ELSEVIER

International Journal of Mass Spectrometry 185/186/187 (1999) 935–948

International Journal of Mass Spectrometry

Computational approach to the proton affinities of Gly_n ($n = 1$–10)

Eric F. Strittmatter, Evan R. Williams*

Department of Chemistry, University of California, Berkeley, CA 94720, USA

Received 3 September 1998; accepted 6 November 1998

Abstract

The proton affinities of a series of polyglycines were calculated as a function of molecular size up to Gly_{10}. Molecular mechanics calculations using the Merck molecular mechanics force field were used to find lowest energy structures. These structures were used as starting geometries for both semiempirical and density functional calculations. Local density approximation density functional theory (DFT) (Slater exchange/VWN correlation or S-VWN, 6-31G*) was used to refine the geometries obtained from the mechanics. B3LYP (6-31G*) energies were calculated using these S-VWN geometries. The results of these calculations are compared to previously measured experimental data. The average deviation between the B3LYP and S-VWN proton affinities and the experimentally measured values of Fenselau and co-workers [J. Am. Soc. Mass Spectrom. 3 (1992) 863] are 4.0 and 2.0 kcal/mol, respectively. Better agreement to the experimentally measured values is obtained if the proton affinities are normalized to that of glycine. As expected, the DFT values are in better agreement than the semiempirical (AM1 and PM3) values. For the semiempirical methods, the average deviation from the proton affinities measured by Fenselau and co-workers (all data normalized to glycine) is ~4.5 kcal/mol. For proton affinities calculated with B3LYP hybrid functionals, this average deviation is only 1.2 kcal/mol (this deviation does not directly reflect the accuracy of the calculations since there are errors in both the experimental and calculated values). For pentaglycine, optimization was performed at the B3LYP 6-311G** level; the proton affinity differed by only 1 kcal/mol over that calculated at the 6-31G* level. This suggests that the lower basis set is sufficient for this application. The energies of the zwitterionic forms of Gly_n ($n = 4$, 5, 7, and 10) were compared to those of the simple protonated form. The zwitterion form of each polyglycine was found to be less stable at all levels of theory. These results suggest that it is possible to obtain accurate thermochemical data using mechanics and DFT calculations even for these relatively large molecules. (Int J Mass Spectrom 185/186/187 (1999) 935–948) © 1999 Elsevier Science B.V.

Keywords: Proton affinity; Polyglycine; Density functional theory; Structure; Zwitterion

1. Introduction

Hydrogen bonding plays an important role in the conformation and function of proteins and peptides [1,2]. Hydrogen bonds can occur both intramolecularly between many different function groups as well as intermolecularly between polar groups and surrounding solvent molecules. Gas-phase studies provide a medium in which to investigate intramolecular hydrogen bonding in the absence of solvent. Ionic hydrogen bonding can play a dominant role in the conformation of gas-phase biomolecule ions [3–9].

* Corresponding author. E-mail: williams@cchem.berkeley.edu

Dedicated to Michael T. Bowers on the occasion of his 60th birthday.

PII S1387-3806(98)14263-X

For example, ion mobility experiments of Bowers and co-workers [3] have shown that singly protonated bradykinin forms a compact ball like structure in which the molecule folds up and the charge site is "solvated" by polar groups in the ion, primarily the carbonyl oxygens of the peptide backbone. Similar results have been reported for a variety of peptides as well as larger proteins [4–9].

These intramolecular interactions can also play a significant role in the charge state distributions observed in electrospray mass spectra [10,11] and in the dissociation pathways and products observed in tandem mass spectrometry experiments. For example, dissociation of singly protonated des-Arg^9-bradykinin (Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe) under low energy conditions results in the formation of complementary b_2 and y_6 ions (corresponding to the cleavage of the Pro-Pro backbone bond) [12]. Of the amino acids in this ion, the most basic is Arg (PA = 251.2 kcal/mol) [13]. The next most basic amino acid is Phe (PA = 220.6 kcal/mol) [13]. In the dissociative transition state, the 6 C-terminal residues effectively compete for the proton despite the fact that arginine, the most basic individual amino acid, is in the N-terminal fragment. This shows cooperative solvation of the proton by several groups in a peptide can effectively stabilize the proton and make retention of the proton by the fragment containing the less basic individual residues a competitive pathway.

One measure of the extent of stabilization provided by these intramolecular hydrogen bonding interactions is the gas-phase basicity (GB) or proton affinity (PA) of a molecule. The GB and PA of molecule B are defined as the negative free energy ($-\Delta G$) and enthalpy ($-\Delta H$), respectively, of

$$B + H^+ \leftrightarrow BH^+ \quad (1)$$

These thermodynamic values can be measured using a variety of different experiment methods including equilibrium [14,15], bracketing [16–18], and kinetic methods [19–21]. The relative merits of these methods are discussed in detail elsewhere [21,22]. From GB measurements, Bowers and co-workers [17] and Kebarle and co-worker [23] independently discovered that the basicities of the diaminoalkanes are significantly higher than those of the corresponding monoaminoalkanes. The higher basicity of the diaminoalkanes is due to cyclization resulting in both amino groups stabilizing the proton.

An excellent example of the effects of intramolecular hydrogen bonding on the GB and PA of peptides [24] are the studies of Lebrilla and co-workers [25], Fenselau and co-workers [26,27] and Cassady and co-workers [28] on these values for polyglycines. Fenselau and co-worker [26,27] used the kinetic method to obtain gas-phase basicities of Gly_n (n = 1–10). Lebrilla and co-worker [25] used the bracketing method for n = 1–5 and Cassady and co-workers [28] used both bracketing and kinetic methods for n up to 6. The experimental values are in excellent agreement for n = 1 and 2 but these values deviate significantly for larger n. For pentaglycine, Fenselau and co-worker [26] and Lebrilla and co-worker [25] reported PA values of 234.0 and 223.4, respectively. This 10.6 kcal/mol difference is well outside the range of errors typically associated with basicity measurements. Even more significantly, the data of Fenselau and co-worker indicate that the PA increases from n = 1 to 10, although the rate diminishes as n approaches 10 [26]. In contrast, the data of Lebrilla and co-worker [25] indicate that the PA plateaus at n = 3; the increase for n = 4–5 is about 2 kcal/mol.

The GB and PA of glycine and small polyglycines have also been investigated using computational methods. Rigorous ab initio calculations have been performed on glycine up to the fourth-order Moller–Plesset (MP4), level and on diglycine up to Hartree–Fock (HF) level with electron correlation corrections [28–30]. Recent calculations of protonation energies of glycine using HF, second-order Moller–Plesset (MP2), MP4 and coupled cluster with single and double substitutions (CCSD) have been reported [30]. The protonation energy calculated at the MP2 6-31+G**//HF 6-31G* level was within 1.0 kcal/mol of that calculated at the MP4 6-31+G**//6-31+G** MP2 level and within 1.9 kcal/mol of the experimental data. For diglycine, MP4 values were extrapolated from HF calculations at the 6-31G* level by using correlation energies from glycine calculations at the

MP4 level [31]. The PA of diglycine at the MP4 6-31G* level (222.5 kcal/mol) is in good agreement with the value measured by Fenselau and co-workers [26,27] (220.7 kcal/mol) but is higher than the values measured by Lebrilla and co-workers [25] and Cassady and co-workers [28] by approximately 4 kcal/mol. These (and other) [32] results indicate that correlated ab initio methods are able to reproduce experimentally determined proton affinities quite well. However, Moller–Plesset calculations are not currently feasible for significantly larger molecules.

HF self-consistent field (SCF) calculations were done on triglycine by Zhang et al. [31]. Several low energy conformers were found using AM1 semiempirical calculations and the lowest energy conformer was optimized at the restricted HF (RHF) 6-31G* level. The calculated PA of 226.9 kcal/mol is in good agreement with the value measured by Fenselau and co-worker [26] (224.0 kcal/mol) but this value is significantly higher than values measured by Cassady and co-workers [28] and Lebrilla and co-worker [25].

Semiempirical methods can be performed on significantly larger molecules, but the accuracy of thermochemical data from these calculations on large molecules is not well known. Beauchamp and co-workers reported PA values for polyglycines up to $n = 5$ calculated at both the AM1 and PM3 levels [33]. Normalized to the measured PA of glycine, the calculated PA of pentaglycine was higher than any of the experimental values by ~15 kcal/mol. Calculations at the PM3 level were also done on the salt-bridge isomer of protonated pentaglycine. The salt-bridge was 17.5 kcal/mol less stable than the simple amino protonated form. Recent ion mobility results by Wyttenbach et al. [7] show that salt-bridge forms of Gly_n ($n = 1–6$) are not favorable for either sodium attached or protonated species. Evidence for stable salt bridges in other peptides and amino acid dimers [12,34] has been reported.

Methods based on the Kohn–Sham density functional theory (DFT) have been used to obtain equilibrium geometries, proton transfer reaction energies, and conformational energy differences [32]. DFT methods have the advantage that they require approximately the same computational time as HF, yet include the effects of electron correlation like Moller–Plesset methods. Previous studies on organic molecules have shown that local density approximation functionals are capable of reproducing experimentally determined geometries of small molecules with high accuracy [35–39]. In addition, hybrid functionals, e.g. B3LYP, can accurately account for stabilization due to van der Waals interactions and hydrogen bonding [37–39].

Dramatic improvements in algorithms used in computational chemistry have occurred. The most time consuming methods part of an ab initio calculation is the atomic integral evaluation [40]. In the recent past, the evaluation of atomic orbitals scaled quartically with respect to the number of basis functions [40]. By taking into account only non-negligible integrals, the scaling of SCF calculations becomes quadratic. Recent advances, such as the continuous fast multipole method [41,42], have been implemented. These advances have further reduced the scaling to near linear [43]. Reducing the scaling has an enormous impact on the time required for a HF or DFT calculation on larger molecules. For example, if a calculation on a molecule containing 10 atoms takes 1 h, a calculation on a 50 atom system would take 26 days under quartic scaling. Based on the linear scaling results of White et al. [41], a calculation on this 50 atom system would take approximately 18 h. With low cost workstations that are currently available, calculations on peptides containing 10–15 residues within practical time limits are possible [42]. In this study, we apply DFT methods to glycine polymers ranging in size from Gly_1 to Gly_{10}. Values for the PAs are calculated and compared to experimentally measured values. To the best of our knowledge, these are the first ab initio calculations on glycine polymers larger than Gly_4.

2. Computational procedures

All the polyglycine molecules and ions were built in the peptide builder subprogram of the Macromodel 6.0 computational package [44]. The N-terminal amino group of the glycine peptide was protonated in

both the nonzwitterion and the zwitterion structures [31,33]. For the zwitterion structures, the C-terminus was deprotonated and the carbonyl oxygen of the adjacent residue was protonated. Monte Carlo Multiple Minimum (MCMM) conformational searching [45] with 1000 steps was employed to find the lowest energy structures for Gly_n (n = 3–5). The systematic unbounded multiple minimum (SUMM) method [46] was used for Gly_n (n = 7, 10) with 2000 and 2500 steps, respectively. These searching techniques create new trial structures from old ones by altering a fraction of the torsional bonds within the molecule. During the search, new trial structures are generated from previous minimal structures. For large molecules, minimal conformers have a high probability of containing at least some of the low energy substructures of the global minimum. The goal of finding the global minimum is accomplished by varying a small number of the bonds randomly until all of the minimal substructures are found. The SUMM method is a variant of Monte Carlo searching in which the entire conformation space is sampled at low resolution at the outset of the search. The SUMM method is more effective for larger molecules (>12 variable torsional bonds) because the systematic methods of searching the conformation space avoids re-exploring regions covered in some earlier part of the search. Conformational searches and minimizations were done using the MMFFs using a dielectric constant of 1.0 unless otherwise stated. MMFFs is a variant of Merck molecular force field (MMFF) that contains parameters which reproduces the structure of sp^2 hybridized nitrogens in crystal structures for proteins [44]. The lowest energy structures obtained from the molecular mechanics calculations were used as starting geometries in the semiempirical and DFT calculations.

Ab initio calculations were done in Qchem v1.0.2 [47] on a DEC alphastation 500 and IBM RS/6000 computer. Local density approximation DFT (Slater exchange/VWN correlation) [48] as well as hybrid functional methods (B3LYP) [49] were used in this study. The geometry convergence criterion for the energy change was set at 1×10^{-4} hartrees. Quantum Coulomb Tree Code (QCTC) [50] was enabled for the S-VWN and B3LYP calculations.

GAMESS was used for all semiempirical calculations. Heats of formation of neutral and protonated polyglycines were taken from semiempirical minimized structures for both AM1 [51] and PM3 [52]. The experimental value for the heat of formation of a proton of 367.2 kcal/mol [13] was used.

Vibrational frequencies for the zero-point and internal energy corrections were calculated within Macromodel by taking the lowest energy structure of Gly_n and minimizing the structure again using the truncated Newton conjugate gradient TNCG method [53] to produce low gradient minimized structure that had real vibrational frequencies. The thermodynamic corrections are calculated at 298 K. DFT frequencies were calculated for Gly_3 at the S-VWN, 6-31G* level.

Experimental measurements of the GB and PA of Gly_n were previously carried out using bracketing and kinetic methods. In both methods, proton transfer reactions take place between Gly_n and reference bases of known PA. Accurate measurement of the PA of Gly_n depends on the accurate knowledge of the PA of these bases. These values have been revised since the original experimental measurements were performed. For comparison, values for the PA of Gly_n were revised using rate constants or abundance ratios reported in the original references [25,26,28] and the newly updated values of PA of the reference bases from the most recent NIST database [13]. The following procedure was used to correct the experimental values to account for this change. For proton affinities measured using the kinetic method, two competitive reactions occur:

$$\mathrm{Gly}_n\mathrm{H}^+ + \mathrm{Base} \xleftarrow{k_1} \mathrm{Base}\cdots\mathrm{H}^+\cdots$$

$$Gly_n \xrightarrow{k_2} \mathrm{Gly}_n + \mathrm{BaseH}^+ \qquad (2)$$

where k_1 and k_2 are the measured rate constants for the dissociation of the proton-bound dimers into the two respective products. Values of ln (k_2/k_1) versus the revised PA of the base were plotted. Least squares regression was used to determine the zero intercepts, from which a new value of the PA of each Gly_n was determined. In the bracketing experiments of Lebrilla and co-worker [25] and Cassady and co-workers [28],

Table 1
Experimentally determined proton affinities (in kcal/mol) for Gly_n (n = 1–10); values in the calculated columns have been adjusted to reflect the revised basicity scale (see the text)

	Cassady and co-workers [28]		Fenselau and co-worker [26]		Lebrilla and co-worker [25]	
n	Reported	Calculated	Reported	Calculated	Reported	Calculated
1	213.5	209.8 ± 2.2	211.6[a]	211.9[b]	215.4	212.9 ± 4.0
2	223.6	217.9 ± 2.5	219.1	220.7 ± 0.8	224.5	218.5 ± 4.0
3	227.2	221.9 ± 2.9	223.1	224.7 ± 0.5	226.0	221.3 ± 4.0
4	233.3	227.7 ± 2.3	227.2	229.7 ± 0.4	226.0	221.3 ± 4.0
5	233.6	228.0 ± 4.6	231.8	234.0 ± 0.7	228.1	223.4 ± 4.0
6	235.7	231.1 ± 4.6	234.4	238.1 ± 0.6	—	
7			237.0	241.7 ± 2.5	—	
10			244.0	249.0 ± 2.5	—	

[a] Value from NIST database 1984.
[b] Value from [13].

the value for the GB of Gly_n was given as the mean of the reference bases that were immediately higher and lower in basicity. Their values were revised using the new values for the GB of the reference bases. The entropy values used to convert basicity to PA are the same ones used by the original authors. The uncertainties reported by the respective authors are listed in Table 1 without modification. These values reflect uncertainties in the PAs of the reference bases and experimental error in the measurements.

3. Results and discussion

3.1. Mechanics calculations

The minimal structures obtained for protonated and neutral Gly_n are shown in Fig. 1. The minimal structure obtained for neutral glycine is in agreement with the structure deduced from previous microwave spectroscopy experiments [54,55]. The structure of protonated glycine is also the same as that reported previously [28]. In this structure, there are two hydrogen bonds between the amino hydrogens and the carbonyl oxygen. For triglycine [Fig. 1(c) and (d)], the optimized geometries are different than those reported by Zhang et al. [31]. The previously reported protonated structure has two of the N1 amino hydrogens bonding to O1 and O3 rather than to O2 and O3 shown in Fig. 1(c). The energies obtained from MMFFs indicates that the structure of Gly_3H^+ [Fig. 1(c)] and of Gly_3 [Fig. 1(d)] are more stable than those reported by Zhang et al. [31] by 1.2 and 0.3 kcal/mol, respectively. These differences in energy are negligible and suggest that both Gly_3 structures are comparable.

The number of possible molecular conformations increases exponentially with molecular size. Finding the global minimum or minima is a key problem in extending high level computational methods to larger systems. One approach (used here) is to search the conformational space using lower level calculations. For even moderate size peptides, molecular mechanics is the only practical method currently available. One possible difficulty with this approach is that the true global minimum may never be reached in the search. A second possibility is that the true global minimum may be found, but discarded due to inaccurate energetic calculations at the lower level. If either of these cases occurs, the subsequent higher level calculations will be inaccurate independent of the level of theory used.

Several methods of searching the conformation space of a molecule are available, including simulated annealing [56], internal [57] and Cartesian [58] coordinate searching. Studies on the effectiveness of conformational searching have focused on the multiple minima problem, i.e. finding all of the local minima within 3 kcal/mol of the very lowest energy structure. For molecules with less than 12 torsion bonds, internal coordinate conformational searching methods are well suited to solving the multiple minima problem [46,59]. This suggests that for Gly_n,

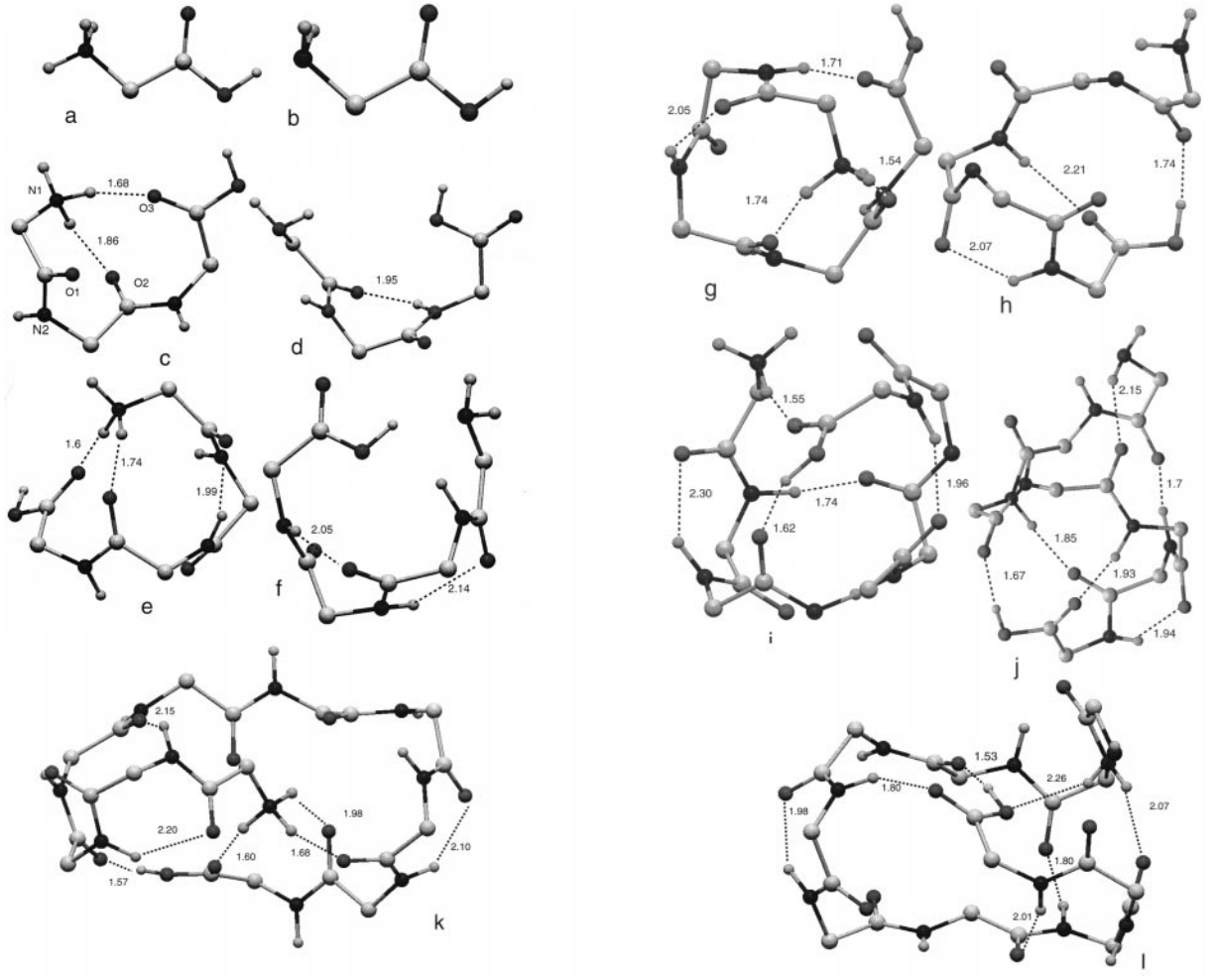

Fig. 1. Minimal structures for Gly_n; (a) $GlyH^+$, (b) Gly, (c) Gly_3H^+, (d) Gly_3, (e) Gly_4H^+, (f) Gly_4, (g) Gly_5H^+, (h) Gly_5, (i) Gly_7H^+, (j) Gly_7, (k) $Gly_{10}H^+$, (l) Gly_{10}. Distances between hydrogen bound atoms are included. Hydrogens bonded to carbons are omitted for clarity.

$n \leq 5$, it is very likely that the global minimal structure (defined by the mechanics force field used) has been found. For molecules with greater than 12 torsional bonds, several methods have been devised that identify regions in conformational space where low energy conformers are likely to be found. These methods increase the probability of finding low energy conformers. These methods have been applied to several cyclic hydrocarbons (up to a 17 carbon rings and 14 torsional bonds) and have satisfactorily identified 99% of the minima [46,59]. Gly_7 and Gly_{10} have nearly the same number of torsional bonds as a C_{17} ring (15 and 20, respectively). These results suggest that it should be possible to find the global minimal structure or a structure within 1–2 kcal/mol of the global minimal structure, on the MMFFs potential energy surface.

A key question is how accurately does MMFFs represent the actual conformational surface of Gly_n? The energies and hence geometries from molecular mechanics strongly depend on the force field used. Studies done on model systems with several force fields have shown that MMFFs is the among the most effective at modeling the energetics of polypeptides. Friesner and co-workers [60] compared mechanics and uncorrelated HF energies to those from LMP2/

cc-pVTZ(-f) calculations for a small set of conformers of Ala_4. MMFFs and Opls-AA [61] were significantly better at reproducing LMP2 values than the other force fields tested. The rms deviation in conformational energy for MMFFs values compared to LMP2/cc-pVTZ(-f) values was 1.40 kcal/mol [60]. The rms deviation for HF was 1.1 kcal/mol. For Ala_4, Friesner and co-workers [60] concluded that MMFFs was "comparable to Hartree–Fock" in root mean square error for relative energetics of conformers. Based on these results, MMFFs was used to find minimal energy structures and higher level calculations were done only on the lowest energy structures. For Gly_n ($n = 3$–5), the error in energies from MMFFs is most likely to be the comparable to that of tetraalanine.

For larger polyglycines, the reliability of mechanics is harder to assess. It is apparent from Fig. 1 that many hydrogen bonding interactions are possible, including interaction of a single hydrogen with two heteroatoms. Mechanics simulations are based on atom centered charges and have no capability for asymmetric or polarizable charge. Thus, mechanics are not expected to model these interactions well [60]. Dielectric constants higher than 1.0 have been suggested for the interior environment of proteins [8] and could partially substitute for polarizability in mechanics calculations. For Gly_7 and Gly_{10}, additional minimizations were carried out for Gly_7 and Gly_{10} with distance dependent dielectric constants up to 1.5 for several of the lowest energy conformers of each. The minimal structure did not change.

3.2. Zero-point and thermal energy corrections

Vibrational frequencies used in the zero-point and internal energy corrections were calculated using molecular mechanics. Higher vibrational states of low frequency modes are more highly populated. Therefore, errors in frequencies below 500 cm^{-1} can potentially produce unreliable internal energies. Halgren [62] compared frequencies for 15 small organic molecules calculated using mechanics to those measured experimentally. The root mean square error between 171 frequencies calculated by MMFF and experimentally measured values was 61 cm^{-1} [62]. Based on these results, MMFF should be capable of providing reasonable frequencies for these thermochemical corrections.

Table 2
Proton affinities (in kcal/mol) of Gly_n at various levels of theory; the structures were minimized with the corresponding methods, except for B3LYP which represents B3LYP6-31G*//S-VWN6-31G*

n	AM1	PM3	S-VWN	B3LYP
1	201.6	202.0	213.0	215.9
3	215.4	216.2	227.1	227.8
4	224.9	223.4	232.8	236.5
5	230.3	229.6	234.6	237.7
7	239.0	233.8	243.9	245.8
10	240.8	248.2	246.6	251.1

As a further test of the reliability of the thermochemical corrections using mechanics frequencies, the zero-point and thermal energy corrections for Gly_3 were calculated with DFT, HF, and mechanics. (These values are given in Table 4.) A scale factor of 0.965 was used to adjust for anharmonicity of the S-VWN frequencies [35,63]. The thermal energy correction calculated from mechanics and DFT frequencies differ by only 0.016 kcal/mol. This indicates that there are no significant systematic differences between the mechanics and DFT derived low frequency modes. The difference in the zero-point energy correction between these two methods is 0.40 kcal/mol. This indicates that some of the mechanics frequencies are too large for the protonated form. A value of 0.4 kcal/mol (140 cm^{-1}) is only a small fraction of the total zero-point energy (128 kcal/mol at 298 K) for triglycine. Lower basis set HF 3-21G calculations differ by only 0.2 kcal/mol from S-VWN calculations. These results suggest that the zero-point and thermal energy can be accurately calculated with mechanics.

3.3. Density functional calculations

The protonation energies of Gly_n are listed in Table 2. Local density functional methods can accurately reproduce the lengths of covalent bonds that have been measured experimentally or those obtained from accurate ab initio methods, such as MP2 and QCISD [35–37,64]. Johnson et al. [64] compared

equilibrium geometries for 32 small neutral molecules using several DFT methods. The mean deviation for S-VWN bond lengths compared to experiment was 0.014 Å. For the 6-31G* basis set, HF, B-LYP, S-VWN had the same absolute deviations for bond lengths within 0.001 Å. For larger organic molecules, studies indicate that S-VWN reproduces experimental geometries as well as or better than HF using the same basis [35–37]. For glycine, geometry optimization at the HF 6-31G* and MP2 6-31G* level resulted in a difference of only one kcal/mol in the final MP2 6-31G* protonation energy. Based on these results, S-VWN 6-31G* should be suitable for geometry optimization.

Hybrid functionals, such as B3LYP, more accurately reproduce experimental energies than HF methods [35–39,65]. For hydrogen bonded dimers and clusters, intermolecular bond distances and stabilization energies approach or are equal to MP2 methods. Single-point energies were calculated using B3LYP using LDA minimized geometries.

3.4. Proton affinities

Values for the PAs of Gly_n were calculated using [66,67]:

$$PA = \Delta E_{DFT} + \Delta(E - E_0) + \Delta E_{ZPE} + 5/2(RT) \quad (3)$$

where ΔE_{DFT} is the protonation energy or the difference in energy between Gly_n and Gly_nH^+ [$E_{DFT}(Gly_n) - E_{DFT}(Gly_nH^+)$], $\Delta(E - E_0)$ is the difference in the internal energy [$(E - E_0)(Gly_n) - (E - E_0)(Gly_nH^+)$] and ΔE_{ZPE} is the difference in the zero-point energy between the neutral and protonated forms of Gly_n[$E_{ZPE}(Gly_n) - E_{ZPE}(Gly_nH^+)$]. Values of ΔE_{DFT}, $\Delta(E - E_0)$ and ΔE_{ZPE} are listed in Tables 3–5 for Gly_n ($n = 1, 10$). The $5/2(RT)$ term corresponds to the classical estimation of the loss of 3 degrees of freedom [$3/2(RT)$] plus the PV term (RT). At 298 K, $5/2\ RT$ is equal to 1.48 kcal/mol. PAs of Gly_n calculated at both DFT levels and both semiempirical levels are given in Table 1.

For glycine, the protonation energies (ΔE_{DFT}) calculated at the S-VWN (220.2 kcal/mol) and B3LYP//S-VWN levels (223.1 kcal/mol) are in good agreement with those calculated by Zhang et al. [30] using B3LYP 6-311++G** (219.15 kcal/mol) and MP4 6-31+G** (220.8 kcal/mol). The slightly higher value with B3LYP//S-VWN is likely due to the smaller basis set used. Adding diffuse functions would likely lower the PA by 1.0–1.5 kcal/mol based on trends described by Zhang et al. [30] for glycine. Zhang et al. reported that the basis set superposition error (BSSE) in MP4 calculations are approximately 1.0 kcal/mol for glycine. They also concluded that this error does not depend strongly on molecular size. In addition, this error should be less for DFT energies than for MP energies [37]. Non-BSSE corrected B3LYP values are within ~1.0 kcal/mol of MP4 BSSE corrected values for glycine [30].

To evaluate whether the basis set used for the larger polyglycines is adequate, geometry optimizations of Gly_5 at the B3LYP 6-311G** level were performed. Single-point energies were calculated using the 6-31G* and 6-311G** basis sets. The value of ΔE_{DFT} with the 6-311G** basis (6-311G** B3LYP

Table 3
Energies (in hartrees, unless noted) at the S-VWN6-31G*//S-VWN

n	Gly_nH^+	Gly_n	ΔE_{DFT}	ΔE_{DFT} (kcal/mol)
1	−282.526 423	−282.175 467	0.350 955	220.2
3	−695.230 175	−694.856 250	0.373 925	234.6
4	−901.640 738	−901.258 940	0.381 799	239.6
5	−1107.997 262	−1107.611 61	0.385 647	242.0
7	−1520.753 427	−1520.352 43	0.400 990	251.6
10	−2139.816 305	−2139.412 52	0.403 804	253.4

Table 4
Energies (in hartrees, unless noted) at the B3LYP6-31G*//S-VWN level

n	Gly_nH^+	Gly_n	ΔE_{DFT}	ΔE_{DFT} (kcal/mol)
1	−284.624 25	−284.268 57	0.355 68	223.2
3	−700.445 43	−700.070 43	0.375 01	235.3
4	−908.362 66	−907.974 81	0.387 85	243.3
5	−1116.268 4	−1115.877 7	0.390 64	245.1
7	−1532.086 60	−1531.682 52	0.404 08	253.6
10	−2155.818 52	−2155.407 36	0.411 16	258.0

Table 5
Zero-point energy and thermal energy corrections (in kcal/mol) for triglycine using ab initio and mechanics generated frequencies

n	ΔE_{ZPE}	$\Delta(E - E_0)$	Level
1	−8.60	−0.11	MMFFs
3	−8.81	−0.18	MMFFs
3	−8.41	−0.20	(S-VWN6-31G*) ν scaled by 0.965
3	−8.63	−0.09	(HF3-21G) ν scaled by 0.91 from [31]
4	−8.12	−0.17	MMFFs
5	−8.81	−0.13	MMFFs
7	−9.05	−0.19	MMFFs
10	−8.29	−0.05	MMFFs

optimization) was only 1.0 kcal/mol higher than that calculate with the 6-31G* basis. The results for glycine and pentaglycine both indicate that the PAs calculated with B3LYP 6-31G* are likely to be within a few kcal/mol of values calculated with the higher basis.

3.5. Comparison to experimental values

The PAs of polyglycines as a function of size, calculated at both semiempirical and both DFT levels, are shown in Fig. 2. Also included in Fig. 2 are the experimental PAs that have been measured previously [25,26,28]. These experimental values have been adjusted to reflect changes in PAs for the reference bases [13]. The procedure to make these adjustments is described in the computational procedures section. For Gly and Gly_3, the semiempirical values are lower than the average experimentally measured proton affinities by ~10 and 7 kcal/mol, respectively. There is better agreement with the experimental values for the larger polymers, although for Gly_5, the 11 kcal/mol range in experimental values makes an accurate comparison difficult. The average deviation between the semiempirical AM1 and PM3 PAs and those measured by Fenselau and co-worker [26] are 6.5 and 6.3 kcal/mol, respectively. The B3LYP values are consistently higher than the experimental values. The average deviation between these values and the values measured by Fenselau and co-worker [26] is 4.0 kcal/mol. This deviation for S-VWN PAs is 2.0 kcal/mol.

In order to reduce the effects of any systematic error that may be present in either the calculation or experiments, the PAs obtained by each of the methods are plotted relative to the PA of glycine (Fig. 3). Possible sources of error in the measured values include the approximated entropy terms used to convert the measured gas-phase basicity to PA (bracketing method) or possible entropy effects in the kinetic methods. Referencing all these values to those of glycine should not only reduce systematic error, but should make possible a more direct comparison of the computational methods ability to accurately reproduce stabilization due to solvation of the charge with increasing molecular size.

When plotted relative to glycine, both semiempirical methods systematically overestimate the PA with larger n (Fig. 3). Similar results were obtained by Campbell et al. [33] who reported semiempirical PAs for Gly_n, $n = 1–5$ normalized to glycine. The AM1 PAs were significantly higher than those determined with PM3. The values of the PA reported for Gly_5 were higher than the values reported by Fenselau and co-worker [26] by 14 and 8 kcal/mol with AM1 and PM3, respectively. The average deviation between the semiempirical values reported here and the experimental values of Fenselau and co-worker [26] for Gly_n is 4.6 and 4.4 kcal/mol for AM1 and PM3, respectively. In contrast to previously reported results [33], we find that there are no significant differences in the PAs calculated using either method, but we do find that these methods overestimate the PAs for the larger polymers in agreement with the findings of Campbell et al. [33].

When plotted relative to glycine, the agreement between the values calculated by both DFT methods with the values measured by Fenselau and co-worker [26] is excellent for all Gly_n. The average and maximum deviation between the B3LYP values and those measured by Fenselau and co-worker [26] are 1.2 and 2.9 kcal/mol, respectively. The average and maximum deviation for S-VWN is 1.7 and 3.5 kcal/mol. The substantially better agreement between the B3LYP values and the values of Fenselau and co-

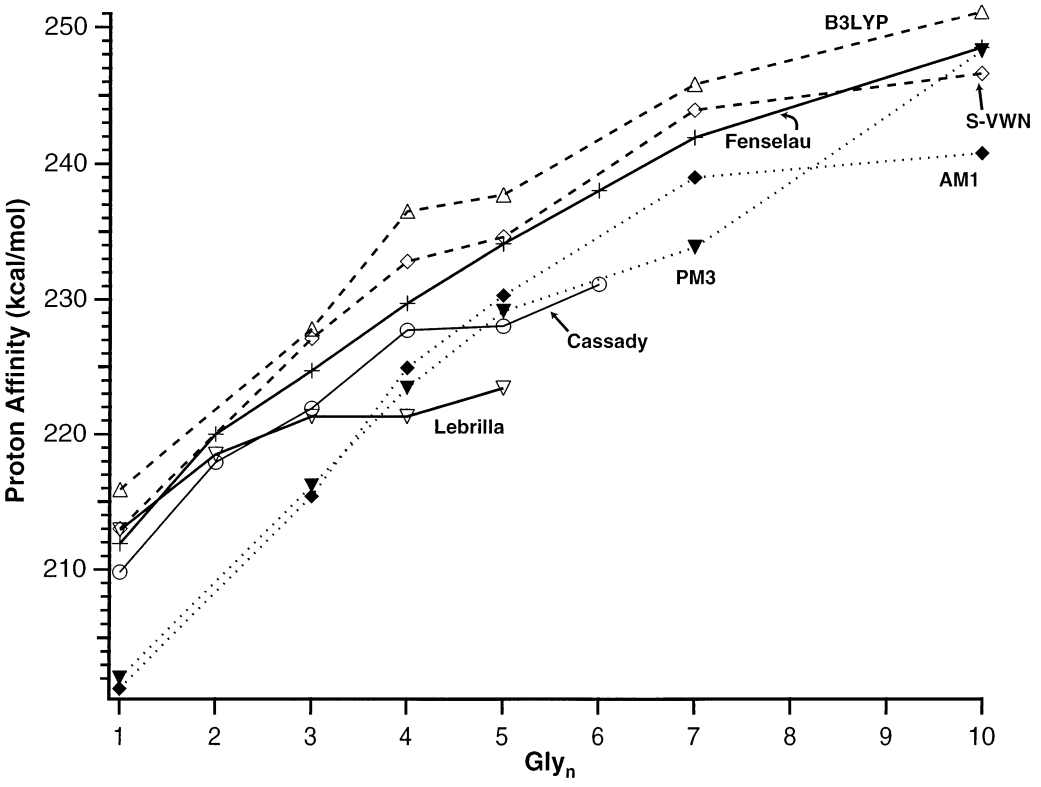

Fig. 2. Proton affinity of Gly_n as a function of size. Experimental values (solid lines) are from Fenselau and co-worker [26] (plus signs), Lebrilla and co-worker [25] (open inverted triangle), Cassady and co-workers [28] (open circle) and have been adjusted to the revised basicity scale [13]. Semiempirical calculations (dotted lines) were done at the AM1 (closed diamond) and PM3 (closed inverted triangle) level, and DFT (dashed lines) calculations were done at the S-VWN 6-31G* (open diamond) and B3LYP 6-31G* (open triangle) level.

worker [26] when the data are normalized to glycine indicates that there is systematic offset in these values. This offset could be due to a number of factors. Adding additional diffuse basis functions to the B3LYP calculations would likely reduce the calculated PAs based on results of Zhang et al. [30] for glycine, and consequently, would reduce this offset. While the agreement between our calculated values and those measured by Fenselau and co-workers is excellent, it should be emphasized that the absolute error in these calculations is greater than the reported deviation with the experimental values. A detailed comparison to the experimental data is made difficult due to the spread in data for Gly_n, $n = 3$–6. However, the close agreement of the calculated values with those values measured by Fenselau and co-worker [26] and Cassady and co-workers [28] suggests that the values measured by Lebrilla and co-worker [25] may be too low for the larger polymers.

3.6. CPU time

With DFT, the time required for calculation of the single-point energies scales near linearly with the number of atoms. Linear scaling has made possible energy calculations of large chain hydrocarbons and biomolecules containing more than 200 atoms [40]. For Gly_n, a correlation coefficient of 0.98 was obtained from a plot of the time required for a single-point energy calculations versus the basis set size. For glycine and Gly_{10}, the time required to obtain B3LYP 6-31G* single-point energies using a DEC Alpha workstation was 8 and 174 min, respectively. In

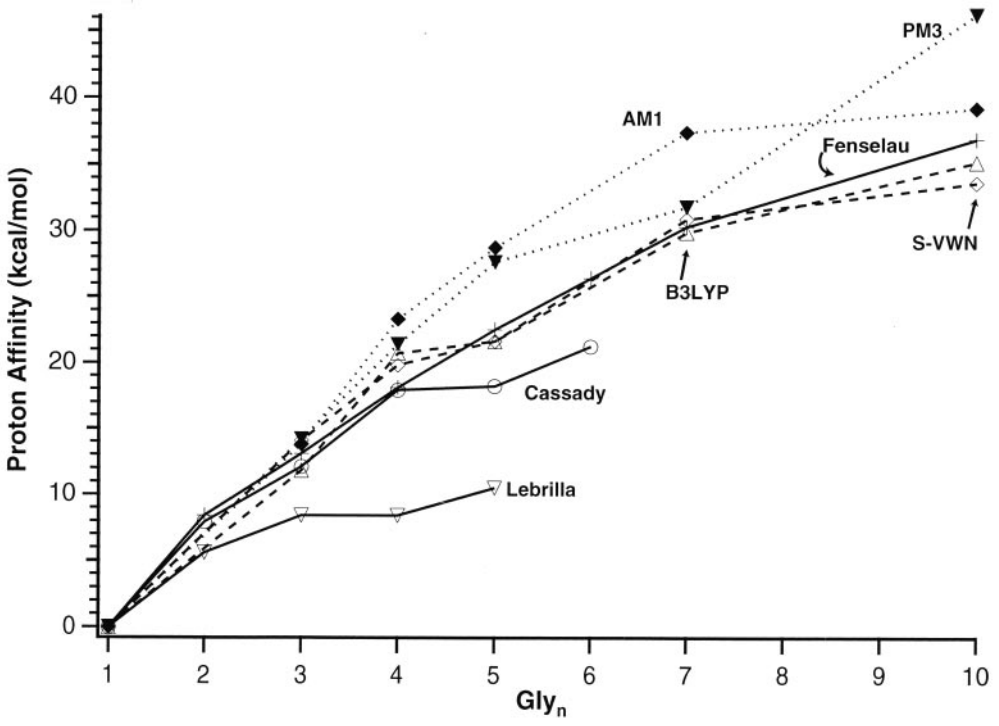

Fig. 3. Proton affinities of Gly_n as a function of size using the same data as in Fig. 2 but normalized to the proton affinity of glycine.

contrast, current geometry optimization techniques do not scale linearly with size. Optimization at the S-VWN 6-31G* level on the larger polyglycine required significant computer time (4 days for Gly_{10}). Calculations using lower basis sets for optimization reduces this time further and may not significantly affect the reliability of the geometries for Gly_n [31].

A significant reduction of computational time can be achieved by using mechanics instead of ab initio calculations to obtain frequencies that are used in the internal and zero-point energy corrections. For Gly_3, the CPU time required for the frequency calculations at the B3LYP level was just under 4 days. For Gly_5, this would require approximately 10 days. However, the internal energy and zero-point energy corrections are only a small fraction of the PA. The comparison of these corrections using mechanics versus DFT frequencies for Gly_3 indicates that calculating frequencies at a higher level is not necessary for accurate corrections.

3.7. Salt-bridge structure and energetics

The relative energies of salt-bridge structures were investigated for Gly_n ($n = 4$, 5, 7 and 10) using both DFT and semiempirical methods. Previous studies indicate that the carbonyl oxygen adjacent to the C-terminal residue of polyglycine is the most basic site among the carbonyl oxygens [31,33]. Salt-bridge structures for these polyglycines were constructed by protonating both this site and the N-terminus, and deprotonating the C-terminus (Fig. 4). Minimum energy structures were determined as described previously. The differences in energy between the zwitterion and the simple protonated form of these polyglycines are given in Table 6. Calculations at all levels indicate that the salt-bridge structure is significantly less stable than the simple amino protonated form. This difference in energy between these two forms is largest with semiempirical calculations.

The smallest difference in energy of the two forms

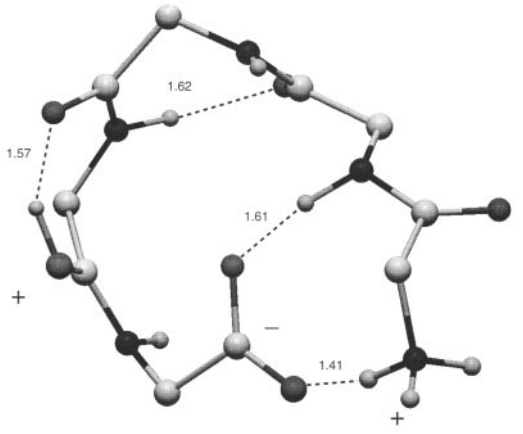

Fig. 4. Salt-bridge structure of Gly_5H^+.

for each polyglycine, except Gly_7, is obtained with S-VWN. Previous studies of the S-VWN method shows that it systematically overestimates the stabilization energy of hydrogen bound molecules [37,64,68]. The salt-bridge structures of polyglycines are more compact than the simple protonated forms [7]. This may result in a slight overestimation of the stabilization from hydrogen bonding for the salt-bridge structures. Increasing the level of theory for Gly_4 and Gly_5 from B3LYP//S-VWN 6-31G* to B3LYP//B3LYP 6-311G** results in an average increase in the relative stability of the simple protonated form by approximately 3 kcal/mol. Higher level calculations were not performed for Gly_7 or Gly_{10} but this same trend is expected for these ions as well. In general, larger basis sets are required to accurately model negative ions as occurs in the salt-bridge form.

Table 6
Difference in energy between the zwitterion and simple protonated form of Gly_n (kcal/mol); positive values indicate that the salt-bridge form is less stable. The DFT basis level is 6-31G* unless otherwise indicated; energies do not include zero-point corrections (see the text)

n	AM1	PM3	S-VWN	B3LYP// S-VWN	B3LYP 6-311G**// 6-311G**
4	+27.2	+22.4	+8.4	+13.6	+15.8
5	+33.0	+22.5	+4.8	+14.5	+18.7
7	+27.2	+16.3	+9.3	+8.2	
10	+27.0	+29.0	+6.0	+9.8	

Zero-point corrections calculated from the mechanics frequencies for the salt-bridge form of Gly_n (n = 4, 5, and 7) relative to the simple protonated form are 2.3, 1.6, and 2.4 kcal/mol, respectively. Thus, it appears that the difference in energy obtained at the B3LYP//S-VWN 6-31G* level is too low by about 4–5 kcal/mol.

Even higher level calculations are required to obtain a more accurate difference between these structures [32]. However, the large energy differences determined at the current level of theory strongly indicate that the salt-bridge form of these polyglycines is not energetically competitive. These results are consistent with the experimental results of Wyttenbach et al. [7], which indicate that Gly_nH^+ and Gly_nNa^+ (n = 1–6) exist in their simple cationized form. It is interesting to note the slight decrease in the energy difference between these two forms of glycine with increasing size. This suggests that there may be some size polyglycine ($n \gg 10$) where the salt-bridge form may become energetically competitive.

4. Conclusions

Values for the proton affinities of Gly_n, n = 1–10, calculated using semiempirical and DFT methods are reported. For the smaller glycines, the agreement between the DFT values reported here, and those calculated at higher levels of theory and those measured experimentally is excellent. This suggests that the proton affinities can be accurately calculated at the level of theory used. For the larger glycines, the accuracy is more difficult to assess due to the significant range in the experimentally reported values and the lack of high level calculations. The best agreement is obtained with Fenselau and co-worker [26] values for comparison. The average deviation between values calculated at the B3LYP//S-VWN 6-31G* level and the experimentally measured values of Fenselau and co-worker [26] is 4.0 kcal/mol. This deviation is only 1.2 kcal/mol when all data is normalized to the PA of glycine. This latter result indicates that B3LYP can accurately take into account intramolecular sol-

vation effects that are responsible for the increasing PA of polyglycine with increasing chain length.

A key assumption in these calculations is that the lowest energy structure or structures close in energy to the lowest energy structure can be identified. The approach used here relies on mechanics calculations and conformation searching to find these low energy structures. The success of this approach depends both on the accuracy of the energetics from the force field used and on the ability to comprehensively explore the conformational energy surface. The latter problem becomes particularly acute with increasing molecular size. In addition to errors due to the mechanics force fields and possible problems with conformation searching, there are also errors associated with DFT calculations using moderate basis sets. While the combined magnitude of these errors is difficult to assess, the excellent agreement between the values of the proton affinity values calculated at the B3LYP//S-VWN 6-31G* level and the experimentally measured values suggest that this approach can be successfully used to obtain thermochemical information on large peptides (~75 atoms) using readily available low-cost workstations.

Acknowledgements

The authors would like to thank Micheal Lee, Dr. David Maurice, and Professor Martin Head-Gordon for their assistance with the Q-chem computational package and Dr. Paul Schnier for helpful comments. The authors acknowledge the generous financial support from the National Science Foundation (grant no. CHE-9726183) and the National Institutes of Health (grant no. IR29GM50336-01A2).

References

[1] C.R. Cantor, P.R. Schimmel, Biophysical Chemistry, Freeman, New York, 1980.
[2] T.E. Creighton, Proteins: Structure and Molecular Properties, 2nd ed., Freeman, New York, 1993.
[3] T. Wyttenbach, G. von Helden, M.T. Bowers, J. Am. Chem. Soc. 118 (1996) 8355.
[4] K.B. Shelimov, D.E. Clemmer, R.R. Hudgins, M.F. Jarrold, J. Am. Chem. Soc. 118 (1996) 2240.
[5] T.D. Wood, R.A. Chorusch, F.M. Wampler, D.P. Little, F.W. McLafferty, Proc. Natl. Acad. Science USA 92 (1995) 2451.
[6] T. Covey, D.J. Douglas, J. Am. Soc. Mass Spectrom. 4 (1993) 616.
[7] T. Wyttenbach, J. Bushnell, M.T. Bowers, J. Am. Chem. Soc. 120 (1998) 5098.
[8] (a) D.S. Gross, E.R. Williams, J. Am. Chem. Soc. 117 (1995) 883. (b) D.S. Gross, P.D. Schnier, S.E. Rodriguez-Cruz, C.K. Fagerquist, E.R. Williams, Proc. Natl. Acad. Science USA 93 (1996) 3143.
[9] K.B. Shelimov, M.F. Jarrold, J. Am. Chem. Soc. 119 (1997) 2987.
[10] (a) P.D. Schnier, D.S. Gross, E.R. Williams, J. Am. Chem. Soc. 117 (1995) 6747. (b) P.D. Schnier, D.S. Gross, E.R. Williams, J. Am. Soc. Mass Spectrom. 6 (105) 1086. (c) P.D. Schnier, W.D. Price, E.R. Williams, ibid. 7 (1996) 972.
[11] M. Miteva, P.A. Demirev, A.D. Karshikoff, J. Phys. Chem. B 101 (1997) 9645.
[12] P.D. Schnier, W.D. Price, R.A. Jockusch, E.R. Williams, J. Am. Chem. Soc. 118 (1996) 7178.
[13] E.P. Hunter, S.G. Lias, in W.G. Mallard, P.J. Linstrom (Eds.), NIST Chemistry WebBook, NIST Standard Reference Database Number 69, March 1998, National Institute of Standards and Technology, Gaithersburg, MD 20899 (http://webbook.nist.gov/chemistry).
[14] J.E. Szulekjo, T.B. McMahon, J. Am. Chem. Soc. 115 (1993) 7839.
[15] M. Mautner, L.W. Sieck, J. Am. Chem. Soc. 113 (1991) 4448.
[16] J. Defrees, R.T. McIver Jr., W.J. Hehre, J. Am. Chem. Soc. 102 (1980) 3334.
[17] D.H. Aue, H.M. Webb, M.T. Bowers, J. Am. Chem. Soc. 95 (1973) 2699.
[18] J.L. Beauchamp, Ann. Rev. Phys. Chem. 22 (1971) 527.
[19] S.A. McLuckey, D. Cameron, R.G. Cooks, J. Am. Chem. Soc. 103 (1981) 1313.
[20] R.G. Cooks, J.S. Patrick, T. Kotiaho, S.A. McLuckey, Mass Spectrom. Rev. 13 (1994) 287.
[21] R.G. Cooks, P.S. Wong, Acc. Chem. Res. 31 (1998) 379.
[22] A.A. Bliznyuk, H.F. Schaefer III, I.J. Amster, J. Am. Chem. Soc. 115 (1993) 5149.
[23] R. Yamdagni, P. Keberle, J. Am. Chem. Soc. 95 (1973) 3504.
[24] A. Harrison, Mass Spectrom. Rev. 16 (1997) 201.
[25] J. Wu, C.B. Lebrilla, J. Am. Chem. Soc. 115 (1993) 3270.
[26] Z. Wu, C. Fenselau, J. Am. Soc. Mass Spectrom. 3 (1992) 863.
[27] Z. Wu, C. Fenselau, Tetrahedron 49 (1993) 9197.
[28] K. Zhang, D.M. Zimmerman, A. Chung-Phillips, C. Cassady, J. Am. Chem. Soc. 115 (1993) 10812.
[29] J.H. Jensen, M.S. Gordon, J. Am. Chem. Soc. 113 (1991) 7917.
[30] K. Zhang, A. Chong-Phillips, J. Phys. Chem. A 102 (1998) 3625.
[31] K. Zhang, C.J. Cassady, A. Chong-Phillips, J. Am. Chem. Soc. 116 (1994) 11512.
[32] W. Hehre, J. Yu, P. Klunzinger, A Guide to Density Func-

tional Calculations in Spartan, Wavefunction, Irvine, CA, 1997.
[33] S. Campbell, M.T. Rodgers, E.M. Marzluff, J.L. Beauchamp, J. Am. Chem. Soc. 117 (1995) 12840.
[34] W.D. Price, R.A. Jockusch, E.R. Williams, J. Am. Chem. Soc. 119 (1997) 11988.
[35] A.A. El-Azhary, H.U. Suter, J. Phys. Chem. 100 (1996) 15056.
[36] A.A. El-Azhary, Spectrochim. Acta A 52 (1996) 33.
[37] Z. Latajka, Y. Bouteiller, J. Chem. Phys. 101 (1994) 9793.
[38] J. Karpfen, J. Phys. Chem. 100 (1996) 13476.
[39] I.A. Topol, S.K. Burt, A.A. Rashin, Chem. Phys. Lett. 247 (1995) 12.
[40] M. Head-Gordon, J. Phys. Chem. 100 (1996) 13213.
[41] C.A. White, B.G. Johnson, P.M.W. Gill, M. Head-Gordon, Chem. Phys. Lett. 230 (1994) 8.
[42] C.A. White, B.G. Johnson, P.M.W. Gill, M. Head-Gordon, Chem. Phys. Lett. 253 (1996) 268.
[43] Linear scaling of computational time has been obtained for one-dimensional molecules, e.g. carbon chains, up to 300 atoms. For three-dimensional molecules, the scaling of computational time required is between linear and quadratic.
[44] F. Mohamadi, N.G.J. Richards, W.C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W.C. Still, J. Comput. Chem. 11 (1990) 440.
[45] M. Saunders, K.N. Houk, Y.D. Wu, W.C. Still, M.J. Lipton, J. Am. Chem. Soc. 112 (1990) 1419.
[46] J.M. Goodman, W.C. Still, J. Comput. Chem. 12 (1991) 1110.
[47] G. Johnson, P.M.W. Gill, M. Head-Gordon, C.A. White, J. Baker, D.R. Maurice, T.R. Adams, J. Kong, M. Challacombe, E. Schwegler, M. Oumi, C. Ochsenfeld, N. Ishikawa, J. Florian, R.D. Adamson, J.P. Dombroski, R.L. Graham, A. Warshel, Q-CHEM, Version 1.0.2, Q-Chem, Inc., Pittsburgh, PA, 1997.
[48] (a) J.C. Slater, Quantum Theory of Molecules and Solids, Vol. 4: The Self Consistant Field for Molecules and Solids, McGraw-Hill New York, 1974. (b) S.H. Vosko, L. Wilks, M. Nusair, Can. J. Phys. 58 (1980) 1200.
[49] D. Becke, J. Chem. Phys. 98 (1993) 5648.
[50] M. Challacombe, E. Schwegler, J. Almlöf, J. Chem. Phys. 104 (1996) 4685.
[51] M.J. Dewar, E.G. Zoebisch, E.F. Healy, J.J. Stewart, J. Am. Chem. Soc. 107 (1985) 3902.
[52] J.J. Stewart, J. Comput. Chem. 9 (1989) 209.
[53] J.W. Ponder, N.G. Richard, J. Comput. Chem. 8 (1987) 1016.
[54] P.D. Godfrey, R.D. Brown, J. Am. Chem. Soc. 117 (1995) 2019.
[55] R.D. Suenram, F.J. Lovas, J. Am. Chem. Soc. 102 (1980) 7180.
[56] S.R. Wilson, W. Cui, J.W. Moskowitz, K.E. Schmidt, J. Comput. Chem. 12 (1991) 342.
[57] M. Lipton, W.C. Still, J. Comput. Chem. 9 (1988) 343.
[58] M. Saunders, J. Am. Chem. Soc. 109 (1987) 3150.
[59] I. Kolossvary, W.C. Guida, J. Am. Chem. Soc. 118 (1996) 5011.
[60] M.D. Beachy, D. Chasman, R.B. Murphy, T.A. Halgren, R.A. Friesner, J. Am. Chem. Soc. 119 (1997) 5908.
[61] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, J. Am. Chem. Soc. 118 (1996) 11225.
[62] T.A. Halgren, J. Comput. Chem. 17 (1996) 553.
[63] G. Rauhut, P.R. Pulay, J. Phys. Chem. 99 (1995) 3093.
[64] J. Johnson, P.M.W. Gill, J.A. Pople, J. Phys. Chem. 98 (1993) 5612.
[65] K. Mohle, H. Hofmann, J. Mol. Model. 4 (1998) 53.
[66] J.E. Del Bene, H.D. Metter, M.J. Frisch, B.J. Luke, J.A. Pople, J. Phys. Chem. 87 (1983) 3279.
[67] J.E. Del Bene, W.B. Person, K. Szczepaniak, J. Phys. Chem. 99 (1995) 10705.
[68] R. Kaschner, D. Hohl, J. Phys. Chem. A 102 (1998) 5111.