Development of solvation theories focused on solvation structure and electronic structure

Daisuke Yokogawa

2008

Preface

The reaction in solvent is common in our life. An infinite number of biochemical reactions in our body, for example, proceed in aqueous phase. From the industrial aspect, many materials are produced by reactions using various type of organic solvents. Although our society and our body make use of a huge number of reactions in solvent everyday, the reaction, where mechanism is fully elucidated by solvation theories, is very limited.

A theoretical approach to clarify the reactions in solvent was started from the end of 19th century. The most difficult but most interesting point in the development of the solvation theories is how to treat the infinite number of variables in solvent system. Many solvent models have been proposed to tackle the difficulty. In the field of solvation chemistry, by using one of the models or by combining some of the models, the reaction in solvent has been discussed mainly based on solvation structure and the reaction field produced by solvent molecules.

Recently, a theoretical approach to study reactions in solvent from a different point of view has been also proceeded. Thanks to the improvement of the computational system, the quantum chemical calculation with solvation effect is becoming possible. In this field, the reaction energies and the geometries of solute molecules are the main targets to be elucidated.

The theoretical study of the reaction in solvent is now active area, where the two fields of solvation chemistry and quantum chemistry are overlapping. With computational chemistry softwares, various type of the reactions in solvent have been studied. However the theoretical approaches in this area seems to be biased toward one field. In most of quantum chemical calculation with solvation effect, for example, the discussion is mainly focused on the electronic structure and solvation energies. To make clear the functions of solvent molecules, the

microscopic character of the solvent, such as solvation structure, must be elucidated.

In this thesis, the solvation theories focused on both of electronic structure of a solute molecule and solvation structure were developed. With these theories, the reaction in solvent can be discussed at molecular level. The author carried out this study in the hope that the theories developed here will enlarge the overlapping area between the solvation chemistry and quantum chemistry and will work well to elucidate the mechanism of the reactions in solvent.

The studies presented in this thesis were carried out at Department of Molecular Engineering, Graduate School of Engineering, Kyoto University from 2003 to 2008. The author would like to express his deepest appreciation to Professor Shigeyoshi Sakaki for his helpful discussions, useful suggestions, and encouragement. Various comments based on theoretical chemistry and physical chemistry were invaluable for this study. The author wishes to express his sincere gratitude to Associate Professor Hirofumi Sato for valuable suggestions, fruitful discussions and strict but heartwarming advices. The development of theories in this thesis could not be possible without his full cooperation. The author also expresses gratitude to Assistant Professor Yoshihide Nakao for his technical advice and valuable comments.

The author is also grateful to Dr. Atsushi Ikeda and Mr. Yu-ya Ohnishi. Scientific talks with them on various occasions were very enjoyable and fruitful. Acknowledgment is also made to all members of the research group of Prof. Shigeyoshi Sakaki. The discussion with them were exciting and generated his motivation of this study.

The author thanks Japan Society for the Promotion of Science (JSPS) for financial support (Fellowship for Japanese Junior Scientists).

Finally, the author sincerely thanks his parents Kiyoshi Yokogawa and Naoko Yokogawa for their understanding, encouragement, and continuous support.

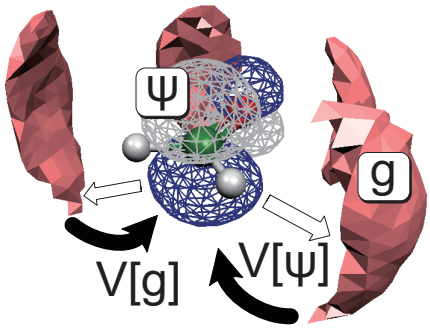
Daisuke Yokogawa

July, 2008

Contents

General Introduction

Solvent molecules play a leading part of reaction in solvent. The electric field produced by solvent has great influence to the reactivity of many reactions and the absorption and emission spectra. In some case, the solvent molecule itself works as a key substrate for chemical reactions and biochemical reactions. With the recent developing theoretical methods and computational systems, these systems are becoming the target of theoretical study.



Scheme 1

The system of the reaction in solvent consists of infinite number of molecules. To characterize such complicated system, what kind of properties should be considered? Because bond breaking and bond forming are included in most of the reactions, electronic structure Ψ of the solute system is of course indispensable (scheme 1). Around the solute system, the solvent molecules move vigorously at room temperature. By averaging the coordinates of the solvent molecules around the solute molecule for a long time, a specific structure, such as hydrogen bonding, is found. Because such a specific structure plays an important role to activate or deactivate the reaction, the solvation structure $g(\mathbf{r})$ is also important property to describe this system (scheme 1).

The electronic structure Ψ and solvation structure $g(\mathbf{r})$ are often considered as independent functions. However, it is not true. The structures Ψ and $g(\mathbf{r})$ produce the electrostatic potential $V[\Psi]$ and $V[g]$ around them. Both of $V[\Psi]$ and $V[g]$ lead to the change of the structures $g(\mathbf{r})$ and Ψ, respectively (see scheme 1). Therefore, the structures Ψ and $g(\mathbf{r})$ are coupled with each other through this potential V.

The reaction in solvent is very complicated system, where solvent system and solute system affect each other. To study the reaction in theoretical manner, how to evaluate the structures Ψ and $g(\mathbf{r})$, as well as the electrostatic potential V plays key roles.

1 Electrostatic potential V

A molecule consists of the positively charged nuclei and negatively charged electrons. These particles produce the electrostatic potential around the molecule. In this section, the author describes how to define the electrostatic potential V as a functional of electronic structure Ψ and solvation structure $g(\mathbf{r})$.

1.1 V as a functional of Ψ

When wave function Ψ is obtained for solute system, the electrostatic potential V is exactly calculated by,

$$V(\mathbf{r}) = \int \frac{\Psi^*(\mathbf{r}')\Psi(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}\mathbf{r}'. \tag{1}$$

This is one of the fundamental property in quantum chemical calculation. By expanding Ψ with atomic orbitals, eq. 1 can be calculated easily [1].

In the classical limit, atomic charge q is widely employed instead of Ψ. The electrostatic potential is replaced by,

$$V(\mathbf{r}) = \sum_i^{N_a} \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|}, \tag{2}$$

where N_a is the number of atoms and $\mathbf{r}_i$ is the position of the atom. In most case, the atomic charge can be determined so that the potential calculated by eq. 2 reproduces $V(\mathbf{r})$ evaluated by eq. 1.

1.2 V as a functional of $g(\mathbf{r})$

The electrostatic potential V generated by solvent molecules is calculated from solvation structure. One of the most difficult points in the calculation is how to treat the solvent system, which has infinite number of variables. To overcome (or avoid) the difficulty, many useful theories such as *molecular simulation*, *integral equation theory*, and *dielectric continuum model* have been proposed (Fig. 1).

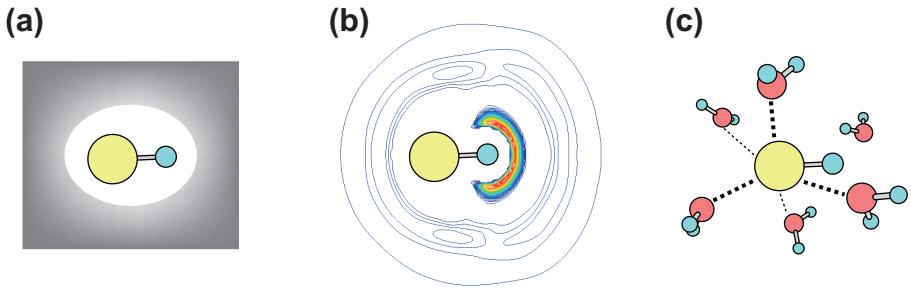


Figure 1: Examples of solvation theories; (a) Dielectric continuum model, (b) Integral equation theories, and (c) molecular simulation.

The most simple method is dielectric continuum model. In this model, the coordinates of all of the solvent molecules are completely averaged (or ignored) and solute molecule is immersed into a cavity embedded in the continuum (Fig. 1(a)). Because the solvent system is characterized solely by a dielectric constant ϵ, the equation to be solved becomes very simple. When a solute molecule is put into a spherical cavity with radius a, the electrostatic potential is expressed with,

$$V(r,\hat{\Omega}) = -\sum_{l=0}^{\infty} \frac{(l+1)(\epsilon-1)}{\epsilon(l+1)+l} \frac{r^l}{a^{2l+1}} \sum_{m=-l}^{l} M_{lm}(\hat{\Omega}), \tag{3}$$

where $\hat{\Omega}$ is Euler angle and M_{lm} is the multipole moment of the solute molecule. The important low-order terms were proposed by Born ($l = 0$) [2], Onsager ($l = 1$) [3], and Abraham *et al.*($l = 2$) [4]. This strategy is very simple and the microscopic properties such as solvation structure cannot be obtained.

In molecular simulation, solvent molecules are treated explicitly (Fig. 1(c)). The coordinate sets of the molecules $(\mathbf{r}_1^{(i)}, \mathbf{r}_2^{(i)}, \cdots, \mathbf{r}_M^{(i)})$ at the i-th step are produced by the Metropolis method

or equation of motion [5], where M is the number of particles in this system. The electrostatic potential produced by the solvent molecules is evaluated by,

$$V(\mathbf{r}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{q_j}{|\mathbf{r}-\mathbf{r}_j^{(i)}|}, \tag{4}$$

where q_j is the atomic charge of the j-th particle and N is the number of the steps calculated in this simulation. If N and M were infinitely large, exact potential could be obtained in principle. However, the calculation in reality is restricted with the finite numbers of N ($10^6 \sim 10^8$) and M ($10^2 \sim 10^4$).

In integral equation theories for liquids (IETs), the coordinate of each solvent molecule is not calculated explicitly. Instead of the coordinate set $(\mathbf{r}_1^{(i)}, \mathbf{r}_2^{(i)}, \cdots, \mathbf{r}_M^{(i)})$, the following pair distribution is employed well [6–12],

$$\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\Xi}\sum_{N=0}^{\infty}\frac{z^N}{(N-2)!}\int\cdots\int \mathrm{d}\mathbf{q}_3\cdots\mathrm{d}\mathbf{q}_N \mathrm{e}^{-\frac{1}{k_BT}U(\mathbf{r}_1,\mathbf{r}_2,\cdots,\mathbf{q}_N)}, \tag{5}$$

where Ξ is the grand partition function, k_B is Boltzmann's constant, U is the intermolecular interaction between molecules, and z is the activity. When a solute molecule is fixed at the origin and the solvent system is homogeneous, the solvation structure around solute molecule $g(\mathbf{r})$ is obtained by $g(\mathbf{r}) = \rho^{(2)}(\mathbf{r})/\rho_u\rho_v$, where $\mathbf{r}$ is the vector from the origin to the solvent site and ρ_u (ρ_v) is the number density of the solute molecules (solvent molecules). With the solvation structure $g(\mathbf{r})$, the electrostatic potential is evaluated by the following equation,

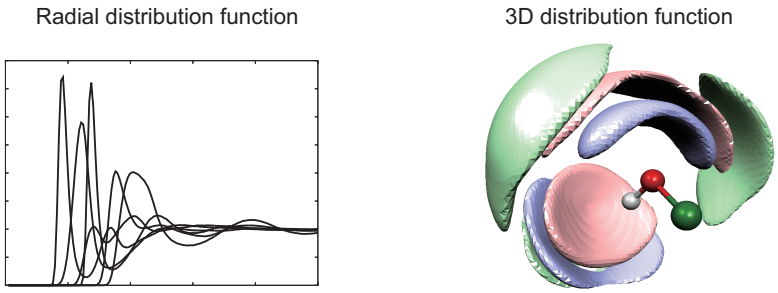
$$V(\mathbf{r}) = \rho_v\sum_{j=1}^{M'}\int\frac{g_j(\mathbf{r}')q_j}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}', \tag{6}$$

where M' is the number of atoms in a solvent molecule and q_j is the atomic charge of these atoms.

The information of the coordinate set $(\mathbf{r}_1^{(i)}, \mathbf{r}_2^{(i)}, \cdots, \mathbf{r}_M^{(i)})$ used in molecular simulation is unnecessarily detailed for many purposes. On the other hand, the coordinates are completely averaged in dielectric continuum model. By introducing $g(\mathbf{r})$, the variables in eq. 6 are dramatically reduced compared to those in eq. 4. Moreover, $g(\mathbf{r})$ is informative enough to analyze the solvation structure at molecular level. The preparation of $g(\mathbf{r})$ is the crucial task to compute $V[g]$ accurately.

2 Solvation structure, g

Solvation structure $g(\mathbf{r})$ is one of the most important properties for the system because it connects directly to the electrostatic potential V as shown in eq. 6.



Scheme 2

What does the solvation structure look like? The structure of solid state is very clear. All of the molecules take the almost fixed coordinates and the structure is completely ordered. On the other hand, in gas phase, the molecules move randomly and the structure around a molecule is completely uniform all over the space. The solvent structure is between them; not completely ordered and not completely uniform.

When the probability of finding two particles at a distance is observed, the 1D solvation structure called radial distribution function (RDF) is evaluated (scheme 2). RDF has been employed to characterize solvation structures for a long time. By neutron scattering and X-ray diffraction, RDFs of liquid systems were obtained by performing Fourier transformation of the static structural factors [13–17]. They can also be obtained by theoretical methods such as molecular simulation [5] and IETs [6]. In the case of Reference Interaction Site Model (RISM) [11, 18], which is one of the most popular IETs, the RDF between sites α and β, $g_{\alpha\beta}$, is calculated by the following equations,

$$g_{\alpha\beta}(r) = h_{\alpha\beta}(r) + 1, \tag{7}$$

$$h_{\alpha\beta}(r) = \sum_{\delta\gamma} \omega_{\alpha\delta} * c_{\delta\gamma} * \omega_{\gamma\beta}(r) + \sum_{\delta\gamma} \omega_{\alpha\delta} * c_{\delta\gamma} * \rho_v h_{\gamma\beta}(r), \tag{8}$$

$$g_{\alpha\beta}(r) = \exp\left[-\frac{1}{k_B T} u_{\alpha\beta}(r) + h_{\alpha\beta}(r) - c_{\alpha\beta}(r)\right], \tag{9}$$

where * denotes the spacial convolution integral, $u_{\alpha\beta}(r)$ is the potential between site α and site β, and $c_{\alpha\beta}(r)$ is the direct correlation function. $\omega_{\alpha\beta}(r)$ is the intramolecular correlation function, which defines the molecular geometry.

RDFs have been well employed to discuss the solvation structure. However, it becomes very difficult to imagine the solvation structure only from RDFs as the number of atoms in a molecule increases. Even in the case of water, the assignment of some peaks in the RDF is very difficult only with the 1D data. This is because the information about the angular part is completely integrated out in RDF.

The computational methods which can directly evaluate 3D solvation structure have also been developed. The solvation structure of liquid water, for example, was discussed in detail with molecular simulation approach [19–21]. In the case of IET, three-dimensional RISM (3D-RISM) [8, 22] has been applied to much larger and more complicated solvation systems than those evaluated by molecular simulation, such as hydration structure around a protein. The solvation structure of the site α around a solute molecule $g_\alpha(\mathbf{r})$ is evaluated with the following equations,

$$g_\alpha(\mathbf{r}) = H_\alpha(\mathbf{r}) + 1 \tag{10}$$

$$H_\alpha(\mathbf{r}) = \sum_\beta C_\beta * \omega_{\beta\alpha}(\mathbf{r}) + \sum_\beta C_\beta * \rho h_{\beta\alpha}(\mathbf{r}) \tag{11}$$

$$g_\alpha(\mathbf{r}) = \exp\left[-\frac{1}{k_B T} u_\alpha(\mathbf{r}) + H_\alpha(\mathbf{r}) - C_\alpha(\mathbf{r})\right], \tag{12}$$

where $u_\alpha(\mathbf{r})$ is the potential function between site α and the solute molecule. This method has been applied to the solvation structures of not only the small systems [23] but also the hydration structure around a protein [24–26]. Although the 3D solvation structure calculated by 3D-RISM or molecular simulation is much more informative than the RDF, high computational cost and long computational time are required to obtain accurate solvation structure.

3 Electronic structure, Ψ

The wave function Ψ with solvation effect is determined by the following equation,

$$\left[\hat{H}^0 + \hat{\mathscr{V}}\right] |\Psi\rangle = E\,|\Psi\rangle\,, \tag{13}$$

where $\hat{H}^0$ is the Hamiltonian of a solute molecule in isolated state and $\hat{\mathscr{V}}$ is the electrostatic interaction operator [27–30]. As shown in scheme 1, the wave function Ψ affects solvation structure $g(r)$ through $V[\Psi]$ and the $g(r)$ affects the wave function Ψ through $V[g(r)]$. The operator $\hat{\mathscr{V}}$ is introduced to incorporate the interaction.

Eq. 13 can be derived from the variation condition on free energy of the system [31, 32]. The total free energy is written by

$$G = \left\langle \Psi|\hat{H}^0|\Psi \right\rangle + E_{\rm nuc} + G_{\rm sol}\left[V[\Psi], \mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_M\right], \tag{14}$$

where $E_{\rm nuc}$ is the nuclear repulsion energy. $G_{\rm sol}$ is the solvation free energy evaluated under the electrostatic potential $V[\Psi]$ and some variables $\mathbf{a}_i$ $(i = 1, \cdots M)$ characterizing solvents (ex. $\mathbf{h}$ and $\mathbf{c}$ in RISM). The trial function to be minimized with the constrains to the orthonormality of the wave function is defined, as follows;

$$\mathscr{L} \equiv G + E(\langle\Psi|\Psi\rangle - 1). \tag{15}$$

Variations with respect to the functions yield

$$\delta\mathscr{L} = \sum_i \left(\frac{\partial G_{\rm sol}}{\partial \mathbf{a}_i}\right)\delta\mathbf{a}_i + 2\left\langle \delta\Psi \left| \hat{H}^0 + \left(\frac{\partial G_{\rm sol}}{\partial V}\right)\left(\frac{\partial V[\Psi]}{\partial \Psi}\right) - E \right| \Psi \right\rangle. \tag{16}$$

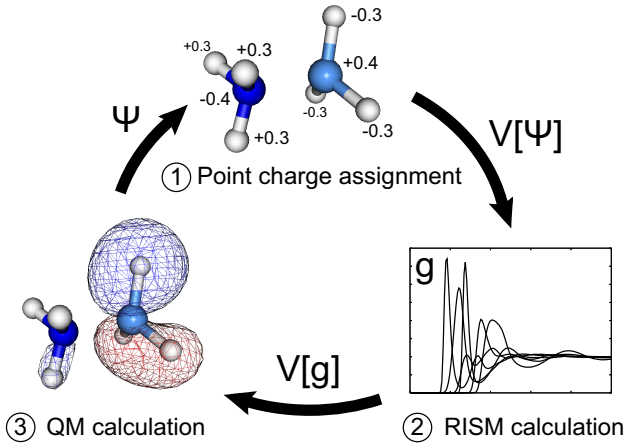
Because $G_{\rm sol}$ is minimum with respect to $\{\mathbf{a}_i\}$ in this scheme, the first term of eq. 16 is 0. Therefore, the optimal wave function is calculated by the following equation,

$$E\,|\Psi\rangle = \left[\hat{H}^0 + \left(\frac{\partial G_{\rm sol}}{\partial V}\right)\left(\frac{\partial V[\Psi]}{\partial \Psi}\right)\right]|\Psi\rangle \tag{17}$$

If the second term in brackets on the right-hand side is replaced by $\hat{\mathscr{V}}$, eq. 17 is equal to eq. 13.

Hybrid approaches have been developed by combining with many solvation theories. In quantum mechanical/molecular mechanical (QM/MM) approach, small systems, such as the

excitation reaction of H_2CO [33], glycine [34], and S_N2 reaction [35] in aqueous phase were investigated precisely. In dielectric continuum approach, Polarizable Continuum Model (PCM) [36–40] has been applied to various physicochemical systems, such as diazine in polar solvent [41], 5-fluorouracil and uracil in acetonitrile and water [42], and solvatochromism of betaine-30 [43]. Although these methods have been widely used, they have some weakpoints; the former method requires much computational cost and time, and the latter method cannot evaluate the solvation structure.



Scheme 3

The hybrid QM calculation with RISM (RISM-SCF) [28, 32, 44, 45] is one of the powerful method to overcome the weakpoints in the QM/MM approach and dielectric continuum model. In the procedure of RISM-SCF, the cycle in scheme 3 is taken; ① point charges on atoms are evaluated after QM calculation, ② RISM calculation is performed with the electrostatic potentials $V[\Psi]$, and ③ QM calculation is carried out with the electrostatic potential $V[g]$ produced by the RDFs. This cycle is repeated until self-consistent structures between Ψ and $g(\mathbf{r})$ are obtained. With this strategy, not only the solvation energies but also 1D solvation structure, RDF can be evaluated. Moreover, the calculation of electrostatic potential $V[g]$ with RDFs is so simple that the computational cost and time are much smaller than those of QM/MM approach. RISM-SCF has been applied to many systems in solution, such as S_N2 type reaction [46, 47], electron transfer reaction [48], proton transfer reaction [49], and

charging process of organic compounds [50].

4 Problems in the calculation of solvation structure and electronic structure

Effective solvation theories have been proposed and applied to many systems to investigate solvation structure and electronic structure with solvation effect theoretically. However, there are still some problems to be solved.

In the analysis of solvation structures, 1D solvation structure (ex. RDF) and 3D solvation structure have been investigated. RDFs have important information about solvation structure and have been widely employed in experimental and theoretical studies. However, because the information of angular part is completely integrated out, it is difficult to image the orientation of solvent molecules from only RDFs as the number of atoms in a molecule increases. 3D solvation structure is more convenient information in this analysis than RDF, but the required computational cost and time are huge even with the recent computer system. To obtain the informative 3D solvation structure with reasonable cost and time, new methods are required.

In the calculation of electronic structure, RISM-SCF is very powerful tool which can calculate not only the electronic structure Ψ but also the solvation structure $g(r)$. The computational cost of RISM-SCF is much smaller than that of QM/MM approach. Moreover, the method can obtain the solvation structure, which cannot be calculated with the dielectric continuum model. This method is becoming a powerful tool to study the reaction in solvent theoretically. However, it has been showed that the calculation of RISM-SCF for complex solute molecules such as metal complexes doesn't converge. To enlarge the applicability of RISM-SCF, different strategies should be developed.

5 Survey of the present thesis

In this study, the author developed the solvation theories focused on both of the solvation structure and electronic structure to overcome the problems. This thesis consists of two parts.

In part I, two theoretical approaches to obtain 3D solvation structure were proposed. In chapters 1 and 2, the method to reconstruct the 3D solvation structure from RDFs was developed. The 3D solvation structure is expanded with real solid harmonics and the expansion coefficients are determined so that the calculated RDFs reproduces the reference RDFs. The method was applied to the solvation structure of typical example, H_2O and more complicated system, neat methanol and DMSO solutions. The results obtained by the present method were compared with the reference solvation structures, which were calculated by molecular dynamics.

In chapter 3, a new method to calculate 3D solvation structure from a first principle was proposed. The method which can evaluate high-quality 3D solvation structure was derived by introducing the information of angular part. The 3D solvation structures were compared with those obtained by previous works and molecular simulation.

In chapter 4, the method which can apply to a large system was derived based on the method proposed in chapter 3. The strategy in this method is so efficient that the solvation structure around a large molecule can be evaluated with reasonable computational time and cost. The solvation structure and the partial molar volumes of amino acids calculated by the present method were compared with those obtained by 3D-RISM. This method was also applied to the calculation of hydration structure around a protein called Fv fragment.

In chapter 5, the hydration structure for a Bacteriorhodopsin (bR) was evaluated with the method described in chapter 4. The bound waters inside bR was calculated and compared with those obtained by X-ray crystallography.

In part II, the development and application of RISM-SCF and the calculation of dielectric continuum model were carried out. In chapter 6, the electronic structures of metal complexes in aqueous phase were evaluated using dielectric continuum model. Because the electronic structure has both of the localized and delocalized characters in aqueous phase, the multireference nature of wave function has to be included. The author employed the two-state model proposed by Farazdel *et al* [51] and combined it with dielectric continuum model. The differ-

ence of electronic structure in aqueous phase among these complexes was explained based on molecular orbital theory.

The dielectric continuum model in chapter 6 worked well for the system. However, it is much better if the solvation structure can be investigated at the same time. To fulfill the requirement, in chapters 7, 8, and 9, RISM-SCF approach was employed.

In chapters 7 and 8, the modified charge assignments were introduced into RISM-SCF cycle described in scheme 3. These methods were applied to metal complexes and a charge-transfer complex in aqueous phase, which are the difficult examples to be calculated by the original RISM-SCF.

In chapter 9, spatial electron density distribution (SEDD) was introduced into the RISM-SCF strategy and proposed a new generation of RISM-SCF (RISM-SCF-SEDD). With the present strategy, the instability of the charge fitting in the original RISM-SCF was removed. RISM-SCF-SEDD was applied to small molecules, H_2O, C_2H_5OH, and HLi. The usefulness of the method was clearly shown by comparing the obtained partial charges and solvation structures with those calculated by the original RISM-SCF.

Bibliography

[1] T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, England, 2000).

[2] M. Born, Z. Phys. **1**, 45 (1920).

[3] L. Onsager, J. Am. Chem. Soc. **58**, 1486 (1936).

[4] R. J. Abraham, L. Cavalli, and K. G. R. Pachler, Mol. Phys. **11**, 471 (1966).

[5] D. Frenkel and B. Smit, *Understanding Molecular simulation: From Algorithms to Applications,* 2nd ed. (Academic, New York, 2002).

[6] J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids,* 2nd ed. (Academic, London, 1986).

[7] M. Ikeguchi and J. Doi, J. Chem. Phys. **103**, 5011 (1995).

[8] D. Beglov and B. Roux, J. Phys. Chem. B **101**, 7821 (1997).

[9] C. M. Cortis, P. J. Rossky, and R. A. Friesner, J. Chem. Phys. **107**, 6400 (1997).

[10] S. Ten-no and S. Iwata, J. Chem. Phys. **111**, 4865 (1999).

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

[12] T. Sumi, T. Imai, and F. Hirata, J. Chem. Phys. **115**, 6653 (2001).

[13] A. H. Narten, M. D. Danford, and H. A. Levy, Discuss. Faraday Soc. **43**, 97 (1967).

[14] M. D. Danford H. A. Levy, J. Am. Chem. Soc. **84**, 3965 (1962).

[15] A. K. Soper and J. Turner, Int. J. Mod. Phys. B **7**, 3049 (1993).

[16] T. Yamaguchi, K. Hidaka, and A. K. Soper, Mol. Phys. **96**, 1159 (1999).

[17] T. Takamuku, M. Tabata, A. Yamaguchi, J. Nishimoto, M. Kumamoto, H. Wakita, and T. Yamaguchi, J. Phys. Chem. B **102**, 8880 (1998).

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

[19] I. M. Svishchev and P. G. Kusalik, Chem. Phys. Lett. **239**, 349 (1995).

[20] I. M. Svishchev and P. G. Kusalik, J. Chem. Phys. **99**, 3049 (1993).

[21] A. De Santis and D. Rocca, J. Chem. Phys. **107**, 9559 (1997).

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

[23] A. Kovalenko and F. Hirata, J. Phys. Chem. B **103**, 7942 (1999).

[24] T. Imai, A. Kovalenko, F. Hirata, Chem. Phys. Lett. **395**, 1 (2004).

[25] T. Imai, R. Hiraoka, A. Kovalenko, and F. Hirata, J. Am. Chem. Soc. **127**, 15334 (2005).

[26] N. Yoshida, S. Phongphanphanee, and F. Hirata, J. Phys. Chem. B **111**, 4588 (2007).

[27] S. Miertus, E.Scrocco, J. Tomasi, Chem. Phys. **55**, 117 (1981).

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

[29] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).

[30] J. Gao, Acc. Chem. Res. **29**, 298 (1996).

[31] R. Cammi and J. Tomasi, J. Chem. Phys. **100**, 7495 (1994).

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

[33] Y. Kawashima, M. Dupuis, and K. Hirao, J. Chem. Phys. **117**, 248 (2002).

[34] H. Takahashi, Y. Kawashima, T. Nitta and N. Matubayasi, J. Chem. Phys. **123**, 124504 (2005).

[35] M. Ohisa, H. Yamataka, M. Dupuis, and M. Aida, Phys. Chem. Chem. Phys. **10**, 844 (2008).

[36] H. Li, C. S. Pomelli, and J. H. Jensen, Theor. Chem. Acc. **109**, 71 (2003).

[37] V. Barone and M. Cossi, J. Phys. Chem. A **102**, 1995 (1998).

[38] E. Cancés, B. Mennucci, and J. Tomasi, J. Chem. Phys. **107**, 3032 (1997).

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

[40] J. Tomasi, B. Mennucci, and R. Cammi, Chem. Rev. **105**, 2999 (2005).

[41] B. Mennucci, J. Am. Chem. Soc. **124**, 1506 (2002).

[42] F. Santoro, V. Barone, T. Gustavsson, and R. Improta, J. Am. Chem. Soc. **128**, 16312 (2006).

[43] M. Caricato, B. Mennucci, and J. Tomasi, Mol. Phys. **104**, 875 (2006).

[44] S. Ten-no, F. Hirata, and S. Kato, Chem. Phys. Lett. **214**, 391 (1993).

[45] *Molecular Theory of Solvation*, edited by F. Hirata (Kluwer, Dordrecht, 2003).

[46] K. Naka, H. Sato, A. Morita, F. Hirata, and S. Kato, Theor. Chem. Acc. **102**, 165 (1999).

[47] H. Sato and S. Sakaki, J. Phys. Chem. A **108**, 1629 (2004).

[48] H. Sato and F. Hirata, J. Phys. Chem. A **106**, 2300 (2002).

[49] S. Yamazaki and S. Kato, Chem. Phys. Lett. **386**, 414 (2004).

[50] H. Sato, Y. Kobori, S. Tero-Kubota, and F. Hirata, J. Chem. Phys. **119**, 2753 (2003).

[51] A. Farazdel, M. Dupuis, E. Clementi, and A. Aviram, J. Am. Chem. Soc. **112**, 4206 (1990).

Part I

Theoretical approach to evaluate three dimensional solvation structure

Chapter 1

A new method to reconstruct 3D spatial distribution function from radial distribution function in solvation structure

1.1 Introduction

To understand most of the chemical reactions correctly, solvent effects should be considered in theoretical analysis. Various methods such as dielectric continuum model, molecular dynamics (MD), and integral equation theory of liquids (IET) [1] have been proposed to investigate solvent effects. These methods are useful to calculate solvation energy. However, 3D information on solvent coordination such as a spatial distribution function (SDF) has not been studied thoroughly, except for limited pioneering works [2–10], despite 3D information being very helpful to understand chemical reactions in solution phase by visualizing the solvation feature [11].

One of those works was reported by Soper *et al*. [7–9]. They expanded the SDF as a function of position vector and orientation of solvent using spherical harmonic functions, and optimized the coefficients, which determine the shape of SDF, with the minimum noise formalism [9]. The equation for the coefficients was solved in an iterative manner. Sato *et al*. [3] presented the "most plausible solvation structure" (MPSS) using the radial distribution functions (RDFs). Simulation techniques such as MD method, three dimensional reference interaction model (3D-RISM) [12, 13], can evaluate SDF directly. However, both of the methods need long computational time to calculate SDF. There are other IETs, such as MOZ theory [14–21], that can analyze three-dimensional structure of solvation, though those results have the approximations

inherent in IETs.

On the other hand, RDFs, which are the most frequently used in discussion, are easily evaluated with much shorter computational time. Thus, it is highly desired to develop the method that easily provides SDF from RDFs.

In this communication, we newly propose an interesting method to reconstruct approximated SDFs of solvent site from RDFs, which are calculated by any solvation theory, by employing the spherical harmonic expansion around each solute site. This expansion leads to simple linear equation and we can obtain the coefficients determining the shape of approximated SDFs by solving the equation. The efficiency of this method is clearly shown here by applying this method to the coordination of solvent water around a water molecule.

1.2 Method and computational details

Method

We begin with SDFs of solvent site s around a solute molecule, $n_s(\mathbf{r})$,

$$\rho_s(\mathbf{r}) = \rho n_s(\mathbf{r}), \tag{1.1}$$

where ρ is the number density of solvent and $\mathbf{r}$ is the position vector in 3D space. The RDF between solute site η and solvent site s is related to SDFs by eq. 1.2;

$$g_{\eta,s}(R) = \frac{1}{4\pi R^2} \int n_s(\mathbf{x} + \mathbf{Q}_\eta)\delta(|\mathbf{x}| - R)d\mathbf{x}, \tag{1.2}$$

where $\mathbf{Q}_\eta$ is the position vector of the η site and R represents the distance between the η site and the s site.

The SDF of solvent *s* site is well approximated by basis functions including real solid harmonics S_{lm} of which center is located on individual solute site η,

$$n_s^\eta(\mathbf{r}) = \left[\sum_i^N f_i^{\eta,s}(|\mathbf{r} - \mathbf{Q}_\eta|) \sum_{l,m} C_{l,m}^{\eta,s,i} S_{l,m}\left(\frac{\mathbf{r} - \mathbf{Q}_\eta}{|\mathbf{r} - \mathbf{Q}_\eta|} \right) \right] + f_f^{\eta,s}(|\mathbf{r} - \mathbf{Q}_\eta|), \tag{1.3}$$

where $\{C_{l,m}^{\eta,s,i}\}$ are coefficients, $f_i^{\eta,s}$ and $f_f^{\eta,s}$ are basis functions in radial part up to *N*. After

substitution of eq. 1.3 into eq. 1.2 and integration over the angular part, we get

$$g_{\eta,s}(R) = \sum_{i}^{N} f_i^{\eta,s}(R) C_{0,0}^{\eta,s,i} + f_f^{\eta,s}(R), \tag{1.4}$$

where $\{C_{0,0}^{\eta,s,i}\}$ are fixed as 1 [22]. The basis functions $f_i^{\eta,s}(R)$ and $f_f^{\eta,s}(R)$ in eq. 1.4 are determined to reproduce $g_{\eta,s}(R)$. $f_i^{\eta,s}(R)$ and $f_f^{\eta,s}(R)$ are used to represent solvation shell and bulk solvent region, respectively. The function $f_f^{\eta,s}(R)$ converges to 1 in $R \to \infty$. Hereafter we call the RDF calculated by solvation theory as reference RDF (RRDF). If we can expand RRDFs exactly by $f_i^{\eta,s}(R)$ and $f_f^{\eta,s}(R)$, RRDFs are reproduced correctly by integrating over the angular part of $n_s^{\eta}(\mathbf{r})$ in eq. 1.3 around solute η site.

It is noted that in each SDF for solvent *s* site, there are expanded SDFs to the number of solute site N_a. The coefficients $\{C_{l,m}^{\eta,s,i}\}$ in eq. 1.3 are determined so that the difference between all these SDFs centered on different solute site becomes the smallest; when a sufficiently large number of real solid harmonics and basis functions are employed, all of the SDFs become almost unique. Here, we define the error function Γ_s as follows:

$$\Gamma_s = \frac{1}{2} \sum_{k}^{N_p} \sum_{\eta \neq \gamma}^{N_a} (n_s^{\eta}(\mathbf{r}_k) - n_s^{\gamma}(\mathbf{r}_k))^2, \tag{1.5}$$

where N_p is the number of the grid points which are prepared around the solute. As the number of basis functions in eq. 1.3 increases, Γ_s decreases. By minimizing eq. 1.5, we can obtain the equation that determines the coefficients $\{C_{l,m}^{\eta,s,i}\}$:

$$\mathbf{C} = -\mathbf{A}^{-1}\mathbf{B}, \tag{1.6}$$

where $\mathbf{C}$ is the vector whose component is $C_{l,m}^{\eta,s,i}$ and the components of $\mathbf{A}$ and $\mathbf{B}$ are given by eq. 1.7a and 1.7b;

$$\begin{aligned} A_{\eta,j,l',m'}^{\gamma,s,i,l,m} = & (N_a \delta_{\eta,\gamma} - 1) \sum_{k=1}^{N_P} f_i^{\eta,s}(|\mathbf{r_k} - \mathbf{Q}_\eta|) f_j^{\gamma,s}(|\mathbf{r_k} - \mathbf{Q}_\gamma|) \\ & \times \quad S_{l,m}\left(\frac{\mathbf{r_k} - \mathbf{Q}_\eta}{|\mathbf{r_k} - \mathbf{Q}_\eta|}\right) S_{l',m'}\left(\frac{\mathbf{r_k} - \mathbf{Q}_\gamma}{|\mathbf{r_k} - \mathbf{Q}_\gamma|}\right), \end{aligned} \tag{1.7a}$$

$$\begin{aligned} B^{\eta,s,i,l,m} = & \sum_{k=1}^{N_P} f_i^{\eta,s}(|\mathbf{r_k} - \mathbf{Q}_\eta|) S_{l,m}\left(\frac{\mathbf{r_k} - \mathbf{Q}_\eta}{|\mathbf{r_k} - \mathbf{Q}_\eta|}\right) \\ & \times \quad \sum_{\gamma} (N_a \delta_{\eta,\gamma} - 1) f_f^{\gamma,s}(|\mathbf{r_k} - \mathbf{Q}_\gamma|). \end{aligned} \tag{1.7b}$$

By solving this linear equation, we can obtain the coefficients without any iterative calculation. In this regard, the present method provides approximated SDFs very easily.

Because we cannot use infinite number of basis functions in practice, the quality of $n_s^\eta(\mathbf{r})$s is not uniform in all space; in other words, $n_s^\eta(\mathbf{r})$ is given in high accuracy where $\mathbf{r}$ is close to solute site η, but it gets worse when $\mathbf{r}$ is distant from η site. To obtain well-balanced SDF $n_s(\mathbf{r})$ from SDFs in eq. 1.3, we used arithmetic average of $n_s^\eta(\mathbf{r})$,

$$n_s(\mathbf{r}) = \sum_{\alpha}^{N_a} n_s^\alpha(\mathbf{r})/N_a. \tag{1.8}$$

The computational procedure is summarized as follows: (1) The first step is to fit $f_{i/f}^{\eta,s}(R)$ to RRDFs, (2) the second step is to evaluate eqs. 1.7a and 1.7b and to solve eq. 1.6, and (3) the third is to take average $n_s^\eta(\mathbf{r})$ with eq. 1.8.

1.2.1 Computational details

We employed the RRDFs calculated by the MD method reported by Jorgensen *et al*. [23]. Also, we evaluated here RRDFs with the extended RISM (XRISM) [24], where the simple point charge (SPC)-like water model [25] was employed. The standard modification for LJ parameters was made in the hydrogen site. The hypernetted chain (HNC)-like closure was used in solving XRISM equation. All calculations were carried out at the temperature 298.15 K and the number density 0.033426 molecule/Å^3. Hereafter, we call the RRDFs of MD and XRISM as RRDFs(MD) and RRDFs(XRISM), respectively.

We used gaussian functions for $f_i^{\eta,s}(R)$ and the following function of eq. 1.9 for $f_f^{\eta,s}(R)$,

$$f_f^{\eta,s}(R) = \frac{1}{\exp(-\alpha^{\eta,s}(R-\beta^{\eta,s}))+1} \quad , \tag{1.9}$$

to approximate RRDFs in the least square fitting technique. In this calculation, we employed the real solid harmonics up to $l = 10$.

1.3 Results and discussion

The RRDFs(XRISM) and RRDFs(MD) of oxygen-oxygen and hydrogen-oxygen are shown in Fig. 1.1(a) as lines **II** and **IV** and Fig. 1.1(b) as lines **II** and **IV**, respectively. The RDFs of

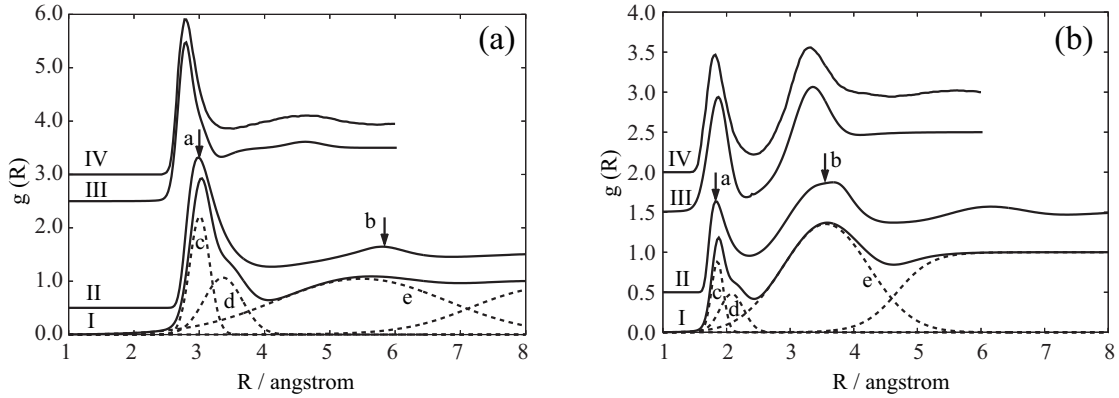


Figure 1.1: Fitted RDFs and RRDFs; (a) oxygen-oxygen and (b) hydrogen-oxygen. Solid lines **I**, **II**, **III**, and **IV** represent fitted RDF(XRISM), RRDF(XRISM), fitted RDF(MD), and RRDF(MD), respectively. Dashed lines represent basis functions used for the fit of RRDF(XRISM).

the XRISM (**I**) and the MD (**III**) methods were constructed by the fitted basis functions, as shown in Fig. 1.1(a) and Fig. 1.1(b), in comparison with **II** and **IV**. We found that three $f_i^{\eta,s}$ functions and one $f_f^{\eta,s}$ function are enough to reproduce well RRDFs (**II** and **IV**). However, we cannot reproduce the second peak (**b** in Fig. 1.1(a) and 1.1(b)) properly, when using only two $f_i^{\eta,s}$ functions and one $f_f^{\eta,s}$ function. One can see that **I** and **III** well reproduce **II** and **IV**, respectively. The basis sets **c** and **e** mainly used to reproduce the peaks **a** and **b** in RRDFs. The second basis set **d** seems to correspond to the distribution of the "interstitial water molecules", which is suggested to be around $R = 3.5\text{Å}$ [5].

The resultant solvent coordinations in 3D, $n_O(\mathbf{r})$ and $n_H(\mathbf{r})$, are used to calculate the charge density by the following equation,

$$\rho_q(\mathbf{r}) = \rho q_O n_O(\mathbf{r}) + 2\rho q_H n_H(\mathbf{r}), \tag{1.10}$$

where q_O and q_H are the charges of the oxygen and the hydrogen of solvent water, respectively. The charge density in the plane including all the atoms (XZ-plane) is shown in Fig. 1.2(a) and that in the bisector plane of the HOH angle (YZ-plane) is shown in Fig. 1.2(b). Because these density maps have C_{2v} symmetry, we divided them into two regions and show only a half of them in these figures. The right hand side is for the MD method and the left hand side for the XRISM. The negative density (solid line) represents $n_O(\mathbf{r})$ and the positive density (dashed

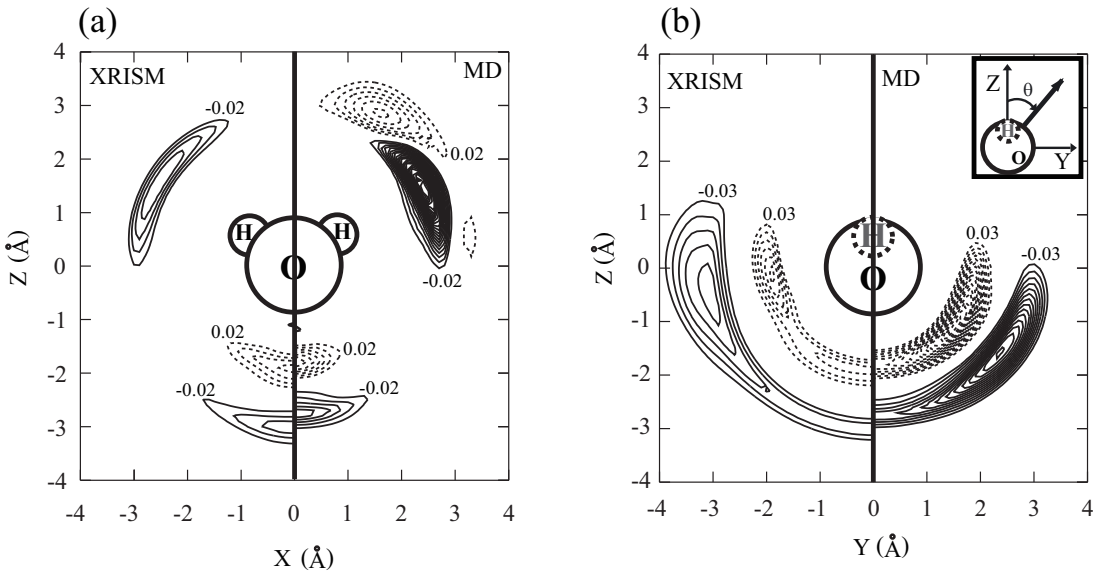


Figure 1.2: Charge distribution ($|e|/\text{Å}^3$); (a) in the plane including the oxygen and two hydrogens and (b) in the bisector plane of HOH angle. Solid line represents negative part (oxygen), and dashed line is for the positive part (hydrogen). The charge distributions constructed by MD and XRISM are shown in the region ($X > 0$) and the region ($X < 0$), respectively.

line) represents $n_H(\mathbf{r})$. In Fig. 1.2(a), one can see the negative charge distribution around the hydrogen of the central water (hereafter, we call it "solute water"). When RRDFs(XRISM) were employed, the maximum of the distribution is located at ($X = -2.5$Å, $Z = 1.6$Å) and the distance between the maximum and the hydrogen of "solute water" is evaluated to be 2.0 Å. When RRDFs(MD) were employed, the maximum is at ($X = 2.4$Å, $Z = 1.4$Å) and the distance is 1.8 Å. Thus, these distributions are attributed to the first peak in Fig. 1.1(b). These negative charge distributions arise from the oxygen that forms hydrogen bond with "solute water". Positive charge distribution and negative one are at ($X = 0.0$, $Z = -2.0$Å) and ($X = 0.0$, $Z = -3.1$Å) when RRDFs(XRISM) are employed and at ($X = 0.0$, $Z = -1.8$Å) and ($X = 0.0$, $Z = -2.8$Å) when RRDFs(MD) are employed. These distributions are attributed to different type of hydrogen bonding solvent, as will be discussed below. Fig. 1.2(b) shows charge distribution on the bisector plane of HOH. The positive distribution and negative one indicate that solvent water molecules are present over the average at the region of $Z < 0$. The angle θ (defined in the right-upper box) of O distribution is evaluated to be

about 102 degree when RRDFs(XRISM) are employed and 124 degree when RRDFs(MD) are employed, respectively. The latter value agrees well with previously reported value (about 130 degree) [10, 14]. The SDF of solvent oxygen site reconstructed by RRDFs(MD) present the θ value close to 125.3 (=$\cos^{-1}(-\frac{1}{\sqrt{3}})$) degree which is the exact value when the oxygen takes perfect tetrahedral network. However, the θ value calculated with RRDFs(XRISM) is somewhat smaller than that of 125.3 degree. This is because the first peak of oxygen shifts to outer region and the first peak of hydrogen shifts to inner region in RRDF(XRISM). We can conclude that this small θ value is attributed not to the present reconstruction method but to RRDFs(XRISM). As shown in Fig. 1.2(a) and Fig. 1.2(b), broad distribution is observed around $(X, Y, Z) = (0.0, 0.0, -1.9)$, $(0.0, 0.0, -3.0)$ when RRDFs(XRISM) were employed and around $(X, Y, Z) = (0.0, 0.0, -1.8)$, $(0.0, 0.0, -2.8)$ when RRDFs(MD) were employed. If water coordination took a simple tetrahedron structure ("standard coordination"), there should be no distribution there and two separated distribution could be found at about $\theta = \pm 125$. These broad distributions indicate a great deal of variation of solvent water from "standard coordination". This continuous distribution of solvent water was also reported by Soper *et al* [8].

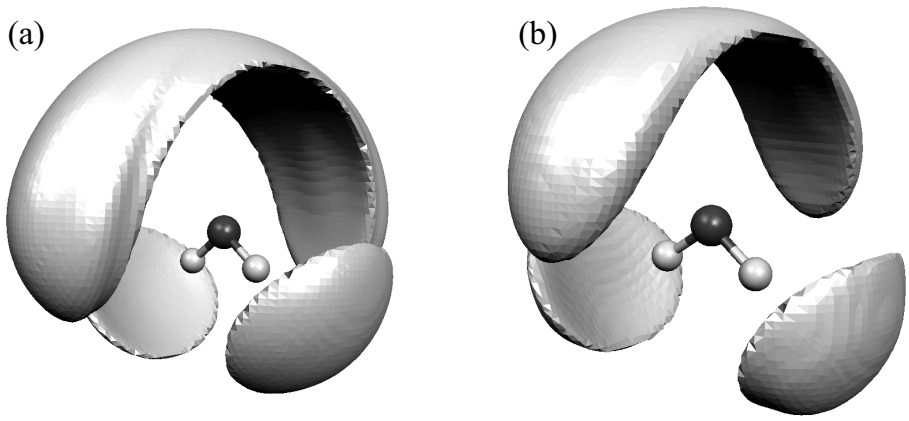


Figure 1.3: Three dimensional SDF map of the oxygen in solvent water using the real solid harmonics up to $l = 10$; (a) constructed from RRDF(XRISM) and (b) constructed from RRDF(MD). The isodensity surface for $n_O(\mathbf{r}) = 1.8$ is shown. This figure was drawn with the help of MOLEKEL package [26].

The three dimensional SDF of $n_o(\mathbf{r}) > 1.80$ reconstructed by RRDFs (XRISM) and RRDFs

(MD) are illustrated in Fig. 1.3(a) and Fig. 1.3(b), respectively. Both SDFs look very similar to each other. All of the lobes correspond to the distribution of the solvent water that forms hydrogen bond with "solute water", as discussed above. This character of the oxygen distribution is essentially the same as those reported by Svishchev *et al*. [10] and Soper *et al* [8]. These results indicate that RRDFs (XRISM) as well as RRDFs (MD) are useful to construct SDFs.

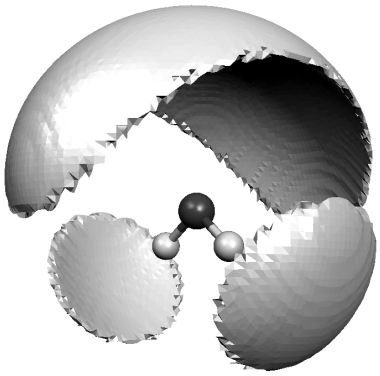


Figure 1.4: The three dimensional SDF map of the oxygen in solvent water constructed by RRDF(XRISM) using the real solid harmonics up to 2. The isodensity surface and the package used are the same as in Fig. 1.3.

In order to check how much the SDFs depend on real solid harmonic expansion, we evaluated the SDF of solvent oxygen site from RRDF(XRISM) using the real solid harmonics up to $l = 2$. The reconstructed 3D SDF $n_o(\mathbf{r})$ is shown as an example in Fig. 1.4. Although edge of the lobe is ambiguous because of the insufficient azimuthal accuracy, the shape and the position of the lobes resemble well those of Fig. 1.3 calculated up to $l = 10$. The XRISM calculation of water followed by reconstruction of SDF with $l = 2$ is performed in a few minutes in a personal computer [27]. Therefore, the present method to reconstruct SDFs from RDFs is very powerful for investigation of solvation structures.

1.4 Conclusion

A new method to reconstruct SDFs from RDFs is presented here. This method was successfully applied to liquid structure of water. In this method, the different expansion from that

by Soper *et al*. leads to the linearized equation with which we can easily obtain approximated SDFs. Our method presents reliable results using a small number of real solid harmonics. This means that the present method can be easily applied to large molecular system.

RISM-SCF can evaluate the solvent structure such as RDFs even for chemical reactions in reasonable computational time [28, 29]. The combination of the present method with the RISM-SCF is one of the powerful methods to evaluate 3D picture of solvation structure.

We will compare this approximated SDFs with SDFs that are directly calculated by MD method in forthcoming full article.

Bibliography

[1] J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 2nd ed. (Academic, London, 1986).

[2] C. Nieto-Draghi, J. B. Avalos, and B. Rousseau, J. Chem. Phys. **118**, 7954 (2003).

[3] H. Sato and F. Hirata, Bull. Chem. Soc. Jpn. **74**, 1831 (2001).

[4] I. M. Svishchev and A. Yu. Zassetsky, J. Chem. Phys. **113**, 7432 (2000).

[5] H. Sato and F. Hirata, J. Chem. Phys. **111**, 8545 (1999), and references therein.

[6] A. De Santis and D. Rocca, J. Chem. Phys. **107**, 9559 (1997).

[7] A. K. Soper, Physica B **213-214**, 448 (1995).

[8] A. K. Soper, J. Chem. Phys. **101**, 6888 (1994).

[9] A. K. Soper, C. Andreani, and M. Nardone, Phys. Rev. E **47**, 2598 (1993).

[10] I. M. Svishchev and P. G. Kusalik, J. Chem. Phys. **99**, 3049 (1993).

[11] B. Ensing, F. Buda, M. C. M. Gribnau, and E. J. Baerends, J. Am. Chem. Soc. **126**, 4355 (2004).

[12] M. Ikeguchi and J. Doi, J. Chem. Phys. **103**, 5011 (1995).

[13] D. Beglov and B. Roux, J. Phys. Chem. B **101**, 7821 (1997).

[14] M. Lombardero, C. Martín, S. Jorge, F. Lado, and E. Lomba, J. Chem. Phys. **110**, 1148 (1999).

[15] J. Richardi, C. Millot, and P. H. Fries, J. Chem. Phys. **110**, 1138 (1999).

[16] J. Richardi, P. H. Fries, R.Fischer, S. Rast, and H. Krienke, Mol. Phys. **93**, 925 (1998).

[17] M. Alvarez, F. Lado, E. Lomba, M. Lombardero, and C. Martín, J. Chem. Phys. **107**, 4642 (1997).

[18] P. H. Fries, J. Richardi, and H. Krienke, Mol. Phys. **90**, 841 (1997).

[19] F. Lado, E. Lomba, and M. Lombardero, J. Chem. Phys. **103**, 481 (1995).

[20] P. H. Fries, W. Kunz, P. Calmettes, and P. Turq, J. Chem. Phys. **101**, 554 (1994).

[21] P. H. Fries and G. N. Patey, J. Chem. Phys. **82**, 429 (1985).

[22] $\{C_{0,0}^{\eta,s,i}\}$ are angular scaling factors of $f_i^{\eta,s}$. The 'solvent shell' of RDF is represented by $f_i^{\eta,s} \times C_{0,0}^{\eta,s,i}$. This means that $C_{0,0}^{\eta,s,i}$ value is arbitrarily taken and we fixed $\{C_{0,0}^{\eta,s,i}\}$ as 1 in this work.

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

[24] F. Hirata, B. M. Pettitt, and P. J. Rossky, J. Chem. Phys. **77**, 509 (1982).

[25] H. J. C. Berendsen, J. P. M. Postma, W. F. von Gunstern, and J. Hermas, in *Intermolecular Forces* , edited by B. Pullman (Reidel, Dordrecht, 1981).

[26] P. Flükiger, H. P. Lüthi, S. Portmann, and J. Weber, *MOLEKEL 4.0* (Swiss Center for Scientific Computing, Manno (Switzerland), 2000).

[27] About 40 Mbytes of memory is necessary. Practically, $N_p = 50^3$ is sufficient and the computational time in this condition is 4.3 sec and 1 hours for l=2 and l=10, respectively (a single CPU (PentiumIV EE) at 3.20GHz). This computational time depends strongly on N_p and l (when $N_p = 100^3$ is used, 32 sec and 8 hours for l=2 and l=10, respectively). Our calculation of SDF by MD, which was performed with 1×10^6 steps suggested by reference [**?**], takes about 10 hour.

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

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

Chapter 2

New evaluation of reconstructed spatial distribution function from radial distribution functions

2.1 Introduction

Local solvation structure such as hydrogen bonding has great influence on the stabilities of compounds and their electronic structures [1]. To investigate such solvation structures, three-dimensional (3D) solvation structure is very useful.

In this regard, 3D solvation structure has been studied by theoretical methods. For instance, 3D structures of pure solvent [2, 3] and binary solvent [4] were calculated directly with molecular dynamics method (MD). Also, the integral equation theory (IET) [5], such as molecular Ornstein-Zernike (MOZ) [6, 7] and 3D reference interaction site model (3D-RISM) [8–10], were proposed to provide the 3D structures.

There is another approach to obtain 3D solvation structure [11–15]. The strategy is to reconstruct 3D solvation structure from a set of one-dimensional (1D) solvation structures such as radial distribution functions (RDFs). Because the 1D solvation structure that is averaged over molecular orientations can be easily presented experimentally and theoretically, the method to reconstruct 3D solvation structure from 1D data is very powerful to analyze solvation structure. Actually, Soper *et al.* expanded the angular pair-correlation functions with spherical harmonics and determined the coefficients with the maximum-entropy method [11–13]. Sato *et al*. proposed most plausible solvation structure (MPSS) method [14], in which they determined the MPSS from a set of RDFs and also represented thermal fluctuation around MPSS

with Gaussian functions. Recently, we have proposed a new method to obtain reconstructed spatial distribution function (RC-SDF) from RDFs [15]. In the method, we expanded spatial distribution function (SDF) using real solid harmonics and Gaussian functions on each solute site. This method was successfully applied to water, in which clear picture of the 3D solvation structure was presented. In the previous work, a few Gaussian functions were employed to represent the radial part so as to reproduce the reference RDFs of water. This strategy is very useful when limited number of Gaussian functions are employed; for example, the RDF of water was reproduced well, as reported. However we need to increase the number of Gaussian functions systematically in order to improve the quality of RC-SDF when the solvation structure is not simple. In our previous method [15], the computational time becomes very long as a number of Gaussian functions increases.

To overcome this weakness, we refined our previous method to adopt a new type of basis sets. We applied the new method to analyze the liquid structures of methanol and DMSO. We selected these solvents as examples, because it is said that the former solvation structure is well-defined and the latter one is broad. The results presented here are discussed in comparison with SDF directly calculated by the MD method.

2.2 Method

SDF of solvent site s, $n_s(\mathbf{r})$, can be expanded at each solute site η by using the real solid harmonics $\{Y_{nm}(\theta,\phi)\}$, as follows,

$$n_s(\mathbf{r})|_\eta \;=\; \sum_{n=0}^{\infty}\sum_{m=-n}^{n} N_{nm}^{\eta,s}(|\mathbf{r}-\mathbf{Q}_\eta|)Y_{nm}(\theta,\phi), \tag{2.1}$$

where $N_{nm}^{\eta,s}(|\mathbf{r}-\mathbf{Q}_\eta|)$ is the radial function around the η site. This function can be expanded with one-dimensional functions $\{f_i^{\eta,s}\}$ and $f^{\eta,s}$,

$$N_{nm}^{\eta,s}(|\mathbf{r}-\mathbf{Q}_\eta|) \simeq \sum_{i}^{N} C_{i,n,m}^{\eta,s} f_i^{\eta,s}(|\mathbf{r}-\mathbf{Q}_\eta|)(1-\delta_{n0}\delta_{m0}) + \sqrt{4\pi} f^{\eta,s}(|\mathbf{r}-\mathbf{Q}_\eta|)\delta_{n0}\delta_{m0} \tag{2.2}$$

where $\{C_{i,n,m}^{\eta,s}\}$ are the coefficients to be determined, $\{f_i^{\eta,s}\}$ are spatially well localized functions which reproduce anisotropic solvent structure ($n, m \neq 0$), and $f^{\eta,s}$ is the function which

represents isotropic one ($n, m = 0$). Insertion of eq. 2.2 into eq. 2.1 gives the following equation,

$$n_s(\mathbf{r})|_\eta = \sum_i^N f_i^{\eta,s}(|\mathbf{r}-\mathbf{Q}_\eta|)\sum_{n=1}^{\infty}\sum_{m=-n}^{n} C_{i,n,m}^{\eta,s} Y_{nm}(\theta,\phi) + f^{\eta,s}(|\mathbf{r}-\mathbf{Q}_\eta|). \quad (2.3)$$

These coefficients $\{C_{i,n,m}^{\eta,s}\}$ are determined under the following conditions. By integrating angular part of $n_s(\mathbf{r})|_\eta$ around the η site, the RDF must be reproduced (**condition 1**). The sets of $\{n_s(\mathbf{r})|_\eta\}$ at each solute site must be consistent with each other in 3D space (**condition 2**). $n_s(\mathbf{r})|_\eta$ must be positive in 3D space (**condition 3**).

The integration of $n_s(\mathbf{r})|_\eta$ around the η site leads to the following equation,

$$\begin{aligned} & \frac{1}{4\pi}\int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta d\theta n_s(\mathbf{r})|_\eta \\ = & \sum_i^N f_i^{\eta,s}(|\mathbf{r}-\mathbf{Q}_\eta|)\sum_{n=1}^{\infty}\sum_{m=-n}^{m=n} \frac{C_{i,n,m}^{\eta,s}}{4\pi}\int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta d\theta Y_{n,m}(\theta,\phi) + f^{\eta,s}(|\mathbf{r}-\mathbf{Q}_\eta|) \\ = & f^{\eta,s}(|\mathbf{r}-\mathbf{Q}_\eta|), \end{aligned} \quad (2.4)$$

where we used the orthogonality condition of real solid harmonics. If $f^{\eta,s}$ is the RDF between the solvent s site and the solute η site, ($f^{\eta,s}(|\mathbf{r}-\mathbf{Q}_\eta|) \equiv g^{\eta,s}(|\mathbf{r}-\mathbf{Q}_\eta|)$), **condition 1** is satisfied.

To satisfy **condition 2**, we determined the coefficients so as to minimize the following Γ' value [15],

$$\Gamma' = \frac{1}{2}\sum_k^{N_p}\sum_{\eta\neq\gamma}^{N_a}(n_s(\mathbf{r}_k)|_\eta - n_s(\mathbf{r}_k)|_\gamma)^2, \quad (2.5)$$

where N_p is the number of grid point around solute site and N_a is the number of solute sites.

To keep $n_s(\mathbf{r})$ positive, we adopted the following strategy. In the calculation of real solid harmonics, we divided a sphere into small regions $\Delta_k (k = 1 \sim M)$ using a reduced grid system [16] (Fig. 2.1) and approximated the real solid harmonics by eq. 2.6,

$$Y_{n,m}(\theta,\phi) \simeq \sum_k U_{k,n,m}\delta_k(\theta,\phi) \quad , \quad (2.6)$$

where the function $\delta_k(\theta,\phi)$ is defined as,

$$\delta_k(\theta,\phi) = \begin{cases} 1 & \text{when } (\theta,\phi) \in \Delta_k \\ 0 & \text{when } (\theta,\phi) \notin \Delta_k \end{cases} . \quad (2.7)$$

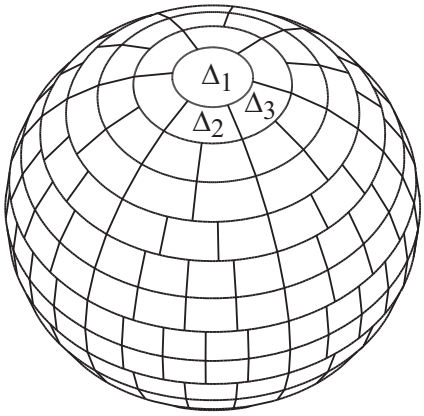


Figure 2.1: Reduced grid sphere in the case of M=254. Note we used M=12302 in the present work.

$\{U_{k,n,m}\}$ are the matrix elements which can be calculated using real solid harmonics. In this work, we used the reduced grid system (M=12302). Insertion of eq. 2.6 into eq. 2.1 leads to the following eq. 2.8 for a given $\mathbf{r}_j = (r_j, \theta_j, \phi_j)$,

$$n_s(\mathbf{r}_j)|_\eta \;\; \simeq \;\; \sum_k^M S_{j,k}^{\eta,s} \delta_k(\theta_j, \phi_j) \;\; , \tag{2.8}$$

where $S_{j,k}^{\eta,s}$ is defined by eq. 2.9,

$$S_{j,k}^{\eta,s} = \sum_i^N f_i^{\eta,s}(|\mathbf{r}_j - \mathbf{Q}_\eta|) \sum_{n=1}^\infty \sum_{m=-n}^n C_{i,n,m}^{\eta,s} U_{k,n,m} + \sqrt{4\pi} U_{k,0,0} f^{\eta,s}(|\mathbf{r}_j - \mathbf{Q}_\eta|) \;\; . \tag{2.9}$$

To satisfy **condition 3**, we employed Γ defined by eq. 2.10 instead of Γ' (eq. 2.5);

$$\Gamma = \frac{1}{2} \sum_k^{N_p} \sum_{\eta \neq \gamma}^{N_a} (n_s(\mathbf{r}_k)|_\eta - n_s(\mathbf{r}_k)|_\gamma)^2 + 2 \sum_{j,k,\eta} \phi(S_{j,k}^{\eta,s}), \tag{2.10}$$

where ϕ is penalty function. The value of the penalty function is 0 when $S_{j,k}^{\eta,s}$ is positive but monotonically increases as the value of $|S_{j,k}^{\eta,s}|$ increases. If $S_{j,k}^{\eta,s}$ is positive all over the space, eq. 2.10 reduces to eq. 2.5. We will discuss the penalty function in detail below. By minimizing Γ, we reach eq. 2.11, with which we can obtain the coefficients $\{C_{i,n,m}^{\eta,s}\}$,

$$\sum_{\beta,j,n',m'} A_{(s)\beta,j,n',m'}^{\alpha,i,n,m} C_{j,n',m'}^{\beta,s} = -B_{(s)}^{\alpha,i,n,m} - \sum_{j,k} \frac{\partial \phi(S_{j,k}^{\alpha,s})}{\partial S_{j,k}^{\alpha,s}} f_i^{\alpha,s}(|\mathbf{r}_j - \mathbf{Q}_\alpha|) U_{k,n,m}, \tag{2.11}$$

where $A_{(s)\beta,j,n',m'}^{\alpha,i,n,m}$ and $B_{(s)}^{\alpha,i,n,m}$ are represented by eqs. 2.12 and 2.13, respectively [17],

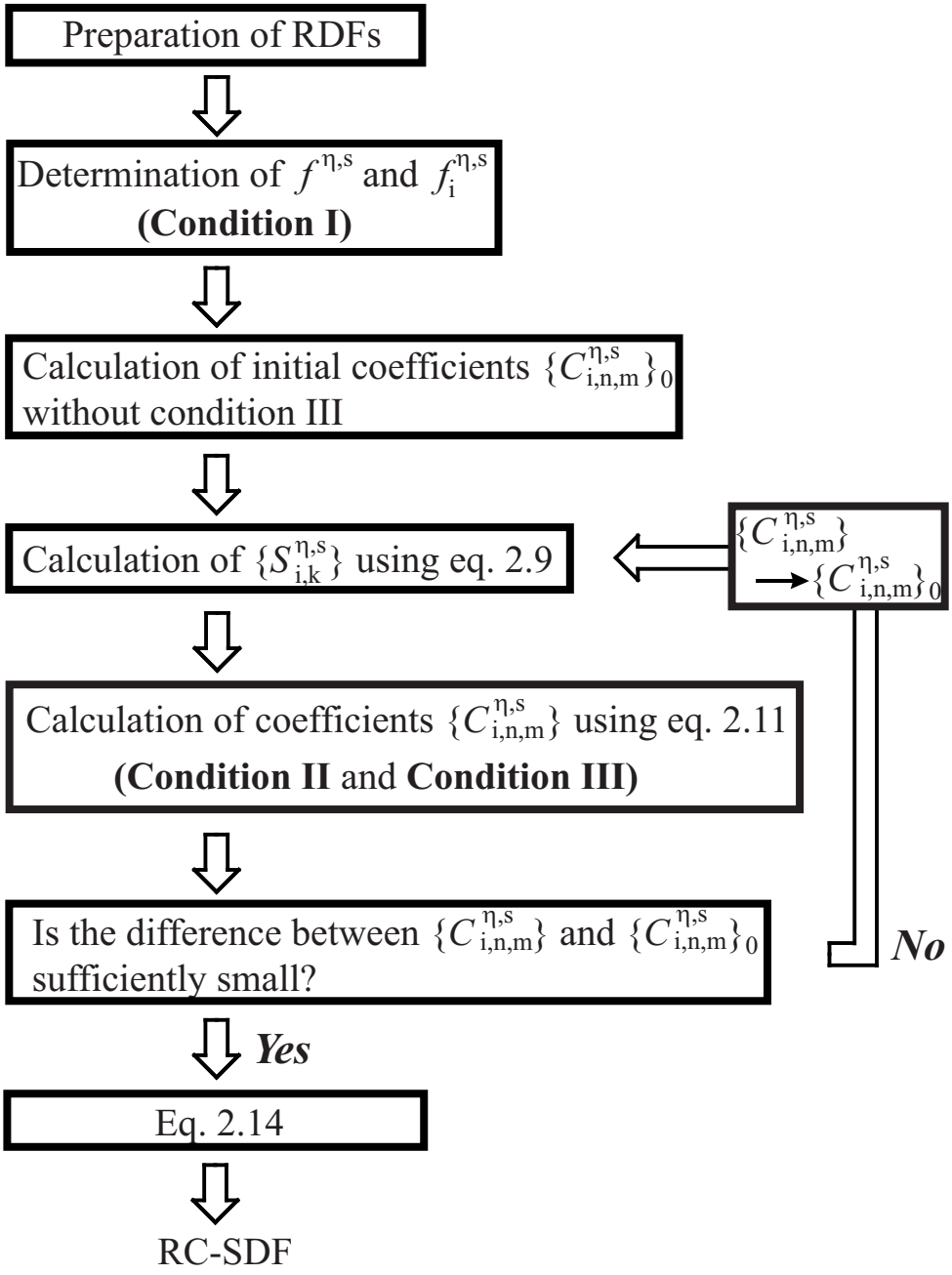
$$\begin{aligned} A_{(s)\beta,j,n',m'}^{\alpha,i,n,m} &= (N_a\delta_{\alpha,\beta}-1)\sum_{k}^{N_p} f_i^{\alpha,s}(|\mathbf{r}_k-\mathbf{Q}_\alpha|)f_j^{\beta,s}(|\mathbf{r}_k-\mathbf{Q}_\beta|) \\ &\quad \times Y_{nm}(\theta_k,\phi_k)Y_{n'm'}(\theta_k,\phi_k) \qquad (2.12) \\ B_{(s)}^{\alpha,i,n,m} &= \sum_{k=1}^{N_p} f_i^{\alpha,s}(|\mathbf{r}_k-\mathbf{Q}_\alpha|)Y_{nm}(\theta_k,\phi_k) \\ &\quad \times \sum_{\gamma}^{N_a}(N_a\delta_{\alpha,\gamma}-1)f^{\gamma,s}(|\mathbf{r}_k-\mathbf{Q}_\gamma|). \qquad (2.13) \end{aligned}$$

In our previous work, **condition III** was not considered because a set of $\{n_s(\mathbf{r})|_\eta\}$ was positive almost all over the space, and the equation to be solved was linear. **Condition III** becomes important as the number of the functions $\{f_i^{\eta,s}\}$ increases. Because of **condition III**, eq. 2.11 is not a linear equation here. Thus, we solved it iteratively. The initial guess is calculated by solving a linear equation (eq. 2.6 in ref [15]) because the results provided correct shape of SDF as shown in our previous work.

The sets of $\{n_s(\mathbf{r})|_\eta\}$ are different from each other at $\mathbf{r}$. To obtain final RC-SDF $n_s(\mathbf{r})$ from these values, we used the following equation with proper weight, $w_\alpha(\mathbf{r})$,

$$\begin{aligned} n_s(\mathbf{r}) &= \sum_{\alpha}^{N_a} w_\alpha(\mathbf{r})n_s(\mathbf{r})|_\alpha \qquad (2.14) \\ \sum_{\alpha}^{N_a} w_\alpha(\mathbf{r}) &= 1\ . \qquad (2.15) \end{aligned}$$

The simple weight, $w_\alpha(\mathbf{r}) = 1/N_a$, was used through our study. Because the quality of $\{n_s(\mathbf{r})|_\eta\}$ far from the origin atom, η becomes worse, better weight function will improve results. This procedure is summarized in Scheme 2.1.


Preparation of RDFs
Determination of $f^{\eta,s}$ and $f_i^{\eta,s}$
(Condition I)
Calculation of initial coefficients $\{C_{i,n,m}^{\eta,s}\}_0$ without condition III
Calculation of $\{S_{i,k}^{\eta,s}\}$ using eq. 2.9
$\{C_{i,n,m}^{\eta,s}\}$
$\rightarrow \{C_{i,n,m}^{\eta,s}\}_0$
Calculation of coefficients $\{C_{i,n,m}^{\eta,s}\}$ using eq. 2.11
(Condition II and Condition III)
Is the difference between $\{C_{i,n,m}^{\eta,s}\}$ and $\{C_{i,n,m}^{\eta,s}\}_0$ sufficiently small?
No
Yes
Eq. 2.14
RC-SDF

Scheme 2.1

2.3 Computational details

In the present method, any kind of functions can be used for $\{f_i^{\eta,s}\}$ in eq. 2.2. In this article, we used roof functions for $\{f_i^{\eta,s}\}$,

$$f_i^{\eta,s}(r) = \begin{cases} 0 & \text{for} \quad 0 \le r \le r_{i-1} \\ \dfrac{r - r_{i-1}}{r_i - r_{i-1}} & \text{for} \quad r_{i-1} \le r \le r_i \\ \dfrac{r_{i+1} - r}{r_{i+1} - r_i} & \text{for} \quad r_i \le r \le r_{i+1} \\ 0 & \text{for} \quad r_{i+1} \le r \end{cases} . \tag{2.16}$$

The following function was used as the penalty function in eq. 2.10.

$$\phi(S) = \begin{cases} \alpha S^4 & (S \le 0) \\ 0 & (S \ge 0) \end{cases} . \tag{2.17}$$

The α value can be determined arbitrary; in this work, we used $\alpha = 15.0$. Although $n_s(\mathbf{r})$ becomes negative with this α in several regions, the negative value is small enough to neglect it (minimum value is -0.11 in this work).

In the calculation of the MD simulation, we use simple rigid potential model for methanol and DMSO with the intermolecular pair-potential. All Lennard-Jones parameters and the fractional charges used for methanol and DMSO are listed in Table 2.1.

Table 2.1: Intermolecular potential parameters

		ϵ/kcal mol^{-1}	σ/Å	charge
Methanol[a]	Oxygen	0.170	3.07	-0.700
	Hydrogen	0.000	0.00	0.435
	Methyl group	0.207	3.775	0.265
DMSO[b]	Sulfur	0.23838	3.40	0.139
	Oxygen	0.07152	2.80	-0.459
	Methyl group	0.29397	3.80	0.160

Molecular geometry: (Methanol) $r_{\mathrm{OH}} = 0.945$Å, $r_{\mathrm{CO}} = 1.430$Å, $\angle\mathrm{COH} = 108.5^\circ$.
(DMSO)[c] $r_{\mathrm{OS}} = 1.53$Å, $r_{\mathrm{SC}} = 1.80$Å, $\angle\mathrm{OSC} = 106.75^\circ$, $\angle\mathrm{CSC} = 97.40^\circ$.
[a] OPLS model [18, 19]. [b] Reference [20]. [c] crystallographic data [21]

MD simulations were carried out within the NVT ensemble at $T = 298.15$ K under 1 atm. In this simulation, the cubic periodic box was filled with 256 molecules. The simulation software used was MOLDY [22]. Temperature was controlled with Nosé-Hoover thermostat [23, 24].

All of 3D figures are drawn with the help of MOLEKEL [25].

2.4 Results and discussion

First, we performed MD calculations of methanol and DMSO to obtain their RDFs. Then, we evaluated RC-SDFs from the RDFs. For the purpose of comparison, we also calculated the SDF from the MD trajectory directly (SDF(MD)).

2.4.1 Methanol

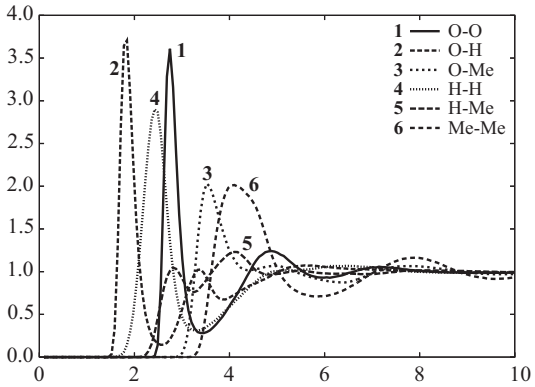
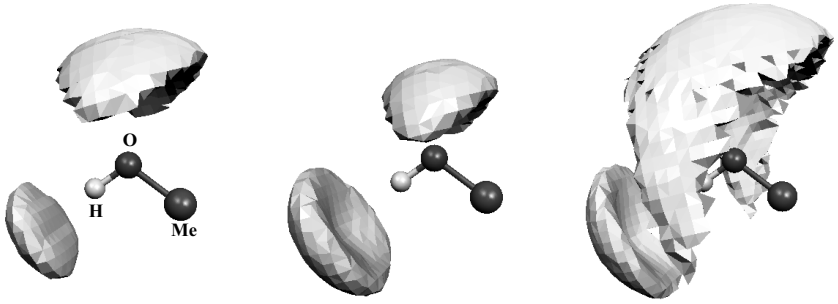


Figure 2.2: Reference RDFs of methanol calculated by MD.

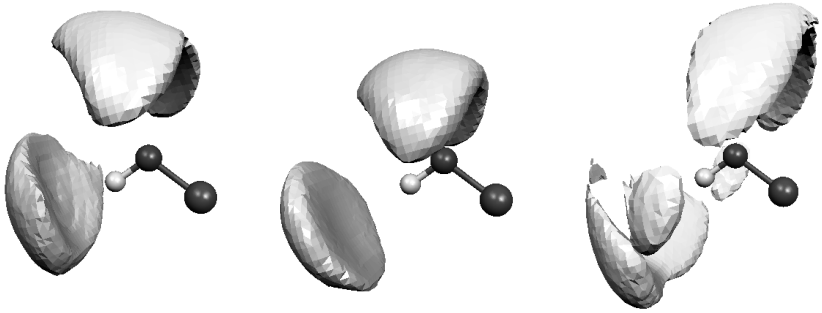
The reference RDFs calculated directly by the MD method are shown in Fig. 2.2. The RDFs between oxygen and oxygen, oxygen and hydrogen, and hydrogen and hydrogen display sharp peaks, while the RDFs between oxygen and methyl group, hydrogen and methyl group, and methyl group and methyl group exhibit broad peaks.

The SDF(MD) and the RC-SDF of oxygen site, hydrogen site, and methyl site are shown in Figs. 2.3 and 2.4, respectively. In SDF(MD), the distribution of oxygen (O) lies nearer to solute hydrogen site than the distribution of hydrogen (H) and lies more separately from oxygen site than the distribution of hydrogen (H). These distributions of SDF(MD) are well reproduced by RC-SDF, as shown in Fig. 2.4. The distribution of methyl group is mainly observed around oxygen and hydrogen sites but it is more delocalized. Both SDF(MD) and RC-SDF present essentially the same distribution of methyl group, while the distribution of RC-SDF is some-



(a) (b) (c)

Figure 2.3: Three dimensional SDFs(MD) map of oxygen (a), of hydrogen (b), and of methyl group (c) in methanol. The isosurfaces of SDFs are drawn at $n(\mathbf{r})$= 3.0.



(a) (b) (c)

Figure 2.4: Three dimensional RC-SDFs map of oxygen (a), of hydrogen (b), and of methyl group (c) in methanol. The isosurfaces are the same as in the Fig. 2.3.

what more localized around solute oxygen site than that of SDF(MD) (compare Fig. 2.3(c) with Fig. 2.4(c)). As shown by these examples, RC-SDF can produce well the distributions of SDF(MD) except for moderately delocalized distribution of methyl group.

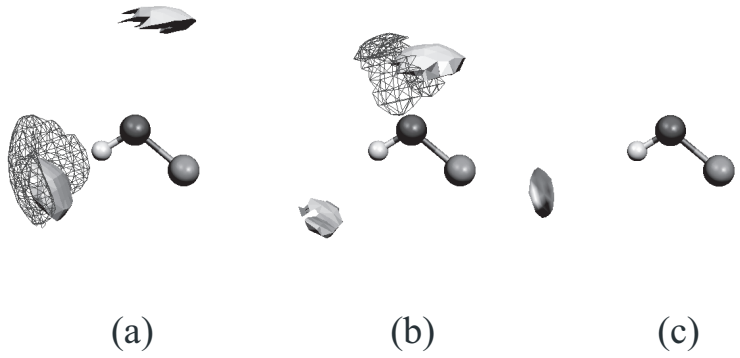


Figure 2.5: Difference maps between RC-SDF and SDF(MD) of oxygen (a), of hydrogen (b), and of methyl group (c) in methanol. The region where the difference ($n_s(\mathbf{r})$ of RC-SDF - $n_s(\mathbf{r})$ of SDF(MD)) is larger than 5.0 is drown with mesh and the region where the difference is less than -5.0 is drown with solid surface.

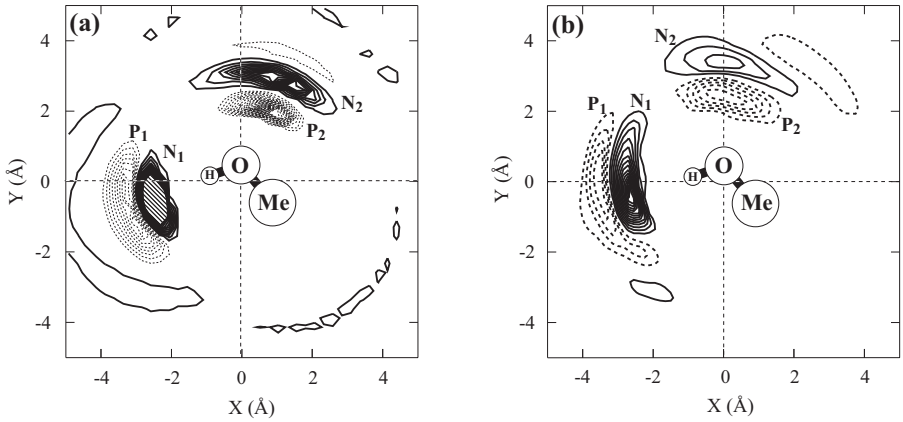


Figure 2.6: Two dimensional map of charge density; (a) The contour of SDF(MD). The interval of contour is 0.010 for positive value and 0.015 for negative value. The region where negative value is less than -0.210 is shaded. (b) The contour of RC-SDF. The interval is the same as in (a).

To make the difference between RC-SDFs and SDFs(MD) clearer, the difference maps ($n_s(\mathbf{r})$ of RC-SDF - $n_s(\mathbf{r})$ of SDF(MD)) are shown in Fig. 2.5. As mentioned above, RC-SDFs present almost same distribution with SDFs(MD).

To analyze the peak positions of these distributions, we show contour maps of charge density calculated by $n_s(\mathbf{r})$ of SDF(MD) and RC-SDF in Fig. 2.6(a) and 2.6(b). The charge density is

represented by eq. 2.18,

$$\rho_q(\mathbf{r}) = \rho \sum_{\gamma}^{N_a} q_\gamma n_\gamma(\mathbf{r}) \ , \tag{2.18}$$

where ρ is the number density of solvent and q_γ is the charge of solvent γ site. In the case of methanol, the negative distribution (solid line) mainly comes from the distribution (O) and the positive one (dashed line) comes from the distributions of hydrogen and methyl group. Because these distributions have little overlap between each other, the positive and negative regions can be recognized as the distributions of hydrogen/methyl group and oxygen group, respectively. By using this charge density maps, three different information can be compiled to one figure. In Fig. 2.6(a), negative distribution is observed about X=-2.34 and Y=-0.75 Å ($\mathbf{N_1}$) and positive distribution ($\mathbf{P_1}$) is outside this negative area. Another positive distribution is observed about X=0.69 and Y=2.08 (Å) ($\mathbf{P_2}$) and negative one ($\mathbf{N_2}$) is outside of this negative area. The considerably localized distributions $\mathbf{N_1}$ and $\mathbf{P_2}$ and their very high peaks indicate that oxygen and hydrogen atoms of solvent form strong bonding interaction with hydrogen and oxygen of solute, respectively. On the other hand, the distributions $\mathbf{N_2}$ and $\mathbf{P_1}$ are very broad. In the case of RC-SDF, the distributions $\mathbf{N_1}$ and $\mathbf{P_2}$ are more localized than the distribution $\mathbf{P_1}$ and $\mathbf{N_2}$, as observed in the case of SDF(MD) (see Figs. 2.6(a) and 2.6(b)). Although the peaks of these distributions are smaller than those of SDF(MD), the peak of RC-SDF are at almost the same position as those of SDF(MD).

To investigate how much the RC-SDFs depend on the order of real solid harmonics used in eq. 2.3, we calculated the distributions of solvent oxygen with real solid harmonics up to order $n = 2$, $n = 6$, and $n = 10$, as shown in Fig. 2.7(a), 2.7(b) and 2.7(c), respectively. [26] In Fig. 2.7(d), SDF(MD) of oxygen is also shown. In Fig. 2.7(a), broad distribution is observed around solute. As n increases, this broad distribution separates into two distributions (Fig. 2.7(b)) concomitantly with increase of the height and the distribution approaches the SDF(MD). The difference maps of RC-SDFs between $n = 2$ and $n = 10$ and between $n = 6$ and $n = 10$ are shown on Fig. 2.8. Although the difference in shape of RC-SDF between $n = 2$ and RC-SDF $n = 6$ is considerably large (Fig. 2.8(a)), the difference between $n = 6$ and $n = 10$

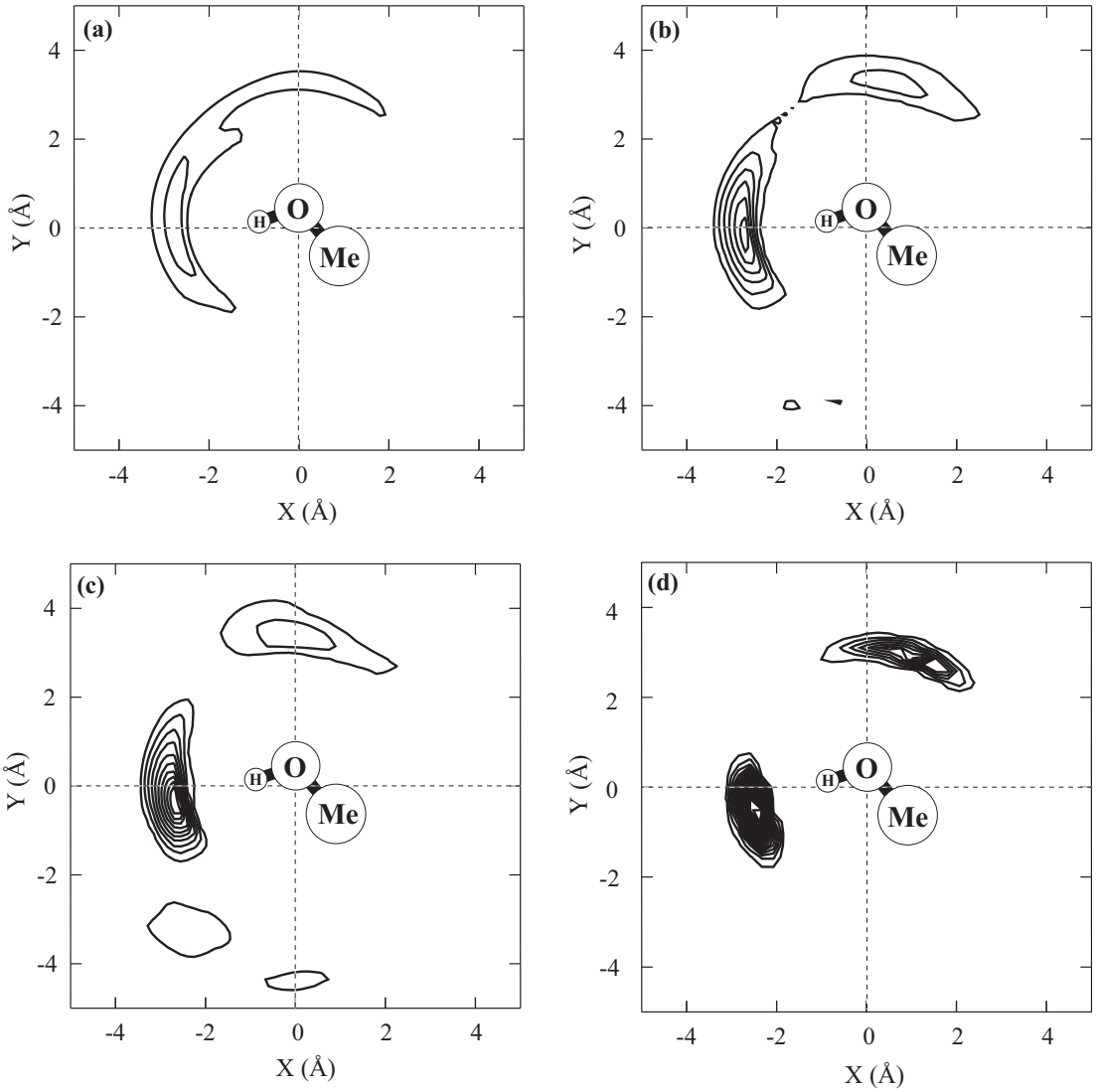


Figure 2.7: Two dimensional RC-SDF map of oxygen in methanol with different angular momentum. The interval of contour value is 2. (a) $n = 2$, (b) $n = 6$, (c) $n = 10$, and (d) the result obtained by MD.

becomes very small, as shown in Fig. 2.8(b). This result indicates that the increase in the order n improve the shape of SDF.

(a) (b)

Figure 2.8: Difference maps of RC-SDF (oxygen) between n=2 and n=10 (a) and n=6 and n=10 (b). The region where the difference is larger than 5.0 is shown.

The computational time of these calculations is very short [27]; for instance, the calculation of RC-SDF with real solid harmonics ($n = 10$) takes about 5 minutes for the distribution of oxygen, about 8 minutes for that of hydrogen, and about 4 minutes for that of methyl group. Although the basis functions and grid size are different from those employed in our previous work [15], the computational time considerably decreases by using roof functions and discrete real solid harmonics [28].

2.4.2 DMSO

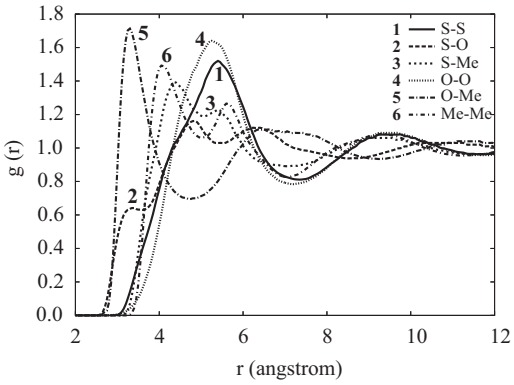


Figure 2.9: Reference RDFs of DMSO calculated by the MD method.

In the case of DMSO, several broad peaks are observed in the reference RDFs (Fig. 2.9), unlike methanol in which the reference RDFs exhibit sharp peaks attributed to hydrogen bonding. These RDFs clearly indicate that the solvation structure becomes more complexed than that of methanol.

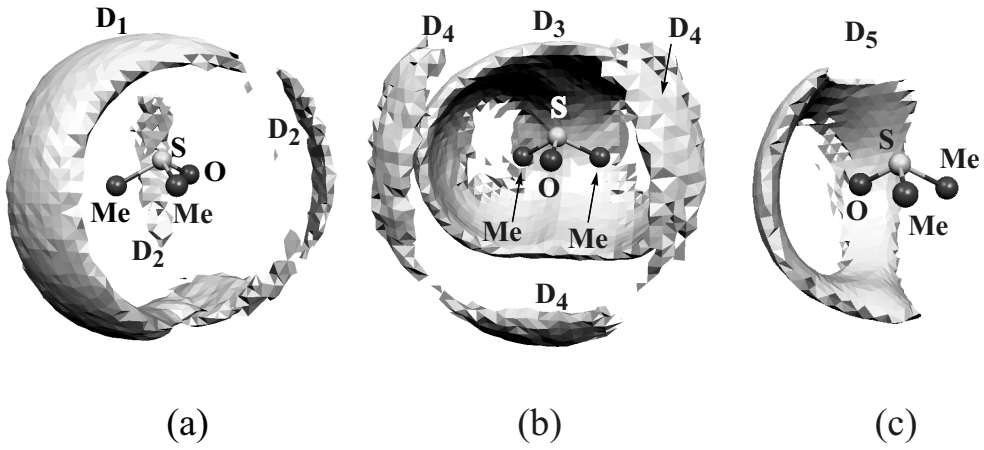


Figure 2.10: Three dimensional SDFs(MD) map of sulfur (a), of oxygen (b), and of methyl group (c). The isosurfaces of SDFs are drawn at $n(\mathbf{r})$= 2.0 for (a) and (b), and 2.9 for (c).

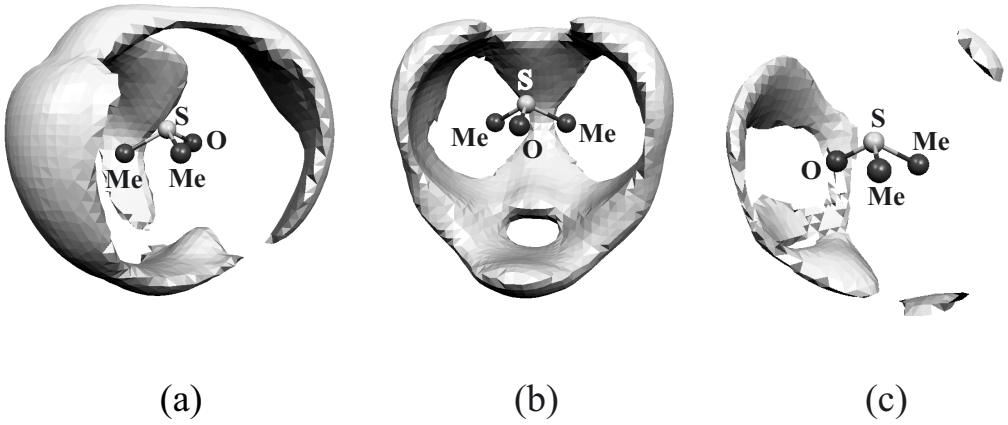


Figure 2.11: Three dimensional RC-SDFs(MD) map of sulfur (a), of oxygen (b), and of methyl group (c). The isosurfaces are the same as in the Fig. 2.10.

The SDFs(MD) and RC-SDFs are shown in Fig. 2.10 and 2.11. The solvation structure by SDFs(MD) (Fig. 2.10) is similar to that of DMSO-water mixture reported by Vishnyakov *et al* [4]. SDF(MD) of sulfur displays broad distribution near solute methyl group ($\mathbf{D_1}$) and very

broad distribution distant from solute ($\mathbf{D_2}$) (see Fig. 2.10(a)). These distributions can be reproduced well by RC-SDF, as shown in Fig. 2.11(a). SDF(MD) of oxygen, on the other hand, presents relatively localized distribution around solute methyl group ($\mathbf{D_3}$) and broad distribution ($\mathbf{D_4}$) distant from solute oxygen site. The relatively localized distribution is attributed to moderately strong electrostatic interaction between oxygen and methyl group. RC-SDF can reproduce well this localized distribution ($\mathbf{D_3}$). However, the broad distribution ($\mathbf{D_4}$) of oxygen is moderately different between SDF(MD) and RC-SDF. The broad distribution of solvent methyl group is presented around solute by SDF(MD), as shown in Fig. 2.10(c). RC-SDF can present well this broad one as shown in Fig. 2.11(c).

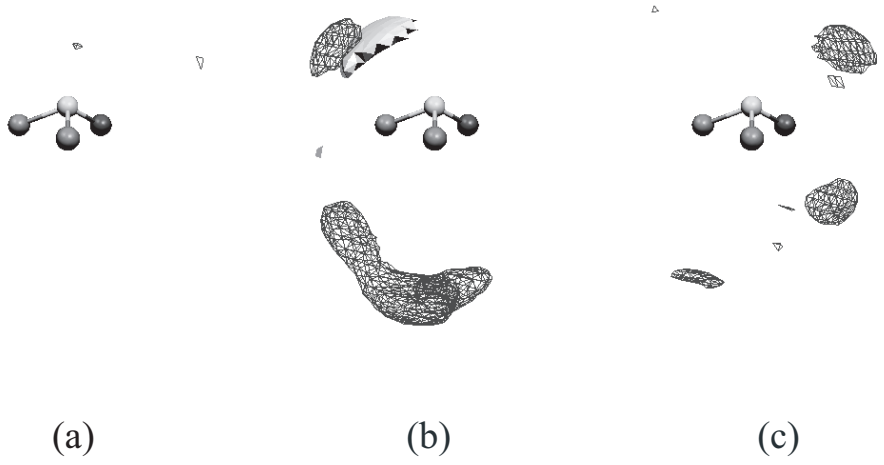


Figure 2.12: Difference maps between RC-SDF and SDF(MD) of sulfur (a), of oxygen (b), and of methyl group (c) in DMSO. The region where the difference ($n_s(\mathbf{r})$ of RC-SDF - $n_s(\mathbf{r})$ of SDF(MD)) is larger than 2.5 is drown with mesh and the region where the difference is less than -2.5 is drown with solid surface.

The difference map ($n_s(\mathbf{r})$ of RC-SDF - $n_s(\mathbf{r})$ of SDF(MD)) are shown in Fig. 2.12. Although broad distributions can not be reproduced by RC-SDF, the important distributions, such as one between solute oxygen and solute methyl groups, are reproduced very well, as mentioned above.

2.5 Conclusions

In our previous work [15], we proposed a new method to calculate RC-SDF from RDFs. In this work, we successfully refined the method by using roof functions and discrete real solid harmonics. Here, RC-SDF is computed under the three conditions described below. By integrating $n_s(\mathbf{r})|_\eta$ around the solute site, the RDF must be reproduced (**condition 1**). The sets of $\{n_s(\mathbf{r})|_\eta\}$ at each solute site must be consistent with each other in 3D space around solute (**condition 2**). $n_s(\mathbf{r})|_\eta$ must be positive in 3D space (**condition 3**). Although the equation to be solved is not linear because of **condition 3**, RC-SDF can be obtained iteratively with reasonable computational time. The obtained RC-SDF can produce not only well-defined 3D solvation structure of methanol but also diffuse one of DMSO.

This method can present RC-SDF from any kind of RDFs with reasonable computational cost. In other words, this method can be combined with methods which present RDFs, such as neutron scattering and RISM-SCF [29, 30]. Combination of these methods provides much clearer understanding of solvation event than the usual RDFs.

Bibliography

[1] J. R. Asher, N. L. Doltsinis, and M. Kaupp, J. Am. Chem. Soc. **126**, 9854 (2004).

[2] I. M. Svishchev and P. G. Kusalik, J. Chem. Phys. **99**, 3049 (1993).

[3] I. M. Svishchev and A. Yu. Zassetsky, J. Chem. Phys. **113**, 7432 (2000).

[4] A. Vishnyakov, A. P. Lyubartsev, and A. Laaksonen, J. Phys. Chem. A **105**, 1702 (2001).

[5] J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids,* 2nd ed. (Academic, London, 1986).

[6] P. H. Fries and G. N. Patey, J. Chem. Phys. **82**, 429 (1985).

[7] M. Lombardero, C. Martín, S. Jorge, F. Lado, and E. Lomba, J. Chem. Phys. **110**, 1148 (1999).

[8] M. Ikeguchi and J. Doi, J. Chem. Phys. **103**, 5011 (1995).

[9] D. Beglov and B. Roux, J. Phys. Chem. B **101**, 7821 (1997).

[10] V. Shapovalov, T. N. Truong, A. Kovalenko, and F. Hirata, Chem. Phys. Lett. **320**, 186 (2000).

[11] A. K. Soper, Physica B **213-214**, 448 (1995).

[12] A. K. Soper, J. Chem. Phys. **101**, 6888 (1994).

[13] A. K. Soper, C. Andreani, and M. Nardone, Phys. Rev. E **47**, 2598 (1993).

[14] H. Sato and F. Hirata, Bull. Chem. Soc. Jpn. **74**, 1831 (2001).

[15] D. Yokogawa, H. Sato, and S. Sakaki, J. Chem. Phys. **123**, 211102 (2005).

[16] K. Takahashi, X. Peng, K. Komine, M. Ohdaira, Y. Abe, T. Sugimura, K. Goto, H. Fuchigami, M. Yamada, and K. Watanabe, Development of Non-hydrostatic Coupled Ocean-Atmosphere Simulation Code on the Earth Simulator,, in: *Proc. 7th International Conference on High Performance Computing and Grid in Asia Pacific Region*, (2004).

[17] Unfortunately, there is a mistype in ref. [15]. The correct element of the matrix $\mathbf{B}$ (eq. 7b) is,

$$B_{(s)}^{\alpha,i,l,m} = \sum_{i}^{N_p} f_i^{\alpha,s}(|\mathbf{r}-\mathbf{Q}_\alpha|) S_{l,m}(\theta,\phi) \sum_{\gamma} (N_a \delta_{\alpha,\gamma} - 1) \times (\sum_{i}^{N} f_i^{\gamma,s}(|\mathbf{r}-\mathbf{Q}_\gamma|) C_{0,0}^{\gamma,s,i} + f_f^{\gamma,s}(|\mathbf{r}-\mathbf{Q}_\gamma|)).$$

In this work, $\sum f_i^{\gamma,s}(|\mathbf{r}-\mathbf{Q}_\gamma|) C_{0,0}^{\gamma,s,i} + f_f^{\gamma,s}(|\mathbf{r}-\mathbf{Q}_\gamma|)$ in the previous work is represented as the function $f^{\gamma,s}(|\mathbf{r}-\mathbf{Q}_\gamma|)$.

[18] W. L. Jorgensen, J. Phys. Chem. **90**, 1276 (1986).

[19] W. L. Jorgensen, J. D. Madura, and C. J. Swenson, J. Am. Chem. Soc. **106**, 6638 (1984).

[20] A. Luzar and D. Chandler, J. Chem. Phys. **98**, 8160 (1993).

[21] R. Thomas, C. B. Shoemaker, and K. Eriks, Acta. Cryst. **21**, 12 (1966).

[22] K. Refson, Comput. Phys. Commun. **126**, 310 (2000).

[23] S. Nosé, J. Chem. Phys. **81**, 511 (1984).

[24] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).

[25] P. Flükiger, H. P. Lüthi, S. Portmann, and J. Weber, *MOLEKEL 4.0* (Swiss Center for Scientific Computing, Manno (Switzerland), 2000).

[26] The total numbers of real solid hamonics are 8 (n=2), 48 (n=6), and 120 (n=10).

[27] All of the calculation were performed with a single CPU (Pentium IV EE) at 3.20 GHz. In the calculation of DMSO, our program requires about 120 Mbytes of memory.

[28] In this work, the number of the coefficients of expansion is 1560, 1920, and 2040 for RC-SDF of oxygen, hydrogen, and methyl group, respectively. The number of grid points is 60^3 for all of the RC-SDFs. In the previous work, the number of the coefficients of expansion and grid points in water is 1080 for RC-SDF of oxygen and 50^3, respectively. The calculation of water is required about 1 hour with the same mathine.

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

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

Chapter 3

An integral equation theory for 3D solvation structure: A new procedure free from 3D Fourier transform.

3.1 Introduction

Solvation has been studied in full detail by experiment and theory because of its great importance in chemistry. In theoretical studies, integral equation theory (IET) for liquids is expected to be a powerful tool for evaluation of micro properties related to solvation structure and thermodynamic properties.

One of the most important and popular IET is RISM [1]. This method, as modified by Hirata et. al. (XRISM), has been applied to many molecular liquid systems [2, 3]. Because of the simplicity of the equation and various range of its applicability, RISM has been combined with other theoretical methods such as quantum method (RISM-SCF) [4, 5], Monte Carlo [6], and solvation structure analysis [7–9].

Three-dimensional Ornstein-Zernike (3D-OZ) [10], 3D-RISM [11, 12] and MSOZ [13] can directly evaluate 3D solvation structure around a solute molecule. Although these methods can provide more accurate local solvation structure than RISM, expensive 3D Fourier transforms are necessary in the calculation. Ten-no et. al. proposed another IET named partial wave (PW) equation theory, which also presents 3D information [14–16]. Recently, Sumi et. al. evaluated molecular orientation using density functional theory (DFT) [17].

In this work, we propose a new procedure to evaluate 3D-correlation functions with real solid harmonics around solute site. Since our procedure employs radial and angular grid struc-

tures instead of a 3D grid, expensive 3D Fourier transforms can be avoided in the calculation of 3D convolution by employing spherical Bessel transforms. In this procedure, a new closure equation based on 3D-HNC closure and fussy cell method proposed by Becke [18] are employed. By decomposition of the 3D correlation function, the site-site interaction between solute and solvent can be evaluated more easily. Such information is useful for clearer understanding of solvation. We have applied this procedure to a typical benchmark system, non-charged/charged HCl model in the present letter. The detail of our procedure is given in Sec. 3.2. Special techniques necessary for employing the procedure are described in Sec. 3.3. The results and conclusions are presented in Secs. 3.4 and 3.5, respectively.

3.2 Method

3.2.1 Ornstein-Zernike type equation

In the 3D-RISM integral equation for a solute-solvent system, total and direct correlation functions in 3D space, H_γ and C_γ, are written as follows [10–12];

$$H_\gamma(\mathbf{r}) = \sum_\delta C_\delta * (\omega_{\delta\gamma} + \rho h_{\delta\gamma})(\mathbf{r}), \tag{3.1}$$

where $\omega_{\delta\gamma}$ is the intra-molecular correlation function between solvent molecule sites, δ and γ, ρ is the number density of solvent molecules, and $*$ denotes convolution in direct space. $\{h_{\delta\gamma}\}$ are the radial site-site correlation functions of bulk solvent and these are usually evaluated with the XRISM theory.

To consider 3D site-site interaction explicitly, we divide the three dimensional correlation functions into the components assigned to each solute site. This division is performed by the following function, in which modified Voronoi cells are employed,

$$\sum_\alpha w_\alpha(\mathbf{r}) = 1. \tag{3.2}$$

This function, proposed by Becke [18], is commonly used for the calculation based on DFT in many ab initio quantum chemistry packages.

The direct correlation function, C_δ, is formally divided into the components around the solute site, α, by the function,

$$\begin{aligned} C_\delta(\mathbf{r}) &= \sum_\alpha w_\alpha(\mathbf{r}) C_\delta(\mathbf{r}) \\ &= \sum_\alpha c_\delta^{(\alpha)}(\mathbf{r}), \end{aligned} \tag{3.3}$$

where superscript α is used to show that $c_{\alpha\delta}^{(\alpha)}(\mathbf{r})$ is a partial component of direct correlation function. It is convenient to define an auxiliary function, $c_{\alpha\delta}^{(\alpha)}(\mathbf{r})$, in which the position vectors are referred to the atomic sites;

$$c_{\alpha\delta}^{(\alpha)}(\mathbf{r}) \equiv c_\delta^{(\alpha)}(\mathbf{r} + \mathbf{R}_\alpha), \forall\alpha. \tag{3.4}$$

where $\mathbf{R}_\alpha$ is the position vector pointing to solute site α. The auxiliary function, $c_{\alpha\delta}^{(\alpha)}(\mathbf{r})$, can be expanded with real-solid harmonics around solute site, α, as follows:

$$c_{\alpha\delta}^{(\alpha)}(\mathbf{r}) = \sum_{l'm'} c_{\alpha\delta,l'm'}^{(\alpha)}(r_\alpha) S_{l'm'}(\hat{r}_\alpha). \tag{3.5}$$

By substituting Eq. 3.3 into Eq. 3.1 and using the notation in Eq. 3.4, the total correlation function is represented with the auxiliary function $c_{\alpha\delta}^{(\alpha)}(\mathbf{r})$:

$$H_\gamma(\mathbf{r}) = \sum_{\alpha\delta} c_{\alpha\delta}^{(\alpha)} * (\omega_{\delta\gamma} + \rho h_{\delta\gamma})(\mathbf{r}). \tag{3.6}$$

Equation 3.6 can be represented by the solute-site centered components, $\{h_{\alpha\gamma,l'm'}^{(\alpha)}\}$, and real-solid harmonics, $\{S_{l'm'}\}$, by using Eq. 3.5, as follows:

$$\begin{aligned} H_\gamma(\mathbf{r}) &= \sum_{\alpha\delta}\sum_{l'm'} \left[c_{\alpha\delta,l'm'}^{(\alpha)} * (\omega_{\delta\gamma} + \rho h_{\delta\gamma}) \right](r_\alpha) S_{l'm'}(\hat{r}_\alpha) \\ &\equiv \sum_{\alpha}\sum_{l'm'} h_{\alpha\gamma,l'm'}^{(\alpha)}(r_\alpha) S_{l'm'}(\hat{r}_\alpha). \end{aligned} \tag{3.7}$$

The auxiliary total correlation function, $H_{\eta\gamma}(\mathbf{r})$, is defined as follows:

$$H_{\eta\gamma}(\mathbf{r}) \equiv H_\gamma(\mathbf{r} + \mathbf{R}_\eta), \forall\eta. \tag{3.8}$$

The auxiliary function can also be expanded with real solid harmonics:

$$H_{\eta\gamma}(\mathbf{r}) = \sum_{l'm'} h_{\eta\gamma,l'm'}(r_\eta) S_{l'm'}(\hat{r}_\eta). \tag{3.9}$$

By following Talman's strategy [19], the components, $\{h^{(\alpha)}_{\alpha\gamma,l'm'}\}$, on each solute center can be connected with $\{h_{\eta\gamma,l'm'}\}$, as follows:

$$h_{\eta\gamma,l'm'}(r_\eta) = \sum_\alpha \sum_{lm} \left[\omega_{lm,l'm'}(\mathbf{R}_{\eta\alpha}) * h^{(\alpha)}_{\alpha\gamma,lm}\right](r_\eta), \tag{3.10}$$

where a new intra-molecular correlation function $\omega_{lm,l'm'}(\mathbf{R}_{\eta\alpha})$ is defined in the k-space, as follows:

$$\omega_{lm,l'm'}(k,\mathbf{R}) = (-1)^l i^{l'+l} \int \mathrm{e}^{i\mathbf{k}\cdot\mathbf{R}} S_{lm}(\hat{k}) S_{l'm'}(\hat{k}) \mathrm{d}\mathbf{\Omega}_k. \tag{3.11}$$

By substituting the components, $\{h^{(\alpha)}_{\alpha\gamma,l'm'}\}$, defined in Eq. 3.7 into Eq. 3.10, OZ-type equation is obtained as follows:

$$h_{\eta\gamma,l'm'}(r_\eta) = \sum_{lm} \sum_{\alpha\beta} \left[\omega_{lm,l'm'}(\mathbf{R}_{\eta\alpha}) * c^{(\alpha)}_{\alpha\beta,lm} * (\omega_{\beta\gamma} + \rho h_{\beta\gamma})\right](r_\eta). \tag{3.12}$$

This equation is reduced to the RISM equation by taking a spherical limit ($l' = 0$, $m' = 0$) of Eq. 3.12.

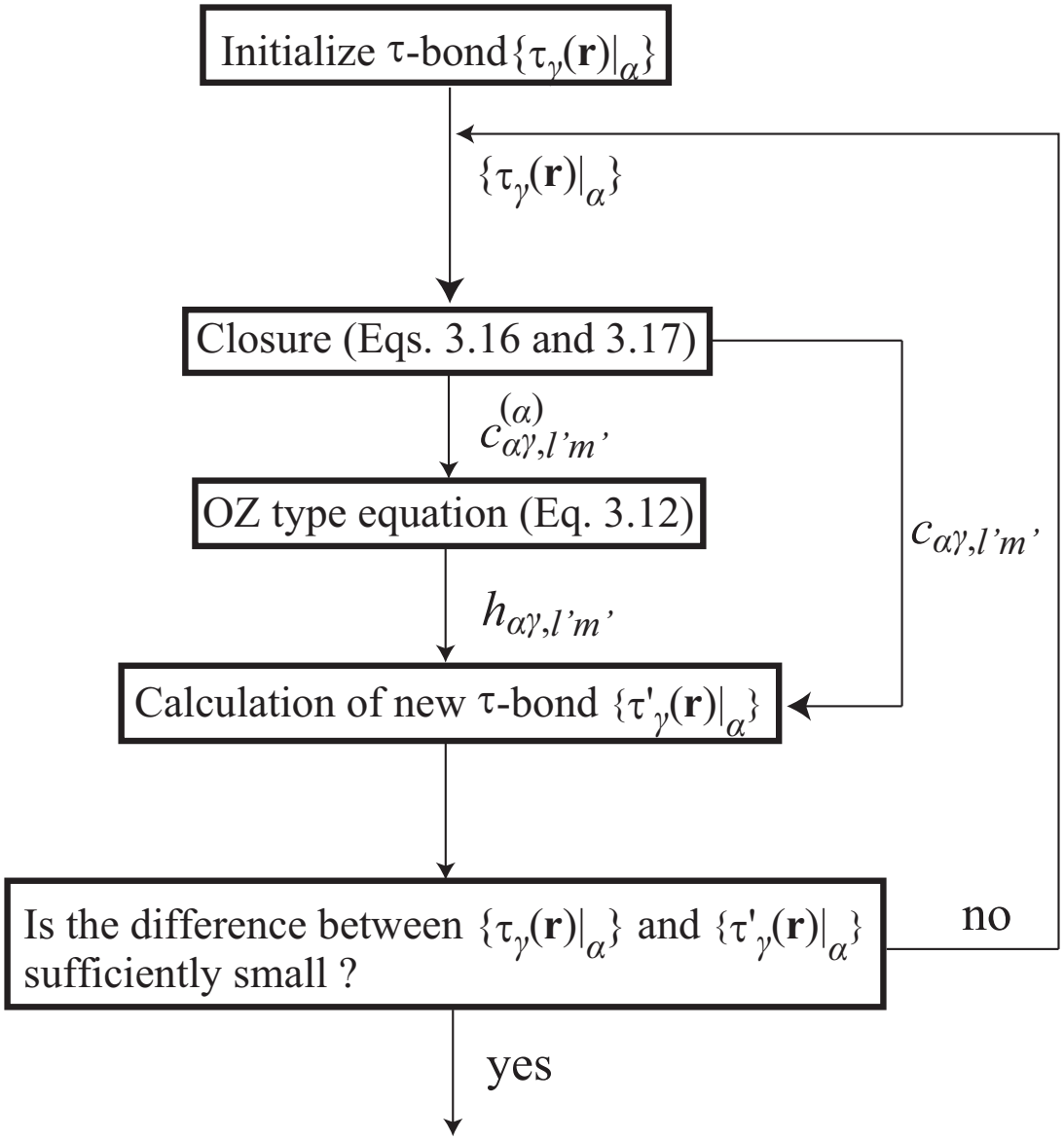
From the standpoint of 3D-RISM, the present procedure can be regarded as a 1D representation of 3D-RISM. On the other hand, Eq. 3.12 can also be derived from the partial Ornstein-Zernike (POZ) scheme. If one of the molecular orientation of Eq. 3.5 in Ref. [16] is spherically averaged and angular momentum indices are applied for another molecular orientation, POZ reduces to Eq. 3.12. In this sense, we can say that the present method implements POZ with angular functions in 2D (spherical Harmonics) for the first time.

3.2.2 Closure relation

In the present work, we start with the following 3D HNC closure [10–12]:

$$C_\delta(\mathbf{r}) = \exp(-\beta u_\delta(\mathbf{r}) + \tau_\delta(\mathbf{r})) - \tau_\delta(\mathbf{r}) - 1, \tag{3.13}$$

where $\beta = 1/kT$, the τ bond is defined as $\tau_\delta = H_\delta - C_\delta$ and $u_\delta(\mathbf{r})$ is the 3D interaction potential between the solvent site δ and the solute.


Initialize τ-bond $\{\tau_\gamma(\mathbf{r})|_\alpha\}$
$\{\tau_\gamma(\mathbf{r})|_\alpha\}$
Closure (Eqs. 3.16 and 3.17)
$c^{(\alpha)}_{\alpha\gamma,l'm'}$
OZ type equation (Eq. 3.12)
$h_{\alpha\gamma,l'm'}$
$c_{\alpha\gamma,l'm'}$
Calculation of new τ-bond $\{\tau'_\gamma(\mathbf{r})|_\alpha\}$
Is the difference between $\{\tau_\gamma(\mathbf{r})|_\alpha\}$ and $\{\tau'_\gamma(\mathbf{r})|_\alpha\}$ sufficiently small ?
no
yes
1D and 3D solvation structures

Scheme 3.1

The auxiliary direct correlation function, $C_{\alpha\delta}(\mathbf{r})$, is expanded around solute site α using real-solid harmonics as follows:

$$\begin{aligned} C_{\alpha\delta}(\mathbf{r}) &\equiv C_{\delta}(\mathbf{r}+\mathbf{R}_{\alpha}) \\ &= \sum_{lm} c_{\alpha\delta,lm}(r_{\alpha}) S_{lm}(\hat{r}_{\alpha}). \end{aligned} \tag{3.14}$$

By using Eqs. 3.8, 3.9, and 3.14, the τ bond can also be expanded around α,

$$\tau_{\alpha\delta}(\mathbf{r}) = \sum_{lm} (h_{\alpha\delta,lm}(r_{\alpha}) - c_{\alpha\delta,lm}(r_{\alpha})) S_{lm}(\hat{r}_{\alpha}). \tag{3.15}$$

To solve Eq. 3.12, the equation which connects $\{h_{\alpha\delta,lm}(r_{\alpha})\}$ with both components of direct correlation functions $\{c_{\alpha\delta,lm}(r_{\alpha})\}$ and the partial direct correlation functions $\{c^{(\alpha)}_{\alpha\delta,lm}(r_{\alpha})\}$ are necessary. Here, $c_{\alpha\delta,lm}(r_{\alpha})$ can be obtained from the following closure relation;

$$\begin{aligned} c_{\alpha\delta,l'm'}(r_{\alpha}) &= \int C_{\alpha\delta}(\mathbf{r}) S_{l'm'}(\hat{r}_{\alpha}) \mathrm{d}\mathbf{\Omega}_{r_{\alpha}} \\ &= \int \left[\exp(-\beta u_{\delta}(\mathbf{r}) + \tau_{\alpha\delta}(\mathbf{r})) - \tau_{\alpha\delta}(\mathbf{r}) - 1\right] S_{l'm'}(\hat{r}_{\alpha}) \mathrm{d}\mathbf{\Omega}_{r_{\alpha}}. \end{aligned} \tag{3.16}$$

By employing the function w_{α} and Eq. 3.3, partial direct correlation function $c^{(\alpha)}_{\alpha\delta,lm}(r_{\alpha})$ can be obtained from the τ bond, like the case of Eq. 3.16.

$$\begin{aligned} c^{(\alpha)}_{\alpha\delta,l'm'}(r_{\alpha}) &= \int w_{\alpha}(\mathbf{r}) C_{\alpha\delta}(\mathbf{r}) S_{l'm'}(\hat{r}_{\alpha}) \mathrm{d}\mathbf{\Omega}_{r_{\alpha}} \\ &= \int w_{\alpha}(\mathbf{r}) \left[\exp(-\beta u_{\delta}(\mathbf{r}) + \tau_{\alpha\delta}(\mathbf{r})) - \tau_{\alpha\delta}(\mathbf{r}) - 1\right] S_{l'm'}(\hat{r}_{\alpha}) \mathrm{d}\mathbf{\Omega}_{r_{\alpha}}. \end{aligned} \tag{3.17}$$

For simple liquids, $c^{(\alpha)}_{\alpha\gamma,lm}(r_{\alpha})$ is equal to $c_{\alpha\gamma,lm}(r_{\alpha})$.

The procedure to solve these equations is summarized in Scheme 3.1.

3.3 Computational details

One-dimensional functions, such as $h_{\eta\gamma,l'm'}$ and $c_{\alpha\gamma,l'm'}$, are defined on logarithmic grids. The integral of Eqs. 3.16 and 3.17 around the solute site is calculated using Gauss-Legendre quadrature. Spherical Bessel transformations of $h_{\eta\gamma,lm}$ and $c_{\alpha\beta,lm}$ are performed using Talman's algorithm [20]. Since this grid system divides 3D space into the radial and angular

parts, the choice of the grid size is more flexible than the cubic lattice. When the cubic lattice is employed for 3D FFT, the number of grids must be 8^N. In our method, the number of the radial grid is 2^N and that of the angular grid can be changed depending on required accuracy. Thus, it is easy for our method to reduce the computational memory.

To apply our method to a charged solute, Ng's method [21] is employed. The 3D interaction potential in Eq. 3.13 is written as follows:

$$u_\gamma(\mathbf{r}) = \sum_\alpha \left[u^{\mathrm{LJ}}_{\alpha\gamma}(\mathbf{r} - \mathbf{R}_\alpha) + u^{\mathrm{El}}_{\alpha\gamma}(\mathbf{r} - \mathbf{R}_\alpha) \right], \tag{3.18}$$

where $u^{\mathrm{LJ}}_{\alpha\gamma}$ and $u^{\mathrm{El}}_{\alpha\gamma}$ are Lennard-Jones part and electrostatic part, respectively. By using multipole expansion, the electrostatic potential is written as follows:

$$u^{\mathrm{El}}_{\alpha\gamma}(\mathbf{r} - \mathbf{R}_\alpha) = q^v_\gamma \sum_{lm} \frac{\sqrt{4\pi}}{2l+1} \frac{S_{lm}(\theta_\alpha, \phi_\alpha)}{|\mathbf{r} - \mathbf{R}_\alpha|^{l+1}} M^u_{\alpha,lm}, \tag{3.19}$$

where θ_α and ϕ_α are determined around the solute site α, q^v_γ is the charge of the solvent site, γ, and the multipole moment $M^u_{\alpha,lm}$ is written with the electron density ρ_α of the solute site α,

$$M^u_{\alpha,lm} = \sqrt{4\pi} \int |\mathbf{r}' - \mathbf{R}_\alpha|^l \rho_\alpha(\mathbf{r}') S_{lm}(\theta'_\alpha, \phi'_\alpha) \mathrm{d}\mathbf{r}'. \tag{3.20}$$

To employ Ng's method (Eq. (3.5) in Ref. [21]), the long range part of the electrostatic potential and the short range part of the partial direct correlation function $c^{(\alpha)}_{\alpha\gamma,lm}$ are written by Eqs. 3.21 and 3.22, respectively:

$$u^{\mathrm{El,long}}_{\alpha\gamma}(|\mathbf{r} - \mathbf{R}_\alpha|) = \frac{q^v_\gamma M^u_{\alpha,00} \mathrm{erf}(|\mathbf{r} - \mathbf{R}_\alpha|)}{|\mathbf{r} - \mathbf{R}_\alpha|} \tag{3.21}$$

$$c^{(\alpha),\mathrm{short}}_{\alpha\gamma,lm}(r_\alpha) = c^{(\alpha)}_{\alpha\gamma,lm}(r_\alpha) + \beta u^{\mathrm{El,long}}_{\alpha\gamma}(r_\alpha)\delta_{l0}\delta_{m0}. \tag{3.22}$$

In this work, we use point charge q^u_α for the solute site α and the following multipole moments:

$$M^u_{\alpha,lm} = \begin{cases} q^u_\alpha & l, m = 0 \\ 0 & l, m \neq 0. \end{cases} \tag{3.23}$$

By replacing $M^u_{\alpha,lm}$ in Eq. 3.19 by Eq. 3.23, simple electrostatic potential is obtained as follows:

$$u^{\mathrm{El}}_{\alpha\gamma}(\mathbf{r} - \mathbf{R}_\alpha) = \frac{q^v_\gamma q^u_\alpha}{|\mathbf{r} - \mathbf{R}_\alpha|}. \tag{3.24}$$

Table 3.1: Lennard-Jones interaction parameters

		ϵ/kcal mol^{-1}	σ/Å
HCl [a]	Chloride	0.5138	3.353
(charged/non-charged)	Hydrogen	0.0397	0.400

[a] Reference [3]

Becke's weight function can be formulated in terms of the following definitions:

$$
\begin{aligned}
w_\alpha(\mathbf{r}) &= \frac{P_{\mathrm{A}}(\mathbf{r})}{\sum_{\mathrm{B}} P_{\mathrm{B}}(\mathbf{r})} && (3.25)\\
P_{\mathrm{A}}(\mathbf{r}) &= \Pi_{\mathrm{B}\neq\mathrm{A}} 0.5\left[1-\nu^{(k)}(r_{\mathrm{A}},r_{\mathrm{B}})\right] && (3.26)\\
\nu^{(l)}(r_{\mathrm{A}},r_{\mathrm{B}}) &= \nu^{(l-1)}(r_{\mathrm{A}},r_{\mathrm{B}})\left(1.5-0.5\left(\nu^{(l-1)}(r_{\mathrm{A}},r_{\mathrm{B}})\right)^2\right) && (3.27)\\
\nu^{(0)}(r_{\mathrm{A}},r_{\mathrm{B}}) &= \mu(r_{\mathrm{A}},r_{\mathrm{B}}) + a_{\mathrm{AB}}\left(1-(\mu(r_{\mathrm{A}},r_{\mathrm{B}}))^2\right) && (3.28)\\
\mu(r_{\mathrm{A}},r_{\mathrm{B}}) &= (r_{\mathrm{A}}-r_{\mathrm{B}})/R_{\mathrm{AB}} && (3.29)\\
r_{\mathrm{A}} &= |\mathbf{r}-\mathbf{R}_{\mathrm{A}}|,\quad r_{\mathrm{B}}=|\mathbf{r}-\mathbf{R}_{\mathrm{B}}|,\quad R_{\mathrm{AB}}=|\mathbf{R}_{\mathrm{A}}-\mathbf{R}_{\mathrm{B}}| && (3.30)\\
a_{\mathrm{AB}} &= \frac{1-\chi_{\mathrm{AB}}^2}{4\chi_{\mathrm{AB}}}\quad (|a_{\mathrm{AB}}\leq 0.5|) && (3.31)\\
\chi_{\mathrm{AB}} &= \sigma_{\mathrm{A}}/\sigma_{\mathrm{B}}. && (3.32)
\end{aligned}
$$

In this work, k is fixed to 4 and σ_{A} (σ_{B}) is the Lennard-Jones parameter of atom A (B).

We calculate the radial distribution functions (RDFs), $g_{\alpha\gamma} = h_{\alpha\gamma,00} + 1$ of charged/non-charged HCl model. The bond length used in this model is 1.3 Å. Calculations are carried out at 210 K and the molecular number density of $\rho = 0.0180$ Å^{-3}. The atomic partial charges used for the charged model are -0.2 e for chloride and +0.2 e for hydrogen. The Lennard-Jones interaction parameters are summarized in Table 3.1. The angular momentum of real solid harmonics up to 5, and 512 and 800 grids for radial and angular parts, respectively, are used.

3.4 Results and discussion

The results of non-charged/charged HCl model are shown in Figs. 3.1 and 3.2, respectively. Cl-Cl RDFs of RISM, POZ, and the present procedure in the non-charged model are in fair

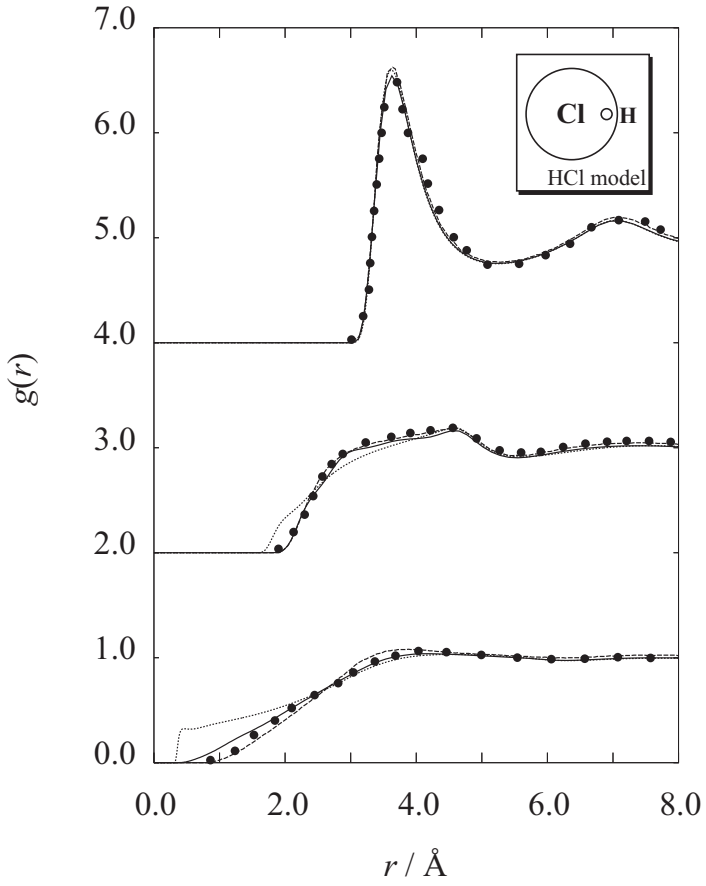


Figure 3.1: Site-site correlation functions, (a) Cl(solute)-Cl(solvent), (b) H(solute)-Cl(solvent), (c) H(solute)-H(solvent), of non-charged HCl model. Solid, dotted, and dashed lines show the results of the present procedure, XRISM, and POZ [15], respectively. Circles denote the Monte Carlo results of Hirata et. al. [3].

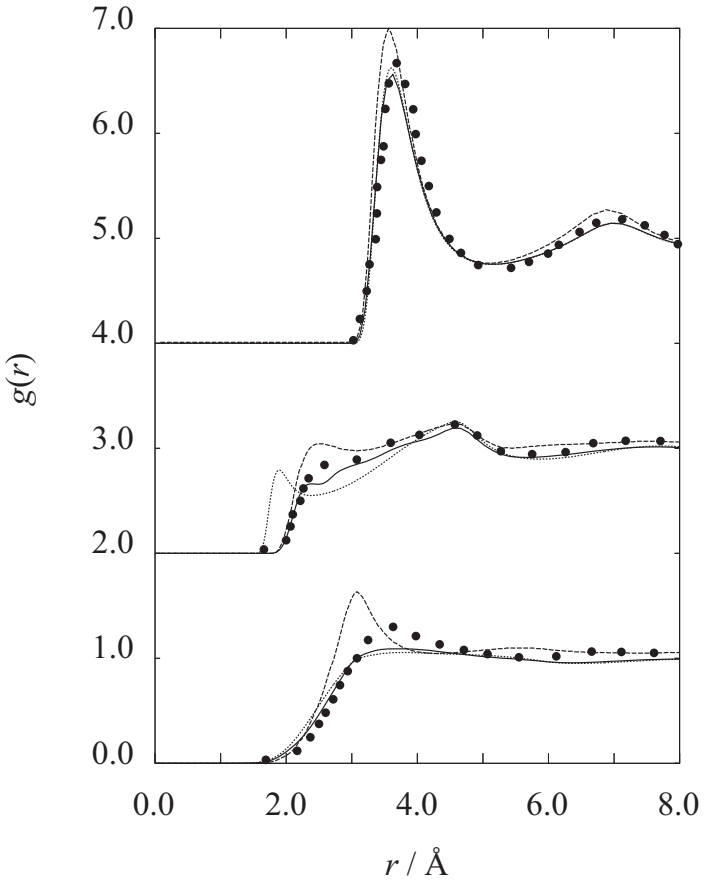


Figure 3.2: Site-site correlation functions, (a) Cl(solute)-Cl(solvent), (b) H(solute)-Cl(solvent), (c) H(solute)-H(solvent), of charged HCl model. See Fig. 3.1 for notations.

agreement with the simulation (Fig. 3.1(a)). The present method can predict the Cl-Cl solvation structure better than POZ for the charged model (Fig. 3.2(a)).

In the present HCl model, the hydrogen site is embedded in the Cl site, as shown in the upper-right corner of Fig. 3.1. For preparing correct H-Cl and H-H RDFs, it is necessary for IET to include intra-molecular interaction in the HCl molecule. The RDF of H-Cl obtained by RISM exhibits an artificial peak at r=1.9Å in charged HCl (Fig. 3.2(b)). When positive hydrogen approaches negative chloride, the hydrogen site cannot come so close to chloride because of Cl-Cl repulsion [13, 15]. The artificial peak shows that RISM cannot evaluate correct intra-molecular interactions. On the other hand, the correct profile of RDF is computed by the present procedure, because it can correctly evaluate H-Cl intra-molecular interaction.

The H-H RDF of the non-charged model obtained by the present procedure can predict very low probability of finding hydrogens within 1 Å, which agrees well with the results of POZ and the simulation. In the charged model, the first peak calculated by the present procedure is almost identical to that of the simulation, though the amplitude is somewhat underestimated.

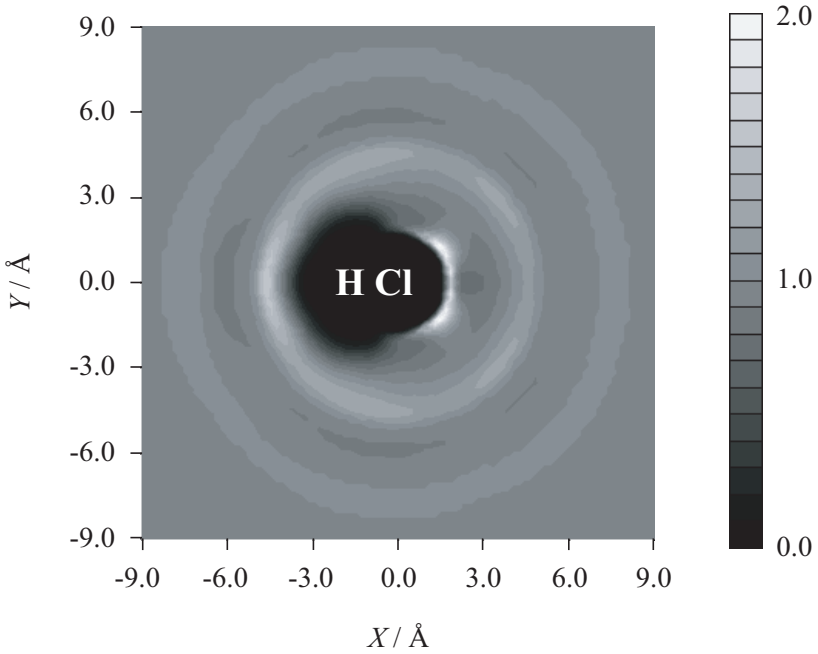


Figure 3.3: 3D solvation structure, $G_{\mathrm{H}}(\mathbf{r})$, of the charged HCl model. Solute chloride and hydrogen nuclei are positioned at (0.0, 0.0) and (-1.3,0.0), respectively.

The present procedure can also evaluate the 3D solvation structure, $G_\alpha(\mathbf{r}) = H_\alpha(\mathbf{r}) + 1$. The hydrogen distribution, $G_{\mathrm{H}}(\mathbf{r})$, for charged HCl model is shown in Fig. 3.3. There is high distribution around the Cl site ($X = 1.1\text{Å} \sim 2.3\text{Å}$, $Y = -1.5\text{Å} \sim 1.5\text{Å}$) because of the strong electrostatic interaction between solvent hydrogen and solute chloride. On the opposite side ($X = -4.7\text{Å}$, $Y = 0.0\text{Å}$), there is a broad distribution. This distribution is derived from the hydrogen which is bonded with the chloride aggregating around solute hydrogen. The 3D map is in fair agreement with the previous results [13, 15].

3.5 Conclusion

We have proposed a new procedure to evaluate 3D solvation structure based on integral equation theory. By employing the expansion of real-solid harmonics, the present procedure does not need expensive Fourier transformations. In the derivation of 3D-HNC like closure, the fussy cell method proposed by Becke is employed. This facilitates the calculation and much clearer understanding of site-site interaction. The present procedure can provide quantitatively accurate radial distribution functions. Some of the RDFs obtained by the present method are in better agreement with those of simulation than those of previous IETs.

In the closure defined in Eqs. (16) and (17), electrostatic potential is represented in 3D coordinate. This will allow us easier use of the electrostatic potential calculated by quantum chemistry.

Bibliography

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

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

[3] F. Hirata, B. M. Pettitt, and P. J. Rossky, J. Chem. Phys. **77**, 509 (1982).

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

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

[6] M. Kinoshita, Y. Okamoto, and F. Hirata, J. Chem. Phys. **110**, 4090 (1999).

[7] H. Sato and F. Hirata, Bull. Chem. Soc. Jpn. **74**, 1831 (2001).

[8] D. Yokogawa, H. Sato, and S. Sakaki, J. Chem. Phys. **123**, 211102 (2005).

[9] D. Yokogawa, H. Sato, and S. Sakaki, J. Chem. Phys. **125**, 114102 (2006).

[10] M. Ikeguchi and J. Doi, J. Chem. Phys. **103**, 5011 (1995).

[11] D. Beglov and B. Roux, J. Phys. Chem. B **101**, 7821 (1997).

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

[13] C. M. Cortis, P. J. Rossky, and R. A. Friesner, J. Chem. Phys. **107**, 6400 (1997).

[14] A. Furuhama and S. Ten-no, J. Chem. Phys. **117**, 4087 (2002).

[15] A. Furuhama and S. Ten-no, Chem. Phys. Lett. **384**, 376 (2004).

[16] S. Ten-no and S. Iwata, J. Chem. Phys. **111**, 4865 (1999).

[17] T. Sumi and H. Sekino, J. Chem. Phys. **125**, 034509 (2006).

[18] A. D. Becke, J. Chem. Phys. **88**, 2547 (1988).

[19] J. D. Talman, Int. J. Quantum. Chem. **93**, 72 (2003).

[20] J. D. Talman, J. Comput. Phys. **29**, 35 (1978).

[21] K.-C. Ng, J. Chem. Phys. **61**, 2680 (1974).

Chapter 4

A highly parallelizable solvation structure theory based on Three-Dimensional Reference Interaction Site Model: Application to biomolecules.

4.1 Introduction

Most of the biochemical reactions proceed in aqueous phase. The waters locate inside and around a protein. Almost all of the proteins can proceed the reaction only in this environment. For example, Bacteriorodopsin, which is a light-driven proton pump, prepare the hydrogen bonding network between hydrated waters and residues inside. Protons are considered to be transfered across the membrane through the network [1, 2]. To clear the functions of these waters, the information for the distribution of them is indispensable.

The experimental study to obtain the hydration structure has been performed using several powerful experimental techniques, such as X-ray diffraction at low temperature and the scattering with very high-power neutron source. Thanks to these studies, the high resolution data, which is accurate enough to discuss the hydration structure around protein, has become available. However, such solvation information is still very limited. Furthermore, almost all of them are in crystal, which is believed to be essentially different from aqueous environment.

Theoretical approach to evaluate the hydration structure has been also performed. The most popular treatment may be molecular simulation. Although the method was applied to some solvation systems, to obtain the hydration structure inside and around protein with high-quality, very long simulation is necessary. Recently a theoretical determination of hydration structure around protein has been done by Imai et al. based on statistical mechanics [3]. They employed

3D reference interaction site model (3D-RISM) [4, 5] for hen egg-white lysozyme and showed that 3D-RISM is capable of computing the solvation structure not only around protein but also inside appropriately. Yoshida et al. showed that 3D-RISM can also correctly reproduce the selective ion binding with human lysozyme [6].

In 3D-RISM calculation, the hydration structure is evaluated on 3D lattice grids. The size of the protein which can be calculated is determined by the number of grid points and the resolution of the data. In principle, it would be possible to compute high-quality hydration structure around large size of the proteins whatever you want by increasing the number of grid size. However, high computational cost and very long calculation are required to obtain such results.

In this work, a new approach based on 3D-RISM, Fragment 3D-RISM, was developed, in which 3D solvation structure is reconstructed from the solvation structures evaluated around each solute site. Since these calculations are performed with a combination of logarithmic grid and real solid harmonics, the number of grid points in the present scheme can be adjusted more flexibly compared to 3D lattice grid employed in the conventional 3D-RISM. Moreover, because the algorithm of the present method readily achieves high parallel performance, the computational time can be reduced dramatically. The total solvation structure is reconstructed with the calculated fragment results.

The organization of this paper is as follows. In Sec. 4.2 and 4.3, the formalism of the present method and the computational details are presented, respectively. In Sec. 4.4, the following three topics are described. First, the parallel performance of the present method is checked. Second, the calculated results in this work are compared with those evaluated by the conventional 3D-RISM in terms of hydration structure and partial molar volume. Finally, the hydration structure around a protein called Fv fragment is evaluated. The computed results are compared with the high-quality results obtained by X-ray crystallography.

4.2 Method

4.2.1 Ornstein-Zernike type equation

3D-RISM equation is written by,

$$H_\alpha(\mathbf{r}) = \sum_\beta C_\beta * \left[\omega^V_{\beta\alpha} + \rho h^V_{\beta\alpha}\right](\mathbf{r}), \tag{4.1}$$

where H_α and C_α are 3D total and direct correlation functions, ω^V is the intramolecular correlation function of solvent, and h^V is the total correlation function of solvent. In the present theory, H_α and C_α are expressed as the sum of reference and residual correlation functions, as follows:

$$H_\alpha(\mathbf{r}) = H^{\mathrm{ref}}_\alpha(\mathbf{r}) + \Delta H_\alpha(\mathbf{r}), \tag{4.2}$$

$$C_\alpha(\mathbf{r}) = C^{\mathrm{ref}}_\alpha(\mathbf{r}) + \Delta C_\alpha(\mathbf{r}). \tag{4.3}$$

The reference correlation functions, H^{ref}_α and C^{ref}_α are defined with 1D correlation functions by

$$H^{\mathrm{ref}}_\alpha(\mathbf{r}) \equiv \sum_{\beta\gamma} c_{\beta\gamma} * \left[\omega^V_{\gamma\alpha} + \rho h^V_{\gamma\alpha}\right](r_\beta), \tag{4.4}$$

$$C^{\mathrm{ref}}_\alpha(\mathbf{r}) \equiv \sum_\beta c_{\beta\alpha}(r_\beta), \tag{4.5}$$

where $r_\beta = |\mathbf{r} - \mathbf{R}_\beta|$ and $\mathbf{R}_\beta$ is the position of the solute atomic site β. Inserting eqs. 4.2, 4.3, 4.4, and 4.5 into eq. 4.1 leads to the Ornstein-Zernike (OZ) type equation for the residual correlation functions,

$$\Delta H_\alpha(\mathbf{r}) = \sum_\beta \Delta C_\beta * \left[\omega^V_{\beta\alpha} + \rho h^V_{\beta\alpha}\right](\mathbf{r}). \tag{4.6}$$

The residual functions, ΔH_α and ΔC_α are divided into the components on each solute atomic site,

$$\Delta H_\alpha(\mathbf{r}) = \sum_\beta w_\beta(\mathbf{r}) \Delta H^{(\beta)}_\alpha(\mathbf{r}_\beta), \tag{4.7}$$

$$\Delta C_\alpha(\mathbf{r}) = \sum_\beta w_\beta(\mathbf{r}) \Delta C^{(\beta)}_\alpha(\mathbf{r}_\beta), \tag{4.8}$$

where $w_\beta(\mathbf{r})$ is the weight function for solute atomic site β at the position $\mathbf{r}$. In this work, $\Delta H_\alpha^{(\beta)}$ and $\Delta C_\alpha^{(\beta)}$ are expanded with real solid harmonics S_{lm} around atom center β, as follows:

$$\Delta H_\alpha^{(\beta)}(\mathbf{r}_\beta) = \sum_{lm} \Delta H_{lm,\alpha}^{(\beta)}(r_\beta) S_{lm}(\hat{\mathbf{r}}_\beta), \tag{4.9}$$

$$\Delta C_\alpha^{(\beta)}(\mathbf{r}_\beta) = \sum_{lm} \Delta C_{lm,\alpha}^{(\beta)}(r_\beta) S_{lm}(\hat{\mathbf{r}}_\beta), \tag{4.10}$$

where $\hat{\mathbf{r}}_\beta$ is a unit vector with its origin. The component of residual total correlation function is approximated by the following equation,

$$\Delta H_{lm,\alpha}^{(\beta)}(r_\beta) \simeq \sum_\delta \Delta C_{lm,\delta}^{(\beta)} * \left[\omega_{\delta\alpha}^V + \rho h_{\delta\alpha}^V\right](r_\beta). \tag{4.11}$$

4.2.2 Closure

In 1D-RISM and 3D-RISM framework, there are many closure equations. Kovalenko and Hirata proposed the following closure (KH closure) for 3D-RISM [7],

$$\begin{aligned} H_\alpha(\mathbf{r}) &= \begin{cases} \exp(\chi_\alpha(\mathbf{r})) - 1 & \text{for } \chi_\alpha(\mathbf{r}) \le 0 \\ \chi_\alpha(\mathbf{r}) & \text{for } \chi_\alpha(\mathbf{r}) > 0 \end{cases} \\ \chi_\alpha(\mathbf{r}) &= -\beta u_\alpha(\mathbf{r}) + H_\alpha(\mathbf{r}) - C_\alpha(\mathbf{r}), \end{aligned} \tag{4.12}$$

where $\beta = 1/k_B T$, k_B is Boltzmann's factor, and $u_\alpha(\mathbf{r})$ is the intermolecular potential function between solute and the solvent site α. To solve eqs. 4.9, 4.10, and 4.11, we elaborated the following KH type closure,

$$\begin{aligned} \Delta H_\alpha^{(\gamma)}(\mathbf{r}_\gamma) &= \begin{cases} \exp(\chi_\alpha^{(\gamma)}(\mathbf{r}_\gamma)) - 1 - H_\alpha^{\mathrm{ref}}(\mathbf{r}_\gamma) & \text{for } \chi_\alpha^{(\gamma)}(\mathbf{r}_\gamma) \le 0 \\ \chi_\alpha^{(\gamma)}(\mathbf{r}_\gamma) - H_\alpha^{\mathrm{ref}}(\mathbf{r}_\gamma) & \text{for } \chi_\alpha^{(\gamma)}(\mathbf{r}_\gamma) > 0 \end{cases} \\ \chi_\alpha^{(\gamma)}(\mathbf{r}_\gamma) &= -\beta u_\alpha(\mathbf{r}_\gamma) + \{H_\alpha^{\mathrm{ref}}(\mathbf{r}_\gamma) + \Delta H_\alpha^{(\gamma)}(\mathbf{r}_\gamma)\} - \{C_\alpha^{\mathrm{ref}}(\mathbf{r}_\gamma) + \Delta C_\alpha^{(\gamma)}(\mathbf{r}_\gamma)\}, \end{aligned} \tag{4.13}$$

where $u_\alpha(\mathbf{r}_\gamma)$ is the intermolecular potential function between solute and the solvent site α around solute site β. The difference between $u_\alpha(\mathbf{r})$ in eq. 4.12 and $u_\alpha(\mathbf{r}_\gamma)$ in eq. 4.13 is only the origins of the vectors $\mathbf{r}$ and $\mathbf{r}_\gamma$.

The procedure of the present method is summarized as follows. The reference correlation functions, H_α^{ref} and C_α^{ref}, are calculated by eqs. 4.4 and 4.5 with the correlation functions obtained by 1D-RISM [8–10] (**step 1**). The residual correlation functions $\Delta H_\alpha^{(\beta)}(\mathbf{r}_\gamma)$, $\Delta C_\alpha^{(\beta)}(\mathbf{r}_\gamma)$

are calculated on solute site β with eqs. 4.9, 4.10, 4.11, and 4.13 (**step 2**). After the step, the solvation structure for the solvent site α ($H_\alpha(\mathbf{r}) + 1$) is evaluated by eqs. 4.2, 4.4, and 4.7 (**step 3**). In this algorithm, **step 2**, which is most time demanding, is highly parallelized because the calculation of the residual correlation functions on each solute site (β) can be treated independently.

4.3 Computational details

The grid set employed in this work is logarithmic grid [11] for radial part and the Lebedev grid [12] for angular part. With the grid set, the convolution in eq. 11 is performed by spherical Bessel transformation [11]. For the weight function $w_\beta(\mathbf{r})$, Becke's function was employed [13]. The calculation of the present method is parallelized with MPICH2 [14].

The geometries of the proteins were taken from PDB data and those of amino acids from $Klotho$ (Biochemical Compounds Declarative Database) [15]. The potential functions for the amino acids and the proteins are united-atom OPLS parameters [16]. For solvent water molecule, SPC-like model was employed [17] with a correction concerning the Lennard-Jones parameters of the hydrogen sites (σ=1.0Å, ϵ=0.056 kcal/mol). To visualize 3D solvation structure, VMD software was used [18].

4.4 Results and discussion

4.4.1 Computational Performance of parallelization

The computational performance of the parallelization was evaluated with a speed up ratio. The benchmark calculation was performed using a small protein, Chignolin [19].

Speed up ratio $S(N)$ using N processors is defined as,

$$S(N) = \frac{\text{Execution time with 1 processor}}{\text{Execution time with } N \text{ processors}} \quad . \tag{4.14}$$

If a program is completely parallelized, $S(N) = N$ (dashed line in Fig. 1), corresponding to linear speed-up. Although the evaluated speed-up ratio $S(N)$ is somewhat smaller than N (filled circle in Fig. 1), the present method showed good performance. Since the execution

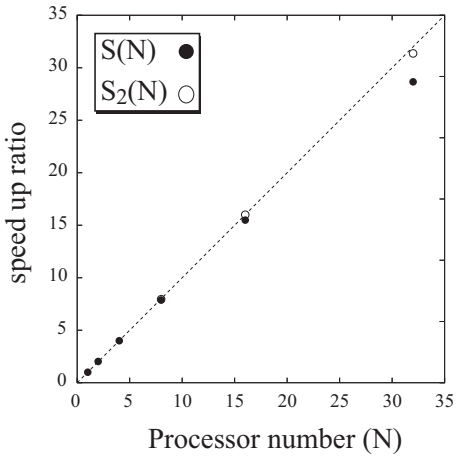


Figure 4.1: Speed up ratio as a function of the number of used processors (N).

time for the calculation is a sum of the times for **step 1** and **step 2**, the ratio of **step 2** $S_2(N)$ is also separately shown in Fig. 4.1, showing the degradation of the computational performance mainly comes from the **step 1**.

In **step 1**, the 1D correlation functions for all solute sites and solvent sites are evaluated at the same time. These functions between solute sites are mixed through the spherical convolution integral in RISM equation. Because transfer of very large data among each processor is required, the performance of this step is degraded. On the other hand, **step 2** does not include the convolution integral between the solute sites, allowing to calculate the functions independently. Note that **step 2** is dominative in the total time and the contributions from **steps 1** and **3** are significantly small. As a consequence, virtually linear speed-up is achieved.

4.4.2 Comparison of the present method with 3D-RISM

The present method can be considered as the approximation of 3D-RISM. To evaluate the accuracy of this scheme, comparison between the present method and the conventional 3D-RISM was performed on hydration structure and partial molar volume (PMV).

The hydration structures around tryptophan in zwitterionic form evaluated by the present method and 3D-RISM are shown in Fig. 4.2. Because tryptophan is a molecule including both of hydrophobic group (benzene ring) and hydrophilic groups ($\mathrm{NH_3^+}$, $\mathrm{CO_2^-}$, and NH in

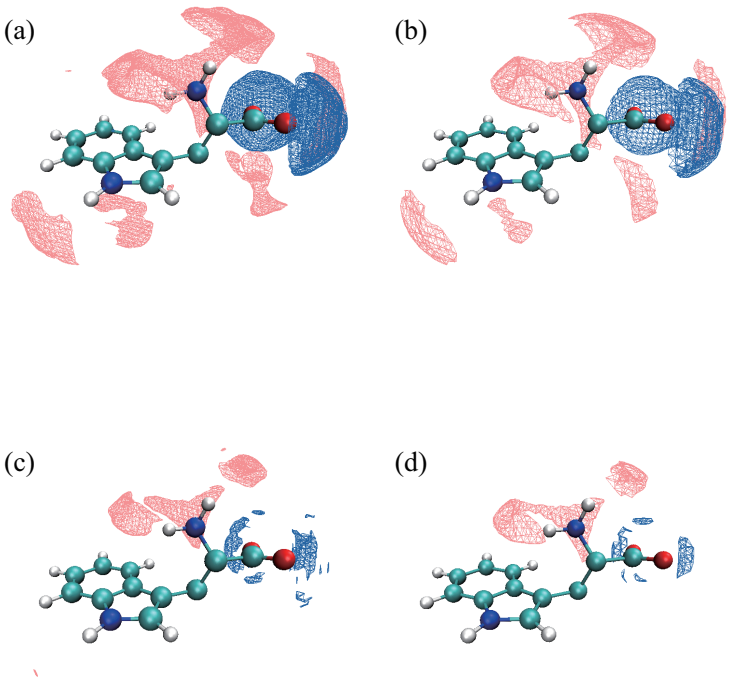


Figure 4.2: 3D hydration structures of oxygen site (pink) and hydrogen site (blue) around tryptophan calculated by the present method (a) and (c) and by 3D-RISM (b) and (d). The isodensity surface of the solvation structure is 3.5 for (a) and (b) and 4.5 for (c) and (d).

pyrrole), it is a good example for the comparison. When the threshold of the isodensity is 3.5 (Figs. 4.2(a) and 4.2(b)), the broad distributions of water hydrogen (blue region) and of water oxygen (pink region) can be seen around hydrophilic groups. The shape of the distributions evaluated by the present method (Fig. 4.2(a)) showed good agreement with that obtained by 3D-RISM (Fig. 4.2(b)). By increasing the threshold, strongly binding hydration waters can be drawn selectively. In Figs. 4.2(c) and 4.2(d), the hydration structures with the threshold of 4.5 are shown. Strongly binding water oxygen and water hydrogen can be seen only around $\mathrm{NH_3^+}$ and $\mathrm{CO_2^-}$, respectively. Therefore, it is concluded that the present method reproduced the distribution of hydration structure evaluated by the original 3D-RISM.

PMV of the solute molecule in molecular solvent, $\bar{V}_u^0$, is expressed in terms of the 3D solute-solvent direct correlation functions by the relation [20, 21]

$$\bar{V}_u^0 = k_B T \chi_T^0 \left(1 - \rho \sum_\beta \int C_\beta(\mathbf{r}) \mathrm{d}\mathbf{r}_\beta \right), \tag{4.15}$$

where χ_T^0 is the isothermal compressibility of the pure solvent. By inserting eq. 4.3 into eq.

4.15, $\bar{V}_u^0$ is expressed with a sum of reference part and residual part, as follows:

$$\begin{aligned} \bar{V}_u^0 &= k_B T \chi_T^0 \left(1 - \rho \sum_{\alpha\beta} \int c_{\alpha\beta}(\mathbf{r}_\gamma) \mathrm{d}\mathbf{r}_\gamma \right) - k_B T \chi_T^0 \rho \sum_{\gamma} \int \Delta C_\gamma(\mathbf{r}_\gamma) \mathrm{d}\mathbf{r}_\gamma \\ &= \bar{V}_u^{\mathrm{ref}} + \Delta \bar{V}_u. \end{aligned} \tag{4.16}$$

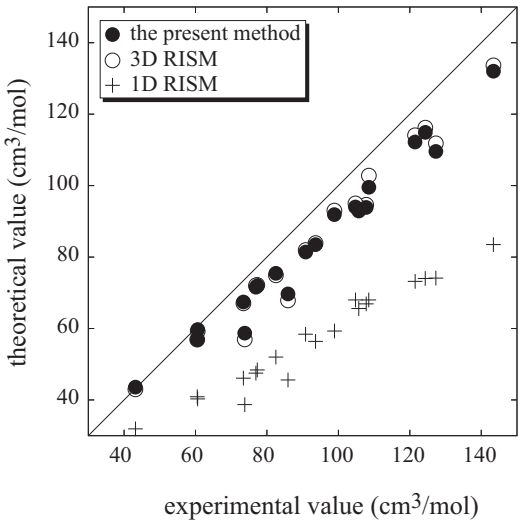


Figure 4.3: PMVs of 20 amino acids calculated by 1D-RISM, the present method, and 3D-RISM in comparison with the corresponding experimental data [22–25].

In Fig. 4.3, the PMVs of 20 amino acids in zwitterionic form calculated by 1D-RISM, the present method, and 3D-RISM are plotted in comparison with the corresponding experimental data [22–25]. If a theoretical value reproduces the experimental one exactly, the data is plotted on the solid line. As shown previously [21], the PMV calculated by 1D-RISM is much smaller than experimental data. On the other hand, the present method considerably improves the agreement with the experimental values. The accuracy is almost the same as that of 3D-RISM. In both of the present method and 3D-RISM, 1D-RISM strategy is used for the description of the solvent-solvent correlations. This may be the reason of the same accuracy of the two 3D type methods.

4.4.3 Hydration structure around a protein

The present method was applied to a protein called Fv fragment, which is a part of an anti-dansyl antibody. The protein was thoroughly studied with X-ray crystallography at low

temperature and high-quality crystal structures were obtained in different conditions [26, 27]. Because the hydration structures around them were also discussed in their works, it is a good system to evaluate the difference between calculated and experimentally observed results. In this work, the hydration structures around only the Fv fragment (unliganded state) [26] and around the Fv fragment which binds with ϵ-dansyl-L-lysine (DNS-lys) (complex state) [27] were evaluated. The geometries for the unliganded state and the complex state were taken from PDB data (PDBID: 2dlf.pdb and 1wz1.pdb). All the water molecules were removed from the data before the computation.

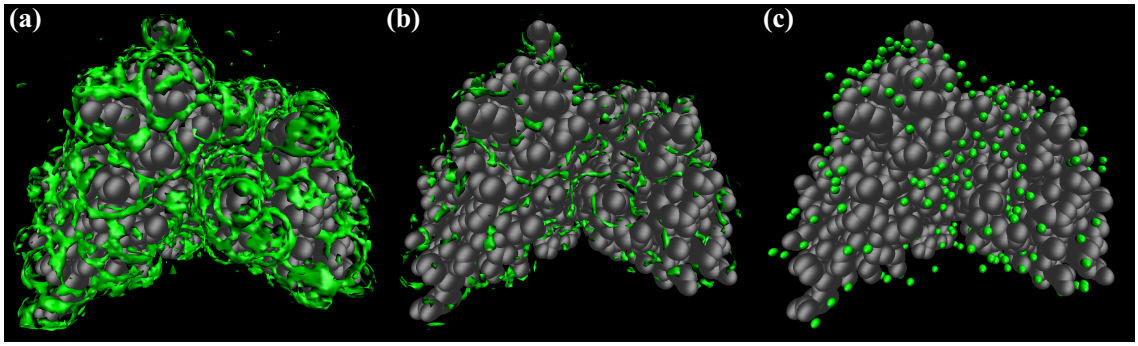


Figure 4.4: 3D hydration structures of oxygen site (green area). The isodensity surfaces of these structures are 2.5 (a) and 4.0 (b), respectively. For the sake of viewability, all of the residues are shown with black spheres.

The hydration structures of water oxygen $G(\mathbf{r})$ around the unliganded state with different threshold of isodensity surfaces are shown in Fig. 4.4 together with the hydration waters determined by X-ray crystallography. When the threshold is 2.5 (Fig. 4.4(a)), the broad hydration structure is observed. The broadening of the distribution reflects the fluctuation of solvent waters at room temperature. The hydration structure with the threshold 4.0 is considerably localized (Fig. 4.4(b)). The distribution shows the area where waters bind strongly with the protein, corresponding to peaks in $G(\mathbf{r})$. These positions are in reasonable agreement with the experimental data, which is obtained at low temperature (Fig. 4.4(c)).

The Fv fragment has a binding pocket with DNS-lys and the structure around the site is greatly different between the unliganded state and complex state. In Fig. 4.5, the calculated hydration structures around the binding site for these states are shown together with the ex-

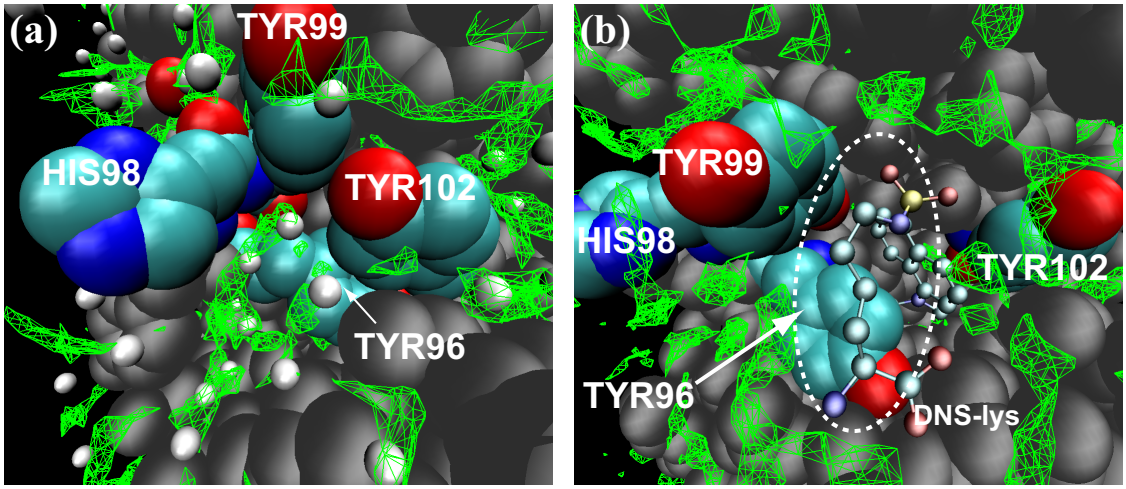


Figure 4.5: The hydration structure of oxygen site (green area) around the binding pocket of DNS-lys for the unliganded state (a) and complex state (b). The isodensity surface is 3.5. The residues except HIS98, TYR96, TYR99, and TYR102 are shown with black spheres. The red, blue, and cyan spheres corresponds to oxygen atom, nitrogen atom, and carbon atom, respectively. The waters determined by X-ray crystallography are shown with white spheres.

perimentaly determined waters (white spheres). In the unliganded state, the binding pocket is surrounded by TYR96, HIS98, TYR99, and TYR102. Around the residues, there are many waters observed experimentally and the positions of them were correctly reproduced by the present method (green mesh area in Fig. 4.5(a)). In the case of the complex state, these residues flip away and the pocket is opened. Along the flip, the hydrophilic groups such as N and O are pushed away and the hydrophobic groups such as benzene rings of TYR96 and TYR99 appear. The distribution of waters can be seen around the binding pocket and there is a hydrophobic space where DNS-lys binds (dotted circle in Fig. 4.5(b)), which is consistent with the previous work [27].

In Fig. 4.6, the bottom of the binding pocket in the complex state is focused to see water oxygen (green) and water hydrogen (white). The positively charged water hydrogen sites are around N and O site of DNS-lys and TYR102, respectively, and the negatively charged water oxygen site is around H site of TYR102, which shows the three hydrogen bondings between DNS-lys and TYR102 (dotted line in Fig. 4.6). These distributions correspond to the water experimentally observed by X-ray diffraction technique at low temperature [27].

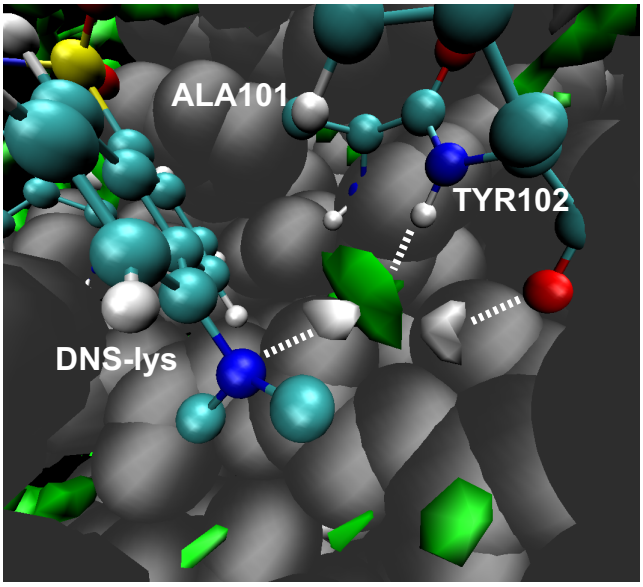


Figure 4.6: The hydration structure of oxygen site (green area) and hydrogen site (white area) in the pocket of DNS-lys, where the isodensity is 3.0. The residues except ALA 101 and TYR102 are shown with black spheres.

4.5 Conclusion

In this article, a new approach, Fragment 3D-RISM, was proposed. The solvation structure is calculated with three steps, the calculation of the reference correlation functions (**step 1**), the residual correlation functions (**step 2**), and the building up of solvation structure from the results obtained by the previous steps (**step 3**). Because the most time demanding step (**step 2**) is highly parallelized, the computational time can be reduced dramatically. Although the present method is regarded as an approximation of 3D-RISM, the hydration structure of tryptophan and PMV of amino acids were almost the same as those evaluated by the original 3D-RISM.

The method was applied to the hydration structure around the binding pocket between Fv fragment and DNS-lys. By drawing the 3D solvation structure with small and large thresholds of isodensity, the fluctuation of waters and tightly binding waters can be investigated. The calculated hydration structure was good agreement with the experimentally observed results. By evaluating not only water oxygen site but also water hydrogen site, which is difficult to

be determined by X-ray crystallography, the direction of the hydrogen bonding was clearly shown.

Bibliography

[1] H. Luecke, B. Schobert, H.-T. Richter, J.-P. Cartailler, and J. K. Lanyi, J. Mol. Biol. **291**, 899 (1999).

[2] H. Kandori, Biochim. Biophys. Acta **1460**, 177 (2000).

[3] T. Imai, R. Hiraoka, A. Kovalenko, and F. Hirata, J. Am. Chem. Soc. **127**, 15334 (2005).

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

[5] D. Beglov and B. Roux, J. Phys. Chem. B **101**, 7821 (1997).

[6] N. Yoshida, S. Phongphanphanee, and F. Hirata, J. Phys. Chem. B **111**, 4588 (2007).

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

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

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

[10] *Molecular Theory of Solvation*, edited by F. Hirata (Kluwer, Dordrecht, 2003).

[11] J. D. Talman, J. Comput. Phys. **29**, 35 (1978).

[12] V. I. Lebedev and D. Laikov, Dokl. Math. **59**, 477 (1999).

[13] A. D. Becke, J. Chem. Phys. **88**, 2547 (1988).

[14] MPICH2 High-performance and widely portable implementation of MPI http://www.mcs.anl.gov/research/projects/mpich2/index.php.

[15] B. Cannon, S. Sadekar, D. Taylor, B. H. Dunford-Shore, W. Sulaman, B. Feng, F. Fabrizio, J. Holcomb, W. Wise, and T. Kazic; 1994–present, http://www.biocheminfo.org/klotho/.

[16] W. L. Jorgensen and J. Tirado-Rives, J. Am. Chem. Soc. **110**, 1657 (1988).

[17] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces* , edited by B. Pullman (Reidel, Dordrecht, 1981).

[18] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).

[19] S. Honda, K. Yamasaki, Y. Sawada, and H. Morii, Structure **12**, 1507 (2004).

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

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

[22] M. Kikuchi, M. Sakurai, and K. Nitta, J. Chem. Eng. Data **40**, 935 (1995).

[23] A. A. Zamyatnin, Annu. Rev. Biophys. Bioeng. **13**, 145 (1984).

[24] Y. Yasuda, N. Tochio, M. Sakurai, and K. Nitta, J. Chem. Eng. Data **43**, 205 (1998).

[25] M. Mizuguchi, M. Sakurai, and K. Nitta, J. Solution Chem. **26**, 579 (1997).

[26] M. Nakasako, H. Takahashi, N. Shimba, I. Shimada, and Y. Arata, J. Mol. Biol. **291**, 117 (1999).

[27] M. Nakasako, T. Oka, M. Mashumo, H. Takahashi, I. Shimada, Y. Yamaguchi, K. Kato, and Y. Arata, J. Mol. Biol. **351**, 627 (2005).

Chapter 5

The position of water molecules in Bacteriorhodopsin: A fragment Three-Dimensional Reference Interaction Site Model study

5.1 Introduction

The information of hydration structure is very fundamental in biosystem [1]. Hydrogen bondings between waters and protein affect protein structure and activity of enzyme. The information is also very useful in drag design since the position of waters in the vicinity of the activity site has great influence on the stabilization of the drug-protein interaction [2, 3] .

The waters in biosystem can be classified into “surface” or “bound”, according to whether they are surrounded by other water molecules or protein [4]. Some proteins contain the bound water molecules inside, which sometimes play an essential role on their functional features. For example, it is well known that hydrogen bonded water molecules play a key role in Bacteriorhodopsin (bR), which is a light-driven proton pump in *Halobacterium salinarum*. To reveal the mechanism of the pump, a huge number of approaches including X-ray crystallography [5], resonance Raman [6] and Fourier transform infrared spectroscopy [7] have been performed. These experimental approaches have elucidated that the hydrogen-bonding networks of these water molecules providing the proton pathway in the pump [7]. Theoretical approaches to study the mechanism have been also performed [8–14]. In most of the studies, the mechanism of bR function was focused and the initial positions of bound waters inside protein were taken

from experimental data.

The theoretical prediction of the water distribution is still very limited. It is extremely difficult task for the molecular simulation such as molecular dynamics to compute the distribution because it is necessary to sample the interaction between the protein and water molecules on the extensive free energy hypersurface. It has been shown that an integral equation theory for molecular liquids, three-dimensional reference interaction site model (3D-RISM) [15, 16], is a powerful tool to study the distribution of bound and surface waters and numerous applications have been carried out [17–21]. We have recently developed a new approach based on 3D-RISM, in which the equations are elaborated so that the high parallel performance can be achieved (fragment 3D-RISM) [22]. Similar to the original 3D-RISM, our method evaluates the 3D solvation structure but the required computational time can be compressed. Although the present theory is regarded as an approximation of 3D-RISM, the obtained distribution function is virtually the same as that by the original method [22]. Furthermore, the new theory is practically free from the grid size since the distribution functions are computed by the expansion around the individual solute site. Actually the functions are described in higher-resolution than the original one.

In the present work, we applied the fragment 3D-RISM to computations of the waters' position in bR and compared with those obtained by X-ray crystallography as well as by previous simulations. After brief description of the method, details of the calculation are explained in section 5.3. The bound waters are discussed in section 5.4.

5.2 Method

3D correlation functions of a solvent site α, H_α and C_α are expressed with reference and residual correlation functions, as follows:

$$H_\alpha(\mathbf{r}) = H_\alpha^{\mathrm{ref}}(\mathbf{r}) + \Delta H_\alpha(\mathbf{r}), \tag{5.1}$$

$$C_\alpha(\mathbf{r}) = C_\alpha^{\mathrm{ref}}(\mathbf{r}) + \Delta C_\alpha(\mathbf{r}). \tag{5.2}$$

For the reference correlation functions, $H_\alpha^{\rm ref}$ and $C_\alpha^{\rm ref}$, 1D correlation functions are employed [22]. The residual functions, ΔH and ΔC, are 3D functions and divided into the components localized on each atomic site of solute,

$$\Delta H_\alpha(\mathbf{r}) = \sum_\beta w_\beta(\mathbf{r}) \Delta H_\alpha^{(\beta)}(\mathbf{r}_\beta), \tag{5.3}$$

$$\Delta C_\alpha(\mathbf{r}) = \sum_\beta w_\beta(\mathbf{r}) \Delta C_\alpha^{(\beta)}(\mathbf{r}_\beta), \tag{5.4}$$

where $w_\beta(\mathbf{r})$ is the weight function for solute atomic site β at the position $\mathbf{r}$. In this work, $\Delta H_\alpha^{(\beta)}$ and $\Delta C_\alpha^{(\beta)}$ are expanded with real solid harmonics S_{lm}, as follows:

$$\Delta H_\alpha^{(\beta)}(\mathbf{r}_\beta) = \sum_{lm} \Delta H_{lm,\alpha}^{(\beta)}(r_\beta) S_{lm}(\hat{\mathbf{r}}_\beta), \tag{5.5}$$

$$\Delta C_\alpha^{(\beta)}(\mathbf{r}_\beta) = \sum_{lm} \Delta C_{lm,\alpha}^{(\beta)}(r_\beta) S_{lm}(\hat{\mathbf{r}}_\beta), \tag{5.6}$$

where $\hat{\mathbf{r}}_\beta$ is a unit vector with its origin at atom β. The component of residual total correlation function can be approximated by the following equation in analogy with the original RISM theory,

$$\Delta H_{lm,\alpha}^{(\beta)}(r_\beta) \simeq \sum_\delta \Delta C_{lm,\delta}^{(\beta)} * \left[\omega_{\delta\alpha}^V + \rho h_{\delta\alpha}^V\right](r_\beta). \tag{5.7}$$

To solve Eqs. 5.5, 5.6 and 5.7, we elaborated the following Kovalenko-Hirata (KH) type closure [23],

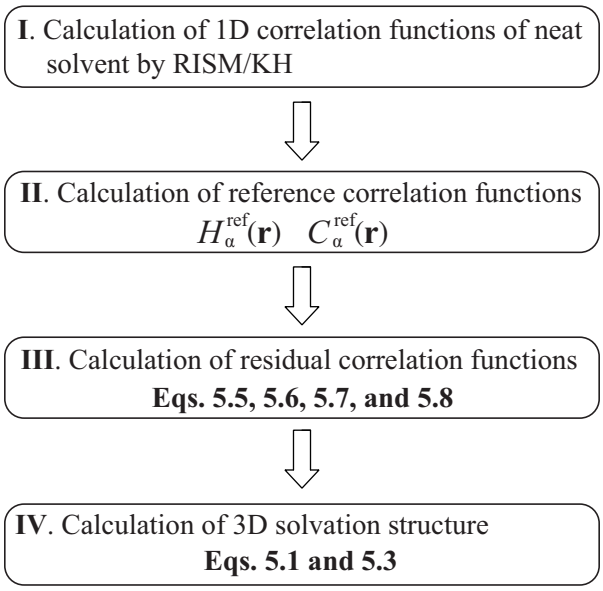
$$\begin{aligned}
\Delta H_\alpha^{(\gamma)}(\mathbf{r}_\gamma) &= \begin{cases} \exp(\chi_\alpha^{(\gamma)}(\mathbf{r}_\gamma)) - 1 - H_\alpha^{\rm ref}(\mathbf{r}_\gamma) & \text{for } \chi_\alpha^{(\gamma)}(\mathbf{r}_\gamma) \le 0 \\ \chi_\alpha^{(\gamma)}(\mathbf{r}_\gamma) - H_\alpha^{\rm ref}(\mathbf{r}_\gamma) & \text{for } \chi_\alpha^{(\gamma)}(\mathbf{r}_\gamma) > 0 \end{cases} \\
\chi_\alpha^{(\gamma)}(\mathbf{r}_\gamma) &= -u_\alpha(\mathbf{r}_\gamma)/k_BT \\
&\quad + \left\{H_\alpha^{\rm ref}(\mathbf{r}_\gamma) + \Delta H_\alpha^{(\gamma)}(\mathbf{r}_\gamma)\right\} - \left\{C_\alpha^{\rm ref}(\mathbf{r}_\gamma) + \Delta C_\alpha^{(\gamma)}(\mathbf{r}_\gamma)\right\},
\end{aligned} \tag{5.8}$$

where k_B is Boltzmann's factor. $u_\alpha(\mathbf{r}_\gamma)$ is the intermolecular potential function between solute and the solvent site α, which is evaluated on the grid points around solute site γ.

Hypernetted chain (HNC) closure is another popular equation used in integral equation theories for liquids. Site-site correlation functions, such as radial distribution functions, can be evaluated well by HNC closure. KH closure sometimes greatly underestimates the site-site

correlation functions, but the process of the numerical solution is much stabler than that of HNC closure.

The flow chart of the present method is shown in Scheme 5.1.



Scheme 5.1

5.3 Computational details

The reference correlation functions were calculated by 1D RISM/KH procedure [24, 25]. Using the converged 1D direct correlation function, the reference correlation functions H_α^{ref} and C_α^{ref} were evaluated. The residual correlation functions were then calculated by Eqs. 5.5, 5.6 and 5.7, coupled with Eq. 5.8. In this approach, the calculation was performed on logarithmic grid for radial part and the Lebedev grid [26] for angular part. With the grid set, the convolution integral in Eq. 5.7 can be calculated by spherical Bessel transformation [27]. To reduce the computational cost per one CPU, calculation was parallelized with MPICH2 [28]. The solvation structure was evaluated with Eqs. (1) and (3) from the obtained residual correlation functions. For visualization of 3D solvation structure, VMD software [29] was used.

5.4 Results and Discussion

The geometry of bR was taken from the PDB data (PDBID: 1c3w) [5] and OPLS parameters [30] were employed. All the water molecules inside the protein in PDB data were removed before the computation. SPC-like model of water was employed [31] with a correction concerning the Lennard-Jones parameters of the hydrogen sites (σ=1.0Å, ϵ=0.056 kcal mol^{-1}). bR is a large molecule and the number of the solute sites is 2221 for the parameter set. However, the required memory size of the present method was about only 850 MB per one CPU, meaning the allowance to perform the computation even with PC cluster. It is noted bR is in the cell membranes in reality, but they are ignored in the the present computation. We believe that the surrounding water or membranes have no effect on the water molecules *inside* the protein.

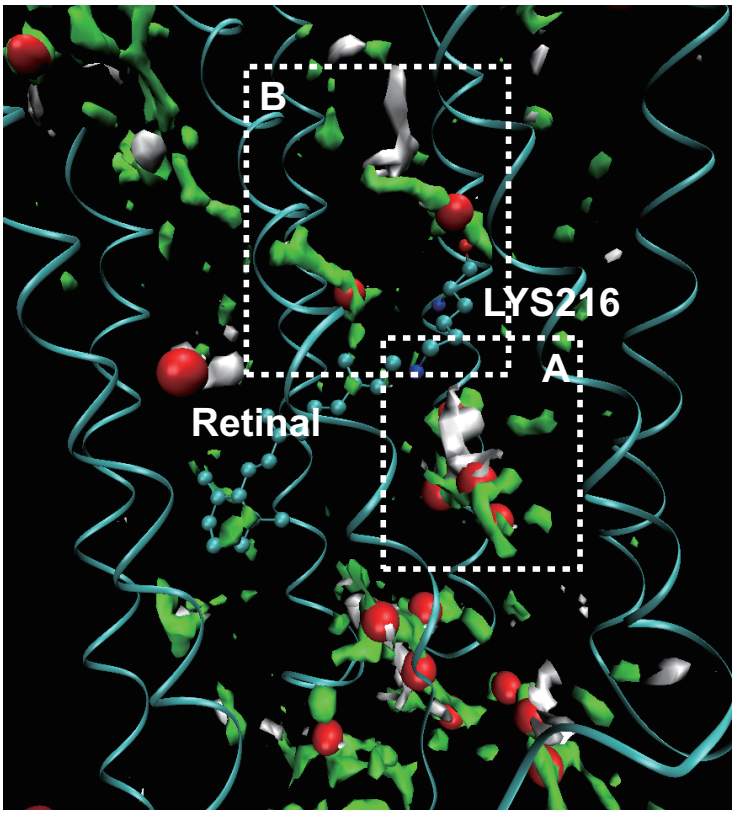


Figure 5.1: 3D distribution of waters inside bR. The green (white) regions correspond to the area where the distribution function of water oxygen (hydrogen) site is larger than 3.2. The bound waters detemined by X-ray crystallography are represented by red spheres.

bR contains all-*trans* retinal, which binds covalently to Lys216 through a protonated Schiff

base linkage. The distribution of waters around retinal calculated by the present method and the water molecules obtained by X-ray diffraction data [5] are shown in Fig. 5.1. Except for Lys216 and retinal, bR is represented by ribbon for the sake of viewability. The scattered green (white) areas shown in the figure indicate where the distribution function of water oxygen (hydrogen) site is greater than the threshold value, 3.2. The bound waters determined by X-ray crystallography are represented by red spheres. Several positions of the comparatively-large area coincide with the experimental data, suggesting that the water distribution is correctly reproduced by the present method. The distribution of waters are not continuous and intermitted by residues. In the case of aquaporins, the bound waters are continuously distributed through out the channel [32]. This may be the large difference between a pump and a channel.

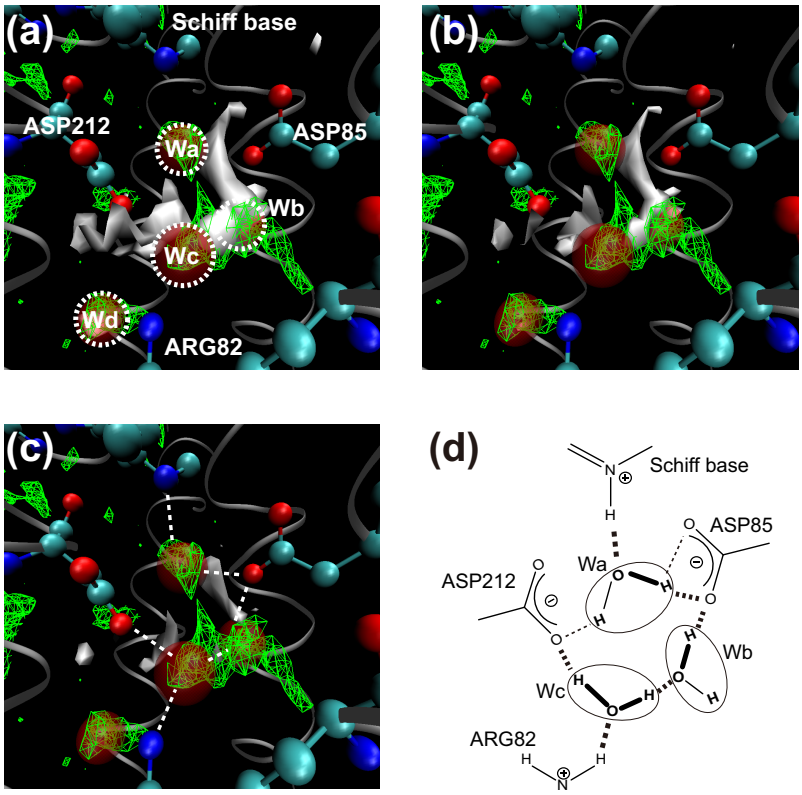


Figure 5.2: 3D solvation structure of water oxygen (green) and hydrogen (white) in **A**. The positions of bound waters determined by X-ray crystallography are shown with dashed line. The white surface show the areas where the distribution of water hydrogen is larger than 2.2 (a), 3.2 (b), and 4.2 (c). In these panels, distribution of water oxygen larger than 3.2 is shown. In panel (c), possible hydrogen bondings are depicted with dashed line. Schematic drawing of this area is shown in (d).

Let us look at closely the regions **A** and **B**, which are enclosed with dashed lines. Fig. 5.2 focuses the water distribution in region **A**, around Schiff base. The scattered green areas shown in the figure indicate where the distribution function of water oxygen site is greater than the threshold value, 3.2. To display the strength of hydrogen bondings clearly, distributions of water hydrogen greater than 2.2, 3.2, and 4.2 are respectively shown in Figs. 5.2(a), 5.2(b), and 5.2(c). The positions of bound waters obtained by X-ray diffraction are also shown with dashed line, **Wa**, **Wb** and **Wc** [5]. Conspicuous localized distributions of oxygen and hydrogen are found surrounded by LYS216, ASP85 and ASP212, which coincides with the result obtained by X-ray crystallography. Shibata et al. proposed that waters **Wa** and **Wb** strongly bind with oxygen site of ASP85, and water **Wc** strongly binds with ASP212 from the FTIR studies [33, 34]. These strong hydrogen bondings are found in Fig. 5.2(c). The broad distributions in Figs. 5.2(a) and 5.2(b) indicate the fluctuation of waters because the present calculation was performed at the condition of room temperature. The schematic picture of the bound waters drawn from these figures is illustrated in Fig. 5.2(d). The thick and thin dotted lines show the strong and weak hydrogen bondings, respectively, which is in good agreement with the network reported by Shibata et al. In the neighborhood of ARG82, the distribution of oxygen site is found (**Wd**) but no hydrogen site can be seen, at least, with the threshold, 4.2. This means the water **Wd** is captured by the residue but its orientation is relatively free compared to aforementioned water molecules, **Wa**, **Wb**, and **Wc**.

Fig. 5.3(a) shows the distributions of water oxygen (green) and water hydrogen (white) sites in region **B**, upside the retinal. There are two main diffuse solvation structures (distributions **I** and **II**). By X-ray crystallography, **We** and **Wf** are detected in **I** and **II**, respectively. **We** in the distribution **I** links ALA215 (in helix G) and TRP182. Schulten et al. reported another water molecule in the vicinity of **We** (see Fig. 1 in their work [11]), although no water molecule is reported in the X-ray crystallography except for **We** [5]. Fig. 5.3(a) shows that there are interaction between water and THR178 (thick dotted line), which may correspond to the water reported by Schulten et al. [11]. Another main distribution (**II**) is continued from LYS216 to ASP96 and THR46 in the vicinity of the helix G backbone, which is distorted from stan-

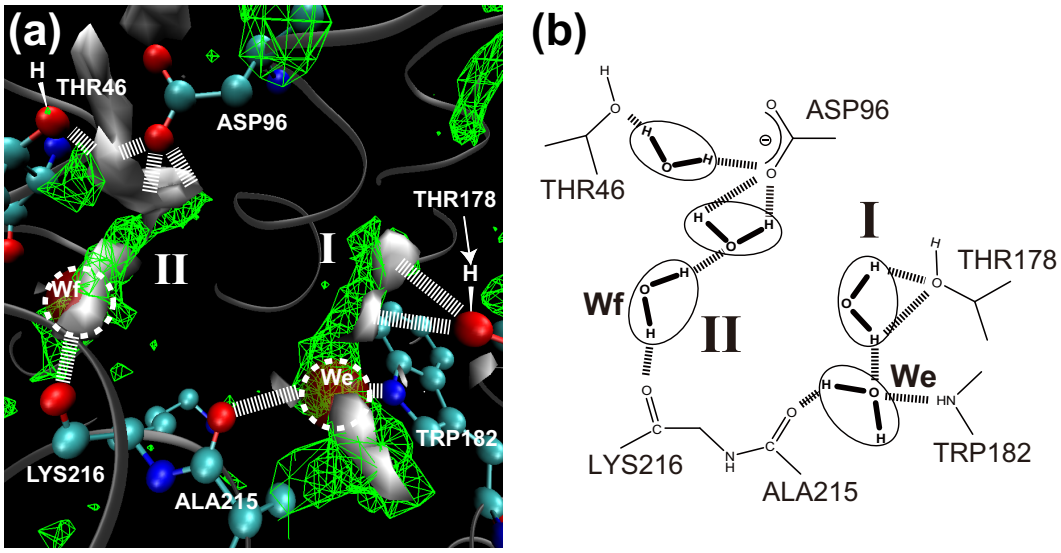


Figure 5.3: (a) 3D solvation structure of water oxygen (green) and hydrogen (white) in **B**. Distributions of water oxygen and water hydrogen larger than 2.5 are shown. (b) Schematic drawing of bound waters proposed from panel (a).

dard α-helical conformation [5]. The water **Wf** binds with the carbonyl group of LYS216. Humphrey et al. proposed other two bound waters in this area [10]. The broad green and white distributions correspond to these waters. The hydrogen bonded waters are shown in Fig. 5.3(b).

Both of **I** and **II** make hydrogen bonding network from the Retinal to ASP96 and to THR46, which is consistent with the previous works' conclusions [5, 10, 11, 14].

5.5 Concluding Remarks

Fragment 3D-RISM was applied to the calculation of the distribution of bound waters in Bacteriorhodopsin. The computed distributions show good agreement with those by X-ray diffraction experiment. The method is highly parallelizable and can sufficiently reduce the required computational cost and time while adequate distribution of water molecules are obtained.

In the neighborhood of the Schiff base, several water molecules captured by residues were found. ASP85 and ASP212 obviously accept hydrogen bonding from neighbor waters because

the direction of oxygen-hydrogen bond in water can be discriminated from the distribution. On the other hand, the water near ARG82 is also captured by the residue but rather freely oriented. Near the G-helix backbone to ASP96, two largely continuous distributions were seen. They are consistent with the bound water molecules reported in molecular simulation study.

The fragment 3D-RISM method is highly efficient with capability to predict the solvation structure concerning sufficiently large bio-molecules.

Bibliography

[1] M. Chaplin, Nature Rev. Mol. Cell Biol. **7**, 861 (2006).

[2] R. A. Engh, H. Brandstetter, G. Sucher, A. Eichinger, U. Baumann, W. Bode, R. Huber, T. Poll, R. Rudolph, W. von der Saal, Structure **4**, 1353 (1996).

[3] J. B. Finley, V. R. Atigadda, F. Duarte, J. J. Zhao, W. J. Brouillette, G. M. Air, M. Luo, J. Mol. Biol. **293**, 1107 (1999).

[4] E. Mayer, Protein Sci. **1**, 1543 (1992).

[5] H. Luecke, B. Schobert, H.-T. Richter, J.-P. Cartailler, J. K. Lanyi, J. Mol. Biol. **291**, 899 (1999).

[6] S. O. Smith, J. Lugtenburg, R. A. Mathies, J. Membr. Biol. **85**, 95 (1985).

[7] H. Kandori, Biochim. Biophys. Acta **1460**, 177 (2000).

[8] S. Hayashi, I. Ohmine, J. Phys. Chem. B **104**, 10678 (2000).

[9] S. Hayashi, E. Tajkhorshid, H. Kandori, K. Schulten, J. Am. Chem. Soc. **126**, 10516 (2004).

[10] W. Humphrey, I. Logunov, K. Schulten, M. Sheves, Biochemistry **33**, 3668 (1994).

[11] J. Baudry, E. Tajkhorshid, F. Molnar, J. Phillips, K. Schulten, J. Phys. Chem. B **105**, 905 (2001).

[12] F. Zhou, A. Windemuth, K. Shulten, Biochemistry **32**, 2291 (1993).

[13] M. Nina, B. Roux, J. C. Smith, Biophys. J. **68**, 25 (1995).

[14] B. Roux, M. Nina, R. Pomès, J. C. Smith, Biophys. J. **71**, 670 (1996).

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

[16] D. Beglov, B. Roux, J. Phys. Chem. B **101**, 7821 (1997).

[17] S. Phongphanphanee, N. Yoshida, F. Hirata, Chem. Phys. Lett. **449**, 196 (2007).

[18] T. Imai, R. Hiraoka, T. Seto, A. Kovalenko, F. Hirata, J. Phys. Chem. B **111**, 11585 (2007).

[19] T. Imai, R. Hiraoka, A. Kovalenko, F. Hirata, Proteins-Structure Function and Bioinformatics **66**, 804 (2007).

[20] T. Imai, R. Hiraoka, A. Kovalenko, F. Hirata, J. Am. Chem. Soc. **127**, 15334 (2005).

[21] N. Yoshida, S. Phongphanphanee, F. Hirata, J. Phys. Chem. B **111**, 4588 (2007).

[22] D. Yokogawa, H. Sato, T. Imai, S. Sakaki, *to be submitted*.

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

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

[25] *Molecular Theory of Solvation*, edited by F. Hirata (Kluwer, Dordrecht, 2003).

[26] V. I. Lebedev, D.N. Laikov, Dokl. Math. **59**, 477 (1999).

[27] J. D. Talman, J. Comput. Phys. **29**, 35 (1978).

[28] MPICH2 High-performance and widely portable implementation of MPI http://www.mcs.anl.gov/research/projects/mpich2/index.php.

[29] W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graphics **14**, 33 (1996).

[30] W. L. Jorgensen, *OPLS and OPLS-AA Parameters for Organic Molecules, Ions, and Nucleic Acids* (Yale University, 1997).

[31] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces* , edited by B. Pullman (Reidel, Dordrecht, 1981).

[32] S. Phongphanphanee, N. Yoshida, F. Hirata, J. Am. Chem. Soc. **130**, 1540 (2008).

[33] M. Shibata, T. Tanimoto, H. Kandori, J. Am. Chem. Soc. **125**, 13312 (2003).

[34] M. Shibata, H. Kandori, Biochemistry **44**, 7406 (2005).

Part II

Quantum chemical calculation with solvation effect

Chapter 6

Localization or delocalization in electronic structure of Creutz-Taube-type complexes in aqueous solution

6.1 Introduction

Mixed-valence complexes containing several metal centers with different oxidation state have received intense theoretical and experimental interests because of their flexible electronic structures and potential ability of molecular electronics [1]. Their electronic structures are explained in terms of a superposition of two localized electronic structures. Robin and Day classified mixed-valence complexes into three classes, namely classes I, II, and III, considering the strength of metal-metal interaction which determines the magnitude of mixing of the two localized electronic structures. In class I, the metal-metal interaction is negligibly weak and the distribution of 'excess electron' or 'hole' is completely localized upon one of the metal centers. In class III, the interaction is strong enough and the distribution is fully delocalized. The interaction in class II is intermediate between them. This classification of the mixed-valence complexes is discussed in many theoretical calculations and such experimental measurements as intervalence charge transfer spectra (IVCT) [1–4]. Both experimental and theoretical works for mixed-valence complexes were summarized in detail by Demadis *et al.* [5] Recently, Reimers *et al.* discussed electronic structure and some physical properties using reorganization energy [6].

Creutz-Taube complex, $[(\mathrm{NH_3})_5\mathrm{Ru-pyrazine-Ru}(\mathrm{NH_3})_5]^{5+}$ (**1** in Fig. 6.1) [7, 8], is one of the typical mixed-valence complexes. There has been heated controversy over how much

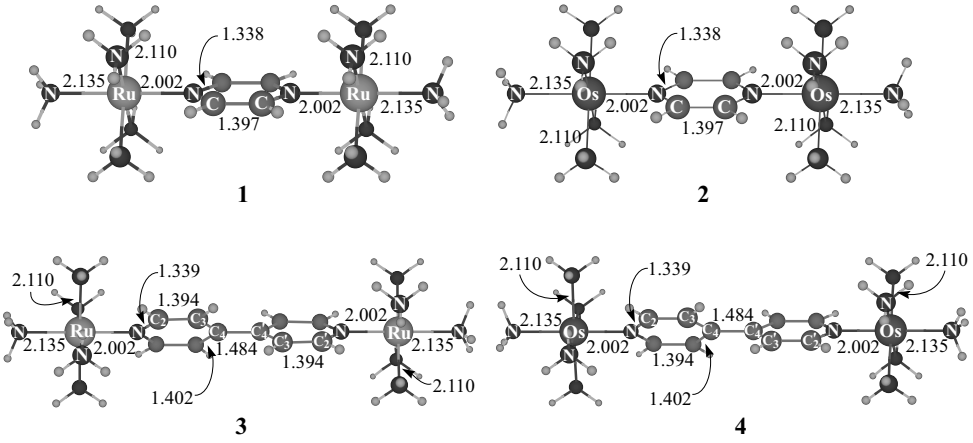


Figure 6.1: Geometries of $[(NH_3)_5Ru-\text{pyrazine}-Ru(NH_3)_5]^{5+}$ (**1**), $[(NH_3)_5Os-\text{pyrazine}-Os(NH_3)_5]^{5+}$ (**2**), $[(NH_3)_5Ru(4,4'-\text{bipyridine})Ru(NH_3)_5]^{5+}$ (**3**), and $[(NH_3)_5Os(4,4'-\text{bipyridine})Os(NH_3)_5]^{5+}$ (**4**) used in this work. (unit; Å)

localized its electronic structure is. In this regard, many experimental and theoretical works have been performed so far to understand the electronic structure. For instance, the electronic structure was investigated by Stark effect [9], near-IR-vis spectra [10], and IVCT spectra. [2] Creutz discussed the relationship between the physical properties of these complexes and the mechanisms of electron transfer processes in her recent review of d^6-d^5 iron, ruthenium and osmium complexes, and showed that Creutz-Taube complex is delocalized while the larger bipyridine-bridged complex is localized in aqueous solution [8]. Density functional theory (DFT) [1, 11–13], MP2 [13], and complete active space SCF (CASSCF) calculations [14] were performed as well. All these studies indicated that the electronic structure is delocalized in this complex due to its strong metal-metal interaction; in other words, this complex belongs to class III.

On the other hand, the electronic structure of 4, 4′-bipyridine-bridged dinuclear Ru complex, $[(NH_3)_5Ru(4,4'-\text{bipyridine})Ru(NH_3)_5]^{5+}$ (**3** in Fig. 6.1), was reported to be quite different from that of **1**. Stark effect [9] and IVCT spectra [3, 4] indicated that the metal-metal interaction of **3** is weak and the electronic structure is substantially localized. Ferreti *et al*. [15] explained this electronic structure and visible spectra by using a four-site vibronic model. Marcus-Hush theory was also applied to evaluate the IVCT spectra of **3** [16–19]. Besides these

studies, only a few computational studies of **3** have been reported to our knowledge, in which DFT [1], CNDO/S [14], and CI methods [20] have been used. In these previous studies, solvation effects were not taken into consideration except for one pioneering work [20], in which the continuum model was employed to incorporate solvent effect. In reality, however, solvation effects should be taken into consideration because the localized electronic structure is significantly stabilized by polar solvent. Another important issue is to consider its multi-reference nature in the electronic structure, which is closely related to mixing of localized wavefunctions. Standard methods such as CASSCF might not be applicable to these mixed-valence complexes in reasonable computing time because of their large sizes; see **3** for example.

In the present article, we theoretically investigated **1**, **3**, pyrazine-bridged dinuclear Os complex, $[(\mathrm{NH_3})_5\mathrm{Os-pyrazine-Os}(\mathrm{NH_3})_5]^{5+}$(**2** in Fig. 6.1), and 4, 4'-bipyridine-bridged dinuclear Os complex, $[(\mathrm{NH_3})_5\mathrm{Os}(4,4'-\mathrm{bipyridine})\mathrm{Os}(\mathrm{NH_3})_5]^{5+}$ (**4** in Fig. 6.1). The complexes, **2** and **4**, are not known experimentally to our knowledge. In fact, to understand the true nature of the mixed-valence complexes, the consideration of vibration coupling and time scale of solvation is indispensable, as reviewed recently. However, it is also important to evaluate theoretically the electronic structure of real molecules of mixed-valence complexes without modeling and static solvation effect. In this work, we evaluated some factors which determine the localization/delocalization of the ions without modeling and tried to relate them with fundamental parameters such as overlap and energy gap. Though our study does not incorporate vibration-coupling and solvation time scale [5], we believe the knowledge of relation between fundamental parameters and localization/delocalization nature is also worthwhile to understand these mixed-valence complexes.

6.2 Method and Computations

6.2.1 Method

As described above, the metal-metal coupling in **1** is considered very strong and that of **3** is considered very weak. The DFT method can be applied to the complexes with strong metal-metal interaction but seems to be difficult to apply to the mixed-valence complexes

with weak metal-metal interaction because the DFT method tends to overestimate delocalized character [13]. CASSCF and CASPT2 methods are believed to be most reliable for this type of compound. However, it is noted also that the mixed-valence complexes are too large to apply the CASSCF method.

In the present work, we employed a method proposed by Farazdel *et al.* [21] to treat the multi-reference nature of the wave function. The first step of this method is to calculate two wave functions, Ψ_A and Ψ_B, by the UHF method with the same geometry, where symmetry-broken UHF orbitals are employed [22][(a),(b)]. In Ψ_A, the excess electron is localized on one metal center, while in Ψ_B it is localized on the other metal center. These Ψ_A and Ψ_B correspond to the non-orthogonal 'diabatic states' [23]. The second step is to construct 'adiabatic' wave functions, Ψ_{E_+} and Ψ_{E_-}, from Ψ_A and Ψ_B, as follows;

$$\Psi_{E_{+/-}} = C_A\Psi_A + C_B\Psi_B. \tag{6.1}$$

Coefficients and the energies $E_{+/-}$ of the adiabatic states can be obtained by solving the following secular equation;

$$\begin{vmatrix} H_{AA} - E & H_{AB} - ES_{AB} \\ H_{AB} - ES_{AB} & H_{BB} - E \end{vmatrix} = 0, \tag{6.2}$$

where $H_{AA} = \langle\Psi_A|H|\Psi_A\rangle, H_{BB} = \langle\Psi_B|H|\Psi_B\rangle, H_{AB} = \langle\Psi_A|H|\Psi_B\rangle$, and $S_{AB} = \langle\Psi_A|\Psi_B\rangle$.

The solvation effects were evaluated by considering the interaction of the point charge and the dipole moment of solute with reaction field, in which the solute was placed in a spherical cavity immersed in a continuous medium with a dielectric constant ϵ. In this situation, the solvation free energy change ΔG is given by eq. 6.3,

$$\Delta G = -\frac{\epsilon - 1}{2\epsilon}\frac{q^2}{a} - \frac{\epsilon - 1}{2\epsilon + 1}\frac{\mu^2}{a^3} \quad , \tag{6.3}$$

where q is total charge, μ is dipole moment, and a is a radius of the spherical cavity which is determined by the method of Wong *et al.* [24]. The ϵ value is taken to be 78.39 throughout the present study to represent aqueous environment. Because the complexes examined possess positive charges, the dipole moment was evaluated with the procedure of Wong *et al.* They divided dipole moment of charged molecule into two parts, μ_e and μ_N, which correspond to

the dipole moment of electrons and that of nuclear charges, respectively. Total dipole moment of the molecule μ is represented by eq. 6.4,

$$\mu = \frac{\mu_e(Q + n_e)}{n_e} + \mu_N, \tag{6.4}$$

where Q and n_e are the total charge and the number of electrons, respectively. In the calculation of μ_e, we used density matrix of the total wave function,

$$P^C_{\mu\nu} = C_A^2 P^A_{\mu\nu} + C_B^2 P^B_{\mu\nu} + 2C_A C_B \det(\mathbf{U}) \det(\mathbf{V}^\dagger) P_{\mu\nu}, \tag{6.5}$$

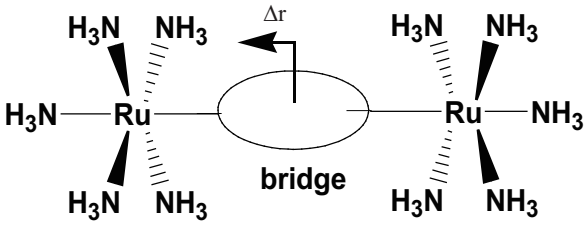
where $P_{\mu\nu}$ is "generalized" density matrix, $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices of the corresponding transformation [25], all of which are defined according to Farazdel *et al* [21]. $P^A_{\mu\nu}$ and $P^B_{\mu\nu}$ are usual density matrices of A and B states, respectively. μ_e was calculated from the partial charge on all the atoms, which is determined so as to reproduce the electrostatic potential evaluated with wavefunctions at each grid point around the solute molecule.

6.2.2 Computational Details

To calculate 'adiabatic' states, we used GAMESS program package [26] with several modifications by us. In all the calculations, core electrons of Ru (up to 3d) and Os (up to 4f) were replaced with effective core potentials (ECPs), where (341/321/31) set was used for valence electrons of Ru and (341/321/21) set was used for those of Os [27]. For C, N, and O, the 6-31G(d) sets were employed and for H the 6-31G set was employed. To check the reliability of this basis set system, electron-transfer matrix element was evaluated with larger basis sets, in which all electron basis sets, [84333/843/75/1] [28] augmented with an f-function (α= 1.235) [29], 6-311G(d), and 6-31G were used for Ru, N, C, and H, respectively. These two different basis set systems presented almost the same value of the electron-transfer matrix elements [21, 30]. Thus, the smaller basis set system was employed throughout the present study.

In **1**, the $\mathrm{Ru-NH_3}$ and $\mathrm{Ru-pyrazine}$ bond distances were taken from the X-ray crystal structure [2], while geometries of $\mathrm{NH_3}$ and pyrazine were optimized by the DFT(B3LYP) [31] method since structural data are not available for these moieties. In the geometry optimization, we used Gaussian 98 program [32]. In **2**, **3**, and **4**, $\mathrm{metal-N(ammonia)}$, $\mathrm{metal-N(pyrazine)}$,

and $\mathrm{metal} - \mathrm{N}(4, 4' - \mathrm{bipyridine})$ distances were taken to be the same as those of **1** because there is no experimental data and our purpose is to compare them in the same situation.



Scheme 6.1

The energy curves were calculated as function of the displacement (Δr) of the bridging ligand from the midpoint of the two metal centers (see Scheme 6.1 for Δr). Along the lines of their procedure, we calculated the diabatic potential energy surface, assuming that the $\mathrm{metal} - \mathrm{NH_3}$ distance did not change along the antisymmetric stretching motion of the bridging ligand. These assumption is reasonable because the displacements of $\mathrm{metal} - \mathrm{NH_3}$ groups have little influence to the potential energy surface [33]. In **3**, the dihedral angle in 4, 4'-bipyridine was fixed to be 40 degree, which was optimized by changing the dihedral angle with an interval of 10 degree. This angle is the same as that reported previously [1]. The effect of dihedral angle on electronic structures will be discussed below.

6.3 Results and Discussion

6.3.1 Potential Energy Curve of Diabatic States

As shown in Fig. 6.2, two symmetry-broken wave functions Ψ_A and Ψ_B are calculated with the UHF method along the reaction coordinate Δr. These two states are degenerate at the symmetrical structure, Δr=0. As shown in Fig. 6.2, SOMOs of these states are almost localized on each metal center. Here, Ψ_A represents the state in which SOMO is almost localized on the metal of the left hand side, and Ψ_B represents the other state. These are 'diabatic' states.

The 'adiabatic' states of ground and excited states are calculated in gas phase by using eq. 6.2, as shown in Fig. 6.3. In all these complexes, the adiabatic state exhibits a single

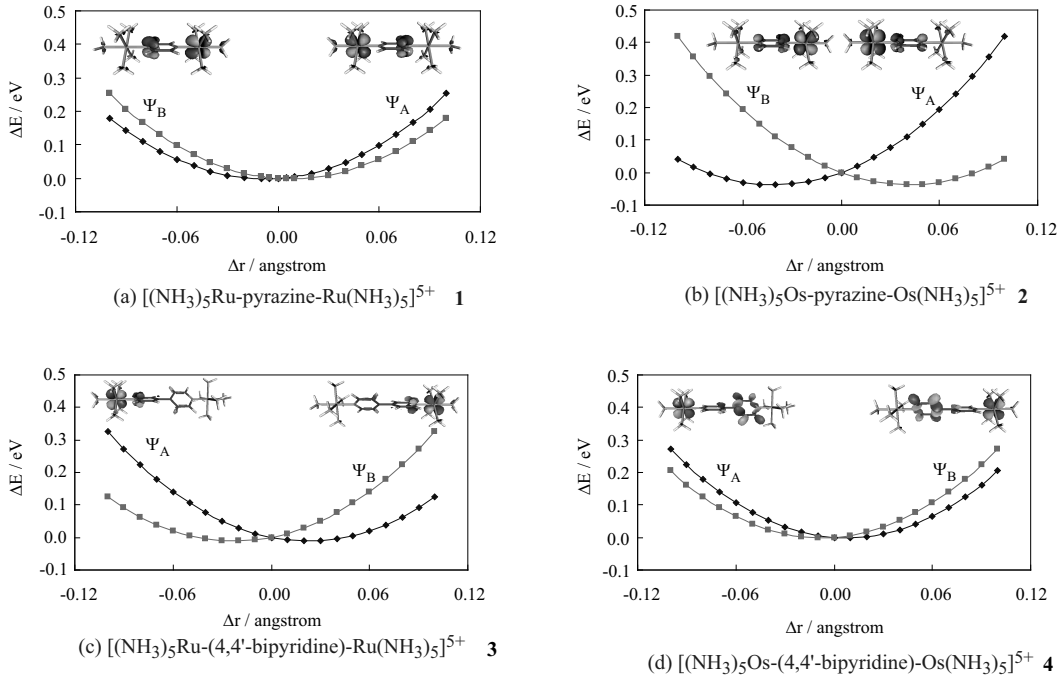


Figure 6.2: Energy curves and SOMOs of two symmetry-broken wave functions, Ψ_A and Ψ_B; (a) $[(NH_3)_5Ru-pyrazine-Ru(NH_3)_5]^{5+}$ **(1)**, (b) $[(NH_3)_5Os-pyrazine-Os(NH_3)_5]^{5+}$ **(2)**, (c) $[(NH_3)_5Ru(4,4'-bipyridine)Ru(NH_3)_5]^{5+}$ **(3)**, and (d) $[(NH_3)_5Os(4,4'-bipyridine)Os(NH_3)_5]^{5+}$ **(4)**.

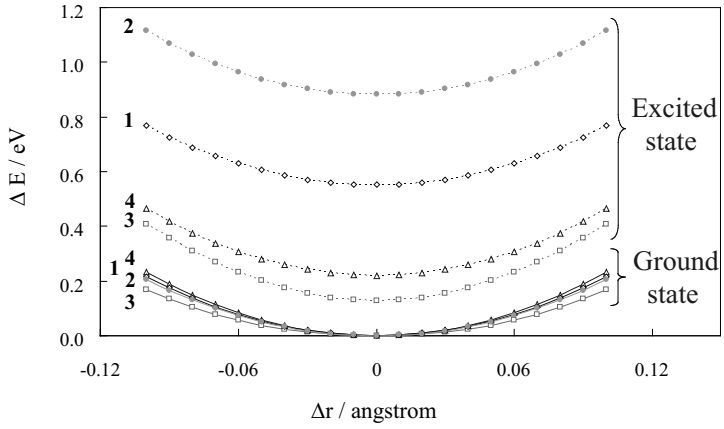


Figure 6.3: Energy curves of diabatic states. Solid and dotted lines represent the energy curves of ground state and excited state, respectively; $\diamond$ $[(NH_3)_5Ru-pyrazine-Ru(NH_3)_5]^{5+}$ **(1)**, $\bullet$ $[(NH_3)_5Os-pyrazine-Os(NH_3)_5]^{5+}$ **(2)**, $\square$ $[(NH_3)_5Ru(4,4'-bipyridine)Ru(NH_3)_5]^{5+}$ **(3)**, and $\triangle$ $[(NH_3)_5Os(4,4'-bipyridine)Os(NH_3)_5]^{5+}$ **(4)**.

minimum at the symmetrical structure (Δr=0), indicating that the electronic structures of all these complexes are delocalized in gas phase. In Fig. 6.3, we can see the energy splitting between two adiabatic state increases in the order **3** $<$ **4** $<$ **1** $<$ **2**. According to the 'two-state' model, the strength of the mixing depends on the difference in energies ($\Delta H = H_{BB} - H_{AA}$) and overlap integral ($S_{AB}$) between two 'diabatic' states: the larger the overlap is and the smaller the difference in energy is, the mixing becomes large. At the seam of crossing between two states, electron transfer matrix, V, is discussed in terms of overlap S_{AB} between two states. The value, V, is calculated by eq. 6.6 using the important parameters, S_{AB}, H_{AB}, H_{AA}, and H_{BB} in eq. 6.2;

$$V = (1 - S_{AB}^2)^{-1} \left[H_{AB} - S_{AB}(H_{AA} + H_{BB})/2 \right] \tag{6.6}$$

Actually, V linearly depends on S_{AB}, as shown in Fig. 6.4. Thus, it is worthwhile to clar-

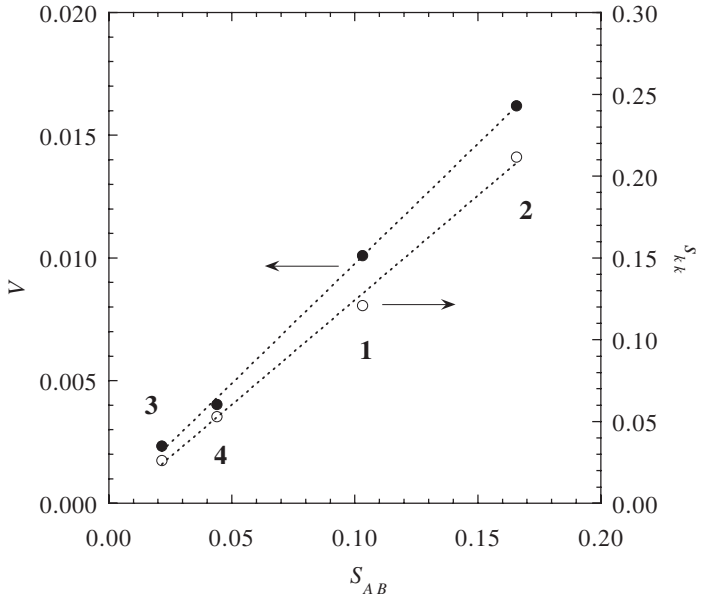


Figure 6.4: Correlation (○) between electron transfer matrix, V, and S_{AB} and correlation (●) between s_{kk} and S_{AB}. Dotted lines are determined by least-square method.

ify what is the origin of the difference in S_{AB}, or the strength of the state-mixing in these complexes examined. In the present 'two-state' model, S_{AB} is defined, as follows [21]:

$$S_{AB} = \langle \Psi_A | \Psi_B \rangle = (\det \mathbf{U})(\det \mathbf{V}^\dagger) \prod_{i=1}^{N} s_{ii} \quad , \tag{6.7}$$

where **U** and **V** are unitary matrices of the corresponding transformation. Notations used here, except for s_{ii}, are the same as those in reference. s_{ii} represents the overlap between corresponding orbitals, $\hat{a}_i$ and $\hat{b}_i$, belonging to each diabatic states.

$$s_{ii} = \langle \hat{b}_i | \hat{a}_i \rangle \quad . \tag{6.8}$$

We found that all s_{ii} is almost 1.0 except for one overlap term, s_{kk}, between two specific orbitals. Consequently, S_{AB} mainly depends on this overlap, as shown in Fig. 6.4.

$$S_{AB} \propto s_{kk} = \langle \hat{b}_k | \hat{a}_k \rangle \quad . \tag{6.9}$$

These key orbitals $\hat{a}_k$ and $\hat{b}_k$ are the corresponding orbitals that are almost same with canonical β-spin HOMO orbitals in **1** $\sim$ **4**. The similar relation was previously reported by Koga *et al* [34]. These $\hat{a}_k$ and $\hat{b}_k$ orbitals are mirror image to each other; one of them is localized on Ru^1/Os^1 site and the other is on Ru^2/Os^2 site. In these orbitals, the d_π orbital expands to the bridge part, as illustrated in Fig. 6.5. Apparently, the overlap integral between $\hat{a}_k$ and $\hat{b}_k$ in **1** and **2** is much larger than in **3** and **4**; in the latter complexes, these orbitals are completely separated and localized on each metal center.

The s_{kk} term is further divided into six parts, as follows [34].

$$\begin{aligned} s_{kk} &= \sum_{I}^{b,m,o} \sum_{J}^{b,m,o} s_{kk}^{IJ} = \sum_{I}^{b,m,o} s_{kk}^{II} + 2 \sum_{I<J}^{b,m,o} s^{IJ} \\ &= s_{kk}^{b-b} + s_{kk}^{m-m} + s_{kk}^{o-o} + s_{kk}^{b-m} + s_{kk}^{b-o} + s_{kk}^{m-o}, \end{aligned} \tag{6.10}$$

where “b”, “m”, and “o” stand for bridge ligand, metal center, and remaining part, respectively. In all complexes, “m-m”, “b-m”, and “b-b” pairs provide dominant contributions to s_{kk}, as shown in Fig. 6.6.

The overlap s_{kk}^{m-m} depends on the metal-metal distance; the longer the distance is, the less the overlap is. In **3** and **4**, the distance is about twice as long as that of **1** and **2**. The longer metal-metal distance in **3** and **4** leads to the significantly smaller s_{kk}^{m-m} than that of **1** and **2**. The overlap s_{kk}^{b-m} between the metal part and the bridge part is mainly determined by the overlap between the metal d_π orbital in $\hat{b}_k$ ($\hat{a}_k$) and the π and π^* orbitals of bridging ligand in

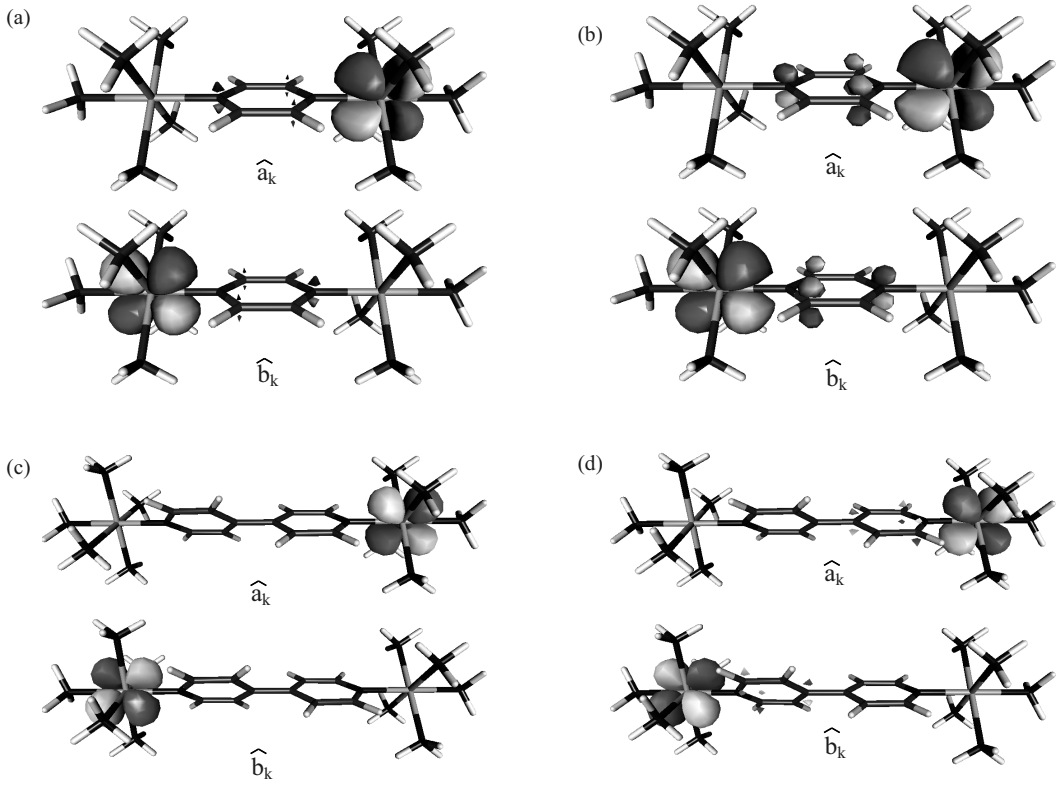


Figure 6.5: Corresponding orbitals $\hat{a}_k$ and $\hat{b}_k$, which are almost the same as the canonical β-spin HOMO orbitals; (a) $[(NH_3)_5Ru - \text{pyrazine} - Ru(NH_3)_5]^{5+}$ **(1)**, (b) $[(NH_3)_5Os - \text{pyrazine} - Os(NH_3)_5]^{5+}$ **(2)**, (c) $[(NH_3)_5Ru(4,4' - \text{bipyridine})Ru(NH_3)_5]^{5+}$ **(3)**, and (d) $[(NH_3)_5Os(4,4' - \text{bipyridine})Os(NH_3)_5]^{5+}$ **(4)**.

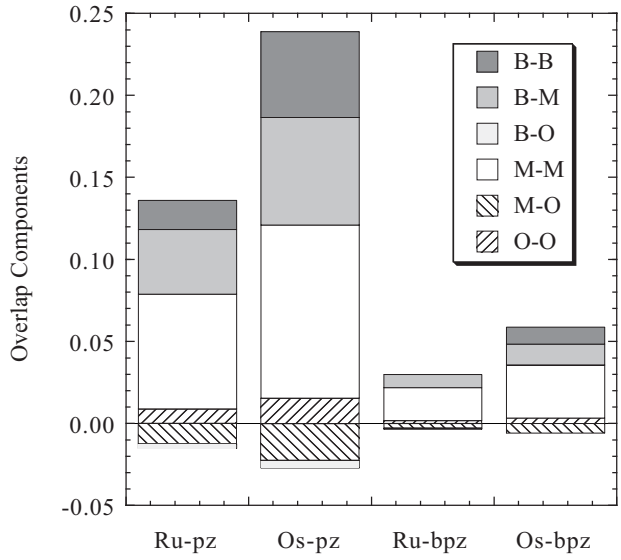


Figure 6.6: Overlap components[a)], s_{kk}^{b-b}, s_{kk}^{b-m}, s_{kk}^{b-o}, s_{kk}^{m-m}, s_{kk}^{m-o}, and s_{kk}^{o-o} of $[(NH_3)_5Ru-$ pyrazine $-\ Ru(NH_3)_5]^{5+}$ **(1)**, $[(NH_3)_5Os-$ pyrazine $-\ Os(NH_3)_5]^{5+}$ **(2)**, $[(NH_3)_5Ru(4,4'-$ bipyridine$)Ru(NH_3)_5]^{5+}$ **(3)**, and $[(NH_3)_5Os(4,4'-$ bipyridine$)Os(NH_3)_5]^{5+}$ **(4)**. a) b, m, and o represent bridge part, metal center, and the other part.

$\hat{a}_k$ ($\hat{b}_k$). Since the d_π orbital of Os more spatially expands than that of Ru [35], the overlaps between the Os d_π orbital and the π^* orbital on bridge ligand of **2** and **4** are larger than those of **1** and **3**. The overlap between 4, 4'-bipyridine π^* orbital of $\hat{b}_k$ ($\hat{a}_k$) and metal d_π orbital of $\hat{a}_k$ ($\hat{b}_k$) in **3** and **4** is much smaller than the overlap between pyrazine $\pi*$ and metal d_π orbitals in **1** and **2**, as easily seen in Figures 5c and 5d. Therefore the order of s_{kk}^{b-m} is $3 < 4 < 1 < 2$.

Interestingly, a remarkable difference in s_{kk}^{b-b} is observed among these complexes, whereas the shapes of the orbitals are very similar to each other. It is likely that because the Os d_π orbital is closer in energy to the pyrazine π^* orbital than the Ru d_π orbital, the π^* orbital more contributes to 'diabatic' state in the Os complex than in the Ru complex [36]. The contribution of π^* orbital to $\hat{a}_k$ and $\hat{b}_k$ was evaluated by the following equation;

$$\phi = C_\pi \phi_\pi + C_{\pi^*} \phi_{\pi^*}, \tag{6.11}$$

where ϕ is the contribution of the bridge moiety to the corresponding orbital $\hat{a}_k$ ($\hat{b}_k$). The ϕ_π and ϕ_{π^*} are canonical orbitals of 4, 4'-bipyridine calculated by the HF method. $C_{\pi^*}^2$ of **3** and **4** are 0.006 and 0.018, respectively. This difference leads to the difference in s_{kk}^{b-b} between **3**

and **4**. In **1** and **2**, $\hat{a}_k$ and $\hat{b}_k$ orbitals exhibit amplitude to some extent on the bridge group, as shown in Figs. 6.5a and 6.5b, which leads to the larger s_{kk}^{b-b} value of **2** than that of **1**.

Summarizing the above discussion, the s_{kk}^{m-m}, s_{kk}^{b-m}, and s_{kk}^{b-b} ($\propto S_{AB}$) increase in the order Ru $<$ Os and in the order **3** and **4** $<$ **1** and **2**. Thus, the energy splitting between two diabatic states increase in the order **3** $<$ **4** $<$ **1** $<$ **2**. These differences in overlap S_{AB} is one of key factors for localized vs delocalized electronic structure in aqueous solution, as will be discussed in the next section.

In Aqueous Solution

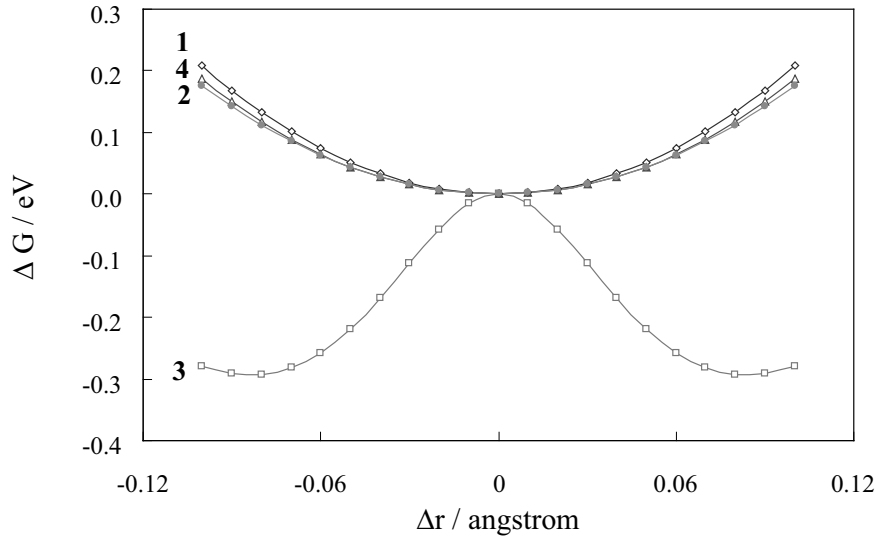


Figure 6.7: Free energy curves of **1**, **2**, **3**, and **4** in aqueous solution; $\diamond$ $[(NH_3)_5Ru - pyrazine - Ru(NH_3)_5]^{5+}$ **(1)**, $\bullet$ $[(NH_3)_5Os - pyrazine - Os(NH_3)_5]^{5+}$ **(2)**, $\square$ $[(NH_3)_5Ru(4,4' - bipyridine)Ru(NH_3)_5]^{5+}$ **(3)**, and $\triangle$ $[(NH_3)_5Os(4,4' - bipyridine)Os(NH_3)_5]^{5+}$ **(4)**.

Free energy curves (FEC) in aqueous solution are shown in Fig. 6.7. In **1** and **2**, FEC possesses a single minimum at the symmetric geometry (Δr=0), as is the case for gas phase. The FEC of **3** has two minima at $\Delta r = \pm 0.08$Å, showing that the electronic structure of **3** is localized in aqueous solution. The key to understand the difference in **3** from the others is dipole moment, which is computed by eq. 6.4. Dipole moment is zero at the point of $\Delta r = 0$ in all the complexes due to the symmetry of the total wavefunction. It increases with increase in Δr. Apparently the dipole moment changes much larger in **3** than in the others, as shown

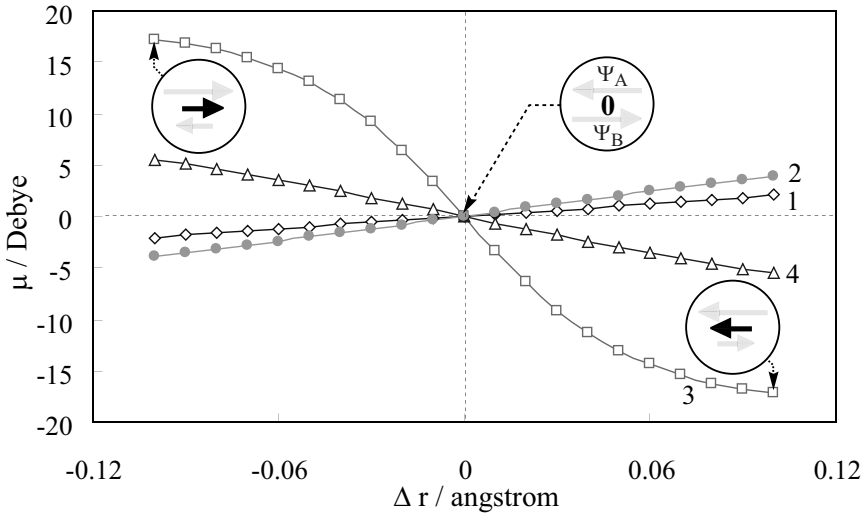


Figure 6.8: Changes of dipole moment along Δr. In solid circle, the dipole moments of 'diabatic' states at Δr = -0.10, 0.0, and + 0.10 are schematically shown; $\diamond$ $[(NH_3)_5Ru-pyrazine-Ru(NH_3)_5]^{5+}$ (**1**), $\bullet$ $[(NH_3)_5Os-pyrazine-Os(NH_3)_5]^{5+}$ (**2**), $\square$ $[(NH_3)_5Ru(4,4'-bipyridine)Ru(NH_3)_5]^{5+}$ (**3**), and $\triangle$ $[(NH_3)_5Os(4,4'-bipyridine)Os(NH_3)_5]^{5+}$ (**4**).

in Fig. 6.8. In **4**, FEC is influenced by the dihedral angle (δ) between two pyridyl rings. When δ is 40 $^\circ$, its electronic structure is delocalized, as shown in Fig. 6.9. However, it becomes localized when δ is 80 $^\circ$. Because the energy difference between the minima at δ =40 $^\circ$and δ =80 $^\circ$, the electronic structure of **4** is between localized and delocalized one (Class II). The effect of the dihedral angle will be discussed below in more detail.

Oh *et al.* studied how much dipole moment of di-ruthenium complexes changes upon going to excited state from ground state in water using electronic absorption (Stark effect) spectroscopy [9]. They reported that the change is about 0 (D) for **1** and 29 (D) for **3**. As clearly shown in Figure 7, the electronic structure of **1** is delocalized at ground state in aqueous solution. In the electronic absorption, the transition should be from the delocalized electronic structure at ground state to the delocalized structure at excited state. On the other hand, the electronic structure of **3** is localized at the equilibrium geometry at ground state in aqueous solution ($\Delta r = 0.08$Å). The dipole moment was evaluated to be -17 (D) and 20 (D) at the ground and excited states, respectively. The calculated change of dipole moment is about 37 (D). These computational results of dipole moment change are consistent with the experimental data.

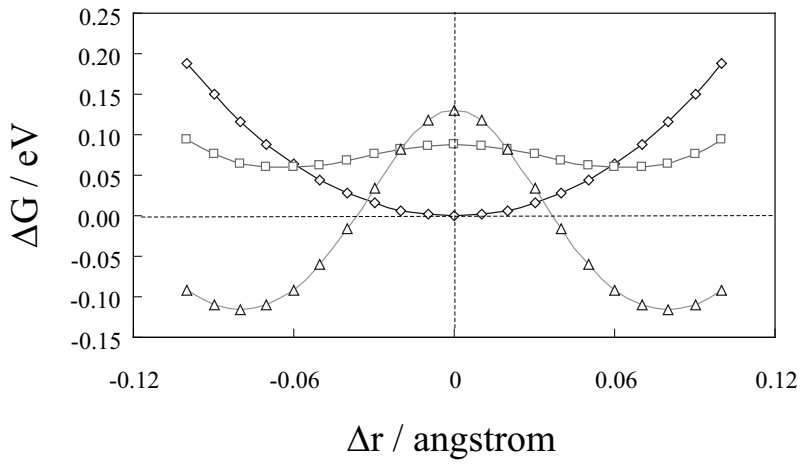


Figure 6.9: Free energy curves of $[(NH_3)_5Os(4,4'-\text{bipyridine})Os(NH_3)_5]^{5+}$ (**4**) at $\delta^{a)}$ =40˚(◇), 70˚(□), and 80˚(△). The standard of free energy is that at 40 ˚and Δr = 0. a) δ is the dihedral angle between two pyridyl planes of 4, 4'-bipyridine.

Change of dipole moment is induced by the mixing ratio of two diabatic states (see eq. 6.5) whose dipole-moment direction is opposite to each other, as illustrated in Fig. 6.8. In the present two-state model, the mixing ratio R depends on H_{AB}, ΔH, and S_{AB}, as represented by eq. 6.12,

$$R = \frac{C_B - C_A}{C_B + C_A} = \sqrt{\frac{1+S_{AB}}{1-S_{AB}}}\tan\theta \quad , \tag{6.12}$$

where θ is given by solving eq. 6.2,

$$\theta = \frac{1}{2}\tan^{-1}\left\{\frac{\sqrt{1-S_{AB}^2}}{2}\frac{\Delta H}{H_{AB} - S_{AB}\left(\frac{H_{AA}+H_{BB}}{2}\right)}\right\} \quad . \tag{6.13}$$

When $\Delta r = 0$, two diabatic states are in the same energy ($\Delta H = 0$), which leads to $R = 0$; this means that two states mix in the same ratio ($C_A = C_B$). In this case, dipole moment is 0. At $\Delta r \neq 0$, on the other hand, the mixing ratio is not equivalent and the dipole moment is induced. As R increases, the localization of adiabatic wave function increases. Fig. 6.10 shows change of R as a function of Δr. In $\Delta r > 0$ the sign of R is positive for **1** and **2** and negative for **3** and **4**. This sign shows which of the state Ψ_A and Ψ_B is dominant in $\Delta r > 0$. In **1** and **2**, Ψ_B is dominant in the adiabatic states, as shown in Figs. 6.2(a) and 6.2(b). On the other hand, Ψ_A is dominant in the adiabatic states of **3** and **4**. One can see that R of **3** changes much larger than those of the others, which means the contribution of one diabatic

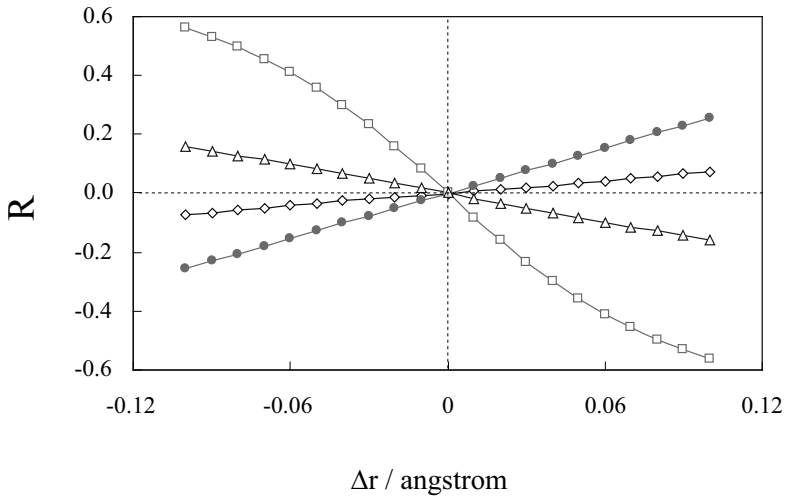


Figure 6.10: Mixing ratio R vs Δr; $\diamond$ $[(NH_3)_5Ru-\text{pyrazine}-Ru(NH_3)_5]^{5+}$, $\bullet$ $[(NH_3)_5Os-\text{pyrazine}-Os(NH_3)_5]^{5+}$, $\square$ $[(NH_3)_5Ru(4,4'-\text{bipyridine})Ru(NH_3)_5]^{5+}$, and $\triangle$ $[(NH_3)_5Os(4,4'-\text{bipyridine})Os(NH_3)_5]^{5+}$.

state considerably increases with increase in Δr; in other words, the adiabatic wavefunction tends to localize on one center in **3** to a more extent than in the others.

Because tan and tan^{-1} are monotonous functions and S_{AB} is very small, the mixing ratio R can be compared with each other using following quantity R';

$$R' = \frac{\Delta H}{H_{AB} - S_{AB}\left(\frac{H_{AA}+H_{BB}}{2}\right)}. \tag{6.14}$$

The larger R' is, the more localized the electronic structure is. As represented by eq. 6.13, the mixing ratio is determined by subtle balance among several parameters such as S_{AB}, H_{AB} etc. It should be emphasized that the solvation energy, which is mainly determined by the dipole moment of the complex, increases enough to stabilize the localized electronic structure when the two coefficients, C_A, C_B, are remarkably different. In **3**, R' is much larger than in the others, because ΔH is the largest and S_{AB} is the smallest, as discussed above. This leads to the much larger dipole moment in **3** than in the others, which further leads to the larger stabilization energy by polar solvent. Thus, the electronic structure of **3** is localized in aqueous solution.

6.3.2 FEC along Rotation of the Bridge Group

In **3** and **4**, two pyridyl rings can rotate around C-N and/or N-N bond axis. The increase in the dihedral angle (δ) decreases the overlap between p_π orbital of C_4 and C_4', which further changes the energy levels of π and π^* orbitals of the bridge group, and therefore, the electronic structure of these complexes is influenced by this rotation. It is interesting to investigate how much the localization/delocalization of the electronic structure depends on the rotation.

FEC of **4** were evaluated at $\delta = 40^\circ$, 70°, and 80°, as shown in Fig. 6.9. Although the electronic structure of **4** is delocalized at $\delta = 40^\circ$, as was discussed above, it is sufficiently localized at 80 $^\circ$. As previously mentioned, the dipole moment, which has great influence on the localization in aqueous solution, is mainly determined by the parameter R' (eq. 6.14). Because the change of the dihedral angle between two pyridyl planes has little influence on the $d_\pi - \pi^*$ interaction, ΔH is almost constant. From eq. 6.14, we can say that the difference in FEC among $\delta = 40^\circ$, 70°, and 80° is mainly governed by the overlap S_{AB} and H_{AB}. Because H_{AB} is almost proportional to S_{AB}, S_{AB} is the main factor that determines the localization/delocalization of these complexes. S_{AB} decreases with an increase in the dihedral angle; for instance S_{AB} is 0.044 at δ =40° ($\Delta r = 0$) but significantly decreases to 0.01 at $\delta = 80^\circ$($\Delta r = 0$). This small S_{AB} at $\delta = 80^\circ$ induces the large dipole moment at $\Delta r \neq 0$, which leads to the localized electronic structure at this angle [37] .

6.4 Conclusions

We have theoretically studied the electronic structures of Creutz-Taube complex and its analogues. They have been attracting great deal of interests in understanding its electronic structure, namely, localization or delocalization. There are two important requirements to understand the electronic structure of the system. One is multi-configurational description in the wave functions, which is caused by a inherent character of mixed-valence metal complex, and the other is solvation effect, which is not negligible. In the present study, we have theoretically investigated these complexes by consideration of "two-state model" based on ab initio molecular orbital theory and dielectric continuum model, and related the localization/delocalization

of the electronic structure with fundamental parameters, such as overlap and energy gaps. Although the calculation is not sufficient for the understanding of the true nature of the ions, our work showed the important factors which determine the localization/delocalization.

It is found that all of the electronic structure of the examined complexes would be delocalized in gas phase, but the electronic structure of **3** with a long bridge, $[(NH_3)_5Ru(4,4'\text{-bipyridine})Ru(NH_3)_5]^{5+}$, shows localized electronic structure in aqueous environment. In **4**, the electronic structure changes as the dihedral angle becomes large. The localized electronic structures of the complexes are interpreted in terms that the magnitude of the mixing of two diabatic states is small; because of large ΔH and small S_{AB}, the mixing ratio R is much larger. Thus, one of two states becomes dominant enough and the dipole moment of the complex significantly increases, which leads to large solvation effects. In the other two complexes, **1** and **2**, π and π^* orbitals in pyrazine and bipyridine interact well with d_π orbital of metal center. As a result, overlap S_{AB} becomes sufficiently large to induce the electron delocalization. In **4**, we wish to propose the possibility that the electronic structure can be designed by introducing some substituents at C_3 and C_3' positions of 4, 4'-bipyridine; such substituents increases the dihedral angle between two pyridyl plane to decrease S_{AB}.

Bibliography

[1] S. B. Braun-Sand and O. Wiest, J. Phys. Chem. A **107**, 285 (2003)

[2] U. Fürholz, H.-B. Bürgi, F. E. Wagner, A. Stebler, J. H. Ammeter, E. Krausz, R. J. H. Clark, M. J. Stead, and A. Ludi, J. Am. Chem. Soc. **106**, 121 (1984).

[3] C. Creutz, Inorg. Chem. **17**, 3723 (1978).

[4] J. E. Sutton, P. M. Sutton, and H. Taube, Inorg. Chem. **18**, 1017 (1979).

[5] K. D. Demadis, C. M. Hartshorn, and T. J. Meyer, Chem. Rev. **101**, 2655 (2001).

[6] J. R. Reimers, Z.-L. Cai, and N. S. Hush, Chem. Phys. **319**, 39 (2005).

[7] C. Creutz and H. Taube, J. Am. Chem. Soc. **95**, 1086 (1973).

[8] C. Creutz, Prog. Inorg. Chem. **30**, 1 (1983).

[9] D. H. Oh, M. Sano, and S. G. Boxer, J. Am. Chem. Soc. **113**, 6880 (1991).

[10] D. H. Oh and S. G. Boxer, J. Am. Chem. Soc. **112**, 8161 (1990).

[11] A. Bencini, I. Ciofini, C. A. Daul, and A. Ferretti, J. Am. Chem. Soc. **121**, 11418 (1999).

[12] Z. Chen, J. Bian, L. Zhang, and S. Li, J. Chem. Phys. **111**, 10926 (1999).

[13] J. Hardesty, S. K. Goh, and D. S. Marynick, J. Mol. Struct. (Theochem) **588**, 223 (2002).

[14] A. Broo and S. Larsson, Chem. Phys. **161**, 363 (1992).

[15] A. Ferretti, R. Improta, A. Lami, and G. Villani, J. Phys. Chem. A **104**, 9591 (2000).

[16] G. C. Allen and N. S. Hush, Prog. Inorg. Chem. **8**, 357 (1967).

[17] N. S. Hush, Prog. Inorg. Chem. **8**, 391 (1967).

[18] J. T. Hupp, Y. Dong, R. L. Blackbourn, and H. Lu, J. Phys. Chem. **97**, 3278 (1993).

[19] K. W. Lau, A. M.-H. Hu, M. H.-J. Yen, E. Y. Fung, S. Grzybicki, R. Matamoros, and J. C. Curtis, Inorg. Chem. Acta **226**, 137 (1994).

[20] I. Cacelli, A. Ferretti, and A. Toniolo, J. Phys. Chem. A **105**, 4480 (2001).

[21] A. Farazdel, M. Dupuis, E. Clementi, and A. Aviram, J. Am. Chem. Soc. **112**, 4206 (1990).

[22] a) To obtain the localized UHF wavefunctions, we adopted the following strategy. At the biggining, we prepared the localized wavefunction for the geometry where one of the metal-bridge distances was taken to be very long but the other was normal. The obtained wavefunction is well-localized. After check of S^2 and spin density, we calculated the wavefunction of the ions whose metal-bridge distance was taken to be moderately shorter, where we employed the above-calculated well-localized UHF wavefunction as initial guess. Until the metal-bridge distance became the same as the real one, we continued the above procedure. This technics presented the well localized broken-symmetry UHF wavefunction; for instance, the spin densities on Ru centers of **1** are 1.09 and -0.01, respectively. The bond order analysis showed that the free valence electrons on one metal and the other at Δr=0.00 are 0.95 and 0.00 for **1**, 0.92 and 0.02 for **2**, 0.93 and 0.00 for **3**, and 0.91 and 0.00 for **4**. These results indicate that the wavefunction is well localized. b) The S^2 values for localized UHF wavefunctions at Δr=0.00 are 0.78 for **1**, 0.80 for **2**, 1.03 for **3**, and 1.04 for **4** (Supporting Information Table S4), indicating that the spin-contamination little occurs in **1** and **2** but somewhat in **3** and **4**. To investigate the nature of spin-contamination of bipyridine systems, we calculated the quartet spin state of the ions because this is considered to contribute considerably to the spin contamination. The unpaired electrons are localized on two metal centers and bridging ligand in the quartet

state. Thus, it is likely that the spin-contamination of the quartet spin state would increase the spin density on the bridging ligand and makes the S^2 values somewhat large.

[23] In the present study, UHF wave functions have been used just as basis functions to construct the total wave function, and it might be unsuitable to call it ‘diabatic basis’ in a precise sense. However, we use the word of ‘diabatic’ or ‘adiabatic’ throughout the manuscript to describe the transformation and mixing nature of the wave functions just for the convenience.

[24] M. W. Wong, M. J. Frisch, and K. B. Wiberg, J. Am. Chem. Soc. **113**, 4776 (1991).

[25] H. F. King, R. E. Stanton, H. Kim, R. E. Wyatt, and R. G. Parr, J. Chem. Phys. **47**, 1936 (1967).

[26] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).

[27] P. J. Hay and W. R. Wadt, J. Chem. Phys. **82**, 270 (1985).

[28] T. Koga, H. Tatewaki, H. Matsuyama, and Y. Satoh, Theor. Chem. Acc. **102**, 105 (1999).

[29] A. W. Ehlers, M. Böhme, S. Dapprich, A. Gobbi, A. Höllwarth, V. Jonas, K. F. Köhler, R. Stegmann, A. Veldkamp, and G. Frenking, Chem. Phys. Letters. **208**, 111 (1993).

[30] Electron-transfer matrix elements calculated with smaller basis set and larger one are 0.276 eV and 0.266 eV, respectively.

[31] A. D. Becke, J. Chem. Phys **98**, 5648 (1993).

[32] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, Montgomery, J. A., Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A.

Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, N. Rega, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, *Gaussian 98,* Revision A9 (Gaussian, Inc., Pittsburgh, PA, 1998).

[33] To calculate potential energy surface (PES), we fixed metal-NH_3 distance and changed only Δr. We examined how much the relaxation of the metal-NH_3 distances influences the PES, as follows: We optimized $Ru-NH_3$ distances with the DFT(B3LYP) method at $\Delta r = 0.00$ and $\Delta r = 0.09$ Å. The maximum difference of equatorial $Ru-NH_3$ distance between these two geometries is negligibly small (0.005 Å), as reported. [11] On the other hand, one of the axial Ru-NH_3 distances at Δr=0.09 Å becomes longer by 0.0157 Å and another one becomes shorter by 0.0249 Å than those at Δr=0.00 Å. To evaluate how much the relaxation of axial $Ru-NH_3$ distances influences PES, we changed the axial $Ru-NH_3$ distances at $\Delta r = 0.09$Å, considering the DFT-optimized $Ru-NH_3$ bond distances. The adiabatic energy, E-, of this geometry is 0.185 eV above the energy at $\Delta r = 0.00$. When $Ru-NH_3$ distances are fixed, its energy is 0.175eV above. Because these two geometries present almost the same energy, it is likely that the relaxation of the $Ru-NH_3$ distance little influences the PES. For **4**, we also examined the influence of the relaxation of the metal-NH_3 distances for PES.

[34] N. Koga, K. Sameshima, and K. Morokuma, J. Phys. Chem. **97**, 13117 (1993).

[35] S. Fraga, K. Saxena, and J. Karwowski, *Handbook of atomic data* , edited by B. Pullman (Elsevier, Amsterdam ; New York, 1976).

[36] To compare the orbital energy of Os with that of Ru, we calculated $[M(NH_3)_5]^{2+}$ (M=Ru or Os) with the HF method. The orbital energies of d_π orbital are -0.6908 eV for Ru and -0.6231 eV for Os, respectively.

[37] This decrease arises from the decrease in S_{kk}^{b-b}, as shown in Fig. 6.6.

Chapter 7

Electronic structure and solvation structure of $[\mathrm{Ru(CN)_6}]^{4-/3-}$ in aqueous solution: A RISM-SCF study

7.1 Introduction

The study of metal complexes in solution phase captures the researcher's interests and their structural properties have been extensively reporeted including experimental and theoretical approaches. In particular, much attention has been payed to the bimetallic cyanide-bridged complexes from the view point of electron transfer chemistry. They are widely used as an intramolecular (inner-sphere) electron transfer system. For instance, the transfer rates have been experimentally determined on $[(NH_3)_5RuNCRu(CN)_5]^-$ complex from femtosecond-spectroscopic investigations, and details of the process as well as the solvation dynamics have been revealed [1]. Numerous structural studies including X-ray diffraction have been also reported so far.

On the other hand, studies on the element of these complexes, i.e. fundamental six-coordinated complexes are rather limited, especially for ruthenium complex [2–4], It is also surprising that theoretical studies are further limited [5–7]. Since it is very perceivable that the electronic structure of the molecule is significantly changed in solution phase, main interests of the current issues are the electronic structure of the complex and solvation structure around it. It should be noted, however, the electronic and solvation structures are strongly coupled in general. A simultaneous approach both from solvation chemistry and quantum chemistry is required to study the present system theoretically.

RISM-SCF theory, that combines the statistical-mechanics of molecular liquids [8, 9] with *ab intio* molecular orbital theory [10–12], is a promising tool to tackle the present subject. In the theory, electronic structure of a solute molecule in solution and solvent distribution around the solute molecule are solved in a self-consistent manner. The methods have been widely applied to a variety of solute-solvent system including transition metal complex [13]. In the present article, we report a theoretical study for the electronic and solvation structure of $[Ru(CN)_6]^{4-}$ and $[Ru(CN)_6]^{3-}$ in aqueous solution based on the RISM-SCF theory. A serious problem in dealing with these systems is the assignment of effective charges that are necessary to carry out RISM calculation. Because of the specially high symmetry of the system, the conventional RISM-SCF procedure can not attain the simultaneous solution of the equation. By implementing an efficient technique, we obtain the electronic structure and solvation structure in a self-consistent manner for the first time.

The organization of this article is as follows. After describing the computational detailes, electronic and solvation structures of $[Ru(CN)_6]^{4-}$ and $[Ru(CN)_6]^{3-}$ are discussed.

7.2 Computational Details

RISM-SCF theory combines two major theoretical elements, the *ab initio* molecular orbital (MO) theory and the RISM integral equation method. In the theory, the solvation effect on the electronic structure of a solute is taken into account in a self-consistent manner, and simultaneous equations for the solute electronic structure and solute-solvent correlation functions are solved by use of the variational principle for the solvation free energy of the system. We recommend referring to the reviews and previous studies [14, 15].

In the RISM-SCF method, the electrostatic potential (ESP) charges, which are determined so as to reproduce the electrostatic potential around a solute molecule, is usually adopted since this set of charges is considered to be suitable to describe the electrostatic interaction between solute and solvent molecules. However, it is widely known that the assignment of ESP charge becomes often difficult because of the ill-posed nature of ESP fitting procedure. This difficulty gets more strained when buried atom exists in the system. In the present case not only the

Table 7.1: Lennard-Jones parameters

atom	σ/Å	ε / kcal mol^{-1}	$q/\|e\|$
Ru	4.68	0.036	[a] —
C	3.65	0.150	[a] —
N	3.20	0.170	[a] —
	water		
O	3.17	0.155	-0.82
H	1.00	0.056	0.41

[a] determined by RISM-SCF method. See Table 7.2.

ruthenium but also carbon atoms are buried and the charge assignment of these atoms are not easy. To overcome this problem, we incorporate the modified charge assignment procedure proposed by Morita et al. [16] into the RISM-SCF computation. The procedure offers a robust definition of the ESP charges.

The MO calculations were performed at the spin-restricted Hartree-Fock (HF) level of theory by using the Dunning-Huzinaga double-zeta basis sets [17] with d-polarization function on carbon (α=0.75) and nitrogen (α=0.80). The standard effective core potential and basis set parameters suggested by Stevens et al., in which 28 inner-shell electrons are replaced with the core potentials, were used for Ru [18]. The density functional theory (DFT) computations were also carried out with the hybridized HF/ Becke/ LYP using VWN formula 5 (B3LYP) for $[Ru(CN)_6]^{3-}$, in which we found very similar tendency to the HF-level computations. The point group of the complex was taken as the octahedral symmetry (O_h), and the geometry was fixed through the study at the X-ray structure [2–4]; R(Ru-C)=2.023Å and R(C-N)=1.157Å. The electronic structure was solved under the assumption of same symmetry for $[Ru(CN)_6]^{4-}$, while D_{4h} was employed for $[Ru(CN)_6]^{3-}$ to deal with the Jahn-Tellar effect.

The grid points to evaluate ESP were distributed around the centers of all the composing atoms based on 194 Lebedev polyhedrons [19] with six equally spaced layers from 10 to 30 Bohr for each directions. Total numbers of grid points were 1836. ϵ [16] was 1.0×10^{-1} and 1.0×10^{-4} for Ru(III) and Ru(II) complexes, respectively. The constraint for the total charge was employed to define the ESP charges.

Table 7.2: Effective atomic charges and contributions to the solvation free energy

	atom	effective charges[a] /$\|e\|$				$\Delta\mu$ /kcal mol^{-1}
		Mulliken		ESP		
$[Ru(CN)_6]^{3-}$	Ru	-0.0738	(-0.0738)	0.3353	(0.3353)	35.46
	C(ax)	-0.0440	(-0.0443)	0.1628	(0.1628)	26.41
	C(eq)	-0.0794	(-0.0794)	0.1638	(0.1638)	26.54
	N(ax)	-0.4353	(-0.4347)	-0.7211	(-0.7211)	-93.76
	N(eq)	-0.4125	(-0.4126)	-0.7185	(-0.7185)	-93.26
					total free energy	-366.13
$[Ru(CN)_6]^{4-}$	Ru	-0.4356	(-0.5147)	0.5677	(0.3188)	80.56
	C(ax)	0.0626	(-0.0409)	0.2998	(0.1269)	58.61
	C(eq)	0.0626	(-0.0409)	0.2998	(0.1269)	58.61
	N(ax)	-0.6567	(-0.5400)	-1.0611	(-0.8467)	-188.50
	N(eq)	-0.6567	(-0.5400)	-1.0611	(-0.8467)	-188.50
					total free energy	-698.74

[a] Values in parenthesis are in gas phase.

RISM equation were solved with the hyper-netted-chain (HNC) approximation. SPC-like water model [20] was employed to describe solvent water. The Lennard-Jones parameter of carbon and nitrogen were the OPLS parameter set [21], and that of ruthenium was the same as our previous study [13]. These are summirised in Table 7.1. The density of water (ρ) was assumed to be 1.0g/cm^3 at a temperature (T) of 298.15K.

All the computations were carried out with our modified version of GAMESS program packages.

7.3 Results and Discussion

7.3.1 The Electronic Structure

The effective charges computed by Mulliken population analysis and by the ESP procedure are shown in Table 7.2 together with the solvation free energy ($\Delta\mu$).

In all the cases, the charges of nitrogen atoms are negative and greater in absolute values than those of the buried atoms. This trend is similar in the two population analyses, although the effective charges derived from them are slightly different. The difference comes from the fact that each method reflects different characteristics of the electronic structure. The

Mulliken analysis represents the distribution of electron density in the solute, while the ESP is determined so as to reproduce the electrostatic potential around the solute. In other words, it is related to how the electronic distribution is seen from solvent molecules. For the results of ESP charges, it can be said the electronic character of the two complexes are similar in the gas phase: the effective charges on the ruthenium and carbon atoms are positive (ca. 0.3 and 0.1, respectively), whereas the charge on nitrogen atoms is notably negative (-0.7 to -0.8). The change of charges on the central ruthenium upon the reduction is not so large, and excess electron is de-localized over the whole complex.

The situation is changed in the solution phase. The solvation considerably affects the electronic structure of $[Ru(CN)_6]^{4-}$ and the effective charges are significantly altered from the gas phase ones. On the contrary, the electronic structure of $[Ru(CN)_6]^{3-}$ is undisturbed and the charges are virtually the same. The quadrupole moments (Q_{xx}, Q_{yy}, Q_{zz}) of this complex are $(2.065, 2.065, -4.131)$ and $(2.092, 2.092, -4.184)$ in the gas and aqueous solution phase, respectively. This electronic structure change is closely related to the hydration structure of these complexes as we will discussed later. In $[Ru(CN)_6]^{3-}$, the axial and equatorial nitrogen and carbon atoms are considered to be different each other in nature because of the Jahn-Tellar effect. But the effective charges of them are almost identical and any further difference is not seen.

How does the solvation affect the orbital energies of these complexes? Figure 7.1 illustrates the change of orbital energies near the frontier orbitals by the solvation. The horizontal axis of the graph is the orbital energy in gas phase, while the vertical axis represents that in aqueous solution. The mark denoted by the symmetry group t_{2g} is degenerated three orbitals in $[Ru(CN)_6]^{4-}$ complex corresponding to d_{xy}, d_{yz} and d_{xz} of the ruthenium $4d$ orbitals. The electronic structure of $[Ru(CN)_6]^{3-}$ is computed with D_{4h} symmetry so as to allow the splitting of these orbitals into two degenerated (e_g) and one singly occupied (b_{2g}) orbitals. The orbital energies in both phases (including some virtual orbitals) show very good correlation in each complex. This means that the orbital energies becomes negatively greater in aqueous solution but the solvation just shifts all the energies equivalently. Presumably, this happens because of

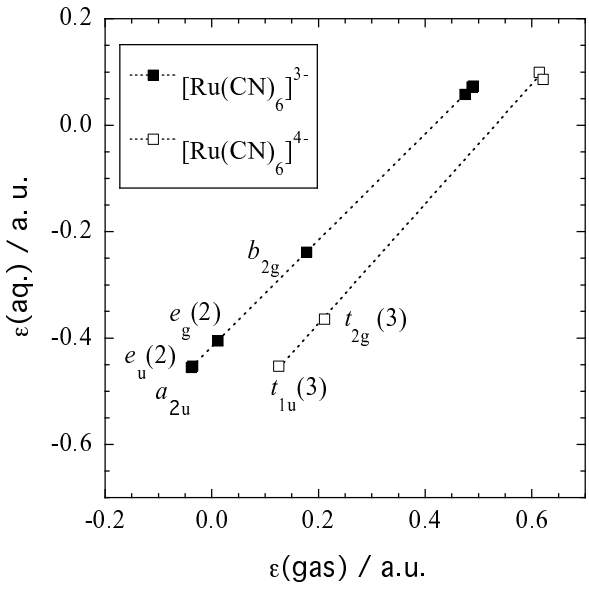


Figure 7.1: Selected orbital energies in the gas phase and in aqueous solution. The occupied orbitals are labeled with their degeneracy (the number in parenthesis).

the specially high symmetry of the present system. It is noted that the total energy represented by the total hamiltonian is not invariant upon transferring from the gas to aqueous solution since the system we are considering is dissipative one. The deepening of orbital energies does not seem to matter much.

$\Delta\mu$ can be decomposed and assigned to contribution from each atom (α) composing the solute molecule, because $\Delta\mu$ is formally expressed as a sum of the site-site contributions.

$$\Delta\mu = \sum_{\alpha} \Delta\mu_{\alpha}, \tag{7.1}$$

and

$$\Delta\mu_{\alpha} = -\rho k_{\mathrm{B}} T \sum_{s} \int d\mathbf{r} \left(c_{\alpha s}(r) - \frac{1}{2} h_{\alpha s}^2(r) + \frac{1}{2} h_{\alpha s}(r) c_{\alpha s}(r) \right), \tag{7.2}$$

where s indicates atoms in solvent molecules, and the functions $h_{\alpha s}$ and $c_{\alpha s}$ are total and direct correlation functions, respectively. $h_{\alpha s}$ is essentially equivalent to the pair correlation function (PCF). Note that $\Delta\mu_{\alpha}$ is not the same as the solvation free energy of an isolated atom α in the solvent. The correlation function used to evaluate $\Delta\mu_{\alpha}$ depends on all other atoms in the solute. As listed in the table, the greatest contribution to the solvation free energy

comes from nitrogens. This is quite a contrast to the hexaammine complex [13], in which the contribution from the embedded ruthenium is dominant. Presumably, the excess electron weakens the effect from the ruthenium and direct electrostatic interaction between nitrogen and solvent plays central role in the solvation process.

7.3.2 The Solvation Structure

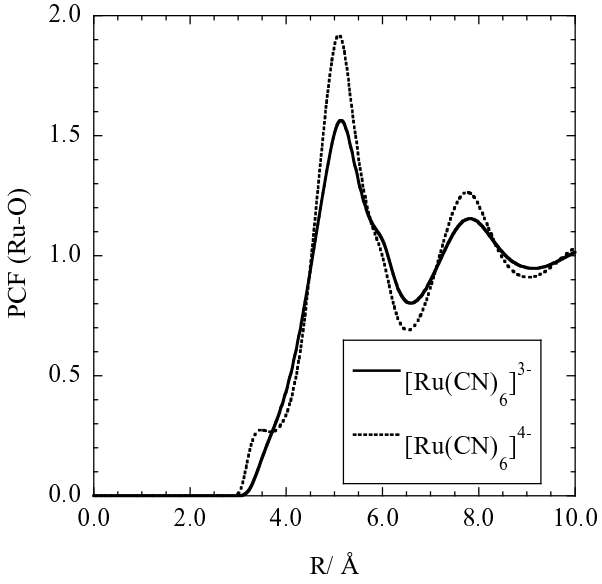


Figure 7.2: PCF between the ruthenium and oxygen atom of solvent water.

A big advantages of the present method is to provide the information of solvation structure in atomic level. Figure 7.2 is PCF between the central ruthenium and oxygen atom of solvent water. In the present model, direct contact of ruthenium and oxygen atoms should appear around $(\sigma_{\mathrm{Ru}} + \sigma_{\mathrm{O}})/2 = 3.9\text{Å}$. A small shoulder found around 3.5Å in [Ru(CN)$_6$]$^{4-}$ corresponds to the sovlation structure of this contact. The conspicuous peaks around $R = 5\text{Å}$ in the both complexes are attributed to the oxygen atom circling around the solute molecule. If the oxygen approaches to the complex along C–N axis, the peak position must be longer. For this results, the oxygen is considered to approach to nitrogen atoms perpendicular to the C–N axis. Since the ligand cyano group in these complexes is linear, there is a space for solvent molecule to enter the area between the ligands. Consequently, it is considered that attractive

interaction in $[Ru(CN)_6]^{4-}$ is strong enough to make a direct contact between the metal and solvent. However, it must be reminded that counter cations are not included in the present computations. In reality, the central metal may attract the cations and solvent molecules tend to be excluded from the nearest neighbour of the metal.

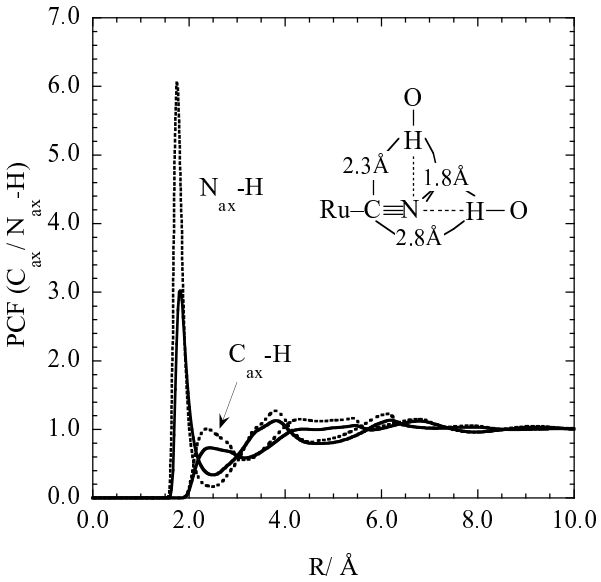


Figure 7.3: PCF between cyano group and hydrogen atom of solvent water. Solid lines are $[Ru(CN)_6]^{3-}$ and dashed lines represent $[Ru(CN)_6]^{4-}$.

Figures 7.3 and 7.4 are PCF around the cyano group. First thing one notices in these figures is the peak around 2.0Å in the N–H PCF. This can be obviously assigned to the hydrogen bonding of solvated water molecule. The distinct peak around 3.0Å in the N–O PCF is another evidence of this hydrogen bonding. As mentioned above, the bonding is not co-linear along the C–N axis if we consider the peak positions. In the C–O PCF, the position of the first peak is found around 3.5Å. It is geometrically impossible to arrange C, N and O atoms in a straight line, since the peak positions of C–O and N–O are too close. A possible solvent location that is consistent with these peak positions is shown in the figure. It is noted, however, that the peak in C–O PCF is rather broad and a small shoulder is seen around 4.0Å. This length is sufficient to align three atoms in a straight line. In summary, there is a strong hydrogen bonding between the nitrogen and hydrogen atoms, but solvent molecules can be placed with wide range of angles in a continuous fashion. Two O–H moieties depicted in the figure correspond to the limits of

these distribution. Similar to the discussion in the effective charges, axial and equatorial atoms are virtually the same and their PCFs are indistinguishable.

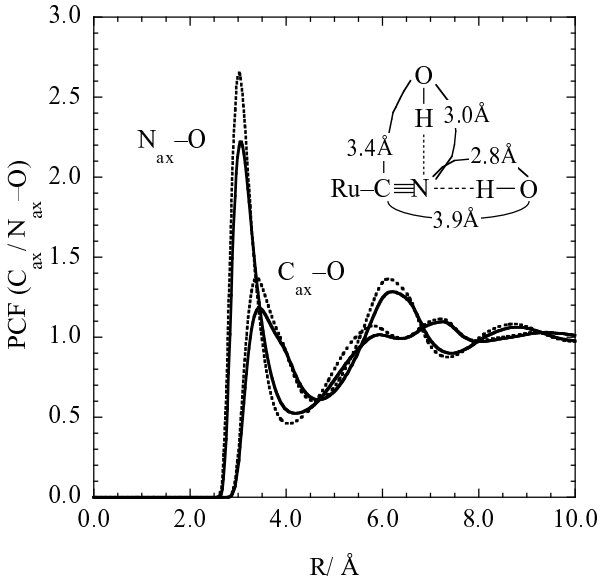


Figure 7.4: PCF between cyano group and oxygen atom of solvent water. See the caption in Fig 7.3.

It is noted that all the peak positions in PCF of $[Ru(CN)_6]^{3-}$ and those in $[Ru(CN)_6]^{4-}$ is very close to each other and only the peak height is slightly greater in $[Ru(CN)_6]^{4-}$ than that in $[Ru(CN)_6]^{3-}$. This implies that the solvation structure around these complexes are similar.

7.4 Conclusions

Solvation and electronic structure for the two metal complexes, $[Ru(CN)_6]^{3-}$ and $[Ru(CN)_6]^{4-}$ in aqueous solution, are studied by means of the ab initio RISM-SCF method.

The electronic structure of these complexes in gas phase are similar from the view point of ESP fitting charge. The electrons tend to show partiality to nitrogen atoms and the buried atoms (ruthenium and carbon) are positively charged. The solvation effect differentiates the electronic structure of them. $[Ru(CN)_6]^{3-}$ is insensitive to the effect and the partial charges assigned on each atom are not changed, whereas the electronic structure of $[Ru(CN)_6]^{4-}$ is significantly polarized by the solvation.

For the results of peak positions of PCF around the two complexes, we can conclude that the

solvation structure is not greatly changed from the oxidation states of the complex although the peak heights, i.e. solvation strength are slightly different each other.

Bibliography

[1] K. Tominaga, D. Kliner, A. Johnson, N. Levinger, and P. Barbara, J. Chem. Phys. **98**, 1228 (1993).

[2] L. Gentil, A. Navaza, J. Olabe, and G. Rigotti, Inorganica Chim. Acta **179**, 89 (1991).

[3] J. Olabe, L. Gentil, G. Rigotti, and A. Navaza, Inorg. Chem. **23**, 4297 (1984).

[4] J. Kautz and R. C. Symes, Acta Cryst. **C59**, i89 (2003).

[5] K. Pierloot, E. V. Praet, L. Vanquickenborne, and B. Roos, J. Phys. Chem. **97**, 12220 (1993).

[6] J. Zeng, N. Hush, and J. Reimers, J. Phys. Chem. **99**, 10459 (1995).

[7] J. Zeng, N. Hush, and J. Reimers, J. Am. Chem. Soc. **118**, 2059 (1996).

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

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

[10] S. Ten-no, F. Hirata, and S. Kato, Chem. Phys. Lett. **214**, 391 (1993).

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

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

[13] H. Sato and F. Hirata, J. Phys. Chem. A **106**, 2300 (2002).

[14] F. Hirata, H. Sato, S. Ten-no, and S. Kato, in *Combined quantum mechanical and molecular mechanical methods* , edited by J. Gao and M. A. Thompson (American Chemical Society, Washington DC, 1998).

[15] F. Hirata, H. Sato, S. Ten-no, and S. Kato, *The RISM-SCF/MCSCF Approach for the Chemical Processes in Solutions* in *"Computational Biochemistry and Biophysics"* , edited by O. M. Becker, A. D. MacKerell Jr, B. Roux, and M. Watanabe (Marcel Dekker, New York, 2001).

[16] A. Morita and S. Kato, J. Phys. Chem. A **106**, 3909 (2002).

[17] T. H. Dunning Jr. and P. J. Hay, Chapter 1 in *"Methods of Electronic Structure Theory"* , edited by H. F. Schaefer III (Plenum, New York, 1977).

[18] W. Stevens, H. Basch, M. Krauss, and P. Jasien, Can. J. Chem. **70**, 612 (1992).

[19] V. I. Lebedev and D. Laikov, Dokl. Math. **59**, 477 (1999).

[20] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces* , edited by B. Pullman (Reidel, Dordrecht, 1981).

[21] W. L. Jorgensen and C. J. Swenson, J. Am. Chem. Soc. **107**, 569 (1985).

Chapter 8

Alternative couplings of solute–solvent interaction in RISM-SCF method

8.1 Introduction

Electronic structure is considerably affected by solute–solvent interaction. In many cases, the electrostatic interaction is the dominative one that governs the electronic structure of the molecules as well as the solvation structure. Numerous representations have been proposed to describe this Coulombic interaction. For example it is replaced with the interaction between the electronic density of the solute and the surrounding media within the framework of dielectric continuum models such as polarisable continuum model (PCM) [1, 2]. In the reference interaction site model-self-consistent field (RISM-SCF) theory, the electrostatic potential of the reaction field at solute atom α produced by the surrounding molecules is expressed as the consequence of the statistically averaged charge distribution.

$$V_\alpha = (\mathbf{V})_\alpha = \rho \sum_s \int \frac{q_s}{r} g_{\alpha s}(r) d\mathbf{r}, \tag{8.1}$$

where q_s is the partial charge on the site s in solvent, ρ is the bulk density of solvent, and $g_{\alpha s}(r)$ is the pair correlation function (PCF) between α and s. The interaction energy (E_{int}) is described as the product of the partial charge assigned on the site in solute ($q_\alpha^{(\mathrm{e})}$) and $(\mathbf{V})$.

$$E_{\mathrm{int}} = \mathbf{V}^t \cdot \mathbf{q}^{(\mathrm{e})} = \sum_\alpha V_\alpha q_\alpha^{(\mathrm{e})}. \tag{8.2}$$

In the original version of RISM-SCF, a set of partial charges is determined so as to reproduce the electrostatic potential (ESP) around the solute with the least square fitting procedure, which

is used for the charge ($\mathbf{q}^{(\mathrm{e})}$) in Eq.(8.2). This is presumably the most reliable choice of the partial charge in the present purpose and RISM-SCF have been successfully applied to broad range of chemical phenomena [3–7]. However, it is known that ESP charge set sometimes shows ill behaviours such as multiple-valued nature.

In the present work, other types of the interaction are examined as an alternative to the description of the solute–solvent interaction.

8.2 Method

In the conventional RISM-SCF procedure, ESP charge ($\mathbf{q}^{(\mathrm{e})}_{\mathrm{ESP}}$) set is used to represent the solute electronic structure. Since the least square fitting is analytically employed the charge set attributed to the electron is described as the following closed-form equation,

$$\mathbf{q}^{(e)}_{\mathrm{ESP}} = -tr\left(\mathbf{a}^{-1}\mathbf{PB}\right) - \frac{-\mathbf{1}^t tr\left(\mathbf{a}^{-1}\mathbf{PB}\right) - N_e}{\mathbf{1}^t\mathbf{a}^{-1}\mathbf{1}}\mathbf{a}^{-1}\mathbf{1}, \tag{8.3}$$

where $\mathbf{P}$ is 'density matrix', $\mathbf{B}$ and $\mathbf{a}$ are defined by

$$(\mathbf{B})_{\mu,\nu,i} = \sum_{k=1}^{\text{grid points}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_k|}\int \frac{\chi^*_\mu(\mathbf{r}')\chi_\nu(\mathbf{r}')}{|\mathbf{r}_k - \mathbf{r}'|}d\mathbf{r}', \tag{8.4}$$

$$(\mathbf{a})_{ij} = \sum_{k=1}^{\text{grid points}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_k||\mathbf{r}_j - \mathbf{r}_k|}. \tag{8.5}$$

N_e is the total numbers of electrons. For the meanings of other notations, we recommend referring to the original paper [5, 6]. The solvated Fock matrix element of the conventional RISM-SCF method

$$\mathbf{F}^{\mathrm{solv}} = \mathbf{F}^{\mathrm{gas}} - \mathbf{V}^t \cdot \left[\mathbf{a}^{-1}\mathbf{B} - \frac{\mathbf{a}^{-1}\mathbf{1}}{\mathbf{1}^t\mathbf{a}^{-1}\mathbf{1}}\left(\mathbf{1}^t\mathbf{a}^{-1}\mathbf{B} - \mathbf{S}\right)\right], \tag{8.6}$$

is derived using this definition of charge set (Method A).

The simplest procedure to derive the charge set is undoubtedly Mulliken population analysis (MPA), in which the charge on atom α is given by,

$$\left(\mathbf{q}^{(e)}_{\mathrm{MPA}}\right)_\alpha = -\sum_{\mu\in\alpha}(\mathbf{PS})_{\mu\mu}\,, \tag{8.7}$$

where $\mathbf{S}$ is overlap matrix. The relationship between these two charge sets is discussed in the Appendix. In the spirit of the conventional RISM-SCF procedure, the solvated Fock operator ($F_{\mu\nu}^{\mathrm{solv}}$) can be defined as follows (Method B).

$$\mathbf{F}^{\mathrm{solv}} = \mathbf{F}^{\mathrm{gas}} - \mathbf{X}, \tag{8.8}$$

where the matrix element related to atoms α and β is given by,

$$X_{\mu\nu} = \frac{1}{2}(V_\alpha + V_\beta)S_{\mu\nu}, \;\; \mu \in \alpha \text{ and } \nu \in \beta. \tag{8.9}$$

As stated above, two representations of the wave function are used in the RISM-SCF theory, $\mathbf{q}_{\mathrm{ESP}}^{(e)}$ and $\mathbf{q}_{\mathrm{MPA}}^{(e)}$. They are obtained from the electronic structure calculation and used for solving RISM equation. Eqs. 8.6 and 8.8 are describing the interaction part of the electronic hamiltonian and related to these charges, respectively. Now, 'dual' type procedure may be possible to be considered (Method C). The Fock operator defined in Eq. (8.6) is used for the electronic structure computations, whilst the Mulliken charge set derived from this Fock operator is used for the RISM computations. The resultant PCF is used for the construction of Eq. 8.6. It is noted that the quantity in the square brackets of Eq. 8.6 can be computed only with the information of the basis functions and the grid points $\{\mathbf{r}_k\}$, which expresses the spatial property around the solute molecule.

The reason to choose this combination is as follows; The electrostatic interaction computed with Eq. 8.8 is generally stronger than the ESP case, Eq. 8.6. This is because $\mathbf{B}$ virtually reflects the extension of electronic clouds of the solute molecule through the grid points, thus the solute–solvent interaction can be treated in a reliable manner. Meanwhile, the absolute values of $\mathbf{q}_{\mathrm{ESP}}^{(e)}$ is slightly greater than $\mathbf{q}_{\mathrm{MPA}}^{(e)}$ in many molecular systems, meaning the polarization of a molecule is somewhat emphasized when using $\mathbf{q}_{\mathrm{ESP}}^{(e)}$. Since the charge set directly governs RISM computations, solvation structure depends on the choice of the charge set. The present choice of combination is very pragmatic but effective way to compute the solvation effect. Such discordance in the description of the interaction is often seen in the framework of QM/MM and causes no problems in our experiences.

Table 8.1: Several properties of water molecule in aqueous solution

Method		gas	A	B	C
Mulliken charge (q_{O}) / $\|e\|$		-0.68	-0.82	-1.20	-0.78
dipole moment / Debye		2.24	2.84	2.71	2.67
$\Delta\mu$ / kcal mol^{-1}		—	-8.66	-17.05	-5.07
E_{reorg} / kcal mol^{-1}		0.00[a]	3.36	10.42	1.74
E_{int} / au		—	-0.42943	-0.52387	-0.33787
orbital energy	(a_1) ϵ_1/au	-20.5565	-20.5861	-20.5784	-20.5803
	(a_1) ϵ_2/au	-1.3463	-1.3746	-1.3857	-1.3694
	(b_1) ϵ_3/au	-0.7109	-0.7403	-0.7455	-0.7348
	(a_1) ϵ_4/au	-0.5758	-0.6185	-0.6128	-0.6093
	(b_2) ϵ_5/au	-0.5029	-0.5371	-0.5412	-0.5302

[a] Total energy is -76.04635 au in gas phase.

In the following section, these three procedures are compared from the view point of physical properties and the convergency.

8.3 Results and Discussion

8.3.1 Benchmark computations on water in aqueous solution

We have applied three types of RISM-SCF method to water molecule in aqueous solution. Hartree-Fock method with DZP basis sets [8] was employed for the electronic structure computations and SPC-like parameters [9] were adopted in the the RISM part. The experimental geometry (R_{OH}=0.9575Å and $\angle$HOH=104.51°) is used. All calculations were carried out at the temperature 298 K and the solvent density 0.03334 molecule/Å^3.

The representative physical properties are summarised in Table 8.1. Mulliken charge assigned on the oxygen and the dipole moment show that the electronic structure of the water molecule in aqueous solution is polarised in all the cases compared with the molecule in the gas phase. The interaction between the solute and solvent in Method B is the strongest, while that in C is the weakest. The conventional procedure (Method A) is intermediate between them. The excess chemical potential ($\Delta\mu$) becomes negatively greater while the reorganisation or polarisation energy (E_{reorg}) becomes positively greater, as the interaction is strengthened. The interaction regularly makes the orbital energies deeper except for ϵ_4 of method A. At this mo-

ment we do not have any clear explanation of the nonstandard depth of the orbital.

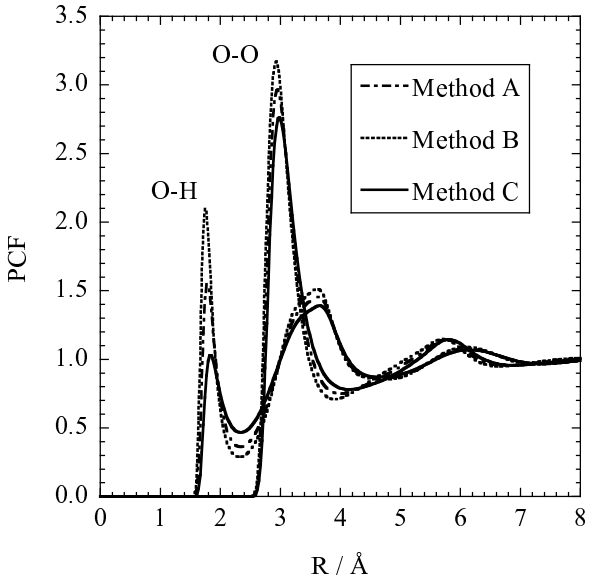


Figure 8.1: PCF around water molecule in aqueous solution computed by three different methods.

The pair correlation function (PCF) computed by the three methods are plotted in Fig. 8.1. The positions of the first peaks in O–H and O–O look very similar, but their heights are slightly different each other. The hydration number is one of the measures to judge the reliability of PCFs. By integrating the O–H PCF up to the first minimum, we obtained 1.90 (Method A), 1.86 (Method B) and 1.55 (Method C) respectively. All these values are less than two, which is good accordance with estimation in a molecular simulation studies (1.6–1.9) [10, 11]. The heights and hydration number show that the interaction computed by Method C is the smallest among the three, which is consists with the properties discussed above. Fig. 8.2 illustrates the convergence profile of the computations of the three methods. Exactly the same algorithm were used for computing the RISM-SCF. The convergence is judged by the root-mean-square-deviation of the electrostatic potentials in the successive iteration cycles. Because of its moderate interaction, computation by Method C quickly converges to the threshold. On the other hand, convergence rate of Method B is two time slower than Method C.

All these results clearly show that the interaction by Method C is weaker than the conventional method, A, while that by B is stronger than A. In other words, the electronic structure

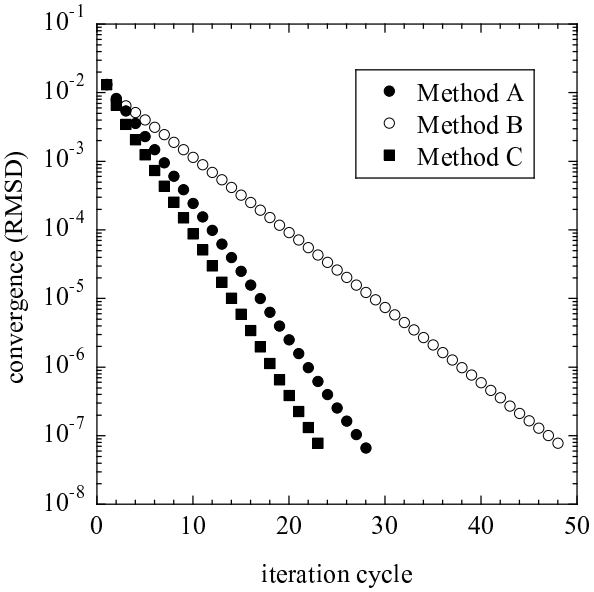


Figure 8.2: Convergence behaviour of the three methods.

by Method B is the most greatly distorted with respect to the gas phase one. Another point we would like to point out is that the solute-solvent interaction is well described by the linear response regime of solvation in many cases. The convergence rate must be faster as weaker the interaction.

8.3.2 A charge-transfer complex NH_3-BH_3

NH_3-BH_3 is known as a typical charge-transfer complex. The sum of Mulliken charges in NH_3 moiety in the gas phase is $+0.26|e|$ at the optimized geometry, exhibiting its charge-transfer character. According to our experience, the RISM-SCF computation diverges when the conventional Method A is applied to this system in aqueous solution. The reason of this behaviour originates from the effective charge assigned on the hydrogen attached to the boron atom (H_B). The value given by ESP procedure is largely negative (-0.2, for example) that attracts hydrogen of liquid water. The height of the first peak in the H_B–H PCF increases, then the polarisation of B–H_B bond is enhanced by the solvation, which again attracts water hydrogen atoms. Because of this endless cycle, the convergence is not usually attained. The situation is the same when using Method B due to its strong interaction.

The convergence can be obtained only when using Method C since its interaction is the most moderate. In the present work, Hartree-Fock method with DZP basis set is carried out. Lennard-Jones parameters are taken from literatures ($\sigma_{\mathrm{B}} = 3.71\text{Å}, \epsilon_{\mathrm{B}} = 0.136\ \mathrm{kcal\,mol^{-1}}$ [12], $\sigma_{\mathrm{N}} = 3.42\text{Å}, \epsilon_{\mathrm{N}} = 0.170\ \mathrm{kcal\,mol^{-1}}$ [13], $\sigma_{\mathrm{HB}} = 2.00\text{Å}, \epsilon_{\mathrm{HB}} = 0.070\ \mathrm{kcal\,mol^{-1}}$ [14]). The parameter for H_N is set to the same as SPC-like water's. Fig. 8.3 shows the PCFs around the solute molecule. All peaks indicate that the description of the hydration is reasonably obtained with this method. Because of the negative charge on the hydrogen atom attached to the boron, H_B–H PCF shows a well-marked peak around 1.5Å. At the same time, hydrogen atom in solvent water is also attracted to the boron atom.

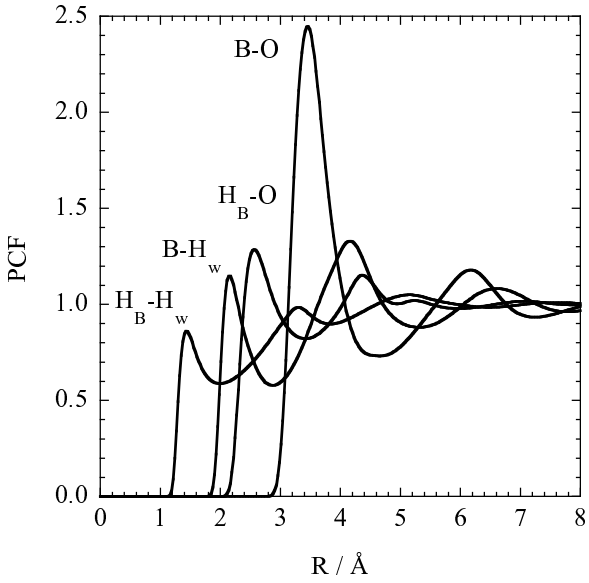


Figure 8.3: PCF around NH_3BH_3. H_W indicates the hydrogen of solvent water

The converged Mulliken charges on H_B, B, N and H_N are $-0.107|e|$, $-0.160|e|$, $-0.526|e|$ and $0.336|e|$, respectively. The corresponding charges in the gas phase are $-0.157|e|$, $0.208|e|$, $-0.715|e|$ and $0.326|e|$, respectively. In aqueous solution, the charge-transfer is strongly enhanced and the sum of Mulliken charges in NH_3 moiety is $+0.48|e|$, which is twice as large as in the gas phase. It is of great interests that even the sign of the effective charge of the boron atom inverts by flowing electrons from NH_3 to BH_3 moiety. The boron atom also attracts electron from the attached hydrogen atoms. These changes in the effective charges indicate that

the electronic structure around the boron is significantly affected by the solvation.

8.4 Conclusions

In the present work, we proposed other type of electrostatic couplings between solute and solvent molecules within the framework of RISM-SCF theory. It is found that the interaction by Mulliken-type method is stronger than the conventional method based on ESP, while that by dual-type method is much more moderate than the others. The dual-type method is very pragmatic but effective way to compute the solvation effect.

Appendix

We consider the relationship between the two charge, $\mathbf{q}^{(e)}_{\mathrm{ESP}}$ and $\mathbf{q}^{(e)}_{\mathrm{MPA}}$. Let us start with considering the first term in Eq. 8.3.

$$\begin{aligned} -tr\left(\mathbf{a}^{-1}\mathbf{PB}\right) &= -\sum_{\beta}\sum_{\mu\nu} a^{-1}_{\alpha\beta} B_{\mu\nu,\alpha} P_{\nu\mu} \\ &= -\sum_{\beta}\sum_{\mu\nu} a^{-1}_{\alpha\beta} \sum_{g} \frac{1}{|\mathbf{r}_\alpha - \mathbf{r}_g|} \langle\mu| \frac{1}{|\mathbf{r}_g - \mathbf{r}|} |\nu\rangle P_{\nu\mu}, \end{aligned} \tag{8.10}$$

where μ and ν are atomic orbitals (basis set) belonging to atoms ξ and η, respectively. Introducing Mulliken-type approximation on the matrix element,

$$\langle\mu| \frac{1}{|\mathbf{r}_g - \mathbf{r}|} |\nu\rangle \sim \frac{1}{2}\left(\frac{1}{|\mathbf{r}_\xi - \mathbf{r}_g|} + \frac{1}{|\mathbf{r}_\eta - \mathbf{r}_g|} \right) \langle\mu|\nu\rangle, \tag{8.11}$$

Eq. 8.10 is rewritten as follows,

$$\begin{aligned} & -\frac{1}{2}\sum_{\alpha}\sum_{\xi\eta}\sum_{\mu\in\xi}\sum_{\nu\in\eta} a^{-1}_{\alpha\beta}\left(a_{\alpha\xi} + a_{\alpha\eta}\right)\langle\mu|\nu\rangle P_{\nu\mu} \\ = & -\frac{1}{2}\sum_{\xi\eta}\sum_{\mu\in\xi}\sum_{\nu\in\eta}\left(\delta_{\beta\xi} + \delta_{\beta\eta}\right)\langle\mu|\nu\rangle P_{\nu\mu} = \sum_{\mu\in\beta}(\mathbf{PS})_{\mu\mu} \equiv \left(\mathbf{q}^{(e)}_{\mathrm{MPA}}\right)_{\beta}. \end{aligned} \tag{8.12}$$

Note that $\mathbf{a}$ is symmetric matrix. Thus the $\mathbf{q}^{(e)}_{\mathrm{ESP}}$ is reduced to $\mathbf{q}^{(e)}_{\mathrm{MPA}}$ by applying Eq. 8.11 to Eq. 8.10, since the second term in Eq. 8.3 disappears by using the same approximation.

Bibliography

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

[2] J. Tomasi, B. Mennucci, and R. Cammi, Chem. Rev. **105**, 2999 (2005).

[3] F. Hirata, H. Sato, S. Ten-no, and S. Kato, in *Combined quantum mechanical and molecular mechanical methods* , edited by J. Gao and M. A. Thompson (American Chemical Society, Washington DC, 1998).

[4] F. Hirata, H. Sato, S. Ten-no, and S. Kato, *The RISM-SCF/MCSCF Approach for the Chemical Processes in Solutions* in *"Computational Biochemistry and Biophysics"* , edited by O. M. Becker, A. D. MacKerell Jr, B. Roux, and M. Watanabe (Marcel Dekker, New York, 2001).

[5] S. Ten-no, F. Hirata, and S. Kato, Chem. Phys. Lett. **214**, 391 (1993).

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

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

[8] T. H. Dunning Jr. and P. J. Hay, Chapter 1 in *"Methods of Electronic Structure Theory"* , edited by H. F. Schaefer III (Plenum, New York, 1977).

[9] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces* , edited by B. Pullman (Reidel, Dordrecht, 1981).

[10] R. D. Mountain, J. Chem. Phys. **90**, 1866 (1989).

[11] R. D. Mountain, J. Chem. Phys. **103**, 3084 (1995).

[12] X. Chen, L. Bartolotti, K. Ishaq, and A. Tropsha, J. Comput. Chem. **15**, 333 (1994).

[13] R. C. Rizzo and W. L. Jorgensen, J. Am. Chem. Soc. **121**, 4827 (1999).

[14] J. Gao and X. Xia, J. Am. Chem. Soc. **115**, 9667 (1993).

Chapter 9

New-generation of the reference interaction site model self-consistent field method: Introduction of spatial electron density distribution to the solvation theory

9.1 Introduction

Quantum molecular orbital calculation (MO calculation) with solvation effect is a fundamental tool in the theoretical study of chemical physics in solution. Many solvation theories have been proposed for investigation of chemical process in solvation phase.

In dielectric continuum model, such as polarizable continuum model (PCM) [1], solvent molecules are replaced by macroscopic media with dielectric constant. The electronic structure is solved in vacuum cavity surrounded by the dielectric continuum. In quantum mechanics/molecular mechanics simulations (QM/MM), the neighboring solvent molecules around a solute molecule are treated explicitly. The electronic structure and solvation structure are calculated by averaging over various solvent configurations. Although these methods have been widely employed, the former oversimplifies microscopic characters of solvent and the latter requires large computational cost for the generation of the solvent configurations. Reference interaction site model self-consistent field (RISM-SCF) [2, 3] is another method, in which solvation structure is provided by an integral equation theories based on statistical mechanics of molecular liquids (RISM) [4, 5]. RISM-SCF offers not only various macroscopic thermodynamic quantities but also microscopic properties such as radial distribution functions (RDFs)

with reasonable computational cost. RISM-SCF has been successfully applied to understand the interplay between the electronic structure and solvation structure [6].

In the treatment of solvation effect, Coulomb interaction between solute and solvent molecules is primarily important factor in most cases. A common representation for the interaction is the sum of pairwise interactions between point charges assigned on each atom. The most popular method to set the charges is the least-square fitting (LSF) procedure, in which the effective charges are determined so that the electrostatic potential (ESP) derived from MO calculation can be reproduced at a set of grid points. Although the LSF procedure, which is employed for the original RISM-SCF [2], is very simple, several weak points have been pointed out so far. For example, the atomic charges depend on the choice of the set of grid points. When buried atoms exist in the molecule, the evaluation of the atomic charges are often ill-behaved. Besides, the representation of point charges neglects spread of electron distribution.

To obtain more realistic Coulomb interaction, another strategy has been used in quantum chemical study, especially in the field of density functional theory. In this strategy, the auxiliary basis sets (ABSs) on each atom are prepared to divide electron density into the components assigned on each atom. Gill, Johnson, Pople, and Taylor proposed a procedure to determine ABSs which reproduce the ESP provided by MO calculation (GJPT procedure) [7]. The great advantage of GJPT procedure is that it treats directly spatial electron density distribution (SEDD) and does not require the set of grid points; it is free from these artificial parameters. As described later, GJPT procedure is very stable to determine the charges even if a buried site is involved in the solute molecule.

In this paper, we propose the new-generation RISM-SCF, in which GJPT procedure is employed. The present method, RISM-SCF explicitly including SEDD (RISM-SCF-SEDD), is much more robust in the connection between RISM and MO calculation than the original version of RISM-SCF and significantly expands the versatility of the RISM-SCF family. In Sec. 9.2, the RISM-SCF-SEDD formalism and the relation between GJPT and LSF procedures are presented. In Sec. 9.3, the computational details of this work are described. The results of H_2O, C_2H_5OH, and HLi evaluated by RISM-SCF-SEDD are shown in Sec. 9.4.

9.2 Method

9.2.1 The formalism of RISM-SCF-SEDD

In GJPT procedure, model electron density $\tilde{\rho}$ is determined so that the ESP calculated by MO calculation can be reproduced, under the constrain of conservation of total number of electron. Gill *et al*. showed that $\tilde{\rho}$ can be obtained by minimizing the following quantity:

$$\Gamma = -2\pi \int\int \left(\rho(\mathbf{r}_1) - \tilde{\rho}(\mathbf{r}_1)\right) |\mathbf{r}_1 - \mathbf{r}_2| \left(\rho(\mathbf{r}_2) - \tilde{\rho}(\mathbf{r}_2)\right) \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 + 2\lambda \left[N_e - \int \tilde{\rho}(\mathbf{r}) \mathrm{d}\mathbf{r} \right], \quad (9.1)$$

where N_e is the number of electrons and ρ is the electron density derived from MO calculation. $\tilde{\rho}(\mathbf{r})$ is represented by the set of ABSs $f_i(\mathbf{r})$ centered on each solute site,

$$\tilde{\rho}(\mathbf{r}) = \sum_i^{N_{\mathrm{ABS}}} d_i f_i(\mathbf{r}), \quad (9.2)$$

where N_{ABS} is the number of ABSs [8]. The expansion coefficients $\mathbf{d}$ in eq. 9.2 can be determined by the following equations,

$$\mathbf{d} = \mathbf{X}^{-1}\mathrm{tr}(\mathbf{PY}) - \lambda \mathbf{X}^{-1}\mathbf{Z}, \quad (9.3)$$

$$\lambda = \frac{\mathbf{Z}^t\mathbf{X}^{-1}\mathrm{tr}(\mathbf{PY}) - N_e}{\mathbf{Z}^t\mathbf{X}^{-1}\mathbf{Z}}, \quad (9.4)$$

using the density matrix $\{P_{\mu\nu}\}$ ($= \sum_i n_i C_{\mu i} C^*_{\nu i}$) calculated from MO coefficients $\{C_{\mu i}\}$ and occupation number n_i. The components of the matrix $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ are defined, as follows:

$$X_{ij} = \int\int f_i(\mathbf{r}_1)|\mathbf{r}_1 - \mathbf{r}_2| f_j(\mathbf{r}_2)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2, \quad (9.5)$$

$$Y_{\mu\nu,i} = \int\int \phi_\mu(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1)|\mathbf{r}_1 - \mathbf{r}_2| f_i(\mathbf{r}_2)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2, \quad (9.6)$$

$$Z_i = \int f_i(\mathbf{r})\mathrm{d}\mathbf{r}, \quad (9.7)$$

where ϕ is the basis function employed in MO calculation.

The effective electrostatic interaction between f_i and solvent is then given by [2]

$$V_i = n^V \sum_\gamma q_\gamma \int\int \frac{f_i(\mathbf{r}' - \mathbf{r}_\alpha)}{|\mathbf{r} - \mathbf{r}'|} h_{\alpha\gamma}(|\mathbf{r} - \mathbf{r}_\alpha|)\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \quad (i \in \alpha), \quad (9.8)$$

where $h_{\alpha\gamma}$ is total correlation function between solute site α and solvent site γ. q_γ is partial charge of solvent site γ, n^V is the number density of solvent, and $\mathbf{r}_\alpha$ is the coordinate of solute

site α. By employing the standard procedure in RISM-SCF [2, 3], the solvated Fock matrix is given by,

$$\mathbf{H}^{\mathrm{solv}} = \mathbf{H}^{\mathrm{gas}} - \mathbf{V}\mathbf{X}^{-1}\mathbf{Y} + \frac{\mathbf{V}\mathbf{X}^{-1}\mathbf{Z}}{\mathbf{Z}^t\mathbf{X}^{-1}\mathbf{Z}}\left[\mathbf{Z}^t\mathbf{X}^{-1}\mathbf{Y} - \mathbf{S}\right], \tag{9.9}$$

where $\mathbf{H}^{\mathrm{gas}}$ is the Fock matrix in gas phase and $\mathbf{S}$ is overlap matrix.

9.2.2 The relationship between GJPT and LSF procedures

In this section we would like to make a brief comment on the relationship between GJPT and LSF procedures. In the standard LSF procedure, atomic population $\mathbf{q}$ is determined by the following equation [2],

$$\mathbf{q} = \mathbf{A}^{-1}\mathrm{tr}(\mathbf{PB}) - \lambda\mathbf{A}^{-1}\mathbf{1}, \tag{9.10}$$

$$\lambda = \frac{\mathbf{1}^t\mathbf{A}^{-1}\mathrm{tr}(\mathbf{PB}) - N_e}{\mathbf{1}^t\mathbf{A}^{-1}\mathbf{1}}. \tag{9.11}$$

The components of $\mathbf{A}$ and $\mathbf{B}$ are defined as follows:

$$A_{\alpha\beta} = \sum_{k=1}^{l} \frac{1}{|\mathbf{r}_k - \mathbf{r}_\alpha||\mathbf{r}_k - \mathbf{r}_\beta|} \tag{9.12}$$

$$B_{\mu\nu\alpha} = \sum_{k=1}^{l} \int \frac{\phi_\mu(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1)}{|\mathbf{r}_k - \mathbf{r}_\alpha||\mathbf{r}_k - \mathbf{r}_1|}\mathrm{d}\mathbf{r}_1, \tag{9.13}$$

where $\mathbf{r}_k$ is the coordinates of grid point and $\mathbf{r}_\alpha$ is those of solute site.

Comparing eqs. 9.3 and 9.10, the stability of the charge-determination depends on the character of $\mathbf{X}^{-1}$ and $\mathbf{A}^{-1}$. In the case of LSF procedure, $\mathbf{A}$ is calculated from the grid set around the solute molecule. Since grid point $\mathbf{r}_k$ is far from $\mathbf{r}_\alpha$ (or $\mathbf{r}_\beta$) in most cases ($|\mathbf{r}_k - \mathbf{r}_M| \gg |\mathbf{r}_{\alpha/\beta} - \mathbf{r}_M|$), eq. 9.12 is,

$$\begin{aligned} A_{\alpha\beta} &= \sum_{k}^{l} \frac{1}{|(\mathbf{r}_k - \mathbf{r}_M) - (\mathbf{r}_\alpha - \mathbf{r}_M)||(\mathbf{r}_k - \mathbf{r}_M) - (\mathbf{r}_\beta - \mathbf{r}_M)|} \\ &\sim \sum_{k} \frac{1}{|\mathbf{r}_k - \mathbf{r}_M|^2} = \mathrm{Const.}\,, \end{aligned} \tag{9.14}$$

where $\mathbf{r}_M$ is arbitrary point in the molecule (ex. the center of mass). Thus all the components of $\mathbf{A}$ tend to be very similar to each other and the behavior of inverse of such matrix sometimes

becomes unstable. [9] On the other hand, the components of eq. 9.5 are much characterized only by the center of ABSs, f_i and f_j. Therefore, the components of $\mathbf{X}$ are very different from each other and $\mathbf{X}^{-1}$ is robustly given compared to $\mathbf{A}^{-1}$ is. The advantage of GJPT procedure relative to LSF procedure is mainly from this different character.

9.3 Computational details

In the present study, normal gaussian functions are employed for ABSs,

$$f_i(\mathbf{r}) = C_i \exp(-\alpha_i r^2), \tag{9.15}$$

where C_i is an appropriate coefficient [10]. Eq. 9.8 is simplified, as follows;

$$V_i = \rho \sum_{\gamma} q_\gamma C_i \left(\frac{\pi}{\alpha_i}\right)^{\frac{3}{2}} \int_0^\infty 4\pi r^2 \frac{\mathrm{erf}(\sqrt{\alpha_i} r)}{r} h_{\alpha\gamma}(r) \mathrm{d}r \quad (i \in \alpha) \tag{9.16}$$

The exponents of the functions α_i and the number of ABSs are determined using the algorithm employed in the GAUSSIAN 03 [8, 11].

RISM and these expressions have been implemented by us in GAMESS [12]. A robust solver for RISM calculation is also implemented (see Appendix). The integration of eqs. 9.5 and 9.6 are calculated using the Obara-Saika recursions [13,14]. All calculations are performed with GAMESS [12] and Gaussian 03 [11].

9.4 Results and discussion

RISM-SCF-SEDD is applied to $\mathrm{H_2O}$, $\mathrm{C_2H_5OH}$, and HLi in aqueous phase. The calculation in this article is performed by restricted Hartree-Fock (RHF) with 6-31G* [15, 16] for $\mathrm{H_2O}$ and $\mathrm{C_2H_5OH}$, and with 6-31G** [17] for HLi [18]. The Lennard-Jones (LJ) parameters are summarized in Table 9.1. For comparison in the charge determination, two sets of grid points are prepared for the LSF procedure in the original RISM-SCF. The grid points employed in this work consist of radial part and angular part; the radial part is prepared from 5 to 50 Bohr (**set A**) and from 10 to 50 Bohr (**set B**) and angular part is based on deltoidal icositetrahedron (vertex 26).

Table 9.1: Lennard-Jones interaction parameters

	σ(Å)	ϵ(kcal/mol)
	H_2O[a]	
O	3.166	0.155
H	1.000	0.056
	C_2H_5OH	
C[b]	3.800	0.050
H[b]	2.500	0.050
O[b]	3.070	0.170
H(OH)[a]	1.000	0.056
	HLi	
H[c]	2.00	0.070
Li[d]	2.126452	0.018279

[a] From Ref. [2],
[b] From Ref. [27],
[c] From Ref. [28],
[d] From Ref. [29]

9.4.1 H_2O

H_2O is one of the typical molecules studied by many chemists. In this section, electrostatic structure (charge and dipole moment) and solvation structure calculated by RISM-SCF-SEDD and the original RISM-SCF are presented.

Table 9.2: Charges and dipole moment for H_2O derived from RISM-SCF-SEDD and the original RISM-SCF with **set A** and **set B**.

	$\mathbf{q}_S$	$\mathbf{q}_A$	$\mathbf{q}_B$
O	-0.974	-0.994	-0.993
H	0.487	0.497	0.496
dipole moment (D)	2.699	2.737	2.747

The charges evaluated by RISM-SCF-SEDD ($\mathbf{q}_S$) and the original RISM-SCF calculated using the **set A** and **set B** grid ($\mathbf{q}_A$ and $\mathbf{q}_B$, respectively) are shown in Table 9.2, where the dipole moment calculated by these methods are also shown. In the case of H_2O, there is little difference between $\mathbf{q}_A$ and $\mathbf{q}_B$. Although absolute value of $\mathbf{q}_S$ is somewhat smaller than $\mathbf{q}_{A/B}$,

these are very similar to each other. The difference in the dipole moment is also very small. By comparing with the experimental value of dipole moment (2.6 D), it is shown that RISM-SCF-SEDD and the original RISM-SCF give reasonable evaluation in electrostatic structures.

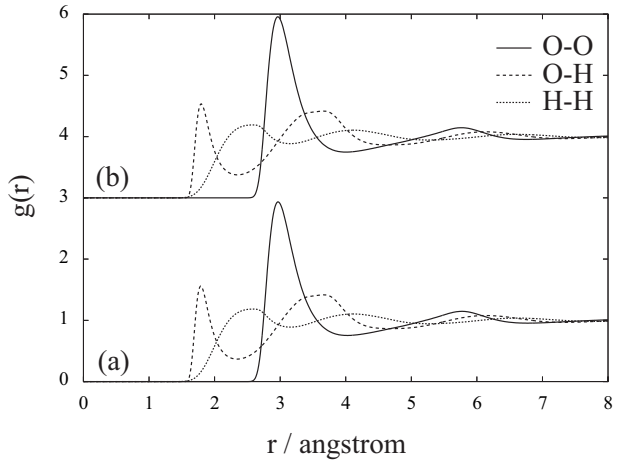


Figure 9.1: RDFs of H_2O derived form (a) RISM-SCF-SEDD and (b) the original RISM-SCF.

RDFs calculated by RISM-SCF-SEDD and those by the original RISM-SCF are shown in Fig. 9.1. The sharp peak located around 1.9 Å corresponds to hydrogen bond between H and O. These methods correctly evaluate the height and the positions of these peaks.

9.4.2 C_2H_5OH

C_2H_5OH has buried sites, C_1 of CH_3 group and C_2 of CH_2 group, and the effective charges of these atoms in gas phase have been studied in detail [19, 20].

$\mathbf{q}_S$ derived from RISM-SCF-SEDD and $\mathbf{q}_{A/B}$ derived from the original RISM-SCF of C_1, C_2, and O are shown in Fig. 9.2. They are plotted along the each RISM-SCF cycle. The charge at iteration cycle = 1 corresponds to that in gas phase. $\mathbf{q}_{A/B}$ significantly depends on the choice of grid sets even in gas phase. $\mathbf{q}_B$ of C_1 is almost zero but $\mathbf{q}_A$ is negative. The difference in charges derived from the grid set becomes large as iteration cycle increases. The change of $\mathbf{q}_A$ from gas phase to aqueous phase is not so large. On the other hand, $\mathbf{q}_B$ monotonously increases or decreases and eventually diverges. Such divergence sometimes occurs in the calculation of the original RISM-SCF when the buried sites exist in a solute molecule. In the case of RISM-

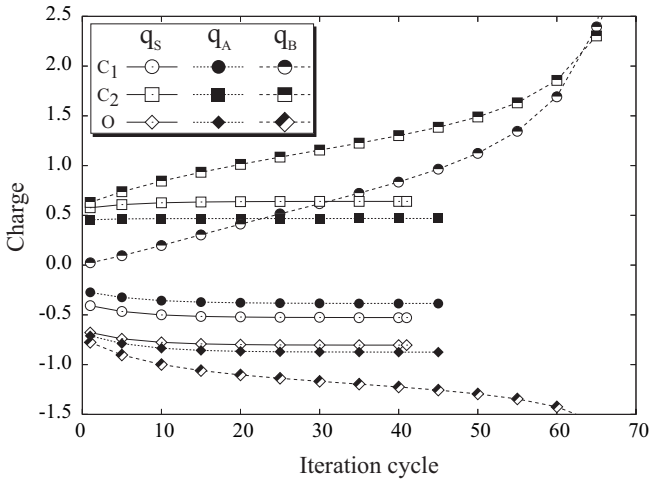


Figure 9.2: The change of $\mathbf{q}_S$, $\mathbf{q}_A$, and $\mathbf{q}_B$ of C_1, C_2, and O along the RISM-SCF iteration cycle.

SCF-SEDD, the grid set is not needed and the converged $\mathbf{q}_S$ is similar to the converged $\mathbf{q}_A$. The stability of $\mathbf{q}_S$ and the independence of grid points show that RISM-SCF-SEDD is superior to the original RISM-SCF when buried sites exist.

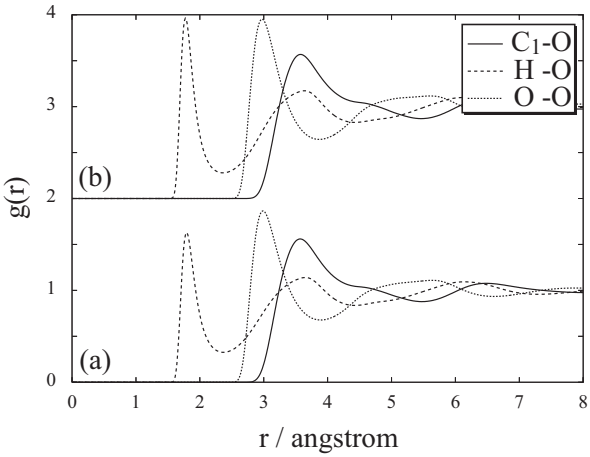


Figure 9.3: RDFs of C_2H_5OH derived form (a) RISM-SCF-SEDD and (b) the original RISM-SCF.

The RDFs calculated by RISM-SCF-SEDD and the original RISM-SCF are shown in Fig. 9.3. Those computed with $\mathbf{q}_S$ and with $\mathbf{q}_A$ look like very similar as in the case of H_2O, while the peaks corresponding to hydrogen bonding (~ 2.0Å) by RISM-SCF-SEDD is somewhat lower than that by the original one.

9.4.3 HLi

Table 9.3: Charges, $\mathbf{q}_S$, $\mathbf{q}_A$, and $\mathbf{q}_N$ for H site of HLi molecule calculated in gas phase and in aqueous phase.

	$\mathbf{q}_S$	$\mathbf{q}_A$	$\mathbf{q}_N$
H in gas phase	-0.756	-0.763	-0.730
H in aqueous phase	-1.044	-1.384	-0.887

HLi is a very simple molecule but the polarization induced by solvent is very large. The natural charges [21, 22] calculated with IEF-PCM [23] ($\mathbf{q}_N$), $\mathbf{q}_S$, and $\mathbf{q}_A$ are shown in Table 9.3. The corresponding gas values are also shown in Table 9.3. In gas phase, the values calculated by all these methods are almost the same with each other. However the charge deviation between H and Li in $\mathbf{q}_A$ is much stronger than those in $\mathbf{q}_S$ and $\mathbf{q}_N$ in aqueous phase.

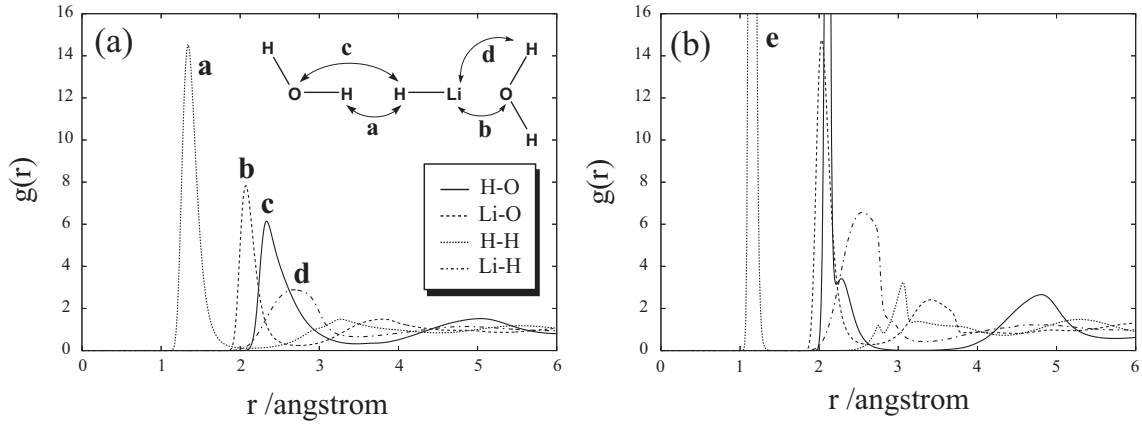


Figure 9.4: RDFs of HLi derived form (a) RISM-SCF-SEDD and (b) the original RISM-SCF. Schematic figures of solvation structure around Li and around H are shown.

RDFs provided by RISM-SCF-SEDD and the original RISM-SCF are shown in Figs. 9.4(a) and (b). The schematic solvation structures are shown in the right-upper side of Fig. 9.4(a). Sharp peaks located around 1.35 (peak **a**) and 2.09 Å (peak **b**) in Fig. 9.4(a) correspond to direct interactions, H-H and O-Li, respectively. They originate from the the strong Coulomb interaction between H-H and O-Li. Compared to peak **a** and peak **b**, the peaks located around 2.35 (peak **c**) and 2.80 Å (peak **d**) are broad, since they correspond to indirect interaction as

shown in the schematic figure. Peak **d** is moderately broad compared to peak **c**. The difference in these peaks shows that solvent H can move around a solute molecule more easily than the solvent O can. The solvation structures by the original RISM-SCF are very different from those by RISM-SCF-SEDD. For example, H-H (peak **e**) and Li-O RDFs, which correspond to direct interaction, are too high. In particular, peak **e** looks like that obtained in solid state. This is because the ESP derived from $\mathbf{q}_A$ is very strong.

In RISM-SCF procedure, ESP is expressed by point charges or ABSs that are determined so as to reproduce the ESP directly computed from the electron density, i.e., molecular orbitals (U_{MO}). The accuracy of the fitted ESP (U_{Fit}) by the point charges or ABSs can be examined by measuring the deviation from the original ESP, $\Delta U = U_{\mathrm{Fit}} - U_{\mathrm{MO}}$. It should be noted that the converged electron densities of RISM-SCF-SEDD and of the original RISM-SCF are slightly different from each other. We thus defined the deviation, ΔU^{SEDD} and ΔU^{ORG}, respectively.

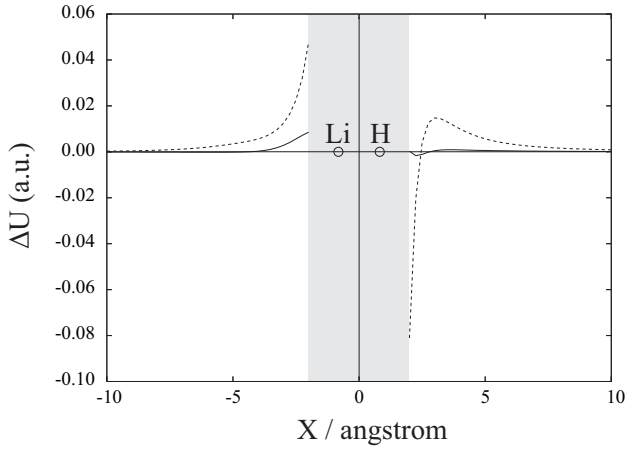


Figure 9.5: The difference of the ESP evaluated by RISM-SCF-SEDD and by the original RISM-SCF from that calculated by QM calculation along HLi bond; solid and dotted line correspond to ΔU^{SEDD} and ΔU^{ORG}. Shaded area show the region where the distance from solute site is shorter than the LJ parameter, $\sigma/2$.

In Fig. 9.5, the ΔU^{SEDD} and ΔU^{ORG} along the H-Li bond are shown. U_{Fit} reproduces U_{MO} very well in the case of RISM-SCF-SEDD. On the other hand, U_{Fit} by the original RISM-SCF ($\mathbf{q}_A$) is considerably different from the U_{MO}: ΔU^{ORG} is strongly positive, especially in the region of $X < 0$ and $2.5 < X < 5.0$ Å, while it is negative in the region close to the solute H ($2.0 < X < 2.5$ Å). These discrepancies seem to be insensitive to the choice of the grid points

and ΔU^{ORG} does not change so much even the grid range is shifted to the shorter distance (from 5 to 20 Bohr). This deviation in the fitted ESP is very crucial to determine the RDFs and is related to unphysical peaks in the original RISM-SCF, such as **e** depicted in Fig. 9.4(b).

9.5 Conclusions

We developed the new-generation of RISM-SCF, RISM-SCF-SEDD. The main advantages of the present method are that it includes explicitly spatial distribution of electron density and that it is grid free and robust compared to the original RISM-SCF. In this article, the independence of the grids and the origin of the stability of the calculation are discussed from the definition of the matrices used in the charge-determination.

RISM-SCF-SEDD was applied to $\mathrm{H_2O}$, $\mathrm{C_2H_5OH}$, and HLi in aqueous phase. The charges derived from the method are very stable and reasonable both in the case of $\mathrm{H_2O}$, which is typical example, and in the case of $\mathrm{C_2H_5OH}$, which has buried sites. In the case of HLi, the polarization in charges between H and Li is strongly enhanced in water. With RISM-SCF-SEDD, the origin of the polarization was clearly discussed from the solvation structures, which is difficult with the original RISM-SCF.

Appendix: A Robust solver for RISM

In RISM, the iterative calculation is needed. When the interaction between solute and solvent is very large, the calculation is sometimes diverge, especially at early stage of the computation. To solve RISM in stable manner, a robust solver is developed in this work.

Hypernetted-chain (HNC) closure is given by,

$$h_{\alpha\beta}(r) = \exp(\chi_{\alpha\beta}(r)) - 1, \tag{A1}$$

$$\chi_{\alpha\beta}(r) = -\frac{1}{k_B T} u_{\alpha\beta}(r) + h_{\alpha\beta}(r) - c_{\alpha\beta}(r) \tag{A2}$$

where $c_{\alpha\beta}(r)$ is the direct correlation function, $h_{\alpha\beta}(r)$ is total correlation function, k_B is Boltzmann constant and $u_{\alpha\beta}(r)$ is the pair potential between sites α and β. (A1) is very unstable when $\chi_{\alpha\beta}(r)$ is large.

With a parameter F, (A1) is rearranged by,

$$h_{\alpha\beta}(r) = \exp\left[F + (\chi_{\alpha\beta}(r) - F)\right] - 1 = \exp(F)\left[\sum_{n=0}\frac{1}{n!}(\chi_{\alpha\beta}(r) - F)^n\right] - 1. \qquad \text{(A3)}$$

When $(\chi_{\alpha\beta}(r) - F)$ is small enough, we can truncate the expansion up to $n = 1$. A new artificial 'closure' is then constructed, as follows;

$$h_{\alpha\beta}(r) = \begin{cases} \exp(F)\left[1 + (\chi_{\alpha\beta}(r) - F)\right] - 1 & (\chi_{\alpha\beta}(r) > F) \\ \exp(\chi_{\alpha\beta}(r)) - 1 & (\chi_{\alpha\beta}(r) \leq F) \end{cases} \qquad \text{(A4)}$$

When $F = 0$, (A4) corresponds to Kovalenko-Hirata type closure [24].

In general, the calculation of total correlation function, $h_{\alpha\beta}(r)$, by KH closure is more robust than that by HNC closure is. To evaluate correlation functions in stable manner especially at the beginning of the RISM iteration, F is gradually increased in a stepwise fashion. In each F value, iterative calculation between RISM and (A4) is performed until the convergence is achieved. When F becomes sufficiently large, the equation is switched from (A4) to the normal HNC closure (A1). This solver is more robust than the previous one used in our original RISM-SCF program.

Bibliography

[1] J. Tomasi, B. Mennucci, and R. Cammi, Chem. Rev. **105**, 2999 (2005).

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

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

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

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

[6] *Molecular Theory of Solvation*, edited by F. Hirata (Kluwer, Dordrecht, 2003).

[7] P. M. W. Gill, B. G. Johnson, J. A. Pople, and S. W. Taylor, J. Chem. Phys. **96**, 7178 (1992).

[8] The number of ABSs ($N_{\rm ABS}$) in eq. 9.2 and the set of exponents $\{\alpha_i\}$ employed in eq. 9.15 are systematically generated by the algorithm used in GAUSSIAN 03 [11], which we completely followed in the present work. At first a set of exponents, $\{\alpha_i'\}$, is prepared from the set of exponents of the primitive atomic orbitals used for the MO calculation of the solute molecule ($\{\beta_i\}$) by the following equation;

$$\alpha_i^{'} = \left\{ \begin{array}{ll} \beta_i & (i=1) \\ 2\beta_i & (i=2\cdots N_p) \end{array} \right. , \tag{9.17}$$

where N_p is the number of the primitive atomic orbitals. The final set of $\{\alpha_i\}$ used in eq. 9.15 is determined after modifying or removing the exponents that are too close to the neighboring ones. N_{ABS} is consequently determined after this modification. The $N_{\rm ABS}$ used in the present work was 17 for $\rm H_2O$, 51 for $\rm C_2H_5OH$, and 14 for HLi, respectively.

[9] This type of problems could be avoided by introducing the penalty function employed in the RESP [20] and PDCNP [25] procedure. However, we would like to emphasize that the obtained charges still depend on the parameters used in the penalty function. In the GJPT procedure, such function is not required.

[10] In this work, the coefficient, C_i, is determined so that $X_{ii} = 1$. If the coefficient is set as

$$C_i = \left(\frac{\alpha_i}{\pi}\right)^{\frac{3}{2}}, \tag{9.18}$$

Z_i in eq. 9.7 becomes simple to be 1.

[11] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02; Gaussian, Inc., Wallingford, CT, 2004.

[12] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).

[13] S. Obara and A. Saika, J. Chem. Phys. **84**, 3963 (1986).

[14] R. Ahlrichs, Phys. Chem. Chem. Phys. **8**, 3072 (2006).

[15] R. Ditchfield, W. J. Hehre, and J. A. Pople, J. Chem. Phys. **54**, 724 (1971).

[16] W. J. Hehre, R. Ditchfield, and J. A. Pople, J. Chem. Phys. **56**, 2257 (1972).

[17] P. C. Hariharan and J. A. Pople, Theoret. Chim. Acta **28**, 213 (1973).

[18] The internal geometry of H_2O is 109.47 degree and 1 Å for the HOH angle and OH distance, respectively [26].

[19] A. Morita and S. Kato, J. Phys. Chem. A **106**, 3909 (2002).

[20] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, J. Phys. Chem. **97**, 10269 (1993).

[21] J. P. Foster and F. Weinhold, J. Am. Chem. Soc. **102**, 7211 (1980).

[22] A. E. Reed, R. B. Weinstock, and F. Weinhold, J. Chem. Phys. **83**, 735 (1985).

[23] E. Cancés, B. Mennucci, and J. Tomasi, J. Chem. Phys. **107**, 3032 (1997).

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

[25] K. Ando, J. Phys. Chem. B **108**, 3940 (2004).

[26] L. X. Dang, J. E. Rice, J. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **113**, 2481 (1991).

[27] P. H. Lee and G. M. Maggiora, J. Phys. Chem. **97**, 10175 (1993).

[28] J. Gao and X. Xia, J. Am. Chem. Soc. **115**, 9667 (1993).

[29] W. L. Jorgensen, *OPLS and OPLS-AA Parameters for Organic Molecules, Ions, and Nucleic Acids* (Yale University, 1997).

General Conclusion

In this thesis, the author developed the methods focused on three-dimensional (3D) solvation structure and electronic structure. The achievements in this thesis are summarized, as follows.

In part I, the author proposed two theoretical methods; one is to reconstruct 3D solvation structure from radial distribution functions (RDFs) and the other is to calculate 3D solvation structure based on statistical mechanics. It was clearly shown that the 3D solvation structures obtained by these method showed the valuable informations, such as the strength of the interaction between solute and solvent molecules and the fluctuation of solvent molecules around a solute molecule.

In chapters 1 and 2, 3D solvation structure is reconstructed by RDFs. The 3D solvation structure is expanded with real solid harmonics S_{lm} and the coefficients are determined so that the RDFs calculated by the coefficients can reproduce the reference RDFs. The equations of the present method are so simple that the 3D distribution are calculated with reasonable computational cost. Moreover, the method can be used as a tool for the analysis in experimental studies because it is possible to employ the RDFs obtained by experimental method as the reference RDFs.

In chapter 3, the method to calculate directly 3D solvation structure was derived based on statistical mechanics. The 3D solvation structure is expanded with real solid harmonics, S_{lm} as in chapters 1 and 2. The coefficients are determined from the equation derived based on statistical mechanics. The present method can be considered as the expansion of Reference Interaction Site Model (RISM) because the equation with $l = 0$ and $m = 0$ corresponds to RISM equation. The 3D distribution evaluated here correctly reproduced the results obtained

by molecular simulation.

In chapter 4, another method to calculate 3D distribution function directly was proposed. Although the method derived in chapter 3 can calculate 3D distribution function accurately, the computational cost for large systems such as hydration structures around a protein becomes very high. By approximating the method in chapter 3 and using parallel computing technique, the author succeeded the reduction of computational cost and time in the present method. The present method was applied to a large protein called as Fv fragment. The evaluated hydration structures reproduced the waters determined by X-ray crystallography very well.

In chapter 5, the method derived in chapter 4 was applied to Bacteriorhodopsin (bR), which is a light-driven proton pump. The present method made it possible to calculate the 3D distribution functions of water oxygen site and water hydrogen site with reasonable computational cost. The hydrogen bonding network obtained by the present method correctly reproduced that proposed in previous works.

In Part II, the quantum mechanical calculation with solvation effect was performed using dielectric continuum model and RISM-SCF scheme.

In chapter 6, the electronic structure of Creutz-Taube complexes in aqueous phase was theoretically studied. There are two important requirements to understand the electronic structure of these complexes. One is a multiconfigurational description in the wave functions between localized state and delocalized state and the other is the solvation effect. In this work, two state model based on ab initio molecular orbital theory and dielectric continuum model were employed. The mechanism of the localization and delocalization of the wave function and the solvation effect on the electronic structure were elucidated by the present method.

In chapter 7, the weakpoint of the original RISM-SCF was overcome by introducing the modified charge assignment procedure proposed by Morita and Kato into RISM-SCF. This method was applied to the electronic structure of $[\mathrm{Ru(CN)_6}]^{4-/3-}$ in aqueous solution. In gas phase, the electronic structures of these complexes are similar to each other from the point of view of fitted point charges on each solute site. The nitrogen atoms tend to be negatively

charged and the point charges of buried sites (ruthenium and carbon atoms) become positive. However, the solvation effect considerably affects the electronic structure of $[\mathrm{Ru(CN)_6}]^{3-}$ compared to $[\mathrm{Ru(CN)_6}]^{4-}$. The electronic distribution between the buried sites and nitrogen atoms is more polarized in $[\mathrm{Ru(CN)_6}]^{4-}$ than in $[\mathrm{Ru(CN)_6}]^{3-}$.

In chapter 8, three algorithms of the charge assignment were examined. In addition to the conventional method employed in the original RISM-SCF, Mulliken-type and dual-type (Mulliken plus conventional) methods were proposed and applied to water. The interaction between solute and solvent waters evaluated by Mulliken-type method was the strongest, while that obtained by dual-type method was the smallest. With the dual-type method, the electronic structure of $\mathrm{NH_3BH_3}$ in water, which cannot be obtained by the original RISM-SCF, was successfully evaluated.

In chapter 9, the weakpoint of the original RISM-SCF was overcome by another strategy. In the present method, the author introduced auxiliary basis sets (ABSs) to incorporate spatial electron density distribution (SEDD) explicitly. By replacing the point charges employed in the original RISM-SCF by the ABSs, the instability in the charge assignment was drastically removed. In the case of a water molecule in aqueous phase, the new generation of RISM-SCF (RISM-SCF-SEDD) obtained reasonable solvation structures and the electronic structure, as the original RISM-SCF does. RISM-SCF-SEDD was also applied to $\mathrm{C_2H_5OH}$ and HLi, which cannot be calculated well by the original method. The obtained charges and solvation structures evaluated by the present method were reasonable.

A huge number of molecules make solvent system very complicated. However the complexity itself is the origin of the variety of the reactions in solvent. To tackle the interesting system, the author developed the methods to calculate 3D solvation structure (in Part I) and the electronic structure with solvation effect (in Part II) in this thesis. The two approaches from the point of view of solvent structure and the electronic structure of solute molecule make it possible to elucidate the mechanism of reactions in solvent at the molecular level theoretically.

List of Publications

Publications included in this thesis

Chapter 1

”A new method to reconstruct three-dimensional spatial distribution function from radial distribution function in solvation structure”

Daisuke Yokogawa, Hirofumi Sato, Shigeyoshi Sakaki

J. Chem. Phys. **2005**, *123*, 211102.

Chapter 2

”New evaluation of reconstructed spatial distribution function from radial distribution functions”

Daisuke Yokogawa, Hirofumi Sato, Shigeyoshi Sakaki

J. Chem. Phys. **2006**, *125*, 114102.

Chapter 3

”An integral equation theory for 3D solvation structure: A new procedure free from 3D Fourier transform”

Daisuke Yokogawa, Hirofumi Sato, Shigeyoshi Sakaki

Chem. Phys. Lett. **2006**, *432*, 595-599.

Chapter 4

”A highly parallelizable solvation structure theory based on Three-Dimensional Reference Interaction Site Model: Application to biomolecules.”

Daisuke Yokogawa, Hirofumi Sato, Takashi Imai, Shigeyoshi Sakaki

J. Chem. Phys. to be submitted.

Chapter 5

”The position of water molecules in Bacteriorhodopsin: A fragment Three-Dimensional Reference Interaction Site Model study”

Daisuke Yokogawa, Hirofumi Sato, Shigeyoshi Sakaki

J. Mol. Liq. submitted.

Chapter 6

”Localization or delocalization in the electronic structure of Creutz-Taube-type complexes in aqueous solution”

Daisuke Yokogawa, Hirofumi Sato, Yoshihide Nakao, Shigeyoshi Sakaki

Inorg. Chem. **2007**, *46*, 1966-1974.

Chapter 7 ”Electronic structure and solvation structure of $[\mathrm{Ru(CN)_6}]^{4-/3-}$ in aqueous solution: A RISM-SCF study”

Hirofumi Sato, Ippei Kawamoto, Daisuke Yokogawa, Shigeyoshi Sakaki

J. Mol. Liq. **2007**, *136*, 194-198.

Chapter 8

”Alternative couplings of solute-solvent interaction in RISM-SCF method”

Hirofumi Sato, Daisuke Yokogawa, Shigeyoshi Sakaki

J. Mol. Liq. **2007**, *136*, 190-193.

Chapter 9

”New generation of the reference interaction site model self-consistent field method: Introduction of spatial electron density distribution to the solvation theory”

Daisuke Yokogawa, Hirofumi Sato, Shigeyoshi Sakaki

J. Chem. Phys. **2007**, *126*, 244504.

Other publications

1. ”Bis(μ-silylene)-bridged dinuclear Rhodium(0) complex and its Palladium(0) and Platinum(0) analogues. Theoretical study of their electronic structure, bonding nature, and interconversion between μ-disilene-bridged form and bis(μ-silylene)-bridged form”
   Singo Nakajima, Daisuke Yokogawa, Yoshihide Nakao, Hirofumi Sato, Shigeyoshi Sakaki
   Organometallics **2004**, *23*, 4672-4681.

2. ”Polyatomic molecules in condensed phase: bond order index and solvation energy studied by RISM-SCF theory”
   Hirofumi Sato, Daisuke Yokogawa, Shigeyoshi Sakaki
   Condens. Matter Phys. **2007**, *10*, 373-380.