



2022

# GAS-PHASE ACID-BASE PROPERTIES AND CONFORMATIONS OF OLIGOPEPTIDES THROUGH MASS SPECTROMETRY AND COMPUTATIONAL STUDIES

Yuntao Zhang  
*University of the Pacific*

Follow this and additional works at: [https://scholarlycommons.pacific.edu/uop\\_etds](https://scholarlycommons.pacific.edu/uop_etds)

 Part of the [Analytical Chemistry Commons](#), and the [Physical Chemistry Commons](#)

---

## Recommended Citation

Zhang, Yuntao. (2022). *GAS-PHASE ACID-BASE PROPERTIES AND CONFORMATIONS OF OLIGOPEPTIDES THROUGH MASS SPECTROMETRY AND COMPUTATIONAL STUDIES*. University of the Pacific, Dissertation. [https://scholarlycommons.pacific.edu/uop\\_etds/3788](https://scholarlycommons.pacific.edu/uop_etds/3788)

This Dissertation is brought to you for free and open access by the Graduate School at Scholarly Commons. It has been accepted for inclusion in University of the Pacific Theses and Dissertations by an authorized administrator of Scholarly Commons. For more information, please contact [mgibney@pacific.edu](mailto:mgibney@pacific.edu).

GAS-PHASE ACID-BASE PROPERTIES AND CONFORMATIONS OF OLIGOPEPTIDES  
THROUGH MASS SPECTROMETRY AND COMPUTATIONAL STUDIES

By

Yuntao Zhang

A Dissertation Submitted to the

Graduate School

In Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Pharmaceutical and Chemical Sciences  
Bioanalytical and Physical Chemistry

University of the Pacific  
Stockton, California

2022

GAS-PHASE ACID-BASE PROPERTIES AND CONFORMATIONS OF OLIGOPEPTIDES  
THROUGH MASS SPECTROMETRY AND COMPUTATIONAL STUDIES

By

Yuntao Zhang

APPROVED BY:

Dissertation Advisor: Jianhua Ren, Ph.D.

Committee Member: Michael C. McCallum, Ph.D.

Committee Member: Miki S. Park, Ph.D.

Committee Member: Jerry Tsai, Ph.D.

Committee Member: Qinliang Zhao, Ph.D.

Department Co-Chairs: Jianhua Ren, Ph.D., and Jerry Tsai, Ph.D.

GAS-PHASE ACID-BASE PROPERTIES AND CONFORMATIONS OF OLIGOPEPTIDES  
THROUGH MASS SPECTROMETRY AND COMPUTATIONAL STUDIES

Copyright 2022

By

Yuntao Zhang

## DEDICATION

This dissertation is dedicated in memory of my father, Baosheng Zhang (1963-2020). He was the kindest and most considerate father for a son. I hope my achievement fulfills his dream for me. This dissertation is also dedicated to my mother, Xiumei Li, and my wife, Baixue Zhang, for their endless love and unconditional support.

## ACKNOWLEDGEMENTS

I am grateful to Dr. Jianhua Ren for guiding me throughout the five years of research and writing this dissertation. She has taught me how to be a researcher and scientist. I appreciate the opportunities she provided me to learn, explore and experiment.

I would also like to thank my committee members (Dr. Mike C. McCallum, Dr. Miki S. Park, Dr. Jerry Tsai, and Dr. Qinliang Zhao) for supporting me throughout my dissertation. Without their guidance and support, I would not have been able to reach this point.

GAS-PHASE ACID-BASE PROPERTIES AND CONFORMATIONS OF OLIGOPEPTIDES  
THROUGH MASS SPECTROMETRY AND COMPUTATIONAL STUDIES

Abstract

By Yuntao Zhang

University of the Pacific  
2022

This dissertation presents a comprehensive study of the peptides of interest to deeper understand the gas-phase acid-base properties in relation to their conformations and chirality. In the first part of the study, two pairs of alanine (A)-based isomeric peptides consisting of a basic probe, lysine (Lys) or 2,3-diaminopropionic acid (Dap), were investigated to understand the nature of the enhanced basicity when the basic probe was moved from the N-terminus to the C-terminus. In the second part of the study, alanine-based peptides containing a cysteine (C) as the acidic probe were investigated to understand the chirality effects on the gas-phase acidity by altering the chiral centers systemically.

Previous studies by mass spectrometry showed that the peptides ALys and AADap have had remarkably higher proton affinity (PA) compared to their isomeric counterparts LysA and DapAA. In this work, conformations, energetics, and molecular properties of the peptide systems have been thoroughly characterized through infrared multiple photon dissociation (IRMPD) spectroscopy and quantum chemical computations utilizing a set of molecular modeling tools. The molecular properties include charge distribution, dipole moment, torsional strain, hydrogen bonding, and non-covalent interaction. Computational studies yielded the lowest energy conformations along with their theoretical infrared (IR) spectra for each of the peptide systems. The resulting theoretical proton affinities are in excellent agreement with experiments. The results also suggest that the relative stability of the protonated peptides is the main source of the difference in the gas-phase basicity between the isomeric peptides. Structurally representative conformations for the protonated peptides were identified by matching the theoretical IR spectra to the corresponding IRMPD spectra. The band features of the IRMPD spectra

were analyzed in detail by vibrational mode decomposition. The N-probe peptide ions, LysAH<sup>+</sup> and DapAAH<sup>+</sup>, adopt diverse backbone geometries and intramolecular hydrogen bonding networks, and rely heavily on the hydrogen bonds for conformational stabilization. In contrast, the C-probe peptide ions, ALysH<sup>+</sup> and AADapH<sup>+</sup>, adopt helical conformations, and benefit from the interaction between the helix macrodipole and the charged NH<sub>3</sub><sup>+</sup> group. The low torsional strain on the Lys sidechain contributes significantly to the conformational stability for ALysH<sup>+</sup> than for LysAH<sup>+</sup>.

The chirality of each residue in CAAA and Ac-CAAA (Ac represents the acetyl group) alters from the L- to the D-form systematically to generate two series of peptides. Qualitative comparison of the gas-phase acidity was achieved through mass spectrometry measurements using the Cooks' kinetic method. The following two acidity ladders from the most acidic to the least acidic were obtained: CAA<sup>d</sup>A > CA<sup>d</sup>AA ~ CAAA > <sup>d</sup>CAAA > C<sup>d</sup>AAA, and Ac-<sup>d</sup>CAAA > Ac-CAA<sup>d</sup>A > Ac-CAAA > Ac-CA<sup>d</sup>AA > Ac-C<sup>d</sup>AAA, where the superscript-d in front of the amino acid symbol indicates the D-form of that residue. In both non-acetylated and acetylated peptides, the gas-phase acidity increases as the D-alanine moves further away from the N-terminal acidic probe cysteine. Inversion of the cysteine residue from the L- to the D-form reduces the gas-phase acidity of the non-acetylated peptide and enhances the gas-phase acidity of the acetylated one. Overall, the change in the gas-phase acidity is likely due to the conformational reorganization in the deprotonated peptides upon chiral inversion.

## TABLE OF CONTENTS

List of Tables .....	11
List of Figures .....	15
List of Abbreviations .....	19
Chapter 1: Introduction .....	20
Lysine and Cysteine in Proteins.....	20
D-Amino Acids.....	24
Ionizable Groups in the Interior of Proteins.....	27
Effects of Structure on the Acidity/Basicity of Model Oligopeptides .....	28
Goal of Study .....	29
Chapter 2: Methods and Materials .....	31
Peptide Synthesis Procedures .....	31
Deprotection.....	32
Coupling.....	32
N-Terminus Acetylation .....	33
Cleavage.....	33
Mass Spectrometry Measurements using the Cooks' Kinetic Method .....	34
Sample Introduction.....	35
Fragmentation of Proton-Bound Dimer .....	35
Standard Kinetic Method .....	36
Extended Kinetic Method .....	39
Infrared Multiple-Photon Dissociation (IRMPD) Spectroscopy.....	41
Computational Methods.....	43
Finding the Lowest Energy Conformations .....	43
Calculating the Theoretical Thermochemical Values .....	49

	9
Calculating the Dot Products and Vector Angles .....	52
Charge Analysis.....	53
Dipole Moment Decomposition.....	54
Torsional Strain Analysis.....	55
Hydrogen Bond Analysis.....	56
RDG Analysis.....	58
IR Matching and Vibrational Mode Decomposition.....	60
Bond Order and Relaxed Force Constant .....	61
<b>Chapter 3: Structural and Energetic Characterization of Peptides Containing Lysine and Homologs .....</b>	<b>63</b>
LysA and ALys.....	66
Features of the IRMPD Spectra of LysAH <sup>+</sup> and ALysH <sup>+</sup> .....	66
Computed PA and GB Values for LysA and ALys .....	68
Computed Structures and IR Spectra of LysA and ALys .....	73
Structural and Energetic Comparison .....	93
IR Decomposition.....	116
Summary.....	139
DapAA and AADap.....	140
Features of the IRMPD Spectra of DapAAH <sup>+</sup> and AADapH <sup>+</sup> .....	141
Computed PA and GB Values for DapAA and AADap .....	143
Computed Structures and IR Spectra of DapAA and AADap .....	148
Structural and Energetic Comparison .....	169
IR Decomposition.....	196
Summary.....	228
<b>Chapter 4: Cysteine-Containing Peptides in the Gas Phase.....</b>	<b>230</b>
Non-Acetylated D-Amino Acid-Containing Peptides .....	233

	10
Acidity Comparison using One Reference Acid at a Single Collision Energy .....	234
Acidity Comparison Using Different Reference Acids at Multiple Collision Energies .....	236
Acetylated D-Amino Acid-Containing Peptides.....	245
Direct Acidity Comparison with Ac-CAAA.....	246
Computed Thermochemical Values of the Acetylated Peptides.....	250
Summary.....	258
Chapter 5: Summary and Conclusions.....	260
References.....	263
Appendix A: Installation and Running of Computational Programs .....	279

## LIST OF TABLES

## Table

1.1	Summary of Deprotonation Enthalpy ( $\Delta_{\text{acidH}}$ ) of Cysteine-Containing Oligopeptides .....	29
1.2	Summary of Proton Affinity (PA) of Oligopeptides Containing Lysine and Homolog .....	29
3.1	Summary of Measured Thermochemical Values of Six Model Peptides .....	66
3.2	Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for LysA and ALys .....	71
3.3	Experimental and Calculated Thermochemical Values of LysA and ALys.....	73
3.4	C→O Vector Angle between the Lys-Ala Amide C=O Group and the Carboxyl C=O Group in C2, C4, and C6.....	84
3.5	Dihedral Angles Across the C <sub>α</sub> of Lysine and C→O Vector Angle between the Ac-Lys and Lys-Ala Amide C=O Groups in C3 and C8 .....	85
3.6	Dihedral Angles across the C <sub>α</sub> of Alanine and C→O Vector Angle between the Ac-Ala and Ala-Lys Amide C=O Groups in C1', C4', C5', C6', and C9' .....	88
3.7	Dihedral Angles across the C <sub>α</sub> of Alanine and C→O Vector Angle between the Ac-Ala and Ala-Lys Amide C=O Groups in C2', C3', C7', C8', and C10' .....	90
3.8	Relative Free Energy Values of Representative Conformations of LysAH <sup>+</sup> and ALysH <sup>+</sup> .....	93
3.9	Relative Free Energy and the Three Types of Charges on the NH <sub>3</sub> <sup>+</sup> Group in the Representative Conformations of LysAH <sup>+</sup> and ALysH <sup>+</sup> .....	94
3.10	Atomic Charges on the Nitrogen and Hydrogens of the NH <sub>3</sub> <sup>+</sup> Group in the Representative Conformations of LysAH <sup>+</sup> and ALysH <sup>+</sup> .....	96
3.11	Dipole Moments of the Representative Conformations of LysAH <sup>+</sup> and ALysH <sup>+</sup> .....	98
3.12	Dihedral Angles on the Lysine Sidechain in the LysAH <sup>+</sup> and ALysH <sup>+</sup> Conformations.....	101
3.13	Torsional Strain on the Lysine Sidechain in the LysAH <sup>+</sup> and ALysH <sup>+</sup> .....	101
3.14	Geometry Information of Hydrogen Bonds with the NH <sub>3</sub> <sup>+</sup> Group for the Representative Conformations of LysAH <sup>+</sup> and ALysH <sup>+</sup> .....	104

	12
3.15 Energetic Information of Hydrogen Bonds with the $\text{NH}_3^+$ Group for the Representative Conformations of $\text{LysAH}^+$ and $\text{ALysH}^+$ .....	104
3.16 Geometric Information and Electron Density for the Interaction between the Lysine N-H Group and the Carboxyl C=O Group in the $\text{ALysH}^+$ Conformations .....	114
3.17 Summary of the Energy Characterizations of the $\text{LysAH}^+$ and $\text{ALysH}^+$ Conformation .....	116
3.18 IRMPD and Assignments and Mode Descriptions with Frequencies Calculated for C2 of $\text{LysAH}^+$ .....	119
3.19 IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C3 of $\text{LysAH}^+$ .....	121
3.20 IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C1' of $\text{ALysH}^+$ .....	123
3.21 IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C6' of $\text{ALysH}^+$ .....	125
3.22 Summary of Amide bond, Carboxyl Group, and Corresponding Hydrogen Bond Information in the $\text{LysAH}^+$ and $\text{ALysH}^+$ Conformations .....	128
3.23 Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for DapAA and AADap .....	145
3.24 Experimental and Calculated Thermochemical Values of DapAA and AADap .....	147
3.25 Dihedral Angles and C→O Vector Angles in C1', C8', and C10' of $\text{AADapH}^+$ .....	163
3.26 Dihedral Angles and C→O Vector Angles in C6' of $\text{AADapH}^+$ .....	165
3.27 Relative Free Energy Values of the Representative Conformations of $\text{DapAAH}^+$ and $\text{AADapH}^+$ .....	170
3.28 Four Types of Charges on the $\text{NH}_3^+$ Group in the Representative Conformations of $\text{DapAAH}^+$ and $\text{AADapH}^+$ .....	171
3.29 O --- H Distance in the Hydrogen Bonds Involving the $\text{NH}_3^+$ in the Representative Conformations of $\text{DapAAH}^+$ and $\text{AADapH}^+$ .....	173
3.30 Atomic Charges on the Nitrogen and Hydrogens of the $\text{NH}_3^+$ Group in the Representative Conformations of $\text{DapAAH}^+$ and $\text{AADapH}^+$ .....	174
3.31 Dipole Moments of the $\text{DapAAH}^+$ and $\text{AADapH}^+$ Representative Conformations.....	175
3.32 Dihedral Angles on the Dap Sidechain in the $\text{DapAAH}^+$ and $\text{AADapH}^+$ Conformations .....	178

	13
3.33 The Torsional Strain on the Dap Sidechain in the DapAAH <sup>+</sup> and AADapH <sup>+</sup> Conformations .....	179
3.34 Geometry Information of the Hydrogen Bonds in the DapAAH <sup>+</sup> and AADapH <sup>+</sup> Conformations .....	180
3.35 Energetic Information of Hydrogen Bonds in the DapAAH <sup>+</sup> and AADapH <sup>+</sup> Conformations .....	182
3.36 Energetic Characterization of the DapAAH <sup>+</sup> and AADapH <sup>+</sup> Representative Conformations .....	196
3.37 IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C1 of DapAAH <sup>+</sup> .....	198
3.38 IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C2 of DapAAH <sup>+</sup> .....	200
3.39 IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C9 of DapAAH <sup>+</sup> .....	203
3.40 IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C1' of AADapH <sup>+</sup> .....	206
3.41 IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C10' of AADapH <sup>+</sup> .....	209
3.42 Summary of Amide Bond, Carboxyl Group, and Corresponding Hydrogen Bond Information in the DapAAH <sup>+</sup> and AADapH <sup>+</sup> Conformations .....	212
3.43 Angles with the Vertex of Dap C <sub>α</sub> in the DapAAH <sup>+</sup> Conformations .....	216
4.1 Reference Acids Used for Acidity Measurements of Non-Acetylated D-Amino Acid-Containing Peptides .....	234
4.2 Branching Ratios of CAAA with Six Selected Reference Acids .....	237
4.3 Branching Ratios of <sup>d</sup> CAAA with Six Selected Reference Acids.....	237
4.4 Branching Ratios of C <sup>d</sup> AAA with Six Selected Reference Acids.....	237
4.5 Branching Ratios of CA <sup>d</sup> AA with Six Selected Reference Acids.....	238
4.6 Branching Ratios of CAA <sup>d</sup> A with Six Selected Reference Acids.....	238
4.7 Branching Ratios of D-Amino Acid-Containing Peptides with Ac-CAAA as the Reference .....	249
4.8 Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for Ac-CAAA and Ac-CAAA <sup>-</sup> .....	252

4.9	Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for Ac- <sup>d</sup> CAAA and Ac- <sup>d</sup> CAAA <sup>-</sup> .....	253
4.10	Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for Ac-C <sup>d</sup> AAA and Ac-C <sup>d</sup> AAA <sup>-</sup> .....	254
4.11	Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for Ac-CA <sup>d</sup> AA and Ac-CA <sup>d</sup> AA <sup>-</sup> .....	255
4.12	Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for Ac-CAA <sup>d</sup> A and Ac-CAA <sup>d</sup> A <sup>-</sup> .....	256
4.13	Summary of Theoretical Energetics and Thermochemical Values for the Acetylated Peptides .....	256

## LIST OF FIGURES

## Figure

1.1	L- (left) and D- (right) configurations of amino acid.....	24
2.1	Wang resin (left) and rink amide resin (right) for peptide synthesis.....	32
2.2	Extended kinetic method (EKM) 1 <sup>st</sup> plot (left) and 2 <sup>nd</sup> plot (right).....	41
2.3	Illustration of the metadynamics method described as “filling the energy wells with computational sand”.....	45
2.4	Schematic representation of the iMTD-GC workflow used for conformational search in CREST.....	48
2.5	Torsional strain function for n-butane.....	56
2.6	Regression plot of hydrogen bond energy (HBE) versus $\rho$ at the BCP of the hydrogen bond for a) the neutral complexes, and b) the charged complexes. ....	58
2.7	Sample RDG scatter plot (left) and the corresponding isosurface diagram (right).....	60
3.1	IRMPD Spectra of LysAH <sup>+</sup> (top) and ALysH <sup>+</sup> (bottom) with major peaks labeled in wavenumber. ....	68
3.2	Computed geometries of LysA at the wB97XD/6-311G** level of theory.....	74
3.3	Computed geometries of ALys at the wB97XD/6-311G** level of theory.....	75
3.4	Overlay of the ten conformations of LysA (left) and ALys (right).....	76
3.5	Computed geometries of LysAH <sup>+</sup> at the wB97XD/6-311G** level of theory.....	78
3.6	Computed IR stick and convoluted spectra (solid light blue) overlaid onto the IRMPD spectrum (dashed black) of LysAH <sup>+</sup> . ....	79
3.7	Computed geometries of ALysH <sup>+</sup> at the wB97XD/6-311G** level of theory.....	80
3.8	Computed IR stick and convoluted spectra (solid dark blue) overlaid onto experimental IRMPD spectrum (dashed black) of ALysH <sup>+</sup> .....	81
3.9	Overlay of the ten conformations of LysAH <sup>+</sup> (left) and ALysH <sup>+</sup> (right). ....	82
3.10	Structure of C1 and C9 (left), C5 and C7 (middle), and C10 (right). ....	83
3.11	Structure of C2 and C4 (left), and C6 (right).....	84
3.12	Structure of C3 and C8.....	84

3.13	Ramachandran plot of C3 (light blue circle) and C8 (green square).....	86
3.14	Boltzmann weighted average spectra (top) and hand-weighted spectra (bottom) overlaid onto the IRMPD spectrum of LysAH <sup>+</sup> .....	87
3.15	Structure of C1' and C6' (left), C4' and C5' (middle), and C9' (right).....	88
3.16	Ramachandran plot of the left-handed helical conformations (dark blue circles, C1', C4', C5', C6') and the right-handed helical conformation (green square, C9').	89
3.17	Structure of C2', C3', C7', C8', and C10'.....	89
3.18	Ramachandran plot of conformation C2', C3', C7', C8' and C10' (red diamond).	91
3.19	Boltzmann weighted average spectra (top) and hand-weighted spectra (bottom) overlaid onto the IRMPD spectrum of ALysH <sup>+</sup> .....	92
3.20	Three types of charges on the NH <sub>3</sub> <sup>+</sup> group in the representative conformations of LysAH <sup>+</sup> and ALysH <sup>+</sup> .....	95
3.21	Dipole moments of the whole molecule (blue), the backbone (green), and the NH <sub>3</sub> <sup>+</sup> group (red) for the representative conformations of LysAH <sup>+</sup> (top) and ALysH <sup>+</sup> (bottom).....	98
3.22	RDG scatter plot (left) and isosurface diagram (right) for C2 (top) and C3 (bottom) of LysAH <sup>+</sup> .....	107
3.23	RDG scatter plot (left) and isosurface diagram (right) for C1' (top) and C6' (bottom) of ALysH <sup>+</sup> .....	110
3.24	Selected conformations of LysAH <sup>+</sup> (top) and ALysH <sup>+</sup> (bottom) with atoms labeled.....	118
3.25	IRMPD spectra of DapAAH <sup>+</sup> (top) and AADapH <sup>+</sup> (bottom) with major peaks labeled in wavenumber.....	143
3.26	Computed geometries of DapAA at the wB97XD/6-311G** level of theory.....	149
3.27	Computed geometries of AADap at the wB97XD/6-311G** level of theory.....	150
3.28	Overlay of the ten conformations of DapAA (left) and AADap (right).....	151
3.29	Computed geometries of DapAAH <sup>+</sup> at the wB97XD/6-311G** level of theory.....	153
3.30	Computed IR stick and convoluted spectra (solid light blue) overlaid onto experimental IRMPD spectrum (dashed black) of DapAAH <sup>+</sup> .....	154
3.31	Computed geometries of AADapH <sup>+</sup> at the wB97XD/6-311G** level of theory.....	155
3.32	Computed IR stick and convoluted spectra (solid blue) overlaid onto experimental IRMPD spectrum (dashed black) of AADapH <sup>+</sup> .....	156

	17
3.33 Overlay of the ten conformations of DapAAH <sup>+</sup> (left) and AADapH <sup>+</sup> (right). .....	157
3.34 Overlay of C1 and C2 (top left), C5 and C7 (top right), C4 (bottom left), and C9 (bottom right). .....	158
3.35 Structure of C3 (left) and C8 (right). .....	160
3.36 Structure of C6 and C10. ....	160
3.37 Boltzmann weighted average spectra (top) and hand-weighted spectra (bottom) overlaid onto the IRMPD spectrum of DapAAH <sup>+</sup> . .....	161
3.38 Structure of C1' and C8' (left), and C10'(right). ....	162
3.39 Ramachandran plot of right-handed helical conformations, C1' and C8' (dark blue circles) and C10' (green squares). ....	163
3.40 Structure of C6' .....	164
3.41 Ramachandran plot of the left-handed helical conformation C6' (green squares).....	165
3.42 Structure of C2' (top left), C4' (top right), C5' and C9' (bottom left), and C3' (bottom right). .....	166
3.43 Structure of C7' .....	168
3.44 Boltzmann weighted average spectra (top) and hand-weighted spectra (bottom) overlaid onto experimental IRMPD spectrum of AADapH <sup>+</sup> . .....	169
3.45 Different types of charges on the NH <sub>3</sub> <sup>+</sup> group in the representative conformations of DapAAH <sup>+</sup> and AADapH <sup>+</sup> . ....	171
3.46 Dipole moments of whole molecule (blue), the backbone (green) and the NH <sub>3</sub> <sup>+</sup> group (red) for representative conformations of DapAAH <sup>+</sup> (top) and AADapH <sup>+</sup> (bottom). ....	176
3.47 RDG scatter plot (left) and isosurface diagram (right) of C1 (top), C2(middle) and C9 (bottom) of DapAAH <sup>+</sup> . ....	185
3.48 RDG scatter plot and isosurface of C1' (top) and C10' (bottom) of AADapH <sup>+</sup> . ....	189
3.49 Selected conformations of DapAAH <sup>+</sup> (top) and AADapH <sup>+</sup> (bottom) with atoms labeled. ....	197
4.1 Structure of E. coli thioredoxin. ....	231
4.2 Product ion spectra resulting from fragmentation of the proton-bound dimer between the non-acetylated peptides and TFMBA at 1.5 eV (E <sub>cm</sub> ) under CID pressure of 0.2 mTorr. (a), [TFMBA <sup>-</sup> ···H <sup>+</sup> ···CAAA <sup>-</sup> ]; (b), [TFMBA <sup>-</sup> ···H <sup>+</sup> ···CAAdA <sup>-</sup> ]; (c), [TFMBA <sup>-</sup> ···H <sup>+</sup> ···CAdAA <sup>-</sup> ]; (d), [TFMBA <sup>-</sup> ···H <sup>+</sup> ···C <sup>d</sup> AAA <sup>-</sup> ]; (e), [TFMBA <sup>-</sup> ···H <sup>+</sup> ··· <sup>d</sup> CAAA <sup>-</sup> ]. ....	236

4.3	Branching ratio plots for non-acetylated D-amino acid-containing peptides using six reference acids. ....	239
4.4	Branching ratio plots of the peptides with CAAA using different reference acids.....	243
4.5	Overlay of branching ratio plots of the peptides with CAAA.....	244
4.6	Fragmentation spectra of the proton-bound dimer between the D-amino acid-containing peptides and the all L-amino acid reference peptide. a) Ac- <sup>d</sup> CAAA and <sup>D3</sup> Ac-CAAA. b) <sup>D3</sup> Ac-C <sup>d</sup> AAA and Ac-CAAA. c) <sup>D3</sup> Ac-CA <sup>d</sup> AA and Ac-CAAA. d) <sup>D3</sup> Ac-CAA <sup>d</sup> A and Ac-CAAA.....	247
4.7	Branching ratio plots of D-amino acid-containing peptides with Ac-CAAA as the reference.....	249
4.8	Semi-quantitative comparison of the $\Delta_{\text{acid}}H$ values of the acetylated peptides. ....	250

## LIST OF ABBREVIATIONS

Ac	Acetyl group
Dap	2, 3-Diaminopropionic Acid
DCM	Dichloromethane
DFT	Density Functional Theory
DIPEA	N, N-Diisopropylethylamine
DMF	N, N-Dimethylformamide
ESI	Electrospray Ionization
FELIX	Free Electron Laser for InfraRed eXperiments
Fmoc	9-fluorenylmethyloxycarbonyl protecting group
FT-ICR	Fourier-transform ion-cyclotron resonance
GB	Gas-Phase Basicity
HBTU	N, N, N', N'-Tetramethyl-O-(1H-benzotriazol-1-yl) uranium hexafluorophosphate
IR	Infrared
MS	Mass Spectrometry
PA	Proton Affinity
Q1	Quadrupole 1
q2	Quadrupole 2 or Collision Cell
Q3	Quadrupole 3
TFA	Trifluoroacetic acid
TIPS	Triisopropylsilane
UV	Ultraviolet
RDG	Reduced Density Gradient

## CHAPTER 1 : INTRODUCTION

Throughout nature, the complex biochemical processes are nearly all catalyzed by proteins and enzymes. Enzymes accelerate chemical reactions and compress their vast timescale onto timescales needed for life. Because life occupies almost every available niche on earth, enzymes have evolved to function in diverse environments (Arcus, van der Kamp, Pudney, & Mulholland, 2020). Despite the different functions and working conditions, they share the same biochemical identity as natural polymer comprised of amino acids. By now, billions of known protein sequences with varying length and amino acid compositions have been identified (Mitchell et al., 2020; Steinegger, Mirdita, & Söding, 2019) and there have been hundreds of amino acids found in nature (Lopez & Mohiuddin, 2020). Among the identified amino acids, only twenty of them are needed to make nearly all proteins in humans and most forms of life on the planet. The unique sequences of proteins grant them diverse geometries and functions. Though proteins are macromolecules and can be as heavy as three million Daltons (Kurzban & Wang, 1988; Maruyama, Kimura, Ohashi, & Kuwano, 1981), their biochemical functions are primarily carried out in a few small regions in their structure, called active sites. In the active sites, the unique spatially arrangements and adjusted chemical and physical properties of the involving amino acids create a three-dimensional space for the biochemical reactions to occur.

### **Lysine and Cysteine in Proteins**

Lysine (symbol Lys or K) is one of the three most common basic amino acids in nature with distinct biochemical and biophysical properties. It plays an important role in many cellular processes, such as catalysis (Bartlett, Porter, Borkakoti, & Thornton, 2002) and protein structure stabilization (Sokalingam, Raghunathan, Soundrarajan, & Lee, 2012). Lys can be found in catalytic sites of enzymes (Gulick, Starai, Horswill, Homick, & Escalante-Semerena, 2003; Sekimoto, Matsuyama, Fukui, & Tanizawa, 1993; Susan L. Schafer et al., 1996) and more often on the surface of proteins with exposure to bulk solvents (Luo, 2018).

Lys has a nucleophilic amino group and a 4-carbon (C4) hydrocarbon sidechain, which are the two key factors for its numerous and unique functions in the biological system. Its  $\epsilon$ -amino group is highly basic and is protonated to a degree of > 99.9% at physiological pH of 7.4. Its C4 hydrocarbon chain is hydrophobic and flexible. It enables Lys to interact in various modes, including ion-pairing salt bridge, hydrogen bonds, cation- $\pi$  interaction, and entropy-driven desolvation of its C4 hydrocarbon sidechain (Luo, 2018). In the work of Gail J. Bartlett, et al. (Bartlett et al., 2002), the involvement of amino acids in catalysis was analyzed for 178 enzyme active sites. Lys was found to play important roles in enzymatic catalysis. It constitutes 9% of the catalytic residues, while its natural abundance in the enzyme dataset is only 5.8%.

Lys is commonly involved in salt bridges, which are important for stabilizing the three-dimensional protein structure (Rothschild & Mancinelli, 2001) and the specificity of the interaction of proteins with other biomolecules (Bosshard, Marti, & Jelesarov, 2004). Post-translational modifications of Lys alter its charged state and biochemical properties. These modifications, such as acetylation, phosphorylation, methylation, sumoylation, succinylation, and ubiquitination, play critical roles in a range of cellular processes (Liu et al., 2014).

Histones are highly basic proteins abundant in Lys (Weng, Zhou, & Greenberg, 2014). The electrostatic attraction between the positively charged Lys in histone and the negatively charged phosphate group in DNA allows the DNA to wind around the histone spool to form a nucleosome. Modifications on the histone lysines are important for epigenetic regulations. Acetylation and phosphorylation are reversible and are often related to the inducible regulation of individual genes. Conversely, methylation is irreversible and seems to be involved in the cellular memory of the transcriptional status by fixing the chromatin organization in a heritable manner (Völkel & Angrand, 2007). The nucleophilic amino group of Lys plays a central role in its methylation reaction using histone lysine methyltransferase (Al Temimi et al., 2019). Its C4 hydrocarbon chain is proved to be the optimal chain length for the lysine methylation process (Temimi et al., 2017).

Ubiquitin is a small, 76-amino acid regulatory protein containing seven Lys residues (Goldstein et al., 1975). Its covalent conjugation to substrate proteins (often in the form of polymer), ubiquitination, governs a variety of biological processes, including cell division, signal transduction, protein trafficking, and quality control (Mukhopadhyay & Riezman, 2007). Secondary ubiquitin molecules are primarily ligated to the  $\epsilon$ -amino group in one of seven Lys residues in the previous ubiquitin molecule. Relying on the existence of a large number of ubiquitinating enzymes, distinct ubiquitin polymers are constructed to encode different signals (W. Li & Ye, 2008). For example, K48-linked ubiquitin chains with a minimum of four subunits act as a proteasome-targeting signal, and K6-linked chains correspond to UV-induced DNA damage (Shekhawat, Pham, Prabakaran, & Strieter, 2014). Specific linkages are achieved by precise spatial orientation of the ubiquitin molecules so that a specific Lys residue is in close proximity to the donor ubiquitin, and its nucleophilic  $\epsilon$ -amino group attacks the donor thiol-ester linkage.

The compound 2, 3-diaminopropionic acid (Dap) is a Lys analog nonproteinogenic amino acid with a C1 hydrocarbon chain instead of a C4 hydrocarbon chain. It has been recognized for its natural occurrence for decades and is produced and utilized by plants (Bell, Tirimanna, EA, & AS, 1965) and bacteria (Thomas, Mitchell, & Blanksby, 2005). Dap is a building block of polypeptide antibiotics, edeine (Rao, 1978), viomycin (Carter et al., 2002), and zwittermicin A (Zhao, Song, Luo, Yu, & Sun, 2008). Dap serves as the precursor and is an important iron-liganding component of siderophores (Beasley, Cheung, & Heinrichs, 2011), which promote iron-restricted growth. It was also found to generate homopoly amino acids by microorganisms, which are rarely observed in nature (Xu et al., 2015). More interestingly, N-terminal Dap-containing peptides are designed to serve as efficient methylglyoxal scavengers to inhibit advanced glycation end product formation (Sasaki et al., 2009).

Cysteine (symbol Cys or C) is an important ionizable amino acid often found in functional sites of proteins, although it is one of the least abundant amino acids (often the least) in organisms (Marino & Gladyshev, 2012). It is involved in many biological processes and cellular functions through its thiol-disulfide exchange pathways (Roos, Foloppe, & Messens, 2013). In the work of Christoph Wiedemann, et al. (Wiedemann, Kumar, Lang, & Ohlenschläger, 2020), the frequency of Cys in proteins was

compared between archaea, bacteria, and eukaryote. It was found that the frequency and the median number of Cys seem to increase during evolution. In the work of Gail J. Bartlett, et al. (Bartlett et al., 2002), Cys constitutes 5.6% of catalytic residues, while its natural abundance is only 1.2%. The catalytic propensity of Cys, the ratio between the residue's frequency among all catalytic residues and its frequency in the overall protein, is 4.6, the second highest among all amino acids. In acid-base catalysis, the easier it is to deprotonate and re-protonate, the faster it will be able to perform its catalytic function. Cys in enzymes is activated through deprotonation as thiolate, which is much more reactive than a neutral thiol as a nucleophile. The sidechain pKa value of free Cys is 8.6, which is the second closest value to the biological pH among all twenty common amino acids. It allows easy deprotonation and re-protonation and contributes significantly to its high catalytic propensity.

Cys is often found buried inside the proteins. When placed on the protein surface, Cys is poorly conserved. When clustered with another Cys residue inside the protein, they behave as polar residues, and appear to be highly conserved (Marino & Gladyshev, 2010). It is well-known that when amino acids are buried in the protein cavity, their sidechain pKa value can be altered from standard solution values (Bartlett et al., 2002), allowing them to catalyze reaction at different conditions. Cys is particularly responsive to the local environment, and its pKa values could be shifted by factors like hydrogen bonds, charge-charge interactions, and helix macrodipoles (Roos et al., 2013).

As a great example to demonstrate the vital role of Cys for life, thioredoxin, is critical for redox regulation of protein function and signaling via thiol redox control (Arnér & Holmgren, 2000). In the oxidizing extracellular environment, proteins are rich in disulfides, while inside the cell, proteins contain many free sulfhydryl groups (Gilbert, 1990). The primary disulfide reducing enzyme for maintaining the cellular proteins in their reduced state is thioredoxin (Arnér & Holmgren, 2000). Thioredoxin is the primary enzyme that maintains a low redox potential and reduces protein disulfide bonds in the cell (Arnér & Holmgren, 2000). It appears to be truly ubiquitous in all living cells, from Archeobacteria to humans (Holmgren, 2003). Two Cys residues are critical for the enzyme function. In human thioredoxin, Cys32 is placed on the hydrophobic surface with a reduced pKa value of 6.3, while Cys35 is buried inside

the protein with an elevated pKa value. After binding to the substrate, Cys32 thiolate in the hydrophobic environment between the substrate and thioredoxin surfaces initiates the process through its nucleophilic attack to the target protein. Attack of the internal Cys35 thiolate to the disulfide-bound Cys32 marks the end of this process and releases the reduced target protein. In both steps, the deprotonated cysteines as thiolates perform nucleophilic attacks and are critical for the enzyme to function properly.

### D-Amino Acids

D-amino acids (DAAs) are mirror images of their corresponding L- amino acids (Figure 1.1). Single DAA has the same physical and chemical properties as its L isomer. However, in nature, L-amino acids are almost exclusively preferred in biological systems (Hein & Blackmond, 2012). This phenomenon is very puzzling. “What is the origin of homochirality in nature” has been collected as one of the one hundred and twenty-five fundamental questions that we are still unable to answer (Kennedy & Norman, 2005).

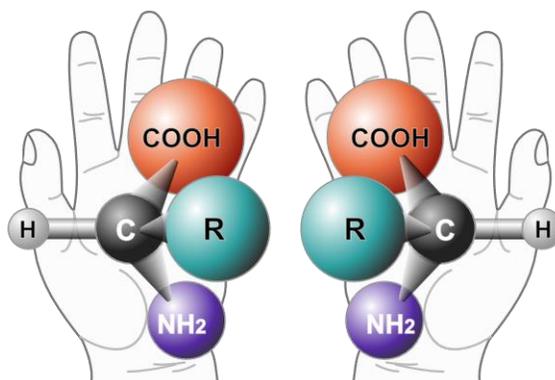


Figure 1.1. L- (left) and D- (right) configurations of amino acid.

Though L-amino acids are generally the default for peptide and protein synthesis, DAAs play critical roles in the biological system (Genchi, 2017). They are found across the whole spectrum of living organisms, from single-cell bacteria to humans (Ollivaux, Soyez, & Toullec, 2014). Bacteria account for a thousand-fold more biomass than humans, and they rely on DAAs to make a strong and elastic polymer called peptidoglycan, the major component of the bacteria cell wall. Peptidoglycan is composed of

polysaccharide chains cross-linked by D-amino acid-containing oligopeptides (Vollmer, Blanot, & De Pedro, 2008). The DAAs heavily influence the amount and strength of the peptidoglycan via their incorporation into the polymer and by regulating enzymes that synthesize and modify it (Cava et al., 2011). DAAs also play important regulatory roles in the bacteria kingdom. D-methionine (D-Met) and D-leucine (D-Leu) are found to regulate cell wall remodeling and biofilm disassembly in aging bacterial communities (Cava et al., 2011). Improved understanding of DAAs is critical to the development of new antibiotics that disrupt cell wall biogenesis.

DAAs are also commonly found in the multicellular organism. For cone snails, the composite venom they use to envenomate their prey contains a wide array of neuroactive oligopeptides. A group of peptides called contryphan serve as the active ingredients of their potent venom, and most of them are found to incorporate D-tryptophan/D-Leucine into their sequences (Massilia, Schininà, Ascenzi, & Polticelli, 2001). Dermorphin is the first biologically active D-amino acid-containing peptide (DAACP) found in amphibians and was discovered in the methanol extract from the skin of South American tree frogs belonging to the genus *Phyllomedusinae* (Melchiorri & Negri, 1996). It contains only five amino acids but is one of the strongest mammalian endogenous analgesic opioids. Dermorphin is a highly potent and nearly specific  $\mu$ -opioid peptide (Mor et al., 1989), more than 30 times more potent than morphine, and has been illegally used in horse racing (“Horse Racing Discovers New Drug Problem, One Linked to Frogs - The New York Times,” n.d.). Afterwards, many other short DAACPs have been isolated from the skin extracts of frogs as naturally occurring peptides with high affinity and selectivity for opioid receptors (Erspamer et al., 1989; Fiori et al., 1997; Kreil et al., 1989; Melchiorri & Negri, 1996; Mor et al., 1989). Defensin-like peptide-2 (DLP-2) is the first biologically active DAACP found in mammals, discovered in the venom from Australia platypus (Torres et al., 2005). It is a 42-amino acid residue polypeptide that contains a D-methionine at position two. Defensin-like peptide-4 (DLP-4) has the identical amino acid sequence as DLP-2, but its amino acids are all in their L-forms. Interestingly, though incorporating the DAA in DLP-2 has minimal effect on the overall fold of the peptide in solution, these two peptides elute at different retention times in reverse-phase HPLC.

The discovery of DAAs in humans can be traced back to five decades ago. Aspartic acid racemization in human tooth enamel had been reported in 1975, and D-aspartic acid was found to accumulate with age and suggested the role of racemization in the aging process (Helfman & Bada, 1975). Afterwards, DAAs have been discovered in various human tissues, including eye lenses (Verzijl et al., 2000), bone (Ohtani, Yamamoto, Matsushima, & Kobayashi, 1998), aorta (Powell et al., 1992), erythrocytes (McFadden & Clarke, 1982), lung (Shapiro, Endicott, Province, Pierce, & Campbell, 1991), ligament (Ritz-Timme, Laumeier, & Collins, 2003), and even brain (Roher et al., 1993; Shapira & Chou, 1987). Recent discoveries have revealed that DAAs are physiologically active and play critical roles in many biological processes, and alteration in the concentration of DAAs might relate to the pathogenesis of some disorders (Fuchs, Berger, Klomp, & De Koning, 2005). For example, D-serine is the most dominant endogenous coagonist for N-methyl-D-aspartate (NMDA) receptor, which is involved in learning and memory in mammals (Wolosker, 2007). Alteration of the NMDA receptor functions is associated with series of disorders such as schizophrenia, ischemia, epilepsy, and neurodegenerative diseases. With analytical instruments of higher sensitivity, all amino acids except glutamic acid were found to present in both L- and D- forms in the mouse brain, and all measured D-amino acid levels in the brain tissue are typically ~10 to 2000 times higher than those in blood levels (Weatherly et al., 2017). Additionally, D-aspartate is important for the dendritic development and survival of newborn neurons and function as a modulator of adult neurogenesis (Kim et al., 2010). As DAAs have biological functions, alternations of the concentrations might relate to the pathogenesis of some disorders, which provide potential diagnostic and therapeutic strategies. Chiral inversion of amyloid-beta has been believed to be involved in pathological development but has been largely ignored because of partially its age dependence and also low abundance (G. Li, DeLaney, & Li, 2019; Roher et al., 1993). Astonishing observations of DAAs in humans have fueled the study on D-amino acid-containing peptides and proteins. Despite that the specific functions of individual DAAs are far from being unraveled, characterizations on how they behave at the fundamental molecular level are important to fully understand DAAs' functionality.

### **Ionizable Groups in the Interior of Proteins**

Charges are inherently more compatible with bulk solvent than the less polar interior of proteins (Isom, Castañeda, Cannon, Velu, & García-Moreno E, 2010). An ionizable group in the less polar interior of a protein is energetically unfavorable and is relatively rare (Isom et al., 2011). Paradoxically, internal ionizable groups are essential for many fundamental biochemical processes, such as energy transduction (Ballmoos, Wiedenmann, & Dimroth, 2009; Rastogi & Girvin, 1999), catalysis (Bartlett et al., 2002; Harris & Turner, 2002; Ho, Ménétret, Tsuruta, & Allen, 2009), and macromolecular assembly (Malevanets et al., 2017).

When placed in different microenvironments, the pKa values and charged states of the amino acids adjust accordingly (Bartlett et al., 2002; Harris & Turner, 2002; Isom et al., 2011). Opposite charges stabilize each other. Placing a basic residue close to an acidic residue tends to raise the pKa of the basic residue and lower the pKa of the acidic one. In this way, the basicity of the basic residue and the acidity of the acidic residue are both enhanced (Gutteridge & Thornton, 2005). Conversely, when placed in a nonpolar environment or near a residue carrying the same charge, the neutral form of the polar residues is preferred. With pKa values adjusted for the surrounding environment, the amino acids become more effective acids and bases (Gutteridge & Thornton, 2005). With easier conversion between their neutral and charged forms in the local environment, they will be able to perform the function faster at various pH values far from their intrinsic pKa values.

In the work of Gail J. Bartlett, et al. (Bartlett et al., 2002), it was found that 92% of the catalytic residues are provided by ionizable amino acids, and over 90% of the catalytic residues are located in clefts, where the effective dielectric response is lowered, and the solvent could be excluded upon binding. The ionizable lysine and cysteine residues in environments with limited solvent access behave differently from their acid-base properties in solvated conditions. Their effective pKa values in the active sites are important for their functionality in the protein active sites and should be characterized carefully. Studies of the acid–base properties of peptides in a solvent-free environment are of great importance.

### **Effects of Structure on the Acidity/Basicity of Model Oligopeptides**

The factors affecting the buried ionizable residues are very difficult to establish. The perturbed pKa values heavily depend on the local microenvironment (Batoon & Ren, 2015). When attempted to measure the pKa values of buried residues by titration of model peptides, the results essentially are from solvated residues. To solve the problem, gas-phase environments which are characteristically solvent-free have been used to resemble the environments for amino acids buried in protein active sites. Studies have shown that the solution-phase conformations of proteins could be retained in the gas phase under mass spectrometry conditions (Katta, Chait, & Carr, 1991; Marcoux & Robinson, 2013; Ruotolo & Robinson, 2006).

In the previous studies of our group, it has been found that the gas-phase acidity/basicity of the peptides is affected by factors including the position of the acidic/basic residue, the peptide length, the sidechain length, and the chirality of the amino acids (Batoon & Ren, 2015, 2016; Ren, Tan, & Harper, 2009; Y. Zhang, Buen, Browne, Mann, & Ren, 2021). Peptides with an N-terminal cysteine are more acidic than their isomeric peptides with a C-terminal cysteine (Ren et al., 2009). Contrarily, peptides with a C-terminal basic residue are more basic than their isomeric peptides with an N-terminal basic residue (Batoon & Ren, 2016). Elongation of the peptide backbone and the sidechain of the probe residue tends to enhance the acidity/basicity of the peptides (Batoon, 2016; Batoon & Ren, 2016; Ren et al., 2009). The gas-phase acidity of a cysteine-containing tripeptide was found to be affected by chiral inversion (Y. Zhang et al., 2021). The measured deprotonation enthalpy of acidity peptides and the proton affinity of basic peptides are summarized in Table 1.1 and 1.2, respectively.

Table 1.1

*Summary of Deprotonation Enthalpy ( $\Delta_{acid}H$ ) of Cysteine-Containing Oligopeptides*

$\Delta_{acid}H$ (kcal/mol)			
<b>CAAAA</b>	319.2 ± 4.0	<b>AAAAC</b>	325.9 ± 2.0
<b>CAAA</b>	319.3 ± 3.0	<b>AAAC</b>	332.2 ± 2.0
<b>CAA</b>	330.5 ± 2.0	<b>AAC</b>	334.6 ± 1.8
<b>CA</b>	331.2 ± 1.8	<b>AC</b>	335.6 ± 1.7
<b>ACA</b>	328.0 ± 2.2	<b>A<sup>d</sup>CA</b>	327.8 ± 2.2

Values obtained from references Ren et al., 2009; Shen & Ren, 2012; Tan & Ren, 2007; Y. Zhang et al., 2021.

Table 1.2

*Summary of Proton Affinity (PA) of Oligopeptides Containing Lysine and Homolog*

PA (kcal/mol)			
<b>LysA</b>	241.5 ± 2.3	<b>ALys</b>	246.3 ± 2.2
<b>DapA</b>	234.5 ± 2.9	<b>ADap</b>	237.0 ± 2.9
<b>LysAA</b>	241.8 ± 2.2	<b>AA<sup>Lys</sup></b>	247.5 ± 2.4
<b>DapAA</b>	238.5 ± 2.2	<b>AA<sup>Dap</sup></b>	240.5 ± 2.2

Values obtained from references Batoon, 2016; Batoon & Ren, 2016.

The molecular modeling has suggested that helical conformations are achievable for peptides with N-terminal Cys or C-terminal Lys/Dap. However, other experimental evidence is not available except for the determined gas-phase acidity and basicity values, and a quantitative evaluation of the contribution of macrodipole interaction to structural stabilization is yet to be investigated. In the study of the effect of chiral inversion on the thermochemical properties of peptides, the observed alternation is subtle between the studied L-/D- Cys peptides. In the context of this dissertation, these problems are addressed.

### Goal of Study

The research presented in this thesis focus on the structural and energetic characterization of acidic and basic oligopeptides. The peptides studied are a series of di-, tri-, and tetra-peptides with a polyalanine backbone and a single acidic or basic probe amino acid at varying positions. Direct gas-phase acidity comparisons among peptides were carried out using mass spectrometry measurements by applying the Cooks' kinetic method. The gas-phase peptide ion fine structures were characterized via infrared multiple photon dissociation (IRMPD) ion spectroscopy by employing the free-electron laser for

infrared experiments (FELIX). Advanced computational modeling was performed to correlate experimental observations and to gain deeper insight into the nature of the structural and energetic features of the peptides.

## CHAPTER 2 : METHODS AND MATERIALS

**Peptide Synthesis Procedures**

All peptides were synthesized in our lab using the solid-phase peptide synthesis (SPPS) techniques (Merrifield, 1963). The peptides are synthesized from the C-terminus to the N-terminus by attaching amino acids one at a time onto the insoluble polymer resin. Unfinished peptides stay connected to the resin. Excess reagents and byproducts are easily washed away with repeated solvent washes. Modifications of the C- or N-terminus and incorporation of specific amino acids are easily achieved.

Synthesis was carried out in a 10 ml polypropylene syringe vessel (Torviq, Niles, MI). The vessel contains a porous polypropylene frit porosity at the bottom of the syringe to keep the resins from being washed out and allow only liquid materials to pass through. After the reaction mixture was drawn into the syringe using the plunger, it was capped with a removable polypropylene cap. Both ends are blocked to ensure the reaction mixture remaining in the vessel throughout the synthesis process. The reaction vessel was mounted onto a mechanical agitator (Model 75 Wrist-Action Shaker, Burrel Scientific, Pittsburgh, PA) to allow adequate mixing as the reactions occurred.

Wang resin was used to yield a carboxyl group at the C-terminus, while rink amide resin was used to yield an amidated C-terminus (Figure 2.1). Before the reaction, 0.1 mmol of the resin was swelled in *N,N*-Dimethylformamide (DMF) for 30 minutes to allow the resin to absorb the subsequent reaction cocktails efficiently. Throughout the synthesis, there were four reaction steps involved. They are deprotection, coupling, acetylation, and cleavage.

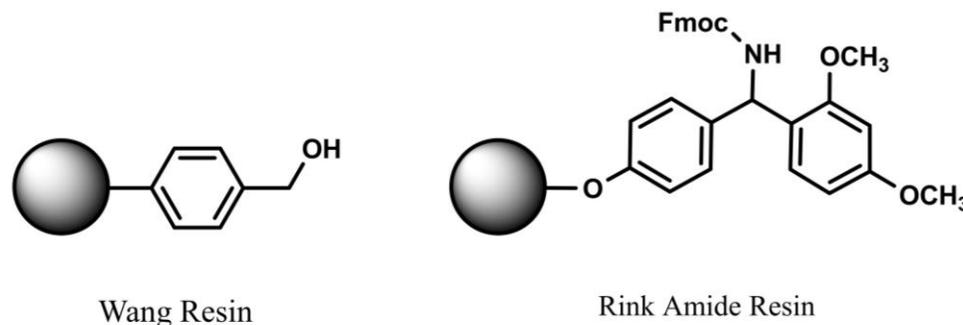


Figure 2.1. Wang resin (left) and rink amide resin (right) for peptide synthesis.

### Deprotection

The reactive  $\text{NH}_2$  group of the rink amide resin and the amino acids are protected with the Fmoc group. To elongate the peptide chain, the Fmoc group needs to be removed. Deprotection was carried out using 5 ml of 20% v/v Piperidine dissolved in DMF. After agitation for 15 minutes, the deprotection solution was drained, and washing was carried out to clean the resin. Alternatively, the deprotonation was carried out three times by using 2 ml of 20% piperidine/DMF each time and agitating at 5 min, 5 min, and 10 min, and followed by washing. 3 ml of methanol (MeOH) was drawn into the vessel, agitated for 30 seconds, and then discarded. Then 3 ml of dichloromethane (DCM) was drawn into the vessel, agitated for 30 seconds, and discarded. Washing with MeOH and DCM was repeated three times to ensure complete removal of the unreacted reaction mixture. An additional wash with DMF was also carried out to restore the condition of the resin for the subsequent reaction.

### Coupling

After deprotection, the reactive group on the resin is exposed and ready to attach a Fmoc protected amino acid. A 1:5:5:9 mole ratio solution of resin: Fmoc protected amino acid: N, N, N', N'-Tetramethyl-O-(1H-benzotriazol-1-yl) uranium hexafluorophosphate (HBTU): N, N-diisopropylethylamine (DIPEA) was mixed and dissolved in 4 ml of DMF. After mixing, the reaction mixture was drawn into the vessel to couple the amino acid onto the resin. The Fmoc group on the  $\text{NH}_2$  group of the amino acid prevents undesired side reactions from happening. Additionally, the reactive

group on the sidechain of the amino acid is protected with acid-labile protecting groups, which are removed during cleavage. After 2 hours of agitation, the reaction mixture was drained. Then washing was carried out as described in the deprotection step.

The deprotection and coupling steps were repeated using different amino acids to create the desired peptide. After reaching the targeted sequence, a final deprotection was carried out to remove the last Fmoc group. If N-terminal acetylation is desired, it is carried out before the cleavage.

### **N-terminus Acetylation**

A 1:5:8 mole ratio solution of Resin: Acetic Anhydride: DIPEA was dissolved in 4 ml of DMF and drawn into the vessel. After agitation of 1 hour, the reaction mixture was drained, and the resin was thoroughly washed as described previously.

### **Cleavage**

After reaching the peptide sequence with desired modifications, the resin was washed twice with DCM to remove the residue DMF on the resin. Because DMF is not volatile, the residue DMF in the resin significantly complicates the purification process. The cleavage cocktail was prepared by dissolving 0.22 g crystalline phenol with 220  $\mu$ l H<sub>2</sub>O and then adding 44  $\mu$ l of triisopropylsilane (TIPS) and 3.8 ml of trifluoroacetic acid (TFA). After agitation of 2 hours, the peptides were cleaved off the resin into the cleavage mixture, with their sidechain protection groups also removed. The cleavage mixture was collected in a 50 ml polypropylene falcon tube.

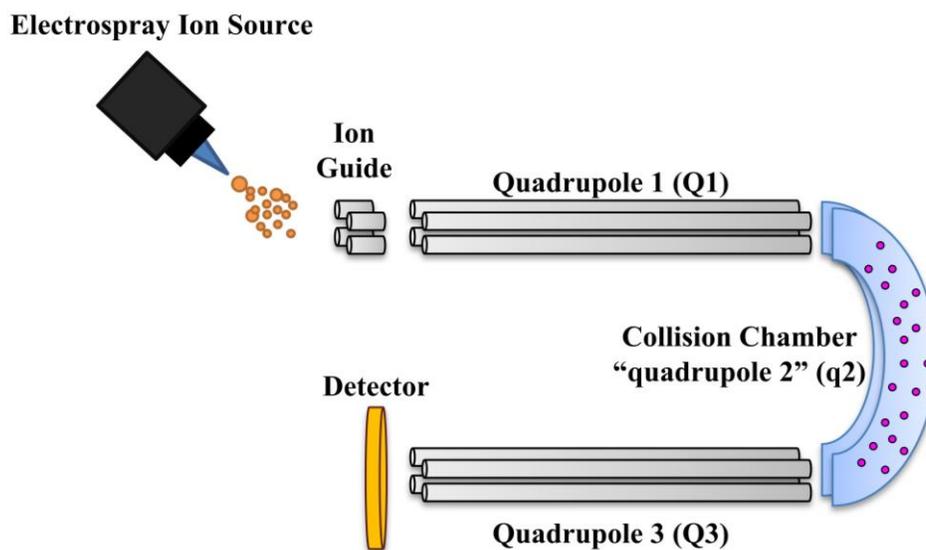
Purification of the product starts with evaporating the TFA with a gentle N<sub>2</sub> flow until a residue volume of about 1ml. Then 20 ml of cold diethyl ether was added to precipitate the peptides and dissolve the impurities. The tube was left in a -20 °C freezer overnight to efficiently separate the peptides from the impurities. Then the solution was centrifuged (HERMLE Labnet Z206A) at 6000 rpm for 10 minutes to pellet the precipitated peptide solids. The liquid containing the impurities was carefully decanted. The solid was washed twice with 10 ml of diethyl ether, centrifuged, and decanted. Then the solid was dissolved in 5-10 ml of ultrapure HPLC grade water and transferred to a 15 ml falcon tube.

The aqueous peptide solution was frozen under liquid N<sub>2</sub> and then lyophilized overnight to make a fluffy white powder. Finally, the peptide was sequenced with mass spectrometry to confirm its identity.

### Mass Spectrometry Measurements using the Cooks' Kinetic Method

Mass spectrometry experiments using the Cooks' kinetic method were performed to measure the gas-phase acidity of the cysteine-containing peptides. The gas-phase basicity measurements of peptides containing lysine and Dap were conducted using the same method by Patrick Batoon, a former graduate student in our lab. The detailed setup for acidity measurements is shown below, and the procedure for basicity measurements has been described in detail in Patrick Batoon's dissertation (Batoon, 2016). In the context of this dissertation, theoretical study is the main focus for the basicity of the peptides.

The primary instrument used in this research was the Varian 320L (Agilent Technologies, Santa Clara, CA) tandem quadrupole mass spectrometer coupled to an electrospray ionization (ESI) source. Data were acquired and recorded using the Varian MS Workstation software package (Version 6.9). A schematic representation of the instrument is shown in Scheme 2.1.



Scheme 2.1. Schematic representation of the Varian 320L mass spectrometer.

## Sample Introduction

Prior to introduction into the mass spectrometer, the samples were prepared with an appropriate concentration, about  $10^{-4}$  to  $10^{-5}$  M of the peptide and the reference acid using a 1:1 (v/v) HPLC grade MeOH/H<sub>2</sub>O. Concentrations were varied to optimize the intensity of the proton-bound dimer between the peptides and the reference acids. In the experiments using reference acids, the proton-bound dimer is [Ref<sup>-</sup>⋯H<sup>+</sup>⋯Pep<sup>-</sup>]. In the experiments between peptides, the proton-bound dimer is [PepA<sup>-</sup>⋯H<sup>+</sup>⋯PepB<sup>-</sup>]. The samples were loaded into a 1 ml syringe (Gastight 1001 TLL SYR, Hamilton, Reno, NV) and introduced into the mass spectrometer through direct infusion using a Chemyx Fusion 100 syringe pump at 10-20 ul/min.

Besides varying the concentration, the ESI source conditions, such as ESI needle voltage, nebulizing gas pressure, drying gas pressure and temperature, shield voltage, and capillary voltage were adjusted to optimize the intensity of the proton-bound dimer. The ESI needle voltage was adjusted between -4.0 and -5.0 kV, and compressed air (Fresno Oxygen, Fresno, CA) was used as the nebulizing gas with a pressure of 30.0 to 50.0 psi. Nitrogen (Ultra-High Purity, Air Products and Chemicals Inc.) was used as the drying gas with a pressure of 10.0 to 16.0 psi and a temperature of 160 °C to 190 °C. Shield voltage has remained at 600 V. Capillary voltage was adjusted between -30 and -45 V.

Before entering the low-pressure (about  $2.8 \times 10^{-6}$  Torr) mass analyzer regions where mass analysis and detection occur, the generated ions were thermalized within the ion guide chamber pressurized at 1 to 2 mTorr with N<sub>2</sub> gas. The relatively high pressure “collisionally cools” the ions to a room-temperature equilibrium state by multiple collisions with the N<sub>2</sub> molecules.

## Fragmentation of Proton-Bound Dimer

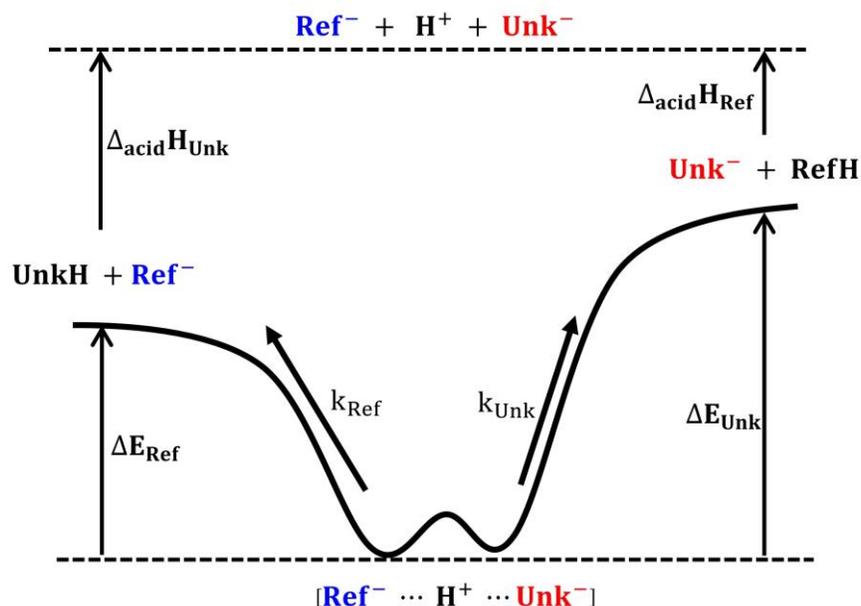
The measurements were conducted using the fragmentation experiments in the mass spectrometer in which the q2 chamber is pressurized at 0.2 to 0.65 mTorr of argon gas. For those experiments, the proton-bound dimer was mass isolated in Q1, and then accelerated into q2 and subjected to collision-induced dissociation (CID) by a range of voltages. The acceleration voltages correspond to the laboratory frame collision energies ( $E_{\text{lab}}$ ), which determine the ion's center-of-mass ( $E_{\text{cm}}$ ) collision energy between

0.5 eV and 3.5 eV. The  $E_{\text{cm}}$  values were calculated through eq 2.1, where  $E_{\text{cm}}$  is the center-of-mass collision energy,  $E_{\text{lab}}$  is the laboratory frame collision energy (acceleration voltage),  $m_{\text{Ar}}$  is the mass of argon, and  $M$  is the mass of the proton-bound dimer ion. The dissociation product ions were analyzed using the third quadrupole (Q3).

$$E_{\text{cm}} = E_{\text{lab}} \frac{m_{\text{Ar}}}{m_{\text{Ar}} + M} \quad (\text{eq 2.1})$$

### Standard Kinetic Method

In the work of R.G. Cooks, et al. (R. G. Cooks & Kruger, 1977), it was found that dissociation of a complex comprised of two different parts loosely bound by a single proton ( $\text{H}^+$ ), i.e., the proton-bond dimer, resulted in the procession of two parallel reaction pathways. Assuming no (or equal in both pathways) reverse activation energy and that only structurally similar compounds are involved, dissociation of the proton-bond dimer gives rise to two species via the two competitive dissociation channels with different kinetic rate constants depending on their relative proton affinity (Xubin Zheng & Cooks, 2002). For demonstration, the dissociation of the proton-bond dimer between the unknown,  $\text{Unk}^-$ , and a reference acid,  $\text{Ref}^-$ , is explained as an example in the following paragraphs and is shown in Scheme 2.2.



Scheme 2.2. Proton-bound dimer activation/dissociation pathways.

$\Delta E_{\text{unk}}$  and  $\Delta E_{\text{Ref}}$  are the threshold dissociation energies to activate the proton-bound dimer and observe the corresponding fragment ions.  $k_{\text{unk}}$  and  $k_{\text{Ref}}$  are the unimolecular kinetic rate constant of the corresponding dissociation reaction.  $\Delta_{\text{acidH}}\text{Unk}$  and  $\Delta_{\text{acidH}}\text{Ref}$  are the deprotonation enthalpies, i.e., the enthalpy changes of the deprotonation reaction of the neutral compounds (Scheme 2.3). The dissociation of the proton-bound dimer results in the competitive formation of the  $\text{Unk}^-$  or  $\text{Ref}^-$  ions and the neutral molecules  $\text{RefH}$  and  $\text{UnkH}$ , respectively (Scheme 2.2). The formation of the products depends on the competitive retention of the proton. The ion with greater relative proton affinity tends to produce the corresponding neutral compound, while the ion with lower relative proton affinity would stay charged and produce a greater ion signal. Using the described interpretation, in Scheme 2.3,  $\text{UnkH}$ , the neutral molecule corresponding to the ion with greater proton affinity, is less acidic than  $\text{RefH}$ , the neutral molecule corresponding to the ion with lower proton affinity.



$$\Delta_{\text{rxn}}\text{H} = \Delta_{\text{acid}}\text{H}_{\text{MH}}$$

$$\Delta_{\text{rxn}}\text{G} = \Delta_{\text{acid}}\text{G}_{\text{MH}}$$

$$\Delta_{\text{acid}}\text{G}_{\text{MH}} = \Delta_{\text{acid}}\text{H}_{\text{MH}} - T\Delta_{\text{acid}}\text{S}_{\text{MH}}$$

Scheme 2.3. Deprotonation reaction of a neutral molecule, MH.

At lower collision energies, the main fragment ions were the deprotonated reference acid ( $\text{Ref}^-$ ) and the deprotonated unknown ( $\text{Unk}^-$ ). At higher collision energies, secondary fragments started to occur. The fragment ion intensities were recorded using the selected reaction monitoring (SRM) mode. Secondary fragment ions were linearly incorporated into the primary fragments to yield  $I_{\text{Unk}^-}$  and  $I_{\text{Ref}^-}$ , which are used to calculate the branching ratios,  $\ln[I_{\text{Unk}^-}/I_{\text{Ref}^-}]$ . Branching ratio plots were constructed by plotting the branching ratios  $\ln[I_{\text{Unk}^-}/I_{\text{Ref}^-}]$  against the collision energies ( $E_{\text{cm}}$ ).

The branching ratios,  $\ln[I_{\text{Unk}^-}/I_{\text{Ref}^-}]$ , are intrinsically regulated by the difference in deprotonation enthalpy between the involving neutral compounds. The standard definition for the unimolecular kinetic rate constant,  $k$ , can be expressed in the form shown in eq 2.2 (Xubin Zheng & Cooks, 2002).

$$k = \frac{RT_{\text{eff}}}{h} \frac{Q^*}{Q} \exp\left(\frac{-\Delta E}{RT_{\text{eff}}}\right) \quad (\text{eq 2.2})$$

In eq 2.2,  $R$  represents the fundamental gas constant,  $T_{\text{eff}}$  is the “effective temperature” of the system (Ervin, 2000) that does not represent the actual temperature at which dissociation occurs, but rather an empirical measure of the internal energy of the system that can be modified by increasing the collision energy or the collision gas pressure,  $h$  represents Planck’s constant,  $Q^*$  is the partition function for internal modes of the activated complex,  $Q$  is the partition function for internal modes of the proton-bound dimer, and  $\Delta E$  is the dissociation energy. The difference in deprotonation enthalpy rises from rearrangements of the branching ratio,  $\ln[I_{\text{Unk}^-}/I_{\text{Ref}^-}]$ , by substituting the ion intensity ratio with the rate constant ratio, as shown in eq 2.3.

$$\begin{aligned}
\ln \left[ \frac{I_{\text{Unk}^-}}{I_{\text{Ref}^-}} \right] &= \ln \frac{k_{\text{Unk}}}{k_{\text{Ref}}} = \ln(k_{\text{Unk}}) - \ln(k_{\text{Ref}}) && \text{(eq 2.3)} \\
&= \ln \left( \frac{RT_{\text{eff}}}{h} \frac{Q_{\text{Unk}}^*}{Q} \exp\left(\frac{-\Delta E_{\text{Unk}}}{RT_{\text{eff}}}\right) \right) - \ln \left( \frac{RT_{\text{eff}}}{h} \frac{Q_{\text{Ref}}^*}{Q} \exp\left(\frac{-\Delta E_{\text{Ref}}}{RT_{\text{eff}}}\right) \right) \\
&= \ln \left( \frac{RT_{\text{eff}}}{h} \frac{Q_{\text{Unk}}^*}{Q} \right) + \ln \left( \exp\left(\frac{-\Delta E_{\text{Unk}}}{RT_{\text{eff}}}\right) \right) - \ln \left( \frac{RT_{\text{eff}}}{h} \frac{Q_{\text{Ref}}^*}{Q} \right) - \ln \left( \exp\left(\frac{-\Delta E_{\text{Ref}}}{RT_{\text{eff}}}\right) \right) \\
&= \ln \left( \frac{RT_{\text{eff}}}{h} \frac{Q_{\text{Unk}}^*}{Q} \right) + \frac{-\Delta E_{\text{Unk}}}{RT_{\text{eff}}} - \ln \left( \frac{RT_{\text{eff}}}{h} \frac{Q_{\text{Ref}}^*}{Q} \right) - \frac{-\Delta E_{\text{Ref}}}{RT_{\text{eff}}} \\
&= \frac{\Delta E_{\text{Ref}} - \Delta E_{\text{Unk}}}{RT_{\text{eff}}} + \ln \left( \frac{RT_{\text{eff}}}{h} \frac{Q_{\text{Unk}}^*}{Q} \times \frac{h}{RT_{\text{eff}}} \frac{Q}{Q_{\text{Ref}}^*} \right) \\
&= \frac{\Delta_{\text{acid}}H_{\text{Ref}} - \Delta_{\text{acid}}H_{\text{Unk}}}{RT_{\text{eff}}} + \ln \left( \frac{Q_{\text{Unk}}^*}{Q_{\text{Ref}}^*} \right)
\end{aligned}$$

Under the simplifying assumption that no (or equal) reverse activation energy, the term  $\ln(Q_{\text{Unk}}^*/Q_{\text{Ref}}^*)$  is equivalent to the reaction entropy difference between the two fragmentation channels,  $(\Delta_{\text{acid}}S_{\text{Unk}} - \Delta_{\text{acid}}S_{\text{Ref}})/R$  or  $\Delta(\Delta S)/R$ . Substitution of the  $\ln(Q_{\text{Unk}}^*/Q_{\text{Ref}}^*)$  by the term  $\Delta(\Delta S)/R$  leads to eq 2.4.

$$\ln \left[ \frac{I_{\text{Unk}^-}}{I_{\text{Ref}^-}} \right] = \ln \frac{k_{\text{Unk}}}{k_{\text{Ref}}} = \frac{\Delta_{\text{acid}}H_{\text{Ref}} - \Delta_{\text{acid}}H_{\text{Unk}}}{RT_{\text{eff}}} + \frac{\Delta(\Delta S)}{R} \quad \text{(eq 2.4)}$$

The entropy term can be canceled under the assumption that the reference compound is structurally similar to the unknown (Xubin Zheng & Cooks, 2002). This leads to the standard version of the kinetic method described in eq 2.5.

$$\ln \left[ \frac{I_{\text{Unk}^-}}{I_{\text{Ref}^-}} \right] = \ln \frac{k_{\text{Unk}}}{k_{\text{Ref}}} = \frac{\Delta_{\text{acid}}H_{\text{Ref}} - \Delta_{\text{acid}}H_{\text{Unk}}}{RT_{\text{eff}}} \quad \text{(eq 2.5)}$$

When multiple reference acids are used, the plot of  $\ln[I_{\text{Unk}^-}/I_{\text{Ref}^-}]$  versus  $\Delta_{\text{acid}}H_{\text{Ref}i}$  should yield a straight line, for which the x-intercept indicates the deprotonation enthalpy of the unknown,  $\Delta_{\text{acid}}H_{\text{Unk}}$ . The standard kinetic method is highly sensitive, and has been used to tackle the subtle changes in the thermochemical values caused by isotope substitution and chirality (Fedick, Bain, Bain, & Cooks, 2017; Nourse & Graham Cooks, 1991; T. I. Williams, Denault, & Cooks, 2001).

### Extended Kinetic Method

In order for the standard kinetic method to hold, the entropy changes for the reaction channels involving each reference and the unknown have to be equal to eliminate entropic effects on the relative

rate constants (Xubin Zheng & Cooks, 2002). This strict requirement has drawn criticism, and the extended kinetic method was developed to address this issue by taking the entropic effects into account (Armentrout, 2000; Cerda & Wesdemiotis, 1996; Cheng, Wu, & Fenselau, 1993; R. Graham Cooks & Wong, 1998; Robert Graham Cooks, Koskinen, & Thomas, 1999; Ervin & Armentrout, 2004; Graham Cooks et al., 1994). In many cases, it is difficult to obtain a series of reference compounds with a similar structure to the unknown. However, it is much easier to obtain the reference compounds dissimilar in structure to the unknown, but structurally similar among themselves. Then the ineligible entropic term,  $\Delta(\Delta S)/R$ , will likely remain constant among them and the plot of  $\ln[I_{\text{Unk}^-}/I_{\text{Ref } i^-}]$  versus  $\Delta_{\text{acid}}H_{\text{Ref } i}$  should still yield a straight line. The branching ratios will reflect the difference in the gas-phase acidity between the involving neutral compounds (eq 2.6 and eq 2.7).

$$\begin{aligned} \ln \left[ \frac{I_{\text{Unk}^-}}{I_{\text{Ref A}^-}} \right] &= \ln \frac{k_{\text{Unk}}}{k_{\text{Ref A}}} && \text{(eq 2.6)} \\ &= \frac{\Delta_{\text{acid}}H_{\text{Ref A}} - \Delta_{\text{acid}}H_{\text{Unk}}}{RT_{\text{eff}}} + \frac{\Delta_{\text{acid}}S_{\text{Unk}} - \Delta_{\text{acid}}S_{\text{Ref A}}}{R} \\ &= \frac{\Delta_{\text{acid}}G_{\text{Ref A}} - \Delta_{\text{acid}}G_{\text{Unk}}}{RT_{\text{eff}}} \end{aligned}$$

$$\begin{aligned} \ln \left[ \frac{I_{\text{Unk}^-}}{I_{\text{Ref B}^-}} \right] &= \ln \frac{k_{\text{Unk}}}{k_{\text{Ref B}}} && \text{(eq 2.7)} \\ &= \frac{\Delta_{\text{acid}}H_{\text{Ref B}} - \Delta_{\text{acid}}H_{\text{Unk}}}{RT_{\text{eff}}} + \frac{\Delta_{\text{acid}}S_{\text{Unk}} - \Delta_{\text{acid}}S_{\text{Ref B}}}{R} \\ &= \frac{\Delta_{\text{acid}}G_{\text{Ref B}} - \Delta_{\text{acid}}G_{\text{Unk}}}{RT_{\text{eff}}} \end{aligned}$$

$$\text{with } \frac{\Delta(\Delta S)}{R} = \frac{\Delta_{\text{acid}}S_{\text{Unk}} - \Delta_{\text{acid}}S_{\text{Ref A}}}{R} = \frac{\Delta_{\text{acid}}S_{\text{Unk}} - \Delta_{\text{acid}}S_{\text{Ref B}}}{R} \neq 0$$

With the branching ratios,  $\ln[I_{\text{Unk}^-}/I_{\text{Ref } i^-}]$ , obtained for multiple references at multiple collision energies, the values of enthalpy ( $\Delta_{\text{acid}}H_{\text{Unk}}$ ) and the reaction entropy difference,  $\Delta(\Delta S)$ , could be determined through two linear plots (Figure 2.2) using the following linear relationship in eq 2.8.

$$\ln \left[ \frac{I_{\text{Unk}^-}}{I_{\text{Ref}^-}} \right] = \frac{\Delta_{\text{acid}}H_{\text{Ref}} - \Delta_{\text{acid}}H_{\text{Avg}}}{RT_{\text{eff}}} + \left[ \frac{-(\Delta_{\text{acid}}H_{\text{Unk}} - \Delta_{\text{acid}}H_{\text{Avg}})}{RT_{\text{eff}}} - \frac{-\Delta(\Delta_{\text{acid}}S)}{R} \right] \quad (\text{eq 2.8})$$

$$\text{with } \Delta_{\text{acid}}H_{\text{Avg}} = \frac{\sum_n \Delta_{\text{acid}}H_{\text{Ref}}}{n}$$

$$\text{and } \Delta(\Delta_{\text{acid}}S) = \Delta_{\text{acid}}S_{\text{Unk}} - \Delta_{\text{acid}}S_{\text{Ref}} = \Delta_{\text{acid}}S_{\text{Unk}} - \Delta_{\text{acid}}S_{\text{Avg}}$$

In the first EKM plot, the branching ratios for the reference acids at the same collision energy ( $E_{\text{cm}}$ ),  $\ln[I_{\text{Unk}^-}/I_{\text{Ref}^-}]$ , are plotted against  $\Delta_{\text{acid}}H_{\text{Ref}} - \Delta_{\text{acid}}H_{\text{Avg}}$ . Each of the  $E_{\text{cm}}$  energy corresponds to a particular  $T_{\text{eff}}$ , which can be empirically determined from the slope of individual regression lines,  $1/RT_{\text{eff}}$ . The Y-axis intercept of each regression line corresponds to the terms in the square bracket in eq 2.8.

The Y-axis intercepts and slopes from the first EKM are used to make the second EKM plot. They are linearly correlated, as shown in eq 2.8. The slope of the regression line in the second EKM plot is the term  $-(\Delta_{\text{acid}}H_{\text{Unk}} - \Delta_{\text{acid}}H_{\text{Avg}})$ , which contains the deprotonation enthalpy value of the unknown. The Y-axis intercept of the regression line in the second EKM plot contains the entropy term.

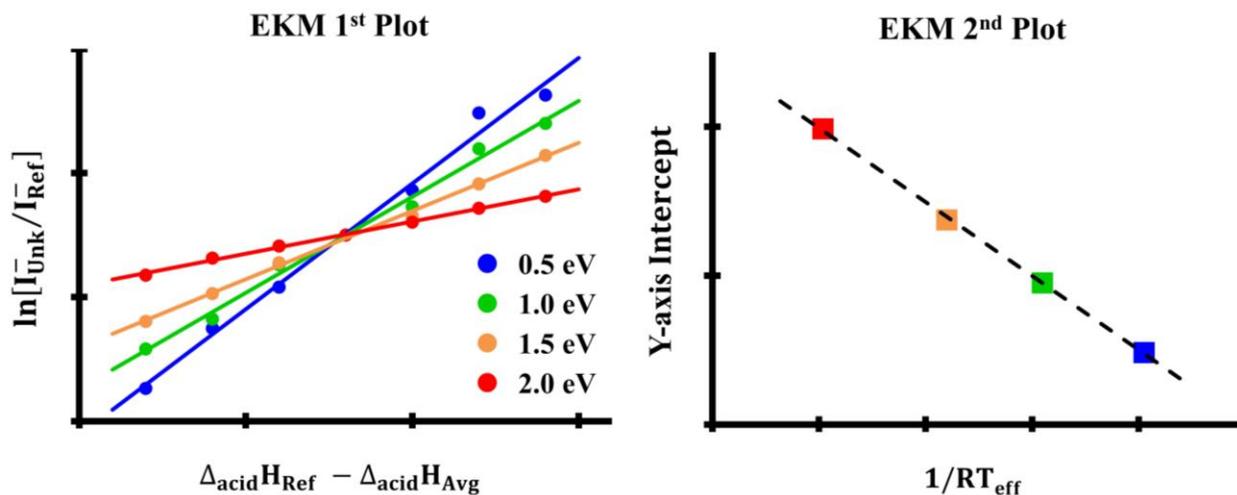
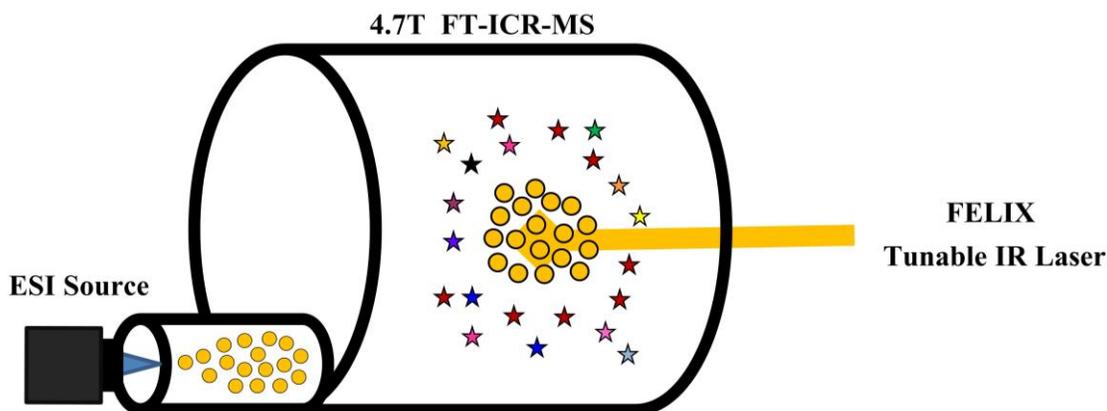


Figure 2.2. Extended kinetic method (EKM) 1<sup>st</sup> plot (left) and 2<sup>nd</sup> plot (right). (The plots show idealized data.)

### Infrared Multiple-Photon Dissociation (IRMPD) Spectroscopy

Gas-phase infrared (IR) spectroscopy was used to obtain the gas-phase IR spectra of the charged peptides, which provide valuable information to elucidate their structures (Eyler, 2009; Polfer, 2011).

Theoretical IR spectra of the computed conformations were matched to the experimental spectrum to identify conformations that dominate the conformational populations of each protonated peptide. The Infrared Multiple-Photon Dissociation Spectroscopy (IRMPD) experiments were conducted using the Free Electron Laser for Infrared eXperiments (FELIX) beamline coupled to a Bruker 4.7 T Fourier-transform ion-cyclotron resonance (FT-ICR) mass spectrometer at Radboud University, Nijmegen, The Netherlands (Martens, Grzetic, Berden, & Oomens, 2015; Polfer & Oomens, 2009; Van Amersfoort et al., 1989). A schematic representation of the IRMPD setup is shown in Scheme 2.4.



Scheme 2.4. Schematic representation of the IRMPD experimental setup.

The charged peptides were introduced into the gas phase by electrospray ionization, and then mass-selected and trapped in the trapping region of the mass spectrometer. A tunable IR laser at a specific wavelength is generated using the FELIX beamline and guided into the ion trap of the FT-ICR mass spectrometer. The ions absorb the incoming IR light if its frequency matches one of the ions' vibrational modes. The absorbed energy is redistributed through intramolecular vibrational redistribution (IVR) and allows reabsorption of another photon of the same frequency. After absorption of multiple photons, the accumulated energy brings the ion over a fragmentation threshold and causes the ions to break apart into fragment ions. The ion intensity of the precursor peptide ion ( $I_{\text{precursor}}$ ) and the fragment ions ( $I_{\text{fragment}}$ ) are recorded and used to calculate the IRMPD yield (eq 2.9) at the specific wavenumber.

The frequency of the IR laser is then adjusted to obtain the IRMPD yield at different frequencies (500 to 2000  $\text{cm}^{-1}$ ). An IRMPD spectrum is obtained by plotting the IRMPD yield against the wavelength or wavenumber.

$$\text{IRMPD Yield} = \frac{\Sigma(I_{\text{fragment}})}{\Sigma(I_{\text{fragment}} + I_{\text{precursor}})} \quad (\text{eq 2.9})$$

### Computational Methods

In conjunction with the various experimental studies, the computational study was carried out to investigate the peptides *in silico*. When executed properly, high accuracy information is obtained, providing valuable insight into the experimental observations. In the context of this research, the desired outcome of computational modeling was to determine the energetically stable and structurally representative conformations that most closely represent the experimental results. The calculated species include the neutral and charged forms of the peptides, ethanethiol, and ethylamine, which are the reference compounds in the isodesmic reactions to predict the gas-phase acidity/basicity.

#### Finding the-Lowest Energy Conformations

For this research, a three-step funneling approach was used to determine the lowest energy conformations of the peptides. It includes an extensive conformational search step at the semiempirical quantum chemical (SQM) level of theory, a robust optimization and screening step at the Hartree-Fock (HF) level of theory, and a final geometry optimization and frequency calculation step at the density functional theory (DFT) level.

The conformational search step began with building the initial peptide structure as a linearly extended strand using the Spartan '14 software (Wavefunction Inc., Irvine, CA) or GaussView 5.0.8 software (Gaussian Inc., Wallingford, CT). The initial structure in XYZ format is used as the seed structure for two approaches to explore the conformational space. The two conformational search approaches (CREST and systematic) are described in detail in the following part of this section.

**The CREST approach.** The first approach utilizes the computational program CREST version 2.10 (abbreviated from Conformer-Rotamer Ensemble Sampling Tool) (Grimme, 2019). The

implemented iMTD-GC workflow in CREST was used for automatic conformer generation and screening to find low-lying conformers. It is a composite approach consisting of meta-dynamics (MTD) sampling, regular molecular dynamic (MD) sampling, and genetic Z-matrix structure crossing (GC) (Pracht, Bohle, & Grimme, 2020). The lowercase i indicates an iterative strategy within the algorithm, meaning if a new lower energy conformer is found at any point during the conformational search, the entire process is restarted with this conformation as an input. The approach heavily relies on the tight-binding method GFN-xTB (short for Geometries, Frequencies, and Noncovalent interactions – extended TightBinding), for fast and robust geometry optimization and energy calculation at a semiempirical quantum chemical (SQM) level of theory.

Meta-dynamics (MTD) is a form of molecular dynamic (MD) simulation, first suggested by Alessandro Laio and Michele Parrinello (Laio & Parrinello, 2002). In the MTD algorithm, the system is described by a set of collective variables (CVs) and a history-dependent potential. The collective variables determine the location of the system in the conformational space, and the history-dependent potential term describes the energy of the system. During the progression of the simulation, positive Gaussian potentials are continuously added to the history-dependent potential term at where the simulation has visited. The added Gaussian potentials discourage resampling of previously visited states, which helps overcome the large energy barrier, and allows a broader exploration of the potential energy surface (PES). It has been informally described as “filling the energy wells with computational sand” (Bernardi, Melo, & Schulten, 2015). A schematic representation of a one-dimensional PES that is filled by additive bias potentials is shown in Figure 2.3.

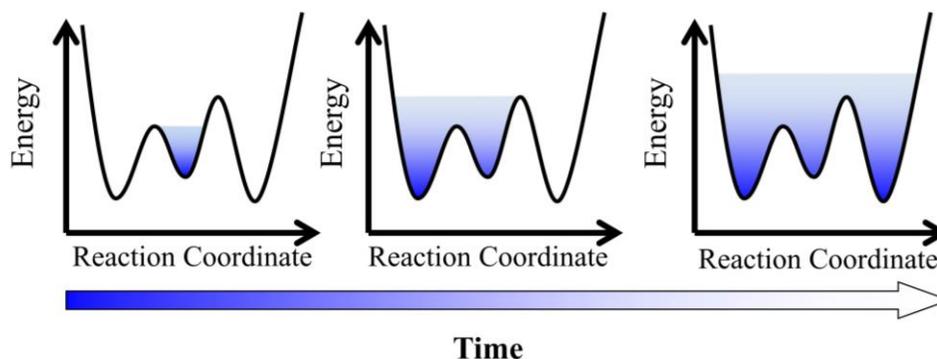


Figure 2.3. Illustration of the metadynamics method described as “filling the energy wells with computational sand”.

In the MTD algorithm implemented in CREST, the collective variables are previous minima on the PES, expressed as standard root-mean-square deviation (RMSD) in Cartesian space between them (Pracht et al., 2020). The RMSD  $\Delta$  between the actual structure in a simulation and a reference structure is given by eq 2.10, where  $r_j$  is a component of the Cartesian space vector of the actual molecule,  $r_j^{\text{ref},i}$  is the corresponding element in reference structure  $i$ , and  $N$  is the number of atoms.

$$\Delta_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (r_j - r_j^{\text{ref},i})^2} \quad (\text{eq 2.10})$$

The quaternion algorithm of Coutsiias et al. (Coutsiias, Seok, & Dill, 2004), is employed for the alignment of the structures and to minimize the RMSD between the two sets of vectors. The history-dependent potential is given in the form of Gaussian-type potentials as shown in eq 2.11, where  $\Delta_i$  are the collective variables, i.e., the atomic RMSDs,  $\alpha_i$  determines the Gaussian potential’s shape,  $n$  is the number of reference structures, and  $k_i$  is the pushing strength. Here,  $\alpha_i$  typically has values between 0.1 and 1.3 Bohr<sup>-2</sup>, which can be seen as the “range” of the bias. The constant  $k_i$  is scaled by the number of atoms  $N$ , where  $k_i/N$  has magnitudes of 0.75 to 3.00 mHartree.

$$V_{\text{bias}} = \sum_i^n k_i e^{-\alpha_i \Delta_i^2} \quad (\text{eq 2.11})$$

The genetic structure crossing is implemented to improve further the efficiency of generating new conformers. In the concept of genetic algorithms, structural elements present only in already generated structures are projected onto a reference structure to create new structures. In the genetic crossing (GC)

algorithm implemented in CREST, the internal (Z-matrix, R) coordinates are automatically generated for all conformers in the existing ensemble, and the energetically lowest structure found so far ( $R_{\text{ref}}$ ) is selected as the reference. The difference in all internal coordinates (i.e., bond length, bond angles, and dihedral angles) between all pairs of structures are calculated and applied to the reference according to eq 2.12, where  $i$  and  $j$  label the pairs of structures, and  $R_{\text{new}}$  is the generated new structure.

$$R_{\text{new}} = R_{\text{ref}} + R_i - R_j \quad (\text{eq 2.12})$$

In this way, structural differences present only in  $i$  and  $j$  are projected onto the reference structure to generate a resulting new conformer. The generated (crossed) structures are checked for any close atomic contacts (i.e., ‘clash’ checks) to filter out unreasonable conformers.

The schematic representation of the iMTD-GC workflow is shown in Figure 2.4. Before the conformational sampling steps, the maximum MTD length is determined based on the molecular size and flexibility of the system, and the default technical settings for MTD sampling are evaluated to make sure the MTDs will run smoothly.

Twelve MTD sampling runs are performed with different pairs of  $k_i$  and  $\alpha_i$  for  $V_{\text{bias}}$  potential. Within the simulations, a new structure is collected into the conformer ensemble and added to the  $V_{\text{bias}}$  potential every 1.0 ps. The collected conformations are reoptimized without the biasing potential in a two-step filtering procedure. The first step is with an energy window of 12 kcal/mol and very crude convergence criteria, where the energy convergence ( $5 \times 10^{-4}$  Hartree) and gradient convergence ( $1 \times 10^{-2}$  Hartree $\cdot\alpha^{-1}$ ) are large. The second step is with an energy window of 6 kcal/mol and tight convergence criteria, where the energy convergence ( $1 \times 10^{-6}$  Hartree) and gradient convergence ( $8 \times 10^{-4}$  Hartree $\cdot\alpha^{-1}$ ) are small. Afterward, the optimized structures in the ensemble are sorted and filtered on the basis of the energy, the atomic Cartesian RMSD, and the molecular rotational constant to remove identical isomers and high energy conformers. If a new structure is found lower in energy than the input, the conformational sampling process is restarted with the new structure. Otherwise, the workflow is continued. The intermediate conformer ensemble is saved to be compared at a later stage.

In the second step of the process, two normal MD sampling at two different temperatures, 400 K and 500 K, are carried out on the three to six lowest energy conformers from the intermediate conformer ensemble. It helps to sample the conformational space around these low energy conformers more extensively. All resulting conformers are reoptimized, sorted, filtered as described in the previous step, and included in the intermediate conformer ensemble. In the final step, genetic Z-matrix structure crossing is performed. It helps to further complete the conformer ensemble and is particularly useful for flexible molecules. If a new lower energy conformer is found, the entire conformational search process is restarted. The restarts are particularly useful for large molecules whose global minimum structure is far from the initial input geometry.

All collected conformers throughout the process are then optimized with very tight convergence criteria where the energy convergence ( $1 \times 10^{-7}$  Hartree) and gradient convergence ( $2 \times 10^{-4}$  Hartree $\cdot\alpha^{-1}$ ) are very small. A final conformer ensemble is created by sorting and filtering conformations with an energy window of 6 kcal/mol and outputted in the Xmol XYZ format. And the lowest energy conformer is written in a separate XYZ file.

A screen module is available in the CREST program to optimize, sort, and filter conformations as a standalone tool. The input and output of the screen module are both in Xmol XYZ format. The conformers are geometry optimized in a multilevel, three-step-filtering procedure by applying loose to tight and then very tight optimization threshold settings and energy windows of 24, 12, and 6 kcal/mol, respectively. To further increase the chance of finding the global minimum of longer peptides, a bash script was written to conduct multiple CREST runs consecutively. The lowest energy structure from the previous CREST run is used as the input for the next run. When the same conformer was determined as the lowest energy structure in three consecutive CREST runs, the search stops and the conformer ensembles from all CREST runs were collected and screened to create the final conformer ensemble. The installation and execution of the CREST and xtb are shown in Appendix 1.

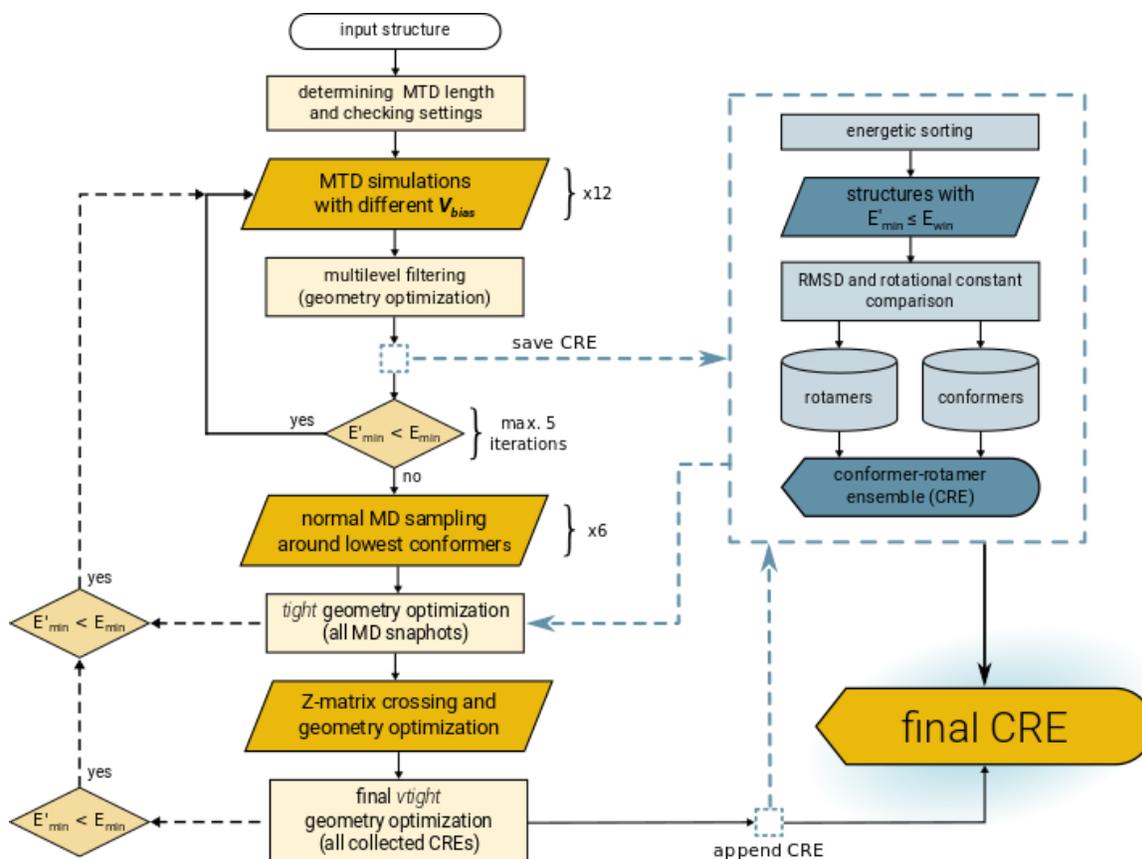


Figure 2.4. Schematic representation of the iMTD-GC workflow used for conformational search in CREST. (Taken from Figure 5 of reference Pracht et al., 2020)

**The systematic approach.** In the second conformational search approach, the conformers are generated systematically utilizing the genrot module in the molclus program developed by Tian Lu (Lu, 2020). The seed conformation is provided as a structure file in XYZ format. A configuration file is created manually to define the bonds to be rotated and the method of rotation. Theoretically, an unlimited number of conformers can be generated by rotating each bond with a very small angle. However, practically, the resulting ensemble for the peptide species contains no more than 100,000 conformations. The time consumption for the downstream work increases linearly with the number of conformers generated in this step. For the peptide systems of interest, all single bonds are rotated every 120 degrees to create three different orientations, and all amide bonds are rotated every 180 degrees to create the cis and trans configurations. The generated conformers are checked for atomic clashes to filter out

unreasonable conformers. The resulting conformers are outputted in the Xmol XYZ format and then optimized, sorted, and filtered using the screen module in the CREST program. Compared to the first approach, this systematic approach screens through a significantly larger number of conformers, takes a much longer time, and was only used for a few peptide systems.

After exhaustive exploration of the conformational space, the output data from both CREST and systematic approaches are combined. Duplicate and high-energy structures were removed to clean up the final ensemble of conformers using the screen module in the crest program. Then the structures in Xmol XYZ format are converted into individual Gaussian Input Data files using the xyz2QC module in the molclus program and are subjected to geometry optimization and energy calculation at the Hartree Fock (HF) level of theory using the 3-21G\* basis set to yield a better energy ladder. This step and the final step of geometry optimization and frequency calculation were carried out using the Gaussian 16' suite of programs (Frisch et al., 2016). After being ranked in free energy, the conformations within appropriate energy windows were kept for the next step.

In the final step, the remaining conformations are subjected to geometry optimization and frequency calculation at high-level density functional theory methods. The calculations were carried out using the B3LYP-D3BJ or wB97XD functionals with a 6-311+G\*\* basis set. In the case of input structures converging into the same final geometry, the degenerate structures were removed. All calculations are ensured that every conformer had a geometry absence of an imaginary frequency, implying true energy minimum.

### **Calculating the Theoretical Thermochemical Values**

The predicted enthalpy and Gibbs free energy values were extracted for each structure using a python script written by Michael D. Browne from our lab. A final set of no more than ten conformations with the lowest free energy were selected to represent an individual peptide species. For the resulting conformations, their Boltzmann probabilities,  $p_i$ , were calculated based on their free energy values using eq 2.13,

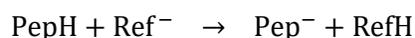
$$p_i = \frac{\exp(-\Delta G_i/RT)}{\sum \exp(-\Delta G_i/RT)} \quad (\text{eq 2.13})$$

Where  $p_i$  represents the Boltzmann probability,  $\Delta G_i$  is the Gibbs free energy relative to the lowest free energy conformation,  $R$  is the fundamental gas constant ( $1.987 \times 10^{-3} \text{ kcal mol}^{-1} \text{ K}^{-1}$ ), and  $T$  is the temperature (298 K). The Boltzmann weighted average enthalpy ( $H_{\text{avg}}$ ) and free energy ( $G_{\text{avg}}$ ) were calculated using eq 2.14 and eq 2.15, respectively,

$$H_{\text{avg}} = \sum(p_i H_i) \quad (\text{eq 2.14})$$

$$G_{\text{avg}} = \sum(p_i G_i) \quad (\text{eq 2.15})$$

where  $p_i$  represents the Boltzmann probability, and  $H_i$  and  $G_i$  are the enthalpy and free energy of individual conformations, respectively. The weighted average enthalpy and free energy values were then used for the acidic peptides to calculate the enthalpy and free energy change of the isodesmic reaction shown in Scheme 2.5.



$$\Delta H_{\text{rxn}} = (\Delta H_{\text{Pep}^-} + \Delta H_{\text{RefH}}) - (\Delta H_{\text{PepH}} + \Delta H_{\text{Ref}^-})$$

$$\Delta H_{\text{rxn}} = (\Delta H_{\text{Pep}^-} - \Delta H_{\text{PepH}}) - (\Delta H_{\text{Ref}^-} - \Delta H_{\text{RefH}})$$

$$\Delta H_{\text{rxn}} = (\Delta H_{\text{H}^+} + \Delta H_{\text{Pep}^-} - \Delta H_{\text{PepH}}) - (\Delta H_{\text{H}^+} + \Delta H_{\text{Ref}^-} - \Delta H_{\text{RefH}})$$

Scheme 2.5. Isodesmic reaction scheme to determine the theoretical deprotonation enthalpy,  $\Delta_{\text{acid}}H_{\text{pep}}$ .

The PepH and RefH represent the neutral peptide and reference acid, and  $\text{Pep}^-$  and  $\text{Ref}^-$  represent the deprotonated peptide and reference acid, respectively. In this reaction, the proton is transferred from the neutral peptide to the anionic reference acid, which was ethanethiol ( $\text{CH}_3\text{CH}_2\text{SH}$ ). By inputting Boltzmann averaged  $\Delta H$  values obtained for the neutral and deprotonated molecules,  $\Delta H_{\text{rxn}}$  can be determined.

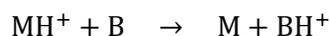
The computed  $\Delta H_{\text{rxn}}$  value in Scheme 2.5 reflects the theoretical difference in deprotonation enthalpy between the peptide and the reference acid (eq 2.16). The gas-phase thermochemical values of

ethanethiol have been well-established, with reported deprotonated enthalpy of  $355.7 \pm 2.1$  kcal/mol and gas-phase acidity of  $348.9 \pm 2.0$  kcal/mol (Janousek, Reed, & Brauman, 1980; Linstrom, 2005). As a result, rearrangement of eq 2.16 to eq 2.17 allows the determination of the theoretical deprotonation enthalpy ( $\Delta_{\text{acid}}H_{\text{pep}}$ ) of the peptide using the computed  $\Delta H_{\text{rxn}}$  value and the reported  $\Delta_{\text{acid}}H_{\text{Ref}}$  value.

$$\Delta H_{\text{rxn}} = \Delta_{\text{acid}}H_{\text{pep}} - \Delta_{\text{acid}}H_{\text{Ref}} \quad (\text{eq 2.16})$$

$$\Delta_{\text{acid}}H_{\text{pep}} = \Delta H_{\text{rxn}} + \Delta_{\text{acid}}H_{\text{Ref}} \quad (\text{eq 2.17})$$

Similarly, the weighted average enthalpy and free energy values were used for the basic peptides to calculate the theoretical protonation enthalpy or proton affinity (PA) and gas-phase basicity (GB) by using the isodesmic reaction shown in Scheme 2.6.



$$\Delta H_{\text{rxn}} = (\Delta H_{\text{M}} + \Delta H_{\text{BH}^+}) - (\Delta H_{\text{MH}^+} + \Delta H_{\text{B}})$$

$$\Delta H_{\text{rxn}} = (\Delta H_{\text{M}} - \Delta H_{\text{MH}^+}) - (\Delta H_{\text{B}} - \Delta H_{\text{BH}^+})$$

$$\Delta H_{\text{rxn}} = (\Delta H_{\text{H}^+} + \Delta H_{\text{M}} - \Delta H_{\text{MH}^+}) - (\Delta H_{\text{H}^+} + \Delta H_{\text{B}} - \Delta H_{\text{BH}^+})$$

Scheme 2.6. Isodesmic reaction scheme to determine the theoretical proton affinity, PA.

The M and B represent the neutral peptide and reference base, and  $\text{MH}^+$  and  $\text{BH}^+$  represent the protonated peptide and protonated reference base, respectively. In this reaction, the proton is transferred from the protonated peptide to the neutral reference base, with a well-established PA value. The reference base used was ethylamine ( $\text{CH}_3\text{CH}_2\text{NH}_2$ ) with a reported proton affinity of 218.0 kcal/mol and gas-phase basicity of 210.0 kcal/mol (Hunter & Lias, 2009; Linstrom, 2005).

In this case, the  $\Delta H_{\text{rxn}}$  is the difference between the PA of the peptide and the PA of the reference base (eq 2.18). By inputting Boltzmann averaged  $\Delta H$  values obtained for the neutral and protonated molecules,  $\Delta H_{\text{rxn}}$  can be determined. Rearrangement of eq 2.18 to eq 2.19 allows the determination of the PA value of the peptide using the computed  $\Delta H_{\text{rxn}}$  and the PA value of the reference base.

$$\Delta H_{\text{rxn}} = PA_{\text{M}} - PA_{\text{B}} \quad (\text{eq 2.18})$$

$$PA_{\text{M}} = \Delta H_{\text{rxn}} + PA_{\text{B}} \quad (\text{eq 2.19})$$

### Calculating the Dot Products and Vector Angles

For structural analysis and description of the peptide species, the dot product between dipole moments is used to quantitatively evaluate their interaction, and the vector angles between diatomic functional groups and dipole moments are used to demonstrate their alignment and relative orientation.

Firstly, the three-dimensional coordinates of the involving functional group vectors are determined by subtracting the coordinates of the starting atom from the corresponding coordinates of the terminal atom. Dipole moment is intrinsically a vector, and its three-dimensional magnitudes are its coordinates in the three perpendicular dimensions.

The dot product between two three-dimensional vectors is the sum of the products of their corresponding coordinates, as shown in Scheme 2.7.

$$\vec{a} = [a_1, a_2, a_3]$$

$$\vec{b} = [b_1, b_2, b_3]$$

$$\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2 + a_3b_3$$

Scheme 2.7. Dot product of two three-dimensional vectors  $\vec{a}$  and  $\vec{b}$ .

Then, the angle between two vectors is determined based on the geometric definition of the dot product. Based on the cosine formula, the dot product of two vectors can be calculated using their magnitudes and the vector angle between them, as shown in Scheme 2.8.

$$\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

$$\|\mathbf{b}\| = \sqrt{b_1^2 + b_2^2 + b_3^2}$$

$$\vec{\mathbf{a}} \cdot \vec{\mathbf{b}} = \|\mathbf{a}\| \times \|\mathbf{b}\| \times \cos \theta$$

Scheme 2.8. Geometric definition of the dot product between two three-dimensional vectors  $\vec{\mathbf{a}}$  and  $\vec{\mathbf{b}}$ .

The  $\|\mathbf{a}\|$  and  $\|\mathbf{b}\|$  denote the length of the vectors  $\vec{\mathbf{a}}$  and  $\vec{\mathbf{b}}$ , and the  $\theta$  is the angle between  $\vec{\mathbf{a}}$  and  $\vec{\mathbf{b}}$ . So, the cosine value of the angle between the two vectors is determined by dividing their dot product by their magnitudes, as shown in eq 2.20. The angle between two vectors is then determined by calculating the inverse cosine.

$$\cos \theta = \frac{\vec{\mathbf{a}} \cdot \vec{\mathbf{b}}}{\|\mathbf{a}\| \times \|\mathbf{b}\|} \quad (\text{eq 2.20})$$

### Charge Analysis

The charged peptides are the experimentally detected and measured species in the mass spectrometry measurement, including the IRMPD and ion mobility experiments. Their stability plays an important role in the observed gas-phase acidities/basicities of peptides. For the peptide systems of interest, protonation, or deprotonation of the sidechain of the amino acid residue concentrates the charge in the protonated or deprotonated moieties, i.e., the  $\text{NH}_3^+$  group in lysine and Dap and the thiolate group ( $\text{S}^-$ ) in cysteine.

To accommodate and diffuse the concentrated charge, hydrogen bonds are formed with the charged moiety. Hydrogen bonds are characterized by a considerable interpenetration of the isolated electronic clouds of the two moieties involved and a certain electron transfer between the two moieties (Alkorta, Rozas, & Elguero, 1998; Z. Zhang, Li, Jiang, & Wang, 2018). The hydrogen bond acceptor (HBA) atom shares its electron density with the hydrogen atom and reduces the atomic charge of the hydrogen (Baranović, Biliškov, & Vojta, 2012). A ‘hydrogen bond acceptor’ is also called an ‘electron donor’ and a ‘hydrogen bond donor (HBD)’ is also called an ‘electron acceptor’ (Alkorta et al., 1998).

Different degrees of accommodation of the charged moiety are reflected in their residue charges. Three types of atomic charges, including Mulliken, NPA, and Hirshfeld (Hirshfeld, 1977), are calculated on each atom of the charged center.

Single point energy calculation was performed on selected conformations at the B3LYP-D3(BJ)/ma-def2-TZVPP level of theory using the Gaussian 16' suite of programs. The generated fchk file contains the structure's high-quality electron density values in the real space. Along with that, natural population analysis (NPA) was also performed. The Mulliken and NPA charges were directly extracted from the calculation output. In the Hirshfeld charge analysis, the charge density at each point in real space is shared in proportion to their free-atom densities at the corresponding distances from the nuclei. The Hirshfeld atomic charge is defined as

$$q_A = - \int w_A(\vec{r}) \Delta\rho(\vec{r}) d\vec{r} \quad (\text{eq 2.21})$$

with

$$\Delta\rho(\vec{r}) = \rho(\vec{r}) - \sum_A \rho_A^0(\vec{r}_A), w_A(\vec{r}) = \frac{\rho_A^0(\vec{r}_A)}{\sum_A \rho_A^0(\vec{r}_A)} \quad (\text{eq 2.22})$$

Where  $\rho(\vec{r})$  is the actual electron density in a molecule,  $\rho_A^0(\vec{r}_A)$  is the spherical electron density of the free atom A.  $\sum_A \rho_A^0(\vec{r}_A)$  and  $\Delta\rho(\vec{r})$  defined in eq 2.22 are also known as promolecular density and deformation density, respectively. The Hirshfeld atomic weighting function  $w_A(\vec{r})$  equals the ratio of free atomic density  $\rho_A^0(\vec{r}_A)$  to promolecular density. With the fchk file from single point energy calculation, the Hirshfeld charge was calculated using the Multiwfn program version 3.8 developed by Tian Lu (Lu & Chen, 2012).

### Dipole Moment Decomposition

In helix conformations, the dipole moment of the molecular backbone can potentially form a stabilizing interaction with the charge on the protonated or deprotonated moiety. To examine this interaction, dipole moment decomposition was performed on the selected conformations. The dipole moment of a molecular fragment,  $D_F$ , composed of certain atoms in a system that is calculated by eq 2.23, where  $A \in F$  represents the atoms belonging to the molecular fragment,  $Z$  is the charge of the nucleus,  $R$

is the position of the nucleus,  $w$  is the Hirshfeld weight function of the atom,  $\rho$  is the electron density, and  $r$  represents the real space coordinates.

$$D_F = \sum_{A \in F} [Z_A R_A - \int w_A(r) \rho(r) r dr] \quad (\text{eq 2.23})$$

The Hirshfeld atomic weighting function for an atom has been defined in eq 2.22. The molecular dipole moment,  $\mu_{\text{whole}}$ , of the peptide species is decomposed into the dipole moment of the molecular fragment excluding the charged moiety, noted as the backbone dipole moment,  $\mu_{\text{backbone}}$ , and the dipole moment of the charged moiety. A built-in module in Multiwfn (Lu & Chen, 2012) was used to calculate the dipole moments of molecular fragments using the fchk file from single point energy calculation at B3LYP-D3(BJ)/ma-def-TZVPP level of theory.

### Torsional Strain Analysis

Different orientations of the lysine sidechain could have very different torsional strains, which significantly affect the stability of the conformation. The torsional strain on the amino acid sidechain is estimated manually based on the torsional strain function of n-Butane (Smith\* & Jaffe, 1996), as shown in eq 2.24.

$$T(\theta) = \sum_{n=1}^j 0.5M_n (1 - \cos(n(\theta + \pi))) \quad (\text{eq 2.24})$$

where  $T(\theta)$  is the torsional strain,  $\theta$  is the C-C-C-C dihedral angle,  $j=8$ ,  $M_1 = 1.1942$ ,  $M_2 = -0.3579$ ,  $M_3 = 3.4117$ ,  $M_4 = -0.2102$ ,  $M_5 = 0.2664$ ,  $M_6 = -0.14226$ ,  $M_7 = 0.10782$ , and  $M_8 = -0.08748$ .

The torsional strain function is also plotted in Figure 2.5.

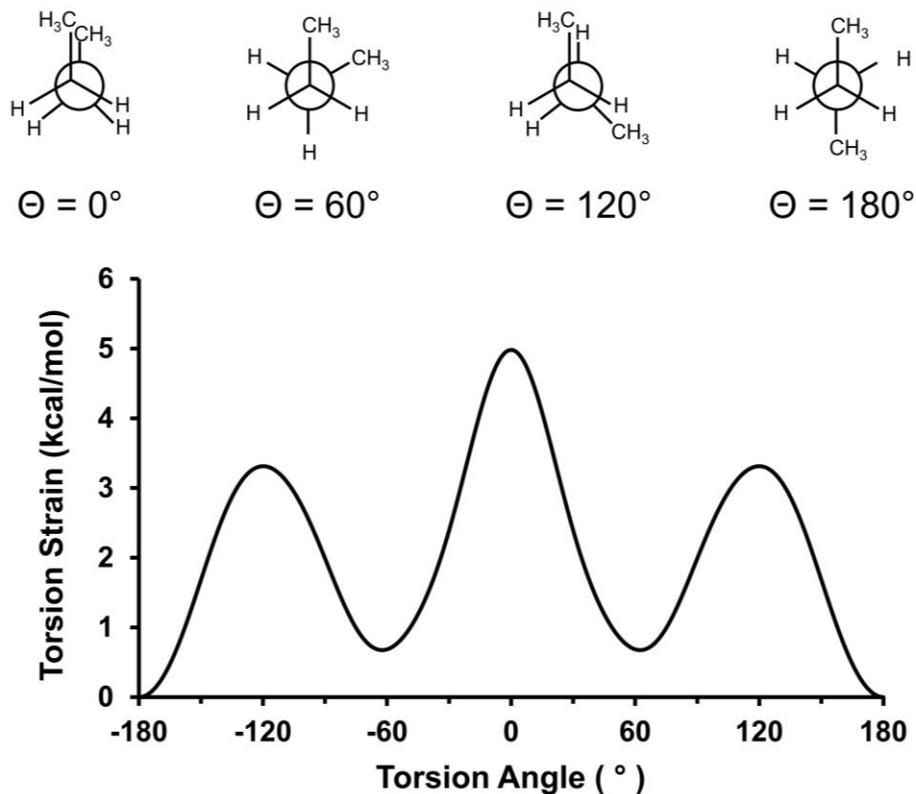


Figure 2.5. Torsional strain function for n-butane.

To calculate the torsional strain on an amino acid's sidechain, the dihedral angles made of only heavy atoms on the sidechain are obtained. The torsional strain of a certain dihedral angle is determined by plugging it into the torsional strain function as the torsion angle. Then the total torsional strain on the amino acid sidechain is calculated by summing the torsional strain of all dihedral angles.

### Hydrogen Bond Analysis

Hydrogen bonds (HBs) are important stabilizing interactions in both neutral and charged peptides and contribute to their different stabilities. One important feature distinguishing HBs from Van der Waals (vdW) interactions is their directional character which forces the bond to be as linear as possible, because linearity maximizes the charge-dipole and dipole-dipole interaction even in the weakest HBs (Batoon, Oomens, Berden, & Ren, 2018; Bertolasi, Gilli, & Gilli, 2011; Taylor & Kennard, 1982). For the peptide systems of interest, the angle in three-atom HBs (e.g.,  $S^- \cdots H-N$ ) and vector angle in four-atom

HBs (e.g., C=O···H–N) are used to characterize the linearity of the interacting groups. The vector angles were calculated based on the method described in section 2.5.3.

To quantitatively evaluate the conformational stability enhancement by H-bonds, the hydrogen bond energy (HBE) was estimated based on high-quality density functional theory wave functions obtained from single-point energy calculation at the B3LYP-D3(BJ)/ma-def2-TZVPP level of theory. In the work of Saeedreza Emamian, et al. (Emamian, Lu, Kruse, & Emamian, 2019), the HBE values of 42 intermolecular hydrogen-bonded dimer complexes with a diverse set of monomers were evaluated at very accurate CCSD(T) level with Truhlar’s “calendar” basis set jul-cc-pVTZ. A fairly satisfactory correlation was found between HBEs and the electron density at the bond critical point (BCP;  $\rho_{\text{BCP}}$ ) of HBs obtained at the B3LYP-D3(BJ)/ma-def2-TZVPP level of theory. The fitted equations for neutral complexes and charged complexes are shown in Figure 2.6. The mean absolute percentage error (MAPE), defined as that shown in eq 2.25, is used as the metrics to portray the magnitude of error in the established linear regressions

$$\text{MAPE} = \frac{\sum \left| \frac{Y_{\text{act}} - Y_{\text{pre}}}{Y_{\text{act}}} \right|}{n} \times 100\% \quad (\text{eq 2.25})$$

Where  $Y_{\text{act}}$  and  $Y_{\text{pre}}$  are actually observed (calculated) data and predicted data, respectively, while  $n$  is the number of samples. The MAPE value of the fitted equations for the neutral complexes and the charged complexes are 14.7 % and 10.0 %, respectively.

With the fchk file from single-point energy calculation at the B3LYP-D3(BJ)/ma-def2-TZVPP level of theory, the  $\rho_{\text{BCP}}$  of HBs in selected conformations of the peptide species were calculated using the Multiwfn program (Lu & Chen, 2012). The energy of H-bonds with the charged moiety was estimated using the fitted equation for the charged complexes, and the energy of other intramolecular H-bonds was estimated using the fitted equation for the neutral complexes. The total hydrogen bond energy of a structure was obtained by summing the HBE of its individual H-bonds.

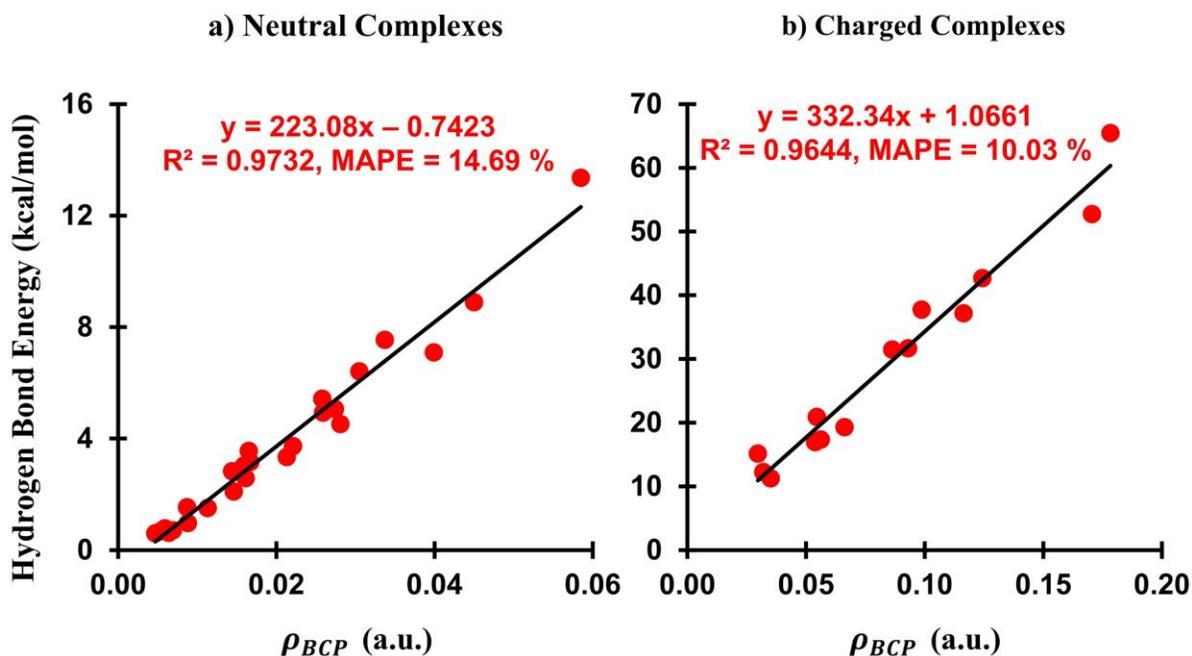


Figure 2.6. Regression plot of hydrogen bond energy (HBE) versus  $\rho$  at the BCP of the hydrogen bond for a) the neutral complexes, and b) the charged complexes. (Reproduced from Figure 3 of Emamian et al., 2019)

### RDG Analysis

Covalent bonds are clearly defined by the three-dimensional molecular structure and can be viewed and analyzed easily. Contrarily, noncovalent interactions, which govern many areas of biology and chemistry, are hidden within voids in the bonding network and are hard to be analyzed. The reduced-density gradient (RDG) approach developed by Erin, R. Johnson, et al. in 2010 (Johnson et al., 2010), has been widely used to detect, visualize, and analyze the complex noncovalent interactions, including van der Waals interactions, hydrogen bonds, and steric repulsions, and was used for the analysis of the peptide systems in this research. The reduced-density gradient (RDG) is a real space dimensionless quantity defined as a function of the electron density and its first derivative, as shown in eq 2.26, where  $\rho$  is the electron density,  $\nabla$  is the gradient operator, and  $|\nabla\rho|$  is the modulus of the electron density gradient.

$$\text{RDG} = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla\rho|}{\rho^{4/3}} \quad (\text{eq 2.26})$$

The  $\rho$  and RDG values in the noncovalent interaction regions are characteristically different from regions around the nucleus, near the chemical bond, and on the molecular edge. Comparatively, the RDG values around the nucleus and on the molecular edge are larger than those in the noncovalent interaction regions. The  $\rho$  values in the noncovalent interaction regions indicate the strength of interaction and are smaller than that in the regions around the nucleus and chemical bonds, but larger than that in the regions on the molecular edge. So based on the RDG and  $\rho$  values, the region and strength of noncovalent interactions in the molecule could be determined.

To distinguish between different types of noncovalent interactions, the sign of  $\lambda_2$ ,  $\text{sign}(\lambda_2)$ , which is the sign of the second eigenvalue of the electron-density Hessian (second derivative) matrix,  $\nabla^2\rho$ , is used in the RDG analysis. In the atoms-in-molecules (AIM) theory, the  $\text{sign}(\lambda_2)$  is negative at the bond critical point (BCP) and in the region around it. Contrarily, the  $\text{sign}(\lambda_2)$  is positive at the critical point of rings or cages and in the surrounding regions. Analysis of the  $\text{sign}(\lambda_2)$  thus helps to discern between different types of noncovalent interactions, whereas the density itself provides information about their strength.

To visualize and semi quantitatively analyze different noncovalent interactions, the RDG vs.  $\text{sign}(\lambda_2)\rho$  scatter plot, and the RDG isosurface diagram are generated correspondingly. With the fchk file from single-point energy calculation at the B3LYP-D3(BJ)/ma-def2-TZVPP level of theory, the RDG and  $\text{sign}(\lambda_2)\rho$  values were at each grid point of a cuboid containing the molecule were calculated using the Multiwfn program. The generated scatter points file was used to create the RDG scatter plot using Gnuplot, (T. Williams & Kelley, 2010) and the cube files were used to create the RDG isosurface diagram using pymol through a batch script written by Yuanhe Li from Peking University (Yuanhe Li, n.d.). An example of RDG scatter plot and RDG isosurface diagram is shown in Figure 2.7.

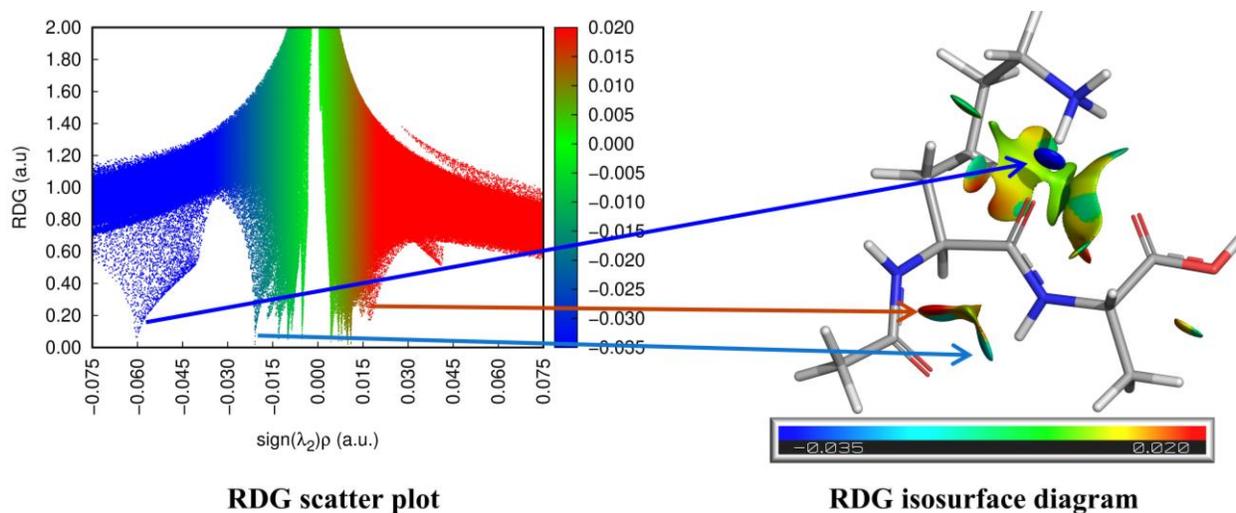


Figure 2.7. Sample RDG scatter plot (left) and the corresponding isosurface diagram (right).

The RDG values are plotted against the  $\text{sign}(\lambda_2)\rho$  values to generate the RDG scatter plot, where the points are colored on a blue-green-red scale according to the  $\text{sign}(\lambda_2)\rho$ , ranging from  $-0.035$  to  $+0.020$  a.u. An RDG spike approaching the x-axis represents an identified noncovalent interaction in the molecule. Blue spikes indicate strong attractive interactions, like hydrogen bonds. Red spikes indicate strong nonbonded overlaps, like steric clashes in rings and cages. The absolute value of the spike's x coordinate is the electron density at the bond critical point,  $\rho_{\text{BCP}}$ , of the corresponding interaction.

To visualize the noncovalent interactions in real-space, RDG isosurface diagrams are created by connecting grid points with an RDG value of 0.5 a.u. in the molecule. The RDG isosurface is colored the same way as the RDG scatter plot, where blue indicates strong attractive interactions and red indicates strong steric clashes. Therefore, the identified noncovalent interactions from the RDG scatter plot could be correlated to the regions of real space on the RDG isosurface diagram. The RDG isosurface diagram provides a much richer visualization of noncovalent interactions than simple pairwise contacts between atoms.

### IR Matching and Vibrational Mode Decomposition

After the final geometry optimization at wB97XD/6-311G\*\* level of theory, the frequency calculation was conducted at the same level of theory to yield accurate wavenumber and absorption (as

epsilon coefficient) of molecular vibrational modes. By applying a convolution function, a theoretical IR spectrum can be generated from the stick spectrum containing the absorption at very specific wavenumbers and aid in the interpretation process. The IR spectra were generated using GaussView 5.0.8 (Gaussian Inc., Wallingford, CT). A broadening parameter of  $15\text{ cm}^{-1}$  at half-width half-max (HWHM) was found to create a resolution that best matches the experimental spectra. After broadening was applied, the theoretical IR spectra were exported with an increment of  $5\text{ cm}^{-1}$ .

Because of the harmonic approximations in quantum chemical calculations, a scaling factor in the range of 0.94 to 0.96 was applied to the theoretical spectra, and then they were plotted and overlaid onto the experimental spectra. Conformations with the theoretical spectra that most closely match the IRMPD spectrum were chosen as the structurally representative structures, and their spectra were used to assign the vibrational modes.

Interpretation of the vibrational spectra in terms of the atomic motions related to each observed signals provides valuable information about the chemical system (Teixeira & Cordeiro, 2018). It helps to elucidate the nature of the experimental and theoretical absorption spectra through vibrational mode analysis. For quantitative interpretation of the vibrational spectra, the vibrational mode automatic relevance determination (VMARD) method, developed by Filipe Teixeira et al. (Teixeira & Cordeiro, 2018), was used to decompose vibrational modes into weighted combinations of displacements along chemically intuitive internal coordinates.

The frequency calculation output file at the DFT level was fed to the VMARD analysis program, VibAnalysis (Teixeira, 2017), to decompose all the computed vibrational modes. Then, each of those internal coordinates is manually assigned to the corresponding normal mode of the functional groups. The normal mode of the internal coordinates with the highest weight is assigned as the primary mode.

### **Bond Order and Relaxed Force Constant**

In addition to vibrational mode decomposition, multiple bond strength descriptors such as bond length, Mayer bond order, and relaxed force constant were used to interpret observed and calculated vibrational frequencies. The Mayer bond order was calculated using the fchk file from single-point

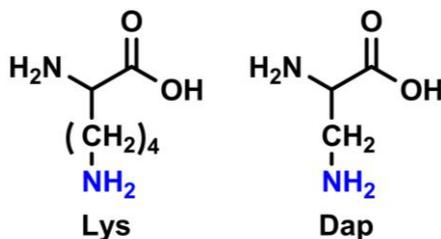
energy calculation at the B3LYP-D3(BJ)/ma-def2-TZVPP level of theory employing the Multiwfn program. The relaxed force constant was calculated using the frequency calculation output from Gaussian 09 utilizing the Compliance program version 3.0.2 (Brandhorst & Grunenberg, 2008, 2010).

### CHAPTER 3 : STRUCTURAL AND ENERGETIC CHARACTERIZATION OF PEPTIDES CONTAINING LYSINE AND HOMOLOGS

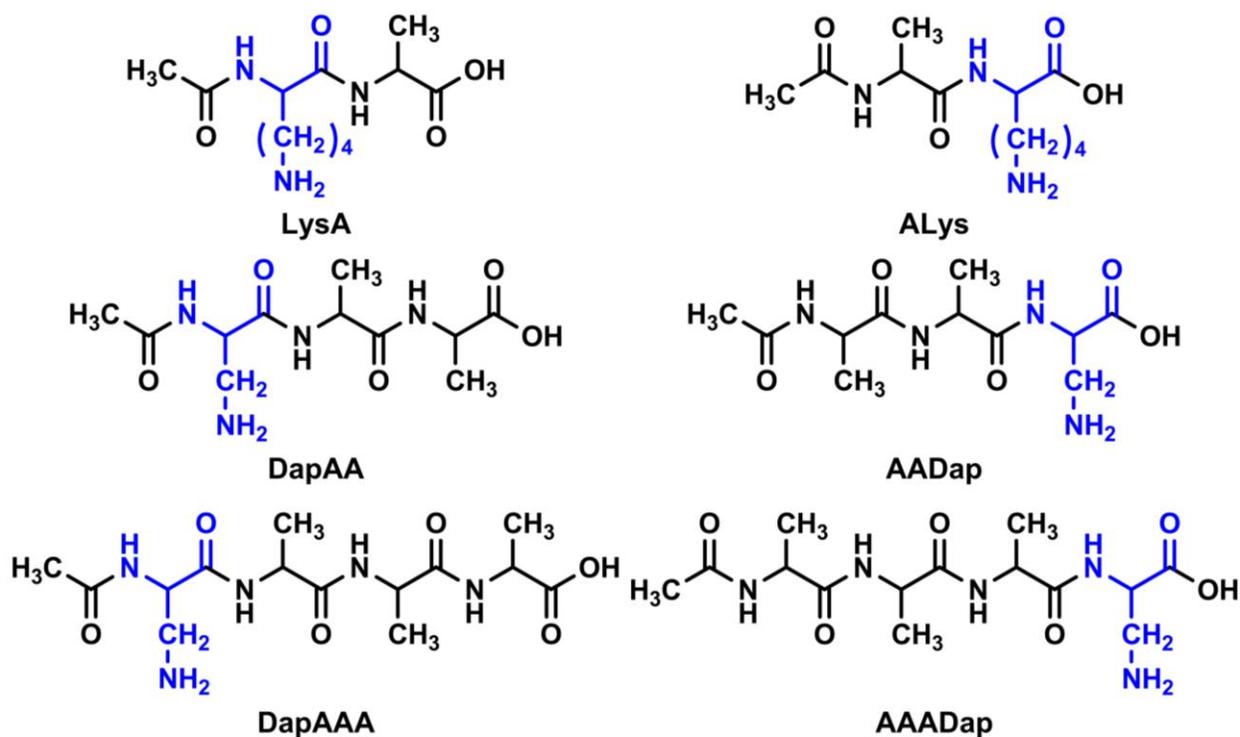
The unique biochemical and biophysical properties of lysine (Lys) mainly originate from its nucleophilic  $\epsilon$ -amino group, which allows it to participate in numerous biochemical reactions. The compound 2,3-diaminopropionic acid (Dap) is a Lys homolog amino acid, with a  $\beta$ -amino group instead of an  $\epsilon$ -amino group. When the amino acids are placed in different biological environments, their pKa values and charged state adjust accordingly (Isom et al., 2011). For example, Lys115 in the acetoacetate decarboxylase is positioned within the hydrophobic interior of the protein. Its pKa value was abnormally low at 7.5, whereas the normal pKa of Lys was 10.5. The reduced pKa value of Lys115 allowed the residue to retain its nucleophilic property as a neutral residue and readily form a Schiff-base intermediate with an incoming substrate at physiological pH (Ho et al., 2009). The exact factors stabilizing the buried ionizable residues are difficult to establish, and the adjusted pKa values are highly conformational dependent. For this reason, our understanding of the fundamental aspects of this phenomenon is limited. Model peptides are free from solvent interactions in the mass spectrometry conditions, which can be assumed to emulate similar conditions in protein active sites with minimum solvent accessibility (Batoon & Ren, 2016). The gas-phase acidity and basicity of oligopeptides can be used as a model to resemble the properties in the active sites of proteins.

In the previous work of our group, a series of model peptides were designed and characterized to study the effects of the position of the basic residue on the gas-phase thermochemistry and conformations of the peptides. The model peptides are composed of an alanine-based backbone, an acetylated N-terminus, and one basic residue as the basic probe. The N-terminus is acetylated to ensure one single protonation site at the side chain of the basic residue. The position of the basic probe is varied by placing it at either the N-terminus (N-Probe) or the C-terminus (C-Probe).

In a pair of dipeptides, the basic probe is Lys. In the other pair of tripeptides, the basic probe is 2,3-diaminopropionic acid (Dap). The structures of the amino acids and the peptides are shown in Scheme 3.1 and Scheme 3.2, respectively.



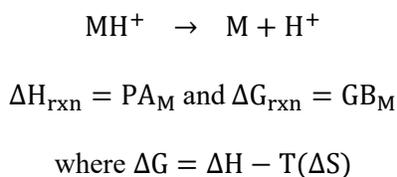
Scheme 3.1. Structures of lysine (left) and 2,3-diaminopropionic acid (right).



Scheme 3.2. Structures of the N-Probe (left) and C-Probe (right) peptides. A represents amino acid alanine.

These model peptides can be protonated by accepting a proton on the primary amino group of the lysine and Dap side chain. To quantitatively evaluate a peptide's ability to acquire a proton, the

thermochemical descriptors of proton affinity (PA) and gas-phase basicity (GB) are used. Proton affinity is the intrinsic basicity of a compound in the gas phase without entropic effects. Alternatively, gas-phase basicity is the Gibbs free energy change of the corresponding reaction taken into the effects of enthalpy and entropy. The definition of PA and GB is shown Scheme 3.3.



Scheme 3.3. Definition of Proton Affinity and Gas-phase Basicity.

High PA and GB values represent large energy release when protonating the neutral molecule, i.e., a high tendency to accept a proton and a large capability to stabilize the charge. The PA and GB values of the six model basic peptides have been determined utilizing the Cooks' extended kinetic method and are listed in Table 3.1 (Batoon, 2016). Interestingly, the C-Probe peptides have significantly higher PA and GB values than their isomeric N-Probe peptides, i.e., the C-Probe peptides are more basic.

This chapter focuses on the characterization of two pairs of isomeric peptides, LysA and ALys, and DapAA and AADap. Structural and energetic characterizations using the infrared multiphoton dissociation (IRMPD) and molecular modeling were carried out on those peptides to understand the observed basicity differences. Conformations that provide thermochemical values matching the experimental results were obtained through molecular modeling. Their structures, energetics, and intramolecular interactions were compared in detail to investigate the source of their basicity differences.

Table 3.1  
*Summary of Measured Thermochemical Values of Six Model Peptides*

Peptide <sup>a</sup>	PA kcal/mol	GB kcal/mol
LysA	241.5 ± 2.3	230.4 ± 2.3
ALys	246.3 ± 2.2	233.6 ± 2.2
DapAA	238.5 ± 2.2	224.2 ± 2.2
AADap	240.5 ± 2.2	229.5 ± 2.2

<sup>a</sup> Data taken from Batoon, 2016 and Batoon & Ren, 2016.

### LysA and ALys

From the mass spectrometry measurements (Table 3.1), the PA and GB values were determined to be  $241.5 \pm 2.3$  and  $230.4 \pm 2.3$  kcal/mol for LysA, and  $246.3 \pm 2.2$  and  $233.6 \pm 2.2$  kcal/mol for ALys. The PA and GB values of LysA are lower than those of ALys by 4.8 and 3.2 kcal mol<sup>-1</sup>, respectively. The differences between the two isomeric dipeptides are remarkable, considering their small sizes. The structural features in relation to the difference in basicity of these two peptides are characterized in detail by analyzing the experimental IRMPD spectra and computational studies.

### Features of the IRMPD Spectra of LysAH<sup>+</sup> and ALysH<sup>+</sup>

Experimental IRMPD spectra of protonated peptides, LysAH<sup>+</sup> and ALysH<sup>+</sup>, are shown in Figure 3.1. The two spectra are characteristically different from each other.

The C-terminal carboxyl C=O stretching band of LysAH<sup>+</sup> has a peak at 1743 cm<sup>-1</sup> and an unresolved shoulder peak at 1762 cm<sup>-1</sup>. The two peaks indicate the existence of multiple conformations with different microenvironments for the C=O group of the C-terminal carboxyl group. Conversely, this band of ALysH<sup>+</sup> has a single sharp peak at 1768 cm<sup>-1</sup>, suggesting its carboxyl C=O group in a single microenvironment. Moreover, the peaks of LysAH<sup>+</sup> are red-shifted by 6 to 25 cm<sup>-1</sup> compared to that of ALysH<sup>+</sup>. It is likely caused by stronger intramolecular interactions involving the carboxyl C=O group in LysAH<sup>+</sup>.

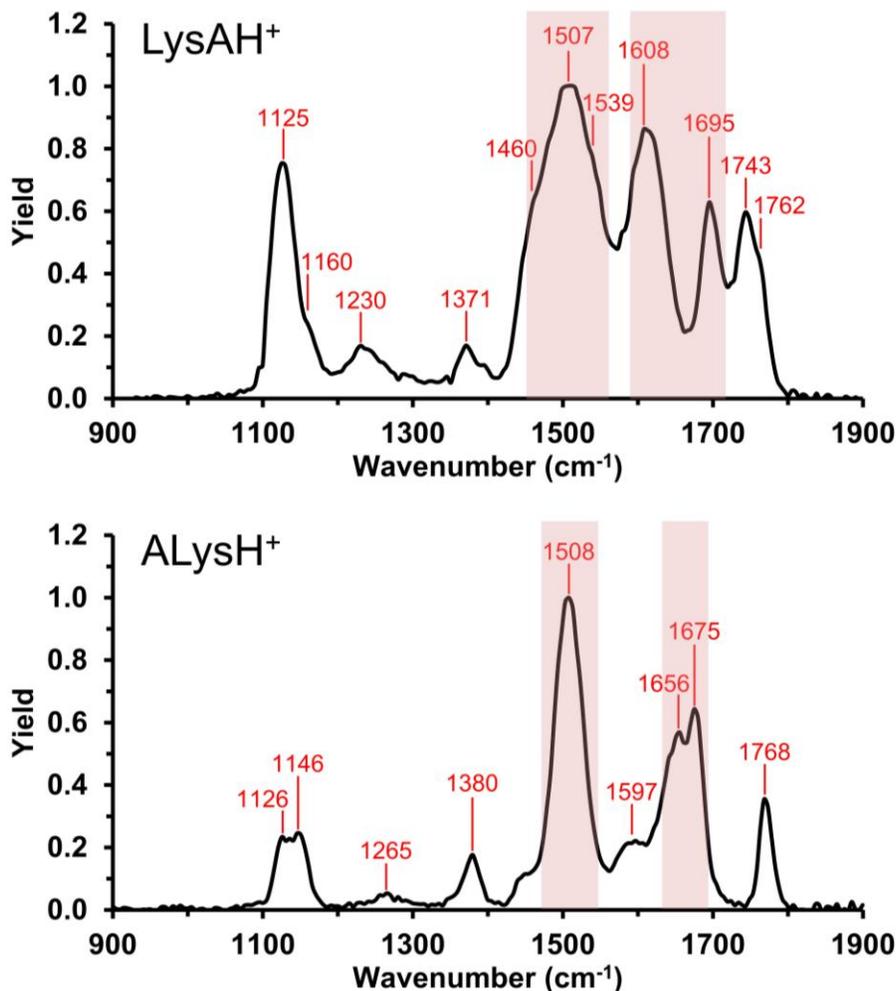
The amide I band of LysAH<sup>+</sup> has two peaks at 1608 and 1695 cm<sup>-1</sup> with a wide separation of 87 cm<sup>-1</sup>. The amide I band of ALysH<sup>+</sup> has two peaks at 1656 and 1675 cm<sup>-1</sup> with a small separation of 19

$\text{cm}^{-1}$ . The wide separation for  $\text{LysAH}^+$  indicates dramatically different microenvironments for the two amide bonds, while the small separation for  $\text{ALysH}^+$  indicates similar microenvironments for the two amide bonds.

The amide II band of  $\text{LysAH}^+$  consists of one broad peak centered at  $1507 \text{ cm}^{-1}$ , and two unresolved shoulder peaks at  $1460 \text{ cm}^{-1}$  and  $1539 \text{ cm}^{-1}$ , respectively. The three peaks further confirm the existence of multiple conformations with different microenvironments for the two amide bonds. The amide II band of  $\text{ALysH}^+$  has one sharp peak at  $1508 \text{ cm}^{-1}$ , indicating a similar microenvironment for the two amide bonds.

Compared to the carboxyl  $\text{C}=\text{O}$  stretching band and the amide I and II bands mentioned above, the low wavenumber bands from  $1000$  to  $1400 \text{ cm}^{-1}$  also provide valuable information but are generally less characteristic to their different structures. The absorption peaks at  $1230$  and  $1371 \text{ cm}^{-1}$  in the  $\text{LysAH}^+$  IRMPD spectrum and at  $1265$  and  $1380 \text{ cm}^{-1}$  in the  $\text{ALysH}^+$  IRMPD spectrum correspond to the amide III and the carboxyl  $\text{C}-\text{O}-\text{H}$  bending mode. The peaks around  $1150 \text{ cm}^{-1}$  correspond to the carboxyl  $\text{C}-\text{O}(\text{H})$  stretching mode.

In summary, the spectral features in the carboxyl  $\text{C}=\text{O}$  stretching region and the amide I and II regions suggest conformations with different intramolecular interaction networks of  $\text{LysAH}^+$ , but conformations with similar intramolecular interactions of  $\text{ALysH}^+$ .



*Figure 3.1.* IRMPD Spectra of LysAH<sup>+</sup> (top) and ALysH<sup>+</sup> (bottom) with major peaks labeled in wavenumber. (The spectral data were obtained from the FELIX Laboratory, Radboud University, The Netherlands. The peaks in the shaded area correspond to the amide I (high wavenumber) and amide II (low wavenumber) bands.)

### Computed PA and GB Values for LysA and ALys

To efficiently explore the conformational space and search for the most favorable conformations of the neutral and protonated peptides, the CREST (Pracht et al., 2020) and systematic conformational search were used at the conformational search step. Detailed setup and procedures of the two approaches can be found in Chapter 2. Compared to the CREST approach (energy calculated at the GFN2-xTB level), the systematic approach found conformations with lower energy by 0.1 and 1.1 kcal/mol for the neutral species of LysA and ALys. For the protonated species, the systematic and CREST approach

found the same lowest energy conformations. After combining the output data from both approaches, duplicate and high-energy conformations were removed to obtain the final ensemble of conformations. In total, 821 to 1,667 conformations for the neutral species and 71 to 95 conformations for the protonated species were generated.

The conformations obtained were optimized at the HF/3-21G\* level of theory and ranked in free energy. Conformations within the energy windows of 4 and 6 kcal/mol for the neutral and protonated species were kept for the next step. The resulting 44 to 97 conformations for the neutral species and 21 to 34 conformations for the protonated species were geometry optimized and frequency calculated at the wB97XD/6-311G\*\* level of theory. After removing the degenerate and high-energy conformations, a final set of ten conformations for each peptide species is used to compute the theoretical thermochemical values based on the equations described in Chapter 2.

Conformations with a good match to the corresponding IRMPD spectrum were selected as structurally representative conformations of the protonated peptide. Their vibrational modes were decomposed into a weighted combination of displacement along chemically meaningful internal coordinates using VibAnalysis (Teixeira & Cordeiro, 2018). Each of those internal coordinates is assigned to a normal mode of the corresponding functional group. The normal mode with the highest weight is appointed as the primary mode. Additionally, the relaxed force constants of the amide bonds were calculated to reveal the strength of the corresponding bonds, which helps explain their IR features.

Finally, single point energy calculation at B3LYP-D3(BJ)/ma-def2-TZVPP level of theory was performed on the selected representative conformations to obtain their high-quality electron density values in the real space, which is required for the subsequent structural and energetic analyses. To characterize the stabilization of the positive charge on the  $\text{NH}_3^+$  group, charge analysis, dipole moment decomposition, and hydrogen bonding analysis were performed. The torsional strain on the Lys and Dap sidechain was estimated and compared to understand the torsional pressure on their hydrocarbon side chain. Intramolecular interactions, including hydrogen bonds, Van Der Waals interactions, steric hindrance, and ring stress, were characterized and visualized through RDG analysis. Lastly, the Mayer

bond order of the amide bonds of interest was calculated to reflect the strength of corresponding bonds, in accordance with the relaxed force constants. Detailed explanations and procedures of the analyses mentioned above are described in Chapter 2.

The calculated energy values for all ten conformations of each peptide species are listed in Table 3.2, which includes the enthalpy/free energy, relative enthalpy/free energy, the Boltzmann probability, and the weighted average of enthalpy/free energy. The ten lowest energy conformations of the neutral and protonated peptides are noted as N1 to N10 for LysA, N1' to N10' for ALys, C1 to C10 for LysAH<sup>+</sup>, and C1' to C10' for ALysH<sup>+</sup>, respectively. The conformations are ranked in free energy. As shown in Table 3.2, the two neutral peptides have about the same enthalpy and free energy (within 0.4 kcal/mol), while LysAH<sup>+</sup> has higher enthalpy (by 4.2 kcal/mol) and higher free energy (by 3.5 kcal/mol) than ALysH<sup>+</sup>, suggesting lower stability of LysAH<sup>+</sup>.

Table 3.2  
Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for LysA and ALys

Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>	Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>
LysA						LysAH <sup>+</sup>					
N1	-896.662742	-896.734249	0.0	0.0	45.5%	C1	-897.053876	-897.122311	0.0	0.0	51.9%
N2	-896.664104	-896.733148	-0.9	0.7	14.2%	C2	-897.050805	-897.121169	1.9	0.7	15.5%
N3	-896.663642	-896.732939	-0.6	0.8	11.4%	C3	-897.053136	-897.12099	0.5	0.8	12.8%
N4	-896.66275	-896.732349	0.0	1.2	6.1%	C4	-897.051342	-897.120198	1.6	1.3	5.5%
N5	-896.663174	-896.732267	-0.3	1.2	5.6%	C5	-897.052562	-897.119968	0.8	1.5	4.3%
N6	-896.660989	-896.732115	1.1	1.3	4.7%	C6	-897.05102	-897.119507	1.8	1.8	2.7%
N7	-896.662922	-896.731799	-0.1	1.5	3.4%	C7	-897.051361	-897.119471	1.6	1.8	2.6%
N8	-896.661264	-896.731794	0.9	1.5	3.4%	C8	-897.050204	-897.119396	2.3	1.8	2.4%
N9	-896.661597	-896.731775	0.7	1.6	3.3%	C9	-897.049712	-897.118805	2.6	2.2	1.3%
N10	-896.661609	-896.731503	0.7	1.7	2.5%	C10	-897.0512	-897.118697	1.7	2.3	1.1%
H <sub>avg</sub> <sup>c</sup>	-896.662869					H <sub>avg</sub> <sup>c</sup>	-897.052799				
G <sub>avg</sub> <sup>c</sup>	-896.733301					G <sub>avg</sub> <sup>c</sup>	-897.121446				
Rel. H <sub>avg</sub> <sup>c</sup> 0						Rel. H <sub>avg</sub> <sup>c</sup> 0					
Rel. G <sub>avg</sub> <sup>c</sup> 0						Rel. G <sub>avg</sub> <sup>c</sup> 0					

<sup>a</sup> Values (kcal/mol) obtained from calculations at the wB97XD/6-311G<sup>\*\*</sup> level of theory.

<sup>b</sup> Boltzmann probability calculated using the equations described in Chapter 2.

<sup>c</sup> Weighted average of enthalpy and free energy (Hartree) of the ten conformations calculated using the equations described in Chapter 2.

<sup>d</sup> Relative enthalpy and free energy (kcal/mol) between LysA and ALys, and between LysAH<sup>+</sup> and ALysH<sup>+</sup>, calculated using the weighted average of H<sub>avg</sub> and G<sub>avg</sub>, respectively.

(Table 3.2 Continued)

Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>	Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>
ALys						ALysH <sup>+</sup>					
N1'	-896.664726	-896.733477	0.0	0.0	32.3%	C1'	-897.060596	-897.127643	0.0	0.0	73.4%
N2'	-896.662203	-896.733086	1.6	0.2	21.4%	C2'	-897.056336	-897.125437	2.7	1.4	7.1%
N3'	-896.660938	-896.732489	2.4	0.6	11.4%	C3'	-897.057733	-897.125436	2.0	1.4	7.1%
N4'	-896.663239	-896.732163	0.9	0.8	8.0%	C4'	-897.056966	-897.125143	2.3	1.6	5.2%
N5'	-896.663907	-896.732097	0.5	0.9	7.5%	C5'	-897.056409	-897.124735	2.6	1.8	3.4%
N6'	-896.663882	-896.731894	0.5	1.0	6.0%	C6'	-897.056859	-897.124328	2.3	2.1	2.2%
N7'	-896.660892	-896.73147	2.4	1.3	3.9%	C7'	-897.054301	-897.123141	4.0	2.8	0.6%
N8'	-896.662913	-896.731358	1.1	1.3	3.4%	C8'	-897.054702	-897.123109	3.7	2.8	0.6%
N9'	-896.662975	-896.731314	1.1	1.4	3.3%	C9'	-897.0554	-897.122369	3.3	3.3	0.3%
N10'	-896.660001	-896.731177	3.0	1.4	2.8%	C10'	-897.053094	-897.122202	4.7	3.4	0.2%
H <sub>avg</sub> <sup>c</sup>	-896.663124					H <sub>avg</sub> <sup>c</sup>	-897.059546				
G <sub>avg</sub> <sup>c</sup>	-896.732691					G <sub>avg</sub> <sup>c</sup>	-897.126948				
Rel. H <sub>avg</sub> <sup>c</sup>	-0.2					Rel. H <sub>avg</sub> <sup>c</sup>	-4.2				
Rel. G <sub>avg</sub> <sup>c</sup>	0.4					Rel. G <sub>avg</sub> <sup>c</sup>	-3.5				

<sup>a</sup> Values (kcal/mol) obtained from calculations at the wB97XD/6-311G\*\* level of theory.

<sup>b</sup> Boltzmann probability calculated using the equations described in Chapter 2.

<sup>c</sup> Weighted average of enthalpy and free energy (Hartree) of the ten conformations calculated using the equations described in Chapter 2.

<sup>d</sup> Relative enthalpy and free energy (kcal/mol) between LysA and ALys, and between LysAH<sup>+</sup> and ALysH<sup>+</sup>, calculated using the weighted average of H<sub>avg</sub> and G<sub>avg</sub>, respectively.

The weighted average of enthalpy and free energy was used to calculate the proton affinity and the gas-phase basicity of the peptides. Ethylamine ( $\text{CH}_3\text{CH}_2\text{NH}_2$ , PA = 218.0 kcal/mol, GB = 210 kcal/mol) (Hunter & Lias, 2009) was used as the reference base. By using the equation described in Chapter 2, the theoretical PA and GB values are computed as 241.3 and 232.0 kcal/mol for LysA and 245.4 and 235.8 kcal/mol for ALys (Table 3.3). The theoretical PA and GB values of LysA are lower than ALys by 4.1 and 3.8 kcal mol<sup>-1</sup>, respectively. The theoretical values agree with the experimental values and suggest a much higher basicity of ALys.

Table 3.3  
*Experimental and Calculated Thermochemical Values of LysA and ALys*

Peptide	PA <sub>expt</sub> kcal/mol	GB <sub>expt</sub> kcal/mol	PA <sub>calc</sub> kcal/mol	GB <sub>calc</sub> kcal/mol
LysA	241.5 ± 2.3	230.4 ± 2.3	241.3	232.0
ALys	246.3 ± 2.2	233.6 ± 2.2	245.4	235.8

The gas-phase basicity is associated with the stability of both neutral and protonated peptides. As indicated by the relative free energy values of (Rel. G<sub>avg</sub>) shown in Table 3.2, LysA and ALys are of similar stability, while ALysH<sup>+</sup> is more stable than LysAH<sup>+</sup> by 3.5 kcal/mol, suggesting that the stability difference in the protonated peptides is the main source of the greater basicity of ALys. The same conclusion could be obtained by comparing the relative enthalpy (Rel. H<sub>avg</sub>) values.

### Computed Structures and IR Spectra of LysA and ALys

**Structural features of LysA and ALys.** As shown in Table 3.2 and mentioned above, the stability of the neutral peptides is similar. Individual structures of the neutral peptides are shown in Figure 3.2 and Figure 3.3. The hydrogen bonds involving the lysine side chain primary amine are indicated in red dashed lines with the N --- H distance labeled. The overlays of the ten conformations of LysA and ALys are shown in Figure 3.4 to demonstrate their structural variations.

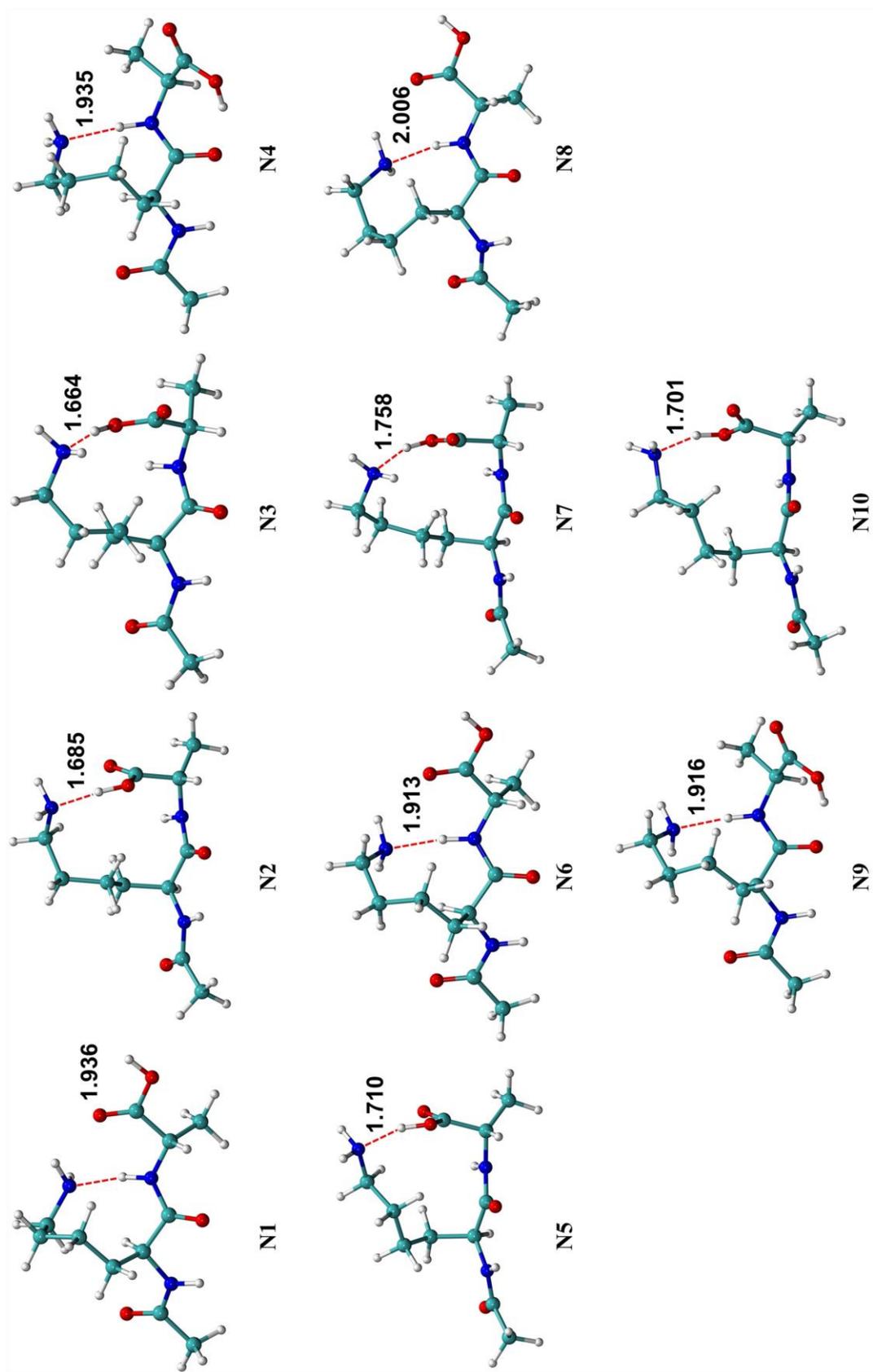


Figure 3.2. Computed geometries of LysA at the wB97XD/6-311G\*\* level of theory. (The dashed lines and the values indicate H-bonding.)

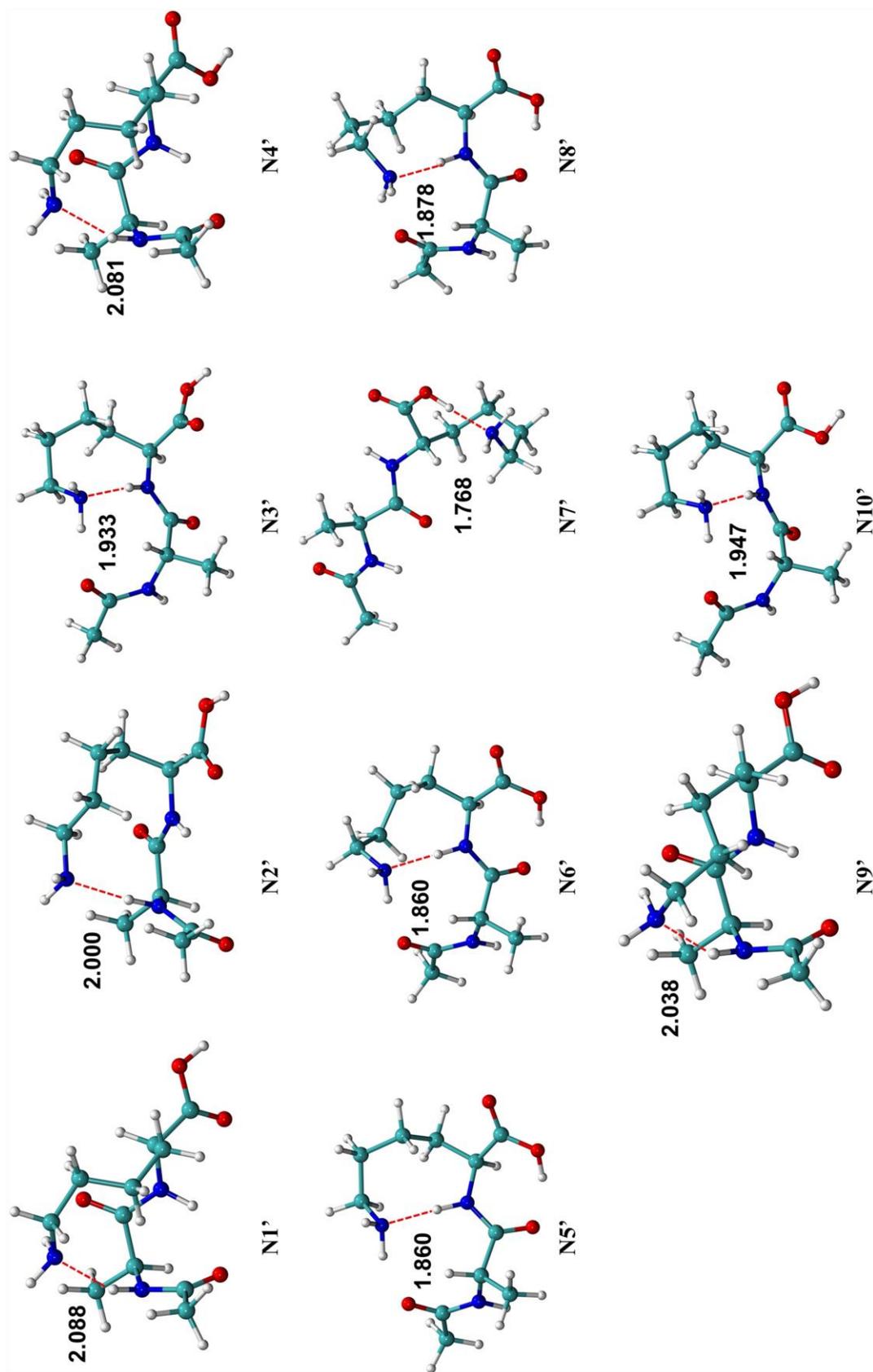
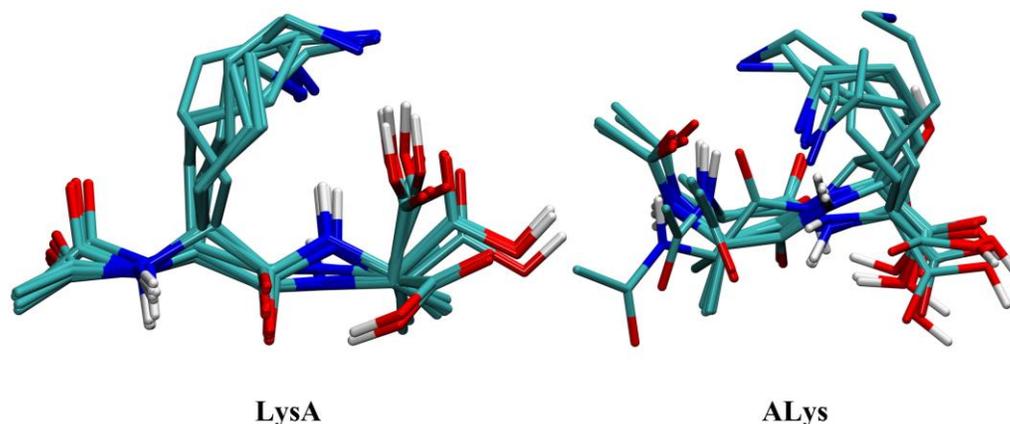


Figure 3.3. Computed geometries of ALys at the wB97XD/6-311G\*\* level of theory. (The dashed lines and the values indicate H-bonding.)



*Figure 3.4.* Overlay of the ten conformations of LysA(left) and ALys(right). (A linear backbone structure is highly conserved among the ten LysA conformations, while in ALys, there are multiple backbone geometries.)

In N2, N3, N5, N7, and N10 of LysA, the lysine sidechain primary amine forms a hydrogen bond with the C-terminal carboxyl OH group. In the other five conformations, N1, N4, N6, N8, and N9 of LysA, the lysine side chain forms a hydrogen bond with the alanine NH group. Because the alanine NH group and the carboxyl OH group are both at the C-terminal side, small orientational adjustments of the flexible lysine sidechain or carboxyl group allow the adoption of those two hydrogen bonding patterns without major variation of the peptide backbone (Figure 3.4).

Contrarily, large backbone variation is necessary for different hydrogen-bonding patterns in ALys. In N7' of ALys, the lysine sidechain primary amine accepts a hydrogen bond from the C-terminal carboxyl group. In N1', N2', N4', and N9' of ALys, it accepts a hydrogen bond from the alanine NH group. The last five conformations, N3', N5', N6', N8', and N10', accept the hydrogen bond from the lysine NH group. The flip of the Ac-Ala amide bond and large orientational adjustments of the lysine sidechain are observed between the ALys conformations, as shown in the overlaid structures.

**Structural features of LysAH<sup>+</sup> and ALysH<sup>+</sup>.** As suggested by the computational results, the dramatically different basicities of LysA and ALys are mainly because of the different stabilities of the protonated peptides, which originate from their different structures. The different features in the IRMPD spectra of LysAH<sup>+</sup> and ALysH<sup>+</sup> also reflect their structural differences. The individual structures of

LysAH<sup>+</sup> and ALysH<sup>+</sup> are shown in Figure 3.5 and Figure 3.7. The hydrogen bonds from the NH<sub>3</sub><sup>+</sup> group are highlighted as the red dashed lines with their H --- O distances labeled. The overlays of the ten conformations of LysAH<sup>+</sup> and ALysH<sup>+</sup> are shown in Figure 3.9 to demonstrate and compare their structural variations. After applying a scaling factor of 0.949 to the computed IR frequencies at the wB97XD/6-311G\*\* level of theory, the theoretical IR spectra from each conformation are overlaid onto the LysAH<sup>+</sup> and ALysH<sup>+</sup> IRMPD spectra shown in Figure 3.6 and Figure 3.8.

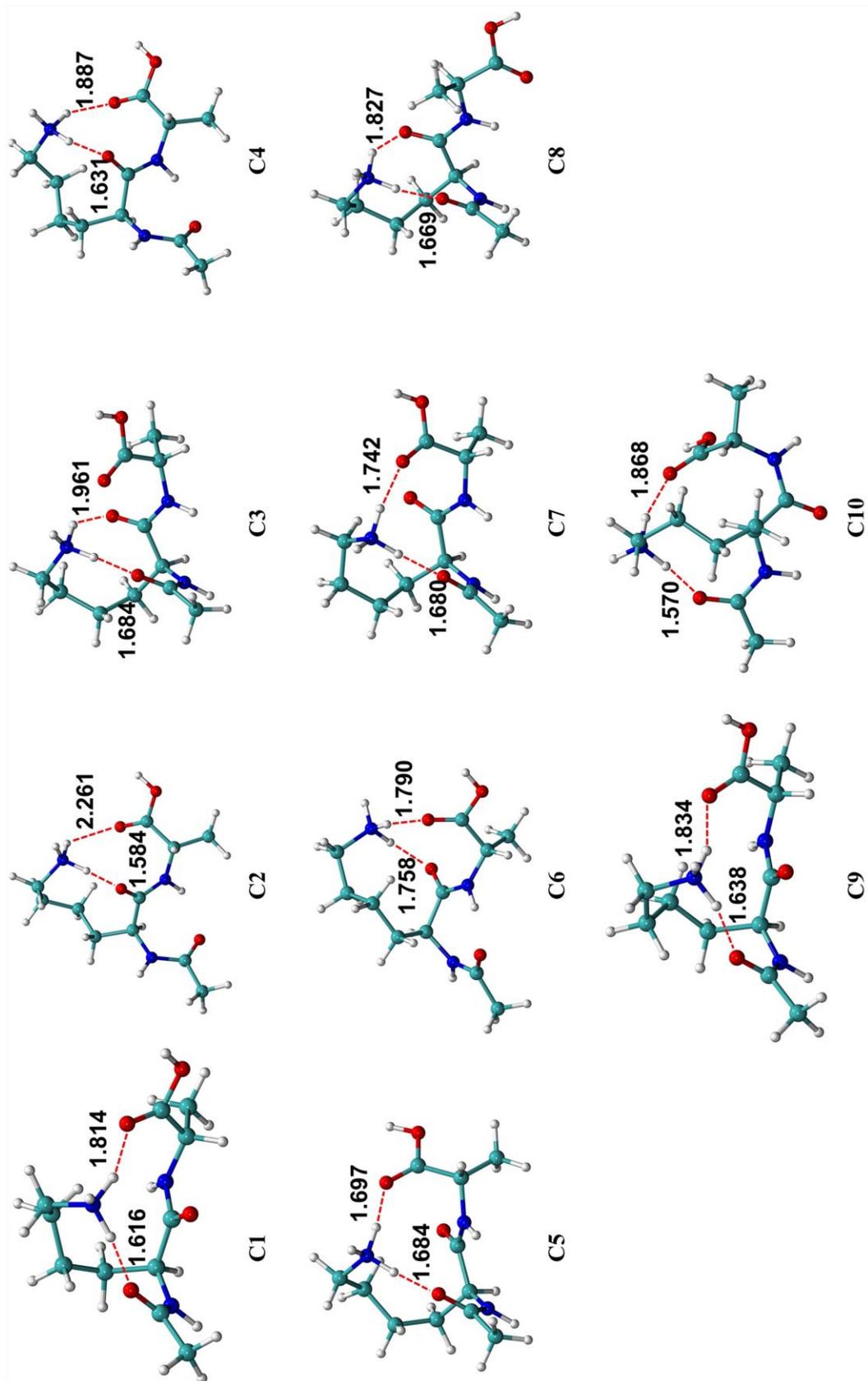


Figure 3.5. Computed geometries of LysAH<sup>+</sup> at the wB97XD/6-311G\*\* level of theory.

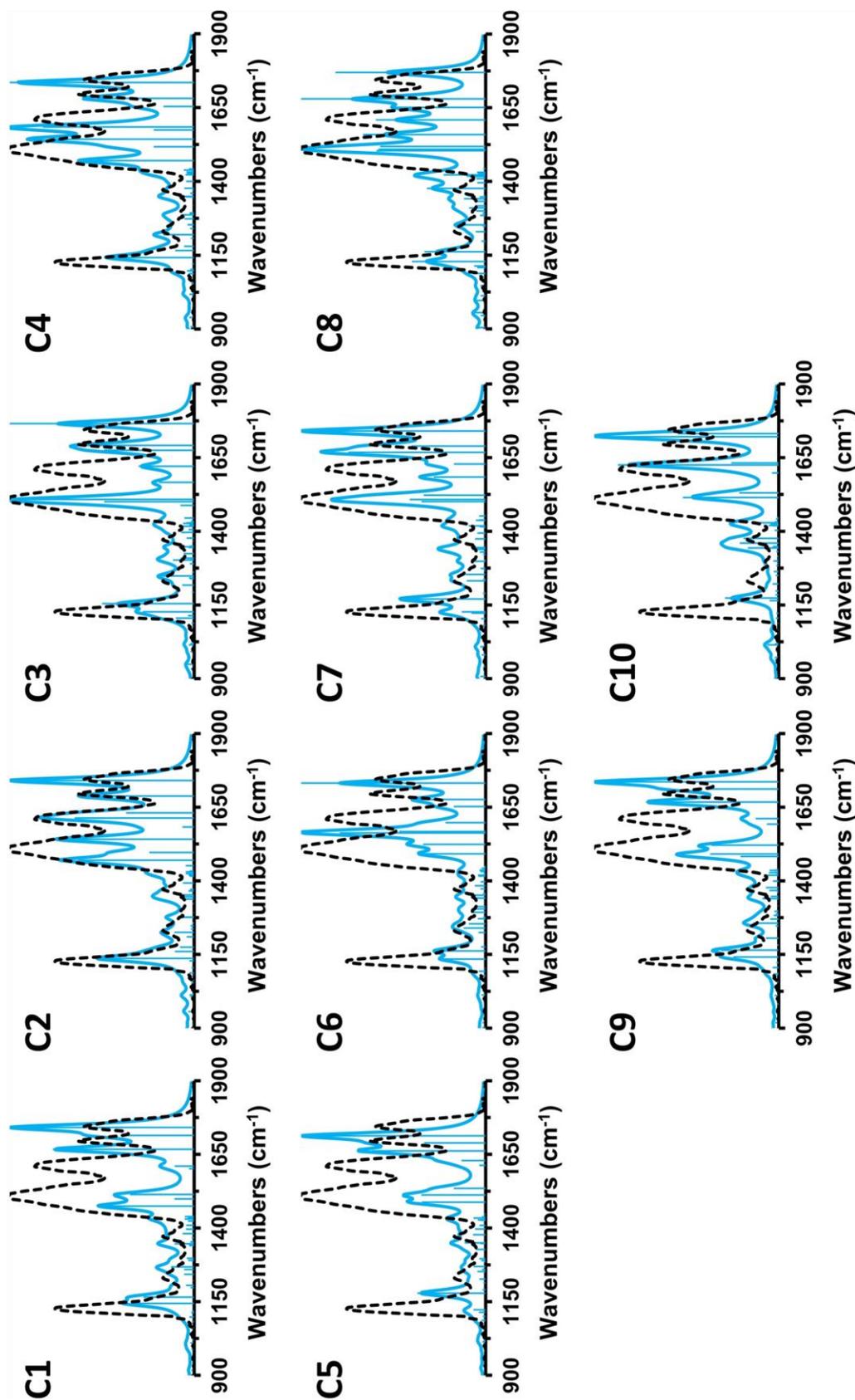


Figure 3.6. Computed IR stick and convoluted spectra (solid light blue) overlaid onto the IRMPD spectrum (dashed black) of LysAH<sup>+</sup>. (The IR frequencies were calculated at the wB97XD/6-311G\*\* level of theory and a scaling factor of 0.949 was applied.)

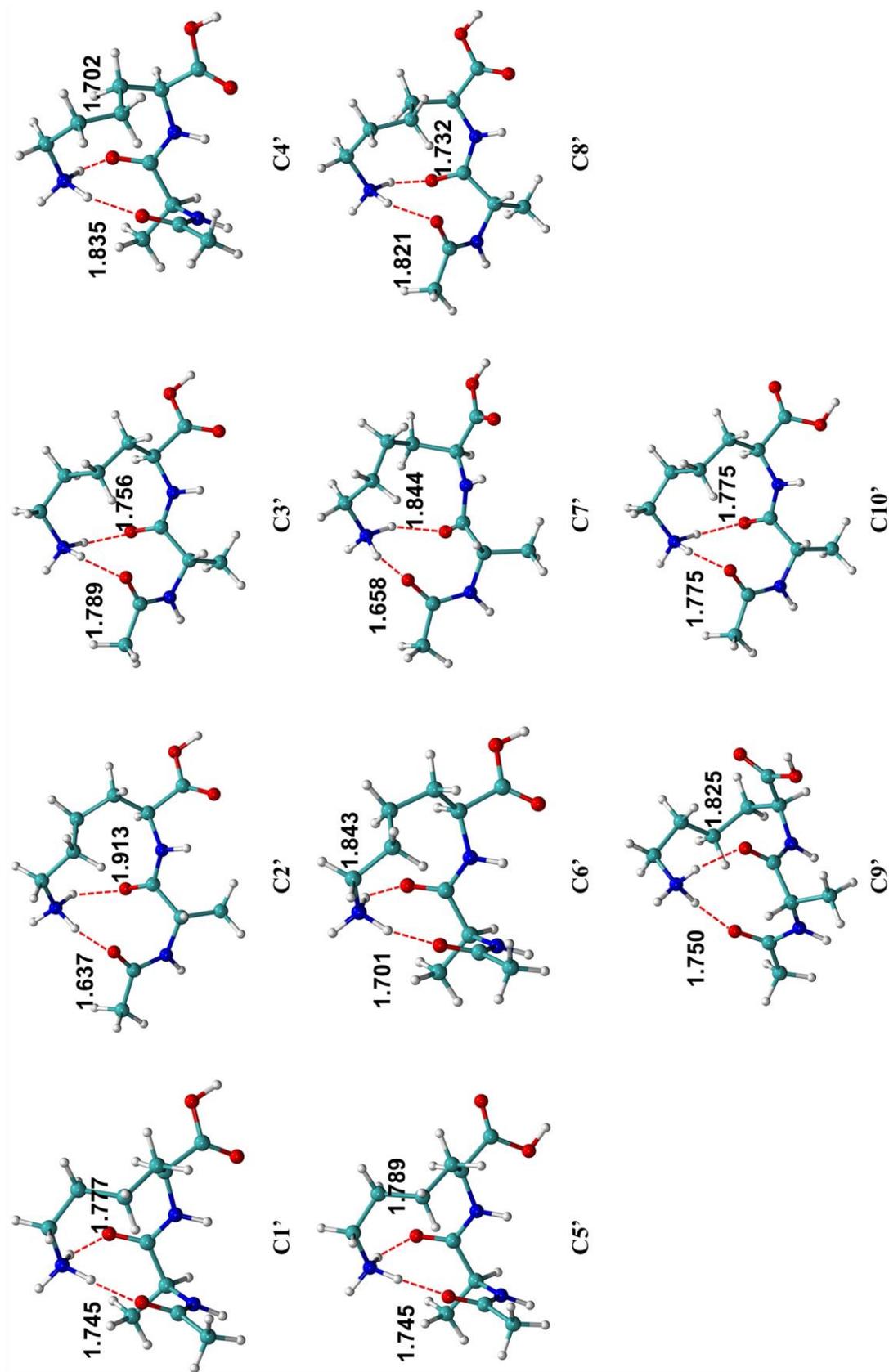


Figure 3.7. Computed geometries of ALysH<sup>+</sup> at the wB97XD/6-311G\*\* level of theory.

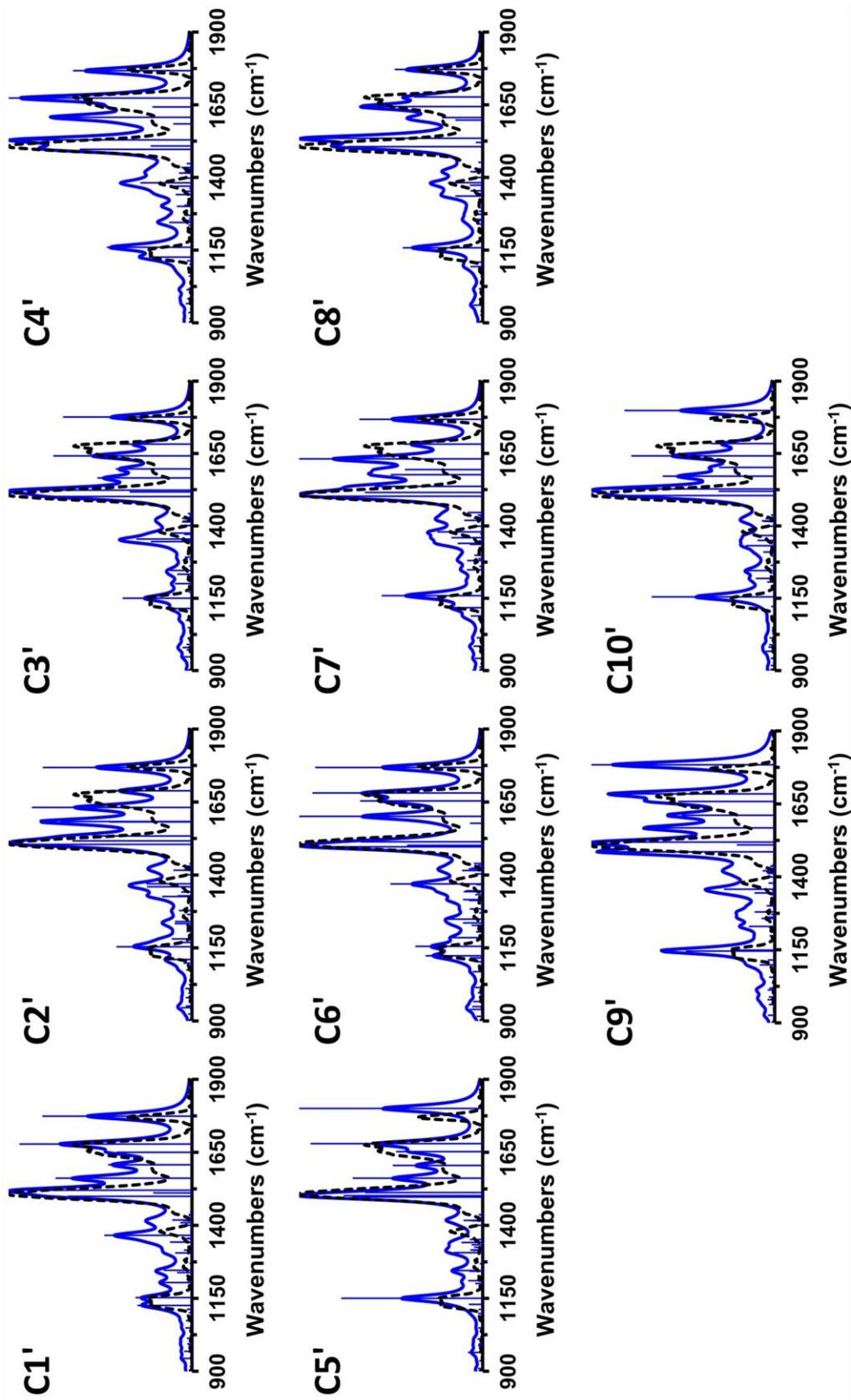


Figure 3.8. Computed IR stick and convoluted spectra (solid dark blue) overlaid onto IRMPD spectrum (dashed black) of  $ALysH^+$ . (The IR frequencies were calculated at the wB97XD/6-311G\*\* level of theory and a scaling factor of 0.949 was applied.)

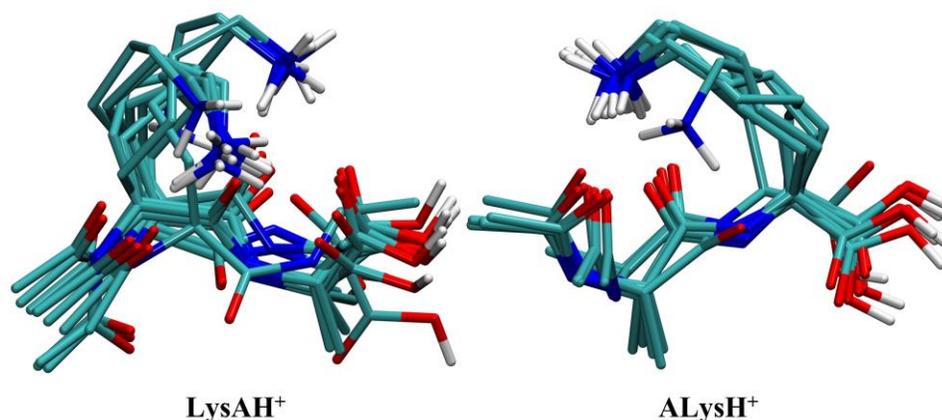


Figure 3.9. Overlay of the ten conformations of LysAH<sup>+</sup> (left) and ALysH<sup>+</sup> (right).

Unlike the monotonous LysA backbone, the conformations of LysAH<sup>+</sup> have diverse backbone structures, i.e., the diversity of LysA backbone geometry increases upon protonation. Contrarily, the backbone geometry of ALys becomes less diverse upon protonation. As shown in Figure 3.5 and Figure 3.9, in most LysAH<sup>+</sup> conformations, the carboxyl C=O group interacts with the NH<sub>3</sub><sup>+</sup> group. Conversely, in none of the ten ALysH<sup>+</sup> conformations, the carboxyl group interacts with the NH<sub>3</sub><sup>+</sup> group (Figure 3.7 and Figure 3.9).

In the first group of LysAH<sup>+</sup> conformations, C1, C5, C7, C9, and C10 (Figure 3.10), the NH<sub>3</sub><sup>+</sup> group forms hydrogen bonds with the Ac-Lys amide C=O group and the carboxyl C=O group. In C1 and C9, the Lys-Ala amide C=O group is placed below the NH<sub>3</sub><sup>+</sup> group, while in C5 and C7, it points away and towards the back of the molecule. The backbone orientation of C10 is in an inversion of C5 and C7 while maintaining the hydrogen-bonding patterns. Comparing the theoretical spectra of this group of conformations to the IRMPD spectrum of LysA<sup>+</sup> (Figure 3.6), poor matching is observed, suggesting the non-existence of this type of hydrogen bonding pattern.

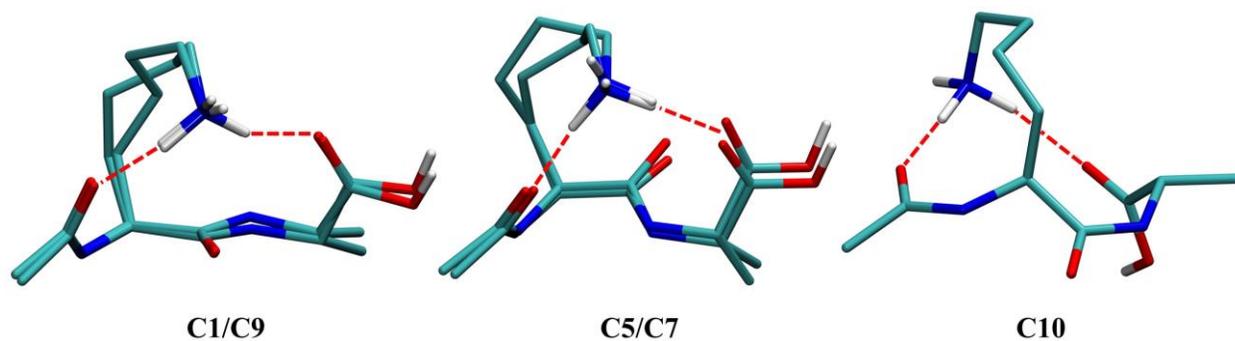


Figure 3.10. Structure of C1 and C9 (left), C5 and C7 (middle), and C10 (right). (The dashed lines indicate H-bonding.)

In the second group of LysAH<sup>+</sup> conformations, C2, C4, and C6 (Figure 3.11), the NH<sub>3</sub><sup>+</sup> group hydrogen bonds with the Lys-Ala amide C=O group and the carboxyl C=O group. In these three conformations, the two C=O bonds interacting with the NH<sub>3</sub><sup>+</sup> group are nearly perpendicular to each other. The C→O vector angle between the Lys-Ala amide C=O group and the carboxyl C=O group in C2, C4 and C6, are calculated and shown in Table 3.4. Detailed explanation and procedures of obtaining the C→O vector angle is described in Chapter 2. Briefly, for each carbonyl group, its C→O vector is obtained by subtracting the coordinates of the carbon from the coordinates of the oxygen. Based on the cosine formula for the vector dot product, the cosine value of the angle between two C→O vectors can be obtained through dividing their dot product by their magnitudes. As shown in Table 3.4, their C→O vector angles are all close to 90° and demonstrate the perpendicular geometry of these two C=O groups.

Comparing the theoretical spectra of this group of conformations with the IRMPD spectrum of LysAH<sup>+</sup> (Figure 3.6), conformation C2 has good match in the carboxyl C=O stretching region and the amide I and II regions, while conformation C4 has slightly worse match, and conformation C6 has no good match.

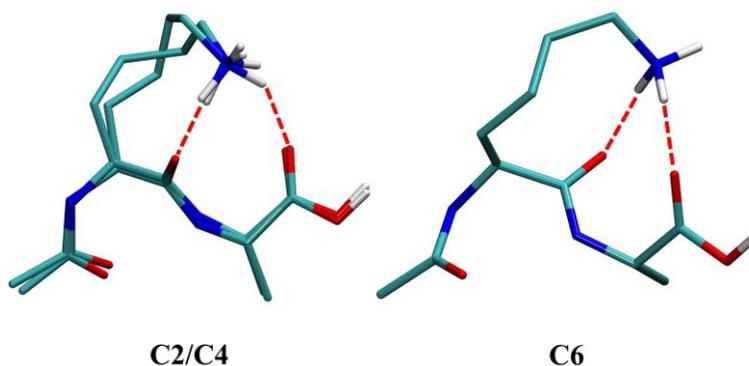


Figure 3.11. Structure of C2 and C4 (left), and C6 (right). (The dashed lines indicate H-bonding.)

Table 3.4

*C*→*O* Vector Angle between the Lys-Ala Amide C=O Group and the Carboxyl C=O Group in C2, C4, and C6

	<b>C→O vector angle (°)</b>
C2	82.4
C4	82.5
C6	72.4

In the last group of LysAH<sup>+</sup> conformations, C3, and C8 (Figure 3.12), the NH<sub>3</sub><sup>+</sup> group hydrogen bonds with the Ac-Lys and Lys-Ala amide C=O groups. In C3, the carboxyl C=O group points towards and stabilizes the NH<sub>3</sub><sup>+</sup> group. In C8, the carboxyl group and the NH<sub>3</sub><sup>+</sup> group are isolated from each other.

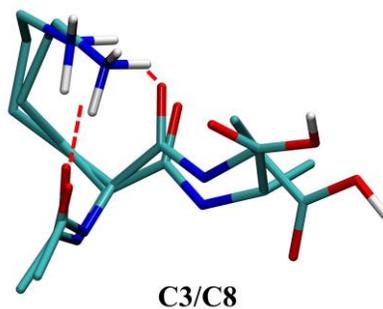


Figure 3.12. Structure of C3 and C8. (The dashed lines indicate H-bonding.)

Additionally, in C3 and C8, the backbone structure around the lysine residue is like in a left-handed helix, and the Ac-Lys and Lys-Ala C=O groups are nearly parallel to each other. Their dihedral angles across the lysine alpha carbon,  $\varphi$ : C(O)–N(H)–C $_{\alpha}$ (Lys)–C(O) and  $\psi$ : N(H)–C $_{\alpha}$ (Lys)–C(O)–N(H), and C→O vector angle between the Ac-Lys and Lys-Ala amide C=O groups are shown in Table 3.5.

Table 3.5

*Dihedral Angles across the C $_{\alpha}$  of Lysine and C→O Vector Angle between the Ac-Lys and Lys-Ala Amide C=O Groups in C3 and C8*

	$\varphi$ (phi) (°)	$\psi$ (psi) (°)	C→O vector angle (°)
	C(O)–N(H)–C $_{\alpha}$ (Lys)–C(O)	N(H)–C $_{\alpha}$ (Lys)–C(O)–N (H)	
C3	54.3	41.1	28.2
C8	44.8	52.0	37.7

Compared to the nearly perpendicular Lys-Ala amide C=O group and carboxyl C=O groups in C2, C4, and C6, the C→O vector angle between the Ac-Lys and Lys-Ala amide C=O groups in C3 and C8 are much smaller, demonstrating their good alignment with each other. To better illustrate their helical geometries, the dihedral angle pairs ( $\varphi$ ,  $\psi$ ) of C3 and C8 are overlaid to the Ramachandran plot in Figure 3.13. As shown, the dihedral angles of C3 and C8 are both positive, and their dihedral pairs both fall in the left-handed helix region. Comparing the theoretical spectra of this group of conformations to the IRMPD spectrum of LysAH<sup>+</sup> (Figure 3.6), both conformations provide good matching in the amide II region, while C3 provides better overall matching than C8.

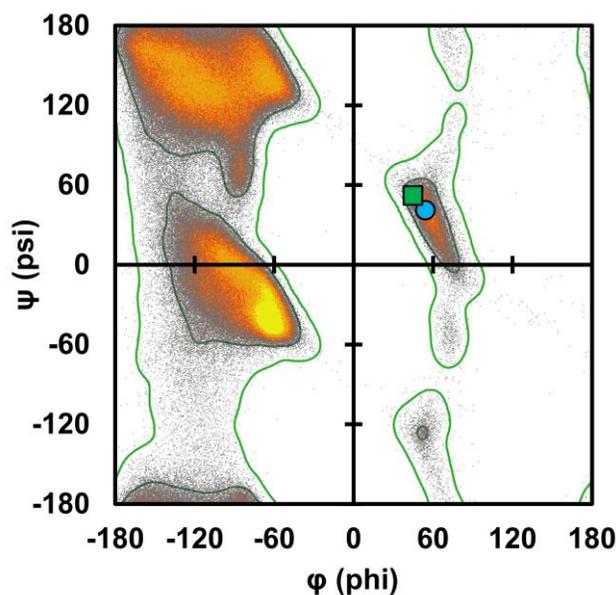
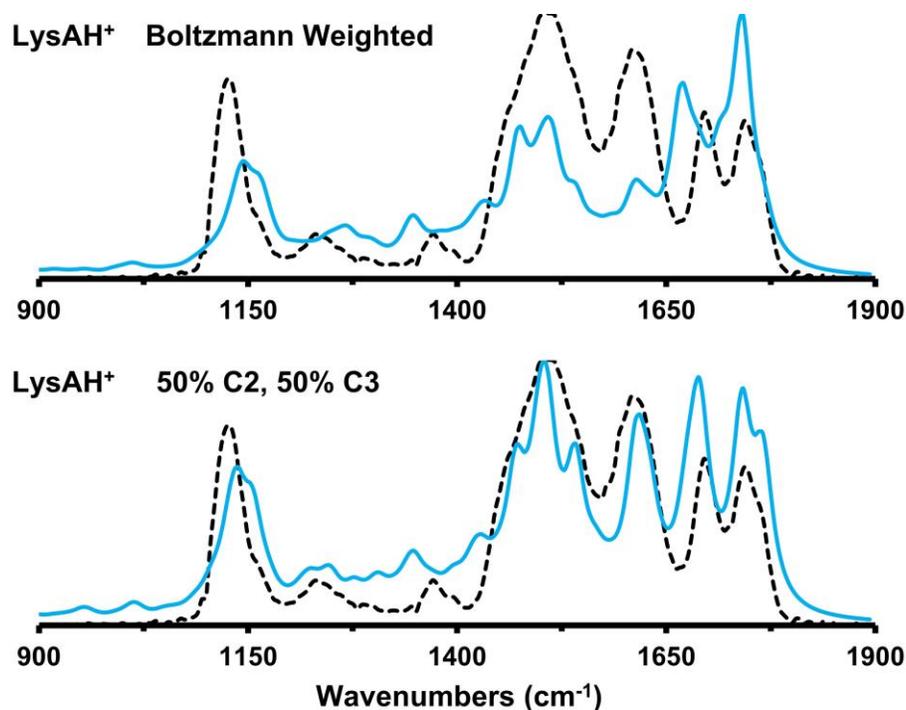


Figure 3.13. Ramachandran plot of C3 (light blue circle) and C8 (green square).

Based on the free energetics, the structure of LysAH<sup>+</sup> could be primarily represented by conformation C1, as it potentially contributes to 51% of the population of LysAH<sup>+</sup>. However, comparing its theoretical spectrum to the IRMPD spectrum of LysAH<sup>+</sup>, poor matching is observed. A Boltzmann weighted averaged spectrum (Figure 3.14) by applying the Boltzmann weights to the full series of calculated spectra does not provide reasonable match. Comparing the theoretical spectra of all ten conformations to the IRMPD spectrum of LysAH<sup>+</sup> (Figure 3.6), no single conformation could provide good match in all regions. However, the spectra of two distinct conformations, C2 and C3, match the IRMPD spectrum in the carboxyl C=O stretching and amide I regions. In the amide II region, a combination of both provides good match. The two computed peaks of C2 match the experimentally observed shoulder peaks at 1460 and 1539 cm<sup>-1</sup>, while the computed peak of C3 matches the center of the amide II band at 1507 cm<sup>-1</sup>. The structural distinction between C2 and C3 (Figure 3.5) is in agreement with the LysAH<sup>+</sup> IRMPD spectrum interpretation that multiple conformations with different intramolecular interaction networks are present. To demonstrate matching, a hand-weighted spectrum with C2 and C3 is shown in Figure 3.14, together with the Boltzmann weighted spectrum.



*Figure 3.14.* Boltzmann weighted average spectra (top) and hand-weighted spectra (bottom) overlaid onto the IRMPD spectrum of LysAH<sup>+</sup>. (The dashed black plot represents the IRMPD spectrum, and the solid light blue plots represent the calculated spectra.)

In all ten ALysH<sup>+</sup> conformations, the NH<sub>3</sub><sup>+</sup> group hydrogen bonds with the Ac-Ala and Ala-Lys amide C=O groups, while the carboxyl group is positioned far from it. Among them, there are two groups of conformations with different backbone geometries. In the first group of conformations, C1', C4', C5', C6', and C9', the backbone geometry around alanine is like in a helix, with the Ac-Ala and Aly-Lys amide C=O groups nearly parallel to each other. The helical backbone geometry around alanine is left-handed in C1', C4', C5' and C6', and is right-handed in C9', shown in Figure 3.15. Their dihedral angles across the alanine alpha carbon,  $\phi$ : C(O)–N(H)–C <sub>$\alpha$</sub> (Ala)–C(O) and  $\psi$ : N(H)–C <sub>$\alpha$</sub> (Ala)–C(O)–N(H), and C→O vector angle between the Ac-Ala and Ala-Lys amide C=O groups are shown in Table 3.6.

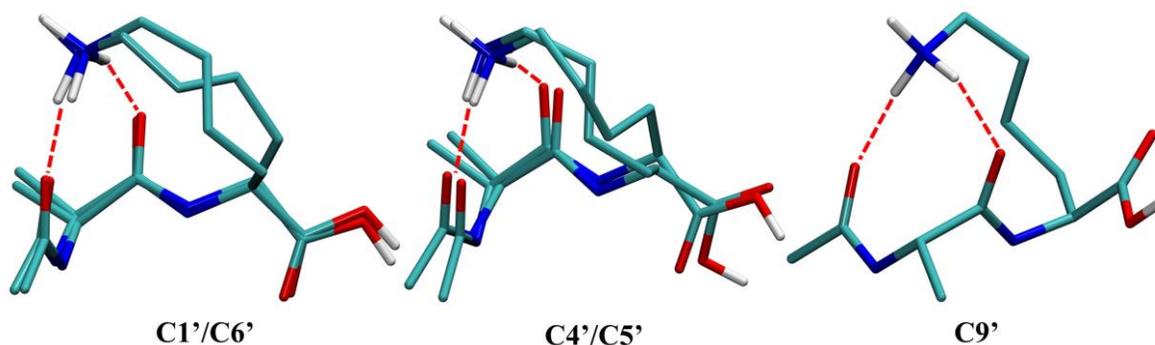


Figure 3.15. Structure of C1' and C6' (left), C4' and C5' (middle), and C9' (right). (The dashed lines indicate H-bonding.)

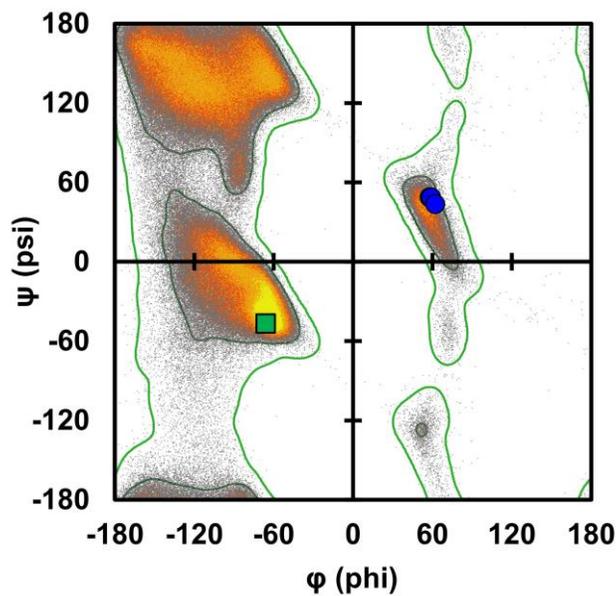
Table 3.6

*Dihedral Angles across the C<sub>α</sub> of Alanine and C→O Vector Angle between the Ac-Ala and Ala-Lys Amide C=O Groups in C1', C4', C5', C6', and C9'*

	$\phi$ (phi) (°)	$\psi$ (psi) (°)	C→O vector angle (°)
	C(O)–N(H)–C <sub>α</sub> (Ala)–C(O)	N(H)–C <sub>α</sub> (Ala)–C(O)–N (H)	
C1'	57.8	48.7	24.0
C4'	58.1	48.8	23.1
C5'	58.9	48.8	24.5
C6'	62.1	43.6	28.6
C9'	-66.1	-46.7	23.7

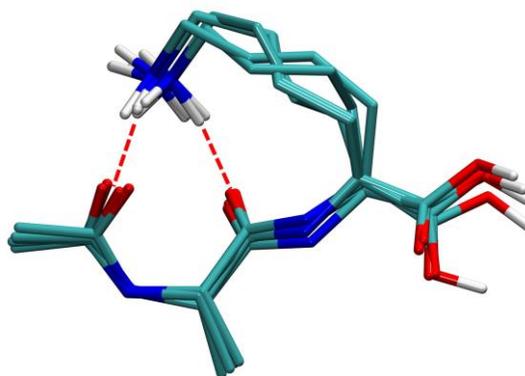
Like in C3 and C8 of LysAH<sup>+</sup>, the C→O vector angle between the Ac-Ala and Ala-Lys amide C=O groups in this group of ALysH<sup>+</sup> conformations, range from 23.1° to 28.6°, demonstrating their good alignment and the helical character of the conformations. As shown in Table 3.6, C9' has a ( $\phi$ ,  $\psi$ ) dihedral pair of (-66.1°, -46.7°), close to the ideal alpha helix dihedral angles of (-60°, -45°), while the others have the opposite dihedral angles close to (60°, 45°). To clearly show their helical geometries, their dihedral angles in Table 3.6 are overlaid to the Ramachandran plot in Figure 3.16. The dihedral angle pairs of C1', C4', C5', and C6' (dark blue circles) fall into the left-hand helix region, while C9' (green square) falls into the right-hand helix region.

Comparison of the theoretical spectra of this group of conformations to the IRMPD spectrum of ALysH<sup>+</sup> suggests that all of them have a partial match, while the spectra of C1' and C6' are almost identical to the experimental spectrum through the whole wavenumber range.



*Figure 3.16.* Ramachandran plot of the left-handed helical conformations (dark blue circles, C1', C4', C5', C6') and the right-handed helical conformation (green square, C9').

In the second group of ALysH<sup>+</sup> conformations, C2', C3', C7', C8', and C10', the two amide C=O groups are nearly perpendicular to each other. Their structures are overlaid and shown in Figure 3.17. The dihedral angles across the alanine alpha carbon,  $\phi$ : C(O)–N(H)–C<sub>α</sub>(Ala)–C(O) and  $\psi$ : N(H)–C<sub>α</sub>(Ala)–C(O)–N(H), and C→O vector angle between the Ac-Ala and Ala-Lys amide C=O groups are shown in Table 3.7.



*Figure 3.17.* Structure of C2', C3', C7', C8', and C10'. (The dashed lines indicate H-bonding.)

Table 3.7

*Dihedral Angles across the C<sub>α</sub> of Alanine and C→O Vector Angle between the Ac-Ala and Ala-Lys Amide C=O Groups in C2', C3', C7', C8', and C10'*

	$\phi$ (phi) (°)	$\psi$ (psi) (°)	C→O vector angle (°)
	C(O)–N(H)–C <sub>α</sub> (Ala)–C(O)	N(H)–C <sub>α</sub> (Ala)–C(O)–N(H)	
C2'	-63.8	166.8	93.5
C3'	-55.8	152.3	92.8
C7'	-62.4	162.7	96.1
C8'	-57.0	155.7	94.3
C10'	-55.0	152.1	93.2

Like in C2, C4 and C6 of LysAH<sup>+</sup>, the C→O vector angles between the Ac-Ala and Ala-Lys amide C=O groups in this group of ALysH<sup>+</sup> conformations range from 92.8° to 96.1°, clearly showing their perpendicular structural relationship. The dihedral angles for this group of conformations are overlaid to the Ramachandran plot to illustrate the backbone geometry around the alanine residue (Figure 3.18). As shown, the dihedral angle pairs ( $\phi$ ,  $\psi$ ) for this group of conformations fall on the edge of the broad beta-sheet region. Compared to the first group of ALysH<sup>+</sup> conformations, this group has a worse match to the IRMPD spectrum of ALysH<sup>+</sup>, though both the carboxyl C=O stretching region and the amide II regions match well.

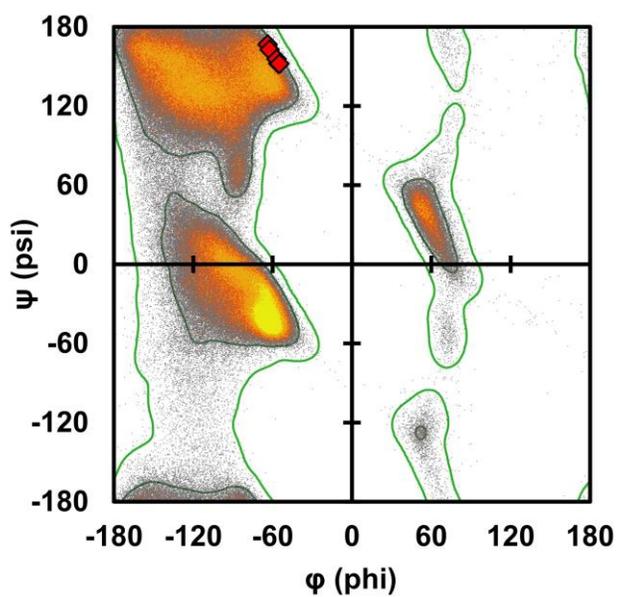
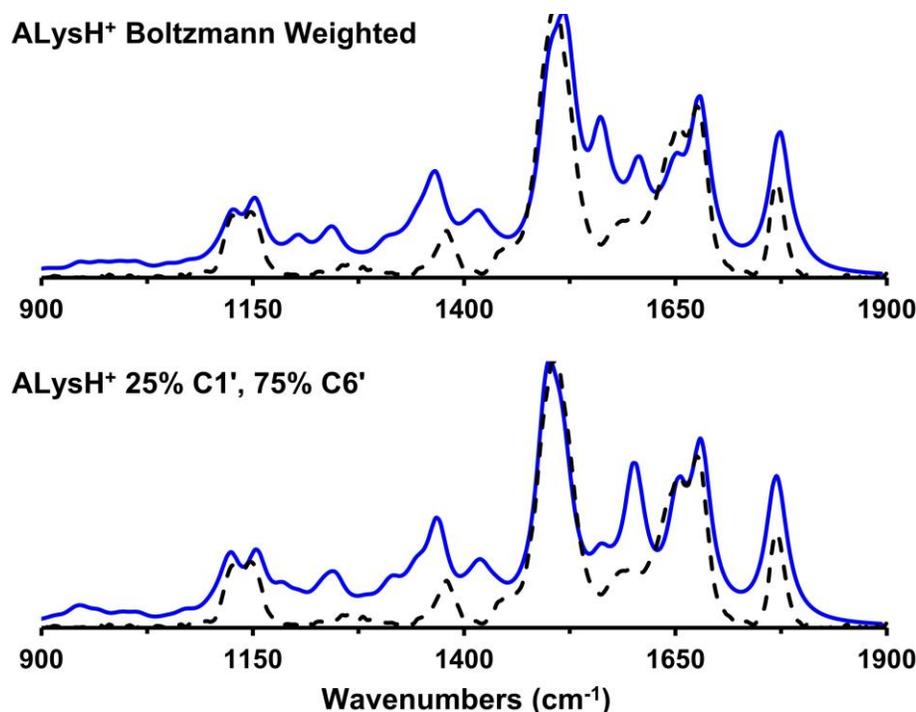


Figure 3.18. Ramachandran plot of conformation C2', C3', C7', C8' and C10' (red diamond).

Unlike the non-representative C1 of LysAH<sup>+</sup>, C1' of ALysH<sup>+</sup> has a great match to the IRMPD spectrum. However, the conformation C6' has an even better match to the IRMPD spectrum. To compare with the Boltzmann weighted average spectrum, a hand-weighted spectrum with C1' and C6' is created and shown in Figure 3.19.



*Figure 3.19.* Boltzmann weighted average spectra (top) and hand-weighted spectra (bottom) overlaid onto the IRMPD spectrum of ALysH<sup>+</sup>. (The dashed black plot represents the IRMPD spectrum, and the solid dark blue plots represent the calculated spectra.)

Compared to the Boltzmann weighted average spectrum, the hand-weighted spectrum with C1' (25%) and C6' (75%) matches the ALysH<sup>+</sup> IRMPD spectrum better. Additionally, as shown in Figure 3.15, conformation C1' and C6' are structurally similar, with major differences in the orientation of the lysine sidechain. On one hand, their similar structures grant C6' a high chance of existence for its easy conversion from the lower energy conformation C1'. On the other hand, their similarity coincides with the interpretation of the ALysH<sup>+</sup> IRMPD spectrum that conformations with similar intramolecular interaction networks are present.

In summary, two distinct conformations, C2 and C3, are used to generate a hand-weighted spectrum with a good match to the LysAH<sup>+</sup> IRMPD spectrum and represent the primary structures of LysAH<sup>+</sup>. Two structurally similar conformations C1' and C6', provide nearly perfect match to the ALysH<sup>+</sup> IRMPD spectrum and represent the primary structures of ALysH<sup>+</sup>.

### Structural and Energetic Comparison

The computed conformations give thermochemical values that match the experimental results, and the observed basicity difference between LysA and ALys mainly comes from the stability difference of the protonated peptides. It is important to investigate the conformational features of LysAH<sup>+</sup> and ALysH<sup>+</sup> thoroughly. Among the set of conformations, C2 and C3 of LysAH<sup>+</sup> and C1' and C6' of ALysH<sup>+</sup> are selected as structurally representative conformations for further characterization (Table 3.8), as they provide good matching to their IRMPD spectra of the corresponding peptides. To understand the stability difference between the selected conformations, multiple types of energetic characterizations were performed, including the charge distribution analysis, dipole moment decomposition (Lu & Chen, 2012), hydrogen bonding analysis (Emamian et al., 2019), torsional strain analysis (Smith\* & Jaffe, 1996), and reduced density gradient (RDG) analysis (Johnson et al., 2010).

Table 3.8

*Relative Free Energy Values of Representative Conformations of LysAH<sup>+</sup> and ALysH<sup>+</sup>*

	<b>Conf.</b>	$\Delta G^a$ <b>(kcal/mol)</b>
LysAH <sup>+</sup>	C2	4.1
	C3	4.2
ALysH <sup>+</sup>	C1'	0.0
	C6'	2.1

<sup>a</sup> Relative free energy between selected conformations for the two peptide systems, LysAH<sup>+</sup> and ALysH<sup>+</sup>, where the energy of C1' is set to be 0.0 kcal/mol.

**Charge distribution analysis and dipole moment decomposition.** Intuitively, a lower energy structure of the protonated isomeric peptide would have a larger capability to remediate the charge, and a more accommodated  $\text{NH}_3^+$  group, which might be reflected as a lower charge density on the  $\text{NH}_3^+$  group. The allocation of the charge density on the  $\text{NH}_3^+$  was investigated. For the  $\text{LysAH}^+$  and  $\text{ALysH}^+$  representative conformations, Mulliken, NPA, and Hirshfeld, three different types of charges were calculated by summing the atomic charges of the  $\text{NH}_3^+$  group. The results are summarized in Table 3.9. For an easy visualization and comparison, the three types of charges are also presented as bar graphs shown in Figure 3.20.

Table 3.9  
*Relative Free Energy and the Three Types of Charges on the  $\text{NH}_3^+$  Group in the Representative Conformations of  $\text{LysAH}^+$  and  $\text{ALysH}^+$*

	<b>Conf.</b>	<b><math>\Delta G</math> (kcal/mol)</b>	<b>Mulliken<sup>a</sup></b>	<b>NPA<sup>a</sup></b>	<b>Hirshfeld<sup>b</sup></b>
$\text{LysAH}^+$	C2	4.1	0.628	0.633	0.466
	C3	4.2	0.632	0.649	0.424
$\text{ALysH}^+$	C1'	0.0	0.636	0.641	0.451
	C6'	2.1	0.610	0.635	0.449

<sup>a</sup> Values (a.u.) obtained from single-point calculations at B3LYP-D3(BJ)/ma-def2-TZVPP level of theory.

<sup>b</sup> Values (a.u.) calculated by Multwfn 3.8 using the fchk file from single-point calculations at B3LYP-D3(BJ)/ma-def2-TZVPP level of theory, with the build-in sphericalized atomic densities in free-states.

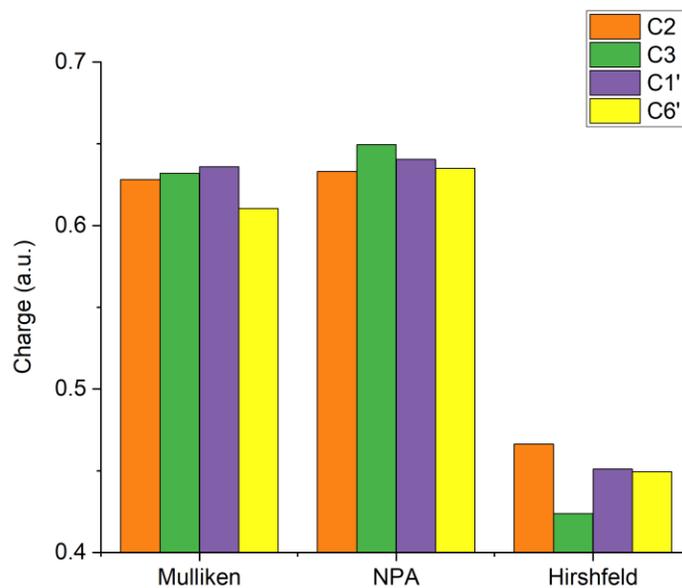


Figure 3.20. Three types of charges on the  $\text{NH}_3^+$  group in the representative conformations of  $\text{LysAH}^+$  and  $\text{ALysH}^+$ .

As shown in Table 3.9 and Figure 3.20, a clear correlation between the charge on the  $\text{NH}_3^+$  group and the relative free energy cannot be established for any of the three types of charges. For example, C1' of  $\text{ALysH}^+$  is the lowest in free energy, but it is ranked the highest in Mulliken and the second-highest in NPA and Hirshfeld charges. Though the charge on the  $\text{NH}_3^+$  group alone cannot explain the relative stability of the representative conformations, they contribute to their final stability differences.

Comparing the charges allocated to each atom on the  $\text{NH}_3^+$  group, the Hirshfeld charge reflects the charge distribution better than Mulliken and NPA charges. In the  $\text{NH}_3^+$  group of all four representative conformations, one hydrogen is not involved in any intramolecular interaction, while the other two are hydrogen-bonded. In hydrogen bonds, the acceptor atom shares its electron density with the hydrogen atom and reduces the atomic charge of the hydrogen (Baranović et al., 2012). Likely, among the three hydrogens of the  $\text{NH}_3^+$  group, the positive charge on isolated hydrogen is higher than the other two. The different atomic charges on the nitrogen and hydrogens of the  $\text{NH}_3^+$  group, together with their atomic labels, are tabulated in Table 3.10.

Table 3.10  
Atomic Charges on the Nitrogen and Hydrogens of the  $\text{NH}_3^+$  Group in the Representative Conformations of  $\text{LysAH}^+$  and  $\text{ALysH}^+$

	Conf.	Atom	Mulliken	NPA	Hirshfeld	
$\text{LysAH}^+$	C2	$\text{N}_{15}$	-0.015	-0.699	0.021	
		$\text{H}_{17}$ (isolated)	0.188	0.422	0.179	
		$\text{H}_{16}$	0.200	0.441	0.157	
			$\text{H}_{18}$	0.255	0.469	0.110
	C3	$\text{N}_{15}$	-0.019	-0.704	0.014	
		$\text{H}_{18}$ (isolated)	0.192	0.426	0.169	
$\text{H}_{16}$		0.251	0.468	0.113		
		$\text{H}_{17}$	0.209	0.459	0.128	
$\text{ALysH}^+$	C1'	$\text{N}_{22}$	-0.022	-0.712	0.019	
		$\text{H}_{23}$ (isolated)	0.190	0.419	0.178	
		$\text{H}_{24}$	0.244	0.472	0.124	
			$\text{H}_{25}$	0.224	0.462	0.130
	C6'	$\text{N}_{22}$	-0.040	-0.709	0.018	
		$\text{H}_{25}$ (isolated)	0.188	0.419	0.179	
$\text{H}_{23}$		0.220	0.457	0.133		
		$\text{H}_{24}$	0.243	0.468	0.120	

<sup>a</sup>The number subscript refers to the atomic label in the molecular model.

As shown, the isolated hydrogens are assigned with lower Mulliken and NPA charges than the hydrogen-bonded hydrogens, which is against the prediction of their relative charges. Contrarily, the isolated hydrogens are allocated with a higher Hirshfeld charge than the hydrogen-bonded hydrogens, in agreement with the prediction. So, for the  $\text{LysAH}^+$  and  $\text{ALysH}^+$  conformations, the Hirshfeld charge better describes the charge distribution than the Mulliken and NPA charge.

Reconsidering the charge on the  $\text{NH}_3^+$  group, C1' and C6' of  $\text{ALysH}^+$  are assigned with similar Hirshfeld charges, which is very likely due to their similar structures. Comparatively, C2 of  $\text{LysAH}^+$  is a different molecule, but its  $\text{NH}_3^+$  group also forms two hydrogen bonds with two parallelly positioned amide C=O groups, like that in C1' and C6' of  $\text{ALysH}^+$ . Additionally, its carboxyl C=O group is placed towards and forms attractive interaction with its  $\text{NH}_3^+$  group. As a result, C2 of  $\text{LysAH}^+$  has a lower charge on the  $\text{NH}_3^+$  group than C1' and C6' of  $\text{ALysH}^+$ . Though the Hirshfeld charges correlate well

with the structural features of the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations, a simple correlation between their relative free energies and the charges on the NH<sub>3</sub><sup>+</sup> group cannot be established.

The molecular dipole moment,  $\mu_{\text{whole}}$  is decomposed into the dipole moment of the molecular fragment excluding the NH<sub>3</sub><sup>+</sup> group, noted as the backbone dipole moment,  $\mu_{\text{backbone}}$ , and the dipole moment of the NH<sub>3</sub><sup>+</sup> group, noted as the NH<sub>3</sub><sup>+</sup> dipole moment,  $\mu_{\text{NH}_3}$ . The decomposition is performed using the Multiwfn program (Lu & Chen, 2012), and allows to analyze the interaction of the molecular backbone dipole moment with the NH<sub>3</sub><sup>+</sup> group. A strong stabilization interaction better stabilizes the positive charge on the NH<sub>3</sub><sup>+</sup> group. The computed  $\mu_{\text{backbone}}$  and  $\mu_{\text{NH}_3}$  values are shown in Table 3.11 and visualized in Figure 3.21.

To qualitatively evaluate the alignment and interaction between the dipole moments of the backbone and the NH<sub>3</sub><sup>+</sup> group, the vector angle and the dot product between  $\mu_{\text{backbone}}$  and  $\mu_{\text{NH}_3}$  are calculated manually and summarized in Table 3.11. Detailed procedures of dipole moment analysis can be found in Chapter 2. Briefly, the fchk file from single point energy calculation at B3LYP-D3(BJ)/ma-def-TZVPP level of theory is fed to Multiwfn 3.8. Dipole moments of the whole molecule or part of the molecule can be calculated by assigning the atoms to be considered. The dipole moments are calculated by subtracting the coordinates of the negative charge center from the coordinates of the positive charge center. The vector angle between dipole moments is obtained in the same way as the C→O angle between amide C=O groups. The dot product between dipole moments is calculated by summing the products of the corresponding coordinates of the two dipole moments.

Table 3.11  
Dipole Moments of the Representative Conformations of LysAH<sup>+</sup> and ALysH<sup>+</sup>

	Conf.	$\Delta G$ (kcal/mol)	$\mu_{\text{backbone}}$ (D)	$\mu_{\text{NH}_3}$ (D)	$\mu_{\text{backbone}}$ (D)	Vector angle between $\mu_{\text{NH}_3}$ and $\mu_{\text{backbone}}$ (°)	Dot product of $\mu_{\text{NH}_3}$ and $\mu_{\text{backbone}}$ (D <sup>2</sup> )
LysAH <sup>+</sup>	C2	4.1	7.75	7.68	2.46	97.46	-2.46
	C3	4.2	2.36	4.89	5.59	155.08	-24.81
ALysH <sup>+</sup>	C1'	0.0	4.94	7.15	3.41	141.87	-19.16
	C6'	2.1	4.34	7.23	4.14	147.54	-25.27

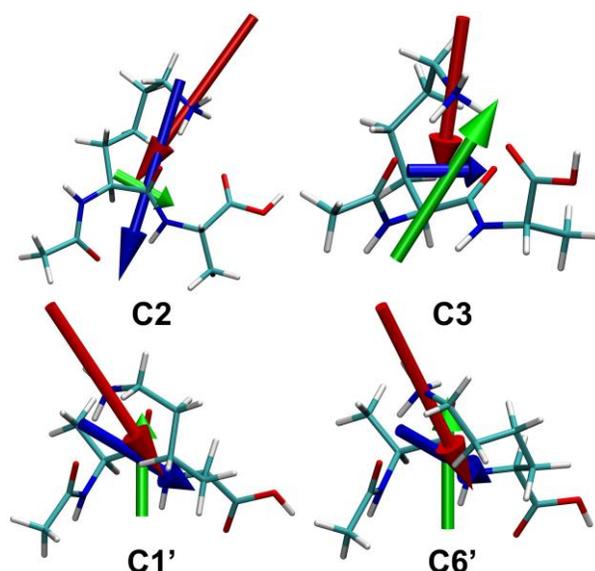


Figure 3.21. Dipole moments of the whole molecule (blue), the backbone (green), and the NH<sub>3</sub><sup>+</sup> group (red) for the representative conformations of LysAH<sup>+</sup> (top) and ALysH<sup>+</sup> (bottom).

In C2 of LysAH<sup>+</sup>,  $\mu_{\text{backbone}}$  does not point toward the NH<sub>3</sub><sup>+</sup> group.  $\mu_{\text{backbone}}$  and  $\mu_{\text{NH}_3}$  are nearly perpendicular to each other, with a vector angle of 97.46°. The small amplitude of their dot product, -2.46 D<sup>2</sup>, reflects the negligible interaction between  $\mu_{\text{backbone}}$  and  $\mu_{\text{NH}_3}$ . As a result, its  $\mu_{\text{whole}}$  is nearly the same as  $\mu_{\text{NH}_3}$ .

Differently, in C3 of LysAH<sup>+</sup>,  $\mu_{\text{backbone}}$  points toward and forms a strong stabilization interaction with the NH<sub>3</sub><sup>+</sup> group. The nearly opposite alignment of  $\mu_{\text{backbone}}$  and  $\mu_{\text{NH}_3}$ , with an angle of 155.08°,

results in a large cancellation of each other. This is reflected in the large amplitude of their dot product of  $-24.81 \text{ D}^2$ . As a result,  $\mu_{\text{whole}}$  is small as  $2.36 \text{ D}$ , much smaller than  $\mu_{\text{backbone}}$  and  $\mu_{\text{NH}_3}$ .

Because of the structural similarity within C1' and C6' of ALysH<sup>+</sup>, the orientation and magnitudes of the dipole moments are comparable. Like in C3 of LysAH<sup>+</sup>,  $\mu_{\text{backbone}}$  in C1' and C6' of ALysH<sup>+</sup> points toward the NH<sub>3</sub><sup>+</sup> group and is nearly oppositely aligned with  $\mu_{\text{NH}_3}$ . As shown, the vector angle between  $\mu_{\text{backbone}}$  and  $\mu_{\text{NH}_3}$  are  $141.87^\circ$  and  $147.54^\circ$  in C1' and C6', respectively. The large amplitude of their dot products demonstrates the large cancellation between  $\mu_{\text{backbone}}$  and  $\mu_{\text{NH}_3}$  and the good interaction of the backbone dipole moment with the NH<sub>3</sub><sup>+</sup> group. Compared to C2 of LysAH<sup>+</sup>, the backbone dipole moment,  $\mu_{\text{backbone}}$ , in C3 of LysAH<sup>+</sup> and the ALysH<sup>+</sup> conformations is nearly oppositely aligned with  $\mu_{\text{NH}_3}$  and stabilizes the terminal NH<sub>3</sub><sup>+</sup> group. Yet, a clear and simple correlation could not be established between the relative energy and the strength of the interaction between the backbone dipole moment and the NH<sub>3</sub><sup>+</sup> group.

In summary, the charge on the NH<sub>3</sub><sup>+</sup> group and the interaction between the backbone dipole moment and the NH<sub>3</sub><sup>+</sup> group could not explain the relative free energies of the LysAH<sup>+</sup> and the ALysH<sup>+</sup> conformations directly, but they may contribute to the different stabilities. By comparing the charges on each hydrogen and the whole NH<sub>3</sub><sup>+</sup> group, the Hirshfeld charge reflects the higher charge on the isolated hydrogen and correlates well with the structural features of the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations. Additionally, in the helical conformations, C3 of LysAH<sup>+</sup>, and C1' and C6' of ALysH<sup>+</sup>, the backbone dipole moment,  $\mu_{\text{backbone}}$  is better orientated to form beneficial interaction with and stabilizes the NH<sub>3</sub><sup>+</sup> group.

**Torsional strain on lysine sidechain.** While examining the structure of the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations, a few C-C-C-C torsional angles on the lysine sidechain were found to be unfavorable. Because the energy difference between eclipsed and gauche geometries is comparable to the energy differences between LysAH<sup>+</sup> and ALysH<sup>+</sup>, the torsional strain on the lysine sidechain is investigated. The torsional strain was estimated using the torsional strain function for n-butane. Detailed procedures can be found in Chapter 2. Briefly, the dihedral angles made of only heavy atoms on the lysine side chain are obtained from their structures. By plugging a dihedral angle into the torsional strain function as the torsion angle, the torsional strain of that dihedral angle is calculated. Individual dihedral angles on the lysine side chain in the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations are listed in Table 3.12. Their corresponding torsional strain values are listed in Table 3.13. There are two dihedral angles that have the C<sub>α</sub>-C<sub>β</sub> bond as the central bond. Their torsional strains are calculated individually and are both taken into account.

The total torsional strain on the lysine sidechain noted as E<sub>torsion</sub> is calculated by summing the torsional strain of all the dihedral angles. Because C1' of ALysH<sup>+</sup> has the smallest E<sub>torsion</sub>, it is assigned as the reference of relative total torsional strain, Rel. E<sub>torsion</sub>, with 0.0 kcal/mol of energy. The Rel. E<sub>torsion</sub> values of the other conformations are calculated accordingly. The E<sub>torsion</sub> and Rel. E<sub>torsion</sub> values of the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations are listed in Table 3.13.

Table 3.12  
Dihedral Angles on the Lysine Sidechain in the  $\text{LysAH}^+$  and  $\text{ALysH}^+$  Conformations

Conf.	Dihedral Angle ( $^\circ$ )					
	$\alpha_1$ N(H)-C $\alpha$ -C $\beta$ -C $\gamma$	$\alpha_2$ C(O)-C $\alpha$ -C $\beta$ -C $\gamma$	$\beta$ C $\alpha$ -C $\beta$ -C $\gamma$ -C $\delta$	$\gamma$ C $\beta$ -C $\gamma$ -C $\delta$ -C $\epsilon$	$\delta$ C $\gamma$ -C $\delta$ -C $\epsilon$ -N(H3)	
LysAH $^+$	-170.822	66.035	-149.495	61.305	47.485	
C3	62.489	-70.660	92.565	-122.153	67.845	
ALysH $^+$	-52.957	-171.976	-67.794	161.807	-67.079	
C6'	-61.766	178.088	93.283	-159.570	79.704	

Table 3.13  
Torsional Strain on the Lysine Sidechain in the  $\text{LysAH}^+$  and  $\text{ALysH}^+$  Conformations

Conf.	Torsion Energy (kcal/mol)						Rel. E $_{\text{torsion}}^a$
	$\alpha_1$ N(H)-C $\alpha$ -C $\beta$ -C $\gamma$	$\alpha_2$ C(O)-C $\alpha$ -C $\beta$ -C $\gamma$	$\beta$ C $\alpha$ -C $\beta$ -C $\gamma$ -C $\delta$	$\gamma$ C $\beta$ -C $\gamma$ -C $\delta$ -C $\epsilon$	$\delta$ C $\gamma$ -C $\delta$ -C $\epsilon$ -N(H3)	E $_{\text{torsion}}$	
LysAH $^+$	0.2	0.7	1.7	0.7	1.1	4.3	1.2
C3	0.7	0.8	2.2	3.3	0.7	7.7	4.6
ALysH $^+$	0.8	0.1	0.7	0.7	0.7	3.1	0.0
C6'	0.7	0.0	2.2	0.9	1.3	5.0	1.9

<sup>a</sup> Relative torsional strain on the lysine sidechain between the selected conformations of  $\text{LysAH}^+$  and  $\text{ALysH}^+$ , where the energy of C1' is set to be 0.0 kcal/mol.

In C2 of LysAH<sup>+</sup>, the  $\beta$  and  $\delta$  dihedral angles are  $-149.495^\circ$  and  $47.485^\circ$ , which are moderately unfavorable. Their torsional strains are 1.7 and 1.1 kcal/mol, both higher than 1.0 kcal/mol. The  $E_{\text{torsion}}$ , and Rel.  $E_{\text{torsion}}$  values of C2 are 4.3 and 1.2 kcal/mol, respectively. C3 of LysAH<sup>+</sup> has the two most unfavorable torsion angles among the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations. The two dihedral angles are  $92.565^\circ$  and  $-122.153^\circ$ , which are just about the second highest peaks in the torsional strain function. Their corresponding torsional strains are 2.2 and 3.3 kcal/mol. As a result, the  $E_{\text{torsion}}$  of C3 is 7.7 kcal/mol, which is the highest among the conformations of LysAH<sup>+</sup> and ALysH<sup>+</sup>, and is higher than that of C1 of ALysH<sup>+</sup> by 4.6 kcal/mol.

The torsional angles in C1' of ALysH<sup>+</sup> are close to  $\pm 60^\circ$  or  $\pm 180^\circ$ , and their torsional strains are all below 1.0 kcal/mol. Its total torsion energy is estimated to be 3.1 kcal/mol, the lowest among the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations. As mentioned before, the main structural difference between C1' and C6' is the orientation of the lysine sidechain. In C6', there are two torsional angles with energy higher than 1.0 kcal/mol. The  $E_{\text{torsion}}$  of C6' is 5.0 kcal/mol, higher than C1' by 1.9 kcal/mol.

In summary, the  $E_{\text{torsion}}$  values are 4.3 and 7.7 kcal/mol for C2 and C3 of LysAH<sup>+</sup>, and 3.1 and 5.0 kcal/mol for C1' and C6' of ALysH<sup>+</sup>, respectively. The Rel.  $E_{\text{torsion}}$  of C3 of LysAH<sup>+</sup>, C1' and C6' of ALysH<sup>+</sup> are 4.6, 0.0, and 1.9 kcal/mol, respectively. These values well correlate with their relative free energies of 4.2, 0.0, and 2.1 kcal/mol, indicating that the torsional strain on the lysine sidechain plays an important role in the stability differences between these three conformations. The Rel.  $E_{\text{torsion}}$  value of C2 is 1.2 kcal/mol, much smaller than its relative free energy of 4.1 kcal/mol. This suggests that besides the torsional strain on the lysine side chain, other factors also contribute to its observed large relative energy, including its high charge density on the NH<sub>3</sub><sup>+</sup> group and the missing strong interaction between the backbone dipole moment with the NH<sub>3</sub><sup>+</sup> group.

**Hydrogen bond characterization.** In LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations, two hydrogen bonds are formed with the NH<sub>3</sub><sup>+</sup> group to stabilize its charge. To qualitatively evaluate the strength of those hydrogen bonds, multiple geometric parameters were collected and analyzed. In addition to a short O --- H distance, linear alignment of the C=O bond and the H-N bond is also beneficial for a stronger hydrogen bonding interaction as it allows a favorable dipole-dipole interaction. Two angles are found to describe the alignment of the C=O bond and the H-N bond well. The alignment angle is the sum of the C=O --- H angle and the O --- H-N angle. The other is the vector angle between the donor vector, C→O, and the acceptor vector, H→N, of the hydrogen bond. The procedures to obtain the vector angle can be found in Chapter 2. Briefly, for each hydrogen bond, its donating group vector, C→O, and accepting group vector, H→N, are obtained by subtracting the coordinates of the initial point atom from the coordinates of the terminal point atom. Based on the cosine formula for vector dot product, the cosine value of the angle between C→O and H→N can be obtained by dividing their dot product by their magnitudes. The C→O and H→N dipole moments are better aligned as the alignment angle approaches 360°, and the vector angle approaches 0°.

Geometric parameters for the hydrogen bonds with NH<sub>3</sub><sup>+</sup> group, including the O --- H distance, the C=O --- H angle, the O --- H-N angle, the alignment angle, and the vector angle, are obtained and tabulated in Table 3.14. As shown, the shorter hydrogen bonds, which have smaller O --- H distances, noted as HB<sub>1</sub>, always have larger alignment angles and smaller vector angles than the longer hydrogen bond, noted as HB<sub>2</sub>. The geometric differences show that HB<sub>1</sub> hydrogen bonds have advantages over HB<sub>2</sub> in both O --- H distance and alignment.

Additionally, quantitative analysis was performed. In the work of Saeedreza Emamian, et al. (Emamian et al., 2019), the hydrogen bond energy (HBE) was found to linearly correlate with the electron density,  $\rho$ , at the bond critical point (BCP). Based on their findings, the HBE value of the hydrogen bonds of interest was calculated for the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations (Table 3.15). The total hydrogen bond energy (HBE<sub>total</sub>) was obtained by summing the HBE of both hydrogen bonds with the NH<sub>3</sub><sup>+</sup> group.

Table 3.14

Geometry Information of Hydrogen Bonds with the  $NH_3^+$  Group for the Representative Conformations of  $LysAH^+$  and  $ALysH^+$

Conf.	Hydrogen Bond	Hydrogen Acceptor	C=O ... H angle (°)	O ... H-N angle (°)	Alignment angle (°)	Vector angle (°)	O ... H Distance (Å)
C2	HB <sub>1</sub>	Lys-Ala amide C=O	121.1	176.4	297.5	60.9	1.584
	HB <sub>2</sub>	Carboxyl C=O	144.0	117.6	261.6	94.3	2.261
C3	HB <sub>1</sub>	Ac-Lys amide C=O	147.1	178.0	325.1	33.8	1.684
	HB <sub>2</sub>	Lys-Ala amide C=O	108.2	136.3	244.5	103.5	1.961
C1'	HB <sub>1</sub>	Ac-Ala amide C=O	163.3	161.4	324.7	35.0	1.745
	HB <sub>2</sub>	Ala-Lys amide C=O	121.0	146.9	267.9	88.5	1.777
C6'	HB <sub>1</sub>	Ac-Ala amide C=O	153.5	163.1	316.7	31.3	1.701
	HB <sub>2</sub>	Ala-Lys amide C=O	115.1	145.5	260.6	98.5	1.843

Table 3.15

Energetic Information of Hydrogen Bonds with the  $NH_3^+$  Group for the Representative Conformations of  $LysAH^+$  and  $ALysH^+$

Conf.	Hydrogen Bond	Electron Density (a.u.)	HBE (kcal/mol)	HBE <sub>total</sub> (kcal/mol)	Rel. HBE <sub>total</sub> (kcal/mol)
C2	HB <sub>1</sub>	6.05E-02	21.2		
	HB <sub>2</sub>	1.31E-02	5.4	26.6	-1.0
C3	HB <sub>1</sub>	4.63E-02	16.4		
	HB <sub>2</sub>	2.51E-02	9.4	25.8	-1.8
C1'	HB <sub>1</sub>	3.85E-02	13.9		
	HB <sub>2</sub>	3.81E-02	13.7	27.6	0.0
C6'	HB <sub>1</sub>	4.41E-02	15.7	27.7	0.1

Because of the geometry differences in both O --- H distance and alignment, two hydrogen bonds can differ dramatically in HBE. So, the HBE values of the two hydrogen bonds with the  $\text{NH}_3^+$  group could serve as quantitative measures of their contribution to stabilizing the  $\text{NH}_3^+$  group through hydrogen bonding interactions.

In C2 of  $\text{LysAH}^+$ ,  $\text{HB}_1$  has an O --- H distance of 1.584 Å, the shortest among all hydrogen bonds in the  $\text{LysAH}^+$  and  $\text{ALysH}^+$  conformations. Its alignment and vector angles are  $297.5^\circ$  and  $60.9^\circ$ , respectively. The HBE of  $\text{HB}_1$  is as high as 21.2 kcal/mol. Contrarily, the O --- H distance in  $\text{HB}_2$  is 2.261 Å, nearly 0.7 Å longer than that of  $\text{HB}_1$ . The C→O and H→N in  $\text{HB}_2$  are poorly aligned with an alignment angle of  $261.6^\circ$  and a vector angle higher of  $94.3^\circ$ . Its HBE is only 5.4 kcal/mol, nearly a quarter of  $\text{HB}_1$ .  $\text{HB}_1$  contributes 80% to stabilizing the  $\text{NH}_3^+$  group through hydrogen bonding interaction, while  $\text{HB}_2$  contributes only 20%.

Compared to C2, the difference between  $\text{HB}_1$  and  $\text{HB}_2$  in C3 in O --- H distance is smaller, while their difference in alignment is larger. The O --- H distance in  $\text{HB}_1$  is 1.684 Å, shorter than that in  $\text{HB}_2$  by 0.276 Å. The alignment angle of  $\text{HB}_1$  is  $325.1^\circ$ , nearly  $100^\circ$  larger than that of  $\text{HB}_2$ . Though the alignment difference becomes larger, the strength difference between  $\text{HB}_1$  and  $\text{HB}_2$  becomes smaller because their difference in O --- H distance gets smaller. In this case, the O --- H distance change has a stronger effect on the hydrogen bond energy than the change in the alignment of the C=O bond and the H-N bond. The contributions of  $\text{HB}_1$  and  $\text{HB}_2$  to stabilizing the  $\text{NH}_3^+$  group through hydrogen bonding interaction are approximately 60% and 40%, which are less different than those in C2.

In C1' of  $\text{ALysH}^+$ ,  $\text{HB}_1$  has an O --- H distance of 1.745 Å, an alignment angle of  $324.7^\circ$ , and an estimated HBE of 13.9 kcal/mol.  $\text{HB}_2$  has an O --- H distance of 1.777 Å, 0.032 Å longer than  $\text{HB}_1$ , and a much less favored alignment angle of  $267.9^\circ$ . However, its HBE is estimated to be 13.7 kcal/mol, nearly the same as  $\text{HB}_1$ . The hydrogen bond geometries in C6' are like that in C1'. The differences in O --- H distance and HBE between  $\text{HB}_1$  and  $\text{HB}_2$  are 0.142 Å and 3.8 kcal/mol. In C1' and C6' of  $\text{ALysH}^+$ ,  $\text{HB}_1$  and  $\text{HB}_2$  contribute almost equally to stabilizing the  $\text{NH}_3^+$  group through hydrogen bonding interactions.

In summary, the  $HBE_{total}$  values are 26.6 and 25.8 kcal/mol for C2 and C3 of LysAH<sup>+</sup>, and 27.6 and 27.7 kcal/mol for C1' and C6' of ALysH<sup>+</sup>. Because C1' of ALysH<sup>+</sup> has the lowest free energy, it is selected as the reference of Rel.  $HBE_{total}$ , where its Rel.  $HBE_{total}$  is set to be 0.0 kcal/mol. The Rel.  $HBE_{total}$  of the other conformations are calculated accordingly. The  $HBE_{total}$  in C6' is nearly the same as that in C1, and it has a Rel.  $HBE_{total}$  of 0.1 kcal/mol. The total hydrogen bond strength in C2 and C3 of LysAH<sup>+</sup> are smaller than that of the ALysH<sup>+</sup> conformations. Their Rel.  $HBE_{total}$  are -1.0 and -1.8 kcal/mol, respectively. The trend of  $HBE_{total}$  is about parallel to the trend of the relative stability of the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations, suggesting an important contribution of  $HBE_{total}$  to their different stabilities.

**RDG analysis.** Though the previous energetic characterizations provide valuable information about the peptides, they only capture a fraction of all intramolecular interactions. RDG analysis (Johnson et al., 2010) is a powerful tool for semi-quantitatively evaluating and visualizing a wide range of noncovalent interactions simultaneously. Detailed procedures to obtain the RDG scatter plot and the isosurface diagram can be found in Chapter 2. The RDG scatter plot and isosurface diagrams for the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations are shown in Figure 3.22 and Figure 3.23.

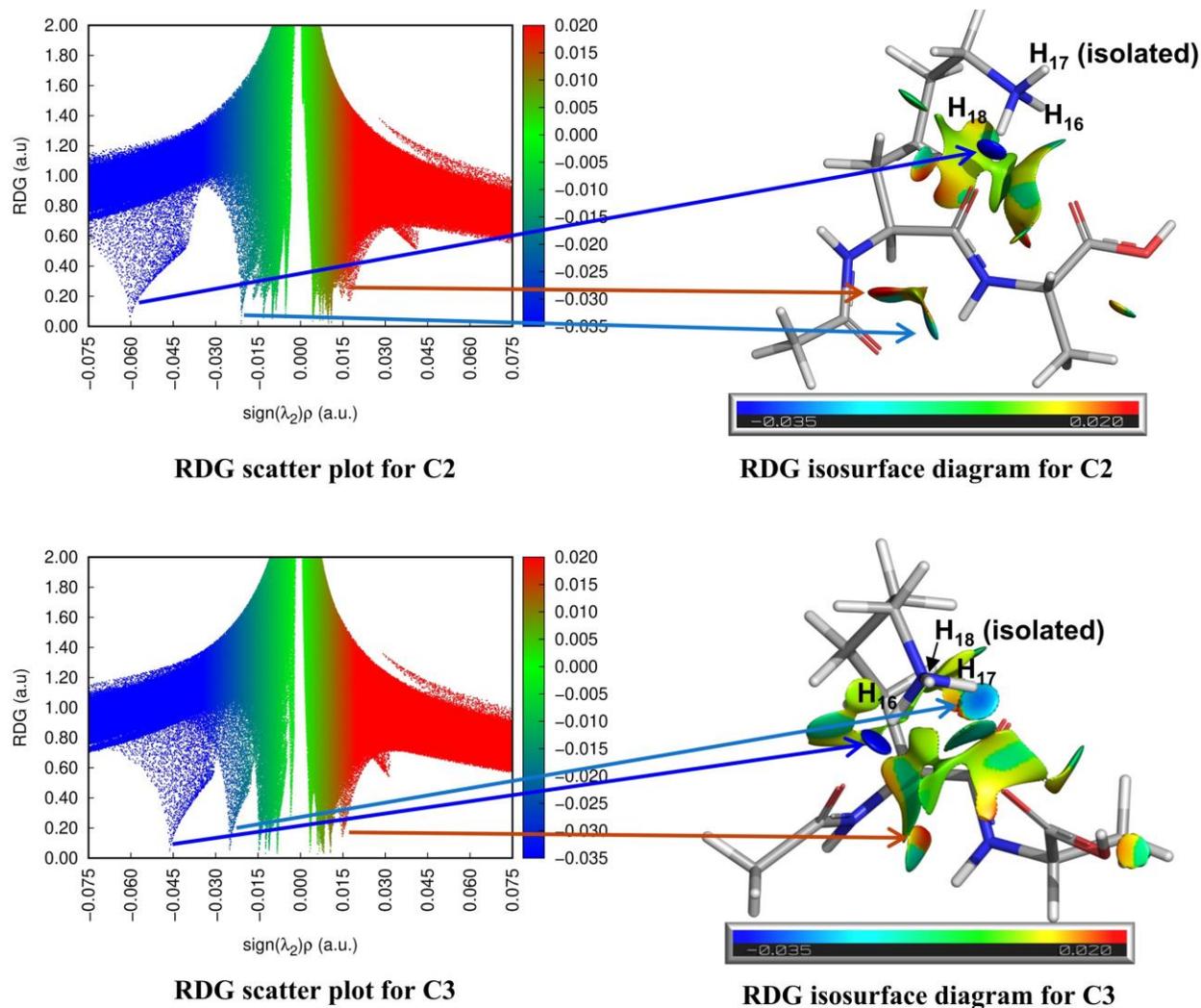


Figure 3.22. RDG scatter plot (left) and isosurface diagram (right) for C2 (top) and C3 (bottom) of LysAH<sup>+</sup>. (The arrows show examples of correlation between the spikes in the RDG scatter plot and the features in the RDG isosurface diagram.)

In C2 of LysAH<sup>+</sup>, there are twelve clearly identifiable attractive interactions between the following pairs of atoms and are listed using bullet points:

- H<sub>18</sub> on the NH<sub>3</sub><sup>+</sup> group and the lysine C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP45}} = 0.61\text{E-}01$  a.u., HBE = 21.2 kcal/mol),
- H<sub>16</sub> on the NH<sub>3</sub><sup>+</sup> group and the carboxyl C=O oxygen (weak hydrogen bond,  $\rho_{\text{BCP48}} = 0.13\text{E-}01$  a.u., HBE = 5.4 kcal/mol),
- the acetyl C=O oxygen and the alanine N-H hydrogen (weak hydrogen bond,  $\rho_{\text{BCP50}} = 0.21\text{E-}01$  a.u., HBE = 8.0 kcal/mol),

- the acetyl C=O oxygen and the lysine alpha hydrogen,
- the lysine C=O oxygen and one of the lysine beta hydrogens,
- the lysine C=O oxygen and one of the lysine gamma hydrogens,
- the lysine C=O oxygen and the carboxyl C=O oxygen ( $\rho_{\text{BCP51}} = 0.10\text{E-}01$  a.u.),
- the lysine C=O oxygen and the alanine alpha hydrogen,
- the carboxyl C=O oxygen and one of the lysine gamma hydrogens ( $\rho_{\text{BCP44}} = 0.53\text{E-}02$  a.u.),
- the carboxyl C-O(H) oxygen and one of the alanine beta hydrogens,
- one of the lysine beta hydrogens and one of the lysine epsilon hydrogens,
- one of the lysine gamma hydrogens and H<sub>18</sub> on the NH<sub>3</sub><sup>+</sup> group.

In C2 of LysAH<sup>+</sup>, there are eight clearly identifiable steric clashes in opened ring structures made of the following groups of atoms and are listed using bullet points:

- from the acetyl C=O oxygen to the lysine alpha hydrogen,
- from the lysine C=O carbon to one of the lysine gamma hydrogens,
- from the lysine C=O oxygen to one of the lysine beta hydrogens,
- from the lysine C=O oxygen to the alanine alpha hydrogen,
- from the lysine C=O oxygen to the carboxyl C=O oxygen,
- from the carboxyl C-O(H) oxygen to one of the alanine beta hydrogens,
- from one of the lysine beta hydrogens to one of the epsilon hydrogens,
- from one of the lysine gamma hydrogens to H<sub>16</sub> on the NH<sub>3</sub><sup>+</sup> group.

In C3 of LysAH<sup>+</sup>, there are thirteen clearly identifiable attractive interactions between the following pairs of atoms and are listed using bullet points:

- H<sub>16</sub> on the NH<sub>3</sub><sup>+</sup> group and the acetyl C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP54}} = 0.46\text{E-}01$  a.u., HBE = 16.4.8 kcal/mol),
- H<sub>17</sub> on the NH<sub>3</sub><sup>+</sup> group and the alanine C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP44}} = 0.25\text{E-}01$  a.u., HBE = 9.4 kcal/mol),
- the NH<sub>3</sub><sup>+</sup> group and the carboxyl C=O oxygen ( $\rho_{\text{BCP46}} = 0.15\text{E-}01$  a.u.),

- the acetyl C=O oxygen and one of the lysine gamma hydrogens,
- the acetyl C=O oxygen and the lysine C=O carbon,
- the acetyl C=O oxygen and the alanine N-H nitrogen,
- the acetyl C=O oxygen and the carboxyl C=O oxygen,
- the lysine C=O oxygen and one of the lysine beta hydrogens,
- the lysine C=O oxygen and one of the lysine delta hydrogens,
- the lysine C=O oxygen and the carboxyl C=O oxygen,
- the lysine C=O oxygen and one of the alanine beta hydrogens,
- the carboxyl C-O(H) oxygen and one of the alanine beta hydrogens,
- the lysine N-H nitrogen and the alanine N-H hydrogen.

In C3 of LysAH<sup>+</sup>, there are eight clearly identifiable steric clashes in opened ring structures made of the following groups of atoms and are listed using bullet points:

- from the acetyl C=O oxygen to the lysine C=O carbon,
- from the acetyl C=O carbon to one of the lysine gamma hydrogens,
- from the lysine N-H nitrogen to the alanine N-H hydrogen,
- from one of the lysine gamma hydrogens to H<sub>16</sub> on the NH<sub>3</sub><sup>+</sup> group,
- from the lysine C=O oxygen to one of the lysine beta hydrogens,
- from the lysine C=O oxygen to one of the alanine beta hydrogens,
- from the lysine C=O carbon to the carboxyl C=O oxygen,
- from the carboxyl C-O(H) oxygen to one of the alanine beta hydrogens.

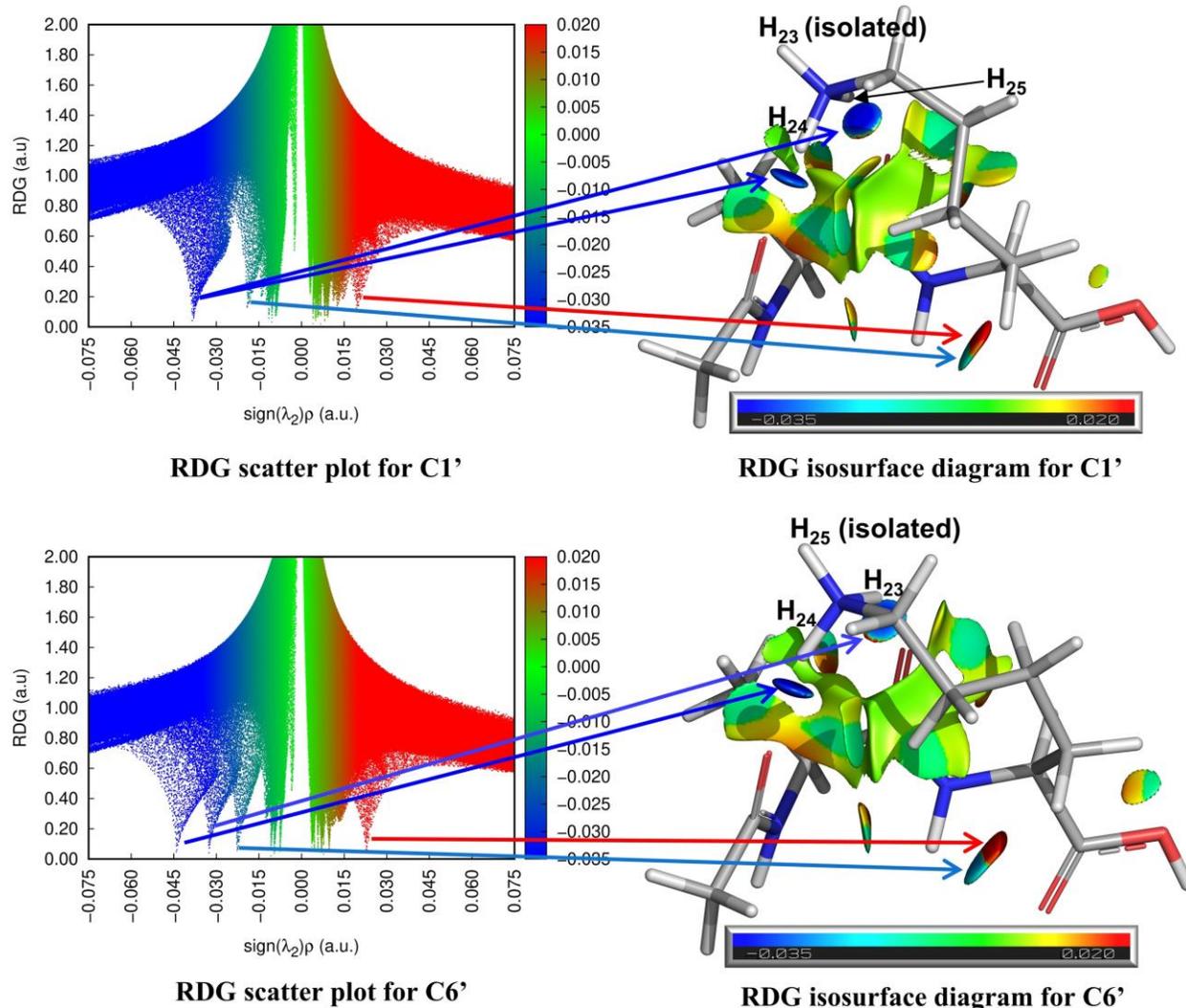


Figure 3.23. RDG scatter plot (left) and isosurface diagram (right) for C1' (top) and C6' (bottom) of ALysH<sup>+</sup>. (The arrows show examples of correlation between the spikes in the RDG scatter plot and the features in the RDG isosurface diagram.)

In C1' of ALysH<sup>+</sup>, there are fourteen clearly identifiable attractive interactions between the following pairs of atoms and are listed using bullet points:

- H<sub>24</sub> on the NH<sub>3</sub><sup>+</sup> group and the acetyl C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP44}} = 0.39\text{E-}01$  a.u., HBE = 13.8 kcal/mol),
- H<sub>25</sub> on the NH<sub>3</sub><sup>+</sup> group and the alanine C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP49}} = 0.38\text{E-}01$  a.u., HBE = 13.7 kcal/mol),
- the acetyl C=O oxygen and one of the alanine beta hydrogens,
- the acetyl C=O oxygen and the alanine C=O carbon,

- the acetyl C=O oxygen and one of the lysine gamma hydrogens ( $\rho_{\text{BCP42}} = 0.12\text{E-}01$  a.u.),
- the alanine C=O oxygen and one of the alanine beta hydrogens,
- the alanine C=O oxygen and the lysine alpha hydrogen,
- the alanine C=O oxygen and one of the lysine delta hydrogens ( $\rho_{\text{BCP50}} = 0.10\text{E-}02$  a.u.),
- the carboxyl C=O oxygen and lysine NH hydrogen ( $\rho_{\text{BCP}} \sim 0.20\text{E-}01$  a.u.),
- the alanine N-H nitrogen and the lysine N-H hydrogen,
- the lysine NH nitrogen and one of the lysine gamma hydrogens,
- one of the alanine beta hydrogens and the  $\text{NH}_3^+$  group,
- $\text{H}_{24}$  on the  $\text{NH}_3^+$  group and one of the lysine gamma hydrogens,
- one of the lysine delta hydrogens and the lysine alpha hydrogen.

In C1' of ALysH<sup>+</sup>, there are eleven clearly identifiable steric clashes in opened ring structures made of the following groups of atoms and are listed using bullet points:

- from the acetyl C=O oxygen to the alanine C=O carbon,
- from the acetyl C=O oxygen to one of the alanine beta hydrogens,
- from the alanine N-H nitrogen to the lysine N-H hydrogen,
- from the alanine C=O oxygen to one of the alanine beta hydrogens,
- from the alanine C=O oxygen to the lysine alpha hydrogen,
- from the lysine N-H nitrogen to one of the lysine gamma hydrogens,
- from the lysine N-H hydrogen to the carboxyl C=O oxygen ( $\rho_{\text{BCP}} \sim 0.20\text{E-}01$  a.u.),
- from the carboxyl C-O(H) oxygen to one of the lysine beta hydrogens,
- from the lysine alpha hydrogen to one of the lysine delta hydrogens,
- from one of the gamma hydrogens to  $\text{H}_{24}$  on the  $\text{NH}_3^+$  group,
- from one of the delta hydrogens to  $\text{H}_{25}$  on the  $\text{NH}_3^+$  group.

In C6' of ALysH<sup>+</sup>, there are fourteen clearly identifiable attractive interactions between the following pairs of atoms and are listed using bullet points:

- H<sub>24</sub> on the NH<sub>3</sub><sup>+</sup> group and the acetyl C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP48}} = 0.44\text{E-}01$  a.u., HBE = 15.7 kcal/mol),
- H<sub>23</sub> on the NH<sub>3</sub><sup>+</sup> group and the alanine C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP46}} = 0.33\text{E-}01$  a.u., HBE = 11.9 kcal/mol),
- the acetyl C=O oxygen and one of the alanine beta hydrogens,
- the acetyl C=O oxygen and the alanine C=O carbon,
- the acetyl C=O oxygen and one of the lysine delta hydrogens ( $\rho_{\text{BCP49}} = 0.95\text{E-}02$  a.u.),
- the alanine C=O oxygen and one of the alanine beta hydrogens,
- the alanine C=O oxygen and the lysine alpha hydrogen,
- the alanine C=O oxygen and one of the lysine gamma hydrogens ( $\rho_{\text{BCP47}} = 0.11\text{E-}01$  a.u.),
- the carboxyl C=O oxygen and the lysine NH hydrogen ( $\rho_{\text{BCP}} \sim 0.23\text{E-}01$  a.u.),
- the carboxyl C-O(H) oxygen and one of the lysine beta hydrogens,
- the alanine N-H nitrogen and the lysine N-H hydrogen,
- the lysine N-H nitrogen and one of the lysine delta hydrogens ( $\rho_{\text{BCP43}} = 0.76\text{E-}02$  a.u.),
- one of the alanine beta hydrogens and the NH<sub>3</sub><sup>+</sup> group,
- H<sub>23</sub> on the NH<sub>3</sub><sup>+</sup> group and one of the lysine gamma hydrogens.

In C6' of ALysH<sup>+</sup>, there are nine clearly identifiable steric clashes in opened ring structures made of the following groups of atoms and are listed using bullet points:

- from the acetyl C=O oxygen to the alanine C=O carbon,
- from the acetyl C=O oxygen to one of the alanine beta hydrogens,
- from the alanine N-H nitrogen to the lysine N-H hydrogen,
- from the alanine C=O oxygen to one of the alanine beta hydrogens,
- from the alanine C=O oxygen to the lysine alpha hydrogen,
- from the lysine N-H hydrogen to the carboxyl C=O oxygen ( $\rho_{\text{BCP}} \sim 0.24\text{E-}01$  a.u.),
- from the carboxyl C-O(H) oxygen to one of the lysine beta hydrogens,

- from one of the lysine gamma hydrogens to H<sub>23</sub> on the NH<sub>3</sub><sup>+</sup> group,
- from one of the lysine delta hydrogens to H<sub>24</sub> on the NH<sub>3</sub><sup>+</sup> group.

As shown in the RDG isosurface diagrams and listed above, the LysAH<sup>+</sup> conformations have less strong attractive interactions than the ALysH<sup>+</sup> conformations. Comparing the appearance of their isosurface diagrams, it is clear that the LysAH<sup>+</sup> conformations have significantly smaller attractive blue-green regions than the ALysH<sup>+</sup> conformations. In contrast, their repulsive yellow-red regions have a higher dominance than the ALysH<sup>+</sup> conformations. These comparisons suggest less attractive and more repulsive noncovalent interactions in the LysAH<sup>+</sup> conformations. The differences in the noncovalent interactions between these conformations contribute to the lower stability of LysAH<sup>+</sup> conformations.

The two ALysH<sup>+</sup> conformations are structurally similar, with their main differences in the orientation of the lysine sidechain. As a result, their RDG isosurface diagrams are also very similar. A semi-quantitative comparison between these two conformations is discussed below to demonstrate the capability of RDG analysis further.

On one hand, the different orientation changes the strength of the two hydrogen bonds with the NH<sub>3</sub><sup>+</sup> group. However, the HBE<sub>total</sub> values of those two conformations are found nearly equal to each other (within 0.1 kcal/mol), which seem not to contribute to their stability difference.

On the other hand, the orientation change of the lysine sidechain affects the other noncovalent interactions besides the hydrogen bonds. In C1', one of the lysine gamma hydrogens interacts with the acetyl C=O oxygen, and one of the lysine delta hydrogens interacts with the alanine C=O oxygen. Compared to C1', the positions of the lysine gamma and delta carbons in C6' are switched. As a result, the interaction partners of the lysine gamma and delta hydrogens are switched in C1' and C6'. One of the lysine delta hydrogens interacts with the acetyl C=O oxygen, and one of the lysine gamma hydrogens interacts with the alanine C=O oxygen. However, the appearances of those attractive interactions on the isosurface are similar, and the switch doesn't seem to increase the stability difference between these two conformations.

There is an identified hydrogen-hydrogen interaction between one of the lysine delta hydrogens and the lysine alpha hydrogen in C1', which is missing in C6'. The interaction is considered to be weak but still may contribute to their stability difference.

Lastly, the interactions between the lysine N-H group and the carboxyl C=O group are also affected. There is an attractive interaction between the lysine N-H hydrogen and the carboxyl C=O oxygen and a steric clash in the half-ring structure from the lysine N-H hydrogen to the carboxyl C=O oxygen. The overall interaction seems attractive, which is reflected in the larger blue region than the red region on the RDG isosurface plates between those two groups. The geometric information and the electron density of the attractive and repulsive interactions between these two groups are summarized in Table 3.16.

Table 3.16  
*Geometric Information and Electron Density for the Interaction between the Lysine N-H Group and the Carboxyl C=O Group in the ALysH<sup>+</sup> Conformations*

<b>Conf.</b>	<b>N-H --- C-O dihedral angle (°)</b>	<b>N --- C Distance (Å)</b>	<b>H --- O Distance (Å)</b>	<b>H --- O Attractive <math>\rho_{BCP}</math> (a.u.)</b>	<b>H-N --- C-O Repulsive <math>\rho_{BCP}</math> (a.u.)</b>
C1'	14.42	2.388	2.261	0.20E-1	0.20E-1
C6'	7.825	2.379	2.139	0.23E-1	0.23E-1

In C6', the N-H group and C=O group are better aligned in C6' with a smaller dihedral angle of 7.83°. The N --- C distances are comparable between these two conformations, while the H --- O distance is shorter in C6', by 0.12 Å. As a result, the attractive interaction between H and O strengthens with an increased electron density of 0.23E-1 a.u. At the same time, the ring steric clash interaction also enhances. A quantitative estimate of the effect of the geometry change on the strength of both interactions is not available. Because the electron density increases for both attractive and repulsive interactions are the same, and there is no clear change in their RDG isosurface, the combined effect on the overall interaction seems to be negligible.

**Summary of energy characterization.** The energetic characterizations of the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations are summarized in Table 3.17. As the reference, C1' of ALysH<sup>+</sup> has a moderate charge on the NH<sub>3</sub><sup>+</sup> group, a strong  $\mu_{\text{backbone}}$  interaction with NH<sub>3</sub><sup>+</sup> group, a low torsional strain on the lysine sidechain, a high HBE<sub>total</sub>, and high dominance of attractive interactions among all noncovalent interactions. The stability difference between C6' and C1' mainly comes from the torsional strain on the lysine sidechain. The other parameters, including the charge density on the NH<sub>3</sub><sup>+</sup> group,  $\mu_{\text{backbone}}$  interaction with the NH<sub>3</sub><sup>+</sup> group, the HBE<sub>total</sub>, and the dominance of attractive intramolecular interactions, are all comparable.

C2 of LysAH<sup>+</sup> has a higher torsional strain on the lysine sidechain and lower HBE<sub>total</sub>, which brings its relative energy to C1' of ALysH<sup>+</sup> to around 2.2 kcal/mol. Furthermore, its  $\mu_{\text{backbone}}$  has minimum interaction with the NH<sub>3</sub><sup>+</sup> group, its Hirshfeld charge is much higher, and its attractive interaction dominance among the noncovalent interactions is lower. These three differences bring its overall relative energy to be higher than that of C1' by 4.1 kcal/mol.

Compared to C1' of ALysH<sup>+</sup>, C3 of LysAH<sup>+</sup> has a much higher torsional strain on the lysine side chain and a much lower HBE<sub>total</sub>. The sum of those two brings its relative energy to 6.4 kcal/mol. The dominance of attractive interactions in C3 of LysAH<sup>+</sup> is less than that in C1' of ALysH<sup>+</sup>. But the better interaction between  $\mu_{\text{backbone}}$  with the NH<sub>3</sub><sup>+</sup> group and a much lower Hirshfeld charge on the NH<sub>3</sub><sup>+</sup> group bring its relative energy down to the computed value of 4.2 kcal/mol compared to that of C1'.

In summary, the energetic characterizations, including the charge analysis, dipole moment decomposition, torsional strain analysis, hydrogen bonding analysis, and the RDG analysis, provide valuable information about the systems of interest. The torsional strain on the lysine sidechain and HBE<sub>total</sub> seem to contribute the most to the observed stability difference between the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations.

Table 3.17  
 Summary of the Energy Characterizations of the LysAH<sup>+</sup> and ALysH<sup>+</sup> Conformations

	Conf.	$\Delta G$ (kcal/mol)	Hirshfeld Charge (a.u.)	Dot Product of $\mu_{\text{NH}_3}$ and $\mu_{\text{backbone}}$	Rel. E <sub>torsion</sub> (kcal/mol)	Rel. HBE <sub>total</sub> (kcal/mol)
LysAH <sup>+</sup>	C2	4.1	0.466	-2.46	1.2	-1.0
	C3	4.2	0.424	-24.81	4.6	4.2
ALysH <sup>+</sup>	C1'	0.0	0.451	-19.16	0.0	0.0
	C6'	2.1	0.449	-25.27	1.9	0.1

### IR Decomposition

IR decomposition of the representative conformations was conducted to understand the underlying structural features in relation to the differences between the IRMPD spectra of LysAH<sup>+</sup> and ALysH<sup>+</sup>. Detailed procedures are available in Chapter 2. Briefly, the frequency calculation output file at the wB97XD/6-311G\*\* level of theory from Gaussian 09W is fed to the VibAnalysis to decompose all of the computed vibrational modes. Each of the vibrational modes is decomposed into a weighted combination of displacement along chemically meaningful internal coordinates. Then, each of those internal coordinates is manually assigned to the corresponding normal mode of the functional groups. The normal mode of the internal coordinates with the highest weight is set as the primary mode. Detailed analysis and comparison are conducted on the carboxyl C=O stretching region and the amide I and II regions, with structurally characteristic and strong absorption bands. The IRMPD band assignments, the wavenumbers calculated, and the mode descriptions are listed in Table 3.18 and Table 3.19, for C2 and C3 of LysAH<sup>+</sup>, and Table 3.20 and Table 3.21 for C1' and C6' of ALysH<sup>+</sup>, where the primary mode is shown in bold. The structure of the selected conformations of LysAH<sup>+</sup> (top) and ALysH<sup>+</sup> (bottom) are shown in Figure 3.26, with all atoms labeled, for easy correlation between the motion description and the mode assignment.

Additionally, the amide bonds and the O --- H distance in their corresponding hydrogen bonds are summarized in Table 3.22. When the amide bond oxygen accepts a hydrogen bond from and shares its electron density with the NH<sub>3</sub><sup>+</sup> group, the amide C=O bond is weakened, and the amide C(O)-N(H) bond

potentially is strengthened. The change could be reflected in the bond length, bond order, and the relaxed force constant, which are all intrinsic measures of the bond strength. The procedures of obtaining these values are described in detail in Chapter 2. Lastly, the wavenumber of the computed vibrational modes strongly depends on the amide bond strength and could be correlated with and verified by the experimentally observed IRMPD peaks.

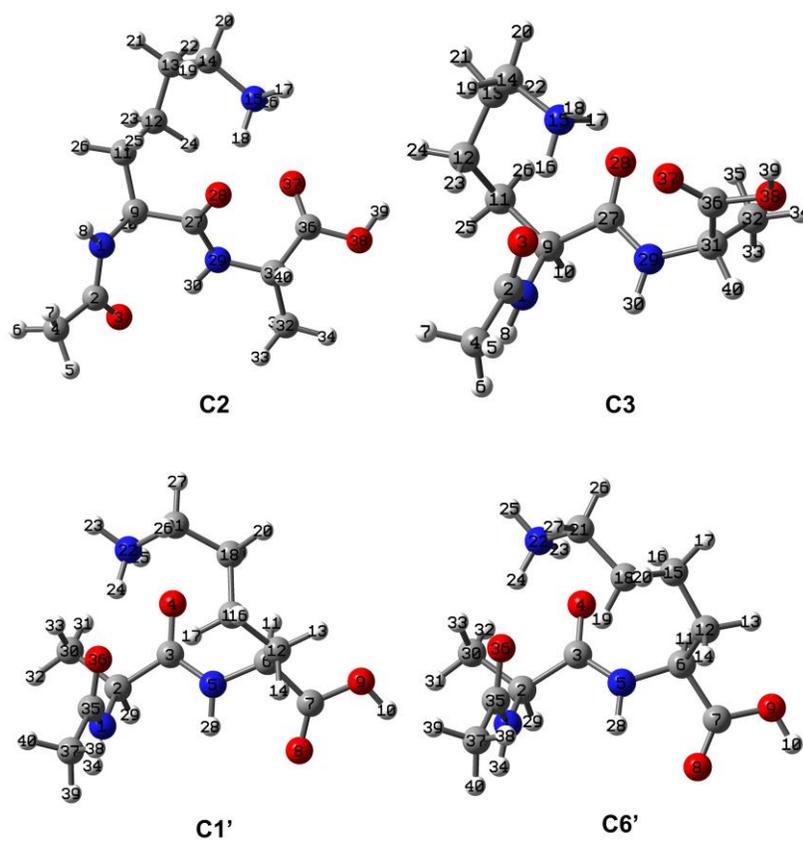


Figure 3.24. Selected conformations of LysAH<sup>+</sup> (top) and ALysH<sup>+</sup> (bottom) with atoms labeled.

Table 3.18

*IRMPD and Assignments and Mode Descriptions with Frequencies Calculated for C2 of LysAH<sup>+</sup>*

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1743	1740	BOND C36 O37	52.7	<b>Carboxyl C=O stretching</b>
		BOND C27 O28	7.4	Lys-Ala Amide I
1695	1688	BOND C2 O3	53.6	<b>Ac-Lys Amide I</b>
		BOND C2 C4	5.9	Ac-Lys Amide I
		BOND C36 O37	5.7	Carboxyl C=O stretching
		BOND C27 O28	5.6	Lys-Ala Amide I
1608	1630	BOND C27 O28	29.8	<b>Lys-Ala Amide I</b>
		BOND C27 N29	10.3	Lys-Ala Amide I
		ANGLE H17 N15 H18	8.5	NH <sub>3</sub> <sup>+</sup> scissoring
	1613	BOND C36 O37	7.5	Carboxyl C=O stretching
		BOND C2 O3	5.5	Ac-Lys Amide I
		BOND C27 O28	22.7	<b>Lys-Ala Amide I</b>
N/A	1583	ANGLE H16 N15 H18	10.1	NH <sub>3</sub> <sup>+</sup> scissoring
		ANGLE H17 N15 H18	9.3	NH <sub>3</sub> <sup>+</sup> scissoring
1539	1541	BOND C27 N29	7.0	Lys-Ala Amide I
		ANGLE H16 N15 H17	23.6	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		ANGLE C14 N15 H18	9.5	NH <sub>3</sub> <sup>+</sup> scissoring
		BOND C27 N29	23.8	<b>Lys-Ala Amide II</b>
		ANGLE C27 N29 H30	10.9	Lys-Ala Amide II
		ANGLE H30 N29 C31	8.1	Lys-Ala Amide II
1507	1496	BOND C9 C27	7.0	Lys-Ala Amide II
		BOND N29 C31	6.5	Lys-Ala Amide II
		BOND C27 O28	6.0	Lys-Ala Amide II
		ANGLE H16 N15 H18	11.4	<b>NH<sub>3</sub><sup>+</sup> umbrella</b>
1460	1470	ANGLE H16 N15 H17	10.0	NH <sub>3</sub> <sup>+</sup> umbrella
		ANGLE H17 N15 H18	9.2	NH <sub>3</sub> <sup>+</sup> umbrella
		ANGLE C14 N15 H18	8.2	NH <sub>3</sub> <sup>+</sup> umbrella
		BOND N1 C2	20.1	<b>Ac-Lys Amide II</b>
1460	1470	ANGLE C2 N1 H8	10.0	Ac-Lys Amide II
		BOND N1 C9	7.6	Ac-Lys Amide II
		ANGLE H8 N1 C9	7.5	Ac-Lys Amide II

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.1).<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.949.<sup>c</sup> Decomposed motions and their percent contributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.18 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
	1367	BOND C36 O38	8.7	<b>Carboxyl C-O-H bending</b>
	1365	BOND C36 O38	7.9	<b>Carboxyl C-O-H bending</b>
1371		BOND C36 O38	12.6	<b>Carboxyl C-O-H bending</b>
	1338	BOND C31 C36	7.9	Carboxyl C-O-H bending
		ANGLE C31 C32 H34	6.8	Ala CH <sub>3</sub> umbrella
		ANGLE H33 C32 H34	5.6	Ala CH <sub>3</sub> umbrella
	1225	BOND C27 N29	10.0	<b>Lys-Ala Amide III</b>
1230		BOND N1 C2	12.3	<b>Ac-Lys Amide III</b>
	1210	BOND C2 O3	7.4	Ac-Lys Amide III
		BOND C2 C4	6.4	Ac-Lys Amide III
1160	1160	BOND C36 O37	6.1	<b>Carboxyl C-O stretching</b>
		BOND C36 O38	5.8	Carboxyl C-O stretching
		BOND C27 N29	5.8	Lys-Ala Amide III
1125	1137	BOND C36 O38	15.4	<b>Carboxyl C-O stretching</b>
		BOND C36 O37	11.2	Carboxyl C-O stretching
		BOND N29 C31	7.9	
		ANGLE C36 O38 H39	6.2	Carboxyl C-O stretching

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.1).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.949.

<sup>c</sup> Decomposed motions and their percent contributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

Table 3.19

*IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C3 of LysAH<sup>+</sup>*

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>	
1762	1766	BOND C36 O37	54.5	<b>Carboxyl C=O stretching</b>	
		BOND C27 O28	9.1	Lys-Ala Amide I	
1695	1692	BOND C27 O28	34.8	<b>Lys-Ala Amide I</b>	
		BOND C2 O3	16.6	Ac-Lys Amide I	
		BOND C36 O37	8.3	Carboxyl C=O stretching	
		BOND C27 N29	6.5	Lys-Ala Amide I	
	1675	BOND C2 O3	39.8	<b>Ac-Lys Amide I</b>	
		BOND C27 O28	18.3	Lys-Ala Amide I	
1608	1620	ANGLE H16 N15 H18	18.6	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>	
		ANGLE H16 N15 H17	18.3	NH <sub>3</sub> <sup>+</sup> scissoring	
		BOND C27 O28	15.1	Lys-Ala Amide I	
		BOND C2 O3	6.3	Ac-Lys Amide I	
N/A	1566	ANGLE H17 N15 H18	15.1	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>	
		BOND C27 O28	12.1	Lys-Ala Amide I	
		ANGLE C14 N15 H16	9.8	NH <sub>3</sub> <sup>+</sup> scissoring	
		ANGLE H16 N15 H18	8.5	NH <sub>3</sub> <sup>+</sup> scissoring	
		ANGLE H16 N15 H17	8.0	NH <sub>3</sub> <sup>+</sup> scissoring	
		1508	BOND C27 N29	18.1	<b>Lys-Ala Amide II</b>
			ANGLE C27 N29 H30	7.8	Lys-Ala Amide II
	ANGLE H30 N29 C31	6.5	Lys-Ala Amide II		
1507	1500	BOND N1 C2	21.9	<b>Ac-Lys Amide II</b>	
		ANGLE C2 N1 H8	9.3	Ac-Lys Amide II	
		ANGLE H8 N1 C9	9.2	Ac-Lys Amide II	
		1488	ANGLE H17 N15 H18	14.6	<b>NH<sub>3</sub><sup>+</sup> umbrella</b>
		ANGLE H16 N15 H18	9.6	NH <sub>3</sub> <sup>+</sup> umbrella	
		BOND C27 N29	9.1	Lys-Ala Amide II	
		ANGLE H16 N15 H17	8.0	NH <sub>3</sub> <sup>+</sup> umbrella	

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.1).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.949.

<sup>c</sup> Decomposed motions and their percent contributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.19 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>	
1371	1377	BOND C36 O38	13.7	<b>Carboxyl C-O-H bending</b>	
		BOND C31 C36	7.6	Carboxyl C-O-H bending	
		ANGLE C36 C31 H40	6.2	Carboxyl C-O-H bending	
		ANGLE O37 C36 O38	5.6	Carboxyl C-O-H bending	
1230	1248	BOND N1 C2	20.6	<b>Ac-Lys Amide III</b>	
		BOND C2 O3	11.8	Ac-Lys Amide III	
		BOND C2 C4	8.6	Ac-Lys Amide III	
		ANGLE C2 N1 H8	7.5	Ac-Lys Amide III	
		BOND N1 C9	6.1	Ac-Lys Amide III	
		1239	1239	BOND C27 N29	7.3
BOND C27 O28	6.0			Lys-Ala Amide III	
1217	BOND C27 N29			12.0	<b>Lys-Ala Amide III</b>
	BOND N29 C31			7.5	Lys-Ala Amide III
1160	1155	BOND C36 O38	14.5	<b>Carboxyl C-O stretching</b>	
		BOND C36 O37	11.3	Carboxyl C-O stretching	
		ANGLE C36 O38 H39	8.3	Carboxyl C-O stretching	
1125	1126	BOND C36 O38	9.9	<b>Carboxyl C-O stretching</b>	
		BOND N29 C31	7.7		
		BOND C36 O37	6.4	Carboxyl C-O stretching	
		BOND C9 C27	6.0		
		BOND C11 C12	5.7		

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.1).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.949.

<sup>c</sup> Decomposed motions and their percent contributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

Table 3.20  
 IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C1' of ALysH<sup>+</sup>

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1768	1774	BOND C7 O8	70.9	<b>Carboxyl C=O stretching</b>
		BOND C7 O9	6.8	Carboxyl C=O stretching
1675	1679	BOND C3 O4	30.5	<b>Ala-Lys Amide I</b>
		BOND C35 O36	22.0	Ac-Ala Amide I
		BOND C3 N5	6.6	Ala-Lys Amide I
1656	1650	BOND C35 O36	35.5	<b>Ac-Ala Amide I</b>
		BOND C3 O4	21.8	Ala-Lys Amide I
		BOND C3 N5	6.5	Ala-Lys Amide I
1597	1607	ANGLE H23 N22 H24	22.6	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		BOND C3 O4	16.6	Ala-Lys Amide I
		ANGLE H24 N22 H25	12.3	NH <sub>3</sub> <sup>+</sup> scissoring
		ANGLE C21 N22 H25	7.1	NH <sub>3</sub> <sup>+</sup> scissoring
N/A	1562	ANGLE H23 N22 H25	14.3	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		BOND C3 O4	13.5	Ala-Lys Amide I
		ANGLE H24 N22 H25	11.6	NH <sub>3</sub> <sup>+</sup> scissoring
		ANGLE C21 N22 H24	7.8	NH <sub>3</sub> <sup>+</sup> scissoring
		ANGLE H23 N22 H24	6.0	NH <sub>3</sub> <sup>+</sup> scissoring
1508	1521	BOND C3 N5	22.7	<b>Ala-Lys Amide II</b>
		ANGLE C6 N5 H28	10.3	Ala-Lys Amide II
		ANGLE C3 N5 H28	5.6	Ala-Lys Amide II
	1511	ANGLE C21 N22 H25	16.1	<b>NH<sub>3</sub><sup>+</sup> umbrella</b>
		ANGLE C21 N22 H24	14.3	NH <sub>3</sub> <sup>+</sup> umbrella
		ANGLE C21 N22 H23	12.5	NH <sub>3</sub> <sup>+</sup> umbrella
		BOND N1 C35	8.0	Ac-Ala Amide II
	1499	BOND N1 C35	20.5	<b>Ac-Ala Amide II</b>
		ANGLE C2 N1 H34	8.8	Ac-Ala Amide II
		BOND C35 C37	6.9	
BOND N1 C2		6.3	Ac-Ala Amide II	
ANGLE H34 N1 C35		5.8	Ac-Ala Amide II	

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.1).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.949.

<sup>c</sup> Decomposed motions and their percent contributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.20 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1380	1366	BOND C7 O9	18.8	<b>Carboxyl C-O-H bending</b>
		BOND C6 C7	9.3	Carboxyl C-O-H bending
		BOND C3 N5	7.9	
		ANGLE C7 O9 H10	7.8	Carboxyl C-O-H bending
		ANGLE O8 C7 O9	6.8	Carboxyl C-O-H bending
1265	1246	BOND N1 C35	18.0	<b>Ac-Ala Amide III</b>
		BOND C35 O36	11.4	Ac-Ala Amide III
		BOND C35 C37	8.1	Ac-Ala Amide III
		ANGLE H34 N1 C35	7.9	Ac-Ala Amide III
		BOND N1 C2	5.6	Ac-Ala Amide III
	1238	BOND C3 N5	7.9	<b>Ala-Lys Amide III</b>
1224	1224	BOND C3 N5	12.3	<b>Ala-Lys Amide III</b>
		BOND C3 O4	7.5	Ala-Lys Amide III
1146	1153	BOND C7 O9	14.4	<b>Carboxyl C-O stretching</b>
		BOND C7 O8	10.3	Carboxyl C-O stretching
		ANGLE C7 O9 H10	5.7	Carboxyl C-O stretching
		BOND N5 C6	5.6	
1126	1126	BOND C7 O9	9.6	<b>Carboxyl C-O stretching</b>
		BOND N5 C6	9.0	
		BOND C2 C30	6.5	

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.1).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.949.

<sup>c</sup> Decomposed motions and their percent contributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

Table 3.21

*IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C6' of ALysH<sup>+</sup>*

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1768	1768	BOND C7 O8	68.7	<b>Carboxyl C=O stretching</b>
		BOND C7 O9	6.7	Carboxyl C=O stretching
1675	1681	BOND C3 O4	33.0	<b>Ala-Lys Amide I</b>
		BOND C35 O36	17.4	Ac-Ala Amide I
		BOND C3 N5	7.7	Ala-Lys Amide I
		ANGLE H23 N22 H24	5.6	NH <sub>3</sub> <sup>+</sup> scissoring
1656	1653	BOND C35 O36	37.2	<b>Ac-Ala Amide I</b>
		BOND C3 O4	17.4	Ala-Lys Amide I
1597	1601	BOND C3 O4	21.3	<b>Ala-Lys Amide I</b>
		ANGLE H23 N22 H24	21.0	NH <sub>3</sub> <sup>+</sup> scissoring
		ANGLE H24 N22 H25	10.9	NH <sub>3</sub> <sup>+</sup> scissoring
		ANGLE H23 N22 H25	5.7	NH <sub>3</sub> <sup>+</sup> scissoring
1577	1577	ANGLE H23 N22 H25	15.0	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		ANGLE H24 N22 H25	11.9	NH <sub>3</sub> <sup>+</sup> scissoring
		BOND C3 O4	10.5	Ala-Lys Amide I
		BOND C35 O36	8.0	Ac-Ala Amide I
		ANGLE C21 N22 H23	6.5	NH <sub>3</sub> <sup>+</sup> scissoring
		ANGLE C21 N22 H24	6.1	NH <sub>3</sub> <sup>+</sup> scissoring

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.1).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.949.

<sup>c</sup> Decomposed motions and their percent contributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.21 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
		ANGLE C21 N22 H24	12.2	<b>NH<sub>3</sub><sup>+</sup> umbrella</b>
	1514	ANGLE C21 N22 H23	10.4	NH <sub>3</sub> <sup>+</sup> umbrella
		ANGLE C21 N22 H25	8.0	NH <sub>3</sub> <sup>+</sup> umbrella
		BOND N1 C35	5.9	Ac-Ala Amide II
		BOND N1 C35	18.2	<b>Ac-Ala Amide II</b>
1508	1501	BOND C3 N5	8.1	Ala-Lys Amide II
		ANGLE H34 N1 C35	6.7	Ac-Ala Amide II
		ANGLE C21 N22 H24	5.6	NH <sub>3</sub> <sup>+</sup> umbrella
		BOND C3 N5	15.7	<b>Ala-Lys Amide II</b>
	1496	BOND N1 C35	8.5	Ac-Ala Amide II
		ANGLE C3 N5 H28	6.6	Ala-Lys Amide II
		BOND N5 C6	6.0	Ala-Lys Amide II
		BOND C7 O9	16.3	<b>Carboxyl C-O-H bending</b>
		BOND C6 C7	7.1	Carboxyl C-O-H bending
1380	1369	ANGLE O8 C7 O9	7.1	Carboxyl C-O-H bending
		ANGLE C7 O9 H10	6.3	Carboxyl C-O-H bending
		BOND C3 N5	6.0	
		ANGLE N5 C6 H11	5.9	

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.1).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.949.

<sup>c</sup> Decomposed motions and their percent contributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.21 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1265	1249	BOND N1 C35	21.7	<b>Ac-Ala Amide III</b>
		BOND C35 O36	13.3	Ac-Ala Amide III
		BOND C35 C37	8.8	Ac-Ala Amide III
		ANGLE C2 N1 H34	8.4	Ac-Ala Amide III
		ANGLE H34 N1 C35	6.2	Ac-Ala Amide III
1236	BOND C3 N5	5.2	<b>Ala-Lys Amide III</b>	
1227	BOND C3 N5	6.5	<b>Ala-Lys Amide III</b>	
	BOND N5 C6	5.7	Ala-Lys Amide III	
1146	1155	BOND C7 O9	15.6	<b>Carboxyl C-O stretching</b>
		BOND C7 O8	12.0	Carboxyl C-O stretching
		ANGLE C7 O9 H10	7.3	Carboxyl C-O stretching
1126	1123	BOND N5 C6	10.1	
		BOND C7 O9	7.8	Carboxyl C-O stretching
		BOND C2 C30	6.8	
		BOND N1 C2	6.0	

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.1).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.949.

<sup>c</sup> Decomposed motions and their percent contributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

Table 3.22

Summary of Amide Bond, Carboxyl Group, and Corresponding Hydrogen Bond Information in the LysAH<sup>+</sup> and ALysH<sup>+</sup> Conformations

Conf.	C(=O)-N(H) Bond				C=O Bond				O --- H Distance <sup>a</sup> (Å)	
	Bond Length <sup>a</sup> (Å)	Bond Order <sup>b</sup>	Relaxed Force Constant <sup>c</sup> (cm <sup>-1</sup> )	Wavenumber <sup>a</sup> (cm <sup>-1</sup> )	Bond Length <sup>a</sup> (Å)	Bond Order <sup>b</sup>	Relaxed Force Constant <sup>c</sup> (cm <sup>-1</sup> )	Wavenumber <sup>a</sup> (cm <sup>-1</sup> )		
LysAH <sup>+</sup>	Ac-Lys	1.366	1.153	6.135	1470	1.221	1.859	11.364	1688	
	C2 Lys-Ala	1.330	1.288	7.692	1541	1.238	1.648	10.417	1630	1.584
	COOH					1.208	1.935	12.500	1740	2.261
C3 Lys-Ala	Ac-Lys	1.357	1.189	6.536	1500	1.226	1.748	11.494	1675	1.684
	COOH	1.347	1.217	7.092	1508	1.222	1.812	11.364	1692	1.961
						1.204	1.964	12.987	1766	
C1' Ala-Lys	Ac-Lys	1.351	1.219	6.849	1499	1.231	1.716	11.236	1650	1.745
	COOH	1.339	1.261	7.299	1521	1.229	1.751	10.989	1679	1.777
						1.201	2.005	13.158	1774	
ALysH <sup>+</sup>	Ac-Lys	1.350	1.216	6.849	1501	1.231	1.715	11.111	1653	1.701
	COOH	1.343	1.248	6.897	1496	1.227	1.777	10.989	1681	1.843
						1.202	1.986	12.987	1768	

<sup>a</sup> Values obtained from the log file of frequency calculations at wB97XD/6-311G\*\* level of theory; the wavenumber is scaled by 0.949.<sup>b</sup> Mayer bond order calculated by Multwfn 3.8 using the fchk file from single-point calculations at B3LYP-D3(BJ)/ma-def2-TZVPP level of theory.<sup>c</sup> Values (mdyn/Å) calculated by Compliance 3.0.2 using log file from frequency calculations at wB97XD/6-311G\*\* level of theory.

**Carboxyl C=O stretching band.** As shown in the IRMPD spectra (Figure 3.1), the carboxyl C=O stretching mode of LysAH<sup>+</sup> has a peak at 1743 cm<sup>-1</sup> and a shoulder peak at 1762 cm<sup>-1</sup>, which both are red-shifted compared to the single peak at 1768 cm<sup>-1</sup> for ALysH<sup>+</sup>. The interactions of the carboxyl group in the LysAH<sup>+</sup> conformations well correlate with the characteristic redshift of the two peaks in its IRMPD spectrum.

In C2 of LysAH<sup>+</sup>, the carboxyl C=O stretching is the primary mode of the computed peak at 1740 cm<sup>-1</sup>, which matches the IRMPD peak at 1743 cm<sup>-1</sup> (Table 3.18). The carboxyl C=O group accepts a hydrogen bond from the NH<sub>3</sub><sup>+</sup> group with an O---H distance of 2.26 Å (Figure 3.5) and an estimated HBE of 5.4 kcal/mol. So, the hydrogen bonding elongates the carboxyl C=O bond to 1.208 Å (Table 3.22), longer than the other three LysAH<sup>+</sup> and ALysH<sup>+</sup> representative conformations. The weakening of its carboxyl C=O bond is also reflected in its bond order of 1.935 and relaxed force constant of 12.500, the lowest among the LysAH<sup>+</sup> and ALysH<sup>+</sup> representative conformations. The weak carboxyl C=O bond down shifts its wavenumber to 1740 cm<sup>-1</sup>, the lowest among the LysAH<sup>+</sup> and ALysH<sup>+</sup> selected conformations.

In C3 of LysAH<sup>+</sup>, the carboxyl C=O stretching mode is the primary mode of the computed peak at 1766 cm<sup>-1</sup>, which matches the unresolved higher wavenumber shoulder peak at 1762 cm<sup>-1</sup> (Table 3.19). Though there is not an identifiable hydrogen bond, its carboxyl C=O group is placed towards and stabilizes the NH<sub>3</sub><sup>+</sup> group (Figure 3.5). So, its carboxyl C=O bond becomes elongated to 1.204 Å (Table 3.22), the second-longest among the LysAH<sup>+</sup> and ALysH<sup>+</sup> representative conformations. Its carboxyl C=O group bond order and relaxed force constant are 1.964 and 12.987, the second-lowest among the LysAH<sup>+</sup> and ALysH<sup>+</sup> representative conformations. As a result, its wavenumber is also red-shifted but by a smaller amount than C2 of LysAH<sup>+</sup>.

Based on the decomposition results, the carboxyl C=O stretching mode is coupled with a lower frequency mode, the Lys-Ala amide I mode, in both LysAH<sup>+</sup> conformations. The percent contributions from the carboxyl C=O stretching mode and the Lys-Ala amide I mode are 52.7% and 7.4% in C2 (Table 3.18), and 54.5% and 9.1% in C3 (Table 3.19), respectively. The coupling seems transmitted through the

interactions of the  $\text{NH}_3^+$  group with the carboxyl C=O group and the Lys-Ala amide group. The stronger hydrogen bonding interaction in C2 might have allowed it to have a better coupling than C3.

In the  $\text{ALysH}^+$  conformations, the carboxyl group is not involved in hydrogen bonding with the highly charged  $\text{NH}_3^+$  group but rather hanging on the edge freely (Figure 3.7). So, their carboxyl C=O bonds are not elongated, and their wavenumbers are less down-shifted. The carboxyl C=O bond length in C1' and C6' of  $\text{ALysH}^+$  are 1.201 and 1.202 Å (Table 3.22), shorter than that in the  $\text{LysAH}^+$  conformations. Their bond orders and relaxed force constants are both higher than those in the  $\text{LysAH}^+$  conformations. In C1' and C6', the carboxyl C=O stretching mode is the primary mode of the computed peaks at 1774 and 1768  $\text{cm}^{-1}$  (Table 3.20 and Table 3.21), respectively, which both match the IRMPD peak at 1768  $\text{cm}^{-1}$ , and are also blue-shifted compared to those in the  $\text{LysAH}^+$  conformations.

Additionally, because of the isolation of the carboxyl group, there is not an identifiable coupling between an amide I mode and the carboxyl C=O stretching mode in the  $\text{ALysH}^+$  conformations. The motions identified from IR decomposition are the carboxyl C=O stretching motion and the carboxyl C-O(H) stretching motion, with their weight of 70.9% and 6.8% in C1', and 68.7% and 6.7% in C6'. The total contributions add up to 77.7% in C1' and 75.4% in C6'. So, the carboxyl C=O stretching mode is more independent in the  $\text{ALysH}^+$  conformations than in the  $\text{LysAH}^+$  conformations.

**Amide I band.** In the amide I region of the IRMPD spectra, LysAH<sup>+</sup> has two widely separated peaks at 1695 and 1608 cm<sup>-1</sup>, while ALysH<sup>+</sup> has two closely positioned peaks at 1675 and 1656 cm<sup>-1</sup>, and one low-intensity hump at 1597 cm<sup>-1</sup>. In this region, C3 has a good match to the LysAH<sup>+</sup> IRMPD spectrum, while C1' and C6' both have a good match to the ALysH<sup>+</sup> IRMPD spectrum. Based on the IR decomposition results in this region, there are four normal vibrational modes: the amide I mode of the two amide bonds and the two asymmetric (scissoring) vibrational modes of the NH<sub>3</sub><sup>+</sup> group. The final set of vibrational modes are the results of complex coupling between these four normal modes. Additionally, the C=O stretching is the primary displacement of the internal coordinate in the amide I mode. So, the strength of the amide C=O bond is discussed in detail to explain the frequency variations between the amide I modes.

In C2 of LysAH<sup>+</sup>, the Ac-Lys amide I mode is the primary mode of the computed peak at 1688 cm<sup>-1</sup>, which matches the experimental peak at 1695 cm<sup>-1</sup> (Table 3.18). It appears to be a significant blue shift. Although it is hydrogen-bonded with the Lys-Ala amide NH group, it is not involved in direct strong hydrogen bonding with the charged NH<sub>3</sub><sup>+</sup> group as in C3 of LysAH<sup>+</sup> and the ALysH<sup>+</sup> conformations. The C=O bond length in the Ac-Lys amide group is 1.221 Å (Table 3.22), the shortest acetyl C=O bonds in all four representative conformations. Its bond order is 1.859, the highest in the acetyl C=O bonds of the LysAH<sup>+</sup> and ALysH<sup>+</sup> representative conformations.

The Lys-Ala amide bond interacts with the NH<sub>3</sub><sup>+</sup> group through a very strong hydrogen bond, which elongates its C=O bond to 1.238 Å, the longest among the LysAH<sup>+</sup> and ALysH<sup>+</sup> representative conformations. The interaction reduces its bond order and relaxed force constant to 1.648 and 10.417, respectively, the lowest among the LysAH<sup>+</sup> and ALysH<sup>+</sup> representative conformations. The Lys-Ala amide I mode is the primary mode of the computed peaks at 1630 and 1613 cm<sup>-1</sup>. In these two vibrational modes, the Lys-Ala amide bond C=O stretching weights are 29.8% and 22.7%, respectively. Taking other Lys-Ala amide I motions into account, the total weights of the Lys-Ala amide I mode in the two computed peaks are 40.1% and 29.7%, respectively. There are also contributions from the NH<sub>3</sub><sup>+</sup> scissoring modes. In the two vibrational modes, the contributions from NH<sub>3</sub><sup>+</sup> scissoring are 8.5% and

19.4%, respectively. This strong coupling is caused by a strong hydrogen bond between the  $\text{NH}_3^+$  group and the Lys-Ala amide bond.

On one hand, the hydrogen bond between the Lys-Ala amide C=O group and the  $\text{NH}_3^+$  group is strong, with an estimated HBE of 21.2 kcal/mol. The strong hydrogen bond ties these two groups together. On the other hand, it also significantly downshifts the Lys-Ala amide I mode, to a frequency close to the  $\text{NH}_3^+$  scissoring mode, allowing a strong coupling. As a result, the Lys-Ala amide I mode is assigned as the primary vibrations of the two computed peaks, though the  $\text{NH}_3^+$  scissoring mode significantly contributes to the computed peak at  $1613\text{ cm}^{-1}$ . The  $\text{NH}_3^+$  scissoring mode is assigned as the primary mode of the computed peak at  $1583\text{ cm}^{-1}$ , which has low absorption and does not correspond to any of the experimentally observed peaks.

In C3 of  $\text{LysAH}^+$ , the Ac-Lys and Lys-Ala amide bonds interact with the  $\text{NH}_3^+$  group through hydrogen bonds with HBE of 16.4 and 9.4 kcal/mol, respectively (Figure 3.5). The Ac-Lys amide group C=O bond is  $1.226\text{ \AA}$ , longer than that in the Lys-Ala amide group. The Ac-Lys and the Lys-Ala amide I modes are the primary mode of the computed peaks at  $1675$  and  $1692\text{ cm}^{-1}$ , respectively (Table 3.18). The bond order of the Ac-Lys amide group C=O bond is predicted to be 1.748, lower than that in the Lys-Ala amide group. The data matches the prediction that stronger hydrogen bonds with the amide C=O group result in longer C=O bonds and lower wavenumber and lower bond order. This feature is also observed for the Ac-Ala and Ala-Lys amide bonds in the  $\text{ALysH}^+$  conformations. Lastly, the relaxed force constants seem not to reflect the bond strength differences correctly. In C3 of  $\text{LysAH}^+$ , and C1' and C6' of  $\text{ALysH}^+$ , the acetyl C=O bond has a higher relaxed force constant than the other amide bond, contradicting the interpretation from the bond length, bond order, and mode wavenumber.

In C3 of  $\text{LysAH}^+$ , a strong coupling between the Ac-Lys and Lys-Ala amide I modes is observed. Summing the weights belonging to each normal mode together, the ratios of contribution from Ac-Lys and Lys-Ala amide I mode to the computed peaks at  $1675$  and  $1692\text{ cm}^{-1}$  are 39.8%:18.3%, and 16.6%:41.3%, respectively.

In the computed mode at wavenumber  $1675\text{cm}^{-1}$ , the Ac-Lys and Lys-Ala amide I mode are out-of-phase, meaning that as one C=O bond extends, the other C=O retracts. At the same time, the isolated hydrogen and the hydrogen interacting with the Lys-Ala amide group vibrate symmetrically, making the  $\text{NH}_3^+$  scissoring mode. In the computed peak at wavenumber  $1692\text{ cm}^{-1}$ , the Ac-Lys and Lys-Ala amide I mode are in-phase, meaning that as one C=O bond extends, the other C=O also extends. The two hydrogens interacting with those two amide bonds on the  $\text{NH}_3^+$  group vibrate symmetrically, making the  $\text{NH}_3^+$  scissoring mode.

Multiple factors might contribute to the strong coupling. The amide I coupling is transmitted through the  $\text{NH}_3^+$  bridge made of the  $\text{NH}_3^+$  group and the hydrogen bonds, where the  $\text{NH}_3^+$  group acts as the gatekeeper. Efficient coupling happens with the following two criteria: both amide I modes are well coupled with the  $\text{NH}_3^+$  scissoring mode, and the motion is transmitted to the other side through the  $\text{NH}_3^+$  scissoring mode. For good coupling with the  $\text{NH}_3^+$  scissoring mode, the strength of the hydrogen bond plays an important role. A stronger hydrogen bond ties the amide C=O oxygen and the  $\text{NH}_3^+$  hydrogen together and allows a stronger coupling between amide I and  $\text{NH}_3^+$  scissoring modes. The geometrical positions and orientations of amide C=O groups relative to the  $\text{NH}_3^+$  group are important for the transmission efficiency of coupling through the  $\text{NH}_3^+$  scissoring mode. The parallel and closely positioned amide C=O groups in C3 seem to allow efficient transmission of in-phase vibrations through the  $\text{NH}_3^+$  bridge using the  $\text{NH}_3^+$  scissoring motions. Because there is a stronger coupling observed in the computed vibrational mode at  $1692\text{ cm}^{-1}$  than in the mode at  $1675\text{ cm}^{-1}$ . Other geometries might be more beneficial for the out-of-phase coupling. Additionally, the two hydrogen bonds of similar strength in C3 down shift their wavenumbers similarly, so their frequencies are close and could be coupled more easily. The similarity in hydrogen bond strength also contributes to the strong coupling.

The  $\text{NH}_3^+$  scissoring mode has been identified as the primary mode of the computed peaks at  $1566$  and  $1620\text{ cm}^{-1}$ , the  $1620\text{ cm}^{-1}$  matches the IRMPD peak at  $1608\text{ cm}^{-1}$ . Their computed absorptions are small, so these two modes seem not contributing to the experimental absorption at  $1608\text{ cm}^{-1}$  in the IRMPD spectrum.

In C1' of ALysH<sup>+</sup>, the Ac-Ala and Ala-Lys amide I mode are the primary mode of the computed peaks at 1650 and 1679 cm<sup>-1</sup>, respectively, which match the IRMPD peaks at 1656 and 1675 cm<sup>-1</sup> (Table 3.20). Similarly, in C6', the Ac-Ala and Ala-Lys amide I mode are the primary mode of the computed peaks at 1653 and 1681 cm<sup>-1</sup>, respectively, which match the IRMPD peaks at 1656 and 1675 cm<sup>-1</sup> (Table 3.21). The computed separations between the two vibrational modes are 29 and 28 cm<sup>-1</sup> for C1' and C6', respectively, which are both larger than the experimentally observed separations of 19 cm<sup>-1</sup>. As shown before, the conformation C1' and C6' are very similar in backbone structure and hydrogen bonding interactions, with the main differences in the orientation of the lysine sidechain. As a result, both conformations have a great match to the IRMPD spectrum in the amide I region.

As discussed in the C3 of LysAH<sup>+</sup> section, the stronger hydrogen bonding to the Ac-Ala amide group results in a slightly longer C=O bond than the Ala-Lys amide group. The change is further reflected in the lower bond order of the C=O bond of the Ac-Ala amide group (Table 3.22). Similar to the observation in C3 of LysAH<sup>+</sup>, strong couplings between the two amide I modes are observed in the two ALysH<sup>+</sup> conformations. They could be rationalized with the same geometrical and energetic reasons. Summing the weights belonging to each normal mode together, for C1' of ALysH<sup>+</sup>, the ratio of contributions from Ac-Ala and Ala-Lys amide I mode to the computed peaks at 1650 and 1679 cm<sup>-1</sup> are 35.5%:28.3%, and 22.0%:37.1%, respectively. For conformation C6' of ALysH<sup>+</sup>, the ratio of contributions from Ac-Ala and Ala-Lys amide I mode to the computed peaks at 1653 and 1681 cm<sup>-1</sup>, are 37.2%:17.4%, and 17.4%:40.7%, respectively.

Like in C3 of LysAH<sup>+</sup>, the Ac-Ala and Ala-Lys amide I mode are in-phase in the computed mode at higher wavenumber and out-of-phase in the computed mode at lower wavenumber. For the higher wavenumber vibrational mode, the two hydrogens on the NH<sub>3</sub><sup>+</sup> group interacting with the two amide bonds vibrate simultaneously, making the NH<sub>3</sub><sup>+</sup> scissoring mode. For the lower wavenumber vibrational mode, the same two hydrogens on the NH<sub>3</sub><sup>+</sup> group vibrate simultaneously, making the NH<sub>3</sub><sup>+</sup> scissoring mode in C1'. While in C6', it is the isolated hydrogen and the hydrogen interacting with the Ala-Lys amide group vibrate simultaneously, making the NH<sub>3</sub><sup>+</sup> scissoring mode. Like in C3 of LysAH<sup>+</sup>, the

coupling is slightly stronger in the in-phase vibrational mode than in the out-of-phase mode. Lastly, compared to C6', the stronger coupling in C1' might be contributed from the closer strength of the two hydrogen bonds.

In C1' of ALysH<sup>+</sup>, the NH<sub>3</sub><sup>+</sup> scissoring is the primary mode of the computed peaks at 1562 and 1607 cm<sup>-1</sup>, the latter peak matches the IRMPD hump at 1597 cm<sup>-1</sup>. In C6' of ALysH<sup>+</sup>, the Ala-Lys amide I mode is the primary mode of the computed peak at 1601 cm<sup>-1</sup>, for its weight of 21.3%. But as shown in the decomposition result, all other internal coordinates correspond to the NH<sub>3</sub><sup>+</sup> scissoring mode, and the highest weight is 21.0%, the same as the Ala-Lys amide I mode. For this reason, the NH<sub>3</sub><sup>+</sup> scissoring mode should be identified as the primary mode for the computed peak at 1601 cm<sup>-1</sup>, which matches the IRMPD hump at 1597 cm<sup>-1</sup>. Additionally, the NH<sub>3</sub><sup>+</sup> scissoring mode is also the primary mode for the computed peak at 1577 cm<sup>-1</sup>.

In C1' of ALysH<sup>+</sup>, both NH<sub>3</sub><sup>+</sup> scissoring modes have high absorptions, while in C6' of ALysH<sup>+</sup>, only the high-frequency mode has high absorption. Based on the IRMPD spectrum of ALysH<sup>+</sup>, the NH<sub>3</sub><sup>+</sup> scissoring modes correspond to the low-intensity hump at 1597 cm<sup>-1</sup>. So experimentally, the NH<sub>3</sub><sup>+</sup> scissoring modes are with low absorption.

**Amide II band.** In the amide II region of the IRMPD spectra, LysAH<sup>+</sup> has one broad peak centered at 1507 cm<sup>-1</sup> and two shoulders at 1460 and 1539 cm<sup>-1</sup>, while ALysH<sup>+</sup> has one sharp peak at 1508 cm<sup>-1</sup>. For LysAH<sup>+</sup> in this region, the computed peaks of C2 match the two shoulder peaks, while the computed peaks of C3 match the peak center. For ALysH<sup>+</sup>, both C1' and C6' have a good match to the experimental spectrum. Based on the decomposition results, there are three normal vibrational modes in this region: the amide II mode of the two amide bonds, and the symmetric (umbrella) vibrational mode of the NH<sub>3</sub><sup>+</sup> group. Compared to the strong coupling in the amide I region, the amide II modes and the NH<sub>3</sub><sup>+</sup> symmetric (umbrella) vibrational mode are mostly independent.

Additionally, compared to the high weight of C=O stretching in the amide I modes, the weight of the C(O)-N(H) stretching in amide II modes is much lower. There are more internal coordinates involved in the amide II mode. As a result, though the C(O)-N(H) stretching is one of the major displacements in

the amide II mode, the correlation of its strength with the amide II wavenumber is not as strong as the correlation of the C=O bond strength with the amide I wavenumber. This dispersed nature might also contribute to the lower coupling between amide II modes.

Lastly, the computed absorption of the  $\text{NH}_3^+$  umbrella vibrational mode is usually lower than the two amide II modes and is always buried under the amide II peaks. Likely, it does not contribute to the observed band in the IRMPD spectrum and is not discussed in detail.

In C2 of  $\text{LysAH}^+$ , the Ac-Lys and Lys-Ala amide II modes are the primary mode of the computed peaks at 1470 and 1541  $\text{cm}^{-1}$ , matching the shoulder peaks at 1460 and 1539  $\text{cm}^{-1}$ , respectively. The  $\text{NH}_3^+$  umbrella vibrational mode is the primary mode of the computed peak at 1496  $\text{cm}^{-1}$ , which matches the peak center at 1507  $\text{cm}^{-1}$ . The decomposition results and investigation of the vibrational modes through GaussView show that no coupling is identifiable between the two amide II modes and the  $\text{NH}_3^+$  umbrella vibrational mode.

Compared to the hydrogen-bonded acetyl C=O groups in C3 of  $\text{LysAH}^+$  and the  $\text{ALysH}^+$  conformations, the isolation of the acetyl C=O bond in C2 grants it a high strength C=O bond and a low strength C(O)-N(H) bond. As a result, the Ac-Lys amide II mode has the lowest wavenumber compared to the other representative conformations of  $\text{LysAH}^+$  and  $\text{ALysH}^+$ . Conversely, the strong hydrogen bond involving the Lys-Ala amide bond weakens the C=O bond and strengthens the C(O)-N(H) bond. Additionally, the N-H bending motion of the Lys-Ala amide II mode is limited because of its hydrogen bond with the Ac-Lys amide C=O oxygen. As a result, the Lys-Ala amide II mode has the highest wavenumber among the representative conformations of  $\text{LysAH}^+$  and  $\text{ALysH}^+$ . The bond length, bond order, and relaxed force constants of the two C(O)-N(H) bonds correlate with their strength.

In C3 of  $\text{LysAH}^+$ , the Ac-Lys and Lys-Ala amide II modes are the primary mode of the computed peaks at 1500 and 1508  $\text{cm}^{-1}$ , respectively, which both match the peak center at 1507  $\text{cm}^{-1}$ . The decomposition results and investigation of the vibrational modes through GaussView show a coupling between the Lys-Ala amide II mode and the  $\text{NH}_3^+$  umbrella vibrational mode, and no coupling is identifiable for the Ac-Lys amide II mode. In C3 of  $\text{LysAH}^+$ , the lower wavenumber of the Ac-Lys

amide II mode does not reflect the strengthening of its C(O)-N(H) bond through a stronger hydrogen bond with the  $\text{NH}_3^+$  group. Moreover, compared to the Lys-Ala amide bond, its C(O)-N(H) bond is longer, and the bond order and the relaxed force constants are lower. It is also observed in the  $\text{ALysH}^+$  conformations. Likely, the substitution groups on both sides of the amide group cause the terminal amide bonds and in-chain amide bonds to have different base C(O)-N(H) bond strength and amide II frequencies. And the hydrogen bonds with the amide oxygen are not strong enough to transmit its effect and cause a change of the C(O)-N(H) bond.

In C1' of  $\text{ALysH}^+$ , the Ac-Ala and Ala-Lys amide II modes are the primary mode of the computed peaks at 1499 and 1521  $\text{cm}^{-1}$ , respectively, which both match the peak centered at 1508 $\text{cm}^{-1}$ . The  $\text{NH}_3^+$  umbrella vibrational mode is the primary mode of the computed peaks at 1511  $\text{cm}^{-1}$ , with a contribution from the Ac-Ala amide II mode of 8.0%. Weak coupling between the Ala-Lys amide II and the  $\text{NH}_3^+$  umbrella vibrational mode is also observed through investigating the  $\text{NH}_3^+$  umbrella vibrational modes in GaussView.

In C6' of  $\text{ALysH}^+$ , the Ac-Ala and Ala-Lys amide II modes are the primary mode of the computed peaks at 1501 and 1496  $\text{cm}^{-1}$ , respectively, which both match the peak centered at 1508 $\text{cm}^{-1}$ . Interestingly, the amide II frequencies and coupling between the three modes differ from C3 of  $\text{LysAH}^+$  and C1' of  $\text{ALysH}^+$ .

In C6' of  $\text{ALysH}^+$ , the Ac-Ala amide II mode has a slightly higher wavenumber than the Lys-Ala amide II mode, which is the opposite of that in C3 of  $\text{LysAH}^+$  and C1' of  $\text{ALysH}^+$ . There are strong couplings between the two amide II modes and the  $\text{NH}_3^+$  asymmetric vibrational mode in C6' of  $\text{ALysH}^+$ , which is not observed in C3 of  $\text{LysAH}^+$  and C1' of  $\text{ALysH}^+$  (Table 3.21).

**Carboxyl C-O-H bending mode.** The peak at  $1371\text{ cm}^{-1}$  in the LysAH<sup>+</sup> IRMPD spectrum and  $1380\text{ cm}^{-1}$  in the ALysH<sup>+</sup> spectrum correspond to the carboxyl C-O-H bending mode where the C<sub>α</sub>-C(=O) stretching is also heavily involved. In this region, the matching from the theoretical spectrum to the IRMPD spectrum is better for ALysH<sup>+</sup> than for LysAH<sup>+</sup>.

In C2 of LysAH<sup>+</sup>, the carboxyl C-O-H bending mode is the primary mode of three computed peaks at  $1338$ ,  $1365$ , and  $1367\text{ cm}^{-1}$ . The last two of which match the observed peak at  $1371\text{ cm}^{-1}$ . But the computed absorptions of those two peaks are low and seem not contributing to the observed IRMPD band at  $1371\text{ cm}^{-1}$ . Though with a higher absorption, the computed peak at  $1338\text{ cm}^{-1}$  is far from the experimental band at  $1371\text{ cm}^{-1}$ . Examination of the three vibrational modes and the others nearby in GaussView reveals extensive coupling of the carboxyl C-O-H bending mode with two less polar alkyl vibrations: the alanine CH<sub>3</sub> umbrella vibrational mode and alanine C<sub>α</sub>-H bending mode. It might contribute to the lower absorption of those three computed peaks.

In C3 of LysAH<sup>+</sup>, the carboxyl C-O-H bending mode is the primary mode of the computed peak at  $1377\text{ cm}^{-1}$ , which matches the IRMPD peak at  $1371\text{ cm}^{-1}$ . The computed mode has low computed absorption and sits at the bottom of a valley in the computed spectrum. Though the decomposition results show that all of the assigned motions correspond to the carboxyl C-O-H bending mode, investigation through GaussView reveals the involvement of the less polar alkyl vibrations: the alanine CH<sub>3</sub> umbrella vibrational mode and alanine C<sub>α</sub>-H bending mode.

Unlike the LysAH<sup>+</sup> conformations, the computed absorption of carboxyl C-O-H bending mode is high for the ALysH<sup>+</sup> conformations. The carboxyl C-O-H bending mode is the primary mode of the computed peak at  $1366\text{ cm}^{-1}$  in C1' and  $1369\text{ cm}^{-1}$  in C6', which both match the IRMPD peak at  $1380\text{ cm}^{-1}$ . Decomposition results and investigation of the vibrational modes in GaussView reveal that the carboxyl C-O-H bending mode is the major contributor to this mode and is mostly independent of other modes, with some involvement of the lysine C<sub>α</sub>-H bending motion. The independent nature of the C-O-H bending mode might contribute to the computed high absorption.

**Amide III mode.** The peak at  $1230\text{ cm}^{-1}$  in the LysAH<sup>+</sup> IRMPD spectrum and  $1265\text{ cm}^{-1}$  in the ALysH<sup>+</sup> IRMPD spectrum correspond to the amide III modes. The experimental peak is better resolved and matched for LysAH<sup>+</sup> than ALysH<sup>+</sup> in this region.

In C2 of LysAH<sup>+</sup>, the Lys-Ala amide III mode is the primary mode of the computed peak at  $1225\text{ cm}^{-1}$ , which matches the IRMPD peak at  $1230\text{ cm}^{-1}$ . In C3 of LysAH<sup>+</sup>, the Ac-Lys amide III mode is the primary mode of the computed peak at  $1248\text{ cm}^{-1}$ , which has high absorption and matches the IRMPD peak at  $1230\text{ cm}^{-1}$ . For the ALysH<sup>+</sup> conformations, the selected computed wavenumber is lower than the IRMPD peak at  $1265\text{ cm}^{-1}$  by more than  $15\text{ cm}^{-1}$ .

Additionally, as shown in Table 3.18 to Table 3.21, the computed peaks with the terminal amide III mode as the primary mode always have multiple identified motions. All weights add up to a significant percentage. But for the peaks with the in-chain amide III vibration as the primary mode, there are always fewer identified motions, and the sum of the weights is always much lower. Investigation of the vibrational modes in GaussView reveals that the terminal amide III mode is more focused, and only the adjacent C-H bending motions are involved. But for the in-chain amide III modes, the longer-range C-H bending modes are also involved, making the vibrational mode more dispersed.

**Carboxyl C-O(H) stretching mode.** The peaks at  $1125$  and  $1160\text{ cm}^{-1}$  in the LysAH<sup>+</sup> IRMPD spectrum and at  $1126$  and  $1146\text{ cm}^{-1}$  in the ALysH<sup>+</sup> IRMPD spectrum correspond to the carboxyl C-O(H) stretching mode, where the carboxyl C<sub>α</sub>-C(O) stretching is not involved. The decomposition results and investigation of those vibrational modes reveal that in the vibrational mode for the peaks at  $1125/1126\text{ cm}^{-1}$ , the motion is more focused on the C-terminus, and the backbone is less involved. But in the vibrational mode for the peaks at  $1160/1146\text{ cm}^{-1}$ , the motion is transmitted through the whole backbone. In this region, the computed spectra of both LysAH<sup>+</sup> and ALysH<sup>+</sup> match their IRMPD spectrum well.

## Summary

The structure and energetics of the isomeric dipeptide pair LysA and ALys have been investigated thoroughly using IRMPD and molecular modeling. The computational results agree with the experimentally observed different basicities and the IRMPD spectra of those two peptides.

The stability difference in the protonated peptides is the main source of their basicity difference. The characteristically different patterns in their IRMPD spectra suggest multiple conformations with various hydrogen bonding networks for LysAH<sup>+</sup> and structurally similar conformations for ALysH<sup>+</sup>. Structurally representative conformations for LysAH<sup>+</sup> and ALysH<sup>+</sup> were found by matching the theoretical IR spectra to their IRMPD spectra. Two structurally distinct conformations, C2 and C3, are selected to represent the structure of LysAH<sup>+</sup> while two structurally similar conformations, C1' and C6', are selected to represent the structure of ALysH<sup>+</sup>. Besides C3 of LysAH<sup>+</sup>, the other three selected conformations all have a left-handed helix backbone geometry, where the backbone dipole moment has a strong interaction with the charged NH<sub>3</sub><sup>+</sup> group. Though the conserved charge on the NH<sub>3</sub><sup>+</sup> group, the interaction of the backbone dipole with the NH<sub>3</sub><sup>+</sup> group, and other nonbonded intramolecular interactions also play important roles, the torsional strain on the lysine sidechain and HBE<sub>total</sub> seem to have the most contribute to the observed stability difference between the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations. The theoretical IR spectra of selected conformations were matched and decomposed to understand the characteristically different IRMPD spectra of LysAH<sup>+</sup> and ALysH<sup>+</sup>. For LysAH<sup>+</sup>, the large separation in the amide I region results from the different strength of the hydrogen bonds with the amide C=O groups. The narrow carboxyl C=O stretching band and the amide I and II bands for ALysH<sup>+</sup> result from low structural diversity and similar H-bond strength with the amide C=O groups. The decomposition results reveal that their adopted different hydrogen-bonding networks strongly affect the bond strength of their carboxyl and amide groups and are reflected in the wavenumber of their carboxyl and amide bands.

### **DapAA and AADap**

From the mass spectrometry measurements (Table 3.1), the PA and GB values were determined to be  $238.5 \pm 2.2$  and  $224.2 \pm 2.2$  kcal/mol for DapAA, and  $240.5 \pm 2.22$  and  $229.5 \pm 2.2$  kcal/mol for AADap (Batoon, 2016). The PA value of DapAA is lower than that of AADap by 2.0 kcal/mol. Though the GB values were calculated from the PA values, the difference in GB values is 5.3 kcal/mol, surprisingly more than twice larger than the difference in PA values. Because of the concern in GB values, the proton affinities will be the primary thermochemical property to rely on.

### Features of the IRMPD Spectra of DapAAH<sup>+</sup> and AADapH<sup>+</sup>

Experimental IRMPD spectra of DapAAH<sup>+</sup> and AADapH<sup>+</sup> are shown in Figure 3.25. The IRMPD spectra of DapAAH<sup>+</sup> and AADapH<sup>+</sup> are characteristically different and reflect their differences in structural features and conformational diversity.

In the C-terminal carboxyl C=O stretching region, DapAAH<sup>+</sup> has a peak at 1769 cm<sup>-1</sup>, while for AADapH<sup>+</sup>, the peak is red-shifted by 12 cm<sup>-1</sup> to 1757 cm<sup>-1</sup>. The difference suggests a weaker interaction with the carboxyl C=O group in DapAAH<sup>+</sup> than in AADapH<sup>+</sup>.

There are three widely separated peaks in the amide I region at 1648, 1682, and 1726 cm<sup>-1</sup> for DapAAH<sup>+</sup>, while there is a single peak at 1678 cm<sup>-1</sup> for AADapH<sup>+</sup>. The large separations in the DapAAH<sup>+</sup> IRMPD spectrum show that the three amide I modes have significantly different frequencies. It may be resulted from different microenvironments for the three amide bonds. Contrarily, the single peak in the AADapH<sup>+</sup> IRMPD spectrum demonstrates similar amide I frequencies and suggests a low diversity of its structure. This feature implies that the three amide bonds are likely to be in a similar microenvironment.

In the amide II region, DapAAH<sup>+</sup> has a broad band with multiple convoluted peaks. There are two well-resolved peaks at 1455 and 1522 cm<sup>-1</sup> and two unresolved peaks at 1480 and 1495 cm<sup>-1</sup>. The band also overlaps with the peaks at lower wavenumbers. Like in the amide I region, the broadness and the number of peaks in the amide II region suggest different microenvironments for the three amide bonds. Conversely, AADapH<sup>+</sup> has a narrow band in this region, with a single peak at 1494 cm<sup>-1</sup>. This feature further suggests that the three amide bonds are likely in similar microenvironments and AADapH<sup>+</sup> has a low conformational diversity.

In the region below the wavenumber of amide II, the peaks are generally less characteristic to the different structures of DapAAH<sup>+</sup> and AADapH<sup>+</sup>. The peak at 1383 cm<sup>-1</sup> in the AADapH<sup>+</sup> IRMPD spectrum is well-resolved and corresponds to the carboxyl C-O-H bending mode. The corresponding peak in the DapAAH<sup>+</sup> IRMPD spectrum could not be easily identified. There are two peaks at 1204 and

1265  $\text{cm}^{-1}$  in the amide III region in the DapAAH<sup>+</sup> IRMPD spectrum, while the corresponding peaks in the AADapH<sup>+</sup> IRMPD spectrum are not well resolved.

In summary, the spectral patterns of the DapAAH<sup>+</sup> and AADapH<sup>+</sup> IRMPD spectra are characteristically different. The amide bonds in DapAAH<sup>+</sup> are likely in different microenvironments and the amide bonds in AADapH<sup>+</sup> reside in a similar microenvironment.

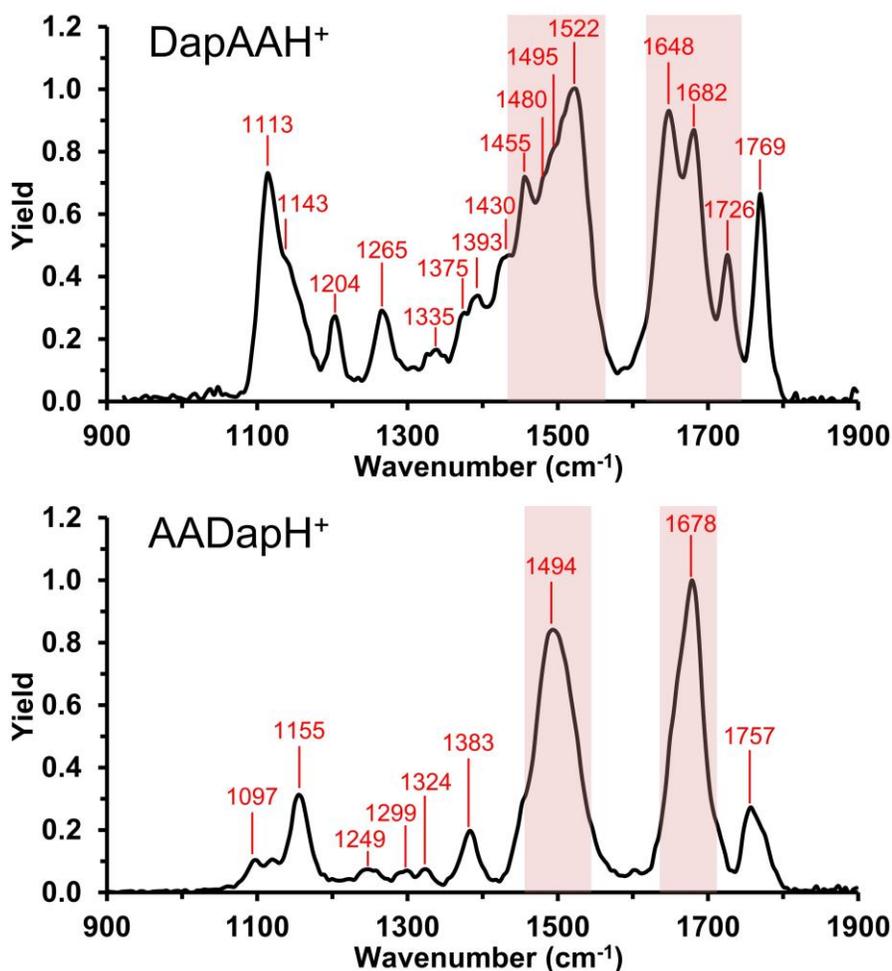


Figure 3.25. IRMPD Spectra of DapAAH<sup>+</sup> (top) and AADapH<sup>+</sup> (bottom) with major peaks labeled in wavenumber. (The spectral data were obtained from the FELIX Laboratory, Radboud University, The Netherlands. The peaks in the shaded area correspond to the amide I (high wavenumber) and amide II (low wavenumber) bands.)

### Computed PA and GB Values for DapAA and AADap

The CREST (Pracht et al., 2020) approach was used for the neutral peptides, and 226 – 293 conformations were generated at the conformational search step. For the protonated peptides, both the CREST and systematic approaches were employed. Detailed computational procedures of the two approaches are described in Chapter 2. Based on the energy calculated at the GFN2-xTB level of theory in CREST, both approaches locate the same lowest energy conformation. After combining the output data from both methods, duplicate and high-energy forms were removed to create the final ensemble of conformations, which gives about 65 – 94 conformations.

For both the neutral and protonated peptides, the conformations obtained from the search step were optimized at the HF/3-21G\* level of theory and ranked in free energy. Conformations within the energy windows of 4 and 6 kcal/mol were kept for the neutral and protonated peptides, respectively. The resulting 35 – 60 conformations for the neutral form and 17 – 22 conformations for the protonated form were geometry optimized and frequency calculated at the wB97XD/6-311G\*\* level of theory. After removing the degenerate and high-energy conformations, a final set of ten conformations for each peptide species is selected and used to compute the theoretical thermochemical values based on the equations described in Chapter 2.

Conformations with a good match to their corresponding IRMPD spectrum were selected as structurally representative conformations of the protonated peptides, and their computed vibrational modes were decomposed into displacements along chemically meaning internal coordinates using VibAnalysis (H. Li et al., 2016). Single point energy calculation at B3LYP-D3(BJ)/ma-def2-TZVPP level of theory was conducted on those conformations. The generated fchk files were used for the charge analysis, dipole moment decomposition (Lu & Chen, 2012), hydrogen bond analysis (Emamian et al., 2019), and RDG analysis (Johnson et al., 2010).

The calculated energy values for all ten conformations of each peptide species are listed in Table 3.23, which includes the enthalpy/free energy, relative enthalpy/free energy, the Boltzmann probability, and the weighted average of enthalpy/free energy. The ten lowest energy conformations of the neutral and protonated peptides are noted as N1 to N10 for DapAA, N1' to N10' for AADap, C1 to C10 for DapAAH<sup>+</sup>, and C1' to C10' for AADapH<sup>+</sup>, respectively. The conformations are ranked in free energy. As shown in Table 3.23, the two neutral peptides have nearly the same free energy (within 0.1 kcal/mol), while DapAAH<sup>+</sup> has higher free energy than AADapH<sup>+</sup> by 1.7 kcal/mol, suggesting higher stability of AADapH<sup>+</sup>. The same trend is also observed by comparing their relative enthalpy (Rel. H<sub>avg</sub>) values. DapAAH<sup>+</sup> is 3.6 kcal/mol higher in enthalpy than AADapH<sup>+</sup>, while the enthalpy difference between DapAA and AADap is much smaller, by 1.2 kcal/mol.

Table 3.23  
 Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for DapAA and AADap

Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>	Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>
DapAA						DapAAH <sup>+</sup>					
N1	-1026.047451	-1026.122387	0.0	0.0	58.3%	C1	-1026.434121	-1026.509331	0.0	0.0	77.1%
N2	-1026.044659	-1026.120941	1.8	0.9	12.6%	C2	-1026.432588	-1026.507464	1.0	1.2	10.7%
N3	-1026.048533	-1026.120359	-0.7	1.3	6.8%	C3	-1026.433177	-1026.507128	0.6	1.4	7.5%
N4	-1026.04478	-1026.120021	1.7	1.5	4.8%	C4	-1026.432027	-1026.505558	1.3	2.4	1.4%
N5	-1026.049784	-1026.11981	-1.5	1.6	3.8%	C5	-1026.428879	-1026.505345	3.3	2.5	1.1%
N6	-1026.0419	-1026.119599	3.5	1.7	3.0%	C6	-1026.428767	-1026.505194	3.4	2.6	1.0%
N7	-1026.046143	-1026.119565	0.8	1.8	2.9%	C7	-1026.428821	-1026.504609	3.3	3.0	0.5%
N8	-1026.043495	-1026.119527	2.5	1.8	2.8%	C8	-1026.430402	-1026.504607	2.3	3.0	0.5%
N9	-1026.042807	-1026.11943	2.9	1.9	2.5%	C9	-1026.430798	-1026.503799	2.1	3.5	0.2%
N10	-1026.047755	-1026.119386	-0.2	1.9	2.4%	C10	-1026.426983	-1026.502594	4.5	4.2	0.1%
H <sub>avg</sub> <sup>c</sup>	-1026.046706					H <sub>avg</sub> <sup>c</sup>	-1026.433689				
G <sub>avg</sub> <sup>c</sup>	-1026.121461					G <sub>avg</sub> <sup>c</sup>	-1026.508765				
Rel. H <sub>avg</sub> <sup>c</sup>	0.0					Rel. H <sub>avg</sub> <sup>c</sup>	0.0				
Rel. G <sub>avg</sub> <sup>c</sup>	0.0					Rel. G <sub>avg</sub> <sup>c</sup>	0.0				

<sup>a</sup> Values (kcal/mol) obtained from calculations at the wB97XD/6-311G\*\* level of theory.

<sup>b</sup> Boltzmann probability calculated using the equations described in Chapter 2.

<sup>c</sup> Weighted average of enthalpy and free energy (Hartree) of the ten conformations calculated using the equations described in Chapter 2.

<sup>d</sup> Relative enthalpy and free energy (kcal/mol) DapAA and AADap, and between DapAAH<sup>+</sup> and AADapH<sup>+</sup>, calculated using the weighted average of H<sub>avg</sub> and G<sub>avg</sub>, respectively.

(Table 3.23 Continued)

Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>	Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>
AADap											
N1'	-1026.048789	-1026.122108	0.0	0.0	22.3%	C1'	-1026.440063	-1026.512176	0.0	0.0	75.2%
N2'	-1026.04874	-1026.12194	0.0	0.1	18.7%	C2'	-1026.4373	-1026.510098	1.7	1.3	8.3%
N3'	-1026.048969	-1026.121925	-0.1	0.1	18.4%	C3'	-1026.438864	-1026.510042	0.8	1.3	7.8%
N4'	-1026.049896	-1026.1217	-0.7	0.3	14.5%	C4'	-1026.435685	-1026.5089	2.7	2.1	2.3%
N5'	-1026.050458	-1026.120838	-1.0	0.8	5.8%	C5'	-1026.435974	-1026.508785	2.6	2.1	2.1%
N6'	-1026.048001	-1026.12083	0.5	0.8	5.8%	C6'	-1026.436165	-1026.508078	2.4	2.6	1.0%
N7'	-1026.047458	-1026.12058	0.8	1.0	4.4%	C7'	-1026.436542	-1026.508034	2.2	2.6	0.9%
N8'	-1026.044366	-1026.120418	2.8	1.1	3.7%	C8'	-1026.435627	-1026.507957	2.8	2.6	0.9%
N9'	-1026.043662	-1026.120343	3.2	1.1	3.4%	C9'	-1026.434859	-1026.507855	3.3	2.7	0.8%
N10'	-1026.04797	-1026.120172	0.5	1.2	2.9%	C10'	-1026.435676	-1026.507695	2.8	2.8	0.7%
H <sub>avg</sub> <sup>c</sup>	-1026.048601					H <sub>avg</sub> <sup>c</sup>	-1026.439374				
G <sub>avg</sub> <sup>c</sup>	-1026.121589					G <sub>avg</sub> <sup>c</sup>	-1026.511512				
Rel. H <sub>avg</sub> <sup>c</sup>	-1.2					Rel. H <sub>avg</sub> <sup>c</sup>	-3.6				
Rel. G <sub>avg</sub> <sup>c</sup>	-0.1					Rel. G <sub>avg</sub> <sup>c</sup>	-1.7				

<sup>a</sup> Values (kcal/mol) obtained from calculations at the wB97XD/6-311G\*\* level of theory.

<sup>b</sup> Boltzmann probability calculated using the equations described in Chapter 2.

<sup>c</sup> Weighted average of enthalpy and free energy (Hartree) of the ten conformations calculated using the equations described in Chapter 2.

<sup>d</sup> Relative enthalpy and free energy (kcal/mol) DapAA and AADap, and between DapAAH<sup>+</sup> and AADapH<sup>+</sup>, calculated using the weighted average of H<sub>avg</sub> and G<sub>avg</sub>, respectively.

The weighted average of enthalpy and free energy were used to calculate the proton affinity and the gas-phase basicity of the peptides. Ethylamine ( $\text{CH}_3\text{CH}_2\text{NH}_2$ , PA = 218.0 kcal/mol, GB = 210 kcal/mol) (Hunter & Lias, 2009) was used as the reference base. By equation described in Chapter 2, the theoretical PA and GB values are 239.5 and 231.4 kcal/mol for DapAA and 241.9 and 233.1 kcal/mol for AADap (Table 3.24). The theoretical PA value of DapAA is lower than AADap by 2.4 kcal mol<sup>-1</sup>. Their theoretical PA values and difference match the experimental results. However, their computed GB values are both significantly higher than the experimental values. The calculated difference in GB values is 1.7 kcal/mol, much smaller than the experimentally determined difference of 5.3 kcal/mol.

As mentioned before, the experimental PA values were used to derive the experimental GB values and should be relied on primarily for the concern in the experimental GB values. Because the computed conformations produce PA values that match the experimental results, the GB values are calculated based on the same set of conformations. The computed GB values are considered a good estimate of basicity values. DapAA is lower in basicity than AADap by about 2 kcal/mol. The greater stability of AADapH<sup>+</sup> is the main source of the higher basicity of AADap.

Table 3.24

*Experimental and Calculated Thermochemical Values of DapAA and AADap*

<b>Peptide</b>	<b>PA<sub>expt</sub> kcal/mol</b>	<b>GB<sub>expt</sub> kcal/mol</b>	<b>PA<sub>calc</sub> kcal/mol</b>	<b>GB<sub>calc</sub> kcal/mol</b>
DapAA	238.5 ± 2.2	224.2 ± 2.2	239.5	231.4
AADap	240.5 ± 2.2	229.5 ± 2.2	241.9	233.1

<sup>a</sup> Values taken from published PhD dissertation (Batoon, 2016).

### Computed Structures and IR Spectra of DapAA and AADap

**Structural features of DapAA and AADap.** As shown in Table 3.23, though their free energy values are different within 0.1 kcal/mol, there is a noticeable difference in the enthalpy of the neutral peptides. The calculated ten lowest energy conformations of the neutral peptides are shown in Figure 3.26 and Figure 3.27. The hydrogen bonds involving the Dap sidechain primary amine are indicated in as red dashed lines with the N --- H distance labeled. The overlays of the ten conformations of DapAA and AADap are shown in Figure 3.28 to demonstrate their structural variations.

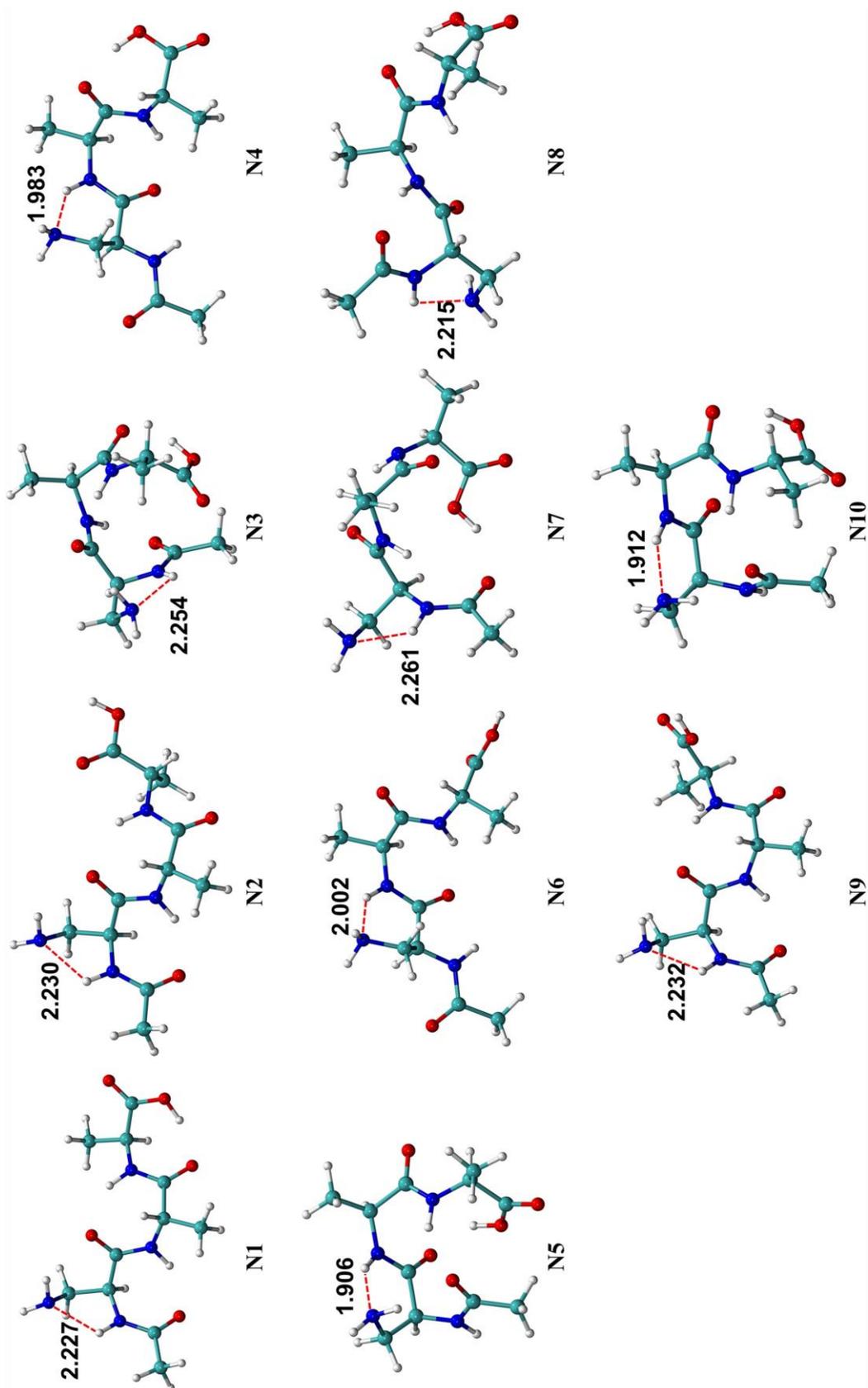


Figure 3.26. Computed geometries of DapAA at the wB97XD/6-311G\*\* level of theory.

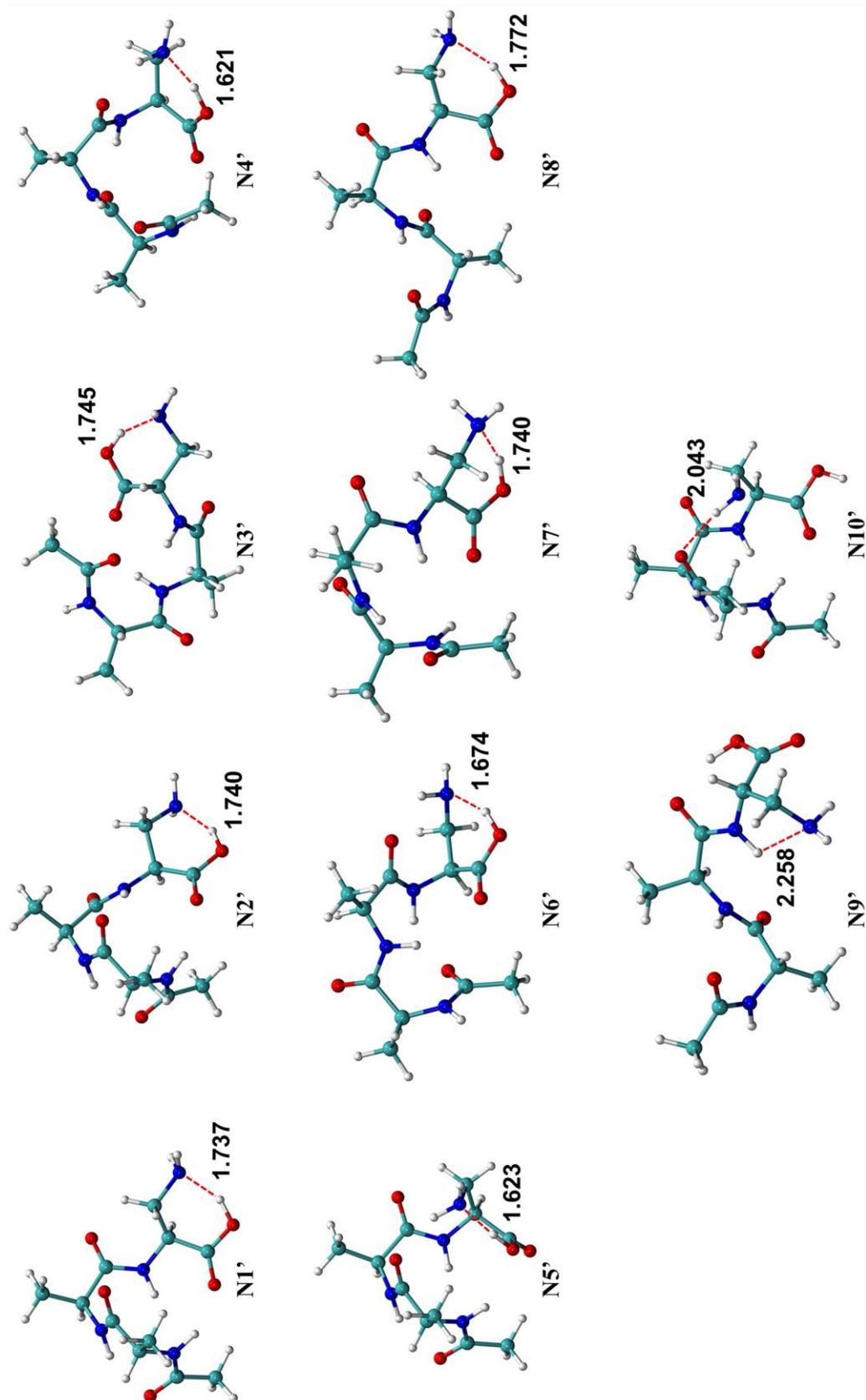


Figure 3.27. Computed geometries of AADap at the wB97XD/6-311G\*\* level of theory.

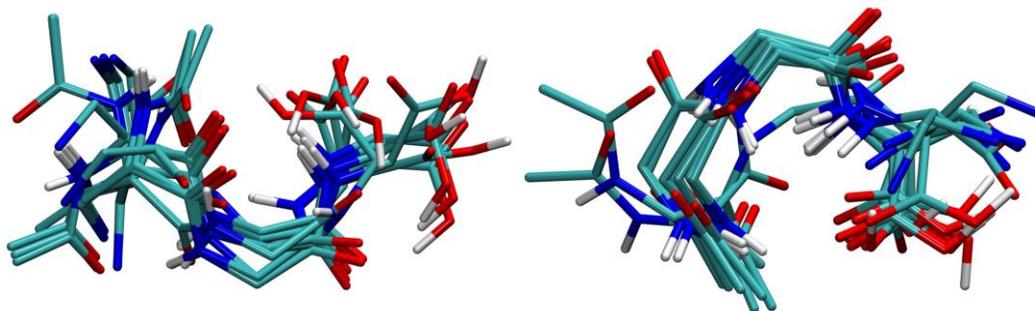


Figure 3.28. Overlay of the ten conformations of DapAA (left) and AADap (right).

In all ten conformations of DapAA (Figure 3.26 and Figure 3.28), the Dap sidechain primary amine functions as an acceptor and forms a hydrogen bond with an amide N-H group on the N-terminus. In N1, N2, N3, N7, N8, and N9, the Ac-Dap amide N-H group acts as the hydrogen bond donor, and in the other four conformations, the Dap-Ala<sup>1st</sup> amide N-H group acts as the hydrogen bond donor. In none of the conformations, the carboxyl group interacts with the Dap sidechain primary amine.

Contrarily, in most AADap conformations except N9' and N10' (Figure 3.27), the Dap sidechain primary amine forms a hydrogen bond with the C-terminal carboxyl O-H group. In conformation N9', the Ala<sup>2nd</sup>-Dap amide N-H group donates a hydrogen bond to the Dap sidechain primary amine. In conformation N10', the Dap sidechain primary amine does not accept a hydrogen bond but rather donates one to the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O group to form a hydrogen bond.

For the hydrogen bonds accepted by the Dap sidechain primary amine, the N --- H distance in the DapAA conformations ranges from 1.906 to 2.261 Å, mostly longer than 2.0 Å. While for the AADap conformations, the N --- H distance ranges from 1.621 to 1.772 Å, all about 1.7 Å. Besides the shorter N --- H distance, the carboxyl O-H group is also a better hydrogen bond donor than the amide N-H group. Considering the primary amine on the Dap side chain being the same, the hydrogen bonds in the AADap conformations are stronger than in the DapAA conformations, which seem to contribute to the lower enthalpy of the AADap conformations.

**Structural features of DapAAH<sup>+</sup> and AADapH<sup>+</sup>.** Compared to DapAA, AADap has a higher basicity resulting from the higher stability of its protonated form. Their stability difference between the two protonated peptides originates from their different structures, which is strongly suggested by their characteristically different IRMPD spectra. The structure and energetics of the protonated peptides are analyzed and compared in detail. The computed conformations of DapAAH<sup>+</sup> and AADapH<sup>+</sup> are shown in Figure 3.29 and Figure 3.31, respectively. The hydrogen bonds involving the NH<sub>3</sub><sup>+</sup> group are highlighted by the red dashed lines with their H --- O distances labeled. The overlays of the ten conformations of DapAAH<sup>+</sup> and AADapH<sup>+</sup> are shown in Figure 3.33 to demonstrate their structural variations and differences. After applying a scaling factor of 0.947 to the computed IR frequencies calculated at the wB97XD/6-311G\*\* level of theory, the theoretical IR spectra from each of the conformations are overlaid onto the DapAAH<sup>+</sup> and AADapH<sup>+</sup> IRMPD spectra and are shown in Figure 3.30 and Figure 3.32.

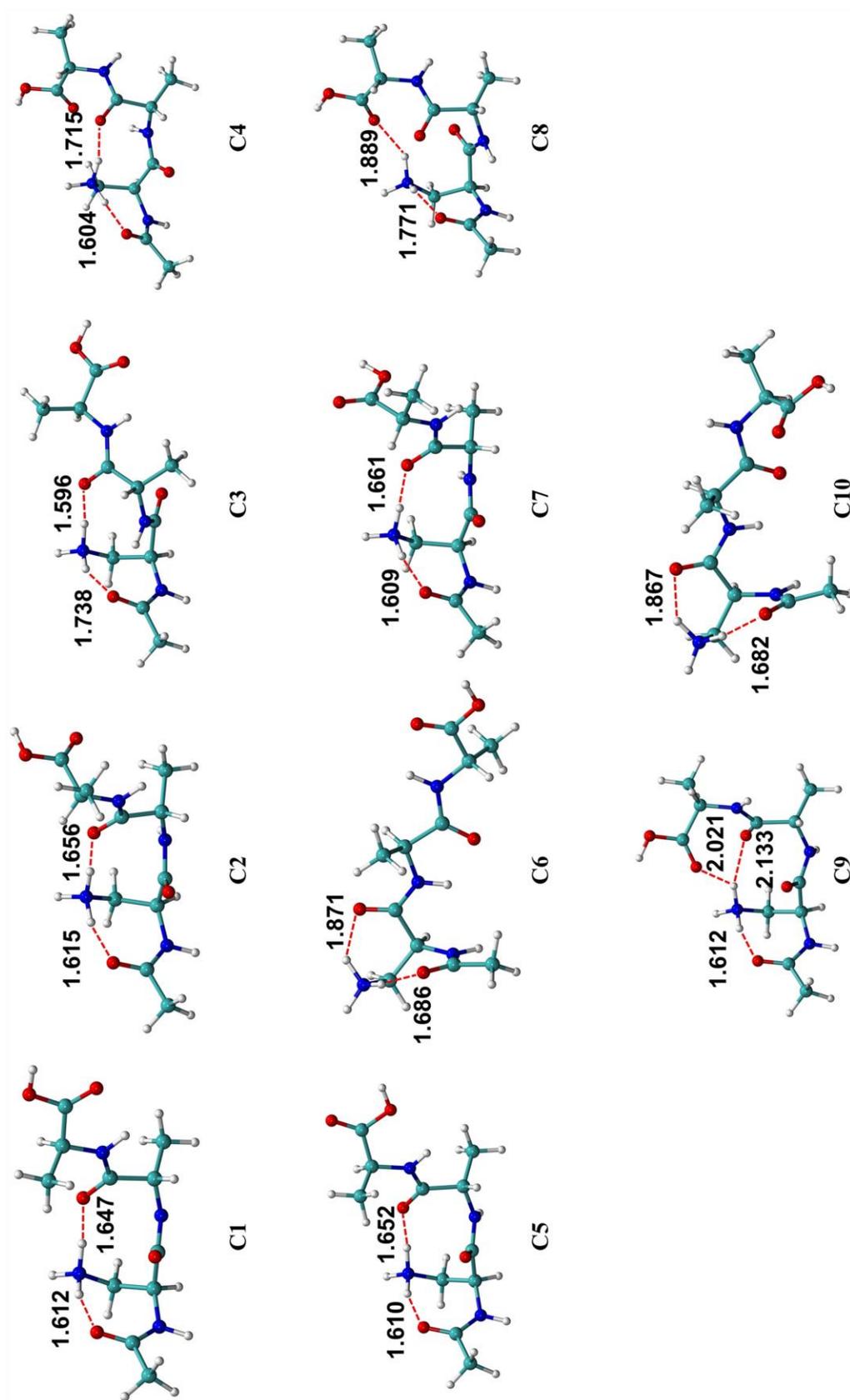


Figure 3.29. Computed geometries of DapAAH<sup>+</sup> at the wB97XD/6-311G\*\* level of theory.

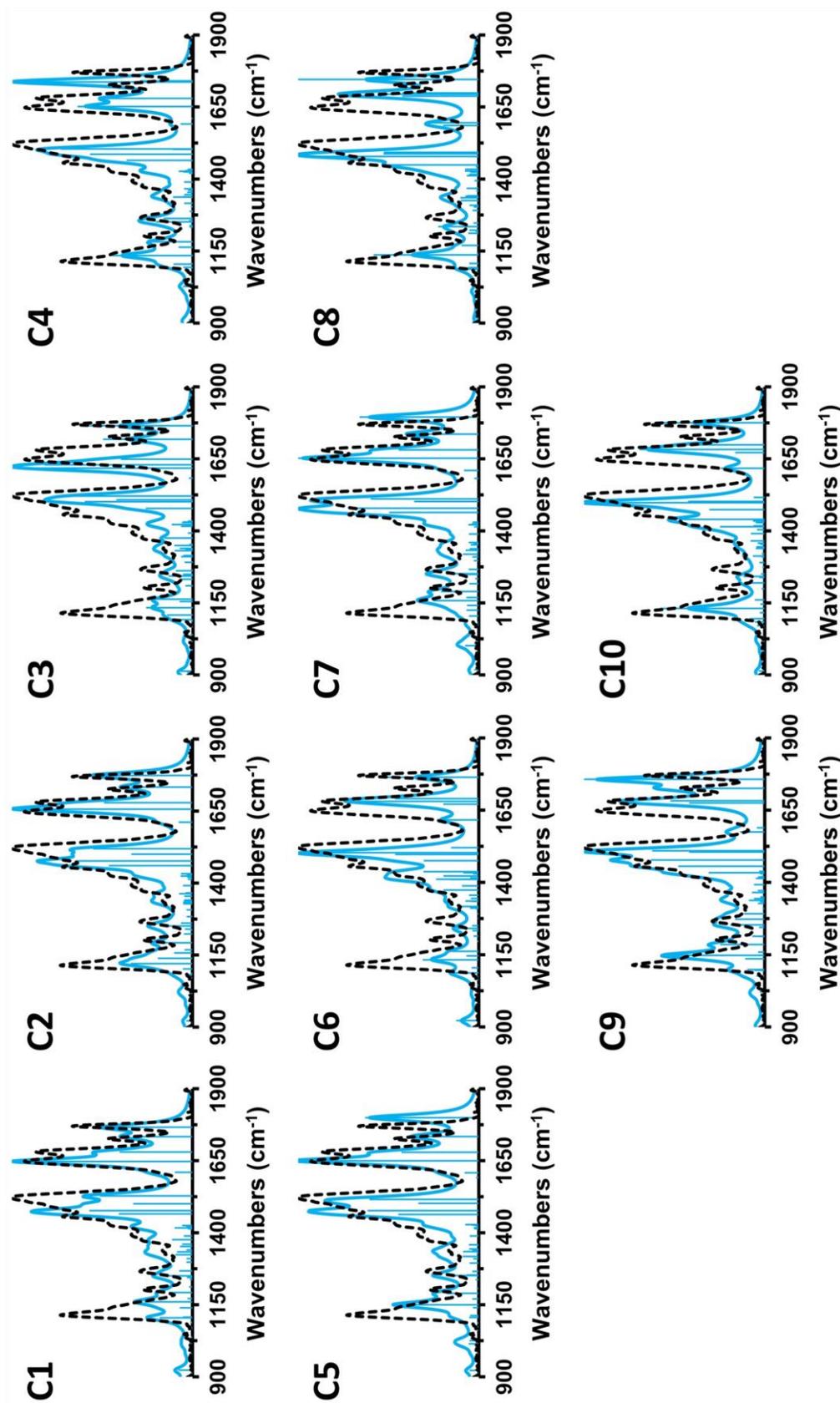


Figure 3.30. Computed IR stick and convoluted spectra (solid light blue) overlaid onto experimental IRMPD spectrum (dashed black) of DapAAH<sup>+</sup>. (The IR frequencies were calculated at the wB97XD/6-311G\*\* level of theory and a scaling factor of 0.947 was applied.)

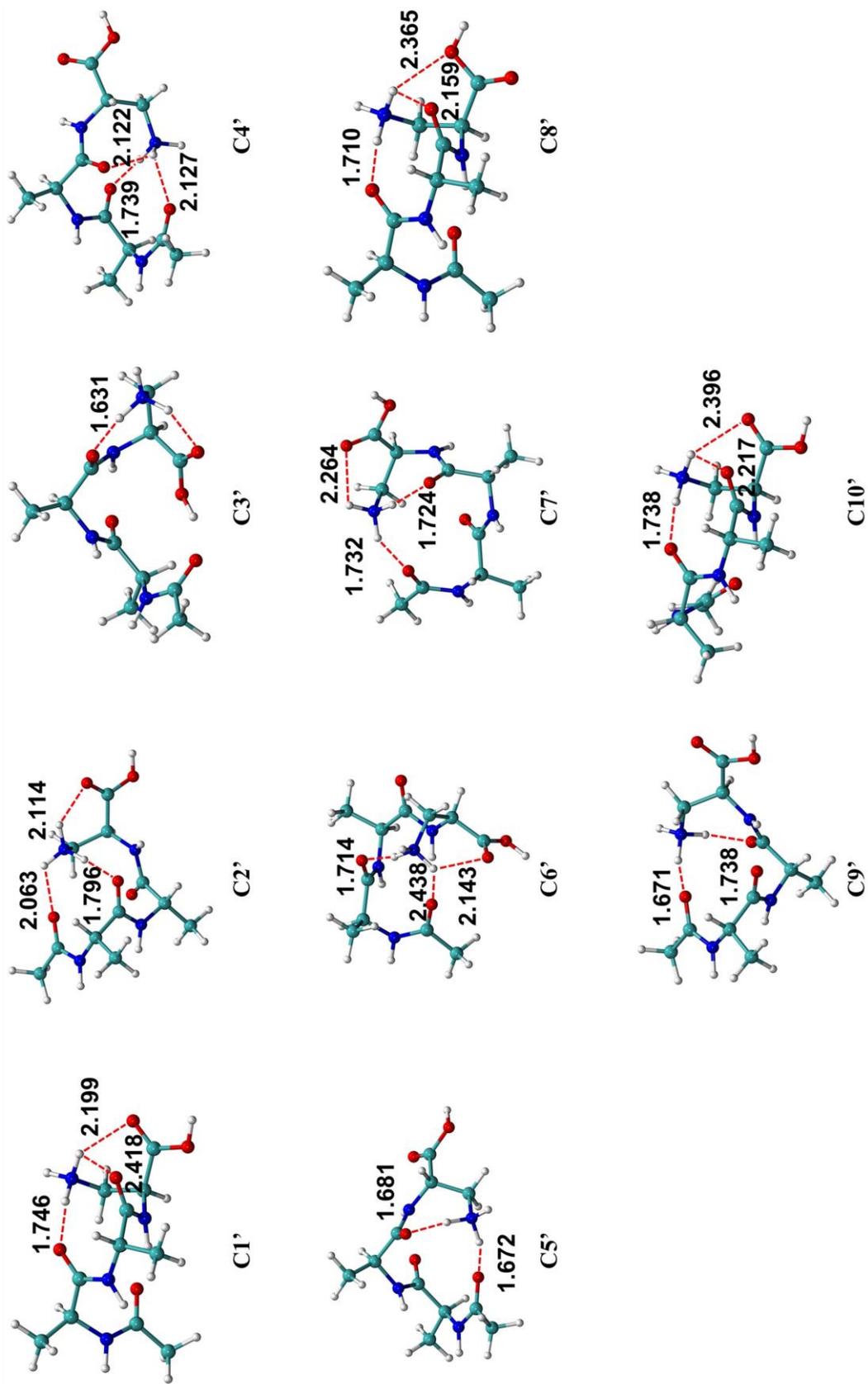


Figure 3.31. Computed geometries of AADapH<sup>+</sup> at the wB97XD/6-311G\*\* level of theory.

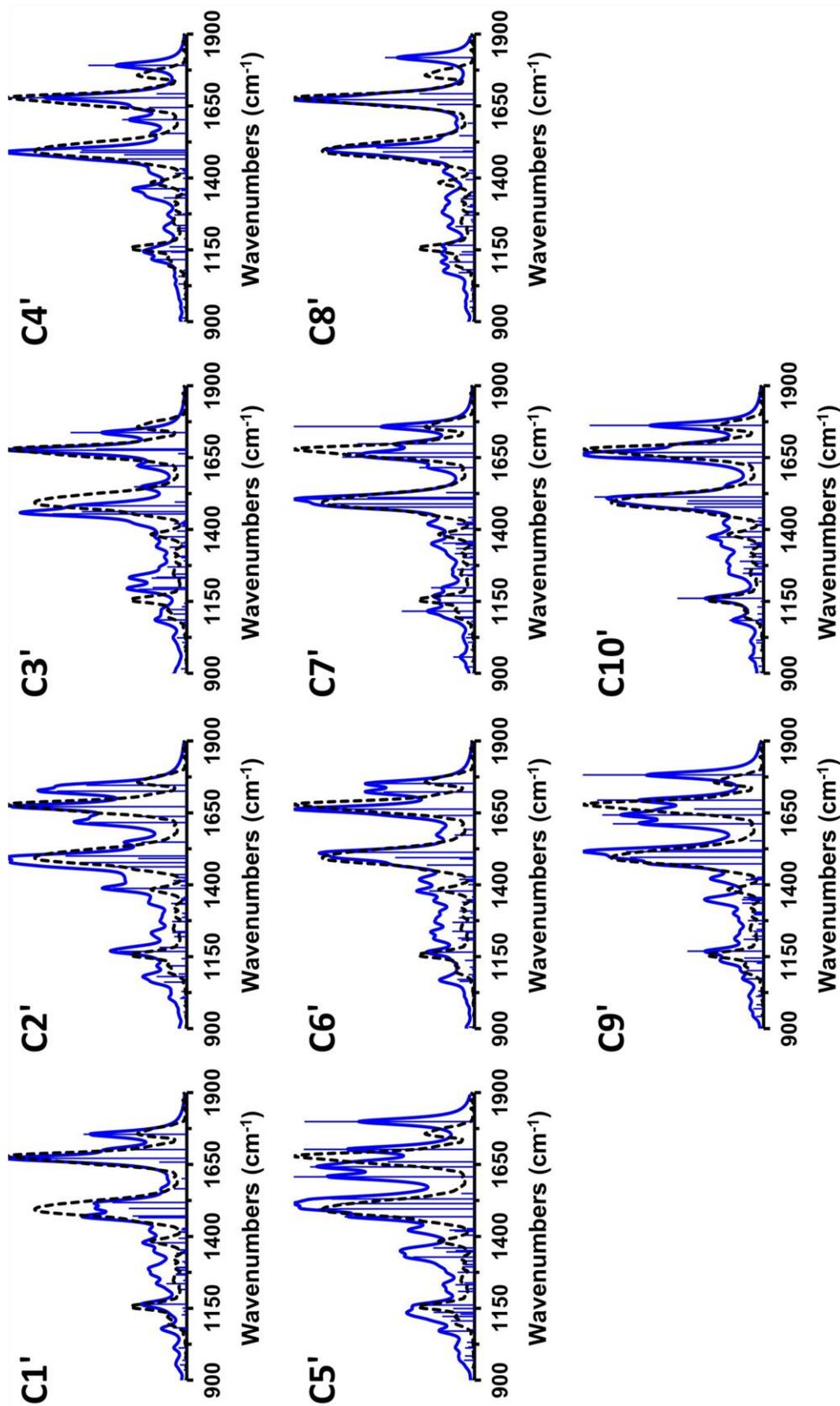


Figure 3.32. Computed IR stick and convoluted spectra (solid blue) overlaid onto experimental IRMPD spectrum (dashed black) of AADapH<sup>+</sup>. (The IR frequencies were calculated at the wB97XD/6-311G\*\* level of theory and a scaling factor of 0.947 was applied.)

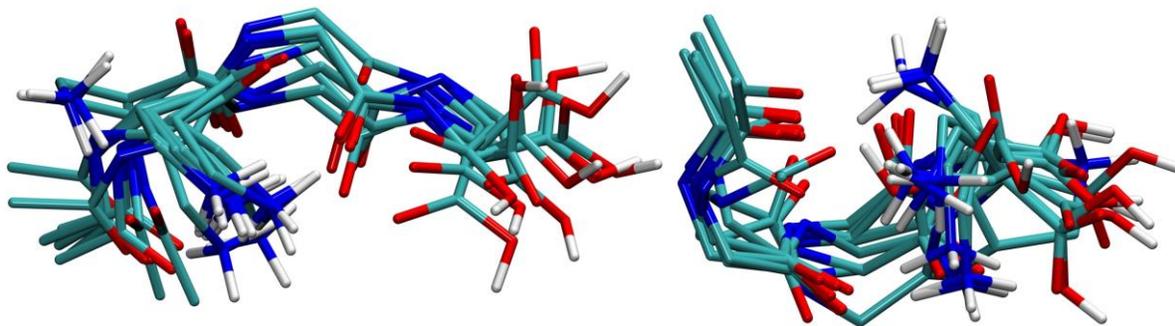


Figure 3.33. Overlay of the ten conformations of DapAAH<sup>+</sup> (left) and AADapH<sup>+</sup> (right).

In all ten conformations of DapAAH<sup>+</sup> (Figure 3.29 and Figure 3.33), the NH<sub>3</sub><sup>+</sup> group donates a hydrogen bond to the Ac-Dap amide C=O group. Like the feature observed in its neutral form, the C-terminal carboxyl group in most DapAAH<sup>+</sup> conformations do not interact with the Dap sidechain. As a result, the DapAAH<sup>+</sup> conformations are extended. Contrarily, in the AADapH<sup>+</sup> conformations, the N-terminus curls up and approaches the NH<sub>3</sub><sup>+</sup> group at the C-terminus. As a result, the AADapH<sup>+</sup> conformations appear to be more compact.

In the first group of DapAAH<sup>+</sup> conformations, C1, C2, C4, C5, C7, and C9 (Figure 3.34), the NH<sub>3</sub><sup>+</sup> group forms hydrogen bonds with the Ac-Dap and Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O groups, while the Dap-Ala<sup>1st</sup> amide C=O group is placed below the NH<sub>3</sub><sup>+</sup> group. The backbone geometry of C1, C2, C5, and C7 is nearly identical. Their main differences are the orientation of the carboxyl group, which hangs on the edge of the molecule. Contrarily, C4 and C9 have their carboxyl group bent towards the middle of the molecule. The resulted torsional pressure on the backbone pushes the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O away from and elongates its hydrogen bond with the NH<sub>3</sub><sup>+</sup> group. In C4, the carboxyl group is placed in the back of the molecule. And the carboxyl C=O group forms a hydrogen bond with the Dap-Ala<sup>1st</sup> amide NH group. As a result, the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O is pushed more to the front, and its hydrogen bond with the NH<sub>3</sub><sup>+</sup> group becomes slightly elongated to 1.715 Å. In C9, the carboxyl group is placed in the front of the molecule. The carboxyl C=O group forms a hydrogen bond with the NH<sub>3</sub><sup>+</sup> group. As a result,

the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O group appears to be pushed towards the back of the molecule, and its hydrogen bond with the NH<sub>3</sub><sup>+</sup> group becomes significantly elongated and weakened.

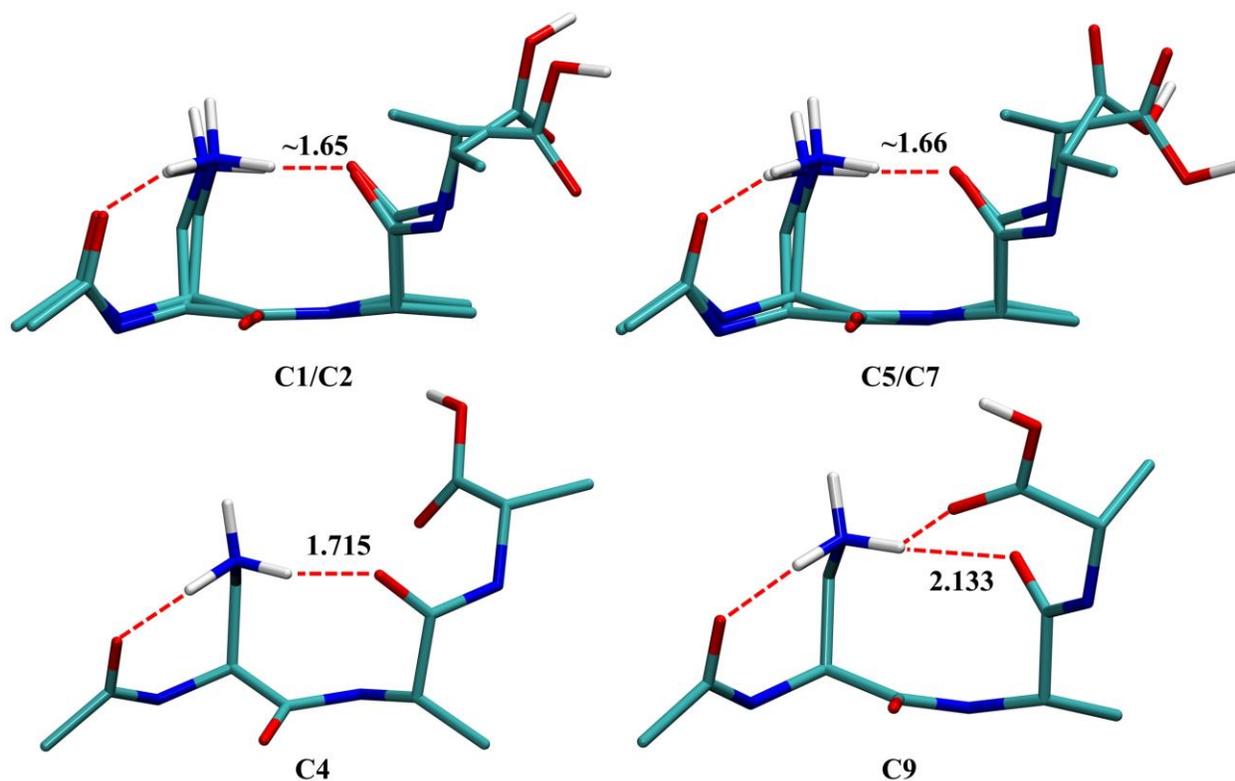


Figure 3.34. Overlay of C1 and C2 (top left), C5 and C7 (top right), C4 (bottom left), and C9 (bottom right). (The dashed lines indicate H-bonding.)

Comparing the theoretical spectra of this group of conformations to the IRMPD spectrum of DapAAH<sup>+</sup> (Figure 3.30), good match with C1 and C2 is observed across the whole range of wavenumbers, especially in the carboxyl C=O stretching region and the amide I region. For C5 and C7, their carboxyl C=O stretching peak is more blue-shifted than C1 and C2 and doesn't match the DapAAH<sup>+</sup> IRMPD spectrum. As shown in Figure 3.34, in C1 and C2, the carboxyl C=O group points downward and interacts with the Ala<sup>2nd</sup> N-H group or the Ala<sup>1st</sup> methyl group. But in C5 and C7, the carboxyl C=O group points upward, making it more isolated than C1 and C2. As a result, their carboxyl C=O groups are less weakened, and their C=O stretching peaks appear at higher wavenumbers.

Because of the hydrogen bond between the carboxyl C=O group and the Dap-Ala<sup>1st</sup> amide N-H group in C4, its carboxyl C=O stretching mode is significantly redshifted and does not match the IRMPD spectrum. Also, its amide II band is significantly redshifted than the IRMPD spectrum, suggesting its non-existence.

Compared to the DapAAH<sup>+</sup> IRMPD spectrum, C9 lacks one peak in the amide I region. But it has a better match than C1 and C2 in the amide II region. Also, its peaks in the carboxyl C=O stretching region and the amide I region match the experimental peaks. The matching results suggest the coexistence of C1, C2, and C9.

In the second group of conformations (Figure 3.35), C3 and C8, the NH<sub>3</sub><sup>+</sup> group forms a hydrogen bond with the Ac-Dap amide C=O group. Unlike in the first group of conformations, their Dap-Ala<sup>1st</sup> amide C=O group is not placed below the NH<sub>3</sub><sup>+</sup> group but rather points toward the back of the molecule.

In C3, the NH<sub>3</sub><sup>+</sup> group donates a hydrogen bond to the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O group, and the carboxyl group is isolated on the edge of the molecule. In C8, the carboxyl group is flipped inside to interact with the NH<sub>3</sub><sup>+</sup> group. As a result, the hydrogen bond between the NH<sub>3</sub><sup>+</sup> group and the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O group is missing in C8. Compared to the DapAAH<sup>+</sup> IRMPD spectrum, the theoretical IR spectra of C3 and C8 are poorly matched, indicating the non-existence of this backbone geometry.

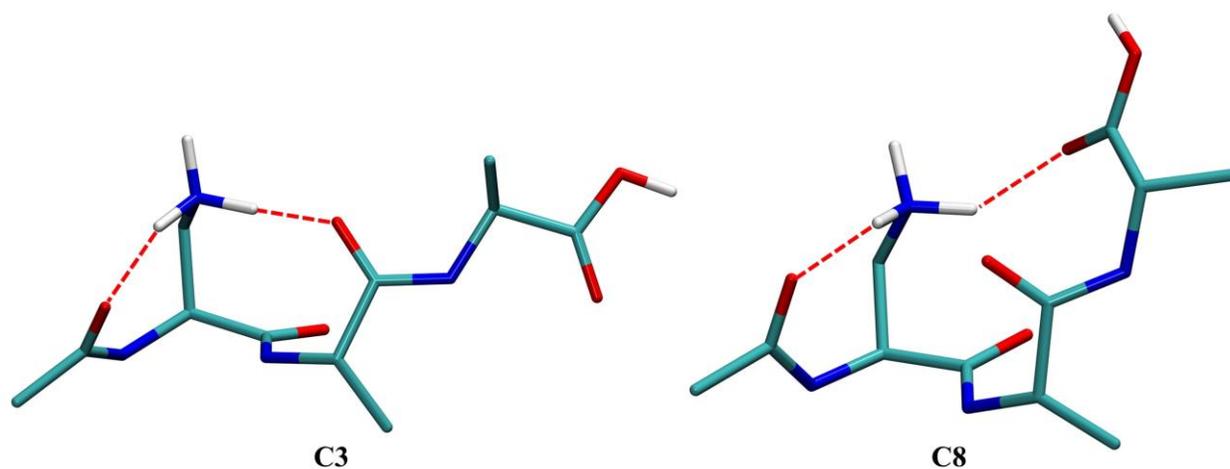


Figure 3.35. Structure of C3 (left) and C8 (right). (The dashed lines indicate H-bonding.)

In the last group of conformations, C6 and C10, the  $\text{NH}_3^+$  group forms hydrogen bonds with the Ac-Dap and Dap-Ala<sup>1st</sup> amide C=O groups, and the rest of the backbone extend almost linearly. Comparing their theoretical IR spectra to the IRMPD spectrum of DapAAH<sup>+</sup> (Figure 3.29), poor matching is observed, suggesting the non-existence of this hydrogen bonding pattern and backbone geometry.

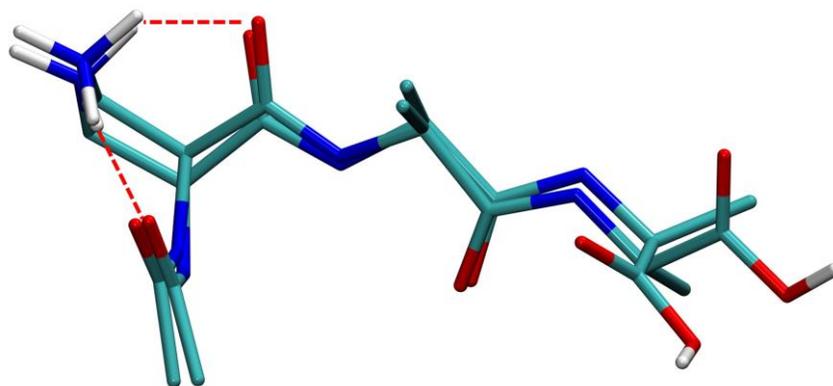
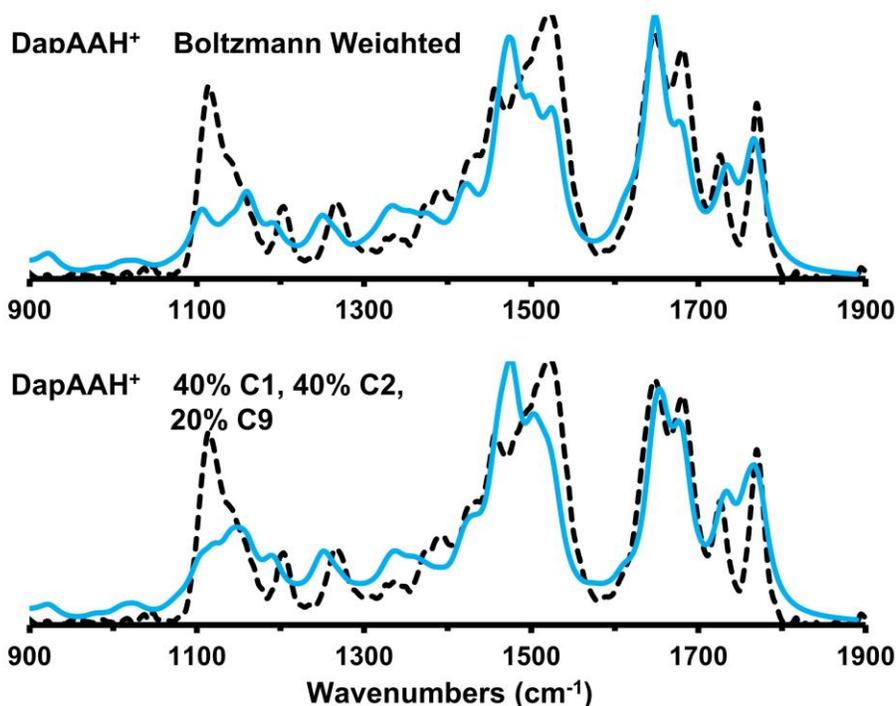


Figure 3.36. Structure of C6 and C10. (The dashed lines indicate H-bonding.)

Based on the free energetics and matching to the IRMPD spectrum, the structure of DapAAH<sup>+</sup> could be represented primarily by C1. However, C2 has similar matching to the IRMPD spectrum across the whole range of wavenumber as C1. And C9 has better matching than C1 and C2 in the amide II

region (Figure 3.30). Additionally, C2 and C9 are structurally similar to C1 (Figure 3.34), allowing easy conversions of C1 to C2 and C9. Therefore, there is a high likelihood of the coexistence of C1, C2, and C9. A hand-weighted spectrum from C1, C2, and C9 is created and shown in Figure 3.37 together with the Boltzmann weighted average spectrum.



*Figure 3.37.* Boltzmann weighted average spectra (top) and hand-weighted spectra (bottom) overlaid onto the IRMPD spectrum of DapAAH<sup>+</sup>. (The dashed black plot represents the IRMPD spectrum, and the solid light blue plots represent the calculated spectra.)

The peaks in the hand-weighted spectrum are less resolved but have a slightly better overall matching to the IRMPD spectrum. All three conformations, C1, C2, and C9, are likely to co-exist and are considered structurally representative conformations of DapAAH<sup>+</sup>.

In the ten conformations of AADapH<sup>+</sup>, there are multiple different backbone geometries and hydrogen-bonding networks. One interesting observation from the AADapH<sup>+</sup> conformations is that four of the ten conformations are helical, with three right-handed (C1', C8' and C10') and one left-handed

(C6'). Another one is that in two conformations (C2' and C7'), all three hydrogens of the  $\text{NH}_3^+$  group are hydrogen-bonded. It is not observed in the DapAAH<sup>+</sup> conformations.

In the first group of AADapH<sup>+</sup> conformations, C1', C8' and C10' (Figure 3.38), the Ala<sup>1st</sup>-Ala<sup>2nd</sup> and the Ala<sup>2nd</sup>-Dap amide C=O groups are nearly parallel to each other, making a right-handed turn at the second alanine, and they both form hydrogen bonds with the  $\text{NH}_3^+$  group. There is also a weak interaction between the  $\text{NH}_3^+$  group and the carboxyl group. The N terminal Ac-Ala<sup>1st</sup> amide group extends the helix geometry and the hydrogen bond network by accepting a hydrogen bond from the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide group.

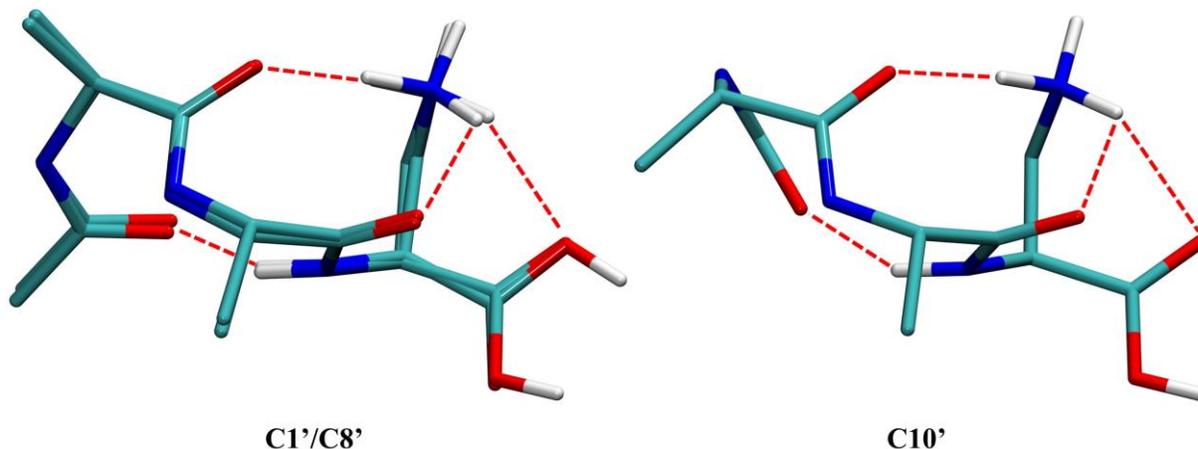


Figure 3.38. Structure of C1' and C8' (left), and C10'(right). (The dashed lines indicate H-bonding.)

To demonstrate the helical geometries of C1', C8', and C10', their dihedral angles ( $\phi$ ,  $\psi$ ) from the C-terminus to the N-terminus are obtained and tabulated in Table 3.25 and overlaid to the Ramachandran plot in Figure 3.39. Their C→O vector angles between adjacent amide C=O and carboxyl C=O groups are listed in Table 3.25 to show their alignments.

Table 3.25  
Dihedral Angles and C→O Vector Angles in C1', C8', and C10' of AADapH<sup>+</sup>

Start Group	End Group	$\phi$ (phi) (°)		$\psi$ (psi) (°)	C→O Vector angle (°)
		C(O)-N(H)-C <sub>α</sub> -C(O)	N(H)-C <sub>α</sub> -C(O)-N(H)		
Ac-Ala <sup>1st</sup>	Ala <sup>1st</sup> -Ala <sup>2nd</sup>	-60.7	-34.8		34.0
C1' Ala <sup>1st</sup> -Ala <sup>2nd</sup>	Ala <sup>2nd</sup> -Dap	-49.2	-40.3		29.7
	Ala <sup>2nd</sup> -Dap carboxyl group				32.1
Ac-Ala <sup>1st</sup>	Ala <sup>1st</sup> -Ala <sup>2nd</sup>	-59.4	-36.5		33.6
C8' Ala <sup>1st</sup> -Ala <sup>2nd</sup>	Ala <sup>2nd</sup> -Dap	-50.2	-38.7		30.3
	Ala <sup>2nd</sup> -Dap carboxyl group				93.6
Ac-Ala <sup>1st</sup>	Ala <sup>1st</sup> -Ala <sup>2nd</sup>	52.1	-131.8		99.9
C10' Ala <sup>1st</sup> -Ala <sup>2nd</sup>	Ala <sup>2nd</sup> -Dap	-51.0	-37.1		33.5
	Ala <sup>2nd</sup> -Dap carboxyl group				31.2

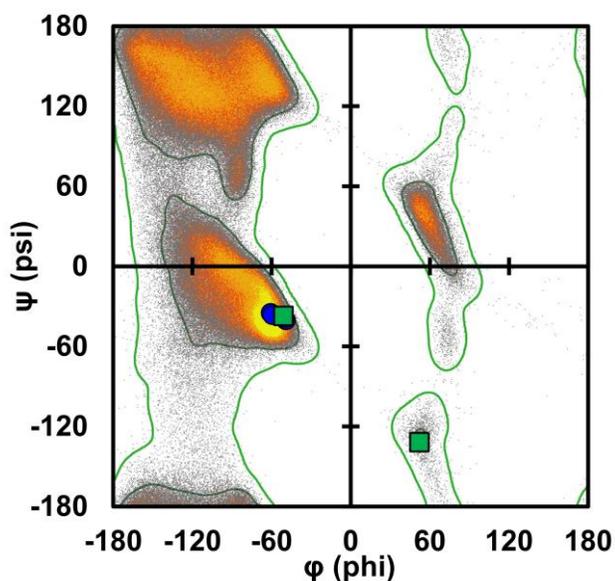


Figure 3.39. Ramachandran plot of right-handed helical conformations, C1' and C8' (dark blue circles) and C10' (green squares).

The dihedral angle pairs of C1' and C8' center around (-55°, -37°) and fall in the right-handed helix region. Their amide C→O vector angles are about 30°, which also demonstrates their helical structures. In C10', the N terminal acetyl group is bent away from the helix axis while maintaining its hydrogen bond with the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide group. As a result, one of the two dihedral angle pairs falls

out of the right-handed helix region, and one of its C→O vector angles increases to 99.9°. Comparing the theoretical IR spectra of this group of conformations to the IRMPD spectrum of AADapH<sup>+</sup>, both C1' and C10' have a good match across the whole range of wavenumbers, supporting their dominance in the structural population. C8' matches perfectly in the amide I and II regions, but it is significantly blue-shifted in the carboxyl C=O stretching region. Unlike C1' and C10', the C-O(H) oxygen in C8' accepts the weak hydrogen bond from the NH<sub>3</sub><sup>+</sup> group rather than the carboxyl C=O group. This grants C8' a stronger carboxyl C=O bond, blueshifts its carboxyl C=O stretching mode. The mismatch of C8' to the IRMPD spectrum rules out its dominance in the conformation population of AADapH<sup>+</sup>.

In C6' of AADapH<sup>+</sup> (Figure 3.40), with two left-handed turns on both alanine, the three amide C=O groups align with each other and form a left-handed helix backbone structure. The Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide group accepts a strong hydrogen bond from one hydrogen of the NH<sub>3</sub><sup>+</sup> group, while the Ac-Ala<sup>1st</sup> C=O group and the carboxyl C=O group form two weak hydrogen bonds with another hydrogen of the NH<sub>3</sub><sup>+</sup> group.

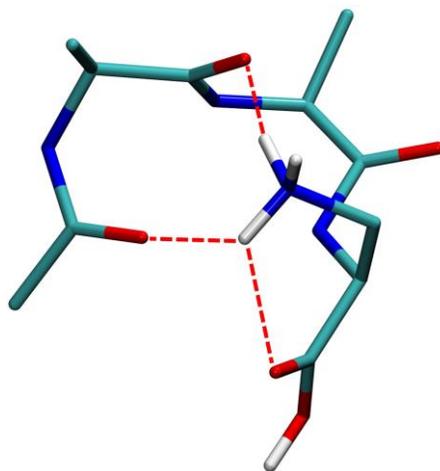


Figure 3.40. Structure of C6'. (The dashed lines indicate H-bonding.)

To demonstrate the left-handed helix structure of C6', its dihedral angles ( $\phi$ ,  $\psi$ ) from the C-terminus to the N-terminus are obtained and tabulated in Table 3.26 and overlaid to the Ramachandran

plot in Figure 3.41. Their C→O vector angles between adjacent amide C=O and carboxyl C=O groups are listed in Table 3.26 to show their alignments.

Table 3.26  
Dihedral Angles and C→O Vector Angles in C6' of AADapH<sup>+</sup>

Start Group	End Group	$\phi$ (phi) (°)		$\psi$ (psi) (°)	C→O Vector angle (°)
		C(O)–N(H)–C <sub>α</sub> –C(O)	N(H)–C <sub>α</sub> –C(O)–N(H)		
Ac-Ala <sup>1st</sup>	Ala <sup>1st</sup> -Ala <sup>2nd</sup>	53.2		50.5	24.9
C6' Ala <sup>1st</sup> -Ala <sup>2nd</sup>	Ala <sup>2nd</sup> -Dap	54.0		32.9	37.1
Ala <sup>2nd</sup> -Dap	carboxyl group				107.1

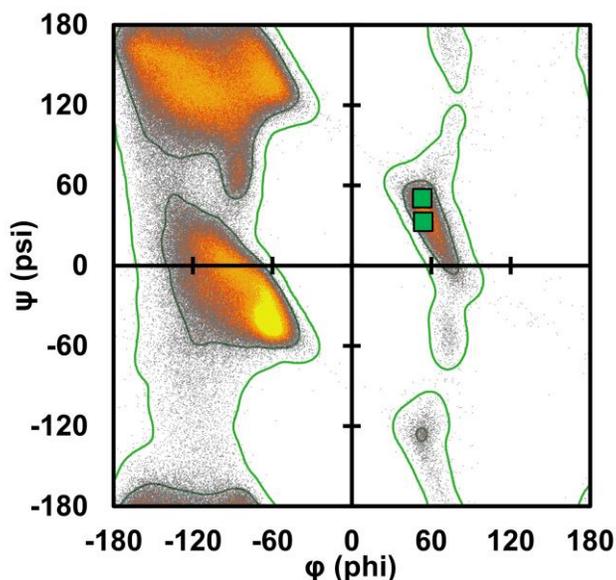


Figure 3.41. Ramachandran plot of the left-handed helical conformation C6' (green squares).

As shown in Table 3.26, the dihedral angle pairs are both around (54°, 40°) and fall in the left-handed helix region. The C→O vector angles between the amide C=O groups are small and demonstrate their good alignments. Comparing its theoretical IR spectrum with the AADapH<sup>+</sup> IRMPD spectrum, it has a good match in the carboxyl C=O stretching region and the amide II region. However, a poor match in the amide I region is observed. The peak on the far right of the amide I region, next to the carboxyl C=O stretching peak, is significantly blue-shifted. Investigation of its vibrational mode in GaussView

reveals that it corresponds to the Ala<sup>2nd</sup>-Dap amide I mode. The Ala<sup>2nd</sup>-Dap amide C=O group in C6' is not involved in any hydrogen bonding, making its C=O strength significantly higher than the other amide C=O groups. As a result, it is significantly blue-shifted and separated from the other amide I modes. The poor match to the IRMPD spectrum suggests the non-existence of this left-handed helix geometry.

In the third group of conformations, C2', C3', C4', C5', and C9' (Figure 3.42.), the Ac-Ala<sup>1st</sup> and Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O group are nearly parallel with each other, while the Ala<sup>2nd</sup>-Dap amide C=O group points to the back of the molecule. Different orientation of the Dap residue allows the NH<sub>3</sub><sup>+</sup> group to establish various hydrogen-bonding networks.

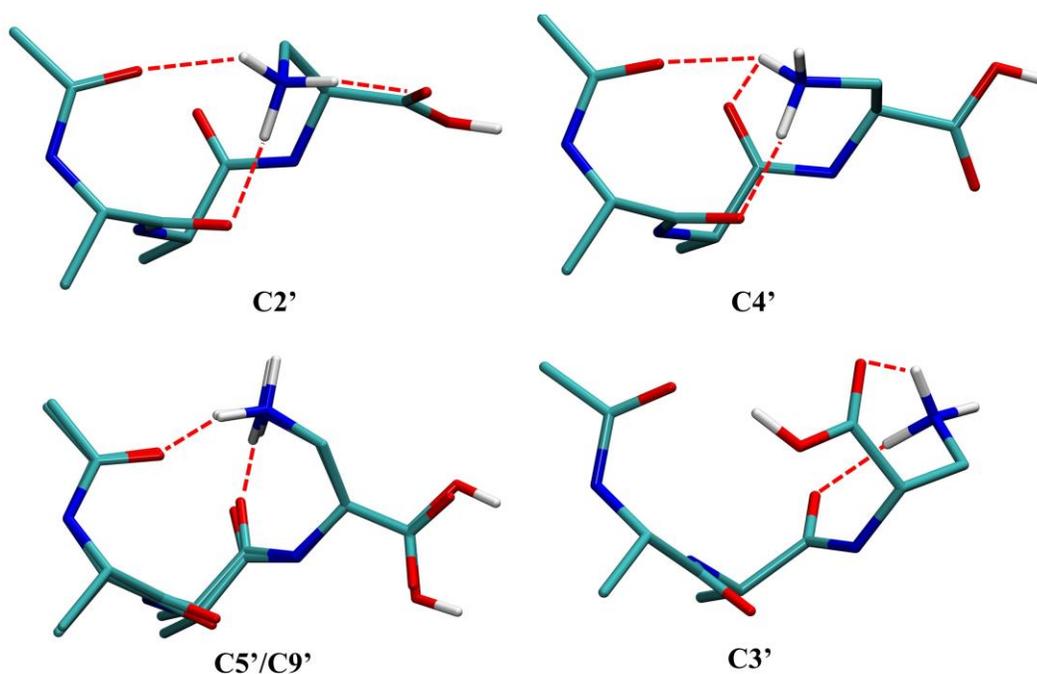


Figure 3.42. Structure of C2' (top left), C4' (top right), C5' and C9' (bottom left), and C3' (bottom right). (The dashed lines indicate H-bonding.)

In C2', the three hydrogens of the NH<sub>3</sub><sup>+</sup> group form hydrogen bonds with the Ac-Ala<sup>1st</sup> and Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O groups and the carboxyl C=O group. Compared to C2', the NH<sub>3</sub><sup>+</sup> group in C4' bends closer to the center of the molecule, losing its interaction with the carboxyl group. As a result, one hydrogen of the NH<sub>3</sub><sup>+</sup> group forms hydrogen bonds with the Ac-Ala<sup>1st</sup> and the Ala<sup>2nd</sup>-Dap amide C=O

groups, while another forms a hydrogen bond with the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O group. In C5' and C9', the NH<sub>3</sub><sup>+</sup> group is position on top of the molecule. Two hydrogen bonds are formed from the NH<sub>3</sub><sup>+</sup> group to the Ac-Ala<sup>1st</sup> and the Ala<sup>2nd</sup>-Dap amide C=O groups. In C3', the NH<sub>3</sub><sup>+</sup> group is position on the right side of the molecule and forms two hydrogen bonds with the Ala<sup>2nd</sup>-Dap amide and the carboxyl C=O groups.

Comparing the theoretical IR spectra of this group of conformations to the IRMPD spectrum of AADapH<sup>+</sup>, poor matchings are observed. Due to the isolation of the carboxyl C=O group in C4', C5', and C9', their carboxyl C=O stretching mode is significantly blue-shifted and does not match the IRMPD result. In C2', C5', and C9', the amide I band consists of three well-separated peaks, contradicting the experimentally observed single sharp peak in this region. In the amide II region, C3' has a significantly red-shifted peak compared to the IRMPD spectrum. Their poor matching to the IRMPD spectrum of AADapH<sup>+</sup> suggests the non-existence of this type of backbone geometry and the corresponding hydrogen-bonding networks.

In C7' (Figure 3.43), the three hydrogens of the NH<sub>3</sub><sup>+</sup> group hydrogen bonded with the Ac-Ala<sup>1st</sup> and the Ala<sup>2nd</sup>-Dap amide C=O groups, and the carboxyl C=O group, respectively. Comparing its spectrum to the AADapH<sup>+</sup> IRMPD spectrum, there are three well-resolved peaks in the amide I region, contradicting the experimentally observed single sharp peak and indicating the non-existence of this structure.

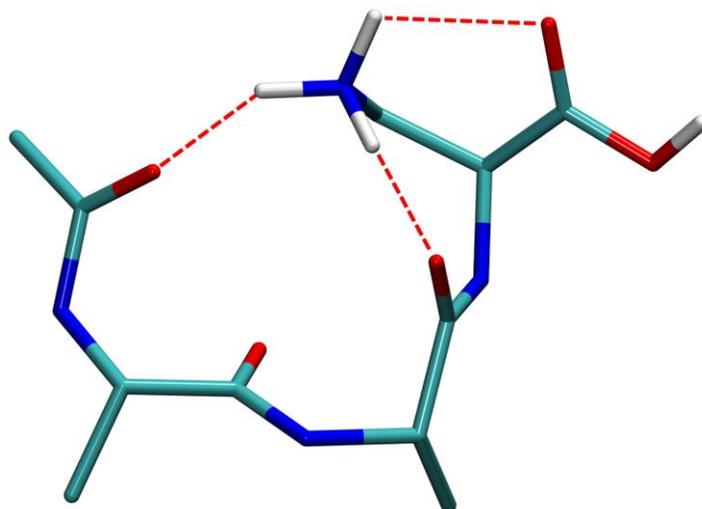
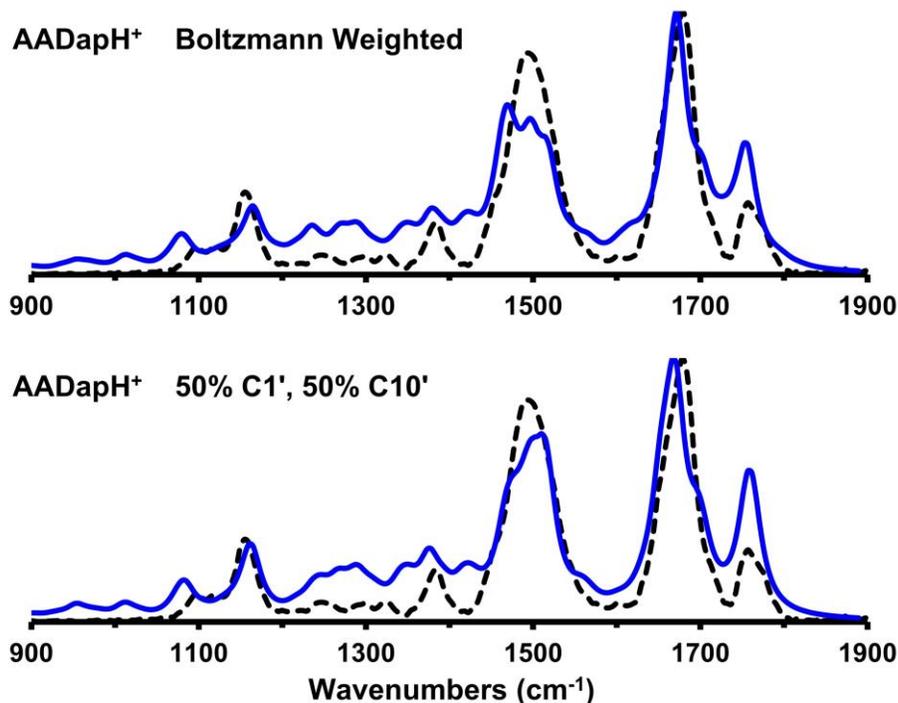


Figure 3.43. Structure of C7'. (The dashed lines indicate H-bonding.)

Based on the free energies, the structure of AADapH<sup>+</sup> could be primarily represented by C1'. However, C10' has a very similar match to the AADapH<sup>+</sup> IRMPD spectrum as C1'. Also, C1' and C10' are structurally similar and conversion from C1' to C10' is easily accessible. So, the high-energy conformation C10' is likely to co-exist with C1'. As a result, the structure of AADapH<sup>+</sup> could be primarily represented by conformation C1' and C10'. A hand-weighted spectrum from C1' and C10' is created and shown with the Boltzmann weighted average spectrum in Figure 3.44. As shown, the hand-weighted spectrum has slightly better matching to the AADapH<sup>+</sup> IRMPD spectrum than the Boltzmann weighted average spectrum.



*Figure 3.44.* Boltzmann weighted average spectra (top) and hand-weighted spectra (bottom) overlaid onto experimental IRMPD spectrum of AADapH<sup>+</sup>. (The dashed black plot represents the IRMPD spectrum, and the solid dark blue plots represent the calculated spectra.)

In summary, three conformations with similar backbone geometries, C1, C2, and C9, are selected as the representative conformations of DapAAH<sup>+</sup>. In those conformations, the Dap-Ala<sup>1st</sup> amide C=O group is not hydrogen bonded as the other two amide C=O groups, which matches the prediction of different microenvironments for the amide groups. Two right-handed helix conformations, C1' and C10', are selected as the representative conformations of AADapH<sup>+</sup>. Their concatenating hydrogen bonding networks match the predicted similar microenvironments for the amide bonds. Detailed IR matching and decomposition analysis are presented in the following sections.

### Structural and Energetic Comparison

As mentioned before, the stability difference in the protonated form is the main source of the different basicities of the two peptides. C1, C2, and C9 for DapAAH<sup>+</sup> and C1' and C10' for AADapH<sup>+</sup> (Table 3.27) are selected as their structurally representative conformations for further investigation. Like those studied for LysAH<sup>+</sup> and ALysH<sup>+</sup>, multiple energetic characterizations were performed on the

chosen conformations of DapAAH<sup>+</sup> and AADapH<sup>+</sup>, including the charge analysis, dipole moment decomposition (Lu & Chen, 2012), torsional strain calculation (Smith\* & Jaffe, 1996), hydrogen bonding analysis (Emamian et al., 2019), and RDG analysis (Johnson et al., 2010).

Table 3.27

*Relative Free Energy Values of the Representative Conformations of DapAAH<sup>+</sup> and AADapH<sup>+</sup>*

	Conf.	$\Delta G^a$
DapAAH <sup>+</sup>	C1	1.8
	C2	3.0
	C9	5.3
AADapH <sup>+</sup>	C1'	0.0
	C10'	2.8

<sup>a</sup>Relative free energy between selected conformations for the two peptide systems, DapAAH<sup>+</sup> and AADapH<sup>+</sup>, where the energy of C1' is set to be 0.0 kcal/mol.

**Charge distribution analysis and dipole moment decomposition.** The charge allocated on the NH<sub>3</sub><sup>+</sup> group is calculated and compared among the selected conformations. It provides valuable information about the different charge distributions in the protonated peptides. Three different types of charge, Mulliken, NPA, and Hirshfeld, on the NH<sub>3</sub><sup>+</sup> group, were calculated by summing the atomic charge on each of the four atoms for the DapAAH<sup>+</sup> and AADapH<sup>+</sup> representative conformations and are listed in Table 3.28. For clear visual comparison, the different types of charges are also illustrated as bar graphs in Figure 3.45.

Table 3.28

Four Types of Charges on the  $\text{NH}_3^+$  Group in the Representative Conformations of DapAAH<sup>+</sup> and AADapH<sup>+</sup>

	Conf.	$\Delta G$ kcal/mol	Mulliken <sup>a</sup>	NPA <sup>a</sup>	Hirshfeld <sup>b</sup>
DapAAH <sup>+</sup>	C1	1.8	0.646	0.669	0.492
	C2	3.0	0.651	0.666	0.490
	C9	5.3	0.636	0.614	0.398
AADapH <sup>+</sup>	C1'	0.0	0.648	0.617	0.402
	C10'	2.8	0.675	0.648	0.418

<sup>a</sup> Values (a.u.) obtained from single point calculations at B3LYP-D3(BJ)/ma-def2-TZVPP level of theory

<sup>b</sup> Values (a.u.) calculated by Multwfn 3.8 using the fchk file from single point calculations at B3LYP-D3(BJ)/ma-def2-TZVPP level of theory, with the build-in sphericalized atomic densities in free-states.

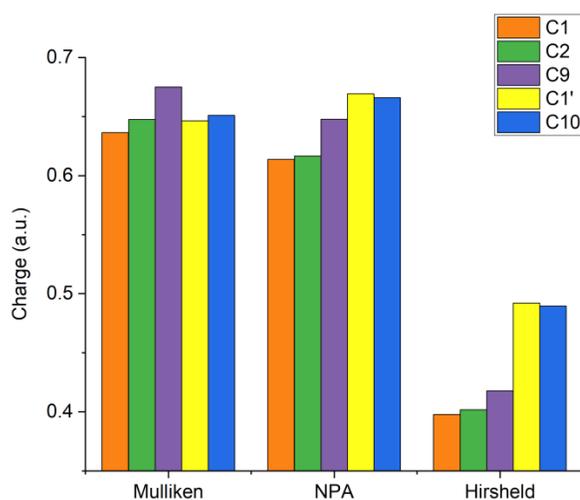


Figure 3.45. Different types of charges on the  $\text{NH}_3^+$  group in the representative conformations of DapAAH<sup>+</sup> and AADapH<sup>+</sup>.

Based on NPA and Hirshfeld charges, the  $\text{NH}_3^+$  group in the AADapH<sup>+</sup> conformations hold a higher charge than the DapAAH<sup>+</sup> conformations. In both cases, C1' of AADapH<sup>+</sup> is ranked with the highest charge, though it is the lowest in free energy. Similar to the observations in LysAH<sup>+</sup> and ALysH<sup>+</sup>, no simple correlation could be obtained between their relative free energies and their charges on the  $\text{NH}_3^+$  group.

The hydrogen bonds with the  $\text{NH}_3^+$  group play important roles in diffusing the charge on the  $\text{NH}_3^+$  group and stabilize the system. Comparison of the O --- H distances in the hydrogen bonds with the

$\text{NH}_3^+$  group would provide insight into their different degrees of charge accommodation. The O --- H distances in the hydrogen bonds with the  $\text{NH}_3^+$  group in the DapAAH<sup>+</sup> and AADapH<sup>+</sup> conformations are shown in Table 3.29. The hydrogen bonding patterns in those conformations (longer O --- H distance) indicate higher charges on the  $\text{NH}_3^+$  group in the AADapH<sup>+</sup> conformations.

Regarding the O---H distances, the hydrogen bonding patterns in C9 of DapAAH<sup>+</sup> and C1' and C10' of AADapH<sup>+</sup> are the same. They all have one strong hydrogen bond with an O --- H distance less than 1.80 Å and two weak hydrogen bonds with an O---H distance longer than 2.00 Å. Compared to C1' and C10' of AADapH<sup>+</sup>, C9 of DapAAH<sup>+</sup> has much shorter O --- H distances in both strong and weak hydrogen bonds. These advantages of C9 allow better charge diffusion from the  $\text{NH}_3^+$  group and support its lower charge than C1' and C10'. This trend is predicted correctly by the NPA and Hirshfeld charges, but not the Mulliken charges.

In C1 and C2 of DapAAH<sup>+</sup>, the  $\text{NH}_3^+$  group form two strong hydrogen bonds with the O --- H distances shorter than 1.66 Å. All three types of charges predict C1 and C2 to have lower charges than C9 of DapAAH<sup>+</sup> and C1' and C10' of AADapH<sup>+</sup>. The charge on the  $\text{NH}_3^+$  group in C1 and C2 is much better diffused through the two strong hydrogen bonds than the three hydrogen bonds system in C9 of DapAAH<sup>+</sup> and C1' and C10' of AADapH<sup>+</sup>.

Table 3.29

*O --- H Distance in the Hydrogen Bonds Involving the NH<sub>3</sub><sup>+</sup> in the Representative Conformations of DapAAH<sup>+</sup> and AADapH<sup>+</sup>*

	<b>Conf.</b>	<b>H Bond Acceptor</b>	<b>O --- H Distance(Å)</b>
DapAAH <sup>+</sup>	C1	Ac-Dap amide group	1.612
		Dap-Ala <sup>1st</sup> amide group	
		Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide group	1.647
		Carboxyl C=O group	
	C2	Ac-Dap amide group	1.615
		Dap-Ala <sup>1st</sup> amide group	
		Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide group	1.656
		Carboxyl C=O group	
	C9	Ac-Dap amide group	1.612
Dap-Ala <sup>1st</sup> amide group			
Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide group		2.133	
Carboxyl C=O group		2.021	
AADapH <sup>+</sup>	C1'	Ac-Ala <sup>1st</sup> amide group	
		Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide group	1.746
		Ala <sup>2nd</sup> -Dap amide group	2.418
		Carboxyl C=O group	2.199
	C10'	Ac-Ala <sup>1st</sup> amide group	
		Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide group	1.738
		Ala <sup>2nd</sup> -Dap amide group	2.217
		Carboxyl C=O group	2.396

Besides the total charge on the NH<sub>3</sub><sup>+</sup> group, attention is also applied to the atomic charge of its individual hydrogens. Similar to the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations, one hydrogen of the NH<sub>3</sub><sup>+</sup> group carries a slightly higher charge because it is isolated rather than hydrogen bonded. Three types of atomic charges of the nitrogen and hydrogens in the NH<sub>3</sub><sup>+</sup> group, together with their atomic labels, are listed in Table 3.30.

Table 3.30

Atomic Charges on the Nitrogen and Hydrogens of the  $\text{NH}_3^+$  Group in the Representative Conformations of  $\text{DapAAH}^+$  and  $\text{AADapH}^+$

	Conf.	Atom <sup>a</sup>	Mulliken	NPA	Hirshfeld
DapAAH <sup>+</sup>	C1	N <sub>8</sub>	-0.028	-0.727	-0.002
		H <sub>9</sub> (isolated)	0.182	0.419	0.176
		H <sub>40</sub>	0.232	0.460	0.116
	C2	H <sub>41</sub>	0.250	0.462	0.107
		N <sub>8</sub>	-0.023	-0.725	0.000
		H <sub>9</sub> (isolated)	0.182	0.419	0.177
	C9	H <sub>40</sub>	0.235	0.460	0.116
		H <sub>41</sub>	0.254	0.462	0.109
		N <sub>8</sub>	-0.015	-0.713	0.010
AADapH <sup>+</sup>	C1'	H <sub>9</sub> (isolated)	0.176	0.421	0.175
		H <sub>40</sub>	0.247	0.464	0.116
		H <sub>41</sub>	0.267	0.475	0.117
	C10'	N <sub>28</sub>	-0.045	-0.691	0.030
		H <sub>40</sub> (isolated)	0.191	0.426	0.186
		H <sub>41</sub>	0.270	0.473	0.123
	C10'	H <sub>29</sub>	0.230	0.461	0.154
		N <sub>28</sub>	-0.031	-0.695	0.029
		H <sub>40</sub> (isolated)	0.191	0.427	0.185
C10'	H <sub>41</sub>	0.268	0.472	0.122	
	H <sub>29</sub>	0.223	0.462	0.154	

<sup>a</sup>The number subscript refers to the atomic label in the molecular model.

Like that for  $\text{LysAH}^+$  and  $\text{ALysH}^+$ , the isolated hydrogen atoms in  $\text{DapAAH}^+$  and  $\text{AADapH}^+$  carry higher Hirshfeld charges than the other two hydrogens, in agreement with the prediction. The Mulliken and NPA charges do not correctly reflect the charge distribution as expected. So, for  $\text{DapAAH}^+$  and  $\text{AADapH}^+$ , the Hirshfeld charge describes the distribution better than the Mulliken and NPA charges.

Dipole moment decomposition (Lu & Chen, 2012) is performed to investigate the interaction of the backbone dipole moment with the  $\text{NH}_3^+$  group. A strong stabilization interaction provides additional accommodation of the charged  $\text{NH}_3^+$  group, besides the hydrogen bonds. The dipole moment of the whole ion,  $\mu_{\text{whole}}$ , is decomposed into the dipole moment of the molecular fragment excluding the  $\text{NH}_3^+$

group, noted as the backbone dipole moment,  $\mu_{\text{backbone}}$ , and the dipole moment of the  $\text{NH}_3^+$  group,  $\mu_{\text{NH}_3}$ .

The three dipole moments for the DapAAH<sup>+</sup> and AADapH<sup>+</sup> conformations are shown in Table 3.34 and visualized in Figure 3.46. To qualitatively evaluate the alignment and interaction between  $\mu_{\text{backbone}}$  and  $\mu_{\text{NH}_3}$ , their vector angle and dot product are manually calculated and summarized in Table 3.34.

Table 3.31

*Dipole Moments of the DapAAH<sup>+</sup> and AADapH<sup>+</sup> Representative Conformations*

	<b>Conf.</b>	$\Delta G$ (kcal/mol)	$\mu_{\text{whole}}$ (D)	$\mu_{\text{NH}_3}$ (D)	$\mu_{\text{backbone}}$ (D)	<b>Vector angle between <math>\mu_{\text{NH}_3}</math> and <math>\mu_{\text{backbone}}</math> (°)</b>	<b>Dot product of <math>\mu_{\text{NH}_3}</math> and <math>\mu_{\text{backbone}}</math> (D<sup>2</sup>)</b>
DapAAH <sup>+</sup>	C1	1.8	7.37	5.77	5.78	100.78	-6.24
	C2	3.0	5.70	5.44	5.72	118.63	-14.91
	C9	5.3	3.42	4.59	5.55	142.14	-20.14
AADapH <sup>+</sup>	C1'	0.0	4.91	8.09	10.95	155.58	-80.68
	C10'	2.8	5.22	7.34	9.25	145.68	-56.06

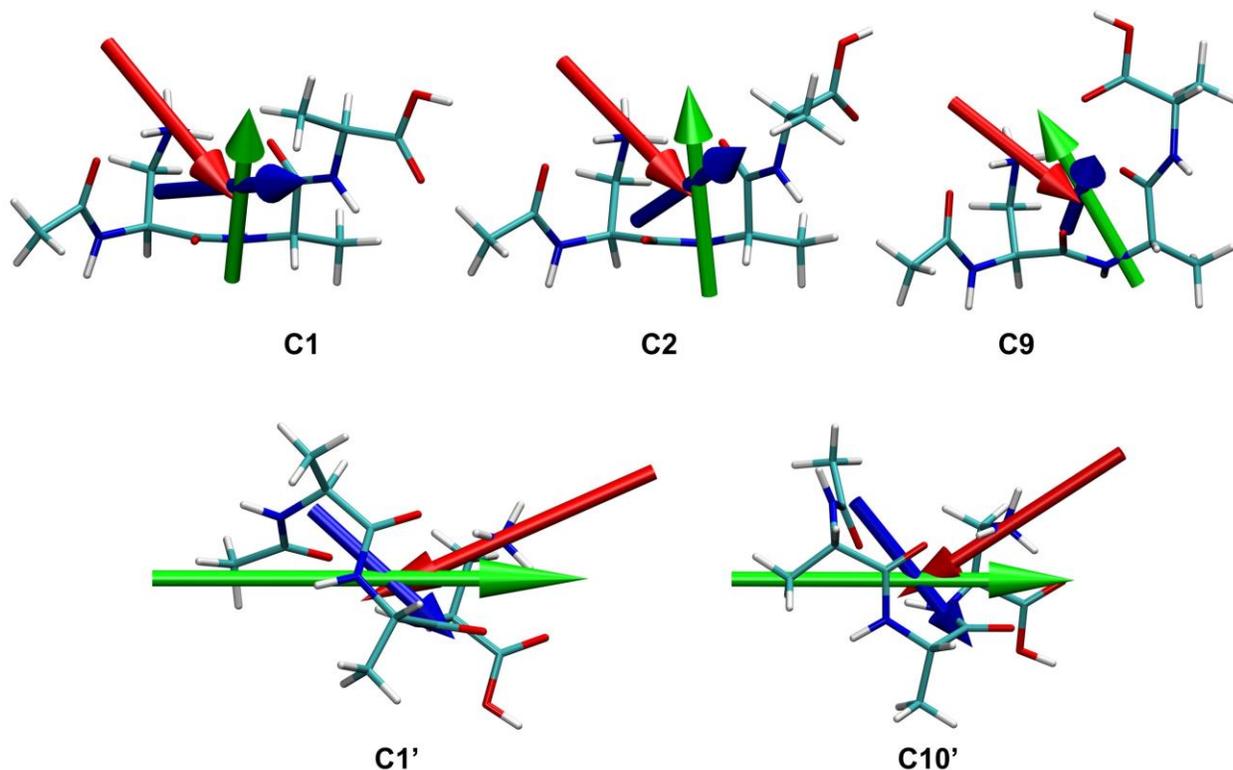


Figure 3.46. Dipole moments of whole molecule (blue), the backbone (green) and the  $\text{NH}_3^+$  group (red) for representative conformations of DapAAH<sup>+</sup> (top) and AADapH<sup>+</sup> (bottom).

The  $\mu_{\text{NH}_3}$  magnitudes in the DapAAH<sup>+</sup> conformations range from 4.59 to 5.77 D. Comparatively, the  $\mu_{\text{NH}_3}$  magnitudes in the AADapH<sup>+</sup> conformations are larger and range from 7.34 to 8.09 D. In the AADapH<sup>+</sup> conformations, the charge on the  $\text{NH}_3^+$  group is less diffused through the hydrogen bonds, and a higher charge is allocated onto the  $\text{NH}_3^+$  group. So, their larger  $\mu_{\text{NH}_3}$  magnitudes also represent a higher necessity for additional accommodation of the  $\text{NH}_3^+$  group in ways besides hydrogen bonds.

Comparatively, their difference in  $\mu_{\text{backbone}}$  is even larger. The  $\mu_{\text{backbone}}$  magnitudes of the DapAAH<sup>+</sup> conformations are similar to their  $\mu_{\text{NH}_3}$ , and range from 5.55 to 5.78 D. In the AADapH<sup>+</sup> conformations, the C=O groups are aligned well. As a result, their  $\mu_{\text{backbone}}$  magnitudes are much larger and range from 9.25 to 10.95 D.

Besides the difference in their magnitudes, the alignment of  $\mu_{\text{NH}_3}$  and  $\mu_{\text{backbone}}$  also plays an important role in the cancellation of  $\mu_{\text{NH}_3}$ . As shown in Figure 3.48,  $\mu_{\text{NH}_3}$  and  $\mu_{\text{backbone}}$  in the AADapH<sup>+</sup> conformations are better oppositely aligned. It allows the  $\mu_{\text{backbone}}$  to better interact with and stabilize the

$\text{NH}_3^+$  group. The vector angles between the  $\mu_{\text{NH}_3}$  and  $\mu_{\text{backbone}}$  in the DapAAH<sup>+</sup> conformations range from 100.78° to 142.14°, while that of the AADapH<sup>+</sup> conformations range from 145.68° to 155.58°. The differences in their magnitudes and alignment result in dramatic differences in the cancellation of  $\mu_{\text{NH}_3}$ . The dot product of  $\mu_{\text{NH}_3}$  and  $\mu_{\text{backbone}}$  in the DapAAH<sup>+</sup> conformations range from -6.24 to -20.14 D<sup>2</sup>, while that in the AADapH<sup>+</sup> conformations range from -56.06 to -80.68 D<sup>2</sup>. The much larger absolute value of their dot products represents the stronger interaction between  $\mu_{\text{NH}_3}$  and  $\mu_{\text{backbone}}$  in the AADapH<sup>+</sup> conformations. As a result, though with much larger  $\mu_{\text{NH}_3}$ , the  $\mu_{\text{whole}}$  in C1' and C10' are 4.91 and 5.22 D, smaller than those of C1 and C2 of DapAAH<sup>+</sup>.

Among the AADapH<sup>+</sup> conformations, the alignment of  $\mu_{\text{NH}_3}$  and  $\mu_{\text{backbone}}$  in C1' is better. Therefore, its  $\mu_{\text{NH}_3}$  is canceled out more efficiently, resulting in a smaller  $\mu_{\text{whole}}$ . In C10', the bent of the acetyl group away from the helix axis decreases the  $\mu_{\text{backbone}}$  interaction with the  $\text{NH}_3^+$  group. The absolute value of its dot product between  $\mu_{\text{NH}_3}$  and  $\mu_{\text{backbone}}$  is 56.06 D<sup>2</sup>, smaller than that for C1'.

In summary, due to their different degrees of charge diffusion through hydrogen bonds, the  $\text{NH}_3^+$  group is allocated with a higher Hirshfeld charge in the DapAAH<sup>+</sup> conformations than the AADapH<sup>+</sup> conformations. The C=O groups are aligned in the helical AADapH<sup>+</sup> conformations, a large backbone dipole moment is created to interact with and stabilizes the  $\text{NH}_3^+$  group at the C-terminus. It compensates for the disadvantages in stabilizing the  $\text{NH}_3^+$  group through hydrogen bonds and enhances the stability of the AADapH<sup>+</sup> conformations.

**Torsional strain on Dap sidechain.** Though there is only one carbon, the torsional strain on the Dap sidechain is calculated and compared between the selected conformations. Like the procedures used for LysAH<sup>+</sup> and ALysH<sup>+</sup>, the dihedral angles on the Dap sidechain are plugged into the torsional strain function for n-butane to calculate their individual torsional strains. The dihedral angles on the Dap sidechain in the DapAAH<sup>+</sup> and AADapH<sup>+</sup> conformations are listed in Table 3.32. Their corresponding torsional strain values are listed in Table 3.33.

The total torsional strain on the Dap sidechain noted as  $E_{\text{torsion}}$  is calculated by summing the torsional strain of all the dihedral angles. C1' of AADapH<sup>+</sup> is assigned as the reference of relative total

torsional strain, Rel.  $E_{\text{torsion}}$ , with 0.0 kcal/mol of energy. The Rel.  $E_{\text{torsion}}$  values of the other conformations are calculated accordingly. The  $E_{\text{torsion}}$  and Rel.  $E_{\text{torsion}}$  values of the DapAAH<sup>+</sup> and AADapH<sup>+</sup> conformations are also listed in Table 3.33.

Table 3.32

*Dihedral Angles on the Dap Sidechain in the DapAAH<sup>+</sup> and AADapH<sup>+</sup> Conformations*

Conf.	Dihedral Angle (°)	
	$\alpha_1$ N(H)-C $_{\alpha}$ -C $_{\beta}$ -N(H3)	$\alpha_2$ C(O)-C $_{\alpha}$ -C $_{\beta}$ -N(H3)
DapAAH <sup>+</sup>	C1	82.67
	C2	82.79
	C9	77.51
AADapH <sup>+</sup>	C1'	63.34
	C10'	57.06

Table 3.33  
*The Torsional Strain on the Dap Sidechain in the DapAAH<sup>+</sup> and AADapH<sup>+</sup> Conformations*

Conf.	Torsional Strain (kcal/mol)				
	$\alpha_1$	$\alpha_2$	$E_{\text{torsion}}$	Rel. $E_{\text{torsion}}$	
	N(H)-C $_{\alpha}$ -C $_{\beta}$ -N(H3)	C(O)-C $_{\alpha}$ -C $_{\beta}$ -N(H3)			
DapAAH <sup>+</sup>	C1	1.5	1.3	2.7	1.4
	C2	1.5	1.3	2.8	1.4
	C9	1.1	1.0	2.1	0.8
AADapH <sup>+</sup>	C1'	0.7	0.7	1.4	0.0
	C10'	0.7	0.8	1.5	0.2

In the AADapH<sup>+</sup> conformations, the dihedral angles on the Dap sidechain are all about  $\pm 60^\circ$ , while in the DapAAH<sup>+</sup> conformations, the dihedral angles are slightly further away from this local minimum. As a result, the torsional strain on the Dap sidechain in the DapAAH<sup>+</sup> conformations are higher. The  $E_{\text{torsion}}$  values are 1.4 and 1.5 kcal/mol for C1' and C10' of AADapH<sup>+</sup>, and 2.7, 2.8, and 2.1 kcal/mol for C1, C2, and C9 of DapAAH<sup>+</sup>. C1' and C10' have nearly the same torsional strain. The Rel.  $E_{\text{torsion}}$  of C1, C2, and C9 are 1.4, 1.4, and 0.8 kcal/mol, respectively.

In the DapAAH<sup>+</sup> conformations, the NH<sub>3</sub><sup>+</sup> group is hydrogen-bonded to the Ac-Dap and Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O groups. The NH<sub>3</sub><sup>+</sup> group is further away from the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O group than the Ac-Dap amide group. The Dap sidechain rotates to the right slightly to form a hydrogen bond with the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O group. It is reflected in the smaller absolute values of the dihedral angles  $\alpha_2$  than  $\alpha_1$ , as shown in Table 3.32. It increases the torsional strain on the Dap sidechain. Though the stability enhancement from hydrogen bonding is greater, the increase of torsional strain on the Dap sidechain counteracts the benefit of the hydrogen bonding. It contributes to the observed lower stability of the DapAAH<sup>+</sup> conformations.

are mult  
formed l  
interacti  
and its t  
bonding  
distance  
obtaine  
hydroge  
total hyc  
system.

Table 3.34  
*Geometry Information of the Hydrogen Bonds in the DapAAH<sup>+</sup> and AADapH<sup>+</sup> Conformations*

Conf.	Hydrogen Bond	Hydrogen Bond Donor	Hydrogen Bond Acceptor	C=O ... H O ... H-N angle (°)	Alignment angle (°)	Vector angle (°)	O ... H Distance (Å)
C1	HB <sub>1</sub>	NH <sub>3</sub> <sup>+</sup> group	Ac-Dap amide C=O	123.1	153.3	276.4	1.612
	HB <sub>2</sub>	NH <sub>3</sub> <sup>+</sup> group	Dap-Ala <sup>1st</sup> amide C=O	133.6	174.9	308.5	1.647
DapAAH <sup>+</sup>	HB <sub>1</sub>	NH <sub>3</sub> <sup>+</sup> group	Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide C=O	123.0	153.0	276.0	1.615
	HB <sub>2</sub>	NH <sub>3</sub> <sup>+</sup> group	Carboxyl C=O	133.5	176.6	310.1	1.656
C9	HB <sub>1</sub>	NH <sub>3</sub> <sup>+</sup> group	Ac-Dap amide C=O	121.4	152.6	273.9	1.612
	HB <sub>2</sub>	NH <sub>3</sub> <sup>+</sup> group	Dap-Ala <sup>1st</sup> amide C=O	103.1	133.2	236.3	2.133
	HB <sub>3</sub>	NH <sub>3</sub> <sup>+</sup> group	Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide C=O	135.5	132.1	267.6	2.021
C1'	HB <sub>1</sub>	Ala <sup>2nd</sup> -Dap amide N-H	Ac-Ala <sup>1st</sup> amide C=O	125.2	151.3	276.5	1.962
	HB <sub>2</sub>	NH <sub>3</sub> <sup>+</sup> group	Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide C=O	140.3	157.0	297.3	1.746
	HB <sub>3</sub>	NH <sub>3</sub> <sup>+</sup> group	Ala <sup>2nd</sup> -Dap amide C=O	105.5	105.1	210.6	2.418
	HB <sub>4</sub>	NH <sub>3</sub> <sup>+</sup> group	Carboxyl C=O	92.6	122.1	214.7	2.199
C10'	HB <sub>1</sub>	Ala <sup>2nd</sup> -Dap amide N-H	Ac-Ala <sup>1st</sup> amide C=O	148.4	141.5	289.9	2.019
	HB <sub>2</sub>	NH <sub>3</sub> <sup>+</sup> group	Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide C=O	142.1	163.0	305.1	1.738
	HB <sub>3</sub>	NH <sub>3</sub> <sup>+</sup> group	Ala <sup>2nd</sup> -Dap amide C=O	104.7	114.1	218.7	2.217
	HB <sub>4</sub>	NH <sub>3</sub> <sup>+</sup> group	Carboxyl C=O	87.3	116.0	203.2	2.396

The  
1 the

there  
ds are  
covalent  
n bond  
ogen



Table 3.35  
Energetic Information of Hydrogen Bonds in the DapAAH<sup>+</sup> and AADapH<sup>+</sup> Conformations

Conf.	Hydrogen Bond	Hydrogen Bond Donor	Hydrogen Bond Acceptor	Electron Density (a.u.)	HBE (kcal/mol)	HBE <sub>total</sub> (kcal/mol)	Rel. HBE <sub>total</sub> (kcal/mol)
C1	HB <sub>1</sub>	NH <sub>3</sub> <sup>+</sup> group	Ac-Dap amide C=O	6.12E-02	21.4		
			Dap-Ala <sup>1st</sup> amide C=O			39.8	7.8
	HB <sub>2</sub>	NH <sub>3</sub> <sup>+</sup> group	Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide C=O	5.22E-02	18.4		
			Carboxyl C=O				
DapAAH <sup>+</sup>	HB <sub>1</sub>	NH <sub>3</sub> <sup>+</sup> group	Ac-Dap amide C=O	6.08E-02	21.3		
			Dap-Ala <sup>1st</sup> amide C=O			39.2	7.2
	HB <sub>2</sub>	NH <sub>3</sub> <sup>+</sup> group	Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide C=O	5.09E-02	18.0		
			Carboxyl C=O				
C9	HB <sub>1</sub>	NH <sub>3</sub> <sup>+</sup> group	Ac-Dap amide C=O	6.11E-02	21.4		
			Dap-Ala <sup>1st</sup> amide C=O			36.1	4.1
	HB <sub>2</sub>	NH <sub>3</sub> <sup>+</sup> group	Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide C=O	1.70E-02	6.7		
	HB <sub>3</sub>	NH <sub>3</sub> <sup>+</sup> group	Carboxyl C=O	2.10E-02	8.1		
C1'	HB <sub>1</sub>	Ala <sup>2nd</sup> -Dap amide N-H	Ac-Ala <sup>1st</sup> amide C=O	2.40E-02	6.1		
	HB <sub>2</sub>	NH <sub>3</sub> <sup>+</sup> group	Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide C=O	3.96E-02	14.2		
	HB <sub>3</sub>	NH <sub>3</sub> <sup>+</sup> group	Ala <sup>2nd</sup> -Dap amide C=O	1.21E-02	5.1	32.0	0.0
	HB <sub>4</sub>	NH <sub>3</sub> <sup>+</sup> group	Carboxyl C=O	1.67E-02	6.6		
AADapH <sup>+</sup>	HB <sub>1</sub>	Ala <sup>2nd</sup> -Dap amide N-H	Ac-Ala <sup>1st</sup> amide C=O	2.09E-02	5.4		
	HB <sub>2</sub>	NH <sub>3</sub> <sup>+</sup> group	Ala <sup>1st</sup> -Ala <sup>2nd</sup> amide C=O	4.05E-02	14.5		
	HB <sub>3</sub>	NH <sub>3</sub> <sup>+</sup> group	Ala <sup>2nd</sup> -Dap amide C=O	1.66E-02	6.6	31.5	-0.5
	HB <sub>4</sub>	NH <sub>3</sub> <sup>+</sup> group	Carboxyl C=O	1.17E-02	4.9		

There are two strong hydrogen bonds in C1 and C2 of DapAAH<sup>+</sup>, noted as HB<sub>1</sub> HB<sub>2</sub>, from the NH<sub>3</sub><sup>+</sup> group. HB<sub>1</sub> has an O --- H distance of about 1.61 Å and a vector angle of 82°. HB<sub>2</sub> has a slightly longer O --- H distance of about 1.65 Å, while its C=O group and H-N group are better aligned with a vector angle of 44°. The estimated HBE is 21.4 kcal/mol for HB<sub>1</sub> and 18.0 kcal/mol for HB<sub>2</sub>. With no other identifiable hydrogen bonding interactions, the HBE of those two hydrogen bonds adds up to an HBE<sub>total</sub> of more than 39 kcal/mol.

C9 also has a short hydrogen bond with the acetyl group, HB<sub>1</sub>, of 21.4 kcal/mol. Compared to C1 and C2, the hydrogen bond with the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide group, HB<sub>2</sub>, in C9 is elongated to 2.133 Å and weakened to 6.7 kcal/mol. The additional hydrogen bond between the NH<sub>3</sub><sup>+</sup> and the carboxyl group, HB<sub>3</sub>, has an HBE of 8.1 kcal/mol. The HBE<sub>total</sub> of the three hydrogen bonds is 36.1 kcal/mol, lower than C1 and C2.

In the AADapH<sup>+</sup> conformations, all the C=O groups are hydrogen-bonded. In both C1' and C10', the NH<sub>3</sub><sup>+</sup> group forms a strong hydrogen bond, HB<sub>2</sub>, with the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide group. This hydrogen bond is stronger than the other three hydrogen bonds in the same conformation. HB<sub>2</sub> is equipped with an O --- H distance around 1.75 Å and a vector angle around 50°. Its HBE is about 14.5 kcal/mol, although weaker than the HB<sub>1</sub> in C1, C2, and C9 of DapAAH<sup>+</sup>. From another hydrogen of the NH<sub>3</sub><sup>+</sup> group, there are two hydrogen bonds, HB<sub>3</sub> and HB<sub>4</sub>, to the Ala<sup>2nd</sup>-Dap amide and the carboxyl C=O groups, with much longer O --- H distances ranging from 2.2 to 2.4 Å. The C=O group and the N-H group of those two hydrogen bonds are poorly aligned, with vector angles of more than 100°. The long O --- H distance and the poor alignment make these two hydrogen bonds very weak, with estimated HBE values from 4.9 to 6.6 kcal/mol.

As the helix extension, there is also a hydrogen bond, HB<sub>1</sub>, between the Ac-Ala<sup>1st</sup> and the Ala<sup>2nd</sup>-Dap amide groups. In C1', the C=O and H-N groups are aligned well with a vector angle of 31.5°. In C10', the acetyl group is bent away from the helix axis, resulting in a vector angle of 59.2°. With an O --- H distance close to 2.0 Å, the HBE of this hydrogen bond is estimated to be 6.1 kcal/mol in C1' and 5.4

kcal/mol in C10'. Though all C=O groups are hydrogen-bonded, the  $HBE_{total}$  of C1' and C10' are 32.0 and 31.5 kcal/mol, respectively, lower than that in the DapAAH<sup>+</sup> conformations.

In summary, the  $HBE_{total}$  values are 39.8, 39.2, and 36.1 kcal/mol for C1, C2, and C9 of DapAAH<sup>+</sup>, and 32.0 and 31.5 kcal/mol for C1' and C10' of AADapH<sup>+</sup>. The conformation with the lowest free energy, C1' of AADapH<sup>+</sup>, is chosen as the reference of Rel.  $HBE_{total}$ , where its Rel.  $HBE_{total}$  is set to be 0.0 kcal/mol. The Rel.  $HBE_{total}$  of the other conformations are calculated accordingly. The  $HBE_{total}$  of C10' is slightly lower than C1', and it has a Rel.  $HBE_{total}$  of -0.5 kcal/mol. The  $HBE_{total}$  of C1, C2, and C9 of DapAAH<sup>+</sup> are much higher than C1', and their Rel.  $HBE_{total}$  values are 7.8, 7.2, and 4.1 kcal/mol, respectively.

Though AADapH<sup>+</sup> conformations have a much lower total strength of the hydrogen bonding interactions, they are lower in free energy. It suggests that there are other beneficial intramolecular interactions in the AADapH<sup>+</sup> conformations. The previously identified strong interaction of the backbone dipole moment with the NH<sub>3</sub><sup>+</sup> group could be a significant contributor. Besides that, other noncovalent interactions might also contribute to their higher stabilities.

**RDG analysis.** The RDG analysis is very useful for visualizing and semi-quantitatively evaluating multiple types of noncovalent interactions simultaneously. It provides valuable information about the Van Der Waals interactions and steric clashes in the system, which are unavailable from other methods. Detailed procedures of the RDG analysis can be found in Chapter 2. The RDG scatter plot and isosurface diagrams for the DapAAH<sup>+</sup> and AADapH<sup>+</sup> conformations are shown in Figure 3.47 and Figure 3.48, respectively. In both the RDG scatter plot and the isosurface diagram, the attractive interactions are colored in blue to green, while the repulsive interactions are colored in brown to red.

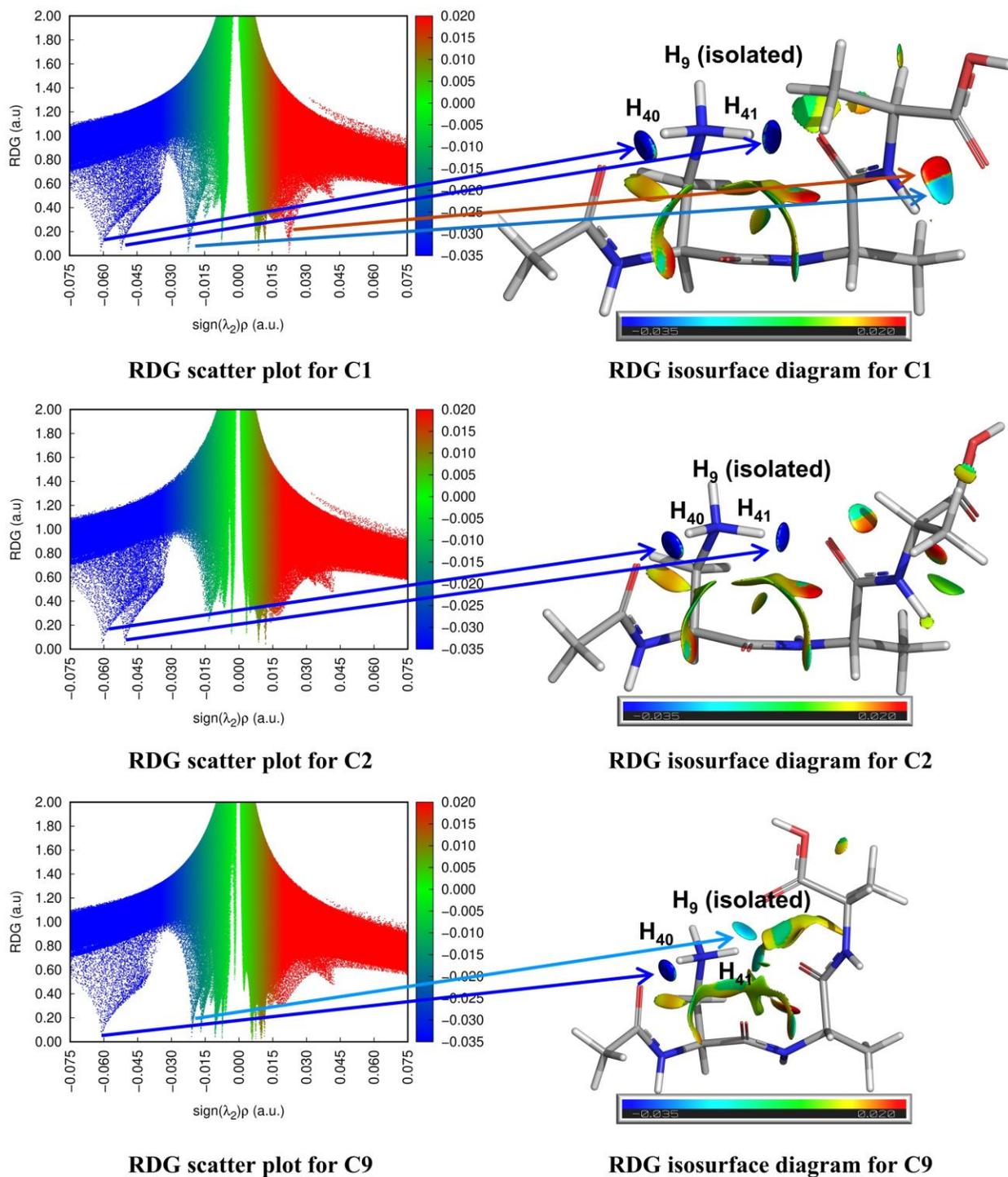


Figure 3.47. RDG scatter plot (left) and isosurface diagram (right) of C1 (top), C2 (middle) and C9 (bottom) of DapAAH<sup>+</sup>. (The arrows show examples of correlation between the spikes in the RDG scatter plot and the features in the RDG isosurface diagram.)

In C1 of DapAAH<sup>+</sup>, there are twelve distinct attractive interactions between the following pairs of atoms and are listed using bullet points:

- H<sub>40</sub> on the NH<sub>3</sub><sup>+</sup> group and the acetyl C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP42}} = 0.61\text{E-}01\text{a.u.}$ , HBE = 21.4 kcal/mol),
- H<sub>41</sub> on the NH<sub>3</sub><sup>+</sup> group and the Ala<sup>1st</sup> C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP44}} = 0.52\text{E-}01\text{a.u.}$ , HBE = 18.4 kcal/mol),
- The acetyl C=O oxygen and the Dap C=O oxygen,
- The Dap C=O oxygen and the Dap N-H nitrogen,
- The Dap C=O oxygen and the Ala<sup>1st</sup> C=O oxygen,
- The Dap C=O oxygen and the Ala<sup>1st</sup> alpha hydrogen,
- The Ala<sup>1st</sup> C=O oxygen and the Ala<sup>1st</sup> N-H nitrogen,
- The Ala<sup>1st</sup> C=O oxygen and the Ala<sup>2nd</sup> alpha hydrogen,
- The Ala<sup>1st</sup> C=O oxygen and one of the Ala<sup>2nd</sup> beta hydrogens,
- The carboxyl C=O oxygen and the Ala<sup>2nd</sup> N-H hydrogen ( $\rho_{\text{BCP}} \sim 0.22\text{E-}01\text{a.u.}$ ),
- The carboxyl C-O(H) oxygen and one of the Ala<sup>2nd</sup> beta hydrogens,
- The Ala<sup>1st</sup> N-H hydrogen and one of the Dap beta hydrogens.

In C1 of DapAAH<sup>+</sup>, there are nine distinct steric clashes in opened ring structures made of the following groups of atoms and are listed using bullet points:

- From the acetyl C=O oxygen to H<sub>40</sub> on the NH<sub>3</sub><sup>+</sup> group,
- From the Dap N-H nitrogen to the Dap C=O oxygen,
- From one of the Dap beta hydrogens to the Ala<sup>1st</sup> N-H hydrogen,
- From the Dap C=O oxygen to the Ala<sup>1st</sup> alpha hydrogen,
- From the Ala<sup>1st</sup> N-H nitrogen to the Ala<sup>1st</sup> C=O oxygen,
- From the Ala<sup>1st</sup> C=O oxygen to the Ala<sup>2nd</sup> alpha hydrogen,
- From the Ala<sup>1st</sup> C=O oxygen to one of the Ala<sup>2nd</sup> beta hydrogens,

- From the Ala<sup>2nd</sup> N-H hydrogen to the carboxyl C=O oxygen ( $\rho_{BCP} \sim 0.22E-01a.u.$ ),
- From one of the Ala<sup>2nd</sup> beta hydrogens to the carboxyl C-O(H) oxygen.

In C2 of DapAAH<sup>+</sup>, there are twelve distinct attractive interactions between the following atom pairs and are listed using bullet points:

- H<sub>40</sub> on the NH<sub>3</sub><sup>+</sup> group and the acetyl C=O oxygen (strong hydrogen bond,  $\rho_{BCP42} = 0.61E-01a.u.$ , HBE = 21.3 kcal/mol),
- H<sub>41</sub> on the NH<sub>3</sub><sup>+</sup> group and the Ala<sup>1st</sup> C=O oxygen (strong hydrogen bond,  $\rho_{BCP44} = 0.51E-01a.u.$ , HBE = 18.0 kcal/mol),
- The acetyl C=O oxygen and the Dap C=O oxygen,
- The Dap C=O oxygen and the Dap N-H nitrogen,
- The Dap C=O oxygen and the Ala<sup>1st</sup> C=O oxygen,
- The Dap C=O oxygen and the Ala<sup>1st</sup> alpha hydrogen,
- The Ala<sup>1st</sup> C=O oxygen and the Ala<sup>1st</sup> N-H nitrogen,
- The Ala<sup>1st</sup> C=O oxygen and the Ala<sup>2nd</sup> alpha hydrogen,
- The carboxyl C=O oxygen and one of the Ala<sup>1st</sup> beta hydrogens,
- The carboxyl C=O oxygen and the Ala<sup>2nd</sup> N-H nitrogen,
- The carboxyl C-O(H) oxygen and one of the Ala<sup>2nd</sup> beta hydrogens,
- The Ala<sup>1st</sup> N-H hydrogen and one of the Dap beta hydrogens.

In C2 of DapAAH<sup>+</sup>, there are the following eight distinct steric clashes in opened ring structures made of the following groups of atoms and are listed using bullet points:

- From the acetyl C=O oxygen to H<sub>40</sub> on the NH<sub>3</sub><sup>+</sup> group,
- From the Dap N-H nitrogen to the Dap C=O oxygen,
- From one of the Dap beta hydrogens to the Ala<sup>1st</sup> N-H hydrogen,
- From the Dap C=O oxygen to the Ala<sup>1st</sup> alpha hydrogen,
- From the Ala<sup>1st</sup> N-H nitrogen to the Ala<sup>1st</sup> C=O oxygen,
- From the Ala<sup>1st</sup> C=O oxygen to the Ala<sup>2nd</sup> alpha hydrogen,

- From the Ala<sup>2nd</sup> N-H hydrogen to the carboxyl C=O oxygen,
- From one of the Ala<sup>2nd</sup> beta hydrogens to the carboxyl C-O(H) oxygen.

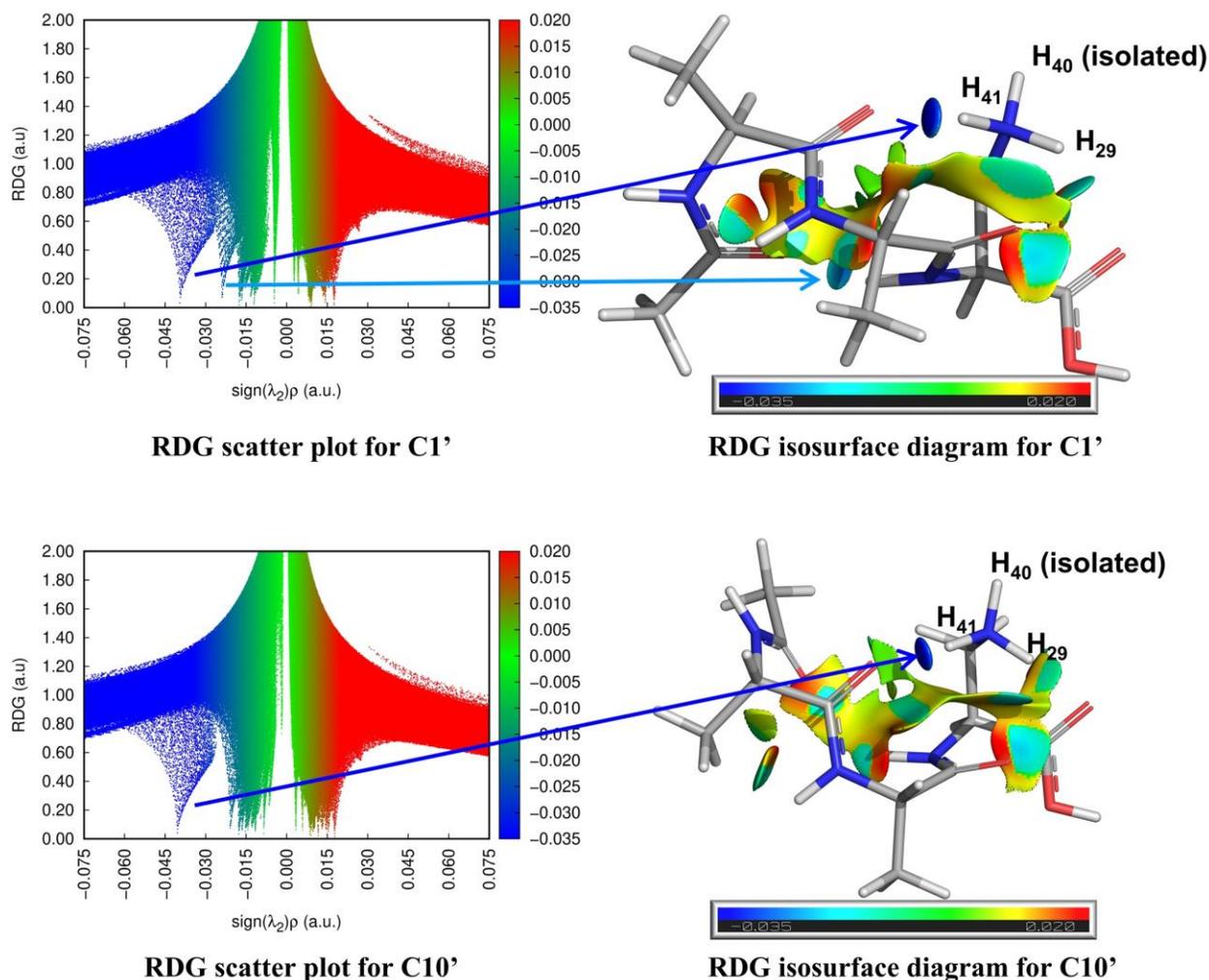
In C9 of DapAAH<sup>+</sup>, there are eleven distinct attractive interactions between the following atom pairs and are listed using bullet points:

- H<sub>40</sub> on the NH<sub>3</sub><sup>+</sup> group and the acetyl C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP}42} = 0.61\text{E-}01\text{a.u.}$ , HBE = 21.4 kcal/mol),
- H<sub>41</sub> on the NH<sub>3</sub><sup>+</sup> group and the Ala<sup>1st</sup> C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP}47} = 0.17\text{E-}01\text{a.u.}$ , HBE = 6.7 kcal/mol),
- H<sub>41</sub> on the NH<sub>3</sub><sup>+</sup> group and the carboxyl C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP}43} = 0.21\text{E-}01\text{a.u.}$ , HBE = 8.1 kcal/mol),
- The acetyl C=O oxygen and the Dap C=O oxygen,
- The Dap C=O oxygen and H<sub>41</sub> on the NH<sub>3</sub><sup>+</sup> group,
- The Dap C=O oxygen and the Ala<sup>1st</sup> alpha hydrogen,
- The Ala<sup>1st</sup> C=O oxygen and the Ala<sup>1st</sup> N-H nitrogen,
- The Ala<sup>1st</sup> C=O oxygen and one of the Dap beta hydrogens ( $\rho_{\text{BCP}50} = 0.75\text{E-}02\text{a.u.}$ ),
- The Ala<sup>1st</sup> C=O oxygen and the carboxyl C=O oxygen ( $\rho_{\text{BCP}46} = 0.11\text{E-}01\text{a.u.}$ ),
- The Ala<sup>1st</sup> C=O oxygen and one of the Ala<sup>2nd</sup> beta hydrogens ( $\rho_{\text{BCP}52} = 0.11\text{E-}01\text{a.u.}$ ),
- The carboxyl C-O(H) oxygen and one of the Ala<sup>2nd</sup> beta hydrogens.

In C9 of DapAAH<sup>+</sup>, there are eight distinct steric clashes in opened ring structures made of the following groups of atoms and are listed using bullet points:

- From the acetyl C=O oxygen to H<sub>40</sub> on the NH<sub>3</sub><sup>+</sup> group,
- From the Dap N-H nitrogen to the Dap C=O oxygen,
- From the Dap C=O oxygen to H<sub>41</sub> on the NH<sub>3</sub><sup>+</sup> group,
- From the Dap C=O oxygen to the Ala<sup>1st</sup> alpha hydrogen,
- From the Ala<sup>1st</sup> N-H nitrogen to the Ala<sup>1st</sup> C=O oxygen,
- From the Ala<sup>1st</sup> C=O oxygen to the carboxyl C=O oxygen,

- From the Ala<sup>1st</sup> C=O oxygen to one of the Ala<sup>2nd</sup> beta hydrogens,
- From one of the Ala<sup>2nd</sup> beta hydrogens to the carboxyl C-O(H) oxygen.



*Figure 3.48.* RDG scatter plot and isosurface of C1' (top) and C10' (bottom) of AADapH<sup>+</sup>. (The arrows show examples of correlation between the spikes in the RDG scatter plot and the features in the RDG isosurface diagram.)

In C1' of AADapH<sup>+</sup>, there are fourteen distinct attractive interactions between the following atom pairs and are listed using bullet points:

- H<sub>41</sub> on the NH<sub>3</sub><sup>+</sup> group and the Ala<sup>1st</sup> C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP45}} = 0.40\text{E-}01\text{a.u.}$ , HBE = 14.2 kcal/mol),

- H<sub>29</sub> on the NH<sub>3</sub><sup>+</sup> group and the Ala<sup>2nd</sup> C=O oxygen (weak hydrogen bond,  $\rho_{\text{BCP42}} = 0.12\text{E-}01$  a.u., HBE = 5.1 kcal/mol),
- H<sub>29</sub> on the NH<sub>3</sub><sup>+</sup> group and the carboxyl C=O oxygen (weak hydrogen bond,  $\rho_{\text{BCP46}} = 0.17\text{E-}01$  a.u., HBE = 6.6 kcal/mol),
- The acetyl C=O oxygen and the Dap N-H hydrogen (weak hydrogen bond,  $\rho_{\text{BCP49}} = 0.24\text{E-}01$  a.u., HBE = 6.1 kcal/mol),
- The acetyl C=O oxygen and one of the Dap beta hydrogens ( $\rho_{\text{BCP49}} = 0.46\text{E-}02$  a.u.),
- The acetyl C=O oxygen and the Ala<sup>1st</sup> C=O carbon,
- The acetyl C=O oxygen and the Ala<sup>2nd</sup> N-H nitrogen,
- The Ala<sup>1st</sup> C=O oxygen and the Ala<sup>2nd</sup> C=O carbon,
- The Ala<sup>1st</sup> C=O oxygen and the Dap N-H nitrogen,
- The Ala<sup>2nd</sup> C=O oxygen and H<sub>41</sub> on the NH<sub>3</sub><sup>+</sup> group,
- The Ala<sup>2nd</sup> C=O oxygen and the carboxyl C=O carbon,
- The carboxyl C=O oxygen and one of the Dap beta hydrogens,
- The Ala<sup>1st</sup> N-H nitrogen and the Ala<sup>2nd</sup> N-H hydrogen,
- The Ala<sup>2nd</sup> N-H nitrogen and the Dap N-H hydrogen.

In C1' of AADapH<sup>+</sup>, there are the following seven distinct steric clashes in opened ring structures made of the following groups of atoms and are listed using bullet points:

- From the acetyl C=O oxygen to the Ala<sup>1st</sup> C=O carbon,
- From the Ala<sup>1st</sup> N-H nitrogen to the Ala<sup>2nd</sup> N-H hydrogen,
- From the Ala<sup>1st</sup> C=O oxygen to the Ala<sup>2nd</sup> C=O carbon,
- From the Ala<sup>2nd</sup> N-H nitrogen to the Dap N-H hydrogen,
- From the Ala<sup>2nd</sup> C=O oxygen to the carboxyl C=O carbon,
- From the carboxyl C=O oxygen to one of the Dap beta hydrogens,
- From the Dap C=O oxygen to H<sub>29</sub> on the NH<sub>3</sub><sup>+</sup> group.

In C10' of AADapH<sup>+</sup>, there are thirteen distinct attractive interactions between the following atom pairs and are listed using bullet points:

- H<sub>41</sub> on the NH<sub>3</sub><sup>+</sup> group and the Ala<sup>1st</sup> C=O oxygen (strong hydrogen bond,  $\rho_{\text{BCP46}} = 0.41\text{E-}01\text{a.u.}$ , HBE = 14.5 kcal/mol),
- H<sub>29</sub> on the NH<sub>3</sub><sup>+</sup> group and the Ala<sup>2nd</sup> C=O oxygen (weak hydrogen bond,  $\rho_{\text{BCP47}} = 0.17\text{E-}01\text{ a.u.}$ , HBE = 6.6 kcal/mol),
- H<sub>29</sub> on the NH<sub>3</sub><sup>+</sup> group and the carboxyl C=O oxygen (weak hydrogen bond,  $\rho_{\text{BCP43}} = 0.12\text{E-}01\text{ a.u.}$ , HBE = 4.9 kcal/mol),
- The acetyl C=O oxygen and the Dap N-H hydrogen (weak hydrogen bond,  $\rho_{\text{BCP50}} = 0.21\text{E-}01\text{ a.u.}$ , HBE = 5.4 kcal/mol),
- The acetyl C=O oxygen and one of the Ala<sup>1st</sup> beta hydrogens,
- The acetyl C=O oxygen and the Ala<sup>1st</sup> C=O carbon,
- The acetyl C=O oxygen and one of the Ala<sup>1st</sup> beta hydrogens ( $\rho_{\text{BCP44}} = 0.21\text{E-}01\text{ a.u.}$ ),
- The Ala<sup>1st</sup> C=O oxygen and the Ala<sup>2nd</sup> C=O carbon,
- The Ala<sup>1st</sup> C=O oxygen and the Dap N-H nitrogen,
- The Ala<sup>2nd</sup> C=O oxygen and the carboxyl C=O carbon,
- The carboxyl C=O oxygen and one of the Dap beta hydrogens,
- The Ala<sup>2nd</sup> N-H nitrogen and the Dap N-H hydrogen,
- The Ala<sup>2nd</sup> N-H hydrogen and two of the Ala<sup>1st</sup> beta hydrogens.

In C10' of AADapH<sup>+</sup>, there are eight distinct steric clashes in opened ring structures made of the following groups of atoms:

- From the acetyl C=O oxygen to the Ala<sup>1st</sup> C=O carbon,
- From the acetyl C=O oxygen to one of the Ala<sup>1st</sup> beta hydrogens,
- From the Ala<sup>1st</sup> C=O oxygen to the Ala<sup>2nd</sup> C=O carbon,
- From the Ala<sup>2nd</sup> N-H nitrogen to the Dap N-H hydrogen,
- From the Ala<sup>2nd</sup> N-H hydrogen to two of the Ala<sup>1st</sup> beta hydrogens,
- From the Ala<sup>2nd</sup> C=O oxygen to the carboxyl C=O carbon,

- From the carboxyl C=O oxygen to one of the Dap beta hydrogens,
- From the carboxyl C=O oxygen to H<sub>29</sub> on the NH<sub>3</sub><sup>+</sup> group.

As listed above and shown in Figure 3.48 besides the hydrogen bonds, there are multiple attractive interactions colored in light blue in the AADapH<sup>+</sup> conformations. While in the DapAAH<sup>+</sup> conformations, there are not many large light blue regions, but rather a few small green regions. That means that the DapAAH<sup>+</sup> conformations are equipped with less and weaker attractive interactions than the AADapH<sup>+</sup> conformations. The dominance of steric clashes among all noncovalent interactions is also higher in the DapAAH<sup>+</sup> conformations. As a result, even though the AADapH<sup>+</sup> conformations have smaller total hydrogen bond energies, the other attractive interactions enhance their stabilities.

There are significant similarities among the DapAAH<sup>+</sup> conformations and among the AADapH<sup>+</sup> conformations. Detailed comparisons of their noncovalent interactions are attempted to understand the effects of their structural differences on their relative stabilities.

In C1 of DapAAH<sup>+</sup>, the Ala<sup>2nd</sup> N-H group and C=O group are placed parallel to each other. As a result, these two groups form one attractive interaction between the H --- O atom pair and one steric clash in the half-ring structure made of the Ala<sup>2nd</sup> residue. This type of geometry and interaction is also observed in the ALysH<sup>+</sup> conformations. Additionally, the Ala<sup>1st</sup> C=O group forms two interactions with the Ala<sup>2nd</sup> alpha and beta hydrogens. A very weak interaction between the Ala<sup>2nd</sup> C-O(H) oxygen and one of the Ala<sup>2nd</sup> beta hydrogens is also identified in the C1 RDG isosurface diagram.

Compared to C1, the carboxyl group in C2 is rotated to the back of the molecule. As a result, the interactions between the Ala<sup>2nd</sup> N-H group and the C=O group are broken. As substitutions, the Ala<sup>2nd</sup> C=O oxygen forms two interactions. One is with the Ala<sup>2nd</sup> N-H nitrogen, which is overall repulsive. As the corresponding RDG isosurface plate shows, the red repulsive region is much darker and larger than the green attractive region. The other one is with one of the Ala<sup>1st</sup> beta hydrogens. As shown, its RDG isosurface plate is dominated in green. This interaction is weak and overall attractive. Considering all the RDG isosurfaces, the dominance of attractive interactions in C2 is lower than C1, contributing to the computed lower stability of C2.

The carboxyl group in C9 is in the front of the molecule and forms a weak hydrogen bond with the  $\text{NH}_3^+$  group. As a result, the hydrogen bond between the  $\text{NH}_3^+$  group and Ala<sup>1st</sup> C=O oxygen becomes weakened. It has been reflected in its  $\text{HBE}_{\text{total}}$  of 36.1 kcal/mol, which is 3.7 kcal/mol lower than C1. The structural difference between C9 and C1 is much more than that between C2 and C1. So are the identified intramolecular interactions. Compared to C1, besides the hydrogen bonds, the green and light blue regions in C9 are larger, while the repulsive regions are not expanded. So overall, the structural feature of C9 grants it a weaker set of hydrogen bonds but a slightly stronger set of other attractive interactions, which counteracts the change in hydrogen bond strength and reduces the stability difference between C1 and C9.

Compared to C1' of AADapH<sup>+</sup>, the divergence of the acetyl group from the helix axis in C10' weakens the hydrogen bond between the acetyl C=O oxygen and the Dap N-H hydrogen. It has been reflected in the lower  $\text{HBE}_{\text{total}}$  of C10' by 0.5 kcal/mol. Additionally, the divergence also caused the loss of two attractive intramolecular interactions. One is between the acetyl C=O oxygen and the Ala<sup>2nd</sup> N-H nitrogen. The other is between the Ala<sup>1st</sup> N-H nitrogen and the Ala<sup>2nd</sup> N-H hydrogen. As substitutions, two interactions are formed from the Ala<sup>1st</sup> beta hydrogen to the acetyl C=O oxygen and the Ala<sup>2nd</sup> N-H hydrogen. The substitutions are weaker than the lost ones. Besides the hydrogen bonds, this difference also contributes to the stability difference between the two AADapH<sup>+</sup> conformations.

**Summary in energy characterization.** The energetic characterizations of the DapAAH<sup>+</sup> and AADapH<sup>+</sup> representative conformations are summarized in Table 3.36. Conformation C1' of AADapH<sup>+</sup> is selected as the reference of the relative free energy,  $\Delta G$ . For easy comparisons, it is also assigned with 0.0 kcal/mol for the Rel.  $E_{\text{torsion}}$  and Rel.  $\text{HBE}_{\text{total}}$ .

As the reference, C1' has the highest charge on the  $\text{NH}_3^+$  group and the largest  $\mu_{\text{backbone}}$  interaction with the  $\text{NH}_3^+$  group. Comparatively, C10' is higher in free energy by 2.8 kcal/mol. Its charge on the  $\text{NH}_3^+$  group and torsional strain on the Dap sidechain are nearly the same as C10'. Its  $\text{HBE}_{\text{total}}$  is less than C1' by 0.5 kcal/mol. So, its disadvantages in the  $\mu_{\text{backbone}}$  interaction with  $\text{NH}_3^+$  group of 24.62 D<sup>2</sup> and the

total strength of other noncovalent interactions identified through RDG analysis contribute about 2 kcal/mol to its lower stability than C1'.

C1 of DapAAH<sup>+</sup> has much stronger hydrogen bonds than C1' of AADapH<sup>+</sup>. As a result, its HBE<sub>total</sub> is much higher, and its Hirshfeld charge on the NH<sub>3</sub><sup>+</sup> group is much lower. These advantages grant C1 a relative free energy of around -7.8 kcal/mol. Its disadvantages in torsional strain on the Dap sidechain bring its relative free energy up by 0.7 kcal/mol to around -7.1 kcal/mol. The other two disadvantages in the  $\mu_{\text{backbone}}$  interaction with NH<sub>3</sub><sup>+</sup> group of 74.44 D<sup>2</sup> and the other noncovalent interactions identified through RDG analysis increases its relative free energy to +1.8 kcal/mol. So, the total contribution of these two disadvantages brings its relative free energy up by about 9 kcal/mol.

C2 is similar to C1 in many aspects. Its Hirshfeld charge on the NH<sub>3</sub><sup>+</sup> group and torsional strain on the Dap sidechain are nearly the same as those of C1. The effect of the Dap sidechain torsional strain and the HBE<sub>total</sub> grants C2 a relative free energy of -6.5 kcal/mol. Its disadvantages in the  $\mu_{\text{backbone}}$  interaction with NH<sub>3</sub><sup>+</sup> group of 65.77 D<sup>2</sup> and the other noncovalent interactions identified through RDG analysis increases its relative free energy to 3.0 kcal/mol. So, the total contribution of these two disadvantages brings its relative free energy up by 9.5 kcal/mol, at the same magnitude as C1.

Compared to C1' of AADapH<sup>+</sup>, C9 of DapAAH<sup>+</sup> has stronger hydrogen bonding networks and higher torsional strain on the Dap sidechain. Combination of both grants C9 a relative free energy of -3.7 kcal/mol. As shown in Table 3.36, it is higher in free energy than C1 by 5.3 kcal/mol. Three factors bring its relative free energy higher by 9 kcal/mol. Two of them are its disadvantages in the  $\mu_{\text{backbone}}$  interaction with NH<sub>3</sub><sup>+</sup> group of 60.54 D<sup>2</sup> and the noncovalent interactions identified through RDG analysis. The other is its advantage of lower Hirshfeld charge on the NH<sub>3</sub><sup>+</sup> group.

Among the DapAAH<sup>+</sup> conformations, C2 is higher in free energy than C1 by 1.2 kcal/mol. C1 and C2 have the same torsional strain on the Dap sidechain, and Hirshfeld charge on the NH<sub>3</sub><sup>+</sup> group. Compared to C1, C2 has a disadvantage of 0.6 kcal/mol in HBE<sub>total</sub>. Its disadvantage in the other noncovalent interactions identified through RDG analysis counteracts its advantage in the  $\mu_{\text{backbone}}$

interaction with the  $\text{NH}_3^+$  group of  $8.67 \text{ D}^2$  and contributes to the rest  $0.6 \text{ kcal/mol}$  of its lower stability than C1.

C9 is higher in free energy than C1 by  $3.5 \text{ kcal/mol}$ . Compared to C1, C9 has a huge disadvantage of  $3.7 \text{ kcal/mol}$  in  $\text{HBE}_{\text{total}}$ , and a marginal advantage of  $0.2 \text{ kcal/mol}$  in the torsional strain on the Dap sidechain. The outcome of these two factors matches the computed free energy difference between C9 and C1. So, its advantages in the  $\mu_{\text{backbone}}$  interaction with the  $\text{NH}_3^+$  group of  $13.9 \text{ D}^2$  and the noncovalent interactions identified through RDG analysis are offset by its higher Hirshfeld charge on the  $\text{NH}_3^+$  group.

In summary, the quantitative evaluation of the torsional strain on the Dap sidechain and the total hydrogen bond strength allows an estimate of the contribution of the  $\mu_{\text{backbone}}$  interaction with the  $\text{NH}_3^+$  group and the noncovalent interactions identified through RDG analysis. Compared to the DapAAH<sup>+</sup> conformations, the AADapH<sup>+</sup> conformations have a significantly lower total strength of the hydrogen bonds. But their advantages in the  $\mu_{\text{backbone}}$  interaction with the  $\text{NH}_3^+$  group and the noncovalent interactions identified through RDG analysis outperform these disadvantages and grant the AADapH<sup>+</sup> conformations higher stabilities.

Table 3.36  
*Energetic Characterization of the DapAAH<sup>+</sup> and AADapH<sup>+</sup> Representative Conformations*

	Conf.	$\Delta G$ (kcal/mol)	Hirshfeld Charge (a.u.)	Dot Product of $\mu_{\text{NH}_3}$ and $\mu_{\text{backbone}}$	Rel. $E_{\text{torsion}}$ (kcal/mol)	Rel. $\text{HBE}_{\text{total}}$ (kcal/mol)
DapAAH <sup>+</sup>	C1	1.8	0.398	-6.24	0.7	7.8
	C2	3.0	0.402	-14.91	0.7	7.2
	C9	5.3	0.418	-20.14	0.4	4.1
AADapH <sup>+</sup>	C1'	0.0	0.492	-80.68	0.0	0.0
	C10'	2.8	0.490	-56.06	0.1	-0.5

### IR Decomposition

The IRMPD spectra of DapAAH<sup>+</sup> and AADapH<sup>+</sup> are characteristically different and suggest their different structures. Conformations with a good match to their IRMPD spectra were found and analyzed from the energy and structure point of view. Like the analysis used for LysAH<sup>+</sup> and ALysH<sup>+</sup>, IR decomposition of those selected conformations is performed to explain their different spectral patterns. Detailed procedures are available in Chapter 2. The IRMPD band assignments, calculated wavenumbers, and mode descriptions are listed in Table 3.37, Table 3.38, and Table 3.39 for C1, C2, and C9 of DapAAH<sup>+</sup>, and in Table 3.40 and Table 3.41 for C1', and C10' of AADapH<sup>+</sup>. The structure of the selected conformations of DapAAH<sup>+</sup> (top) and AADapH<sup>+</sup> (bottom) are shown in Figure 3.49, with all atoms labeled, for easy correlation between the motion description and the mode assignment.

Additionally, the amide bonds and the O --- H distance in their corresponding hydrogen bonds are summarized in Table 3.42. The amide bond strength is affected by its position in the peptide and its surrounding microenvironment. Their strength is attempted to be correlated with their corresponding computed wavenumber. Multiple intrinsic measures of the bond strength, including the bond length, bond order, relaxed force constant, are used to justify the computed wavenumbers.

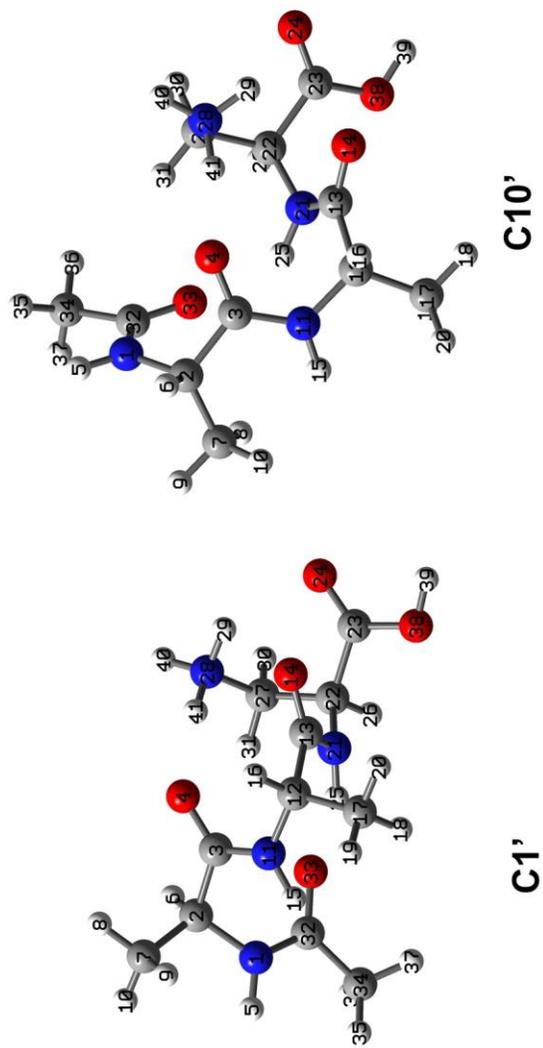
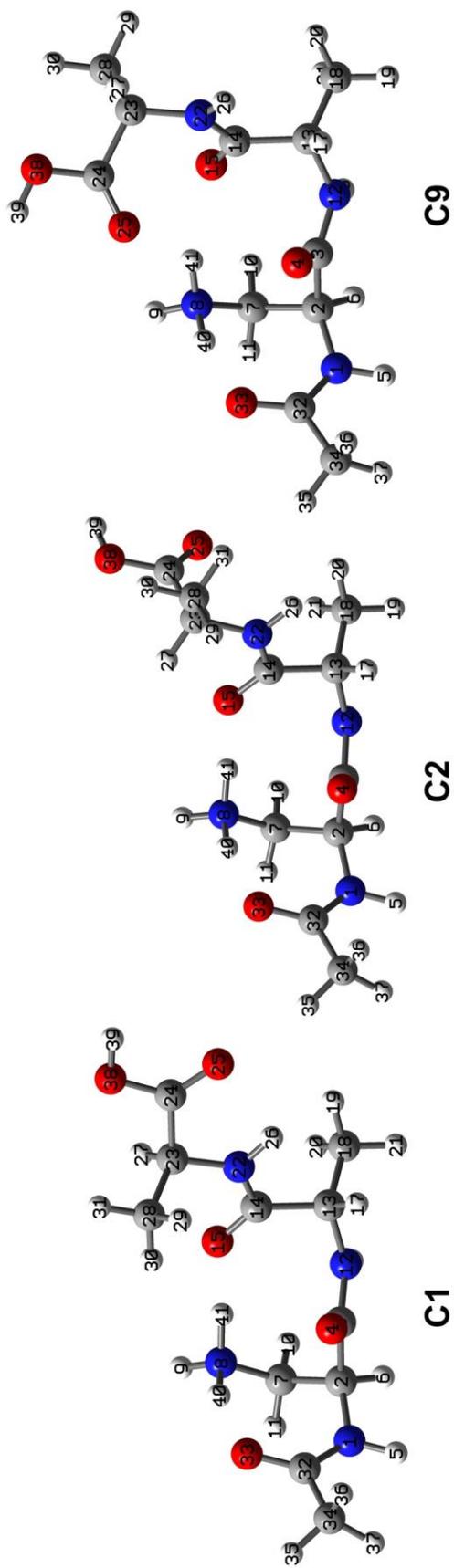


Figure 3.49. Selected conformations of DapAAH<sup>+</sup> (top) and AADapH<sup>+</sup> (bottom) with atoms labeled.

Table 3.37

*IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C1 of DapAAH<sup>+</sup>*

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1769	1767	BOND C24 O25	68.6	<b>Carboxyl C=O stretching</b>
		BOND C24 O38	5.7	Carboxyl C=O stretching
1726	1734	BOND C3 O4	55.2	<b>Dap-Ala<sup>1st</sup> Amide I</b>
		BOND C32 O33	9.3	Ac-Dap Amide I
		BOND C3 N12	7.4	Dap-Ala <sup>1st</sup> Amide I
1682	1680	BOND C32 O33	43.6	<b>Ac-Dap Amide I</b>
		BOND C3 O4	10.0	Dap-Ala <sup>1st</sup> Amide I
		BOND C32 C34	5.8	Ac-Dap Amide I
1648	1647	BOND C14 O15	40.6	<b>Ala<sup>1st</sup>-Ala<sup>2nd</sup> Amide I</b>
		BOND C14 N22	14.4	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
N/A	1610	ANGLE H40 N8 H41	11.9	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		ANGLE H9 N8 H40	8.9	NH <sub>3</sub> <sup>+</sup> scissoring
		ANGLE H9 N8 H41	7.4	NH <sub>3</sub> <sup>+</sup> scissoring
		TORSION O15 C14 N22 H26	6.0	
N/A	1589	ANGLE C7 N8 H41	10.8	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		ANGLE H9 N8 H40	9.9	NH <sub>3</sub> <sup>+</sup> scissoring
		ANGLE H9 N8 H41	8.6	NH <sub>3</sub> <sup>+</sup> scissoring
		BOND C32 O33	7.4	Ac-Dap Amide I
		BOND C14 O15	6.2	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.37 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1522	1528	BOND C14 N22	23.3	<b>Ala<sup>1st</sup>-Ala<sup>2nd</sup> Amide II</b>
		BOND C14 O15	11.2	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
		ANGLE C23 N22 H26	8.8	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
		BOND N22 C23	7.0	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
		BOND C13 C14	6.7	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
		ANGLE C14 N22 H26	6.2	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
1495	1500	BOND N1 C32	17.3	<b>Ac-Dap Amide II</b>
		ANGLE C2 N1 H5	8.8	Ac-Dap Amide II
		ANGLE H5 N1 C32	7.4	Ac-Dap Amide II
1480	1476	BOND C3 N12	19.1	<b>Dap-Ala<sup>1st</sup> Amide II</b>
		ANGLE C3 N12 H16	14.0	Dap-Ala <sup>1st</sup> Amide II
		BOND N12 C13	7.1	Dap-Ala <sup>1st</sup> Amide II
		ANGLE O4 C3 N12	6.6	Dap-Ala <sup>1st</sup> Amide II
1455	1465	ANGLE C7 N8 H40	9.5	<b>NH<sub>3</sub><sup>+</sup> umbrella</b>
		BOND C32 O33	7.7	Ac-Dap Amide II
		ANGLE C7 N8 H41	7.4	NH <sub>3</sub> <sup>+</sup> umbrella
		BOND N1 C32	5.8	Ac-Dap Amide II
1375	1376	BOND C24 O38	15.7	<b>Carboxyl C-O-H bending</b>
		BOND C14 N22	7.8	
		ANGLE O25 C24 O38	7.1	Carboxyl C-O-H bending
		BOND C23 C24	6.9	Carboxyl C-O-H bending
		ANGLE H29 C28 H31	6.1	Ala <sup>2nd</sup> CH <sub>3</sub> umbrella
		ANGLE C24 O38 H39	5.7	Carboxyl C-O-H bending

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

Table 3.38

*IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C2 of DapAAH<sup>+</sup>*

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1769	1774	BOND C24 O25	69.7	<b>Carboxyl C=O stretching</b>
		BOND C23 C24	5.7	Carboxyl C=O stretching
1726	1733	BOND C3 O4	53.9	<b>Dap-Ala<sup>1st</sup> Amide I</b>
		BOND C32 O33	9.1	Ac-Dap Amide I
		BOND C3 N12	6.5	Dap-Ala <sup>1st</sup> Amide I
1682	1679	BOND C32 O33	41.3	<b>Ac-Dap Amide I</b>
		BOND C3 O4	9.9	Dap-Ala <sup>1st</sup> Amide I
		BOND C14 O15	6.2	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
1648	1657	BOND C14 O15	42.7	<b>Ala<sup>1st</sup>-Ala<sup>2nd</sup> Amide I</b>
		BOND C14 N22	11.0	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
N/A	1611	ANGLE H40 N8 H41	12.4	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		ANGLE H9 N8 H41	12.3	NH <sub>3</sub> <sup>+</sup> scissoring
		BOND C32 O33	6.5	Ac-Dap Amide I
		ANGLE H9 N8 H40	6.4	NH <sub>3</sub> <sup>+</sup> scissoring
N/A	1595	ANGLE H9 N8 H40	9.8	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		ANGLE H9 N8 H41	9.4	NH <sub>3</sub> <sup>+</sup> scissoring
		ANGLE C7 N8 H41	9.4	NH <sub>3</sub> <sup>+</sup> scissoring
		ANGLE H40 N8 H41	6.0	NH <sub>3</sub> <sup>+</sup> scissoring

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.38 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1522	1519	BOND C14 N22	18.5	<b>Ala<sup>1st</sup>-Ala<sup>2nd</sup> Amide II</b>
		ANGLE C14 N22 H26	7.5	Ala <sup>1st</sup> - Ala <sup>2nd</sup> Amide II
		ANGLE C23 N22 H26	7.2	Ala <sup>1st</sup> - Ala <sup>2nd</sup> Amide II
		ANGLE C14 N22 C23	6.0	Ala <sup>1st</sup> - Ala <sup>2nd</sup> Amide II
		TORSION N22 C23 C24 O25	5.7	
		BOND N22 C23	5.7	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
1495	1500	BOND N1 C32	15.6	<b>Ac-Dap Amide II</b>
		ANGLE C2 N1 H5	8.9	Ac-Dap Amide II
		ANGLE H5 N1 C32	6.2	Ac-Dap Amide II
		BOND C32 C34	6.2	Ac-Dap Amide II
		BOND N1 C2	5.9	Ac-Dap Amide II
1480	1477	BOND C3 N12	19.5	<b>Dap-Ala<sup>1st</sup> Amide II</b>
		ANGLE C3 N12 H16	12.2	Dap-Ala <sup>1st</sup> Amide II
		BOND N12 C13	8.5	Dap-Ala <sup>1st</sup> Amide II
		ANGLE C13 N12 H16	6.5	Dap-Ala <sup>1st</sup> Amide II
1455	1460	ANGLE C7 N8 H40	11.3	<b>NH<sub>3</sub><sup>+</sup> umbrella</b>
		ANGLE C7 N8 H41	8.9	NH <sub>3</sub> <sup>+</sup> umbrella
		BOND C32 O33	6.5	Ac-Dap Amide II
1375	1366	BOND C24 O38	9.3	<b>Carboxyl C-O-H bending</b>
		ANGLE C13 C18 H21	6.7	Ala <sup>1st</sup> CH <sub>3</sub> umbrella
		ANGLE C13 C18 H20	6.6	Ala <sup>1st</sup> CH <sub>3</sub> umbrella
	ANGLE C13 C18 H19	6.5	Ala <sup>1st</sup> CH <sub>3</sub> umbrella	
1363	1363	BOND C24 O38	10.4	<b>Carboxyl C-O-H bending</b>
		ANGLE H19 C18 H20	9.3	Ala <sup>1st</sup> CH <sub>3</sub> umbrella
		ANGLE H20 C18 H21	9.0	Ala <sup>1st</sup> CH <sub>3</sub> umbrella
		ANGLE H19 C18 H21	7.7	Ala <sup>1st</sup> CH <sub>3</sub> umbrella

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.38 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1265	1251	BOND N1 C32	16.8	<b>Ac-Dap Amide III</b>
		BOND C32 O33	9.9	Ac-Dap Amide III
		BOND N1 C2	7.1	Ac-Dap Amide III
		BOND C32 C34	7.0	Ac-Dap Amide III
		ANGLE C2 N1 H5	6.6	Ac-Dap Amide III
N/A	1223	BOND C14 N22	16.2	<b>Ala<sup>1st</sup>-Ala<sup>2nd</sup> Amide III</b>
		ANGLE C14 N22 H26	8.7	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide III
		BOND C14 O15	7.6	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide III
		BOND N22 C23	6.1	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide III
1204	1193	BOND C3 N12	11.0	<b>Dap-Ala<sup>1st</sup> Amide III</b>
		BOND N12 C13	10.9	Dap-Ala <sup>1st</sup> Amide III
		ANGLE C13 N12 H16	5.9	Dap-Ala <sup>1st</sup> Amide III
1143	1157	BOND N22 C23	7.1	<b>Ala<sup>2nd</sup> N(H)-C<sub>α</sub> stretching</b>
		BOND C24 O38	6.5	Carboxyl C-O stretching
	1138	BOND C24 O25	5.9	Carboxyl C-O stretching
		BOND C3 N12	8.9	<b>Dap-Ala<sup>1st</sup> Amide III</b>
		BOND C24 O38	6.5	Carboxyl C-O stretching
		BOND C3 O4	6.4	Dap-Ala <sup>1st</sup> Amide III
1113	1120	BOND C24 O38	13.5	<b>Carboxyl C-O stretching</b>
		BOND C24 O25	8.7	Carboxyl C-O stretching
		BOND N22 C23	7.5	Ala <sup>2nd</sup> N(H)-C <sub>α</sub> stretching
	1102	BOND C14 O15	6.7	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide III
		BOND C7 N8	8.0	<b>Dap C<sub>β</sub>-N(H3) Stretching</b>
1102	1102	ANGLE C7 N8 H9	7.3	NH <sub>3</sub> <sup>+</sup> wobble
		ANGLE C7 N8 H40	6.9	NH <sub>3</sub> <sup>+</sup> wobble
		BOND N1 C2	6.2	Dap N(H)-C <sub>α</sub> stretching

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the ωB97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

Table 3.39  
*IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C9 of DapAAH<sup>+</sup>*

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1769	1757	BOND C24 O25	50.4	<b>Carboxyl C=O stretching</b>
		BOND C14 O15	9.5	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
		BOND C3 O4	7.8	Dap-Ala <sup>1st</sup> Amide I
1726	1726	BOND C3 O4	46.4	<b>Dap-Ala<sup>1st</sup> Amide I</b>
		BOND C32 O33	10.7	Ac-Dap Amide I
		BOND C24 O25	8.2	Carboxyl C=O stretching
1682	1683	BOND C32 O33	31.3	<b>Ac-Dap Amide I</b>
		BOND C14 O15	14.9	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
		BOND C3 O4	8.3	Dap-Ala <sup>1st</sup> Amide I
1682	1675	BOND C14 O15	34.4	<b>Ala<sup>1st</sup>-Ala<sup>2nd</sup> Amide I</b>
		BOND C32 O33	15.5	Ac-Dap Amide I
		BOND C14 N22	8.2	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
		BOND C24 O25	6.5	Carboxyl C=O stretching
N/A	1589	ANGLE H9 N8 H40	17.5	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		ANGLE C7 N8 H41	9.5	NH <sub>3</sub> <sup>+</sup> scissoring
		BOND C14 O15	7.4	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
		BOND C32 O33	7.4	Ac-Dap Amide I
N/A	1575	ANGLE H9 N8 H41	13.3	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		ANGLE H40 N8 H41	10.6	NH <sub>3</sub> <sup>+</sup> scissoring
		BOND α3 O4	8.4	Dap-Ala <sup>1st</sup> Amide I
		ANGLE C7 N8 H9	6.9	NH <sub>3</sub> <sup>+</sup> scissoring

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the ωB97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.39 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1522	1509	BOND C14 N22	20.9	<b>Ala<sup>1st</sup>-Ala<sup>2nd</sup> Amide II</b>
		ANGLE C23 N22 H26	10.0	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
		BOND N22 C23	7.0	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
		ANGLE C14 N22 H26	7.0	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
		BOND C13 C14	6.4	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
		ANGLE O15 C14 N22	6.1	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
1495	1505	BOND N1 C32	17.1	<b>Ac-Dap Amide II</b>
		ANGLE C2 N1 H5	10.5	Ac-Dap Amide II
		ANGLE H5 N1 C32	8.1	Ac-Dap Amide II
		BOND N1 C2	6.9	Ac-Dap Amide II
		BOND C32 C34	6.4	Ac-Dap Amide II
1480	1478	BOND C3 N12	16.6	<b>Dap-Ala<sup>1st</sup> Amide II</b>
		ANGLE C13 N12 H16	9.7	Dap-Ala <sup>1st</sup> Amide II
		ANGLE C3 N12 H16	8.5	Dap-Ala <sup>1st</sup> Amide II
1455	1456	ANGLE H10 C7 H11	8.5	<b>Dap CH<sub>2</sub> Scissoring</b>
		ANGLE H40 N8 H41	7.6	<b>NH<sub>3</sub><sup>+</sup> umbrella</b>
		ANGLE C7 N8 H40	7.1	NH <sub>3</sub> <sup>+</sup> umbrella
		BOND C32 O33	5.9	Ac-Dap Amide II
1430	1433	ANGLE H10 C7 H11	11.5	<b>Dap CH<sub>2</sub> Scissoring</b>
		ANGLE C7 N8 H40	6.8	<b>NH<sub>3</sub><sup>+</sup> umbrella</b>
		BOND C32 O33	6.0	Ac-Dap Amide II
1375	1372	BOND C24 O38	14.3	<b>Carboxyl C-O-H bending</b>
		BOND C23 C24	6.9	Ala <sup>2nd</sup> C <sub>α</sub> -C(O) stretching
		ANGLE N22 C23 H27	6.7	Ala <sup>2nd</sup> C <sub>α</sub> -H bending

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.39 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1265	1253	BOND N1 C32	14.9	<b>Ac-Dap Amide III</b>
		BOND C32 O33	8.4	Ac-Dap Amide III
		BOND N1 C2	7.2	Ac-Dap Amide III
		ANGLE C2 N1 H5	6.7	Ac-Dap Amide III
1204	1214	BOND C14 N22	17.6	<b>Ala<sup>1st</sup>-Ala<sup>2nd</sup> Amide III</b>
		BOND N22 C23	10.1	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide III
		BOND C14 O15	9.2	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide III
		ANGLE C23 N22 H26	6.4	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide III
1185	1185	BOND C3 N12	12.7	<b>Dap-Ala<sup>1st</sup> Amide III</b>
		BOND N12 C13	6.8	Dap-Ala <sup>1st</sup> Amide III
1143	1147	BOND C24 O38	13.1	<b>Carboxyl C-O stretching</b>
		BOND C24 O25	9.8	Carboxyl C-O stretching
		ANGLE C24 O38 H39	6.7	Carboxyl C-O stretching
1113	1098	BOND C24 O38	9.1	<b>Carboxyl C-O stretching</b>
		BOND N22 C23	8.4	Ala <sup>2nd</sup> N(H)-C <sub>α</sub> stretching
		BOND C24 O25	6.1	Carboxyl C-O stretching
		BOND C7 N8	10.6	<b>Dap C<sub>β</sub>-N(H3) Stretching</b>
1096	1096	BOND N1 C2	9.5	Dap N(H)-C <sub>α</sub> stretching
		ANGLE C7 N8 H9	6.6	NH <sub>3</sub> <sup>+</sup> wobble
		ANGLE C7 N8 H40	5.8	NH <sub>3</sub> <sup>+</sup> wobble

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

Table 3.40  
*IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C1' of AADapH<sup>+</sup>*

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1757	1755	BOND C23 O24	58.1	<b>Carboxyl C=O stretching</b>
		BOND C23 O38	6.8	Carboxyl C=O stretching
		BOND C13 O14	6.7	Ala <sup>2nd</sup> -Dap Amide I
	1702	BOND C13 O14	51.4	<b>Ala<sup>2nd</sup>-Dap Amide I</b>
		BOND C23 O24	5.9	Carboxyl C=O stretching
1678	1671	BOND C32 O33	33.7	<b>Ac-Ala<sup>1st</sup> Amide I</b>
		BOND C3 O4	23.2	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
	1658	1658	BOND C3 O4	31.8
BOND C32 O33			23.0	Ac-Ala <sup>1st</sup> Amide I
BOND C3 N11			6.7	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
N/A	1601	ANGLE H40 N28 H41	23.8	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		ANGLE H29 N28 H40	11.9	NH <sub>3</sub> <sup>+</sup> scissoring
		BOND C23 O24	6.9	Carboxyl C=O stretching
		ANGLE C27 N28 H41	6.7	NH <sub>3</sub> <sup>+</sup> scissoring
		ANGLE C27 N28 H29	6.7	NH <sub>3</sub> <sup>+</sup> scissoring
		BOND C3 O4	6.5	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
N/A	1566	ANGLE H29 N28 H40	15.4	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		ANGLE H29 N28 H41	14.9	NH <sub>3</sub> <sup>+</sup> scissoring
		BOND C13 O14	9.6	Ala <sup>2nd</sup> -Dap Amide I
		ANGLE H40 N28 H41	6.7	NH <sub>3</sub> <sup>+</sup> scissoring
		BOND C3 O4	5.9	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
		ANGLE C27 N28 H40	5.7	NH <sub>3</sub> <sup>+</sup> scissoring

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.40 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
		BOND C13 N21	8.5	<b>Ala<sup>2nd</sup>-Dap Amide II</b>
	1518	BOND C3 N11	7.9	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
		ANGLE C13 N21 H25	7.5	Ala <sup>2nd</sup> -Dap Amide II
		ANGLE C22 N21 H25	6.5	Ala <sup>2nd</sup> -Dap Amide II
		BOND C3 N11	16.9	<b>Ala<sup>1st</sup>-Ala<sup>2nd</sup> Amide II</b>
	1497	ANGLE C3 N11 H15	8.7	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
		BOND C13 N21	5.8	Ala <sup>2nd</sup> -Dap Amide II
1494		BOND N1 C32	17.6	<b>Ac-Ala<sup>1st</sup> Amide II</b>
		ANGLE H5 N1 C32	7.5	Ac-Ala <sup>1st</sup> Amide II
	1471	ANGLE C2 N1 H5	7.2	Ac-Ala <sup>1st</sup> Amide II
		BOND C32 C34	6.2	Ac-Ala <sup>1st</sup> Amide II
		BOND N1 C2	5.7	Ac-Ala <sup>1st</sup> Amide II
		ANGLE C27 N28 H41	15.5	<b>NH<sub>3</sub><sup>+</sup> umbrella</b>
	1465	ANGLE C27 N28 H29	12.2	NH <sub>3</sub> <sup>+</sup> umbrella
		ANGLE C27 N28 H40	7.7	NH <sub>3</sub> <sup>+</sup> umbrella
		BOND C23 O38	22.3	<b>Carboxyl C-O-H bending</b>
		BOND C22 C23	10.8	Carboxyl C-O-H bending
1383	1378	ANGLE O24 C23 O38	8.8	Carboxyl C-O-H bending
		ANGLE C23 O38 H39	8.4	Carboxyl C-O-H bending
		BOND C13 N21	6.5	Ala <sup>2nd</sup> -Dap Amide III

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.40 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1757	1755	BOND C23 O24	58.1	<b>Carboxyl C=O stretching</b>
	1266	BOND C13 N21	12.8	<b>Ala<sup>2nd</sup>-Dap Amide III</b>
		BOND C12 C13	6.1	Ala <sup>2nd</sup> -Dap Amide III
		BOND C3 N11	9.4	<b>Ala<sup>1st</sup>-Ala<sup>2nd</sup> Amide III</b>
	1235	BOND C3 O4	6.9	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide III
1249		BOND C13 N21	6.8	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide III
		BOND N1 C32	5.8	Ac-Ala <sup>1st</sup> Amide III
		BOND N1 C32	12.9	<b>Ac-Ala<sup>1st</sup> Amide III</b>
	1215	BOND C32 O33	8.0	Ac-Ala <sup>1st</sup> Amide III
		BOND C3 N11	6.5	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide III
		BOND C23 O38	19.1	<b>Carboxyl C-O stretching</b>
1155	1164	BOND C23 O24	18.2	Carboxyl C-O stretching
		ANGLE C23 O38 H39	10.1	Carboxyl C-O stretching
		BOND N21 C22	7.1	Dap N(H)-C <sub>α</sub> stretching
1097	1082	BOND N21 C22	9.8	<b>Dap N(H)-C<sub>α</sub> stretching</b>
		BOND C27 N28	6.1	Dap C <sub>β</sub> -N(H3) Stretching

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

Table 3.41

*IRMPD Band Assignments and Mode Descriptions with Frequencies Calculated for C10' of AADapH<sup>+</sup>*

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1757	1762	BOND C23 O24	65.40%	<b>Carboxyl C=O stretching</b>
		BOND C23 O38	7.50%	Carboxyl C=O stretching
1696	1696	BOND C13 O14	46.80%	<b>Ala<sup>2nd</sup>-Dap Amide I</b>
		BOND C13 N21	6.40%	Ala <sup>2nd</sup> -Dap Amide I
		BOND C32 O33	6.10%	Ac-Ala <sup>1st</sup> Amide I
		BOND C32 O33	39.60%	<b>Ac-Ala<sup>1st</sup> Amide I</b>
1678	1667	BOND C3 O4	17.20%	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
		BOND C3 O4	35.90%	<b>Ala<sup>1st</sup>-Ala<sup>2nd</sup> Amide I</b>
	1652	BOND C32 O33	16.70%	Ac-Ala <sup>1st</sup> Amide I
		BOND C3 N11	8.20%	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
N/A	1631	ANGLE H40 N28 H41	19.20%	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		BOND C3 O4	9.10%	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I
		BOND C13 O14	8.00%	Ala <sup>2nd</sup> -Dap Amide I
N/A	1556	BOND C23 O24	6.40%	Carboxyl C=O stretching
		ANGLE H29 N28 H40	18.30%	<b>NH<sub>3</sub><sup>+</sup> scissoring</b>
		BOND C13 O14	11.50%	Ala <sup>2nd</sup> -Dap Amide I
		ANGLE C27 N28 H41	7.30%	NH <sub>3</sub> <sup>+</sup> scissoring
		BOND C3 O4	5.70%	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide I

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.41 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
1513		BOND C13 N21	10.60%	<b>Ala<sup>2nd</sup>-Dap Amide II</b>
		ANGLE C13 N21 H25	7.70%	Ala <sup>2nd</sup> -Dap Amide II
1500		BOND N1 C32	14.80%	<b>Ac-Ala<sup>1st</sup> Amide II</b>
		ANGLE C2 N1 H5	7.80%	Ac-Ala <sup>1st</sup> Amide II
		BOND C3 N11	6.60%	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
		BOND N1 C2	6.00%	Ac-Ala <sup>1st</sup> Amide II
1494	1489	BOND C3 N11	15.30%	<b>Ala<sup>1st</sup>-Ala<sup>2nd</sup> Amide II</b>
		ANGLE C3 N11 H15	7.70%	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide II
1478		ANGLE H29 N28 H41	12.50%	<b>NH<sub>3</sub><sup>+</sup> umbrella</b>
		ANGLE H40 N28 H41	9.50%	NH <sub>3</sub> <sup>+</sup> umbrella
		ANGLE H29 N28 H40	8.90%	NH <sub>3</sub> <sup>+</sup> umbrella
		ANGLE C27 N28 H41	7.70%	NH <sub>3</sub> <sup>+</sup> umbrella
1383	1374	BOND C23 O38	21.40%	<b>Carboxyl C-O-H bending</b>
		BOND C22 C23	9.40%	Carboxyl C-O-H bending
		ANGLE O24 C23 O38	9.30%	Carboxyl C-O-H bending
		ANGLE C23 O38 H39	8.20%	Carboxyl C-O-H bending
		BOND C13 N21	7.70%	Ala <sup>2nd</sup> -Dap Amide II
		ANGLE C23 C22 H26	5.90%	Carboxyl C-O-H bending

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

(Table 3.41 Continued)

Obs. <sup>a</sup>	Calc. <sup>b</sup>	Motion Description <sup>c</sup>	Percent <sup>c</sup>	Mode Assignment <sup>d</sup>
		BOND C13 N21	12.50%	<b>Ala<sup>2nd</sup>-Dap Amide III</b>
	1264	BOND C12 C13	6.90%	Ala <sup>2nd</sup> -Dap Amide III
		ANGLE C22 N21 H25	6.20%	Ala <sup>2nd</sup> -Dap Amide III
1249	1248	BOND N1 C32	10.90%	<b>Ac-Ala<sup>1st</sup> Amide III</b>
		BOND C32 O33	6.60%	Ac-Ala <sup>1st</sup> Amide III
	1243	BOND N1 C32	7.60%	<b>Ac-Ala<sup>1st</sup> Amide III</b>
	1219	BOND C3 N11	13.20%	<b>Ala<sup>1st</sup>-Ala<sup>2nd</sup> Amide III</b>
		BOND C3 O4	8.60%	Ala <sup>1st</sup> -Ala <sup>2nd</sup> Amide III
1155	1161	BOND C23 O38	20.90%	<b>Carboxyl C-O stretching</b>
		BOND C23 O24	19.60%	Carboxyl C-O stretching
		ANGLE C23 O38 H39	10.40%	Carboxyl C-O stretching
		BOND N21 C22	6.20%	Dap N(H)-C $\alpha$ stretching
1097	1084	BOND N21 C22	8.70%	<b>Dap N(H)-C<math>\alpha</math> stretching</b>
		ANGLE C27 N28 H40	7.00%	NH <sub>3</sub> <sup>+</sup> wobble
		BOND C27 N28	6.20%	Dap C $\beta$ -N(H3) Stretching
		BOND C23 O38	6.00%	Carboxyl C-O stretching

<sup>a</sup> Peak position taken from the experimental IRMPD spectrum (Figure 3.25).

<sup>b</sup> Calculated IR frequencies at the  $\omega$ B97X-D/6-311G\*\* level and scaled by 0.947.

<sup>c</sup> Decomposed motions and their percent attributions for the computed vibrational modes, obtained from VibAnalysis. The notation BOND represents stretching motion, ANGLE represents in-plane bending motion, and TORSION represents out-of-plane bending motion.

<sup>d</sup> Each internal coordinate is assigned to a normal mode of the corresponding functional group. Primary mode of each computed vibrational mode is indicated in bold.

Table 3.42  
 Summary of Amide Bond, Carboxyl Group, and Corresponding Hydrogen Bond Information in the DapAAH<sup>+</sup> and AADapH<sup>+</sup> Conformations

Conf.	C(=O)-N(H) Bond			C=O Bond			O ... H Distance (Å)		
	Bond Length <sup>a</sup> (Å)	Bond Order <sup>b</sup>	Relaxed Force Constant <sup>c</sup> (cm <sup>-1</sup> )	Bond Length <sup>a</sup> (Å)	Bond Order <sup>b</sup>	Relaxed Force Constant <sup>c</sup> (cm <sup>-1</sup> )			
DapAAH <sup>+</sup>	Ac-Dap	1.364	1.211	5.988	1.226	1.684	11.111	1774	1.612
	Dap-Ala <sup>1st</sup>	1.356	1.165	6.711	1.209	1.947	12.500	1831	
	Ala <sup>1st</sup> -Ala <sup>2nd</sup>	1.329	1.340	7.937	1.234	1.583	10.753	1740	1.647
	COOH				1.202	1.845	13.158	1865	
C2	Ac-Dap	1.363	1.213	6.061	1.226	1.679	11.236	1773	1.615
	Dap-Ala <sup>1st</sup>	1.355	1.175	6.711	1.209	1.927	12.500	1830	
	Ala <sup>1st</sup> -Ala <sup>2nd</sup>	1.337	1.339	7.299	1.230	1.630	10.989	1749	1.656
	COOH				1.200	1.886	13.333	1873	
C9	Ac-Dap	1.362	1.213	6.250	1.225	1.968	11.236	1777	1.612
	Dap-Ala <sup>1st</sup>	1.358	1.164	6.289	1.212	1.893	12.195	1822	
	Ala <sup>1st</sup> -Ala <sup>2nd</sup>	1.344	1.262	7.194	1.224	1.670	11.364	1769	2.133
	COOH				1.205	1.799	12.821	1855	2.021
AADapH <sup>+</sup>	Ac-Ala <sup>1st</sup>	1.362	1.160	6.173	1.226	1.673	11.364	1765	1.962
	Ala <sup>1st</sup> -Ala <sup>2nd</sup>	1.342	1.195	7.143	1.229	1.674	11.236	1751	1.746
	Ala <sup>2nd</sup> -Dap	1.358	1.226	6.369	1.218	1.721	11.765	1798	2.418
	COOH				1.204	1.809	12.658	1853	2.199
C10 <sup>*</sup>	Ac-Ala <sup>1st</sup>	1.352	1.217	6.897	1.227	1.643	11.364	1761	2.019
	Ala <sup>1st</sup> -Ala <sup>2nd</sup>	1.343	1.241	7.194	1.229	1.645	11.111	1745	1.738
	Ala <sup>2nd</sup> -Dap	1.356	1.254	6.329	1.220	1.731	11.494	1790	2.217
	COOH				1.203	1.817	12.658	1861	2.396

<sup>a</sup> Values obtained from the log file of frequency calculations at wB97XD/6-311G\*\* level of theory; the wavenumber is scaled.

<sup>b</sup> Mayer bond order calculated by Multwfn 3.8 using the fchk file from single point calculations.

<sup>c</sup> Values (mdyn/Å) calculated by Compliance 3.0.2 using log file from frequency calculations at wB97XD/6-311G\*\* level of theory.

**Carboxyl C=O stretching band.** As shown in the IRMPD spectra, the carboxyl C=O stretching mode of DapAAH<sup>+</sup> has a peak at 1769 cm<sup>-1</sup>. It is blue-shifted than the peak at 1757 cm<sup>-1</sup> for AADapH<sup>+</sup> by 12 cm<sup>-1</sup>.

The carboxyl group is isolated from the NH<sub>3</sub><sup>+</sup> group in C1 and C2 of DapAAH<sup>+</sup> while interacts with the NH<sub>3</sub><sup>+</sup> group in C1' and C10' of AADapH<sup>+</sup>. This characteristic difference in the involvement of interactions matches the relative shift of the carboxyl C=O stretching mode of DapAAH<sup>+</sup> and AADapH<sup>+</sup>.

In C1 of DapAAH<sup>+</sup>, the carboxyl C=O stretching mode is the primary mode of the computed peak at 1767 cm<sup>-1</sup>, which matches the IRMPD peak at 1769 cm<sup>-1</sup>. As identified through the RDG analysis, the carboxyl C=O group interacts with the Dap N-H group both attractively and repulsively. Besides this, it is not involved in other intramolecular interactions.

In C2 of DapAAH<sup>+</sup>, the carboxyl C=O stretching mode is the primary mode of the computed peak at 1774 cm<sup>-1</sup>, which matches the IRMPD peak at 1769 cm<sup>-1</sup>. The computed wavenumber is higher than that in C1. As identified through RDG analysis, its carboxyl C=O group interacts with one of the Ala<sup>1st</sup> beta hydrogens and the Ala<sup>2nd</sup> N-H nitrogen. But these interactions are weaker than those in C1. The weaker interactions grant C2 a slightly stronger carboxyl C=O bond and a slightly higher wavenumber. The difference of the carboxyl C=O bond strength between C1 and C2 is also reflected in their C=O bond length, bond order, and relaxed force constants. As shown in Table 3.42, the carboxyl C=O bond length in C2 is 1.200 Å, shorter than 1.202 Å in C1. Correspondingly, the carboxyl C=O bond order and relaxed force constant in C2 are 1.886 and 13.333, higher than C1.

In C1 and C2, the carboxyl group does not interact with the NH<sub>3</sub><sup>+</sup> group. No coupling is observed between the carboxyl C=O stretching mode and the amide I modes. The identified internal coordinates of the computed vibrational mode in both C1 and C2 all belong to the carboxyl C=O stretching mode, and their total contribution adds up to around 75%.

In C9 of DapAAH<sup>+</sup>, the carboxyl C=O group interacts with the NH<sub>3</sub><sup>+</sup> group through a weak hydrogen bond with an O --- H distance of 2.021 Å and an HBE of 8.1 kcal/mol. As a result, its carboxyl C=O bond becomes more elongated and weakened than C1 and C2. It is reflected in its much lower

computed wavenumber. In C9, the carboxyl C=O stretching mode is the primary mode of the computed peak at 1757 cm<sup>-1</sup>. It is red-shifted than C1 and C2, and has a poor match to the IRMPD peak at 1769 cm<sup>-1</sup>. Its carboxyl C=O bond length is 1.205 Å, the longest among the DapAAH<sup>+</sup> conformations. Its carboxyl C=O bond order and relaxed force constant are 1.799 and 12.821, respectively, which are lower than those of C1 and C2. Additionally, coupling to the carboxyl C=O stretching mode from the Ala<sup>1st</sup>-Ala<sup>2nd</sup> and Dap-Ala<sup>1st</sup> amide I mode is observed. The contribution from the carboxyl C=O stretching mode, and the Ala<sup>1st</sup>-Ala<sup>2nd</sup> and Dap--Ala<sup>1st</sup> amide I mode are 50.4%, 9.5%, and 7.8%. It is likely because all three C=O oxygens interact with H<sub>41</sub> on the NH<sub>3</sub><sup>+</sup> group.

The carboxyl C=O group interacts with the NH<sub>3</sub><sup>+</sup> group through a weak hydrogen bond with an O---H distance and HBE of 2.199 Å and 6.6 kcal/mol in C1', and 2.396 Å and 4.9 kcal/mol in C10'. Additionally, there is a strong attractive interaction between the Ala<sup>2nd</sup>-Dap amide C=O oxygen and the carboxyl C=O carbon. The interactions weaken the carboxyl C=O bond. As a result, the carboxyl C=O stretching mode of C1' and C10' are both red-shifted. The carboxyl C=O stretching mode is the primary mode of the computed peak at 1755 cm<sup>-1</sup> for C1' and at 1762 cm<sup>-1</sup> for C10'. Both match the IRMPD peak at 1757 cm<sup>-1</sup>. Compared to C10', the stronger hydrogen bond with the NH<sub>3</sub><sup>+</sup> group in C1' resulted in a weaker carboxyl C=O bond and a lower computed wavenumber. This difference is also reflected in their carboxyl C=O bond length and bond orders. The carboxyl C=O bond order in C1 is 1.809, lower than 1.817 in C10'.

Additionally, both the carboxyl C=O group and the Ala<sup>2nd</sup>-Dap amide C=O group form weak hydrogen bonds with H<sub>29</sub> of the NH<sub>3</sub><sup>+</sup> group, and their frequencies are close to each other. A coupling is observed in the IR decomposition result of C1'. The contribution of Ala<sup>2nd</sup>-Dap amide I mode in C1' is 6.7%. Lastly, summing the weights belonging to each normal mode together, the carboxyl C=O stretching mode's contribution in the computed vibrational modes is 64.9% in C1' and 72.9% in C10', both lower than that in C1 and C2 of DapAAH<sup>+</sup>.

In summary, the isolation of the carboxyl group in C1 and C2 of DapAAH<sup>+</sup> grants them a stronger carboxyl C=O bond. While in AADapH<sup>+</sup> conformations, the carboxyl C=O group interacts with

the  $\text{NH}_3^+$  group and the Ala<sup>2nd</sup>-Dap amide C=O group. The difference in the involvement of interactions causes DapAAH<sup>+</sup> to have blue-shifted bands than AADapH<sup>+</sup> in the carboxyl C=O stretching region.

**Amide I band.** In the amide I region, there are three well-resolved peaks at 1648, 1682, and 1726  $\text{cm}^{-1}$  for DapAAH<sup>+</sup>, and a single peak at 1678  $\text{cm}^{-1}$  for AADapH<sup>+</sup>. Good matchings are observed for the representative conformations of DapAAH<sup>+</sup> and AADapH<sup>+</sup>. Unlike the involvement of the two  $\text{NH}_3^+$  scissoring vibrational modes in LysAH<sup>+</sup> and ALysH<sup>+</sup>, only the amide I mode of the three amide bonds contributes to the observed absorption in the IRMPD spectra.

In C1 of DapAAH<sup>+</sup>, the Dap-Ala<sup>1st</sup> amide bond is the only amide bond not involved in hydrogen bonding. So, its amide C=O group is the strongest, with a C=O bond length of 1.209 Å (Table 3.42), the shortest among the three amide bonds. Its higher bond order and relaxed force constant reflect its higher strength than the other two amide bonds. As a result, its amide I mode is the primary mode of the computed peak at 1734  $\text{cm}^{-1}$ , which matches the highest wavenumber IRMPD peak in the amide I region at 1726  $\text{cm}^{-1}$  (Table 3.37).

The  $\text{NH}_3^+$  group hydrogen bonds with the Ac-Dap amide C=O group with an O --- H distance of 1.612 Å and an HBE of 21.4 kcal/mol. Comparatively, the hydrogen bond between the  $\text{NH}_3^+$  group and the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O group is weaker, with an O --- H distance of 1.647 Å and an HBE of 18.4 kcal/mol. Their involvement in hydrogen bonds causes their amide I frequencies to be lower than that of the Dap-Ala<sup>1st</sup> amide bond.

A stronger hydrogen bond elongates the corresponding amide C=O bond more, resulting in a lower amide I frequency. However, the Ac-Dap amide C=O bond length is 1.226 Å, shorter than 1.234 Å of the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O bond. The Ac-Dap amide C=O bond order and relaxed force constant are higher than the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O bond. It suggests a higher strength of the Ac-Dap amide C=O bond.

The Ac-Dap amide I mode is the primary mode of the computed peak at a higher wavenumber of 1680  $\text{cm}^{-1}$ , which matches the IRMPD peak at 1682  $\text{cm}^{-1}$ . The Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide I mode is the primary mode of the computed peak at a lower wavenumber of 1647  $\text{cm}^{-1}$ , which matches the IRMPD peak at

1648  $\text{cm}^{-1}$  (Table 3.37). The matching also suggests a higher strength of the Ac-Dap amide C=O bond. Correspondingly, the Ac-Dap amide C(O)-N(H) bond is found to be longer and weaker than the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C(O)-N(H) bond (Table 3.42).

The same phenomenon has been observed in the other two DapAAH<sup>+</sup> conformations (Table 3.38 and 3.39). Their corresponding computed wavenumbers are all around 1680  $\text{cm}^{-1}$  with almost no variation. This consistency matches the same hydrogen bond strength between H<sub>40</sub> on the NH<sub>3</sub><sup>+</sup> group and the Ac-Dap amide C=O oxygen.

Multiple factors might contribute to the higher strength of the Ac-Dap amide C=O bond and the lower strength of the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O group.

The angles with the vertex of the Dap C<sub>α</sub> in the DapAAH<sup>+</sup> conformations are listed in Table 3.43. Compared to the angle of 109.5° in ideal tetrahedral geometries, the N(H)-C<sub>α</sub>-C<sub>β</sub> angles are significantly higher, while the N(H)-C<sub>α</sub>-H<sub>α</sub> and H<sub>α</sub>-C<sub>α</sub>-C<sub>β</sub> are lower. It suggests that the Dap sidechain and the N(H)-C<sub>α</sub> are pushed toward the back of the molecule, which might strengthen the N(H)-C<sub>α</sub> bond and, in turn, weaken the N(H)-C(O) bond.

Table 3.43  
*Angles with the Vertex of Dap C<sub>α</sub> in the DapAAH<sup>+</sup> Conformations*

	<b>Conf.</b>	<b>N(H)-C<sub>α</sub>-C<sub>β</sub></b>	<b>N(H)-C<sub>α</sub>-C(O)</b>	<b>N(H)-C<sub>α</sub>-H<sub>α</sub></b>	<b>C<sub>β</sub>-C<sub>α</sub>-C(O)</b>	<b>C(O)-C<sub>α</sub>-H<sub>α</sub></b>	<b>H<sub>α</sub>-C<sub>α</sub>-C<sub>β</sub></b>
DapAAH <sup>+</sup>	C1	114.79	110.87	106.23	110.11	108.88	105.61
	C2	114.81	110.88	106.26	110.11	108.89	105.55
	C9	115.72	111.74	105.98	108.11	109.96	105.03
AADapH <sup>+</sup>	C1'	110.50	111.92	109.70	110.31	104.48	109.77
	C10'	111.59	109.06	106.34	116.11	103.93	109.12

As shown in Table 3.42, the vector angle between C→O and H→N in the hydrogen bond with the Ac-Dap amide C=O group is 81.9°. In comparison, the vector angle in the hydrogen bond with the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide group is only 43.7°. The worse hydrogen bond geometry for the Ac-Dap amide bond

decreases its hydrogen bond strength and contributes to the observed higher strength of the Ac-Dap amide C=O bond.

The different positions of those two amide bonds might also contribute to this phenomenon. As in DapAAH<sup>+</sup>, the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide bond is next to the electron-withdrawing carboxyl group, while the Ac-Dap amide bond is at the N-terminus. However, this phenomenon is not observed in the LysAH<sup>+</sup> and ALysH<sup>+</sup> conformations.

Besides the difference in the amide C=O bond strength, the Ac-Dap amide I mode coupling with the higher frequency Dap-Ala<sup>1st</sup> amide I mode also contributes to its higher wavenumber. As shown in Table 3.37, there is a 10.0 % contribution from the Dap-Ala<sup>1st</sup> amide I mode to the computed vibrational mode at 1680 cm<sup>-1</sup>, while the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide I mode is the only identified mode contributing to the computed vibrational mode at 1647 cm<sup>-1</sup>.

C1 and C2 of DapAAH<sup>+</sup> have the same hydrogen bonding network and exhibit similar patterns in the amide I region. Like C1, the Ac-Dap and Dap-Ala<sup>1st</sup> amide I modes are the primary mode of the computed peak at 1679 and 1733 cm<sup>-1</sup>, which match the IRMPD peak at 1682 and 1726 cm<sup>-1</sup>, respectively. A 9.9% contribution of Dap-Ala<sup>1st</sup> amide I mode to the computed peak at 1682 cm<sup>-1</sup> is observed in C2, demonstrating a similar extent of coupling between the Ac-Dap and Dap-Ala<sup>1st</sup> amide I mode as in C1. The Ac-Dap amide C=O bond in C2 is stronger than the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O bond as indicated by its shorter C=O bond length, and higher bond order, and relaxed force constant.

Compared to C1, the hydrogen bond with the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O group is weaker in C2. Its O --- H distance is 1.656 Å, longer than that in C1 by 0.01 Å. As a result, the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O in C2 is slightly stronger than that in C1. It is reflected in its shorter C=O bond length of 1.230 Å and higher bond order and relaxed force constant of 1.630 and 10.989 than C1. The Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide I mode is the primary mode of the computed peak at 1657 cm<sup>-1</sup>, which is higher than that in C1 by 10 cm<sup>-1</sup> and has a slightly worse match with the IRMPD peak at 1648 cm<sup>-1</sup>.

In C9 of DapAAH<sup>+</sup>, the Dap-Ala<sup>1st</sup> amide I mode is the primary mode of the computed peak at 1726 cm<sup>-1</sup>, which is lower than that of C1 and C2, and matches the IRMPD peak at 1726 cm<sup>-1</sup>. As

identified through the RDG analysis, there is an attractive interaction between the Dap-Ala<sup>1st</sup> amide C=O oxygen and the NH<sub>3</sub><sup>+</sup> group. This interaction is absent in C1 and C2. It weakens the Dap-Ala<sup>1st</sup> amide C=O bond and downshifts its amide I wavenumber.

Like that in C1 and C2, C9 also has close to 10% contribution from the Ac-Dap amide I mode to the computed peak at 1726 cm<sup>-1</sup>. Additionally, an 8.2% contribution from the carboxyl C=O stretching mode is observed. Because Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide mode is not involved, this coupling is transmitted from the Dap-Ala<sup>1st</sup> amide bond to the Ac-Dap amide I bond through the backbone and then goes through the NH<sub>3</sub><sup>+</sup> bridge.

The Ac-Dap and Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide I modes are the primary mode of the computed peaks at 1683 and 1675 cm<sup>-1</sup>, which both match the IRMPD peak at 1682 cm<sup>-1</sup>. Like that in C1 and C2, though the Ac-Dap amide C=O group is involved in a stronger hydrogen bond, the corresponding computed wavenumber is close to 1680 cm<sup>-1</sup>. Compared to C1 and C2, the hydrogen bond between the NH<sub>3</sub><sup>+</sup> group and the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide bond is weaker in C9. As a result, its Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O bond is less weakened, and the corresponding computed wavenumber in C9 is higher than that in C1 and C2. So the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide I frequencies in DapAAH<sup>+</sup> conformations correlate well with their different hydrogen bond strength with the NH<sub>3</sub><sup>+</sup> group.

Additionally, the coupling between the Ac-Dap and Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide I mode is stronger in C9. Summing the weights belonging to each mode together, the ratio of contributions from Ac-Dap and Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide I mode to the computed peaks at 1675 and 1683 cm<sup>-1</sup> are 15.5%:42.6%, and 31.3%:14.9%. The corresponding ratios are 0%:55% and 49.4%:0% in C1 and 0%:53.7% and 41.3%:6.2% in C2. Their difference in coupling between these conformations originates from their structural differences.

In C1 and C2, the hydrogen bonds between the NH<sub>3</sub><sup>+</sup> group and the Ac-Dap and Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O groups are strong, with HBE ranging from 18.0 to 21.4 kcal/mol. Therefore, the amide C=O oxygens are tied tightly with their corresponding NH<sub>3</sub><sup>+</sup> hydrogens. Though no contribution of the NH<sub>3</sub><sup>+</sup> scissoring mode is identified through the IR decomposition, investigation of the computed vibrational

modes in GaussView reveals that the  $\text{NH}_3^+$  scissoring motion is involved. However, no coupling between the Ac-Dap and  $\text{Ala}^{1\text{st}}\text{-Ala}^{2\text{nd}}$  amide I modes is observed.

The hydrogen bond to the Ac-Dap in C9 is as strong as that in C1 and C2. In comparison, its hydrogen bond to the  $\text{Ala}^{1\text{st}}\text{-Ala}^{2\text{nd}}$  amide C=O group is much weaker than C1 and C2. Though with a weaker interaction, the  $\text{Ala}^{1\text{st}}\text{-Ala}^{2\text{nd}}$  amide I mode couples with the  $\text{NH}_3^+$  scissoring mode. Furthermore, the Ac-Dap and  $\text{Ala}^{1\text{st}}\text{-Ala}^{2\text{nd}}$  amide I modes in C9 are coupled through the  $\text{NH}_3^+$  bridge.

In C1 and C2, the  $\text{Ala}^{1\text{st}}\text{-Ala}^{2\text{nd}}$  amide C=O groups are far from and orientated nearly opposing to each other. In C9, these two C=O groups are further away but less opposing to each other. The structural positioning and orientation of the Ac-Dap and  $\text{Ala}^{1\text{st}}\text{-Ala}^{2\text{nd}}$  amide C=O groups relative to the  $\text{NH}_3^+$  group and each other in C1 and C2 does not seem to transmit the motion effectively. In C9, though the Ac-Dap and  $\text{Ala}^{1\text{st}}\text{-Ala}^{2\text{nd}}$  amide C=O groups are further away, their orientations seem better for the coupling.

In the  $\text{AADapH}^+$  conformations, the Ac- $\text{Ala}^{1\text{st}}$  amide I mode is the primary mode of the computed peak at  $1671\text{ cm}^{-1}$  in C1' and at  $1667\text{ cm}^{-1}$  in C10'. Both match the IRMPD peak at  $1678\text{ cm}^{-1}$ . Their Ac- $\text{Ala}^{1\text{st}}$  amide C=O group forms a weak hydrogen bond with the  $\text{Ala}^{2\text{nd}}\text{-Dap}$  amide N-H group. The O --- H distance and HBE values are  $1.962\text{ \AA}$  and  $9.0\text{ kcal/mol}$  in C1' and  $2.019\text{ \AA}$  and  $8.0\text{ kcal/mol}$  in C10, respectively. Though involved in much weaker hydrogen bonds than the Ac-Dap amide groups in the  $\text{DapAAH}^+$  conformations, the Ac- $\text{Ala}$  amide C=O bonds in the  $\text{AADapH}^+$  conformations are clearly weakened by the weak intramolecular interactions they participate. Their amide I wavenumbers around  $1670\text{ cm}^{-1}$  are lower than the Ac-Dap amide I wavenumber around  $1680\text{ cm}^{-1}$  in the  $\text{DapAAH}^+$  conformations.

Besides a longer O --- H distance, the vector angle between the Ac- $\text{Ala}^{1\text{st}}$  amide C=O group and the  $\text{Ala}^{2\text{nd}}\text{-Dap}$  amide H-N group is  $59.2^\circ$  in C10' larger than  $31.5^\circ$  in C1'. The larger vector angle in C10' is caused by the bent away of the acetyl group from the helix axis and is not beneficial for the hydrogen bond. The Ac- $\text{Ala}^{1\text{st}}$  amide C=O bond in C1' is weakened more than in C10' by this hydrogen bond. Also, based on the decomposition results, a higher contribution of the lower frequency  $\text{Ala}^{1\text{st}}\text{-Ala}^{2\text{nd}}$  amide I mode is observed in C1' than in C10'. The Ac- $\text{Ala}^{1\text{st}}$  amide I frequency in C1' should be lower

than that in C10'. However, as mentioned before, the computed Ac-Ala<sup>1st</sup> amide I mode in C1' is at 1671 cm<sup>-1</sup>, marginally higher than that of 1667 cm<sup>-1</sup> in C10'.

It is likely caused by the other intramolecular interactions that involve the Ac-Ala<sup>1st</sup> amide C=O group. As identified through the RDG analysis, the Ac-Ala<sup>1st</sup> amide C=O oxygen in C1' and C10' is engaged in attractive interactions to the Dap beta hydrogen and the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide group. The darker blueish color suggests that these two interactions in C10' are stronger than that in C1'. The stronger attractive interaction in C10' attributes to a stronger weakening of its Ac-Ala<sup>1st</sup> amide C=O bond and lowers its amide I frequency. Compared to C1', the Ac-Ala<sup>1st</sup> amide C=O group in C10 is involved in a weaker hydrogen bond but a stronger set of two other intramolecular interactions. As a result, its computed Ac-Ala<sup>1st</sup> amide I wavenumber is lower than that in C1'.

The Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O group in the AADapH<sup>+</sup> conformations forms a strong hydrogen bond with the NH<sub>3</sub><sup>+</sup> group. The O --- H distance and HBE values are 1.746 Å and 14.2 kcal/mol in C1' and 1.738 Å and 14.5 kcal/mol in C10, respectively. As a result, its C=O bond becomes significantly weakened, and its amide I mode is red-shifted. The Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide I mode is the primary mode of the computed peaks at 1658 cm<sup>-1</sup> in C1' and 1652 cm<sup>-1</sup> in C10'. Both are red-shifted than the experimental peak at 1678 cm<sup>-1</sup>.

Based on the decomposition results of the above-stated computed vibrational modes, the contribution of the Ac-Ala<sup>1st</sup> amide I mode is 23.0% in C1' and 16.7% in C10'. It indicates a stronger coupling between those two adjacent amide bonds in C1', which is also observed in the decomposition result of the computed Ac-Ala vibrational modes. The better coupling in C1' seems to benefit from the alignment of those two amide bonds along the helix direction.

The hydrogen bond between the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O group and the NH<sub>3</sub><sup>+</sup> group in C1' is weaker than that in C10'. As a result, its Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C=O bond is slightly stronger, reflected in its higher bond order and relaxed force constant. Taking its coupling with the Ac-Ala<sup>1st</sup> amide I mode into account, the wavenumber of the computed Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide I vibrational mode in C1' is slightly higher than that in C10'.

The Ala<sup>2nd</sup>-Dap amide C=O group is hydrogen-bonded with the NH<sub>3</sub><sup>+</sup> group through a weak hydrogen bond with an O --- H distance and HBE of 2.418 Å and 5.1 kcal/mol in C1', and 2.217 Å and 6.6 kcal/mol in C10'. These two hydrogen bonds are weaker than those between the Ac-Ala<sup>1st</sup> amide C=O group and the Ala<sup>2nd</sup>-Dap amide N-H group. As a result, the Ala<sup>2nd</sup>-Dap amide I mode is the primary mode of the computed peaks at 1702 cm<sup>-1</sup> in C1' and 1696 cm<sup>-1</sup> in C10', which are the highest wavenumber among the computed amide I modes in each conformation.

Because the Ala<sup>2nd</sup>-Dap amide C=O and the carboxyl C=O groups are hydrogen-bonded with the same hydrogen on the NH<sub>3</sub><sup>+</sup> group, and there is also an attractive interaction between the Ala<sup>2nd</sup>-Dap amide oxygen and the carboxyl carbon. The Ala<sup>2nd</sup>-Dap amide I mode is found to couple with the carboxyl C=O stretching in C1'. In C10', a weak coupling with the Ac-Ala<sup>1st</sup> amide I mode is found. It seems to be transmitted through the hydrogen bond between the Ac-Ala<sup>1st</sup> amide C=O group and the Ala<sup>2nd</sup>-Dap amide N-H group.

The computed largest separation between amide I peaks is 87 cm<sup>-1</sup> in C1, 76 cm<sup>-1</sup> in C2, and 51 cm<sup>-1</sup> in C9, significantly larger than 44 cm<sup>-1</sup> in C1' and C10'. The large separations in the DapAAH<sup>+</sup> conformations match the experimental separations in the DapAAH<sup>+</sup> IRMPD spectrum of 88 cm<sup>-1</sup>. For the representative conformations of both DapAAH<sup>+</sup> and AADapH<sup>+</sup>, the two NH<sub>3</sub><sup>+</sup> scissoring vibrational modes are located in the valley between amide I and II bands. They have a low absorption, and do not contribute to any observed IRMPD peaks.

**Amide II band.** In the amide II region, DapAAH<sup>+</sup> has two well-resolved peaks at 1455 and 1522 cm<sup>-1</sup>, and two unresolved peaks at 1480 and 1495 cm<sup>-1</sup>. In the same region, AADapH<sup>+</sup> has a single peak at 1494 cm<sup>-1</sup>. Based on the decomposition results, there are mainly four normal vibrational modes in this region: the amide II mode of the three amide bonds and the umbrella vibrational mode of the NH<sub>3</sub><sup>+</sup> group. Similar to the observations in LysAH<sup>+</sup> and ALysH<sup>+</sup>, the amide II modes and the NH<sub>3</sub><sup>+</sup> umbrella vibrational mode are mostly independent of each other, and the contribution of the C(O)-N(H) stretching in amide II modes is much lower than that of C=O stretching motion in the amide I modes.

In C1 of DapAAH<sup>+</sup>, the Dap-Ala<sup>1st</sup> amide II mode is the primary mode of the computed peak at 1476 cm<sup>-1</sup>, which matches the IRMPD peak at 1480 cm<sup>-1</sup>. The Dap-Ala<sup>1st</sup> amide bond is not involved in hydrogen bonding, and its C(O)-N(H) bond stays unaffected, is the weakest among all three amide bonds, and serves as a reference in the following discussions. Its C(O)-N(H) bond length, bond order, and relaxed force constant are 1.356 Å, 1.165, and 6.711, respectively. Similar values are observed for the Dap-Ala<sup>1st</sup> amide C(O)-N(H) bond in C2 and C9 of DapAAH<sup>+</sup>.

As mentioned in the amide I mode analysis, the Ac-Dap amide C=O bonds of the DapAAH<sup>+</sup> conformations are surprisingly strong. Accordingly, their Ac-Dap amide C(O)-N(H) bonds are unusually weak. But still, their Ac-Dap amide II frequencies are higher than that of the Dap-Ala<sup>1st</sup> amide group because of their involvement in hydrogen bonding.

In C1, the Ac-Dap amide II mode is the primary mode of the computed peaks at 1500 cm<sup>-1</sup>, which matches the IRMPD peaks at 1495 cm<sup>-1</sup>. Though it is involved in a stronger hydrogen bond, its computed frequency is lower than that of the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide bond. The surprising properties of the Ac-Dap amide bond are not well understood at the time of writing.

The Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide II mode is the primary mode of the computed peaks at 1528 cm<sup>-1</sup>, which is the highest among all three amide II modes and matches the IRMPD peaks at 1522 cm<sup>-1</sup>. The Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide bond forms a very strong hydrogen bond with the NH<sub>3</sub><sup>+</sup> group, and its C(O)-N(H) bond becomes significantly shortened to a bond length of 1.329 Å, much shorter than 1.356 of the Dap-Ala<sup>1st</sup> amide group. Its higher strength is also reflected in its higher bond order of 1.340 and lower relaxed force constant of 7.937 than the corresponding values for the Dap-Ala<sup>1st</sup> amide group.

The umbrella vibrational mode of the NH<sub>3</sub><sup>+</sup> group is the primary mode of the computed peak at 1465 cm<sup>-1</sup>, which matches the IRMPD peak at 1455 cm<sup>-1</sup>. In this computed mode, a total of 13.5% contributions from the Ac-Dap amide II mode is observed. The coupling between the NH<sub>3</sub><sup>+</sup> umbrella mode and the Ac-Dap amide II mode seems transmitted through their strong hydrogen bond.

The wavenumber and decomposition results of the computed vibrational modes in C2 are similar to those in C1. The Dap-Ala<sup>1st</sup> and Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide II modes are the primary modes of the computed

peak at 1477 and 1519  $\text{cm}^{-1}$ , which match the IRMPD peaks at 1480 and 1522  $\text{cm}^{-1}$ . With the contradiction of a strong hydrogen bond of 21.3 kcal/mol and a long C(O)-N(H) bond of 1.363 Å, the Ac-Dap amide II mode is assigned as the primary mode of the computed peak at 1500  $\text{cm}^{-1}$ , between those for the Dap-Ala<sup>1st</sup> and Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide II modes. The NH<sub>3</sub><sup>+</sup> umbrella mode is the primary mode of the computed peak at 1460  $\text{cm}^{-1}$ , which matches the IRMPD peak at 1455  $\text{cm}^{-1}$ .

The Ac-Dap and Dap--Ala<sup>1st</sup> amide I modes in C9 are the primary mode of the computed peaks at 1505 and 1478  $\text{cm}^{-1}$ , which match the IRMPD peaks at 1495 and 1480  $\text{cm}^{-1}$ . Unlike C1 and C2, the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide group in C9 is involved in a much weaker hydrogen bond with the NH<sub>3</sub><sup>+</sup> group, and its amide C(O)-N(H) bond is less strengthened than in C1 and C2. As a result, its Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide II mode is the primary mode of the computed peak at 1509  $\text{cm}^{-1}$ . It is lower than that in C1 and C2, and has a worse match to the IRMPD peak at 1522  $\text{cm}^{-1}$ .

Though coupled with the Ac-Dap amide II mode as that in C1 and C2, the NH<sub>3</sub><sup>+</sup> umbrella mode in C9 couples more strongly with the Dap CH<sub>2</sub> scissoring mode. Summing the weights belonging to each mode together, the ratio of contribution of the NH<sub>3</sub><sup>+</sup> umbrella mode and the Dap CH<sub>2</sub> scissoring mode to the computed peaks at 1433 and 1456  $\text{cm}^{-1}$  are 6.8%:11.5%, and 14.7%:8.5%.

Considering their total contributions, the Dap CH<sub>2</sub> scissoring mode and the NH<sub>3</sub><sup>+</sup> umbrella mode are the primary mode of the peak at 1433 and 1456  $\text{cm}^{-1}$ , which matches the IRMPD peaks at 1430 and 1455  $\text{cm}^{-1}$ . The strong coupling of Dap CH<sub>2</sub> scissoring mode with the umbrella mode of the highly charged NH<sub>3</sub><sup>+</sup> group contributes to the high absorption of the peak at 1433  $\text{cm}^{-1}$ . It is not observed in the other DapAAH<sup>+</sup> conformations and strongly suggests the existence of C9.

As mentioned in the amide I mode analysis, the Ala<sup>2nd</sup>-Dap amide C=O groups in C1' and C10' involve in the weakest hydrogen bond and becomes weakened than the other two amide groups. As a result, their amide I frequencies are the highest. Interestingly, their computed amide II frequencies are also the highest, even though the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide C(O)-N(H) bonds are the shortest. It is caused by the hydrogen bond between the Ala<sup>2nd</sup>-Dap amide N-H hydrogen and the Ac-Ala<sup>1st</sup> amide C=O oxygen. Upshift of amide II frequency has been observed in model peptides where the N-H group is hydrogen

bonded to a water molecule (Myshakina, Ahmed, & Asher, 2008). The N-H bending motion is an important part of the amide II motion. The hydrogen bond involving the N-H hydrogen hinders its movement. As a result, the Ala<sup>2nd</sup>-Dap amide II frequency is significantly high. In C1' and C10', the Ala<sup>2nd</sup>-Dap amide II mode is the primary mode of the computed peaks at 1518 and 1513 cm<sup>-1</sup>, respectively, which are the highest among their amide II frequencies.

In C1', the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide group involves in the strongest hydrogen bond with the NH<sub>3</sub><sup>+</sup> group. Its C(O)-N(H) bond becomes significantly shortened to a bond length of 1.342 Å, the shortest among the amide groups. As a result, its Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide II mode is the primary mode of the computed peak at 1497 cm<sup>-1</sup>, which matches the IRMPD peak at 1494 cm<sup>-1</sup>. Additionally, a coupling is observed between the Ala<sup>1st</sup>-Ala<sup>2nd</sup> and Ala<sup>2nd</sup>-Dap amide II modes in C1'. Summing the weights belonging to each mode together, the ratio of contribution between Ac-Ala<sup>1st</sup> and Ala<sup>1st</sup>-Ala<sup>2nd</sup> toward the amide II modes at 1497 and 1518 cm<sup>-1</sup> are 25.6%:5.8%, and 7.9%:22.5%.

The Ac-Ala<sup>1st</sup> amide group forms a weak hydrogen bond with the Ala<sup>2nd</sup>-Dap N-H hydrogen, and its C(O)-N(H) bond is not strengthened and has a bond length of 1.362 Å, the longest among the three amide bonds. The Ac-Ala<sup>1st</sup> amide II is the primary mode of the computed peak at 1471 cm<sup>-1</sup>, which is the most red-shifted and has the worst match to the IRMPD peak at 1494 cm<sup>-1</sup> among the three amide II modes.

Compared to C1', the computed wavenumber of the Ac-Ala<sup>1st</sup> amide II mode in C10' is significantly higher. Its Ac-Ala<sup>1st</sup> amide II is the primary mode of the computed peak at 1500 cm<sup>-1</sup>, higher than 1471 cm<sup>-1</sup> of C1' by 29 cm<sup>-1</sup>, and has a better match to the IRMPD peak at 1494 cm<sup>-1</sup>. As mentioned in the analysis of the AADapH<sup>+</sup> amide I modes, though the Ac-Ala<sup>1st</sup> amide group in C10' forms a weaker hydrogen bond with the Ala<sup>2nd</sup>-Dap amide group, it involves in other stronger intramolecular interactions. As a result, its Ac-Ala<sup>1st</sup> amide C=O bond is weakened while the C(O)-N(H) bond becomes strengthened.

Compared to C1', C10' is missing a strong attractive interaction between the Ac-Ala<sup>1st</sup> amide N-H nitrogen and the Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide N-H hydrogen. This interaction weakens the Ac-Ala<sup>1st</sup> amide C(O)-N(H) bond in C1' and lowers its amide II wavenumber. Comparatively, the missing interaction in

C10' keeps its Ac-Ala<sup>1st</sup> amide C(O)-N(H) bond strong and brings a higher Ac-Ala<sup>1st</sup> amide II wavenumber. The Ac-Ala<sup>1st</sup> amide C(O)-N(H) bond length in C10' is 1.352 Å, shorter than 1.362 in C1' by 0.01 Å. It also reflects the higher strength of the Ac-Ala<sup>1st</sup> amide C(O)-N(H) bond.

The NH<sub>3</sub><sup>+</sup> umbrella mode is the primary mode of computed mode at the peak of 1465 cm<sup>-1</sup> in C1' and 1478 cm<sup>-1</sup> in C10'. Both are more red-shifted and have a poor match to the experimental IRMPD peak at 1494 cm<sup>-1</sup>.

**Carboxyl C-O-H bending mode.** The peak at 1375 cm<sup>-1</sup> in the DapAAH<sup>+</sup> IRMPD spectrum and 1383 cm<sup>-1</sup> in the AADapH<sup>+</sup> IRMPD spectrum correspond to the carboxyl C-O-H bending mode. The decomposition results (Table 3.37 to 3.41) show that the carboxyl C-O(H) stretching motion and the C<sub>α</sub>-C(O) stretching motion are heavily involved. In the DapAAH<sup>+</sup> conformations, the absorption of the computed peaks is low and does not match the high absorption observed in the DapAAH<sup>+</sup> IRMPD spectrum. In the AADapH<sup>+</sup> conformations, the wavenumber and absorption of computed peaks both match those shown in the AADapH<sup>+</sup> IRMPD spectrum.

In C1 of DapAAH<sup>+</sup>, the carboxyl C-O-H bending mode is the primary mode of the computed peak at 1376 cm<sup>-1</sup>, and the frequency matches the IRMPD peak at 1375 cm<sup>-1</sup>. The mode decomposition result and motion validation reveal that the Ala<sup>2nd</sup> C<sub>α</sub>-H bending mode and the Ala<sup>2nd</sup> CH<sub>3</sub> umbrella mode are also involved, which are resulted from the strong carboxyl C<sub>α</sub>-C(O) bending motion.

In C2 of DapAAH<sup>+</sup>, the carboxyl C-O-H bending mode is the primary mode of the computed peaks at 1363 and 1366 cm<sup>-1</sup>. These two peaks are more red-shifted than those of C1 and have a worse match to the IRMPD peak at 1375 cm<sup>-1</sup>. The decomposition results show that the Ala<sup>1st</sup> CH<sub>3</sub> umbrella mode is heavily involved. As identified through the RDG analysis, there is an attractive interaction between the carboxyl C=O group and the Ala<sup>1st</sup> CH<sub>3</sub> group. It helps to transmit the coupling.

In C9 of DapAAH<sup>+</sup>, the carboxyl C-O-H bending mode is the primary mode of the computed peak at 1372 cm<sup>-1</sup> which matches the IRMPD peak at 1375 cm<sup>-1</sup>. Like those in C1, the Ala<sup>2nd</sup> C<sub>α</sub>-H bending mode and the Ala<sup>2nd</sup> CH<sub>3</sub> umbrella mode are heavily involved because of the strong carboxyl C<sub>α</sub>-C(O) stretching motion.

In the AADapH<sup>+</sup> conformations, the carboxyl C-O-H bending mode is the primary mode of the computed peak at 1378 cm<sup>-1</sup> in C1' and 1374 cm<sup>-1</sup> in C10'. Both match the IRMPD peak at 1383 cm<sup>-1</sup>. Investigation of their motions reveals that the Dap C<sub>α</sub>-H bending mode is also involved. However, their involvement of the non-polar alkyl vibrational modes is much less than that in the DapAAH<sup>+</sup> conformations. It is reflected in the higher contribution of the carboxyl C-O-H bending mode to the computed modes. The carboxyl C-O-H bending motion contribution adds up to 50.3% in C1' and 54.2% in C10'. They are much higher than those in the DapAAH<sup>+</sup> conformations. The less involvement of non-polar alkyl vibrational modes likely contributes to the high absorption of AADapH<sup>+</sup> in this region.

**Amide III modes.** There are two well-resolved and separated peaks at 1204 and 1265 cm<sup>-1</sup> in the DapAAH<sup>+</sup> IRMPD spectrum in this region. The computed peaks in C1, C2, and C9 all have a good match to the IRMPD spectrum. For AADapH<sup>+</sup>, there is one broad peak centered at 1249 cm<sup>-1</sup>, and the computed spectra from C1' and C10' show a reasonable match to the IRMPD spectrum.

In C1, the Ac-Dap and Dap-Ala<sup>1st</sup> amide III modes are the primary mode of computed peaks at 1250 and 1193 cm<sup>-1</sup>, which match the IRMPD peaks at 1265 and 1204 cm<sup>-1</sup>, respectively. The Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide III mode is the primary mode of the computed peak at 1212 cm<sup>-1</sup>, which has a predicted low absorption and matches the valley between the two observed IRMPD peaks. As the C(O)-N(H) stretching is an important motion in the amide III mode, its strength affects the amide III frequency. The Dap-Ala<sup>1st</sup> amide group has the weakest C(O)-N(H) bond. As a result, its amide III wavenumber is the lowest.

Like those in C1, the Ac-Dap and Dap-Ala<sup>1st</sup> amide III modes are the primary mode of computed peaks at 1251 and 1193 cm<sup>-1</sup> in C2, and 1253 and 1185 cm<sup>-1</sup> in C9, which match the IRMPD peaks at 1265 and 1204 cm<sup>-1</sup>. The Ala<sup>1st</sup>-Ala<sup>2nd</sup> amide III mode is the primary mode of the computed peak at 1223 cm<sup>-1</sup> in C2, and 1214 cm<sup>-1</sup> in C9, which have a predicted low absorption and match the valley between the two observed IRMPD peaks.

In C1', the Ac-Ala<sup>1st</sup>, Ala<sup>1st</sup>-Ala<sup>2nd</sup>, and the Ala<sup>2nd</sup>-Dap amide III modes are the primary mode of the computed peaks at 1215, 1235, and 1266 cm<sup>-1</sup>, respectively, which all fall under the broad IRMPD

peak centered at  $1249\text{ cm}^{-1}$ . Similarly, the Ac-Ala<sup>1st</sup>, Ala<sup>1st</sup>-Ala<sup>2nd</sup>, and the Ala<sup>2nd</sup>-Dap amide III modes in C10' are the primary mode of the computed peaks at  $1248, 1219, 1264\text{ cm}^{-1}$ .

**Carboxyl C-O(H) stretching.** In this region, there is one main peak at  $1113\text{ cm}^{-1}$  and a shoulder peak at  $1143\text{ cm}^{-1}$  in the DapAAH<sup>+</sup> IRMPD spectrum. The computed spectra have a reasonable match to the IRMPD spectrum. The decomposition results show that there are three major vibrational modes involved in this region: the carboxyl C-O(H) stretching, Dap C<sub>β</sub>-N(H3) stretching, and Ala<sup>2nd</sup> N(H)-C<sub>α</sub> stretching.

In the AADapH<sup>+</sup> IRMPD spectrum, there is one major peak at  $1155\text{ cm}^{-1}$  and one minor peak at  $1097\text{ cm}^{-1}$ . The involved motions for the AADapH<sup>+</sup> conformations are the carboxyl C-O(H) stretching, Dap C<sub>β</sub>-N(H3) stretching, and Dap N(H)-C<sub>α</sub> stretching. Comparatively, better matching is achieved for AADapH<sup>+</sup> than DapAAH<sup>+</sup> in this region.

In C1, the carboxyl C-O(H) stretching mode is the primary mode of the computed peak at  $1160\text{ cm}^{-1}$ , which matches the IRMPD peak at  $1143\text{ cm}^{-1}$ . The decomposition result shows that the computed mode is dominated by the carboxyl C-O(H) stretching mode with a 32.8% contribution and has limited contribution from other motions. The Dap C<sub>β</sub>-N(H3) stretching motion and N(H)-C<sub>α</sub> stretching motion are involved in the computed vibrational mode at  $1107\text{ cm}^{-1}$ , which matches the IRMPD peak at  $1113\text{ cm}^{-1}$ . As the decomposition shows, the NH<sub>3</sub><sup>+</sup> group wobbles on the top of the Dap sidechain in this computed mode. The Ala<sup>2nd</sup> N(H)-C<sub>α</sub> stretching motion and carboxyl C-O(H) stretching motion are involved in the computed vibrational mode at  $1104\text{ cm}^{-1}$ , which matches the IRMPD peak at  $1113\text{ cm}^{-1}$ .

Compared to C1, the carboxyl group in C2 is less isolated. As a result, C2 exhibits a different pattern in this region. The carboxyl C-O(H) stretching dominates the computed peak at  $1157\text{ cm}^{-1}$ , which matches the IRMPD peak at  $1143\text{ cm}^{-1}$ . But this computed mode is very dispersed. The identified contribution of the carboxyl C-O(H) stretching mode is 12.4%, much lower than 32.8% in the computed peak at  $1160\text{ cm}^{-1}$  in C1.

The computed mode at  $1138\text{ cm}^{-1}$ , which matches the IRMPD peak at  $1143\text{ cm}^{-1}$ , is more dispersed. The decomposition result and the investigation of its motion in GaussView reveal that the

backbone from Dap-Ala<sup>1st</sup> amide bond to the carboxyl group are all involved. The carboxyl C-O(H) stretching mode is the primary mode of the computed mode at 1120 cm<sup>-1</sup>, which matches the IRMPD peak at 1113 cm<sup>-1</sup>. As the decomposition results show, the Dap N(H)-C<sub>α</sub> stretching motion is also observed in this mode. The N-terminal geometry of C2 is similar to C1. The computed peak at 1102 cm<sup>-1</sup> in C2 is like the computed peak at 1107 cm<sup>-1</sup> in C1. As the Dap C<sub>β</sub>-N(H3) and N(H)-C<sub>α</sub> bonds stretch and retract, the NH<sub>3</sub><sup>+</sup> group wobbles on the top of the Dap sidechain.

In C9, the carboxyl C-O(H) stretching mode is the primary mode of the computed peaks at 1147 cm<sup>-1</sup>, which matches the IRMPD peak at 1143 cm<sup>-1</sup>. The decomposition result shows that the carboxyl C-O(H) stretching mode dominates this computed mode and has a total identified contribution of 29.6%. The carboxyl C-O(H) stretching mode is also the primary mode of a very dispersed computed peak at 1098 cm<sup>-1</sup>, which matches the IRMPD peak at 1113 cm<sup>-1</sup>. Like that in C1 and C2, the computed peak at 1096 cm<sup>-1</sup> in C9 involves the motion of multiple bonds of the Dap residue, and also involves the Dap C<sub>β</sub>-N(H3) and N(H)-C<sub>α</sub> stretching motion and the NH<sub>3</sub><sup>+</sup> group wobbling motion.

The carboxyl C-O(H) stretching is the primary mode of the computed peak at 1164 cm<sup>-1</sup> in C1', and 1161 cm<sup>-1</sup> in C10', which both match the IRMPD peak at 1155 cm<sup>-1</sup>. A small contribution from the Dap N(H)-C<sub>α</sub> stretching is observed from the decomposition results.

Similar to that of DapAAH<sup>+</sup> conformations, wobbling of the NH<sub>3</sub><sup>+</sup> group on the top of the Dap side chain is observed in the AADapH<sup>+</sup> conformations. The stretching modes involved indicate the Dap C<sub>β</sub>-N(H3) stretching and Dap N(H)-C<sub>α</sub> stretching. The resulting computed peak shows up at 1082 cm<sup>-1</sup> in C1' and 1084 cm<sup>-1</sup> in C10', which both match the IRMPD peak at 1097 cm<sup>-1</sup>.

## Summary

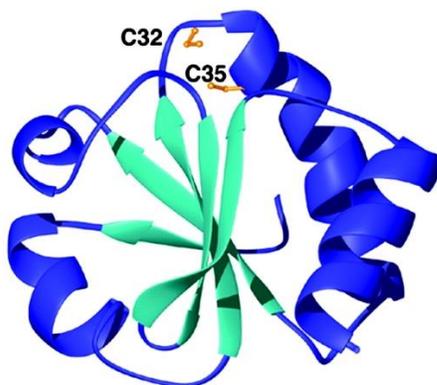
Structural and energetic characterizations of the isomeric tripeptide pair DapAA and AADap have been carried out using IRMPD and computational chemistry. Good agreement was achieved between the experimental and theoretical proton affinities. The stability difference between the protonated peptides was found to be the main source of the experimental observed different basicities. Characteristically different structures of the protonated peptides were revealed by the results from IRMPD and

computational studies. Helical conformations were found to dominate the conformation population of AADapH<sup>+</sup>, while DapAAH<sup>+</sup> seem to adopt coiled conformations.

In the DapAAH<sup>+</sup> conformations, the positive charge is mainly stabilized through the hydrogen bonds with the NH<sub>3</sub><sup>+</sup> group. Contrarily, in the AADapH<sup>+</sup> conformations, the hydrogen bonds with the NH<sub>3</sub><sup>+</sup> group are much weaker, while the interaction of the backbone dipole moment with the NH<sub>3</sub><sup>+</sup> group is very strong and plays an important role in enhancing the stability of the AADapH<sup>+</sup> conformations. The characteristically different IRMPD spectra of DapAAH<sup>+</sup> and AADapH<sup>+</sup> reflect their adopted different hydrogen-bonding networks and indicate differences in structural diversity and bond strength of the carboxyl and amide groups. The broad and complex IRMPD spectrum of DapAAH<sup>+</sup> results from its diverse conformations and hydrogen-bonding networks, while the narrow and clean IRMPD spectrum of AADapH<sup>+</sup> results from less diverse helical conformations and the same hydrogen-bonding network.

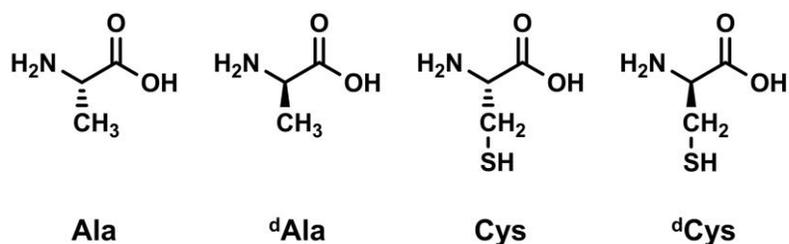
## CHAPTER 4 : CYSTEINE-CONTAINING PEPTIDES IN THE GAS PHASE

Thioredoxin is an enzyme that regulates the cellular redox potential through reducing the protein disulfide bonds. It is critical for redox regulation of protein function and signaling (Arnér & Holmgren, 2000), and is truly ubiquitous in all living cells (Holmgren, 2003). The function of thioredoxin relies on two cysteine (Cys, C) residues in its active site (Figure 4.1). C32 sits at the N-terminus of an  $\alpha$ -helix motif, and has an abnormally low pKa value of 6.3 in human thioredoxin. (Julie, Clore, & Angela, 1992) The low pKa value allows it to act as nucleophile during catalysis of disulfide exchange and is correlated with the redox potential of the enzyme (Mössner, Iwai, & Glockshuber, 2000). An electrostatic interaction between the C32 thiolate anion and the dipole of the active site helix has been proposed to explain the lowered pKa of the nucleophilic thiol group (Katti, LeMaster, & Eklund, 1990). Alanine-based peptides are known for their propensity to form helices (Oommachen, Ren, & McCallum, 2008). In the previous study from our group, a series of Cys-containing polyalanine peptides have been designed to model the thiolate – helix dipole interaction. It has been shown that the N-terminal Cys peptides are significantly more acidic than the corresponding isomeric C-terminal peptides. The thiolate anion of the N-terminal Cys peptides form helical conformations while that of the N-terminal Cys peptides adopt coiled structures (Ren et al., 2009; Tan & Ren, 2007).



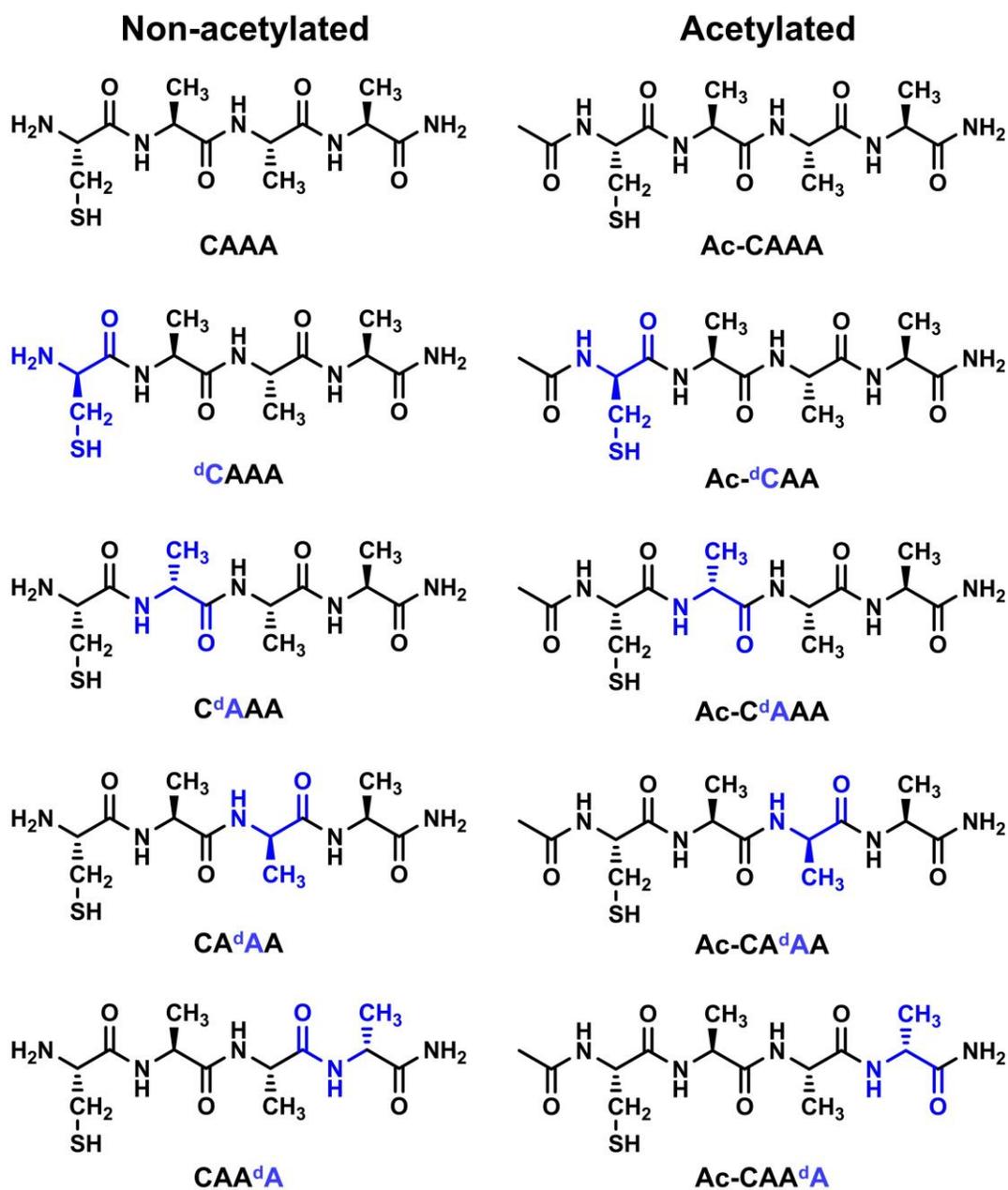
*Figure 4.1.* Structure of *E. coli* thioredoxin. The active site cysteines (C32 and C35) are indicated in yellow. (Adopted from reference Kumar, Tabor, & Richardson, 2004.)

Building upon the established helical oligopeptides, the effect of chirality on the thermochemical properties of Cys-containing poly-alanine peptides in the gas-phase was investigated by incorporating D-amino acids into the sequence. In the previous study from our group, subtle acidity alternation was observed upon chiral inversion of the amino acid in a tripeptide (Table 1.1). In the context of this thesis, model tetrapeptides with Cys at the N-terminus were designed and characterized using the Cooks' kinetic method. All of the peptides are composed of a single Cys as an acidic probe, a poly-alanine (Ala, A) backbone, and an amidated C-terminus. The C-terminus is amidated to ensure a single deprotonatable site at the sidechain of the cysteine residue. The structures of involved amino acids are shown in Scheme 4.1, where the symbol d denotes a D-form amino acids.



Scheme 4.1. Structures of amino acids used in the peptides.

Systematic investigation on the effect of chirality on the gas-phase acidity was performed using two series of tetra-peptides (Scheme 4.2). All peptides share the same sequence: CAAA and differ in the chirality of the amino acids and the N-terminal acetylation. In the first series, there are five peptides, and their N-termini are kept as a free amino group. Among them, one peptide does not contain any D-amino acid and acts as a reference peptide, and the other four peptides contain one D-amino acid but at the four different positions. The second series of the peptides are generated by acetylating the N-terminus of the corresponding peptides in the first series. In total, ten peptides were synthesized and characterized through systematically alternating the chirality of each amino acid in the sequence.



Scheme 4.2. Cys-containing poly-Ala tetrapeptides with a D-amino acid colored in blue.

### Non-Acetylated D-Amino Acid-Containing Peptides

In the first series of tetrapeptides, the N-terminus is kept as a free amino group. Compared to the acetylated peptides in the second series, the moiety on the N-terminal side of Cys is much smaller and has a minimal effect on the geometry of the moiety on the C-terminal side of Cys.

Mass spectrometry experiments were conducted to determine the gas-phase acidity of the five peptides using the kinetic method. The experiments were carried out under identical instrumental conditions. The nebulizing gas pressure, drying gas pressure, and collision cell pressure are 41 psi, 18 psi, and 0.2 mTorr, respectively. The drying gas temperature is 160 °C. The ESI voltage, shield voltage, and capillary voltage are -4.0 kV, -600 V, and -40 V, respectively. The six reference acids used in the measurements are listed in Table 4.1.

Table 4.1

*Reference Acids Used for Acidity Measurements of Non-Acetylated D-Amino Acid-Containing Peptides*

Reference Acid	Abbreviation	$\Delta_{\text{acidH}}$	$\Delta_{\text{acidG}}$	$\Delta_{\text{acidS}}$
		(kcal/mol)	(kcal/mol)	(cal/mol K)
Cyanoacetic Acid	CYA	330.3	323.7	22.1
Dichloroacetic Acid	DCA	328.4	321.9	21.8
Dibromoacetic Acid	DBA	328.3	321.3	23.5
4-Nitrobenzoic Acid	4NBA	328.1	321.1	23.5
3,5-Bis(trifluoromethyl) Benzoic Acid	TFMBA	324.4	317.4	23.5
Trifluoroacetic Acid	TFA	323.8	317.4	21.5

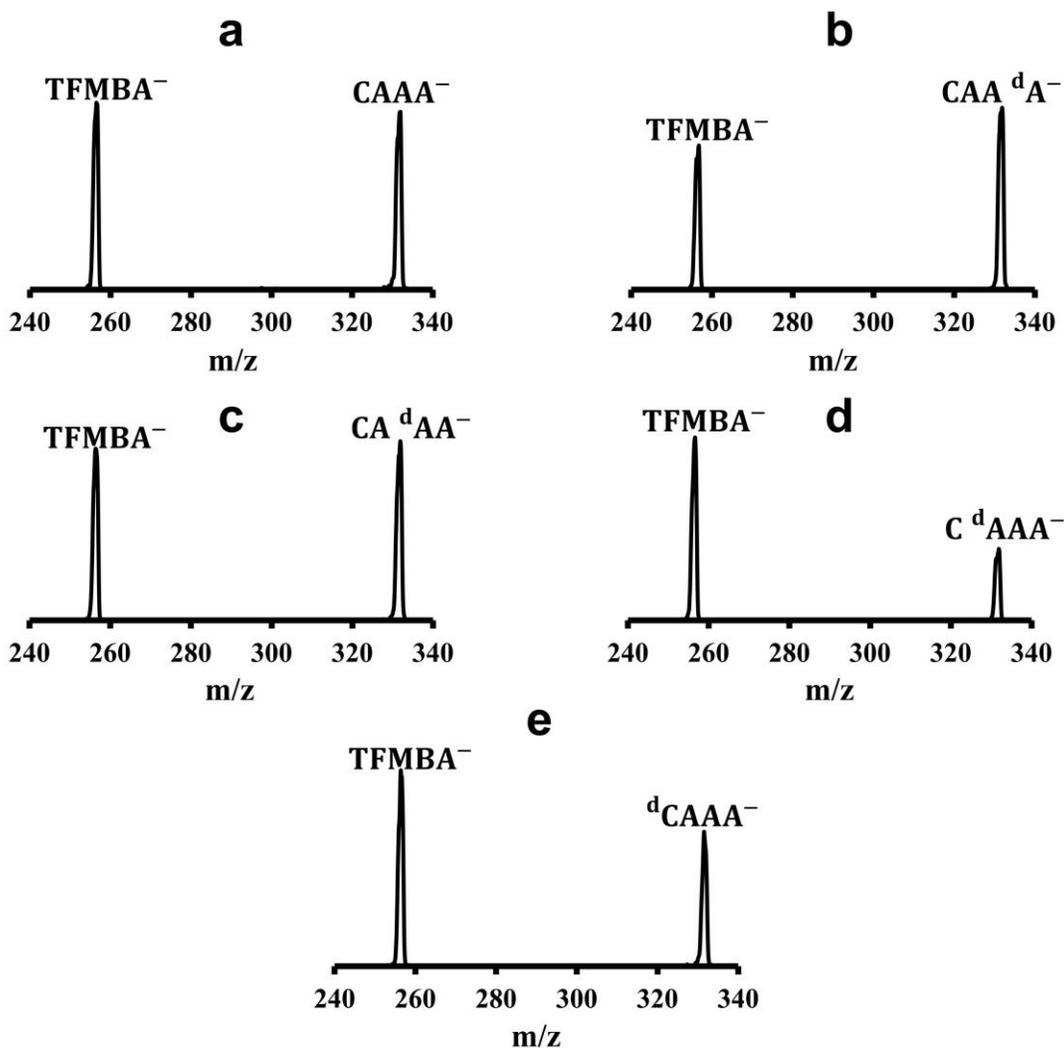
Values obtained from the NIST Chemistry Webbook, Linstrom, 2005.

### Acidity Comparison using One Reference Acid at a Single Collision Energy

Product ion analyses were conducted by fragmenting the proton-bound dimer between the peptides and 3,5-Bis(trifluoromethyl) Benzoic Acid (TFMBA) at 1.5 eV ( $E_{\text{cm}}$ ) in the profile mode. Their product ion spectra are shown individually in Figure 4.1. The relative intensity of the peptide ions compared to TFMBA is an intuitive and direct indicator of their different acidities. The peptide with a higher relative intensity is more acidic, while the peptide with a lower relative intensity is less acidic.

In the product ion spectrum of  $[\text{TFMBA}^- \cdots \text{H}^+ \cdots \text{CAA}^{\text{d}}\text{A}^-]$  (Figure 4.1 (b)), the intensity of  $\text{CAA}^{\text{d}}\text{A}^-$  is higher than that of  $\text{TFMBA}^-$ . Comparatively, in the product ion spectra of  $[\text{TFMBA}^- \cdots \text{H}^+ \cdots \text{CAAA}^-]$  and  $[\text{TFMBA}^- \cdots \text{H}^+ \cdots \text{CA}^{\text{d}}\text{AA}^-]$ , the intensity of the peptide is comparable to

the intensity of TFMBA<sup>-</sup>. While in the product ion spectra of [TFMBA<sup>-</sup>...H<sup>+</sup>...C<sup>d</sup>AAA<sup>-</sup>] and [TFMBA<sup>-</sup>...H<sup>+</sup>...<sup>d</sup>CAAA<sup>-</sup>], the peptide ions C<sup>d</sup>AAA<sup>-</sup> and <sup>d</sup>CAAA<sup>-</sup> both have lower intensity than TFMBA<sup>-</sup>. Comparatively, the intensity difference between C<sup>d</sup>AAA<sup>-</sup> and TFMBA<sup>-</sup> is larger than that between <sup>d</sup>CAAA<sup>-</sup> and TFMBA<sup>-</sup>. Based on the relative intensity of the peptides, an approximate acidity rank from the most acidic peptide to the least acidic peptide can be established as: CAA<sup>d</sup>A > CA<sup>d</sup>AA ~ CAAA > <sup>d</sup>CAAA > C<sup>d</sup>AAA.



*Figure 4.2.* Product ion spectra resulting from fragmentation of the proton-bound dimer between the non-acetylated peptides and TFMBA at 1.5 eV ( $E_{cm}$ ) under CID pressure of 0.2 mTorr. (a),  $[\text{TFMBA}^- \cdots \text{H}^+ \cdots \text{CAAA}^-]$ ; (b),  $[\text{TFMBA}^- \cdots \text{H}^+ \cdots \text{CAAdA}^-]$ ; (c),  $[\text{TFMBA}^- \cdots \text{H}^+ \cdots \text{CA}^d\text{AA}^-]$  (d),  $[\text{TFMBA}^- \cdots \text{H}^+ \cdots \text{C}^d\text{AAA}^-]$ ; (e),  $[\text{TFMBA}^- \cdots \text{H}^+ \cdots {}^d\text{CAAA}^-]$ .

### Acidity Comparison Using Different Reference Acids at Multiple Collision Energies

In order to achieve semi-quantitative comparison of the gas-phase acidity among the model peptides, collision induced experiments were carried out at ten collision energies. For each peptide, six reference acids are used to form the proton bound dimers. The intensities of the peptide ions, reference ions, and their secondary fragment ions were collected to calculate the branching ratios, as described in Chapter 2. The branching ratios,  $\ln[I_{\text{Pep}}^- / I_{\text{Ref}}^-]$ , for the five peptides with all reference acids are summarized in tables from Table 4.2 to Table 4.6.

Table 4.2  
Branching Ratios of CAAA with Six Selected Reference Acids

	Collision Energy, $E_{cm}$ (eV)									
	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75
CYA	4.55	4.01	3.42	3.02	2.81	2.76	2.81	3.04	3.43	3.92
DCA	1.66	1.40	0.97	0.67	0.47	0.34	0.22	0.16	0.16	0.22
DBA	0.89	0.64	0.30	0.02	-0.13	-0.30	-0.48	-0.60	-0.60	-0.55
4NBA	3.90	3.71	3.18	2.68	2.34	2.06	1.80	1.61	1.51	1.49
TFMBA	1.02	0.84	0.50	0.23	0.03	-0.18	-0.38	-0.51	-0.56	-0.58
TFA	-1.33	-1.75	-2.10	-2.34	-2.46	-2.49	-2.52	-2.54	-2.45	-2.31

Table 4.3  
Branching Ratios of <sup>d</sup>C AAA with Six Selected Reference Acids

	Collision Energy, $E_{cm}$ (eV)									
	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75
CYA	4.19	3.62	3.07	2.70	2.54	2.52	2.58	2.79	3.25	3.76
DCA	1.31	1.00	0.60	0.29	0.12	0.02	-0.09	-0.14	-0.13	-0.05
DBA	0.45	0.22	-0.07	-0.31	-0.47	-0.60	-0.77	-0.89	-0.89	-0.83
4NBA	3.50	3.15	2.63	2.19	1.89	1.63	1.42	1.25	1.16	1.16
TFMBA	0.49	0.33	0.05	-0.20	-0.39	-0.56	-0.75	-0.88	-0.91	-0.94
TFA	-1.67	-2.09	-2.42	-2.67	-2.76	-2.77	-2.81	-2.83	-2.73	-2.58

Table 4.4  
Branching Ratios of C<sup>d</sup>AAA with Six Selected Reference Acids

	Collision Energy, $E_{cm}$ (eV)									
	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75
CYA	3.50	3.09	2.61	2.29	2.15	2.18	2.29	2.56	3.01	3.54
DCA	0.74	0.47	0.13	-0.17	-0.30	-0.37	-0.45	-0.49	-0.47	-0.38
DBA	-0.15	-0.33	-0.56	-0.77	-0.89	-0.99	-1.13	-1.24	-1.24	-1.17
4NBA	2.86	2.60	2.10	1.72	1.46	1.26	1.06	0.92	0.84	0.84
TFMBA	-0.06	-0.21	-0.44	-0.66	-0.80	-0.96	-1.12	-1.23	-1.26	-1.26
TFA	-2.30	-2.69	-3.00	-3.16	-3.23	-3.21	-3.21	-3.20	-3.10	-2.93

Table 4.5  
*Branching Ratios of CA<sup>d</sup>AA with Six Selected Reference Acids*

	Collision Energy, E <sub>cm</sub> (eV)									
	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75
CYA	4.61	4.06	3.45	3.05	2.85	2.79	2.83	3.05	3.47	3.94
DCA	1.79	1.46	1.03	0.67	0.48	0.36	0.24	0.18	0.19	0.25
DBA	0.87	0.63	0.31	0.06	-0.12	-0.29	-0.47	-0.59	-0.60	-0.55
4NBA	4.16	3.82	3.25	2.74	2.40	2.11	1.85	1.67	1.56	1.54
TFMBA	1.08	0.87	0.54	0.25	0.05	-0.16	-0.37	-0.50	-0.54	-0.57
TFA	-1.45	-1.79	-2.15	-2.38	-2.48	-2.51	-2.56	-2.55	-2.45	-2.31

Table 4.6  
*Branching Ratios of CAA<sup>d</sup>A with Six Selected Reference Acids*

	Collision Energy, E <sub>cm</sub> (eV)									
	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75
CYA	4.92	4.35	3.68	3.25	3.03	2.95	2.99	3.20	3.60	4.09
DCA	2.12	1.76	1.29	0.90	0.69	0.55	0.41	0.35	0.34	0.39
DBA	1.19	0.96	0.59	0.31	0.12	-0.08	-0.27	-0.40	-0.41	-0.38
4NBA	4.39	4.09	3.46	2.92	2.53	2.20	1.93	1.73	1.61	1.57
TFMBA	1.39	1.16	0.79	0.48	0.25	0.03	-0.20	-0.35	-0.41	-0.44
TFA	-1.09	-1.46	-1.82	-2.07	-2.20	-2.24	-2.29	-2.32	-2.22	-2.09

The branching ratio plots of the five peptides with each of the six reference acids are overlaid and shown in Figure 4.2. CAAA does not contain d-amino acid, and is colored in green. The colors for the other four D-amino acid containing peptides are assigned and are indicated in the plots. From the highest to the lowest branching ratio, the colors are assigned as red, pink, purple, and blue for CAA<sup>d</sup>A, CA<sup>d</sup>AA, <sup>d</sup>CAAA, and C<sup>d</sup>AAA, respectively.

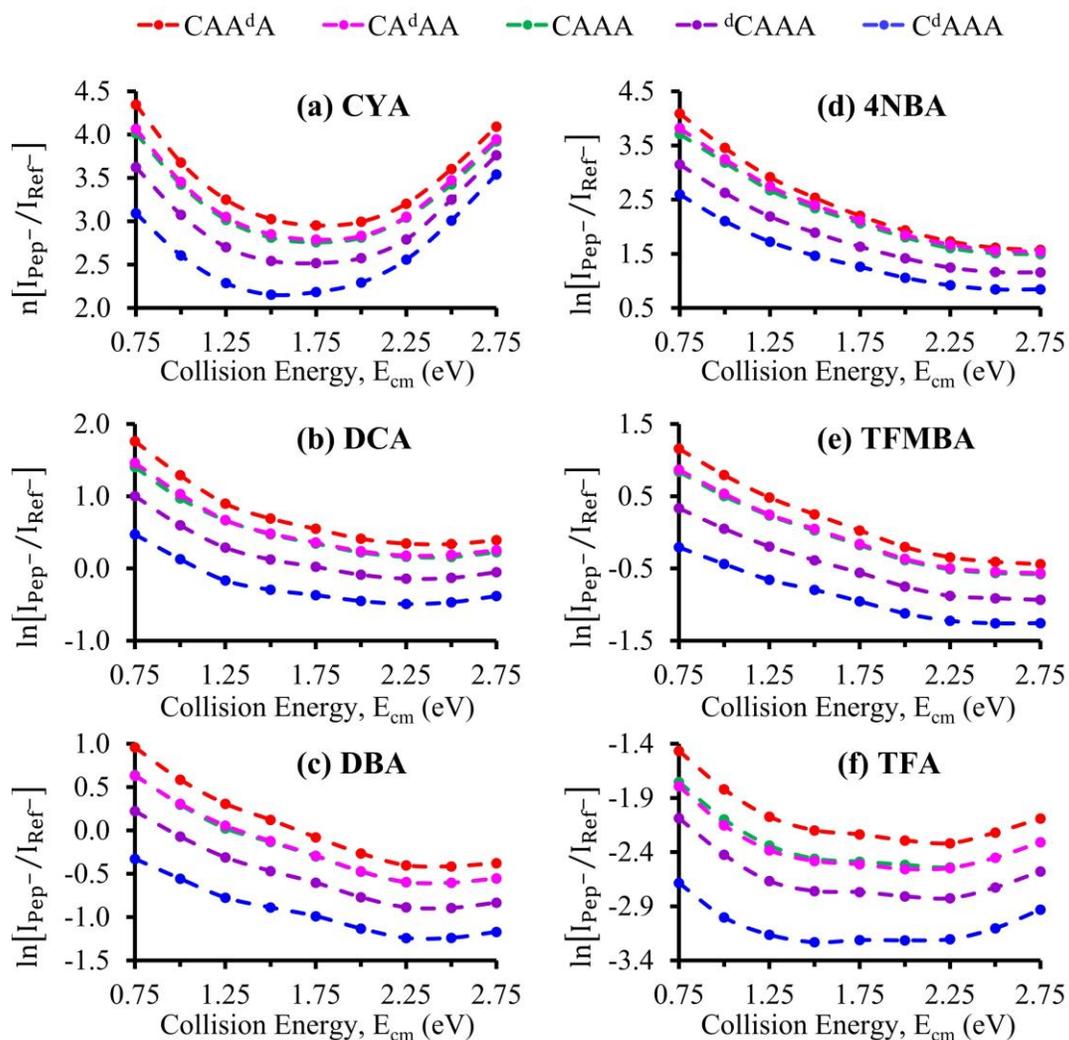


Figure 4.3. Branching ratio plots for non-acetylated D-amino acid-containing peptides using six reference acids.

As a qualitative measure, the branching ratios hold important information about the relative acidity of the peptides. When compared using the same reference acid, higher (or more positive) branching ratios indicate a greater acidity of the peptide. The assigned colors are indicative of the peptides' acidity. The “hot” colors, red and pink, are used for the more acidic peptides, while the “cold” colors, purple and blue, correspond to the less acidic peptides.

No matter which reference acid is used for comparison, CAA<sup>d</sup>A in red always has the highest branching ratio values, while C<sup>d</sup>AAA in blue always has the lowest values across the whole collision

energy range. Also, regardless of the reference acid used, the branching ratios of CAAA and CA<sup>d</sup>AA are always almost equal to each other, while the branching ratios of <sup>d</sup>CAAA in purple always lie between that of CAAA and C<sup>d</sup>AAA. The branching ratio values strongly indicate the following acidity trend from the most acidic to the least acidic: CAA<sup>d</sup>A (red) > CA<sup>d</sup>AA (pink) ~ CAAA (green) > <sup>d</sup>CAAA (purple) > C<sup>d</sup>AAA (blue). The acidity trend is the same as that obtained from the product ion analyses of the peptides with TFMBA at single collision energy.

Additionally, the shape of the branching ratio plots across the collision energy range contains valuable information about the entropy change during the dissociation of the proton-bound dimer between the peptides and the reference acids.

In the standard kinetic method scheme, where two dissociating parts are structurally similar, the entropy term  $\Delta(\Delta S)/R$  is negligible and will not change as the collision energy increases (eq 2.4). As a result, the branching ratios approaches zero as the collision energy increases. But as shown in Figure 4.2 (b, c, e), the branching ratio plots with DCA, DBA, and TFMBA go across the zero axis, while the branching ratios with TFA in Figure 4.2 (f) become more negative as the collision energy increases. The reference acids are carboxylic acids and are structurally different from the thiol group in the peptides. The entropy term is not negligible and makes the branching ratio plots with the reference acids not approaching the zero axis.

In the extended kinetic method scheme, the reference acids are selected to be structurally similar to each other but not necessarily similar to the analyte of interest. So, their entropy term in eq 2.4 will likely to be the same. For this reason, the reference acids are all carboxylic acids. In ideal cases where the entropy term for the reference acids is the same, their branching ratio plot for the same analyte will have similar shapes. However, the branching ratio plots for the same peptide using the six reference acids are actually in different shapes.

As an example, in Figure 4.2 (a), the branching ratio for CAA<sup>d</sup>A in red with cyanoacetic acid (CYA) decreases between 0.75 eV and 1.50 eV and increases between 1.75 eV and 2.75 eV. This rounding bottom curvature is characteristically different from that for the other five reference acids. In

Figure 4.2 (f), the branching ratio for CAA<sup>d</sup>A with trifluoroacetic acid (TFA) increases at higher energies, but the plot curvature is flatter than that of CYA. For the other four reference acids in Figure 4.2 (b – d), the branching ratios decrease in the range of collision energies. The different curvatures indicate that entropy terms (eq 2.4) for the six reference acids are not the same and would affect the quantitative determination of the gas-phase deprotonation enthalpy ( $\Delta_{\text{acid}}H_{\text{pep}}^{\circ}$ ) and acidity ( $\Delta_{\text{acid}}G_{\text{pep}}^{\circ}$ ) using the extended kinetic method.

The peptides are structurally similar, with the only difference in the chirality of one or two alpha carbons. A high degree of similarity is observed when comparing the branching ratio plots for the peptides using the same reference acid. As shown in Figure 4.2 (a), the branching ratio plots for all peptides with CYA have the rounding bottom curvature. The branching ratio plots with the other five reference acids are also characteristically similar. The similar branching ratio plot curvatures indicate that the peptides' entropy change in the dissociation of the proton-bound dimer are similar.

Another way to compare the relative acidity between peptides is to use one peptide as a “peptide reference acid”. In this way, the entropy term becomes negligible. CAAA is selected as the “peptide reference acid”. The branching ratio between a peptide and CAAA is obtained by subtracting the branching ratio of CAAA to reference (e.g. CYA) from the branching ratio of a peptide to the same reference acid (CYA). Details are shown in Scheme 4.3. Because the same reference acid is used, the enthalpy ( $\Delta_{\text{acid}}H_{\text{Ref}}$ ) and entropy ( $\Delta_{\text{acid}}S_{\text{Ref}}$ ) terms for the reference acid are canceled out. The branching ratios with CAAA reflects a “direct” acidity comparison between the peptides and CAAA. The branching ratios with CAAA were calculated using all selected reference acids and graphed in Figure 4.3. The color is coded the same way as that shown in Figure 4.2.

$$\ln \left[ \frac{I_{\text{Pep}^-}}{I_{\text{Ref}^-}} \right] = \frac{\Delta_{\text{acidHRef}} - \Delta_{\text{acidHPep}}}{RT_{\text{eff}}} + \frac{\Delta_{\text{acidSPep}} - \Delta_{\text{acidSRef}}}{R}$$

$$\ln \left[ \frac{I_{\text{CAAA}^-}}{I_{\text{Ref}^-}} \right] = \frac{\Delta_{\text{acidHRef}} - \Delta_{\text{acidHCAAA}}}{RT_{\text{eff}}} + \frac{\Delta_{\text{acidSCAAA}} - \Delta_{\text{acidSRef}}}{R}$$

$$\ln \left[ \frac{I_{\text{Pep}^-}}{I_{\text{Ref}^-}} \right] - \ln \left[ \frac{I_{\text{CAAA}^-}}{I_{\text{Ref}^-}} \right] = \ln \left[ \frac{I_{\text{Pep}^-}}{I_{\text{CAAA}^-}} \right]$$

$$= \frac{\Delta_{\text{acidHRef}} - \Delta_{\text{acidHPep}}}{RT_{\text{eff}}} + \frac{\Delta_{\text{acidSPep}} - \Delta_{\text{acidSRef}}}{R}$$

$$- \frac{\Delta_{\text{acidHRef}} - \Delta_{\text{acidHCAAA}}}{RT_{\text{eff}}} - \frac{\Delta_{\text{acidSCAAA}} - \Delta_{\text{acidSRef}}}{R}$$

$$\ln \left[ \frac{I_{\text{Pep}^-}}{I_{\text{CAAA}^-}} \right] = \frac{\Delta_{\text{acidHCAAA}} - \Delta_{\text{acidHPep}}}{RT_{\text{eff}}} - \frac{\Delta_{\text{acidSCAAA}} - \Delta_{\text{acidSPep}}}{R}$$

Where  $\frac{\Delta_{\text{acidSCAAA}} - \Delta_{\text{acidSPep}}}{R} \approx 0$

Scheme 4.3. Determination of the branching ratio of a peptide with CAAA.

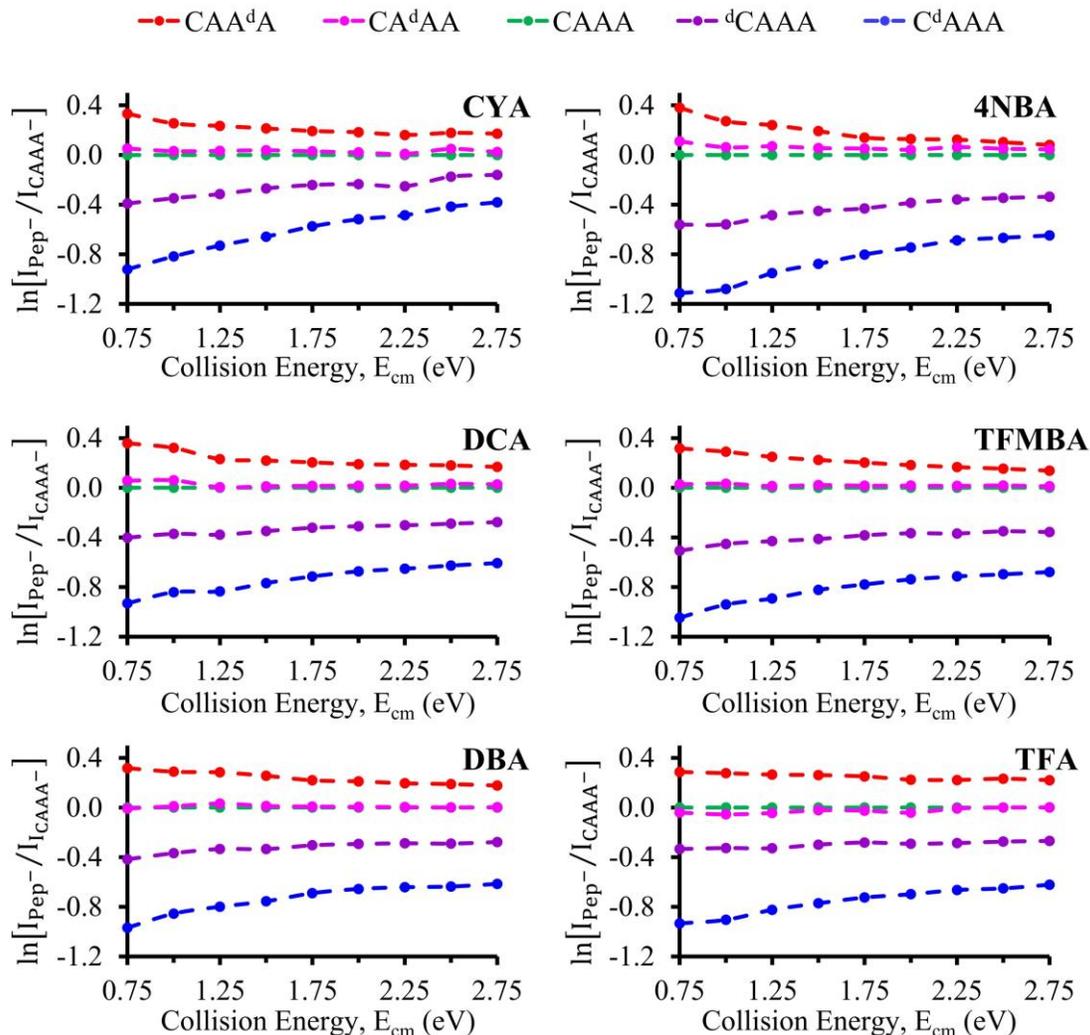


Figure 4.4. Branching ratio plots of the peptides with CAAA using different reference acids.

Compared to the branching ratio plots with the reference acids, the branching ratio plots with CAAA stay in the same order. The relative positions of the branching ratio plots remain the same and reflect the relative acidity of the peptides.

Additionally, the branching ratio plots with CAAA are characteristically similar and do not vary with the reference acids. The independence of the branching ratio plots from reference acids reflects the fact that the branching ratios with CAAA are intrinsically governed by the difference between the thermochemical properties of the peptides. The branching ratios with CAAA are overlaid in Figure 4.4 to demonstrate the variation when using different reference acids. The plot with the same color includes

branching ratios determined using six reference acids. It appears that small variations are observed when different reference acids are used to determine the branching ratio with CAAA. But consistently, the branching ratios for CAA<sup>d</sup>A are positive and the largest, while the branching ratios for C<sup>d</sup>AAA are negative and the smallest.

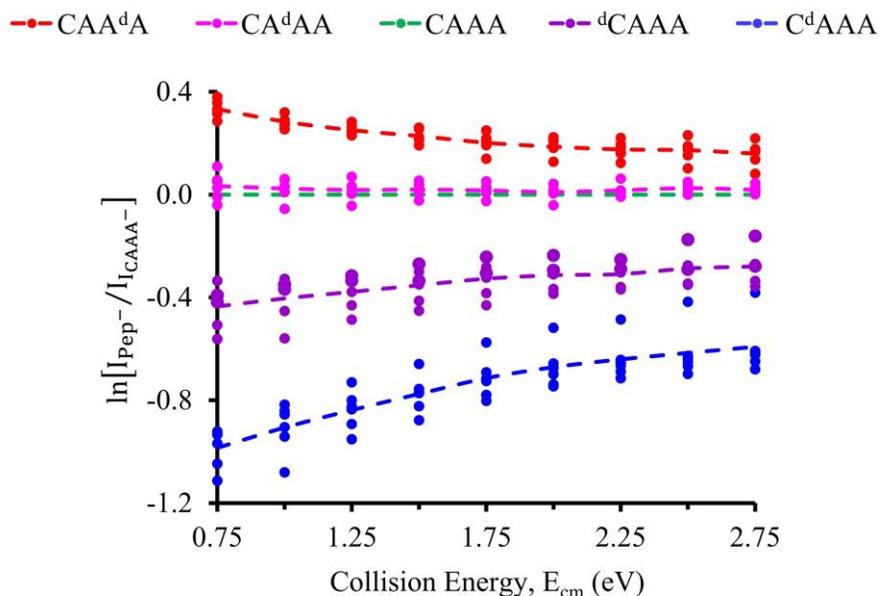


Figure 4.5. Overlay of branching ratio plots of the peptides with CAAA. (The plot with the same color contains the branching ratios determined from six reference acids. The lines connect the average of the branching ratios.)

Furthermore, as shown in Figures 4.3 and 4.4, the branching ratios approach zero as the collision energies increase. This feature matches the expected appearance of branching ratio plots in the standard kinetic method scheme. If the compounds involved in the two competing dissociation channels are structurally similar, the entropy term can be canceled. As a result, the branching ratios are inversely correlated with the temperature and will approach the zero axis as the collision energy increases.

As shown in Scheme 4.3, firstly, the enthalpy ( $\Delta_{\text{acid}}H_{\text{Ref}}$ ) and entropy ( $\Delta_{\text{acid}}S_{\text{Ref}}$ ) terms for the reference acid are canceled out because the same reference acid is used to calculate the branching ratio with CAAA. Then the entropy terms of the peptides ( $\Delta_{\text{acid}}S_{\text{Pep}}$  and  $\Delta_{\text{acid}}S_{\text{CAAAA}}$ ) are assumed to be equal

and get canceled out, leaving the branching ratios depending on the enthalpy terms and the collision energy only. The assumption is suggested by the characteristically similar branching ratio plot shapes of the peptides with the same reference acid. The approaching zero-axis feature in the branching ratio plots with CAAA further validates the assumption that the entropy effect gets eliminated through the mathematical treatment of taking the difference of the branching ratios with the same reference.

In summary, the acidity comparison of the non-acetylated peptides was firstly achieved through product ion analysis of the proton-bound dimer of the peptides with one reference acid at one collision energy. And the following acidity order from most acidic to the least acidic:  $CAA^dA > CA^dAA \sim CAAA > {}^dCAAA > C^dAAA$  was achieved. More extensive comparisons were performed using six different reference acids at ten different collision energies. And the same acidity ladder of the peptides was obtained.

When different reference acids are used, the shape of the branching ratio plots are characteristically different. These features reveal that though all reference acids are carboxylic acids, their entropy change are different during the dissociation of the proton-bound dimer with the peptide. “Direct” acidity comparison of the peptides with CAAA was achieved by taking the difference of the peptides’ branching ratios with that of CAAA using the same reference acid. The entropy change of the peptides during the dissociation of the proton-bound dimer was suggested to be similar from the appearance of their branching ratio plots with CAAA.

Among the d-alanine containing peptides, acidity suppression is observed after chiral inversion of the alanine closest to the acidic cysteine, while acidity enhancement is observed after chiral inversion of the alanine the furthest from the acidic cysteine.

### **Acetylated D-Amino Acid-Containing Peptides**

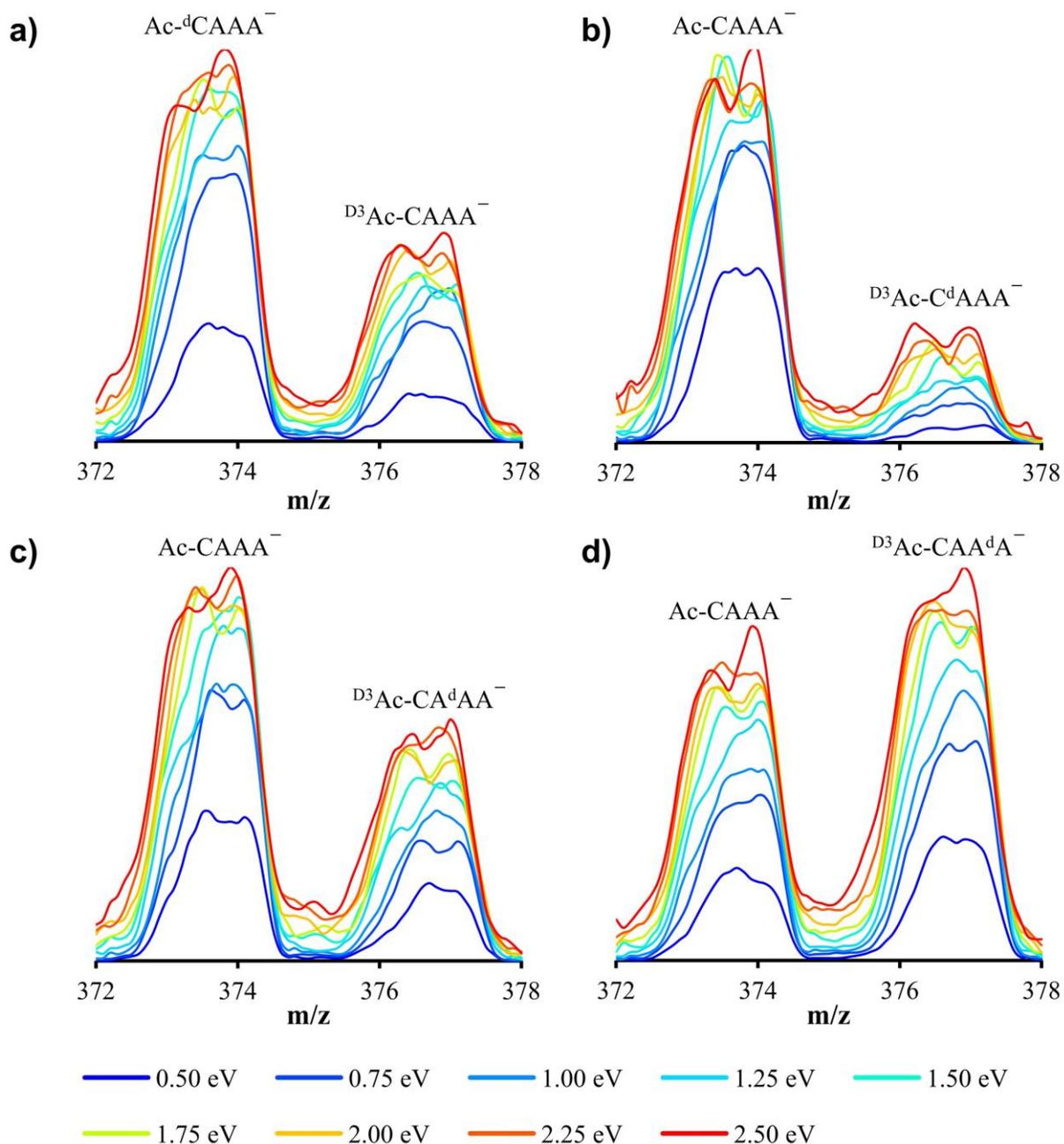
As shown in Scheme 4.2, the peptides in the second group share the same sequence, chirality, and C-terminus modification with the first group of peptides. They differ from the first group of peptides in the way that their N-terminus is acetylated. The acetylated N-terminus is much larger than the free amino

group, and would potentially form strong interaction with the moiety on the C-terminal side of Cys. As a result, the acetylation could affect the overall geometry and acidity of the peptides.

The acetylation was carried out using the regular acetic anhydride and deuterated acetic anhydride. The peptides from acetylation using the deuterated acetic anhydride have three deuterium atoms on the acetyl group. They are three Daltons heavier than the peptides from acetylation using regular acetic anhydride. The mass difference allows direct acidity comparison between the peptides through fragmentation of their proton-bound dimers. The kinetic isotope effect of the deuterium atoms is negligible and will not affect the rates of the two competing fragmentation channels of the proton-bound dimer. In the primary isotope kinetic effect, a bond to the isotopically-labeled atom is formed or broken during the reaction (Westheimer, 1961). In the deuterium labeled peptides, the C-H(D) bonds in the acetyl group stay intact during the fragmentation process of the proton-bound dimer. Therefore, there is no primary kinetic isotope effect. Secondary isotope effects is considered to be mainly due to changes in C-H(D) force constants near the reaction site upon going from reactants to the transition state of the rate-determining step of the reaction (Matsson & Westaway, 1999; Westaway, 2006). Because the deuterium atoms are located far from the dissociation site, the secondary kinetic isotope effect is also negligible.

### **Direct Acidity Comparison with Ac-CAAA**

Direct acidity comparison of the D-amino acid-containing peptides with Ac-CAAA was carried out by fragmenting their proton-bound dimer with Ac-CAAA. Firstly, the product ion analysis was carried out in the profile mode to gain a qualitative measure of their different acidities under identical instrument conditions. The nebulizing gas pressure, drying gas pressure, and collision cell pressure were 41 psi, 18 psi, and 0.2 mTorr, respectively. The drying gas temperature is 160 °C. The ESI voltage, shield voltage, and capillary voltage were -4.0 kV, -600 V, -and 40 V, respectively. The spectra were recorded in profile mode and are shown in Figure 4.5. The collision energy ( $E_{cm}$ ) ranges from 0.5 eV to 2.5 eV, with 0.25 eV increment. The spectra are color coded based the collision energies. As the collision energy increases, the color of spectra changes from “cold” blue to “hot” red.



*Figure 4.6.* Fragmentation spectra of the proton-bound dimer between the D-amino acid-containing peptides and the all L-amino acid reference peptide. a)  $\text{Ac-dCAAA}^-$  and  $\text{D}^3\text{Ac-CAAA}^-$ . b)  $\text{D}^3\text{Ac-C}^d\text{AAA}^-$  and  $\text{Ac-CAAA}^-$ . c)  $\text{D}^3\text{Ac-CA}^d\text{AA}^-$  and  $\text{Ac-CAAA}^-$ . d)  $\text{D}^3\text{Ac-CAA}^d\text{A}^-$  and  $\text{Ac-CAAA}^-$ .

Firstly, as the collision energy increases, the intensity of both involved peptides increases. This is expected because at higher energies, more proton-bound dimers are fragmented to produce the peptide ions. In the fragmentation experiments involving the D-Cys containing peptide,  $\text{Ac-dCAAA}^-$  and

deuterium-labeled reference peptide  $^{D3}\text{Ac-CAAA}$  are used. The fragmentation spectra of the proton bound dimer between  $\text{Ac-}^d\text{C AAA}^-$  and  $^{D3}\text{Ac-C AAA}^-$  are shown in Figure 4.5 (a). At all collision energies from 0.5 eV to 2.5 eV, the intensity of  $\text{Ac-}^d\text{C AAA}^-$  is higher than that of  $^{D3}\text{Ac-C AAA}^-$ . It suggests that during the dissociation of  $[\text{Ac-}^d\text{C AAA}^- \cdot \text{H}^+ \cdot ^{D3}\text{Ac-C AAA}^-]$ , the proton in between is less likely to be acquired by  $\text{Ac-}^d\text{C AAA}^-$ . As a result,  $\text{Ac-}^d\text{C AAA}$  is more acidic than  $^{D3}\text{Ac-C AAA}$ .

The D-Ala containing peptides with a deuterated acetyl group are used in the fragmentation experiments with the non-deuterated reference peptide,  $\text{Ac-C AAA}$ . The corresponding fragmentation spectra are shown in Figure 4.5 (b–d). In the fragmentation experiments for  $^{D3}\text{Ac-C}^d\text{ AAA}$ , the intensity of  $^{D3}\text{Ac-C}^d\text{ AAA}^-$  is significantly lower than that of  $\text{Ac-C AAA}^-$ , indicating a significantly lower acidity of  $\text{Ac-C}^d\text{ AAA}$  compared to that of  $\text{Ac-C AAA}$ . Comparatively, in the fragmentation experiments for  $^{D3}\text{Ac-CA}^d\text{ AA}$ , the intensity of  $^{D3}\text{Ac-CA}^d\text{ AA}^-$  is lower than that of  $\text{Ac-C AAA}^-$  by a smaller amount. It suggests that  $\text{Ac-CA}^d\text{ AA}$  is less acidic than  $\text{Ac-C AAA}$ , but more acidic than  $\text{Ac-C}^d\text{ AAA}$ . In the fragmentation experiments for  $^{D3}\text{Ac-CAA}^d\text{ A}$ , the intensity of  $^{D3}\text{Ac-CAA}^d\text{ A}$  is higher than that of  $\text{Ac-C AAA}$ , suggesting the higher acidity of  $\text{Ac-CAA}^d\text{ A}$ . As the chiral inversion center moves away from the N-terminal Cys to the C-terminus, the acidity of the D-alanine containing peptides increases. The same trend was observed in the non-acetylated peptides.

As indicated by their higher intensity than  $\text{Ac-C AAA}^-$ , both  $\text{Ac-}^d\text{C AAA}$  and  $\text{Ac-CAA}^d\text{ A}$  are more acidic than  $\text{Ac-C AAA}$ . Comparatively, the intensity difference for  $\text{Ac-}^d\text{C AAA}$  is larger. It suggests that  $\text{Ac-}^d\text{C AAA}$  is more acidic than  $\text{Ac-CAA}^d\text{ A}$ . The results of the relative intensities of the peptides suggest the following acidity order from the most acidic to the least acidic:  $\text{Ac-}^d\text{C AAA} > \text{Ac-CAA}^d\text{ A} > \text{Ac-C AAA} > \text{Ac-CA}^d\text{ AA} > \text{Ac-C}^d\text{ AAA}$ .

The fragmentation experiments were also performed in the centroid mode, and the intensities of the peptide ions were recorded. The results were used to obtain the branching ratios of the D-amino acid-containing peptides with  $\text{Ac-C AAA}$ ,  $\ln[I_{\text{pep}}^-/I_{\text{Ac-C AAA}}^-]$ . The branching ratios with  $\text{Ac-C AAA}$  as the reference are summarized in Table 4.7 and plotted in Figure 4.6. The branching ratio plots are color coded based on their values. From the highest to the lowest branching ratios, the peptides follow an order

of: Ac-<sup>d</sup>CAAA, Ac-CAA<sup>d</sup>A, Ac-CA<sup>d</sup>AA and Ac-C<sup>d</sup>AAA and are colored in red, pink, purple and blue, respectively.

Table 4.7  
Branching Ratios of D-Amino Acid-Containing Peptides with Ac-CAAA as the Reference

	Collision Energy, E <sub>cm</sub> (eV)								
	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5
Ac- <sup>d</sup> CAAA	0.90	0.82	0.76	0.72	0.68	0.66	0.63	0.62	0.60
Ac-CAA <sup>d</sup> A	0.32	0.31	0.29	0.26	0.27	0.25	0.23	0.22	0.21
Ac-CAAA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ac-CA <sup>d</sup> AA	-0.79	-0.73	-0.68	-0.64	-0.59	-0.56	-0.54	-0.52	-0.50
Ac-C <sup>d</sup> AAA	-2.18	-1.98	-1.80	-1.65	-1.50	-1.38	-1.29	-1.23	-1.16

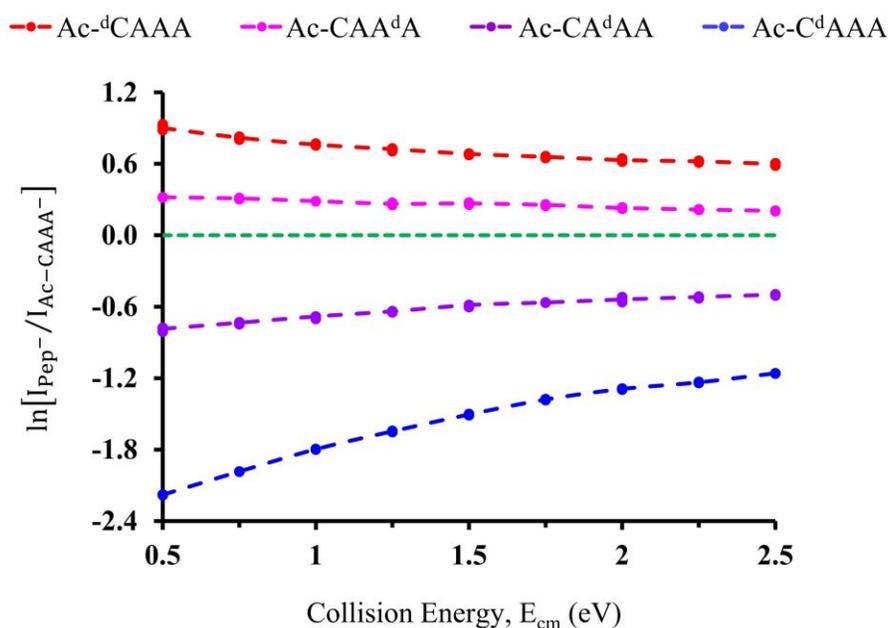


Figure 4.7. Branching ratio plots of D-amino acid-containing peptides with Ac-CAAA as the reference.

Peptides with higher branching ratios are more acidic when Ac-CAAA is used as the common reference. As shown in Figure 4.6, the branching ratios of Ac-<sup>d</sup>CAAA and Ac-CAA<sup>d</sup>A are consistently higher than zero, while those of Ac-CA<sup>d</sup>AA and Ac-C<sup>d</sup>AAA are lower than zero. Comparatively, the branching ratios for Ac-C<sup>d</sup>AAA are the smallest. The acidity order based on the branching ratio plots is

the same as that obtained from the product ion analysis in the profile mode:  $\text{Ac-}^{\text{d}}\text{CAAA} > \text{Ac-CAA}^{\text{d}}\text{A} > \text{Ac-CAAA} > \text{Ac-CA}^{\text{d}}\text{AA} > \text{Ac-C}^{\text{d}}\text{AAA}$ .

The deprotonation site in the D-amino acid-containing peptides and the reference peptide are both the thiol groups on the cysteine sidechain. So, the two parts in the proton-bound dimers are structurally similar, and the entropy difference between the two fragmentation channels is expected to be negligible. The prediction is supported by the approaching zero-axis feature of the branching ratio plots for the D-amino acid-containing peptides. As a result, the branching ratios reflect the deprotonation enthalpy ( $\Delta_{\text{acid}}\text{H}$ ) differences between the D-amino acid-containing peptides and Ac-CAAA, as shown in eq 2.5. A semi-quantitative comparison of the  $\Delta_{\text{acid}}\text{H}$  values could be achieved by comparing their branching ratios with Ac-CAAA. Based on the branching ratio values shown in Table 4.7, the  $\Delta_{\text{acid}}\text{H}$  difference between Ac- $^{\text{d}}\text{CAAA}$  and Ac-CAAA is about three times of that between Ac-CAA $^{\text{d}}\text{A}$  and Ac-CAAA. Similarly, the  $\Delta_{\text{acid}}\text{H}$  difference between Ac-C $^{\text{d}}\text{AAA}$  and Ac-CAAA is about 2.5 times of that between Ac-CA $^{\text{d}}\text{AA}$  and Ac-CAAA. The acidity of Ac-CAAA seems to lie in the middle between that of Ac- $^{\text{d}}\text{CAAA}$  and Ac-CA $^{\text{d}}\text{AA}$ . A schematic representation of the relative  $\Delta_{\text{acid}}\text{H}$  values is shown in Figure 4.8.

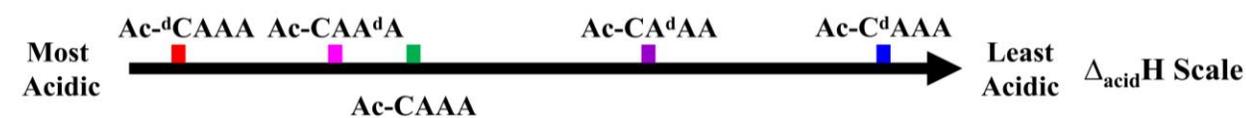


Figure 4.8. Semi-quantitative comparison of the  $\Delta_{\text{acid}}\text{H}$  values of the acetylated peptides.

### Computed Thermochemical Values of the Acetylated Peptides

The thermochemical values of the acetylated peptides were computed through theoretical studies. At the conformational search step, the CREST conformational search was used for both the neutral and deprotonated peptides. In total, 133 to 224 conformations for the neutral and 65 to 210 conformations for the deprotonated peptides were generated. The conformations obtained were optimized at the HF/3-21G\* level of theory and ranked in free energy. Conformations within the energy windows of 5 and 6 kcal/mol for the neutral and deprotonated species were kept for the final step. The resulting 25 to 53

conformations for the neutral and 12 to 23 conformations for the deprotonated peptides were geometry optimized and frequency calculated at the wB97XD/6-311+G\*\* level of theory. After removing the degenerate and high-energy conformations, a final set of ten conformations for each peptide species is used to compute the theoretical thermochemical values based on the equations described in Chapter 2.

The calculated energy values for all ten conformations of each peptide species are listed in tables from Table 4.8 to Table 4.12, which includes the enthalpy/free energy, relative enthalpy/free energy, the Boltzmann probability, and the weighted average of enthalpy/free energy. The ten lowest energy conformations of the neutral and deprotonated peptides are noted as N1 to N10 and A1 to A10, respectively. The conformations are ranked in free energy. The calculated energetics and thermochemical values of the five peptides are summarized in Table 4.13.

Table 4.8  
 Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for Ac-CAAA and Ac-CAAA<sup>-</sup>

Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>	Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>
Ac-CAAA						Ac-CAAA <sup>-</sup>					
N1	-1596.301565	-1596.387648	0.0	0.0	47.4%	A1	-1595.78959	-1595.876339	0.0	0.0	82.6%
N2	-1596.299609	-1596.387466	1.2	0.1	39.1%	A2	-1595.78598	-1595.873629	2.3	1.7	4.7%
N3	-1596.299711	-1596.385641	1.2	1.3	5.6%	A3	-1595.78691	-1595.873201	1.7	2.0	3.0%
N4	-1596.300491	-1596.38521	0.7	1.5	3.6%	A4	-1595.78789	-1595.873075	1.1	2.0	2.6%
N5	-1596.298698	-1596.384087	1.8	2.2	1.1%	A5	-1595.78642	-1595.873035	2.0	2.1	2.5%
N6	-1596.298169	-1596.383975	2.1	2.3	1.0%	A6	-1595.7854	-1595.872881	2.6	2.2	2.1%
N7	-1596.298048	-1596.383924	2.2	2.3	0.9%	A7	-1595.78728	-1595.872727	1.5	2.3	1.8%
N8	-1596.298097	-1596.383879	2.2	2.4	0.9%	A8	-1595.78403	-1595.871172	3.5	3.2	0.3%
N9	-1596.296506	-1596.382908	3.2	3.0	0.3%	A9	-1595.7864	-1595.870822	2.0	3.5	0.2%
N10	-1596.296625	-1596.3825	3.1	3.2	0.2%	A10	-1595.78459	-1595.87042	3.1	3.7	0.2%
H <sub>avg</sub> <sup>c</sup>	-1596.300506					H <sub>avg</sub> <sup>c</sup>	-1595.78906				
G <sub>avg</sub> <sup>c</sup>		-1596.38721				G <sub>avg</sub> <sup>c</sup>		-1595.875774			
Rel. H <sub>avg</sub> <sup>c</sup>	0.0					Rel. H <sub>avg</sub> <sup>c</sup>	0.0				
Rel. G <sub>avg</sub> <sup>c</sup>		0.0				Rel. G <sub>avg</sub> <sup>c</sup>		0.0			
ΔH(EtSH)	-477.9387822					ΔH(EtS-)	-477.374309				
ΔG(EtSH)		-477.9710607				ΔG(EtS-)		-477.405477			

<sup>a</sup> Values obtained from calculations at the wB97XD/6-311+G\*\* level of theory.

<sup>b</sup> Boltzmann probability calculated using the equations described in Chapter 2.

<sup>c</sup> Weighted average of enthalpy and free energy (Hartree) of the ten conformations calculated using the equations described in Chapter 2.

<sup>d</sup> Relative enthalpy and free energy (kcal/mol) between Ac-CAAA and Ac-CAAA<sup>-</sup> and between Ac-CAAA<sup>-</sup> and Ac-CAAA<sup>-</sup> calculated using the weighted average of H<sub>avg</sub> and G<sub>avg</sub>, respectively.

Table 4.9  
 Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for Ac-<sup>d</sup>CAAA and Ac-<sup>d</sup>CAAA<sup>-</sup>

Ac- <sup>d</sup> CAAA		Ac- <sup>d</sup> CAAA <sup>-</sup>					
Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH ΔG P <sub>i</sub> (%) <sup>b</sup>	Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH ΔG P <sub>i</sub> (%) <sup>b</sup>
N1	-1596.301011	-1596.385692	0.0 0.0 32.1%	A1	-1595.7911	-1595.880094	0.0 0.0 84.4%
N2	-1596.301001	-1596.385404	0.0 0.2 23.6%	A2	-1595.79134	-1595.878372	-0.1 1.1 13.6%
N3	-1596.297847	-1596.384673	2.0 0.6 10.9%	A3	-1595.78801	-1595.876189	1.9 2.5 1.3%
N4	-1596.298218	-1596.384365	1.8 0.8 7.9%	A4	-1595.78947	-1595.875402	1.0 2.9 0.6%
N5	-1596.296809	-1596.384063	2.6 1.0 5.7%	A5	-1595.78536	-1595.872742	3.6 4.6 0.0%
N6	-1596.296387	-1596.383921	2.9 1.1 4.9%	A6	-1595.78331	-1595.871719	4.9 5.3 0.0%
N7	-1596.299368	-1596.38387	1.0 1.1 4.6%	A7	-1595.78313	-1595.869252	5.0 6.8 0.0%
N8	-1596.297504	-1596.383682	2.2 1.3 3.8%	A8	-1595.78212	-1595.868447	5.6 7.3 0.0%
N9	-1596.298562	-1596.383668	1.5 1.3 3.8%	A9	-1595.78181	-1595.868018	5.8 7.6 0.0%
N10	-1596.294985	-1596.38337	3.8 1.5 2.7%	A10	-1595.78049	-1595.867068	6.7 8.2 0.0%
H <sub>avg</sub> <sup>c</sup>	-1596.299511			H <sub>avg</sub> <sup>c</sup>	-1595.79108		
G <sub>avg</sub> <sup>c</sup>		-1596.384928		G <sub>avg</sub> <sup>c</sup>		-1595.879776	
Rel. H <sub>avg</sub> <sup>c</sup>	0.6			Rel. H <sub>avg</sub> <sup>c</sup>	-1.3		
Rel. G <sub>avg</sub> <sup>c</sup>		1.4		Rel. G <sub>avg</sub> <sup>c</sup>		-2.5	
ΔH(EtSH)	-477.9387822			ΔH(EtS <sup>-</sup> )	-477.374309		
ΔG(EtSH)		-477.9710607		ΔG(EtS <sup>-</sup> )		-477.405477	

<sup>a</sup> Values obtained from calculations at the wB97XD/6-311+G\*\* level of theory.

<sup>b</sup> Boltzmann probability calculated using the equations described in Chapter 2.

<sup>c</sup> Weighted average of enthalpy and free energy (Hartree) of the ten conformations calculated using the equations described in Chapter 2.

<sup>d</sup> Relative enthalpy and free energy (kcal/mol) between Ac-CAAA and Ac-<sup>d</sup>CAAA, and between Ac-CAAA<sup>-</sup> and Ac-<sup>d</sup>CAAA<sup>-</sup> calculated using the weighted average of H<sub>avg</sub> and G<sub>avg</sub>, respectively.

Table 4.10  
 Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for Ac-C<sup>d</sup>AAA and Ac-C<sup>d</sup>AAA<sup>-</sup>

Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>	Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>
Ac-C <sup>d</sup> AAA						Ac-C <sup>d</sup> AAA <sup>-</sup>					
N1	-1596.30419	-1596.390168	0.0	0.0	46.5%	A1	-1595.78939	-1595.877635	0.0	0.0	90.5%
N2	-1596.303353	-1596.389278	0.5	0.6	18.1%	A2	-1595.78806	-1595.874626	0.8	1.9	3.7%
N3	-1596.302329	-1596.389278	1.2	0.6	18.1%	A3	-1595.7869	-1595.874111	1.6	2.2	2.2%
N4	-1596.30285	-1596.388498	0.8	1.0	7.9%	A4	-1595.78368	-1595.873928	3.6	2.3	1.8%
N5	-1596.301388	-1596.388249	1.8	1.2	6.1%	A5	-1595.78636	-1595.87344	1.9	2.6	1.1%
N6	-1596.299844	-1596.387175	2.7	1.9	1.9%	A6	-1595.78431	-1595.87279	3.2	3.0	0.5%
N7	-1596.299492	-1596.386112	2.9	2.5	0.6%	A7	-1595.78375	-1595.871357	3.5	3.9	0.1%
N8	-1596.297891	-1596.385961	4.0	2.6	0.5%	A8	-1595.78348	-1595.871068	3.7	4.1	0.1%
N9	-1596.297039	-1596.384624	4.5	3.5	0.1%	A9	-1595.78452	-1595.87054	3.1	4.5	0.0%
N10	-1596.296913	-1596.384433	4.6	3.6	0.1%	A10	-1595.78509	-1595.868972	2.7	5.4	0.0%
H <sub>avg</sub> <sup>c</sup>	-1596.30326					H <sub>avg</sub> <sup>c</sup>	-1595.78911				
G <sub>avg</sub> <sup>c</sup>		-1596.389477				G <sub>avg</sub> <sup>c</sup>		-1595.877293			
Rel. H <sub>avg</sub> <sup>c</sup>	-1.7					Rel. H <sub>avg</sub> <sup>c</sup>	0.0				
Rel. G <sub>avg</sub> <sup>c</sup>		-1.4				Rel. G <sub>avg</sub> <sup>c</sup>		-1.0			
ΔH(EtSH)	-477.9387822					ΔH(EtS <sup>-</sup> )	-477.374309				
ΔG(EtSH)		-477.9710607				ΔG(EtS <sup>-</sup> )		-477.405477			

<sup>a</sup> Values obtained from calculations at the wB97XD/6-311+G\*\* level of theory.

<sup>b</sup> Boltzmann probability calculated using the equations described in Chapter 2.

<sup>c</sup> Weighted average of enthalpy and free energy (Hartree) of the ten conformations calculated using the equations described in Chapter 2.

<sup>d</sup> Relative enthalpy and free energy (kcal/mol) between Ac-CAAA and Ac-C<sup>d</sup>AAA, and between Ac-CAAA<sup>-</sup> and Ac-C<sup>d</sup>AAA<sup>-</sup> calculated using the weighted average of H<sub>avg</sub> and G<sub>avg</sub>, respectively.

Table 4.11  
 Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for Ac-CA<sup>d</sup>AA and Ac-CA<sup>d</sup>AA<sup>-</sup>

Ac-CA <sup>d</sup> AA		Ac-CA <sup>d</sup> AA <sup>-</sup>									
Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>	Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>
N1	-1596.300255	-1596.388278	0.0	0.0	36.3%	A1	-1595.7905	-1595.877087	0.0	0.0	60.6%
N2	-1596.303679	-1596.387859	-2.1	0.3	23.3%	A2	-1595.78693	-1595.876377	2.2	0.4	28.5%
N3	-1596.303888	-1596.387857	-2.3	0.3	23.2%	A3	-1595.78815	-1595.875261	1.5	1.1	8.7%
N4	-1596.302323	-1596.386854	-1.3	0.9	8.0%	A4	-1595.78639	-1595.87345	2.6	2.3	1.3%
N5	-1596.298069	-1596.385871	1.4	1.5	2.8%	A5	-1595.78535	-1595.872447	3.2	2.9	0.4%
N6	-1596.30106	-1596.385659	-0.5	1.6	2.3%	A6	-1595.7859	-1595.871915	2.9	3.2	0.3%
N7	-1596.298654	-1596.385469	1.0	1.8	1.8%	A7	-1595.78452	-1595.871274	3.8	3.6	0.1%
N8	-1596.299877	-1596.385226	0.2	1.9	1.4%	A8	-1595.7826	-1595.870409	5.0	4.2	0.1%
N9	-1596.294735	-1596.38415	3.5	2.6	0.5%	A9	-1595.78127	-1595.86907	5.8	5.0	0.0%
N10	-1596.295935	-1596.383958	2.7	2.7	0.4%	A10	-1595.78041	-1595.867328	6.3	6.1	0.0%
H <sub>avg</sub> <sup>c</sup>	-1596.301941					H <sub>avg</sub> <sup>c</sup>	-1595.78918				
G <sub>avg</sub> <sup>c</sup>		-1596.387711				G <sub>avg</sub> <sup>c</sup>		-1595.876632			
Rel. H <sub>avg</sub> <sup>c</sup>	-0.9					Rel. H <sub>avg</sub> <sup>c</sup>	-0.1				
Rel. G <sub>avg</sub> <sup>c</sup>		-0.3				Rel. G <sub>avg</sub> <sup>c</sup>		-0.5			
ΔH(EtSH)	-477.9387822					ΔH(EtSH)	-477.374309				
ΔG(EtSH)		-477.9710607				ΔG(EtSH)		-477.405477			

<sup>a</sup> Values obtained from calculations at the wB97XD/6-311+G\*\* level of theory.

<sup>b</sup> Boltzmann probability calculated using the equations described in Chapter 2.

<sup>c</sup> Weighted average of enthalpy and free energy (Hartree) of the ten conformations calculated using the equations described in Chapter 2.

<sup>d</sup> Relative enthalpy and free energy (kcal/mol) between Ac-CAAA and Ac-CA<sup>d</sup>AA, and between Ac-CAAA<sup>-</sup> and Ac-CA<sup>d</sup>AA<sup>-</sup> calculated using the weighted average of H<sub>avg</sub> and G<sub>avg</sub>, respectively.

Table 4.12  
 Summary of Calculated Energetics and Boltzmann Distributions of the Conformations Obtained for Ac-CAA<sup>d</sup>A and Ac-CAA<sup>d</sup>A<sup>-</sup>

Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>	Conf.	H <sup>a</sup>	G <sup>a</sup>	ΔH	ΔG	P <sub>i</sub> (%) <sup>b</sup>
Ac-CAA <sup>d</sup> A						Ac-CAA <sup>d</sup> A <sup>-</sup>					
N1	-1596.299994	-1596.386951	0.0	0.0	63.4%	A1	-1595.7923	-1595.878181	0.0	0.0	54.5%
N2	-1596.300875	-1596.385472	-0.6	0.9	13.2%	A2	-1595.78757	-1595.877305	3.0	0.5	21.5%
N3	-1596.296583	-1596.384926	2.1	1.3	7.4%	A3	-1595.79099	-1595.877292	0.8	0.6	21.3%
N4	-1596.298185	-1596.38483	1.1	1.3	6.7%	A4	-1595.78987	-1595.874996	1.5	2.0	1.9%
N5	-1596.298182	-1596.384326	1.1	1.6	3.9%	A5	-1595.78801	-1595.874195	2.7	2.5	0.8%
N6	-1596.297924	-1596.383483	1.3	2.2	1.6%	A6	-1595.78237	-1595.869897	6.2	5.2	0.0%
N7	-1596.295084	-1596.383246	3.1	2.3	1.2%	A7	-1595.78363	-1595.869173	5.4	5.7	0.0%
N8	-1596.297142	-1596.382966	1.8	2.5	0.9%	A8	-1595.78168	-1595.868879	6.7	5.8	0.0%
N9	-1596.29646	-1596.382864	2.2	2.6	0.8%	A9	-1595.78083	-1595.868633	7.2	6.0	0.0%
N10	-1596.296777	-1596.382802	2.0	2.6	0.8%	A10	-1595.78144	-1595.868596	6.8	6.0	0.0%
H <sub>avg</sub> <sup>c</sup>	-1596.29949					H <sub>avg</sub> <sup>c</sup>	-1595.79092				
G <sub>avg</sub> <sup>c</sup>		-1596.386155				G <sub>avg</sub> <sup>c</sup>		-1595.87771			
Rel. H <sub>avg</sub> <sup>c</sup>	0.6					Rel. H <sub>avg</sub> <sup>c</sup>	-1.2				
Rel. G <sub>avg</sub> <sup>c</sup>		0.7				Rel. G <sub>avg</sub> <sup>c</sup>		-1.2			
ΔH(EtSH)	-477.9387822					ΔH(EtS <sup>-</sup> )	-477.374309				
ΔG(EtSH)		-477.9710607				ΔG(EtS <sup>-</sup> )		-477.405477			

<sup>a</sup> Values obtained from calculations at the wB97XD/6-311+G\*\* level of theory.

<sup>b</sup> Boltzmann probability calculated using the equations described in Chapter 2.

<sup>c</sup> Weighted average of enthalpy and free energy (Hartree) of the ten conformations calculated using the equations described in Chapter 2.

<sup>d</sup> Relative enthalpy and free energy (kcal/mol) between Ac-CAA and Ac-CAA<sup>d</sup>A, and between Ac-CAAA<sup>-</sup> and Ac-CAA<sup>d</sup>A<sup>-</sup> calculated using the weighted average of H<sub>avg</sub> and G<sub>avg</sub>, respectively

Table 4.13  
*Summary of Theoretical Energetics and Thermochemical Values for the Acetylated Peptides*

	$\Delta_{\text{acidG}}$	$\Delta_{\text{acidH}}$	Neutral		Anion	
			Rel. $G_{\text{avg}}$	Rel. $H_{\text{avg}}$	Rel. $G_{\text{avg}}$	Rel. $H_{\text{avg}}$
Ac-CAAA	314.9	322.4	0.0	0.0	0.0	0.0
Ac- <sup>d</sup> CAAA	311.0	320.5	1.4	0.6	-2.5	-1.3
Ac-C <sup>d</sup> AAA	315.4	324.1	-1.4	-1.7	-1.0	0.0
Ac-CA <sup>d</sup> AA	314.7	323.3	-0.3	-0.9	-0.5	-0.1
Ac-CAA <sup>d</sup> A	313.0	320.6	0.7	0.6	-1.2	-1.2

The deprotonation enthalpy ( $\Delta_{\text{acidH}}$ ) and gas-phase acidity ( $\Delta_{\text{acidG}}$ ) values were calculated using the isodesmic reaction scheme with ethanethiol described in Chapter 2. The relative enthalpy (Rel.  $H_{\text{avg}}$ ) and free energy (Rel.  $G_{\text{avg}}$ ) values were calculated by comparing with the corresponding values of Ac-CAAA.

Based on the computed deprotonation enthalpy ( $\Delta_{\text{acidH}}$ ) values, the acidity order of the peptides from the most acidic to the least acidic is Ac-<sup>d</sup>CAAA > Ac-CAA<sup>d</sup>A > Ac-CAAA > Ac-CA<sup>d</sup>AA > Ac-C<sup>d</sup>AAA, which is in consistent with the experimental observations. Quantitatively, the deprotonation enthalpy values of Ac-<sup>d</sup>CAAA and Ac-CAA<sup>d</sup>A are smaller than that of Ac-CAAA by 1.9 and 1.8 kcal/mol, respectively. Their computed acidity differences with Ac-CAAA are nearly the same, which is different from the semi-quantitative evaluation of their  $\Delta_{\text{acidH}}$  values from the branching ratios, where the  $\Delta_{\text{acidH}}$  value difference between Ac-<sup>d</sup>CAAA and Ac-CAAA is about three times of that between Ac-CAA<sup>d</sup>A and Ac-CAAA. More computational work should be performed to investigate the observed inconsistency. The deprotonation enthalpy ( $\Delta_{\text{acidH}}$ ) values of Ac-CA<sup>d</sup>AA and Ac-C<sup>d</sup>AAA are higher than that of Ac-CAAA by 0.9 and 1.8 kcal/mol, respectively. Their  $\Delta_{\text{acidH}}$  value differences are in agreement with the semi-quantitative evaluations, where the  $\Delta_{\text{acidH}}$  value difference between Ac-C<sup>d</sup>AAA and Ac-CAAA is about 2.5 times of that between Ac-CA<sup>d</sup>AA and Ac-CAAA.

Based on the relative enthalpy (Rel.  $H_{\text{avg}}$ ) and free energy (Rel.  $G_{\text{avg}}$ ) values, both the neutral and deprotonated forms contribute to the observed acidity difference between the peptides. For example, the

Rel.  $H_{\text{avg}}$  values for the neutral and anion form of Ac- $^d$ CAAA are 0.6 and -1.3 kcal/mol, respectively. It means that the neutral Ac- $^d$ CAAA is higher in energy and less stable than the neutral Ac-CAAA, while the Ac- $^d$ CAAA anion is lower in energy and more stable than the Ac-CAAA anion. The energy difference in both neutral and anions contribute notably to the computed higher acidity of 1.9 kcal/mol of Ac- $^d$ CAAA compared to that of Ac-CAAA. For Ac-CA $^d$ AA, the Rel.  $H_{\text{avg}}$  values of the neutral and anion are -0.9 and -0.1 kcal/mol, respectively. So, a larger energy difference is found between the neutral species of Ac-CA $^d$ AA and Ac-CAAA, while their anions have nearly the same energy.

In summary, direct acidity comparisons were achieved through the fragmentation experiments of the proton-bound dimers with the reference peptide Ac-CAAA. The obtained acidity ladder from the most acidic to the least acidic is Ac- $^d$ CAAA > Ac-CAA $^d$ A > Ac-CAAA > Ac-CA $^d$ AA > Ac-C $^d$ AAA. The two parts in the proton-bound dimers between the peptides and Ac-CAAA are structurally similar and their entropy changes during the dissociation process are similar. The branching ratios of the D-amino acid-containing peptides with Ac-CAAA mainly depend on their deprotonation enthalpy values and approach zero as the collision energy increases. The computed acidity trend is the same as that from experiments, but discrepancies were observed when compared with the semi-quantitative values. Further investigations are needed to fully understand the experimental observations.

### Summary

Among the non-acetylated peptide, an order of acidity was established from the most acidic to the least acidic as: CAA $^d$ A > CA $^d$ AA ~ CAAA >  $^d$ CAAA > C $^d$ AAA. The trend was achieved from the product ion analysis of all peptides with a single reference acid at one collision energy. When multiple references acids and different collision energies were used for the product ion analysis, the same trend was achieved.

The branching ratio plots of the peptides with the reference acids suggest different entropy effects among the reference acids. The branching ratios of the peptides with CAAA were determined by taking the difference of their branching ratios with the same reference acids. The branching ratio plots of the

peptides with CAAA approach the zero axis as collision energy is increased. This suggests the similarity in the entropy term.

Direct acidity comparison between the acetylated peptides was achieved, and the following acidity order was established from the most acidic to the least acidic,  $\text{Ac-}^d\text{CAAA} > \text{Ac-CAA}^d\text{A} > \text{Ac-CAAA} > \text{Ac-CA}^d\text{AA} > \text{Ac-C}^d\text{AAA}$ . In the dissociation of the proton-bound dimers between the peptides, the two parts are structurally similar, and the entropy differences between the two fragmentation channels are negligible.

## CHAPTER 5 : SUMMARY AND CONCLUSIONS

In this thesis, a series of model oligopeptides containing an acidic or a basic probe residue were characterized via mass spectrometry, ion spectroscopy, and computational studies. In the model peptides, Lys and Dap were used as the basic probe, and Cys was used as the acidic probe.

The peptides with the basic probe residue on the C-terminus have significantly higher proton affinity (PA) and gas phase basicity (GB) than their isomeric peptides with the probe residue on the N-terminus. The observed gas-phase basicity difference between the dipeptides LysA and ALys is as large as 3 kcal/mol, which is significant considering their small sizes. The conformations of the protonated peptides, LysAH<sup>+</sup> and ALysH<sup>+</sup>, were characterized using the IRMPD ion spectroscopy in conjunction with computation.

The IRMPD spectrum of LysAH<sup>+</sup> has broad and overlapped peaks in the regions of carboxyl C=O stretching and amide I and II. While the IRMPD spectrum of ALysH<sup>+</sup> has well resolved and narrow peaks. Based on their characteristically different patterns in their IRMPD spectra, LysAH<sup>+</sup> was predicted to exist in multiple conformations with much different geometries, while ALysH<sup>+</sup> adopts structurally similar conformations. Through extensive conformational search at semiempirical quantum chemical (SQM) levels of theory and geometry optimizations at DFT level of theory, lowest energy conformations along with energies and theoretical infrared (IR) spectra were obtained. The stability difference in the protonated peptides was found to be the main source of their basicity differences. The structurally representative conformations for LysAH<sup>+</sup> and ALysH<sup>+</sup> were found by matching the theoretical IR spectra to their experimental IRMPD spectra. Two structurally distinct conformations were selected to represent the structure for LysAH<sup>+</sup>. Among which one appears to be a left-handed helix and the other one adopts an extended linear shape. Two structurally similar left-handed helical conformations were selected to represent the structure for ALysH<sup>+</sup>. Energetic and property characterizations, including charge analysis, dipole moment decomposition, torsional strain analysis, hydrogen bonding analysis, and the reduced density gradient (RDG) analysis, were applied to the selected conformations to reveal the nature of their

relative stabilities. Although the conserved charge on the  $\text{NH}_3^+$  group and the interaction of the backbone dipole with  $\text{NH}_3^+$  play important roles, the torsional strain on the lysine sidechain and the hydrogen bonding interactions seems to have the most contributions to the observed relative stability between conformations of  $\text{LysAH}^+$  and  $\text{ALysH}^+$ .

The isomeric tripeptide pair DapAA and AADap were characterized from the structural and energetic point of view using IRMPD and computational chemistry. From the mass spectrometry measurements, DapAA was found to be less basic than AADap by 2.0 kcal/mol. The IRMPD spectrum of  $\text{DapAAH}^+$  are characteristically more complex, indicating the existence of multiple conformations with different hydrogen-bonding networks. While the IRMPD spectrum of  $\text{AADapH}^+$  has well resolved sharp peaks, and suggests structurally similar conformations. The results from computational studies were found to be in good agreement with the experimental proton affinities. The stability difference between the protonated peptides is the main source of the experimental observed difference in basicities between DapAA and AADap. Three coiled conformations with similar backbone geometries, but different hydrogen bonding networks were selected as the representative conformations for  $\text{DapAAH}^+$ . Two structurally similar right-handed helical conformations were selected as the representative conformations for  $\text{AADapH}^+$ . The concatenating hydrogen bonding networks in the helical conformations provide similar microenvironments for the amide bonds and contributes to the observed narrow IRMPD bands.

In the  $\text{DapAAH}^+$  conformations, the positive charge is mainly stabilized through strong hydrogen bonds with the  $\text{NH}_3^+$  group, and the residual charge on the  $\text{NH}_3^+$  group is small. Contrarily, in the  $\text{AADapH}^+$  conformations, the hydrogen bonds with the  $\text{NH}_3^+$  group are much weaker, and the residual charge is much larger. The interaction of the backbone dipole moment with the  $\text{NH}_3^+$  group is very strong and plays an important role in enhancing the stability of the  $\text{AADapH}^+$  conformations.

IR decomposition of the selected representative conformations of the four peptides shows that their differences in preferred geometry and hydrogen-bonding networks are well-reflected in their characteristically different IRMPD spectra patterns. The N-probe peptides adopt diverse backbone

structures and hydrogen bonding networks, while the C-probe peptides adopt helical conformations with similar microenvironments for the amide bonds.

Two series of Cys-containing poly-alanine peptides were used to characterize the effect of chirality on the gas-phase acidity of peptides by systematically alternating the chirality of each residue in the sequence. For the non-acetylated peptides, the acidity comparison was firstly achieved through product ion analysis of the proton-bound dimer of the peptides with one reference acid at one collision energy. This was followed by a more extensive characterization using six reference acids at ten different collision energies. The following acidity order was achieved, from the most acidic to the least acidic:  $CAA^dA > CA^dAA \sim CAAA > ^dC AAA > C^dAAA$ . The product ion branching ratios of the peptides with individual reference acids were used to determine the branching ratios of two peptide ions. This allows a direct comparison of the gas-phase acidities of the isomeric peptides with the reference peptide CAAA. The entropy term is negligible during the dissociation of the proton-bound dimers of the peptides with CAAA. This is suggested by the appearance of their branching ratio plots. Among the D-alanine containing peptides, acidity suppression is observed after chiral inversion of the alanine closest to the cysteine residue, while acidity enhancement is observed after chiral inversion of the alanine furthest from the cysteine residue.

For acetylated peptides, deuterium labeling on the acetyl group allows direct comparison of the gas-phase acidity between the peptides. Through the fragmentation experiments of the proton-bound dimers containing an analyte peptide and the reference peptide Ac-CAAA, the gas-phase acidity ladder was obtained as  $Ac-^dC AAA > Ac-CAA^dA > Ac-CAAA > Ac-CA^dAA > Ac-C^dAAA$ . The two parts in the proton-bound dimers are structurally similar and their entropy changes during the dissociation process are similar. The branching ratios of the D-amino acid-containing peptides with Ac-CAAA mainly depend on their deprotonation enthalpies and the branching ratio plots approach to zero as the collision energy increases. The computed acidity trend agrees with the experimental observations in general. However, discrepancies were observed when compared with the semi-quantitative values. Further investigations are needed to fully understand the experimental observations.

## REFERENCES

- Al Temimi, A. H. K., Amadajais-Groenen, H. I. V., Reddy, Y. V., Blaauw, R. H., Guo, H., Qian, P., & Mecinović, J. (2019). The nucleophilic amino group of lysine is central for histone lysine methyltransferase catalysis. *Communications Chemistry* 2019 2:1, 2(1), 1–14.  
<https://doi.org/10.1038/s42004-019-0210-8>
- Alkorta, I., Rozas, I., & Elguero, J. (1998). Non-conventional hydrogen bonds. *Chemical Society Reviews*, 27(2), 163–170. <https://doi.org/10.1039/a827163z>
- Arcus, V. L., van der Kamp, M. W., Pudney, C. R., & Mulholland, A. J. (2020). Enzyme evolution and the temperature dependence of enzyme catalysis. *Current Opinion in Structural Biology*, 65, 96–101.  
<https://doi.org/10.1016/j.sbi.2020.06.001>
- Armentrout, P. B. (2000). Entropy measurements and the kinetic method: A statistically meaningful approach. *Journal of the American Society for Mass Spectrometry*, 11(5), 371–379.  
[https://doi.org/10.1016/S1044-0305\(00\)00102-1](https://doi.org/10.1016/S1044-0305(00)00102-1)
- Arnér, E. S. J., & Holmgren, A. (2000). Physiological functions of thioredoxin and thioredoxin reductase. *European Journal of Biochemistry*, 267(20), 6102–6109. <https://doi.org/10.1046/J.1432-1327.2000.01701.X>
- Ballmoos, C. von, Wiedenmann, A., & Dimroth, P. (2009). Essentials for ATP Synthesis by F1F0 ATP Synthases. [Http://Dx.Doi.Org/10.1146/Annurev.Biochem.78.081307.104803](http://Dx.Doi.Org/10.1146/Annurev.Biochem.78.081307.104803), 78, 649–672.  
<https://doi.org/10.1146/ANNUREV.BIOCHEM.78.081307.104803>
- Baranović, G., Biliškov, N., & Vojta, D. (2012). Characterization of intramolecular hydrogen bonds by atomic charges and charge fluxes. *Journal of Physical Chemistry A*, 116(32).  
<https://doi.org/10.1021/jp306070x>
- Bartlett, G. J., Porter, C. T., Borkakoti, N., & Thornton, J. M. (2002). Analysis of Catalytic Residues in Enzyme Active Sites. *Journal of Molecular Biology*, 324(1), 105–121.  
[https://doi.org/10.1016/S0022-2836\(02\)01036-7](https://doi.org/10.1016/S0022-2836(02)01036-7)

- Batoon, P. (2016). *Thermochemical Differences in Lysine and Lysine-Homolog Containing Oligopeptides: Determination of Basicity and Gas-Phase Structure Through Mass Spectrometry, Infrared Spectroscopy, and Computational Chemistry*
- Batoon, P., Oomens, J., Berden, G., & Ren, J. (2018). Conformations of Protonated AlaDap and DapAla Characterized by IRMPD Spectroscopy and Molecular Modeling. *Journal of Physical Chemistry B*, 122(8), 2191–2202. <https://doi.org/10.1021/acs.jpcc.7b10435>
- Batoon, P., & Ren, J. (2015). Proton affinity of dipeptides containing alanine and diaminobutyric acid. *International Journal of Mass Spectrometry*, 378, 151–159. <https://doi.org/10.1016/j.ijms.2014.07.025>
- Batoon, P., & Ren, J. (2016). Proton Affinity of Isomeric Dipeptides Containing Lysine and Non-Proteinogenic Lysine Homologues. *Journal of Physical Chemistry B*, 120(32), 7783–7794. <https://doi.org/10.1021/ACS.JPCB.6B03776>
- Beasley, F. C., Cheung, J., & Heinrichs, D. E. (2011). Mutation of L-2,3-diaminopropionic acid synthase genes blocks staphyloferrin B synthesis in *Staphylococcus aureus*. *BMC Microbiology* 2011 11:1, 11(1), 1–12. <https://doi.org/10.1186/1471-2180-11-199>
- Bell, E. A., Tirimanna, A. S. L., EA, B., & AS, T. (1965). Associations of amino acids and related compounds in the seeds of forty-seven species of *Vicia*: their taxonomic and nutritional significance. *Biochemical Journal*, 97(1), 104. <https://doi.org/10.1042/BJ0970104>
- Bernardi, R. C., Melo, M. C. R., & Schulten, K. (2015). Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochimica et Biophysica Acta*, 1850(5), 872. <https://doi.org/10.1016/J.BBAGEN.2014.10.019>
- Bertolasi, V., Gilli, P., & Gilli, G. (2011). Hydrogen Bonding and Electron Donor–Acceptor (EDA) Interactions Controlling the Crystal Packing of Picric Acid and Its Adducts with Nitrogen Bases. Their Rationalization in Terms of the pKa Equalization and Electron-Pair Saturation Concepts. *Crystal Growth and Design*, 11(7), 2724–2735. <https://doi.org/10.1021/CG101007A>

- Bosshard, H. R., Marti, D. N., & Jelesarov, I. (2004). Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *Journal of Molecular Recognition*, *17*(1), 1–16. <https://doi.org/10.1002/JMR.657>
- Brandhorst, K., & Grunenberg, J. (2008). How strong is it? The interpretation of force and compliance constants as bond strength descriptors. *Chemical Society Reviews*, *37*(8), 1558–1567. <https://doi.org/10.1039/B717781J>
- Brandhorst, K., & Grunenberg, J. (2010). Efficient computation of compliance matrices in redundant internal coordinates from Cartesian Hessians for nonstationary points. *Journal of Chemical Physics*, *132*(18), 184101. <https://doi.org/10.1063/1.3413528>
- Carter, J. H., Bus, R. H. Du, Dyer, J. R., Floyd, J. C., Rice, K. C., & Shaw, P. D. (2002). Biosynthesis of viomycin. I. Origin of  $\alpha\beta$ -diaminopropionic acid and serine. *Biochemistry*, *13*(6), 1221–1227. <https://doi.org/10.1021/BI00703A026>
- Cava, F., Lam, H., Pedro, M. A. de, Waldor, M. K., De Pedro, M. A., & Waldor, M. K. (2011). Emerging knowledge of regulatory roles of d-amino acids in bacteria. *Cellular and Molecular Life Sciences*, *68*(5), 817–831. <https://doi.org/10.1007/s00018-010-0571-8>
- Cerda, B. A., & Wesdemiotis, C. (1996). Li<sup>+</sup>, Na<sup>+</sup>, and K<sup>+</sup> binding to the DNA and RNA nucleobases. Bond energies and attachment sites from the dissociation of metal ion-bound heterodimers. *Journal of the American Chemical Society*, *118*(47), 11884–11892. <https://doi.org/10.1021/ja9613421>
- Cheng, X., Wu, Z., & Fenselau, C. (1993). Collision Energy Dependence of Proton-Bound Dimer Dissociation: Entropy Effects, Proton Affinities, and Intramolecular Hydrogen-Bonding in Protonated Peptides. *Journal of the American Chemical Society*, *115*(11), 4844–4848. <https://doi.org/10.1021/ja00064a052>
- Cooks, R. G., & Kruger, T. L. (1977). Intrinsic Basicity Determination Using Metastable Ions. *Journal of the American Chemical Society*, *99*(4), 1279–1281. <https://doi.org/10.1021/ja00446a059>

- Cooks, R. Graham, & Wong, P. S. H. H. (1998). Kinetic Method of Making Thermochemical Determinations: Advances and Applications. *Accounts of Chemical Research*, 31(7), 379–386. <https://doi.org/10.1021/ar960242x>
- Cooks, Robert Graham, Koskinen, J. T., & Thomas, P. D. (1999). The kinetic method of making thermochemical determinations. *Journal of Mass Spectrometry*, 34, 85–92.
- Coutsias, E. A., Seok, C., & Dill, K. A. (2004). Using quaternions to calculate RMSD. *Journal of Computational Chemistry*, 25(15), 1849–1857. <https://doi.org/10.1002/jcc.20110>
- Emamian, S., Lu, T., Kruse, H., & Emamian, H. (2019). Exploring Nature and Predicting Strength of Hydrogen Bonds: A Correlation Analysis Between Atoms-in-Molecules Descriptors, Binding Energies, and Energy Components of Symmetry-Adapted Perturbation Theory. *Journal of Computational Chemistry*, 40(32), 2868–2881. <https://doi.org/10.1002/JCC.26068>
- Erspamer, V., Melchiorri, P., Falconieri-Erspamer, G., Negri, L., Corsi, R., Severini, C., ... Kreil, G. (1989). Deltorphins: a family of naturally occurring peptides with high affinity and selectivity for delta opioid binding sites. *Proceedings of the National Academy of Sciences of the United States of America*, 86(13), 5188. <https://doi.org/10.1073/PNAS.86.13.5188>
- Ervin, K. M. (2000). Microcanonical analysis of the kinetic method. The meaning of the “effective temperature.” *International Journal of Mass Spectrometry*, 195–196, 271–284. [https://doi.org/10.1016/S1387-3806\(99\)00176-1](https://doi.org/10.1016/S1387-3806(99)00176-1)
- Ervin, K. M., & Armentrout, P. B. (2004). Systematic and random errors in ion affinities and activation entropies from the extended kinetic method. *Journal of Mass Spectrometry*, 39(9), 1004–1015. <https://doi.org/10.1002/jms.682>
- Eyler, J. R. (2009). Infrared multiple photon dissociation spectroscopy of ions in penning traps. *Mass Spectrometry Reviews*, 28(3), 448–467. <https://doi.org/10.1002/mas.20217>
- Fedick, P. W., Bain, R. M., Bain, K., & Cooks, R. G. (2017). Chiral Analysis by Tandem Mass Spectrometry Using the Kinetic Method, by Polarimetry, and by <sup>1</sup>H NMR Spectroscopy. *Journal of Chemical Education*, 94(9), 1329–1333. <https://doi.org/10.1021/ACS.JCHEMED.7B00090>

- Fiori, A., Cardelli, P., Negri, L., Savi, M. R., Strom, R., & Erspamer, V. (1997). Deltorphin transport across the blood-brain barrier. *Proceedings of the National Academy of Sciences*, *94*(17), 9469–9474. <https://doi.org/10.1073/pnas.94.17.9469>
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., ... others. (2016). *Gaussian 16*. Gaussian, Inc. Wallingford, CT.
- Fuchs, S. A., Berger, R., Klomp, L. W. J., & De Koning, T. J. (2005). D-Amino acids in the central nervous system in health and disease. *Molecular Genetics and Metabolism*, *85*(3), 168–180. <https://doi.org/10.1016/j.ymgme.2005.03.003>
- Genchi, G. (2017). An overview on d-amino acids. *Amino Acids*, *49*(9), 1521–1533. <https://doi.org/10.1007/s00726-017-2459-5>
- Gilbert, H. F. (1990). Molecular and cellular aspects of thiol-disulfide exchange. *Advances in Enzymology and Related Areas of Molecular Biology*, *63*, 69–172. <https://doi.org/10.1002/9780470123096.ch2>
- Goldstein, G., Scheid, M., Hammerling, U., Schlesinger, D. H., Niall, H. D., & Boyse, E. A. (1975). Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *Proceedings of the National Academy of Sciences of the United States of America*, *72*(1), 11. <https://doi.org/10.1073/PNAS.72.1.11>
- Graham Cooks, R., Patrick, J. S., Kotiaho, T., McLuckey, S. A., Cooks, R. G., Patrick, J. S., ... McLuckey, S. A. (1994). Thermochemical determinations by the kinetic method. *Mass Spectrometry Reviews*, *13*(4), 287–339. <https://doi.org/10.1002/mas.1280130402>
- Grimme, S. (2019). Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *Journal of Chemical Theory and Computation*, *15*(5), 2847–2862. Retrieved from <https://pubs.acs.org/doi/full/10.1021/acs.jctc.9b00143>
- Gulick, A. M., Starai, V. J., Horswill, A. R., Homick, K. M., & Escalante-Semerena, J. C. (2003). The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A. *Biochemistry*, *42*(10), 2866–2873. <https://doi.org/10.1021/bi0271603>

- Gutteridge, A., & Thornton, J. M. (2005). Understanding nature's catalytic toolkit. *Trends in Biochemical Sciences*, 30(11), 622–629. <https://doi.org/10.1016/J.TIBS.2005.09.006>
- Harris, T. K., & Turner, G. J. (2002). Structural Basis of Perturbed pKa Values of Catalytic Groups in Enzyme Active Sites. *IUBMB Life*, 53(2), 85–98. <https://doi.org/10.1080/15216540211468>
- Hein, J. E., & Blackmond, D. G. (2012). On the origin of single chirality of amino acids and sugars in biogenesis. *Accounts of Chemical Research*, 45(12), 2045–2054. <https://doi.org/10.1021/ar200316n>
- Helfman, P. M., & Bada, J. L. (1975). Aspartic acid racemization in tooth enamel from living humans. *Proceedings of the National Academy of Sciences of the United States of America*, 72(8), 2891–2894. <https://doi.org/10.1073/pnas.72.8.2891>
- Hirshfeld, F. L. (1977). Bonded-atom fragments for describing molecular charge densities. *Theoretica Chimica Acta* 1977 44:2, 44(2), 129–138. <https://doi.org/10.1007/BF00549096>
- Ho, M.-C., Ménétret, J.-F., Tsuruta, H., & Allen, K. N. (2009). The origin of the electrostatic perturbation in acetoacetate decarboxylase. *Nature* 2009 459:7245, 459(7245), 393–397. <https://doi.org/10.1038/nature07938>
- Holmgren, A. (2003). THIOREDOXIN. <https://doi.org/10.1146/Annurev.Bi.54.070185.001321>, 54(1), 237–271. <https://doi.org/10.1146/ANNUREV.BI.54.070185.001321>
- Horse Racing Discovers New Drug Problem, One Linked to Frogs - The New York Times. (n.d.). Retrieved October 13, 2021, from <https://www.nytimes.com/2012/06/20/sports/horse-racing-discovers-new-drug-problem-one-linked-to-frogs.html>
- Hunter, E. P. L., & Lias, S. G. (2009). Evaluated Gas Phase Basicities and Proton Affinities of Molecules: An Update. *Journal of Physical and Chemical Reference Data*, 27(3), 413. <https://doi.org/10.1063/1.556018>
- Isom, D. G., Castañeda, C. A., Cannon, B. R., E., B. G.-M., Castañeda, C. A., Cannon, B. R., ... García-Moreno, B. E. (2011). Large shifts in pKa values of lysine residues buried inside a protein. *Proceedings of the National Academy of Sciences of the United States of America*, 108(13), 5260–5265. <https://doi.org/10.1073/pnas.1010750108>

- Isom, D. G., Castañeda, C. A., Cannon, B. R., Velu, P. D., & García-Moreno E, B. (2010). Charges in the hydrophobic interior of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(37), 16096–16100. <https://doi.org/10.1073/pnas.1004213107>
- Janousek, B. K., Reed, K. J., & Brauman, J. I. (1980). Electron photodetachment from mercaptyl anions (RS<sup>-</sup>). Electron affinities of mercaptyl radicals and the sulfur-hydrogen bond strength in mercaptans. *Journal of the American Chemical Society*, *102*(9), 3125–3129.
- Johnson, E. R., Keinan, S., Mori-Sánchez, P., Contreras-García, J., Cohen, A. J., & Yang, W. (2010). Revealing noncovalent interactions. *Journal of the American Chemical Society*, *132*(18), 6498–6506. <https://doi.org/10.1021/ja100936w>
- Julie, D. F. K., Clore, G. M., & Angala, M. G. (1992). Relationship Between Electrostatics and Redox Function in Human Thioredoxin: Characterization of pH Titration Shifts Using Two-Dimensional Homo- and Heteronuclear NMR. *Biochemistry*, *31*(13), 3442–3452. <https://doi.org/10.1021/bi00128a019>
- Katta, V., Chait, B. T., & Carr, S. (1991). Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry. *Rapid Communications in Mass Spectrometry*, *5*(4), 214–217. <https://doi.org/10.1002/rcm.1290050415>
- Katti, S. K., LeMaster, D. M., & Eklund, H. (1990). Crystal structure of thioredoxin from *Escherichia coli* at 1.68 Å resolution. *Journal of Molecular Biology*, *212*(1), 167–184. [https://doi.org/10.1016/0022-2836\(90\)90313-B](https://doi.org/10.1016/0022-2836(90)90313-B)
- Kennedy, D., & Norman, C. (2005). What don't we know? *Science*, *309*(5731), 75. <https://doi.org/10.1126/SCIENCE.309.5731.75>
- Kim, P. M., Duan, X., Huang, A. S., Liu, C. Y., Ming, G., Song, H., & Snyder, S. H. (2010). Aspartate racemase, generating neuronal D-aspartate, regulates adult neurogenesis. *Proceedings of the National Academy of Sciences*, *107*(7), 3175–3179. <https://doi.org/10.1073/PNAS.0914706107>
- Kreil, G., Barra, D., Simmaco, M., Erspamer, V., Falconieri Erspamer, G., Negri, L., ... Melchiorri, P. (1989). Deltorphin, a novel amphibian skin peptide with high selectivity and affinity for  $\delta$  opioid

- receptors. *European Journal of Pharmacology*, 162(1), 123–128. [https://doi.org/10.1016/0014-2999\(89\)90611-0](https://doi.org/10.1016/0014-2999(89)90611-0)
- Kumar, J. K., Tabor, S., & Richardson, C. C. (2004). Proteomic analysis of thioredoxin-targeted proteins in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 101(11), 3759–3764. <https://doi.org/10.1073/PNAS.0308701101>
- Kurzban, G. P., & Wang, K. (1988). Giant polypeptides of skeletal muscle titin: Sedimentation equilibrium in guanidine hydrochloride. *Biochemical and Biophysical Research Communications*, 150(3), 1155–1161. [https://doi.org/10.1016/0006-291X\(88\)90750-4](https://doi.org/10.1016/0006-291X(88)90750-4)
- Laio, A., & Parrinello, M. (2002). Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20), 12562–12566. <https://doi.org/10.1073/PNAS.202427399>
- Li, G., DeLaney, K., & Li, L. (2019). Molecular basis for chirality-regulated A $\beta$  self-assembly and receptor recognition revealed by ion mobility-mass spectrometry. *Nature Communications*, 10(1). <https://doi.org/10.1038/s41467-019-12346-8>
- Li, H., Anuwongcharoen, N., Malik, A., Prachayasittikul, V., Wikberg, J., Nantasenamat, C., ... Nantasenamat, C. (2016). Roles of d-Amino Acids on the Bioactivity of Host Defense Peptides. *International Journal of Molecular Sciences*, 17(7), 1023. <https://doi.org/10.3390/ijms17071023>
- Li, W., & Ye, Y. (2008). Polyubiquitin chains: functions, structures, and mechanisms. *Cellular and Molecular Life Sciences : CMLS*, 65(15), 2397. <https://doi.org/10.1007/S00018-008-8090-6>
- Linstrom, P. J. (2005). NIST chemistry webbook. *Http://Webbook. Nist. Gov.*
- Liu, Z., Wang, Y., Gao, T., Pan, Z., Cheng, H., Yang, Q., ... Xue, Y. (2014). CPLM: a database of protein lysine modifications. *Nucleic Acids Research*, 42(Database issue), D531. <https://doi.org/10.1093/NAR/GKT1093>
- Lopez, M. J., & Mohiuddin, S. S. (2020). Biochemistry, Essential Amino Acids. In *StatPearls*. StatPearls Publishing. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/32496725>
- Lu, T. (2020). *Molclus Program*. Retrieved from <http://www.keinsci.com/research/molclus.html>

- Lu, T., & Chen, F. (2012). Multiwfn: A multifunctional wavefunction analyzer. *Journal of Computational Chemistry*, 33(5), 580–592. <https://doi.org/10.1002/JCC.22885>
- Luo, M. (2018). Chemical and Biochemical Perspectives of Protein Lysine Methylation. *Chemical Reviews*, 118(14), 6656. <https://doi.org/10.1021/ACS.CHEMREV.8B00008>
- Malevanets, A., Chong, P. A., Hansen, D. F., Rizk, P., Sun, Y., Lin, H., ... Wodak, S. J. (2017). Interplay of buried histidine protonation and protein stability in prion misfolding. *Scientific Reports* 2017 7:1, 7(1), 1–14. <https://doi.org/10.1038/s41598-017-00954-7>
- Marcoux, J., & Robinson, C. V. (2013). Twenty years of gas phase structural biology. *Structure*, 21(9), 1541–1550. <https://doi.org/10.1016/j.str.2013.08.002>
- Marino, S. M., & Gladyshev, V. N. (2010). Cysteine Function Governs Its Conservation and Degeneration and Restricts Its Utilization on Protein Surfaces. *Journal of Molecular Biology*, 404(5), 902–916. <https://doi.org/10.1016/j.jmb.2010.09.027>
- Marino, S. M., & Gladyshev, V. N. (2012). Analysis and functional prediction of reactive cysteine residues. *Journal of Biological Chemistry*, 287(7), 4419–4425. <https://doi.org/10.1074/JBC.R111.275578>
- Martens, J. K., Grzetic, J., Berden, G., & Oomens, J. (2015). Gas-phase conformations of small polyprolines and their fragment ions by IRMPD spectroscopy. *International Journal of Mass Spectrometry*, 377(1), 179–187. <https://doi.org/10.1016/J.IJMS.2014.07.027>
- Maruyama, K., Kimura, S., Ohashi, K., & Kuwano, Y. (1981). Connectin, an elastic protein of muscle. identification of “titin” with connectin. *Journal of Biochemistry*, 89(3), 701–709. <https://doi.org/10.1093/oxfordjournals.jbchem.a133249>
- Massilia, G. R., Schininà, M. E., Ascenzi, P., & Polticelli, F. (2001). Contryphan-Vn: A novel peptide from the venom of the Mediterranean snail *Conus ventricosus*. *Biochemical and Biophysical Research Communications*, 288(4), 908–913. <https://doi.org/10.1006/bbrc.2001.5833>

- Matsson, O., & Westaway, K. C. (1999). Secondary Deuterium Kinetic Isotope Effects and Transition State Structure. *Advances in Physical Organic Chemistry*, 31(C), 143–248.  
[https://doi.org/10.1016/S0065-3160\(08\)60194-X](https://doi.org/10.1016/S0065-3160(08)60194-X)
- McFadden, P. N., & Clarke, S. (1982). Methylation at D-aspartyl residues in erythrocytes: possible step in the repair of aged membrane proteins. *Proceedings of the National Academy of Sciences*, 79(8), 2460–2464. <https://doi.org/10.1073/PNAS.79.8.2460>
- Melchiorri, P., & Negri, L. (1996). *The dermorphin peptide family*. 27(7), 1099–1107.  
[https://doi.org/10.1016/0306-3623\(95\)02149-3](https://doi.org/10.1016/0306-3623(95)02149-3)
- Merrifield, R. B. (1963). Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society*, 85(14), 2149–2154. <https://doi.org/10.1021/ja00897a025>
- Mitchell, A. L., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., ... Finn, R. D. (2020). MGnify: The microbiome analysis resource in 2020. *Nucleic Acids Research*, 48(D1), D570–D578. <https://doi.org/10.1093/nar/gkz1035>
- Mor, A., Delfour, A., Sagan, S., Amiche, M., Pradelles, P., Rossier, J., & Nicolas, P. (1989). Isolation of dermenkephalin from amphibian skin, a high-affinity ( $\delta$ -selective opioid heptapeptide containing a D-amino acid residue. *FEBS Letters*, 255(2), 269–274. [https://doi.org/10.1016/0014-5793\(89\)81104-4](https://doi.org/10.1016/0014-5793(89)81104-4)
- Mössner, E., Iwai, H., & Glockshuber, R. (2000). Influence of the pK(a) value of the buried, active-site cysteine on the redox properties of thioredoxin-like oxidoreductases. *FEBS Letters*, 477(1–2), 21–26. [https://doi.org/10.1016/S0014-5793\(00\)01738-5](https://doi.org/10.1016/S0014-5793(00)01738-5)
- Mukhopadhyay, D., & Riezman, H. (2007). Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science*, 315(5809), 201–205. <https://doi.org/10.1126/SCIENCE.1127085>
- Myshakina, N. S., Ahmed, Z., & Asher, S. A. (2008). Dependence of Amide Vibrations on Hydrogen Bonding. *The Journal of Physical Chemistry. B*, 112(38), 11873. <https://doi.org/10.1021/JP8057355>

- Nourse, B. D., & Graham Cooks, R. (1991). Proton affinity determinations using the kinetic method in an ion trap mass spectrometer. *International Journal of Mass Spectrometry and Ion Processes*, 106(C), 249–272. [https://doi.org/10.1016/0168-1176\(91\)85022-E](https://doi.org/10.1016/0168-1176(91)85022-E)
- Ohtani, S., Yamamoto, T., Matsushima, Y., & Kobayashi, Y. (1998). Changes in the amount of D-aspartic acid in the human femur with age. *Growth, Development, and Aging : GDA*, 62(4), 141–148. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10219704>
- Ollivaux, C., Soyeze, D., & Toullec, J.-Y. (2014). Biogenesis of <sc>d</sc>-amino acid containing peptides/proteins: where, when and how? *Journal of Peptide Science*, 20(8), 595–612. <https://doi.org/10.1002/psc.2637>
- Oommachen, S., Ren, J., & McCallum, C. M. (2008). Stabilizing helical polyalanine peptides with negative polarity or charge: Capping with cysteine. *Journal of Physical Chemistry B*, 112(18), 5702–5709. <https://doi.org/10.1021/jp073315a>
- Polfer, N. C. (2011). Infrared multiple photon dissociation spectroscopy of trapped ions. *Chemical Society Reviews*, 40(5), 2211–2221. <https://doi.org/10.1039/c0cs00171f>
- Polfer, N. C., & Oomens, J. (2009). Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance. *Mass Spectrometry Reviews*, 28(3), 468–494. <https://doi.org/10.1002/MAS.20215>
- Powell, J. T., Vine, N., Crossman, M., JT, P., N, V., & M, C. (1992). On the accumulation of D-aspartate in elastin and other proteins of the ageing aorta. *Atherosclerosis*, 97(2–3), 201–208. [https://doi.org/10.1016/0021-9150\(92\)90132-Z](https://doi.org/10.1016/0021-9150(92)90132-Z)
- Pracht, P., Bohle, F., & Grimme, S. (2020). Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Physical Chemistry Chemical Physics*, 22(14), 7169–7192. <https://doi.org/10.1039/c9cp06869d>
- Rao, S. L. N. (1978). A sensitive and specific colorimetric method for the determination of  $\alpha,\beta$ -diaminopropionic acid and the *Lathyrus sativus* neurotoxin. *Analytical Biochemistry*, 86(2), 386–395. [https://doi.org/10.1016/0003-2697\(78\)90762-5](https://doi.org/10.1016/0003-2697(78)90762-5)

- Rastogi, V. K., & Girvin, M. E. (1999). Structural changes linked to proton translocation by subunit c of the ATP synthase. *Nature* 1999 402:6759, 402(6759), 263–268. <https://doi.org/10.1038/46224>
- Ren, J., Tan, J. P., & Harper, R. T. (2009). Gas-Phase Acidities of Cysteine-Polyalanine Peptides I: A3,4CSH and HSCA3,4. *Journal of Physical Chemistry A*, 113(41), 10903–10912. <https://doi.org/10.1021/JP903594A>
- Ritz-Timme, S., Laumeier, I., & Collins, M. (2003). Age estimation based on aspartic acid racemization in elastin from the yellow ligaments. 117(2), 96–101. Retrieved from <https://link.springer.com/article/10.1007/s00414-002-0355-2>
- Roher, A. E., Lowenson, J. D., Clarke, S., Wolkow, C., Wang, R., Cotter, R. J., ... Greenberg, B. D. (1993). Structural alterations in the peptide backbone of  $\beta$ -amyloid core protein may account for its deposition and stability in Alzheimer's disease. *Journal of Biological Chemistry*, 268(5), 3072–3083. [https://doi.org/10.1016/s0021-9258\(18\)53661-9](https://doi.org/10.1016/s0021-9258(18)53661-9)
- Roos, G., Foloppe, N., & Messens, J. (2013). Understanding the pKa of redox cysteines: The key role of hydrogen bonding. *Antioxidants and Redox Signaling*, 18(1), 94–127. <https://doi.org/10.1089/ars.2012.4521>
- Rothschild, L. J., & Mancinelli, R. L. (2001). Life in extreme environments. *Nature* 2001 409:6823, 409(6823), 1092–1101. <https://doi.org/10.1038/35059215>
- Ruotolo, B. T., & Robinson, C. V. (2006). Aspects of native proteins are retained in vacuum. *Current Opinion in Chemical Biology*, 10(5), 402–408. <https://doi.org/10.1016/j.cbpa.2006.08.020>
- Sasaki, N. A., Garcia-Alvarez, M. C., Wang, Q., Ermolenko, L., Franck, G., Nhiri, N., ... Potier, P. (2009). N-Terminal 2,3-diaminopropionic acid (Dap) peptides as efficient methylglyoxal scavengers to inhibit advanced glycation endproduct (AGE) formation. *Bioorganic & Medicinal Chemistry*, 17(6), 2310–2320. <https://doi.org/10.1016/J.BMC.2009.02.018>
- Sekimoto, T., Matsuyama, T., Fukui, T., & Tanizawa, K. (1993). Evidence for lysine 80 as general base catalyst of leucine dehydrogenase. *Journal of Biological Chemistry*, 268(36), 27039–27045. [https://doi.org/10.1016/S0021-9258\(19\)74215-X](https://doi.org/10.1016/S0021-9258(19)74215-X)

- Shapira, R., & Chou, C. H. J. (1987). Differential racemization of aspartate and serine in human myelin basic protein. *Biochemical and Biophysical Research Communications*, *146*(3), 1342–1349.  
[https://doi.org/10.1016/0006-291X\(87\)90797-2](https://doi.org/10.1016/0006-291X(87)90797-2)
- Shapiro, S. D., Endicott, S. K., Province, M. A., Pierce, J. A., & Campbell, E. J. (1991). Marked longevity of human lung parenchymal elastic fibers deduced from prevalence of D-aspartate and nuclear weapons-related radiocarbon. *Journal of Clinical Investigation*, *87*(5), 1828.  
<https://doi.org/10.1172/JCI115204>
- Shekhawat, S. S., Pham, G. H., Prabakaran, J., & Strieter, E. R. (2014). Simultaneous Detection of Distinct Ubiquitin Chain Topologies by 19F NMR. *ACS Chemical Biology*, *9*(10), 2229–2236.  
<https://doi.org/10.1021/CB500589C>
- Shen, J., & Ren, J. (2012). Gas phase acidity of a cysteine residue in small oligopeptides. *International Journal of Mass Spectrometry*, *316–318*, 147–156. <https://doi.org/10.1016/J.IJMS.2012.01.001>
- Smith\*, G. D., & Jaffe, R. L. (1996). Quantum Chemistry Study of Conformational Energies and Rotational Energy Barriers in n-Alkanes. *Journal of Physical Chemistry*, *100*(48), 18718–18724.  
<https://doi.org/10.1021/JP960413F>
- Sokalingam, S., Raghunathan, G., Soundrarajan, N., & Lee, S.-G. (2012). A Study on the Effect of Surface Lysine to Arginine Mutagenesis on Protein Stability and Structure Using Green Fluorescent Protein. *PLOS ONE*, *7*(7), e40410. <https://doi.org/10.1371/JOURNAL.PONE.0040410>
- Steinegger, M., Mirdita, M., & Söding, J. (2019). Protein-level assembly increases protein sequence recovery from metagenomic samples manifold. *Nature Methods*, *16*(7), 603–606.  
<https://doi.org/10.1038/s41592-019-0437-4>
- Susan L. Schafer, William C. Barrett, §, Abraham T. Kallarakal, Bharati Mitra, John W. Kozarich, ¶ and, John A. Gerlt\*, §, ... Kenyon, G. L. (1996). Mechanism of the Reaction Catalyzed by Mandelate Racemase: Structure and Mechanistic Properties of the D270N Mutant. *Biochemistry*, *35*(18), 5662–5669. <https://doi.org/10.1021/BI960174M>

- Tan, J. P., & Ren, J. (2007). Determination of the Gas-Phase Acidities of Cysteine-Polyalanine Peptides Using the Extended Kinetic Method. *Journal of the American Society for Mass Spectrometry*, *18*(2), 188–194. <https://doi.org/10.1016/J.JASMS.2006.09.012>
- Taylor, R., & Kennard, O. (1982). Crystallographic Evidence for the Existence of C···O, C-H-N, and C-H-Cl Hydrogen Bonds. *J. Am. Chem. Soc.*, *104*(19), 5063. <https://doi.org/10.1021/JA00383A012>
- Teixeira, F. (2017). *VibAnalysis - Tools for performing vibrational analysis on molecular systems*. Retrieved from <https://github.com/teixeirafilipe/vibAnalysis>
- Teixeira, F., & Cordeiro, M. N. D. S. (2018). Improving Vibrational Mode Interpretation Using Bayesian Regression. *Journal of Chemical Theory and Computation*, *15*(1), 456–470. <https://doi.org/10.1021/ACS.JCTC.8B00439>
- Temimi, A. H. K. Al, Reddy, Y. V., White, P. B., Guo, H., Qian, P., & Mecinović, J. (2017). Lysine Possesses the Optimal Chain Length for Histone Lysine Methyltransferase Catalysis. *Scientific Reports 2017 7:1*, *7*(1), 1–12. <https://doi.org/10.1038/s41598-017-16128-4>
- Thomas, M. C., Mitchell, T. W., & Blanksby, S. J. (2005). A comparison of the gas phase acidities of phospholipid headgroups: Experimental and computational studies. *Journal of the American Society for Mass Spectrometry*, *16*(6), 926–939. <https://doi.org/10.1016/j.jasms.2005.02.019>
- Torres, A. M. M., Tsampazi, C., Geraghty, D. P. P., Bansal, P. S. S., Alewood, P. F. F., & Kuchel, P. W. W. (2005). D-amino acid residue in a defensin-like peptide from platypus venom: Effect on structure and chromatographic properties. *Biochemical Journal*, *391*(2), 215–220. <https://doi.org/10.1042/BJ20050900>
- Van Amersfoort, P. W., Best, R. W. B. B., Van der Geer, C. A. J. J., Mastop, W. J. I., Meddens, B. J. H. H., van der Meer, A. F. G. G., ... van der Wiel, M. J. (1989). Felix: The Dutch Free Electron Laser for Infrared Experiments. *Nuclear Inst. and Methods in Physics Research, A*, *285*(1–2), 67–70. [https://doi.org/10.1016/0168-9002\(89\)90427-0](https://doi.org/10.1016/0168-9002(89)90427-0)

- Verzijl, N., DeGroot, J., Thorpe, S. R., Bank, R. A., Shaw, J. N., Lyons, T. J., ... TeKoppele, J. M. (2000). Effect of collagen turnover on the accumulation of advanced glycation end products. *Journal of Biological Chemistry*, 275(50), 39027–39031. <https://doi.org/10.1074/JBC.M006700200>
- Völkel, P., & Angrand, P. O. (2007). The control of histone lysine methylation in epigenetic regulation. *Biochimie*, 89(1), 1–20. <https://doi.org/10.1016/J.BIOCHI.2006.07.009>
- Vollmer, W., Blanot, D., & De Pedro, M. A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*, 32(2), 149–167. <https://doi.org/10.1111/j.1574-6976.2007.00094.x>
- Weatherly, C. A., Du, S., Parpia, C., Santos, P. T., Hartman, A. L., & Armstrong, D. W. (2017). D - Amino Acid Levels in Perfused Mouse Brain Tissue and Blood: A Comparative Study. *ACS Chemical Neuroscience*, 8(6), 1251–1261. <https://doi.org/10.1021/acscemneuro.6b00398>
- Weng, L., Zhou, C., & Greenberg, M. M. (2014). Probing Interactions between Lysine Residues in Histone Tails and Nucleosomal DNA via Product and Kinetic Analysis. *ACS Chemical Biology*, 10(2), 622–630. <https://doi.org/10.1021/CB500737Y>
- Westaway, K. C. (2006). Using kinetic isotope effects to determine the structure of the transition states of SN2 reactions. *Advances in Physical Organic Chemistry*, 41, 217–273. [https://doi.org/10.1016/S0065-3160\(06\)41004-2](https://doi.org/10.1016/S0065-3160(06)41004-2)
- Westheimer, F. H. (1961). The magnitude of the primary kinetic isotope effect for compounds of hydrogen and deuterium. *Chemical Reviews*, 61(3), 265–273. <https://doi.org/10.1021/cr60211a004>
- Wiedemann, C., Kumar, A., Lang, A., & Ohlenschläger, O. (2020). Cysteines and Disulfide Bonds as Structure-Forming Units: Insights From Different Domains of Life and the Potential for Characterization by NMR. *Frontiers in Chemistry*, 280.
- Williams, T. I., Denault, J. W., & Cooks, R. G. (2001). Proton affinity of deuterated acetonitrile estimated by the kinetic method with full entropy analysis. *International Journal of Mass Spectrometry*, 210–211(211), 133–146. [https://doi.org/10.1016/S1387-3806\(01\)00455-9](https://doi.org/10.1016/S1387-3806(01)00455-9)
- Williams, T., & Kelley, C. (2010). gnuplot homepage. Retrieved November 6, 2021, from Gnuplot 4.4: an interactive plotting program, website: <http://www.gnuplot.info/>

Wolosker, H. (2007). NMDA Receptor Regulation by D-serine: New Findings and Perspectives.

*Molecular Neurobiology* 2007 36:2, 36(2), 152–164. <https://doi.org/10.1007/S12035-007-0038-6>

Xu, Z., Sun, Z., Li, S., Xu, Z., Cao, C., Xu, Z., ... Xu, H. (2015). Systematic unravelling of the biosynthesis of poly (L-diaminopropionic acid) in *Streptomyces albus* PD-1. *Scientific Reports*, 5(1), 1–10. <https://doi.org/10.1038/srep17400>

Xubin Zheng, & Cooks, R. G. (2002). Thermochemical Determinations by the Kinetic Method with Direct Entropy Correction†. *Journal of Physical Chemistry A*, 106(42), 9939–9946. <https://doi.org/10.1021/JP020595F>

Yuanhe Li. (n.d.). Rendering an RDG-filled isosurface with HSV blending mode in PyMOL. Retrieved October 9, 2021, from <http://bbs.keinsci.com/thread-15756-1-1.html>

Zhang, Y., Buen, Z., Browne, M. D., Mann, Y. S., & Ren, J. (2021). Subtle chirality effects of a D/l-Cysteine on the intrinsic acidity and conformation of isomeric tripeptides ACA and AdCA. *International Journal of Mass Spectrometry*, 469, 116685. <https://doi.org/10.1016/J.IJMS.2021.116685>

Zhang, Z., Li, D., Jiang, W., & Wang, Z. (2018). The electron density delocalization of hydrogen bond systems. *Https://Doi.Org/10.1080/23746149.2018.1428915*, 3(1), 298–315. <https://doi.org/10.1080/23746149.2018.1428915>

Zhao, C., Song, C., Luo, Y., Yu, Z., & Sun, M. (2008). 1-2,3-Diaminopropionate: One of the building blocks for the biosynthesis of Zwittermicin A in *Bacillus thuringiensis* subsp. *kurstaki* strain YBT-1520. *FEBS Letters*, 582(20), 3125–3131. <https://doi.org/10.1016/j.febslet.2008.07.054>

## APPENDIX A: INSTALLATION AND RUNNING OF COMPUTATIONAL PROGRAMS

Installation of xtb and crest for CREST conformational search in CentOS7

1. Download the latest version of xtb from: <https://github.com/grimme-lab/xtb/releases>

▼ Assets 8

 <a href="#">xtb-6.4.1-linux-x86_64.tar.xz</a>	17.8 MB
 <a href="#">xtb-6.4.1-linux-x86_64.tar.xz.sha256</a>	96 Bytes
 <a href="#">xtb-6.4.1-linux-x86_64.tar.xz.sig</a>	119 Bytes
 <a href="#">xtb-6.4.1.tar.xz</a>	949 KB
 <a href="#">xtb-6.4.1.tar.xz.sha256</a>	83 Bytes
 <a href="#">xtb-6.4.1.tar.xz.sig</a>	119 Bytes
 <a href="#">Source code (zip)</a>	
 <a href="#">Source code (tar.gz)</a>	

Accessed on Oct 6<sup>th</sup>, 2021

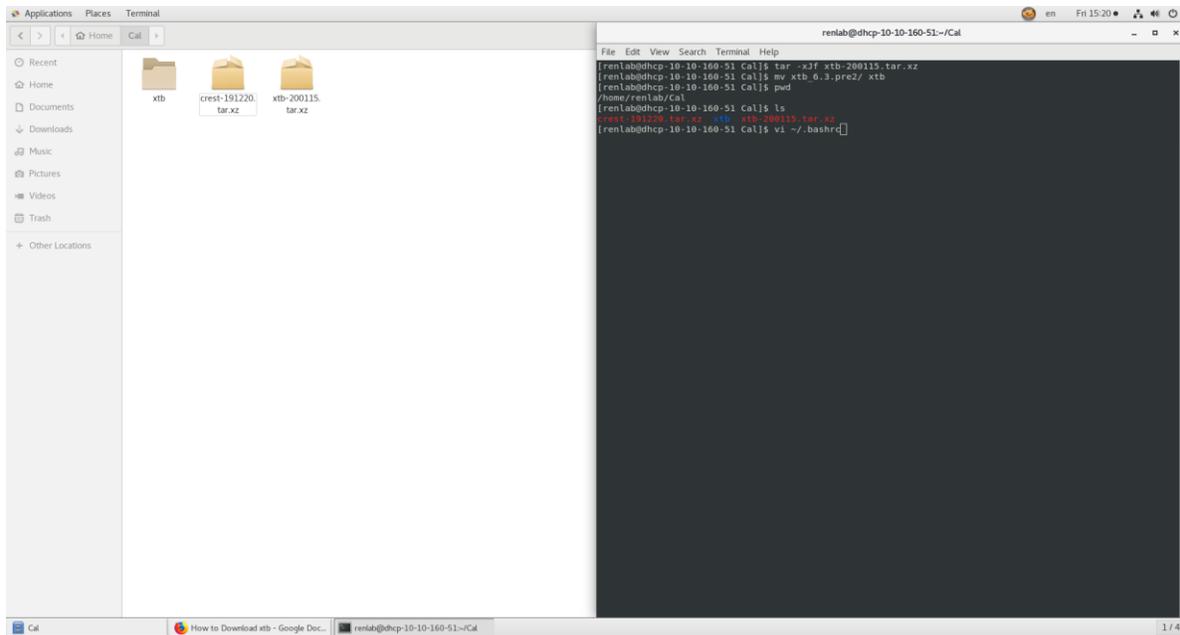
2. Download the latest version of crest from: <https://github.com/grimme-lab/crest/releases>

▼ Assets 3

 <a href="#">crest-binary-2.11.1.zip</a>	10.5 MB
 <a href="#">Source code (zip)</a>	
 <a href="#">Source code (tar.gz)</a>	

Accessed on Oct 6<sup>th</sup>, 2021

3. Move downloaded files into /home/renlab/Cal
4. Open terminal by right clicking and selecting “Open Terminal”
5. Unzip file using terminal command: **tar -xJf xtb-200115.tar.xz**
6. Rename folder: **mv xtb\_6.3.pre2/ xtb**
7. Show current directory: **pwd**
8. Add the directory of xtb to PATH variable
  - a. Open .bashrc file: **vi ~/.bashrc**



- b. Press “i” to enter insert mode
- c. Enter the following on a new line:
  - i. **#following lines added for xtb <2021/10/06> # for reference purposes**
  - ii. **export PATH=\$PATH:/home/renlab/Cal/xtb/bin**
  - iii. **export XTBPATH=/home/renlab/Cal/xtb/share/xtb**
  - iv. **export OMP\_NUM\_THREADS=4 # threads the computer has**
  - v. **export MKL\_NUM\_THREADS=4 # threads the computer has**
  - vi. **export OMP\_STACKSIZE=1000m**
  - vii. **ulimit -s unlimited**

```

renlab@dhcp-10-10-160-51:~/Cal
File Edit View Search Terminal Help
# .bashrc
# Source global definitions
if [ -f /etc/bashrc ]; then
    . /etc/bashrc
fi

# Uncomment the following line if you don't like systemctl's auto-paging feature:
# export SYSTEMD_PAGER=

# User specific aliases and functions

# added for xtb. 2020/02/13
export PATH=$PATH:/home/renlab/Cal/xtb/bin
export XTBPATH=/home/renlab/Cal/xtb/share/xtb
export OMP_NUM_THREADS=4
export MKL_NUM_THREADS=4
export OMP_STACKSIZE=1000m
ulimit -s unlimited

```

- d. Exit insert mode by pressing ESC
  - e. Save and exit the file by typing: **:wq**
9. Run **xtb** to ensure it works
- a. If error message with “No geometry given, so there is nothing to do, abnormal termination of xtb” then the program is ready to accept data input.

```

renlab@dhcp-10-10-160-51:~/Cal
File Edit View Search Terminal Help
[renlab@dhcp-10-10-160-51 Cal]$ tar -xJf xtb-200115.tar.xz
[renlab@dhcp-10-10-160-51 Cal]$ mv xtb.6.3.pre2/ xtb
[renlab@dhcp-10-10-160-51 Cal]$ pwd
/home/renlab/Cal
[renlab@dhcp-10-10-160-51 Cal]$ ls
crest-191220  xtb  xtb-200115  tar.xz
[renlab@dhcp-10-10-160-51 Cal]$ vi ~/.bashrc
[renlab@dhcp-10-10-160-51 Cal]$ xtb
#ERROR! No geometry given, so there is nothing to do.
abnormal termination of xtb
[renlab@dhcp-10-10-160-51 Cal]$

```

How to download crest on CentOS (Linux)

Terminal commands in **bold**

## Procedure

10. Open terminal by right clicking and selecting “Open Terminal”
11. Make folder “crest” in desired location: **mkdir crest1**
12. Unzip file using terminal command: **tar -xJf crest-191220.tar.xz**
13. Rename folder: **mv crest crest1**
14. Rename folder again: **mv crest1 crest**
15. Add the directory of xtb to PATH variable
  - a. Open .bashrc file: **vi ~/.bashrc**
  - b. Press “i” to enter insert mode
  - c. Enter the following on a new line:
    - i. **#added for crest <insert date>** This is for reference purposes

```

renlab@dhcp-10-10-160-51:~/Cal
File Edit View Search Terminal Help
# .bashrc
# Source global definitions
if [ -f /etc/bashrc ]; then
    . /etc/bashrc
fi
# Uncomment the following line if you don't like systemctl's auto-paging feature:
# export SYSTEMD_PAGER=
# User specific aliases and functions
# added for xtb. 2020/02/13
export PATH=$PATH:/home/renlab/Cal/xtb/bin
export XTBPATH=/home/renlab/Cal/xtb/share/xtb
export OMP_NUM_THREADS=4
export MKL_NUM_THREADS=4
export OMP_STACKSIZE=1000m
ulimit -s unlimited
# added for crest. 2020/02/21
export PATH=$PATH:/home/renlab/Cal/crest

```

- ii. **export PATH=\$PATH:/home/renlab/Cal/crest**
    - iii. Exit insert mode by pressing esc
  - d. Save and exit the file by typing: **:wq**
16. Run crest to ensure it works