Polyglycine Conformational Analysis: Calculated vs Experimental Gas-Phase Basicities and Proton Affinities

Alice Chung-Phillips*

Department of Chemistry and Biochemistry, Miami University, Oxford, Ohio 45056

Received: December 15, 2004; In Final Form: May 3, 2005

Structures of neutral and protonated polyglycines (Gly_n and Gly_nH^+ with $n = 1-6$) in the vicinity of global energy minima were calculated using the density functional theory at the B3LYP/6-311++G** (A) and B3LYP/6-31+G** (B) levels. Ninety-three structures were chosen for conformation and protonation studies. Geometries of the peptides are found to vary from open chains to multiple rings. Intramolecular hydrogen bonding is deduced to be the driving force for conformational stability. The preferred protonation sites are shown to be the terminal nitrogen atom and its adjacent amide oxygen atom. Structural series are developed according to geometrical form, hydrogen bonding, and protonation site. Physical factors that influence the relative electronic and thermodynamic stabilities of different structural series are examined. To obtain ab initio values of highest quality for gas-phase basicity (GB) and proton affinity (PA), electronic energies for $n = 1-6$ and thermal corrections to Gibbs free energy and enthalpy for $n = 1-3$ were calculated at level A, supplemented by thermal corrections for $n = 4-6$ at level B. Calculated GB and PA values are compared with mass spectral results obtained by the kinetic method (KM) and reaction bracketing (RB). The KM results and the ab initio values derived from structurally compatible pairs of lowest free energies are generally in good agreement, but the RB results for GB are lower by 2−8 kcal/mol for $n = 2-6$. Several reaction pathways are proposed to elucidate the experimental results. On the basis of theoretical structures consistent with the measurements, it is concluded that KM mostly samples the neutral and protonated structures of highest populations at thermal equilibrium, whereas RB targets those with sterically most accessible sites for protonation and deprotonation.

Introduction

The simplest peptide containing n residues is the polyglycine Gly_n with the molecular formula $NH_2CH_2(CONHCH_2)_{n-1}COOH$. Devoid of side chains and their functional groups, Gly_n forms the backbones of amino acids, peptides, and proteins.[1] Scientific findings from rigorous investigations on Gly_n are important to the study of a wide range of biological systems.

The biological activities of a peptide depend on its three-dimensional structure and locations of basic sites. In the gas phase, the preferred conformations, favored protonation sites, and pathways of proton migration from one site to another are the intrinsic structural properties of a peptide and its protonated ions. A versatile experimental tool to study gas-phase ion chemistry of biomolecules is mass spectrometry.[2] In the positive ion analysis, the protonated ion is of primary concern. The location of the proton affects the fragmentation pattern of the ion which in turn provides structural information for the identification of the unknown peptide.[3] Two intrinsic thermodynamic properties of a peptide M are gas-phase basicity (GB) and proton affinity (PA), which can be measured quantitatively as the respective $-\Delta G$ and $-\Delta H$ of the protonation reaction $M + H^+ \rightarrow MH^+$. The GB and PA values of a number of oligopeptides were measured by the kinetic method (KM) and reaction bracketing (RB) in the past decade. A critical review on this topic was given by Harrison.[4] For polyglycines Gly_n, Wu and Fenselau[5] estimated the GBs and PAs of $n = 2-10$ using KM, Wu and Lebrilla[6] determined the GBs of $n = 1-5$ using RB, and Zimmerman and Cassady[7] measured the GBs of $n = 1-6$ using both KM and RB. Comparisons of the reported data showed substantial discrepancies between the KM and RB results.[4]

The most direct approach to find the energies and structures of neutral and protonated molecules is to apply the ab initio molecular orbital theory based on quantum mechanics.[8] A comprehensive review on ab initio calculations of amino acids and peptides by Schäfer, Newton, and Jiang[9] provides a valuable source of references. For gaseous glycine and its protonated ions, Gly and $GlyH^+$, the level of theory[8] progressed from Hartree−Fock (HF), second-order Møller−Plesset perturbation (MP2), to Becke 3-parameter-Lee−Yang−Parr (B3LYP) functional[10] of the density functional theory (DFT), in combination with small (3-21G) to large (6-311++G**) basis sets. Representative topics ranged from conformational analysis[11−16] to intramolecular proton migration.[17,18] But for Gly_n and Gly_nH^+ with $n > 3$, rigorous analyses were rarely attempted due to a lack of practical procedures to circumvent the seemingly insurmountable work required by the ab initio approach.

The first GB and PA calculations at the HF/3-21G and HF/6-31G* levels were carried out by Zhang et al. of this laboratory to provide pertinent data for Gly, Gly_2, Gly_3, Ala, Ala_2, GlyAla, AlaGly, and their protonated species.[7,19,20] The work proceeded to higher theoretical levels for Gly and $GlyH^+$,[16,17,21] resulting in the best calculated GB and PA values for glycine, 203.5 and 211.1 kcal/mol, at the composite level MP4/6-311+G(3df,2p) over MP2/6-311+G** geometries. The ab initio values are in excellent agreement with the NIST values,[22] 203.7 and 211.9 kcal/mol, evaluated by Hunter and Lias. From other laboratories, Strittmatter and Williams[23] computed six PAs of Gly_n ($n = 1$,

* To whom correspondence should be addressed. E-mail: philliac@muohio.edu.

10.1021/jp040758v CCC: $30.25
Published on Web 06/10/2005

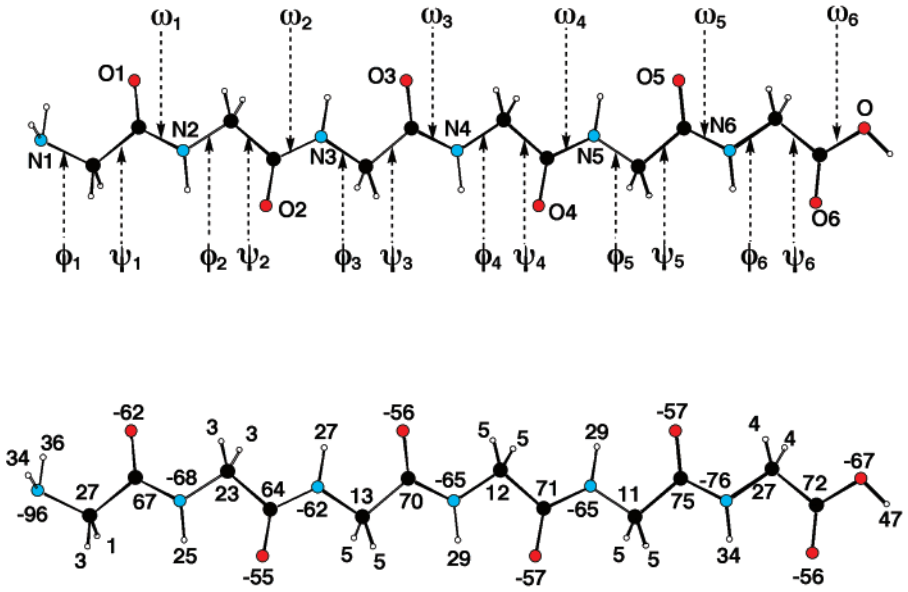

Figure 1. The extended form of hexaglycine: conformational dihedral angles and notations for nitrogen and oxygen atoms (top); atomic partial charges in 10^{-2} e at the B3LYP/6-311++G** level (bottom). Atoms are identified by color (H, none; C, black; N, blue; O, red).

3−5, 7, and 10) at the B3LYP/6-31G* level starting from Merck molecular force field (MMFFs)[24] geometries; the plot of their calculated PAs vs *n* resembles well the plot from mass spectral PAs of Wu and Fenselau. In a mechanistic study of proton migration and tautomerism in Gly_3H^+, Rodriquez et al.[25] deduced the GB and PA values of Gly_3 at the B3LYP/6-31++G** level; their values agree favorably with mass spectral values.

This work is part of a continuing project to bring ab initio applications to biomolecules.[26] Recent advances in supercomputer technology have facilitated accurate studies of molecules of the size of hexaglyine using the B3LYP method with large basis sets. The present objective is to find relevant structures of Gly_n and Gly_nH^+ ($n = 1-6$) for conformation and protonation studies. To achieve this goal a proficient algorithm for optimizing peptide geometries using internal coordinates is developed and conformational potential energy surfaces (PESs) around the global minima of the respective species are searched. The resulting structures supply the source data for an in-depth analysis of conformational properties, exploration of pathways for protonation and deprotonation, rigorous calculations of GB and PA, and a rational explanation at the molecular level for the highly different GB values measured by KM and RB. The calculated structures and related studies bring new knowledge and physical insight to polyglycines.

The work represents the first major attempt to carry out accurate ab initio calculations for neutral and singly protonated peptides containing more than three residues. The extensive compilation of structural and energetic data, accompanied by simple schemes developed for conformational analysis and protonation mechanisms, provide important references to future theoretical and experimental studies of gas-phase ion chemistry of polypeptides.

Computational Methods

Theoretical Levels. The B3LYP and MP2 levels of theory with basis sets comparable or larger than 6-31+G** have been found reasonably accurate.[25,27,28] Although MP2 incorporates electron correlation more completely than B3LYP, MP2/6-31+G** geometry optimizations for $n > 3$ would overburden the current computing capacity. For this reason, B3LYP/6-311++G** (A) and B3LYP/6-31+G** (B) are used. Our previous work on glycine demonstrated that level A yields better protonation energy and geometry than MP2/6-31+G** and gives results comparable to those of MP2/6-311++G**.[21]

Initial Geometries. The extended form of hexaglycine in Figure 1 is used to illustrate a polyglycine structure. The conformation (top) is specified by the conformational dihedral angles (CDAs) using the conventional symbols φ (phi), ψ (psi), and ω (omega) for peptides.[1] The structure may be viewed regionally in terms of conformational units $i = 1-6$ determined by the corresponding φ_i, ψ_i, and ω_i. In this study the low-energy conformers on the PESs of Gly_n and Gly_nH^+ with $n = 1-6$ are searched using internal coordinates for geometries. A *z*-matrix[8] that contains the CDAs explicitly is constructed to give a precise definition to the peptide conformation. The *z*-matrix elements are sequenced to attain maximum ease in transferring geometrical parameter values of individual conformational units from one conformer to another. Using this procedure a library of low-energy conformers is built from glycine to hexaglycine.[26c] The stationary point of lowest electronic energy is designated as the global minimum.

Gaussian Calculations. The Gaussian 98 computer program is employed to carry out all requisite calculations.[29] For clarity, the Gaussian commands are *italicized*. Given a trial structure, geometry optimization (*opt* or *opt = calcall*) using the self-consistent-field iterative procedure is carried out to determine the optimized geometry and electronic energy, followed by calculations of harmonic vibrational frequencies ν_i (*freq* or *calcall*) to produce the relevant thermal corrections. All calculated structures satisfy the default convergence criteria of the Gaussian program. A local minimum has all positive ν_i, while each transition state (TS) has only one negative ν_i.

The *freq* or *calcall* procedure involves analytical force constants and therefore requires significantly greater computer memory and longer execution time than the default *opt* procedure. In fact, the memory requirement constitutes the bottleneck to ab initio applications to large molecular systems. Given the available computer resources, it is feasible to optimize geometries at level A up to $n = 6$. To calculate frequencies, level A is practical up to $n = 3$, but a lower level (level B) needs be used for $n = 4-6$.

TABLE 1: Relative Electronic Energies (ΔE_e), Thermal Corrections (ΔG_{tc}), and Gibbs Free Energies (ΔG) of Polyglycine Structures at the B3LYP/6-311++G (A), B3LYP/6-31+G** (B), and Composite (A/B) Levels, in kcal/mol[a]**

glycine				diaglycine				triglycine			
structure	ΔE_e(A)	ΔG_{tc}(A)	ΔG(A)	structure	ΔE_e(A)	ΔG_{tc}(A)	ΔG(A)	structure	ΔE_e(A)	ΔG_{tc}(A)	ΔG(A)
1e	0.00	0.00	0.00	**2n**	−1.64	0.36	−1.28	**3f**	−3.87	2.94	−0.93
1m	0.42	0.46	0.89	**2f**	−1.61	0.32	−1.29	**3m**	−3.01	4.04	1.04
1c	1.45	0.11	1.56	**2m**	−1.51	2.32	0.81	**3e**	0.00	0.00	0.00
1b	1.52	−0.60	0.92	**2e**	0.00	0.00	0.00	**3oh**	−235.92	10.93	−224.98
1d	5.61	−0.24	5.37	**2nT**	2.19	1.20	3.39	**3fh**	−235.31	13.42	−221.89
1dT	12.62	−0.99	11.64	**2eh**	−229.22	9.89	−219.33	**3lh**	−233.62	10.60	−223.02
1eh	−219.14	9.12	−210.02	**2lh**	−228.17	9.72	−218.45	**3eh**	−233.19	10.65	−222.54
1bh	−214.66	8.52	−206.15	**2mh**	−226.63	10.07	−216.56	**3mh**	−230.46	11.40	−219.06
tetraglycine				pentaglycine				hexaglycine			
structure	ΔE_e(A)	ΔG_{tc}(B)	ΔG(A/B)	structure	ΔE_e(A)	ΔG_{tc}(B)	ΔG(A/B)	structure	ΔE_e(A)	ΔG_{tc}(B)	ΔG(A/B)
4g	−6.67	7.84	1.17	**5g**	−10.32	9.81	−0.51	**6g**	−10.85	12.00	1.15
4f	−5.98	5.94	−0.44	**5f3**	−9.36	10.63	1.27	**6g2**	−8.43	12.36	3.93
4m	−3.78	5.37	1.59	**5m**	−3.88	6.89	3.01	**6m**	−3.18	8.62	5.44
4e	0.00	0.00	0.00	**5e**	0.00	0.00	0.00	**6e**	0.00	0.00	0.00
4gh	−245.14	16.77	−228.37	**5gh**	−251.20	18.98	−232.22	**6f3h**	−254.92	20.63	−234.30
4oh	−239.27	10.79	−228.48	**5f2h**	−250.13	17.80	−232.33	**6g2h**	−251.37	20.10	−231.27
4lh	−236.17	10.75	−225.42	**5lh**	−237.55	11.23	−226.32	**6lh**	−238.41	11.40	−227.16
4mh	−233.18	12.90	−220.28	**5mh**	−234.55	14.63	−219.92	**6mh**	−233.33	16.26	−217.07

[a] See Figures 2−6 and Tables 1S, 2S. ΔG(A) = ΔE_e(A) + ΔG_{tc}(A) and ΔG(A/B) = ΔE_e(A) + ΔG_{tc}(B).

Electron population analysis (*pop*) and atomic partial charges (*pop* = *chelpg*) are included in the discussion. An example for the CHELPG charges[30] is provided for Gly_6 in Figure 1 (bottom). For mechanistic studies, the TS is found using *opt* = *qst2* accompanied by optimized geometries for the initial and final states or *opt* = *ts* starting with a trial TS geometry. In some cases the TS is verified using the intrinsic reaction coordinate (*irc*). For protonation calculations, the basis set superposition error (BSSE) is evaluated using *opt massage*, which is carried to convergence for most Gly_n with $n = 1-3$ and up to about 15 cycles otherwise. In the situation where the ghost-atom is located in a congested area, causing severe energy fluctuation, the result from the first cycle is used, which is equivalent to a single-point (sp) BSSE.

Computational Results

Ninety-three Gly_n and Gly_nH^+ structures (10 for $n = 1$, 18 for $n = 2$, 22 for $n = 3$, 16 for $n = 4$, 13 for $n = 5$, and 14 for $n = 6$) were selected for conformation and protonation studies. Values of electronic energy E_e, zero-point energy E_{ZP}, and thermal corrections H_{tc} and G_{tc} to enthalpy, H, and Gibbs free energy, G, at 298.15 K and 1 atm are provided in Tables 1S and 2S of Supporting Information. The directly computed values are shown in Table 1S for the extended conformers of Gly_n (**1e**−**6e**) chosen as the reference structures; values relative to the reference values are listed for all Gly_n and Gly_nH^+ structures for $n = 1-6$ in Table 2S. The ΔE_e, ΔG_{tc}, and ΔG data of interest to energy analysis are presented in Table 1. The CDAs of the structures are provided in Table 3S. Examples of calculated bond lengths, bond angles, and dihedral angles are provided in the output *z*-matrixes for the neutral, N- and O-protonated triglycines in Table 4S.

Structures representing lower to lowest E_e and G and those deemed to have significant presence in the KM and RB measurements are shown in Figures 2−6 for $n = 1-6$. Additional structures of low E_e are presented in Figure 7 for $n = 2$, 3, 4, and 6. For most structures, the graph shows the N-terminus on the left and the main chain extending horizontally to the right before bending to the left. But for the folded structures of $n = 4-6$, a clearer view of H-bonds is obtained by flipping the original graph from right to left by 180° and turning clockwise by 90° in the other two directions.

In Figures 2−6 the neutral group is placed before the protonated group; members within each group are presented in order of decreasing electronic stability. Structures are named n*x*, n*y*h, and n*z*T for the neutral, protonated, and TS species where *x*, *y*, and *z* are indicators of certain geometrical and physical properties. The Gly_n minima are exemplified by *e*, *c*, and *m* in the extended form, and *f* and *g* in the folded form. For the Gly_nH^+ minima, *eh*, *fh*, and *gh* are named for amino N-protonations with "h" added to the parent neutral structures *e*, *f*, and *g*, but the names *lh*, *oh*, and *mh* are used to indicate different modes of O-protonations. Each TS species has "T" appended to the name of the most relevant local minimum in question. In Figure 7 the structures are grouped for recognition; ΔG data are provided for comparisons with those in Table 1.

Conformational Analysis

Ab initio conformational analysis of peptides are generally carried out for model neutral peptides that replace the terminal NH_2CH_2 and CH_2COOH by H or CH_3 to replicate the main-chain conformations of large peptides.[31,32] In recent years greater attention has been given to the intermolecular interactions of small peptides or peptide zwitterions with polar solvent molecules.[33] As for the *bona fide* gas-phase Gly_n and Gly_nH^+ where the two terminal groups are kept intact, much is to be learned about the influence of intramolecular H-bonds on conformational stability. In fact, the exceptional physical dexterity and distinctive chemical properties of the terminal groups bring new possibilities and complexity to gas-phase conformations.

As a guide to conformational analysis, 11 structural series (*e*, *c*, *m*, *f*, *g*, *eh*, *lh*, *oh*, *mh*, *fh*, *gh*) covering 56 energy minima are catalogued in Table 2. The combined series notations *ec* (for *e* plus *c*), *fg* (for *f* plus *g*), and *fgh* (for *fh* plus *gh*) are used occasionally in plotting and discussion. Structures in each series generally follow a similar hydrogen bonding(H-bonding) pattern as the chain length (n) increases, but the relative stability among structures of the same n in different series may change significantly as n increases. A comparison of the relative electronic (ΔE_e) and free (ΔG) energies of the neutral series *e*,

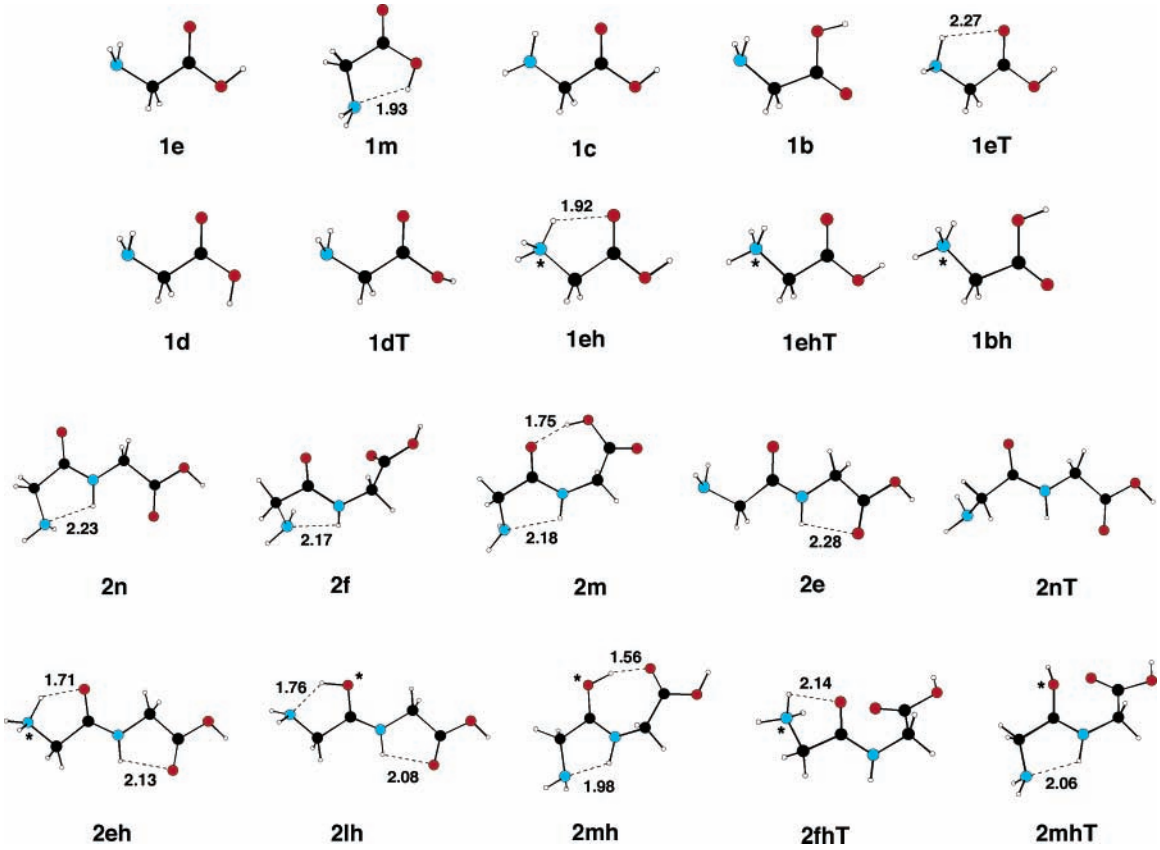

Figure 2. Structures of glycine **1e**−**1bh** and diglycine **2n**−**2mhT** optimized at the B3LYP/6-311++G** level. H-bond lengths in Å are shown by dashed lines. The protonated atom is marked by an asterisk.

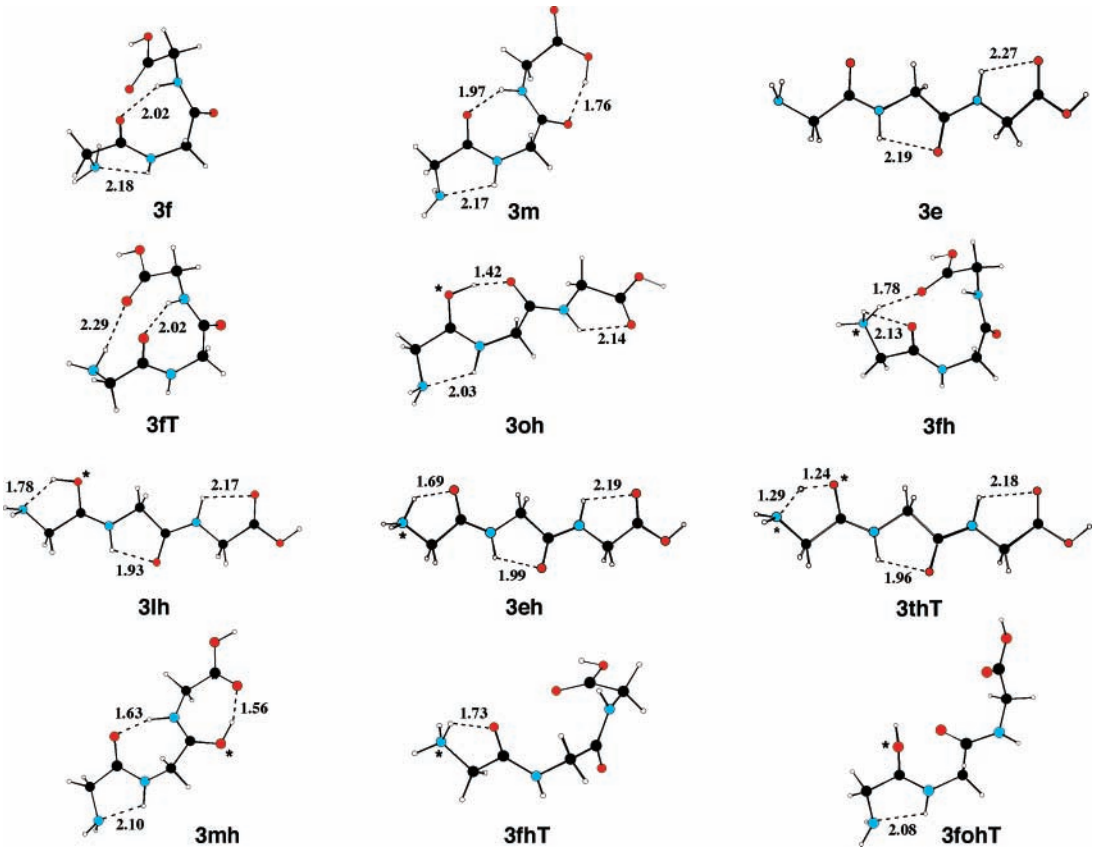

Figure 3. Triglycine structures **3f**−**3fohT** optimized at the B3LYP/6-311++G** level.

m, and *fg* is made in Figure 8 to show the role of H-bonding on stabilities. Free energy plots of the neutral and protonated series *e*, *c*, *m*, *fg*, *eh*, *lh*, *oh*, *mh*, and *fgh* are shown in Figure 9 to reveal the structures of lower to lowest *G* important in

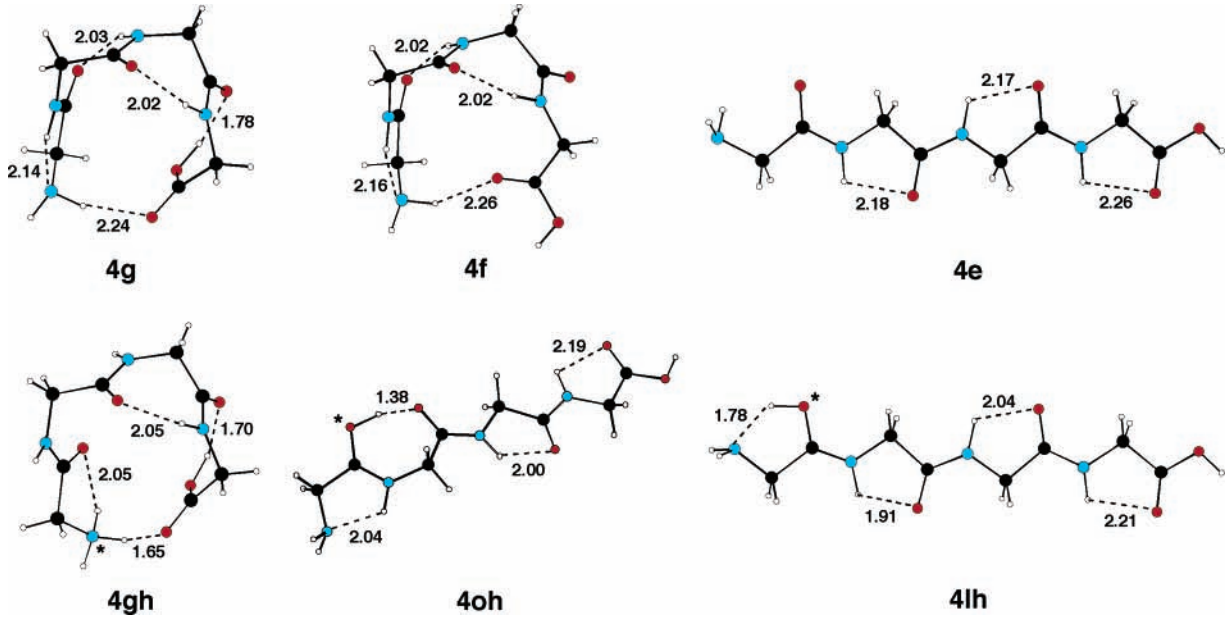

Figure 4. Tetraglycine structures **4g**−**4lh** optimized at the B3LYP/6-311++G** level.

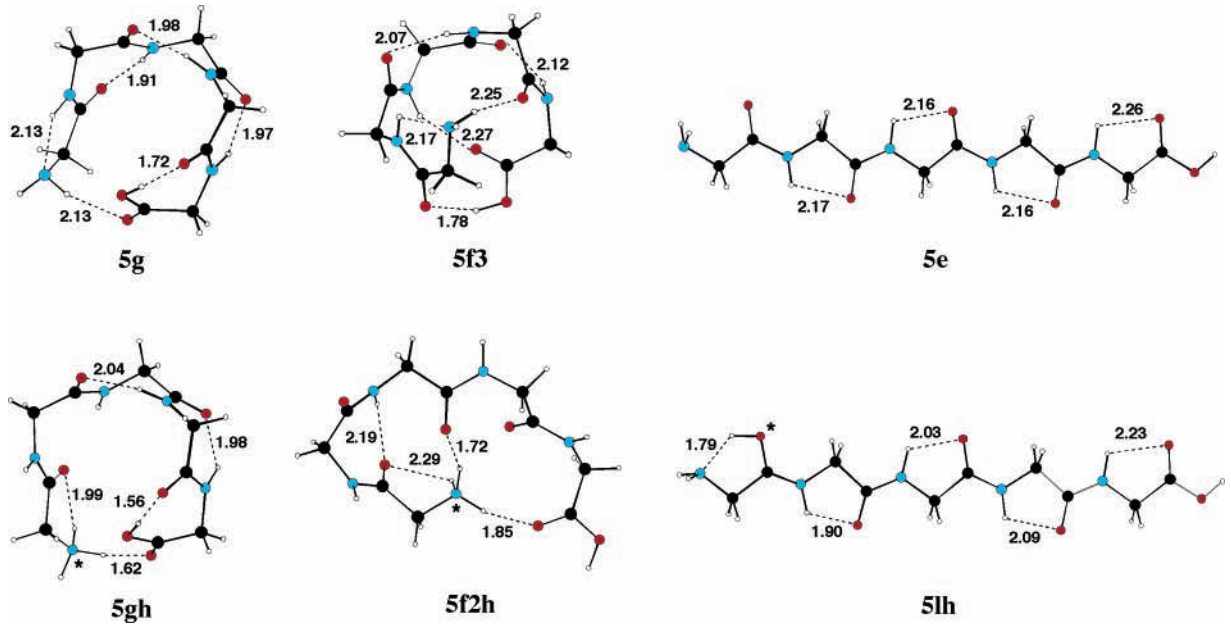

Figure 5. Pentaglycine structures **5g**−**5lh** optimized at the B3LYP/6-311++G** level.

experimental measurements. Here conformational stability is made synonymous with electronic stability and thermodynamic stability with free energy stability. In the following a general description of H-bonds and the dispositions of the terminal groups is given before specific topics are discussed.

The H-bond is represented by C_m(X−H···Y) which has a ring-like structure containing m atoms closed by H···Y. Atom labels (Figure 1) may be employed for the H-donor X and acceptor Y to identify a specific H-bond. The H···Y distance r is the H-bond length which has been used widely as an indicator of bond strength. In Figures 2−7, r < 2.3 Å is shown with a dashed line to indicate the presence of a normal to strong H-bond.[34] A protonated H-bond is noted with an asterisk, C_m*-(X*−H···Y), where X* is the protonation site. The main-chain, ring-closing, and N1-protonated H-bonds of the NH···O type, are designated simply as C_m and C_m*. Special designations are given to certain interactions involving terminal groups and the OH···O, OH···N, and NH···N types. (Here "main-chain" refers to the segment between terminal groups and "ring-closing" applies to a m-membered ring with m > 7.)

The N-terminal −NH_2 can be *cis* or *trans* to its adjacent amide O along the NCCO chain: *cis* as in **2e** and *trans* as in **2n**. The barrier of the coupled internal rotations around φ_1 and ψ_1 between the two minima is **2nT** (irN):

$$\mathbf{2e}\ (0.00) \rightarrow \mathbf{2nT}\ (3.39) \rightarrow \mathbf{2n}\ (-1.28) \qquad \text{(irN)}$$

where ΔG in kcal/mol are provided in parentheses. The energies show a moderate G-barrier (3−4 kcal/mol) on flipping −NH_2 from the "up" to the "down" position. At the N-terminus, there are three kinds of interactions between the amino group and its adjacent amide group: bifurcated NH_2···O, $C_5(NH_2)$ in *e*, single NH···O, C_5(NH) in *c*, and NH···N involving the amino N lone pair, C_5(N1) in *m*, *oh*, *mh*, *f*, and *g* with the exception of **6g2**.

The C-terminal −COOH can be *cis* or *trans* along the OCOH chain: *cis* as in **1e** and *trans* as in **1d**. The barrier of internal

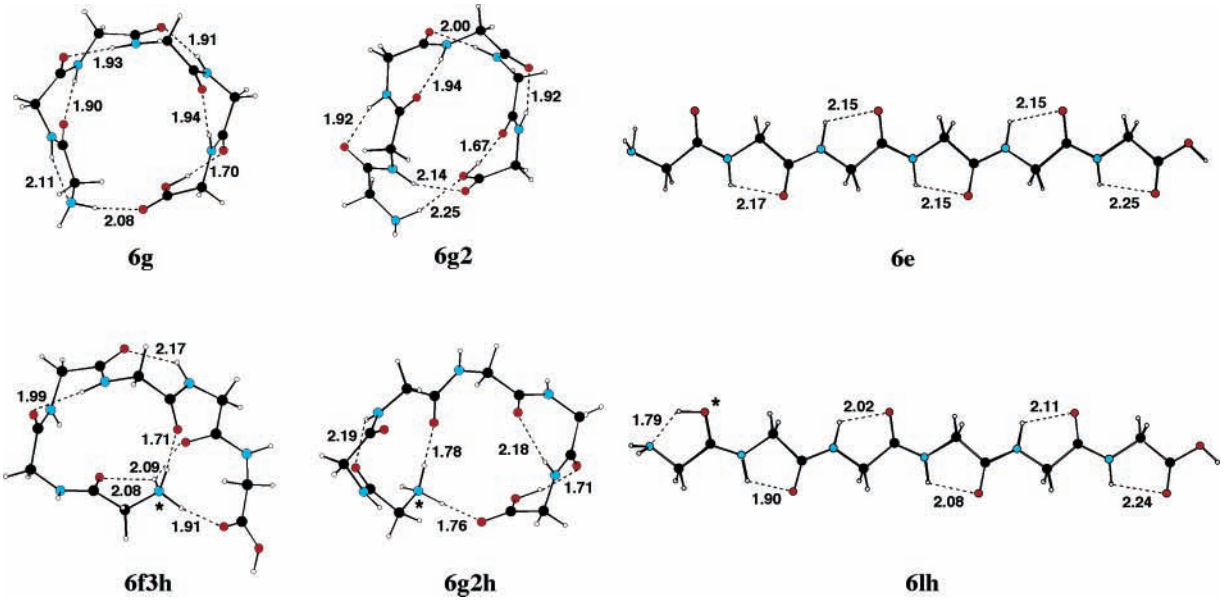

Figure 6. Hexaglycine structures **6g**−**6lh** optimized at the B3LYP/6-311++G** level.

rotation around ω_1 between the two minima is **1dT** (irC):

$$\textbf{1e}\ (0.00) \rightarrow \textbf{1dT}\ (11.64) \rightarrow \textbf{1d}\ (5.37) \qquad \text{(irC)}$$

showing a larger *G*-barrier (11−12 kcal/mol) on converting *cis* to *trans*. At the C-terminus, the O−H···O═C attraction in *cis*-COOH is named C_4(OH), which is taken as a H-bond despite its occurrence in a functional group. This terminal C_4(OH) appears in all the extended series (except *m*) and in the folded series *f* and *fh*. Conversion from *cis*- to *trans*-COOH releases the hydroxyl OH to form a H-bond with the adjacent amide O, OH···O, as the C_7(OH) in the *m*, *g*, and *gh* series. The C_7(OH) is stronger than C_4(OH) on account of a shorter *r*(H···O) in the former.

Neutral Series. The neutral series (*e*, *c*, *m*, *f*, and *g*) are simple to decipher with regard to correlating conformational stability with H-bonding. A casual inspection of the Gly_n main-chain conformations in Figures 2−6 reveals the frequent occurrence of the 5- and 7-membered NH···O bonds, C_5 and C_7. The C_5 bonds are prominent in the extended *ec* which evolve into "repeated C_5"[32] in the larger structures (e.g., **1e** → **6e**). As the C_7 bonds increase with "repeated ($C_{7,eq}$, $C_{7,ax}$)"[31] in the folded *fg*, the stability of *fg* increases relative to *ec*. This observation implies that C_7 is a stronger bond than C_5 and is confirmed in part by the shorter *r* in C_7 compared with that in C_5. When the folded structures begin to form the ring-closing C_{14}, C_{17}, and C_{20}, the ring-like structures over 4, 5, and 6 residues emerge, respectively. The single-ring *f* or *g* carries repeated C_7 plus one C_m with $m > 7$, while the multiple-ring *fs* or *gs* contains "*s*" such rings. Clearly the C_7 and C_{3n+2} bonds are responsible for the significantly greater stability of *fg* vs *ec*. The extended *m* contains "repeated $C_{7,eq}$" in an open-chain form;[32] its stability is intermediate between those of *ec* and *fg*. This can be explained in part by the stronger C_7 in *m* as compared with the C_5 in *ec*, and a lack of C_{3n+2} in *m* in contrast to the presence of such bonds in *fg*. Between the two-folded series, *g* is expected to be more stable than *f* owing to a stronger C_7(OH) than C_4(OH). The analysis is consistent with the relative stabilities of the four major series from tetra- to hexaglycines shown in Table 1: $g > f > m > e$. Among the extended series, the structures of *c* are φ_1-rotamers of *e*: the *c* series is introduced along with the *eT* series (not shown in Table 2) in the discussion of deprotonation mechanisms. The relative stability is $e > c$, indicating a more stable bifurcated $C_5(NH_2)$ in *e* than the single C_5(NH) in *c* at the N-terminus.

Factors other than H-bonding that influence conformational stability can be broadly described as structural strain to destabilize and electronic enhancement to stabilize. Steric repulsion occurs when nonbonded atoms are in close proximity, e.g., when φ_i or ψ_i approaches 0°. Electronic strain occurs when π-electron delocalization among the conjugated covalent bonds in the planar amide or carboxyl group is disrupted, i.e., when ω_i deviates from 180° or ω_n deviates from 0° or 180°. For the smaller peptides, minimizing structural strain becomes relevant to attain greater electronic stability. Examples include **1e** vs **1m** and the open-chain **2f** and **3f** vs **2m** and **3m** (Table 1).

One distinct electronic enhancement for peptides is the "π-bond cooperativity"[34] that induces electron transfer and polarization between favorably oriented adjacent NH···O and NH···O bonds. For example, the extended *e* exhibits noticeable stabilization as the H-bond chain elongates, evidenced by the decreasing *r* of C_5 as a result of linking the planar C_5 bonds. This is not the case for the extended *m* which has virtually constant *r* of C_7 from **3m** to **6m**. On the other hand, there is a significant decrease in all the H-bond lengths for the folded *g* from **4g** to **6g**, partly due to a contraction of the peptide ring size to form the ring-closing C_{14}, C_{17}, and C_{20}. Despite the increasing ring strain, a greater increase in the stability of *g* relative to that of *e* is seen as a result of a greater increase in H-bonding attraction from the decreasing *r*. The relative electronic stability, $fg > m > e$ for $n = 4-6$, is displayed vividly in Figure 8. Overall, the major driving force for conformational stability is H-bonding although other factors may become important at times.

Protonated Series. The basicity of different protonation sites in a peptide is expected to follow the trend (t1):

$$\text{amino N} > \text{amide O} > \text{amide N} > \text{carboxyl carbonyl O} \qquad \text{(t1)}$$

which is supported by the GB values of NIST in kcal/mol: 210 for the N-protonation of ethylamine, 189.1 for the O-protonation of formamide, 175 for the N-protonation of formamide, and 169.8 for the carbonyl O-protonation of formic acid.[22] Separate calculations at the B3LYP/6-311++G** level for the respective GBs yield 210.0, 189.4, 172.6, and 168.2 kcal/mol.[26c] These

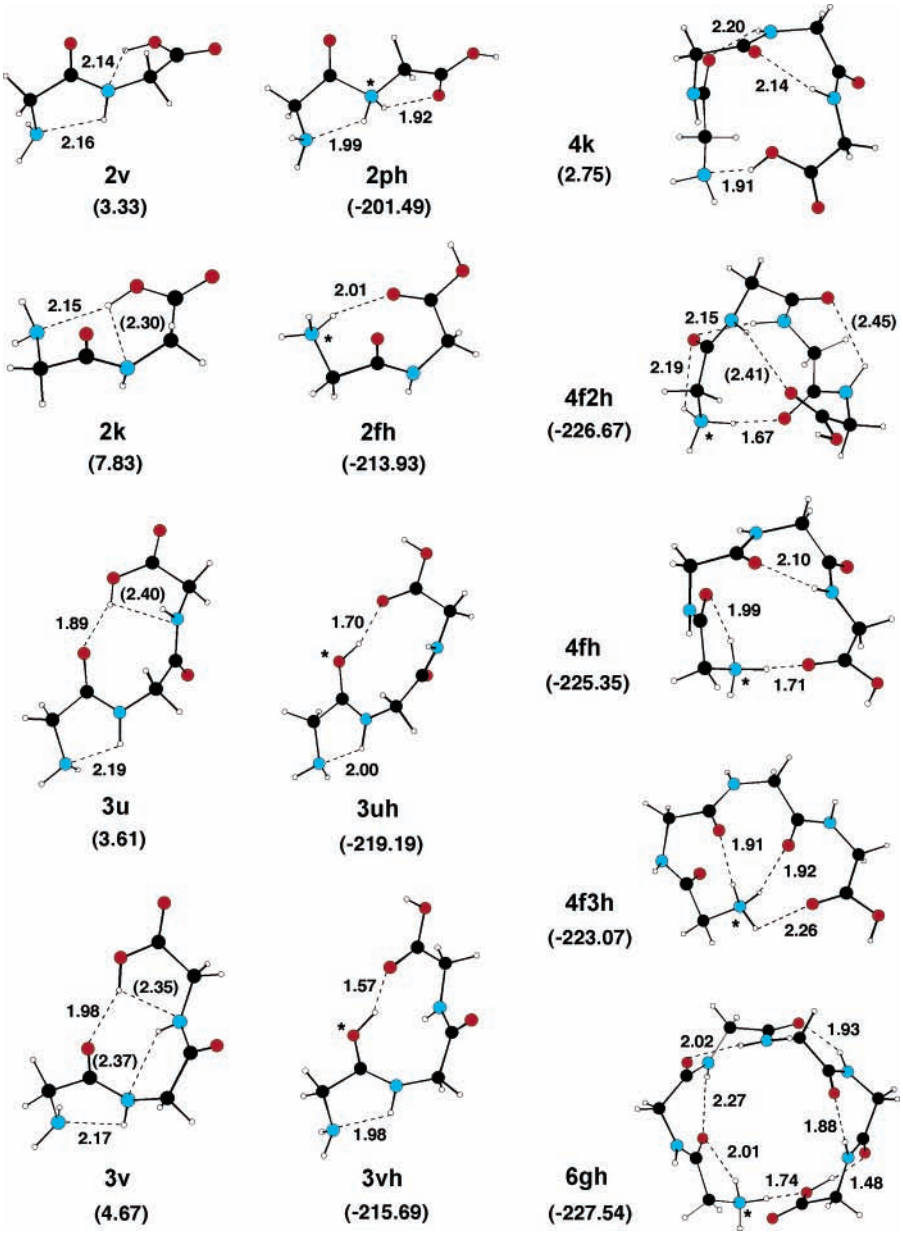

Figure 7. Additional structures of diglycine **2v**−**2fh**, triglycine **3u**−**3vh**, tetraglycine **4k**−**4f3h**, and hexaglycine **6gh** optimized at the B3LYP/6-311++G** level. The free energy value in kcal/mol is shown below the name of each structure (cf. Table 1): ΔG(A) for $n = 2$ and 3; ΔG(A/B) for $n = 4$ and 6.

values reflect the relative gain in the overall stability of each positive ion on forming the new X*−H covalent bond in the absence of H-bonding as shown in trend (t2):

N*−H at amino N > O*−H at amide O >
N*−H at amide N > O*−H at carboxyl carbonyl O (t2)

Note that the omission of carboxyl hydroxyl O in the two trends is due to the destruction of the C−O bond on forming the H_2O component in the resulting positive ion complex.[12a] The calculated GB for the hydroxyl O-protonation of formic acid is 151.3 kcal/mol,[26c] a value low enough to make the hydroxyl O an unlikely basic site for protonation.

Protonations of the structures in C_5- and C_7-based extended series *e* and *m* and the C_7-induced folded series *f* and *g* generate the protonated series *eh*, *lh*, *oh*, *mh*, *fh*, and *gh*. The most stable structures are found to result from protonations at N1 and O1: this is consistent with the expectations that the amino N and amide O atoms are the respective first- and second-most basic atoms which in turn produce the respective strongest and next strongest N*−H and O*−H covalent bonds (cf. t1 and t2). The major factor that determines which one of the two sites is more basic is likely to be the increase in H-bonding brought by protonation. For N1-protonation, the geometrical freedom of the terminal $-N^*H_3$ facilitates H-bond formation of varying bond length and strength of the NH···O type, designated as C_5*, C_8*, C_{11}*, C_{14}*, C_{17}*, and C_{20}* in *eh* and *fgh*. Note all N1-protonated species with folded structures (*fgh*) adopt *cis*-NCCO at the N-terminus which requires a conversion from *trans*-NCCO in the parent neutral species (*fg*) via internal rotations (cf. irN in reverse over a *G*-barrier of 4−5 kcal/mol).

In the case of O1-protonation, two fairly localized C_5*(O1*−H···N1) and C_7*(O1*−H···O2) interactions, designated as C_5*(O1) and C_7*(O1), are found to be prominent. The C_5*(O1) is formed with the neighboring N1; this bond is seen in *lh*. The C_7*(O1), formed with the neighboring O2, has exceptionally short *r* (1.42 Å in **3oh** and 1.38 Å in **4oh**) and consequently is

TABLE 2: Structural Series of Selected Polyglycine Energy Minima[a]

NCCO	OCOH	series	structures	H-bonds[b]
			extended forms	
cis	*cis*	*e*	**1e**−**6e**	$C_5(NH_2)$, $(n-1)C_5$, $C_4(OH)$
		c	**1c**−**6c**	$C_5(NH)$, $(n-1)C_5$, $C_4(OH)$
trans	*trans*	*m*	**2m**−**6m**	$C_5(N1)$, $(n-2)C_7$, $C_7(OH)$
cis	*cis*	*eh*	**1eh**−**6eh**	C_5*, $(n-1)C_5$, $C_4(OH)$
		lh	**2lh**−**6lh**	C_5*(O1), $(n-1)C_5$, $C_4(OH)$
trans	*cis*	*oh*	**3oh**−**6oh**	$C_5(N1)$, C_7*(O1), $(n-2)C_5$, $C_4(OH)$
trans	*cis*	*mh*	**2mh**−**6mh**	$C_5(N1)$, $(n-2)C_7$, C_7*[O$(n-1)$], $C_4(OH)$
			folded forms	
trans	*cis*	*f*	**2f**, **3f**, **4f**	$C_5(N1)$, $(n-2)C_7$, C_{3n+2}, $C_4(OH)$
			5f3	$C_5(N1)$, $2C_7$, C_{11}, C_{14}, C_{16}, $C_4(OH)$
trans	*trans*	*g*	**4g**, **5g**, **6g**	$C_5(N1)$, $(n-2)C_7$, C_{3n+2}, $C_7(OH)$
			6g2	$4C_7$, C_{17}, C_{20}, $C_7(OH)$
cis	*cis*	*fh*	**2fh**, **3fh**, **4fh**	C_5*, $(n-3)C_7$, C_{3n+2}*, $C_4(OH)$
			4f2h	C_5*, C_7, C_{11}*, (C7), (C11), $C_4(OH)$
			4f3h	C_8*, C_{11}*, C_{14}*, $C_4(OH)$
			5f2h	C_5*, C_7, C_{11}*, C_{17}*, $C_4(OH)$
			6f3h	C_5*, $2C_7$, C_{14}*, C_{17}*, C_{20}*, $C_4(OH)$
cis	*trans*	*gh*	**4gh**, **5gh**, **6gh**	C_5*, $(n-3)C_7$, C_{3n+2}*, $C_7(OH)$
			6g2h	$2C_7$, C_{11}*, C_{20}*, $C_7(OH)$

[a] The series are distinguished by the conformations of atom chains NCCO and OCOH at the termini and the number and types of H-bonds. [b] All H-bond lengths <2.3 Å except $C_5(NH_2)$, $C_5(NH)$, $C_4(OH)$, and the (C7) and (C11) in **4f2h**. For **2f** and **3f**, omit C_{3n+2}. For **2fh**, omit C_5* and $(n-3)C_7$.

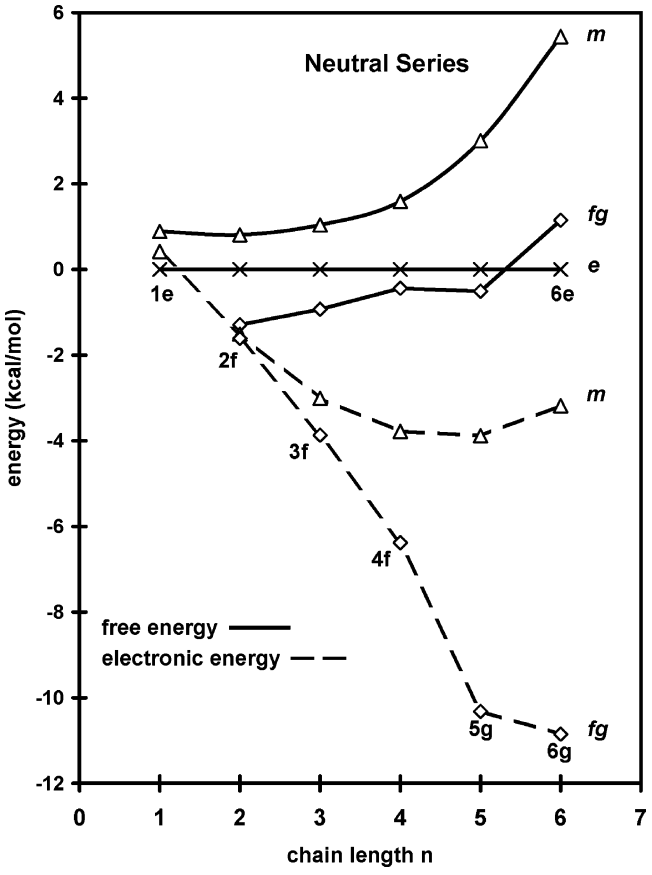

Figure 8. Relative electronic and free energies of polyglycines, Gly_n with $n = 1-6$, for the neutral structural series *e*, *m*, and *fg*. Names of selected structures are shown. Ranges of energy values in kcal/mol: electronic energy from −12 to 2; free energy from −2 to 6.

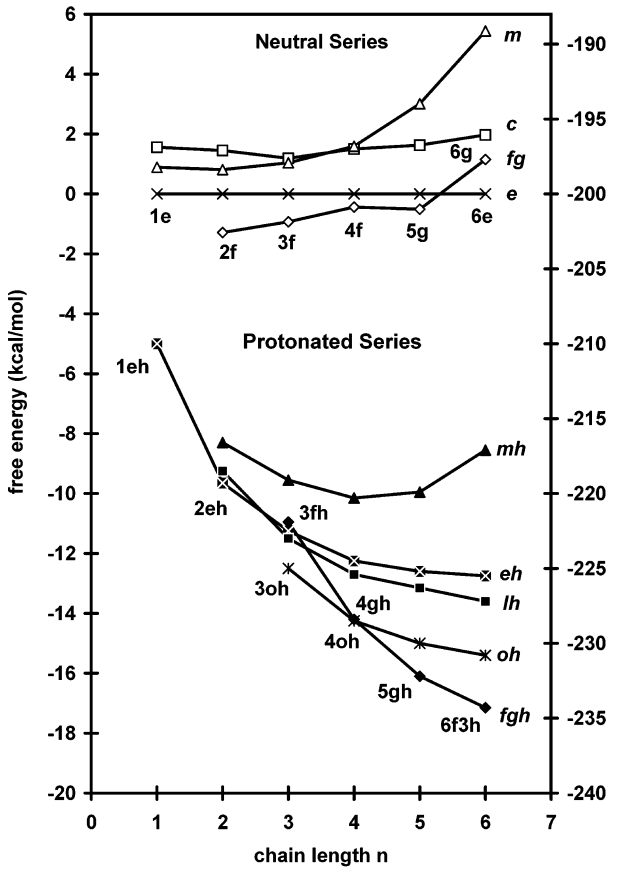

Figure 9. Relative free energies of polyglycines, Gly_n and Gly_nH^+ with $n = 1-6$, for the neutral structural series *e*, *c*, *m*, and *fg* (top) and protonated structural series *eh*, *lh*, *oh*, *mh*, and *fgh* (bottom). Names of selected structures are shown. Ranges of energy values in kcal/mol: neutral series, left axis, from −2 to 6; protonated series, right axis, from −240 to −205.

exceptionally strong. The formation of either H-bond further enhances molecular stability by allowing greater π-electron delocalization in the peptide linkage OCN involving O1,[7] exemplified by the respective C−O1 and C−N2 bond lengths of **3lh** (1.29 and 1.30 Å) vs **3eh** (1.23 and 1.32 Å). Despite the strong H-bond strength and the induced electronic enhancement, the formation of a new O*−H brings less stabilization to *lh* and *oh* than the new N*−H to *eh* and *fgh* (cf. t2). Moreover, the repeated C_5 in *oh* are weaker than the repeated C_7 in *fgh*. After balancing these different factors **3oh** turns out to be the only O1-protonated structure amidst the N1-protonated **1eh**, **2eh**,

4gh, **5gh**, and **6f3h** in the group of protonated structures with lowest-E_e.

The O2-protonated H-bond in **3mh**, C_7*(O2*−H···O3) or C_7*(O2), involves the carboxyl carbonyl O as the H-acceptor. The *r* of C_7*(O2) in **3mh** (1.56 Å) is longer than the *r* of C_7*(O1) in **3oh** which involves an amide O as the H-acceptor. On this basis it can be generalized that C_7*(O3)−C_7*(O5) in the respective **4mh**−**6mh** are all weaker H-bonds than the C_7*(O1) in **4oh**−**6oh**.

The relative stabilities of the protonated structures are affected by many factors, among which are (a) the stabilization brought by forming the new covalent bond (e.g., N*−H vs O*−H), (b) the strength of the protonated H-bond [e.g., C_7*(O1) vs C_{3n+2}*], (c) the increased conjugation in the peptide bond linkage from amide O-protonation [e.g., C_5*(O1) vs C_5*], (d) the strength of the main-chain H-bonds (e.g., C_7 vs C_5), and (e) the effect of "π-bond cooperativity" (e.g., *lh* vs *mh*).

Ring Series. Special mention is in order for the ring structures of the *fg* and *fgh* series which were not well-known in the past. Upon folding with repeated ($C_{7,eq}$, $C_{7,ax}$), the limits for single-ring formations are reached around $n = 6$: both **6g** and **6gh** exhibit noticeable ring strain. The multiple-ring structures begin to develop at $n = 4$, **4f2h** and **4f3h**, and rapidly become competitive in stability as shown by **5f3**, **5f2h**, **6g2**, **6f3h**, and **6g2h** (Figures 2−7 and Table 1). The most complex multiple structures are represented by **5f3** which has a most compact shape and **6f3h** which is exceptionally stable. The N1-protonated structures with all three H atoms of $-N^*H_3$ engaged in H-bonding are **4f3h**, **5f2h**, and **6f3h**. The moderately complex structures are **4f2h**, **6g2**, and **6g2h**. A complex multiple-ring structure usually has a soccer-ball shape made of rigid, intertwining rings of atoms and is distinctly different from a single-ring structure such as **4f**, **4g**, **5g**, **6g**, **2fh**, **3fh**, **4fh**, **4gh**, **5gh**, or **6gh**.

An oligoglycine of *n*-residues in an *s*-ring conformation is composed of *s* single-ring conformations sharing parts with one another. Brief analyses of three multiple-ring conformers are given next using H-bonds (Table 2) for illustration. (1) The **4f2h** ion, which contains C_5*, C_7, C_{11}*, (C7), and (C11), can be seen as portions of **3fT** [(C7), (C11)] on top and **3fh** (C_5*, C_{11}*) at the bottom with a gain of C_7. [The terminal C_4(OH) is omitted for convenience.] This 2-ring Gly_4H^+ is therefore a conglomerate of two single rings, Gly_3 and a Gly_3H^+, stacked together. Note the H-bonding pattern of the near parallel pair, (C_7, C_{11}) or (C_7, C_{11}*), is also seen in **5f3**, **5f2h**, and **6g2h**. (2) The **4f3h** ion with C_8*, C_{11}*, and C_{14}* appears to grow from **2fh** (C_8*) to **3fh** (C_{11}*) to **4fh** (C_{14}*) with a loss of C_5* and C_7: the composite ion is taken to be a 3-ring Gly_4H^+ from overlapping the single rings Gly_2H^+, Gly_3H^+, and Gly_4H^+. Due to the limited chain length of **4f3h** the carbonyl O2, O3, and O4 are made into the H-acceptors for $-N^*H_3$. In larger peptide ions with a fully H-bonded $-N^*H_3$, main-chain H-bonds can be added to increase the overall stability of the ion. Examples include the 2-ring **5f2h** gaining one C_7 on forming C_5*, C_{11}*, and C_{17}* with O1, O3, and O5 as acceptors and the 3-ring **6f3h** adding one C_5* and two C_7 on forming C_{14}*, C_{17}*, and C_{20}* with O4, O5, and O6. (3) The neutral **6g2** ($4C_7$, C_{17}, C_{20}) can be recognized as two superimposed rings of **5g** ($3C_7$, C_{17}) and **6g** ($4C_7$, C_{20}) with slight modifications on C_{17} (from N1−H···O5 to N2−H···O6) and C_{20} (from N1−H···O6 to N1−H···O) and a loss of C_5(N1). [The terminal C_7(OH) is omitted for convenience.] Finally, the H-bond lengths of multiple-ring peptides are highly variable as a result of geometrical constriction. In particular the ring-closing bond lengths (in Å) fluctuate dramatically: C_{14}* from 1.67 in **4f2h**, to 1.71 in **4fh**, and to 2.26 in **4f3h**; C_{17}* from 1.62 in **5gh** to 1.85 in **5f2h**; and C_{20}* at 1.91 in **6f3h** vs 1.76 in **6g2h**.

Series of Secondary Choice. Three additional groups of structures were investigated: the smaller members and associated H-bonds of interest are provided in Figure 7 as examples. (1) Neutral conformers that contain motifs of type II β-turn and 3_{10} helix:[31,32] the precursors **3u** and **3v**, respectively, with C_{10}(OH) for C_{10}(O−H···O1), C_5(O−H···N3), and C_5(N3−H···N2). In larger peptides the OH···O and OH···N interactions are replaced by the NH···O and NH···N interactions. (2) The O1-protonated species containing C_m*(O1) with $m > 7$: **3uh** and **3vh** with C_{10}*(O1) for C_{10}*(O1*−H···O). (3) The neutral species containing C_m(O−H···N1): **1m** previously with $m = 5$; **2k** with $m = 8$; and **4k** with $m = 14$. (4) The amide N-protonated species: **2ph** with C_5*(N2*−H···O2). In this case the original π-conjugation in the C−N2 peptide bond is destroyed. The new structures in Figure 7 are found less stable in free energy than the top two best of those in Table 1 and Figures 2−6. Nonetheless, the additional information adds breadth and depth to the overall discussion.

Hydrogen Bonding

While conformational analysis can be carried out without knowing precisely the extent of H-bonding contribution, it is worthwhile to attempt a direct evaluation of the individual H-bond strength based on the physical attributes of the atoms and geometry involved. An independent knowledge of the relative strength of different H-bonds will help understand analytically their collective influence on the electronic and free energies.

Several indicators of H-bond strength emerge from crystallographic data of biological structures[34] and ab initio calculations on dimers of small hydrides:[35] bond electron population *p*, bond length *r*, and bond energy *B*, all between the two H and Y atoms in X−H···Y. The following guidelines on H-bonds are found useful here:[34,35] *p*(H···Y) in *e*, 0.01−0.03 typically and >0.10 for a strong bond; *r*(H···Y) in Å, 3.0−1.5 for a weak-to-normal bond and 1.5−1.2 for a strong bond; and *B*(H···Y) in kcal/mol, <5 for a weak bond and >10 for a strong bond. In this analysis *p* is the Mulliken overlap population[8] and $B = -E$ where *E* is approximated as an electrostatic attraction between the CHELPG charges q_H and q_B separated by *r*. The *p*, q_H, q_B, and *r* data[29] of 20 bond energy terms selected from structures with simple bonding patterns are presented in Table 3. The derivation of *E* follows Coulomb's law as shown below.

The CHELPG charges are atomic partial charges derived from molecular electrostatic potentials.[30] The molecular dipole moment (in debye) calculated from CHELPG charges replicates closely the ab initio value calculated from the molecular electronic distribution and atomic nuclear charges, e.g., 12.12 (CHELPG) vs 12.33 (ab initio) for the hexaglycine structure in Figure 1. Considering that a large part of H-bonding is electrostatic, the CHELPG charges are employed to estimate the energy by means of the equation, $E = c \cdot q_H \cdot q_B / r$, where *c* is a calibration factor. Two glycine conformers are used to fix *c*: **1e** which has a O−H···O═C interaction in *cis*-COOH, C_4(OH) at *r*(H···O) = 2.30 Å; and **1d** which resembles **1e** in all respects except for the *trans*-COOH at *r*(H···O) = 3.01 Å, without an apparent C_4(OH). In this model the *E* of C_4(OH) in *trans*-COOH is assumed zero. As **1e** is 5.61 kcal/mol more stable than **1d** in E_e, the *E* of C_4(OH) in **1e** is assigned the value −5.61 kcal/mol. Using this *E* value and the *q* and *r* data for C_4(OH) from Table 3, $c = 0.00455$ kcal·Å/(mol·10^{-4}·e^2) is obtained for

TABLE 3: Hydrogen Bonding (X−H···Y) in Selected Polyglycine Structures at the B3LYP/6-311++G Level[a,b]**

type	structure	X−H···Y	p(H···Y)	r(H···Y)	q_H	q_Y	B(H···Y)
			neutral				
C_4(OH)	**1e**	O−H···O1	0	2.30	48	−59	5.6
C_5(N1)	**2f**	N2−H···N1	−3	2.17	25	−89	4.7
C_5	**3e**	N2−H···O2	6	2.19	25	−55	2.9
C_7	**3f**	N3−H···O1	11	2.02	42	−64	6.1
C_7(OH)	**4g**	O−H···O3	13	1.78	41	−58	6.1
C_{10}(OH)	**3v**	O−H···O1	19	1.89	46	−68	7.5
C_{14}(OH)	**4k**	O−H···N1	−7	1.91	30	−59	4.2
C_{14}	**4g**	N1−H···O4	6	2.24	42	−60	5.1
C_{17}	**5g**	N1−H···O5	8	2.13	43	−58	5.3
C_{20}	**6g**	N1−H···O6	5	2.08	45	−61	6.0
			protonated				
C_5*	**3eh**	N1*−H···O1	15	1.68	30	−53	4.3
C_8*	**2fh**	N1*−H···O2	10	2.01	31	−61	4.3
C_5*(O1)	**3lh**	O1*−H···N1	12	1.78	36	−85	7.8
C_7*(O1)	**3oh**	O1*−H···O2	19	1.42	54	−59	10.2
C_{10}*(O1)	**3vh**	O1*−H···O3	4	1.70	52	−63	8.8
C_7*(O2)	**3mh**	O2*−H···O3	15	1.56	53	−58	9.0
C_{11}*	**3fh**	N1*−H···O3	11	1.78	22	−63	3.5
C_{14}*	**4gh**	N1*−H···O4	26	1.65	28	−60	4.6
C_{17}*	**5gh**	N1*−H···O5	22	1.62	32	−62	5.6
C_{20}*	**6gh**	N1*−H···O	5	1.74	27	−67	4.7

[a] Units: electron population p and atomic charges q in 10^{-2} e, length r in Å, and energy B in kcal/mol. [b] C_5(N2) for C_5(N3−H···N2) in **3v**: $p = 1$, $r = 2.37$, $q_H = 21$, $q_{N2} = -26$, and $B = 1.0$.

calculating the B values in Table 3. Note the p(H···O) of C_4(OH) in **1e** is zero, a correct condition for the electrostatic model. But in most C_m and C_m* of polyglycines the p values are positive, showing some degree of covalency. In such cases the B values are underestimated.

The calculated indicator values in Table 3 fall in the following ranges: −0.07 to 0.26 e for p; 2.3 to 1.4 Å for r; and 3 to 10 kcal/mol for B. The ranges reflect mostly weak to normal H-bonds based on the guidelines for r and B. But for the protonated H-bonds in Table 3, 80% have $p > 0.1$ e and only 10% show a $B > 10$ kcal/mol, which is not expected from the guideline for p. Obviously, the calculated B values for most C_m* are underestimated owing to a significant presence of covalency in H-bonding.

On the basis of B and taking into account p and r in certain groupings, the relative strengths of the neutral and protonated H-bonds are proposed as follows:

$$C_{10}(OH), C_7(OH), C_7 > C_{20}, C_{17}, C_{14}, C_4(OH) > C_5(N1), C_{14}(OH), C_5$$

$$C_7*(O1), C_7*(O2), C_{10}*(O1) > C_5*(O1), C_{17}*, C_{14}*, C_{20}* > C_5*, C_8*, C_{11}*$$

which are consistent with the deductions drawn from the preceding conformational analysis. In particular, the main-chain C_5 is shown to be weak and C_7 is estimated to be ∼3 kcal/mol more stable than C_5: this helps explain the greater stability of m and fg over e. The ring-closing bonds C_{14}, C_{17}, and C_{20} are reasonably large to help establish the significant difference in stability between fg and m. The strong C_7(OH) involving the terminal OH (>6 kcal/mol) effectively puts f second to g in stability. Finally, the C_7*(O1) is indisputably the strongest H-bond (>10 kcal/mol) responsible for making **3oh** the lowest ΔE_e and both **3oh** and **4oh** the lowest ΔG of the protonated tri- and tetraglycines. As regards the relative strengths of H-bond types, OH···O > NH···O > OH···N > NH···N for neutral bonding and O*H···O > O*H···N > N*H···O for protonated bonding are expected from the greater acidity of OH over NH and O*H over N*H. One important piece of data for NH···N is provided in footnote b of Table 3.

For the gas-phase peptides containing more than three residues the driving force to form stable conformers is intramolecular H-bonding. This is the operating principle upon which the low-energy conformers were constructed by the "z-matrix" approach. The initial strategy was to maximize the number of H-bonds by engaging every proton donor to a prospective acceptor with the use of appropriate CDAs. The best examples are **4g**, **5g**, and **6g** for which the initial CDAs were set up to connect all amide and terminal groups by H-bonds in a continuous pattern so that maximal electron polarization may be attained to gain maximal stabilization. The fact that **4g**, **5g**, and **6g** were calculated to be the most stable neutral conformers for $n = 4-6$ confirms the success of the "z-matrix" approach as well as the importance of H-bonding in locating global energy minima.

Thermodynamically, H-bonding decreases entropy, increases the thermal correction G_{tc}, and ultimately increases the free energy G. A glance at Table 1 finds that more or stronger H-bonds yield larger ΔG_{tc}, i.e., ΔG_{tc} is usually greater for multiple rings than single rings and greater for the C_7-based series than the C_5-based series. In Figure 8 a measure of ΔG_{tc} can be visually assessed by comparing the solid (ΔG) and the dashed (ΔE_e) line plots pertaining to the same series. The difference in the two plots, $\Delta G - \Delta E_e$, gives an indication of how the ΔG_{tc} of the series changes as the peptide increases in size. The most critical finding is the faster increase in the ΔG_{tc} of fg and m relative to e, consistent with the rapid increase of H-bonding energies in fg and m relative to e. The resulting thermodynamic stability (ΔG) indicates that m is the least stable, fg is most stable from $n = 2-5$, while e increases in stability slowly but surely relative to fg. In fact, at $n = 6$, **6e** has lower G than **6g**, showing how H-bonding makes **6g** most stable electronically but not thermodynamically. In Figure 9, variations of thermodynamic stability (ΔG) for both the neutral and protonated series are shown.

Structures of Lowest Electronic and Free Energies

The calculated data and deductions presented in the preceding discussion on conformational analysis and H-bonding have built a body of evidence to establish the leading neutral and protonated structures of Table 1 and Figures 2−6 to be those at or near the global electronic energy minima of the respective species. The order of stabilities can be verified by the number and relative strengths of H-bonds, structural strain and electronic enhancement, and other factors intrinsic to the conformers. In view of the fact that electronic stability does not directly translate into thermodynamic stability, rationalization has been sought and assurance is given to identify the structures with lower to lowest free energies (Figures 8 and 9). At this point the exhaustive search for the most stable conformers is ended.

For glycine, the most stable and abundant neutral and protonated structures are indisputably **1e** and **1eh**.[11−14,16] For di- and triglycines, sufficient data were presented previously on the relative basicities of different protonation sites.[7,19,20] Yet, several most important structures (**2n**, **2f**, **3s**, and **3oh**) seem to have been overlooked. In the case of Gly_2, the φ_2-rotamers **2n** and **2f** have nearly the same stability but individually represent the conformers of lowest E_e and G, respectively. The φ_3-rotamer of **3f**, **3s** in Tables 2S and 3S, is the next lowest-G conformer among Gly_3. For the O1-protonated Gly_3H^+, Rodriquez et al. first noted the exceptionally short H-bond length of C_7*(O1) in their conformer "**4**" but predicted conformer "**2**" with C_5*(O1)

TABLE 4: Protonation and Deprotonation Pathways Using Structures of Glycine, Diglycine, and Triglycine as Examples[a]

protonations

path	neutral	site	protonated
p1	**1e**	N1	[1ehT] → **1eh**
p1O	**1m**	O1 → N1	**1eh**
p2	**2f**	N1	[2fhT] → **2eh**
p2O	**2n**	O1	**2nh** (0.0) → **2lhT** (7.3) → **2lh** (−5.0)
p3	**3f** → [3fT]	N1 → O1	**3fh** (0.0) → **3fhT** (4.2) → **3lh** (−1.1) → **3lhT** (11.1) → **3nh** (4.4) → **3ohT** (10.6) → **3oh** (−3.1)
p3t	**3e**	N1 → O1	**3eh** (0.0) → **3thT** (0.6) → **3lh** (−0.4)

KM dissociations

path	site	neutral	←	(dimer)	→	protonated
d2k	N1	**2f**		[2fhT]		**2eh**
d3Nk	N1	**3f**		[3fhT]		**3fh**
d3Ok	O1	**3f**		[3fohT]		**3oh**

RB deprotonations

path	protonated	site	neutral
d1b	**1eh**	N1	**1e**; **1c** (0.0) → **1eT** (0.5) → **1e** (−1.6)
d1Ob	**1eh**	O → O1	**1m**
d2Ob	**2lh** → [2lhT]	O1	[2nT] → **2n**
d3Nb	**3fh**	N1	[3fT] → **3f**
d3tb	**3lh**	O1→ N1	**3c** (0.0) → **3eT** (0.6) → **3e** (−1.2)

[a] See text for details. Relative Gibbs free energies in kcal/mol deduced from Table 2S are enclosed in parentheses.

as the conformer of lowest *G* on the basis of an energy profile for tautomerism (Scheme 1, in reference 25). Note that "**4**" differs from **3oh** in the −COOH orientation and "**2**" resembles **3lh**. In this work, **3oh** instead of **3lh** is shown to be the lowest-*G* conformer under the condition of thermal equilibrium.

As for the larger polyglycines, $n = 4-6$, the *fg* and *fgh* series are likely to be brand new while *e*, *eh*, and *lh* are simple extensions of known structures of lower *n*. The CDAs for the *e*, *m*, and *fg* structures are consistent with those obtained from model studies by Böhm[31] and Schäfer et al.[32] Some of the large cyclic bonds, C_m and C_m* with $m > 7$, appeared in the MMFFs conformations of the Gly_n/Gly_nH^+ pairs for $n = 3-5$ by Strittmatter and Williams (cf. Figure 1 of ref 23).

With the knowledge that all major conformational features and protonation sites have been examined and on the basis of the calculated data of ΔE_e and ΔG, the following statements are made. The leading neutral structures **1e**, **2n**, **3f**, **4g**, **5g**, and **6g** and protonated structures **1eh**, **2eh**, **3oh**, **4gh**, **5gh**, and **6f3h** are deduced as the global minima. Most of the same structures also have the lowest free energy; the exceptions are to be replaced by **2f**, **4f**, **6e**, **4oh**, and **5f2h** (cf. Table 1). But structurally, **6g** is more compatible with **6f3h** than **6e** and **5gh** is more compatible with **5g** than **5f2h**. In the interest of protonation studies both **6g** and **5gh** are retained. The resulting neutral/protonated pairs **1e**/**1eh**, **2f**/**2eh**, **3f**/**3oh**, **4f**/**4oh**, **5g**/**5gh**, and **6g**/**6f3h** are taken to be the best representative pairs (cf. Figure 9). Atomic Cartesian coordinates for the six pairs are listed in Table 5S.

Protonation and Deprotonation Pathways

By definition the GB and PA of a peptide M are the ΔG_r and ΔH_r of the protonation reaction r1

$$M + H^+ \rightarrow MH^+ \qquad \text{(r1)}$$

The mass spectral methods for measuring these quantities concern primarily the structure of a M or MH^+ species in a dynamic process of protonation or deprotonation. The KM method measures the rates of two competing dissociation reactions (r2) from a proton-bound dimer between the unknown M and a known base B

$$M + BH^+ \leftarrow (MHB^+) \rightarrow MH^+ + B \qquad \text{(r2)}$$

followed by an application of the absolute rate theory to determine the ΔH_r and ΔG_r of r1.[5] The RB method brackets the GB of the unknown M between those of the known bases B in deprotonation reactions (r3) using MH^+ as the reactant:

$$MH^+ + B \rightarrow M + BH^+ \qquad \text{(r3)}$$

where B represents each of the two bases with GBs above and below the GB of M.[7]

To correlate theoretical calculations with experimental processes, reaction paths (r1′−r3′) are constructed for some specific pairs of M/MH^+ in the presence of B and BH^+ in Table 4. The protonation path r1′ is portrayed by the structural changes of the neutral structure **i**, M(i), before and after reacting with BH^+ to produce the protonated structure **j**, MH^+(j):

$$M(i) + BH^+ \rightarrow (\text{intermediates}) \rightarrow MH^+(j) + B \qquad \text{(r1′)}$$

The "intermediates" are weakly bound or nonbonded clusters such as $[M(\kappa)\cdot H^+B]$ and $[MH^+(\lambda)\cdot B]$ in which $M(\kappa)$ and $MH^+(\lambda)$, or simply $[\kappa]$ and $[\lambda]$, represent conformations resembling those of some nearby local minimum or TS structures κ and λ. The path may be separated into neutral and protonated regions in the abbreviated expression: **i** → $[\kappa]$ ⋯ $[\lambda]$ → **j**, where B and BH^+ are omitted for clarity. Analogously, the KM dissociation paths r2′ are abbreviated as **i** ← $[\gamma]$ → **j** where $[\gamma]$, the peptide portion of the dimer ion, tends to be $[\lambda]$ to the right and $[\kappa]$ to the left. Structural compatibility between κ and λ, and hence **i** and **j**, appears to be a prerequisite. For convenience the protonated conformation $[\lambda]$ is adopted. The RB deprotonation route r3′ is depicted as: **j** → $[\mu]$ ⋯ $[\nu]$ → **k**, involving structures with sterically accessible protonation and deprotonation sites. The present goal is to identify κ, λ, γ, μ, and ν as road signs for the three types of proposed routes.

Several reaction paths using structures of glycine, diglycine, and triglycine are given in Table 4 as examples. The primary protonation routes involving the most abundant structures in

N1 and N1 → O1 protonations are *p1* (**1e**/**1eh**), *p2* (**2f**/**2eh**), and *p3* (**3f**/**3oh**): these are precursors to three KM dissociation paths *d2k*, *d3Nk*, and *d3Ok* and two RB deprotonation paths *d1b* and *d3Nb*. The secondary routes concerning O1 → N1, O1, and N1 → O1 protonations among structures of sufficient abundance are *p1O* (**1m**/**1eh**), *p2O* (**2n**/**2lh**), and *p3t* (**3e**/**3lh**): these precede the respective RB paths *d1Ob*, *d2Ob*, and *d3tb*. All these are conceptual pathways to interpret mass spectral results. A brief guide to the proposed paths is given below using Figures 2−7 and Table 3S as visual aids and Tables 1 and 2S for free energy references.

Glycine. The protonation path *p1* is straightforward. Adding H^+ to the lone pair of N1 in **1e** yields [1ehT] which relaxes to **1eh** with the formation of C_5*. The deprotonation of **1eh**, path *d1b*, involves the N1 of $-N^*H_3$. Removing one of the two H atoms not H-bonded to O1 leads to **1eT** or **1eT′** which relaxes to **1e**. If the H-bonded H atom is removed, **1c** or **1c′** results. The low-barrier **1eT** in the path between **1c** and **1e** should allow **1e** to reach thermal equilibrium. Note that **1c′** and **1eT′**, the mirror images of **1c** and **1eT**, undertake a symmetry-equivalent path. Thus, all pathways of *d1b* lead to **1e**/**1eh**, the RB pair for $n = 1$. (Introduction of **1c′** and **1eT′** serves as a reminder to the existence of all symmetry related structures in polyglycines.)

The O1-protonation of glycine prefers **1m**, *p1O*, with H^+ approaching the carbonyl O of *trans*-COOH in the direction of *cis*-$H^+\cdots O{=}C{-}O{-}$. The bonding interaction with H^+ triggers a "spontaneous" H migration from the hydroxyl O to N1 along the H-bonding path C_5(O−H···N1) to form **1eh**.[7,16] Schematically, the mechanism may be expressed in terms of the six atoms involved directly in the migration:

$$H^+ + O{=}C{-}O{-}H\cdots N \rightarrow (H{-}O\underline{\cdots}C\underline{\cdots}O{-}H\cdots N)^+ \rightarrow (H{-}O{-}C{=}O\cdots H{-}N)^+$$

The deprotonation, *d1Ob*, reverses the mechanism above by removing the acidic H from the *cis*-COOH of **1eh**; this triggers a "spontaneous" H migration from N1* to O1 along the H-bonding path C_5*(N1*−H···O1) to form **1m**. The word "spontaneous" refers to an intramolecular H migration with no apparent barrier. The RB pair for *d1Ob* is **1m**/**1eh**.

Among the selected peptides nine pairs have the structural properties for spontaneous H migrations: **1m**/**1eh** and **2k**/**2fh** using the O···H···N1 path, **3u**/**3uh** and **3v**/**3vh** taking the O···H···O1 path, and **nm**/**nmh** ($n = 2-6$) adopting the O···H···O ($n - 1$) path. All these pairs are responsive to RB measurements due to the spatial accessibility of −COOH for protonation and deprotonation. As for measuring GB, only **1m**/**1eh**, **2m**/**2mh**, and **3m**/**3mh** are viable because the initial structures in the preceding protonation steps (**1m** at O1, **2f** at O1, and **3s** at O2) are present in sufficient abundance to ensure measurable quantities of the protonated structures in the deprotonation steps (**1eh**, **2mh**, and **3mh**, all at the hydroxyl O).

Diglycine. The N1-protonation of **2f**, *p2*, is physically demanding: first the $-NH_2$ in **2f** moves halfway up as in **2nT** to make room for the H^+ transfer to N1; next the resulting $-N^*H_3$ rises further to [2fhT] which has one H atom forming C_5*(N1*−H···O1) and another poised for "C_8*", a weak N1*−H···O2 interaction with −COOH; and finally "C_8*" breaks up and stretches into **2eh** by strengthening the existing C_5* and forming a new C_5(N2−H···O2). Note the initial breakup of C_5(N2−H···N1) in **2f** is made easier by a lack of covalency in H-bonding ($p = -0.03$ *e* in Table 3).

In the following discussion the KM dimer ion, MHB^+ in (r2), is presented in two structural forms: the "simple" form that has only one contact between M and B, the M···H···B bridge, and an "ideal" form that has the bridge and additional H-bonding between M and B. An ideal form is proposed first because it is designed to yield more stable dissociation products. For example, the ideal Gly_2HB^+ in *d2k* may be visualized as [2fhT]···B with $N^‡-H^‡$ from B intercepting the "C_8*" in [2fhT] to form a 10-membered ring closed by N1···H···$N^‡-H^‡$···O2. Obviously, the role of $N^‡-H^‡$···O2 is to stabilize the cluster and to break apart when the dimer ion dissociates. Dissociations take place by a cleavage of N1···H to form **2f** and separately by a cleavage of H···$N^‡$ to yield **2eh**. The KM pair for $n = 2$ is **2f**/**2eh**.

Next, consider the simple Gly_2HB^+ which has N1···H···$N^‡$ only. The absence of $N^‡-H^‡$···O2 facilitates a φ_2-rotation to form the C-terminus C_5 in **2eh** prior to dissociations and a stay at **2e** after severing N1···H. Consequently, the conversion of **2e** to the more stable **2f** is skipped on forming the **2e**/**2eh** pair.

The O1-protonation of **2n**, *p2O*, sees the H^+ transfer from the "left" to produce **2nh**, followed by crossing the barrier **2lhT** to reach the more stable **2lh**. The protonated structures **2nh** and **2lhT** are conformationally similar to the neutral structures **2n** and **2nT**, except for the extra H atom bonded to O1 in the conjugated segment H−O1*···C···N2−H. Along the path the principal rotational changes in ψ_1 take the $-NH_2$ in **2nh** ($\psi_1 = 0°$) to halfway up in **2lhT** ($\psi_1 = -80°$) and all way up in **2lh** ($\psi_1 = 180°$), parallel to the changes of **2n** to **2nT** to **2e**. As for the deprotonation of **2lh** in *d2Ob*, the $-NH_2$ goes halfway down to [2lhT] to make room for the capture of the H atom bonded to O1. The departure of H^+ from the O1 in [2lhT] leaves behind [2nT] which directs the $-NH_2$ all way down to **2n**. The RB pair for $n = 2$ is **2n**/**2lh**.

Triglycine. The search for reaction paths are more difficult because of the increased number of geometrical variables. As a result, the proposed paths are less precise. The N1-protonation of **3f**, *p3*, is postulated to have the $-NH_2$ moving halfway up to [3fT] to make room for the H^+ transfer to N1 to form **3fh**. The deprotonation of **3fh**, *d3Nb*, removes the only non H-bonded H atom in $-N^*H_3$ to become [3fT] and continues with a ψ_1-rotation to take the resulting $-NH_2$ all way down to form **3f**.

With the knowledge that the N1-protonated **3fh** is less stable than the O1-protonated **3oh**, *p3* is extended from **3fh** to **3oh**, parallel to the Scheme 1 of Rodriquez et al.[25] (vide supra):

1 (0.0) → **TS(1→2)** (4.3) → **2** (−1.3) →
TS(2→3) (11.1) → **3** (4.3) → **TS(3→4)** (11.0) → **4** (−1.2)

where ΔG in kcal/mol are enclosed in parentheses. Despite a difference in the basis sets (6-31++G** vs 6-311++G**) and minor differences in the conformations of some members, the two sets of calculations are in good agreement. Assuming the barrier at **3fhT**, 4 kcal/mol, is high enough to obstruct the conversion of **3fh** to **3lh** before deprotonation in *d3Nb*, the RB pair for $n = 3$ is assigned **3f**/**3fh**.

Like the **2fhT** in diglycine, the conformation of **3fhT** in *p3* has C_5*(N1*−H···O1) and a favorably oriented $-N^*H_3$ relative to −COOH for forming the C_{11}* in **3fh**. The KM dissociation paths of *d3Nk* are therefore analogous to those of *d2k*. The ideal dimer ion is [3fhT]···B which contains a 13-membered ring closed by N1···H···$N^‡-H^‡$···O3. Sequential cleavages of the left N1···H and right H···$N^‡$ in the N1···H···$N^‡$ bridge lead to **3f** and **3fh**, respectively. The KM path for the N1-protonation of **3f** is **3f**/**3fh**.

The O1-protonation of diglycine in the confirmed route, **2f** → [2mhT] → **2mh**, prompted the search of a direct route to

O1-protonation of triglycine, **3f** → [3fohT] → **3oh**. The conformations of **3f** and **2mhT** were used initially for the synthesis of the intermediate **3fohT**. The KM path *d3Ok* is introduced with a dimer ion [3fohT]···B in the ideal form of a 12-membered ring closed by O1···H···N‡−H‡···O3. Immediately following the cleavage of O1···H, a ψ_2-rotation creates C_7(N3−H···O1) to form **3f**. After breaking H···N‡, the strong C_7*(O1*−H···O2) snaps in place while a φ_3-rotation lowers −COOH to form C_5(N3−H···O3) in **3oh**. The KM pair for $n = 3$ is assigned **3f**/**3oh**.

The tautomerism in *p3* also projects the formation of **3lh** from the N1 → O1 protonation of **3f**. Although **3lh** is 2 kcal/mol higher than **3oh** in *G*, it is insulated from **3oh** by two high barriers at **3lhT** and **3ohT**. The KM pair **3f**/**3lh** would call for a dimer ion conformationally closer to [3fhT]···B than [3fohT]···B (Figure 3) and would appear physically more accessible than **3f**/**3oh**.

Larger Polyglycines. Considering the rather complex routing for *p3*, similar investigations for $n = 4-6$ are not attempted. Meanwhile, *d3Ok* and *d3Nk* may be consulted in devising the KM paths for the **4f**/**4oh** and **5g**/**5gh** pairs that result from O1- and N1-prontonations, respectively. Clearly, any KM path for the **6g**/**6f3h** pair would seem highly speculative.

When the *e* series emerges in the same low-*G* region as *fg* in Figure 9, protonations of **4e**−**6e** need be addressed. The following pathways concerning **3e** were developed as models. The N1-protonation of **3e** in *p3t* produces **3eh** which converts to **3lh** after a H migration from N1 to O1 across a low barrier at **3thT** (Figure 3). The deprotonation of **3lh** in *d3tb* yields either **3c** or **3c′** which converts to **3e** via a low barrier **3eT** or **3eT′**. Similar conversions for **neh** and **nc** with $n = 4$, 5, and 6 were found with the respective *G*-barriers of 0.2, 0.3, and 0.2 at **nthT** and 0.4, 0.5, and 0.3 at **neT**, all in kcal/mol. These barriers are lower than the 0.6 at **3thT** and 0.6 at **3eT** for triglycine and thus promise faster conversions to the respective **nlh** and **ne**. Considering further that **nlh** is unlikely to convert to the more stable protonated structures on account of high barriers exemplified by *p3*, the most likely RB pairs for $n = 4-6$ would be **ne**/**nlh**. Henceforth, the RB paths for $n = 4-6$ are formally assigned as *dntb*, preceded by the protonation routes *pnt*.

Gas-Phase Basicities and Proton Affinities

Ab Initio Calculations. The protonation reaction of an ideal gas involving the selected pair of neutral (i) and protonated (j) structures is expressed as

$$\mathrm{Gly}_n(\mathrm{i}) + \mathrm{H}^+ \rightarrow \mathrm{Gly}_n\mathrm{H}^+(\mathrm{j})$$

The ΔG_r and ΔH_r of this reaction at 298.15 K and 1 atm in kcal/mol are

$$\Delta G_r = G(\mathrm{j}) - G(\mathrm{i}) + 6.28 \quad (1)$$

$$\Delta H_r = H(\mathrm{j}) - H(\mathrm{i}) - 1.48 \quad (2)$$

where the constants result from H^+ and PV work.[21] After correcting for BSSE, the calculated GB and PA associated with the i/j pair are:

$$\mathrm{GB} = -(\Delta G_r + \mathrm{BSSE}) \quad (3)$$

$$\mathrm{PA} = -(\Delta H_r + \mathrm{BSSE}) \quad (4)$$

The ΔG_r of eq 1 is calculated from the ΔG terms for i and j in Table 1, using ΔG(A) at level A for $n = 1-3$ but ΔG(A/B) at level A/B for $n = 4-6$. Analogously, the ΔH_r of eq 2 is calculated with $\Delta H(\mathrm{A}) = \Delta E_e(\mathrm{A}) + \Delta H_{tc}(\mathrm{A})$ and $\Delta H(\mathrm{A/B}) = \Delta E_e(\mathrm{A}) + \Delta H_{tc}(\mathrm{B})$ using the ΔE_e(A), ΔH_{tc}(A), and ΔH_{tc}(B) data from Table 2S. For $n = 4-6$, the errors from substituting ΔG_{tc} and ΔH_{tc} of level A by the less accurate level B for the i/j pair are likely to be smaller than 0.2 kcal/mol. (See footnote b of Table 2S.) The BSSE values in kcal/mol are generally accurate to better than 0.001 for $n = 1-3$ and 0.01 for $n = 4-6$.

In the previous study on glycine the conformational equilibrium effect (CEE) is included by calculating the *G* and *H* terms of i and j in eqs 1 and 2 as weighted averages of contributions from all low-*G* conformers based on Boltzmann distributions.[21] The present approach is to choose a structurally compatible i/j pair present in the highest population for the calculations. In this study the equilibrium populations for glycine ($n = 1$) at 298 K are 64% **1e**, 14% **1m**, 13% **1b**, and 9% total for **1c** and **1c′** for Gly, and for $GlyH^+$ 99.9% **1eh** and 0.1% **1bh**. Incorporating these populations, the GB of glycine including CEE is: $\langle GB \rangle$ = GB(**1e**/**1eh**) + 0.38 kcal/mol. As the polyglycine increases in size ($n = 2 \rightarrow 6$), the contribution from the protonated conformers to CEE becomes increasingly significant. A more effective cancellation of contributions from the neutral and protonated conformers is expected to result in a minimal overall CEE correction to GB.[21]

The GB and PA values calculated for the six best representative pairs are entered as the first entries in the two "calcd" columns of Table 5 for each Gly_n. These values may be taken as the "best" values corresponding to the "best" pairs. The protonation sites are N1 for glycine **1e**/**1eh** and diglycine **2f**/**2eh**, O1 for triglycine **3f**/**3oh** and tetraglycine **4f**/**4oh**, and N1 for pentaglycine **5g**/**5gh** and hexaglycine **6g**/**6f3h**. Note that N1-protonation is preferred; the greater stability of the O1-protonated **3oh** and **4oh** over the folded **3fh** and **4gh** is mainly due to the exceptionally strong H-bond, C_7*(O1*−H···O2). Taking into account the omission of CEE and the use of level B for G_{tc} and H_{tc} in the case of $n = 4-6$, the error estimates for the best values are within ±0.5 kcal/mol for $n = 1-3$ and ±1.0 kcal/mol for $n = 4-6$.

The next 10 "backup" pairs in Table 5 should have sufficient populations to be accountable since at least one member of each pair has the lowest or near-lowest *G*. Five additional pairs that yield comparable GB values but consist of less stable structures are included in footnote b.

Experimental Results. The experimental values in the "KM" and "RB" columns of Table 5 were adjusted to the current NIST basicity scale[22] from the originally reported values.[5,7,19] The linear regression procedures employed previously for KM[5,23] were followed here. Among the 32 PA and GB values for the reference amines, all have four significant figures except the GBs of ethylamine and triethylamine. Deleting triethylamine from the data set for the GB of pentaglycine improves the correlation coefficient from 0.984 for 270 kcal/mol to 0.9996 for 226.6 kcal/mol (vide infra).

Experimental uncertainties in kcal/mol were reported within ±0.8 for both the KM[5a] and RB[7] measurements on polyglycines and estimated to be no better than ±1 for each reference base in prior measurements.[4] Conservative estimates for the total uncertainties would therefore fall within ±2 for the KM and ±3 for the RB values.

The measured GB and PA values for Gly_n are dependent on experimental designs. To interpret an experimental outcome, pathways become relevant with regard to identifying the dominant i/j pair being measured. The primary protonation

TABLE 5: Calculated Proton Affinities and Gas-Phase Basicities for Polyglycines: Comparisons with Experiments[a,b]

structures		protonation			proton affinities		gas-phase basicities		
i	j	site	pathways	BSSE	calcd	KM[c]	calcd	KM[c]	RB[d]
				glycine					
1e	**1eh**	N1	*p1*, *d1b*	0.6	210.9	−	203.1	−	202.5
1m	**1eh**	O1	*p1O*, *d1Ob*	0.2	211.9		204.4		
				diglycine					
2f	**2eh**	N1	*p2*, *d2k*	0.4	220.0	220.2	211.4	212	
2e	**2eh**	N1	(cf. *d2k*)	0.6	221.3		212.5		
2n	**2lh**	O1	*p2O*, *d2Ob*	0.2	219.6		210.7		209.7
				triglycine					
3f	**3oh**	O1	*p3*, *d3Ok*	0.3	225.9	224.6	217.4	216.8	
3f	**3lh**	O1	(cf. *d3Nk*, *d3Ok*)	0.3	223.2		215.5		
3f	**3fh**	N1	*d3Nk*, *d3Nb*	0.4	223.4		214.3		213.6
				tetraglycine					
4f	**4oh**	O1	(cf. *d3Ok*)	−0.3	228.3	229.7	222.1	221.7	
4g	**4gh**	N1	(cf. *d3Nk*)	0.3	231.2		223.0		
4e	**4lh**	O1	*p4t*, *d4tb*	0.0	230.9		219.1		219.4
				pentaglycine					
5g	**5gh**	N1	(cf. *d3Nk*)	0.3	233.8	234.8	225.1	227	
5g	**5f2h**	N1	−	0.3	232.2		225.2		
5e	**5lh**	O1	*p5t*, *d5tb*	0.1	230.7		219.9		219.6
				hexaglycine					
6g	**6f3h**	N1	−	−0.6	237.2	238.2	229.8	231	
6e	**6lh**	O1	*p6t*, *d6tb*	−0.2	231.8		221.1		222.8

[a] All values in kcal/mol. See Table 4. [b] The "calcd" values are based on eqs 3 and 4. Structural pairs of lower populations including (BSSE, PA, GB): **2m**/**2mh** (0.2, 218.6, 210.9), **3e**/**3eh** (0.6, 225.3, 215.7), **3m**/**3mh** (0.0, 221.6, 213.8), **5f3**/**5f2h** (sp 1.0, 232.5, 226.3), and **6g2**/**6g2h** (sp −0.5, 235.9, 229.4). [c] Experimental values by the kinetic method, ref 5, adjusted from linear regression calculations using values of ref 22. [e] Experimental values by reaction bracketing, refs 7 and 19, adjusted by C. J. Cassady using values of ref 22.

routes *p1*−*p3* have been proposed as the most probable physical routes for checking the KM and RB values. To begin, each measured GB value is compared with the best value in Table 5. If there is a good numerical agreement and the proposed pathway is consistent with the measurement, the measured value is listed on the same line as the best value. If not, the pair taken to be the next highest in population with the appropriate pathway is examined for matching. Following this procedure the KM and RB values are entered into Table 5.

All the best KM pairs coincide with the best pairs for $n = 2-6$. The excellent agreement is not surprising in view of the fact that the unknown peptide usually contains enough functional groups to ensure the formation of a stable proton-bound dimer with the reference base. Naturally the most abundant dimer ions would contain the conformations related to the most abundant i or j of the best i/j pair at thermal equilibrium.

Contrary to KM, the best RB pairs are all different from the best pairs except glycine. Obviously RB targets different populations from the KMs when seeking sterically accessible basic sites in i or protonated sites in j.[4] In the RB pairs for $n = 1-3$, **1e**/**1eh**, **2n**/**2lh**, and **3f**/**3fh**, the amino N1 lone pair in **1e**, amide O1 lone pair in **2n**, and the non H-bonded N1*−H in the terminal $-N^*H_3$ of **1eh** and **3fh** are indeed geometrically and chemically favorable sites for protonation or deprotonation. The strongest support to the supposition of steric factor comes with the RB pairs for $n = 4$ and 5, **4e**/**4lh** and **5e**/**5lh**, resulting from the pathways *p4t*/*d4tb* and *p5t*/*d5tb*. The protonation routes choose the unencumbered N1 lone pairs of the extended **4e** and **5e**, instead of the H-bonded N1 lone pairs of the more populous **4f** and **5g**, for easier access to the N1 basic site. A peculiarity in the measured GBs in kcal/mol has been noted, i.e., the GB increase for $n = 4 \rightarrow 5$ is only 0.2, which is significantly smaller than the incremental increase of 5.8 for $n = 3 \rightarrow 4$. Yet, this peculiarity is confirmed by the calculated GB increase of 0.8 for $n = 4 \rightarrow 5$ vs 4.8 for $n = 3 \rightarrow 4$ using the proposed structures for the RB paths. The verification provides the best evidence of a successful theoretical modeling of an experimental process. Compared with KM, RB measures a smaller portion of the sample and the dominant structures in the measured portion can be quite different. The latter is confirmed by the significant differences in the GB values reported by the two methods: the RB values are lower than the KM values by 2 kcal/mol in diglycine to 8 kcal/mol in hexaglycine.

It has been well accepted that the measured basicities represent a number of conformations.[4,5a,7] In this study 21 structural pairs are shown as acceptable candidates for GB calculations (Table 5), among which 11 have already been assigned as the best KM and RB pairs. The present task is to scrutinize the relevance of the remaining 10 pairs which are expected to contribute less to the observed values due to lower populations. Extrapolating from the prior discussion on KM pathways, the KM pairs **2e**/**2eh** and **3f**/**3lh** appear to be energetically less favored than the best pairs but physically easier to accomplish. The **3e**/**3eh** and **4g**/**4gh** pairs have precedents among those already proposed, while "simple" dimer ions may be suggested to explain the pathways of **5g**/**5f2h**, **5f3**/**5f2h**, and **6g2**/**6g2h** that contain multiple-ring structures. The RB pairs **1m**/**1eh**, **2m**/**2mh**, and **3m**/**3mh** share the structural properties of "spontaneous H migration" exemplified by the *p1O*/*d1Ob* paths.

Comparisons of Results. The deviation of calculated value from the mass spectral value listed on the same line in Table 5, $\delta X = X(\text{calcd}) - X(\text{exptl})$ where $X =$ PA or GB in kcal/mol, gives a measure on how closely the proposed theoretical structures and pathways depict the experimental process. A close agreement imparts credibility to both experiment and theory. With regard to the proton affinities by KM, $-1.4 \le \delta\text{PA} \le 1.3$ is within the estimated experimental uncertainty of ±2. The results for gas-phase basicity fare better: $-1 < \delta\text{GB} < 1$ for KM and $-0.6 \le \delta\text{GB} \le 1.0$ for RB, with one exception each. The largest δGB for KM, 2 for pentaglycine, could reflect a failure to form the ideal dimer ion for the **5g**/**5gh** pair. It could

also be a problem with significant figures in the regression analysis (vide supra). Note a smaller deviation, $\delta GB = 1.5$, would result if the problematic data point were removed. The largest δGB for RB, 1.7 for hexaglycine, is likely an anomaly in view of the low reaction efficiency during the measurement and the computational difficulty in locating the energy minimum in a flat PES. In either case the largest deviation is within the estimated limits of ± 2 for KB and ± 3 for RB.

In general, a structural analysis of the KM measurement is difficult, owing to the ambiguity in the dimer ion and its dissociation products. The RB approach, on the other hand, is straightforward. As for the theoretical PA and GB, there is the inherent advantage of cancellation of errors when energy differences are calculated between neutral and protonated peptides of closely related structures (cf. eqs 1−4). The experimental designs of both KM and RB happen to favor the structural compatibility that helps improve the accuracy of theoretical results. This is one reason for the excellent GB values calculated at the B3LYP/6-311++G** level for glycine[21] and the model compounds (ethylamine, formamide, and formic acid)[26c] compared with the benchmark ab initio value[21] and the NIST values,[22] respectively.

Summary and Concluding Remarks

Ninety-three ab initio structures, including 20 transition states, are derived for neutral and singly protonated polyglycines with one to six residues at the B3LYP/6-311++G** and B3LYP/6-31+G** levels. Intramolecular H-bonds are shown to play a major role in conformational stability. Relative strengths of H-bonds are established from calculated values of interatomic distance, electron population, and electrostatic attraction between the H atom and its acceptor. Effects of H-bonding on the electronic and thermodynamic stability of the peptides are demonstrated. Structurally compatible neutral/protonated pairs are selected from those with lower to lowest Gibbs free energies for protonation calculations. Model protonation and deprotonation mechanisms are developed for the selected pairs.

Most of the mass spectral values of gas-phase basicity measured by the KM and RB methods are found to fall within ± 1 kcal/mol of the ab initio values calculated from theoretical structures proposed for the experimental processes. The good agreement confirms the premise that KM measures structurally compatible pairs in largest abundance, while RB measures structures with sterically accessible acidic and basic sites present in largest abundance. The agreement also gives credence to the identity and relative stability of the predicted structures.

For the protonation studies of glycine through hexaglycine, mass spectral data provide the critical clues for the theoretical search of the most stable neutral and protonated species. In return, quantum chemical theories offer accurate structures and energies for interpreting experimental measurements. There is clearly a dynamic synergy between theory and experiments.

Acknowledgment. I am indebted to Dr. Carolyn J. Cassady for several helpful correspondences and Mr. Scott E. Patterson for valuable assistance with the molecular figures. My deep appreciation goes to Dr. David B. Phillips for his continuing help and support on the many applications of personal computers to this work. Computing services on the DEC AXP and high-volume laser printing were provided by the Miami University Computing and Information Services. The financial support from the National Institute of General Medical Sciences (Grant: 2 R15-GM52670-02) and a substantial award of computing services on the Cray SV1 from the Ohio Supercomputer Center are gratefully acknowledged.

Supporting Information Available: Values of E_e, E_{ZP}, H_{tc}, G_{tc}, and G for the reference structures **1e**−**6e** in Table 1S (page S1); the corresponding energy values of 93 structures relative to those of the reference structures for $n = 1-6$ in Table 2S (pages S2−S4); conformational dihedral angles for the structures in Table 3S (pages S5−S7); the optimized *z*-matrixes of triglycine structures **3f**, **3fh**, and **3oh** in Table 4S (pages S8−S9); atomic Cartesian coordinates for the six best representative Gly_n/Gly_nH^+ pairs in Table 5S (pages S10−S16). This material is available free of charge via the Internet at hhtp://pubs.acs.org.

References and Notes

(1) Branden, C.; Tooze, J. *Introduction to Protein Structure*, 2nd ed.; Garland Publishing: New York, 1999.

(2) (a) Cassady, C. J. In *The Amide Linkage: Structural Significance in Chemistry, Biochemistry, and Material Science*; Greenberg, A., Breneman, C. M., Liebman, J. F., Eds.; Wiley-Interscience: London, 2000; pp 463−494. (b) Jarrold, M. F. *Annu. Rev. Phys. Chem.* **2000**, *51*, 179−207.

(3) (a) Biemann, K.; Martin, S. A. *Mass Spectrom. Rev.* **1987**, *6*, 1−76. (b) Biemann, K. In *Mass Spectrometry*; McCloskey, J. A., Ed.; *Methods in Enzymology*, Vol. 193; Academic Press: San Diego, 1990; pp 455−479. (c) Papayannopoulos, I. A. *Mass Spectrom. Rev.* **1995**, *14*, 49−73.

(4) Harrison, A. G. *Mass Spectrom. Rev.* **1997**, *16*, 201−217.

(5) (a) Wu, Z.; Fenselau, C. *J. Am. Soc. Mass Spectrom.* **1992b**, *3*, 863−866. (b) Wu, Z.; Fenselau, C. *Tetrahedron* **1993**, *49*, 9197−9206.

(6) Wu, Z.; Lebrilla, C. B. *J. Am. Chem. Soc.* **1993**, *115*, 3270−3275.

(7) Zhang, K.; Zimmerman, D. M.; Chung-Phillips, A.; Cassady, C. J. *J. Am. Chem. Soc.* **1993**, *115*, 10812−10822.

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

(9) Schäfer, L.; Newton, S. Q.; Jiang, X. In *Molecular Orbital Calculations for Biological Systems*; Sapse, A.-M., Ed.; Oxford University Press: New York, 1998; pp 181−224.

(10) (a) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789. (b) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100. (c) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200−206. (d) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(11) Jensen, J. H.; Gordon, M. S. *J. Am. Chem. Soc.* **1991**, *113*, 7917−7924.

(12) (a) Bouchonnet, S.; Hoppilliard, Y. *Org. Mass Spectrom.* **1992**, *27*, 71−76. (b) Jensen, F. *J. Am. Chem. Soc.* **1992**, *114*, 9533−9537.

(13) Császár, A. G. *J. Am. Chem. Soc.* **1992**, *114*, 9568−9575.

(14) Hu, C.-H.; Shen, M.; Schaefer, H. F., III. *J. Am. Chem. Soc.* **1993**, *115*, 2923−2929.

(15) Baron, V.; Adamo, C.; Leij, F. *J. Chem. Phys. A* **1995**, *102*, 364−370.

(16) Zhang, K.; Chung-Phillips, A. *J. Comput. Chem.* **1998**, *19*, 1862−1876.

(17) Zhang, K.; Chung-Phillips, A. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 382−395.

(18) Balta, B.; Basma, M.; Aviyente, V.; Zhu, C.; Lifshitz, C. *Int. J. Mass Spectrom.* **2000**, *201*, 69−85.

(19) Zhang, K.; Cassady, C. J.; Chung-Phillips, A. *J. Am. Chem. Soc.* **1994**, *116*, 11512−11521.

(20) Cassady, C. J.; Carr, S. R.; Zhang, K.; Chung-Phillips, A. *J. Org. Chem.* **1995**, *60*, 1704−1712.

(21) Zhang, K.; Chung-Phillips, A. *J. Phys. Chem. A* **1998**, *102*, 3625−3634.

(22) (a) Hunter, E. P.; Lias S. G. *J. Phys. Chem. Ref. Data* **1998**, *27*, 413−656. (b) Hunter, E. P.; Lias S. G. *Proton Affinities Evaluation in NIST Chemistry Webbook, NIST Standard Reference Database Number 69*; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaisthersburg, MD, March, 2003 (http://webbook.nist.gov).

(23) Strittmatter, E. F.; Williams, E. R. *Int. J. Mass Spectrom.* **1999**, *185/186/187*, 935−948.

(24) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 720−729.

(25) Rodriquez, C. F.; Cunje, A.; Shoeib, T.; Chu, I. K.; Hopkinson, A. C.; Siu, K. W. M. *J. Am. Chem. Soc.* **2001**, *123*, 3006−3012 and references therein.

(26) In addition to glycine (refs 16, 17, and 21), peptides (refs 7, 19, and 20), and carbohydrates (refs a and b below), the work is continuing on peptides (ref c below). (a) Jebber, K. A.; Zhang, K.; Cassady, C. J.; Chung-Phillips, A. *J. Am. Chem. So.* **1996**, *118*, 10515−10524. (b) Chung-Phillips,

A.; Chen, Y. Y. *J. Phys. Chem. A* **1999**, *103*, 953−964. (c) Chung-Phillips, A. (unpublished results).

(27) Del Bene, J. E.; Person, W. B.; Szczepaniak, K. *J. Phys. Chem.* **1995**, *99*, 10705−18707.

(28) Jákli, I.; Perczel, A.; Farkas, Ö; Császár, A. G.; Sosa, C.; Csizmadia, I. G. *J. Comput. Chem.* **2000**, *21*, 626−655.

(29) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, J.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A.9; Gaussian, Inc.: Pittsburgh, PA, 1998.

(30) (a) Chirlian, L. E.; Franci, M. M. *J. Comput. Chem.* **1987**, *8*, 894−905. (b) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.

(31) Bohm, H. J. *J. Am. Chem. Soc.* **1993**, *115*, 6152−6158.

(32) Schäfer, L.; Newton, S. Q.; Cao, M.; Peeters, A.; Alsenoy, C. V.; Wolinski, K.; Momany, F. A. *J. Am. Chem. Soc.* **1993**, *115*, 272−280.

(33) (a) Jensen, J. H.; Gordon, M. S. *J. Am. Chem. Soc.* **1995**, *117*, 8159−8170. (b) Rodriquez, C. F.; Cunje, A.; Shoeib, T.; Chu, I. K.; Ho pkinson, A. C.; Siu, K. W. M. *J. Phys. Chem. A* **2000**, *104*, 5023−5028. (c) Nielsen, P. A.; Norrby, P.-O.; Liljefors, T.; Rega, N.; Barone, V. *J. Am. Chem. Soc.* **2000**, *122*, 3151−3155 and references therein.

(34) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: New York, Berlin, Heidelberg, 1991.

(35) Scheiner, S. In *Reviews of Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1991; Vol. 2, pp 165−219.