# Theoretical Study of the Tautomeric/Conformational Equilibrium of Aspartic Acid Zwitterions in Aqueous Solution

**Peter I. Nagy*,† and Béla Noszál‡**

*Department of Medicinal and Biological Chemistry, The University of Toledo, Toledo, Ohio 43606-3390, and Institute of Pharmaceutical Chemistry, Semmelweis University, H-1092 Budapest, Hungary*

*Received: February 17, 2000; In Final Form: May 2, 2000*

The present study is the first attempt at the theoretical calculation of the tautomeric/conformational equilibrium for a zwitterionic α-amino acid in aqueous solution. The gas-phase zwitterionic structure for aspartic acid (Asp) does not correspond to a local minimum on the potential energy surface. The optimized geometry was determined for the dihydrated zwitterion at the HF/6-31G* level. Single-point calculations for the isolated zwitterion, with the geometry optimized for the dihydrate, were performed up to the QCISD(T)/6-31G* and MP2/6-311++G** levels. Vibrational-frequency-dependent thermal corrections were calculated for the dihydrates. Relative hydration free energies of the isomeric species were calculated using the free energy perturbation method implemented in Monte Carlo simulations. In a second approach, relative free energies were calculated using the self-consistent isodensity polarizable continuum model (SCIPCM). In aqueous solution, aspartic acid forms two tautomeric zwitterions (α and $\beta$), with at least three (two N−C−C−C gauche and one trans) stable rotamers for each. Calculated by the ab initio/Monte Carlo method, the relative value of log $K$ for the deprotonation equilibria of the α and $\beta$ zwitterions is 1.07 ± 0.36, as compared to the experimental value of 1.26 at $T = 298$ K and $p = 1$ atm. The calculations predicted (with one exception) the rotamer preference of the six studied species in accord with a previous experiment. Relative in-solution free energies obtained with the SCIPCM and ab initio/Monte Carlo methods are similar at the Hartree−Fock level for the lowest-free-energy α and $\beta$ conformers. Furthermore, the SCIPCM method predicts the free energy separation of the $\beta$-gauche conformers in basic agreement with the experiment. Results suggest that consideration of the correlation energy is more important for comparisons of different tautomers of the Asp zwitterion than for considerations of conformers of a given tautomer.

## Introduction

Protonation states of organic acids and bases are known to be sensitive to the pH in aqueous solution. Biologically very important small molecules, typically amino acids, exist in which both acidic and basic functional groups are present in a single molecule.

An important problem is the structure of amino acids with zero net charge. There is general agreement that the simplest α-amino acids exist in the gas phase in the so-called neutral form. In contrast, the predominating species is generally agreed to be the zwitterion in neutral aqueous solution. The zwitterionic form of gas-phase glycine has been the subject of extended theoretical studies for several years. Upon ab initio HF/6-31G* geometry optimization, a zwitterionic structure was found for isolated glycine,[1] but detailed studies using larger basis sets[2] and the MP2 level of geometry optimization[3] indicated that the zwitterionic form of isolated glycine does not correspond to a stationary point on the gas-phase potential energy surface (PES). For the 1:1 glycine−water complex, Ding and Krogh-Jespersen[4] found that the zwitterionic form corresponded to a local energy minimum already at the HF/4-31G level and that the charge-separated form remained stable in calculations with larger basis sets including the HF/6-31G* level. Although the zwitterionic structure can exist in the 1:1 complex in the gas-phase, it is still less stable than the neutral glycine−water dimer. Jensen and Gordon pointed out, however, that the zwitterion and two water molecules give rise to a potential energy minimum.[1b]

A more complicated case is provided if the amino acid side chain also contains a protonating/deprotonating group. Recently, on the basis of B3LYP/6-31G* and MP2/6-31G* calculations, the zwitterionic gas-phase structure was suggested for arginine, with the proton is located at the guanidino group.[5] Because of the $C_3$ spacer in the side chain, a flexible ring system can be formed, which stabilizes the zwitterion with intramolecular hydrogen bond(s). However, infrared cavity ringdown laser absorption spectroscopy (IR-CRLAS) experiments on jet-cooled arginine did not find the zwitterionic form in the gas phase.[6] HF/6-311++G** geometry optimization for zwitterionic $\gamma$-amino butyric acid (GABA), also having a flexible aliphatic $-(CH_2)_3-$ spacer, led to a proton jump from the $NH_3^+$ group to the $COO^-$ group, and only the neutral form was found to be stable in the gas phase.[7] (See the following article.)

Amino acids have been extensively studied in aqueous solution as well. Most theoretical investigations have been performed for zwitterionic glycine,[8] including recent calculations by Tuñón et al.[9a] and Nagaoka et al.[9b] for proton transfer from the neutral form to the zwitterion. Conformational changes for the pentapeptide bis(penicillamine) enkephalin[10] and tryptophan[11] were followed in molecular dynamics simulations and for $\gamma$-amino butyric acid[7] and serine[12a] using the Monte Carlo method. The continuum dielectric approximation has been

---

* Author to whom correspondence should be addressed.

† The University of Toledo.

‡ Semmelweis University.

10.1021/jp0006227 CCC: $19.00 © 2000 American Chemical Society
Published on Web 07/01/2000

applied for studying the zwitterionic conformers of alanine,[12b] serine,[12b] and glutamine.[13] Calculations of Raman optical activity for L-alanine[12c] and vibrational spectra for glutamine[13] also utilized the continuum water approach. Infrared spectra and conformers of zwitterionic L-tryptophan were studied in a solid solvent of KBr.[14] Solvation free energy was determined using the RISM theory for the alanine dipeptide in a 1 M NaCl solution.[15] Rotamer populations in aqueous solution were analyzed experimentally using $^1$H NMR coupling constants for histidine,[16] phenylalanine,[16] aspartic acid,[16,17a] serine,[16,17b] cystein,[17b] and seleno-cystein.[17b]

The aim of the present study is to compare theoretical calculations with available experimental results for the conformational/tautomeric equilibrium of aspartic acid zwitterion in aqueous solution. Noszál and Sándor[17a] determined equilibrium constants for the differently protonated forms of aspartic acid (Asp) and asparagine in the pH range of 0.0−13.2 in aqueous solution. Aspartic acid forms a dianion at high pH and a monocation at low pH, thus existing in species of different stoichiometry. The net charge is zero both for the so-called neutral form and for its zwitterionic tautomers. Amino acids exist in the zwitterionic form in aqueous solution at the isoelectric pH. There are, however, two tautomeric zwitterionic structures for Asp, and each may have different rotamers. Even for studies of only the zwitterions, a complicated tautomer/conformer equilibrium must be considered.

Recently, several theoretical analyses have been performed in this laboratory to characterize the conformational/tautomeric equilibrium and solution structures of small, biologically important molecules.[18] Combined ab initio (gas-phase)/Monte Carlo (solution) calculations have been carried out for, among others, histamine,[19a] $\gamma$-hydroxy butyric acid,[19b] and dopamine.[19c] In all of these cases, the studied molecules, neutral or protonated, formed stable species in the gas-phase; thus, it was possible to obtain the internal free energy for an isolated system by performing ab initio calculations. Zwitterionic nicotonic acid isomers also turned out to form stable species in the gas phase.[20] Isolated zwitterionic amino acids are, however, not stable according to both experimental and theoretical results, as discussed above. Thus, the previously used combination of ab initio calculations for the isolated species and Monte Carlo simulations for the relative solvation free energies of the tautomers/conformers *with the gas-phase optimized geometry* is not possible for modeling aspartic acid in aqueous solution. In the present study, a new approach has been devised and tested against experimental results.

## Methods and Calculations

The total relative free energy, $\Delta G_{\text{tot}}$, of the Asp tautomers/conformers in aqueous solution is broken into two parts in the present approximation, as follows:

$$\Delta G_{\text{tot}} = \Delta G(\text{int}) + \Delta G(\text{sol}) \qquad (1)$$

where $\Delta G$(int) and $\Delta G$(sol) are the relative internal and solvation free energy contributions, respectively, to the total relative free energy.

To calculate the two terms, a reasonable reference structure must be selected. If the isolated species, generally called the gas-phase structure, is stable, then it can be conveniently taken as the reference species. This choice is, however, not mandatory. Because the zwitterionic form does not correspond to a local energy minimum on the PES for an isolated amino acid, the zwitterion + 2 water molecule system has been selected as the gas-phase reference structure in the present study. The minimum number of water molecules was chosen on the basis of the results of Jensen and Gordon.[1b] In the starting arrangements, the two water molecules were placed in positions from which favorable hydrogen bonds could be formed with the carboxylate oxygens. Consideration of additional water molecules hydrating the $NH_3^+$ group raises difficulties. Although the acceptance of three more water molecules around the $NH_3^+$ group, in its trans conformation with respect to the $\beta$-COO$^-$ group (Figure 1f), would be reasonable, at least one of the waters would be repelled from the first hydration sphere of the ionic sites in the gauche conformers with intramolecular N−H···O−C hydrogen bond(s). This (these) water molecule(s) would be in hydrogen bonds with other water molecule(s), thus putting the calculations on an unequal footing.

Also, on the basis of the results by Jensen and Gordon, the Asp·2$H_2O$ system, subsequently referred to as the dihydrate, may correspond to an energy minimum, enabling the calculation of vibrational frequencies and then zero-point energy and thermal corrections. Water molecules are expected to be strongly bound at the ionic sites of the zwitterion in aqueous solution. Optimization for the dihydrate could reflect some geometric changes for the zwitterion in the presence of nearby water molecules. Furthermore, the strong hydrogen bonds in the dihydrate will effect the vibrational frequencies for all three units. Consideration of this frequency shift in calculating internal free energy corrections for the reference system is a new feature of the present approach. Frequency-dependent thermal corrections would become more questionable if they were obtained for systems with more than two water molecules, including those that are hydrogen-bonded only to other water molecule(s) (see above).

Although the frequency-dependent corrections must be calculated for the dihydrate (similar involvement of all three molecules in the low-frequency normal modes prevented the separate assignment of many modes either to the zwitterion or to the water molecules), single-point energies have been calculated for the isolated zwitterion with the geometry optimized in the dihydrate. Relative energies of the different dihydrate structures contain a fraction of the total relative solvation energy in solution (see next section). Solvation of the dihydrate as a rigid unit would overemphasize the role of the ordered, minimum-energy arrangement of the zwitterion and two close water molecules. Thus, the approach was applied that normal frequencies and the thermal corrections depending on them were calculated for the optimized dihydrate, whereas the quantum chemical internal energy was calculated for the isolated zwitterion. The relative free energy of solvation in Monte Carlo simulations was determined also for the tautomers/conformers of the zwitterion with its geometry optimized in the dihydrate.

Ab initio geometry optimizations for the dihydrated zwitterion were performed at the HF/6-31G* level,[21a] utilizing the Gaussian 98 package[22] running on a T90 computer at the Ohio Supercomputer Center. Characteristic geometric parameters for the zwitterion are summarized in Table 1a. Single-point calculations for the zwitterion tautomers/conformers, at geometries optimized in the dihydrates, were performed at the B3LYP/6-31G*,[23] MP2/6-31G*,[21a] and QCISD(T)/6-31G*[24] levels. When the 6-311++G** basis set was used,[21a] calculations were carried out at the HF, B3LYP, and MP2 levels (Table 2). Zero-point energy and thermal corrections (in the rigid rotator−harmonic oscillator approximation[25]) were obtained using normal frequencies calculated at the HF/6-31G* level for the local-energy-minimum structures of dihydrates (Table 3). One trans and two gauche

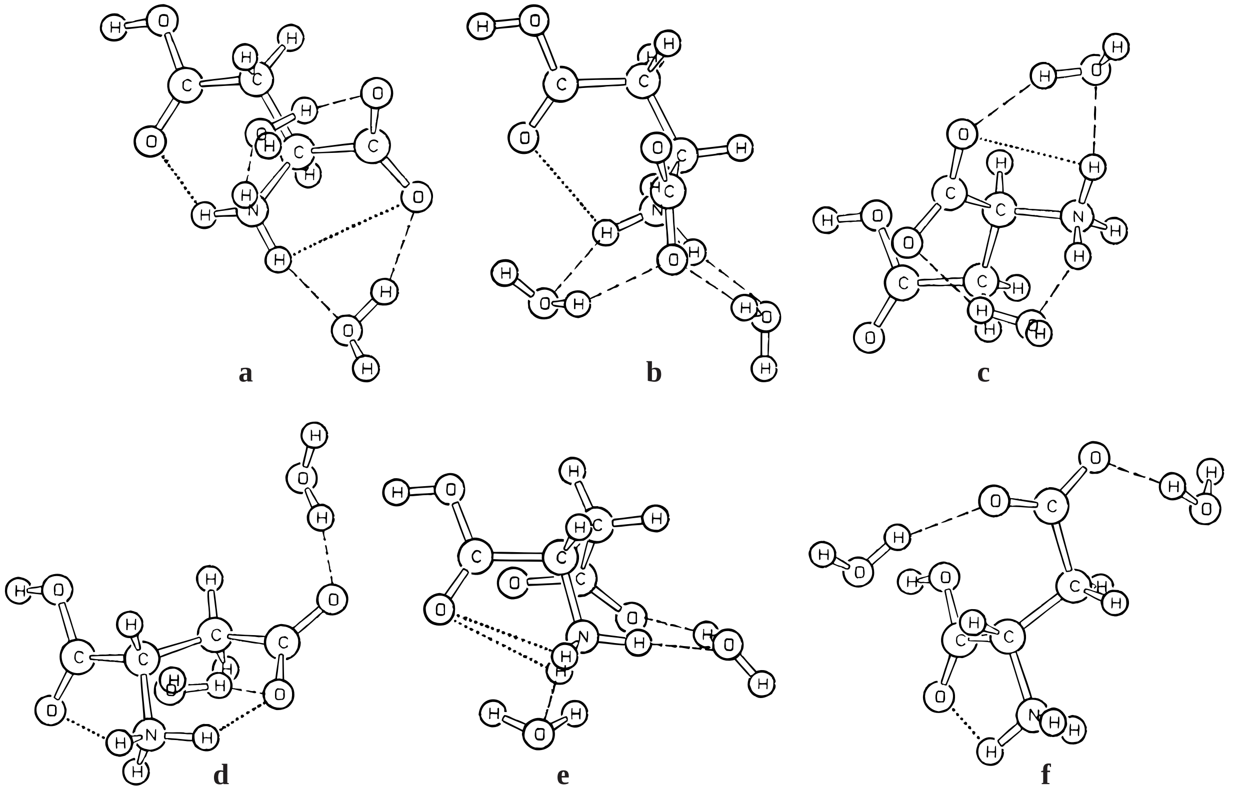


**Figure 1.** (a) HF/6-31G* optimized geometry for the aspartic acid dihydrate with α-$COO^-$. The conformation is α-gauche 1, based on the N$-$C$_\alpha-$C$_\beta-$C arrangement. Letters α or $\beta$ for Figure 1a$-$f indicate the location of the carboxylate group on atoms C$_\alpha$ or C$_\beta$, respectively. Intramolecular hydrogen bonds are indicated by dotted lines (for values, see Table 1a). In the intermolecular O$-$H bonds (dashed line), the first and second letters refer to the water atom and to the zwitterion, respectively. Hydrogen bond distances are in angstroms. Upper water: H$-$O = 1.88, O$-$H = 1.91. Lower water: H$-$O = 1.83, O$-$H = 1.85. (b) Optimized α-gauche 2 dihydrate. Hydrogen bonds for left-side water: H$-$O = 2.01, O$-$H = 1.89. Hydrogen bonds for right-side water: H$-$O = 1.87, O$-$H = 1.86. (c) Optimized α-trans dihydrate. Hydrogen bonds for upper water: H$-$O = 1.83, O$-$H = 1.83. Hydrogen bonds for lower water: H$-$O = 1.96, O$-$H = 1.91. (d) Optimized $\beta$-gauche 1 dihydrate. Hydrogen bond for upper water: H$-$O = 1.93. Hydrogen bond for lower water: H$-$O = 1.98. (e) Optimized $\beta$-gauche 2 dihydrate. Hydrogen bond for left-side water: O$-$H = 1.89. Hydrogen bonds for right-side water: O$-$H = 1.85, H$-$O = 1.82. (f) Optimized $\beta$-trans dihydrate. Hydrogen bond for left-side water: H$-$O = 1.87. (There is no hydrogen bond to the seemingly close α-COOH. The O(w)···HO distance is 3.59 Å.) . Hydrogen bond for right-side water: H$-$O = 1.88.

conformations of the zwitterion were localized for each of the α and $\beta$ tautomers (Figure 1, Table 2). A scaling factor of 0.9 was applied in calculating the relative zero-point energy (Δ*Z*PE) term.[21a] Scott and Radom[21b] found optimal scaling factors in the range of 0.895$-$0.905 for vibrational frequencies obtained at the HF level with basis sets increasing from 3-21G up to 6-311G(df,p). HF/6-31G* optimization for the isolated zwitterion resulted in local energy minima for the conformers of the α-tautomer. The calculations always led to the neutral form starting from the $\beta$-tautomer. The geometric changes and energy results obtained for the α structure will be discussed in relation to Table 4a.

Monte Carlo simulations[26] for the aqueous solution of the Asp zwitterion were performed by using the BOSS 3.6 software.[27] The program runs on a Silicon Graphics Indigo 2 workstation at the University of Toledo. Calculations were performed in the *NpT* (isobaric$-$isothermal ensembles) at $T =$ 298 and $p = 1$ atm. The 24 × 24 × 24 Å$^3$ periodic box contained one solute and 502 TIP4P water molecules.[28] Preferential sampling proportional to $1/(R^2 + c)$ was applied, where the value of the constant $c$ was set to 120. $R$ is the distance between the C$_\alpha$ and the oxygen atom of the water molecule to be displaced in the generation of a new configuration. The solvent$-$solvent and solute$-$solvent cutoff radii were set to 8.5 and 12 Å, respectively. The ICUT = 2 option was used in calculating the solute$-$solvent interactions. This option allows for a solvent molecule to be seen by the solute, if any distances between an explicit solute atom and the water oxygen are not larger than 12 Å. In the equilibration and averaging phases, 3500 and 5000 K configurations, respectively, were generated. A solute move was attempted every 50 steps, and volume changes were allowed in every 1000 steps.

The united-atom model was used for the CH and $CH_2$ groups of Asp, while all other atoms were considered explicitly. Steric parameters of the 12-6-1 OPLS potential function[29] were taken from the program library. Atomic charges for the zwitterion were derived by the CHELPG procedure[30] upon fitting charges to the HF/6-31G* electrostatic potential of the *isolated* zwitterion. Relative solvation free energies were calculated using the free-energy perturbation method, as implemented in Monte Carlo simulations.[31] Geometric and OPLS potential parameters were calculated using a linear function depending on the perturbation step parameter, $\lambda$

$$\mathrm{par}(\lambda) = \lambda\,\mathrm{par(a)} + (1-\lambda)\,\mathrm{par(b)} \qquad (2)$$

where the system is transformed from state a ($\lambda = 1$) to state b ($\lambda = 0$). Double-wide sampling was applied with changes for $\lambda$ in the 0.015$-$0.05 range. Conformers of a given tautomer were transformed along real physical pathways, meaning changes were basically made in the torsion angles. Transformation of the α(trans) structure to $\beta$(trans) took place along a nonphysical pathway by developing and annihilating a hydrogen atom on the carboxylate and the neutral carboxylic group, respectively.

**TABLE 1: Optimized Geometric Parameters for the Asp Zwitterions**[a]

(a) Optimized Geometric Parameters for the Gas-Phase Asp-Zwitterion Dihydrates

| | α-COO$^-$ | | | $\beta$-COO$^-$ | | |
|---|---|---|---|---|---|---|
| | g1 | g2 | t | g1 | g2 | t |
| $NC_\alpha C_\beta C_{carb}$ | −59.6 | 66.0 | 175.8 | −55.6 | 51.5 | −155.3 |
| $C_{carb}C_\alpha C_\beta C_{carb}$ | −178.9 | −57.4 | 58.4 | −175.6 | −68.8 | 85.9 |
| $O{=}C_{carb}C_\alpha C_\beta$ | | | | 107.2 | 130.5 | 134.9 |
| $^-O{-}C_{carb}C_\alpha C_\beta$ | 21.9<br>−161.4 | −25.9<br>154.2 | 8.3<br>−175.6 | | | |
| $C_\alpha C_\beta C_{carb}{=}O$ | 23.0 | −28.5 | −153.6 | | | |
| $C_\alpha C_\beta C_{carb}$-$O^-$ | | | | 32.0<br>−149.3 | 65.9<br>−105.7 | 22.2<br>−160.5 |
| $C{-}C_{carb}OH$ | −177.7 | 179.9 | −177.8 | 176.6 | 177.4 | −175.0 |
| (N)H···$O_\beta$ | 2.05 | 2.41 | | 1.58 | 2.79<br>2.83 | |
| (N)H···$O_\alpha$ | 2.72 | 2.52 | 2.57 | 2.26 | 2.55<br>2.56 | 2.19 |

(b) Optimized Geometric Parameters for the Asp Zwitterion in Aqueous Solution

| | α-COO$^-$ | $\beta$-COO$^-$ | |
|---|---|---|---|
| | g1 | g1 | g2 |
| $NC_\alpha C_\beta C_{carb}$ | −65.0 | −57.1 | 60.2 |
| $C_{carb}C_\alpha C_\beta C_{carb}$ | 171.9 | −178.2 | −60.9 |
| $O{=}C_{carb}C_\alpha C_\beta$ | | 111.7 | 138.5 |
| $^-O{-}C_{carb}C_\alpha C_\beta$ | −50.5<br>130.9 | | |
| $C_\alpha C_\beta C_{carb}{=}O$ | 6.5 | | |
| $C_\alpha C_\beta C_{carb}$-$O^-$ | | 11.4<br>−169.9 | 0.3<br>−179.4 |
| $C{-}C_{carb}OH$ | −178.6 | 178.3 | −177.3 |
| (N)H···$O_\beta$ | 2.30 | 1.90 | 2.08 |
| (N)H···$O_\alpha$ | 2.37 | 2.55 | 2.42 |

[a] Torsion angles in degrees. (N)H···O distances in angstroms indicate the short(est) H···O distance(s) *within* the aspartate zwitterion at the optimized geometry of the dihydrate. Subscripts α and β refer to the corresponding −COO(H) group containing the considered O atom.

**TABLE 2: Relative Energies of the Asp-zwitterion with Geometry Optimized in the Dihydrate**[a]

| | 6-31G* | | | | 6-311++G** | | |
|---|---|---|---|---|---|---|---|
| | HF | B3LYP | MP2 | QCISD(T) | HF | B3LYP | MP2 |
| | | | | α-COO$^-$ | | | |
| g1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| g2 | −2.22 | −0.90 | −2.27 | -2.24 | −1.41 | −0.32 | −1.77 |
| t | 8.18 | 8.67 | 7.40 | 7.56 | 8.30 | 8.47 | 6.82 |
| | | | | $\beta$-COO$^-$ | | | |
| g1 | −3.04 | −6.62 | −5.29 | −4.64 | −2.61 | −5.35 | −5.14 |
| g2 | 12.20 | 13.10 | 11.10 | 11.46 | 11.71 | 12.06 | 9.49 |
| t | 23.39 | 23.29 | 24.10 | 24.16 | 22.37 | 22.03 | 22.23 |

[a] Energies are in kilocalories per mole. Geometries were optimized at the HF/6-31G* level.

**TABLE 3: Relative Zero-Point Energies and Thermal Corrections**[a]

| | 0.9Δ*Z*PE | Δ*H*(*T*) | *T*Δ*S*(*T*) | *T*Δ$S_{vibr}$(*T*) | Δ$G_{th}$(*T*) |
|---|---|---|---|---|---|
| | | | gas phase[b] | | |
| | | | α-COO$^-$ | | |
| g1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| g2 | 0.02 | −0.07 | −0.61 | −0.50 | 0.56 |
| t | −0.18 | 0.03 | −0.34 | −0.28 | 0.19 |
| | | | $\beta$-COO$^-$ | | |
| g1 | −1.48 | 0.51 | 1.45 | 1.37 | −2.41 |
| g2 | 0.44 | −0.37 | −1.83 | −1.65 | 1.90 |
| t | −1.43 | 0.88 | 2.73 | 2.74 | −3.29 |
| | | | aqueous solution[c] | | |
| | | | α-COO$^-$ | | |
| g1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | | $\beta$-COO$^-$ | | |
| g1 | 0.09 | −0.18 | −0.56 | −0.55 | 0.46 |
| t | 0.10 | −0.16 | −0.28 | −0.18 | 0.21 |

[a] Energies are in kilocalories per mole. Geometries optimized at the HF/6-31G* level. $\Delta G(\text{thermal}) = \Delta G_{th}(T) = 0.9\Delta ZPE + \Delta H(T) - T\Delta S(T)$ at $T = 298$ K. [b] Calculated for the Asp-zwitterion dihydrate in the gas phase. [c] Calculated for the pure Asp zwitterion in aqueous solution.

When this nonphysical transformation was performed, eq 2 was still applied. A summary of the solvation results is provided in Table 4a. The number of hydrogen bonds, derived by integration of the solute−solvent pair-energy distribution functions, and coordination numbers, obtained by integration of the radial distribution functions, are summarized in Table 5.

In a second approach, the relative free energies were calculated using the self-consistent isodensity polarizable continuum model[32] (SCIPCM) at the HF/6-31G* and HF/6-311++G**//HF/6-31G* levels. Because of slow convergence in the geometry optimization, the method was applied only to three zwitterionic species in aqueous solution. The isodensity contour parameter was set to 0.001 au, and 512 grid points were considered for the surface charges. The solvent dielectric constant was set to 78.3. The geometry was optimized at the HF/6-31G* level, placing the zwitterionic solute in a cavity formed within the water solvent. Starting Asp geometries were taken from the α-g1, β-g1, and β-g2 gas-phase dihydrates (Figure 1a, d, and e). Numerical frequency analysis was performed for the optimized structures in aqueous solution, and the relative zero-point energies and thermal corrections were evaluated. The results are summarized in Tables 1b, 3, and 4b.

## Results and Discussion

**Geometry.** Two gauche and one trans arrangements of the $NC_\alpha C_\beta C_{carb}$ moiety (see Scheme 1; g1, g2, and t rotamers designated by t, h, g, respectively, in ref 17a) have been

**TABLE 4: Relative Energies and Free Energies in Aqueous Solution for the Zwitterionic Asp Isomers**

(a) Relative Energies and Free Energies in Aqueous Solution for the Zwitterionic Asp Isomers with Geometry Optimized in the Gas-Phase Dihydrate[a]

| | $\Delta E$(A+2W, HF) | $\Delta G$(sol) | $\Delta E$(A, HF) + $\Delta G$(sol)[b] | $\Delta E$(A, MP2) + $\Delta G$(sol)[b] | $\Delta G_{\mathrm{tot}}$[b,c] |
|---|---|---|---|---|---|
| | | | α-COO⁻ | | |
| g1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| g2 | 1.43 | 4.05 ± 0.33 | 1.83 | 2.28 | 2.84 |
| t | 7.18 | −1.56 ± 0.41 | 6.62 | 5.26 | 5.45 |
| | | | β-COO⁻ | | |
| g1 | 10.22 | 8.87 ± 0.57 | 5.83 | 3.73 | 1.32 |
| g2 | 5.85 | −5.01 ± 0.67 | 7.19 | 4.48 | 6.38 |
| t | 34.37 | −16.55 ± 0.45 | 6.84 | 5.68 | 2.39 |

(b) Relative Free Energies for the Zwitterionic Asp Isomers in Aqueous Solution[d]

| | $\Delta G_{\mathrm{tot}}$/HF/6-31G* | | $\Delta G_{\mathrm{tot}}$/HF/6-311++G** | |
|---|---|---|---|---|
| | gas phase[e] | solution[f] | gas phase[e] | solution[f] |
| | | α-COO⁻ | | |
| g1 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | β-COO⁻ | | |
| g1 | 3.42 | 3.78 | 3.85 | 3.50 |
| g2 | 9.09 | 3.87 | 8.60 | 4.37 |

[a] Energies are in kilocalories per mole. $\Delta G$(sol) stands for the relative solvation free energy in aqueous solution from Monte Carlo simulations. $\Delta E$(A+2W, HF) shows the relative energies of the optimized Asp zwitterion + 2 water system calculated at the HF/6-31G* level. The $\Delta E$(A, HF) and $\Delta E$(A, MP2) terms (see Table 2) stand for the relative internal energies of the Asp zwitterion (without waters) at the HF/6-31G* and MP2/6-311++G** levels, respectively. [b] Terms have the same standard deviation as indicated for $\Delta G$(sol). [c] $\Delta G_{\mathrm{tot}} = \Delta E(\mathrm{A, MP2}) + \Delta G(\mathrm{sol}) + \Delta G_{\mathrm{th}}(T)$ at $T = 298$ K, $p = 1$ atm. Corresponding values are taken from Tables 2 and 3. [d] $\Delta G_{\mathrm{tot}}$ in kilocalories per mole. $\Delta G_{\mathrm{th}}(T)$ terms (included in $\Delta G_{\mathrm{tot}}$) are taken from Table 3. [e] Zwitterionic Asp geometry optimized in the gas-phase dihydrate. [f] Zwitterionic Asp geometry optimized in aqueous solution by using the SCIPCM method.

**TABLE 5: Number of Hydrogen Bonds and Coordination Numbers for the Asp Zwitterion in Aqueous Solution**

| | $N_{\mathrm{HB}}$ ($E \leq E_{\mathrm{ul}}$)[a] | N/$O_w$ | (N)H/$O_w$ | =O/$H_w$ | −O/$H_w$ | $-O^-/H_w$ | H/$O_w$ | ΣH/O |
|---|---|---|---|---|---|---|---|---|
| | | | | α-COO⁻ | | | | |
| g1 | 8.8 (−5) | 3.6 | 3.0 | 1.1 | undef[b] | 4.9 | 1.0 | 10.0 |
| g2 | 9.4 (−5) | 5.0 | 3.6 | 1.0 | 0.4 | 4.0 | 1.0 | 10.0 |
| t | 9.7 (−5) | 3.1 | 3.0 | 1.2 | 0.3 | 4.5 | 1.0 | 10.0 |
| | | | | β-COO⁻ | | | | |
| g1 | 7.5 (−5) | 3.0 | 2.0 | 1.4 | 0.2 | 4.1 | 1.0 | 8.7 |
| g2 | 8.6 (−6) | 3.5 | 2.8 | 1.4 | 0.7 | 4.9 | 1.0 | 10.8 |
| t | 10.0 (−6) | 3.8 | 3.1 | 1.0 | undef[b] | 5.3 | 1.0 | 10.4 |

[a] $E_{\mathrm{ul}}$ values in parentheses stand for the upper limits in integration of the solute−solvent pair-energy distribution function. [b] The radial distribution function does not show a minimum below 2.5 Å, preventing the calculation of the coordination number by integration of the rdf until its first minimum.

investigated in gas-phase dihydrates for both tautomers (Figure 1a−f). The $NH_3^+$ group can form two different hydrogen bonds with the oxygens of the neutral β-COOH. The energy of the α-g1 dihydrate was higher by 4.25 kcal/mol when the N−H···OH bond was formed as compared to the structure with the N−H···O=C intramolecular bond. The corresponding value for the optimized isolated α-zwitterion (as mentioned previously, the zwitterionic structures turned out to correspond to local energy minima for the α tautomer) is 4.99 kcal/mol. Thus, for further consideration, the α-g1 structure with the N−H···O=C hydrogen bond was selected (Figure 1a).

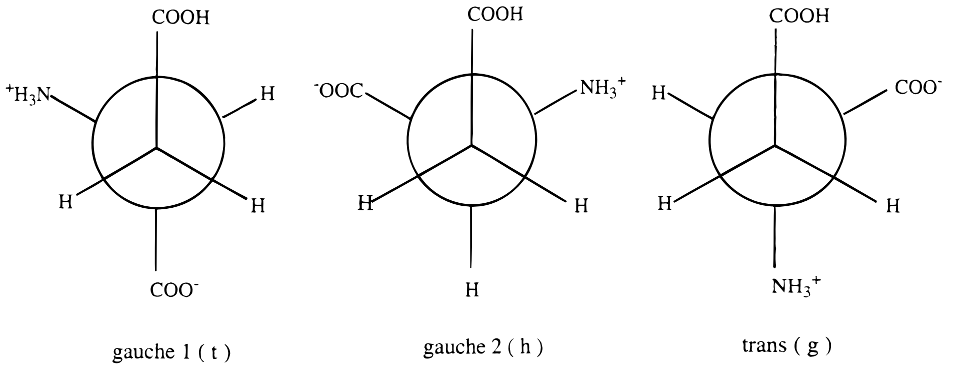
Table 1a shows nearly classical torsion angles for the gauche and trans arrangements of the $NC_\alpha C_\beta C_{\mathrm{carb}}$ moiety in the α-tautomer. This suggests that the strong hydrogen bond between the $NH_3^+$ and α-COO⁻ groups and the coexisting NH···O=C hydrogen bond with the neutral carboxyl have only a moderate effect on the $NC_\alpha C_\beta C_{\mathrm{carb}}$ torsion angle. In contrast, the torsion angle deviates by 25° from the classical 180° value for the β-t structure (Figure 1f), indicating a strong electrostatic interaction between the remote ionic sites.

The difference of the two $^-O{-}C_{\mathrm{carb}}C_\alpha C_\beta$ torsion angles is about 180° for each α-zwitterion conformer, indicating a nearly planar α-carboxylate group. The difference is only 172° for the $C_\alpha C_\beta C_{\mathrm{carb}}{-}O^-$ torsion angles with the β-g2 structure (Figure 1e). Although the shortest (N)H···$O_\beta$ separation is about 2.8 Å, thus being larger than a hydrogen-bond distance, deviation from planarity for β-carboxylate suggests a nonnegligible interaction of the ionic sites.

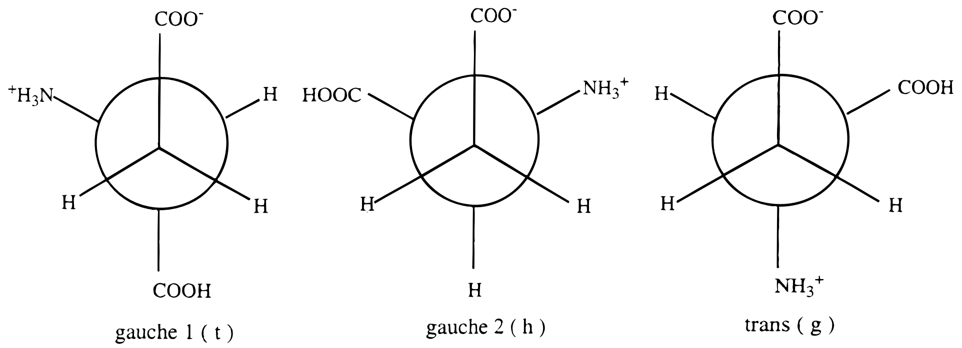
The intramolecular (N)H···$O_\beta$ distance is conspicuously short, at 1.58 Å for the β-g1 isomer (Figure 1d). This indicates the favorable orientation of the β-carboxylate group with respect to the $NH_3^+$ group because of the flexibility of the $NC_\alpha C_\beta C_{\mathrm{carb}}$ chain. No such flexible spacer is available for the α-carboxylate group with a fairly constant (N)H···$O_\alpha$ separation in the 2.5−2.7 Å range.

Not indicated in Table 1a, but worth mentioning, is the change in the $^-O{-}C_{\mathrm{carb}}C_\alpha C_\beta$ torsion angle for the α-zwitterion upon optimization without the two water molecules. For this tautomer, the torsion angles are −39°, −30°, and −43° for the g1, g2, and t rotamers, respectively. The deviation is 61° for g1, but is

**SCHEME 1**



alpha - zwitterionic conformers



beta - zwitterionic conformers

only 4° for g2. This finding will be discussed more in relation to the relative solvation free energies.

As far as the intermolecular interactions in dihydrates are concerned, Figure 1a−f shows that the water molecules surround one or both ionic sites without forming water−water hydrogen bonds. The shortest O(w)···O(w) distance is about 4 Å. In Figure 1a, b, and c, the water molecules are on different sides of the plane defined by the N−$C_\alpha$−C(α-carboxyl) atoms. Each water molecule is involved in two hydrogen bonds: it is an acceptor in the bond to the $NH_3^+$ group and a donor to the carboxylate. The intermolecular O···H distances are in the range of 1.8−2.0 Å. In Figure 1a and c, both C−O and two N−H bonds are involved in hydrogen bonds with the water molecules. The third N−H bond is in an intramolecular bond to the $\beta$-carbonyl oxygen in Figure 1a (see Table 1a), but remains unbound in Figure 1c. In the α-t conformer, the N−H bond points away from the $\beta$-COOH. For α-g2 (Figure 1b), two different N−H bonds, but the same C−O bond, are involved in the four hydrogen bonds to the water molecules. (The carboxylate oxygen in question also forms the NH···$O_\alpha$ = 2.52 Å intramolecular hydrogen bond with the proton partially covered in Figure 1b.)

For the α-carboxylate conformers, the distance of the ionic sites is not strongly affected upon rotation of the $\beta$-COOH group. The water molecules can form a “nest” with the ionic sites and can easily form two hydrogen bonds to each of them. Rotation of the ionic $\beta$-carboxylic group (thus the $\beta$-$COO^-$ system) results in largely different (N)H···$O_\beta$ distances. Upon formation of the very short intramolecular bond in the $\beta$-g1 isomer, water molecules are repelled out of this part of space and form hydrogen bonds only with the $\beta$-$COO^-$ group (Figure 1d). The structure also allows the formation of the intramolecular (N)H···$O_\alpha$ bond with the neutral O═C group of the α-COOH (see Table 1a). In contrast, there are two N−H···O(w) hydrogen bonds in Figure 1e, and there is only one bond to the carboxylate group. This structure may not be the global minimum for the $\beta$-g2 dihydrate, a result that would partially explain the calculated high relative free energy for this species. In the $\beta$-t isomer (Figure 1f), the water molecules follow the $-COO^-$ group, and each of them forms one hydrogen bond to that group. Because the protonated amine site remains without stabilizing hydrogen bond in this structure, the relative energy is the highest for this conformer, considering either the dihydrate or the isolated zwitterion in the single-point energy calculations.

The optimized geometric parameters obtained with the continuum solvent SCIPCM method are collected in Table 1b. Whereas changes in the $NC_\alpha C_\beta C_{carb}$ and $C_{carb}C_\alpha C_\beta C_{carb}$ torsional angles are moderate, up to 10°, much larger deviations have been found for the torsional angles defining the position of the −COO(H) planes. For the α-g1 isomer, the $^-O-C_{carb}C_\alpha C_\beta$ torsional angles change from 21.9° and −161.4° to −50.5° and 130.9°, respectively (Table 1a and b). Thus, the basic planarity of the carboxylate group is conserved in aqueous solution, but the plane of the group is rotated by about 70°. This rotation, however, should not be considered as being entirely due to solvent effects. As mentioned above, the HF/6-31G* geometry optimization for the isolated α-g1 isomer produced a $^-O-C_{carb}C_\alpha C_\beta$ torsional angle of −39°; thus the solvent effect for this species seems to be only about 11°. The remarkable difference in torsional angles obtained for the gas-phase dihydrate, on one hand, and the gas-phase and in-solution aspartic acid α-g1 species on the other hand should be attributed to the explicit consideration of two close and strongly bound water molecules. This difference raises the question about the geometry to be considered as relevant for in-solution calculations.

The change of the $C_\alpha C_\beta C_{carb}-O^-$ torsional angle is about 20° for the $\beta$-g1 tautomer/conformer, but is large, about 65°, for the $\beta$-g2 species. In the absence of gas-phase results for the isolated zwitterionic $\beta$-tautomers, the solvent effect on the equilibrium geometries cannot be estimated for these structures. A possible reason for the large geometric change in $\beta$-g2 is that the two water molecules close to the $\beta$-g2 carboxylate largely affect the equilibrium geometry and result in the structure shown in Figure 1e. On the basis of the energy results, however, it is more likely that the dihydrate structure in Figure 1e is not the global minimum for this tautomer/conformer.

Disregarding the two water molecules and considering the continuum solvent environment around the cavity also lead to changes in the intramolecular NH$\cdots$O distances. The considerably different NH$\cdots$O$_\alpha$ and NH$\cdots$O$_\beta$ values for $\alpha$-g1 in the gas phase become almost equal, 2.30−2.37 Å, after the SCIPCM optimization. The short intramolecular hydrogen bond of 1.58 Å extends to 1.90 Å for $\beta$-g1, and one of the two, long (N)H$\cdots$O$_\beta$ bonds for $\beta$-g2 becomes a regular hydrogen bond of 2.08 Å. The ionic sites with solute geometries given in Table 1a and b are exposed to hydration to a different extent, and these differences may be reflected in the tautomer/conformer stabilities calculated with the ab initio/Monte Carlo and the SCIPCM methods.

**Internal Energy and Free Energy.** Results of the single-point energy calculations for all zwitterionic species with geometries optimized in the dihydrates at the HF/6-31G* level are summarized in Table 2. Because the zwitterion contains a negatively charged carboxylate site, some concern may arise regarding the geometry optimization performed with a basis set without diffuse functions. The zwitterion is a zero-net-charge system, and this may be the reason Ding and Krogh-Jespersen found only small changes in the optimized geometric parameters for the 1:1 glycine−water system when the basis set was upgraded from HF/4-31G to HF/6-31G++G**.[4] In our previous study for the zwitterionic pyridine carboxylic acids,[20] the optimized geometry obtained with the HF/6-311++G** basis set showed only minor changes as compared to the result of the HF/6-31G* optimization.

Table 2 shows that the MP2/6-31G* and QCISD(T)/6-31G* single-point relative energies are fairly close. Both the HF and B3LYP level calculations with the 6-31G* basis set show remarkable deviations from them mainly when considering the gauche rotamers. When the 6-311++G** basis set is used, the relative energies generally decrease in absolute value. This is a trend that was already observed in a detailed analysis for the pyridine carboxylic acid[20] and was also found for the glycine-water dimer.[4] Because of the similarity of the MP2 and high-level QCISD(T) results when the 6-31G* basis set is used, the MP2/6-311++G**//HF/6-31G* relative energies were considered in the final determination of the relative internal free energies.

The calculations with different basis sets and levels are in general qualitative agreement. They find, unanimously, that the $\beta$-g1 structure with its strong internal hydrogen bond is the most favorable species of the six isomers considered. The second-lowest-energy structure is $\alpha$-g2, followed by $\alpha$-g1. The calculations, unanimously again, predict high relative energy for the trans rotamers mainly with the $\beta$-zwitterion.

Relative frequency-dependent corrections for the gas-phase isomers (Table 3) are of moderate magnitude for the conformers of the $\alpha$-tautomer. Both the $\Delta Z$PE and $\Delta H(T)$ values are almost negligible for the $\alpha$-zwitterion. The nonnegligible $T\Delta S(T)$ total entropy-term is mainly due to changes in its vibrational entropy component, $T\Delta S_{\mathrm{vibr}}(T)$. Because the translational entropy is constant for isomeric systems, nearly equal values of $T\Delta S(T)$ and $T\Delta S_{\mathrm{vibr}}(T)$ allow only small changes in the rotational entropy. Vibrational entropy changes are expected mainly because of changes in the lowest-energy torsional vibrations of the zwitterion and because of different intermolecular vibrational frequencies. The resulting largest thermal correction, $\Delta G_{\mathrm{th}}(T)$ is, however, no more than 0.56 kcal/mol, which suggests a fairly constant structure for the dihydrate with only a small effect due to rotation of $\beta$-COOH. Although there are a number of “soft” vibrational modes in the dihydrate and estimation of the absolute entropy contribution based on harmonic frequencies[33] may lead to erroneous value for the present system, relative contributions are expected to give more reliable values.

Correction terms, in contrast to those for the $\alpha$-tautomer, are unusally large for the $\beta$-zwitterion. The $\Delta Z$PE term is remarkably decreased for g1 and t, indicating an overall red shift of the vibrations. The lower frequencies allow easier thermal excitation of the vibrations, $\Delta H(T)$ is positive, and the system becomes less ordered, as concluded from the remarkable entropy increase due mainly to the vibrational entropy. The $\Delta Z$PE term is positive for g2, indicating the opposite tendency and leading to changes of signs for all other terms. The final $\Delta G_{\mathrm{th}}(T)$ value is on the order of some relative isomer energies. It indicates that consideration of the vibration-related energy corrections may be very important. In fact, calculation of the total relative free energy without considering these terms would not give results in reasonable agreement with some experimental values (see next section).

Frequency-dependent corrections for the in-solution optimized $\beta$-g1 and $\beta$-g2 structures show basic deviations from their gas-phase values (Table 3). It is less surprising for the $\beta$-g2 species, for which the optimized gas-phase dihydrate structure, as was mentioned above, may not correspond to the most favorable arrangement of the Asp$\cdot$2H$_2$O system. The deviation is, however, noteworthy in the case of the $\beta$-g1 species. All terms are of opposite sign in the two sets of calculations. Each individual term is considerable for the gas-phase system, and together they result in a total contribution of −2.41 kcal/mol to the relative free energy. In contrast, the small correction terms have an overall positive contribution of 0.46 kcal/mol to the relative free energy in the continuum solvent optimization. This difference emphasizes again the possible importance of the explicit consideration of the strongly bound water molecules.

**In-Solution Euilibrium.** $\Delta G$(sol) values in Table 4a show the relative solvation free energies obtained in the perturbation calculations within the Monte Carlo simulations. Rotation of the neutral carboxylic group affects the hydration free energies moderately for the $\alpha$-conformers. Both gauche rotamers form intramolecular hydrogen bonds; thus, the difference of 4 kcal/mol indicates different appreciation of these bonds by the solvent environment. The $\Delta G$(sol) value is reasonably negative for the $\alpha$-t rotamer. This means that the trans structure without an intramolecular hydrogen bond, and thus with two more opened sites for forming hydrogen bonds with the solvent, is preferably hydrated in comparison with the internally bound $\alpha$-g1 rotamer.

The differences in $\Delta G$(sol) values are, however, large for the $\beta$-zwitterion rotamers, and the trend is opposite compared to that of the internal terms. The most favorable hydration was calculated for the $\beta$-t rotamer. This is not surprising, because the NH$_3^+$ and COO$^-$ groups are farthest from each other in this tautomer/rotamer. The $\beta$-g2 structure is hydrated about 14 kcal/mol more favorably than the $\beta$-g1 rotamer. As was discussed in the previous section, the hydrogen bond in $\beta$-g1 is very short; thus, water molecules have no chance to hydrate the ionic sites in the N−H$\cdots$O region. The region is more opened for $\beta$-g2 (see distances in Table 1a), and an even slightly better hydration potential for a zwitterion leads to a remarkably more negative solvation free energy.

Table 4a provides relative energies for the optimized dihydrates, as well. The $\Delta E$(A+2W, HF) values are the results from the HF/6−31-G* optimizations. The third column shows the sum of the relative internal energies of the isolated zwitterion, $\Delta E$(A, HF), calculated at the HF/6-31G* level, and the $\Delta G$(sol) solvation term. Values in the first and third columns are in surprisingly good agreement for the α-zwitterion. A tempting interpretation could be that most of the solvent effects are due to the difference in hydration by the closest solvent environment, basically modeled by the dihydrate. The similarity of the values is not maintained, however, for the $\beta$-gauche rotamers, and it breaks down dramatically for the $\beta$-t structure. This latter result is least surprising, as interaction with two close water molecules cannot represent the solvent effect for two separated ionic sites.

The $\Delta G$(sol) values contain corrections for the long-range electrostatic interactions. Because of the large solute−solvent cutoff radius of 12 Å taken around every explicit atom (including the united CH and $CH_2$ atoms) in the Monte Carlo simulations, the volume seen by the solute corresponds approximately to a sphere of $R = 14$ Å. Corrections are to be calculated for the electrostatic interaction of the solute and the solvent out of this sphere. The solute was considered as a dipole in the center of the cavity, and the Onsager correction[34] was applied. This approach seems to be crude even for a sphere with $R = 14$ Å, provided that the dipole moments are in the range of 11−15 D. The differences, however, may be much more realistic, amounting to up to 0.3 kcal/mol. The generalized Born formula[35] has been successfully applied recently for a set of atomic charges representing cationic dopamine conformers, and it showed only small deviations from the more precise calculations performed with the PCM continuum method.[36] For overall zero-net-charge systems the Onsager approach may work as well.

The $\Delta G_{\mathrm{tot}}$ values are the final results of our free energy calculations. The relative free energy values in solution are completely different from the calculated internal energies (Table 2, last column). The deviations are basically due to the $\Delta G$(sol) terms, but nonnegligible contributions are from the $\Delta G_{\mathrm{th}}(T)$ corrections. The final results suggest that the α-g1 is the lowest-free-energy species in solution, followed by the $\beta$-g1 isomer.

Noszál and Sándor[17a] determined the protonation microconstants for the aspartic acid α and $\beta$ zwitterions. Because the present study provides with relative free energies, only the difference of the equilibrium constants can be estimated. The log $K$ values for the $\mathrm{HAsp}^- + \mathrm{H}^+ \rightleftharpoons$ Asp (zwitterion) equilibrium are 3.63 and 2.37, for the α and $\beta$ zwitterions, respectively. Because

$$\log K(\alpha) - \log K(\beta) = -[G(\alpha) - G(\beta)]/2.3RT \qquad (3)$$

the difference can be theoretically calculated at $T = 298$ and $p = 1$atm. On the basis of the populations calculated for the six species of the α and $\beta$ zwitterions, the $-[\mathrm{G}(\alpha) - \mathrm{G}(\beta)]$ free energy difference is 1.45 ± 0.49 kcal/mol, corresponding to $\Delta(\log K) = 1.07 \pm 0.36$. If only the $\beta$-g1 and α-g1 tautomers/rotamers are compared, the calculated $\Delta(\log K)$ value, 0.97 ± 0.42, is still close to the experimental difference of 1.26.

This good agreement with the experimental relative log $K$ values could only be achieved because the frequency-dependent corrections were taken in consideration. Without the $\Delta G_{\mathrm{th}}(T)$ terms in Table 3, the relative free energy of the $\beta$-g1 species is 3.73 kcal/mol, and the calculated $\Delta(\log K)$ is 2.74, more than double the experimental value. Disregarding the $\Delta G_{\mathrm{th}}(T)$ terms would lead to the practical disappearance of the $\beta$-t rotamer from the equilibrium mixture. The agreement with experimental results provides a strong support for the selected zwitterion + 2 water model in calculating internal terms.

The calculations are not as successful in estimating the rotamer populations for the two tautomers. In-solution NMR measurements by Noszál and Sándor[17a] found 43−45% g1, 40−43% g2, and 14−16% t rotamers for the α-zwitterion. The composition with the $\beta$-tautomer is 47−49% g1, 38−42% g2, and 12−14% t. Relative free energies in Table 4a indicate qualitative agreement with the experimental population, with one exception. The trend is correct for the α-zwitterion, but the $\beta$-trans form was computed as more populated than the gauche-2 rotamer, in contrast to the experimental results.

To improve the results, additional solvation calculations have been performed. The torsion angle for the α-carboxylate differs by about 60° whether it was determined for the isolated zwitterion or for the dihydrate (see Geometry section). In one set of calculations, we performed a transformation of α-g1 to α-g2 using HF/6-31G* geometries optimized for the isolated zwitterion. (Atomic charges were fit to the relevant electrostatic potentials.) The $^{-}\mathrm{O}-\mathrm{C}_{\mathrm{carb}}\mathrm{C}_{\alpha}\mathrm{C}_{\beta}$ torsion angle changed from −38.9° to −30.2°, and the calculated $\Delta G$(sol) value was 3.67 kcal/mol. For comparison, in calculations utilizing the zwitterion geometry from the dihydrate, the $^{-}\mathrm{O}-\mathrm{C}_{\mathrm{carb}}\mathrm{C}_{\alpha}\mathrm{C}_{\beta}$ torsion angle changes from 21.9° in g1 to −25.9° in g2, and the corresponding $\Delta G$(sol) value is 4.05 kcal/mol (Table 4a). Thus, a change of the torsion angle that is larger by about 40° in the latter case resulted in a relative free energy increase of 0.4 kcal/mol. A second set of calculations used the HF/6-31G* optimized isolated α-g2 zwitterion geometry, and the carboxylate group was simply rotated by 60°, resulting in $\Delta G(\mathrm{sol}) = 1.12$ kcal/mol. These calculations show that the torsion angle for the carboxylate group is a sensitive geometric parameter and can increase free energy changes by up to 1 kcal/mol. The torsion angle relevant in solution may not be determined from hydrates with a limited number of water molecules around the zwitterion. The geometric change, even allowing for an effect of as much as 1 kcal/mol, still does not explain the overestimation of the α-g2 rotamer relative to the α-g1 rotamer.

The reason for the, at best qualitative, agreement between calculations and experiment lies partly in the opposite trend for the internal and solvation free energy contributions. For example, the relative value for the $\beta$-t species was calculated as $(22.23 - 3.29) - 16.55 \pm 0.45 = 2.39 \pm 0.45$ kcal/mol. Thus, the difference of the two numbers is smaller by an order of magnitude than any terms in the subtraction. To reproduce the g1/g2 experimental population, a free energy difference of about 0.1 kcal/mol should have been obtained. Considering the many approximations in our approach, such a precise determination of the relative free energy terms should not be expected. To reproduce the experimental g1/t = 3−4 population ratio, a calculated relative free energy of 0.6−0.8 kcal/mol would have been required. This relative free energy range is almost the uncertainty in the free energy perturbation calculations for the present system.

A comparison of the ab initio/Monte Carlo and SCIPCM calculations (Table 4b) shows that, despite the very different thermal corrections obtained in the gas-phase and in-solution optimizations for the $\beta$ tautomers, the total relative free energy is still close both at the HF/6-31G* and HF/6-311++G** levels for the $\beta$-g1 species. The value shows some basis-set dependence, but the relative $\beta$-g1 free energy agrees within 0.36 kcal/mol in the two types of calculations with each basis set. This result suggests that changes in the optimized geometry and differences in thermal corrections due to explicit versus continuum consideration of the near-solvent environment have compensating effects. The SCIPCM calculations predict, however, 3.5−3.8 kcal/mol relative free energy for $\beta$-g1, a range that is significantly above the best value of 1.32 kcal/mol in Table 4a. Taking the lower limit of 3.50 kcal/mol, the calculated value of $\Delta(\log K) = 2.57$ (eq 3) is more than 1 log $K$ unit larger than the experimental value, which was closely estimated by the ab initio/Monte Carlo method. The possible reason for this overestimation is the neglect of the correlation effect. The best value in Table 4a was obtained at the MP2/6-311++G** level, including a relative correlation energy of about 2.5 kcal/mol (Table 2). We could not estimate this energy within SCIPCM, as no MP2 calculations are available for this method in Gaussian 98 at present.

The SCIPCM method produces, however, a good free energy difference between the $\beta$-g2 and the $\beta$-g1 conformers at the HF/6-31G* level. The result is less satisfactory with the larger, HF/6-311++G**, basis set, but the relatively small free energy increase for $\beta$-g2 as compared to $\beta$-g1 is basically in agreement with the experiment: there are only slightly more $\beta$-g1 than $\beta$-g2 conformers in the equilibrium mixture. On the basis of this computational finding, it can be concluded that the correlation effect is more important for comparisons of different tautomers ($\alpha$ and $\beta$) of the Asp zwitterion than for considerations of the conformers of a given tautomer.

In summary, this study is the first attempt at calculating chemical equilibrium in aqueous solution for tautomeric zwitterion structures of an amino acid. Despite the weaknesses described above, the ab initio/Monte Carlo calculations reproduced well the difference in the logarithm of deprotonation constants of the zwitterion tautomers and reproduced most of the trend for the rotamer populations for the specific tautomers. A possible strategy for the future is to optimize the geometry of the zwitterionic species in solution using the SCIPCM method and to calculate thermal corrections by obtaining normal frequencies for the in-solution structure. The relative internal energy should be determined by single-point calculations, at the possible highest level, *in the gas phase at the geometry obtained in solution.* Then, by assigning atomic charges to the solute (using the procedure described in the Method section or using, for example, CM2 charges[37]), relative solvation free energies for the tautomers/conformers can be obtained in Monte Carlo simulations. In this approach, the vibrational changes upon the direct solute−water interaction will not be considered. Comparison with the experimental results could suggest the most appropriate computational approach.

**Solution Structure.** Radial distribution functions (rdf's) obtained in Monte Carlo simulations showed the main course, as was determined previously for −COOH[38] and for the $-COO^-$ and $-NH_3^+$ groups.[26c] The effect of the strong intramolecular hydrogen bond(s) can be concluded from analysis of the rdf's of atoms involved in H-bond formation. The total number of hydrogen bonds, $N_{HB}$, for a species and the coordination numbers around selected atoms also reflect the fact that hydration is partially hindered at some sites for a tautomer/rotamer.

The calculated number of the hydrogen bonds (Table 5) depends on the accepted upper limit for the solute−solvent pair energy when the distribution function is integrated. This limit, as the first minimum of the curve, was not clear in some cases. Taking the limit at −5 and −6 kcal/mol, $N_{HB}$ is 8−10 for most species. It is well seen that $\beta$-g1 forms the smallest number, 7.5, of solute−water hydrogen bonds because of the strong internal $NH\cdots{}^-O$ bonding. $N_{HB}$ is larger by 2.5 for $\beta$-t (even with a lower upper limit accepted for determination of the number of intermolecular hydrogen bonds), which reflects the opening of two more sites, a carboxylate oxygen and an N−H site, for hydration. Only small changes in $N_{HB}$ have been calculated for the $\alpha$-zwitterion. The largest value was obtained for the trans rotamer, but the number of hydrogen bonds changed by less than a unit value. This indicates that the number of the closest and most strongly bound water molecules is hardly changed upon rotation of the neutral −COOH group.

This latter conclusion is supported by the constant value of the total H/O coordination number for the $\alpha$-zwitterion. The sum of the $(N)H/O_w$, $=O/H_w$, $-O/H_w$, $-O^-/H_w$, and $H/O_w$ coordination numbers, $\Sigma H/O$ was calculated to 10 for each rotamer. The number of water molecules in the vicinity of the neutral −COOH group is nearly constant, as seen from the values around 1 for both the $=O/H_w$ and $H/O_w$ coordination numbers. Hydration coordination numbers for the ionic sites show a complementary character: $(N)H/O_w$ increases and $-O^-/H_w$ decreases for g2, but the total H/O number does not change. It should be emphasized, however, that $\Sigma H/O$ is always larger than $N_{HB}$, meaning that not all water molecules hydrating the solute contribute to the number of hydrogen bonds formed. The $N/O_w$ value shows remarkable changes in the range of 3−5, indicating reorganization of the water structure around the $NH_3^+$ group upon rotation of the −COOH group, but this process does not show correlation with the formation of intermolecular hydrogen bonds.

The $\Sigma H/O$ value is below 10 for the $\beta$-g1 rotamer and above 10 for the $\beta$-g2 and $\beta$-t rotamers. The decrease in the total $\Sigma H/O$ to 8.7 is mainly due to the decrease of the $(N)H/O_w$ value to 2 from the general value of about 3. Also in this case, the reason for this finding is the strong intramolecular hydrogen bond. The $-O^-/H_w$ value is 5.3 for $\beta$-t, the largest for all six species in the table. (The $-O^-/H_w$ coordination number is the sum of the values for the two carboxylate oxygens.) The value of 5.3 reflects the maximum O/H coordination number for a free and *remote* carboxylate group of aspartic acid. In fact, the largest number is only 4.9 for the $\alpha$-tautomer, where the somewhat bent intramolecular hydrogen bond is always present between the protonated amine and the carboxylate groups. The maximum $-O^-/H_w$ value is still smaller than the carboxylate coordination numbers of 6.8−7.2 calculated for formate and acetate[26c] and reflects the effect of the $\beta$-t $NH_3^+$ group with a torsion angle remarkably less than 180° (Table 1a.)

## Summary

The present study is the first attempt at the theoretical calculation of the tautomeric/conformational equilibrium for a zwitterionic $\alpha$-amino acid in aqueous solution. On the basis of previous computational experience for glycine[1−4] and results in the present study for aspartic acid, the zwitterionic structure does not correspond to a local energy minimum on the potential energy surface of the isolated species.

An assumption has been accepted here that the zwitterionic geometry, relevant in aqueous solution, can be reasonably determined by geometry optimization of the dihydrated zwitterion. Aspartic acid forms two tautomeric zwitterions with at least three (two N−C−C−C gauche and one trans) stable rotamers. Their populations were determined by NMR spectroscopy in aqueous solution, in line with determination of several microconstants for the protonation/deprotonation equilibria.[17a]

The zwitterion geometry was determined in the dihydrate at the ab initio HF/6-31G* level. Single-point calculations for the

isolated zwitterion were performed up to QCISD(T)/6-31G* and MP2/6-311++G** levels. Thermal corrections for the dihydrate at $T = 298$ and $p = 1$ atm were calculated using HF/6-31G* based vibrational frequencies. The corrections are relatively small for the α-zwitterion rotamers but are considerable for the $\beta$ structures. The largest contributions to the relative internal free energy stem from changes in the zero-point energy and the vibrational entropy.

Relative solvation free energies were calculated using the free energy perturbation method implemented in Monte Carlo simulations. The sign of the solvation terms is opposite to that for the internal free energy, and the two terms are of the same order of magnitude for any rotamers. The total relative free energy in solution is, thus, generally a relatively small difference of two large numbers. It may contribute to the limited accuracy of the present calculations.

Nonetheless, the calculated relative log $K$ for the deprotonation equilibria of the α and $\beta$ zwitterions is $1.07 \pm 0.36$, in good agreement with the experimental value of 1.26. This result is considered to be strong support for the overall approach applied in this study. On the other hand, the ab initio/Monte Carlo calculations predicted only the rotamer preference (with one exception) for the six studied species in agreement with the experiment.

Calculations performed with the self-consistent isodensity polarizable continuum model produced similar relative free energies to those by the ab initio/Monte Carlo method at the Hartree−Fock level for the lowest-free-energy α and $\beta$ conformers. Furthermore, the SCIPCM method predicted the $\beta$-g2 and $\beta$-g1 free energy separation in basic agreement with experiment. Results suggest that consideration of the correlation energy is more important for comparisons of different tautomers of the Asp zwitterion than for considerations of conformers of a given tautomer.

Solution structure simulations indicate that 8−10 water molecules are in strong hydrogen bonds with the solute. The calculated values correctly reflect the effect of the strong intramolecular hydrogen bonds. Reasonable values have been obtained for the number of solute−water hydrogen bonds involving free carboxylate groups or those bonded internally.

**Acknowledgment.** P.I. Nagy is thankful to Professor Jorgensen for the use of the BOSS 3.6 program and to Professor Hadad for valuable discussions regarding the SCIPCM calculations. The computer time used in the ab initio calculations was granted by the Ohio Supercomputer Center, which is also thanked.

## References and Notes

(1) (a) Langlet, J.; Caillet, J.; Evleth, E.; Kassab, E. Modelling of Molecular Structures and Properties. In *Studies in Physical and Theoretical Chemistry*; Rivail, J. L., Ed.; Elsevier: Amsterdam, 1990; Vol. 71, p 345. (b) Jensen, J. H.; Gordon, M. S. *J. Am. Chem. Soc.* **1995**, *117*, 8159.

(2) Ding, Y.; Krogh-Jespersen, K. *Chem. Phys. Lett.* **1992**, *199*, 261.

(3) Yu, D.; Armstrong, D. A.; Rauk, A. *Can. J. Chem.* **1992**, *70*, 1762.

(4) Ding, Y.; Krogh-Jespersen, K. *J. Comput. Chem.* **1996**, *17*, 338.

(5) Price, W. D.; Jockush, R. A.; Williams, E. R. *J. Am. Chem. Soc.* **1997**, *119*, 11988.

(6) Chapo, C. J.; Paul, J. B.; Provencal, R. A.; Roth, K.; Saykally, R. J. *J. Am. Chem. Soc.* **1998**, *120*, 12956.

(7) Ramek, M.; Nagy, P. I. *J. Phys. Chem. A* **2000**, *104*, 6844.

(8) See, for example: (a) Clementi, E.; Cavallone, F.; Scordamaglia, R. *J. Am. Chem. Soc.* **1977**, *99*, 5531. (b) Romano, S.; Clementi, E. *Int. J. Quantum Chem.* **1978**, *14*, 839. (c) Bonaccorsi, R.; Palla, P.; Tomasi, J. *J. Am. Chem. Soc.* **1984**, *106*, 1945. (d) Mezei, M.; Mehrotra, P. K.; Beveridge, D. L. *J. Biomol. Struct. Dyn.* **1984**, *2*, 1. (e) Alagona, G.; Ghio, C.; Kollman, A. *J. Mol. Struct. (THEOCHEM)* **1988**, *166*, 385. (f) Alagona, G.; Ghio, C. *J. Mol. Liquids* **1990**, *47*, 139.

(9) (a) Tuñón, I.; Silla, E.; Millot, C.; Martins-Costa, M. T. C.; Ruiz-López, M. F. *J. Phys. Chem. A* **1998**, *102*, 8673. (b) Nagaoka, M.; Okuyama-Yoshida, N.; Yamabe, T. *J. Phys. Chem. A* **1998**, *102*, 8202.

(10) (a) Smith, P. E.; Dang, L. X.; Pettitt, B. M. *J. Am. Chem. Soc.* **1991**, *113*, 67. (b) Smith, P. E.; Pettitt, B. M. *J. Am. Chem. Soc.* **1991**, *113*, 6029. (c) Smith, P. E.; Marlow, G. E.; Pettitt, B. M. *J. Am. Chem. Soc.* **1993**, *115*, 7493.

(11) Gordon, H. L.; Jarrell, H. C.; Szabo, A. G.; Willis, K. J.; Somorjai, R. L. *J. Phys. Chem.* **1992**, *96*, 1915.

(12) (a) Romano, S.; Clementi, E. *Int. J. Quantum Chem.* **1980**, *17*, 1007. (b) Kikuchi, O.; Matsuoka, T.; Sawahata, H.; Takahashi, O. *J. Mol. Struct. (THEOCHEM)* **1994**, *305*, 79. (c) Yu, G.-S.; Freedman, T. B.; Nafie, L. A.; Deng, Z.; Polavarapu, P. L. *J. Phys. Chem.* **1995**, *99*, 835.

(13) Ramírez, F. J.; Tuñón, I.; Silla, E. *J. Phys. Chem. B* **1998**, *102*, 6290.

(14) Cao, X.; Fischer, G. *J. Phys. Chem. A* **1999**, *103*, 9995.

(15) Kinoshita, M.; Okamoto, Y.; Hirata, F. *J. Comput. Chem.* **1998**, *19*, 1724.

(16) Fujiwara, S.; Ishizuka, H.; Fundano, S. *Chem. Lett.* **1974**, 1281.

(17) (a) Noszál, B.; Sándor, P. *Anal. Chem.* **1989**, *61*, 2631. (b) Noszál, B.; Guo, W.; Rabenstein, D. L. *J. Phys. Chem.* **1991**, *95*, 9609.

(18) Nagy, P. I. *Recent Res. Dev. Phys. Chem.* **1999**, *3*, 1.

(19) (a) Nagy, P. I.; Durant, G. J.; Hoss, W. P.; Smith, D. A. *J. Am. Chem. Soc.* **1994**, *116*, 4898. (b) Nagy, P. I.; Flock, M.; Ramek, M. *J. Chem. Phys. A* **1997**, *101*, 5942. (c) Nagy, P. I.; Alagona, G.; Ghio, C. *J. Am. Chem. Soc.* **1999**, *121*, 4804.

(20) Nagy, P. I.; Takács-Novák, K. *J. Am. Chem. Soc.* **1997**, *119*, 4999.

(21) (a) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986. (b) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502.

(22) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.6; Gaussian, Inc., Pittsburgh, PA, 1998.

(23) (a) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785. (b) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(24) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.

(25) McQuarrie, D. *Statistical Mechanics*; Harper & Row: New York, 1976.

(26) (a) Jorgensen, W. L.; Madura, J. D. *J. Am. Chem. Soc.* **1983**, *105*, 1407. (b) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 1489. (c) Jorgensen, W. L.; Gao, J. *J. Phys. Chem.* **1986**, *90*, 2174.

(27) Jorgensen, W. L. *Biochemical and Organic Simulation System User's Manual*; BOSS, version 3.6; Yale University: New Haven, CT, 1995.

(28) (a) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926. (b) Jorgensen, W. L.; Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381.

(29) Tirado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(30) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 316.

(31) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.

(32) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098.

(33) East, A. L. L.; Radom, L. *J. Chem. Phys.* **1997**, *106*, 6655.

(34) (a) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486. (b) Beveridge, D. L.; Schnuelle, G. W. *J. Chem. Phys.* **1975**, *79*, 2562.

(35) See, for example: (a) Constanciel, R.; Contreras, R. *Theor. Chim. Acta* **1984**, *65*, 1. (b) Kozaki, T.; Morihashi, K.; Kikuchi, O. *J. Am. Chem. Soc.* **1989**, *111*, 1547. (c) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127. (d) Cremer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213. (e) Hawkins, G. D.; Cremer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824.

(36) (a) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117. (b) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.

(37) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820.

(38) Nagy, P. I.; Dunn, W. J., III; Alagona, G.; Ghio, C. *J. Phys. Chem.* **1993**, *97*, 4628.