
A Reevaluation of Computed Proton Affinities for the Common α -Amino Acids

Scott Gronert, David C. Simpson, and Keyanna M. Conner

Department of Chemistry, Virginia Commonwealth University, Richmond, Virginia, USA

The proton affinities of the 20 common amino acids have been computed at the G3MP2 level using structures derived from broad conformational searches at a variety of levels including G3MP2. In some cases, the conformational surveys identified more stable species than had been used in previous studies of proton affinities, though the differences in energy are sometimes rather small. The present values are likely the most reliable measure of amino acid proton affinities in the gas phase. An analysis of differences between these values and those obtained experimentally via the kinetic method indicates that the extraction of proton affinities from kinetic method data can potentially lead to large errors linked to the estimation of relative protonation entropies. (J Am Soc Mass Spectrom 2009, 20, 2116–2123) © 2009 Published by Elsevier Inc. on behalf of American Society for Mass Spectrometry

In 2006 in this journal, Paizs and coworkers presented a comprehensive computational study of the gas-phase proton affinities (PAs) of the common amino acids [1]. As noted by these authors, there is substantial variation in the experimental values reported for these species. The origin of this variability is most likely related to the difficulties of completing thermodynamic measurements in the gas phase on materials with low volatility. As a result, many of the experimental values are from kinetic method measurements and are subject to errors associated with the kinetic method approximation (as well as cumulative errors associated with the development of ladder-type scales). Therefore, in these systems, high-level computational work offers an attractive means of resolving differences in the various experimental measurements.

While working on a computational study of the conformational preferences of neutral, gaseous amino acids, we realized that it would be possible to improve on the accuracy of the proton affinities reported by Paizs and coworkers. First, their computational strategy relied on the B3LYP approach to identify global minima and, often, they identified conformations for the neutral amino acids that did not match with those reported in previous computational and experimental studies (apparently in some cases this was the result of the search strategy and in others the computational method). For example, Paizs and coworkers concluded that glycine prefers a structure with an internal hydrogen bond between the carboxylic acid and amine groups. High-level *ab initio*, other B3LYP data [2], and experimental data suggest a structure with a *syn* carboxyl group and a weak interaction between the amine hydrogens and

the carbonyl [3–5]. While the energetic impact of this problem is generally minor, we feel it is important to report values based on the most appropriate conformations that can be obtained within a computational method. For the protonated amino acids, this problem is less pronounced because the hydrogen bonding interactions are much stronger; most computational methods identify similar structures as the global minimum. Second, values reported by Paizs and coworkers are at 0 K. We have added a correction for the thermal energy of a proton, which amounts to about 1.5 kcal/mol at 298 K, as well as for the neutral and protonated amino acid. Although these corrections cancel to some extent and while the general picture presented by Paizs is correct, we found significant differences in computed proton affinities for some systems when these issues were addressed. Therefore, we believe a reevaluation is useful. Here, we present computational proton affinities for the amino acids based on extensive conformational searches, including broad surveys at the highly accurate G3MP2 composite level of theory.

Methods

A key problem in complex, multifunctional species is determining global energy minima. One would like to survey all possible conformations at a high level of theory, but this is not practical in large systems. Instead, one must survey broad conformational surfaces at modest levels of theory and hope that the global minimum will be included in the set of structures retrieved from the survey. In practice, this requires retrieving a relatively large set of structures for subsequent analysis at a higher level of theory. In the present case, we used the Monte Carlo search engine in Spartan02 [6] along with the PM3 method to generate up to 250 structures (ranked in stability) for further analysis. In the case of

Address reprint requests to Dr. S. Gronert, Department of Chemistry, Virginia Commonwealth University, 1001 West Main Street, Richmond, Virginia, 23284, USA. E-mail: sgronert@vcu.edu

smaller amino acids, fewer than 250 structures fell within the energy cutoff of 30 kcal/mol relative to the global minimum. In some cases (arginine, phenylalanine, and tyrosine), we extended our survey to 2500 structures, keeping the best 500. In doing so, we found additional conformers, and for one tautomer of arginine, a more stable conformer. Next, each structure in the set of best conformations was subjected to a single point calculation at the B3LYP/6-31+G(d,p) level in Gaussian 03 [7]. The best 50 structures from these calculations were then subjected to G3MP2 calculations. In the case of the larger amino acids (Arg, His, Lys, Phe, Trp, Tyr), a two-step approach was applied that initially used a simplified composite method. After optimization at the MP2/6-31G* level, an MP4/6-31G* calculation was completed and then corrected for basis set expansion using the same extended basis set used in the G3MP2 calculations (gtmp2large). The approach is outlined in eq 1. The energies were then corrected for zero-point vibrational energy at the HF/6-31G* level scaled by 0.9135 [8]. The best 10 conformations from these calculations were then submitted to full G3MP2 calculations. The reported PA values correspond to the protonation of the conformation with the most favorable calculated free-energy producing the product with the most favorable calculated free-energy (structures in Supplementary Material, which can be found in the electronic version of this article.) The logic behind this decision was that experimental measures of the proton affinities are naturally linked to a protonation process controlled by free-energy rather than enthalpy changes. In several instances, this led to a different conformation being identified as the global minimum. No effort was made to improve the quality of the free-energy calculations in Gaussian 03, which suffer from errors related to the treatment of restricted rotors, and the reported PAs are based on single conformations rather than the ensemble of structures expected to be present at room-temperature. It was assumed that these corrections would have only a minor effect on the accuracy of the computed PAs. Finally, it should be noted that the G3MP2 method uses a modest basis set for optimizations (6-31G(d)), so the ability to characterize hydrogen bonding can be somewhat comprised by the geometry employed in the higher-level calculations.

$$E(\text{composite}) = E(\text{MP4/6-31G}^*) + E(\text{MP2/gtmp2large}) - E(\text{MP2/6-31G}^*) \quad (1)$$

Assessment of Theoretical Approach

At the core of the strategy are the assumptions that the simple PM3 method can identify the global minimum within its 250 best structures, and that the B3LYP method, though unable to always identify the experimentally-determined global minimum for some amino acids, will have the global minimum at the G3MP2 level within its best 50 structures. This strategy differs most markedly from Paizs and coworkers [1] in that they relied on the B3LYP calculations to identify the global minimum and then the high-level ab initio calculations (G2MP2) were applied to a single conformation for each species. The obvious danger in that approach is that the B3LYP method can favor a conformation that is significantly less stable than the conformation that is the global minimum at the higher level of theory.

As a test of the general accuracy of the G3MP2 method with respect to the protonation of amino acid functional groups, several simple compounds were investigated (Table 1) [9]. The accuracy is very good for the expected site of protonation, but some errors are seen with respect to groups expected to provide hydrogen bond acceptors (carboxylic acids and amides); however, all the G3MP2 values are within the expected uncertainties of the experimental values (about ± 2 kcal/mol for most species in the NIST table). In the case of the proton-bound dimer complex of trimethylamine, the computed binding energy is quite good. Values at the B3LYP/6-311++G(d,p) level are also included. The accuracy is roughly the same for the PAs, but the value for the complex is not as good. This is not surprising given that DFT is known to have difficulties with the mid-range van der Waals attractions that are likely in such a complex [10]. Given that mid-range van der Waals should be important in larger amino acids, this is a significant issue.

In the present study, we have opted for the G3MP2 method in the conformational surveys and PA calculations. There are several reasons for this decision that justify the enormous computational effort required

Table 1. G3MP2 values for representative functional groups^a

Base	PA (G3MP2)	PA (B3LYP) ^b	PA (literature) ^c
Methylamine	215.3	215.3	214.9
Imidazole	225.6	226.1	225.3
Guanine	235.8	237.5	235.7
Acetic acid	189.1	188.0	187.3
<i>N,N</i> -dimethylacetamide	218.3	216.9	217.0
[N(CH ₃) ₃] ₂ H ⁺ cluster	22.6 (20.8) ^d	19.7 (19.0) ^d	22.0, 22.6, 22.5 ^d

^aValues in kcal/mol.

^bB3LYP/6-311++G(d,p).

^cReference [9].

^dDissociation energy. Values in parentheses include counterpoise correction as implemented in Gaussian 03.

compared to a B3LYP approach. First, the G3MP2 global minima for neutral amino acids match structures inferred from microwave and other gas-phase spectroscopic methods. The examples include glycine [3, 5], alanine [11–15], isoleucine [16], leucine [17], serine [15, 18], tryptophan [19], and valine [20, 21]. However, there are some examples in the literature of global minima identified with the B3LYP functional that do not match experimental structures (although many do match). Along with those from Paizs, problems have been seen with valine [2] and serine [22]. This could be related to search routines or the B3LYP functional. Second, G3MP2 is much more accurate than B3LYP in general, wide-ranging surveys. For example, with the extended G2 test set, G3 methods have average absolute errors in the 1 kcal/mol range whereas B3LYP is nearly 4 kcal/mol [23]. This is not a truly fair comparison in the present context because B3LYP is known to perform very well for PAs, but not so well for some heats of formation (the largest component of the G2 test set). However, because we are looking at species with complex interactions and limited, accurate experimental data for validation, it is prudent to use a method with the greatest generality. Finally, it is well documented that B3LYP struggles with mid-range van der Waals interactions (see above). Given that the larger amino acids can have significant internal van der Waals interactions, a method that characterizes them reasonably well is desirable.

To explore in greater detail the variation between the B3LYP and G3MP2 conformational surfaces, we completed a full conformational search on glutamine using the B3LYP/6-31+G(d,p) and G3MP2 methodologies (enthalpies at 298 K are used in each case). The results are displayed in Figure 1. In the graph, we have compared energies relative to the global enthalpy minimum, which is the same conformation for glutamine with both methods. Generally, there is good correspondence between the two methods, but the G3MP2 method tends to indicate a larger variation in conformational energies. In addition, there are significant differences in some families of structures. For example, conformations 16–22 represent structures with essentially a linear conformation and no internal hydrogen bonding aside from a syn carboxyl. They are disfavored to a much greater extent at G3MP2, presumably because this method produces more stabilization for the internal hydrogen bond found in the cyclic global minimum (carboxylic acid/amide carbonyl interaction).

Results and Discussion

Computed Proton Affinities

For a fair comparison, each of the structures identified in the Paizs paper as a global minimum was recomputed using our G3MP2 approach (Table S1, Supplementary Content). There are two general reasons why different conformations were identified in the two datasets: (1) the conformation identified by the G3MP2

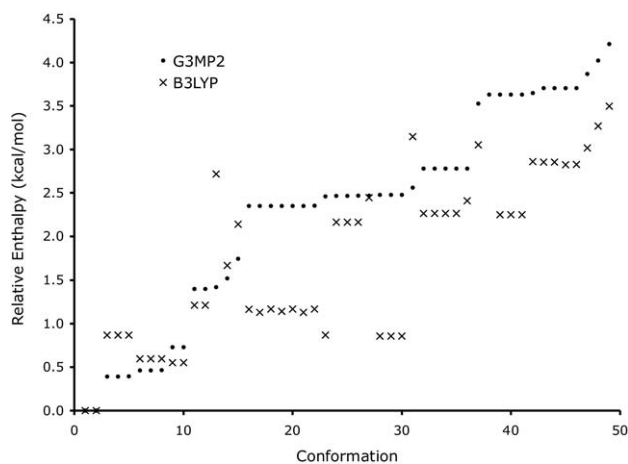


Figure 1. Conformational energies of glutamine relative to the global minimum at the G3MP2 and B3LYP/6-31+G(d,p) levels. Conformations are ordered based on G3MP2 energies. Redundant conformations are observed due to optimizations converging to the same structures. In some cases, optimizations with the two methods led to different conformations despite starting with the same structure.

search has a more favorable enthalpy and free-energy than the one chosen by Paizs with the B3LYP search, or (2) both search approaches identified the same species as the enthalpic global minimum, but we chose to use the conformation with the lowest free-energy rather than enthalpy as the global minimum. The latter scenario leads to the use of an enthalpically less stable species in the proton affinity calculations than the one chosen by Paizs; however as noted above, this conformation is the most relevant in comparisons to experiments. A full discussion of differences in the conformations is presented in the Supplementary Content.

For clarity, we will use the PA values directly reported by Paizs in his paper in the following discussion and therefore they are at the G2MP2 rather than the G3MP2 level. This is a minor effect that amounts to a difference of less than 0.1 kcal/mol for alanine. The values are shown in Table 2 and include G3MP2 values based on the lowest enthalpy as well as the lowest free-energy species. In terms of PAs, the average difference between the present results and those of Paizs is -1.0 kcal/mol (Paizs - present). The largest deviations are ~ 2.0 kcal/mol for Gln, His, and Lys. As noted above, our values include a correction for the enthalpy of the proton, 1.5 kcal/mol, which increases the computed PA. We also include thermal corrections for the other neutral and protonated amino acids. The effect is less significant (average impact = -0.4 kcal/mol on the PAs), but it does vary from amino acid to amino acid. Finally, our choice of 0.9135 for the scaling factor also tends to shift the PAs lower by about 0.2 kcal/mol. We prefer this scaling factor because it gives the greatest accuracy for zero-point energies, a key factor in the PA calculation, and is very close to the optimum scaling factor for low-frequency vibrations, an important contributor to computed free energies [8]. Overall, the

Table 2. Computed values for amino acid proton affinities (kcal/mol)

AA	G3MP2 ^a	Paizs			Maksic ^c			Leszczynski ^d			
		G2MP2 ^b	Dev.	B3LYP ^b	Dev.	Dev.	B3LYP	Dev.	MP2	Dev.	
Gly	211.9	211.8	-0.1	211.4	-0.5	210.5	-1.4	211.5	-0.4	211.9	0.0
Ala	215.5	215.2	-0.3	215.0	-0.5	214.0	-1.5	215.6	0.1	215.1	-0.4
Cys	215.6	214.1	-1.5	214.4	-1.2	213.2	-2.4	216.3	0.7	215.8	0.2
Ser	218.1	217.7	-0.4	217.3	-0.8	216.0	-2.1	217.8	-0.3	217.7	-0.4
Leu	218.5	218.2	-0.3	218.2	-0.3	216.8	-1.7	218.9	0.4	218.2	-0.3
Val	218.5	218.0	-0.5	217.7	-0.8	215.2	-3.3	218.9	0.4	218.0	-0.5
Asp	219.0	218.5	-0.5	218.0	-1.0	217.3	-1.7	219.9	0.9	218.8	-0.2
Ile	219.5	218.8	-0.7	218.7	-0.8	216.5	-3.0	220.0	0.5	219.0	-0.5
Thr	219.7 (219.4)	218.5	-1.2	218.7	-1.0	217.5	-2.2	220.2	0.5	219.2	-0.5
Phe	221.2 (220.7)	219.8	-1.4	220.6	-0.6	223.3	2.1	222.7	1.5	223.1	1.9
Tyr	222.0 (221.5)	220.2	-1.8	220.9	-1.1	221.9	-0.1	223.5	1.5	223.8	1.8
Asn	223.6 (223.7)	222.4	-1.2	224.2	0.6	223.2	-0.4	224.5	0.9	222.7	-0.9
Met	223.8	223.7	-0.1	224.0	0.2	221.7	-2.1	225.1	1.3	224.5	0.7
Trp	224.7			223.8	-0.9	220.7	-4.0	226.2	1.5	226.1	1.4
Pro	225.1	223.8	-1.3	224.0	-1.1	221.9	-3.2	226.1	1.0	225.9	0.8
Glu	226.5	224.7	-1.8	225.8	-0.7	225.9	-0.6	226.6	0.1	226.3	-0.2
Gln	232.6 (232.2)	230.5	-2.1	231.9	-0.7	233.0	0.4	235.2	2.6	232.6	0.0
His	233.9	231.8	-2.1	233.7	-0.2	229.8	-4.1	236.4	2.5	234.0	0.1
Lys	239.1 (237.7)	237.1	-2.0	237.3	-1.8	237.8	-1.3	238.4	-0.7	240.0	0.9
Arg	250.1	250.2	0.1	253.3	3.2	248.6	-1.5	252.8	2.7	249.6	-0.5
Average Dev. ^e			-1.0		-0.5		-1.7		0.9		0.2

^aValues in parenthesis from calculation based on lowest enthalpy species. If a single value is given, both criteria gave the same conformation.

^bReference [1].

^cReference [24].

^dReference [25].

^eDeviation relative to G3MP2 computed values.

average deviation is in line with these systematic corrections and, therefore, subtle differences arise from the conformation choices. Interestingly, the amino acids where we found some of the larger differences in conformational energies have the smallest deviations in PAs because the conformational effect is countered by the systematic corrections. Histidine is a unique case where we identified significantly more stable conformations for both the neutral and protonated amino acids, canceling each other for the most part.

Data are also included in Table 2 from an early systematic study by Maksic and coworkers [24] at the MP2 level, and a more recent one by Leszczynski and coworkers [25] at both the MP2 and B3LYP levels. In each case, extensive manual searches of the conformational space were used to identify global minima. The data from Maksic vary the most from the other computational studies and although they did not publish the geometric details of their proposed global minima, it is likely they were working with different structures in some cases. This interpretation is reinforced by the fact that Leszczynski proposed quite different PA values using the same theoretical approach (MP2). The B3LYP data from Leszczynski parallel those from Paizs, but are generally shifted somewhat higher because they include thermal corrections. Interestingly, Leszczynski's MP2 PA values are within 1 kcal/mol of the G3MP2 values, with the exception of the Phe, Tyr, and Trp, which differ from the G3MP2 values by 1.5–2.0 kcal/mol. These results suggest that the two ab initio methods (MP2 and G3MP2) are providing surprisingly sim-

ilar assessments of the PAs, though higher levels of correlation (i.e., G3MP2) may be needed to analyze the aromatic systems. If thermal corrections are taken into account, the six sets of computational PAs in Table 2 provide strong consensus values for the amino acids. The variations across the sets can be attributed to differences in the methodologies as well as different choices for global minima. However, they all present the same general picture. Proton affinities have also been reported for individual amino acids at various levels of theory. They incorporate a range of search methods for global minima but, with a few exceptions, provide additional support for the values in Table 2 [26–40].

Comparison to Experimental Values

In Table 3, the computed PA values from the present study are reported along with those from Harrison [41], Tabet [42], Bouchoux [34, 43–45], and the NIST website [9]. For Harrison's scale, we have adopted his values that included an estimated entropy correction for cases where protonation was expected to lead to cyclization of the cation (values listed parenthetically in his paper). We excluded the correction for tryptophan because the computational data rule out an interaction with the side chain and the formation of a cyclic cation. The average deviations of this collection of scales from the present G3MP2 scale ranges from 1.0 to 2.4 kcal/mol. It is interesting to note that in the great majority of cases, the previous values are smaller (indicating less basic) than

Table 3. Computed and experimental values for amino acid proton affinities (kcal/mol)

AA	G3MP2 ^a	Harrison ^b	Dev.	Tabet ^c	Dev.	Bouchoux ^d	Dev.	NIST ^e	Dev.
Gly	211.9	210.5	-1.4	212.3	0.4	211.8	-0.1	211.9	0.0
Ala	215.5	214.2	-1.3	214	-1.5	213.8	-1.7	215.5	0.0
Cys	215.6	214	-1.6	214.4	-1.2			215.9	0.3
Ser	218.1	215.2	-2.9	216.4	-1.7	216.2	-1.9	218.6	0.5
Leu	218.5	217.4	-1.1	217.8	-0.7			218.6	0.1
Val	218.5	216.5	-2.0	217.6	-0.9			217.6	-0.9
Asp	219.0	216.4	-2.6	217.4	-1.6	221.5	2.5	217.2	-1.8
Ile	219.5	218.6	-0.9	218.4	-1.1			219.3	-0.2
Thr	219.7	219.5	-0.2	218.5	-1.2			220.5	0.8
Phe	221.2	219.5	-1.7	219.8	-1.4			220.6	-0.6
Tyr	222.0	220.9	-1.1	220.3	-1.7			221	-1.0
Asn	223.6	223.6	0.0	223	-0.6	230.7	7.1	222	-1.6
Met	223.8	224.1	0.3	221.7	-2.1	224.1	0.3	223.6	-0.2
Trp	224.7	223.9	-0.8	223.8	-0.9			226.8	2.1
Pro	225.1	222.1	-3.0	223.9	-1.2	221.3	-3.8	220	-5.1
Glu	226.5	226.4	-0.1	223.3	-3.2	234.7	8.2	218.2	-8.3
Gln	232.6	225	-7.6	226.2	-6.4	236.2	3.6	224.1	-8.5
His	233.9	234.5	0.6	229.2	-4.7	238.1	4.2	236.1	2.2
Lys	239.1	235.6	-3.5	228.6	-10.5	237.6	-1.5	238	-1.1
Arg	250.1	247.8	-2.3			251.2	1.1	251.2	1.1
Average Dev. ^f			-1.7		-2.4		1.7		-1.2

^aBased on lowest free energy species. See Table 2 for values based on lowest enthalpy.

^bReference [41].

^cReference [42].

^dReferences [34,43–45].

^eReference [9].

^fDeviation relative to G3MP2 computed values.

the G3MP2 values. The Paizs paper provides a good discussion of the cases where the experimental proton affinities differ significantly from the computational ones, so only a brief discussion will be provided here. We will use a deviation of 2.5 kcal/mol to define a potentially problematic value. Paizs used the same cutoff, and it is reasonable because experimental PAs generally have uncertainties of this magnitude. With this criteria, the following experimental measurements are questionable based on the G3MP2 results: serine (Harrison), aspartic acid (Harrison, Bouchoux), asparagine (Bouchoux), proline (Harrison, NIST, Bouchoux; however a kinetic method value from Poutsma is very close to the G3MP2 value [38]), glutamic acid (Tabet, NIST, Bouchoux), glutamine (Harrison, Tabet, NIST, Bouchoux), histidine (Tabet, Bouchoux), and lysine (Harrison, Tabet). As noted by Paizs, many of the problematic species yield cations with strong internal hydrogen bonds and cyclic structures. The inability to adequately correct for entropy effects in the formation of these cations is a likely source of error for them (all experimental measures directly probe the free-energy of protonation so assumptions tend to be made in the extraction of the enthalpy of protonation from these data). The most recent data from Tabet [42] and Bouchoux [43, 44] employ the Cooks kinetic method [46, 47], and in Bouchoux's case, make explicit corrections for entropy changes based on the energy dependence of the dissociation processes used in the kinetic method.

The large deviations observed in Table 3 for the complex amino acids (i.e., those expected to exhibit

strong internal hydrogen bonding via cyclic structures) brings into some question the reliability of experimental methods to estimate accurate protonation enthalpies when the substrates and their cations can form internal, hydrogen-bonded networks. A good example is histidine. From a computational point of view, this is not a particularly difficult system. The imidazole ring is rigid, and a limited number of conformations are possible with hydrogen bonding to the other functional groups in the amino acid. It is true that the imidazole can adopt tautomeric forms in histidine, but this can be addressed by searching the conformational space for both. Finally, the G3MP2 method does a fine job on the proton affinity of imidazole (Table 1). As a result, the computational value for histidine, 233.9 kcal/mol, has a high probability of being reasonably accurate. The experimental values vary from 229.2 (Tabet) [42] to 238.1 (Bouchoux) [44]. The first value can be discounted because entropy corrections were not made, but the second is quite high. In contrast, Wu and Fenselau [48] in 1994 obtained a value (234.0 kcal/mol) close to the computed one using an earlier formulation of the kinetic method. It is interesting to note that Bouchoux, and Wu and Fenselau, used similar references compounds and found similar protonation free energies (i.e., gas-phase basicities) for histidine, but the entropy corrections from their implementations of the kinetic method are very different and lead to the large difference in reported proton affinities. Paizs also noted the discrepancy with Bouchoux's value and analyzed it on the basis of the entropy correction [1].

An interesting pair is asparagine and glutamine. Based on the calculations, these homologous amino acids have proton affinities that differ by 9 kcal/mol, a stunning difference for such similar compounds. The difference results from the ability of protonated glutamine to form a very strong internal hydrogen bond between the ammonium and the side-chain amide, whereas the shorter side-chain length in asparagine does not allow this interaction to occur to a significant extent. A hidden difference is that the internally hydrogen-bonded structure of protonated glutamine is severely restricted in internal rotations and probably pays almost the full entropy cost expected for ring formation in a covalent system or strongly hydrogen bonded system (~20 eu). As a result, the 9 kcal/mol advantage in enthalpy could be reduced to about 3 kcal/mol in terms of free-energy at 298 K (less at higher effective temperatures). This may explain why so many of the experimental studies in the literature indicate that glutamine is only a few kcal/mol more basic than asparagine. The entropy effect with these amino acids is not reproduced in the kinetic method studies. Tabet [42] did not include entropy corrections and found that glutamine's PA was about 3 kcal/mol above asparagine, but 6.4 kcal/mol below the computed value. Bouchoux [44] found a 5.5 kcal/mol difference in proton affinities, but their values for asparagine and glutamine are 7.1 and 3.6 kcal/mol over the computed values. In addition, their entropy corrections are similar for the two amino acids. The high entropy correction they used for asparagine is a significant contributor to the deviation from the computed value. Paizs noted that earlier data from the Harrison review could be reconciled with these amino acids if arbitrary, but logical corrections were made for entropy effects [1]. In any case, none of the reported values for glutamine are within 3.5 kcal/mol of the computed value (they range from 8 kcal/mol lower to 3.6 kcal/mol higher).

Glutamic acid is another unusual case. The NIST database value is clearly errant in that most experimental and computed values are at least 4 kcal/mol higher. The computed values, Harrison's scale, and Tabet's scale all place glutamic acid's PA about 6–7 kcal/mol above aspartic acid (1 kcal/mol on the NIST scale). This is much smaller than the glutamine/asparagine pair because the internal hydrogen bonds are weaker than in the amide systems. This is also reflected in the fact that our glutamic acid structure and Paizs's have very different hydrogen bonding patterns, yet similar stabilities. However, Bouchoux [44] finds that glutamic acid's PA is 13 kcal/mol greater than that of aspartic acid, driven mainly by a very large entropy correction for glutamic acid (almost twice as large as lysine and glutamine). Again, it is difficult to reconcile this PA value for glutamic acid with the computed one, and the reason for the discrepancy mainly rests with the experimental entropy correction.

Perspective on Experimental Amino Acid Proton Affinities

The determination of the proton affinities of the common amino acids has proven to be a vexing experimental problem. After years of study, there has not been a coherent dataset from a single group that appears to be free of problematic values for at least a few amino acids (based on comparison to computed values and other experimental datasets). The best set to date appears to be Harrison's careful reanalysis of data from multiple groups. The most recent kinetic method approaches have shown mixed success with the most complicated amino acids (those with side chains capable of forming cyclic protonation products) and clearly contain some values that are significantly in error. For example, the two entropy-corrected kinetic method values for histidine (234.0 [48] and 238.1 [44] kcal/mol) differ by 4 kcal/mol. For glutamine, the only entropy-corrected kinetic method value is nearly 4 kcal/mol higher than the computed one. With glutamic acid, the one entropy-corrected kinetic method value is over 8 kcal/mol away from the computed one and includes a counterintuitive entropy correction (in this case, Tabet's uncorrected kinetic method value is closer to the computed one). The situation with lysine is much more satisfactory and all of the entropy-corrected kinetic method values (238 [48], 240.5 [37], and 237.6, [44] kcal/mol) are within about 2 kcal/mol of the present value. There is also consistency with arginine and the values from Bouchoux [34] and Fenselau [49] are both within 1 kcal/mol of the present value. If one considers the seven amino acids that have a reasonable possibility of side-chain involvement in a cyclic, protonated species (arginine, lysine, histidine, glutamine, asparagine, glutamic acid, and aspartic acid), seven of the 11 entropy-corrected kinetic method values in the literature are within about 3 kcal/mol of the present values. It should be noted that very good consistency is generally seen in the simpler amino acids. Nonetheless, there is a disappointing level of precision and accuracy in the published data for the complicated amino acids. These data suggest that there are still issues with reliability in the experimental determination of the proton affinities of complex, nonvolatile, multifunctional species, and literature values must be viewed carefully. The key problems are that nonvolatility eliminates equilibrium approaches for the most part and the complexity of the species can potentially push the basic assumptions of the kinetic method to or beyond their limits.

Under these circumstances, one might question whether attempting to experimentally determine the proton affinities of multifunctional species, such as amino acids, is a worthwhile enterprise. Although gas-phase chemists have generally focused on determining reaction enthalpies in their studies, it is not clear that these values are the most useful ones in defining the acidity or basicity of a particular compound. In most phenomenologic applications, it is the reaction free-

energy that is operative. That is why acidities and basicities in solution are generally cast in terms of pK_a , a free-energy term. It may be the case that free-energy terms could also provide the most useful measures of basicity of multi-functional compounds in the gas-phase. As the present data indicate, attempts to extract enthalpies (PAs) from kinetic method data can potentially lead to large errors and results that do not reflect the effective basicity of the multifunctional species.

Conclusion

The G3MP2 values listed in Table 2 likely provide the most accurate computed proton affinities to date. Although many of the values are similar to those from Paizs and other workers, there are significant variations in some cases. However, it must be noted that these values in this study refer to comparisons involving single conformations and do not take into account the anticipated ensemble of multiple conformations with relatively similar stabilities that would be present experimentally. That stated, it is a reasonable assumption that these values provide the best measure at this time of the proton affinities of the amino acids. The relative values, however, are not likely to give a good measure of free-energy differences (gas-phase basicities) because large entropy corrections are expected.

Acknowledgments

The authors gratefully acknowledge access to the VCU High-Performance Computing Center. The authors also acknowledge generous support from the National Science Foundation (CHE-0716147) and from the University Pierre et Marie Curie (invited professor program).

Appendix A Supplementary Material

Supplementary material associated with this article may be found in the online version at doi:10.1016/j.jasms.2009.07.006.

References

- Bleilholder, C.; Suhai, S.; Paizs, B. Revising the Proton Affinity Scale of the Naturally Occurring Alpha-Amino Acids. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1275–1281.
- Jones, C. M.; Bernier, M.; Carson, E.; Colyer, K. E.; Metz, R.; Pawlow, A.; Wischow, E. D.; Webb, I.; Andriole, E. J.; Poutsma, J. C. Gas-Phase Acidities of the 20 Protein Amino Acids. *Int. J. Mass Spectrom.* **2007**, *267*, 54–62.
- Hu, C. H.; Shen, M. Z.; Schaefer, H. F. Glycine Conformational Analysis. *J. Am. Chem. Soc.* **1993**, *115*, 2923–2929.
- Suenram, R. D.; Lovas, F. J. Millimeter Wave Spectrum of Glycine—a New Conformer. *J. Am. Chem. Soc.* **1980**, *102*, 7180–7184.
- Godfrey, P. D.; Brown, R. D. Shape of Glycine. *J. Am. Chem. Soc.* **1995**, *117*, 2019–2023.
- Spartan '02 1.0.8: Johnson, J. A.; Deppmeir, B. J.; Driessen, A. J.; Hehre, W. J.; Klunzinger, P. B.; Pham, I. N.; Wantanabe, M.; Wavefunction: Irvine, CA, 2002.
- Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A. Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Pomper, R.; Stratmann, R. E.; Yazayev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; P. M. W. Gill; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03 Revision B04; Gaussian, Inc.: Pittsburgh, PA, 2003.
- Scott, A. P.; Radom, L. Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Moller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem.* **1996**, *100*, 16502–16513.
- Hunter, E. P.; Lias, S. G. In *Nist Standard Reference Database Number 69*; Mallard, W. G., Linstrom, P. J., Eds., National Institute of Standards and Technology (<http://webbook.nist.gov>): Gaithersburg MD, 2008.
- Grimme, S. Seemingly Simple Stereoelectronic Effects in Alkane Isomers and the Implications for Kohn-Sham Density Functional Theory. *Angew. Chem. Int. Ed.* **2006**, *45*, 4460–4464.
- Blanco, S.; Lesarri, A.; Lopez, J. C.; Alonso, J. L. The Gas-Phase Structure of Alanine. *J. Am. Chem. Soc.* **2004**, *126*, 11675–11683.
- Powis, I.; Rennie, E. E.; Hergenbahn, U.; Kugeler, O.; Bussy-Socrate, R. Investigation of the Gas-Phase Amino Acid Alanine by Synchrotron Radiation Photoelectron Spectroscopy. *J. Phys. Chem. A* **2003**, *107*, 25–34.
- Cao, M.; Newton, S. Q.; Pranata, J.; Schaefer, L. Ab-Initio Conformational Analysis of Alanine. *THEOCHEM* **1995**, *332*, 251–267.
- Godfrey, P. D.; Firth, S.; Hatherley, L. D.; Brown, R. D.; Pierlot, A. P. Millimeter-Wave Spectroscopy of Biomolecules—Alanine. *J. Am. Chem. Soc.* **1993**, *115*, 9687–9691.
- Gronert, S.; O'Hair, R. A. J. Ab-Initio Studies of Amino-Acid Conformations. I. The Conformers of Alanine, Serine, and Cysteine. *J. Am. Chem. Soc.* **1995**, *117*, 2071–2081.
- Lesarri, A.; Sanchez, R.; Cocinero, E. J.; Lopez, J. C.; Alonso, J. L. Coded Amino Acids in Gas Phase: The Shape of Isoleucine. *J. Am. Chem. Soc.* **2005**, *127*, 12952–12956.
- Cocinero, E. J.; Lesarri, A.; Grabow, J. U.; Lopez, J. C.; Alonso, J. L. The Shape of Leucine in the Gas Phase. *Chem. Phys. Chem.* **2007**, *8*, 599–604.
- Blanco, S.; Sanz, M. E.; Lopez, J. C.; Alonso, J. L. Revealing the Multiple Structures of Serine. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 20183–20188.
- Compagnon, I.; Hagemester, F. C.; Antoine, R.; Rayane, D.; Broeyer, M.; Dugourd, P.; Hudgins, R. R.; Jarrold, M. F. Permanent Electric Dipole and Conformation of Unsolvated Tryptophan. *J. Am. Chem. Soc.* **2001**, *123*, 8440–8441.
- Lesarri, A.; Cocinero, E. J.; Lopez, J. C.; Alonso, J. L. The Shape of Neutral Valine. *Angew. Chem. Int. Ed.* **2004**, *43*, 605–610.
- Shirazian, S.; Gronert, S. The Gas-Phase Conformations of Valine: An Ab Initio Study. *THEOCHEM* **1997**, *397*, 107–112.
- Miao, R.; Jin, C.; Yang, G. S.; Hong, J.; Zhao, C. M.; Zhu, L. G. Comprehensive Density Functional Theory Study on Serine and Related Ions in Gas Phase: Conformations, Gas Phase Basicities, and Acidities. *J. Phys. Chem. A* **2005**, *109*, 2340–2349.
- Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Assessment of Gaussian-3 and Density-Functional Theories on the G3/05 Test Set of Experimental Energies. *J. Chem. Phys.* **2005**, *123*, 124107.1–124107.12.
- Maksic, Z. B.; Kovacevic, B. Towards the Absolute Proton Affinities of 20 α -Amino Acids. *Chem. Phys. Lett.* **1999**, *307*, 497–504.
- Dinadayalane, T. C.; Sastry, G. N.; Leszczynski, J. Comprehensive Theoretical Study Towards the Accurate Proton Affinity Values of Naturally Occurring Amino Acids. *Int. J. Quantum Chem.* **2006**, *106*, 2920–2933.
- Leng, Y. M.; Zhang, M. L.; Song, C.; Chen, M. L.; Lin, Z. J. A Semi-Empirical and Ab Initio Combined Approach for the Full Conformational Searches of Gaseous Lysine and Lysine-H₂O Complex. *THEOCHEM* **2008**, *858*, 52–65.
- Lioe, H.; O'Hair, R. A. J.; Gronert, S.; Austin, A.; Reid, G. E. Experimental and Theoretical Proton Affinities of Methionine, Methionine Sulfoxide, and Their N- and C-Terminal Derivatives. *Int. J. Mass Spectrom.* **2007**, *267*, 220–232.
- Andriole, E. J.; Colyer, K. E.; Cornell, E.; Poutsma, J. C. Proton Affinity of Canavanine and Canaline, Oxyanalogs of Arginine and Ornithine, from the Extended Kinetic Method. *J. Phys. Chem. A* **2006**, *110*, 11501–11508.
- Kocacevic, B.; Rozman, M.; Klasinc, L.; Szrc, D.; Maksic, Z. B.; Yanez, M. Gas-Phase Structure of Protonated Histidine and Histidine Methyl Ester: Combined Experimental Mass Spectrometry and Theoretical Ab Initio Study. *J. Phys. Chem. A* **2005**, *109*, 8329–8335.
- Mezzache, S.; Afonso, C.; Pepe, C.; Karoyan, P.; Fournier, F.; Tabet, J. C. Proton Affinity of Proline and Modified Prolines Using the Kinetic Method: Role of the Conformation Investigated by Ab Initio Calculations. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1626–1632.
- Topol, I. A.; Burt, S. K.; Toscano, M.; Russo, N. Protonation of Glycine and Alanine: Proton Affinities, Intrinsic Basicities and Proton Transfer Path. *THEOCHEM* **1998**, *430*, 41–49.
- Uggerud, E. The Unimolecular Chemistry of Protonated Glycine and the Proton Affinity of Glycine: A Computational Model. *Theoret. Chem. Accts.* **1997**, *97*, 313–316.

33. Bouchonnet, S.; Hoppilliard, Y. Proton and Sodium-Ion Affinities of Glycine and Its Sodium-Salt in the Gas-Phase—Ab Initio Calculations. *Org. Mass Spectrom.* **1992**, *27*, 71–76.
34. Bouchoux, G.; Desaphy, S.; Bourcier, S.; Malosse, C.; Bimbong, R. N. B. Gas-Phase Protonation Thermochemistry of Arginine. *J. Phys. Chem. B* **2008**, *112*, 3410–3419.
35. Chen, M. L.; Lin, Z. J. Ab Initio Studies of Aspartic Acid Conformers in Gas Phase and in Solution. *J. Chem. Phys.* **2007**, *127*, 154314.1–154314.10.
36. Huang, Z. J.; Lin, Z. J.; Song, C. Protonation Processes and Electronic Spectra of Histidine and Related Ions. *J. Phys. Chem. A* **2007**, *111*, 4340–4352.
37. Schroeder, O. E.; Andriole, E. J.; Carver, K. L.; Colyer, K. E.; Poutsma, J. C. Proton Affinity of Lysine Homologues from the Extended Kinetic Method. *J. Phys. Chem. A* **2004**, *108*, 326–332.
38. Kuntz, A. F.; Boynton, A. W.; David, G. A.; Colyer, K. E.; Poutsma, J. C. The Proton Affinity of Proline Analogs Using the Kinetic Method with Full Entropy Analysis. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 72–81.
39. Rak, J.; Skurski, P.; Simons, J.; Gutowski, M. Low-Energy Tautomers and Conformers of Neutral and Protonated Arginine. *J. Am. Chem. Soc.* **2001**, *123*, 11695–11707.
40. Sun, W.; Kinsel, G. R.; Marynick, D. S. Computational Estimates of the Gas-Phase Basicity and Proton Affinity of Glutamic Acid. *J. Phys. Chem. A* **1999**, *103*, 4113–4117.
41. Harrison, A. G. The Gas-Phase Basicities and Proton Affinities of Amino Acids and Peptides. *Mass Spectrom. Rev.* **1997**, *16*, 201–217.
42. Afonso, C.; Modeste, F.; Breton, P.; Fournier, F.; Tabet, J. C. Proton Affinities of the Commonly Occurring L-Amino Acids by Using Electrospray Ionization-Ion Trap Mass Spectrometry. *Eur. J. Mass Spectrom.* **2000**, *6*, 443–449.
43. Bouchoux, G.; Salpin, J. Y. Gas-Phase Basicity of Glycine, Alanine, Proline, Serine, Lysine, Histidine, and Some of Their Peptides by the Thermokinetic Method. *Eur. J. Mass Spectrom.* **2003**, *9*, 391–402.
44. Bouchoux, G.; Buisson, D. A.; Colas, C.; Sablier, M. Protonation Thermochemistry of α -Amino Acids Bearing a Basic Residue. *Eur. J. Mass Spectrom.* **2004**, *10*, 977–992.
45. Desaphy, S.; Malosse, C.; Bouchoux, G. Gas-Phase Basicity of Methionine. *J. Mass Spectrom.* **2008**, *43*, 116–125.
46. Cooks, R. G.; Kruger, T. L. Intrinsic Basicity Determination Using Metastable Ions. *J. Am. Chem. Soc.* **1977**, *99*, 1279–1281.
47. McLuckey, S. A.; Cameron, D.; Cooks, R. G. Proton Affinities from Dissociations of Proton-Bound Dimers. *J. Am. Chem. Soc.* **1981**, *103*, 1313–1317.
48. Wu, Z. C.; Fenselau, C. Gas-Phase Basicities and Proton Affinities of Lysine and Histidine Measured from the Dissociation of Proton-Bound Dimers. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 777–780.
49. Wu, Z. C.; Fenselau, C. Proton Affinity of Arginine Measured by the Kinetic Approach. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 403–405.