Supplementary Information

Statistical Thermodynamics for Unexpectedly Large Difference between Disaccharide Stereoisomers in Terms of Solubility in Water

Simon Hikiri,[1,2] Tomohiko Hayashi,[2] Mitsunori Ikeguchi,[3,4] and Masahiro Kinoshita*[2]

[1]Graduate School of Science, Chiba University, 1-33 Yayoi-cho, Inage, Chiba 263-8522, Japan
[2]Institute of Advanced Energy, Kyoto University, Uji, Kyoto 611-0011, Japan
[3]Graduate School of Medical Life Science, Yokohama City University, 1-7-29, Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan
[4]RIKEN Medical Sciences Innovation Hub Program, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

*To whom correspondence should be addressed. Tel: +81-774-38-3503. E-mail: kinoshit@iae.kyoto-u.ac.jp.

1. Molecular Dynamics (MD) Simulations

1.1. Force fields

We test three different force fields: CHARMM36,[1] GROMOS 45X4 (X=A in the aqueous solution and X=B in vacuum),[2] and GLYCAM.[3] Hereafter, CHARMM36, GROMOS 45X4 are referred to simply as CHARMM and GROMOS, respectively. CH, CH_2, and CH_3 groups are treated as united atoms in GROMOS. The parameters of potentials (partial charges and LJ parameters assigned to the constituent atoms) in CHARMM and GROMOS are collected in Table S1: The parameters for cellobiose are exactly the same as those for maltose. The arrangements of the atoms named in Table S1 are illustrated in Figure S1. In GLYCAM, the parameters are separately prepared for the first and second glucose units. There are slight differences between the parameter sets for cellobiose and maltose.

Table S1. Atomic partial charges (q) and Lennard-Jones potential parameters (σ and ε) of CHARMM and GROMOS force fields.

	CHARMM			GROMOS		
Atom type	q [−]	σ [Å]	ε [kcal/mol]	q [−]	σ [Å]	ε [kcal/mol]
C1 (1st glucose unit)	0.290	1.782	0.032	–	–	–
C1 (2nd glucose unit)	0.340	1.782	0.032	–	–	–
C1	–	–	–	0.232	2.510	0.023
C2	0.140	1.782	0.032	0.232	2.510	0.023
C3	0.140	1.782	0.032	0.232	2.510	0.023
C4 (1st glucose unit)	0.140	1.782	0.032	–	–	–
C4 (2nd glucose unit)	0.090	1.782	0.032	–	–	–
C4	–	–	–	0.232	2.510	0.023
C5	0.110	1.782	0.032	0.376	2.510	0.023
C6	0.050	1.791	0.056	0.232	2.035	0.098
H (ring)	0.090	1.194	0.045	–	–	–
H (hydroxymethyl)	0.090	1.194	0.035	–	–	–
H (hydroxyl)	0.420	0.200	0.046	0.410	0.000	0.000
O1	−0.650	1.572	0.192	−0.538	1.477	0.203
O2	−0.650	1.572	0.192	−0.642	1.477	0.203
O3	−0.650	1.572	0.192	−0.642	1.477	0.203
O4 (1st glucose unit))	−0.650	1.470	0.192	−0.642	1.477	0.203
O4 (2nd glucose unit)	−0.360	1.572	0.100	−0.360	1.477	0.203
O5	−0.400	1.470	0.100	−0.480	1.477	0.203
O6	−0.650	1.572	0.192	−0.642	1.477	0.203

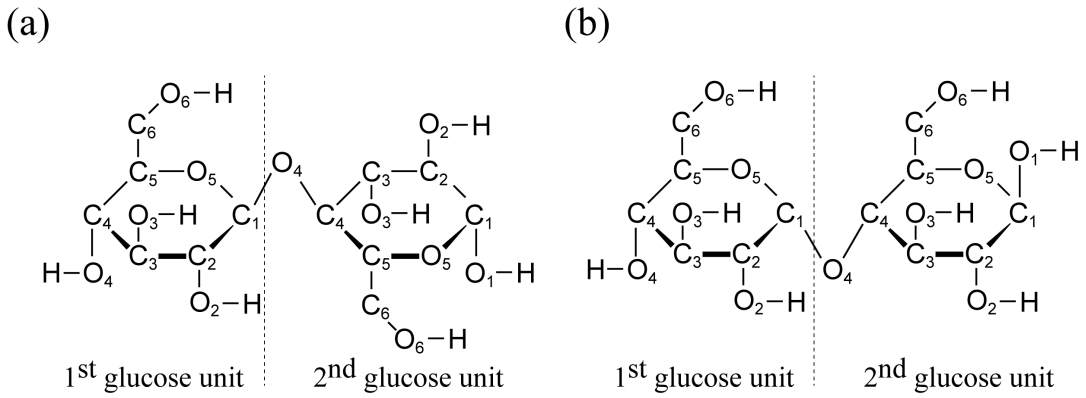

Figure S1. Arrangement of atoms named in Table S1. (a) Cellobiose. (b) Maltose. Hydrogen atoms which is bound to carbon atoms are not drawn.

Table S2. Atomic partial charges (q) and Lennard-Jones potential parameters (σ and ε) of GLYCAM force field for the first glucose unit (1st glucose unit; see Figure S1).

	Cellobiose			Maltose		
Atom type	Q [−]	σ [Å]	ε [kcal/mol]	q [−]	σ [Å]	ε [kcal/mol]
C1	0.384	1.700	0.109	0.509	1.700	0.109
C2	0.310	1.700	0.109	0.246	1.700	0.109
C3	0.284	1.700	0.109	0.286	1.700	0.109
C4	0.276	1.700	0.109	0.254	1.700	0.109
C5	0.225	1.700	0.109	0.283	1.700	0.109
C6	0.282	1.700	0.109	0.276	1.700	0.109
H (on C1)	0.000	1.147	0.016	0.000	1.147	0.016
H (on C2−C6)	0.000	1.236	0.016	0.000	1.236	0.016
H (on O2)	0.437	0.178	0.030	0.437	0.178	0.030
H (on O3)	0.432	0.178	0.030	0.427	0.178	0.030
H (on O4)	0.440	0.178	0.030	0.436	0.178	0.030
H (on O6)	0.424	0.178	0.030	0.418	0.178	0.030
O2	−0.718	1.533	0.210	−0.713	1.533	0.210
O3	−0.709	1.533	0.210	−0.699	1.533	0.210
O4	−0.714	1.533	0.210	−0.710	1.533	0.210
O5	−0.471	1.500	0.170	−0.574	1.500	0.170
O6	−0.688	1.533	0.210	−0.682	1.533	0.210

Table S3. Atomic partial charges (q) and Lennard-Jones potential parameters (σ and ε) of GLYCAM force field for the second glucose unit (2nd glucose unit: see Figure S1).

	Cellobiose			Maltose		
Atom type	q [−]	σ [Å]	ε [kcal/mol]	q [−]	σ [Å]	ε [kcal/mol]
C1	0.384	1.700	0.109	0.384	1.700	0.109
C2	0.310	1.700	0.109	0.310	1.700	0.109
C3	0.284	1.700	0.109	0.284	1.700	0.109
C4	0.276	1.700	0.109	0.276	1.700	0.109
C5	0.225	1.700	0.109	0.225	1.700	0.109
C6	0.282	1.700	0.109	0.282	1.700	0.109
H (on C1)	0.000	1.147	0.016	0.000	1.147	0.016
H (on C2−C6)	0.000	1.236	0.016	0.000	1.236	0.016
H (on O1)	0.445	0.178	0.030	0.445	0.178	0.030
H (on O2)	0.437	0.178	0.030	0.437	0.178	0.030
H (on O3)	0.432	0.178	0.030	0.432	0.178	0.030
H (on O6)	0.424	0.178	0.030	0.424	0.178	0.030
O1	−0.639	1.533	0.210	−0.639	1.533	0.210
O2	−0.718	1.533	0.210	−0.718	1.533	0.210
O3	−0.709	1.533	0.210	−0.709	1.533	0.210
O4	−0.468	1.500	0.170	−0.468	1.500	0.170
O5	−0.471	1.500	0.170	−0.471	1.500	0.170
O6	−0.688	1.533	0.210	−0.688	1.533	0.210

1.2. Simulation procedures

The molecular dynamics (MD) simulations for the disaccharides are carried out in the following manner. The initial conformation (structure) of a disaccharide molecule is constructed using the CHARMM 41b1 program package[4] for CHARMM and using the LEaP module in the AMBER 2017 program package[5] for GROMOS and GLYCAM. We have found that the conformations constructed via these two different routes are essentially the same. The following procedure is then followed using the AMBER 2017 program package[5] for all the three force fields.

The MD simulation in vacuum is performed at 298 K for a disaccharide molecule. The MD simulation in water is conducted for a disaccharide molecule immersed in the SPC/E water[6] on the basis of the NPT ensemble at 298 K and 1 atm under periodic boundary conditions. The minimum distance between the disaccharide surface and the box edge is initially set at ~27 Å that is about 9.6 times larger than the molecular diameter of water 2.8 Å. The number of water molecules and the initial box size are given in Table S4. The lengths of the bonds with H-atoms are constrained with the SHAKE algorithm.[7] The electrostatic interactions are calculated using the particle-mesh Ewald

(PME) method.[8] The real-space cut-off, spline order, and Ewald tolerance in the PME method are set at 8 Å, 4, and 10^{-5}, respectively. The cut-off for the Lennard-Jones (LJ) potential is 12 Å. The time step is 2.0 fs. The Langevin thermostat[9] and the Berendsen barostat[10] are employed for the temperature and pressure regulations, respectively.

To achieve the NPT ensemble at 298 K and 1 atm, the temperature and pressure equilibration of the system is carried out as follows. First, for removing the overlaps of the disaccharide and water molecules, the initial positions of water molecules are slightly modified using the steepest descent and conjugated gradient methods. By this modification, the potential energy of the system becomes sufficiently low. We then perform the subsequent temperature and pressure equilibration. We generate the velocities in accordance with a Maxwell distribution at 298 K and perform the temperature equilibration for 50 ps under the NVT condition. The pressure equilibration is then conducted at 298 K and 1 atm for 50 ps under the NPT condition. The disaccharide conformation thus obtained is referred to as "pre-equilibrated conformation" (we note that this conformation is quite similar to the initial one).

Starting from the pre-equilibrated conformation, we perform a 10-ns equilibration run followed by a 270-ns production run. These runs are conducted with the NPT ensemble at 298 K and 1 atm. During the production run, snapshot structures of a disaccharide molecule are stored every 100 ps and used in the calculation of thermodynamic quantities other than the conformational entropy: The number of the snapshot structures is 2700. (Refer to Section 2.4 for the calculation of the conformational entropy.) For estimating the statistical error for a thermodynamic quantity, we employ a block average by dividing the trajectory data into nine segments. In each segment, we have 300 snapshot structures and take an average of the 300 values of the quantity. The average of the nine averaged values for the segments is adopted as the theoretically calculated one. The block-average method possesses the advantage that we can check whether the calculated value converges sufficiently well as the MD simulation proceeds.

For a cellobiose molecule in vacuum, there are two peaks in the distribution of probability for φ, an angle associated with the glycosidic linkage (see Figure S4). Here, when we examined the value of Z in M_T snapshot structures obtained by the MD simulation and observed that Z equals Z_0 in M snapshot structures, "probability for Z at $Z=Z_0$ (or probability of $Z=Z_0$)" is M/M_T. (When Z is a continuous amount, the probability is calculated for discretized Z.) The two peaks occur at $\varphi\sim-75°$ and 50°. The probability of $\varphi\sim50°$ is much lower than that of $\varphi\sim-75°$. Only in the simulation using GLYCAM, however, we encounter the following problem: After the simulation time exceeds 100 ns, the molecular conformation is frozen with $\varphi\sim66°$. For this reason, we perform the production run only for 100 ns. This is why the results from CHARMM and GROMOS are mainly discussed and the result from GLYCAM is presented as additional information.

Table S4. Number of water molecules and initial box size.

	Force field	Number of water molecules	Initial box size: $x \times y \times z$ (Å^3)
Cellobiose	CHARMM	8000	65.32×65.32×65.32
	GROMOS	9238	65.77×65.77×65.77
	GLYCAM	8632	66.46×71.65×67.14
Maltose	CHARMM	8000	64.29×64.29×64.29
	GROMOS	8932	65.05×65.05×65.05
	GLYCAM	8606	67.70×70.41×66.82

1.3. Root mean square deviation (RMSD) from initial structure

The change of RMSD from the pre-equilibrated structure in terms of all the atoms during the MD simulation is shown in Figures S2 (CHARMM) and S3 (GROMOS). During the 270-ns production run, the RMSD fluctuates only within the small range 0.3–3.0 Å. In the case of cellobiose, for instance, there are roughly two stationary points in terms of the RMSD. The conformations around RMSD~1 and ~2.8 Å in each plot correspond to $\varphi \sim -75°$ and 50°, respectively.

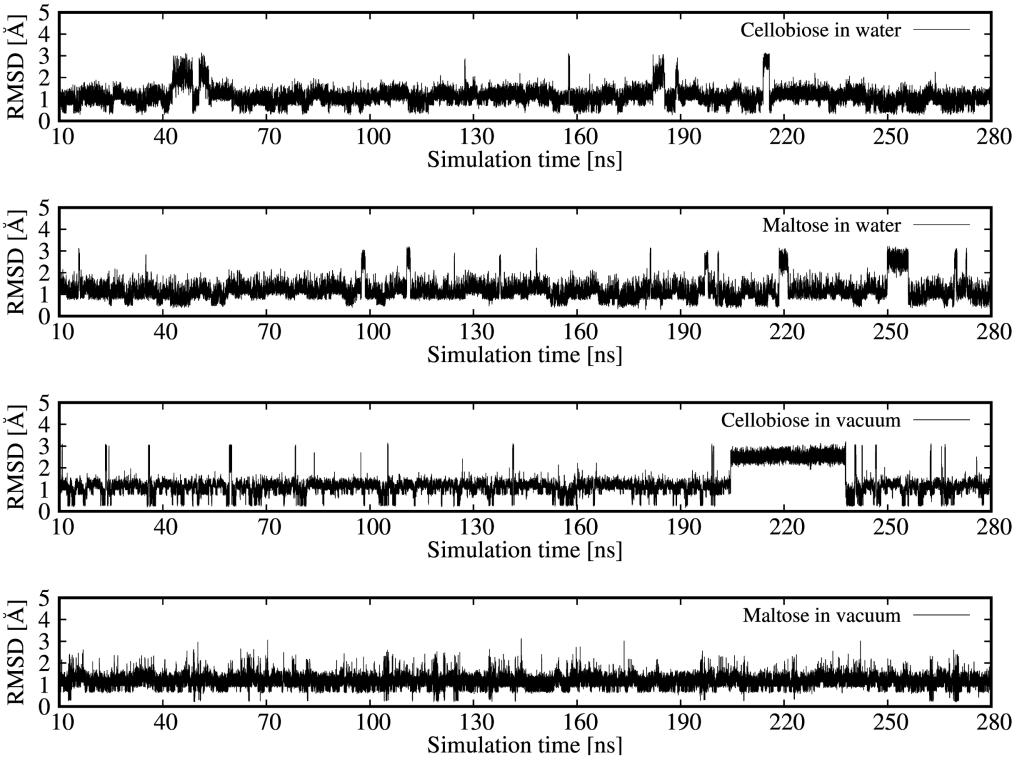

Figure S2. Root mean square deviation (RMSD) from the pre-equilibrated conformation in terms of all the atoms in the disaccharide molecule. The force field employed is CHARMM.

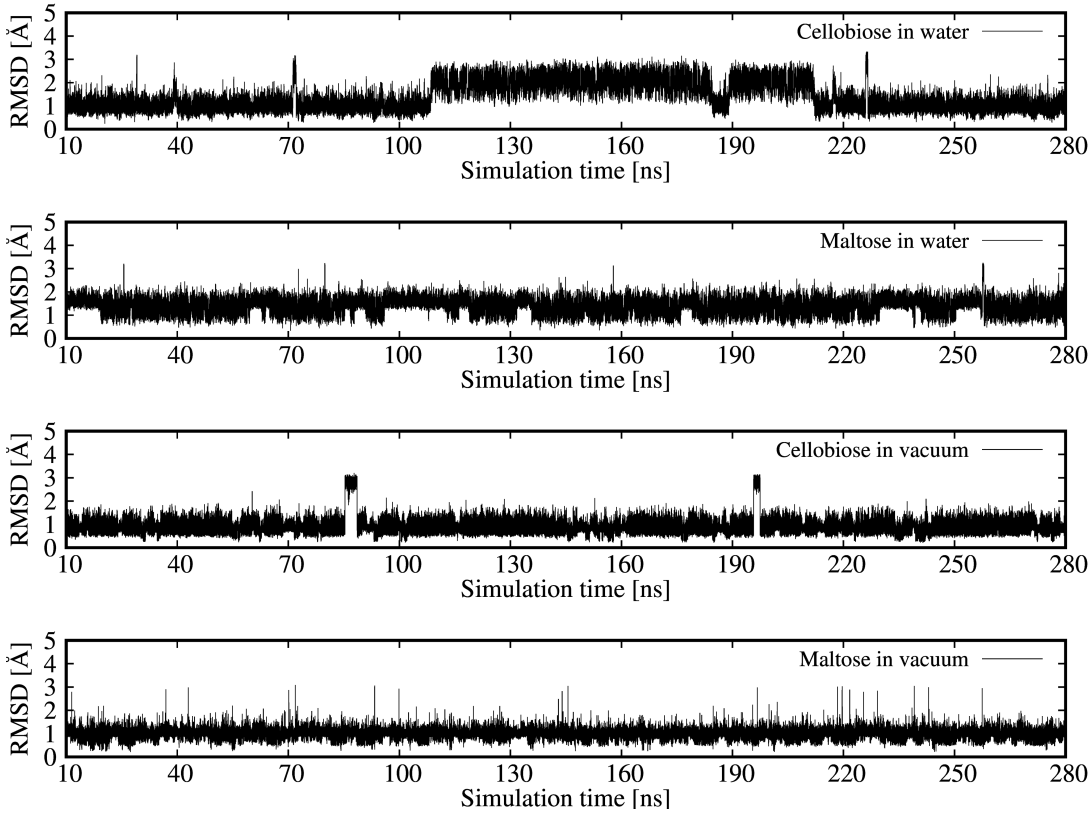

Figure S3. Root mean square deviation (RMSD) from the pre-equilibrated conformation in terms of all the atoms in the disaccharide molecule. The force field employed is GROMOS.

1.4. Examination of simulation results

Figure S4 shows the distribution of probability for each of the two torsion angles (φ and ψ) on the glycosidic bonds, which are calculated using CHARMM. Overall, the frequencies are quite similar to those reported by another research group.

The most important quantity is the difference between cellobiose and maltose in terms of the key free-energy function $\mu^*-\mu^{g\circ}$: $\Delta(\mu^*-\mu^{g\circ})=(\mu^*-\mu^{g\circ})_A-(\mu^*-\mu^{g\circ})_B$ (the subscripts "A" and "B" denote "cellobiose" and "maltose", respectively). To check if the simulation time is long enough, we plot $\Delta(\mu^*-\mu^{g\circ})$ calculated against the simulation time in Figure S5 where the snapshot structures sampled until "Time" are utilized in the calculation. $\Delta(\mu^*-\mu^{g\circ})$ calculated is well converged especially when CHARMM is employed as the force field. In the figure, we also show that the difference between

cellobiose and maltose in terms of the entropic component, $-T\Delta S_D = -T(S_{CW}-S_C)_A-(S_{CW}-S_C)_B$, is also well converged.

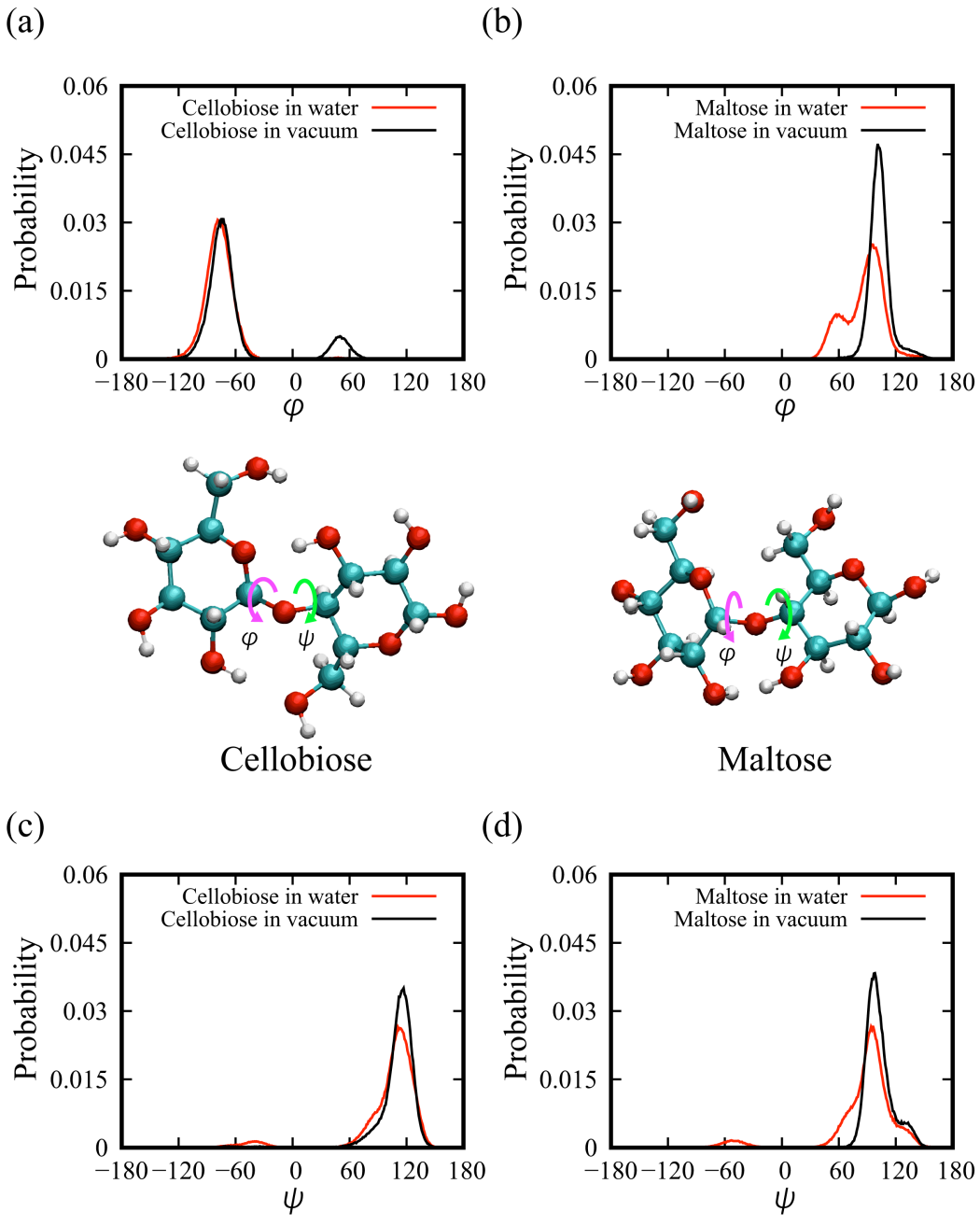

Figure S4. Distribution of probability for each of the two torsion angles (φ and ψ) on the glycosidic bonds in vacuum or in water. The force field employed is CHARMM.

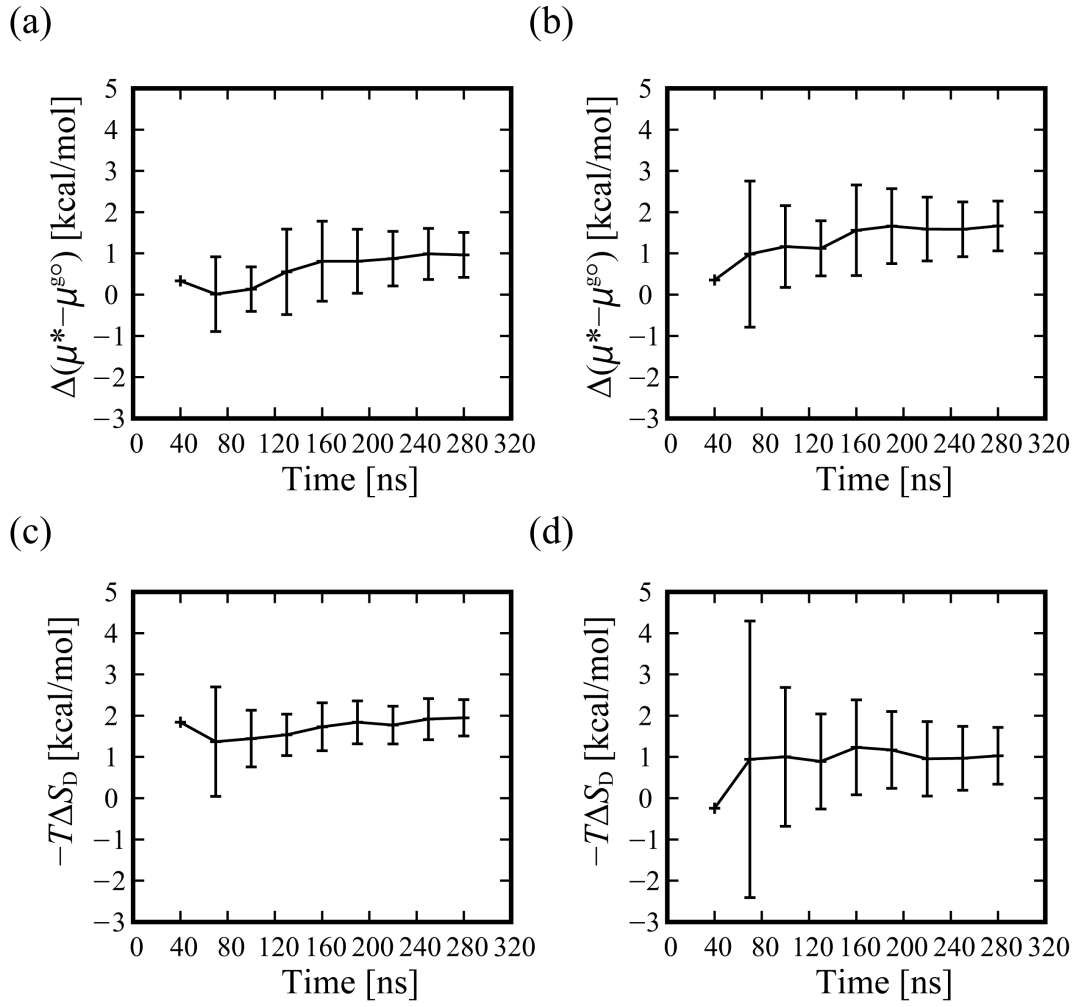

Figure S5. Convergent behavior of $\Delta(\mu^*-\mu^{g\circ})$. $\Delta(\mu^*-\mu^{g\circ})$="$(\mu^*-\mu^{g\circ})$ of cellobiose" – "$(\mu^*-\mu^{g\circ})$ of maltose". (a) CHARMM. (b) GROMOS. Convergent behavior of $-T\Delta S_D$. (c) CHARMM. (d) GROMOS. The error bar indicates 95% confidence interval.

2. Theoretical Methods for Calculating Hydration Energy, Hydration Entropy, and Conformational Entropy

2.1. Two types of integral equation theories for molecular liquids

In an integral equation theory (IET),[11] the Ornstein-Zernike (OZ) equation and a closure equation are derived from the system partition function on the basis of classical statistical mechanics. In the case of bulk solvent of a single component, for example, the temperature, number density, and solvent-solvent interaction potential form the input data. By numerically solving the OZ equation

coupled with the closure equation, we can calculate the direct and total correlation functions, microscopic structure of the solvent, and thermodynamic quantities. The average value of a physical quantity is calculated essentially for an infinitely large system and an infinitely large number of system configurations. The solvent structure near a solute and thermodynamic quantities of solvation (i.e., solvation free energy, entropy, and energy), can also be calculated. In the analysis of solvation properties of a solute, it is assumed that the solute is inserted into a solvent under isochoric condition at infinite dilution. In the first step, the bulk solvent is treated and the solvent-solvent correlation functions are calculated. In the second step, the solute-solvent correlation functions are calculated using the solvent-solvent correlation functions and the solute-solvent interaction potential as part of the input data. For a molecular liquid like water as the solvent, there are two types of IETs: the reference interaction site model (RISM) theory[12–16] and the angle-dependent integral equation (ADIE) theory.[17–22] In what follows, we briefly summarize the two types of IETs applied to bulk water.

In the RISM and related theories,[12–16] a water molecule is represented by atomic sites referred to as "interaction sites" (i.e., a water molecule has three sites of an oxygen and two hydrogens: O, H, and H). The OZ equation is expressed by the site-site correlation functions and the intramolecular correlation function representing the molecular structure of water and referred to as "site-site OZ (SSOZ) equation". The closure equation is also expressed by the site-site correlation functions. A site-site correlation function is dependent only on the distance between centers of two interaction sites, which is advantageous simplicity. In this study, we employ a modified version of the SPC/E (cSPC/E) model[23] for water, the dielectrically consistent version referred to as "dielectrically consistent RISM (DRISM) theory",[12] and the Kovalenko-Hirata (K-H) closure equation.[15] The DRISM theory is the most reliable among the RISM and related theories. The value of $\rho_S d_S^3$ is set at 0.7317 (ρ_S is the number density of bulk water and d_S=2.8 Å) and T=298 K.

On the other hand, the ADIE theory[17–22] explicitly takes account of the dependence of a correlation function on the orientations of water molecules. Therefore, a correlation function is dependent not only on the distance between centers of two water molecules but also on their orientations. The orientation of a water molecule is represented by three Euler angles. The water-water correlation is a function of six independent variables. The SSOZ equation is approximate, whereas the OZ equation in the ADIE theory is exact. In the ADIE theory, however, since the computational burden in numerically handling the six-variable functions is unacceptably heavy, the OZ and closure equations must be reduced using mathematical techniques such as the rotational-invariant expansion of correlation functions. In this study, a water molecule is modeled as a hard sphere with diameter d_S=0.28 nm ($\rho_S d_S^3$=0.7317) in which a point dipole and a point quadrupole of tetrahedral symmetry are embedded.[17,18] The hypernetted-chain (HNC) approximation is adopted as the closure equation. The influence of molecular polarizability of water is taken into account by employing the self-consistent mean field theory.[17,18] It has been corroborated that the results from the ADIE/HNC theory are in very good agreement with those from Monte Carlo and

MD simulations for hydrophobic hydration.[22,24]

Here, we summarize the characteristics of the DRISM and ADIE theories exhibited when an analysis on solute hydration is undertaken. In general, the result from the DRISM theory is less accurate than that from the ADIE theory.[22,24] However, the hydration energy of a hydrophilic solute can be calculated with sufficient accuracy even by the DRISM theory,[25,26] though the hydration entropy is significantly less accurate.[24] The great advantage of the DRISM theory is that it can treat a large solute with polyatomic structure. When a polyatomic solute is treated, a solute-water site-site correlation function, which is defined for an atom in the solute with a fixed structure (this is the solute) and an interaction site in a water molecule, is three dimensional. Hence, the theory for analyzing hydration of a polyatomic solute is referred to as "three-dimensional RISM (3D-RISM) theory".[13–16] On the other hand, the ADIE theory can practically treat only a spherical solute due to the numerical complexity mentioned above. Our solution of this problem is the following: For a solute molecule with a fixed structure, its hydration energy is calculated by the 3D-RISM theory; and its hydration entropy is calculated by combining the ADIE theory with our morphometric approach (MA)[30,31] which is capable of relating the hydration entropy of the solute molecule to its polyatomic structure with quantitative accuracy.

2.2. Calculation of hydration energy of a solute molecule with a fixed structure

In the 3D-RISM theory,[13–16] the SSOZ and closure equations are expressed by the 3D correlation functions. The calculation procedure is as follows. First, the site-site correlation functions for bulk water are calculated using the DRISM theory[12] coupled with the K-H closure equation.[15] The solute-water site-site correlation functions are then obtained by numerically solving the 3D-RISM theory where the K-H closure equation[15] is employed. (The SSOZ and K-H closure equations were written in our earlier publications.[24,27–29]) The numerical solution is performed on a 3D cubic grid. The grid spacing (Δx, Δy, and Δz) is set at 0.5 Å, and the grid resolution ($N_x \times N_y \times N_z$) is chosen such that the minimum distance between the solute surface and the box edge reaches 14 Å. It has been confirmed that the spacing is sufficiently small and the box size ($N_x\Delta x$, $N_y\Delta y$, $N_z\Delta z$) is large enough for the calculation result to become identical within convergence tolerance.

The hydration free energy of the solute μ_H is calculated using the modified version of the Singer-Chandler formula.[32] The hydration entropy S_H is evaluated through the analytical differentiation of μ_H with respect to T. The hydration energy ε_H is calculated from

$$\varepsilon_H = \mu_H + TS_H. \quad \text{(S2)}$$

We note that μ_H is independent of the solute insertion condition, isochoric or isobaric, though ε_H and S_H are not.[21]

2.3. Calculation of hydration entropy of a solute molecule with a fixed structure

Fortunately, it has been shown that S_H is fairly insensitive to the solute-water interaction potential[32] and influenced primarily by the geometric characteristics of the solute structure.[27–31] Therefore, a solute molecule with a fixed structure (this is the solute) can be modeled as a set of fused, neutral hard spheres, in which case the diameter of each atom in the solute molecule is set at the corresponding value of an LJ potential parameter σ.

The geometric characteristics of the solute polyatomic structure can be taken into account by only its four geometric measures with quantitative accuracy, which is a great advantage of the MA.[30,31] The four measures are the excluded volume denoted by V_{ex}, solvent-accessible surface area A, and integrated mean and Gaussian curvatures of the accessible surface, X and Y, respectively. S_H is expressed as the linear combination of the four measures referred to as "morphometric form"[27–31] (k_B is the Boltzmann constant):

$$S_H/k_B=C_1V_{ex}+C_2A+C_3X+C_4Y. \tag{S3}$$

The four coefficients (C_1–C_4), which are dependent only on the solvent species and its thermodynamic state, are determined beforehand for the simplest geometries: isolated hard-sphere solutes with various diameters.

The calculation comprises the following steps:

(1) Calculate S_H of an isolated hard-sphere solute (S_{IHSS}) with diameter d_U using the ADIE theory.[16–21] Consider different values of d_U in the range, $0.6\le d_U/d_S\le 10$, to obtain a sufficiently large set of data for S_H and d_U. The numerical solution of the basic equations is carried out using the robust, highly efficient algorithm developed by Kinoshita and coworkers.[19,34] (The basic equations, the OZ and HNC closure equations, were written in earlier publications.[17–22,24,27–29]) S_H is evaluated through Eq. (S1) where μ is calculated using the Morita-Hiroike formula[35,36] adapted to molecular liquids.

(2) Determine C_1–C_4 by applying the least-squares method to the following equation:

$$S_{IHSS}/k_B=C_1(4\pi R^3/3)+C_2(4\pi R^2)+C_3(4\pi R)+C_4(4\pi),\ \ R=(d_U+d_S)/2. \tag{S4}$$

Equation (S4) is the morphometric form for isolated spherical solutes.

(3) Calculate V_{ex}, A, X, and Y of the biomoecule using an extended version of Connolly's algorithm.[37,38] The input data are the (x, y, z) coordinates of the center of each atom in the solute molecule and its diameter D. D is set at σ taken from the force field employed.

(4) Calculate S_H of the solute molecule from Eq. (S3) to which C_1–C_4 determined in step (2) are substituted.

The high accuracy of the MA was corroborated in our earlier works.[30,31] We tested a hard-sphere solvent[31] and a solvent that is quite similar to the LJ liquid.[31] We calculated the solvation entropy of protein G with hundreds of different structures via the two routes: the 3D integral equation theory (3D-IET)[39,40] and a combination of the radial-symmetric IET (RSIET)[11] and the MA. The polyatomic structure of the protein can explicitly be treated by the 3D-IET. The error of the RSIET-MA combination was smaller than ±2%.

In the morphometric form hinging on the Hadwiger theorem,[41] the four coefficients are equal to thermodynamic quantities of pure bulk solvent. In the case of the solvation free energy μ, for instance, the first and second coefficients are the pressure P and the surface tension γ, respectively. However, the Hadwiger theorem is valid only for an infinitely large solute. As argued in our earlier works,[31,43,45] the form becomes problematic when it is applied to a nonpolar solute immersed in water. At P=1 atm, the EV term is negligibly small, and the form is approximated by $\mu \sim \gamma A$ (μ>0). However, γ becomes larger as T is lowered with the result that μ increases and the hydrophobicity is strengthened. This conflicts with the experimental evidence that at low temperatures the hydrophobicity is weakened and a protein is denatured.[31,43–45] No such problem arises in our form, because the four coefficients are calculated using the IET.

In general, when the 3D-RISM theory is exclusively applied to the calculation of the free-energy change upon binding of two biomolecules, it usually becomes positive and significantly large.[28,42] This serious problem is overcome by replacing the 3D-RISM theory by the ADIE-MA combination in the calculation of S_H.

2.4. Calculation of conformational entropy of solute molecules

During the production run, snapshot conformations of a disaccharide molecule are stored every 1 ps and used in the calculation of conformational entropy: The number of the snapshot conformations is 270000. In the Boltzmann-quasi-harmonic (BQH) approximation method,[46-48] the conformational entropy S_{BQH} is expressed as

$$S_{BQH} = -k_B \Sigma_i^{3N-6} \int dq_i \rho(q_i) \ln \rho(q_i) + (\ln |\boldsymbol{C}|)/2 \quad \text{(S5)}$$

where N is the number of atoms, $\rho(q_i)$ is the probability-density function (PDF) of i-th coordinate q_i, and $\boldsymbol{C}$ is the correlation coefficient matrix. The coordinate system is composed of bond, angle, and torsion terms, and the number of coordinates is $3N$−6. The first term is the so-called "independent term". It is calculated by a summation of the exact one-dimensional Boltzmann entropy which is described using $\rho(q_i)$. The second term is the so-called "correlation term". It is equivalent to the second term in the quasi-harmonic approximation.[49] The calculation procedure is as follows. For

estimating the statistical error for $-T\Delta S_D$, we employ a block average by dividing the trajectory data into nine segments. $-T\Delta S_D$ is calculated using 30000 snapshot conformations in each segment, and the value averaged over the nine segments is adopted as the theoretically calculated one. The first term of Eq. (S5) is calculated using the 30000 snapshot conformations. The numerical solution is performed using the histogram binning method.[47] The width of bin for bond is set at 0.025 Å, and those for angle and torsion are set at 1 degree. $\boldsymbol{C}$ in the second term is also calculated using the snapshot conformations. We emphasize that, for the calculation of conformational entropy, the BQH approximation using internal coordinate system including improper torsion[50] is far superior to the other methods which are currently available (e.g., quasi-harmonic and related computational methods[46,49–52]).

2.5. Comparison between 3D-RISM and MD results in terms of hydration energy

We calculate the hydration energy $\varepsilon_{H,MD}$ of maltose by the MD simulation based on NPT ensemble using CHARMM. The procedure is as follows:

(1) Using 2700 snapshot configurations of the equilibrated maltose-water system, we calculate the system energy E_{System} comprising the conformational energy of maltose, maltose-water interaction energy, and water-water interaction energy. The conformational energy of maltose is denoted by $E_{Maltose}$.

(2) We then perform the MD simulation for pure water. After the equilibrium is reached, we calculate the pure-water energy E_{Water} using 2700 snapshot configurations. The numbers of water molecules in (1) and (2) are the same.

(3) $\varepsilon_{H,MD}$ is obtained as $\varepsilon_{H,MD}=E_{System}-(E_{Maltose}+E_{Water})$.

We note that $\varepsilon_{H,MD}$ is the hydration energy under the isobaric condition. On the other hand, the hydration energy obtained by the 3D-RISM theory, ε_H, is calculated under the isochoric condition. Therefore, the hydration energy under the isobaric condition from the 3D-RISM theory $\varepsilon_{H,IET}$ to be compared with $\varepsilon_{H,MD}$ is given by[21]

$$\varepsilon_{H,IET}/(k_BT)=\varepsilon_H/(k_BT)+\Omega-PV_P/(k_BT), \quad \text{(S6a)}$$

$$\Omega=(\alpha^*/\kappa_T^*)V_P/d_S^3. \quad \text{(S6b)}$$

Here, P is the system pressure and V_P is the partial molar volume of maltose (i.e., the change in system volume caused upon insertion of maltose into water under the isobaric condition). The dimensionless parameters, α^* and κ_T^*, which depend only on the properties of pure bulk water, are defined as $\alpha^*=\alpha T$ and $\kappa_T^*=\kappa_T k_B T/d_S^3$ where α is the isobaric thermal expansion coefficient and κ_T is the isothermal compressibility. $\alpha^*/\kappa_T^*=0.894$ at $T=298$ K.[53,54] We calculate V_P using the 3D-RISM theory. At $P=1$ atm, $PV_P/(k_BT)$ is negligibly small.

The result obtained is as follows: ε_{H}=−58.87±0.11, $\varepsilon_{H,IET}$=−50.04±0.11, and $\varepsilon_{H,MD}$=−52.2±12.1 (the unit is kcal/mol). Considering the fact that the water model in the 3D-RISM theory is not exactly the same as that in the MD simulation, we conclude that the agreement between $\varepsilon_{H,IET}$ and $\varepsilon_{H,MD}$ is remarkably good.

However, the large error bars (the 95% confidence intervals) in the MD result is rather puzzling. A problem in the MD simulation is the following: E_{System}=−86153.8±7.0, E_{Water}=−86359.0±6.6, $E_{Maltose}$=257.4±0.2, and $E_{Maltose}$+E_{Water})=−86101.6±6.6 (the unit is kcal/mol); cancellation of significant digits occurs in $\varepsilon_{H,MD}$=E_{System}−($E_{Maltose}$+E_{Water}); and calculation of $\varepsilon_{H,MD}$ with sufficient accuracy is a rather difficult task. It seems that a simulation time which is much longer than 270 ns is required for the production run. In our opinion, the MD simulation is not suited to the present study which attempts to elucidate $\Delta(\mu^{*}-\mu^{g\circ})$=0.87 kcal/mol. The IET is much more suited. In the simulation time of 270 ns, we can generate a sufficiently large ensemble of molecular structures of a disaccharide in water for calculating a thermodynamic quantity of hydration using the IET.

References

(1) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., Jr. *J. Chem. Theory Chomput.* **2009**, *5*, 2353–2370.

(2) Lins, R. D.; Hünenberger P. H. *J. Comput. Chem.* **2005**, *26*, 1400–1412.

(3) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655.

(4) Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(5) Case, D. A.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R.M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 2017*; University of California: San Francisco, 2017.

(6) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(7) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(8) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(9) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
(10) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
(11) Hansen, J.-P.; McDonald, I. R.; *Theory of simple liquids*, 3rd ed.; Academic Press: London, 2006.
(12) Perkyns, J. S.; Pettitt, B. M. *J. Chem. Phys.* **1992**, *97*, 7656–7666.
(13) Beglov, D.; Roux, B. *J. Chem. Phys.* **1995**, *103*, 360–364.
(14) Beglov, D.; Roux, B. *J. Chem. Phys.* **1996**, *104*, 8678–8689.
(15) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **1999**, *110*, 10095–10112.
(16) Ratkova, E. L.; D. S. Palmer; Fedorov, M. V. *Chem. Rev.* **2015**, *115*, 6312–6356.
(17) Kusalik, P. G.; Patey, G. N. *J. Chem. Phys.* **1988**, *88*, 7715–7738.
(18) Kusalik, P. G.; Patey, G. N. *Mol. Phys.* **1988**, *65*, 1105–1119.
(19) Kinoshita, M.; Bérard, D. R. *J. Comput. Phys.* **1996**, *124*, 230–241.
(20) Kinoshita, M.; Iba, S.; Harada, M. *J. Chem. Phys.* **1996**, *105*, 2487–2499.
(21) Cann, N. M.; Patey, G. N. *J. Chem. Phys.* **1997**, *106*, 8165–8195.
(22) Kinoshita, M. *J. Chem. Phys.* **2008**, *128*, 024507.
(23) Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. *J. Chem. Theory Comput.* **2010**, *6*, 607−624.
(24) Hayashi, T.; Oshima, H.; Harano, Y.; Kinoshita, M. *J. Phys.: Condens. Matter* **2016**, *28*, 344003.
(25) Truchon, J-F.; Pettitt, B. M.; Labute, P. *J. Chem. Theory Comput.* **2014**, *10*, 934–941.
(26) Oshima, H.; Kinoshita. M. *J. Comput. Chem.* **2016**, *37*, 712–723.
(27) Hayashi, T.; Oshima, H.; Mashima, T.; Nagata, T.; Katahira, M.; Kinoshita, M. *Nucleic Acids Res.* **2014**, *42*, 6861–6875.
(28) Hayashi, T.; Oshima, H.; Yasuda, S.; Kinoshita, M. *J. Phys. Chem. B* **2015**, *119*, 14120–14129.
(29) Oshima, H.; Hayashi, T.; Kinoshita, M. *Biophys. J.* **2016**, *110*, 2496–2506.
(30) Roth, R.; Harano, Y.; Kinoshita, M. *Phys. Rev. Lett.* **2006**, *97*, 078101.
(31) Oshima, H.; Kinoshita, M. *J. Chem. Phys.* **2015**, *142*, 145103.
(32) Singer, S. J.; Chandler, D. *Mol. Phys.* **1985**, *55*, 621–625.
(33) Imai, T.; Harano, Y.; Kinoshita, M.; Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **2006**, *125*, 024911.
(34) Kinoshita, M.; Harada, M. *Mol. Phys.* **1994**, *81*, 1473–1488.
(35) Morita, T. *Prog. Theor. Phys.* **1960**, *23* 829–845.
(36) Morita, T.; Hiroike, K. *Prog. Theor. Phys.* **1961**, *25* 537–578.
(37) Connolly, M. L. *Science* **1983**, *221*, 709–713.
(38) Connolly, M. L., *J. Am. Chem. Soc.* **1985**, *107*, 1118–1124.
(39) Ikeguchi, M.; Doi, J. *J. Chem. Phys.* **1995**, *103*, 5011−5017.

(40) Kinoshita, M. *J. Chem. Phys.* **2002**, *116*, 3493−3501.
(41) König, P.-M.; Roth, R.; Mecke, K. R. *Phys. Rev. Lett.* **2004**, *93*, 160601.
(42) Hayashi, T.; Matsuda, T.; Nagata, T.; Katahira, M.; Kinoshita, M. *Phys. Chem. Chem. Phys.* **2018**, *20*, 9167−9180.
(43) Yoshidome, T.; Kinoshita, M. *Phys. Chem. Chem. Phys.* **2012**, *14*, 14554−14566.
(44) Kinoshita, M. *Biophys. Rev.* **2013**, *5*, 283−293.
(45) Kinoshita, M. and Hayashi, T. *Phys. Chem. Chem. Phys.* **2017**, *19*, 25891−25904.
(46) Di Nola, A.; Berendsen, H.; Edholm O. *Macromolecules* **1984**, *17*, 2044–2050.
(47) Harpole K. W.; Sharp K. A. *J. Phys. Chem. B* **2011**, *115*, 9461–9472.
(48) Hikiri, S.; Yoshidome, T; Ikeguchi M. *J. Chem. Theory Comput.* **2016**, *12*, 5990–6000.
(49) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325–332.
(50) McQuarrie, D. A. *Statistical mechanics*, University Science Books: California, 2000.
(51) Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617–621.
(52) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289–6292.
(53) Wagner, W.; Pruß, A. *J. Phys. Chem. Ref. Data*, **2002**, *31*, 387–535.
(54) Shvab, I.; Sadus, R. J. *J. Chem. Phys.*, **2013**, *139*, 194505.