広島大学学位請求論文

Theoretical Study on the Hydration Effect of

Organic Molecule in Aqueous Solution

水溶液中の有機分子の水和効果についての理論的研究

2014 年

広島大学大学院理学研究科

化学専攻

Hideo Doi

土居　英男

目次

１．主論文

Theoretical Study on the Hydration Effect of Organic Molecule in Aqueous Solution

水溶液中の有機分子の水和効果についての理論的研究

土居　英男

２．公表論文

(1) Hydration of Adamantane Skeleton: Water Assembling around Amantadine and Halo-substituted Adamantanes

H. Doi, M. Aida, *Chemistry Letters*. **2013**, 42, 292–294.

(2) A new variant of multicanonical Monte Carlo algorithm with specifying the temperature range and its application to the hydration free energy change of fluorinated methane derivatives

H. Doi, M. Aida, *Chemical Physics Letters*. **2014**, 595-596, 55-60.

主論文

Theoretical Study on the Hydration Effect of Organic Molecule in Aqueous Solution

Hideo Doi

Department of Chemistry, Graduate School of Science,

Hiroshima University

Acknowledgement

I am deeply grateful to Professor Misako Aida for kindly guidance and a lot of encouragement throughout my studies.

I am deeply grateful to Professor Taka-aki Ishibashi in University of Tsukuba for helpful comments and valuable discussions about my research.

I am deeply grateful to Associate Professor Kazumasa Okada and Assistant Professor Yukiteru Katsumoto, for helpful suggestions.

I am grateful to Professor Takayuki Ebata, Professor Katsuyoshi Yamasaki, and all of the professors and staffs in the department of chemistry in Hiroshima University.

I want to thank Dr. Masato Tanaka, Dr. Tomoki Yoshida, Dr. Masayuki Ohisa, Dr. Hidenori Miyamoto, Mr. Dai Akase for valuable discussions and comments.

I also want to thank Mr. Akihiro Maeda, Dr. Shunsuke Mieda, Mr. Hiroshi Ando, and all the members of quantum chemistry research group.

Contents

Chapter 1.

General Introduction

1.1 Introduction

There are innumerable organic compounds in this world. The scientists have classified a large number of the properties of the organic compounds but most of the properties have not been reached perfect comprehension because the many molecular properties depend on many factors. One of the most important properties is a hydration property. For example, a significant fraction of the human body is water. In the drug discovery, the hydrophobic property and hydrophilic property is important. In our body, there are hydrophilic molecules and hydrophobic molecules, e.g. lipid bilayer. Moreover it is difficult experimentally to see hydrophobic molecule in the water molecules. In a few decades, computers as a calculator have been growing in performance. The computers are able to carry out a numerical simulation. The theoretical approach with a computational simulation has been a powerful tool to give interpretations of hydration mechanism with microscopic level.

To study hydration, I calculate the water distribution around the solute molecule at the finite temperature. The substitution effects of solute molecules affect the water distribution around the solute molecule.

1.2 Create canonical ensemble

Even if the high level of theory is employed to calculate the energy of a solvent configuration, the solvent effects are not able to discuss form the configuration because the solvent structure has countless configurations at finite temperature. Therefore we need to take an ensemble average. Monte Carlo (MC) simulation can give a canonical ensemble which leads to a partition function, Helmholtz energy of the system and solvent structures at finite temperature.

Molecular dynamics (MD) simulation [3] can give a canonical ensemble too. In MD simulation, a new configuration is generated by applying Newton's equations of motion to all atoms of a system. Thus an MD method requires the gradients of the potential energy surface to calculate the forces acting between the atoms.

However, in the MD simulation, the temperature controlling is complicated. The best way for controlling the temperature does not exist. Nosé-Hoover thermostat, Nosé-Hoover chains thermostat, Berendsen thermostat and Andersen thermostat. These all method is a common method. Any method has the good feature, respectively.

On the other hand, in the MC simulation, the temperature controlling is simple. This feature is good point of MC simulation.

1.3 Monte Carlo method with Metropolis-Hastings algorithm

MC method is a statistical method with random sampling. Metropolis-Hastings algorithm [1,2] for MC method is a useful to make a canonical ensemble of the system including the solute molecule and the many solvent molecules. This method is useful for obtaining an ensemble from a known probability distribution function.

1.4 Calculate Helmholtz energy changes

Free energy perturbation (FEP) theory is a method for calculation the changes in Helmholtz energy between two states. In other words, FEP theory gives the Helmholtz energy changes that result from conversion of atom or functional group.

In CaMC simulation of a system at temperature T and an inverse temperature β, the difference in Helmholtz energy between a reference system i and a target system j, ΔA^{ij}, is given by

$$\Delta A^{ij} = A^j - A^i = \frac{\ln\left\langle \exp\{-\beta\,(E^j - E^i)\}\right\rangle_{(i)}}{-\beta} \qquad (1)$$

The bracket $\langle \ldots \rangle_{(i)}$ denotes the ensemble average. To sample the configurations, we adopt the "double wide" sampling.

In FEP calculations, a coupling parameter λ_i was used to linearly scale all geometrical and force-field parameters χ from the parameters χ_{Ini} for initial state to the parameters χ_{Fin} for final state as follows:

$$\chi_i = \lambda_i \chi_{\mathrm{Fin}} + (1 - \lambda_i)\chi_{\mathrm{Ini}}$$

1.5 Non-Boltzmann sampling

There are many non-Boltzmann sampling method. The aim of these methods is to sample the rare event in the Boltzmann distribution. Therefore the probability distribution in these methods is different from the distribution in the Boltzmann distribution.

First, the replica exchange method was developed by Swendsen and Wang. [4] This method was proposed for simulations of systems with quenched random interactions.

Second, the simulated tempering method was developed by Marinari and Paris. [5] This method was proposed as a new global optimization method for a system with a rough free energy landscape. In this method, the temperature is variable.

Third, the multicanonical Monte Carlo (MuMC) method was developed by Berg and Neuhaus [6]. They applied this method to simulate first-order phase transitions.

Fourth, the multibaric-multithermal method was developed by Okumura and Okamoto. [7] This method can sample configurations to calculate isobaric-isothermal ensemble averages at any pressure and any temperature.

Many other methods exist in addition to the above.

1.6 References and Notes

[1] METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. J. CHEM. PHYS. 1953, 21, 1087.

[2] HASTINGS, W. K. BIOMETRIKA 1970, 57, 97.

[3] B.J. ALDER, T.E. WAINWRIGHT, J. CHEM. PHYS. 1959, 31, 459.

[4] SWENDSEN, R.; WANG, J. PHYS. REV. LETT. 1986, 57, 2607.

[5] MARINARI, E.; PARISI, G. EUROPHYS. LETT. 1992, 19, 451.

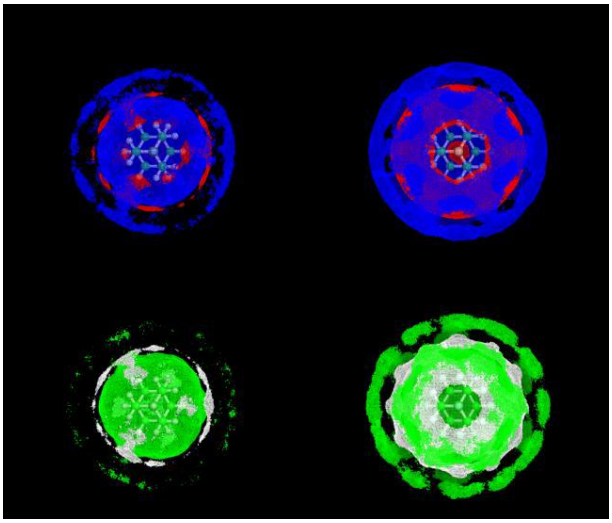
[6] BERG, B.; NEUHAUS, T. PHYS. REV. LETT. 1992, 68, 9.

[7] OKUMURA, H.; OKAMOTO, Y. CHEM. PHYS. LETT. 2004, 383, 391.

Chapter 2.

Hydration around the adamantane and the adamantane derivatives

Hydration of Adamantane Skeleton: Water Assembling around Amantadine and Halo-substituted Adamantanes



2.1 Introduction

Adamantane ($C_{10}H_{16}$) is a highly symmetrical tricyclic aliphatic hydrocarbon, that is the simplest of the diamondoids, and the dipole moment of adamantane is 0 Debye. Adamantane has been known as a valuable building block of polymers to enhance the thermal stability or to improve their physical properties.[1,2] Adamantyl-bearing compounds are also used in the field of medicinal chemistry and drug development, and some of them are in clinical use. [3] Adamantyl group is used to modify the water solubility and the lipophilicity of an original compound that may be highly water soluble or poorly soluble. By incorporating an adamantly group into the structure, the membrane solubility can be changed to a value that may be clinically useful.

The n-octanol/water partition coefficient, log P,[4] is used as an index of the lipophilicity of a compound, and is often used to evaluate the bioavailability of candidate molecules for therapy. A positive value of log P indicates a preference of the compound for lipid phase, and a negative value for water. It is thought that a compound should have a reasonable value of log P to be a good candidate for a new drug. The value of log P of adamantane is positive, yet lower than that of n-decane, which is also composed of 10 carbon atoms. [5] This feature affords adamantyl-bearing compounds more tractable solubility properties, compared to decane-substituted analogues, which may be much less water soluble. [3]

Amantadine (1-aminoadamantane) is known as a blocker of the M2 channel activity of influenza A virus, [6] and also as an antagonist against Parkinson's disease. [7] The estimated values of log P of amantadine [5] and adamantane [3] are 2.22 and 4.24, respectively, indicating that the ratio of the water solubility to the lipophilicity of amantadine is roughly 100 fold larger than that of adamantane.

Although the water solubility of adamantane is very low, it is still slightly 'soluble' in water, which is related to the hydrophobic hydration. [8] Previously we showed [9] that the water molecules in the first solvent shell around adamantane are oriented in such a way that the dipole

direction of water is tangential to the skeleton surface, and the hydrogen bonding network is formed among the water molecules in the first shell as well as between the first shell and the outer shell.

For the development of new drugs and new materials utilizing adamantyl groups, it is needed to estimate how attractive the adamantane skeleton is for water. For this end, we perform Monte Carlo (MC) simulations of some adamantane derivatives surrounded by water molecules, and show how the water distribution changes by a substituent group.

For the MC simulation, we made the molecular mechanics (MM) parameters of the solutes as follows. First, we optimized the geometries of the target molecules, using ab initio molecular orbital (MO) method. The target molecules in this letter are adamantane, 1-chloroadamantane, 1-fluoroadamantane, amantadinium and perfluoroadamantane. Adamantane, as well as perfluoroadamantane, has Td symmetry. Each of the other three adamantane derivatives has C3v symmetry. The MP2/aug-cc-pVDZ level of theory was used for the geometry optimization using Gaussian09 program package. [10] The stationary structures of those molecules were confirmed by means of the normal mode analysis. The optimized structural parameters are in excellent agreement with the experimental values. [11,12] Then we calculated the atomic charges of the adamantane derivatives using natural population analysis (NPA). [13] The structural parameters and the atomic charges obtained from the ab initio MO calculations are listed in the Appendix A. [20] The Lennard-Jones parameters determined by Freindorf and Gao [14] were applied to the atoms of the adamantane derivatives. The TIP3P water potential function [15] was employed for the solvent molecules.

2.2 Computational details

We coded our own program to create the configuration of water molecules so as to form the NVT ensemble (at constant number N, volume V and temperature T) using the MC method based

on the Metropolis-Hastings algorithm. [16] For each system, 500 water molecules were distributed around a solute molecule in a solvation sphere with a radius of 15.4Å. The density of a solute with 500 water molecules in this sphere was 1 g cm^{-3}. The temperature was set at 300K.

For each system, the simulation was done with the following procedure. First, 10^8 MC steps simulation was done for equilibrium. Afterward 2×10^{10} MC steps simulation was performed and used for the analysis.

The number density of an oxygen atom of solvent water is expressed as

$$n_{\mathrm{O}}(x,y,z)=\frac{\langle N_{\mathrm{O}}(x,\delta x,y,\delta y,z,\delta z)\rangle}{\delta x\delta y\delta z}. \qquad (1)$$

The numerator of the right side of Eq. (1) denotes the count average of a solvent O atom in x~x+δx, y~y+δy and z~z+δz. This was sampled every time when all water molecules had moved. The number density is in atoms $Å^{-3}$. For solvent H atom, the number density $n_{\mathrm{H}}(x,y,z)$ was defined similarly.

2.3 Results and Discussion

It should be noted that the O number density, $n_{\mathrm{O}}(x,y,z)$, of bulk water (1 g cm^{-3}) is uniform and 0.033 atoms $Å^{-3}$. In Figure. 1, we plot a red point where the O number density is higher than 0.09 atoms $Å^{-3}$.

It is remarkable that the distribution is not uniform in both of the O distribution (Figure. 1) and the H distribution (Figure. 3) for each of the adamantane derivatives. ‘Front,’ ‘side’ and ‘back’ are indicated in Figure. 2. The water distribution is affected by the interaction between the solute and surrounded water molecules, so that the water distribution possesses a similar symmetry as the solute. These characteristics cannot be seen in the radial distributions of O number density

(Figure. A2) nor H number density (Figure. A3), since a radial distribution gives an averaged view around a solute.

As listed in Table 1, for each of the adamantine derivatives, the position of the first peak of H distribution is close to that of O distribution, indicating that the water molecules in the first hydration shell are tangential to the skeleton surface. This is one of the characteristics of hydrophobic hydration. [17]

The water distribution around adamantane is shown in Figure. 1a for oxygen and Figure. 3a for hydrogen. The first solvation shell is observed: it is composed of "patches" of solvent water molecules which are laid over the adamantane surface. The O number density around the highest population region is almost 3 times higher than that of bulk water. It is noteworthy that the O density of the first hydration shell is higher than the bulk water (see Table 1), indicating that water molecules are assembled around adamantane.

The water distribution around perfluoroadamantane is shown in Figure. 1e for oxygen and Figure. 3e for hydrogen. Since a C-F bond in perfluoroadamantane is around 0.25Å longer than a C-H bond in adamantane (see Tables A1(a) and A2(a)), the molecular diameter of perfluoroadamantane is larger than that of adamantane. Note that the atomic charges of perfluoroadamantane and adamantane are "inside out," as shown in Tables A1(b) and A2(b). The carbon atoms of adamantane are negative, while those of perfluoroadamantane are positive. In spite of the reversed charge distribution, the first solvation shell of perfluoroadamantane looks like a cage composed of patches, in which solvent water molecules are laid over the molecular surface. The O number density of the first hydration shell of perfluoroadamantane is higher than that of adamantane. Perfluoroadamantane attracts water molecules more strongly than adamantane does.

For hydrophobic hydration, the hydration free energy change of alkane can be decomposed into the free energy of the cavity formation and the free energy change due to attractive van der Waals and electrostatic interactions. [18] The former (cavity component) is large and positive, and

the latter is large in magnitude and negative being dominated by the attractive component of the solute-solvent energy change. For alkane, generally, the attractive component is slightly smaller in magnitude than the cavity component resulting in the positive but small hydration free energy. Comparing perfluoroadamantane with adamantane, the cavity component of the former is larger than the latter because the molecular skeleton of the former is larger, and the attractive component of the former is larger in magnitude than the latter because the former attracts more water molecules than the latter.

The water distribution around 1-chloroadamantane is shown in Figure. 1b for oxygen and Figure. 3b for hydrogen. The atom at the center in the front view is chlorine. A circular distribution of water O atom is observed around Cl atom, yet no specific H distribution around Cl is observed. This indicates that the Cl atom is not strong enough to fix the position of water H atoms by hydrogen bonding.

The water distribution around 1-fluoroadamantane is shown in Figure. 1c for oxygen and Figure. 3c for hydrogen. The atom at the center in the front view is fluorine. The F atom in 1-fluoroadamantane attracts water molecules. The distribution of water H atoms around F is divided into two layers, indicating that two hydrogen atoms of a water molecule distribute separately. One of H atoms of a water molecule is strongly hydrogen-bonded to the F atom, and another H atom points outer. This is a characteristic of hydrophilic hydration, although the first hydration shell of the adamantane skeleton has a hydrophobic character, similar to that of adamantane.

The water distribution around amantadinium is shown in in Figure. 1d for oxygen and Figure. 3d for hydrogen. The characteristics in the water distribution around amantadinium is similar to that around 1-fluoroadamantane. The H atoms of the ammonium cation attracts water molecules, which is a characteristic of hydrophilic hydration, although the first hydration shell around the adamantane skeleton exhibits a hydrophobic character, similar to that of adamantane.

These 1-substituted adamantane derivatives (1-chloroadamantane, 1-fluoroadamantane, and amantadinium) can be regarded as amphiphilic molecules. Adamantane and perfluoroadamantane are hydrophobic molecules. Because of the significant increase in water density near the cavity, each of these adamantane derivatives can be said to be wet in a microscopic aspect. [19] The visualization of the number density around a solute makes it possible to show clearly the hydration pattern and the halo-substitution effect around a large molecule such as adamantane.

2.4 References and Notes

1 R. C. FORT JR., P. V. R. SCHLEYER, CHEM.REV. 1964, 64, 277.

2 I. A. NOVAKOV, B. S. ORLINSON, POLYMER SCI.C 2005, 47, 50.

3 J. LIU, R. D. OBANDO, V. LIAO, T. LIFA, R. CODD, EUR. J. MED. CHEM. 2011, 46, 1949.

4 C. HANSCH, P. P. MALONEY, T. FUJITA, R. M. MUIR, NATURE, 1962, 194, 178.

5 P. RUELLE, CHEMOSPHERE, 2000, 40, 457.

6 J. R. SCHNELL, J. J. CHOU, NATURE, 2008, 451, 591.

7 L. V. KALIA, S. K. KALIA, M.W. SALTER, LANCET NEUROLOGY, 2008, 7, 742.

8 R. SCHMID, MON. CHEMIE, 2001, 132, 1295.

9 M. OHISA, M. AIDA, CHEM. PHYS. LETT. 2011, 511, 62.

10 GAUSSIAN 09, REVISION B.1, M. J. FRISCH, G. W. TRUCKS, H. B. SCHLEGEL, G. E. SCUSERIA, M. A. ROBB, J. R. CHEESEMAN, G. SCALMANI, V. BARONE, B. MENNUCCI, G. A. PETERSSON, H. NAKATSUJI, M. CARICATO, X. LI, H. P. HRATCHIAN, A. F. IZMAYLOV, J. BLOINO, G. ZHENG, J. L. SONNENBERG, M. HADA, M. EHARA, K. TOYOTA, R. FUKUDA, J. HASEGAWA, M. ISHIDA, T. NAKAJIMA, Y. HONDA, O. KITAO, H. NAKAI, T. VREVEN, J. A. MONTGOMERY, JR., J. E. PERALTA, F. OGLIARO, M. BEARPARK, J. J. HEYD, E. BROTHERS, K. N. KUDIN, V. N. STAROVEROV, R. KOBAYASHI, J. NORMAND, K. RAGHAVACHARI, A. RENDELL, J. C. BURANT, S. S. IYENGAR, J. TOMASI, M. COSSI, N. REGA, J. M. MILLAM, M. KLENE, J. E. KNOX, J. B. CROSS, V. BAKKEN, C. ADAMO, J. JARAMILLO, R. GOMPERTS, R. E. STRATMANN, O. YAZYEV, A. J. AUSTIN, R. CAMMI, C. POMELLI, J. W. OCHTERSKI, R. L. MARTIN, K. MOROKUMA, V. G. ZAKRZEWSKI, G. A. VOTH, P. SALVADOR, J. J. DANNENBERG, S. DAPPRICH, A. D. DANIELS, Ö. FARKAS, J. B. FORESMAN, J. V. ORTIZ, J. CIOSLOWSKI, AND D. J. FOX, GAUSSIAN, INC., WALLINGFORD CT, 2009.

11 I. HARGITTAI, K. HEDBERG, J. CHEM. SOC. D, 1971, 22, 1499.

12 I. HARGITTAI, J. BRUNVOLL, T. SONODA, T. ABE, H. BABA, J. MOL. STRUCT., 1998, 445, 55.

13 A. E. REED, R. B. WEINSTOCK, F. WEINHOLD, J. CHEM. PHYS. 1985, 83, 735 .

14 M. FREINDORF, J. GAO, J. COMPUT. CHEM., 1996, 17, 386.

15 W. L. JORGENSEN, J. CHANDRASEKHAR, J. D. MADURA, R. W. IMPEY, M. L. KLEIN, J. CHEM. PHYS., 1983, 79, 926.

16 N. METROPOLIS, A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, E. TELLER, J. CHEM. PHYS., 1953, 21, 1087.

17 B. GUILLOT, Y. GUISSANI, S. BRATOS, J. CHEM. PHYS., 1991, 95, 3643.

18 E. GALLICCHIO, M. M. KUBO, R. M. LEVY, J. PHYS. CHEM. B, 2000, 104, 6271.

19 D. CHANDLER, NATURE, 2005, 437, 640.

20 APPENDIX A

2.5 Tables and Figures

FIGURE CAPTIONS

Figure 1. Oxygen number density of solvent water molecules around adamantane derivatives: (a) adamantane, (b) 1-chloroadamantane, (c) 1-fluoroadamantane, (d) amantadinium (e) perfluoroadamantane. Threshold to plot is 0.09 atoms $Å^{-3}$.

Figure 2. Molecular structure of adamantane derivative. 'Front' and 'back' are viewed along the C3 axis of the molecule.

Figure 3. Hydrogen number density of solvent water molecules around adamantane derivatives: (a) adamantane, (b) 1-chloroadamantane, (c) 1-fluoroadamantane, (d) amantadinium (e) perfluoroadamantane. Threshold to plot is 0.12 atoms $Å^{-3}$.

Table 1. The positions (in Å) of the first maximum and of the first minimum for each of the radial distributions of O and H number densities (Figs. A2 and A3), the hydration number, and the O number density (in atoms $Å^{-3}$) of the first hydration shell for each of the adamantane derivatives: (a) adamantane, (b) 1-chloroadamantane, (c) 1-fluoroadamantane, (d) amantadinium and (e) perfluoroadamantane. The hydration number of a target molecule is defined as the number of O atoms in the first hydration shell, which is inside the first minimum of the radial distribution of O number density.

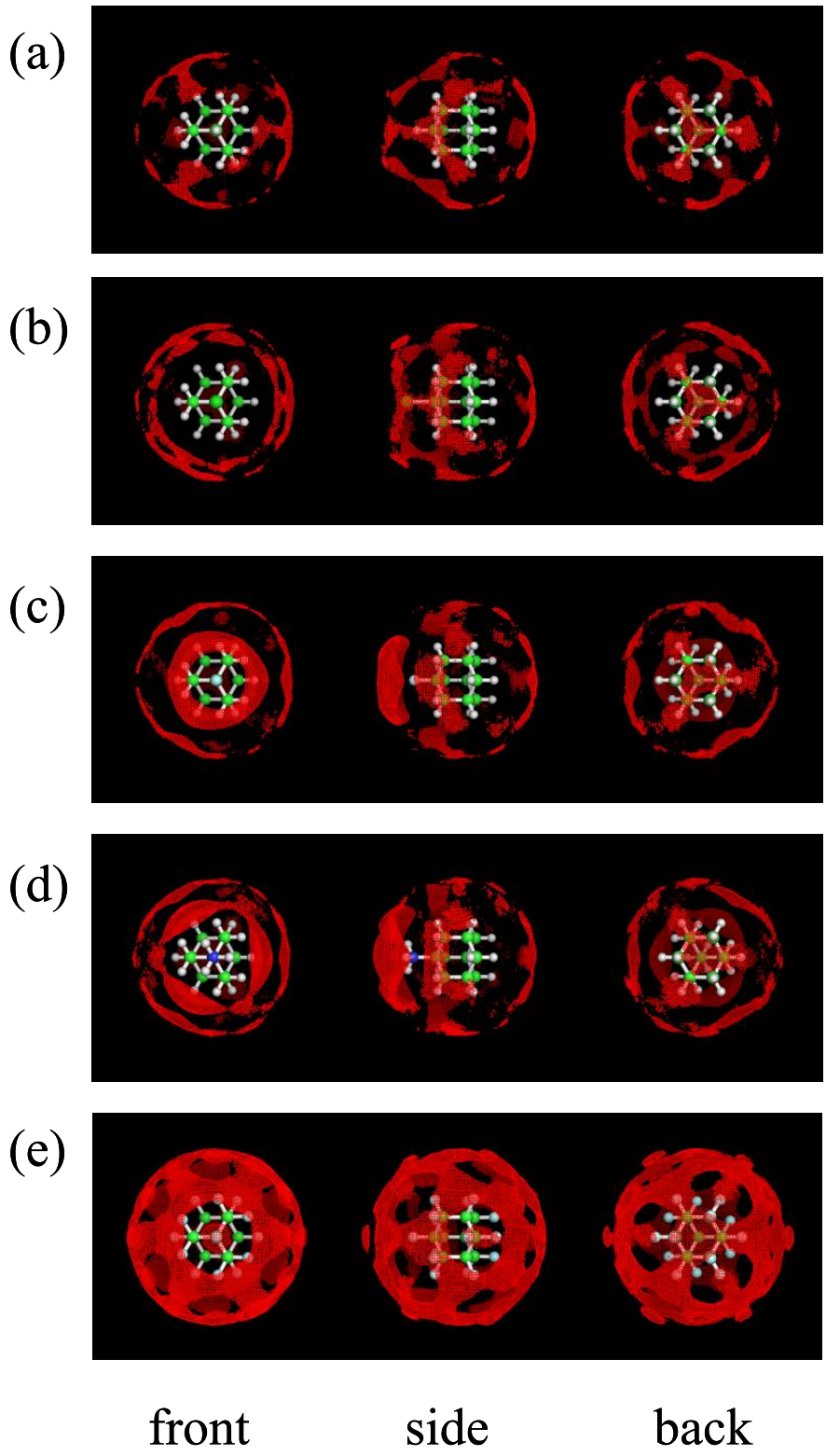


Figure 1.

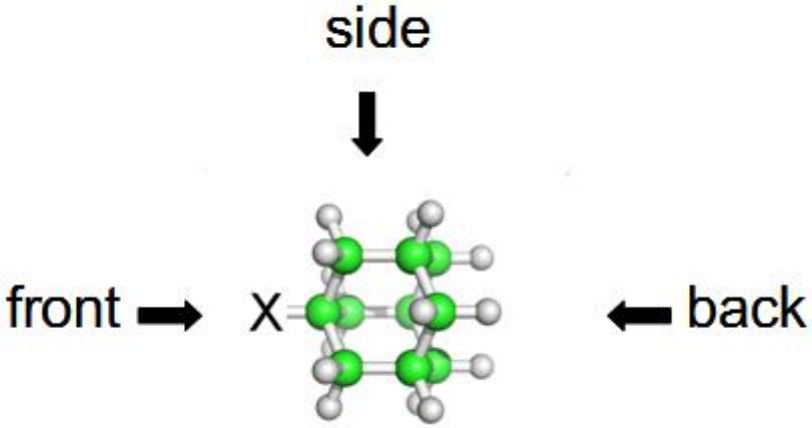


Figure 2.

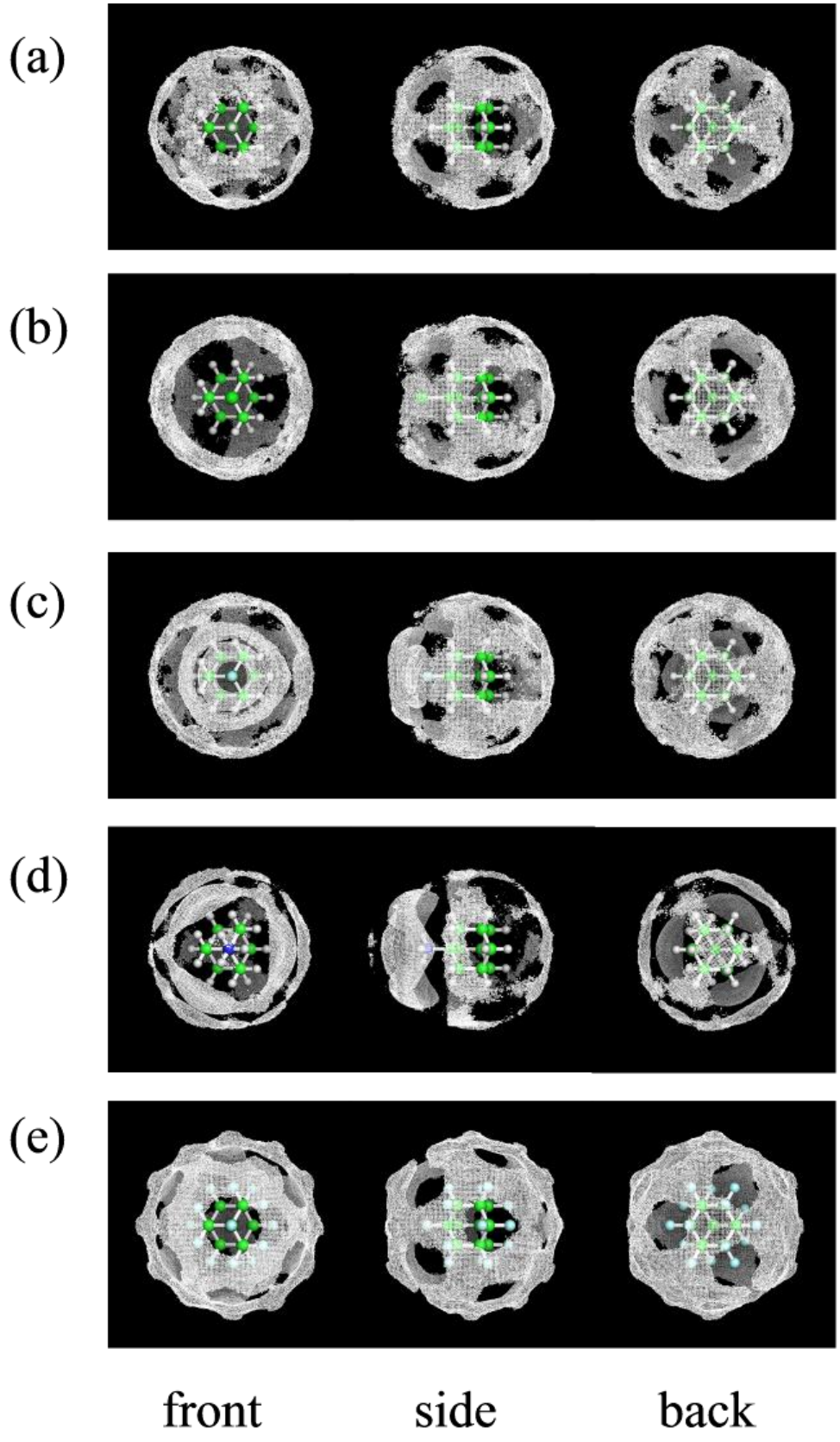


Figure 3.

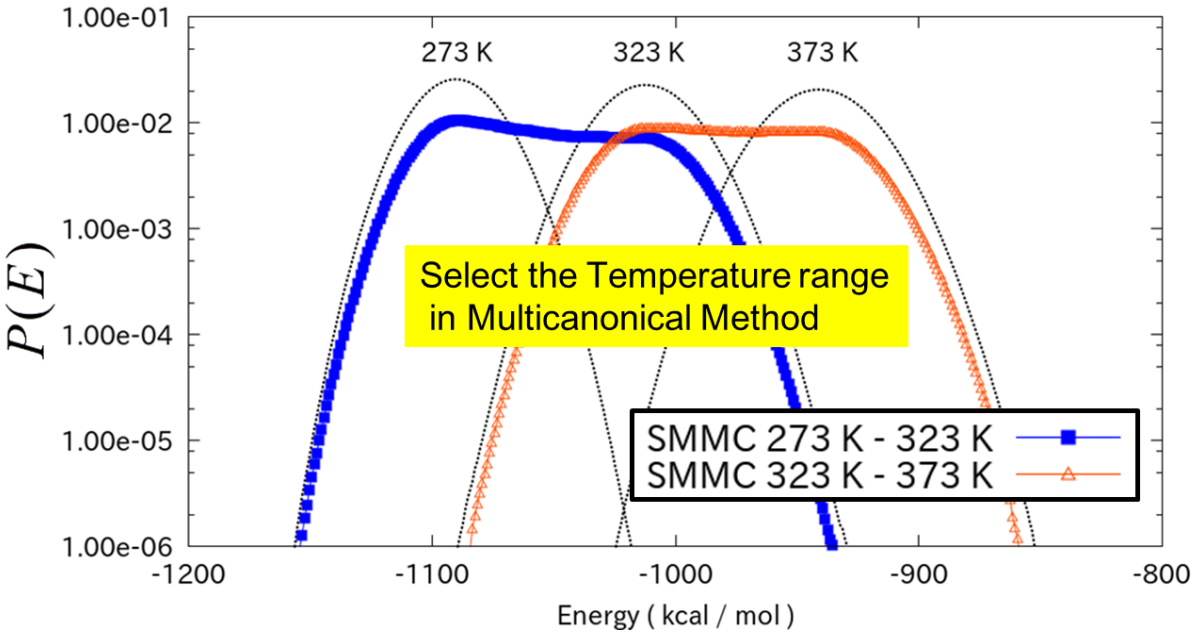
Table 1.

	First maximum		First minimum		Hydration number	Number density
	H	**O**	**O**	**H**		
(a)	5.05	5.15	6.65	6.95	36.7	0.037
(b)	5.15	5.15	6.65	7.05	35.6	0.036
(c)	5.15	5.15	6.65	6.95	36.6	0.037
(d)	5.15	5.15	6.65	6.95	36.4	0.036
(e)	5.45	5.55	7.05	7.15	43.6	0.040

Chapter 3.

The temperature controlling method for multicanonical ensemble and its application to the hydration free energy change

A new variant of multicanonical Monte Carlo algorithm with specifying the temperature range and its application to the hydration free energy change of fluorinated methane derivatives



3.1 Introduction

Computer simulations play important roles in many fields of science nowadays, since high-performance computers are available. In the field of computational chemistry, it is now a kind of routine work to calculate an optimized geometry and its energy of a middle-sized molecule at 0 K by means of ab initio molecular orbital (MO) method with a fairly large basis set taking account of the electron correlation effect. It is still not easy, however, to simulate a distribution of molecular structures, free energy change or its temperature dependence of a complex system. To obtain such physical quantities, two major categories exist in simulation methods. One is molecular dynamics (MD) simulation [1] and the other is Monte Carlo (MC) simulation [2].

In MD simulation, a new configuration is generated by applying Newton's equations of motion to all atoms of a system. Thus an MD method requires the gradients of the potential energy surface to calculate the forces acting between the atoms. The time evolution of a system is obtained from MD simulations, and the correlation function can be calculated [3]. The integration process in MD method is troublesome in a system such as a first order phase transition that has the discontinuity in energy.

In MC simulation, a new configuration is generated by selecting a molecule randomly, followed by translating or rotating it. Whether a new configuration is accepted or not is determined by the Metropolis algorithm [2]. Averages of structures or thermodynamic quantities in an equilibrium state are calculated after creating enough configurations. The MC method requires only the total energy of a system. The gradients of the potential energy surface are not needed. Therefore the MC method is suitable to generate the canonical ensemble and the grand canonical ensemble of a system and can be applied to phase transition phenomena [4].

The canonical Monte Carlo (CaMC) method was developed by Metropolis et al. [2]. By applying this method, an equilibrium state can be obtained at a finite temperature. The multicanonical Monte Carlo (MuMC) method was developed by Berg and Neuhaus [5]. They

applied this method to simulate first-order phase transitions. The canonical-ensemble averages are obtained as functions of temperature by the single-histogram [6] and/or the multiple-histogram [7, 8] reweighting techniques. An extension of the multiple-histogram is the weighted histogram analysis method (WHAM) [8]. The multicanonical method is widely used [9], especially in biological and chemical systems, such as protein folding problem [10] and free energy landscape [11]. Several variants of MuMC method have been developed such as the following. The multibaric–multithermal method can sample configurations to calculate isobaric-isothermal-ensemble averages at any pressure and any temperature [12, 13]. The partial multicanonical method can sample a wide range of a specific part of potential energy [14, 15]. The multi-overlap method gives a flat probability distribution in the multidimensional dihedral-angle distance space [16, 17]. Thus, the fact that many kinds of variants exist indicates that MuMC method is powerful and can be applied to many phenomena.

The multicanonical method has the advantage of generating low energy configurations and high energy configurations equally, making use of non-Boltzmann probability weight factor. It is difficult, however, to obtain the adequate weight factor. The replica-exchange method reduces the difficulty in the weight-factor determination. The number of required replicas, however, increases as the square root of the system size [18]. In most of the methods aiming to create multicanonical ensemble [19, 20], the energy range that is treated as multicanonical ensemble has to be determined. The problem is that the way of this determination is complicated. Since the energy of a system depends on each system, the energy range in each system and each simulation run has to be determined each time.

Here we propose a new method to overcome the difficulty in determining the weight factor, in which not the energy range but the temperature range is controlled in MuMC simulation. It is worthy of note that the same temperature range can be used in many simulation runs to create the multicanonical ensemble of a system. The multicanonical method and the variants which have

been developed hitherto can set the range of *extensive variable* such as energy, while our new method can set the range of temperature, which is *intensive variable*. This feature distinguishes the present method from the other variants. The new method makes the multicanonical method more powerful and more convenient. We show the application of this method to the free energy perturbation (FEP) calculation [21] to compare the hydration free energy changes of methane derivatives.

3.2 Monte Carlo Sampling Algorithms

Here, we first review some familiar methods of canonical Monte Carlo and multicanonical Monte Carlo algorithms. And then we present a new variant of multicanonical Monte Carlo method.

3.2.1 Canonical algorithm

In CaMC sampling, the probability $P_{\mathrm{B}}(E)$ of a system having energy E is given by Eq. (1), where $n(E)$ represents the density of states with energy E, and $W_{\mathrm{B}}(E)$ is the Boltzmann factor, $\exp(-\overline{\beta}E)$. The inverse temperature $\overline{\beta}$ is defined as Eq. (2).

$$P_{\mathrm{B}}(E) \propto n(E) \times W_{\mathrm{B}}(E) \tag{1}$$

$$\overline{\beta} = \frac{1}{kT} \tag{2}$$

Here, k is Boltzmann constant, and T is temperature.

CaMC simulation is performed by the Metropolis algorithm [2]. The probability $P_{\mathrm{B}}(E)$ can be determined by a histogram, $h(E)$, which is obtained from a simulation run, as follows:

$$P_{\mathrm{B}}(E) = \frac{h(E)}{\sum h(E)} \tag{3}$$

The energy space is divided into small bins. Then, $n(E)$ can be determined by Eq. (4).

$$n(E) \propto \frac{P_{\mathrm{B}}(E)}{W_{\mathrm{B}}(E)} \tag{4}$$

3.2.2 Multicanonical algorithm

A multicanonical Monte Carlo simulation is carried out in a way that satisfies the following condition [5, 22]:

$$P_{\mathrm{mu}}(E) \propto n(E) \times W_{\mathrm{mu}} = \text{const.} \tag{5}$$

Here, $W_{\mathrm{mu}}(E)$ is the multicanonical weight factor. According to Eq. (5), the multicanonical weight factor is proportional to the inverse of the density of states, while it is not *a priori* known. Since the weight is given by the Boltzmann factor in the case of CaMC, we can calculate the density of states by performing CaMC simulation and then estimate the multicanonical weight factor [23]. The parameters $\alpha(E)$ and $\beta(E)$ are introduced in the weight factor as follows:

$$W_{\mathrm{mu}}(E) = \exp\{-(\overline{\beta} + \beta(E))E - \alpha(E)\} \tag{6}$$

Taking account of equations (5) and (6), the parameters $\alpha(E_i)$ and $\beta(E_i)$, at discretized E_i, should satisfy the following equation:

$$\beta(E_i)E_i + \alpha(E_i) = \ln h(E_i) + \text{const.} \tag{7}$$

These parameters can be obtained by the following iterative procedure. First, the initial values of $\alpha(E)$ and $\beta(E)$ are set at zero. A trial MuMC simulation at an initial temperature is performed and the probability distribution, the histogram $h(E_i)$, is calculated. The parameters $\alpha(E_i)$ and $\beta(E_i)$ can be estimated from the two adjacent points in the histogram. The subsequent Monte Carlo simulation is carried out with thus obtained weight factor $W_{\mathrm{mu}}(E)$ (Eq. (6)). Then, $h(E)$ is renewed, and the parameters $\alpha(E)$ and $\beta(E)$ are updated from Eq. (7).

The next trial MuMC is carried out with the updated weight factor. This cycle of simulations is repeated until the probability distribution $P_{\mathrm{mu}}(E)$ becomes flat and covers a sufficiently wide range in the energy space. Thereafter the actual MuMC simulation is carried out.

Instead of using both of $\alpha(E)$ and $\beta(E)$, only $\alpha(E)$ can be used: i.e., $\beta(E)$ is fixed to zero. The weight factor is calculated by the linear interpolation of the histograms between neighboring bins in the simulation. In this case, the histogram at each bin derived after the simulation is independently calculated. Sometimes there is no sampling for bins located near the end of the energy range. If sampling is missing inside the histograms, a value is interpolated by linear interpolation between neighboring bins.

The final $P_{\mathrm{mu}}(E)$ is converted to the Boltzmann probability distribution $P_{\mathrm{B}}(E)$ at any temperature, by the following reweighting formula [5, 22]:

$$P_{\mathrm{B}}(E) = \frac{W_{\mathrm{mu}}^{-1}(E) P_{\mathrm{mu}}(E) \exp(-\overline{\beta}E)}{\sum_{\mathrm{E}} W_{\mathrm{mu}}^{-1}(E) P_{\mathrm{mu}}(E) \exp(-\overline{\beta}E)}$$

$$= \frac{n(E) \exp(-\overline{\beta}E)}{\sum_{\mathrm{E}} n(E) \exp(-\overline{\beta}E)} \tag{8}$$

The ensemble average of any physical quantity A at any temperature T is obtained as follows [9]:

$$\langle A \rangle_T = \frac{\sum_E A(E) n(E) \exp(-\overline{\beta}E)}{\sum_E n(E) \exp(-\overline{\beta}E)} \tag{9}$$

3.2.3 A new variant: Selected-Multicanonical Monte Carlo Sampling

In a multicanonical Monte Carlo simulation which is performed often hitherto, the temperature is not defined: it is treated like a parameter. The transition probability from E_i to

E_j $(E_j > E_i)$ in MuMC simulation is described as $W_{\mathrm{mu}}(E_j)/W_{\mathrm{mu}}(E_i)$. Here we introduce our own variant where we use the temperature range explicitly to control a simulation.

The temperature-like parameter $T'(i,j)$ and $\overline{\beta}$-like parameter $\beta'(i,j)$ are defined by equations (10) and (11). In a simulation, $\overline{\beta}$ is set at the initial temperature. The transformation of Eq. (10) gives Eq. (12).

$$-\beta'(i,j)(E_j - E_i) = -(\overline{\beta} + \beta(E_j))E_j - \alpha(E_j) - \{-(\overline{\beta} + \beta(E_i))E_i - \alpha(E_i)\} \tag{10}$$

$$\beta'(i,j) = \frac{1}{kT'(i,j)} \tag{11}$$

$$\beta'(i,j) = \overline{\beta} + \frac{\beta(E_j)E_j - \beta(E_i)E_i}{E_j - E_i} + \frac{\alpha(E_j) - \alpha(E_i)}{E_j - E_i} \tag{12}$$

In MuMC simulation in which we use the ensembles with a temperature range from minimum T_{min} ($\overline{\beta}_{\mathrm{min}}$) to maximum T_{max} ($\overline{\beta}_{\mathrm{max}}$), we impose the following condition (Eq. (13)) for any i and j, (here, $E_i < E_j$).

$$\overline{\beta}_{\mathrm{max}} \le \beta'(i,j) \le \overline{\beta}_{\mathrm{min}} \tag{13}$$

The condition is equivalent to Eq. (14). See the Appendix B for the derivation.

$$\overline{\beta}_{\mathrm{max}} - \overline{\beta} \le \frac{\beta(E_{i+1})E_{i+1} - \beta(E_i)E_i}{E_{i+1} - E_i} + \frac{\alpha(E_{i+1}) - \alpha(E_i)}{E_{i+1} - E_i} \le \overline{\beta}_{\mathrm{min}} - \overline{\beta} \tag{14}$$

The energy E_{i+1} is a neighboring bin of higher energy. Here, instead of using $\alpha(E)$ and $\beta(E)$, we use only $\alpha(E)$: i.e., $\beta(E)$ is fixed to zero.

The gradient of $\alpha(E)$ is defined as follows.

$$\alpha'(E_i) = \frac{\alpha(E_{i+1}) - \alpha(E_i)}{E_{i+1} - E_i} \tag{15}$$

The parameters $\alpha_{\mathrm{imp}}'(E)$ and $\alpha_{\mathrm{imp}}(E)$ are the parameters $\alpha'(E)$ and $\alpha(E)$ that satisfy Eq. (14), and these are defined as follows:

$$\alpha'_{\mathrm{imp}}(E_i) = \begin{cases} \overline{\beta}_{\min} - \overline{\beta} & \text{if } \alpha'(E_i) > \overline{\beta}_{\min} - \overline{\beta} \\ \overline{\beta}_{\max} - \overline{\beta} & \text{if } \alpha'(E_i) < \overline{\beta}_{\max} - \overline{\beta} \\ \alpha'(E_i) & \text{otherwise} \end{cases}$$
$$\alpha_{\mathrm{imp}}(E_{i+1}) = (E_{i+1} - E_i)\alpha'_{\mathrm{imp}}(E_i) + \alpha_{\mathrm{imp}}(E_i) \tag{16}$$

Then, we have the following relation:

$$\overline{\beta}_{\max} \le \overline{\beta} + \alpha'_{\mathrm{imp}}(E) \le \overline{\beta}_{\min} \tag{17}$$

We define the selected-multicanonical weight factor $W_{\mathrm{smu}}(E)$ as follows:

$$W_{\mathrm{smu}}(E) = \exp\{-\overline{\beta}E - \alpha_{\mathrm{imp}}(E)\} \tag{18}$$

This method controls the energy range treated as multicanonical through controlling the weight factor $W_{\mathrm{smu}}(E)$. Therefore it is necessary to specify not the energy range but the temperature range. The energy depends on a system concerned. The temperature does not depend on a specific system. Moreover the ensemble never goes out of the temperature range. This is the advantage over any other methods which specify the range of energy. This method generates the ensemble by 'selecting' the configurations which fit the desired temperature range. We call the new method "selected-multicanonical" Monte Carlo (SMMC).

It can happen that there is no sampling for a bin located near the end of the energy range. In a low energy range where sampling is missing, the weight factor is treated at temperature $T_{\min}$. In a high energy range where sampling is missing, the weight factor is treated at temperature $T_{\max}$.

In SMMC simulation, the subsequent Monte Carlo simulation is carried out with the selected-multicanonical weight factor $W_{\mathrm{smu}}(E)$ instead of $W_{\mathrm{mu}}(E)$. Other processes in a simulation are the same as those in MuMC simulation (Section 2-2). The final probability

distribution $P_{\mathrm{smu}}(E)$ is converted to the Boltzmann probability distribution $P_{\mathrm{B}}(E)$ at temperature from $T_{\min}$ to $T_{\max}$ by Eq. (8), as in MuMC simulation. The processes except calculating the weight factor are the same as those in MuMC simulation. The implementation of the small modification as shown here in a MuMC simulation code makes the multicanonical method more usable.

3.3 Computational Methods

3.3.1 Free Energy Perturbation Calculation

In CaMC simulation of a system at temperature T and an inverse temperature $\overline{\beta}$, the difference in Helmholtz energy between a reference system i and a target system j, ΔA^{ij}, is given by

$$\Delta A^{ij} = \frac{\ln\left\langle \exp\{-\overline{\beta}(E^j - E^i)\}\right\rangle_{(i)}}{-\overline{\beta}} \tag{19}$$

The bracket $\langle \ldots \rangle_{(i)}$ denotes the ensemble average. To sample the configurations, we adopt the "double wide" sampling [24].

In SMMC simulation, we calculate ΔA^{ij} at temperature T_m $(m = 1, \cdots M)$ and $\overline{\beta}_{T_m}$ by the following 3 step procedure.

1. $\sum \exp\{-\overline{\beta}_{T_m}(E^j - E_k^i)\}$ is calculated at discretized E_k by summing up $\exp\{-\overline{\beta}_{T_m}(E^j - E_k^i)\}$ value at each Monte Carlo step.

2. The average of $\exp\{-\overline{\beta}_{T_m}(E^j - E_k^i)\}$ is calculated by dividing it by $h(E_k)$.

3. The ensemble average of $\left\langle \exp\{-\overline{\beta}_{T_m}(E^j - E^i)\}\right\rangle_{(i)}$ at temperature T_m is calculated by using Eq. (9).

We coded our own MC program to perform FEP calculations with CaMC and SMMC sampling algorithms. To show the applicability of the SMMC simulation, we used this program to obtain the hydration Helmholtz energy changes of methane derivatives by CaMC as well as by SMMC. We compare the computational results of both methods, and confirm the performance of the SMMC method, as will be shown below.

3.3.2 Hydration Helmholtz Energy Changes of Methane Derivatives

To perform MC simulation, molecular mechanics (MM) parameters of the solutes were created as follows. First, we optimized the geometries of the target molecules, using ab initio molecular orbital (MO) method. The target molecules in this work are methane, fluoromethane, difluoromethane, fluoroform and tetrafluoromethane. Methane and tetrafluoromethane have T_d symmetry. Fluoromethane and fluoroform have C_{3v} symmetry. Difluoromethane has C_{2v} symmetry. The MP2/aug-cc-pVDZ level of theory was used for the geometry optimization using Gaussian09 program package [25]. The stationary structures of those molecules were confirmed by the normal mode analysis. Then, we calculated the atomic charges of the methane derivatives by means of natural population analysis (NPA) [26]. The obtained charges are shown in the Appendix B. The Lennard-Jones parameters determined by Freindorf and Gao [27] were applied to the atoms of methane derivatives. The Lennard-Jones parameters between a solute atom and a solvent atom were decided by the geometric mean. The TIP3P water potential function [28] was employed for the solvent molecules. In Monte Carlo simulation, we use 125 TIP3P water molecules in a solvation sphere with a radius of 9.65 Å, in which the density of the system is 1 g cm^{-3}. In CaMC simulation, the temperature was set at from 273 K to 373 K with an increment of 25 K. In SMMC simulation, the temperature ranges were set at 273 K~323 K, 298 K~348 K, 323 K~373 K and 273 K~373 K. The initial temperature was set at the middle temperature in the

temperature range for each case. For example, for a simulation of the temperature range 273 K~323 K, the initial temperature was set at 298 K.

For each system, CaMC simulation was performed with the following procedure. First, a simulation with 5×10^7 MC steps was performed for equilibrium. Afterward, a simulation of 5×10^9 MC steps was performed and used for the analysis. The SMMC simulation was performed with the following procedure. First, a simulation with 10^8 MC steps was performed to calculate the weight factor. We performed this simulation 10 times. In the 10th time of the simulation, we confirmed that the probability distribution became flat and covered a wide range in the energy space. Afterward, a simulation of 10^{10} MC steps was performed and used for the analysis.

The FEP calculation was carried out with the following procedure in CaMC calculation. First, a simulation with 5×10^7 MC steps was performed for equilibrium. Afterward, a simulation of 4×10^9 MC steps was performed and used for the analysis.

The FEP calculation with SMMC simulation was performed with the following procedure. First, a simulation with 10^8 MC steps was performed to calculate the weight factor. We performed this simulation 10 times. In the 10th time of the simulation, we confirmed that the probability distribution became flat and covered a wide range in the energy space. Afterward, a simulation of 10^9 MC steps was performed and used for the analysis.

In FEP calculations, a coupling parameter λ_i was used to linearly scale all geometrical and force-field parameters from the parameters χ_{Ini} for methane to the parameters χ_{Fin} for methane derivative as follows:

$$\chi_i = \lambda_i \chi_{\mathrm{Fin}} + (1-\lambda_i)\chi_{\mathrm{Ini}} \qquad (20)$$

The inclement of the coupling parameter is set as $\Delta\lambda_i = 0.05$.

3.4 Results and Discussion

3.4.1 Comparison between Canonical Monte Carlo and Selected-Multicanonical Monte Carlo Simulations

First, we show whether the SMMC method works well or not. To this end, we compare the simulation results of CaMC and SMMC using a system of methane with 125 water molecules.

Figure 1 shows the probability distributions by CaMC and SMMC simulations at some temperatures between 273 K and 373 K. Each probability distribution that was generated with SMMC method at a certain temperature range is flat and covers the energy range corresponding to the canonical ensembles at the temperature range.

In Figure 2, the values of $\overline{\beta}_{\text{imp}}$ in SMMC simulations of various temperature ranges are plotted. $\overline{\beta}_{\text{imp}}$ is defined as follows:

$$\overline{\beta}_{\text{imp}} = \overline{\beta} + \alpha'_{\text{imp}}(E) \tag{21}$$

The values of $\overline{\beta}$ at 273 K, 298 K, 323 K, 348 K and 373 K are 1.84, 1.69, 1.56, 1.45 and 1.35 mol/kcal, respectively [29]. According to Eq. (17), the following conditions should be satisfied:

SMMC simulation from 273 K to 323 K: $1.56 \le \overline{\beta}_{\text{imp}} \le 1.84$,

SMMC simulation from 323 K to 373 K: $1.35 \le \overline{\beta}_{\text{imp}} \le 1.56$,

SMMC simulation from 298 K to 348 K: $1.45 \le \overline{\beta}_{\text{imp}} \le 1.69$, and

SMMC simulation from 273 K to 373 K: $1.35 \le \overline{\beta}_{\text{imp}} \le 1.84$.

As shown in Figure 2, the parameter $\overline{\beta}_{\text{imp}}$ is actually in the specified range in each simulation. The SMMC method works correctly: It complies with the range of simulation temperature that we specify, and generates only the desired ensemble.

Figure 3 shows the probability distribution which is obtained by CaMC simulation at each temperature, and the reweighted probability distribution which is obtained by SMMC simulation at the temperature range from 273 K to 373 K. The probability distribution at each temperature that was calculated from the SMMC simulation is in very good agreement with the probability distribution from CaMC simulation at the corresponding temperature.

The mathematical formulas of SMMC (Section 2.3.) resemble those of replica-exchange multicanonical algorithm (REMUCA) [30, 31]. In the case of REMUCA, the energy range is specified in a simulation, and the temperature is not cared during the simulation. In the case of SMMC, the temperature range is specified without caring about the energy range. Both methods may give a same result only when the configuration space which is sampled corresponding to the specified energy range in REMUCA happens to be the same as the configuration space which is sampled corresponding to the specified temperature range in SMMC.

Next, we show if SMMC works well in free energy calculations. To this end, we calculated and compared the Helmholtz energy difference between methane and fluoromethane in aqueous solution using both of CaMC and SMMC. As shown in Figure 4, the computational free energy differences at some selected temperatures using CaMC compare very well with those using SMMC at the corresponding temperatures.

3.4.2 Differences in Hydration Helmholtz Energy of Fluorinated Methanes

We calculated the differences of free energy ΔA from CH_4 to CH_3F, CH_3F to CHF_3 and CHF_3 to CF_4 with SMMC method. The results are summarized in Table 1. The simulation results reweighted at some selected temperatures in 273 K~373 K are shown in the Appendix B. The experimental values were estimated from the differences of the standard-state free energy of solvation ΔG [32].

Hydration free energy of CH_3F relative to that of CH_4 is negative, while that of CF_4 relative to CHF_3 is positive. One fluorine atom substituted in CH_4 causes the stability in hydration free energy, while one hydrogen atom substituted in CF_4 causes the stability in hydration free energy. The simulation results of the differences in the hydration free energy caused by fluorination are qualitatively consistent with the experimental values.

3.5 Conclusion

We developed a new variant of multicanonical method, SMMC, in which we set the temperature range when we obtain the weight factor. Deciding the energy range to obtain the weight factor is not necessary. The probability distribution by SMMC method is flat and covers a certain range in energy space. The SMMC method is powerful and can be used to generate and select the ensemble in any other multicanonical MC techniques, such as multicanonical replica-exchange method [33] or other variants [12-17].

3.6 References and Notes

[1] B.J. ALDER, T.E. WAINWRIGHT, J. CHEM. PHYS. 31 (1959) 459.

[2] N. METROPOLIS, A.W. ROSENBLUTH, M.N. ROSENBLUTH, A.H. TELLER, E. TELLER, J. CHEM. PHYS. 21 (1953) 1087.

[3] J.T. TITANTAH, M. KARTTUNEN, J. AM. CHEM. SOC. 134 (2012) 9362.

[4] L.A. ROWLEY, D. NICHOLSON, N.G. PARSONAGE, J. COMPUT. PHYS. 17 (1975) 401.

[5] B.A. BERG, T. NEUHAUS, PHYS. LETT. B 267 (1991) 249.

[6] A.M. FERRENBERG, R.H. SWENDSEN, PHYS. REV. LETT. 61 (1988) 2635.

[7] A.M. FERRENBERG, R.H. SWENDSEN, PHYS. REV. LETT. 63 (1988) 1195.

[8] S. KUMAR, D. BOUZIDA, R.H. SWENDSEN, P.A. KOLLMAN, J.M. ROSENBERG, J. COMPUT. CHEM. 13 (1992) 1011.

[9] Y. OKAMOTO, J. MOL. GRAPH. MODEL. 22 (2004) 425.

[10] J.P. ULMSCHNEIDER, W.L. JORGENSEN, J. AM. CHEM. SOC. 126 (2004) 1849.

[11] K. SAYANO, H. KONO, M.M. GROMIHA, A. SARAI, J. COMPUT. CHEM. 21 (2000) 954.

[12] H. OKUMURA, Y. OKAMOTO, CHEM. PHYS. LETT. 383 (2004) 391.

[13] H. OKUMURA, Y. OKAMOTO, PHYS. REV. E 70 (2004) 026702.

[14] H. OKUMURA, J. CHEM. PHYS. 129 (2008) 124116.

[15] H. OKUMURA, PHYS. CHEM. CHEM. PHYS. 13 (2011) 114.

[16] B.A. BERG, H. NOGUCHI, Y. OKAMOTO, PHYS. REV. E 68 (2003) 036126.

[17] S.G. ITOH, Y. OKAMOTO, CHEM. PHYS. LETT. 400 (2004) 308.

[18] K. HUKUSHIMA, K. NEMOTO, J. PHYS. SOC. JAPAN 65 (1996) 1604.

[19] T. TERADA, Y. MATSUO, A. KIDERA, J. CHEM. PHYS. 118 (2003) 4306.

[20] R. JONO, Y. WATANABE, K. SHIMIZU, T. TERADA, J. COMPUT. CHEM. 31 (2010) 1168.

[21] R.W ZWANZIG, J. CHEM. PHYS. 22 (1954) 1420.

[22] B. A. BERG, T. NEUHAUS, PHYS. REV. LETT. 68 (1992) 9.

[23] U.H.E. HANSMANN, Y. OKAMOTO, ANN. REV. COMP. PHYS 6 (1999) 129.

[24] W.L. JORGENSEN, C. RAVIMOHAN, J. CHEM. PHYS. 83 (1985) 3050.

[25] M.J. FRISCH ET AL., GAUSSIAN 09, REVISION C.01, GAUSSIAN INC., WALLINGFOLD, CT, 2009.

[26] A.E. REED, R.B. WEINSTOCK, F. WEINHOLD, J. CHEM. PHYS. 83 (1985) 735.

[27] M. FREINDORF, J. GAO, J. COMPUT. CHEM. 17 (1996) 386.

[28] W.L. JORGENSEN, J. CHANDRASEKHAR, J.D. MADURA, R.W. IMPEY, M.L. KLEIN, J. CHEM. PHYS. 79 (1983) 926.

[29] P.J. MOHR, B.N. TAYLOR, D.B. NEWELL, REV. MOD. PHYS. 84 (2012) 1527.

[30] A. MITSUTAKE, Y. SUGITA, Y. OKAMOTO, J. CHEM. PHYS. 118 (2003) 6664.

[31] A. MITSUTAKE, Y. SUGITA, Y. OKAMOTO, J. CHEM. PHYS. 118 (2003) 6676.

[32] J. HINE, P.K. MOOKERJEE, J. ORG. CHEM. 40 (1975) 292.

[33] Y. SUGITA, Y. OKAMOTO, CHEM. PHYS. LETT. 329 (2000) 261.

3.7 Tables and Figures

Table 1. Relative hydration free energies (in kcal/mol) of fluorinated methane derivatives.

	ΔA (calc)	ΔG (exp) [a)]
CH_4 to CH_3F	–8.2	–2.2
CH_3F to CHF_3	–1.8	1.0
CHF_3 to CF_4	3.7	2.4

a) Experimental values taken from [32].

Figure captions

Figure 1. The probability distributions by CaMC and SMMC simulation runs for the interaction between methane and water molecules. The results of CaMC (5×10^9 MC steps) are shown by the dotted lines. The simulation temperature for each CaMC simulation is written near the dotted lines. The results of SMMC are shown by the colored symbols. The width of the energy bin is 1.0 kcal/mol.

Figure 2. The parameters of $\overline{\beta}_{\mathrm{imp}}$ obtained by a preliminary Monte Carlo simulation are shown by the colored solid lines. For comparison, the probability distributions from CaMC (5×10^9 MC steps) at 273 K, 323 K and 373 K are shown with the dotted lines.
(a) The SMMC simulations at the temperature ranges of 273 K ~ 323 K and 323 K ~ 373 K.
(b) The SMMC simulations at the temperature ranges of 298 K ~ 348 K and 273 K ~ 373 K.

Figure 3. The reweighted probability distribution from SMMC and the probability distribution in CaMC simulation. Solid gray line indicates the result of CaMC simulation at each temperature. The colored symbols indicate the reweighted histograms that are reweighted at each temperature. The reweighted histograms were calculated from the SMMC simulation results at 273 K~ 373 K.

Figure 4. The Helmholtz energy difference ΔA between methane and fluoromethane in aqueous solution. The blue circle indicates the result from CaMC simulation at each temperature. The solid line with red triangles indicates the result from SMMC simulation at the temperature range of 273 K ~ 373 K.

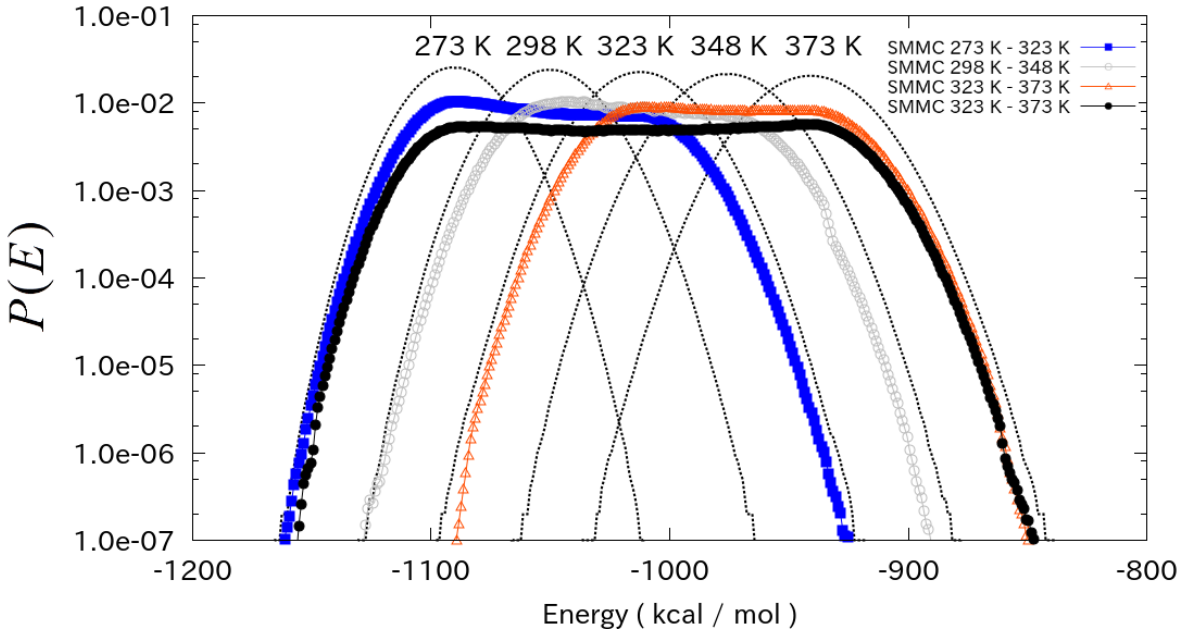

1.0e-01
1.0e-02
1.0e-03
1.0e-04
1.0e-05
1.0e-06
1.0e-07
P(E)
273 K 298 K 323 K 348 K 373 K
SMMC 273 K - 323 K
SMMC 298 K - 348 K
SMMC 323 K - 373 K
SMMC 323 K - 373 K
-1200
-1100
-1000
-900
-800
Energy (kcal / mol)

Figure 1.

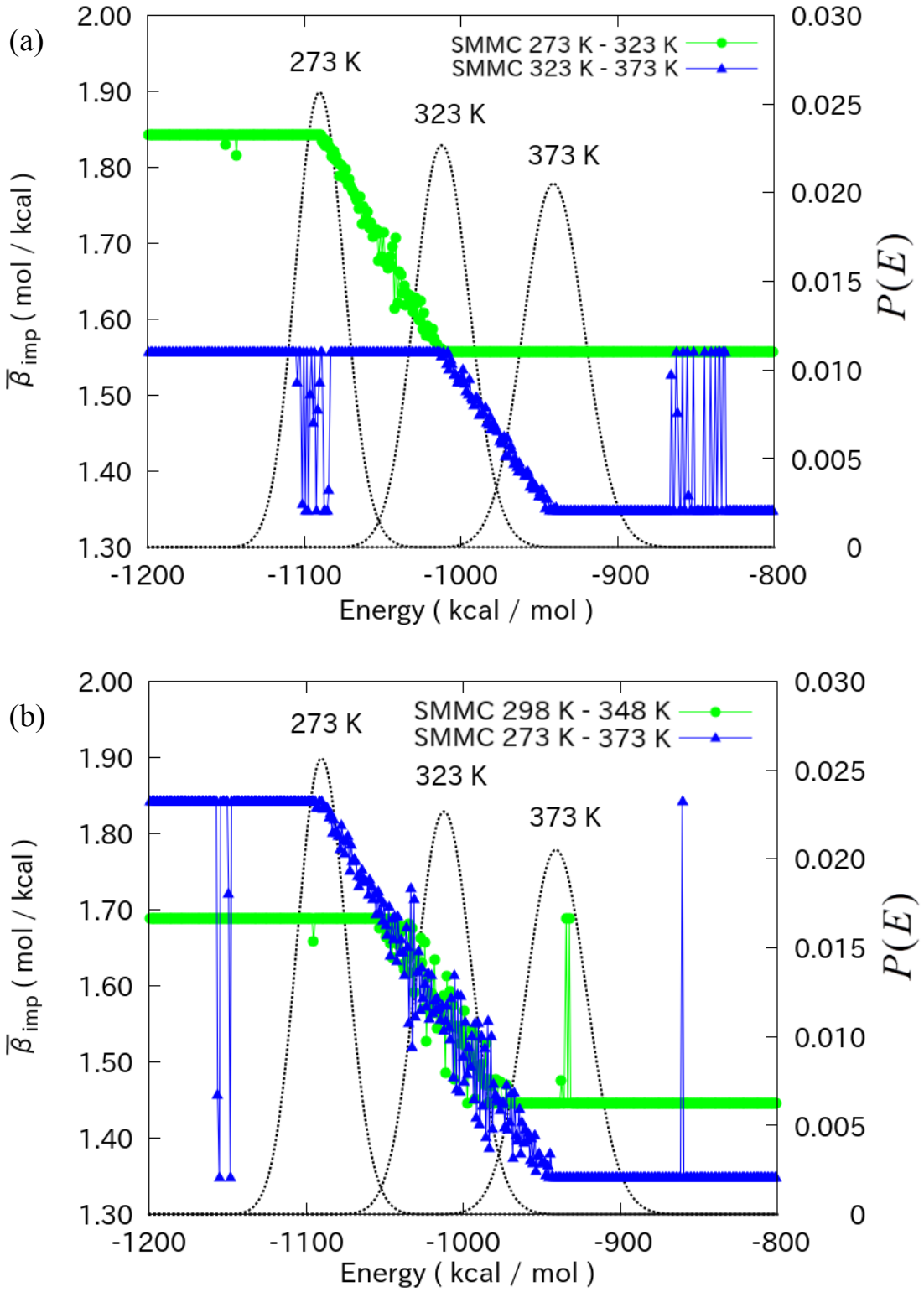

(a)
273 K
323 K
373 K
SMMC 273 K - 323 K
SMMC 323 K - 373 K
2.00
1.90
1.80
1.70
1.60
1.50
1.40
1.30
0.030
0.025
0.020
0.015
0.010
0.005
0
-1200
-1100
-1000
-900
-800
Energy (kcal / mol)
$\overline{\beta}_{imp}$ (mol / kcal)
$P(E)$
(b)
273 K
323 K
373 K
SMMC 298 K - 348 K
SMMC 273 K - 373 K
2.00
1.90
1.80
1.70
1.60
1.50
1.40
1.30
0.030
0.025
0.020
0.015
0.010
0.005
0
-1200
-1100
-1000
-900
-800
Energy (kcal / mol)
$\overline{\beta}_{imp}$ (mol / kcal)
$P(E)$

Figure 2.

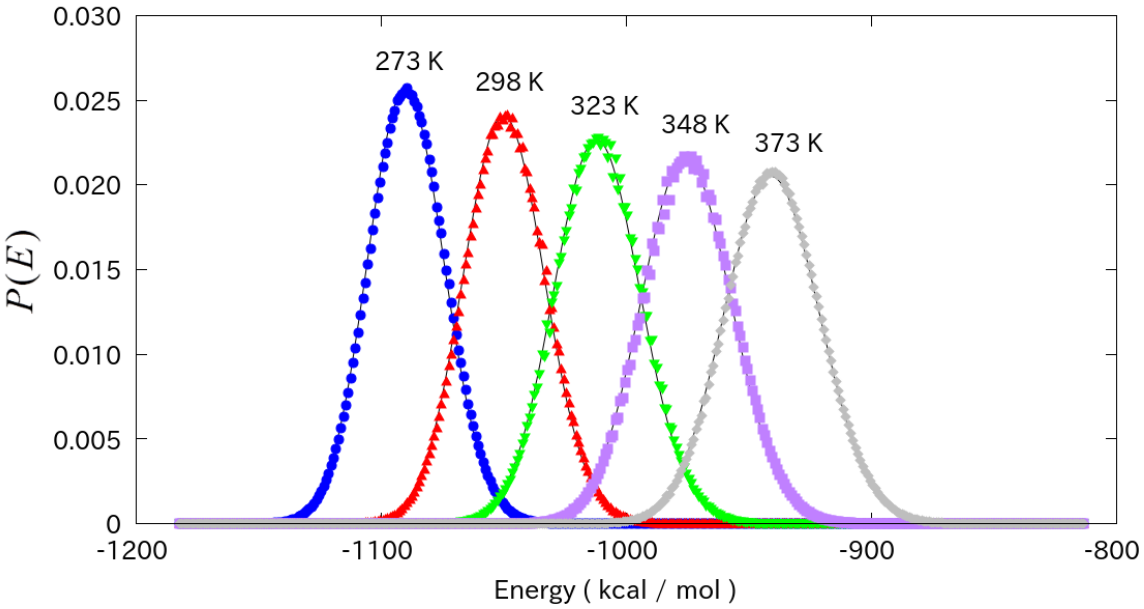

0.030
0.025
0.020
0.015
0.010
0.005
0
P(E)
273 K
298 K
323 K
348 K
373 K
-1200
-1100
-1000
-900
-800
Energy (kcal / mol)

Figure 3.

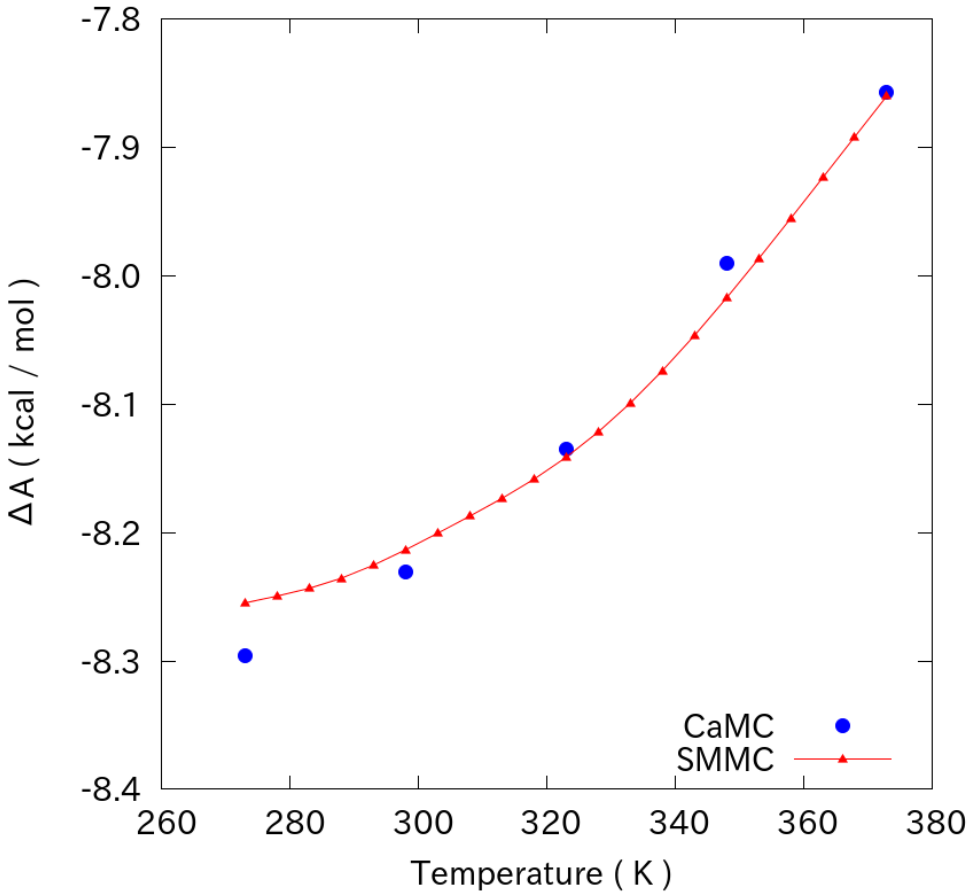

-7.8
-7.9
-8.0
-8.1
-8.2
-8.3
-8.4
ΔA (kcal / mol)
260
280
300
320
340
360
380
Temperature (K)
CaMC
SMMC

Figure 4.

Chapter 4.

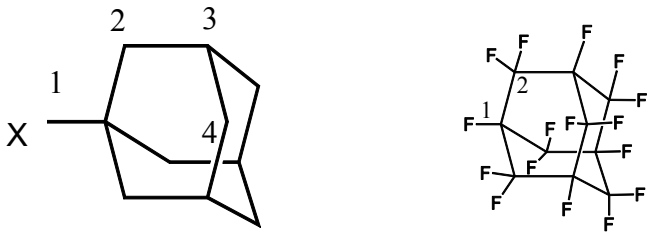
General Conclusion

In this thesis, I showed how to represent solvation structure around organic molecule. I showed that the substitution effect affects the hydration structure. I calculated the hydration Helmholtz energy change between the non-substituted molecule and substituted molecules. I compared the change of the hydration structures with the change of the hydration Helmholtz energy changes. The hydration structures are consistent with the Helmholtz energy changes.

I developed the method that is suited for calculation of physical quantity that depending on the temperature. This method is useful for not only computational chemists but also experimentalists.

Appendix A

(a) (b)



X = H, Cl, F, NH_3^+

Figure A1. Target molecules and atomic numbering

(a) X=H adamantane (T_d symmetry)

X=Cl 1-adamantyl chloride (C_{3v} symmetry)

X=F 1-adamantyl fluoride (C_{3v} symmetry)

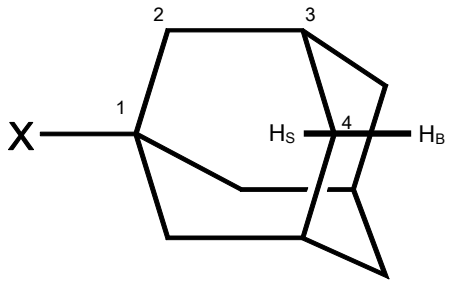
X=NH_3^+ amantadinium cation (C_{3v} symmetry)

(b) perfluoroadamantane (T_d symmetry)

Table A1. Bond lengths (in Å) and atomic charges of adamantane derivatives.

(a) Calculated Bond lengths (in Å) of adamantane derivatives (MP2/aug-cc-pVDZ). Experimental data are taken from ref.11.

	N–H	C_1–X	C_1–C_2	C_2–H_2	C_2–C_3	C_3–H_3	C_3–C_4	C_4–H_S	C_4–H_B
X = H		1.1044	1.5401	1.1051					
(exp.[11])		1.112	1.542	1.112					
X = Cl		1.8258	1.5299	1.1034	1.5425	1.1038	1.5396	1.1045	1.1045
X = F		1.4256	1.5250	1.1032	1.5427	1.1035	1.5402	1.1045	1.1044
X = NH_3^+	1.0286	1.5288	1.5299	1.1057	1.5440	1.1017	1.5397	1.1040	1.1023



(b) Calculated atomic charges (NPA) of adamantane derivatives (MP2/aug-cc-pVDZ).

	H (in X)	X	C_1	C_2	H_2	C_3	H_3	C_4	H_S	H_B
X = H		0.230	–0.264	–0.420	0.221					
X = Cl		–0.061	–0.056	–0.441	0.237	–0.261	0.238	–0.422	0.225	0.226
X = F		–0.402	0.355	–0.455	0.233	–0.262	0.236	–0.421	0.225	0.225
X = NH_3^+	0.456	–0.726	0.107	–0.448	0.239	–0.264	0.260	–0.424	0.260	0.248

Table A2. Bond lengths (in Å) and atomic charges of perfluoroadamantane.

(a) Calculated Bond lengths (in Å) of perfluoroadamantane (MP2/aug-cc-pVDZ). Experimental data are taken from ref.12.

	C_1–F_1	C_1–C_2	C_2–F_2
perfluoroadamantane	1.3692	1.5520	1.3555
(exp.[12])	1.363	1.560	1.340

(b) Calculated NPA charges of adamantane derivatives (MP2/aug-cc-pVDZ).

	C_1	F_1	C_2	F_2
perfluoroadamantane	0.193	-0.321	0.746	-0.330

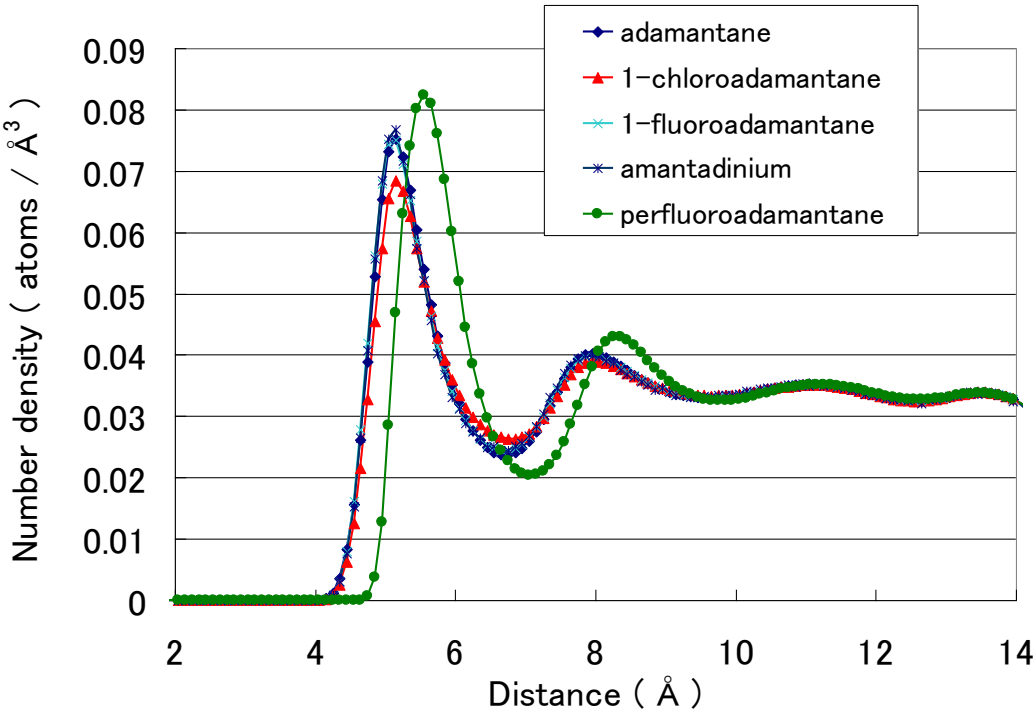


Figure A2. The oxygen radial distribution functions of water. The origin is the center of the carbon atoms of adamantane skeleton for each of the adamantane derivatives.

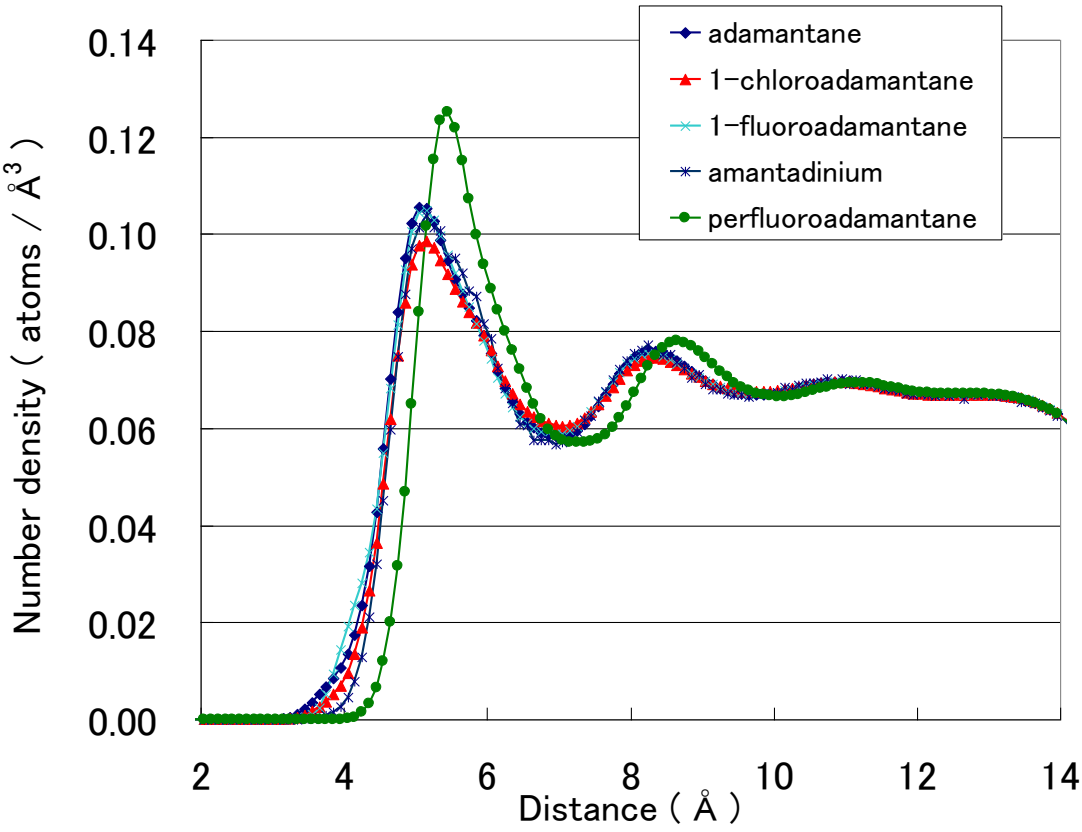


Figure A3. The hydrogen radial distribution functions of water. The origin is the center of the carbon atoms of adamantane skeleton for each of the adamantane derivatives.

Appendix B

B-1: Derivation of eq. (14)

The temperature-like parameter $T'(i,j)$ and $\overline{\beta}$-like parameter $\beta'(i,j)$ are defined by equations (10) and (11), and eq. (10) is transformed into eq. (12).

$$\beta'(i,j) = \overline{\beta} + \frac{\beta(E_j)E_j - \beta(E_i)E_i}{E_j - E_i} + \frac{\alpha(E_j) - \alpha(E_i)}{E_j - E_i} \tag{12}$$

We impose the condition for any *i* and *j*, as in eq. (13):

$$\overline{\beta}_{\max} \le \beta'(i,j) \le \overline{\beta}_{\min} \tag{13}$$

It is equivalent to eq. (14):

$$\overline{\beta}_{\max} - \overline{\beta} \le \frac{\beta(E_{i+1})E_{i+1} - \beta(E_i)E_i}{E_{i+1} - E_i} + \frac{\alpha(E_{i+1}) - \alpha(E_i)}{E_{i+1} - E_i} \le \overline{\beta}_{\min} - \overline{\beta} \tag{14}$$

The derivation of this equation is as follows.

To simplify the formula, we set

$$A = \beta_{\max} - \overline{\beta} \quad \text{, and} \quad B = \beta_{\min} - \overline{\beta} \quad . \tag{a}$$

Then eq. (14) is described as:

$$A(E_{i+1} - E_i) \le \beta(E_{i+1})E_{i+1} - \beta(E_i)E_i + \alpha(E_{i+1}) - \alpha(E_i) \le B(E_{i+1} - E_i) \tag{b}$$

We take the following summation for all sides of eq. (b):

$$\sum_{k=i}^{j-1}(E_{k+1} - E_k)A \le \sum_{k=i}^{j-1}\{\beta(E_{k+1})E_{k+1} - \beta(E_k)E_k + \alpha(E_{k+1}) - \alpha(E_k)\} \le \sum_{k=i}^{j-1}(E_{k+1} - E_k)B \tag{c}$$

Thus,

$$A(E_j - E_i) \le \beta(E_j)E_j - \beta(E_i)E_i + \alpha(E_j) - \alpha(E_i) \le B(E_j - E_i) \tag{d}$$

$$A \le \frac{\beta(E_j)E_j - \beta(E_i)E_i}{E_j - E_i} + \frac{\alpha(E_j) - \alpha(E_i)}{E_j - E_i} \le B \tag{e}$$

$$\overline{\beta}_{\max} - \overline{\beta} \le \frac{\beta(E_j)E_j - \beta(E_i)E_i}{E_j - E_i} + \frac{\alpha(E_j) - \alpha(E_i)}{E_j - E_i} \le \overline{\beta}_{\min} - \overline{\beta} \qquad \text{(f)}$$

Eq. (f) is eq.(13). Therefore, if eq. (14) is satisfied, the imposed condition (eq. (13)) is satisfied.

Table B1. NPA charges of the solute atoms

	C	**H**	**F**
CH_4	-0.838	0.210	---
CH_3F	-0.093	0.168	-0.410
CHF_3	1.004	0.135	-0.380
CF_4	1.457	---	-0.364

The MP2/aug-cc-pVDZ level of theory was used for the geometry optimization.

The atomic charges of the methane derivatives were calculated by means of natural population analysis (NPA).

Table B2. Relative free energy ΔA (kcal/mol) from CH_4 to CH_3F, CH_3F to CHF_3 and CHF_3 to CF_4 at some temperatures between 273K and 373K.

	Computational ΔA (kcal/mol) at temperature				
	273K	298K	323K	348K	373K
CH_4 to CH_3F	-8.25	-8.21	-8.14	-8.02	-7.86
CH_3F to CHF_3	-1.91	-1.77	-1.67	-1.58	-1.51
CHF_3 to CF_4	3.48	3.67	3.81	3.94	4.05