

COMPUTATIONAL STUDIES OF THE FUNDAMENTAL THERMODYNAMIC
PROPERTIES OF AMINO ACIDS AND SMALL PEPTIDES

by

MICHELE LEIGH STOVER

DAVID A. DIXON, COMMITTEE CHAIR

CAROLYN J. CASSADY
DANIEL J. GOEBBERT
MICHAEL K. BOWMAN
C. HEATH TURNER

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Chemistry
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2015

Copyright Michele Leigh Stover 2015
ALL RIGHTS RESERVED

ABSTRACT

In 2011, the Human Proteome Project (HPP) was launched with the main goal of experimentally mapping the entire human proteome. Determining a protein's sequence is a first step to understanding its structure and function, which are of great importance in biological, biochemical, and medical studies. The sequencing of peptides and proteins by mass spectrometry (MS) has become a major tool in proteomics research because it is a cost effective and highly reproducible analytical technique. 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) and matrix-assisted laser desorption ionization (MALDI), 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. Thermodynamic values, such as the gas-phase acidity (GA or ΔG_{acid}), the ΔG for the deprotonation reaction $\text{AH} \rightarrow \text{A}^- + \text{H}^+$, 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 peptides. Changing the protonation state can impact the hydrogen bonding in the molecule and result in decreased or increased properties such as solubility, hydrophobicity, and electrostatic interactions. By combining these experimentally obtained results with the results from high level electronic structure theory calculations, an improved understanding of the structures and energetics of

polypeptides can be obtained. Herein we describe computational studies using reliable, correlated molecular orbital methods of the gas-phase properties, including acidities and heats of formation, and solution-phase acidities of amino acids (AAs) and small peptides including substituted molecules such as AA amides and phosphorylated AAs. This dissertation will focus on the prediction of fundamental thermodynamic properties of AAs and small peptides that are of interest in the area of biochemistry, proteomics, and the Human Proteome Project.

DEDICATION

I dedicate this dissertation to my parents, Tim and Cecilia Stover. I could not have done this without your unwavering love, prayers, and support. I love you.

LIST OF ABBREVIATIONS AND SYMBOLS

AA	Amino acid
Aib	Tri-2-methylalanine, carbon methylated alanine
Ala	Alanine
Arg	Arginine
Asp	Aspartic acid
ATFP	4-amino-2,3,5,6-tetrafluoropyridine
aug-cc-pVXZ	Augmented, correlation consistent, polarized valence n zeta basis sets, where X = double (D), triple (T) or quadruple (Q)
aug-cc-pwCVXZ	Augmented, correlation consistent, polarized weighted core-valence X (D, T, Q or 5) zeta basis sets
B3LYP	Becke 93 (exchange), Lee-Yang-Parr (correlation) DFT functional
CBS	Complete basis set
CCSD(T)	Coupled cluster singles, doubles, and disconnected triples
CGTO	Contracted Gaussian type orbital
COSMO	Conductor-like screening model
COSMO-RS	Conductor-like screening model for real solvents
CI	Configuration interaction
CID	Collision induced dissociation
CISD	Configuration interaction singles and doubles
CPMD	Car-Parrinello molecular dynamics

CV	Core valence
Cys	Cysteine
DFT	Density functional theory
DMF	N,N-dimethylformamide
DMSO	Dimethyl sulfoxide
DZVP2	DFT optimized double zeta valence basis set with polarization functions
ΔE_{CBS}	Complete basis set energy change
ΔE_{CV}	Core valence energy change
ΔE_{Rel}	Scalar relativistic energy change
ΔE_{SO}	Spin orbit energy change
ΔE_{SR}	Scalar relativistic correction calculated as the MVD expectation values
ΔE_{ZPE}	Zero point energy change
ECD	Electron capture dissociation
ETD	Electron transfer dissociation
ESI	Electrospray ionization
FPD	Feller-Peterson-Dixon composite approach
FC	Frozen core approximation, inner shell electrons are excluded
FULL	All electrons are included in correlation calculation
FT-ICR	Fourier transform ion cyclotron resonance
FT-IR	Fourier transform infrared spectroscopy
FT-Raman	Fourier transform Raman spectroscopy
G	Gibb's free energy
Gn	Gaussian- <i>n</i> theory where n = 2, 3 or 4

G3B3	Variant of G3 theory, geometry obtained at B3LYP level
G3(MP2)	Variant of G3 theory which uses MP2 theory instead of MP4 theory
G3(MP2)B3	Variant of G3(MP2) theory, geometry obtained at B3LYP level
ΔG_{aq}	Aqueous deprotonation Gibbs free energy (solution free energy)
ΔG_{298} , ΔG , or GA	Gas-phase acidity or relative energy difference between neutrals
$\Delta\Delta G$	Relative free energy difference between two anions
$\Delta\Delta G_{\text{solv}}$	Aqueous solvation free energy
GB	Gas-phase basicity
GGA	Generalized gradient approximation functional
Glu	Glutamic acid
Gly	Glycine
GTO	Gaussian type orbital
ΔH_{298} , ΔH	Gas-phase enthalpy or relative enthalpy difference between neutrals
H	Enthalpy
ΔH_f 0 K	Heat of formation at 0K
ΔH_f 298 K	Heat of formation at 298K
ΔH_{rxn}	Reaction enthalpy change
$\Delta\Delta H$	Relative free energy difference between two anions
HF	Hartree-Fock
HGP	Human Genome Project
His	Histidine
HLC	Higher level correction
HPB	4-(4-hydroxyphenyl)-2-butanone

HPP	Human Proteome Project
HUPO	Human Proteome Organization
IR	Infrared
IRMPD	Infrared Multiple Photon Dissociation
K	Equilibrium constant
k_{exp}	Experimentally determined reaction rate constant
LCAO	Linear combination of atomic orbitals
LDA	Local density approximation functional
Lys	Lysine
MALDI	Matrix-assisted laser desorption ionization
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MD	Molecular dynamics
MO	Molecular orbital
MPX	Møller-Plesset perturbation theory where $X=2^{\text{nd}}$ (2) or 4^{th} (4) order
MVD	Mass-velocity and Darwin operators
Nma	Nitrogen methylated alanine
NMR	Nuclear magnetic resonance
OMe	Methyl Ester
p	Phospho-
PES	Potential energy surface or photoelectron spectroscopy
Phe	Phenylalanine
$\text{p}K_{\text{a}}$	Negative logarithm of the acid dissociation constant

PTM	Post-translational modification
QCISD(T)	Quadratic CI singles, doubles, and disconnected triples
R	Gas constant
RE	Reaction efficiency
RRKM	Rice-Ramsperger-Kassel-Marcus
R/UCCSD(T)	Open-shell CCSD(T)
S	Entropy
Sar	Sarcosine, nitrogen methylated glycine
SCF	Self-consistent field
SCRf	Self-consistent reaction field
Ser	Serine
SO	Spin orbit
SR	Scalar relativistic
STO	Slater type orbital
T	Temperature
TAE	Total atomization energy
THF	tetrahydrofuran
Thr	Threonine
Tyr	Tyrosine
VDZ	Valence double zeta basis set
ZPE	Zero point energy
...	Hydrogen bond
Å	Angstrom

ACKNOWLEDGMENTS

I would like to thank all of the people who have guided, inspired, and helped me throughout my graduate career. First of all, I would like to thank my research advisor Dr. David A. Dixon for his tremendous support, guidance, and patience. His hands on approach helped me to become an improved and more disciplined researcher. I will forever be grateful for the many opportunities he afforded me to develop my teaching and presentation skills. I would also like to thank Dr. Carolyn Cassady who is the PI on my research funding and essentially served as a secondary advisor. Her willingness to share her expertise was invaluable to my comprehension of the research project and my growth as a scientist.

I would like to thank my other committee members, Dr. Daniel J. Goebbert, Dr. Michael K. Bowman, and Dr. C. Heath Turner, for their challenging questions, guidance, and support. I would also like to thank all of my current and past group members who have become my family here in Tuscaloosa. I greatly appreciate their unconditional support and willingness to help in both my professional and personal life.

I would like to thank Dr. Samantha S. Bokatzian and Chelsea E. Plummer, members of Dr. Cassady's group, who spearheaded the experimental side of this research project. I would also like to thank the undergraduates who have worked alongside me on this project: Sean R. Miller, John T. Davis, Michael A. Raddatz, Johan H. Both, Maranda Burns, Ashley S. McNeil, Cody H. Savage, William C. Jackson, and Jake D. Owen. Additionally, I would like to thank the

contributors to this work not previously mentioned: Dr. Virgil E. Jackson, Dr. Myrna H. Matus, and Margaret A. Adams, co-authors of chapter 2.

I am very thankful for the financial support that made my graduate studies possible at the University of Alabama. I would like to thank the National Science Foundation (NSF) for awarding me the GAANN (Graduate Assistance in Areas of National Need) fellowship for one year and for funding my research project. I would also like to thank the Howard Hughes Medical Institute (HHMI) for awarding me a teaching fellowship for two years. I am also grateful to the Alabama Supercomputing Center (ASA) for providing the majority of the computational resources used in this work. I would also like to thank The University of Alabama and the Department of Chemistry for financial support and assistance over the past five years.

Last, but not least, I would like to thank my family and friends for their love, support, and prayers throughout my PhD years. Special thanks are owed to Ms. Rhonda Robertson and Dr. Marcus Steele for their encouragement and support throughout my undergraduate career in chemistry. Without them, I know I would not be where I am today. I would like to thank my parents again and my siblings, Tiffany and Michael, for their love, support, and encouragement over the years. I would also like to thank the close group of friends here in Tuscaloosa who have made this journey so memorable: CJ Pruitt, Virgil Jackson, Chelsea McMillen, Monica Vasiliu, Ted Garner, Ashley Jolly, Alie Mallet, Bradley Smith, Luis Flores, and Katrina Staggemeier. CJ Pruitt deserves special thanks for putting up with the most and adding happiness, laughter, and childlike amusement to my life. I would like to thank Jay and Donita Henderson and Gail Sappington for pouring into me spiritually and emotionally throughout my graduate career. Finally, I would like to thank Willie and Meri Newell, Kaci Robertson, and Jennifer Lillico for their continued friendship, support, and encouragement over the years.

CONTENTS

ABSTRACT.....	ii
DEDICATION.....	iv
LIST OF ABBREVIATIONS AND SYMBOLS	v
ACKNOWLEDGMENTS	x
LIST OF TABLES.....	xv
LIST OF FIGURES	xvii
CHAPTER 1 INTRODUCTION.....	1
1.1 Background.....	1
1.2 Computational Chemistry	3
1.2.1 Molecular Orbital Theory	4
1.2.2 G3(MP2).....	6
1.2.3 Density Functional Theory	8
1.3 Computational Thermodynamics.....	10
1.4 Chapter Descriptions.....	12
References.....	15
CHAPTER 2 FUNDAMENTAL THERMOCHEMICAL PROPERTIES OF AMINO ACIDS: GAS-PHASE AND AQUEOUS ACIDITIES AND GAS- PHASE HEATS OF FORMATION	21
2.1 Introduction.....	21
2.2 Computational Methods.....	24

2.3	Results and Discussion	28
2.4	Conclusions.....	40
	References.....	66
	Appendix.....	71
CHAPTER 3	AN EXPERIMENTAL AND COMPUTATIONAL INVESTIGATION INTO THE GAS-PHASE ACIDITIES OF TYROSINE AND PHENYL-ALANINE: THREE STRUCTURES FOR DEPROTON- ATED TYROSINE.....	79
3.1	Introduction.....	79
3.2	Experimental and Computational Methods	82
3.3	Results and Discussion	87
3.4	Conclusions.....	101
	References.....	133
	Appendix.....	141
CHAPTER 4	AN EXPERIMENTAL AND COMPUTATIONAL STUDY OF THE GAS-PHASE ACIDITIES OF THE COMMON AMINO ACID AMIDES.....	146
4.1	Introduction.....	146
4.2	Experimental and Computational Methods	148
4.3	Results and Discussion	150
4.4	Conclusions.....	158
	References.....	175
	Appendix.....	181
CHAPTER 5	GAS- AND SOLUTION-PHASE ACIDITIES OF PHOSPHORYLATED AMINO ACIDS AND THEIR AMIDES.....	188
5.1	Introduction.....	188

5.2	Computational and Experimental Methods	191
5.3	Results and Discussion	194
5.4	Conclusions.....	210
	References.....	247
	Appendix.....	254
CHAPTER 6	GAS-PHASE DEPROTONATION OF THE PEPTIDE BACKBONE FOR TRIPEPTIDES AND THEIR METHYL ESTERS WITH HYDROGEN AND METHYL SIDE CHAINS	260
6.1	Introduction.....	260
6.2	Experimental and Computational Methods	262
6.3	Results and Discussion	265
6.4	Conclusions.....	275
	References.....	301
	Appendix.....	305
CHAPTER 7	HEATS OF FORMATION AND ACIDITIES OF THE DIAMINO- BENZENES AND HEATS OF FORMATION OF CARBOXYLIC ACIDS, PHENOL, AND CRESOLS	310
7.1	Introduction.....	310
7.2	Computational Methods.....	310
7.3	Results and Discussion	311
7.4	Conclusions.....	314
	References.....	319
	Appendix.....	322
CHAPTER 8	CONCLUSIONS	323

LIST OF TABLES

2.1	G3MP2 Acidities of the Amino Acids in kcal/mol.....	42
2.2	Calculated Acidities of Tyrosine and Cysteine in kcal/mol at Different Computational Levels	44
2.3	Acidities of the 5 Rare Amino Acids in kcal/mol.....	45
2.4	Calculated G3MP2 pK _a 's of the Amino Acids in kcal/mol	46
2.5	Calculated Heats of Formation at the G3MP2 level of the Amino Acids in kcal/mol	48
2.6	Calculated Heats of Formation at the G3MP2 level of the 5 Rare Amino Acids in kcal/mol.....	49
3.1	G3(MP2) Gas-Phase Acidities of the Carboxylic Acids and Phenols in kcal/mol	104
3.2	Calculated Gas Phase Acidities of <i>o</i> -, <i>m</i> -, and <i>p</i> - Cresols in kcal/mol at Different Computational Levels	105
3.3	Experimental, Theoretical, and Literature Gas-Phase Acidities in kcal/mol for Tyrosine, Phenylalanine, and HPB	106
3.4	Reaction Efficiencies for the Proton Transfer Reactions of Tyrosine, Phenylalanine, and HPB with Reference Compounds.....	108
3.5	Reactions of Trimethylsilyl Azide with Deprotonated Tyrosine Electrosprayed from Various Solvents	110
3.6	Relative Free Energies in kcal/mol of Solvated Tyrosine	111
4.1	Experimental and Computational GAs of the Amino Acid Amides in kcal/mol	160
5.1	Lowest Energy Gas-Phase Acidities at the G3(MP2) level in kcal/mol.....	213
5.2	Calculated Acidities in kcal/mol of Phosphoserine at Different Computational Levels. All with respect to neutral 1	214

5.3	Calculated Acidities in kcal/mol of Phosphothreonine at Different Computational Levels. All with respect to neutral 1	215
5.4	Vibrational Frequencies (ν , cm^{-1}) of Phosphoserine Anions at the B3LYP/aug-cc-pvdz level	216
5.5	Vibrational Frequencies (ν , cm^{-1}) of Phosphothreonine Anions at the B3LYP/aug-cc-pvdz level	217
5.6	N-H, O-H, and Hydrogen Bond Lengths of Phosphoserine Anions in Å at Different Computational Levels	218
5.7	N-H, O-H, and Hydrogen Bond Lengths of Phosphothreonine Anions in Å at Different Computational Levels	219
5.8	Gas-Phase Acidities in kcal/mol at the G3(MP2) level and Aqueous pKa's Obtained with COSMO Parameterization of Phosphorylated Amino Acids and Amides	220
5.9	Vibrational Frequencies (ν , cm^{-1}) of Phosphotyrosine Anions at the B3LYP/aug-cc-pvdz level	222
5.10	Calculated Heats of Formation at 298 K in kcal/mol at the G3(MP2) level of Phosphorylated Amino Acids and Amides with Two Different Computational Approaches	223
6.1	Reaction Efficiencies for the Proton Transfer reactions of Deprotonated Tripeptides with Reference Compounds.....	277
6.2	Experimental and G3(MP2) Theoretical GAs in kcal/mol for Tripeptides	278
6.3	Reaction Efficiencies for the Proton Transfer Reactions of Deprotonated Tripeptide Methyl Esters with Reference Compounds.....	279
6.4	Experimental and G3(MP2) Theoretical GAs in kcal/mol for Peptide Methyl Esters.....	280
7.1	Calculated Gas-Phase Acidities of <i>o</i> -, <i>m</i> -, and <i>p</i> - Diaminobenzenes in kcal/mol at Different Computational Levels	315
7.2	Calculated Heats of Formation at 298 K in kcal/mol at the G3(MP2) level of the Phenols and Diaminobenzenes	316
7.3	Calculated Heats of Formation at 298 K in kcal/mol at the G3(MP2) level of Carboxylic Acids	317

LIST OF FIGURES

2.1	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.....	50
2.2	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.....	63
3.1	Structures of Tyrosine, Phenylalanine, and HPB.....	112
3.2	G3(MP2) optimized structures for carboxylic acids and anions. Important hydrogen bonds in Å.....	113
3.3	G3(MP2) optimized structures for phenol, <i>o</i> -, <i>m</i> -. and <i>p</i> -cresol and anions.....	116
3.4	G3(MP2) results showing the structures of the most stable neutral acid and the lowest energy anions resulting from proton loss. Energies are in kcal/mol. Hydrogen bond distances in Å. The yellow arrows show the deprotonation site. (a) Tyrosine (b) Phenylalanine (c) HPB	117
3.5	G3(MP2) optimized structures for phenylalanine conformers. Relative energies in the gas and solution phases are given in kcal/mol	120
3.6	G3(MP2) optimized structures for tyrosine conformers. Relative energies in the gas and solution phases are given in kcal/mol	122
3.7	G3(MP2) optimized higher energy amino acid anions. Energies in kcal/mol. Hydrogen bond distances in Å. The yellow arrows show the deprotonation site. (a) tyrosine (b) phenylalanine	125
3.8	Reactant loss curve for the reaction of deprotonated tyrosine with ethyl cyanoacetate, which is present at a constant pressure of 8.9×10^{-8} mbar. The solvent is 49.5:49.5:1 (v/v/v) CH ₃ OH:H ₂ O:NH ₄ OH. The logarithm of [Tyr - H] ⁻ intensity is plotted as a function of reaction time. The experimental data points (black circles) are fitted to an equation involving the sum of two exponential decays	126
3.9	Reactant loss curve for the reaction of deprotonated tyrosine with 4-amino-2,3,5,6-tetrafluoropyridine (ATFP), which is present at a constant pressure of 5.6×10^{-8} mbar. The solvent is 99:1 (v/v) CH ₃ CN:H ₂ O. The logarithm of [Tyr - H] ⁻	

	intensity is plotted as a function of reaction time. The experimental data points (black circles) are fitted to an equation involving the sum of two exponential decays.....	127
3.10	Percentage of product ions from the reaction of deprotonated tyrosine with trimethylsilylazide (TMSN ₃) as a function of time that the ions accumulate in the hexapole. The presence of m/z 252 indicates a phenoxide deprotonated ion structure, while m/z 42 indicates a carboxylate structure. The solvent systems are: (a) 74.5:24.5:1 CH ₃ OH:H ₂ O:NH ₄ OH, (b) 98:2 CH ₃ CN:H ₂ O, and (c) 99:1 CH ₃ CN:H ₂ O.....	128
3.11	B3LYP/aug-cc-pvdz optimized tyrosine anions. Important hydrogen bonds in Å. Relative energies in the gas and solution phases are given in kcal/mol	129
3.12	B3LYP/aug-cc-pvdz optimized tyrosine anions with explicit waters. Important hydrogen bonds in Å. Relative energies in the gas and solution phases in kcal/mol. (a) two explicit waters (b) three explicit waters (c) six explicit waters.....	130
4.1	Structures of an Amino Acid, Amino Acid Amide, and Tripeptide	162
4.2	Reactant ion loss curves for deprotonated proline amide reacting with (a) 1,2,4,5-tetrafluorobenzene (GA = 353.3 kcal/mol) at a constant pressure of 7.8 × 10 ⁻⁸ mbar and (b) 3-methylpyrazole (GA = 348.3 kcal/mol) a constant pressure of 9.0 × 10 ⁻⁸ mbar. The logarithmic [M-H] ⁻ intensity is plotted versus reaction time. In (a) the experimental data points (black circles) are fit to an equation involving the sum of two exponentials, while (b) involves a single exponential fit. Hydrogen bond distances in the structures of (a) are given in Å.....	163
4.3	The most stable neutral and two anions for aliphatic side chain amino acid amides at the G3(MP2) level. Important hydrogen bond distances are given in Å. mc-cis = main chain deprotonation with cis-like orientation and mc-trans = main chain deprotonation with trans-like orientation (Note this is for the orientation of the N-H bond in the [-C(=O)NH] ⁻ group. All energetic values are in kcal/mol. (a) alanine (b) glycine (c) isoleucine (d) leucine (e) phenylalanine (f) proline (g) valine	164
4.4	The most stable neutral and two anions for asparagine, glutamine, lysine, methionine, serine, threonine, and arginine amide at the G3(MP2) level. sc = side chain deprotonation. See Figure 4.3 caption for additional details. (a) asparagine (b) glutamine (c) lysine (d) methionine (e) serine (f) threonine (g) arginine.....	168
4.5	The most stable neutral and two anions for the amino acid amides with acidic side chains at the G3(MP2) level. mc-cis/sc = main chain deprotonation with cis-like orientation and proton sharing between the main and side chains and mc-trans = main chain deprotonation with trans-like orientation and proton sharing between	

	the main and side chains. See Figure 4.3 caption for additional details. (a) aspartic acid (b) glutamic acid (c) tyrosine (d) cysteine (e) tryptophan (f) histidine.....	172
5.1	Optimized structures of pSer, pThr, and anions at the G3(MP2) level. GAs and relative energies are in kcal/mol. Important hydrogen bond distances, given in Å, for the [pSer-H] ⁻ and [pThr-H] ⁻ anions are given in Tables 6 and 7 respectively	224
5.2	Infrared spectra of deprotonated phospho-serine and -threonine anions. IRMPD spectrum is on top. Anions [A] – [F]: calculated IR spectra at the B3LYP/aug-cc-pvdz level. Relative energies of isomers at the G3(MP2) level of theory	226
5.3	Optimized structures of phosphotyrosine and corresponding anions at the G3(MP2) level. GAs and relative energies are given in kcal/mol. Important hydrogen bond distances are given in Å	228
5.4	Infrared spectra of deprotonated phosphotyrosine. IRMPD spectrum is on top. Anions [A] – [D] & [F]: calculated IR spectra at the B3LYP/aug-cc-pvdz level. Relative energies of isomers at the G3(MP2) level of theory.....	229
5.5	Optimized structures of phospho-serine, -threonine, and -tyrosine amide neutrals and corresponding anions at the G3(MP2) level. GAs and relative energies are in kcal/mol. Important hydrogen bond distances are given in Å	230
5.6	Optimized structures of the phosphorylated amino acids and their anions at the G3(MP2) level. GAs and relative energies are in kcal/mol. Important hydrogen bond distances are given in Å	233
5.7	Optimized structures of the phosphorylated amino acid amides and their anions at the G3(MP2) level. GAs and relative energies are in kcal/mol. Important hydrogen bond distances are given in Å.....	240
6.1	Structures of tripeptides (X=H) and their methyl esters (X=CH ₃)	281
6.2	G3(MP2) calculated structures for neutral and deprotonated tripeptide acids (a) GlyGlyGly (b) GlyAlaGly (c) AlaGlyAla (d) AlaAlaAla (e) AibAibAib (f) SarSarSar (g) NmaNmaNma. Arrows indicate the deprotonation site. Bond lengths are in Å	282
6.3	G3(MP2) optimized structures of the higher energy neutrals. Energies in kcal/mol.....	289
6.4	G3(MP2) calculated structures for neutral and deprotonated tripeptide methyl esters (a) GlyGlyGly-OMe (b) GlyAlaGly-OMe (c) AlaAlaAla-OMe. Arrows indicate the deprotonation site. Bond lengths are in angstroms. These peptide methyl esters deprotonate readily by ESI	291

6.5	G3(MP2) calculated structures for neutral and deprotonated tripeptide methyl esters (a) AlaGlyAla-OMe (b) AibAibAib-OMe (c) SarSarSar-OMe. Arrows indicate the deprotonation site. Bond lengths are in angstroms. These peptide methyl esters do not deprotonate well by ESI	294
6.6	Van der waals surfaces and solvent accessible surfaces for neutral (a) AlaGlyAla-OMe (b) AibAibAib-OMe (c) SarSarSar-OMe	297
6.7	Van der waals surfaces and solvent accessible surfaces for the tripeptides that readily deprotonated with ESI	298
7.1	G3(MP2) optimized structures for <i>o</i> -, <i>m</i> - and <i>p</i> -diaminobenzene and anions. Important hydrogen bonds in Å	318

CHAPTER 1: INTRODUCTION

1.1 Background In 1990, the Human Genome Project (HGP) was started by the National Institute of Health and the Department of Energy of the United States with the goal of sequencing the entire human genome.¹ After the successful completion of the HGP in 2003,^{2,3} the Human Proteome Project (HPP) was planned by the Human Proteome Organization (HUPO) as the international follow-up project to the HGP building on the newly obtained genetic blueprint of the human body. The HPP was launched at the 2011 World Congress of Proteomics in Geneva, Switzerland, with the main goal of experimentally mapping the entire human proteome.⁴ The human body contains 20,300 protein-encoding genes, and currently there is no experimental data on one-third of these genes and only limited data on the other two-thirds. The HPP was built upon three research pillars: mass spectrometry, antibody capture, and knowledge databases. Through the combination of these three approaches, the goal of the HPP is to improve our overall understanding of cellular human biology, which can be applied to future medicinal diagnostic, therapeutic and preventative approaches.^{5,6,7}

Determining a protein's sequence is a first step to understanding its structure and function, which are of great importance in biological, biochemical, and medical studies. Mass spectrometry (MS) has become a major tool in proteomics research because it is a cost effective and highly reproducible analytical technique that allows for sensitive and rapid identification of peptides and proteins.⁸ The effectiveness of a mass spectrometer relies heavily on the appropriate selection of the type of mass analyzer with respect to the molecules being studied as

well as the source of the ions. The analyzers, which can be used separately, or paired together to perform tandem mass spectrometry (MS/MS), differ in key parameters such as sensitivity and resolution. With modern mass spectrometers, it is very easy to switch the sign of the voltages to perform either positive or negative ion mode studies. However, to date, biomolecule sequencing by MS/MS has primarily employed positive 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. Since the mid-1990s, a number of studies of the fragmentation of deprotonated peptides have been reported.^{9,10,11,12,13,14,15,16}

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)¹⁷ and matrix-assisted laser desorption ionization (MALDI),¹⁸ involve the addition and removal of protons. The sites where proton transfer reactions occur can affect the fragmentation patterns of peptide ions, which consequently impacts the sequence information that can be obtained from mass spectrometry experiments.^{19,20,21} The study of gas-phase proton transfer reactions provides unique insights into the structures and energetics of the peptides. Changing the protonation state can impact the hydrogen bonding in the molecule and result in decreased or increased properties such as solubility, hydrophobicity, and electrostatic interactions.^{22,23,24} By combining these experimentally obtained results with electronic structure theory computations, an improved understanding of the biological activities of polypeptides can be obtained. Herein we describe computational studies using reliable, correlated molecular orbital methods of the gas-phase properties, including acidities and heats of formation, of amino acids and small peptide including substituted molecules. This dissertation will focus on the prediction of fundamental thermodynamic properties of amino acids and small peptides that are

of interest in the area of biochemistry, proteomics, and the Human Proteome Project.

1.2 Computational Chemistry The goal of this work is to use highly accurate computational chemistry methods to predict thermodynamic properties of peptides. Molecular properties that can be calculated reliably include molecular geometries, thermodynamic properties such as acidities and heats of formation, dipole moments, and vibrational frequencies. These data can be used to develop reaction mechanisms and potentially even the rate constants using transition state or Rice-Ramsperger-Kassel-Marcus (RRKM)-type theories. Computational methods provide a way to study unstable molecules and even transition states. The field of computational chemistry grows with the increase in computing power and the broad availability of supercomputers as well as improvements in methods and algorithms.

Computational chemistry can be divided into two main categories: molecular mechanics and electronic structure theory. Molecular mechanics, or molecular dynamics (MD), is a classical computational method that is useful for determining the thermodynamic properties of large molecules over time. MD theory solves Newton's equations of motion explicitly by using classical force fields to describe the potential energy surface.²⁵ The second approach which is used in this dissertation is electronic structure theory which is the solution of the time-independent Schrödinger equation. This theory can be divided into three categories, ab-initio molecular orbital theory, density functional theory, and semi-empirical methods. Semi-empirical methods use fitting parameters obtained from experimental data whereas ab-initio methods only use physical constants.²⁶ The time-independent Schrödinger equation has the form:

$$\hat{H} \Psi = E \Psi \quad (1.1)$$

where \hat{H} is the molecular Hamiltonian, the total energy operator representing the kinetic energy and electrostatic interactions, Ψ is the wavefunction, and E is the total energy of the molecular

system. Because nuclei are heavier and move much more slowly than electrons, we can use the Born-Oppenheimer approximation, which assumes that the nuclei are fixed. By using the Born-Oppenheimer approximation, we can separate the electron and nuclear motion and define the electronic Hamiltonian \hat{H}_{elec} as:

$$\hat{H}_{elec} = - \sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{|R_A - r_i|} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{|r_j - r_i|} \quad (1.2)$$

where N represents the number of electrons, M represents the number of nuclei, ∇ is the gradient operator, A represents the nuclei, Z is the atomic number, i and j represent different electrons, and R and r represent the coordinates of the nuclei and electrons respectively. The first term is repulsive and is the kinetic energy of the electrons, the second term is the nuclear-electron attractive energy, and the third term is the electron-electron repulsive energy.²⁷ The use of first derivatives of the energy with respect to the nuclear coordinates $\{dx_i\}$ allows for geometry optimization. When $dE/\{dx_i\} = 0$, the second derivatives provide the curvature of the potential energy surface and are used to determine if the structure is at a minimum, a transition state, or a more complex saddle-point. There are two main approaches to solving the Schrodinger equation: molecular orbital (MO) theory and density functional theory (DFT).

1.2.1 Molecular Orbital Theory The simplest and most fundamental approach to solving the Schrödinger equation is Hartree-Fock (HF) theory. HF approximates the wavefunction of a molecule with N electrons as a Slater determinant in terms of molecular orbitals. The unknown molecular orbitals are constructed as a linear combination of the known atomic orbitals (LCAO). There are two main types of atomic orbitals, otherwise known as basis functions: Slater Type Orbitals (STOs) and Gaussian Type Orbitals (GTOs). STOs are more accurate than GTOs because GTOs do not properly represent the behavior near the nucleus. However, STOs are

computationally expensive, so GTOs are used instead. The issue with the form of the GTOs can be mitigated by using a linear combination of GTOs to mimic an STO creating a contracted GTO (CGTO) and achieving a greater level of accuracy. Thus, GTOs are the typical basis sets used in electronic structure theory calculations.

Currently, there are a substantial number of GTO basis sets available with varying numbers of basis functions. When choosing the proper basis set for a particular calculation, two ideas must be considered and balanced: computational cost and accuracy. The smallest basis set is defined as the minimum number of basis functions, or atomic orbitals, required to represent all of the electrons on an atom. Doubling the number of basis functions for each atomic orbital results in a double zeta (DZ) basis set whereas only doubling the valence orbitals associated with bonding results in a split valence basis set or valence double zeta (VDZ) basis set. The basis sets can be further expanded to be of triple, quadruple, or quintuple- ζ quality, etc.. Polarization basis sets include additional functions, for example p functions on hydrogen and d functions on heavy atoms, to allow more spatial variation in the basis set. An example of a basis set with polarization is a double zeta polarized (DZP) basis set. Diffuse basis sets involve adding functions with smaller exponents to better represent when electrons are far away from the nucleus such as in anions or transition states.

There are many different types of basis sets available in the literature. The basis sets developed by Pople and co-workers²⁸ are split-valence basis sets that explicitly state the number of GTO's in the CGTO and have the form X-YZG where X describes the core and Y and Z describe the valence. The polarization functions are described by a '*' or by explicitly stating the functions added in a parenthesis after the G whereas diffuse functions are represented by a '+'. For example, the 6-31+G* or 6-31+G(d) basis set is a split-valence double-zeta basis set with

both diffuse and polarization functions on the non-hydrogenic atoms. The core orbital is a CGTO made up of 6 GTOs, the inner part of the valence is a CGTO made up of 2 GTOs, and the outer part of the valence is a single GTO. The Dunning correlation consistent basis sets²⁹ were optimized using configuration interaction singles and doubles (CISD) wavefunctions. The functions are represented as cc-pVXZ where the cc stands for correlation consistent, the pV stands for polarized valence, and XZ stands for X-zeta where X=D, T, Q, 5..... Adding diffuse functions to this basis set is represented by the prefix 'aug'. In this dissertation, the basis sets primarily used were the 6-31G(*d*) Pople basis set, the polarized double zeta DGauss basis set DZVP2,^{30,31} and the aug-cc-pVXZ Dunning basis set where X=D or T.

The HF method gives a self-consistent field (SCF) solution because the wavefunction is iteratively converged until the expansion coefficients in the molecular orbitals are converged to a specific tolerance. However, HF theory neglects correlations between electrons, especially in a single orbital due to missing the Coulomb hole, which can lead to an inaccurate description of the electronic structure of the molecule. Corrections can be made to improve the accuracy of the solution of the Schrödinger equation, but the inclusion of such corrections results in a much more computationally expensive approach. Post-HF methods, methods that include electron correlation, usually begin with the HF wavefunction as a reference and then perform additional steps to obtain an improved energy value. Electron-correlation methods include configuration interaction (CI),^{32,33,34} Møller-Plesset (MP) perturbation theory,³⁵ coupled cluster (CC) theory,^{36,37} and multi-reference methods,³⁸ which are a subset of CI methods.

1.2.2 G3(MP2) In order to avoid a computationally expensive CCSD(T)^{39,40,41} calculation with a large basis set, the *G_n*, or Gaussian-*n*, methods⁴² were developed. These methods are composite approaches that attempt to describe the energies of molecules within chemical accuracy, ± 2

kcal/mol, by overcoming the deficiencies of HF theory and including terms to describe the correlation energy. The G_n methods calculate the total energy of a system at 0 K by performing a series of energy calculations at different levels of theory. G3(MP2)⁴³ is a computationally cheaper variation of Gaussian-3 (G3) theory⁴⁴ in which the effects of basis set extension are obtained from calculations at the second-order Møller-Plesset (MP2) level. The following sequence is used:

- (1) Optimization and frequency calculation at the HF/6-31G(d) level of theory.
- (2) Optimization at the MP2(FULL)/6-31(d) level of theory.
- (3) Single point calculation at the QCISD(T,FC)/6-31G(d)/MP2(FULL)/6-31G(d) level of theory
- (4) Single point calculation at the MP2(FC)/G3MP2large//MP2(FULL)/6-31G(d) level of theory.

The total energy equation for G3(MP2) at 0 K has the form:

$$E_0[G3(MP2)] = QCISD(T)/6 - 31G(d) + \Delta E(MP2) + \Delta E(SO) + E(HLC) + E(ZPE) \quad (1.3)$$

where the total energy is based on a higher level calculation at the quadratic configuration interaction (QCISD(T)) level of theory,⁴⁵ which is similar to a coupled cluster (CCSD(T)) calculation, and a series of additional corrections. $\Delta E(MP2)$ is the correction for the basis set extension at the MP2 level and is given by:

$$\Delta E(MP2) = [E(MP2/G3MP2large)] - [E(MP2/6 - 31G(d))] \quad (1.4)$$

where the G3MP2large basis set⁴⁶ is a modification of the 6-311+G(3df,2p) basis set from G2(MP2) theory. It is necessary to include a spin orbit relativistic correction $\Delta E(SO)$ which lowers the sum of the atomic energies decreasing the total atomization energies. These corrections are included by replacing energies that correspond to an average over the available spin multiplets with energies for the lowest multiplets, as most electronic structure codes are only capable of producing spin multiplet averaged wavefunctions. In this dissertation, this

correction is only needed for calculating heats of formation. The zero-point correction, $E(ZPE)$, is obtained by scaling the SCF frequencies from the HF/6-31G(*d*) level by 0.8929. The calculated harmonic frequencies need to be scaled to account for anharmonic effects because SCF harmonic frequencies overestimate experimental frequencies by up to $\sim 12\%$.⁴⁷ The higher level correction, $E(HLC)$, is added to account for any other remaining deficiencies in the calculation such as correlation energy between spin-paired electrons present in single bonds. For molecules, the $E(HLC)$ is given as:

$$E(HLC) = -An_B - B(n_\alpha - n_\beta) \quad (1.5)$$

where A and B are 9.279 and 4.471 mhartrees respectively and are numbers chosen to give the smallest error to the energy after fitting to experiment, and n_α and n_β are the number of alpha and beta valence electrons respectively. Again, the HLC is only relevant to the calculation of heats of formation.

In this dissertation, the thermodynamic calculations were performed with the G3(MP2) method because G3(MP2) has an advantage over density functional theory (DFT) methods in terms of reliable predictions for these types of compounds. The correlated molecular orbital methods in G3(MP2) perform better in the prediction of hydrogen bond energies as well as steric non-bonded interactions than do most widely-used DFT exchange-correlation functionals.

1.2.3 Density Functional Theory Density functional theory (DFT) is a relatively inexpensive computational method that is founded on the Hohenburg-Kohn theorem⁴⁸ which states that the ground-state electronic energy is a functional of the ground state electron density. Based on this idea, the general energy expression for DFT is given as:

$$E[\rho] = T_s[\rho] + V_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \quad (1.6)$$

where T_s is the kinetic energy for the non-interacting system, ρ is the electron density, V_{ne} is the nuclear-electronic potential energy, J is the electron-electron repulsive energy, and E_{xc} is the exchange-correlation energy.⁴⁹ The exchange-correlation functional describes the many-body effects of a non-classical system and is defined as:

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (V_{ee}[\rho] - J[\rho]) \quad (1.7)$$

where T and V_{ee} are the kinetic energy and electron-electron repulsive energy for the interacting system respectively. Since the exact form of E_{xc} is unknown, the accuracy of DFT is directly dependent upon the quality of the functional that is used. There are a wide variety of functionals currently available including local density approximation (LDA) functionals,^{50,51} generalized gradient approximation (GGA) functionals,^{52,53,54} and hybrid functionals.⁵⁵ LDA functionals depend only on electronic density whereas GGA depends on the electronic density and its gradient. Hybrid functionals are comprised of a portion of HF exchange in addition to exchange and correlation from a variety of other sources. The B3LYP (Becke, three parameter, Lee-Yang-Parr) exchange-correlation functional^{56,57} is currently the, most popular hybrid functional and has the form:

$$E_{xc}^{B3LYP} = aE_x^{HF} + (1 - a)E_x^{LDA} + b\Delta E_x^{B88} + cE_c^{LYP} + (1 - c)E_c^{VWN} \quad (1.8)$$

where the first three terms correspond to exchange functionals and the last two correspond to correlation functionals. The parameters a, b, and c have been fit by comparison to experimental thermodynamic data and are equal to 0.20, 0.72, and 0.81 respectively.

In this dissertation, the B3LYP exchange-correlation functional and the DZVP2 basis set^{30,31} were used to perform initial geometry optimizations on a wide range of structures. 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. DFT performs reasonably well for predicting geometries and is a good first approximation of determining the potential energy surfaces (PES) of the molecule. Because of the issues with the common functionals in DFT,⁵⁸ we used the G3MP2 method as described in Section 1.6.

1.3 Computational Thermodynamics Thermodynamics is used to describe energy changes associated with various chemical and physical changes. According to the first law of thermodynamics, the change in internal energy (ΔU) is defined as the sum of heat (q) and work (w). If the system is at constant pressure, then we can define enthalpy (H), the heat content, as:

$$H = U + PV \quad \text{or} \quad \Delta H = \Delta U + P\Delta V \quad (1.9)$$

According to the second law of thermodynamics, the entropy (S), or measure of disorder in the system, can be defined as:

$$T\Delta S = \Delta U + w \quad (1.10)$$

The Gibb's free energy (G) combines enthalpy and entropy into one value to predict whether a process is spontaneous or not at constant pressure and temperature and is defined as:

$$G = H - TS \quad \text{or} \quad \Delta G = \Delta H - T\Delta S \quad (1.11)$$

In this dissertation, the thermodynamic values of interest are the gas-phase acidities, deprotonation enthalpies, and heats of formation as well as aqueous acidities. The gas-phase acidity (GA or ΔG_{acid}) of a molecule AH is the ΔG of reaction 1.12:



The GAs were calculated using the energy values from the G3(MP2) method because previous work showed that this method is in agreement with experimental values and coupled cluster CCSD(T)/CBS values to within about 1 kcal/mol.⁵⁹ The change in free energy can be used to calculate the equilibrium constant (K):

$$K = e^{-\frac{\Delta G}{RT}} \quad (1.13)$$

where R is the gas constant and T is the temperature in Kelvin. The total entropy is the sum of the entropy contributions due to the electronic, vibrational, translation, and rotational motions as a function of temperature. The entropies were calculated using the statistical mechanical expressions for a harmonic oscillator, rigid rotor description of a molecule.⁶⁰

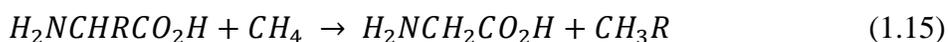
Theoretical pK_a values in aqueous solution can be calculated by combining the gas-phase acidities with single point (at the optimum gas-phase geometry) self-consistent reaction field calculations⁶¹ with the COSMO parameterization.⁶² The Gibbs free energy for deprotonation in aqueous solution (ΔG_{aq}) was calculated from the gas-phase free energy and the aqueous solvation free energy. The COSMO solvation energy is calculated as the sum of the electrostatic energies (polarized solute - solvent) and the non-electrostatic 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 the following equation:

$$pK_a' = pK_a(HA) + \Delta G_{aq}/(2.303RT) \quad (1.14)$$

where ΔG_{aq} is the solution free energy, R is the gas constant, and T is the temperature at 298 K.

The molecular heats of formation at 0 K were calculated from the atomization energies at the G3(MP2) level using 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, 65.66 ± 0.06 kcal/mol for S, 75.42 ± 0.02 kcal/mol for P, 51.63 ± 0.001 kcal/mol for H).⁶³ Heats of formation at 298 K were calculated by following the procedures outlined by Curtiss et al.⁶⁴ The heats of formation were also calculated from a set of isodesmic reactions which are defined as "... chemical changes in which there is retention of the number of bonds of a given formal type, but with a change in their

relation to one another”⁶⁵ at the G3MP2 level using the calculated heat of formation of glycine. The heat of formation of glycine was calculated by using the Feller-Peterson-Dixon (FPD) composite approach^{66,67,68,69,70} developed in the Dixon group in collaboration with Peterson and Feller at Washington State University. This method is based on CCSD(T) calculations^{37,39,40,41} extrapolated to the complete basis set limit using the correlation consistent basis sets plus additional corrections. For example, for the amino acids, the isodesmic reaction used methane as a reactant and glycine as a product where the reaction is:



where R represents the side-chain that is specific to each amino acid. This leads to the energy expression given as:

$$\Delta H_f(\text{rxn}) = \Delta H_f(\text{product}) + \Delta H_f(\text{glycine}) + \Delta H_{\text{rxn}}(\text{G3(MP2)}) - \Delta H_f(\text{CH}_4) \quad (1.16)$$

where the heats of formation of the product, glycine, and CH₄ are taken from experiment.

1.4 Chapter Descriptions Chapters 2 and 3 describe computational studies of common L-amino acids as well as some so-called rare amino acids. Chapter 2 describes the prediction of the gas-phase and aqueous acidities as well as the heats of formation for the neutral compounds of the 20 common L-amino acids, five so-called rare amino acids, and their corresponding anions. 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. Chapter 3 describes further experimental mass spectrometry and computational G3(MP2) studies of tyrosine and phenylalanine in the gas phase in order to determine the site of proton removal from tyrosine. Three GA values were predicted for tyrosine at the G3(MP2) level of theory and

confirmed the results of studies of proton transfer reactions in a Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer.

Chapters 4 and 5 describe the experimental mass spectrometry and computational studies of modified amino acids. Chapter 4 focuses on using the G3(MP2) level of theory and proton transfer reactions with an FT-ICR mass spectrometer to determine the gas-phase acidities of the 20 L-amino acid amides. The amino acid amides were formed by modification of the C-terminal carboxylic acid (-CO₂H) to the amide (-CONH₂) which provides a better model of amino acid residues in peptides as most residues do not have a terminal carboxylic acid end group.

Chapter 5 focuses on the phosphorylation of amino acids, a common post-translational modification (PTM) in proteins which allows for the function of the protein to be altered after the cell has been encoded. Phosphorylation involves the substitution of a phosphate group on the side chain of an amino acid residue adding a second site for proton loss and increasing the potential for hydrogen bonding capacity. The GAs of ten phosphorylated amino acids and their corresponding amides have been predicted at the G3(MP2) level of theory. To confirm the validity of the G3(MP2) results, the GAs of the three most common phosphorylated amino acids, phospho-serine, -threonine, and -tyrosine, and their amides have been determined experimentally by proton transfer reactions with an FT-ICR mass spectrometer. Additionally, the IR spectra for deprotonated phospho-serine, -threonine, and -tyrosine have been calculated at the B3LYP/aug-cc-pvdz level of theory and compared to previously obtained IRMPD experimental spectra.

Chapter 6 focuses on six tripeptides composed of glycine and alanine residues (GlyGlyGly, GlyAlaGly, AlaGlyAla, AlaAlaAla, AibAibAib, and SarSarSar) and their methyl esters. The backbone substituted AibAibAib tripeptide was generated by replacing the hydrogens on the peptide backbone with methyl groups. The SarSarSar tripeptide, or N-methyl glycine, was

generated by methylating the central and C-terminal amide nitrogens and one of the N-terminal NH₂ hydrogens. Their GAs were predicted at the G3(MP2) level of theory and measured by proton transfer reactions in a FT-ICR mass spectrometer. Additionally, the van der Waals and solvent accessible surfaces were generated for each of the neutral tripeptides to study conformational and steric hindrance effects.

Chapter 7 focuses on the predictions of the gas-phase acidities and heats of formation of ortho, meta, and para substituted diaminobenzenes as there is substantial interest in the disubstituted benzenes as monomers for polymers. Additionally, the heats of formation of 11 carboxylic acids, phenol and the three cresols have also been predicted. The G3(MP2) level of theory was used for the heats of formation calculations whereas the GAs for the diaminobenzenes were calculated at a variety of different computational levels.

References

- ¹ Chial, H. DNA Sequencing Technologies Key to the Human Genome Project. *Nature Education* **2008**, *1*, 219.
- ² Lander E. S., Linton L. M., Birren B., et al. Initial Sequence Analysis of the Human Genome. *Nature* **2001**, *409*, 860-921.
- ³ Venter J. C., Adams M. D., Myers E. W., et al. The Sequence of the Human Genome. *Science* **2001**, *291*, 1304-1351.
- ⁴ Omenn, G. S. The HUPO Human Proteome Project (HPP), a Global Health Research Collaboration. *Central Asian Journal of Global Health* **2012**, *1*.
- ⁵ A Gene-Centric Human Proteome Project *Mol. Cell Proteomics* **2010**, *9*, 427-429.
- ⁶ The Call of the Human Proteome. *Nat. Methods* **2010**, *7*, 661.
- ⁷ Legrain, P.; Aebersold, R.; Archakov, A.; et al. The Human Proteome Project: Current State and Future Direction. *Mol. Cell Proteomics* **2011**, *10*, M111.0099993.
- ⁸ Nilsson, T.; Mann, M.; Aebersold, R.; Yates, J. R.; Bairoch, A.; Bergeron, J. J. M. Mass Spectrometry in High-Throughput Proteomics: Ready for the Big Time. *Nat. Methods* **2010**, *7*, 681-685.
- ⁹ Ewing, N. P.; Cassady, C. J. Dissociation of Multiply-Charged Negative Ions for Hirudin (54-65), Fibrinopeptide B, and Insulin A (Oxidized). *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 105-116.
- ¹⁰ Jai-nhuknan, J.; Cassady, C. J. Negative Ion Postsource Decay Time-of-Flight Mass Spectrometry of Peptides Containing Acidic Amino Acid Residues. *Anal. Chem.* **1998**, *70*, 5122-5128.
- ¹¹ Brinkworth, C. S.; Dua, S.; Bowie, J. H. Backbone Cleavages of [M-H]⁻ Anions of Peptides. New Backbone Cleavages Following Cyclisation Reactions of a C-terminal [CONH]⁻ Group with Ser Residues and the Use of the Gamma Backbone Cleavage Initiated by Gln to Differentiate Between Lys and Gln Residues. *Eur. J. Mass Spectrom.* **2002**, *8*, 53-66.
- ¹² Maclean, M. J.; Brinkworth, C. S.; Bilusich, D.; Bowie, J. H.; Doyle, J. R.; Llewellyn, L. E.; Tyler, M. J. New Caerin Antibiotic Peptides from the Skin Secretion of the Dainty Green Tree Frog *Litoria gracilentata*. Identification using Positive and Negative Ion Electrospray Mass Spectrometry. *Toxicon Mass Spectrometry in Toxinology - A 21st-Century Technology for the Study of Biopolymers from Venoms* **2006**, *47*, 664-675.
- ¹³ Marzluff, E. M.; Campbell, S.; Rodgers, M. T.; Beauchamp, J. L. Low Energy Dissociation Pathways of Deprotonated Peptides in the Gas Phase. *J. Am. Chem. Soc.* **1994**, *116*, 7787-7796.

-
- ¹⁴ Harrison, A. G. Sequence-Specific Fragmentation of Deprotonated Peptides Containing H or Alkyl Side Chains. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 1-13.
- ¹⁵ Bowie, J. H.; Brinkworth, C. S.; Dua, S. Collision-Induced Fragmentations of the (M-H)-Parent Anions of Underivatized Peptides: An Aid to Structure Determination and Some Unusual Negative Ion Cleavages. *Mass Spectrom. Rev.* **2002**, *21*, 87-107.
- ¹⁶ Bilusich, D.; Bowie, J. H. Fragmentations of (M-H)- Anions of Underivatized Peptides. Part 2: Characteristic Cleavages of Ser and Cys and of Disulfides and Other Post-Translational Modifications, Together with Some Unusual Internal Processes. *Mass Spectrom. Rev.* **2009**, *28*, 20-34.
- ¹⁷ Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F. Electrospray Ionization-Principles and Practice. *Mass Spectrom. Rev.* **1990**, *9*, 37-70.
- ¹⁸ Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers. *Anal. Chem.* **1991**, *63*, 1193A-1203A.
- ¹⁹ Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and Localized Protons: A Framework for Understanding Peptide Dissociation. *J. Mass Spectrom.* **2000**, *35*, 1399-1406.
- ²⁰ Cox, K. A.; Gaskell, S. J.; Morris, M.; Whiting, A. Role of the Site of Protonation in the Low-Energy Decompositions of Gas-Phase Peptide Ions. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 522-531.
- ²¹ Johnson, R. S.; Martin, S. A.; Biemann, K. Collision-Induced Fragmentation of (M + H)⁺ Ions of Peptides. Side Chain Specific Sequence Ions. *Int. J. Mass Spectrom. Ion Proc.* **1988**, *86*, 137-154.
- ²² Ghelis, C.; Yon, J. *Protein Folding*, Academic Press: New York, 1982.
- ²³ McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press: Cambridge, 1987.
- ²⁴ Margoulies, M.; Greenwood, F. C. *Structure-Activity Relationships of Protein and Polypeptide Hormones*, Excerpta Medica Foundation: Amsterdam, 1972.
- ²⁵ Cai, W.; Li, J.; Yip, S. Molecular Dynamics. In *Comprehensive Nuclear Materials*; Konings, R. J. M., Ed.; Elsevier: Amsterdam, 2012; pp 249-265.
- ²⁶ Thiel, W. Semiempirical Quantum-Chemical Methods in Computational Chemistry. In *Theory and Applications of Computational Chemistry: The First Forty Years*; Dykstra, C.; Frenking, G.; Kim, K.; Suseria, G., Eds.; Elsevier: Amsterdam, 2005; pp 559-580.

-
- ²⁷ Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Publications, Inc.: New York, 1996.
- ²⁸ Ditchfield, R.; Hehre, W.J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724-728.
- ²⁹ Dunning, Thomas H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007-1023.
- ³⁰ Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. Optimization of Gaussian-Type Basis Sets for Local Spin Density Functional Calculations Part I. Boron through Neon, Optimization Technique and Validation. *Can. J. Chem.* **1992**, *70*, 560-571.
- ³¹ Sosa, C.; Andzelm, J.; Elkin, B. C.; Wimmer, E.; Dobbs, K. D.; Dixon, D.A. A Local Density Functional Study of the Structure and Vibrational Frequencies of Molecular Transition-Metal Compounds. *J. Phys. Chem.* **1992**, *96*, 6630-6636.
- ³² Pople, J.A.; Seeger, R.; Krishnan, R. Variational Configuration Interaction Methods and Comparison with Perturbation Theory. *Int. J. Quantum Chem.* **1977**, *12*, 149-163.
- ³³ Raghavachari, K.; Schlegel, H. B.; Pople, J. A. Derivative Studies in Configuration-Interaction Theory," *J. Chem. Phys.* **1980**, *72*, 4654-4655.
- ³⁴ Raghavachari, K.; Pople, J. A. Calculation of One-Electron Properties using Limited Configuration-Interaction Techniques. *Int. J. Quantum Chem.* **1981**, *20*, 1067-1071.
- ³⁵ Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618-622.
- ³⁶ Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. Electron Correlation Theories and Their Application to the Study of Simple Reaction Potential Surfaces. *Int. J. Quantum Chem.* **1978**, *14*, 545-560.
- ³⁷ Bartlett, R.J.; Musial, M. Coupled-cluster Theory in Quantum Chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291-352.
- ³⁸ Bartlett, R. J. To Multireference or Not to Multireference: That is the Question? *Int. J. Mol. Sci.* **2002**, *3*, 579-603.
- ³⁹ Purvis III, G. D.; Bartlett, R. J. A Full Coupled-Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples. *J. Chem. Phys.* **1982**, *76*, 1910-1918.
- ⁴⁰ Raghavachari, K.; Trucks, G. W. Pople, J. A. Head-Gordon, M. A Fifth-order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.* **1989**, *157*, 479-483.

-
- ⁴¹ Watts, J. D.; Gauss, J.; Bartlett, R. J. Coupled-Cluster Methods with Noniterative Triple Excitations for Restricted Open-Shell Hartree-Fock and other General Single Determinant Reference Functions. Energies and Analytical Gradients. *J. Chem. Phys.* **1993**, *98*, 8718-8733.
- ⁴² Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gn Theory. *WIREs Comput. Mol. Sci.* **2011**, *1*, 810-825.
- ⁴³ Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. Gaussian-3 Theory using Reduced Moller-Plesset Order. *J. Chem. Phys.* **1999**, *110*, 4703-4709.
- ⁴⁴ Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. Gaussian-3 (G3) Theory for Molecules Containing First and Second-Row Atoms. *J. Chem. Phys.* **1998**, *109*, 7664-7776.
- ⁴⁵ Pople, J. A.; Head-Gordon, N.; Raghavachari, K. Quadratic Configuration Interaction. A general Technique for Determining Electron Correlation Energies. *J. Chem. Phys.* **1987**, *87*, 5968-5976.
- ⁴⁶ The G3MP2Large basis set, G3(MP2) total energies of the molecules, and a complete tabulation of deviations with experiment can be downloaded from the website <http://chemistry.anl.gov/compmat/g3theory.htm>.
- ⁴⁷ Grev, R. S.; Janssen, C. L.; Schaefer, H. F. Concerning Zeropoint Vibrational Energy Corrections to Electronic Energies. *J. Chem. Phys.* **1991**, *95*, 5128-5132.
- ⁴⁸ Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864-B870.
- ⁴⁹ Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133-A1138.
- ⁵⁰ Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200-1211.
- ⁵¹ Perdew, J. P.; Wang, Y. Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy. *Phys. Rev. B* **1992**, *45*, 13244-13249.
- ⁵² Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. Atoms, Molecules, Solids, and Surfaces: Application of the Generalized Gradient Approximation for Exchange and Correlation. *Phys. Rev. B* **1992**, *46*, 6671-6687.
- ⁵³ Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098-3100.

-
- ⁵⁴ Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865-3868.
- ⁵⁵ Becke, A. D. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372-1377.
- ⁵⁶ Becke, A. D. Density-Functional Thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- ⁵⁷ Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Physical Review B.* **1988**, *37*, 785-789.
- ⁵⁸ Cohen, A. J.; Mori-Sanchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289-320.
- ⁵⁹ Gutowski, K. E.; Dixon, D. A. Ab Initio Prediction of the Gas- and Solution-Phase Acidities of Strong Bronsted Acids: The Calculation of pK(a) Values less than -10. *J. Phys. Chem. A* **2006**, *110*, 12044-12054.
- ⁶⁰ McQuarrie, D. A. *Statistical Mechanics*. University Science Books: Sausalito, CA, **2000**.
- ⁶¹ Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Methods. *Chem. Rev.* **2005**, *105*, 2999-3093.
- ⁶² Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc. Perkin Trans.* **1993**, *2*, 799-805.
- ⁶³ 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.
- ⁶⁴ Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *J. Chem. Phys.* **1997**, *106*, 1063-1079.
- ⁶⁵ Hehre, W. J.; Radom, L.; Pople, J. A. Molecular Orbital Theory of the Electronic Structure of Organic Compounds. *J. Am. Chem. Soc.* **1970**, *92*, 4796-4801.
- ⁶⁶ Feller, D.; Dixon, D. A. Extended Benchmark Studies of Coupled Cluster Theory through Triple Excitations. *J. Chem. Phys.* **2001**, *115*, 3484-3496.
- ⁶⁷ 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. On the Enthalpy of Formation of Hydroxyl Radical and Gas-Phase Bond Dissociation Energies of Water and Hydroxyl. *J. Phys. Chem. A* **2002**, *106*, 2727-2747.

⁶⁸ Dixon, D. A.; Feller, D.; Peterson, K. A. Heats of Formation and Ionization Energies of NH_x, x=0-3. *J. Chem. Phys.* **2001**, *115*, 2576-2581.

⁶⁹ Pollack, L.; Windus, T. L.; de Jong, W. A.; Dixon, D. A. Thermodynamic Properties of the C₅, C₆, and C₈ *n*-Alkanes from Ab Initio Electronic Structure Theory. *J. Phys. Chem. A* **2005**, *109*, 6934-6938.

⁷⁰ Feller, D.; Peterson, K. A.; Dixon, D. A. A Survey of Factors Contributing to Accurate Theoretical Predictions of Atomization Energies and Molecular Structures. *J. Chem. Phys.* **2008**, *129*, 204015-204046.

CHAPTER 2: FUNDAMENTAL THERMOCHEMICAL PROPERTIES OF AMINO ACIDS: GAS-PHASE AND AQUEOUS ACIDITIES AND GAS-PHASE HEATS OF FORMATION

2.1 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.^{1,2,3,4,5,6} Bowie and coworkers^{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.^{12,13,14} Thermodynamic values, such as the gas-phase acidity (GA or ΔG_{acid}), the ΔG for the deprotonation reaction $\text{AH} \rightarrow \text{A}^- + \text{H}^+$, 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.^{15,16,17}

A range of studies on positive ions of the amino acids have been performed.^{18,19,20,21,22,23,24,25,26,27} For example, the proton affinities ($\text{B} + \text{H}^+ \rightarrow \text{BH}^+$) of the amino acids have been reported at the G2MP2 level²⁵ 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 McIver²⁸ used the proton transfer equilibrium method to measure the GAs of glycine and alanine.²⁹ Kass and coworkers³⁰ 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 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 coworkers³² included glycine in their proton transfer equilibrium measurements of the GAs of ninety-six aliphatic carboxylic acids. Bowie and coworkers³³ 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. We³⁴ 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 G3MP2³⁵ 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 coworkers³⁷ determined the acidities of glutamic and aspartic acid by extended kinetic methods and noted potential issues with this approach. Poutsma and co-workers³⁸ 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-workers⁴⁰ 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 carboxylate anion structure. Subsequently, Oomens and co-workers⁴² 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 coworkers^{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) calculations⁴⁵ with the COSMO parameterization⁴⁶ 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 5 so-called⁴⁷ common rare amino acids were calculated.

2.2 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 CO₂H group from those used in the proton affinity study, they are similar enough that we can exploit the previous conformational sampling.²⁷

We have shown³⁶ that calculations at the MP2/CBS (CBS = complete basis set) level with the augmented correlation-consistent basis sets up through the quadruple level⁵³ 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) method⁵⁴ 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(CH₃CO₂H) 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 non-bonded 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 CO₂H 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 Gaussian09 as shown in the Appendix. 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 NH₂ 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 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 calculations⁴⁵ using the COSMO parameterization.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 non-electrostatic 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 Equation 1:

$$\text{p}K_{\text{a}}' = \text{p}K_{\text{a}}(\text{HA}) + \Delta G_{\text{aq}}/(2.303RT) \quad (2.1)$$

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 pK_a 's.³⁶

The heat of formation of glycine was calculated by using a composite approach⁵⁶ developed in our group and Washington State University, which is based on CCSD(T) calculations^{57,58,59,60} 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 (Equation 2):⁶²

$$E(n) = E_{\text{CBS}} + A \exp[-(n-1)] + B \exp[-(n-1)^2] \quad (2.2)$$

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 Hamiltonian⁶⁴ for the CISD (configuration interaction with single and double excitations) wavefunction 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 Equation 3.

$$\Sigma D_{0,0\text{K}} = \Delta E_{\text{CBS}} + \Delta E_{\text{Rel}} + \Delta E_{\text{CV}} + \Delta E_{\text{ZPE}} + \Delta E_{\text{SO}} \quad (2.3)$$

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, 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.

2.3 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 CO₂H 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 CO₂H group led to a more stable conformer of the carboxylate ion. Removal of a proton from the methyl side chain or the main chain NH₂ groups led to proton transfer to generate less stable carboxylate ions. For asparagine, removal of a proton from the main chain NH₂ group gave a carboxylate ion which was more stable than the one formed by removal of a proton from the main chain CO₂H group. For glutamine, removal of a proton from the main chain CO₂H 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 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 CO₂H 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 NH₂ and the N in the imidazole ring ($r(\text{N}\dots\text{H}) = 2.17$ 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 CO₂⁻ group. The lowest energy structure of the anion has an N-H distance of 1.06 Å and the higher energy conformer has an 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 NH₂ 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 CO₂H 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 CH₂ in the τ tautomer led to a CH⁻ anion with strong hydrogen bonding between CO₂H and the N in the ring ($r(\text{N}\dots\text{H}) = 1.79$ 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.

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 $\Delta H_{298} = 338.3$ kcal/mol. Removal of a proton from the side chain OH gave a phenoxide anion with a similar acidity, $\Delta H_{298} = 340.5$ 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 values⁷⁰ 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 study³⁰ 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 Appendix, 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 CO₂ 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 CO₂ 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 CO₂H 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 towards 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 coworkers³⁰ where there was not a hydrogen between the SH and the O and instead the SH is hydrogen bonded to the NH₂ 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 NH₂...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. 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 T₁ 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 wavefunction 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 coworkers,^{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 CO₂H and main chain NH₂ (1.73 Å) and between the main chain NH₂ and the side chain N (1.73 Å). Deprotonation of a proton from the main chain CO₂H group of the neutral [a] led to the most stable conformer of the carboxylate ion. Removal of a proton from the main chain NH₂ group, the side chain NH amino group, and the side chain terminal NH₂ group led to structures similar to that of abstraction of a proton from the main chain CO₂H group. For neutral [b], removal of a proton from the main chain CO₂H 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 NH₂ in neutral [b] led to a NH⁻ anion with strong hydrogen bonding between the main chain CO₂H and main chain NH₂ (1.75 Å) and between the main chain NH₂ 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-workers³⁸ 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 NH₂ 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 $O\cdots H(O)$ 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 $r(O\cdots H(O)) = 1.53$ 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 ($r(O\cdots H(O)) = 2.02$ 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 $r(O\cdots H(O)) = 1.53$ Å). In cysteine, the proton is shared between S and O ($r(OH) = 1.04$ Å and $r(O\cdots H(O)) = 1.94$ Å). In histidine π tautomer, the proton is shared between N in the ring and O ($r(OH) = 1.06$ Å and $r(O\cdots H(O)) = 1.64$ 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 CO₂H 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.

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 CO₂H 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 $r(\text{O}\cdots\text{H}(\text{O}))$ 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 ($r(\text{OH}) = 1.03$ Å and $r(\text{O}\cdots\text{H}(\text{O})) = 1.65$ 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.^{71,72,73,74}) 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 experimental⁷⁶ 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 parameterization and the gas phase non-zwitterion structures can be compared to the experimental⁷⁶ values for the zwitterions in aqueous solution. These values correspond to loss of a proton from the main chain ammonium group in the neutral zwitterionic form ($NH_3^+CHRCO_2^-$) 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 non-zwitterion 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 non-zwitterionic structure about 5 kcal/mol higher in free energy than the zwitterion. The remaining amino acids have the non-zwitterionic 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 pK_a 's for the non-zwitterionic structures are consistent with the results of Coote and Ho⁷⁷ who used similar approaches to calculate the pK_a 's of carbonic acids for biochemical applications. Sprik and co-workers⁷⁵ 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 pK_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 non-zwitterionic 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 pK_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).



This leads to the energy expression given in Equation 5

$$\Delta H_f(\text{reaction}) = \Delta H_f(\text{product}) + \Delta H_f(\text{glycine}) + \Delta H_{\text{rxn}}(\text{MP2}) - \Delta H_f(\text{CH}_4) \quad (5)$$

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 ($\Delta H_f^{298} = -92.6$ 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 also within 2 kcal/mol of the available experimental values^{66,78,79,80}. The calculated heats of formation from the isodesmic reactions are within 2 kcal/mol of the available experimental values. Sagadeev et al.⁸¹ predicted the heats of formation of the amino acids using a parameterized 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 $\Delta H_f(\text{histidine})$ to be too positive by ~ 12 kcal/mol and $\Delta H_f(\text{tryptophan})$ to be too positive by ~ 8 kcal/mol.

Heats of formation of the 5 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-methyllysine.

2.4 Conclusions 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 $-\text{CO}_2\text{H}$ acid group to form the $-\text{CO}_2^-$ 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 $-\text{CO}_2\text{H}$ 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 $\text{p}K_a$ values were used to estimate the free energy difference between the zwitterion and non-

zwitterion forms in solution. Asparagine, lysine, phenylalanine, proline, and tyrosine have the non-zwitterionic structure about 5 kcal/mol higher in free energy than the zwitterions, whereas the zwitterionic and nonzwitterionic 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.

Acknowledgement 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. Dixon 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.

Table 2.1. G3MP2 Acidities of the Amino Acids in kcal/mol.

Amino Acid	ΔH_{298}	ΔG_{298}	ΔH_{298} calc ³⁸	ΔH_{298} expt ³⁸	ΔH_{298} expt ³³
Alanine	342.3	334.6	342.3	341.8 \pm 1.9	340.6 \pm 2.1 ²⁸
Arginine a A	330.1	322.0	331.5	330.1 \pm 2.2	332.0 \pm 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 \pm 2.2	331.7 \pm 3.1
Asparagine B	339.9	332.7			
Aspartic Acid A ³⁴	322.4	315.4	321.5	321.5 \pm 3.3	320.3 \pm 1.4 ³⁷
Aspartic Acid B ³⁴	333.1	325.9			
Cysteine A	334.4	327.1	333.7 ³⁰	333.4 \pm 2.2 ³⁰	332.9 \pm 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 \pm 5.0	322.7 \pm 1.4 ³⁷
Glutamic Acid B ³⁴	330.8	324.3			
Glutamine A	328.0	321.9	329.3	331.0 \pm 2.6	331.7 \pm 3.1
Glutamine B	351.9	345.8			
Glycine ³⁴	342.9	335.3	342.7	342.7 \pm 2.2	342.5 \pm 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 \pm 2.6	331.0 \pm 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 \pm 1.9	338.9 \pm 3.1
Leucine A	340.3	333.5	341.3	339.1 \pm 2.4	339.1 \pm 3.1
Leucine B	384.7	377.5			
Lysine A WRT neutral a	336.9	329.9	338.2	338.4 \pm 1.7	337.5 \pm 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			

Table 2.2. 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	
	ΔΔH	0.2	2.2	2.0	2.4	1.8	1.8	1.3	1.9	
	ΔΔG	0.3	2.5	2.1	2.7	1.7	1.6	1.1	1.7	
Cys COO ⁻ anion [B]	ΔH	335.8	336.6	336.4	336.3	336.3	336.6 ^b	333.7 ^c	336.3 ^c	
	ΔG	327.7	328.2	328.3	328.0	328.4	328.2 ^b	325.4 ^c	328.1 ^c	
Cys COO ⁻ anion [C]	ΔH	336.4	336.3	336.6	336.0	d	336.5 ^b	333.2/333.3 ^c	336.1/336.1 ^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]	ΔΔH	-0.6	0.3	-0.2	0.3	d	0.1	0.4 ^c	0.2 ^c
	[B] - [C]	ΔΔG	-1.0	-0.4	-0.7	-0.3	d	-0.6	-0.2 ^c	-0.3 ^c
Cysteine S ⁻ anion [A]	ΔH	333.8	334.4	333.2	334.0	333.6	334.1 ^b	330.1/330.1 ^c	332.9/333.0 ^c	
	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]	ΔΔH	-2.6	-1.9	-3.4	-2.0	d	-2.5	-3.1/-3.2 ^c	-3.2/-3.1 ^c
	[A] - [C]	ΔΔG	-1.9	-1.5	-2.9	-1.6	d	-2.0	-2.8/-2.8 ^c	-2.7/-2.7 ^c
	[A] - [B]	ΔΔH	-2.0	-2.2	-3.2	-2.3	d	-2.6	-3.6 ^c	-3.6 ^c
[A] - [B]	ΔΔG	-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.

Table 2.3. 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

Table 2.4. 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			

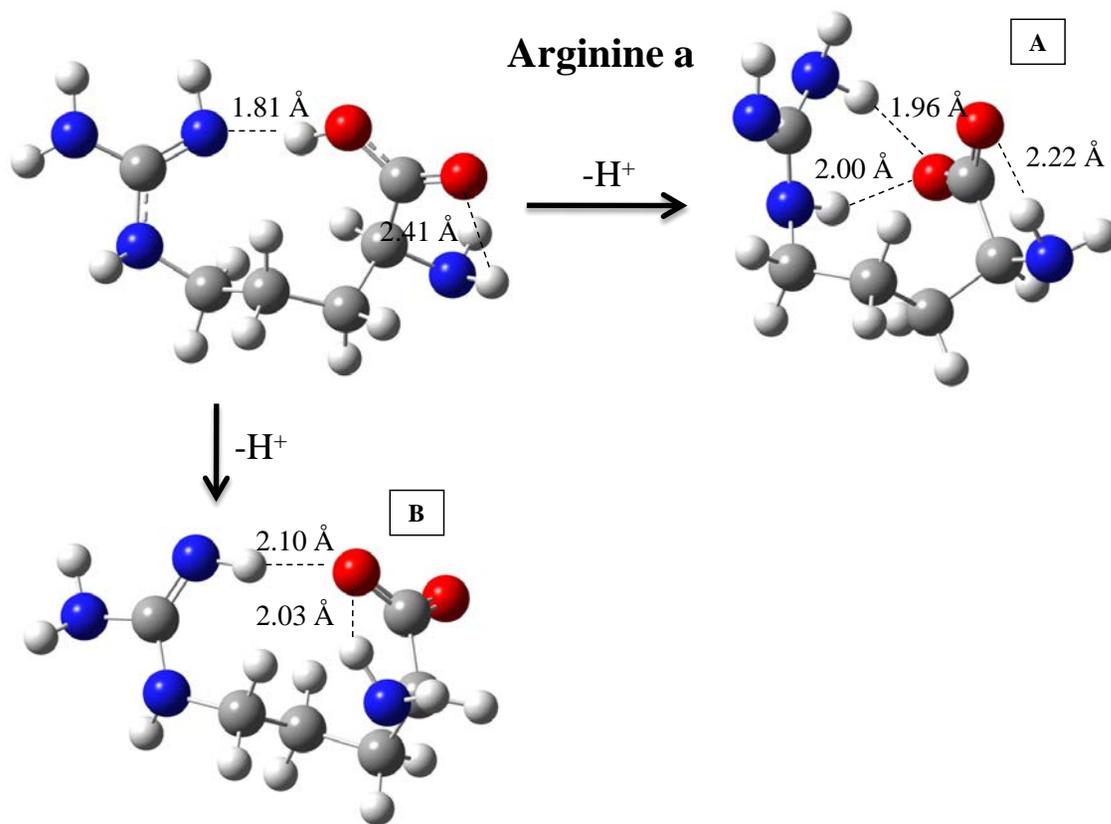
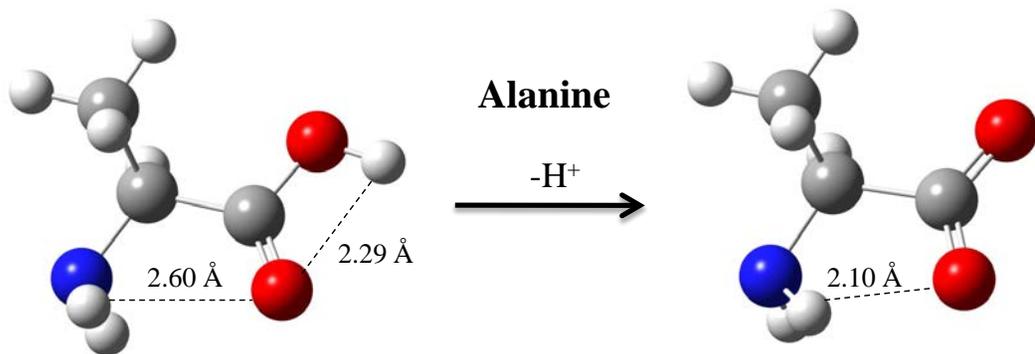
Table 2.5. Calculated Heats of Formation at the G3MP2 level of the Amino Acid in kcal/mol.

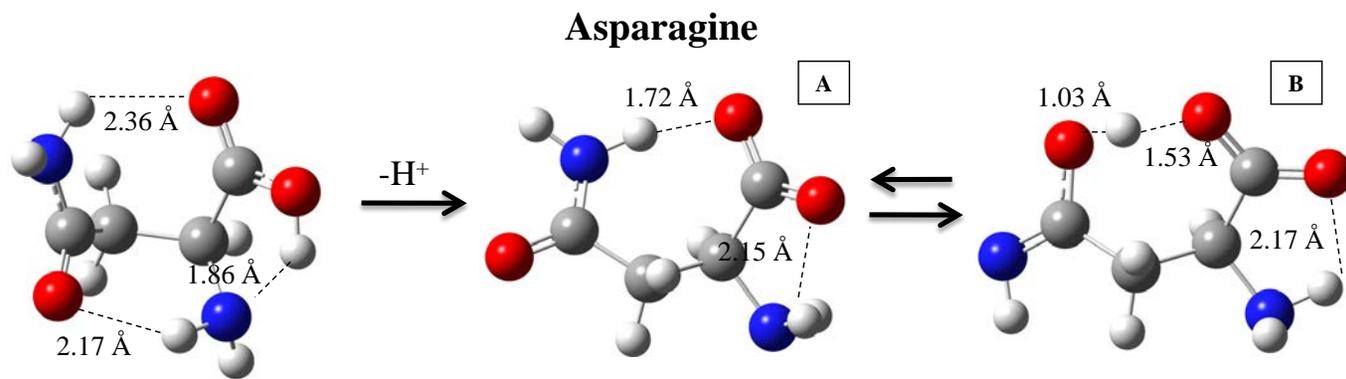
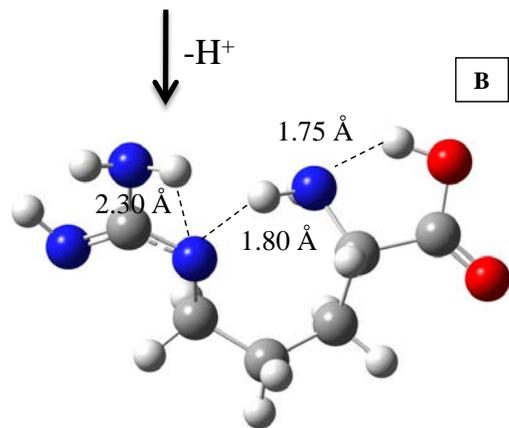
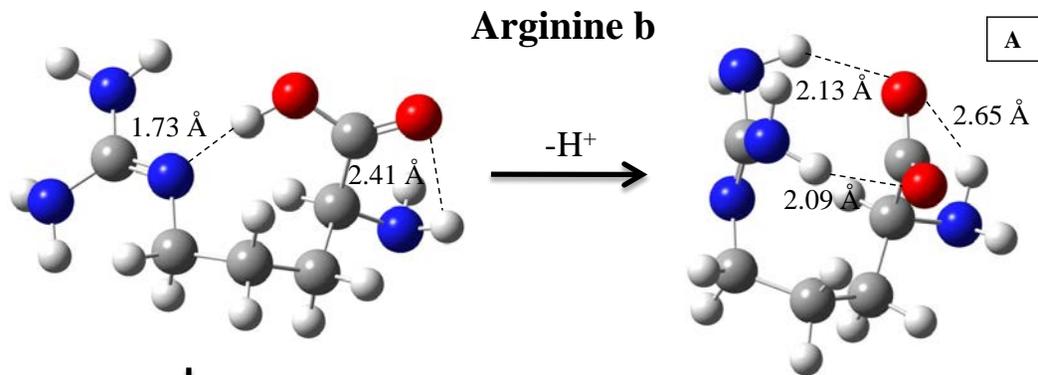
Amino Acid	ΔH_f 0 K	ΔH_f 298 K TAE	ΔH_f 298 K calc ⁸¹	Isodesmic product	ΔH_f 298 K Isodesmic	ΔH_f 298 K Experiment
Alanine	-94.8	-100.7	-99.4	ethane	-100.2	-99.1 ± 1.0 ⁷⁸
Arginine a	-80.1	-93.0				
Arginine b	-81.8	-94.6				
Asparagine	-136.7	-144.4	-141.1	propanamide	-145.8	
Aspartic Acid	-182.1	-188.9	-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

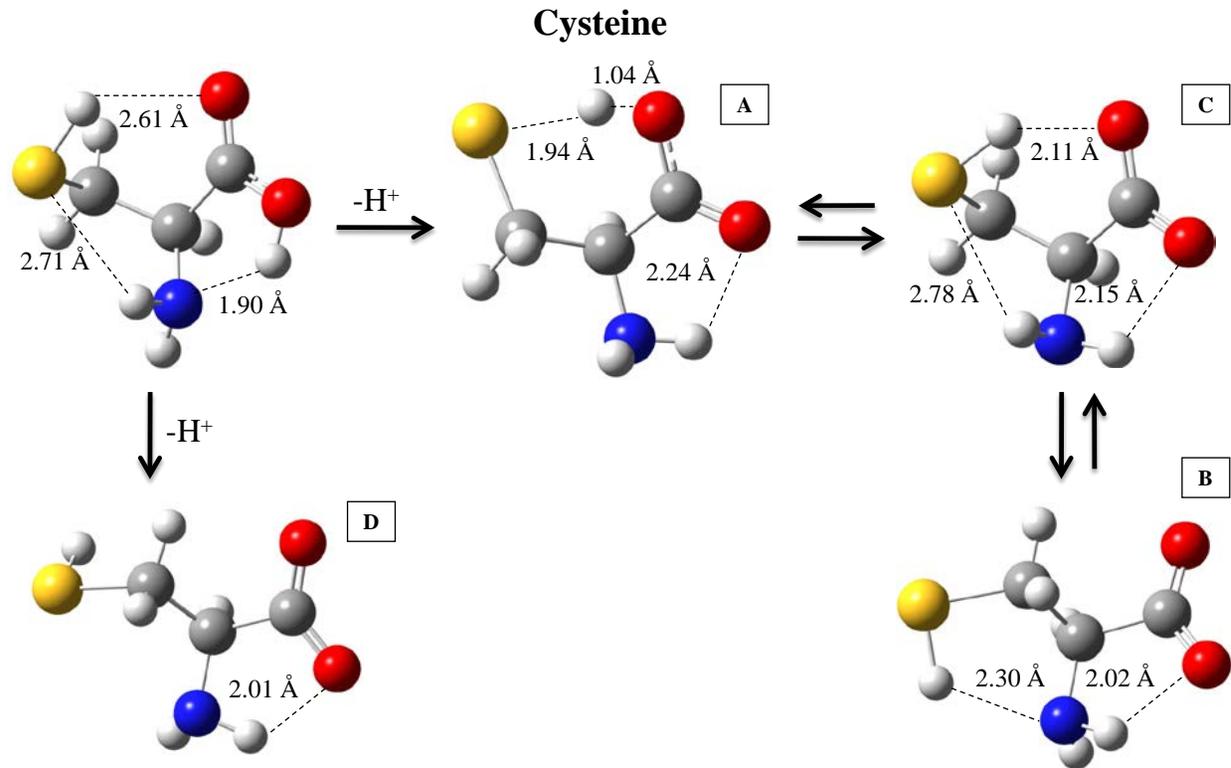
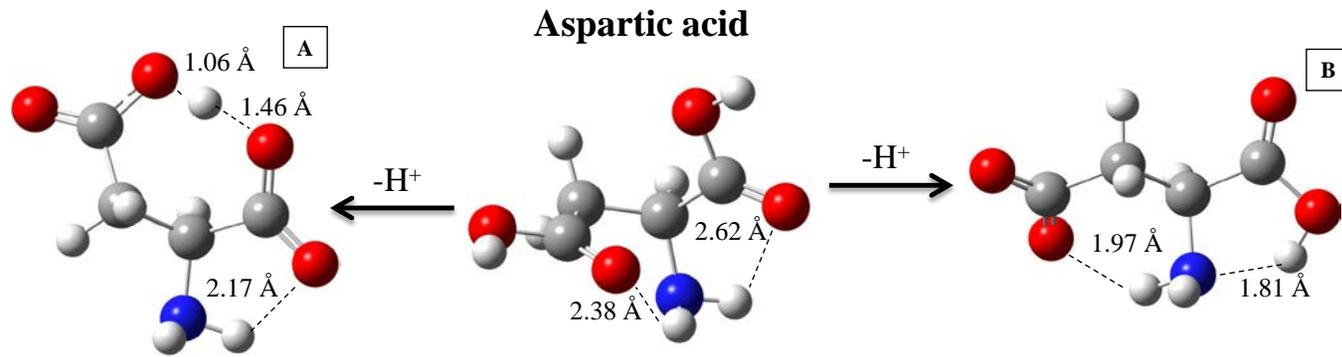
^a The products are pyrrolidine and acetic acid, not glycine

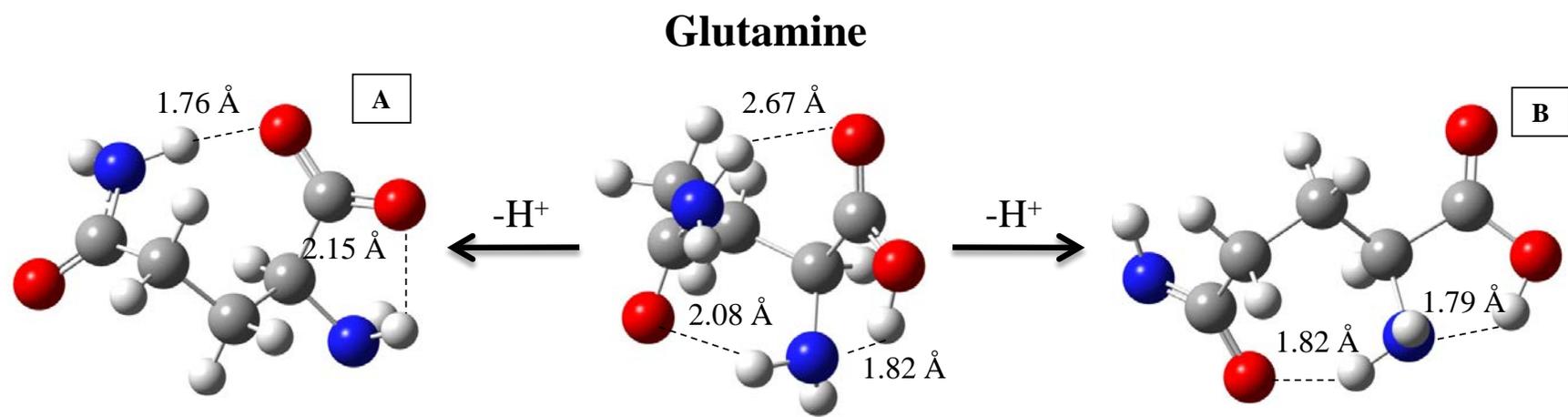
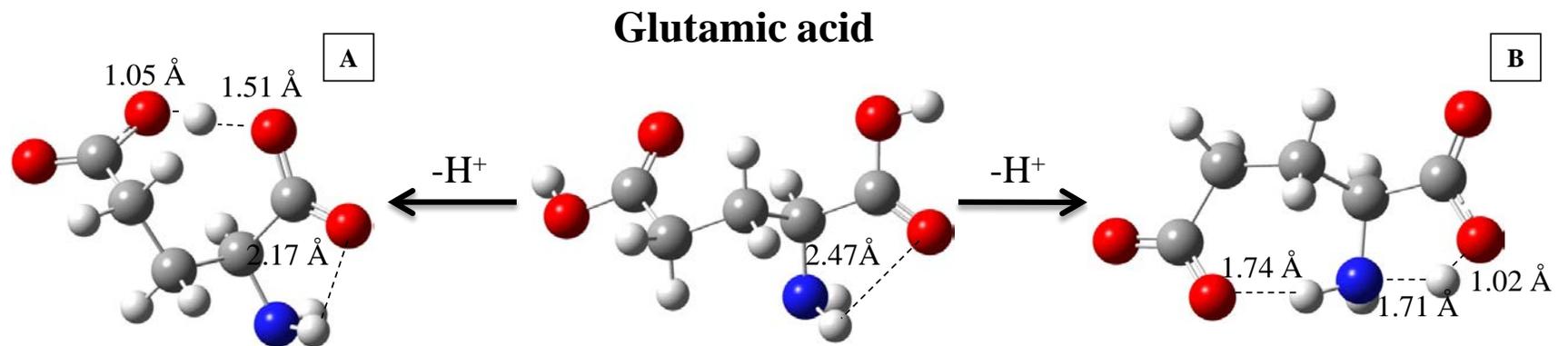
Table 2.6. Calculated Heats of Formation at the G3MP2 level of the 5 Rare Amino Acids in kcal/mol.

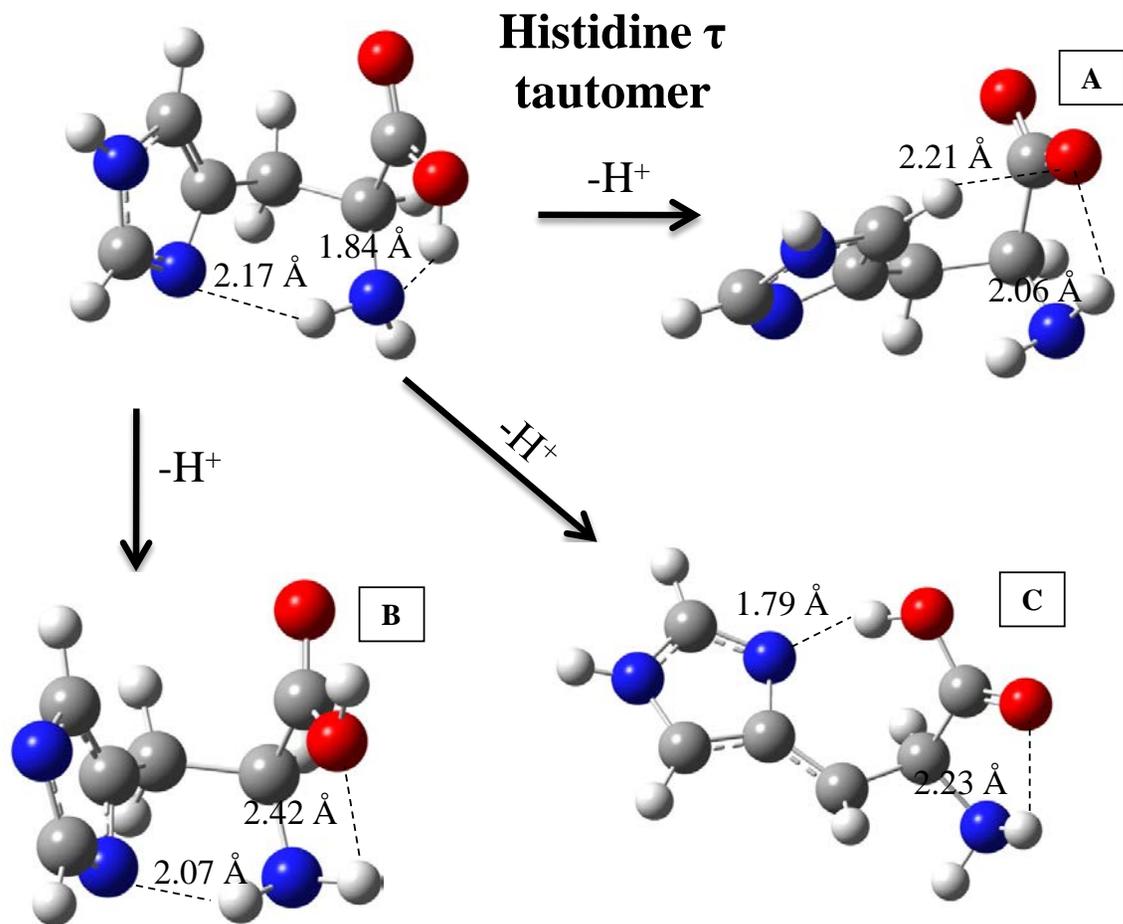
Amino Acid	ΔH_f 0 K	ΔH_f 298 K TAE	ΔH_f 298 K calc ⁸¹	ΔH_f 298 K Isodesmic
3-Methylhistidine	-59.7	-69.6		-68.5
4-Hydroxyproline	-120.5	-128.8	-129.2	-129.4
5-Hydroxylysine	-135.6	-147.7	-145.9	-148.3
N-Methyllysine	-93.2	-106.1		-106.1
Pyroglutamic acid	-124.9	-131.7		



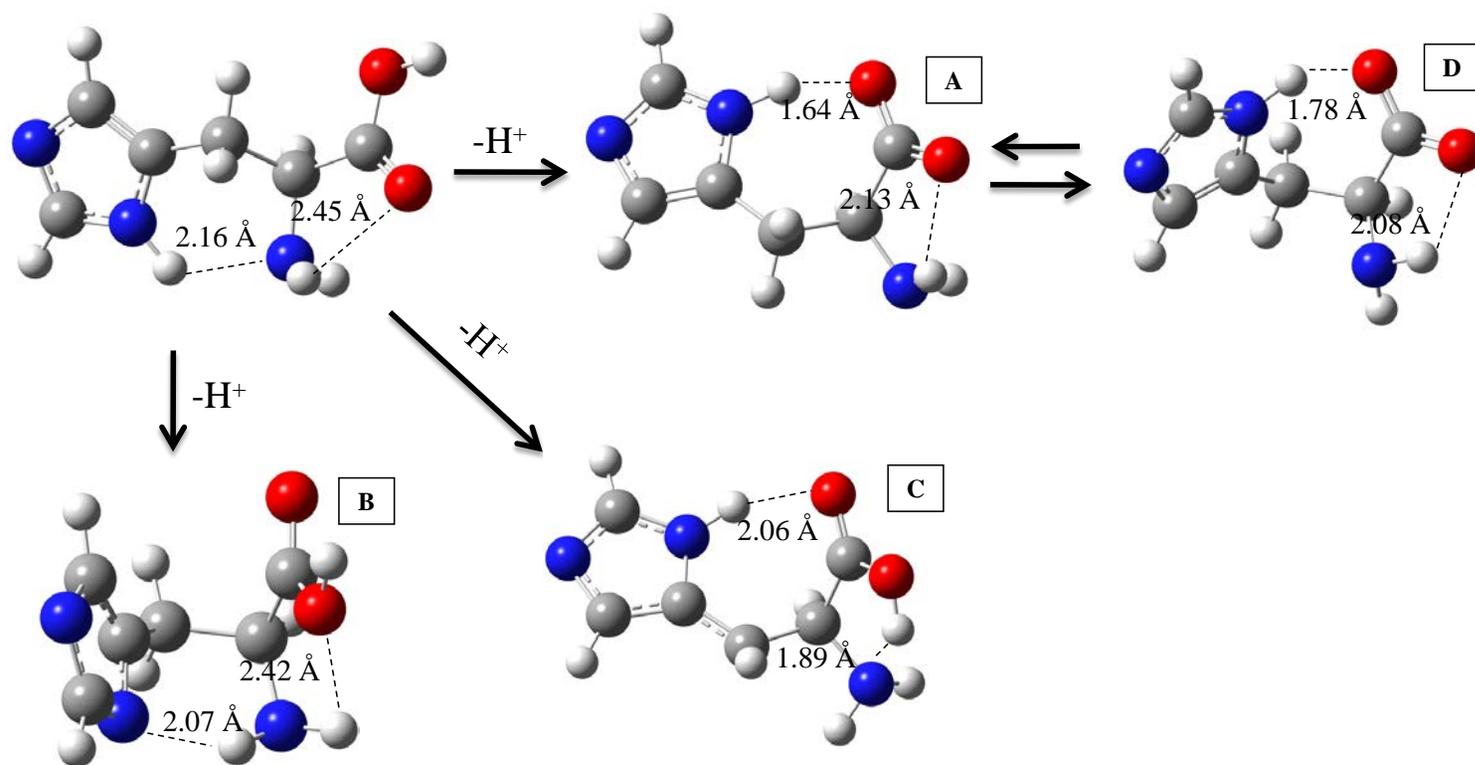




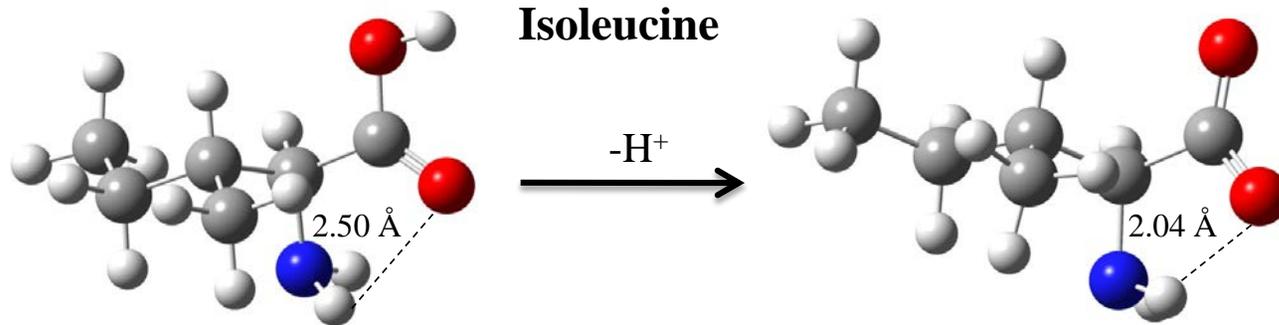




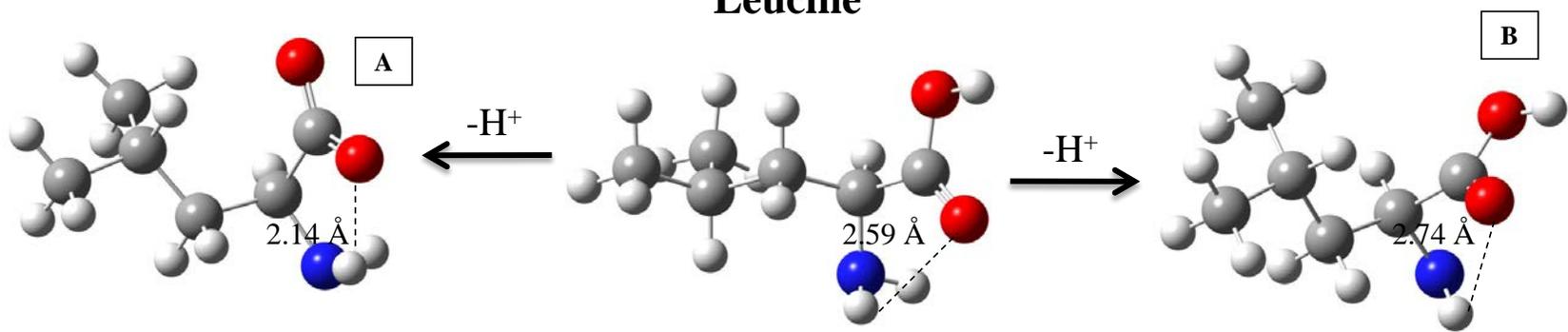
Histidine π tautomer



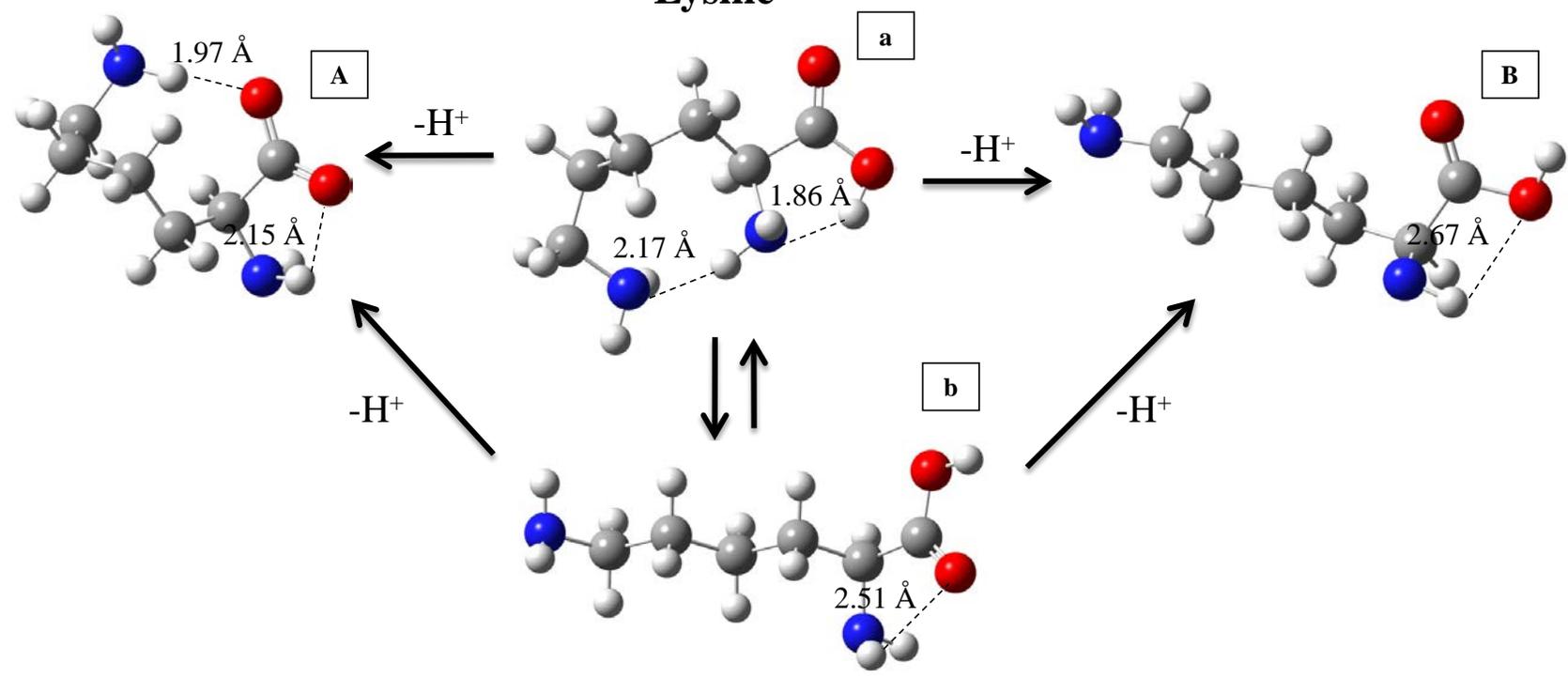
Isoleucine



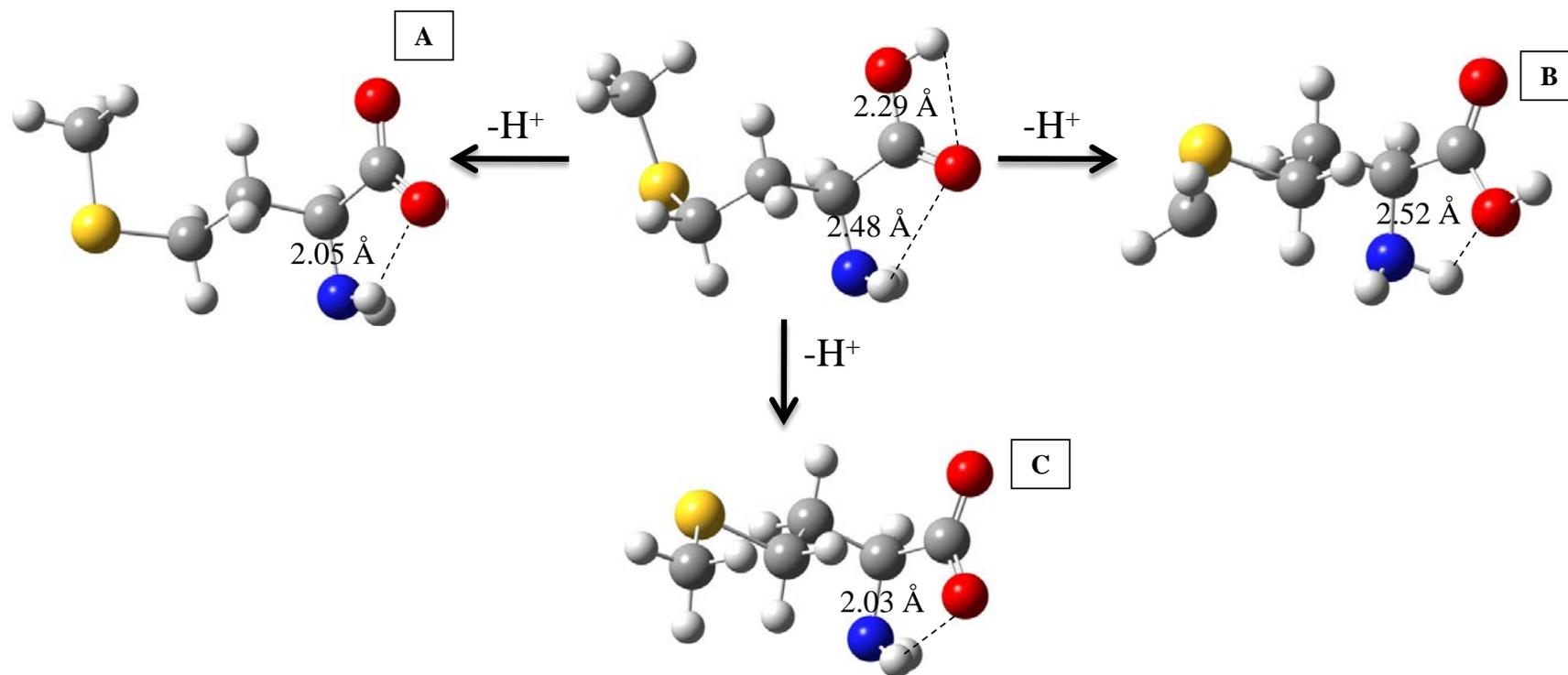
Leucine

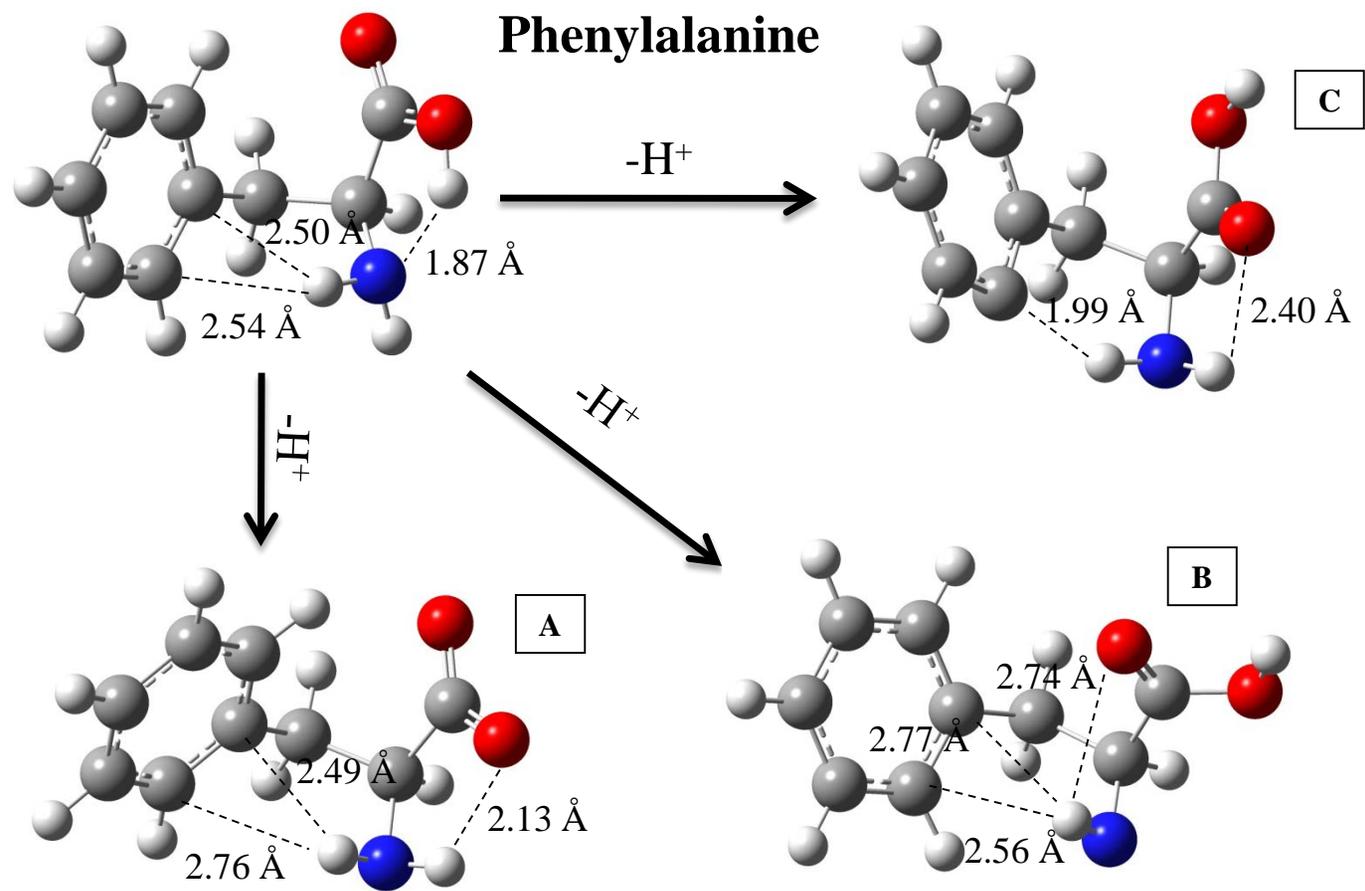


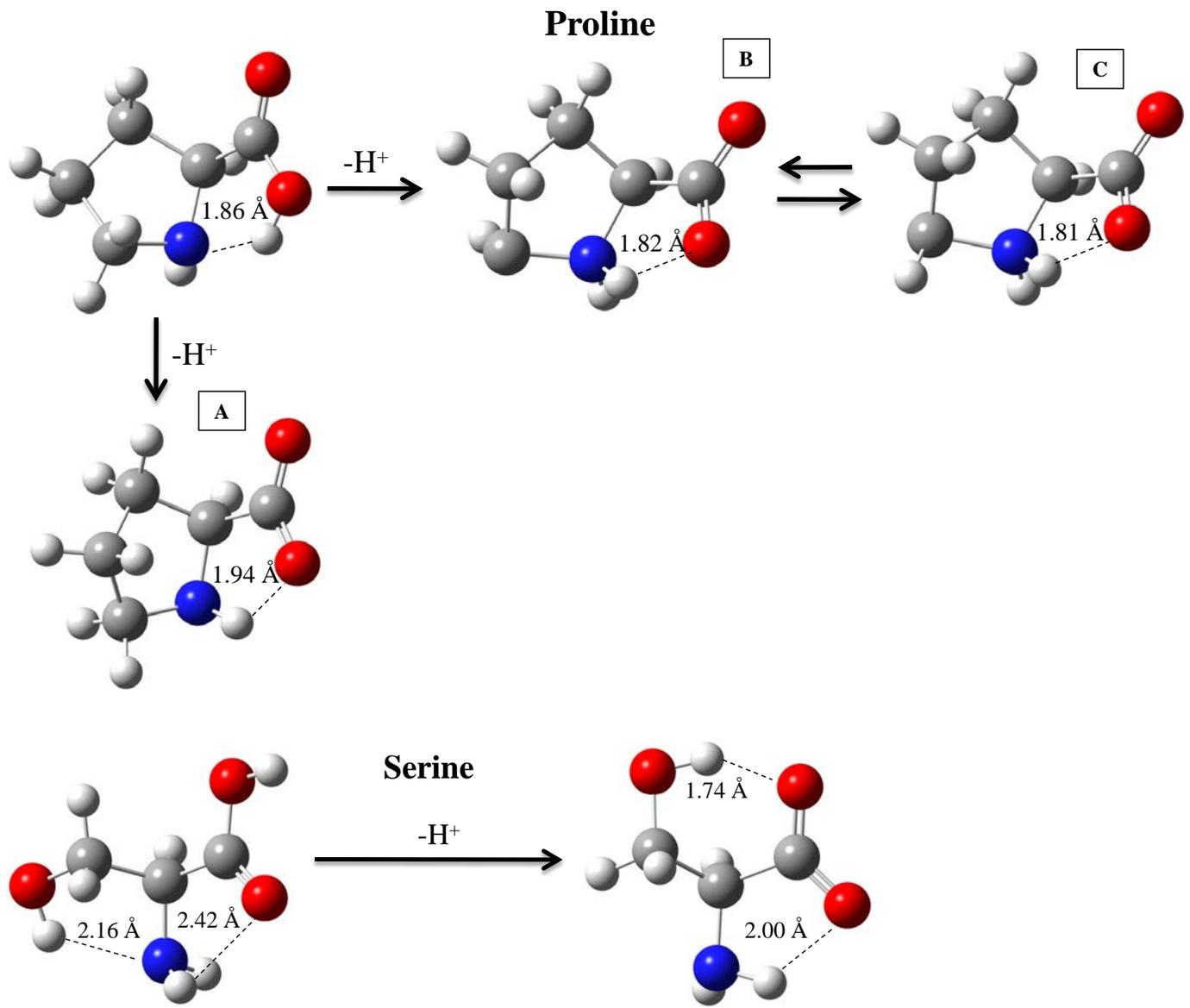
Lysine

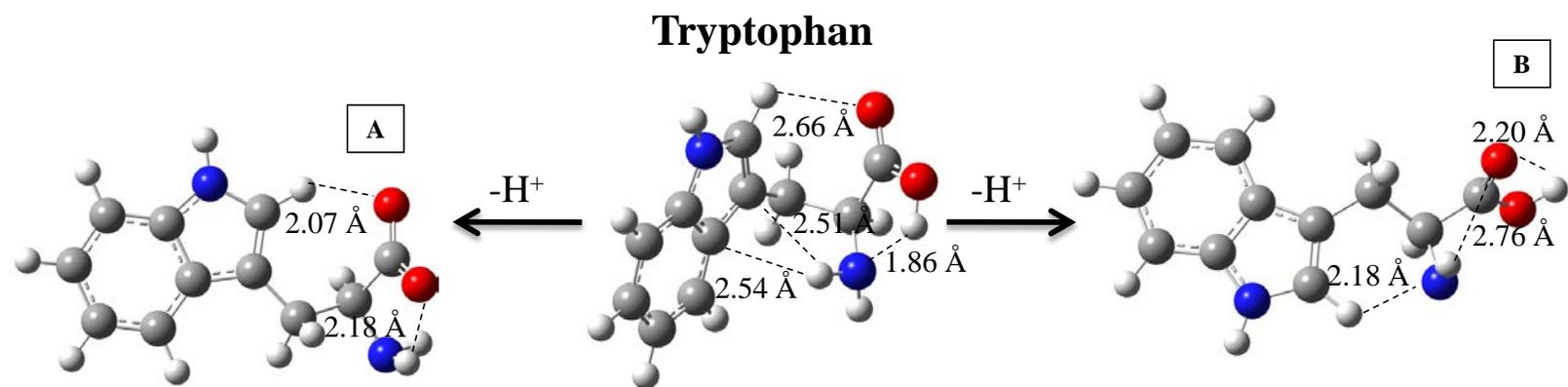
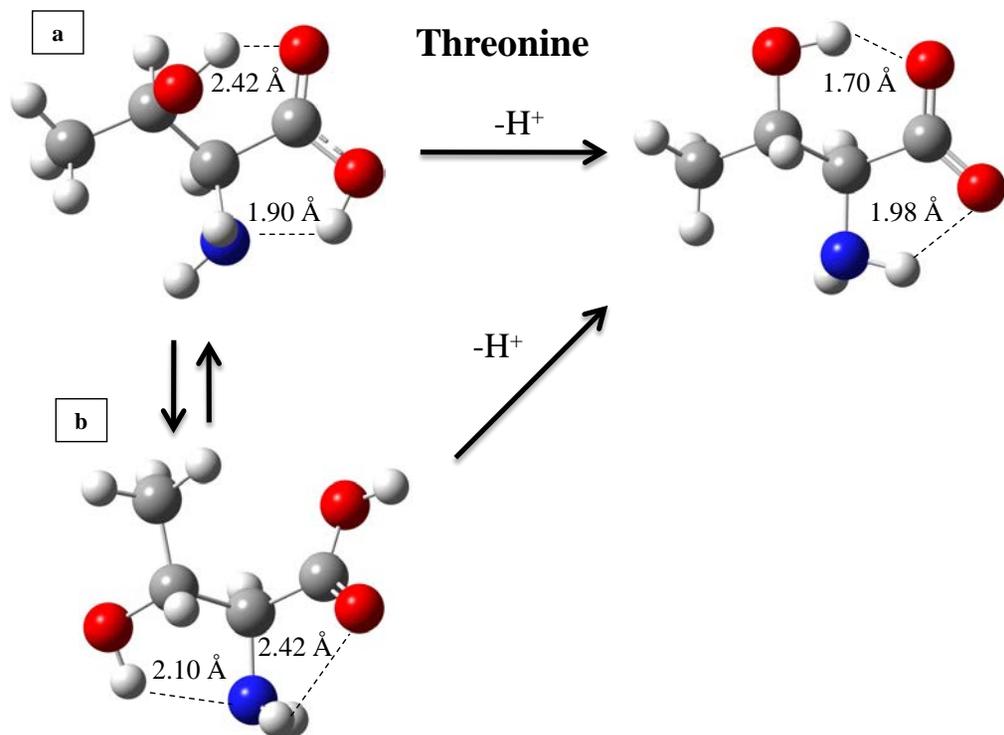


Methionine

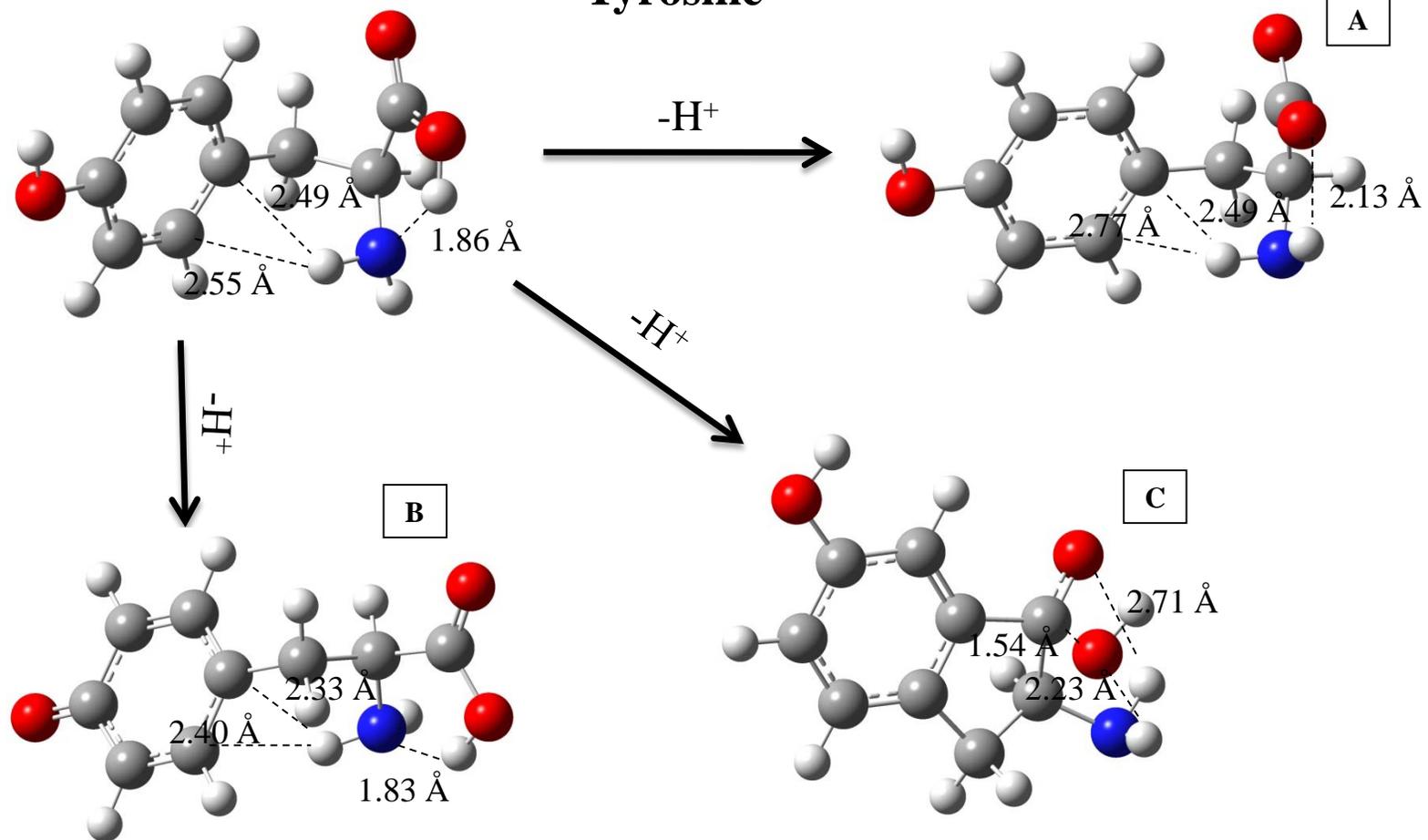








Tyrosine



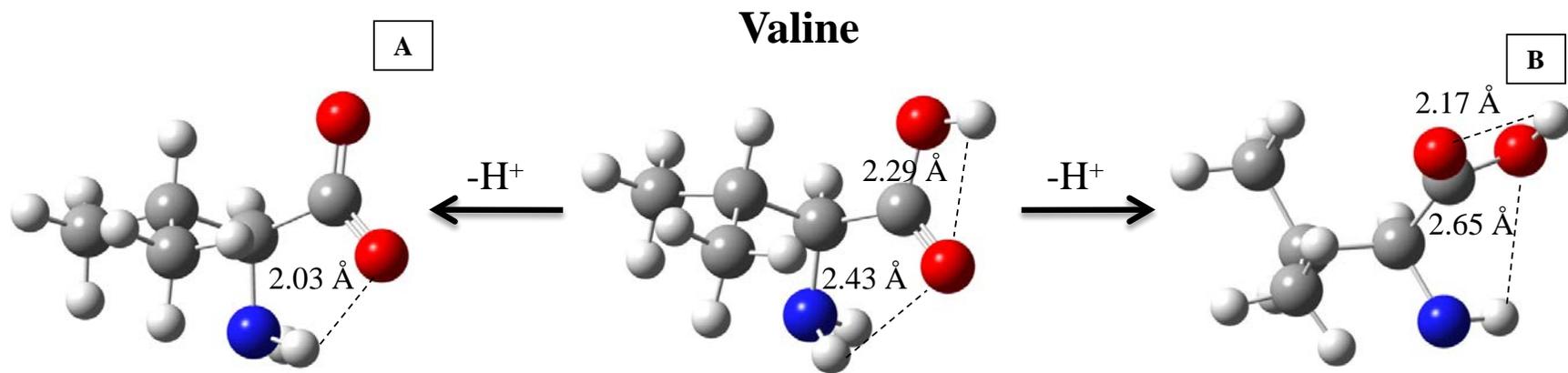
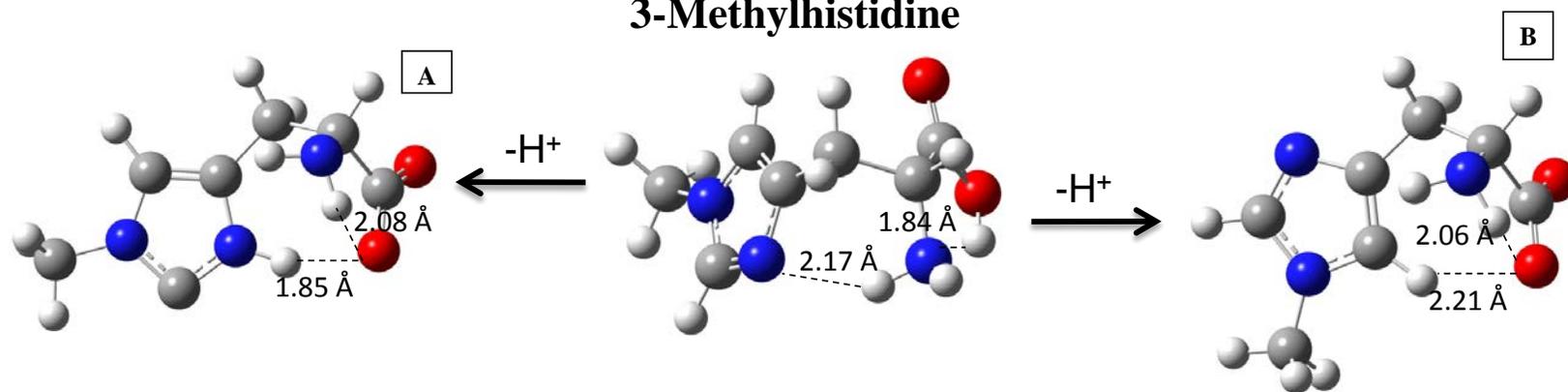
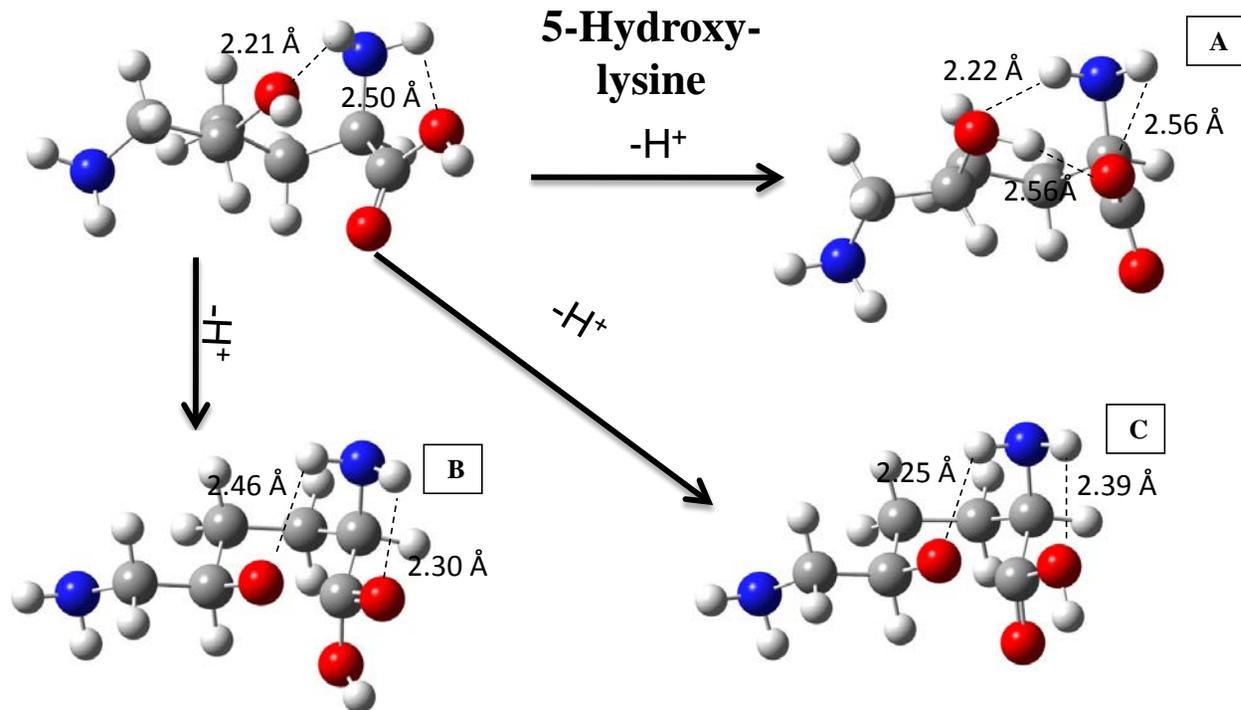


Figure 2.1. Optimized structures of the amino acids and their lowest energy anions at the G3MP2 (MP2(full)/6-31G(d)) level. Important hydrogen bond distances are given in Å.

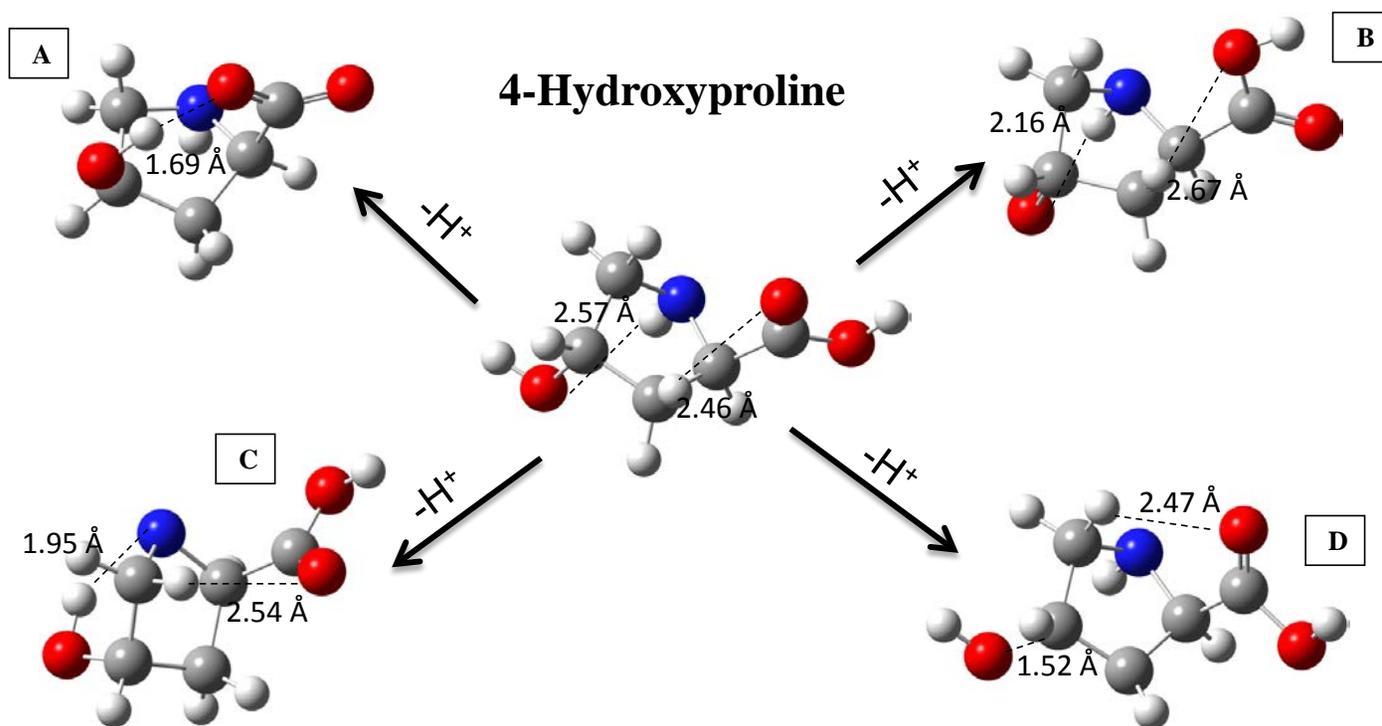
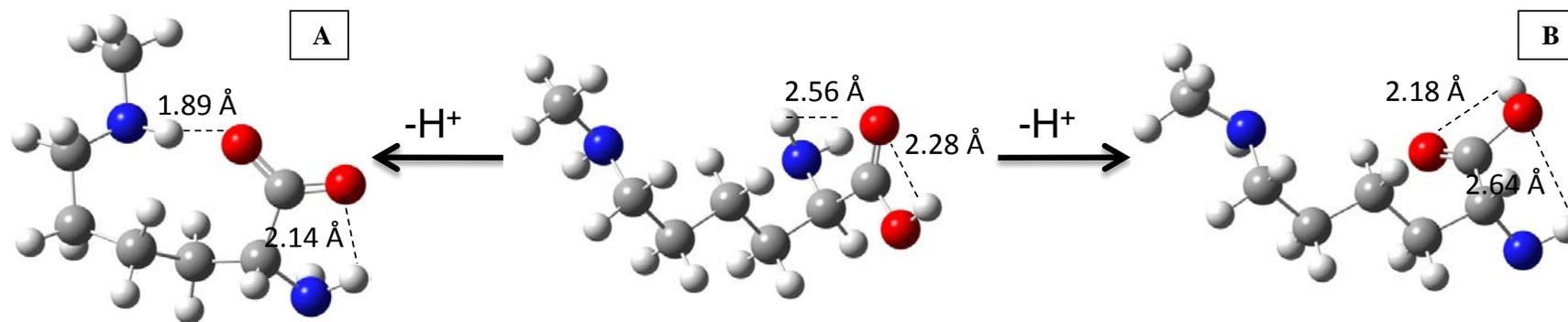
3-Methylhistidine



5-Hydroxy-lysine



N-Methyllysine



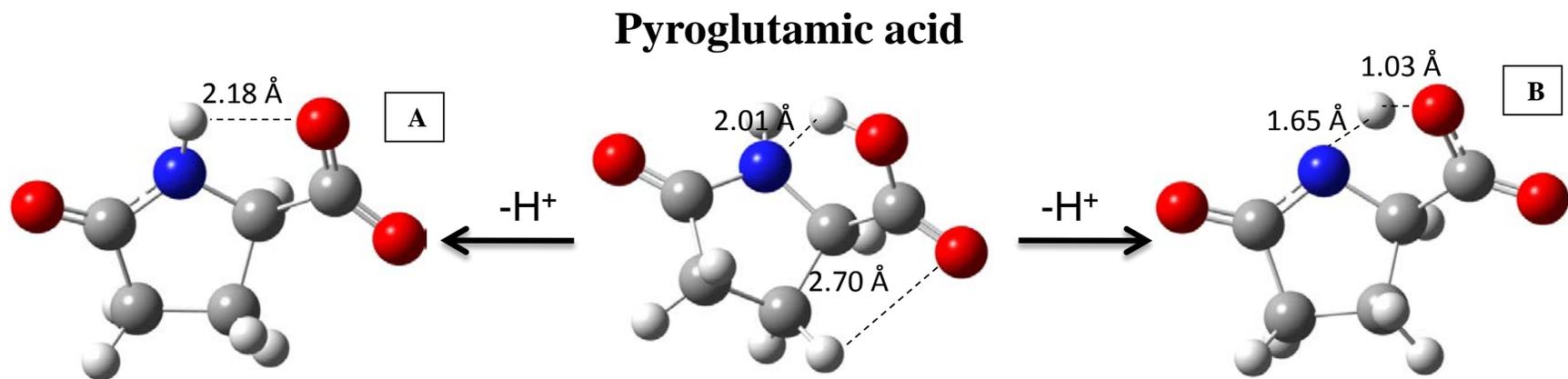


Figure 2.2. Optimized structures of the five “rare” amino acids and their lowest energy anions at the G3MP2 level. Important hydrogen bond distances are given in Å.

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.; Brechi, 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. *Journal of Molecular Structure: 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 A. L. Lehninger, *Biochemistry, 2th 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. *Physical Review 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, 20899, June 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 III, G. D.; 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, Volume I, H to V*; U.S. National Bureau of Standards Circular 467; U.S. Department of Commerce, National Technical Information Service, COM-72-50282: Washington, D C, 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. *Journal of Chemical Physics* **1999**, 110, 7650 – 7657.
- ⁶⁹ Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *Journal of Chemical Physics* **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 Yor, 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.

CHAPTER 2 APPENDIX: FUNDAMENTAL THERMOCHEMICAL PROPERTIES OF AMINO ACIDS: GAS-PHASE AND AQUEOUS ACIDITIES AND GAS-PHASE HEATS OF FORMATION

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 which differ from ref 27, and ΔH_{298} and ΔG_{298} acidity comparison for acetic acid and phenol.

Table A2.1. H_{298} and G_{298} Values for all Neutral Amino Acids and Anions at the G3MP2 level (a.u.).

Amino Acid	H_{298}	G_{298}
Alanine neutral	-323.284633	-323.323491
Alanine anion	-322.741542	-322.780211
Arginine neutral a	-605.637941	-605.691940
Arginine a A anion	-605.114307	-605.168795
Arginine a B anion	-605.098379	-605.153117
Arginine neutral b	-605.640535	-605.695158
Arginine b A anion	-605.100727	-605.154292
Arginine b B anion	-605.079220	-605.133476
Asparagine neutral	-491.781589	-491.827290
Asparagine A anion	-491.255931	-491.301666
Asparagine B anion	-491.242351	-491.287049
Aspartic Acid neutral	-511.649157	-511.694813
Aspartic Acid A anion	-511.137798	-511.182248
Aspartic Acid B anion	-511.120753	-511.165416
Cysteine neutral	-721.041257	-721.083050
Cysteine A anion	-720.510790	-720.551812
Cysteine B anion	-720.507278	-720.550025
Cysteine C anion	-720.507756	-720.549445
Cysteine D anion	-720.504052	-720.546774
Glutamic Acid neutral	-550.882892	-550.933153
Glutamic Acid A anion	-550.372341	-550.418904
Glutamic Acid B anion	-550.358166	-550.406268
Glutamine neutral	-531.013129	-531.063358
Glutamine A anion	-530.492709	-530.540288
Glutamine B anion	-530.454760	-530.502276
Histidine τ tautomer neutral	-547.979128	-548.027307
Histidine τ tautomer A anion	-547.438113	-547.486305
Histidine τ tautomer B anion	-547.432387	-547.480009

Histidine τ tautomer C anion	-547.376763	-547.425841
Histidine π tautomer neutral	-547.977604	-548.026129
Histidine π tautomer A anion	-547.459604	-547.507311
Histidine π tautomer B anion	-547.432387	-547.480009
Histidine π tautomer C anion	-547.384604	-547.432235
Histidine π tautomer D anion	-547.454676	-547.501461
Isoleucine neutral	-440.986550	-441.035695
Isoleucine anion	-440.445771	-440.494006
Leucine neutral	-440.986751	-441.036066
Leucine A anion	-440.446873	-440.494651
Leucine B anion	-440.375983	-440.424482
Lysine neutral a	-496.255944	-496.306510
Lysine neutral b	-496.254548	-496.307806
Lysine A anion	-495.721431	-495.770841
Lysine B anion	-495.643917	-495.696107
Methionine neutral	-799.503878	-799.554933
Methionine A anion	-798.966285	-799.017239
Methionine B anion	-798.880338	-798.931415
Methionine C anion	-798.965965	-799.016997
Phenylalanine neutral	-553.935927	-553.985535
Phenylalanine A anion	-553.398896	-553.448532
Phenylalanine B anion	-553.325721	-553.375661
Phenylalanine C anion	-553.316307	-553.366006
Proline neutral	-400.553954	-400.595426
Proline A anion	-400.013755	-400.054359
Proline B anion	-399.939773	-399.980598
Proline C anion	-399.937218	-399.978101
Serine neutral	-398.426440	-398.467583
Serine anion	-397.898795	-397.938559
Threonine neutral a	-437.666767	-437.710078
Threonine neutral b	-437.666311	-437.710368

Threonine anion	-437.139650	-437.182428
Tryptophan neutral	-685.313123	-685.367614
Tryptophan A anion	-684.773461	-684.828371
Tryptophan B anion	-684.704096	-684.759058
Tyrosine neutral	-629.084721	-629.137110
Tyrosine A anion	-628.547995	-628.600502
Tyrosine B anion	-628.544388	-628.596587
Tyrosine C anion	-628.515061	-628.564214
Valine neutral	-401.753755	-401.799009
Valine A anion	-401.213395	-401.258056
Valine B anion	-401.141748	-401.185964
3-Methylhistidine neutral	-587.210058	-587.262529
3-Methylhistidine A anion	-586.665638	-586.718391
3-Methylhistidine B anion	-586.636000	-586.687174
5-Hydroxylysine neutral	-571.401559	-571.454460
5-Hydroxylysine A anion	-570.874993	-570.926955
5-Hydroxylysine B anion	-570.843360	-570.894180
5-Hydroxylysine C anion	-570.841368	-570.890084
N-Methyllysine neutral	-535.477851	-535.534524
N-Methyllysine A anion	-534.944014	-534.997563
N-Methyllysine B anion	-534.865826	-534.921593
4-Hydroxyproline neutral	-475.696175	-475.740771
4-Hydroxyproline A anion	-475.164889	-475.206875
4-Hydroxyproline B anion	-475.116615	-475.159789
4-Hydroxyproline C anion	-475.105241	-475.147681
4-Hydroxyproline D anion	-475.071982	-475.115862
Pyroglutamic acid neutral	-474.535716	-474.578579
Pyroglutamic acid A anion	-474.014879	-474.057363
Pyroglutamic acid B anion	-473.999180	-474.041002

Table A2.2. H_{298} and G_{298} of Tyrosine and Cysteine at Different Computational Levels (a.u.).

Amino Acid	Prop	B3LYP	G3B3	G3	G4	MP2/aT
Tyrosine	H_{298}	-629.974389	-629.597616	-629.584486	-629.699231	a
neutral	G_{298}	-630.026427	-629.649740	-629.636876	-629.750915	a
Tyr CO_2^-	H_{298}	-629.436718	-629.061006	-629.048116	-629.162500	a
Anion [A]	G_{298}	-629.488757	-629.113007	-629.100625	-629.214024	a
Tyr O^-	H_{298}	-629.436432	-629.057650	-629.044228	-629.159569	a
Anion [B]	G_{298}	-629.488216	-629.109751	-629.096426	-629.211369	a
Cysteine	H_{298}	-721.922011	-721.578556	-721.570310	-721.650110	-720.816779
neutral	G_{298}	-721.963604	-721.620240	-721.612103	-721.691746	-720.857990
Cys COO^-	H_{298}	-721.388275	-721.044436	-721.037221	-721.158456	a
Anion [B]	G_{298}	-721.431417	-721.086983	-721.079434	-721.116572	a
Cys COO^-	H_{298}	-721.389312	-721.044829	-721.036687	b	-720.287957
Anion [C]	G_{298}	-721.429699	-721.085915	-721.078908	b	-720.329086
Cysteine S^-	H_{298}	-721.392463	-721.049958	-721.040362	-721.120776	-720.293120
Anion [A]	G_{298}	-721.432872	-721.090561	-721.081385	-721.161089	-720.333528

^a Only single point MP2/aug-cc-pVTZ were performed

^b Only one structure could be obtained at the G4 level.

Table A2.3. Energy Components to Calculate the H298 Energies of Cysteine and Tyrosine at the MP2 and CCSD(T) levels (a.u.).

Amino Acid	CCSD(T)/aT //B3LYP total energy	E(Therm) at Gn level	E(ZPE) at Gn level	CCSD(T)/aT //MP2/aT total energy	E(Therm) at MP2/aT level	E(ZPE) at MP2/aT level
Tyr neutral	-629.059615	0.203569 ^a	0.191612 ^a	-629.060195	c	c
Tyr CO ₂ ⁻ [A]	-628.509126	0.189834 ^a	0.177897 ^a	-628.509799	c	c
Tyr O ⁻ [B]	-628.505726	0.189154 ^a	0.177374 ^a	-628.506118	c	c
Cys neutral	-721.033841	0.113503 ^b	0.105509 ^b	-721.034412	0.117047	0.109339
					0.116437 ^d	0.108671 ^d
Cys COO ⁻ [B]	-720.486738	0.100391 ^b	0.092167 ^b	-720.487079	0.102699 ^d	0.094778 ^d
Cys COO ⁻ [C]	-720.487022	0.100502 ^b	0.092620 ^b	-720.487489	0.103291	0.095692
					0.102701 ^d	0.095049 ^d
Cys S ⁻ [A]	-720.493402	0.103063 ^b	0.095491 ^b	-720.493879	0.104676	0.097367
					0.104125 ^d	0.096747 ^d

^a Calculated at the G4 level.

^b Calculated at the G3 level.

^c Only single point MP2/aug-cc-pVTZ were performed.

^d Calculated at MP2/aug-cc-pVDZ level.

Table A2.4. H_{298} and G_{298} Values for all Neutral Amino Acids at the G3MP2 level (a.u.) which Differ from ref. 27.

Amino Acid	H_{298} (Ref. 27) a.u.	$\Delta H_{298} = H_{298}(\text{ref27})$ - $H_{298}(\text{current})$ in a.u.	ΔH_{298} kcal/mol	G_{298} (Ref. 27) a.u.	$\Delta G_{298} = G_{298}(\text{ref27})$ - $G_{298}(\text{current})$ in a.u.	ΔG_{298} kcal/mol
Cysteine neutral	-721.039540	0.001717	1.1	-721.082441	0.000609	0.4
Glutamine neutral	-531.012312	0.000817	0.5	-531.062414	0.000944	0.6
Isoleucine neutral	-440.985901	0.000649	0.4	-441.034957	0.000738	0.5
Phenylalanine neutral	-553.935181	0.000746	0.5	-553.985579	-0.000044	0.0
Tyrosine neutral	-629.083586	0.001135	0.7	-629.136871	0.000239	0.1

Table A2.5. Acidity Comparison for Acetic Acid and Phenol (kcal/mol).

Property	Acetic Acid ΔH_{298}	Acetic Acid ΔG_{298}	Phenol ΔH_{298}	Phenol ΔG_{298}
G3MP2	348.4	340.3	349.5	342.1
G3B3	348.2	340.1	349.5	342.0
G3	348.0	339.9	349.5	342.1
G4	347.8	339.8	349.3	341.7
B3LYP	348.6	340.5	349.7	342.2
CCSD(T)/CBS	348.3	341.5		
Expt(ICR) ^{a,b}	348.7 ± 2	341.8 ± 2	349.2 ± 2.5	342.4 ± 2
Expt(HPMS) ^{a,c}	348.7 ± 3	341.5 ± 2	350.4 ± 2.5	343.4 ± 2
Expt(CID) ^d	348.2 ± 1		348.0 ± 1	

^a Most values derived from experimental free energy from: ICR: Lias, S. G.; Bartmess, J. E.;

Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. *J. Phys. Chem. Ref. Data* **1988**, *17* (Suppl. 1).

^b Original ICR reference: Fujio, M.; McIver, R. T., Jr.; Taft, R. W. *J. Am. Chem. Soc.* **1981**, *103*, 4017-4029; Taft, R. W.; Topsom, R. D. *Prog. Phys. Org. Chem.* **1987**, *16*, 1-84.

^c Original HPMS reference: Cumming, J. B.; Kebarle, P. *Can. J. Chem.* **1978**, *56*, 1-9; Kebarle, P.; McMahon, T. B. *J. Am. Chem. Soc.* **1977**, *99*, 2222-2230.

^d CID reference: Angel, L. A.; Ervin, K. M. *J. Phys. Chem. A* **2006**, *110*, 10392-10403.

CHAPTER 3: AN EXPERIMENTAL AND COMPUTATIONAL INVESTIGATION INTO THE GAS-PHASE ACIDITIES OF TYROSINE AND PHENYLALANINE: THREE STRUCTURES FOR DEPROTONATED TYROSINE

3.1 Introduction The amino acids tyrosine and phenylalanine often have similar properties because both possess aromatic, hydrophobic side chains. Their structures (shown in Figure 1) are similar with tyrosine having a 4-hydroxybenzyl group at the side chain and phenylalanine having a benzyl side chain. Tyrosine possesses two acidic functionalities capable of deprotonation, the carboxylic acid group and the phenolic hydroxyl group. In the gas phase, most amino acids and peptides deprotonate at the C-terminal carboxylic acid group (if such a group is present). Exceptions include aspartic acid and glutamic acid, which have carboxylic acid groups located on the side chain and can deprotonate readily on either the C-terminus or the side chain.¹ Another exception is cysteine, which has been shown to deprotonate on the side chain,^{2,3,4} with theory³ and experiment⁴ indicating that a proton is shared between the deprotonated sulfur at the side chain and the deprotonated C-terminal carboxylate group. An infrared spectroscopy study suggested that the carboxylic acid group was deprotonated in cysteine, but this still can allow for the shared hydrogen between the two anionic sites.⁵ In the context of a peptide backbone, the carboxylic acid functionality becomes part of the amide linkage except at the C-terminus, thus removing a likely deprotonation site. However, deprotonation of peptides lacking highly acidic sites has been observed experimentally.^{6,7} Thus, alternative deprotonation sites are accessible using common mass spectrometry ionization techniques such as electrospray ionization (ESI).

Kass and coworkers have studied deprotonated tyrosine experimentally and theoretically, with a focus on determining if the deprotonated ion has a carboxylate or phenoxide structure.^{8,9} These researchers developed a useful ion/molecule reaction involving trimethylsilyl azide (TMSN₃) to distinguish between carboxylate and phenoxide ions.^{8,9} They employed this chemical probe to study the effects of the solvent system on deprotonated tyrosine structures produced by ESI and found that phenoxide is the favored when the solvent includes methanol, while carboxylate dominates when acetonitrile or acetonitrile/water solvents are used.⁸ Disparate results from photoelectron spectroscopy (PES) experiments indicate that the carboxylate ion is dominant.⁹ Kass and coworkers attributed this inconsistency between their TMSN₃ and PES experiments to different ESI source configurations. The PES experiments, which found carboxylate ions, used a home-built ESI source in which ions leaving the skimmer region pass through an ion guide for 100 ms. The ion/molecule reactions with TMSN₃, which found primarily phenoxide ions, employed a commercial ESI source where ions leaving the skimmer are accumulated in a hexapole trap for 1-5 seconds prior to their introduction into the mass analyzer. Using density functional theory (DFT) calculations (B3LYP/aug-cc-pVDZ), Kass and coworkers found a carboxylate structure to be energetically favored by 0.2 kcal/mol over a phenoxide structure.⁹ This very small difference would suggest the existence of both structures in the gas phase. These researchers interpreted their results as deprotonated tyrosine undergoing a structural change as the compound goes from a solution-phase neutral to a gas-phase anion.⁸

Oomens and coworkers performed infrared multiphoton dissociation (IRMPD) on several deprotonated amino acids, including tyrosine.⁵ The IRMPD spectra were compared to calculated spectra at the DFT B3LYP/6-31++G** level of theory. The amino acids were dissolved in 80:20 (v/v) ratio of methanol (CH₃OH) and water (H₂O), which is similar to the solvent conditions

where Tian and Kass⁸ observed increased phenoxide production. The ESI source used by Oomens and coworkers⁵ had a very similar commercial design to the source used by Kass and coworkers⁹ for the TMSN₃ reactions; both sources employed hexapole accumulations times of several seconds. However, the experimental IRMPD results show that deprotonated tyrosine is a carboxylate in the gas-phase, with no evidence of phenoxide. We previously predicted³ that tyrosine is deprotonated at the C-terminal carboxylate site using a variety of correlated molecular orbital (MO) theory methods up through CCSD(T)/aug-cc-pVTZ.^{10,11} Our higher level calculations show that the carboxylate ion is more stable than the phenoxide ion by ~2 kcal/mol, which is consistent with the IRMPD experimental results. Using lower level DFT, Li et al.¹² have found that the experimental IRMPD spectra of Oomens and coworkers⁵ are best matched to the calculated IR spectra for the carboxylate ion.

Gas-phase acidity (GA or ΔG_{acid}) values are an important tool in understanding structure, reactivity, and fragmentation behavior of compounds in mass spectrometry. The GA is the Gibbs free energy change (ΔG) for the reaction: $\text{AH} \rightarrow \text{A}^- + \text{H}^+$. GAs or deprotonation enthalpies (ΔH_{acid}) have been determined for the amino acids by O'Hair and coworkers¹³ and Poutsma and coworkers¹⁴ using the kinetic and extended kinetic methods, respectively, which involve collision-induced dissociation (CID) of a proton-bound dimer containing the analyte and a reference compound of known acidity. Poutsma and coworkers also calculated ΔH_{acid} for the common amino acids using a hybrid DFT approach with B3LYP functional combinations.¹⁴ We have determined the GAs of glutamic acid and aspartic acid with DFT and molecular orbital (MO) computational approaches, as well as experimentally using the thermokinetic method,¹⁵ which involves measurement of the rates of ion/molecule reactions between the deprotonated ion and reference compounds of known GA.¹ In addition, we predicted the GAs for the twenty

common and five rare amino acids using the G3(MP2) method and obtained results consistent with experiment and with other calculations.³ Kass and coworkers found the GA of tyrosine to be 332.5 ± 1.5 kcal/mol using equilibrium GA measurements in a dual cell Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR).⁹

In our recent study of the gas-phase and aqueous acidities of the amino acids, tyrosine was studied at eight computational levels involving both DFT and MO theory.³ In all cases, a gas-phase deprotonated carboxylate structure was predicted to be more stable on the free energy scale by 1.7 to 2.7 kcal/mol than a deprotonated phenoxide structure. However, as discussed above, gas-phase equilibrium proton transfer reactions and ion/molecule reactions involving TMSN₃ have found the phenoxide structure to be produced in the greatest abundance from protic solvents using ESI.^{8,9} This difference was the motivation for the current study. Our goal was to experimentally find the lowest energy, most acidic carboxylate form of gas-phase deprotonated tyrosine.

3.2 Experimental and Computational Methods *Mass Spectrometry* All experiments were performed with a Bruker Daltonics (Billerica, MA, USA) BioApex 7T FT-ICR mass spectrometer. The amino acids were in the L-stereoisomer. Samples were prepared at 60 μ M in solvents containing various ratios of CH₃OH, ultrapure H₂O, acetonitrile (CH₃CN), acetone, dimethyl sulfoxide (DMSO), tetrahydrofuran (THF), and N-dimethylformamide (DMF). All of the organic solvents were HPLC or LC-MS grade, except DMSO, which was ACS grade. For some solutions, 1% (by volume) of ammonium hydroxide (NH₄OH) was added to promote deprotonation. For experiments in which it was important to have a low water content, the solvents were also dried with 3A pore size molecular sieves.

Analyte solutions were introduced into an Apollo API source using a syringe pump set to deliver $\sim 90 \mu\text{L/hr}$. Electrospray ionization (ESI) employed dried air as a heated (225°C) counter and parallel current drying gas. (The air was dried with Labclear (Oakland, CA, USA) refillable gas filter employing RK-400 molecular sieves with a 13x pore size; the sieves also contained Drierite as a color changing indicator to denote the presence of water. The sieves were dried for at least 15 hours prior to experimental use in an oven at 200°C .) The ESI needle was grounded, while the capillary entrance and end plate were at a potential of 3.5-4.0 kV for negative ion mode analysis. Unless otherwise noted, ions produced by ESI were accumulated in a hexapole for 600-700 ms before being transported to the reaction cell by electrostatic focusing. The time used to transport ions into the cell was 10 ms.

Deprotonated ions, $[\text{M} - \text{H}]^-$, were isolated using correlated frequency ion ejection techniques.¹⁶ Because the ions are exposed to constant pressures of neutral reactants during this isolation period, the time involved in ion isolation was kept to a minimum and was never more than 50 ms. The isolated precursor ions were then allowed to react with a reference compound introduced to the ICR cell through a leak valve at a constant pressure. Each of the ions selected for study were reacted with a series of reference compounds of known GA.¹⁷ The reference compound pressures were measured using an ion gauge that was calibrated with the proton transfer reaction between protonated glycine and N,N-dimethylformamide, which has an experimental rate constant of $8.19 (\pm 1.07) \times 10^{-10} \text{ cm}^3/\text{molecules}\cdot\text{s}$.¹⁸ Pressures were corrected for reactant gas ionization efficiency,¹⁹ which involved polarizabilities calculated by atomic hybrid parameter procedures.²⁰ Reference compound pressures were in the range of $(1-10) \times 10^{-8}$ mbar. For the reactions of each precursor ion with each reference compound, at least one

pressure used was from the upper half of this range and at least one pressure was from the lower half of this range; the higher and lower pressures differed by at least a factor of 4 ($\times 4$).

Reaction rate constants, k_{exp} , were obtained by observing the pseudo-first-order decay in reactant ion intensity as a function of trapping time. In cases where deprotonation was in competition with proton-bound dimer formation, k_{exp} was obtained by fitting the experimental ion intensity data as discussed previously.²¹ For experiments in which non-linear pseudo-first-order kinetics plots (bimodal plots) indicated the presence of two ion structures reacting at two different rates, the data was fit to the sum of two exponential decays using the program Sigma Plot by Systat Software Inc. (San Jose, CA, USA). The fraction that each exponential contributes to the fit directly relates to the relative abundance of that ion structure, while the slope of each exponential decay is used to calculate the reaction rate constant (in the same manner that this information is used to calculate rate constants from a unimodal kinetics fit). This procedure has been used in the past to obtain rate constants and gas-phase basicities for systems containing two or three ion structures reacting at different rates. This procedure has been used in the past to obtain rate constants^{22,23,24,25,26} and gas-phase basicities^{27,28,29} for systems containing two or three ion structures reacting at different rates.

The ratio of the experimental rate constant to the thermal capture rate constant^{30,31} yields a reaction efficiency (RE). A RE of 0.269 was used as a break point, where a reaction is considered to become exoergic and a GA value is assigned. This selection of 0.269 comes from the work of Bouchoux and coworkers,^{15,32,33,34} which has been termed the “thermokinetic method.” We have found that this method provides excellent agreement between experimental and theoretical GAs for several amino acids and small peptides.^{1,6}

Computational Methods The calculations were performed at the DFT and correlated MO theory levels with the program Gaussian-09.³⁵ The geometries were initially optimized at the DFT level with the B3LYP exchange-correlation functional^{36,37} and the DZVP2 basis set.³⁸ Vibrational frequencies were calculated to show that the structures were minima. A range of structures were optimized to determine the most stable conformers chosen by sampling many conformations with and without hydrogen bonds. A substantial number of low energy conformers were found for the neutrals as discussed below. In our previous work on predicting the GAs of amino acids^{1,3,39} and organic acids,³⁹ we showed that the high level G3(MP2) correlated MO method⁴⁰ gave agreement for the acidities with the experimental values to within about ± 1 kcal/mol and also agreement with higher level CCSD(T) calculations extrapolated to the complete basis set limit with additional corrections.^{41,42,43,44,45} G3(MP2) has an additional advantage over DFT methods in terms of reliable predictions for these types of compounds because the correlated MO methods in G3(MP2) perform better in the prediction of hydrogen bond energies as well as steric non-bonded interactions than do most widely-used DFT exchange-correlation functionals. Thermal corrections to the enthalpies and the free energies were calculated in the scaled harmonic oscillator, rigid approximation⁴⁶ using the geometries and frequencies obtained in the G3(MP2) calculations (Hartree-Fock/6-31G* level).

Organic Acid and Alcohol Benchmarks As a further benchmark of the computational methods in use, the GAs have been predicted for 11 carboxylic acids (formic, acetic, propanoic, isobutyric, trimethylacetic, butanoic, isovaleric, tert-butylacetic, pentanoic, isohexanoic, and tert-butylacetic) and 4 phenols (phenol, and *ortho*-, *meta*-, and *para*-cresol). The optimized structures of the neutrals and resulting anions of the 11 carboxylic acids are shown in the Figure 2. Conformational searches were done and the lowest energy conformer of the neutral and anion

was used for the acidity calculations. In all cases, except for isobutyric acid, the orientation of the methyl groups in the neutrals and their corresponding anions were identical. In isobutyric acid, a rotation of the two methyl groups occurred to produce the most stable anion.

Excellent agreement is found with the available experimental ΔH_{acid} and GA values of the organic acids except for the results from Muftakhov et al.⁴⁷ (Table 1). Muftakhov et al. underestimate the deprotonation enthalpies of formic and acetic acid by ~5 kcal/mol in both cases. For formic acid, the reaction enthalpy was also calculated at a composite CCSD(T)/CBS level^{41,42,43,44,45} to further benchmark the ability of G3(MP2) to calculate the acidities for these molecules. The components for the CCSD(T) atomization energies are given in the Appendix. The CCSD(T) results for formic and acetic acid³⁹ (Table 1) are in excellent agreement with the GAs calculated at the G3(MP2) level and the experimental values. The CCSD(T) calculations¹⁰ were done with the augmented correlation consistent basis sets¹¹ and the MOLPRO⁴⁸ and NWChem⁴⁹ program systems.

The results show that the carbon chain length has only a very small effect on the acidity. Comparison of the acidities of butanoic and isobutyric acids, which both have 4 carbons in the chain, the additional methyl group in isobutyric acid resulted in an increased acidity of 0.5 kcal/mol. These trends follow for the other carboxylic acids as well. Carboxylic acids with two methyl groups are slightly more acidic and range from 338.5 to 338.9 kcal/mol, and those with three methyl groups are even more acidic, ranging from 337.4 to 337.8 kcal/mol.

The optimized structures for the neutrals and anions of phenol and the three cresols are shown in Figure 3. The calculated values of the phenol and cresol acidities are all within the experimental error bars as shown in Table 1. The experimental results show that the acidity of the cresols decreases from *ortho* to *meta* to *para* with the acidity of phenol falling between *o*-

cresol and *m*-cresol giving the order *o*-cresol > phenol > *m*-cresol > *p*-cresol. The difference in acidities is only 2 kcal/mol from *o*-cresol to *p*-cresol. As the acidity difference is small, further calculations were performed to demonstrate that *p*-cresol was not lower in energy at a different computational level. The energy differences between the *ortho*, *meta*, and *para* anions were calculated at different Gx computational levels,⁵⁰⁻⁵³ as shown in Table 2. *o*-Cresol was predicted to be ~1 kcal/mol more acidic than *m*-cresol with all methods except for G3B3 and G3MP2B3 where *o*-cresol was predicted to be more acidic by ~2 kcal/mol. *o*-Cresol was predicted to be more acidic than *p*-cresol by 1.5-2 kcal/mol.

3.3 Results and Discussion *Calculated GAs and Structures for Tyrosine, Phenylalanine and 4-(4-Hydroxyphenyl)-2-butanone (HPB)* For deprotonated tyrosine, we have previously predicted at the G3(MP2) level that the carboxylate anion is more stable than the phenoxide by 2.5 kcal/mol.³ The structure of the carboxylate is folded to maximize hydrogen bonding. The question exists as to whether there are additional higher energy carboxylate conformers with a similar elongated structure to that of the phenoxide. Figure 4 shows the optimized structures of the lowest energy conformer of the two amino acids and HPB, and low energy conformers for the deprotonated species. Unless noted, all energies refer to free energies at 298 K. The multiple low energy conformers of the neutral amino acids are shown in Figures 5 and 6 for phenylalanine and tyrosine respectively. In all cases, the lowest energy folded and unfolded structures differ by less than 0.5 kcal/mol. Table 3 shows the experimental and G3(MP2) calculated GAs for these molecules.

There are 22 conformers for tyrosine with energies within 2 kcal/mol of the lowest energy structure. Weak hydrogen bonding between the amine and two carbons on the phenolic ring plays an important role in the neutral, leading to a folded structure. For the tyrosine anion,

we studied 10 different structures. Three low energy structures (Figure 4a) were found that differ by only 2.5 kcal/mol, two carboxylate conformers and one phenoxide isomer, in contrast to the large number of low-lying structures in the neutral. A folded structure is predicted to be the lowest energy structure for the carboxylate anion with ring C-H's interacting with the CO₂⁻ group. The carboxylate anion has a higher energy unfolded conformer 1.0 kcal/mol less stable than the folded carboxylate conformer. The lowest energy conformer for the phenoxide anion is unfolded, is 2.5 kcal/mol less stable than the lowest energy folded carboxylate conformer, and is 1.5 kcal/mol less stable than the unfolded carboxylate conformer on the free energy scale at 298 K. In terms of $\Delta H_{298}(\text{gas})$, the energy difference between the unfolded anions is smaller, only 0.3 kcal/mol.

As shown in Table 3, we have previously used a variety of Gx methods^{51,52} and CCSD(T) theory to predict the energy difference between the folded carboxylate and phenoxide sites of tyrosine. We used these same methods to calculate the relative energy values for the unfolded carboxylate anion. In all cases, the folded carboxylate anion is predicted to be 1.7 to 2.7 kcal/mol more stable than the phenoxide and the unfolded carboxylate is between the folded carboxylate and the unfolded phenoxide. Higher energy deprotonation sites include the aromatic ring, the α carbon, and the amine group and these are shown in Figure 7 for the amino acids.

Phenylalanine has 13 low energy conformers within 2 kcal/mol of the lowest energy folded structure. Deprotonation of phenylalanine (Figure 4b) forms two carboxylate structures that have G3(MP2) acidities differing by 0.3 kcal/mol with the lowest energy conformer being folded.

Neutral HPB has two low energy conformers and the lowest energy structure is unfolded. The lowest energy phenoxide anion is folded and 2.3 kcal/mol more stable than the unfolded ion.

Experimental GAs and Structures Produced from a Protic Solvent System Our experimental studies began with proton transfer ion/molecule reactions to bracket the GAs of tyrosine, phenylalanine, and HPB. This work was performed using a protic solvent mixture of 49.5:49.5:1 (v/v/v) CH₃OH:H₂O:NH₄OH, which is a routine ESI solvent system that often generates intense negative ions. Table 4 shows the reference compounds used in this study, their GAs, and the measured reaction efficiencies. From these results, experimental GAs were assigned. Our experimental and computational GA values are summarized in Table 3, along with literature values from other studies on these compounds.

The least acidic compound studied was HPB, which has an experimental GA of 339.6 ± 3.0 kcal/mol. (A higher GA value indicates a less acidic compound.) This agrees with our G3(MP2) calculated value of 337.8 kcal/mol and the computational result is well within the experimental error bars. HPB was chosen for comparison to tyrosine and phenylalanine because this compound has no carboxylic acid group and, therefore, must deprotonate at its phenolic hydroxyl group.

Phenylalanine, which contains no phenolic side chain but has a carboxylic acid group at the C-terminus, is more acidic than HPB and has an experimental GA of 332.5 ± 2.2 kcal/mol, which agrees within experimental error with our G3(MP2) value of 330.7 kcal/mol. For HPB and phenylalanine, all experimental kinetics plots were linear, suggesting the presence of only one major deprotonated ion structure. If multiple structures exist, the experiments indicate that they are very similar in energy; for example, the 0.3 kcal/mol energy difference found by G3(MP2) calculations between the folded and unfolded conformers of deprotonated phenylalanine would be difficult to distinguish experimentally. There is reasonable agreement of both the calculated and experimental GAs with the experimental GA for phenylalanine of 329.6

± 3.0 kcal/mol previously reported by O'Hair et al.¹³ The fact that HPB is less acidic than phenylalanine is consistent with phenols generally being less acidic than carboxylic acids. For example, the GA of phenol is 343.4 ± 2.0 kcal/mol, while the GA of acetic acid is 341.5 ± 2.0 .⁵⁴ Recent calculated GAs by Angel and Ervin⁵⁵ based on experimental enthalpies for loss of a proton are 341.5 ± 1.0 and 339.9 ± 1.7 for phenol and acetic acid, respectively, are consistent with the above values. In both cases the error bars overlap but the experimental results and the theoretical results in Table 1 are all consistent in the ordering between acetic acid and phenol.

Tyrosine is an interesting case because there are two potential sites of deprotonation. Using the $\text{CH}_3\text{OH}:\text{H}_2\text{O}:\text{NH}_4\text{OH}$ solvent system, two ion populations were observed in the reactions with ethyl cyanoacetate (GA = 333.6 ± 2.0 kcal/mol¹⁷) and 4-amino-2,3,5,6-tetrafluoropyridine (ATFP, GA = 332.8 ± 2.0 kcal/mol¹⁷). To illustrate this bimodal reactivity, Figure 8 gives a semi-logarithmic plot of the intensity of deprotonated tyrosine, $[\text{Tyr} - \text{H}]^-$, as a function of reaction time with ethyl cyanoacetate. The experimental data (black circles) are an excellent fit to an equation involving the sum of two exponentials (black line). From the reaction efficiency data of Table 4, the less acidic deprotonated tyrosine structure accounts for ~27% of the ions and has a GA of 333.5 ± 2.4 kcal/mol, while the more acidic structure accounts for ~73% of the ions and has a GA of 332.4 ± 2.2 kcal/mol. This is the first time that experimental GAs have been separately obtained for the different deprotonated tyrosine structures. Using an equilibrium method, Kass and coworkers reported a single GA value of tyrosine of 332.5 ± 1.5 kcal/mol,⁹ which is in excellent agreement with our GA value relating to the more acidic structure. Using the kinetic method of CID on a proton bound dimer, O'Hair et al. obtained a value of 329.5 ± 3.0 kcal/mol,¹³ which agrees to within experimental error. O'Hair's value is in excellent agreement with our calculated G3(MP2) value for the folded carboxylate.

The proton transfer reaction data clearly shows two structures for deprotonated tyrosine, but does not distinguish between carboxylate and phenoxide ions. To identify these structures, we employed the ion/molecule reaction with TMSN_3 developed by Kass and coworkers.^{8,9} Using a series of model compounds to characterize reactivity, they found that deprotonated tyrosine with the carboxylate structure reacts with TMSN_3 to form the azide ion, N_3^- (m/z 42), and a neutral tyrosine trimethylsilyl ester (TyrOTMS); in contrast, the phenoxide structure of deprotonated tyrosine silylates at the phenol group, producing TMSOTyrCO_2^- (m/z 252) and the neutral HN_3 . In our work, for ion generation from a protic $\text{CH}_3\text{OH}:\text{H}_2\text{O}:\text{NH}_4\text{OH}$ solvent system, the TMSN_3 reaction showed that the phenoxide structure accounts for $67(\pm 7)\%$ of the ions, with the remaining $33(\pm 7)\%$ of the ions being carboxylate.

Taken together, the proton transfer and the TMSN_3 reactivity data indicate that the more abundant and more acidic deprotonated tyrosine species ($\sim 70\%$ of the ions) has a phenoxide structure and yields an experimental GA of 332.4 ± 2.2 kcal/mol. This corresponds to structure C of Figure 4a, unfolded phenoxide, which involves a calculated GA of 332.9 kcal/mol. The less abundant and less acidic deprotonated tyrosine species ($\sim 30\%$) has a carboxylate structure and an experimental of GA of 333.5 ± 2.4 kcal/mol. This corresponds to structure B of Figure 4a, unfolded carboxylate, which involves a calculated GA of 331.4 kcal/mol. These structures were generated from a polar solvent system of 49.5:49.5:1 (v/v/v) $\text{CH}_3\text{OH}:\text{H}_2\text{O}:\text{NH}_4\text{OH}$. Kass and coworkers found this same phenoxide/carboxylate ratio of 70:30 when tyrosine ions were produced from a solution of 75:25 (v/v) $\text{CH}_3\text{OH}:\text{H}_2\text{O}$.^{8,9} Thus, our results and those of Kass both show that the deprotonated phenoxide ion is the dominant gas-phase species when tyrosine is electrosprayed from a protic solvent system.

Our experiments found the GA involving the phenoxide structure to be 0.5 kcal/mol lower than the GA with the carboxylate structure. This differs from the G3(MP2) ordering, which predicts the unfolded carboxylate structure to yield a lower GA by 1.5 kcal/mol. There are several possible experimental reasons for this deviation. First, the proton transfer reactions involving the two deprotonated tyrosine reactant ion structures may have different activation barrier heights, which could lead to a reaction being under kinetic control as opposed to thermodynamic control. This issue is known to affect proton transfer studies of several organic molecules.^{56,57,58} In addition, the reaction intermediate is a proton bound dimer involving deprotonated tyrosine and an acidic reference compound, $[(\text{Tyr} - \text{H})^- \cdot \text{H}^+ \cdot (\text{A} - \text{H})^-]$. This is essentially the same type of complex that is used in the kinetic method of gas-phase basicity (GB) and GA determinations, which involves collision-induced dissociation of the dimer.^{59,60} For several small organic acids, deviations from expected GAs or GBs using the kinetic method have been attributed to the dimer adapting a slightly higher energy structure in situations where double hydrogen bonding exists within the dimer or where the analyte conformation preferred in the dimer differs from that of its monomer form.^{61,62,63} For example, Fournier et al.⁶² found that the amino acid glutamic acid has bimodal dissociation plots during a GA study using the kinetic method. Using calculations at the G3(MP2) and OLYP/aug-cc-pVTZ levels, they concluded that glutamic acid could exist within the dimer in both zwitterionic and non-zwitterionic forms. Finally, steric hindrance of the reactive site, either from bulky substituents or folded conformations with hydrogen bonding, can lessen the ability of a neutral to access an ion's reactive site, which lowers the reaction efficiency and results in a lower GA assignment.^{21,64,65}

As a check of the computational method, other approaches were used to predict the energy difference between the two carboxylate conformers and phenoxide isomer as shown in

Table 3. The predicted energy differences are essentially independent of the computational method. Geometries were optimized at the MP2/aug-cc-pVTZ level with diffuse functions and higher order polarization, in contrast to the G3 geometries. Relative energies were calculated at the CCSD(T)/aug-cc-pVTZ level using optimized MP2/aug-cc-pVTZ geometries with thermal corrections, entropies and zero point energies calculated at the MP2/aug-cc-pVDZ level. The CCSD(T) results are good to at least ± 1 kcal/mol for the relative energies.

Effects of the Solvent on Deprotonated Tyrosine: A New Carboxylate Structure Because the experiments with the CH₃OH:H₂O:NH₄OH solvent system did not result in our finding the low energy carboxylate structure predicted by theory, additional experiments were performed. Kass and coworkers had reported that the identity of the solvent used in the ESI experiment has a dramatic effect on the phenoxide/carboxylate ratio for deprotonated tyrosine.^{8,9} They found that protic solvent systems containing CH₃OH favored phenoxide formation, while aprotic solvent systems containing CH₃CN favored the carboxylate structure. In addition, Schröder et al.⁶⁶ found that the phenoxide/carboxylate ratio for deprotonated *p*-hydroxybenzoic acid depends on the solvent, pH, and concentration of the solution being electrosprayed.

Using the TMSN₃ ion/molecule reaction to distinguish between phenoxide and carboxylate ions, we electrosprayed tyrosine from a variety of solvents. These included seven solvent systems used by Tian and Kass,⁸ plus 13 additional solvent systems. The results of these experiments are shown in Table 5. Exclusively (100 %) carboxylate ions were found for the entirely aprotic solvents of pure CH₃CN, pure acetone, and a mixture of DMF, THF, and DMSO. An addition of 1% CH₃OH or H₂O to aprotic CH₃CN has almost no effect, yielding 99-100% carboxylate ions. However, 1% addition of the base NH₄OH results in 35% of the deprotonated tyrosine ions having a phenoxide structure. Addition of 2% CH₃OH or H₂O also has a

noteworthy effect, causing about 50-70% of the ions to be phenoxide. In general, phenoxide ions appear in abundance when NH_4OH or the protic solvents CH_3OH or H_2O are used. Our results generally agree with those of Tian and Kass⁸ although they concluded that CH_3OH greatly affects the phenoxide/carboxylate ion ratio but H_2O has no effect, while we found that H_2O and CH_3OH have a similar large effect on the ratio. For example, as our data in Table 5 indicate, ESI from a solution of 50% CH_3CN and 50% of either CH_3OH or H_2O both result in 63% phenoxide ions and 37% carboxylate ions.

Since aprotic solvents yield almost exclusively carboxylate ions, we performed proton transfer reactions on deprotonated tyrosine produced by ESI from a solvent of 99:1 $\text{CH}_3\text{CN}:\text{H}_2\text{O}$. The mixture included 1% H_2O solution because the presence of a trace of water greatly increased the solubility of tyrosine; however, the water content was tightly controlled and the LC-MS grade CH_3CN was dried with molecular sieves to insure that only the 1% water that we added was present. Reactions with TMSN_3 were performed multiple times (including in close time proximity to the proton transfer reactions) and always showed that this solvent system forms 99-100% carboxylate ions and 1 to 0% phenoxide ions.

The reaction efficiencies for proton transfer reactions of tyrosine anions produced from this aprotic solvent are given in Table 4. For all six reference compounds, two deprotonated tyrosine ion populations reacting at different rates were observed. To illustrate this, Figure 9 shows the semi-logarithmic plot of the intensity of $[\text{Tyr} - \text{H}]^-$ as a function of reaction time with the reference compound was ATPF. The plot indicates that ~71% of the ions are reacting with ATPF (albeit slowly), while the remaining ions are almost non-reactive. Using the data of Table 4, the faster reacting (less acidic) species was found to account for $61(\pm 11)\%$ of the ions and to involve a GA of 333.5 ± 2.4 kcal/mol. The slower reacting (more acidic) species includes

39(\pm 11)% of the ions and yields a GA of 324.7 ± 3.6 kcal/mol. Because there is \sim 9 kcal/mol difference in acidities, these two species are readily distinguished by ion/molecule reactions.

This most acidic structure for deprotonated tyrosine, yielding an experimental GA of 324.7 ± 3.6 kcal/mol, was not produced from the protic solvent system. This ion only forms in aprotic solvent and the TMSN₃ reactions indicate that it has a carboxylate structure. Therefore, this is the lowest energy deprotonated tyrosine with a carboxylate structure that our calculations had previously predicted. The G3(MP2) calculated GA involving the folded carboxylate, structure A of Figure 4a, is 330.4 kcal/mol. This value is 5.7 kcal/mol higher than the experimental GA, which is a greater deviation than we typically find between experiment and theory. However, the calculations indicate that this structure is highly folded with the deprotonated carboxylate group participating in two hydrogen bonds. Due to its highly folded structure, the experimental GA value relating to this structure is likely to have more error than the other experimental GA values obtained this study because, as noted above, steric hindrance of the reactive site can yield erroneously low reaction efficiencies.^{21,64,65} Still, reactions with TMSN₃ show that carboxylate ions are produced almost exclusively from aprotic solvents, while proton transfer reactions find a carboxylate structure to be the lowest energy gas-phase conformer or isomer of deprotonated tyrosine

The different deprotonated tyrosine ion structures generated from protic versus aprotic solvents may relate to a different tyrosine neutral structure in the two types of solvents. In aqueous solution, amino acids exist in their zwitterionic form in \sim 10,000 times greater abundance than their non-zwitterion form. In general, the zwitterionic forms of amino acids dominate in protic solvents, where they are stabilized by strong hydrogen bonding to the solvent.⁶⁷ Therefore, in our protic solvent system of CH₃OH:H₂O:NH₄OH neutral tyrosine is a

zwitterion. Consequently, during the ESI process, the site that loses a proton to form deprotonated tyrosine is the ammonium group of the zwitterion. In contrast, for aprotic solvents there is no consistency with regard to whether a neutral amino acid exists in zwitterionic or non-zwitterionic form. The form relates to both the solvent and the amino acid, and is difficult to predict. In general, the larger the side chain of an amino acid, the less likely it is to exist as a zwitterion in aprotic solvents because steric hindrance will limit solvation.^{67,68} Due to the low solubility of tyrosine in most solvents, almost nothing is known about the form of its neutral in non-aqueous solvents. However, tyrosine has one of the largest amino acid side chains, which suggests that its neutral exists as a non-zwitterion in aprotic solvents. Therefore, it is likely that in an aprotic solvent, the ESI process is deprotonating the most acidic site of the non-zwitterion of tyrosine, which is the carboxylic acid group ($pK_a = 6.15$ for carboxylate deprotonation and $pK_a = 10.37$ for phenolate deprotonation³).

To test the premise that the neutral form of tyrosine differs in protic and aprotic solvents (and that this might account for different gas-phase ion structures), attempts were made to study the solution-phase structure(s) of tyrosine using FT-IR, FT-Raman, and NMR. Solutions saturated with tyrosine were analyzed in solvent systems ranging from 100% aprotic to 100% protic with 0-5% acid or base additive. Unfortunately, we were not able to obtain any usable data with these spectroscopic techniques because of the low solubility of tyrosine in all solvents tested (including water⁶⁹). The fact that tyrosine can be studied by mass spectrometry in these solvents is testimony to the fact that mass spectrometry can analyze very dilute solutions.

The second, faster reacting tyrosine anion population generated from aprotic solvent is also interesting. Reactions with $TMSN_3$ indicate that the aprotic solvent produces 99% carboxylate ions, but the experimental GA from this second ion population exactly matches the

experimental GA relating to the phenoxide ion found in protic solvent. This could mean that the less acidic (more elongated) carboxylate structure is converting to the phenoxide due to interactions with the references compounds. In particular, AFTP is only ~1 kcal/mol more acidic than deprotonation of tyrosine to generate the unfolded carboxylate. During collisions of AFTP neutrals with tyrosine anions a proton bound dimer intermediate, $[(\text{Tyr} - \text{H})^- \cdot \text{H}^+ \cdot (\text{AFTP} - \text{H})^-]$, could form. This dimer might dissociate with conversion of a tyrosine ion from carboxylate to phenoxide. Such a process would not yield a change in mass and, therefore, would not have been apparent in the mass spectra.

Additional support for a conversion between structures is provided by the AFTP reaction data for tyrosine ions produced from protic solvent. As seen in Table 4, for tyrosine anions produced from protic solvent reacting with AFTP the measured reaction efficiency is 0.53 ± 0.25 (mean \pm standard deviation). The high standard deviation illustrates that this data is much less reproducible than all of the other reaction efficiencies that were measured. For this reaction only, the efficiency varied with pressure and was a relatively poor fit to the sum of two exponentials. This data suggested that the ion population was changing in composition as we were studying it. Thus, data for the reactions of AFTP with deprotonated tyrosine ions produced from both protic and aprotic solvents suggest that a gas-phase conversion between carboxylate and phenoxide ion structures can occur.

Deprotonated Tyrosine Structure as a Function of Hexapole Accumulation Time The ratio of carboxylate to phenoxide ions was studied as a function of accumulation time for ions in the hexapole that immediately follows our ESI source. Previously, Kass and coworkers⁹ had attributed the inconsistency between their TMSN₃ and PES experiments to different ESI source configurations, with shorter accumulation times (100 ms) favoring a carboxylate structure and

longer times (1-5 s) favoring a phenoxide structure. Our mass spectrometer has a hexapole for ion accumulation and its timescale can be easily manipulated, allowing us to study accumulation times similar to those for the two instruments used by Kass in the PES and TMSN₃ experiments.

Our proton transfer reactions that bracketed GAs in both protic and aprotic solvent systems used a hexapole accumulation time of 0.6-0.7 s. This time was selected because it generally provided the optimal signal intensity for deprotonated tyrosine. In ESI, ions are produced continually but the FT-ICR mass analyzer requires a pulsed packet of ions. To convert a continuous stream of ions into a pulse, ions leaving the ESI source are trapped in the hexapole where they remain until moved into the FT-ICR cell. During the entire accumulation time, the hexapole can accept new ions. Thus, some ions are trapped in the hexapole for almost the entire accumulation time while other more recently formed ions only spend a few milliseconds in the hexapole.

Because proton transfer reactions with ethyl cyanoacetate can readily distinguish the carboxylate and phenoxide tyrosine structures generated from protic solvent systems, reactions with this reference compound were used to monitor the ion populations as a function of hexapole accumulation time. With an extremely short accumulation time of 1 ms, the faster reacting species (i.e., carboxylate) nearly disappears, resulting in 7% carboxylate ions and 93% phenoxide ions. With an accumulation time of 0.7 s (from the data of Table 1), there is 24% carboxylate and 76% phenoxide. For a very long accumulation time of 5 s, the ratios of the two populations remained at ~24:76.

The effects of accumulation time for ions produced from various solvents were also studied using TMSN₃ reactions to measure the carboxylate and phenoxide abundances. Data obtained for three solvent systems is shown in Figure 10. Figure 10a involves a protic solvent

system of 74.5:24.5:1 (v/v/v) CH₃OH:H₂O:NH₄OH. At very short accumulation times the ion population is 55% phenoxide and 45% carboxylate, but this stabilizes at ~70% phenoxide and ~30% carboxylate after ~0.7 s. Similar effects were observed with other protic solvent systems. Figure 10c shows data obtained with an aprotic solvent of 99:1 (v/v) CH₃OH:H₂O. At short accumulation times, the ions are almost exclusively carboxylate; at a long time of 5 s, ~3% of the ions are phenoxide. Figure 10b, involving 98:2 (v/v) CH₃OH:H₂O, is especially interesting because within less than 1 second of accumulation time the ion ratio goes from ~55% carboxylate/ 45% phenoxide to the reverse. This experiment was repeated several times over a period of months and the result was always the same.

The general trend is that as the ions spend more time in the hexapole, the relative abundance of phenoxide ions increases and carboxylate ions decreases. This is especially pronounced when the ESI solvent system contains at least 2% of a protic solvent such as H₂O or CH₃OH. A possibility is a preferential retention of phenoxide ions or a preferential loss of carboxylate ions, although it is difficult to envision a reason for either of these events to occur. Another possibility is that a conversion between structures is occurring. Presumably, this would be the less stable carboxylate ion (experimental GA of 333.5 kcal/mol) converting to the slightly more stable phenoxide ion (experimental GA of 332.4 kcal/mol), just as these ions appear to convert during the reactions with the reference compound ATRP.

Hexapole ion guides are generally known for collisional cooling of ions but a slight heating of accumulated ions can occur due to a radial stratification effect. This effect is most pronounced for ions of high charge and high mass-to-charge ratio.^{70,71,72} Deprotonated tyrosine ions are singly charged and of relatively low mass, making addition of energy during the

hexapole trapping process less likely. However, very little energy would be required to convert between the deprotonated tyrosine isomers and conformers.

If a conversion is occurring, it may be facilitated by solvent interactions with ions in the hexapole. The pressure in the hexapole is in the 10^{-3} mbar range, with the primary component being dried air (the ESI drying and nebulizer gas) along with an unknown amount of solvent.⁷³ The fact that the change in carboxylate/phenoxide ratio is most pronounced in the presence of H₂O supports the involvement of protic solvent (which can hydrogen bond to tyrosine) in an isomerization process. Tian and Kass⁸ proposed that there was conversion between the two isomeric structures in the presence of CH₃OH. They attributed this effect to a gas-phase relay mechanism^{74,75} in which CH₃OH simultaneously coordinates to the deprotonated carboxylic acid and neutral phenol sites, protonating the carboxylate while deprotonating the phenol.

We studied structures for the starting point of the proposed relay mechanism at the DFT level (B3LYP/ aug-cc-pVDZ) for the folded and unfolded carboxylate and phenoxide structures with the addition of two, three, or six waters. Relative energies were calculated for ΔG_{gas} and ΔG_{aq} and are given in Table 6. The optimized structures of the solvated tyrosine anions and anions with two, three, or six explicit waters are shown in Figures 11 and 12. 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 obtained from self-consistent reaction field calculations.⁷⁶ The solvation energy was calculated as the sum of the electrostatic energies (polarized solute - solvent) and the non-electrostatic energies using the COSMO parameterization⁷⁷ at the B3LYP/aug-cc-pVDZ level using the gas-phase geometries obtained at this level. Solvents used include H₂O, CH₃OH and CH₃CN with respective dielectric constants of 78.39, 32.63, and 36.64.⁷⁸ In the gas phase, the folded carboxylate and unfolded phenoxide

anions are ~1 kcal/mol higher in energy than the unfolded carboxylate structure. Without the addition of any explicit waters of solvation, the phenoxide anions are still higher in energy than the carboxylate anions independent of the solvent used. When using a two- or three-water relay mechanism, the folded carboxylate anion is more stable in the gas phase whereas the unfolded carboxylate is more stable in aqueous solution. In both cases, the phenoxide anions are higher in energy than the carboxylate anions by ~4 to 9 kcal/mol. Extra water molecules were added to better solvate the O⁻ and COO⁻ sites on the anions. The addition of six waters resulted in the unfolded and folded carboxylate structures having approximately the same energy in the gas phase whereas in aqueous solution the unfolded carboxylate was still more stable. The unfolded phenoxide structure is more stable with the additional waters and has approximately the same relative energy as the folded carboxylate anion, ~4 kcal/mol. The stability in solution of the unfolded structure is likely due to an effective larger dipole moment than for the folded structure. These results show that there is additional chemistry occurring, leading to the experimental observations.

3.4 Conclusions Gas-phase deprotonation of the amino acid tyrosine produces three structures: a species deprotonated at the phenolic side chain and two conformers deprotonated at the C-terminal carboxylic acid group. To the best of our knowledge, this is the first time that three distinct structures have been observed when removing a proton from the same small organic precursor molecule by mass spectrometry. The lowest energy, most stable structure is deprotonated at the carboxylic acid group and G3(MP2) calculations indicate that the structure is highly folded with extensive hydrogen bonding. This structure was only observed experimentally during ESI from aprotic solvents, where tyrosine may exist in a non-zwitterionic form. The second lowest energy structure found experimentally, which is several kcal/mol higher in energy,

is deprotonated at the phenol group and is present when the solvent system contains at least 2% of a protic compound. In protic solvent, neutral tyrosine exists as a zwitterion. The third structure is a carboxylate ion, which G3(MP2) calculations indicate to be unfolded. This carboxylate ion is only ~1.5 kcal/mol higher in energy than the phenolate ion and the two structures appear to be converting during ion/molecule reactions with some acidic reference compounds and also during ion accumulation in a hexapole. G3(MP2) calculations found that the three structures differ in energy by ~2.5 kcal/mol, yet all three are readily distinguished experimentally by proton transfer ion/molecule reactions. Experimental conditions, such as solvent and time that ions are accumulated in a hexapole, greatly affect the abundance ratios of the deprotonated tyrosine structures. In contrast, the aromatic amino acid phenylalanine, which has a benzyl side chain, was found experimentally to have one structure when electrosprayed from a protic solvent. This is consistent with the results of G3(MP2) calculations, which indicate that a folded carboxylate structure is only 0.3 kcal/mol more stable than an unfolded carboxylate structure.

Acknowledgement This research 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, and CHE-1308348. Samantha S. Bokatzian thanks the University of Alabama's National Alumni Association for a License Tag Fellowship. Michele L. Stover and Chelsea E. Plummer thank the U.S. Department of Education for a Graduate Assistance in Areas of National Need fellowship (DoEd-GAANN grant P200A100190). Michele L. Stover also thanks the Howard Hughes Medical Institute (HHMI) for a teaching fellowship through the Precollege and Undergraduate Science Education Program. David A. Dixon thanks the Robert Ramsay Fund from the University of Alabama for

partial support. The assistance of Dr. Ken Belmore and Dr. Russell Timkovich with the NMR experiments is gratefully acknowledged. Some of the CCSD(T) calculations were performed at the Molecular Sciences Computing Facility in the W. R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by DOE's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory, operated for the DOE by Battelle.

Table 3.1. G3(MP2) Gas-Phase Acidities of the Carboxylic Acids and Phenols in kcal/mol.

Compound	Acid R group	ΔH_{298} calc	ΔH_{298} expt ^a	ΔG_{298} calc	ΔG_{298} expt ^a
Formic acid	H	344.7 344.5 ^b	345.2 ± 2.9 ⁵⁴	337.3 337.0 ^b	338.2 ± 2.0 ⁵⁴
			345.4 ± 2.2 ⁷⁹		338.4 ± 2.0 ⁷⁹
			345.3 ± 2.2 ⁸⁰		338.3 ± 2.0 ⁸⁰
			346.2 ± 1.2 ⁸¹		339.2 ± 1.5 ⁸¹
			340.1 ± 4.6 ⁴⁷		
Acetic acid	CH ₃	348.4 348.3 ^b	348.6 ± 2.1 ⁵⁴	340.3 341.5 ^b	341.5 ± 2.0 ⁵⁴
			348.7 ± 2.2 ⁷⁹		341.7 ± 2.0 ⁷⁹
			348.1 ± 2.2 ⁸²		341.1 ± 2.0 ⁸²
			348.2 ± 1.4 ⁵⁵		
			343.2 ± 0.7 ⁴⁷		
Propanoic acid	CH ₂ CH ₃	347.9	347.4 ± 2.9 ⁵⁴ 347.4 ± 2.2 ⁸⁰	340.0	340.3 ± 2.0 ⁵⁴ 340.4 ± 2.0 ⁸⁰
Isobutyric acid	CH(CH ₃) ₂	346.7	346.0 ± 2.1 ⁸⁰	338.9	339.0 ± 2.0 ⁸⁰
Trimethylacetic acid	C(CH ₃) ₃	345.3	345.0 ± 2.1 ⁸³	337.7	338.0 ± 2.0 ⁸³
			344.6 ± 2.1 ⁸⁰		337.6 ± 2.0 ⁸⁰
Butanoic acid	CH ₂ CH ₂ CH ₃	347.3	346.5 ± 2.2 ⁵⁴	339.4	339.5 ± 2.0 ⁵⁴
			346.8 ± 2.0 ⁸⁴		339.5 ± 2.0 ⁸⁰
Isovaleric acid	CH ₂ CH(CH ₃) ₂	345.9	346.7 ± 2.1 ⁸²	338.5	339.7 ± 2.0 ⁸²
			345.5 ± 2.1 ⁸⁰		338.5 ± 2.0 ⁸⁰
Tert-butyl-acetic acid	CH ₂ C(CH ₃) ₃	344.8	345.1 ± 2.1 ⁸² 344.8 ± 2.1 ⁸⁰	337.4	338.1 ± 2.0 ⁸² 337.8 ± 2.0 ⁸⁰
Pentanoic acid	CH ₂ CH ₂ CH ₂ CH ₃	347.1	346.1 ± 2.4 ⁵⁹	339.1	339.1 ± 2.3 ⁵⁹
			346.2 ± 2.1 ⁸⁰		339.2 ± 2.0 ⁸⁰
Isohexanoic acid	CH ₂ CH ₂ CH(CH ₃) ₂	346.5		338.7	
Tert-butyl-propanoic acid	CH ₂ CH ₂ C(CH ₃) ₃	346.2		337.8	
Phenol		349.5	348.3 ± 1.7 ⁸⁵	342.1	341.5 ± 1.8 ⁸⁵
			348.0 ± 1.0 ⁵⁵		343.4 ± 2.0 ⁵⁴
			350.3 ± 2.3 ⁵⁴		342.3 ± 2.0 ⁸⁶
			349.1 ± 2.1 ⁸⁶		340.8 ± 1.9 ⁸⁹
			350.8 ± 0.9 ⁸⁷		
			350.4 ± 3.1 ⁸⁸		
<i>o</i> -cresol		348.6	347.5 ± 1.9 ⁸⁹	341.4	342.7 ± 2.0 ⁹⁰
			350.7 ± 2.9 ⁹⁰		342.0 ± 2.0 ⁷⁹
<i>m</i> -cresol		349.9	349.5 ± 2.2 ⁷⁹	342.7	343.8 ± 2.0 ⁹⁰
			350.7 ± 2.3 ⁹⁰		342.7 ± 2.0 ⁷⁹
<i>p</i> -cresol		350.7	349.5 ± 2.1 ⁷⁹	342.9	344.7 ± 2.0 ⁹⁰
			351.6 ± 2.3 ⁹⁰		343.4 ± 2.0 ⁸⁶
			350.2 ± 2.1 ⁸⁶		343.4 ± 2.0 ⁸⁶
			350.2 ± 2.1 ⁷⁹		343.4 ± 2.0 ⁷⁹

^a Numbers in superscript provide the reference for each literature value. ^b CCSD(T) values.

Table 3.2. Calculated Gas Phase Acidities of *o*-, *m*-, and *p*- Cresols in kcal/mol at Different Computational Levels

Substituted Anions	Property	B3LYP	G3	G3MP2	G3B3	G3MP2B3	G4	G4MP2	MP2/aD
<i>o</i> -cresol	ΔH	349.4	348.7	348.6	348.6	348.6	348.5	348.5	343.7
	ΔG	342.8	341.4	341.4	341.1	341.1	340.9	340.9	336.2
<i>m</i> -cresol	ΔH	350.3	349.8	349.9	349.8	349.8	349.6	349.6	345.2
	ΔG	343.3	342.7	342.7	343.0	342.9	342.0	342.1	337.5
<i>p</i> -cresol	ΔH	351.0	350.7	350.7	350.7	350.7	350.6	350.6	345.9
	ΔG	343.5	342.9	342.9	343.2	343.2	343.0	343.0	338.0
<i>m</i> -cresol – <i>o</i> -cresol	$\Delta\Delta H$	0.9	1.1	1.3	1.2	1.2	1.1	1.1	1.5
	$\Delta\Delta G$	0.5	1.3	1.3	1.9	1.8	1.1	1.2	1.3
<i>p</i> -cresol – <i>o</i> -cresol	$\Delta\Delta H$	1.6	2.0	2.1	2.1	2.1	2.1	2.1	2.2
	$\Delta\Delta G$	0.7	1.5	1.5	2.1	2.1	2.1	2.1	1.8

Table 3.3. Experimental, Theoretical, and Literature Gas-Phase Acidities in kcal/mol for Tyrosine, Phenylalanine, and HPB.

	Experimental Method	Deprotonation Site	Conformation	Experimental		Computational Method	Theoretical		
				$\Delta G_{298}^{\text{gas}}$ (GA)	$\Delta H_{298}^{\text{gas}}$		$\Delta G_{298}^{\text{gas}}$	$\Delta H_{298}^{\text{gas}}$	
Tyrosine									
Present Work	Thermokinetic Bracketing	carboxylate	Unfolded	333.5 ± 2.4	– ^a	G3(MP2)	331.4	340.2	
		phenolate	Unfolded	332.4 ± 2.2	–		332.9 ^b	340.5 ^b	
		carboxylate	Folded	324.7 ± 3.6	–		330.4 ^b	338.3 ^b	
			carboxylate	Unfolded			G3B3	331.1	339.9
			phenolate	Unfolded				332.6 ^b	340.3 ^b
			carboxylate	Folded				330.5 ^b	338.3 ^b
			carboxylate	Unfolded			G3	331.2	340
			phenolate	Unfolded				332.9 ^b	340.5 ^b
			carboxylate	Folded				330.2 ^b	338.1 ^b
			carboxylate	Unfolded			G4	331.2	339.6
			phenolate	Unfolded				332.3 ^b	340.1 ^b
			carboxylate	Folded				330.6 ^b	338.3 ^b
		carboxylate	Unfolded			CCSD(T) /aT//MP2/aT	331.3	339.7	
		phenolate	Unfolded				332.3 ^b	340.1 ^b	
		carboxylate	Folded				330.6 ^b	338.2 ^b	
Kass ^c	Equilibrium	carboxylate		332.5 ± 1.5	340.7 ± 1.5	B3LYP	–	339.5	
		phenolate		–	–		–	338.3	
		carboxylate		–	–	G3B3	–	339.9	
		phenolate		–	–		–	339.0	
Poutsma ^d	Extended Kinetic	–		–	337.7 ± 2.6	B3LYP/6-31+G*	–	339.2	
O'Hair ^e	Kinetic	–		329.5 ± 3.0	336.4 ± 3.1	–	–	–	

Phenylalanine								
Present Work	Thermokinetic	carboxylate	Folded	332.5 ± 2.2	–	G3(MP2)	330.7	338.5
	Bracketing	carboxylate	Unfolded				331	339.6
Poutsma ^e	Extended Kinetic	carboxylate		–	338.9 ± 4.3	B3LYP/6-31+G*	–	338.7
O'Hair ^e	Kinetic	carboxylate		329.6 ± 3.0	336.5 ± 3.1	–	–	–
4-(4-Hydroxyphenyl)-2-butanone (HPB)								
Present Work	Thermokinetic	phenolate	Folded	339.6 ± 3.0	–	G3(MP2)	337.8	344
	Bracketing	phenolate	Unfolded				340.1	347.5

^a – indicates that no value was reported. ^b From reference ⁴. ^c From reference ⁹. ^d From reference ¹⁴. ^e From reference ¹³.

Table 3.4. Reaction Efficiencies for the Proton Transfer Reactions of Tyrosine, Phenylalanine, and HPB with Reference Compounds.

Reference Compound	GA ^a (kcal/mol)	Average Reaction Efficiency (\pm Standard Deviation)			
		HPB ^b	Tyrosine (Protic Solvent) ^c	Tyrosine (Aprotic Solvent) ^d	Phenylalanine
Chloroform	349.9 \pm 2	0.01 \pm 0.00	– ^e	–	–
4-Trifluoromethyl aniline	346.0 \pm 2	–	–	–	–
Phenol	342.3 \pm 2	–	–	–	–
Acetic acid	341.1 \pm 2	0.07 \pm 0.01	–	–	–
		BREAK^f			
Formic acid	339.2 \pm 2	0.34 \pm 0.01	–	–	–
Isovaleric acid	338.5 \pm 2	0.41 \pm 0.11	–	–	–
Trimethylacetic acid	337.6 \pm 2	0.41 \pm 0.06	NR ^g	NR	NR
p-Chlorophenol	336.2 \pm 2	0.55 \pm 0.04	NR	NR	NR
Ethyl cyanoacetate	333.6 \pm 2	–	0.23 \pm 0.06 (24 \pm 9%) ^h 0.03 \pm 0.04 (76 \pm 9%)	0.028 \pm 0.019 (56 \pm 7%) NR (44 \pm 7%)	0.007 \pm 0.001
			BREAK		
4-Amino-2,3,5,6-tetrafluoropyridine	332.8 \pm 2	–	0.53 \pm 0.25 (30 \pm 8%) 0.07 \pm 0.01 (70 \pm 8%)	0.033 \pm 0.011 (61 \pm 11%) 0.005 \pm 0.004 (39 \pm 11%)	0.02 \pm 0.00
			BREAK	BREAK	BREAK
3-Trifluoromethyl phenol	332.4 \pm 2	–	0.29 \pm 0.02	0.29 \pm 0.03 (58 \pm 13%) 0.007 \pm 0.002 (42 \pm 13%)	0.36 \pm 0.02
3,3,3-Trifluoropropionic acid	326.9 \pm 2	–	0.75 \pm 0.09	0.38 \pm 0.04 (59 \pm 3%) 0.13 \pm 0.04 (41 \pm 3%)	0.79 \pm 0.08
				BREAK	

Difluoroacetic acid	323.8 ± 2	–	–	0.45 ± 0.10 (46 ± 13%)	–
				0.33 ± 0.08 (54 ± 13%)	
Pentafluorophenol	320.8 ± 2	–	–	1.2 ± 0.27 (58 ± 15%)	–
				0.51 ± 0.03 (42 ± 15%)	

^a All reference compound GAs were obtained from reference ¹⁷. ^b HPB stands for 4-(4-Hydroxyphenyl)-2-butanone.

^c Protic solvent system of 49.5: 49.5: 1 (v/v/v) CH₃OH:H₂O:NH₄OH. ^d Aprotic solvent system of 99: 1 (v/v) CH₃CN: H₂O.

^e "–" indicates that no reaction was performed. ^f "BREAK" indicates the point where experimental GA was assigned.

^g "NR" indicates that no reaction was observed. ^h Two reaction efficiencies and the relative abundances of each are listed for reactions in which bimodal kinetics plots indicate the existence of two ion populations reacting at two different rates.

Table 3.5. Reactions of Trimethylsilyl Azide with Deprotonated Tyrosine Electrospayed from Various Solvents.

Solvent A ^a		Solvent B ^a		Solvent C ^a		Ratios of TMSN ₃ Product Ions	
Name	%	Name	%	Name	%	% Phenoxide	% Carboxylate
H ₂ O	100					62	38
H ₂ O	99	NH ₄ OH	1			68	32
H ₂ O	50	CH ₃ OH	50			61	39
H ₂ O	49.5	CH ₃ OH	49.5	NH ₄ OH	1	67	33
H ₂ O	25	CH ₃ OH	75			65	35
H ₂ O	24.5	CH ₃ OH	74.5	NH ₄ OH	1	71	29
H ₂ O	49.5	CH ₃ CN	49.5	NH ₄ OH	1	56	44
H ₂ O	50	CH ₃ CN	50			63	37
CH ₃ CN	49.5	CH ₃ OH	49.5	NH ₄ OH	1	67	33
CH ₃ CN	50	CH ₃ OH	50			63	37
CH ₃ CN	97	H ₂ O	2	NH ₄ OH	1	70	30
CH ₃ CN	98	H ₂ O	1	NH ₄ OH	1	53	47
CH ₃ CN	98	H ₂ O	2			49	51
CH ₃ CN	98	CH ₃ OH	1	NH ₄ OH	1	3	97
CH ₃ CN	99	NH ₄ OH	1			35	65
CH ₃ CN	99	CH ₃ OH	1			1	99
CH ₃ CN	99	H ₂ O	1			0	100
CH ₃ CN	100					0	100
Acetone	100					0	100
DMF	45	THF	45	DMSO	10	0	100

^a All solvent compositions are % by volume.

Table 3.6. Relative Free Energies in kcal/mol of Solvated Tyrosine.

Free Energy Type ^a	# of waters	Folded Carboxylate	Unfolded Carboxylate	Unfolded Phenoxide	Folded Phenoxide
$\Delta\Delta G_{\text{gas}}$	0	1.0	0.0	1.3	8.4
$\Delta\Delta G_{\text{solv}}(\text{H}_2\text{O})$	0	1.5	0.0	6.4	10.9
$\Delta\Delta G_{\text{solv}}(\text{CH}_3\text{OH})$	0	1.6	0.0	6.5	11.0
$\Delta\Delta G_{\text{solv}}(\text{CH}_3\text{CN})$	0	1.4	0.0	6.5	10.9
$\Delta\Delta G_{\text{gas}}$	2	0.0	6.5	4.3	6.0
$\Delta\Delta G_{\text{solv}}(\text{H}_2\text{O})$	2	2.5	0.0	5.8	8.5
$\Delta\Delta G_{\text{gas}}$	3	0.0	6.2	4.0	7.1
$\Delta\Delta G_{\text{solv}}(\text{H}_2\text{O})$	3	2.5	0.0	8.8	8.7
$\Delta\Delta G_{\text{gas}}$	6	0.0	0.5	4.4	7.8
$\Delta\Delta G_{\text{solv}}(\text{H}_2\text{O})$	6	4.0	0.0	4.4	8.1

^a $\Delta\Delta G_{\text{gas}}$ = relative free energy in the gas phase. $\Delta\Delta G_{\text{solv}}(\text{H}_2\text{O})$ = relative free energy in aqueous solution modeled by a self-consistent reaction field. $\Delta\Delta G_{\text{solv}}(\text{CH}_3\text{OH})$ = relative free energy in methanol solution modeled by a self-consistent reaction field. $\Delta\Delta G_{\text{solv}}(\text{CH}_3\text{CN})$ = relative free energy in acetonitrile solution modeled by a self-consistent reaction field.

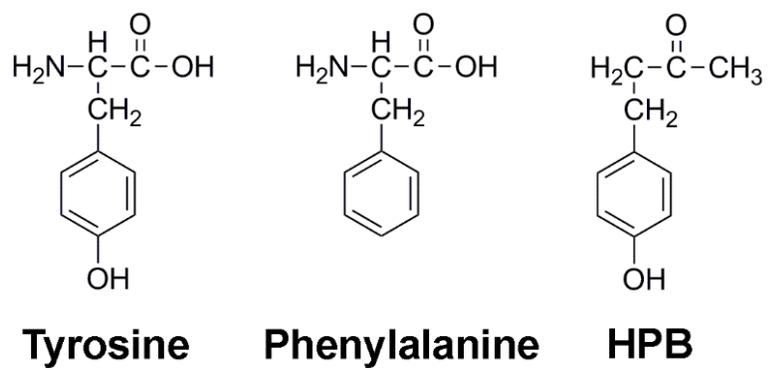
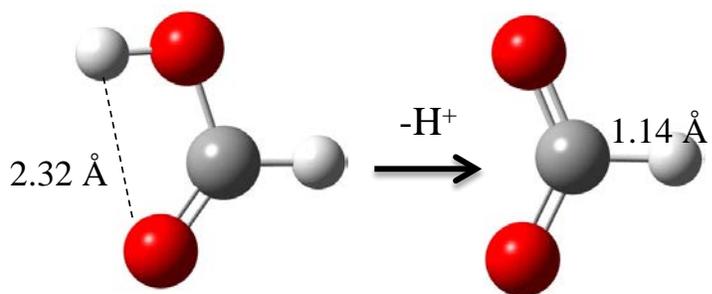
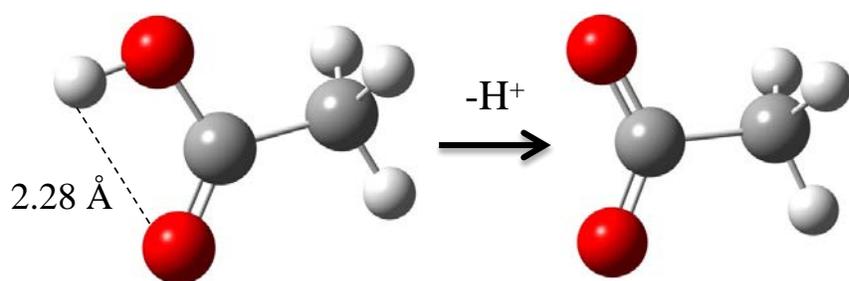


Figure 3.1 Structures of Tyrosine, Phenylalanine, and HPB

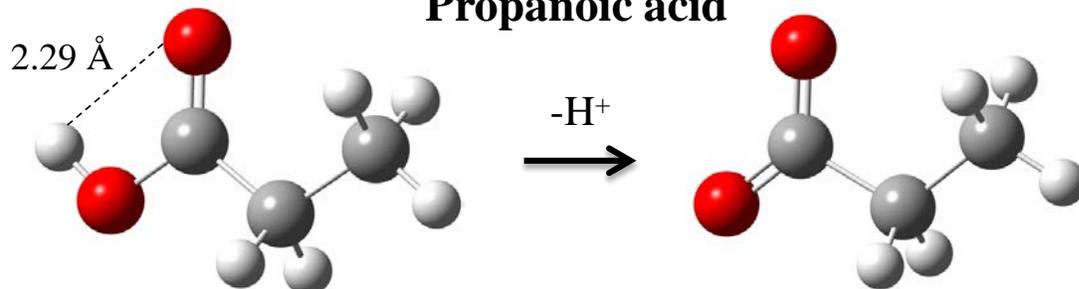
Formic acid



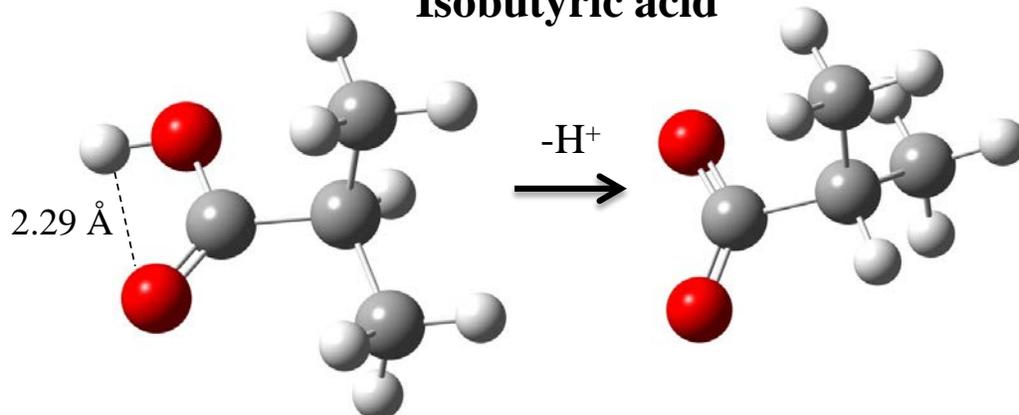
Acetic acid

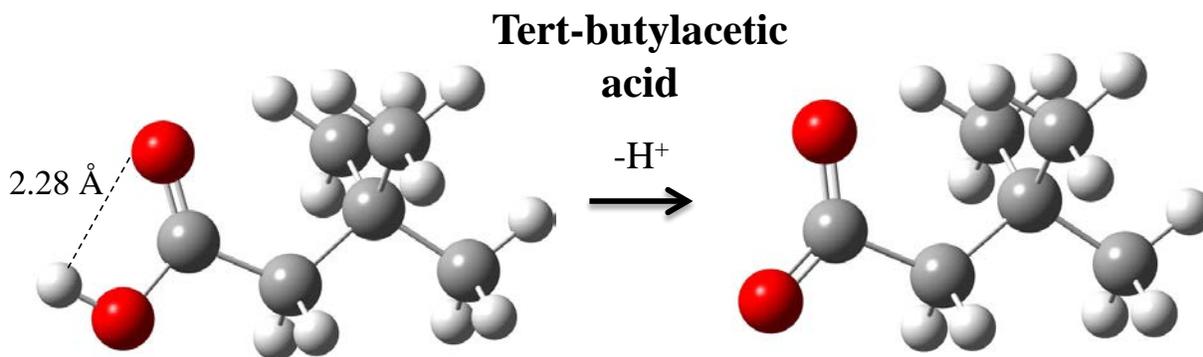
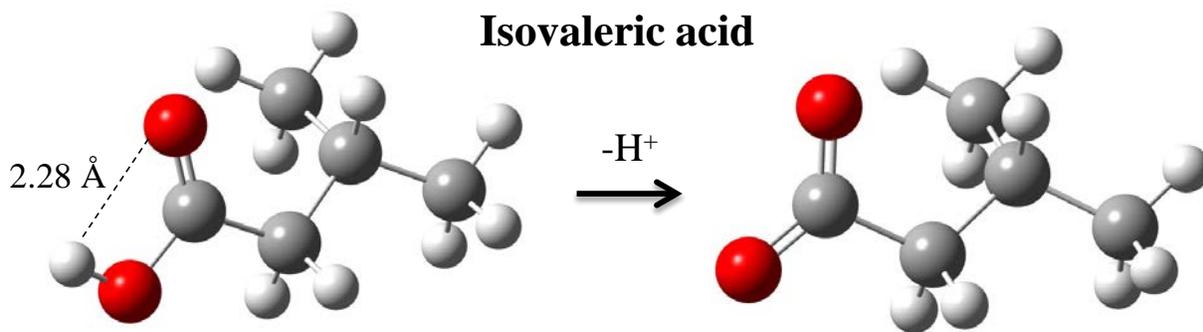
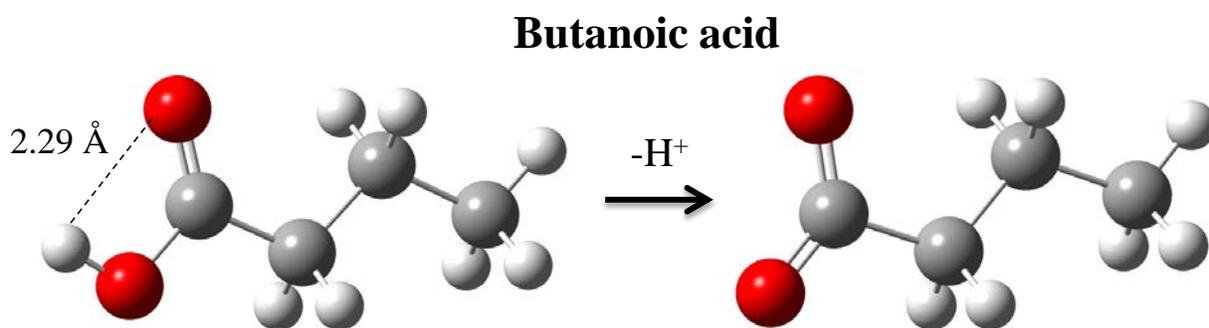
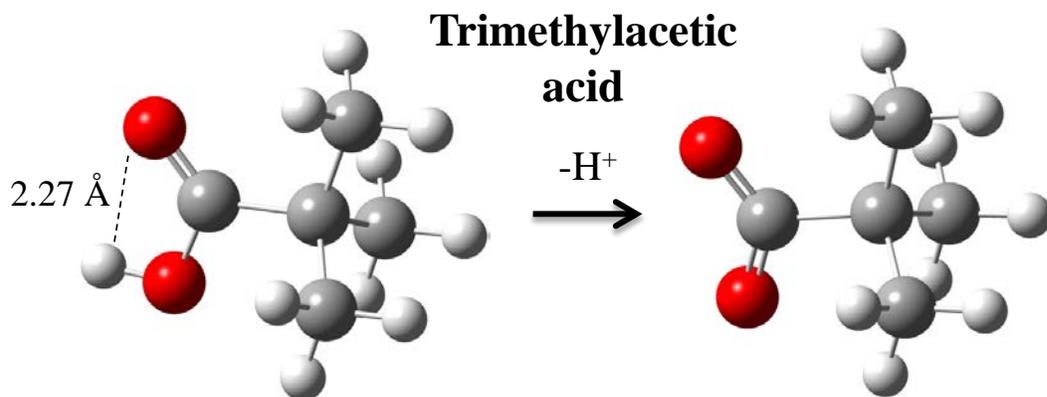


Propanoic acid



Isobutyric acid





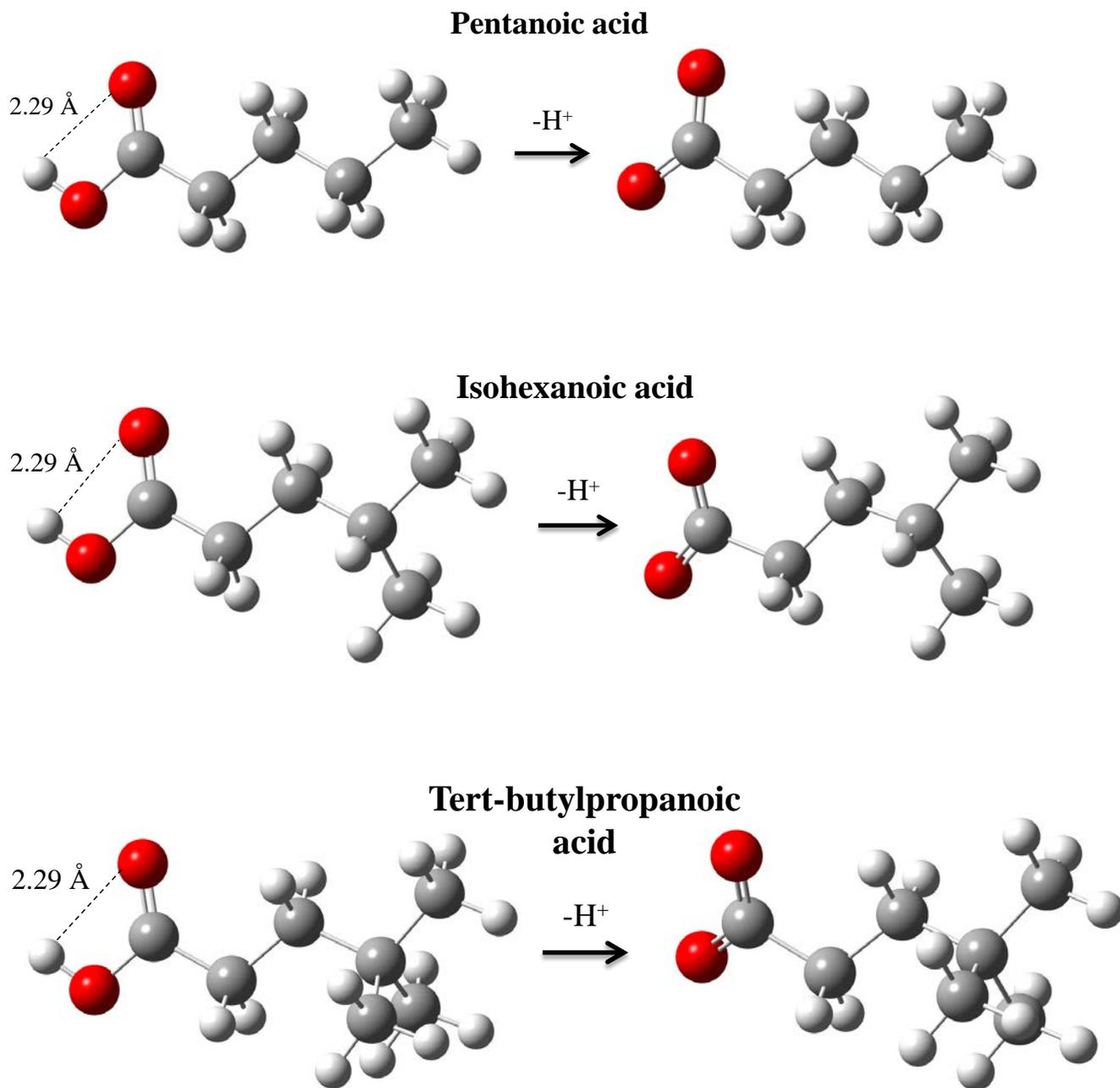


Figure 3.2. G3(MP2) optimized structures for carboxylic acids and anions. Important hydrogen bonds in Å.

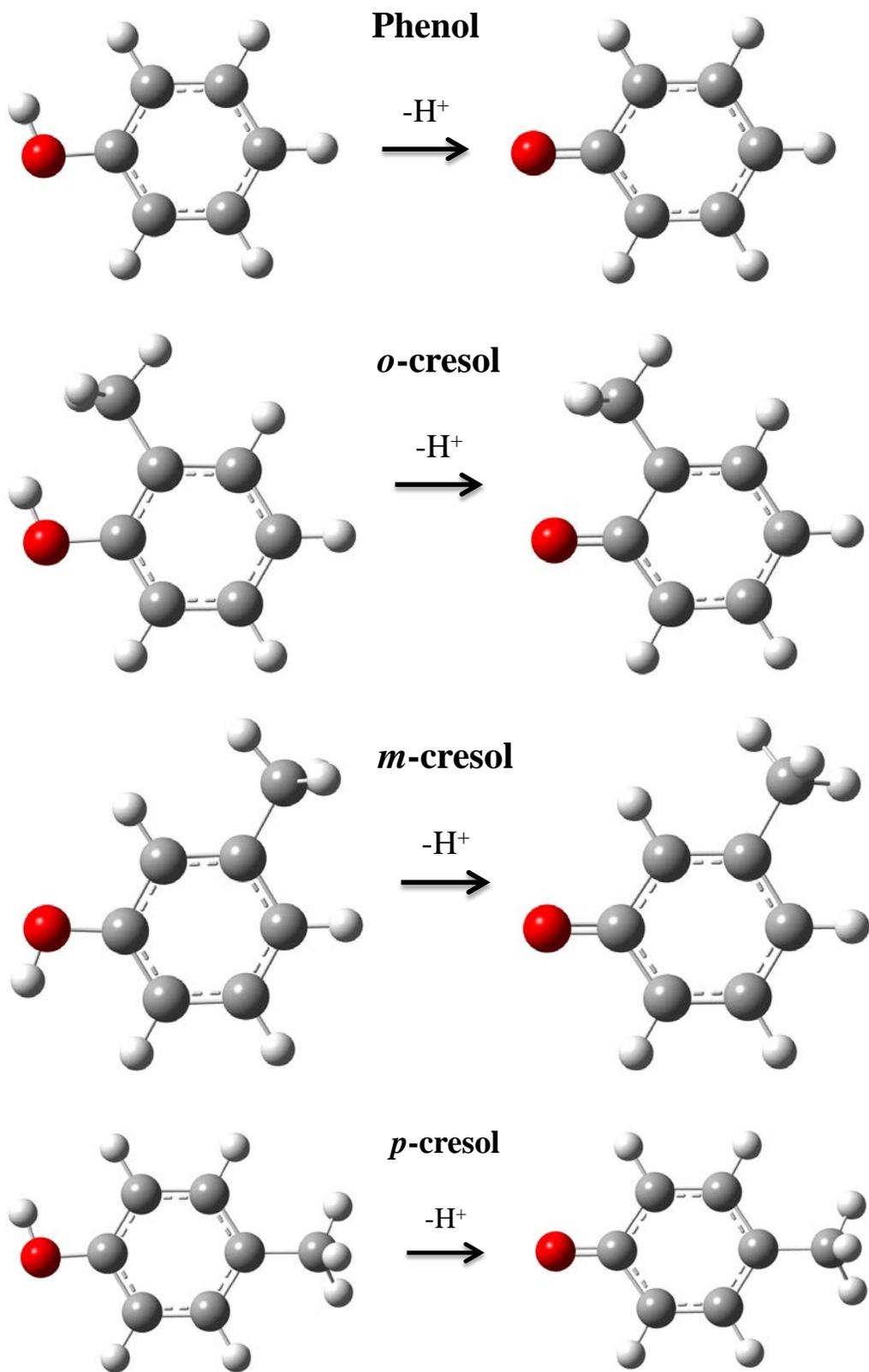
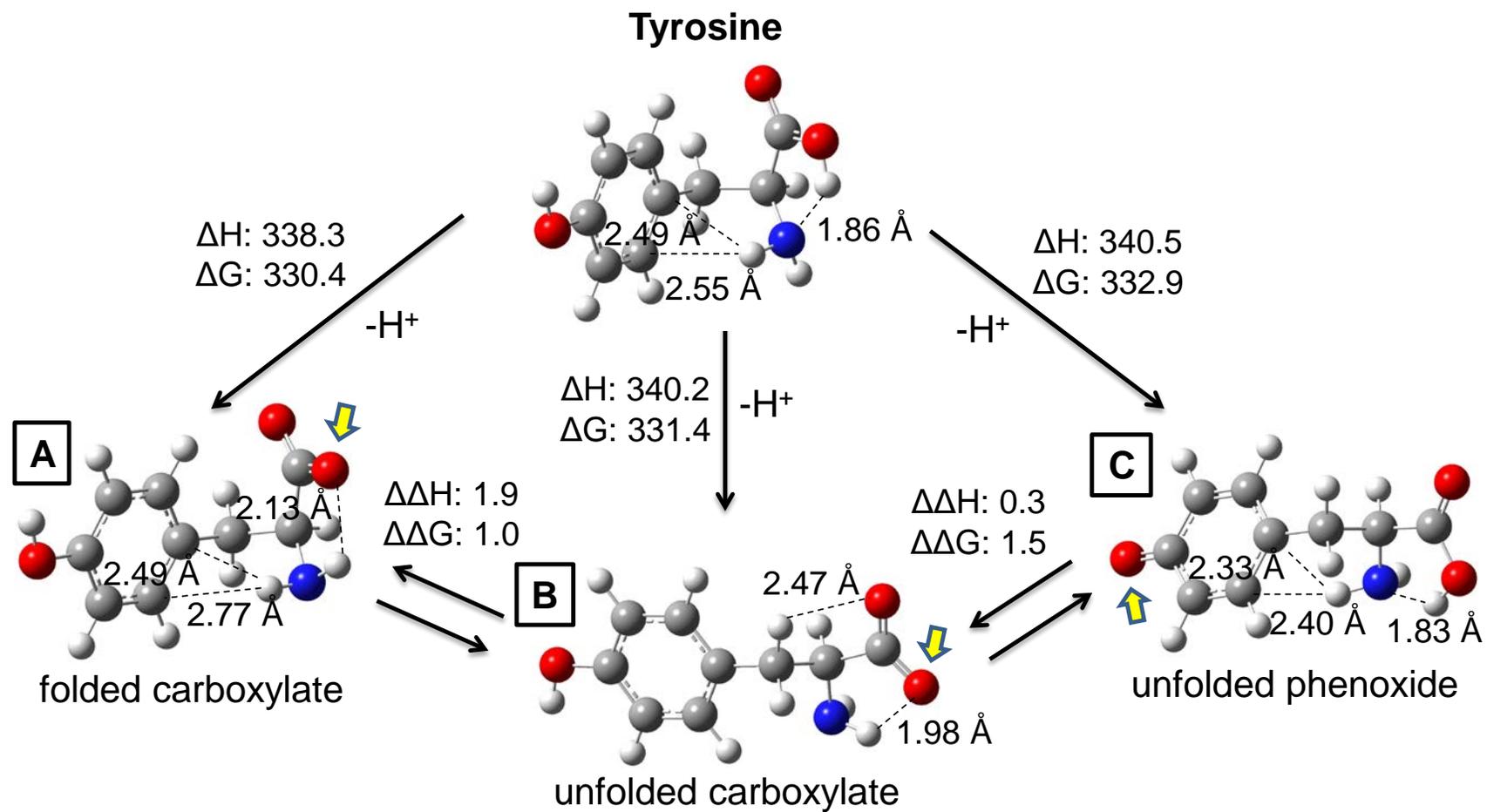
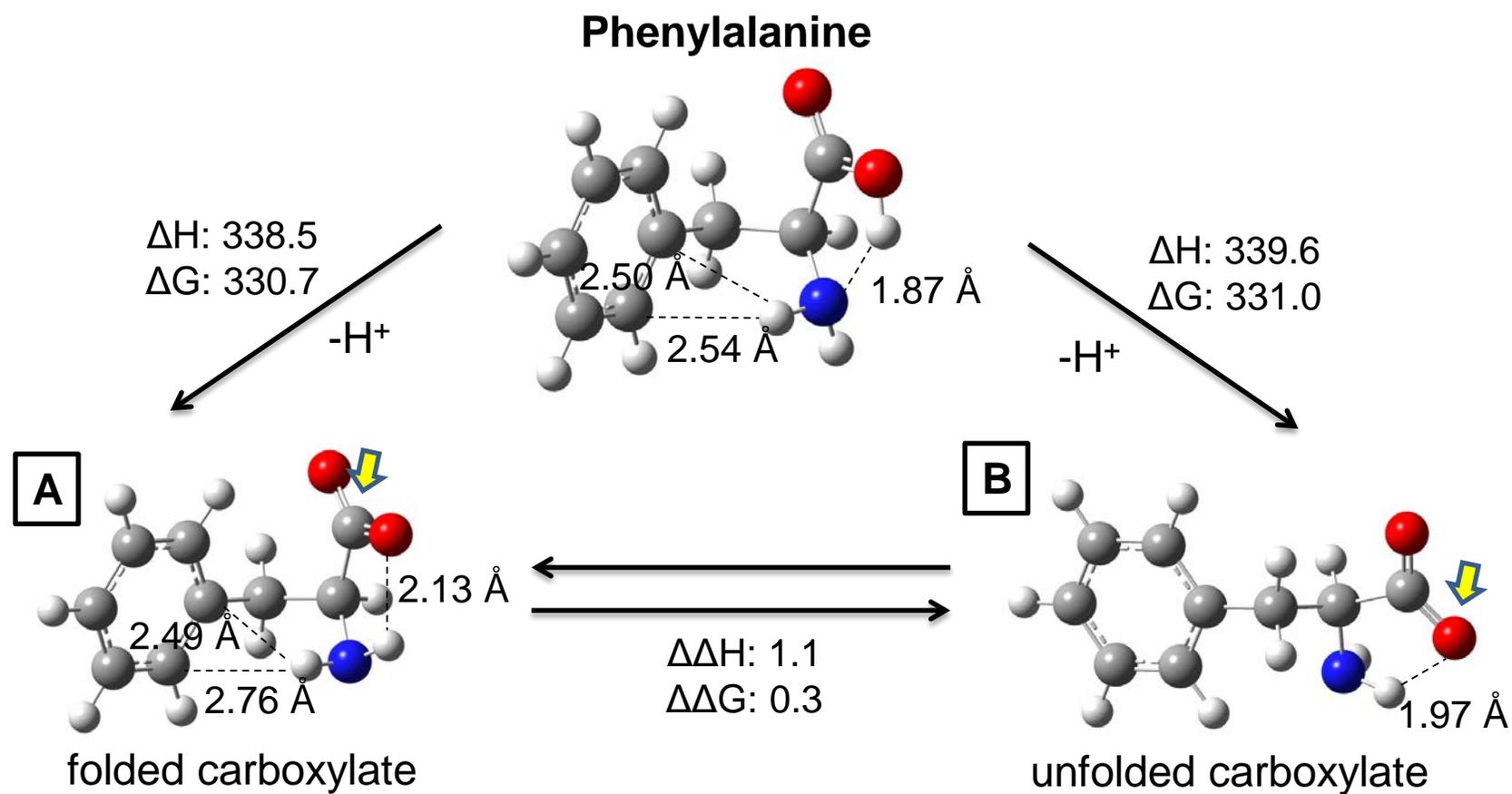


Figure 3.3. G3(MP2) optimized structures for phenol, *o*-, *m*-. and *p*-cresol and anions.

(a)



(b)



(c)

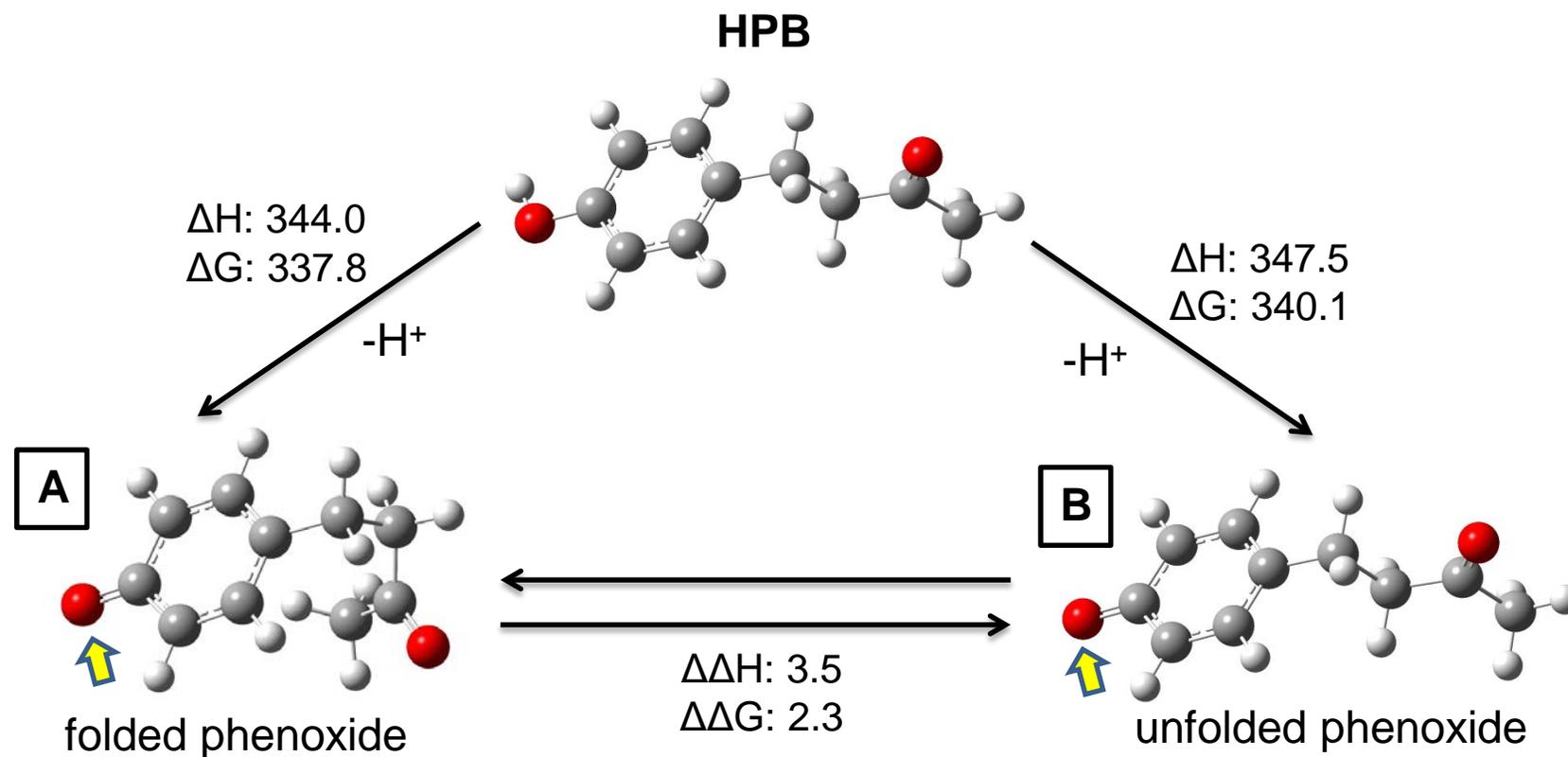
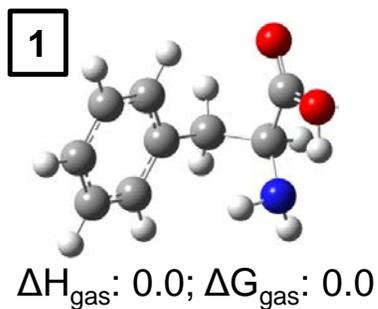
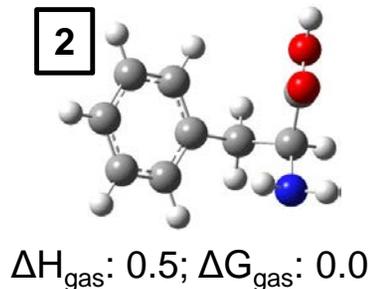


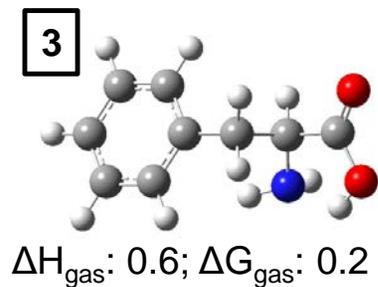
Figure 3.4. G3(MP2) results showing the structures of the most stable neutral acid and the lowest energy anions resulting from proton loss. Energies are in kcal/mol. Hydrogen bond distances in Å. The yellow arrows show the deprotonation site. (a) Tyrosine (b) Phenylalanine (c) HPB.



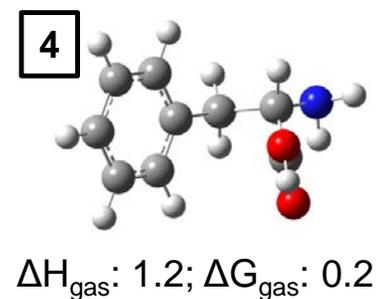
$\Delta\Delta G_{\text{solv}}: 0.0$ (H_2O)
 $\Delta\Delta G_{\text{solv}}: 0.0$ (CH_3OH)
 $\Delta\Delta G_{\text{solv}}: 0.0$ (CH_3CN)



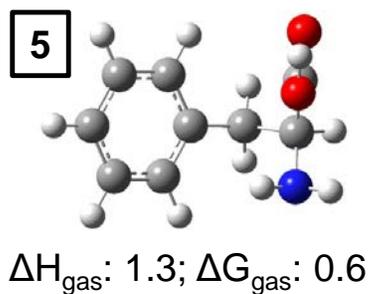
$\Delta\Delta G_{\text{solv}}: 3.3$ (H_2O)
 $\Delta\Delta G_{\text{solv}}: 3.0$ (CH_3OH)
 $\Delta\Delta G_{\text{solv}}: 3.2$ (CH_3CN)



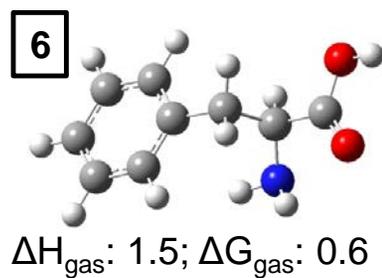
$\Delta\Delta G_{\text{solv}}: 0.2$ (H_2O)
 $\Delta\Delta G_{\text{solv}}: 0.3$ (CH_3OH)
 $\Delta\Delta G_{\text{solv}}: 0.6$ (CH_3CN)



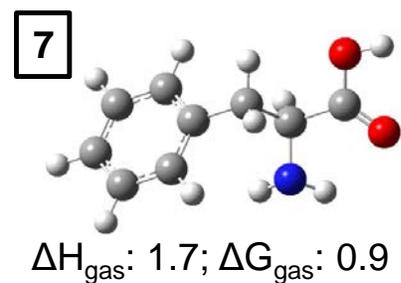
$\Delta\Delta G_{\text{solv}}: 3.8$ (H_2O)
 $\Delta\Delta G_{\text{solv}}: 3.5$ (CH_3OH)
 $\Delta\Delta G_{\text{solv}}: 3.8$ (CH_3CN)



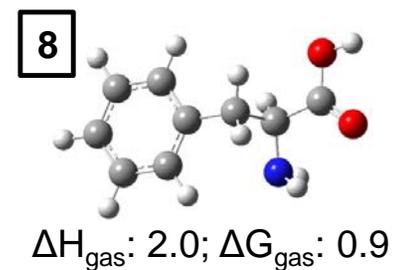
$\Delta\Delta G_{\text{solv}}: 4.1$ (H_2O)
 $\Delta\Delta G_{\text{solv}}: 3.9$ (CH_3OH)
 $\Delta\Delta G_{\text{solv}}: 4.1$ (CH_3CN)



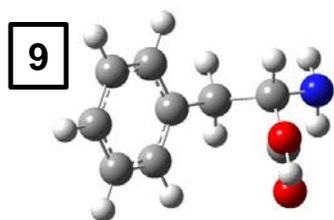
$\Delta\Delta G_{\text{solv}}: 3.7$ (H_2O)
 $\Delta\Delta G_{\text{solv}}: 3.4$ (CH_3OH)
 $\Delta\Delta G_{\text{solv}}: 3.6$ (CH_3CN)



$\Delta\Delta G_{\text{solv}}: 4.2$ (H_2O)
 $\Delta\Delta G_{\text{solv}}: 3.9$ (CH_3OH)
 $\Delta\Delta G_{\text{solv}}: 4.2$ (CH_3CN)

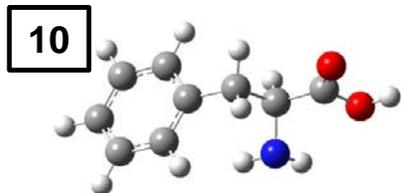


$\Delta\Delta G_{\text{solv}}: 4.6$ (H_2O)
 $\Delta\Delta G_{\text{solv}}: 4.3$ (CH_3OH)
 $\Delta\Delta G_{\text{solv}}: 4.7$ (CH_3CN)



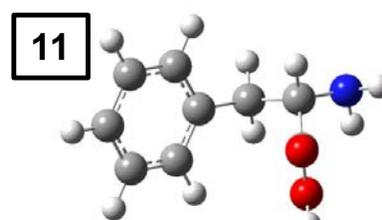
$$\Delta H_{\text{gas}}: 2.1; \Delta G_{\text{gas}}: 1.1$$

$$\begin{aligned} \Delta\Delta G_{\text{solv}}: 3.9 (\text{H}_2\text{O}) \\ \Delta\Delta G_{\text{solv}}: 3.6 (\text{CH}_3\text{OH}) \\ \Delta\Delta G_{\text{solv}}: 3.8 (\text{CH}_3\text{CN}) \end{aligned}$$



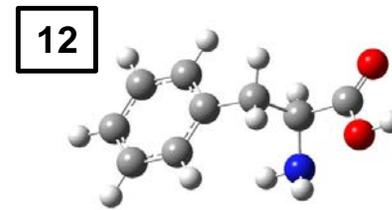
$$\Delta H_{\text{gas}}: 2.0; \Delta G_{\text{gas}}: 1.2$$

$$\begin{aligned} \Delta\Delta G_{\text{solv}}: 4.2 (\text{H}_2\text{O}) \\ \Delta\Delta G_{\text{solv}}: 4.0 (\text{CH}_3\text{OH}) \\ \Delta\Delta G_{\text{solv}}: 4.2 (\text{CH}_3\text{CN}) \end{aligned}$$



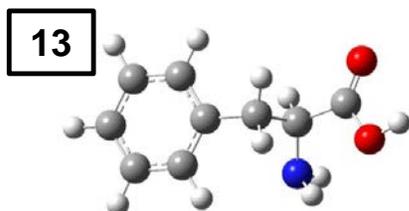
$$\Delta H_{\text{gas}}: 2.2; \Delta G_{\text{gas}}: 1.3$$

$$\begin{aligned} \Delta\Delta G_{\text{solv}}: 3.9 (\text{H}_2\text{O}) \\ \Delta\Delta G_{\text{solv}}: 3.6 (\text{CH}_3\text{OH}) \\ \Delta\Delta G_{\text{solv}}: 3.9 (\text{CH}_3\text{CN}) \end{aligned}$$



$$\Delta H_{\text{gas}}: 2.5; \Delta G_{\text{gas}}: 1.5$$

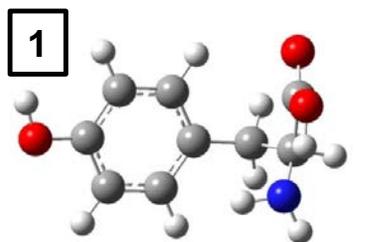
$$\begin{aligned} \Delta\Delta G_{\text{solv}}: 4.2 (\text{H}_2\text{O}) \\ \Delta\Delta G_{\text{solv}}: 3.9 (\text{CH}_3\text{OH}) \\ \Delta\Delta G_{\text{solv}}: 4.1 (\text{CH}_3\text{CN}) \end{aligned}$$



$$\Delta H_{\text{gas}}: 3.3; \Delta G_{\text{gas}}: 1.8$$

$$\begin{aligned} \Delta\Delta G_{\text{solv}}: 4.7 (\text{H}_2\text{O}) \\ \Delta\Delta G_{\text{solv}}: 4.4 (\text{CH}_3\text{OH}) \\ \Delta\Delta G_{\text{solv}}: 4.8 (\text{CH}_3\text{CN}) \end{aligned}$$

Figure 3.5. G3(MP2) optimized structures for phenylalanine conformers. Relative energies in the gas and solution phases are given in kcal/mol.

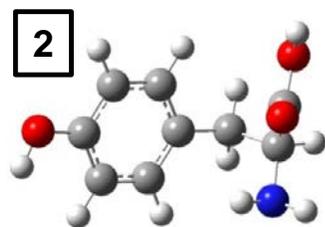


$$\Delta H_{\text{gas}}: 0.0; \Delta G_{\text{gas}}: 0.0$$

$$\Delta\Delta G_{\text{solv}}: 0.0 (\text{H}_2\text{O})$$

$$\Delta\Delta G_{\text{solv}}: 0.0 (\text{CH}_3\text{OH})$$

$$\Delta\Delta G_{\text{solv}}: 0.0 (\text{CH}_3\text{CN})$$

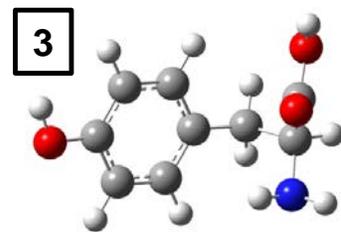


$$\Delta H_{\text{gas}}: 0.7; \Delta G_{\text{gas}}: 0.1$$

$$\Delta\Delta G_{\text{solv}}: 3.0 (\text{H}_2\text{O})$$

$$\Delta\Delta G_{\text{solv}}: 2.8 (\text{CH}_3\text{OH})$$

$$\Delta\Delta G_{\text{solv}}: 3.6 (\text{CH}_3\text{CN})$$

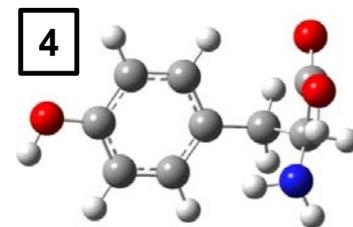


$$\Delta H_{\text{gas}}: 0.7; \Delta G_{\text{gas}}: 0.2$$

$$\Delta\Delta G_{\text{solv}}: 3.0 (\text{H}_2\text{O})$$

$$\Delta\Delta G_{\text{solv}}: 2.8 (\text{CH}_3\text{OH})$$

$$\Delta\Delta G_{\text{solv}}: 3.0 (\text{CH}_3\text{CN})$$

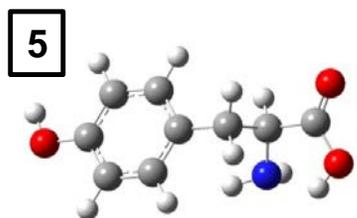


$$\Delta H_{\text{gas}}: 0.3; \Delta G_{\text{gas}}: 0.3$$

$$\Delta\Delta G_{\text{solv}}: -0.5 (\text{H}_2\text{O})$$

$$\Delta\Delta G_{\text{solv}}: -0.4 (\text{CH}_3\text{OH})$$

$$\Delta\Delta G_{\text{solv}}: -0.3 (\text{CH}_3\text{CN})$$

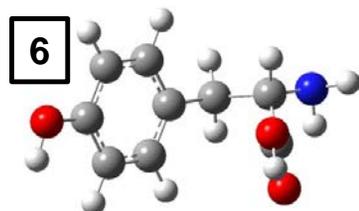


$$\Delta H_{\text{gas}}: 0.9; \Delta G_{\text{gas}}: 0.4$$

$$\Delta\Delta G_{\text{solv}}: 0.2 (\text{H}_2\text{O})$$

$$\Delta\Delta G_{\text{solv}}: 0.1 (\text{CH}_3\text{OH})$$

$$\Delta\Delta G_{\text{solv}}: 0.3 (\text{CH}_3\text{CN})$$

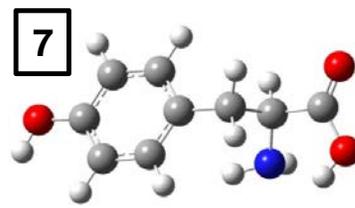


$$\Delta H_{\text{gas}}: 1.5; \Delta G_{\text{gas}}: 0.4$$

$$\Delta\Delta G_{\text{solv}}: 3.7 (\text{H}_2\text{O})$$

$$\Delta\Delta G_{\text{solv}}: 3.4 (\text{CH}_3\text{OH})$$

$$\Delta\Delta G_{\text{solv}}: 3.6 (\text{CH}_3\text{CN})$$

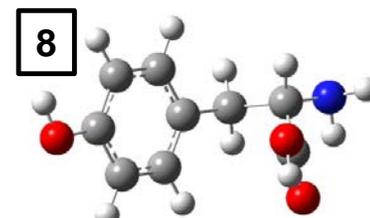


$$\Delta H_{\text{gas}}: 1.0; \Delta G_{\text{gas}}: 0.5$$

$$\Delta\Delta G_{\text{solv}}: 0.2 (\text{H}_2\text{O})$$

$$\Delta\Delta G_{\text{solv}}: 0.1 (\text{CH}_3\text{OH})$$

$$\Delta\Delta G_{\text{solv}}: 0.3 (\text{CH}_3\text{CN})$$

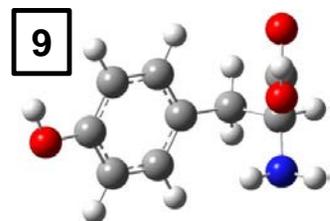


$$\Delta H_{\text{gas}}: 1.5; \Delta G_{\text{gas}}: 0.5$$

$$\Delta\Delta G_{\text{solv}}: 3.7 (\text{H}_2\text{O})$$

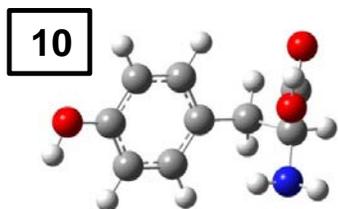
$$\Delta\Delta G_{\text{solv}}: 3.3 (\text{CH}_3\text{OH})$$

$$\Delta\Delta G_{\text{solv}}: 3.6 (\text{CH}_3\text{CN})$$



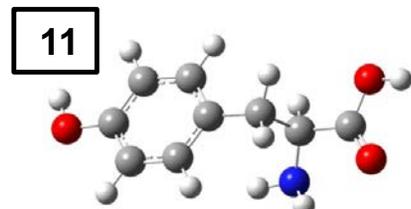
$\Delta H_{\text{gas}}: 1.5; \Delta G_{\text{gas}}: 0.8$

$\Delta\Delta G_{\text{solv}}: 4.0 (\text{H}_2\text{O})$
 $\Delta\Delta G_{\text{solv}}: 3.8 (\text{CH}_3\text{OH})$
 $\Delta\Delta G_{\text{solv}}: 3.9 (\text{CH}_3\text{CN})$



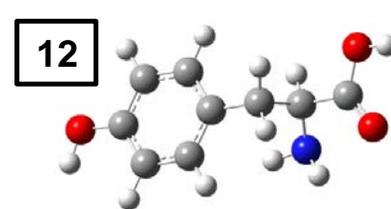
$\Delta H_{\text{gas}}: 1.5; \Delta G_{\text{gas}}: 0.8$

$\Delta\Delta G_{\text{solv}}: 3.9 (\text{H}_2\text{O})$
 $\Delta\Delta G_{\text{solv}}: 3.7 (\text{CH}_3\text{OH})$
 $\Delta\Delta G_{\text{solv}}: 3.8 (\text{CH}_3\text{CN})$



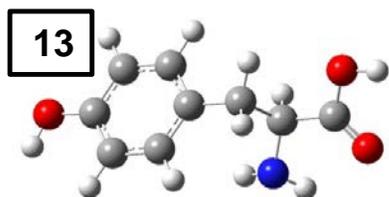
$\Delta H_{\text{gas}}: 1.9; \Delta G_{\text{gas}}: 0.9$

$\Delta\Delta G_{\text{solv}}: 3.5 (\text{H}_2\text{O})$
 $\Delta\Delta G_{\text{solv}}: 3.2 (\text{CH}_3\text{OH})$
 $\Delta\Delta G_{\text{solv}}: 3.4 (\text{CH}_3\text{CN})$



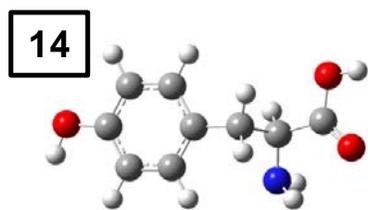
$\Delta H_{\text{gas}}: 1.9; \Delta G_{\text{gas}}: 0.9$

$\Delta\Delta G_{\text{solv}}: 3.5 (\text{H}_2\text{O})$
 $\Delta\Delta G_{\text{solv}}: 3.2 (\text{CH}_3\text{OH})$
 $\Delta\Delta G_{\text{solv}}: 3.4 (\text{CH}_3\text{CN})$



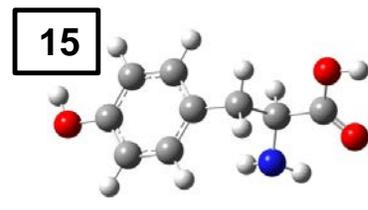
$\Delta H_{\text{gas}}: 2.0; \Delta G_{\text{gas}}: 1.1$

$\Delta\Delta G_{\text{solv}}: 4.0 (\text{H}_2\text{O})$
 $\Delta\Delta G_{\text{solv}}: 3.8 (\text{CH}_3\text{OH})$
 $\Delta\Delta G_{\text{solv}}: 4.0 (\text{CH}_3\text{CN})$



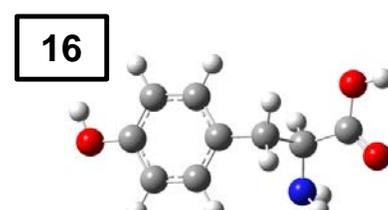
$\Delta H_{\text{gas}}: 2.3; \Delta G_{\text{gas}}: 1.1$

$\Delta\Delta G_{\text{solv}}: 4.4 (\text{H}_2\text{O})$
 $\Delta\Delta G_{\text{solv}}: 4.1 (\text{CH}_3\text{OH})$
 $\Delta\Delta G_{\text{solv}}: 4.4 (\text{CH}_3\text{CN})$



$\Delta H_{\text{gas}}: 2.1; \Delta G_{\text{gas}}: 1.2$

$\Delta\Delta G_{\text{solv}}: 3.8 (\text{H}_2\text{O})$
 $\Delta\Delta G_{\text{solv}}: 3.7 (\text{CH}_3\text{OH})$
 $\Delta\Delta G_{\text{solv}}: 3.9 (\text{CH}_3\text{CN})$



$\Delta H_{\text{gas}}: 2.4; \Delta G_{\text{gas}}: 1.3$

$\Delta\Delta G_{\text{solv}}: 4.6 (\text{H}_2\text{O})$
 $\Delta\Delta G_{\text{solv}}: 4.2 (\text{CH}_3\text{OH})$
 $\Delta\Delta G_{\text{solv}}: 4.5 (\text{CH}_3\text{CN})$

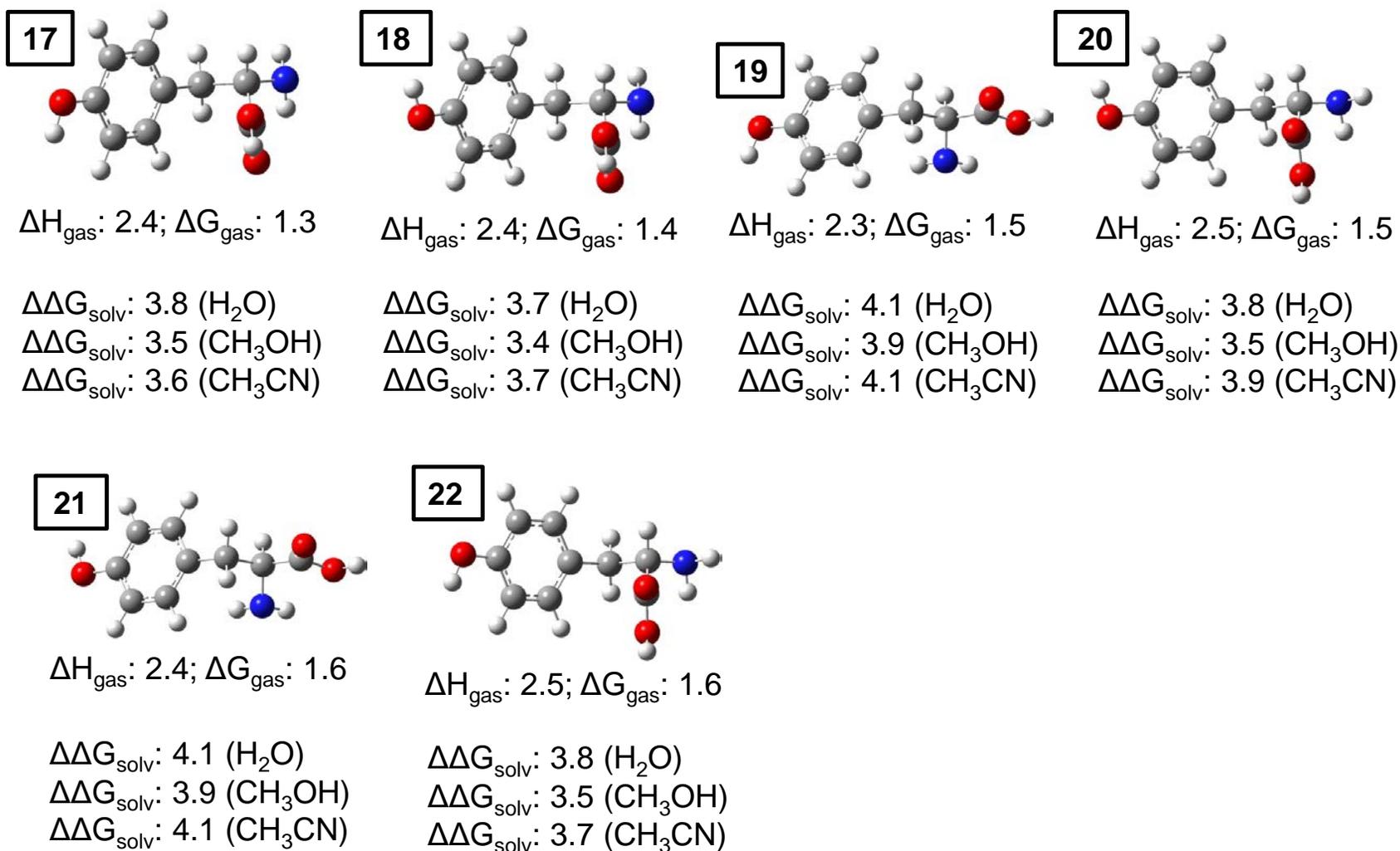
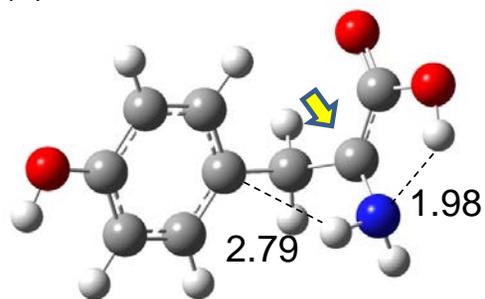
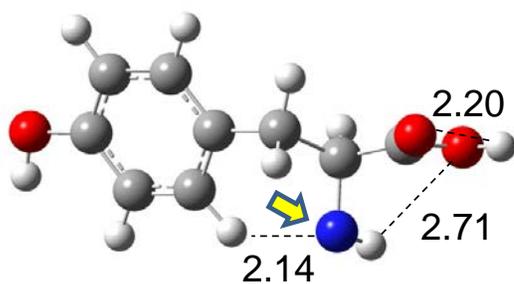


Figure 3.6. G3(MP2) optimized structures for tyrosine conformers. Relative energies in the gas and solution phases are given in kcal/mol.

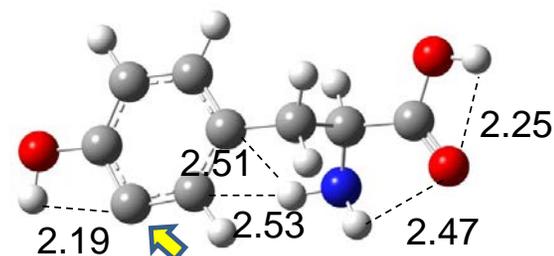
(a)



$\Delta H: 366.3; \Delta G: 358.0$

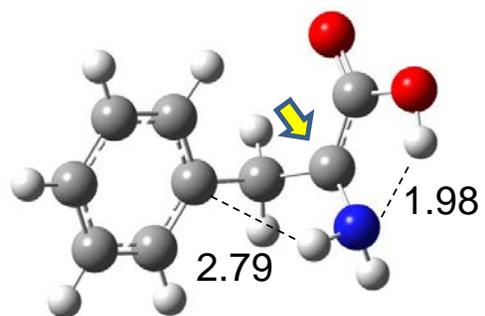


$\Delta H: 381.8; \Delta G: 373.7$

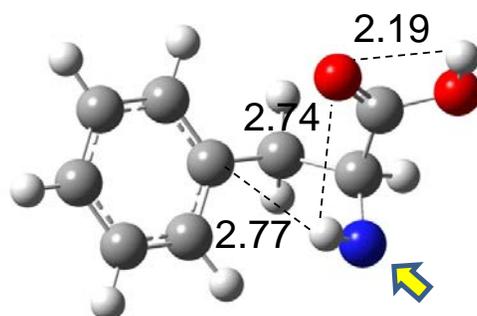


$\Delta H: 382.1; \Delta G: 373.9$

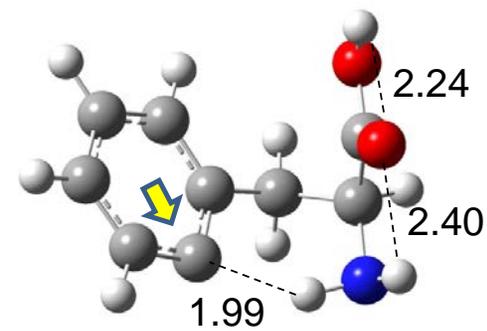
(b)



$\Delta H: 365.5; \Delta G: 357.4$



$\Delta H: 384.4; \Delta G: 376.4$



$\Delta H: 390.3; \Delta G: 382.5$

Figure 3.7. G3(MP2) optimized higher energy amino acid anions. Energies in kcal/mol. Hydrogen bond distances in Å. The yellow arrows show the deprotonation site. (a) tyrosine (b) phenylalanine

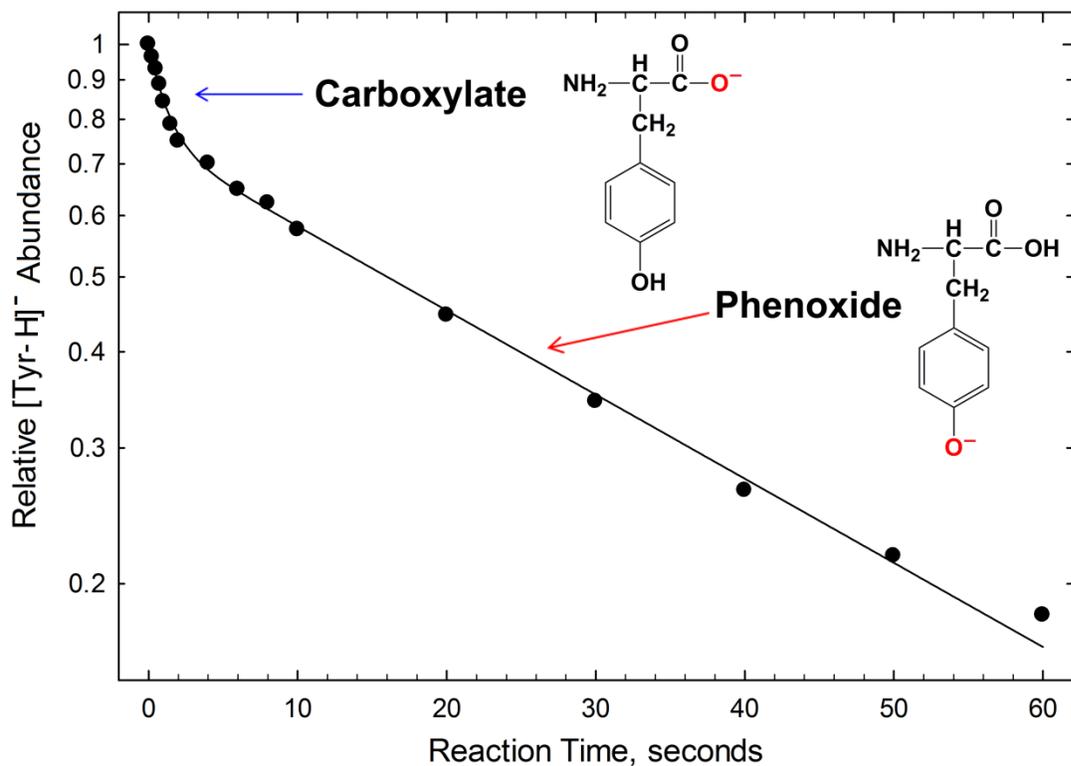


Figure 3.8. Reactant loss curve for the reaction of deprotonated tyrosine with ethyl cyanoacetate, which is present at a constant pressure of 8.9×10^{-8} mbar. The solvent is 49.5:49.5:1 (v/v/v) $\text{CH}_3\text{OH}:\text{H}_2\text{O}:\text{NH}_4\text{OH}$. The logarithm of $[\text{Tyr} - \text{H}]^-$ intensity is plotted as a function of reaction time. The experimental data points (black circles) are fitted to an equation involving the sum of two exponential decays.

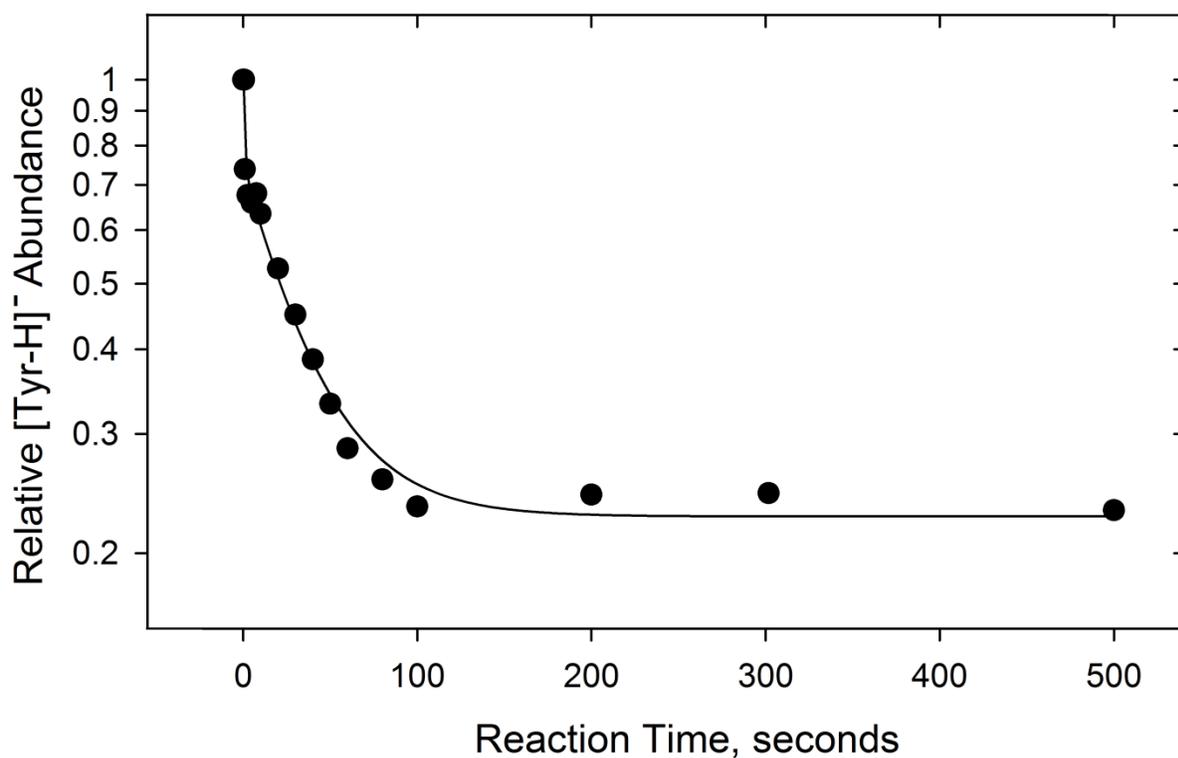


Figure 3.9. Reactant loss curve for the reaction of deprotonated tyrosine with 4-amino-2,3,5,6-tetrafluoropyridine (ATFP), which is present at a constant pressure of 5.6×10^{-8} mbar. The solvent is 99:1 (v/v) $\text{CH}_3\text{CN}:\text{H}_2\text{O}$. The logarithm of $[\text{Tyr} - \text{H}]^-$ intensity is plotted as a function of reaction time. The experimental data points (black circles) are fitted to an equation involving the sum of two exponential decays.

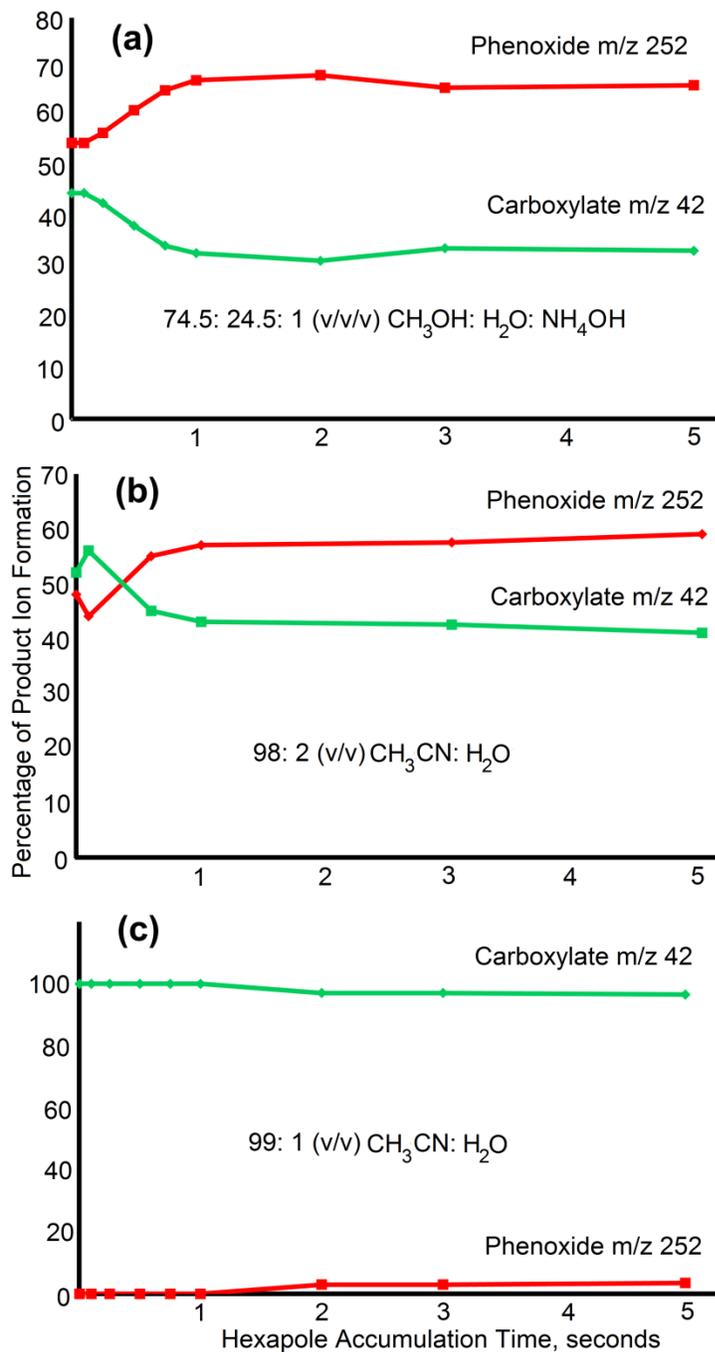


Figure 3.10. Percentage of product ions from the reaction of deprotonated tyrosine with trimethylsilyazide (TMSN₃) as a function of time that the ions accumulate in the hexapole. The presence of m/z 252 indicates a phenoxide deprotonated ion structure, while m/z 42 indicates a carboxylate structure. The solvent systems are: (a) 74.5:24.5:1 CH₃OH:H₂O:NH₄OH, (b) 98:2 CH₃CN:H₂O, and (c) 99:1 CH₃CN:H₂O.

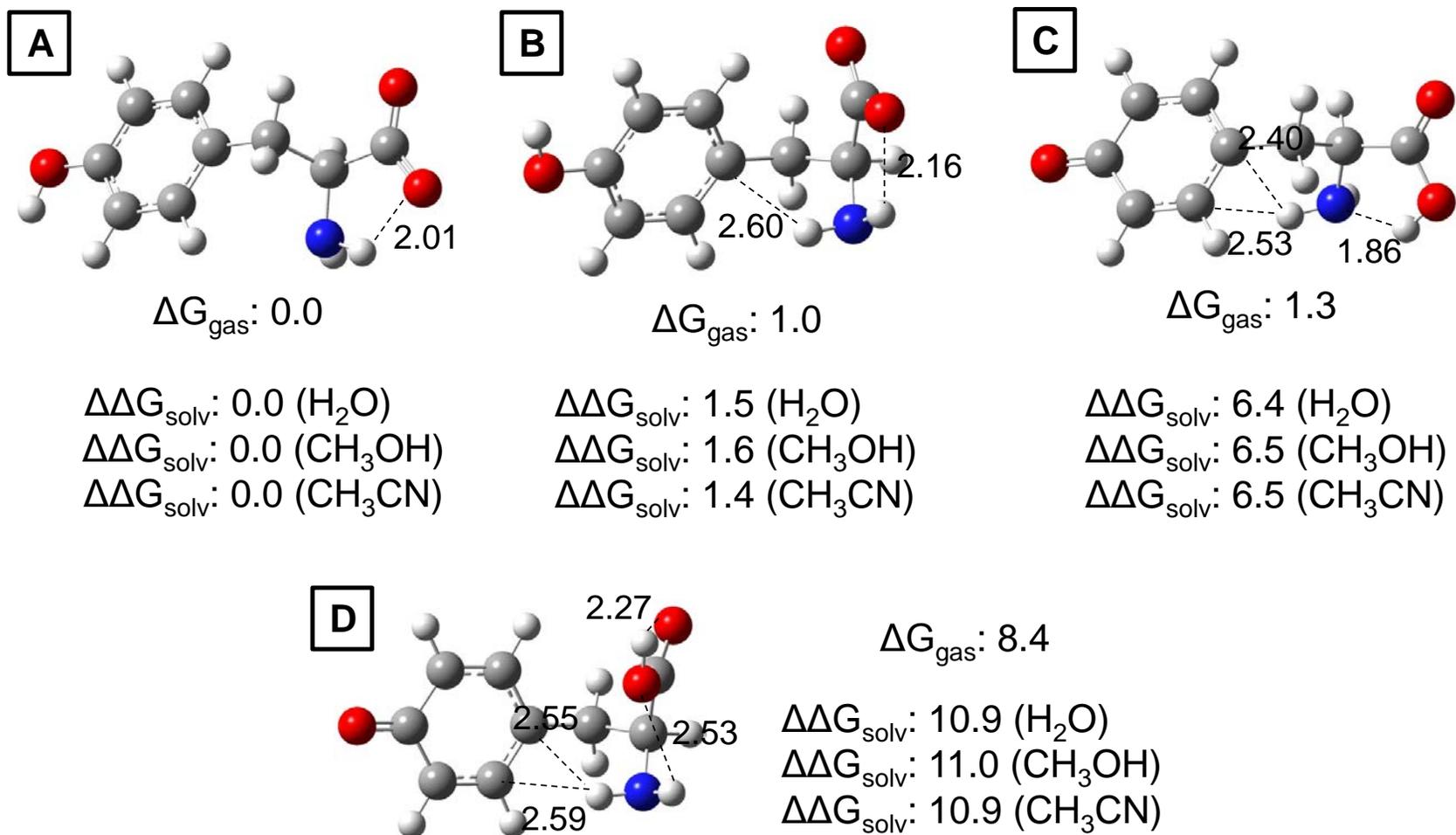
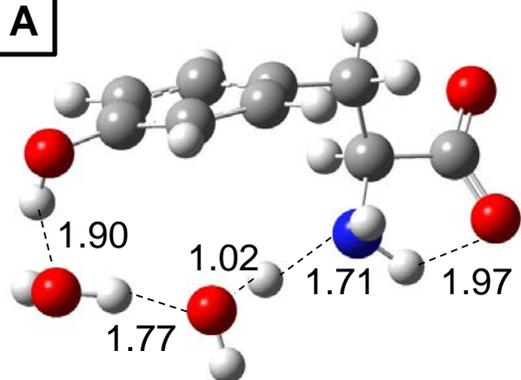


Figure 3.11. B3LYP/aug-cc-pvdz optimized tyrosine anions. Important hydrogen bonds in Å. Relative energies in the gas and solution phases are given in kcal/mol.

(a)

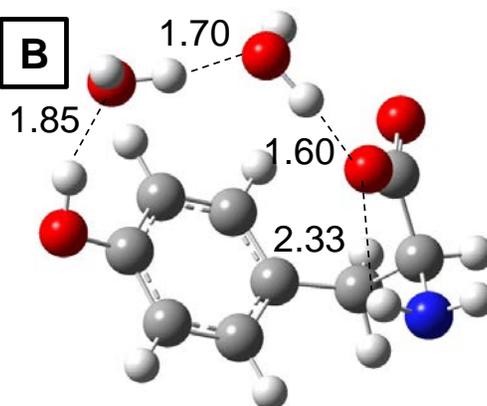
A



$$\Delta G_{\text{gas}}: 6.5$$

$$\Delta\Delta G_{\text{solv}}: 0.0 (\text{H}_2\text{O})$$

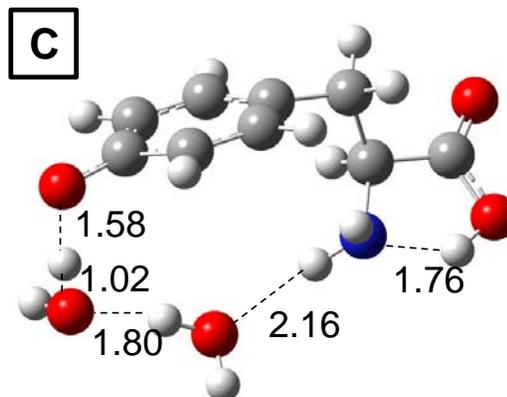
B



$$\Delta G_{\text{gas}}: 0.0$$

$$\Delta\Delta G_{\text{solv}}: 2.5 (\text{H}_2\text{O})$$

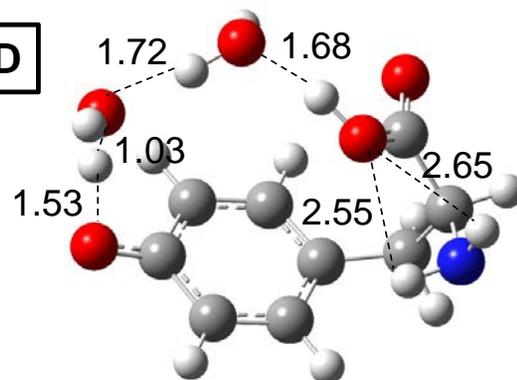
C



$$\Delta G_{\text{gas}}: 4.3$$

$$\Delta\Delta G_{\text{solv}}: 5.8 (\text{H}_2\text{O})$$

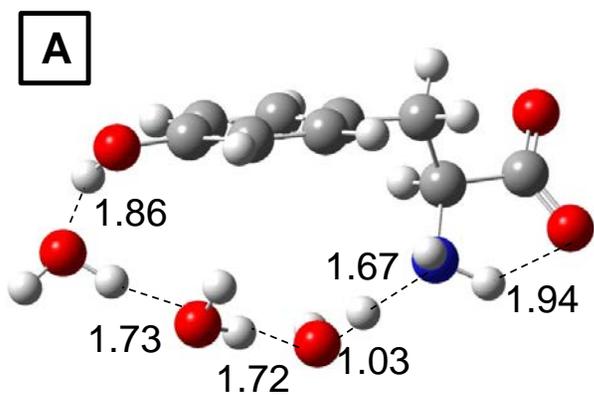
D



$$\Delta G_{\text{gas}}: 6.0$$

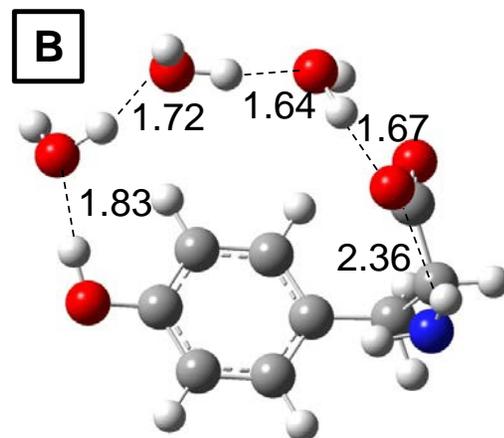
$$\Delta\Delta G_{\text{solv}}: 8.5 (\text{H}_2\text{O})$$

(b)



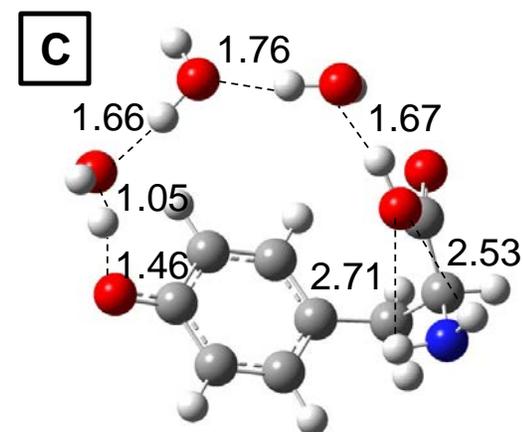
$$\Delta G_{\text{gas}}: 6.2$$

$$\Delta\Delta G_{\text{solv}}: 0.0 (\text{H}_2\text{O})$$



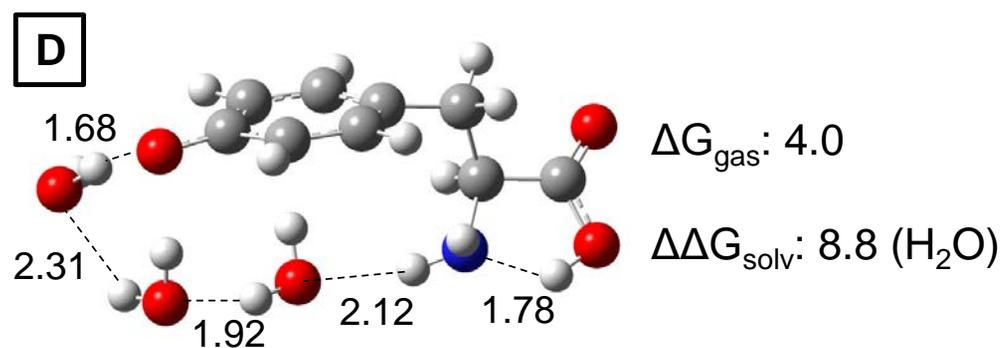
$$\Delta G_{\text{gas}}: 0.0$$

$$\Delta\Delta G_{\text{solv}}: 2.5 (\text{H}_2\text{O})$$



$$\Delta G_{\text{gas}}: 7.1$$

$$\Delta\Delta G_{\text{solv}}: 8.7 (\text{H}_2\text{O})$$



$$\Delta G_{\text{gas}}: 4.0$$

$$\Delta\Delta G_{\text{solv}}: 8.8 (\text{H}_2\text{O})$$

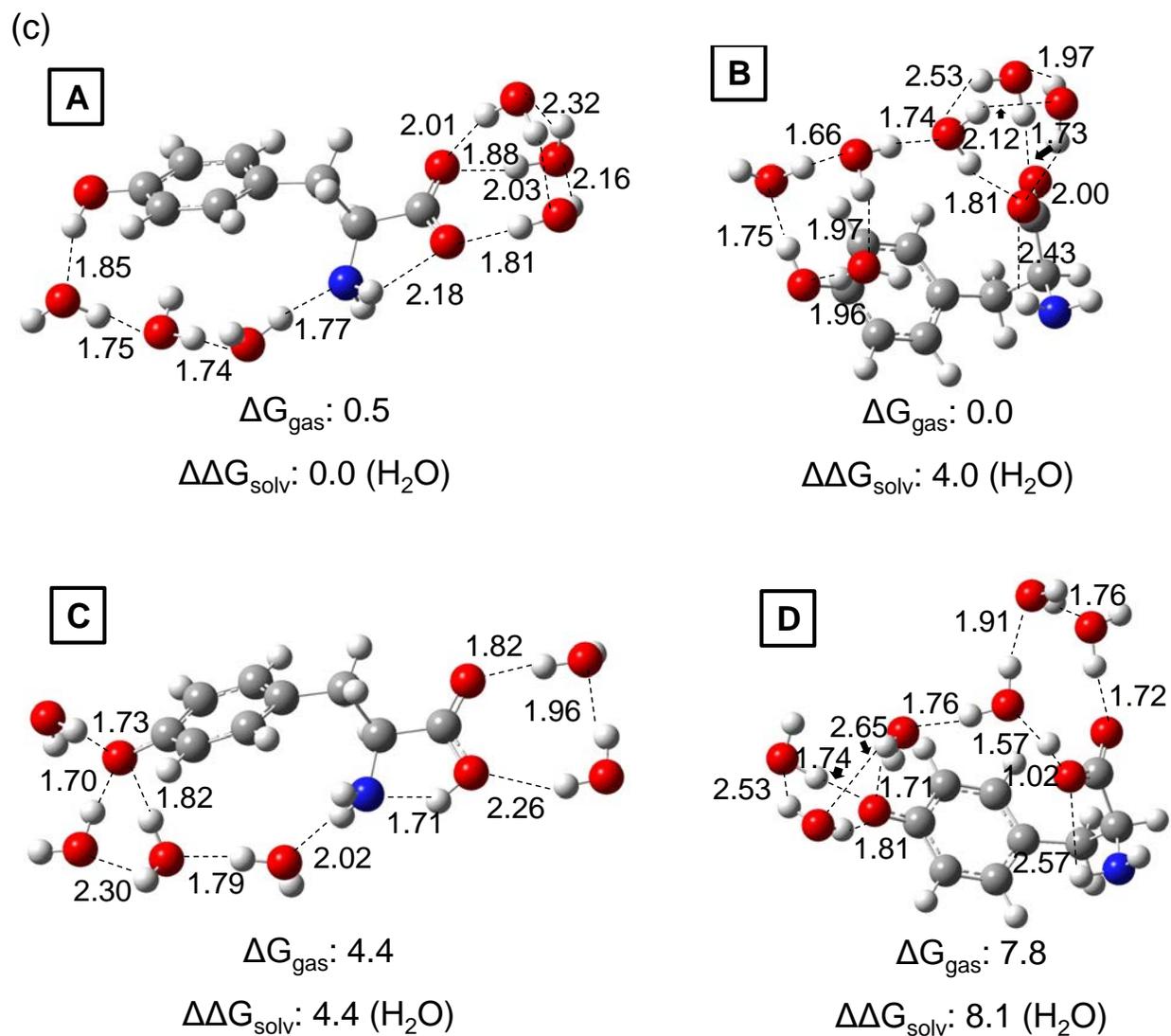


Figure 3.12. B3LYP/aug-cc-pvdz optimized tyrosine anions with explicit waters. Important hydrogen bonds in Å. Relative energies in the gas and solution phases in kcal/mol. (a) two explicit waters (b) three explicit waters (c) six explicit waters

References

- ¹ Li, Z.; Matus, M. H.; Velazquez, H. A.; Dixon, D. A.; Cassady, C. J. Gas-Phase Acidities of Aspartic Acid, Glutamic Acid, and their Amino Acid Amides. *Int. J. Mass Spectrom.* **2007**, *265*, 213-223.
- ² Tian, Z.; Pawlow, A.; Poutsma, J. C.; Kass, S. R. Are Carboxyl Groups the Most Acidic Sites in Amino Acids? Gas-Phase Acidity, H/D Exchange Experiments, and Computations on Cysteine and its Conjugate Base. *J. Am. Chem. Soc.* **2007**, *129*, 5403-5407.
- ³ Stover, M. L.; Jackson, V.; Matus, M.; Adams, M.; Cassady, C. J.; Dixon, D. A. Fundamental Thermochemical Properties of Amino Acids: Gas-Phase and Aqueous Acidities and Gas-Phase Heats of Formation. *J. Phys. Chem. B* **2012**, *116*, 2905-2916.
- ⁴ DeBlase, A. F.; Kass, S. R.; Johnson, M. A. On the Character of the Cyclic Ionic H-Bond in Cryogenically Cooled Deprotonated Cysteine. *Phys. Chem. Chem. Phys.* **2014**, *16*, 4569-4575.
- ⁵ Oomens, J.; Steill, J. D.; Redlich, B. Gas-Phase IR Spectroscopy of Deprotonated Amino Acids. *J. Am. Chem. Soc.* **2009**, *131*, 4310-4319.
- ⁶ Bokatzian-Johnson, S. S.; Stover, M. L.; Dixon, D. A.; Cassady, C. J. Gas-Phase Deprotonation of the Peptide Backbone for Tripeptides and their Methyl Esters with Hydrogen and Methyl Side Chains. *J. Phys. Chem. B* **2012**, *116*, 14844-14858.
- ⁷ Gao, J.; Cassady, C. J. Negative Ion Production from Peptides and Proteins by Matrix-Assisted Laser Desorption/ionization Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 4066-4072.
- ⁸ Tian, Z.; Kass, S. R. Does Electrospray Ionization Produce Gas-Phase Or Liquid-Phase Structures? *J. Am. Chem. Soc.* **2008**, *130*, 10842-10843.
- ⁹ Tian, Z.; Wang, X. B.; Wang, L. S.; Kass, S. R. Are Carboxyl Groups the most Acidic Sites in Amino Acids? Gas-Phase Acidities, Photoelectron Spectra, and Computations on Tyrosine, p-Hydroxybenzoic Acid, and their Conjugate Bases. *J. Am. Chem. Soc.* **2009**, *131*, 1174-1181.
- ¹⁰ Bartlett, R. J.; Musial, M. Coupled-Cluster Theory in Quantum Chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291-352.
- ¹¹ Kendall, R. A.; Dunning Jr., T. H.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796-6806.
- ¹² Li, H.; Lin, Z.; Luo, Y. Gas-Phase IR Spectroscopy of Deprotonated Amino Acids: Global or Local Minima? *Chem. Phys. Lett.* **2014**, *598*, 86-90.
- ¹³ O'Hair, R. A. J.; Bowie, J. H.; Gronert, S. Gas-Phase Acidities of the Alpha-Amino Acids. *Int. J. Mass Spectrom. Ion Proc.* **1992**, *117*, 23-36.

-
- ¹⁴ 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.
- ¹⁵ Bouchoux, G.; Salpin, J. Y. Re-Evaluated Gas Phase Basicity and Proton Affinity Data from the Thermokinetic Method. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 932-936.
- ¹⁶ de Koning, L. J.; Nibbering, N. M. M.; van Orden, S. L.; Laukien, F. H. Mass Selection of Ions in a Fourier Transform Ion Cyclotron Resonance Trap using Correlated Harmonic Excitation Fields (CHEF). *Int. J. Mass Spectrom. Ion Proc.* **1997**, *165/166*, 209-219.
- ¹⁷ Bartmess, J. E. In *Negative Ion Energetics Data*; Linstrom, P. J., Ed.; NIST Chemistry WebBook, NIST Standard Reference Database Number 69; National Institute of Standards and Technology: Gaithersburg MD, 20899, <http://webbook.nist.gov>, (retrieved June 11, 2014).
- ¹⁸ Zhang, K.; Cassady, C. J.; Chung-Phillips, A. Ab Initio Studies of Neutral and Protonated Triglycines: Comparison of Calculated and Experimental Gas-Phase Basicity. *J. Am. Chem. Soc.* **1994**, *116*, 11512-11521.
- ¹⁹ Bartmess, J. E.; Georgiadis, R. M. Empirical Methods for Determination of Ionization Gauge Relative Sensitivities for Different Gases. *Vacuum* **1983**, *33*, 149-153.
- ²⁰ Miller, K. J. Additivity Methods in Molecular Polarizability. *J. Am. Chem. Soc.* **1990**, *112*, 8533-8542.
- ²¹ Zhang, K.; Zimmerman, D. M.; Chung-Phillips, A.; Cassady, C. J. Experimental and Ab Initio Studies of the Gas-Phase Basicities of Polyglycines. *J. Am. Chem. Soc.* **1993**, *115*, 10812-10822.
- ²² Gong, S.; Camara, E.; He, F.; Green, M. K.; Lebrilla, C. B. Chiral Recognition and the Deprotonation Reaction of Gas-Phase Cytochrome c Ions. *Int. J. Mass Spectrom.* **1999**, *185-187*, 401-412.
- ²³ Jankiewicz, B. J.; Vinueza, N. R.; Kirkpatrick, L. M.; Gallardo, V. A.; Li, G.; Nash, J. J.; Kenttämaa, H. I. Does the 2,6-Didehydropyridinium Cation Exist? *J. Phys. Org. Chem.* **2013**, *26*, 707-714.
- ²⁴ He, F.; Marshall, A. G.; Freitas, M. A. Assignment of Gas-Phase Dipeptide Amide Hydrogen Exchange Rate Constants by Site-Specific Substitution: GlyGly. *J Phys Chem B* **2001**, *105*, 2244-2249.
- ²⁵ Watkins, M. A.; Nelson, E. D.; Tichy, S. E.; Kenttämaa, H. I. Rearrangement of Phenylcarbene Radical Cation to Dehydrotropylium Cation. *Int. J. Mass Spectrom.* **2006**, *249-250*, 1-7.

-
- ²⁶ Schaaff, T. G.; Stephenson, J. L.; McLuckey, S. A. Gas Phase H/D Exchange Kinetics: DI Versus D₂O. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 167-171.
- ²⁷ Gross, D. S.; Schnier, P. D.; Rodriguez-Cruz, S. E.; Fagerquist, C. K.; Williams, E. R. Conformations and Folding of Lysozyme Ions in Vacuo. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 3143-3148.
- ²⁸ Cassady, C. J.; Carr, S. R. Elucidation of Isomeric Structures for Ubiquitin [M+12H]¹²⁺ Ions Produced by Electrospray Ionization Mass Spectrometry. *J. Mass Spectrom.* **1996**, *31*, 247-254.
- ²⁹ Ewing, N. P.; Cassady, C. J. Effects of Cysteic Acid Groups on the Gas-Phase Reactivity and Dissociation of [M+4H]⁴⁺ Ions from Insulin Chain B. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 928-940.
- ³⁰ Su, T.; Chesnavich, W. J. Parametrization of the Ion-Polar Molecule Collision Rate-Constant by Trajectory Calculations. *J. Chem. Phys.* **1982**, *76*, 5183-5185.
- ³¹ Su, T. Erratum: Trajectory Calculations of Ion-Polar Molecule Capture Rate Constants at Low Temperatures. *J. Chem. Phys.* **1988**, *89*, 5355-5355.
- ³² Bouchoux, G.; Salpin, J. Y.; Leblanc, D. A Relationship Between the Kinetics and Thermochemistry of Proton Transfer Reactions in the Gas Phase. *Int. J. Mass Spectrom. Ion Proc.* **1996**, *153*, 37-48.
- ³³ 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.
- ³⁴ Bouchoux, G.; Buisson, D.; Bourcier, S.; Sablier, M. Application of the Kinetic Method to Bifunctional Bases ESI Tandem Quadrupole Experiments. *Int. J. Mass Spectrom.* **2003**, *228*, 1035-1054.
- ³⁵ Gaussian 09, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.

-
- ³⁶ Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- ³⁷ Parr, R. G.; Yang, W. In *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.
- ³⁸ Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. Optimization of Gaussian-Type Basis Sets for Local Spin Density Functional Calculations. Part I. Boron through Neon, Optimization Technique and Validation. *Can. J. Chem.* **1992**, *70*, 560-571.
- ³⁹ Gutowski, K. E.; Dixon, D. A. Ab Initio Prediction of the Gas- and Solution-Phase Acidities of Strong Bronsted Acids: The Calculation of pKa Values Less than -10. *J. Phys. Chem. A* **2006**, *110*, 12044-12054.
- ⁴⁰ Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. Gaussian-3 Theory using Reduced Moller-Plesset Order. *J. Chem. Phys.* **1999**, *110*, 4703-4709.
- ⁴¹ Dixon, D. A.; Feller, D.; Peterson, K. A. In *A Practical Guide to Reliable First Principles Computational Thermochemistry Predictions Across the Periodic Table*; Wheeler, R. A., Section Ed. Tschumper, G. S., Eds.; Annual Reports in Computational Chemistry; Elsevier: Amsterdam, 2012; Vol. 8, pp 1-28.
- ⁴² Feller, D.; Peterson, K. A.; Dixon, D. A. Further Benchmarks of a Composite, Convergent, Statistically Calibrated Coupled-Cluster-Based Approach for Thermochemical and Spectroscopic Studies. *Mol. Phys.* **2012**, *110*, 2381-2399.
- ⁴³ Peterson, K. A.; Feller, D.; Dixon, D. A. Chemical Accuracy in Ab Initio Thermochemistry and Spectroscopy: Current Strategies and Future Challenges. *Theor. Chem. Acc.* **2012**, *131*, 1-20.
- ⁴⁴ Feller, D.; Peterson, K. A.; Dixon, D. A. A Survey of Factors Contributing to Accurate Theoretical Predictions of Atomization Energies and Molecular Structures. *J. Chem. Phys.* **2008**, *129*, 204105/1-204105/32.
- ⁴⁵ Feller, D.; Dixon, D. A. Extended Benchmark Studies of Coupled Cluster Theory through Triple Excitations. *J. Chem. Phys.* **2001**, *115*, 3484-3496.
- ⁴⁶ McQuarrie, D. A. In *Statistical Mechanics*; Harper & Row: New York, USA, 1976.
- ⁴⁷ Muftakhov, M. V.; Vasil'ev, Y. V.; Mazunov, V. A. Determination of Electron Affinity of Carbonyl Radicals by Means of Negative Ion Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1104-1108.
- ⁴⁸ MOLPRO, version 2010.1, a package of *ab initio* programs, Knowles, P. J.; Manby, F. R.; Schütz, M.; Celani, P.; Knizia, G.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn,

A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M.; Wolf, A. See <http://www.molpro.net>.

⁴⁹ Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. NWChem: A Comprehensive and Scalable Open-Source Solution for Large Scale Molecular Simulations. *Comput. Phys. Commun.* **2010**, *181*, 1477-1489.

⁵⁰ Curtiss, L. A.; Raghavachari, K. Gaussian-3 (G3) Theory for Molecules Containing First and Second-Row Atoms. *J. Chem. Phys.* **1998**, *109*, 7764-7776.

⁵¹ Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-3 Theory using Density Functional Geometries and Zero-Point Energies. *J. Chem. Phys.* **1999**, *110*, 7650-7657.

⁵² Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 Theory. *J. Chem. Phys.* **2007**, *126*, 084108/1-084108/12.

⁵³ Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 Theory using Reduced Order Perturbation Theory. *J. Chem. Phys.* **2007**, *127*, 124105/1-124105/8.

⁵⁴ Cumming, J. B.; Kebarle, P. Summary of Gas Phase Acidity Measurements Involving Acids AH. Entropy Changes in Proton Transfer Reactions Involving Negative Ions. Bond Dissociation Energies D(A-H) and Electron Affinities EA(A). *Can. J. Chem.* **1978**, *56*, 1-9.

⁵⁵ Angel, L. A.; Ervin, K. M. Gas-Phase Acidities and O-H Bond Dissociation Enthalpies of Phenol, 3-Methylphenol, 2,4,6-Trimethylphenol, and Ethanoic Acid. *J Phys Chem A* **2006**, *110*, 10392-10403.

⁵⁶ Harrison, A. G.; Tu, Y. Site of Protonation of N-Alkylanilines. *Int. J. Mass Spectrom.* **2000**, *195/196*, 33-43.

⁵⁷ Harrison, A. G. In *Chemical Ionization Mass Spectrometry*; CRC Press: Boca Raton, FL, 1992.

⁵⁸ Lau, Y. K.; Kebarle, P. Substituent Effects on the Intrinsic Basicity of Benzene: Proton Affinities of Substituted Benzenes. *J. Am. Chem. Soc.* **1976**, *98*, 7452-7453.

⁵⁹ McLuckey, S. A.; Cameron, D.; Cooks, R. G. Proton Affinities from Dissociations of Proton-Bound Dimers. *J. Am. Chem. Soc.* **1981**, *103*, 1313-1317.

⁶⁰ Cheng, X.; Wu, Z.; Fenselau, C. Collision Energy Dependence of Proton-Bound Dimer Dissociation: Entropy Effects, Proton Affinities, and Intramolecular Hydrogen Bonding in Protonated Peptides. *J. Am. Chem. Soc.* **1993**, *115*, 4844-4848.

-
- ⁶¹ Hauptert, L. J.; Poutsma, J. C.; Wenthold, P. G. The Curtin-Hammett Principle in Mass Spectrometry. *Acc. Chem. Res.* **2009**, *42*, 1480-1488.
- ⁶² Fournier, F.; Afonso, C.; Fagin, A. E.; Gronert, S.; Tabet, J. C. Can Cluster Structure Affect Kinetic Method Measurements? the Curious Case of Glutamic Acid's Gas-Phase Acidity. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1887-1896.
- ⁶³ Jia, B.; Angel, L. A.; Ervin, K. M. Threshold Collision-Induced Dissociation of Hydrogen-Bonded Dimers of Carboxylic Acids. *J. Phys. Chem. A* **2008**, *112*, 1773-1782.
- ⁶⁴ Buker, H.-H.; Grutzmacher, H.-F. A Fourier Transform-Ion Cyclotron Resonance Study of Steric Effects on Proton Transfer Reactions of Polyalkyl Benzenes. *Int. J. Mass Spectrom. Ion Proc.* **1991**, *109*, 95-104.
- ⁶⁵ Attina, M.; Cacace, F. Concerning Sterical Hindrance to Deprotonation of Gaseous Polyalkylbenzenium Ions. *Int. J. Mass Spectrom. Ion Proc.* **1992**, *120*, R1-R4.
- ⁶⁶ Schroder, D.; Budesinsky, M.; Roithova, J. Deprotonation of p-Hydroxybenzoic Acid: Does Electrospray Ionization Sample Solution or Gas-Phase Structures? *J. Am. Chem. Soc.* **2012**, *134*, 15897-15905.
- ⁶⁷ Hughes, D. L.; Bergan, J. J.; Grabowski, E. J. J. Amino Acid Chemistry in Dipolar Aprotic Solvents: Dissociation Constants and Ambident Reactivity. *J. Org. Chem.* **1986**, *51*, 2579-2585.
- ⁶⁸ Cheung, E. T. Substituent Effects on the Tautomerization of Amino Acids, MS Thesis, Texas Tech University, Lubbock, TX, 1995.
- ⁶⁹ Lide, D. R., Ed.; In *CRC Handbook of Chemistry and Physics*; CRC Press: Cleveland, OH, 1992-1993.
- ⁷⁰ Tolmachev, A. V.; Udseth, H. R.; Smith, R. D. Radial Stratification of Ions as a Function of Mass to Charge Ratio in Collisional Cooling Radio Frequency Multipoles used as Ion Guides or Ion Traps. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1907-1913.
- ⁷¹ McDonnell, L. A.; Giannakopoulos, A. E.; Derrick, P. J.; Tsybin, Y. O.; Hakansson, P. A. Theoretical Investigation of the Kinetic Energy of Ions Trapped in a Radio-Frequency Hexapole Ion Trap. *Eur. J. Mass Spectrom.* **2002**, *8*, 181-189.
- ⁷² Tolmachev, A. V.; Udseth, H. R.; Smith, R. D. Modeling the Ion Density Distribution in Collisional Cooling RF Multipole Ion Guides. *Int. J. Mass Spectrom.* **2003**, *222*, 155-174.
- ⁷³ Easterling, M.L. (Bruker Daltonics FT-ICR Product Manager) **March 28, 2012**.
- ⁷⁴ Campbell, S.; Rodgers, M.; Marzluff, E.; Beauchamp, J. Deuterium Exchange Reactions as a Probe of Biomolecule Structure. Fundamental Studies of Gas Phase H/D Exchange Reactions of

Protonated Glycine Oligomers with D₂O, CD₃OD, CD₃CO₂D, and ND₃. *J. Am. Chem. Soc.* **1995**, *117*, 12840-12854.

⁷⁵ Gard, E.; Green, M. K.; Bregar, J.; Lebrilla, C. B. Gas-Phase Hydrogen/Deuterium Exchange as a Molecular Probe for the Interaction of Methanol and Protonated Peptides. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 623-631.

⁷⁶ Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999-3094.

⁷⁷ Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc. Perkin Trans. 2* **1993**, 799-805.

⁷⁸ Cammi, R.; Mennucci, B.; Tomasi, J. In *Computational Modelling of the Solvent Effects on Molecular Properties: An Overview of the Polarizable Continuum Model (PCM) Approach*; Leszczynski, J., Ed.; Computational Chemistry: Reviews of Current Trends; World Scientific: Singapore, 2003; Vol. 8, pp 1-1-79.

⁷⁹ Fujio, M.; McIver, R. T.; Taft, R. W. Effects of the Acidities of Phenols from Specific Substituent-Solvent Interactions. Inherent Substituent Parameters from Gas-Phase Acidities. *J. Am. Chem. Soc.* **1981**, *103*, 4017-4029.

⁸⁰ Caldwell, G.; Renneboog, R.; Kebarle, P. Gas-Phase Acidities of Aliphatic Carboxylic Acids, Based on Measurements of Proton Transfer Equilibria. *Can. J. Chem.* **1989**, *67*, 611-618.

⁸¹ Kim, E. H.; Bradforth, S. E.; Arnold, D. W.; Metz, R. B.; Neumark, D. M. Study of HCO₂ and DCO₂ by Negative Ion Photoelectron Spectroscopy. *J. Chem. Phys.* **1995**, *103*, 7801-7814.

⁸² Taft, R. W.; Topsom, R. D. The Nature and Analysis of Substituent Electronic Effects. *Prog. Phys. Org. Chem.* **1987**, *16*, 1-84.

⁸³ Jinfeng, C.; Topsom, R. D.; Headley, A. D.; Koppel, I.; Mishima, M.; Taft, R. W.; Veji, S. Acidities of Substituted Acetic Acids. *J. Mol. Struct. : THEOCHEM* **1988**, *168*, 141-146.

⁸⁴ Norrman, K.; McMahan, T. B. Intramolecular Solvation of Carboxylate Anions in the Gas Phase. *J. Phys. Chem. A* **1999**, *103*, 7008-7016.

⁸⁵ Richardson, J. H.; Stephenson, L. M.; Brauman, J. I. Photodetachment of Electrons from Phenoxides and Thiophenoxide. *J. Am. Chem. Soc.* **1975**, *97*, 2967-2970.

⁸⁶ Bartmess, J. E.; Scott, J. A.; McIver, R. T. Scale of Acidities in the Gas-Phase from Methanol to Phenol. *J. Am. Chem. Soc.* **1979**, *101*, 6046-6056.

⁸⁷ Gunion, R. F.; Gilles, M. K.; Polak, M. L.; Lineberger, W. C. Ultraviolet Photoelectron Spectroscopy of the Phenide, Benzyl and Phenoxide Anions, with Ab Initio Calculations. *Int. J. Mass Spectrom. Ion Proc.* **1992**, *117*, 601-620.

⁸⁸ DeTuri, V. F.; Ervin, K. M. Proton Transfer between Cl^- and $\text{C}_6\text{H}_5\text{OH}$. O-H Bond Energy of Phenol. *Int. J. Mass Spectrom. Ion Proc.* **1998**, *175*, 123-132.

⁸⁹ Angel, L. A.; Ervin, K. M. Competitive Threshold Collision-Induced Dissociation: Gas-Phase Acidity and O-H Bond Dissociation Enthalpy of Phenol. *J. Phys. Chem. A* **2004**, *108*, 8346-8352.

⁹⁰ McMahon, T. B.; Kebarle, P. Intrinsic Acidities of Substituted Phenols and Benzoic Acids Determined by Gas-Phase Proton-Transfer Equilibria. *J. Am. Chem. Soc.* **1977**, *99*, 2222-30.

**APPENDIX: AN EXPERIMENTAL AND COMPUTATIONAL INVESTIGATION INTO
THE GAS-PHASE ACIDITIES OF TYROSINE AND PHENYLALANINE: THREE
STRUCTURES FOR DEPROTONATED TYROSINE**

Additional information from the electronic structure calculations including total H_{298} and G_{298} energies from Gaussian09 for all neutrals and anions at the G3(MP2) and B3LYP/aug-cc-pvdz levels, and CCSD(T) atomization energies use for the benchmark study of the acidity of formic acid.

Table A3.1. H_{298} and G_{298} Total Energies for all Organic Acids, Alcohols, and Anions at the G3(MP2) level (a.u.)

Amino Acid	H_{298}	G_{298}
Formic acid neutral	-189.532514	-189.560700
Formic acid anion	-188.985540	-189.013210
Acetic acid neutral	-228.777179	-228.809713
Acetic acid anion	-228.224278	-228.257380
Propanoic acid neutral	-268.010445	-268.046429
Propanoic acid anion	-267.458382	-267.494628
Isobutyric acid neutral	-307.245013	-307.284326
Isobutyric acid anion	-306.694884	-306.734246
Trimethylacetic acid neutral	-346.482023	-346.523936
Trimethylacetic acid anion	-345.934111	-345.975687
Butanoic acid neutral	-307.243191	-307.282925
Butanoic acid anion	-306.692049	-306.732033
Isovaleric acid neutral	-346.479363	-346.521889
Isovaleric acid anion	-345.930481	-345.972478
Tert-butylacetic acid neutral	-385.715587	-385.760651
Tert-butylacetic acid anion	-385.168411	-385.212900
Pentanoic acid neutral	-346.476177	-346.519632
Pentanoic acid anion	-345.925471	-345.969299
Isohexanoic acid neutral	-385.711498	-385.758216
Isohexanoic acid anion	-385.161634	-385.208504
Tert-butylpropanoic acid neutral	-424.948715	-424.998117
Tert-butylpropanoic acid anion	-424.399389	-424.449750
Phenol neutral	-306.972942	-307.008706
Phenol anion	-306.418358	-306.453494
<i>o</i> -cresol neutral	-346.210874	-346.250560
<i>o</i> -cresol anion	-345.657634	-345.696524
<i>m</i> -cresol neutral	-346.210423	-346.251864
<i>m</i> -cresol anion	-345.655243	-345.695720
<i>p</i> -cresol neutral	-346.209488	-346.250320
<i>p</i> -cresol anion	-345.652918	-345.693839

Table A3.2. Components for CCSD(T) Atomization Energies in kcal/mol of Formic Acid

neutral/anion	$\Delta E_{\text{CBS}}^{\text{a}}$	$\Delta E_{\text{ZPE}}^{\text{b}}$	$\Delta E_{\text{CV}}^{\text{c}}$	$\Delta E_{\text{SR}}^{\text{d}}$	$\Delta E_{\text{SO}}^{\text{e}}$	$\Sigma D_0(0 \text{ K})^{\text{f}}$
HCO ₂ H	497.56	-20.99	1.42	-0.79	-0.52	476.67
HCO ₂ ⁻	459.65	-12.52	1.34	-0.81	-0.52	447.14

^a The atomic asymptotes were calculated with the R/UCCSD(T) method. The energies were extrapolated to the complete basis set limit using $E(n) = E_{\text{CBS}} + A \exp[-(n-1)] + B \exp[-(n-1)^2]$ with $n = 2$ (aVDZ), 3 (aVTZ), and 4 (aVQZ),

^b The zero point energies were taken as 0.5 the sum of the frequencies from calculations at the MP2/aug-cc-pVTZ level.

^c Core-valence corrections were obtained at the CCSD(T)/aug-cc-pwCVTZ level.

^d Scalar relativistic correction was evaluated from the expectation values for the two dominant terms in the Breit-Pauli Hamiltonian (the mass-velocity and one-electron Darwin (MVD) corrections) from configuration interaction singles and doubles (CISD) calculations with a aug-cc-pVTZ basis set.

^e Correction due to the incorrect treatment of the atomic asymptotes as an average of spin multiplets. Values are from Moore, C. E. *Atomic energy levels as derived from the analysis of optical spectra, Volume 1, H to V*; U.S. National Bureau of Standards Circular 467, COM-72-50282; U.S. Department of Commerce, National Technical Information Service: Washington, D.C., 1949.

^f Total atomization energies.

Table A3.3. H_{298} and G_{298} Total Energies for Phenylalanine, Tyrosine, HPB, and Anions at the G3(MP2) level (a.u.)

Amino Acid	H_{298}	G_{298}
Phenylalanine (Phe)	-553.935927	-553.985535
Folded carboxylate anion	-553.398896	-553.448532
Unfolded carboxylate anion	-553.397057	-553.447963
Alpha carbon deprotonated anion	-553.355803	-553.406006
Amine deprotonated anion	-553.325721	-553.375661
Aromatic ring deprotonated anion	-553.316307	-553.366006
Tyrosine (Tyr)	-629.084721	-629.137110
Folded carboxylate anion	-628.547995	-628.600502
Unfolded carboxylate anion	-628.544951	-628.599001
Unfolded phenoxide anion	-628.544388	-628.596587
Alpha carbon deprotonated anion	-628.503416	-628.556540
Amine deprotonated anion	-628.478709	-628.531628
Aromatic ring deprotonated anion	-628.478211	-628.531330
4-(4-hydroxyphenyl)-2-butanone (HPB)	-537.868775	-537.923205
Folded phenoxide anion	-537.322868	-537.371202
Unfolded phenoxide anion	-537.31738	-537.371202

Table A3.4. H_{298} and G_{298} Total Energies for Solvated Tyrosine Anions at the B3LYP/aug-cc-pvdz level (a.u.)

# of waters	Anion	H_{298}	G_{298}
0	Folded Carboxylate	-629.373294	-629.425478
0	Unfolded Carboxylate	-629.373448	-629.427004
0	Unfolded Phenoxide	-629.373232	-629.424937
0	Folded Phenoxide	-629.361213	-629.413565
2	Folded Carboxylate	-782.255190	-782.318501
2	Unfolded Carboxylate	-782.244668	-782.308990
2	Unfolded Phenoxide	-782.246337	-782.311604
2	Folded Phenoxide	-782.246097	-782.308151
3	Folded Carboxylate	-858.692494	-858.763522
3	Unfolded Carboxylate	-858.683171	-858.753582
3	Unfolded Phenoxide	-858.686674	-858.757140
3	Folded Phenoxide	-858.683059	-858.752243
6	Folded Carboxylate	-1087.998516	-1088.087044
6	Unfolded Carboxylate	-1087.994726	-1088.086318
6	Unfolded Phenoxide	-1087.981879	-1088.079985
6	Folded Phenoxide	-1087.983021	-1088.074549

CHAPTER 4: AN EXPERIMENTAL AND COMPUTATIONAL STUDY OF THE GAS-PHASE ACIDITIES OF THE COMMON AMINO ACID AMIDES

4.1 Introduction Amino acid amides contain an amide group in the place of the carboxylic acid functionality located at the C-terminus of amino acids. For amino acids, the carboxylic acid is generally the most acidic site on the molecule.¹ However, an amino acid amide should provide a better representation of the energetic values associated with amino acid residues incorporated into a peptide chain. As shown in Figure 1, after the formation of a peptide bond, the carboxylic acid group is no longer available at an internal residue of a peptide. When studying deprotonated peptides or proteins, unless an amino acid has an acidic side chain, the site most readily available for deprotonation is the C-terminal carboxylic acid group.^{2,3} Because the majority of amino acid residues are not located at the termini of peptides, understanding the acidity of amino acid amides gives insight when evaluating potential sites of deprotonation for internal residues.

The acidities of amino acid residues are important to the field of proteomics. Gas-phase acidity (GA) and gas-phase basicity (GB) values can assist in explaining many attributes of gas-phase behavior, including conformation, reactivity, and fragmentation.^{4,5} The GA, or ΔG_{acid} , of a molecule is the Gibbs free energy change (ΔG) of the reaction: $\text{AH} \rightarrow \text{A}^- + \text{H}^+$. Although tandem mass spectrometry (MS/MS) techniques for sequencing proteins and peptides are often conducted in the positive ion mode, the negative ion mode can provide either greater or complementary information on sequence.^{6,7,8,9,10,11,12} Acidic peptides often deprotonate more easily than they protonate,^{6,13,14,15} and the negative mode is better suited for sequencing acidic peptides. Since peptide fragmentation pathways are usually charge-directed,^{4,16} amino acid

amides can provide valuable acidity and deprotonation site information to assist in understanding the fragmentation patterns in negative ion mode MS/MS techniques for peptide and protein sequencing.

Several of the twenty common amino acids deprotonate readily at their side chains in the gas phase. Glutamic acid and aspartic acid contain side chain carboxylic acid groups that deprotonate and these acidic groups are also present in their amino acid amide forms.² The deprotonated structure of tyrosine has been studied extensively experimentally and computationally by us,¹⁷ as well as by the Kass^{18,19,20} and Oomens²¹ groups. Tyrosine deprotonates at both the C-terminal carboxylic acid group and the phenolic side chain. Cysteine has been investigated theoretically and experimentally and has been found to form a thiolate ion by side chain deprotonation with a strong hydrogen bond to it from the carboxylate group.^{1,21,22} In addition, Ren and coworkers have observed cysteine side chain deprotonation in experimental and computational GA studies of small cysteine-containing peptide amides.^{23,24,25,26} These studies indicate that residue side chains may deprotonate in the gas phase even in the absence of a side chain carboxylic acid group. In addition, amide nitrogens along a peptide backbone can deprotonate in the absence of acidic side chains; we have demonstrated this in an experimental and computational study of the GAs for glycine- and alanine-containing tripeptide methyl esters.³

Additional evidence for the deprotonation of amino acid side chains is found in MS/MS experiments involving collision-induced dissociation (CID) on deprotonated peptide ions. For example, residues with hydroxyl side chains (i.e., serine, threonine, and tyrosine) display characteristic neutral losses when deprotonated peptides are subjected to negative ion mode CID.^{6,10,27,28,29,30,31} Proposed mechanisms for these processes involve side chain hydroxyl groups

that are deprotonated. Aspartic acid^{10,12} and glutamic acid³² residues induce specific peptide backbone cleavages by negative ion mode CID that suggest deprotonation at their side chains. Another example of side chain deprotonation was observed by Grzetic and Oomens via IRMPD,³³ as well as Bowie and coworkers by CID on asparagine-containing peptides.^{10,34} Both groups proposed a succinamide structure that deprotonates at the asparagine side chain and allows charge-directed backbone cleavage.

Although the GAs of amino acids has been the focus of several studies,^{1,2,17,35,36} the only previous report of GAs for amino acid amides is our work with glutamic acid amide and aspartic acid amide.² In this prior report, our experimental and computational G3(MP2) level GA values were in excellent agreement and the calculations confirmed that both amides undergo side chain deprotonation at their carboxylic acid groups. In the current study, we expand our work to include the eighteen other common amino acid amides whose GAs and lowest energy deprotonation sites are currently unknown.

4.2 Experimental and Computational Methods *Mass Spectrometry* All experiments employed a Bruker Daltonics (Billerica, MA, USA) BioApex 7T Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer. The amino acids were in the L-stereoisomer. All solutions were at 60 μ M in a solvent of volume ratio of 49.5:49.5:1 CH₃OH:H₂O:NH₄OH. The 1% ammonium hydroxide was added to facilitate deprotonation of the amides in the solution phase prior to electrospray ionization (ESI). For glycine amide and alanine amide, addition experiments were performed using 1% sodium hydroxide. In the case of cysteine amide, detritylation was performed,³⁷ as well as reduction of disulfide linkage.³⁸ Amino acid amide solutions were introduced into an Apollo API source using a syringe pump set to deliver ~115 μ L/hr. Solutions were ionized by ESI, which utilized air as the drying gas. The ESI needle was

grounded, and the capillary entrance and end plate were at a potential of 3.5-4.0 kV for optimal negative ion formation.

Deprotonated amide ions were isolated by correlated frequency resonance ejection techniques³⁹ and reacted with a series of neutral reference compounds with known GAs. Autoionization of the anions was not observed in the experiment or in the computations. Reference compounds were introduced through a leak valve at constant pressures, which were in the range of $(1.0-11) \times 10^{-8}$ mbar. Ion/molecule reaction times in the FT-ICR cell were varied from 0-300 seconds. Pressures were measured by a calibrated ion gauge,^{17,40} and the pressure of each reference compound was corrected for its ionization efficiency,⁴¹ which is determined by polarizabilities calculated with atomic hybrid parameter procedures.⁴²

The pseudo-first-order decay of precursor ion intensity as a function of reaction time was utilized to obtain experimental rate constants from which GA values were assigned. For experiments involving non-linear pseudo-first order kinetic behavior resulting in bimodal plots, data were fitted to the sum of two exponential decays using SigmaPlot by Systat Software Inc. (San Jose, CA, USA).^{17,40} The fraction that each exponential contributed to the fit directly related to the relative abundance of that ion structure. To assign GA values, the ratio of the experimental rate constant to the thermal rate capture rate constant^{43,44} provided a reaction efficiency (RE). A RE of 0.269 was used as a break point, where a reaction is considered to become exoergic and a GA value was assigned. This selection of 0.269 comes from the work of Bouchoux and coworkers,^{45,46,47,48} which is known as the “thermokinetic method.” In our past work, the thermokinetic method of obtaining GA values has yielded excellent agreement for experimental and computational GAs for small biomolecules.^{2,3,17}

Computational Methods The calculations were performed at the density functional theory (DFT) and correlated molecular orbital (MO) theory levels with the program Gaussian-09.⁴⁹ A range of structures were initially optimized at the DFT level with the B3LYP exchange-correlation functional^{50, 51} and the DZVP2 basis set⁵² to determine the most stable conformers of the neutrals and anions by sampling many conformations with and without hydrogen bonds. Vibrational frequencies were calculated to show that the structures were minima. A substantial number of low energy conformers were found for the neutrals and anions as discussed below. Our previous predictions of the GAs of amino acids^{1,2,17} and organic acids,⁵³ at the correlated G3(MP2) molecular orbital theory level⁵⁴ are in agreement with the experimental values to within about ± 1 kcal/mol and with higher level CCSD(T) calculations extrapolated to the complete basis set limit with additional corrections^{55,56,57,58,59} at optimized MP2/aug-cc-pVTZ geometries and with CCSD(T)/aug-cc-pVTZ calculations at the same MP2 geometries for larger molecules. G3(MP2) has an additional advantage over DFT methods in terms of reliable predictions for these types of compounds because of the important role that hydrogen bonding plays in controlling the lowest energy structures; the correlated MO methods in G3(MP2) perform better in the prediction of hydrogen bond energies as well as steric non-bonded interactions than do most widely-used DFT exchange-correlation functionals. Thermal corrections to the enthalpies and the free energies were calculated in the harmonic oscillator, rigid rotor approximation⁶⁰ using the geometries and scaled frequencies obtained at the Hartree-Fock/6-31G* level of theory from the G3(MP2) calculations.

4.3 Results and Discussion The experimental and G3(MP2) GA values obtained in this study are summarized in Table 1. For purposes of comparison and completeness, Table 1 also includes our previously published GAs for aspartic acid amide and glutamic acid amide. (Note that a

lower GA numerical value means a more acidic compound.) There is excellent agreement between the experimental and theoretical GAs. The values agree to within experiment error except for arginine amide, where the lowest energy calculated GA is only 0.3 kcal/mol below the experimental error range. As discussed below, Table 1 provides the G3(MP2) GA for the formation of the most stable anion (most acidic). In addition, G3(MP2) GAs are provided for higher energy conformers and/or isomers that have a slightly higher energy GA (less acidic). For some amino acid amides, the experimental data also supports the presence of two deprotonated ion conformers and the assignment of two GAs. The reaction efficiencies for each amino acid amide reacting with the appropriate references compounds (from which the experimental GAs were assigned) are given in the appendix.

Amino Acid Amides Containing Aliphatic Side Chains The least acidic amino acid amides have aliphatic side chains that are non-polar and contain only carbon and hydrogen atoms. The aliphatic amino acid amides are those derived from glycine, alanine, proline, valine, leucine, isoleucine, and phenylalanine. For these amino acid amides, only C-terminal amide nitrogen deprotonation is possible because there is no heteroatom on the side chain. The glycine and alanine amides could not be studied experimentally because they did not produce an adequate deprotonated ion signal by ESI. Attempts to improve the anionic signal by addition of sodium hydroxide to the solution undergoing ESI were unsuccessful. Apparently, the ESI environment was insufficient to deprotonate glycine amide and alanine amide, which were the smallest and among the least acidic compounds involved in this study (G3(MP2) GAs of 351.5 kcal/mol and 351.1 kcal/mol, respectively).

The least acidic amino acid amide that could be studied experimentally was proline amide, which has a cyclic aliphatic side chain. Proline amide yielded experimental data showing

two ion populations reacting at two different rates. Both experiment and theory provided two GA values that differed by 5-7 kcal/mol. The lowest energy experimental GA of 350.2 ± 2.5 kcal/mol has a G3(MP2) counterpart of 350.8 kcal/mol, and the higher energy experimental GA of 357.7 ± 4.0 kcal/mol correlates with a G3(MP2) value of 355.4 kcal/mol. Figure 2(a) provides an example of the bimodal experimental data from the reaction of deprotonated proline amide with the reference compound 1,2,4,5-tetrafluorobenzene, whose GA of $353.3 \text{ kcal/mol}^{61}$ is between the two GA values for proline amide, thus allowing the two ion populations to be distinguished. Figure 2(b) shows experimental data for 3-methylpyrazole (GA = 348.3 kcal/mol⁶¹), which is a more acidic reference compound than either proline amide population. In Figure 2(b), both proline amide ion populations react readily and they cannot be distinguished experimentally. For proline amide, the majority of the ions ($65 \pm 10 \%$) are involved in the more acidic and lower energy process.

Structures for the most stable neutral and two anions for the aliphatic side chain amino acid amides are shown in Figure 3. Important hydrogen bond distances are shown in Å. The two anion structures are designated as either mc-cis or mc-trans, main chain deprotonation of the C-terminal amide nitrogen with cis- or trans-like orientation of the N-H bond in the $[-C(=O)NH]^-$ group, with mc-cis being the most acidic structure. The neutral structure for proline amide, Figure 3(f), is stabilized by a $O \cdots H(N)$ hydrogen bond within the $-C(=O)NH_2$ group (2.53 Å) and by a hydrogen bond between the $-C(=O)NH_2$ and the $-NH-$ in the pyrrole ring (2.17 Å). The calculations predict that there are two C-terminal deprotonated conformers with different orientations of hydrogen bonds. In the lowest energy cis-like anion, the loss of the hydrogen from the $-C(=O)NH_2$ terminal group results in a shorter internal $O \cdots H(N)$ hydrogen bond (2.43 Å) in the $[-C(=O)NH]^-$ group. In addition, the NH on the ring reorients to form a

bond with the $[-C(=O)NH]^-$ group, $N(H)\cdots(N)CO$ (2.07 Å). In the higher energy trans-like anion, the rotation of the H away from the O results in only one strong hydrogen bond between the pyrrole NH and the $[-C(=O)NH]^-$ group, $N(H)\cdots(O)CN$ (1.95 Å).

The other aliphatic side chain amides (valine, leucine, isoleucine, and phenylalanine amides) also yielded kinetic data indicating two ion populations reacting at two different rates. The two GA values for each amide differ by 4-6 kcal/mol. For the valine, leucine, and isoleucine amides, the lower energy (more acidic) GA is 348-350 kcal/mol and the higher energy GA is 352-355 kcal/mol. The two GAs for the aromatic phenylalanine amide are comparable to those for the non-aromatic aliphatic amides.

Both experiment and theory show that as the size of the aliphatic side chain increases, the amide becomes slightly more acidic. The lowest energy (more acidic) process was always observed experimentally in the greatest abundance, generally accounting for ~65 % of the ion population produced by ESI. The lowest energy anionic conformer for the aliphatic side chains and phenylalanine amide have a cis-like orientation of the $[-C(=O)NH]^-$ group with a $O\cdots H(N)$ hydrogen bond with distances of ~2.4 Å. The higher energy conformer is generated by rotating the H away from the O and breaking this hydrogen bond. This rotation results in a trans-like orientation of the $[-C(=O)NH]^-$ group and the GAs of the higher energy conformer are 4-6 kcal/mol higher (less acidic) than the cis-like orientation of the lowest energy conformer.

Amino Acid Amides Containing Amide, Basic, Hydroxyl, and Thioester Side Chains The amino acid amides discussed in this section have side chains that contain heteroatoms capable of deprotonation, but only deprotonation of the main chain C-terminal amide group was predicted and observed. The experimental and computational results for these amides are comparable to the results for the aliphatic amides. Again, two ion populations were found to react at two rates,

providing GA values that differ by 5-6 kcal/mol, with the lowest energy most acidic population accounting for ~65 % of the ions formed by ESI. The amino acid amides in this category have side chains that are basic (lysine), hydroxyl (serine and threonine), thioester (methionine), and amide (asparagine and glutamine). We discuss arginine separately at the end of this section.

For these amino acid amides, the calculations predict the presence of both the cis-like (more hydrogen bonded, lower energy) and trans-like (less hydrogen bonded, higher energy) structures. Because these compounds contain side chain heteroatoms that are capable of hydrogen bonding, they can fold into more stabilized structures. Consequently, these amino acid amides are more acidic than the amides with aliphatic side chains. As seen in Table 1, for these amino acid amides, the GA relating to the lowest energy anion is in the range of 340-348 kcal/mol, while the GA for the higher energy process ranges from 346-352 kcal/mol.

The most stable neutrals and two differing energy anions are shown in Figure 4 for these amino acid amides. The presence of an accessible oxygen or nitrogen on the side chain that is available for hydrogen bonding increases the acidity of the amides by ~8 kcal/mol. Serine, threonine, asparagine, and glutamine were determined to have nearly equal GA values, varying only in a 1.9 kcal/mol range computationally. Asparagine and glutamine amides were found to be slightly more acidic than the hydroxyl-containing serine and threonine amides. The $-C(=O)NH_2$ group on the side chains of asparagine and glutamine amides allow for a more compact, folded structure that results from 4 to 5 hydrogen bonds in the neutrals and anions. Serine and threonine amides have relatively short side chains that contain a hydroxyl group that readily participates in hydrogen bonding. The most stable neutral results from four hydrogen bonds whereas the two anions have only three. For methionine amide, which contains a thioester functionality, the side chain length is similar in size to the aliphatic amides leucine and

isoleucine. However, the neutral methionine amide folds more readily upon itself, increasing the acidity. Lysine amide contains a longer aliphatic chain (four methylenes) with an amino $-NH_2$ group, while methionine has a shorter aliphatic chain (two methylenes) with an $-SH$ group. This combination leads to a GA for lysine amide that is 2.3 kcal/mol less acidic than methionine amide. For lysine amide, the lowest energy neutral and anion structures contain very similar hydrogen bonding networks with three hydrogen bonds between the side chain $-NH_2$, the N-terminal $-NH_2$, and the C-terminal amide group.

For arginine amide, even though the two low energy conformers of the neutral differ by the arrangements of the hydrogens on the guanidine functional group, they are essentially isoenergetic. The two stable neutral conformers and three anions of arginine amide are shown in Figure 4(g). The lowest energy structure for the anion is the result of deprotonation from the C-terminal amide group as a cis-like conformer with a predicted GA of 339.4 kcal/mol. Deprotonation from the side chain guanidine functional group is only 1.4 kcal/mol higher and both values are within the bounds of the lowest energy experimental GA of 340.7 ± 3.0 kcal/mol. Thus, both of these low energy predicted structures are within the experimental error bar. The higher energy experimental GA is 347.4 ± 3.2 kcal/mol and is predicted by theory to correspond to a trans-like amide deprotonated structure.

Amino Acid Amides with Acidic Side Chains Six amino acid amides have side chains that could be sufficiently acidic to deprotonate in the gas phase. These side chains have carboxylic acid groups (aspartic and glutamic acids), a phenolic group (tyrosine), a thiol group (cysteine), and cyclic nitrogen-containing groups (histidine and tryptophan). Proton transfer reactions involving these six amino acid amides gave linear pseudo-first order kinetic behavior, indicating the

presence of one predominant deprotonated ion structure. The experiments also leave open the possibility of two or more structures with near identical GAs.

The most acidic species are glutamic acid amide and aspartic acid amide, with calculated GAs of 326.4 kcal/mol and 325.9 kcal/mol, respectively, which we have previously studied experimentally and computationally and found to undergo exclusively side chain deprotonation.²

The data clearly show that deprotonation of the –OH group in tyrosine amide and of the –SH group in cysteine amide occurs preferentially over deprotonation of the C-terminal amide functionality. This is consistent with the fact that deprotonation of the –OH group in tyrosine amino acid is slightly less favored than deprotonation of the C-terminal –CO₂H and that deprotonation of –SH occurs in the amino acid cysteine.^{1,17,18,19,20,21,22} Tyrosine amide deprotonation has an experimental GA of 336.4 ± 2.7 kcal/mol, which is in excellent agreement with the calculated GA of 336.5 kcal/mol involving an ion deprotonated at the side chain. G3(MP2) calculations found that deprotonation of tyrosine amide at the C-terminal amide group results in a GA of 347.4 kcal/mol which is ~11 kcal/mol higher in energy. This higher energy deprotonated ion was not observed experimentally. For side chain deprotonation of cysteine amide, the experimental GA is 335.6 ± 3.8 kcal/mol and the G3(MP2) GA is 332.9 kcal/mol. The calculations predicted that C-terminal amide deprotonation results in a GA of 342.6 kcal/mol, which is ~10 kcal/mol more energetic than side chain deprotonation and again was not observed experimentally. As a consequence of their acidic side chains, these compounds are several kcal/mol more acidic than the amino acid amides discussed earlier.

The most stable neutral conformers and the two anion isomers of the amino acid amides with acidic side chains are shown in Figure 5. For tyrosine amide, the most stable neutral as well as the higher energy anions have extensively folded and hydrogen bonded structures which allow

the amide and NH₂ groups (where the NH has a positive charge) to interact with the π cloud on the benzene ring. These cation- π type interactions are well established in the literature for compounds containing benzene rings.⁶² The lowest energy phenoxide anion prefers an unfolded structure with no interactions with the benzene π cloud. However, a folded structure for the phenoxide anion where a NH proton interacts with the benzene π cloud was found to only be 0.4 kcal/mol higher in energy. For cysteine amide, both the lowest energy neutral and anion are folded structures, with the neutral containing four hydrogen bonds while the lowest energy anion only contains three. As noted previously, this ability to make a compact, folded structure may help to stabilize the negative charge as it does in cysteine amide, but clearly is not important in tyrosine amide.

Tryptophan and histidine amides display one experimental deprotonated ion population, suggesting side chain deprotonation. The most stable neutral of tryptophan amide is extensively folded with multiple hydrogen bonds between the ring and the C-terminal amide and N-terminal -NH₂ groups. The lowest energy structure for tryptophan amide is predicted to be deprotonated at the nitrogen of the indole side chain and is 12.1 kcal/mol more stable than the structure involving C-terminal amide deprotonation. The lowest energy anion maintains the extensive hydrogen bonding network of the neutral. Deprotonation of the C-terminal amide group leads to a higher energy structure with far less hydrogen bonding.

Histidine amide is an interesting case because the side chain is basic and protonates readily in positive ion mode mass spectrometry.^{63,64,65,66} In the amino acid, the τ tautomer is more stable than the π tautomer with $\Delta G(298\text{ K}) = 1\text{ kcal/mol}$.¹ We predict the τ tautomer neutral amide to be more stable than the π tautomer neutral amide by 2.5 kcal/mol. Side chain deprotonation of histidine amide has an experimental GA is $330.8 \pm 4.8\text{ kcal/mol}$. For both of

the histidine amide tautomers π and τ , deprotonation of the $-\text{NH}-$ in the imidazole ring generated the same lowest energy anion with three hydrogen bonds. The GA with respect to the less stable π tautomer neutral is 328.8 kcal/mol and the GA with respect to the more stable τ tautomer neutral is 331.3 kcal/mol. Both of the calculated GA values are within the experimental error bar. The higher energy anion resulting from deprotonation of the C-terminal amide in the π tautomer leads to a GA that is 4.4 kcal/mol less acidic than the lowest energy anion. The anion derived by deprotonation of the C-terminal amide in the τ tautomer is 20 kcal/mol higher in energy than the most stable anion. These results differ from those of deprotonation of the amino acids tryptophan and histidine which deprotonate at the carboxylic acid group.^{1,21,35}

4.4 Conclusions Gas-phase acidities for the twenty common amino acid amides have been obtained for the first time. Six of the amides (aspartic acid, glutamic acid, tyrosine, cysteine, tryptophan, and histidine amides) undergo the lowest energy deprotonation at their side chains. This suggests that the side chains of these six amino acid residues can also readily deprotonate during negative ion mode mass spectrometry analysis of peptides and proteins and may influence charge directed fragmentation pathways during MS/MS experiments. The remaining fourteen common amino acid amides were found experimentally and computationally to each have two major ion populations, with corresponding GAs that differ by 4-7 kcal/mol. G3(MP2) molecular orbital calculations indicate that the two populations are conformers that both deprotonate at the C-terminal amide nitrogen. The lowest energy structures are predicted to have a cis-like conformation of the $[-\text{C}(=\text{O})\text{NH}]^-$ group whereas the higher energy structures have a trans-like conformation. This is consistent with our previous work on the GAs and gas-phase structures of tripeptide methyl esters which found that amide nitrogens along a peptide backbone can readily deprotonate during negative ion mode electrospray ionization mass spectrometry.³

Acknowledgement This research 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, and CHE-1308348. Samantha S. Bokatzian thanks the University of Alabama's National Alumni Association for a License Tag Fellowship. Michele L. Stover and Chelsea E. Plummer thank the U.S. Department of Education for a Graduate Assistance in Areas of National Need fellowship (DoEd-GAANN grant P200A100190). Michele L. Stover also thanks the Howard Hughes Medical Institute (HHMI) for a teaching fellowship through the Precollege and Undergraduate Science Education Program. David A. Dixon thanks the Robert Ramsay Fund from the University of Alabama for partial support.

Table 4.1. Experimental and Computational GAs of the Amino Acid Amides in kcal/mol.

Amino Acid Amide	Lowest Energy Calc. GA ^a	Lowest Energy Expt. GA	Higher Energy Calc. GA ^f	Higher Energy Expt. GA	Lowest Energy Calc. ΔH	Higher Energy Calc. ΔH
Alanine	351.1	^e	355.2	^e	358.8	362.5
Arginine	339.4	342.9 \pm 3.2	344.6 ^g	347.4 \pm 3.2	347.3	352.2
Asparagine	340.2	340.7 \pm 3.0	346.1	346.5 \pm 2.7	348.0	353.5
Aspartic acid	325.9 ^{b,c}	326.5 \pm 3.6 ^b	331.6 ^h	ⁿ	332.9	338.2
Cysteine	332.9 ^c	335.6 \pm 3.8	342.6 ⁱ	ⁿ	340.1	350.1
Glutamine	340.2	340.8 \pm 3.0	344.7	346.6 \pm 2.7	345.8	349.6
Glutamic acid	326.4 ^{b,c}	328.7 \pm 4.8 ^b	332.0 ^j	ⁿ	334.0	339.1
Glycine	351.5	^e	355.8	^e	359.7	363.6
Histidine	331.3 ^c	330.8 \pm 4.8	335.7 ^{a,k}	ⁿ	338.8	342.2
Isoleucine	348.3	347.9 \pm 2.6	352.7	352.1 \pm 3.2	355.1	359.2
Leucine	349.5	347.8 \pm 2.6	353.8	352.7 \pm 3.2	357.3	361.1
Lysine	347.5	348.3 \pm 2.6	352.5	352.1 \pm 3.2	355.8	358.8
Methionine	345.2	345.0 \pm 2.7	349.6	347.7 \pm 2.5	352.4	357.2
Phenylalanine	347.8	346.5 \pm 2.7	353.0	350.1 \pm 3.0	354.6	360.5
Proline	350.8	350.2 \pm 2.5	355.4	357.7 \pm 4.0	357.8	362.2
Serine	342.1	343.3 \pm 3.2	347.8	347.6 \pm 3.2	349.3	355.0
Threonine	341.7	342.8 \pm 3.2	347.4	347.4 \pm 3.2	349.2	354.7
Tryptophan	336.1 ^c	335.8 \pm 3.8	348.2 ^{a,l}	ⁿ	336.1	355.5
Tyrosine	336.5 ^{c,d}	336.4 \pm 2.7	347.4 ^{a,m}	ⁿ	344.2	354.7
Valine	349.3	350.2 \pm 2.5	354.0	355.0 \pm 4.0	356.4	361.3

^a Unless otherwise noted, the deprotonated ion has mc-cis structure. mc-cis = main chain deprotonation with cis-like orientation.

^b Previous work by our group. Reference 2. ^c sc structure. sc = side chain.

^d A higher energy folded sc structure was also predicted: $\Delta H=344.3$; $\Delta G=336.9$.

^e No deprotonated ion signal was obtained experimentally and therefore no experimental GA could be assigned.

^f Unless otherwise noted, the deprotonated ion has mc-trans structure. mc-trans = main chain deprotonation with trans-like orientation.

^g A lower energy sc structure was also predicted: $\Delta H=348.0$; $\Delta G=340.8$.

^h mc-cis/sc shared structure. A higher energy mc-trans/sc shared structure was also predicted: $\Delta H=342.5$; $\Delta G=336.1$.

ⁱ mc-trans/sc shared structure. A mc-cis/sc shared structure was predicted to have similar energies: $\Delta H=349.7$; $\Delta G=342.9$.

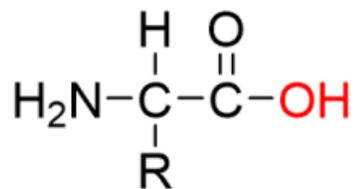
^j mc-cis/sc shared structure. A higher energy mc-trans/sc shared structure was also predicted: $\Delta H=341.8$; $\Delta G=335.0$.

^k A higher energy mc-trans structure was also predicted: $\Delta H=349.7$; $\Delta G=343.1$.

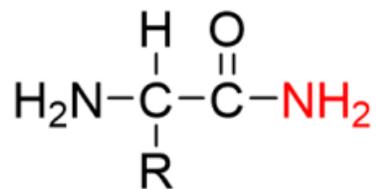
^l A higher energy mc-trans structure was also predicted: $\Delta H=360.4$; $\Delta G=353.0$.

^m A higher energy mc-trans structure was also predicted: $\Delta H=360.4$; $\Delta G=352.9$.

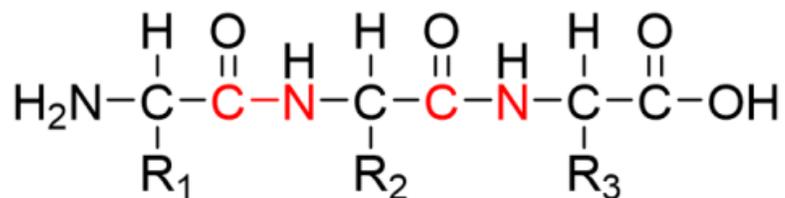
ⁿ Only one ion population was observed experimentally and only one experimental GA could be assigned.



Amino Acid



Amino Acid Amide



Locations of Peptide Bonds

Figure 4.1 Structures of an Amino Acid, Amino Acid Amide, and Tripeptide

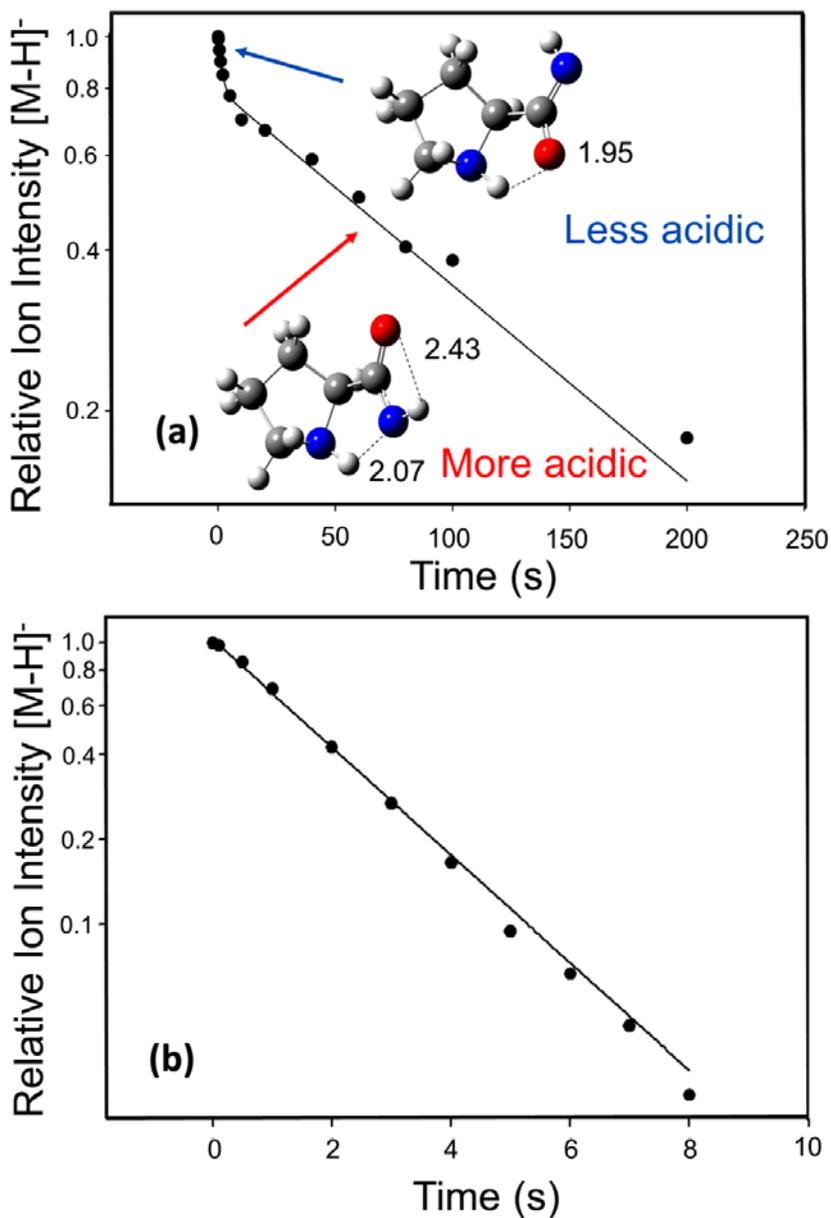
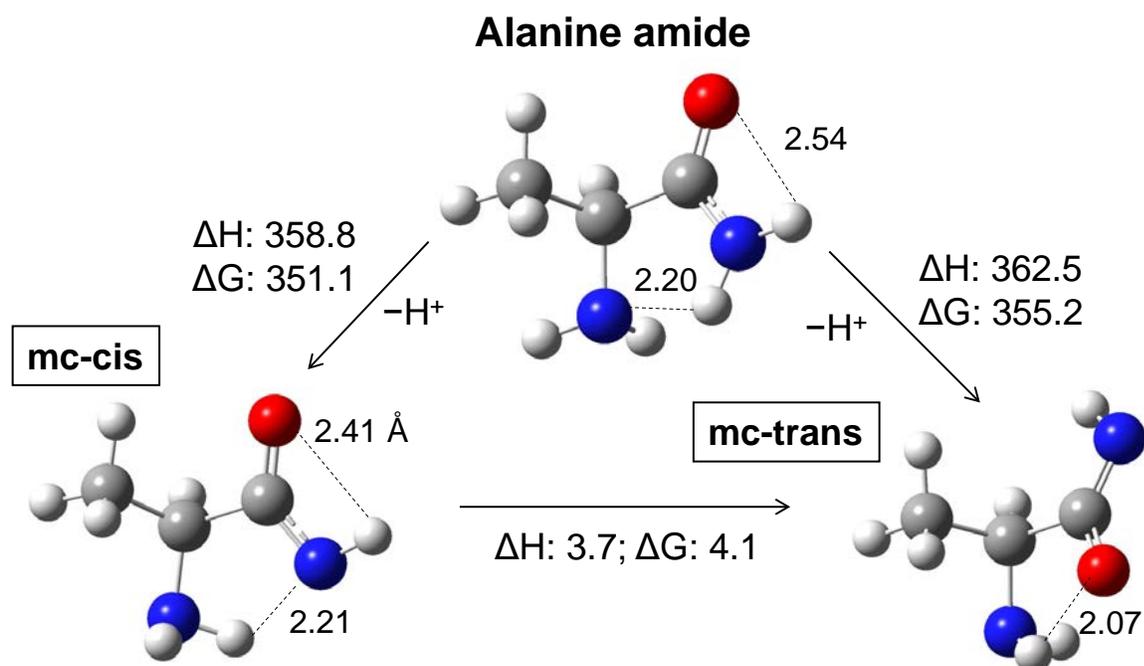
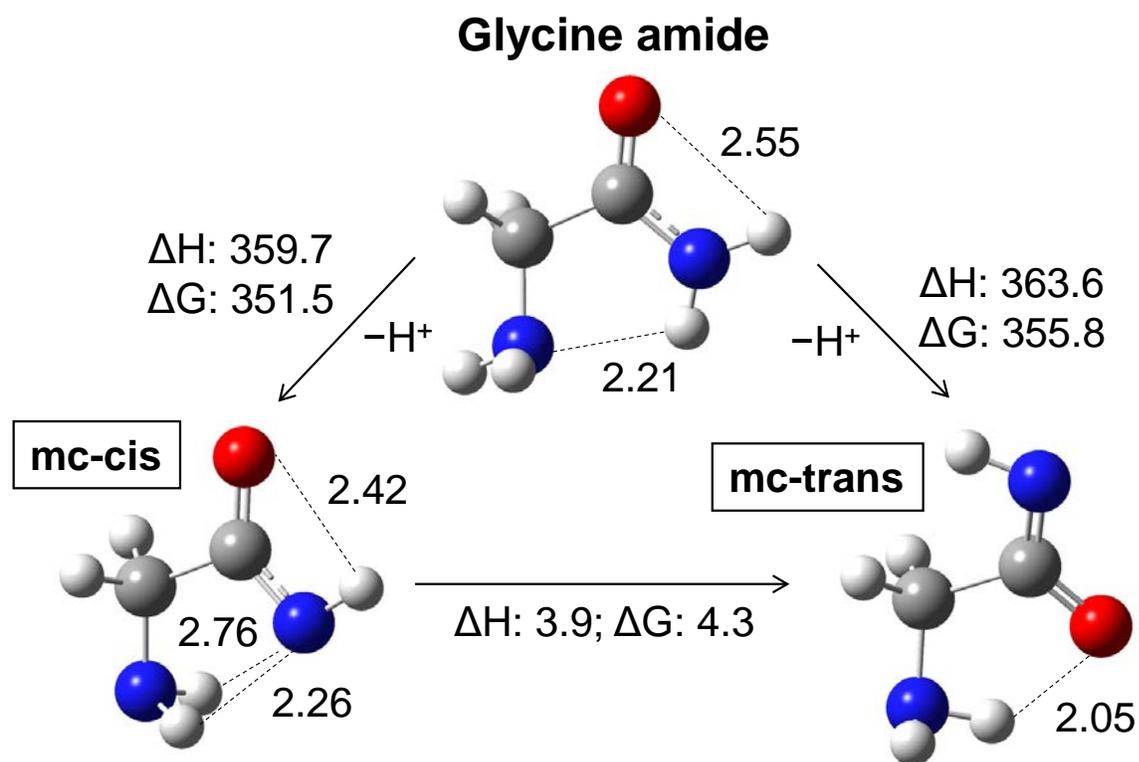


Figure 4.2. Reactant ion loss curves for deprotonated proline amide reacting with (a) 1,2,4,5-tetrafluorobenzene (GA = 353.3 kcal/mol) at a constant pressure of 7.8×10^{-8} mbar and (b) 3-methylpyrazole (GA = 348.3 kcal/mol) a constant pressure of 9.0×10^{-8} mbar. The logarithmic [M-H]⁻ intensity is plotted versus reaction time. In (a) the experimental data points (black circles) are fit to an equation involving the sum of two exponentials, while (b) involves a single exponential fit. Hydrogen bond distances in the structures of (a) are given in Å.

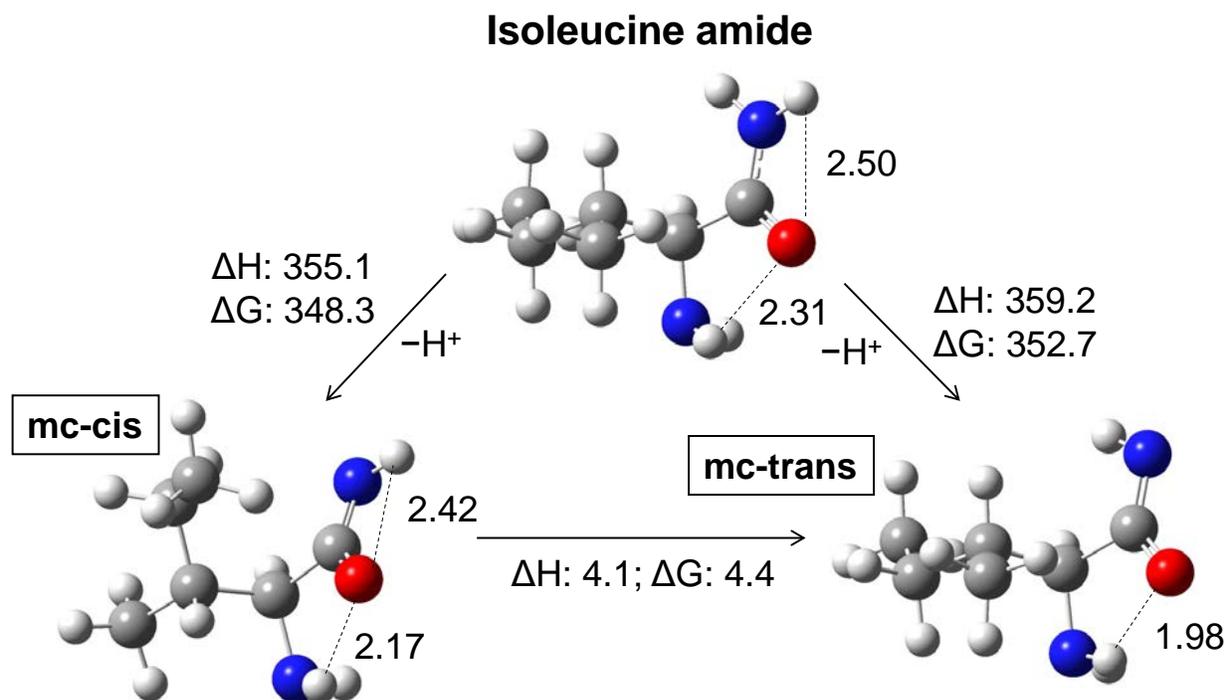
(a)



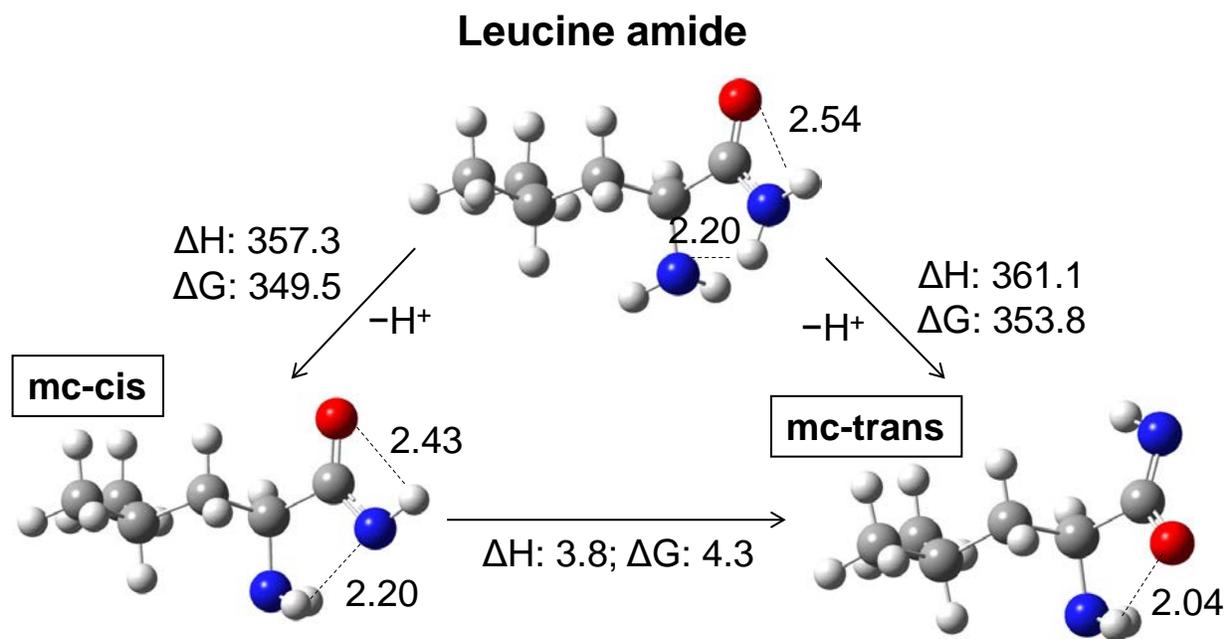
(b)



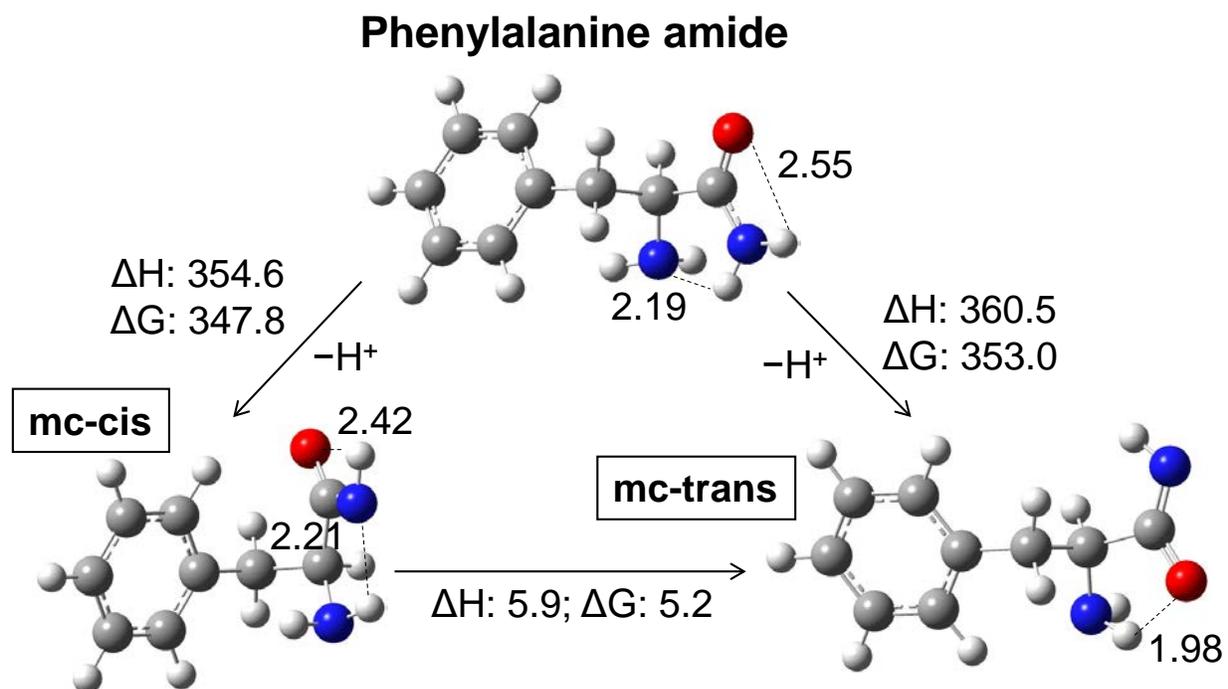
(c)



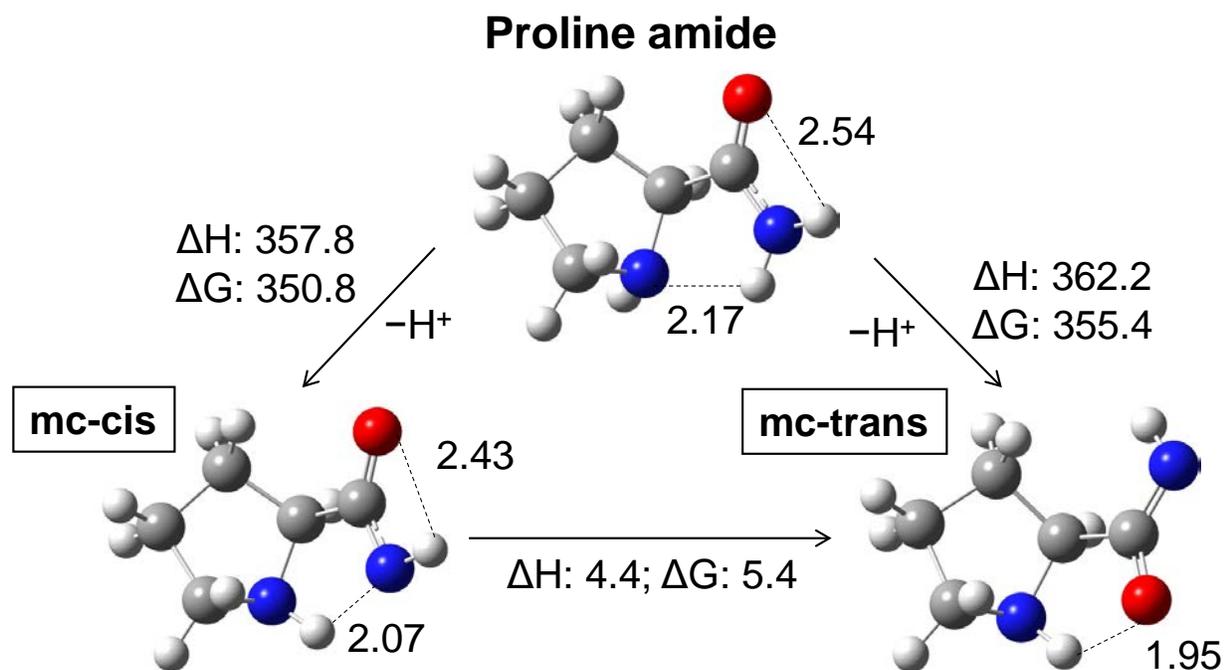
(d)



(e)



(f)



(g)

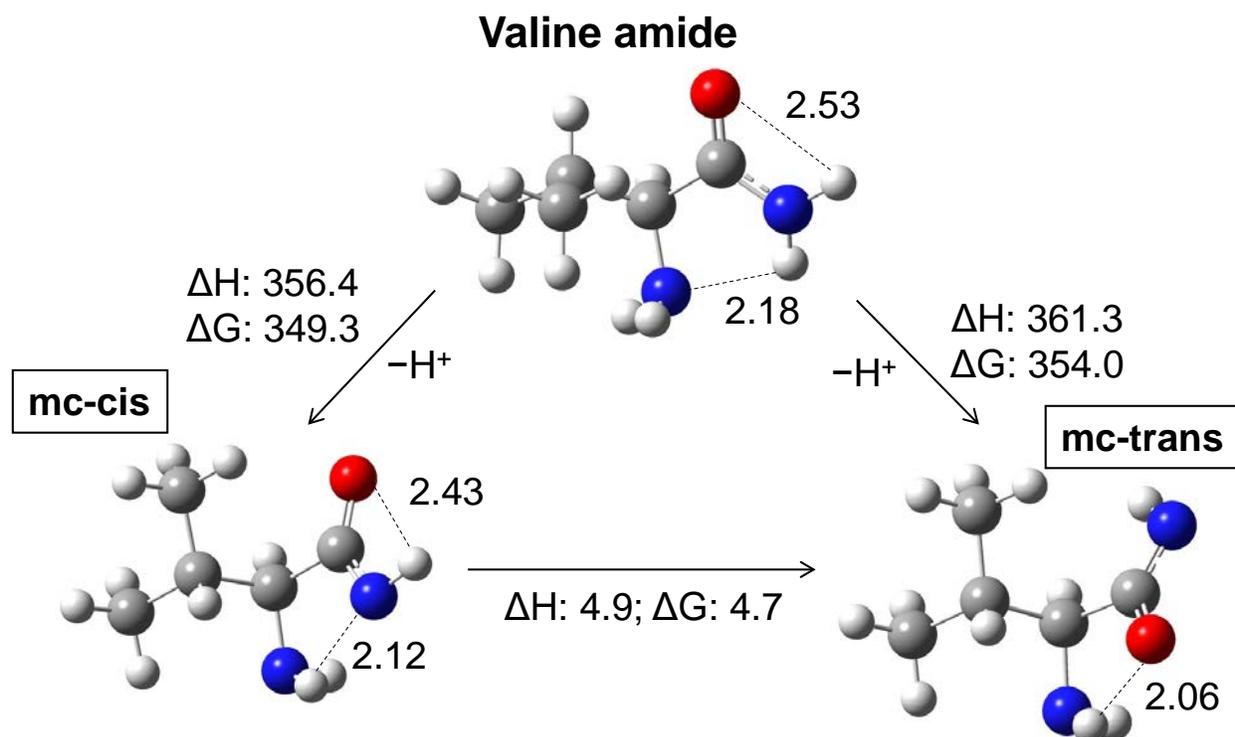
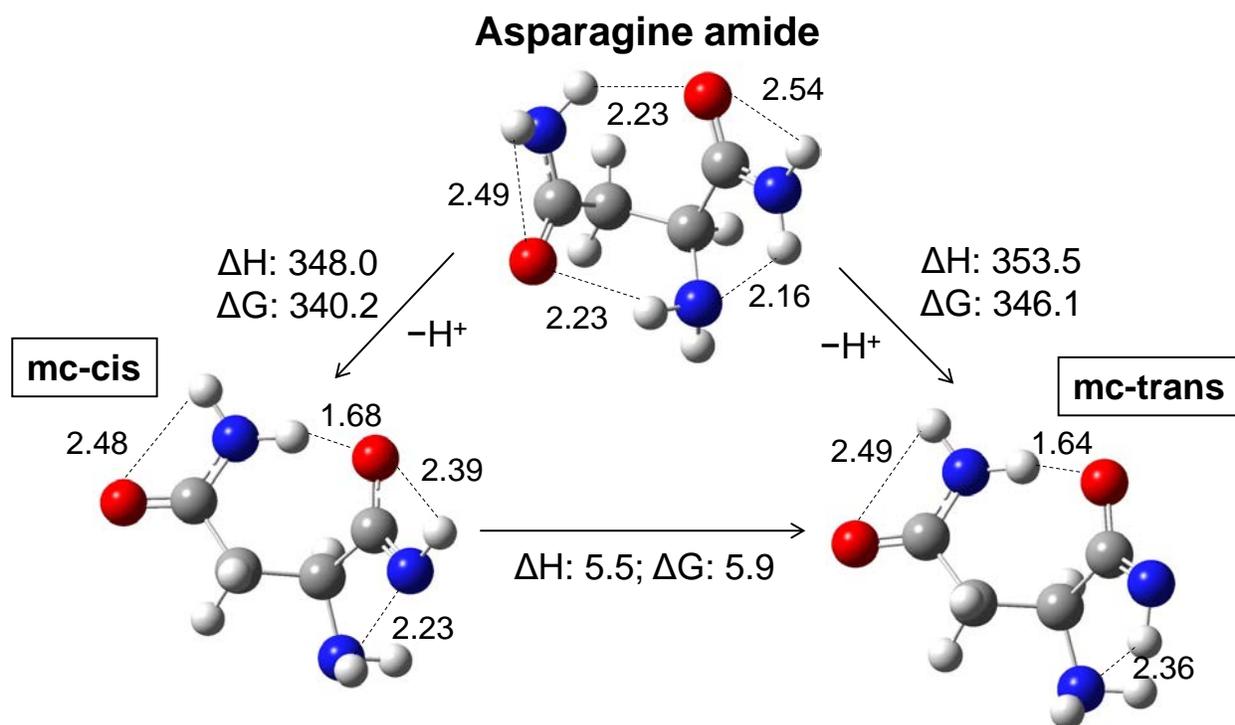
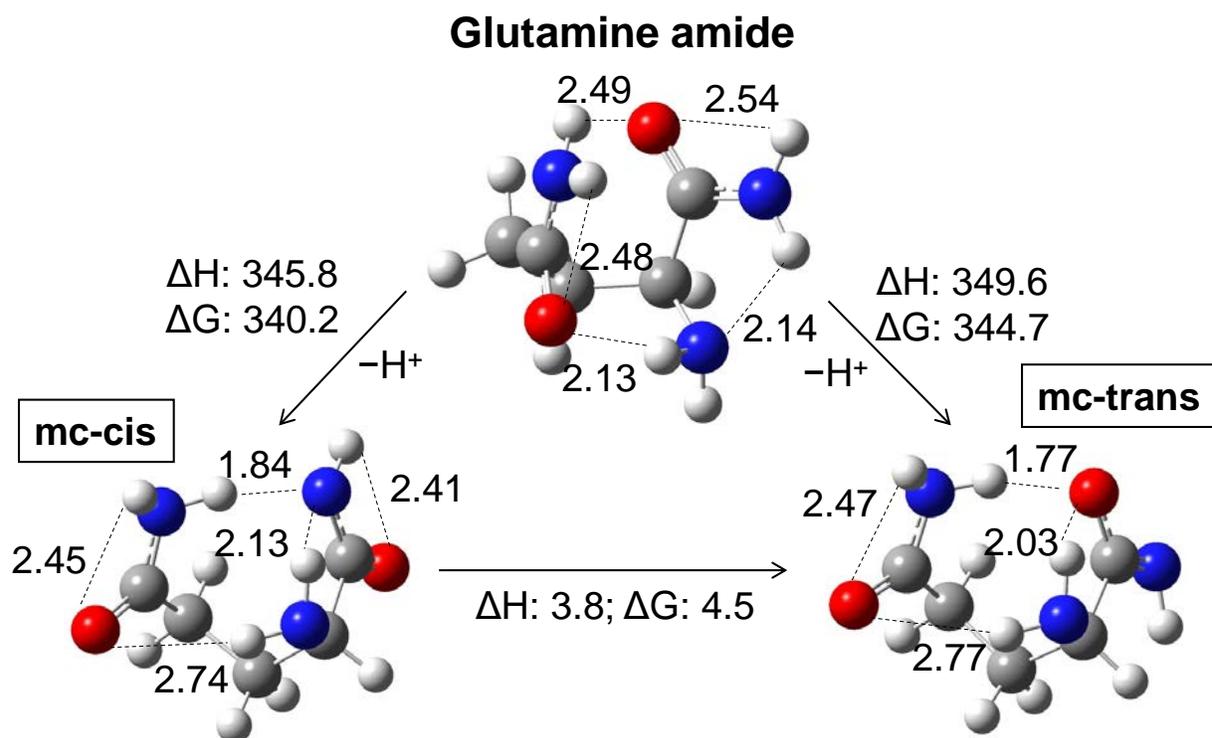


Figure 4.3. The most stable neutral and two anions for aliphatic side chain amino acid amides at the G3(MP2) level. Important hydrogen bond distances are given in Å. mc-cis = main chain deprotonation with cis-like orientation and mc-trans = main chain deprotonation with trans-like orientation (Note this orientation is of the N-H bond in the $[-\text{C}(=\text{O})\text{NH}]^-$ group). All energetic values are in kcal/mol. (a) alanine (b) glycine (c) isoleucine (d) leucine (e) phenylalanine (f) proline (g) valine

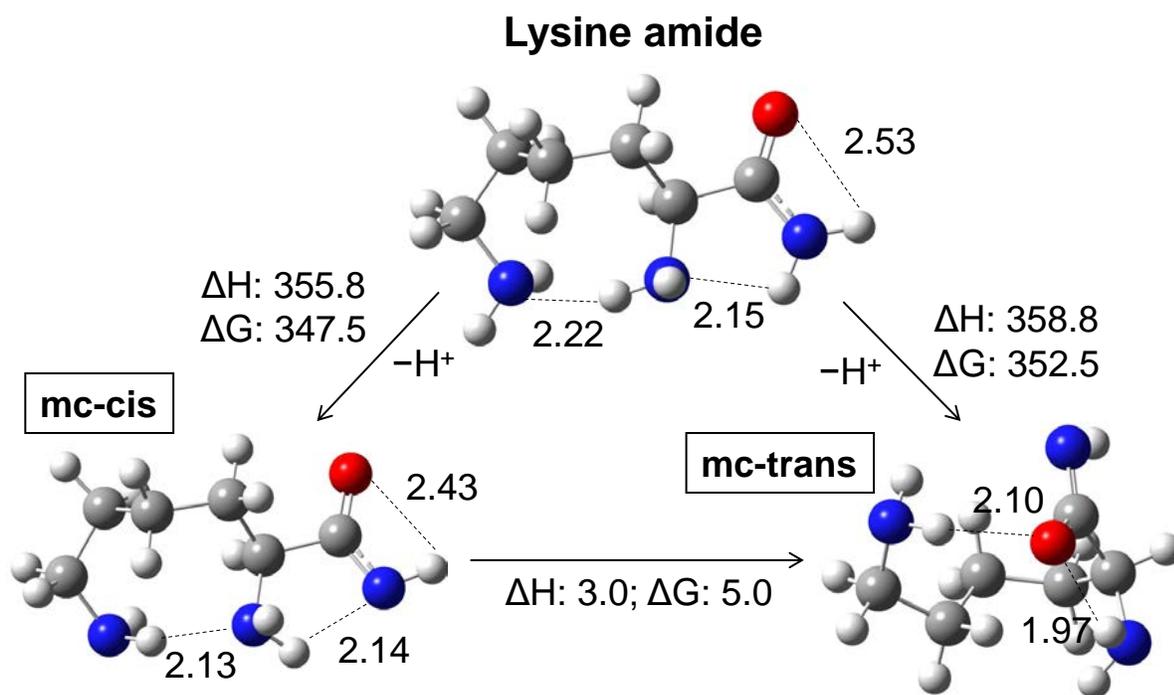
(a)



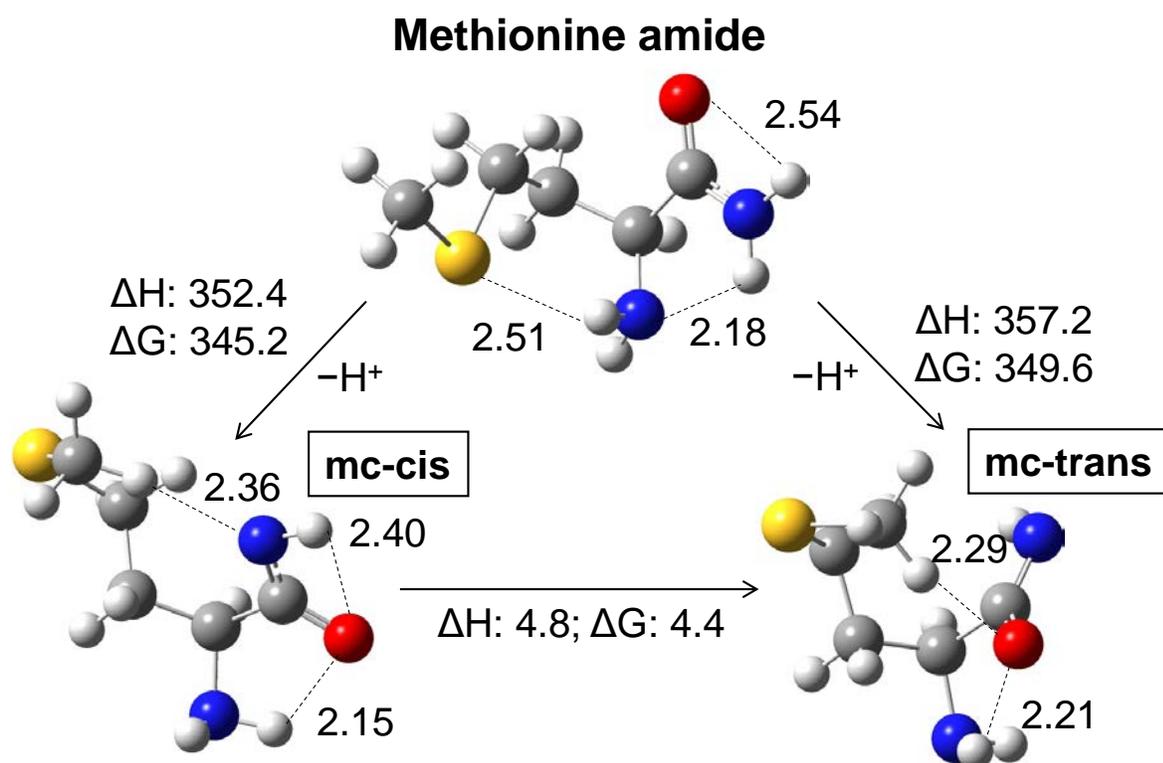
(b)



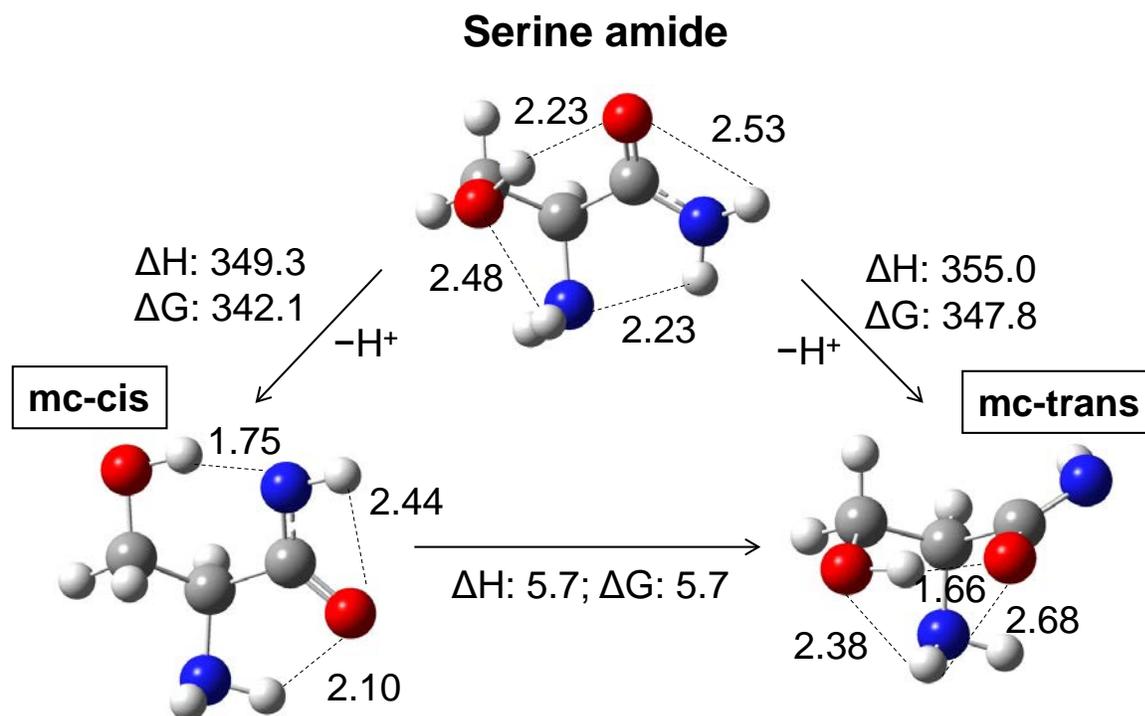
(c)



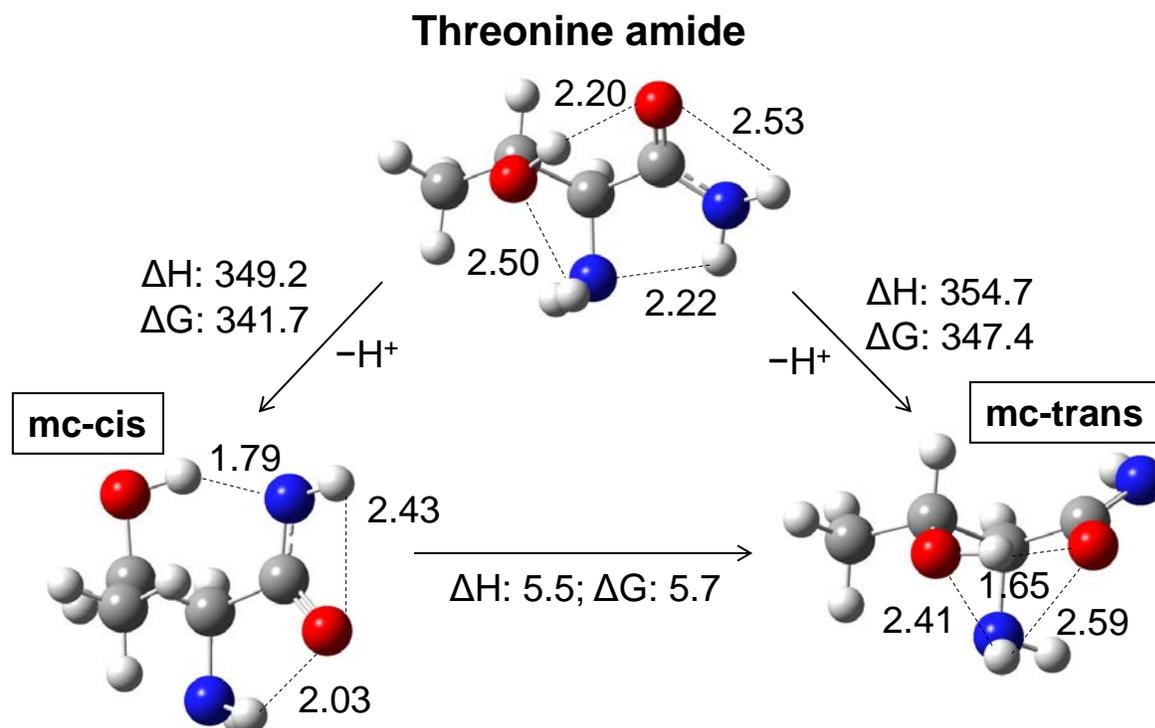
(d)



(e)



(f)



(g)

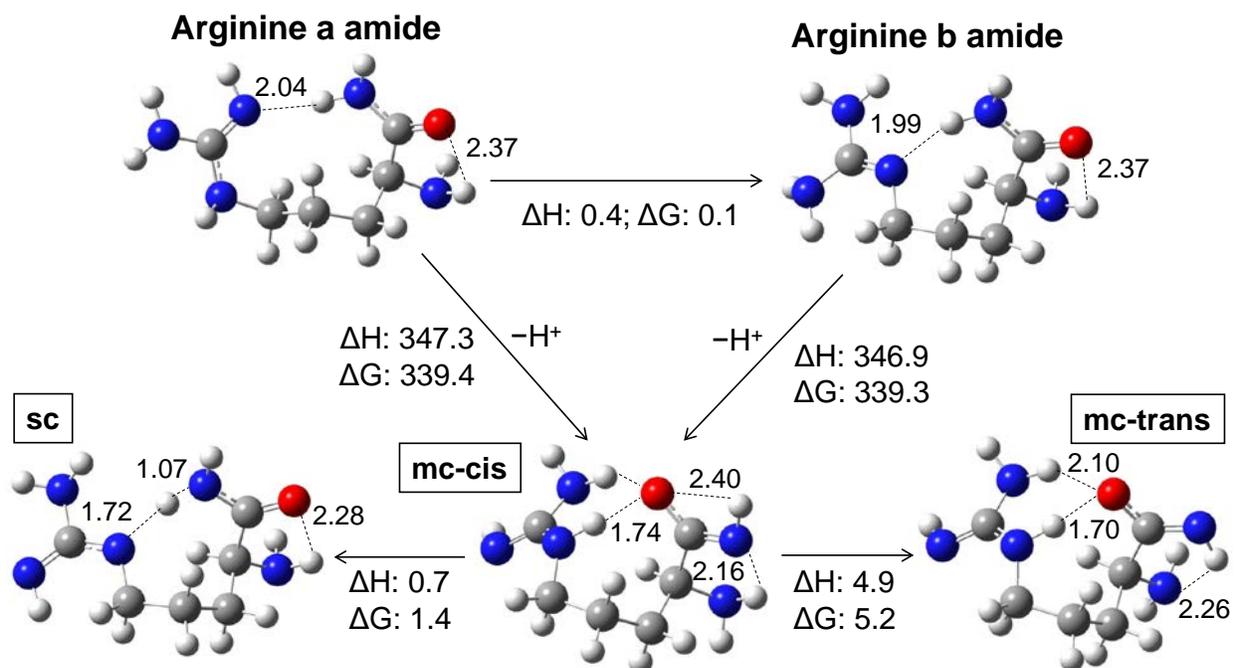
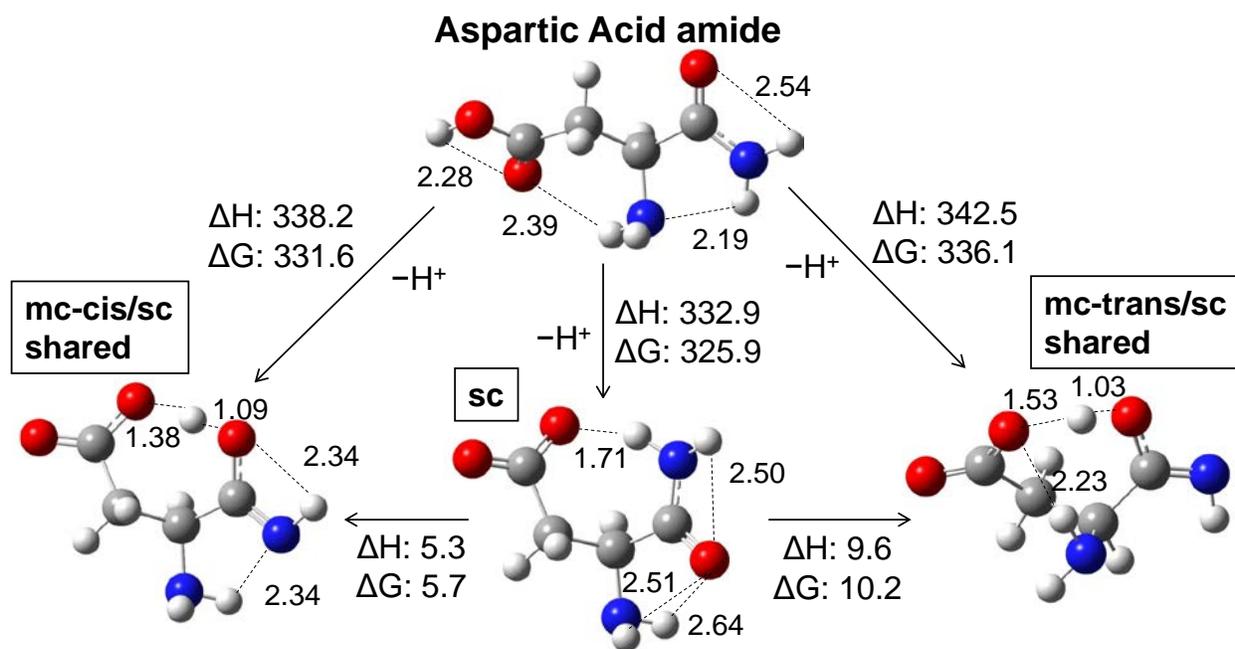
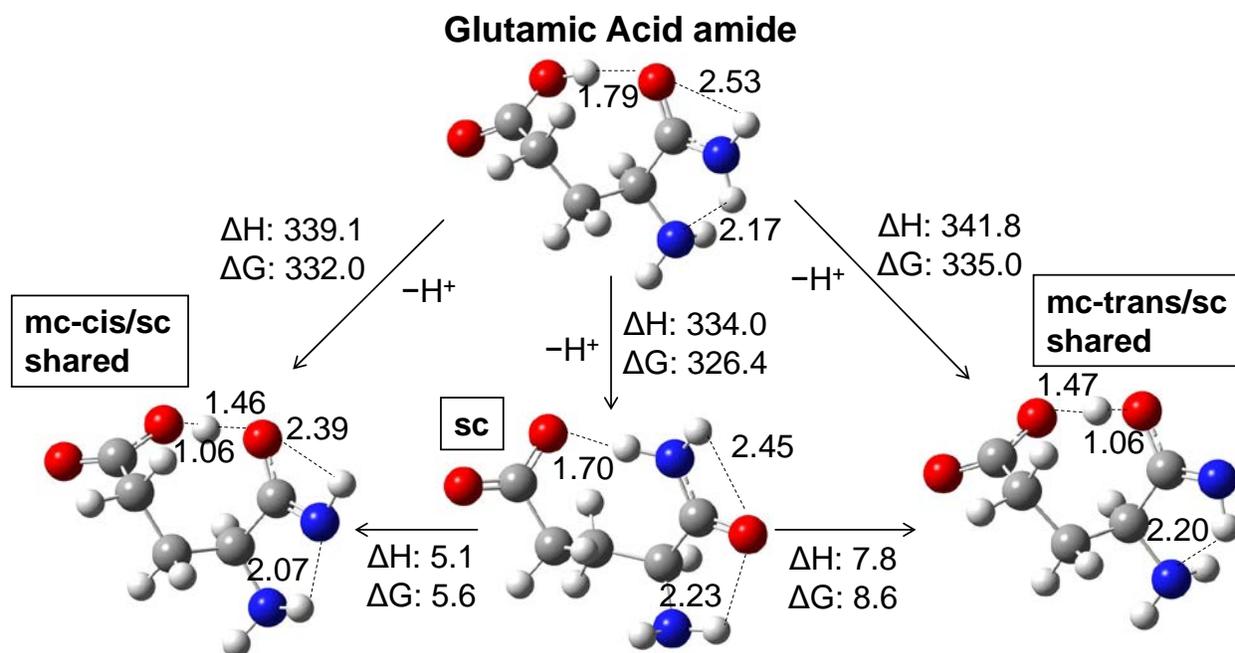


Figure 4.4. The most stable neutral and two anions for asparagine, glutamine, lysine, methionine, serine, threonine, and arginine amide at the G3(MP2) level. sc = side chain deprotonation. See Figure 4.3 caption for additional details. (a) asparagine (b) glutamine (c) lysine (d) methionine (e) serine (f) threonine (g) arginine

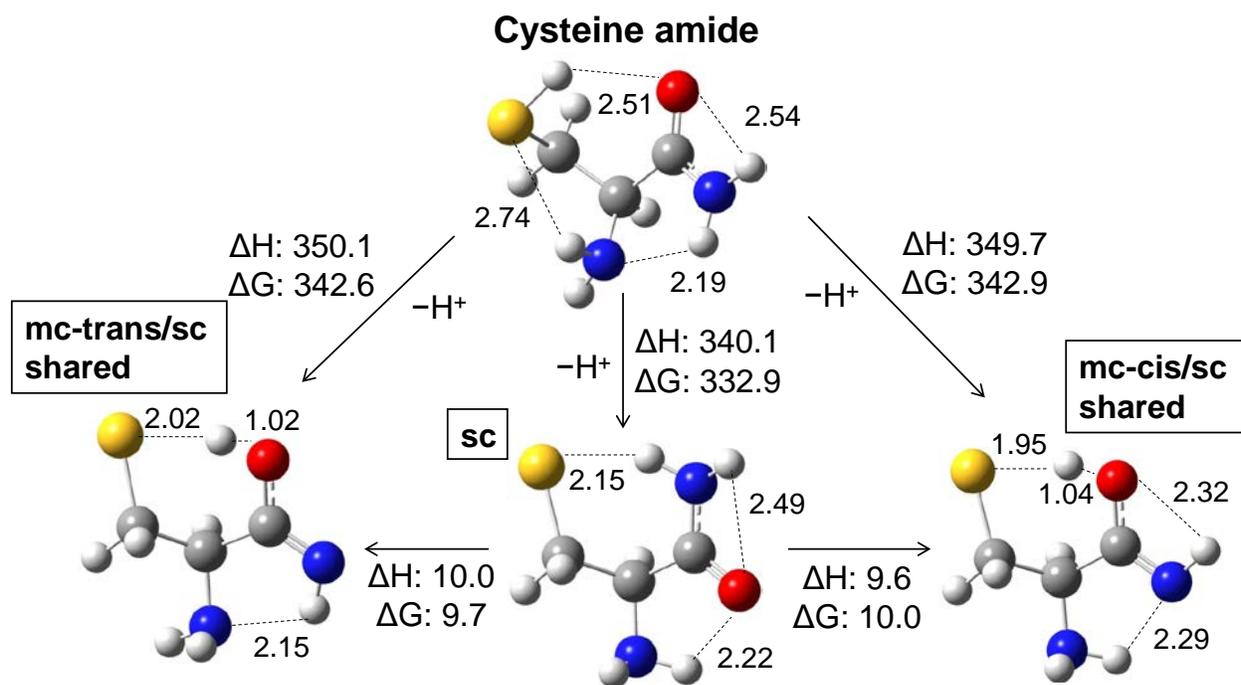
(a)



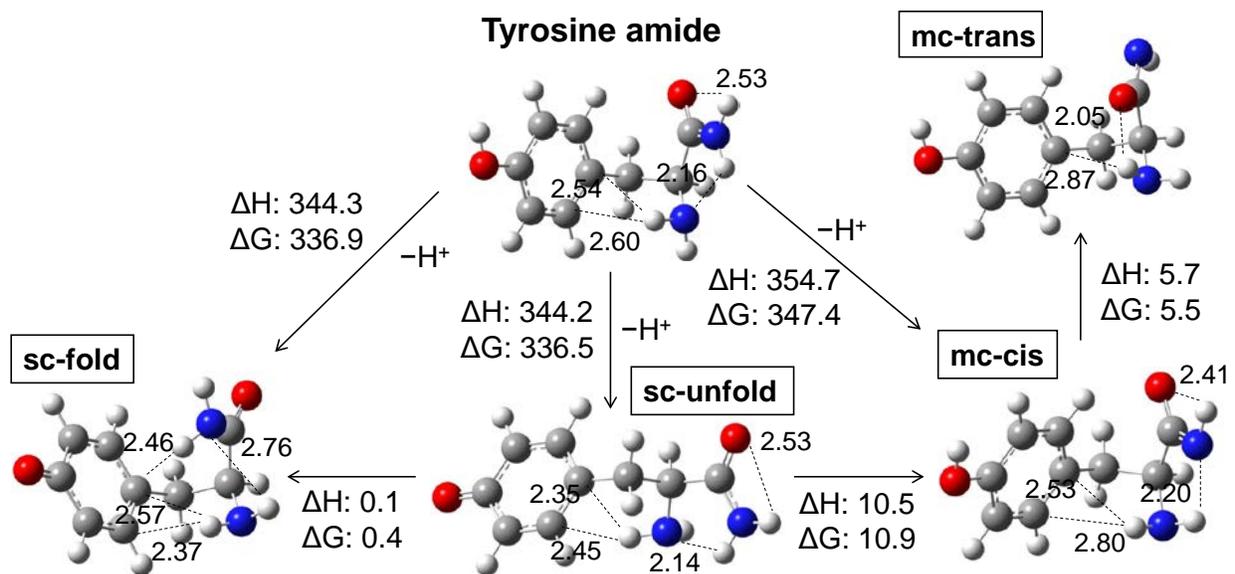
(b)



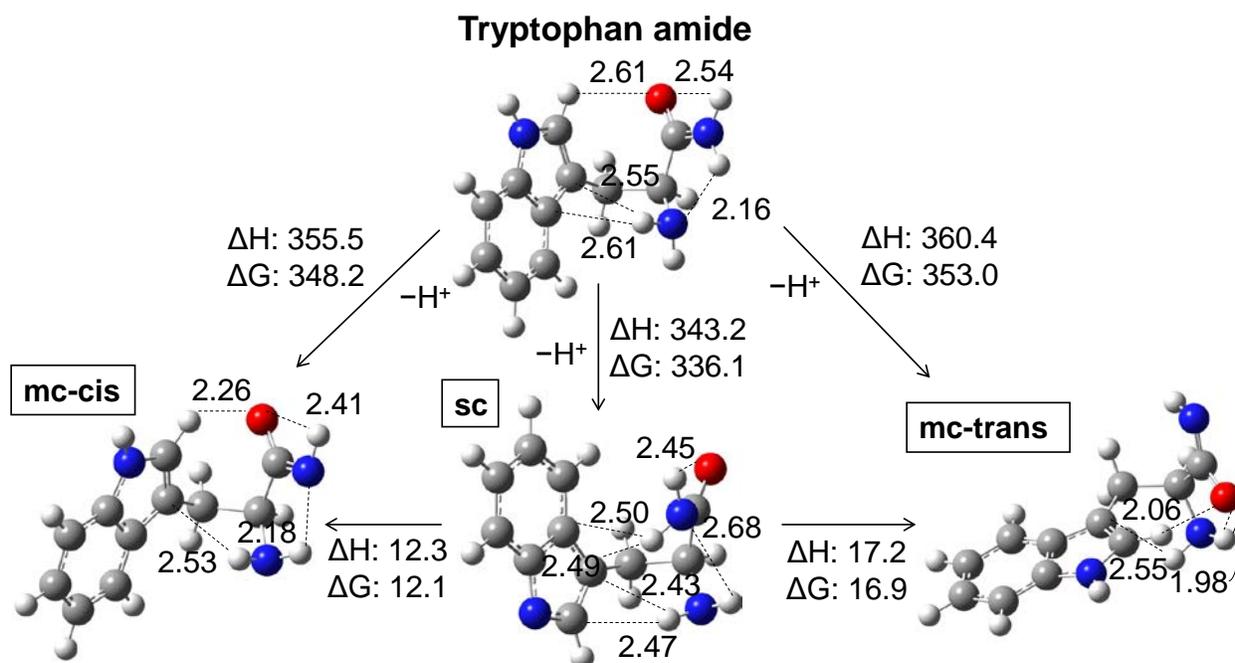
(c)



(d)



(e)



(f)

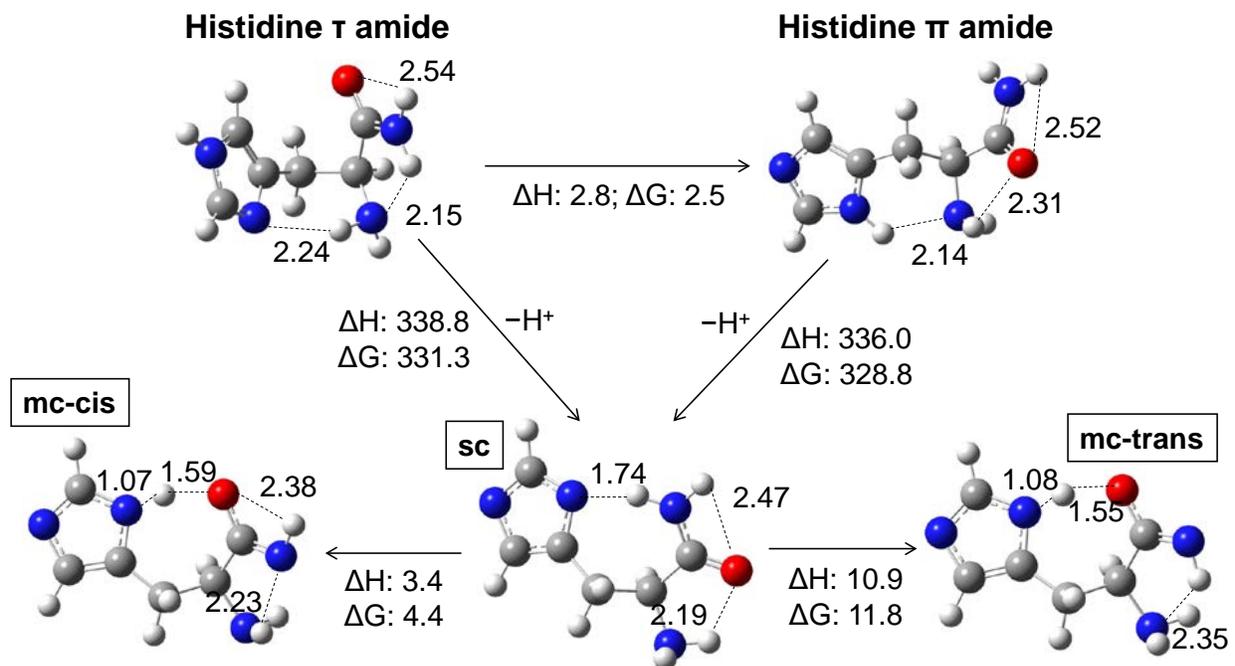


Figure 4.5. The most stable neutral and two anions for the amino acid amides with acidic side chains at the G3(MP2) level. mc-cis/sc = main chain deprotonation with cis-like orientation and proton sharing between the main and side chains and mc-trans = main chain deprotonation with trans-like orientation and proton sharing between the main and side chains. See Figure 4.3 caption for additional details. (a) aspartic acid (b) glutamic acid (c) tyrosine (d) cysteine (e) tryptophan (f) histidine

References

- ¹ Stover, M. L.; Jackson, V.; Matus, M.; Adams, M.; Cassady, C. J.; Dixon, D. A. Fundamental Thermochemical Properties of Amino Acids: Gas-Phase and Aqueous Acidities and Gas-Phase Heats of Formation. *J. Phys. Chem. B* **2012**, *116*, 2905-2916.
- ² Li, Z.; Matus, M. H.; Velazquez, H. A.; Dixon, D. A.; Cassady, C. J. Gas-Phase Acidities of Aspartic Acid, Glutamic Acid, and their Amino Acid Amides. *Int. J. Mass Spectrom.* **2007**, *265*, 213-223.
- ³ Bokatzian-Johnson, S. S.; Stover, M. L.; Dixon, D. A.; Cassady, C. J. Gas-Phase Deprotonation of the Peptide Backbone for Tripeptides and their Methyl Esters with Hydrogen and Methyl Side Chains. *J. Phys. Chem. B* **2012**, *116*, 14844-14858.
- ⁴ Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J. Am. Chem. Soc.* **1996**, *118*, 8365-8374.
- ⁵ Zhang, X.; Ewing, N. P.; Cassady, C. J. Effects of Basic Site Proximity on Deprotonation and Hydrogen/Deuterium Exchange Reactions for Model Dodecapeptide Ions Containing Lysine and Glycine. *Int. J. Mass Spectrom.* **1998**, *175*, 159-171.
- ⁶ Ewing, N. P.; Cassady, C. J. Dissociation of Multiply-Charged Negative Ions for Hirudin (54-65), Fibrinopeptide B, and Insulin A (Oxidized). *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 105-116.
- ⁷ Chen, Y.; Watson, H. M.; Gao, J.; Sinha, S. H.; Cassady, C. J.; Vincent, J. B. Characterization of the Organic Component of Low-Molecular-Weight Chromium-Binding Substance and its Binding of Chromium. *J. Nutr.* **2011**, *141*, 1225-1232.
- ⁸ Andrezza, H. J.; Wang, T.; Bagley, C. J.; Hoffmann, P.; Bowie, J. H. Negative Ion Fragmentations of Deprotonated Peptides. the Unusual Case of isoAsp: A Joint Experimental and Theoretical Study. Comparison with Positive Ion Cleavages. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 1993-2002.
- ⁹ Boontheung, P.; Brinkworth, C. S.; Bowie, J. H.; Baudinette, R. V. Comparison of the Positive and Negative Ion Electrospray Mass Spectra of some Small Peptides Containing Polyglutamate. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 287-292.
- ¹⁰ Bowie, J. H.; Brinkworth, C. S.; Dua, S. Collision-Induced Fragmentations of the [M-H]⁻ Parent Anions of Underivatized Peptides: An Aid to Structural Determination and some Unusual Negative Ion Cleavages. *Mass Spectrom. Rev.* **2002**, *21*, 87-107.
- ¹¹ Boontheung, P.; Alewood, P. F.; Brinkworth, C. S.; Bowie, J. H.; Wabnitz, P. A.; Tyler, M. J. Negative Ion Electrospray Mass Spectra of Caerulin Peptides: An Aid to Structural Determination. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 281-286.

-
- ¹² Men, L.; Wang, Y. Fragmentation of the Deprotonated Ions of Peptides Containing Cysteine, Cysteine Sulfinic Acid, Cysteine Sulfonic Acid, Aspartic Acid, and Glutamic Acid. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 777-784.
- ¹³ Gao, J.; Cassady, C. J. Negative Ion Production from Peptides and Proteins by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 4066-4072.
- ¹⁴ Jai-nhuknan, J.; Cassady, C. J. Negative Ion Postsource Decay Time-of-Flight Mass Spectrometry of Peptides Containing Acidic Amino Acid Residues. *Anal. Chem.* **1998**, *70*, 5122-5128.
- ¹⁵ Crizer, D. M.; Xia, Y.; McLuckey, S. A. Transition Metal Complex Cations as Reagents for Gas-Phase Transformation of Multiply Deprotonated Polypeptides. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1718-1722.
- ¹⁶ Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Brechi, L. A. Mobile and Localized Protons: A Framework for Understanding Peptide Dissociation. *J. Mass Spectrom.* **2000**, *35*, 1399-1406.
- ¹⁷ Bokatzian, S. S.; Stover, M. L.; Plummer, C. E.; Dixon, D. A.; Cassady, C. J. An Experimental and Computational Investigation into the Gas-Phase Acidities of Tyrosine and Phenylalanine: Three Structures for Deprotonated Tyrosine. *J. Phys. Chem. B* **2014**, *118*, 12630-12643.
- ¹⁸ Tian, Z.; Kass, S. R. Gas-Phase Versus Liquid-Phase Structures by Electrospray Ionization Mass Spectrometry. *Angew. Chem, Int. Ed.* **2009**, *48*, 1321-1323.
- ¹⁹ Tian, Z.; Wang, X. B.; Wang, L. S.; Kass, S. R. Are Carboxyl Groups the most Acidic Sites in Amino Acids? Gas-Phase Acidities, Photoelectron Spectra, and Computations on Tyrosine, P-Hydroxybenzoic Acid, and their Conjugate Bases. *J. Am. Chem. Soc.* **2009**, *131*, 1174-1181.
- ²⁰ Tian, Z.; Kass, S. R. Does Electrospray Ionization Produce Gas-Phase Or Liquid-Phase Structures? *J. Am. Chem. Soc.* **2008**, *130*, 10842-10843.
- ²¹ Oomens, J.; Steill, J. D.; Redlich, B. Gas-Phase IR Spectroscopy of Deprotonated Amino Acids. *J. Am. Chem. Soc.* **2009**, *131*, 4310-4319.
- ²² Tian, Z.; Pawlow, A.; Poutsma, J. C.; Kass, S. R. Are Carboxyl Groups the most Acidic Sites in Amino Acids? Gas-Phase Acidity, H/D Exchange Experiments, and Computations on Cysteine and its Conjugate Base. *J. Am. Chem. Soc.* **2007**, *129*, 5403-5407.
- ²³ Shen, J.; Ren, J. Gas Phase Acidity of a Cysteine Residue in Small Oligopeptides. *Int. J. Mass Spectrom.* **2012**, *316-318*, 147-156.

-
- ²⁴ Morishetti, K. K.; De Suan Huang, B.; Yates, J. M.; Ren, J. Gas-Phase Acidities of Cysteine-Polyglycine Peptides: The Effect of the Cysteine Position. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 603-614.
- ²⁵ Ren, J.; Tan, J. P.; Harper, R. T. Gas-Phase Acidities of Cysteine-Polyalanine Peptides I: A(3,4)CSH and HSCA(3,4). *J. Phys. Chem. A* **2009**, *113*, 10903-10912.
- ²⁶ Tan, J.; Ren, J. Determination of the Gas-Phase Acidities of Cysteine-Polyalanine Peptides using the Extended Kinetic Method. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 188-194.
- ²⁷ Waugh, R. J.; Bowie, J. H. A Review of the Collision Induced Dissociations of Deprotonated Dipeptides and Tripeptides. an Aid to Structure Determination. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 169-173.
- ²⁸ Brinkworth, C. S.; Dua, S.; Bowie, J. H. Backbone Cleavages of [M-H]⁻ Anions of Peptides. Cyclisation of Citropin 1 Peptides Involving Reactions between the C-Terminal [CONH]⁻ Residue and Backbone Amide Carbonyl Groups. A New Type of Beta Cleavage: A Joint Experimental and Theoretical Study. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 713-721.
- ²⁹ Marzluff, E. M.; Campbell, S.; Rodgers, M. T.; Beauchamp, J. L. Low-Energy Dissociation Pathways of Small Deprotonated Peptides in the Gas Phase. *J. Am. Chem. Soc.* **1994**, *116*, 7787-7796.
- ³⁰ Reiter, A.; Teesch, L. M.; Zhao, H.; Adams, J. Gas-Phase Fragmentations of Anionic Complexes of Serine- and Threonine-Containing Peptides. *Int. J. Mass Spectrom. Ion Proc.* **1993**, *127*, 17-26.
- ³¹ Pu, D.; Cassady, C. J. Negative Ion Dissociation of Peptides Containing Hydroxyl Side Chains. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 91-100.
- ³² Li, Z.; Yalcin, T.; Cassady, C. J. C-Terminal Amino Acid Residue Loss for Deprotonated Peptide Ions Containing Glutamic Acid, Aspartic Acid, Or Serine Residues at the C-Terminus. *J. Mass Spectrom.* **2006**, *41*, 939-949.
- ³³ Grzetic, J.; Oomens, J. Effect of the Asn Side Chain on the Dissociation of Deprotonated Peptides Elucidated by IRMPD Spectroscopy. *Int. J. Mass Spectrom.* **2013**, *354-355*, 70-77.
- ³⁴ Waugh, R. J.; Bowie, J. H.; Gross, M. L. Collision-Induced Dissociations of Deprotonated Peptides - Dipeptides Containing Asn, Arg, and Lys. *Aust. J. Chem.* **1993**, *46*, 693-702.
- ³⁵ 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.

-
- ³⁶ O'Hair, R. A. J.; Bowie, J. H.; Gronert, S. Gas-Phase Acidities of the Alpha-Amino Acids. *Int. J. Mass Spectrom. Ion Proc.* **1992**, *117*, 23-36.
- ³⁷ Anonymous In *Cleavage, Deprotection, and Isolation of Peptides after Fmoc Synthesis*; Applied Biosystems: 1998; .
- ³⁸ Scigelova, M.; Green, P.; Giannakopoulos, A.; Rodger, A.; Crout, D.; Derrick, P. A Practical Protocol for the Reduction of Disulfide Bonds in Proteins Prior to Analysis by Mass Spectrometry. *Eur. J. Mass Spectrom.* **2001**, *7*, 29-34.
- ³⁹ de Koning, L. J.; Nibbering, N. M. M.; van Orden, S. L.; Laukien, F. H. Mass Selection of Ions in a Fourier Transform Ion Cyclotron Resonance Trap using Correlated Harmonic Excitation Fields (CHEF). *Int. J. Mass Spectrom. Ion Proc.* **1997**, *165/166*, 209-219.
- ⁴⁰ Zhang, K.; Cassady, C. J.; Chung-Phillips, A. Ab Initio Studies of Neutral and Protonated Triglycines: Comparison of Calculated and Experimental Gas-Phase Basicity. *J. Am. Chem. Soc.* **1994**, *116*, 11512-11521.
- ⁴¹ Bartmess, J. E.; Georgiadis, R. M. Empirical Methods for Determination of Ionization Gauge Relative Sensitivities for Different Gases. *Vacuum* **1983**, *33*, 149-153.
- ⁴² Miller, K. J. Additivity Methods in Molecular Polarizability. *J. Am. Chem. Soc.* **1990**, *112*, 8533-8542.
- ⁴³ Su, T.; Chesnavich, W. J. Parametrization of the Ion-Polar Molecule Collision Rate-Constant by Trajectory Calculations. *J. Chem. Phys.* **1982**, *76*, 5183-5185.
- ⁴⁴ Su, T. Erratum: Trajectory Calculations of Ion-Polar Molecule Capture Rate Constants at Low Temperatures. *J. Chem. Phys.* **1988**, *89*, 5355-5355.
- ⁴⁵ Bouchoux, G.; Salpin, J. Y.; Leblanc, D. A Relationship between the Kinetics and Thermochemistry of Proton Transfer Reactions in the Gas Phase. *Int. J. Mass Spectrom. Ion Proc.* **1996**, *153*, 37-48.
- ⁴⁶ Bouchoux, G.; Salpin, J. Y. Re-Evaluated Gas Phase Basicity and Proton Affinity Data from the Thermokinetic Method. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 932-936.
- ⁴⁷ 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.
- ⁴⁸ Bouchoux, G.; Buisson, D.; Bourcier, S.; Sablier, M. Application of the Kinetic Method to Bifunctional Bases ESI Tandem Quadrupole Experiments. *Int. J. Mass Spectrom.* **2003**, *228*, 1035-1054.

-
- ⁴⁹ Gaussian 09, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009..
- ⁵⁰ Becke, A. D. Density-Functional Thermochemistry. III. the Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- ⁵¹ Parr, R. G.; Yang, W. In *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989; .
- ⁵² Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. Optimization of Gaussian-Type Basis Sets for Local Spin Density Functional Calculations. Part I. Boron through Neon, Optimization Technique and Validation. *Can. J. Chem.* **1992**, *70*, 560-571.
- ⁵³ Gutowski, K. E.; Dixon, D. A. Ab Initio Prediction of the Gas- and Solution-Phase Acidities of Strong Bronsted Acids: The Calculation of pKa Values Less than -10. *J. Phys. Chem. A* **2006**, *110*, 12044-12054.
- ⁵⁴ Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. Gaussian-3 Theory using Reduced Moller-Plesset Order. *J. Chem. Phys.* **1999**, *110*, 4703-4709.
- ⁵⁵ Dixon, D. A.; Feller, D.; Peterson, K. A. In *A Practical Guide to Reliable First Principles Computational Thermochemistry Predictions Across the Periodic Table*; Wheeler, R. A., Section Ed. Tschumper, G. S., Eds.; Annual Reports in Computational Chemistry; Elsevier: Amsterdam, 2012; Vol. 8, pp 1-28.
- ⁵⁶ Feller, D.; Peterson, K. A.; Dixon, D. A. Further Benchmarks of a Composite, Convergent, Statistically Calibrated Coupled-Cluster-Based Approach for Thermochemical and Spectroscopic Studies. *Mol. Phys.* **2012**, *110*, 2381-2399.
- ⁵⁷ Peterson, K. A.; Feller, D.; Dixon, D. A. Chemical Accuracy in Ab Initio Thermochemistry and Spectroscopy: Current Strategies and Future Challenges. *Theor. Chem. Acc.* **2012**, *131*, 1-20.
- ⁵⁸ Feller, D.; Peterson, K. A.; Dixon, D. A. A Survey of Factors Contributing to Accurate Theoretical Predictions of Atomization Energies and Molecular Structures. *J. Chem. Phys.* **2008**, *129*, 204105/1-204105/32.

-
- ⁵⁹ Feller, D.; Dixon, D. A. Extended Benchmark Studies of Coupled Cluster Theory through Triple Excitations. *J. Chem. Phys.* **2001**, *115*, 3484-3496.
- ⁶⁰ McQuarrie, D. A. In *Statistical Mechanics*; Harper & Row: New York, USA, 1976; .
- ⁶¹ Bartmess, J. E. In *Negative Ion Energetics Data*; Linstrom, P. J., Ed.; NIST Chemistry WebBook, NIST Standard Reference Database Number 69; National Institute of Standards and Technology: Gaithersburg MD, 20899, <http://webbook.nist.gov>, (retrieved April 14, 2015).
- ⁶² Gallivan, J. P.; Dougherty, D. A. Cation- π Interactions in Structural Biology. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9459-9464.
- ⁶³ Carr, S. R.; Cassady, C. J. Gas-Phase Basicities of Histidine and Lysine and their Selected Di- and Tripeptides. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 1203-1210.
- ⁶⁴ Wu, Z.; 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.
- ⁶⁵ Pu, D.; Clipston, N. L.; Cassady, C. J. A Comparison of Positive and Negative Ion Collision-Induced Dissociation for Model Heptapeptides with One Basic Residue. *J. Mass Spectrom.* **2010**, *45*, 297-305.
- ⁶⁶ Xia, Y.; Gunawardena, H. P.; Erickson, D. E.; McLuckey, S. A. Effects of Cation Charge-Site Identity and Position on Electron-Transfer Dissociation of Polypeptide Cations. *J. Am. Chem. Soc.* **2007**, *129*, 12232-12243.

CHAPTER 4 APPENDIX: AN EXPERIMENTAL AND COMPUTATIONAL STUDY OF THE GAS-PHASE ACIDITIES OF THE COMMON AMINO ACID AMIDES

H_{298} and G_{298} total energies for all neutrals and anions at the G3(MP2) level and reaction efficiency data for ion/molecule reactions used to assign experimental GA values.

Table A4.1. H_{298} and G_{298} Total Energies for the Amino Acid Amides and their Anions at the G3(MP2) level (a.u.).

Amino Acid	H_{298}	G_{298}
Alanine amide neutral	-303.416132	-303.455272
Alanine amide anion mc-cis	-302.846671	-302.885792
Alanine amide anion mc-trans	-302.840751	-302.879148
Arginine a amide neutral	-585.768853	-585.824175
Arginine b amide neutral	-585.768293	-585.823967
Arginine anion mc-cis	-585.217768	-585.273286
Arginine anion mc-trans	-585.209918	-585.265055
Arginine anion sc	-585.216596	-585.271012
Asparagine amide neutral	-471.914221	-471.960632
Asparagine amide anion mc-cis	-471.362079	-471.408516
Asparagine amide anion mc-trans	-471.353182	-471.399036
Aspartic acid amide neutral	-491.781663	-491.82821
Aspartic acid amide anion sc	-491.253441	-491.29887
Aspartic acid amide anion mc-cis/sc shared	-491.245073	-491.289685
Aspartic acid amide anion mc-trans/sc shared	-491.238276	-491.282615
Cysteine amide neutral	-701.173981	-701.216810
Cysteine amide anion sc	-700.634323	-700.676355
Cysteine amide anion mc-cis/sc shared	-700.618497	-700.660770
Cysteine amide anion mc-trans/sc shared	-700.614935	-700.655955
Glutamine amide neutral	-511.143515	-511.195253
Glutamine amide anion mc-cis	-510.594824	-510.643056
Glutamine amide anion mc-trans	-510.588694	-510.635906
Glutamic acid amide neutral	-531.016116	-531.063987
Glutamic acid amide anion sc	-530.486174	-530.533898
Glutamic acid amide anion mc-trans/sc shared	-530.478107	-530.524952
Glutamic acid amide anion mc-cis/sc shared	-530.473737	-530.520086
Glycine amide neutral	-264.179190	-264.214396
Glycine amide anion mc-cis	-263.608300	-263.644198
Glycine amide anion mc-trans	-263.602181	-263.637374
Histidine π amide neutral	-528.107414	-528.157154
Histidine τ amide neutral	-528.111916	-528.161115
Histidine anion sc	-527.574385	-527.623142
Histidine anion mc-cis	-527.568834	-527.616182
Histidine anion mc-trans	-527.557064	-527.604405

Isoleucine amide neutral	-421.114805	-421.165219
Isoleucine amide anion mc-cis	-420.551241	-420.600162
Isoleucine amide anion mc-trans	-420.544797	-420.593129
Leucine amide neutral	-421.117645	-421.167287
Leucine amide anion mc-cis	-420.550647	-420.600345
Leucine amide anion mc-trans	-420.544512	-420.593507
Lysine amide neutral	-476.386768	-476.438933
Lysine amide anion mc-cis	-475.822151	-475.871861
Lysine amide anion mc-trans	-475.817396	-475.867224
Methionine amide neutral	-779.633847	-779.684994
Methionine amide anion mc-cis	-779.074651	-779.124929
Methionine amide anion mc-trans	-779.067050	-779.117824
Phenylalanine amide neutral	-534.067328	-534.118727
Phenylalanine amide anion mc-cis	-533.504640	-533.554409
Phenylalanine amide anion mc-trans	-533.495153	-533.546214
Proline amide neutral	-380.685221	-380.728101
Proline amide anion mc-cis	-380.117468	-380.159045
Proline amide anion mc-trans	-380.110348	-380.151755
Serine amide neutral	-378.559506	-378.600403
Serine amide anion mc-cis	-378.005276	-378.045283
Serine amide anion mc-trans	-377.996111	-378.036190
Threonine amide neutral	-417.799329	-417.843330
Threonine amide anion mc-cis	-417.245246	-417.288721
Threonine amide anion mc-trans	-417.236441	-417.279663
Tryptophan amide neutral	-665.444563	-665.500159
Tryptophan amide anion sc	-664.899942	-664.954612
Tryptophan amide anion mc-cis	-664.880324	-664.935305
Tryptophan amide anion mc-trans	-664.872643	-664.927598
Tyrosine amide neutral	-609.216665	-609.270055
Tyrosine amide anion sc-unfolded	-608.670440	-608.723830
Tyrosine amide anion sc-folded	-608.670289	-680.723123
Tyrosine amide anion mc-cis	-608.653713	-608.706366
Tyrosine amide anion mc-trans	-608.644691	-608.697722
Valine amide neutral	-381.884897	-381.930354
Valine amide anion mc-cis	-381.319309	-381.363697
Valine amide anion mc-trans	-381.311440	-381.356206

Table A4.2. Reaction Efficiencies for the Proton Transfer Reactions of Proline, Valine, Leucine, Isoleucine, Lysine, Phenylalanine, Methionine, and Arginine Amides with Neutral Reference Compounds

Reference Compound	GAa (kcal/Mol)	Average Reaction Efficiency (\pm Standard Deviation)							
		Proline	Valine	Leucine	Isoleucine	Lysine	Phenylalanine	Methionine	Arginine
Cyclohexanone	358.7 \pm 2	0.25 \pm 0.02 (37 \pm 11%) ^b	0.15 \pm 0.06 (37 \pm 10%)	— ^c	—	—	—	—	—
		NR ^d (63 \pm 11%)	NR (63 \pm 10%)						
		BREAK ^e	BREAK						
Acetophenone	354.5 \pm 2	0.77 \pm 0.01 (28 \pm 5%)	0.29 \pm 0.10 (36 \pm 4%)	—	—	—	—	—	—
		0.0001 \pm 0.0003 (72 \pm 5%)	0.0006 \pm 0.0006 (64 \pm 4%)						
1,2,4,5-Tetrafluoro-benzene	353.3 \pm 2	0.0051 \pm 0.0021	0.52 \pm 0.08 (29 \pm 10%)	0.26 \pm 0.15 (31 \pm 15%)	0.19 \pm 0.10 (27 \pm 14%)	0.14 \pm 0.12 (37 \pm 12%)	—	—	—
			0.0019 \pm 0.0010 (71 \pm 10%)	0.0014 \pm 0.0005 (69 \pm 15%)	0.0013 \pm 0.0007 (73 \pm 14%)	0.0022 \pm 0.0020 (63 \pm 12%)			
				BREAK	BREAK	BREAK			
Pyrrole	350.9 \pm 2	0.088 \pm 0.016	0.089 \pm 0.15	0.30 \pm 0.09 (28 \pm 8%)	0.29 \pm 0.17 (29 \pm 11%)	0.29 \pm 0.10 (28 \pm 15%)	0.15 \pm 0.08 (34 \pm 7%)	0.09 \pm 0.02 (40 \pm 18%)	—
				0.009 \pm 0.001 (72 \pm 8%)	0.006 \pm 0.001 (71 \pm 11%)	0.009 \pm 0.002 (72 \pm 15%)	0.0003 \pm 0.0002 (66 \pm 7%)	0.007 \pm 0.002 (60 \pm 18%)	
			BREAK				BREAK		
Chloroform	349.9 \pm 2	0.37 \pm 0.13	0.36 \pm 0.12	0.38 \pm 0.09 (33 \pm 12%)	0.38 \pm 0.13 (30 \pm 6%)	0.39 \pm 0.14 (35 \pm 5 %)	0.29 \pm 0.12 (29 \pm 11%)	0.12 \pm 0.05 (36 \pm 12%)	—
				0.038 \pm 0.013	0.09 \pm 0.02	0.03 \pm 0.04	0.005 \pm 0.003	0.04 \pm 0.02	

				(67±12%)	(70 ± 6%)	(65 ± 5%)	(71 ± 11%)	(64 ± 12%)	
				0.64±0.14	0.68± 0.19	0.69± 0.20	0.34± 0.08	0.19± 0.01	0.13±0.02
				(30±10%)	(32 ± 17%)	(38 ±10%)	(33 ± 16%)	(38 ± 9%)	(42±20%)
3-Methyl pyrazole	348.3±2	0.41±0.16	0.39 ± 0.13	0.087± 0.024	0.13 ±0.04	0.08± 0.04	0.021± 0.003	0.13± 0.08	0.002± 0.002
				(70 ±10%)	(68 ± 17%)	(62 ±10%)	(67 ± 16%)	(62 ± 8%)	(58±20%)
				BREAK	BREAK	BREAK		BREAK	BREAK
							0.59± 0.18	0.36± 0.03	0.30±0.11
1-Butane- thiol	347.4±2	0.58±0.20	0.55 ± 0.18	0.32±0.20	0.43± 0.12	0.28 ±0.02	(36 ±8 %)	(35 ± 3 %)	(37±13%)
							0.09± 0.03	0.21± 0.11	0.08±0.05
							(64 ± 8%)	(65 ± 3%)	(63±13%)
4- Trifluoro- methyl aniline	346.0±2	—	—	0.49±0.23	0.54± 0.06	0.36 ±0.05	0.21± 0.06	0.62± 0.16	0.44±0.20
								(36 ± 8 %)	(34 ± 7%)
								0.22± 0.15	0.11±0.08
								(64 ± 8%)	(66 ± 7%)
							BREAK	BREAK	
Indene	344.6±2	—	—	—	—	—	0.38± 0.11	0.28± 0.03	0.20±0.15
Phenol	342.3±2	—	—	—	—	—	0.51± 0.14	0.68± 0.30	0.29±0.03
									BREAK
Acetic Acid	341.1±2	—	—	—	—	—	0.74± 0.21	0.77± 0.23	0.45±0.14

^a All reference compound GAs were obtained from reference 1 (Bartmess, J. E. In *Negative Ion Energetics Data*; Linstrom, P. J., Ed.; NIST Chemistry WebBook, NIST Standard Reference Database Number 69; National Institute of Standards and Technology: Gaithersburg MD, 20899, <http://webbook.nist.gov>, (retrieved April 14, 2015)).

^bTwo reaction efficiencies and the relative abundances of each ion population are listed; these reactions exhibited bimodal kinetics indicating two ion populations reacting at different rates.

^c—" indicates no reaction was performed. ^d "NR" indicates that no reaction, or a reaction with an efficiency <<< 0 was obtained.

^e "BREAK" indicates point where experimental GA value was assigned.

Table A4.3. Reaction Efficiencies for the Proton Transfer Reactions of Arginine, Threonine, Serine, Glutamine, and Asparagine Amides with Neutral Reference Compounds

Reference Compound	GA ^a (kcal/mol)	Average Reaction Efficiency (\pm Standard Deviation)				
		Arginine	Threonine	Serine	Glutamine	Asparagine
3-Methyl pyrazole	348.3 \pm 2	0.13 \pm 0.02 ^b (42 \pm 20%)	0.11 \pm 0.08 (39 \pm 12%)	0.12 \pm 0.02 (41 \pm 13%)	0.08 \pm 0.05 (38 \pm 10%)	0.06 \pm 0.09 (39 \pm 15%)
		0.002 \pm 0.002 (58 \pm 20%)	NR ^c (61 \pm 12%)	NR (59 \pm 13%)	NR (62 \pm 10%)	NR (61 \pm 15%)
		BREAK ^d	BREAK	BREAK		
1-Butanethiol	347.4 \pm 2	0.27 \pm 0.11 (37 \pm 13%)	0.27 \pm 0.09 (36 \pm 10%)	0.29 \pm 0.07 (38 \pm 9%)	0.17 \pm 0.11 (34 \pm 7%)	0.15 \pm 0.03 (31 \pm 6%)
		0.08 \pm 0.05 (63 \pm 13%)	0.05 \pm 0.03 (64 \pm 10%)	0.08 \pm 0.02 (62 \pm 9%)	NR (66 \pm 7%)	NR (69 \pm 6%)
					BREAK	BREAK
4-Trifluoromethyl aniline	346.0 \pm 2	0.44 \pm 0.20 (34 \pm 7%)	0.48 \pm 0.14 (33 \pm 8%)	0.53 \pm 0.10 (36 \pm 5%)	0.34 \pm 0.02 (37 \pm 7%)	0.33 \pm 0.13 (40 \pm 9%)
		0.11 \pm 0.08 (63 \pm 7%)	0.09 \pm 0.05 (67 \pm 8%)	0.12 \pm 0.08 (64 \pm 5%)	0.008 \pm 0.002 (63 \pm 7%)	0.006 \pm 0.001 (60 \pm 9%)
Indene	344.6 \pm 2	0.20 \pm 0.15 BREAK	0.19 \pm 0.08 BREAK	0.22 \pm 0.12 BREAK	0.05 \pm 0.02	0.03 \pm 0.01
Phenol	342.3 \pm 2	0.29 \pm 0.03	0.29 \pm 0.09	0.31 \pm 0.09	0.14 \pm 0.05	0.17 \pm 0.10
Acetic Acid	341.2 \pm 2	0.45 \pm 0.14	0.41 \pm 0.03	0.58 \pm 0.18	0.24 \pm 0.08 BREAK	0.24 \pm 0.06 BREAK
Formic Acid	339.1 \pm 2	0.69 \pm 0.10	0.77 \pm 0.20	0.92 \pm 0.21	0.38 \pm 0.04	0.37 \pm 0.08
Isovaleric Acid	338.5 \pm 2	– ^e	–	–	0.60 \pm 0.18	0.55 \pm 0.14
Trimethylacetic Acid	337.6 \pm 2	–	–	–	0.83 \pm 0.15	0.79 \pm 0.18

^a GAs were obtained from reference (1). See table A4.1. ^bTwo reaction efficiencies and the relative abundances of each ion

population; indicates two ion populations reacting at different rates. ^c“–” indicates no reaction was performed. ^d“NR” indicates that no reaction, or a reaction with an efficiency $\lll 0$. ^e“BREAK” indicates point where experimental GA value was assigned.

Table A4.4. Reaction Efficiencies for the Proton Transfer Reactions of Tryptophan, Cysteine, Tyrosine, and Histidine Amides with Neutral Reference Compounds

Reference Compound	GA ^a (kcal/mol)	Average Reaction Efficiency (\pm Standard Deviation)			
		Tryptophan	Cysteine	Tyrosine	Histidine
Acetic Acid	341.2 \pm 2	– ^b	–	0.005 \pm 0.004	
Formic Acid	339.1 \pm 2	0.007 \pm 0.004	0.005 \pm 0.002	0.11 \pm 0.03	–
Isovaleric Acid	338.5 \pm 2	0.08 \pm 0.03	0.05 \pm 0.01	0.13 \pm 0.05	–
Trimethylacetic Acid	337.6 \pm 2	0.15 \pm 0.07	0.13 \pm 0.04	0.22 \pm 0.10	–
		BREAK^c	BREAK	BREAK	
Ethyl cyanoacetate	333.6 \pm 2	0.42 \pm 0.11	0.41 \pm 0.08	0.39 \pm 0.20	0.02 \pm 0.02
4-amino-2,3,5,6-tetrafluoropyridine	332.8 \pm 2	0.71 \pm 0.16	0.68 \pm 0.13	0.48 \pm 0.17	0.09 \pm 0.03
3-Trifluormethyl phenol	332.4 \pm 2	–	–	1.05 \pm 0.09	0.11 \pm 0.09
					BREAK
3,3,3-Trifluoro-propionic acid	326.9 \pm 2	–	–	–	0.66 \pm 0.13

^a All reference compound GAs were obtained from reference (1). See table A4.1. ^b“–” indicates no reaction was performed.

^c“BREAK” indicates point where experimental GA value was assigned.

CHAPTER 5: GAS- AND SOLUTION-PHASE ACIDITIES OF PHOSPHORYLATED AMINO ACIDS AND THEIR AMIDES

5.1 Introduction Post-translational modifications (PTMs) allow for the function of a protein to be altered through covalent modification, e.g., the transformation of a side chain functionality into another one.^{1,2} Phosphorylation is a common PTM and plays an important role in various cellular processes such as cell signaling.³ Phosphorylation involves the substitution of a phosphate group on the side chain of an amino acid residue, which adds a site for proton loss and increases its capacity for hydrogen bonding.⁴ Because phosphorylation occurs after the protein has been encoded, the process of phosphorylation enables the cell to meet changes in its environment. In eukaryotes, phosphorylation most commonly occurs at the hydroxyl side chain of serine, threonine, and tyrosine. The relative abundances of phosphoserine (pSer), phosphothreonine (pThr), and phosphotyrosine (pTyr) in the human proteome have been estimated to be 90%, 10%, and 0.05% respectively.⁵ Additional amino acids including cysteine, histidine, arginine, lysine, aspartic acid, glutamic acid, and glycine have been shown to undergo phosphorylation.⁶

Phosphorylation is a reversible process and can have a very short lifetime, which makes it a challenging process to analyze. Additionally, the multitude of potential sites for phosphorylation and the potential for isoforms to exist in the proteome further complicate the analysis. Mass spectrometry is an important method in the analysis of phosphoproteomes because of its ability to identify and sequence phosphorylated peptides and determine the sites of

phosphorylation.⁷ The use of soft ionization techniques such as electrospray ionization (ESI)⁸ and matrix-assisted laser desorption ionization (MALDI)⁹, allow for the study of proton transfer reactions. Studying the gas-phase acidities (GAs), ΔG of $AH \rightarrow A^- + H^+$, of phosphopeptides provides detailed information on structural conformations and energetics. Because phosphopeptides contain a very acidic phosphate group, negative ion mode mass spectrometry could provide a more suitable environment for studying their behavior in the gas phase.¹⁰

The most commonly used tandem mass spectrometry (MS/MS) techniques for identifying phosphopeptides are collision-induced dissociation (CID), electron capture dissociation (ECD), and electron transfer dissociation (ETD).⁷ Phosphopeptides create many problems for detection and identification with MS because their strong acidities lead to lower positive ion mode ionization efficiencies by ESI and MALDI.^{11,12,13} Additionally, sequence information can be lost as a result of the absence of product ions consistent with the phosphate group or the migration of the phosphate group.^{14,15,16,17,18,19,20} A range of studies have been performed using negative ion mode CID to sequence phosphopeptides and have shown that the phosphorylated residues can be identified and characterized based on the types of neutral loss the peptides undergo and the negative ions formed.^{10, 11,15,21,22,23,24,25,26,27,28,29,30,31,32,33}

Despite the interest in phosphorylated peptides and amino acids, there has not been much work done to study the gas-phase thermochemical properties of individually phosphorylated amino acids. Ohanessian and coworkers³⁴ used infrared multiple photon dissociation (IRMPD) spectroscopy together with density functional theory (DFT) at the B3LYP/6-31+G* level to determine the vibrational signatures of protonated $[pSer+H]^+$, $[pThr+H]^+$, and $[pTyr+H]^+$. They suggest that the vibrational modes of phosphate groups between 900 – 1300 cm^{-1} can be used to identify the site of phosphorylation in phosphopeptides as the P–O–H and P=O bands directly

relate to the amino acid to which the phosphate group is bound. Subsequently, Ohanessian and coworkers³⁵ applied these same methods to report the IRMPD spectra of a protonated phosphodipeptide. Because the phosphate group is expected to be quite acidic and deprotonated in the gas phase, Maitre and coworkers³⁶ used IRMPD to study the IR spectra of deprotonated [pSer-H]⁻, [pThr-H]⁻, and [pTyr-H]⁻. They first measured the IRMPD spectrum of H₂PO₄⁻ to obtain the characteristic frequencies of the deprotonated phosphate bending and stretching modes. They then used these frequencies together with calculated vibrational spectra at the DFT/6-31+G* and DFT/6-311+G** levels of theory to determine that the site of deprotonation for all three phosphorylated amino acid anions was the phosphate group. They predicted that serine and threonine have a folded structure, whereas tyrosine has an extended structure. Gaigeot and coworkers^{37,38} performed DFT-based Car-Parrinello molecular dynamics (CPMD) simulations at room temperature on deprotonated [pSer-H]⁻ and protonated [pSer+H]⁺ to predict the effects of temperature and to obtain anharmonic frequency corrections. By comparing the spectra of [pSer-H]⁻ and [pSer+H]⁺, they predicted the main vibrational modes that can be used to determine the protonation state of the phosphate, carboxylate, and amino groups. They suggest that temperatures effects will increase as the size of the phosphorylated peptide increases. Additional IRMPD studies³⁹ have been performed on monohydrated [pTyr•H₂O+H]⁺ together with calculations at the B3LYP-D/SVP, M06/6-31G(*d,p*), and MP2/SVP levels of theory to investigate the effect of adding water to the phosphorylated amino acids on the vibrational spectra. These studies showed that the added water molecule creates a hydrogen bonding bridge between the phosphate and ammonium groups resulting in significant changes to the characteristic frequencies. In 1996, Hearn and coworkers⁴⁰ performed potentiometric investigations to determine the pK_a of pSer and found a value of 2.19. In 2010, Smiechowski⁴¹

performed DFT with the SMD solvation model calculations to predict the pK_a of pSer to be 1.6 in good agreement with experiment.

Building on our previous work on amino acids and peptides,^{42,43,44,45} we extended our studies to the phosphorylated amino acids. The GAs of ten phosphorylated amino acids and their corresponding amides have been predicted at the G3(MP2) level of theory. In addition to our work, there are a number of other studies of the gas-phase acidities of the amino acids.^{46,47,48} Because of the importance of pSer, pThr, and pTyr in the human proteome and the potential for the formation of a gas-phase zwitterion (a salt bridge formed by an ammonium group between two anionic centers), we have performed additional calculations at the MP2 level with augmented correlation-consistent basis sets⁴⁹ up through the triple- ζ to further study the hydrogen bonding present in the anionic structures. Theoretical pK_a values in aqueous solution were predicted by self-consistent reaction field⁵⁰ calculations using the COSMO parameterization.⁵¹ Additionally, the heats of formation of the neutral phosphorylated amino acids and phosphorylated amino acid amides were calculated from the G3(MP2) atomization energies and isodesmic reactions using their corresponding non-phosphorylated L-common amino acid and amino acid amide parents respectively.

5.2 Computational and Experimental Methods *Computational Methods* Calculations were performed at the density functional theory (DFT) and correlated molecular orbital (MO) theory levels with the programs Gaussian-03 and 09.⁵² Geometries were initially optimized at the DFT level with the B3LYP exchange-correlation functional^{53, 54} and the DZVP2 basis set.⁵⁵ 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. Ten phosphorylated L-amino acids (arginine,

aspartic acid, cysteine, glutamic acid, glycine, histidine, lysine, serine, threonine, and tyrosine), their corresponding amides, and the corresponding anions were studied. A range of deprotonation sites were studied. Extensive conformational sampling was performed using density functional theory (DFT) to search the conformational space which is complicated by the flexibility of the amino acid and the $\text{OPO}_3\text{H}_2/\text{OPO}_3\text{H}^-$ groups. Our previous predictions of the GAs of amino acids and peptides^{42,43,44,45} at the correlated G3(MP2) molecular orbital theory level⁵⁶ are in agreement with the experimental values to within about ± 1 kcal/mol and with higher level CCSD(T) calculations extrapolated to the complete basis set limit with additional corrections.^{57,58,59,60,61} G3(MP2) has an advantage over DFT methods with commonly used functionals in terms of reliable predictions for these types of compounds as the correlated molecular orbital methods in G3(MP2) perform better in the prediction of hydrogen bond energies as well as steric non-bonded interactions than do most widely used DFT exchange-correlation functionals.

Theoretical $\text{p}K_a$ 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 calculations⁵⁰ using the COSMO parameterization.⁵¹ The Gibbs free energy for deprotonation in aqueous solution (ΔG_{aq}) at 298 K 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 non-electrostatic 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 $\text{p}K_a$ values in aqueous solution were calculated using Equation 1:

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

where ΔG_{aq} is the solution free energy, R is the gas constant, and $T = 298 \text{ K}$ is the temperature. We report our $\text{p}K_{\text{a}}$ values for the phosphorylated amino acids and amides relative to phosphonoxyacetic acid (1.19) and phosphonoxyacetamide (1.45) respectively to minimize errors in the prediction of the $\text{p}K_{\text{a}}$'s.⁶²

The heats of formation of the amino acids were calculated from atomization energies at the G3(MP2) level and from a set of isodesmic reactions at the G3(MP2) level using the calculated heat of formation of the amino acids and amino acid amides.

Experimental Methods All experiments were performed using a Bruker Daltonics (Billerica, MA, USA) BioApex 7T Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer. Compounds were the L-stereoisomer. All solutions were at $60 \mu\text{M}$ in a solvent of volume ratio of 49.5:49.5:1 $\text{CH}_3\text{OH}:\text{H}_2\text{O}:\text{NH}_4\text{OH}$. The 1% ammonium hydroxide helped to assist with solution-phase deprotonation prior to electrospray ionization (ESI). Analyte solutions were introduced into an Apollo API source using a syringe pump set to deliver $\sim 125 \mu\text{L/hr}$. Solutions were ionized by ESI, with a drying gas temperature of $220 \text{ }^\circ\text{C}$. The ESI needle was grounded, and the capillary entrance and end plate were at a potential of 3.5-4.0 kV for optimal negative ion formation.

Deprotonated phosphorylated ions were isolated by correlated frequency resonance ejection techniques⁶³ and reacted with neutral reference compounds with well-established GAs.⁶⁴ Reference compounds were introduced through a leak valve at constant pressures, which were in the range of $(1.0\text{-}17) \times 10^{-8} \text{ mbar}$. Pressures were measured by a calibrated ion gauge,^{45,65} and the pressure of each reference compound was corrected for its ionization efficiency,⁶⁶ which is determined by polarizabilities calculated with atomic hybrid parameter procedures.⁶⁷

The pseudo-first-order decay of precursor ion intensity as a function of reaction time was utilized to obtain experimental rate constants from which GA values were assigned. To assign GA values, the ratio of the experimental rate constant to the thermal rate capture rate constant^{68,69} provided a reaction efficiency (RE). A RE of 0.269 was used as a break point, where a reaction is considered to become exoergic and a GA value can be assigned. This selection of 0.269 comes from the work of Bouchoux and coworkers,^{70,71,72,73} is known as the “thermokinetic method.” In our past work, the thermokinetic method of obtaining GA values has yielded excellent agreement for experimental and computational GAs for amino acids and small peptides.^{42,44,45,74}

5.3 Results and Discussion The lowest energy G3(MP2) and experimental GAs are given in Table 1 for ten phosphorylated AAs (arginine, aspartic acid, cysteine, glutamic acid, glycine, histidine, lysine, serine, threonine, and tyrosine) as well as their phosphorylated amides. The previous work done by our group on the GAs for the L-common amino acids⁴³ and their amides⁷⁴ has been included for comparison. Excellent agreement is found between the G3(MP2) and the available experimental GAs and will be discussed in further detail below. For phosphoserine, -threonine, and -tyrosine, phosphorylation occurs at the hydroxyl group on the side chain. The amino acids serine and threonine have very similar structures with threonine only differing from serine by the presence of a methyl group on the side chain. Thus, the results of phosphoserine and phosphothreonine will be discussed together.

Phosphoserine (pSer) and Phosphothreonine (pThr): IR spectra, GAs, and hydrogen bonding

For pSer and pThr, two low energy neutral conformers with basically the same structure were predicted (Figure 1). The lowest energy neutral conformer, pSer-1 (pThr-1), has one strong hydrogen bond (~ 1.8 to 1.9 Å) between the $-\text{PO}_3\text{H}_2$ and the $-\text{NH}_2$ groups and a weaker

hydrogen bond (~ 2.5 to 2.6 Å) between the $-\text{NH}_2$ and the $-\text{CO}_2\text{H}$ groups. In the higher energy neutral conformer, pSer-2 (pThr-2), the $-\text{CO}_2\text{H}$ group forms strong hydrogen bonds with both the $-\text{PO}_3\text{H}_2$ and $-\text{NH}_2$ groups (~ 1.8 and 1.9 Å respectively). A weaker hydrogen bond (2.4 Å) is formed between the $-\text{NH}_2$ and the $-\text{PO}_3\text{H}_2$ groups. Deprotonation of pSer and pThr generated six low energy anion conformers/isomers with the sites of deprotonation being the phosphate and/or carboxylate groups. The experimental GAs of pSer and pTyr are 308.5 ± 3.4 and 308.2 ± 3.4 kcal/mol respectively. For pSer, the G3(MP2) GAs of anion structure structures pSer-A, pSer-B, pSer-C, and pSer-D all fall within the experimental error, whereas for pThr, only the G3(MP2) calculated GAs of anion structures pThr-A and pThr-B are in agreement with experiment.

The most stable anion structure, pSer-A and pThr-A, has partial proton transfer between the phosphate and carboxylate groups with the proton closer to the carboxylate, so this is labelled as ‘phosphate’ deprotonation. The higher energy anion structure, pSer-B and pThr-B, is a more folded structure than pSer-A and pThr-A with partial proton transfer again for ‘phosphate’ deprotonation. Anion structure pSer-C and pThr-C is an unfolded ‘phosphate’ deprotonated structure. As anion structures pSer-A (pThr-A), pSer-B (pThr-B), and pSer-C (pThr-C) are all generated from phosphate deprotonation, we predict that the anion observed experimentally by the proton transfer reactions is deprotonated at the phosphate group. Higher energy anion structures pSer-D (pThr-D) and pSer-E (pThr-E) are carboxylate deprotonated structures with pSer-D (pThr-D) exhibiting partial proton transfer between the $-\text{CO}_2^-$ and $-\text{PO}_3\text{H}_2$. The highest energy anion structure pSer-F (pThr-F) is zwitterionic (salt-bridge like) with an $-\text{NH}_3^+$ group sandwiched between the $-\text{PO}_3\text{H}^-$ and $-\text{CO}_2^-$ groups and held together by strong hydrogen bonds. A detailed discussion of the hydrogen bonding in these structures is given below.

Due to the unusual occurrence of a small gas-phase zwitterionic anion, we further optimized the structures of the lowest energy neutral and six anions for pSer and pTyr at the MP2/aug-cc-pVnZ level of theory with $n = D$ and T . The energy differences between the six anion sites are shown in Table 2 for pSer and Table 3 for pTyr. Regardless of the computational method used, structure pSer-B (pThr-B) is predicted to be higher in energy than structure pSer-A (pThr-A) on the free energy scale, although this can be reversed for the enthalpy at the MP2 level. In any case these energy differences at the MP2 level are small, on the order of 0.5 kcal/mol. Structure pThr-A is predicted to be lower in energy than structure pThr-B in terms of free energy and enthalpy by 0.8 and 1.3 kcal/mol respectively at the MP2/aT level. Structure pSer-C is predicted to be 1.1 kcal/mol higher in energy than structure pSer-A at the MP2/aT level and structure pThr-C is predicted to be 3.9 kcal/mol higher in energy than structure pThr-A. Thus, only structures pSer-A (pThr-A), pSer-B (pThr-B), and pSer-C are relevant in terms of their energies. All three anion structures pSer-A (pThr-A), pSer-B (pThr-B), and pSer-C are ‘phosphate’ deprotonated. The most interesting result was the decrease in the energy difference between the higher energy zwitterionic anion structure pSer-F (pThr-F) and structure pSer-A (pThr-A), from 4.3 to 2.5 kcal/mol for pSer and from 5.3 to 3.6 kcal/mol for pThr at the MP2/aT levels of theory. The relative energy differences at the MP2/aD only differ from the energy differences already given at the MP2/aT levels of theory by 0.1 kcal/mol. Thus the use of the lower levels of geometry optimization in the G3(MP2) composite calculations is not significantly biasing the predictions.

Maitre and coworkers³⁶ compared their IRMPD spectrum of deprotonated [pSer-H]⁻ and [pThr-H]⁻ with the calculated spectra of their lowest energy PO₃H⁻, COO⁻, and zwitterionic anions. Their three structures correspond to our anion structures pSer-B (pThr-B), pSer-E (pThr-

E), and pSer-F (pThr-F), respectively. They did not report the deprotonated phosphate anion structures pSer-A (pThr-A) and pSer-C (pThr-C). We calculated the IR spectra for the six [pSer-H]⁻ and [pThr-H]⁻ anion structures pSer-A (pThr-A) to pSer-F (pThr-F) at the B3LYP/aug-cc-pvdz level. These calculated spectra as well as the experimental IRMPD spectrum for deprotonated [pSer-H]⁻ and [pThr-H]⁻ obtained by Maitre and coworkers are given in Figure 2. Tables 4 and 5 contain the position of the experimental bands, the calculated frequencies of PO₃H⁻ anion structures pSer-A (pThr-A), pSer-B (pThr-B), and pSer-C (pThr-C) as well as that of the zwitterionic anion structure pSer-F (pThr-F) at the B3LYP/aug-cc-pvdz level, the ratio between the experimental and calculated frequencies, and the previously proposed assignments of the IRMPD bands by Maitre and coworkers for deprotonated [pSer-H]⁻ and [pThr-H]⁻. Maitre and coworkers suggested that the experimental IRMPD spectrum corresponds to their lowest energy PO₃H⁻ anion, our structure pSer-B (pThr-B). However, the calculated IR spectra for anion structures pSer-A (pThr-A) and pSer-C (pThr-C) also show good agreement with the IRMPD spectrum. Additionally, we note that the calculated IR spectra for the zwitterionic anion structure pSer-F (pThr-F) shows good agreement. Because the relative energy differences between anion structures pSer-F and pSer-A is only ~ 2.5 kcal/mol at the MP2/aT level of theory, the anionic zwitterion structure pSer-F could be present at about 1%. We agree with Maitre and coworkers that the experimental spectra of deprotonated [pSer-H]⁻ and [pThr-H]⁻ are composed of mostly phosphate deprotonated anions. However, our calculations show that these PO₃H⁻ anions can be present as several different low energy conformers. At equilibrium in the gas phase at 298 K, the ratio for structures pSer-A, pSer-B, and pSer-C relative to pSer-A is 1.0:0.36:0.16 at the MP2/aT level whereas the ratio for structures pThr-A, pThr-B, and pThr-C relative to pThr-A is 1.0:0.11:0.01. Thus, the experimental IRMPD spectra of deprotonated deprotonated

[pSer-H]⁻ and [pThr-H]⁻ are probably composed of a mixture of PO₃H⁻ anion conformers.

To further study the formation and stabilization of the anionic zwitterion structure pSer-F (pThr-F), the relevant hydrogen bonding was examined at the different computation levels and the bond lengths are given for the six pSer anion structures in Table 6. The atom labels for the anions are given in Figure 1. For the six pThr anion structures, the table of hydrogen bonding values is given in the Table 7. Because the structures of the pThr anions are basically the same as the pSer anions, the hydrogen bonding results are essentially the same following all of the same trends and having approximately the same bond distances. Thus, only the pSer hydrogen bond distances are discussed in detail.

In general, the results show that there is not much change in key bond lengths at the B3LYP, MP2/6-31G(*d*) from G3(MP2), MP2/aD, and MP2/aT levels of theory. For the pSer anion structures pSer-A, pSer-B, pSer-D, and pSer-E, two hydrogens are bound to the nitrogen (H1 and H18) and both have a distance of 1.01 to 1.03 Å, respectively. A zwitterionic anion structure cannot be formed in any of these cases as the third hydrogen needed to form the zwitterion is shared between the -CO₂⁻ and the -PO₃H⁻ groups. Anion structures pSer-A and pSer-D are similar as are anions pSer-B and pSer-E with the major difference being the rotation of the phosphate group to allow for hydrogen bonding with the -NH₂ group in pSer-B and pSer-E. Anion structures pSer-A and pSer-B are ‘phosphate’ deprotonation and anion structures pSer-D and pSer-E are ‘carboxylate’ deprotonation. In all four cases, there is one hydrogen bond formed between the hydrogen H1 in the -NH₂ group and oxygen O5 in the carboxylate group. For all four anions, there is no O5⋯H18 hydrogen bond ($r \sim \geq 3$ Å). The (O5⋯H1) hydrogen bond in pSer-D and pSer-E is ~ 2 to 2.2 Å showing that the conformational difference between the structures has no effect on the hydrogen bond distance. The (O5⋯H1) hydrogen bond

distance for pSer-A and pSer-B is weaker, ~ 2.3 and 2.6 \AA respectively. The slight increase in the hydrogen bond distance in pSer-B is likely due to the presence of a hydrogen bond of O5 with H8. In pSer-C, the $-\text{NH}_2$ group is stabilized between the $-\text{CO}_2\text{H}$ and the $-\text{PO}_3\text{H}^-$ groups.

Structure pSer-F is formed from the additional hydrogen bond between the nitrogen and H8. Structure pSer-C is similar to structure pSer-F with the only difference being the transfer of the proton H8 from O5 in the carboxylic acid group to the nitrogen to form the $-\text{NH}_3^+$ group. In both pSer-C and pSer-F structures, H1 is bound to the nitrogen with an undistorted bond distance of $\sim 1.02 \text{ \AA}$ and is not involved in hydrogen bonding with the $-\text{PO}_3\text{H}^-$ and $-\text{CO}_2^-$ groups, ($r \geq 3.1 \text{ \AA}$). The (N-H18) bond is elongated with respect to the (N-H1) bond, ~ 1.04 and 1.09 \AA for C and F respectively showing partial transfer of the proton to the $-\text{PO}_3\text{H}^-$ group. In structure pSer-C, the (O5-H8) and (N \cdots H8) bond distances are ~ 1.01 and 1.73 \AA showing that the proton is bound to the carboxylic acid group whereas in structure pSer-F the (O5 \cdots H8) and (N-H8) bond distances are ~ 1.71 and 1.06 \AA . This shows that the proton is partially transferred to the $-\text{CO}_2^-$ group but is much closer to the nitrogen forming the zwitterionic structure.

Phosphotyrosine (pTyr): IR spectra, GAs, and hydrogen bonding For pTyr, two low energy neutral conformers and six corresponding low energy anion conformer/isomers were found (Figure 3). The lowest energy neutral, pTyr-1, has only one strong hydrogen bond between the $-\text{CO}_2\text{H}$ and the $-\text{NH}_2$ groups ($r(\text{N}\cdots\text{H}(\text{O})) = 1.9 \text{ \AA}$). In the higher energy neutral, pTyr-2, the hydrogen in the $-\text{CO}_2\text{H}$ group is rotated away from the nitrogen, so the $-\text{NH}_2$ group rotates to form a weaker hydrogen bond with the $-\text{CO}_2\text{H}$ group ($r(\text{N}(\text{H})\cdots\text{O}) = 2.6 \text{ \AA}$). Deprotonation of pTyr has an experimental GA of $<307.5 \text{ kcal/mol}$ as shown in Table 1. Only our lowest energy anion structure pThr-A is in agreement with the experimental value with a calculated GA of 306.7 kcal/mol at the G3(MP2) level. The next lowest energy anion structure pTyr-B has a

calculated GA of 310.7 kcal/mol at the G3(MP2) level (Figure 3). The higher energy GAs for the [pTyr-H]⁻ anions are given in Table 8.

The most stable anion structure pThr-A is a carboxylate deprotonated structure. The benzene ring distorts to allow the -CO₂⁻ group to hydrogen bond with the -PO₃H₂ group. Both of the hydrogens in the phosphate group form strong hydrogen bonds with one oxygen in the carboxylate group ($r(\text{O}(\text{H})\cdots\text{O}) = 1.6$ and 1.7 \AA). Additionally, this oxygen forms a weaker hydrogen bond with the NH₂ ($r(\text{N}(\text{H})\cdots\text{O}) = 2.4 \text{ \AA}$). Anion structure pThr-B is similar to pThr-A except that the phosphate group is deprotonated. The -CO₂H group forms one strong hydrogen bond with the -PO₃H⁻ group (1.7 \AA) and maintains the weaker hydrogen bond (2.4 \AA) with the -NH₂ group found in anion structure pThr-A. Anion structure pThr-C is an unfolded phosphate deprotonated structure, so there is no hydrogen bonding present between the -CO₂H and -PO₃H⁻ groups. Both -PO₃H⁻ anion structures pThr-B and pThr-C are 4.0 kcal/mol higher in energy than the lowest energy structure pThr-A. Anion structures pThr-D and pThr-E are zwitterionic with -NH₃⁺, -PO₃H⁻, and -CO₂⁻ groups. Anion structure pThr-D is a folded zwitterionic structure stabilized through hydrogen bonding with the -PO₃H⁻ and -CO₂⁻ groups and is predicted to be 11.0 kcal/mol higher in energy than anion structure pThr-A. Anion structure pThr-E is an unfolded zwitterionic anion that is only stabilized through hydrogen bonding with the -CO₂⁻ group. This structure could only be optimized at the B3LYP/6-31+G(*d*) and HF/6-31G(*d*) levels of theory. Thus the G3(MP2) method had to be adjusted to obtain the calculated GA value shown in Figure 3. Since this structure could not be optimized at the MP2/6-31G(*d*) level of theory, the optimized HF geometry was used to perform the single point QCISD(T) and MP2/Large calculations. The approximate G3(MP2) GA value for anion pThr-E is 19.3 kcal/mol higher in energy than anion pThr-A. Anion structure pThr-F is an unfolded carboxylate anion and is 20.3

kcal/mol higher in energy than pThr-A at the G3(MP2) level.

Maitre and coworkers³⁶ measured the IRMPD spectrum of deprotonated phosphotyrosine. They compared the experimental spectrum to their calculated spectra for their lowest energy PO_3H^- , COO^- , and zwitterionic anions. These three structures correspond to our anion structures pThr-C, pThr-E, and pThr-F respectively. They did not find either of the folded anion structures pThr-A or pThr-B which have distorted benzene rings and hydrogen bonding between the phosphate and carboxylate groups. Based on our experimental GA determined by proton transfer reactions, anion structure pThr-A is the only structure that matches the experimental GA. The calculated IR spectra for the $[\text{pTyr-H}]^-$ anion structures pThr-A to pThr-D and pThr-F at the B3LYP/aug-cc-pvdz level as well as Maitre and coworkers experimental IRMPD spectrum are given in Figure 4. Anion structure pThr-E is not shown as it could not be optimized at this level of theory. A table containing the position of the experimental bands, the calculated frequencies of the anion structures pThr-A, pThr-B, pThr-C, and pThr-D at the B3LYP/aug-cc-pvdz, the ratio between the experimental and calculated frequencies, and the previously proposed assignments of the IRMPD bands by Maitre and coworkers for $[\text{pTyr-H}]^-$ is given in the Table 9. Maitre and coworkers suggest that the experimental IR spectrum corresponds to their lowest energy PO_3H^- anion, which is our anion structure pThr-C. Whereas their calculated spectrum does match the IRMPD spectrum relatively well, we have shown that anion structures pThr-B to pThr-F do not match the experimental GA. This suggests that the IRMPD experiment is missing the ground state distorted benzene structure. In our previous work⁴³ on the amino acid acidities, we showed that the IRMPD experiments⁷⁵ did not yield the lowest energy structure for cysteine. Our calculated G3(MP2) results as well as experiments using the extended kinetic and gas-phase equilibrium methods⁴⁸ found that the ground state anion resulted from deprotonation of the side

chain –SH group. In contrast, the IRMPD results showed that the lowest energy anion was generated from deprotonation of the carboxylic acid. We suggested that the complications in the interpretation of the IRMPD spectra could originate from the partial proton transferring we predicted between the side chain thiol and carboxylic acid groups. In the case of phosphotyrosine, the conditions used to prepare the anion in the IRMPD experiment are likely to be biased against the folded anion.

pSer, pThr, and pTyr amide GAs The phosphorylated amides were formed by converting the carboxylic acid functional group (CO₂H) into an amide. We have used this substitution previously to better understand side chain acidity, as well backbone acidity, as the proton can no longer be lost from the carboxylic acid to form the acid (CO₂⁻).⁷⁴ The optimized geometries of the low energy neutrals and corresponding anions for pSer, pThr, and pTyr amides and their relative energies at the G3(MP2) level are given in Figure 5. The experimental and lowest energy calculated GA for each phosphorylated amino acid amide are given in Table 1. The calculated GAs of the higher energy anions are given in Table 8. Because the CO₂⁻ anion can no longer be formed, the lowest energy GAs result from deprotonation of the phosphate group.

For pSer amide, the lowest energy neutral, pSer(amide)-1, the amide group is involved in two hydrogen bonds. It forms one hydrogen bond with the NH₂ group (2.18 Å) and one strong hydrogen bond with the phosphate group (1.75 Å). In the higher energy neutral structure, pSer(amide)-2, the –NH₂ group is involved in two hydrogen bonds, one with the –C(O)NH₂ group (2.15 Å) and one with the –PO₃H₂ group (1.70 Å). Both of the PO₃H⁻ anion structures pSer(amide)-A and pSer(amide)-B are stabilized by two hydrogen bonds with A being ~ 3 kcal/mol more stable than pSer(amide)-B. The experimental GA of pSer amide is 314.0 ± 4.3 kcal/mol and is in excellent agreement with the predicted value for our lowest energy anion

structure pSer(amide)-A. Anion structure pSer(amide)-C is the zwitterionic anion and is substantially higher in energy by 16 kcal/mol. Thus, a low energy zwitterionic anion is only seen with phosphoserine as there is no second anionic site in the amide.

For pThr amide, a single low energy neutral and three anions were found. The neutral and higher energy anion structures contain three hydrogen bonds whereas the lowest energy anion structure only contains two. The experimental GA of pThr amide is 315.3 ± 4.3 kcal/mol and is in excellent agreement with both of our calculated anion structures pThr(amide)-A and pThr(amide)-B, which differ in energy by 0.6 kcal/mol. As for pSer amide, anion structure pThr(amide)-C is only shown to further demonstrate that the zwitterionic anion will not form with the amides.

For pTyr amide, two low energy neutrals and only one low energy anion were found. The lowest energy neutral pTyr(amide)-1, is an unfolded structure with only one hydrogen bond between the $-C(O)NH_2$ and $-NH_2$ groups (2.17 Å) whereas the higher energy neutral structure, pTyr(amide)-2, is folded with two hydrogen bonds, one between the $-C(O)NH_2$ and $-NH_2$ groups (2.24 Å) and one between the $-C(O)NH_2$ and $-PO_3H_2$ groups (1.82 Å). The experimental GA of pTyr amide is <307.5 kcal/mol, consistent with our calculated GA of 307.0 kcal/mol.

Other phosphorylated amino acid and amide GAs The most stable neutral structures of phosphorylated arginine, aspartic acid, cysteine, glutamic acid, glycine, histidine, and lysine and their anions are shown in Figure 6. The corresponding phosphorylated amides were also studied for these amino acids and the most stable neutral and anion structures are shown in Figure 7. For phosphorylated amino acids and their amides, the lowest energy GAs are given in Table 1 and the higher energy GAs are given in Tables 8 and 10 for the acids and amides respectively.

Phosphorylated glycine was studied as glycine is the smallest amino acid with

phosphorylation occurring at the terminal NH_2 group. For pGly only one lowest energy neutral and anion were found, whereas for pGly(amide), two low energy neutrals and one lowest energy anion were found. For pGly, the lowest energy anion is deprotonated at the phosphate group with partial proton transfer between the $-\text{PO}_3\text{H}^-$ and $-\text{CO}_2\text{H}$ groups. For pGly(amide) the lowest energy anion is also phosphate deprotonation, but there is no partial proton transfer. The two low energy anions pGly(amide)-1 and pGly(amide)-2 are within 1.4 kcal/mol of one another and contain only one strong hydrogen bond. The GA of pGly(amide) is predicted to be 6.8 kcal/mol higher in energy than pGly.

In phosphoarginine, phosphorylation occurs at the NH_2 group on the side chain guanidine group. The lowest energy neutral, pArg-1, is a zwitterion generated from the transfer of a hydrogen from the phosphate group to the guanidine side chain to generate $-\text{PO}_3\text{H}^-$ and $-\text{NH}_2^+$ groups respectively. The pArg-1 structure is stabilized by strong hydrogen bonds between the $-\text{CO}_2\text{H}$ and $-\text{NH}$ groups (1.98 Å) and the $-\text{CO}_2\text{H}$ and $-\text{PO}_3\text{H}^-$ groups (1.72 Å). The next higher energy neutral isomer pArg-2 is also a zwitterion with $-\text{NH}_2^+$ and $-\text{N}^-$ groups on the side chain and is 10.8 kcal/mol higher in energy than pArg-1. The pArg-2 structure is stabilized by strong hydrogen bonds between the $-\text{CO}_2\text{H}$ and N^- groups (1.73 Å) and the $-\text{CO}_2\text{H}$ and $-\text{PO}_3\text{H}_2$ groups (1.84 Å). The non-zwitterion neutral pArg-3 is 13.4 kcal/mol higher in energy than pArg-1 and is stabilized by only one strong hydrogen bond between the $-\text{CO}_2\text{H}$ and $-\text{PO}_3\text{H}_2$ groups (1.80 Å). The lowest energy anion structure pArg-A is an elongated zwitterion with a $-\text{NH}_2^+$ sandwiched between the $-\text{PO}_3\text{H}^-$ and $-\text{CO}_2^-$ groups with partial proton transfer between the $-\text{CO}_2^-$ and NH groups. Anion structures pArg-B and pArg-D are folded non-zwitterion structures generated from phosphate deprotonation and are 0.9 and 1.4 kcal/mol higher in energy than pArg-A respectively. Anion structure pArg-C is a zwitterion with a similar structure to pArg-A that is 1.2

kcal/mol higher in energy. The main difference is the rotation of the $-\text{PO}_3\text{H}^-$ group away from the $-\text{NH}_2^+$ group resulting in a structure with a fewer number of hydrogen bonds. The highest energy anion structure pArg-E is another example of a zwitterionic anion with $-\text{PO}_3\text{H}^-$, $-\text{NH}_2^+$, and $-\text{N}^-$ groups and is 1.7 kcal/mol higher in energy than pArg-A. All of the anion structures are held together by multiple strong hydrogen bonds between the phosphate, carboxylate and nitrogen groups.

For phosphoarginine amide, all three neutral structures follow the same trend as predicted for pArg. The lowest energy neutral pArg(amide)-1 and pArg(amide)-2 are zwitterions and correspond to structures pArg-1 and pArg-2 respectively. The major differences arise from their relative energies with pArg(amide)-2 only 5.7 kcal/mol higher in energy than pArg(amide)-1. Neutral pArg(amide)-3 is non-zwitterionic and the highest energy neutral, 6.5 kcal/mol higher in energy than pArg(amide)-1. Thus, even though the structures are very similar, the energy differences between the neutral amides are approximately half of what was predicted for the acids. Deprotonation of the phosphate group generates all of the low energy anion structures pArg(amide)-A to pArg(amide)-E. Only one structure, pArg(amide)-D, is a zwitterionic anion with $-\text{PO}_3\text{H}^-$, $-\text{NH}_2^+$, and $-\text{N}^-$ groups and is 1.4 kcal/mol higher in energy than pArg(amide)-A.

Phospho-aspartic and -glutamic acids are phosphorylated at their side chain carboxylic acid groups. For phosphoaspartic acid and its corresponding amide, three low energy neutrals and two low energy anions are predicted. The GAs for pAsp and pAsp(amide) are predicted to be the same, less than 0.2 kcal/mol apart. The lowest energy neutrals for pAsp are within 1.2 kcal/mol of each other whereas for pAsp(amide) they are even closer in energy, within 0.5 kcal/mol. All of the neutral structures for pAsp and pAsp(amide) contain four to five hydrogen bonds resulting in very stable structures. For Asp and Asp(amide), the low energy anions are

generated from phosphate deprotonation. The lowest energy anion structure for the acid, pAsp-A, is only 0.7 kcal/mol lower in energy than the higher energy anion structure pAsp-B whereas for the amide, anion structure pAsp(amide)-B is 1.7 kcal/mol higher in energy than anion structure pAsp(amide)-A. The pAsp anions contain three hydrogen bonds and the pAsp(amide) anions contain four. For pGlu, four neutrals and one anion were predicted whereas for pGlu(amide), two low energy neutrals with relatively no energy difference and one anion were predicted. For pGlu, the four neutrals are within 1.7 kcal/mol of each other with pGlu-2 is predicted to be 0.6 kcal/mol higher on the free energy scale. Unlike what was predicted for phosphoaspartic acid, the predicted GA of pGlu(amide) is 4.6 kcal/mol higher in energy than pGlu. For both pGlu and pGlu(amide), the lowest energy anions are generated from phosphate deprotonation with the resulting structures having three and four hydrogen bonds respectively.

For phosphocysteine, phosphorylation occurs at the thiol group on the side chain. For pCys, three low energy neutrals and five low energy anions were predicted. These low energy anions were generated from phosphate deprotonation and are all within ~ 1 kcal/mol of the lowest energy anion structure pCys-A. The lowest energy anion structure pCys-A does not contain the shortest hydrogen bond distance or the highest number of hydrogen bonds. Only two strong hydrogen bonds are predicted between the $-\text{PO}_3\text{H}^-$, $-\text{NH}_2$ and $-\text{CO}_2\text{H}$ groups (1.67 and 1.95 Å). For pCys(amide), two low energy neutrals and four low energy anions were found. All four anions are generated by phosphate deprotonation. Structure pCys(amide)-B is only 0.2 kcal/mol higher in energy than the lowest energy structure pCys(amide)-A whereas structure pCys(amide)-C is 1.3 kcal/mol higher in energy than structure pCys(amide)-A. Both the lowest energy neutral, pCys(amide)-1 and the lowest energy anion structure pCys(amide)-A contain four hydrogen bonds between the phosphate, amide, and NH_2 groups.

For phosphohistidine, phosphorylation occurred at the NH in the imidazole ring. As in our previous work,⁴³ we studied the π and τ tautomers. For pHis- τ , three low energy neutrals and two low energy anions were predicted whereas for pHis- τ (amide), three low energy neutrals and anions were predicted. For both pHis- π and pHis- π (amide), two low energy neutrals and anions were found. In both the acid and amide cases, the GAs of the two tautomers, pHis- τ (pHis- τ (amide)) and pHis- π (pHis- π (amide)), are predicted to be within 1 kcal/mol. The lowest energy anions for pHis- π (amide), pHis- τ , and pHis- τ (amide) result from phosphate deprotonation. In the lowest energy anion structure for pHis- π , pHis- π -A, the anion fragments into a carboxylated amine and a phosphorylated imidazole ring connected by strong hydrogen bonds. Two partial proton transfers occur between the $-\text{PO}_3\text{H}_2$ group and the $-\text{CO}_2^-$ group. The higher energy anion structure pHis- π -B is a phosphate deprotonated structure that is 1.4 kcal/mol higher in energy in terms of free energy and 2.0 kcal/mol lower in energy in terms of enthalpy.

In phospholysine, phosphorylation occurs at the side chain $-\text{NH}_2$ group. For pLys, two low energy neutrals and three low energy anions were predicted whereas for pLys(amide), three low energy neutrals and two low energy anions were predicted. For pLys, the two neutrals have similar structures and only differ in energy by 0.3 kcal/mol. For pLys(amide), Lys(amide)-1 is lowest energy neutral in terms of free energy by 0.5 kcal/mol. Neutral structure pLys(amide)-3 is the lowest energy neutral in terms of enthalpy. However, all three neutrals are within 1 kcal/mol of one another in terms of enthalpy and free energy which is within the predictive accuracy of the G3(MP2) method for such relative energies. The GA of pLys(amide) is predicted to be 1.3 kcal/mol higher in energy than pLys. For both pLys and pLys(amide), the lowest energy anions are generated from phosphate deprotonation. For pLys(amide), both anion structures pLys(amide)-A and pLys(amide)-B are the same energy in terms of enthalpy, but pLys(amide)-A

is lower in terms of free energy by 0.5 kcal/mol. For pLys, the anion structures pLys-A and pLys-B only differ by 0.1 kcal/mol with pLys-B having partial proton transfer between the $-\text{PO}_3\text{H}^-$ and $-\text{CO}_2\text{H}$ groups. Anion structure pLys-C is a zwitterion with a NH_3^+ sandwiched between the $-\text{PO}_3\text{H}^-$ and $-\text{CO}_2^-$ groups. Structure pLys-C has partial proton transfer between the $-\text{PO}_3\text{H}^-$ and NH_3^+ groups and is predicted to be 4.1 kcal/mol higher in energy than pLys-A.

Comparison of Acidities of Phosphorylated Species with Non-Phosphorylated Species In

summary, the GAs of the phosphorylated amino acids are 13 to 27 kcal/mol more acidic than their corresponding non-phosphorylated amino acids. The smallest change in acidity of a phosphorylated side chain is predicted for the Asp/pAsp and Arg/pArg pairs (~ 13 kcal/mol) and the largest change is predicted for the His/pHis and Tyr/pTyr pairs (~ 23 kcal/mol). The difference between Gly and pGly is 27 kcal/mol, but this is for backbone phosphorylation. pSer, pThr, pGlu, and pLys are all ~ 17 kcal/mol more acidic than their corresponding non-phosphorylated amino acids. The phosphohistidine τ and π tautomers have almost equivalent GAs, while non-phosphorylated amino acids have ~ 15 kcal/mol difference with the π tautomer being the most acidic. The phosphorylated amino acids are also 15 to 20 kcal/mol more acidic than phosphoric acid (GA = 322.2 kcal/mol) and more similar to sulfuric acid (GA = 304.6 kcal/mol), a strong gas-phase acid.^{76,77} Specifically pAsp and pAsp(amide) have acidities equivalent to sulfuric acid, whereas pGlu and pHis are more acidic than sulfuric acid. In general, the phosphorylated amino acids are 1 to 7 kcal/mol more acidic than their corresponding amides. For pTyr vs. pTyr(amide) and pAsp vs. pAsp(amide), the acids and amides have essentially the same acidities within < 0.5 kcal/mol. For pArg vs. pArg(amide) and pLys vs. pLys(amide), the energy differences are only slightly larger, ~ 1 kcal/mol. The largest energy difference is predicted between pGly and pGly(amide) where the amide is 6.8 kcal/mol less acidic than the

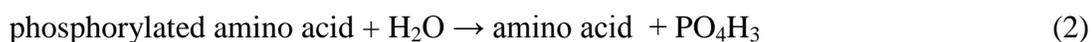
acid. This is not surprising as the change here is for backbone phosphorylation. The phosphorylated amino acid amides are 22 to 30 kcal/mol more acidic than their corresponding non-phosphorylated amino acid amides with pLys(amide) and pGly(amide) being ~ 35 kcal/mol more acidic than Lys(amide) and Gly(amide) respectively.

Phosphorylated amino acid and corresponding amide pK_a's in aqueous solution The calculated pK_a's of the phosphorylated amino acids and corresponding amides are summarized in the Table 8. The predicted pK_a's were obtained using the self-consistent reaction field approach with the COSMO parameterization. The pK_a values for the phosphorylated acids and amides are reported relative to phosphonoxyacetic acid (1.19) and phosphonoxyacetamide (1.45) respectively to minimize errors in the prediction of the pK_a's.⁶²

The neutrals that were studied, except for pArg, are non-zwitterionic in solution. Thus, the predicted pK_a's could be used to derive the difference between the energies of the zwitterionic and non-zwitterionic forms in solution if the experimental pK_a's of the zwitterions are known as they are for the common amino acids following our previous work.⁴³ The difference in the pK_a values for the zwitterionic and nonzwitterionic neutrals is predicted to be 1.9 favoring the zwitterion. Since the anion is the same when a proton is lost, we can estimate that the zwitterion is only 2.6 kcal/mol more stable than the non-zwitterionic structure in aqueous solution at 298 K. In general, the pK_a values for the phosphorylated amino acids and their corresponding amides are within 3 pK units. Neutral pHis- π is more stable in solution than neutral pHis- τ for both the acid and the amide and the pK_a difference is in the above range. Neutral pCys is significantly more solvated than neutral pCys(amide) so the acid's pK_a is ~ 7 pK units higher than that of the amide. pSer, pGlu(amide), pAsp(amide), pCys(amide), pHis- τ (amide), and pHis- π (amide) are all predicted to be very acidic with pK_a values ≤ 0 . For pSer,

pThr, and pTyr, the most acidic anion in solution is the zwitterion. This provides further support for the importance of zwitterions in phosphorylated amino acids. For pArg and pArg(amide), the neutrals are zwitterionic and have larger solvation energies so a high pK_a of around 10 is predicted.

Phosphorylated and corresponding amide gas-phase heats of formation The heats of formation of the neutral phosphorylated amino acids and amides were calculated from the atomization energies at the G3(MP2) level and from a set of isodesmic reactions at the G3(MP2) level using the calculated heat of formation of the amino acids and amino acid amides (Reaction 2).



This leads to the energy expression given in Equation 3:

$$\Delta H_f(\text{reaction}) = \Delta H_f(\text{AA}) + \Delta H_f(\text{PO}_4\text{H}_3) + \Delta H_{\text{rxn}}(\text{G3(MP2)}) - \Delta H_f(\text{H}_2\text{O}) \quad (3)$$

where the heat of formation of water ($\Delta H_f = -57.8 \pm 0.01$) is taken from experiment.⁷⁸ For the reaction with glycine, the experimental heat of formation was used ($\Delta H_f = -93.3 \pm 1.1$)⁷⁹ which is in excellent agreement with the CCSD(T)/CBS heat of formation calculated using the Feller-Peterson-Dixon (FPD) method ($\Delta H_f^{298} = -92.6$ kcal/mol).⁴³ The heats of formation for the remaining amino acids and the amino acid amides were taken from our previous work.^{43,74} A table of the heats of formation of these amino acid amides is given in the Appendix. The heats of formation of the neutral phosphorylated amino acids and amides are given in Table 10. In every case except for pTyr(amide), the heats of formation calculated from the atomization energies and isodesmic reactions are within 1 kcal/mol of each other for each phosphorylated amino acid and corresponding amide. In pTyr(amide), the difference is slightly larger, 2.4 kcal/mol. These are the first available reliable values for these quantities.

5.4 Conclusion The gas-phase acidities for ten phosphorylated amino acids and their amides

were predicted, providing the first reliable set of these values. The calculated GAs at the G3(MP2) level for pSer, pThr, and pTyr and their amides were compared to GAs determined by proton transfer reaction in a mass spectrometer. Excellent agreement was found between the predicted GAs and the experimental values. The lowest energy anions for pSer and pThr are generated from phosphate deprotonation. For pTyr, the most stable anion is a carboxylate deprotonated structure in which the benzene ring is distorted to allow for strong hydrogen bonding between the carboxylate and phosphate groups. The infrared spectra of the anions of pSer, pThr, and pTyr were calculated and compared to experimental IRMPD spectra. For deprotonated $[\text{pSer-H}]^-$ and $[\text{pThr-H}]^-$ good agreement is found between the experimental spectra and the calculated spectra for our lowest energy anion structure. For pTyr, good agreement is not found between the calculated and experimental spectra with our lowest energy anion and instead corresponds to a higher energy phosphate deprotonated structure. However, the experimental and predicted GAs are in good agreement with each other for the lowest energy structure pTyr-A of the anion and the calculated GA for structure pTyr-C is not in good agreement with the experimental GA. This suggests that the conditions used to generate the ion $[\text{pTyr-H}]^-$ from which the IRMPD spectrum is obtained did not generate the lowest energy structure.

In general, the phosphorylated amino acids are 13 to 27 and 1 to 7 kcal/mol more acidic than their corresponding non-phosphorylated amino acids and phosphorylated amino acid amides respectively. This energy difference between the phosphorylated amino acids and their amides differs from what was found for the simple amino acids and shows the importance of strong hydrogen bonding between the phosphate group and the carboxylate groups in the anions. The phosphorylated amino acids are 15 to 20 kcal/mol more acidic than phosphoric acid and can

become similar to sulfuric acid, a strong gas phase acid. In general, the pK_a values for the phosphorylated amino acids and their corresponding amides are within 3 pK units. The heats of formation of the neutral phosphorylated amino acids and amides calculated from the G3(MP2) atomization energies and an isodesmic reaction are within 1 kcal/mol of one another except for pTyr(amide) which is slightly large, ~ 2.5 kcal/mol.

In the gas phase, amino acids tend to adopt a neutral structure whereas they adopt a zwitterionic structure in solution.⁴ A number of experimental and theoretical studies have been done on the glycine zwitterion and have shown that it does not exist in the gas phase as it is much higher in energy and unstable.^{80,81,82} Arginine has also been the source of controversy over the existence of a gas-phase anion with evidence on both sides of the argument.^{83,84} In 2000, Kass and coworkers⁸⁵ performed photodetachment experiments on small zwitterionic anions in the gas phase. They showed that the attractive forces experienced by the negatively charged groups were greater than the repulsive forces leading to the zwitterion having a higher binding energy than the non-zwitterion. For pSer and pThr, we see the potential for a gas-phase deprotonated zwitterion to form. Additionally, we see that in pArg, the lowest energy neutral and anion prefers a zwitterionic structure whereas in pArg(amide), only the neutral prefers to be a zwitterion.

Acknowledgement This research is supported by NSF under CHE-1308348. D. A. Dixon thanks the Robert Ramsay Fund from the University of Alabama for its partial support.

Table 5.1. Lowest Energy Gas-Phase Acidities at the G3(MP2) level in kcal/mol.

Compound	ΔG_{298}^a expt	ΔH_{298}	ΔG_{298}	ΔG_{298} expt	ΔH_{298}	ΔG_{298}
	L-Amino Acid ⁴³			Phosphorylated Amino Acid		
serine	325.8±3.0 ⁴⁶	332.6	325.7	308.5 ± 3.4	315.7	309.2
threonine	325.2±3.0 ⁴⁶	332.3	324.8	308.2 ± 3.4	314.2	308.3
tyrosine	329.5±3.0 ⁴⁶ 332.5±1.5 ⁴⁷	338.3	330.4	<307.5	314.5	306.7
arginine	323.2±2.3 ⁴⁶	330.1	322.0	-	317.8	309.1
aspartic acid	315.3±3.3 ⁴²	322.4	315.4	-	308.8	302.4
cysteine	326.0±3.0 ⁴⁶ 327.5±2.1 ⁴⁸	334.4	327.1	-	312.8	306.5
glutamic acid	318.2±3.7 ⁴²	321.9	316.4	-	305.9	299.6
glycine	335.1±3.0 ⁴⁶ 335.5±2.0 ⁸¹	342.9	335.3	-	315.0	308.3
histidine	324.1±3.0 ⁴⁶	328.9	321.4	-	305.6	298.5
lysine	330.5±3.0 ²	336.9	329.9	-	319.6	312.5
	Amino Acid Amide ⁷⁴			Phosphorylated Amino Acid Amide		
serine	343.3±3.2	349.3	342.1	314.0 ± 4.3	320.4	313.2
threonine	342.8±3.2	349.2	341.7	315.3 ± 3.3	321.5	314.2
tyrosine	336.4±2.7	344.2	336.5	<307.5	314.9	307.0
arginine	342.9±3.2	347.3	339.4	-	317.5	310.2
aspartic acid	326.5±3.6 ⁴²	332.9	325.9	-	310.2	302.2
cysteine	335.6±3.8	340.1	332.9	-	317.8	310.3
glutamic acid	325.7±4.8 ⁴²	331.7	325.7	-	312.1	304.2
glycine	-	359.7	351.5	-	322.3	315.1
histidine	330.8±4.8	336.0	328.8	-	309.7	303.0
lysine	348.3±2.6	355.8	347.5	-	319.4	313.8

^a $\Delta G_{298} = GA$.

Table 5.2. Calculated Acidities in kcal/mol of Phosphoserine at Different Computational Levels. All with respect to neutral 1.

Amino Acid	Prop	B3LYP ^a	G3(MP2)	MP2/aD ^b	MP2/aT ^c
pSer PO ₃ H ⁻ [A]	ΔH	313.8	315.7	312.6	313.6
	ΔG	307.1	309.2	306.0	307.1
pSer PO ₃ H ⁻ [B]	ΔH	315.6	316.0	312.1	313.2
	ΔG	310.6	310.5	306.5	307.7
pSer PO ₃ H ⁻ [C]	ΔH	315.5	317.9	314.1	315.4
	ΔG	308.8	310.7	307.0	308.2
pSer COO ⁻ [D]	ΔH	316.1	317.5	313.6	314.2
	ΔG	309.9	311.7	307.8	308.4
pSer COO ⁻ [E]	ΔH	316.7	318.0	313.6	314.4
	ΔG	311.5	312.8	308.4	309.2
pSer zwitt [F]	ΔH	316.3	320.1	315.0	316.2
	ΔG	310.1	313.5	308.4	309.6
A	ΔΔH	0	0	0	0
A	ΔΔG	0	0	0	0
B	ΔΔH	1.8	0.3	-0.5	-0.4
B	ΔΔG	3.5	1.3	0.5	0.6
C	ΔΔH	1.7	2.2	1.5	1.8
C	ΔΔG	1.7	1.5	1.0	1.1
D	ΔΔH	2.3	1.8	1.0	0.6
D	ΔΔG	2.8	2.5	1.8	1.3
E	ΔΔH	2.9	2.5	1.0	0.8
E	ΔΔG	4.4	3.6	2.4	2.1
F	ΔΔH	2.5	4.4	2.4	2.6
F	ΔΔG	3.0	4.3	2.4	2.5

^aB3LYP = B3LYP/DZVP2. ^baD = aug-cc-pVDZ. ^caT = aug-cc-pVTZ.

Table 5.3. Calculated Acidities in kcal/mol of Phosphothreonine at Different Computational Levels. All with respect to neutral 1.

Amino Acid	Prop	B3LYP ^a	G3(MP2)	MP2/aD ^b	MP2/aT ^c
pThr PO ₃ H ⁻ [A]	ΔH	313.3	314.2	311.0	312.0
	ΔG	307.0	308.3	305.0	306.0
pThr PO ₃ H ⁻ [B]	ΔH	315.1	315.5	311.6	312.8
	ΔG	309.9	310.0	306.1	307.3
pThr PO ₃ H ⁻ [C]	ΔH	316.3	318.5	314.8	315.9
	ΔG	309.1	311.7	307.9	309.1
pThr COO ⁻ [D]	ΔH	318.0	318.4	314.8	315.2
	ΔG	311.9	312.3	308.7	309.2
pThr COO ⁻ [E]	ΔH	316.1	317.4	313.2	313.8
	ΔG	310.8	312.4	308.2	308.8
pThr zwitt [F]	ΔH	316.5	320.2	315.2	316.3
	ΔG	309.9	313.6	308.5	309.6
A	ΔΔH	0	0	0	0
A	ΔΔG	0	0	0	0
B	ΔΔH	1.8	1.3	0.6	0.8
B	ΔΔG	0.5	1.7	1.1	1.3
C	ΔΔH	2.9	4.3	3.8	3.9
C	ΔΔG	2.1	3.4	2.9	3.1
D	ΔΔH	4.7	4.2	3.8	3.2
D	ΔΔG	4.9	4.0	3.7	3.2
E	ΔΔH	2.8	3.2	2.2	1.8
E	ΔΔG	3.8	4.1	3.3	2.8
F	ΔΔH	3.2	6.0	4.2	4.3
F	ΔΔG	2.9	5.3	3.5	3.6

^a B3LYP = B3LYP/DZVP2. ^b aD = aug-cc-pVDZ. ^c aT = aug-cc-pVTZ.

Table 5.4. Vibrational Frequencies (ν , cm^{-1}) of Phosphoserine Anions at the B3LYP/aug-cc-pvdz level.

Bands	Vibrational modes	ν IRMPD ^{a,36}	pSer PO ₃ H ⁻ [A]		pSer PO ₃ H ⁻ [B]		pSer PO ₃ H ⁻ [C]		pSer zwitt [F]	
			ν Calc ^b	Ratio ^c	ν Calc ^b	Ratio ^c	ν Calc ^b	Ratio ^c	ν Calc ^b	Ratio ^c
N	ν_s CO–P, $\delta_{\text{rock.}}$ NH ₂	738 [0.02]	670 (5)	1.101	667 (106)	1.106	690 (123)	1.070	706 (124)	1.045
	$\delta_{\text{rock.}}$ NH ₂ , ν CO–P	812 [0.08]	709 (106)	1.145	719 (63)	1.129	738 (25)	1.100	767 (40)	1.059
M	ν P–OH, ν C–N	836 [0.08]	761 (240)	1.099	768 (92)	1.089	750 (271)	1.115	774 (279)	1.080
I	ν_s PO ₂ ⁻	1028 [0.03]	1001 (177)	1.027	970 (165)	1.060	1018 (189)	1.010	987 (205)	1.042
H	$\delta_{\text{wag.}}$ NH ₂ and CH ₂	1052 [0.04]	1045 (201)	1.007	1041 (107)	1.011	1053 (100)	0.999	1042 (213)	1.010
G	δ P–OH	1108 [0.07]	1115 (65)	0.994	1142 (86)	0.970	1088 (141)	1.018	1136 (117)	0.975
E	ν_{as} PO ₂ ⁻	1291 [0.21]	1230 (429)	1.050	1247 (424)	1.035	1240 (377)	1.041	1227 (480)	1.057
D	$\delta_{\text{rock.}}$ NH ₂	1419 [0.3]	1383 (8)	1.026	1388 (11)	1.022	1387 (12)	1.023	1392 (6)	1.019
C	δ C–O–H	1461 [0.03]	1507 (273)	0.969	1497 (256)	0.976	1487 (395)	0.983	1492 (150)	0.979
B	$\delta_{\text{sciss.}}$ NH ₂	1610 [0.03]	1655 (31)	0.973	1636 (37)	0.984	1679 (12)	0.959	1650 (31)	0.976
A	ν C=O	1728 [0.08]	1749 (427)	0.988	1749 (444)	0.988	1795 (441)	0.963	1713 (422)	1.009

^aThe values in brackets are the IRMPD yields. ^bThe values in parentheses are the infrared intensities in km/mol.

^cThe ratio obtained from the experimental frequencies divided by the calculated frequencies.

Table 5.5. Vibrational Frequencies (ν , cm^{-1}) of Phosphothreonine Anions at the B3LYP/aug-cc-pvdz level.

Bands	Vibrational modes	ν IRMPD ^{a,36}	pThr PO ₃ H ⁻ [A]		pThr PO ₃ H ⁻ [B]		pThr PO ₃ H ⁻ [C]		pThr zwitt [F]	
			ν Calc ^b	Ratio ^c	ν Calc ^b	Ratio ^c	ν Calc ^b	Ratio ^c	ν Calc ^b	Ratio ^c
N	ν CO-P	773 [0.07]	751 (49)	1.029	729 (71)	1.060	707 (39)	1.093	722 (51)	1.071
M	ν P-OH	842 [0.31]	806 (191)	1.045	777 (102)	1.084	754 (272)	1.117	781 (260)	1.078
L	δ NH ₂ , ν C-N	962 [0.23]	969 (47)	0.993	961 (77)	1.001	950 (127)	1.013	937 (141)	1.027
I	ν_s PO ₂ ⁻	1000 [0.16]	1001 (166)	0.999	982 (211)	1.018	1028 (169)	0.973	993 (188)	1.007
H	ν CO-P, $\delta_{\text{wag.}}$ NH ₂	1038 [0.30]	1057 (192)	0.982	1043 (70)	0.995	1059 (137)	0.980	1061 (68)	0.978
G	δ P-OH	1106 [0.10]	1107 (59)	0.999	1108 (64)	0.998	1115 (51)	0.992	1154 (113)	0.958
F	ν C-OH, δ CH, δ NH ₂	1257 [0.87]	1272 (143)	0.988	1247 (433)	1.008	1236 (68)	1.017	1230 (370)	1.022
E	ν_{as} PO ₂ ⁻	1299 [0.74]	1298 (276)	1.001	1266 (71)	1.026	1242 (295)	1.046	1321 (201)	0.983
D	$\delta_{\text{rock.}}$ NH ₂	1419 [0.30]	1402 (26)	1.012	1395 (13)	1.017	1393 (15)	1.019	1392 (17)	1.019
C	δ C-O-H	1444 [0.24]	1507 (277)	0.958	1496 (262)	0.965	1496 (413)	0.965	1492 (199)	0.968
B	$\delta_{\text{sciss.}}$ NH ₂	1576 [0.12]	1658 (33)	0.951	1634 (34)	0.965	1679 (15)	0.939	1654 (10)	0.953
A	δ C=O	1713 [0.08]	1727 (334)	0.992	1745 (409)	0.982	1789 (428)	0.958	1715 (394)	0.999

^aThe values in brackets are the IRMPD yields. ^bThe values in parentheses are the infrared intensities in km/mol.

^cThe ratio obtained from the experimental frequencies divided by the calculated frequencies.

Table 5.6. N-H, O-H, and Hydrogen Bond Lengths of Phosphoserine Anions in Å at Different Computational Levels.

Amino Acid	Method	N-H1	O5...H1	O17...H1	N-H18	O5...H18	O17...H18	N...(-)H8	O5...(-)H8	N...H16
pSer PO ₃ H ⁻ [A]	B3LYP ^a	1.022	2.266	NB	1.018	3.080	NB	NB	NB	NB
	G3(MP2) ^b	1.023	2.324	NB	1.021	2.922	NB	NB	NB	NB
	MP2/aD ^c	1.025	2.380	NB	1.022	2.976	NB	NB	NB	NB
	MP2/aT ^d	1.018	2.319	NB	1.014	3.000	NB	NB	NB	NB
pSer COO ⁻ [D]	B3LYP ^a	1.022	2.035	NB	1.016	3.402	NB	NB	NB	NB
	G3(MP2) ^b	1.025	2.144	NB	1.021	2.984	NB	NB	NB	NB
	MP2/aD ^c	1.027	2.161	NB	1.022	3.097	NB	NB	NB	NB
	MP2/aT ^d	1.019	2.023	NB	1.011	3.395	NB	NB	NB	NB
pSer PO ₃ H ⁻ [B]	B3LYP ^a	1.023	2.643	2.003	1.017	3.413	NB	NB	1.037	2.169
	G3(MP2) ^b	1.026	2.596	1.920	1.020	3.344	NB	NB	1.024	2.060
	MP2/aD ^c	1.030	2.613	1.929	1.021	3.350	NB	NB	1.031	2.084
	MP2/aT ^d	1.022	2.596	1.931	1.013	3.332	NB	NB	1.034	2.084
pSer COO ⁻ [E]	B3LYP ^a	1.023	2.162	2.887	1.016	3.511	NB	NB	1.532	1.764
	G3(MP2) ^b	1.027	2.018	2.795	1.018	3.401	NB	NB	1.646	1.721
	MP2/aD ^c	1.028	2.111	2.711	1.019	3.439	NB	NB	1.556	1.725
	MP2/aT ^d	1.021	2.082	2.795	1.012	3.431	NB	NB	1.646	1.722
pSer PO ₃ H ⁻ [C]	B3LYP ^a	1.016	NB	3.278	1.038	NB	1.806	1.736	1.012	NB
	G3(MP2) ^b	1.017	NB	3.312	1.038	NB	1.821	1.736	1.015	NB
	MP2/aD ^c	1.018	NB	3.309	1.042	NB	1.778	1.728	1.012	NB
	MP2/aT ^d	1.011	NB	3.272	1.037	NB	1.761	1.715	1.010	NB
pSer zwitterion [F]	B3LYP ^a	1.017	3.081	3.076	1.090	NB	1.508	1.059	1.715	NB
	G3(MP2) ^b	1.020	3.108	3.153	1.079	NB	1.562	1.061	1.713	NB
	MP2/aD ^c	1.019	3.098	3.106	1.088	NB	1.513	1.061	1.705	NB
	MP2/aT ^d	1.013	3.082	3.078	1.089	NB	1.491	1.055	1.701	NB

^aB3LYP = B3LYP/DZVP2. ^bG3(MP2) = MP2/6-31G(d). ^caD = aug-cc-pVDZ. ^daT = aug-cc-pVTZ. NB = not bonded.

Table 5.7. N-H, O-H, and Hydrogen Bond Lengths of Phosphothreonine Anions in Å at Different Computational Levels.

Amino Acid	Method	N-H1	O5...H1	O17...H1	N-H18	O5...H18	O17...H18	N...(-)H8	O5...(-)H8	N...H16
pThr PO ₃ H ⁻ [A]	B3LYP ^a	1.021	2.258	NB	1.018	3.131	NB	NB	NB	NB
	G3MP2 ^b	1.023	2.225	NB	1.020	3.087	NB	NB	NB	NB
	MP2/aD ^c	1.025	2.256	NB	1.021	3.126	NB	NB	NB	NB
	MP2/aT ^d	1.018	2.211	NB	1.013	3.133	NB	NB	NB	NB
pThr COO ⁻ [D]	B3LYP ^a	1.024	2.073	NB	1.017	3.229	NB	NB	NB	NB
	G3MP2 ^b	1.025	2.094	NB	1.021	3.069	NB	NB	NB	NB
	MP2/aD ^c	1.028	2.112	NB	1.021	3.151	NB	NB	NB	NB
	MP2/aT ^d	1.020	2.056	NB	1.013	3.194	NB	NB	NB	NB
pThr PO ₃ H ⁻ [B]	B3LYP ^a	1.023	2.631	2.011	1.017	3.413	NB	NB	1.038	2.147
	G3MP2 ^b	1.026	2.587	1.918	1.020	3.344	NB	NB	1.025	2.043
	MP2/aD ^c	1.030	2.610	1.927	1.021	3.350	NB	NB	1.032	2.061
	MP2/aT ^d	1.022	2.593	1.928	1.013	3.332	NB	NB	1.035	2.061
pThr COO ⁻ [E]	B3LYP ^a	1.023	2.148	2.940	1.016	3.510	NB	NB	1.534	1.765
	G3MP2 ^b	1.027	2.010	2.828	1.018	3.400	NB	NB	1.634	1.726
	MP2/aD ^c	1.027	2.112	2.686	1.019	3.444	NB	NB	1.536	1.735
	MP2/aT ^d	1.021	2.086	2.790	1.012	3.436	NB	NB	1.539	1.732
pThr PO ₃ H ⁻ [C]	B3LYP ^a	1.015	NB	3.251	1.038	NB	1.823	1.692	1.016	NB
	G3MP2 ^b	1.017	NB	3.324	1.037	NB	1.842	1.712	1.017	NB
	MP2/aD ^c	1.018	NB	3.339	1.041	NB	1.792	1.715	1.014	NB
	MP2/aT ^d	1.011	NB	3.294	1.036	NB	1.770	1.692	1.013	NB
pThr zwitt [F]	B3LYP ^a	1.017	3.104	3.081	1.088	NB	1.518	1.063	1.666	NB
	G3MP2 ^b	1.020	3.128	3.166	1.076	NB	1.584	1.063	1.685	NB
	MP2/aD ^c	1.019	3.108	3.125	1.086	NB	1.528	1.064	1.680	NB
	MP2/aT ^d	1.013	3.093	3.091	1.087	NB	1.503	1.058	1.670	NB

^aB3LYP = B3LYP/DZVP2. ^bG3MP2 = MP2/6-31G(d). ^caD = aug-cc-pVDZ. ^daT = aug-cc-pVTZ. NB = not bonded

Table 5.8. Gas-Phase Acidities in kcal/mol at the G3(MP2) level and Aqueous pKa's Obtained with COSMO Parameterization of Phosphorylated Amino Acids and Amides.^a

Compound	Phosphorylated Amino Acids			Phosphorylated Amino Acid Amides		
	$\Delta H_{298 \text{ gas}}$	$\Delta G_{298 \text{ gas}}$	pK _a (298)	$\Delta H_{298 \text{ gas}}$	$\Delta G_{298 \text{ gas}}$	pK _a (298)
arginine A (neut1) ^b	317.8	309.1	11.11	317.5	310.2	11.47
arginine B (neut1)	317.5	310.0	14.28	317.8	310.6	8.62
arginine C (neut1)	318.7	310.3	11.21	318.9	311.3	11.42
arginine D (neut1)	318.5	310.5	18.29	318.8	311.6	12.48
arginine E (neut1)	319.1	310.8	15.32	321.8	312.2	10.73
arginine A (neut2) ^c	306.1	298.3	-1.82	311.8	304.5	-2.36
arginine B (neut2)	305.8	299.2	-1.35	312.1	305.0	-5.21
arginine C (neut2)	307.0	299.5	-1.72	313.3	305.6	-2.41
arginine D (neut2)	306.8	299.7	5.36	313.1	305.9	-1.35
arginine E (neut2)	307.4	300.0	2.40	316.2	306.5	-3.10
arginine A (neut3) ^d	303.3	295.7	-3.59	310.9	303.7	-4.13
arginine B (neut3)	303.0	296.6	-0.43	311.2	304.1	-4.20
arginine C (neut3)	304.2	296.9	-3.49	312.3	304.8	-4.18
arginine D (neut3)	304.0	297.1	3.59	312.2	305.0	-3.12
arginine E (neut3)	304.6	297.4	0.62	315.3	305.7	-4.88
aspartic acid A	308.8	302.4	2.71	310.2	302.2	-0.81
aspartic acid B	309.2	303.1	2.59	311.2	303.9	0.48
cysteine A	312.8	306.5	6.45	317.8	310.3	-2.44
cysteine B	313.5	306.6	7.34	318.1	310.5	-0.50
cysteine C	312.4	306.7	4.98	320.6	311.6	-1.39
cysteine D	314.4	306.9	4.75	321.2	312.0	-2.57
cysteine E	314.5	307.6	6.90			
glutamic acid	305.9	299.6	1.39	312.1	304.2	0.10
glycine	315.0	308.3	4.16	322.3	315.1	4.37
histidine τ A	305.6	298.5	-1.32	309.7	303.0	-1.69
histidine τ B	308.1	302.0	0.03	310.8	304.1	-3.61
histidine τ C				310.4	304.5	-0.52

histidine π A	309.6	299.6	11.69	310.3	302.6	0.19
histidine π B	307.6	301.0	5.46	311.5	303.1	-1.45
lysine A	319.6	312.5	7.57	319.4	313.8	4.39
lysine B	319.5	312.6	9.30	319.4	314.3	4.38
lysine C	323.1	316.6	3.32			
serine A	315.7	309.2	4.65	320.4	313.2	3.12
serine B	316.0	310.5	4.00	324.5	316.1	2.65
serine C	317.9	310.7	4.69	337.0	329.3	11.07
serine D	317.5	311.7	9.67			
serine E	318.0	312.8	3.62			
serine F	320.1	313.5	0.29			
threonine A	314.2	308.3	4.25	321.5	314.2	5.16
threonine B	315.5	310.0	6.22	321.9	314.8	3.39
threonine C	318.5	311.7	8.83	337.4	330.2	11.91
threonine D	318.4	312.3	4.04			
threonine E	317.4	312.4	4.03			
threonine F	320.2	313.6	1.68			
tyrosine A	314.5	306.7	9.42	314.9	307.0	1.93
tyrosine B	315.3	310.7	7.29			
tyrosine C	317.7	310.7	3.29			
tyrosine D	324.8	317.6	7.56			
tyrosine E ^c	333.2	326.0	2.49			
tyrosine F	336.0	327.0	14.39			

^a Bold values are the most acidic site in solution.

^b Zwitterion neutral with a protonated guanidine side chain nitrogen and deprotonated phosphate group.

^c Zwitterion neutral where the guanidine side chain contains one protonated and one deprotonated nitrogen.

^d Non-zwitterion neutral.

^e GA values were obtained at a hybrid G3(MP2) approach using the Hartree-Fock geometry. Lowest pK_a value only obtained with the HF geometry.

Table 5.9. Vibrational Frequencies (ν , cm^{-1}) of Phosphotyrosine Anions at the B3LYP/aug-cc-pvdz level.

Bands	Vibrational modes	ν IRMPD ^{a, 36}	pTyr COO ⁻ [A]		pTyr PO ₃ H ⁻ [B]		pTyr PO ₃ H ⁻ [C]		pTyr zwitt [D]	
			ν Calc ^b	Ratio ^c	ν Calc ^b	Ratio ^c	ν Calc ^b	Ratio ^c	ν Calc ^b	Ratio ^c
H	$\delta_{\text{wag.}} \text{NH}_2$	845 [0.09]	–	–	854 (127)	0.989	842 (44)	1.004	852 (75)	0.992
G	$\delta \text{P-OH}$	1049 [0.08]	861 (118)	1.218	1020 (115)	1.028	982 (87)	1.068	1018 (119)	1.030
F	$\nu_s \text{PO}_2^-$	1098 [0.43]	926 (141)	1.186	1065 (78)	1.031	1041 (279)	1.055	1067 (58)	1.029
E	Ring def., $\delta \text{P-O-C}$	1271 [0.44]	1228 (344)	1.035	1199 (314)	1.060	1290 (311)	0.985	1251 (225)	1.016
D	$\delta \text{C-O-H}$	1329 [0.39]	1345 (227)	0.988	1236 (254)	1.075	1428 (428)	0.931	1323 (207)	1.005
C	$\delta_{\text{ip}} \text{C-H}$	1502 [0.07]	1422 (61)	1.056	1529 (104)	0.982	1533 (219)	0.980	1524 (107)	0.986
B	$\delta_{\text{ip}} \text{C-H}$, ring. def.	1596 [0.03]	1525 (64)	1.047	1647 (43)	0.969	1646 (89)	0.970	1656 (70)	0.964
A	$\nu \text{C=O}$	1769 [0.17]	1658 (488)	1.067	1742 (317)	1.015	1813 (405)	0.976	1716 (415)	1.031

^aThe values in brackets are the IRMPD yields. ^bThe values in parentheses are the infrared intensities in km/mol .

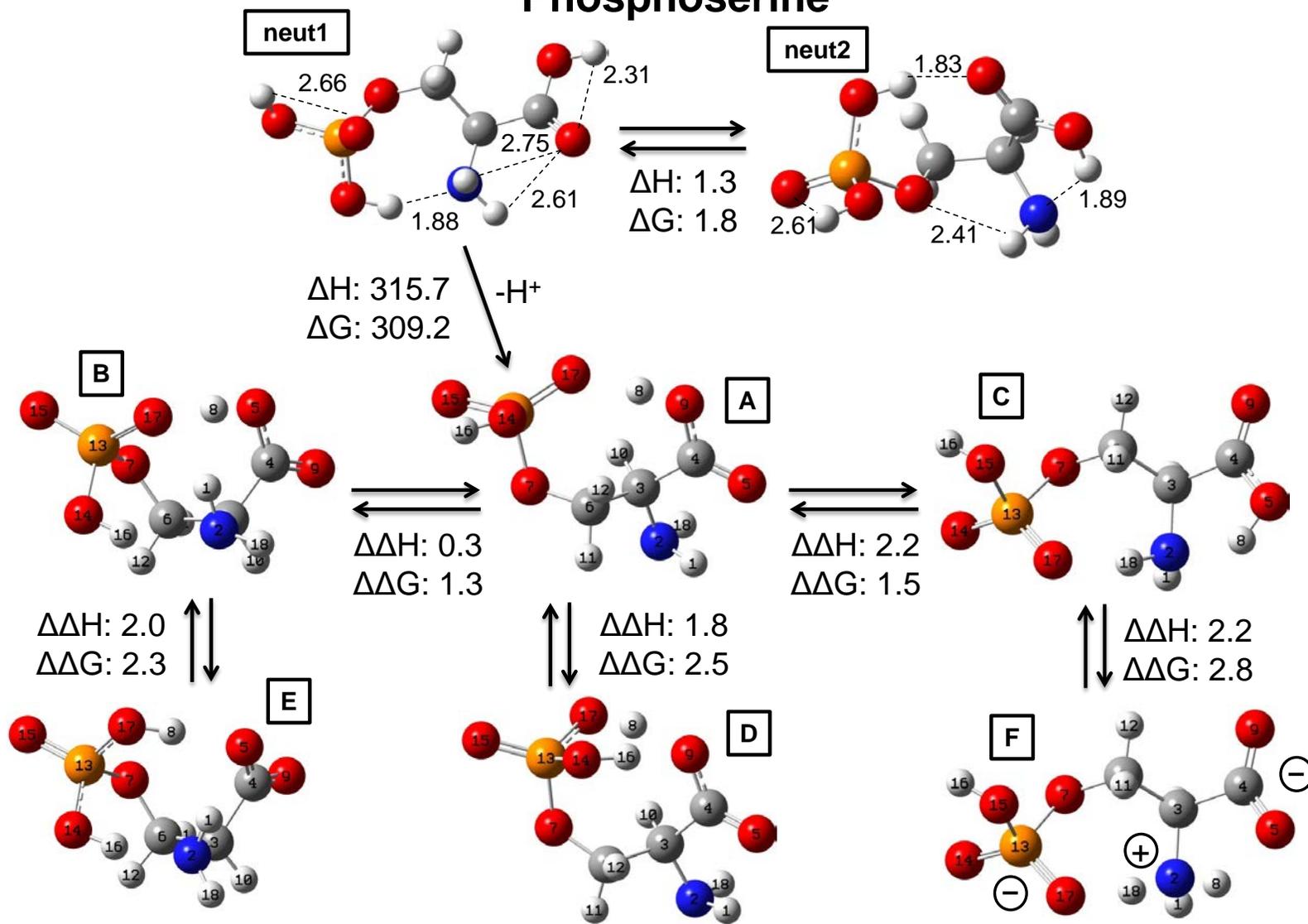
^cThe ratio obtained from the experimental frequencies divided by the calculated frequencies.

Table 5.10. Calculated Heats of Formation at 298 K in kcal/mol at the G3(MP2) level of Phosphorylated Amino Acids and Amides with Two Different Computational Approaches.

Amino Acids	ΔH_f Phosphorylated Amino Acid		ΔH_f Phosphorylated Amino Acid Amide	
	TAE	Isodesmic	TAE	Isodesmic
arginine (neut1)	-310.6	-310.2	-261.2	-260.8
arginine (neut2)	-298.9	-298.5	-255.5	-255.1
arginine (neut3)	-296.1	-295.7	-254.6	-254.2
aspartic acid	-394.6	-394.0	-350.9	-350.6
cysteine	-297.6	-297.2	-255.1	-254.6
glutamic acid	-403.2	-402.6	-361.5	-361.1
glycine	-300.6	-300.5 ^a	-256.7	-256.3
histidine τ	-271.4	-271.0	-227.5	-227.1
histidine π	-268.4	-268.1	-223.0	-222.6
lysine	-320.3	-319.9	-273.2	-272.8
serine	-353.3	-352.9	-309.6	-309.2
threonine	-362.4	-362.0	-320.4	-320.0
tyrosine	-328.0	-327.6	-283.5	-281.1

^a Used the experimental heat of formation of glycine

Phosphoserine



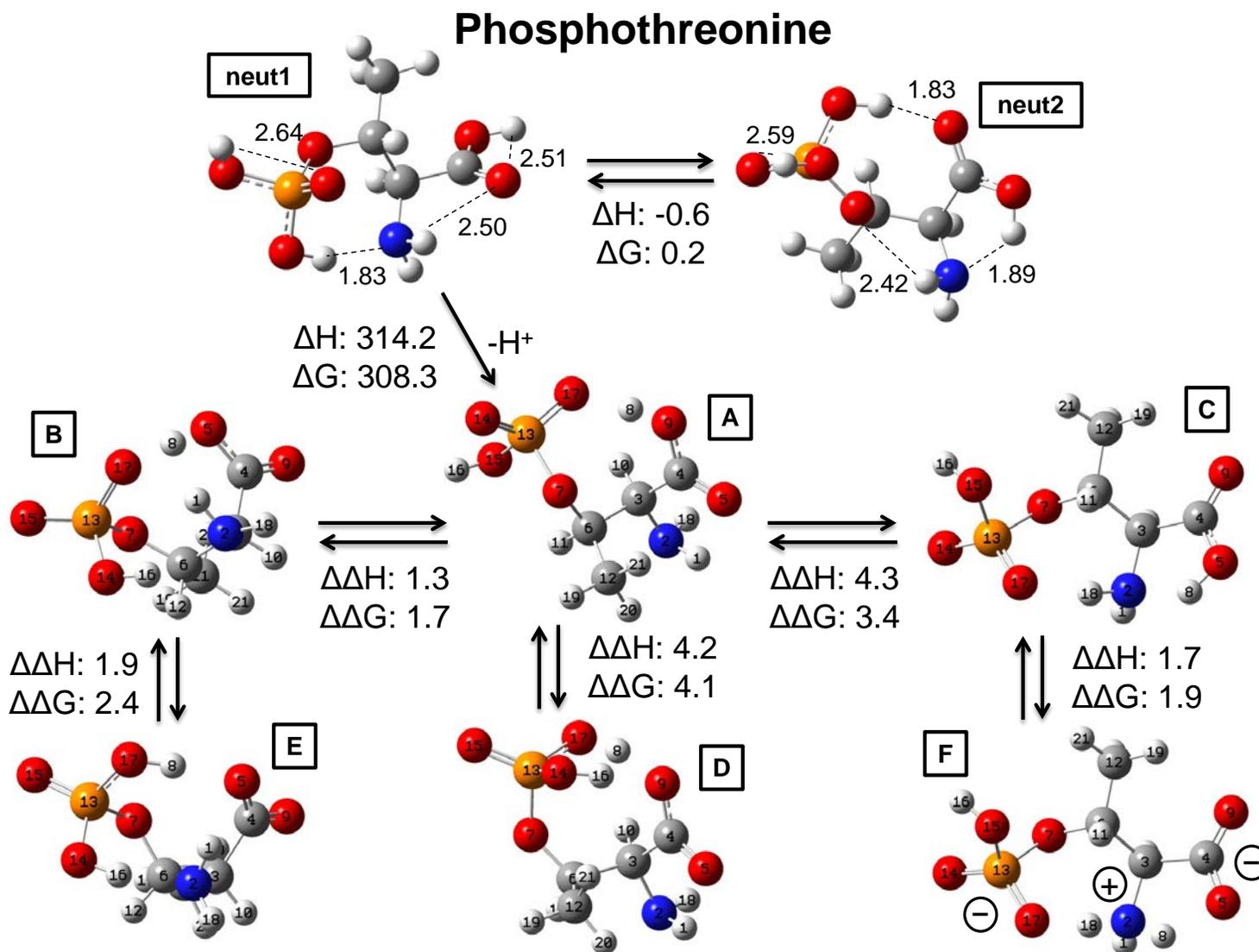
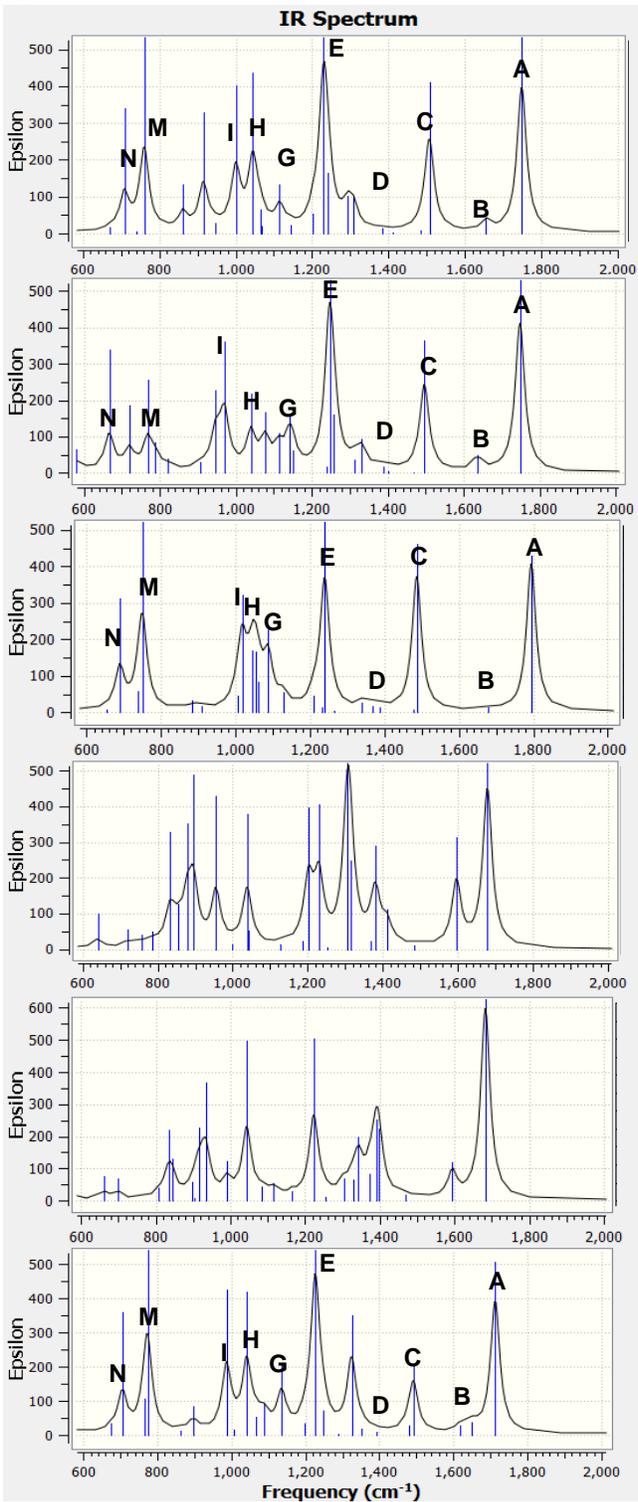
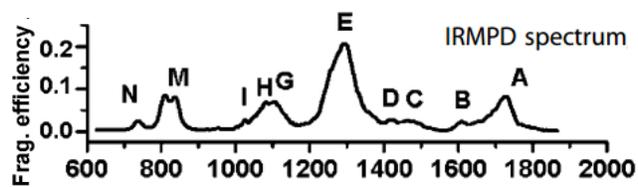
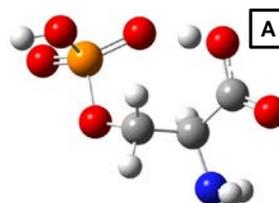


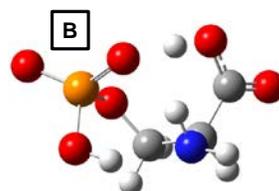
Figure 5.1. Optimized structures of pSer, pThr, and anions at the G3(MP2) level. GAs and relative energies are in kcal/mol. Important hydrogen bond distances, given in Å, for the [pSer-H]⁻ and [pThr-H]⁻ anions are given in Tables 6 and 7 respectively.



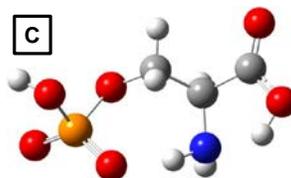
Deprotonated phosphoserine [pSer-H]⁻



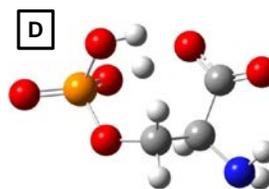
$\Delta\Delta H: 0.0$
 $\Delta\Delta G: 0.0$



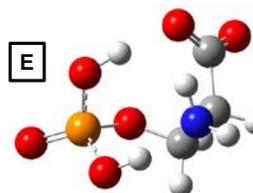
$\Delta\Delta H: 0.3$
 $\Delta\Delta G: 1.3$



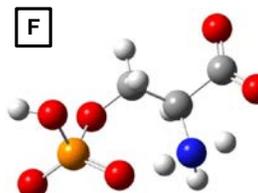
$\Delta\Delta H: 2.2$
 $\Delta\Delta G: 1.5$



$\Delta\Delta H: 1.8$
 $\Delta\Delta G: 2.5$



$\Delta\Delta H: 2.5$
 $\Delta\Delta G: 3.6$



$\Delta\Delta H: 4.4$
 $\Delta\Delta G: 4.3$

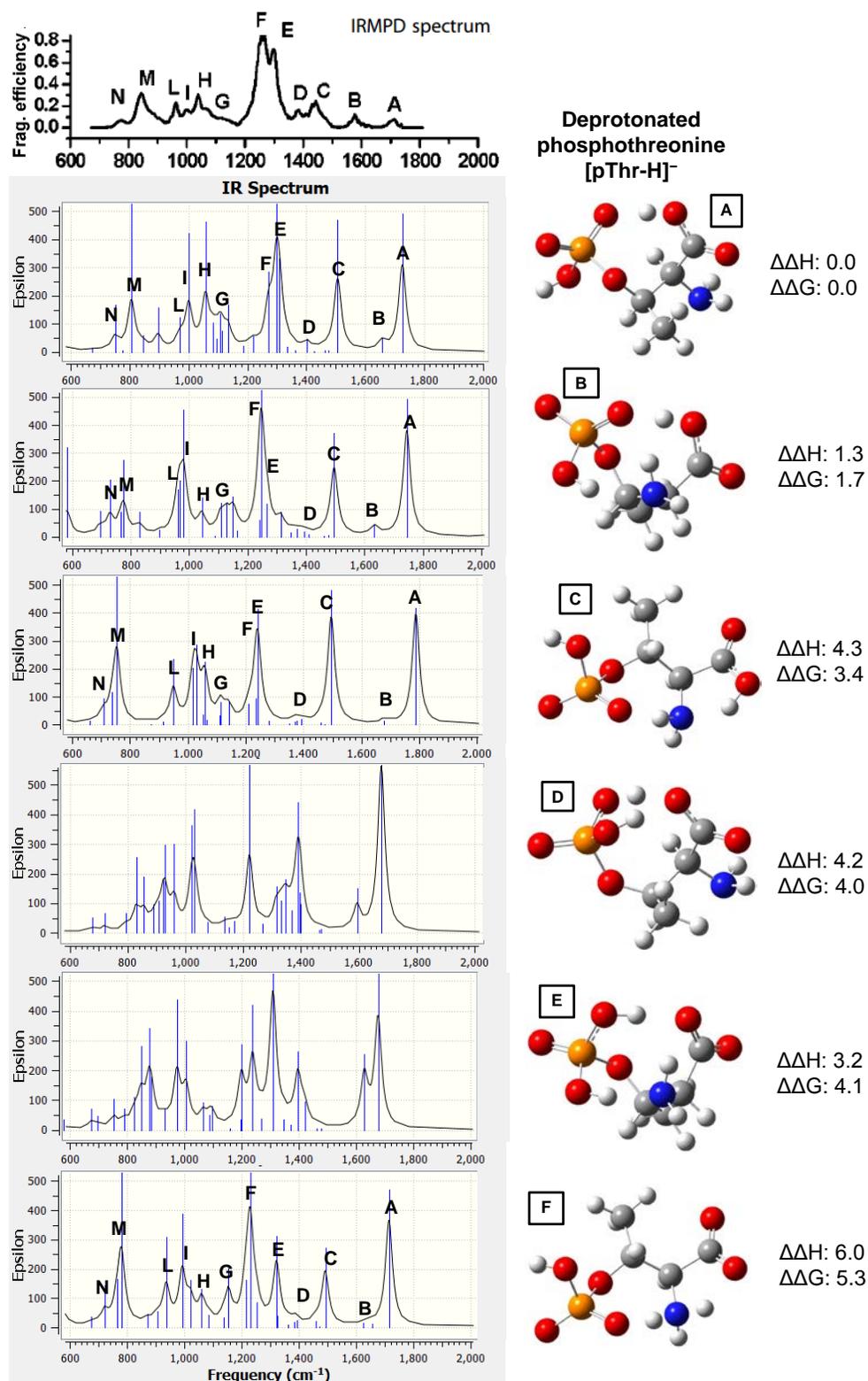


Figure 5.2. Infrared spectra of deprotonated phospho-serine and -threonine anions. IRMPD spectrum is on top. Anions [A] – [F]: calculated IR spectra at the B3LYP/aug-cc-pvdz level. Relative energies of isomers at the G3(MP2) level of theory.

Phosphotyrosine

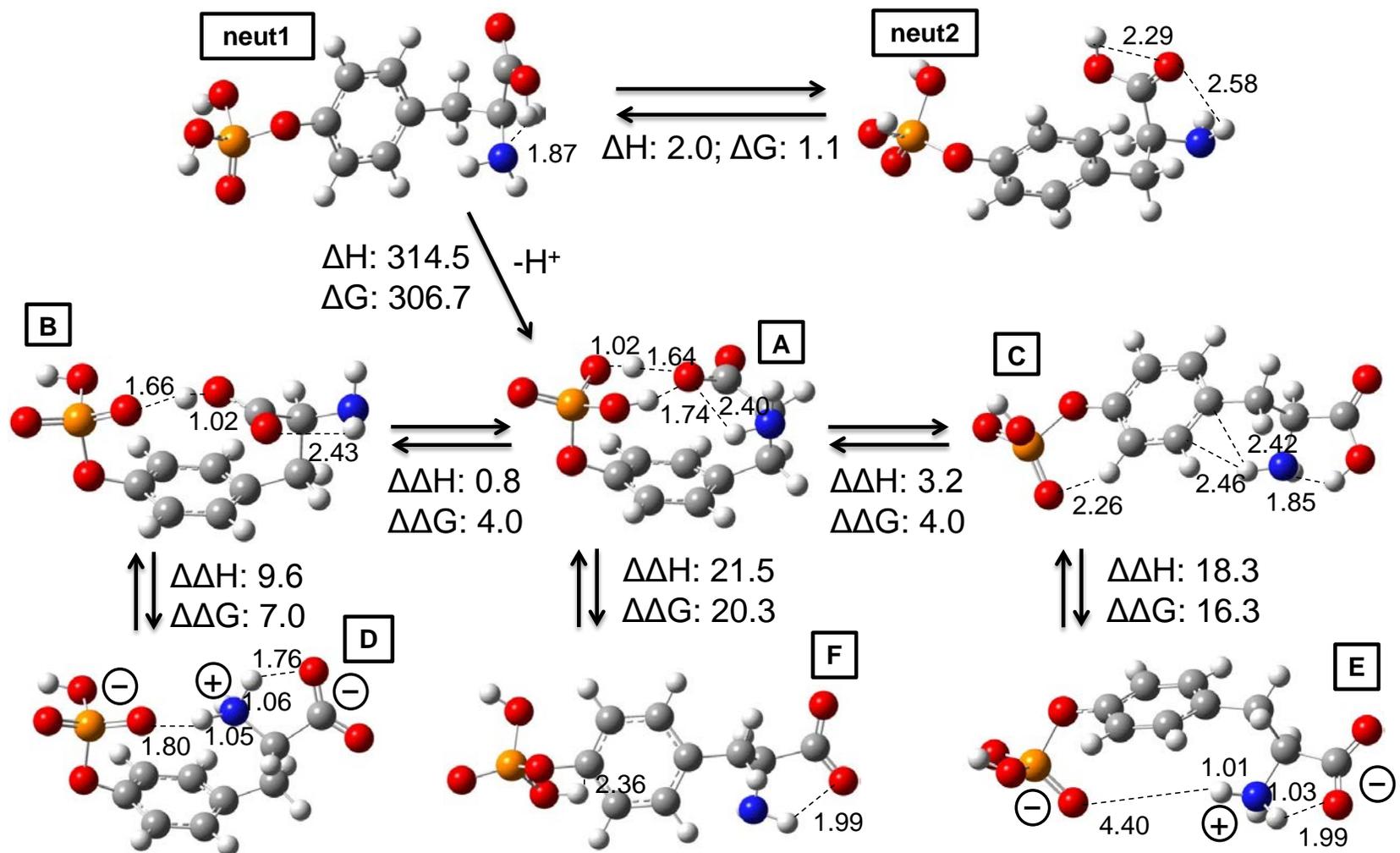


Figure 5.3. Optimized structures of phosphotyrosine and corresponding anions at the G3(MP2) level. GAs and relative energies are given in kcal/mol. Important hydrogen bond distances are given in Å.

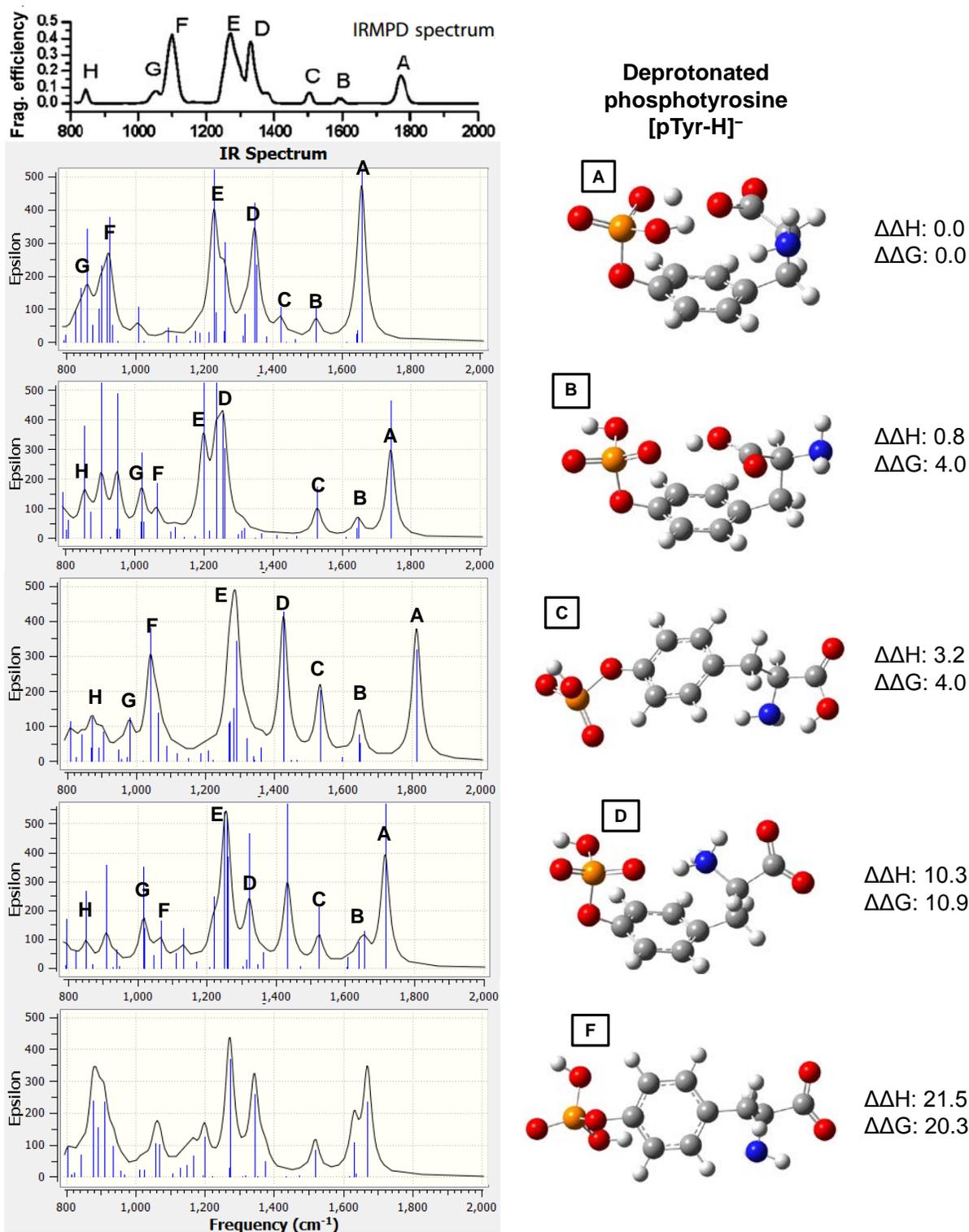
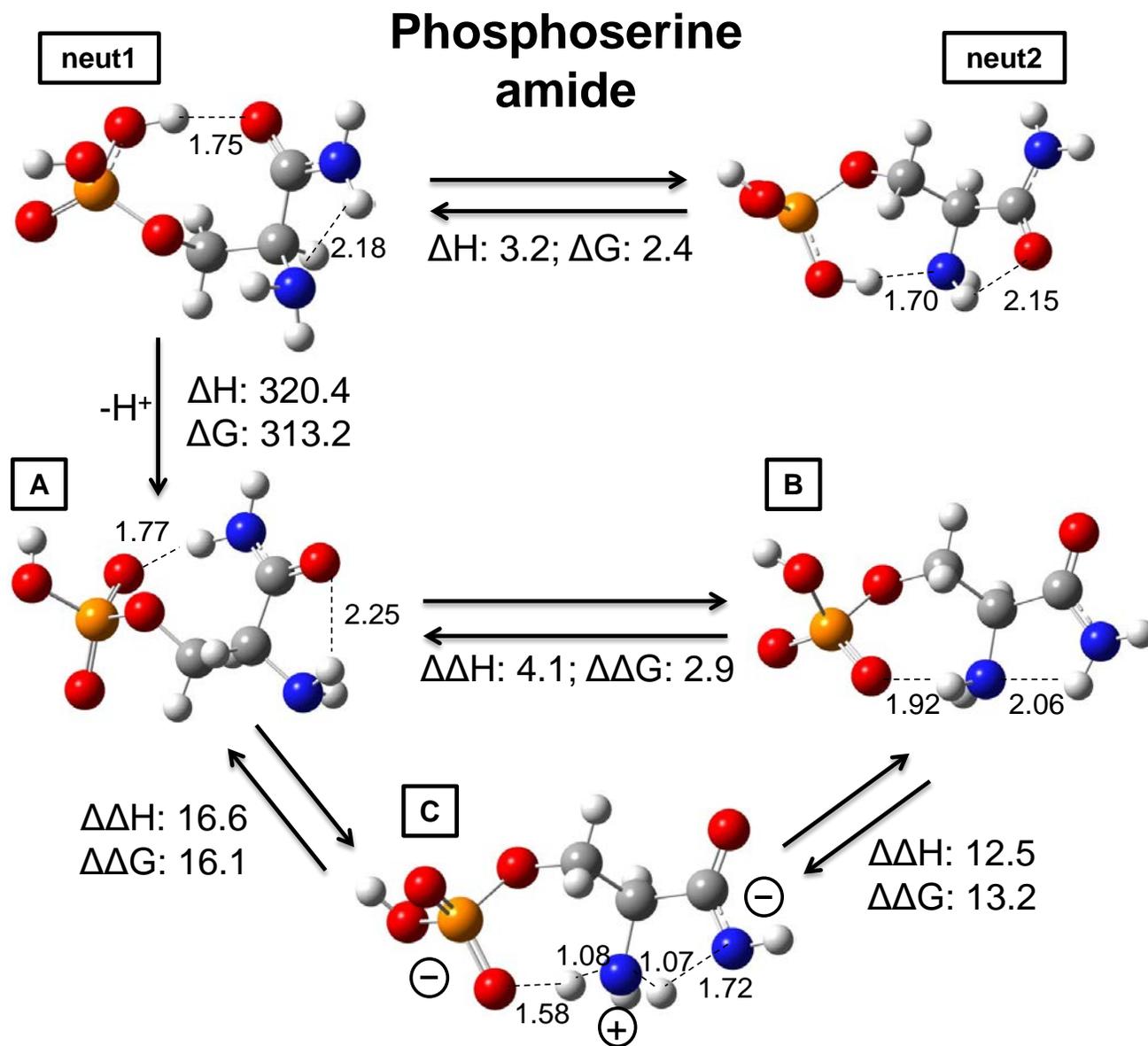
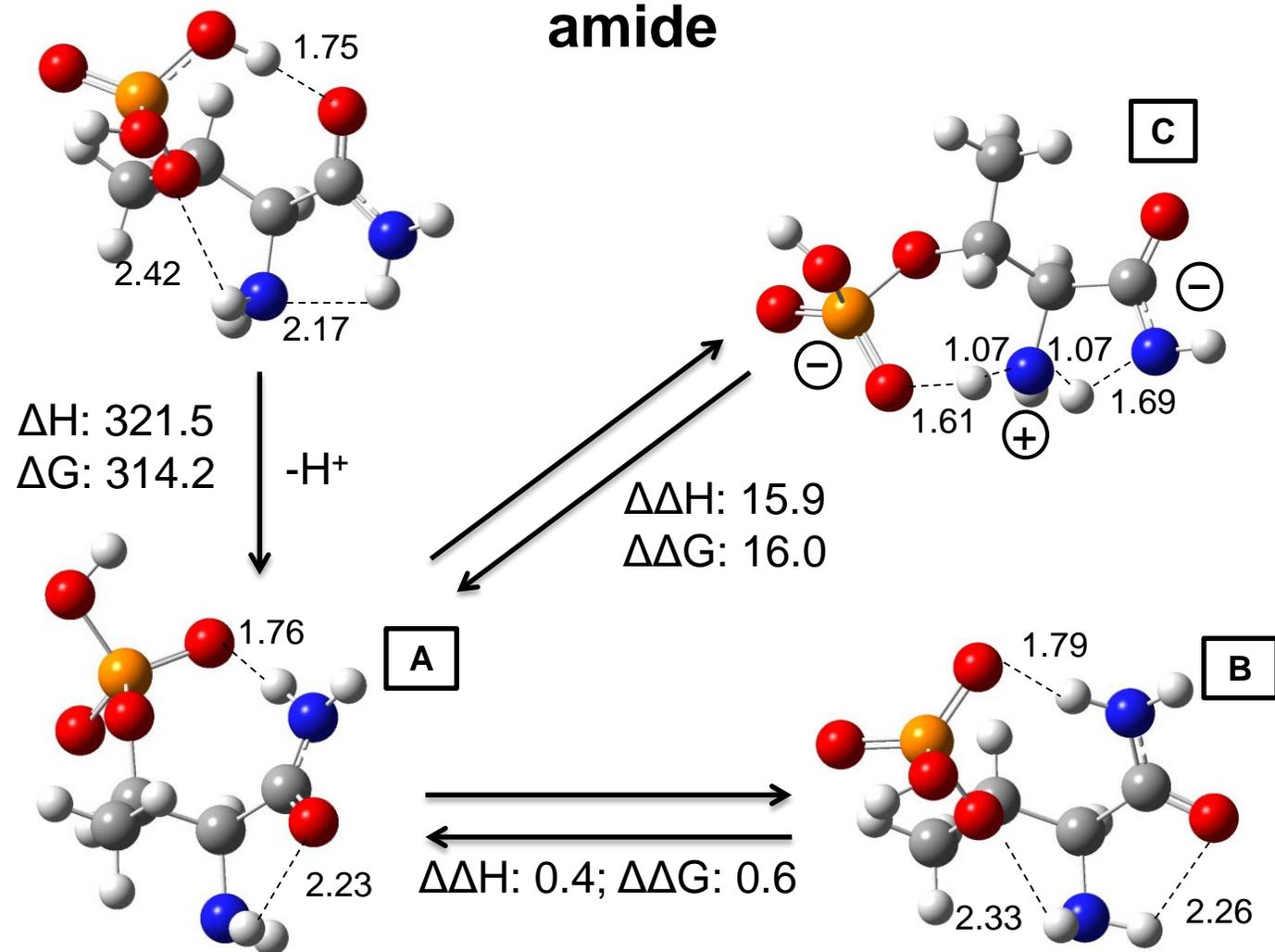


Figure 5.4. Infrared spectra of deprotonated phosphotyrosine. IRMPD spectrum is on top. Anions [A] – [D] & [F]: calculated IR spectra at the B3LYP/aug-cc-pvdz level. Relative energies of isomers at the G3(MP2) level of theory.



Phosphothreonine amide



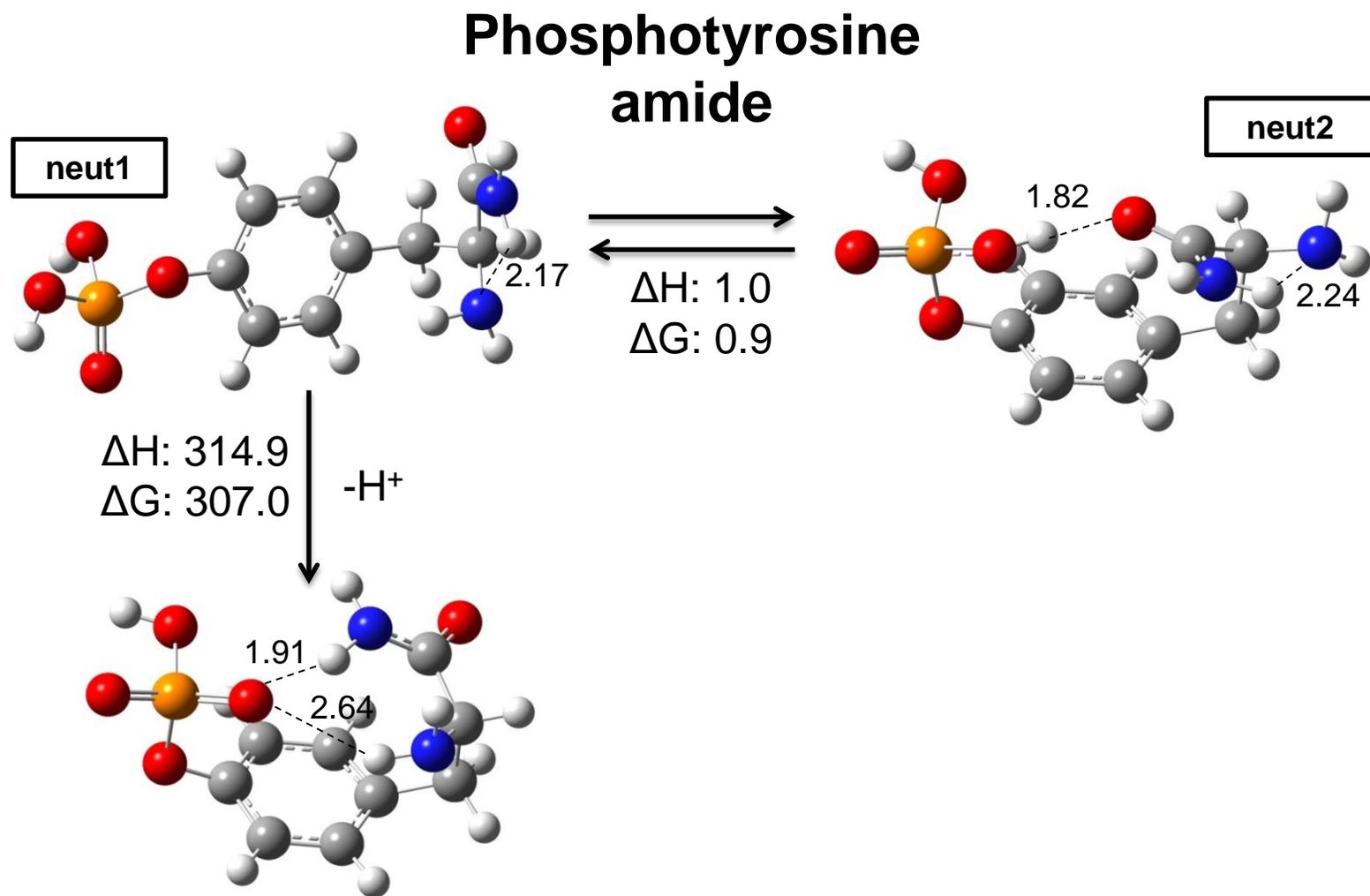
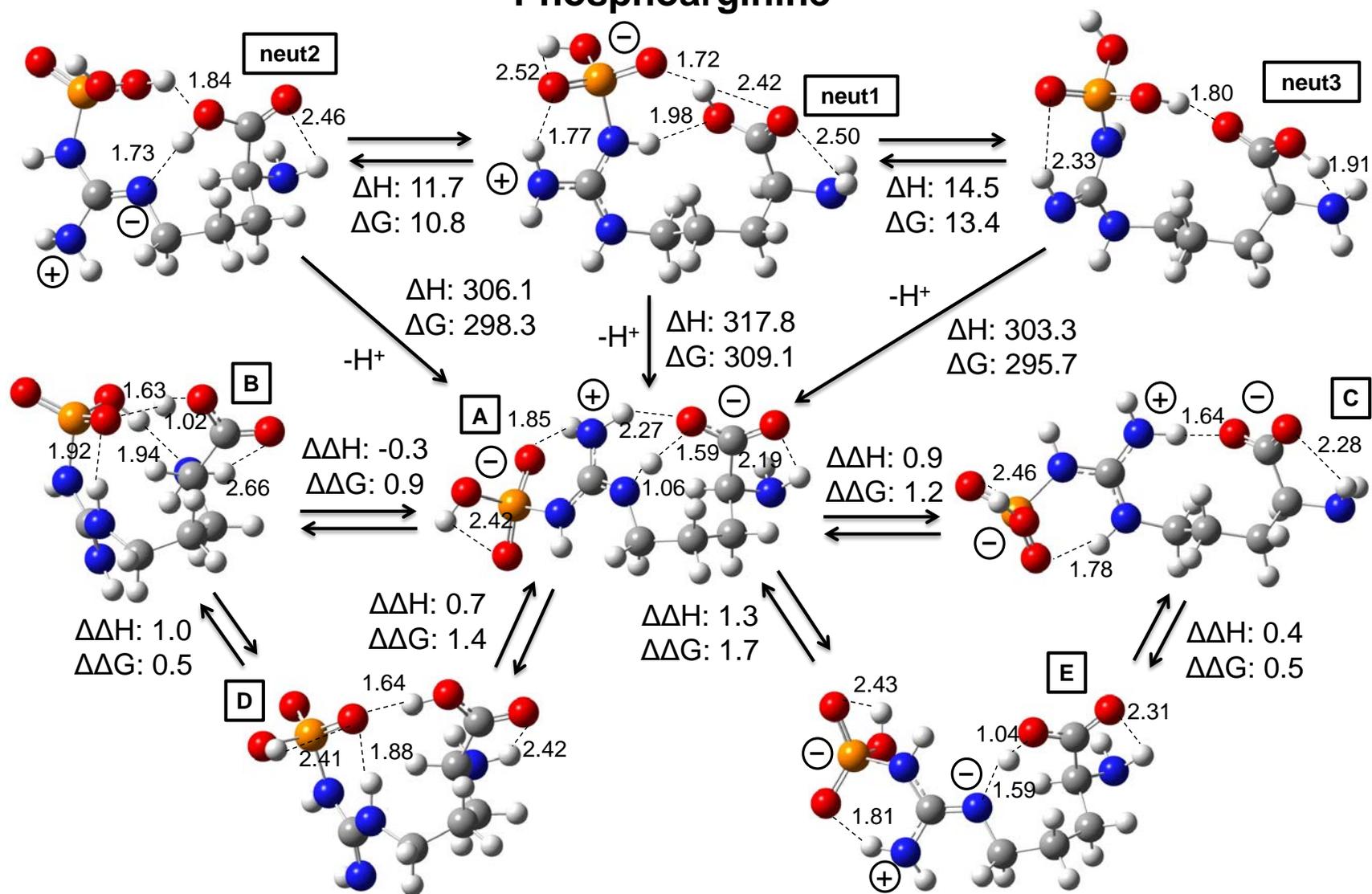
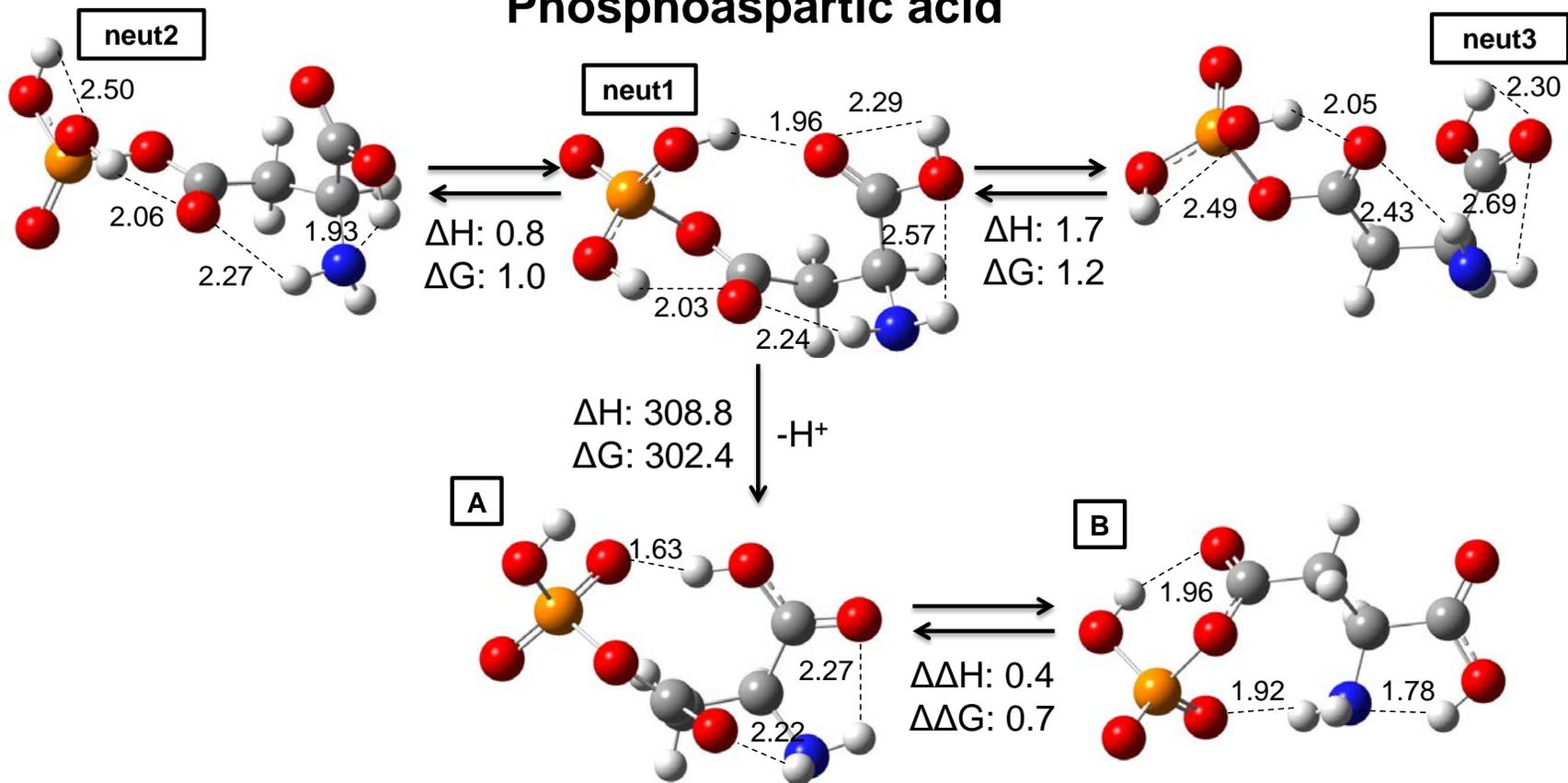


Figure 5.5. Optimized structures of phospho-serine, -threonine, and -tyrosine amide neutrals and corresponding anions at the G3(MP2) level. GAs and relative energies are in kcal/mol. Important hydrogen bond distances are given in Å.

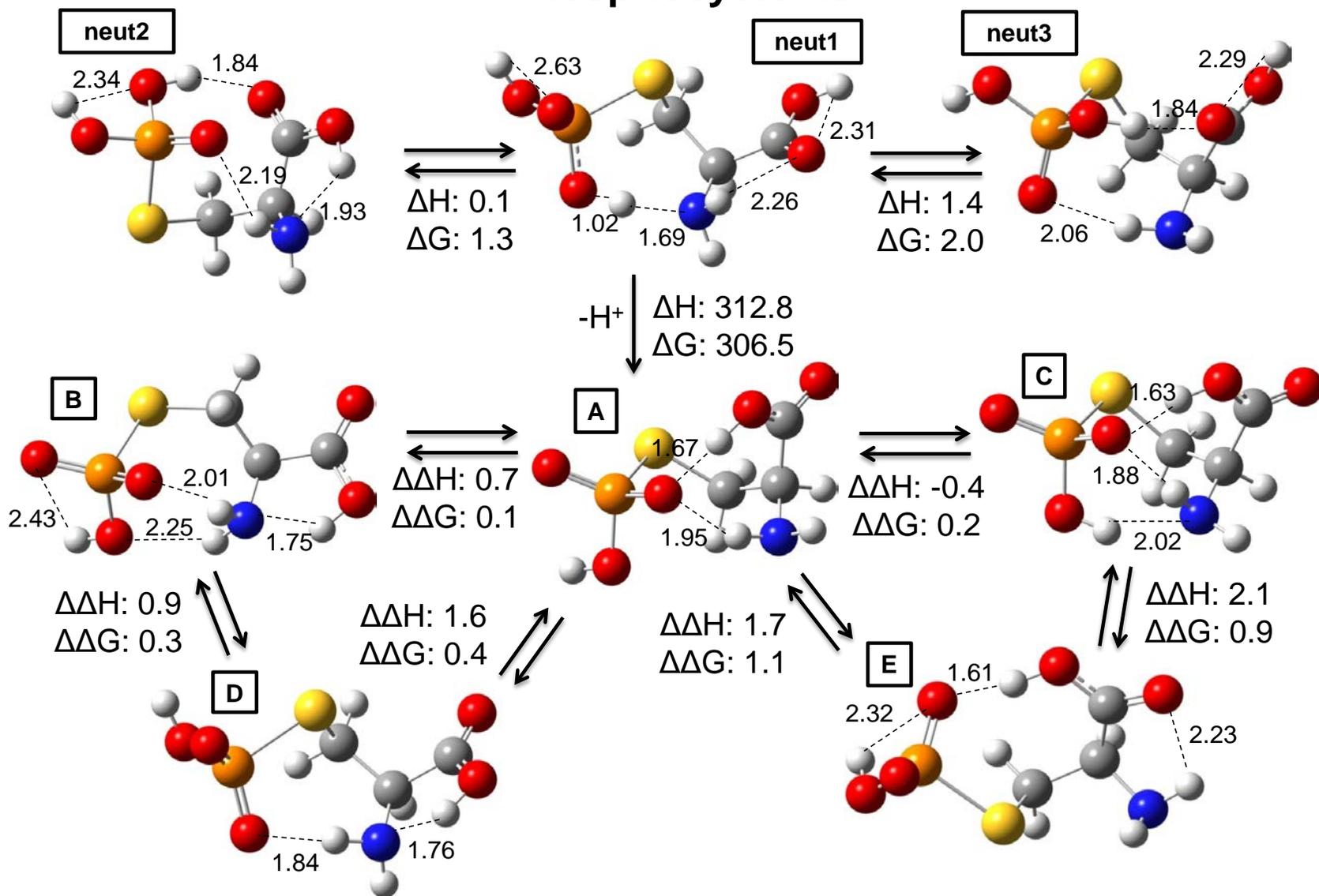
Phosphoarginine



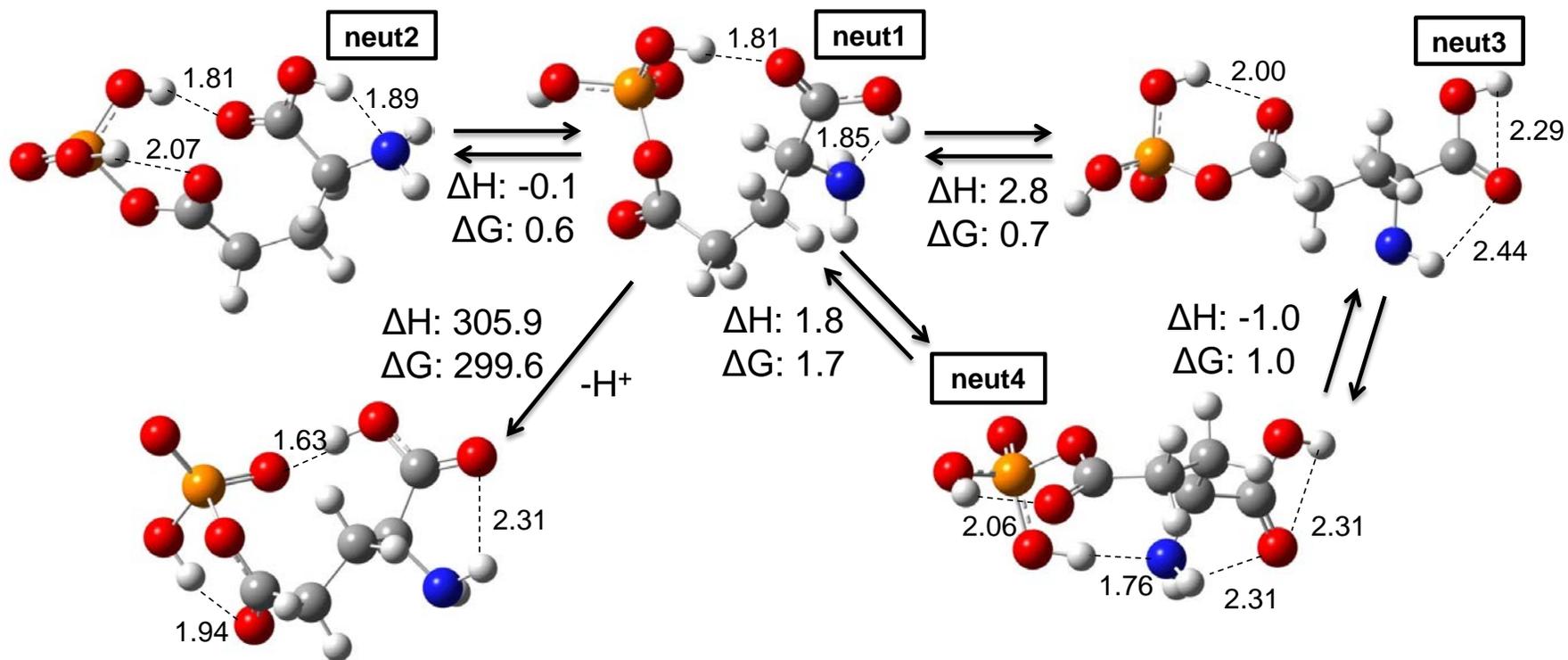
Phosphoaspartic acid



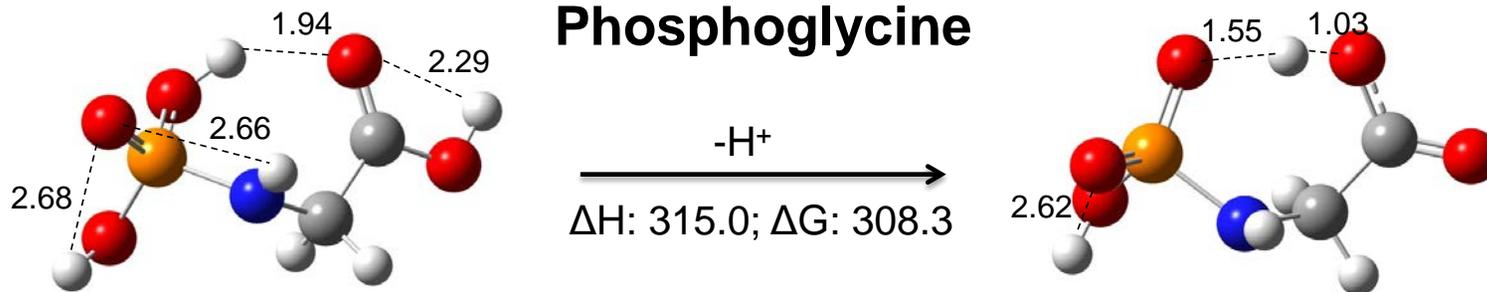
Phosphocysteine



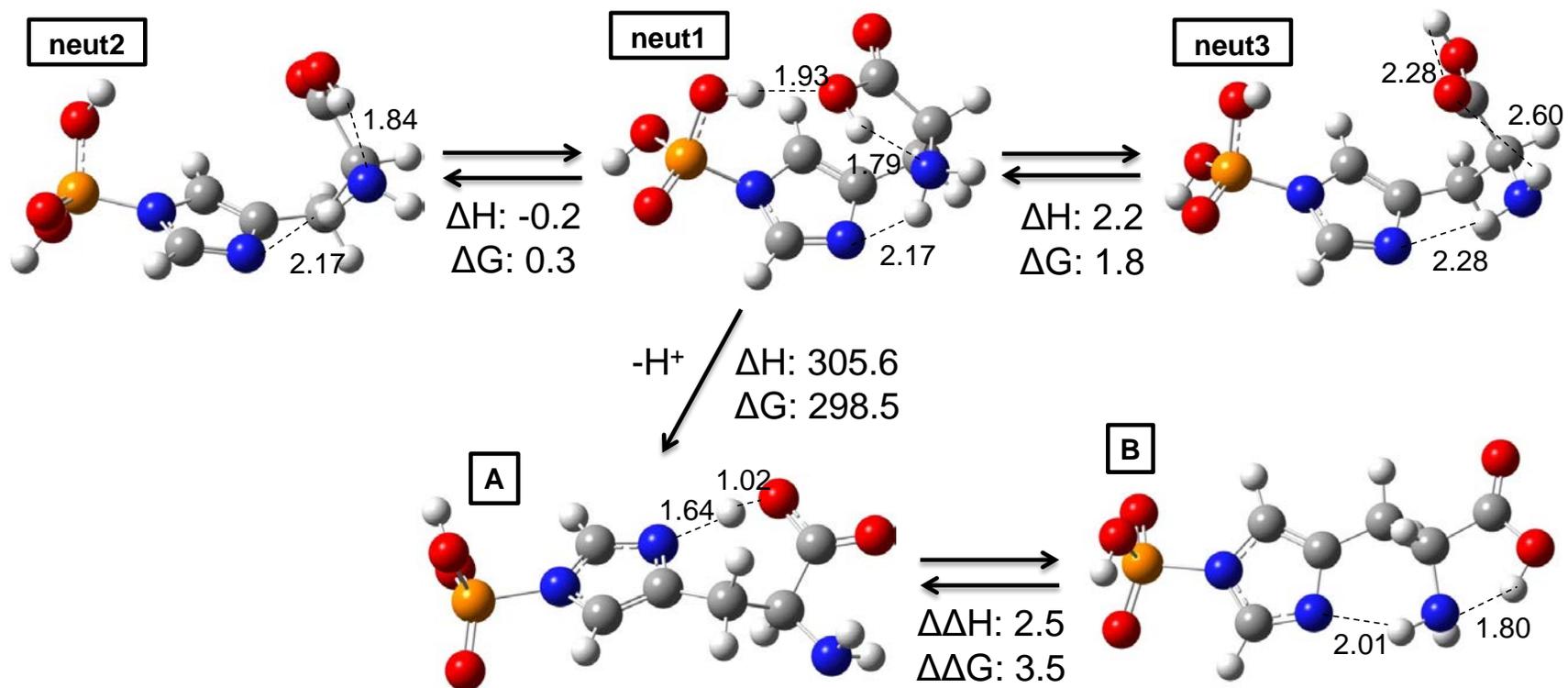
Phosphoglutamic acid



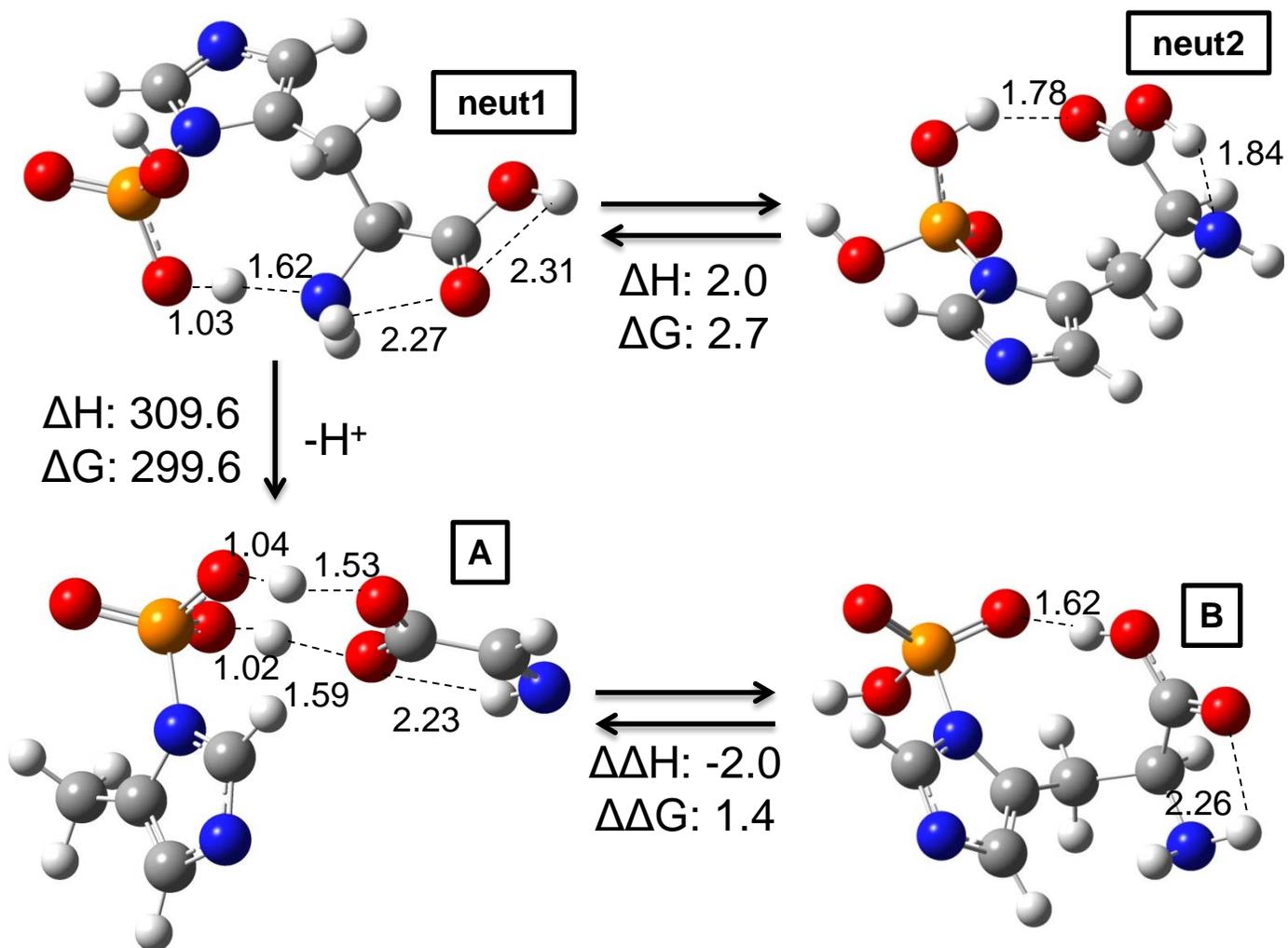
Phosphoglycine



Phosphohistidine τ



Phosphohistidine π



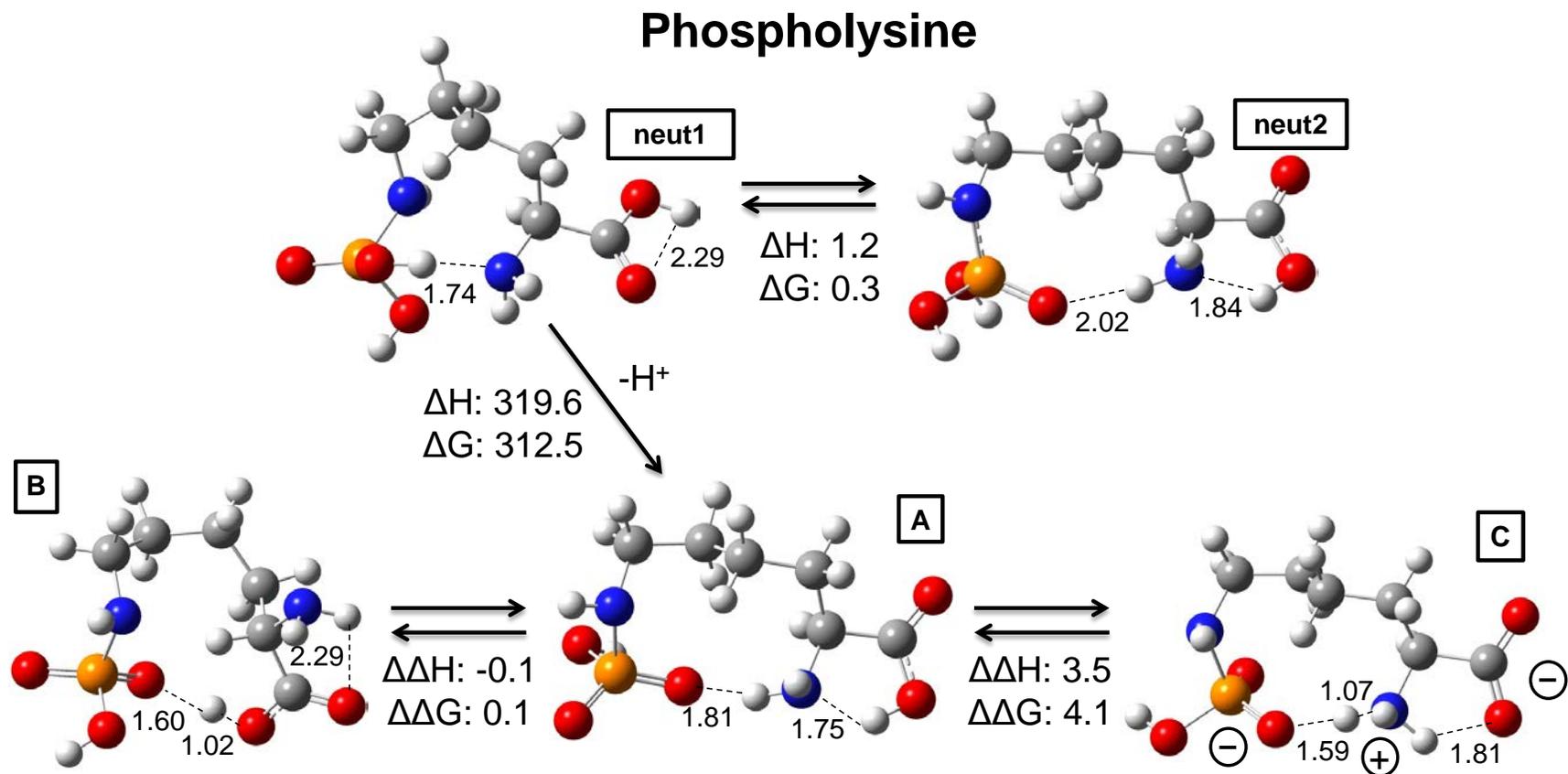
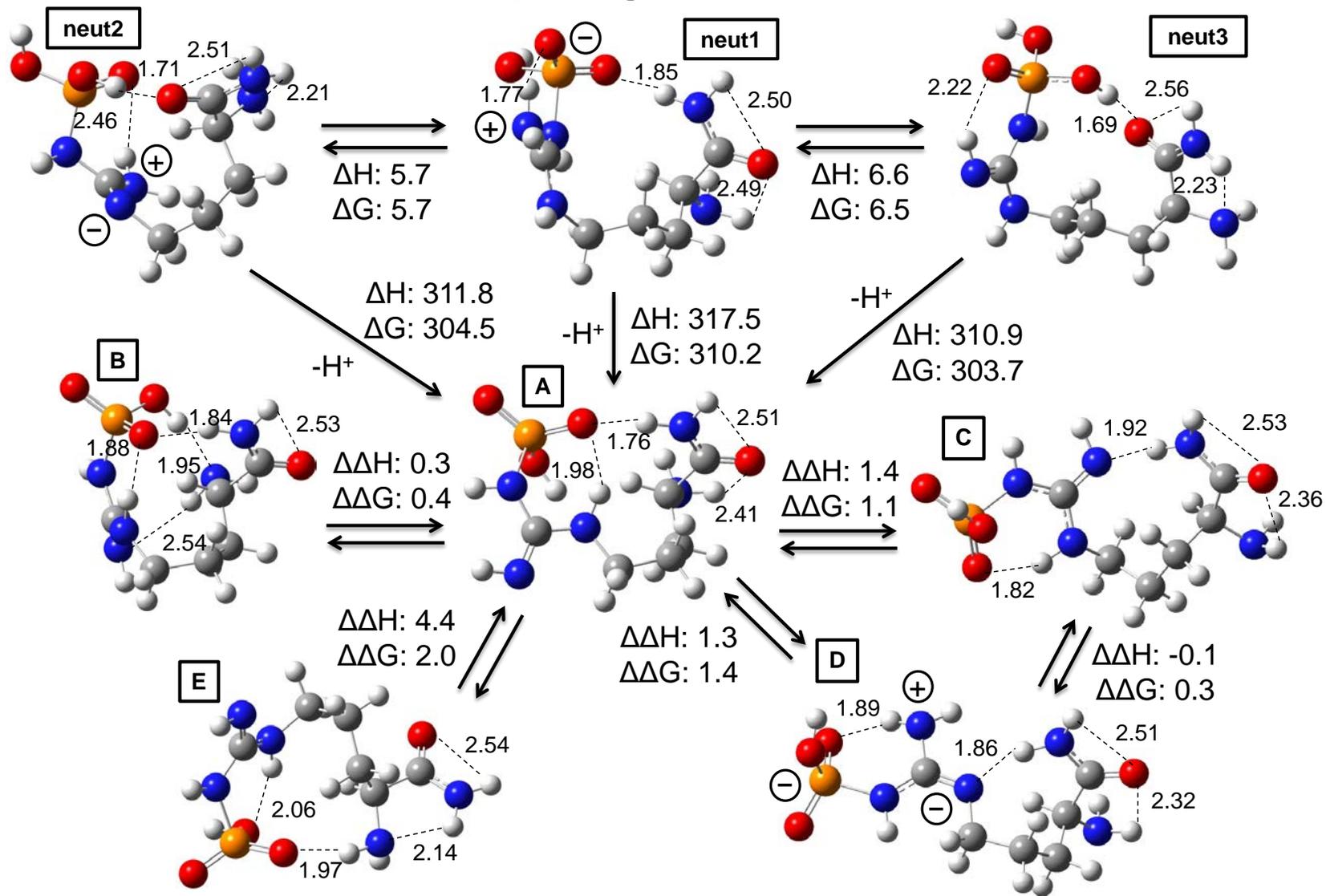
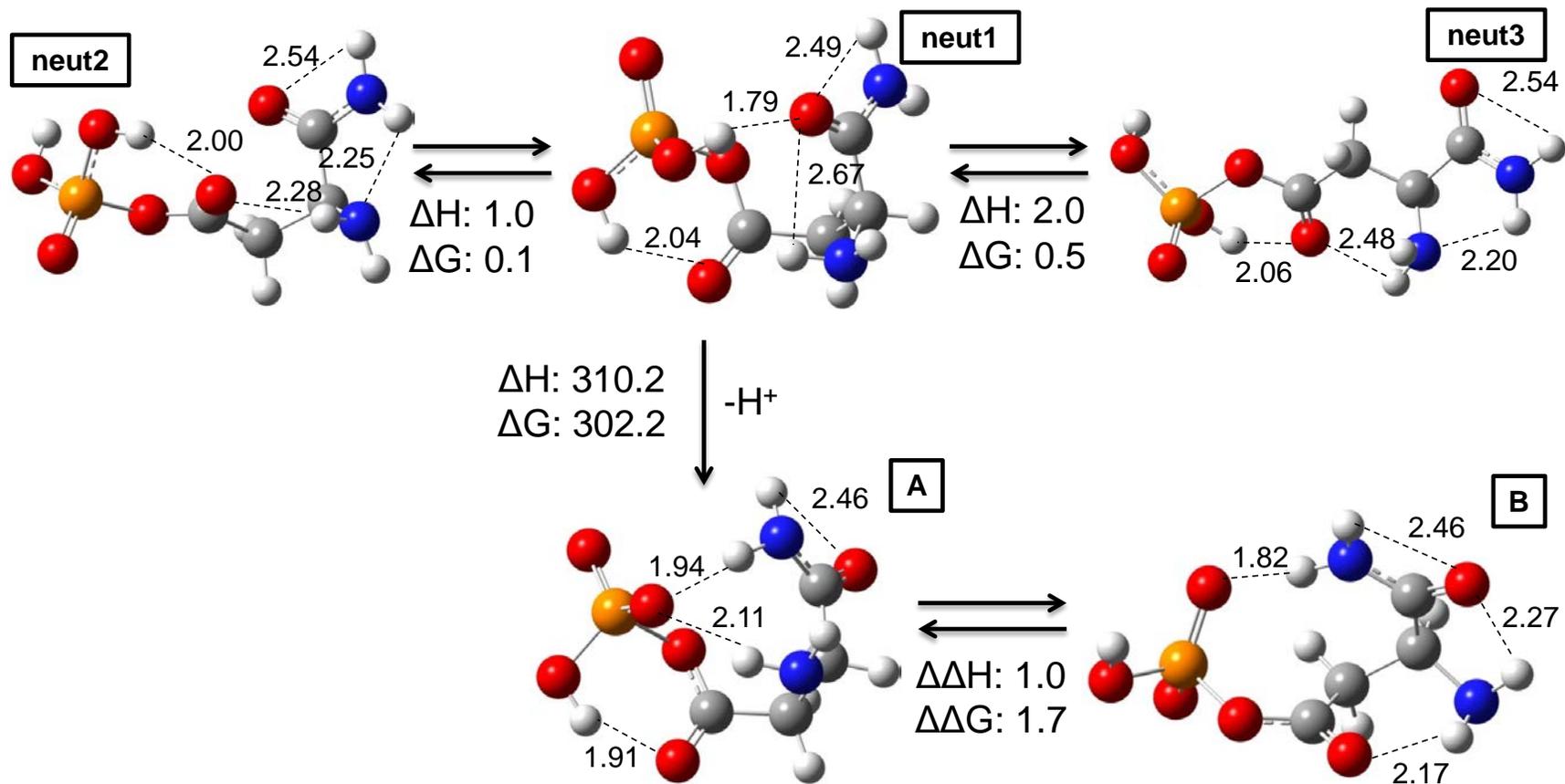


Figure 5.6. Optimized structures of the phosphorylated amino acids and their anions at the G3(MP2) level. GAs and relative energies are in kcal/mol. Important hydrogen bond distances are given in Å.

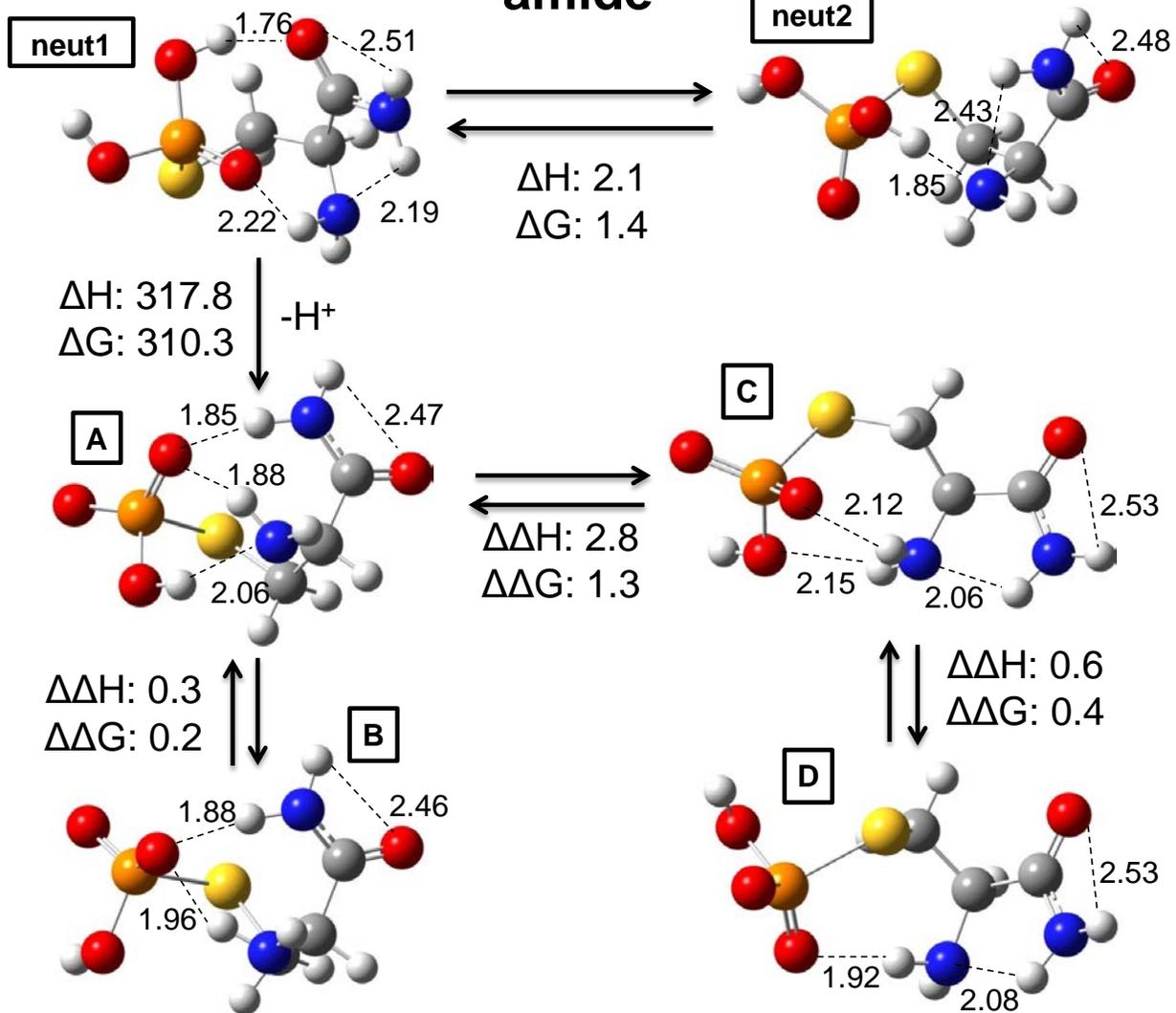
Phosphoarginine amide



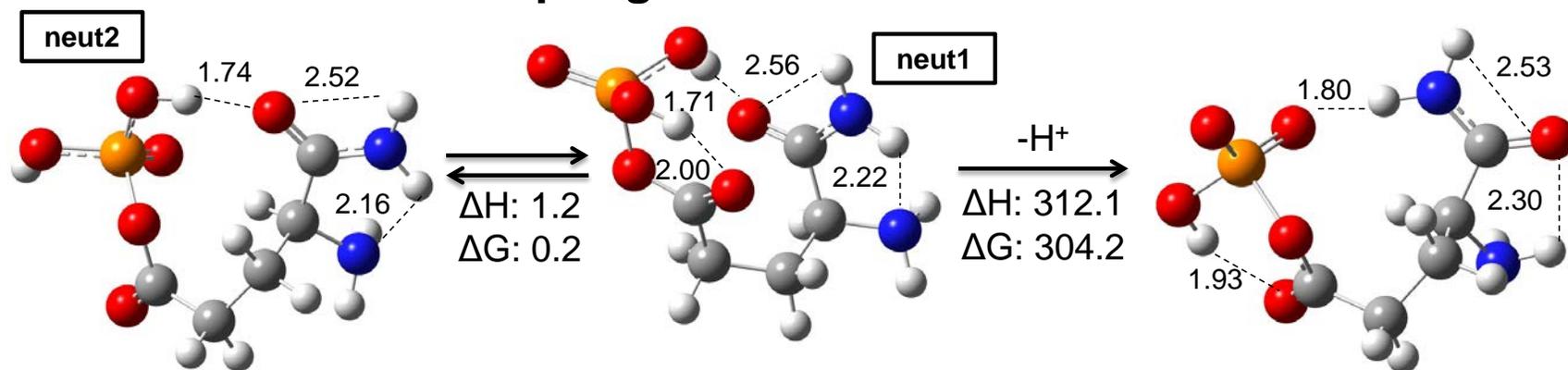
Phosphoaspartic acid amide



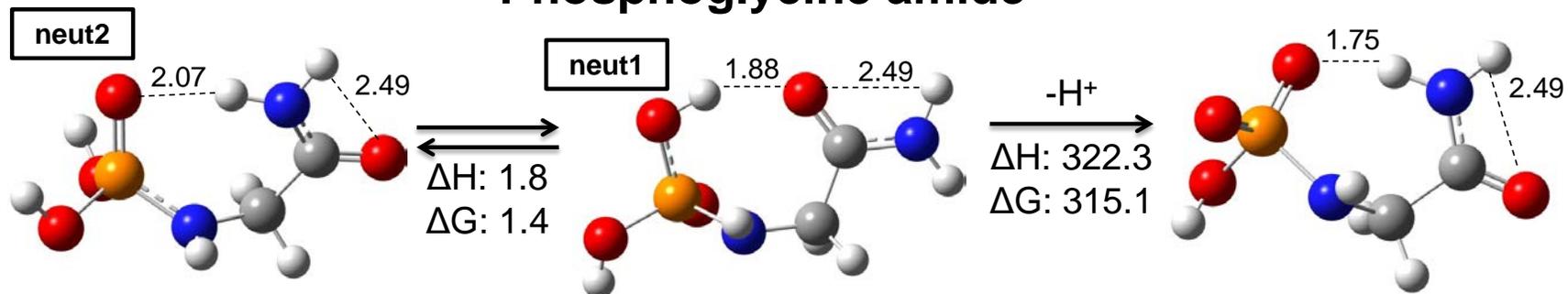
Phosphocysteine amide



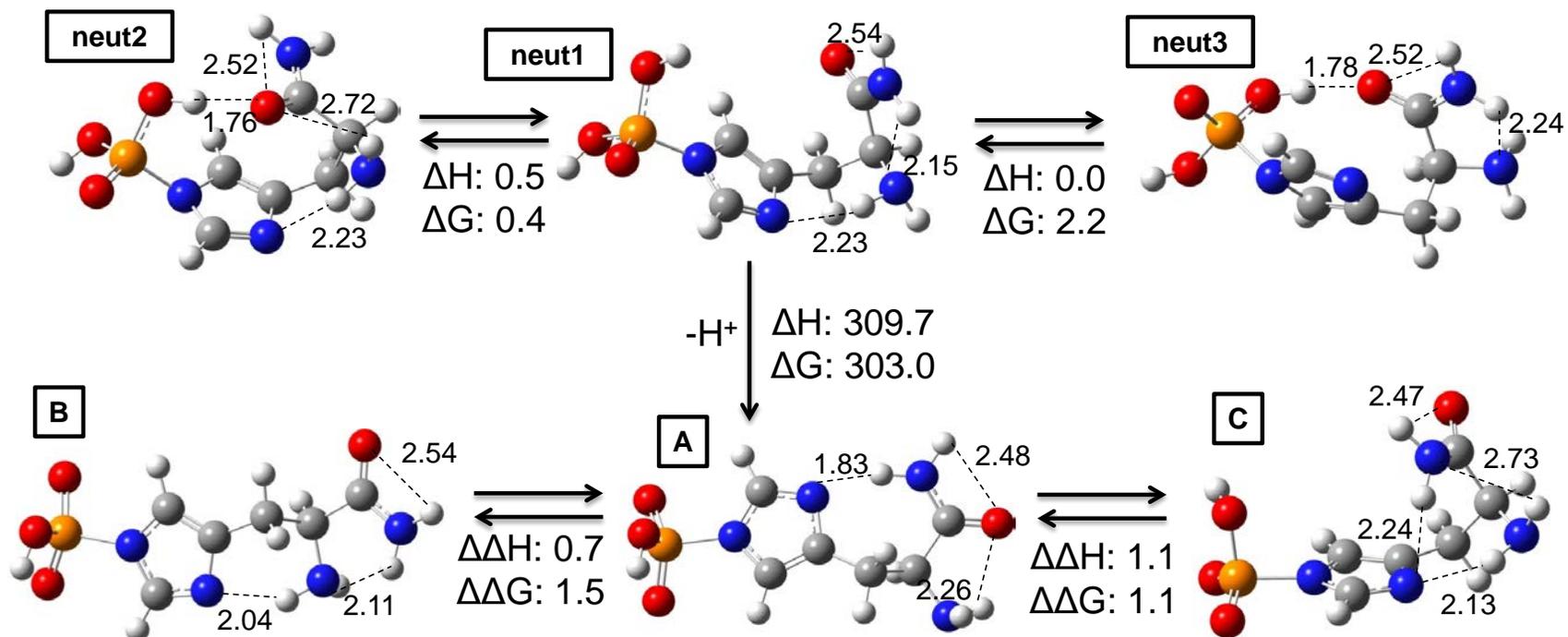
Phosphoglutamic acid amide



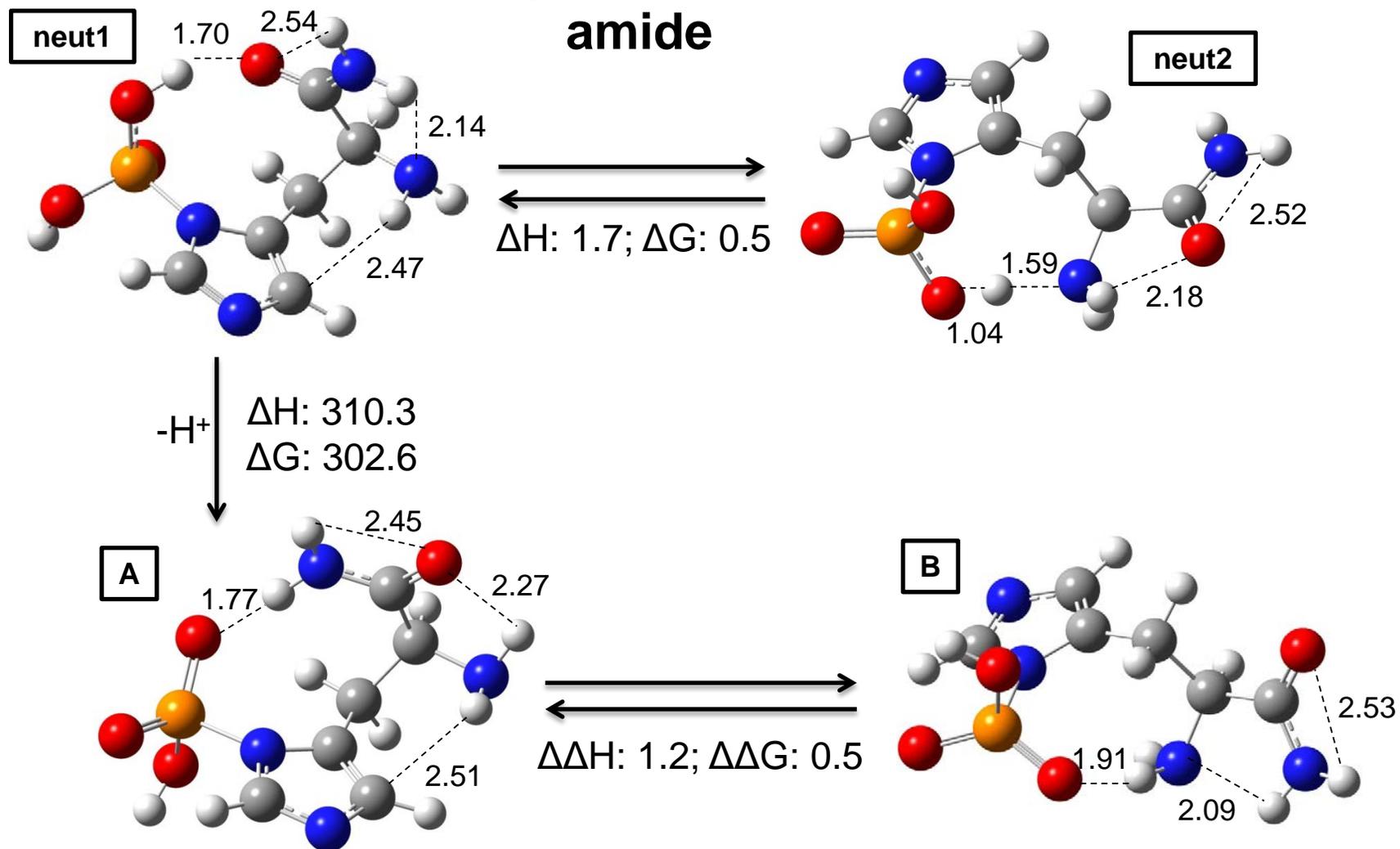
Phosphoglycine amide



Phosphohistidine τ amide



Phosphohistidine π amide



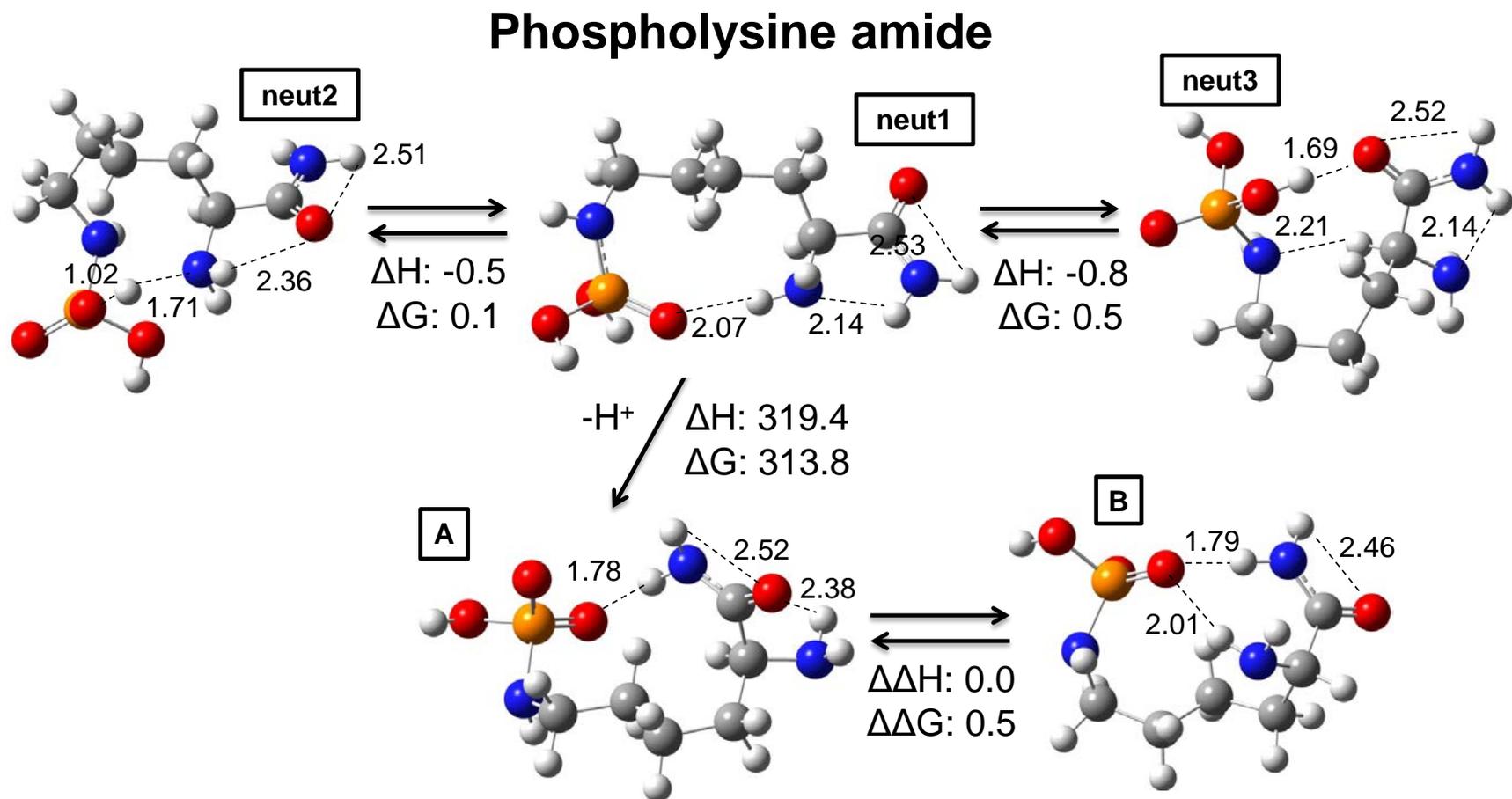


Figure 5.7. Optimized structures of the phosphorylated amino acid amides and their anions at the G3(MP2) level. GAs and relative energies are in kcal/mol. Important hydrogen bond distances are given in Å.

References

- ¹ Prabakaran, S.; Lippens, G.; Steen, H.; Gunawardena, J. Post translational Modification: Nature's Escape from Genetic Imprisonment and the Basis for Dynamic Information Encoding. *WIREs Syst Biol Med* **2012**, *4*, 565-583.
- ² Wang, Y-C., Peterson, S. E.; Loring, J. F. Protein Post-Translational Modification and Regulation of Pluripotency in Human Stem Cells. *Cell Research*, **2014**, *24*, 143-160.
- ³ Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **2002**, *298*, 1912-1916, 1933-1934.
- ⁴ Berg, J. M.; Tymoczko, J. L., Stryer, L. Biochemistry, 5th ed.; W H Freeman,: New York, 2002.
- ⁵ Engholm-Keller, K.; Larsen, M. R. Technologies and Challenges in Large-Scale Phosphoproteomics. *Proteomics* **2013**, *13*, 910-931.
- ⁶ Ciesla, J.; Fraczek, T.; Rode, W. Phosphorylation of Basic Amino Acid Residues in Proteins: Important but Easily Missed. *Acta Biochim Pol* **2011**, *58*, 137-147.
- ⁷ Palumbo, A. M.; Smith, S. A.; Kalcic, C. L.; Dantus, M.; Stemmer, P. M.; Reid, G. E. Tandem Mass Spectrometry Strategies for Phosphoproteome Analysis. *Mass Spectrom. Rev.* **2011**, *30*, 600-625.
- ⁸ Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F. Electrospray Ionization – Principles and Practice. *Mass Spectrom. Rev.* **1990**, *9*, 37-70.
- ⁹ Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers. *Anal. Chem.* **1991**, *63*, 1193A-1203A.
- ¹⁰ Janek, K.; Wenschuh, H.; Bienert, M; Krause, E. Phosphopeptide Analysis by Positive and Negative Ion Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun Mass Spectrom* **2001**, *15*, 1593-1599.
- ¹¹ Edelson-Averbukh, M.; Pipkorn, R.; Lehmann, W. D. Phosphate Group-Driven Fragmentation of Multiply Charged Phosphopeptide Anions. Improved Recognition of Peptides Phosphorylated at Serine, Threonine, Or Tyrosine by Negative Ion Electrospray Tandem Mass Spectrometry. *Anal. Chem.* **2006**, *78*, 1249-1256.
- ¹² Zhou, W.; Merrick, B. A.; Khaledi, M. G.; Tomer, K. B. Detection and Sequencing of Phosphopeptides Affinity Bound to Immobilized Metal Ion Beads by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 273-282.
- ¹³ Gunawardena, H. P.; Emory, J. F.; McLuckey, S. A. Phosphopeptide Anion Characterization Via Sequential Charge Inversion and Electron-Transfer Dissociation. *Anal. Chem.* **2006**, *78*, 3788-3793.

-
- ¹⁴ Annan, R. S.; Carr, S. A. Phosphopeptide Analysis by Matrix-Assisted Laser Desorption Time-of-Flight Mass Spectrometry. *Anal. Chem.* **1996**, *68*, 3413-3421.
- ¹⁵ Huddleston, M. J.; Annan, R. S.; Bean, M. F.; Carr, S. A. Selective Detection of Phosphopeptides in Complex Mixtures by Electrospray Liquid-Chromatography Mass-Spectrometry. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 710-719.
- ¹⁶ Reid, G. E.; Simpson, R. J.; O'Hair, R. A. J. Leaving Group and Gas Phase Neighboring Group Effects in the Side Chain Losses from Protonated Serine and its Derivatives. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 1047-1060.
- ¹⁷ Schlosser, A.; Pipkorn, R.; Bossemeyer, D.; Lehmann, W. D. Analysis of Protein Phosphorylation by a Combination of Elastase Digestion and Neutral Loss Tandem Mass Spectrometry. *Anal. Chem.* **2001**, *73*, 170-176.
- ¹⁸ Moyer, S. C.; Cotter, R. J.; Woods, A. S. Fragmentation of Phosphopeptides by Atmospheric Pressure MALDI and ESI/ion Trap Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 274-283.
- ¹⁹ Palumbo, A. M.; Reid, G. E. Evaluation of Gas-Phase Rearrangement and Competing Fragmentation Reactions on Protein Phosphorylation Site Assignment using Collision Induced Dissociation-MS/MS and MS³. *Anal. Chem.* **2008**, *80*, 9735-9747.
- ²⁰ Medzihradszky, K. F.; Trinidad, J. C. Unusual Fragmentation of Pro-Ser/Thr-Containing Peptides Detected in Collision-Induced Dissociation Spectra. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 602-607.
- ²¹ Beck, A.; Deeg, M.; Moeschel, K.; Schmidt, E. K.; Schleicher, E. D.; Voelter, W.; Haring, H. U.; Lehmann, R. Alkaline Liquid chromatography/electrospray Ionization Skimmer Collision-Induced Dissociation Mass Spectrometry for Phosphopeptide Screening. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2324-2333.
- ²² Annan, R. S.; Huddleston, M. J.; Verma, R.; Deshaies, R. J.; Carr, S. A. A Multidimensional Electrospray MS-Based Approach to Phosphopeptide Mapping. *Anal. Chem.* **2001**, *73*, 393-404.
- ²³ Tholey, A.; Reed, J.; Lehmann, W. D. Electrospray Tandem Mass Spectrometric Studies of Phosphopeptides and Phosphopeptide Analogues. *J. Mass Spectrom.* **1999**, *34*, 117-123.
- ²⁴ Ding, J.; Burkhart, W.; Kassel, D. B. Identification of Phosphorylated Peptides from Complex Mixtures using Negative-Ion Orifice-Potential Stepping and Capillary Liquid chromatography/electrospray Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 94-98.
- ²⁵ Beck, A.; Moeschel, K.; Deeg, M.; Haring, H.; Voelter, W.; Schleicher, E. D.; Lehmann, R. Identification of an in Vitro Insulin Receptor Substrate-1 Phosphorylation Site by Negative-Ion LC/ES-API-CID-MS Hybrid Scan Technique. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 401-405.

-
- ²⁶ Neubauer, G.; Mann, M. Parent Ion Scans of Large Molecules. *J. Mass Spectrom.* **1997**, *32*, 94-98.
- ²⁷ Edelson-Averbukh, M.; Pipkorn, R.; Lehmann, W. D. Analysis of Protein Phosphorylation in the Regions of Consecutive Serine/Threonine Residues by Negative Ion Electrospray Collision-Induced Dissociation. Approach to Pinpointing of Phosphorylation Sites. *Anal. Chem.* **2007**, *79*, 3476-3486.
- ²⁸ Edelson-Averbukh, M.; Shevchenko, A.; Pipkorn, R.; Lehmann, W. D. Gas-Phase Intramolecular Phosphate Shift in Phosphotyrosine-Containing Peptide Monoanions. *Anal. Chem.* **2009**, *81*, 4369-4381.
- ²⁹ Edelson-Averbukh, M.; Shevchenko, A.; Pipkorn, R.; Lehmann, W. D. Discrimination between Peptide O-Sulfo- and O-Phosphotyrosine Residues by Negative Ion Mode Electrospray Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 2256-2268.
- ³⁰ Tran, T. T. N.; Wang, T.; Hack, S.; Bowle, J. H. Diagnostic Cyclisation Reactions which Follow Phosphate Transfer to Carboxylate Anion Centres for Energised $[M-H]^-$ Anions of pTyr-Containing Peptides. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 2489-2499.
- ³¹ Tran, T. T. N.; Wang, T.; Hack, S.; Hoffmann, P.; Bowie, J. H. Can Collision-Induced Negative-Ion Fragmentations of $[M-H]^-$ Anions be used to Identify Phosphorylation Sites in Peptides?. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 3537-3548.
- ³² Wang, T.; Andreatza, H. J.; Bilusich, D.; Bowie, J. H. Negative Ion Fragmentations of Deprotonated Peptides Containing Post-Translational Modifications. an Unusual cyclization/rearrangement Involving Phosphotyrosine; a Joint Experimental and Theoretical Study. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 1669-1677.
- ³³ Andreatza, H. J.; Fitzgerald, M.; Bilusich, D.; Hoffmann, R.; Hoffmann, P.; Eichinger, P. C. H.; Bowie, J. H. Characteristic Negative Ion Fragmentations of Deprotonated Peptides Containing Post-Translational Modifications: Mono-Phosphorylated Ser, Thr and Tyr. A Joint Experimental and Theoretical Study. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3305-3312.
- ³⁴ Correia, C. F.; Balaj, P. O.; Scuderi, D.; Maitre, P.; Ohanessian, G. Vibrational Signatures of Protonated, Phosphorylated Amino Acids in the Gas Phase. *J. Am. Chem.Soc.* **2008**, *130*, 3359-3370.
- ³⁵ Correia, C. F.; Clavaguera, C.; Erlekam, U.; Scuderi, D.; Ohanessian, G. IRMPD Spectroscopy of a Protonated, Phosphorylated Dipeptide. *ChemPhysChem* **2008**, *9*, 2564-2573.
- ³⁶ Scuderi, D.; Correia, C. F.; Balaj, O. P.; Ohanessian, G.; Lemaire, J.; Maitre, P. Structural Characterization by IRMPD Spectroscopy and DFT Calculations of Deprotonated Phosphorylated Amino Acids in the Gas Phase. *ChemPhysChem* **2009**, *10*, 1630-1641.

-
- ³⁷ Cimas, A.; Maitre, P.; Ohanessian, G.; Gaigeot, M.-P. Molecular Dynamics and Room Temperature Vibrational Properties of Deprotonated Phosphorylated Serine. *J. Chem. Theory Comput.* **2009**, *5*, 2388-2400.
- ³⁸ Cimas, A.; Gaigeot, M.-P. DFT-MD and Vibrational Anaharmonicities of a Phosphorylated Amino Acid. Success and Failure. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3501-3510.
- ³⁹ Scuderi, D.; Bakker, J. M.; Durand, S.; Maitre, P.; Sharma, P.; Martens, J. K.; Nicol, E.; Clavaguera, C.; Ohanessian, G. Structure of Singly Hydrated, Protonated Phospho-tyrosine. *Int. J. Mass Spectrom.* **2011**, *308*, 338-347.
- ⁴⁰ Zachariou, M.; Traverso, I.; Spiccia, L.; Hearn, M. T. W. Potentiometric Investigations into the Acid-Base and Metal Ion Binding Properties of Immobilized Metal Ion Affinity Chromatographic (IMAC) Adsorbents. *J. Phys. Chem.* **1996**, *100*, 12680-12690.
- ⁴¹ Smiechowski, M. Theoretical pK_a Prediction of *O*-phosphoserine in Aqueous Solution. *Chem. Phys. Lett.* **2010**, *501*, 123-129.
- ⁴² Li, Z.; Matus, M. H.; Velazquez, H. A.; Dixon, D. A.; Cassady, C. J. Gas-Phase Acidities of Aspartic Acid, Glutamic Acid, and their Amino Acid Amides. *Int. J. Mass Spectrom.* **2007**, *265*, 213-223.
- ⁴³ Stover, M. L.; Jackson, V.; Matus, M.; Adams, M.; Cassady, C. J.; Dixon, D. A. Fundamental Thermochemical Properties of Amino Acids: Gas-Phase and Aqueous Acidities and Gas-Phase Heats of Formation. *J. Phys. Chem. B* **2012**, *116*, 2905-2916.
- ⁴⁴ Bokatzian-Johnson, S. S.; Stover, M. L.; Dixon, D. A.; Cassady, C. J. Gas-Phase Deprotonation of the Peptide Backbone for Tripeptides and their Methyl Esters with Hydrogen and Methyl Side Chains. *J. Phys. Chem. B* **2012**, *116*, 14844-14858.
- ⁴⁵ Bokatzian-Johnson, S. S.; Stover, M. L.; Plummer, C. E.; Dixon, D. A.; Cassady, C. J. An experimental and Computational Investigation into the Gas-Phase Acidities of Tyrosine and Phenylalanine: Three Structures for Deprotonated Tyrosine. *J. Phys. Chem. B* **2014**, *118*, 12630-12643.
- ⁴⁶ O'Hair, R. J.; Bowie, J. H.; Gronert, S. Gas Phase Acidity of the Alpha-Amino Acids. *Int. J. Mass Spectrom. Ion Proc.* **1992**, *117*, 23-36.
- ⁴⁷ Tian, Z. X.; Wang, X. B.; Wang, L. S.; Kass, S. R. Are Carboxyl Groups the Most Acidic Sites in Amino Acids? Gas-Phase Acidities, Photoelectron Spectra, and Computations on Tyrosine, *p*-Hydroxybenzoic Acid, and their Conjugate Bases *J. Am. Chem. Soc.* **2009**, *131*, 1174-1181.
- ⁴⁸ Tian, Z. X.; Pawlow, A.; Poutsma, J. C.; Kass, S. R. Are Carboxyl Groups the Most Acidic Sites in Amino Acids? Gas-Phase Acidity, H/D Exchange Experiments, and Computations on Cysteine and its Conjugate Base *J. Am. Chem. Soc.* **2007**, *129*, 5403-5407.

-
- ⁴⁹ Kendall, R. A.; Dunning, Jr., T. H.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796-6806.
- ⁵⁰ Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999-3093.
- ⁵¹ Klamt, A.; Schüürmann, G. J. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *Chem. Soc. Perkin Trans.* **1993**, *2*, 799-805.
- ⁵² Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Fox, D. J. *Gaussian 09*, revision C.01; Gaussian, Inc.: Wallingford, CT, 2010.
- ⁵³ Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- ⁵⁴ Lee, C.; Yang, W.; Parr, R. G. Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy. *Physical Review B.* **1988**, *37*, 785-789.
- ⁵⁵ Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. Optimization of Gaussian-Type Basis-Sets for Local Spin-Density Functionals Calculations. 1. Boron Through Neon, Optimization Technique and Validation. *Can. J. Chem.* **1992**, *70*, 560-571.
- ⁵⁶ Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. Gaussian-3 Theory Using Reduced Moller-Plesset Order. *J. Chem. Phys.* **1999**, *110*, 4703-4709.
- ⁵⁷ Dixon, D. A.; Feller, D.; Peterson, K. A. In *A Practical Guide to Reliable First Principles Computational Thermochemistry Predictions Across the Periodic Table*; Wheeler, R. A., Section Ed. Tschumper, G. S., Eds.; Annual Reports in Computational Chemistry; Elsevier: Amsterdam, 2012; Vol. 8, pp 1-28.
- ⁵⁸ Feller, D.; Peterson, K. A.; Dixon, D. A. Further Benchmarks of a Composite, Convergent, Statistically Calibrated Coupled-Cluster-Based Approach for Thermochemical and Spectroscopic Studies. *Mol. Phys.* **2012**, *110*, 2381-2399.

-
- ⁵⁹ Peterson, K. A.; Feller, D.; Dixon, D. A. Chemical Accuracy in Ab Initio Thermochemistry and Spectroscopy: Current Strategies and Future Challenges. *Theor. Chem. Acc.* **2012**, *131*, 1-20.
- ⁶⁰ Feller, D.; Peterson, K. A.; Dixon, D. A. A Survey of Factors Contributing to Accurate Theoretical Predictions of Atomization Energies and Molecular Structures. *J. Chem. Phys.* **2008**, *129*, 204105/1-204105/32.
- ⁶¹ Feller, D.; Dixon, D. A. Extended Benchmark Studies of Coupled Cluster Theory through Triple Excitations. *J. Chem. Phys.* **2001**, *115*, 3484-3496.
- ⁶² ChemAxon pKa Calculator Plugin. Marvin 6.0, 2015, <http://www.chemaxon.com>
- ⁶³ de Koning, L. J.; Nibbering, N. M. M.; van Orden, S. L.; Laukien, F. H. Mass Selection of Ions in a Fourier Transform Ion Cyclotron Resonance Trap using Correlated Harmonic Excitation Fields (CHEF). *Int. J. Mass Spectrom. Ion Proc.* **1997**, *165/166*, 209-219.
- ⁶⁴ Bartmess, J. E. In *Negative Ion Energetics Data*; Linstrom, P. J., Ed.; NIST Chemistry WebBook, NIST Standard Reference Database Number 69; National Institute of Standards and Technology: Gaithersburg MD, 20899, <http://webbook.nist.gov>, (retrieved April 14, 2015).
- ⁶⁵ Zhang, K.; Cassady, C. J.; Chung-Phillips, A. Ab Initio Studies of Neutral and Protonated Triglycines: Comparison of Calculated and Experimental Gas-Phase Basicity. *J. Am. Chem. Soc.* **1994**, *116*, 11512-11521.
- ⁶⁶ Bartmess, J. E.; Georgiadis, R. M. Empirical Methods for Determination of Ionization Gauge Relative Sensitivities for Different Gases. *Vacuum* **1983**, *33*, 149-153.
- ⁶⁷ Miller, K. J. Additivity Methods in Molecular Polarizability. *J. Am. Chem. Soc.* **1990**, *112*, 8533-8542.
- ⁶⁸ Su, T.; Chesnavich, W. J. Parametrization of the Ion-Polar Molecule Collision Rate-Constant by Trajectory Calculations. *J. Chem. Phys.* **1982**, *76*, 5183-5185.
- ⁶⁹ Su, T. Erratum: Trajectory Calculations of Ion-Polar Molecule Capture Rate Constants at Low Temperatures. *J. Chem. Phys.* **1988**, *89*, 5355-5355.
- ⁷⁰ Bouchoux, G.; Salpin, J. Y.; Leblanc, D. A Relationship between the Kinetics and Thermochemistry of Proton Transfer Reactions in the Gas Phase. *Int. J. Mass Spectrom. Ion Proc.* **1996**, *153*, 37-48.
- ⁷¹ Bouchoux, G.; Salpin, J. Y. Re-Evaluated Gas Phase Basicity and Proton Affinity Data from the Thermokinetic Method. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 932-936.
- ⁷² 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.

-
- ⁷³ Bouchoux, G.; Buisson, D.; Bourcier, S.; Sablier, M. Application of the Kinetic Method to Bifunctional Bases ESI Tandem Quadrupole Experiments. *Int. J. Mass Spectrom.* **2003**, *228*, 1035-1054.
- ⁷⁴ Plummer, C. E.; Stover, M. L.; Bokatzian, S. S.; Davis, J. T.; Dixon, D. A.; Cassady, C. J. An Experimental and Computational Study of the Gas-Phase Acidities of the Common Amino Acid Amides. *J. Phys. Chem. B* (submitted, May 2015)
- ⁷⁵ Oomens, J.; Steill, J. D.; Redlich, B. Gas-Phase IR Spectroscopy of Deprotonated Amino Acids. *J. Am. Chem. Soc.* **2009**, *131*, 4310-4319.
- ⁷⁶ Alexeev, Y.; Windus, T. L.; Zhan, C.-G.; Dixon, D. A. Accurate Heats of Formation and Acidities for H₃PO₄, H₂SO₄, and H₂CO₃ From Ab Initio Electron Structure Calculations. *Int. J. Quant. Chem.* **2005**, *102*, 775-784 erratum, **2005**, *104*, 379-380.
- ⁷⁷ Gutowski, K. E.; Dixon, D. A. Ab Initio Prediction of the Gas- and Solution-Phase Acidities of Strong Brønsted Acids: The Calculation of pK_a Values Less Than -10. *J. Phys. Chem. A*, **2006**, *110*, 12044-12054.
- ⁷⁸ Chase, M. W., Jr. NIST-JANAF Thermochemical Tables, 4th ed.; Journal of Physical Chemistry Reference Data, Monograph 9; American Institute of Physics: Woodbury, NY, 1998; Suppl 1, pp 1-1951.
- ⁷⁹ Ngauv, S. N.; Sabbah, R.; Laffitte, M. Thermodynamique de composés azotes. III. Etude thermochimique de la glycine et de la l- α -alanine. *Thermochim. Acta*, **1977**, *20*, 371-380.
- ⁸⁰ Suenram, R. D.; Lovas, F. J. Millimeter Wave Spectrum of Glycine. *J Mol Spectros.* **1978**, *72*, 372-382.
- ⁸¹ Locke, M. J.; McIver, R. T., Jr. Effect of Solvation on the Acid/Base Properties of Glycine. *J. Am. Chem. Soc.* **1983**, *105*, 4226-4232.
- ⁸² Ding, Y.; Kroug-Jespersen, K. The Glycine Zwitterion does not Exist in the Gas Phase: Results from a Detailed Ab Initio Electronic Structure Study. *Chem. Phys. Lett.* **1992**, *199*, 261-266.
- ⁸³ Price, W. D.; Jockusch, R. A.; Williams, E. R. Is Arginine a Zwitterion in the Gas Phase? *J. Am. Chem. Soc.* **1997**, *119*, 11988-11989.
- ⁸⁴ Chappo, C. J.; Paul, J. B.; Provencal, R. A.; Roth, K.; Saykally, R. J. Is Arginine Zwitterionic or Neutral in the Gas Phase? Results from IR Cavity Ring Down Spectroscopy. *J. Am. Chem. Soc.* **1998**, *120*, 12956-12957.
- ⁸⁵ Wang, X.-B.; Broadus, K. M.; Wang, L.-S.; Kass, S. R. Photodetachment of the First Zwitterionic Anions in the Gas Phase: Probing Intramolecular Coulomb Repulsion and Attraction. *J. Am. Chem. Soc.* **2000**, *122*, 8306-8306.

**CHAPTER 5 APPENDIX: GAS- AND SOLUTION-PHASE ACIDITIES OF
PHOSPHORYLATED AMINO ACIDS AND THEIR AMIDES**

H_{298} and G_{298} total energies for all neutrals and anions at the G3(MP2) level and the heats of formation of the amino acid amides.

Table A5.1. H_{298} and G_{298} Total Energies for the Phosphorylated Amino Acids and Anions at the G3(MP2) level (a.u.).

Amino Acid	H_{298}	G_{298}
Phosphoarginine neutral 1	-1172.739730	-1172.805972
Phosphoarginine neutral 2	-1172.739549	-1172.805767
Phosphoarginine neutral 3	-1172.736111	-1172.803323
Phosphoarginine anion A	-1172.254309	-1172.320544
Phosphoarginine anion B	-1172.254707	-1172.319129
Phosphoarginine anion C	-1172.252852	-1172.318613
Phosphoarginine anion D	-1172.253246	-1172.318311
Phosphoarginine anion E	-1172.252245	-1172.317889
Phosphoaspartic acid neutral 1	-1078.750732	-1078.808426
Phosphoaspartic acid neutral 2	-1078.749463	-1078.806882
Phosphoaspartic acid neutral 3	-1078.748081	-1078.806486
Phosphoaspartic acid anion A	-1078.261015	-1078.316515
Phosphoaspartic acid anion B	-1078.260363	-1078.315338
Phosphocysteine neutral 1	-1288.138768	-1288.192742
Phosphocysteine neutral 2	-1288.138609	-1288.190677
Phosphocysteine neutral 3	-1288.136495	-1288.189482
Phosphocysteine anion A	-1287.642688	-1287.694279
Phosphocysteine anion B	-1287.641497	-1287.694124
Phosphocysteine anion C	-1287.643233	-1287.694005
Phosphocysteine anion D	-1287.640141	-1287.693724
Phosphocysteine anion E	-1287.639891	-1287.692555
Phosphoglutamic acid neutral 1	-1117.989221	-1118.048779
Phosphoglutamic acid neutral 2	-1117.989408	-1118.047782
Phosphoglutamic acid neutral 3	-1117.984826	-1118.047606
Phosphoglutamic acid neutral 4	-1117.986333	-1118.046004
Phosphoglutamic acid anion	-1117.504057	-1117.561368
Phosphoglycine neutral	-851.152015	-851.199247
Phosphoglycine anion	-850.651499	-850.697074
Phosphohistidine τ neutral 1	-1115.075788	-1115.138223
Phosphohistidine τ neutral 2	-1115.076070	-1115.137782
Phosphohistidine τ neutral 3	-1115.072356	-1115.135306

Phosphohistidine τ anion A	-1114.591098	-1114.652469
Phosphohistidine τ anion B	-1114.587224	-1114.647023
Phosphohistidine π neutral 1	-1115.080489	-1115.139915
Phosphohistidine π neutral 2	-1115.077235	-1115.135661
Phosphohistidine π anion A	-1114.589418	-1114.652481
Phosphohistidine π anion B	-1114.592600	-1114.650301
Phospholysine neutral 1	-1063.368110	-1063.429875
Phospholysine neutral 2	-1063.366255	-1063.429437
Phospholysine anion A	-1062.861115	-1062.921857
Phospholysine anion B	-1062.861280	-1062.921678
Phospholysine anion C	-1062.855526	-1062.915373
Phosphoserine neutral 1	-965.542866	-965.595468
Phosphoserine neutral 2	-965.540858	-965.592593
Phosphoserine anion A	-965.042117	-965.092781
Phosphoserine anion B	-965.041604	-965.090650
Phosphoserine anion C	-965.038639	-965.090267
Phosphoserine anion D	-965.039277	-965.088804
Phosphoserine anion E	-965.038411	-965.086949
Phosphoserine anion F	-965.035143	-965.085814
Phosphothreonine neutral 1	-1004.782264	-1004.838200
Phosphothreonine neutral 2	-1004.782534	-1004.836938
Phosphothreonine anion A	-1004.283862	-1004.336952
Phosphothreonine anion B	-1004.281871	-1004.334191
Phosphothreonine anion C	-1004.276986	-1004.331453
Phosphothreonine anion D	-1004.277258	-1004.330519
Phosphothreonine anion E	-1004.278770	-1004.330365
Phosphothreonine anion F	-1004.274309	-1004.328489
Phosphotyrosine neutral 1	-1196.191537	-1196.257812
Phosphotyrosine neutral 2	-1196.188281	-1196.256035
Phosphotyrosine anion A	-1195.692770	-1195.759070
Phosphotyrosine anion B	-1195.691500	-1195.752734
Phosphotyrosine anion C	-1195.687563	-1195.752642
Phosphotyrosine anion D	-1195.676268	-1195.741607
Phosphotyrosine anion E	-1195.662967	-1195.728298
Phosphotyrosine anion F	-1195.658425	-1195.726771

Table A5.2. . H_{298} and G_{298} Total Energies for the Phosphorylated Amino Acid Amides and Anions at the G3(MP2) level (a.u.).

Amino Acid	H_{298}	G_{298}
Phosphoarginine neutral 1	-1152.882874	-1152.950038
Phosphoarginine neutral 2	-1152.873854	-1152.941006
Phosphoarginine neutral 3	-1152.872364	-1152.939627
Phosphoarginine anion A	-1152.379274	-1152.445677
Phosphoarginine anion B	-1152.378854	-1152.445000
Phosphoarginine anion C	-1152.377011	-1152.443929
Phosphoarginine anion D	-1152.377187	-1152.443512
Phosphoarginine anion E	-1152.372339	-1152.442485
Phosphoaspartic acid neutral 1	-1058.884336	-1058.941198
Phosphoaspartic acid neutral 2	-1058.882795	-1058.941065
Phosphoaspartic acid neutral 3	-1058.881225	-1058.940333
Phosphoaspartic acid anion A	-1058.392371	-1058.449640
Phosphoaspartic acid anion B	-1058.390756	-1058.446872
Phosphocysteine neutral 1	-1268.274236	-1268.326709
Phosphocysteine neutral 2	-1268.270881	-1268.324482
Phosphocysteine anion A	-1267.770179	-1267.822195
Phosphocysteine anion B	-1267.821827	-1267.769644
Phosphocysteine anion C	-1267.765665	-1267.820089
Phosphocysteine anion D	-1267.764807	-1267.819493
Phosphoglutamic acid neutral 1	-1098.126027	-1098.184428
Phosphoglutamic acid neutral 2	-1098.124181	-1098.184170
Phosphoglutamic acid anion	-1097.630996	-1097.689687
Phosphoglycine neutral 1	-831.285318	-831.332701
Phosphoglycine neutral 2	-831.282385	-831.330464
Phosphoglycine anion	-830.774074	-830.820507
Phosphohistidine τ neutral 1	-1095.206559	-1095.269385
Phosphohistidine τ neutral 2	-1095.205835	-1095.268747
Phosphohistidine τ neutral 3	-1095.206486	-1095.265934
Phosphohistidine τ anion A	-1094.715410	-1094.776443
Phosphohistidine τ anion B	-1094.713585	-1094.774742
Phosphohistidine τ anion C	-1094.714206	-1094.774100

Phosphohistidine π neutral 1	-1095.213840	-1095.272340
Phosphohistidine π neutral 2	-1095.211207	-1095.271603
Phosphohistidine π anion A	-1094.721764	-1094.780143
Phosphohistidine π anion B	-1094.719741	-1094.779237
Phospholysine neutral 1	-1043.496225	-1043.560709
Phospholysine neutral 2	-1043.497092	-1043.560519
Phospholysine neutral 3	-1043.497550	-1043.559914
Phospholysine anion A	-1042.989620	-1043.050549
Phospholysine anion B	-1042.989618	-1043.049882
Phosphoserine neutral 1	-945.676612	-945.728995
Phosphoserine neutral 2	-945.671485	-945.725175
Phosphoserine anion A	-945.168313	-945.219902
Phosphoserine anion B	-945.161928	-945.215233
Phosphoserine anion C	-945.141817	-945.192838
Phosphothreonine neutral 1	-984.918634	-984.974016
Phosphothreonine anion A	-984.408647	-984.463267
Phosphothreonine anion B	-984.407945	-984.462354
Phosphothreonine anion C	-984.383241	-984.437849
Phosphotyrosine neutral 1	-1176.323942	-1176.391264
Phosphotyrosine neutral 2	-1176.322329	-1176.389880
Phosphotyrosine anion	-1175.824516	-1175.892079

Table A5.3. Heats of Formation in kcal/mol at the G3(MP2) level of the Amino Acid Amides

Amino Acid Amide	ΔH_f 298 K
arginine	-47.6
aspartic acid	-144.5
cysteine	-50.1
glutamic acid	-150.5
glycine	-48.1
histidine τ	-21.6
histidine π	-18.7
lysine	-62.5
serine	-94.1
threonine	-103.5

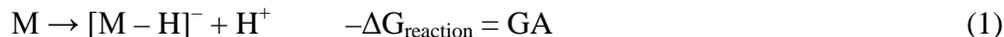
CHAPTER 6: GAS-PHASE DEPROTONATION OF THE PEPTIDE BACKBONE FOR TRIPEPTIDES AND THEIR METHYL ESTERS WITH HYDROGEN AND METHYL SIDE CHAINS

6.1 Introduction Proton transfer processes are of importance to the biological activities, three-dimensional structures, and physical properties of peptides in solution.^{1,2,3} Protonation and deprotonation is also important in gas-phase experiments employing mass spectrometry, which has been widely used for peptide sequencing in the past two decades.^{4,5,6} Although not as commonly employed as protonated peptide fragmentation, deprotonated peptide dissociation can also be used for sequencing.^{7,8,9,10,11,12}

Deprotonation of a peptide is generally accepted to take place at the C-terminal carboxylic acid group or other acidic sites such as the side chains of glutamic acid, aspartic acid, and cysteine residues.^{12,13,14} However, gas-phase deprotonation is quite universal and even neutral and basic peptides deprotonate by electrospray ionization (ESI), matrix-assisted laser desorption ionization (MALDI), and fast atom bombardment (FAB).^{7,15,16,17} For example, highly basic peptides without an acidic site (e.g., myosin kinase inhibiting peptide and substance P) deprotonate readily in the gas phase.¹⁵ Therefore, typical mass spectrometry ionization techniques are capable of deprotonating a peptide somewhere along the backbone, such as at the amide nitrogen or methylene carbon. Deprotonation of a peptide at a position other than the C-terminus may have a significant effect on fragmentation pathways because most dissociation in mass spectrometry is charge-directed.^{18,19}

Exploration of alternate deprotonation sites for peptides may enhance the collective understanding of the mechanistic pathways of peptide dissociation. Various mechanisms proposed for deprotonated peptide fragmentations have involved both deprotonation at backbone methylene carbons and at backbone amide nitrogens.^{8,9,11,20,21,22,23,24,25,26} For example, cleavage of the C-N bond to form negative mode c-ions has been proposed to involve nitrogen deprotonation, while (O=C)-C bond cleavage to form negative mode b- and y-ions has been suggested to involve carbon deprotonation.²⁷ In addition, the energetics of these deprotonation processes have not been clearly established. Using $\Delta G^{\circ}_{\text{acid}}$ values for small molecules to estimate deprotonation of the peptide backbone suggests that deprotonation at a CH group adjacent to a carbonyl function requires ~5 kcal/mol less than deprotonation of a NH group adjacent to a carbonyl.²⁷ Harrison and coworkers²⁸ studied the formation of "b₂ ion structures derived by the loss of neutral glycine from deprotonated triglycine (GlyGlyGly) at the B3LYP and MP2 levels with the 6-31+G(d) basis set.

Determination of the gas-phase acidity (GA) for peptides that do not possess traditional acidic sites is important to understanding the energetic basis for deprotonation along the backbone. GA is defined as the negative Gibbs free energy change for the deprotonation reaction shown in Reaction 1 at 298 K:



Several experimental studies have measured the GAs of the 20 standard amino acids.^{13,29,30,31,32,33,34} The GAs of these amino acids have also been calculated at G3MP2 and lower levels.^{13,29,30,31,34,35,36,37,38,39,40,41} However, GA values for amino acids are of limited value in rep-

representing residues in peptides because the acidic site on most amino acids is the C-terminal carboxylic acid group, which is absent unless the residue is located at the peptide's C-terminus.

There have been few reports of the GAs of neutral peptides. Ren and coworkers^{14,42,43,44} have performed experimental and computational studies of the GAs and structures of several small peptides containing a cysteine residue. The highly acidic cysteine side chain serves as the deprotonation site. The position of the cysteine residue in the peptide has been found to affect the GA values. In addition, the calculated structures indicate that extensive hydrogen bonding stabilizes the structures.

The goal of the current work is to explore deprotonation along the peptide backbone. Six model tripeptides and their methyl esters were studied both experimentally and computationally. The amino acid side chains contained either hydrogens or methyl groups. Hydrogens were replaced with methyl groups at specific sites along the peptide backbone in order to obtain information about deprotonation sites and their effects on acidity and structure.

6.2 Experimental and Computational Methods *Mass Spectrometry* All experiments were performed on a Bruker (Billerica, MA, USA) Bruker BioApex 7e FT-ICR mass spectrometer with a 7.0 T superconducting magnet. Peptide solutions were prepared at 60 μM in a solvent system of 50:50:1 methanol:water:ammonium hydroxide and introduced to an Apollo API source (Bruker Daltonics, Billerica, MA) using a syringe pump set to deliver $\sim 90 \mu\text{L/hr}$. Electrospray ionization (ESI) employed a 3.5-4.0 kV potential with air as a heated (225 $^{\circ}\text{C}$) counter and parallel current drying gas. Ions were allowed to accumulate in a hexapole for $\sim 700 \text{ ms}$ before being transported to the ICR cell by electrostatic focusing.

Deprotonated ions, $[M - H]^-$, were isolated with correlated frequency ion ejection techniques⁴⁵ and then allowed to react with a reference compound that was introduced to the ICR cell at constant pressure through a leak valve. Each of the ions selected for study was reacted with a series of reference compounds that have known GAs.⁴⁶ Neutral pressures were in the range of $(1-20) \times 10^{-8}$ mbar and were measured with a calibrated ionization gauge.⁴⁷

Reaction rate constants, k_{exp} , were determined by observing the pseudo-first-order decay in reactant ion intensity as a function of time. In cases where deprotonation was in competition with proton-bound dimer formation, k_{exp} was determined by fitting the experimental reaction data as discussed previously.⁴⁸ Reported reaction efficiencies (RE) are the ratio of k_{exp} to collision rate constants that were obtained from the thermal capture trajectory calculation procedure of Su and Chesnavich.^{49,50} An RE value of 0.269 was used as the “break point,” where a reaction becomes exoergic and the GA is assigned; this selection of a break point follows the work of Bouchoux et al.⁵¹ and has been discussed previously.¹³

Peptide synthesis All peptides were used in their L- forms. Triglycine (GlyGlyGly), trialanine (AlaAlaAla), glycylalanyl glycine (GlyAlaGly), and alanyl glycylalanine (AlaGlyAla) were acquired from Sigma-Aldrich (St. Louis, MO, USA). Tri-2-methylalanine (AibAibAib) was synthesized in our laboratory by standard Fmoc procedures⁵² on an Advanced ChemTech Model 90 peptide synthesizer (Louisville, KY, USA). Our attempts to synthesize trisarcosine (SarSarSar) in-house proved unsuccessful and this peptide was custom synthesized by Neo Bioscience (Cambridge, MA, USA).

A simple acid-catalyzed esterification with methanol was employed to generate the methyl ester forms of the six tripeptides. This equilibrium process was driven to ester formation by

chemically removing water via reaction with acetic anhydride.⁵³ Typical conditions involved mixing 20 μL of concentrated hydrochloric acid, 100 μL of acetic anhydride, and 500 μL of methanol and allowing the reaction to proceed for 10 min at room temperature. The product was used without purification, except for isolation of the desired deprotonated molecular ion during mass spectrometry experiments.

Computational Methods Calculations were performed at the density functional theory (DFT) and correlated molecular orbital (MO) theory levels with the program Gaussian-09.⁵⁴ The geometries were initially optimized at the DFT level with the B3LYP exchange-correlation functional^{55,56} and the DZVP2 basis set.⁵⁷ The DFT calculations with this DFT optimized basis set were used to examine different conformations and to provide good starting structures for the more computationally expensive G3MP2 calculations. A range of conformers, in general up to 10, with as many hydrogen bonds as possible were examined starting from initial geometries based on our experience with the geometries of the amino acids.⁴⁰ In a number of cases, when there is no hydrogen bonding present due to methylation, fewer structures are possible and thus fewer were investigated. Vibrational frequencies were calculated to show that the structures are 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. The most stable conformers were determined by optimizing a range of structures. In our previous work on the GAs of amino acids^{13, 40} and inorganic and organics acids,⁵⁸ the high level G3(MP2) correlated molecular orbital method⁵⁹ gave agreement for the acidities with the experimental values to within about ± 1 kcal/mol. G3(MP2) has an additional advantage over DFT methods in terms of reliable predictions for these types of compounds because the correlated molecular orbital methods in G3(MP2)

perform better in the prediction of hydrogen bond energies as well as steric non-bonded interactions than do most widely-used DFT exchange-correlation functionals. Van der Waals surfaces and solvent accessible surfaces^{60,61,62,63} were generated using the program Jmol.⁶⁴

6.3 Results and Discussion Six model tripeptides were chosen for this experiment: GlyGlyGly, GlyAlaGly, AlaGlyAla, AlaAlaAla, AibAibAib, and SarSarSar. Their structures are shown in Figure 1. Trisarcosine, SarSarSar, is structurally similar to triglycine, GlyGlyGly, except that each of the amide nitrogens along the peptide backbone are methylated and the N-terminal nitrogen has one methyl group plus one hydrogen. (Sarcosine is also known as N-methylglycine.) Tri-2-methylalanine, AibAibAib, is similar to trialanine, AlaAlaAla, except that each of the alpha carbons along the peptide backbone has two methyl groups (i.e., no hydrogens present on any of the alpha carbons). The methyl esters of each of these peptides were also studied with the C-terminal carboxylic acid group (-COOH) converted to a methyl ester group (-COOCH₃).

Electrospray ionization (ESI) produced abundant $[M - H]^-$ for all six of the tripeptides. The peptide methyl esters produced much less abundant $[M - H]^-$, with the exception of AlaAlaAla-OMe which produced a $[M - H]^-$ signal of comparable (and sometimes greater) intensity than the peptide acids. AlaAlaAla-OMe was also included in a previous study by Harrison²³ using deprotonated peptides with alkyl and hydrogen side chains. In that work, he reported the ability of AlaAlaAla-OMe to deprotonate readily. The fact that peptide methyl esters generally do not deprotonate as intensely as their acid forms (with the exception of AlaAlaAla-OMe) is reasonable because the peptide acids have a highly acidic carboxylic acid group at the C-terminus and this deprotonation site is lacking from the methyl esters, which must deprotonate at a backbone site.

Extremely low ESI signals were produced for $[M - H]^-$ from AlaGlyAla-OMe, AibAibAib-OMe, and SarSarSar-OMe. This was also observed when using an ESI source on another mass spectrometer, a Bruker (Billerica, MA, USA) HCTUltra high capacity quadrupole ion trap. These three methyl esters did not generate sufficient $[M - H]^-$ for study by ion/molecule reactions and are thus not included in the experimental GA determinations. They are, however, included in our computational study.

Experimental and theoretical GAs and structures of tripeptides Table 1 lists the reference compounds and reaction efficiencies for deprotonation reactions with the tripeptides. Experimental GAs for each of the tripeptides are shown in Table 2 together with the G3(MP2) calculated GAs for each of the possible deprotonation sites. The lowest-energy G3(MP2) calculated structures for the neutral and deprotonated peptides are shown in Figure 2. As a benchmark of the G3MP2 method, we calculated the GAs of some of the relevant reference compounds. The calculated GA of 323.6 kcal/mol for difluoroacetic acid is in excellent agreement with the experimental value and the calculated value of 322.3 kcal/mol is within the error bars of the experimental GA of pentafluorophenol (Table 1).⁴⁶ As a check on this latter value, we calculated the GA of phenol to be 342.1 kcal/mol, in excellent agreement with the experimental value (Table 3).⁴⁶

The GAs of all of the tripeptides were experimentally bracketed between difluoroacetic acid (GA = 323.8 kcal/mol⁴⁶) and pentafluorophenol (GA = 320.7 kcal/mol⁴⁶). This yields experimental GAs within a 1.2 kcal/mol range of 321.4-322.6 kcal/mol. The peptide GA values are in excellent agreement with the G3(MP2) values. As expected, the calculations predicted the most favorable site of deprotonation to be the C-terminal carboxylic acid group. These peptides are about 10 kcal/mol more acidic (lower GA values) than the amino acids glycine (G3(MP2) GA =

335.3 kcal/mol¹³) and alanine (G3(MP2) GA = 334.6 kcal/mol⁴⁰). Thus, as the length of the peptide chain increases, the ability of the peptide to stabilize the charge increases and consequently the acidity at the C-terminus increases. This is likely due to increased hydrogen bonding stabilizing the CO₂⁻ anion site that is not available in the monomer. This is analogous to basicity of the N-terminus increasing as the size of the peptide chain increases for polyglycines.^{48, 65-67} In addition, Ren and coworkers^{14, 42-44} have found that for small peptides containing cysteine residues the peptide becomes more acidic as the chain length increases; they also attributed this to enhanced hydrogen bonding for larger peptide chains.

The experimental GA of GlyGlyGly is 322.5 ± 2.3 kcal/mol. The calculated GA for GlyGlyGly of 323.3 kcal/mol is in excellent agreement and involves deprotonation at the C-terminus. The amide NH closest to the C-terminus is predicted to have a GA of 329.2 kcal/mol, which is only 5.7 kcal/mol less acidic than the carboxylic acid group of the C-terminus. The lowest energy calculated structures for neutral and C-terminally deprotonated GlyGlyGly are shown in Figure 2(a). Neutral GlyGlyGly forms a wrapped structure with four hydrogen bonds. The N-terminus forms a hydrogen bond to the C-terminus (2.31 Å), the C-terminal amide forms a hydrogen bond with the N-terminal carbonyl oxygen (2.00 Å), the central amide forms a hydrogen bond with the N-terminus (2.13 Å), and the C-terminus forms a hydrogen bond between the two oxygens (2.30 Å). Deprotonated GlyGlyGly is slightly more compact than the neutral with three strong hydrogen bonds. This is due to the fact that the N-terminal carboxyl oxygen rotates away from the C-terminal amide essentially breaking that hydrogen bond and allowing the hydrogen bond between the N-terminus and the C-terminus to be shortened to 1.91 Å. The

hydrogen bond between the central amide and the N-terminus is retained (2.06 Å) and a new, strong hydrogen bond forms between the C-terminal amide and the C-terminus (1.93 Å).

All of the calculated GA values are in excellent agreement with experiment to within 1.4 kcal/mol as seen in Table 2. AlaGlyAla is the most acidic peptide with a calculated GA of 321.0 kcal/mol while AlaAlaAla and GlyAlaGly have a slightly more basic GA of 321.6 kcal/mol. (Lower GA values are more acidic.) SarSarSar is the least acidic peptide with a calculated GA of 323.7 kcal/mol; GlyGlyGly is only 0.2 kcal/mol more acidic. To complete this series, G3(MP2) calculations were performed on the tripeptide of N-methylalanine (NmaNmaNma). The GA of NmaNmaNma is essentially the same as that of C-methylalanine (AibAibAib) showing that methyl substitution on the backbone C or N affects the acidity in the same way. The lowest energy N-methyl neutrals SarSarSar and NmaNmaNma were compared to their corresponding lowest energy isomer C-methyl neutrals AlaAlaAla and AibAibAib, respectively. In SarSarSar versus AlaAlaAla and NmaNmaNma versus AibAibAib, both comparisons showed that the C-methyl isomers were lower in free energy than their respective N-methyl isomers by 30.4 and 43.5 kcal/mol, respectively.

The GlyAlaGly, AlaGlyAla, AlaAlaAla, and AibAibAib lowest energy neutrals and anions basically have the same structure as neutral and anionic GlyGlyGly, respectively, including the same hydrogen bonds. This indicates that the hydrogen and methyl substituents have little effect on the structure itself. The AibAibAib anion is an exception because the additional methyl groups cause the lowest energy structure to be more linear due to steric interactions. Strong hydrogen bonds exist between the central amide and C-terminal carbonyl oxygen (1.91 Å) and the C-terminal amide and C-terminus (1.80 Å). The bond lengths and structures for these peptides

are given in Figure 2(b-e). Because of the methylated nitrogens in SarSarSar, the amide groups are no longer able to hydrogen bond to the carbonyl oxygens. For SarSarSar, the lowest energy neutral only contains one hydrogen bond, the C-terminus forms a bond between the two oxygens (2.30 Å). In anionic SarSarSar, a hydrogen bond forms between the N-terminus and the C-terminus (1.94 Å). For the lowest energy NmaNmaNma neutral, a very strong hydrogen bond exists between the C-terminus and the C-terminal carbonyl oxygen (1.72 Å). Anionic NmaNmaNma contains the same hydrogen bond as the SarSarSar anion, except the bond is longer by 0.24 Å. The structures for SarSarSar and NmaNmaNma are shown in Figure 2(f-g).

Two low energy neutral structures were predicted at the DFT and G3(MP2) levels for the neutral tripeptides. At the composite G3(MP2) level (energies based on MP2/6-31G(d) geometries), the neutral structure with 3 hydrogen bonds is higher in energy than the neutral structure with 4 hydrogen bonds for all of the tripeptides in terms of the free energy and enthalpy by 0.6 to 4.5 kcal/mol. As the difference in energy at the G3(MP2) level for all of the tripeptides, except for GlyGlyGly, is less than approximately 1.5 kcal/mol, both structures could be present in the gas phase. For GlyGlyGly, the energy differences are higher ($\Delta H = 2.7$ and $\Delta G = 4.5$ kcal/mol), so only the structure with the four hydrogen bonds should be present in the gas phase.

The DFT results can be compared to the composite G3(MP2) results to benchmark DFT's ability to predict the lowest energy neutral structures for these neutral tripeptides. The comparison showed that DFT predicted the wrong structure to be of lowest energy when compared to the G3(MP2) results. The lowest energy neutral predicted by DFT had only 3 hydrogen bonds, whereas the one predicted by G3(MP2) had 4 hydrogen bonds. At the DFT level, the neutrals with 4 hydrogen bonds were predicted to be higher in energy than the neutrals with 3 hydrogen

bonds in AlaGlyAla, AlaAlaAla, and AibAibAib in terms of both the free energy and the enthalpy by 1.0 to 2.0 kcal/mol. For GlyAlaGly and SarSarSar, the structures with 4 hydrogen bonds are higher in energy than the ones with 3 hydrogen bonds by ~ 0.5 kcal/mol for the enthalpy, but the free energy reverses the ordering with the structure with 3 hydrogen bonds being higher in energy by approximately 1 kcal/mol. For GlyGlyGly, the two structures are isoenergetic at the DFT level. The hydrogen bonds between the central amide and N-terminus and the C-terminal amide and the N-terminal carbonyl oxygen were retained in both the G3(MP2) and DFT structures with the difference between the two structures arising from the orientation of the hydrogen in the C-terminus. In the G3(MP2) structure, the N-terminus formed a bond to the C-terminus and the C-terminus formed a bond between the two oxygens. In the DFT structure, the hydrogen in the C-terminus was rotated away from the double bonded oxygen and formed a hydrogen bond with the C-terminal carbonyl oxygen. The structures of the higher energy neutrals at the G3(MP2) level are given in the Figure 3.

Deprotonation from the N-terminal carbon or the N-terminus (sites 6 and 7) led to the formations of 5 and 6 membered rings, respectively, for all tripeptides. Deprotonation at the N-terminal carbon forms an electron rich carbon so that rotations about various bonds occur to close to a 5-member ring by bonding to the C-terminal carbonyl carbon. Deprotonation at the N-terminus leads to bond rotations so that a 6-member ring can be formed by bonding to the C-terminal carbonyl carbon. A larger free energy can also be seen in these ring molecules due to the decrease in entropy. Anions that are created from deprotonating the N-terminus that do not form a ring, as found for GlyGlyGly, have a higher GA value (~ 400 kcal/mol). The GAs of methane and ammonia, obtained by combining the experimental heats of formation with the entropy

corrections calculated at the DFT level, are 409.6 and 397.0 kcal/mol, respectively.^{46, 68} The acidity calculated from deprotonating at the end terminus (i.e., the acidity of the amine group) and is comparable to that of ammonia.⁴⁶

Experimental and theoretical GAs and structures of tripeptide methyl esters The GAs of the peptide methyl esters GlyGlyGly-OMe, AlaAlaAla-OMe, and GlyAlaGly-OMe have been determined for the first time. Converting the C-terminal carboxylic acid group to a methyl ester removes the most acidic site on the tripeptide, thus forcing deprotonation to occur at another site. Table 3 lists the reference compounds and reaction efficiencies for the deprotonation reactions with the methyl ester peptides. Table 4 lists the experimental GAs of each of the peptide methyl esters and the G3(MP2) calculated GAs for each of the deprotonation sites that were explored. The calculations show that the most favorable backbone deprotonation sites are the amide nitrogens. Figure 4 gives the lowest energy calculated structures for these neutral and deprotonated peptide methyl esters.

The GA of GlyGlyGly-OMe was experimentally determined to be 338.1 ± 2.3 kcal/mol. G3(MP2) calculations show deprotonation of the amide nitrogens and the N-terminus to be very close in energy. The central amide nitrogen has a G3(MP2) calculated GA of 340.1 kcal/mol and the C-terminal amide nitrogen has a GA of 342.1 kcal/mol. Thus, ESI on GlyGlyGly-OMe is mostly likely causing deprotonation at the central backbone nitrogen site. The experimental and calculated GAs are within experimental error.

Figure 4(a) reveals that neutral GlyGlyGly-OMe adopts a conformation similar to that of GlyGlyGly except that only 3 hydrogen bonds are present because of methylation of the C-terminus. The N-terminus forms a hydrogen bond to the C-terminus (2.28 Å), the C-terminal

amide forms a hydrogen bond with the N-terminal carbonyl oxygen (2.01 Å), and the central amide forms a hydrogen bond with the N-terminus (2.13 Å). When deprotonation occurs at the central NH of GlyGlyGly-OMe, the resulting structure is more open than the neutral molecule due to the fact that the hydrogen bond between the N-terminus and C-terminus is broken. Hydrogen bonding is still observed between the C-terminal NH and the N-terminal carbonyl oxygen (1.72 Å), and the N-terminus has a hydrogen bond to the central nitrogen, which is the site of deprotonation (2.31 Å). Delocalization of the negative charge occurs over the central nitrogen and the adjacent carbonyl to form an amidate structure given by the resonance structures $C(=O)-N^- \leftrightarrow C(-O^-)=N$.

The calculated GAs for GlyAlaGly-OMe and AlaAlaAla-OMe agree with experiment to within 2.0 kcal/mol for both the central and C-terminal amides (Table 4). Deprotonation from the central amide leads to a difference of only 0.1 kcal/mol between GlyAlaGly-OMe and AlaAlaAla-OMe and results in the most acidic anion for the GlyAlaGly methyl ester peptide. Deprotonation at the C-terminal amide in AlaAlaAla-OMe leads to a more acidic anion by 1.5 kcal/mol. The lowest energy structure for neutral GlyAlaGly-OMe has the same 3 hydrogen bonds as found for GlyGlyGly-OMe with only small changes in these bond lengths. The lowest energy structure for neutral AlaAlaAla-OMe is similar, but has only two hydrogen bonds because of the loss of the hydrogen bond between the N-terminus and the C-terminus. The lowest energy GlyAlaGly-OMe anion results in the same structure as found for the GlyGlyGly-OMe anion. Anionic AlaAlaAla-OMe is still a very closed structure with a different set of hydrogen bonds. The central amide now hydrogen bonds to the C-terminus and the N-terminus is hydrogen bonded to the N-terminal carbonyl oxygen. 5- and 6-member rings are formed in the tripeptide me-

thyl esters when deprotonation occurs at the N-terminal carbonyl carbon or the N-terminus, respectively, just as found for the non-esterified tripeptide acids.

Peptide methyl esters not deprotonated by ESI Three of the methyl esters studied, AlaGlyAla-OMe, AibAibAib-OMe, and SarSarSar-OMe, produced only extremely weak $[M - H]^-$ by ESI and, consequently, their GAs could not be determined experimentally using ion/molecule reactions. These species were studied computationally and Figure 5 shows the lowest energy structures of their neutral and deprotonated forms.

Abundant deprotonation by ESI requires that the analyte be more acidic than the solvent. This is obviously the case here. The reported GAs for the components of the solvent system used in this research are: GA of methanol = 376.02 ± 0.62 kcal/mol and GA of water = 383.74 ± 0.06 kcal/mol.⁴⁶ These GA values are much higher (less acidic) than the calculated GAs of AlaGlyAla-OMe, AibAibAib-OMe, and SarSarSar-OMe, which are included in Table 4. These methyl esters are much more acidic than the solvent system, thus solvent acidity cannot be the cause for their inability to form $[M - H]^-$.

The calculations show that the GAs of AlaGlyAla-OMe and AibAibAib-OMe are not the cause of the tripeptide methyl esters' inability to deprotonate by ESI. For AlaGlyAla-OMe, the calculated GAs are 340.7 and 341.0 kcal/mol for the central and C-terminal amides respectively. These GAs are both within 1 kcal/mol of the GA for GlyGlyGly-OMe and very similar to the GAs of analogous sites in AlaAlaAla-OMe and GlyAlaGly-OMe; these latter three tripeptide methyl esters deprotonate readily by ESI. The calculated GAs of the central amides in AibAibAib-OMe show that only the central amide would deprotonate with a GA of 339.9 kcal/mol with the GA of the C-terminal amide ~15 kcal/mol higher in energy. The GA for SarSarSar-OMe is 350.1

kcal/mol, which is ~10 kcal/mol higher in energy than GlyGlyGly-OMe. Thus, with the potential exception of SarSarSar-OMe, the values of the GAs are not the reason that these anions are not observed.

Steric and conformational interactions within the neutral peptide are another viable reason for poor $[M - H]^-$ formation. Conformation is known to affect the ability of ESI to protonate and deprotonate peptides (i.e., affect the charge state distribution produced by ESI).⁶⁹⁻⁷³ More compact conformations, where potential deprotonation sites are shielded and inaccessible to solvent molecules, may not readily deprotonate.

To study steric accessibility for proton removal, van der Waals surfaces and solvent accessible surfaces were generated for all of the tripeptides and their methyl esters. The surfaces for AlaGlyAla-OMe, AibAibAib-OMe, and SarSarSar-OMe are shown in Figure 6; the surfaces for the remaining peptides are shown in Figure 7. The van der Waals surface for AlaGlyAla-OMe shows that the C-terminal amide is covered by the N-terminal carbonyl oxygen because of a hydrogen bond (2.05 Å). The solvent accessible surface for AlaGlyAla-OMe shows that the orientation of the methyl groups in the terminal alanine residues and the wrapped nature of the peptide block the central amide from deprotonation. Comparison of AlaGlyAla-OMe with AlaAlaAla-OMe, which did deprotonate, shows that the orientation of the N-terminus with respect to the C-terminus can prevent deprotonation. In AlaAlaAla-OMe, the N-terminus is rotated away from the C-terminus; however, in AlaGlyAla-OMe, the end terminus has a very strong hydrogen bond to the double bonded oxygen in the C-terminus that could be preventing deprotonation. The van der Waals surface for AibAibAib-OMe shows that the two methyl groups on each carbon in the backbone completely cover the C-terminal and central amide sites preventing

deprotonation from occurring. The solvent accessible surface further justifies this conclusion. For SarSarSar-OMe, methylation of the central amide nitrogens and methylation of one hydrogen on the N-terminus essentially removes all sites of deprotonation. The van der Waals surface for SarSarSar-OMe shows that the only hydrogen available for deprotonation is blocked by the methyl group on the C-terminal carbon, consistent with its higher GA as noted above.

6.4 Conclusions The GAs of six tripeptides (GlyGlyGly, GlyAlaGly, AlaGlyAla, AlaAlaAla, AibAibAib, and SarSarSar) with neutral side chains and their methyl esters have been determined for the first time. Experimental and calculated GA values are in excellent agreement. The standard C-terminal acid (-COOH) peptides have very similar GAs, all falling within a 1.2 kcal/mol range. These tripeptides are about 13 kcal/mol more acidic than the amino acids glycine and alanine. Deprotonation occurs at the C-terminus, as confirmed by the computational results. The three peptide methyl esters that could be studied experimentally (GlyGlyGly-OMe, GlyAlaGly-OMe, and AlaAlaAla-OMe) also have very close GA values, all falling within a 2.0 kcal/mol range. Due to the lack of a traditional acidic site (i.e. -COOH), deprotonation of the methyl esters must be occurring at a backbone site on the peptide. High level G3MP2 computations predict that both C-terminal and central amide nitrogens (NHs) are the energetically favorable site for backbone deprotonation except for GlyGlyGly-OMe where the central NH is the most favorable. Thus, when mechanisms are proposed for the dissociation of deprotonated peptides during tandem mass spectrometry (MS/MS) processes, amide nitrogens should be considered as the most likely sites of backbone deprotonation. AlaGlyAla-OMe, AibAibAib-OMe, and SarSarSar-OMe peptides were also studied; however, ESI did not show deprotonation. The inability of ESI to deprotonate SarSarSar-OMe, which lacks amide hydrogens that can be abstracted, provides

additional confirmation that amide nitrogens are the site of peptide backbone deprotonation by ESI. The calculated structures indicate that steric effects prevent deprotonation by ESI for AlaGlyAla and AibAibAib-OMe.

Acknowledgement This work is supported by the National Science Foundation (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. Samantha S. Bokatzian-Johnson thanks the University of Alabama's National Alumni Association for a License Tag Fellowship. Michele L. Stover thanks the U.S. Department of Education for a Graduate Assistance in Areas of National Need Fellowship (DoEd-GAANN grant P200A100190). David A. Dixon thanks the Robert Ramsay Fund from the University of Alabama for partial support.

Table 6.1. Reaction Efficiencies for the Proton Transfer reactions of Deprotonated Tripeptides with Reference Compounds.

Ref Compound	GA ^a (kcal/mol)	Average Reaction Efficiency (\pm standard deviation)					
		GlyGlyGly	GlyAlaGly	AlaGlyAla	AlaAlaAla	AibAibAib	SarSarSar
Trifluoropropionic acid	327.0 \pm 1.1	0.05 \pm 0.01	0.05 \pm 0.00	0.06 \pm 0.05	0.04 \pm 0.00	0.04 \pm 0.01	0.08 \pm 0.01
Difluoroacetic acid	323.8 \pm 2.0	0.06 \pm 0.02	0.06 \pm 0.01	0.04 \pm 0.01	0.04 \pm 0.01	0.04 \pm 0.01	0.07 \pm 0.00
		BREAK^b	BREAK	BREAK	BREAK	BREAK	BREAK
Pentafluorophenol	320.7 \pm 2.0	0.45 \pm 0.05	0.48 \pm 0.02	0.43 \pm 0.02	0.48 \pm 0.03	0.34 \pm 0.02	0.57 \pm 0.02
Trifluoroacetic acid	317.4 \pm 2.0	0.52 \pm 0.06	0.53 \pm 0.09	0.72 \pm 0.06	0.51 \pm 0.08	0.55 \pm 0.11	0.72 \pm 0.06

^a All reference compound GAs were obtained from reference 46.

^b "BREAK" indicates the point where the experimental GA was assigned.

Table 6.2. Experimental and G3(MP2) Theoretical GAs in kcal/mol for Tripeptides.

Peptide	Experimental	G3MP2						
		N-terminus (NH ₂)	N-terminal C _α H	Central NH	Central C _α H	C-terminal NH	C-terminal C _α H	C-terminus (COOH)
GlyGlyGly-OH	322.1 ± 2.5	364.2	350.6	334.7	344.5	329.2	344.3	323.3
GlyAlaGly-OH	322.3 ± 2.5	358.5	354.4	337.6	356.3	335.7	347.0	321.6
AlaGlyAla-OH	322.0 ± 2.5	360.3	358.4	340.3	355.3	338.1	348.9	321.0
AlaAlaAla-OH	322.2 ± 2.5	357.0	355.6	338.1	354.3	340.3	345.4	321.6
AibAibAib-OH	321.4 ± 2.5	b	a	330.4	a	328.9	a	322.1
SarSarSar-OH	322.6 ± 2.5	353.2	347.0	a	358.2	a	348.7	323.7
NmaNmaNma-OH	N/A ^c	322.0	322.9	a	329.5	a	337.8	322.5

^a This site does not exist in the peptide.

^b Attempts to deprotonate at this site in the calculations resulted in a reversion back to deprotonation at the central NH.

^c N/A = not available. Experimental GA was not studied.

Table 6.3. Reaction Efficiencies for the Proton Transfer Reactions of Deprotonated Tripeptide Methyl Esters with Reference Compounds.

Reference Compound	GA ^a (kcal/mol)	Average Reaction Efficiency (\pm standard deviation)		
		GlyGlyGly-OMe	GlyAlaGly-OMe	AlaAlaAla-OMe
phenol	342.3 \pm 2.0	0.05 \pm 0.01	0.02 \pm 0.004	0.04 \pm 0.02
acetic acid	341.1 \pm 2.0	0.025 \pm 0.002	0.02 \pm 0.005	0.02 \pm 0.001
formic acid	339.1 \pm 1.5	0.11 \pm 0.06	0.10 \pm 0.04	0.09 \pm 0.02
isovaleric acid	338.5 \pm 2.0	0.23 \pm 0.04	0.09 \pm 0.04	0.05 \pm 0.03
		BREAK^b		
trimethylacetic acid	337.6 \pm 2.0	0.31 \pm 0.18	0.22 \pm 0.15	0.12 \pm 0.02
			BREAK	BREAK
p-chlorophenol	336.2 \pm 2.0	0.32 \pm 0.01	0.41 \pm 0.03	0.34 \pm 0.01
3-trifluoromethyl phenol	332.4 \pm 2.0	– ^c	1.02 \pm 0.09	1.09 \pm 0.08

^a All reference compound GAs were obtained from reference 46.

^b "BREAK" indicates the point where the experimental GA was assigned.

^c "–" indicates no experiment was performed.

Table 6.4. Experimental and G3(MP2) Theoretical GAs in kcal/mol for Peptide Methyl Esters.

Peptide	Experimental	G3MP2					
		N-terminus (NH ₂)	N-terminal C _α H	Central NH	Central C _α H	C-terminal NH	C-terminal C _α H
GlyGlyGly-OMe	338.1 ± 2.3	363.0	358.1	340.1	358.7	342.1	348.7
GlyAlaGly-OMe	337.2 ± 2.4	358.1	354.4	338.7	357.0	339.2	350.0
AlaGlyAla-OMe	N/A ^b	361.8	358.2	340.7	362.3	341.0	352.2
AlaAlaAla-OMe	336.7 ± 2.4	358.4	355.0	338.6	355.4	337.1	347.9
AibAibAib-OMe	N/A ^b	363.2	^a	339.9	^a	354.9	^a
SarSarSar-OMe	N/A ^b	350.1	354.1	^a	354.7	^a	351.6

^a This site does not exist in the peptide.

^b Experimental GA could not be determined by ion/molecule reactions because peptide could not be deprotonated by ESI.

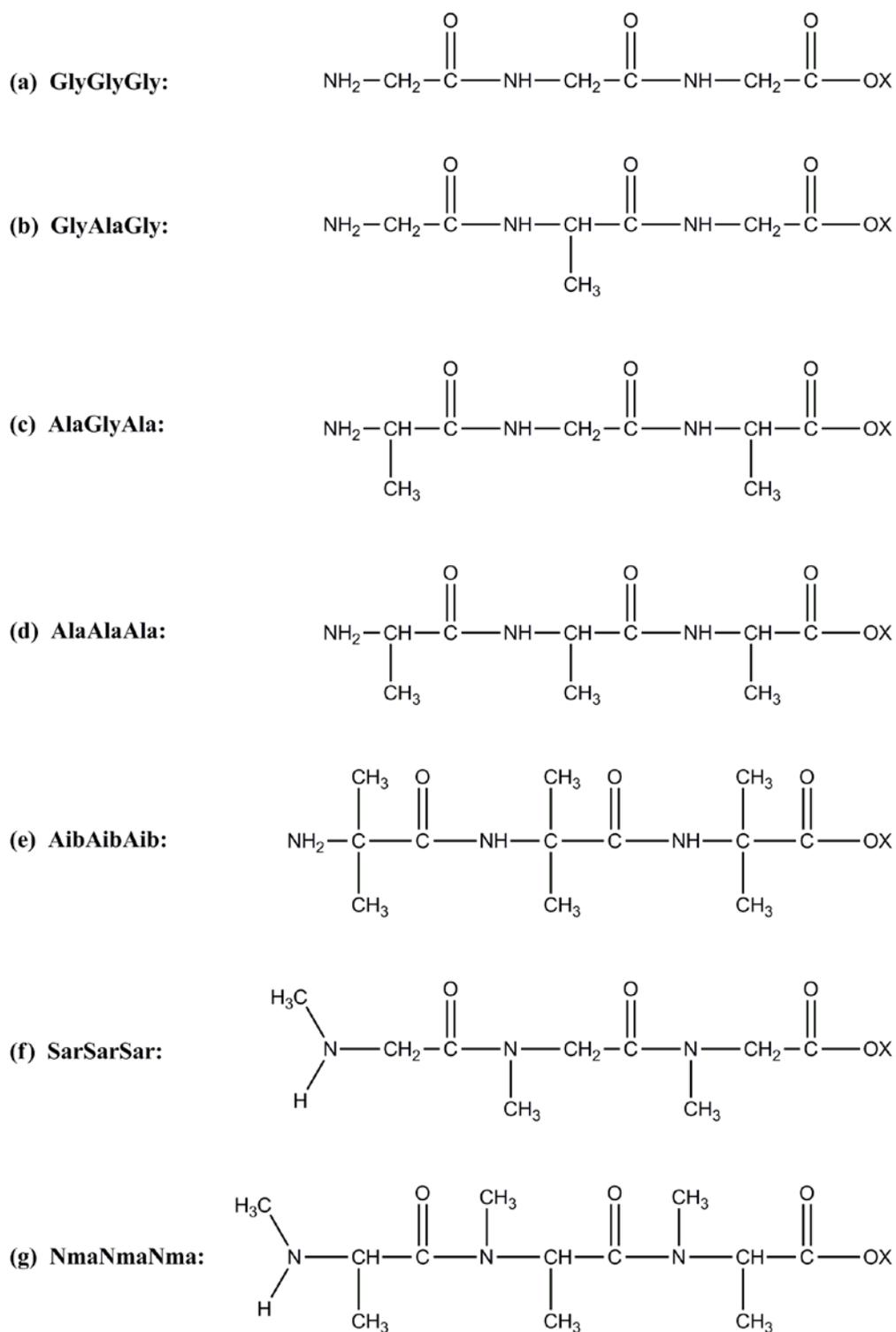
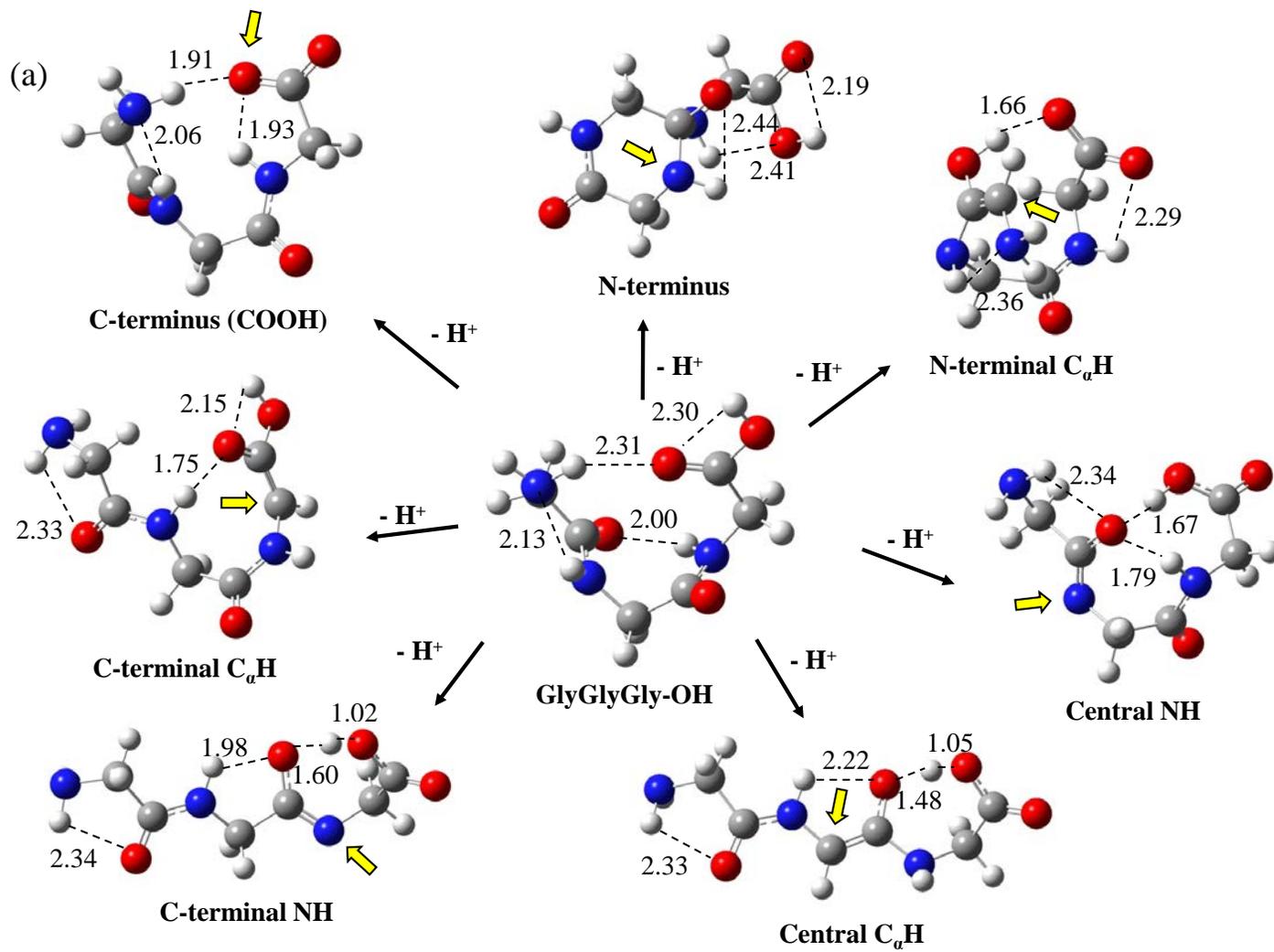
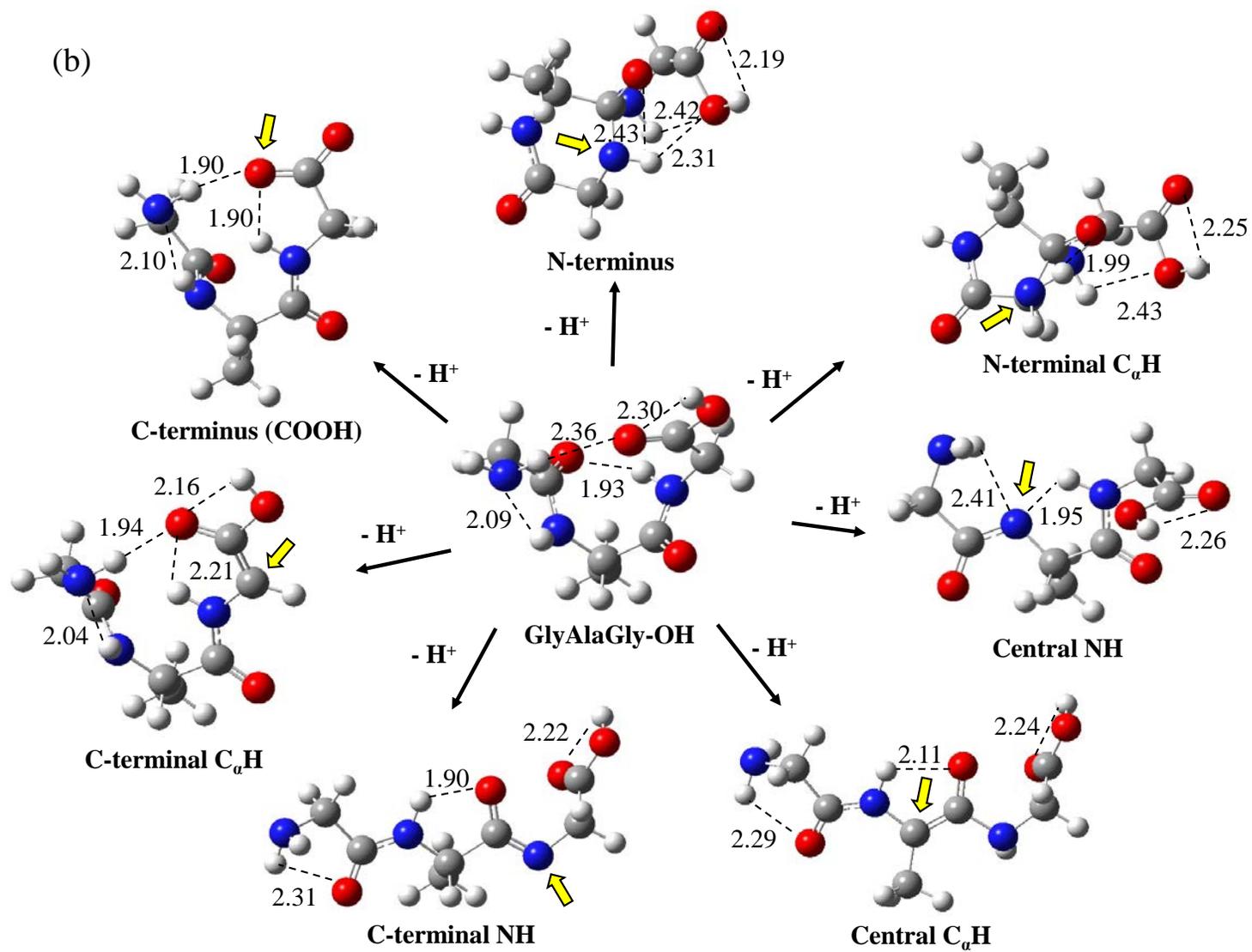
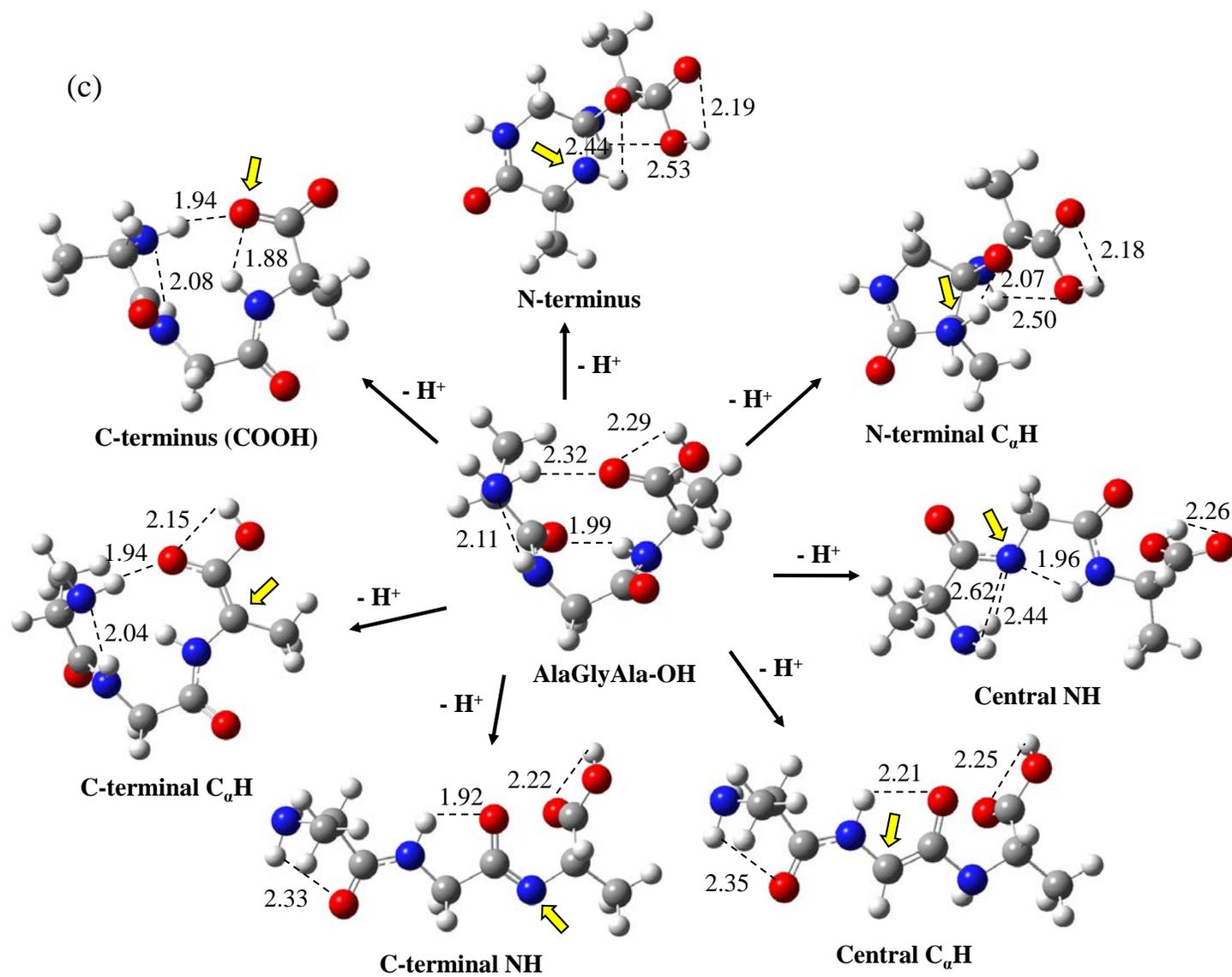
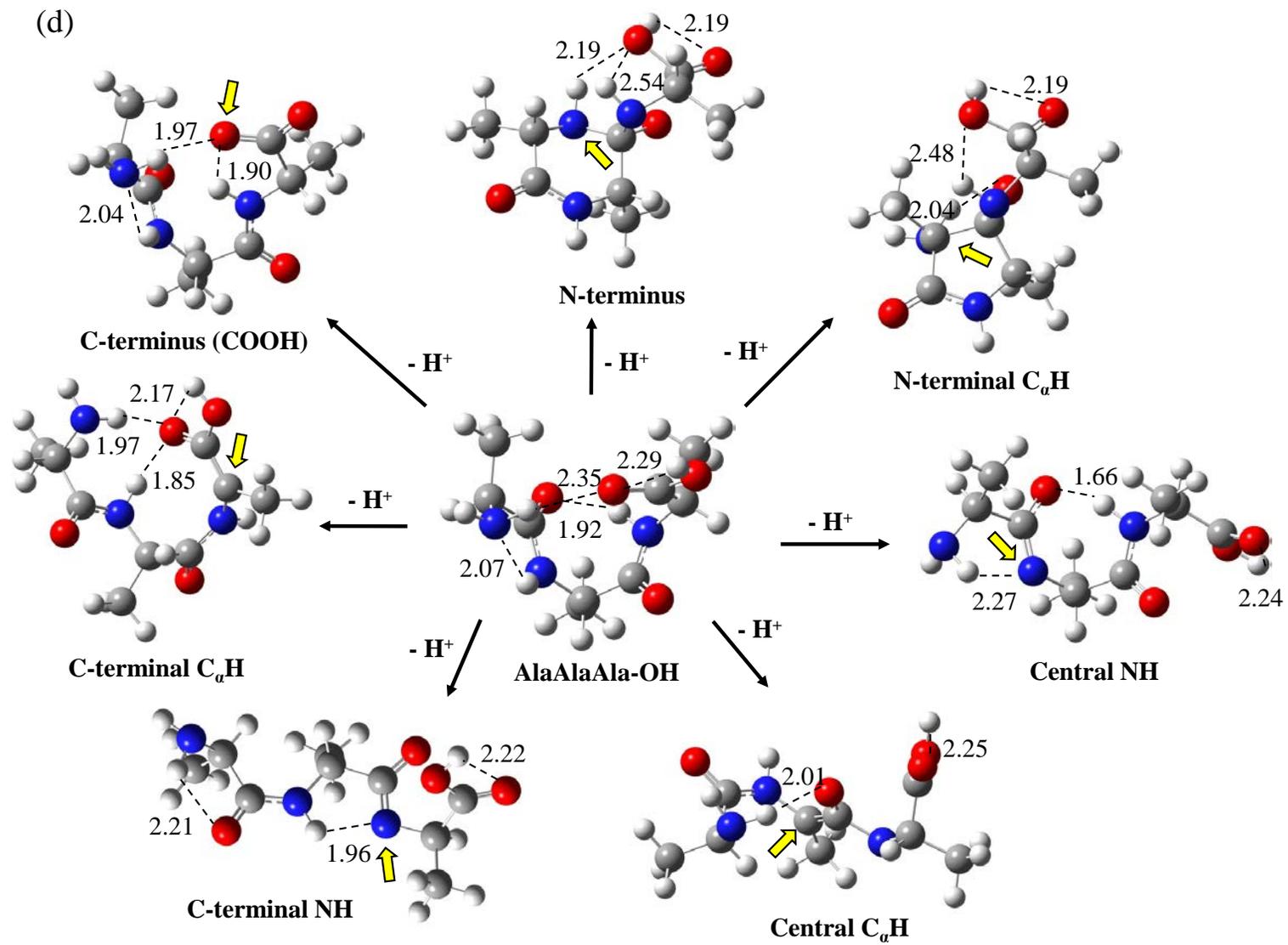


Figure 6.1. Structures of tripeptides (X=H) and their methyl esters (X=CH₃)

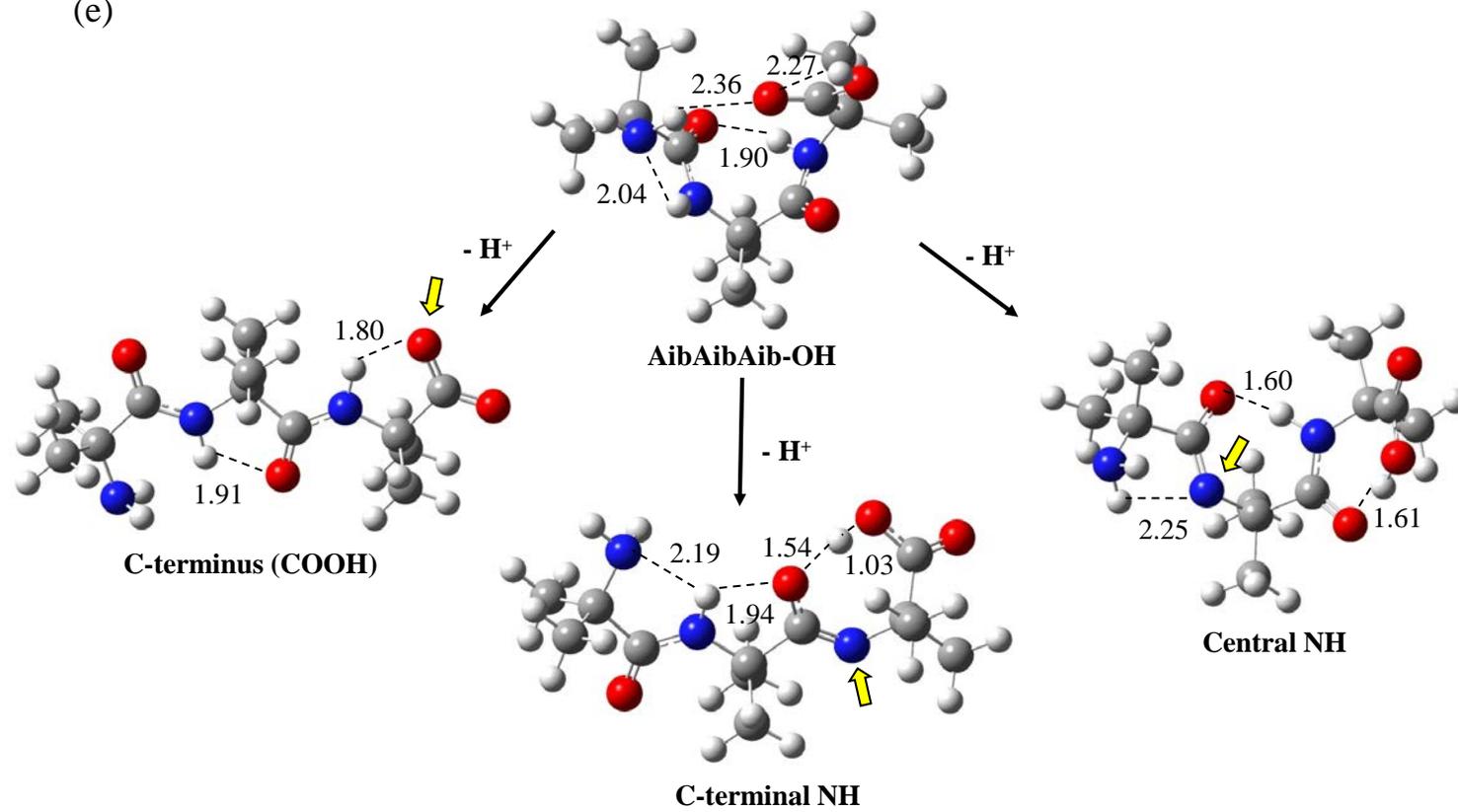


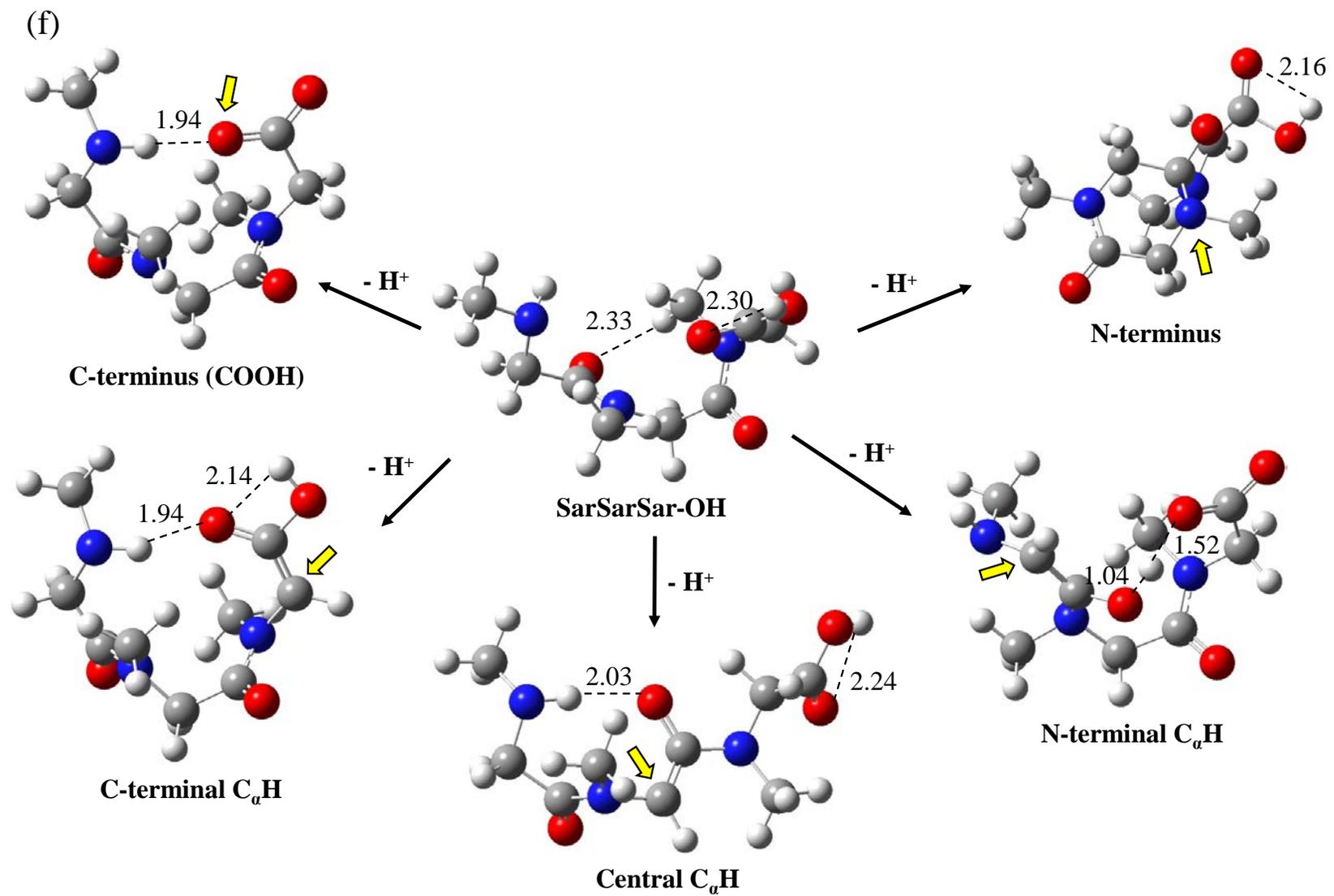






(e)





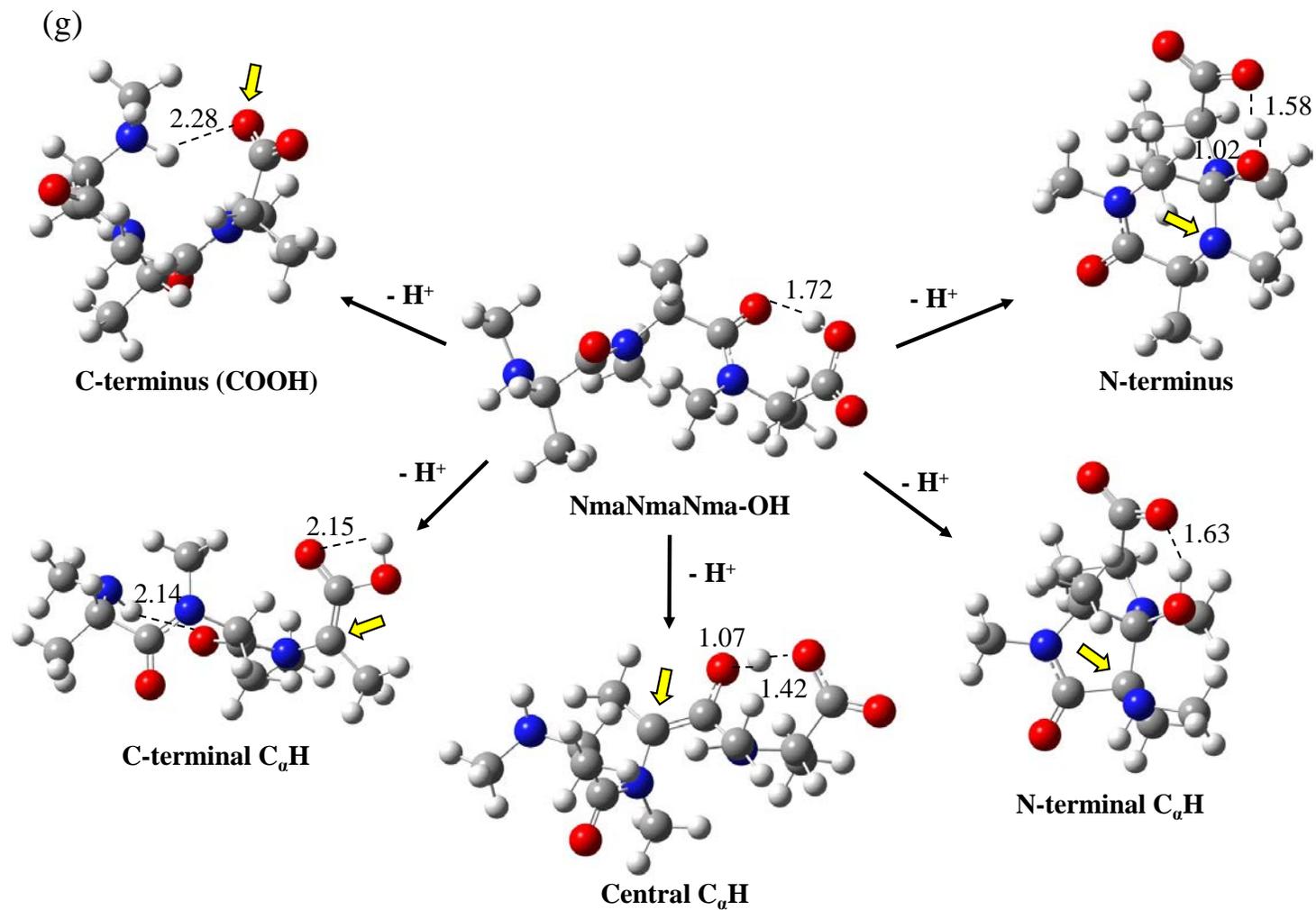
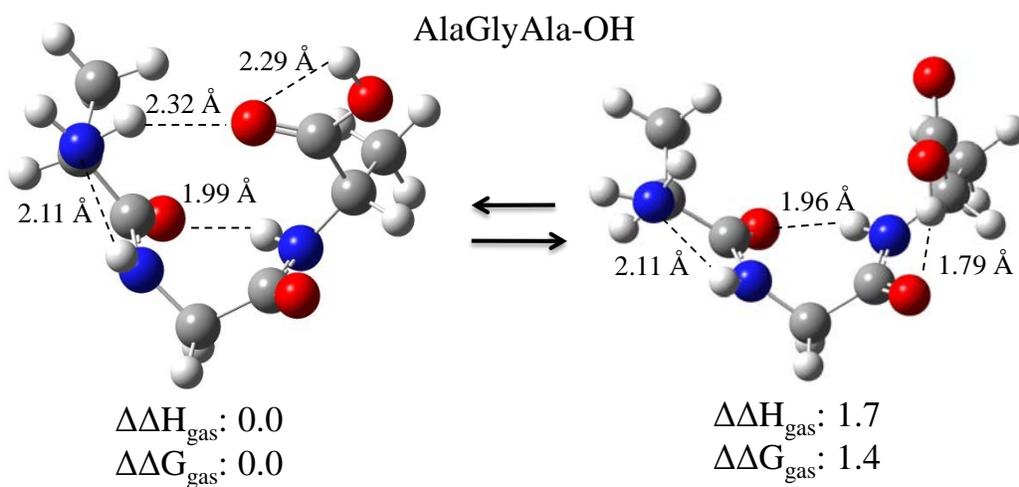
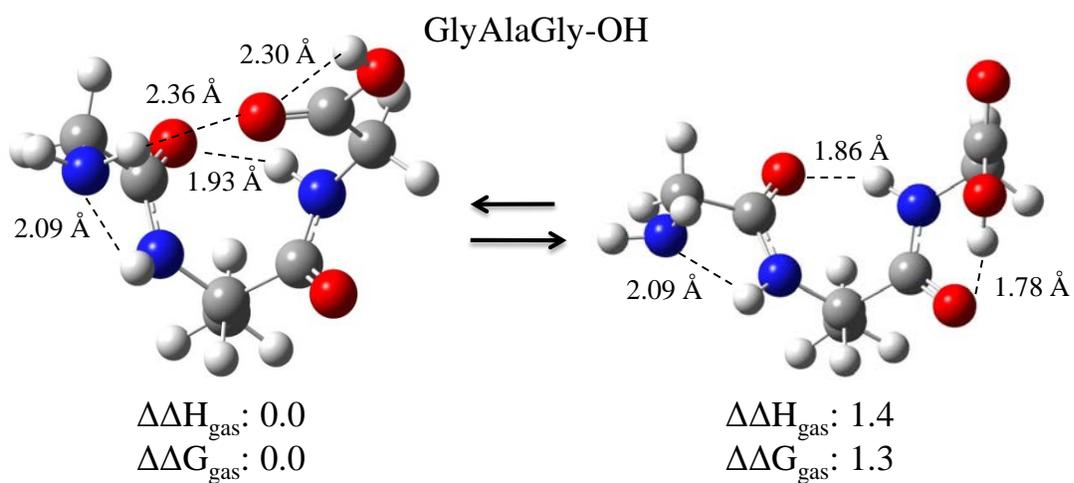
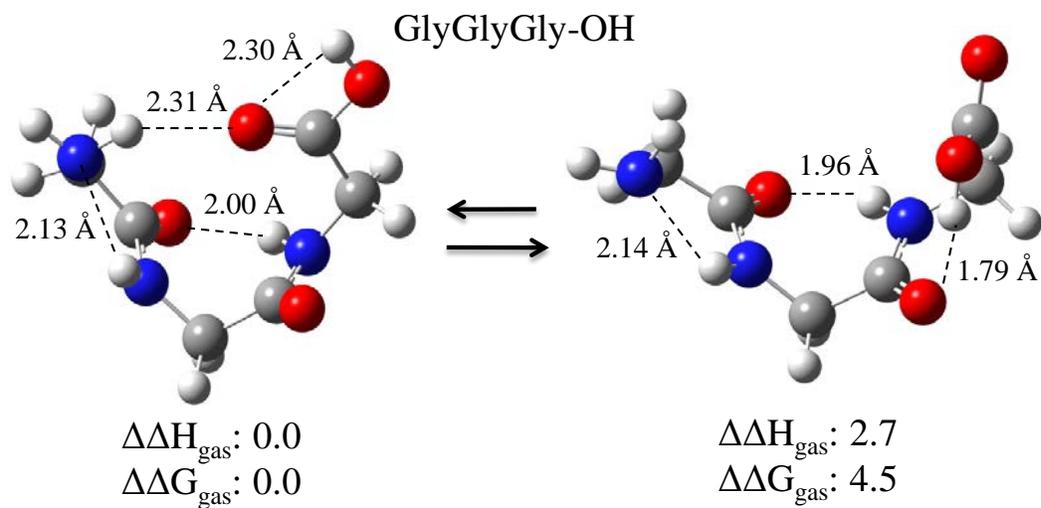


Figure 6.2. G3(MP2) calculated structures for neutral and deprotonated tripeptide acids (a) GlyGlyGly (b) GlyAlaGly (c) AlaGlyAla (d) AlaAlaAla (e) AibAibAib (f) SarSarSar (g) NmaNmaNma. Arrows indicate the deprotonation site. Bond lengths are in Å.



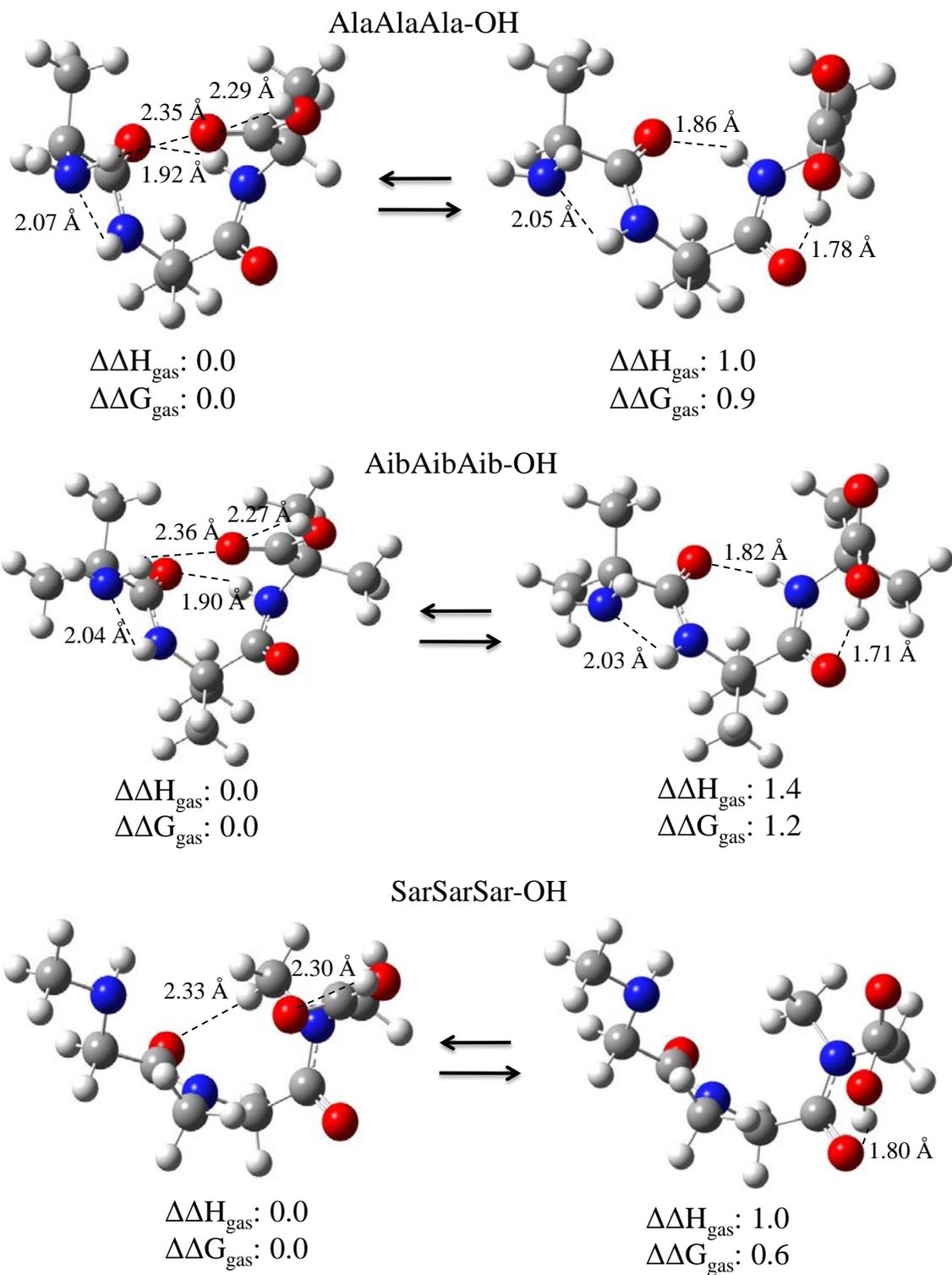
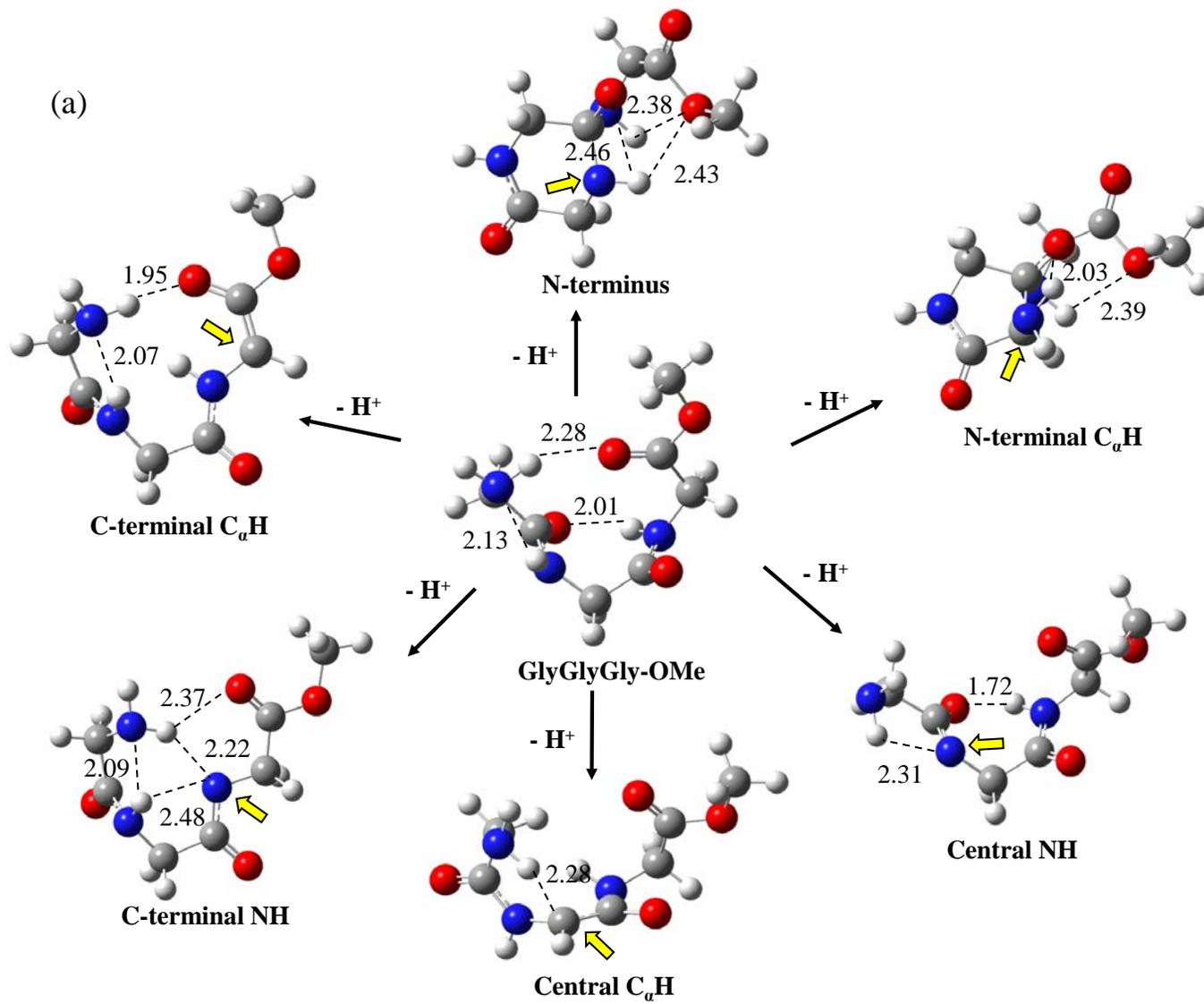
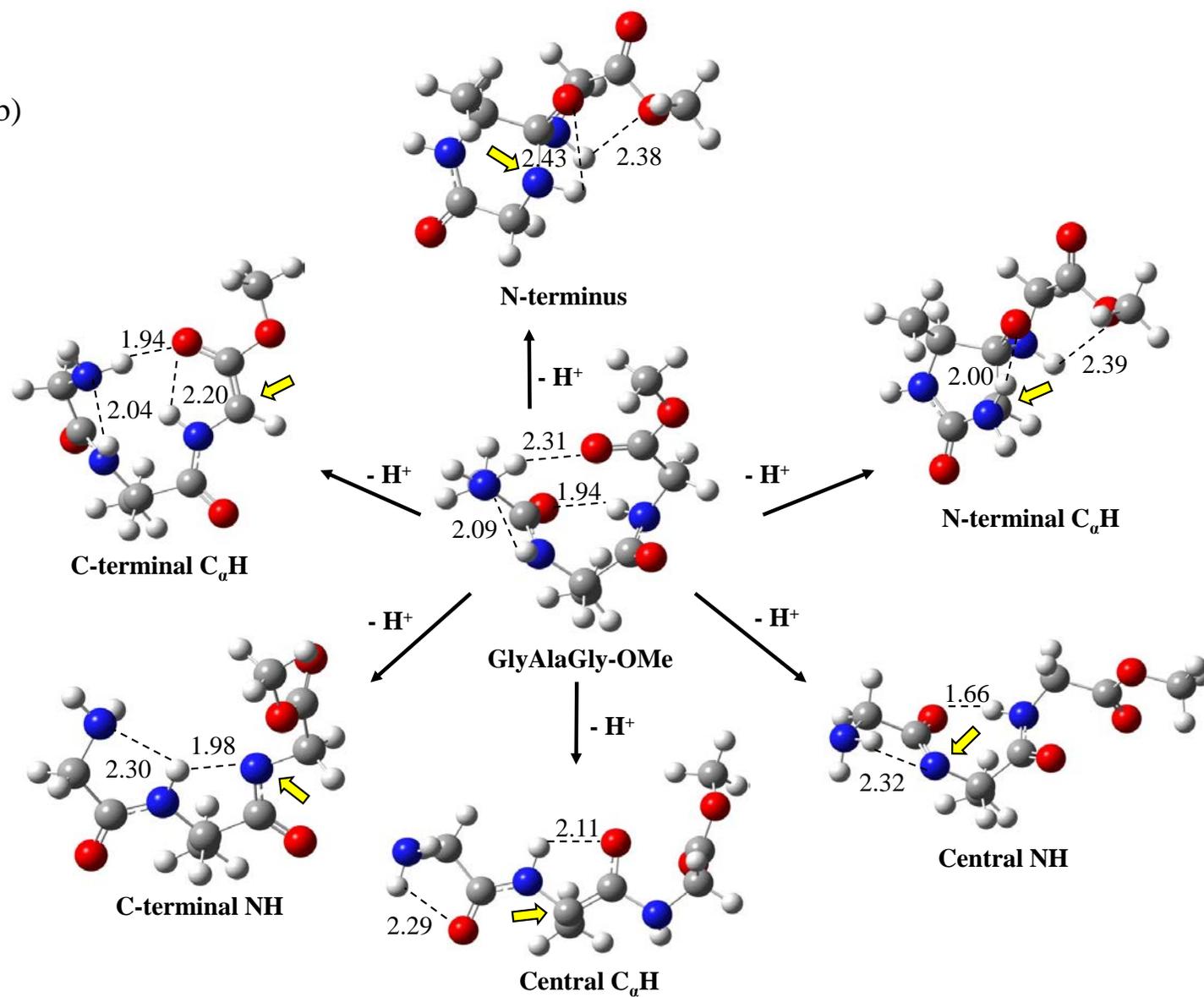


Figure 6.3. G3(MP2) optimized structures of the higher energy neutrals. Energies in kcal/mol.



(b)



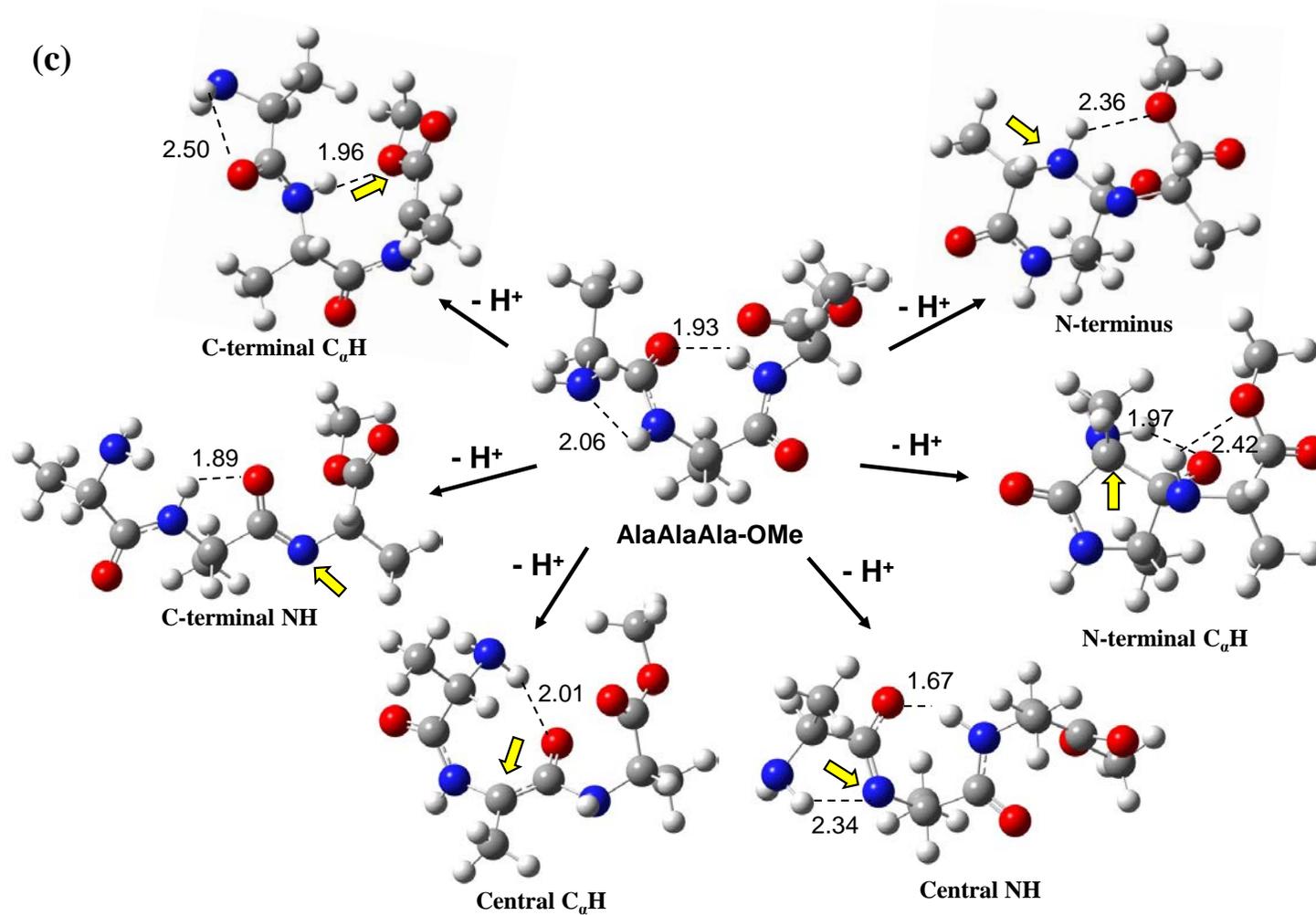
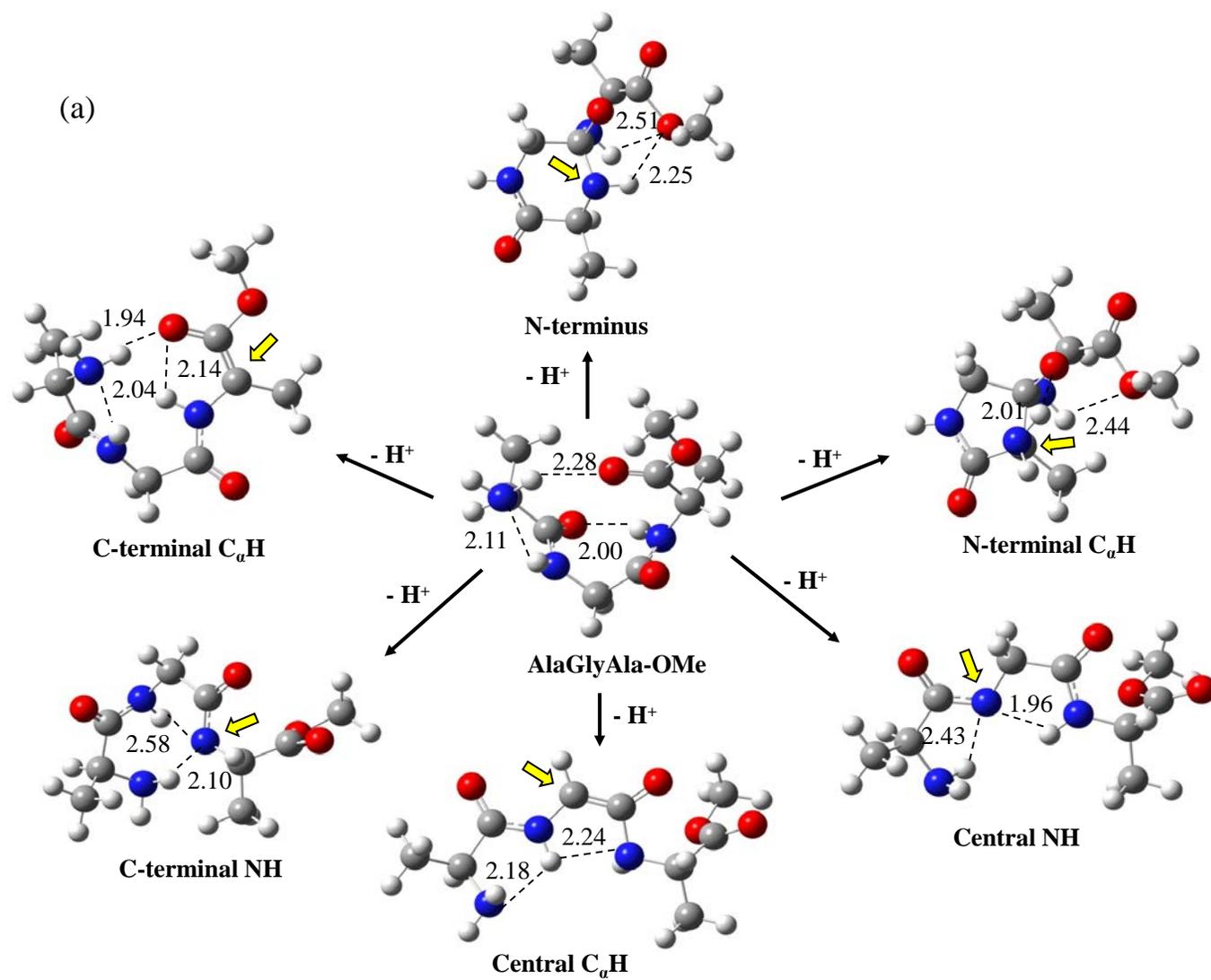
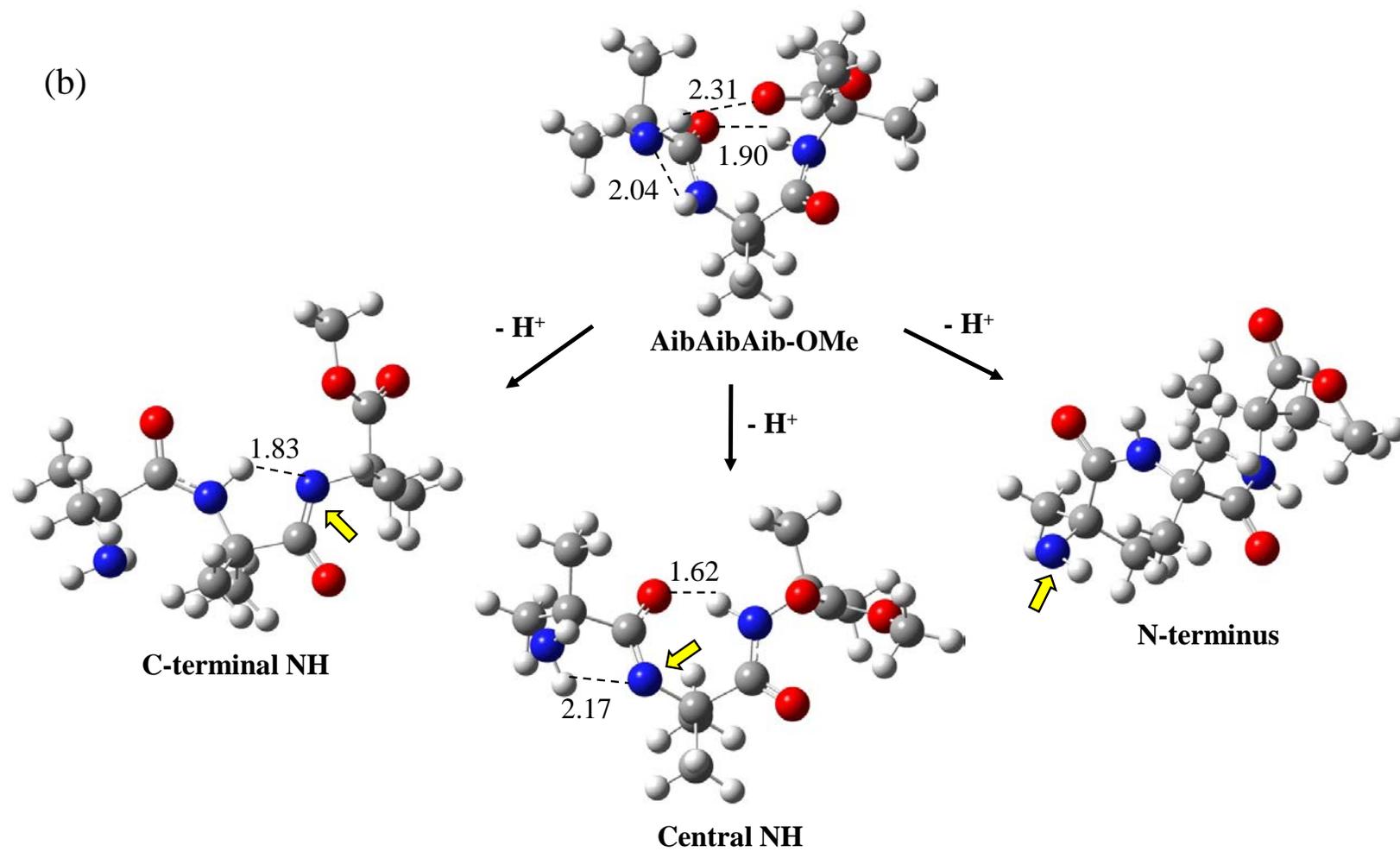


Figure 6.4. G3(MP2) calculated structures for neutral and deprotonated tripeptide methyl esters (a) GlyGlyGly-OMe (b) GlyAlaGly-OMe (c) AlaAlaAla-OMe. Arrows indicate the deprotonation site. Bond lengths are in angstroms. These peptide methyl esters deprotonate readily by ESI.



(b)



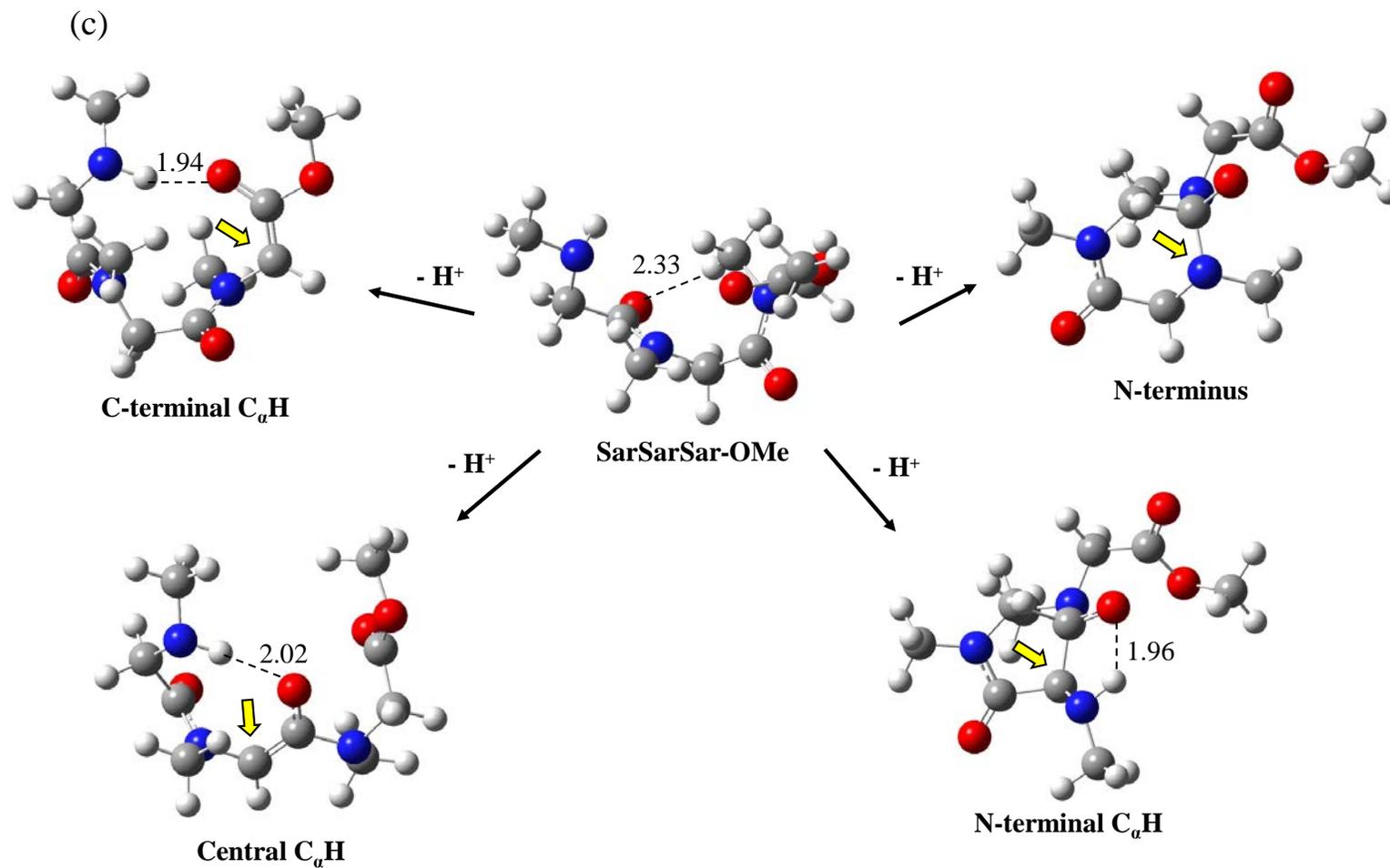


Figure 6.5. G3(MP2) calculated structures for neutral and deprotonated tripeptide methyl esters (a) AlaGlyAla-OMe (b) AibAibAib-OMe (c) SarSarSar-OMe. Arrows indicate the deprotonation site. Bond lengths are in angstroms. These peptide methyl esters do not deprotonate well by ESI.

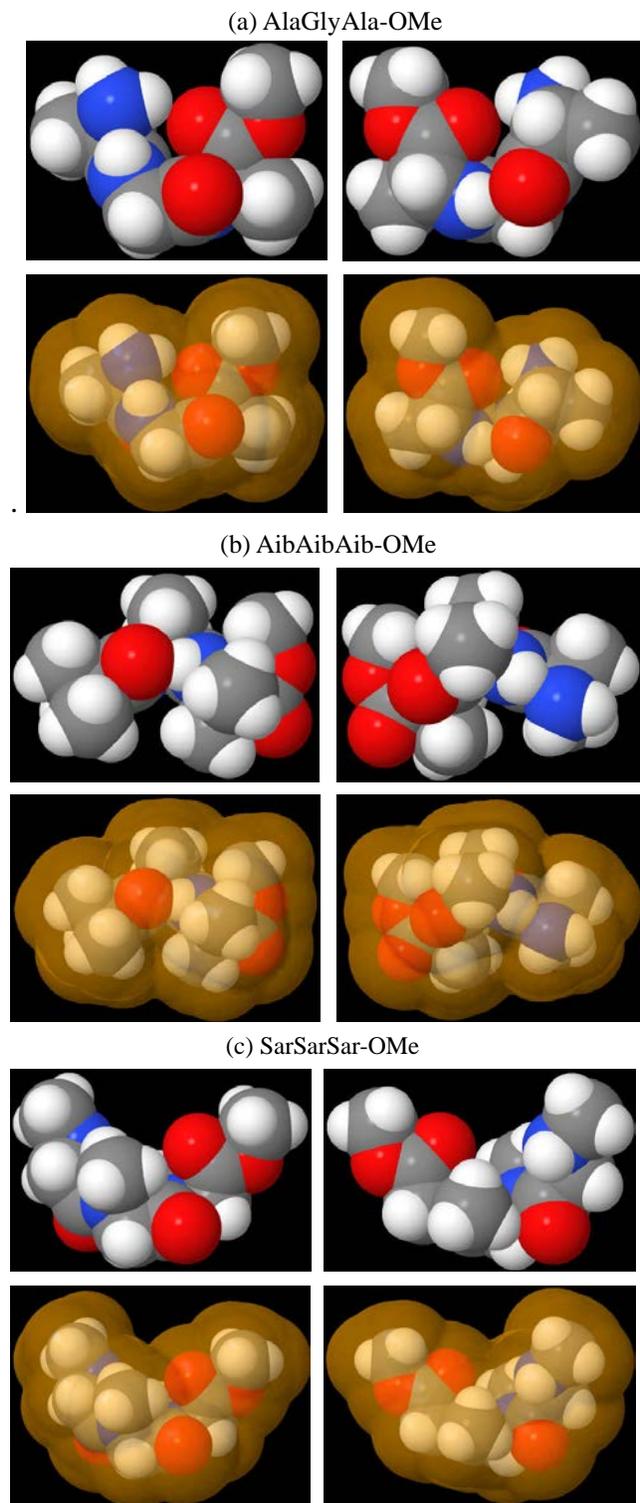
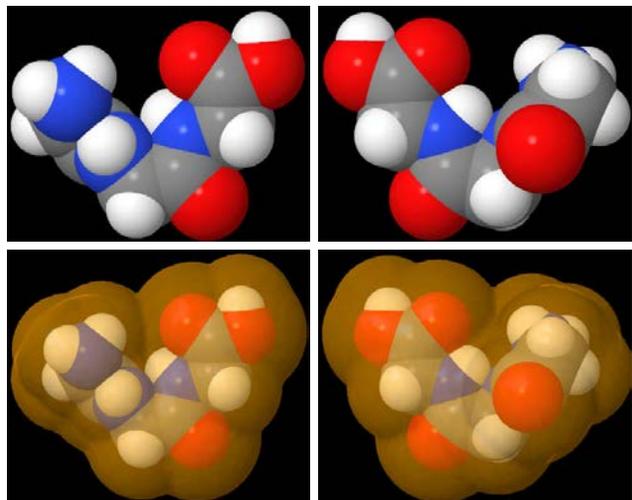
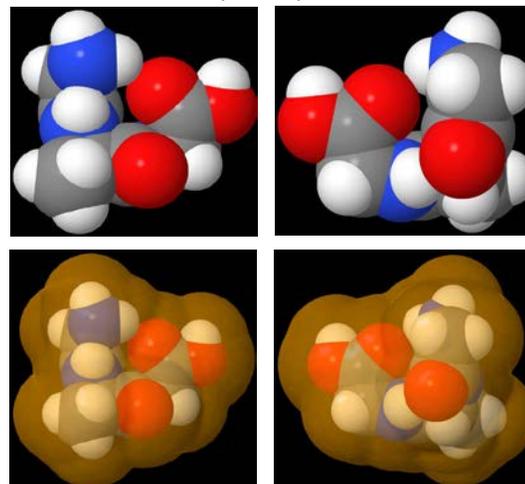


Figure 6.6. Van der Waals surfaces and solvent accessible surfaces for neutral (a) AlaGlyAla-OMe (b) AibAibAib-OMe (c) SarSarSar-OMe.

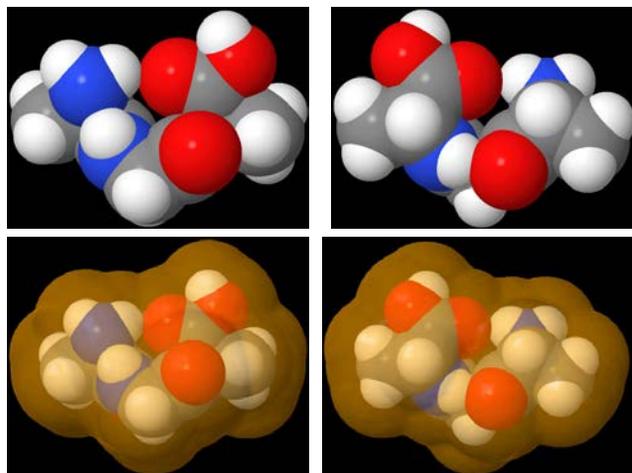
GlyGlyGly-OH



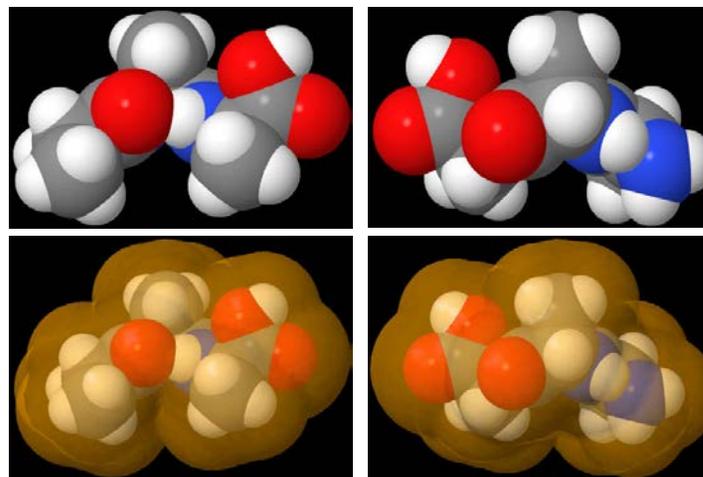
GlyAlaGly-OH



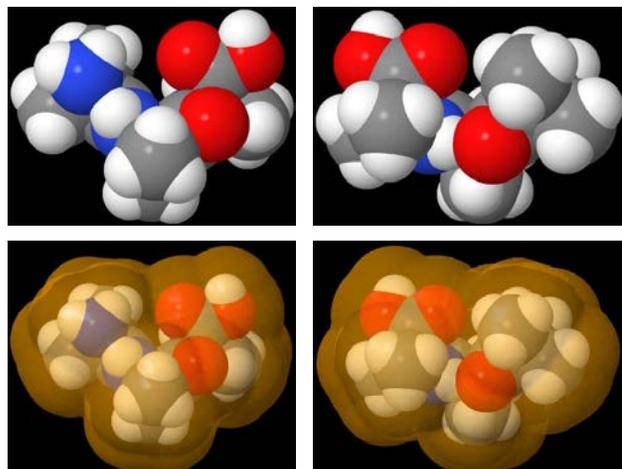
AlaGlyAla-OH



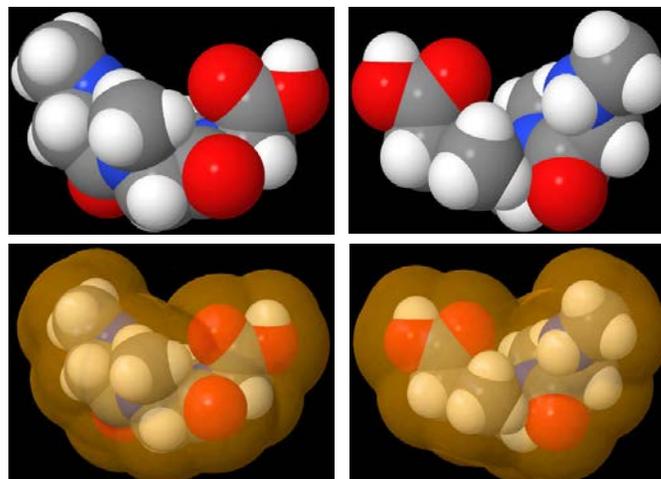
AlaAlaAla-OH



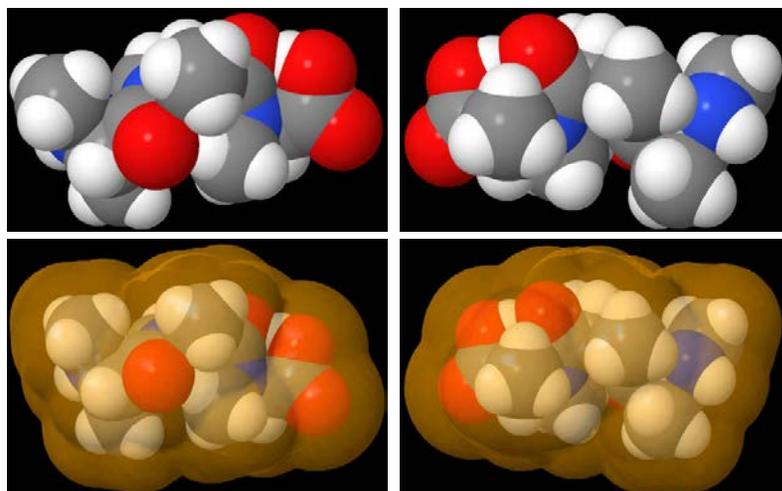
AibAibAib-OH



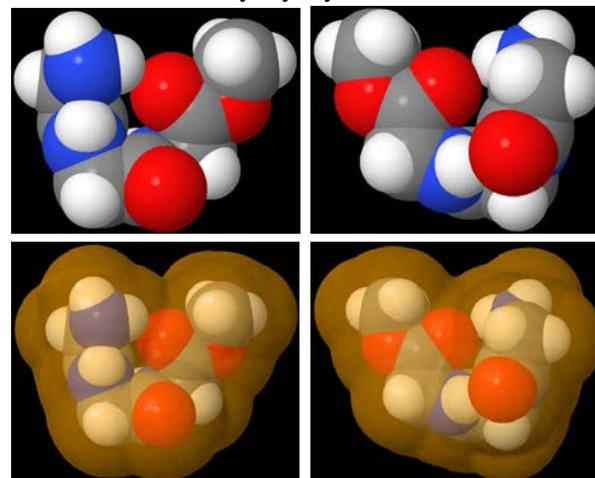
SarSarSar-OH



NmaNmaNma-OH



GlyGlyGly-OMe



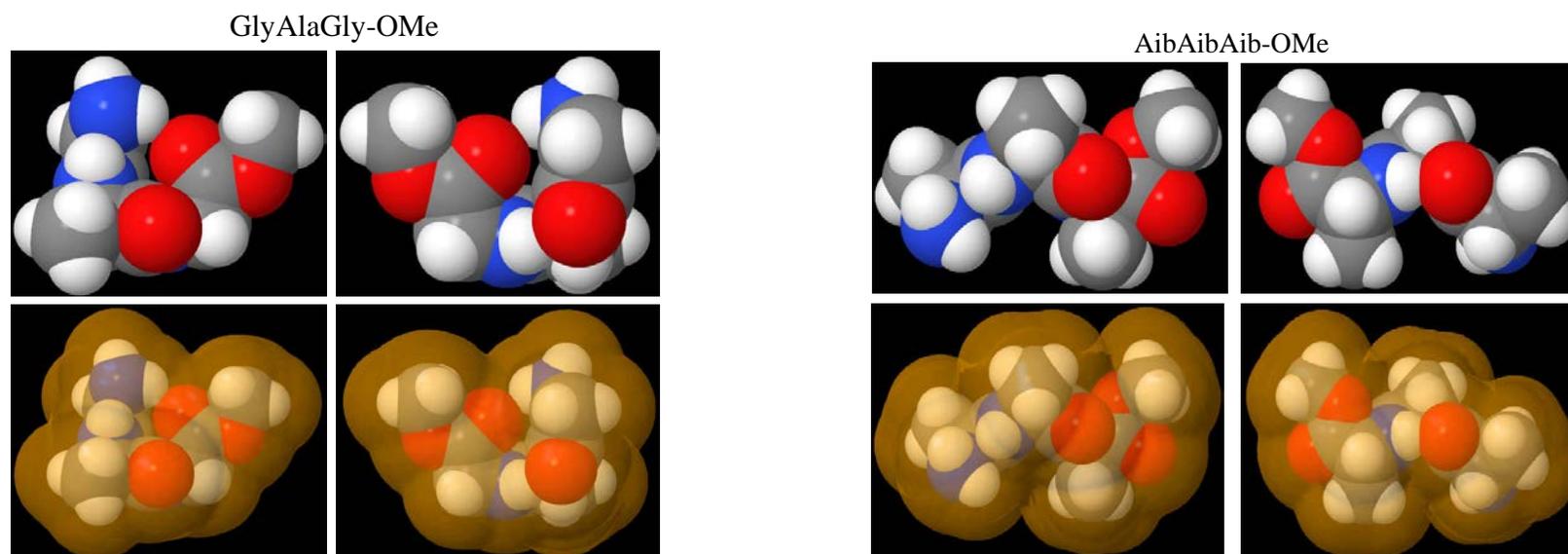


Figure 6.7. Van der waals surfaces and solvent accessible surfaces for the tripeptides that readily deprotonated with ESI.

References

- ¹ Ghelis, C.; Yon, J. In *Protein Folding*; Academic Press: New York, 1982.
- ² McCammon, J. A.; Harvey, S. C. In *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.
- ³ Margoulies, M.; Greenwood, F. C. In *Structure-Activity Relationships of Protein and Polypeptide Hormones*; Excerpta Medica Foundation: Amsterdam, 1972.
- ⁴ Wysocki, V. H.; Resing, K. A.; Zhang, Q. F.; Cheng, G. L. *Methods* **2005**, *35*, 211-222.
- ⁵ Papayannopoulos, I. A. *Mass Spectrom. Rev.* **1995**, *14*, 49-73.
- ⁶ Biemann, K. *Meth. Enzymol.* **1990**, *193*, 455-479.
- ⁷ Eckersley, M.; Bowie, J. H.; Hayes, R. N. *Org. Mass Spectrom.* **1989**, *24*, 597-602.
- ⁸ Waugh, R. J.; Bowie, J. H. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 169-173.
- ⁹ Pu, D.; Clipston, N. L.; Cassady, C. J. *J. Mass Spectrom.* **2010**, *45*, 297-305.
- ¹⁰ Ewing, N. P.; Cassady, C. J. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 105-116.
- ¹¹ Harrison, A. G.; Young, A. B. *Int. J. Mass Spectrom.* **2006**, *255-256*, 111-122.
- ¹² Marzluff, E. M.; Campbell, S.; Rogers, M. T.; Beauchamp, J. L. *J. Am. Chem. Soc.* **1994**, *116*, 7787-7796.
- ¹³ Li, Z.; Matus, M. M.; Velazquez, H. A.; Dixon, D. A.; Cassady, C. J. *Int. J. Mass Spectrom.* **2007**, *265*, 213-223.
- ¹⁴ Tan, J. P.; Ren, J. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 188-194.
- ¹⁵ Gao, J.; Cassady, C. J. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 4066-4072.
- ¹⁶ Bradford, A. M.; Waugh, R. J.; Bowie, J. H.; Vollmer, D. L.; Gross, M. L. *Int. J. Mass Spectrom. Ion Processes* **1994**, *136*, 143-153.
- ¹⁷ Raftery, M. J.; Bowie, J. H. *Int. J. Mass Spectrom. Ion Processes* **1988**, *85*, 167-186.
- ¹⁸ Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Brechi, L. A. *J. Mass Spectrom.* **2000**, *35*, 1399-1406.
- ¹⁹ Summerfield, S.; Gaskell, S. *Int. J. Mass Spectrom.* **1997**, *165*, 509-521.

-
- ²⁰ Bradford, A. M.; Waugh, R. J.; Bowie, J. H. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 677-685.
- ²¹ Steinborner, S. T.; Bowie, J. H. *Rapid Commun Mass Spectrom.* **1996**, *10*, 1243-1247.
- ²² Steinborner, S. T.; Bowie, J. H. *Rapid Commun Mass Spectrom.* **1997**, *11*, 253-258.
- ²³ Harrison, A. G. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 1-13.
- ²⁴ Bokatzian-Johnson, S. S.; Stover, M. L.; Dixon, D. A.; Cassady, C. J. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1544-1557.
- ²⁵ Wang, T.; Tran, T. T. N.; Calabrese, A. N.; Bowie, J. H. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1832-1840.
- ²⁶ Chass, G. A.; Marai, C. N. J.; Setiadi, D. H.; Csizmadia, I. G.; Harrison, A. G. *J. Mol. Struct. : THEOCHEM* **2004**, *675*, 149-162.
- ²⁷ Clipston, N. L.; Jai-nhuknan, J.; Cassady, C. J. *Int. J. Mass Spectrom.* **2003**, *222*, 363-381.
- ²⁸ Harrison, A. G.; Young, A. B. *J. Mass. Spectrom.* **2005**, *40*, 1173-1186.
- ²⁹ O'Hair, R. A. J.; Bowie, J. H.; Gronert, S. *Int. J. Mass Spectrom.* **1992**, *117*, 23-36.
- ³⁰ 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.
- ³¹ Tian, Z.; Pawlow, A.; Poutsma, J. C.; Kass, S. R. *J. Am. Chem. Soc.* **2007**, *129*, 5403-5407.
- ³² Locke, M. J.; McIver Jr., R. T. *J. Am. Chem. Soc.* **1983**, *105*, 4226-4232.
- ³³ Caldwell, G.; Renneboog, R.; Kebarle, P. *Can. J. Chem.* **1989**, *67*, 611-618.
- ³⁴ Fournier, F.; Afonso, C.; Fagin, A. E.; Gronert, S.; Tabet, J. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1887-1896.
- ³⁵ Tian, Z.; Wang, X.; Wang, L.; Kass, S. R. *J. Am. Chem. Soc.* **2009**, *131*, 1174-1181.
- ³⁶ Woo, H.; Lau, K.; Wang, X.; Wang, L. *J Phys Chem A* **2006**, *110*, 12603-6.
- ³⁷ Miao, R.; Jin, C.; Yang, G.; Hong, J.; Zhao, C.; Zhu, L. *J Phys Chem A* **2005**, *109*, 2340-2349.
- ³⁸ Tian, Z.; Kass, S. R. *J. Am. Chem. Soc.* **2008**, *130*, 10842-10843.
- ³⁹ Oomens, J.; Steill, J. D.; Redlich, B. *J. Am. Chem. Soc.* **2009**, *131*, 4310-4319.

-
- ⁴⁰ Stover, M. L.; Jackson, V.; Matus, M.; Adams, M.; Cassady, C. J.; Dixon, D. A. *J. Phys. Chem. B* **2012**, *116*, 2905-2916.
- ⁴¹ Skurski, P.; Gutowski, M.; Barrios, R.; Simons, J. *Chem. Phys. Lett.* **2001**, *337*, 143-150.
- ⁴² Morishetti, K. K.; Huang, B.; De Suan; Yates, J. M.; Ren, J. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 603-614.
- ⁴³ Ren, J.; Tan, J. P.; Harper, R. T. *J Phys Chem A* **2009**, *113*, 10903-10912.
- ⁴⁴ Shen, J.; Ren, J. *Int. J. Mass Spectrom.* **2012**, *316-318*, 147-156.
- ⁴⁵ de Koning, L. J.; Nibbering, N. M. M.; van Orden, S. L.; Laukien, F. H. *Int. J. Mass Spectrom.* **1997**, *165*, 209-219.
- ⁴⁶ Bartmess, J. E. In *Negative Ion Energetics Data*; Linstrom, P. J., Mallard, W. G., Eds.; NIST Chemistry WebBook, NIST Standard Reference Database Number 69; National Institute of Standards and Technology: Gaithersburg MD, 20899, <http://webbook.nist.gov>, (last updated March 29, 2012; retrieved September 24, 2012).
- ⁴⁷ Cassady, C. J.; Wronka, J.; Kruppa, G. H.; Laukien, F. H. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 394-400.
- ⁴⁸ Zhang, K.; Zimmerman, D. M.; Chung-Phillips, A.; Cassady, C. J. *J. Am. Chem. Soc.* **1993**, *115*, 10812-10822.
- ⁴⁹ Su, T.; Chesnavich, W. J. *J. Chem. Phys.* **1982**, *76*, 5183-5185.
- ⁵⁰ Su, T. *J. Chem. Phys.* **1988**, *89*, 5355-5355.
- ⁵¹ Bouchoux, G.; Salpin, J. Y. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 932-936.
- ⁵² Chan, W. C.; White, P. D. In *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*; Oxford University Press: New York, 2000; pp 345.
- ⁵³ Greene, T. W.; Wuts, P. G. M., Eds.; In *Protective Groups in Organic Synthesis*; John Wiley & Sons, Inc.: New York, 1991.
- ⁵⁴ M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Jr Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayshi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S.

Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuna, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox. Gaussian 09, Revision B.0; Gaussian, Inc, Wallingford, CT, 2010.

⁵⁵ Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

⁵⁶ Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785-789.

⁵⁷ Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem.* **92**, *70*, 560-571.

⁵⁸ Gutowski, K. E.; Dixon, D. A. *J. Phys. Chem. A* **2006**, *110*, 12044-12054.

⁵⁹ Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703-4709.

⁶⁰ Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379-IN4.

⁶¹ Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548-558.

⁶² Richmond, T. J. *J. Mol. Biol.* **1984**, *178*, 63-89.

⁶³ Connolly, M. L. *J. Mol. Graphics* **1993**, *11*, 139-141.

⁶⁴ Jmol: An open-source Java viewer for chemical structures in 3D. <http://www.jmol.org/> (accessed September 1, 2012).

⁶⁵ Cassady, C. J.; Carr, S. R.; Zhang, K.; Chung-Phillips, A. *J. Org. Chem.* **1995**, *60*, 1704-1712.

⁶⁶ Wu, J. Y.; Lebrilla, C. B. *J. Am. Chem. Soc.* **1993**, *115*, 3270-3275.

⁶⁷ Wu, Z.; Fenselau, C. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 863-866.

⁶⁸ Chase, M. W., Jr. *J. Phys. Chem. Ref. Data* **1998**, *Monograph 9*.

⁶⁹ Loo, R. R. O.; Smith, R. D. *J. Mass Spectrom.* **1995**, *30*, 339-347.

⁷⁰ Schnier, P. D.; Gross, D. S.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 1086-1097.

⁷¹ Price, W.; Schnier, P.; Williams, E. *Anal. Chem.* **1996**, *68*, 859-866.

⁷² Williams, E. *J. Mass Spectrom.* **1996**, *31*, 831-842.

⁷³ Hoaglund-Hyzer, C.; Counterman, A.; Clemmer, D. *Chem. Rev.* **1999**, *99*, 3037-3079.

**CHAPTER 6 APPENDIX: GAS-PHASE DEPROTONATION OF THE PEPTIDE
BACKBONE FOR TRIPEPTIDES AND THEIR METHYL ESTERS WITH HYDROGEN
AND METHYL SIDE CHAINS**

H₂₉₈ and G₂₉₈ values for all tripeptide and tripeptide methyl ester neutrals and anions at the G3MP2 level and higher energy picture of GlyGlyGly N-terminus anion that did not form a ring at the G3MP2 level

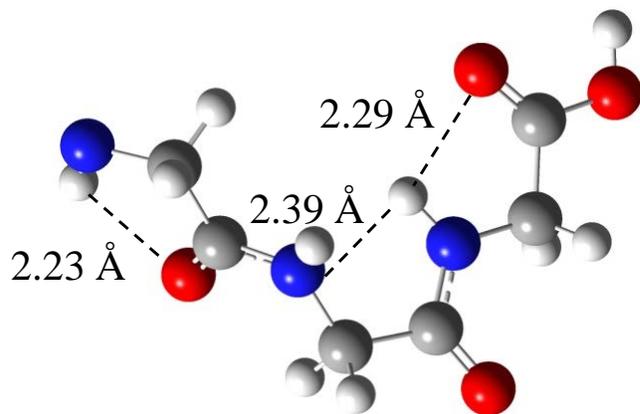
Table A6.1. H_{298} and G_{298} values for all Tripeptide and Tripeptide Methyl Ester Neutrals and Anions at the G3MP2 level.

Tripeptide	H_{298}	G_{298}
GlyGlyG-OH	-699.490986	-699.548232
C-terminus anion	-698.972274	-699.029886
C-terminal $C_{\alpha}H$ anion	-698.939639	-698.996708
C-terminal NH anion	-698.962525	-699.020830
Central $C_{\alpha}H$ anion	-698.938816	-698.996365
Central NH anion	-698.956301	-699.012068
N-terminal $C_{\alpha}H$ anion	-698.931935	-698.986779
N-terminus anion	-698.915294	-698.965062
GlyAlaGly-OH	-738.728872	-738.788848
C-terminus anion	-738.206915	-738.266311
C-terminal $C_{\alpha}H$ anion	-738.165968	-738.225787
C-terminal NH anion	-738.181239	-738.243878
Central $C_{\alpha}H$ anion	-738.148992	-738.211031
Central NH anion	-738.178525	-738.240852
N-terminal $C_{\alpha}H$ anion	-738.158587	-738.214135
N-terminus anion	-738.154773	-738.207604
AlaGlyAla-OH	-777.970969	-778.034033
C-terminus anion	-777.449801	-777.512554
C-terminal $C_{\alpha}H$ anion	-777.405910	-777.468082
C-terminal NH anion	-777.419439	-777.485241
Central $C_{\alpha}H$ anion	-777.392649	-777.457807
Central NH anion	-777.415862	-777.481788
N-terminal $C_{\alpha}H$ anion	-777.397057	-777.452884
N-terminus anion	-777.393869	-777.449869
AlaAlaAla-OH	-817.205337	-817.271507
C-terminus anion	-816.683415	-816.748967
C-terminal $C_{\alpha}H$ anion	-816.645683	-816.711086
C-terminal NH anion	-816.652608	-816.719257

Central C _α H anion	-816.629935	-816.696904
Central NH anion	-816.655901	-816.722648
N-terminal C _α H anion	-816.635929	-816.694758
N-terminus anion	-816.633566	-816.692582
AibAibAib-OH	-934.923055	-934.996548
C-terminus anion	-934.398395	-934.473236
C-terminal NH anion	-934.388423	-934.462428
Central NH anion	-934.386443	-934.460024
SarSarSar-OH	-817.154462	-817.223027
C-terminus anion	-816.631246	-816.697198
C-terminal C _α H anion	-816.591417	-816.657340
Central C _α H anion	-816.574537	-816.642138
N-terminal C _α H anion	-816.594902	-816.660097
N-terminus anion	-816.590420	-816.650214
NmaNmaNma-OH	-934.853027	-934.927156
C-terminus anion	-934.328823	-934.403197
C-terminal C _α H anion	-934.302879	-934.378890
Central C _α H anion	-934.317286	-934.392005
N-terminal C _α H anion	-934.334347	-934.402598
N-terminus anion	-934.336558	-934.404082
GlyGlyGly-OMe	-738.713128	-738.773866
C-terminal C _α H anion	-738.146417	-738.208110
C-terminal NH anion	-738.157485	-738.218607
Central C _α H anion	-738.130528	-738.192208
Central NH anion	-738.159195	-738.221927
N-terminal C _α H anion	-738.136648	-738.193113
N-terminus anion	-738.131532	-738.185426
GlyAlaGly-OMe	-777.947403	-778.011243
C-terminal C _α H anion	-777.379090	-777.443443
C-terminal NH anion	-777.394404	-777.460611
Central C _α H anion	-777.366343	-777.432296

Central NH anion	-777.395747	-777.461471
N-terminal C _α H anion	-777.376938	-777.436384
N-terminus anion	-777.371765	-777.430610
AlaGlyAla-OMe	-817.189858	-817.256797
C-terminal C _α H anion	-816.618924	-816.685562
C-terminal NH anion	-816.634748	-816.703309
Central C _α H anion	-816.600228	-816.669442
Central NH anion	-816.633988	-816.703845
N-terminal C _α H anion	-816.614754	-816.675974
N-terminus anion	-816.610105	-816.670202
AlaAlaAla-OMe	-856.424180	-856.494184
C-terminal C _α H anion	-855.860008	-855.929730
C-terminal NH anion	-855.875416	-855.946915
Central C _α H anion	-855.848190	-855.917813
Central NH anion	-855.873829	-855.944609
N-terminal C _α H anion	-855.854594	-855.918415
N-terminus anion	-855.849825	-855.912950
AibAibAib-OMe	-974.142664	-974.220278
C-terminal NH anion	-973.589876	-973.668639
Central NH anion	-973.566124	-973.644755
N-terminus anion	-973.558509	-973.631411
SarSarSar-OMe	-856.372833	-856.445445
C-terminal C _α H anion	-855.804667	-855.875062
Central C _α H anion	-855.789755	-855.861524
N-terminal C _α H anion	-855.811143	-855.877583
N-terminus anion	-855.807087	-855.871179

Figure A6.1. Higher energy pic of GlyGlyGly N-terminus anion that did not form a ring at the G3MP2 level.



GlyGlyGly N-terminus anion

CHAPTER 7: HEATS OF FORMATION AND ACIDITIES OF THE DIAMINOBENZENES AND HEATS OF FORMATION OF CARBOXYLIC ACIDS, PHENOL, AND CRESOLS

7.1 Introduction There is substantial interest in the disubstituted benzenes as monomers for polymers, yet they are not always straightforward to synthesize.¹ For example, dinitration of benzene followed by reduction can readily lead to m-diaminobenzene which is used in Nomex® aramid. P-diaminobenzene is the key monomer for Kevlar® and is difficult to synthesize with toxic intermediates. A possible way to make the para-isomer would be to isomerize the meta isomer to the para isomer if the energetics are favorable. Previously, we predicted¹ that order of stability is o- > m- > p- for the diaminobenzenes at the MP2/DZ+D(C) level. As our values for the cresols are in very good agreement with experiment considering the experimental error bars, we calculated the energies of the o-, m-, and p-substituted diaminobenzenes. The heats of formation of the diaminobenzenes were recently determined by Riberio da Silva and coworkers² experimentally using static bomb calorimetry and computationally at the G3(MP2) level. In this work, we calculated the gas-phase acidities (GAs) by deprotonating one of the amine groups. We recalculated the heats of formation of the diaminobenzenes at the G3(MP2) level as part of our study of the acidities.

7.2 Computational Methods Calculations were performed at the density functional theory (DFT) and correlated molecular orbital (MO) theory levels with the programs Gaussian-09.³ Geometries were initially optimized at the DFT level with the B3LYP exchange-correlation functional^{4,5} and the DZVP2 basis set.⁶ 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 the lowest energy structures for the ortho, meta, and para amine-substituted diaminobenzenes expanding on our previous work^{7,8} on the gas-phase acidities of strong organic acids, phenol, and the three cresols. The prior work showed that calculations at the MP2/CBS (CBS = complete basis set) level with the augmented correlation-consistent basis sets up through the quadruple level⁹ predicted the acidities of organic acids to better than 4 kcal/mol with the calculated values more acidic than the experimental values. The prior work also showed that the G3(MP2) method¹⁰ improved the agreement for the acidities with the experimental values and/or the coupled cluster CCSD(T)/CBS (complete basis set) values to within about 1 kcal/mol. G3(MP2) has an additional advantage over DFT methods in terms of reliable predictions for these types of compounds as the correlated molecular orbital methods in G3(MP2) perform better in the prediction of hydrogen bond energies as well as steric non-bonded interactions than do most widely-used DFT exchange-correlation functionals.

The heats of formation of the ortho, meta, and para amine-substituted diaminobenzenes and their anions were calculated from atomization energies at the G3(MP2) level. Additionally, the G3(MP2) calculated heats of formation of the previously studied 11 organic acids, phenol, and three cresols were also predicted.

7.3 Results and Discussion *Diaminobenzenes* Optimized structures of the diaminobenzene neutrals and anions are given in Figure 1. Hydrogen bonding was only present in the *o*-diaminobenzene neutral and anion between the amines. In the neutral there are two hydrogen bonds of equal length (2.45 Å) with each amine forming a bond to the other. In the anion, there is only one hydrogen bond from the NH₂ to the NH with a distance of 2.05 Å. For *m*- and *p*-

diaminobenzenes, two structures were studied, where the NH₂ groups are pointed in the same and opposite directions. The rotation of the amine group had a negligible effect on the GAs and gas-phase heats of formation, ~ 1 kcal/mol.

As in our previous work,⁸ the gas-phase acidities for the *ortho*, *meta*, and *para* diaminobenzenes have been calculated at different computational levels.^{11,12,13,14} The GAs of the diaminobenzenes and the relative energy differences between the anions are given in Table 2. At the G3(MP2) level, the *ortho* isomer is the most stable in the gas phase with the *meta* isomer 0.4 kcal/mol higher in energy and the *para* isomer 2.6 kcal/mol higher in energy on the enthalpy scale at 298 K. On the free energy scale at 298 K, the *ortho* isomer is still the most stable and the *meta* and *para* isomers are 0.3 and 1.8 kcal/mol respectively higher in energy on the free energy scale. The current results confirm the ordering given by the previous lower level MP2 calculations¹ at 0 K and the other G3MP2 results.² The results are consistent with the previous conclusion¹ that it will not be possible to use a catalyst to convert meta-diaminobenzene into para-diaminobenzene due to thermodynamic constraints.

For the diaminobenzenes at all Gx levels of theory, the *ortho* substituted diaminobenzene was ~ 3 to 3.5 kcal/mol more acidic with respect to deprotonation enthalpy and free energy than the corresponding *meta* isomer. *o*-diaminobenzene is more acidic than *p*-diaminobenzene by ~5.5 to 6 kcal/mol in all cases except for the G3 and G3MP2 values where the energy difference was slightly larger, ~7 kcal/mol.. The large differences in energy led to an easy assignment of the acidities of the substitutions (*ortho* > *meta* > *para*) which are the same results as found for the three cresols. Aromatic amines are more acidic than aliphatic amines by ~ 20 kcal/mol. For example, *p*-diaminobenzene is by 31 and 24 kcal/mol more acidic than methylamine and dimethylamine respectively.^{15,16}

The G3(MP2) calculated and experimental² heats of formation of *o*-, *m*-, and *p*-diaminobenzenes and anions are given in Table 1. The predicted relative energy differences of the neutrals and anions are also given. The experimental heats of formation of the anions were calculated using the experimental heats of formation of the neutral² and proton ($\Delta H_f(\text{H}^+) = 365.7$ kcal/mol)¹⁷ and the calculated deprotonation enthalpy. The energy expression is given as:

$$\Delta H_f(\text{anion}) = \Delta H_{\text{rxn}}(\text{G3(MP2)}) - \Delta H_f(\text{H}^+) + \Delta H_f(\text{neutral}) \quad (1)$$

Excellent agreement is found between the calculated and experimental results, within ~ 1 kcal/mol, so-called ‘chemical accuracy’.

Gas-Phase Heats of Formation of Carboxylic Acids, Phenol, and Cresols The calculated and experimental heats of formation of phenol and the three cresols and carboxylic acid neutrals are given in Tables 2 and 3 respectively. The optimized structures for these molecules are given in our previous work.⁸ For the carboxylic acids, the calculated and experimental heats of formation^{18,19,20,21,22,23,24,25,26,27,28,29} are all within 1 kcal/mol except for pentanoic acid where some of the older values²¹ differ from the calculated values by ~ 2 kcal/mol. Of the carboxylic acid neutrals studied, there were three sets of isomers, butanoic and isobutyric acid, isovaleric and trimethylacetic acid, and isohexanoic and tert-butylacetic acid. For all three cases, the differences in the heats of formation of the isomers are dependent upon the size of the molecule. For the smallest isomers, there was not a large change, ~ 1 kcal/mol, and this almost triples for the largest isomers to ~ 3.5 kcal/mol.

For the phenols, the calculated and experimental heats of formation^{30,31,32,33,34} are all within 1 kcal/mol except for *m*-cresol where an older heat of formation³² obtained from the heat of formation of the solid and the heat of sublimation is ~ 3 kcal/mol greater than the calculated value. The three other experimental values for *m*-cresol do match the calculated value suggesting

that there is an error in the experimental value that deviates. The calculated values show that *o*-cresol is more stable than *m*-cresol and *p*-cresol by 0.3 and 0.9 kcal/mol respectively. This differs from the experimental trend which has *m*-cresol as more stable than *o*-cresol by ~1 kcal/mol. At all Gx levels of theory, *o*-cresol was ~ 1 kcal/mol more stable than *m*-cresol in enthalpy and free energy except for the G3B3 and G3MP2B3 levels of theory where the change in energy was larger, ~ 2 kcal/mol. *o*-Cresol was ~ 2 kcal/mol more stable than *p*-cresol in terms of the enthalpy and free energy for the Gx levels of theory except for the G3 and G3MP2 levels of theory where the change in energy was smaller, ~ 1.5 kcal/mol. .

7.4 Conclusions Gas-phase acidities and heats of formation for the three diaminobenzenes and anions were predicted at the G3(MP2) level of theory. For the diaminobenzenes, *ortho* substitution generated both the most stable neutral and anion. The GAs of the three diaminobenzenes were determined to provide further comparison for the trends developed from the study of the cresols at a variety of computational levels. The trend in the GAs of the diaminobenzenes was exactly the same as that of the cresols, (*ortho* > *meta* > *para*). For the cresols, *ortho* substitution generated the most stable neutral on the enthalpy scale whereas *meta* substitution generated the most stable neutral isomer on the free energy scale.

Acknowledgement This work was supported in part by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, U.S. Department of Energy (DOE) under grant no. DE-FG02-03ER15481 (Catalysis Center Program). D. A. Dixon thanks the Robert Ramsay Fund from the University of Alabama for its partial support.

Table 7.1. Calculated Gas-Phase Acidities of *o*-, *m*-, and *p*- Diaminobenzenes in kcal/mol at Different Computational Levels.

Substituted Anions	Property	B3LYP	G3	G3MP2	G3B3	G3MP2B3	G4	G4MP2	MP2/aD
<i>o</i> -diaminobenzene	ΔH	367.6	365.9	365.1	365.5	364.8	365.1	364.5	360.2
	ΔG	360.2	357.9	357.1	358.0	357.3	357.6	357.0	352.7
<i>m</i> -diaminobenzene	ΔH	370.9	368.9	368.2	368.6	367.8	368.3	367.7	363.8
	ΔG	363.4	361.5	360.8	361.1	360.3	360.8	360.1	356.3
<i>p</i> -diaminobenzene	ΔH	373.3	371.6	370.9	371.3	370.5	371.0	370.4	366.5
	ΔG	365.7	364.7	364.0	363.4	362.8	363.1	362.5	358.6
<i>m</i> -diaminobenzene –	$\Delta\Delta H$	3.3	3.0	3.1	3.1	3.0	3.2	3.2	3.6
<i>o</i> -diaminobenzene	$\Delta\Delta G$	3.2	3.6	3.7	3.1	3.0	3.2	3.1	3.6
<i>p</i> -diaminobenzene –	$\Delta\Delta H$	5.7	5.7	5.8	5.8	5.7	5.9	5.9	6.3
<i>o</i> -diaminobenzene	$\Delta\Delta G$	5.5	6.8	6.9	5.4	5.5	5.5	5.5	5.9

Table 7.2. Calculated Heats of Formation at 298 K in kal/mol at the G3(MP2) level of the Phenols and Diaminobenzenes.

Molecule	ΔH_f neutral	ΔH_f expt neutral	$\Delta\Delta H$ neutral	$\Delta\Delta G$ neutral	ΔH_f anion	ΔH_f expt anion	$\Delta\Delta H$ anion	$\Delta\Delta G$ anion
O-diaminobenzene	21.7	20.7 ± 0.4^2	0.0	0.0	19.9	20.1 ± 0.4	0.0	0.0
M-diaminobenzene	22.1	21.4 ± 0.4^2	0.4	0.3	23.3	23.9 ± 0.4	3.4	3.9
P-diaminobenzene	24.3	23.8 ± 0.4^2	2.6	1.8	28.2	29.0 ± 0.4	8.3	8.7
Phenol	-23.1	-22.8^{30} -22.5^{32} -23.1 ± 0.2^{33} -23.0 ± 0.1^{34}			-40.7	-39.2 ± 0.1		
O-cresol	-31.4	-30.7 ± 0.2^{33} -30.7 ± 0.2^{34}	0.0	0.9	-50.4	-47.8 ± 0.2	0.0	0.0
M-cresol	-31.1	-32.0^{30} -28.0^{31} -31.6 ± 0.3^{33} -31.9 ± 0.3^{34}	0.3	0.0	-49.7	-47.7 ± 0.3	1.5	0.5
P-cresol	-30.5	-30.0 ± 0.4^{33} -29.9 ± 0.4^{34}	0.9	1.0	-46.8	-44.9 ± 0.4	3.0	1.7

Table 7.3. Calculated Heats of Formation at 298 K in kal/mol at the G3(MP2) level of Carboxylic Acids.

Carboxylic Acid	Formula	ΔH_f 298 K G3MP2 TAE	ΔH_f 298 K Experiment
			-90.5 ± 0.1 ^{19,25}
			-90.4 ^{25,26}
Formic acid	CH ₂ O ₂	-90.0	-90.6 ± 0.1 ^{19,27}
			-90.6 ^{26,27}
			-90.5 ²⁶
			-104.1 ± 1.0 ^{19,22}
Acetic acid	C ₂ H ₄ O ₂	-102.5	-103.4 ± 0.4 ^{19,24}
			-103.5 ± 0.4 ^{19,27}
			-103.2 ± 0.4 ^{19,29}
Propanoic acid	C ₃ H ₆ O ₂	-107.7	-108.7 ± 0.8 ^{18,19}
			-108.9 ± 0.5 ^{19,27}
Isobutyric acid (2-Methylpropanoic acid)	C ₄ H ₈ O ₂	-113.9	
Trimethylacetic acid (2,2-Dimethylpropanoic acid)	C ₅ H ₁₀ O ₂	-121.5	
Butanoic acid	C ₄ H ₈ O ₂	-112.7	-113.7 ± 1.0 ^{19,27}
Isovaleric acid (3-Methylbutanoic acid)	C ₅ H ₁₀ O ₂	-119.8	-120.1 ± 1.6 ²³
Tert-butylacetic acid	C ₆ H ₁₂ O ₂	-127.0	
			-119.7 ± 1.0 ^{20,21}
			-115.7 ± 1.7 ^{21,23}
Pentanoic acid	C ₅ H ₁₀ O ₂	-117.8	-119.0 ± 1.0 ^{21,27}
			-118.7 ± 1.0 ^{21,28}
Isohexanoic acid (4-Methylpentanoic acid)	C ₆ H ₁₂ O ₂	-124.4	
Tert-butyl-propanoic acid	C ₇ H ₁₄ O ₂	-132.2	

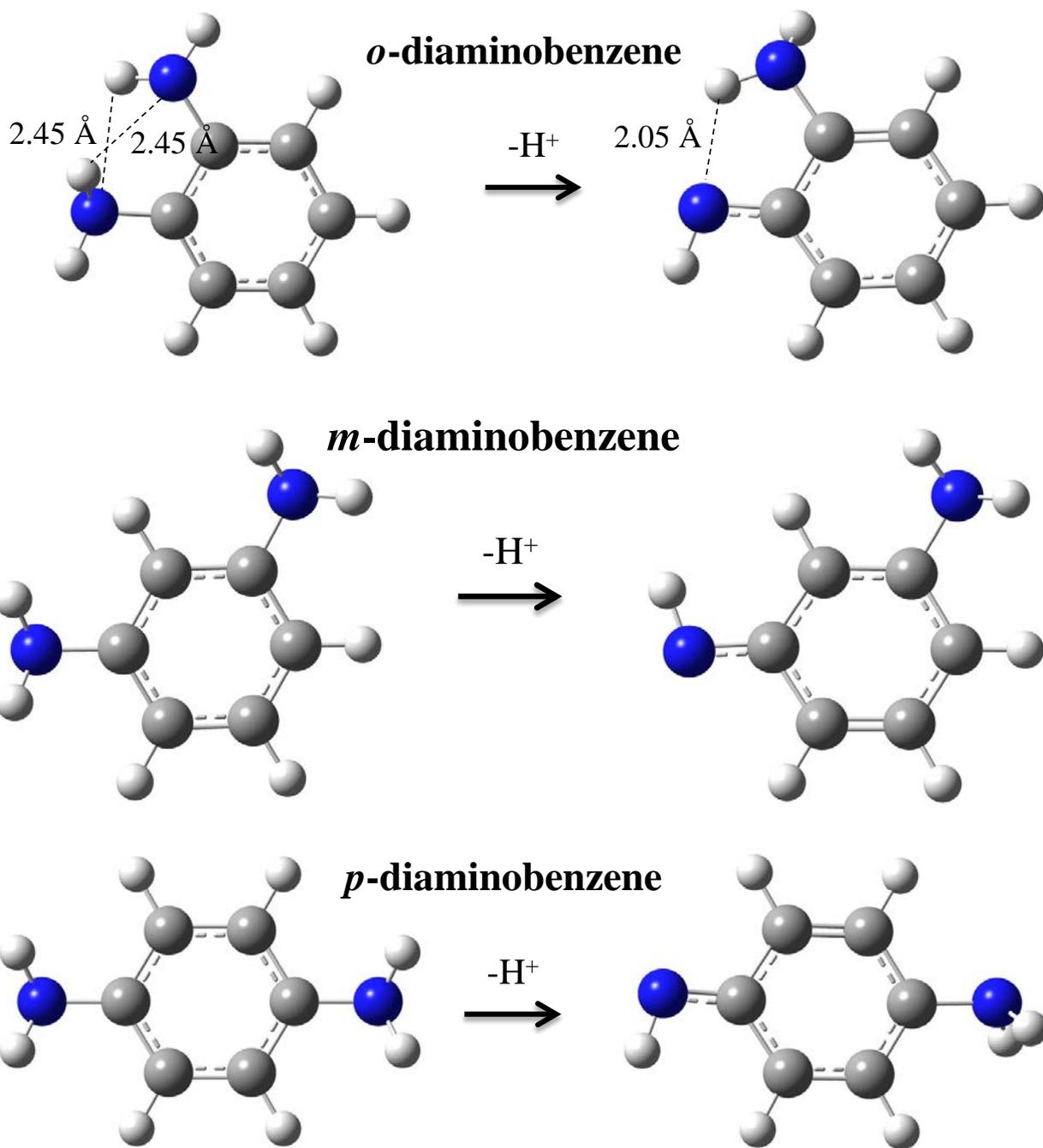


Figure 7.1. G3(MP2) optimized structures for *o*-, *m*- and *p*-diaminobenzene and anions. Important hydrogen bonds in Å.

References

- ¹ "The Relative Energies of Amino, Hydroxo and Fluoro Disubstituted Benzenes," D. A Dixon, in *ACS Symposium Series 404, Computer Applications in Applied Polymer Science II*, Ed. T. Provder, **1989**, Chapt. 15, p. 147-159.
- ² Santos, A. F. L. O. M.; Ribeiro da Silva, M. A. V. Diaminobenzenes: An Experimental and Computational Study. *J. Phys Chem. B*, **2011**, 115, 4939-4948.
- ³ Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Fox, D. J. *Gaussian 09*, revision C.01; Gaussian, Inc.: Wallingford, CT, 2010.
- ⁴ Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, 98, 5648-5652.
- ⁵ Lee, C.; Yang, W.; Parr, R. G. Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy. *Physical Review B*. **1988**, 37, 785-789.
- ⁶ Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. Optimization of Gaussian-Type Basis-Sets for Local Spin-Density Functionals Calculations. 1. Boron Through Neon, Optimization Technique and Validation. *Can. J. Chem.* **1992**, 70, 560-571.
- ⁷ Gutowski, K. E.; Dixon, D. A. Ab Initio Prediction of the Gas- and Solution-Phase Acidities of Strong Bronsted Acids: The Calculation of pK_a Values Less Than -10. *J. Phys. Chem. A* **2006**, 110, 12044-12054.
- ⁸ Bokatzian-Johnson, S. S.; Stover, M. L.; Plummer, C. E.; Dixon, D. A.; Cassady, C. J. An experimental and Computational Investigation into the Gas-Phase Acidities of Tyrosine and Phenylalanine: Three Structures for Deprotonated Tyrosine. *J. Phys. Chem. B* **2014**, 118, 12630-12643.
- ⁹ Kendall, R. A.; Dunning, T. H. Jr.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, 96, 6796-6806.
- ¹⁰ Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. Gaussian-3 Theory Using Reduced Moller-Plesset Order. *J. Chem. Phys.* **1999**, 110, 4703-4709.

-
- ¹¹ Curtiss, L. A.; Raghavachari, K. Gaussian-3 (G3) Theory for Molecules Containing First and Second-Row Atoms. *J. Chem. Phys.* **1998**, *109*, 7764-7776.
- ¹² Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-3 Theory using Density Functional Geometries and Zero-Point Energies. *J. Chem. Phys.* **1999**, *110*, 7650-7657.
- ¹³ Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 Theory. *J. Chem. Phys.* **2007**, *126*, 084108/1-084108/12.
- ¹⁴ Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 Theory using Reduced Order Perturbation Theory. *J. Chem. Phys.* **2007**, *127*, 124105/1-124105/8.
- ¹⁵ Radisic, D.; Xu, S. J.; Bowen, K. H. Photoelectron spectroscopy of the anions, CH_3NH^- and $(\text{CH}_3)_2\text{N}^-$ and the anion complexes, $\text{H}^-(\text{CH}_3\text{NH}_2)$ and $(\text{CH}_3)_2\text{N}^-[(\text{CH}_3)_2\text{NH}]$. *Chem. Phys. Lett.* **2002**, *354*, 9-13.
- ¹⁶ MacKay, G. I.; Hemsworth, R. S.; Bohme, D. K. Absolute Gas-Phase Acidities of the CH_3NH_2 , $\text{C}_2\text{H}_5\text{NH}_2$, $(\text{CH}_3)_2\text{NH}$, and $(\text{CH}_3)_3\text{N}$. *Can. J. Chem.* **1976**, *54*, 1624-1630.
- ¹⁷ Chase, M. W., Jr. NIST-JANAF Thermochemical Tables, 4th ed.; Journal of Physical Chemistry Reference Data, Monograph 9; American Institute of Physics: Woodbury, NY, 1998; Suppl 1, pp 1-1951.
- ¹⁸ Schjanberg, E. Die Verbrennungswarmen Und Die Refraktionsdaten Einiger Chlorsubstituierter Fettsauren und Ester. *Z. Phys. Chem. Abt. A*, **1935**, *172*, 197-233.
- ¹⁹ Konicek, J.; Wadso, I. Enthalpies of Vaporization of Organic Compounds. VII. Some Carboxylic Acids. *Acta Chem. Scand.*, **1970**, *24*, 2612-2626.
- ²⁰ Schjanberg, E. Die Verbrennungswarmen Und Die Refraktionsdaten Einiger Chlorsubstituierter Fettsauren Und Ester. *Z. Phys. Chem. Abt. A*, **1935**, *172*, 197-233.
- ²¹ Kruif, C. G.; Oonk, H. A. J. Enthalpies of Vaporization and Vapour Pressures of Seven Aliphatic Carboxylic Acids. *J. Chem. Thermodyn.*, **1979**, *11*, 287-290.
- ²² Carson, A. S.; Skinner, H. A. 201. Carbon-Halogen Bond Energies in the Acetyl Halides. *J. Chem. Soc.*, **1949**, *0*, 936-939.
- ²³ Hancock, C. K.; Watson, G. M.; Gilby, R. F. Heats of Combustion of Five-Carbon Fatty Acids and their Methyl Esters. *J. Phys. Chem.*, **1954**, *58*, 127-129.
- ²⁴ Evans, F. W.; Skinner, H. A. The Heat of Combustion of Acetic Acid. *Trans. Faraday Soc.*, **1959**, *55*, 260-261.
- ²⁵ Sinke, G. C. The Heat of Formation of Formic Acid. *J. Phys. Chem.*, **1959**, *63*, 2063-2063.

-
- ²⁶ Guthrie, J. P. Hydration of Carboxamides. Evaluation of the Free Energy Change for Addition of Water to Acetamide and Formamide Derivatives. *J. Am. Chem. Soc.*, **1974**, *96*, 3608-3615.
- ²⁷ Lebedeva, N. D. Heats of Combustion of Monocarboxylic Acids. *Russ. J. Phys. Chem. (Engl. Transl.)*, **1964**, *38*, 1435-1437.
- ²⁸ Adriaarise, N.; Dekker, H., Coops, J. Heats of Combustion of Normal Saturated Fatty Acids and their Methyl Esters. *Rec. Trav. Chim. Pays/Bas*, **1965**, *84*, 393-407.
- ²⁹ Steele, W. V.; Chirico, R. D.; Cowell, A. B.; Knipmeyer, S. E.; Nguyen, A. Thermodynamic Properties and Ideal-Gas Enthalpies of Formation for 2-aminoisobutyric acid (2-methylalanine), acetic acid (4-methyl-3-penten-2-one), 4-methylpent-1-ene, 2,2'-bis(phenylthio)propane, and glycidyl phenyl ether (1,2-epoxy-3-phenoxypropane). *J. Chem. Eng. Data*, **1997**, *42*, 1052,1066.
- ³⁰ Badoche, M. No 19. – Chaleurs de Combustion du Phenol, du-m-cresol et del leurs ethers; par M. Marius BADOCHÉ. *Bull. Soc. Chim. Fr.*, **1941**, *8*, 212-220.
- ³¹ Cox, J. D. The Heats of Combustion of Phenol and the Three Cresols. *Pure Appl. Chem.*, **1961**, *2*, 125-128.
- ³² Pushin, N. A. Heats of Combustion and Heats of Formation of Isomeric Organic Compounds. *Bull. Soc. Chim. Belgrade*, **1954**, *19*, 531-547.
- ³³ Parks, G. S.; Manchester, K. E.; Vaughan, L. M. Heats of combustion and Formation of Some Alcohols, Phenols, and Ketones. *J. Chem. Phys.*, **1954**, *22*, 2089-2090.
- ³⁴ Andon, R. J. L.; Biddiscombe, D. P.; Cox, J. D.; Handley, R.; Harrop, D.; Herington, E. F. G.; Martin, J. F. Thermodynamic Properties of Organic Oxygen Compounds. Part I. Preparation and Physical Properties of Pure Phenol, Cresols, and Xylenols. *J. Chem. Soc.*, **1960**, 5246-5254.

**CHAPTER 7 APPENDIX: HEATS OF FORMATION AND ACIDITIES OF THE
DIAMINOBENZENES AND HEATS OF FORMATION OF CARBOXYLIC ACIDS,
PHENOL, AND CRESOLS**

Table A7.1. H_{298} and G_{298} Total Energies for all Neutral Diaminobenzenes and Anions at the G3(MP2) level (a.u.)

Molecule	H_{298}	G_{298}
<i>o</i> -diaminobenzene neutral	-342.390074	-342.429255
<i>o</i> -diaminobenzene anion	-341.810540	-341.850145
<i>m</i> -diaminobenzene neutral	-342.389443	-342.428853
<i>m</i> -diaminobenzene anion	-341.805080	-341.843893
<i>p</i> -diaminobenzene neutral	-342.385977	-342.426322
<i>p</i> -diaminobenzene anion	-341.797264	-341.836312

CHAPTER 8: CONCLUSIONS

The goal of this dissertation is to provide detailed and reliable thermodynamic quantities of the properties of negative ions of amino acids and peptides, including different substituents, by using the G3(MP2) composite correlated molecular orbital theory method. These values have been carefully benchmarked against the mass spectrometry experiments performed in the Cassady group at The University of Alabama or against the available experimental data from the literature. Overall, excellent agreement between our calculated results and the available experimental data was found. Thus we expect our calculated gas phase acidities to be good to ± 1 kcal/mol. Our calculated data provides unique insights into the molecular structures that cannot be obtained from experiment including the important role of hydrogen bonding in determining the lowest energy structure and hence the acidity. In general, the proton is lost from the carboxylic acid group to form the carboxylate anion; however, if this site is unavailable, then deprotonation from the side chains of amino acids and backbones of peptides to be possible. Gas-phase heats of formation and solution-phase acidities were also determined for amino acids and phosphorylated amino acid amides to provide a more detailed picture of the thermodynamic properties of these molecules. The thermodynamic results can then be used by experimentalists to improve negative ion-based mass spectrometry proteomics.

Chapter 2 describes studies of the 20 common L-amino acids. The gas-phase acidities (GAs) 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 carboxylic acid site whereas cysteine is predicted to be deprotonated at the thiol group with proton transfer between the $-\text{CO}_2\text{H}$ and $-\text{S}^-$ groups. In general, deprotonation of the $-\text{CO}_2\text{H}$ group results in GAs between 329 and 335 kcal/mol showing that substituent effects are not very large. For the acidic amino acids aspartic and glutamic acid, the GAs are lower, ~ 316 kcal/mol. Self-consistent reaction field (SCRFF) calculations with the COSMO parameterization were used to predict the pK_a 's in aqueous solution. The differences in the pK_a values were used to estimate the free energy difference between the zwitterion and non-zwitterion 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 and are within 4 kcal/mol of the available experimental values. Further calculations were performed on 5 rare amino acids to predict their GAs, heats of formation, and pK_a values. The results show that substitution effects on these rare amino acids are relatively small as the GAs only differ from their parent amino acids by ≤ 7 kcal/mol.

Chapter 3 describes further computational and experimental studies on tyrosine and phenylalanine in the gas phase to better determine the site of deprotonation in tyrosine. Three deprotonated structures were revealed for tyrosine by a combination of correlated molecular orbital theory and mass spectrometry. The structures were predicted at the G3(MP2) level of theory and distinguished experimentally by ion/molecule reactions involving proton transfer and trimethylsilyl azide. The gas-phase acidities predicted at the G3(MP2) level agree well with the results from the proton transfer reactions. The lowest energy structure, which was only observed

experimentally using electrospray ionization from aprotic solvents, is deprotonated at the carboxylic acid group and is predicted to be highly folded. A second unfolded carboxylate structure is 8.8 kcal/mol higher in energy and primarily forms from protic solvents. Protic solvents also yield a structure deprotonated at the phenolic side chain, which experiments find to be intermediate in energy to the two carboxylate forms. The G3MP2 calculations were done prior to the experiment and indicate that the three structures differ in energy by only 2.5 kcal/mol, yet they are readily distinguished experimentally. For the phenylalanine, which lacks a phenolic group, only one deprotonated structure was observed experimentally when electrosprayed from protic solvent. This agrees with G3(MP2) calculations that find the folded and unfolded carboxylate forms to differ by 0.3 kcal/mol.

Chapter 3 also reports additional benchmarking of our computational G3(MP2) method. Because of the difficulty of getting amino acids into the gas phase, the amino acid anions are generated in solution and electrosprayed into the mass spectrometer. The gas phase acidity is thus determined by a one-way bracketing technique using acids whose acidities are known from experiment or highly accurate calculations. However, there is often a lack of good experimental data on the gas phase-acidities of molecules needed for bracketing studies in ion molecule reactions to determine the gas-phase acidities of biomolecules, for example, amino acids and peptides. Gas- and solution-phase acidities and heats of formation for 11 carboxylic acids, phenol, and *ortho*-, *meta*-, and *para*-cresol were predicted to study the effects of the methyl groups on the acidities and stabilities of the molecules. Substitutions were studied with the phenolic rings to determine the importance of the location of the methyl group in the cresols. The lowest energy neutral and anionic conformers were predicted at the correlated composite G3MP2 molecular orbital theory level. Excellent agreement is found with the available experimental

ΔH_{acid} and GA values further benchmarking our computational method. For the carboxylic acids, the effect of the length of the carbon chain and the number of methyl groups present were studied. In general, the addition of methyl groups to the carboxylic acids led to an increase in the gas and solution phase acidities. For the cresols, *ortho* substitution generated the most stable neutral on the enthalpy scale whereas *meta* substitution generated the most stable neutral isomer on the free energy scale. Deprotonation of *O*-cresol led to the most stable anionic isomer on both the enthalpy and free energy scales. *Ortho* and *para* substitution generated the isomers with the greatest acidity in the gas and solution phases respectively.

As most residues in peptides do not have a terminal carboxylic acid end group, the amino acid amides were studied as they provide a better model of amino acid residues in peptides as described in Chapter 4. By building on previous work done on aspartic and glutamic acid amides, correlated molecular orbital theory at the G3(MP2) level and proton-transfer reactions in a Fourier transform ion cyclotron resonance mass spectrometer were used to predict the gas phase acidities (GAs) of the remaining amino acid amides. The amino acid amides were formed by modification of the C-terminal carboxylic acid (-CO₂H) to the amide (-CONH₂). For compounds whose most acidic site is the C-terminal amide nitrogen, two ions populations were observed experimentally with GAs that differ by 4-7 kcal/mol. The lower energy, more acidic structure accounts for the majority of the ions (~65 %) formed experimentally by electrospray ionization. G3(MP2) calculations predict that the lowest energy anionic conformer has a cis-like orientation of the [-C(=O)NH]⁻ group whereas the higher energy, less acidic conformer corresponds to a trans-like orientation of the [-C(=O)NH]⁻ group. These two distinct conformers were predicted for compounds with aliphatic, amide, basic, hydroxyl, and thioether side chains. The most acidic compounds studied were the amino acid amides of tyrosine, cysteine,

tryptophan, and histidine. These species (as well as the previously studied aspartic and glutamic acid amides) undergo side chain deprotonation. For tyrosine, cysteine, and tryptophan amides, G3(MP2) calculations predict that C-terminal amide deprotonation requires 10-12 kcal/mol more energy than side chain deprotonation. For all four amino acid amides with acidic side chains, the experimental GAs correlate to side chain deprotonation.

Phosphorylation is a common post-translation modification in proteins and is involved in cell signaling. The gas- and solution- phase acidities and gas-phase heats of formation have been predicted at the G3(MP2) level of theory for ten phosphorylated amino acids and their corresponding amides providing the first reliable set of these values as described in Chapter 5. The most common phosphorylated amino acids are phospho-serine (pSer), -threonine (pThr), and -tyrosine (pTyr) where phosphorylation occurs at the side chain hydroxyl group. The gas phase acidities (GAs) of these three phosphorylated amino acids and their corresponding amides were determined by using proton transfer reactions in a Fourier transform ion cyclotron mass spectrometer by the Cassidy group. Excellent agreement was found between the experimental GAs and the predicted GAs for these six molecules. The phosphate group was found to be the deprotonation site for pSer and pThr whereas deprotonation from the carboxylic acid generated the lowest energy anion for pTyr. The infrared spectra were calculated for six low energy anions of pSer, pThr, and pTyr and compared to the experimental IRMPD spectra. For deprotonated $[\text{pSer-H}]^-$ and $[\text{pThr-H}]^-$ good agreement is found between the experimental spectra and the calculated spectra for our lowest energy anion structure. For pTyr, good agreement is not found between the calculated and experimental spectra with our lowest energy anion and instead corresponds to a higher energy phosphate deprotonated structure. However, the experimental and predicted GAs are in good agreement with each other for the lowest energy structure pTyr-A of

the anion and the calculated GA for structure pTyr-C is not in good agreement with the experimental GA. This suggests that the conditions used to generate the ion [pThr-H]⁻ from which the IRMPD spectrum is obtained did not generate the lowest energy structure. In general, the phosphorylated amino acids are 13 to 27 and 1 to 7 kcal/mol more acidic than their corresponding non-phosphorylated amino acids and phosphorylated amino acid amides respectively. The energy differences in the GAs between the phosphorylated amino acid and their amides differ from what was found for the common L-amino acids and shows the importance of strong hydrogen bonding between the phosphate group and the carboxylate groups in the anions. Self-consistent reaction field (SCRf) calculations with the COSMO parameterization were used to predict the pK_a's in aqueous solution for the ten phosphorylated amino acid and amides. In general, the pK_a values for the phosphorylated amino acids and their corresponding amides are within 3 pK units. The heats of formation of the neutral compounds were calculated from the atomization energies at the G3(MP2) level and isodesmic reactions and excellent agreement was found between the two methods.

After thoroughly studying the simple amino acids, the same methods used previously were applied to study tripeptides composed of unsubstituted and substituted glycine (Gly) and alanine (Ala) residues as described in Chapter 6. The gas-phase acidities (GAs) of six tripeptides (GlyGlyGly, GlyAlaGly, AlaGlyAla, AlaAlaAla, AibAibAib, and SarSarSar) and their methyl esters were predicted at the G3(MP2) level of theory and obtained by proton transfer reactions in a Fourier transform ion cyclotron resonance mass spectrometer by the Cassidy group. The backbone substituted AibAibAib tripeptide was generated by replacing the hydrogens on the peptide backbone with methyl groups whereas the SarSarSar tripeptide, or N-methyl glycine, was generated by methylating the central and C-terminal amide nitrogens and one of the N-

terminal NH₂ hydrogens. All six acidic tripeptides have GAs in the range of 321.0-323.7 kcal/mol and deprotonation occurs at the carboxylic acid. The tripeptides are ~10 kcal/mol more acidic than the parent amino acids glycine and alanine. This is consistent with the extensive hydrogen bonding that was found in the tripeptide structures. For the methyl esters, deprotonation occurs at the peptide backbone. G3(MP2) calculations show that the most energetically favored site of deprotonation is an amide nitrogen, with the central amide being generally preferred. Nitrogen deprotonation requires 10-20 kcal/mol less energy than deprotonation at a methylene carbon. Only three of the methyl esters (GlyGlyGly-OMe, GlyAlaGly-OMe, and AlaAlaAla-OMe) deprotonate experimentally by electrospray ionization. Experimental GAs for these esters are in the range of 336.7-338.1 kcal/mol, in excellent agreement with the calculated G3(MP2) values. Experimental GAs could not be obtained for the other three methyl esters (AlaGlyAla-OMe, AibAibAib-OMe, and SarSarSar-OMe) because they did not produce sufficient deprotonated molecular ions. Trisarcosine methyl ester, SarSarSar-OMe, cannot be deprotonated at a central amide nitrogen because methyl groups are present at these sites; consequently, it has a high G3(MP2) GA value (less acidic) of 350.6 kcal/mol for deprotonation at the N-terminal nitrogen. For AlaGlyAla-OMe and AibAibAib-OMe, calculations of van der Waals and solvent accessible surfaces reveal that methyl groups are blocking the amide nitrogen sites. Therefore, conformational and steric hindrance effects are limiting the ability of these peptide methyl esters to deprotonate in the mass spectrometer.

To provide additional data for the carboxylic acid benchmarking study, the gas-phase acidities and heats of formation of ortho, meta, and para substituted diaminobenzenes and anions were calculated at the G3(MP2) level of theory as described in Chapter 7. The gas-phase heats of formation for the carboxylic acids, phenol, and three cresols are also predicted. For the

diaminobenzenes, *ortho* substitution generated the most stable neutral and anion. The trend in the gas-phase acidities of the diaminobenzenes was exactly the same as that of the cresols, (*ortho* > *meta* > *para*). Excellent agreement is found between the calculated and experimental heats of formation for the diaminobenzenes, within ~1 kcal/mol. For the cresols, *ortho* substitution generated the most stable neutral on the enthalpy scale whereas *meta* substitution generated the most stable neutral isomer on the free energy scale at 298 K. *Ortho* and *para* substitution generated the isomers with the greatest acidity in the gas and solution phases respectively. In general, for the carboxylic acids, the calculated and experimental heats of formation are all within 1 kcal/mol.