{2.1}$$

where η is the dynamic viscosity of the fluid, $\nabla = \partial/\partial\mathbf{r}$ is the gradient operator, $p(\mathbf{r})$ represents the pressure at position $\mathbf{r}$, $\delta(\mathbf{r})$ is the Dirac delta function centered at the target particle, and V is the volume of the cubic cell. Here, $1/V$ represents the force $\mathbf{F}$ over the entire volume V of the cubic cell, which ensures that the net force is zero in the cell.

Taking the divergence of both sides of the Stokes equation results in the following equation:

$$\nabla^2 p(\mathbf{r}) = -\mathbf{F}\cdot\nabla\left[\left(\delta(\mathbf{r}) - \frac{1}{V}\right)\right], \tag{2.2}$$

where $\nabla\cdot\nabla^2\mathbf{v}(\mathbf{r}) = 0$ because $\nabla\cdot\mathbf{v}(\mathbf{r}) = 0$.

In Fourier space, Eqs. (2.1) and (2.2) transform into

$$\eta k^2\tilde{\mathbf{v}}_k = i\mathbf{k}\tilde{p}_k + (1-\delta_k)\mathbf{F}, \tag{2.3}$$

and

$$k^2\tilde{p}_k = i(1-\delta_k)\mathbf{k}\cdot\mathbf{F}, \tag{2.4}$$

where $\mathbf{k} = 2\pi L^{-1}(l_x, l_y, l_z)$, $\tilde{\mathbf{v}}_k = \int_V \mathbf{v}(\mathbf{r})\exp(i\mathbf{k}\cdot\mathbf{r})\,d\mathbf{r}$, $\tilde{p}_k = \int_V p(\mathbf{r})\exp(i\mathbf{k}\cdot\mathbf{r})\,d\mathbf{r}$, and $\delta_k = 1$ if $\mathbf{k} = 0$; otherwise, $\delta_k = 0$.

Substituting the expression for $\tilde{p}_k$ from Eq. (2.4) into Eq. (2.3) gives

$$\tilde{\mathbf{v}}_k = \frac{1-\delta_k}{\eta k^2}\left(\mathbf{1} - \frac{\mathbf{kk}}{k^2}\right)\cdot\mathbf{F}, \tag{2.5}$$

where $\mathbf{1}$ is the 3×3 unit matrix. The inverse Fourier transform of $\tilde{\mathbf{v}}_k$ gives

$$\mathbf{v}(\mathbf{r}) = \mathbf{T}_{\mathrm{pbc}}(\mathbf{r})\cdot\mathbf{F}, \tag{2.6}$$

where $\mathbf{T}_{\mathrm{pbc}}(\mathbf{r}) = \sum_{\mathbf{k}\neq 0}\frac{\exp(-i\mathbf{k}\cdot\mathbf{r})}{\eta k^2 V}\left(\mathbf{1} - \frac{\mathbf{kk}}{k^2}\right)$ is the Oseen mobility tensor for a periodic system. If an infinite non-periodic system were assumed, Eq. (2.6) is given by

$$\mathbf{v}_\infty(\mathbf{r}) = \mathbf{T}_0(\mathbf{r})\cdot\mathbf{F}, \tag{2.7}$$

where $\mathbf{T}_0(\mathbf{r}) = \frac{1}{8\pi\eta r}\left(\mathbf{1} + \frac{\mathbf{rr}}{r^2}\right)$ is the Oseen mobility tensor for an infinite system.

In the infinite system, the mean velocity of the target particle, $\mathbf{u}_0$, is given by:

$$\mathbf{u}_0 = \frac{D_\infty}{k_\mathrm{B}T}\mathbf{1}\cdot\mathbf{F}. \tag{2.8}$$

Based on the Kirkwood-Riseman theory, the mean velocity of the target particle in the periodic system, $\mathbf{u}_\mathrm{pbc}$, is given by:

$$\mathbf{u}_\mathrm{pbc} = \left[\frac{D_\infty}{k_\mathrm{B}T}\mathbf{1} - \lim_{r\to 0}\{\mathbf{T}_\mathrm{PBC}(\mathbf{r}) - \mathbf{T}_0(\mathbf{r})\}\right]\cdot\mathbf{F}. \tag{2.9}$$

Thus, the diffusion coefficient for the periodic system is given by:

$$\begin{aligned} D_\mathrm{pbc} &= \frac{1}{3}\mathrm{Tr}\Big[D_\infty\mathbf{1} - k_\mathrm{B}T\lim_{r\to 0}\big\{\mathbf{T}_\mathrm{PBC}(\mathbf{r}) - \mathbf{T}_0(\mathbf{r})\big\}\Big] \\ &= D_\infty + \frac{k_\mathrm{B}T}{6\pi\eta}\lim_{r\to 0}\left[\sum_{\mathbf{k},\mathbf{k}\neq 0}^{\infty}\frac{4\pi\exp(-i\mathbf{k}\cdot\mathbf{r})}{k^2V} - \frac{1}{|\mathbf{r}|}\right]. \end{aligned} \tag{2.10}$$

Here, $\frac{1}{k^2+m^2}$ is the Fourier transform of the Green's function, which gives the Yukawa potential, $e^{-m\mathbf{r}}/4\pi|\mathbf{r}|$, or the Coulomb potential, $1/4\pi|\mathbf{r}|$, for $m = 0$ in the infinite system [46]. In the periodic system, the $\sum_\mathbf{k}\frac{4\pi\exp(-i\mathbf{k}\cdot\mathbf{r})}{k^2V}$ gives $\sum_\mathbf{n} 1/|\mathbf{r}+\mathbf{n}L|$.

Thus, Eq. (2.10) becomes:

$$D_\mathrm{pbc} = D_\infty + \frac{k_\mathrm{B}T}{6\pi\eta}\left(\sum_{\mathbf{n},\mathbf{n}\neq 0}^{\infty}\frac{1}{\mathbf{n}L} - \sum_{\mathbf{k}=0}^{0}\frac{4\pi}{k^2V}\right). \tag{2.11}$$

This equation can be solved using Ewald techniques, which finally gives:

$$\begin{aligned} D_\mathrm{pbc} &= D_\infty + \frac{k_\mathrm{B}T}{6\pi\eta}\left(-\frac{2\alpha}{\sqrt{\pi}} + \sum_{\mathbf{n}\neq 0}\frac{\mathrm{erfc}(\alpha n)}{n} + \sum_{\mathbf{k}\neq 0}\frac{4\pi\exp\left(-k^2/4\alpha^2\right)}{k^2V} - \frac{\pi}{V\alpha^2}\right) \\ &= D_\infty - \frac{k_\mathrm{B}T\xi}{6\pi\eta L}. \end{aligned} \tag{2.12}$$

This equation has also been obtained by Dünweg and Kremer [47].

2.2.2 Higher-order correction term for Yeh–Hummer method

Based on Hashimoto's solution to the Stokes equation, Yeh and Hummer mentioned that the above equation includes a higher-order correction term, which I will discuss in this subsection.

In Ref. [42], Hashimoto derived an equation for the mean velocity u as follows:

$$u = u_0 - \frac{B}{\eta L^3}, \tag{2.13}$$

where B is a coefficient depending on R_h. Hashimoto mentioned that the term $-\frac{B}{\eta L^3}$ is proportional to the volume fraction of the solvent: $4\pi R_\mathrm{h}^3/3L^3$. Also, Fushiki showed that the mean velocity has a higher-order term proportional to the volume fraction of the solvent [40]. Thus, the relationship between u and u_0 is given as:

$$u = u_0 \left(1 + \frac{4\pi R_\mathrm{h}^3}{3L^3}\right). \tag{2.14}$$

Therefore, based on Eqs. (2.8) and (2.9), I can modify the Yeh–Hummer method as follows:

$$D_\mathrm{pbc} = D_\infty \left(1 + \frac{4\pi R_\mathrm{h}^3}{3L^3}\right) - \frac{k_\mathrm{B}T\xi}{6\pi\eta L}. \tag{2.15}$$

Using the SE law: $D_\infty = k_\mathrm{B}T/6\pi\eta R_\mathrm{h}$, this equation provides the Yeh–Hummer method with the higher-order correction term. Therefore, the higher-order correction term originates from the system-size dependence of the mean velocity.

Here, I will discuss the origin of the system-size dependence of the mean velocity. A volume flow Q enters through the Y-Z plane of the simulation cell, and the mean velocity in the x-direction can be expressed as:

$$u = \frac{Q}{S}, \tag{2.16}$$

where S is the cross-sectional area of the flow path. In PBC, Q also exits through the counterpart of the Y-Z plane at the same time. Therefore, Q must be conserved. If the solute can be regarded as infinitesimal, S is L^2. Then, the mean velocity corresponding to u_0 is:

$$u_0 = \frac{Q}{L^2}. \tag{2.17}$$

However, in fact, the finite volume of the solute $4\pi R^3/3$ decreases the volume of the flow path $S \cdot L$ as follows:

$$S \cdot L = L^3 - \frac{4\pi R^3}{3}, \tag{2.18}$$

which gives:

$$S = L^2 - \frac{4\pi R^3}{3L}. \tag{2.19}$$

Therefore, the mean velocity considering the finite volume of the solute can be expressed as:

$$u = \frac{Q}{L^2 - \frac{4\pi R^3}{3L}}. \tag{2.20}$$

Substituting Eq. (2.17) into this equation gives:

$$u = u_0 \frac{L^2}{L^2 - \frac{4\pi R^3}{3L}} = u_0 \frac{1}{1 - \frac{4\pi R^3}{3L^3}} = \frac{u_0 \left(1 + \frac{4\pi R^3}{3L^3}\right)}{1 - \left(\frac{4\pi R^3}{3L^3}\right)^2}. \tag{2.21}$$

Since $\left(\frac{4\pi R^3}{3L^3}\right)^2 \approx 0$, I finally obtain Eq. (2.14). Therefore, the system-size dependence of the mean velocity, as well as the higher-order correction term, originates from the narrowing of the flow path.

The above discussions finally yield:

$$D_\infty = D_{\mathrm{pbc}} + \frac{k_\mathrm{B} T \xi}{6\pi\eta L} - \frac{2 k_\mathrm{B} T R_\mathrm{h}^2}{9\eta L^3}. \tag{2.22}$$

Here, D_{pbc} corresponds to D_{MD} obtained from MD simulations with PBC. The third term on the right-hand side of Eq. (2.22) is smaller than the second term when L is sufficiently large. Thus, we can find some approximation levels in this equation. The diffusion coefficients in the infinite system are defined as follows:

1. The diffusion coefficient obtained through Fushiki's method, D_∞^{F}:

$$D_\infty^{\mathrm{F}} = D_{\mathrm{MD}} + \alpha \frac{1}{L}, \tag{2.23}$$

2. The diffusion coefficient obtained through the Yeh–Hummer method, D_∞^{YH1}:

$$D_\infty^{\mathrm{YH1}} = D_{\mathrm{MD}} + \frac{k_\mathrm{B} T \xi}{6\pi\eta L}, \tag{2.24}$$

3. The diffusion coefficient obtained through the Yeh–Hummer method with the higher-order correction term, D_∞^{YH2}:

$$D_\infty^{\mathrm{YH2}} = D_{\mathrm{MD}} + \frac{k_\mathrm{B} T \xi}{6\pi\eta L} - \frac{2 k_\mathrm{B} T R_\mathrm{h}^2}{9\eta L^3}. \tag{2.25}$$

These equations correspond to Eqs. (1.38), (1.39), and (1.40), respectively. In the following sections, these equations will be applied to the results obtained from MD simulations.

Table 2.1: System parameters

the simulation cell size L[nm]	the number of water molecules	the number of Na^+ ions
chignolin[a]		
2.49092	471	2
2.94867	807	2
4.06195	2171	2
5.00068	4084	2
6.00696	7106	2
6.98039	11173	2
8.00418	16865	2
9.01075	24078	2
β-lactoglobulin[a]		
5.99188	6334	8
6.96927	10407	8
9.01441	23394	8
11.00137	43140	8
12.01388	56409	8
12.95841	70970	8
14.00551	89805	8
14.96803	109777	8
pure water		
2.48582	512	0

[a] The number of proteins is fixed to 1. ;

2.3 MD simulation

I conducted molecular dynamics (MD) simulations under NVT conditions to calculate the diffusion coefficients for a protein in an aqueous solution. Two proteins, chignolin [48] and β-lactoglobulin (BLG) [49], were adopted as small and large proteins, respectively. The TIP4P2005 model [50] was used for water molecules. To achieve charge neutrality, Na^+ ions were added to the solution. For proteins and ions, the all-atom optimized potentials for liquid simulations (OPLS-AA) force field was adopted [51]. The simulations were conducted under ambient conditions at $T = 298.15$ K and $P = 1$ bar. The temperature was maintained using the stochastic velocity rescaling thermostat [52] with a relaxation time of 0.1 ps. The box length, L, was determined as the average cell size obtained from NPT simulations conducted at $T = 298.15$ K and $P = 1$ bar. The pressure was maintained using the Parrinello–Rahman barostat [53]. The parameters are summarized in Table 2.1. In addition, MD simulations of pure water under NVT conditions were conducted to calculate the solvent viscosity. All NVT simulations were run for 200 ns after equilibration. GROMACS software was used for these simulations, using particle mesh Ewald (PME) summation [54, 55, 56]. The equations of motion were integrated with the leap-frog algorithm, using a time step of 2 fs.

The mean-square displacement (MSD), $\langle |r(t) - r(0)|^2 \rangle$, of each protein was calculated from the MD trajectory. In this chapter, the diffusion coefficient D_{pbc} is defined as:

$$D_{\mathrm{MD}} = \frac{1}{6} \lim_{t \to \infty} \frac{\partial}{\partial t} \langle |r(t) - r(0)|^2 \rangle, \tag{2.26}$$

where $r(t)$ denotes the position of the protein at time t. The value of D_{MD} was determined from the slope of the MSD. Specifically, the MSD data in the time ranges of 40 ps to 80 ps (for chignolin) and 60 ps to 100 ps (for BLG) were used to calculate D_{MD}. After D_{MD} was obtained, the diffusion coefficients for an infinite system were corrected using Eqs. (2.23), (2.24), and (2.25), respectively.

The shear viscosity η was calculated using the Green–Kubo relation: [44, 57]

$$\eta = \lim_{t\to\infty} \frac{V}{k_{\mathrm{B}}T} \int_0^t \frac{1}{5} \sum_{i=1}^{5} \langle P_i(t) P_i(0) \rangle \, dt, \tag{2.27}$$

where V is the system volume, T is the temperature, and P_i represents each of the five independent components of the traceless stress tensor, defined as $P_{i=1\sim5} = [(P_{xx} - P_{yy})/2, (P_{yy} - P_{zz})/2, P_{xy}, P_{yz}, P_{zx}]$. The viscosity η was calculated by integrating the correlation function up to 10 ps. In this study, error values represent the 95% confidence interval.

2.4 Results and Discussion

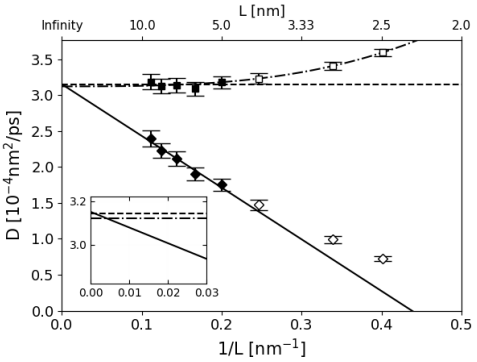

Figure 2.1: The diffusion coefficients of chignolin are shown as a function of the inverse box length $1/L$ [nm^{-1}] for systems containing N = 471–24078 water molecules, sodium ions, and chignolin (L = 2.49092 nm–9.01075 nm). Diamonds represent the uncorrected diffusion coefficients D_{MD} obtained from MD simulations for each box size. Squares correspond to the corrected diffusion coefficients calculated using Eq. (2.24). The straight solid and dashed lines are estimated from the closed diamonds and closed squares, respectively. The dash–dot line is plotted using Eq. (2.28), with parameters determined from the three open squares. The inset enlarges the range between 0.0 nm^{-1} and 0.03 nm^{-1}.

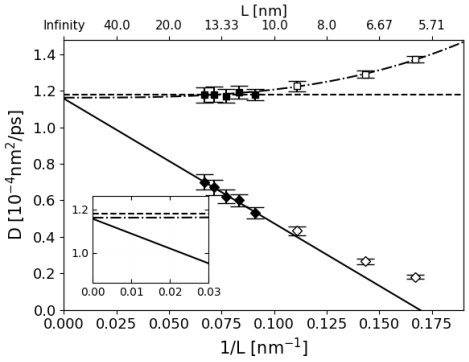

Figure 2.2: The diffusion coefficients of $\beta-$lactoglobulin (BLG) are shown as a function of the inverse box length $1/L$ [nm^{-1}] for systems containing $N = 6334$–109777 water molecules, sodium ions, and BLG ($L = 5.99188$ nm–14.96803 nm). Diamonds represent the uncorrected diffusion coefficients D_{MD} obtained from MD simulations for each box size. Squares correspond to the corrected diffusion coefficients calculated using Eq. (2.24). The straight solid and dashed lines are estimated from the closed diamonds and closed squares, respectively. The dash–dot line is plotted using Eq. (2.28), with parameters determined from the three open squares. The inset enlarges the range between 0.0 nm^{-1} and 0.03 nm^{-1}.

Fig. 2.1 shows the system-size dependence of the diffusion coefficients of chignolin in an aqueous solution. The diffusion coefficient D_{MD} increases as the cell size L becomes larger (see closed and open diamonds). As the cell size approaches infinity, the value converges to the diffusion coefficient of an infinite system at $L = \infty$. By performing a linear fit to the closed diamonds, D_{∞}^{F} can be obtained from the intercept. Squares (both closed and open symbols) represent the corrected diffusion coefficients calculated using the simplified Yeh–Hummer method (Eq. (2.24)) for each box size with the viscosity of pure water $(8.68 \pm 0.11) \times 10^{-4}\,\mathrm{Pa \cdot s}$. In the region represented by closed squares, the plotted diffusion coefficients exhibit a constant value comparable to D_{∞}^{F}. However, in the region represented by open squares, the corrected values are not constant. As $1/L$ increases, the deviation becomes more pronounced, particularly larger than 0.25 nm^{-1}.

Fig. 2.2 shows the system-size dependence of the diffusion coefficients for β-lactoglobulin (BLG) in an aqueous solution. A similar trend is observed in the open squares, with the deviation becoming more pronounced for values larger than 0.1 nm^{-1}.

If the simplified Yeh–Hummer method [Eq. (2.24)] is applied to the open diamonds, the resulting $D_{\infty}^{\mathrm{YH1}}$s deviate from the true value. Specifically, the ratios of the $D_{\infty}^{\mathrm{YH1}}$s between the smallest and largest cells are 1.13 for chignolin and 1.16 for BLG. These are non-negligible overestimations. In contrast, applying the Fushiki method [Eq. (2.23)] to the open diamonds leads to a significant underestimation of D_{∞}. These results suggest that when using macromolecules, neglecting the higher-order correction term may induce significant errors. Although this is-

sue can be addressed by using a sufficiently large cell, such a brute-force approach becomes increasingly difficult as the size of the macromolecule increases.

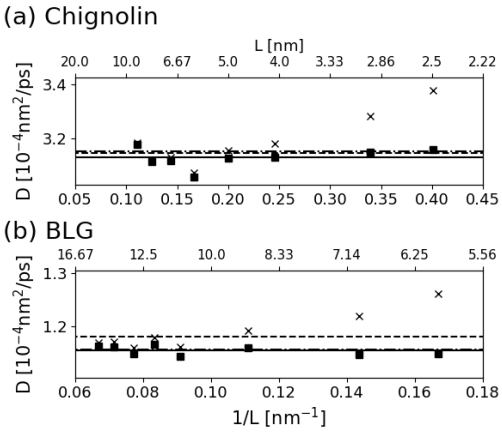

Figure 2.3: (a) The diffusion coefficient of chignolin as a function of $1/L$. (b) The diffusion coefficient of BLG as a function of $1/L$. In each figure, the squares represent corrected diffusion coefficients estimated using each uncorrected diffusion coefficient D_{MD} and Eq. (2.25) with $R(D_{\infty}^{\mathrm{YH2}})$. The straight solid lines are obtained from the average of the squares. The straight dash–dot and dashed lines show the previous results of the Fushiki and Yeh–Hummer methods, respectively. The crosses indicate corrected diffusion coefficients estimated using each uncorrected diffusion coefficient D_{MD} and Eq. (2.25) with the radius of gyration $\langle R_g \rangle$.

Before further discussion, I examined the Yeh–Hummer method with the higher-order correction term [Eq. (2.25)]. However, Eq. (2.25) explicitly contains the hydrodynamic radius R_{h}. The R_{h} will now be determined independently of Eq. (2.25). As shown above, for small proteins, such as chignolin and BLG, it is possible to prepare systems with sufficiently large L. For both proteins, the left-most set of five data points (closed squares) is constant. The average $D_{\infty}^{\mathrm{YH1}}$ values obtained from these five points are listed in Table 2.2. The Fushiki method is also applied to the corresponding five data points (closed diamonds), and the resulting D_{∞}^{F} values are also shown in Table 2.2. The $D_{\infty}^{\mathrm{YH1}}$ and D_{∞}^{F} values are nearly identical for both proteins. Using the SE law with each $D_{\infty}^{\mathrm{YH1}}$, I obtained the R_{h}s. Subsequently, I calculated D_{∞} using Eq. (2.25) with these R_{h} values. The results are plotted as squares in Fig. 2.3 (a) and (b). Both plots remain constant and agree well with D_{∞}^{F} (dash–dot line) and $D_{\infty}^{\mathrm{YH1}}$ (dashed line). This result suggests that the deviations in estimated values obtained from Eq. (2.24) can be corrected by the higher-order correction term.

In evaluating Eq. (2.25), it is necessary to have an accurate hydrodynamic radius R_{h}. Although each protein has an average gyration radius $\langle R_g \rangle$, which can be easily obtained from MD simulations, $\langle R_g \rangle$s cannot simply replace R_{h}s. I calculated the values of $\langle R_g \rangle$ for each protein (chignolin: 0.562 nm, BLG: 1.50 nm) and adopted these gyration radii as R_{h} in Eq. (2.25). The results are plotted as crosses in Figs. 2.3(a) and (b). However, the deviations from the true

values still remain. This indicates that gyration radii are not suitable for evaluating Eq. (2.25). By using the typical relationship between the hydrodynamic radius and the gyration radius for a spherical macromolecule (see Eq. (1.15)), improvements may be possible under certain conditions. In fact, using the relation, gyration radii provide values close to the actual hydrodynamic radius, 1.94 nm for BLG and 0.726 nm for chignolin. However, experimental data show that proteins of similar size can have widely varying diffusion coefficients [1, 2, 3, 4, 5, 6, 7, 8, 9], which indicates that $\langle R_g \rangle$ is not appropriate for estimating D_∞.

Next, I introduce a method that does not rely on estimating the particle size. Since the above analysis demonstrated that D_∞^{YH} depends on the system size, it is replaced with $D_{\mathrm{YH}}(x)$, where $x = 1/L^3$. Following this substitution, D_{MD} can be eliminated from Eq. (2.25) using Eq. (2.24). Thus, Eq. (2.25) can be reformulated as follows:

$$D_{\mathrm{YH}}(x) = D_\infty + \beta(R)x, \tag{2.28}$$

where the constant $\beta(R)$ is defined as $\frac{2k_{\mathrm{B}}TR_{\mathrm{h}}^2}{9\eta_{\mathrm{sol}}}$. The parameters D_∞ and $\beta(R_{\mathrm{h}})$ were determined by fitting the data. The estimated value of D_∞ obtained from Eq. (2.28) is denoted as D_0^{YH2}. Although this equation involves two parameters, D_0^{YH2} and $\beta(R_{\mathrm{h}})$, it is effectively a single-parameter model because there exists a relationship between R_{h} and D_0^{YH2}. To facilitate the linear fitting of $D_{\mathrm{YH}}(x)$, $\beta(R)$ is introduced as an additional parameter. The parameters D_∞ and $\beta(R_{\mathrm{h}})$ were obtained by fitting to the data points represented by open squares, corresponding to D_{YH} in small systems. The resulting lines, determined by these parameters, are shown as dash–dot lines in Figs. 2.1 and 2.2.

Table 2.2 lists the obtained D_0^{YH2} and the R_{h}s derived from D_0^{YH2} and $\beta(R)$. The agreement between $R_{\mathrm{h}}(D_0^{\mathrm{YH2}})$ and $R_{\mathrm{h}}(\beta)$ exhibits the consistency in Eq. (2.28). Additionally, the table also includes radii calculated using the Fushiki and Yeh–Hummer methods with sufficiently large cells. The results from the four methods are in close agreement. This consistency indicates that the method based on Eq. (2.28) is effective when constructing a sufficiently large basic cell, as required by the Fushiki and Yeh–Hummer methods, is impractical.

Now, I will compare the present results with the experimental data. The diffusion coefficient of BLG in pH6 HCl solution of heavy water (D_2O) is 0.946×10^{-4} nm^2/ps, as measured by pulsed field gradient NMR spectroscopy[58]. Considering the difference in viscosity between TIP4P2005 water and D_2O, the expected value for the diffusion coefficient of BLG is calculated as $0.946 \times 10.9/8.68 \times 10^{-4} = 1.19 \times 10^{-4}$ nm^2/ps. Here, I used a value of 10.9 $\times10^{-4}$ Pa·s for the viscosity of D_2O[59]. It should be noted that the experimental system could contain dimers, and the solvent in the experiment was D_2O. Furthermore, the molecules used in the present study are only a model. While further comparisons are necessary, no significant discrepancies

Table 2.2: Diffusion coefficients for D_∞ and hydrodynamic radius R_h in each method.[a]

	Fushiki Method		Yeh–Hummer Method(Eq. (2.24))		Eq. (2.28)(Eq. (2.25))		
protein	$D_\infty^\mathrm{F}[10^{-4}\mathrm{nm}^2/\mathrm{ps}]$	$R(D_\infty^\mathrm{F})[\mathrm{nm}]$	$D_\infty^\mathrm{YH1}[10^{-4}\mathrm{nm}^2/\mathrm{ps}]$	$R(D_\infty^\mathrm{YH1})[\mathrm{nm}]$	$D_\infty^\mathrm{YH2}[10^{-4}\mathrm{nm}^2/\mathrm{ps}]$	$R(D_\infty^\mathrm{YH2})[\mathrm{nm}]$	$R(\beta)[\mathrm{nm}]$
chignolin	3.15	0.798	3.14	0.800	3.12	0.806	0.832
BLG	1.16	2.17	1.18	2.13	1.16	2.16	2.05

[a] $R(D_\infty)$ are obtained using the Stokes–Einstein law from each D_∞.;

are observed between the experimental data and the simulation results (see Table 2.2).

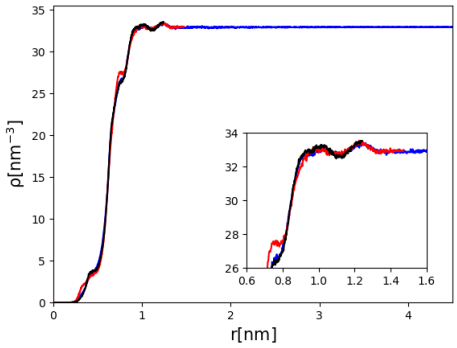

Figure 2.4: The radial density distribution function of water molecules (oxygen) around the center of the protein molecule (chignolin). The systems with $L = 2.49092$ nm (black line), $L = 2.94867$ nm (red line), and $L = 9.01075$ nm (blue line) are shown. The inset provides an enlarged view of the region around $r = 1.0$ nm.

The hydrodynamic effects of the basic cell size have been discussed. However, the finite size also affects the hydration structure. To investigate the effect of the static structure, I calculated the density distribution functions of water around the center of the protein. Fig. 2.4 shows the density distribution $\rho(r)$ around chignolin. For $L = 2.94867$ nm and $L = 9.01075$ nm, the density approaches the bulk value at the tail of the function. In contrast, for $L = 2.49092$ nm, $\rho(r)$ does not reach the bulk value at the tail. The hydration shell breaks at the peak of the oscillation around $r = 1.2$ nm. Despite the incomplete hydration shell, the diffusion coefficients still obey the hydrodynamic law described by Eq. (2.24). Fig. 2.5 shows the density distribution $\rho(r)$ around BLG. For the smallest basic cell size, $L = 5.99188$ nm, $\rho(r)$ oscillates around $r = 3.0$ nm, indicating that the complete hydration shell cannot form within the simulation box. However, the diffusion coefficients again adhere to Eq. (2.24). These results suggest that the effect of the solvation structure on diffusion is relatively minor when the macromolecule has sufficient solvation layers within the simulation box. Fig. 2.5 also shows $\rho(r)$ for three different values of L. Although the short-range structure of $\rho(r)$ differs slightly among the cases, no significant effect on Eq. (2.24) is observed. For instance, peaks around $r = 0.5$ nm, corresponding to water molecules at the ligand-binding site of BLG, vary in shape among the cases. However, the diffusion coefficients remain in agreement with Eq. (2.24). This indicates

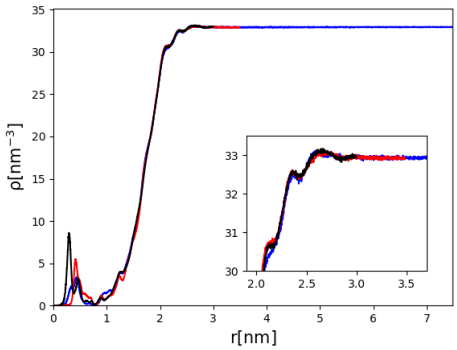

Figure 2.5: The radial density distribution function of water molecules (oxygen) around the center of the protein molecule (BLG). The systems with $L = 5.99188$ nm (black line), $L = 6.96927$ nm (red line), and $L = 14.96803$ nm (blue line) are shown. The inset provides an enlarged view of the region around $r = 2.0$ nm.

that the effect of binding water is negligible, as the number of water molecules at the binding site remains constant in this case.

2.5 Conclusion

MD simulations were carried out to evaluate a higher-order correction term [the third term on the right-hand side of Eq. (2.25)] and to obtain the diffusion coefficient of a globular protein in an aqueous solution at the dilute limit. Although the higher-order term was shown in Yeh and Hummer's paper, the simplified Yeh–Hummer formula [Eq. (2.25)] has been widely adopted for corrections. When the solute molecule becomes larger, such as in the case of proteins, the higher-order term appears to be important. Therefore, I assessed the higher-order correction term to analyze the diffusion of large molecules, such as proteins. In many studies, the simplified Yeh–Hummer formula [Eq. (2.24)] has been widely adopted and generally provides an adequate estimate of the diffusion coefficient for an infinite system when the system size is sufficiently large. However, my MD results demonstrated that the diffusion coefficient corrected by the simplified formula depends on the basic cell size. Thus, the simplified formula was insufficient in these cases. The deviations were non-negligible when the cell sizes were small. These deviations were explained by the higher-order correction term for the Yeh–Hummer method. The formula Eq. (2.28), which is derived from Eq. (2.24), is more powerfull than the simplified formula because it significantly reduces the computational cost of calculating diffusion coefficients for solute molecules at infinite dilution. This reduction becomes increasingly important as the solute size increases.

Chapter 3

Decomposition of friction coefficients to analyze solvation effects on diffusion[60]

3.1 Introduction

As shown in the previous chapter, the Yeh–Hummer method with the higher-order correction term allows us to accurately calculate protein diffusion coefficients for an infinite system from MD simulations for a given interaction model under thermodynamic conditions. MD simulations with the correction method must become a remarkable tool for studying solvation effects in diffusion coefficients because MD simulations offer advantages in conducting hydration analyses and investigating hypothetical solute models. These analyses provide important information for discussing solvation effects. On the other hand, a diffusion coefficient is a coarse-grained value for a Brownian particle. For example, the value obtained experimentally cannot be devided into individual contributions.

As explained in Sections 1.4 to 1.7 of Chapter 1, it would be helpful to decompose the friction coefficients, which are proportional to the inverse of the diffusion coefficient, into individual contributions from specific sites. In fact, the diffusion coefficient has not been decomposed into individual contributions in the context of this study. Here, we should discuss the friction coefficient.

The friction coefficient γ is defined as the integral of the friction kernel $\Gamma(t)$, which is related to the correlation function of random forces through the fluctuation-dissipation theorem [32] [see Section 1.4, Eq. (1.19), and Eq. (1.21)]. Therefore, the friction coefficient for a molecule

composed of N sites, such as atoms and functional groups, can be decomposed as follows:

$$\gamma = \sum_{i=0}^{N} \int_0^\infty \frac{1}{k_\mathrm{B}T} \langle f_\mathrm{R}^i(0) \cdot F_\mathrm{R}(\tau) \rangle \, d\tau \tag{3.1}$$

Here, $\mathbf{f}_\mathrm{R}^i(t)$ represents the random force acting on the i-site. If we define

$$K^i(t) = \frac{1}{k_\mathrm{B}T} \langle f_\mathrm{R}^i(0) \cdot F_\mathrm{R}(\tau) \rangle \tag{3.2}$$

as the contribution of the i-site to the friction kernel, the contribution of the i-site to the friction coefficient, κ^i, can be defined as follows:

$$\kappa^i = \int_0^\infty K^i(\tau) \, d\tau \tag{3.3}$$

Thus, if κ^i can be calculated, it is theoretically possible to decompose the friction coefficient into individual contributions. In this thesis, I refer to K^i and κ^i as the contribution kernel and contribution value for the i-site, respectively. In fact, using a similar definition for the contribution of the i-th component to the friction kernel, Yamaguchi et al. have decomposed the friction coefficients into contributions from the repulsive and attractive forces acting on a particle [61, 62]. However, their theory for a single particle cannot be directly applied to the decomposition of the friction coefficient for a polyatomic molecule.

Since proteins have rough and flexible surfaces, it is difficult to analyze and discuss the contribution of each site. Before studying proteins, we adopted fullerenols as an alternative to proteins. Since fullerenols also have mosaic surfaces and their shape is not flexible, we can expect that their analysis would be simpler than that of proteins. Several researchers have investigated the diffusion coefficients of fullerenols using molecular dynamics (MD) simulations [63, 64, 65]. These authors found that the values decreased with the number of OH groups and attributed this decrease to the increased number of hydrogen bonds. Their findings are consistent with the arguments of Terazima and coworkers [1, 2, 3, 4, 5, 6, 7, 8, 9]. However, the friction coefficient has not been decomposed into contributions from each site. Therefore, the aim of this study is to discuss the friction coefficient on the basis of the decomposition.

In this chapter, I present a method to decompose the friction coefficient into contributions from specific sites on a polyatomic molecule using MD simulations. First, I will introduce the method to calculate the contribution values κ^i from MD simulations. Second, the method will be applied to MD simulations with models of a fullerene or a fullerenol in an aqueous solution. Finally, the dependence of the friction coefficient on the number of OH groups will be discussed based on the analysis.

3.2 Method to calculate the contribution values κ^i from MD simulations

3.2.1 The relation about the contribution kernel K^i

Before obtaining the contribution values κ^i, it is necessary to first determine the contribution kernel K^i. Unfortunately, the contribution kernel, defined as the correlation function of the random forces, cannot be directly calculated from MD simulations. Thus, I derive another expression for K^i, and I revisit the Generalized Langevin Equation (GLE) introduced in Section 1.4:

$$F(t) = -\int_0^t d\tau \Gamma(\tau) v(\tau - t) + F_{\mathrm{R}}(t). \tag{3.4}$$

Using K^i, the GLE can also be decomposed as:

$$F(t) = \sum_{i=1}^{N} \Big[-\int_0^t d\tau K^i(\tau) v(\tau - t) + f^i_{\mathrm{R}}(t) + f^i_{\mathrm{IN}}(\mathbf{q}(t)) \Big], \tag{3.5}$$

which gives:

$$f^i(t) = -\int_0^t d\tau K^i(\tau) v(\tau - t) + f^i_{\mathrm{R}}(t) + f^i_{\mathrm{IN}}(\mathbf{q}(t)), \tag{3.6}$$

where $f^i(t)$ and $f^i_{\mathrm{IN}}(\mathbf{q}(t))$ are the total force and the intramolecular force acting on the i-th site (or atom). The intramolecular forces are determined by the solute conformation $\mathbf{q}(t)$ at time t and are defined as the coordinate derivative of the intramolecular potential $V_{\mathrm{sol}}(\mathbf{q}(t))$, as follows:

$$f^i_{\mathrm{IN}}(\mathbf{q}(t)) = -\frac{\partial V_{\mathrm{sol}}(\mathbf{q}(t))}{\partial q^i}, \tag{3.7}$$

where q^i is a coordinate of the i-th site (or atom). For example, in the all-atom MD simulations, the spring potentials, such as bonds, angles, and dihedrals, are parts of the intramolecular potential. We can calculate $V_{\mathrm{sol}}(\mathbf{q}(t))$ in vacuum. However, because the solvent affects $V_{\mathrm{sol}}(\mathbf{q}(t))$, the calculation of $V_{\mathrm{sol}}(\mathbf{q}(t))$, and therefore $f^i_{\mathrm{IN}}(\mathbf{q}(t))$, is difficult. Thus, I first define the extramolecular force as:

$$f^i_{\mathrm{EX}}(t) = f^i(t) - f^i_{\mathrm{IN}}(\mathbf{q}(t)) = -\int_0^t d\tau K^i(\tau) v(\tau - t) + f^i_{\mathrm{R}}(t), \tag{3.8}$$

to derive another expression for K^i. The problem of the intramolecular force $f^i_{\mathrm{IN}}(\mathbf{q}(t))$ will be addressed in the next subsection.

To extract the contribution kernel, $f^i_{\mathrm{EX}}(0)$ is multiplied by Eq. (3.4) to compute the ensemble

average, i.e.:

$$\langle f^i_{\mathrm{EX}}(0)\cdot F(t)\rangle = -\int_0^t d\tau\Gamma(\tau)\langle f^i_{\mathrm{EX}}(0)\cdot v(\tau-t)\rangle + \langle f^i_{\mathrm{R}}(0)\cdot F_{\mathrm{R}}(t)\rangle, \tag{3.9}$$

and applying Eq. (3.2) gives the following equation:

$$K^i(t) = \frac{1}{k_{\mathrm{B}}T}\left[C^{f(\mathrm{EX,i})\mathrm{F}}(t) + \int_0^t d\tau\Gamma(\tau)C^{f(\mathrm{EX,i})\mathrm{v}}(t-\tau)\right], \tag{3.10}$$

where $C^{f(\mathrm{EX,i})\mathrm{v}}(t) = \langle f^i_{\mathrm{EX}}(0)\cdot v(t)\rangle$, $C^{f(\mathrm{EX,i})\mathrm{F}}(t) = \langle f^i_{\mathrm{EX}}(0)\cdot F(t)\rangle$. If we have $C^{f(\mathrm{EX,i})\mathrm{v}}(t)$, $C^{f(\mathrm{EX,i})\mathrm{F}}(t)$, and $\Gamma(t)$, the contribution kernels can be calculated using Eq. (3.10). Here, $\Gamma(t)$ can be calculated using methods proposed by previous researchers [66, 67]. However, the calculations of $C^{f(\mathrm{EX,i})\mathrm{v}}(t)$ and $C^{f(\mathrm{EX,i})\mathrm{F}}(t)$ require certain techniques. Next, I will introduce the method to obtain $C^{f(\mathrm{EX,i})\mathrm{v}}(t)$ and $C^{f(\mathrm{EX,i})\mathrm{F}}(t)$.

3.2.2 Method to obtain $C^{f(\mathrm{EX,i})\mathrm{v}}(t)$ and $C^{f(\mathrm{EX,i})\mathrm{F}}(t)$

As explained in the previous subsection, the direct calculation of $C^{f(\mathrm{EX,i})\mathrm{v}}(t)$ and $C^{f(\mathrm{EX,i})\mathrm{F}}(t)$ is difficult because calculating $f^i_{\mathrm{IN}}(\mathbf{q}(t))$ and $f^i_{\mathrm{EX}}(0)$ is also challenging. Thus, I introduce a method to calculate $C^{f(\mathrm{EX,i})\mathrm{v}}(t)$ and $C^{f(\mathrm{EX,i})\mathrm{F}}(t)$ without directly calculating f^i_{EX} or f^i_{IN}.

If we have another MD trajectory with the same solute conformations $\mathbf{q}_0$, which corresponds to the solute conformation $\mathbf{q}(0)$ in the original MD trajectory, but with different solvent configurations, $C^{f(\mathrm{EX,i})\mathrm{v}}(t)$ can be calculated as follows:

$$\begin{aligned}
&C^{f(\mathrm{EX,i})\mathrm{v}}(t)\\
&= \langle f^i_{\mathrm{EX}}(0)\cdot v(t)\rangle = \langle f^i_{\mathrm{EX}}(0)\cdot v(t)\rangle - \langle f^i_{\mathrm{EX,ano}}\cdot v(t)\rangle\\
&= \langle [f^i_{\mathrm{EX}}(0) - f^i_{\mathrm{EX,ano}}]\cdot v(t)\rangle\\
&= \langle [f^i_{\mathrm{EX}}(0) - f^i_{\mathrm{EX,ano}} + f^i_{\mathrm{IN}}(\mathbf{q}_0) - f^i_{\mathrm{IN,ano}}(\mathbf{q}_0)]\cdot v(t)\rangle\\
&= \langle [f^i_{\mathrm{EX}}(0) + f^i_{\mathrm{IN}}(\mathbf{q}(0)) - \{f^i_{\mathrm{EX,ano}} + f^i_{\mathrm{IN,ano}}(\mathbf{q}_0)\}]\cdot v(t)\rangle\\
&= \langle [f^i(0) - f^i_{\mathrm{ano}}]\cdot v(t)\rangle,
\end{aligned} \tag{3.11}$$

where f^i_{ano}, $f^i_{\mathrm{EX,ano}}$, and $f^i_{\mathrm{IN,ano}}(\mathbf{q}_0)$ are the total, extramolecular, and intramolecular forces acting on the i-th site (or atom) in another MD trajectory. Here, $f^i_{\mathrm{IN}}(\mathbf{q}_0) - f^i_{\mathrm{IN,ano}}(\mathbf{q}_0) = 0$ because the same conditions for the solvent yield the same intramolecular forces, and $\langle f^i_{\mathrm{EX,ano}}\cdot v(t)\rangle = 0$, provided the solvent flow is independent of the original MD trajectory. Similarly,

$C^{f(\mathrm{EX,i})\mathrm{F}}(t)$ can be calculated as follows:

$$C^{f(\mathrm{EX,i})\mathrm{F}}(t) = \langle [f^i(0) - f^i_{\mathrm{ano}}] \cdot F(t) \rangle. \tag{3.12}$$

Therefore, we can calculate $C^{f(\mathrm{EX,i})\mathrm{v}}(t)$ and $C^{f(\mathrm{EX,i})\mathrm{F}}(t)$ using $f^i - f^i_{\mathrm{ano}}$ instead of $f^i_{\mathrm{EX}} = f^i - f^i_{\mathrm{IN}}$.

Here, f^i can be obtained directly from the original MD trajectory. In contrast, f^i_{ano} can be calculated using additional MD simulations in which the solute conformation is frozen at the target solute conformation $\mathbf{q}_0$, while the water molecules move freely. In these simulations, the correction for the velocity of the center of mass is turned off, allowing the solute molecule's center to move freely relative to the solvent molecules. This approach is known as the belly approach [68]. In this thesis, I refer to MD simulations using the belly approach as 'belly simulations'.

In the belly simulation, each solute atom is assumed to be trapped in an infinitely strong bias potential $W(\mathbf{q})$, which is a function of the solute conformation $\mathbf{q}$ with the following conditions:

$$W(\mathbf{q}_0) = 0, \text{ and } -\frac{\partial}{\partial \mathrm{q}^i} W(\mathbf{q}_0) = 0, \tag{3.13}$$

where q^i is a coordinate of the i-th solute atom. In constrained simulations, the relationship between the mean force $f^i_{\mathrm{solv}}(\mathbf{q})$ and the mean force in constrained simulations $f^i_{\mathrm{solv,res}}(\mathbf{q})$ is:

$$f^i_{\mathrm{solv}}(\mathbf{q}) = f^i_{\mathrm{solv,res}}(\mathbf{q}) - \frac{\partial}{\partial \mathrm{q}^i} W(\mathbf{q}), \tag{3.14}$$

where $f^i_{\mathrm{solv}}(\mathbf{q}_0) = f^i_{\mathrm{solv,res}}(\mathbf{q}_0)$ using Eq. (3.13). Hence, f^i_{ano} can be obtained from the belly simulation without altering the intramolecular force term $f^i_{\mathrm{IN}}(\mathbf{q}_0)$.

A single belly simulation provides one $f^i_{\mathrm{ano}}(\mathbf{q}_0)$. To calculate the averages of $C^{f(\mathrm{EX,i})\mathrm{v}}(t)$ and $C^{f(\mathrm{EX,i})\mathrm{F}}(t)$, multiple belly simulations are required to obtain restart points for $f^i_{\mathrm{ano}}(\mathbf{q}_0)$. However, the equilibration steps for each belly simulation require high computational cost. Thus, I connected those belly simulations using several quasistatic steps to reduce the equilibration steps. Details are provided in Appendix E.

Finally, we can now calculate $C^{f(\mathrm{EX,i})\mathrm{v}}(t)$, $C^{f(\mathrm{EX,i})\mathrm{F}}(t)$, and $\Gamma(t)$. Using these, the contribution kernel K^i can be calculated from Eq. (3.10), and the contribution value κ^i can be obtained from Eq. (3.3). In the following sections, the method introduced in this section will be applied to MD simulations of fullerenol in an aqueous solution.

3.3 MD simulation

I conducted MD simulations to obtain diffusion properties, including the friction coefficient. The simulation system consisted of 1000 SPC/E water [69] molecules and a solute in a cubic box with periodic boundary conditions. We also performed MD simulations of systems with solute-to-water molecule ratios of 1/1500 and 1/2200 to confirm the system size effect [41]. Fullerene and fullerenols were adopted as the solute molecules. The structures are shown in Fig. 3.1. The force field Generalized Amber Force Field 2 (GAFF2) was used for the solute molecules. The solute parameters were generated by AmberTool [70] with the AM1–BCC charge model [71]. Before parameterization, the structure was optimized using the HF/6-31G* level in Gaussian 16 [72]. Cases where all the solute charge parameters were set to zero were also investigated. In this study, these solutes are referred to as 'nonpolar'.

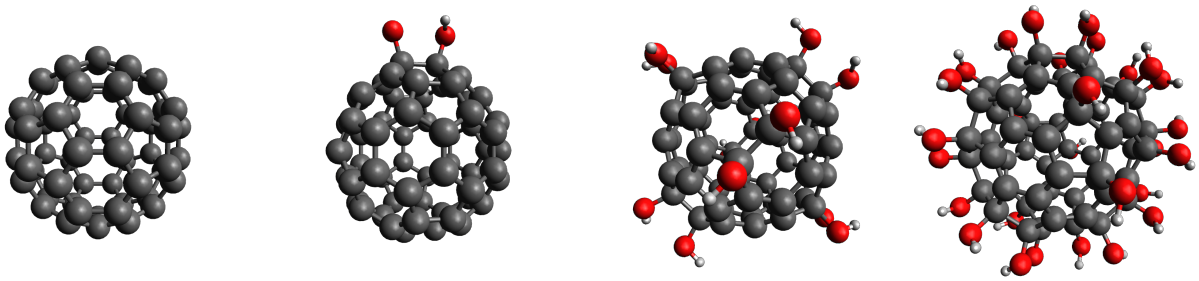

Figure 3.1: The structures of fullerene and fullerenols. From left to right, C_{60}, $C_{60}(OH)_2$, $C_{60}(OH)_{12}$, and $C_{60}(OH)_{28}$.

Production runs were performed under NVT conditions. Ambient conditions of $T = 298.15$ K and $P = 1$ bar (10^5 Pa) were adopted. The temperature was controlled by the stochastic velocity rescaling thermostat [52], coupled with a time constant of 0.5 ps for the solvent only. This thermostat does not influence the solute friction kernel [73]. Each simulation cell size, L, as well as the volume L^3, was determined by the average cell size under NPT conditions at a pressure of $P = 1$ bar, controlled by a stochastic cell rescaling barostat [74]. After equilibration, production simulations were conducted using more than 40 independent trajectories for a total of 1 or 3 μs. All MD simulations were performed using GROMACS 2021.4 [75] with particle mesh Ewald summation (PME) [56]. A cutoff of 1.2 nm was applied to the direct space of PME and the non-bonded van der Waals interactions. All solute bonds were kept fixed using the LINCS algorithm [76], and the bonds and angles of the water molecules were kept fixed using the SETTLE algorithm [77]. The equations of motion were integrated using the leap-frog algorithm with a time step of 2 fs.

For the extraction of the friction kernel $\Gamma(t)$, I used the iterative method developed by Kowalik et al. [67]. We can calculate the friction kernel using a numerical method [66] or

Laplace transform methods [78, 79] based on the first-kind Volterra integral equation:

$$m\dot{C}^{vv}(t) = m\langle v(0) \cdot F(t)\rangle = -\int_0^t d\tau \Gamma(\tau) C^{vv}(\tau - t), \tag{3.15}$$

which is derived from the GLE by multiplying Eq. (1.18) by the initial velocity $v(0)$ and taking an ensemble average [80, 81]. However, these calculations can lead to stability problems [82]. To improve the algorithms, Kowalik et al. integrated Eq. (3.15) and proposed the following equation:

$$mC^{vv}(t) = -\int_0^t d\tau G(\tau) C^{vv}(\tau - t) + mC^{vv}(0), \tag{3.16}$$

where $G(t)$ is the running integral over the friction kernel defined as

$$G(t) = \int_0^t d\tau \Gamma(\tau). \tag{3.17}$$

They demonstrated that the friction kernel can be obtained stably by first calculating $G(t)$ from Eq. (3.16) and then determining the slope of $G(t)$ [67]. The iterative method is expressed by the following discretized equation with time step Δt and frame number $j = 0, 1, 2, \ldots$:

$$G_j = \frac{2}{\Delta t C_0^{vv}} \left[mC_0^{vv} - mC_j^{vv} - \sum_{k=0}^{j-1} \Delta t G_j C_{j-k}^{vv} w_{j,k} \right], \tag{3.18}$$

where $G_j = G(j\Delta t)$, $C_j^{vv} = C^{vv}(j\Delta t)$, m is the mass of the solute, and $w_{j,k}$ is a weighting factor of the trapezoidal rule defined as $w_{j,j} = w_{j,0} = 1/2$ and $w_{j,k} = 1$ otherwise. After obtaining $G(t)$ from Eq. (3.18), the friction kernel $\Gamma(t)$ can be calculated as follows:

$$\Gamma_j = \begin{cases} \dfrac{G_1}{\Delta t} & (j = 0) \\ \dfrac{G_{j+1} - G_{j-1}}{2\Delta t} & (j \geq 1) \end{cases}, \tag{3.19}$$

where $\Gamma_j = \Gamma(t)$.

After obtaining $C^{f(\mathrm{EX,i})\mathrm{v}}(t)$, $C^{f(\mathrm{EX,i})\mathrm{F}}(t)$, and $\Gamma(t)$ from Eqs. (3.11), (3.12), and (3.16), I calculated the contribution kernels $K^i(t)$ using the discretized version of Eq. (3.10):

$$K_j^i = \frac{1}{mC_0^{vv}} \left[C_j^{f(\mathrm{EX,i})\mathrm{F}} + \sum_{k=0}^{j-1} \Delta t \Gamma_j C_{j-k}^{f(\mathrm{EX,i})\mathrm{v}} w_{j,k} \right], \tag{3.20}$$

where $K_j^i = K(j\Delta t)$, $C_j^{f(\mathrm{EX,i})\mathrm{F}} = C^{f(\mathrm{EX,i})\mathrm{F}}(j\Delta t)$, and $C_j^{f(\mathrm{EX,i})\mathrm{v}} = C^{f(\mathrm{EX,i})\mathrm{v}}(j\Delta t)$. Here, the equipartition theorem, $k_\mathrm{B}T = mC_0^{vv}$, was used. The contribution values κ^i were then obtained

using Eq. (3.3). The contribution function of a specific site, such as an OH group, can be obtained directly from Eq. (3.20) by replacing f^i_{EX} with the force acting on the center of mass of that site.

Since the diffusion and friction coefficients depend on the simulation cell size (see the previous chapter), I applied the correction proposed by Yeh and Hummer [41]. In this chapter, the results for different cell sizes with water-to-solute molecule ratios of 1:1000, 1:1500, and 1:2200 were analyzed. Afterward, I obtained the hydrodynamic radius from the D_0 of the 1:2200 ratio using Eq. (2.24) and corrected all D_{MD} values using Eq. (2.25). I used the averaged value of the three D_0s with different ratios. After obtaining D_0 and D_{MD}, the contribution values κ^i were scaled as follows:

$$\kappa^i_0 = \frac{D_{\mathrm{pbc}}}{D_0}\kappa^i \tag{3.21}$$

to obtain the contribution values of the infinite system, κ^i_0.

In this chapter, the diffusion coefficient D_{MD} and shear viscosity η were calculated using the Green–Kubo relation [44, 57]:

$$D_{\mathrm{MD}} = \frac{k_{\mathrm{B}}T}{mC^{vv}(0)}\int_0^\infty d\tau C^{vv}(\tau), \tag{3.22}$$

and Eq. (2.27). In Eq. (3.22), the value was scaled by $k_{\mathrm{B}}T/mC^{vv}(0)$ to unify with the ideal temperature T. The correlation functions, such as C^{vv}, $C^{f(\mathrm{EX,i})\mathrm{v}}$, and $C^{f(\mathrm{EX,i})\mathrm{F}}$, were averaged over all independent directions: x, y, and z.

3.4 Results and Discussion

I calculated the contribution values of $C_{60}(OH)_2$, $C_{60}(OH)_{12}$, and $C_{60}(OH)_{28}$ in water. To assess the effects of electrostatic interactions, I also performed calculations for nonpolar solutes, where the solute's charge parameters were set to zero. These solutes are referred to as nonpolar solutes, for example, nonpolar $C_{60}(OH)_2$. To highlight the difference, I will occasionally refer to the former solutes as polar solutes, for example, polar $C_{60}(OH)_2$.

First, I calculated the contribution kernels by decomposing the friction kernels using Eq. (3.20). The friction kernels were calculated using Eqs. (3.18) and (3.19) from the velocity autocorrelation functions [67]. Fig. 3.2 indicates the contribution kernels (left panels) and their integrated results (right panels) for each solute: $C_{60}(OH)_2$, nonpolar $C_{60}(OH)_2$, $C_{60}(OH)_{12}$, nonpolar $C_{60}(OH)_{12}$, $C_{60}(OH)_{28}$, and nonpolar $C_{60}(OH)_{28}$. The red and green solid lines represent the results for the COH site (a carbon atom with an OH group) and the C site (a carbon atom), respectively.

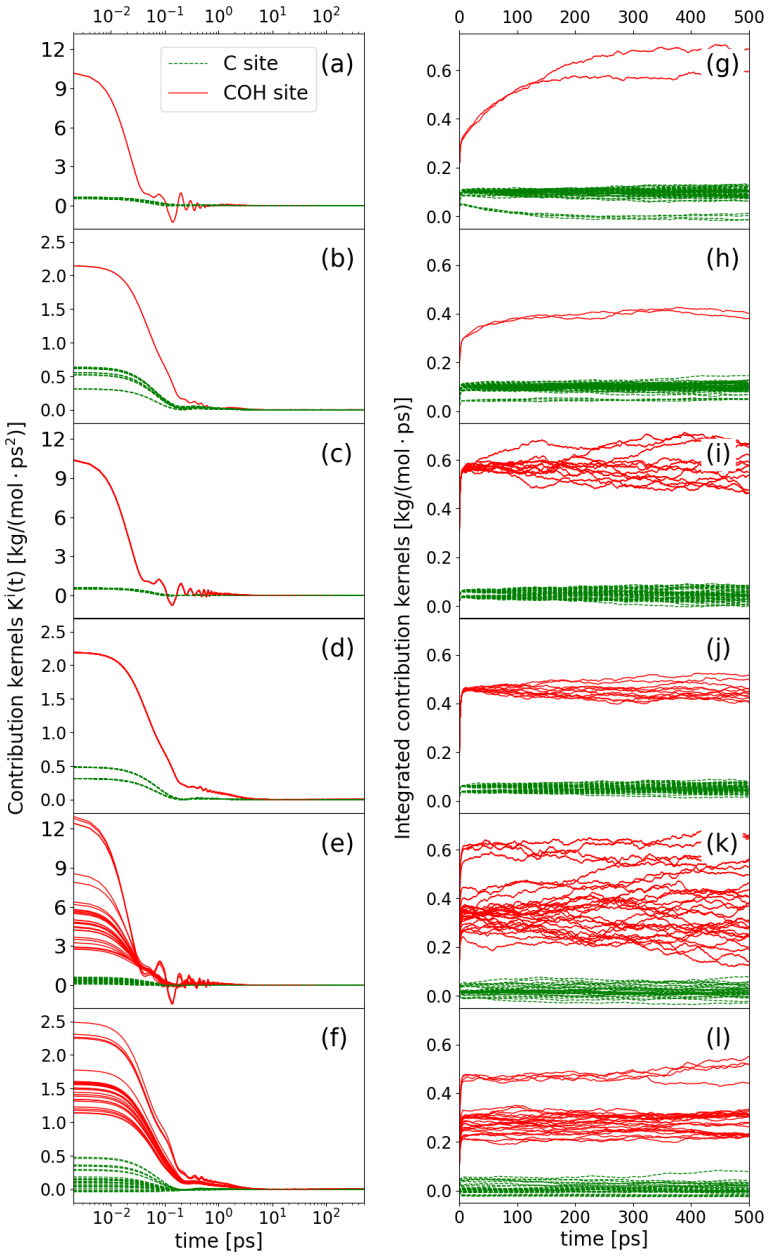

Figure 3.2: The contribution kernels (left panels) and the integrated contribution kernels (right panels) for 60 C sites or the COH sites in fullerenols are shown. From top to bottom: $C_{60}(OH)_2$ [(a) and (g)], nonpolar $C_{60}(OH)_2$ [(b) and (h)], $C_{60}(OH)_{12}$ [(c) and (i)], nonpolar $C_{60}(OH)_{12}$ [(d) and (j)], $C_{60}(OH)_{28}$ [(e) and (k)], and nonpolar $C_{60}(OH)_{28}$ [(f) and (l)]. The red solid lines and green dashed lines correspond to the results for the COH site and the C site, respectively.

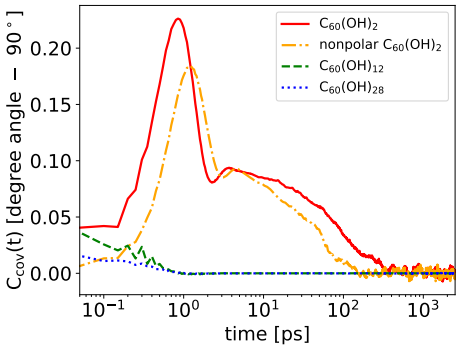

Figure 3.3: The angular correlation functions between the velocity and the molecular orientation defined by Eq. (3.23) are shown. The red solid line, orange dash-dotted line, green dashed line, and blue dotted line are the results for $C_{60}(OH)_2$, nonpolar $C_{60}(OH)_2$, $C_{60}(OH)_{12}$, and $C_{60}(OH)_{28}$, respectively.

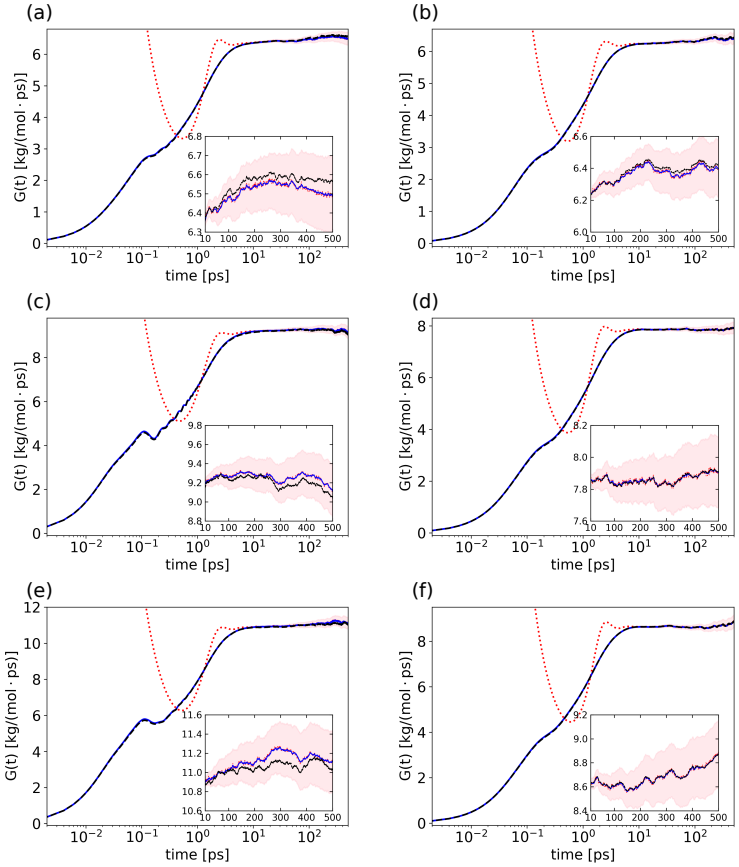

Figure 3.4: The summation of integrated contribution kernels (black dashed lines), the integrated friction kernels (blue solid lines) and $k_BT/\int C^{vv}(\tau)d\tau$ (red dotted lines) for (a) $C_{60}(OH)_2$, (b) nonpolar $C_{60}(OH)_2$, (c) $C_{60}(OH)_{12}$, (d) nonpolar $C_{60}(OH)_{12}$, (e) $C_{60}(OH)_{28}$, and (f) nonpolar $C_{60}(OH)_{28}$. The pink areas represents the 95% confidence interval for the integrated friction kernels. The inset enlarges the range between 10 ps and 500 ps.

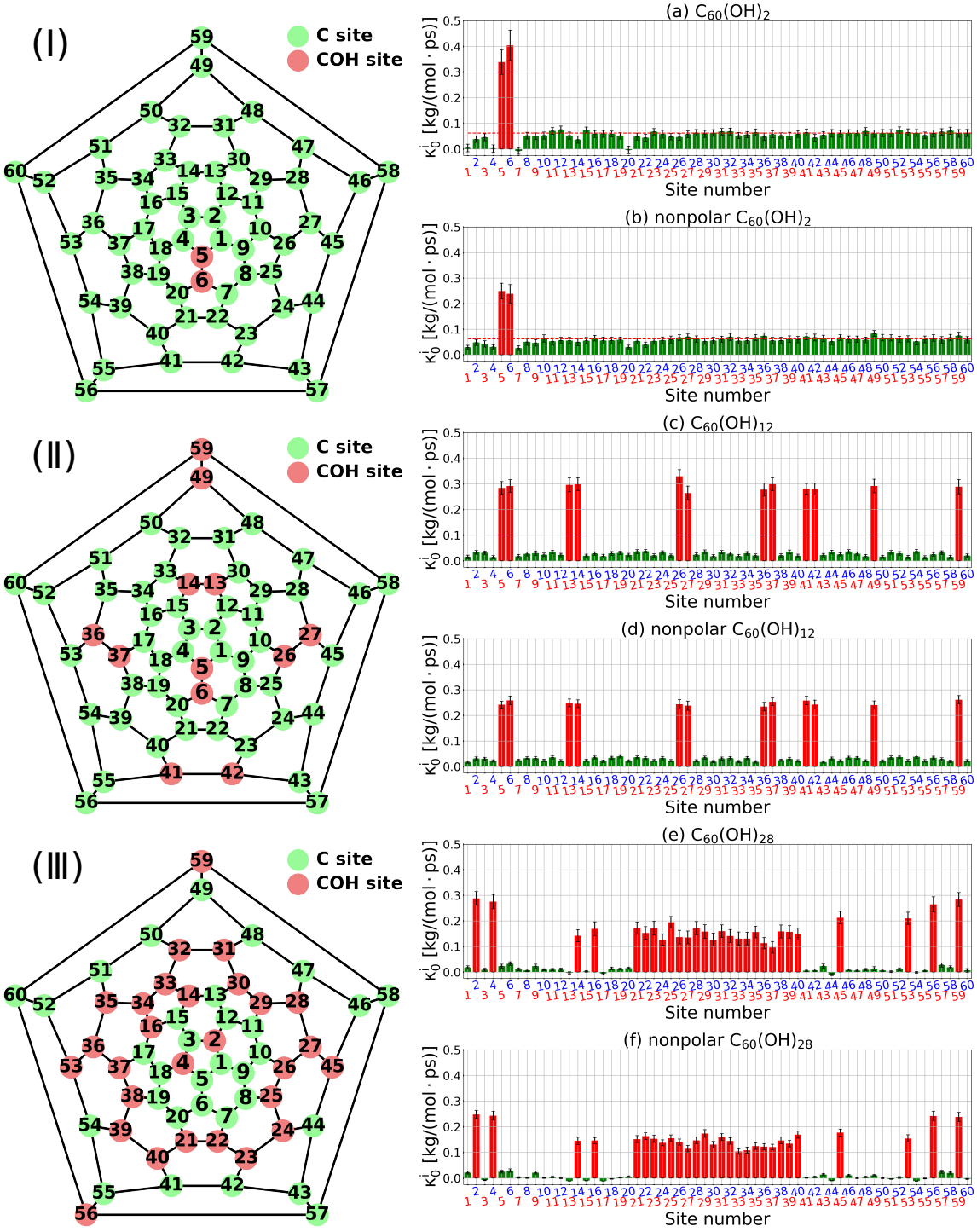

Figure 3.5: The contribution values for 60 sites (C or COH) are shown for (a) $C_{60}(OH)_2$, (b) nonpolar $C_{60}(OH)_2$, (c) $C_{60}(OH)_{12}$, (d) nonpolar $C_{60}(OH)_{12}$, (e) $C_{60}(OH)_{28}$, and (f) nonpolar $C_{60}(OH)_{28}$. The red bars and green bars represent the contribution values for the COH sites and C sites, respectively. Error bars denote the 95% confidence interval. The site numbers correspond to the numbering on the fullerene expanded maps in the left panels, where red and green circles represent the COH and C sites, respectively. In the left panels, from top to bottom: (I) $C_{60}(OH)_2$, (II) $C_{60}(OH)_{12}$, and (III) $C_{60}(OH)_{28}$. In (a) and (b), the red dashed line shows the friction coefficient per C atom on the fullerene, calculated by dividing the total friction coefficient of the fullerene by 60.

Each integrated contribution kernel plateaus at a constant value, representing the contribution value of each site [see Fig. 3.2 (right panels)]. These behaviors are consistent with the trends observed in the contribution kernels [see Fig. 3.2 (left panels)]. As the structural asymmetry increases, the relaxation times of the contribution kernels become longer. For polar and nonpolar $C_{60}(OH)_2$, the integrated contribution functions plateau around 300 ps, while other functions have shorter relaxation times, around 10 ps, which is comparable to that of the friction kernel of C_{60}. This slowing is likely due to the higher friction in the hemisphere containing COH sites compared to the opposite hemisphere. Consequently, a rotation caused by a significant drag force on one side seems to induce a flow around the molecule. In other words, this slowing is attributed to the asymmetric structure.

This assumption is supported by the results of the angular correlation function between the velocity and the molecular orientation, which is defined as:

$$C_{\mathrm{cov}}(t) = \frac{1}{n_{\mathrm{co}}} \sum_{i=1}^{n_{\mathrm{co}}} \left\{ \arccos\left(\left\langle \frac{\mathbf{v}(0)}{\|\mathbf{v}(0)\|} \cdot \frac{\mathbf{r}_{\mathrm{co}}^{i}(t)}{\|\mathbf{r}_{\mathrm{co}}^{i}(t)\|} \right\rangle \right) - 90^{\circ} \right\}, \tag{3.23}$$

where n_{co} is the total number of the carbon-oxygen bonds in the fullerenol, $\mathbf{r}_{\mathrm{co}}^{i}$ is the i-th carbon-oxygen bond vector, and $\|\cdot\|$ represents the norm. Here, $C_{\mathrm{cov}}(t) = 0^{\circ}$ indicates no correlation. Fig. 3.3 shows the results for $C_{60}(OH)_2$ (red), nonpolar $C_{60}(OH)_2$ (orange), $C_{60}(OH)_{12}$ (green), and $C_{60}(OH)_{28}$ (blue). The functions for $C_{60}(OH)_{12}$ and $C_{60}(OH)_{28}$ quickly converge to zero, whereas those for polar and nonpolar $C_{60}(OH)_2$ decay to zero around 300 ps and 200 ps, respectively. These results are consistent with the relaxation times observed in the contribution kernels. Thus, the long relaxation time in the contribution kernel of $C_{60}(OH)_2$ is attributed to its asymmetric structure.

The summation of the contribution kernels must equal the friction kernel. This requirement also applies to the summation of integrated contribution kernels and the integrated friction kernels. Fig. 3.4 shows the summation of integrated contribution kernels (black dashed line), the integrated friction kernel (blue solid line), and $k_{\mathrm{B}}T/\int C^{vv}(\tau)d\tau$ (red solid line) for each solute. While the summation of integrated contribution kernels exhibits slight deviations from the integrated friction kernels, these deviations are within the numerical error. These results confirm that the summation of contribution values matches the friction coefficient within numerical accuracy. Additionally, the integrated friction kernel and the summation of integrated contribution kernels converge to the value of $k_{\mathrm{B}}T/\int C^{vv}(\tau)d\tau$ at large t. Notably, $k_{\mathrm{B}}T/\int C^{vv}(\tau)d\tau$ approaches $\gamma = k_{\mathrm{B}}T/D$, as defined by the Einstein-Smoluchowski (ES) equation. Therefore, the friction coefficients and the summation of contribution values satisfy the ES equation.

Next, I show the contribution values for the 60 C or COH sites for each solute in Fig. 3.5

(right panels). These values were calculated from the integrated contribution kernels and scaled using Eq. (3.21). The red and green bars represent the contribution values of the COH and C sites, respectively. The site numbers correspond to those on the fullerene expanded maps shown in Fig. 3.5 (left panels), where red and green circles represent the COH and C sites, respectively.

Let us first examine the case of $C_{60}(OH)_2$. In Fig. 3.5(a), the contribution values for C sites that are not next to COH sites are almost the same. These values are comparable to the friction coefficient per one C atom on fullerene, i.e., the friction coefficient of fullerene divided by 60, which is shown as the red dashed line in Fig. 3.5(a). A similar trend is observed for nonpolar $C_{60}(OH)_2$ (see the red dashed line in Fig. 3.5(b)). These results confirm that the contributions from C sites have been correctly accounted for. In contrast, the contribution values of COH sites and their adjacent C sites deviate from this constant value. The contribution values of COH sites are significantly larger, while those of adjacent C sites are smaller, compared to non-adjacent C sites or C sites in fullerene.

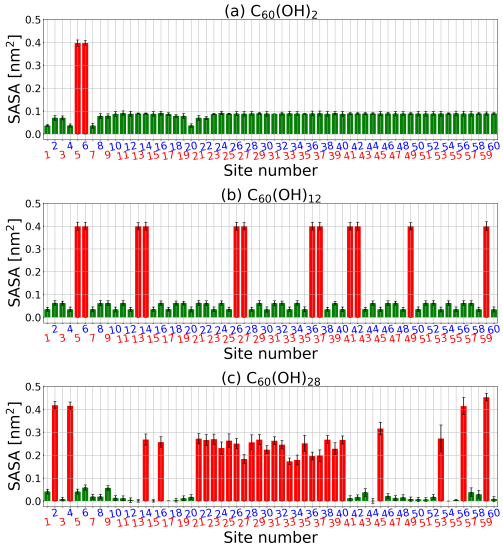

Figure 3.6: The solvent-accessible surface area (SASA) of 60 sites in (a) $C_{60}(OH)_2$, (b) $C_{60}(OH)_{12}$, and (c) $C_{60}(OH)_{28}$. The red and green bars represent the SASA of the COH and C site, respectively. The error bars denote the 95% confidence interval. The site numbers correspond to the numbering on the fullerene expanded maps in the Fig. 3.5 (left panels).

The solvent-accessible surface area (SASA) provides a rough explanation for the observed differences. In Fig. 3.6(a), the SASA of 60 sites in $C_{60}OH_2$ is shown. SASAs were calculated using the GROMACS software (gmx sasa) [75, 83] with Bondi's van der Waals radii [84]. Here, the solvent molecules were approximated as hard spheres with a radius of 0.14 nm. The results indicate a strong correlation between friction (contribution values) and the SASA: as the SASA increases, the friction also increases. The magnitude of the SASA is related to surface roughness.

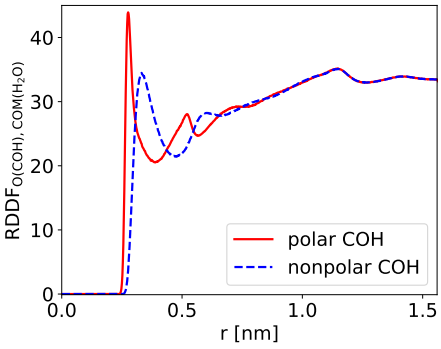

Figure 3.7: The radial density distribution functions between the center of oxygen atom in the COH site and the mass center of water are shown for polar (red solid line) and nonpolar (blue dashed line) $C_{60}(OH)_2$, respectively.

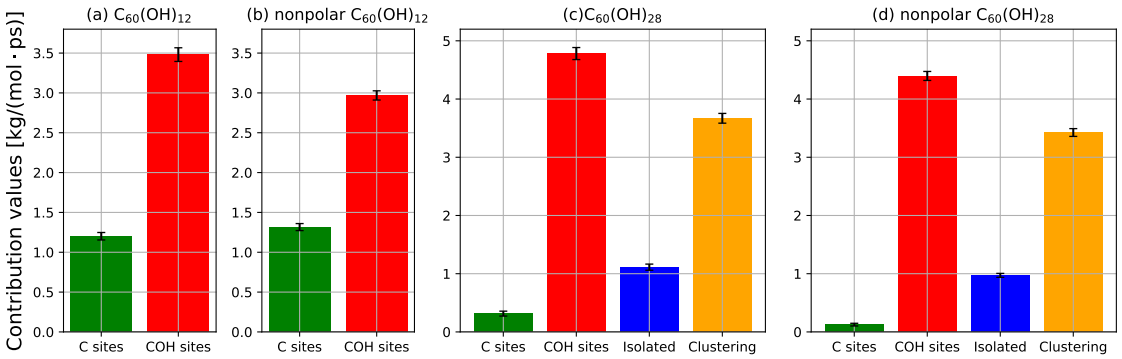

Figure 3.8: The total contribution values for specific sites in (a) $C_{60}(OH)_{12}$, (b) nonpolar $C_{60}(OH)_{12}$, (c) $C_{60}(OH)_{28}$ and (d) nonpolar $C_{60}(OH)_{28}$ are shown. The green bar, red bar, blue bar, and orange bar represent the values for C sites, COH sites, isolated COH sites and clustering COH sites, respectively. The error bars denote the 95% confidence interval.

In other words, these results indicate that surface roughness can significantly affect the friction or diffusion coefficient.

The differences in contribution values cannot be fully explained by the SASA. For instance, the COH sites exhibit contribution values approximately 1.5 times larger than those of the nonpolar COH sites [see Figs. 3.5(a) and (b)]. This is unexpected because the SASA of the polar COH sites is predicted to be smaller due to hydrogen bonding between the charged sites. As shown in Fig. 3.7, the radial density distribution function (RDDF) between the oxygen atom in the COH site and the mass center of water shows that water molecules are located closer to the polar COH sites compared to the nonpolar ones. These results suggest that the differences in contribution values of the COH sites cannot be attributed solely to their surface roughness but are likely due to the effects of hydrogen bonding.

Also, in the cases of $C_{60}(OH)_{12}$ and $C_{60}(OH)_{28}$, similar trends in the contribution values of

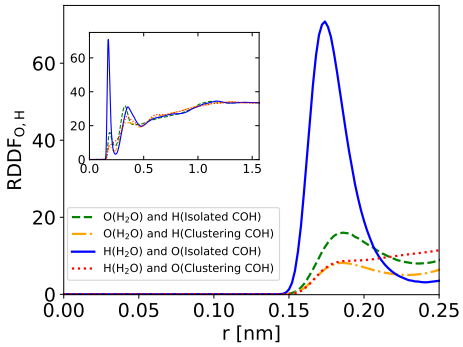

Figure 3.9: The radial density distribution functions (RDDF) for the oxygen of the COH site interacting with hydrogen of water for the clustering COH sites (red dotted line) and the isolated COH sites (blue solid line) are shown. Similarly, the RDDFs for the oxygen of water interacting with the oxygen of the COH site for the clustering COH sites (orange dash-dotted line) and the isolated COH sites (green dashed line) are shown. The RDDFs in the range $r = 0$ – 0.25 nm are mainly shown. The inset shows the entire RDDF range.

the COH sites were observed. To clarify the differences, the total contribution values for the C and COH sites in $C_{60}(OH)_{12}$ and $C_{60}(OH)_{28}$ were calculated. Fig. 3.8 shows the results, where the green bars represent the total contribution values of the C sites, and the red bars represent those of the COH sites. For both $C_{60}(OH)_{12}$ and $C_{60}(OH)_{28}$, the contribution values of the COH sites exceed those of the nonpolarized COH sites. To investigate this trend, the number of hydrogen bonds was counted based on a rough geometric criterion. Table 3.1 summarizes the average number of hydrogen bonds between the solutes and the solvent. The increases in the contribution values per hydrogen bond, determined based on the number of hydrogen bonds, are 0.095, 0.032, and 0.021 kg/(mol·ps) for $C_{60}(OH)_2$, $C_{60}(OH)_{12}$, and $C_{60}(OH)_{28}$, respectively. $C_{60}(OH)_2$ exhibits a significantly larger value than the other solutes (the reason will be addressed later). However, the values for $C_{60}(OH)_{12}$ and $C_{60}(OH)_{28}$ are comparable, indicating that the increase in the contribution value for the COH sites is approximately proportional to the number of hydrogen bonds. These results suggest that hydrogen bonds can also affect the friction or diffusion coefficient.

To further investigate the role of hydrogen bonds, I classified the COH sites on $C_{60}(OH)_{28}$ into two groups: clustered COH sites (site numbers: 14, 16, 21 to 40, 45, 53) and isolated COH

Table 3.1: The average number of hydrogen bonds (HB) between solute and solvent.

Solute	Number of HB
$C_{60}(OH)_2$	2.7
$C_{60}(OH)_{12}$	16.2
$C_{60}(OH)_{28}$	18.2

sites (site numbers: 2, 4, 56, 59). Here, the site numbers correspond to the numbering on the fullerene expanded map [see Fig. 3.5].

Fig. 3.8(c) and (d) also show the total contribution values for the clustered and isolated COH sites in both polarized and nonpolar $C_{60}(OH)_{28}$. The value of polar $C_{60}(OH)_{28}$ is higher than that of the nonpolar one. A similar trend is observed for the clustered sites (orange bars). However, the increase in contribution value per site for the clustered COH sites (1.0×10^{-2} kg/[mol·ps]) is significantly smaller than that for the isolated COH sites (3.5×10^{-2} kg/[mol·ps]) or for the COH sites in $C_{60}(OH)_{12}$ (4.3×10^{-2} kg/[mol·ps]).

The difference between the clustered and isolated COH sites can be explained by examining their hydration structures. Fig. 3.9 shows the RDDF between the COH sites and water molecules. The RDDFs for the hydrogen of the COH sites interacting with the oxygen of water are represented by orange dash-dotted lines for the clustered COH sites and green dashed lines for the isolated COH sites. Similarly, the RDDFs for the oxygen of the COH sites interacting with the hydrogen of water are represented by red dotted lines for the clustered COH sites and blue solid lines for the isolated COH sites. The height of the first peak in these RDDFs is related to the likelihood of hydrogen bond formation between the water and the COH sites.

In Fig. 3.9, the RDDFs for the clustered COH sites (red and orange lines) exhibit lower first peaks compared to those for the isolated COH sites (blue and green lines). This significant difference in the first peak is prominently observed in the RDDFs between the hydrogen atoms of the COH sites and the oxygen atoms of water molecules. These results indicate that the reduced accessible surface for solvent water molecules, combined with the formation of internal hydrogen bonds among the clustered COH sites, leads to a decrease in the number of hydrogen bonds formed between water molecules and the clustered COH sites. Similar discussions about internal hydrogen bonds have been presented by Chaban and Fileti [63]. The above analysis supports their discussions. In contrast, isolated COH sites are more effective than clustered COH sites in increasing the contribution value for friction due to hydrogen bonds formed between the COH sites and water molecules. These findings suggest that the number of hydrogen bonds influences the contribution value for friction and further support the argument by Terazima and co-workers regarding the reduction of diffusion coefficients caused by hydrogen bond formation between proteins and water molecules [1, 2, 3, 4, 5, 6, 7, 8, 9].

In the discussion above, I demonstrated that the presence of OH groups increases the contribution values of those sites. However, COH sites also influence the contributions of other nearby sites, particularly those adjacent to the COH site. For instance, in the case of $C_{60}(OH)_2$, the contribution values of C sites (sites 1, 4, 7, and 20) located near the COH site decrease (see Fig. 3.5(a)). Conversely, when the COH sites are nonpolar, the contribution values of the same C sites (sites 1, 4, 7, and 20) show a slight increase (see Fig. 3.5(b)). In both cases, It can be

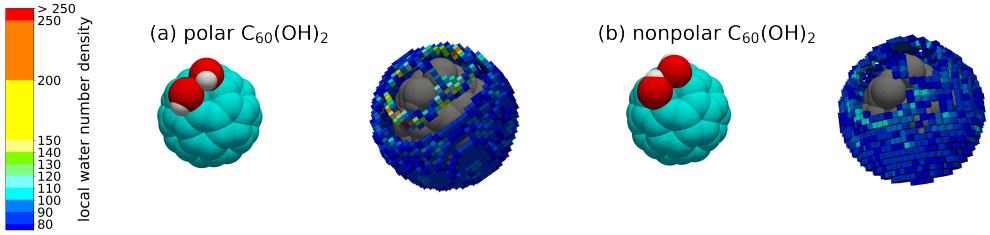

Figure 3.10: The spatial density distribution functions (SDDF) of water around (a) polar and (b) nonpolar $C_{60}(OH)_2$ are shown using a 0.05 nm cubic grid resolution. In (a) and (b), the images on the left show the camera angles and the images on the right show the SDDF grids with the local water number density larger than 75. Each grid is classified by the local water number density, and colored according to the left color bar.

explained by the reduction in the SASA, which is consistent with the earlier discussion on the relation between SASAs and frictions. Furthermore, the difference between polar and nonpolar COH sites may be attributable to the difference in hydration structure around the COH site.

Thus, to examine the hydration structure around the COH sites, I calculated the spatial density distribution function (SDDF) of water molecules around $C_{60}(OH)_2$. Figure 3.10 shows the SDDF for (a) polar and (b) nonpolar $C_{60}(OH)_2$, using a cubic grid resolution of 0.05 nm. Here, grids with a local number density of water larger than 75 are visualized. In nonpolar $C_{60}(OH)_2$, high-density grids appear around the neck of the COH site. In contrast, in polar $C60(OH)_2$, these high-density grids are absent in the neck region but appear near the hydrogen-bonding sites of the polar OH groups. Those results indicate the attraction of the polar OH groups reduces the accessibility of water molecules to nearby C sites. Consequently, the contribution values of C sites next to the polar OH groups are lower than those of nonpolar $C_{60}(OH)_2$.

The COH site in $C_{60}(OH)_{28}$ also reduces the contribution values for the C site compared to the C site in fullerene. However, the influence of polar COH sites is weaker compared to those of the other fullerenols. This suggests that the relative impact of these effects can reverse as the number of OH groups increases. The reversal is clearly shown by the green bars in Fig. 3.8(c) and (d). This phenomenon is also attributed to differences in the hydration structure. In Figure 3.11, the RDDFs between the C site and the mass center of water for polar (red solid line) and nonpolar (blue solid line) cases of each solute are compared. For $C_{60}(OH)_{28}$, the RDDF in the polar case remains consistently higher than that in the nonpolar case up to about 0.8 nm [see Fig. 3.11(c)]. In contrast, this behavior is not observed for the other fullerenols [see Figs. 3.11(a) and (b)]. Therefore, the increased contribution value of the C site in polar $C_{60}(OH)_{28}$ would be caused by the higher water density around the C site. This increase in water density can be attributed to charge polarization at the C site, which is induced by the attachment of multiple OH groups. While this effect is relatively minor for small molecules,

it could be significant for larger graphene structures, such as nanotubes, due to cumulative effects.

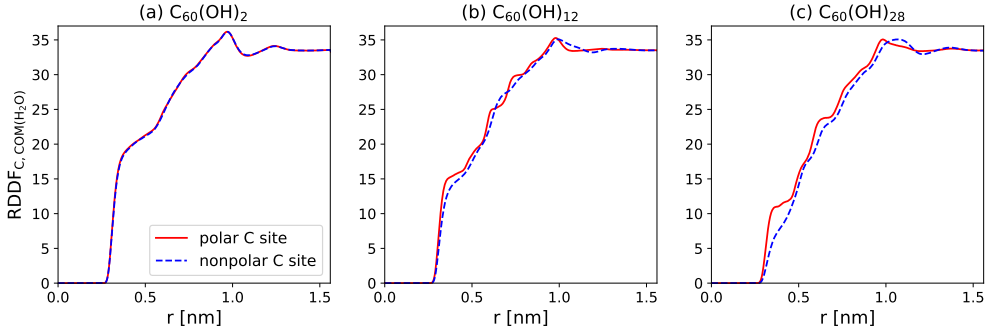

Figure 3.11: The radial density distribution functions (RDDF) between the C site and the mass center of water in both the polar (red solid line) and nonpolar (blue dashed line) are shown for (a) $C_{60}(OH)_2$, (b) $C_{60}(OH)_{12}$ and (c) $C_{60}(OH)_{28}$, respectively.

Now, I will discuss the diffusion properties of the whole molecule. The friction coefficients, diffusion coefficients, and hydrodynamic radii for fullerenols are shown in Fig. 3.12, where the closed and open circles represent the values for the polar and nonpolar solutes, respectively. These diffusion properties were corrected using the Yeh–Hummer method [Eq. (2.25)] to estimate the values for the infinite system. Here, the finite-size corrections have large contributions, such as about 50 %, to the diffusion properties of infinite systems. However, for the systems with the same number of water molecules, the corrections are comparable and do not change the order of relative magnitude. Additionally, under periodic boundary conditions, the conservation of the angular momentum of the system does not hold. It might affect the translational diffusion of $C_{60}(OH)_2$ due to the interdependence of rotational and translational diffusions in asymmetric structures. Thus, it should be noted that I assume this effect is not significant in the following discussions.

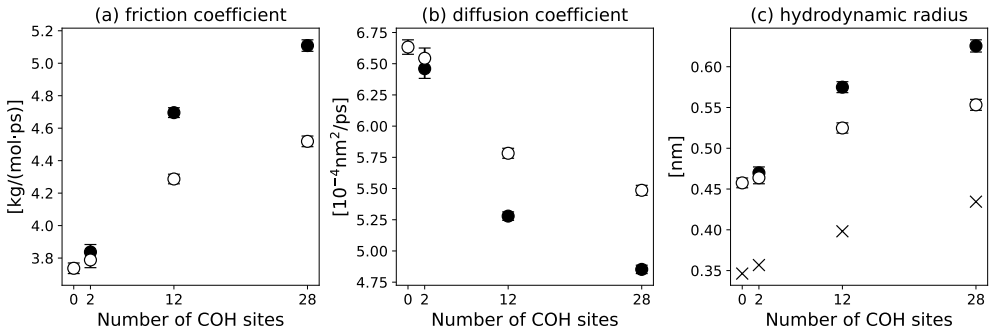

Figure 3.12: The diffusion properties of fullerene and fullerenols are shown as a function of the number of COH sites, including (a) friction coefficients, (b) diffusion coefficients, and (c) hydrodynamic radii. The closed and opened circles represent the results for the polar and nonpolar solute, respectively. In (c), the crosses represent the gyration radii. The error bars denote the 95% confidence interval.

In Figs. 3.12 (a), (b), and (c), as the number of OH groups increases, the friction coefficients and hydration radii increase, while the diffusion coefficients decrease. These trends are consistent with each other, as described by the SE law and the Einstein-Smoluchowski equation. Particularly, the differences in diffusion properties are pronounced for polar solutes. For instance, the diffusion coefficient of polar $C_{60}(OH)_{28}$ is about 1.13 times smaller than that of its nonpolar counterpart, where the difference is larger than the numerical errors. Based on the above discussions, these differences can be attributed to hydrogen bondings between the molecular surface and water, as well as surface roughness.

The trends in the fullerenol diffusivity with an increasing number of OH groups observed in this study are qualitatively consistent with previous simulation studies [63, 64, 65], which also reported a decrease in diffusion coefficients as the number of OH groups increased. However, the magnitude of diffusivity changes in our results were smaller than those of the 5 to 6-fold decrease reported in those studies. This relatively smaller change might arise from differences in the MD simulation setups, including force field, cell size, barostats, thermostat, and simulation time length. Direct comparison of these results with ours is not possible because diffusion coefficients depend on the simulation cell size [41] and barostat [85]. In the present study, cell-size corrections were applied [41], which is an advantage of this study.

As explained in Section 1.2, the relationship between the hydrodynamic and gyration radii strongly depends on the molecular shape, which is characterized by the constant ρ as:

$$R_{\mathrm{g}} = \rho \cdot R_{\mathrm{h}}, \tag{3.24}$$

where ρ is predicted to be $\sqrt{3/5} \approx 0.775$ for spherical molecules [25] [see Eq. (1.15)]. In Fig. 3.12(c), the gyration radii are also shown as crosses, plotted only for polar solutes because

they are identical to those for nonpolar solutes. The constant ρ ranges from 0.693 to 0.785, consistent with the spherical nature of fullerene and fullerenols. Thus, the hydrodynamic radii are influenced by the molecular radii. Specifically, the hydrodynamic radii of nonpolar solutes exhibit behaviors almost identical to their gyration radii. For example, the radius ratio between fullerene and nonpolar $C_{60}(OH)_{28}$ is about 1.2 for both hydrodynamic and gyration radii. In contrast, the radius ratio for fullerene and polar $C_{60}(OH)_{28}$ is about 1.37, which is larger than that of the gyration radius. Similar trends are observed for $C_{60}(OH)_2$ and $C_{60}(OH)_{12}$. This indicates that the hydrodynamic radius accounts not only for the molecular radius but also for interactions between the molecular surface and water.

To quantify the increase in the friction coefficient per hydrogen bond, I divided the difference in friction coefficients between polar and nonpolar fullerenols by the number of hydrogen bonds (see Table 3.1). The results are 0.018, 0.025, and 0.035 kg/(mol·ps) for $C_{60}(OH)_2$, $C_{60}(OH)_{12}$, and $C_{60}(OH)_{28}$, respectively. The value for $C_{60}(OH)_{28}$ is higher than the others, which may be caused by strong interactions between C sites and water. Indeed, the corresponding value for $C_{60}(OH)_{28}$ in the discussion of contribution values per hydrogen bond was 0.021 kg/(mol·ps), which is comparable to the above results for the other fullerenols. These results indicate that friction coefficients are approximately proportional to the number of hydrogen bonds.

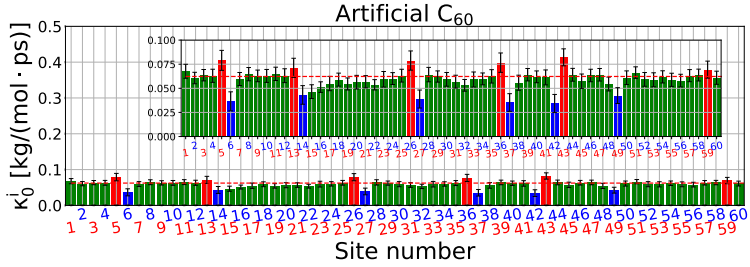

Figure 3.13: The contribution value of 60 sites on the artificial fullerene model is shown. The red bars (Sites 5, 13, 26, 36, 43, and 59) and the blue bars (Sites 6, 14, 27, 37, 42, and 49) represents the positively and negatively charged sites, respectively. The other bars (green bars) indicate the normal carbon sites with zero charge. The error bars denote the 95% confidence interval. The red dashed line indicates the friction coefficient per C atom on fullerene obtained by dividing the fullerene friction coefficient by 60. The insets enlarge the range between 0 and 0.1 kg/(mol· ps).

Finally, I will report on an artificial fullerene model with charged sites. Six positive charges were placed on Sites 5, 13, 26, 36, 43, and 59, and six negative charges on Sites 6, 14, 27, 37, 42, and 49. These site numbers correspond to those on the fullerene expanded maps (Fig. 3.5). Each positively charged site is adjacent to a negatively charged site, and the absolute value of each charge is 0.3031. This charge value was determined based on the bond moment of OH in SPC/E water [69].

Fig. 3.13 shows the contribution values of the 60 sites on the artificial fullerene model. The red bars (Sites 5, 13, 26, 36, 43, and 59) and blue bars (Sites 6, 14, 27, 37, 42, and 49) represent the positively and negatively charged sites, respectively. The green bars correspond to the normal C sites with zero charge. In Fig. 3.13, the green bars are identical to the friction coefficient per C atom on fullerene, calculated by dividing the friction coefficient of fullerene by 60 (see the red dashed line). However, the charged sites deviate from this value because of electrostatic interactions between water and the charged sites. Before this study, I assumed that the contribution values for the charged sites would increase similarly to those for fullerenol. However, the increase in contribution values for the positively charged sites was not significant. On the contrary, the contribution values for the negatively charged sites were smaller than those of the normal C sites (green bars and the red dashed line). Consequently, the friction coefficient for the artificial fullerene model (3.57 ± 0.03 kg/(mol·ps)) was smaller than that of fullerene (3.74 ± 0.03 kg/(mol $\cdot$ ps)).

The above results suggest that electrostatic interactions between water and charged sites do not always increase friction. One possible explanation is a decrease in surface roughness. Since the charges are buried inside the sphere, water molecules may not effectively interact with the charged sites. Specifically, the negatively charged sites and the positive charges of water are located at the center of the van der Waals sphere, which makes direct interactions difficult. In other words, surface roughness, including structural features on the surface, is important for interactions between charged sites and water molecules, as buried charges cannot increase the accessible surface area for water. Interestingly, the friction slightly decreases due to charges buried on the macromolecular surface. This may be because the buried OH groups disrupt the hydrogen-bonding network among water molecules, which leads to a reduction in the local viscosity near the macromolecular surface. Future studies are needed to clarify this effect.

3.5 Conclusion

In this chapter, I introduce a method to decompose the friction coefficient into the individual contributions of the specific sites, i.e., the contribution values κ^i. The decomposition method is formulated in the framework of the GLE [30]. MD simulations were performed using fullerenols [$C_{60}(OH)_2$, $C_{60}(OH)_{12}$, and $C_{60}(OH)_{28}$] in aqueous solution. The proposed method successfully decomposed the friction coefficients into their individual contributions, which indicates the diffusion properties depend on the surface characteristics of the fullerenols. The analysis revealed a strong correlation between the obtained contribution values and SASA, which suggests that surface roughness is an important factor for the diffusion coefficient. Significant differences in diffusivity were observed with increasing numbers of OH groups in fullerenols. However, the

changes in diffusion are larger in the case of fullerenols with polar OH groups. In particular, the diffusion coefficients of fullerenol with 28 OH groups decreased by a factor of about 1.37 over fullerene. This difference is comparable to diffusivity differences accompanied by conformation changes of protein induced by a light reaction observed in the experiments of Terazima and coworkers; for example, the LOV2 -linker is about 1.3 times [6]. Thus, the results support that the minor changes on the macromolecule surface can change the diffusion coefficients significantly. Comparisons between the results of polar and non-polar fullerenols revealed that electrostatic interactions also influence the contribution values. Then, the analysis of the number of hydrogen bonds showed that the effect of hydrogen bonds on the friction coefficient is proportional to the number of hydrogen bonds. The increased number of the non-polar OH group on the hydrodynamic radius is the same extent as that of the charge addition in the OH group. It indicates the effects of charge are comparable with that of surface roughness in the present results. However, the amount of hydrogen bonds between the OH group and the water depends on the level of exposure to the water. Even if the OH group has enough charges, the group buried on the fullerene surface cannot form hydrogen bonds with the water, and the effect on the hydrodynamic radius remains only slight. From these results, I conclude that it is important for the macromolecule diffusivity to consider the surface roughness, as well as the direct interaction between solvent and solute, such as hydrogen bonding.

Chapter 4

Summary

In this thesis, I construct a method using MD simulations to assess the solvation effects on the diffusion coefficient. I also analyzed the MD simulations of fullerene or fullerenol in an aqueous solution to assess the relation between diffusion coefficient and hydration. This study is summarized as follows.

In Chapter 2, MD simulations were performed to evaluate the system-size dependence of the diffusion coefficients. I focused on the higher-order correction term for the method proposed by Yeh and Hummer and applied it to correct the diffusion coefficient of protein obtained from MD simulation. Chignolin and β-lactoglobulin (BLG) were adopted as a protein model. The results show that the correction method based on the higher-order correction term enables the correction of diffusion coefficients for infinite systems, even using small simulation cells. The obtained value for BLG was 1.16×10^{-4} nm^2/ps. This value is comparable to 1.19×10^{-4} nm^2/ps expected from the experimental result[58, 59]. Therefore, the correction method is a powerful tool for the precise calculation of protein diffusion coefficients while reducing computational costs.

In Chapter 3, A method to decompose the friction coefficients into the contribution of individual components on the solute surface was proposed and applied to the MD trajectories of fullerenols [$C_{60}(OH)_2$, $C_{60}(OH)_{12}$, and $C_{60}(OH)_{28}$] in ambient water. The friction coefficient is inversely proportional to the diffusion coefficient. The MD results show that the diffusion coefficient decreases, while the friction coefficient increases, as the number of OH groups increases. By using the decomposition method, the individual components were obtained and analyzed. The contribution value is the contribution of each component on the solute surface to the friction coefficient. The results reveal the strong correlation between SASA and the contribution values and the proportional relationship between the number of hydrogen bonds and the friction coefficient. It indicates that both surface roughness and solute-solvent interactions are significant factors of the diffusion coefficients. As a consequence of those factors, the

difference rate in the diffusion coefficient between C_{60} and $C_{60}(OH)_{28}$ was about 1.37 times, which is comparable to about 1.3 times observed in the experiment result given by Terazima et al. for proteins. These results supported the discussion of Terazima et al. that minor changes on the macromolecule surface, such as the number change of the hydrogen bonds, can change the diffusion coefficient significantly.

Acknowledgements

I am grateful to thank Prof. Ryo Akiyama for guiding my study. I would like to thank Prof. Masahide Terazima (Kyoto University), Prof. Akira Yoshimori (Niigata University), Prof. Yuki Uematsu (Kyushu Institute of Technology), and Mr. Masaaki Nagao for their comments and discussion on my study, as well as, Prof. Yuka Nakamura (Niigata University) for her comments. I also thank the members of the Chemical Physics and Biophysics group at Kyushu University. Finally, I am grateful to my family for their warm support and encouragement.

I would also like to acknowledge the support of the Kyushu University Leading Human Resources Development Fellowship Program. The computations were performed using the Research Center for Computational Science, Okazaki, Japan (Project: 24-IMS-C083) and the Research Institute for Information Technology, Kyushu University, Japan, respectively.

Appendices

Appendix A – Fick law

The molecular concentrations at positions described by the Cartesian coordinates x, and $x+\delta x$ are $c(x)$ and $c(x+\delta x)$, respectively, where δx is an infinitesimal. At each position, half of the molecules move to the left, and half move to the right with equal probability. Then, the flow J through a unit cross section s per unit time τ is

$$J = -\frac{\{c(x+\delta x) - c(x)\} \cdot s\delta x}{2s \cdot \tau} = -\frac{\delta x^2}{2\tau}\frac{c(x+\delta x) - c(x)}{\delta x} = -D\frac{dc}{dx}, \tag{4.1}$$

where a constant D is constant: $\delta^2/(2\tau)$.

Appendix B – Why can we ignore the inertia term when viscosity is high?

We define the Laplace transform of the velocity as $\mathcal{L}(v) = \int_0^\infty v(t)e^{-st}dt$. Then, the Laplace transform of Eq. (1.3) can be expressed as follow:

$$m(s\mathcal{L}(v) - v_0) = -\gamma\mathcal{L}(v) + \frac{K_{\mathrm{tmp}}}{s}, \tag{4.2}$$

where $v_0 = v(0)$. From eq.(4.2), we obtain

$$\mathcal{L}(v) = \frac{K_{\mathrm{tmp}}/\gamma}{s} + \frac{v_0 - K_{\mathrm{tmp}}/\gamma}{s + \gamma/m}, \tag{4.3}$$

By performing the inverse Laplace transform on Eq. (4.3), we obtain the equation for the time variation of the velocity:

$$v(t) = K_{\mathrm{tmp}}/\gamma + (v_0 - K_{\mathrm{tmp}}/\gamma)e^{-\gamma/m \cdot t}. \tag{4.4}$$

When $\gamma \gg 0$, eq. (1.4) is obtained.

Appendix C – GLE for Brownian motion [30, 31]

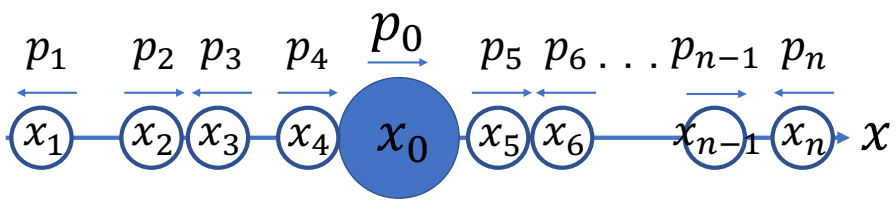

Figure 4.1: The one-dimensional Brownian motion.

We suppose a solute molecule (Brownian particle) surrounded by n solvent molecule in the one dimension as shown in Fig. 4.1. The positions and momentums at time t for the molecules are defined as $\mathbf{x(t)} = \{x_0(t), x_1(t), x_2(t) \cdots x_n(t)\}$ and $\mathbf{p(t)} = \{p_0(t), p_1(t), p_2(t) \cdots p_n(t)\}$, respectively. Here, $x_0(t)$ and $p_0(t)$ are assigned for solute, the others are assigned for solvents. Then, the Newton equation is

$$\frac{d\mathbf{p}}{dt} = -\frac{\partial V(\mathbf{x})}{\partial \mathbf{x}}, \tag{4.5}$$

where $V(\mathbf{x})$ is a potential of interactions among molecules. The total energy of system , i.e. Hamiltonian H, is

$$H = \sum \frac{\mathbf{p}^2}{2\mathbf{m}} + V(\mathbf{x}), \tag{4.6}$$

where $\mathbf{m}$ is a mass of molecules and $\sum$ denotes summation. Using the Hamiltonian H, the Newton equation can be rewritten to the Hamilton equation:

$$\frac{\partial H}{\partial \mathbf{p}} = \frac{d\mathbf{x}}{dt}, \frac{\partial H}{\partial \mathbf{x}} = -\frac{d\mathbf{p}}{dt} \tag{4.7}$$

If we set the origin of position to the center of solute molecule, i.e. $x_0 = 0$, a time differential of an arbitral physical quantity for solute molecule $A(\mathbf{x}, \mathbf{p})$ is expressed as:

$$\frac{dA}{dt} = \sum_{i=1}^{n} \left(\frac{\partial A}{\partial x_i}\frac{dx_i}{dt} + \frac{\partial A}{\partial p_i}\frac{dp_i}{dt} \right) \tag{4.8}$$

$$= \sum_{i=1}^{n} \left(\frac{\partial A}{\partial x_i}\frac{\partial H}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial H}{\partial x_i} \right), \tag{4.9}$$

where we use eq. (4.7). Here, we introduce an operator:

$$\mathrm{i}\hat{L} = \sum_{i=1}^{n} \left(\frac{\partial H}{\partial p_i}\frac{\partial}{\partial x_i} - \frac{\partial H}{\partial x_i}\frac{\partial}{\partial p_i} \right), \tag{4.10}$$

where $\mathrm{i} = \sqrt{-1}$. Then, eq. (4.9) is

$$\frac{dA}{dt} = \mathrm{i}\hat{L}A. \tag{4.11}$$

This equation is called Liouville equation which is also a sort of equation of motion. By solving this equation, we obtain

$$A(t) = \exp(\mathrm{i}\hat{L}t)A(0), \tag{4.12}$$

where $A(0)$ is the initial value. In the equilibrium system, the initial time can be set arbitrarily because of the stationarity, i.e, $\langle A(t)\rangle = \langle A(t+\tau)\rangle$. Here, the operator $\mathrm{i}\hat{L}$ gives the time derivative of a variable, while the operator $\exp(\mathrm{i}\hat{L}t)$ give the value of the variable as a function of time t. Using those operators, we obtain the equation of motion for the solute molecule as follow:

$$\frac{dp_0(t)}{dt} = \exp(\mathrm{i}\hat{L}t)\mathrm{i}\hat{L}p_0(0). \tag{4.13}$$

From now on, we obtain the GLE from eq. (4.13) using the projection operator method. we introduce two projection operators $\hat{P}$ and $\hat{Q}$ as follows:

$$\hat{P}A(t) = \frac{\langle A(t)\cdot p_0(0)\rangle}{\langle p_0(0)^2\rangle}p_0(0) \tag{4.14}$$

and

$$\hat{Q} = 1 - \hat{P}, \tag{4.15}$$

where $\langle\cdots\rangle$ denotes an ensemble average. Here, $\hat{P}$ is the projection operator which gives the value of a variable $A(t)$ projected onto $p_0(0)$, while $\hat{Q}$ is the projection operator which gives the value of A(t) orthogonal to $p_0(0)$. Using the two projection operator, we decompose the terms $\mathrm{i}\hat{L}q_0(0) = dq_0(0)/dt$ and $\exp(\mathrm{i}\hat{L}t)$, respectively.

First, we focus on the term $\mathrm{i}\hat{L}q_0(0)$. Using $(\hat{P}+\hat{Q}) = 1$, Eq. (4.13) is rewritten as follows:

$$\frac{dp_0(t)}{dt} = \exp(\mathrm{i}\hat{L}t)\hat{P}\mathrm{i}\hat{L}p_0(0) + \exp(\mathrm{i}\hat{L}t)\hat{Q}\mathrm{i}\hat{L}p_0(0). \tag{4.16}$$

Form eq. (4.14), the first term is

$$\exp(\mathrm{i}\hat{L}t)\hat{P}\mathrm{i}\hat{L}p_0(0) = \exp(\mathrm{i}\hat{L}t)\frac{\langle \frac{dp_0(0)}{dt}\cdot p_0(0)\rangle}{\langle p_0(0)^2\rangle}p_0(0). \tag{4.17}$$

Here, $\langle \frac{dp_0(0)}{dt} \cdot p_0(0) \rangle$=0 because of $\langle \frac{dp_0(0)}{dt} \cdot p_0(\tau) \rangle = -\langle \frac{dp_0(0)}{dt} \cdot p_0(-\tau) \rangle$, which is derived as follow:

$$\left\langle \frac{dp_0(0)}{dt} \cdot p_0(\tau) \right\rangle = \lim_{T\to\infty} \frac{1}{T} \int_0^T \frac{dp_0(t)}{dt} \cdot p_0(t+\tau) dt \tag{4.18}$$

$$= \lim_{T\to\infty} \frac{1}{T} \{ [p_0(t) \cdot p_0(t+\tau)]_0^T - \int_0^T p_0(t) \cdot \frac{dp_0(t+\tau)}{dt} dt \} \tag{4.19}$$

$$= - \lim_{T\to\infty} \frac{1}{T} \int_0^T p_0(t) \cdot \frac{dp_0(t+\tau)}{dt} dt \tag{4.20}$$

$$= - \lim_{T\to\infty} \frac{1}{T} \int_0^T \frac{dp_0(u)}{dt} \cdot p_0(u-\tau) du \tag{4.21}$$

$$= -\left\langle \frac{dp_0(0)}{dt} \cdot p_0(-\tau) \right\rangle, \tag{4.22}$$

where $u = t + \tau$. Therefore, the first term is zero in this case. the first term corresponds to the term iwA in Mori's paper[30], and becomes important in the case of $[p_0(t) \cdot p_0(t+\tau)]_0^T \neq 0$, such as non-equilibrium systems. However, it is unnecessary for our purpose.

Next, we focus on the term $\exp(\mathrm{i}\hat{L}t)$, Using $(\hat{P} + \hat{Q}) = 1$, Eq. (4.17) is rewritten as follow:

$$\frac{dp_0(t)}{dt} = \exp([\mathrm{i}\hat{L}\hat{P} + \mathrm{i}\hat{L}\hat{Q}]t)\hat{Q}\mathrm{i}\hat{L}p_0(0). \tag{4.23}$$

Here, the Laplace transform of $\exp([\hat{P}\mathrm{i}\hat{L} + \hat{Q}\mathrm{i}\hat{L}]t)$ is

$$\int_0^\infty \exp([\hat{P}\mathrm{i}\hat{L} + \hat{Q}\mathrm{i}\hat{L}]t) \exp(-st) dt = \frac{1}{s - (\hat{P}\mathrm{i}\hat{L} + \hat{Q}\mathrm{i}\hat{L})} \tag{4.24}$$

$$= \frac{1}{s - \hat{Q}\mathrm{i}\hat{L}} + \frac{1}{s - \hat{Q}\mathrm{i}\hat{L}} \frac{\hat{P}\mathrm{i}\hat{L}}{s - (\hat{P} + \hat{Q})\mathrm{i}\hat{L}} \tag{4.25}$$

$$= \frac{1}{s - \hat{Q}\mathrm{i}\hat{L}} + \frac{1}{s - \hat{Q}\mathrm{i}\hat{L}} \frac{\hat{P}\mathrm{i}\hat{L}}{s - \mathrm{i}\hat{L}}. \tag{4.26}$$

Then, the inverse Laplace transform of Eq.(4.26) gives

$$\exp([\hat{P}\mathrm{i}\hat{L} + \hat{Q}\mathrm{i}\hat{L}]t) = \exp(\hat{Q}\mathrm{i}\hat{L}) + \int_0^t \exp(\mathrm{i}\hat{L}t)\hat{P}\mathrm{i}\hat{L} \exp(\hat{Q}\mathrm{i}\hat{L}[t-\tau]) d\tau. \tag{4.27}$$

Therefore, Eq. (4.23) is rewritten as follow:

$$\frac{dp_0(t)}{dt} = \exp(\hat{Q}\mathrm{i}\hat{L}t)\hat{Q}\mathrm{i}\hat{L}p_0(0) + \int_0^t \exp(\mathrm{i}\hat{L}t)\hat{P}\mathrm{i}\hat{L}\exp(\hat{Q}\mathrm{i}\hat{L}[t-\tau])\hat{Q}\mathrm{i}\hat{L}p_0(0)d\tau \tag{4.28}$$

$$= F_\mathrm{R}(t) + \int_0^t \exp(\mathrm{i}\hat{L}t)\hat{P}\mathrm{i}\hat{L}F_\mathrm{R}(t-\tau)d\tau, \tag{4.29}$$

$$= F_\mathrm{R}(t) + \int_0^t \frac{1}{\langle p_0(0)^2\rangle}\langle \mathrm{i}\hat{L}F_\mathrm{R}(t-\tau)\cdot p_0(0)\rangle p_0(t)d\tau, \tag{4.30}$$

$$= F_\mathrm{R}(t) - \int_0^t \frac{1}{\langle p_0(0)^2\rangle}\langle F_\mathrm{R}(t-\tau)\cdot \mathrm{i}\hat{L}p_0(0)\rangle p_0(t)d\tau, \tag{4.31}$$

$$= F_\mathrm{R}(t) - \int_0^t \frac{1}{\langle p_0(0)^2\rangle}\langle F_\mathrm{R}(t-\tau)\cdot (\hat{P}+\hat{Q})\frac{dp_0(0)}{dt}\rangle p_0(t)d\tau, \tag{4.32}$$

$$= F_\mathrm{R}(t) - \int_0^t \frac{1}{\langle p_0(0)^2\rangle}\langle F_\mathrm{R}(t-\tau)\cdot \exp(\hat{Q}\mathrm{i}\hat{L}\cdot 0)\hat{Q}\mathrm{i}\hat{L}p_0(0)\rangle p_0(t)d\tau, \tag{4.33}$$

$$= F_\mathrm{R}(t) - \int_0^t \frac{1}{\langle p_0(0)^2\rangle}\langle F_\mathrm{R}(t-\tau)\cdot F_\mathrm{R}(0)\rangle p_0(t)d\tau, \tag{4.34}$$

where $F_\mathrm{R}(t) = \exp(\hat{Q}\mathrm{i}\hat{L}t)\hat{Q}\mathrm{i}\hat{L}p_0(0)$. Here, we use the relations in the equilibrium state: $\langle \mathrm{i}\hat{L}F_\mathrm{R}(t-\tau)\cdot p_0(0)\rangle = -\langle F_\mathrm{R}(t-\tau)\cdot \mathrm{i}\hat{L}p_0(0)\rangle$, $\hat{P}\frac{dp_0(0)}{dt} = 0$ (See Eqs. (4.18) $\sim$ (4.22)).

Finally, we obtain the GLE for Brownian motions as follows:

$$mf(t) = -\int_0^t \Gamma(t-\tau)\cdot v(\tau)d\tau + F_\mathrm{R}(t), \tag{4.35}$$

where m is the mass of the solute, $f(t)$ is the force acting on the solute, $v(t)$ is the velocity of the solute, and $\Gamma(\tau)$ is a friction kernel defined as

$$\Gamma(\tau) = \frac{1}{\langle mv(0)^2\rangle}\langle F_\mathrm{R}(0)\cdot F_\mathrm{R}(\tau)\rangle. \tag{4.36}$$

As we can see the above, the GLE is derived mathematically from the Newton equations, which is the basic form of the equation of motion, without any approximations. Therefore, the GLE is rigorous equation in the classical system. if we assume Markov process in which the time variation of the $v(t)$ is enough larger than the relaxation time of $\Gamma(t)$, the GLE is

transformed into the Langevin equation as follow:

$$mf(t) = -v(t)\int_0^\infty \Gamma(t-\tau)d\tau + F_{\mathrm{R}}(t), \tag{4.37}$$

$$= -v(t)\int_{-\infty}^0 \Gamma(t')dt' + F_{\mathrm{R}}(t), \tag{4.38}$$

$$= -v(t)\int_0^\infty \Gamma(t')dt' + F_{\mathrm{R}}(t), \tag{4.39}$$

$$= -\gamma v(t) + F_{\mathrm{R}}(t), \tag{4.40}$$

where $t' = t - \tau$, $\gamma = \int_0^\infty \Gamma(t')dt'$. Here, we use $\Gamma(t) = \Gamma(-t)$. Therefore, in the GLE, the convolution integral represents the drag force which takes account into memory effect, while $F_{\mathrm{R}}(t)$ is random force. Then, a friction coefficient is defined as:

$$\gamma = \int_0^\infty \Gamma(\tau)d\tau. \tag{4.41}$$

A random force is uncorrelated with the initial velocity:

$$\langle v(0) \cdot F_{\mathrm{R}}(\tau)\rangle = 0. \tag{4.42}$$

As we can see the definition of friction kernel, the random force and the friction kernel are connected to each other through the the fluctuation-dissipation theorem[32]:

$$\langle F_{\mathrm{R}}(0) \cdot F_{\mathrm{R}}(\tau)\rangle = k_{\mathrm{B}}T\Gamma(\tau). \tag{4.43}$$

Appendix D – Molecular dynamics (MD) simulation [44]

In classical system, the atomic motions can be described by the Newtonian equations. A position vector in a Cartesian coordinate system is expressed as follows:

$$\mathbf{r}_i = x_i\mathbf{e}_x + y_i\mathbf{e}_y + z_i\mathbf{e}_z, \tag{4.44}$$

where $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$ are the unit vectors in the directions of the respective axes. By defining the interactions among atoms, the potential energy $U(\mathbf{r}_N)$ can be calculated. When the kinetic energy is denoted as $K(\mathbf{v}_N)$, the total energy H is given by:

$$H(\mathbf{v}_N, \mathbf{r}_N) = K(\mathbf{v}_N) + U(\mathbf{r}_N). \tag{4.45}$$

The equation of motion for a atom i under a force $\mathbf{F}_i$ is:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i, \quad i = 1, 2, \ldots, N. \tag{4.46}$$

Here, the force can be calculated as the derivative of the system's potential energy, i.e,

$$\mathbf{F}_i = -\frac{\partial U(\mathbf{r}_N)}{\partial \mathbf{r}_i}. \tag{4.47}$$

Specifically, when the interaction between atom i and atom j is the two body interactions $\phi(r_{ij})$, the force from j to i can be calculated by:

$$\mathbf{F}_{ij} = -\frac{\partial \phi(\mathbf{r}_N)}{\partial \mathbf{r}_{ij}} \frac{\mathbf{r}_i - \mathbf{r}_j}{r_{ij}}, \tag{4.48}$$

where $\phi(r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$.

In molecular dynamics (MD) simulations, the Newtonian equations of motion expressed in Eq. (4.46) are numerically integrated using finite difference methods to approximate particle trajectories. Among several finite difference methods, the Verlet method is mainly used in MD simulations for solutions.

Now, I discribe the Verlet method. First, by expanding the position vector $\mathbf{r}_i$ at $t \pm \Delta t$ using the Taylor series, we obtain:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t) + \frac{\Delta t^2}{2} \frac{\mathbf{F}_i(t)}{m} + O(\Delta t^3), \tag{4.49}$$

$$\mathbf{r}_i(t - \Delta t) = \mathbf{r}_i(t) - \Delta t \mathbf{v}_i(t) + \frac{\Delta t^2}{2} \frac{\mathbf{F}_i(t)}{m} + O(\Delta t^3). \tag{4.50}$$

Adding and subtracting these two equations gives:

$$\mathbf{r}_i(t + \Delta t) + \mathbf{r}_i(t - \Delta t) = 2\mathbf{r}_i(t) + \Delta t^2 \frac{\mathbf{F}_i(t)}{m} + O(\Delta t^4), \tag{4.51}$$

$$\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t - \Delta t) = 2\Delta t \mathbf{v}_i(t) + O(\Delta t^3). \tag{4.52}$$

From these equations, the position at time $t + \Delta t$ and the velocity at time t are given as:

$$\mathbf{r}_i(t + \Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \Delta t^2 \frac{\mathbf{F}_i(t)}{m} + O(\Delta t^4), \tag{4.53}$$

$$\mathbf{v}_i(t) = \frac{1}{2\Delta t} \{\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t - \Delta t)\} + O(\Delta t^2). \tag{4.54}$$

By preparing the initial conditions of positions, the particle positions and velocities can be calculated using these equations. Several improved versions of the Verlet method exist, one of

which is the leapfrog method. The leapfrog method uses the expansion of the velocity $\mathbf{v}_i$ at $t \pm \Delta t/2$:

$$\mathbf{v}_i(t + \Delta t/2) = \mathbf{v}_i(t) + \frac{\Delta t}{2}\frac{\mathbf{F}_i(t)}{m} + O(\Delta t^2), \tag{4.55}$$

$$\mathbf{v}_i(t - \Delta t/2) = \mathbf{v}_i(t) - \frac{\Delta t}{2}\frac{\mathbf{F}_i(t)}{m} + O(\Delta t^2). \tag{4.56}$$

Note that the $O(\Delta t^2)$ term in the velocity expansion corresponds to $O(\Delta t^3)$ for the position expansion, and thus it can be neglected. Rewriting these equations, we obtain:

$$\mathbf{v}_i(t \pm \Delta t/2) - \mathbf{v}_i(t) = \pm\frac{\Delta t}{2}\frac{\mathbf{F}_i(t)}{m}. \tag{4.57}$$

Substituting this result into Eq. (4.53), we derive the following update equations:

$$\mathbf{v}_i(t + \Delta t/2) = \mathbf{v}_i(t - \Delta t/2) + \Delta t\frac{\mathbf{F}_i(t)}{m}, \tag{4.58}$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t\mathbf{v}_i(t + \Delta t/2). \tag{4.59}$$

In comparison to Eqs. (4.53) and (4.54), where the velocity is not explicitly used in position updates, the leapfrog method uses the velocity directly. This feature makes the leapfrog method effective for controlling temperature. Here, the velocity is staggered by a half-step relative to the position, which is the origin of the name 'leapfrog'. If the velocity at integer time steps t is required, it can be obtained by averaging the half-step velocities using Eq. (4.57):

$$\mathbf{v}_i(t) = \frac{1}{2}\left\{\mathbf{v}_i(t + \Delta t/2) + \mathbf{v}_i(t - \Delta t/2)\right\}. \tag{4.60}$$

Here, let us consider an example of MD simulation for a 2D fluid where the pair potential in Eq. (4.48) is set to the Lennard-Jones potential:

$$\phi(r_{ij}) = 4\epsilon\left(\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right)$$

(see Fig. 4.2). After providing the initial positions and velocities, the MD simulation can be performed by repeating the following processes (process numbers correspond to those in Fig. 4.2):

1. Compute the forces based on Eq. (4.48) using the particle positions (structure).

2. Update the velocities using the calculated forces based on Eq. (4.58).

3. Update the positions using the velocities based on Eq. (4.59). Here, when assuming bulk fluids, periodic boundary conditions (PBC) are applied. In PBC, if a particle crosses the simulation cell boundary, it is reallocated at the opposite boundary.

By repeating this series of steps, time-dependent data on particle positions, velocities, and forces can be obtained.

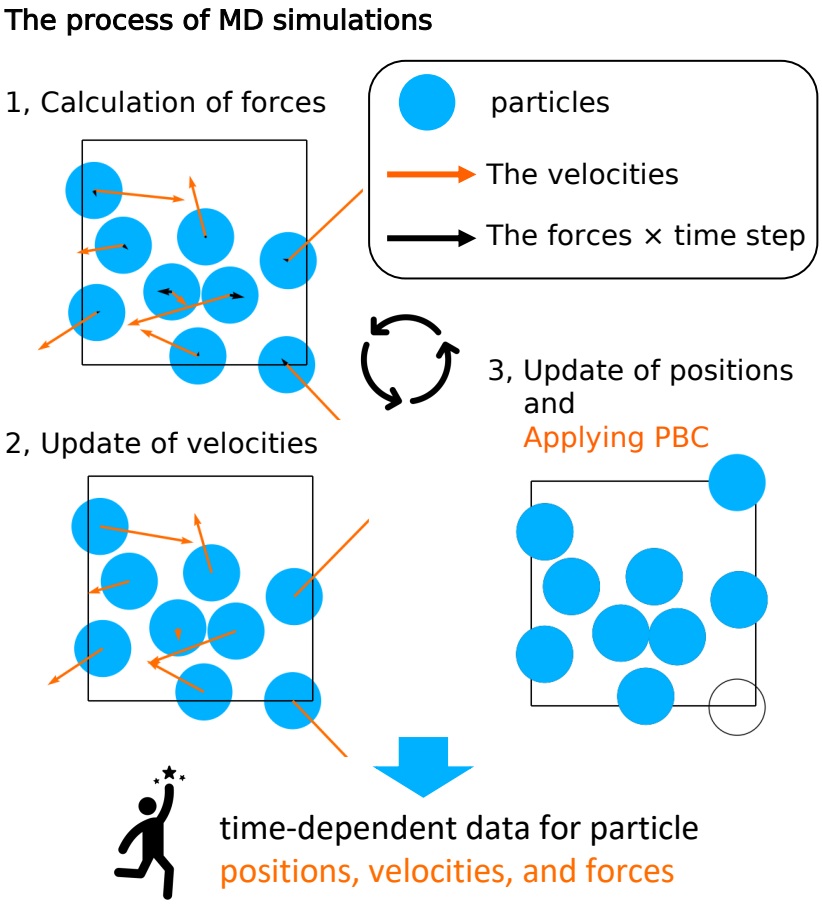

Figure 4.2: An example of MD simulation.

Appendix E – Belly simulations

1. **Acquisition of the target solute conformations:** $[\mathbf{q}_0(0), \mathbf{q}_0(1), \ldots, \mathbf{q}_0(n)]$ are obtained from the original MD trajectory. The original MD trajectory was generated during production runs under the standard NVT ensemble.

2. **Equilibration of the belly simulation:** The solute coordinates are frozen at the target solute conformation $\mathbf{q}(0)$, while the solvent molecules equilibrate.

3. **Production run of the belly simulation:** The solute coordinates remain frozen at the target solute conformation $\mathbf{q}(0)$ in the first case, while the solvent moves freely for n_{prd} steps. The forces acting on each atom from the solvents are averaged to calculate f^i_{ano}. To reduce noise in f^i_{ano}, the obtained forces are corrected using the following equation:

$$\sum_{i=0}^{N} f^i_{\mathrm{ano}} = \sum_{i=0}^{N} \left[f^i_{\mathrm{ano}} - \frac{m^i_{\mathrm{a}}}{m} f_{\mathrm{sum}} \right],$$

where m^i_{a} is the mass of the i-th site (or atom), and f_{sum} is the sum of the obtained forces: $f_{\mathrm{sum}} = \sum_{i=0}^{N} f^i_{\mathrm{ano,obtained}}$. This correction does not alter the essential nature of $f^i_{\mathrm{IN,ano}}(\mathbf{q}_0)$.

4. **Quasistatic steps:** The solute coordinates are slowly moved to the next target solute conformation, $\mathbf{q}(1)$ in the first case, over n_{step} steps using the following equation:

$$\mathbf{X}^j = \frac{(n_{\mathrm{step}} - j)\mathbf{q}_{\mathrm{pre}} + j \cdot \mathbf{q}_{\mathrm{next}}}{n_{\mathrm{step}}},$$

where $\mathbf{X}^j$, $\mathbf{q}_{\mathrm{pre}}$, and $\mathbf{q}_{\mathrm{next}}$ represent the coordinates of the j-th solute atom, the previous target solute conformation, and the next target solute conformation, respectively. The n_{step} is determined as follows:

$$n_{\mathrm{step}} = \max\left(\mathrm{disp}(\mathbf{q}_{\mathrm{next}}, \mathbf{q}_{\mathrm{pre}})\right) \cdot n_{\mathrm{buffer}},$$

where max and disp are operators used to obtain the maximum value and the displacement of each atom between the previous and next solute conformations, respectively. This ensures that the solute atom displacement per step is less than $1/n_{\mathrm{buffer}}$ nm. During these steps, the solvent molecules move freely.

5. **Iteration:** Steps 3 and 4 are repeated to calculate $f^i_{\mathrm{ano}}(\mathbf{q}(k))$ at each k-th target solute conformation $\mathbf{q}(k)$.

In step 1, each structure was superposed upon that of the previous target solute conformation to minimize the translational and rotational motion. In step 3, the obtained forces can be restored to their original orientation by projecting them onto the solute coordinate in the original MD trajectory. In step 4, the quasistatic process was assumed. Similar procedures have been used in the studies of mean force dynamics (MFD) [86, 87, 88, 89]. The thermodynamic conditions were the same as those of the original MD simulation. The correction for the center of mass translational velocity, which is enabled by default in the GROMACS 2021.4 software, was turned off. I used $n_{\mathrm{prd}} = 100$ and $n_{\mathrm{buffer}} = 5000$. To calculate steps 3–5, I prepared a modified version

of GROMACS 2021.4 [75]. The code was modified to conduct steps 3–5 continuously. The modified code is available at `https://github.com/TomoyaIwashita/gromacs2021_belly`. In Fig. 4.3, I show schematic drawings for each step, where the solute is represented by a 2-dimensional linear four-atomic molecule instead of fullerenols.

After obtaining the values $\{f^i_{\mathrm{ano}}(\mathbf{q}_0(0)), f^i_{\mathrm{ano}}(\mathbf{q}_0(1)), \cdots, f^i_{\mathrm{ano}}(\mathbf{q}_0(n))\}$, I can calculate $C^{f(\mathrm{EX,i})\mathrm{v}}(t)$ and $C^{f(\mathrm{EX,i})\mathrm{F}}(t)$ using Eqs. (3.11) and (3.12). In this thesis, the interval of the restart points on $(f^i - f^i_{\mathrm{ano}})$ was set to 0.1 ps.

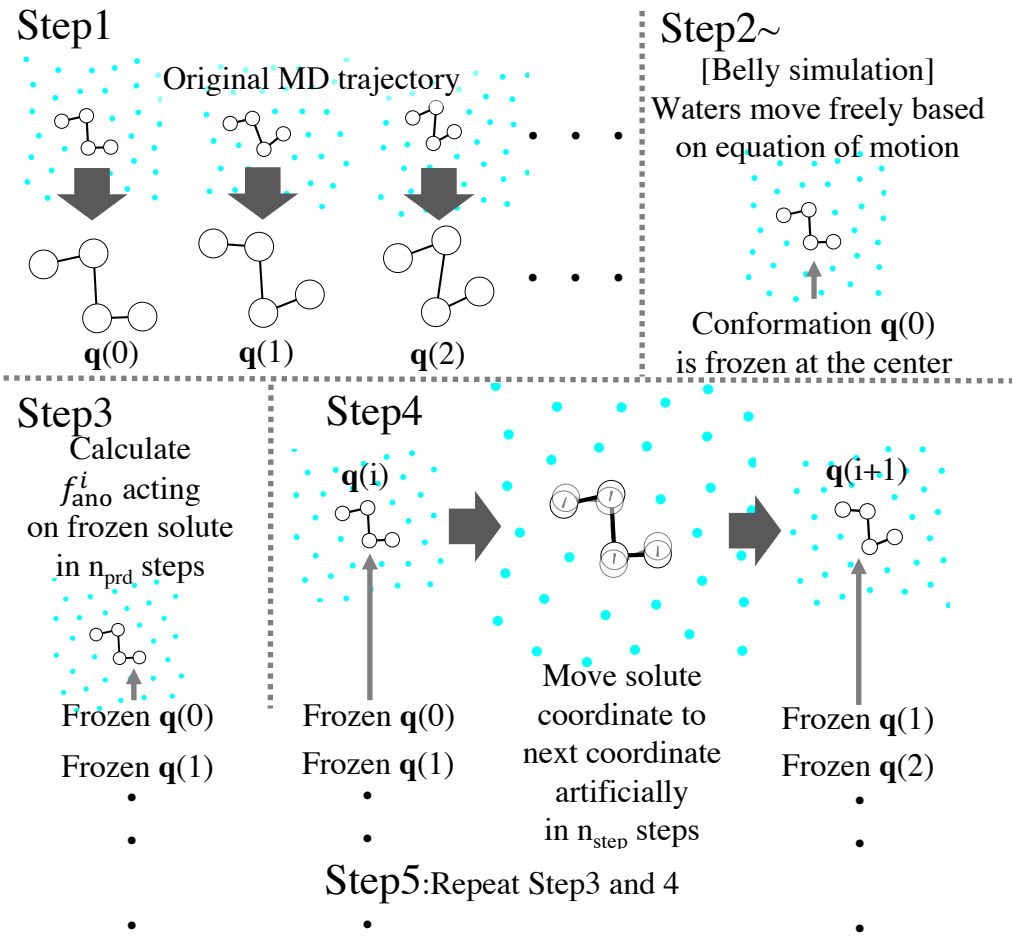

Figure 4.3: Schematic drawings for steps 1 $\sim$ 5. The linear four-atomic molecule is the solute, and the small circles denote the solvent molecules.

References

[1] K. Takeshita, Y. Imamoto, M. Kataoka, F. Tokunaga, and M. Terazima. Themodynamic and transport properties of intermediate states of the photocyclic reaction of photoactive yellow protein. *Biochemistry*, 41(9):3037–3048, 2002. PMID: 11863442.

[2] J. Choi and M. Terazima. Denaturation of a protein monitored by diffusion coefficients: Myoglobin. *J. Phys. Chem. B*, 106(25):6587–6593, 2002.

[3] S. Nishida, T. Nada, and M. Terazima. Kinetics of intermolecular interaction during protein folding of reduced cytochrome ¡em¿c¡/em¿. *Biophys. J.*, 87(4):2663–2675, 2004.

[4] S. Nishida, T. Nada, and M. Terazima. Hydrogen bonding dynamics during protein folding of reduced cytochrome ¡em¿c¡/em¿: Temperature and denaturant concentration dependence. *Biophys. J.*, 89(3):2004–2010, 2024/11/05 2005.

[5] K. Inoue, N. Baden, and M. Terazima. Diffusion coefficient and the secondary structure of poly-l-glutamic acid in aqueous solution. *J. Phys. Chem. B*, 109(47):22623–22628, 2005. PMID: 16853945.

[6] T. Eitoku, Y. Nakasone, D. Matsuoka, S. Tokutomi, and M. Terazima. Conformational dynamics of phototropin 2 LOV2 domain with the linker upon photoexcitation. *J. Am. Chem. Soc.*, 127(38):13238–13244, 2005. PMID: 16173753.

[7] Y. Nakasone, T. Eitoku, D. Matsuoka, S. Tokutomi, and M. Terazima. Dynamics of conformational changes of arabidopsis phototropin 1 LOV2 with the linker domain. *J. Mol. Bio.*, 367(2):432–442, 2007.

[8] K. Inoue, M. Tsuda, and M. Terazima. Photoreverse reaction dynamics of octopus rhodopsin. *Biophys. J.*, 92(10):3643–3651, 2007.

[9] Y. Nakasone, Y. Kawasaki, M. Konno, K. Inoue, and M. Terazima. Time-resolved detection of light-induced conformational changes of heliorhodopsin. *Phys. Chem. Chem. Phys.*, 25:12833–12840, 2023.

[10] A. Fick. Ueber diffusion. *Ann. Phys.*, 170(1):59–86, 1855.

[11] A. Fick. On liquid diffusion. *Journal of membrane science*, 100(1):33–38, 1995.

[12] A. Einstein. Eine neue bestimmung der moleküldimensionen. *Ann. Phys.*, 324(2):289–306, 1906.

[13] A. Einstein. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Ann. Phys.*, 322(8):549–560, 1905.

[14] A. Einstein. Zur theorie der brownschen bewegung. *Ann. Phys.*, 324(2):371–381, 1906.

[15] R. Penrose. *Einstein's Miraculous Year: Five Papers That Changed the Face of Physics*. Princeton University Press, 2005.

[16] M. von Smoluchowski. Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen. *Ann. Phy.*, 326(14):756–780, 1906.

[17] P. Langevin. Sur la théorie du mouvement brownien. *C. R. Acad. Sci. (Paris)*, 146:530–533, 1908.

[18] Don S. Lemons and A. Gythiel. Paul Langevin's 1908 paper "On the Theory of Brownian Motion" ["Sur la théorie du mouvement brownien," C. R. Acad. Sci. (Paris) 146, 530–533 (1908)]. *Am. J. Phys.*, 65(11):1079–1081, 11 1997.

[19] J. Perrin. *Les Atomes*. Félix Alcan, Paris, 1913.

[20] J. Perrin. *Atoms*. Constable, London, 1916.

[21] Marco A. Catipovic, Paul M. Tyler, Josef G. Trapani, and Ashley R. Carter. Improving the quantification of Brownian motion. *Am. J. Phys.*, 81(7):485–491, 07 2013.

[22] J Stetefeld, S. A. McKenna, and T. R. Patel. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys. Rev.*, 8(4):409–427, 2016.

[23] R. Foord, E. Jakeman, C. J. Oliver, E. R. Pike, R. J. Blagrove, E. Wood, and A. R. Peacocke. Determination of diffusion coefficients of haemocyanin at low concentration by intensity fluctuation spectroscopy of scattered laser light. *Nature*, 227(5255):242–245, 1970.

[24] International Organization for Standardization. Iso 22412:2017 - particle size analysis — dynamic light scattering (dls). ISO standard, 2017. This document reached stage 90.92 on 2022-08-17, TC/SC: ISO/TC 24/SC 4, ICS: 19.120.

[25] W. Burchard, M. Schmidt, and W. H. Stockmayer. Information on polydispersity and branching from combined quasi-elastic and intergrated scattering. *Macromolecules*, 13(5):1265–1272, 1980.

[26] M. Terazima. Time-dependent intermolecular interaction during protein reactions. *Phys. Chem. Chem. Phys.*, 13:16928–16940, 2011.

[27] S. Ohishi, N. Shimizu, K. Mihara, and M. Imamoto, Y .and Kataoka. Light induces destabilization of photoactive yellow protein. *Biochemistry*, 40(9):2854–2859, 2001. PMID: 11258896.

[28] J. S. Khan, Y. Imamoto, M. Harigai, M. Kataoka, and M. Terazima. Conformational changes of pyp monitored by diffusion coefficient: Effect of n-terminal ¡em¿α¡/em¿-helices. *Biophys. J.*, 90(10):3686–3693, 2024/11/05 2006.

[29] M. Nakasako, T. Iwata, D. Matsuoka, and S. Tokutomi. Light-induced structural changes of lov domain-containing polypeptides from arabidopsis phototropin 1 and 2 studied by small-angle x-ray scattering. *Biochemistry*, 43(47):14881–14890, 2004. PMID: 15554695.

[30] H. Mori. Transport, Collective Motion, and Brownian Motion. *Prog. Theor. Phys.*, 33(3):423–455, 03 1965.

[31] D. J. Evans and G. Morriss. *Statistical Mechanics of Nonequilibrium Liquids*. Cambridge University Press, 2 edition, 2008.

[32] R. Kubo. The fluctuation-dissipation theorem. *Rep. Prog. Phys.*, 29(1):255, jan 1966.

[33] Y. Nakamura, A. Yoshimori, and R. Akiyama. A perturbation theory for friction of a large particle immersed in a binary solvent. *J. Phys. Soc. Jpn.*, 81(Suppl.A):SA026, 2012.

[34] Y. Nakamura, A. Yoshimori, and R. Akiyama. Effects of the solvation structure on diffusion of a large particle in a binary mixture studied by perturbation theory. *J. Mol. Liq.*, 200:85–88, 2014. Proceedings of the 7th Mini-Symposium on Liquids "Liquid-Liquid Phase Separation and Related Topics on Liquids".

[35] Y. Nakamura, A. Yoshimori, and R. Akiyama. Perturbation theory of large-particle diffusion in a binary solvent mixture. *J. Phys. Soc. Jpn.*, 83(6):064601, 2014.

[36] Y. Nakamura, A. Yoshimori, R. Akiyama, and T. Yamaguchi. Stick boundary condition at large hard sphere arising from effective attraction in binary hard sphere mixtures. *J. Chemput. Phys.*, 148(12):124502, 2018.

[37] Y. Nakamura, A. Yoshimori, and R. Akiyama. Solvation effects on diffusion processes of a macromolecule: Accuracy required for radial distribution function to calculate diffusion coefficient. *J. Chemput. Phys.*, 154(8):084501, 2021.

[38] J. R. Schmidt and J. L. Skinner. Hydrodynamic boundary conditions, the Stokes–Einstein law, and long–time tails in the Brownian limit. *J. Chem. Phys.*, 119(15):8062–8068, 10 2003.

[39] J. R. Schmidt and J. L. Skinner. Brownian motion of a rough sphere and the stokes–einstein law. *J. Phys. Chem. B*, 108(21):6767–6771, 2004.

[40] M. Fushiki. System size dependence of the diffusion coefficient in a simple liquid. *Phys. Rev. E*, 68:021203, Aug 2003.

[41] I-C. Yeh and G. Hummer. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *J. Phys. Chem. B*, 108(40):15873–15879, 2004.

[42] H. Hasimoto. On the periodic fundamental solutions of the stokes equations and their application to viscous flow past a cubic array of spheres. *J. Fluid Mech.*, 5(2):317–328, 1959.

[43] T. Iwashita, M. Nagao, A. Yoshimori, M. Terazima, and R. Akiyama. Usefulness of higher-order system-size correction for macromolecule diffusion coefficients: A molecular dynamics study. *Chem. Phys. Lett.*, 807:140096, 2022.

[44] M. P. Allen and D. J. Tildesley, editors. *Computer simulation of liquids*. Oxford, Clarendon, 1987.

[45] S. von Bülow, M. Siggel, M. Linke, and G. Hummer. Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. *PNAS*, 116(20):9843–9852, 2019.

[46] 今村 勤. 物理とグリーン関数（物理数学シリーズ； 4）. 岩波書店, 2016.

[47] B. Dünweg and K. Kremer. Molecular dynamics simulation of a polymer chain in solution. *J. Chem. Phys.*, 99(9):6983–6997, 11 1993.

[48] S. Honda, K. Yamasaki, Y. Sawada, and H. Morii. 10 residue folded peptide designed by segment statistics. *Structure*, 12(8):1507–1518, 2004.

[49] J. Loch, A. Polit, A. Górecki, P. Bonarek, K. Kurpiewska, M. Dziedzicka-Wasylewska, and K. Lewiński. Two modes of fatty acid binding to bovine β-lactoglobulin—crystallographic and spectroscopic studies. *J. Mol. Recognit.*, 24(2):341–349, 2011.

[50] J. L. F. Abascal and C. Vega. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.*, 123(23):234505, 2005.

[51] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996.

[52] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101, 2007.

[53] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182–7190, 1981.

[54] M. J. Abraham, D. van der Spoel, E. Lin-dahl, and B. Hess. the gromacs development team. gromacs user manual version 2019.4.

[55] M. J. Abraham, D. van der Spoel, E. Lin-dahl, and B. Hess. the gromacs development team. gromacs user manual version 2018.1.

[56] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An n·log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[57] D. Alfè and M. J. Gillan. First-principles calculation of transport coefficients. *Phys. Rev. Lett.*, 81:5161–5164, Dec 1998.

[58] Da-Mi Jung and Susan E. Ebeler. Investigation of binding behavior of α- and β-ionones to β-lactoglobulin at different ph values using a diffusion-based noe pumping technique. *J. Agric. Food Chem.*, 51(7):1988–1993, 2003. PMID: 12643663.

[59] International Association for the Properties of Water and Steam. Release on the iapws formulation 2020 for the viscosity of heavy water. *IAPWS R17-20*, 2020.

[60] T. Iwashita, Y. Uematsu, M. Terazima, and R. Akiyama. Decomposition of friction coefficients to analyze hydration effects on a C60(OH)n. *J. Chem. Phys.*, 161(24):244907, 12 2024.

[61] T. Yamaguchi and Y. Kimura. Effects of solute-solvent and solvent-solvent attractive interactions on solute diffusion. *Mole. Phys.*, 98(19):1553–1563, 2000.

[62] T. Yamaguchi, N. Matubayasi, and M. Nakahara. A mode-coupling approach to the attractive interaction effect on the solute diffusion in liquids. *J. chem. phys.*, 115(1):422–432, 07 2001.

[63] V. V. Chaban and E. E. Fileti. Which fullerenols are water soluble? systematic atomistic investigation. *New J. Chem.*, 41:184–189, 2017.

[64] Sonanki Keshri and B. L. Tembe. Thermodynamics of hydration of fullerols $[C_{60}(OH)_n]$ and hydrogen bond dynamics in their hydration shells. *J. Chem. Phys.*, 146(7):074501, 02 2017.

[65] Cleiton M., Eudes E. F., and Roberto R. Assessing the solvation mechanism of $C_{60}(OH)_{24}$ in aqueous solution. *Chem. Phys. Lett.*, 507(4):244–247, 2011.

[66] J. E. Straub, M. Borkovec, and B. J. Berne. Calculation of dynamic friction on intramolecular degrees of freedom. *J. Phy. Chem.*, 91(19):4995–4998, 1987.

[67] B. Kowalik, J. O. Daldrop, J. Kappler, J. C. F. Schulz, A. Schlaich, and R. R. Netz. Memory-kernel extraction for different molecular solutes in solvents of varying viscosity in confinement. *Phys. Rev. E*, 100:012126, Jul 2019.

[68] L. X. Dang, Kenneth M. Merz Jr., and P. A. Kollman. Free energy calculations on protein stability: Thr-157. fwdarw. val-157 mutation of t4 lysozyme. *J. Am. Chem. Soc.*, 111(22):8505–8508, 1989.

[69] S. Chatterjee, P. G. Debenedetti, F. H. Stillinger, and R. M. Lynden-Bell. A computational investigation of thermodynamics, structure, dynamics and solvation behavior in modified water models. *J. Chem. Phys.*, 128(12):124511, 03 2008.

[70] D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman. Amber 2022. University of California, San Francisco 2022.

[71] A. Jakalian, D. B. Jack, and C. I. Bayly. Fast, efficient generation of high-quality atomic charges. am1-bcc model: Ii. parameterization and validation. *J. Comput. Chem.*, 23(16):1623–1641, 2002.

[72] 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, and D. J. Fox. Gaussian˜16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.

[73] J. O. Daldrop, B. G. Kowalik, and R. R. Netz. External potential modifies friction of molecular solutes in water. *Phys. Rev. X*, 7:041065, Dec 2017.

[74] M. Bernetti and G. Bussi. Pressure control using stochastic cell rescaling. *J. Chem. Phys.*, 153(11):114107, 09 2020.

[75] M. Abraham, B. Hess, D. van der Spoel, and E. Lindahl. Gromacs – 2021.4. The GROMACS development teams at the Royal Institute of Technology and Uppsala University, Sweden. 2021.

[76] B. Hess, H. Bekker, H. J. C. Berendsen, and Johannes G. E. M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[77] S. Miyamoto and P. A. Kollman. Settle: An analytical version of the shake and rattle algorithm for rigid water models. *J. Comput. Chem.*, 13(8):952–962, 1992.

[78] T. W. Allen., O. S. Andersen, and B. Roux. Energetics of ion conduction through the gramicidin channel. *PNAS*, 101(1):117–122, 2004.

[79] A. B. Manonov., M. G. Kurnikova, and R. D. Coalson. Diffusion constant of k+ inside gramicidin a: A comparative study of four computational methods. *Biophys. Chem.*, 124(3):268–278, 2006. Ion Hydration Special Issue.

[80] B. J. Berne and G. D. Harp. *On the Calculation of Time Correlation Functions*, pages 63–227. John Wiley & Sons, Ltd, 1970.

[81] H. K. Shin, C. Kim, P. Talkner, and E. K. Lee. Brownian motion from molecular dynamics. *Chem. Phys.*, 375(2):316–326, 2010. Stochastic processes in Physics and Chemistry (in honor of Peter Hänggi).

[82] O. F. Lange and H. Grubmüller. Collective Langevin dynamics of conformational motions in proteins. *J. Chem. Phys.*, 124(21):214903, 06 2006.

[83] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, and M. Scharf. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.*, 16(3):273–284, 1995.

[84] A. V. Bondi. van der waals volumes and radii. *J. Phys. Chem*, 68(3):441–451, 1964.

[85] J. T. Bullerjahn, S. von Bülow, M. Heidari, J. Hénin, and G. Hummer. Unwrapping npt simulations to calculate diffusion coefficients. *Journal of Chemical Theory and Computation*, 19(11):3406–3417, 2023. PMID: 37257090.

[86] L. Rosso, P. Mináry, Z. Zhu, and M. E. Tuckerman. On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles. *J. Chem. Phys.*, 116(11):4389–4402, 03 2002.

[87] Luca M. and Eric V-E. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem. Phys. Lett.*, 426(1):168–175, 2006.

[88] Alessandro L. and Michele P. Escaping free-energy minima. *PNAS*, 99(20):12562–12566, 2002.

[89] T. Morishita, S. G. Itoh, H. Okumura, and M. Mikami. Free-energy calculation via mean-force dynamics using a logarithmic energy landscape. *Phys. Rev. E*, 85:066702, Jun 2012.