THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

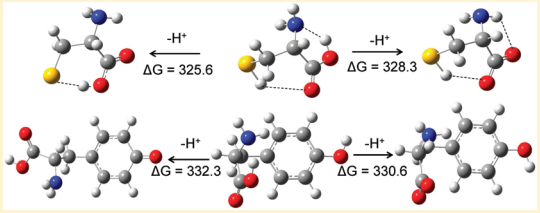
Fundamental Thermochemical Properties of Amino Acids: Gas-Phase and Aqueous Acidities and Gas-Phase Heats of Formation

Michele L. Stover, Virgil E. Jackson, Myrna H. Matus, Margaret A. Adams, Carolyn J. Cassady, and David A. Dixon*

Chemistry Department, The University of Alabama, Shelby Hall, Box 870336, Tuscaloosa, Alabama 35487-0336, United States

Supporting Information

ABSTRACT: The gas-phase acidities of the 20 L-amino acids have been predicted at the composite G3(MP2) level. A broad range of structures of the neutral and anion were studied to determine the lowest energy conformer. Excellent agreement is found with the available experimental gas-phase deprotonation enthalpies, and the calculated values are within experimental error. We predict that tyrosine is deprotonated at the CO_2H site. Cysteine is predicted to be deprotonated at the SH but the proton on the CO_2H is shared with the S^- site. Self-consistent reaction field (SCRF) calculations with the COSMO parametrization were used to predict the pK_a's of the non-zwitterion form in aqueous solution. The differences in the non-zwitterion pK_a values were used to estimate the free energy difference between the zwitterion and nonzwitterion forms in solution. The heats of formation of the neutral compounds were calculated from atomization energies and isodesmic reactions to provide the first reliable set of these values in the gas phase. Further calculations were performed on five rare amino acids to predict their heats of formation, acidities, and pK_a values.



INTRODUCTION

The sequencing of peptides and proteins by mass spectrometry has become a major tool in proteomics research. Gaining information on sequence is frequently a first step to understanding protein structure and function, which are of great importance in biological, biochemical, and medical studies. Biomolecule sequencing by tandem mass spectrometry (MS/MS) has most often employed fragmentation of positive ions. To date, mass spectral studies of the dissociation of negative peptide ions have been far less common than studies of positive peptide ions. Complementing positive ion studies with negative ion studies can readily increase the amount of peptide sequence information obtained without adding significantly to either the cost or time involved because, with modern mass spectrometers, it is very easy to perform negative ion studies by simply switching the sign of the voltages. A number of studies of the fragmentation of deprotonated peptides have been reported.[1−6] Bowie and co-workers[7,8] reported the use of collision-induced dissociation (CID)[9] to characterize the backbone and side chain fragmentations of $(M-H)^-$ anions of underivatized peptides.

The analysis of biomolecules by mass spectrometry requires an understanding of proton transfer reactions because the two most commonly used ionization techniques, electrospray ionization (ESI)[10] and matrix-assisted laser desorption ionization (MALDI),[11] involve the addition and removal of protons. The sites of proton transfer reactions can affect the fragmentation patterns of peptide ions, which consequently impacts the sequence information that can be obtained from mass spectrometry experiments.[12−14] Thermodynamic values, such as the gas-phase acidity (GA or ΔG_{acid}), the ΔG for the deprotonation reaction $AH \rightarrow A^- + H^+$, provide valuable information to help in understanding the less studied negative ion peptide fragmentation mode by mass spectrometry. The study of gas-phase proton transfer reactions provides unique insights into the structures and energetics of the amino acids. Changing the protonation state can impact the hydrogen bonding in the molecule, which is critical to the determination of the three-dimensional structures and biological activities of polypeptides and proteins. Changing the protonation state can also result in decreased or increased properties such as solubility, hydrophobicity, and electrostatic interactions.[15−17]

A range of studies on positive ions of the amino acids have been performed.[18−27] For example, the proton affinities ($B + H^+ \rightarrow BH^+$) of the amino acids have been reported at the G2MP2 level[25] to improve the experimental scale of Harrison.[26] Subsequently, Gronert and co-workers reported the amino acid proton affinities at the G3MP2 level to further refine these values.[27] In this latter work, the proton affinities were reported as the difference between the lowest free energy structure of the neutral and of the protonated species. However, thermodynamic studies of negatively charged amino acids have been less frequent. Locke and McIver[28] used the proton transfer equilibrium method to measure the GAs of glycine and alanine.[29] Kass and co-workers[30] used the extended kinetic and gas-phase equilibrium methods together with density functional theory (DFT) calculations at the B3LYP/aug-cc-pVDZ level and molecular orbital theory calculations at the G3B3 level to determine the acidity of cysteine. They showed that the side

Received: July 29, 2011
Revised: November 22, 2011
Published: January 25, 2012

chain SH group is 3.1 kcal/mol more acidic than the main chain carboxylic acid. Wang and co-workers used photoelectron spectroscopy in combination with electronic structure calculations and found that the cysteine anion is in the thiolate form.[31] Kebarle and co-workers[32] included glycine in their proton transfer equilibrium measurements of the GAs of ninety-six aliphatic carboxylic acids. Bowie and co-workers[33] determined the GAs of nineteen amino acids from the kinetic method of collision-induced dissociation (CID) on a proton bound dimer. However, they were unable to measure the GAs of aspartic acid and glutamic acid, the two most acidic amino acids of the nineteen, because their low volatilities prevented gas-phase equilibrium measurements from being performed. We[34] reported the acidities of glutamic and aspartic acid from ion cyclotron resonance proton transfer reaction bracketing measurements and predicted the acidities of these two amino acids and glycine at the G3MP2[35] level to complete the set. The G3MP2 computational method was chosen on the basis of calculations of the acidities of very strong acids when compared to those calculated at the CCSD(T)/complete basis set level.[36] Gronert and co-workers[37] determined the acidities of glutamic and aspartic acid by extended kinetic methods and noted potential issues with this approach. Poutsma and co-workers[38] recently reported the GAs of all of the amino acids from mass spectrometry experiments using the kinetic and extended kinetic methods. They also calculated the GAs at the B3LYP//6-311++G** level. Miao et al. calculated the GA of serine using density functional theory (DFT) at the B3LYP/6-311+G level.[39] Kass, Wang and co-workers[40] used FT-ICR, anion photoelectron spectroscopy, and DFT B3LYP/aug-cc-pVDZ and G3B3 calculations to show that in tyrosine the phenolic hydroxyl group has about the same acidity as the carboxyl group. They showed that deprotonation is preferred at the OH group as compared to the carboxyl group in a 70:30 ratio. Later, the Kass group noted that the observed gas-phase structure depended on the solvent used in the electrospray ionization process.[41] Use of methanol as the solvent led to the phenoxide anion structure in the gas phase whereas use of acetonitrile led to the carboylate anion structure. Subsequently, Oomens and co-workers[42] used gas-phase infrared multiple photon dissociation (IRMPD) spectroscopy to study the structure of the anions of aspartic, cysteine, glutamic, phenylalanine, serine, tryptophan, and tyrosine amino acids. They found that the carboxylate structure is preferred for all of these anions. In the specific case of tyrosine, they showed that the structure was not dependent on the nature of the solvent used for the electrospray, in contrast to the results of the Kass group. Gutowski and co-workers[43,44] used a simple genetic algorithm to identify the lowest energy conformers of canonical (neutral), zwitterionic, and protonated arginine at the MP2 and DFT levels.

We have predicted the gas-phase acidities of the normal amino acids at the G3MP2 level. Self-consistent reaction field (SCRF) calculations[45] with the COSMO parametrization[46] were used to predict the pK_a's in aqueous solution. The heats of formation of the neutral compounds were calculated from the G3MP2 atomization energies and as well as isodesmic reactions using glycine to provide the first reliable set of these values for the amino acids in the gas phase. In addition, the same quantities for the five so-called[47] common rare amino acids were calculated.

COMPUTATIONAL METHODS

The calculations were performed at the density functional theory (DFT), and correlated molecular orbital (MO) theory levels with the programs Gaussian 03 and 09.[48] The geometries were initially optimized at the DFT level with the B3LYP exchange-correlation functional[49,50] and the DZVP2 basis set.[51] Vibrational frequencies were calculated to show that the structures were minima and to provide zero-point and thermal corrections to the enthalpy and entropies so that free energies could be calculated for direct comparison to experiment. We optimized a range of structures for the 17 amino acids not studied by us previously as well as the rare amino acids to determine the most stable structures. This followed on the work of Marynick and co-workers on the structures of glutamic acid.[52] All amino acids were studied in their L stereoisomer. Protons were removed from different sites on the main chain and the side chain of different conformers of the neutral structures to provide starting structures and the resulting anion structures were then optimized. A range of structures for each of the anions was studied. The main sites of deprotonation were on oxygen, nitrogen, and sulfur substituents, and some C−H deprotonation was also examined. Our conformational sampling was performed based on the likely chemical structures. Although the structures of the neutral differ in terms of the conformation of the CO_2H group from those used in the proton affinity study, they are similar enough that we can exploit the previous conformational sampling.[27]

We have shown[36] that calculations at the MP2/CBS (CBS = complete basis set) level with the augmented correlation-consistent basis sets up through the quadruple level[53] predicted the acidities of organic acids to better than 4 kcal/mol with the calculated values more acidic than the experimental values. In our previous work, it was shown that the G3(MP2) method[54] improved the agreement for the acidities with the experimental values and/or the coupled cluster CCSD(T)/CBS values to within about 1 kcal/mol. For example, the G3MP2 value for GA(CH_3CO_2H) is 340.3 kcal/mol at 298 K, the experimental value is 341.5 ± 2.0 kcal/mol,[55] and the MP2/CBS value is 337.2 kcal/mol. Because of our previous benchmarks, we used the G3MP2 approach to calculate the gas-phase acidities for the amino acids. G3MP2 has an additional advantage over DFT methods in terms of reliable predictions for these types of compounds as the correlated molecular orbital methods in G3MP2 perform better in the prediction of hydrogen bond energies as well as steric nonbonded interactions than do most widely used DFT exchange-correlation functionals. The G3MP2 structures chosen were the lowest in term of the free energy at 298 K. The structures of the neutral compounds differ in some cases from those previously reported,[27] mostly in terms of the orientation of the hydrogen on the CO_2H group. We recalculated all of the lowest energy G3MP2 structures from ref 27. which were obtained using Gaussian03 and the total energy values changed somewhat in Gaussian 09 as shown in the Supporting Information. The critical difference in the structures, which usually amounts to at most 1 kcal/mol, is where the hydrogen on the carboxylic group prefers to hydrogen bond. We find that about 50% of the time this hydrogen prefers to hydrogen bond to the carbonyl oxygen in the acid group and about 50% of the time to another site on the amino acid such as the terminal NH_2 group. The G3MP2 energy calculations use the MP2(full)/6-31G(d) geometry. We can compare this geometry to that optimized at the MP2/aug-cc-pVTZ level where the latter has diffuse functions for glycine. The lack of diffuse functions in the G3MP2 calculations does not affect the geometry parameters of the glycine anion by more than 0.005 Å. The values of the C−O, C−O(···H), and N−H···O bond

dx.doi.org/10.1021/jp207271p | J. Phys. Chem. B 2012, 116, 2905−2916

distances at the MP2(full)/6-31G(d) are 1.257, 1.263, and 2.121 Å, respectively, and the corresponding MP2/aug-cc-pVTZ values are 1.259, 1.268, and 2.119 Å.

Theoretical pK_a values in aqueous solution were calculated by combining the gas-phase acidities with single point (at the optimum gas phase geometry) self-consistent reaction field calculations[45] using the COSMO parametrization.[46] The Gibbs free energy for deprotonation in aqueous solution (ΔG_{aq}) was calculated from the gas-phase free energy and the aqueous solvation free energy. The solvation energy is calculated as the sum of the electrostatic energies (polarized solute−solvent) and the nonelectrostatic energies. A dielectric constant of 78.39 corresponding to that of bulk water was used in the COSMO calculations at the B3LYP/aug-cc-pVDZ level using the gas-phase geometries obtained at this level. The pK_a values in aqueous solution were calculated using eq 1

$$pK_a' = pK_a(HA) + \Delta G_{aq}/(2.303RT) \quad (1)$$

where ΔG_{aq} is the solution free energy, R is the gas constant, and T = 298 K is the temperature. We report our pK_a values relative to the well-established value for acetic acid (HA) with pK_a = 4.76 to minimize errors in the prediction of pKa's.[36]

The heat of formation of glycine was calculated by using a composite approach[56] developed in our group and Washington State University, which is based on CCSD(T) calculations[57−60] extrapolated to the complete basis set limit using the correlation consistent basis sets. Single point CCSD(T) calculations were performed with the aug-cc-pVnZ basis sets for n = D, T, Q at the MP2/aug-cc-pVTZ geometry. The open-shell atomic calculations were done with the R/UCCSD(T) approach where a restricted open shell Hartree−Fock (ROHF) calculation was initially performed and the spin constraint was then relaxed in the coupled cluster calculation.[61] The CCSD(T) energies with n = D, T, Q were extrapolated to the complete basis set (CBS) limit using a mixed Gaussian/exponential formula (eq 2)[62]

$$E(n) = E_{CBS} + A\ \exp[-(n-1)] + B\ \exp[-(n-1)^2] \quad (2)$$

The cardinal numbers for the n = D, T, Q basis sets are 2, 3, and 4. Core−valence correlation corrections (ΔE_{CV}) were calculated at the CCSD(T) level with the aug-cc-pVTZ basis set for H, the aug-cc-pwCVTZ basis set for the first row atoms.[63] Scalar relativistic corrections (ΔE_{SR}) were calculated as the expectation values of the mass-velocity and Darwin operators (MVD) from the Breit−Pauli Hamiltonian[64] for the CISD (configuration interaction with single and double excitations) wave function with the aug-cc-pVTZ basis set. The atomic spin−orbit corrections (SO) were calculated from the experimental values for the ground states of the atoms.[65] The total atomization energies (TAEs) at 0 K were calculated from eq 3

$$\Sigma D_{0,0K} = \Delta E_{CBS} + \Delta E_{Rel} + \Delta E_{CV} + \Delta E_{ZPE} + \Delta E_{SO} \quad (3)$$

The molecular heat of formation at 0 K were calculated from the calculated TAE and the heats of formation of the atoms at 0 K (58.98 ± 0.02 kcal/mol for O, 169.98 ± 0.1 kcal/mol for C, 112.53 ± 0.02 kcal/mol for N, and 51.63 ± 0.001 kcal/mol for H).[66] Heats of formation at 298 K were calculated by following the procedures outlined by Curtiss et al.[67]

The heats of formation of the amino acids were calculated from atomization energies at the G3MP2 level and from a set of isodesmic reactions at the G3MP2 level using the calculated heat of formation of glycine.

RESULTS AND DISCUSSION

Gas-Phase Acidities of the Common Amino Acids: Structures. The most stable neutral structures for the L-forms of the amino acids alanine, arginine, asparagine, cysteine, glutamine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine, and their anions are shown in Figure 1. Important hydrogen bond distances are shown. Amino acids with more than one anion structure are designated with capital letters, with [A] being the most acidic structure, and amino acids with more than one neutral structure are designated with lower case letters. The acidities are summarized in Table 1. As previously reported,[34] glutamic acid, aspartic acid, and glycine all prefer to lose the proton from the main chain CO_2H group to form the COO^- anion. All anions exhibit strong hydrogen bonding and all calculated acidities fall within the error bars of the experimental numbers in literature.

Protons were removed from the main chain carboxylic acid as well as selected sites on the backbone or the side chain substituent. For alanine, removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion. Removal of a proton from the methyl side chain or the main chain NH_2 groups led to proton transfer to generate less stable carboxylate ions. For asparagine, removal of a proton from the main chain NH_2 group gave a carboxylate ion which was more stable than the one formed by removal of a proton from the main chain CO_2H group. For glutamine, removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion. Removal of a proton from the main chain NH_2 group led to proton transfer to generate a less stable carboxylate ion. For isoleucine, removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion. Removal of a proton from the main chain NH_2 group and the side chain methyl group led to proton transfer to generate less stable carboxylate ions. For leucine, all sites of deprotonation are shown; removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion. For lysine, two stable neutral structures were found: a and b. Lysine neutral [a] led to the most stable ΔG_{298} anions, and lysine neutral [b] led to the most stable ΔH_{298} anions. Removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion. Removal of a proton from the side chain NH_2 led to proton transfer to generate a less stable carboxylate ion. For methionine, removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion. Removal of a proton from the main chain NH_2 led to proton transfer to generate a less stable carboxylate ion. For phenylalanine, all sites of deprotonation are shown. Removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion. For proline, removal of a proton from the NH in the ring led to proton transfer to generate the most stable carboxylate ion conformer. For serine, removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion. Removal of a proton from main chain NH_2 and the side chain OH, led to proton transfer to generate less stable carboxylate ions. For threonine, two stable neutral structures were found: a and b. Threonine neutral [a] led to the most stable ΔG_{298} anions, and threonine neutral [b] led to the most stable

dx.doi.org/10.1021/jp207271p | J. Phys. Chem. B 2012, 116, 2905−2916

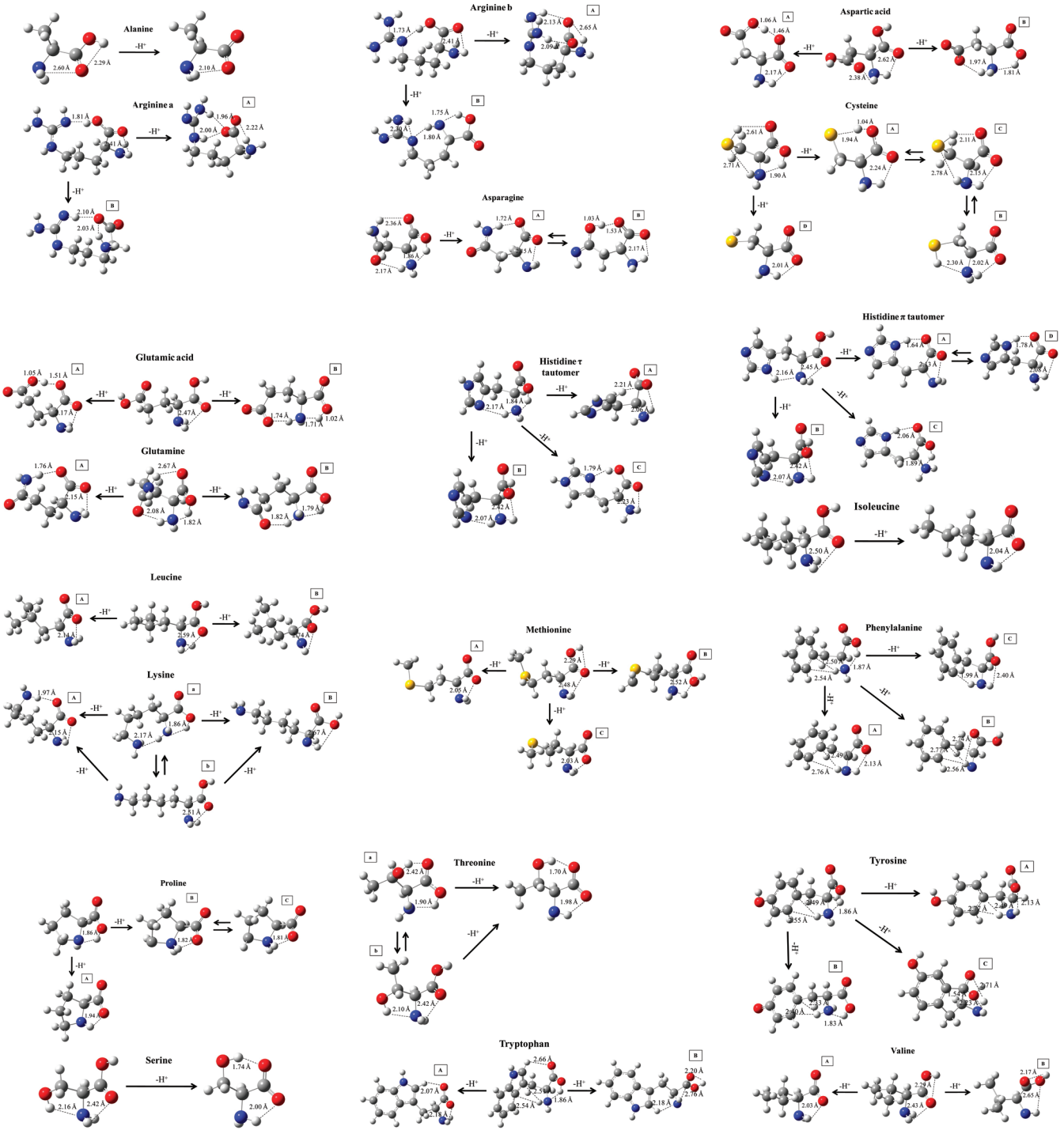


Figure 1. Optimized structures of the amino acids and their lowest energy anions at the G3MP2 (MP2(full)/6-31G(d)) level. Important hydrogen bond distances are given in Å.

ΔH_{298} anions. Removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion. Removal of a proton from the main chain NH_2 and the side chain OH, led to proton transfer to generate less stable carboxylate ions. For tryptophan, all sites of deprotonation are shown. For valine, all sites of deprotonation are shown. Removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion.

For neutral histidine, only one imidazolium N has a hydrogen, so two tautomeric forms exist: the Nε2-protonated τ tautomer and the Nδ1-protonated π tautomer. We predict the neutral τ tautomer to be more stable than the neutral π tautomer with $\Delta H(298)$ = 1.0 kcal/mol and $\Delta G(298)$ = 0.7 kcal/mol. Thus, there is 23% π tautomer and 77% τ tautomer at equilibrium in the gas phase at 298 K. In the more stable neutral τ tautomer, there is a strong hydrogen bond between the main chain NH_2 and the N in the imidazole ring (r(N···H) = 2.17 Å). Deprotonation of the carboxylate group of the neutral π tautomer led to the most stable anion with partial proton transfer from the imidazole ring N−H to the CO_2^- group. The lowest energy structure of the anion has an N−H distance of 1.06 Å and the higher energy conformer has an

dx.doi.org/10.1021/jp207271p | J. Phys. Chem. B 2012, 116, 2905−2916

Table 1. G3MP2 Acidities of the Amino Acids in kcal/mol

amino acid	ΔH_{298} gas	ΔG_{298} gas	ΔH_{298} calc[38]	ΔH_{298} gas expt[38]	ΔH_{298} gas expt[33]
alanine	342.3	334.6	342.3	341.8 ± 1.9	340.6 ± 2.1[28]
arginine a A	330.1	322.0	331.5	330.1 ± 2.2	332.0 ± 3.1
arginine a B	340.1	331.8			
arginine b A	340.2	333.1			
arginine b B	353.7	346.2			
asparagine A	331.3	323.5	330.8	331.0 ± 2.2	331.7 ± 3.1
asparagine B	339.9	332.7			
aspartic acid A[34]	322.4	315.4	321.5	321.5 ± 3.3	320.3 ± 1.4[37]
aspartic acid B[34]	333.1	325.9			
cysteine A	334.4	327.1	333.7[30]	333.4 ± 2.2[30]	332.9 ± 3.1
cysteine B	336.6	328.2			
cysteine C	336.3	328.6			
cysteine D	338.6	330.2			
glutamic acid A[34]	321.9	316.4	322.4	322.2 ± 5.0	322.7 ± 1.4[37]
glutamic acid B[34]	330.8	324.3			
glutamine A	328.0	321.9	329.3	331.0 ± 2.6	331.7 ± 3.1
glutamine B	351.9	345.8			
glycine[34]	342.9	335.3	342.7	342.7 ± 2.2	342.5 ± 2.1[28]
histidine τ tautomer A	342.4	334.6			
histidine τ tautomer B	343.6	336.4			
histidine τ tautomer C	380.9	372.5			
histidine π tautomer A	328.9	321.4	328.4	328.6 ± 2.6	331.0 ± 3.1
histidine π tautomer B	346.0	338.5			
histidine π tautomer C	376.0	368.5			
histidine π tautomer D	332.0	325.0			
isoleucine	340.8	333.6	340.8	340.1 ± 1.9	338.9 ± 3.1
leucine A	340.3	333.5	341.3	339.1 ± 2.4	339.1 ± 3.1
leucine B	384.7	377.5			
lysine A WRT neutral a	336.9	329.9	338.2	338.4 ± 1.7	337.5 ± 3.1
lysine B WRT neutral a	385.5	376.8			
lysine A WRT neutral b	336.0	330.7			
lysine B WRT neutral b	384.7	377.6			
methionine A	338.8	331.1	337.5	336.3 ± 2.2	335.8 ± 3.1
methionine B	392.8	385.0			
methionine C	339.0	331.3			
phenylalanine A	338.5	330.7	338.7	338.9 ± 4.3	336.5 ± 3.1
phenylalanine B	384.4	376.4			
phenylalanine C	390.3	382.5			
proline A	340.5	333.2	341.8	342.0 ± 2.2	341.8 ± 3.1
proline B	386.9	379.5			
proline C	388.5	381.1			
serine	332.6	325.7	332.7	332.5 ± 5.3	332.7 ± 3.1
threonine WRT neutral a	332.3	324.8	333.9	331.7 ± 2.4	332.2 ± 3.1
threonine WRT neutral b	332.0	325.0			
tryptophan A	340.1	332.1	339.9	339.6 ± 2.2	337.0 ± 3.1
tryptophan B	383.7	375.6			
tyrosine A	338.3	330.4	339.1	337.7 ± 2.6	336.5 ± 3.1
tyrosine B	340.5	332.9			
tyrosine C	358.9	353.2			
valine A	340.6	333.2	341.8	342.0 ± 1.9	339.4 ± 3.1
valine B	385.5	378.4			

N−H bond distance of 1.04 Å consistent with less proton transfer and a higher energy. Removal of a proton from the NH in the imidazole ring of the π tautomer led to a structure with strong hydrogen bonding between the main chain NH_2 and the N in the imidazole ring (2.07 Å); this structure was ∼17 kcal/mol less stable than the most stable carboxylate ion conformer. Deprotonation of the CO_2H group in the τ tautomer led to a less stable conformer of the carboxylate ion (∼13 kcal/mol) than the same deprotonation in the less stable π tautomer neutral. Deprotonation of the side chain CH_2 in the τ tautomer led to a CH^- anion with strong hydrogen bonding between CO_2H and the N in the ring (r(N···H) = 1.79 Å); this structure is ∼40 kcal/mol higher in energy than the most stable carboxylate ion conformer derived from the π tautomer. Thus it is likely that the observed gas phase anion is derived from the less stable neutral π tautomer.

dx.doi.org/10.1021/jp207271p | J. Phys. Chem. B 2012, 116, 2905−2916

Table 2. Calculated Acidities of Tyrosine and Cysteine in kcal/mol at Different Computational Levels

amino acid	prop	B3LYP	G3MP2[35]	G3B3	G3[35]	G4	CCSD(T) /aT//B3LYP	MP2/aT	CCSD(T) /aT//MP2/aT
Tyr CO_2^-	ΔH	338.9	338.3	338.3	338.1	338.3	338.3[a]	335.8	338.2[a]
	ΔG	331.1	330.4	330.5	330.2	330.6	330.6[a]	328.1	330.6[a]
Tyr O^-	ΔH	339.1	340.5	340.3	340.5	340.1	340.1[a]	337.1	340.1[a]
	ΔG	331.4	332.9	332.6	332.9	332.3	332.2[a]	329.2	332.3[a]
	$\Delta\Delta H$	0.2	2.2	2.0	2.4	1.8	1.8	1.3	1.9
	$\Delta\Delta G$	0.3	2.5	2.1	2.7	1.7	1.6	1.1	1.7
Cys COO^-	ΔH	335.8	336.6	336.4	336.3	336.3	336.6[b]	333.7[c]	336.3[c]
anion [B]	ΔG	327.7	328.2	328.3	328.0	328.4	328.2[b]	325.4[c]	328.1[c]
Cys COO^-	ΔH	336.4	336.3	336.6	336.0	d	336.5[b]	333.2/333.3[c]	336.1/336.1[c]
anion [C]	ΔG	328.7	328.6	329.0	328.3	d	328.8[b]	325.6/325.6[c]	328.3/328.4[c]
[B] - [C]	$\Delta\Delta H$	−0.6	0.3	−0.2	0.3	d	0.1	0.4[c]	0.2[c]
[B] - [C]	$\Delta\Delta G$	−1.0	−0.4	−0.7	−0.3	d	−0.6	−0.2[c]	−0.3[c]
cysteine S^-	ΔH	333.8	334.4	333.2	334.0	333.6	334.1[b]	330.1/330.1[c]	332.9/333.0[c]
anion [A]	ΔG	326.8	327.1	326.1	326.7	326.7	326.8[b]	322.8/322.8[c]	325.6/325.7[c]
[A] - [C]	$\Delta\Delta H$	−2.6	−1.9	−3.4	−2.0	d	−2.5	−3.1/-3.2[c]	−3.2/-3.1[c]
[A] - [C]	$\Delta\Delta G$	−1.9	−1.5	−2.9	−1.6	d	−2.0	−2.8/-2.8[c]	−2.7/-2.7[c]
[A] - [B]	$\Delta\Delta H$	−2.0	−2.2	−3.2	−2.3	d	−2.6	−3.6[c]	−3.6[c]
[A] - [B]	$\Delta\Delta G$	−0.9	−1.1	−2.2	−1.3	d	−1.4	−2.6[c]	−2.4[c]

[a]Thermal corrections, entropies and zero-point energies from the G4 calculations. [b]Thermal corrections, entropies and zero-point energies from the G3 calculations. [c]Thermal corrections, entropies and zero-point energies at the MP2/aug-cc-pVDZ level. [d]Only one structure could be obtained at the G4 level.

As noted in the Introduction, different research groups have found different sites for deprotonation in tyrosine and cysteine. In our calculations for tyrosine, removal of a proton from the main chain NH_2, led to proton transfer to generate a carboxylate ion; removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion with ΔH_{298} = 338.3 kcal/mol. Removal of a proton from the side chain OH gave a phenoxide anion with a similar acidity, ΔH_{298} = 340.5 kcal/mol. Due to the significant questions raised in the literature about the structure of the lowest energy anion, we further explored the potential energy surface for tyrosine. As shown in Table 2, we used a variety of Gx[68,69] methods to predict the energy difference between the two anion sites. In all cases, the carboxylate ion is predicted to be more stable by 1.8 to 2.4 kcal/mol in terms of the enthalpy and by 1.7 to 2.7 kcal/mol on the free energy scale. In order to further benchmark these calculations, we calculated the relative energy at the CCSD(T) level with the aug-cc-pVDZ and aug-cc-pVTZ basis sets using the optimized MP2/aug-cc-pVTZ and B3LYP/DZVP2 geometries. Only the CCSD(T)/aug-cc-pVTZ results are given as this is a larger basis set at the CCSD(T) (or QCISD(T)) level than those in the Gx methods. The CCSD(T)/aug-cc-pVTZ results are very comparable to all of the Gx results. The quality of the CCSD(T)/aug-cc-pVTZ electronic energy difference for the two isomers should be better than ±1 kcal/mol. The T_1 diagnostic values[70] at the aug-cc-pVDZ/CCSD(T) level are 0.0139 for the tyrosine neutral, 0.0160 for the O^- anion and 0.0152 for the COO^- anion showing that these structures are dominated by a single reference configuration. There is no effect on the relative energies for the two different optimized geometries with all of the different methods that include a level of correlation beyond MP2. Thus, all of our results show that the carboxylate ion is more stable than the phenoxide anion by just under 2 kcal/mol. This value is consistent with the experimental IRMPD results.[42] This result suggests that the observations of Kass and co-workers involve kinetic control of their observed species. It is important to note that the B3LYP prediction of the energy difference between the two anion isomers is far too small, 0.2 to 0.3 kcal/mol, which may have biased the previous interpretations of the experimental data.[40,41] An earlier computational study[30] on the site of deprotonation at the B3LYP/aug-cc-pVDZ and G3B3 levels included additional correction factors taken as the difference between the calculated values and experiment for acetic acid and phenol. These authors found a larger difference for phenol than for acetic acid and used this correction factor to change the order of deprotonation to the phenol site. As shown in the Supporting Information (Table S5), the experimental data for acetic acid and phenol are not accurate enough to warrant the use of such a correction factor as the results depend on the choice of experimental data.

For cysteine, a proton is shared between the S and an O on the CO_2 group in the anion. This was determined after extensive searching of the conformational space in the anion. The question of deprotonation site thus reduces to which atom does the proton prefer to be closest. The proton prefers to be closest to the CO_2 site so the result looks like deprotonation from the SH group (structure [A]). In the lowest energy structure at the MP2/aug-ccpVTZ level (the B3LYP/DZVP2 geometry is very similar), the O−H distance is 1.05 Å, which is elongated from a normal CO_2H distance by 0.06 Å; the corresponding hydrogen bond S−H distance is 1.90 Å. In the anion structure where the H is bonded to the S (structure [C]), the S−H distance is 1.35 Å as compared to a normal S−H distance of 1.34 Å; the SH...O hydrogen bond distance is 2.14 Å. Thus there is little transfer of the proton toward the COO^- when it is bonded to S and, in this arrangement of the atoms, the S holds the H more tightly than does the O. A different lowest energy structure for the carboxylate ion was reported by Kass and co-workers[30] where there was not a hydrogen between the SH and the O and instead the SH is hydrogen bonded to the NH_2 group. In this carboxylate ion (structure [B]), the S−H distance is 1.34 Å, the SH...N hydrogen bond distance is 2.35 Å, and the NH_2···O hydrogen bond distance is 2.02 Å. The results in Table 2 show that the energies of cysteine-[B} and cysteine-[C] are very close.

dx.doi.org/10.1021/jp207271p | J. Phys. Chem. B 2012, 116, 2905−2916

The free energy favors [B] over [C] by −0.2 to −0.7 kcal/mol for the correlated molecular orbital theory methods. The structures obtained using B3LYP predict [B] to be more stable than [C] in terms of ΔH_{298} whereas the other methods predict that [C] is more stable in terms of ΔH_{298}, including CCSD(T)/aug-cc-pVTZ. The energy differences are so small as to make it difficult to determine the most stable anion with the negative charge localized on the carboxylate group. In summary, the most stable anion has the anion mostly localized on the S and all four anions have energies within 4 kcal/mol of each other. The T_1 diagnostic values at the aug-cc-pVDZ/CCSD(T) level (aug-cc-pVTZ/CCSD(T) in parentheses) are 0.0149 (0.0138) for the cysteine neutral, 0.0159 (0.0145) for the S^- anion, 0.0172 (0.0157) for the COO^- anion [B], and 0.0170 (0.0155) for the COO^- anion [C], again showing that a single reference wave function is dominant. Our results consistently show that the sulfur anion is ∼2 kcal/mol more stable than the carboxylate ion, which is consistent with other experiments but not with the IRMPD results.[42] The interpretation of the IRMPD spectra could be complicated if the proton is partially transferred and potentially shuttling between the S^- and the COO^- sites. We did search for a transition state between the two structures but this was only found at the Hartree−Fock level. We were unable to converge to a meaningful transition state at the aug-cc-pVDZ/MP2 or aug-cc-pVTZ/MP2 levels. This suggests that this region of the potential energy surface is very anharmonic and a detailed study of this region, which is beyond our current scope, is needed to resolve the differences in the various experimental interpretations.

For arginine, the two lowest energy neutral structures, previously determined by Gutowski and co-workers,[43,44] and their associated anions were studied: a and b. In the more stable neutral [b], there is a strong hydrogen bond between the main chain CO_2H and main chain NH_2 (1.73 Å) and between the main chain NH_2 and the side chain N (1.73 Å). Deprotonation of a proton from the main chain CO_2H group of the neutral [a] led to the most stable conformer of the carboxylate ion. Removal of a proton from the main chain NH_2 group, the side chain NH amino group, and the side chain terminal NH_2 group led to structures similar to that of abstraction of a proton from the main chain CO_2H group. For neutral [b], removal of a proton from the main chain CO_2H group led to a less stable conformer of the carboxylate ion (∼11 kcal/mol) than the same deprotonation in the less stable neutral [a]. Deprotonation of the end terminus NH_2 in neutral [b] led to a NH^- anion with strong hydrogen bonding between the main chain CO_2H and main chain NH_2 (1.75 Å) and between the main chain NH_2 and the side chain N (1.80 Å).

Gas-Phase Acidities of the Common Amino Acids: Energetics. We first compare our calculated acidities with those of Poutsma and co-workers[38] at the B3LYP level. Good agreement was found for all of the amino acids. We note that some of our conformers are of lower energy than those from the B3LYP study. For example, rotation of the side chain in the valine and threonine carboxylic acid anions led to more stable anions and hence an enhanced acidity. In the valine and threonine anions, rotating the side chain at the terminal C led to a slight increase in the stability of the anion. In the lysine anion, rotating the side chain to allow the end terminus NH_2 to bond to the COO^- resulted in a more stable structure.

Excellent agreement is found with the available experimental gas-phase deprotonation enthalpies and the calculated values are all within the experimental error bars. Our results confirm the amino acid GA scale and suggest that the reported experimental error bars are larger than they need to be. For example, the reported error bars for phenylalanine and serine can be reduced to less than ±2 kcal/mol.

The lowest energy structures of the parent neutral and the most stable anions generated by deprotonation are all stabilized by hydrogen bonds with O···H(O) hydrogen bond distances of no more than ∼2 Å. In all cases but asparagine, phenylalanine, proline, and tyrosine, the shortest hydrogen bond length is found in the lowest energy structure for the anion, thereby stabilizing the most acidic site. In asparagine, a proton is partially transferred in the second most stable anionic conformer/isomer so that it is shared between two oxygen molecules with r(OH) = 1.03 Å and r(O···H(O)) = 1.53 Å. In phenylalanine, the shortest hydrogen bond by about 0.15 Å is found in the least stable anionic conformer. In phenylalanine, the shortest hydrogen bond is between NH_2 and C in the ring (r(O···H(O)) = 2.02 Å) in the least acidic anion. In asparagine, cysteine, and the histidine π tautomer, a proton is partially transferred in the most stable anionic conformer. In asparagine, the proton is shared between O and O (r(OH) = 1.03 Å and r(O···H(O)) = 1.53 Å). In cysteine, the proton is shared between S and O (r(OH) = 1.04 Å and r(O···H(O)) = 1.94 Å). In histidine π tautomer, the proton is shared between N in the ring and O (r(OH) = 1.06 Å and r(O···H(O)) = 1.64 Å).

In general, in the most acidic structure, the proton is lost from the CO_2H acid group at the C-terminus to form the CO_2^- anionic group. Removal of a proton from the CO_2H group results in gas-phase acidities (ΔG_{acid}) between 329 and 334 kcal/mol showing that substituent effects are not very large except for arginine, asparagine, cysteine, glutamine, histidine, serine, and threonine where strong hydrogen bonds form with the side chain groups leading to lower gas-phase acidities between 320 and 327 kcal/mol (i.e., stronger gas-phase acids). In all cases but cysteine, the most acidic site was generated by removal of the proton from the CO_2H group. In cysteine, the proton is shared between the side-chain S and CO_2^- groups and is closer to the S. In arginine [A] and asparagine [A], the CO_2^- group is stabilized by hydrogen bonding by the main and side chain NH_2 groups. In the histidine π tautomer [A] and [B] and glutamine, hydrogen bonds with the side chain NH stabilize the CO_2^- group. In serine and threonine, the CO_2^- group is stabilized by hydrogen bonds with the side chain OH groups. In proline [B] and [C], deprotonation of a CH_2 in the ring leads to more positive gas-phase acidities, i.e., lower acidity. In most cases, loss of the proton from the most stable structure of the neutral leads to the most stable structure of the anion except for arginine, glutamine, histidine, lysine, and serine, where rotations in the side chain result in greater bonding and stability.

Gas-Phase Acidities of the Rare Amino Acids. We studied the so-called "rare" amino acids which involve methyl and hydroxyl substitution and hydroxyl elimination at the side chain. Figure 2 shows the most stable structures for neutral 3-methylhistidine, 5-hydroxylysine, *N*-methyllysine, 4-hydroxyproline, pyroglutamic acid, and their anions respectively. The values of the acidities and corresponding unsubstituted normal amino acid acidities are given in Table 3.

For 3-methylhistidine, removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion. Removal of a proton from the main chain NH_2 or the methyl group led to proton transfer to generate the carboxylate ion. For 5-hydroxylysine, removal of a proton from the main chain CO_2H group led to the a more stable conformer

of the carboxylate ion. Removal of a proton from side chain NH_2 led to proton transfer to generate an O^- anion on the side chain which is of higher energy. For *N*-methyllysine, removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion. Removal of a proton from the NH in the ring led to proton transfer to generate the carboxylate ion. For 4-hydroxyproline, all sites of deprotonation are shown. Removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion. For pyroglutamic acid, removal of a proton from the main chain CO_2H group led to a more stable conformer of the carboxylate ion. Removal of a proton from the ring led to proton transfer to generate the carboxylate ion.

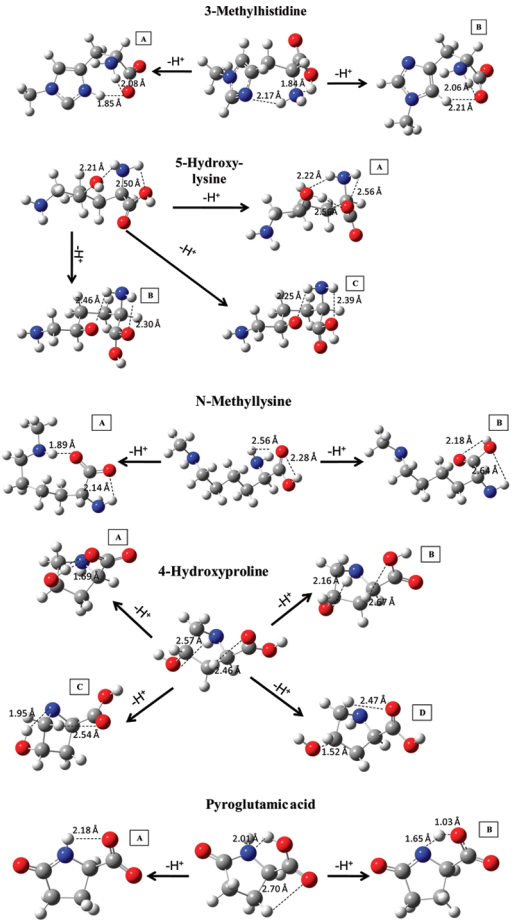


Figure 2. Optimized structures of the five "rare" amino acids and their lowest energy anions at the G3MP2 level. Important hydrogen bond distances are given in Å.

The effect of methyl group substitution to form 3-methylhistidine is small, leading to a GA that is 0.6 kcal/mol larger than that of histidine. The effect of the methyl group substitution to form *N*-methyllysine has no change in the GA with respect to lysine neutral b. Substitution of an OH group for H in lysine and proline to generate 5-hydroxylysine and 4-hydroxyproline, respectively, blocks the site from which the proton is removed in the parent. Thus, the GA reverts to that of the CO_2H group and values of ~335 kcal/mol are found consistent with the higher end of the amino acid acidity scale for removal of a proton from the carboxylic acid site. The cyclization of glutamic acid to produce pyroglutamic acid leads to a GA that is 4.4 kcal/mol larger than that of glutamic acid.

Slightly longer r(O···H(O)) hydrogen bond lengths for the shortest hydrogen bond were found in all of the neutral and anionic structures for the rare amino acids (~2.5 Å). In all cases but N-methyllysine, the shortest hydrogen bond length is found in the least stable anionic conformer. In *N*-methyllysine, the shortest hydrogen bond length is found in the most stable anionic conformer. In pyroglutamic acid, a proton is shared between two atoms in the least stable anionic conformer/isomer. The proton is shared between N in the ring and O (r(OH) = 1.03 Å and r(O···H(O)) = 1.65 Å).

Amino Acid pK_a's in Aqueous Solution. The calculated and experimental pK_a's are summarized in Table 4. The amino acids in the gas phase are not zwitterions (there has been some discussion as to whether arginine is a zwitterions in the gas phase but recent analyses suggest that it is not[71−74]) whereas they are zwitterions in solution. We first compare the side chain pK_a's of the amino acids with neutral ionizable side chains including cysteine, aspartic acid, glutamic acid, and tyrosine. The calculated pK_a's of the side chain S^- in cysteine [A], side chain O^- in tyrosine [B], and the CO_2^- side chains in aspartic acid and glutamic acid are in agreement with the experimental[76] pK_a's within ~1 pK_a unit. This suggests that the zwitterionic form of the backbone is not strongly affecting the side chain pK_a's. Similar good agreement for the side chain pK_a's was found for cysteine, aspartic acid and tyrosine by Sprik and co-workers.[75]

The predicted pK_a's obtained using the self-consistent reaction field approach with the COSMO parametrization and the gas phase nonzwitterion structures can be compared to the experimental[76] values for the zwitterions in aqueous

Table 3. Acidities of the 5 Rare Amino Acids in kcal/mol

amino acid	ΔH_{298}	ΔG_{298}	ΔH_{298} parent	ΔG_{298} parent	pK_a(298) calc COSMO	pK_a(298) calc COSMO-RS
3-methylhistidine A	343.1	335.2	341.0	333.2	5.94	6.11
3-methylhistidine B	361.7	354.8			22.48	22.31
5-hydroxylysine A	331.9	324.7	337.3	331.3	2.23	2.75
5-hydroxylysine B	351.8	345.3			9.26	9.42
5-hydroxylysine C	353.0	347.9			16.55	17.70
N-methyllysine A	336.5	330.7	337.3	331.3	6.24	7.46
N-methyllysine B	385.5	378.3	384.8	377.7	34.11	34.45
4-hydroxyproline A	334.9	328.7	341.5	333.6	0.93	2.00
4-hydroxyproline B	365.2	358.3			22.51	23.38
4-hydroxyproline C	372.3	365.9			33.76	34.64
4-hydroxyproline D	393.2	385.9			46.47	46.60
pyroglutamic acid A	328.3	320.8	321.7	315.3	2.55	2.59
pyroglutamic acid B	338.2	331.1			14.06	14.14

dx.doi.org/10.1021/jp207271p | *J. Phys. Chem. B* 2012, 116, 2905−2916

solution. These values correspond to loss of a proton from the main chain ammonium group in the neutral zwitterionic from ($NH_3^+CHRCO_2^-$) leading to the same anion as in our case starting from NH_2CHRCO_2H. The differences in the pK_a values can be used to estimate the free energy difference between the zwitterion and nonzwitterion forms in solution. The higher pK_a corresponds to the more stable species, the zwitterion. The calculations predict that the zwitterionic and nonzwitterionic forms of arginine are within better than 1 kcal/mol of each other. A number of amino acids (asparagine, lysine, phenylalanine, proline, and tyrosine) have the nonzwitterionic structure about 5 kcal/mol higher in free energy than the zwitterion. The remaining amino acids have the nonzwitterionic structures even higher in energy as compared to the zwitterions with the largest differences for aspartic and glutamic acid. The magnitude of the main chain carboxylate

Table 4. Calculated G3MP2 pK_a's of the Amino Acids in kcal/mol

amino acid	pK_a(298) calc COSMO	pK_a(298) calc COSMO-RS	pK_a(298) expt(1)[76]	ΔG(1) kcal/mol	side chain expt[76]	side chain calc[75]
alanine	4.43	4.77	9.868	7.4		
arginine A	8.88	9.71	8.991	0.2	12.1	
arginine B	6.63	7.22				
arginine A	15.32	16.32				
arginine B	27.71	27.78				
asparagine A	5.54	5.56	8.73	4.4		
asparagine B	12.33	13.38				
aspartic acid A[34]	0.75	1.48	10.002	11.4		
aspartic acid B	3.97	4.60			3.86/3.900	
cysteine A	9.59	9.43			8.00/8.36	10.7 ± 1.8
cysteine B	6.22	5.26				
cysteine C	6.74	6.80				
cysteine D	6.17	5.80	10.74	6.2		
glutamic acid A[34]	0.26	1.41	9.96	12.2		
glutamic acid B	2.78	3.88			4.07/4.30	
glutamine A	4.51	4.83	9.00	6.1		
glutamine B	18.81	18.86				
glycine[34]	4.89	5.26	9.778	6.7		
histidine τ tautomer A	5.07	6.98				
histidine τ tautomer B	15.10	16.69				
histidine τ tautomer C	40.23	43.75				
histidine π tautomer A	4.91	5.93				
histidine π tautomer B	15.10	15.18				
histidine π tautomer C	36.26	40.57				
histidine π tautomer D	4.17	5.47	9.28	6.9		
isoleucine	4.50	5.11	9.758	7.2		
leucine A	4.74	5.53	9.744	6.8		
leucine B	33.64	33.92				
lysine A WRT neutral a	6.13	6.84	9.07		10.82	
lysine B WRT neutral a	37.65	37.79				
lysine A WRT neutral b	5.75	7.27		4.5		
lysine B WRT neutral b	37.28	38.22				
methionine A	3.15	4.86				
methionine B	41.51	42.17				
methionine C	2.17	4.32	9.08	9.4		
phenylalanine A	5.96	6.13	9.31	4.6		
phenylalanine B	38.01	38.60				
phenylalanine C	46.92	45.89				
proline A	6.84	6.85	10.640	5.2		
proline B	38.99	38.76				
proline C	41.78	41.52				
serine	2.82	3.49	9.209	8.7		
threonine WRT neutral a	3.62	3.76	9.100	7.5		
threonine WRT neutral b	2.44	3.07				
tryptophan A	4.69	7.56	9.33	6.3		
tryptophan B	38.70	38.76				
tyrosine A	6.15	5.84	8.67	3.4		5.1 ± 1.0
tyrosine B	10.37	11.31			10.07/11.01	9.7 ± 1.8
tyrosine C	24.52	24.79				
valine A	4.36	4.80	9.719	7.3		
valine B	33.80	38.03				

dx.doi.org/10.1021/jp207271p | J. Phys. Chem. B 2012, 116, 2905−2916

Table 5. Calculated Heats of Formation of the Amino Acid in kcal/mol

amino acid	ΔH_f 0 K	ΔH_f 298 K G3MP2 TAE	ΔH_f 298 K calc[81]	isodesmic product	ΔH_f 298 K isodesmic rxn	ΔH_f 298 K experiment
alanine	−94.8	−100.7	−99.4	ethane	−100.2	−99.1 ± 1.0[78]
arginine a	−80.1	−93.0				
arginine b	−81.8	−94.6				
asparagine	−136.7	−144.4	−141.1	propanamide	−145.8	
aspartic acid	−182.1	−188.9	−188.0	propionic acid	−189.6	
cysteine	−88.0	−94.4	−90.4	ethylthiol	−93.6	
glutamic acid	−186.4	−194.5	−193.0	butanoic acid	−195.0	
glutamine	−139.6	−148.6	−146.1			
glycine	−87.4	−91.9	−93.2			−93.3 ± 1.1[78]
histidine τ	−57.0	−65.8				
histidine π	−56.2	−64.8	−53.0			
isoleucine	−107.8	−117.9	−116.3	isopentane	−117.6	
leucine	−107.9	−118.1	−116.3	isopentane	−117.7	−116.4 ± 1.2[66]
lysine a	−95.9	−107.9				
lysine b	−95.6	−107.0	−106.0			
methionine	−93.7	−102.5	−98.5	methylpropylthiol	−101.6	−98.8 ± 1.0[80]
phenylalanine	−68.1	−77.1	−72.2	ethylbenzene	−75.0	
proline	−83.0	−91.1	−89.2	*a*	−92.5	−87.5 ± 1.0[79]
serine	−131.7	−138.2	−135.7	ethanol	−138.2	
threonine a	−139.8	−147.9	−144.3	isopropanol	−147.9	
threonine b	−139.7	−147.6				
tryptophan	−49.5	−59.6	−51.9			
tyrosine	−109.7	−118.9	−115.2	*p*-ethylphenol	−117.2	
valine	−104.2	−112.9	−111.4	isobutane	−112.5	−108.8 ± 1.0[66]

[a]The products are pyrrolidine and acetic acid, not glycine.

pK_a's for the nonzwitterionic structures are consistent with the results of Coote and Ho[77] who used similar approaches to calculate the pK_a's of carbonic acids for biochemical applications. Sprik and co-workers[75] calculated the pK_a of the main chain carboxylate ion group for tyrosine using a DFT-based molecular dynamics approach with the inclusion of water molecules in the simulation and predict a pK_a of 5.1 ± 1.0.

The values of the calculated p*K*a's for the main chain carboxylate ion for the rare amino acid are summarized in Table 3. Assuming a pK_a of 9 for the zwitterionic form, we predict that the nonzwitterionic forms are higher in energy by ∼4 (*N*-methyllysine) to 11 kcal/mol (4-hydroxyproline) higher in energy than the corresponding zwitterion. The predicted pK_a's are within ∼1.5 p*K*a units of the corresponding unsubstituted amino acid for 3-methylhistidine, 5-hydroxylysine, and pyroglutamic acid showing a modest substituent effect on the pK_a. The effect of the substituent on the pK_a is larger for the other two acids.

Amino Acid Gas-Phase Heats of Formation. Heats of formation of the amino acids were calculated from the G3MP2 atomization energies and by an isodesmic reaction approach using glycine as a product and CH_4 as the reactant where possible (reaction 4).

$$H_2NCHRCO_2H + CH_4 \rightarrow H_2NCH_2CO_2H + CH_3R \quad (4)$$

This leads to the energy expression given in eq 5

$$\Delta H_f(\text{reaction}) = \Delta H_f(\text{product}) + \Delta H_f(\text{glycine}) + \Delta H_{rxn}(\text{MP2}) - \Delta H_f(CH_4) \quad (5)$$

where the heats of formation of the product (Table 5), glycine and CH_4 are taken from experiment. For comparison, the CCSD(T)/CBS heat of formation of glycine (ΔH_f^{298} = −92.6 kcal/mol) differs from experiment (−93.3 ± 1.1 kcal/mol)[78] by 0.7 kcal/mol and is within the experimental error bars; the G3MP2 value for glycine differs by 1.4 kcal/mol from experiment. The heats of formation are summarized in Table 5. The heats of formation from the atomization energies are within 4 kcal/mol of the available experimental values.[66,78−80] The calculated heats of formation from the isodesmic reactions are also within 4 kcal/mol of the available experimental values. Sagadeev et al.[81] predicted the heats of formation of the amino acids using a parametrized group additivity method. The group additivity values are within 4 kcal/mol of our values except for tryptophan and histidine. The group additivity approach predicts ΔH_f(histidine) to be too positive by ∼12 kcal/mol and ΔH_f(tryptophan) to be too positive by ∼8 kcal/mol.

Heats of formation of the five rare amino acids were calculated from the total atomization energies and provide the first good estimates of these values (Table 6). The heats of formation from isodesmic reactions are within ∼1 kcal/mol of the calculated heats of formation from the total atomization energies for 3-methylhistidine, 4-hydroxyproline, 5-hydroxylysine, and *N*-methylysine.

Table 6. Calculated Heats of Formation of the 5 Rare Amino Acids in kcal/mol

amino acid	ΔH_f 0 K	ΔH_f 298 K G3MP2 TAE	ΔH_f 298 K calc[81]	ΔH_f 298 K isodesmic rxn
3-methylhistidine	−59.7	−69.6		−68.5
4-hydroxyproline	−120.5	−128.8	−129.2	−129.4
5-hydroxylysine	−135.6	−147.7	−145.9	−148.3
N-methyllysine	−93.2	−106.1		−106.1
pyroglutamic acid	−124.9	−131.7		

dx.doi.org/10.1021/jp207271p | *J. Phys. Chem. B* 2012, 116, 2905−2916

CONCLUSION

We optimized a range of structures for 17 of the 20 common L-amino acids and 5 rare amino acids to determine the most stable structures which, in general, involve the strongest hydrogen bonding. Protons were removed from a variety of positions on the main chain and side chains of these acids. Excellent agreement is found with the available experimental gas-phase deprotonation enthalpies and our results confirm the amino acid GA scale and can be used to substantially reduce the error bars for these values. In general, the proton is lost from the $-CO_2H$ acid group to form the $-CO_2^-$ anionic group, except for cysteine where the proton is partially shared between the S^- and CO_2^- groups and is closer to the sulfur. We confirm that the proton is lost from the CO_2H group in tyrosine. Removal of a proton from the $-CO_2H$ group results in gas-phase acidities between 329 and 335 kcal/mol showing that substituent effects are not very large except for arginine [a], asparagine, glutamine, histidine π tautomer, serine, and threonine [a] and [b] where strong hydrogen bonds form with the side chain groups leading to lower gas-phase acidities between 321 and 326 kcal/mol. For aspartic acid and glutamic acid, the substituents lead to lower GA values (~316 kcal/mol). The differences in the pK_a values were used to estimate the free energy difference between the zwitterion and nonzwitterion forms in solution. Asparagine, lysine, phenylalanine, proline, and tyrosine have the nonzwitterionic structure about 5 kcal/mol higher in free energy than the zwitterions, whereas the zwitterionic and non-zwitterionic forms of arginine are within better than 1 kcal/mol of each other. Heats of formation from the total atomization energies (TAEs) and isodesmic heats of formation of the common amino acids are within 4 kcal/mol of the available experimental values for the TAEs and within 2 kcal/mol, except for methionine, proline, and valine, for the isodemic reaction approach.

ASSOCIATED CONTENT

Supporting Information

Cartesian coordinates of B3LYP/DZVP2 optimized geometries in angstroms, H_{298} and G_{298} values for all neutral amino acids and anions at the G3MP2 level, H_{298} and G_{298} of tyrosine and cysteine for the two sites at different computational levels, CCSD(T) total energies for tyrosine and cysteine, H_{298} and G_{298} values for all neutral amino acids and anions at the G3MP2 level, including those that differ from ref 27, and ΔH_{298} and ΔG_{298} acidity comparison for acetic acid and phenol. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dadixon@bama.ua.edu.

ACKNOWLEDGMENTS

This work is supported by NSF under CHE-0848470, which is funded by the American Recovery and Reinvestment Act (ARRA) and is jointly sponsored by NSF's Analytical & Surface and Experimental Physical Chemistry sections. This work is also supported by the Research Experiences for Undergraduates (REU) program funded by the National Science Foundation (NSF). This was awarded through the Summer Undergraduate Research Program (SURP) at the University of Alabama. D.A.D. thanks the Robert Ramsay Fund from the University of Alabama for its partial support and the Department of Energy, Office of Basic Energy Sciences, Geosciences Program.

REFERENCES

(1) Ewing, N. P.; Cassady, C. J. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 105−116.
(2) Jai-nhuknan, J.; Cassady, C. J. *Anal. Chem.* **1998**, *70*, 5122−5128.
(3) Brinkworth, C. S.; Dua, S.; Bowie, J. H. *Eur. J. Mass Spectrom.* **2002**, *8*, 53−66.
(4) MacLean, M. J.; Brinkworth, C. S.; Bilusich, D.; Bowie, J. H.; Doyle, J. R.; Llewellyn, L. E.; Tyler, M. *Toxicon Mass Spectrometry in Toxinology - A 21st-Century Technology for the Study of Biopolymers from Venoms* **2006**, *47*, 664−675.
(5) Marzluff, E. M.; Campbell, S.; Rodgers, M. T.; Beauchamp, J. L. *J. Am. Chem. Soc.* **1994**, *116*, 7787−7796.
(6) Harrison, A. G. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 1−13.
(7) Bowie, J. H.; Brinkworth, C. S.; Dua, S. *Mass Spectrom. Rev.* **2002**, *21*, 87−107.
(8) Bilusich, D.; Bowie, J. H. *Mass Spectrom. Rev.* **2009**, *28*, 20−34.
(9) McLuckey, S. A.; Cameron, D.; Cooks, R. G. *J. Am. Chem. Soc.* **1981**, *103*, 1313−1317.
(10) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F. *Mass Spectrom. Rev.* **1990**, *9*, 37−70.
(11) Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1991**, *63*, 1193A−1203A.
(12) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. *J. Mass Spectrom.* **2000**, *35*, 1399−1406.
(13) Cox, K. A.; Gaskell, S. J.; Morris, M.; Whiting, A. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 522−531.
(14) Johnson, R. S.; Martin, S. A.; Biemann, K. *Int. J. Mass Spectrom. Ion Proc.* **1988**, *86*, 137−154.
(15) Ghelis, C.; Yon, J. *Protein Folding*; Academic Press: New York, 1982.
(16) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.
(17) Margoulies, M.; Greenwood, F. C. *Structure-Activity Relationships of Protein and Polypeptide Hormones*; Excerpta Medica Foundation: Amsterdam, 1972.
(18) Cassady, C. J.; Carr, S. R.; Zhang, K.; Chung-Phillips, A. *J. Org. Chem.* **1995**, *60*, 1704−1712. Carr, S. R.; Cassady, C. J. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 1203−1210.
(19) Wu, J.; Lebrilla, C. B. *J. Am. Chem. Soc.* **1993**, *115*, 3270−3275.
(20) Gorman, G. S.; Speir, J. P.; Turner, C. A.; Amster, I. J. *J. Am. Chem. Soc.* **1992**, *114*, 3986−3988.
(21) Wu, Z.; Fenselau, C. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 403−405.
(22) Sun, W.; Kinsel, G. R.; Marynick, D. S. *J. Phys. Chem. A* **1999**, *103*, 4113−4117.
(23) Noguera, M.; Rodriguez-Santiago, L.; Sodupe, M.; Bertran, J. *J. Mol. Struct. THEOCHEM* **2001**, *537*, 307−318.
(24) Dinadayalane, T. C.; Sastry, G. N.; Leszczynski, J. *Int. J. Quantum Chem.* **2006**, *106*, 2920−2933.
(25) Bleiholder, C.; Shuai, S.; Paizs, B. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1275−1281.
(26) Harrison, A. G. *Mass Spectrom. Rev.* **1997**, *16*, 201−217.
(27) Gronert, S.; Simpson, D. C.; Conner, K. M. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2116−2123.
(28) Locke, M. J.; McIver, R. T. Jr. *J. Am. Chem. Soc.* **1983**, *105*, 4226−4232.
(29) Meot-Ner, M.; Hunter, E. P.; Field, F. H. *J. Am. Chem. Soc.* **1979**, *101*, 686−689.
(30) Tian, Z.; Poutsma, J. C.; Pawlow, A.; Kass, S. R. *J. Am. Chem. Soc.* **2007**, *129*, 5403−5407.
(31) Woo, H. K.; Lau, K. C.; Wang, X. B.; Wang, L. S. *J. Phys. Chem. A* **2006**, *110*, 12603−12606.
(32) Caldwell, G.; Renneboog, R.; Kebarle, P. *Can. J. Chem.* **1989**, *67*, 611−618.
(33) O'Hair, R. A. J.; Bowie, J. H.; Gronert, S. *Int. J. Mass Spectrom. Ion Proc.* **1992**, *117*, 23−36.

(34) Li, Z.; Velazquez, H. A.; Hernandez-Matus, M.; Dixon, D. A.; Cassady, C. J. *Int. J. Mass Spectrom.* **2007**, *265*, 213−223.
(35) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703−4709.
(36) Gutowski, K. E.; Dixon, D. A. *J. Phys. Chem. A* **2006**, *110*, 12044−12054.
(37) Fournier, F.; Afonso, Cl; Fagin, A. E.; Gronert, S.; Tabet, J-C *J. Am. Soc. Mass. Spectrom* **2008**, *19*, 1887−1897.
(38) 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. *Int. J. Mass Spectrom.* **2007**, *267*, 54−62.
(39) Miao, R.; Jin, C.; Yang, G.; Hong, J.; Zhao, C.; Zhu, L. *J. Phys. Chem. A* **2005**, *109*, 2340−2349.
(40) Tian, Z.; Wang, X.-B.; Wang, L.-S.; Kass, S. R. *J. Am. Chem. Soc.* **2009**, *131*, 1174−1181.
(41) Tian, Z.; Kass, S. R. *J. Am. Chem. Soc.* **2008**, *130*, 10842−10843.
(42) Oomens, J.; Steill, J. D.; Redlich, B. *J. Am. Chem. Soc.* **2009**, *131*, 4310−4319.
(43) Skurski, P.; Gutowski, M.; Barrios, R.; Simons, J. *Chem. Phys. Lett.* **2001**, *337*, 143−150.
(44) Rak, J.; Skurski, P.; Simons, J.; Gutowski, M. *J. Am. Chem. Soc.* **2001**, *123*, 11695−11707.
(45) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3093.
(46) Klamt, A.; Schüürmann, G. J. *Chem. Soc. Perkin Trans.* **1993**, *2*, 799−805.
(47) Lehninger, A. L. *Biochemistry*, 2nd ed.; Worth Publishers Inc.: New York, 1970.
(48) 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.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuna, 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.; Cioslowshi, 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.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
(49) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(50) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(51) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem.* **1992**, *70*, 560−571.
(52) Sun, W.; Kinsel, G. R.; Marynick, D. S. *J. Phys. Chem. A* **1999**, *103*, 4113−4117.
(53) Kendall, R. A.; Dunning, T. H. Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(54) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703−4709.
(55) Bartmess, J. E., Linstrom, P. J., Mallard, W. G., Eds.; *Negative Ion Energetics Data*; NIST Chemistry WebBook, NIST Standard Reference Database Number 69; National Institute of Standards and Technology: Gaithersburg MD, 2005; http://webbook.nist.gov.
(56) Feller, D.; Dixon, D. A. *J. Chem. Phys.* **2001**, *115*, 3484−3496. Ruscic, B.; Wagner, A. F.; Harding, L. B.; Asher, R. L.; Feller, D.; Dixon, D. A.; Peterson, K. A.; Song, Y.; Qian, X.; Ng, C.; Liu, J.; Chen, W.; Schwenke, D. W. *J. Phys. Chem. A* **2002**, *106*, 2727−2747. Dixon, D. A.; Feller, D.; Peterson, K. A. *J. Chem. Phys.* **2001**, *115*, 2576−2581. Pollack, L.; Windus, T. L.; de Jong, W. A.; Dixon, D. A. *J. Phys. Chem. A* **2005**, *109*, 6934−6938. Feller, D.; Peterson, K. A.; Dixon, D. A. *J. Chem. Phys.* **2008**, *129*, 204015−204046.
(57) Purvis, G. D. III; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910−1918.
(58) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479−483.
(59) Watts, J. D.; Gauss, J.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 8718−8733.
(60) Bartlett, R. J.; Musial, M. *Rev. Mod. Phys.* **2007**, *79*, 291−352.
(61) Deegan, M. J. O.; Knowles, P. *J. Chem. Phys. Lett.* **1994**, *227*, 321−326. Knowles, P. J.; Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1993**, *99*, 5219−5227. Rittby, M.; Bartlett, R. J. *J. Phys. Chem.* **1988**, *92*, 3033−3036.
(62) Peterson, K. A.; Woon, D. E.; Dunning, T. H. Jr. *J. Chem. Phys.* **1994**, *100*, 7410−7415.
(63) Woon, D. E.; Dunning, T. H. Jr. *J. Chem. Phys.* **1995**, *103*, 4572−4585. Peterson, K. A.; Dunning, T. H. Jr. *J. Chem. Phys.* **2002**, *117*, 10548−10560.
(64) Davidson, E. R.; Ishikawa, Y.; Malli, G. L. *Chem. Phys. Lett.* **1981**, *84*, 226−229.
(65) Moore, C. E. *Atomic Energy Levels as Derived from the Analysis of Optical Spectra*; *Vol. 1, H to V*; U.S. National Bureau of Standards Circular 467; U.S. Department of Commerce, National Technical Information Service: Washington, DC, 1949.
(66) Chase, M. W. Jr. NIST-JANAF *Thermochemical Tables*, 4th ed.; *J. Phys. Chem. Ref. Data, Mono. 9*; American Institute of Physics: Woodbury, NY, 1998; Suppl 1.
(67) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063−1079.
(68) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650−7657.
(69) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *126*, 84108−84108·12.
(70) Lee, T. J.; Rice, J. E.; Scuseria, G. E.; Schaefer, H. F. *Theor. Chim. Acta* **1989**, *75*, 81−89. Lee, T. J.; Taylor, P. R. *Int. J. Quantum Chem. Symp.* **1989**, *23*, 199−207. Lee, T. J. *Chem. Phys. Lett.* **2003**, *372*, 362−367.
(71) Price, W. D.; Jockusch, R. A.; Williams, E. R. *J. Am. Chem. Soc.* **1997**, *119*, 11988−11989.
(72) Forbes, M. W.; Bush, M. F.; Polfer, N. C.; Oomens, J.; Dunbar, R. C.; Williams, E. R.; Jockusch, R. A. *J. Phys. Chem. A* **2007**, *111*, 11759−11770.
(73) Bush, M. F.; Prell, J. S.; Saykally, R. J.; Williams, E. R. *J. Am. Chem. Soc.* **2007**, *129*, 13544−13553.
(74) O'Brien, J. T.; Prell, J. S.; Berden, G.; Oomens, J.; Williams, E. R. *Int. J. Mass Spec* **2010**, *297*, 116−123.
(75) Mangold, M.; Rolland, L.; Costanzo, F.; Sprik, M.; Sulpizi, M.; Blumberger. *J. Chem. Theory Comput.* **2011**, *7*, 1951−1961.
(76) Harris, D. C. *Quantitative Chemical Analysis*; W. H. Freeman: New York, 2007; Chapter 10, pp 180−183.
(77) Ho, J. M.; Coote, M. L. *J. Chem. Theory Comput.* **2009**, *5*, 295−306.
(78) Ngauv, S. N.; Sabbah, R.; Laffitte, M. *Thermochim. Acta* **1977**, *20*, 371−380.
(79) Sabbah, R.; Laffitte, M. *Bull. Soc. Chim. Fr.* **1978**, *1*, 50−52.
(80) Sabbah, R.; Laffitte, M. *Thermochim. Acta* **1978**, *23*, 192−195.
(81) Sagadeev, E. V.; Gimadeev, A. A.; Barabanov, V. P. *Russ. J. Phys. Chem. A* **2010**, *84*, 209−214.