Ab Initio Studies of Neutral and Protonated Triglycines: Comparison of Calculated and Experimental Gas-Phase Basicity

Kui Zhang, Carolyn J. Cassady,* and Alice Chung-Phillips*

Contribution from the Department of Chemistry, Miami University, Oxford, Ohio 45056

Received May 16, 1994[⊗]

Abstract: Ab initio Hartree-Fock (HF) molecular orbital calculations were carried out for triglycine and its six singly-protonated species (at the three different nitrogen and carbonyl oxygen sites). Minimum-energy structures were first determined using the 3-21G basis set with full geometry optimizations; single-point calculations were next performed in the larger 6-31G* basis set to gain better electronic protonation energies. Based on the 6-31G*//3-21G electronic protonation energies, the relative basicity of different atomic sites is as follows: amino N > amide carbonyl O > carboxylic carbonyl O > amide N. To obtain proton affinity (PA) and gas-phase basicity (GB) values for triglycine, electronic energies at the 6-31G*//6-31G* level and zero-point and thermal energies at the 3-21G//3-21G level were calculated for the most stable structures of the neutral and amino N-protonated species. The resulting theoretical PA and GB are 227.9 and 219.6 kcal/mol; the latter is in close agreement with the experimental GB of 218.5 ± 2.4 kcal/mol which was obtained by proton transfer reactions in a Fourier transform ion cyclotron resonance mass spectrometer. Among the calculated structures, several ring forms (C_5, C_7, C_{10}, and C_{11}) arising from intramolecular hydrogen bonding are identified. A novel intramolecular interaction between an amide carbonyl oxygen atom and a trivalent carbon atom is also uncovered. Effects of H-bonds and the geometry of the relevant protonated peptide bond on the relative stability of protonated species are discussed in detail. The 6-31G*//3-21G approach is shown to be a practical alternative to 6/31G*//6-31G* for obtaining reasonable theoretical estimates of GB with relative economy. Finally, comparisons are made between experimental and theoretical data on the thermodynamic and structural aspects of the protonation of triglycine.

Introduction

The mechanism of protonation of large molecules in biological processes is currently of interest. Major thermodynamic parameters for the protonation reaction

$$M + H^+ \rightarrow MH^+ \quad (1)$$

are proton affinity (PA) and gas-phase basicity (GB) which correspond to the negative of the respective enthalpy and Gibbs free energy changes for protonation ($-\Delta H$ and $-\Delta G$). While experimental measurements of PA and GB are becoming one of the most active areas of research in the gas-phase ion chemistry of biomolecules,[1-6] ab initio calculations have only recently been applied to supplement experiments with energetic, thermodynamic, and structural data of the neutral and protonated species.[1,6]

In our previous experimental and ab initio studies of the GBs of polyglycines, Gly_n ($n = 1-6$), Hartree-Fock (HF) calculations were carried out for glycine (Gly), diglycine (Gly_2), and their protonated species using the 6-31G* basis set with full geometry optimizations.[1] Zero-point energies and changes in enthalpy and entropy from 0 to 298 K were also obtained for the subsequent GB calculations. To further improve the calculated electronic energies, a significantly larger basis set, 6-31+G**, was used for Gly and electron correlation was incorporated using the Møller-Plesset (MP) perturbation method up to and including the fourth order (MP4) for both Gly and Gly_2. The final theoretical GB values for Gly and Gly_2 were 206.3 and 215.2 kcal/mol, in excellent agreement with the experimental values of 206.2 ± 2.2 and 215.3 ± 2.5 kcal/mol. The remarkable accuracy of high-level ab initio results gives impetus to extend the calculations to triglycine and its protonated species (Gly_3 and Gly_3H^+). Triglycine, $NH_2CH_2CONHCH_2CONHCH_2COOH$, is especially important because it is the first peptide with a main chain long enough to replicate certain structural features of large peptides and possibly proteins.

In contrast to the monomer and dimer of the polyglycine series, the trimer is a much more complex system and offers a far greater challenge to an ab initio investigation. Computationally, the cost of using a conventional mainframe or supercomputer is substantial if a basis as large as 6-31G* is used for geometry optimizations or a MP treatment is included to a satisfactory level. As a result, one has to resort to less accurate theoretical models. Structurally, the addition of one more glycine unit to the main chain increases the number of conformational isomers many-fold; it also affords more opportunities for long-range intramolecular hydrogen bonding (H-bonding). To find the global minimum of the relevant conformational energy surface amidst these complications would be very difficult. As this study involves protonation, the task is further multiplied by the number of available protonation sites (N and O atoms).

The focus of recent ab initio studies on tripeptides is on model structures that mimic the interior of a protein chain, e.g., *N*-formyl-ala-ala-amide and *N*-acetylglycylglycine *N'*-methylamide.[7,8] So far as a real tripeptide is concerned, the work by Wright and Borkman more than a decade ago seems to be the

[⊗] Abstract published in *Advance ACS Abstracts,* November 15, 1994.

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

(2) Wu, J.; Lebrilla, C. B. *J. Am. Chem. Soc.* **1993**, *115*, 3270.

(3) Wu, Z.; Fenselau, C. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 403.

(4) Gorman, G. S.; Amster, I. J. *J. Am. Chem. Soc.* **1993**, *115*, 5729.

(5) McKiernan, J. W.; Beltrame, C. E. A.; Cassady, C. J. *J. Am. Soc. Mass Spectrom.* **5**, 718.

(6) Bliznyuk, A. A.; Schaefer, H. F.; Amster, I. J. *J. Am. Chem. Soc.* **1993**, *115*, 5149.

(7) Van Alsenoy, C.; Cao, M.; Newton, S. Q.; Teppen, B.; Perczel, A.; Csizmadia, I. G.; Momany, F. A.; Schafer, L. *J. Mol. Struct. (Theochem)* **1993**, *286*, 149 and references cited therein.

0002-7863/94/1516-11512$04.50/0

only one in the literature.[9] In their study, SCF energies were calculated for the neutral and zwitterionic forms of triglycine in an extended trans conformation using fixed bond lengths and angles and a basis set comparable to STO-3G. To our best knowledge, no ab initio study on a protonated triglycine has been reported.

The aim of this work is to find the most stable structures for triglycine and its singly-protonated species at every available basic site using the ab initio approach with geometry optimizations. The calculated electronic energies will determine the relative ease of protonation at different basic sites, while the calculated structures for different molecular species will reveal the extent and pattern of H-bonding. These findings will provide useful insight to the energetics and geometries of larger peptides. From the calculated energies and thermodynamic terms, theoretical PA and GB values for triglycine will be deduced.

The theoretical GB of triglycine will next be compared with an experimental GB value obtained by proton transfer reactions in a Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR). In addition, experimental results of the dissociation of protonated triglycine will be compared with the calculated structures. Favorable comparisons will give support to the credibility of both theoretical calculations and experimental measurements.

Computational Methods

The Gaussian 92 program was used to carry out all calculations.[10,11] Major calculations involved geometry optimizations at the HF level using the split-valence 3-21G basis in search of the most stable structures of the neutral triglycine and six singly-protonated species at the three nitrogen and three carbonyl oxygen sites. The terminal carboxylic hydroxyl oxygen site was not considered because of known complications related to protonation.[12]

Due to the large number of basis functions involved in each calculation, it was computationally impractical to take the option of CalcFC or CalcAll in the Gaussian program for geometry optimization,[11] i.e., to request analytic HF force constants for an efficient search of an energy minimum via geometry optimization. Under the less-than-ideal conditions of employing forces determined by numerical differentiation[11] and the presence of a large number of conformational minima in each molecular species, a strategy outlined below was adopted in seeking the global energy minimum or an energy minimum relevant to this protonation study.

An extended chain structure was chosen as the starting geometry for the neutral species. The semiempirical AM1 model was first used to generate various stable structures by scanning the low-energy portion of the potential energy surface (PES) that represents internal rotations about pertinent C−C and C−N bonds. Geometries for the lower-energy conformers generated by AM1 were next used as the trial geometries to initiate ab initio HF/3-21G geometry optimizations. From these trials, the 3-21G structure with the lowest energy was identified. Confirmatory tests for this "lowest energy" conformer consisted of a new round of geometry optimizations starting with the geometry of the conformer but with altered dihedral angles. If the subsequent optimizations led to the same structure or new structures of higher energies only, the search was completed. Otherwise, the resulting new "lowest-energy" conformer was subjected to the same confirmatory tests. The final lowest-energy conformer was then taken to be the most stable 3-21G structure for the neutral species. Verification of the final structure as a minimum on the PES was made by a vibrational frequency calculation that yields all positive frequencies.[11]

For a protonated species, the 3-21G geometry for the most stable structure of the neutral species was used as the starting geometry with the proton added to a particular site. A procedure similar to, but less extensive than, the one described above for the neutral species was followed in search of the most stable conformation of each protonated species. The final geometry was verified as a minimum-energy structure by the uniformly positive eigenvalues of the Hessian matrix.[11]

After the most stable structures of the seven species were decided at the 3-21G//3-21G level, zero-point energies (based on harmonic vibrational frequencies) and thermodynamic properties were obtained for two species: the Gly_3 and Gly_3H^+ structures of greatest stability. These values plus the corresponding electronic protonation energy were used to calculate ideal-gas proton affinity and basicity at 298.15 K and 1 atm.

Upon discovering that the calculated GB at the 3-21G//3-21G level was more than 10 kcal/mol higher than the experimental value, single-point HF/6-31G* calculations were performed at the HF/3-21G geometries for all seven structures. In addition, HF/6-31G* calculations with full geometry optimization were carried out for the two structures representing the neutral and most stable protonated species. Employing the same 3-21G//3-21G zero-point and thermal energies but the new 6-31G*//3-21G or 6-31G*//6-31G* electronic energy, the resulting GB was found to agree more closely with experiment than the original 3-21G//3-21G value (*vide infra*).

In addition to seeking the most stable conformation for each species, the optimized geometry for the extended chain form of Gly_3 was also determined for use as a reference at both the HF/3-21G and HF/6-31G* levels. All relevant MP4 calculations were performed as MP4 = FULL. Among all the calculations, the largest case was optimizing the geometry of amino N-protonated Gly_3H^+ at the HF/6-31G* level. This optimization involved 219 basis functions and required over 80 cycles to reach convergence.

Experimental Methods

All experiments were performed using a Bruker CMS-47X Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR) equipped with an external ion source and a 4.7 T superconducting magnet.[13] Protonated triglycine ions were produced in the external source using a Phrasor Scientific fast atom bombardment (FAB) gun[14] which employs a 6−10 kV beam of xenon atoms and ions. Triglycine was dissolved in glycerol with trifluoroacetic acid added as necessary to increase ion abundance.

The procedure for obtaining experimental GBs from proton transfer reactions has been discussed in detail previously.[1] Following movement of ions into the FT-ICR cell by electrostatic focusing, the trapped ions were thermalized by collisional cooling with argon. Protonated triglycine ions (Gly_3H^+) were mass selected by resonant frequency ejection techniques and allowed to undergo deprotonation reactions with a series of reference compounds (B) of known GB, reaction 2.

$$Gly_3H^+ + B \rightarrow Gly_3 + BH^+ \qquad (2)$$

These reactions were conducted with static pressures of reference compounds in the range of $(3-30) \times 10^{-8}$ Torr. The FT-ICR cell was maintained at room temperature for all measurements. The presence or absence of proton transfer, as determined by comparing experimental and theoretical[15] reaction rate constants, enabled the GB of triglycine to be bracketed. Reference compound GBs were obtained from the basicity ladder of Meot-Ner and Sieck[16] or from the

(8) For recent computational studies of model tripeptides, see: (a) Sapse, A. M.; Daniels, S. B.; Erickson, B. W. *Tetrahedron* **1988**, *44*, 999. (b) Perczel, A.; Kajtar, M.; Marcoccia, J. F.; Csizmadia, I. G. *J. Mol. Struct. (Theochem)* **1991**, *232*, 291. (c) Schafer, L.; Newton, S. Q.; Cao, M.; Peeters, A.; van Alsenoy, C.; Wolinski, K.; Momany, F. A. *J. Am. Chem. Soc.* **1993**, *115*, 272. (d) Bohm, H. J. *J. Am. Chem. Soc.* **1993**, *115*, 6152.

(9) Wright, L. R.; Borkman, R. F. *J. Phys. Chem.* **1982**, *86*, 3956.

(10) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab initio Molecular Orbital Theory*; Wiley: New York, 1986.

(11) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. A.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. *Gaussian 92*; Gaussian: Pittsburgh, 1992.

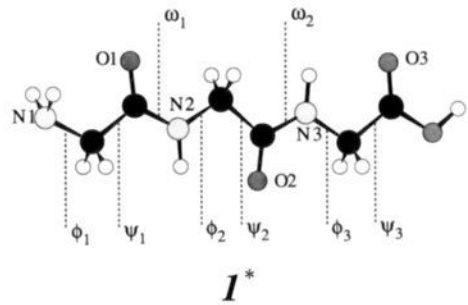
(12) Bouchonnet, S.; Hoppilliard, Y. *Org. Mass Spectrom.* **1992**, *27*, 71.

(13) Kofel, P.; Allemann, M.; Kellerhals, H. P.; Wanczek, K. P. *Int. J. Mass. Spec. Ion. Proc.* **1985**, *65*, 97.

(14) Perel, J.; Faull, K.; Mahoney, J. F.; Tyler, A. N.; Barchas, J. D. *Am. Lab.* **1984**, *Nov.*, 94.

(15) Su, T.; Bowers, M. T. *Int. J. Mass Spectrom. Ion Phys.* **1973**, *12*, 347.

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



*1**

Figure 1. Definitions of dihedral angles (ϕ_1, ψ_2, ω_1, ϕ_2, ψ_2, ω_2, ϕ_3, and ψ_3) and protonation sites (N1, O1, N2, O2, N3, and O3) for triglycine. Atoms are identified by shading (H, none; C, dark; N, light; O, medium) or size (C > N > O > H). This is the extended form of neutral triglycine (**1***), a reference structure.

compilation of Lias *et al.*[17] and adjusted to the Meot-Ner and Sieck ladder. The experimental ladder of GBs and PAs is currently in a state of revision;[16,18] experimental values discussed in this paper may need to be lowered by a few kilocalories per mole at a later date if an accepted revision of the basicity ladder is achieved.

In low-energy collision-induced dissociation (CID) experiments,[19] mass-selected Gly_3H^+ were excited by single-frequency pulses and allowed to undergo collisions with a static pressure of argon at $(1-3) \times 10^{-7}$ Torr. The collision energy was varied from 0 to 200 eV (lab). In order to ensure single-collision conditions, the time allowed for collisions was limited to 50 ms.

Results and Discussion

Geometries and Energies. The conformation of a triglycine may be described by torsional angles ϕ, ψ, and ω used conventionally for a peptide.[20] The angles ϕ and ψ pertain to atoms along the main chain, while ω refers specifically to the amide group. The dihedral angle is positive when measured clockwise from AB to CD in a group of four atoms ABCD.

In Figure 1, the torsional angles and protonation sites relevant to this study are defined using the extended structure, **1***, of a neutral triglycine for reference. Starting from the terminal H−N bond from the left and ending at the terminal O−H bond on the right, the torsional angles (ϕ_1, ψ_1), (ϕ_2, ψ_2), and (ϕ_3, ψ_3) relate to the respective first, second, and third glycine units. Likewise, ω_1 and ω_2 are the dihedral angles assigned to the first and second amide groups. Atom labels for protonation sites follow the same order: (N1, O1), (N2, O2), and (N3,O3) refer to the nitrogen and oxygen atoms in the respective glycine units.

In a Hartree−Fock calculation the SCF energy (E_{SCF}) represents the electronic energy. Relative stability of different molecular species in the same class of compounds is often expressed in terms of relative energies, which are differences between SCF energies (ΔE_{SCF}) in this case. With regard to protonation, it is convenient to set E_{SCF} of the neutral species as zero of energy.

The HF/3-21G calculations for the neutral and six protonated species of triglycine yielded numerous stable structures. Among the many low-energy conformers generated in separate searches for each species, the structure with the lowest energy was taken to represent the most stable structure. The seven most stable structures are shown in Figure 2: **1** for the neutral species and **2**−**7** for the protonated species. The specific protonation site in each protonated species is marked with an atom label;

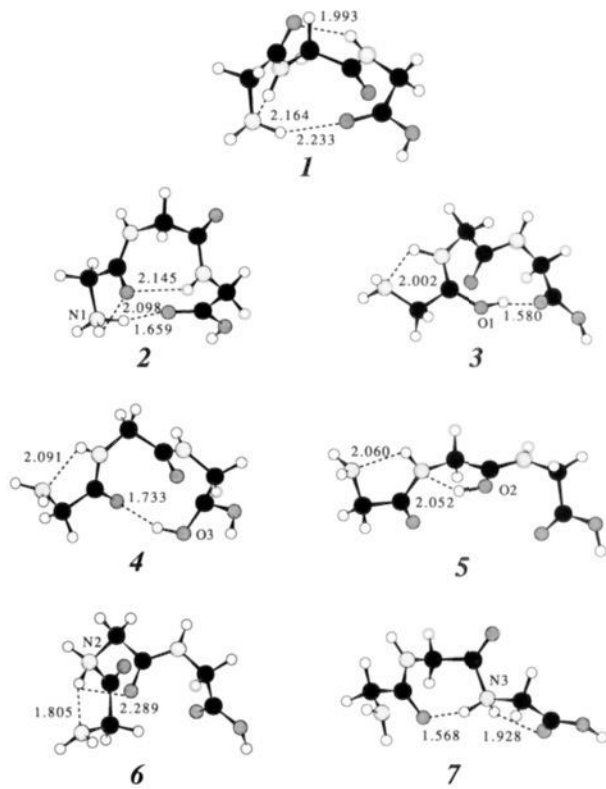


Figure 2. HF/3-21G minimum-energy structures for the neutral (**1**) and protonated (**2**−**7**) triglycines. Pertinent site of protonation is denoted by an atom label. Intramolecular hydrogen bond is shown with a dotted line; the bond distance is given in angstroms.

important H-bonds with bonding distances below 2.3 Å are drawn with dotted lines.

Conformations of species **1*** and **1**−**7** as described by the torsional angles and the corresponding SCF energies (3-21G//3-21G) are given in Table 1. Setting the zero of energy at the most stable neutral species Gly_3(**1**), relative energies of the protonated species Gly_3H^+(**I**) for **I** = **2**, ..., **7** are listed in order of decreasing magnitude.

The torsional angles ϕ, ψ, and ω for structures **1**−**7** *show* considerable flexibility of the triglycine backbone. This can be seen by comparison of the angles of each species to those of the reference structure **1***. Hydrogen bonding, especially those linking distant fragments, affects the torsional angles significantly. Different folding patterns emerge as a result. Judging from the overall shape and local twists and bends, structures **2**−**7** appear to be reasonable representations of the most likely products of the reactant structure **1** as depicted in Figure 2.

SCF energies obtained from single-point calculations using the larger basis 6-31G* at the optimized 3-21G geometries (6-31G*//3-21G) are also given in Table 1. The order of stability resulting from the two basis sets differs only in the positions of **4** and **5**: **5** is more stable than **4** according to 6-31G*. Accepting the 6-31G* electronic protonation energies as more accurate (*vide infra*), the new order of stability becomes **2** > **3** > **5** > **4** > **6** > **7**.

SCF energies resulting from full geometry optimizations for species **1***, **1**, and **2** using the larger 6-31G* basis directly are also shown in Table 1. The corresponding molecular figures with complete atom labels and the optimized geometrical parameters (bond lengths, bond angles, and dihedral angles) are provided as supplementary material (Figure S-1 and Table S-I). For comparison, the geometrical parameters for the same three structures optimized in the 3-21G basis are also provided (Table

(17) Lias, S. G.; Liebman, J. F.; Levin, R. D. *J. Phys. Chem. Ref. Data* **1984**, *13*, 695.
(18) Szulejko, J. E.; McMahon, T. B. *J. Am. Chem. Soc.* **1993**, *115*, 7839.
(19) Cody, R. B.; Freiser, B. S. *Int. J. Mass Spectrom. Ion Phys.* **1982**, *41*, 199.
(20) Ramakrishnan, C.; Ramachandran, G. N. *Biophys. J.* **1965**, *5*, 909.

Table 1. Minimum-Energy Structures of Neutral and Protonated Triglycines[a]

	Gly_3(1*)	Gly_3(1)	Gly_3H^+(2)	Gly_3H^+(3)	Gly_3H^+(4)	Gly_3H^+(5)	Gly_3H^+(6)	Gly_3H^+(7)
protonation site:	–	–	N1	O1	O3	O2	N2	N3
dihedral angles								
ϕ_1	57.8	66.4	46.6	113.0	115.5	98.1	113.5	162.4
ψ_1	−179.9	31.2	147.5	0.6	−0.6	19.0	16.5	−31.8
ϕ_2	180.0	82.0	80.5	−51.2	−38.1	−113.1	−69.2	−71.9
ψ_2	180.0	−68.6	−76.1	−154.9	132.0	−174.9	147.5	85.1
ϕ_3	180.0	−95.1	−101.1	−64.3	13.8	70.3	67.1	−59.4
ψ_3	180.0	−178.8	−177.9	−172.2	−136.9	175.3	174.8	−178.5
ω_1	180.0	171.8	179.3	−178.8	−178.5	164.3	177.1, 60.6[b]	−175.7
ω_2	180.0	−176.5	179.5	178.8	177.6	179.6	177.1	−104.9, 139.1[b]
SCF energies (3-21G//3-21G)								
E_{SCF}	−692.570060	−692.580714	−692.970713	−692.963999	−692.946171	−692.939052	−692.933566	−692.923240
ΔE_{SCF}	6.69	0.00	−244.73	−240.52	−229.33	−224.86	−221.42	−214.94
SCF energies (6-31G*//3-21G)								
E_{SCF}	−696.458619	−696.460464	−696.833642	−696.825537	−696.804378	−696.813948	−696.800009	−696.793888
ΔE_{SCF}	1.16	0.00	−234.17	−229.09	−215.81	−221.81	−213.07	−209.23
SCF energies (6-31G*//6-31G*)								
E_{SCF}	−696.463559	−696.468206	−696.842636					
ΔE_{SCF}	2.92	0.00	−234.96					

[a] Angles in deg, E_{SCF} in au, ΔE_{SCF} in kcal/mol. See text and Figures 1 and 2. [b] For the two N−H bonds resulting from protonation on the N atom.

Table 2. Important Structural Features in Neutral and Protonated Triglycines[a]

A. Ring Forms vs Hydrogen Bonding

structure	C_5	C_7	C_{10}	C_{11}
1	N1···H−N2	O1···H−N3		N1−H···O3
2 (N1)	$N1^+$−H···O1	O1···H−N3		$N1^+$−H···O3
3 (O1)	N1···H−N2		$O1^+$−H···O3	
4 (O3)	N1···H−N2		O1···H−$O3^+$	
5 (O2)	N1···H−N2 N2···H−$O2^+$			
6 (N2)	N1···H−$N2^+$ $N2^+$−H···O2			
7 (N3)	$N3^+$−H···O3	O1···H−$N3^+$		

B. C=O and C−N Bond Lengths in Amide Groups[b]

	N-terminal (O1=C−N2)		C-terminal (O2=C−N3)	
structure	C=O	C−N	C=O	C−N
1	1.230	1.340	1.217	1.360
2 (N1)	1.227	1.326	1.212	1.362
3 (O1)	1.286*	1.291*	1.215	1.348
4 (O3)	1.242	1.330	1.297**	1.296**
5 (O2)	1.209**	1.392**	1.294*	1.287*
6 (N2)	1.184*	1.539*	1.220	1.333
7 (N3)	1.237	1.348	1.186*	1.509*

[a] Bond length in Å. See Figures 1 and 2 and Table 1 for notations. In specifying the structure, the protonation site is enclosed in parentheses. See text for definitions and discussion. [b] 3-21G//3-21G values. Bond lengths with significant deviations from the normal C=O and C−N values of an amide group are marked with one or two asterisks.

S-II). Note that the 3-21G geometries closely resemble the 6-31G* geometries. Therefore, the ΔE_{SCF} values calculated at the 6-31G*//3-21G level are expected to be good approximations of the 6-31G*//6-31G* values (*vide infra*). To show the degree of convergence in geometry optimization, the maximum internal coordinate residual force for each optimized structure (three HF/6-31G* and eight HF/3-21G) is included as supplementary material (Table S-III). Finally, the Cartesian coordinates for all eleven optimized geometries are listed for completeness (Tables S-IV and V). The optimized geometries of these peptide systems are of considerable interest for molecular modeling parameter development.[7]

The ab initio results presented here for a tripeptide and its protonation products are new and extensive. To gain a better understanding of the numerous structural forms and their influence on molecular stability requires an in-depth and systematic study. Several pertinent topics are explored below.

Hydrogen Bonding. The seven structures in Figure 2 display sixteen important intramolecular hydrogen-bonding interactions (Y···H−X). They are variations of four basic types: N···H−N, O···H−N, N···H−O, and O···H−O. As a consequence of H-bonding, 5-, 7-, 10-, and 11-membered rings appear; they are called C_5, C_7, C_{10}, and C_{11} ring forms. The H-bonds involved in different rings are listed in Table 2A and defined in the Appendix.

Compared with glycine and diglycine,[1] triglycine has a longer chain length and therefore a greater flexibility to form long-range H-bonding. In a protonated species, the transferred proton has added geometrical reach and freedom in the choice of donor atom (Y) in forming Y···H−X. These are reasons for the increase in the number of H-bonds, the emergence of the larger C_{10} and C_{11} ring forms, and some significantly shorter interaction distances (Y···H) among the triglycine species. The normal Y···H distance is around 2.0 Å, but there are five distances below 1.9 Å shown here. These five H-bonds are exceptionally strong attractions. (See the Appendix for more details.)

The neutral triglycine **1** contains three ring forms of different sizes: the folded structure is in stark contrast to the extended structure of neutral diglycine.[1] This result confirms the general expectation that a large peptide prefers a folded conformation. Each protonated species contains at least two ring forms. Those having large rings (**2**–**4**) seem to have greater stability than those without (**5**–**7**). (See the Appendix.) The H-bonds in each structure as exhibited in Figure 2 will be briefly described below; obvious relationships between H-bonds of each protonated species and its parent neutral species **1** will be mentioned.

Neutral species **1**: This species is a compactly folded structure supported by N1–H···N2, O2–H···N3, and N1–H···O3 bonding in the respective C_5, C_7, and C_{11} ring forms. Note especially that C_{11} holds together the terminal amino and carboxylic groups and is therefore the largest ring in a triglycine.

N1-protonated species **2**: Adding the proton to the terminal amino N does not seem to disturb the rest of the main chain. The three ring forms of **1** are fairly well maintained except for switching N1–H···N2 in the C_5 ring to N1–H···O1 and flipping the C_7 ring.

O1-protonated species **3**: Protonation at O1 breaks up the C_7 and C_{11} rings of **1** to form a new C_{10} ring with $O1^+$–H···O3 bonding.

O3-protonated species **4**: **4** appears to be created by **3** with a "migration" of the proton from O1 to O3 to form O1···H–$O3^+$.

O2-protonated species **5**: Protonation at the centrally located O2 disengages the two larger rings in **1**; however, the new $O2^+$–H bond engages itself in N2···H–$O2^+$ to form a smaller C_5 ring at the center.

N2-protonated species **6** or N3-protonated species **7**: Effects from protonating the amide N2 or N3 are harder to analyze. The lengthening of the amide C–N bond due to protonation, as well as the placement of the new $N2^+$–H or $N3^+$–H bond, may interfere sufficiently with the original ring forms in **1**. The H-bonds in **6** and **7** are variations of C_5 and C_7 already discussed.

It needs to be emphasized that structure **1** represents the most folded conformation computed for a tripeptide known to date. In the previous study on the neutral triglycine,[9] the extended conformer **1*** was assumed as the lowest-energy conformation because of a lack of information on intramolecular H-bonding. In this work **1*** is shown to be higher in energy than **1** at all three levels (Table 1). This may be explained by the fact that the H-bonds in **1*** are weaker than those in **1**. A close examination of **1*** reveals 1,5 interactions NH_2···O (bifurcated) on the left, N2–H···O2 in the center, and N3–H···O3 on the right with respective H···Y distances at 2.765, 2.145, and 2.220 Å. The bifurcated interactions (with each H···O distance larger than 2.7 Å) are definitely weaker interactions than the average intramolecular H-bonding in these systems (with a H···Y distance around 2 Å)[1] and are therefore expected to be weaker than any one of the three H-bonds in the folded conformer ***1*** (Figure 2). Folding obviously stabilizes the neutral triglycine.

The foregoing analysis on the H-bonds suggests that structures for the neutral species **1** and the three lower-energy protonated species **2**–**4** have the characteristic features of global minima. These are optimally folded conformations supported by long-range H-bonds with short bonding distances which infer extra stability. The situation is not as clear-cut for those of the higher-energy protonated species **5**–**7**, each of which is likely to represent a local minimum close to the transition state that lies in the direct path to protonation of **1** at the respective protonation site.

Previous findings on peptide conformations center around model systems which are not entirely compatible to those here. Typically in computational studies, a model peptide is one that eliminates the terminal amino group and replaces the terminal carboxylic group with an amide unit. The H-bonds are exclusively of the type between an amide carbonyl oxygen (C=O) in one peptide unit and an amide hydrogen (N–H) in another peptide unit. The ring size for the C=O···H–N interaction depends on how far apart the two involved peptide units are. For the ring forms of interest here, double C_5, double C_7, and C_{10} (β- or U-turn) conformations have all been shown computationally to be stable in model tripeptides.[7,8] Due to the configuration of the real tripeptide under study here, however, the only molecular fragment that may be compared directly with the previous model studies is that associated with the central pair of dihedrel angles ϕ_2, ψ_2. Individual C_5, C_7, and C_{10} conformations were also observed in solution studies of model peptide systems.[21-23] In protein crystals, the smaller C_5 and C_7 forms are rare, but the larger C_{10} rings do exist—the larger rings are believed to be energetically more favored because of less ring strain.[21]

The H-bonds involved in the C_5 and C_7 rings in this study are characteristic of small peptides and protonated peptides. Yet, some of the H-bonds can be of general interest. For example, the C_5 ring in the N1-protonated species **2**, which connects the terminal ammonium ion ($-NH_3^+$) to the oxygen of amide bond 1, is a plausible model of internal H-bonding in zwitterions. Likewise, the C_7 ring in the neutral species **1** or N1-protonated species **2**, which links amide bonds 1 and 2, can conceivably occur in the bulk of a large peptide or protein. These two C_7 rings also resemble the $C7^{eq}$ conformations described in the model tripeptide *N*-formyl-ala-ala-amide.[7] (The C_7 ring in the N3-protonated species **7** also has similar dihedral angles ϕ_2, ψ_2 but involves a tetrahedral protonated N3 instead of a neutral amide N3.) The C_{10} ring in the O3-protonated species **4** bears analogy to the β-turn. The C_{11} ring in the neutral species **1** is the theoretically possible but rarely observed stable conformation.

Peptide Bonds. In a polypeptide molecule the peptide bond C–N binds two amino acid units. It is part of the amide group "–CO–NH–" widely known to be trans and nearly planar due to the resonance O=C–NH ↔ O–C=NH. In the language of molecular orbitals, there is partial π-electron delocalization across the O–C–N framework. The amide N atom is nearly sp^2 hybridized; hence the amide C–N bond is shorter than a C–N single bond where N is sp^3 hybridized.

The angle ω is the dihedral angle for the group of four atoms O–C–N–H associated with the peptide bond. A trans and planar conformation for this group would correspond to $\omega = 180°$. In Table 1 the HF/3-21G values of ω_1 and ω_2 for the two peptide bonds in the triglycines are listed. Those involving protonations at the amide N atoms (ω_1 in **6** and ω_2 in **7**) are expected to deviate markedly from 180° as a result of changing hybridization on N from sp^2 to sp^3. Among the remaining ω values, there are eight within ±3° and three within ±10° of

(21) For a brief review on experimental conformational studies of model peptides, see: Dado, G. P.; Gellman, S. H. *J. Am. Chem. Soc.* **1994**, *116*, 1054.

(22) For solution studies of C_5 and C_7 conformations in model dipeptides, see: (a) Avignon, M.; Houng, P. V.; Lascombe, J.; Marraud, M.; Neel, J. *Biopolymers* **1969**, *8*, 69. (b) Bystrov, V. F.; Portnova, S. L.; Tsetlin, V. I.; Ivanov, V. T.; Ovchinnikov, Y. A. *Tetrahedron* **1969**, *25*, 493. (c) Bystrov, V. F.; Portnova, S. L.; Balashova, T. A.; Tsetlin, V. I.; Ivanov, V. T.; Kostetzky, P. V.; Ovchinnikov, Y. A. *Tetrahedron Lett.* **1969**, *59*, 5225. (d) Zuk, W. M.; Freedman, T. B.; Nafie, L. A. *Biopolymers* **1989**, *28*, 2025.

(23) For NMR studies on conformations of tripeptide model systems in solution see: Boussard, G.; Marraud, M.; Neel, J. *J. Chim. Phys.* **1974**, *71*, 46.

180°; there is one severe distortion (ω_1 = 164.3° in **5**), a consequence of an unusual, concerted "push and pull" at the two ends of the N−H bond by two separate H-bonding interactions (Figure 2). Interestingly, protonation on the amide O atom (ω_1 in **3** or ω_2 in **5**) has no significant effect on ω as conjugation over the HO−C−NH segment, HO⋯C⋯NH, is expected. We therefore conclude that the conformation of a normal "−CO−NH−" group, one that is not perturbed by N-protonation, remains typically trans and nearly planar.

The small deviation of ω from 180° in a normal amide group observed here is consistent with existing experimental and computational findings on the nonplanarity of "−CO−NH−". X-ray studies on a series of globular proteins show an average deviation of 3.5° and individual deviations as much as 10°.[24] Recently, ab initio calculations using different basis sets up to 6-311++G** on acetamide (as a model compound of the peptide bond) also show the O−C−N−H torsional angle being several degrees away from 180°.[25]

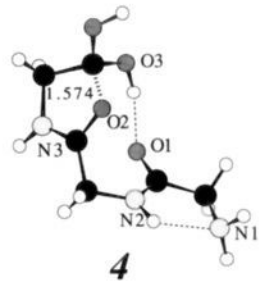
In this study the CO and CN bond lengths in "−CO−NH−" are influenced greatly by the electronic states on the C, O, and N atoms and intramolecular nonbonded interactions. To show some of these effects, the bond lengths of the two amide groups in structures **1**−**7** are listed in Table 2B. Note the conventional notation O=C−NH for an amide group is now used for convenience: the C=O and C−N bonds here are not fully double and single bonds because of partial conjugation.

Among the seven triglycines in Figure 2, there are 14 amide groups with 14 pairs of C=O and C−N bonds. Four of these pairs are "perturbed" by either O- or N-protonation; their corresponding bond lengths are marked with single asterisks. Two of the remaining ten pairs are also "perturbed" by unusual circumstances; their values are marked with double asterisks. There are eight pairs that appear to be "unperturbed" or normal.

Based on the HF/3-21G calculations shown in Table 2B for the eight normal pairs (those without asterisks), the average bond lengths are 1.225 ± 0.009 Å for C=O and 1.343 ± 0.011 Å for C−N. They compare favorably with results from recent HF/4-21G calculations on the model tri- and hexapeptides.[8c] Adding polarization functions to the basis set reduces C=O by 0.03 Å but has very little effect on C−N.[8d] (See also Table S-I and S-II in the supplementary material with this paper.) The 3-21G and 4-21G values cited here for the C=O and C−N bonds are in reasonable agreement with the respective average values of 1.229 and 1.335 Å in solid-state proteins.[26]

As mentioned above, protonation at the amide O atom in species **3** and **4** creates the conjugated segment HO⋯C⋯NH which equalizes the CN and CO bond distances (to around 1.29 Å). On the other hand, protonation at the amide N atom in species **6** and **7** destroys the resonance in the amide group and leads to a dramatic lengthening of the CN bonds (to greater than 1.5 Å). These trends are definitive and consistent.

There are two "anomalies" in the bond lengths arising from unusual nonbonded interactions. One is species **5** discussed earlier. The "push and pull" actions of the two H-bonds distort the planarity of the amide group 1 enough to yield a noticeable shortening and lengthening of the respective C=O and C−N bonds. The other one is the amide group 2 in species **4** which is expected to be "unperturbed" as protonation occurs elsewhere (on the terminal O3 atom). Rather, the segment O2−C−N3



4

Figure 3. A rotated figure of Gly_3H^+(**4**) in Figure 2 to show a new type of intramolecular interaction involving the segment O2−C−N3−C−C. The O2⋯C distance for this interaction is 1.574 Å.

appears to be conjugated, as the CO and CN bond lengths are shown to be practically equal in Table 2B. (The complete geometry of species **4** is provided in Table S-V, supplementary material.) A close examination reveals that O2 is indeed "perturbed" by the terminal C atom in forming an "intramolecular C-bond".

In Figure 3, a better perspective of species **4** for this unique bonding is displayed. A nearly planar 5-membered ring is formed over the segment O2−C−N3−C−C. The ring is closed by the O2⋯C interaction with a short separation of 1.574 Å. At this short distance O2 acts like the electron donor while C becomes the acceptor.

The O2⋯C interaction triggers two unexpected events. First, O3-protonation no longer invokes conjugation in the HO−C−OH segment as in diglycine.[1] This is evidenced by the pyramidal conformation of the terminal C with significantly longer C−O bond lengths than their counterparts in diglycine.[1] Second, the amide group 2 undergoes conjugation along the (C)O−C−NH segment, (C)O⋯C⋯NH, much like the familiar HO⋯C⋯NH, as shown by similar CO and CN bond lengths.

Further comparisons of HF/6-31G* Mulliken overlap populations for both bonded and nonbonded atom pairs in species **4** indicate that the population for the O2⋯C pair (0.27 e) is four times larger than those for each of the H-bonded pairs, N1⋯H (0.07 e) and O1⋯H (0.08 e), and about one-half of the C−O bonds (0.62 and 0.69 e) at the C-terminus. This implies that the new O2⋯C interaction acts more like a one-half coordinate covalent bond. Thus, this "intramolecular C-bond" represented by "⟩C=O⋯C(OH)$_2$−" is far stronger than a typical "intramolecular H-bond".

A re-examination of Figure 3 reveals that the new O2⋯C interaction lies nearby the H-bond between O1 and the added proton on O3 in forming the C_{10} ring. An interesting question arises as to whether the O2 and terminal C atoms are brought to proximity as a result of this C_{10} ring closure. In view of the stronger "C-bond" relative to a H-bond as discussed above, the answer must be that the new O2⋯C interaction is independent of the "synergistic" effect of this H-bonding in the immediate neighborhood. To test the hypothesis that this "C-bond" is indeed independent of the nonbonding environment specific to species **4**, separate HF/3-21G and HF/6-31G* calculations with full geometry optimizations were carried out on the model species $RCONHCH_2C(OH)_2^+$ with R = H and CH_3 in the conformation similar to the relevant fragment in species **4**.[27] The respective O2⋯C distances for R = H and CH_3 are 1.561 and 1.574 Å in the 3-21G basis and 1.516 and 1.495 Å in the 6-31G* basis. These relatively short interatomic distances unambiguously establish the existence of this "C-bond" outside of the protonated triglycine environment.

(24) Ramakrishnan, C.; Balasubramanian, R. *Int. J. Peptide Protein Res.* **1972**, *4*, 79.

(25) Wong, M. W.; Wiberg, K. B. *J. Phys. Chem.* **1992**, *96*, 668.

(26) Benedetti, E. In *Peptides: Proceedings of the 5th American Peptide Symposium*, San Diego, CA, June 20−24, 1977; Goodman, M., Meienhofer, J., Eds.; Wiley: New York, 1977; p 257.

(27) Zhang, K.; Chung-Phillips, A. unpublished results.

Protonation Energies. The SCF energies in Table 1 may be used to estimate the electronic protonation energy (PE) of triglycine corresponding to each basic site as follows:

$$Gly_3(\mathbf{1}) + H^+ \rightarrow Gly_3H^+(\mathbf{I}) \qquad I = \mathbf{2}, ..., \mathbf{7} \qquad (3)$$

for which

$$\begin{aligned} PE &= -[E_{elec}(Gly_3H^+) - E_{elec}(Gly_3)] \\ &= -\Delta E_{elec} = -\Delta E_{SCF} \end{aligned} \qquad (4)$$

Here the electronic energy (E_{elec}) is equated to the Hartree–Fock energy (E_{SCF}) because of the omission of electron correlation (E_{corr}). Otherwise, $E_{elec} = E_{SCF} + E_{corr}$. Note that PE is a positive quantity; a larger PE implies a greater ability for a given basic site to accept a proton. Based on the 6-31G*//3-21G values for the PEs, the relative protonation ability of different basic sites is N1 > O1 > O2 > O3 > N2 > N3. In terms of chemically distinct basic sites, the relative basicity becomes amino N > amide carbonyl O > carboxylic carbonyl O > amide N.

The amino N1 is the most favored protonation site because the formation of the terminal ammonium ion ($-NH_3^+$) gives great stability to the protonated species **2**. In the protonation reaction **1** → **2**, none of the amide groups are perturbed and H-bonding remains nearly the same. Thus, a PE of 234 kcal/mol reflects mainly the stabilization due to the formation of the N^+-H bond in a molecule the size of triglycine.

Protonations at the amide carbonyl O1 and O2 reduce the PEs by 5 and 12 kcal/mol, respectively. Species **3** (O1) is more stable than species **5** (O2) because of a very strong H-bond formed over the large C_{10} ring.

Compared with the carboxylic carbonyl O3, the amide O1 and O2 are more favored for bonding with H^+. This could be explained by a larger negative charge at the O^- end of the resonance structure $O^--C{=}N^+H$ (in the O1 and O2 cases) than at the O^- end in $O^--C{=}O^+H$ (in the case of O3), owing to a larger electronegativity on oxygen than nitrogen in the former structure. Note that upon protonation, the conjugation in either HO–C–NH or HO–C–OH gives stability additional to the O^+-H bond formation. Unfortunately, this explanation no long holds when O3-protonation leads to an "abnormal" structure **4** (Figure 3) which exhibits conjugation in (C)O–C–NH and an exceptionally strong "⟩C=O···C(OH)$_2$−" interaction. Even so, the O1-protonated species **2** is still some 14 kcal/mol more stable than O3-protonated species **4**.

Protonating the amide N destroys the partial conjugation in the original amide group and upsets somewhat the gain in stability due to the N^+-H bond formation. The amide N2- and N3-protonated species **6** and **7** are 20–23 kcal/mol smaller in PE than the amino N1-protonated species **2** in which no amide group is disrupted. The large difference is partially attributed to the extra stability attained by the long-range C_{11} H-bonding in **2**.

The relative basicities of different sites discussed so far are completely consistent with the previous findings on diglycine.[1] Ambiguity exists, however, between the amide N and the carboxylic carbonyl O. The N site is favored for diglycine,[1] whereas the O site is favored for triglycine as a result of the exceptional stability of the "C-bond".

From the observations made in this work and the previous study,[1] the following general guidelines may be formulated for estimating the protonation ability of a basic site. In decreasing order of influence, preference to a particular site X for protonation depends on (with examples enclosed in parentheses) the following: (1) the stability brought to the molecular system in forming the new X^+-H bond (N–H > O–H); (2) the additional stability arising from conjugation effect (HO–C–NH > HO–C–OH); and (3) the overall strength of H-bonding and other significant nonbonded interactions in the protonated species (species **2** > **3**, **4** > **5**, **6**, **7**). Using these general guidelines, the relative protonation ability of the four sites in diglycine and six sites in triglycine may be explained satisfactorily. The conclusion is shown below.

Basicity: amino N > amide carbonyl O > amide N, carboxylic carbonyl O

Table 3. Proton Affinity (PA) and Gas-Phase Basicity (GB) of Triglycine[a]

A. Thermodynamic Properties for Relevant Species[b]

property	Gly_3(**1**)	Gly_3H^+(**2**)
E_{ZP}	130.325	139.807
$E - E_0$	7.834	7.740
S	112.268	110.307

B. Calculated and Experimental Results

result	PA	GB
calculated[c]		
3-21G//3-21G	237.67	229.32
6-31G*//3-21G	227.11	218.76
6-31G*//6-31G*	227.90	219.55
experimental		218.5 ± 2.4

[a] PA and GB in kcal/mol at 298.15 K and 1 atm. See text for details. [b] Calculated at the 3-21G//3-21G level. E_{ZP} and E in kcal/mol; S in cal/(K·mol). [c] The same thermodynamic data in part A above are used. Results correspond to the three different ΔE_{SCF} in Table 1.

Gas-Phase Basicity. The systematic procedure described previously[1] is followed here for finding the theoretical proton affinity (PA) and gas-phase basicity (GB). The protonation reaction corresponding to the measured GB of triglycine is assumed to involve only the most stable neutral species (**1**) and the protonation product species of the lowest energy (**2**):

$$Gly_3(\mathbf{1}) + H^+ \rightarrow Gly_3H^+(\mathbf{2}) \qquad (5)$$

To calculate the change of Gibbs free energy (ΔG_r), enthalpy (ΔH_r), and entropy (ΔS_r) for this reaction at 298 K and 1 atm, the following equations are used:

$$\Delta G_r = \Delta H_r - T\Delta S_r \qquad (6)$$

$$\begin{aligned} \Delta H_r &= E(Gly_3H^+) - E(Gly_3) - E(H^+) + \Delta(PV) \\ &= \Delta E_{SCF} + \Delta E_{ZP} + \Delta(E - E_0) - 1.48 \text{ kcal/mol} \end{aligned} \qquad (7)$$

$$\begin{aligned} -T\Delta S_r &= -T[S(Gly_3H^+) - S(Gly_3) - S(H^+)] \\ &= -T\Delta S + 7.76 \text{ kcal/mol} \end{aligned} \qquad (8)$$

Here the E_{SCF} listed in Table 1 is taken as E_{elec} because E_{corr} is omitted. E_{ZP} is the zero-point energy calculated from harmonic vibrational frequencies. E and E_0 are internal energies at 298 and 0 K. S is the absolute entropy at 298 K. Finally, $\Delta W = W(Gly_3H^+) - W(Gly_3)$, where W represents E_{SCF}, E_{ZP}, $E - E_0$, or S. The necessary thermodynamic properties for Gly_3(**1**) and Gly_3H^+(**2**) at the 3-21G//3-21G level are listed in Table 3A. Using these data and after scaling ΔE_{ZP} by a factor of 0.91,[28]

(28) Grev, R. S.; Janssen, C. L.; Schaefer, H. F., III *J. Chem. Phys.* **1991**, *94*, 5128.

the equations are simplified to

$$\Delta H_r = \Delta E_{SCF} + 7.06 \text{ kcal/mol} \quad (9)$$

$$-T\Delta S_r = 8.35 \text{ kcal/mol} \quad (10)$$

Substituting the ΔE_{SCF} values of Table 1 and applying the relations

$$PA = -\Delta H_r \quad (11)$$

$$GB = -\Delta G_r \quad (12)$$

theoretical values corresponding to the three different sets of electronic energy (3-21//3-21G, 6-31G*//3-21G, and 6-31G*//6-31G*) are obtained. The calculated results are compared to the experimental GB in Table 3B. Specifically, the theoretical GBs derived from the 6-31G* basis, 218.76 and 219.55 kcal/mol, are in close agreement with experiment, while the GB of 3-21G is some 10 kcal/mol too high.

From eqs 3−12, PA and GB are shown to relate directly to "$-\Delta E_{SCF}$", the electronic protonation energy (PE). It is important to note that the PE calculated from the 6-31G* basis is some 10 kcal/mol more accurate than the PE derived from the 3-21G basis according to their differences from the experimental value (*vide supra*). This conclusion drawn from the triglycine study is further supported by similar comparisons of protonation energies for glycine and diglycine[27,1] and the α-D- and β-D-glucose.[29]

Finally, how reliable is the 6-31G*//6-31G* value for the GB reported here for triglycine? The question focuses on three major issues in the present procedure: (a) the use of 3-21G//3-21G vibrational frequencies; (b) the adequacy of the 6-31G* basis; and (c) the neglect of electron correlation. These issues are addressed separately below.

To investigate the dependence of zero-point energy and thermodynamic properties on the size of a basis, the term ΔE_{therm} is introduced for $Gly_n \rightarrow Gly_nH^+$,

$$\Delta E_{therm} = \Delta E_{ZP} + \Delta(E - E_0) - T\Delta S \quad (13)$$

Thus, for $Gly_n + H^+ \rightarrow Gly_nH^+$,

$$\Delta G_r = \Delta E_{SCF} + \Delta E_{therm} + 6.28 \text{ kcal/mol} \quad (14)$$

When the level of theory is changed from 3-21G//3-21G to 6-31G*//6-31G*, the ΔE_{therm} value is reduced by 0.7 kcal/mol for glycine ($n = 1$) and 0.2 kcal/mol for diglycine ($n = 2$).[27] These small differences suggest that ΔE_{therm} is insensitive to a basis change and that the use of 3-21G//3-21G vibrational frequencies is adequate in the 6-31G*//6-31G* calculation of GB for triglycine ($n = 3$). Using the average of the two reductions (0.5 kcal/mol) for triglycine, a corrected GB value at the 6-31G*//6-31G* level would be around (219.6 + 0.5) or 220.1 kcal/mol.

To calculate PE, it would be desirable to have diffuse functions on non-hydrogen atoms and polarization functions on the hydrogen atoms. A good compromise would be a basis such as 6-31+G**, which is the same as 6-31+G(d,p).[30,1] Calculations indicate that on going from 6-31G*//6-31G* to 6-31+G**//6-31+G**, the PE decreases by 1.17 and 0.93 kcal/mol for glycine[1] and diglycine,[27] respectively.[31] Taking the average of 1.1 kcal/mol as an estimate for triglycine, the GB at the hypothetical 6-31+G**//6-31+G** level would be approximately (220.1 − 1.1) or 219.0 kcal/mol.

For electron correlation, only one piece of relevant data for the glycine series is available: the PE of glycine is decreased by 1.00 kcal/mol on going from HF/6-31+G** to MP4/6-31+G** at the HF/6-31+G** geometries.[1,31] In other words, ΔE_{corr} amounts to 1.00 kcal/mol at the MP4 level in the 6-31+G** basis. Assuming the same correlation correction holds for protonation in chemically similar systems,[1] the GB of triglycine at the hypothetical MP4/6-31+G**//6-31+G** level is estimated to be (219.0 − 1.0) or 218.0 kcal/mol.

The last consideration is the use of a basis larger than 6-31+G** such as 6-31+G(2d,2p); the latter comprises an additional set of polarization functions on each atom than the former.[30] To deduce the effects from this change, calculations on the model compound methylamine (CH_3NH_2) have been performed.[27,29] At the HF/6-31+G** geometries, ΔE_{corr} amounts to 1.31 kcal/mol for protonation of CH_3NH_2 at the MP4 level in the 6-31+G** basis, comparable to the 1.00 kcal/mol obtained under the same conditions for glycine. However, on changing to the larger basis and maintaining the same HF/6-31+G** geometries, an additional reduction of 2.18 kcal/mol in the PE of CH_3NH_2 results on going from MP4/6-31+G** to MP4/6-31+G(2d,2p). This result is consistent with the previous findings on ammonia (NH_3).[30] Again, assuming this term is roughly additive and transferable in PE calculations, the calculated GB for triglycine would be (218 − 2) or 216 kcal/mol at the hypothetical MP4/6-31+G(2d,2p)//6-31+G** level.

Future PA and GB Calculations. A previous ab initio study of model glycine and alanine dipeptides indicates that the HF/3-21G geometries are reasonably reliable, but their relative energies are not as good.[32] The assumption implicit in several computational studies on model peptides using the 4-21G//4-21G approach is that 4-21G, a small split-valence set like 3-21G, is reasonably reliable for both geometry and energy.[7,8c] The general consensus has been that a split-valence basis such as 3-21G or 4-21G yields reasonable geometries for the global minima of molecular species. To accurately differentiate the geometries and energies of closely related conformers of the same molecular species, a larger basis (including polarization functions and diffuse functions) and a correlation-gradient geometry optimization procedure are needed.[32,33]

The present work on triglycine yields results consistent with earlier findings. Data in Figures 1 and S-1 and Tables 1, S-I, and S-II for the neutral species **1*** and **1** and the amino N-protonated species **2** indicate that the HF/3-21G geometries are indeed reasonably close to the HF/6-31G* geometries; the small differences in ΔE_{SCF} between the 6-31G*//3-21G and 6-31G*//6-31G* values (1.76 and −0.79 kcal/mol from Table 1) are particularly illuminating.[31,32] The 3-21G basis used in this work is comparable to 4-21G used previously[7,8c] with respect to geometry, while 6-31G*, which is a larger basis than 3-21G or 4-21G and includes polarization functions, is shown to give more accurate electronic protonation energy.

(29) (a) Cassady, C. J. *Gas Phase Basicities of Small Carbohydrates*; The 41st ASMS Conference on Mass Spectrometry, San Francisco, CA, May 1993. (b) Jebber, K. A.; Chung-Phillips, A. *Computational Studies of Neutral and Protonated α-D- and β-D-Glucose*; Abstracts of the 208th National Meeting of the American Chemical Society, Washington, DC, August 1994; American Chemical Society: Washington, DC, 1994.

(30) Del Bene, J. E.; Shavitt, I. *J. Phys. Chem.* **1990**, *94*, 5514.

(31) See Table V of ref 1. Note four typographical corrections in Table VIB for glycine in ref 1: the 6-31G*//6-31G* (scaled) and MP4/6-31+G-(d,p)//6-31+G(d,p) (scaled) values should be 208.4 and 208.7 kcal/mol for N-protonation and 195.1 and 192.0 kcal/mol for O-protonation.

(32) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C. L., II; Pople, J. A. *J. Am. Chem. Soc.* **1991**, *113*, 5989.

(33) Frey, R. F.; Coffin, J.; Newton, S. Q.; Ramek, M.; Cheng, V. K. W.; Momany, F. A.; Schafer, L. *J. Am. Chem. Soc.* **1992**, *114*, 5369.

In estimating the theoretical PA and GB, small deficiencies in the HF/3-21G geometries may not be serious. The reason is that PA and GB relate to the difference in electronic energy between two chemically distinct species, which is about two orders of magnitude larger than the energy differences between two different conformers of the same species. For example, the 6-31G*//3-21G value for ΔE_{SCF} in Table 1 is −234.17 kcal/mol for Gly_3(**1**) → Gly_3H^+(**2**), which is two orders of magnitude larger than 1.16 kcal/mol for Gly_3 (**1**) → Gly_3(**1***).

As compared with 6-31G*//6-31G*, it is apparent that 6-31G*//3-21G represents a significant saving of computation resources and brings nearly the same results. For this reason, the 6-31G*//3-21G approach is recommended for studying protonations of comparable or larger peptides. Further improvements on the calculated PA or GB from inclusion of electron correlation in a basis larger than 6-31G* may be estimated as shown above for triglycine: a reduction of 3−6 kcal/mol from the 6-31G*//3-21G value can be expected.

Comparison of Theory to Experiments. Recently, the gas-phase basicities of polyglycines, Gly_n (n = 1−6), have been determined in our laboratory using proton transfer reactions in a Fourier transform ion cyclotron resonance mass spectrometer.[1] The GB of triglycine was bracketed between isopropylamine (GB = 218.1 kcal/mol[16]) and pyridine (GB = 219.8 kcal/mol[16]), yielding a value of 218.9 ± 2.9 kcal/mol.[1] Because triglycine is the focus of the present paper, we have pinpointed the GB of triglycine further by performing additional proton transfer reactions with *sec*-butylamine (GB = 218.8 kcal/mol[16,17]). Proton transfer to *sec*-butylamine was observed, placing the GB of triglycine at 218.5 ± 2.4 kcal/mol. Using the theoretically determined entropy term, this experimental value yields a PA of 226.8 ± 2.4 kcal/mol. These results are in agreement with our theoretical GB and PA for triglycine of 219.6 and 227.9 kcal/mol, respectively.

The major source of experimental GB values in the literature is a compilation by Lias *et al.*[17] Recently, two other basicity schemes have been proposed to refine the most basic part of this ladder.[16,18] These studies are in agreement that an upward revision to the ladder of Lias *et al.*[17] is needed, but disagree about the magnitude of this adjustment. All experimental GB and PA values given in the previous paragraph are referenced to one of the recent scales, the Meot-Ner and Sieck[16] basicity ladder. If the earlier scale of Lias *et al.*[17] is used, the experimental GB and PA values for triglycine obtained from our study are 211.4 ± 2.4 and 219.7 ± 2.4 kcal/mol, respectively. These values may be considered as lower limits for the experimental basicities. It is likely that a revised basicity ladder will have values that are intermediate to those currently found in the Lias compilation[17] and the Meot-Ner and Sieck[16] work.

Three other recent experimental studies have also focused on the gas-phase basicities and proton affinities of polyglycines. Our experimental and theoretical GB and PA values are intermediate to the values reported in these studies. (For purposes of comparison, all values have been referenced to the Meot-Ner and Sieck[16] ladder; note that if the Lias *et al.*[17] scale is used, all GBs and PAs discussed in this paragraph are lower by approximately 8 kcal/mol.) Wu and Lebrilla[2] obtained a GB for triglycine of 215.0 ± 2.0 kcal/mol using proton transfer reactions in a procedure similar to that employed in our study. These workers also reported a PA of 226.0 kcal/mol for triglycine using an entropy term derived from semiempirical AM1 calculations. Fenselau and co-workers obtained a PA of 230.6 kcal/mol using a kinetic method of collision-induced dissociation (CID) on proton bound dimers containing triglycine and various reference compounds.[3] Using a modified version of this procedure, which varies the collision energies and considers entropy effects, this group has recently revised their value for the PA of triglycine upward to 236.4 kcal/mol.[34]

To provide additional information about the protonation process, low-energy CID was performed in our laboratory on protonated triglycine ions. Single-collision conditions were employed to minimize any rearrangement that might occur during these experiments. The lowest energy process involves cleavage at the central carbonyl to produce a b_2 ion (using Roepstorff and Fohlman nomenclature[35] as modified by Biemann[36]), reaction 15.

$$Gly_3H^+ \rightarrow NH_2CH_2(C{=}O)NHCH_2(C{=}O)^+ + Gly \quad (15)$$

In the dissociation of Gly_3H^+, the second product to appear as the collision energy is increased involves cleavage at the N-terminal carbonyl producing a y_2 ion, reaction 16.

$$Gly_3H^+ \rightarrow NH_2CH{=}C{=}O + Gly_2H^+ \quad (16)$$

Dissociation to produce b and y ions is typical of the processes observed for low molecular weight peptides that do not contain highly basic amino acid residues.[37] This suggests a mobile hydrogen at the protonation site, which is consistent with the high degree of intramolecular H-bonding that is found in our ab initio structures. Other minor products also form; however, at higher collision energies $NH_2{=}CH_2^+$ dominates. This is analogous to the formation of $NR_2{=}CH_2^+$ immonium ions that occurs in alkylamine CID spectra[38] and also suggests that the terminal amino group is a major protonation site for triglycine.

Concluding Remarks

In solution, the three-dimensional structure and biological activity of peptides is affected by internal hydrogen bonding, which in turn is related to the locations of basic sites. The present ab initio study shows that even the simplest tripeptide, triglycine, has a complex series of stable protonated structures in the gas phase which involve extensive intramolecular hydrogen bonding. For the most basic protonation site, the terminal amino nitrogen, a gas-phase basicity (GB) of 219.6 kcal/mol was calculated. An experimental GB of 218.5 ± 2.4 kcal/mol was obtained by proton transfer reactions in a mass spectrometer. This close agreement between experimental and theoretical results provides an important check on the validity of both approaches.

Acknowledgment. The computational work was supported by Miami University through the Miami Computing and Information Services. We thank Mr. Stephen M. Moore for his valuable assistance with the Gaussian 92 program. Financial support of the mass spectrometry research by the National Institutes of Health (R15-GM47657-01A1) and the Ohio Board of Regents Academic Challenge Program is gratefully acknowledged.

(34) Cheng, X.; Wu, Z.; Fenselau, C. *J. Am. Chem. Soc.* **1993**, *115*, 4844.
(35) Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11*, 601.
(36) Biemann, K. *Biomed Environ. Mass Spectrom.* **1988**, *16*, 99.
(37) (a) Lippstreu-Fisher, D. L.; Gross, M. L. *Anal. Chem.* **1985**, *57*, 1174. (b) Hunt, D. F.; Yates, J. R., III; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233. (c) Yeh, R. W.; Grimley, J. M.; Bursey, M. M. *Biol. Mass Spectrom.* **1991**, *20*, 443.
(38) Reiner, E. J.; Harrison, A. G; Bowen, R. D. *Can. J. Chem.* **1989**, *67*, 2081.

Appendix: A Detailed Analysis of Hydrogen Bonding

The ring composition of each H-bond (Y···H−X or X−H···Y) in Table 2A is given below.

C_5: N1···H−N2 and N1···H−$N2^+$, 1,5 interaction in N−C−C−NH; N2···H−$O2^+$, 1,5 interaction in N−C−C−OH; $N1^+$−H···O1, $N2^+$−H···O2, and $N3^+$−H···O3, 1,5 interaction in HN−C−C=O.

C_7: O1···H−N3 and O1···H−$N3^+$, 1,7 interaction in O=C−N−C−C−NH.

C_{10}: $O1^+$−H···O3 and O1···H−$O3^+$, 1,10 interaction in HO−C−N−C−C−N−C−C=O.

C_{11}: N1−H···O3 and $N1^+$−H···O3, 1,11 interaction in HN−C−C−N−C−C−N−C−C=O.

Here, sequencing of atoms in an H-bond follows the way atoms appear in Figure 2 from left to right; the numerical label for each N or O atom corresponds to the specification in Figure 1; protonated N or O in the donor bond (X−H) is labeled as N^+ or O^+. Using this system of notations, the location of the H-bond in a triglycine chain and the nature of the donor can be easily identified. The atom labels also convey the size of the ring. For example, the notation N1−H···O3 implies linking N in glycine unit 1 to O in glycine unit 3 and therefore this H-bond must correspond to a long-range interaction.

It would be instructive to compare these results for triglycines to those of the lower members of the series. Due to shorter chain lengths, glycines and diglycines have mainly C_5 form for H-bonding, with only one instance of a C_8 form in a Gly_2H^+ species that links the terminal ammonium and carboxylic groups.[1] The average H-bonding distance is around 2.0 Å.[1] In the case of triglycines, most H-bonding distances in the C_5 and C_7 rings are "normal", i.e., around 2.0 Å. The exception is the C_7 ring in Gly_3H^+(**7**), which shows a H-bonding distance as low as 1.568 Å. This is due to protonation at N2 that brings a change in the hybridization on N from sp^2 to sp^3, breaks the planarity constraint, and gives the proton a greater access to the carbonyl oxygen.

The important structural forms emerging from this triglycine study are the larger rings, C_{10} and C_{11}, present in the neutral (**1**) and the more stable protonated (**2**−**4**) species. A most distinctive feature in these larger rings is the closeness of approach between the acceptor atom (Y) and hydrogen: the Y···H distance drops to 1.659 Å in **2**, 1.580 Å in **3**, and 1.733 Å in **4**, instead of the normal distance around 2.0 Å. Clearly, the H-bond with such short bonding distances represents exceptionally strong attractions. In view of the greater stability of **2**−**4** relative to **5**−**7**, it may be assumed that C_{10} and C_{11} are able to stabilize triglycines better than C_5 and C_7.

The predominance of the smaller ring forms (C_5 in particular) in triglycine does not necessarily contradict the assumption that larger rings are energetically preferred. The reason that larger rings fail to materialize in certain species is owing to geometrical or configurational constraints. For instance, C_{11} links the terminal N and O atoms to form N1−H···O3; protonation at a position on the backbone of the ring (O2 in **5**, N2 in **6**, or N3 in **7**) obviously proves disruptive, sterically or otherwise. On the other hand, C_{10} requires a protonated O atom to form $O1^+$−H···O3 or O1···H−$O3^+$; a lack of either $O1^+$ or $O3^+$ in **5**, **6**, and **7** precludes the existence of this H-bond.

Supplementary Material Available: Figure S-1: HF/6-31G* minimum-energy structures for the extended (**1***) and folded (**1**) neutral triglycines and the amino N-protonated triglycine (**2**) with complete atom labels; Table S-I: HF/6-31G* Optimized Geometrical Parameters for Triglycines (**1***, **1**, and **2**); Table S-II: HF/3-21G Optimized Geometrical Parameters for Triglycines (**1***, **1**, and **2**); Table S-III: Maximum Internal Coordinate Residual Forces in Optimized Structures; Table S-IV: HF/6-31G* Cartesian Coordinates for Triglycines (**1***, **1**, and **2**); Table S-V: HF/3-21G Cartesian Coordinates for Triglycines (**1***, **1**, ..., and **7**) (11 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, and can be ordered from the ACS; see any current masthead page for ordering information.