# Theoretical Studies of Chemical Processes in Multi-Component Solution Systems Based on Integral Equation Theory for Molecular Liquids

Kentaro Kido

2012

# Preface

Electrolyte solutions are ubiquitously utilized in various chemical fields. Electrolytes often dramatically alter chemical phenomena in solution, which means that the cation and anion yielded by dissociation of the electrolyte play a key role in the phenomena. In electrolyte solutions, a great number of both solvent molecules and ions surrounds and interacts with a solute system. An observed data about the system is determined as a result of the diverse interactions. Base on the macroscopic data, it is difficult to directly understand the role of electrolyte at molecular level.

Similarly, it is not easy to theoretically investigate a chemical process in electrolyte solution. There is no established molecular model to treat it and the development is still hot topic in chemistry. The aim of this thesis is to propose a model of multi-component molecular solution system. In this thesis, based on an integral equation theory for molecular liquids (IET), several developments of theoretical tool to study a molecular-level role of electrolyte in chemical phenomena in the solution and the applications were performed.

The thesis consists of two parts and contains seven chapters. Chapters 1 and 7 are the general introduction and conclusions, respectively. In part I, RISM theory, which is a one-dimensional formalism of IET, was fully extended to a multi-component molecular solvent system. The theory was applied to chemical processes in acidic aqueous (chapters 2 and 3) and an electrolyte solution (chapters 4) and provided new molecular-level information about the processes. It is difficult to study each process with Monte Carlo and molecular dynamics simulations which are the most widely used models of a chemical process in solution because a treatment of long-ranged electrostatic interactions in electrolyte solution is not easy, different

from a pure solvent system. In part II, based on MC-MOZ method (a representation of three-dimensional IET), methods to accurately evaluate solvation free energy (chapter 5) and to improve the computational parallelization performance (chapter 6) were proposed.

The studies in the thesis were carried out at Department of Molecular Engineering, Graduate School of Engineering, Kyoto University from 2005 to 2012. The supervisor of the studies is Professor Hirofumi Sato. Firstly, the author would like to express his deepest appreciation to Professor Hirofumi Sato for his valuable suggestions, fruitful discussions and warm encouragement. All studies in this thesis could not be possible without his full cooperation. The author would like to express his sincerely appreciation to Professor emeritus Shigeyoshi Sakaki for his various and energetic comments and advices based on physical chemistry and quantum chemistry.

The author also expresses gratitude to Associate Professor Yoshihide Nakao for his technical advice and valuable comments. The author is also grateful to Dr. Daisuke Yokogawa, Mr. Hideo Ando, Mr. Kenji Iida and Mr. Seigo Hayaki. Scientific talks with them on various occasions were very enjoyable and fruitful. Acknowledgement is also made to all members of the research group of Prof. Hirofumi Sato and Professor emeritus Shigeyoshi Sakaki. Especially, the informal seminars with them were great opportunities to learn many things for the author.

Finally, the author sincerely thanks his mother Keiko Kido, his grandmother Sumako Kido, and his wife Yoko Kido for their understanding, encouragement, and support from all sides.

# Contents

# Chapter 1

# General Introduction

Electrolyte plays a key role in various chemical processes in solution. For example, in electrochemistry, it is an important material composing of liquid media in battery system because the battery performance directly depends on the mobility of ion species in the liquid media. In synthetic chemistry, chemical reactions proceeding only in acid or base solutions are countless, which shows that the electrolyte solution provides an environment to control reactions. Similarly, electrolytes deeply relate on biological fields since a living body is basically filled up with an electrolyte aqueous solution. Biological functions such as proton pump and ion channel are driven by an interaction between specific proteins and electrolyte ions. These examples indicate that electrolyte solutions are not mere liquid media.

It is well-known that by a kind and concentration of electrolyte, chemical processes in solution are affected. For example, a mobility of cation in solution is dependent on both charge and molecular structure of the counter-anion. Hence, a molecularity of ions is essential to understand the behavior and role. Ion channels allow specific kinds of ion to pass through it and the process is quite sensitive to electrolyte concentration. In addition, the electrolyte-concentration dependency shows that a great number of solvent molecules as well as ions associate with the process, which indicates that intermolecular interactions with electrolyte and solvent molecules are also indispensable to investigate the chemical process in solution. However, it is not easy to obtain a microscopic picture of chemical process in solution using an observed data determined by overall intermolecular many-body interactions. One of the ef-

fective approaches is a theoretical procedure complementary with experimental methods based on molecular-level models of electrolyte solution.

In this thesis, an integral equation theory for molecular liquids (IET) is employed as the theoretical tool. IET is a promising framework to study chemical phenomena in general multi-component solution systems. Similarly, molecular simulations are another powerful tool and are applied to diverse and complicate solution systems in practice. But, there are few molecular simulation studies dealing with an effect on a chemical process of both a kind and concentration of electrolyte because of the computational difficulty to treat electrolyte solutions. Before IET framework is presented, the shortcomings in molecular simulation are briefly pointed out below.

## 1.1 Molecular simulations for electrolyte solution systems and the difficulty

Molecular simulations, Monte Carlo (MC)[1] and molecular dynamics (MD)[2] simulations, are the most successful approaches for chemical processes in solution. In MC and MD frameworks, a chemical process is represented by a large number of molecules treated by classical mechanics in a simulation box and is represented by trajectories generated by either Metropolis method (MC)[1] or an equation of motion (MD). Therefore, in principle, molecular simulations can be applied to various chemical processes in electrolyte solution.

In fact, until the late 1970s, there have already been a great number of individual studies at different concentrations for diverse aqueous solutions including simple electrolytes such as alkali or alkaline earth halide[3]. The main subject of the studies is an investigation of static nature of an electrolyte solution with a specific concentration in terms of coordination numbers of ions and liquid structures. In these studies, effects of the concentration have been not treated. Actually, there is few studies investigating the effect with molecular simulations. For instance, NaCl concentration effects in the aqueous solution were studied by an extensive MD simulation[4]. In the paper, the effects on static and dynamical properties of molecule in the solution are systematically investigated as well as the affects of computational conditions (the

length of trajectory, ion model and finite cutoff radii of intermolecular potential). In addition to the study, similar investigations were reported using MD technique[5, 6]. Recently, an implicit solution model involving (coarse-grained) solute and electrolytes in dielectric continuum solvent is often utilized to study an electrolyte concentration effect of chemical processes in solution[7, 8]. However, these types of molecular simulations were considerably limited.

Different from the case of pure solvents, more strong and long-ranged electrostatic interactions exist in electrolyte solution. Coulombic interaction is an essential factor to characterize chemical processes in electrolyte solution but the accurate evaluation is difficult since a simple cutoff method is not applicable due to the Debye length which is longer than usual simulation box lengths. As the result, a total number of molecules in the box becomes larger, compared with pure solvents. Similarly, a huge number of molecules is required to define a low electrolyte concentration. Since the mobility of ions is much lower than that of non-ionic molecules, a significantly long trajectory is required in order to obtain accurate statistics of ionic properties (so-called *sampling* problem). Computational techniques to adequately treat the issues are quite costly and are still non-trivial in current era.

## 1.2 IETs and an extend to molecular multi-component solvent systems

IET is an alternative framework to study chemical processes in solution based on statistical mechanics and provides solvation structure $g(\mathbf{r})$, pair correlation function. Here, $\mathbf{r}$ represents relative distance and orientation between molecules in solution (six dimensions). $g(\mathbf{r})$ is one of the most fundamental functions because any thermodynamic quantity can be described as this function. In addition, $g(\mathbf{r})$ is related to a radial distribution function (RDF) with the measurements such X-ray diffraction[9], inelastic neutron scattering[10], light scattering[11] and electron diffraction[12] experiments. In the theory, a procedure to obtain $g(\mathbf{r})$ is free from the *sampling* procedure, and the analytical representation of the pair correlation functions is available to calculate excess chemical potential (solvation free energy (SFE)) which is a key quantity to elucidate chemical processes in solution, while in molecular simulation technique

the calculation requires a quite high computational costs and time.

In IET framework, a solution is characterized by $g(\mathbf{r})$, which means that the number density of molecule in solution is used to represent the system instead of the number of molecule. This treatment enables us to systematically investigate pH and salt concentration effects. This situation is very contrast to molecular simulations; A number of molecule in simulation box defines solution density and electrolyte concentration. Furthermore, the concentration of co-solvent molecule is limited by available computational costs. This issue is crucial for an investigation of pH effects. For instance, in the aqueous solution at pH=6.0 about $6.0\times10^7$ water molecules are included for a proton species. The costs and calculation time causing by this number are beyond a power of current computer.

In 1914, L. S. Ornstein and F. Zernike introduced the following OZ equation[13]

$$h(\mathbf{r}) = c(\mathbf{r}) + c(\mathbf{r}) * \rho h(\mathbf{r}) \tag{1.1}$$

where $h(\mathbf{r}) = g(\mathbf{r}) - 1$ and $c(\mathbf{r})$ are total and direct correlation functions of molecules, respectively. Asterisk means a convolution integral. Since these two functions are both unknown, a closure equation,

$$h(\mathbf{r}) + 1 = \exp\left\{-\frac{1}{k_{\mathrm{B}}T}u(\mathbf{r}) + h(\mathbf{r}) - c(\mathbf{r}) + b(\mathbf{r})\right\}, \tag{1.2}$$

is coupled with OZ equation. $u(\mathbf{r})$ and $b(\mathbf{r})$ denote an intermolecular interaction potential and bridge function, respectively. $k_{\mathrm{B}}$ and $T$ are Boltzmann constant and temperature, respectively. This OZ/closure formalism has been applied to liquids composed of simple molecule[14]. However, due to the difficulty to numerically solve OZ/closure equation, it is not practical to investigate chemical phenomena in solution with the formalism.

### 1.2.1 Reference interaction site model (RISM)

In 1972, D. Chandler and H. C. Andersen derived RISM equation[15] by using a site-site representation of general OZ equation (eq.(1.1));

$$\mathbf{h} = \mathbf{w} * \mathbf{c} * \mathbf{w} + \mathbf{w} * \mathbf{c} * \boldsymbol{\rho}\mathbf{h} \tag{1.3}$$

where $\mathbf{h}$ and $\mathbf{c}$ are matrix forms of total and direct correlation functions, respectively. $\mathbf{w}$ and $\boldsymbol{\rho}$ denote the intramolecular correlation function and the number density matrix. These matrices are completely defined by molecular geometry and density. Compared with OZ/closure equations, these equations coupled with a closure are easy to solve because the matrix elements, namely site-site pair correlation functions, are the functions of site-site distances (one dimension). Intermolecular orientations are represented with $\mathbf{w}$ matrix. By F. Hirata, *et al.* this equation was extended to more *realistic* liquids which involved an electrostatic intermolecular interaction in 1981[16]. RISM becomes one of the most popular formalisms in IETs.

Taking the number density of solute molecule infinite dilute, one can separate eq.(1.3) into the solvent-solvent (VV, eq.(1.4)) and solute-solvent (UV, eq.(1.5)) parts[17].

$$\mathbf{h}^{\mathrm{VV}} = \mathbf{w}^{\mathrm{VV}} * \mathbf{c}^{\mathrm{VV}} * \left(\mathbf{w}^{\mathrm{VV}} + \boldsymbol{\rho}^{\mathrm{V}}\mathbf{h}^{\mathrm{VV}}\right) \tag{1.4}$$

$$\mathbf{h}^{\mathrm{UV}} = \mathbf{w}^{\mathrm{UU}} * \mathbf{c}^{\mathrm{UV}} * \left(\mathbf{w}^{\mathrm{VV}} + \boldsymbol{\rho}^{\mathrm{V}}\mathbf{h}^{\mathrm{VV}}\right) \tag{1.5}$$

where $\boldsymbol{\rho}^{\mathrm{V}}$ represents the number density of solvent molecule. Let us focus on the eq.(1.4) to consider multi-component solvent systems. In the case of pure water, the intramolecular correlation function ($\mathbf{w}^{\mathrm{VV}}_{\mathrm{water}}$) and the number density matrix ($\boldsymbol{\rho}^{\mathrm{V}}_{\mathrm{water}}$) are written by

$$\mathbf{w}^{\mathrm{VV}}_{\mathrm{water}} = \begin{pmatrix} w_{\mathrm{OO}} & w_{\mathrm{OH_1}} & w_{\mathrm{OH_2}} \\ w_{\mathrm{H_1O}} & w_{\mathrm{H_1H_1}} & w_{\mathrm{H_1H_2}} \\ w_{\mathrm{H_2O}} & w_{\mathrm{H_2H_1}} & w_{\mathrm{H_2H_2}} \end{pmatrix}, \quad \boldsymbol{\rho}^{\mathrm{V}}_{\mathrm{water}} = \begin{pmatrix} \rho_{\mathrm{O}} & 0 & 0 \\ 0 & \rho_{\mathrm{H_1}} & 0 \\ 0 & 0 & \rho_{\mathrm{H_2}} \end{pmatrix}. \tag{1.6}$$

If there are components A and B such as acids, bases or salts in addition to water molecule in the liquid, the corresponding intramolecular correlation function and the number density matrix are expressed by the following block diagonal matrices[18]

$$\mathbf{w}^{\mathrm{VV}}_{\mathrm{multi}} = \begin{pmatrix} \mathbf{w}^{\mathrm{VV}}_{\mathrm{water}} & & \mathbf{0} \\ & \mathbf{w}^{\mathrm{VV}}_{\mathrm{A}} & \\ \mathbf{0} & & \mathbf{w}^{\mathrm{VV}}_{\mathrm{B}} \end{pmatrix}, \quad \boldsymbol{\rho}^{\mathrm{V}}_{\mathrm{multi}} = \begin{pmatrix} \boldsymbol{\rho}^{\mathrm{V}}_{\mathrm{water}} & & \mathbf{0} \\ & \boldsymbol{\rho}^{\mathrm{V}}_{\mathrm{A}} & \\ \mathbf{0} & & \boldsymbol{\rho}^{\mathrm{V}}_{\mathrm{B}} \end{pmatrix}. \tag{1.7}$$

Of course, these representations are not limited in aqueous solutions and are available for any solvent molecule. Similarly, if components C, D, $\cdots$ exist in the same liquid, the intramolecular correlation function and the number density matrix are furthermore extended to put the corresponding matrices of components C, D $\cdots$ into the block diagonal elements. Here, it is emphasized that in this framework the parameter essentially needed to apply to electrolyte

solutions is only a concentration of co-solvent. This fact is crucial for pH effects because approximately $6.0\times10^7$ water molecules for a proton species are required to realize aqueous solution at pH=6 if the number of molecules in solvent is directly treated.

### 1.2.2 Calculations of thermodynamic properties based on IET framework

In IET framework, $h(\mathbf{r})$ and $c(\mathbf{r})$ allow us to analytically compute any thermodynamic quantity. For example, excess internal energy $U^{\mathrm{ex}}$ of a fluid is provided by the following equation[11],

$$\frac{U^{\mathrm{ex}}}{N_{\mathrm{A}}} = \sum_{s}^{\text{solvent}} \sum_{s'}^{\text{solvent}} \rho_s \int d\mathbf{r}\, \frac{1}{2} u_{ss'}^{\mathrm{VV}} (h_{ss'}^{\mathrm{VV}} + 1) \tag{1.8}$$

where $N_{\mathrm{A}}$ is Avogadro constant and $u_{ss'}^{\mathrm{VV}}$ denotes solvent-solvent intermolecular interaction potential between site $s$ and $s'$.

Properties of solute molecule can be also written as a function of these pair correlation functions. By coupling RISM theory with Kirkwood-Buff theory[19], a partial molar volume of solute molecule, $V$ is expressed as follows[20].

$$V = k_{\mathrm{B}} T \chi_T \left\{ 1 - \sum_{\alpha}^{\text{solute}} \sum_{s}^{\text{solvent}} \rho_s \tilde{c}_{\alpha s}^{\mathrm{UV}}(0) \right\}. \tag{1.9}$$

where $\chi_T$ and $\tilde{c}_{\alpha s}^{\mathrm{UV}}$ denote the isothermal compressibility of solvent and the Fourier transform of $c_{\alpha s}^{\mathrm{UV}}$, respectively. The expression of SFE by Gaussian fluctuation approximation, $\Delta\mu^{\mathrm{GF}}$ is given as follows[22].

$$\Delta\mu^{\mathrm{GF}} = k_{\mathrm{B}} T \sum_{\alpha}^{\text{solute}} \sum_{s}^{\text{solvent}} \rho_s \int d\mathbf{r} \left\{ -c_{\alpha s}^{\mathrm{UV}} - \frac{1}{2} h_{\alpha s}^{\mathrm{UV}} c_{\alpha s}^{\mathrm{UV}} \right\}. \tag{1.10}$$

Similarly, if RISM equation is coupled with hyper-netted chain (HNC) closure ($b(\mathbf{r}) = 0$ in eq.(1.2)), a SFE, $\Delta\mu^{\mathrm{HNC}}$ of solute molecule is represented by[21]

$$\Delta\mu^{\mathrm{HNC}} = k_{\mathrm{B}} T \sum_{\alpha}^{\text{solute}} \sum_{s}^{\text{solvent}} \rho_s \int d\mathbf{r} \left\{ \frac{1}{2} \left( h_{\alpha s}^{\mathrm{UV}} \right)^2 - c_{\alpha s}^{\mathrm{UV}} - \frac{1}{2} h_{\alpha s}^{\mathrm{UV}} c_{\alpha s}^{\mathrm{UV}} \right\}. \tag{1.11}$$

However, absolute values of SFE evaluated by this formula almost always overestimate the corresponding experimental ones due to the poor description of cavitation energy. The systematical overestimations are caused by the approximation of HNC closure[23, 24].

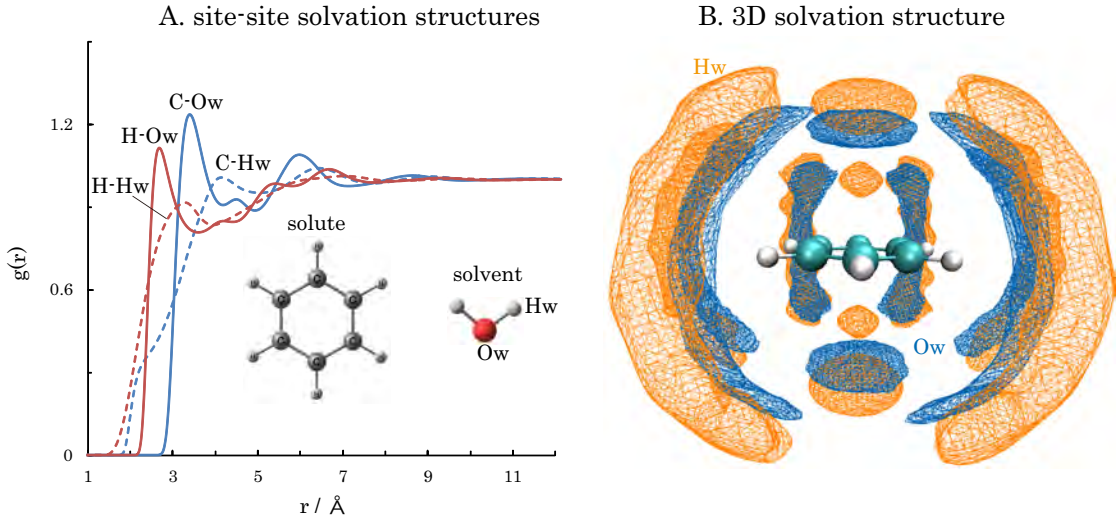


Figure 1.1: An example of site-site (A) and 3D (B) solvation structures; the hydration structures near benzene molecule computed by RISM/HNC (A) and MC-MOZ/HNC (B).

In order to obtain more accurate SFE, A. Kovalenko and F. Hirata[23] proposed a repulsive bridge correction (RBC) by introducing $b_{\alpha s}^{\mathrm{RBC}}$, a solute-solvent repulsive bridge function. This study is one of the efforts beyond HNC approximation. RISM equation coupled with the HNC closure involving the function give a SFE ($\Delta\mu^{\mathrm{RBC}}$);

$$\Delta\mu^{\mathrm{RBC}} = \Delta\mu^{\mathrm{HNC}} + k_{\mathrm{B}}T \sum_{\alpha}^{\text{solute}} \sum_{s}^{\text{solvent}} \rho_s \int d\mathbf{r} \left(h_{\alpha s}^{\mathrm{UV}} + 1\right) \left(\exp b_{\alpha s}^{\mathrm{RBC}} - 1\right), \quad (1.12)$$

$$\exp b_{\alpha s}^{\mathrm{RBC}} = \prod_{s' \neq s}^{\text{solvent}} w_{s's}^{\mathrm{VV}} * \exp\left(-\frac{u_{\alpha s'}^{\mathrm{R}}}{k_{\mathrm{B}}T}\right). \quad (1.13)$$

The second term in the right hand side of eq.(1.12) is negative for any case because $h_{\alpha s}^{\mathrm{UV}} + 1 \geq 0$ and $\exp b_{\alpha s}^{\mathrm{RBC}} - 1 \leq 0$. So $\Delta\mu^{\mathrm{RBC}}$ is always lower than $\Delta\mu^{\mathrm{HNC}}$. This formula is successfully applied to various solute molecules in aqueous solution. The formalism of eq.(1.13) is exact for diatomic-type solvent molecules but is not adequate for more realistic ones. In practice, a SFE in several organic solvent such as benzene is severely underestimated by this equation.

## 1.3 Three-dimensional (3D) formalisms in IETs

For a complicated-shaped solute molecule like protein, it is difficult to analyze the solvation structure based on the one-dimensional pair correlation function such as RDF. In this case, a

higher dimensional pair correlation function is convenient to discuss the distribution of solvent molecules around the solute. Several 3D representations of OZ equation (eq.(1.1)) have been already elaborated by employing 3D-lattice grids[25]. OZ-type equation[26], which is called 3D-RISM theory, is one of the representations and becomes the most popular 3D expression of IETs. The OZ-type equation is as follows.

$$h_s(\mathbf{r}_s) = \sum_{s'}^{\text{solvent}} c_{s'} * \left(w_{s's}^{\text{VV}} + \rho_s h_{s's}^{\text{VV}}\right)(\mathbf{r}_s) \tag{1.14}$$

where $h_s(\mathbf{r}_s)$ and $c_s(\mathbf{r}_s)$ represent total and direct correlation functions of solvent site $s$. The vector $\mathbf{r}_s$ means a 3D coordinate of site $s$. By coupling with an adequate closure, $h_s(\mathbf{r}_s)$ and $c_s(\mathbf{r}_s)$ are determined. This formalism corresponds to the 3D version of solute-solvent RISM equation (eq.(1.5)). The reliability has been shown in the various applications from simple atomic solutes to biochemical systems such as protein in solution[27]. Of course, in this framework the analytical expressions are available to compute thermodynamic properties. Although employing 3D-lattice as the spatial coordinate is quite straightforward, the approach needs 3D-fast Fourier transformation (3D-FFT) which consumes computational time and resources.

Very recently, D. Yokogawa *et al.* proposed an alternative approach to obtain 3D solvation structures, multi-center molecular OZ (MC-MOZ)[28]. In MC-MOZ approach, both $h_s(\mathbf{r}_s)$ and $c_s(\mathbf{r}_s)$ are separated to reference (ref) and residue (res) parts, namely, $f_s(\mathbf{r}_s) = f_s^{\text{ref}}(\mathbf{r}_s) + f_s^{\text{res}}(\mathbf{r}_s)$ $(f = h,\ c)$. The reference terms of $h_s(\mathbf{r}_s)$ and $c_s(\mathbf{r}_s)$ are given by the solution of eq.(1.5).

$$h_s^{\text{ref}}(\mathbf{r}_s) = \sum_{\alpha}^{\text{solute}} \sum_{s'}^{\text{solvent}} c_{\alpha s'}^{\text{UV}} * \left(w_{s's}^{\text{VV}} + \rho_s h_{s's}^{\text{VV}}\right) \tag{1.15}$$

$$c_s^{\text{ref}}(\mathbf{r}_s) = \sum_{\alpha}^{\text{solute}} c_{\alpha s}^{\text{UV}} \tag{1.16}$$

The residue terms are represented by the summation of those assigned on each solute site, and

the assigned residue terms are expanded with real solid harmonics $Y_{lm}(\theta_s, \phi_s)$;

$$\begin{aligned} f_s^{\mathrm{res}}(\mathbf{r}_s) &= \sum_{\alpha}^{\mathrm{solute}} f_{\alpha s}^{\mathrm{res}}(\mathbf{r}_s) \\ &= \sum_{\alpha}^{\mathrm{solute}} \sum_{lm} f_{\alpha s}^{\mathrm{res},lm}(r_s) Y_{lm}(\theta_s, \phi_s) \quad (f = h,\ c). \end{aligned} \tag{1.17}$$

Substituting eqs.(1.15), (1.16) and (1.17) into eq.(1.14), one obtains

$$h_{\alpha s}^{\mathrm{res},lm}(r_s) = \sum_{s'}^{\mathrm{solvent}} c_{\alpha s'}^{\mathrm{res},lm} * \left( w_{s's}^{\mathrm{VV}} + \rho_s h_{s's}^{\mathrm{VV}} \right)(r_s). \tag{1.18}$$

Since this equation can be independently solved for each solute site, a highly computational parallelization can be generated on the calculations. Furthermore, thanks to above treatment of pair correlation functions, MC-MOZ does not require 3D-FFT procedure to compute the convolution integral. However, the routine to construct the reference term, which solves the solute-solvent RISM equation, is not site-independently performed, the parallelization efficiency becomes lower as the number of processors increases[28].

## 1.4 Survey of the present thesis

In this thesis, based on IET framework, chemical processes in multi-component solution system are investigated in terms of both solvation structure and free energy. The thesis includes two parts (seven chapters). Chapters 1 (here) and 7 are the general introduction and the general conclusion, respectively. Part I is consists of chapters 2, 3 and 4. The part describes RISM theory extended to a multi-component solution systems and the applications. Part II includes chapters 5 and 6. In this part, the author develops new methods to estimate SFE and to achieve the fully computational parallelization of whole MC-MOZ procedure.

Chapter 2 proposes a molecular level theory to predict p$K_{\mathrm{a}}$ based on RISM theory extended to multi-component molecular solvent systems. At the present, the most common approach to calculate p$K_{\mathrm{a}}$ is a free energy evaluation of proton dissociation in dielectric continuum model of pure water, $\epsilon = 80$, which completely ignores ionic influence, *i.e.* ionic strength. In the chapter, by regarding an acidic aqueous solution as three component system including water, proton species (cation) and anion, aqueous solutions with desired pH can be theoretically

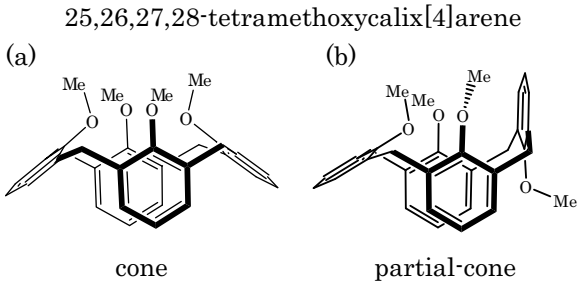


Figure 1.2: Chemical structures of CXA (cone (a) and partial-cone (b) conformers).

realized. The pH is controlled by the number density of the proton species. To our best knowledge, this is the first theoretical attempt to directly evaluate p$K_{\mathrm{a}}$ under the condition where ionic influence is explicitly taken into account by using statistical molecular theory.

In chapter 3, toward accurate computations of p$K_{\mathrm{a}}$ and p$K_{\mathrm{b}}$ from the first principles, a systematic assessment based on accurate evaluation of free energy is presented by using RISM self-consistent field spatial electronic density distribution (RISM-SCF-SEDD) method[30, 31, 32] which is a hybrid framework of RISM and a standard *ab initio* MO calculation. Because the process is essentially bond formation and dissociation in aqueous solution, both of highly-sophisticated quantum chemistry and an accurate evaluation of SFE are necessary to appropriately compute p$K_{\mathrm{a}}$ and p$K_{\mathrm{b}}$. In this chapter, the proton dissociation of carboxyl and amino groups in glycine molecule in aqueous solution is examined. As the *ab initio* MO calculation, Hartree-Fock method, density functional theory and coupled cluster method are employed.

In chapter 4, a cation capture by 25,26,27,28-tetramethoxycalix[4]arene (CXA) in sodium and potassium perchlorate chloroform solutions is studied with RISM-SCF-SEDD employing multi-component version RISM and MP2 which is a sophisticated *ab intio* method. $^{1}$H NMR experiment shows that CXA captures a sodium cation by the cone conformer depicted in Figure 1.2(a) and a potassium cation by the partial-cone (Figure 1.2(b))[34, 35]. In the chapter, by computing and analyzing the free energy difference to capture the cation, an origin of the conformational change is investigated. Similarly, with obtained RDFs between CXA and the cation, the position of captured cation near CXA is discussed.

Chapter 5 describes a new method to estimate SFE. As pointed out above, the formalism

presented in eqs.(1.12) and (1.13) is not valid for more complicate-shaped solvent molecules than dumbbell-like one such as organic solvents. This is because an exclusive volume of the complicate-shaped solvent molecules is not accurately estimated in the original core repulsive bridge function. In this chapter, by modifying the function, the author develops a new method applicable to such solvents with a high accuracy. As the benchmark, the present method is applied to various solute molecules in aqueous, chloroform and benzene solutions.

In chapter 6, a revised version of MC-MOZ method is developed. As mentioned in the previous section, while in original MC-MOZ the residue term are fully parallelized for each solute site, the reference term (RISM) can not be site-independently computed, which is the cause of low computational efficiency on the parallelization[28]. Several types of function were examined as the reference part to investigate the numerical robustness of the method in this chapter. As the benchmark, the method is applied to water, benzene in aqueous solution and single-walled carbon nanotube in chloroform solution.

# Bibliography


[1] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).

[2] B. Alder, T. E. Wainwright, *J. Chem. Phys.*, *31*, 459 (1959).

[3] for example, K. Heinzinger, *Physica*, **131B**, 196 (1985).

[4] (a) A. P. Lyubartsev, A. Laaksonen, *J. Phys. Chem.*, **100**, 16410 (1996). (b) A. P. Lyubartsev, A. Laaksonen, *Phys. Rev. E*, **55**, 5689 (1997).

[5] R. Chitra, P. E. Smith, *J. Phys. Chem. B*, **104**, 5854 (2000).

[6] S. Chowdhuri, A. Chandra, *J. Chem. Phys.*, **115**, 3732 (2001).

[7] (a) S. Gavryushov, P. Linse, *J. Phys. Chem. B*, **110**, 10878 (2006). (b) S. Gavryushov, *J. Phys. Chem. B*, **110**, 10888 (2006).

[8] A. Savelyev, G. A. Papoian, *J. Phys. Chem. B*, **113**, 7785 (2009).

[9] W. J. Huisman, J. F. Peters, J. W. Derks, H. G. Ficke, D. L. Abernathy, J. F. van der Veen, *Rev. Sci. Instrum.*, **68**, 4169 (1997).

[10] T. Bausenwein, H. Bertagnolli, A. David, K. Goller, H. Zweier, K. Todheide, P. Chieux, *J. Chem. Phys.*, **101**, 672 (1994).

[11] J. P. Hansen, I. R. McDonald, *Academic Press*, San Diego (1986).

[12] P. G. Ackermann, J. E. Mayer, *J. Chem. Phys.*, **4**, 377 (1936).

[13] L. S. Ornstein, F. Zernike, *Proc. R. Acad. Sci. Amsterdam*, **17**, 793 (1914).

[14] (a) L. Blum, A. J. Torruella, *J. Chem. Phys.* **56**, 303 (1972). (b) L. Blum, *J. Chem. Phys.*, **57**, 1862 (1972). (c) L. Blum, *J. Chem. Phys.*, **58**, 3295 (1973). (d) P. H. Fries, G. N. Patey, *J. Chem. Phys.*, **82**, 429 (1985). (e) P. G. Kusalik, G. N. Patey, *Mol. Phys.*, **65**, 1105 (1988). (f) J. Richardi, C. Millot, P. H. Fries, *J. Chem. Phys.*, **110**, 1138 (1999).

[15] (a) D. Chandler, H. C. Andersen, *J. Chem. Phys.*, **57**, 1930 (1972). (b) D. Chandler, H. C. Andersen, *J. Chem. Phys.*, **57**, 2626 (1972).

[16] (a) F. Hirata, P. J. Rossky, *Chem. Phys. Lett.*, **83**, 329 (1981). (b) F. Hirata, B. M. Pettitt, P. J. Rossky, *J. Chem. Phys.*, **77**, 509 (1982).

[17] F. Hirata, P. J. Rossky, B. M. Pettitt, *J. Chem. Phys.*, **78**, 4133 (1983).

[18] M. Kinoshita, F. Hirata, *J. Chem. Phys.*, **106**, 5202 (1997).

[19] J. G. Kirkwood and F. P. Buff, *J. Chem. Phys.* **19**, 774 (1951).

[20] (a) T. Imai, M. Kinoshita, F. Hirata, *J. Chem. Phys.*, **112**, 9469 (2000). (b) Y. Harano, T. Imai, A. Kovalenko, M. Kinoshita, F. Hirata, *J. Chem. Phys.*, **114**, 9506 (2001).

[21] S. J. Singer, D. Chandler, *Mol. Phys.*, **55**, 621 (1985).

[22] D. Chandler, Y. Singh, D. M. Richardson, *J. Chem. Phys.*, **81**, 1975 (1984).

[23] A. Kovalenko, F. Hirata, *J. Chem. Phys.*, **113** 2793 (2000).

[24] (a) D. Chandler, *Mol. Phys.* **31** 1213 (1976). (b) P. T. Cummings, C. G. Gray, D. E. Sullivan, *J. Phys. A*, *14*, 1483 (1981).

[25] (a) D. Beglov, B. Roux, *J. Chem. Phys.*, **103**, 360 (1995). (b) M. Ikegami, J. Doi, *J. Chem. Phys.*, **103**, 5011 (1995).

[26] (a) D. Beglov, B. Roux, *J. Phys. Chem. B*, **101**, 7821 (1997). (b) A. Kovalenko, F. Hirata, *Chem. Phys. Lett.* **290** 237 (1998).

[27] (a) A. Kovalenko, F. Hirata, *J. Chem. Phys.* **112** 10391 (2000). (b) A. Kovalenko, F. Hirata, *J. Chem. Phys.* **112** 10403 (2000). (c) N. Yoshida, P. Saree, Y. Maruyama, T. Imai, F. Hirata, *J. Am. Chem. Soc.* **128** 12042 (2006). (d) Y. Kiyota, R. Hiraoka, N. Yoshida, Y. Maruyama, T. Imai, F. Hirata, *J. Am. Chem. Soc.* **131** 3852 (2009).

[28] (a) D. Yokogawa, H. Sato, T. Imai, S. Sakaki, *J. Phys. Chem.*, **130**, 064111 (2009). (b) D. Yokogawa, H. Sato, S. Sakaki, *J. Mol. Liq.*, **147**, 112 (2009). (c) K. Hirano, D. Yokogawa, H. Sato, S. Sakaki, *J. Phys. Chem. B*, **114**, 7935 (2010).

[29] S. Ten-no, F. Hirata, S. Kato, *J. Chem. Phys.*, **100**, 7443 (1994).

[30] (a) S. Maw, H. Sato, S. Ten-no, F. Hirata, *Chem. Phys. Lett.*, **276**, 20 (1997). (b) H. Sato, F. Hirata, *J. Chem. Phys.*, **111**, 8545 (1999). (c) H. Sato, F. Hirata, *J. Phys. Chem. A*, **102**, 2603 (1998). (d) H. Sato, F. Hirata, *J. Phys. Chem. B*, **103**, 6596 (1999). (e) H. Sato, F. Hirata, S. Sakaki, *J. Phys. Chem. A*, 108, 2097 (2004). (f) N. Yoshida, R. Ishizuka, H. Sato and F. Hirata, *J. Phys. Chem. B*, **110**, 8451 (2006).

[31] H. Sato, F. Hirata, S. Kato, *J. Chem. Phys.*, **105**, 1546 (1996).

[32] D. Yokogawa, H. Sato, S. Sakaki, *J. Chem. Phys.*, **126**, 244504 (2007).

[33] A. Ikeda, S. Shinkai, *Chem. Rev.*, **97**, 1713 (1997).

[34] K. Iwamoto, A. Ikeda, K. Araki, T. Harada, S. Shinkai, *Tetrahedron*, **49**, 9937 (1993).

[35] T. Harada, M. J. Rudziński, S. Shinkai, *J. Chem. Soc., Perkin. Trans. 2*, **12**, 2109 (1992).

# Part I

# An extend of RISM theory to a multi-component system containing molecular co-solvents and the applications

# Chapter 2

# A first principle theory for p$K_{\rm a}$ prediction at molecular level: pH effects based on explicit solvent model

## 2.1 Introduction

pH is one of typical macroscopic parameters characterizing aqueous solutions. It affects the electronic structure of solvated molecules and even controls the chemical processes. p$K_a$, which controls structure and function of proteins, is directly related to this quantity. A majour goal in the present study is to evaluate p$K_{\rm a}$ of the dissociation equilibrium of an amino acid, glycine molecule, in hydrochloric acid (HCl(aq)) solution, absolutely from the first principle at molecular level. In general, the common theoretical strategy to obtain p$K_{\rm a}$ is to compute the free energy difference between the protonated and deprotonated forms in liquid water ($\Delta G$) that is mostly represented by dielectric continuum with $\varepsilon$=80,[1] followed by assuming the simple relationship.

$$\mathrm{p}K_{\rm a} = \frac{\Delta G}{\ln 10 \cdot RT}. \tag{2.1}$$

It is noted that $\Delta G$ is usually evaluated as a constant value in *pure* water. A few other strategies to compute p$K_{\rm a}$ have also be proposed.[2] However, aqueous solution in reality should be considered in a different way; by changing the concentration of the acid, pH of the aqueous solution is defined. An amino acid and bulk solvent vie with each other to get possession of the protonated species and this competition is balanced at a specific ionic-concentration condition, namely p$K_{\rm a}$. This means that, the concentrations of ionic species in the solution affect chem-

ical equilibrium; $\Delta G$ can vary from the concentration condition to condition, pH, *i.e.*, $\Delta G$ is regarded as a function of pH. By definition, p$K_{\mathrm{a}}$ corresponds to the 'balanced' condition and is determined as the inflection point of the titration curve, which makes a correlation between pH and $\Delta G$, namely the fraction in the dissociation.

Although pH is regarded as one of fundamental properties of aqueous solution, theoretical or computational realization of pH effect at molecular level had been unfeasible. This is because that a huge system size is required to represent pH condition if proton species are explicitly treated in standard molecular simulations such as molecular dynamics. For example, an aqueous solution at pH = 6.0 involves $6.0 \times 10^7$ water molecules per a proton, which may be beyond the capability of simulation size in the present computational environment. Evaluation of thermodynamic quantities like free energy is time-consuming, and, furthermore, all of aqueous systems with different proton concentrations must be treated at the same accuracy to elucidate pH effect in a chemical process.

The key to overcoming this difficulty is to use statistical mechanics for molecular liquids. In the present study we employed the reference interaction site model (RISM), which was developed by Chandler and Andersen[3] and extended by Rossky and Hirata[4]. In RISM theory, the number of density (number of molecules in a unit volume, Å$^3$) is employed to express the concentration of solution, leading to avoid direct counting of the particle numbers in the target system. It should also be emphasized that RISM is free from the so-called sampling problem because of its strong background in statistical mechanical theory. The method has been combined with ab initio molecular orbital theory (RISM-SCF),[5, 6] and many chemical equilibria including a series of hydrogen halides in aqueous solution[7], liquid water[8] and p$K_{\mathrm{w}}$[9, 10] at various thermodynamic conditions and tautomerization[11]. Recently, Kinoshita and Hirata applied the theory to electrolyte solution systems.[12] In the present study, we developed a molecular level theory to handle pH effect in aqueous solution for the first time. Another representative achievement of the present work is to realize titration-curve computed from the first principle at molecular level of theory.

In this work, we carried out a simple and honest approach to compute p$K_{\mathrm{a}}$ as follows; free

energy change $\Delta G$ for acidic pH range ($0 \leq$ pH $\leq 7$) was computed with a molecular level treatment based on the statistical mechanical RISM theory, and p$K_{\rm a}$ was then computed by plotting the titration curve from $\Delta G$'s. As expalined below, the obtained result shows a good agreement with experimental knowledge[13]. Thanks to the molecular level description of the theory, the solvation structure and its change as function of pH can be obtained in terms of pair correlation functions (PCFs), equivalently to radial distribution functions (RDFs), $g(r)$.

Because of its inherent difficulty, systematic investigation of pH effect based on molecular simulation method has never been carried out. To our best knowledge, the present study is the first report to elucidate the pH effect fully based on a molecular level description

## 2.2 Theory and Computational Method

RISM equations [3] are given in the matrix forms as

$$\mathbf{h}^{\rm VV} = \mathbf{w}^{\rm V} * \mathbf{c}^{\rm VV} * \mathbf{w}^{\rm V} + \mathbf{w}^{\rm V} * \mathbf{c}^{\rm VV} * \boldsymbol{\rho} * \mathbf{h}^{\rm VV}, \tag{2.2}$$

$$\mathbf{h}^{\rm UV} = \mathbf{w}^{\rm U} * \mathbf{c}^{\rm UV} * \mathbf{w}^{\rm V} + \mathbf{w}^{\rm U} * \mathbf{c}^{\rm UV} * \boldsymbol{\rho} * \mathbf{h}^{\rm VV}, \tag{2.3}$$

where asterisks represent convolution integrals and subscripts V and U denote solvent and solute, respectively. $\mathbf{w}^{\rm U/V}$ is the intramolecular correlation function specifying the geometry of chemical species. $\mathbf{h}^{\rm UV/VV}$ and $\mathbf{c}^{\rm UV/VV}$ are matrixes of intermolecular total and direct correlation functions, respectively. The former matrix element is equivalent to pair correlation function representing liquid structure but its definition is $h_{ab}(r) = g_{ab}(r) - 1$. By coupled with a closure equation, RISM equation is usually solved as simultaneous equation. Please refer to the reviews or textbooks[14] for the details if the readers were not familiar with the statistical theory for molecular liquids. Since we need to deal with polyatomic cations and anions, $\mathbf{w}^{\rm V}$ is expressed by a block diagonal matrix. $\boldsymbol{\rho}$ is the diagonal matrix comprising the number densities of solvent components.

$$\mathbf{w}^{\rm V} = \begin{pmatrix} \mathbf{w}^{\rm water} & & \mathbf{0} \\ & \mathbf{w}^{\rm cation} & \\ \mathbf{0} & & w^{\rm anion} \end{pmatrix}, \quad \boldsymbol{\rho} = \begin{pmatrix} \boldsymbol{\rho}^{\rm water} & & \mathbf{0} \\ & \boldsymbol{\rho}^{\rm cation} & \\ \mathbf{0} & & \rho^{\rm anion} \end{pmatrix}. \tag{2.4}$$

In the present study, HCl aqueous solution system is assumed and $H_3O^+$ is employed as the model of proton (cation) and $\mathbf{w}^{\rm cation}$ and $\boldsymbol{\rho}^{\rm cation}$ are represented in 4×4 matrix since the cation consists of four atoms. This is a straightforward extension of the treatment by Kinoshita et al [12], where monoatomic cations and anions were considered. Water and anion are expressed as three and mono atomic molecules and $w^{\rm anion}$ is set to 1 in the above equation. Obviously, its extension to basic solution is straightforward; all you need to do is change the density of $OH^-$ instead of the cation. The matrix elements of diagonal $\boldsymbol{\rho}$ is determined from the concentration.

$$\begin{aligned}[\text{cation}] &= 10^{-\text{pH}}\ \text{mol/L} = 10^{-\text{pH}}\frac{N_{\rm A}}{10^{27}}\ \text{molecule/Å}^3 = 6.02\times 10^{-(4+\text{pH})}\ \text{molecule/Å}^3 \\ &= (\boldsymbol{\rho}^{\rm cation})_{ii} = \rho^{\rm cation}, \quad (\text{where } i = \text{O, H, H and H}) \end{aligned} \tag{2.5}$$

where $N_{\rm A}$ is Avogadro constant and pH is the 'input' of the computation. Since HCl molecules are fully dissociative, $\rho^{\rm anion} = \rho^{\rm Cl^-}$ equals to $\rho^{\rm cation}$. The number density of water molecule ($\rho^{\rm water}$) is taken from the literature [15]. Under this condition of aqueous solution, let us consider equilibrium of glycine between the cationic form ($Gly^+$) and the zwitterion form ($Gly^\pm$).

$$\text{Gly}^+ \rightleftharpoons \text{Gly}^\pm + \text{H}^+. \tag{2.6}$$

Note that the equilibrium between the zwitterion and the deprotonated anion forms ($Gly^\pm \rightleftharpoons Gly^- + H^+$) is not necessary to be considered because $Gly^-$ hardly exists in the present acidic condition. The free energy change of the equilibrium is described as follows;

$$-\frac{\Delta G}{RT} = \ln\frac{[\text{Gly}]^\pm[\text{H}]^+}{[\text{Gly}]^+} = \ln\frac{[\text{Gly}]^\pm}{[\text{Gly}]^+} - (\ln 10)\text{pH}. \tag{2.7}$$

The fraction, $\theta$, is defined as the probability to find the protonated form of glycine molecule,

$$\theta = \frac{[\text{Gly}]^+}{[\text{Gly}]^\pm + [\text{Gly}]^+}. \tag{2.8}$$

Combining these two equations, the following expression is obtained.

$$\theta = \left(1 + 10^{\text{pH} - \dfrac{\Delta G}{\ln 10\cdot RT}}\right)^{-1}. \tag{2.9}$$

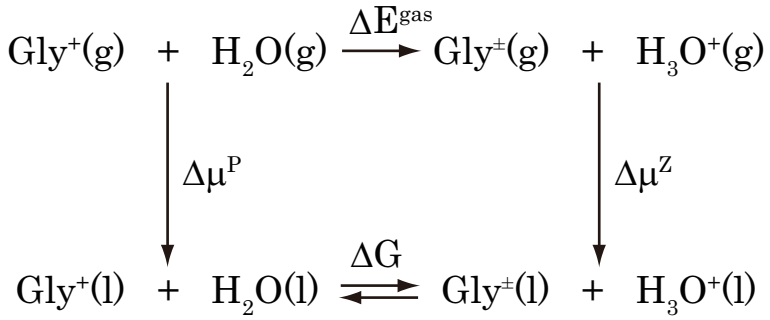


Figure 2.1: The thermodynamic cycle used in this study in order to estimate ΔG. $\Delta\mu^{\mathrm{P}}$ and $\Delta\mu^{\mathrm{Z}}$ is the solvation free energies of cationic and zwitterionic forms of glycine molecule, respectively.

An explicit treatment of proton and its solvation free energy, which appears in Eq.(2.6), are difficult because of its quantum mechanical character. Hence, solvation free energy of four species ($\mathrm{Gly}^{+}$, $\mathrm{H_2O}$, $\mathrm{Gly}^{\pm}$ and $\mathrm{H_3O}^{+}$) were individually computed in different pH condition by changing ionic concentration ($\mathrm{Cl}^{-}$ and $\mathrm{H_3O}^{+}$), and $\Delta G$ is then evaluated using the thermodynamic cycle shown in Figure 2.1.

$$\Delta G = \Delta G^{\mathrm{gas}} + (\Delta\mu^{\mathrm{Z}} - \Delta\mu^{\mathrm{P}}) = \Delta G^{\mathrm{gas}} + \Delta\Delta\mu, \tag{2.10}$$

where $\Delta\mu^{\mathrm{P}}$ and $\Delta\mu^{\mathrm{Z}}$ are the solvation free energies of protonated and zwitterionic forms in Figure 2.1, respectively. In RISM framework, these quantities can be computed from $\mathbf{h}^{\mathrm{UV}}$ and $\mathbf{c}^{\mathrm{UV}}$ matrix components [16].

$$\Delta\mu^{\mathrm{X}} = k_B T \sum_{\alpha}^{\mathrm{U}} \sum_{\gamma}^{\mathrm{V}} \rho_\gamma \int 4\pi r^2 dr \{ \frac{1}{2} {h_{\alpha\gamma}^{\mathrm{UV}}}^2 \Theta(-h_{\alpha\gamma}^{\mathrm{UV}}) - c_{\alpha\gamma}^{\mathrm{UV}} - \frac{1}{2} c_{\alpha\gamma}^{\mathrm{UV}} h_{\alpha\gamma}^{\mathrm{UV}} \}, \text{ (X=P or Z)} \tag{2.11}$$

where $\Theta$ is heaviside step function. $\Delta G^{\mathrm{gas}}$ is the free energy change between $\mathrm{Gly}^{\pm}$ and $\mathrm{Gly}^{+}$ in gas-phase and equals to 70.84kcal/mol (B3LYP/6-311G(2d,p) level)[17], which is regarded as a constant value throughout the present study. Note that $\Delta G$ defined in Eq. (2.10) were computed at different pH conditions, in which different number of ionic species such as $\mathrm{H_3O}^{+}$, $\mathrm{Cl}^{-}$ are contained. Hence it is not necessary to be a constant in the present treatment. Actually $\Delta G$ depends on pH because of the contributions from the cosolvents, cation and anions as shown below. Now, the change of $\theta$ as function of pH is nothing except for the titration curve

Table 2.1: Potential parameters for HCl(aq) components

| molecule | site | charge /\|e\| | $\sigma$ / Å | $\epsilon$ / (kcal/mol) |
|---|---|---|---|---|
| [a]$H_2O$ | O | -0.820 | 3.166 | 0.155 |
| | H | 0.410 | 1.000 | 0.056 |
| [b]$H_3O^+$ | O | -0.656 | 3.166 | 0.155 |
| | H | 0.552 | 1.000 | 0.056 |
| [c]$Cl^-$ | | -1.000 | 3.620 | 0.448 |

[a] modified SPC model [25]. [b] ref.[9]. [c] ref.[7].

Table 2.2: potential parameters for protonated and zwitterion forms of glycine molecules.[a]

| | $Gly^+$ | $Gly^\pm$ | | |
|---|---|---|---|---|
| site | charge/\|e\| | | $\sigma$ / Å | $\epsilon$ / kcal mol$^{-1}$ |
| $H_N2$ | 0.315 | 0.311 | 1.000 | 0.056 |
| $H_N1$ | 0.288 | 0.304 | 1.000 | 0.056 |
| N | -0.080 | -0.175 | 3.250 | 0.170 |
| C | -0.152 | -0.133 | 3.500 | 0.066 |
| H(-C) | 0.142 | 0.114 | 2.500 | 0.030 |
| C(=O) | 0.751 | 0.843 | 3.750 | 0.105 |
| O(=C) | -0.608 | -0.851 | 2.960 | 0.210 |
| O | -0.592 | -0.831 | 2.960 | 0.210 |
| H(-O) | 0.506 | | 1.000 | 0.056 |

[a] The parameters $\sigma$ and $\epsilon$ for heavy atoms are standard OPLS taken from Ref. [18].

and p$K_a$ can be determined as the inflection point ($\partial^2\theta/\partial(\mathrm{pH})^2 = 0$) of the drawn curve. The inflection point was numerically determined on the fitting curve in the present study.

For the computations, the standard pair potential function consisted of Coulomb and Lennard-Jones term as intermolecular interaction function were employed to solve RISM equation coupled with Kovalenko-Hirata (KH) closure [19]. Temperature was set to 298.15K. All the parameters of solute and solvent molecules are collected in Tables 2.1 and 2.2. The geometries of glycine molecules were optimized in pure water by using RISM-SCF-SEDD method[6] ((B3LYP/6-311G(2d, p), KH closure)),[20] in which the solvation effect plays an essential role to obtain the stable structure of the zwitterion form. When a standard geometry optimization in the gas phase is employed, any stable structure does not exist and the proton transfers from the amino to the carboxyl group without barrier, spontaneously leading to the neutral. In aqueous solution, the energy profile is dramatically changed and the zwitterion form exists as a stable species. The contribution from other conformers is negligible because their free energy

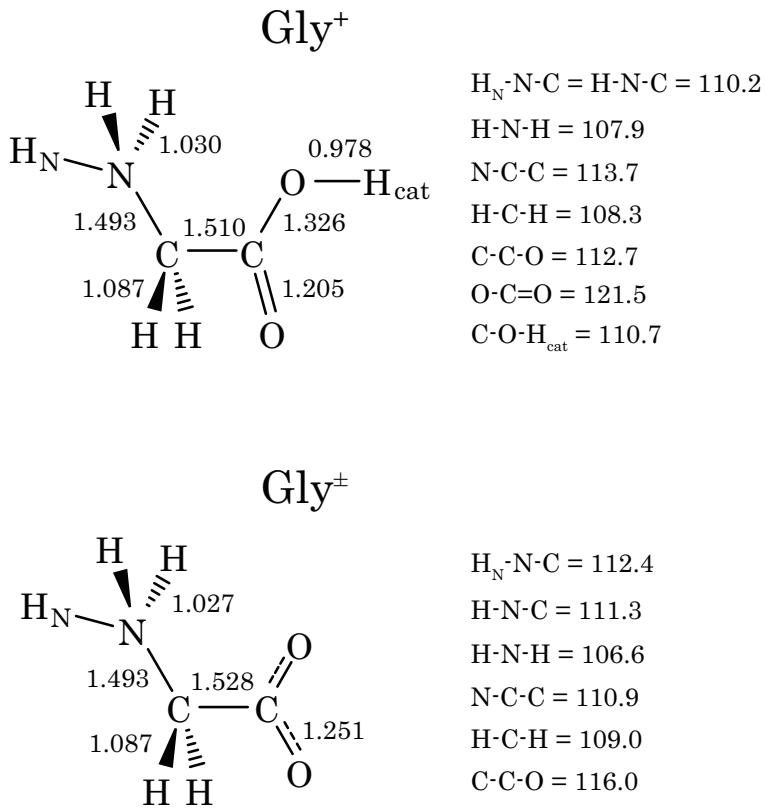


Figure 2.2: The geometric paremeters and site labels of each form in optimized glycine molecule. Bond distances and angles are given by angstrom and degree unit, respectively.

were higher than 5.0 kcal/mol.

## 2.3 Results and discussion

### 2.3.1 Free energy change and p$K_{\mathrm{a}}$

Figure 2.3 is the free energy change as a function of pH. For the sake of better understanding, the deviation from the $\Delta G$ at pH=7.0 ($\Delta\Delta G$) is plotted.[21] As has been emphasized repeatedly, $\Delta G$ is not necessary to be a constant over a wide range of pH condition. In fact $\Delta G$ slightly decreases ($\sim$ 1 kcal/mol) as pH decreases at extreamly low region (pH $<$ 1.0), although the line is fairly flat, namely, $\Delta G$ is virtually regarded as a constant for a wide range of pH.

In Eq. (2.11), the summation for solvent index ($\gamma$) run over water, cation and anion. The

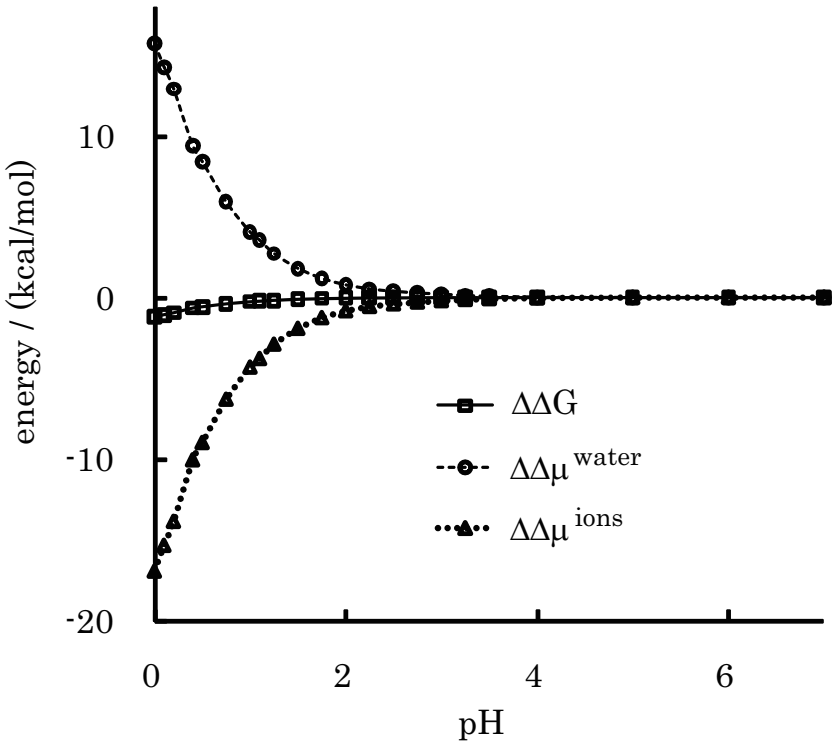


Figure 2.3: ΔG and its components ($\Delta\Delta\mu^{\rm water}$ and $\Delta\Delta\mu^{\rm ions}$) plotted as function of pH, where the ΔG at pH=7.0 is taken to be standard.

contributions from bulk water and ion (cation and anion) are formally evaluated by taking respective indexes in the summation, *i.e.*, O, H, H for water and the others,

$$\Delta\mu^{\rm X} = \Delta\mu^{\rm water} + \Delta\mu^{\rm ions}, \quad \text{and hence,} \quad \Delta\Delta\mu^{\rm X} = \Delta\Delta\mu^{\rm water} + \Delta\Delta\mu^{\rm ions}, \quad \text{(X=P or Z)} \tag{2.12}$$

To closely look at the decrease of $\Delta G$ found at pH $< 1.0$, the change of these two components ($\Delta\Delta\mu^{\rm water}$ and $\Delta\Delta\mu^{\rm ions}$) were also plotted. Each solvation free energies at pH=7.0 were chosen as the respective standard. Both of $\Delta\Delta\mu^{\rm water}$ (dash line) and $\Delta\Delta\mu^{\rm ions}$ (dot line) are flat over a wide range of pH but considerably change for low pH region. They are nearly canceled out each other especially in the region of $1 <$ pH $< 2$, but the contribution from ionic species ($\Delta\Delta\mu^{\rm ions}$) is slightly dominative at the low pH range (pH $< 1$). The decrease of $\Delta G$ could be regarded as a concequence of breaking down of detailed balances between the two contributions. It is worthwhile that each contribution is more than ten kcal/mol especially at low pH condition though the sum of them becomes very small.

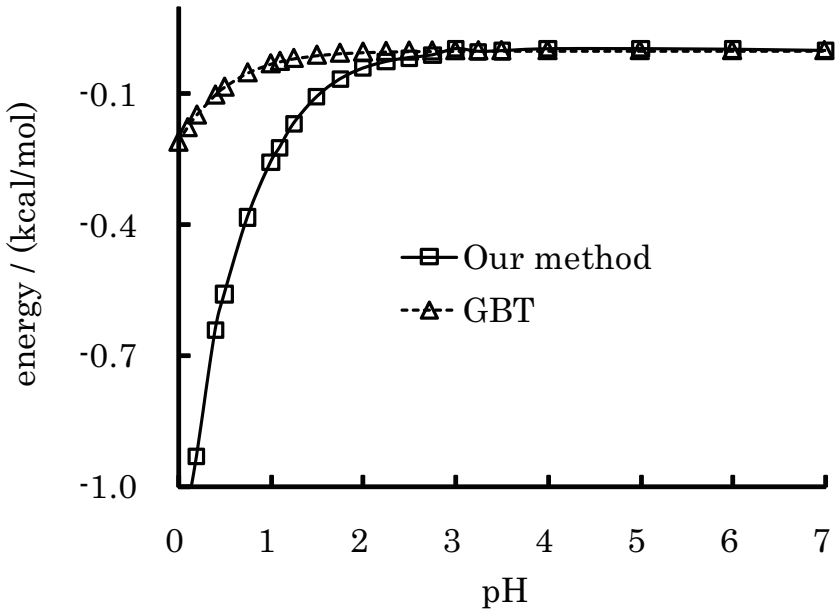


Figure 2.4: The comparison of $\Delta\Delta$Gs plotted as a function of pH by the present method (solid line) and GBT (dash line).

$\Delta\Delta G$ may be also evaluated based on a simplified procedure such as generalized Born thoery (GBT). In the theory, solvation free energy consists of cavitation energy, van der Waals interaction energy and polarization energy[22]. The last term is expressed as follows,

$$\Delta\mu^{\mathrm{GBT}} = -\frac{1}{2}\sum_{i,j}\left(1-\frac{e^{-\kappa f_{ij}}}{\epsilon}\right)\frac{q_i q_j}{f_{ij}}, \tag{2.13}$$

where $\epsilon$ is dielectric constant of solvent water fixed at 80.0, $\kappa$ is Debye-Hückel screening parameter, and $q_i$ is effective charge assigned to each atom.

$$\kappa = \sqrt{\frac{2e^2 I}{\varepsilon k_B T}}, \quad f_{ij} = \sqrt{r_{ij}^2 + \sigma_i\sigma_j e^{\frac{r_{ij}^2}{4\sigma_i\sigma_j}}}, \tag{2.14}$$

$r_{ij}$ denotes the distance between site $i$ and $j$, Born radii $\sigma_i$ and $q_i$ were set to the same values in Table 2.1 and Table 2.2. The ionic strength of the system ($I$) is given by the well-known formula using the molarity of ion ($b_i$) or the number of density ($\rho_i$),

$$I = \frac{1}{2}\sum_i b_i Z_i^2 = \frac{1}{2}\sum_i\left(\rho_i\frac{10^{27}}{N_{\mathrm{A}}}\right)Z_i^2. \tag{2.15}$$

In the present analysis, the contributions from the cavitation and from the van der Waals interaction are constant as pH variation, and the free energy change is attributed to that of $\kappa$

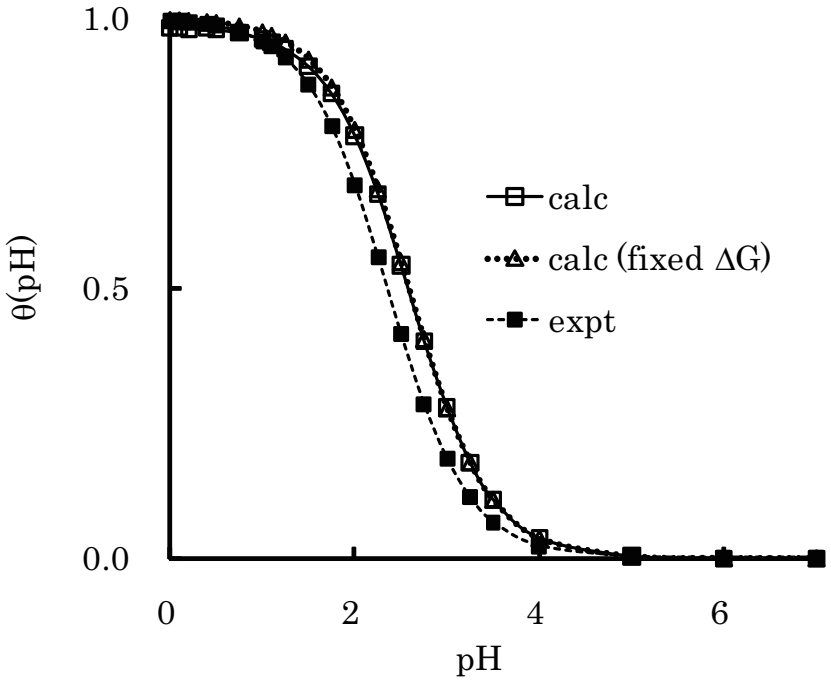


Figure 2.5: The titration curves calculated by using the present method (solid and dots lines) and experimental value (dash line). The dots line indicates the titration curve computated from a fixed ΔG at pH = 7.0.

($I$) in Eq. (2.13). Figure 2.4 shows the $\Delta\Delta G$ obtained by the present method and $\Delta\Delta\mu^{\rm GBT}$. Their behaviors agree well, especially at moderate pH condition, indicating that the decrease at low pH is related to the significance of $I$, although the evaluation by GBT is weaker than the present all-atomic-type treatment.

Finally, the titration curve is obtained (Figure 2.5) by insertion of $\Delta G$ into Eq. (2.9). The pH value at the inflection point of the curve corresponding to p$K_{\rm a}$ value is 2.58, being good agreement with the experimental value (2.35)[13]. It is noted that assumption of regarding $\Delta G$ as a constant over all region of pH is very reliable. If one uses a value that is calculated at pH=7.0 (3.53 kcal/mol) in Eq. (2.9), the drawn titration curve (dotted line) is almost identical and the same p$K_{\rm a}$ is obtained. The difference is seen only for low pH range where the ionic contributions are significant. But the low pH, high $\theta$ region is obviously far from the inflection point. That means, in general, $\Delta G$ calculated in pure water (pH=7.0) is expected to be valid for p$K_{\rm a}$ prediction because ionic strength does not significantly affect the proton release at

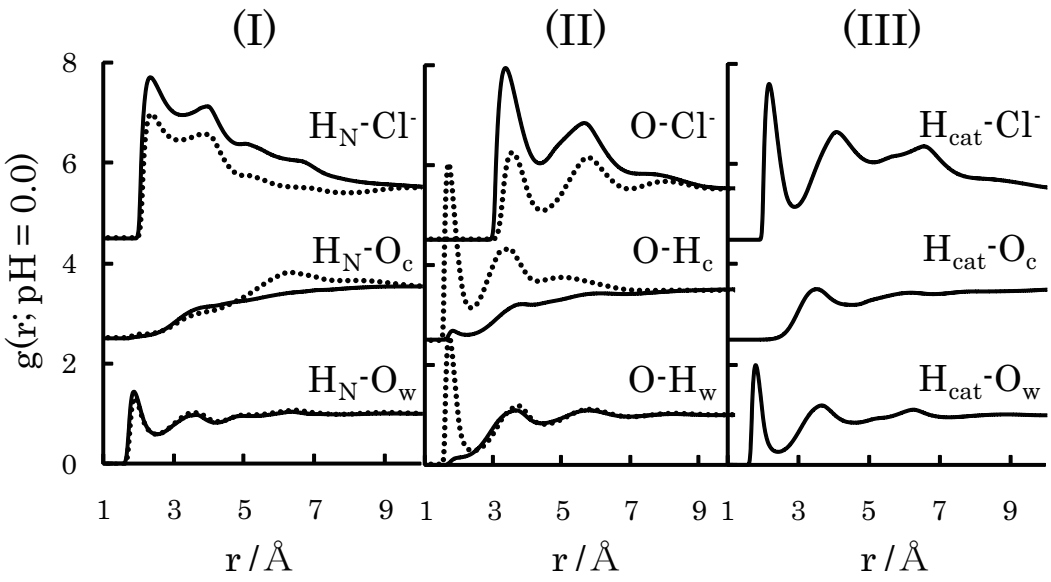


Figure 2.6: The solvation strucrure near cationic (solid line) and zwitterionic (dots line) forms at pH = 0.0. Panel I, II and III show RDFs around $H_N$, O and $H_{cat}$ sites, respectively. $O_w$ and $H_w$ denote the oxygen and hydrogen sites in water molecule, respectively. $O_c$ and $H_c$ denote the oxygen and hydrogen sites in $H_3O^+$ ion, respectively.

normal p$K_a$ region. For a molecule whose p$K_a$ is in low pH range (pH $< 2$), free energy can not be regarded as a constant any more and its change caused by the ionic strength must be taken into account.

### 2.3.2 Solvation structures around amino and carboxyl groups

One of the great advantages of the present method is to shedding light on the molecular level mechanism of solvation phenomenon. In statistical mechanics, solvation structure is characterized by radial distribution function (RDF) as shown in Figure 2.6. Here we pay attention at the solvation structure change from the protonated form ($Gly^+$) to zwitterionic form ($Gly^{\pm}$), and the following three representative atoms (sites) are focused; (I) hydrogen atom attaching to the amine group ($H_N$), (II) oxygen atom in carboxyl group (O) and (III) hydrogen atom attaching to carboxyl group ($H_{cat}$).

In the first set concerning $H_N$ (I), remarkable changes upon the protonation are found in the interaction with solvent $Cl^-$ and with oxygen in $H_3O^+$ ($O_c$). The former RDFs ($H_N$–$Cl^-$) have distinct peaks at 2.5Å corresponding to the direct contact of ion, while $H_N$–$O_c$ RDFs are flat

without any characteristic peaks. Upon the changing from $Gly^+$ (solid line) to $Gly^\pm$ (dotted line), profiles at 2.5Å $\sim$ 9.0Å in $H_N$–$Cl^-$ pair and those at 3.8Å $\sim$ 9.0Å in $H_N$–$Cl^-$ pair are remarkably changed, suggesting that both of cation and anion located at the second solvation shell are affected by the protonation. By contrast, RDFs of $H_N$–$O_w$ (solvent water oxygen) pair remain unchanged. The peaks at 2.0 Å is a typical indication of a hydrogen bond, showing that the hydration structure hardly changes by the protonation. On the deprotonation, net charge assigned to the amino group decreases by $0.068|e|$. This change weakens the attractive interaction with $Cl^-$ as well as repulsive one with $H_3O^+$. Presumably, the attractive interaction is strong enough even in the $Gly^\pm$ and the first solvation shell is fully occupied with hydrogen bonding water molecule and $Cl^-$. The difference between $Gly^+$ and $Gly^\pm$ appears only in the second solvation shell.

Let us move to the carboxyl group. (Ⅱ) and (Ⅲ) show the RDFs around O and $H_{cat}$ sites, respectively. In the latter case, only in $Gly^+$ form is plotted because $H_{cat}$ itself is removed by the deprotonation to form $Gly^\pm$. RDF in (III) exhibits typical hydrogen bonding characters, and the negatively charged $Cl^-$ and $O_w$ show conspicuous peaks around 2.0Å. On the other hand, the positively charged $H_3O^+$ can not approach to $H_{cat}$ and the corresponding peak is missing. It is important and interesting that the second peaks of $O_w$ and $O_c$ ($\sim$3.5Å) are in the almost the same positions, suggesting that both of them construct the hydrogen-bonding network in a similar manner. In other words, $H_3O^+$ is naturally embedded in the liquid water structure though it may not become hydrogen-bonding accepter.

The most drastic changes are seen in RDFs of O–$H_c$ and O–$H_w$ as depicted in (II). Very sharp peaks exist near 2.0Å in $Gly^\pm$ form. Both of solvent water and $H_3O^+$ serve hydrogen bonding to O through their hydrogen atoms. These peaks disappear on the deprotonation and RDFs of $Gly^+$ form are very flat, especially in O–$H_c$ pair because of the direct repulsion between O–$Cl^-$ RDFs shows a peak around 3.0Å. The first peak in $Gly^+$ (solid line) corresponds to the aforementioned hydrogen bonding to $H_{cat}$ attached to O. By the deprotonation to form $Gly^\pm$ (dotted line), repulsive interaction betwen O and $Cl^-$ is enhanced and RDF decreases in the wide range (3.5Å $\sim$8.0Å). This discussion is largely consistent with several neutron diffraction

experiments of amino acid in aqueous solution.[23]

### 2.3.3 pH dependence of solvation structure

Finally, pH effect on the solvation structure is addressed. Probably, one of the most standard ways to analyze the change of solvation structures is to look at the difference between RDFs.

$$\Delta g(r) = g^{\mathrm{X}}(r) - g^{\mathrm{std}}(r), \tag{2.16}$$

or the difference between spatial densities.

$$\Delta \rho(r) = \rho^{\mathrm{X}} g^{\mathrm{X}}(r) - \rho^{\mathrm{std}} g^{\mathrm{std}}(r), \tag{2.17}$$

where $g^{\mathrm{X}}(r)$ denotes RDF computed at pH=X condition, $\rho$ is the number of density focused and 'std' stands for the standard pH condition. Unfortunately, these simple choices are not adequate enough in the present pH-dependency study, because there are two major contributions to affect the solvation structure. One is attributed to a trivial contribution based solely on the density difference, the other is the non-trivial term related to the change in solvation structure such as hydrogen-bonding network alteration that deserves attention. Remember that the number of density of ionic species is changed in more than $10^7$ order, and the definition of Eq. (2.17), for instance, is meaningless when different order $\rho$'s are compared because one of the terms readily becomes negligible. In fact, it was hard to analyze $\Delta g(r)$ or $\Delta \rho(r)$ (not shown). Here we introduce the following function to separate the contribution from the trivial contribution by taking the logarithm of distribution function.

$$\Delta^{\mathrm{X}}(r) \equiv \log\left\{\rho^{\mathrm{X}} g^{\mathrm{X}}(r)\right\} - \log\left\{\rho^{\mathrm{std}} g^{\mathrm{std}}(r)\right\} = (\mathrm{X}^0 - \mathrm{X}) + \log \frac{g^{\mathrm{X}}(r)}{g^{\mathrm{std}}(r)}. \tag{2.18}$$

where $\mathrm{X}^0 = -\log \rho^{\mathrm{std}}$ is the standard pH (std) and set to 6.0 throughout the study and X$= -\log \rho$. In the right-hand-side of the equation, the first term represents the contribution from the trivial density difference and the second is a non-trivial contribution arising from solvation structure change. If two RDFs obtained at pH=X and at $\mathrm{X}^0$ are exactly the same, the second term becomes zero. The first term is non-essential and virtually regarded as a constant.[24]

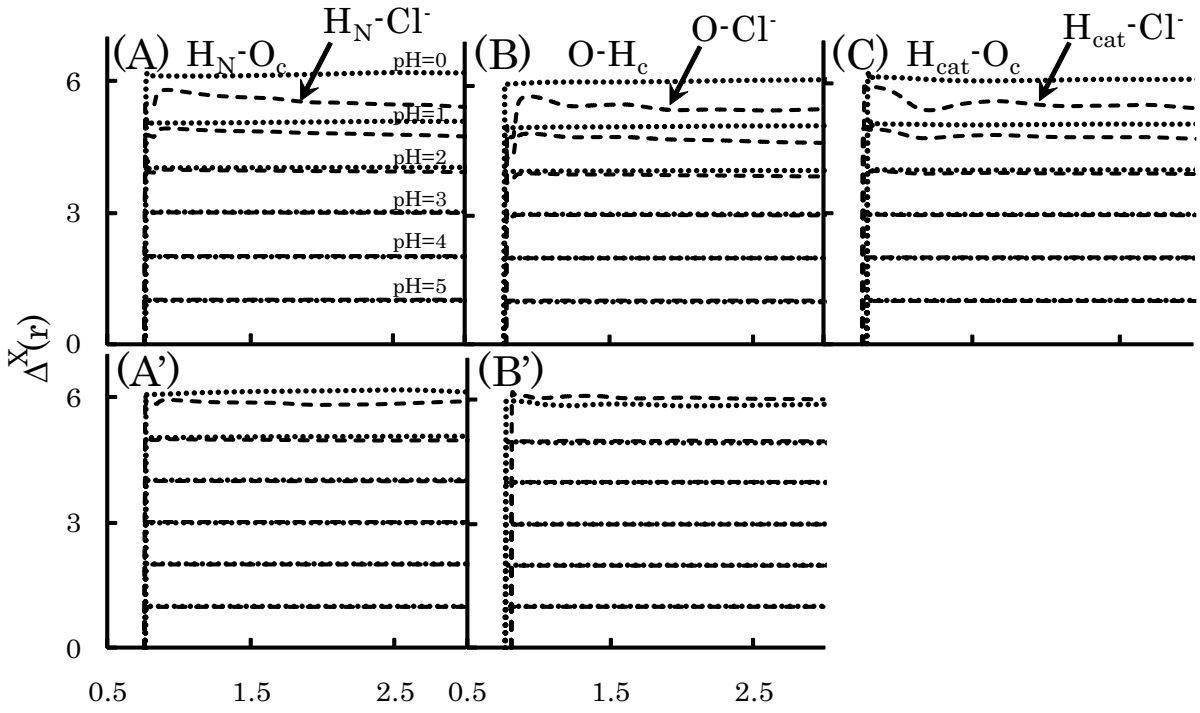


Figure 2.7: $\Delta^X(r)$'s near cationic (upper three panels) and zwitterionic (lower two panels) forms. Panel A(A'), B(B') and C show the $\Delta^X(r)$'s around $H_N$, O and $H_{cat}$ sites, respectively. The horizontal axises in all panel are scaled length, in which site-site distance is divided by the corresponding LJ parameter ($\sigma_{\alpha\gamma}$).

The functions $\Delta^{\mathrm{X}}(r)$ are plotted in Figure 2.7. The upper three panels respectively correspond to (A) $H_N$, (B) O and (C) $H_{cat}$ in $Gly^+$. The lower two to (A') $H_N$ and (B') O in $Gly^{\pm}$. In all the cases at higher pH conditions, the functions ($\Delta^2(r) \sim \Delta^5(r)$) are very simple, step-function like one. The means the RDFs are very similar even at different pH conditions, clearly indicating the ion strength is not essential in these conditions. As increasing the ion concentration to lower pH, $\Delta^{\mathrm{X}}(r)$ begin to exhibit variety of changes. The probabilities to find $Cl^-$ in the vicinity of $H_N$, O and $H_{cat}$ of $Gly^+$ (upper panel) are significantly reduced compared to high pH condition. It should not be confused that number of $Cl^-$ is still greater than that in higher pH condition. In this low pH limit, even if the ionic concentration is increased tenfold, the repulsive interaction between $Cl^-$ is too strong not to reach 10-times concentration, at least in the vicinity of $Gly^+$. This suggests that individual hydration of $Cl^-$ in bulk aqueous environment is preferable. The situation looks very different in $Gly^{\pm}$ case and the concentration

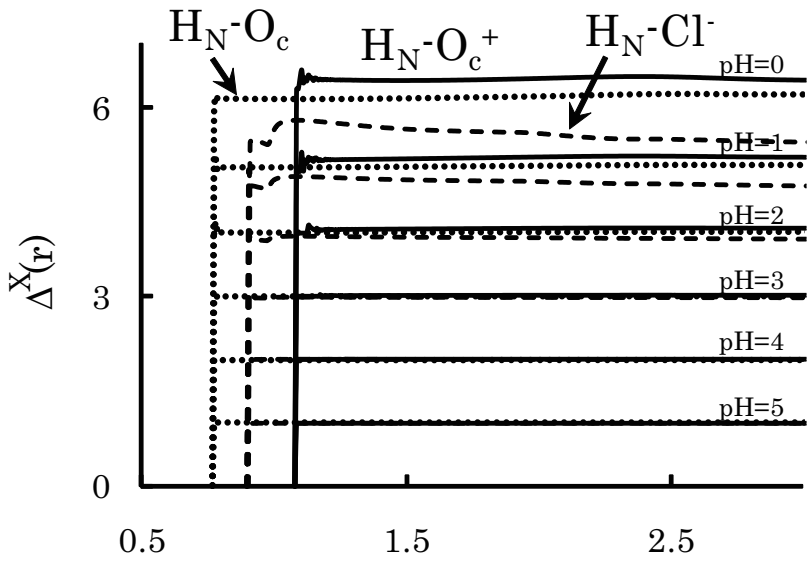


Figure 2.8: $\Delta^X(r)$ near $H_N$ site in cationic form; solid: $O_c^+$, dots: $O_c$, dash: $Cl^-$. The horizontal axises in all panel are scaled length, in which site-site distance is divided by the corresponding LJ parameter ($\sigma_{H_N\gamma}$).

of $Cl^-$ is monotonically increased, but remember that the 'std' RDF is different from the $Gly^+$ case.

On the contrary, the functions of $\Delta^X(r)$ for $O_c$ and $H_c$ always show step-function like profile, meaning that solvation structure is hardly changed over wide range of pH conditions. To clarify this difference, we performed a computation with hypothetical molecule ($O_c^+$), in which all the hydrogen atoms are removed from $H_3O^+$. Since the molecular size of $H_3O^+$ and $O_c^+$ are virtually the same, the difference is only the charge distribution in 'molecule', attributed to the presence of hydrogen. Computed $\Delta^X(r)$ for $O_c^+$ show pH dependence in low pH condition (Figure 2.8), which clearly indicates hydrogen is the key to remain the solvation structure of $H_3O^+$. In another words, hydrogen bonding is preserved even in high ionic concentration condition and $H_3O^+$ may be embedded in hydrogen bonding network though it only acts as hydrogen donor.

## 2.4 Conclusions

In the present article, we have proposed a procedure to theoretically realize aqueous solution with desired pH condition, and to predict p$K_a$ absolutely from the first principle at molecular level based on a statistical mechanics for molecular liquids, RISM. From the computed free energy changes on the protonation of glycine at various pH conditions, a titration curve of this equilibirium and p$K_a$ can be obtained, which shows excellent agreement with experimental data. We found that ionic strength is not so serious at the moderate pH condition, but it becomes crucial at pH $< 2$, suggesting that approximately p$K_a$ prediction using free energy evaluation in *pure water* is safely used in general case. The hydrogen-bonding solvation structure is also significantly affected only at low pH region.

# Bibliography


[1] For example, Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.*, **1991**, *95*, 5610.

[2] (a) Onufriev, A.; Case, D. A.; Ullmann, G. M. *Biochem.*, **2001**, *40*, 3413. (b) Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comp. Chem.* **2004**, *25*, 2038.

[3] Chandler, D.; Andersen, H. C. *J. Chem. Phys.*, **1972**, *57*, 1930.

[4] (a) Hirata, F.; Rossky, P. J. *Chem. Phys. Lett.*, **1981**, *83*, 329. (b) Hirata, F.; Rossky, P. J.; Pettitt, B. M. *J. Chem. Phys.*, **1983**, *78*, 4133.

[5] (a) Ten-no, S.; Hirata, F.; Kato, S. *J. Chem. Phys.*, **1994**, *100*, 7443. (b) Sato, H.; Hirata, F.; Kato, S. *J. Chem. Phys.*, **1996**, *105*, 1546.

[6] Yokogawa, D.; Sato,H.; Sakaki, S. *J. Chem. Phys.*, **2007**, *126*, 244504.

[7] Sato, H.; Hirata, F. *J. Am. Chem. Soc.*, **1999**, *121*, 3460.

[8] (a) Maw, S.; Sato, H.; Ten-no, S.; Hirata, F. *Chem. Phys. Lett.*, **1997**, *276*, 20. (b) Sato, H.; Hirata, F. *J. Chem. Phys.*, **1999**, *111*, 8545.

[9] Sato, H.; Hirata, F. *J. Phys. Chem. A*, **1998**, *102*, 2603.

[10] (a) Sato, H.; Hirata, F. *J. Phys. Chem. B*, **1999**, *103*, 6596. (b) Yoshida, N.; Ishizuka, R.; Sato, H.; Hirata, F. *J. Phys. Chem. B*, **2006**, *110*, 8451.

[11] Sato, H.; Hirata, F.; Sakaki, S. *J. Phys. Chem. A*, **2004**, *108*, 2097.

[12] Kinoshita, M.; Hirata, F. *J. Chem. Phys.*, **1997**, *106*, 5202.

[13] Wada,G.; Tamura, E.; Okina, M.; Nakamura, M. *Bull. Chem. Soc. Jpn.*, **1982**, *55*, 3064.

[14] (a) Hansen, J. P.; McDonald, I. R. *Theory of Simple Liquids*, Academic Press, (1990). (b) Hirata, F. Ed., *Molecular Theory of Solvation (Understanding Chemical Reactivity)*, Kluwer-Springer, (2004). (c) Mennucci, B.; Cammi, R. Eds., *Continuum Solvation Models in Chemical Physics: From Theory to Applications*, John Wiley & Sons Inc (2008).

[15] Lide, D. R. ed., *CRC Handbook of Chemistry and Physics*, **87th Edition**, 8-67 (2003).

[16] Singer, S. J.; Chandler, D. *Mol. Phys.*, **1985**, *55*, 621.

[17] Total energies of $Gly^{\pm}$ and $Gly^{+}$ are respectively, -284.328575 and -284.724782 au, and corresponding free-energy corrections are -284.276147 and -284.658692 au. Those for the proton are -76.693589 and -76.677919. It is also noted that addition of diffuse functions does not alter the change. $\Delta G^{gas}$ by 6-311++G(2d,p) basis set gives 70.70 kcal/mol and 6-311G(2d,p) is 70.84 kcal/mol.

[18] Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.*, **1988**, *110*, 1657.

[19] Kovalenko, A.; Hirata, F. *J. Chem. Phys.*, **1999**, *110*, 10095.

[20] In RISM-SCF-SEDD computations, total energies for the solvated molecule are -286.767761 ($Gly^{\pm}$) and -278.912913 ($Gly^{+}$) were obtained.

[21] $\Delta\mu^{\mathrm{P}} = \Delta\mu(\mathrm{H_2O}) + \Delta\mu(\mathrm{Gly})^{+}$ and $\Delta\mu^{\mathrm{Z}} = \Delta\mu(\mathrm{H_3O^{+}}) + \Delta\mu(\mathrm{Gly})^{\pm}$. Respective energies are -2.66 ($\Delta\mu(\mathrm{H_2O})$), -48.08 ($\Delta\mu(\mathrm{Gly})^{+}$), -70.09 ($\Delta\mu(\mathrm{H_3O^{+}})$) and -47.97 kcal/mol ($\Delta\mu(\mathrm{Gly})^{\pm}$).

[22] (a) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.*, **1990**, *112*, 6127. (b) Srinivasan, J.; Trevathan,M. W.; Beroza, P.; Case, D. A. *Theo. Chem. Acc.*, **1999**, *101*, 426.

[23] (a) Kameda Y.; Ebata H.; Usuki T.; Uemura O.; Misawa M. *Bull. Chem. Soc. Jpn.*, **1994**, 67, 3159. (b) Sugawara K.; Kameda Y.; Usuki T.; Uemura O.; Fukunaga T. *Bull. Chem.*

*Soc. Jpn.*, **2000**, 73, 1967. (c) Sasaki M.; Kameda Y.; Yaegashi M.; Usuki T. *Bull. Chem. Soc. Jpn.*, **2003**, 76, 2293. (d) McLain, S. E.; Soper, A. K.; Watts, A. *J. Phys. Chem. B*, **2006**, 110, 21251.

[24] The function looks like step function because the second term becomes zero near the origin. Since the rise of RDF is always closer to the origin as higher pH condition, the function is definable without singularity under the condition of X<std.

[25] Sato, H.; Sakaki, S. *J. Phys. Chem. A*, **2002**, 106, 2300.

# Chapter 3

# Systematic assessment on aqueous $pK_a$ and $pK_b$ of an amino acid base on RISM-SCF-SEDD method: Toward first principles calculations

## 3.1 Introduction

It is well recognized that accurate computation of p$K_{\mathrm{a}}$ and p$K_{\mathrm{b}}$ is formidable though they are fundamental quantities to characterize amino acid and proteins. This is mainly attributed to the difficulty in evaluation of free energy change on the proton attachment or detachment in aqueous solution. Because they are essentially regarded as bond formation and dissociation in solution phase, a high-level electronic structure theory coupled with solvent effect is required to properly describe the phenomenon. Dielectric continuum model [1] may be insufficient because the hydrogen bonding is responsible for the events. Hence QM/MM or its alternative method utilizing highly sophisticated electronic structure theory is appropriate to attack the problem. Yagasaki et al. reported QM/MM treatment on p$K_{\mathrm{w}}$ [2], but the evaluation of the absolute value is still one of the most challenging problems.

On performing molecular simulation such as molecular dynamics, intermolecular electrostatic interaction is a key to express a specific interaction such as hydrogen bonding. While a great number of simulations have been performed so far [3, 4, 5], it is non-trivial whether point-charge (PC) representation, which is commonly utilized, is sufficient or not to adequately represent p$K_{\mathrm{a}}$ and p$K_{\mathrm{b}}$. Recently, Takahashi *et al.* reported that PC model could not properly evaluate the free energy for isomerization of glycine in aqueous solution [6, 7]. They pointed out that PC model seriously overestimates the solvation free energy of zwitterionic

form, leading to incorrect reaction free energy. The effects of spreading electrons (also called SEDD; spatial electron density distribution) is essential for the adequate description [7]. There is a similar difficulty in the computation of p$K_{\mathrm{a}}$ and p$K_{\mathrm{b}}$ but systematic assessment has never been employed because of the following reasons; (1) Applicable method incorporating solvent effect with SEDD treatment is very limited. (2) Evaluation of accurate free energy change in solution phase is required. Furthermore, a uniformly accurate evaluation over a variety of computational levels such as basis sets is necessary to properly perform a systematic assessment. The latter leads a difficulty in standard molecular simulation methods because the elimination of uneven statistical sampling errors at different computational levels is not readily achieved in general.

In the present study, RISM-SCF is carried out to evaluate p$K_{\mathrm{a}}$ and p$K_{\mathrm{b}}$. RISM-SCF is an alternative to QM/MM method and has been successfully used for a variety of chemical processes [9]. As differentiated from the original RISM-SCF theory [10], RISM-SCF-SEDD, which is the recently developed new generation RISM-SCF theory [11], enables us to directly deal with SEDD. Because of its inherent character based on statistical mechanics for molecular liquid (RISM; reference interaction site model) [12, 13], solvation free energies (excess chemical potential) with various level computations can be systematically compared without sampling error. It should also be emphasized that the method allows us to employ a highly sophisticated ab initio molecular orbital theory such as CCSD(T) within the same framework of QM/MM. In particular, p$K_{\mathrm{a}}$ and p$K_{\mathrm{b}}$ of glycine in aqueous solution are considered using various level of computations. Furthermore, the obtained wave functions are analyzed in terms of the resonance structures [14] and the differences between gas phase and aqueous solution are discussed.

Glycine is the simplest amino acids and often utilized as a useful molecule to understand a stability of zwitterion in aqueous media[7, 8, 25, 27]. These studies so far focused on the energy difference between the neutral and zwitterionic forms, namely the intramolecular proton transfer. In this paper, a special attention is paid to the role of bulk water surrounding the glycine. An accurate estimation and systematic assessment of absolute values of the p$K_{\mathrm{a}}$ and

p$K_{\rm b}$ are indispensable to understand the protonation and deprotonation process in reality.

## 3.2 Method

RISM [12, 13], RISM-SCF [10, 11] and RISM-SCF-SEDD [9] are molecular level theories of liquid based on statistical mechanics. They provide solvation free energy and solvation structure in terms of site-site correlation function such as radial distribution function (RDF). Since the details of the theory have been documented elsewhere [15], here we just provide an outline of the procedure. In the theory, total energy of the infinitely dilute solution of chemical species $i$, is defined as follows:

$$G_i = E_i^{\rm solute} + \delta\mu_i = E_i^{\rm isolated} + E_i^{\rm reorg} + \delta\mu_i. \tag{3.1}$$

$E_i^{\rm reorg}$ is defined as the difference in energy between a solute molecule in solvent ($E_i^{\rm solute}$) and that in the isolated state ($E_i^{\rm isolated}$).

$$E_i^{\rm reorg} = E_i^{\rm solute} - E_i^{\rm isolated} = \langle\Psi|H|\Psi\rangle - \langle\Psi_0|H|\Psi_0\rangle, \tag{3.2}$$

where $|\Psi\rangle$ and $|\Psi_0\rangle$ are wave functions in solution and in the gas phase, respectively. $\delta\mu_{\rm i}$ is computed by the following equation [18] coupled with Kovalenko-Hirata (KH) closure [19],

$$\delta\mu_i = \frac{\rho}{2\beta}\sum_{\alpha}\sum_{\gamma}^{\rm solvent}\int_0^{\infty} 4\pi r^2 dr\{h_{\alpha\gamma}^2\Theta(-h_{\alpha\gamma}) - 2c_{\alpha\gamma} - h_{\alpha\gamma}c_{\alpha\gamma}\}, \tag{3.3}$$

where $\rho$ is the number density of water, $\beta = 1/k_{\rm B}T$, and $\Theta$ is Heaviside step function. $h_{\alpha\gamma}$ and $c_{\alpha\gamma}$ are respectively total and direct correlation functions between $\alpha$(solute site) and $\gamma$(solvent site).

In the present study, the following equilibria were considered.

$$^{+}{\rm H_3NCH_2COOH}\ ({\rm Gly^{+}}) + {\rm H_2O} \rightleftharpoons\ ^{+}{\rm H_3NCH_2COO^{-}}\ ({\rm Gly^{\pm}}) + {\rm H_3O^{+}}, \tag{3.4}$$

and

$${\rm H_2NCH_2COO^{-}}\ ({\rm Gly^{-}}) + {\rm H_2O} \rightleftharpoons {\rm Gly^{\pm}} + {\rm OH^{-}}. \tag{3.5}$$

By regarding the species concerning the reaction as a 'solute' in infinitely dilute aqueous solution [17], p$K_{\rm a}$ is computed by considering the difference in free energy of equilibrium Eqs. (3.4) and (3.5) in *pure* water, which is directly computed with the present molecular level theory.

$$\mathrm{p}K_{\rm a} = \frac{\Delta G_{\rm a}}{2.303RT} - \log[\mathrm{H_2O}], \quad \mathrm{p}K_{\rm b} = \frac{\Delta G_{\rm b}}{2.303RT} - \log[\mathrm{H_2O}], \tag{3.6}$$

where $\Delta G_{\rm a}$ and $\Delta G_{\rm b}$ are obtained by considering a standard thermodynamic cycle. Note that we use "$\Delta$" for changes of quantities associated with the chemical reaction (proton transfer) and "$\delta$" for changes due to solvation.

$$\begin{aligned} \Delta G_{\rm a} &= \left(G_{\mathrm{Gly^{\pm}}} + G_{\mathrm{H_3O^+}}\right) - \left(G_{\mathrm{Gly^+}} + G_{\mathrm{H_2O}}\right) \\ &= \Delta G_{\rm a}^{\rm gas} + \Delta G_{\rm a}^{\rm solv}, \end{aligned} \tag{3.7}$$

$\Delta G_{\rm a}^{\rm gas}$ is the free energy difference of the equilibrium Eq. (3.4) in gas phase defined as a sum of $\Delta E_{\rm a}^{\rm isolated}$ and the correction from the standard frequency analysis procedure. $\Delta G_{\rm a}^{\rm solv}$ is essentially attributed to the difference in solvation free energy, $\Delta\delta\mu_{\rm a}$,

$$\Delta\delta\mu_{\rm a} = \left(\delta\mu_{\mathrm{Gly^{\pm}}} + \delta\mu_{\mathrm{H_3O^+}}\right) - \left(\delta\mu_{\mathrm{Gly^+}} + \delta\mu_{\mathrm{H_2O}}\right), \tag{3.8}$$

and the contribution from $\Delta E_{\rm a}^{\rm reorg}$ is added. Yagasaki et al. reported that a correction ($\Delta E^{\rm corr}$) is necessary to properly evaluate the effect of electron transfer to bulk water (-9.0 kcal/mol), which is beyond the scope of hybrid-type method[2]. Then the correction is also added to evaluation of $\Delta G_{\rm a}^{\rm solv}$. $\Delta G_{\rm b}$ is evaluated in a similar manner.

$$\begin{aligned} \Delta G_{\rm b} &= \left(G_{\mathrm{Gly^{\pm}}} + G_{\mathrm{OH^-}}\right) - \left(G_{\mathrm{Gly^-}} + G_{\mathrm{H_2O}}\right) \\ &= \Delta G_{\rm b}^{\rm gas} + \Delta G_{\rm b}^{\rm solv}. \end{aligned} \tag{3.9}$$

Equations (3.7) and (3.9) correspond to SEDD model, in which electronic structure and its spatial electron density distribution are directly treated. In PC model, the same expression (3.3) is used for the evaluation of solvation free energy, but a set of point charge was employed ($\delta\mu^{\rm PC}$). Since the wave function does not explicitly appear in this model, the related contributions ($\Delta E_{\rm a}^{\rm reorg}$, $\Delta E_{\rm b}^{\rm reorg}$ and $\Delta E^{\rm corr}$) are not necessary to be included. The $\Delta G_{\rm a}^{\rm solv}$ in Eq. (3.7)

Table 3.1: Basis sets and their indices use in the present study.

| index | the number of Gauss function of $Gly^+$ | basis set |
|---|---|---|
| I | 105 | 6-31G(d,p) |
| II | 105 | cc-pVDZ |
| III | 131 | 6-311G(d,p) |
| IV | 161 | 6-311G(2d,p) |
| V | 179 | 6-311G(2d,2p) |
| VI | 191 | 6-311G(3d,p) |
| VII | 227 | 6-311G(3d,3p) |
| VIII | 241 | 6-311G(3df,p) |
| IX | 265 | cc-pVTZ |
| X | 277 | 6-311G(3df,3p) |
| XI | 179 | aug-cc-pVDZ |
| XII | 247 | 6-311+G(3d,3p) |
| XIII | 253 | 6-311++G(3d,3p) |
| XIV | 297 | 6-311+G(3df,3p) |
| XV | 303 | 6-311++G(3df,3p) |
| XVI | 425 | aug-cc-pVTZ |

and $\Delta G_{\mathrm{b}}^{\mathrm{solv}}$ in Eq. (3.9) are simply calculated as follows.

$$\Delta G_{\mathrm{a}}^{\mathrm{solv}} = \Delta\delta\mu_{\mathrm{a}}^{\mathrm{PC}}, \quad \Delta G_{\mathrm{b}}^{\mathrm{solv}} = \Delta\delta\mu_{\mathrm{b}}^{\mathrm{PC}}. \tag{3.10}$$

### 3.2.1 Computational Details

Examined basis sets and their indices used in the present study were listed in Table 3.1. KH closure was coupled with RISM equation in both method. Lennard-Jones parameters used for glycine were the standard OPLS and collected in Table 3.2. The temperature was 298.15K and the density was 1.00 g/cm$^3$ (= 0.033426 molecule/Å$^3$). All the molecular geometries of chemical species appeared in Eqs. (3.4) and (3.5), i.e. three forms of glycine, water, hydronium and hydroxyl ions, were optimized in aqueous solution using RISM-SCF-SEDD method with DFT (B3LYP) functional. It should be mentioned that the optimization of $Gly^{\pm}$ in gas phase was impossible because the proton in the amino group automatically transferred to the carboxyl group without barrier in the optimization. The zwitterion-form does not exist in gas phase and the solvation effect is essential to stabilize this form. In addition to DFT, CCSD(T) combined

Table 3.2: Potential parameters for solvent and solute species.

| molecule | site | charge / \|e\| | $\sigma$ / Å | $\epsilon$ / kcal mol$^{-1}$ |
|---|---|---|---|---|
| [a]$H_2O$ | O | -0.820 | 3.166 | 0.155 |
| | H | 0.410 | 1.000 | 0.056 |
| [b]$H_3O^+$, [b]$OH^-$ | O | | 3.166 | 0.155 |
| | H | | 1.000 | 0.056 |
| [c]$Gly^+$, $Gly^{\pm}$ and $Gly^-$ | H1,H2 | | 1.000 | 0.056 |
| | H3 ($Gly^+$, $Gly^{\pm}$ only) | | 1.000 | 0.056 |
| | N4 | | 3.250 | 0.170 |
| | C5 | | 3.500 | 0.066 |
| | H6,H7 | | 2.500 | 0.030 |
| | C8 | | 3.750 | 0.105 |
| | O9, O10 | | 2.900 | 0.210 |
| | H11 ($Gly^+$ only) | | 1.000 | 0.056 |

[a] SPC like model [26]. [b] ref.[28]. [c] ref.[29].

with RISM-SCF-SEDD method was also performed at the given geometry to compute energy components of $\Delta G_{\rm a}$ and $\Delta G_{\rm b}$. Hereafter, we will call this evaluation procedure SEDD model, where the spatial distribution is fully taken into account. It is well known that Mulliken gross charge is not adequate to represent the electrostatic field around a molecule. Hence a set of point charge (PC) obtained from the RISM-SCF-SEDD procedure was adopted, where the charge was given as the sum of expansion coefficients $d_{\rm i}$ on the site *i* [11]. A conventional RISM was then employed to evaluate $\Delta G_{\rm a}^{\rm solv}$ and $\Delta G_{\rm b}^{\rm solv}$ using this PC (PC model). Note that PC model also depends upon the basis set used in the DFT (B3LYP) computation.

All calculations were carried out by GAMESS package[20] modified by our laboratory.

## 3.3 Results and Discussion

### 3.3.1 p$K_{\rm a}$, p$K_{\rm b}$ ands structural characterization

Figure 3.1 plots the basis set dependency of $\Delta G_{\rm a}^{\rm gas}$ and $\Delta G_{\rm b}^{\rm gas}$ computed with various level of theory, namely DFT (diamonds), HF (triangles) and CCSD(T) (squares). The DFT result of $\Delta G_{\rm b}^{\rm gas}$ shows remarkable dependency compared to the $\Delta G_{\rm a}^{\rm gas}$, but each of them respectively converges to a constant value with larger basis set. In particular, addition of diffuse function is

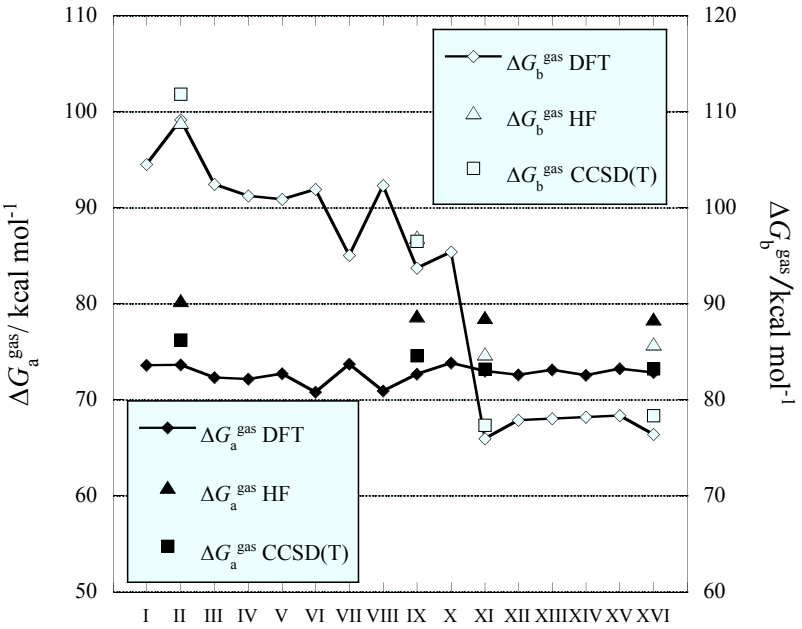


Figure 3.1: Basis set dependency of $\Delta G_{\mathrm{a}}^{\mathrm{gas}}$ and $\Delta G_{\mathrm{b}}^{\mathrm{gas}}$.

crucial to properly compute $\Delta G_{\mathrm{b}}^{\mathrm{gas}}$. This may be related to the spatial extension of negatively charged carbonyl group in $\mathrm{Gly}^{-}$. It is also noted that DFT and CCSD(T) give very close value with larger basis set, suggesting that DFT provides sufficiently accurate result in the present system. HF result is not much far from the these values but slightly overestimates.

Figure 3.2 shows p$K_{\mathrm{a}}$ (upper) and p$K_{\mathrm{b}}$(lower) computed with SEDD and PC models. In evaluation of p$K_{\mathrm{a}}$ with SEDD models, both of DFT (diamonds) and CCSD(T) (squares) results are very similar to each other and reasonably agree with the experimental value (2.34[21]), especially with larger basis sets. On the contrary, despite being reasonable with middle size basis set, PC model (circles) gives largely negative values with larger basis sets. Obviously this is related to the difference in the description of charge distribution between the two models. From the detailed analysis of the solvation free energy $\Delta\delta\mu_{\mathrm{a}}$, we found that the discrepancy originated from the solvation free energy of $\mathrm{Gly}^{\pm}$ as described below. In a similar way, p$K_{\mathrm{b}}$ in the lower panel also exhibits serious discrepancy between SEDD and PC models. While SEDD models with larger basis set show reasonable agreement with the experimental value (4.42[21, 22]), the results by PC model are considerably greater, especially with larger basis

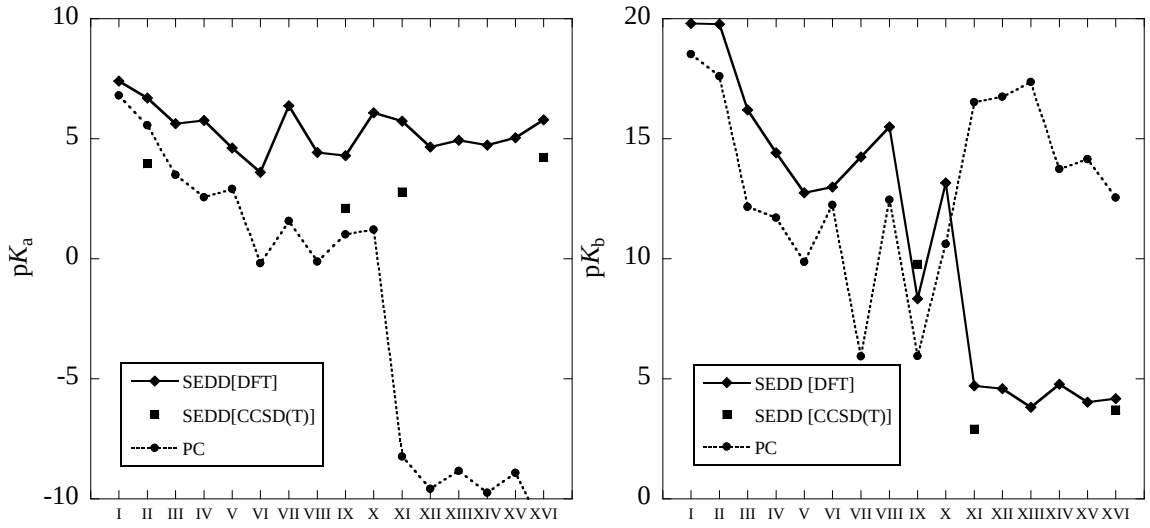


Figure 3.2: The basis set dependency of $pK_a$ (upper panel) and $pK_b$ (lower panel) evaluated by SEDD (B3LYP and CCSD(T) methods) and PC model.

sets. These indicate that the spatial extension of electronic distribution with diffuse basis set is difficult to be represented in PC model, which is not capable to describe the electric field in the vicinity of solute molecules and to correctly evaluate $pK_a$ and $pK_b$.

The molecular structure and electronic structure of the glycine are focused on to look closely at the difference between SEDD and PC model. Figure 3.3 depicts the optimized structures of $Gly^+$, $Gly^\pm$ and $Gly^-$ computed by B3LYP/XII. $pK_a$ and $pK_b$ computed with this level of theory are satisfactory close to their experimental values as described above. The obtained geometrical parameters show good agreement with X-ray [23] and neutron diffraction [24] experiments. For example, the computed bond lengths, 1.528Å(C5–C8), 1.498Å(C5–N), 1.254Å(C8–O9) and 1.251Å(C8–O10) in $Gly^\pm$, show very good agreement with corresponding experimental values (1.525Å, 1.475Å, 1.252Å and 1.251Å, respectively) [24].

The dipole moment is remarkably enhanced in aqueous solution, especially on $Gly^\pm$. Takahashi et al [7] pointed out that a significant increase to 15.1 Debye due to the solvation effect, and Sun et al. recently reported the average dipole moment is around 16.5 Debye in aqueous solution [25]. The present result (16.2 Debye) shows very good agreement with these values. In the figure, the assigned charges in PC model are indicated in parentheses. The dipole moments evaluated with these charges are 9.9 ($Gly^+$), 15.1 ($Gly^\pm$) and 6.4 ($Gly^-$) Debye. It

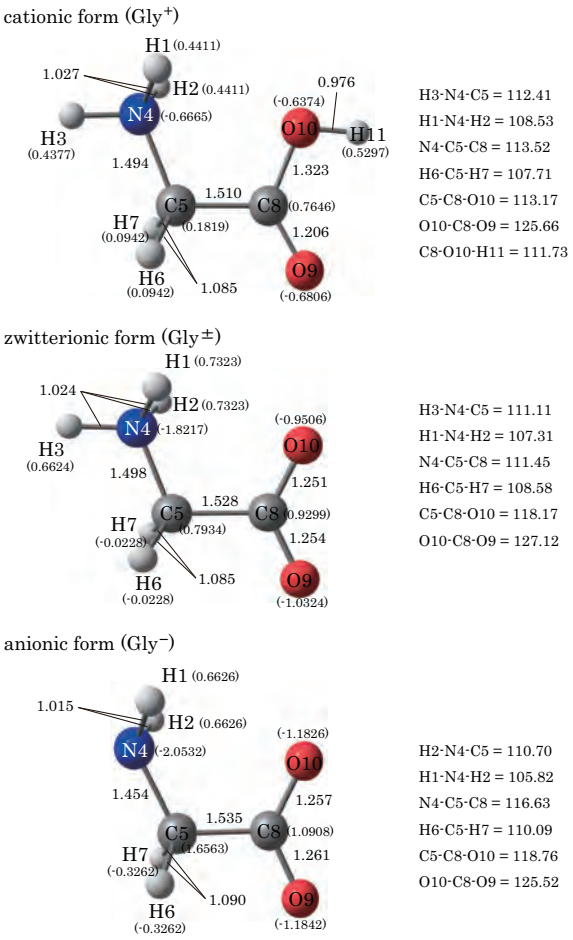


Figure 3.3: The optimized geometrical parameters in three forms of glycine molecules in aqueous solution. The values in parentheses are the point charge obtained with B3LYP/XII level (unit:|e|). Bond distances and angles are given by Å and degree unit, respectively.

is interesting to note that they are very close to those computed directly from wave function, 10.6, 16.2 and 6.5 Debye, respectively.

Because the contribution from the proton transfer reaction ($\Delta G_{\mathrm{a}}^{\mathrm{gas}}$ and $\Delta G_{\mathrm{b}}^{\mathrm{gas}}$) is the same in SEDD and PC models, the discrepancy in p$K_{\mathrm{a}}$ and p$K_{\mathrm{b}}$ arises from the solvation free energy, $\Delta G_{\mathrm{a}}^{\mathrm{solv}}$ and $\Delta G_{\mathrm{b}}^{\mathrm{solv}}$. The discussion is more clearly shown in Figure 3.4, where the differences of the solvation free energy ($\delta\mu_i$) in Eq.(3.3) obtained with SEDD and PC are compared. In the RISM framework, the electrostatic interaction is described in the site-site interaction and multipole interaction beyond dipole moment is treated. As shown in the figure, the computed $\delta\mu_i$'s of $H_2O$, $H_3O^+$ and $Gly^+$ respectively resemble both with PC and SEDD

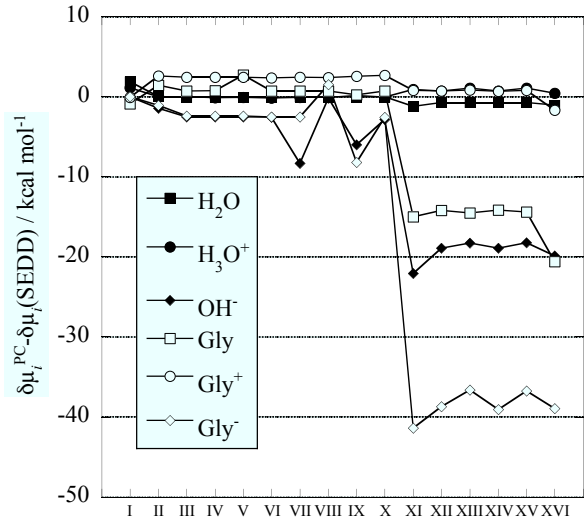


Figure 3.4: Basis set dependency of the difference in solvation free energy between SEDD ($\delta\mu_i$) and PC ($\delta\mu_i^{\mathrm{PC}}$) models ($i$=$H_2O$, $H_3O^+$, $OH^-$, Gly, $Gly^+$ and $Gly^{\pm}$).

models, indicating that PC model properly express the solute-solvent interaction, at least with the same accuracy provided with SEDD model. But the situation is very different in the anions and zwitterion. With the diffuse basis set, the difference in $\delta\mu_i$ between PC and SEDD models becomes considerably larger. Hence we conclude that the discrepancy between these models may be largely attributed to $\delta\mu_{\mathrm{Gly}^{\pm}}$ in the evaluation of $\Delta G_{\mathrm{a}}^{\mathrm{solv}}$, and to $\delta\mu_{\mathrm{Gly}^-}$ and $\delta\mu_{\mathrm{OH}^-}$ in $\Delta G_{\mathrm{b}}^{\mathrm{solv}}$. The total hydration free energy ($\delta\mu_{\mathrm{Gly}^{\pm}}$) computed with PC model is smaller by 15 kcal/mol than that with SEDD method, which is very close to 16 kcal/mol reported by QM/MM-ER method[7].

The hydration free energy may be further analyzed in terms of functional group using the formal partition of $\delta\mu_i$. Since the equation (3.3) is described as the sum of the contribution from the solute site ($\alpha$), it is formally possible to partition into each site contribution. Figure 3.5 depicts the site contributions to $\delta\mu_{\mathrm{Gly}^{\pm}}$ and $\delta\mu_{\mathrm{Gly}^-}$ computed at B3LYP/XII level. In the upper panel, the discrepancy between the two models is mainly attributed to the contribution of carboxyl group. The discrepancy becomes greater in the anion shown in the lower panel. Not only the carboxyl group, but amino and methylene groups also cause the difference of more than 40 kcal/mol in total $\delta\mu_i$.

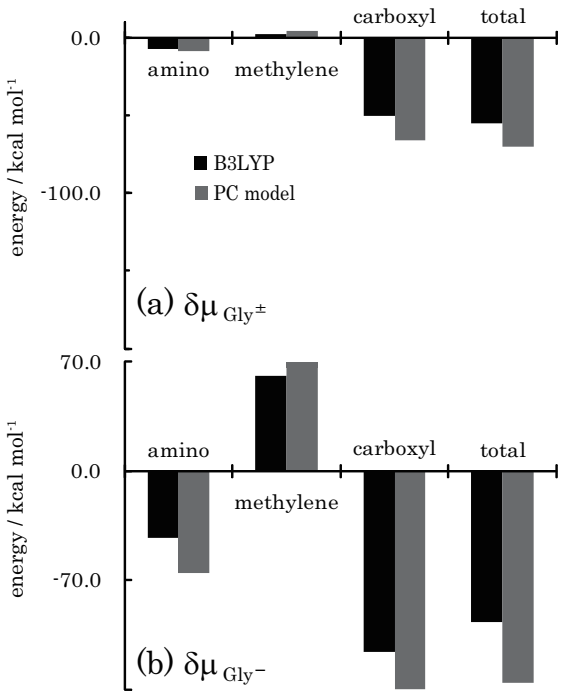


Figure 3.5: Comparison of the site ($\alpha$ in Eq. (3.3)) contributions from the moiety. (a) $\delta\mu_{\mathrm{Gly}^{\pm}}$ and (b) $\delta\mu_{\mathrm{Gly}^{-}}$ obtained with SEDD and PC models.

### 3.3.2 Hydration structures near carboxyl and amino groups

Figure 3.6 shows RDFs around the amino and carboxyl groups in $\mathrm{Gly}^{\pm}$ (left panel) and $\mathrm{Gly}^{-}$ (right panel) . The solid lines denote RDFs computed with SEDD method (B3LYP/XII) and the dots lines PC model, respectively. Ow and Hw represent the oxygen and hydrogen site of solvent water molecule.

In the left panel, H3–Ow and N4–Ow exhibit a typical hydrogen bonding between $\mathrm{Gly}^{\pm}$ and solvent water. Both of the solid and dots lines look like each other, indicating that the hydration structures around the amino group ($-\mathrm{NH}_3^+$) are very similar in SEDD and PC models. But the structure in the vicinity of carboxyl group ($-\mathrm{COO}^-$) is slightly different between the two models. The height of O9–Hw RDF computed with PC model is evidently higher than those with SEDD model, which is consistent with the analysis on solvation free energy mentioned above. In addition, the RDFs by SEDD model are consistent with the previous theoretical [7], [8](f), [25] and experimental [27] studies in terms of the first peak positions (H3-Ow,

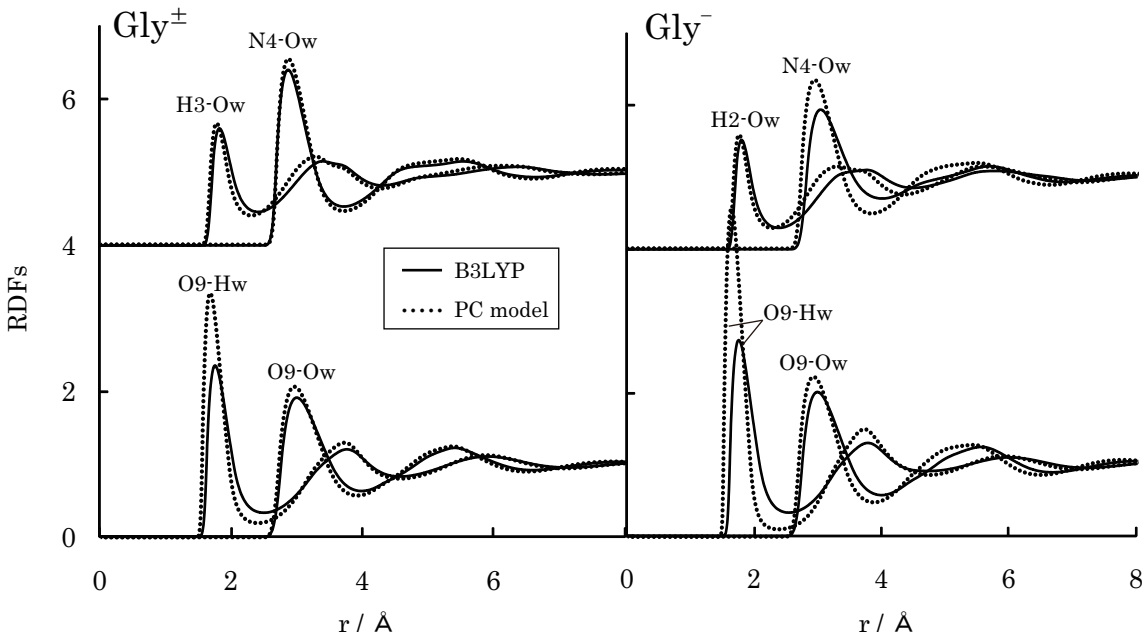


Figure 3.6: RDFs near the amino and carboxyl groups in $Gly^{\pm}$ (left panel) and $Gly^{-}$ (right panel). Solid and dots lines are respectively computed with SEDD and PC model.

1.83Å; N4-Ow, 2.87Å; O9-Hw, 1.76Å; O9-Ow, 3.00Å) and hydration numbers (1.15 around H3 calculated from H3-Ow RDF; 1.60 around O9 calculated from O9-Hw RDF).

The right panel shows the hydration structure around $Gly^{-}$ (right panel). Although the first peaks in H2–Ow RDF look similar between the two model, the second peak (3.0–4.0Å) in PC model is shifted to shorter distance. The peak of N4–Ow is clearly different. The hydration structure around the amino group ($–NH_2$) is not properly described with PC model. As shown in Figure 3.5, the contribution from the amino group is one of the error sources of the $pK_b$ estimation in PC model. The difference becomes much greater in the carboxyl group ($–COO^{-}$). While the O9–Ow look similar in the both models, the first peak of O9–Hw RDF computed with PC is considerably high compared to that with SEDD model. Moreover, its position is shifted into the shorter distance. We concluded therefore that SEDD model is necessary to obtain an appropriate hydration structure. In other words, the hydration around carboxyl group is overemphasized in PC model.

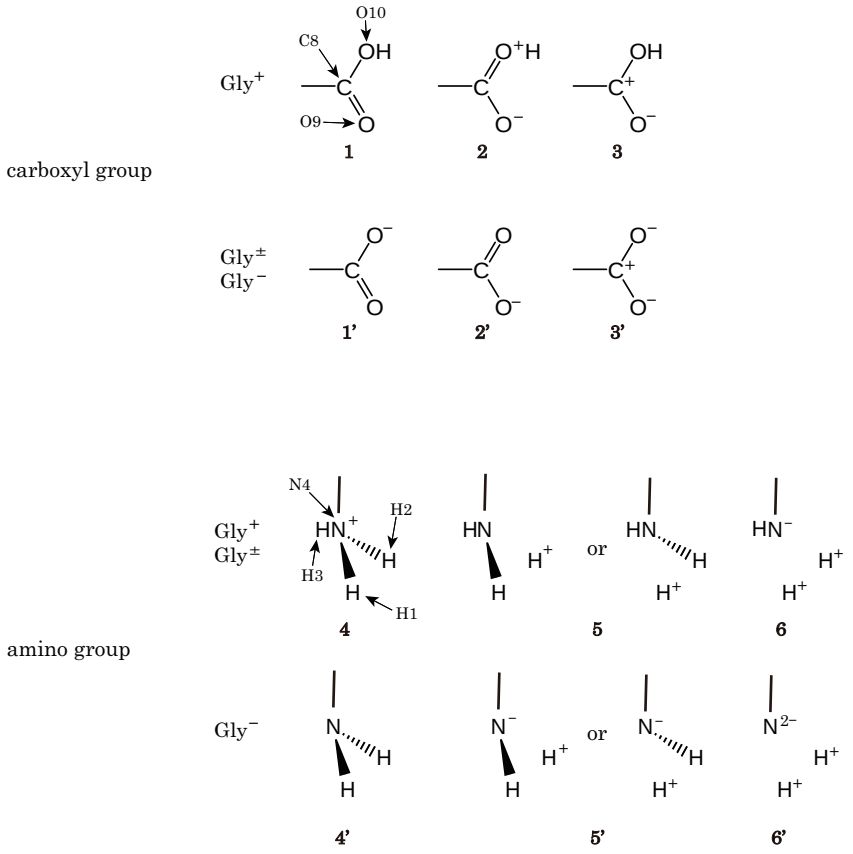


Figure 3.7: The resonance structures of carboxyl (upper) and amino (lower) groups in $Gly^{+}$, $Gly^{\pm}$ and $Gly^{-}$ molecules.

### 3.3.3 Electronic structures of amino and carboxyl group in view of resonance structures

Figure 3.7 illustrates representative resonance structures of carboxyl (upper) and amino (lower) groups. At the end of this work, the analysis on the electronic structure is briefly described. The full description on the computed weights is found in Figure 3.8. Let us start with the carboxyl group. As shown in the scheme, three resonance structures are expected to represent the electronic structure. For the $Gly^{+}$, where a proton is attached to the carboxyl group, the weights for **1**, **2** and **3** are respectively 32.5, 7.6 and 41.8 %. The absence of the proton slightly changes the weight; 21.0(**1**′), 19.6(**2**′) and 41.1(**3**′) % are computed for the $Gly^{\pm}$. The weights for $Gly^{-}$ are essentially the same with $Gly^{\pm}$. **1**′ and **2**′ in $Gly^{\pm}$ and $Gly^{-}$ are distinguished by the orientation with respect to the amino group but practically equivalent

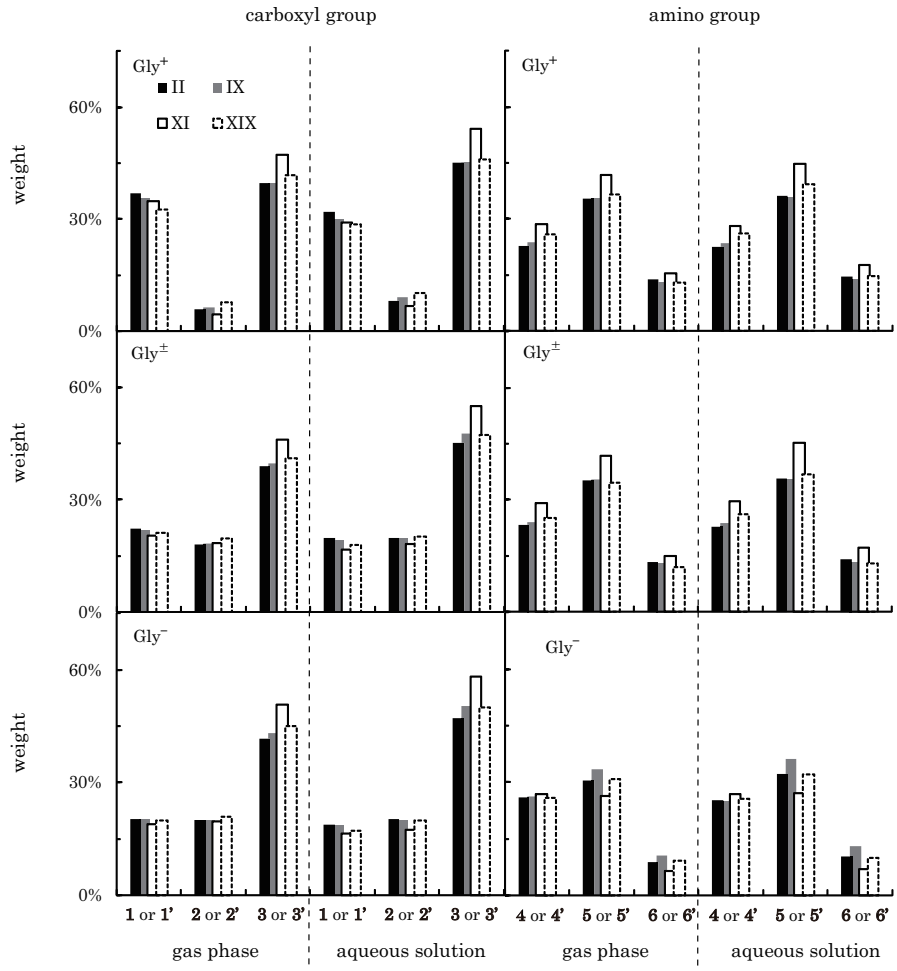


Figure 3.8: The basis set dependencies of resonance weights for carboxyl (left panels) and amino (right panels) groups in $Gly^{+}$, $Gly^{\pm}$ and $Gly^{-}$ molecules in gas phase and aqueous solution. The changes of weights for the both groups are small, whose interpretations therefore are the same using any basis sets shown here.

contribution.

It may be interesting to point out that the bond between O10 and H in $Gly^{+}$ shows covalent (**1**) rather than ionic (**2**) character. In aqueous solution, the ionic character is evidently enhanced but the difference is not significant. In any cases, the contribution from the charge-separated structure (**3** and **3**′) is dominative and increased by about 5 % by solvation effect.

Next is the amino group. For the wavefunction of $Gly^{+}$, the weight of **4**, **5** and **6** are respectively 25.9, 36.6 and 13.0 %. A similar results are obtained in $Gly^{\pm}$ (25.8, 30.8 and 9.2 %, respectively), implying little effect from the carboxyl group. In contrast to carboxyl group, the

charge-separated structure (**6**) is less contributed and the main structure is **5** or **4**, in which a positive charge is localized on a specific site. It should also be mentioned that the electronic structure of the amino group in aqueous solution is almost the same as in the gas phase, which is different from the carboxyl group case. Presumably, the hydration around the amino group is not strongly structured compared to the vicinity of the carboxyl group.

## 3.4 Conclusions

p$K_{\mathrm{a}}$ and p$K_{\mathrm{b}}$ of glycine in aqueous solution were computed using two different treatments of intermolecular electrostatic interaction, SEDD and PC model. RISM-SCF-SEDD method coupled with HF and other highly sophisticated theory, namely DFT(B3LYP) and CCSD(T), provided very good agreement with the experimental values for p$K_{\mathrm{a}}$ and p$K_{\mathrm{b}}$ when an appropriate basis sets with a diffuse function was employed. However, PC model exhibited considerable discrepancy, notably with large basis set. The poor representation of PC was caused by serious underestimation of the hydration free energy and the hydration structures are remarkably different between SEDD and PC models, especially around carboxyl group in $\mathrm{Gly}^{\pm}$ and around both carboxyl and amino groups in $\mathrm{Gly}^{-}$. It is essential for an appropriate prediction of p$K_{\mathrm{a}}$ and p$K_{\mathrm{b}}$ to employ accurate intermolecular electrostatic potential near the functional groups. The spatial distribution of electrons is indispensable, especially in the case of molecules containing either a conjugated or negatively charged group. We suggest using a highly sophisticated method and at least triple-$\zeta$ class basis sets with diffuse function is desirable to obtain accurate p$K_{\mathrm{a}}$ and p$K_{\mathrm{b}}$.

# Bibliography


[1] Coote, M. L.; Ho, J. Theor Chem Acc 2009, 125, 3.

[2] Yagasaki, T.; Iwahashi, K.; Saito, S.; Ohmine, I. J Chem Phys 2005, 122, 144504.

[3] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impay, R. W.; Klein, M. L. J Am Chem Soc 1983, 79, 926.

[4] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.

[5] Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. J Am Chem Soc 1984, 106, 765.

[6] Takahashi, H.; Matubayasi, N.; Nakahara, M.; Nitta, T. J Chem Phys 2004, 121, 3989.

[7] Takahashi, H.; Kawashima, Y.; Nitta, T.; Matubayasi, N. J Chem Phys 2005, 123, 124504.

[8] (a) Bonaccorsi, R.; Palla, P.; Tomasi, J. J Am Chem Soc 1984, 106, 1945. (b) Jensen, H. J.; Gordon S. M. J Am Chem Soc 1995, 117, 8159. (c) Ding, Y.; Krogh-Jesperson, K. J Comput Chem 1996, 17, 338. (d) Tortonda, R. F.; Pascual-Ahuir, L. J.; Silla E.; Tuñón, I. Chem Phys Lett 1996 260, 21. (e) Okuyama-Yoshida, N.; Kataoka, K.; Nagaoka, M.; Yamabe, K. J Chem Phys 2000 113, 3519. (f) Leung, K.; Rempe, B. S. J Chem Phys 2005 122, 184506.

[9] (a) Iida, K.; Yokogawa, D.; Sato, H.; Sakaki, S. Chem Phys Lett 2007, 443, 264. (b) Hayaki, S.; Yokogawa, D.; Sato, H.; Sakaki, S. Chem Phys Lett 2008, 458, 329. (c)

Hayaki, S.; Kido, K.; Yokogawa, D.; Sato, H.; Sakaki, S. J Phys Chem B 2009, 113, 8227. (d) Iida, K.; Yokogawa, D.; Sato, H.; Sakaki, S. Phys Chem Chem Phys 2009, 11, 8556. (e) Hayaki, S, Kido, K.; Yokogawa, D.; Sato, H.; Sakaki, S. Phys Chem Chem Phys , 2010, 12, 1822.

[10] (a) Ten-no, S.; Hirata, F.; Kato, S. J Chem Phys 1994, 100, 7443. (b) Sato, H.; Hirata, F.; Kato, S. J Chem Phys 1996, 105, 1546.

[11] (a) Yokogawa, D.; Sato, H.; Sakaki, S. J Chem Phys 2007, 126, 244504. (b) Yokogawa, D.; Sato, H.; Sakaki, S. J Chem Phys 2009, 131, 214504.

[12] Chandler, D.; Andersen, H. C. J Chem Phys 1972, 57, 1930.

[13] (a) Hirata, F.; Rossky, P.; Chem Phys Lett 1981, 83, 329. (b) Hirata, F.; Rossky, P.; Pettitt, B. M. J Chem Phys , 1983, 78, 4133.

[14] (a) Ikeda, A.; Nakao, Y.; Sato, H.; Sakaki, S. J Phys Chem A 2006, 110, 9028. (b) Ikeda, A.; Yokogawa, D.; Sato, H.; Sakaki, S. Chem Phys Lett 2007, 424, 449. (c) Ikeda, A.; Yokogawa, D.; Sato, H.; Sakaki, S. Int J Quantum Chem 2007, 107, 3132.

[15] (a) Hansen, J. P.; McDonald, I. R. Theory of Simple Liquids, Academic Press, 1990. (b) Hirata, F. Ed. Molecular Theory of Solvation Understanding Chemical Reactivity, Kluwer-Springer, 2004. (c) Mennucci, B.; Cammi, R. Eds. Continuum Solvation Models in Chemical Physics: From Theory to Applications, John Wiley & Sons Inc, 2008.

[16] Kido, K.; Sato, H.; Sakaki, S. J Phys Chem B 2009, 113, 10509.

[17] (a) Sato, H.; Hirata, F. J Phys Chem A 1998, 102, 2603. (b) Sato, H.; Hirata, F. J Phys Chem B 1999, 103, 6596. (c) Yoshida, N.; Ishizuka, R.; Sato, H.; Hirata, F. J Phys Chem B 2006, 110,8451.

[18] Singer, J. S.; Chandler, D. Mol Phys 1985, 55, 621.

[19] Kovalenko, A.; Hirata, F. J Chem Phys 1999, 110, 10095.

[20] Ver.11, Apr. 2008(R1) Schmidt, W. M.; Baldridge, K. K.; Boatz, A. J.; Elbert S. T.; Gordon, S. M.; Jensen, H. J.; Koseki, S.; Matsunaga, N.; Nguyen, A. K.; Su, J. S.; Windus, L. T.; Dupuis, M.; Montgomery, A. J J Comput Chem 1993, 14, 1347.

[21] Lide, D. R. Ed. CRC Handbook of Chemistry and Physics 87th Edition, 2003;7-1.

[22] $pK_{a2}$ is an acidic dissociation constant of the amino group and 9.58 found in ref.19. We therefore obtained $pK_b$ using the value; $pK_b = pK_w - pK_{a2}$ = 14.0 - 9.58 = 4.42.

[23] (a) Albrecht, G.; Corey, B. R. J Am Chem Soc 1939, 61, 1087. (b) Marsh, E. R. Acta Cryst 1958,11,654.

[24] Power, F. L.; Turner, E. K.; Moore, H. F. Acta Cryst 1976, B32, 11.

[25] Sun, J.; Bousquet, D.; Forbert, H.; Marx, D. J Chem Phys 2010, 133, 114508.

[26] Sato, H.; Sakaki, S. J Phys Chem A 2002, 106, 2300.

[27] (a) Kameda, Y.; H. Ebata, H.;, Usuki, T.; Uemura, O.; Misawa, M. Bull Chem Soc Jpn 1994 67, 3159. (b) Kameda, Y.; Mori, T.; Nishiyama, T.; Usuki, T.; Uemura O. Bull Chem Soc Jpn 1996 69, 1495. (c) Sugawara, K; Kameda, Y.; Usuki, T.; Uemura, O.; Fukunaga, T. Bull Chem Soc Jpn 2000, 73, 1967. (d) Kameda, Y.; Fukuhara, K.; Mochizuki, K.; Naganuma, H.; Usuki, T.; Uemura, O. J Non-Cryst Sol 2002 312-314, 433. (e) Sasaki, M.; Kameda, Y.; Yaegashi, M.; Usuki, T. Bull Chem Soc Jpn 2003, 76, 2293. (f) McLain, S. E.; Soper, A. K.; Watts, A. J Phys Chem B 2006, 110, 21251.

[28] Sato, H.; Hirata, F. J Phys Chem A 1998, 102, 2603.

[29] Jorgensen, W. L.; Tirado-Rives, J. J Am Chem Soc 1988, 110, 1657.

# Chapter 4

# A molecular level study of selective cation capture by a host-guest mechanism for 25,26,27,28-tetramethoxycalix[4]arene in $MClO_4$ solution (M=Na, K)

## 4.1 Introduction

A specific interaction between a few molecules is often referred as molecular recognition. Channel proteins, which are typical examples in nature, selectively recognize cation such as proton, sodium or potassium ion. Essentially the same process is also found in much simpler chemical systems called host-guest complex. Calix[$n$]arene[1] consisting of $n$ anisole units is one of the representatives. It takes several conformations with respect to the anisole-unit direction, and the conformational change is directly related to the selective cation capture. 25,26,27,28-tetramethoxycalix[4] arene (the molecule is simply called "calixarene" in this paper) is a simple but a fundamental derivative: It captures the lithium and sodium ions with the four anisole units in the "cone" conformer. The "partial-cone" conformer, which is produced by the rotation of one unit, recognizes the potassium and argentum ions[2] as shown in 4.1. These conformers are in equilibrium when metal salt is absent[3].

Since the calixarene is solved into electrolyte solution, the alkali cation plays two different roles in the system, *i.e.*, the guest and co-solvent. While the direct interaction between calixarene and cation is basically interpreted in *ab initio* quantum chemistry, many-body interaction among the host-guest molecules and the surrounding solvent including cation is described

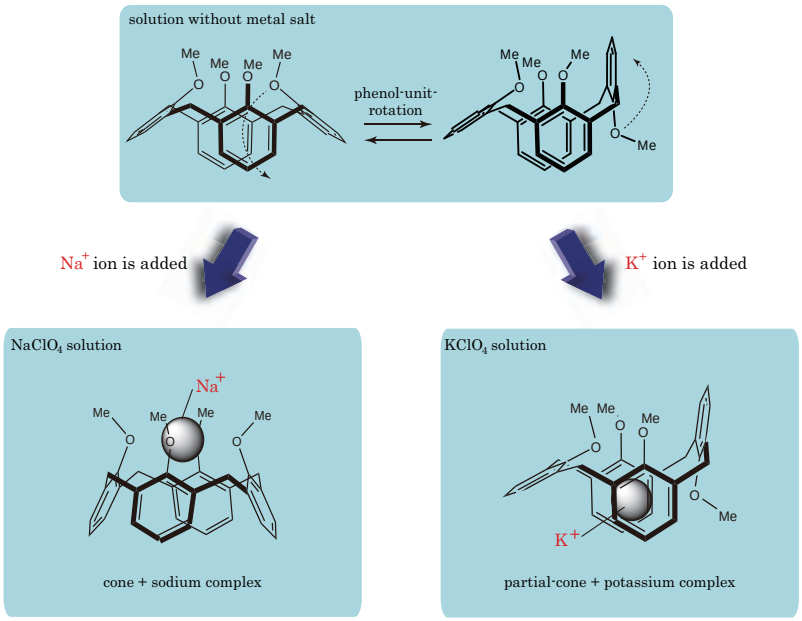


Figure 4.1: Scheme of cone and partial-cone conformers. The empty calixarene is in the equilibrium in the solution phase with respect to the anisole unit rotation, and it recognizes sodium cation by the cone and potassium by the partial-cone.

in statistical mechanics. It is an interplay between them that governs the mechanism of cation recognition. In addition, it is well known that the counter anion also plays a crucial role to characterize the system[13]. Because of this difficulty, computational studies on calixarene and related derivatives have been very limited so far. Only few studies are reported utilizing a standard quantum chemical calculation with dielectric continuum solvents[4], QM/MM[5] and classical molecular simulations[3, 6]. To our best knowledge, there is no report on the electrolyte solution containing calixarenes due to the strong Coulombic interaction among ions, which generally makes it infeasible to accurately evaluate free-energy of the system based on molecular simulation technique.

In the present study, a hybrid methodology of quantum chemistry and statistical mechanics for molecular liquids is utilized to elucidate the recognition mechanism. We employ a new generation reference interaction site model self-consistent field (RISM-SCF)[7, 8] approach that enables us to treat spatial electron density distribution (SEDD). Thanks to the inherent nature of RISM theory, solvation free energy can be computed in the analytical fashion[9], together

with molecular level information of solvent distribution including ions. RISM-SCF-SEDD is a promising tool to treat solvation effect at molecular level, which is regarded as an alternative to QM/MM method. The recently developed multi-component solvent version has been successfully applied to Diels-Alder reaction[10] and SN2 reaction[11] in room-temperature ionic liquids, as well as $pK_a$ prediction in HCl aqueous solution[12]. In particular, computation of solution with a wide range of pH was realized for the first time in this $pK_a$ study.

A goal of the present study is to reveal the molecular picture of the selectivity through the accurate computation of the association free-energies between calixarene and cation. Because of the importance of the dispersion interaction between host-guest molecules[4][d], MP2 method was coupled with RISM-SCF-SEDD method. This is the first report on the molecular recognition of calix[$n$]arene in a realistic solution model based on a first principle theory for molecular system.

## 4.2 Methodology and Computational Details

RISM-SCF-SEDD is a hybrid approach of quantum chemistry and statistical mechanics for molecular liquids. It provides solvation free energy and solvation structure in terms of radial distribution functions (RDFs) that are consistent with electronic structure of the solute molecule. Since the details of the method have been already described elsewhere, please refer to the papers[14, 15] and textbooks[16] if readers are not familiar with the method.

Let us consider the association constant $K$ concerning the capture of cation $M^+$ by the calixarene X[2],

$$\mathrm{X} + \mathrm{M}^+ \overset{K}{\rightleftharpoons} \mathrm{X}\cdot\mathrm{M}^+ \ (\mathrm{M} = \mathrm{Na},\ \mathrm{K}), \quad \text{and} \quad \log K = -\frac{\delta G}{RT\ln 10}. \tag{4.1}$$

$\delta G$ is the free energy difference of the equilibrium. Note that $\delta G < 0$ means that the cation is "captured" by the calixarene. In RISM-SCF-SEDD theory, the free energy of a solvated species $a$ is described as,

$$G_a = G_a^{\mathrm{gas}} + E_a^{\mathrm{reorg}} + \Delta\mu_a, \tag{4.2}$$

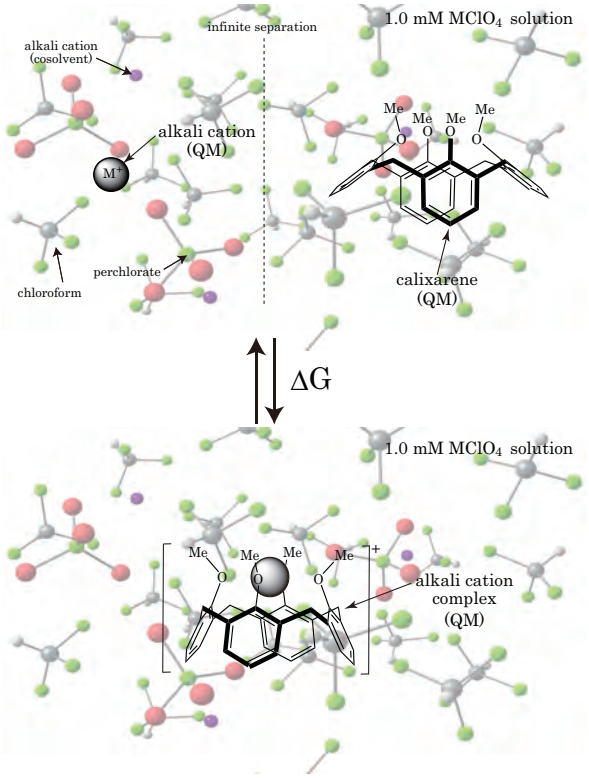


Figure 4.2: The computational scheme of RISM-SCF-SEDD method for the cone conformer. While the calixarene and its cation complex represented by schematic structures are treated by quantum chemical method, electrolyte (sodium perchlorate) solution depicted by ball-stick model is dealt with RISM theory.

where $G_a^{\mathrm{gas}}$ is the free energy of $a$ in the isolated state. $E_a^{\mathrm{reorg}}$ denotes the electronic reorganization (distortion) energy of $a$ originated from the coupling between the electronic structure and solvation structure.[7, 8] $\Delta\mu_a$ is the excess chemical potential (solvation free energy), which is directly computed with RISM theory. Hence $\delta G$ is given by the sum of these three components,

$$\begin{aligned}\delta G &= (G^{\mathrm{gas}}_{\mathrm{X\cdot M^+}} - G^{\mathrm{gas}}_{\mathrm{X}} - G^{\mathrm{gas}}_{\mathrm{M^+}}) + (E^{\mathrm{reorg}}_{\mathrm{X\cdot M^+}} - E^{\mathrm{reorg}}_{\mathrm{X}} - E^{\mathrm{reorg}}_{\mathrm{M^+}}) + (\Delta\mu_{\mathrm{X\cdot M^+}} - \Delta\mu_{\mathrm{X}} - \Delta\mu_{\mathrm{M^+}}) \\ &= \delta G^{\mathrm{gas}} + \delta E^{\mathrm{reorg}} + \delta G^{\mathrm{solv}}.\end{aligned} \tag{4.3}$$

Calixarene X takes two conformations on the capture, cone and partial-cone, and their difference in free energy is also necessary. For the sake of convenience, the relative energy with respect to the partial-cone conformer, $\Delta\delta G$, is introduced. (We use "$\delta$" for changes of quantities associated with the complex formation and "$\Delta$" for the difference between the two

Table 4.1: Potential parameters for calixarene and solvent.

| molecule | site | charge / \|e\| | $\sigma$ / Å | $\epsilon$ / kcal mol$^{-1}$ |
|---|---|---|---|---|
| calixarene[6] | C (benzen ring) | [a]– | 3.550 | 0.0700 |
| | C (methylene) | [a]– | 3.500 | 0.0660 |
| | H (benzen ring) | [a]– | 2.420 | 0.0300 |
| | H (methylene) | [a]– | 2.500 | 0.0300 |
| | O | [a]– | 3.000 | 0.1700 |
| | C (methyl) | [a]– | 3.520 | 0.0670 |
| | H (methyl) | [a]– | 2.500 | 0.0300 |
| chloroform[25] | CH | 0.4200 | 3.800 | 0.0800 |
| | Cl | -0.1400 | 3.470 | 0.3000 |
| sodium[26] | Na | 1.0000 | 3.328 | 0.0028 |
| potassium[26] | K | 1.0000 | 4.739 | 0.0033 |
| perchlorate[27] | $Cl_a$ | 1.0812 | 3.471 | 0.2650 |
| | O | -0.5203 | 2.960 | 0.2100 |

[a] Automatically determined through RISM-SCF-SEDD procedure.

isomers.) Namely, if $\Delta\delta G < 0$, the cation is preferably captured by the cone, and if $\delta\Delta G > 0$ indicates that the cation is recognized by partial-cone. The corresponding components are also given by,

$$\Delta\delta G = \Delta\delta G^{\mathrm{gas}} + \Delta\delta E^{\mathrm{reorg}} + \Delta\delta G^{\mathrm{solv}}. \tag{4.4}$$

Geometry optimizations were carried out by Hartree-Fock (HF) method followed by MP2 energy evaluation. As shown below, dispersion interaction through MP2 was crucial to describe the association energy. Basis sets of C, H and O atoms were TZV(d,p)[17] and alkali metals (Na and K) 6-311+G(d). We dealt with the electrolyte solution consisting of three components; chloroform ($CHCl_3$), alkali metal cation and perchlorate anion ($ClO_4^-$). The dielectric consistent RISM equation (DRISM[18]) coupled with Kovalenko-Hirata closure[19] was employed and the dielectric constant of chloroform was set 4.81[20]. Potential parameters used in the present study were collected in Table 4.1. The temperature was 298.15K and the density of chloroform was 1.479 g/cm$^3$ (= 0.07460 molecule/Å$^3$)[21]. The concentration of alkali metal salt was fixed on 1.000 mM (= $6.022\times10^{-7}$ molecule/Å$^3$)[2]. For the comparison of the association free energy evaluation, PCM (polarization continuum model)[23] was also employed

Table 4.2: Computed $\delta G$ and $\Delta\delta G$ by MP2 and HF (in parentheses).

| | RISM-SCF-SEDD | | | | PCM | | | |
|---|---|---|---|---|---|---|---|---|
| | $Na^+$ | | $K^+$ | | $Na^+$ | | $K^+$ | |
| $\delta G$ (cone) | -44.94 | (-43.85) | -26.41 | (-21.27) | -9.51 | ( -8.25) | -0.35 | ( 7.92) |
| $\delta G$ (partial-cone) | -41.25 | (-32.71) | -31.31 | (-18.55) | -11.37 | ( -2.91) | -15.51 | ( 0.18) |
| $\Delta\delta G$ | -3.69 | (-11.14) | 4.90 | ( -2.72) | 1.86 | ( -5.34) | 15.16 | ( 7.74) |

All the values are given in kcal/mol.

Table 4.3: Energy components of $\delta G$ by MP2 and HF (in parentheses). The values are given in kcal/mol.

| | $Na^+$ | | | | $K^+$ | | | |
|---|---|---|---|---|---|---|---|---|
| | $\delta G^{\mathrm{gas}}$ | $\delta E^{\mathrm{reorg}}$ | $\delta G^{\mathrm{solv}}$ | $\delta G$ | $\delta G^{\mathrm{gas}}$ | $\delta E^{\mathrm{reorg}}$ | $\delta G^{\mathrm{solv}}$ | $\delta G$ |
| cone | -65.84 | -1.77 | 22.67 | -44.94 | -40.27 | -1.53 | 15.39 | -26.41 |
| | (-65.11) | (-1.42) | (22.67) | (-43.85) | (-35.45) | (-1.21) | (15.39) | (-21.27) |
| partial-cone | -69.47 | -2.31 | 30.53 | -41.25 | -54.11 | -2.22 | 25.02 | -31.31 |
| | (-61.89) | (-1.35) | (30.53) | (-32.71) | (-42.33) | (-1.23) | (25.02) | (-18.55) |

as the environments in the vicinity of calixarene and the alkali metal complexes, where the neat chloroform was assumed for the dielectric constant ($\varepsilon = 4.90$).

## 4.3 Results and Discussion

### 4.3.1 The cation selectivity

The computed free energy and its components are listed in Table 4.2 and Table 4.3, respectively. Using RISM-SCF-SEDD, $\delta G$ of the sodium-cation capture in cone conformer is -44.94 kcal/mol, which is lower than that by the partial-cone (-41.25 kcal/mol). Hence $\Delta\delta G$ becomes negative (-3.69 kcal/mol), indicating that the sodium cation prefers to associate with the cone conformer. The situation is of contrast in the potassium cation case. $\Delta\delta G$ is positive (4.90 kcal/mol), which means that the partial-cone captures the cation. These computed selectivity agrees with the result from $^1$H NMR experiment[2]. Note that the both of $\Delta\delta G$'s calculated with RISM-SCF-SEDD HF method are negative (-11.14 and -2.72 kcal/mol). That is, HF method failed to reproduce the selectivity in the potassium case. These facts suggest that a sophisticated method such as MP2 coupled with the influence of electrolyte solution at molecular level is required to properly describe the non-covalent, weak host-guest interaction.

Table 4.4: The decomposition of $\Delta\delta G$ obtained by MP2 and HF (in parentheses) methods.

| | | $Na^+$ | | $K^+$ | |
|---|---|---|---|---|---|
| $\Delta\delta G^{\rm gas}$ | $\Delta\delta E^{\rm dist}$ | 12.43 | ( 7.61) | 12.55 | (10.35) |
| | $\Delta\delta E^{\rm int}$ | -9.67 | (-11.69) | 1.78 | (-2.97) |
| | $\Delta\delta E^{\rm thrm}$ | 0.87 | ( 0.87) | -0.49 | (-0.49) |
| | subtotal ($\Delta\delta G^{\rm gas}$) | 3.63 | ( -3.22) | 13.84 | ( 6.88) |
| $\Delta\delta E^{\rm reorg}$ | | 0.54 | (-0.07) | 0.69 | ( 0.02) |
| $\Delta\delta G^{\rm solv}$ | | -7.86 | (-7.86) | -9.63 | (-9.63) |
| total ($\Delta\delta G$) | | -3.69 | (-11.15) | 4.90 | (-2.73) |

The values are given in kcal/mol.

Based on Eq. (4.3), $\Delta\delta G$ provided by RISM-SCF-SEDD is decomposed into three components (Table 4.4). Although both of $\Delta\delta G^{\rm gas}$'s by MP2 are positive, value of the potassium cation (13.8 kcal/mol) is considerably greater than that of the sodium (3.6 kcal/mol). Both of the solvation free energies ($\Delta\delta G^{\rm solv}$) are negative (-7.9 kcal/mol and -9.6 kcal/mol), and the contributions from the coupling term ($\Delta\delta E^{\rm reorg}$) are very small. It is therefore concluded that the balance between $\Delta\delta G^{\rm gas}$ and $\Delta\delta G^{\rm solv}$ basically determined the sign of $\Delta\delta G$, namely the cation selectivity. Because $\Delta\delta G^{\rm gas}$ is great enough in the potassium cation, the relative stability of the isomers in the gas phase remains unchanged in solution phase.

$\delta G^{\rm gas}$ is further decomposed into three components. One is distortion energy required to alter the geometry of calixarene from the isolated-state to the complex ($\delta E^{\rm dist}$). The interaction energy between the cation and the distorted calixarene ($\delta E^{\rm int}$) is the the second one. The third is the thermodynamic correction in gas phase obtained by standard frequency analysis ($\delta E^{\rm thrm}$).

$$\delta G^{\rm gas} = \delta E^{\rm dist} + \delta E^{\rm int} + \delta E^{\rm thrm}. \tag{4.5}$$

The difference between cone and partial-cone is similarly given by,

$$\Delta\delta G^{\rm gas} = \Delta\delta E^{\rm dist} + \Delta\delta E^{\rm int} + \Delta\delta E^{\rm thrm}. \tag{4.6}$$

$\Delta\delta E^{\rm dist}$ is attributed to the geometrical change in the calixarene, the contributions are essentially very similar (12.4 kcal/mol and 12.6 kcal/mol). The main difference between the cations arises from the interaction energy ($\Delta\delta E^{\rm int}$); -9.7 kcal/mol for sodium and 1.8 kcal/mol for

Table 4.5: The energy components of $\delta G^{\mathrm{gas}}$ by MP2 and HF (in parentheses).

| | $Na^+$ | | | | $K^+$ | | | |
|---|---|---|---|---|---|---|---|---|
| | cone | | partial-cone | | cone | | partial-cone | |
| $\delta E^{\mathrm{dist}}$ | 12.97 | ( 10.02) | 0.54 | ( 2.41) | 15.37 | ( 12.67) | 2.82 | ( 2.32) |
| $\delta E^{\mathrm{int}}$ | -89.92 | (-86.24) | -80.25 | (-74.54) | -65.76 | (-58.24) | -67.54 | (-55.26) |
| $\delta E^{\mathrm{thrm}}$ | 11.11 | ( 11.11) | 10.24 | ( 10.24) | 10.12 | ( 10.12) | 10.61 | ( 10.61) |
| total ($\delta G^{\mathrm{gas}}$) | -65.84 | (-65.11) | -69.47 | (-61.89) | -40.27 | (-35.45) | -54.11 | (-42.33) |

All the values are given in kcal/mol.

potassium, respectively. $\Delta\delta E^{\mathrm{int}}$'s computed by HF method (-11.7 kcal/mol for sodium and -3.0 kcal/mol for potassium) are smaller, which is the main source of the qualitatively incorrect results. Table 4.5 lists the change of the energy components on the complex formation. Since $\delta E^{\mathrm{dist}}$ is evidently greater in the cone conformer, the geometry of calixarene is changed on the formation. In contrast, small value of $\delta E^{\mathrm{dist}}$ indicates that the partial-cone does not significantly alter its geometry on the capture. All $\delta E^{\mathrm{int}}$ show great negative values. In any cases, the sodium cation is more strongly captured than potassium. It is pointed out that $\delta E^{\mathrm{int}}$ calculated by MP2 method is greater than those by HF method and the difference is remarkable in the partial-cone complex. As described hereinbelow the cation is closer to the benzene ring in the partial-cone complex. The difference between MP2 and HF probably arises from the dispersion interaction that is believed to be adequately treated only in the MP2 method.

### 4.3.2 Optimized Structures

The difference in the interaction energy between the two cations may be understood based on the structures. Figure 4.3 shows the optimized structure of the cone conformer and its complexes. It is interesting to note that the cone without cation (empty cone) possesses $C_{2V}$ symmetry and $C_{4V}$ structure is the saddle point of the $C_{2V}$–$C_{2V}$ interconversion[22]. Two of the four anisole units lean toward inside of the macrocycle, and the distances between facing oxygen atoms are 3.4 Å (O1$\cdots$O3) and 5.5 Å (O2$\cdots$O4). On capturing sodium ion, both of these distances between oxygen atoms become 4.5 Å and the complex deforms to possess $C_{4V}$ symmetry. The cation is located at the equidistance position from the four oxygen atoms and

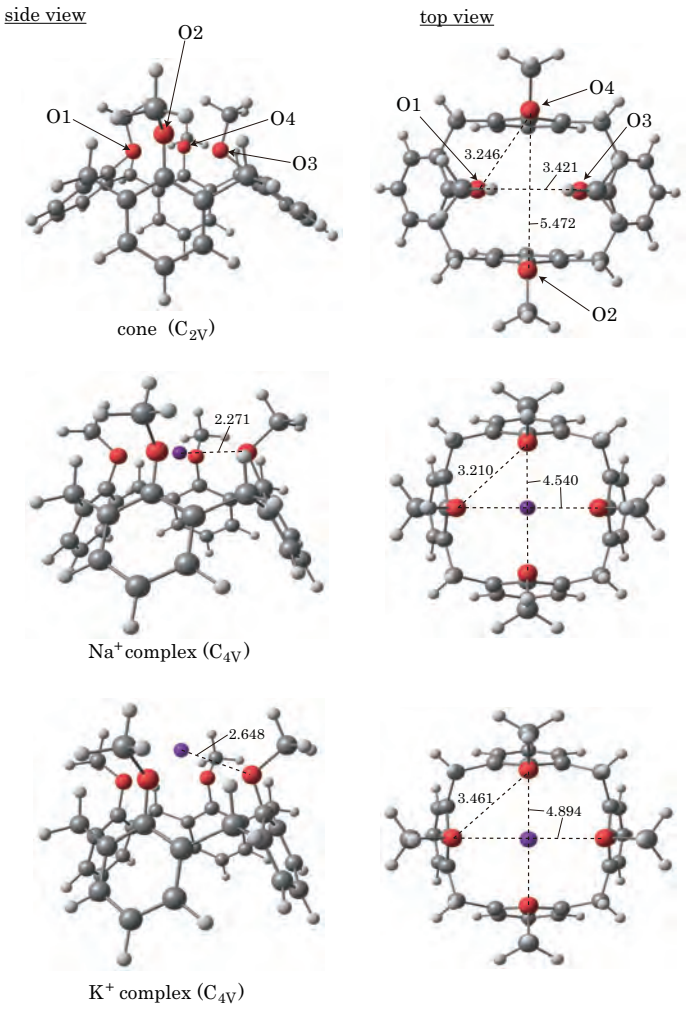


Figure 4.3: Optimized structures in the gas phase of cone conformer and its alkali-metal complexes. The lengths and angles are given by Å and degree units, respectively.

lies in the same plane.

The potassium cation is also equally distant from the oxygen atoms, but it is deviated about 1.0 Å above the plane. O$\cdots$K distances (2.6 Å) is longer by 0.4 Å than O$\cdots$Na. Indeed, the size of macrocycle is expanded and the distance between oxygen atoms is increased to 4.9Å. As shown in the figure, the macrocycle is stretched in the potassium-cation capture while it shrinks in the sodium case compared to the empty cone. The difference between ions is explained from the view point of their radius, which may be a reason that the electrostatic interaction becomes weaker in O$\cdots$K. Although the aforementioned $\delta E^{\mathrm{dist}}$ coincidentally look similar both in the cations, the physical background of the deformation is significantly different. By starting from

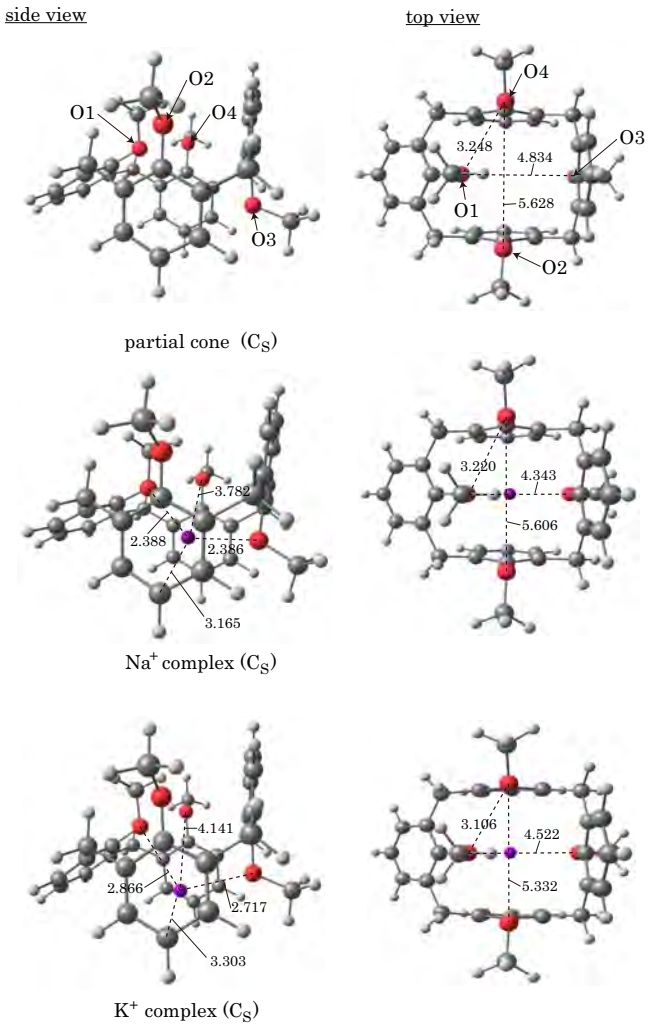


Figure 4.4: Optimized structures in the gas phase of partial-cone conformer and its alkali-metal complexes. The units are the same in Figure 4.3.

several different structures, other complex with different geometry was searched, but all the computations reached the same geometry. Hence, the position of sodium and that of potassium are respectively unique in the complexes.

The optimized structures of the partial-cone and its complex are shown in Figure 4.4. Except for the inverted anisole unit (including O3), the geometry is basically similar to the cone. For instance, O2$\cdots$O4 distance (5.6Å) in the empty structure is very close to 5.5Å in the cone. Unlike the cone complex, however, the structure remains unchanged on the capture of cation. In particular, the geometry of the sodium complex is similar to the empty partial-cone, though

the anisole-units including O1 and O3 slightly change their position to capture the cation. The distances of O1$\cdots$Na and O3$\cdots$Na are nearly equal ($\simeq$2.4Å) and slightly longer than those in the cone complex. The distance between O2 (or O4) and the cation (4.1Å) is longer by 0.3Å and the cation is located at closer to the benzene rings. It is interesting that the potassium cation is 'deeply' captured compared to the sodium.

### 4.3.3 Distributions of cations near the cone and partial-cone conformers

RISM-SCF-SEDD method provides information on the solvation structure in terms of RDF, which is a function of the distance ($r$) between atom $\alpha$ in calixarene (or the complex) and $\gamma$ in solvent of electrolyte solution. Solvation structures around cone, partial-cone and their complexes are characterized by the peak positions and heights of the corresponding RDFs.

**Sodium cation near methoxy group**

Figure 4.5 displays the RDFs of sodium cation in the vicinity of the calixarene and its complex in $NaClO_4$ solution. Black lines are RDFs for the empty conformers, and blue ones for their sodium complexes. Let us start with the distribution of cation near oxygen atom (upper panel). Due to the symmetry, O1 and O2 are the unique site in the cone conformer, these being respectively equivalent to O3 and O4 (see Figure 4.3). For the empty-cone, a distinct peak is found at 2.34 Å in O1–Na (black solid lines). Since the position of the peak agrees well with the distances in the optimized structure of cone-Na complex (2.27Å) shown in Figure 4.3, it may be assigned to the captured sodium cation. In fact, the peak virtually disappears when the cone-Na complex is immersed into solvent (blue solid lines), especially in the range of $r$ ¡ 5Å. The electrostatic field generated by the captured cation prevents other sodium cation from approaching to the calixarene complex. It should be noted that the peak width is considerably broad, indicating that the sodium cation fluctuates wildly. This may contrast markedly with the impression from the optimized structure in the gas phase, where the cation is tightly bounded at a specific position.

The RDF for the partial-cone is exhibited with dashed lines in the figure. O3 becomes different from O1 by the rotation of anisole unit (see Figure 4.4), and is plotted with dashed-

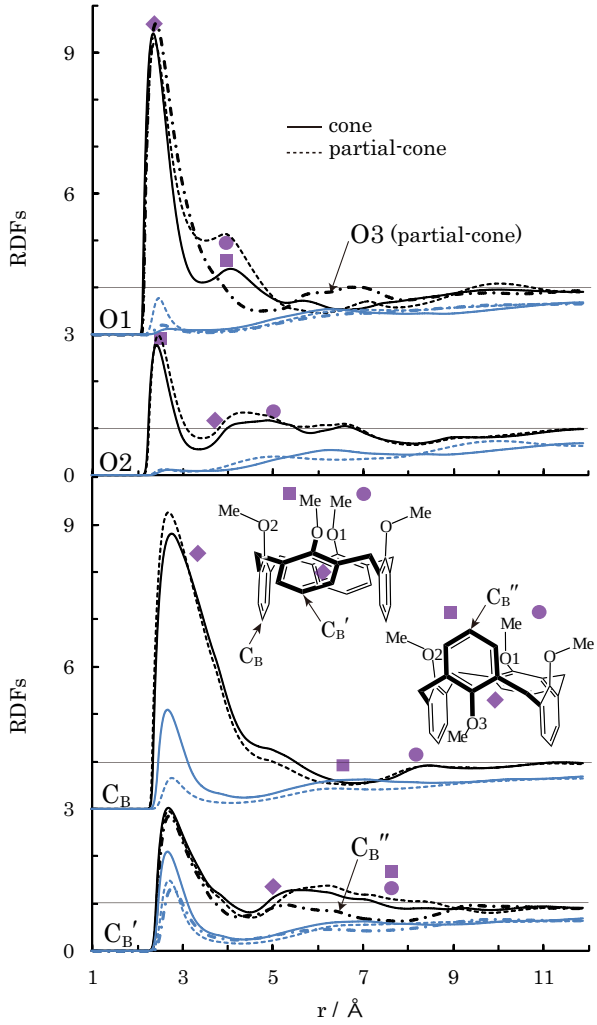


Figure 4.5: RDFs of sodium cation near O1, O3 and O2 (upper panel) atoms, and near $C_B$, $C'_B$ and $C''_B$ (lower panel). RDFs for the empty cone and partial-cone conformers are shown with black lines, while blue lines correspond to those for the sodium complexes.

dotted lines. For the empty-partial-cone, the first peak of O1, O2 and O3 RDFs locates 2.37Å, 2.45Å and 2.40Å, respectively (black dashed and dashed-dotted lines). The heights of RDFs for the cation-complex (blue lines) are considerably lower than the corresponding black ones, indicating that the partial-cone also captures only one sodium cation, namely the sodium cation seems to exist inside of the partial-cone conformer.

Based on the geometrical consideration, the position of sodium cation can be understood more clearly. The inset exhibits plausible positions of sodium cation, and the distance between these positions and oxygen atoms are also indicated with the same symbol on the RDFs. As

seen in the figure, a few positions are consistent with the RDFs. The position of diamond is central of the internal space surrounded by the benzene rings, corresponding to the first peak of O1–Na RDF as well as to the second peak (shoulder) of O2–Na RDF. This is very similar to the gas-phase optimized structure.On the other hand, the position of square (and circle) close to the methoxy groups is consistent both with the first peak of O2–Na RDF and the second peak of O1–Na RDF. Actually, this is outside of the calixarene, and the distribution is assigned to cation in the bulk solution. Note that all these distributions are vanished if one cation occupies the internal space (blue lines). While the diamond is located inside of the calixarene and evidently the captured cation, the lowering of RDFs in O2–Na and O3–Na can be interpreted due to the long-range electrostatic field generated by the captured cation as previously described.

**Sodium cation near the benzene ring**

The lower panel of Figure 4.5 shows RDFs of sodium cation near the terminus carbon atoms of benzene ring. As illustrated in the figure, $C_B$ and $C'_B$ are common to the cone and partial-cone, and $C''_B$ is assigned only for the partial-cone due to the symmetry. For the empty conformers, high and broad peak is found at 2.75Å in $C_B$-Na RDF both in cone (black solid line) and in partial-cone (black dashed line). These peaks are much higher than those in $C'_B$-Na RDF (black solid and dashed lines in the bottom), suggesting that sodium cation can be found with high probability in the area sandwiched by the benzene rings (diamond). The distances between $C'_B \cdots C'_B$ (9.89Å in the optimized structure) is much longer than that between $C_B \cdots C_B$ (5.41Å), and $C'_B$'s are relatively exposed to solvent. Although RDFs for the cation complexes (blue solid and dashed lines) are different from those for the empty calixarenes, the change in $C'_B$-Na RDFs is not significant compared to that in $C_B$-Na. Hence the cation exists near the $C'_B$ even after capturing the cation. In other words, $C'_B$-Na RDFs for the empty is mainly attributed to cation in bulk solvent. The situation is partly similar to $C_B$, but great reduction in RDFs is observed on the capture, suggesting that sodium cation should be found in the interior spaces in the both conformers (diamond) though much broad feature of carbon–Na RDFs undoubt-

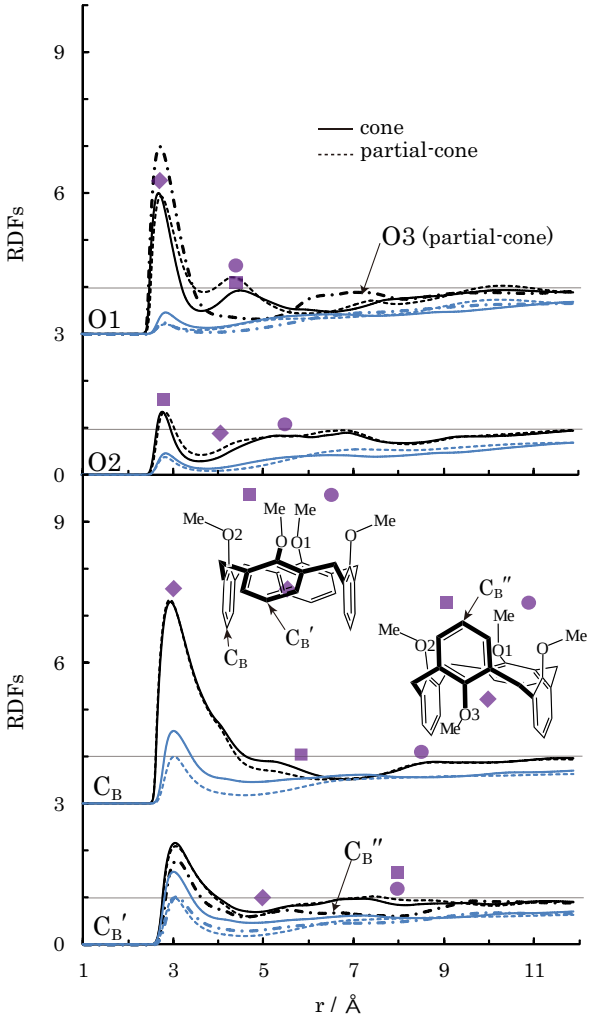


Figure 4.6: RDFs of potassium cation near O1, O3 and O2 (upper panel) atoms, and near $C_B$, $C_B'$ and $C_B''$ (lower panel). RDFs for the empty cone and partial-cone conformers are shown with black lines, while blue lines correspond to those for the potassium complexes.

edly indicates obscure distribution of cation near benzene rings. Based on the free energy difference, sodium cation is stable only with the cone conformer as discussed in the previous section. The present computation suggests that both of cone and partial-cone conformers have capability to capture the cation, but the system finally comes to stable equilibrium with dominant distribution of Na-cone conformer, guided by the free energy difference.

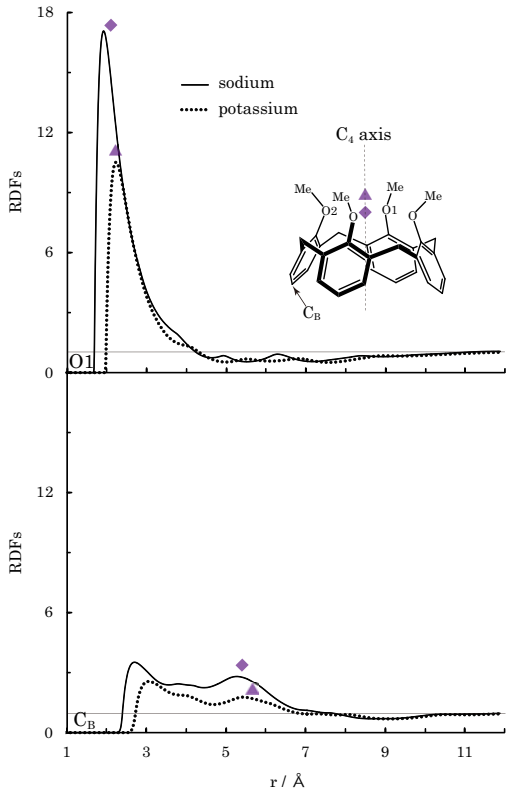


Figure 4.7: RDFs of sodium (solid line) and potassium (dots line) cations near O1 and $C_B$ atoms in cone conformer with the geometry of the cation complex removed the cation.

**The distribution of potassium cation**

Figure 4.6 shows distribution of potassium cation. The upper panel displays RDFs near oxygen atoms (O1, O2 and O3), and lower one near the terminus carbons. Similar to the sodium cation case, black lines corresponding to the empty calixarenes of cone and partial-cone indicate the certain probability to find potassium cation inside of the calixarenes. The dramatic lowering of the peaks in the complexes (blue lines) means that both conformers can not capture more than two potassium cations.

In the empty cone (black solid lines), the position of peak of O1–K (2.68Å) is close to the corresponding distance in the cone-potassium complex (2.65Å, Figure 4.4). Similar to the sodium case, the plausible positions of cation are estimated, namely square (bulk), circle (bulk) and diamond (captured) shown in the figure. The position of diamond indicates the captured potassium cation.

A similar discussion is applicable to the RDFs of empty partial-cone (black dashed and dashed-dotted lines). The positions of the first peak are 2.73Å (O1) and 2.72Å (O3), respectively corresponding to the O1–K distance (2.87Å) and O3–K distance (2.72Å) in the optimized structure of partial-cone-potassium complex (Figure 4.4). Different from the cone conformer, the partial-cone structure is not drastically changed on the capture. Because the potassium cation is captured in the sandwiched area of partial-cone, O3 atom plays an important role. Presumably, this is the reason that the first peak of O3–K RDF is the highest among RDFs shown in this panel.

RDFs near $C_B$, $C'_B$ and $C''_B$ in both conformers are shown in lower panel of the figure. Both of the solid and dash lines for $C_B$–K RDF are quite similar each other, and their peaks found around 2.93Å suggest that the potassium cation distributes in the sandwiched area of both conformers (diamond). Similar to the sodium cation case, it is the free energy difference that governs the capture of the potassium cation. On the other hand, different from the sodium, peak heights are generally low, namely the probability to find the potassium cation is smaller than that in the sodium case. As mentioned in previous section, this is probably caused by the difference in the cation radius. The size of macrocycle should be expanded to capture the bigger potassium cation though the structure was fixed at the empty one.

On the capture of cation, the symmetry of calixarene is changed from $C_{2v}$ to $C_{4v}$. Figure 4.7 shows RDFs around the empty-cones, whose structures were prepared by removing cation from the optimized structures shown in Figure 4.3. Since all the oxygen atoms are equivalent, unique high peak is found at 1.92Å in the sodium system and 2.22Å in the potassium one. These are very close to the gas-phase optimized structures. Compared to the RDFs in the low symmetry, the width of the first peak remains unchanged while the second peak is almost vanished. Hence the peak evidently corresponds to the captured cation. The lower-panel displays RDF between cation and terminus carbon ($C_B$). The peak height is similar to that of $C'_B$ in the $C_{2v}$ complex, suggesting that no cation exists in the internal space surrounded by benzene rings. Since the position of the second peak is consistent with the captured cation, the first peak is mainly attributed to the cation in bulk electrolyte.

## 4.4 Conclusion

The selective cation capture by 25,26,27,28-tetramethoxycalix[4]arene (calixarene) was investigated using a hybrid method of *ab initio* MO calculation and statistical mechanics for molecular-level electrolyte solution, called RISM-SCF-SEDD method. The cation selectivity was successfully reproduced based on the free energy estimations with a higher level MO calculation (MP2). This means that a sophisticated MO calculation coupled with the information of molecular level solvation is necessary to accurately treat the selectivity.

The decomposition of the obtained free energies demonstrates that an interplay between free energy change in gas phase ($\Delta\delta G^{\mathrm{gas}}$) and the difference in solvation free energy ($\Delta\delta G^{\mathrm{solv}}$) is crucial to determine the selectivity. Because $\Delta\delta G^{\mathrm{solv}}$'s are similar between two cations, $\Delta\delta G^{\mathrm{gas}}$ is essential to the process. Further decomposition of $\Delta\delta G^{\mathrm{gas}}$ reveals that $\Delta\delta E^{\mathrm{int}}$ is a key quantity for the selectivity.

Solvation structure is discussed at molecular level based on the computed RDFs. The analysis shows that the cations are captured with similar structures to the optimized ones in the gas phase. It is important, however, that the peaks are considerably broad, indicating that the cations fluctuate wildly and never fix at specific position. The present computation suggests that the cation can be captured both by cone and partial-cone conformers at first, but the system finally comes to stable equilibrium with the corresponding dominant distribution of conformer.

# Bibliography


[1] Ikeda, A.; Shinkai, S. *Chem. Rev.*, **1997**, *97*, 1713.

[2] Iwamoto, K.; Ikeda, A.; Araki, K.; Harada, T.; Shinkai, S. *Tetrahedron*, **1993**, *49*, 9937.

[3] Harada, T.; Rudziński, M. J., Shinkai, S. *J. Chem. Soc., Perkin. Trans. 2*, **1992**, *12*, 2109.

[4] (a) Bernardino, J. R.; Cabral, C. J. B. *J. Phys. Chem. A*, **1999**, *103*, 9080. (b) Hay, P. B.; Nicholas, B. J.; Feller, D. *J. Am. Chem. Soc.*, **2000**, *122*, 10083. (c) Bernardino, J. R.; Cabral, C. J. B. *Supramolecular Chemistry*, **2002**, *14*, 57. (d) Macias, T. A.; Norton, E. J.; Evanseck, D. J. *J. Am. Chem. Soc.*, **2003**, *125* 2351. (e) Wanno, B.; Sang-aroon, W.; Tuntulani, T.; Polpoka, B.; Ruangpornvisuti, V. *J. Mol. Struct.(THEOCHEM)*, **2003**, *629*, 137. (f) Alemán, C.; den Otter, K. W.; Tolpekina, V. T.; Briels, J. W. *J. Org. Chem.*, **2004**, *69*, 951. (g) Alemán, C.; Casanovas, J. *J. Phys. Chem. A*, **2005**, *109*, 8049.

[5] (a) Varnek, A.; Wipff, G. *J. Mol. Struct.(THEOCHEM)*, **1996**, *363*, 67. (b) Golebiowski, J.; Lamare, V.; Martins-Costa, C. T. M.; Millot, C.; Ruiz-López, F. M. *Chem. Phys.*, **2001**, *272*, 47.

[6] (a) Grootenhuis, J. D. P.; Kollman, A. P.; Groenen, C. L.; Reinhoudt, N. D.; van Hummel, J. G.; Ugozzoli, F.; Andreetti, D. G. *J. Am. Chem. Soc.*, **1990**, *112*, 4165. (b) Lipkowitz, B. K.; Pearl, G. *J. Org. Chem.*, **1993**, *58*, 6729. (c) Lauterbanch, M.; Engler, E.; Muzet, N.; Troxler, L.; Wipff, G. *J. Phys. Chem. B*, **1998**, *102*, 245. (d) McDonald, A. N.; Duffy, M. E.; Jorgensen, L. W. *J. Am. Chem. Soc.*, **1998**, *120*, 5104.

[7] (a) Ten-no, S.; Hirata, F.; Kato, S. *J. Chem. Phys.*, **1994**, *100*, 7443. (b) Sato, H.; Hirata, F.; Kato, S. *J. Chem. Phys.*, **1996**, *105*, 1546.

[8] Yokogawa, D.; Sato,H.; Sakaki, S. *J. Chem. Phys.*, **2007**, *126*, 244504.

[9] Singer, J. S.; Chandler, D. *Mol. Phys.*, **1985**, *55*, 621.

[10] Hayaki, S.; Kido, K.; Yokogawa, D.; Sato, H.; Sakaki, S. *J. Phys. Chem. B*, **2009**, *113*, 8227.

[11] Hayaki, S.; Kido, K.; Sato, H.; Sakaki, S. *J. Phys. Chem. Chem. Phys.*, **2010**, *12*, 1822.

[12] Kido, K.; Sato, H.; Sakaki, S. *J. Phys. Chem. B*, **2009**, *113*, 10509.

[13] Beer, D. P.; Gale, A. P. *Angew. Chem. Int. Ed.*, **2001**, *40*, 486.

[14] Chandler, D.; Andersen, H. C. *J. Chem. Phys.*, **1972**, *57*, 1930.

[15] (a) Hirata, F.; Rossky, P. *J. Chem. Phys. Lett.*, **1981**, *83*, 329. (b) Hirata, F.; Rossky, P. J.; Pettitt, B. M. *J. Chem. Phys.*, **1983**, *78*, 4133. (c) Kinoshita, M.; Hirata, F. *J. Chem. Phys.*, **1997**, *106*, 5202.

[16] (a) Hansen, J. P. and McDonald, I. R.; *Theory of Simple Liquids*, Academic Press, (1990). (b) F. Hirata Ed., *Molecular Theory of Solvation (Understanding Chemical Reactivity)*, Kluwer-Springer, (2004). (c) Mennucci, B. and Cammi, R. Eds., *Continuum Solvation Models in Chemical Physics: From Theory to Applications*, John Wiley & Sons Inc (2008).

[17] Dunning, H. T. *J. Chem. Phys.* **1971**, *55* 716.

[18] (a) Perkyns, J.; Pettitt, M. B. *Phys. Chem. Lett.*, **1992**, *190*, 626. (b) Perkyns, J.; Pettitt, M. B. *J. Chem. Phys.*, **1992**, *97*, 7656.

[19] Kovalenko, A.; Hirata, F. *J. Chem. Phys.*, **1999**, *110*, 10095.

[20] Lide, D. R., Ed; *CRC Handbook of Chemistry and Physics, 87th ed.*; CRC press: Boca Raton, FL, 2003; p**6**-*157*.

[21] Lide, D. R., Ed; *CRC Handbook of Chemistry and Physics, 87th ed.*; CRC press: Boca Raton, FL, 2003; p**8**-*540*.

[22] Ikeda, A.; Tsuzuki, H.; Shinkai, S. *J. Chem. Soc., Perkin. Trans. 2*, **1994**, *10*, 2073.

[23] carried out by GAUSSIAN03 package. The universal force field (UFF) are employed as the atomic radii and the solvent type is chloroform. The option “surface = VDW” was also used.

[24] Araki, K.; Shimizu, H.; Shinkai, S. *Chem. Lett.*, **1993**, *2*, 205.

[25] Jorgensen, L. W.; Briggs, M. J.; Contreras, L. M. *J. Phys. Chem.*, **1990**, *94*, 1683.

[26] Aqvist, J. *J. Chem. Phys.*, **1990**, *94*, 8021.

[27] Liu, X.; Zhang, S.; Zhou, G.; Wu, G.; Yuan. X.; Yao, X. *J. Phys. Chem. B*, **2006**, *110*, 12062.

# Part II

# Developlements of a method to evaluate solvation free energy and of a new algorithm of MC-MOZ method

# Chapter 5

# A modified repulsive bridge correction to accurate evaluation of solvation free energy in an integral equation theory for molecular liquids

## 5.1 Introduction

Solvation free energy (SFE) is one of the most fundamental thermodynamic quantities to understand chemical processes in solution. However, the accurate computation is not so straightforward because complex integration is necessary with respect to numerous degrees of freedom composing the system. One of the most popular approaches is based on molecular simulation such as molecular dynamics[1], in which a fully numerical integration is employed for the intermolecular interactions over sufficient number of solvent configurations. In principle, the computed free energy is guaranteed to properly approach the exact value of system[2, 3, 4] with increasing the system size but a high computational cost is usually required. It is also important to note that a sufficient sampling is necessary to achieve statistical convergence (so-called sampling problem). Dielectric continuum model[5], in which solvent is represented with the dielectric continuum, is one of the simplest and widely adopted methods to obtain SFE. The evaluation is, however, essentially based on the electromagnetic theory. An integral equation theory for molecular liquids[6, 7, 8, 9] is an alternative to evaluate SFE, which is based on analytical treatment of the thermodynamic integration. Thanks to this nature, the method does not require a large amount of computational resources and is free from the sampling problem, although the accuracy depends on approximation.

A good example was shown by A. Kovalenko and F. Hirata[10]. In this work, hydration

free energy of a series of rare gases was computed using hyper-netted chain (HNC) closure coupled with Ornstein-Zernike (OZ) type equation. As expected, the computed trend was in contradiction with the experiment because of overestimation of hydrophobicity. In these systems, hydrogen atom of solvent water is practically inside of oxygen atom region and the exclusive volume of water is not adequately described. They proposed a promising procedure to correct this shortcoming and improve the accuracy of SFE called repulsive bridge correction (RBC). This procedure was successfully applied to various solute molecules from simple hydrocarbons[10] to proteins[11]. But the applications are almost limited to aqueous solution systems. As shown below, the correction lose its shine when it is applied to other solvation system since the approximated treatment gets worse as increasing the size of solvent molecule.

In the present study, we generalize RBC to evaluate accurate SFE for a wide range of solvent systems. As the benchmarks, the method is applied to solution systems including water, chloroform and benzene solvents. The organization of this article is as follows: In Section 5.2, after summarizing RBC, the detailed of the present method is described. The computational results to demonstrate the reliability are shown in Section 5.3, followed by conclusions in Section 5.4.

## 5.2 Method

### 5.2.1 Repulsive Bridge Correction

It is our intent to assume the readers' familiarity of integral equation theories of molecular liquids; otherwise, textbooks[12, 13, 14] and reviews[15] are available. Let us begin with closure equation in 3D space.

$$h_s(\mathbf{r}) + 1 = \exp\left[-\beta u_s(\mathbf{r}) + h_s(\mathbf{r}) - c_s(\mathbf{r}) + b_s(\mathbf{r})\right] \tag{5.1}$$

where $h_s(\mathbf{r})$ and $c_s(\mathbf{r})$ respectively represent three-dimensional total and direct correlation functions of solvent site $s$. $u_s(\mathbf{r})$ is solute-solvent interaction potential. $\beta$ is $1/k_{\mathrm{B}}T$, where $k_{\mathrm{B}}$ and $T$ are Boltzmann constant and temperature. $b_s(\mathbf{r})$ is bridge function that is usually unknown.

In RBC formalism, $b_s(\mathbf{r})$ is assumed to be the following equation.

$$\exp\{b_s(\mathbf{r})\} = \prod_{s' \neq s}^{n} \int d\mathbf{\Omega}_{s'} \ \exp\left[-\beta u_{s'}^{\mathrm{R}}(\mathbf{r}_{s'}, \mathbf{\Omega}_{s'})\right] \simeq \prod_{s' \neq s}^{n} w_{ss'} * \exp\left[-\beta u_{s'}^{\mathrm{R}}(\mathbf{r}_{s'})\right] \tag{5.2}$$

where $n$ is the number of interaction sites in the solvent molecule. $\mathbf{r}_s$ and $\mathbf{\Omega}_s$ denote relative position and orientation of the solvent site $s$. By the orientational averaging of the entire core repulsion $u_{s'}^{\mathrm{R}}$, the final equation is given using the solvent-solvent intramolecular correlation function, $w_{ss'}$ . Asterisk in the above equation is a convolution integral. Using Eq. (5.1) and OZ-type equation, SFE ($\Delta\mu$) is obtained with thermodynamic perturbation technique[16],

$$\Delta\mu = \Delta\mu^{\mathrm{HNC}} + k_{\mathrm{B}}T \sum_{s}^{\text{solvent}} \rho_s \int d\mathbf{r}_s \left\{h_s(\mathbf{r}_s) + 1\right\} \left[\exp\{b_s(\mathbf{r}_s)\} - 1\right], \tag{5.3}$$

where

$$\Delta\mu^{\mathrm{HNC}} = k_{\mathrm{B}}T \sum_{s}^{\text{solvent}} \rho_s \int d\mathbf{r}_s \left\{\frac{1}{2}h_s^2(\mathbf{r}_s) - c_s(\mathbf{r}_s) - \frac{1}{2}h_s(\mathbf{r}_s)c_s(\mathbf{r}_s)\right\}. \tag{5.4}$$

This is the original expression of HNC+RBC introduced by Kovalenko et al[10]. It is noted that Eq. (5.3) is reduced to Eq. (5.4) when the bridge function is set to zero.

### 5.2.2 Generalization of RBC

RBC is a promising approach to evaluate accurate SFE for aqueous solution system. As pointed out in the original paper, the approximation is valid for water solvent because the hydrogen-hydrogen correlation is not crucial for the correction effect. To put it another way, the correction may not be suitable for complicated-shape solvent molecule. In fact, the treatment does not work satisfactorily for benzene and chloroform solvents as shown below.

Eq. (5.2) indicates that the correction is described as a superposition of repulsive contribution from solute ($u_{s'}^{\mathrm{R}}$) through *all* the other site ($s'$) in the same solvent molecule, namely the indirect repulsive effect is considered using $\omega_{ss'}$. This may be a good approximation for small molecule such as water. But the reliability of this simple orientational averaging through $\omega_{ss'}$ becomes worse as solvent shape becomes complex. If the size of molecule is sufficiently large, the contributions through other sites should not be uniform and become complex. While the contribution from neighboring site is necessary, those from further sites should be eliminated.

Here, Eq. (5.2) is written as follows by introducing a filtering factor $P_{ss'}$.

$$\exp\{b_s(\mathbf{r})\} = \prod_{s' \neq s}^{n} w_{ss'} * \exp\left[-\beta P_{ss'} u_{s'}^{\mathrm{R}}(\mathbf{r}_{s'})\right] \quad \text{and} \quad P_{ss'} \equiv 1. \tag{5.5}$$

It is convenient and intelligible to represent the factor as square matrix $\mathbf{P}$ with a dimension of $n$. In the above RBC, all the elements of matrix $\mathbf{P}$ are one. If $\mathbf{P} = \mathbf{E}$ (unit matrix), the right hand side of Eq. (5.5) is reduced to be zero, namely corresponding to HNC closure. This means that all the atoms composing the solvent molecule are virtually independent. Hence $\mathbf{P}$ matrix may be regarded as to bridge these two limits, *i.e.*, RBC and HNC. Although *all* the atomic pairs in solvent molecule are explicitly taken into account in RBC, it would be reasonable to suppose that the contribution only from bonding sites participates in RBC because repulsive potential is usually short-ranged function.

$$\mathbf{P} = \mathbf{E} + \mathbf{P}^{\mathrm{CB}} \tag{5.6}$$

where $\mathbf{P}^{\mathrm{CB}}$ is defined as follows.

$$P_{ss'}^{\mathrm{CB}} = \begin{cases} 1 & (\text{if there is chemical bond between } s \text{ and } s') \\ 0 & (\text{others}) \end{cases} \tag{5.7}$$

The matrix element is uniquely defined. For example, $\mathbf{P}^{\mathrm{CB}}$ for aqueous (water) and chloroform (united model) solutions are respectively presented by

$$\mathbf{P}^{\mathrm{CB}}(\mathrm{water}) = \begin{array}{c} \\ \mathrm{O} \\ \mathrm{H} \\ \mathrm{H} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{O} & \mathrm{H} & \mathrm{H} \end{array} \\ \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \end{array}, \tag{5.8}$$

$$\mathbf{P}^{\mathrm{CB}}(\mathrm{chloroform}) = \begin{array}{c} \\ \mathrm{CH} \\ \mathrm{Cl} \\ \mathrm{Cl} \\ \mathrm{Cl} \end{array} \begin{array}{c} \begin{array}{cccc} \mathrm{CH} & \mathrm{Cl} & \mathrm{Cl} & \mathrm{Cl} \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array}. \tag{5.9}$$

The present approach called chemical bond-RBC (CB-RBC) unifies RBC and HNC. As shown below this treatment remarkably improves the evaluation of SFE for a variety of solvent systems, specially organic solvent.

Table 5.1: Potential parameters of solvent water, chloroform and benzene.

| molecule | site | charge / \|e\| | $\sigma$ / Å | $\epsilon$ / kcal mol$^{-1}$ |
|---|---|---|---|---|
| water | O | -0.8340 | 3.151 | 0.1520 |
| (TIP3P[25, 26]-like) | H | 0.4170 | 1.000 | 0.0560 |
| chloroform[27] | CH | 0.4200 | 3.800 | 0.0800 |
| | Cl | -0.1400 | 3.470 | 0.3000 |
| benzene[27] | C | -0.1030 | 3.550 | 0.0700 |
| | H | 0.1030 | 2.420 | 0.0300 |

### 5.2.3 Computational Details

In this study, two types of function are employed for $u_s^{\mathrm{R}}$. One is the repulsive term in Lennard-Jones (LJ) potential,

$$\sum_{\alpha}^{\text{solute}} 4\epsilon_{\alpha s}\left(\frac{\sigma_{\alpha s}}{r_{\alpha s}}\right)^{12} \quad (\mathrm{LJ12}) \tag{5.10}$$

where $r_{\alpha s}$ denotes the distance between site $\alpha$ (solute) and $s$ (solvent). The other is the repulsive part of Weeks-Chandler-Andersen (WCA) separation of LJ potential[17],

$$\sum_{\alpha}^{\text{solute}}\left[4\epsilon_{\alpha s}\left\{\left(\frac{\sigma_{\alpha s}}{r_{\alpha s}}\right)^{12}-\left(\frac{\sigma_{\alpha s}}{r_{\alpha s}}\right)^{6}\right\}+\epsilon_{\alpha s}\right]\Theta\left(2^{\frac{1}{6}}\sigma_{\alpha s}-r_{\alpha s}\right) \quad (\mathrm{WCA}) \tag{5.11}$$

where $\Theta$ means Heaviside step function.

The temperature is 298.15K and the densities of aqueous, chloroform and benzene solutions are set to 1.000g/cm$^3$ (0.033426 molecule/Å$^3$), 1.479g/cm$^3$ (0.007460 molecule/Å$^3$)[18] and 0.877g/cm$^3$ (0.006774 molecule/Å$^3$)[19]. LJ parameters of solvent molecules are listed in Table 5.1. The geometries and LJ parameters of solute molecules are taken from Refs. 30 and 22.

Multi-center molecular OZ (MC-MOZ) method[9] was solved by coupling with HNC closure. In this study, radial 512 (logarithm) and angular 302 (Lebedev) grid points are used to solve MC-MOZ/HNC equation for all calculations. The degree of real spherical harmonics expansions, $l$, is set to 14.

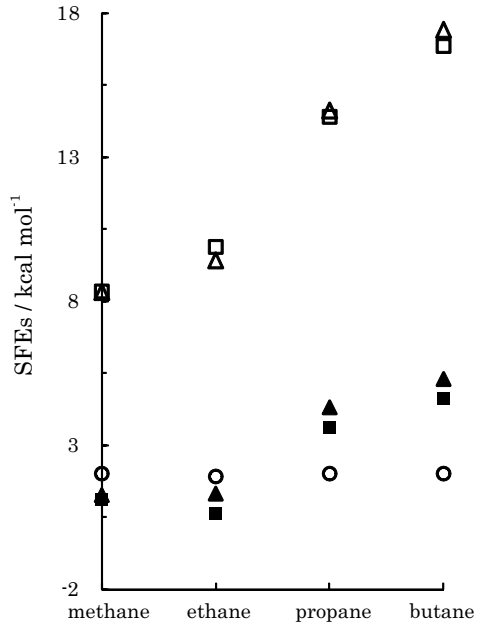


Figure 5.1: Hydration free energy of methane, ethane, propane and butane computed by MC-MOZ (open and close triangles) methods, compared with 3D-RISM results (open and close squares)[10]. The open and close symbols represent $\Delta\mu^{\mathrm{HNC}}$ and $\Delta\mu$, respectively. Circles denote the corresponding experimental values.

## 5.3 Results and Discussion

### 5.3.1 Preliminary examination

Similar to 3-dimensional reference interaction site model (3D-RISM) theory[8], MC-MOZ provides 3D solvation structure although the latter is suitable for parallel computation. Since RBC has been applied only to 3D-RISM, SFE is evaluated using MC-MOZ to check the applicability of the correction. Figure 5.1 shows SFEs of a series of hydrocarbon models in SPC/E-like water computed by MC-MOZ (open and close triangles) compared with 3D-RISM (open and close squares) taken from Ref. 10. As displayed in the figure, both of $\Delta\mu^{\mathrm{HNC}}$ (open symbols) and the corrected $\Delta\mu$ (closed symbols) are virtually identical, indicating the original RBC is applicable to MC-MOZ.

### 5.3.2 Aqueous solution

Table 5.2: Computed SFEs in aqueous solution with HNC, the original RBC, the present CB-RBC and MC methods compared with experiments (unit: kcal/mol).

| molecule | HNC | RBC | | CB-RBC | | MC | Exptl. |
|---|---|---|---|---|---|---|---|
| | | LJ12 | WCA | LJ12 | WCA | | |
| Methanol | 1.17 | -4.94 | -3.03 | -4.64 | -2.78 | -4.01 | -5.1 |
| Methylamine | 1.23 | -4.58 | -2.82 | -4.30 | -2.59 | -3.30 | -4.6 |
| Acetonitrile | 4.07 | -3.96 | -1.15 | -3.55 | -0.82 | -3.83 | -3.9 |
| Dimethylether | 7.36 | -0.10 | 2.14 | 0.30 | 2.48 | -1.85 | -1.9 |
| Methanethiol | 6.50 | -0.57 | 1.88 | -0.22 | 2.17 | -0.72 | -1.2 |
| Chloromethane | 6.68 | -0.69 | 2.00 | -0.32 | 2.30 | -0.32 | -0.6 |
| Ethane | 9.84 | 3.26 | 5.19 | 3.64 | 5.52 | 0.83 | 2.0 |
| Acetamide | -3.27 | -11.30 | -8.73 | -10.94 | -8.43 | -10.00 | -9.7 |
| Acetic acid | 0.07 | -8.30 | -5.57 | -7.91 | -5.25 | -7.40 | -6.7 |
| Acetone | 4.99 | -3.69 | -1.01 | -3.25 | -0.64 | -4.45 | -3.8 |
| Methyl acetate | 6.17 | -3.37 | -0.45 | -2.89 | -0.04 | -5.17 | -3.3 |
| Benzene | 12.31 | 3.13 | 5.89 | 3.65 | 6.33 | -1.07 | -0.8 |
| Pyridine | 8.24 | -1.03 | 1.85 | -0.54 | 2.26 | -3.91 | -4.7 |
| Dichloromethane | 8.95 | -0.62 | 3.22 | -0.14 | 3.59 | -0.49 | -1.4 |
| Fluoromethane | 6.30 | 0.35 | 2.21 | 0.67 | 2.48 | -0.87 | -0.2 |
| Trichloromethane | 11.57 | 0.19 | 4.98 | 0.75 | 5.41 | 0.14 | -1.1 |
| Ethene | 8.67 | 2.65 | 4.54 | 2.99 | 4.83 | 0.34 | 1.3 |
| Cyclohexane | 18.84 | 8.54 | 11.45 | 9.14 | 11.99 | 0.22 | 1.2 |
| 2-propanol | 8.99 | 0.23 | 2.86 | 0.70 | 3.26 | -5.08 | -4.8 |
| Dimethylamine | 7.98 | 0.44 | 2.69 | 0.85 | 3.04 | -2.61 | -4.3 |
| Dimethyl sulfide | 11.48 | 2.89 | 5.69 | 3.36 | 6.09 | -0.48 | -1.4 |
| Tart-butyl alcohol | 11.41 | 1.77 | 4.60 | 2.29 | 5.05 | -6.00 | -4.5 |
| Trimethylamine | 12.72 | 4.01 | 6.57 | 4.51 | 7.00 | -1.43 | -3.2 |
| Butane | 16.31 | 7.12 | 9.76 | 7.65 | 10.23 | 0.57 | 2.1 |
| Propane | 13.37 | 5.45 | 7.76 | 5.91 | 8.16 | 0.60 | 2.0 |
| Ethylamine | 6.06 | -1.31 | 0.90 | -0.93 | 1.22 | -5.19 | -4.5 |
| Ethanol | 6.16 | -1.35 | 0.94 | -0.96 | 1.28 | -4.93 | -5.0 |
| Pronene | 11.55 | 4.10 | 6.36 | 4.53 | 6.72 | -0.02 | 1.3 |
| Acetoaldehyde | 5.00 | -2.61 | -0.12 | -2.22 | 0.20 | -3.79 | -3.5 |
| Naphthalene | 19.35 | 7.24 | 10.94 | 7.94 | 11.54 | -1.85 | -2.4 |
| Phenol | 11.11 | 1.05 | 4.21 | 1.59 | 4.67 | -5.01 | -6.6 |
| Tetrahydrofuran | 11.15 | 1.91 | 4.63 | 2.42 | 5.07 | -2.99 | -3.5 |
| RMSD | 11.76 | 3.97 | 6.21 | 4.34 | 6.57 | 1.02 | |

Table 5.2 lists SFE of 32 solute molecules in aqueous solution, hydration free energies. SFEs computed by HNC, the original RBC, the present method (CB-RBC), Monte Carlo (MC)

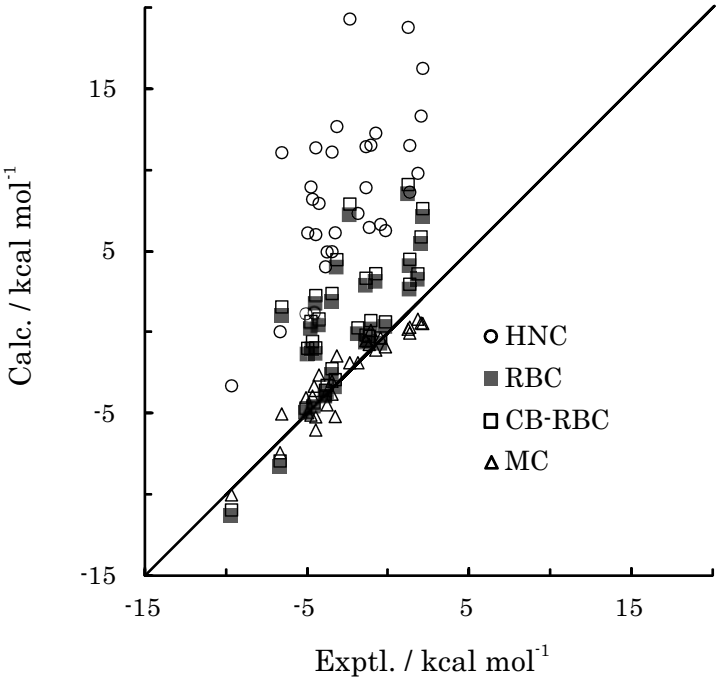


Figure 5.2: Comparison between calculated and experimental SFEs. LJ12 was adopted in the original RBC and the present CB-RBC.

simulation[20, 21, 22] are shown with their corresponding experimental values[23]. The same data is plotted in Figure 5.2. The two axis are taken as the calculated and experimental SFEs; If a symbol in this figure locates in the area upper the line, the SFE is overestimated. When an underestimated SFE is given, the position of the symbol is lower than the line. HNC (circle) results are obviously distributed at upper region. Actually, the root mean square deviations (RMSDs) with respect to the experimental values are 11.76kcal/mol (HNC), 3.97kcal/mol (RBC, LJ12), 4.34kcal/mol (CB-RBC, LJ12) and 1.02kcal/mol (MC). Note that, although the highest accuracy is achieved in the MC simulation, reported data is adjusted using both solvent-accessible surface area (SASA) and some optimized parameters utilizing the experimental SFEs. As mentioned in the previous reports[10], HNC considerably overestimates the observations more than 10kcal/mol, and they are adequately corrected both by RBC and CB-RBC, although these corrected values are still overestimated. It is suggested that a charge polarization of solute molecule is crucial, especially in aqueous media, to accurately evaluate SFEs.

Table 5.3: The SFEs in chloroform solution evaluated by HNC, the present, the original and MC methods. The unit is kcal/mol.

| molecule | HNC | RBC | | CB-RBC | | MC | Exptl. |
|---|---|---|---|---|---|---|---|
| | | LJ12 | WCA | LJ12 | WCA | | |
| Methanol | -0.65 | -7.75 | -5.96 | -4.37 | -3.31 | -3.32 | -3.32 |
| Methylamine | -0.79 | -7.63 | -5.96 | -4.35 | -3.37 | -2.85 | -3.44 |
| Acetonitrile | -2.02 | -10.96 | -8.56 | -6.78 | -5.32 | -4.85 | -4.49 |
| Acetamide | -4.16 | -13.23 | -10.91 | -9.05 | -7.64 | -6.58 | -6.98 |
| Acetic acid | -2.90 | -12.18 | -9.75 | -7.90 | -6.41 | -4.90 | -4.63 |
| Acetone | -2.45 | -12.14 | -9.79 | -7.64 | -6.21 | -5.58 | -4.96 |
| Methyl acetate | -2.87 | -13.35 | -10.84 | -8.52 | -6.97 | -5.62 | -4.88 |
| Benzene | -1.65 | -11.73 | -9.39 | -7.00 | -5.58 | -4.76 | -4.61 |
| Pyridine | -2.59 | -12.83 | -10.38 | -8.06 | -6.56 | -5.39 | -6.58 |
| Cyclohexane | -1.57 | -12.51 | -10.13 | -7.45 | -6.00 | -5.30 | -4.48 |
| Dimethylamine | -1.04 | -9.52 | -7.56 | -5.52 | -4.34 | -3.77 | -3.77 |
| Trimethylamine | -1.34 | -10.92 | -8.76 | -6.44 | -5.13 | -4.49 | -3.98 |
| Acetoaldehyde | -1.33 | -9.84 | -7.68 | -5.85 | -4.54 | -4.44 | -3.65 |
| Phenol | -3.06 | -13.89 | -11.30 | -8.88 | -7.28 | -5.88 | -7.10 |
| RMSD | 2.81 | 6.65 | 4.40 | 2.34 | 1.07 | 0.66 | |

One might notice that SFE by CB-RBC is slightly higher than RBC. This is simply attributed to the treatment in Eq. (5.3): Because $(h_s + 1) \geq 0$ and $\exp(b_s) - 1 \leq 0$ are satisfied for any solvent site $s$, the correction (second term in this equation) always contributes to SFE negatively. Because of the nature of $\mathbf{P}$ matrix, this correction is partially effective in CB-RBC. Namely, the correction in the original RBC is slightly greater in negative than that in the present CB-RBC. The difference is quite small in practice, indicating that the effect of the core repulsion is dominated by the oxygen atom in water solvent. This is consistent with the discussion on the empirical approach by Roux et. al[24].

The comparison of SFEs computed with LJ12 and WCA is shown in Table 5.2. Both of them provide similar values but the difference could not be negligibly small: the LJ12 RMSD is 4.34 kcal/mol whereas the RMSD by WCA is 6.57 kcal/mol. Since there is no clear criterion to select the core repulsive function, it is difficult to judge between them. Anyway, these numerical results show that LJ12 function provides more accurate SFEs than WCA in the case of aqueous solution.

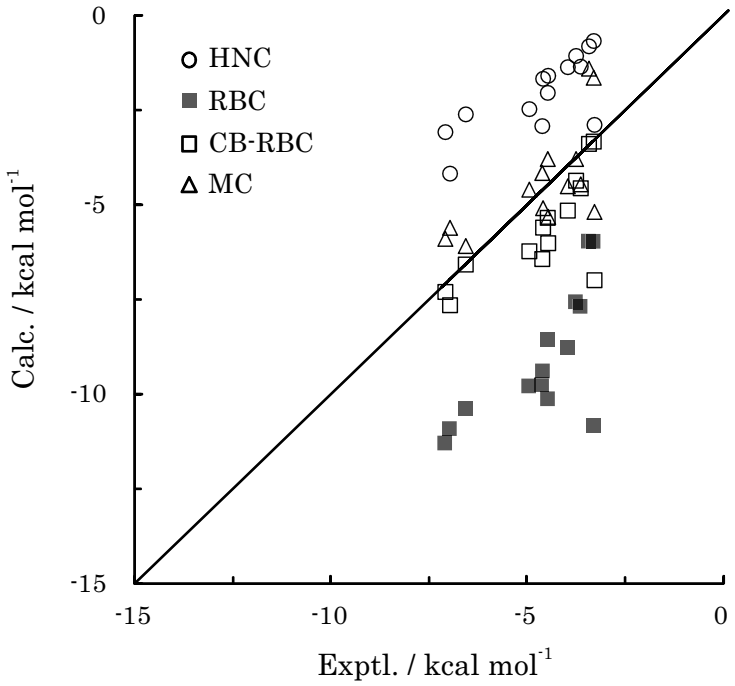


Figure 5.3: Comparison between calculated and experimental SFEs. WCA was adopted in the original RBC and the present CB-RBC.

### 5.3.3 Chloroform and Benzene solutions

In chloroform solution, the validity of the present method is investigated for 14 solute molecules. Similarly, 12 solute molecules are examined in benzene solution. The numerical results in chloroform solution computed by HNC, the original RBC, the present CB-RBC and MC[22] procedures are listed in Table 5.3. The RMSD of HNC results is 2.81kcal/mol, which is remarkably smaller than that in aqueous solution, the accuracy is not poor so much. But all SFEs of HNC (circles) somewhat overestimates the corresponding experimental values[23] as shown in Figure 5.3.

Different from the case of aqueous solution, the original RBC (closed squares) considerably underestimated SFEs (the symbols locate the lower area than the line). The values get worse compared to the HNC ones, (RMSD is 4.40kcal/mol), indicating that the correction dose not work in practice. On the other hand, as expected, the CB-RBC method (open squares) gives very reasonable results (RMSD is 1.07kcal/mol when WCA was adopted to the repulsive func-

Table 5.4: The SFEs in benzene solution evaluated by HNC, the present and the original methods. The unit is kcal/mol.

| molecule | HNC | RBC | | CB-RBC | | Exptl. |
|---|---|---|---|---|---|---|
| | | LJ12 | WCA | LJ12 | WCA | |
| Methanol | 1.35 | -29.44 | -24.33 | -6.65 | -4.27 | -2.58 |
| Methylamine | 1.40 | -28.33 | -23.55 | -6.21 | -4.01 | -2.66 |
| Benzene | 2.26 | -40.03 | -33.74 | -9.12 | -5.99 | -4.55 |
| Pyridine | 1.20 | -41.64 | -35.01 | -10.56 | -7.16 | -5.28 |
| Cyclohexane | 3.36 | -41.69 | -35.33 | -9.05 | -5.87 | -4.05 |
| 2-propanol | 2.13 | -37.68 | -31.58 | -8.69 | -5.67 | -3.48 |
| Dimethylamine | 2.08 | -34.03 | -28.59 | -7.44 | -4.84 | -3.01 |
| Trimethylamine | 2.62 | -37.43 | -31.59 | -8.16 | -5.29 | -2.80 |
| Ethylamine | 1.77 | -33.33 | -27.88 | -7.51 | -4.91 | -2.73 |
| Ethanol | 1.81 | -33.57 | -28.03 | -7.64 | -4.96 | -3.42 |
| Phenol | 1.08 | -44.00 | -37.01 | -11.47 | -7.85 | -7.12 |
| Water | 0.63 | -21.58 | -17.28 | -4.97 | -3.06 | -1.71 |
| RMSD | 5.64 | 32.03 | 26.25 | 4.55 | 1.77 | |

tion). Although the corrected value slightly underestimates the experiments, the accuracy of the present approach is comparable to that of MC simulation displayed with open triangles (RMSD is 0.66kcal/mol). As the same in the aqueous system, the SFE by MC simulation is modified result of a free-energy perturbation with SASA and 3-adjusted parameters. Here, it should be emphasized that the present CB-RBC procedure achieves high accuracy without such a treatment, *i.e.*, simply based on thermodynamic integration method of SFE. As listed in the table, RMSD is slightly worse (2.34 kcal/mol) when LJ12 was adopted. This is opposite trend to the aqueous solution.

Finally, computed SFEs in benzene solution by HNC, the original RBC and the present CB-RBC are summarized in Table 5.4. Unfortunately, simulation results could not be found. As seen in Figure 5.4, the original RBC has a large margin of error similar to the case of chloroform solution. RMSDs of RBC and CB-RBC are 26.25kcal/mol (WCA) and 1.77 kcal/mol (WCA), respectively. The significant underestimation by RBC may be attributed to the large size of benzene. As increasing the number of sites in solvent, the correction becomes negatively greater. In contrast, the present CB-RBC exhibits an excellent agreement with the

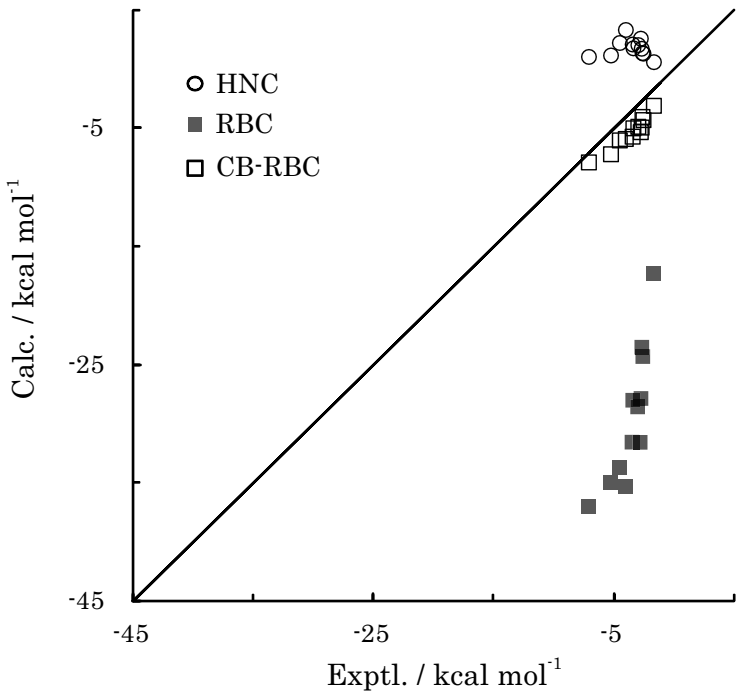


Figure 5.4: Comparison between calculated and experimental SFEs. WCA was adopted in the original RBC and the present CB-RBC.

experimental SFEs[23].

## 5.4 Conclusions

In the present work, a generalization of RBC is reported to accurately evaluate the solvation free energy based on the integral equation theory for molecular liquids. Utilizing the bonding information in the solvent molecule, an efficient way is proposed to correctly take the repulsive bridge correction into account. SFEs in chloroform and benzene solutions are underestimated by the original RBC.

Aqueous solution (32 solute molecules), chloroform solutions (14 solute molecules) and benzene solutions (12 solute molecules) were examined. In the aqueous solution, SFEs are accurately computed using the original RBC and the present CB-RBC within 4.5kcal/mol of RMSD from the experiments. In order to improve the agreement, the effect of charge polarization of solute molecule is necessary, which is not dealt with in the present study.

In chloroform and benzene solutions, however, the original RBC severely underestimates the SFEs. On the other hand, the present CB-RBC achieves a excellent agreement with the experimental values (the RMSDs are less than 2kcal/mol). These results clearly indicate the validity of the present procedure for various solvents.

# Bibliography


[1] P. Kollman, *Chem. Rev.*, **93**, 2395 (1993).

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

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

[4] N. Matubayasi, M. Nakahara, *J. Chem. Phys.*, **119**, 9686 (2003).

[5] J. Tomasi, M. Persico, *Chem. Rev.*, **94**, 2027 (1994).

[6] D. Chandler, H. C. Andersen, *J. Chem. Phys.*, **57**, 1930 (1972).

[7] F. Hirata, P. J. Rossky, *Chem. Phys. Lett.*, **83**, 329 (1981).

[8] A. Kovalenko, F. Hirata, *Chem. Phys. Lett.*, **290**, 237 (1998).

[9] D. Yokogawa, H. Sato, T. Imai, S. Sakaki, *J. Chem. Phys.*, **130**, 064111 (2009).

[10] A. Kovalenko, F. Hirata, *J. Chem. Phys.*, **113**, 2793 (2000).

[11] T. Imai, Y. Harano, M. Kinoshita, A. Kovalenko, F. Hirata, *J. Chem. Phys.*, **125**, 024911 (2006).

[12] J. P. Hansen and I. R. McDonald Theory of Simple Liquids (1990) Academic Press.

[13] F. Hirata (Ed.), Molecular Theory of Solvation (Understanding Chemical Reactivity), Kluwer-Springer, 2004.

[14] B. Mennucci and R. Cammi (Eds.) Continuum Solvation Models in Chemical Physics: From Theory to Applications, John Wiley & Sons Inc. 2008.

[15] J. J. Howard, B. M. Pettitt, *J. Stat. Phys.*, **145**, 441 (2011).

[16] S. J. Singer, D. Chandler, *Mol. Phys.*, **55**, 621 (1985).

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

[18] D. R. Lide (Ed.), CRC Handbook of Chemistry and Physics, 87th ed. CRC press: Boca Raton, FL, (2003) p**3**-540.

[19] D. R. Lide (Ed.), CRC Handbook of Chemistry and Physics, 87th ed. CRC press: Boca Raton, FL, (2003) p**3**-36.

[20] H. A. Carlson, T. B. Nguyen, M. Orozco, W. L. Jorgensen, *J. Compt. Chem.*, **14**, 1240 (1993).

[21] W. L. Jorgensen, J. Tirado-Rives, *Drug Discovery Des.*, **3**, 123 (1995).

[22] N. A. McDonald, H. A. Carlson, W. L. Jorgensen, *J. Phys. Org. Chem.*, **10**, 563 (1997).

[23] G. D. Hawkins, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B*, **102**, 3257 (1998) (supporting information).

[24] Q. Du, D. Beglov, B. Roux, *J. Phys. Chem. B*, **104**, 796 (2000).

[25] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.*, **79**, 926 (1983).

[26] W. L. Jorgensen, J. D. Madura, *Mol. Phys.*, **99**, 1381 (1985).

[27] W. L. Jorgensen, J. M. Briggs, M. L. Contreras, *J. Phys. Chem.*, **94**, 1683 (1990).

# Chapter 6

# The development of a revised version of multi-center molecular Ornstein-Zernike equation

## 6.1 Introduction

Integral equation theory for molecular liquids [2, 3] is a powerful tool to obtain liquid and solvation structures. In 3-dimensional (3D) formalism, solvent distribution in the vicinity of the intended molecule is provided by Ornstein-Zernike (OZ)-type equations. 3-dimensional reference interaction site model (3D-RISM) [4] is one of the representatives of the approach. Accuracy and validity of the method are demonstrated in a variety of applications including simple atomic ion and biological systems like protein in solution phase [5, 6, 7, 8]. 3D-RISM is capable of adequately computing the solvation structure not only around a protein but also inside.

Recently, an alternative approach called multi-center molecular OZ (MC-MOZ) method is proposed [9, 10, 11]. In MC-MOZ, a solvation structure is expressed as the sum of two terms, namely reference and residue terms; After obtaining the former term by solving the standard RISM equation for radial distribution, the latter residue term representing the angular distribution is computed. Because the residue term is described as a set of distribution functions, which are localized in the vicinity of each atom (site) in the solute molecule such as protein, the method is suitable for parallel computing of 3D solvation structure. However, since the reference-term computation was not able to be carried out site-independently, the total parallel efficiency became lower as the number of processors increases in practice.

In this letter, several other functions were examined as the reference term to investigate the

numerical robustness of the method. Although it is possible to arbitrarily set the reference term, the computed 3D distribution function should be uniquely obtained in principle. More specifically, the residue term is expected to complement the difference caused by the reference. Another aim of the present study is to achieve the full-parallelization of the computation. Solvation structures of water (TIP3P-like), benzene in aqueous solution and single-walled carbon nanotube (SW-CNT, 1024 atoms) in chloroform solution are calculated using the new reference, and the obtained 3D solvent distributions are compared with the original MC-MOZ procedure.

## 6.2 Method

The detail of MC-MOZ has been described in the previous paper [9]. It is our intent to assume the readers' familiarity of MC-MOZ as well as the statistical mechanics of molecular liquids. If readers are not familiar, please refer to textbook [12, 13, 14], some reviews and papers [2, 3, 4, 15]. Here, only the reference terms are focused on because the corresponding residue terms are computed through the same procedure in the previous paper [9]. On the computations, two types of functions are prepared as the reference term.

$$h_s^{\mathrm{ref}}(\mathbf{r}) = \sum_{\alpha}^{\mathrm{solute}} \sum_{s'}^{\mathrm{solvent}} c_{\alpha s'}^{\mathrm{method}} * (w_{s's}^{\mathrm{V}} + \rho_s h_{s's}^{\mathrm{V}})(r_{\alpha s}) \tag{6.1}$$

$$c_s^{\mathrm{ref}}(\mathbf{r}) = \sum_{\alpha}^{\mathrm{solute}} c_{\alpha s}^{\mathrm{method}}(r_{\alpha s}) \tag{6.2}$$

where $h_s^{\mathrm{ref}}(\mathbf{r})$ and $c_s^{\mathrm{ref}}(\mathbf{r})$ are the reference terms of total and direct correlation functions for solvent site $s$, respectively. $w_{ss'}^{\mathrm{V}}$ and $h_{ss'}^{\mathrm{V}}$ denote the solvent-solvent intramolecular and total correlation functions, respectively. $\rho_s$ is the number density of solvent and the asterisk represents a convolution integral.

In original MC-MOZ procedure, standard 1-dimensional (1D) RISM is employed to prepare the reference term. $c_{\alpha s}^{\text{1D-RISM}}(r_{\alpha s})$ is obtained by solving the corresponding solute-solvent 1D-RISM equation [15] coupled with an adequate closure. Then this function is plugged into the

right hand side of Eqs. (6.1) and (6.2) to evaluate the residue term.

$$c_{\alpha s}^{\mathrm{method}}(r_{\alpha s}) = c_{\alpha s}^{\text{1D-RISM}}(r_{\alpha s}). \tag{6.3}$$

As mentioned above, the procedure can not be independently performed for each solute site because $c_{\alpha s}^{\text{1D-RISM}}(r_{\alpha s})$ depend on the information of other solute sites. This is the cause for the low parallel efficiency. In the present study, following four functions are examined for $c_{\alpha s}^{\mathrm{method}}$ to prepare the reference term. The first one is just zero;

$$c_{\alpha s}^{\mathrm{I}}(r_{\alpha s}) = 0. \tag{6.4}$$

$h_s^{\mathrm{ref}}(\mathbf{r})$ is generated using Eq.(6.1), hence both of $c_s^{\mathrm{ref}}(\mathbf{r})$ and $h_s^{\mathrm{ref}}(\mathbf{r})$ are zero value functions. The second one is based on analytical solution of OZ/Percus-Yevick (PY) equations for hard spheres [16, 17, 18].

$$c_{\alpha s}^{\mathrm{II}}(r_{\alpha s}) = \begin{cases} -\lambda_0 + 6\eta\lambda_2 \left(\dfrac{r_{\alpha s}}{\sigma_{\alpha s}}\right) - \dfrac{1}{2}\eta\lambda_1 \left(\dfrac{r_{\alpha s}}{\sigma_{\alpha s}}\right)^3 & (r_{\alpha s} < \sigma_{\alpha s}) \\ 0 & (r_{\alpha s} \geq \sigma_{\alpha s}) \end{cases} \tag{6.5}$$

$$\lambda_0 = 6\eta\lambda_2 - \frac{1}{2}\eta\lambda_1, \quad \lambda_1 = \frac{(1+2\eta)}{(1-\eta)^4}, \quad \lambda_2 = \frac{(1+1/2\eta)}{(1-\eta)^4}$$

where $\sigma_{\alpha s} = (\sigma_\alpha + \sigma_s)/2$ ($\sigma$ is the Lennard-Jones (LJ) radius) and $\eta$ denotes a packing fraction that is set to 0.30 in this study. In a similar manner, $c_s^{\mathrm{ref}}(\mathbf{r})$ and $h_s^{\mathrm{ref}}(\mathbf{r})$ are obtained using Eqs. (6.1) and (6.2). The third one is generated by mean spherical approximation (MSA) with LJ potential

$$c_{\alpha s}^{\mathrm{III}}(r_{\alpha s}) = \begin{cases} 0 & (r_{\alpha s} < \sigma_{\alpha s}) \\ -\dfrac{1}{k_{\mathrm{B}}T} u^{\mathrm{LJ}}(r_{\alpha s}) & (r_{\alpha s} \geq \sigma_{\alpha s}) \end{cases} \tag{6.6}$$

where $k_{\mathrm{B}}$ and $T$ represent Boltzmann constant and temperature. $u^{\mathrm{LJ}}(r_{\alpha s})$ denotes the LJ potential between site $\alpha$ and $s$. Here, it should be emphasized that these new reference terms no longer depends on the information of other solute sites, namely, these computations can be site-independently performed. Furthermore, they can be prepared in non-iterative manner whereas it is necessary to iteratively solve 1D-RISM equation in the original procedure.

Table 6.1: Potential parameters for water(TIP3P-like), chloroform, benzene and SW-SNT.

| molecule | site | charge / \|e\| | $\sigma$ / Å | $\epsilon$ / kcal mol$^{-1}$ |
|---|---|---|---|---|
| water (TIP3P[23, 24]-like) | Ow | -0.8340 | 3.151 | 0.1520 |
| | Hw | 0.4170 | 1.000 | 0.0560 |
| benzene[25] | C | -0.1030 | 3.550 | 0.0700 |
| | H | 0.1030 | 2.420 | 0.0300 |
| SW-CNT | C | 0.0000 | 3.851 | 0.1050 |
| chloroform[25] | CH | 0.4200 | 3.800 | 0.0800 |
| | Cl | -0.1400 | 3.470 | 0.3000 |

The last one is an iterative but independent of other solute sites. A 1D-RISM-like equation is considered.

$$h_{\alpha s}(r_{\alpha s}) = \sum_{\gamma}^{\text{solute}} \sum_{s'}^{\text{solvent}} w_{\alpha\gamma} * c^{\text{IV}}_{\alpha s'} * \left(w^{\text{V}}_{s's} + \rho_s h^{\text{V}}_{s's}\right)(r_{\alpha s}), \tag{6.7}$$

where $h_{\alpha s}(r_{\alpha s})$ and $w_{\alpha\gamma}$ represent solute-solvent total and solute-solute intramolecular correlation functions, respectively. By numerically solving this equation coupled with an adequate closure, $c^{\text{IV}}_{\alpha s}(r_{\alpha s})$ is obtained. Then $h^{\text{ref}}_s(\mathbf{r})$ is calculated using eq.(6.1). Note that the index of $c^{\text{IV}}_{\alpha s'}$ in this equation is *not* $\gamma s'$, corresponding to the standard 1D-RISM equation. Although there is no physical meaning of this description, it enables us to compute $c^{\text{IV}}_{\alpha s'}$ in a site-independent manner. Different from above three methods (I, II and III), $c^{\text{IV}}_{\alpha s}(r_{\alpha s})$ implicitly involves the contribution of solute geometry through the matrix $w_{\alpha\gamma}$ and that of solute-solvent interaction.

Potential parameters used in the present study for water (TIP3P[23, 24]-like), benzene [25] and SW-CNT [26, 27] are summarized in Table 6.1. In SW-CNT (the CNT index = (16,0)), the C-C distance and the CCC angle are 1.418Å and 120.0 degree, respectively. The diameter and the cylindrical length are 12.52Å and 68.00Å, respectively. For all calculations, the solution temperature is set to 298.15K. The number densities of water and chloroform are 1.000g/cm$^3$(= 0.033426 molecule/Å$^3$) and 1.479 g/cm$^3$(= 0.007460 molecule/Å$^3$) [36], respectively.

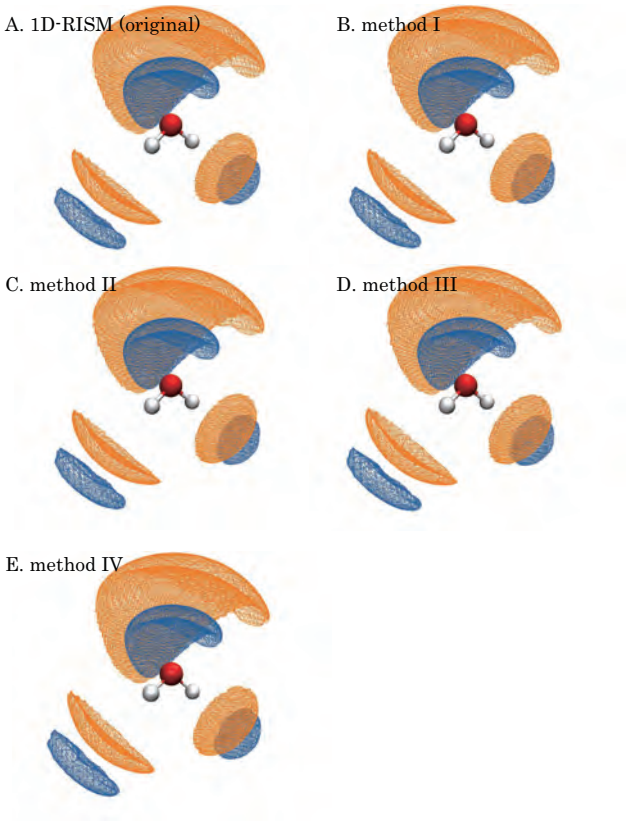


Figure 6.1: 3-dimensional hydration structures near water molecule computed by original (1D-RISM, A), method I (B), II (C), III (D) and IV (E) reference terms. Blue (isovalue: 1.8) and orange (isovalue: 3.0) iso-surfaces represent the distributions of hydrogen and oxygen atoms of solvent water molecules, respectively.

## 6.3 Results and Discussion

### 6.3.1 water and benzene

Three dimensional hydration structures around water (Fig.6.1) and benzene (Fig.6.2) molecules are obtained by MC-MOZ theory coupled with HNC closure. The degree, $l$, is set to 14 for the real spherical harmonics. 512 (logarithm) and 302 (Lebedev) grid points are used as the radial and angular grids, respectively.

The final results, total correlation functions, were obtained as the sum of the reference and residue terms ($h_s^{\text{1D-RISM}}(\mathbf{r}_k)$, $h_s^{\text{I}}(\mathbf{r}_k)$, $h_s^{\text{II}}(\mathbf{r}_k)$, $h_s^{\text{III}}(\mathbf{r}_k)$ and $h_s^{\text{IV}}(\mathbf{r}_k)$) after the computation. Fig.6.1 shows the hydration structures near water molecule computed by all the reference terms: the original (panel A), method I (panel B), II (panel C), III (panel D) and IV (panel E).

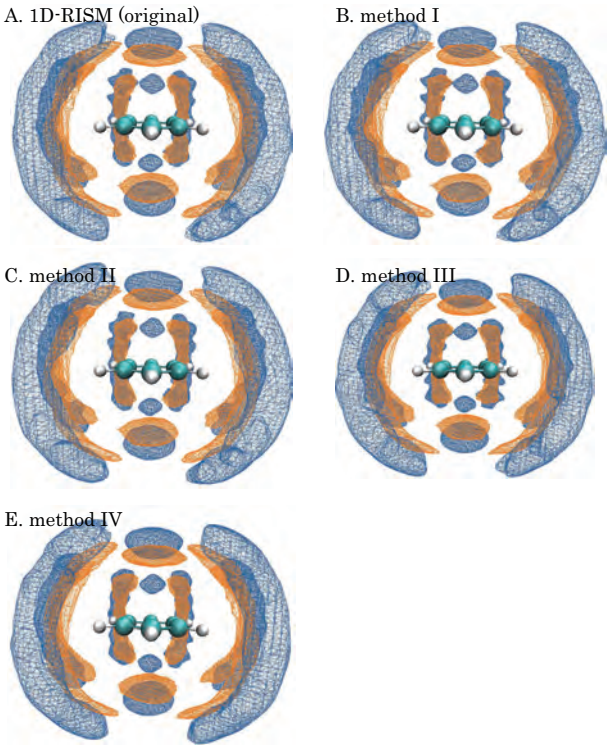


Figure 6.2: 3-dimensional hydration structures near benzene molecule computed by original (1D-RISM, A), method I (B), II (C), III (D) and IV (E) reference terms. Blue (isovalue: 1.5) and orange (isovalue: 3.0) iso-surfaces represent the distributions of hydrogen and oxygen atoms of solvent water molecules, respectively.

They look quite similar each other. In order to quantify the deviation, mean absolute deviation (MAD) on all the grid points is introduced,

$$\mathrm{MAD} = \sum_{s}^{N_\mathrm{v}} \sum_{k}^{N_\mathrm{g}} \frac{\left| h_s^i(\mathbf{r}_k) - h_s^{\text{1D-RISM}}(\mathbf{r}_k) \right|}{N_\mathrm{v} N_\mathrm{g}} \quad (i = \mathrm{I},\ \mathrm{II},\ \mathrm{III\ and\ IV}), \tag{6.8}$$

where $N_\mathrm{v}$ and $N_\mathrm{g}$ denote the numbers of solvent site and total grid points (radial part $\times$ angular part), respectively. Note that MAD is the averaged difference of $h_s^i(\mathbf{r})$ with respect to $h_s^{\mathrm{1D-RISM}}(\mathbf{r})$ on all over the grid points used in the computation. These values of the method I, II, III and IV are $2.12 \times 10^{-3}$, $2.54 \times 10^{-3}$, $2.19 \times 10^{-3}$ and $3.03 \times 10^{-3}$, respectively. They are sufficiently small, hence all the new reference terms provide essentially the same result to the original procedure. In fact, the excess chemical potentials of hydration (hydration free energies, HFEs) and partial molar volumes (PMVs) computed by integration of total and direct correlation functions [19, 20, 21], are virtually identical; HFEs are -2.29 kcal/mol (orig-

inal), -2.29 kcal/mol (I), -2.30 kcal/mol (II), -2.29 kcal/mol (III) and -2.29 kcal/mol (IV). And PMVs are 14.6 $cm^3$/mol (original), 14.6 $cm^3$/mol (I), 14.6 $cm^3$/mol (II), 14.6 $cm^3$/mol (III) and 14.6 $cm^3$/mol (IV). The experimental values of HFE and PMV are -6.3 kcal/mol [28, 29] and 18.0 $cm^3$/mol [30], respectively. Since molecular simulation study using essentially the same potential parameters provides accurate HFE [31], these errors are probably caused by the HNC closure, which does not accurately describe an exclusive volume of solvent molecule [32, 33, 34, 35]. We have checked the effect of the tolerance to solve 1D-RISM equation for judging the convergence and $l$ on MAD. The former ($10^{-10}$ for all calculations in the present study) does not affect MAD at all, and the latter shows slight dependency. For example, the MAD by method I with respect to $h_s^{\text{1D-RISM}}$ ($l = 14$) is gradually improved by increasing $l$: $3.69 \times 10^{-3}$ ($l = 11$), $3.34 \times 10^{-3}$ ($l = 12$) and $3.16 \times 10^{-3}$ ($l = 13$).

Hydration structures near benzene molecule are presented in Fig.5.4. Similar to the water molecule case, the structures by new reference terms (panel B, C, D and E) resemble that by the original one (panel A). The MADs are $1.42 \times 10^{-2}$ (I), $1.78 \times 10^{-2}$ (II), $1.57 \times 10^{-2}$ (III) and $6.92 \times 10^{-3}$ (IV). Compared with the corresponding values of water case, they are slightly greater but hardly affect the HFE estimations; 12.21 kcal/mol (original), 12.21 kcal/mol (I), 12.22 kcal/mol (II), 12.21 kcal/mol (III) and 12.21 kcal/mol (IV). Similarly, PMVs are 54.0 $cm^3$/mol (original), 54.1 $cm^3$/mol (I), 54.0 $cm^3$/mol (II), 54.0 $cm^3$/mol (II) and 54.1 $cm^3$/mol (IV).

These facts show that MC-MOZ is numerically robust enough, at least for these small molecule systems, and any kinds of initial function can be adopted as the reference term. While the new reference function can be prepared non-iteratively, the residue term is determined through a numerical iterative procedure to satisfy the OZ-type equation. The numbers of iterations are different among each solute site because the residue term assigned to this site is independently computed as mentioned above. In the case of water molecule, the averaged iteration numbers are 254 (original), 281 (I), 280 (II), 210 (III) and 207 (IV). Similarly for the benzene calculation the numbers are 203 (original), 197 (I), 197 (II), 200 (III) and 200 (IV). These results suggest that the convergence does not strongly depend on the choice of the

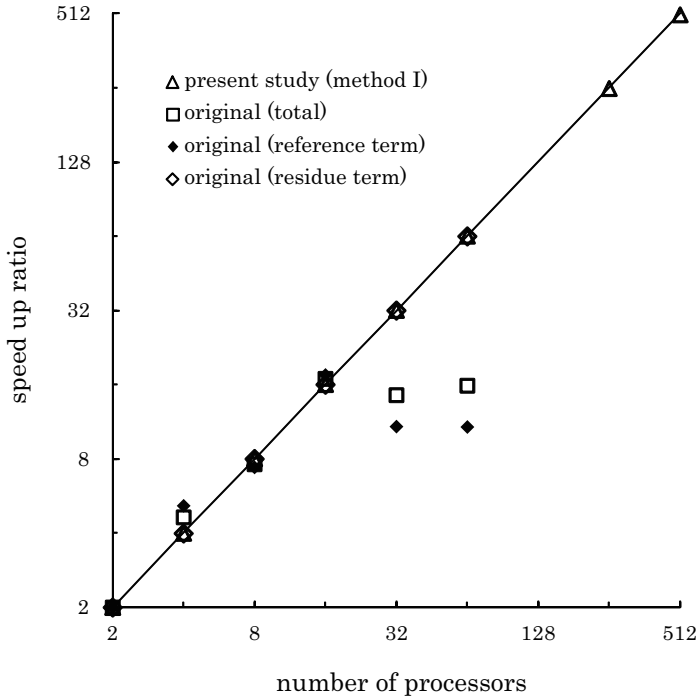


Figure 6.3: The parallel efficiencies in SW-CNT calculations using the original (open squares) and method I (open triangles) as the reference. The subcomponents of the original procedure (reference and residue) are shown.

reference term so much.

### 6.3.2 SW-CNT

Kovalenko-Hirata closure[22] is employed to compute the solvation structure around SW-CNT molecule. As the degree of a real spherical harmonics $l = 11$ is used. The numbers of radial and angular grid points are 512 and 194, respectively. The SW-CNT calculations were performed by CPUs of SGI Altix in Institute for Molecular Science (Okazaki Japan).

To check the parallel efficiency of the calculation, speed up ratio, $S(n)$, is defined as follows.

$$S(n) = 2 \times \frac{[\text{execution time with 2 processors}]}{[\text{execution time with } n \text{ processors}]}, \tag{6.9}$$

where $n$ is the number of processors. If parallelization is perfect, $S(n)$ is identical with $n$. $S(n)$ for the SW-CNT calculations are shown in Fig.6.3. As expected, the present method (I, open triangles) realizes remarkable parallelizations. On the other hand, $S(n)$ by the original procedure (open squares) reach a plateau on $n > 16$. As pointed out in the previous paper[9],

each processor must summarize large-size data from other sites in the each iteration step on the construction of the 1D-RISM reference term (closed diamonds). So this process basically determines the total speed up ratio, especially when a large number of processors is used. In contrast, the procedure of residue term is well-parallelized (open diamonds). As the results, almost all of the total execution times (more than 80%) are consumed by the reference term routine, which makes the parallelization less efficient.

Finally, let us move the discussion of solvation structure around SW-CNT. Different from the cases of water and benzene molecules, the distributions obtained by method I, II and III do not resemble that of the original 1D-RISM reference so much. Their MADs are $1.16\times10^{-1}$ (I), $2.00 \times 10^{-1}$ (II) and $9.91 \times 10^{-2}$ (III). These values are indeed larger than those of water and benzene molecules. On the other hand, MAD of method IV is sufficiently small, $1.10 \times 10^{-2}$, which is similar to the benzene molecule case.

The result indicates that starting from the original 1D-RISM and from method IV reach the same result while the others give different numerical solutions. It is difficult to fully understand this behavior but one possible explanation is the insufficiency of $l$. In other words, the numerical solution of MC-MOZ procedure could not depend on the reference term at the limit of an infinite number of $l$. In order to confirm this, we also performed additional calculation using the greater $l$(= 14) and finer angular grid (the number of points is 302 for a site). Fig.6.4 shows the distributions of CH site in chloroform solvent near SW-CNT. MAD of method IV (panel E) becomes sufficiently small ($9.27 \times 10^{-3}$), indicating that the numerical solution of method IV is virtually the same with the original one (panel A). Similarly, MADs of method I (panel B), II (panel C) and III (panel D) become smaller; $9.43 \times 10^{-2}$, $1.59 \times 10^{-1}$ and $8.24 \times 10^{-2}$, respectively. In fact, the improvement is about 20%. Therefore, the deviation seems to come from rather slow convergence on $l$. It should be emphasized that the original MC-MOZ gives the same result to 3D-RISM as shown in the previous study. By solving 1D-RISM equation, many-body effect is undoubtedly integrated into the reference term. Methods I and III are, however, purely site-independent procedure, meaning that the effect from other site is not taken into account at all. For the simple molecules like water, this insufficiency is re-

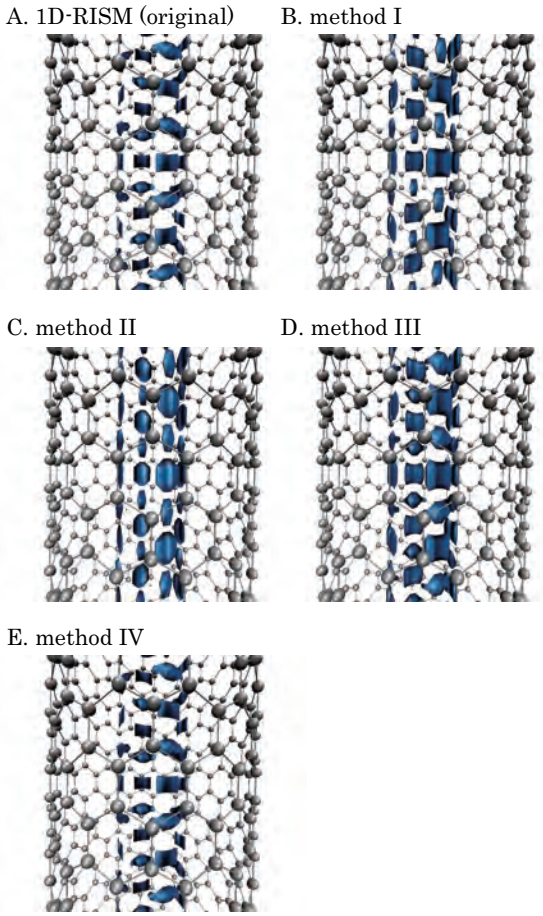


Figure 6.4: 3-dimensional solvation structures inside SW-CNT calculated by original (panel A), method I (panel B), II (panel C), III (panel D) and IV (panel E) reference terms. The iso-surfaces (blue, isovalue: 5.2) represent the distributions of CH site of solvent chloroform molecule.

covered by the residue term. In contrast, both of method II and IV deal with many-body effect (Note that method II utilizes OZ/PY analytical solution for hard sphere system). A possible reason of the superiority of method IV may come from the intramolecular correlation function both for solute-solute and solvent-solvent. The other reason is the effect of partial charges on solvent sites.

## 6.4 Conclusion

In this letter, the four functions were examined as the reference part to investigate the numerical robustness of MC-MOZ method. As the benchmark tests, the hydration structures

around water and benzene molecules are computed. All the reference functions give virtually the same distribution and thermodynamic quantities. Thanks to the site-independent character of the newly proposed references, fully-parallelization is achieved by utilizing them. On the computation of solvation structure around SW-CNT, the speed up ratio of the proposed procedure showed a linear scaling for number of processors. The comparison of the solvation structure near SW-CNT suggests that the reference term by method IV is the promising to efficiently compute 3D solvation structure.

# Bibliography


[1] For example, C. Vieille, G. J. Zeikus, Microbiol. Mol. Biol. Rev. 65 (2001) 1.

[2] D. Chandler, H. C. Andersen, J. Chem. Phys. 57 (1972) 1930.

[3] F. Hirata, P. J. Rossky, Chem. Phys. Lett. 83 (1981) 329.

[4] A. Kovalenko, F. Hirata, Chem. Phys. Lett. 290 (1998) 237.

[5] A. Kovalenko, F. Hirata, J. Chem. Phys. 112 (2000) 10391.

[6] A. Kovalenko, F. Hirata, J. Chem. Phys. 112 (2000) 10403.

[7] N. Yoshida, P. Saree, Y. Maruyama, T. Imai, F. Hirata, J. Am. Chem. Soc. 128 (2006) 12042.

[8] Y. Kiyota, R. Hiraoka, N. Yoshida, Y. Maruyama, T. Imai, F. Hirata, J. Am. Chem. Soc. 131 (2009) 3852.

[9] D. Yokogawa, H. Sato, T. Imai, S. Sakaki, J. Chem. Phys. 130 (2009) 064111.

[10] D. Yokogawa, H. Sato, S. Sakaki, J. Mol. Liq. 147 (2009) 112.

[11] K. Hirano, D. Yokogawa, H. Sato, S. Sakaki, J. Phys. Chem. B 114 (2010) 7935.

[12] J. P. Hansen and I. R. McDonald Theory of Simple Liquids (1990) Academic Press.

[13] F. Hirata Ed. Molecular Theory of Solvation (Understanding Chemical Reactivity) (2004) Kluwer-Springer.

[14] B. Mennucci and R. Cammi Eds. Continuum Solvation Models in Chemical Physics: From Theory to Applications (2008) John Wiley & Sons Inc.

[15] F. Hirata, P. J. Rossky, B. M. Pettitt, J. Chem. Phys. 78 (1983) 4133.

[16] M. S. Wertheim, Phys. Rev. Lett. 10 (1963) 321.

[17] M. S. Wertheim, J. Math. Phys. 5 (1964) 643.

[18] J. L. Lebowitz, Phys. Rev. 133 (1964) A895.

[19] S. J. Singer, D. Chandler, Mol. Phys. 55 (1985) 621.

[20] T. Imai, M. Kinoshita, F. Hirata, J. Chem. Phys. 112 (2000) 9469.

[21] Y. Harano, T. Imai, A. Kovalenko, M. Kinoshita, F. Hirata, J. Chem. Phys. 114 (2001) 9506.

[22] A. Kovalenko, F. Hirata, J. Chem. Phys. 110 (1999) 10095.

[23] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, J. Chem. Phys. 79 (1983) 926.

[24] W. L. Jorgensen, J. D. Madura, Mol. Phys. 99 (1985) 1381.

[25] W. L. Jorgensen, J. M. Briggs, M. L. Contreras, J. Phys. Chem. 94 (1990) 1683.

[26] J. H. Walther, R. L. Jaffe, T. Halicioglu, P. Koumoutsakos, J. Phys. Chem. B 105 (2001) 9980.

[27] T. Werder, J. H. Walther, R. L. Jaffe, T. Halicioglu, P. Koumoutsakos, J. Phys. Chem. B 107 (2003) 1345.

[28] C. H. Reynolds, J. Chem. Inf. Comput. Sci. 35 (1995) 738.

[29] C. J. Cramer, D. G. Truhlar, J. Comput.-Aided Mol. Des. 6 (1993) 629.

[30] D. P. Kharakoz, J. Sol. Chem. 21 (1992) 569.

[31] W. L. Jorgensen, J. Tirado-Rives, Perspect. Drug Discov. 3 (1995) 123.

[32] D. Chandler, Mol. Phys. 31 (1976) 1213.

[33] P. T. Cummings, C. G. Gray, D. E. Sullivan, J. Phys. A 14 (1981) 1483.

[34] M. Kinoshita, F. Hirata, J. Chem. Phys. 106 (1997) 5202.

[35] A. Kovalenko, F. Hirata, J. Chem. Phys. 113 (2000) 2793.

[36] Lide, D. R., Ed CRC Handbook of Chemistry and Physics, 87th ed. CRC press: Boca Raton, FL, (2003) p**8**-540.

# Chapter 7

# General Conclusion

In this thesis, the author studied chemical processes in electrolyte solutions by utilizing an IET framework. In part I, the author extended RISM to multi-component molecular solvent system and the theory was applied to proton dissociation and cation capture processes in solution. In part II, based on MC-MOZ method, the author developed a method to accurately evaluate solvation free energies (SFEs) and an algorithm to achieve a fully computational parallelization of MC-MOZ procedure. The achievements in this thesis are summarized, as follows.

In chapter 2, the author proposed a method to treat pH effects and computed p$K_{\mathrm{a}}$ from first principle calculations at molecular level. In multi-component version of RISM theory, an acidic aqueous solution was regarded as three component system including water, $H_3O^+$ and the counter-anion. Under this assumption, one can obtain the desired pH solution by setting the density of $H_3O^+$ molecules. This procedure was applied to a protonation process of carboxyl group in glycine in HCl aqueous solution. p$K_{\mathrm{a}}$ was determined as the inflection point of the titration curve, showing excellent agreement with the experimental value. Because in the vicinity of the inflection point, the effect of ionic species on a free-energy change of the protonation process was negligible, the free-energy change in pure water is quite similar to that of the inflection point. This result supports that a computation of p$K_{\mathrm{a}}$ by free energy estimation in pure water is adequate in general case.

In chapter 3, levels of theory to accurately evaluate a proton dissociation constant were assessed with RISM-SCF-SEDD method by taking dissociations of amino ($pK_b$) and carboxyl ($pK_a$) groups in glycine in aqueous solution as an example. Molecular orbital (MO) calculations, HF, DFT(B3LYP) and CCSD(T) methods were employed. In addition, two models to treat intermolecular electrostatic interaction between glycine species and solvent water, SEDD and point charge (PC) models, were used to estimate SFE. Both $pK_a$ and $pK_b$ computed by SEDD model showed good agreement with the experimental values for any *ab initio* method, while PC model provided poor $pK_a$ and $pK_b$ results. This was because serious underestimation of the hydration free energies of $NH_2$-$CH_2$-$COO^-$ and $NH_3^+$-$CH_2$-$COO^-$, which suggests that PC model does not properly describe an intermolecular electrostatic potential near negatively charged and electron-delocalized functional groups such as $COO^-$.

In chapter 4, the selective cation capture by the conformation change of 25,26,27,28-tetramethoxycalix[4]arene (CXA) was investigated using multi-component version of RISM-SCF-SEDD combined with MP2. The obtained free energy change ($\Delta\delta G$) was consistent with the observed cation selectivity by NMR experiment. The free-energy decomposition analysis showed that the selectivity was determined by the balance between $\Delta\delta G^{\mathrm{gas}}$ and $\Delta\delta G^{\mathrm{solv}}$: In the sodium capture, $\Delta\delta G^{\mathrm{gas}} < \Delta\delta G^{\mathrm{solv}}$, but $\Delta\delta G^{\mathrm{gas}} > \Delta\delta G^{\mathrm{solv}}$ in the potassium capture. Based on radial distribution functions (RDFs) of the cations around CXA, the positions where the cations were captured were discussed.

In chapter 5, the author proposed an accurate method to compute SFE based on MC-MOZ procedure. The method corresponds to a modified version of repulsive bridge correction developed by A. Kovalenko and F. Hirata. In their original formalism, there was a shortcoming that SFE was seriously underestimated in solvents with complicate molecular shape. In order to improve the underestimation, the repulsive bridge function was modified by utilizing the geometry of solvent molecule. The benchmark tests for aqueous (32 solutes), chloroform (14 solutes) and benzene (12 solutes) solutions provided quite reasonable results. The mean errors from the experimental values were 4.34kcal/mol (aqueous), 1.07kcal/mol (chloroform) and 1.77kcal/mol (benzene).

In chapter 6, an algorithm to obtain a full parallelization in the computation of MC-MOZ method was developed. In the method, the three dimensional pair correlation functions are separated into the reference and the residue terms. The calculation routine of residue term was sufficiently parallelized since the residue term was independently constructed for each solute site. But the procedure of reference term (solving RISM equation) was not because in order to solve the equation, the information of all solute sites must be gathered in the iterative routine. In the chapter, as reference term, the author examined four type functions which could be site-independently prepared. These functions were applied to water, benzene in aqueous solution and single-walled carbon nanotube in chloroform solution. Through the qualitative and quantitative examination of the functions, a promising function was found. At the same time, the high efficiency of parallelization on the computations was achieved.

Chemical process in multi-component solution systems widely appears in various chemical fields. In the process, co-solvent molecules such as ions in electrolyte solution often play an essential role. The elaboration of molecular models is an important issue to elucidate the role at molecular level. The issue is still non-trivial due to a complexity of chemical process in solution. In part I, the author proposed molecular-level tools to treat chemical processes in multi-component solutions based on IET framework and the validity was shown in the several applications. In addition, the author believes that developed methods in part II undoubtedly contribute to quantitatively reliable descriptions of chemical phenomena in solution at molecular level and become promising for investigations of the chemical phenomena involving a large chemical system since the methods are suitable for the parallel computations using a supercomputer with huge number of CPU.

# List of Publications

## Publications included in this thesis

### Chapter 2

”A first principle theory for p$K_{\mathrm{a}}$ prediction at molecular level:
pH effects based on explicit solvent model”
Kentaro Kido, Hirofumi Sato, Shigeyoshi Sakaki
*J. Phys. Chem. B* **2009**, *113*, 10509-10514.

### Chapter 3

”Systematic assessment on aqueous $pK_a$ and $pK_b$ of an amino acid base on RISM-SCF-SEDD method: Toward first principles calculations”
Kentaro Kido, Hirofumi Sato, Shigeyoshi Sakaki
*Int. J. Quant. Chem.* **2012**, *112*, 103-112.

### Chapter 4

”A molecular level study of selective cation capture by a host-guest mechanism for 25,26,27,28-tetramethoxycalix[4]arene in $MClO_4$ solution (M=Na, K)”
Kentaro Kido, Hirofumi Sato, Shigeyoshi Sakaki
*submitted.*

### Chapter 5

”A modified repulsive bridge correction to accurate evaluation of solvation free energy in an integral equation theory for molecular liquids”
Kentaro Kido, Daisuke Yokogawa, Hirofumi Sato
*submitted.*

**Chapter 6**

”Development of a revised version of multi-center molecular Ornstein-Zernike equation”

Kentaro Kido, Daisuke Yokogawa, Hirofumi Sato

*Chem. Phys. Lett.* **2012**, *531*, 223-228.

## Other publications

1. ”A Theoretical Analysis of a Diels-Alder Reaction in Ionic Liquids”
   Seigo Hayaki, Kentaro Kido, Hirofumi Sato, Shigeyoshi Sakaki
   *J. Chem. Phys. B* **2009**, *113*, 8227-8230.

2. ”Ab initio study on S(N)2 reaction of methyl p-nitrobenzenesulfonate and chloride anion in [mmim][PF(6)]”
   Seigo Hayaki, Kentaro Kido, Hirofumi Sato, Shigeyoshi Sakaki
   *Phys. Chem. Chem. Phys.* **2010**, *12*, 1822-1826.