JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

On the Origins of Core−Electron Chemical Shifts of Small Biomolecules in Aqueous Solution: Insights from Photoemission and *ab Initio* Calculations of Glycine$_{aq}$

Niklas Ottosson,[*,†] Knut J. Børve,[*,‡] Daniel Spångberg,[*,§] Henrik Bergersen,[†] Leif J. Sæthre,[‡] Manfred Faubel,[∥] Wandared Pokapanich,[†] Gunnar Öhrwall,[⊥] Olle Björneholm,[†] and Bernd Winter[#]

[†]Department of Physics and Astronomy, Uppsala University, Box 516, SE-751 20 Uppsala, Sweden

[‡]Department of Chemistry, University of Bergen, NO-5007 Bergen, Norway

[§]Department of Materials Chemistry, Uppsala University, Box 538, SE-751 21 Uppsala, Sweden

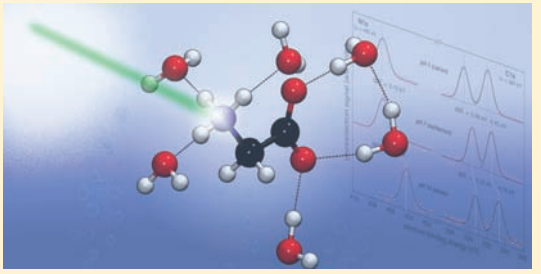
[∥]Max-Planck-Institut für Dynamik und Selbstorganisation, Bunsenstrasse 10, D-37073 Göttingen, Germany

[⊥]MAX-lab, Lund University, Box 118, SE-221 00 Lund, Sweden

[#]Helmholtz-Zentrum Berlin für Materialien und Energie and BESSY, Albert-Einstein-Strasse 15, D-12489 Berlin, Germany

Supporting Information

ABSTRACT: The local electronic structure of glycine in neutral, basic, and acidic aqueous solution is studied experimentally by X-ray photoelectron spectroscopy and theoretically by molecular dynamics simulations accompanied by first-principle electronic structure and spectrum calculations. Measured and computed nitrogen and carbon 1s binding energies are assigned to different local atomic environments, which are shown to be sensitive to the protonation/deprotonation of the amino and carboxyl functional groups at different pH values. We report the first accurate computation of core-level chemical shifts of an aqueous solute in various protonation states and explicitly show how the distributions of photoelectron binding energies (core-level peak widths) are related to the details of the hydrogen bond configurations, i.e. the geometries of the water solvation shell and the associated electronic screening. The comparison between the experiments and calculations further enables the separation of protonation-induced (covalent) and solvent-induced (electrostatic) screening contributions to the chemical shifts in the aqueous phase. The present core-level line shape analysis facilitates an accurate interpretation of photoelectron spectra from larger biomolecular solutes than glycine.



1. INTRODUCTION

Determination of the electronic and geometric structure of biomolecules in general, and of proteins in particular, in aqueous environments is of great importance for advancing our understanding of a wide range of biological processes. These systems tend to be highly sensitive to the dynamic interactions with solvent molecules, which themselves are part of the fluctuating water hydrogen-bonded network. Such effects may be rather subtle in terms of total energy variations but can be essential for biologic function. In contrast, the pH value of the solution has a much larger influence on the energetics and electronic structure of the solvated biomolecules since the availability of protons determines the distribution of charge states through protonation/deprotonation of the titratable functional groups.

For these reasons amino acids, being the building blocks of proteins, have been studied intensively both experimentally and theoretically. Experimental electronic structure studies of amino acids in aqueous solution are however scarce; a single X-ray absorption (XA)[1] and one X-ray emission (XE)[2] spectroscopic study have been reported for glycine in water. Photoelectron (PE) spectroscopy has only recently become applicable for the study of aqueous solutions,[3,4] and one single core-level PE study of an amino acid, lysine,[5] in water has been reported to date. In this work we use photoelectron spectroscopy to study aqueous glycine, the smallest of the amino acids, as it is well suited as a model system for biomolecular hydration. Several PE reports do exist for gas- and solid-phase glycine,[6−11] but as pointed out above, such studies are not very useful for predicting molecular electronic properties in an aqueous environment. Figure 1 shows structures of the molecule in the gas phase, together with the three aqueous forms: cationic, zwitterionic, and anionic glycine—the distribution of which is determined by the pH of the solution.

In the following we use liquid-jet photoelectron spectroscopy to measure the nitrogen and carbon 1s binding energies (BEs) of cationic, zwitterionic and anionic glycine and, thereby, quantify

Received: November 17, 2010
Published: February 14, 2011

dx.doi.org/10.1021/ja110321q | *J. Am. Chem. Soc.* 2011, 133, 3120–3130

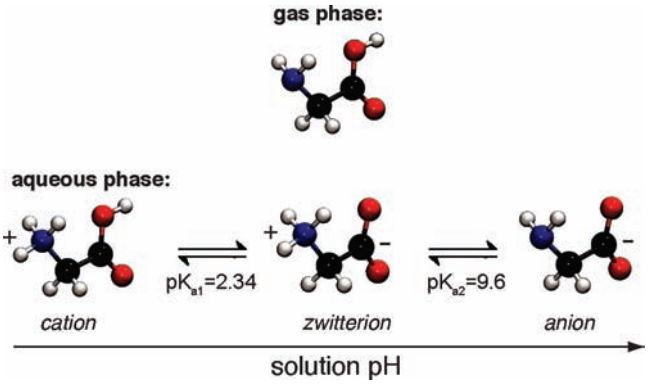


Figure 1. Structures of glycine in the gas phase and in aqueous solution: oxygen in red, carbon in black, nitrogen in blue, and hydrogen atoms in white. When dissolved in water, glycine adopts the zwitterionic form, where a proton is transferred from the carboxylic group to the amine group. For acidic solutions the carboxylic group is protonated, yielding the cationic form, and for basic solution the anion dominates where both the amino and the carboxyl groups are deprotonated.

electronic and geometric structural effects as a function of pH and the hydrogen bonding pattern. While much information can be retrieved from the experimental spectra alone, we will invest considerable effort into performing spectrum calculations from simulated hydration structures of glycine. Beyond enabling a more accurate interpretation of the photoemission data this approach also provides a deeper insight into the various contributions to the resulting spectrum. One goal of the present work is to investigate the level of theory needed to reproduce the characteristic features in the experimental photoelectron spectrum of the three forms of aqueous glycine. We have performed electronic structure calculations of solvated glycine with increasing complexity, starting with a description of the aqueous solvent as a polarizable continuum and gradually including an increasing number of explicit water molecules. Further, spectrum calculations from a large number of geometries for zwitterionic glycine, obtained by molecular dynamics (MD) simulations, enable a correlation analysis between binding energies and geometrical parameters of the solvation shell. It is demonstrated that the N1s BE distribution of the ammonium group displays considerable broadening due to a varying degree of polarization screening in the final state, depending on hydrogen-bond lengths to acceptor water molecules in the first solvation shell, while both the methylene and carboxylic C1s energies are less sensitive to the explicit solvation structure—in good agreement with the experimentally observed photoelectron widths. Finally, the explicit calculation of core-ionized final states allows us in a unique way to disentangle the contributions of protonation-induced (covalent) electronic structure effects from solvent-induced (electrostatic) polarization screening to the resulting chemical shifts. It is shown that the respective contributions to the chemical shifts are distinctly different in the photoemission from anionic and cationic species.

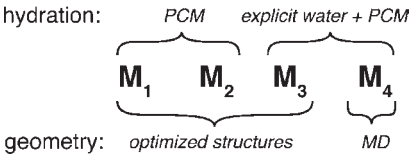
The present study of glycine clearly shows that valuable information on the hydration structure and energetics of small biomolecules in aqueous solution can be extracted from core-level spectra, if accompanied by high-level structure and spectrum calculations. Our results thus greatly encourage the application of liquid photoemission as a tool to access important site-specific chemical information from complex biomolecules in aqueous solution, complementary to standard tools, such as NMR and IR spectroscopy.

2. METHODS

2.1. Experimental Details. The X-ray photoelectron spectroscopy experiments from 0.5 molal (m) glycine aqueous solutions at pH values 1.0, 7.0, and 13.0 were performed using a 15 μm liquid jet at the I411 undulator beamline at MAX-lab, the Swedish national synchrotron facility, Lund. Experimental details have been described previously.[12,13] Briefly, the propagation of the synchrotron light beam is perpendicular to both the liquid jet and the central axis of the hemispherical electron analyzer, which is mounted at 54.7° (the so-called "magic angle") relative to the polarization plane, intended to minimize angular distribution effects.[14] Emitted photoelectrons need to travel around 1 mm in the main chamber, before reaching the 0.5 mm orifice entrance of the hemispherical electron energy analyzer, which is differentially pumped to $\sim 1 \times 10^{-6}$ mbar. The velocity of the jet was ~100 m/s, and the temperature of the solutions was kept at approximately 10 °C. The total experimental energy resolution was better than 300 meV for all core-level measurements.

Solutions were prepared from mixing highly demineralized water and glycine, obtained commercially from Sigma-Aldrich (>99%, used without further purification). Acidic and basic pH's were adjusted by the addition of hydrochloric acid and sodium hydroxide, respectively. Measured N1s and C1s spectra were fitted using Voigt line shapes where the respective lifetime widths were fixed to the experimental gas-phase values from the literature,[15] while the Gaussian width was free for each photoline. Electron binding energies reported here are with respect to vacuum, and calibration was made against the $1b_1$ state of liquid water.[16]

2.2. Computational Details. *2.2.1. Geometries of Solvated Glycine.* First-principle calculations have been performed of the geometric and electronic structure of gas-phase glycine and its three molecular forms in the aqueous phase, i.e., cationic (Gly^+), zwitterionic ($Gly^{+,-}$), and anionic (Gly^-). For the aqueous species, a number of different models have been employed, differing both in how the water solvent is included and in how the geometries are obtained. We label the theoretical models in the present work as M_1—M_4:



The water solvent was included in two principally different ways, either by implicit solvation only (M_1 and M_2) using the polarizable continuum model (PCM) as developed by Tomasi and co-workers[17] or by placing the molecule explicitly hydrogen bonded to a number of water molecules into a PCM cavity (M_3 and M_4). For cationic and anionic glycine in water the effect of the charge of the molecule must be carefully addressed. In the case of implicit solvation only, calculations are performed for Gly^+ and Gly^-, without (M_1) and with (M_2) the respective Cl^- and Na^+ counterions—the latter resulting in charge-neutral systems. Ground-state geometry optimizations of all these systems were performed within the PCM field. When the charge is not compensated for, the polarization of the PCM field will be overestimated due to lack of long-range screening of the charge. Including the counterions reduces these problems, although direct contact between the counterions is not likely to be a representative geometry for the real solutions. This situation is greatly improved in model M_3 where six explicit water molecules are included. Input geometries were based on those reported by Derbel et al.,[18] from which water molecules were removed down to the six molecules directly bonded to glycine. For calculation of cationic and anionic glycine ground-state structures the respective counterions were explicitly included. Chloride and sodium ions (given that HCl and NaOH was used to adjust the pH) were

coordinated to the water molecule at the largest distance from the charged group of the glycine (no direct ion pair). All structures were thereafter optimized within the PCM field.

Optimization of ground-state geometries in models M_1-M_3 was performed using hybrid density functional theory with the B3LYP functional as implemented in Gaussian 03.[19] The UA0 scheme for assigning atomic radii (United Atom Topological Model applied on atomic radii of the UFF force field for heavy atoms as implemented in Gaussian 03) was adopted for all PCM models. The glycine molecule was described in flexible atom-centered Gaussian basis sets based on Dunning's triple-ζ bases[20] and augmented by a double set of polarization functions on all atoms[21,22] as well as a set of diffuse primitive *s* and *p* functions centered at the carboxylic carbon and oxygens (exponents α_s = 0.0456 and α_p = 0.03344 on C, and α_s = 0.08458 and α_p = 0.05654 on O). For the solvent water molecules Dunning's reduced double-ζ plus polarization (DZP)[23] basis was used. We will in the following refer to this mixed basis as modified TZ.

Although model M_3 probably captures an essential part of the interactions of glycine with the (first-shell) waters rather accurately, a single optimized structure cannot describe the dynamical range of configurations in real liquid water. To investigate how fluctuations in the hydrogen-bond pattern affect the associated PE spectrum, we have computed structures using molecular dynamics simulations (M_4), and we have chosen zwitterionic glycine as a representative case. Molecular dynamics simulations of one glycine zwitterion and 864 water molecules were performed using a cubic simulation box and standard periodic boundary conditions. The temperature was set to 300 K and the pressure to 0 Pa using Nosé−Hoover extended system dynamics[24,25] to obtain the NPT ensemble. The velocity Verlet integrator[26] was used with a time step of 0.5 fs. The water molecules were kept rigid using the RATTLE algorithm[27] using successive over-relaxation to increase convergence speed,[28] while the glycine molecule was fully flexible. Polarizable shell models were used for both water and glycine molecules.[29] The shells were given small masses, and their instantaneous positions were found using adiabatic dynamics.[30] The water−water, glycine−glycine, and water−glycine interaction model parameters were determined by fitting to *ab initio* data computed for gas-phase clusters at the counterpoise-corrected[31] MP2/aug-cc-pVTZ level of theory. The *ab initio* calculations were performed with the GAMESS[32] and Turbomole[33] program suites. Basis sets were obtained from the EMSL database.[34] Details of the model will be published in a separate article, but the interaction model can be found in the Supporting Information. Ewald lattice sums were used for the Coulomb interactions,[35] and short-range interactions were cutoff at 12 Å. The system was equilibrated for 0.5 ns followed by a production run of 0.5 ns. The resulting density of the system was 0.996 g/cm^3.

2.2.2. Binding Energy Calculations. Using ground-state geometries and basis sets as outlined above, vertical shifts in N1s and C1s ionization energies relative to those of the neutral gas-phase molecule were computed as described in the following. Initial- and final-state electronic energies were obtained using Gaussian 09[36] at the B3LYP level of theory, in both cases calculated in the ground-state geometries. For the core-ionized molecules, the effect of the ionized core on the valence electron structure is represented by Stevens et al.'s effective core potentials (ECP),[37] scaled to account for only one electron in the 1s shell.[38] Moreover, the implicit solvent was described using a nonequilibrium PCM (NEPCM) field based on the initial-state configurations. Within this NEPCM approach, only electronic relaxation of the solvent is allowed while the slow component of the dielectric polarization−which is associated with reorientation of molecular dipoles−is frozen to the value in the initial (ground) state. This is in essence the sudden approximation, frequently used to describe the photoemission process.[14] The NEPCM approach has previously been successfully applied to the valence photoionization of simple aqueous cations.[39] Since the calculations for the core-ionized species are based on the use of an effective core potential rather than an explicit core hole, absolute ionization energies cannot be obtained. However, it is possible to obtain relative ionization energies, i.e. chemical shifts, with high accuracy.[38] In this work, we calculate the shifts in core ionization energies relative to the same core in the gas-phase molecule, treated on the same level of theory. Since the experimental gas phase BEs are known,[10,11] they can thus be used to anchor the calculated values on an absolute scale, allowing a direct and quantitative comparison between experimental and computed energies. Notice though that the comparison thereby is only accurate within the experimental uncertainty of the reported binding energies by Plekan *et al.* for gas-phase glycine, which was estimated to ±100 meV for the C1s and N1s photolines.[10,11]

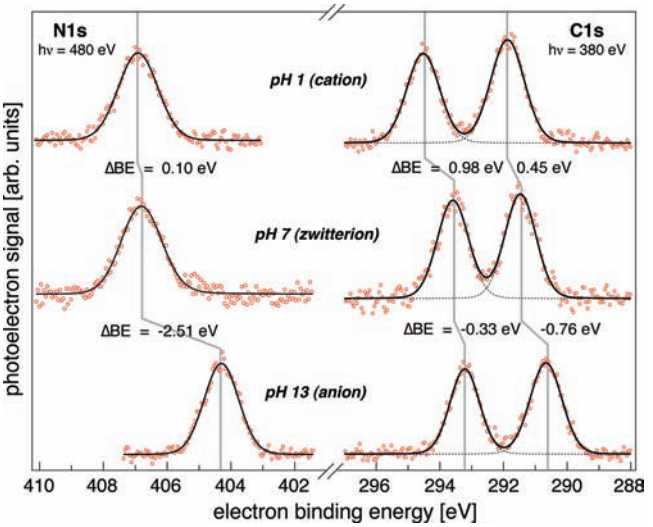


Figure 2. N1s and C1s photoelectron spectra of aqueous glycine recorded at pH 1 (top), pH 7 (middle), and pH 13 (bottom). Circles are experimental data points, and the lines represent fits; dashed lines are the individual components, and the solid lines show the resulting total fit. Energy shifts are given relative to the zwitterionic form.

From the MD simulation of zwitterionic glycine, 100 snapshots were selected separated by 5 ps in time. For each snapshot, water molecules within a given (water oxygen) distance from $Gly^{+,-}$ were determined to generate clusters of a center glycine zwitterion surrounded by its 6, 8, 10, 12, 14, or 16 closest water molecules, respectively. Vertical ionization energies were computed for each such cluster using the nonequilibrium PCM + explicit solvation model with the same methodology as that for the M_3 structures. The ionization energies determined in this way were sorted into histograms with a bin size of 0.1 eV, and expectation values and the full widths at half-maximum (fwhm) were determined by fitting to Gaussian functions.

3. RESULTS AND DISCUSSION

3.1. N1s and C1s Photoelectron Spectra. Nitrogen 1s and carbon 1s PE spectra of glycine in water, for three different pH values, are shown in Figure 2 together with peak fits performed along the lines described in the Methods section. All values are further summarized in Table 1. The spectra on the left side of the figure show the evolution of the N1s BE as a function of the charge state of the molecule. The middle-left spectrum pertains to the neutral solution, at pH = 7, and yields 406.81 eV BE. At this pH almost all glycine molecules are in the zwitterionic form, i.e., with one ammonium (NH_3^+) moiety and one carboxylate (COO^-) moiety. Increasing the pH to 13, which transforms the majority of the molecules to the anionic form, results in a

Table 1. Summary of All Measured N1s and C1s Binding Energies and Peak Widths (fwhm) for Aqueous Glycine[a]

	N1s		$C1s_{COO}$		$C1s_{CH2}$	
	BE	fwhm	BE	fwhm	BE	fwhm
gas phase[b]	405.4	0.89	295.2	0.71	292.3	0.77
$cation_{aq}$	406.91	1.45	294.53	1.10	291.88	1.17
$zwitterion_{aq}$	406.81	1.40	293.55	1.07	291.43	1.10
$anion_{aq}$	404.30	1.20	293.22	1.05	290.67	1.10

[a] Values are given in electron volts. [b] From ref 10.

negative shift in the N1s BE of −2.51 eV. This large chemical shift arises from deprotonation of the ammonium group. In contrast, a decrease of pH from 7 to 1 leads to a very small chemical shift, +0.10 eV, of opposite sign. This reflects that the local charge state of the amino group is not altered in that case. At a first level of approximation, the magnitude of the chemical shift correlates with the amount of electron charge withdrawn from the nitrogen site in the ground state.

C1s PE spectra of the same solutions are shown at the right side of Figure 2. For each pH value two distinct peaks are observed, arising from ionization of the methylene (CH_2) and the carboxyl group, respectively. The carboxyl C1s peak is found at higher binding energies for all protonation states due to the higher electronegativity of the carboxylic oxygens. For the zwitterionic form, at pH 7 (Figure 2, right center), the respective BEs obtained from peak fits are 291.43 and 293.55 eV. For the anionic form, at pH 13, the shifts upon deprotonation of the amino group are −0.33 and −0.76 eV for the carboxylate and methylene carbons, respectively. Both of these shifts are significantly smaller than the case for the N1s and larger for the methylene carbon than for the carboxylate carbon, thus correlated with the distance to the site of deprotonation. For cationic glycine, at pH 1, the C1s ionization energies are instead 291.88 and 294.53 eV. Protonation of the carboxylate group of zwitterionic glycine thus leads to a +0.98 eV C1s BE shift for carboxylic carbon while only +0.45 eV for methylene. Again, these effects are smaller than the corresponding N1s chemical shifts for protonation of the amino group. This is partly due to the larger spatial distribution of the negative charge on the carboxylate group in the zwitterionic form as compared to that of the rather localized positive charge at the amino group.

In addition to the chemical shifts there is also information available in the line widths of the respective core lines; the experimental full widths at half-maximum (fwhm) are given in Table 1. All lines are considerably broader in water compared to the gas phase, as expected from the existence of a distribution of hydration patterns. To extract information from these values alone is difficult, as the line width receives a contribution not only from lifetime and experimental broadening but also from unresolved vibrations in the ionized final state that might have a different character in the gas and aqueous phases. However, it is clear that the N1s lines are generally broader than the C1s and that the width is somehow connected to the protonation state of the amine group; the cation and zwitterion N1s fwhm's are both approximately 1.4 eV (both structures share the protonated NH_3^+ group) while the transformation into the anionic state with its neutral amine group yields the significantly lower value of 1.2 eV. It is therefore reasonable to expect the line widths to be influenced by the hydration structure around the functional groups, a question to which we will return in more detail in the next section.

Table 2. Summary of Computed N1s and C1s Chemical Shifts for the Cationic, Zwitterionic and Anionic Forms of Glycine at Various Levels of Theory[a]

N1s shift	M_1	ΔM_1	M_2	ΔM_2	M_3	ΔM_3	M_3*	ΔM_3*	Exp.
$cation_{aq}$	3.74	2.23	2.29	0.78	1.18	−0.33	1.28	−0.23	1.51
$zwitterion_{aq}$	2.29	0.88	2.29	0.88	0.95	−0.46	1.05	−0.36	1.41
$anion_{aq}$	−1.96	−0.86	−1.78	−0.68	−1.73	−0.63	−1.67	−0.57	−1.1
$C1s_{COO}$ shift	M_1	ΔM_1	M_2	ΔM_2	M_3	ΔM_3	M_3*	ΔM_3*	Exp.
$cation_{aq}$	0.2	0.87	−0.14	0.53	−0.35	0.32	−0.42	0.25	−0.67
$zwitterion_{aq}$	−2.05	−0.4	−2.05	−0.4	−1.68	−0.03	−1.74	−0.09	−1.65
$anion_{aq}$	−3.31	−1.33	−2.78	−0.8	−2.29	−0.31	−2.35	−0.37	−1.98
$C1s_{CH2}$ shift	M_1	ΔM_1	M_2	ΔM_2	M_3	ΔM_3	M_3*	ΔM_3*	Exp.
$cation_{aq}$	0.7	1.12	0.15	0.57	−0.18	0.24	−0.17	0.25	−0.42
$zwitterion_{aq}$	−0.74	0.13	−0.74	0.13	−0.89	−0.02	−0.86	0.01	−0.87
$anion_{aq}$	−2.56	−0.93	−2.24	−0.61	−1.92	−0.29	−1.92	−0.29	−1.63

[a] Experimental shifts are given for reference and the deviation from the experimental values are tabulated in the adjacent column to each model. Values are given in electron volts. For details on the different models M_1−M_3, see the main text.

3.2. Calculated Structures and Core-Level Binding Energies. In the following we present and discuss the performance of our various computational approaches in reproducing the experimental N1s and C1s binding energies of solvated glycine.

3.2.1. Polarizable Contiunuum. In the first approach we treat the solvent entirely as a polarizable continuum. Two situations are considered; the single glycine molecule (M_1) and the molecule plus its counterion (M_2), the latter yielding charge neutrality of the total system for anionic and cationic glycine in water. Calculated N1s and C1s chemical shifts for the different approaches are presented in Table 2, and the deviations from the respective experimental shifts are given in the adjacent columns. The overall performance of these simple nonequilibrium PCM models is rather poor. In the case of the bare glycine cation the ammonium N1s BE is overestimated by 2.23 eV and the carboxylic and methylene C1s BEs are overestimated by 0.87 and 1.12 eV, respectively. The agreement with experiment improves significantly upon inclusion of counterions, i.e. in model M_2, but still the cation N1s BE and the anionic carboxyl C1s BE remain overestimated and underestimated, respectively, by 0.8 eV. Also, the C1s BE difference for methylene and carboxyl carbon is poorly reproduced; in the zwitterionic case it is overestimated by >0.5 eV.

The performance of the implicit solvation models for glycine is hence overall unsatisfactory in predicting N1s and C1s chemical shifts for the present system. This failure is partly due to an exaggerated polarization of the solvent continuum by a single charge, which explains the slightly better performance when the counterions are included. In order to reproduce the effect of the water hydrogen bonding on core-level BEs we must however obviously proceed to better models. We begin by inclusion of explicit water molecules.

3.2.2. Explicit First-Shell Water Molecules Plus Polarizable Continuum. In model M_3 we approximate the first hydration shell by six water molecules, three coordinating the NH_2/NH_3^+ groups and the remaining three waters bonded to the COOH/COO^- groups. The effect of the solvent outside the first shell is approximated by a PCM field. The optimized gas-phase, cationic,

dx.doi.org/10.1021/ja110321q | J. Am. Chem. Soc. 2011, 133, 3120–3130

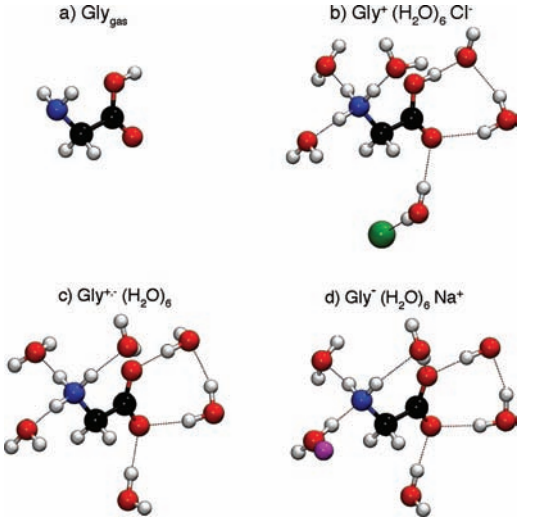


Figure 3. Explicit first-solvation shell structures of glycine used for calculation of chemical shifts, obtained from geometry optimization at the B3LYP/modified TZ level of theory (see text for details). (a) Gas-phase structure for reference; (b) cationic $Gly^+(H_2O)_6Cl^-$, (c) zwitterionic $Gly^{+\cdot-}(H_2O)_6$, and (d) anionic $Gly^-(H_2O)_6Na^+$. While the six explicit water molecules constitute the first solvation shell, a PCM field accounts for long-range solvation effects.

zwitterionic, and anionic structures used to calculate core−electron BEs at this level of theory are displayed in Figure 3 (details on the geometries are presented in the Supporting Information).

All calculated core-level shifts obtained from the structures given in Figure 3 are presented in Table 2. In order to more easily compare the experimental and computed shifts we refer to the graphical compilation of the most relevant set of calculated BEs, shown in Figure 4. Overall, computed energies quite well capture the trends in the shifts between the different protonation states and are far superior to the simpler model M_1 and M_2.[40] The performance is clearly best for the zwitterionic case for which we calculate the N1s, $C1s_{COO}$, and $C1s_{CH2}$ shifts to 0.95, −1.68, and −0.89 eV, respectively, relative to the gas phase. While the N1s value is somewhat too low, the carbon shifts are nearly exact; deviations are −0.46, −0.03, and −0.02 eV, respectively. For the ionic forms the agreement is somewhat poorer. For the cationic form the C1s shifts are about 0.3 eV overestimated while the calculations of the anion underestimates the shifts by about the same value. The rather uniform error in the N1s values for all forms however raises the question whether our model leads to a systematic error in the BE calculations. Alternatively, the fully relaxed structures of Figure 3 may not be representative for the average local H-bonding configurations of glycine in water. Furthermore, six explicit waters may be too few to capture the essential influence of the first solvation shell well enough. Driven by these concerns we have performed analogous calculations on a higher level of theory, B3LYP/aug-cc-pVTZ, using the same structures as shown in Figure 3—we denote this model M_3*. The resulting shifts have been included in Figure 4 as well as in Table 2 and show a small improvement relative to the lower-level theory. Notably, the C1s splitting is very well reproduced for all molecular forms. The deviations from experimental values are 100 meV or less, for each structure, showing that important features of the intramolecular electronic structure must have been captured accurately. In addition to testing the influence of basis set size we have further tried out two other recently

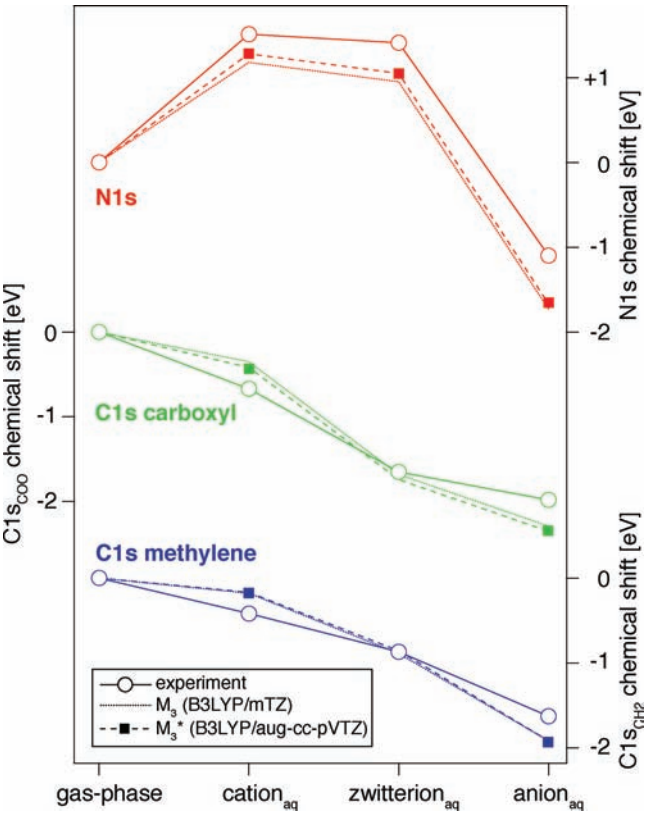


Figure 4. Calculated N1s and C1s chemical shifts from the structures in Figure 3 at the B3LYP/modified TZ and B3LYP/aug-cc-pVTZ levels of theory, together with the experimental values. The shifts are given relative to the corresponding BEs for gas-phase glycine.

developed hybrid DFT functionals, namely X3LYP[41] and M06,[42] which are known to describe hydrogen-bonding rather well. These values coincide with the B3LYP N1s and C1s shifts within ±50 meV. Hence, the systematic errors in the N1s shifts are rather unlikely to arise from too small a set of basis functions or from specific shortcomings associated with the B3LYP functional. Consequently, the optimized structures of Figure 3 are most likely not good enough models of the average hydrogen-bonding configuration in aqueous solutions. However, the overall satisfying agreement between experimental and computed N1s and C1s core level shifts of aqueous glycine at the M_3/M_3* level of theory suggests that the spectrum calculations are in principle sound, and we therefore proceed to obtain more accurate geometries from MD simulations.

3.2.3. Spectrum Calculations from MD Simulated Structures. Figure 5a displays calculated N1s and C1s ionization energy histograms of zwitterionic glycine, obtained from MD simulated geometries as described in the computational details, as function of the number of explicitly included water molecules. The shifts of the center of each distribution (vertical BEs) and associated peak widths (fwhm) are separately presented in Table 3. Ionization energies were calculated using the modified TZ basis set for clusters inside a PCM cavity, containing 6, 8, 10, 12, 14, and 16 water molecules around a single $Gly^{(+,-)}$—this model is termed M_4. While the dependence on the number of included water molecules is dramatic for N1s it is much less pronounced for carbon. Upon inspection of the geometries the reason for this behavior becomes apparent; when only the six

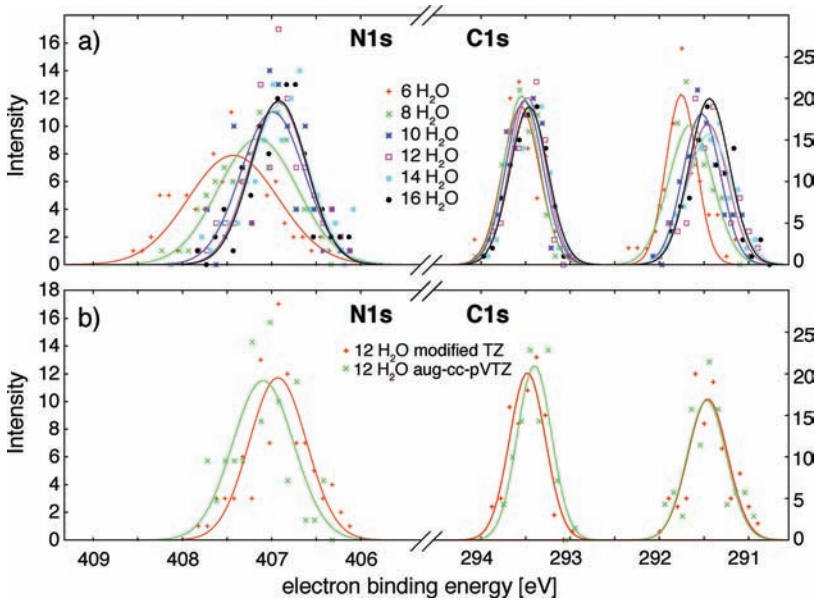


Figure 5. Distribution of calculated N1s and C1s photoelectron binding energies of aqueous zwitterionic glycine, based on 100 snapshots from the MD simulations. The calculated values from the 100 snapshots were put in a histogram with a bin size of 0.1 eV. Gaussian functions were fitted to these histograms. The histograms are shown as markers in the figure, while the lines show the Gaussian fits. (a) Distributions obtained at the B3LYP/modified TZ level of theory (M_4), where for each snapshots the closest 6, 8, 10, 12, 14, and 16 water molecules were included explicitly in the calculations. Convergence is met around 12 water molecules. (b) Distribution obtained for 12 water molecules, but for two different basis sets, comparing the modified TZ and the more extended aug-cc-pTVZ basis set (models M_4/M_4*).

Table 3. Chemical Shifts of the Vertical BE and Associated Full Widths at Half Maximum of Zwitterionic Glycine Obtained from Spectrum Calculation of Snapshots from MD Simulations[a]

			shift	fwhm
N1s	modified TZ (M_4)	6 H_2O	2.02	1.19
		8 H_2O	1.75	1.04
		10 H_2O	1.58	0.84
		12 H_2O	1.52	0.75
		14 H_2O	1.50	0.76
		16 H_2O	1.51	0.74
	aug-cc-pVTZ (M_4*)	12 H_2O	1.69	0.81
$C1s_{COO}$	modified TZ (M_4)	6 H_2O	−1.65	0.49
		8 H_2O	−1.65	0.47
		10 H_2O	−1.68	0.49
		12 H_2O	−1.72	0.47
		14 H_2O	−1.74	0.50
		16 H_2O	−1.74	0.50
	aug-cc-pVTZ (M_4*)	12 H_2O	−1.80	0.45
$C1s_{CH2}$	modified TZ (M_4)	6 H_2O	−0.54	0.40
		8 H_2O	−0.65	0.58
		10 H_2O	−0.77	0.49
		12 H_2O	−0.84	0.53
		14 H_2O	−0.86	0.59
		16 H_2O	−0.86	0.53
	aug-cc-pVTZ (M_4*)	12 H_2O	−0.83	0.52

[a] Values are given in electron volts.

closest waters are extracted from the MD simulations, certain configurations lack full coordination on the ammonium side, due to the shorter water−glycine bond distances on the carboxylate end, which results in a too high N1s BE due to incomplete final state polarization screening. When the number of waters is increased these low-energy configurations disappear, and as a result the BE distribution narrows and the vertical ionization energy red-shifts. Hence, if an automated (unbiased) selection of the nearest water molecules is made from a large number of MD snapshots, 6 waters are not sufficient to saturate the most important glycine−water hydrogen bonds.

Both energies and fwhm's appear to be converged for clusters containing 12 H_2O molecules using the modified TZ basis set. For comparison, Figure 5b therefore shows the spectrum calculated with the aug-cc-pVTZ basis set with the same number of water molecules; this model is termed M_4*. The shifts calculated with the M_4 model are larger by about 0.2 eV for N1s compared to those obtained with M_4* and are lower by about 0.05 eV for $C1s_{COO}$, i.e. thus exhibiting a similar basis set dependence as observed for the structures in Figure 3. Compared to experimental values (Table 1), the overall performance of the MD model (M_4) is very satisfactory with a root-mean-square (rms) deviation of only 77 and 185 meV from the experimental values for the modified TZ and aug-cc-pVTZ calculations, respectively. Most importantly, compared to the shift obtained from the zwitterionic structure in Figure 3c, the N1s chemical shift is now reproduced much better—instead of being about 400 meV too low, the vertical N1s BEs from the MD structures are on average 100 meV too high. We can thereby conclude that the optimized structures obtained in model M_3 belong to the low-energy side of the rather broad BE distributions.

The calculated M_4/M_4* shifts agree with observations almost within the error bars (±100 meV) of the reference values for

dx.doi.org/10.1021/ja110321q | J. Am. Chem. Soc. 2011, 133, 3120–3130

gas-phase glycine.[10,11] The agreement is in accordance with experience from a study of 70 accurately determined C1s shifts in saturated, olefinic, and aromatic hydrocarbons. In this study, the ECP/B3LYP approach to core-level binding energies was found to reproduce the observed shifts with mean absolute errors of almost 130 meV (using cc-pVTZ basis sets) and just above 80 meV (with Dunning's tzp sets).[43]

3.3. Correlation Analysis between Structural Parameters and Core-Ionization Energies. Beyond providing a set of more accurate geometries for solvated $Gly^{+,-}$ compared to the single optimized structures in Figure 3, the BE distributions obtained from the MD simulations provide an opportunity to investigate the origins of the experimental core-level line widths in terms of solvation geometries. As previously noted, the experimental N1s fwhm of zwitterionic glycine is substantially larger than that for C1s, and this is also observed in the spectrum calculations from the MD snapshots. We have therefore extracted correlations between the respective BEs and suitable structure parameters, analyzing the results obtained from the M_4* model containing the closest 12 water molecules in an NEPCM field.

While we would like to explore possible correlations between inner-shell binding energies and structural parameters for the water molecules in the first solvation sphere around the zwitterionic glycine molecule, the number of binding energies (3) and geometry variables (16, see below) suggest that univariate regression analysis is both an impractical and insufficient tool. On the other hand, the structural parameters show a high degree of correlation among themselves, making multiple linear regression in terms of the original parameters prone to produce models with low interpretability. A better approach is to carry out regression in terms of orthogonal variables (also known as latent variables or principal components), reflecting the covariance among the structural parameters. These can be defined in different ways, each giving rise to regression models with its own strengths and disadvantages. In partial least-squares (PLS) regression analysis, the regression is carried out in terms of latent variables that are based on covariance both among the geometry variables and between those and the binding energies. In many cases, PLS gives models with better prediction ability and fewer components than its competitors.[44−46]

The variable space in our analysis consists of eight distances and eight angles, characterizing the hydrogen-bonding configurations around the ammonium as well as the carboxylate groups, in addition to computed shifts in the $C1s_{COO}$, $C1s_{CH2}$, and N1s ionization energies. The geometry variables are denoted rNH-O#*n*, aNHO#*n*, rCO-HO#*n* and aCO-H-O#*n*, where *n* is an integer running from 1 to 4, and referring to the four water molecules closest to each of the two charged groups of the zwitterion. Further details about the data treatment are included in the Supporting Information; here we summarize the findings from the PLS regression analysis.

PLS component 1 explains 19% of the variance in the 16 (independent) geometry variables and 25% of the variance in the three binding energies. Its composition may be seen from Figure 6, where both geometry and energy variables are projected onto a plane spanned by the first and second PLS components. Component 1 is dominated by the combined distance between nitrogen and the three closest water molecules (rN1 + rN2 + rN3) minus the sum of NH···O angles involving the same three water molecules. These are the geometry parameters describing the solvation shell of the ammonium moiety. The dependent variables are also included in the figure, and the shift in N1s BE is

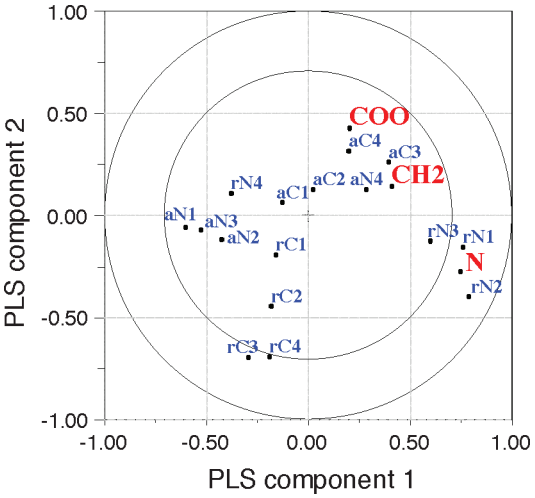


Figure 6. Projection of centered and standardized geometry and energy variables onto the two first PLS components. COO, $C1s_{COO}$; CH2, $C1s_{CH2}$; N, N1s; rN, distance NH-Ow; aN, angle N-H···O_w; rC, distance CO···H_w; aC, angle C-O···H_w, where O_w (H_w) is a water oxygen (hydrogen). The water molecules coordinated at the ammonium and carboxyl ends of glycine are numbered according to increasing value of rN and rC, respectively, and this numeral is appended to each marker, associating rN1 and aN1 with the water molecule that is coordinated the closest to the ammonium moiety.

seen to be strongly and positively correlated with PLS component 1. Thus, looser coordination of the three water molecules closest to the ammonium moiety increases the N1s binding energy, which may be rationalized in terms of withdrawal of the negatively charged oxygen end of the water molecules. However, a slight but yet consistent trend in the opposite direction is found for the fourth, more loosely bound water molecule: the N1s BE increases when this water molecule approaches. This indicates that closer approach of the fourth water molecule causes a destabilization of the inner three water molecules, which in turn results in an increase of the N1s BE.

The second PLS component explains 9% of the variance in the binding energy shifts and 10% of the variance in the geometry variables. From Figure 6 we see that PLS component 2 can be characterized as a carboxyl analogue to component 1, with a strong emphasis on the two less strongly bound water molecules (3 and 4). PLS component 2 correlates positively with the BE of the carboxyl carbon, to a lesser extent with that of methylene, and negatively with that of nitrogen. Since the water molecules coordinate to the carboxyl moiety through a positively charged hydrogen, this leads to a positive shift in the carboxyl C1s binding energy.

Figure 7 shows the respective binding energies as a function of the PLS component, which in each case explains most of the variance of this BE. As is evident from Figure 7a, PLS component 1 describes the main variation in N1s BE due to geometric variations in the solvation shell, implying that the N1s BE is quite susceptible to variation in the hydrogen-bonding structure at the ammonium site. The C1s binding energies vary within a narrower interval, and more importantly, the systematic variations with the geometry parameters for the solvation shell make up a smaller part of the total energy range than in the case for N1s. The methylene C1s data are more sensitive to PLS component 1 than 2 or, in other words, more sensitive to the details in the solvation shell at the ammonium end than at the carboxyl entity.

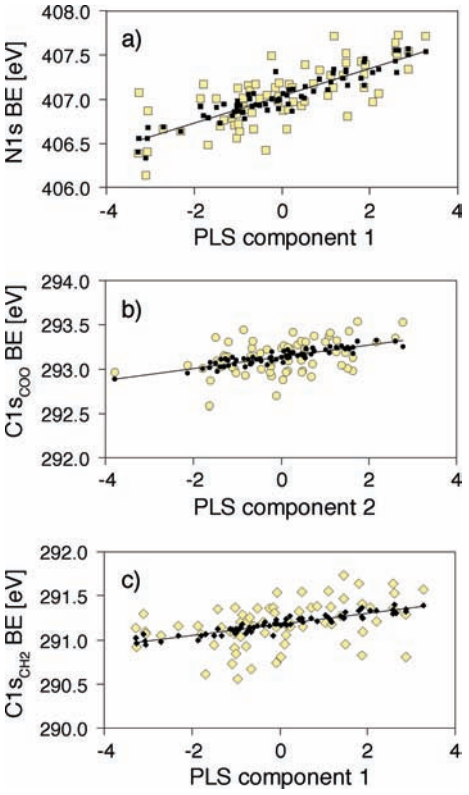


Figure 7. Core-level binding energies (BE, in eV) from electronic structure calculations (open markers) and as modeled by a two-component PLS model (filled markers) in terms of 16 geometric variables, for N1s (top; squares), carboxyl C1s (middle; circles), and methylene C1s (bottom; diamonds). The energy data are plotted vs the projection of each geometry onto the PLS component that is the most important for describing variations in the respective binding energy.

What is the physical origin of these correlations? To disentangle this, we have inspected the respective initial- and final-state contributions to the BE shift. We find that the initial-state energies do not seem to correlate with the parameters describing the hydrogen-bonding structure and therefore conclude that configurational line broadening is dominated by final-state polarization screening of the solvent. This provides a direct explanation as to the origin of the narrower N1s line shape for Gly^- compared to $Gly^{+,-}$ and Gly^+ as observed in the experiments (see Table 1): Since the polarization screening scales with the square of the (local) final-state charge, the BE correlation with the shortest $NH_3^+\cdots O$ hydrogen bond distance should be approximately three times stronger in the case of cationic ammonium (leading to a doubly charged NH_3^{2+} final state) compared to the neutral amine (with a singly charged NH_2^+ final state). This explains, at least in part, the experimentally observed narrowing of the N1s line by 0.2 eV when going from zwitterionic to anionic glycine.

Along the same lines of reasoning, we can identify at least two reasons for the narrower distributions of the C1s lines: First, as the carboxylate acts as a hydrogen-bond acceptor the electronic screening from the hydrogen-bond donating waters will be less efficient (due to the low electron density at the H_2O hydrogen sites). Second, in the ionized state, the carboxylate group is neutral, making the energy less sensitive to positions of the nearest waters. This results in a less pronounced configurational broadening and explains the narrower C1s line shapes compared to the N1s lines as observed in the experiments. These results show that PE line shapes from aqueous solutes is dependent on the charge state and can be sensitive to relatively small variations in local hydrogen-bond structure, at least in the case when the ionized site is involved in hydrogen-bond donation.

3.4. Decomposition of the Chemical Shifts upon Solvation. The resulting chemical shifts of glycine upon hydration in solution of a given pH value have two interconnected but principally different origins: structural modifications of the molecules due to protonation/deprotonation of the functional groups and polarization screening of the solvent. The former effect, which is directly connected to the solution pH, involves covalent bonding between the proton and the functional groups and results in a strong electronic perturbation in both the initial and final states. The solvent induced screening is more electrostatic in nature and dependent on water configurations and the local charge of the ionized atom. To disentangle these two contributions is challenging, and to separate the various contributions it would be useful to study glycine in the identical set of charge and conformational states, both in vacuum and in aqueous solution. Experimentally, this is not feasible though, because the gas-phase structure of glycine is unstable in water, and conversely, the ionic forms are not easily produced in vacuum. Calculations of the ionic forms in vacuum, denoted Gly^+_{vac}, $Gly^{+,-}_{vac}$, and Gly^-_{vac} respectively, can however bridge this gap.

Figure 8 shows the changes in the respective initial state (blue) and N1s-ionized final state energies (red) of the ionic forms in vacuum (center part) and in solution (right part), with reference to the neutral glycine molecule in the gas phase (left side). Vertical arrows represent the respective energy differences, i.e. the binding energies, and a difference in the arrow lengths relative to the gas phase identifies the chemical shifts. Comparison between the neutral molecule and the ions in vacuum thus identifies the effect of protonation/deprotonation of the functional groups on the initial and final state energies. Analogously, comparison between the ionic forms in vacuum and in solution reveals the effect caused by the solvent polarization screening, in both the initial and final states. For direct comparison between the various forms it is necessary to compare isoelectronic systems, which is why we have chosen to consider the implicit solvent model only (M_1) for the solvated forms. Although the resulting energies are not accurate, we are here primarily interested in qualitatively understanding the origins of the various shifts.

We are now equipped to disentangle the different contributions to the N1s chemical shifts of the three forms of glycine in water. In comparing the zwitterion to the neutral molecule in vacuum, we find that both the initial and final state energies are higher. However, due to the internal Coulomb repulsion associated with a localized +2 charge on the ammonium group in the N1s ionized zwitterion, the final state energy increase is larger than that of the initial state, resulting in a substantial positive chemical shift. When taking the zwitterion from vacuum into solution both the initial and final states are lowered in energy. Due to the total charge of +1 in the final state, the energy is lowered more than in the case of the neutral ground state—this in turns lowers the binding energy. Hence there are two counteracting effects contributing to the total chemical shift upon solvation; the structural rearrangement increases the N1s BE while the polarization screening of the solvent lowers it. Apparently, the first contribution is dominating since the BE of the zwitterion in the aqueous phase is indeed higher compared to the neutral molecule in the gas phase—see Figure 4 and Table 1 for the corresponding experimental results.

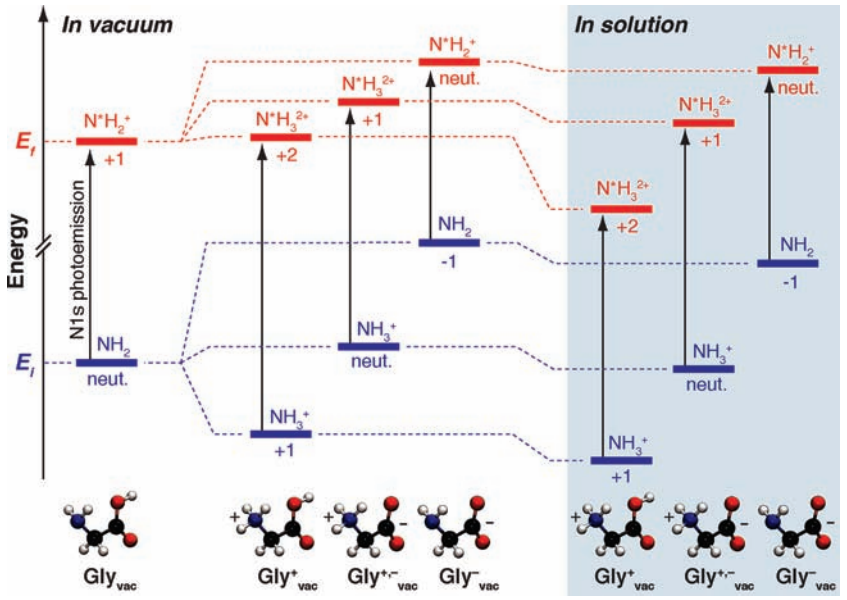


Figure 8. Schematic decomposition of the different contributions to the aqueous glycine N1s shifts relative to gas phase, obtained from DFT calculations (see text for details). The left part shows the initial and N1s ionized final state for the neutral molecule in the gas phase while the corresponding levels for the ionic forms in vacuum are given in the center part. The initial and final state energies of the solvated forms are given on the right-hand side. This scheme allows for disentangling the contribution of structural modifications due to protonation/deprotonation and solvent polarization screening to the total chemical shift.

Turning the neutral molecule into a cation in vacuum drastically lowers the initial-state energy—a result of the increased nuclear potential. The final-state energy is in turn nearly unaffected by the protonation, reflecting the very unfavorable situation of having a doubly charged system unshielded in the gas phase. The situation is reversed in the case of the anion; here the initial state is significantly higher in energy due to the loss of the potential from the departed proton. This also increases the final-state energy, but since the residue is neutral (instead of +2 as in the case of the cation) this increase is not as large as that of the ground state. This means that the protonation/deprotonation of the neutral molecule leads to an increase of the N1s BE for the cation but to a decrease for the anion. In simple continuum models, such as the Born solvation model,[47] the solvation free energy scales with the square of the ionic charge, implying that the sign of the charge has no effect on the polarization screening. Accordingly, placing the anion and cation in solution has very similar effects on the respective ground state. While the cationic N1s ionized state is doubly positively charged, leading to a significant lowering of the energy upon solvation, the corresponding state of the anion is instead neutral which results in very similar final state energies in vacuum and in solution. As previously demonstrated by Weber et al. for photoemission of aqueous alkali-halides,[3] this charge asymmetry in the photoemission process of cations (+1 → +2) compared to anions (−1 → neutral) leads to a negative chemical shift upon solvation of the former case but to an increase for the latter.

A fairly precise picture as to the origins of the N1s chemical shifts of the aqueous forms of glycine is thus emerging. The directions of the shifts, i.e. positive shifts for the cation and zwitterion, while negative for the anion, is primarily a structural effect due to the protonation and deprotonation of the amine and carboxyl functional groups. The polarization screening from the solvent in all cases counteracts this effect - in the cationic case this is primarily a final-state effect, while the anionic solvation shift is dominated by initial-state contributions. With regard to the experimental data in Table 1, we conclude that structural effects caused by protonation/deprotonation of the amine group dominate the N1s chemical shifts of aqueous glycine.

We could similarly analyze the respective energy contributions to the C1s ionization of aqueous glycine, and the procedure is fully analogous to N1s. We are not going to do this here, and instead only discuss the C1s splitting, i.e., the energy difference of the C1s methylene and carboxyl photolines for the different protonation states of glycine. As already mentioned, the methylene C1s BE is consistently lower than that of the carboxyl carbon. This is a consequence of the difference in electronegativity of the ligands; the carboxylic oxygens are more electronegative than the amino nitrogen. In the gas phase the C1s splitting is 2.9 eV. For zwitterionic glycine, this effect is however counteracted by both the negative charge on the carboxylate group (reducing the ligands electronegativity) and the excess proton on the ammonium (increasing its electronegativity). The difference in BE is found to be 2.12 eV for the neutral aqueous solution, and in solid glycine the corresponding shift is 2.1 eV (measured for multilayers and corrected for photoelectron charging).[7] These values are nearly identical because in both cases the molecules are zwitterionic. It is by the same reasoning that the C1s splitting is expected to increase upon decreasing (due to the protonation of the carboxylate group) and increasing pH (due to the deprotonation of the ammunium group), respectively, which is indeed observed experimentally. The corresponding values are 2.66 and 2.55 eV, respectively.

4. CONCLUSIONS

Aqueous glycine has been studied by core-level photoelectron spectroscopy at pH = 1, 7, and 13; it is found that the respective charge state of the molecule is associated with a unique N1s and C1s binding energy. To rationalize the experimental findings, first-principle calculations have been performed for the gas-phase

molecule and for solvated cationic, zwitterionic, and anionic glycine. We have shown that it is important to use a non-equilibrium polarizable continuum model (NEPCM) and to include explicit waters in the first hydration shell to accurately capture the chemical shifts associated with solvation. In the case of zwitterionic aqueous glycine we have further performed spectrum calculations using uncorrelated snapshots from an MD simulation. In addition to significantly improving the results for the N1s chemical shift, these calculations enabled a correlation analysis between structural parameters and resulting binding energies. This provided insights into the origins of the configurational broadening mechanism of the core-level lines and specifically helped to explain why the nitrogen 1s lines of the cationic and zwitterionic forms are broader than that of the anion, to further rationalize the substantially narrower carbon 1s lines. Finally, the various contributions of protonation/deprotonation and solvent induced screening to the resulting chemical shifts of the various ionic forms in solution could be singled out.

In conclusion, the knowledge obtained from the comparison between the calculated and experimental photoelectron spectrum of glycine in the present work represents an important step forward in understanding the origins of core−electron spectral features of solvated biomolecules in general. Our results will thus impact future photoemission studies of complex solutes in aqueous solution.

■ ASSOCIATED CONTENT

Supporting Information. Complete refs 19 and 36 as well as structures (for each charge state where the geometry was optimized separately) with absolute energies. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

niklas.ottosson@fysik.uu.se; knut.borve@kj.uib.no; daniels@mkem.uu.se

■ ACKNOWLEDGMENT

The authors would like to acknowledge financial support by the Knut and Alice Wallenberg Foundation, Swedish Scientific Council (VR), The Swedish Foundation for strategic research (SSF), Göran Gustafsson's Foundation, Carl Trygger's Foundation, the Nordic Research Board, the Meltzer Foundation of the University of Bergen, the NanoUiB program at the University of Bergen, and computing grants from SNIC-UPPMAX under Project p2006027 and from the Norwegian high-performance computing programme (NOTUR). B.W. gratefully acknowledges support by the Deutsche Forschungsgemeinschaft (Project WI 1327/3-1). Ricardo Marinho is acknowledged for his contributions during the early stages of this work, as well as Johan Gråsjö and Jan-Erik Rubensson for many stimulating discussions.

■ REFERENCES

(1) Messer, B. M.; Cappa, C. D.; Smith, J. D.; Wilson, K. R.; Gilles, M. K.; Cohen, R. C.; Saykally, R. J. *J. Phys. Chem. B* **2005**, *109*, 5375.

(2) Gråsjö, J.; Andersson, E.; Forsberg, J.; Duda, L.; Henke, E.; Pokapanich, W.; Björneholm, O.; Andersson, J.; Pietzsch, A.; Hennies, F.; Rubensson, J.-E. *J. Phys. Chem. B* **2009**, *113*, 16002.

(3) Weber, R.; Winter, B.; Schmidt, P. M.; Widdra, W.; Hertel, I. V.; Dittmar, M.; Faubel, M. *J. Phys. Chem. B* **2004**, *108*, 4729.

(4) Winter, B.; Faubel, M. *Chem. Rev.* **2006**, *106*, 1176.

(5) Nolting, D.; Aziz, E. F.; Ottosson, N.; Faubel, M.; Hertel, I. V.; Winter, B. *J. Am. Chem. Soc.* **2007**, *129*, 14068.

(6) Slaughter, A. R.; Banna, M. S. *J. Phys. Chem.* **1988**, *92*, 2165.

(7) Löfgren, P.; Krozer, A.; Lausmaa, J.; Kasemo, B. *Surf. Sci.* **1997**, *370*, 277.

(8) Hasselström, J.; Karis, O.; Weinelt, M.; Wassdahl, N.; Nilsson, A.; Nyberg, M.; Pettersson, L. G. M.; Samant, M. G.; Stöhr, J. *Surf. Sci.* **1998**, *407*, 221.

(9) Nyberg, M.; Hasselström, J.; Karis, O.; Wassdahl, N.; Weinelt, M.; Nilsson, A.; Pettersson, L. G. M. *J. Chem. Phys.* **2000**, *112*, 5420.

(10) Plekan, O.; Feyer, V.; Richter, R.; Coreno, M.; de Simone, M.; Prince, K. C.; Carravetta, V. *J. Phys. Chem. A* **2007**, *111*, 10998.

(11) Plekan, O.; Feyer, V.; Richter, R.; Coreno, M.; de Simone, M.; Prince, K. C.; Carravetta, V. *Chem. Phys. Lett.* **2007**, *442*, 429.

(12) Bergersen, H.; Marinho, R. R. T.; Pokapanich, W.; Lindblad, A.; Björneholm, O.; Saethre, L. J.; Öhrwall, G. *J. Phys.: Condens. Matter* **2007**, 19.

(13) Winter, B. *Nucl. Instrum. Methods Phys. Res., Sect. A* **2009**, *601*, 139.

(14) Hüfner, S. *Photoelectron Spectroscopy*; Springer-Verlag: Berlin Heidelberg, 1995.

(15) Coville, M.; Thomas, T. D. *Phys. Rev. A* **1991**, *43*, 6053.

(16) Winter, B.; Weber, R.; Widdra, W.; Dittmar, M.; Faubel, M.; Hertel, I. V. *J. Phys. Chem. A* **2004**, *184*, 2625.

(17) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327.

(18) Derbel, N.; Hernandez, B.; Pflüger, F.; Liquier, J.; Geinguenaud, F.; Jaidane, N.; Lakhdar, Z. B.; Ghomi, M. *J. Phys. Chem. B* **2007**, *111*, 1470.

(19) Frisch, M. J.; et al. *Gaussian03*, Gaussian, Inc.: Wallingford, CT, 2004.

(20) Dunning, T. H. *J. Chem. Phys.* **1971**, *55*, 716.

(21) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.

(22) Frisch, M. J.; Pople, J. A. *J. Chem. Phys.* **1984**, *80*, 3265.

(23) Dunning, T. H.; Hay, P. J. *Methods of Electronic Structure Theory*; Plenum Press: 1977.

(24) Nose, S. *Mol. Phys.* **1984**, *52*, 255.

(25) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.

(26) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637.

(27) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24.

(28) Barth, E.; Kuczera, K.; Leimkuhler, B. J.; Skeel, R. D. *J. Comput. Chem.* **1995**, *16*, 1192.

(29) Rick, S. W.; Stuart, S. J. *Reviews in Computational Chemistry*, Vol. 18; Wiley-VCH, John Wiley and Sons, Inc.: New York: 2002; p 89.

(30) Mitchell, P. J.; Fincham, D. *J. Phys.: Condens. Matter* **1993**, *5*, 1031.

(31) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

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

(33) *TURBOMOLE* V6.0, 2009, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989−2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com

(34) Basis sets were obtained from the Extensible Computational Chemistry Environment Basis Set Database, Version 10/12/01, as developed and distributed by the Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory which is part of the Pacific Northwest Laboratory, P.O. Box 999, Richland, Washington 99352, USA, and funded by the U.S. Department of Energy. The Pacific Northwest Laboratory is a multiprogram laboratory operated by Battelle Memorial Institue for the U.S. Department of Energy under Contract DE-AC06-76RLO 1830. Contact David Feller or Karen Schuchardt for further information.

(35) Perram, J. W.; Petersen, H. G.; Leeuw, S. W. *Mol. Phys.* **1988**, *65*, 875.
(36) Frisch, M. J.; et al. *Gaussian09*; Gaussian, Inc.: Wallingford, CT, 2009.
(37) Stevens, W. J.; Basch, H.; Krauss, M. *J. Chem. Phys.* **1980**, *72*, 650.
(38) Karlsen, T.; Børve, K. J. *J. Chem. Phys.* **2000**, *112*, 7986.
(39) Jagoda-Cwiklik, B.; Slavicek, P.; Nolting, D.; Winter, B.; Jungwirth, P. *J. Phys. Chem. B* **2008**, *112*, 7355.
(40) We note in passing that the chemical shifts obtained from the structures in Figure 3 at the highest level of theory (B3LYP/aug-cc-pVTZ) using Koopman's theorem, i.e. that BE = $-\varepsilon_n$, where n denotes the orbital energy of the nth orbital, is highly unsatisfactory. As this approach inherently neglects all final state relaxation, resulting in about 1 eV too large shifts for the aqueous phase, we are little surprised that the performance is significantly poorer compared to modeling the core-hole explicitly in the final state of the same geometrical structures. Especially noteworthy is the unacceptable predictions of the splittings between the $C1s_{CH2}$ and $C1_{COO}$ BEs; even in the gas phase the experimental value of 2.9 eV is underestimated by 0.4 eV.
(41) Xu, X.; Goddard, W. A., III. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2673.
(42) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
(43) Holme, A.; Børve, K. J.; Sæthre, L. J.; Thomas, T. D. *To be published*.
(44) Martens, H.; Naes, T. *Multivariate Calibration*; Wiley: 1989.
(45) Höskuldsson, A. *J. Chemometrics* **1988**, *2*, 211.
(46) Sirius ver. 8.1., Pattern Recognition Systems A.S., Bergen, Norway.
(47) Born, M. *Z. Phys.* **1920**, *1*, 45.