Theoretical Study on Microscopic Solvation Effects in Glycine Aqueous Solution

Yukichi Kitamura

Contents

Chapter 1

General Introduction

1.1. Introduction

Modern chemistry has been developed on the basis of the concept “molecule” which consists of an assembly of atoms connected each other by chemical (covalent) bonds among them, although a super-“molecule” might be bound not only by chemical bonds but also by non-covalent bonds. Thus, chemical phenomena can be explained through the molecular properties and by their mutual transformations. Incidentally, it is well known that the very small (**micro**) particles, such as nuclei and electrons, show the strange behaviors that are impossible to be described by classical mechanics (CM) for the larger (**macro**) systems; instead, they should be described by quantum mechanics (QM). In the context of QM, all elementary particles (and light) exhibit the properties of not only particles but also waves (cf., the wave–particle duality of matter), which is the characteristic of so-called “quantum” systems. Today, it is indispensable and common in modern chemistry to deal with molecules as such quantum systems that consist of a pair of CM nuclei and QM electrons, for understanding chemical phenomena deeply.

The mathematical formulations of QM are described typically by the time-dependent Schrödinger equation (SE), which is in general a linear partial differential equation, generating the time-evolution of the electron's wave function

Ψ.

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi, \tag{1.1}$$

where i is the imaginary unit, $\hbar$ is the reduced Planck constant, and $\hat{H}$ is the molecular Hamiltonian operator and characterizes the total molecular energy of any given wave function depending on the situation. For bound states of molecules, the time-dependent SE can be separated out, providing the time-independent SE for electrons in a molecule,

$$\hat{H}\Psi = E\Psi. \tag{1.2}$$

Here the Hamiltonian itself is independent on time. Moreover, the wave functions that are obtained by solving the time-independent SE can form a number of standing waves, called stationary states (also called "molecular orbitals").

In principle, the wave function is the most complete description (the physical properties and time evolution of quantum systems) that can be obtained from a molecular system. The only systems that can be solved exactly, however, are those composed of only one or two particles, where the latter transforms the problem into a pseudo-particle with a reduced mass ($\mu = m_1 m_2 / (m_1 + m_2)$). Thus, systems composed of more than two particles cannot be analytically solved. Nevertheless, it is possible to achieve an approximate separation of the degree of freedom based on physical properties. For many-particle systems, two separations of the degrees of freedom are often introduced into quantum system: the Born-Oppenheimer approximation (for motions of nuclei and electrons) and the Hartree-Fock (HF) approximation (for interactions between two electrons), where the average electron-electron repulsion is incorporated.

In spite of introduction of such approximation, it remains difficult to solve SE generally, because it is necessary to perform a very large number of mathematical operations to obtain the results with chemical accuracy. Accordingly, the applications of QM to the understanding of practical targets in the field of chemistry are available with explosive advances in electronic computers. In addition to development of hardware and programmatic technique, several excellent theoretical methods have been also developed to overcome the limitations of HF method and to improve the computational efficiency; for example, Many-Body Perturbation Theory (MBPT), Configuration Interaction (CI), Coupled Cluster (CC), density functional theory (DFT), and so on. Recently, the results by extensive high-level calculations have been reported for atoms and small molecules [1, 2], which is correct over 40 digits.

On the other hand, since most of chemical processes proceed in solution, it is also essentially important to take the solvent effects into consideration. As a simple extension from the traditional QM calculations, the supermolecule model which involves a few explicit solvent molecules has been often utilized to include the solvent effects naively. However, these calculations are exquisitely sensitive to the configuration of solvent molecules. Also, the supermolecule model is not able to sufficiently consider both long-range solvation effects and the thermal contribution.

Under the circumstances, computer simulation methods are suitable approach to understand chemical processes in condensed systems and can provide us with microscopic information. Molecular dynamics (MD) method [3], one of computer simulation methods, solves numerically a set of Newtonian equations of motion. MD simulations are suitable for studying time-dependent phenomena (such as substance transport, diffusion and vibrational spectra), and can yield information about both

static and dynamic physical properties by assuming the ergodicity of a given molecular system in equilibrium (the time average of the equilibrium MD calculation is usually utilized instead of the ensemble average).

In this thesis, I have selected the simplest amino acid, glycine, as model system, in particularly the neutral-forms having several conformers which can be converted by rotation of three dihedral angles (see **Figure 1.1**). I have investigated an explicit solvent effect on (i) the free energy (FE) difference between their conformers and (ii) vibrational spectra in solution. To extract the explicit solvent effects, I have employed the dielectric continuum model (DCM) method and the quantum mechanical / molecular mechanical (QM/MM) method for present research. In the former study, the relative FEs of conformers in solution were found to become smaller than the relative potential energies (PEs) in the isolated state due to the electrostatic compensation of PE destabilization. In the latter one, I have proposed a new theoretical methodology of vibrational spectra in solution, employing two kinds of analytical Hessians. I have applied it to glycine molecule and it is shown that the present approaches would provide us with a plausible tool to analyze vibrational properties.

1.2. Theoretical Methods for Quantum System in Condensed Phase

Liquid solutions play in fact a fundamental role in chemistry, and a number of theoretical methods have been developed to take successfully the solvent effects into consideration. Additionally, from an aspect of the evolution of scientific research, chemistry shares with physics, biology, engineering, geology, and all of the other

branches of sciences, and these evolutions depend ultimately on the widespread availability of efficient computers.

As the theoretical method to consider the quantum description (for electrons) into the quantum system in condensed systems, there is the well-known Car-Parrinello MD (CP-MD) method with a number of applications for the solution chemistry [4]. In the CP-MD method, the motions of molecule are considered as a semi-classical dynamics approach (where the electrons are treated quantum mechanically while the motions of nucleus are treated classically) and the evolution of a system is described by solving the (extended) Lagrange functions introducing the wave function parameters as variables with fictive "masses" in the dynamics. Even with the CP technique, however, the direct applications of the CP-MD method are still computationally very expensive and are restricted to those systems with a small number of atoms. Therefore, CP-MD method can only give a limited sampling of the phase space.

In the circumstances, a multiscale model has been proposed for the molecular simulation that can deal with more complex systems and longer simulation time. The central concept of a multiscale model can be translated into a simple formal expression (Eq. 1.3). The whole system is partitioned into two parts which are defined as the solute system (S) and the solvent bath (B). The total Hamiltonian consists of the partial Hamiltonians of these two S and B and the interaction one between S and B, as follows,

$$\hat{H}^{\mathrm{SB}}\left(\mathbf{R}^{\mathrm{s}},\mathbf{R}^{\mathrm{b}}\right)=\hat{H}^{\mathrm{S}}\left(\mathbf{R}^{\mathrm{s}}\right)+\hat{H}^{\mathrm{B}}\left(\mathbf{R}^{\mathrm{b}}\right)+\hat{H}^{\mathrm{int}}\left(\mathbf{R}^{\mathrm{s}},\mathbf{R}^{\mathrm{b}}\right), \tag{1.3}$$

where $\mathbf{R}^{\mathrm{s}}$ and $\mathbf{R}^{\mathrm{b}}$ indicate the each set of variables or parameters for the S and B parts, respectively. The total computational cost can be drastically reduced by

introducing an approximate description into B part which occupy the largest part in whole one.

If the B is handled by a continuum dielectric, such models are called a dielectric continuum model (DCM) and the polarizable continuum model (PCM) [5] is often used as one of DCM. Although the PCM is useful in the case of requiring the high-accuracy QM calculation for the low calculation cost (e.g. an excited state and a heavy metal complex systems), it is well known that the PCM has some problems such as the incorrectness of the intermolecular interactions of hydrogen bonds (HBs) between solutes and solvent molecules. Also, the basic QM continuum model can treat the solvation effects caused by the surrounding environment that is a very dilute and isotropic solution at equilibrium (given temperature and pressure).

On the other hand, the quantum mechanical / molecular mechanical (QM/MM) method [6] is often utilized as another treatment that only the reactive parts (S) in the whole system are treated quantum-mechanically while the other parts (B) are treated molecular-mechanically to reduce the computational cost drastically. Moreover, the MD method combined with the QM/MM method (QM/MM-MD method) [7] is accordingly very useful as the statistical sampling method for solution chemistry. The microscopic solute-solvent interactions in solution can be considered with relatively low cost and the QM/MM-MD method can be also applied to the behavior of molecules under non-equilibrium condition.

1.2.1. Dielectric Continuum Model (DCM)

A DCM considers the solvent as a uniform polarizable medium with a dielectric constant ε and with solute molecule placed in a suitably shaped hole (cavity)

in the medium. Since the number of the degrees of freedom of B are not explicitly taken into account, the B part can be parameterized by ε only, while the S part is described by the coordinates of electrons and nuclei, i.e., $\mathbf{r}^{N_{\mathrm{ele}}} = \{\mathbf{r}_1, \cdots, \mathbf{r}_{N_{\mathrm{ele}}}\}$ and $\mathbf{R}^{N_{\mathrm{nuc}}} = \left\{\mathbf{R}_1, \cdots, \mathbf{R}_{N_{\mathrm{nuc}}}\right\}$, the latter denoting a set of all the N_{nuc} Cartesian coordinate vectors of the solute molecule. The Hamiltonian in solution (Eq. (1.3)) can be expressed as an effective Hamiltonian $\hat{H}_{\mathrm{eff}}^{\mathrm{SB}}$ as follows,

$$\hat{H}_{\mathrm{eff}}^{\mathrm{SB}}\left(\mathbf{R}^{\mathrm{s}}, \varepsilon\right) = \hat{H}^{\mathrm{S}}\left(\mathbf{R}^{\mathrm{s}}\right) + \hat{H}^{\mathrm{int}}\left(\mathbf{R}^{\mathrm{s}}, \varepsilon\right), \tag{1.4}$$

$$\hat{H}^{\mathrm{int}}\left(\mathbf{R}^{\mathrm{s}}, \varepsilon\right) = \hat{H}^{\mathrm{int}}\left(\mathbf{R}^{N_{\mathrm{nuc}}}, \mathbf{r}^{N_{\mathrm{ele}}}, \varepsilon\right) = \sum_{\alpha} Z_{\alpha} \Phi_{\sigma}\left(\mathbf{R}_{\alpha}\right) - \sum_{i} \Phi_{\sigma}\left(\mathbf{r}_i\right). \tag{1.5}$$

Here, $\Phi_{\sigma}\left(\mathbf{r}\right)$ is the electrostatic field generated by the polarized dielectric, $\varepsilon\left(\mathbf{r}\right)$, at the position **r**. Furthermore, the solute-solvent interaction (electrostatic) contribution to the total energy is obtained by

$$\left(H^{\mathrm{S}} + H^{\mathrm{int}}\right)\Psi^{(\mathrm{f})} = \left(E^{\mathrm{S}} + E^{\mathrm{int}}\right)\Psi^{(\mathrm{f})}, \tag{1.6a}$$

and

$$E^{\mathrm{int}} = \int \Psi^{(\mathrm{f})*} \hat{H}^{\mathrm{int}} \Psi^{(\mathrm{f})} d\mathbf{r}^{N_{\mathrm{ele}}} = \int \rho\left(\mathbf{r}\right) \Phi_{\sigma}\left(\mathbf{r}\right) d\mathbf{r}; \quad d\mathbf{r}^{N_{\mathrm{ele}}} \equiv \prod_{i}^{N_{\mathrm{ele}}} d\mathbf{r}_i, \tag{1.6b}$$

where, an index (f) indicates that the function is provided by the final (converged) solution of the solute system placed in the cavity, and $\rho\left(\mathbf{r}\right)$ is the sum of the nuclear charge distribution $\rho_{\mathrm{nuc}}\left(\mathbf{r}\right)$ and the electron density function $\rho_{\mathrm{ele}}\left(\mathbf{r}\right)$ at the position **r**,

$$\rho\left(\mathbf{r}\right) = \rho_{\mathrm{nuc}}\left(\mathbf{r}\right) + \rho_{\mathrm{ele}}\left(\mathbf{r}\right) = \sum_{\alpha} Z_{\alpha} \delta\left(\mathbf{r} - \mathbf{R}_{\alpha}\right) - \sum_{i}^{N_{\mathrm{ele}}} \int \Psi^{(\mathrm{f})*} \Psi^{(\mathrm{f})} \delta\left(\mathbf{r} - \mathbf{r}_i\right) d\mathbf{r}^{N_{\mathrm{ele}}}. \tag{1.7}$$

Since $\hat{H}^{\mathrm{int}}$ is a one-electronic operator, the calculation cost is not time-consuming.

The QM continuum model requires to solve the general Poisson equation

within a QM framework. The general Poisson equation can be simplified to

$$\begin{cases} -\nabla^2\Phi(\mathbf{r}) = 4\pi\rho(\mathbf{r}) = 4\pi\left(\rho_{\text{nuc}}(\mathbf{r}) + \rho_{\text{ele}}(\mathbf{r})\right): & \varepsilon(\mathbf{r}) = 1 \quad \mathbf{r} \in V_{\text{in}} \\ -\varepsilon\nabla^2\Phi(\mathbf{r}) = 0: & \varepsilon(\mathbf{r}) = \varepsilon \quad \mathbf{r} \in V_{\text{out}} \end{cases} \tag{1.8}$$

where $\Phi(\mathbf{r})$ is the total electrostatic field (which is the sum of the electrostatic potential $\Phi_S(\mathbf{r})$ for the QM solute molecules and $\Phi_\sigma(\mathbf{r})$ generated by the polarized dielectric), V_{in} and V_{out} are the volumes inside and outside the cavity, respectively. Also, it is necessary to satisfy two boundary conditions.

$$\begin{cases} \left[\Phi_{\text{in}}(\mathbf{r}^*)\right] = \left[\Phi_{\text{out}}(\mathbf{r}^*)\right] = 0 \\ \dfrac{\partial}{\partial\mathbf{n}}\left[\Phi_{\text{in}}(\mathbf{r}^*)\right] - \varepsilon\dfrac{\partial}{\partial\mathbf{n}}\left[\Phi_{\text{out}}(\mathbf{r}^*)\right] = 0 \end{cases} \tag{1.9}$$

These conditions express the continuity of the potential across the cavity surface $\mathbf{r}^*$. Eq. (1.7)-(1.9) are the basic elements to use in the elaboration of solvation methods according to standard electrostatics, but there is not an analytical solution. The DCMs have different options to describe Φ_σ and the electrostatic interaction energy between ρ and Φ_σ. Thus, the electrostatic problem are solved by the numerical approaches following an iterative procedure, which are classified into six categories, (1) the apparent surface charge (ASC) methods, (2) the multipole expansion methods, (3) the generalized Born approximation (GBA), (4) the image charge (IC) methods, (5) the finite element (FE) methods, and (6) the finite difference (FD) methods.

The solute-solvent interactions correlated with the quantum system are limited to those of electrostatic origin (Eq. (1.6)). Other contributions (exchange/dispersion) are considered by adding the (free) energy components after solving the electrostatic problem, and correspond to the cavity energy (entropic contributions and loss of

solvent-solvent van der Waals interactions) and stabilization energy (due to van der Waals interactions between the solute and solvent). These terms are parameterized by the total solvent accessible surface (SAS) area or constant ξ specific for each atom type.

$$G_{\text{cavity}} + \Delta G_{\text{dispersion}} = \gamma \text{SAS} + \beta \tag{1.10a}$$

$$G_{\text{cavity}} + \Delta G_{\text{dispersion}} = \sum_{i}^{\text{atoms}} \xi_i S_i \tag{1.10b}$$

Here, γ, β and ξ are parameters being determined by fitting to experimental solvation data. Finally, the solvation free energy may thus be written as in Eq. (1.11).

$$\Delta G_{\text{solvation}} = \Delta G_{\text{cavity}} + \Delta G_{\text{dispersion}} + \Delta G_{\text{elec}} \tag{1.11}$$

1.2.2. QM/MM-MD method

For the condensed system, *ab initio* QM/MM-MD method has been often adopted to make several configurations of the instantaneous solvent structures reflected explicitly in the solute electronic state. In the *ab initio* QM/MM-MD method, the simultaneous equations of motion of the solute-solvent system are written as follows,

$$m_A \frac{d^2\mathbf{R}_A}{dt} = -\nabla_A V(\mathbf{R}^N, \mathbf{R}^{N_B}), \tag{1.12a}$$

$$M_M \frac{d^2\mathbf{R}_M}{dt} = -\nabla_M V(\mathbf{R}^N, \mathbf{R}^{N_B}), \tag{1.12b}$$

where m_A and $\mathbf{R}_A$ are the mass of a solute QM atom *A* and its Cartesian coordinate vectors, and M_M and $\mathbf{R}_M$ are the (effective) mass of a solvent MM particle *M* and its Cartesian coordinate vectors, respectively, while $V(\mathbf{R}^N, \mathbf{R}^{N_B})$ is the total system

potential energy function, $\mathbf{R}^N$ and $\mathbf{R}^{N_B}$ denoting a set of all the atomic Cartesian coordinate vectors of the *N*-atomic solute molecule and those of solvent molecules consisting of N_B atoms as a whole. $V(\mathbf{R}^N, \mathbf{R}^{N_B})$ is subsequently obtained clearly as Eq. (1.12) by treating the following total Hamiltonian $\hat{H}$:

$$\hat{H} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM}, \tag{1.13}$$

where, in the RHS, the first two terms $\hat{H}_{QM}$ and $\hat{H}_{MM}$ stand for the standard Hamiltonian of the QM and the MM systems, respectively, while the last QM/MM term $\hat{H}_{QM/MM}$ denotes the interaction between the QM and MM system, and is expressed as a sum of electrostatic (est) and van der Waals (vdW) terms:

$$\hat{H}_{QM/MM} = \hat{H}^{est}_{QM/MM} + \hat{H}^{vdW}_{QM/MM}. \tag{1.14}$$

The est term $\hat{H}^{est}_{QM/MM}$ is obtained by

$$\hat{H}^{est}_{QM/MM} = \sum_M q_M \hat{V}_{QM}(\mathbf{R}_M) = \sum_M q_M \left[-\sum_s \frac{1}{R'_{sM}} + \sum_A \frac{Z_A}{R_{AM}} \right], \tag{1.15}$$

where $\hat{V}_{QM}(\mathbf{R}_M)$ is represented as the sum of coulomb interactions induced by the electrons and the cores of QM atoms. In these expressions, q_M is the point charge on the *M*-th MM atom in solvent molecules, R'_{sM} ($=|\mathbf{r}_s - \mathbf{R}_M|$) is the distance between the *s*-th QM electron of the solute molecule and the *M*-th MM atom, Z_A is the core charge of the *A*-th QM atom, and R_{AM} ($=|\mathbf{R}_A - \mathbf{R}_M|$) is the distance between the *A*-th QM atom and the *M*-th MM one. The vdW term $H^{vdW}_{QM/MM}$ is expressed using a number of 6-12 Lennard-Jones (LJ) functions:

$$H_{\mathrm{QM/MM}}^{\mathrm{vdW}} = \sum_{A}\sum_{M} \varepsilon_{AM}\left[\left(\frac{r_{AM}}{R_{AM}}\right)^{12} - 2\left(\frac{r_{AM}}{R_{AM}}\right)^{6}\right], \tag{1.16}$$

where ε_{AM} and r_{AM} are a couple of LJ parameters for the *A*-th QM atom interacting with the *M*-th MM atom. The general AMBER force field (GAFF) [8] was often used to describe LJ interactions between the solute QM and solvent MM molecules. Hence, the above total system potential energy *V* is then obtained by

$$V(\mathbf{R}^N, \mathbf{R}^{N_{\mathrm{B}}}) = \left\langle \Psi(\mathbf{R}^N) \middle| \hat{H}_{\mathrm{QM}}(\mathbf{R}^N) + \hat{H}_{\mathrm{QM/MM}}(\mathbf{R}^N, \mathbf{R}^{N_{\mathrm{B}}}) \middle| \Psi(\mathbf{R}^N) \right\rangle + \hat{H}_{\mathrm{MM}}$$

$$= V_{\mathrm{SB}} + V_{\mathrm{MM}} \tag{1.17a}$$

where the instantaneous wave functions Ψ and the instantaneous solute energy V_{SB} in solution are obtained by solving the following associated Schrödinger equation,

$$\left[\hat{H}_{\mathrm{QM}} + \hat{H}_{\mathrm{QM/MM}}\right]\left|\Psi\right\rangle = V_{\mathrm{SB}}\left|\Psi\right\rangle, \tag{1.17b}$$

at an instantaneous solution structure ($\mathbf{R}^N(t)$, $\mathbf{R}^{N_{\mathrm{B}}}(t)$) obtained by solving Eq. (1.12).

1.3. Statistical Mechanical Understanding of Chemical Processes in Solvation and Solution Reactions

1.3.1. Free Energy Methods

There are some methods to obtain the free energy (FE) changes along the given reaction coordinates, such as the umbrella sampling method [9], free energy perturbation (FEP) method [10], thermodynamic integration (TI) method [11], energy representation (ER) method [12, 13]. All the thermodynamic quantities can be derived,

in principle, in closed expressions by using the partition function Q for the system. However, for a collection of many interacting particles (e.g., at each thermodynamic state in condensed phase), it is not possible to construct it explicitly. Nevertheless, it is possible to estimate their *differences* in Q and those *derivatives* of Q by a representative sample of the system. In general, the configurations to estimate Q are generated by either molecular dynamics (MD) or Monte Carlo (MC) procedures.

1.3.1.1. Free Energy Perturbation Method

The partition function Q is obtained by summing over all energy states E_i for the whole system. Classically, because of the closely spaced energy levels, the discrete sum over energy states, that is a definition of Q, can be replaced by an integral over all coordinates (**q**) and momenta (**p**), called the phase space,

$$Q = \sum_i^{\infty} e^{-E_i/k_{\mathrm{B}}T} = \int e^{-E(\mathbf{q},\mathbf{p})/k_{\mathrm{B}}T} d\mathbf{q}d\mathbf{p} \,. \tag{1.18}$$

The thermodynamic functions, such as the internal energy U, Helmholtz free energy A (=U-TS) and entropy S can be, thus, calculated from Q.

$$U = k_{\mathrm{B}}T^2 \left(\frac{\partial \ln Q}{\partial T} \right)_V , \tag{1.19a}$$

$$A = U - TS = -k_{\mathrm{B}}T \ln Q \,, \tag{1.19b}$$

$$S = \frac{U - A}{T} = k_{\mathrm{B}}T \left(\frac{\partial \ln Q}{\partial T} \right)_V + k_{\mathrm{B}} \ln Q \,. \tag{1.19c}$$

In order to calculate the partition function Q, we need to know all possible quantum states for the system, but it is computationally difficult to obtain them within a reasonable statistical error for these quantities.

In free energy perturbation (FEP) method, we can consider the energy difference between two states A and B described by two different energy functions E^{A}

and E^{B}, respectively. The FE difference is obtained by Eq. (1.20) involving a ratio of the corresponding partition functions,

$$A_{\mathrm{A}} - A_{\mathrm{B}} = -k_{\mathrm{B}}T\left(\ln Q_{\mathrm{A}} - \ln Q_{\mathrm{B}}\right) = -k_{\mathrm{B}}T\ln\frac{Q_{\mathrm{A}}}{Q_{\mathrm{B}}}. \tag{1.20}$$

If the ergodic hypothesis is valid for the present system, a macroscopic observable quantity can be calculated as an average over a corresponding microscopic one.

$$\left\langle \Delta A \right\rangle_{M} = k_{\mathrm{B}}T\ln\left(\frac{1}{M}\sum_{i=1}^{M} e^{\left(E_i^{\mathrm{B}} - E_i^{\mathrm{A}}\right)/k_{\mathrm{B}}T}\right) = k_{\mathrm{B}}T\ln\left\langle e^{\Delta E_{\mathrm{BA}}/k_{\mathrm{B}}T}\right\rangle_{M}. \tag{1.21}$$

Actually, a number of intermediate states connecting between A and B are introduced to obtain the FE difference with sufficiently statistical precision, and can be described in terms of a coupling parameter λ ($0 \le \lambda \le 1$). If the energy difference $E^{\mathrm{B}} - E^{\mathrm{A}}$ is sufficiently small compared with $k_{\mathrm{B}}T$, the ensemble average will show much better convergence than the approach employing directly the partition function.

1.3.1.2. Thermodynamic Integration Method

In the thermodynamic integration (TI) method, the partition function and FE are considered as functions of λ,

$$A(\lambda) = -k_{\mathrm{B}}T\ln Q(\lambda), \tag{1.22}$$

and the FE difference ΔA is estimated by using the differentiation of the expression (eq. (1.22)),

$$\frac{\partial A}{\partial \lambda} = -\frac{k_{\mathrm{B}}T}{Q}\frac{\partial Q}{\partial \lambda} = \frac{\partial V}{\partial \lambda}. \tag{1.23}$$

Then, by integrating over λ, ΔA is approximately evaluated by the summation of its discrete estimations,

$$\Delta A = \sum_i \left\langle \frac{\partial V(\lambda_i)}{\partial \lambda_i} \right\rangle \Delta\lambda_i \,. \tag{1.24}$$

When selecting a proper coordinate **s** as the coupling parameter λ, the FE change ΔA_i for changing from $\mathbf{s}_i$ to $\mathbf{s}_{i+1}$ (= $\mathbf{s}_i + \Delta\mathbf{s}_i$) can be represented by

$$\Delta A_i = A_{i+1} - A_i = \left\langle \frac{\partial V(\mathbf{s}_i)}{\partial \mathbf{s}_i} \right\rangle_i \Delta\mathbf{s}_i = \left\langle \mathbf{F}^{\mathrm{FE}}(\mathbf{s}_i) \right\rangle_i \Delta\mathbf{s}_i \,, \tag{1.25}$$

where $\langle \cdots \rangle_i$ donote the time average constraining the solute molecule so as to correspond to the interpolated structures $\mathbf{s}_i$ and $\left\langle \mathbf{F}^{\mathrm{FE}}(\mathbf{s}_i) \right\rangle_i$ is the average force exerted on the solute molecule at $\mathbf{s}_i$. Finally, ΔA is obtained as

$$\Delta A = \sum_{i=1}^{N_\mathrm{p}} \Delta A_i = -\int_{s_1}^{s_{N_\mathrm{p}}} \left\langle \mathbf{F}^{\mathrm{FE}}(\mathbf{s}_i) \right\rangle_i d\mathbf{s}_i \,. \tag{1.26}$$

1.3.2. Free Energy Gradient (FEG) Methods

The reaction path on the multidimensional potential energy surfaces (PESs) of the solution reaction system can be traced by using the explicit supermolecule models or implicit PCM. However, these PESs are too high-dimensional, and then it is difficult to search the TS structure with respect to all degrees of freedom for the whole system. In such case, the free energy surface (FES) has been often estimated approximately along only a certain reaction coordinate since a vast number of solvent molecules participate in solution reaction system. However, for purpose to identify accurately these structures on the FES, the stable-state (SS) and TS geometries should be optimized with respect to all degrees of freedom of the solute in solution.

Under the circumstances, being analogous to the (potential) energy gradient method on the Born-Oppenheimer PES obtained by the molecular orbital (MO) theory, Nagaoka and co-workers have developed the free energy gradient (FEG) method [14-19] which utilizes the forces on the FES. In equilibrium system, the ensemble average of an observable is to be replaced as the time average calculated over the trajectory in the MD method, provided its trajectory is ergodic. Then the forces $\mathbf{F}^{\mathrm{FE}}(\mathbf{q}^{\mathrm{S}})$ on the FES are equal to the time average of the forces acting on each atom of the solute molecule with the geometry $\mathbf{q}^{\mathrm{S}}$.

$$\mathbf{F}^{\mathrm{FE}}(\mathbf{q}^{\mathrm{S}}) = -\frac{\partial A(\mathbf{q}^{\mathrm{S}})}{\partial \mathbf{q}^{\mathrm{S}}} = -\left\langle \frac{\partial V_{\mathrm{SB}}(\mathbf{q}^{\mathrm{S}})}{\partial \mathbf{q}^{\mathrm{S}}} \right\rangle, \tag{1.27}$$

where $A(\mathbf{q}^{\mathrm{S}})$ is the Helmholtz FE, and $V_{\mathrm{SB}}(\mathbf{q}^{\mathrm{S}})$ is the sum of both the solute potential energy and the interaction energy between the solute species and solvent molecules (cf. Eq. (1.17)). The brackets denote the equilibrium ensemble average,

$$\langle \cdots \rangle = \frac{\int d\mathbf{q}^{\mathrm{B}} (\cdots) \exp(-\beta \mathrm{V}(\mathbf{q}^{\mathrm{S}}))}{\int d\mathbf{q}^{\mathrm{B}} \exp(-\beta \mathrm{V}(\mathbf{q}^{\mathrm{S}}))}, \tag{1.28}$$

where $\mathbf{q}^{\mathrm{B}}$ denotes the solvent coordinates as a whole. By the FEP theory, the FE difference at an optimization cycle *i*, ΔA_i, is described as follows.

$$\Delta A_i = A_{i+1} - A_i = -k_{\mathrm{B}} T \ln \left\langle \exp\left[-\beta \left\{ V_{\mathrm{SB}}(\mathbf{q}^{\mathrm{S}}_{i+1}) - V_{\mathrm{SB}}(\mathbf{q}^{\mathrm{S}}_{i}) \right\} \right] \right\rangle_i, \tag{1.29}$$

where $\mathbf{q}^{\mathrm{S}}_i$ and $\mathbf{q}^{\mathrm{S}}_{i+1}$ are solute structures at optimization step *i* and *i*+1, respectively. The subscript *i* in the average $\langle \cdots \rangle_i$ in Eq. (1.29) means that the average is taken over the sampling at $\mathbf{q}^{\mathrm{S}}_i$. The optimization cycle was repeated for the FEP treatment, until

the following condition, i.e., the zero gradient condition, is satisfied,

$$\left\langle \frac{\partial V_{\mathrm{SB}}(\mathbf{q}^{\mathrm{s}})}{\partial \mathbf{q}^{\mathrm{s}}} \right\rangle_i \approx 0 . \tag{1.30}$$

1.4. Computational Procedure for Free Energy Gradient (FEG) Method

1.4.1. Structure Optimization via Free Energy Gradient (FEG) Method

The free energy gradient (FEG) method [14-19], utilizing the force on the FES, has been often adopted to obtain the equilibrium structures in solution with full optimization with respect to all degrees of freedom of the solute molecules. By time-averaging forces acting on each constituent atom of a solute molecule over the equilibrium distribution with respect to all the solvent molecules, the forces on the FES, $\mathbf{F}\left(\mathbf{q}^{\mathrm{s}}\right)$, are obtained as a function of solute's Cartesian coordinates $\mathbf{q}^{\mathrm{s}}$.

In the FEG method with the following Broyden-Fletcher-Goldfarb-Shanno (BFGS) procedure [20-23], the $(i+1)$th reactant structure $\mathbf{q}^{\mathrm{s}}_{i+1}$ is taken as,

$$\mathbf{q}^{\mathrm{s}}_{i+1} = \mathbf{q}^{\mathrm{s}}_i + \Delta\mathbf{q}^{\mathrm{s}}_i = \mathbf{q}^{\mathrm{s}}_i - \alpha_i \mathbf{H}'_i \mathbf{F}\left(\mathbf{q}^{\mathrm{s}}_i\right), \tag{1.31}$$

where α_i is the step length, and the matrix $\mathbf{H}'_i$ is the approximate representation of the inverse Hessian matrix, which is defined as,

$$\mathbf{H}'_i = \left[\mathbf{I} - \frac{\mathbf{s}_i \mathbf{y}_i^T}{\mathbf{y}_i^T \mathbf{s}_i} \right] \mathbf{H}'_{i-1} \left[\mathbf{I} - \frac{\mathbf{y}_i \mathbf{s}_i^T}{\mathbf{y}_i^T \mathbf{s}_i} \right] + \frac{\mathbf{s}_i \mathbf{s}_i^T}{\mathbf{y}_i^T \mathbf{s}_i}, \tag{1.32}$$

$$\mathbf{s}_i = \mathbf{q}_i^{\mathrm{s}} - \mathbf{q}_{i-1}^{\mathrm{s}}, \tag{1.33}$$

and

$$\mathbf{y}_i = \mathbf{F}\left(\mathbf{q}_i^{\mathrm{s}}\right) - \mathbf{F}\left(\mathbf{q}_{i-1}^{\mathrm{s}}\right). \tag{1.34}$$

Hence, adopting the BFGS method as the numerical scheme of structural optimization procedure, the FEG method is executed through the following successive steps,

(Step 1) Start with the initial structure $\mathbf{q}_k^{\mathrm{s}}$, $k = 0$.

(Step 2) For $\mathbf{q}_k^{\mathrm{s}}$, calculate the free energy change ΔA_k and the forces $\mathbf{F}\left(\mathbf{q}_k^{\mathrm{s}}\right)$ on the FES.

(Step 3) Using the forces $\mathbf{F}\left(\mathbf{q}_k^{\mathrm{s}}\right)$, update the molecular structure by such a difference as

$$\Delta \mathbf{q}_k^{\mathrm{s}} = \alpha_k \mathbf{H}'_k \mathbf{F}\left(\mathbf{q}_k^{\mathrm{s}}\right), \tag{1.35}$$

to find the stationary point. If the force $\mathbf{F}\left(\mathbf{q}_k^{\mathrm{s}}\right)$ is satisfied with the zero-gradient condition (Eq. (1.30)), then, finish the optimization procedure. If not,

(Step 4) Set $\mathbf{q}_{k+1}^{\mathrm{s}} = \mathbf{q}_k^{\mathrm{s}} + \Delta \mathbf{q}_k^{\mathrm{s}}$, $k = k+1$ and return to step 2.

Actually, the L-BFGS algorithm [24] have been often adopted, which is a limited memory variant of BFGS, using the ALGLIB libraries [25].

1.4.2. Vibrational Frequency Analysis via Free Energy Gradient (FEG) method

In the simple approximation, taking thermal motions of the solute and solvent molecules in equilibrium state into consideration, the molecular vibrations can be described as those of harmonic oscillators. The whole energy, E, is expanded as a function of the nuclear coordinates, $\mathbf{R}$, in a Taylor series around the equilibrium structure as follows,

$$E(\mathbf{R}) = E(\mathbf{R}_0) + \frac{dE}{d\mathbf{R}}(\mathbf{R}-\mathbf{R}_0) + \frac{1}{2}\frac{d^2E}{d\mathbf{R}^2}(\mathbf{R}-\mathbf{R}_0)^2 + \frac{1}{6}\frac{d^3E}{d\mathbf{R}^3}(\mathbf{R}-\mathbf{R}_0)^3 + \cdots$$
$$\approx \frac{1}{2}\frac{d^2E}{d\mathbf{R}^2}(\mathbf{R}-\mathbf{R}_0)^2. \quad (1.36)$$

where $\mathbf{R}_0$ are the nuclear coordinates in the equilibrium geometry. The first term may be taken as zero since this is just the total energy of the ground states, and only the third term (the second order deviation) remains as the non-zero one in the harmonic approximation.

In quantum mechanics, the vibrational energy levels ε_n for a one-dimensional harmonic oscillator are obtained,

$$\varepsilon_n = \left(n + \frac{1}{2}\right)h\nu, \quad (1.37a)$$

$$\nu = \frac{1}{2\pi}\sqrt{k_{\mathrm{MW}}}, \quad (1.37b)$$

where n is a vibrational quantum number and ν is the vibrational frequency, which is expressed by using the mass-weighted force constant k_{MW}. The vibrational partition function for an *N*-atomic molecule Q_{vib}, is expressed in general as

$$Q_{\mathrm{vib}} = \prod_{i=1}^{3N-6}\left(\frac{e^{-h\nu_i/2k_BT}}{1-e^{-h\nu_i/k_BT}}\right). \quad (1.38)$$

Given the partition function, the thermodynamic functions such as Helmholtz FE, can

be calculated from the vibrational frequency.

In the FEG method, similarly as the traditional treatment in gas phase, we can approximately estimate the contributions of the vibrational motion to the Helmholtz FE A_{vib}, which are defined by using the effective vibrational frequencies in solution as follows,

$$A_{\mathrm{vib}} = -k_{\mathrm{B}}T \ln Q_{\mathrm{vib}} = \sum_{i=1}^{3N}\left(\frac{1}{2}h\nu_i + k_{\mathrm{B}}T \ln\left(1-e^{-h\nu_i/k_{\mathrm{B}}T}\right)\right) = A^{\mathrm{ZPE}} + A^{\mathrm{HM}}. \tag{1.39}$$

On the right hand side, while the first term is the contribution of the effective zero-point energy (ZPE) correction, the second one is that of the effective harmonic motion (HM) correction.

The effective normal vibrational frequencies are obtained from the Hessian matrix on the FES, which is a $3N \times 3N$ matrix containing all the second derivatives of the free energy with respect to the solute coordinates. For the analytic method, the Hessian can be expressed as follows [7].

$$\mathbf{H}^{\mathrm{FE}} = \frac{\partial^2 A\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial \mathbf{q}^{\mathrm{s}} \partial \mathbf{q}^{\mathrm{s}}} = \left\langle \frac{\partial^2 V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial \mathbf{q}^{\mathrm{s}} \partial \mathbf{q}^{\mathrm{s}}} \right\rangle - \frac{1}{k_{\mathrm{B}}T}\left[\left\langle \frac{\partial V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial \mathbf{q}^{\mathrm{s}}} \frac{\partial V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)^T}{\partial \mathbf{q}^{\mathrm{s}}} \right\rangle - \left\langle \frac{\partial V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial \mathbf{q}^{\mathrm{s}}} \right\rangle \left\langle \frac{\partial V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial \mathbf{q}^{\mathrm{s}}} \right\rangle^T \right], \tag{1.40a}$$

where the superscript T denotes the transposition. To reduce the simulation time, the FE Hessian matrix elements were calculated under the following approximation,

$$\frac{\partial^2 A\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial \mathbf{q}^{\mathrm{s}} \partial \mathbf{q}^{\mathrm{s}}} \approx \left\langle \frac{\partial^2 V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial \mathbf{q}^{\mathrm{s}} \partial \mathbf{q}^{\mathrm{s}}} \right\rangle, \tag{1.40b}$$

since the contribution of the second term in Eq. (1.40a) was considered small enough in comparison with that of the first term [26-28]. Then, by diagonalizing the

mass-weighted FE Hessian matrix, we obtained the effective normal vibrational frequencies of the solute molecule in solution.

1.4.3. Concept of "Dual" Approach to Vibrational Spectra

The vibrational frequency analysis (VFA) in condensed phase, which is based on the FEG methodology, can characterize the averaged behavior of the solute in solution, i.e. the peak vibrational frequencies and representative normal-mode motions in vibrational spectra. The VFA in FEG methodology then requires the mean Hessian matrix in solution, which is obtained by time-averaging along equilibrium QM/MM-MD trajectory. It is the effective Hessian matrix on the free energy surface (FES), called the Free Energy Hessian matrix (FE-Hessian). Moreover, not only these quantities but also the other observable properties (e.g., the band widths and the overlap between peaks) can be discussed by the "dual" approach to vibrational spectra that was developed in this thesis as an extension of the previous approach.

In the dual approach, an assumption is introduced to generate the vibrational spectra in condensed phase. In fact, this approach postulates that these spectra can be reproduced using only a short-time dynamics in condensed phase, meaning that the vibrational spectra are simply reconstructed by the ensemble of vibrational eigenvalues obtained through the instantaneous curvatures of the PES. In other words, the vibrational motions are described in terms of a collection of (harmonic) oscillators; i.e., an instantaneously decoupled second-order description of the vibrational motions of solute molecule along the time-dependent trajectory. The method using such

approximation is called the “instantaneous normal mode (INM)” one [29-34].

By applying the INM approximation, “the molecular motions are governed by a set of the instantaneous normal modes which can interpret rigorously the dynamics at enough short time (in practice, being in the subpicosecond range)” [29]. The difference between the INM prediction and that by the exact time-domain method using the MD simulation becomes clear by comparing the velocity autocorrelation function (VAF) [29, 30], which is connected with the vibrational spectra via Fourier transformation. It was reported that the agreement between those two was quite good for several molecular systems [31, 32], and it would be considered that the INM method can qualitatively reproduce the exact results including anharmonic effects occurring in the short time range. One advantage of the INM method is to provide physical insights that are not available from the usual VAF approach which is based on the averaging over a long-time simulation. Moreover, it can be applied to the calculation of spectra in non-equilibrium systems [33, 34].

Figure 1.2 shows the schematic flowchart of the dual approach that uses two kinds of Hessian matrices. First, we execute the QM/MM-MD sampling to obtain a number of configurations. Next, we analytically calculate the Hessian matrix at each configuration, diagonalizing it. Thus, we can obtain the approximate vibrational spectra by making use of the weighted distribution of instantaneous vibrational frequencies. They are called the instantaneous normal mode (INM) Hessian. The dual method, employing the FE- and the INM-Hessian, enables to obtain vibrational spectra and effective normal modes at once without additional hessian calculation.

The present approach has three merits; (1) We can estimate the Hessian matrix with high efficiency and accuracy by analytically estimating them, (2) obtain the

spectra including explicit solvation effects and (3) investigate the microscopic origin of the spectra. By using complementarily two kinds of Hessian matrices, i.e., FE- and INM Hessian, the dual approach can provide a systematic way for investigating the influence of microscopic solvation to the vibrational motions.

1.5. Outline

In the present thesis, I have investigated the explicit solvent effects on two physical properties of neutral-form glycine in water, (i) the free energy (FE) stability of the conformational isomers and (ii) the shape of vibrational spectra by using the QM/MM-MD method. I have also employed the dielectric continuum model (DCM) method for the present system to extract the microscopic solvation effects.

In chapter 2, I have investigated the FE difference between their conformers. As a result, the relative FEs of conformers in water were found to become smaller than the potential energies (PEs) in the isolated state due to the electrostatic energy compensation of PE destabilization which is introduced with increases in the dipole moment, with only a clear exception of conformer II that has the intramolecular hydrogen bond. The conformer II is stabilized by the non-electrostatic solute-solvent interaction energy (vdW contribution) and their behaviors can be explained by the microscopic solvation effects in water.

In chapter 3, I have proposed a new theoretical methodology for vibrational spectra in solution, the “dual” approach, employing two kinds of analytical Hessians. I have applied it to NF-glycine aqueous solution and investigated the vibrational shift and broadening of spectra induced by the ambient water molecules. From the comparison with the results by experiment and two theoretical methods (i.e. PCM and

QM/MM-MD method), it is concluded that the inclusion of the explicit solvent effect must be essential to understand qualitatively the microscopic origin of the experimental vibrational shifts in solution. Moreover, the microscopic origins of the broadening of spectra were discussed from the point of view of (instantaneous) electron density fluctuation of the functional group.

Finally, it is concluded that the present approaches would provide us a plausible tool to analyze vibrational properties separately to the contributions from the original modes.

In chapter 4, the general conclusion was provided including future perspectives.

References

[1] H. Nakashima, H. Nakatsuji, J. Chem. Phys. 127 (2007) 224104.

[2] H. Nakashima, H. Nakatsuji, J. Chem. Phys. 128 (2008) 154107.

[3] A. Rahman, F. H. Stillinger, J. Chem. Phys. 55 (1971) 169.

[4] R. Car, M. Parrinello, Phys. Rev. Lett. 55 (1985) 2471.

[5] B. Pullman, B. S. Miertus, C. Perahia, Theor. Chim. Acta. 50 (1979) 317.

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

[7] M. J. Field, P. A. Bash, M. Karplus, J. Comp. Chem. 11 (1990) 700.

[8] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, J. Comp. Chem. 25 (2004) 1157.

[9] G. M. Torrie, J. P. Valleau, J. Comp. Phys. 23 (1977) 187.

[10] R. W. Zwanzig, J. Chem. Phys. 22 (1954) 1420.

[11] J. G. Kirkwood, J. Chem. Phys. 3 (1935) 300.

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

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

[14] N. Okuyama-Yoshida, M. Nagaoka, T. Yamabe, Int. J. Quantum Chem. 70 (1998) 95.

[15] N. Okuyama-Yoshida, K. Kataoka, M. Nagaoka, T. Yamabe, J. Chem. Phys. 113 (2000) 3519.

[16] H. Hirao, Y. Nagae, M. Nagaoka, Chem. Phys. Lett. 348 (2001) 350.

[17] Y. Nagae, Y. Oishi, N. Narse, M. Nagaoka, J. Chem. Phys. 119 (2003) 7972.

[18] M. Nagaoka, Y. Nagae, Y. Koyano, Y. Oishi, J. Phys. Chem. A 110 (2006) 4555.

[19] Y. Koyano, N. Takenaka, Y. Nakagawa, M. Nagaoka, Bull. Chem. Soc. Jpn. 83 (2010) 486.

[20] C. G. Broyden, Comp, Jour. 12 (1969) 94.

[21] R. Fletcher, Comp. Jour. 13 (1970) 317.

[22] D. Goldfarb, Math. Comp. 24 (1970) 23.

[23] D. F. Shanno, Math. Comp. 24 (1970) 647.

[24] D. C. Liu, J. Nocedal, Math. Prog. 45 (1989) 503.

[25] S. Bochkanov, Alglib.net - Web Resource. http://www.alglib.net/

[26] N. Takenaka, Y. Kitamura, Y. Koyano, T. Asada, M. Nagaoka, Theor. Chem. Acc. 130 (2011) 215.

[27] Y. Koyano, N. Takenaka, Y. Nakagawa, M. Nagaoka, Bull. Chem. Soc. Jpn. 31 (2010) 2628.

[28] I. F. Galván, M. L. Sánchez, M. E. Martín, F. J. Olivares del Valle, M. A. Aguilar,

J. Chem. Phys. 118 (2003) 255.

[29] M. Buchner, B. M. Ladanyi, R. M. Stratt, J. Chem. Phys. 97 (1992) 8522.

[30] M. Cho, G. R. Fleming, S. Saito, I. Ohmine, R. M. Stratt, J. Chem. Phys. 100 (1994) 6672.

[31] J. T. Kindt, C. A. Schmuttenmaer, J. Chem. Phys. 106 (1997) 4389.

[32] J. Kim, U. W. Schmitt, J. A. Gruetzmacher, G. A. Voth, N. E. Scherer, J. Chem. Phys. 737 (2002) 737.

[33] A. Bastida, M. A. Soler, J. Zúñiga, A. Requena, A. Kalstein, S. Fernández-Alberti, J. Chem. Phys. 132 (2010) 224501.

[34] M. H. Farag, J. Zúñiga, A. Requena, A. Bastida, J. Chem. Phys. 138 (2013) 205102.

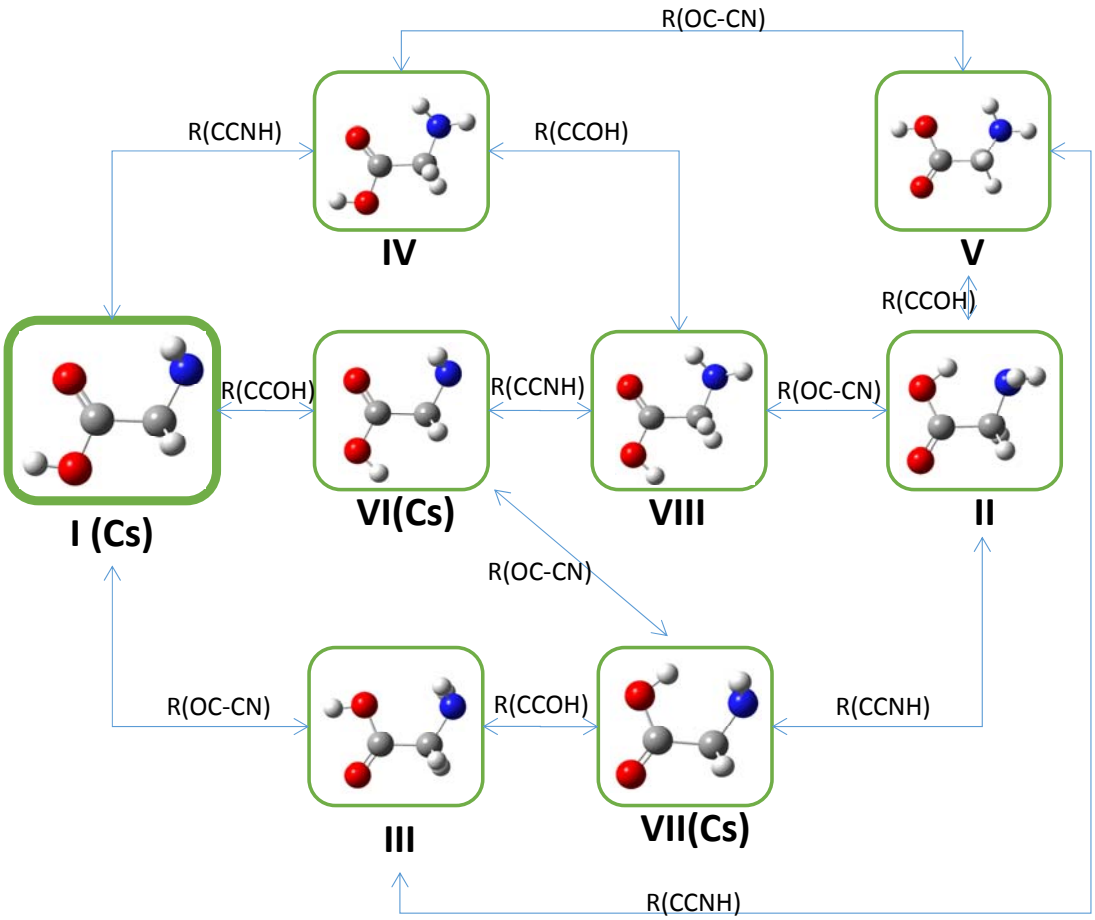


Figure 1.1. Isomerization map of neutral-form glycine.

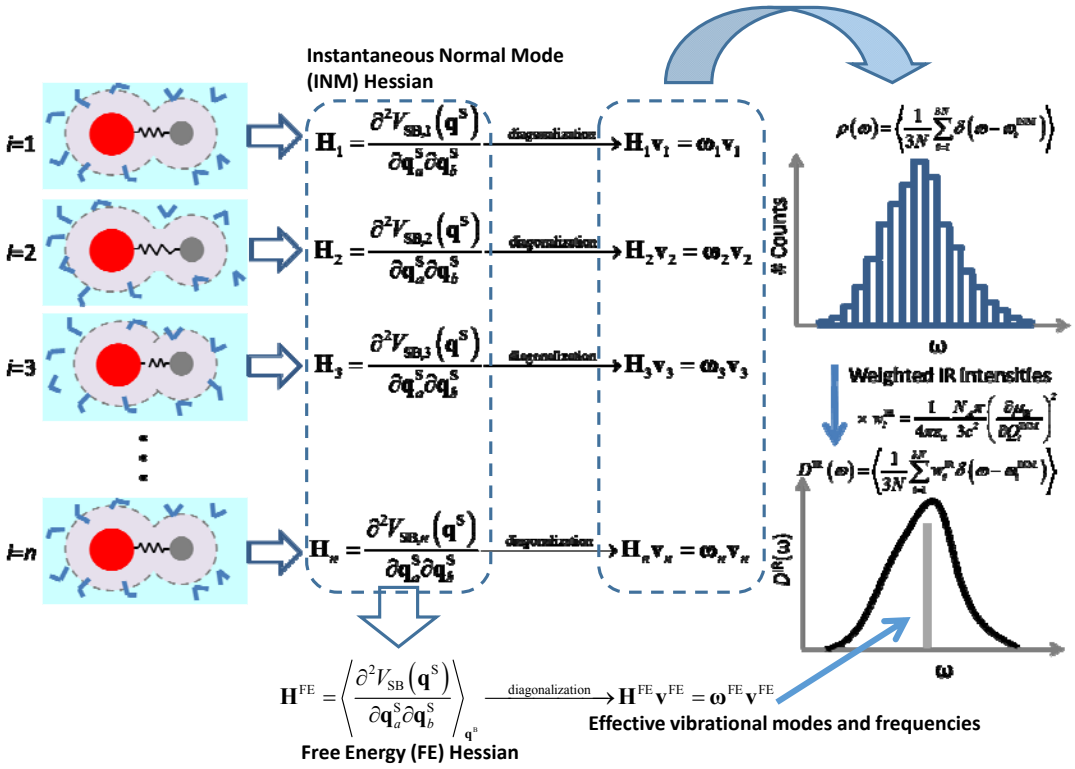


Figure 1.2. Schematic Flow chart of the dual approach.

Chapter 2

On the Smoothing of Free Energy Landscape of Solute Molecules in Solution: A Demonstration of the Stability of Glycine Conformers via Ab Initio QM/MM Free Energy Calculation

2.1. Introduction

Physical properties, reaction selectivity and dynamical behaviors of biological species, such as peptides, proteins and nucleic acids, are closely related to their configurations. Most of their higher-order structures are microscopically determined on the basis of both intra- and intermolecular interactions among those species and solvent water molecules under the thermodynamic environment. To microscopically understand the general role of amino-acid residues in biomolecular interactions, the simplest α-amino acid, glycine, is still a good target of studies since it shows conformational isomerization. The molecular properties and solvation behaviors of glycine molecule, therefore, have so far been investigated by various experiential [1-10] and theoretical [11-37] methods for its pivotal biochemical importance.

In the gas phase, its experimental equilibrium conformer was found to take the *trans* structure (Conformer I in **Figure 2.1**) of the neutral form (NF) [1,2]. In aqueous solutions, on the other hand, it is in its chemical equilibrium state where both NF and

the zwitterionic (ZW) form coexist. However, since ZW becomes much more stable than NF, its equilibrium conformer of NF has never been experimentally identified, although many conformers of NF should exist. It has also been shown that free energetic stability of NF should be considered in understanding the proton transfer process of glycine from ZW to NF [35,36]. Therefore, it should be important to theoretically investigate the free energetic stability of NF glycine in solution.

A number of theoretical methods have so far been developed to take solvent effects into consideration. Among them, the dielectric continuum models (DCMs) are such methods that approximately replace solvent molecules around the solutes by a dielectric continuum [38]. The DCMs are often used when highly accurate QM calculation is required for the solutes with relatively low calculation cost, which is generally increased by only 10 - 40 % in comparison with the QM calculations in an isolated state [17,42]. However, they have such problems as inaccuracy in the direct intermolecular interactions of hydrogen bonds (HBs) between solutes and solvent molecules [39,40]. On the other hand, the quantum mechanical/molecular mechanical (QM/MM) method is often utilized as another method that treats only the reactive parts in the whole solution system quantum mechanically and other parts molecular-mechanically. The molecular dynamics (MD) method combined with the QM/MM method (QM/MM-MD method) [41] can legitimately consider HB interactions more explicitly as a sum of the interatomic interactions between solutes and solvent molecules for instantaneous structural arrangements of those solvent molecules.

I have investigated in this chapter the free energetic stability of the conformational isomers of NF glycine in solution by the ab initio QM/MM-MD

method. I have focused on eight stable NF structures (**Figure 2.1**) obtained in the gas phase and in solution with a DCM model, i.e., the polarizable continuum model (PCM) [38], or more correctly, the conductor-like polarizable continuum model (CPCM) [42]. Then, I have calculated the potential energies (PEs) in the gas phase and free energies (FEs) in solution by both the CPCM and QM/MM-MD methods, relative to the most stable conformer I. In the gas phase, the PEs of NF glycine conformers become less stabilized substantially as the dipole moments increase, while, in aqueous solution, the FEs is stabilized in proportion to the PEs in the gas phase. This is because the stabilization due to the solute-solvent electrostatic interaction should adequately compensate for the PE destabilization, being accompanied by the increase in the dipole moments. Furthermore, the FE landscape of NF glycine conformers by the QM/MM-MD method is found to become rather flat in comparison with that by the CPCM method. The present QM/MM-MD method can account for the microscopic solvation effect, such as the HB structuring among the solute glycine and solvent water molecules.

After briefly introduction of the methodology in Section 2.2, the results and discussion are presented in Section 2.3. The main results are summarized in Section 2.4. The QM/MM-MD method and the thermodynamic integration (TI) method are explained in Chapter 1 and **Figure A1** in Appendix showed the numerical convergence of the free energy calculation.

2.2. Theory and Computational Methods

2.2.1. Solute-Solvent Interaction Energy

In order to elucidate the solvation effect, I have estimated the total solute-solvent interaction energy $E_{\rm int}$ between the solute glycine and the ambient solvent water molecules.

According to the CPCM, I can define $E_{\rm int}$ to be

$$E_{\rm int}^{\rm CPCM}\left(\mathbf{R}_{\rm CPCM}^{N}\right) \equiv \left\{\left\langle \Psi^{\rm (f)} \middle| \hat{H}_{\rm QM}\left(\mathbf{R}_{\rm CPCM}^{N}\right) + \hat{V}^{\rm (f)}\left(\mathbf{R}_{\rm CPCM}^{N}\right) \middle| \Psi^{\rm (f)} \right\rangle - \left\langle \Psi^{\rm gas}\left(\mathbf{R}_{\rm CPCM}^{N}\right) \middle| \hat{H}_{\rm QM}\left(\mathbf{R}_{\rm CPCM}^{N}\right) \middle| \Psi^{\rm gas}\left(\mathbf{R}_{\rm CPCM}^{N}\right) \right\rangle \right\} + \left(E_{\rm dis}\left(\mathbf{R}_{\rm CPCM}^{N}\right) + E_{\rm rep}\left(\mathbf{R}_{\rm CPCM}^{N}\right)\right) \tag{2.1a}$$

$$= \left\{E^{\rm (f)} - E_{\rm gas}^{\rm QM}\left(\mathbf{R}_{\rm CPCM}^{N}\right)\right\} + E_{\rm vdW}^{\rm CPCM}\left(\mathbf{R}_{\rm CPCM}^{N}\right) \tag{2.1b}$$

$$= E_{\rm el}^{\rm CPCM}\left(\mathbf{R}_{\rm CPCM}^{N}\right) + E_{\rm vdW}^{\rm CPCM}\left(\mathbf{R}_{\rm CPCM}^{N}\right), \tag{2.1c}$$

where $E_{\rm gas}^{\rm QM}(\mathbf{R}_{\rm CPCM}^{N})$ and $\Psi^{\rm gas}(\mathbf{R}_{\rm CPCM}^{N})$ are defined as the one-point SCF energy and the wave function of an isolated solute with the molecular structure $\mathbf{R}_{\rm CPCM}^{N}$ obtained by the CPCM;

$$\hat{H}_{\rm QM}\left|\Psi^{\rm gas}(\mathbf{R}_{\rm CPCM}^{N})\right\rangle = E_{\rm gas}^{\rm QM}(\mathbf{R}_{\rm CPCM}^{N})\left|\Psi^{\rm gas}(\mathbf{R}_{\rm CPCM}^{N})\right\rangle, \tag{2.2}$$

and $V^{\rm (f)}\left(\mathbf{R}_{\rm CPCM}^{N}\right)$ is the interaction potential, which is related to the thermal-average distribution function of the solvent molecules, being defined by

$$V^{\rm (f)} = \sum_{A} Z_A \Phi_\sigma(\mathbf{R}_A) - \sum_{s} \Phi_\sigma(\mathbf{r}_s)\,. \tag{2.3}$$

Here $\Phi_\sigma(\mathbf{r})$ is the electrostatic field generated by the polarizable dielectric at the position $\mathbf{r}$ in general, $\mathbf{r}_s$ is the position of the s-th QM electron of the solute molecule and Z_A and $\mathbf{R}_A$ are the core charge and the position of the A-th QM atom,

respectively. The wave functions $\Psi^{(\mathrm{f})}$ and the eigenvalues $E^{(\mathrm{f})}$ are obtained by solving the following associated Schrödinger equation:

$$\left[\hat{H}_{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)+\hat{V}^{(\mathrm{f})}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)\right]\left|\Psi^{(\mathrm{f})}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)\right\rangle = E^{(\mathrm{f})}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)\left|\Psi^{(\mathrm{f})}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)\right\rangle. \tag{2.4a}$$

$E_{\mathrm{el}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ and $E_{\mathrm{vdW}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ are the electrostatic and the non-electrostatic contributions to $E_{\mathrm{int}}^{\mathrm{CPCM}}$, respectively. I have added an index (f) to indicate that the solutions of Eqs. (2.3) and (2.4) are often determined iteratively, and the final converged solution is our ultimate interest. The eigenvalue of the polarized solute,

$$E_{0}^{(\mathrm{f})}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)=\left\langle\Psi^{(\mathrm{f})}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)\right|\hat{H}_{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)\left|\Psi^{(\mathrm{f})}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)\right\rangle, \tag{2.4b}$$

instead of $E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$, is usually adopted in Eq. (2.1) as the reference to estimate the interaction energy. However, as I intend to evaluate the interaction energy accompanying the solvation process from gas to solution, $E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ is used as the reference state. The value of $E_{\mathrm{el}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ in Eq. (2.1c) is estimated by the energy difference between those in the polarizable continuum $E^{(\mathrm{f})}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ and the isolated state ($E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$) for the same molecular structure $\mathbf{R}_{\mathrm{CPCM}}^{N}$, and $E_{\mathrm{vdW}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ is obtained as the sum of the dispersion energy $E_{\mathrm{dis}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ and repulsion energy $E_{\mathrm{rep}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$.

Finally, the FE differences between various conformers, $\Delta A_{\mathrm{sol}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$, results in

$$\Delta A_{\mathrm{sol}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)=E_{\mathrm{int}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)+E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)+E_{\mathrm{cav}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right). \tag{2.5}$$

It should be noted the non-electrostatic terms in the CPCM method, which

are a sum of the cavitation and van der Waals solute-solvent interaction terms, i.e., $E_{\mathrm{cav}}^{\mathrm{CPCM}}$ and $E_{\mathrm{vdW}}^{\mathrm{CPCM}}$, are included using empirical functions (cf. Eq. (1.10)).

The values of $\Delta A_{\mathrm{sol}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ are compared with $\Delta A_{\mathrm{sol}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ relative to that of the most stable Conformer I in **Table 2.1**. For direct comparison with the QM/MM results, the cavitation energy $E_{\mathrm{cav}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$, which is found to be nearly equal as shown in **Table 2.1**, is hereafter omitted from Eq. (2.5).

On the other hand, the E_{int} value (the total solute-solvent interaction energy) can be expressed in the present QM/MM-MD framework (The methodology of the QM/MM-MD method are explained in Chapter 1.2.) by

$$E_{\mathrm{int}}^{\mathrm{QM/MM}}(\mathbf{R}_{\mathrm{CPCM}}^{N}) \equiv \left\langle V_{\mathrm{SB}} \right\rangle_{\mathbf{R}_{\mathrm{CPCM}}^{N}} - E_{\mathrm{gas}}^{\mathrm{QM}}(\mathbf{R}_{\mathrm{CPCM}}^{N}) \tag{2.6a}$$

$$= \left\langle \left\langle \Psi \middle| \hat{H}_{\mathrm{QM}} + \hat{H}_{\mathrm{QM/MM}} \middle| \Psi \right\rangle \right\rangle_{\mathbf{R}_{\mathrm{CPCM}}^{N}} - \left\langle \Psi^{\mathrm{gas}}(\mathbf{R}_{\mathrm{CPCM}}^{N}) \middle| \hat{H}_{\mathrm{QM}}(\mathbf{R}_{\mathrm{CPCM}}^{N}) \middle| \Psi^{\mathrm{gas}}(\mathbf{R}_{\mathrm{CPCM}}^{N}) \right\rangle \tag{2.6b}$$

$$= \left\{ \left\langle \left\langle \Psi \middle| \hat{H}_{\mathrm{QM}} + \hat{H}_{\mathrm{QM/MM}}^{\mathrm{est}} \middle| \Psi \right\rangle \right\rangle_{\mathbf{R}_{\mathrm{CPCM}}^{N}} - \left\langle \Psi^{\mathrm{gas}}(\mathbf{R}_{\mathrm{CPCM}}^{N}) \middle| \hat{H}_{\mathrm{QM}}(\mathbf{R}_{\mathrm{CPCM}}^{N}) \middle| \Psi^{\mathrm{gas}}(\mathbf{R}_{\mathrm{CPCM}}^{N}) \right\rangle \right\}$$
$$+ \left\langle \left\langle \Psi \middle| \hat{H}_{\mathrm{QM/MM}}^{\mathrm{vdW}} \middle| \Psi \right\rangle \right\rangle_{\mathbf{R}_{\mathrm{CPCM}}^{N}} \tag{2.6c}$$

$$= \left\{ E_{\mathrm{sol}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right) - E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right) \right\} + \left\langle \left\langle \Psi \middle| \hat{H}_{\mathrm{QM/MM}}^{\mathrm{vdW}} \middle| \Psi \right\rangle \right\rangle_{\mathbf{R}_{\mathrm{CPCM}}^{N}} \tag{2.6d}$$

$$= E_{\mathrm{el}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right) + E_{\mathrm{vdW}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right), \tag{2.6e}$$

where $\left\langle V_{\mathrm{SB}} \right\rangle_{\mathbf{R}_{\mathrm{CPCM}}^{N}}$ is the solute energy V_{SB} (cf. Eq. (1.17), which can be seen in Chapter 1.2.) averaged over the equilibrium solution structures under the constraint of the solute structure, $\mathbf{R}_{\mathrm{CPCM}}^{N}$, obtained by the CPCM method, and $E_{\mathrm{sol}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ is defined as the time-averaged QM energy in the solvated state with the same molecular

structure $\mathbf{R}_{\mathrm{CPCM}}^{N}$;

$$E_{\mathrm{sol}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)=\left\langle\left\langle\Psi\left|\hat{H}_{\mathrm{QM}}+\hat{H}_{\mathrm{QM/MM}}^{\mathrm{est}}\right|\Psi\right\rangle\right\rangle_{\mathbf{R}_{\mathrm{CPCM}}^{N}}. \tag{2.6f}$$

Namely, the electrostatic contribution $E_{\mathrm{el}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ can be estimated from the time-average of the energy difference between the energy in the solvated state, $\left\langle\left\langle\Psi(\mathbf{R}_{\mathrm{CPCM}}^{N})\left|\hat{H}_{\mathrm{QM/MM}}^{\mathrm{est}}\right|\Psi(\mathbf{R}_{\mathrm{CPCM}}^{N})\right\rangle\right\rangle_{\mathbf{R}_{\mathrm{CPCM}}^{N}}$ (cf. Eq. (1.15) in Chapter 1.2.) and $E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$, minus the vdW contribution $E_{\mathrm{vdW}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$, which is estimated as a time average of the LJ interaction energy, $\left\langle H_{\mathrm{QM/MM}}^{\mathrm{vdW}}\right\rangle_{\mathbf{R}_{\mathrm{CPCM}}^{N}}$.

2.2.2. Computational Details

In order to execute the direct ab initio QM/MM-MD simulation, I used the AMBER-GAUSSIAN interface (AG-IF) [44], by combining the MD simulation program package AMBER 9.0 [45] with the ab initio MO program GAUSSIAN03 [46]. The QM/MM-MD simulations were performed for a whole system including a QM system of one glycine molecule and an MM system of 760 TIP3P water molecules [43] in a rectangular simulation cell with linear dimensions $L_x = 28.56$ Å, $L_y = 27.95$ Å, and $L_z = 28.93$ Å, which had been adjusted initially by executing the classical MD simulation with the *NPT* ensemble at 1 atm. In the classical MD simulation for the effective atomic charges of the solute glycine molecule, I employed the ESP atomic charges obtained by the Merz-Kollman scheme [47,48] at any coordinates. The three-dimensional periodic boundary condition was imposed. The temperature was controlled to 300 K by the Berendsen algorithm so that the system

might be maintained to be an *NVT* ensemble. The nonbonded cutoff distance was chosen as 10.0 Å.

I estimated the FE differences $\Delta A_{\mathrm{sol}}^{\mathrm{QM/MM}}$ among the various conformational isomers of NF glycine in solution by using the Thermodynamic integration (TI) method combined with the ab initio QM/MM-MD method. The detailed methodology of TI method is described in Chapter 1.3. I have interpolated the reactant and product structures along the cooperative coordinate $\mathbf{s}$ that consists of internal coordinates (bond lengths, bond angles and dihedral ones) of the stable state solute structures. Actually, the structures interpolated these NF conformers, optimized by the CPCM method at the MP2/6-31+G(d,p) level of theory, were utilized to be to the interpolated coordinates, s_i's. The number of partitions, N_{p}, was determined so that each change in the potential energy, $\Delta E_i^{\mathrm{CPCM}}$, caused by a change from s_i to s_{i+1} fall within 0.6 kcal mol^{-1} ($\fallingdotseq RT$ at 300K). The average value of N_{p} ranged from 20 to 80, typically 30 in this case. After equilibration, the FE changes were estimated by sampling runs executed for 10 ps with a time step of 0.5 fs at any coordinates. In order to obtain the accurate solute-solvent interaction energy ΔE_{int} between the solute glycine and the solvent water molecules in the solution, the sampling runs were then taken for 50 ps with a time step of 1.0 fs. In these calculations, the structures optimized by the CPCM method were also used in the QM/MM-MD method.

2.3. Results and Discussion

2.3.1. Potential Energies and Dipole Moments of NF Glycine Conformers in the Gas Phase: A Propensity Rule

Table 2.1 shows the PE, $E_{\text{gas}}^{\text{QM}}\left(\mathbf{R}_{\text{gas}}^{N}\right)$, of eight NF glycine conformers (I-VIII) (**Figure 2.1**) in the gas phase, relative to the PE of the most stable conformer I. The nomenclature proposed by Császár for the glycine conformers I-VIII [21] is also shown. The conformers were found classified into two groups according to the magnitude of $E_{\text{gas}}^{\text{QM}}\left(\mathbf{R}_{\text{gas}}^{N}\right)$. The first-group conformers, I-V, are more stable than the second group, VI-VIII, by approximately 3 kcal mol^{-1}. Such a difference in the stability between the two groups can be attributed, in principle, to that in the orientation of the OH group of the carboxy group (-COOH), i.e., as to whether it turns outward or inward, with the exception of II, as discussed below.

More precisely, the hydroxy group (-OH) of glycine in the second group (VI-VIII) turns to the hydrogen atoms of alkyl (>CH_2) or amino (-NH_2) group, making the glycine molecule unstable as a whole because of the repulsive interaction among them. On the contrary, the -OH group in the first conformer group (I-V) is free from the repulsion. Similarly to the earlier theoretical studies [1,3,10], the relative stabilities among the conformers in the gas phase have in fact been explained on the basis of the number of intramolecular hydrogen bonds [1]. For example, Hu et al. [11] discussed the relative stability in relation to the number difference of intramolecular hydrogen bonds (HBs) of each conformer by assuming that each might have several intramolecular HBs (involving the carboxylic acid hydrogen and amino nitrogen,

OH···N, involving the amino hydrogen and carboxylic acid or carbonyl group oxygen, NH···O).

I note further that the relative stability difference exists within the first group, between one subgroup (I or III) and the other (IV or V), except for Conformer II, and that this behavior should be ruled precisely by the relative arrangement of the outward -OH group against the $-NH_2$ group, i.e., *trans* (I and III) or *cis* (IV and V). The former becomes more stable than the latter presumably due to the stronger attractive electrostatic interaction between the oxygen atom (O3) and hydrogen atoms (H9 and H10) in the amino group. According to the electron density analysis (EDA) based on the atoms in molecules (AIM) theory or the natural bond orbital (NBO) theory, however, the stability of Conformer I was explained in terms of hyperconjugation, which enhances the charge transfer (CT) stabilization energy between the occupied orbital n(O4) of the lone pair electrons and the adjacent empty π^*(C1-O3) orbital (i.e., 44.99 kcal mol^{-1} (B3LYP/6-311++G(3d,3p)), while the CT stabilization energy in Conformer IV is as small as i.e., 5.04 kcal mol^{-1} (B3LYP/6-311++G(3d,3p)) [14]. Similarly, it is concluded that only Conformer II forms a clear intramolecular HB, as implied by an additional stabilization term between n(N5) and σ^*(O4-H6), i.e., 10.43 kcal mol^{-1} (B3LYP/6-311++G(3d,3p)) [14]. This is the reason why Conformer II is the second stable in spite of the inward orientation of the OH group of the carbonyl group.

The dipole moments of the eight conformers I-VIII are also shown in **Table 2.1** because of our interest in the influence of the water solvent on the characteristics of conformers in the solution. I note such significant propensity that the PEs are destabilized with the increase in the dipole moment, with only a clear exception of Conformer II, which shows a lower relative energy 1.18 kcal mol^{-1} and a larger dipole

moment 6.32 D (MP2/6-31+G(d,p)).

2.3.2. Free Energetic Stability of NF Glycine Conformers in Solution: The CPCM method

I have also shown in **Table 2.1** the dipole moments $\mu^{(\mathrm{f})}\left(\mathbf{R}^{N}_{\mathrm{CPCM}}\right)$, eigenvalues $E^{(\mathrm{f})}\left(\mathbf{R}^{N}_{\mathrm{CPCM}}\right)$ (cf. Eq. (2.4)), and the FE $\Delta A^{\mathrm{CPCM}}_{\mathrm{sol}}\left(\mathbf{R}^{N}_{\mathrm{CPCM}}\right)$ values (Eq. (2.5)) between the stable state structures within the CPCM method relative to the corresponding values $E^{(\mathrm{f})}\left(\mathbf{R}^{N}_{\mathrm{CPCM}}\right)$ and $\Delta A^{\mathrm{CPCM}}_{\mathrm{sol}}\left(\mathbf{R}^{N}_{\mathrm{CPCM}}\right)$ of Conformer I. In comparison with the eigenvalues $E^{\mathrm{QM}}_{\mathrm{gas}}\left(\mathbf{R}^{N}_{\mathrm{gas}}\right)$ at the equilibrium structures in the gas phase, the corresponding $E^{\mathrm{QM}}_{\mathrm{gas}}\left(\mathbf{R}^{N}_{\mathrm{CPCM}}\right)$ values of all the conformers in the solution, obtained by the CPCM method, are in general enhanced. The dipole moments in the CPCM method, $\mu^{(\mathrm{f})}\left(\mathbf{R}^{N}_{\mathrm{CPCM}}\right)$, also exceed those in the gas phase, $\mu^{\mathrm{gas}}\left(\mathbf{R}^{N}_{\mathrm{gas}}\right)$. It is found in addition that the amount of deviation of the dipole moment from the gas phase to that by the CPCM method, $\mu^{(\mathrm{f})}\left(\mathbf{R}^{N}_{\mathrm{CPCM}}\right)-\mu^{\mathrm{gas}}\left(\mathbf{R}^{N}_{\mathrm{gas}}\right)$, correlates positively with that of the potential energy, i.e., $E^{\mathrm{QM}}_{\mathrm{gas}}\left(\mathbf{R}^{N}_{\mathrm{CPCM}}\right)-E^{\mathrm{QM}}_{\mathrm{gas}}\left(\mathbf{R}^{N}_{\mathrm{gas}}\right)$, leading to the observation that the variations $\Delta A^{\mathrm{CPCM}}_{\mathrm{sol}}\left(\mathbf{R}^{N}_{\mathrm{CPCM}}\right)$ of the eight conformers are relatively smaller than those of $E^{\mathrm{QM}}_{\mathrm{gas}}\left(\mathbf{R}^{N}_{\mathrm{gas}}\right)$, with the exception of Conformer II.

For clarification of this observation due to the solvation effect, I have evaluated the solute-solvent interaction energy $E^{\mathrm{CPCM}}_{\mathrm{int}}\left(\mathbf{R}^{N}_{\mathrm{CPCM}}\right)$ with its two contributions

$E_{\mathrm{el}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ and $E_{\mathrm{vdW}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ as well (**Table 2.1**). In fact, the magnitude of $E_{\mathrm{int}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ is larger in accordance with that of the dipole moment $\mu^{(\mathrm{f})}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$. It was further found by comparison of $E_{\mathrm{el}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ and $E_{\mathrm{vdW}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ that $E_{\mathrm{int}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ consists mainly of the difference in the electrostatic contribution $E_{\mathrm{el}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$. It can thus be understood that the stabilization due to the electrostatic interaction E_{el} with the solvent should sufficiently compensate the potential energy destabilization ($=E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)-E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{gas}}^{N}\right)$) of the solute molecule, accompanied by the increase in the dipole moments, $\mu^{(\mathrm{f})}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)-\mu^{\mathrm{gas}}\left(\mathbf{R}_{\mathrm{gas}}^{N}\right)$. Our present statement is based on the assumption that the cavitation energy $E_{\mathrm{cav}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ is identical for all the NF conformers.

However, there is an exception to the above observation. The $\Delta A_{\mathrm{sol}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ value of Conformer II, 2.80 kcal mol^{-1}, is much larger than the $E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ (=0.69 kcal mol^{-1}). According to the molecular structure of Conformer II, this can be ascribed to the unstable solute-solvent interaction energy because of the existence of the intramolecular HB interaction between N5 and H6.

2.3.3. Free Energetic Stability of NF Glycine Conformers in Solution: The QM/MM Method

The FE values by the QM/MM-MD method, $\Delta A_{\mathrm{sol}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$, are smaller than those by the CPCM method, $\Delta A_{\mathrm{sol}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$. For instance, the corresponding relative solvation FE values of the least stable conformer (VII) are 1.7 and 4.42 kcal mol^{-1}, respectively. This could be attributed to the difference in the HB interactions between the solute glycine molecule and the ambient solvent water molecules. Actually, the solute-solvent interaction energies by the QM/MM-MD method, $E_{\mathrm{int}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$, are in general larger than those by the CPCM method $E_{\mathrm{int}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ (**Table 2.1**). In order to investigate the characteristic difference between the QM/MM-MD and CPCM methods, I calculated the eigenvalues $E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ in the gas phase with the molecular structure of the stable state by the CPCM method, $\mathbf{R}_{\mathrm{CPCM}}^{N}$, as plotted in **Figure 2.2**, and the solute-solvent interaction energies $E_{\mathrm{int}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ as a function of $E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$. In this case, the solute-solvent interaction energies by the CPCM method $E_{\mathrm{int}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ (open squares) are proportional to $E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$. This is a support for that $E_{\mathrm{int}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ depends mainly on only the electrostatic contribution $E_{\mathrm{el}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ (**Table 2.1**).

On the other hand, the $E_{\mathrm{int}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ values based on the QM/MM-MD method (open circles) are not linearly proportional to the $E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ values. In particular, the absolute $E_{\mathrm{int}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ values of unstable conformers (VI-VIII) are found to be larger than those of $E_{\mathrm{int}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$. Such differences in the

solute-solvent interaction energies can be attributed to the difference in the orientation of the OH group in the carbonyl group (-COOH), being similar to the discussion on the relative stability of the conformers in the gas phase. It is recognized that the difference in the orientation of the OH group affects on the HB orientation of the solvent water molecules around the carbonyl group (-COOH) of the solute glycine molecule. In addition, $E_{\mathrm{int}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ is regarded as consisting of both contributions of $E_{\mathrm{el}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ and $E_{\mathrm{vdW}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ (Eqs. (2.1c) and (2.6e)). In Conformer II, for instance, $E_{\mathrm{el}}^{\mathrm{QM/MM}}$ and $E_{\mathrm{vdW}}^{\mathrm{QM/MM}}$ values are 3.21 and −4.85 kcal mol^{-1}, respectively, which are qualitatively different in their characteristics from the $E_{\mathrm{el}}^{\mathrm{CPCM}}$ and $E_{\mathrm{vdW}}^{\mathrm{CPCM}}$ values, 2.19 and 0.05 kcal mol^{-1}, by the CPCM method (**Table 2.1**). This observation can be explained by the microscopic solvation effect in the QM/MM-MD method, which explicitly deals with many solvent water molecules around the solute glycine molecule. Therefore, a more explicit treatment of the ambient solvent water molecules is required for a more precise description of the solute-solvent interactions in solution.

The FE of NF glycine conformers (I–VIII) in solution obtained by the CPCM method and the QM/MM-MD method are summarized in **Figure 2.3**. The relative FE values by the CPCM method, $\Delta A_{\mathrm{sol}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$, are larger than those by the QM/MM-MD method, $\Delta A_{\mathrm{sol}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$. Namely, the $\Delta A_{\mathrm{sol}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ values might be overestimated, similarly to those in the previous investigations by various methods in the DCMs [17-20]. For instance, the FE value of the least stable conformer VII exceeded 4.0 kcal mol^{-1} by using SCIPCM method with B3LYP/6-31+G(d) level [18]. In contract to these previous investigations [17-20], I suggest that the FE

landscape of the NF glycine conformers in solution is close to flat (|Relative FEs| ≤ 1.7 kcal mol^{-1}). Therefore, it is imperative for more accurate discussion of the FE stability to explicitly treat the solvent water molecules around the solute glycine molecule.

2.4. Concluding Remarks

In the present study, I have focused on eight stable conformer structures of NF glycine molecule obtained in the gas phase and in the solution by the CPCM method and calculated the PEs in the gas phase and FEs in the solution obtained by the CPCM and QM/MM-MD methods. The PEs of the NF conformers in the gas phase are substantially destabilized as the dipole moments increase, with only a clear exception of Conformer II, which has the intramolecular HB interaction between the amino group ($-NH_2$) and carboxy (-COOH) groups. In solution, on the other hands, the relative changes in FEs become smaller than that in the PEs in the gas phase; as a results of the stabilization due to the electrostatic interaction with the ambient solvent water molecules, which should adequately compensate the PE destabilization, being accompanied by the increase in the dipole moments. Furthermore, inclusion of the solvent effect explicitly or atomistically by the ab initio QM/MM-MD method causes the NF conformers to be more stabilized free-energetically in comparison with those by the CPCM method. This observation can be explained by the microscopic solvation effect in the QM/MM-MD method, which explicitly deals with solvent water molecules around the solute glycine molecule. As a result, it can be concluded that the present FE landscape of NF glycine conformers in the solutions is found to be rather flat in comparison with that by the DCMs. This propensity should also be confirmed

for several other amino acids.

References

[1] R.D. Suenram, F.J. Lovas, J. Am. Chem. Soc. 102 (1980) 7180.

[2] K. Iijima, K. Tanaka, S. Onuma, J. Mol. Struct. 246 (1991) 257.

[3] I.D. Reva, A.M. Plokhotnichenko, S.G. Stepanian, A.Y. Ivanov, E.D. Radchenko, G.G. Sheina, Y.P. Blagoi, Chem. Phys. Lett. 232 (1995) 141.

[4] K. Machida, A. Kagayama, Y. Saito, Y. Kuroda, T. Uno, Spectrochim. Acta. 33A (1977) 569.

[5] S.A.S. Ghazanfar, D.V. Myers, J.T. Edsall, J. Am. Chem. Soc. 86 (1964) 3439.

[6] M. Sheinblatt, J. Am. Chem. Soc. 92 (1970) 2505.

[7] G. Wada, E. Tamura, M. Okina, M. Nakamura, Bull. Chem. Soc. Jpn. 55 (1982) 3064.

[8] M.A. Slifkin, S. M. Ali, J. Mol. Liq. 28 (1984) 215.

[9] M. Sasaki, Y. Kameda, M. Yaegashi, T. Usuki, Bull. Chem. Soc. 76 (2003) 2293.

[10] M.T. Parsons, Y. Koga, J. Chem. Phys. 123 (2005) 234504.

[11] C.-H. Hu, M. Shen, H.F. Schaefer III, J. Am. Chem. Soc. 115 (1993) 2923.

[12] L.E. Pacios, P.C. Gómez, J. Comp, Chem. 22 (2001) 702.

[13] L.E. Pacios, O. Gálvez, P.C. Gómez, J. Phys. Chem. A. 105 (2001) 5232.

[14] W. Wang, X. Pu, W. Zheng, N.-B. Wong, A. Tian, Chem. Phys. Lett. 370 (2003) 147.

[15] S.M. Bachrach, J. Phys. Chem. A 112 (2008) 3722.

[16] T. Yoshikawa, H. Motegi, A. Kakizaki, T. Takayanagi, M. Shiga, M. Tachikawa, Chem. Phys. 365 (2009) 60.

[17] T.N. Truong, E.V. Stefanovich, J. Chem. Phys. 103 (1995) 3709.

[18] P. Selvarengan, P. Kolandaivel, J. Mol. Struct. (Theochem) 617 (2002) 99.

[19] F.R. Tortonda, J.L. Pascual-Ahuir, E. Silla, I. Tuñón, J. Mol. Struct. (Theochem) 623 (2003) 203.

[20] H.-W. Ke, L. Rao, X. Xu, Y.-J. Yan, J. Theo. Comp. Chem. 7 (2008) 889.

[21] A.G. Császár, J. Am. Chem. Soc. 114 (1992) 9568.

[22] Y. Ding, K. Krogh-Jespersen, J. Comp. Chem. 17 (1996) 338.

[23] F.R. Tortonda, J.L. Pascual-Ahuir, E. Silla, I. Tuñón, Chem. Phys. Lett. 260 (1996) 21.

[24] F.R. Tortonda, J.L. Pascual-Ahuir, E. Silla, I. Tuñón, F. J. Ramírez, J. Chem. Phys. 109 (1998) 592.

[25] E. Kassab, J. Langlit, E. Evleth, Y. Akacem, J. Mol. Struct. (Theochem) 531 (2000) 267.

[26] B. Balta, V. Aviyente, J. Comp. Chem. 25 (2004) 690.

[27] C.M. Aikens, M.S. Gordon, J. Am . Chem. Soc. 128 (2006) 12835.

[28] K. Hong-Wei, R. L. Li, X. Xin, Y. Yi-Jing, Sci. China. Chem. 53 (2010) 383.

[29] M. Nagaoka, N. Okuyama-Yoshida, T. Yamabe, J. Phys. Chem. A. 102 (1998) 8202.

[30] I. Tuñón, E. Silla, C. Millot, M.T.C. Martins-Costa, M.F. Ruiz-López, J. Phys. Chem. A 102 (1998) 8673.

[31] P. Bandyopadhyay, M.S. Gordon, B. Mennucci, J. Tomasi, J. Chem. Phys. 116 (2002) 5023.

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

[33] S. Gusarov, T. Ziegler, A. Kovalenko, J. Phys. Chem. A 110 (2006) 6083.

[34] Z. Lu, Y. Zhang, J. Chem. Theo. Comp. 4 (2008) 1237.

[35] N. Takenaka, Y. Kitamura, Y. Koyano, T. Asada, M. Nagaoka, Theor. Chem. Acc. 130 (2011) 215.

[36] K. Leung, S.B. Rempe, J. Chem. Phys. 122 (2005) 184506.

[37] J. Sun, D. Bousquet, H. Forbert, D. Marx, J. Chem. Phys. 133 (2010) 114508.

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

[39] O. Acevedo, W.L. Jorgensen, Org. Lett. 6 (2004) 2881.

[40] O. Acevedo, W.L. Jorgensen, J. Am. Chem. Soc. 128 (2006) 6141.

[41] M.J. Field, P.A. Bash, M. Karplus, J. Comp. Chem. 11 (1990) 700.

[42] V. Barone, M. Cossi, J. Chem. A. 102 (1998) 1995.

[43] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, J. Chem. Phys. 79 (1983) 926.

[44] T. Okamoto, K. Yamada, Y. Koyano, T. Asada, N. Koga, M. Nagaoka, J. Comp. Chem. 32 (2010) 932.

[45] D.A. Case et al., AMBER 9, University of California, San Francisco, 2006.

[46] M.J. Frisch et al., Gaussian03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.

[47] U.C. Singh, P.A. Kollman, J. Comp. Chem. 5 (1984) 129.

[48] B.H. Besler, K.M. Merz Jr. P.A. Kollman, J. Comp. Chem. 11 (1990) 431.

FIGURES

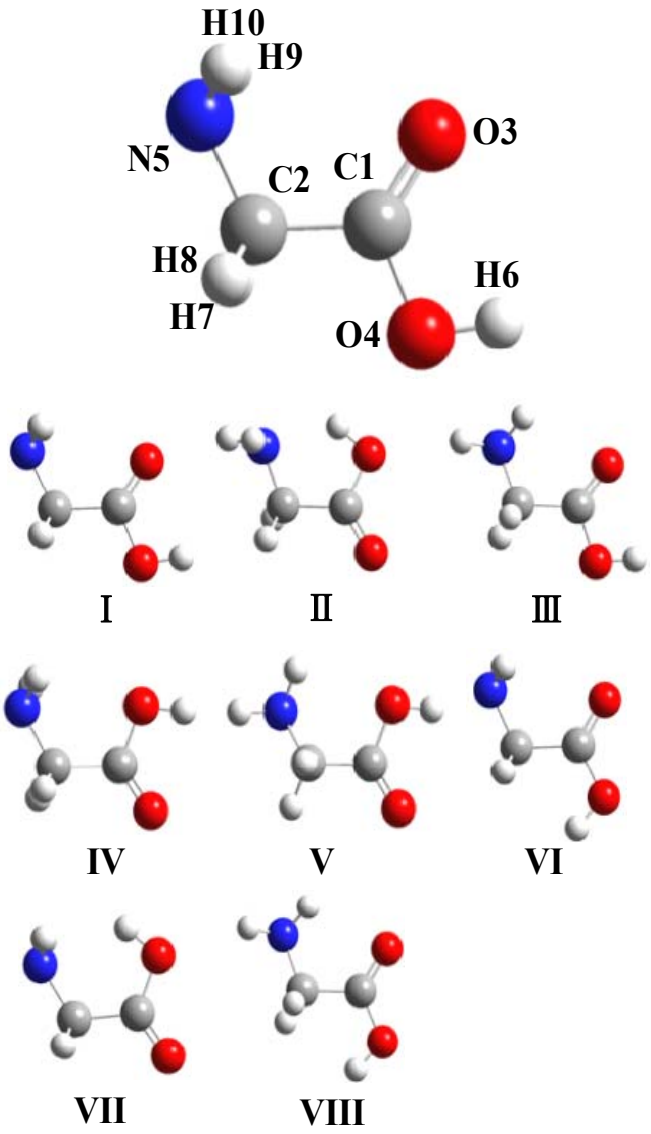


Figure 2.1. Conformers of the neutral form of glycine in gas phase and in solution.

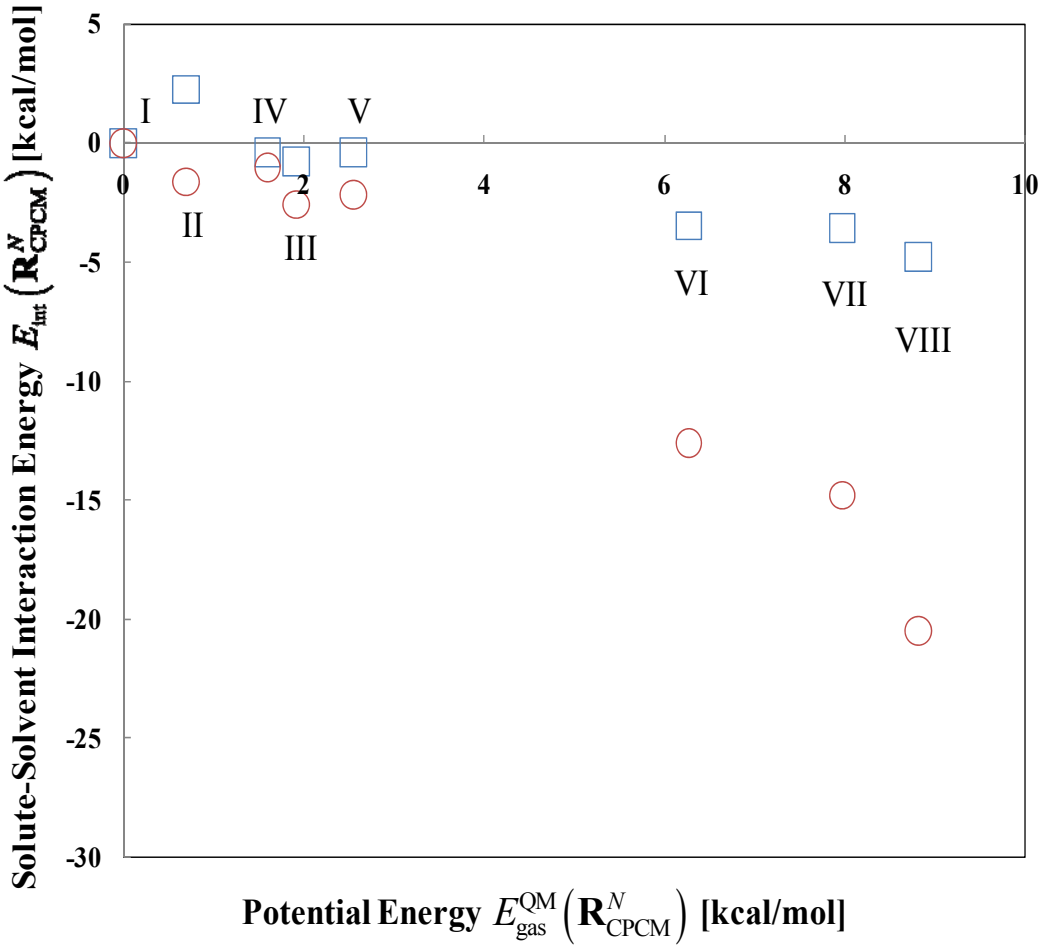


Figure 2.2. Solute-solvent interaction energies $E_{\text{int}}\left(R^N_{\text{CPCM}}\right)$ between the glycine molecule and ambient water molecules of neutral form glycine conformers (I-VIII) as a function of the potential energies $E^{\text{QM}}_{\text{gas}}\left(R^N_{\text{CPCM}}\right)$ in the gas phase (Open square □: CPCM, Open circles ◯: QM/MM-MD). Their geometries were optimized by the CPCM at MP2/6-31+G(d,p) level of theory.

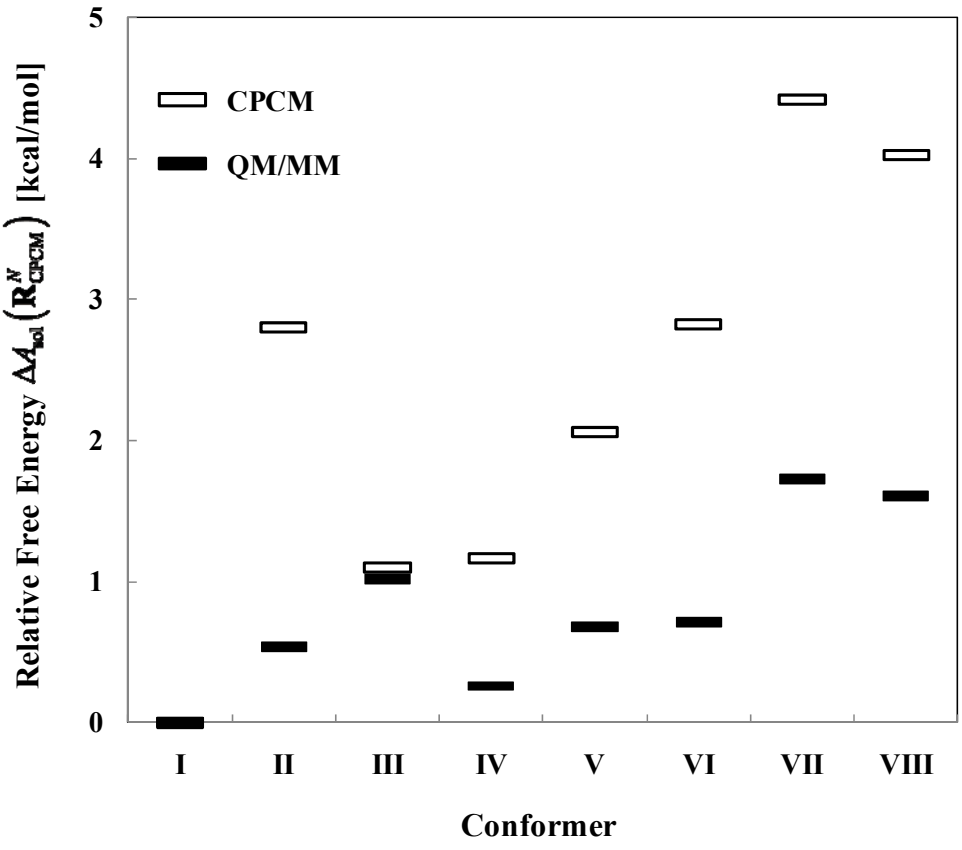


Figure 2.3. Relative free energy $\Delta A_{\mathrm{sol}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)$ of the neutral glycine conformers (I–VIII) in the solutions obtained by the CPCM and QM/MM-MD methods at the MP2/6-31+G(d,p) level of theory.

TABLE

Table 2.1. Relative energies (E and ΔA_{sol}) and dipole moments (μ) of the neutral glycine conformers (I–VIII) in gas phase and in solution obtained by the CPCM method and QM/MM-MD method at the MP2/6-31+G(d,p) level of theory*.

The present nomenclature			I	II	III	IV	V	VI	VII	VIII
The nomenclature proposed by Császár [21]			Ip	IIn	IVn	IIIn	Vn	VIp	VIIp	VIIIn
Gas[i]		$E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{gas}}^{N}\right)^{a}$	0	1.18	1.40	1.56	2.33	5.89	6.80	7.48
		$\mu^{\mathrm{gas}}\left(\mathbf{R}_{\mathrm{gas}}^{N}\right)^{j}$	1.29	6.32	2.47	1.98	3.00	3.61	5.00	5.04
Solution[i]		$E^{(\mathrm{f})}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{b}$	0	2.88	1.18	1.20	2.15	2.74	4.33	4.02
		$E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{a}$	0	0.69	1.92	1.60	2.55	6.27	7.97	8.81
		$\mu^{(\mathrm{f})}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{j}$	2.00	8.24	3.78	3.11	3.92	5.01	6.35	7.61
		$\Delta A_{\mathrm{sol}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{c}$	0	2.8	1.1	1.2	2.1	2.8	4.4	4.0
	CPCM	$E_{\mathrm{int}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{d}$	0	2.24	-0.76	-0.37	-0.41	-3.47	-3.57	-4.76
		$E_{\mathrm{el}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{e}$	0	2.19	-0.74	-0.40	-0.40	-3.53	-3.64	-4.78
		$E_{\mathrm{vdW}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{f}$	0	0.05	-0.02	0.03	-0.01	0.06	0.07	0.02
		$E_{\mathrm{cav}}^{\mathrm{CPCM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{g}$	0	-0.13	-0.05	-0.05	-0.08	0.04	0.03	-0.02
	QM/MM	$E_{\mathrm{sol}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{h}$	0	3.90	-0.52	0.58	0.59	-7.38	-6.70	-12.38
		$E_{\mathrm{gas}}^{\mathrm{QM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{a}$	0	0.69	1.92	1.60	2.55	6.27	7.97	8.81
		$\mu^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{j}$	2.43	8.29	3.83	3.39	3.98	4.94	6.28	7.96
		$\Delta A_{\mathrm{sol}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{c}$	0	0.5	1.0	0.3	0.7	0.7	1.7	1.6
		$E_{\mathrm{int}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{d}$	0	-1.64	-2.58	-1.02	-2.16	-12.6	-14.8	-20.5
		$E_{\mathrm{el}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{e}$	0	3.21	-2.44	-1.02	-1.96	-13.6	-14.7	-21.2
		$E_{\mathrm{vdW}}^{\mathrm{QM/MM}}\left(\mathbf{R}_{\mathrm{CPCM}}^{N}\right)^{f}$	0	-4.85	-0.14	0.00	-0.20	1.01	-0.06	0.72

[a] Gas-phase potential energy. [b] The eigenvalue of the polarized solute. [c] Relative free energy. [d] Solute-solvent interaction energy. [e] The electrostatic interaction of E_{int}. [f] The non-electrostatic interaction of E_{int}. [g] The cavitation energy. [h] The time averages of eigenvalue by QM/MM method. [i] All the values of energies are tabulated relative to those of the most stable conformer I, in kcal/mol. [j] These are in Debye. * Note that the structures optimized in the gas phase, $\mathbf{R}_{\mathrm{gas}}^{N}$, have been used only in the gas phase, while the structures optimized by the CPCM method, $\mathbf{R}_{\mathrm{CPCM}}^{N}$, have been used in both CPCM method and the QM/MM-MD one.

Chapter 3

Dual Approach to Vibrational Spectra in Solution: Microscopic Influence of Hydrogen Bonding to the State of Motion of Glycine in Water

3.1. Introduction

Vibrational spectroscopy is one of the most widely used experimental techniques for determining molecular structures, for the identification and characterization of molecules, and for monitoring the chemical reaction processes [1-3]. In general, because the Infrared (IR) and Raman peaks are often overlapped due to the spectral line broadening, it is difficult to assign peaks in spectra to particular vibrational modes based on such vibrational spectra. On the other hand, from the theoretical viewpoint, all normal modes and their frequencies of a certain molecule can be automatically obtained by the vibrational frequency analysis (VFA). In fact, the VFA has been generally utilized as the reliable tools for the identification of the experimental spectra in gas phase [4, 5]. However, it is still difficult to obtain theoretically such information on the vibrational motions of the chemical species in condensed phase, because of the solvent-induced vibrational coupling by the

solute-solvent interactions and their complicated correlations between the vibrational motions.

Until now, a number of theoretical methods have been developed to take solvent effects into consideration. Among them, the dielectric continuum models (DCMs), which are such methods that replace the solvent molecules around the solute by a dielectric continuum [6], are often used in the cases of requiring highly-accurate QM calculation with relatively low calculation cost. However, they have several problems due to the neglect of direct intermolecular interactions such as the hydrogen bonds (HBs) between solutes and solvent molecules [7-9]. On the contrary, the quantum mechanical/molecular mechanical (QM/MM) method [10] is often utilized as a useful alternative method that only the reactive parts in the whole solution system are treated quantum mechanically while the other part is molecular mechanically. The molecular dynamics (MD) method combined with the QM/MM method (QM/MM-MD method) [11], therefore, can consider legitimately solute-solvent interactions explicitly for instantaneous structural rearrangements of the surrounding solvent molecules.

Under the circumstances, the free energy gradient (FEG) method [12-19] combined with the QM/MM-MD method has been proposed as the structure optimization method on a multidimensional free energy surface (FES) to obtain stable states (SS) and transition states (TS) in condensed phase. In this method, the effective equilibrium structures in solution are obtained as such a molecular structure satisfied with the zero-force condition on the FES, which is analogous to the traditional method on the Born Oppenheimer potential energy surface (PES) using ab initio molecular orbital (MO) calculation. As with the FEG force, since the Hessian on the FES is

similarly obtained, the VFA with FE-Hessian (VFA-FEH) can be executed to obtain both effective normal modes and their vibrational frequencies in solution through the diagonalization of the mass-weighted Hessian matrix on the FES (FE-Hessian). Namely, I can obtain the vibrational frequencies taking into consideration microscopic solute-solvent interaction and thermal fluctuation of the condensed environment by such VFA combined with the QM/MM-MD method. In addition, this method has such an advantage that vibrational frequencies and their fluctuations can be systematically obtained without preparing the empirical vibrational Hamiltonian [20-22] so as to reproduce the vibrational frequencies.

On the other hand, the instantaneous normal mode (INM) analysis [23-26] has been proposed to obtain the approximate vibrational spectra through the diagonalization of the Hessian matrix at each instantaneous configuration of the solute molecule (INM-Hessian). The INM spectra are defined by an ensemble-average of instantaneous $3N$ normal mode eigenvalues over their distribution. By combining it with the QM/MM-MD method, the VFA with the INM-Hessian (VFA-INMH) provides the complementary information that cannot be obtained by VFA-FEH approach. While the latter approach provides a set of effective normal modes and their vibrational frequencies in solution, the former does an instantaneously decoupled second-order description of the vibrational motions of solute molecule along the time-dependent trajectory. Accordingly, the VFA-INMH approach is very useful to clarify the "instantaneous" solvent effects, determining a number of instantaneous vibrational frequencies and vectors.

In this chapter, I have proposed a new theoretical methodology employing both the VFA-FEH and the VFA-INMH approaches within the QM/MM framework. In our

previous study [16, 17], the QM/MM Hessians have been estimated numerically, requiring a lot of calculation times. However, to reduce these calculation costs, they are estimated in the present study by evaluating new analytically-derived expressions of the vdW components in QM/MM Hessian. For a preliminary application using the present dual VFA approach, the vibrational spectra of H_2O molecule in liquid water have been reported [19]. Then, I have explained the dual VFA approach clearly and applied it to the neutral form of glycine molecule in aqueous solution, which is still a good target of studies for its pivotal biochemical importance [9, 27]. Further, I have investigated the microscopic origin of vibrational frequency shifts and vibrational spectra of the glycine molecule in aqueous solution.

In this chapter, I explained the methodology for the VFAs in solution in Section 3.2. In the results and discussion in Section 3.3, I first discussed the computational efficiency and accuracy of the present VFA employing the analytical QM/MM Hessian in Section 3.3.1. Next, in Section 3.3.2, I showed that the calculated vibrational frequency shifts by the VFA-FEH approach are in quite good agreement with the experimental ones. To estimate the importance of the explicit solvent effect, I have also shown comparatively the calculation results by the VFA with the CPCM-Hessian (VFA-CPCM) via the conductor-like polarizable continuum model (CPCM) method. As a result, it was found that the vibrational shifts are attributed to not only the structural relaxation but also the explicit solute-solvent interactions. Finally, by using the VFA-INM approach, I have investigated the approximate IR spectra in aqueous solution considering together the densities of states (DOS) of the vibrational INMs. By comparing the spectroscopic features with the solvation structure around glycine molecule, it was clearly recognized that the IR spectra are

influenced by the hydrogen bond (HB) formation between the solute molecule and the first hydration water molecules. Furthermore, by the natural bond orbital (NBO) analysis, the instantaneous vibrational spectral shift of the hydroxyl group and that of the carbonyl one can be differently explained on the basis of two different effects, i.e., the electron density effect in the bonding MO, and the hyperconjugation effect between the MOs of the lone electron pair and the antibonding MO, respectively. Finally, the concluding remarks were provided in Section 3.4.

3.2. Theory and Computational Methods

3.2.1. Analytical Hessian in the QM/MM Formalism

In the QM/MM treatment of reactive solute molecules in solution, only the reactive parts in the whole solution system are usually treated quantum mechanically, while the other parts are molecular mechanically. The total system potential energy V is obtained by

$$V = \langle \Psi | \hat{H}_{\mathrm{QM}} + \hat{H}_{\mathrm{QM/MM}} | \Psi \rangle + V_{\mathrm{MM}} \tag{3.1a}$$

$$= \langle \Psi | \hat{H}_{\mathrm{QM}} + \hat{H}^{\mathrm{est}}_{\mathrm{QM/MM}} | \Psi \rangle + \hat{H}^{\mathrm{vdW}}_{\mathrm{QM/MM}} + V_{\mathrm{MM}} \tag{3.1b}$$

$$= V_{\mathrm{SB}} + V_{\mathrm{MM}} \tag{3.1c}$$

where $\hat{H}_{\mathrm{QM}}$ stands for the standard Hamiltonian of the QM systems and $\hat{H}_{\mathrm{QM/MM}}$ denotes the interaction between the QM and the MM system. The electrostatic (est) interaction of the polarized QM density with an external charge distribution (e.g., a set

of MM point charges) is estimated by the electrostatic embedding scheme, which is implemented by adding the contribution of the MM point charges to the 1-electron Hamiltonian ($\hat{H}^{\mathrm{est}}_{\mathrm{QM/MM}}$ in Eq. (1.15) in Chapter 1.2.), while the van der Waals (vdW) interaction is typically described by a Lennard–Jones potential ($\hat{H}^{\mathrm{vdW}}_{\mathrm{QM/MM}}$ in Eq. (1.16) in Chapter 1.2.). Here, V_{SB} denotes the sum of both the potential energy of the solute and the interaction energy between the solute and solvent molecules.

The Hessian $\mathbf{H}$ in the QM/MM method, which is a $3N \times 3N$ matrix containing all the second order derivatives of the potential energy V with respect to the mass-weighted (mw) solute's Cartesian coordinates $\mathbf{q}^{\mathrm{s}}_{\mathrm{mw}}$, can be expressed as follows,

$$\mathbf{H} = \frac{\partial^2 V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial \mathbf{q}^{\mathrm{s}}_{\mathrm{mw}} \partial \mathbf{q}^{\mathrm{s}}_{\mathrm{mw}}} = \left[\left\{ \frac{\partial^2 \left\langle \Psi \middle| \hat{H}_{\mathrm{QM}} + \hat{H}^{\mathrm{est}}_{\mathrm{QM/MM}} \middle| \Psi \right\rangle}{\partial q_{\alpha A} \partial q_{\beta B}} \right\} + \left\{ \frac{\partial^2 \hat{H}^{\mathrm{vdW}}_{\mathrm{QM/MM}}}{\partial q_{\alpha A} \partial q_{\beta B}} \right\} \right], \tag{3.2a}$$

with

$$\mathbf{q}^{\mathrm{s}}_{\mathrm{mw}} = \{q_{\alpha A}\} = \left\{ \sqrt{m_1} x_1, \sqrt{m_1} y_1, \sqrt{m_1} z_1, \cdots, \sqrt{m_A} \alpha_A, \cdots, \sqrt{m_N} x_N, \sqrt{m_N} y_N, \sqrt{m_N} z_N \right\}, \tag{3.2b}$$

where subscripts A and B are the indices of the QM atoms (which have a value from 1 to N), and α and β denote the x, y and z components, and m_A is the mass of the A-th QM atom. The Hessian $\mathbf{H}$ consists of the components originating in the est and the vdW contribution. The est components are analytically computed by the program GAUSSIAN [28], where approximate methods of electronic state calculations are available, e.g., the Hartree-Fock (HF), second-order Møller-Plesset perturbation theory (MP2), configuration interaction with singles (CIS), complete active space self-consistent field (CASSCF) method and all the density functional theory (DFT)

methods. On the other hand, since the vdW components are expressed as such analytical functions of those coordinate variables on the same atom (*A*=*B*) as follows,

$$\frac{\partial^2 \hat{H}^{\mathrm{vdW}}_{\mathrm{QM/MM}}}{\partial q_{\alpha A}\partial q_{\beta A}} = \begin{cases} \sum_M \varepsilon_{AM}\left[24\left(7\left(\frac{r_{AM}}{R_{AM}}\right)^{12} - 4\left(\frac{r_{AM}}{R_{AM}}\right)^{6}\right)\cdot\frac{\left(q_{\alpha A}-q_{\alpha M}\right)^2}{R_{AM}^4}\right. \\ \qquad\qquad \left. -12\left(\left(\frac{r_{AM}}{R_{AM}}\right)^{12} - \left(\frac{r_{AM}}{R_{AM}}\right)^{6}\right)\cdot\frac{1}{R_{AM}^2}\right] & (\alpha=\beta) \\ \sum_M \varepsilon_{AM}\left[24\left(7\left(\frac{r_{AM}}{R_{AM}}\right)^{12} - 4\left(\frac{r_{AM}}{R_{AM}}\right)^{6}\right)\cdot\frac{\left(q_{\alpha A}-q_{\alpha M}\right)\left(q_{\beta A}-q_{\beta M}\right)}{R_{AM}^4}\right] & (\alpha\neq\beta) \end{cases} \quad (3.2c)$$

where *M* denotes an integral number discriminating the MM atoms, and ε_{AM} and r_{AM} are a couple of LJ parameters for the *A*-th QM atom interacting with the *M*-th MM atom, and R_{AM} is the distance between the *A*-th QM atom and the *M*-th MM one. In the case of glycine molecule ($N=10$), only 90 components remain as the non-zero elements in the Hessian matrix.

3.2.2. Vibrational Frequency Analysis (VFA) in Solution

3.2.2.1. Effective Normal Modes of Vibration and Vibrational Frequencies Using FE-Hessian: VFA-FEH Analysis

In order to obtain the effective normal modes $\{Q_i\}$ and their vibrational frequencies $\{\omega_i\}$ in solution, I have proposed the VFA using the Hessian on the FES ("FE-Hessian"), i.e., the VFA-FEH analysis. In this analysis, all the molecular vibrations are described under the harmonic approximation on the FES. Then, the FE-Hessian $\mathbf{H}^{\mathrm{FE}}$ can be analytically expressed as follows [12],

$$\mathbf{H}^{\mathrm{FE}}\left(\mathbf{q}^{\mathrm{s}}\right)=\frac{\partial^2 A\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}}=\left\langle\frac{\partial^2 V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}}\right\rangle$$
$$-\frac{1}{k_{\mathrm{B}}T}\left[\left\langle\frac{\partial V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}}\frac{\partial V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)^{T}}{\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}}\right\rangle-\left\langle\frac{\partial V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}}\right\rangle\left\langle\frac{\partial V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}}\right\rangle^{T}\right], \tag{3.3}$$

where the brackets $\langle\cdots\rangle$ denote the time average that is equal to the equilibrium ensemble average and the superscript T denotes the transposition. Since the contribution of the second term of the force fluctuation in Eq. (3.3) was considered small enough in comparison with that of the first term [16-18], the approximation to neglect the second term in Eq. (3.3) was adopted in present study,

$$\frac{\partial^2 A\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}}\approx\left\langle\frac{\partial^2 V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}}\right\rangle, \tag{3.4}$$

thereby FE-Hessian matrix $\mathbf{H}^{\mathrm{FE}}$ is analytically obtained as an equilibrium ensemble average of the instantaneous Hessian matrix $\mathbf{H}$ in solution.

The effective vibrational frequencies $\{\omega_i\}$ are estimated by diagonalizing of the FE-Hessian matrix $\mathbf{H}^{\mathrm{FE}}$,

$$\mathbf{U}^{-1}\mathbf{H}^{\mathrm{FE}}\mathbf{U}=\mathbf{K}\mathbf{I}=\{k_i\}\mathbf{I}, \tag{3.5a}$$

and

$$\omega_i=\frac{1}{2\pi c}\sqrt{k_i}, \tag{3.5b}$$

and

$$\{Q_i\}=\left\{\sum_{A=1}^{N}\sum_{\alpha}^{x,y,z}u_{i,A\alpha}q_{A\alpha}\right\}=\mathbf{U}\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}, \tag{3.5c}$$

where $\mathbf{U}$ and $\mathbf{I}$ are 3*N*-dimensional unitary and identity matrices, respectively. In Eq. (3.5a), $\mathbf{K}$ is a matrix whose diagonal elements consists of a set of the

eigenvalues $\{k_i\}$, and c is the speed of light in Eq. (3.5b). A set of $3N$ effective normal modes of vibration $\{Q_i\}$ can be also obtained from eigenvectors corresponding to those effective vibrational frequencies ω_i.

3.2.2.2 Approximate Vibrational Spectra Using INM-Hessian: VFA-INMH Analysis

In order to elucidate the solvation effects on vibrational properties in aqueous environment, I propose another VFA analysis called the VFA-INMH analysis, i.e., the VFA with the instantaneous Hessian ("INM-Hessian") combined with the QM/MM-MD method, calculating the approximate vibrational spectra via the instantaneous normal mode (INM) analysis [23-26]. The INMs are defined by diagonalizing the INM-Hessian matrix $\mathbf{H}^{\mathrm{INM}}$ at any instantaneous configuration $\mathbf{R}'$, and the eigenvalues $\left\{k_i^{\mathrm{INM}}\right\}$ are obtained as

$$\left\{k_i^{\mathrm{INM}}\right\}_{\mathbf{R}'}\mathbf{I}=\mathbf{U}_{\mathbf{R}'}^{-1}\mathbf{H}^{\mathrm{INM}}\mathbf{U}_{\mathbf{R}'}, \tag{3.6a}$$

with

$$\mathbf{H}^{\mathrm{INM}}=\left.\frac{\partial^2 V_{\mathrm{SB}}\left(\mathbf{q}^{\mathrm{s}}\right)}{\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}\partial\mathbf{q}_{\mathrm{mw}}^{\mathrm{s}}}\right|_{\mathbf{R}'}, \tag{3.6b}$$

and

$$\omega_i^{\mathrm{INM}}=\frac{1}{2\pi c}\sqrt{k_i^{\mathrm{INM}}}. \tag{3.6c}$$

At a finite temperature, most of instantaneous configurations $\mathbf{R}'$ do not correspond exactly to any local minimum on the PES. Consequently, since the forces are usually nonzero and all the eigenvalues of the INM Hessian matrix are not

necessarily non-negative, those normal modes with the negative eigenvalues, are called unstable modes [23, 24]. They are corresponding to such modes with imaginary frequencies, to show the instantaneous motions on an upside-down potential, and are associated with a barrier between two stable wells (e.g., the double wells) or shoulder of a single well [24-26].

To characterize the ensemble-averaged vibrational frequency distribution, the vibrational INM DOS [23-26] are defined as follows,

$$\rho(\omega) = \left\langle \frac{1}{3N - N_{\mathrm{m}}} \sum_{i=1}^{3N-N_{\mathrm{m}}} \delta\left(\omega - \omega_i^{\mathrm{INM}}\right) \right\rangle, \tag{3.7}$$

Here, N_{m} shows the number of normal modes with imaginary frequencies. In addition, to obtain the infrared (IR) spectra, I define the vibrational INM DOS weighted by the dipole moment derivatives $\partial \mu_{\mathbf{R}'} / \partial Q_i^{\mathrm{INM}}$, i.e., with respect to the corresponding INM coordinates $Q_i^{\mathrm{INM}} \left(= \mathbf{U}_{\mathbf{R}'} \mathbf{q}_{\mathrm{mw}}^{\mathrm{s}} \right)$ as follows,

$$D^{\mathrm{IR}}(\omega) = \left\langle \frac{1}{3N - N_{\mathrm{m}}} \sum_{i=1}^{3N-N_{\mathrm{m}}} w_i^{\mathrm{IR}} \delta\left(\omega - \omega_i^{\mathrm{INM}}\right) \right\rangle, \tag{3.8a}$$

with

$$w_i^{\mathrm{IR}} = \frac{1}{4\pi\varepsilon_0} \frac{N_{\mathrm{A}}\pi}{3c^2} \left(\frac{\partial \mu_{\mathbf{R}'}}{\partial Q_i^{\mathrm{INM}}} \right)^2, \tag{3.8b}$$

where ε_0, N_{A} and $\mu_{\mathbf{R}'}$ are vacuum permittivity, Avogadro's number and the net dipole moment at an instantaneous configuration $\mathbf{R}'$, respectively. Actually, I utilize the instantaneous IR intensities calculated with GAUSSIAN09 [28] as w_i^{IR}, and the ensemble-averaged IR intensities I_{aq} in solution are estimated as follows,

$$I_{\mathrm{aq}}(\omega)=\frac{D^{\mathrm{IR}}(\omega)}{\rho(\omega)}=\frac{\left\langle \frac{1}{3N-N_{\mathrm{m}}}\sum_{i=1}^{3N-N_m} w_i^{\mathrm{IR}}\delta\left(\omega-\omega_i^{\mathrm{INM}}\right)\right\rangle}{\left\langle \frac{1}{3N-N_{\mathrm{m}}}\sum_{i=1}^{3N-N_m} \delta\left(\omega-\omega_i^{\mathrm{INM}}\right)\right\rangle}, \tag{3.9}$$

where $\delta\omega$ is the width of a bin for integration, being set at 1 cm^{-1} in this study. I_{aq} means the absorption coefficient for IR spectra, which is independent of temperature, and has an inherent value with each vibrational mode. Then, $D^{\mathrm{IR}}(\omega)$ can be represented by

$$D^{\mathrm{IR}}(\omega)=I_{\mathrm{aq}}(\omega)\rho(\omega), \tag{3.10}$$

and therefore the total spectral line of the approximate IR spectra contributed by two components: I_{aq} and ρ, while I_{aq} has a significant influence on the peak height. On the other hand, the broadening of spectra only depends on ρ which has the temperature dependence.

The vibrational INM DOS is obtained as the ensemble-averaged value of instantaneous normal modes as a function of vibrational frequency, and can provide information on vibrational fluctuations induced by the solvation effects. However, the VFA-INMH analysis cannot necessarily identify any general feature of microscopic vibrational motions in the vibrational spectroscopy because such a normal mode might take an individual different characteristic depending on the instantaneous configuration. On the other hand, in the VFA-FEH analysis, the effective normal modes and their vibrational frequencies at the equilibrium structure in solution are uniquely obtained. Therefore, by using these two kinds of VFAs using FE- and INM-Hessians, I can assign the peaks in the vibrational spectra to the microscopic

vibrational motions, and take some insights into the microscopic solvation effects on the spectroscopic line shape.

3.2.3. Computational Details

The QM/MM-MD simulations were performed for a whole system, which consist of one QM glycine molecule and 759 MM water molecules in a rectangular simulation cell with linear dimensions L_x = 28.56 Å, L_y = 27.95Å, L_z = 28.93 Å. The simulation cell box was adjusted after executing the classical MM-MD simulation with *NPT* ensemble at 1 atm. For estimation of the effective atomic charges of the solute glycine molecule, I employed the ESP atomic charges obtained by Merz-Kollman scheme [29, 30]. Then, the three-dimensional periodic boundary condition was imposed. The temperature was controlled to 300 K by the Berendsen algorithm so that the whole system might be maintained to be an *NVT* ensemble. The nonbonded cutoff distance was chosen as 10.0 Å. The leap-frog algorithm was used to integrate a set of simultaneous equations of motion. After equilibration run, I have obtained the equilibrium structure of the glycine molecule in aqueous solution to be the equilibrium structure optimized by the standard FEG method (the details of the optimization procedure are shown in Chapter 1.4.1.). A set of ten sampling runs for an FEG optimization cycle were executed concurrently each for 50 ps with a time step 1.0 fs. At the optimized structure, for estimating Hessian matrices for the VFAs, I used 50,000 snapshots, which were collected from the total 500 ps MD trajectory with the time interval of 10 fs. To improve the statistics, this 500 ps MD calculation was also executed concurrently with ten independent starting solvent configurations with the optimized glycine structure.

For the purpose to execute the classical MM-MD and *ab initio* QM/MM-MD simulation, I used the AMBER-GAUSSIAN interface (AG-IF) [31], combining the MD simulation program package AMBER 9.0 [32] with the *ab initio* MO program GAUSSIAN03 [33]. In the present study, a solute glycine molecule was taken as a QM part and described at MP2 method with 6-31+G(d,p) basis set. On the other hand, the MM solvent water molecules were taken as the remaining MM part where each water molecule is represented by the rigid TIP3P model [34].

3.3. Results and Discussion

3.3.1. Computational Efficiency and Accuracy of Analytical Hessian Expressions

First, I show the computational efficiency and accuracy of the present analytical Hessian expressions, where the QM/MM Hessians were obtained by estimating analytically-derived formula (Eq.3.2), i.e., the analytical VFA approach. In our previous numerical VFA approach [16, 17], the second derivatives of the energies (the numerical Hessians) were obtained by the central difference method using the analytical first derivatives. To discuss the difference of the computational performance between the analytical and the numerical VFA approaches in solution, I have estimated CPU times and numerical errors of the vibrational frequencies, which were calculated by the sequential ten configurations after equilibration. For the benchmark calculations, I have employed a series of *n*-hydrocarbon C_nH_{2n+2} (n = 1,3,5,7) and the glycine molecule in aqueous solution. Here, I have adopted at MP2/6-31+G(d, p) and TIP3P water model in two systems, and the geometries of *n*-hydrocarbon C_nH_{2n+2} were

obtained with the equilibrium structures in gas phase, while that of the glycine molecule was optimized by the FEG method.

Figure 3.1 shows that the CPU times for *n*-hydrocarbon (C_nH_{2n+2}, n=1,3,5,7) and glycine molecule by the numerical and analytical VFA approaches. As a result, the CPU times by the analytical VFA approach were less than those by the numerical VFA one in any cases, being drastically reduced up to their 15 ~ 20 %. On the other hand, that of glycine molecule by the analytical VFA approach was reduced to about 40 % of the numerical VFA one. This is because I used for glycine more basis functions per atom (cf. $C_2H_5NO_2$, N_{basis}=120), requiring a high computational cost in the analytical approach, as compared with for *n*-hydrocarbon molecules (cf. C_3H_8, N_{basis}=97). In other words, such difference about computational efficiency between two systems, i.e. *n*-hydrocarbons and glycine molecule, should be due to the system-size dependence of the analytical and the numerical VFA approach. Nevertheless, it can be said in general that the present analytical VFA approach is very efficient from the aspect of the computational efficiency in comparison with the numerical VFA approach.

With respect to the numerical accuracy, I have shown in **Table 3.1** the numerical errors of vibrational frequencies in two ways, i.e., the maximum (MAX) absolute value among all the frequency differences

$$\mathrm{MAX} = \max\left\{ \left| \omega_i^{\mathrm{num}} - \omega_i^{\mathrm{ana}} \right| i \in \{1,2,\cdots,3N-6\} \right\}, \tag{3.11a}$$

and their root-mean-square (RMS) value,

$$\mathrm{RMS} = \sqrt{\sum_{i=1}^{3N-6} \left(\omega_i^{\mathrm{num}} - \omega_i^{\mathrm{ana}} \right)^2 \Big/ (3N-6)}\,, \tag{3.11b}$$

using the exact vibrational frequencies obtained by the analytical VFA approach in three different system sizes, i.e., methane, heptane and glycine molecules in solution.

As for methane molecule in solution, the MAX and the RMS value were 18 and 6 cm^{-1}, respectively. On the other hand, in gas phase, their numerical errors (i.e., MAX and RMS) were 0.3 and 0.1 cm^{-1}, respectively. The errors in solution were considerably large as compared with those in gas phase (< 1 cm^{-1}). In previous study [48,49], it was reported that the average deviations from the experimental values by the MP2 method range from +27 to +92 cm^{-1} in small molecules, i.e. H_2N_2, H_2CO and NH_3, and the numerical errors in solution were not negligible when compared with experimental values. In general, with the increase of the system size, the numerical errors became significantly increased. In addition, it is exceptional that the error of glycine molecule was larger than that of heptane molecule in spite of its smaller molecule size. This can be attributed to the presence of the stronger hydration around a glycine molecule, which increases the solute-solvent contribution in Hessian, bringing about the large numerical errors. Namely, the numerical errors would be highly influenced by difference of the surrounding environment, i.e., the microscopic solute-solvent interactions.

3.3.2. Vibrational Frequency by VFA Using FE-Hessian: Importance of Inclusion of Explicit Solvation Effect

Table 3.2 shows a triple of typical vibrational frequencies (ω_{gas}, ω_{CPCM}, ω_{FE}) of glycine molecule in gas phase and in aqueous solution with their vibrational

frequency shifts induced by the solute-solvent interactions ($\Delta\omega_{CPCM}$, $\Delta\omega_{FE}$), which are scaled by the recommended factor of 0.9486 [35]. In addition, they were compared with the experimental values $\Delta\omega_{exp}$ obtained by the Fourier transform infrared (FT-IR) spectroscopy measurement of glycine-$(H_2O)_n$ (n=1,2) [4] and glycine-(1-methyluracil) H-bond complexes [5]. It should be noted that the vibrational frequencies by the FE- and the CPCM-Hessian are calculated at the optimized structures by the FEG and the CPCM method, respectively.

Both $\Delta\omega_{CPCM}$ and $\Delta\omega_{FE}$ for the stretching modes (>1650 cm^{-1}) with the exception of those of CH_2 group were red-shifted, and $\Delta\omega_{FE}$ agreed well with $\Delta\omega_{exp}$ in comparison with $\Delta\omega_{CPCM}$ (see **Table 3.2**). On the other hand, for the four bending modes (≤1650 cm^{-1}), ω_{FE} s showed the large blue-shifts in solution. According to the previous experimental studies [36-45], such tendency is reasonable and common in the intermolecular hydrogen bonding (HB) systems. In particular, ω_{FE} of γ(COH) showed the largest blue-shift (+391 cm^{-1}) in the four and was clearly consistent with $\Delta\omega_{exp}$ (+416 cm^{-1}). In contrast to $\Delta\omega_{FE}$, $\Delta\omega_{CPCM}$ for the bending modes were very small in any cases, showing that the inclusion of explicit effect of HBs is inevitable to reproduce vibrational frequency shifts for the bending modes. It was understood, therefore, that the present VFA-FEH approach can provide the more reasonable estimations of vibrational frequency shifts induced by the solvent molecules in solution.

In order to clarify the microscopic origin of such vibrational frequency shifts $\Delta\omega_{FE}$, I have investigated two essential effects, i.e., the structural relaxation of glycine molecule itself and the explicit solvation structure around it. According to the comparison between the geometries in gas phase and in aqueous solution (see **Table A1** in Appendix), the OH bond length in the –OH group (r(O-H)) was found to increase from 0.973 Å in gas phase to 1.005 Å in aqueous solution (FEG method). Similarly, the NH and the CO bond lengths in the $-NH_2$ and >CO groups (r(N-H) and r(C=O)) increased, although their changes (+0.006 Å and +0.009 Å) were almost negligible in comparison to that of the OH one (+0.032 Å). In relation to the increase of such bond lengths, the stretching-modes were found to show red-shifts in aqueous solution (see **Table 3.2**). On the other hand, while the CH bond lengths of the $>CH_2$ group (r(C–H)) are almost unchanged (-0.001 Å), both its symmetric and asymmetric modes showed slight blue-shifts. In addition, all the bond lengths by FEG method show tendency to increase from those in gas phase, except the CH bond lengths, while these bond lengths by CPCM method also increased (+0.001 Å), their relating stretching modes $\nu_{as}(CH_2)$ and $\nu_s(CH_2)$ showing red-shift, i.e., $\Delta\omega_{CPCM}$ in **Table 3.2**. The changes of CH bond lengths are due to the nonelectrostatic intermolecular interaction with the closest water molecule. It is considered, therefore, that the description of nonelectrostatic interactions by the CPCM method is inadequate for considering these contributions. The $\Delta\omega_{FE}$ for stretching-modes were closely related to the bond length elongation and such tendency also corresponds to the previous experimental results in the various HB systems [36-45]. Thus, it is considered that the mechanism of frequency shifts of the stretching-modes would be attributed to the

softening (or hardening) of vibrational frequency modes due to the bond length elongation (or shortening). However, since there is almost no correlation between $\Delta\omega_{FE}$ of the bending-modes (see **Table 3.2**) and the bond angles (see **Table A1** in the Appendix), the large blue-shifts of ω_{FE} of the bending-modes in aqueous solution cannot be explained by only the changes in bond angles of the glycine molecule.

Next, I have shown in **Figure 3.2** the spatial distributions of oxygen (red) and hydrogen (gray) atoms of solvent water molecules in the first hydration shell around the glycine molecule. It can be clearly seen that the maxima of the oxygen atom distribution are located on the extended line of the O–H and N–H bond of glycine molecule (see **Figure 3.2**). In addition, the hydrogen atom distributions are also high around the area located on the direction extended from the N atom of $–NH_2$ group and O atom of >CO group toward their electron lone-pairs of each atom (see **Figure A2** in Appendix). It was found that the surrounding water molecules clearly form HBs with the –OH, $–NH_2$ and >CO groups. In particular, it was suggested that there are several strong HBs between the –OH group and solvent water molecules. For quantitative estimation of the strength of such HBs, **Figure 3.3** shows the relative frequency distributions of the four kinds of HB angles ($\theta_{OH\cdots Ow}$, $\theta_{CO\cdots Hw}$, $\theta_{N\cdots HOw}$ and $\theta_{NH\cdots Ow}$). The $\theta_{OH\cdots Ow}$ shows the clear maximum value at ~170 deg, with the peak width sharper than those of other three angles. Actually, the time-averaged HB numbers of OH···Ow, CO···Hw, N···HOw and NH···Ow were estimated to be 1.03, 0.41, 0.28 and 0.58, respectively. The differences among the HB strengths of functional groups can be interpreted due to those among the proton affinities of donor-accepters [46], and they are in agreement with the general tendencies in previous studies [47]. It is

understood that the magnitudes of vibrational frequency shifts of stretching modes (ν(CO), $\nu_{as}(NH_2)$, ν(OH)) were found to increase in the same order as the HB numbers (see **Table 3.2**).

In contrast to the present VFA approach, the CPCM method, a representative DCM (dielectric continuum model), cannot include the explicit solute-solvent interactions such as HBs. According to the comparison between $\Delta\omega_{CPCM}$ with $\Delta\omega_{FE}$ for the bending-modes, such large blue shifts of ω_{FE} in aqueous solution can be understood to be brought about by the microscopic solute-solvent interactions. The mechanism must be explicitly explained by the steric hindrance by the HB water molecules (see **Figure 3.2**), which induce the modification of the original PES of the solute and consequently bring about the shifts in the vibrational frequencies. Since the bending motions generally occur oscillatorily orthogonal to the HB direction, the stronger HB formation should make their motions more difficult to occur. As a result, it is considered that the bending-modes including –OH and $–NH_2$ groups show the large blue-shifts in aqueous solution. Moreover, the experimental vibrational shift of Gly-(1-methyluracil) cluster agreed well with our calculated value. This fact supports that their blue-shifts are brought about by a few HB water molecules. From the above comparison with the results by experiment and the CPCM method, it is concluded that the inclusion of explicit solvent effect must be essential to understand even qualitatively the microscopic origin of the experimental vibrational shifts in solution.

3.3.3. Vibrational Spectra by the VFA-INMH Analysis: Strong Correlation between the Explicit HBs and Vibrational Spectra

In order to clarify the influences of the explicit solvent water molecules on the vibrational spectra, I have investigated them in aqueous solution, considering the vibrational INM DOS (see Eq.(3.7)). **Figure 3.4** shows the vibrational INM DOS $\rho(\omega)$ of the glycine molecule in aqueous environment in the high-frequency range (2800–3600 cm^{-1}), where a number of typical stretching modes are observed. Since such peak positions agreed with the ω_{FE} values, each peak was appropriately assigned to the corresponding vibrational mode obtained by the VFA-FEH approach (see **Table 3.2**). It can be reasonably understood, therefore, that the relatively broad peak in the range of 3000 – 3160 cm^{-1} should correspond to the stretching motion of the –OH group (3077 cm^{-1}), where its full width at half-maximum (FWHM) is ~80 cm^{-1}. On the contrary, in the range of 3280 – 3440 cm^{-1}, there were two less broad peaks with each FWHM (~30 cm^{-1}) less than a half the value of –OH group. Actually, it was found that such two peaks, i.e., 3322 and 3416 cm^{-1}, correspond to the symmetric and the asymmetric modes of $–NH_2$ group, respectively. In contrast to the above three modes, the two modes of $>CH_2$ group, i.e., 3028 and 2972 cm^{-1}, were found to become very sharp since they were hardly influenced electrostatically even in aqueous environment.

Next, I have shown in **Figure 3.5** the approximate IR spectra as the vibrational INM DOS $D^{IR}(\omega)$ (Eq.(3.8a)) weighed with w_i^{IR} of each vibrational mode (see in

parentheses in **Table 3.3**). The IR intensities in aqueous solution I_{aq} were found to increase considerably in comparison to those in gas phase I_{gas} with the exception of that of $\nu_{\mathrm{s}}(\mathrm{CH_2})$ (see **Table 3.3**). In particular, since ΔI of the –OH group was the largest among those of all the vibrational modes, the corresponding IR peak showed the maximum in the high-frequency range although the height of $\rho(\omega)$ was the lowest (**Figure 3.4**). Thus, it was recognized in the present system that the intrinsic IR intensity of each vibrational mode has a significant influence on the peak height. In addition, it can be said that, in the present VFA methods, the spectroscopic features, e.g., peak intensity and broadening of IR spectra, were consistently reproduced with the observation by the common IR spectroscopy measurements [36-45].

Finally, I have investigated the microscopic origin of the strong broadening of spectra for the –OH group. It should be noted that, since the geometry of the solute molecule was fixed here at the equilibrium structure in aqueous solution, the peak width is influenced clearly by only the fluctuation of electron density that depends on the instantaneous solvent configurations. In particular, such fluctuation should be induced by the HB formation with the closer water molecules around the glycine molecule. Then, I have investigated the correlation between the vibrational spectra and instantaneous configuration of those water molecules. For the purpose, I have focused on the IR spectra of the C=O and O–H stretching modes. **Figure 3.6** shows the scatter plots of the intermolecular HB distances d_{HB} with the nearest-neighboring O (or H) atoms in the water molecule (the left ordinate) and $D^{\mathrm{IR}}(\omega)$ (the right ordinate), where the abscissa is taken as the vibrational frequency. In this case, I have also

shown the linearly-fitted curves, i.e., $d_{\text{O-H}\cdots\text{O}_\text{w}}$ =0.0031 $\nu(\text{OH})$ −7.9532 and $d_{\text{C=O}\cdots\text{H}_\text{w}}$ = −0.0249 $\nu(\text{CO})$ +43.936. As a result, $\nu(\text{OH})$ was positively correlated with the O–H···O_w distance $d_{\text{O-H}\cdots\text{O}_\text{w}}$, with its correlation coefficient R = 0.901, indicating the strong correlation. This positive correlation indicates that $\nu(\text{OH})$ is red-shifted as the O–H···O_w distance becomes smaller, which is ascribed to the weakening of O–H bond due to the decrease of the electron density of O–H bond induced by the negative point-charge on the O atom of a closer water molecule. On the other hand, the slope of the fitted curve for $\nu(\text{CO})$ became negative, showing relatively weak correlation (R = −0.623). Contrary to an above tendency of the O–H group, $\nu(\text{CO})$ is blue-shifted as the intermolecular HB distance (the C=O···H_w distance $d_{\text{C=O}\cdots\text{H}_\text{w}}$) becomes smaller. The electron densities in the C=O σ and π bonding orbitals increase slightly due to the polarization but these changes are smaller than the change in the O–H bonding orbital. To interpret and explain this vibrational frequency changes occurring upon HB formation, I performed the natural bond orbital (NBO) analysis in three systems, i.e., isolated system and two polarized systems with each point charge (system I and system II). Here, the latter two systems are the simplest HB models polarized with a point-charge expressing the nearest-neighboring atom (O or H). **Table 3.4** shows the hyperconjugative interaction energies $E(2)$ that are estimated by the second order perturbation approach. $E(2)$s in system I scarcely change by the polarization of O–H group. On the contrary, in system II, the LP(O3) → π*(C1-O4) contribution increases due to the polarization of C=O group, giving clear evidence of stabilization originating

in the hyperconjugation between lone electron pair of O3 and C=O π* antibonding orbitals. This interaction should bring about the blue-shift of $\nu(\mathrm{CO})$. In conclusion, by the polarization, $\nu(\mathrm{CO})$ is blue-shifted due to the stabilization by the hyperconjugative interaction, while $\nu(\mathrm{OH})$ is red-shifted by the destabilization of the corresponding orbitals.

As a result, the vibrational frequencies were found to clearly correlate with the HB distance d_{HB}. Considering that the HB number of –OH group increased in comparison to that of >CO group (see **Figure 3.2**), the broadening of IR spectra can be related to the HB strength.

3.4. Conclusion Remarks

In this study, I have proposed a new theoretical methodology of vibrational spectra in solution, employing two VFAs (vibrational frequency analyses) using analytical Hessians on the FES (FE-Hessian) and instantaneous ones (INM-Hessian), i.e., VFA-FEH and VFA-INMH. First, I compared the computational performance between analytical and numerical VFA approaches, and confirmed that the present analytical VFA approach is more efficient than the conventional numerical VFA approach. Second, by using the VFA-FEH approach, I investigated the vibrational frequencies and their fluctuations of glycine molecule in aqueous solution under the explicit environment consisting of ambient solvent water molecules. As a result, the calculated vibrational frequencies were quantitatively in good agreement with the experimental ones. For the stretching-mode, the vibrational frequencies of –OH, $–NH_2$ and >CO groups were red-shifted in aqueous solution, while those of $>CH_2$ group

were slightly blue-shifted. This is mainly due to the softening (or hardening) of vibrational motions influenced by the structural relaxation in solution, i.e., the bond length elongating (or shortening). On the other hand, the bending-modes including –OH and $–NH_2$ groups showed the large blue shifts, being consistent to the experimental observations. This is explained by the hindrance of motion by the adjacent solvent water molecules. Namely, such vibrational shifts are attributed to not only the structural relaxation but also the explicit atomistic solute-solvent interactions.

Third, by using the VFA-INMH approach, I have investigated the approximate IR spectra in solution through the vibrational INM densities of states (DOS). Since such peak positions agreed with the effective vibrational frequencies by the VFA-FEH approach, each peak was appropriately assigned to the corresponding vibrational mode. In particular, since the IR intensity for the –OH group was largest in the high-frequency range, the peak height considerably increased. Namely, such peak heights in the IR spectra must be strongly influenced by the intrinsic IR intensity. Moreover, according to the relationships between the solvent frequency fluctuation and the HB (hydrogen bond) distance, the broadening of IR spectra could be related to the HB strength. In particular, the instantaneous vibrational changes of the –OH group and >CO one can be explained on the basis of two different mechanisms, i.e., the change of electron density in OH orbital and hyperconjugation between the lone electron pair of oxygen atom of >CO group and the C=O π^* antibonding orbital, respectively. This is considered an essential clue to understand the general feature of experimental vibrational spectra in the various HB systems.

In the present study, I have discussed about the IR spectra depending on the electron density induced by the instantaneous solvent configurations, by combining

with VFA-FEH approach. However, more realistically, the present method should be executed successively for a set of “instantaneous” structures obtained through an equilibrium (or non-equilibrium) QM/MM-MD trajectory without fixing the structures of solute molecules themselves. Moreover, by using the VFA-INMH approach, any complicated vibrational spectra, which are overlapped by the several vibrational modes, can be, in principle, separated to the contributions of individual vibrational modes. Finally, it is concluded that the present approaches would provide us a plausible tool to analyze vibrational properties separately to the contributions from the original modes and are quite useful to interpret the microscopic origin of the experimental vibrational spectra.

References

[1] B. Yan, J. B. Fell, G. J. Kumaravel, J. Org. Chem. 61 (1996) 7467.

[2] E. T. J. Nibbering, H. Fidder, E. Pines, Annu. Rev. Phys. Chem. 56 (2005) 337.

[3] D. Stead, G. Carbone, P. O'Brien, K. R. Campos, I. Coldham, A. Sanderson, J. Am. Chem. Soc. 132 (2010) 7260.

[4] R. Ramaekers, J. Pajak, B. Lambie, G. Maes, J. Chem. Phys. 120 (2004) 4182.

[5] B. Boeckx, G. Maes, J. Phys. Chem. B 116 (2012) 11890.

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

[7] O. Acevedo, W. L. Jorgensen, Org. Lett. 6 (2004) 2881.

[8] O. Acevedo, W. L. Jorgensen, J. Am. Chem. Soc. 128 (2006) 6141.

[9] Y. Kitamura, N. Takenaka, Y. Koyano, M. Nagaoka, Chem. Phys. Lett. 514 (2011) 261.

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

[11] M. J. Field, P. A. Bash, M. Karplus, J. Comp. Chem. 11 (1990) 700.

[12] N. Okuyama-Yoshida, M. Nagaoka, T. Yamabe, Int. J. Quant. Chem. 70 (1998) 95.

[13] M. Nagaoka, N. Okuyama-Yoshida, T. Yamabe, J. Phys. Chem. A 102 (1998) 8202.

[14] N. Okuyama-Yoshida, M. Nagaoka, T. Yamabe, J. Phys. Chem. A 102 (1998) 285.

[15] N. Okuyama-Yoshida, K. Kataoka, M. Nagaoka, T. Yamabe, J. Chem. Phys. 113 (2000) 3519.

[16] Y. Koyano, N. Takenaka, Y. Nakagawa, M. Nagaoka, Bull. Chem. Soc. Jpn. 83 (2010) 486.

[17] N. Takenaka, Y. Kitamura, Y. Koyano, T. Asada, M. Nagaoka, Theor. Chem. Acc. 130 (2011) 215.

[18] I. F. Galván, M. L. Sánchez, M. E. Martín, F. J. Olivares del Valle, M. A. Aguilar, J. Chem. Phys. 118 (2003) 255.

[19] H. C. Georg, T. S. Fernandes, S. Canuto, N. Takenaka, Y. Kitamura, M. Nagaoka, In *Practical Aspects of Computational Chemistry III*; Leszczynski, J., Shukla, M. K., Eds.; Springer; New York, USA, 2014; pp 231-247.

[20] D. W. Oxtoby, D. Levesque, J. –J. Weis, J. Chem. Phys. 68 (1978) 5528.

[21] C. P. Lawrence, J. L. Skinner, J. Chem. Phys. 117 (2002) 8847.

[22] C. P. Lawrence, J. L. Skinner, J. Chem. Phys. 118 (2003) 264.

[23] M. Buchner, B. M. Ladanyi, R. M. Stratt, J. Chem. Phys. 97 (1992) 8522.

[24] M. Cho, G. R. Fleming, S. Saito, I. Ohmine, R. M. Stratt, J. Chem. Phys. 100 (1994) 6672.

[25] J. T. Kindt, C. A. Schmuttenmaer, J. Chem. Phys. 106 (1997) 4389.

[26] A. Melzer, A. Schella, J. Schablinski, D. Block, A. Piel, Phys. Rev. E 87 (2013) 033107.

[27] I. Tuñón, E. Silla, M. F. Ruiz-López, Chem. Phys. Lett. 321 (2000) 433.

[28] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels,

O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09, Revision A.02; Gaussian Inc.: Wallingford, CT, 2009.

[29] U. C. Singh, P. A. Kollman, J. Comp. Chem. 5 (1984) 129.

[30] B. H. Besler, K. M. Merz Jr., P. A. Kollman, J. Comp. Chem. 11 (1990) 431.

[31] T. Okamoto, K. Yamada, Y. Koyano, T. Asada, N. Koga, M. Nagaoka, J. Comp. Chem. 32 (2010) 932.

[32] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz Jr., D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. E. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. E. A. F. Schafmeister, W. S. Ross, P. A. Kollman, AMBER 9; University of California: San Francisco, 2006.

[33] 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, J. A. Pople, Gaussian 03, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

[34] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, J. Chem. Phys. 79 (1983) 926.

[35] J. P. Merrick, D. Moran, L. Radom, J. Phys. Chem. A 111 (2007) 11683.

[36] R. M. Badger, S. H. Bauer, J. Chem. Phys. 5 (1937) 839.

[37] R. M. Badger, J. Chem. Phys. 8 (1940) 288.

[38] C. M. Huggins, G. C. Pimentel, J. Phys. Chem. 60 (1956) 1615.

[39] A. Allerhand, P. von R. Schleyer, J. Am. Chem. Soc. 85 (1963) 371.

[40] R. S. Drago, N. O'Bryan, G. C. Vogel, J. Am. Chem. Soc. 92 (1970) 3924.

[41] K. Mizuno, Y. Miyashita, Y. Shindo, H. Ogawa, J. Phys. Chem. 99 (1995) 3225.

[42] M. Freda, G. Onori, A. Santucci, J. Phys. Chem. B 105 (2001) 12714.

[43] S. Tanabe, T. Ebata, M. Fujii, N. Mikami, Chem. Phys. Lett. 215 (1993) 347.

[44] J. Kubelka, T. A. Keiderling, J. Phys. Chem. A 105 (2001) 10922.

[45] T. L. Jansen, J. Knoester, J. Chem. Phys. 124 (2006) 044502.

[46] S. Grabowski, J. Chem. Rev. 111 (2011) 2597.

[47] K. Leung, S. B. Rempe, J. Chem. Phys. 122 (2005) 184506.

[48] N. C. Handy, R. D. Amos, J. F. Gaw, J. E. Rice, E. D. Simandiras, Chem. Phys. Lett. 120 (1985) 151.

[49] R. J. Harrison, G. B. Fitzgerald, W. D. Laidig, R. J. Bartlett, Chem. Phys. Lett. 124 (1986) 291.

FIGURES

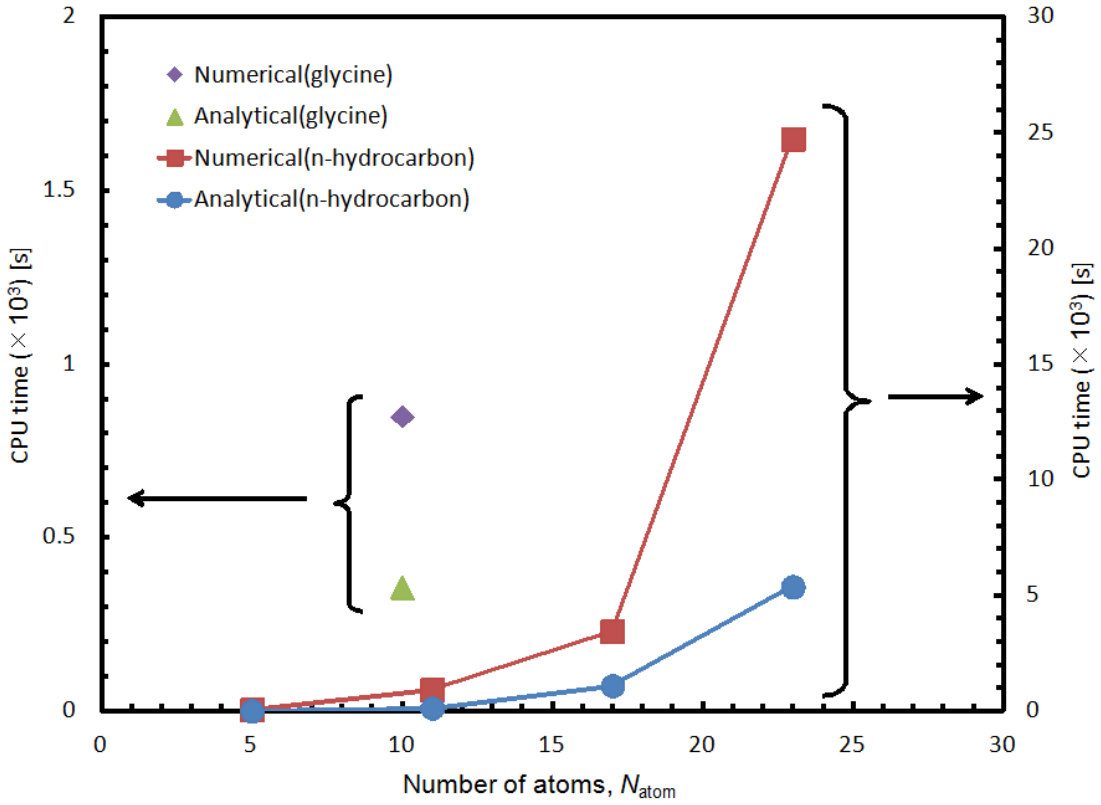
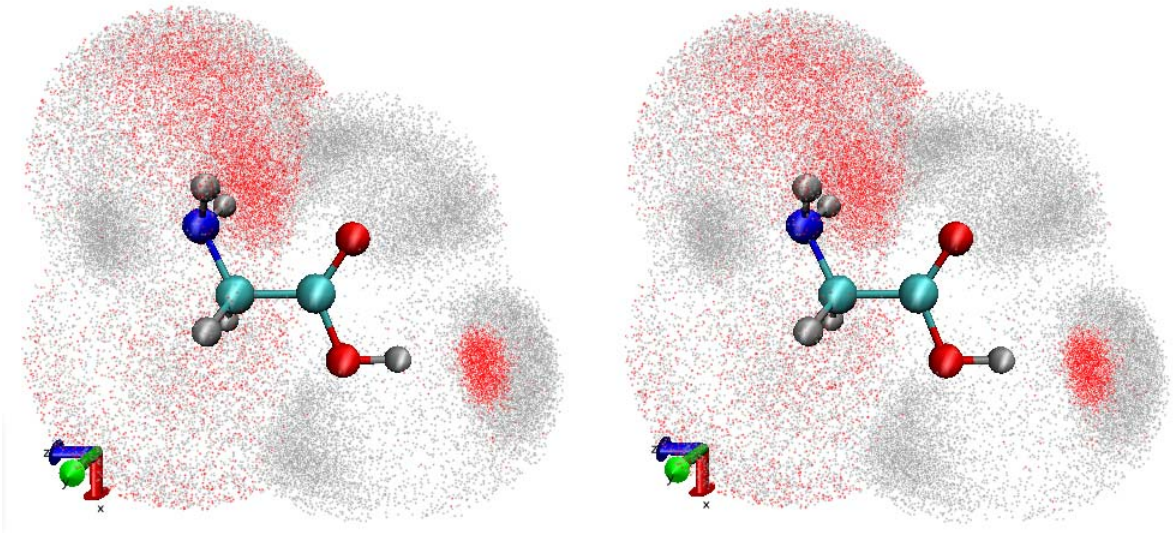


Figure 3.1. CPU times for n-hydrocarbons (C_nH_{2n+2}, n=1,3,5,7) and glycine molecule. It was obtained by averaging the calculation time for ten configurations. All the calculations were carried out with an Intel Xeon E3-1270 (3.4GHz) machine. Note that the right and left ordinate axes represent the CPU times for n-hydrogens and glycine molecule, respectively.

(a) Side View



(b) Top View

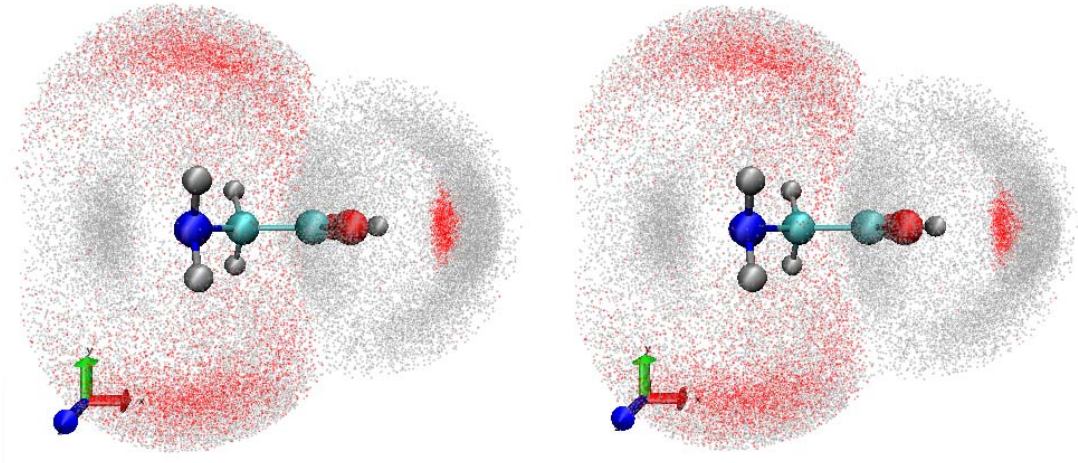


Figure 3.2. Stereo view of the spatial distribution of oxygen and hydrogen atoms of a water molecule in the first hydration shell around the solute glycine molecule, (a) Side view and (b) Top view. The red points indicate the oxygen atom of water molecule, and the grey one indicate the hydrogen atom.

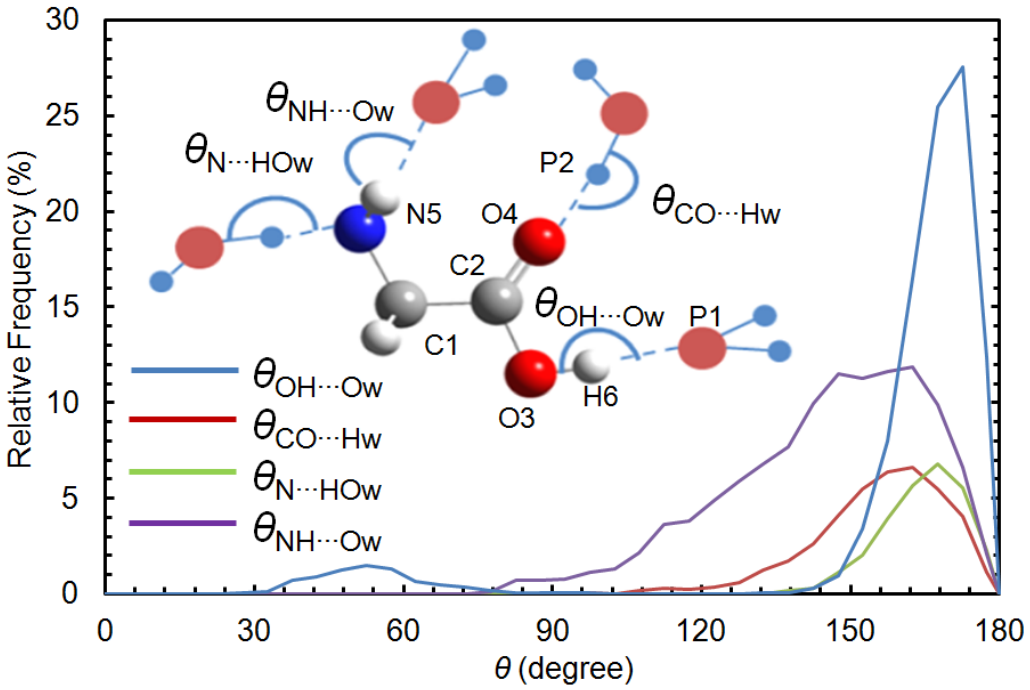


Figure 3.3. Relative frequency distribution of the hydrogen bond angles.

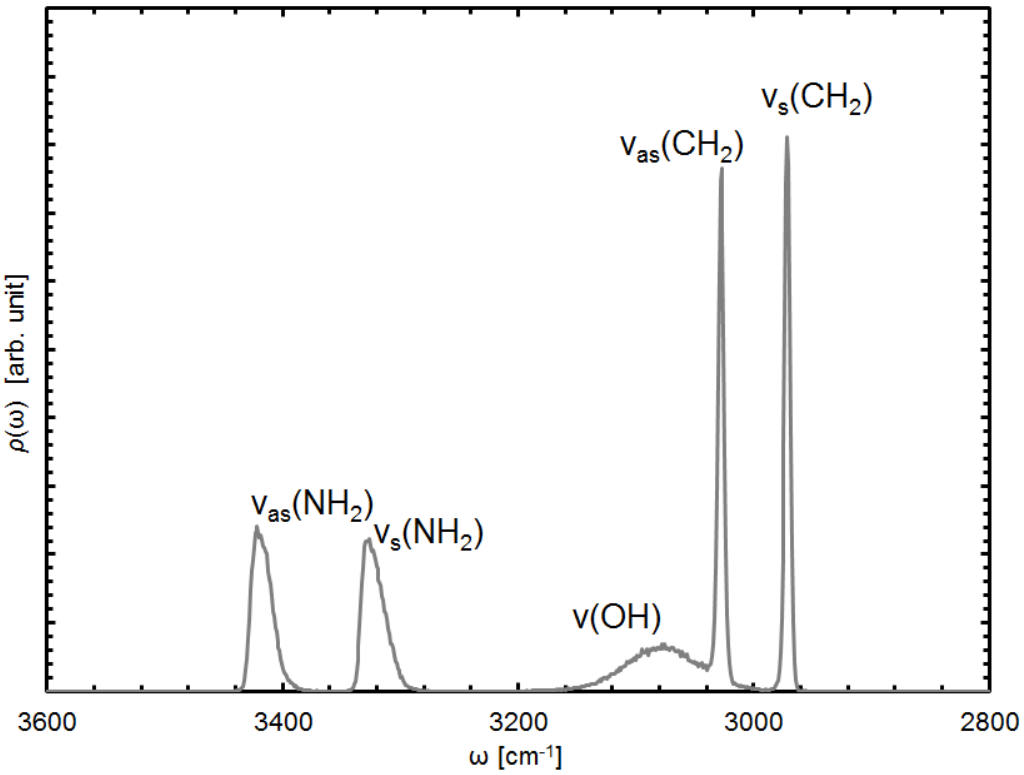


Figure 3.4. The vibrational INM DOS $\rho(\omega)$ in high-frequency range.

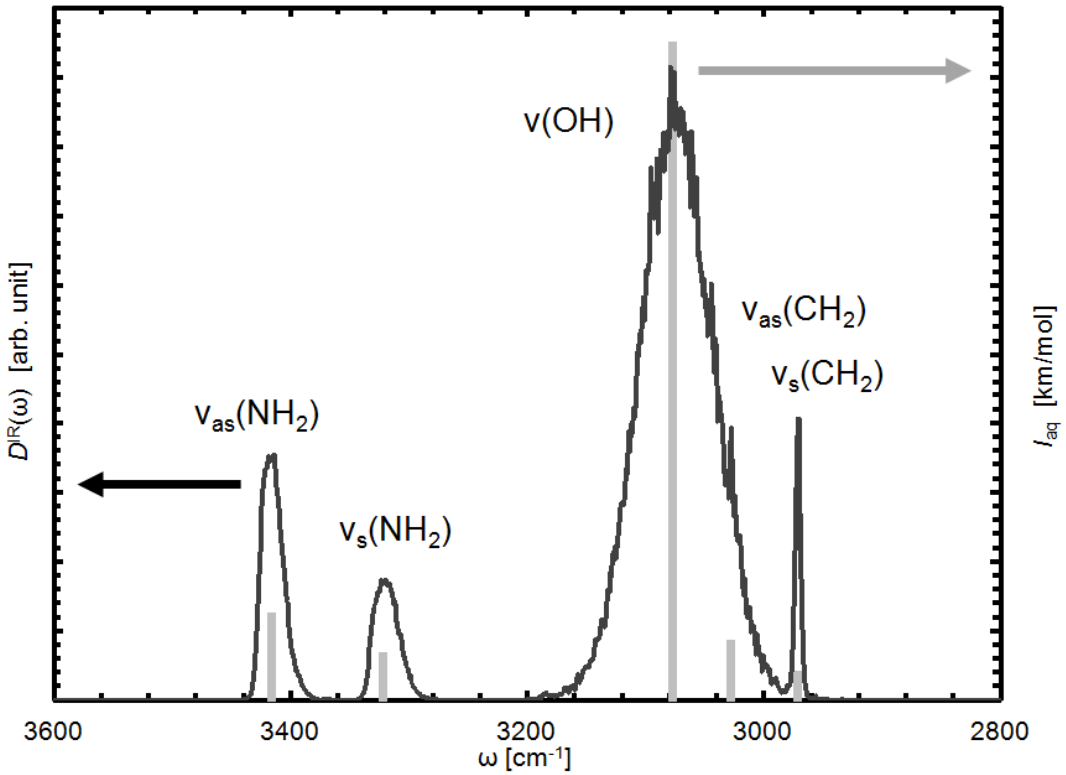
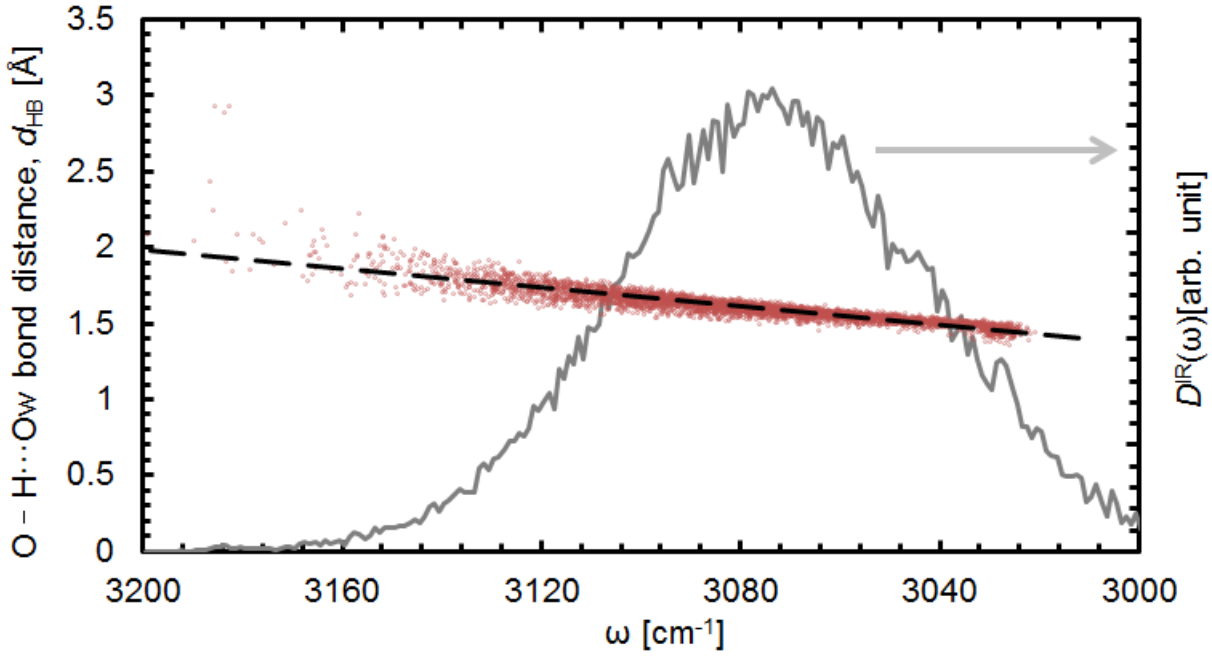


Figure 3.5. The approximate IR spectra: the vibrational INM DOS $D^{\mathrm{IR}}(\omega)$ weighed with w_i^{IR} in high-frequency range. Those average IR intensities I_{aq} at the vibrational frequencies estimated by the VFA-FEH approach are also indicated by the gray bar.

a) O–H stretching mode



b) C=O stretching mode

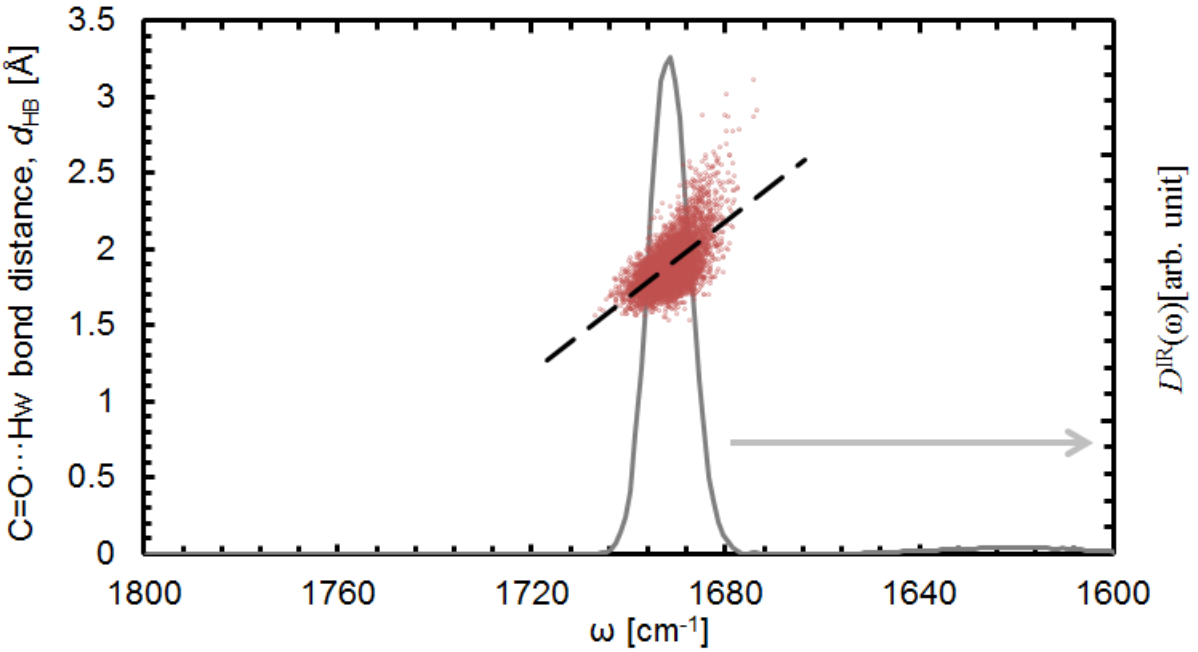


Figure 3.6. IR spectra of O–H and C=O stretching modes. Scatter plots of the lengths of nearest-neighboring H or O atom of water molecules versus the O–H and C=O stretching frequencies. Linear fits (dashed lines) are (a) $d_{\mathrm{O\text{-}H\cdots O_w}} = 0.0031\,\nu(\mathrm{OH}) - 7.9532$; R = 0.901 and (b) $d_{\mathrm{C{=}O\cdots H_w}} = -0.0249\,\nu(\mathrm{CO}) + 43.936$; $R = -0.623$.

TABLES

Table 3.1. Numerical errors of vibrational frequencies [cm^{-1}] (the maximum (MAX) absolute value among all the frequency differences and their root mean of square (RMS) value) in three different system sizes, i.e., methane, heptane and glycine in solution. The system sizes are denoted by the number of atoms and the degree of basis sets.

component	Numerical error [cm^{-1}]		System size	
	MAX[a]	RMS[b]	N_{atom}	N_{basis}
Methane	18	8	5	39
Glycine	122	29	10	120
Heptane	109	22	23	213

[a] $\mathrm{MAX} = \max\left\{\left|\omega_i^{\mathrm{num}} - \omega_i^{\mathrm{ana}}\right| \; i \in \{1,2,\ldots,3N-6\}\right\}$

[b] $\mathrm{RMS} = \sqrt{\sum_{i=1}^{3N-6}\left(\omega_i^{\mathrm{num}} - \omega_i^{\mathrm{ana}}\right)^2 \Big/ (3N-6)}$

Table 3.2. Calculated vibrational frequencies[a] ω (cm^{-1}) and their shifts from in gas phase to in aqueous solution $\Delta\omega$ (cm^{-1}).

Mode	Gas	In water		Frequency shift		
	ω_{gas}	ω_{CPCM}	ω_{FE}	$\Delta\omega_{CPCM}$ [b]	$\Delta\omega_{FE}$ [c]	$\Delta\omega_{exp}$
Stretching						
$\nu_{as}(NH_2)$	3476	3351	3416	-125	-60	-13[d]
$\nu_{s}(NH_2)$	3381	3262	3322	-119	-59	
ν(OH)	3570	3047	3077	-523	-493	-320[d]
$\nu_{as}(CH_2)$	3011	2990	3028	-21	+17	
$\nu_{s}(CH_2)$	2958	2936	2972	-22	+14	
ν(CO)	1708	1657	1691	-51	-17	-39[d]
Bending						
δ(COH)	1341	1232	1338	-9	+97	
δ(COH)+ $\gamma_s(NH_2)$	886	897	990	+11	+104	
γ(COH)	606	563	997	-43	+391	+416[e]
$\gamma_{as}(NH_2)$	225	221	505	-4	+280	

[a] Normal vibrational frequencies (in cm^{-1}) scaled by 0.9418 (Ref. 35).

[b] $\Delta\omega_{CPCM} = \omega_{CPCM} - \omega_{gas}$

[c] $\Delta\omega_{FE} = \omega_{FE} - \omega_{gas}$

[d] FTIR data for Gly-(H_2O) and Gly-(H_2O)$_2$ cluster in Ar matrix [4].

[e] FTIR data for Gly-(1-methyluracil) cluster in Ar matrix [5].

Table 3.3. Calculated IR Intensities I (km/mol) and shifts from in gas phase to in aqueous solution ΔI (km/mol).

Mode	i	Gas phase	In water	Intensity shift
		$I_{\text{gas}}\left(\omega_{\text{gas},i}\right)$	$I_{\text{aq}}\left(\omega_{\text{FE},i}\right)$	ΔI^{*}
$\nu_{as}(NH_2)$	1	9.9 (3476)	41.7 (3416)	+31.8
$\nu_{s}(NH_2)$	2	2.9 (3381)	21.9 (3322)	+19.0
ν(OH)	3	73.9 (3570)	331.0 (3077)	+257.1
$\nu_{as}(CH_2)$	4	4.6 (3011)	28.1 (3028)	+23.5
$\nu_{s}(CH_2)$	5	13.3 (2958)	12.4 (2972)	-0.9
ν(CO)	6	260.6 (1708)	390.3 (1691)	+129.7

$^{*}\ \Delta I = I_{\text{aq}} - I_{\text{gas}}$

Table 3.4. NBO analysis: The hyperconjugative interaction energies estimated by the second order perturbation approach *E*(2) (kcal/mol), and their differences between the isolated system and two polarized systems with each point charge (System I and System II). (The names of atoms in the parentheses and point charges follow those in **Figure 3.3**).

Type[a]	Isolated	System I[b]		System II[c]	
Donor/Acceptor	*E*(2)	*E*(2)	Δ	*E*(2)	Δ
LP_1(O3)/BD_1*(C1-O4)	11.12	11.12	0.00	11.14	0.02
LP_2(O3)/BD_2*(C1-O4)	65.5	65.49	-0.01	65.72	0.22
LP_1(O4)/RY_1*(C1)	21.82	21.82	0.00	21.82	0.00
LP_2(O4)/BD_1*(C1-C2)	21.93	21.93	0.00	21.93	0.00
LP_2(O4)/BD_1*(C1-O3)	40.46	40.46	0.00	40.39	-0.07
LP_1(N5)/BD_1*(C1-C2)	12.17	12.17	0.00	12.13	-0.04

Chapter 4

General Conclusion

In this thesis, to understand the explicit solvation effects, I have investigated the (i) free energy (FE) stability and (ii) vibrational spectra of the neutral-form (NF) glycine in aqueous solution by employing two theoretical methods, i.e. the dielectric continuum model (DCM) method and the quantum mechanical/molecular mechanical (QM/MM) method.

In chapter 2, the FE stability of NF-glycine conformers and the contribution of microscopic solvation effects in those systems were investigated. In the CPCM method, the electrostatic contribution of solute-solvent interaction energy was underestimated due to the lack of explicit hydrogen bonds among solute and solvent molecules. On the other hand, in Conformer II that has one intramolecular hydrogen bond (HB), the vdW contribution was overestimated, because the CPCM method cannot consider the correct contributions by the relatively weak solvation of Conformer II. Consequently, the relative FE differences among conformers in aqueous solution were smaller than the PEs in the isolated conformers.

In chapter 3, I have proposed a new theoretical methodology for vibrational spectra in solution, employing two kinds of analytical Hessians. I have applied it to glycine molecule and discussed the microscopic solvation effects on the vibrational frequency shifts and broadening of the vibrational spectra in solution. From comparing with the experimental results those by two kinds of theoretical methods (the CPCM

and the QM/MM-MD method), it was shown that the inclusion of explicit solvent effect must be essential to understand even qualitatively the microscopic origin of the experimental vibrational shifts in solution, especially those of bending modes. Moreover, to clarify the microscopic origin of the broadening of each peak corresponding to the vibrational modes, I have investigated the correlation between the instantaneous vibrational shifts and the configuration of the closer water molecules around the glycine molecule. In proportion to the intermolecular HB distance, it was found that the stretching mode of hydroxyl group shows the larger red-shift while one of carbonyl group does blue-shift. Their behaviors were discussed on the basis of the electron density analysis, and suggested that they originated in the different mechanism.

The development of electronic computers has allowed a systematic development of computational chemistry and wide applications of those theoretical techniques to among chemical problems. The art of developing and applying computer programs for solving them has brought about great breakthroughs in the field of chemistry and its environs. Moreover, theoretical chemistry has large overlaps with (theoretical and experimental) condensed matter physics and molecular physics, and this research field must develop more and more in future while interrelating closely with each other. So, I would be happy if my research finding (both scientific suggestion and the developed method) in the present thesis could make a meaningful contribution to the science and its society.

Chapter 5

Appendix

5.1. Convergence of Free Energy Difference

In order to know the simulation time necessary to obtain the statistical accuracy of numerical values of the free energy difference, I did the preliminary survey and it was found that the free energy change for Δs_1, $\Delta A_{\mathrm{sol},\,i=1}^{\mathrm{QM/MM}}$ converges until 7.5 ps (15,000 steps) at the time step 0.5 fs (**Figure A1**).

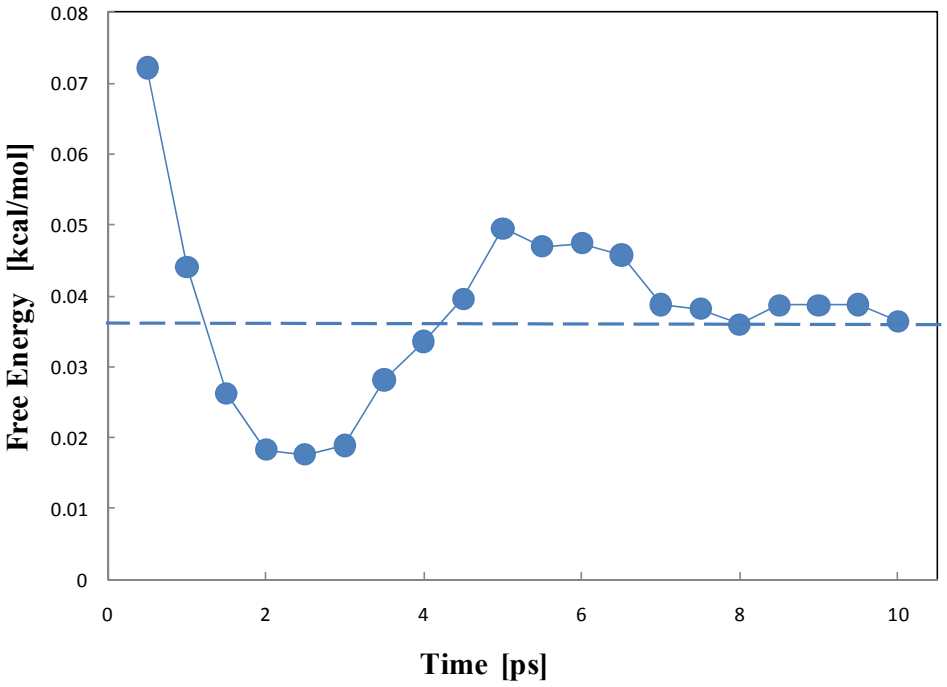


Figure A1. The convergence of the free energy difference $\Delta A_{\mathrm{sol},\,i=1}^{\mathrm{QM/MM}}$ (cf. Eq. A8) at an interpolated structure i=1 between Conformer I (i=0) and IV (i=80). The abscissa represents the simulation time for the statistical averaging.

5.2. Structural Change in Aqueous Solution

For the purpose to obtain the equilibrium structures in aqueous solution, I have optimized the molecular structure of the glycine molecule by the FEG method. **Table A1** shows the internal coordinates (bond lengths and bond angles) of the equilibrium structures in gas phase and in aqueous solution. The internal coordinates were arranged in order of those in gas phase and in aqueous solution (CPCM method and FEG method), and were written in the bold letters if there is relatively a large difference between in gas phase and in aqueous solution.

Table A1. Equilibrium bond lengths (Å), bond angles (degree) and the differences between those in gas phase and in aqueous solution.

Parameter	Gas phase	Solution		Δ	
		CPCM	FEG	CPCM	FEG
Bond length					
r(C=O)	1.221	1.223	1.230	+0.002	+0.009
r(C-H)	1.091	1.092	1.090	+0.001	-0.001
r(O-H)	0.973	0.997	1.005	**+0.024**	**+0.032**
r(N-H)	1.014	1.020	1.020	+0.006	+0.006
Bond Angle					
θ(COH)	106.7	108.5	110.4	+1.8	+3.7
θ(CNH)	110.3	110.1	110.3	-0.2	0.0

5.3. Spatial Distribution of Water Molecules: Cross-Sectional View

To help the understanding the solvation structure around glycine molecule, the spatial distribution of oxygen (red) and hydrogen (gray) atoms of water molecules is

shown in **Figure A2**, displayed in the cross-sectional view of x-y plane. The distribution of the oxygen atoms was found to clearly locate around the hydrogen atom of –OH group, while those of the hydrogen atoms can be observed around the nitrogen atom of $–NH_2$ group and the oxygen atom of >CO group.

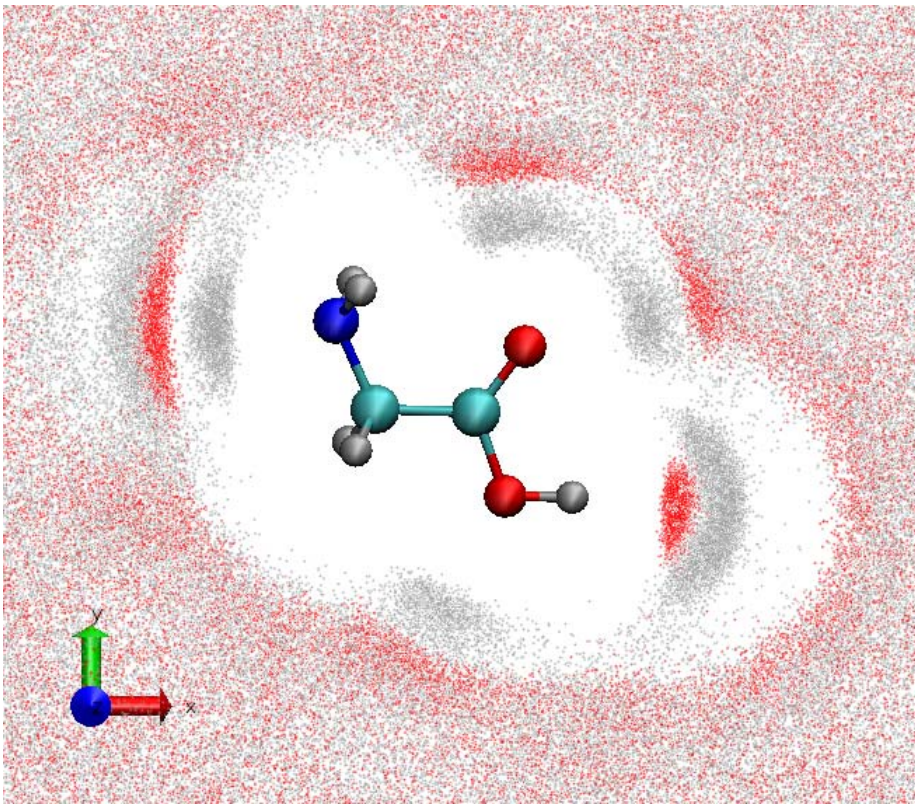


Figure A2. Spatial distribution of oxygen (red) and hydrogen (gray) atoms of water molecules around glycine molecule in the cross-sectional view of x-y plane

5.4. Convergence of the Effective Frequencies

To confirm the convergence of the effective vibrational frequencies in solution, I have shown the time variation of the mean value $\omega_{\mathrm{FE}}(t)$ and the standard error of the mean (SEM) $\sigma_{\mathrm{M}}(t)$ in the glycine molecule. The SEM is estimated:

$$\sigma_{\mathrm{M}}(t) = \frac{s(t)}{\sqrt{N_{\mathrm{traj}}}} \quad , \tag{5.1}$$

where N_{traj} is a total number of trajectories executed with independent initial solvent configurations, and s is the standard deviation (SD) defined by

$$s^2(t) = \frac{1}{N_{\mathrm{traj}}} \sum_{I=1}^{N_{\mathrm{traj}}} \left(\omega^I(t) - \langle \omega(t) \rangle_{N_{\mathrm{traj}}} \right)^2 . \tag{5.2}$$

Here, it should be noted that $\omega^I(t)$ means the effective vibrational frequency averaged over a time step t in the Ith MD simulation, and $\langle \omega(t) \rangle_{N_{\mathrm{traj}}}$ is the mean value averaged over N_{traj} number of trajectories:

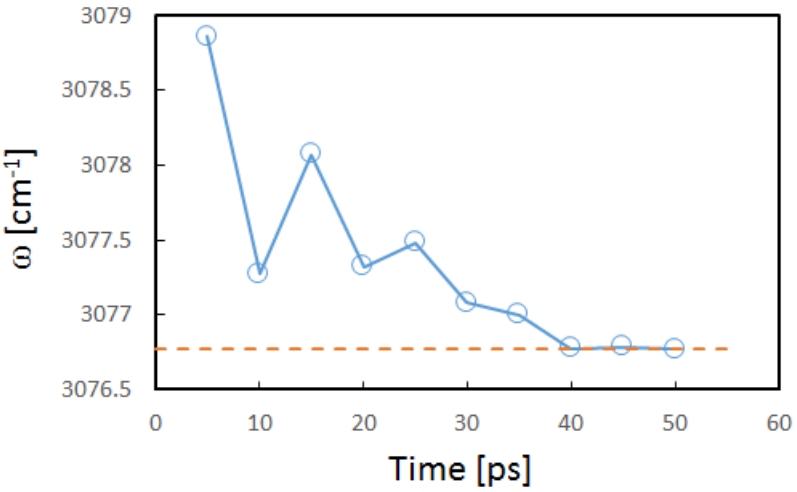
$$\langle \omega(t) \rangle_{N_{\mathrm{traj}}} = \frac{1}{N_{\mathrm{traj}}} \sum_{I=1}^{N_{\mathrm{traj}}} \omega^I(t). \tag{5.3}$$

In the present study, 10 sampling MD runs for VFA analysis were used ($N_{\mathrm{traj}} = 10$).

Figure A3 and **Figure A4** show the temporal changes in $\omega_{\mathrm{FE}}(t)$ and $\sigma_{\mathrm{M}}(t)$ of the glycine molecule, respectively. Until reaching the convergence, $\omega_{\mathrm{FE}}(t)$ of O-H stretching mode showed quite large fluctuations, while that of C=O stretching mode changes to a small extent. The tendencies of other stretching modes were similar to that of C=O stretching mode. In any case, it have been confirmed that $\omega_{\mathrm{FE}}(t)$ converges sufficiently enough by 50 [ps]. Moreover, in **Table A2**, it was found that

the converged values of σ_M take a value within the range of $0.08 \leq \sigma_M \leq 0.66$ [cm^{-1}], showing that our calculation results have sufficient statistical precision to discuss the vibrational shifts by the solvation effects.

(a) O-H stretching mode



(b) C=O stretching mode

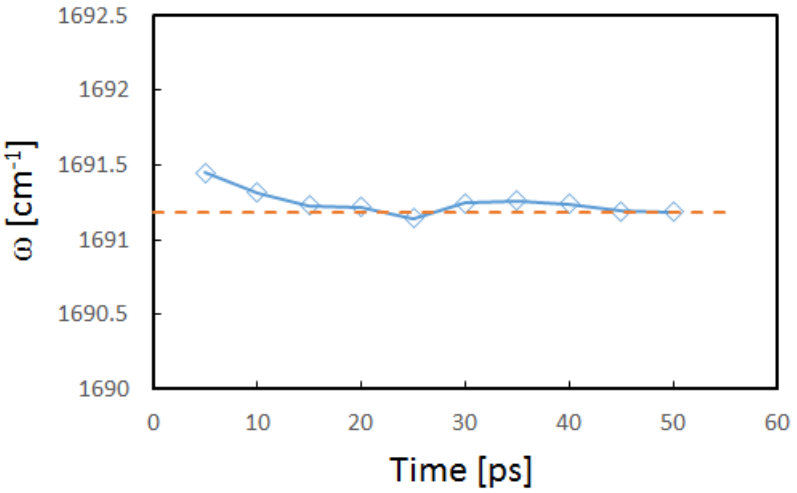


Figure A3. Temporal changes of effective vibrational frequencies $\omega_{FE}(t)$ averaged over 10 MD trajectories, (a) the O-H stretching mode and (b) the C=O stretching mode. The broken lines (---) are the averaged evaluation at t=50 [ps].

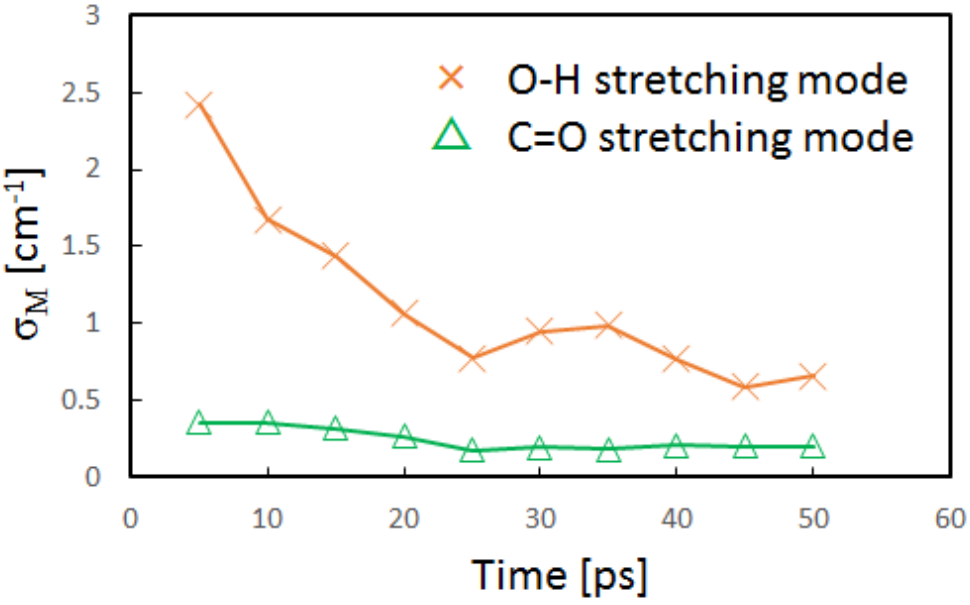


Figure S4. Temporal changes of the standard errors of the mean (SEM) averaged over 10 MD trajectories in case of the O-H and C=O stretching mode.

Table A2. The SEMs of the stretching mode averaged over 10 MD trajectories.

Mode	σ_M [cm^{-1}]
$\nu_{as}(NH_2)$	0.20
$\nu_{s}(NH_2)$	0.24
$\nu(OH)$	0.66
$\nu_{as}(CH_2)$	0.10
$\nu_{s}(CH_2)$	0.08
$\nu(CO)$	0.21

Acknowledgements

This thesis is the summary of my researches in the Graduate School of Information Science, Nagoya University.

I would like to thank many people who supported me to complete this thesis by their invaluable assistance and discussions. First, I would like to express my deep gratitude to Prof. Masataka Nagaoka who gives me kind guidance and continuous encouragement for instructive discussions and initiation into the research publication. I would like to thank also Prof. Nobuaki Koga, Prof. Hedong Zhang, Prof. Satoru Iuchi and Prof. Koji Yasuda who instructed me in theories of electronic states and molecular dynamics. Further, I thank all of collaborators in Nagaoka's laboratory for their many important advices and discussions. Finally, I would like to thank my parents and brother for their "watching over" fondly and supporting me in my research life during the doctor's course.

This thesis was partially supported by a Grant-in-Aid for the 21st Century COE program "Frontiers of Computational Science", the Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sport, Science, and Technology (MEXT) in Japan and also by Core Research for Evolutional Science and Technology (CREST) "High Performance Computing for Multi-scale and Multi-physics Phenomena" and "Establishment of Molecular Technology towards the Creation of New Functions" from the Japan Science and Technology Agency (JST). I also acknowledges the support from Japan Society for the Promotion of Science (JSPS) by

the Research Fellowship for Young Scientist.

Publication List

1. Norio Takenaka, Yukichi Kitamura, Yoshiyuki Koyano, Toshio Asada and Masataka Nagaoka, “Reaction Path Optimization and Vibrational Frequency Analysis via Ab Initio QM/MM Free Energy Gradient (FEG) Method: Application to Isomerization Process of Glycine in Aqueous Solution”, Theor. Chem. Acc. 130 (2011) 215.
DOI: 10.1007/s00214-011-0962-4

2. Yukichi Kitamura, Norio Takenaka, Yoshiyuki Koyano and Masataka Nagaoka, “On the Smoothing of Free Energy Landscape of Solute Molecules in Solution: A Demonstration of the Stability of Glycine Conformers via Ab Initio QM/MM Free Energy Calculation”, Chem. Phys. Lett. 514 (2011) 261.
DOI: 10.1016/j.cplett.2011.08.061

3. Norio Takenaka, Yukichi Kitamura, Yoshiyuki Koyano and Masataka Nagaoka, “The Number-Adaptive Multiscale QM/MM Molecular Dynamics Simulation: Application to Liquid Water”, Chem. Phys. Lett. 524 (2012) 56.
DOI: 10.1016/ j.cplett.2011.12.053

4. Norio Takenaka, Yukichi Kitamura, Yoshiyuki Koyano and Masataka Nagaoka, “An Improvement in Quantum Mechanical Description of Solute-Solvent Interactions in Condensed Systems via the Number-Adaptive Multiscale Quantum Mechanical/Molecular Mechanical-Molecular Dynamics Method: Application to Zwitterionic Glycine in Aqueous Solution”, J. Chem. Phys. 137 (2012) 024501.
DOI: 10.1063/1.4732307

5. Yukichi Kitamura, Norio Takenaka, Yoshiyuki Koyano and Masataka Nagaoka, “Dual Approach to Vibrational Spectra in Solution: Microscopic Influence of Hydrogen Bonding to the State of Motion of Glycine in Water”, J. Chem. Theory Comp. 10 (2014) 3369.
DOI: 10.1021/ct500235a

6. H. C. Georg, T. S. Fernandes, S. Canuto, N. Takenaka, Y. Kitamura and M. Nagaoka, "A Combination of the Sequential QM/MM and the Free Energy Gradient Methodologies with Applications", In "*Practical Aspects of Computational Chemistry III*", Eds. M. Shukla and J. Leszczynski, Springer, 2014
ISBN 978-1-4899-7445-7