Comprehensive studies on the tautomerization of glycine: a theoretical study

Chang Kon Kim, Byung-Ho Park, Hai Whang Lee and Chan Kyung Kim*

The tautomerization process of glycine between the neutral (NE) and zwitterionic (ZW) forms in aqueous solution was explored theoretically using the conductor-like polarizable continuum model (CPCM) by adopting the PAULING cavity model at the B3LYP, MP2 and CCSD levels with the 6-311+G(d,p) basis set. The tautomerization of glycine is unable to be predicted satisfactorily within the equilibrated framework of the CPCM method. Instead, in this study, three plausible non-equilibrated solvation situations were assumed: (S-1) one water molecule attached to the transferring proton in the ZW moves together with the transferring proton; (S-2) one water molecule attached to the transferring proton in the ZW remains motionless at a fixed position near the NH₂ fragment at the TS structure; and (S-3) proton transfer occurs without changing the position of the surrounding water molecules from their initial state, the ZW form, in the eight water clusters. Although the calculation of (S-3) failed, the Gibbs free energies of activation for tautomerization from the ZW to NE, \( \Delta G^\ddagger(ZW \rightarrow NE) \), was well consistent with the experimental findings in the hypothetical non-equilibrated solvation states of (S-1) and (S-2). This suggests that non-equilibrium solvation is essential to explain the observed experimental data.

Introduction

In metabolism, amino acids play important roles as the building blocks of proteins, and many experimental\(^{1–4}\) and theoretical studies\(^{5–19}\) have been reported. Among the amino acids, glycine has attracted considerable attention in theoretical studies\(^{5–18}\) owing to its small size and the availability of experimental data.\(^{1–4}\) Moreover, it has been reported that the neutral form (NE) of glycine exists in the gas phase but the zwitterionic form (ZW) is unstable in vacuo.\(^{5–18}\) On the other hand, the ZW of glycine predominates in crystalline or in aqueous solutions.\(^{2,3}\) Therefore, solvent (or environmental) effects should be included in theoretical studies\(^{20–25}\) to examine glycine chemistry.\(^{5–18}\) In previous work,\(^{14}\) the conformations of glycine and inter-conversion processes between them were examined theoretically in the gas phase and in aqueous solution, and it was shown that the conductor-like polarizable continuum model (CPCM)\(^{26,27}\) adopting the cavity model with explicit hydrogen(s), such as the BOND\(^{28}\) or PAULING\(^{29}\) cavity model, could be applied to aqueous solution. On the other hand, the Gibbs free energies of activation for tautomerization going from the ZW to NE forms, \( \Delta G^\ddagger(ZW \rightarrow NE) \), were underestimated compared to the experimental value,\(^{1,3}\) even though the \( \Delta G^\ddagger(ZW \rightarrow NE) \) value was fairly consistent with the experimental one relative to any other theoretical studies reported previously,\(^{8,9,12,16}\) i.e. the experimental \( \Delta G^\ddagger(ZW \rightarrow NE) \) value was reported to be 14.6 kcal mol\(^{-1}\),\(^{1,2}\) but the \( \Delta G^\ddagger(ZW \rightarrow NE) \) value obtained at the CPCM-CCSD/6-311+G(d,p) level was 9.2 kcal mol\(^{-1}\).\(^{1,14}\) Tortonda et al. reported that the activation energy barrier for tautomerization was only 3.5 kcal mol\(^{-1}\) at the MP2(Full)/6-31+G(d,p) level in aqueous solution using the continuum model.\(^{6,18}\) On the other hand, this theoretical barrier was too low to compare with the experimental value.\(^{1,2}\) The activation barrier for the transformation from the NE to ZW for the proton transfer with a single water molecule was reported to be 15.56 kcal mol\(^{-1}\) at the MP2(Full)/6-31+G(d,p) level in the gas phase, which is much higher than the experimentally determined value, 7.3 kcal mol\(^{-1}\).\(^{6}\) This experimental value was obtained by considering the \( \Delta G^\ddagger(ZW \rightarrow NE) \) and \( \Delta G^\ddagger(ZW \rightarrow NE) \) values of 14.6 and 7.3 kcal mol\(^{-1}\), respectively. Similarly, the \( \Delta G^\ddagger(ZW \rightarrow NE) \) value reported by Tuñón et al. was too low (only 5.42 kcal mol\(^{-1}\)) at the PCM-B3LYP/6-31+G(d,p) level.\(^{17}\) Therefore, they suggested that the experimental barrier of 7.3 kcal mol\(^{-1}\) corresponds to the reorientation of the H-atom in the –COOH group of the neutral conformers. This suggestion appears questionable because the activation barriers for the reorientation of the –COOH group between the NE conformers could be reversely too high compared to the experimental finding of 7.3 kcal mol\(^{-1}\). In previous work, the highest activation barrier among the inter-conversions between the NE conformers was 13.8 kcal mol\(^{-1}\).\(^{1,14}\) On the other hand, the activation free...
energies for the ZW → NE and NE → ZW conversions were reported to be 16.9 and 8.5 kcal mol⁻¹, respectively, using empirical valence bond (EVB) molecular dynamics simulations, which is in good agreement with the experimental data.¹⁶ This appears to be fortuitous because the parameters of the EVB method were obtained from the activation barriers that were overestimated considerably at the Hartree–Fock level of theory. Therefore, comparisons of the activation barriers between the experimental and theoretical studies are unclear.

In this study, the tautomerization of glycine was examined in more detail to understand the precise mechanism because Kim et al. were only concerned with a direct intra-molecular proton transfer between ZW and NE.¹⁴ On the other hand, the tautomerization process could proceed through many other possible processes, i.e. a proton exchange process between the solvent (water) molecule and the solute (glycine) producing an anionic intermediate (AN) of glycine and hydronium ions, and inter-molecular proton transfer by a participating water molecule, as shown in Scheme 1.

### Results and discussion

As reported previously,¹⁴ the ΔG°(ZW → NE) values determined using the direct proton transfer process, (A) in Scheme 1, were 9.2 and 8.8 kcal mol⁻¹ at the CCSD and CCSD(T) levels with the 6-311+G(d,p) basis set, respectively, in aqueous solution. On the other hand, the calculated values appeared to be somewhat lower than the experimental value of 14.6 kcal mol⁻¹. Therefore, the possibility that tautomerization proceeds via two different paths, (B) and (C) in Scheme 1, cannot be excluded. Especially, the path (B) could be a plausible process, because the autoionization processes of ammonia²¹ and hydrogen halides²² in water are similar to path (B) and these processes were studied extensively. Table 1 summarizes the ΔG°(ZW → NE) values for path (B) calculated at various levels. Tautomerization via a proton exchange process with AN and hydronium ion as intermediates, path (B), could be excluded because the Gibbs free energy changes from the ZW + H₂O cluster to AN + H₃O⁺, ΔG_{AN-ZW}, were much more unfavorable than the experimental ΔG°(ZW → NE) values. For example, the ΔG_{AN-ZW} values were 35.5 and 35.7 kcal mol⁻¹ at the CCSD and CCSD(T) levels, respectively. Moreover, the Gibbs free energies of the intermediates, AN and hydronium ion, were also 26.3 and 26.9 kcal mol⁻¹ higher, respectively, than that for the direct intra-molecular proton transfer, path (A), at the CCSD and CCSD(T) levels. This suggests that the tautomerization of ZW ⇄ NE in glycine chemistry cannot occur through acid–base equilibrium.

On the other hand, Table 2 shows that the ΔG°(ZW → NE) values for the inter-molecular proton transfer with a participating water molecule, path (C), 14.8 and 14.0 kcal mol⁻¹ at the

---

**Scheme 1**

(A) Direct intra-molecular proton transfer.

(B) Proton exchange process between water and glycine via the AN and hydronium ion.

(C) Inter-molecular proton transfer with a participating water molecule.

---

**Computational methods**

All the structures studied in aqueous solution were fully optimized without geometrical constraints using the CPCM method²⁶,²⁷ adopting the PAULING cavity model²⁹ at the B3LYP,³⁰,³¹ MP2,³² and CCSD⁻¹,³⁴ levels with the 6-311+G(d,p) basis set. The optimized structures were then characterized by frequency calculations at the B3LYP and MP2 levels. The energetics were then refined at the CCSD(T)/6-311+G(d,p) level using the geometries optimized at the CCSD/6-311+G(d,p) level. In the CPCM method, non-electrostatic terms are important because the computed energies depend on the cavity size, which is one of the major components of the non-electrostatic terms.²⁶,²⁷ Therefore, in this study, the calculated Gibbs free energies in aqueous solution were obtained using eqn (1), where G is the Gibbs free energy of solvation including the non-electrostatic terms, such as cavitation and dispersion. In eqn (1), E_{el} is the electronic energy in the gas phase on the geometry optimized at the CPCM calculation, and E_{ZPV}, E_{TH}, and S are the zero-point vibration energy, thermal energy and entropy terms, respectively, which were obtained from CPCM calculations. The CCSD(T) calculations in aqueous solution were performed at the gas-phase CCSD(T) level on the geometries at the CPCM-CCSD/6-311+G(d,p) or CPCM-B3LYP/6-311+G(d,p) levels, and the G values at the CCSD(CCSD) level were used because the CCSD(T) calculations were not applicable to the CPCM method. All the calculations were performed with the Int(grid = ultrafine) option using the Gaussian 03 program.³⁵

\[
G(\text{at } 298 \text{ K}) = E_{el} + E_{ZPV} + E_{TH} + PV - TS + G_s
\]

\[
= E_{el} + G_{corr} + G_s
\]
CPCMC-CCSD and CPCMC-CCSD(T) levels, respectively, agreed well with the experimental data. Therefore, the tautomerization of glycine between ZW and NE could be expected to proceed via path (C). This expectation, however, would give rise to a serious problem. Although the ΔG‡ values of path (C) agreed well with the experimental findings, the tautomerization should proceed via path (A) because of the ~4 kcal mol⁻¹ lower barrier for path (A).

To understand these problems, this study examined the tautomerization of direct proton transfer with one water molecule similar to path (C). In this process, it was assumed that the water molecule is simply to solvate the transferring proton, as shown by (D1) in Scheme 2. Recently, the solvation effects by a single water molecule were well discussed by Kolaski and coworkers. However, no plausible TS structures could be located in this process. Instead, the TS structures converged to (D2) in Scheme 2, in which the solvating water molecule was hydrogen-bonded to one hydrogen on the -NH₂ functional group far apart from the transferring proton. Consequently, the ΔG‡ values via this process were similar to those via path (A) at all the theoretical levels employed. For example, the ΔG‡ values for paths (A) and (D2), respectively, at the CPCM-B3LYP/6-311+G(3df,2p) level. The TS structure of (D2) might be the result of a problem with the CPCM method used in this study. The clusters with eight water molecules were studied in the gas phase to examine this phenomenon in more detail. Fig. 1 shows the optimized structures at the B3LYP level of theory.
with 6-311+G(d,p) and 6-311+G(3df,2p) basis sets. Again, the fully optimized TS structures were similar to (D2), i.e. the solvating water molecules were far apart from the transferring proton. Therefore, the TS structure similar to (D2) was the intrinsic nature of the tautomerization of glycine and not a problem with the CPCM method.

The structures optimized using the CPCM method in the traditional condition might, in principle, correspond to an equilibrium or at least quasi-equilibrium state. In general, the proton transfer process normally takes place on a very short time scale but the orientational relaxation time of the solvent molecules might take place approximately 1000 times later than the time for proton transfer. This suggests that the tautomerization TS between the ZW and NE of glycine might be in the non-equilibrium state of solvent structures because tautomerization typically corresponds to proton transfer. If this is true, the strategy considered above could not be described adequately in terms of the equilibrium hypothesis. Kurz and Kurz proposed an interesting situation for proton transfer in solution, i.e. the fast protonic motions could be resisted by a force that would not be present if solvent relaxation was fast. According to their work, if the activated complex has a non-equilibrated environment, the deviation of a solvent configuration from its equilibrium state could be expected to be toward a configuration that is appropriate for an internal structure, where the proton is half-transferred. Moreover, they showed that their model is related both to the earlier qualitative suggestions by Schowen et al., Ritchie and the Marcus theory of proton transfer reactions. On the other hand, some theoretical studies also used continuum models for the non-equilibrium solvation effects on proton transfer. In particular, Tortonda et al. showed that the activation barrier for the NE → ZW conversion of serine changes substantially in the range of 1.5–5.5 kcal mol⁻¹ when the non-equilibrated solvent effects are considered. On the other hand, they concluded that proton transfer in the tautomerization of serine is extremely favorable with a negligible activation barrier, which is similar to that of glycine shown by the quantum mechanical–molecular mechanics (QM–MM) simulation. Instead, to account for the differences in activation barriers between the experimental and theoretical results for the tautomerization of amino acids, they suggested that the essential activation barrier for producing ZW would not have originated from proton transfer itself but by a conformational change between the NE conformers. Although it is possible that the essential activation barrier for producing ZW originated from a conformational change between the NE conformers, a previous study showed that the activation barriers for inter-conversions between the NE conformers are much higher than the experimental ΔG°(ZW → NE) values at the sophisticated level of theory, such as CCSD and CCSD(T) levels. Therefore, the suggestion of Tuñón et al. appears questionable.

In previous work, theoretical results from the CPCM method using the PAULING cavity model agreed well with the experimental findings, such as the differences in Gibbs free energy between ZW and NE of glycine, ΔGZW-NE in aqueous solution. The ΔGZW-NE values were defined as the Gibbs free energy differences between the ZW and the most stable NE in aqueous solution. Therefore, it is expected that the kinetic quantities, such as the ΔG°(ZW → NE) obtained using the same method, are also consistent with the experimental findings. On the other hand, as noted above, the ΔG°(ZW → NE) values were underestimated considerably (by more than 4 kcal mol⁻¹) even at the CPCM-CCSD and CCSD(T) levels using the 6-311+G(d,p) basis set. Nevertheless, these values agreed reasonably well with the experimental data compared to those reported previously. To understand the reasons for the underestimation, the non-equilibrium solvation effects at the TS were examined by assuming three situations: (S1), one water molecule attached to the transferring proton in the ZW moves together with the transferring proton; (S2), one water molecule attached to the transferring proton in the ZW remains motionlessly at a fixed position near the NH₂ fragment at the TS structure; and (S3), proton transfer occurred without changing the position of the surrounding water molecules from its initial state, the ZW form, in the eight water clusters.

(S1) was assumed to be a non-equilibrated solvent configuration because the water molecule attached to the proton moves while the proton is transferring. To observe this motion, the bond distance, dO→H, between the transferring proton and oxygen atom on the moving water molecule was fixed arbitrarily. These structures do not correspond to the

![Image](48x466 to 284x729)
stationary (and/or an equilibrated) species. Interestingly, the species with $d_{O-H}$ in the threshold range of 2.03–2.45 Å have only one negative eigenvalue in the Hessian matrix at the CPCM-B3LYP/6-311+G(d,p) level of theory, and the corresponding normal mode of vibration was found to be the transferring motion of the proton. In contrast, the species with only one negative eigenvalue in the Hessian matrix could not be located when $d_{O-H}$ is out of the threshold range (<2.0 Å or >2.5 Å). This suggests that the species with $d_{O-H}$ in the threshold range have similar characteristics to the TS in the equilibrated solvent state. Therefore, the species becomes one of the TSs in the non-equilibrated solvent state. Fig. 2 shows the relationship between $\Delta G^s(ZW \rightarrow NE)$ and $d_{O-H}$ at the CPCM-CCSD/6-311+G(d,p)//CPCM-B3LYP/6-311+G(d,p) level. The calculated $\Delta G^s(ZW \rightarrow NE)$ values decreased gradually with lengthening the $d_{O-H}$ value. Table 3 lists the detailed energetic results.

Starting from the equilibrium solvent state, (D2), the species with $d_{O-H}$ = 2.03 Å becomes an actual TS in the non-equilibrium state when proton transfer takes place in a very short time. This is a reasonable assumption because the $d_{O-H}$ in Zw was 1.993 Å, which changes little at the non-equilibrium TS. Therefore, the $\Delta G^s(ZW \rightarrow NE)$ of 13.4 kcal mol$^{-1}$ obtained when $d_{O-H} = 2.03$ Å agrees well with the experimental findings.$^{1,2}$

In (S-2), the solvent configuration was assumed to be frozen in the initial state (the Zw conformer) during rapid proton transfer. In this process, the inter-atomic distances of an oxygen atom on the solvating water molecule to the N and O atoms of glycine were fixed. Interestingly, the calculated species had only one negative eigenvalue in the Hessian matrix and the corresponding normal modes of vibration were the transferring motion of the proton. Therefore, the species might be expected to be one of the TSs in the non-equilibrated solvent state, as noted above for (S-1). The $\Delta G^s(ZW \rightarrow NE)$ for (S-2) was 17.0 kcal mol$^{-1}$ at the CPCM-CCSD//CPCM-B3LYP level, which is also consistent with the experimental findings. Fig. 3 shows the optimized structure for (S-2).

![Fig. 3](image_url) The non-equilibrated solvent transition state for S-2. Bond lengths are in Å.

### Table 3

Energetic results as a function of $d_{O-H}$ in situation (S-1) at the CPCM-CCSD/6-311+G(d,p) level of theory

<table>
<thead>
<tr>
<th>$d_{O-H}$ (Å)</th>
<th>$E_{rel}^s$ (kcal mol$^{-1}$)</th>
<th>$G_{corr}^{b,c}$ (kcal mol$^{-1}$)</th>
<th>$\Delta G_s^d$ (kcal mol$^{-1}$)</th>
<th>$G^b$ (kcal mol$^{-1}$)</th>
<th>$E_{ref}^d$ (kcal mol$^{-1}$)</th>
<th>$\delta G_s^d$ (kcal mol$^{-1}$)</th>
<th>$\Delta G^{s,d,e}$ (kcal mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.03</td>
<td>-360.07031</td>
<td>0.0625</td>
<td>-33.5</td>
<td>-360.05752</td>
<td>6.5</td>
<td>0.2</td>
<td>13.4</td>
</tr>
<tr>
<td>2.05</td>
<td>-360.07058</td>
<td>0.06575</td>
<td>-33.6</td>
<td>-360.05836</td>
<td>6.3</td>
<td>0.1</td>
<td>12.8</td>
</tr>
<tr>
<td>2.10</td>
<td>-360.07133</td>
<td>0.06526</td>
<td>-33.7</td>
<td>-360.05980</td>
<td>5.8</td>
<td>0.0</td>
<td>11.9</td>
</tr>
<tr>
<td>2.15</td>
<td>-360.07644</td>
<td>0.06415</td>
<td>-31.1</td>
<td>-360.06185</td>
<td>2.6</td>
<td>2.6</td>
<td>10.6</td>
</tr>
<tr>
<td>2.20</td>
<td>-360.07835</td>
<td>0.06602</td>
<td>-30.4</td>
<td>-360.06076</td>
<td>1.4</td>
<td>3.3</td>
<td>11.3</td>
</tr>
<tr>
<td>2.25</td>
<td>-360.07990</td>
<td>0.06558</td>
<td>-30.4</td>
<td>-360.06185</td>
<td>1.0</td>
<td>3.3</td>
<td>10.7</td>
</tr>
<tr>
<td>2.35</td>
<td>-360.07974</td>
<td>0.06546</td>
<td>-30.5</td>
<td>-360.06296</td>
<td>0.5</td>
<td>3.2</td>
<td>10.0</td>
</tr>
<tr>
<td>2.40</td>
<td>-360.08025</td>
<td>0.06480</td>
<td>-30.5</td>
<td>-360.06404</td>
<td>0.2</td>
<td>3.2</td>
<td>9.3</td>
</tr>
<tr>
<td>2.45</td>
<td>-360.08061</td>
<td>0.06517</td>
<td>-30.5</td>
<td>-360.06398</td>
<td>0.0</td>
<td>3.3</td>
<td>9.3</td>
</tr>
<tr>
<td>Zw</td>
<td>-360.06792</td>
<td>0.07325</td>
<td>-52.8</td>
<td>-360.07882</td>
<td>8.0</td>
<td>-19.1</td>
<td>—</td>
</tr>
<tr>
<td>TS</td>
<td>-360.07757</td>
<td>0.06789</td>
<td>-26.2</td>
<td>-360.05147</td>
<td>1.9</td>
<td>7.5</td>
<td>—</td>
</tr>
</tbody>
</table>

$^a$ Geometries at the CPCM-B3LYP/6-311+G(d,p) level were used. $^b$ Values are in kcal mol$^{-1}$. $^c$ Values at the CPCM-B3LYP/6-311+G(d,p) level were used. $^d$ Values in kcal mol$^{-1}$. $^e$ $\Delta G_s^d = G(D1) - G(ZW·H_2O)$.
Nonetheless, the non-equilibrated states using one solvent (water) molecule with the arbitrarily fixed distances could not be the real stationary species on the potential energy surface (PES). Therefore it might be insufficient to explain any non-equilibrated solvation states clearly. To overcome this problem, the clusters with eight water molecules as shown in Fig. 1 were examined in aqueous solution. Fig. 1 shows, as is generally expected, that the eight solvent water molecules in the ZW form were bound more tightly owing to its zwitterionic character than those in the NE form. These water molecules should reorganize gradually during the tautomerization process on going from the ZW form to the NE form (or vice versa), if the tautomerization occurs under an equilibrated solvation state. Unfortunately, the fully optimized TS structure similar to (b) in Fig. 1 could not be located in aqueous solution despite numerous attempts to find it. The difficulty has arisen due to a sudden structural change of solvated water molecules in the vicinity of a plausible TS structure. However, we managed to locate a second-order saddle point, (D3), with two negative eigenvalues in the Hessian matrix at the CPCM-B3LYP/6-311+G(d,p) level\(^{46}\) as shown in Fig. 4. Normal mode analyses show that one negative frequency (\(\nu_1 = -756.1\) cm\(^{-1}\)) corresponds to the proton transfer, but the second one (\(\nu_2 = -78.3\) cm\(^{-1}\)) corresponds to the fluctuation of the eight solvent water molecules (see Fig. 4b for normal modes of vibrations). Note that the difference in Gibbs free energy between the NE form and (D3) was only 1.9 kcal mol\(^{-1}\). This value was nearly the same as those obtained from the CPCM calculations\(^{14}\) and much lower than the experimental value of 7.3 kcal mol\(^{-1}\). Again, this could strongly imply that the tautomerization of glycine proceeds via the non-equilibrated solvation states. Nevertheless, the clusters with eight water molecules might be still arbitrary, and thus a more detailed study is in progress in our laboratory.

Finally, in (S-3), structures corresponding to the TS and NE were inserted into the cavity formed by the eight water molecules by removing the ZW in the eight water clusters of the ZW form. To perform this operation, the root-mean square deviation (RMSD) of the four heavy atoms (N and O atoms directly involved in the proton transfer and two carbon backbone atoms) was minimized and the RMSD-minimized structures were placed inside the cavity formed by the ZW. As expected from Fig. 1, the ZW formed the tightest hydrogen-bonded network owing to its zwitterionic character among the three structures, and inserting the TS structure allowed bumping between the moving proton and the hydrogen atom in the solvent shell (bond distance = 0.937 Å), which makes this plausible structure highly unstable (\(\Delta G^\ddagger(ZW \rightarrow NE) = 102.49\) kcal mol\(^{-1}\)). This shows that the procedure using eight water clusters failed to provide a reasonable non-equilibrium description.

**Conclusions**

From the theoretical studies on the tautomerization of glycine, non-equilibrium solvation is essential to explain the observed experimental data. In the CPCM methodology, the user cannot control the solvent configuration of the continuum model. The only way to accomplish this is to introduce ancillary water molecule(s) to consider the non-equilibrated solvent states. The following conclusions were drawn from this study:

(i) The tautomerization of glycine can proceed via inter-molecular proton transfer with a participating water molecule, (C) in Scheme 1, if the process occurs in an equilibrated solvent state.

(ii) The tautomerization process can occur via direct intra-molecular proton transfer, (S-1) or (S-2) without active catalytic water molecule(s), if the process occurs in a non-equilibrated solvent state.

(iii) The proton exchange process, such as (B) in Scheme 1, could not occur because the sum of the energies of the AN + hydronium ion is higher than that of the ZW + H\(_2\)O.

**Acknowledgements**

This study was supported by INHA University. One of the authors (C. K. Kim) thanks Dr. Noham Weinberg for his helpful discussions.
References

46. The CPCM options were changed to RMIN = 0.5 and OFAC = 0.8 due to convergence problems in the cluster calculations.