

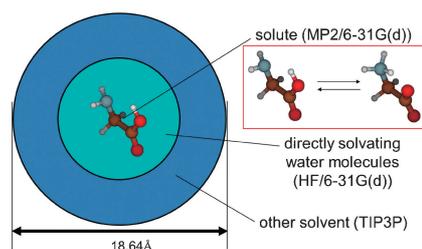
# Helmholtz Energy Change between Neutral and Zwitterionic Forms of Glycine in Aqueous Solution Using Ab Initio Expanded QM/MM-MC with QM Solvent

Hidenori Miyamoto<sup>1,2</sup> and Misako Aida<sup>\*1,2</sup>

<sup>1</sup>Center for Quantum Life Sciences (QuLiS), Hiroshima University,  
1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526

<sup>2</sup>Department of Chemistry, Graduate School of Science, Hiroshima University,  
1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526

(Received May 1, 2013; CL-130412; E-mail: maida@hiroshima-u.ac.jp)



Schematic representation of the theoretical levels which are used to calculate the Helmholtz free energy change of glycine tautomerization in the expanded QM/MM-MC method.

REPRINTED FROM

**Chemistry  
Letters**

Vol.42 No.9 2013 p.1010–1012

CMLTAG  
September 5, 2013

The Chemical Society of Japan

## Helmholtz Energy Change between Neutral and Zwitterionic Forms of Glycine in Aqueous Solution Using Ab Initio Expanded QM/MM-MC with QM Solvent

Hidehiko Miyamoto<sup>1,2</sup> and Misako Aida<sup>\*1,2</sup>

<sup>1</sup>Center for Quantum Life Sciences (QuLiS), Hiroshima University,  
1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526

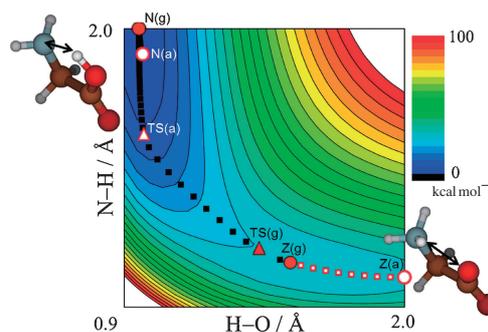
<sup>2</sup>Department of Chemistry, Graduate School of Science, Hiroshima University,  
1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526

(Received May 1, 2013; CL-130412; E-mail: maida@hiroshima-u.ac.jp)

We present an extended methodology of the quantum mechanical/molecular mechanical framework with Metropolis Monte Carlo method (QM/MM-MC), in which close surrounding solvent molecules are included in QM subsystem to evaluate quantitatively the solvation free energy change. The neutral (N) and zwitterionic (Z) structures of glycine in aqueous solution were optimized in the combination procedure of simulated annealing with MC at MM level and QM/MM-vib geometry optimization. Helmholtz energy change between N and Z forms of glycine in aqueous solution was calculated stepwise with expanded QM/MM-MC method, and the computational result was in good agreement with the experimental value.

Glycine is a molecule that can form an intramolecular hydrogen bond and may convert from a neutral (N) form to a zwitterionic (Z) form in an aqueous solution. Since the conversion between N and Z of glycine can be considered as a prototype of the tautomerization of amino acid, the process and the energy profile of the N–Z conversion have been extensively studied.<sup>1–7</sup> Experimentally, the free energy of Z form was estimated to be 7.27 kcal mol<sup>−1</sup> lower than that of N form,<sup>8</sup> and the free energy barrier height from Z to N was estimated to be 14.6 kcal mol<sup>−1</sup>.<sup>9</sup> Recently, we have shown in the quantum mechanical/molecular mechanical framework with Metropolis Monte Carlo method (QM/MM-MC method) that the free energy barrier height of a water-mediated tautomerization path is higher than that of the direct tautomerization path in aqueous environment and that it is not “proton” but “hydrogen atom” that transfers.<sup>7</sup>

In the current work, we aim at performing the quantitative calculation of the N–Z free energy difference of glycine in aqueous solution. To this end, we present an extended QM/MM-MC method to calculate the Helmholtz free energy change. Solvation effect cannot be elucidated simply by taking account of only a few solvent molecules in a few selected configurations of the solvent, even if high level of theory is used. It is necessary to treat many solvent molecules and many solvent configurations explicitly. This is why several methods such as QM/MM-MC method have been adopted. If MC simulations are to be performed to sample solvent configurations satisfactorily, a lower level of theory must be used. If a lower level of theory is used throughout for free energy calculations, however, a resultant computational value may not be reliable. In this Letter, the following three points are addressed to evaluate quantitatively the free energy change in aqueous solution. The first is how to obtain a solute structure in aqueous solution, which may be largely different from that in the gas phase. The second is how to include close surrounding water molecules in QM subsystem in free



**Figure 1.** Potential energy surface of glycine at the HF/6-31G(d) level. The red full circles (N(g) and Z(g)) and the red full triangle (TS(g)) indicate N, Z, and TS in the gas phase, respectively. The white circles (N(a) and Z(a)) and the white triangle (TS(a)) indicate N, Z, and TS in aqueous solution, respectively.

energy calculations. We present a new methodology here, which we call the “expanded QM/MM-MC method.” The last is how to treat the solute with higher level QM in the free energy evaluation. Thus, we present the multistage evaluation of Helmholtz energy change.

At the beginning, we calculated the potential energy surface of a glycine molecule in the gas phase at the HF/6-31G(d) level of theory (Figure 1). By means of vibrational frequency calculations, all the stationary structures were confirmed. The IRC pathway connecting N, TS, and Z structures was calculated. In Figure 1, N and Z structures are indicated by red full circles, and TS structure is indicated by a red full triangle. Some points along the IRC pathway are indicated by black full squares. Note that the hydrogen transfer proceeds in a very narrow ravine. Ab initio MO calculations were performed using Gaussian 03.<sup>10</sup>

To obtain the optimized N and Z structures in an aqueous solution, we put glycine in an aqueous environment with 101 water molecules in a sphere such that the density of the solvent is 1 g cm<sup>−3</sup> with a radius of 9.32 Å. Since many solvent molecules are included in the system, an abundance of local minima may exist. To find a more stable structure, we take the procedure of the combination of the simulated annealing<sup>11</sup> and QM/MM-vib geometry optimization, as follows.

For the geometry optimization in the aqueous environment, we use the QM/MM-vib level. Here, a glycine molecule is treated at the QM level of theory (HF/6-31G(d)), and other 101 solvent water molecules are treated at the MM level of theory (TIP3P<sup>12</sup>) incorporated with the vibrational potential.<sup>13</sup> Total QM/MM energy of a system is expressed as follows.<sup>14,15</sup>

$$E^{\text{QM/MM}} = E_{\text{QM}} + E_{\text{QM-MM}} + E_{\text{MM-MM}} + E_{\text{vib}} \quad (1)$$

Here,  $E_{\text{QM}}$  is the energy of QM subsystem in the presence of the effective charges of MM solvent molecules.  $E_{\text{QM-MM}}$  is the interaction energy between QM subsystem and MM molecules.  $E_{\text{MM-MM}}$  is the interaction energy between MM molecules.  $E_{\text{vib}}$  is the total of the vibrational energy of each MM water molecule. The set of parameters used for the QM/MM calculation is the same as those used in our previous work,<sup>7</sup> except for the vibrational energy, which is calculated from the quartic force field.<sup>13</sup>

The Metropolis Monte Carlo (MC) algorithm was used for the simulated annealing process, in which a high temperature (298 K, in this work) was given to the system, allowing it to equilibrate, then it was cooled in steps of 14.9 K until the temperature is zero, and 505000 configurations were generated randomly at each temperature. In an annealing process, a solute structure was fixed, and MM/MM calculations were performed, where the second and the third terms of eq 1 were used for energy calculations, and the electrostatic interaction term between solute and solvent was calculated by the Coulomb interactions between the CHELPG<sup>16</sup> using HF/6-31G(d) of the solute atoms and the partial charges of the solvent atoms.

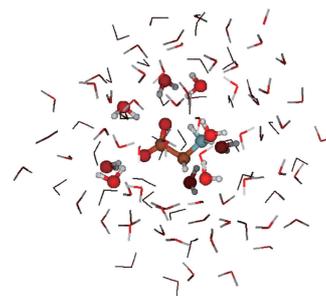
In the procedure of QM/MM geometry optimization combined with simulated annealing, several cycles of the simulated annealing process were performed, and then all the positions of the molecules (glycine and 101 MM water molecules) were optimized in the framework of QM/MM-vib level. This procedure was repeated several times. Then, we obtained an optimized geometry of glycine in 101 water molecules. The geometrical parameters of thus-obtained N and Z forms in the aqueous solution are plotted in Figure 1, with the white circles. Note that the Z structure in the aqueous solution is very much different from that in the gas phase (Figure S1<sup>17</sup>). The XYZ coordinates of these structures are listed in Tables S1 and S2.<sup>17</sup> In the gas phase, one of H atoms of the NH<sub>3</sub> group is strongly hydrogen-bonded to one of O atoms of OOC group. In the aqueous solution, they are separated so that both groups are exposed to solvent molecules. The H...O distance, 2.05 Å, is longer than that reported previously.<sup>4</sup> This may be because we performed the combination procedure of simulated annealing and geometry optimization several times to obtain Z form in aqueous solution.

The routes between the structures (N or Z) in the gas phase and the corresponding structures in aqueous solution were created by dividing the differences in the coordinates linearly. White squares in Figure 1 indicate the route from the Z structure in the gas phase to that in aqueous solution. Since the selection of a reaction path does not affect the N–Z Helmholtz energy difference, we can take any reaction path in principle. We use the IRC path with extended steps (black and white squares in Figure 1) to evaluate the Helmholtz energy difference between the N and Z structures in aqueous solution by means of the free energy perturbation method. The solvation effect is the target of the current work, and solvent configurational contributions must be important. We assume the system obeys classical statistical mechanics.

The difference  $\Delta A_{ij}$  in the classical Helmholtz free energy of the system between glycine structures  $i$  and  $j$  is given by

$$\Delta A_{ij} = -kT[\ln\langle \exp\{-(E_j^{\text{QM/MM}} - E_i^{\text{QM/MM}})/kT\} \rangle_{(i)}] \quad (2)$$

The bracket  $\langle \dots \rangle_{(i)}$  denotes the ensemble average of the difference in total QM/MM energies of glycine (in the structures



**Figure 2.** A snapshot of solvent configuration. Every water molecule is replaced by QM, when an atomic distance between the solute and the water molecule is less than 2.45 Å. The molecules included in the QM subsystem are drawn with the ball-and-stick model.

$i$  and  $j$ ) with the solvent molecules in configurations originating from the Metropolis sampling around glycine in the structure  $i$  in a solvation sphere with a radius of 9.32 Å.

To sample solvent configurations, we adopted the double perturbation approach.<sup>18</sup> In this approach, at a solute structure, 2020000 solvent configurations were generated (NVT ensemble at  $T = 298$  K) by MM/MM calculations. Since only the energy differences of the solute structures of  $i$  and  $j$  with the same solvent configuration based on the solute structure  $i$  are taken into account in the free energy calculations (eq 2), we consider that the NVT ensemble generated from MM/MM calculations with TIP3P is adequate. From the generated 2020000 configurations, 2020 were selected randomly to be used for the *expanded* QM/MM calculations, where *the solute and directly solvating water molecules were treated as QM* (HF/6-31G(d)). We regard a water molecule as “directly solvating,” when the atomic distance between the solute and the water molecule is less than 2.45 Å. A snapshot of solvent configuration around Z form of glycine is shown in Figure 2. The QM/MM and free energy calculations were performed using HONDO.<sup>19</sup>

Thus-obtained Helmholtz free energy change of the system can be divided<sup>20</sup> into the free energy change of the solute and the solvation free energy change as follows.

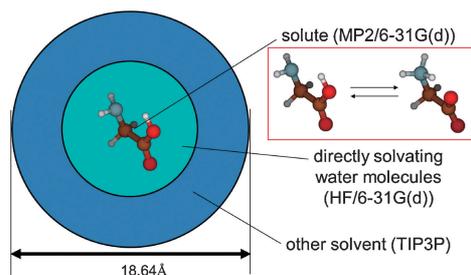
$$\Delta A_{\text{system}}^{\text{HF/HF/MM}} = \Delta A_{\text{solute}}^{\text{HF}} + \Delta \Delta A_{\text{solvation}}^{\text{HF/MM}} \quad (3)$$

The superscript indicates the level of theory used for the evaluation for each term. Since our free energy difference calculations are performed with a fixed geometry of the solute at each of the tautomerization route, the free energy change of the solute is equal to the potential energy change of the solute along the tautomerization. Then, we obtain

$$\Delta A_{\text{system}}^{\text{HF/HF/MM}} = \Delta E_{\text{solute}}^{\text{HF}} + \Delta \Delta A_{\text{solvation}}^{\text{HF/MM}} \quad (4)$$

The potential energy change of the solute,  $\Delta E_{\text{solute}}^{\text{HF}}$ , is obtained by performing ab initio MO calculations at HF/6-31G(d) level. Thus, the solvation free energy change,  $\Delta \Delta A_{\text{solvation}}^{\text{HF/MM}}$ , is obtained by subtracting  $\Delta E_{\text{solute}}^{\text{HF}}$  from  $\Delta A_{\text{system}}^{\text{HF/HF/MM}}$ . The obtained solvation free energy change is plotted with the reversed sign using blue diamonds in Figure 4.

The following is the last stage of our evaluation of free energy change. Since it is probable that the energy differences among the structures of the solute are strongly sensitive to the electron correlation effect, we perform MP2/6-31G(d) calculations for the solute structures. Then, we add the MP2 energy



**Figure 3.** Schematic representation of the theoretical levels which are used to calculate the Helmholtz free energy change of glycine tautomerization in the expanded QM/MM-MC method.

change to the solvation free energy change  $\Delta\Delta A_{\text{solvation}}^{\text{HF/MM}}$ . Thus, we obtain the Helmholtz free energy change, in which the solute is treated with MP2/6-31G(d):

$$\Delta A_{\text{system}}^{\text{MP2/HF/MM}} = \Delta E_{\text{solute}}^{\text{MP2}} + \Delta\Delta A_{\text{solvation}}^{\text{HF/MM}} \quad (5)$$

This is our final expression for Helmholtz free energy change in this Letter. Schematic representation is shown in Figure 3.

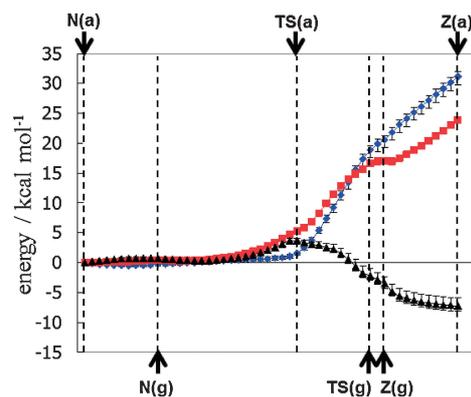
We used 52 points along the route of direct hydrogen transfer of glycine in the aqueous solution, for each of which 2020 expanded QM/MM-MC calculations were performed, where a glycine molecule with directly solvating water molecules was calculated at HF/6-31G(d) surrounded by the rest of water molecules of TIP3P. Then, the replacement of QM energy of HF/6-31G(d) with that of MP2/6-31G(d) led us to the Helmholtz free energy change according to eq 5. The computational result is shown in Figure 4. The standard deviations at each step are indicated for Helmholtz free energy change (black) and solvation free energy change (blue).

TS structure in the aqueous solution, which corresponds to the highest point on the Helmholtz free energy change, is plotted in the solute potential energy surface with the white triangle (Figure 1). It is interesting to note that the TS structure is closer to the N form in the aqueous solution than in the gas phase. The calculated free energy barrier height from Z to N is 11.0 kcal mol<sup>-1</sup>, which is slightly lower than the experimental estimation.<sup>9</sup> One of the reasons for this may be the insufficiency of the theoretical level used to calculate the solute energy change along the route. The solute energies of the intermediate steps of hydrogen atom transfer may depend on the theory used: MP2 level may not be good enough in the current case.

In the current work, we treated directly solvating water molecules (when they are within 2.45 Å of the solute) as QM (HF/6-31G(d)), which resulted in larger stabilization in solvation free energy than that of the previous work<sup>7</sup> where all solvent water molecules were treated as MM. When a solute is ionized, water molecules surrounding the solute must be polarized, and they need to be treated as QM.

Regarding the tautomerization route, there may be other paths than the one considered here. Furthermore, there may be various states corresponding to N form and Z form of glycine in aqueous solution. In the current work, we do not extend our calculations to those other routes or other states. Concerning the N states, it has been suggested that the free energy differences between those states are rather small.<sup>21</sup>

In aqueous solution, our computational Helmholtz free energy of the Z form of glycine is lower than that of the



**Figure 4.** Helmholtz free energy change (black full triangles), solute potential energy change (red full squares), and solvation free energy change (blue full diamond) along the hydrogen-transfer route of glycine in aqueous solution, according to eq 5. The solvation free energy change is plotted with the reversed sign. N(a), TS(a), and Z(a) indicate N, TS, and Z forms in aqueous solution, respectively. N(g), TS(g), and Z(g) indicate those in the gas phase. The standard deviations at each step are indicated for Helmholtz free energy change (black) and solvation free energy change (blue).

N form by 7.25 kcal mol<sup>-1</sup>, and it compares well with the experimental value (7.27 kcal mol<sup>-1</sup>).<sup>8</sup> The multistage strategy, which we have presented in this Letter, is a practical and promising way to evaluate the Helmholtz free energy change in solution quantitatively; enough solvent configurations are generated at MM level, a stable solute geometry is optimized fully in aqueous solution, solvation free energy is evaluated from selected configurations treating the solute with directly solvating water molecules at HF level, and the solute energy change is evaluated at higher level of theory.

#### References and Notes

- M. Nagaoka, N. Okuyama-Yoshida, T. Yamabe, *J. Phys. Chem. A* **1998**, *102*, 8202.
- K. Leung, S. B. Rempe, *J. Chem. Phys.* **2005**, *122*, 184506.
- C. M. Aikens, M. S. Gordon, *J. Am. Chem. Soc.* **2006**, *128*, 12835.
- N. Takenaka, Y. Kitamura, Y. Koyano, T. Asada, M. Nagaoka, *Theor. Chem. Acc.* **2011**, *130*, 215.
- C. H. Choi, S. Re, M. Feig, Y. Sugita, *Chem. Phys. Lett.* **2012**, *539–540*, 218.
- N. Takenaka, Y. Kitamura, Y. Koyano, M. Nagaoka, *J. Chem. Phys.* **2012**, *137*, 024501.
- H. Miyamoto, M. Aida, *Chem. Lett.* **2013**, *42*, 598.
- G. Wada, E. Tamura, M. Okina, M. Nakamura, *Bull. Chem. Soc. Jpn.* **1982**, *55*, 3064.
- M. A. Slifkin, S. M. Ali, *J. Mol. Liq.* **1984**, *28*, 215.
- M. J. Frisch, et al., *Gaussian 03 (Revision A.02)*, Gaussian, Inc., Wallingford CT, **2004**.
- S. Kirkpatrick, C. D. Gelatt, Jr., M. P. Vecchi, *Science* **1983**, *220*, 671.
- W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
- R. J. Bartlett, I. Shavitt, G. D. Purvis, *J. Chem. Phys.* **1979**, *71*, 281.
- M. Aida, H. Yamataka, M. Dupuis, *Int. J. Quantum Chem.* **2000**, *77*, 199.
- M. Dupuis, M. Aida, Y. Kawashima, K. Hirao, *J. Chem. Phys.* **2002**, *117*, 1242.
- C. M. Breneman, K. B. Wiberg, *J. Comput. Chem.* **1990**, *11*, 361.
- Supporting Information is available electronically on the CSJ-Journal Web site, <http://www.csj.jp/journals/chem-lett/index.html>.
- M. Dupuis, G. K. Schenter, B. G. Garrett, E. E. Arcia, *THEOCHEM* **2003**, *632*, 173.
- M. Dupuis, *HONDO2004*, based on HONDO-95 available from the Quantum Chemistry Program Exchange, Indiana University.
- M. Ohisa, H. Yamataka, M. Dupuis, M. Aida, *Phys. Chem. Chem. Phys.* **2008**, *10*, 844.
- Y. Kitamura, N. Takenaka, Y. Koyano, M. Nagaoka, *Chem. Phys. Lett.* **2011**, *514*, 261.