J. Phys. Chem. A **1998,** *102,* 3625−3634

Gas-Phase Basicity of Glycine: A Comprehensive ab Initio Study

Kui Zhang and Alice Chung-Phillips*

Department of Chemistry, Miami University, Oxford, Ohio 45056

Received: March 6, 1998

Geometry-optimized structures for the most stable conformers of glycine and protonated glycine were obtained using the Hartree−Fock and second-order Møller−Plesset perturbation (MP2) methods with the 3-21G*, 6-31G*, 6-31G**, 6-31+G**, 6-311G**, 6-311+G**, and 6-311++G** basis sets. Analyses of results indicate that the MP2/6-31G*, MP2/6-31+G**, and MP2/6-311+G** levels of theory are more suited for protonation studies. Considerations were given to potential applications of single-point calculations using higher correlation methods such as MP4, QCISD(T), and CCSD(T) and larger basis sets including 6-311+G-(3df,2p) and aug-cc-pVTZ. An ideal gas basicity of 203.5 kcal/mol at 298.15 K, which was calculated at the MP4/6-31+G(2d,2p) composite level for electronic properties and at the MP2/6-31G* level for thermodynamic properties with corrections of basis-set superposition error and conformational equilibrium effect, is shown to be sufficiently accurate by systematic deductions. This theoretical value is in good agreement with the lower of the two mass spectrometric values, 202.5 and 207.0 kcal/mol, assigned as the gas-phase basicity (GB) of glycine based on two different basicity scales. Comparisons with GB calculations on ammonia and methylamine reveal that certain protonation properties remain fairly constant among molecules undergoing amino N-protonations. Several findings from this study help formulate practical strategies for calculating the GBs of larger molecules, including the use of density functional theory.

Introduction

The acid−base properties of a peptide affect physicochemical activities such as solubility, hydrophobicity, and electrostatic interactions which directly impact the biological activity of the peptide in a living system. The gas-phase basicity (GB) and proton affinity (PA) are the major thermodynamic parameters essential for a quantitative understanding of the intrinsic properties of a peptide in the absence of solvents. While experimental GB measurements are becoming one of the most active areas of research in the gas-phase ion chemistry of biomolecules,[1−10] ab initio calculations have been performed only sparingly to supplement experiments with energetic, thermodynamic, and structural data of the relevant neutral and protonated species.[3,4,8,9] One important function of an accurate theoretical calculation is to provide the thermodynamic $T\Delta S$ term that relates the experimental GB to PA in a protonation reaction.[3,11]

Glycine, the smallest amino acid that forms the backbone of peptides, is uniquely suited to play the role of a model compound in the study of biomolecules. In recent years the geometries and energies of different conformers of gaseous glycine have been extensively studied by experimental and ab initio methods.[12−22] Current advances in mass spectrometric measurements of the GBs of glycine and its peptides have further provided the impetus for theoretical investigations on the protonations of these compounds.[23−25] Most recently we have completed an ab initio conformational analysis of amino N- and carbonyl O-protonated glycine in the gas phase.[26]

As an extension of previous GB studies on glycine, alanine, mixed dipeptides of glycine and alanine, and triglycine (Gly, Ala, GlyGly, GlyAla, AlaGly, and GlyGlyGly),[3,4,8] we present here a comprehensive ab initio study of the GB of glycine calculated from low to high levels of theory. The objective is to assess the ability of each level to calculate this GB accurately. Following those who pioneered ab initio protonation calculations on small molecules,[27−31] we study glycine in regard to the relative influence of basis set, electron correlation, basis-set superposition error, and conformational equilibrium on the accuracy of the calculated GB. Since protonation calculations on larger compounds (e.g., oligopeptides[4] and small carbohydrates[9]) must be limited to the lowest level of theory that yields reliable GBs, we hope this study will provide important information in these regards.

Presently the experimental GB of glycine based on the mass spectrometric measurement by Cassady and co-workers[5,8] may be assigned two different values: 202.5 ± 2.8 kcal/mol by the scale of Lias et al.[5b,32] and 207.0 ± 3.1 kcal/mol by the scale of Meot-Ner et al.[8,33] A recent study by Szulejko et al.[34] suggests that the Lias scale is too low and the Meot-Ner scale is too high for compounds more basic than ammonia. As a result the true GB of glycine is expected to be between 202.5 and 207.0 kcal/mol. In this study we wish to establish a theoretical GB accurate enough to serve as the benchmark for the experimental values. We hope our finding will help decide which basicity scale is more appropriate for assigning the absolute GBs of amino acids and peptides.

Computational Methods

The Gibbs free energy change (ΔG) for the protonation reaction of glycine (Gly) at 298.15 K and 1 atm,

$$\mathrm{Gly} + \mathrm{H}^+ \rightarrow \mathrm{GlyH}^+ \qquad (1)$$

is calculated by means of the following equations:[3]

$$\Delta \mathrm{G} = \Delta H - T\Delta S \qquad (2)$$

$$\Delta H = \Delta E_{\mathrm{e}} + \Delta E_{\mathrm{ZP}} + \Delta(E - E_0) - 1.48\ \mathrm{kcal/mol} \qquad (3)$$

S1089-5639(98)01405-4 CCC: $15.00

$$T\Delta S = (298.15\ \text{K})[S(\text{GlyH}^{+}) - S(\text{Gly})] - 7.76\ \text{kcal/mol} \quad (4)$$

The ΔE_e, ΔE_{ZP}, and $\Delta(E - E_0)$ terms refer to the change Gly → GlyH^+ in electronic energy (E_e), zero-point energy (E_{ZP}), internal energy change ($E - E_0$) from 0 to 298.15 K, and entropy (S) at 298.15 K. The constant 1.48 kcal/mol is the sum of translational energy of H^+ and PV work from the reaction. The constant 7.76 kcal/mol equals the entropy of H^+ at 298.15 K. The gas-phase basicity (GB) and proton affinity (PA) are defined as the negative of the Gibbs free energy change and enthalpy change: GB $= -\Delta G$ and PA $= -\Delta H$.

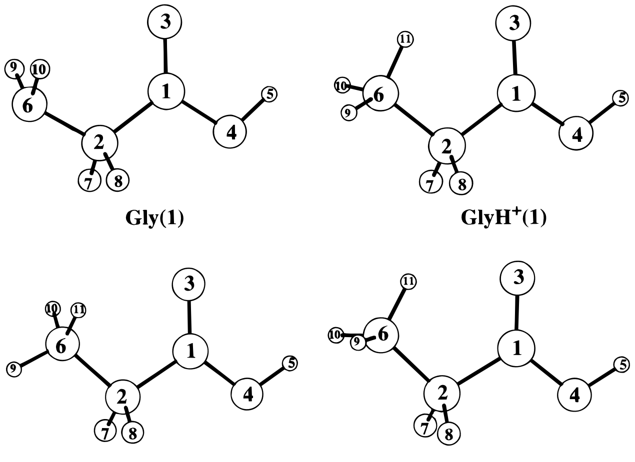
We calculated the GB values based on the most stable conformers Gly(1) and GlyH^+(1) optimized at both Hartree−Fock (HF) and second-order Møller−Plesset perturbation (MP2) levels of theory with standard basis sets ranging from 3−21G to 6-311++G**.[35,36] We also made preliminary investigations on the use of the density functional theory (DFT) for protonation calculations. Although some of the structures optimized at the levels of interest to us were already published,[8,15−17,23−25] the relevant calculations were repeated for systematic evaluations. To improve the calculated electronic energies, higher level calculations including the MP4, quadratic configuration interaction [QCISD(T)], and coupled-cluster [CCSD(T)] methods and basis sets with added polarization functions up to 6-311+G-(2df,2pd) were performed on selected optimized geometries. In addition to electronic energies, harmonic vibrational frequencies were computed at the HF/3-21G*, HF/6-31G*, and MP2/6-31G* optimized levels to provide thermodynamic quantities required by the calculation of ΔG at 298.15 K and 1 atm. To correct for the basis set superposition error, electronic energies of Gly-(1) were recalculated with a "ghost proton" present at selected levels. The contribution of conformational equilibria to ΔG at 298.15 K was estimated by taking into account appropriate stable conformers, Gly(i) and GlyH^+(i) for $i > 1$. Finally, protonation properties for ammonia and methylamine were calculated at levels comparable to or higher than those for glycine to provide data for comparison and discussion; results are presented in Tables S-1 and S-2 as Supporting Information. For these two molecules the largest basis is the augmented correlation-consistent polarized triple-split valence set aug-cc-pVTZ. The ab initio calculations were carried out using the Gaussian 92/94 programs.[36] Both the core and valence electrons were included in all correlation treatments.

Results and Discussion

Structures. The MP2/6-31G* structures for the most stable conformers of glycine and protonated glycine, Gly(1) and GlyH^+(1), and two transition-state conformers of protonated glycine, GlyH^+(T1) and GlyH^+(T2), are shown in Figure 1. Those of the transition states were taken from a companion study[26] to show the different geometries encountered in this study for the global minimum GlyH^+(1).

The neutral species Gly(1) belongs to the point group C_s with atoms N6, C2, C1, O3, O4, and H5 in the symmetry plane; its conformation is characterized by a bifurcated H-bond linking two amino hydrogens to the carbonyl oxygen ($NH_2 \cdots O$). This symmetrical conformation is supported by microwave[12b,20a] and electron diffraction[14] experiments and by ab initio calculations based on HF,[1a,13,15] MP2,[16,17] CCSD(T),[18] and DFT[19] methods.

The conformation of the N-protonated species GlyH^+(1) appears less definitive. While MP2/6-31G* optimizations[25,26] indicate an asymmetric (C_1) geometry with a single H-bond (NH⋯O) for GlyH^+(1) in Figure 1, HF/3-21G and HF/6-31G* optimizations[3,8,23] for the global minimum result in geometries resembling the MP2 transition states GlyH^+(T2) and GlyH^+(T1), respectively, with C_s symmetry of different hydrogen-bonding arrangements. Subsequent HF and MP2 optimizations using basis sets larger than 6-31G* show no change in the respective C_s and C_1 conformations. These changes in the conformation of GlyH^+(1), due to changes in the basis set (from HF/3-21G to HF/6-31G*) and correlation level (from HF/6-31G* to MP2/6-31G*), are in accord with previous findings on certain low-energy conformers of the neutral glycine Gly(i) for $i > 1$.[13,16]



Gly(1) GlyH^+(1)

GlyH^+(T1) GlyH^+(T2)

Figure 1. MP2/6-31G* optimized structures for the most stable conformers of glycine and protonated glycine, Gly(1) and GlyH^+(1), and two transition-state conformers of protonated glycine, GlyH^+(T1) and GlyH^+(T2).

Basis Sets and Correlation Levels. The electronic protonation energy (PE), defined as the negative of ΔE_e, has the same sign and order of magnitude as the calculated GB or PA. A variation in PE can therefore be correlated with a variation in GB or PA. There have been several in-depth studies on the effects of basis set and electron correlation on the calculated PAs of small molecules.[27−30] Here we examine systematically these effects on the PE of glycine.

Electronic and protonation energies of Gly(1) and GlyH^+(1) calculated at different theoretical levels are listed in Table 1. To highlight special trends, 16 PEs are selected for plotting in Figure 2. Each theoretical level is specified by two parts, L2//L1: L1 in column 1 is the level for obtaining the optimized geometry, and L2 in column 2 is the level for calculating E_e. When L2 is the same as L1, the level is referred to as an optimized level.

We first discuss the HF and MP2 optimized levels and the correlated levels of higher order. For geometry optimizations, the basis sets are 3-21G, 6-31G*, 6-31G**, 6-31+G**, 6-311G**, 6-311+G**, and 6-311++G**. These sets are chosen to show a gradual progression in the number and composition of the basis functions. To investigate correlation enhancements beyond MP2, the HF and MP2 geometries are employed as host geometries for MP4, QCISD(T), and CCSD(T) single-point (SP) calculations.

The 6-311++G** basis set is included as a reference for setting the upper limit in basis size for geometry optimizations; previously it has been shown by Császár[17] to yield calculated properties of glycine at the MP2/6-311++G** optimized level in close agreement with experimental values. From Table 1 we find that the HF and MP2 PEs from 6-311++G** differ from those of 6-311+G** by less than 0.1 kcal/mol. We

TABLE 1: Electronic Protonation Energies for the Most Stable Conformer of Glycine[a]

geometry	level	E_e (hartree) Gly(1)	E_e (hartree) GlyH$^+$(1)	PE (kcal/mol)[b]
HF/3-21G	HF/3-21G	−281.247498	−281.616293	231.42 (1)
	HF/6-31G*	−282.827640	−283.183154	223.09
HF/6-31G*	HF/6-31G*	−282.831096	−283.186989	223.33 (2)
	MP2/6-31G*	−283.596549	−283.951430	222.69
	MP4/6-31G*	−283.651270	−284.007657	223.64
	MP2/6-31+G**	−283.685670	−284.036691	220.27
HF/6-31G**	HF/6-31G**	−282.848342	−283.207426	225.33 (3)
HF/6-31+G**	HF/6-31+G**	−282.858138	−283.212177	222.16 (4)
HF/6-311G**	HF/6-311G**	−282.917278	−283.274600	224.22 (5)
HF/6-311+G**	HF/6-311+G**	−282.924826	−283.278526	221.95 (6)
HF/6-311++G**	HF/6-311++G**	−282.925015	−283.278788	222.00
MP2/6-31G*	MP2/6-31G*	−283.619194	−283.975231	223.42 (7)
	MP4/6-31G*	−283.674334	−284.031748	224.28 (12)
	QCISD(T)/6-31G*	−283.671842	−284.029319	224.32
	CCSD(T)/6-31G*	−283.670767	−284.028231	224.31
	MP2/6-31+G**	−283.688689	−284.040490	220.76
MP2/6-31G**	MP2/6-31G**	−283.666296	−284.027498	226.66 (8)
	MP4/6-31G**	−283.723410	−284.086319	227.73 (13)
MP2/6-31+G**	MP2/6-31+G**	−283.689007	−284.040714	220.70 (9)
	MP4/6-31+G**	−283.747521	−284.100912	221.76 (14)
	MP2/6-31+G(2d,2p)	−283.797132	−284.147207	219.68
MP2/6-311G**	MP2/6-311G**	−283.867289	−284.224152	223.93 (10)
	MP4/6-311G**	−283.932429	−284.290976	224.99 (15)
MP2/6-311+G**	MP2/6-311+G**	−283.882985	−284.233184	219.75 (11)
	MP4/6-311+G**	−283.948786	−284.300688	220.82 (16)
	MP2/6-311+G(2d,2p)	−283.960018	−284.309278	219.16
	MP2/6-311+G(2df,2pd)	−284.055550	−284.405171	219.39
	MP2/6-311+G(3df,2p)	−284.067018	−284.415066	218.40
MP2/6-311++G**	MP2/6-311++G**	−283.883500	−284.233836	219.84
B3LYP/6-311++G**	B3LYP/6-311++G**	−284.529605	−284.878845	219.15

[a] For Gly(1) + H$^+$ → GlyH$^+$(1), PE = $-\Delta E_e$, where $E_e(H^+) = 0$ and $\Delta E_e = E_e(\mathrm{GlyH}^+) - E_e(\mathrm{Gly})$. [b] Numbers in parentheses identify the points on the line plots in Figure 2.

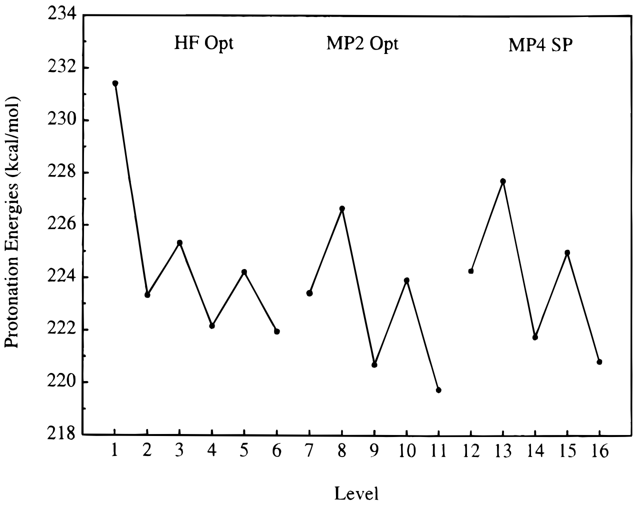


Figure 2. Electronic protonation energies for the most stable conformer of glycine calculated at various theoretical levels. [HF optimized levels: **1**, HF/3-21G; **2**, HF/6-31G*; **3**, HF/6-31G**; **4**, HF/6-31+G**; **5**, HF/6-311G**; and **6**, HF/6-311+G**. MP2 optimized levels: **7**, MP2/6-31G*; **8**, MP2/6-31G**; **9**, MP2/6-31+G**; **10**, MP2/6-311G**; and **11**, MP2/6-311+G**. MP4 single-point levels: **12**, MP2/6-31G*; **13**, MP2/6-31G**; **14**, MP2/6-31+G**; **15**, MP2/6-311G**; and **16**, MP2/6-311+G**. See Table 1 for details.]

therefore accept the smaller 6-311+G** set as the upper basis limit. Likewise, the QCISD(T) and CCSD(T) methods are included as representatives of fairly complete correlation models. We find that the QCISD(T)/6-31G* and CCSD(T)/6-31G* PEs are only 0.04 kcal/mol different from the MP4/6-31G* PE for glycine (Table 1). Additional calculations on ammonia and methylamine yield PE differences of less than 0.4 kcal/mol for MP4 vs QCISD(T) and CCSD(T) using the 6-31G* and 6-311+G** basis sets (Table S-1). On the basis of these small differences, we accept the lower MP4 level as the upper correlation limit.

We next discuss trends exhibited by the PEs as a function of basis set and the correlation method in Figure 2. The 16 selected levels are separated into three groups based on theoretical models: six HF optimized (HF Opt) levels, five MP2 optimized (MP2 Opt) levels, and five MP4 single-point (MP4 SP) levels. In each group the levels are arranged in order of increasing basis size with the corresponding PEs connected in a line plot. On the basis of the work on ammonia by Frisch et al.,[27a] we anticipate that the larger the basis, the lower the PE, and the closer the calculated PA or GB approaches experiment.

The zigzag plots of the three groups show similar patterns. There is indeed a general decrease of PE as the basis expands; but there are several irregularities. The sharpest decline occurs at the beginning of the "HF Opt" group, points 1 → 2, where an 18 kcal/mol drop from 3-21G to 6-31G* is seen. This translates into an 18 kcal/mol greater deviation from experiment for the HF/3-21G value as compared with the HF/6-31G* value. The principal deficiency of 3-21G, as compared with 6-31G*, is obviously the lack of d polarization functions on the oxygen and nitrogen atoms to provide a proper description of these highly electronegative atoms.

We observe in all three groups two similar aberrations in the basis expansions 6-31G* → 6-31G** and 6-31+G** → 6-311G**; both pairs reverse the expected trend of lowering PE. These correspond to points 2 → 3, 4 → 5, 7 → 8, 9 → 10, 12 → 13, and 14 → 15. The rise in PE is about 2 kcal/mol in "HF Opt" and is increased to more than 3 kcal/mol in "MP2 Opt" and "MP4 SP". The disproportionate PEs from the

6-31G** and 6-311G** sets are likely to be caused by the p polarization functions on the hydrogen atoms, which strengthen the N-terminus ionic N−H bonds and hydrogen bonding ($NH_2\cdots O$ or $NH\cdots O$) in GlyH$^+$(1) more than their neutral counterparts in Gly(1). (See Figure 1.) In other words, the basis functions in these two sets are unbalanced with respect to representing ionicity and hydrogen bonding. The addition of sp diffuse functions on the oxygen and nitrogen atoms to form the next larger sets 6-31+G** and 6-311+G** appears to correct this imbalance as the calculated PEs are significantly improved.[29b]

A closer inspection of each line plot reveals two descending paths: (1) the high-PE path, 3-21G → 6-31G** → 6-311G**, corresponding to points 1 → 3 → 5, 8 → 10, and 13 → 15; and (2) the low PE path, 6-31G* → 6-31+G** → 6-311+G** for points 2 → 4 → 6, 7 → 9 → 11, and 12 → 14 → 16. As the latter employs diffuse functions on non-hydrogens favorable to protonation calculations,[28,29b] we use the low-PE path in the subsequent GB calculations.

Following the low-PE paths of the “HF Opt” and “MP2 Opt” groups, we note that the MP2 PEs become increasingly lower than the HF PEs of the same basis as the basis size increases. In other words, when the basis is larger, incorporating electron correlation in the theoretical model makes a greater impact on PE. Since a MP2 optimized level with a basis larger than 6-31G* yields a lower PE than its HF counterpart, it is more advantageous to pursue MP2 optimizations for bases larger than 6-31G*.

Comparisons of the five “MP2 Opt” PEs of glycine to the corresponding five “MP4 SP” values show a nearly constant increase of ca. 1 kcal/mol in PE on going from MP2 to MP4. Further support is provided by the PE results shown for the protonations of NH_3 and CH_3NH_2 in Table S-1. It is important to note that enhancement in the correlation level from MP2 to MP4 is not affected much by molecular size, basis size, and hydrogen bonding.

We last discuss the use of “composite level” to obtain more accurate electronic energy by properly combining calculations that improve both correlation level and basis set. [See, for example, the G2(MP2)[37] and the focal-point (fp)[38] procedures.] Briefly, energy at the composite level "M2/B2" is deduced from energies calculated at the M2/B1, M1/B2, and M1/B1 levels on the same geometry:

$$E_e(\text{“M2/B2”}) = E_e(\text{M2/B1}) + E_e(\text{M1/B2}) - E_e(\text{M1/B1}) \tag{5}$$

Here M2 is a correlated level higher than M1, and B2 is a basis set larger than B1. The cost for E_e(“M2/B2”) is much lower than E_e(M2/B2) from a direct M2/B2 calculation at the same geometry. In this study MP4 is the only higher correlated level (M2) taken into consideration. The selection of the extended bases (B2) is to be expounded below.

In previous studies[28,29b] it has been shown that PEs computed with double-split valence functions may be brought to better agreement with experiments by including diffuse functions, but reliable values are assured only when triple-split valence basis sets are used. Furthermore, the standard bases used as a benchmark for obtaining accurate PAs are on the order of 6-311+G(2df,2pd) and 6-311++G(3df,3pd) at MP2 or higher correlated levels. To test PE convergence, four basis sets extended from 6-31+G** and 6-311+G** with additional polarizations functions are investigated. Relevant energy quantities resulting from MP2 SP calculations using the new sets are shown in Table 1. The PEs of double-split valence sets 6-31+G** and 6-31+G(2d,2p) at 220.70 and 219.68 kcal/mol give a hint of rapid convergence. The PEs of triple-split valence sets 6-311+G**, 6-311+G(2d,2p), 6-311+G(2df,2pd), and 6-311+G(3df,2p) at 219.75, 219.16, 219.39, and 218.40 kcal/mol show quite convincingly that convergence is nearly at hand. Lacking the computing power to test even larger basis sets, the 6-31+G(2d,2p) and 6-311+G(3df,2p) bases are tentatively taken as upper limits for the respective double-split and triple-split valence basis sets in PE calculations.

TABLE 2: Thermodynamic Properties for the Most Stable Conformers of Glycine and Protonated Glycine[a]

property	level[c]	Gly(1)	GlyH$^+$(1)	Gly(1) → GlyH$^+$(1)
E_{ZP} (kcal/mol)	HF/3-21G	53.134	63.066	9.932
	HF/6-31G*	54.381	64.211	9.830
	MP2/6-31G*	51.132	60.310	9.178
$E - E_0$ (kcal/mol)	HF/3-21G	3.412	3.338	−0.074
	HF/6-31G*	3.341	3.442	0.101
	MP2/6-31G*	3.464	3.496	0.032
S (cal/(K mol))	HF/3-21G	73.743	73.347	−0.396
	HF/6-31G*	73.470	74.894	1.424
	MP2/6-31G*	74.488	74.935	0.447
G_{therm} (kcal/mol)[b]	HF/3-21G	30.556	39.673	9.117
	HF/6-31G*	31.705	40.362	8.656
	MP2/6-31G*	31.170	39.922	8.752

[a] All quantities at 1 atm; E_{ZP} at 0 K; and E, S, and G_{therm} at 298.15 K. The last column presents the difference in the calculated property: $\Delta M = M(\text{GlyH}^+) - M(\text{Gly})$. [b] $G_{\text{therm}} = E_{ZP} + (E - E_0) + RT - TS$, where E_{ZP} is scaled by 0.9135 for HF/3-21G and HF/6-31G* and 0.9646 for MP2/6-31G*. [c] Optimized level.

Geometries. The geometrical parameters corresponding to the optimized levels (column 1 of Table 1) are provided in Tables S-3 and S-4 for the neutral and protonated glycine as Supporting Information. For the neutral species, geometries obtained previously have been thoroughly discussed;[13,16−18] our results complement those published by Schäfer et al.[12a,13] and by Hu et al.[18] On the other hand, geometries for the protonated species are relatively new; the MP2 results in particular are important reference material for this emerging field of gas-phase ion chemistry. In view of the importance of these species as model compounds and the popularity of the chosen basis sets in computational studies of medium-sized organic molecules, a detailed discussion on the effects of the basis set and electron correlation on calculated geometries is presented in Appendix S in the Supporting Information.

Thermodynamic Properties. The thermodynamic properties E_{ZP}, $E - E_0$, and S of the neutral and protonated glycine, calculated with the two smallest basis sets (3-21G and 6-31G*), are presented in Table 2. The thermal contribution to Gibbs free energy, G_{therm}, a collective term that includes enthalpy and entropy (footnote *b* of Table 2), is also shown for the individual species. The quantity directly entering the GB calculation is the ΔG_{therm} term for the change Gly(1) → GlyH$^+$(1) in the last column.

In calculating G_{therm}, E_{ZP} is scaled by the factor 0.9135 for the HF levels and 0.9646 for the MP2 levels. These two factors were derived by Pople et al.[39] from fitting the zero-point energies based on harmonic frequencies calculated from the HF/6-31G* and MP2/6-31G* levels to the experimental zero-point energies. The scaling factor compensates for the inherent deficiencies in the theoretical model with respect to anharmonicity, electron correlation, and basis set. The $E - E_0$ and S terms in G_{therm}, however, are calculated with harmonic frequencies without corrections, as their contributions to GB are minor.[3]

In the past we estimated the thermodynamic properties either at the same level as or at a level below the level used for

TABLE 3: Gas-Phase Basicities and BSSE Corrections for the Most Stable Conformer of Glycine: Comparisons with Experiments[a,b]

geometry	level[c]		GB[d]	BSSE	GB(B)
HF/3-21G	HF/3-21G		216.02	3.65	212.37
	HF/6-31G*		207.69	0.57	
HF/6-31G*	HF/6-31G*		208.39	0.46	207.93
	MP2/6-31G*		207.75 (1)		
	MP4/6-31G*		208.70 (2)		
	MP2/6-31+G**		205.33 (3)		
	"MP4/6-31+G**"		206.28 (4)		
MP2/6-31G*	MP2/6-31G*		208.39 (5)	1.89	
	MP4/6-31G*		209.25 (6)	2.04	
	MP2/6-31+G**		205.73 (7)	2.90	
	"MP4/6-31+G**"	(*w*)	206.59 (8)	3.05	203.54
MP2/6-31+G**	MP2/6-31+G**		205.67 (9)	2.91	
	MP4/6-31+G**		206.73 (10)	3.01	
	MP2/6-31+G(2d,2p)		204.65 (11)	2.44	
	"MP4/6-31+G(2d,2p)"	(*x*)	205.71 (12)	2.54	203.17
MP2/6-311+G**	MP2/6-311+G**		204.72 (13)	2.77	
	MP4/6-311+G**		205.79 (14)	2.92	
	MP2/6-311+G(2d,2p)		204.13	1.99	
	MP2/6-311+G(2df,2pd)		204.36	1.95	
	MP2/6-311+G(3df,2p)		203.37 (15)	1.11	
	"MP4/6-311+G(3df,2p)"	(*y*)	204.44 (16)	1.26	203.18
		Gas-Phase Basicity			
GB(B,C)[e]	"MP4/6-31+G**"	(*w*)			203.9
	"MP4/6-31+G(2d,2p)"	(*x*)			203.5
	"MP4/6-311+G(3df,2p)"	(*y*)			203.5
experimental	NIST[f]				203.7
	mass spectrometry[g]				202.5 ± 2.8
		Proton Affinity			
PA(B,C)	"MP4/6-31+G(2d,2p)"	(*x*)			211.1
experimental	NIST[f]				211.8

[a] For Gly(1) + $H^+ \rightarrow GlyH^+(1)$ at 298.15 K and 1 atm. All quantities are in kcal/mol. [b] GB = PE $-$ ΔG_{therm} $-$ 6.28; GB(B) = GB $-$ BSSE; GB(B,C) = GB(B) $-$ CEE; PA = GB $-$ $T\Delta S$; PA(B) and PA(B,C) are defined similarly. The PE and ΔG_{therm} values are from Tables 1 and 2. [c] Quotes denote a composite level. Letter in parentheses identifies a level of special interest. [d] Numbers in parentheses identify the points on the line plots in Figure 3. [e] The CEE and $T\Delta S$ values are -0.34 and -7.63, respectively. [f] Reference 32c. [g] Reference 5b.

calculating E_e. For example, ΔG_{therm} at HF/3-21G, but E_e at HF/6-31G* or higher, was employed for simple peptides and glucose.[4,8,9] It is therefore instructive to compare the three sets of thermodynamic properties for glycine. Note especially that the ΔG_{therm} values become comparable at all three levels after scaling ΔE_{ZP}: the HF/3-21G value is ca. 0.5 kcal/mol higher than the two 6-31G* values, whereas the HF/6-31G* and MP2/6-31G* values agree to within 0.1 kcal/mol. In the subsequent GB calculations, the ΔG_{therm} value derived from a particular geometry is applied to levels involving the same geometry. When a geometry is optimized with a basis larger than 6-31G*, the ΔG_{therm} value derived for the 6-31G* geometry is used for the larger calculation to cut cost.

Similar calculations for ammonia and methylamine, at the three levels for glycine plus the higher levels MP2/6-31+G** and MP2/6-311+G**, show close agreement in the ΔG_{therm} values (Table S-2). These results indicate that with proper scaling the ΔG_{therm} value calculated at any level is about the same for a given molecule. The major reason for the very small differences in ΔG_{therm} from different levels is due to the very small changes in the vibrational frequencies based on very slightly changed geometries. This reason justifies in part the use of ΔG_{therm} from a lower level as a cost-cutting measure.

Another worthy observation is a fairly constant ΔG_{therm} from one molecule to another (Tables 2 and S-2). The average ΔG_{therm} in kcal/mol, over the HF/6-31G* and MP2/6-31G* values, is 8.7 for glycine, 9.2 for methylamine, and 9.8 for ammonia. Thus, ΔG_{therm} may be considered a "transferrable" protonation property. In the event that direct calculations of vibrational frequencies become too costly for an amino N-protonation study, an approximate value of 9 kcal/mol deduced from glycine and methylamine may be used.

Gas-Phase Basicity. For the protonation reaction 1, a more compact way to express ΔG than eqs 2−4 is

$$\Delta G = \Delta E_e + \Delta G_{\text{therm}} + 6.28 \text{ kcal/mol} \qquad (6)$$

where

$$\Delta G_{\text{therm}} = \Delta E_{\text{ZP}} + \Delta(E - E_0) - (298.15 \text{ K})[S(\text{GlyH}^+) - S(\text{Gly})] \qquad (7)$$

The major component of ΔG is ΔE_e (Table 1), which is about 20-fold larger than ΔG_{therm} (Table 2). At ordinary temperatures (around 300 K) the accuracy of the calculated ΔG depends more on the accuracy of the calculated ΔE_e than ΔG_{therm}.

The GBs are presented in Table 3 for selected levels. The first three levels are of practical interest to ab initio calculations of larger molecules. The HF/3-21G optimized level yields an unacceptably high GB of 216.02 kcal/mol, but the next SP level, HF/6-31G*//HF/3-21G*, results in a much improved 207.69 kcal/mol. The latter is only 0.30 kcal/mol different from the GB of the HF/6-31G* optimized level. The similarity between the two HF/6-31G* values, despite the noticeable difference between the HF/3-21G and HF/6-31G* geometries of $GlyH^+(1)$ at the ammonium terminus, is favorable to GB calculations of larger molecules, as 3-21G may be the only affordable basis for the geometry optimization step.

To explore special trends, 16 GBs corresponding to four composite levels and their components are plotted in Figure 3.

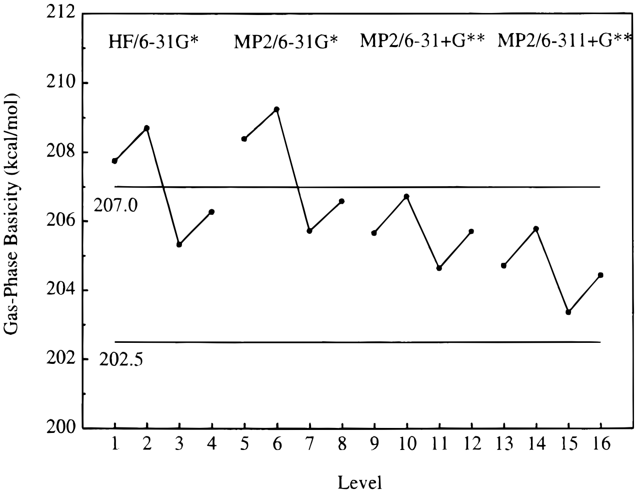


Figure 3. Gas-phase basicities for the most stable conformer of glycine calculated at various theoretical levels. [HF/6-31G* geometries: **1**, MP2/6-31G*; **2**, MP4/6-31G*; **3**, MP2/6-31+G**; and **4**, "MP4/6-31+G**". MP2/6-31G* geometries: **5**, MP2/6-31G*; **6**, MP4/6-31G*; **7**, MP2/6-31+G**; and **8**, "MP4/6-31+G**". MP2/6-31+G** geometries: **9**, MP2/6-31+G**; **10**, MP4/6-31+G**; **11**, MP2/6-31+G(2d,2p); and **12**, "MP4/6-31+G(2d,2p)". MP2/6-311+G** geometries: **13**, MP2/6-311+G**; **14**, MP4/6-311+G**; **15**, MP2/6-311+G(3df,2p); and **16**, "MP4/6-311+G(3df,2p)". See Table 4 for details. Horizontal straight lines: experimental values based on two different basicity scales.]

The two experimental values are drawn as two horizontal lines to show the high and low limits of the absolute GB. Theoretical levels selected for plotting are separated into four groups based on geometries: HF/6-31G*, MP2/6-31G*, MP2/6-31+G**, and MP2/6-311+G**. In each group, the four points follow the same sequence M1/B1, M2/B1, M1/B2, and "M2/B2" as defined above for eq 5. Note that all four line plots have the tilted "Z" pattern. Each "Z" consists of four points which follow three general trends: (a) a nearly constant rise of ca. 1 kcal/mol from point 1 to point 2 for the change in correlated level MP2 → MP4 with the same basis, indicating greater correlation stabilizes $GlyH^+$ more than Gly; (b) a drop of 1−3 kcal/mol from point 1 to point 3 for the change in basis B1 → B2 at the same MP2 level, showing a larger basis stabilizes Gly more than $GlyH^+$; (c) a cancellation of the two opposing effects from upgrading correlation and basis yields an overall decrease in GB from point 1 to point 4 for the combined change, (MP2, B1) → (MP4, B2). An interesting consequence of trend (c) is the nearly identical GBs of MP2/6-31+G** and "MP4/6-31+G(2d,2p)" in the third plot (points 9 and 12) and the very similar GBs of MP2/6-311+G** and "MP4/6-311+G(3df,2p)" in the fourth plot (points 13 and 16). A shortcut, therefore, is to take the GBs of MP2/6-31+G** and MP2/6-311+G** optimized levels (points 9 and 13) to represent those derived from the respective higher composite levels (points 12 and 16) without going through the additional MP4 SP and MP2 SP calculations.

The similarity in the first two plots, which correspond to the same composite level "MP4/6-31+G**" but based on two different geometries (HF/6-31G* and MP2/6-31G*), simply points out the possible advantage of using a lower level geometry to achieve nearly the same improvement. This similarity, of course, depends on how "similar" the two geometries are. Tables S-3 and S-4 show that the HF/6-31G* and MP2/6-31G* geometries are reasonably similar except for the conformation of the ammonium terminus, which obviously has an insignificant impact on the overall energy of the protonated species.

The GBs of the three composite levels "MP4/6-31+G**" (*w*), "MP4/6-31+G(2d,2p)" (*x*), and "MP4/6-311+G(3df,2p)" (*y*) are 206.59, 205.71, and 204.44 kcal/mol. The values show convergence to lower GB as the basis expands at the MP4 level. There is a clear signal that the theoretical GB is heading toward the lower of the two experimental values, 202.5 kcal/mol.

The procedure for computing GB as described above is incomplete owing to its neglect of the basis set superposition error (BSSE) and the conformational equilibrium effect (CEE). In the following two sections we investigate BSSE and CEE as they apply to the GB calculation of glycine. These are corrections that can be significant in principle but have not been applied explicitly to glycine and larger molecules.

BSSE Correction. In previous PA studies on small molecules such as CH_4, NH_3, H_2O, HCN, and CO_2, it has been demonstrated that BSSE is a significant quantity that should be taken into account when high accuracy is desired.[28,29a,30] In this glycine study the BSSE corrections are estimated for the majority of the levels (Table 3). The E_e of neutral glycine is computed with a "ghost proton" present, i.e., using the basis set for the protonated glycine:

$$\mathrm{BSSE} = E_e(\text{Gly in neutral basis}) - E_e(\text{Gly in protonated basis}) \quad (8)$$

After correcting for BSSE, the change in the electronic energy for Gly → $GlyH^+$ becomes

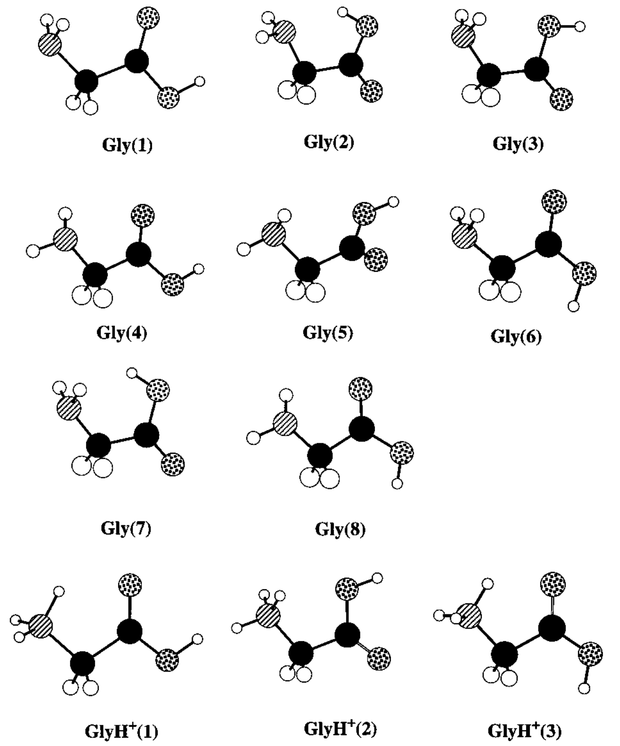
$$\Delta E_e + \mathrm{BSSE} = E_e(\text{GlyH}^+ \text{ in protonated basis}) - E_e(\text{Gly in protonated basis}) \quad (9)$$

Geometrical parameters for the ghost atom "H11" of Gly(1) are derived from optimizations on the ammonium ion (NH_4^+) at the relevant theoretical levels. Details are provided in Appendix S in the Supporting Information.

Results shown in Table 3 indicate that BSSE at the HF level is large for the 3-21G basis (~4 kcal/mol) but small for 6-31G* or larger basis (<0.6 kcal/mol). On the other hand, BSSE values at the MP2 and MP4 levels are similar for a given basis (around 2−3 kcal/mol). The data also show that BSSE at a given correlated level decreases with increasing basis size. At the MP2 level, for example, BSSE decreases as the basis expands: 2.77 → 1.99 → 1.95 → 1.11 kcal/mol for 6-311+G** → 6-311+G(2d,2p) → 6-311+G(2df,2pd) → 6-311+G(3df,2p) at the MP2/6-311+G** geometry.

Finally, the BSSE values for the three composite levels *w*, *x*, and *y* are 3.05, 2.54 and 1.26 kcal/mol, respectively, which show convergence to lower corrections as the basis expands at the MP4 level. After correcting the GB of Table 3 with BSSE, i.e., GB(B) = GB − BSSE, the GB(B) values for the three composite levels become almost equal, i.e., 203.54, 203.17, and 203.18 kcal/mol (Table 3). Based on GB(B), level *x*, which is "MP4/6-31+G(2d,2p)"//MP2/6-31+G**, is taken as the best practical level for the GB calculation since it represents the lowest level at which GB(B) converges.

For comparison with glycine, the BSSE results for ammonia and methylamine are presented in Table S-1. Note the excellent agreement between the BSSE values of all three molecules at any given theoretical level. For example, the BSSEs are 2.44, 2.49, and 2.38 kcal/mol, respectively, for glycine, methylamine, and ammonia at the MP2/6-31+G(2d,2p) level. This agreement seems to suggest that the BSSE correction to the GB of an $R{-}NH_2$ molecule with regard to N-protonation is nearly independent of the type and size of substitute R. If this is the case, the BSSE corrections obtained here may be used as



Gly(1) **Gly(2)** **Gly(3)**

Gly(4) **Gly(5)** **Gly(6)**

Gly(7) **Gly(8)**

GlyH$^+$(1) **GlyH$^+$(2)** **GlyH$^+$(3)**

Figure 4. MP2/6-31G* optimized structures for low-energy stable conformers of glycine and protonated glycine, Gly(i), $i = 1, ..., 8$, and GlyH$^+$(i), $i = 1, ..., 3$.

correction constants in future GB calculations for amines, amino acids, and peptides.

CEE Correction. The CEE contribution to the ΔG of protonation may be defined as

$$\mathrm{CEE} = \langle G \text{ of GlyH}^+\rangle - \langle G \text{ of Gly}\rangle \tag{10}$$

where $\langle G\rangle$ represents the average Gibbs free energy over $G(i)$ of contributing conformers i relative to $G(1)$ of the lowest energy conformer 1:

$$\langle G\rangle = \sum_i p(i)\, G(i) \tag{11}$$

The equilibrium population $p(i)$ for each conformer i at temperature T is calculated by a Boltzmann expression[15]

$$p(i) = \frac{\mathrm{e}^{-G(i)/RT}}{\sum_i \mathrm{e}^{-G(i)/RT}} \tag{12}$$

To our knowledge, the potentially important contribution of CEE has not been explicitly examined in previous protonation studies. To study this effect on glycine, the $p(i)$ values for the eight neutral species and three N-protonated species shown in Figure 4 were evaluated using eqs 10−12 with the MP2/6-31G* values for E_e, E_{ZP}, $E - E_0$, and S (Table S-5). Those for Gly(6), Gly(7), Gly(8), and GlyH$^+$(3) conformers turned out to be negligible. We then eliminated these four conformers and searched for the energy minima of the remaining conformers at the higher MP2/6-31+G** level.

The $E_e(i)$ of the five neutral and two protonated glycine conformers of the lowest energies, requisite for the equilibrium calculations at the "MP4/6-31+G(2d,2p)"//MP2/6-31+G** level, are presented in Table 4. Note that BSSEs for individual neutral conformers are not explicitly included in $G(i)$ (footnote *a* of Table 4) on the assumption that they are all similar to the BSSE of Gly(1). The resulting $p(i)$ values depict an equilibrium composition of ~70% Gly(1) and ~20% Gly(3) for the neutral glycine and roughly 100% GlyH$^+$(1) for the protonated glycine. Using the $p(i)$ and $G(i)$ values of Table 4, we obtained 0.347 kcal/mol for $\langle G$ of Gly$\rangle$ but only 0.006 kcal/mol for $\langle G$ of GlyH$^+\rangle$ because of the dominant presence of GlyH$^+$(1) at room temperature. From eq 10, the CEE correction to the GB of glycine is determined to be −0.341 kcal/mol, a relatively small quantity.

TABLE 4: Electronic Energies, Gibbs Free Energies, and Equilibrium Populations for Low-Energy Conformers of Glycine and Protonated Glycine Calculated at the MP2/6-31+G Geometries[a]**

conformer (i)	MP2/6-31+G** $E_e(i)$	MP4/6-31+G** $E_e(i)$	MP2/6-31+G(2d,2p) $E_e(i)$
Gly(1)	0.000	0.000	0.000
Gly(2)	0.726	0.869	0.392
Gly(3)	1.491	1.459	1.588
Gly(4)	1.323	1.369	1.307
Gly(5)	2.194	2.187	2.485
GlyH$^+$(1)	0.000	0.000	0.000
GlyH$^+$(2)	3.695	3.632	4.442

conformer	"MP4/6-31+G(2d,2p)"		
(i)	$E_e(i)$	$G(i)$	$p(i)$
Gly(1)	0.000	0.000	68.0
Gly(2)	0.535	1.288	7.7
Gly(3)	1.556	0.802	17.6
Gly(4)	1.353	1.451	5.9
Gly(5)	2.478	2.610	0.8
GlyH$^+$(1)	0.000	0.000	99.9
GlyH$^+$(2)	4.379	3.883	0.1

[a] Units: E_e and G in kcal/mol; p in %. G and p are at 298.15 K and 1 atm. Relative to $G(1) = 0.000$ for Gly(1) and GlyH$^+$(1), $G(i) = E_e(i) + G_{\mathrm{therm}}(i)$, where E_e in this table and G_{therm} in Table S-5 are used.

The CEE correction generally differs from molecule to molecule. For small molecules such as ammonia and methylamine, there is one energy minimum for the neutral or protonated species. For amino acids more complex than glycine, conformers analogous to Gly(2), Gly(3), and GlyH$^+$(2) may or may not be important contributors to CEE because of the side chain at C2 (Figures 1 and 4). In peptides, conformers with hydrogen bonding between the amino or ammonium group and the carbonyl or hydroxyl group located at two or more residues away must involve ringlike structures larger than the five-membered type in glycine. Examples of different conformational possibilities for small peptides can be found in several recent publications.[3,4,8,40−43] More importantly, conformers of a protonated peptide are expected to be closer in energy to one another than the special case of glycine, for which GlyH$^+$(2) is some 4 kcal/mol higher than GlyH$^+$(1). In diglycine, an example has been found for two amino N-protonated species being less than 0.5 kcal/mol apart.[3] In larger compounds, therefore, we expect contributions made by protonated conformers to CEE to be as significant as those by neutral conformers, and a cancellation of these two contributions results in a minimal overall CEE correction to GB.

Comparisons with Experiments. After correcting the GB of Table 3 with BSSE and CEE, i.e., GB(B,C) = GB − BSSE − CEE, the GB(B,C) values for the three composite levels of interest are shown in Table 3. The ideal-gas basicity of glycine is 203.5 kcal/mol at the best practical level, "MP4/6-31+G(2d,2p)"//MP2/6-31+G**. Based on the body of evidence gathered in this investigation, we believe this value to be

TABLE 5: Selected Electronic Protonation Energies and BSSE Corrections for Ammonia and Methylamine: Comparisons with Experiments[a]

			NH_3			CH_3NH_2		
geometry	level		PE	BSSE	PE(B)	PE	BSSE	PE(B)
MP2/6-311+G**	MP4/6-311+G**	(*Mx*)	214.76	3.34	211.42	225.27	3.15	222.12
	QCISD(T)/6-311+G**	(*My*)	214.99	3.31	211.68	225.57	3.14	222.43
	CCSD(T)/6-311+G**	(Mz)	215.03	3.31	211.72	225.61	3.12	222.49
MP2/6-31+G**	MP2/6-31+G(2d,2p)	(*Bx*)	211.91	2.38	209.53	223.17	2.49	220.68
MP2/6-311+G**	MP2/6-311+G(3df,2p)	(*By*)	210.78	1.18	209.60	221.87	1.18	220.69
	MP2/aug-cc-pVTZ	(*Bz*)	211.26	1.62	209.64	222.49	1.77	220.72
MP2/6-31+G**	"MP4/6-31+G(2d,2p)"	(*x*)	213.05	2.50	210.55	224.33	2.60	221.73
MP2/6-311+G**	QCISD(T)/6-311+G(3df,2p)	(*y*)	211.97	1.14	210.83	223.27	1.14	222.13
	CCSD(T)/aug-cc-pVTZ	(*z*)	212.45	1.49	210.96	223.94	1.69	222.25
			Gas-Phase Basicities					
GB(B)[b]	"MP4/6-31+G(2d,2p)"	(*x*)			194.6			206.3
	QCISD(T)/6-311+G(3df,2p)	(*y*)			194.8			206.7
	CCSD(T)/aug-cc-pVTZ	(*z*)			194.9			206.8
experimental[c]	NIST				195.8			206.6
			Proton Affinities					
PA(B)	"MP4/6-31+G(2d,2p)"	(*x*)			202.8			213.8
	QCISD(T)/6-311+G(3df,2p)	(*y*)			203.0			214.3
	CCSD(T)/aug-cc-pVTZ	(*z*)			203.2			214.4
experimental	NIST				204.1			214.9

[a] All quantities are in kcal/mol. PE, BSSE, and thermodynamic data are from Tables S-1 and S-2. PE(B) = PE − BSSE. GB and PA are calculated for 298.15 K and 1 atm. See footnotes *b* and *c* of Table 3. [b] Level *x* is the same as level *x* for glycine in Table 3; MP2/6-31G* values for $\Delta G_{\mathrm{therm}}$ and $T\Delta S$ are 9.71 and −8.22 for NH_3 and 9.06 and −7.53 for CH_3NH_2. For levels *y* and *z*, MP2/6-311+G** values for $\Delta G_{\mathrm{therm}}$ and $T\Delta S$ are 9.75 and −8.22 for NH_3 and 9.13 and −7.59 for CH_3NH_2. [c] Reference 32c.

sufficiently accurate to serve as a benchmark for the experimental GB values. Clearly, this theoretical GB favors the lower mass spectrometric GB assigned by the Lias scale (202.5 kcal/mol).[5b] It is in excellent agreement with the evaluated experimental GB of 203.7 kcal/mol by Hunter and Lias published recently by the National Institute of Standards and Technology (NIST).[32c] Furthermore, the respective theoretical GBs of ammonia and methylamine at the same level, after corrected for BSSE, are 194.6 and 206.4 kcal/mol, in good agreement with the NIST values of 195.8 and 206.6 kcal/mol (Table 5 below).[32c] Using the MP2/6-31G* value for the $T\Delta S$ term, the corresponding PAs are calculated to be 211.1, 213.9, and 202.8 kcal/mol for glycine, methylamine, and ammonia, respectively, which are in reasonable agreement with the NIST values of 211.8, 214.9, and 204.1. A brief discussion on error limits is presented in the last section.

Assuming the theoretical GB of glycine (203.5 kcal/mol) is the true GB, the mass spectrometric GB based on the Meot-Ner scale (207.0 kcal/mol) is 3.5 kcal/mol too high. In our previous publications on the GBs of glycine, alanine, and their peptides,[3,4,8] the experimental GBs were assigned according to the Meot-Ner scale and therefore were too high. The calculated GBs were of the HF/6-31G* variety without BSSE and CEE corrections and therefore were also too high. The good agreement (in absolute magnitude) found earlier between the experimental and calculated values for these compounds is therefore coincidental.

Large Molecules. The data presented above for glycine may be used to set up a strategy for obtaining reasonably reliable GBs of peptides. We follow a procedure suggested previously[8,9] to bring the GB of a large molecule calculated at a "low level" to a value comparable to that of a "high level" by using a correction factor. The factor equals the GB difference between these two levels calculated for a related but smaller compound (i.e., a model compound). Suppose the "low level" refers to ΔE_{e} at HF/6-31G* and $\Delta G_{\mathrm{therm}}$ at HF/3-21G, while the "high level" corresponds to ΔE_{e} at "MP4/6-31+G(2d,2p)" and $\Delta G_{\mathrm{therm}}$ at MP2/6-31G*. Using glycine as the model compound the "low level" GB is 207.9 kcal/mol and the "high level" GB is 205.7 kcal/mol, which yield a GB correction of −2.2 kcal/mol (Tables 1 and 2). This correction may be supplemented by the BSSE change from HF/6-31G* to "MP4/6-31+G(2d,2p)" for a total GB(B) correction of −4.3 kcal/mol (Table 3). The factor, −4.3 kcal/mol, can now be used to correct the published "low-level" GBs of GlyGly, GlyAla, AlaGly, and GlyGlyGly[8] to improve their agreement with the mass spectrometric GBs adjusted to the new Lias scale.[32c] In this procedure, the error lies mainly in the nontransferability of the ΔE_{e} change from HF/6-31G* to MP2/6-31+G(2d,2p) owing to different molecular size and hydrogen bonding, as the MP2 to MP4, BSSE, and $\Delta G_{\mathrm{therm}}$ changes are intrinsically transferrable from one molecule to another. The strategy as described may be adapted for other types of protonation.

Density Functional Theory. DFT has been applied successfully to the conformational analysis[19] and interpretation of electron momentum spectroscopy experiments[22] on glycine. To explore the potential use of DFT for protonation studies, we applied the DFT (B3LYP) procedure[44] with the 6-311++G** basis set for geometry optimizations of the lowest energy conformers of Gly and $GlyH^+$. The resulting B3LYP/6-311++G** geometrical parameters are in reasonable agreement with those of MP2/6-311++G** (Tables S-3 and S-4), except for the dihedral angles of the ammonium hydrogen atoms in the protonated species. In fact, B3LYP/6-311++G** produces a $GlyH^+$(T2) conformation for the global minimum similar to HF/3-21G (Figure 1). The encouraging result is the DFT protonation energy, 219.15 kcal/mol, which is in good agreement with the MP2 PE (219.84 kcal/mol) and superior to the HF PE (222.00 kcal/mol) evaluated from the same 6-311++G** basis set (Table 1). Similar conclusions are reached for the B3LYP/6-311+G** PEs of NH_3 and CH_3NH_2 as compared with their MP2 and HF counterparts (Table S-1). In view of the economy of using DFT for geometry optimizations, as opposed to MP2, we recommend DFT (B3LYP) for protonation calculations of large molecules.

Correlation-Consistent Sets. In several benchmark calculations on small molecules Dunning and co-workers showed that the convergence behavior of the correlation-consistent basis sets cc-pV*n*Z and aug-cc-pV*n*Z ($n = 2-5$) is remarkably systematic. Moreover, the "complete basis set" estimated from these sets yields calculated properties in excellent agreement with experiments when used with a high-level correlation method such as CCSD(T).[45] For this protonation study, consider CCSD(T)/aug-cc-pVTZ (z) as the ideal level where the basis is augmented with more diffuse functions than a standard basis such as 6-311++G**.[29c] To estimate the differences in GB(B) between the ideal level z and (1) "MP4/6-31+G(2d,2p)" (x) as our best practical level and (2) QCISD(T)/6-311+G(3df,2p) (y) representing our upper-limit level, we carried out numerous SP calculations for NH_3 and CH_3NH_2 employing correlation methods and basis sets relevant to x, y, and z (Tables S-1 and 5).

The first six lines of data in Table 5 show variations of PE, BSSE, and PE(B) with respect to changes in correlated level (*Mx*, *My*, *Mz*) while keeping the same basis 6-311+G** and changes in basis (*Bx*, *By*, *Bz*) at the same correlated level MP2. Based on PE(B), the BSSE corrected PE, we note the maximum variations in kcal/mol are (1) −0.37 for MP4 → CCSD(T) and −0.14 for 6-31+G(2d,2p) → aug-cc-pVTZ and (2) −0.06 for QCISD(T) → CCSD(T) and −0.04 for 6-311+G(3df,2p) → aug-cc-pVTZ. Assuming the variations in PE(B) are roughly additive, these numbers yield variations about −0.5 for $x \rightarrow z$ and −0.1 for $y \rightarrow z$. These theoretical variations are small compared with experimental errors (e.g., ca. 3 kcal/mol for the mass spectrometric value on glycine[5,8] in Table 3). We therefore conclude that levels x, y, and z all give very similar PE(B) and consequently very similar GB(B) and PA(B), as shown in Table 5. Yet, in terms of cost, only level x is practical for glycine.

The NIST values cited as experimental PAs in Tables 3 and 5 were evaluated with a complex procedure that depends on the available experimental data. (See, for example, the steps taken to evaluate the PA of ammonia, which has been used as a standard for assigning absolute values of PA.[32a]) The error limits for experimental PAs or GBs are usually around 3 kcal/mol (cf. Table 3). The advantage of using a theoretical approach to determine a gas-phase property is obviously the rigor attained by not having to make imprecise corrections for the physicochemical interferences to the sample during experiments. The errors in ab initio calculations may be estimated by systematic deductions based on the quality of basis set and correlation treatment employed. Using the GB(B) of CCSD(T)/aug-cc-pVTZ as the limit, we deduce that the GB(B) from MP4/6-31+G(2d,2p) is about 0.5 kcal/mol too high, while an upgrade to QCISD(T)/6-311+G(3df,2p) brings it close to the limiting value (Table 5). The possible scatter in BSSE due to uncertainty in the geometrical parameters of the ghost atom may add another 0.5 kcal/mol to the error (Appendix S). Thus an error limit of 1 kcal/mol may be assigned to the GB(B,C) of glycine from level x (Table 3). Extrapolating from Dunning's work,[45] we expect the error to be reduced to zero when a complete basis set is used with the CCSD(T) wave functions.

Summary

The calculated gaseous structures of the most stable conformers of glycine and protonated glycine are, respectively, symmetric with $NH_2\cdots O$ bonding and asymmetric with $NH\cdots O$ bonding at the MP2 level with the 6-31G* or larger basis sets. Analyses have been made on the electronic energies and geometries of these two structures calculated at different levels of theory. The gas-phase basicity (GB) has been determined to a high level of accuracy. These results, supplemented by those of ammonia and methylamine, lead to the following conclusions of practical importance to amino N-protonation calculations.

1. The HF/3-21G geometries may be used to obtain reasonable thermodynamic properties and as host geometries for HF/6-31G* calculations of electronic energies.
2. The HF/6-31G* geometries are similar to HF geometries of larger basis sets. Correlated-level calculations with extended basis sets at the HF/6-31G* geometries generally lead to improved accuracy.
3. The MP2/6-31G** and MP2/6-311G** optimized levels, as well as their HF counterparts, yield electronic protonation energies (PEs) significantly larger than those from basis sets of comparable size. The 6-31G** and 6-311G** basis sets are therefore not recommended for protonation studies.
4. The MP2/6-31G*, MP2/6-31+G**, and MP2/6-311+G** optimized levels, taken together as a series of MP2 levels with increasing basis size, produce PEs and geometrical parameters that converge smoothly to a high level of accuracy. The HF counterparts show similar trends but reach a lower level of accuracy. As a compromise between cost and accuracy, MP2/6-31+G** is recommended for general protonation studies.
5. The thermal contribution to GB, $\Delta G_{\mathrm{therm}}$, changes insignificantly from one level to another and very little from one molecule to another.
6. Correlation enhancement from MP2 to MP4 (using the same basis at the same geometry) raises the PE consistently by about 1 kcal/mol for different molecules. Basis expansion is shown to lower the PE effectively toward better accuracy; but the rate of improvement differs with basis sets and molecules. At a composite level where correlation enhancement and basis expansion are built in for the purpose of improving accuracy, these two effects cancel to a significant degree. Further correlation enhancement from MP4 to higher correlated levels such as QCISD(T) or CCSD(T) has relatively insignificant effect on the calculated PE.
7. The basis set superposition error (BSSE) at any correlated level is a significant quantity even with a relatively large basis set (ca. > 1 kcal/mol). The BSSE at a given theoretical level stays fairly constant among different molecules. The BSSE-corrected PE, PE(B), stays fairly constant for basis sets comparable to or larger than 6-31+G(2d,2p) at a given correlated level.
8. This study is the first to examine the impact of BSSE and the conformational equilibria effect (CEE) on the calculated GB of a medium-sized molecule such as glycine. For the calculated GB of glycine at the MP4/6-31+G(2d,2p)//MP2/6-31+G** composite level, 205.7 kcal/mol, BSSE brings a downward adjustment of 2.5 kcal/mol, much larger than the CEE upward adjustment of 0.3 kcal/mol. These corrections result in a GB of 203.5 kcal/mol, which is in good agreement with the mass spectrometric GB of 202.5 kcal/mol.

Concluding Remarks

This study provides a comprehensive guide to ab initio calculations of gas-phase basicity from low to high levels of theory. It also demonstrates how a theoretical GB may be used as a benchmark for evaluating the experimental GB. We hope these findings will build confidence in similar applications to compounds beyond glycine, alanine, and their peptides.

Acknowledgment. We thank Dr. Carolyn J. Cassady and Dr. Sharon G. Lias for providing revised experimental GB values of relevant molecules prior to publication, Dr. S. Mark

Cybulski for helpful discussions on the BSSE correction, and Dr. Alan Isaacson for critical comments on the CEE calculation. We acknowledge the computational support of the Ohio Supercomputer Center (OSC) and Miami Computing and Information Services (MCIS) and the financial assistance of National Institute of the General Medical Sciences (Grant R15-GM52670-01).

Supporting Information Available: Electronic energies, protonation energies, BSSE corrections, and thermodynamic properties for the protonations of ammonia and methylamine in Tables S-1 and S-2; optimized geometrical parameters for the most stable conformers of glycine and protonated glycine in Tables S-3 and S-4; electronic energies, thermodynamic properties, and equilibrium populations for low-energy conformers of glycine and protonated glycine at the MP2/6-31G* optimized level; and discussions on geometrical and BSSE parameters in Appendix S (8 pages). Ordering information is given on any current masthead page.

References and Notes

(1) (a) Wu, J.; Lebrilla, C. B. *J. Am. Soc. Soc.* **1993**, *115*, 3270. (b) Wu, J.; Lebrilla, C. B. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 91. (c) Wu, J.; Gard, E.; Bregar, J.; Green, M. K.; Labrilla, C. B. *J. Am. Chem. Soc.* **1995**, *117*, 9900.

(2) (a) Bliznyuk, A. A.; Schaefer, H. F., III.; Amster, I. J. *J. Am. Chem. Soc.* **1993**, *115*, 5149. (b) Gorman, G. S.; Amster, I. J. *J. Am. Chem. Soc.* **1993**, *115*, 5729.

(3) Zhang, K.; Zimmerman, D. M.; Chung-Phillips, A.; Cassady, C. J. *J. Am. Chem. Soc.* **1993**, *115*, 10812.

(4) Zhang, K.; Cassady, C. J.; Chung-Phillips, A. *J. Am. Chem. Soc.* **1994**, *116*, 11512.

(5) (a) McKiernan, J. W.; Beltrame, C. E. A.; Cassady, C. J. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 718. (b) Cassady, C. J. Private communication, 1997.

(6) Wu, Z.; Fenselau, C. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 777.

(7) Bojesen, G.; Breindahl, T. *J. Chem. Soc., Perkin Trans. 2.* **1994**, 1029.

(8) Cassady, C. J.; Carr, S. R.; Zhang, K.; Chung-Phillips, A. *J. Org. Chem.* **1995**, *60*, 1704.

(9) Jebber, K. A.; Zhang, K.; Cassady, C. J.; Chung-Phillips, A. *J. Am. Chem. Soc.* **1996**, *118*, 10515.

(10) (a) Carr, S. R.; Cassady, C. J. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 1203. (b) Ewing, N. P.; Zhang, X.; Cassady, C. J. *J. Mass Spectrom.* **1996**, *31*, 1345.

(11) Somogyi, A.; Wysocki, V. H.; Mayer, I. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 704.

(12) (a) Schäfer, L.; Sellers, H. L.; Lovas, F. J.; Suenram, R. D. *J. Am. Chem. Soc.* **1980**, *102*, 6566. (b) Suenram, R. D.; Lovas, F. J. *J. Am. Chem. Soc.* **1980**, *102*, 7180.

(13) Ramek, M.; Chang, V. K. W.; Frey, R. F.; Newton, S. Q.; Schäfer, L. *J. Mol. Struct.* **1991**, *235*, 1.

(14) Iijima, K.; Tanaka, K.; Onuma, S. *J. Mol. Struct.* **1991**, *246*, 257.

(15) Jensen, J.; Gordon, M. S. *J. Am. Chem. Soc.* **1991**, *113*, 7917.

(16) (a) Frey, R. F.; Coffin, J.; Newton, S. Q.; Ramek, M.; Chang, V. K. W.; Momany, F. A.; Schäfer, L. *J. Am. Chem. Soc.* **1992**, *114*, 5369. (b) Ramek, M.; Momany, F. A.; Schäfer, L. *J. Mol. Struct.* **1996**, *375*, 189.

(17) (a) Császár, A. G. *J. Am. Chem. Soc.* **1992**, *114*, 9568. (b) Császár, A. G. *J. Mol. Struct.* **1995**, *346*, 141.

(18) Hu, C.-H.; Shen, M.; Schaefer, H. F., III. *J. Am. Chem. Soc.* **1993**, *115*, 2923.

(19) Barone, V.; Adamo, C.; Lelj, F. *J. Chem. Phys.* **1995**, *102*, 364.

(20) (a) Godfrey, P. D.; Brown, R. D. *J. Am. Chem. Soc.* **1995**, *117*, 2019. (b) Godfrey, P. D.; Brown, R. D.; Rodgers, F. M. *J. Mol. Struct.* **1996**, *376*, 65.

(21) Reva, I. D.; Plokhotnichenko, A. M.; Stepanian, S. G.; Ivanov, A. Y.; Radchenko, E. D.; Sheina, G. G.; Blagoi, Y. P. *Chem. Phys. Lett.* **1995**, *232*, 141; Erratum. *Chem. Phys. Lett.* **1995**, *235*, 617.

(22) Neville, J. J.; Zheng, Y.; Brion, C. E. *J. Am. Chem. Soc.* **1996**, *118*, 10533.

(23) Bouchonnet, S.; Hoppilliard, Y. *Org. Mass Spectrom.* **1992**, *27*, 71.

(24) Jensen, F. *J. Am. Chem. Soc.* **1992**, *114*, 9533.

(25) Yu, D.; Rauk, A.; Armstrong, D. A. *J. Am. Chem. Soc.* **1995**, *117*, 1789.

(26) Zhang, K.; Chung-Phillips, A. Submitted for publication.

(27) (a) Frisch, M. J.; Del Bene, J. E.; Raghavachari, K.; Pople, J. A. *Chem. Phys. Lett.* **1981**, *83*, 240. (b) Del Bene, J. E.; Frisch, M. J.; Raghavachari, K.; Pople, J. A. *J. Phys. Chem.* **1982**, *86*, 1529.

(28) DeFrees, D. J.; McLean, A. D. *J. Comput. Chem.* **1986**, *7*, 321.

(29) (a) Del Bene, J. E.; Shavitt, I. *J. Phys. Chem.* **1990**, *94*, 5514. (b) Del Bene, J. E.; Aue, D. H.; Shavitt, I. *J. Am. Chem. Soc.* **1992**, *114*, 1631. (c) Del Bene, J. E. *J. Phys. Chem.* **1993**, *97*, 107.

(30) Komornicki, A.; Dixon, D. A. *J. Chem. Phys.* **1992**, *97*, 1087.

(31) Smith, B. J.; Radom, L. *J. Am. Chem. Soc.* **1993**, *115*, 4885.

(32) (a) Lias, S. G.; Liebman, J. F.; Levin, R. D. *J. Phys. Chem. Ref. Data* **1984**, *13*, 695. (b) Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. *J. Phys. Chem. Ref. Data Suppl.* **1988**, 17. (c) Hunter, E. P.; Lias, S. G. Evaluated Gas-Phase Basicities and Proton Affinities of Molecules: An Update. *J. Phys. Chem. Ref. Data* (to be published). See also: NIST Standard Reference Database Number 69, Mallard, W. G.; Linstrom, P. J., Eds.; National Institute of Standards and Technology: Gaithersburg, MD, 1997.

(33) Meot-Ner (Mautner), M.; Sieck, L. W. *J. Am. Chem. Soc.* **1991**, *113*, 4448.

(34) Szulejko, J. E.; McMahon, T. B. *J. Am. Chem. Soc.* **1993**, *115*, 7839.

(35) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley & Sons: New York, 1986. See relevant references therein.

(36) See, for example: Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery. J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*; Gaussian, Inc.: Pittsburgh, 1994.

(37) Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 1293.

(38) East, A. L. L.; Allen, W. D. *J. Chem. Phys.* **1993**, *99*, 4638.

(39) Pople, J. A.; Scott, A. P.; Wong, M. W.; Radom, L. *Isr. J. Chem.* **1993**, *33*, 345.

(40) Schäfer, L.; Newton, S. Q.; Cao, M.; Peeters, A.; Alsenoy, C. V.; Wolinski, K.; Momany, F. A. *J. Am. Chem. Soc.*, **1993**, *115*, 272.

(41) Alsenoy, C. V.; Cao, M.; Newton, S. Q.; Teppen, B.; Perczel, A.; Csizmadia, I. G.; Momany, F. A.; Schäfer, L. *J. Mol. Struct. (THEOCHEM)* **1993**, *286*, 149.

(42) Böhm. H. J. *J. Am. Chem. Soc.* **1993**, *115*, 6152.

(43) Gould, I. R.; Cornell, W. D.; Hillier, I. H. *J. Am. Chem. Soc.* **1994**, *116*, 9250.

(44) See, for example: Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(45) See references cited in: (a) Peterson, K. A.; Dunning, T. H., Jr. *J. Chem. Phys.* **1997**, *106*, 4119. (b) Wilson, A. K.; Dunning, T. H., Jr. *J. Chem. Phys.* **1997**, *106*, 8718.