UNIVERSITY OF CALIFORNIA
RIVERSIDE

Mass Spectrometry Based Characterization of Protein Structure and Peptide Isomerization

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Yuanqi Tao

December 2014

Dissertation Committee:
Dr. Ryan Julian, Chairperson
Dr. Yinsheng Wang
Dr. Huiwang Ai

The Dissertation of Yuanqi Tao is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

Completing a Ph.D. degree is a long, but happy journey. I am very thankful that I can have all the supports from my teachers and professors, colleagues, family and friends. My interest in chemistry first started in high school, from my chemistry teachers Mrs. Yingji Yin and Mr. Jungang Hao. They were very instrumental in my decision to choose chemistry as my college major. In my third year at Nankai University, I joined Professor Zhengpu Zhang's lab in the Institute of Polymer Chemistry. After working in this group and learning about polymer synthesis, I joined Professor Xiuping Yan's group in the Research Center for Analytical Science. While working in the Yan lab, I did research under the supervision of Dr. Hefang Wang, from whom I learned several analytical chemistry techniques. Both Dr. Yan and Dr. Zhang are excellent scientists whose sincerity and kindness helped develop me into the chemist and person I am today. Through their guidance, I decided to pursue an advance degree in analytical chemistry. Also, I deeply appreciate the conversation with Dr. Shen Lin from Nankai University, who got his PhD from Purdue University, in analytical chemistry with a focus in mass spectrometry. His vast knowledge in the field prompted me to join Dr. Ryan Julian's research group at UC Riverside.

Firstly, I must thank Ryan for teaching and training over the last five years. From him I have learned how to be both an efficient and conscientious scientist. Ryan has taught me to never jump to hasty conclusions in regards to any scientific question. Whenever I encountered any problems, he would inspire me with excellent ideas while still

encouraging me to think independently and solve the problem by myself. There were a lot of times that the experiments did not work, and if I complained to Ryan, he would say, “well, this is not the end of the world… negative results are results.” Moreover, he would encourage me to find valuable information from a “failed” experiment and to propose solutions or new ideas. Ryan motivated me to express myself, communicate with other people and be enthusiastic and confident about my research and future career. It was my privilege to have had Ryan as my mentor during my graduate study; thank you very much.

Besides my PhD advisor, I would like to thank Dr. Yinsheng Wang, Dr. Huiwang Ai, Dr. Christopher Bardeen, Dr. Cindy Larive, Dr. Quan Cheng, Dr. Sharon Walker, as my SYRE exam, Oral exam, and final dissertation committee, for their comments and kind help.

I would like to express my sincere gratitude to all my colleagues, who are excellent researchers as well as reliable friends. I had a great time working with you during the last five years and wish you the best of luck in all of your future endeavors. I enjoyed all the discussions and conversions I had with Dr. Benjamin Moore on a broad range of topics, including science, programming and hobbies such as Ben’s fascination for helicopters. I would like to thank previous lab members who used to train me using the instrument and give me scientific suggestions. Dr. Tony Ly, an extremely knowledgeable and productive scientist and an expert on mass spectrometry and protein science. Dr. Jolene Diedrich, a super nice person and excellent scientist, thank you for all the discussion about the proteomics research and advice for my postdoc application. Dr. Qingyu Sun, a very good

friend with passionate personality, thank you for helping make my transition into graduate school and my move to Riverside a smooth one. I also would like to thank Dr. Thu Huong Pham for all the interesting discussions and kind encouragement. Dr. Arun Agarwal, I learned a lot from his enthusiasm! Xing Zhang, a smart scientist with tremendous talent, who is always willing to discuss any scientific questions. Omar Hamdy, a scientist with excellent communication and management skills. Nathan Hendricks, I really loved your Valentine's Day idea! Christopher Nellessen, thanks for the advice about job hunting. Yana Lyon, I truly wish we could have met earlier! Hyojik Yang, a very hard-working student, I wish you all the best in the graduate study. Eric Knudsen, who happens to be extremely nice and polite, and offered me a lot of help during my first year. I thank you all for your support and friendship through the years. I would also like to thank the undergraduate student, Neil Quebbemann, for synthesizing the peptides and helping with the data analysis in Chapter 2. Again, I am so lucky to have had all of you as my colleagues.

I would also like to recognize the collaborators who I had the privilege to work with over the years. Thank you to Professor Joe Loo for his helpfulness and kindness in allowing me to perform ECD experiments on the FT-ICR instrument at UCLA. I thank Dr. Richard Kondrat and Mr. Ronald New for helping me run experiments on the Time-of-Flight mass spectrometer in the UCR mass spectrometry facility. I thank Dr. Songqin Pan for helping me with the Orbitrap Fusion instrument at Center for Plant Cell Biology. I thank Prof. Daniel Gallie from biology department for letting me use the sonicator and centrifuge. I also need to express my gratitude to Dr. Cindy Larive, Dr. Wenwan Zhong,

Dr. Chia-en Chang, Dr. Gregory Beran, Dr. Yinsheng Wang, for their guidance during graduate division courses.

Last, but not least, I would like to thank my parents and my friends. Family and friendship are priceless. I know my parents will always be there for me no matter what happens. Their unconditional love and support provided the foundation of my graduate study and future career. I would like to thank my cousins, my grandparents, my aunts and uncles, and everyone in the family. I cannot imagine I would have made it this far without your support. I thank my lovely friends, especially Dr. Lichao Zhang, Dr. Yangyang Liu, soon-to-be Dr. Yuanyuan Shen, soon-to-be Dr. Qian Liu, who have been there for me through the ups and the downs over the past ten years. We share similar life experiences, pursue similar dreams and will keep supporting each other along the way. I am very lucky to have many friends here at UC Riverside: Dr. Wenting Hou, Dr. Yang Li, Dr. Fang Jia, Dr. Bo Zhao, soon-to-be Dr. Haiyu Zhang, you are my family here in the US. I am grateful for how we take care of each other, and also explore new things in the life together. Although we are heading down new paths, I will cherish the precious memories we had in Riverside for the rest of my life. Finally, I thank my fiancé Chen Yang, who has been supportive, understanding, and encouraging. Words cannot express my gratefulness for having you by my side over years and everything you have done for me-- thank you.

I thank the following institutions for funding: UC Riverside, the National Institutes of Health (NIH R01 GM08406) and the National Science Foundation Career Award (RRJ,

CHE-0747481). The peptides used in Chapter 2 were synthesized by Neil Quebbemann and Neil also contributed to a portion of the experiments and data analysis.

The text of this dissertation, in part or in full, is a reprint of the material as they appear in the following publications:

Ch. 2: Tao, Y.; Quebbemann, N. R.; Julian, R. R., *Anal. Chem.* **2012**, *84*, 6814-6820.

Ch. 3: Tao, Y.; Julian, R. R., *Ana. Chem.* 2014, DOI: 10.1021/ac502296c

Ch. 4: Tao, Y.; Julian, R. R., *J. Am. Soc. Mass. Spectrom.***2013**, *24*, 1634-1640.

Ch. 5: Tao, Y.; Julian, R. R., *Biochemistry* **2012**, *51*, 1796-1802.

ABSTRACT OF THE DISSERTATION

Mass Spectrometry Based Characterization of Protein Structure and Peptide Isomerization

by

Yuanqi Tao

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, December 2014
Dr. Ryan R. Julian, Chairperson

Structure is the key factor in protein function. Mass spectrometry has been shown to be a powerful technique for exploring protein structures and post translational modifications. This dissertation describes the application of mass spectrometry based techniques to investigate peptide isomerization and protein solution phase structure.

The first half of the dissertation focuses on the identification of subtle post translational modifications of protein: epimerization and isomerization. The presence of a single D-amino acid in a peptide is very difficult to detect. Mass spectrometry-based approaches rely on differences in fragmentation between peptide epimers. The success of this approach is dependent on the structural sensitivity of the fragmentation method. Radical Directed Dissociation (RDD) is particularly sensitive to the structure of the ion being fragmented. It is demonstrated herein that RDD provides significantly better chiral discrimination than collisional induced dissociation (CID). The combination of RDD and CID is further applied for the identification of peptide epimerization/isomerization in crystallin proteins. The sites that undergo the greatest degree of isomerization in

crystallin correspond to disordered regions, which may have important implications for the chaperone functionality of the proteins within the context of aging.

The second half of the dissertation utilizes the ability of 18-crown-6 (18C6) to form noncovalent complexes with cationic groups to examine protein structures in solution. It is demonstrated with model peptides that the 18C6 adduct stability is increased if intramolecular charge complexation is inhibited by steric or competitive binding. Molecular mechanics and dissociation experiments demonstrated that significant structural changes occur upon loss of 18C6 in model peptides. Collisional activation of protein-18C6 complex indicates that lower charge states represent structures that are not similar to gas phase idealized states. Therefore, 18C6 can behave as a pseudo-solvent molecule and preserve protein solution phase structure. 18C6 attachment can also be applied to investigate protein electrostatic surface structure. It is shown that proteins can have completely different surface structures despite the fact that the backbone structures are similar. The results illustrated a new technique to characterize protein surface structure and dynamics.

TABLE OF CONTENTS

Chapter 1 1
CHARACTERIZATION OF PROTEIN STRUCTURE AND PTMS BY MASS SPECTROMETRY 1
1.1 Protein Characterization by Mass Spectrometry 1
1.2 Post Translational Modification Analysis 2
1.3 Protein Tertiary Structures 7
References 13
Chapter 2 17
DISCRIMINATING D-AMINO ACID CONTAINING PEPTIDES BY RADICAL DIRECTED DISSOCIATION USING MASS SPECTROMETRY 17
2.1 Introduction 17
2.2 Experimental Methods 21
2.2.1 Materials 21
2.2.2 Peptide Synthesis 21
2.2.3 Radical Precursor Synthesis 22
2.2.4 Mass Spectrometry and Radical Directed Dissociation 23
2.3 Results and Discussion 24
2.3.1 D-serine Peptides 24
2.3.2 D-Asp Peptides. 31
2.3.3 D-Ala Peptides. 35
2.3.4 Discrimination Abilities 40
2.4 Conclusions 43
References 44
Chapter 3 47
ISOMER PROTEOMICS: A CASE STUDY OF EPIMERIZATION AND ISOMERIZATION IN CRYSTALLINS 47
3.1 Introduction 47
3.2 Experimental Methods 51

3.2.1 Materials 51
3.2.2 Peptide and Radical Precursor Synthesis 52
3.2.3 Protein Extraction and Digestion 52
3.2.4 Mass Spectrometry and Radical Directed Dissociation 53
3.2.5 LC-MS Data Acquisition and Analyisis 53
3.2.6 Calculation of R Values 55
3.3 Results and Discussion 56
3.3.1 General Approach 56
3.3.2 Isomer Separation 58
3.3.3 Data Analysis 62
3.3.4 Sheep Crystallins 65
3.3.5 Discrimination Abilities 72
3.3.6 Isomerization and Functionality 79
3.4 Conclusions 80
References 82
Supporting Information 86
Chapter 4 91
FACTORS THAT INFLUENCE COMPETITIVE INTERMOLECULAR SOLVATION OF PROTONATED GROUPS IN PEPTIDES AND PROTEINS IN THE GAS PHASE 91
4.1 Introduction 91
4.2 Experimental Methods 93
4.2.1 Materials 93
4.2.2 Peptide Synthesis and Iodination 94
4.2.3 Mass Spectrometry 94
4.2.4 Molecular Mechanics Calculations 95
4.3 Results and Discussion 96
4.3.1 Mass Shift of Crown Complex in the Ion Trap 96
4.3.2 Standard Peptides 98
4.3.3 Molecular Dynamics Simulations 105

4.3.4 Protein 18C6 Adduct Stability....107
4.4 Conclusions....116
References....118
Chapter 5....121
EXAMINING PROTEIN SURFACE STRUCTURE IN HIGHLY CONSERVED SEQUENCE VARIANTS WITH MASS SPECTROMETRY....121
5.1 Introduction....121
5.2 Experimental Methods....125
5.2.1 Protein Samples and Purification....125
5.2.2Mass Spectrometry....125
5.2.3 Protein Structures....127
5.3 Results and Discussion....127
5.3.1 Insulin....127
5.3.2 Cytochrome c (cytc)....130
5.3.3 Lysozyme....134
5.3.4 Comparative Analysis....137
5.3.5 Protein Denaturation....140
5.4 Conclusions....145
References....146
Chapter 6....149
CONCLUDING REMARKS....149

LIST OF FIGURES

Figure 2.1 (a) PD of the two serine epimers of ^{1}DVG<u>S</u>NK-NH_2. The major difference is the -30Da serine side chain loss. (b) CID of -30 loss from PD of [^{1}DVG<u>S</u>NK-NH_2 + H]$^+$ (D-ser epimer). A series of b_3, b_4-30, b_5-30 fragments are observed, while b_3-30 ion is not present in the spectrum, which verifies that the -30 loss is from serine. 25

Figure 2.2 (a) CID of re-isolated radical peptide [(^{1}DVG<u>S</u>NK-NH_2)• +H]$^+$ formed in Figure 2.1a. The inset highlights differences in the losses of water and NH_3. (b) Traditional CID of protonated DVG<u>S</u>NK-NH_2. 27

Figure 2.3 (a) PD-CID of [^{1}G<u>S</u>WD + H]$^+$ and (b) CID of [G<u>S</u>WD + H]$^+$. RDD shows a better chiral discrimination ability with an R_{Chiral} value of 12±1 (peak 294.17/b_3•) while the number for CID is 1.85±0.05 (b_3-18/y_2). 29

Figure 2.4 (a) PD-CID of [^{1}GL<u>S</u>FA-NH_2 + H]$^+$ and (b) CID of [GL<u>S</u>FA-NH_2 + H]$^+$. Relative intensities of b_4• are significantly different between the two epimers. The R_{Chiral} value is 32±4 (b_4•/c_3) for RDD and 6.9±0.2 (b_4-18/-17) for CID. 30

Figure 2.5 (a) PD of [^{1}IQTGL<u>D</u>ATHAER+H]$^+$ epimers. (b) PD of [^{1}IQTGL<u>D</u>ATHAER+2H]$^{2+}$ epimers. The D-form shows abundant −56 Da loss from Leu for both charge states. Bold down arrow indicates precursor ion. 32

Figure 2.6 (a) CID of [IQTGL<u>D</u>ATHAER+H]$^+$ epimers. (b) CID of [IQTGL<u>D</u>ATHAER+2H]$^{2+}$ epimers. (c) CID of [IQTGL<u>D</u>ATHAER+3H]$^{3+}$ epimers. This demonstrated that RDD provides better chiral discrimination than CID in different charge states. 33

Figure 2.7 (a) PD-CID of [^{1}<u>D</u>AEFR + H]$^+$ and (b) CID of [<u>D</u>AEFR + H]$^+$ epimers. RDD shows a better chiral discrimination ability with a R_{Chiral} value of 6.2±0.3 (peak z_4^+/-44) while the number for CID is 2.33±0.07 (y_4^+/-17). 34

Figure 2.8 (a) PD-CID of [(^{1}Y<u>A</u>FGYPS-NH_2)• − H]$^-$ epimers. (b) CID of [Y<u>A</u>FGYPS-NH_2 + H]$^+$ epimers show similar spectra. 36

Figure 2.9 (a) PD of [Y<u>A</u>FGYPS-NH_2 + ***3*** + H]$^+$ epimers. (b) CID of the radical complex generated by loss of iodine in panel a. The insets highlight differences in the ratio of radical/canonical peptide generated by loss of the crown adduct. Bold down arrow indicates precursor ion. 38

Figure 2.10 (a) PD of [Y<u>A</u>FDVVG-NH_2 + ***3*** + H]$^+$ and (b) PD-CID of [Y<u>A</u>FDVVG-NH_2+ ***3*** +H]$^+$ epimers. The hydrogen abstraction efficiency is limited in the D-Ala peptide. 39

Figure 2.11 PD-CID of [<u>A</u>STTTNYT-NH_2+3+H]$^+$ epimers. The radical transfer efficiency difference is very easy to identify. 40

Figure 2.12 Calibration curve for PD of ^{1}DVG<u>S</u>NK-NH_2 (a) and PD of Y<u>A</u>FDVVG-NH_2 + ***3*** complex (b). 43

Figure 3.1 LC-MS/MS results for a mixture of the four isomers of synthetic peptide DAEFR (a) before (b) after modification. ^{I}X represents the 4-iodo-benzoic acid modification of X where X is any amino acid. Underlined residues correspond to D-amino acid, and bold residues correspond to isoaspartic acid. The inset of (b) shows the structure of 4-iodo-benzoic acid. The four isomers can be easily separated after covalent modification by 4-iodo-benzoic acid. 60

Figure 3.2 (a) LC chromatogram of a peptide mixture containing several synthetic peptides including ^{I}DVGSNK-NH_2, ^{I}DVG<u>S</u>NK-NH_2 and ^{I}LDLAGR. The ^{I}DVGSNK epimers cannot be completely separated. (b) MS^3 (RDD) spectra at 35.68min and 36.49min (the blue asterisks) of (a). The two spectra are significantly different, indicating the LC peak contains two peptides. (c) LC chromatogram of the same run as (a) during a later elution time. (d) MS^3 (RDD) spectra at 72.27min and 72.87min (the blue asterisks) of (c). The two spectra are almost identical. 61

Figure 3.3 Sequence coverage (orange), degree of isomerization (blue) and degree of epimerization (green) for αA-crystallin, αB-crystallin, and βB3- crystallin. The white bar represents the full protein sequence. 66

Figure 3.4 (a) LC chromatogram for peptide TVLDSGISEVR in sheep eye crystallin digestion mixture. (b) CID of [TVLDSGISEVR + 2H]$^{2+}$ at 24.76min, 26.59min, 28.53min separately. The spectra at 24.76min and 28.53min are identified as L-isoAsp and D-Asp containing peptide by comparing with (c). The R values for each of these spectra relative to the synthetic peptides spectra are 1.3 and 1.7. The red 70

Figure 3.5 (a) LC chromatogram for peptide ^{I}TVLDSGISEVR in sheep eye crystallin digestion mixture after modification. (b) RDD spectra of [^{I}TVLDSGISEVR + 2H]$^{2+}$ at 64.68min, 67.57min, 70.44min separately. The three spectra are identified as L-isoAsp, D-isoAsp, and L- Asp containing peptide by comparing with (c). The R values for each of these spectra relative to the synthetic peptides spectra are of 2.6, 2.4 and 3.1. The red

stars represent fragments from a co-eluting peptide. (c) RDD spectra of four synthetic peptides L-Asp, D- Asp, L-isoAsp, and D-isoAsp ^{I}TVLDSGISEVR. 71

Figure 3.6 LC chromatogram for peptide HFSPEDLTVK in sheep eye crystallin digestion mixture. (b) CID spectra of $[HFSPEDLTVK + 2H]^{2+}$ at 32.44min and 35.39min, separately. The R_{isomer} value is 1.5 (y_9^+ & -18). (c) CID spectra of $[HFSPEDLTVK + H]^+$ at 32.64 and 35.53min. The R_{isomer} value is 7.3 (b_8^+ & y_5^+). The isomer discrimination is achieved at +1 charge state. (d) LC chromatogram for peptide HFSPEDLTVK^{I} in sheep eye crystallin digestion mixture after modification. (e) RDD spectra of $[HFSPEDLTVK^{I} + 2H]^{2+}$ at 61.97 and 66.25min. The R_{isomer} value is 25.3 (y_6^+& y_4^+). 73

Figure 3.7 (a) LC chromatogram for peptide Ac-AEQHSAPEQAAAGK in sheep eye crystallin digestion mixture (b) CID spectra of $[Ac\text{-}AEQHSAPEQAAAGK + 2H]^{2+}$ at 22.18min and 22.48min, separately. The two spectra are identified as L-serine and D-serine containing peptide by comparing with (c). (c) CID spectra of two synthetic peptides L-serine and D-serine Ac-AEQHSAPEQAAAGK. 75

Figure 3.8 Relationship between retention time difference in HPLC and the R_{isomer} value of peptide isomers listed in Table 3.1, for CID (a) and RDD (b). 78

Figure 3.9 Isomerization ratio of αA and αB crystallins. Different colors indicate the three structural regions of crystallin, with the N-terminal region in orange, the alpha crystallin domain in blue, and the C-terminal extension in purple. The black asterisks represent aspartic acid residues and the white asterisks represent serine residues. The black stars (peptide 89-99 in αA and peptide 124-149 in αB) indicate regions where isomerization was detected but not quantified due to incomplete separation by HPLC. . 80

Figure 4.1 (a) Full MS spectrum of peptide MRFA. The exact mass is 524.27 Da for $[MRFA+H]^+$ (b) Full MS spectrum of peptide MRFA and 18C6 acquired on a different day under different operating conditions. The exact mass is 788.42Da for $[MRFA+18C6+H]^+$. The complex exhibits significant mass deviation (0.59 Da) while the peptide peak is observed at correct mass (deviation within 0.2 Da). (c) CID of $[MRFA+18C6+H]^+$. The arrow represents the precursor ion. The peptide mass is correct after CID. (d) Full mass spectrum of MRFA and 18C6 in time-of-flight instrument (zoom in between 780Da and 800Da). The mass and isotope patterns agree well with predicted values. 97

Figure 4.2 Isolation windows for a series of Ac-KG_x (x = 0 to 5) peptides that have been complexed with 18C6. The predicted masses and isotope distributions are shown as red lines. The mass shifts increase as more glycine residues are added to the peptide. 99

Figure 4.3 TOF mass spectrum of the mixture of 18C6 and Ac-KG_x (x = 1 to 5) peptides. The masses and isotope distributions agree well with predicted values for the 18C6 complex with Ac-KG_x (x = 1 to 3). The 18C6 complex with Ac-KGGGG has very low relative abundance while the Ac-KGGGGG-crown complex cannot be observed. 100

Figure 4.4 (a) Precursor ion survival as a function of excitation voltage/degrees of freedom. Again additional glycine residues lead to decreased stability. (b) The trends in peptide complex stability determined independently by mass shifting and collisional activation for Ac-KGx peptides. The y axis on the right represents the reciprocal of the activation voltage (normalized by the number of degrees of freedom) to induce 50% dissociation of the crown complex by CID..... 101

Figure 4.5 Isolation windows (10Da) for a series of Ac-RG_x (x = 0 to 5) peptides that have been complexed with 18C6. The predicted masses and isotope distributions are shown as red lines. The mass shifts increase as more glycine residues are added to the peptide..... 102

Figure 4.6 Isolation windows (10 Da) for a series of G_x (x = 2 to 5) peptides that have been complexed with 18C6. The predicted masses and isotope distributions are shown as red lines. The mass shifts and peak widths are reduced compared to Ac-RG_x and Ac-KG_x peptides. 103

Figure 4.7 (a) Mass shifts for the primary isotopic peaks obtained from isolation windows of 18C6 adducts with Ac-RG_x, Ac-KG_x, and G_x peptides. Reduced intramolecular charge solvation in polyglycine leads to greater 18C6 adduct stability and smaller mass shifts. (b) The magnitude of the mass shifting is reduced if the scan speed of the instrument is decreased. 104

Figure 4.8 Lowest energy structures for GGGG, GGGG–18C6, Ac-KGGG, and Ac-KGGG–18C6 from molecular dynamics conformational searches. The hydrogen bond distances between H and O atoms are shown for each structure. The peptide structures change significantly when 18C6 is attached. The interactions between peptide backbone and side chain are stronger for Ac-KGGG than for GGGG. 106

Figure 4.9 Distributions of remaining 18C6 adducts following collisional activation of the four adduct peak from different proteins in 50/50 water/methanol. (a) Cytochrome c from charge state +9 to +15. (b) Ubiquitin from charge state +6 to +10. (c) apomyoglobin from charge state +11 to +18. In general higher charge states retain more 18C6 adducts. 109

Figure 4.10 CID spectra of cytochrome c at +9 charge state with four-adduct peaks at various excitation energies. The black arrows represent precursor ions. Peaks are labeled by "charge state-number of crown adducts". 15% NCE is sufficient to produce the bare protein. 111

Figure 4.11 CID of $[KKKKK+4(18C6)+4H]^{4+}$. The black arrow represents the precursor ion. Peaks are labeled by "charge state - number of 18C6 adducts". The loss of one crown from +4 charge state is observed yields the primary product. 112

Figure 4.12 CID spectra of cytochrome c at +14 charge state with four-adduct peaks at various excitation energies. The black arrows represent precursor ions. Peaks are labeled by "charge state-number of crown adducts". The activation of +14 charge state mainly leads to the loss of two crowns even at higher activation energies. 113

Figure 4.13 MS^3 spectrum showing collisional activation of the 14-2 peak of cytochrome c. Naked protein is not the preferred product. Peaks are labeled by "charge state - number of 18C6 adducts". The precursor ion was generated by collisional activation of the 14-4 peak. 114

Figure 4.14 CID of holomyoglobin with four crown adducts for the (a) +10 and (b) +11 charge states. Heme loss occurs primarily after 18C6 is lost, suggesting a protective function. Numbers refer to charge state and number of 18C6 adducts, i.e. 11-1 is the +11 charge state with a single 18C6 adduct. The black arrows represent the precursor ions.116

Figure 5.1 (a) Backbone structural alignment for the three variants of insulin. (b) Sequence alignment for human, porcine and bovine insulin from top to bottom, respectively. Sequence variation (white), basic residues (blue) and acidic residues (red) are highlighted. (c) ESI-MS spectrum of insulin from bovine, porcine, human and 18C6 in water. Peaks are labeled by (charge state) – (number of 18C6 adducts). The results are similar for all three proteins. 129

Figure 5.2 Sequence alignment for horse, bovine, pigeon, and yeast cytc from top to bottom, respectively. 130

Figure 5.3 (a) Backbone structures of horse, bovine, and yeast cytc. Displayed atoms represent basic and acidic sequence variations relative to horse cytc. (b) Surface electrostatic distributions derived from crystal structures. The colors represent charge polarity (positive charge, blue; negative charge, red). 131

Figure 5.4 SNAPP distributions for four variants of cytc for the (a) +9 and (b) +10 charge states. Bovine, horse, and pigeon cytc exhibit similar results, while yeast cytc has a distinct distribution..... 134

Figure 5.5 Backbone structural alignment of hen and human lysozyme. 135

Figure 5.6 Surface electrostatic distributions derived from crystal structures. The colors represent charge polarity (positive charge, blue; negative charge, red). 135

Figure 5.7 Sequence alignment of hen and human lysozyme (top and bottom, respectively). The blue stars indicate basic and acidic sequence variations. 136

Figure 5.8 SNAPP distributions for hen and human lysozyme, for the +11 and +12 charge states. 137

Figure 5.9 ESI-MS spectra of horse cytc in (a) 100% water (b) 50/50 water/methanol and (c) 50/50 water/MeOH with 1% acid. No charge state distribution shift was observed for 50/50 water/MeOH, while a significant shift was observed with 0.1% acid. 140

Figure 5.10 DESI mass spectra of (a) horse, (b) bovine, (c) pigeon, and (d) yeast cytc in 100% water and 50/50 water/methanol mixture. The yeast variant is significantly unfolded in 50% methanol. 143

Figure 5.11 SNAPP distributions of +9 charge state for horse, bovine, and pigeon cytc in water and 50/50 water/MeOH. Partial denaturation in the presence of methanol when compared to that in water. 144

LIST OF SCHEMES

Scheme 1.1 Peptide backbone fragmentation nomenclature 5

Scheme 1.2 (a) Structure of 18-crown-6 (18C6). (b) Three hydrogen bonds formed between primary amines and 18C6. 10

Scheme 2.1 Three photolabile chromophores. 20

Scheme 2.2 Mechanism for 30 Da loss from serine side chain 26

Scheme 3.1 Workflow to identify peptide isomers in a protein digestion mixture by LC-MS/MS 57

Scheme 5.1 Diagram of the Liquid DESI Source 127

LIST OF TABLES

Table 2.1 Comparison of chiral recognition factors among different dissociation methods42

Table 3.1 Identified peptide isomers from sheep eye lens digest in ovine database.....66

Table 3.2 Identified ovine crystallin proteins from bovine database 68

Table 3.3 R_{isomer} value of the identified isomer peptides for bothd RDD and CID 77

Table 5.1 Comparison of Relative Changes in SNAPP versus Sequence.....139

Table S 3.1 Source ions for R and S values 86

Table S3.2 Detailed list for all the peptides identified in sheep eye crystallin from Ovine database.....88

List of Acronyms

MS – Mass Spectrometry

m/z – Mass to Charge ratio

ESI – ElectroSpray Ionization

PTM – Post Translational Modification

NMR – Nuclear magnetic resonance

LC – Liquid Chromatography

LC-MS – Liquid Chromatography couple with Mass Spectrometry

18C6 – 18-Crown-6 ether

SNAPP – Selective Noncovalent Adduct Protein Probing

CCS – Collisional Cross Section

RDD – Radical Directed Dissociation

PD – PhotoDissociation

CID – Collision Induced Dissociation

ECD – Electron Capture Dissociation

ETD – Electron Transfer Dissociation

Cytc – cytochrome c

HDX – Hydrogen/Deuterium Exchange

MALDI – Matrix Assisted Laser Desorption/Ionization

DESI –Desorption ElectroSpray Ionization

MeOH – Methanol

NCE – Normalized Collisional Energy

DOF – Degree Of Freedom

TOF – Time Of Flight

Chapter 1

CHARACTERIZATION OF PROTEIN STRUCTURE AND PTMS BY MASS SPECTROMETRY

1.1 Protein Characterization by Mass Spectrometry

Mass spectrometry measures the mass to charge ratio (m/z) of a molecular ion. Other than the m/z of the entire molecule, the ion can be fragmented in the mass spectrometer and the masses of these fragment ions can be detected (tandem mass spectrometry). The fragmentation pattern (including both bond cleavage sites and relative abundances of fragment ions) provides structural information about the analyte molecule. Since early the 1900s, mass spectrometry has been widely used to analyze small organic molecules. In the last few decades, the rapid development of instrumentation has empowered the characterization of large biomolecules by mass spectrometry, especially after the invention of two revolutionary soft ionization techniques, electrospray (ESI)[1] and matrix assisted laser desorption ionization (MALDI).[2]

Proteins, as the most abundant biomolecules in living organisms, play a wide range of functional roles, including catalyzing reactions, structural support, transportation, storage, signaling, etc. Although proteins are made of only twenty types of common amino acids, they can adopt a wide variety of structures and conformations, which determines the protein functionality. Mass spectrometry is one of the major techniques used to characterize protein structures. Generally, the identification of proteins by mass spectrometry is achieved by comparing the masses of the proteolytic peptides or their tandem mass spectra with those theoretically calculated from proteome or genome

database (bottom-up proteomics).[3] By coupling liquid chromatography (LC) and mass spectrometry (LC-MS), thousands of proteins can be identified in a complex biological system such as cell lysate.[4] Merely protein identification, however, is only the first step in proteomics studies. With isotope labeling, or tandem mass tags, or label-free techniques, target peptides or proteins can also be quantified in unknown samples, providing another important dimension of information. Furthermore, mass spectrometry can also reveal protein tertiary and quaternary structures via labeling techniques and limited digestion.[5] Nevertheless, full protein characterization remains challenging due to the intricacy of the proteome and the complexity of protein structures. The purpose of this dissertation is to develop mass spectrometry based techniques to identify/quantify subtle post translational modifications and characterize protein tertiary structures.

1.2 Post Translational Modification Analysis

One of the essential mechanisms that modulates protein structure and function is post translational modification (PTM). In biological systems, a protein is synthesized via the translation of mRNA and different amino acids are incorporated based on the genomic information. After translation, some proteins are modified by enzymatic reactions before achieving their final functionality. These modifications are known as post-translational modifications. This explains why the human proteome is much greater than the human genome. Hundreds of PTMs are known to occur physiologically, including phosphorylation, glycosylation, acetylation, disulfide bonds, deamidation, methylation, etc. All these PTMs have a remarkable influence on the function of a protein. Unlike the protein primary structure, PTMs cannot be directly determined from genetic information.

Reliable methods for protein characterization and PTM analysis are therefore needed to accurately identify the modification type and location.

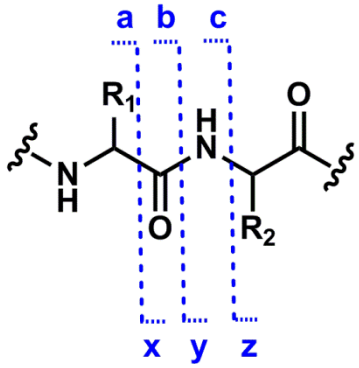
PTMs can be divided into two broad classes, enzyme-catalyzed modifications and spontaneous modifications. The addition and elimination of the majority of PTMs on a protein are catalyzed by enzymes prior to the protein arrival of the final destination in the cell. Usually the enzyme catalyzed modifications are necessary for the protein function. However, some PTMs occur spontaneously *in vivo*, which can result in damage to the protein structures and the cell.[6] These non-enzymatic modifications are relatively abundant in long-lived proteins, such as enamelin from teeth, crystallin from lens and collagen from bones.[7] For example, free radicals can react with proteins and lead to oxidation of several amino acids such as methionine and cysteine. During the aging process these oxidative modifications accumulate, which are detrimental to the protein function.[8] It is believed that oxidative stress is a major cause of several diseases, especially age-related illness, such as cancer, Alzheimer's disease, Parkinson's disease, diabetics, etc.[9] Another example is disulfide bonding, which can also occur upon aging. The unexpected disulfide bonds can cause misfolding and aggregation of protein structure, leading to dysfunctionality.[10] Therefore, analysis of the age-related PTMs is necessary and important for understanding the pathology of aging.

PTM identification using mass spectrometry is relatively straight forward: the amino acid sequence of a peptide can be obtained from subtracting the masses of adjacent product ions[11] and chemical modifications can be directly observed by mass shifts.[12]

Isomerization and epimerization, however, are exceptions. Epimerization refers to the process of inverting the chiral center of one amino acid in a peptide or protein. Although almost every amino acid in the eukaryotic proteins exists in the L form, D-amino acid containing peptides are known to display important biological activities.[13] Aspartic acid isomerization refers to a spontaneous chemical reaction that leads to the formation of several isomeric forms of aspartic acid through a cyclic succinimide intermediate.[14] Both epimerization and isomerization can significantly affect protein structure and function, and are also associated with numerous diseases. One notable example is beta-amyloid; chiral inversion of Ser 26 in the peptide can result in a toxic truncated fragment which is related to the pathology of Alzhimer's diseases.[15,16] In addition, the accumulation of D-amino acid and iso-Asp in proteins with increasing age has been detected in various human tissues such as eye lenses, bones, brain, teeth, etc.[17-19] Especially, isomerization in the eye lens protein can induce protein aggregation and lens opacification, leading to cataract disease. [20,21] The detection of these modifications is very difficult due to the identical mass between the original and modified peptides. Comparing to other age related PTMs (deamination, disulfide bonds, oxidation, etc.), isomerization and epimerization have been reported much less in the literature.

As previously mentioned, the fragmentation pattern of a peptide upon dissociation in the mass spectrometer is structurally relevant. A variety of fragmentation techniques have been reported to discriminate peptide epimers or isomers.[22] Typically the fragmentation is achieved by collision induced dissociation (CID).[23] In CID, the positively charged target ions undergo repeated collision with an inert collision gas until fragmentation occurs. The

dissociation pathway is related to the proton mobility and the cleavages are strongly influenced by peptide sequences and the positions of charged side chains.[24] Recently, electron capture dissociation (ECD)[25] and electron transfer dissociation (ETD)[26] have been invented and shown to be powerful techniques for protein characterization and PTM analysis. In these experiments, free electrons or radical anions are introduced to interact with a multiply charged target peptide/protein, creating radicals and breaking backbones. Unlike CID, during which the loss of PTMs occurs frequently, ECD and ETD fragments retain labile PTMs such as phosphorylation and glycosylation. Typically, CID cleaves the peptide amide backbone and yields b and y ions, while the radical dissociation methods generate a, c and z ions. Generic nomenclature for peptide backbone fragmentation is shown in Scheme 1.1.



Scheme 1.1. Peptide backbone fragmentation nomenclature

Another radical based dissociation method, radical directed dissociation (RDD) has been recently applied to elucidate protein tertiary structures.[27] In RDD, peptides/proteins are covalently labeled by a photolabile chromophore and irradiation of the labeled ion with a UV laser generates a radical at the modification site. The radical can migrate along the peptide backbone or side chains and then lead to peptide fragmentation upon

collisional activation. The RDD fragmentation pattern is very structurally sensitive for several reasons: First, the radical migration process is related to the protein structure and the RDD fragmentation occurs in the vicinity of the radical.[27, 28] Secondly, different from the mobile protons in CID and the radicals in ECD or ETD, the position of the initial radical in RDD is fixed based on the chemical modification.[29] Thirdly, only one radical is present in RDD, while several mobile protons can be responsible for CID fragmentation in multiply charged peptides.

Described in Chapter 2 is the application of RDD for the discrimination of peptide epimers. Based on the structural sensitivity, it is demonstrated that RDD can successfully distinguish the subtle structural change between two peptide epimers and provides better chiral discrimination ability than all the other previously reported dissociation methods. However, Chapter 2 only examined standard peptide epimers. The significance of this project lies in the possibility of detecting amino acid isomerization in proteins. This could reveal how isomerization affects protein function and explore the correlation between isomerization and certain diseases. Chapter 3 describes "isomer proteomics", which is the study of protein epimerization and isomerization in a complex biological sample. The primary difference between isomer proteomics and traditional proteomics is that each m/z of interest must be sampled multiple times to allow for comparison of tandem mass spectra. The approach was illustrated by the examination of long-lived crystallin proteins extracted from a sheep eye lens in Chapter 3. Numerous sites of epimerization and isomerization were identified from the whole lens extract using a combination of RDD and CID following separation by liquid chromatography. The results demonstrated for the

first time using both LC and MS for the detection of peptide isomers in a large-scale proteomics analysis.

1.3 Protein Tertiary Structures

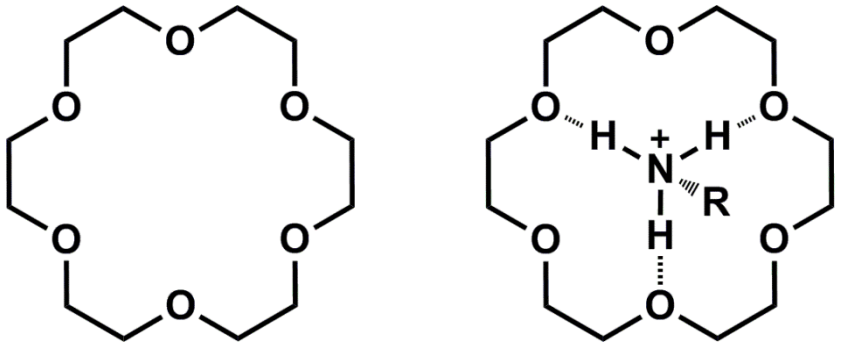
Several methods have been applied to investigate protein structures. For example, Edman degradation[30] has been widely used for peptide sequencing, with the following identification of the single amino acid by UV spectroscopy. For secondary structures, circular dichroism spectroscopy is a common method to determine alpha helix, beta-sheet and random coil structures in a protein.[31] The protein three-dimensional structures can be accurately determined by nuclear magnetic resonance (NMR) and X-ray crystallography. However, there are limitations for each method. Edman degradation can only provide information about primary structure and the procedure is very time-consuming. X-ray crystallography yields unambiguous information about protein structures but the intramolecular interactions in solution can be overlooked in the crystal structure due to the noncovalent intermolelcular interactions with adjacent proteins. NMR is capable of detecting very fine structural variation or highly dynamic states, but the application is restricted due to the complicated sample preparation process.

Mass spectrometry has been demonstrated to be a powerful technique for the characterization of protein structures. Covalent labeling techniques are widely used to study protein solvent accessibility, protein-protein interaction, protein-ligand binding, etc. Hydrogen-deuterium exchange (HDX) is one of the most common covalent labeling methods. In HDX, proteins are dissolved in deuterated solvents during a certain period of

time which result in the substitution of the covalent bonded hydrogens on the protein to deuterium atoms.[32] The labile hydrogen atoms such as those on side chains exchange much faster than those on amide backbones and the exchange rates can be monitored by mass spectrometry. Proteins in compact structures have relative less solvent accessibility and a smaller amount exchanged hydrogens than unfolded states. Combining with LC, HDX-MS has been applied for probing protein structures under various circumstances such as conformational changes upon ligand binding,[33] large macromolecular assembly,[34] protein aggregation,[35] structural dynamics, etc.[36,37] Besides HDX, amino acid-specific covalent labeling is another valuable technique to examine protein structures.[38] In these experiments, certain amino acid side chains (primary amine, carboxylic acid, serine, tyrosine, etc.) of a protein are chemically modified, following by digestion and LC-MS for the identification of the modification site. The reactivity of amino acids in proteins depends on the solvent accessibility of that certain residue and thus the site-specific structural information can be obtained. Additionally, in the last two decades, ion mobility coupled to mass spectrometry has provided a new area in the structural analysis of large protein complex.[39,40] Ion mobility separates protein ions based on their ability to travel through a drift tube. The drift tube is applied with electric field which affects the drift velocity based on protein charges. The tube is also filled with buffer gas to induce collisions with protein ions.[41] Under the influence of both electric fields and collisional gas, the drift time of a protein correlates with the mass, charge, shape and size. A compact protein that has a smaller collision cross section (CCS) will have a shorter drift time. Although ion mobility cannot directly provide protein structural information in

detail, this method opens the way for comprehensive structural analysis for large protein assemblies and aggregated protein systems.[42]

A critical concern for protein structural characterization by mass spectrometry is that the experiments are conducted in the gas phase, while these biomolecules carry out their functions in aqueous solution. Therefore, it is necessary to ensure that the solution phase structures are preserved during the transition. One of the most common ionization methods is electrospray ionization (ESI).[1] During electrospray, high voltage is applied to the liquid sample and a mist of highly charged droplets is generated. The size of these droplets is reduced by solvent evaporation or coulomb explosion, finally yielding analyte ions.[43] As a gentle ionization method, ESI does not fragment the sample molecules, which allows weakly bonded complex to be transferred and detected.[44] However, detailed mechanism of the ESI process at a molecular level is not known.[45] There is continuous debate about the degree to which protein structure is preserved in the gas phase after the transition by ESI. Recent work with ion mobility experiments[46] revealed that the interactions between charged lysine side chain and backbone carbonyls can dramatically disturb the tertiary structure of a protein in the absence of solvent. However, with the charged residues capped with 18-Crown-6, the protein solution phase structures are preserved for certain charge states.



Scheme 1.2 (a) Structure of 18-crown-6 (18C6). (b) Three hydrogen bonds formed between primary amines and 18C6.

18-crown-6 (18C6) is a small molecule that binds cationic species as shown in Scheme 1.2. When mixing with proteins or peptides in solution, 18C6 can noncovalently attach to basic sites (Lys/Arg side chains and N-terminus) on the surface of the protein/peptide during ESI.[47] Such attachment has been applied for peptide charge stabilization,[47] radical generation,[28] protein structural analysis, [48] etc. Chapter 4 demonstrates using mass spectrometry that 18C6 can protect protein solution phase native structure by solving the charged sites. The stability of the 18C6 attachment to the peptide is highly dependent on the peptide sequence and structure. Experiments with model peptides with different cation groups and different length have shown that the 18C6 adduct stability is decreased if more intramolecular charge complexation is involved. Examination of the loss of 18C6 adducts from proteins following collisional activation reveals that lower charge states lose the most 18C6. These results are in agreement with the ion mobility experiments that upon the addition of 18C6, protein adopt more compact and native-like structures in the gas phase than in the absence of 18C6, indicating that the non-covalent attachment

of 18C6 to protein can solvate the positively charged side chain in a similar way as solvent molecules.[46]

In this regard, 18C6 may serve the function of protecting solution phase protein structure. Accordingly, an 18C6 based technique, known as SNAPP (selective noncovalent adduct protein probing) has been developed for the examination of protein solution phase structure by mass spectrometry.[48,49] In this method, 18C6 is used as a recognition molecule to non-covalently attach solvent accessible basic residues in proteins via three hydrogen bonds (Scheme 1.2b). The interaction is weak in solution, but becomes strong in the gas phase during the ESI process as the solvent molecules rapidly evaporated. The unique pattern of the number and relative intensities of protein-18C6 complexes is called SNAPP distribution. Previous results have demonstrated that the 18C6 attachment to charged side chains is interfered by intramolecular interactions such as hydrogen bonds and salt bridges.[50] Thus, the SNAPP distributions do not solely rely on the number of basic residues or the protein sequence, but also probe the overall protein conformation.

SNAPP is complementary to other common existing mass spectrometry based techniques for protein structural elucidation. For example, the SNAPP distribution is very sensitive to the local chemical environment for the basic residue side chains; in HDX, however, the side chain information is lost due to the back exchange effect. Covalent chemical modification of specific amino acids is suspect to induce protein structural perturbation, while SNAPP experiments rely on the noncovalent 18C6 attachment to the protein basic residues, which does not affect the overall protein structure. Although

SNAPP cannot provide detailed information about specific amino acid side chains, this technique is highly suitable for perturbing the protein structural change in different environments (adding organic solvent, interacting with ligands, etc).

Described in Chapter 5 is the application of SNAPP technique for evaluating the structural consequences of point mutations in naturally occurring sequence variants from different species. Results in Chapter 5 demonstrate that mutations that lead to perturbations to the electrostatic surface structure of a protein affect noncovalent attachment and are easily observed with SNAPP. Comparing with the other techniques described previously, SNAPP experiments are relatively easy to perform, require minimal sample consumption, and can provide a facile route for protein surface structure analysis.

References

[1] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** *246*, 64-71.

[2] Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10000 Daltons. *Anal. Chem.* **1988,** *60*, 2299-2301.

[3] Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., III, Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113*, 2343-2394.

[4] Yates, J. R.; Ruse, C. I.; Nakorchevsky, A., Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **2009**, *11*, 49-79.

[5] Kriwacki, R.; Reisdorph, N.; Siuzdak, G., Protein structure characterization with mass spectrometry. *Spectroscopy- Int. J.* **2004,** *18*, 37-47.

[6] Cloos, P. A. C.; Christgau, S., Post-translational modifications of proteins: implications for aging, antigen recognition, and autoimmunity. *Biogerontology* **2004,** *5*, 139-158.

[7] Ritz-Timme, S.; Collins, M. J., Racemization of aspartic acid in human proteins. *Ageing Res. Rev.* **2002,** *1*, 43-59.

[8] Khansari, N.; Shakiba, Y.; Mahmoudi, M., Chronic inflammation and oxidative stress as a major cause of age-related diseases and cancer. *Recent Pat Inflamm Allergy Drug Discov.* **2009,** *3*, 73-80.

[9] Uttara, B.; Singh, A. V.; Zamboni, P.; Mahajan, R. T., Oxidative Stress and Neurodegenerative Diseases: A Review of Upstream and Downstream Antioxidant Therapeutic Options. *Curr. Neuropharmacol.* **2009,** *7*, 65-74.

[10] Shringarpure, R.; Davies, K. J. A., Protein turnover by the proteasome in aging and disease. *Free Radic. Biol. Med.* **2002,** *32*, 1084-1089.

[11] Standing, K. G., Peptide and protein de novo sequencing by mass spectrometry. *Curr. Opin. Struct. Biol.* **2003,** *13*, 595-601.

[12] Seo, J.; Lee, K. J., Post-translational modifications and their biological functions: Proteomic analysis and systematic approaches. *J. Biochem. Mol. Biol.* **2004,** *37*, 35-44.

[13] Kreil, G., D-amino acids in animal peptides. *Annu. Rev. Biochem.* **1997,** *66*, 337-345.

[14] Geiger, T.; Clarke, S., Deamidation, isomerization, and racemization and at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *J. Bio. Chem.* **1987,** *262*, 785-794.

[15] Kaneko, I.; Morimoto, K.; Kubo, T., Drastic neuronal loss in vivo by beta-amyloid racemized at Ser(26) residue: conversion of non-toxic [D-Ser(26)]beta-amyloid 1-40 to toxic and proteinase-resistant fragments. *Neuroscience* **2001,** *104*, 1003-1011.

[16] Kubo, T.; Nishimura, S.; Kumagae, Y.; Kaneko, I., In vivo conversion of racemized beta-amyloid ([D-Ser(26)]A beta 1-40) to truncated and toxic fragments ([D-Ser(26)]A beta 25-35/40) and fragment presence in the brains of Alzheimer's patients. *J. Neurosci. Res.* **2002,** *70*, 474-483.

[17] Ritz-Timme, S.; Collins, M. J., Racemization of aspartic acid in human proteins. *Ageing Res. Rev.* **2002,** *1*, 43-59.

[18] Fujii, N., D-amino acid in elderly tissues. *Biol. Pharm. Bull.* **2005,** *28*, 1585-1589.

[19] Cloos, P. A. C.; Fledelius, C., Collagen fragments in urine derived from bone resorption are highly racemized and isomerized: a biological clock of protein aging with clinical potential. *Biochem. J.* **2000,** *345*, 473-480.

[20] Moreau, K. L.; King, J. A., Protein misfolding and aggregation in cataract disease and prospects for prevention. *Trends Mol. Med.* **2012,** *18*, 273-282.

[21] Sharma, K. K.; Santhoshkumar, P., Lens aging: Effects of crystallins. *Biochim. Biophys. Acta* **2009,** *1790*, 1095-1108.

[22] Hurtado, P. P.; O'Connor, P. B., Differentiation of isomeric amino acid residues in proteins and peptides using mass spectrometry. *Mass Spectrom. Rev.* **2012**, *31*, 609-625.

[23] Wells, J. M.; McLuckey, S. A., Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol,* **2005**, *402*, 148-185.

[24] Downard, K. M.; Biemann, K., The effect of charge state and the localization of charge on the collision induced dissociation of peptide ions. *J. Am. Soc. Mass Spectrom.* **1994,** *5*, 966-975.

[25] Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **1998,** *120*, 3265-3266.

[26] Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *P. Natl. Acad. Sci. U S A* **2004,** *101*, 9528-9533.

[27] Ly, T.; Julian, R. R., Elucidating the tertiary structure of protein ions *in vacuo* with site specific photoinitiated radical reactions. *J. Am. Chem. Soc.* **2010**, *132*, 8602-8609.

[28] Sun, Q. Y.; Nelson, H.; Ly, T.; Stoltz, B. M.; Julian, R. R., Side chain chemistry mediates backbone fragmentation in hydrogen deficient peptide radicals. *J. Proteome Res.* **2009,** *8*, 958-966.

[29] Ly, T.; Julian, R. R., Residue-specific radical-directed dissociation of whole proteins in the gas phase. *J. Am. Chem. Soc.* **2008,** *130*, 351-358.

[30] Edman, P., Method for determination of the amino acid sequence in peptides. *Acta Chem. Scand.* **1950,** *4*, 283-293.

[31] Whitmore, L.; Wallace, B. A., Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* **2008,** *89*, 392-400.

[32] Katta, V.; Chait, B. T., Hydrogen deuterium exchange electrospray ionization mass spectrometry – a method for probing protein conformational changes in solution. *J. Am. Chem. Soc.* **1993,** *115*, 6317-6321.

[33] Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; He, Y. J.; Hendrickson, C. L.; Marshall, A. G.; Griffin, P. R., Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry. *Anal. Chem.* **2006,** *78*, 1005-1014.

[34] Tuma, R.; Coward, L. U.; Kirk, M. C.; Barnes, S.; Prevelige, P. E., Hydrogen-deuterium exchange as a probe of folding and assembly in viral capsids. *J. Mol. Biol.* **2001,** *306*, 389-396.

[35] Zhang, Y.; Rempel, D. L.; Zhang, J.; Sharma, A. K.; Mirica, L. M.; Gross, M. L., Pulsed hydrogen-deuterium exchange mass spectrometry probes conformational changes in amyloid beta (A beta) peptide aggregation. *P. Natl. Acad. Sci. U S A* **2013**, *110*, 14604-14609.

[36]Engen, J. R., Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS. *Anal. Chem.* **2009,** *81*, 7870-7875.

[37] Englander, J. J.; Del Mar, C.; Li, W.; Englander, S. W.; Kim, J. S.; Stranz, D. D.; Hamuro, Y.; Woods, V. L., Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proc. Natl. Acad. Sci. U S A* **2003,** *100*, 7057-7062.

[38] Mendoza, V. L.; Vachet, R. W., Probing protein structure by amino acid- specific covalent labeling and mass spectrometry. *Mass Spectrom. Rev.* **2009,** *28*, 785-815.

[39] Zhong, Y.; Hyung, S.-J.; Ruotolo, B. T., Ion mobility-mass spectrometry for structural proteomics. *Expert Rev. Proteomics* **2012**, *9*, 47-58.

[40] Hoofnagle, A. N.; Resing, K. A.; Ahn, N. G., Protein analysis by hydrogen exchange mass spectrometry. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 1-25.

[41] Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E., The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* **2014**, *6*, 281-294.

[42] Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. R., Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* **2010,** *39*, 1633-1655.

[43] Wortmann, A.; Kistler-Momotova, A.; Zenobi, R.; Heine, M. C.; Wilhelm, O.; Pratsinis, S. E., Shrinking droplets in electrospray ionization and their influence on chemical equilibria. *J. Am. Soc. Mass Spectrom.* **2007,** *18*, 385-393.

[44] Loo, J. A., Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **1997,** *16*, 1-23.

[45] Kebarle, P.; Verkerk, U. H., Electrospray: From ions in solution to ions in the gas phase, what we know now. *Mass Spectrom. Rev.* **2009,** *28*, 898-917.

[46] Warnke, S.; von Helden, G.; Pagel, K., Protein structure in the gas phase: the influence of side-chain microsolvation. *J. Am. Chem. Soc.* **2013,** *135*, 1177-1180.

[47] Julian, R. R.; Beauchamp, J. L., Site specific sequestering and stabilization of charge in peptides by supramolecular adduct formation with 18-crown-6 ether by way of electrospray ionization. *Int. J. Mass Spectrom.* **2001,** *210*, 613-623.

[48] Ly, T.; Julian, R. R., Using ESI-MS to probe protein structure by site-specific noncovalent attachment of 18-crown-6. *J. Am. Soc. Mass Spectrom.* **2006,** *17*, 1209-1215.

[49] Ly, T.; Julian, R. R., Protein-Metal Interactions of Calmodulin and alpha-Synuclein Monitored by Selective Noncovalent Adduct Protein Probing Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2008,** *19*, 1663-1672.

[50] Liu, Z. J., Cheng, S. J., Gailie, D. R., and Julian, R. R. Exploring the mechanism of selective noncovalent adduct protein probing mass spectrometry utilizing site-directed mutagenesis to examine ubiquitin. *Anal. Chem.* **2008**, *80*, 3846−3852.

Chapter 2

DISCRIMINATING D-AMINO ACID CONTAINING PEPTIDES BY RADICAL DIRECTED DISSOCIATION USING MASS SPECTROMETRY

2.1 Introduction

Although eukaryotic proteins and peptides normally consist solely of L-amino acids, peptides containing D-amino acids with known biological activities have also been discovered in amphibians, invertebrates, and mammals.[1,2] Inversion of the chirality of a single amino acid from L to D yields a peptide epimer which can exhibit significantly altered biological function. For example, the deltorphins[3] and dermorphin[4] are peptides found in the skin of a South American frog, where the second amino acid is converted to the D-isomer by post translational modification. These peptides show high affinity and selectivity for opioid receptors, and the D-configurations are essential for biological response. In addition, D-Ser and D-Asp have both been detected in human tissues where it is hypothesized that spontaneous racemization occurs with aging.[5] The presence of the D-amino acids in proteins has also been correlated with abnormal protein aggregation or unfolding. For example, the racemization of serine residues in β-Amyloid protein may be involved in the degeneration of neuronal cells and related to the pathology of Alzheimer's disease.[6] In each of these examples, inversion of chirality is inconspicuous in an analytical sense because it does not result in a change in the molecular weight and cannot be easily identified by standard amino acid sequencing methods. Therefore,

reliable and sensitive methods for identifying and characterizing peptide epimers are needed.

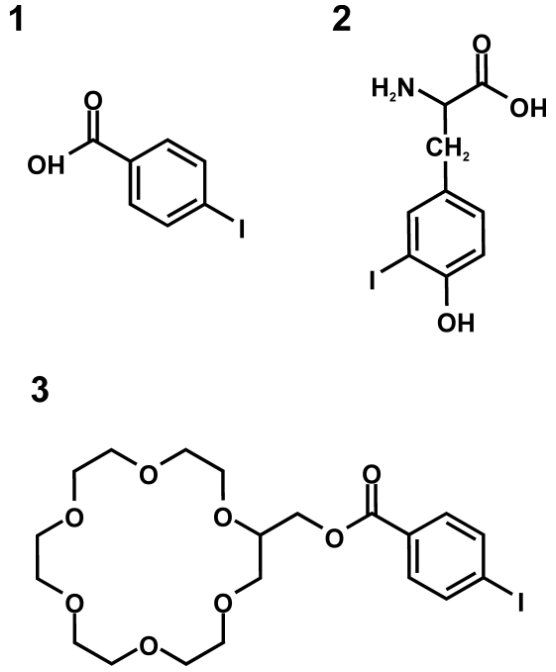
Methods for detecting the chirality of amino acid residues in peptides include liquid/gas chromatography,[7] NMR,[8] and stereoselective enzymatic digestion.[9] Frequently these methods require chiral reagents or chiral stationary phases which are not widely available. Although mass spectrometry (MS) is a common and very powerful method for carrying out proteomics research, differentiating peptide epimers with MS is challenging because the chiral inversion does not lead to any change in mass/charge. Several strategies have been developed to overcome this difficulty. Tao et al. reported enantiomeric differentiation of single D- and L- amino acids with MS via the kinetic method.[10] Hydrogen/Deuterium exchange has been combined with tandem MS to identify the racemization sites in immunoglobulin proteins.[11] Chiral discrimination can also be achieved by comparing the relative abundances of fragments obtained following individual activation of each epimer. For this method to be successful both epimers must be available and the relative intensity of some of the fragments must be substantially different between the two fragmentation spectra. Both collision induced dissociation (CID)[12,13,14] and electron capture dissociation (ECD)[15,16] have been successfully employed to distinguish the presence of D amino acids in peptides using this approach; however, discrimination is difficult with both methods. Fragmentation in CID occurs largely through proton initiated pathways. Due to proton mobility and multiplicity, structural sensitivity in CID can be limited. In the case of ECD, structural sensitivity is primarily

dependent on the noncovalent hydrogen bonding network, which is poor in short peptides.

An alternative MS based dissociation method, radical directed dissociation (RDD) has been used previously to elucidate protein tertiary structure [17] or for site specific cleavage.[18, 19] In these experiments, proteins or peptides are labeled by chromophores that include a carbon-iodine or carbon-sulfur bond. Irradiation of the precursor ion with a 266nm laser homolytically cleaves the C-I or C-S bond and yields a radical specifically at the modification site. Collisional activation of this radical species leads to various backbone dissociation and side chain loss products. Migration of the radical dictates which fragments will be observed and is highly dependent on peptide structure.[20] Consequently, RDD fragmentation patterns are typically very structurally informative, which should provide a greater potential for identifying subtle changes to peptide structure relative to ECD and CID.

Another advantage of RDD is that there are numerous, nonequivalent methods for generating radical peptides. Both covalent and noncovalent chemistry can be used; several chromophores relevant to the present work are shown in Scheme 2.1.[21] The peptide N-terminus or lysine side chains can be modified by **1**. Alternatively, if the peptide contains a tyrosine residue, then single iodine can be easily installed as shown in **2**. A third method combines the recognition ability of 18-crown-6 (18C6) ether with a photolabile radical precursor (see **3**). The 18C6 portion of **3** can noncovalently attach to basic sites during electrospray by simply adding **3** to the solution containing the peptide.[22, 23] Photoactivation of the complex with 266 nm light leads to the loss of $I^{\bullet}$ and

generates a phenyl radical, which can subsequently migrate to the peptide (or not) prior to dissociation of the noncovalent complex following subsequent collisional activation. Both the radical transfer probability and the subsequent fragmentation of the radical peptide are potentially sensitive to peptide conformation. To simplify reading this manuscript, traditional one letter codes are used for amino acids with underlined residues representing sites of epimerization.



Scheme 2.1 Three photolabile chromophores.

Herein, eight peptides containing naturally occurring D-alanine, D-serine or D-aspartic acid residues were examined by RDD and compared with all L-amino acid epimers. Several different photolabile radical precursor chemistries were used, and in all cases excellent discrimination between epimers was obtained by at least one method. It is demonstrated that RDD can easily discriminate between epimers in various charge states,

including multiply charged ions and positively and negatively charged ions. RDD can successfully discriminate peptide epimers of various chain lengths (4-12 residues were tested herein). In addition, for some peptides radical transfer efficiency from **3** (following photoactivation) was shown to easily discriminate between epimers, illustrating an alternate route for identification that does not involve fragmentation. The results clearly demonstrate that RDD provides a diverse new array of powerful tools for detecting subtle changes to peptide structure caused by chiral inversion of a single amino acid.

2.2 Experimental Methods

2.2.1 Materials

Organic solvents and reagents were purchased from Sigma-Aldrich (St. Louis, MO) or Acros Organics (Geel, Belgium) and used without further purification. Water was purified to 18.2 MΩ using a Millipore 147 (Billerica, MA) Direct-Q system. The D amino acids and resin were purchased from Ana Spec (Fremont, CA). Peptides were purified by reversed-phase HPLC using a Biobasic C18 column (Thermo Scientific) prior to quantitative analysis.

2.2.2 Peptide Synthesis

All the Peptides except D-alanine Y<u>A</u>FDVVG-NH_2 (deltorphin C) were synthesized manually using standard fmoc procedures[24] with Rink Amide Resin or Wang Resin being employed as the solid support. Amino acids with protected side chains were used when needed. Deltorphin C was purchased from American peptide company (Sunnyvale, CA).

2.2.3 Radical Precursor Synthesis

Procedures to synthesize the three labeling compounds shown in Scheme 2.1 are as follows:

4-iodobenzoic acid modification (**1**). After synthesis of a peptide chain was completed, but prior to cleavage from the resin, half of the sample was further modified to add 4-iodobenzoic acid to the N-terminus of the peptide. Peptides modified in this fashion have a superscripted **1** on the N-terminal residue (i.e. $^{\mathbf{1}}$DVGSNK). To the reaction vessel 5mL dimethylformamide (DMF) was added and bubbled with N_2 for 15 minutes to swell the resin, and then the DMF was drained. Afterwards, a mixture of equal moles of tetramethylaminium hexafluorophosphate (HCTU) and 4-iodobenzoic acid dissolved in 2.5mL of DMF were added to reaction vessel and bubbled with N_2 for 6 minutes. The resin was rinsed with 5mL of methanol and dried for 20 minutes. Finally, the resin was transferred to a glass vial and the peptide was cleaved.

Iodination of tyrosine in peptides (**2**). Peptides were iodinated at room temperature using a previously published method.[25] Sodium iodide and chloramine-T were used as the iodine source and the oxidizing reagent, respectively. After 1 min reaction, sodium metabisulfite was added to the solution to quench the iodination reaction. Stoichiometric quantities of iodination reagents were used to avoid any side reactions (1:1:1:2 molar ratio of peptide/sodium iodide/chloramine-T/sodium metabisulfite). Iodinated peptides were purified by removal of excess iodination reagents via peptide trap (Michrom Bioresource Inc.).

2-(hydroxymethyliodobenzoylester-18-Crown-6 ether (**3**) was prepared by the previously reported procedure.[21] 0.50 mmol DCC in 5.0 mL dioxane was added to a 50 mL round bottom flask containing 0.50 mmol of 4-iodobenzoic acid and 0.50 mmol 2-hydroxymethyl-18-crown-6 ether. A catalytic amount of DMAP (~10 mg) was added. After a 12 h reaction period, a crystalline hair-like precipitate was observed. The precipitate was removed by filtration. The filtrate was then evaporated over nitrogen.

2.2.4 Mass Spectrometry and Radical Directed Dissociation

Solutions containing ~10 μM peptide in 50/50 methanol/water were infused into an LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, San Jose,) with a standard ESI source. No acid or buffer was added to the solutions for positive mode. 0.5% ammonium hydroxide or 1mM ammonium acetate buffer was added to solutions to enhance the signal for negative mode. The mass spectrometer has been modified with the addition of a flash-pumped Nd:YAG laser (Continuum, Santa Clara, CA) to the back of the instrument. The laser fires fourth harmonic pulses (266nm) through a quartz window at the back of the instrument into the ion trap. Photodissociation (PD) of the covalently/noncovalently modified peptides always produces the loss of iodine as the most abundant peak. Further MS^3 experiments were performed by re-isolation followed by and collision induced dissociation (CID) of the radical species. The isolation window width of MS^2 and MS^3 experiments was set to 3-4 Da. For noncovalent radical delivery, **3** was added at 5 times the peptide concentration to the solution prior to electrospraying, and the MS^n type experiments described below were performed on the resulting noncovalent peptide-crown complexes.

2.3 Results and Discussion

2.3.1 D-serine Peptides

DVGSNK is a small β-amyloid peptide fragment. Racemization at the serine residue has been found to be an age-dependent process and related to the pathology Alzheimer's disease. Figure 2.1a shows the PD spectra of the two serine epimers of $^{\mathbf{1}}$DVG<u>S</u>NK-NH_2 ($^{\mathbf{1}}$X denotes N-terminal modification of X by **1** where X is any amino acid). The most abundant peak for both peptides is loss of 127 Da from the precursor ion, which corresponds to homolytic cleavage of the C-I bond. In addition to the loss of iodine, an intense peak resulting from a loss of 30 Da from the radical peptide is observed for the D-serine peptide. The matching loss is barely observed in the L-serine epimer. Chiral inversion at serine is the only difference between the two peptides, and the side chain of serine can undergo a loss of 30 Da by RDD via the mechanism shown in Scheme 2.2. Tandem MS^3 experiments (Figure 2.1b) confirm that the loss of 30 Da originates from the serine side chain. Previous results have demonstrated that the backbone and side chain fragments resulting from PD (without subsequent collisional activation) are typically localized to the vicinity where the radical is initially produced.[18,19] In this case, the D-serine side chain is fortuitously in close proximity to the N-terminus of peptide, while the L-serine side chain is not. Excellent discrimination between the two epimers is therefore afforded.

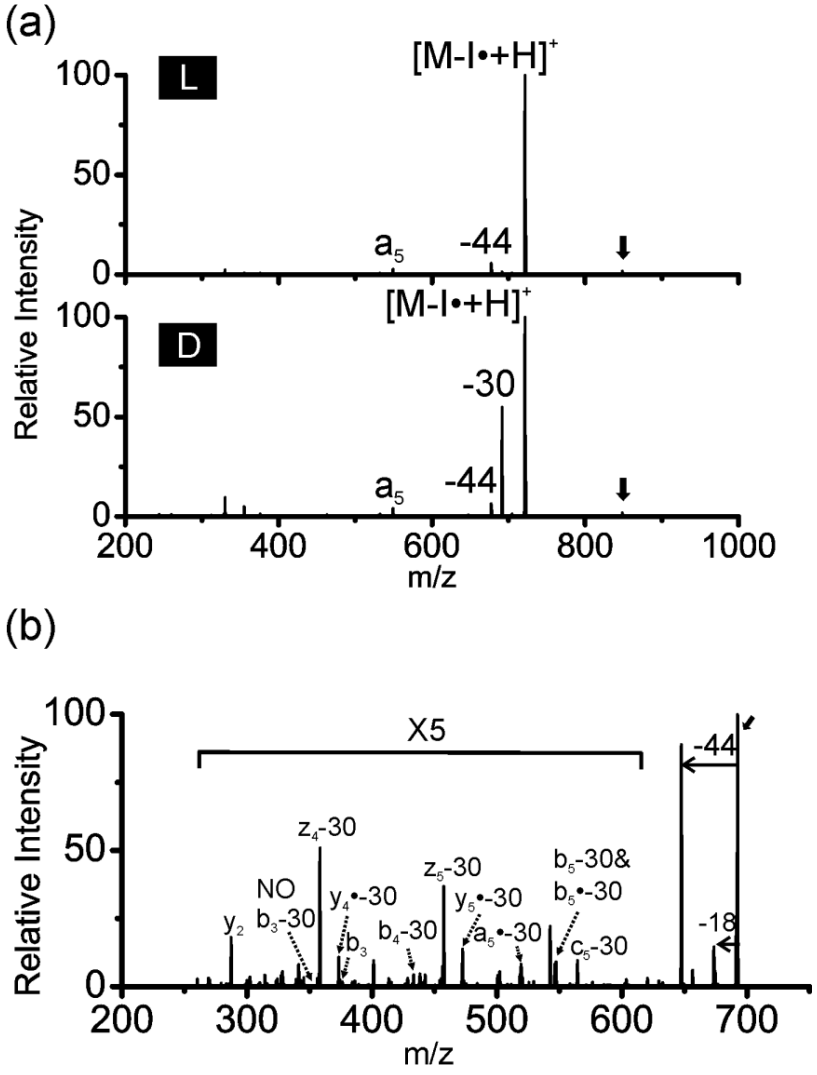
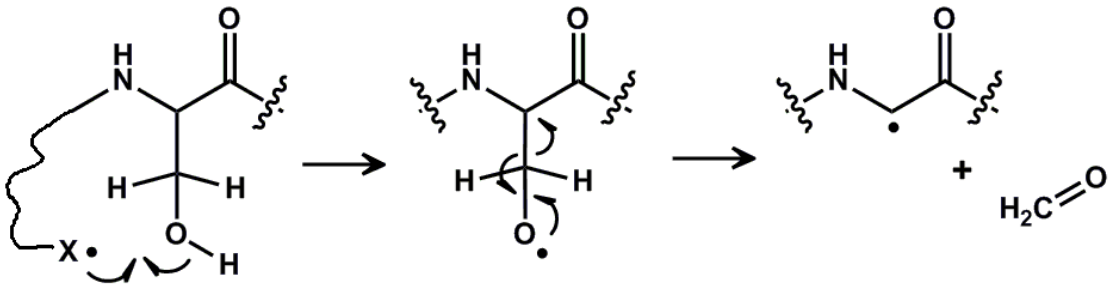


Figure 2.1 (a) PD of the two serine epimers of ^{1}DVG<u>S</u>NK-NH_2. The major difference is the -30Da serine side chain loss. (b) CID of -30 loss from PD of [^{1}DVG<u>S</u>NK-NH_2 + H]$^+$ (D-ser epimer). A series of b_3, b_4-30, b_5-30 fragments are observed, while b_3-30 ion is not present in the spectrum, which verifies that the -30 loss is from serine.



Scheme 2.2. Mechanism for 30 Da loss from serine side chain

The MS^3 spectra generated by collisional activation of the radical peptides created by the loss of 127 Da are shown in Figure 2.2a. Loss of 44 Da is prevalent for both epimers and minor backbone fragmentations are also observed in roughly equal abundances. Interestingly, the loss of 30 Da from serine side chain is not present in either spectrum. It is likely that the fraction of radical which migrates promptly to the serine side chain accounts for the -30 Da observed in the PD spectrum, and that subsequent collisional activation favors migration to more thermodynamically favorable destinations. This phenomenon has been observed previously where PD and PD-CID have yielded different fragmentation patterns.[18] Importantly, the PD-CID spectra of the two epimers also show significant differences in the relative losses of NH_3 and H_2O. The D-serine peptide is dominated by water loss, which is a minor loss for the L-serine peptide.

It is useful to quantitatively compare the degree of difference obtained by comparison of epimers in this fashion. An R value dependent on the ratios of observed peaks similar to that originally reported for chiral recognition by Tao and coworkers[10] is typically used. We will report R_{Chiral} values as defined in equation (2.1).[14,16] R_D and R_L represent ratios of the abundances of a pair of fragment ions which have the largest difference between

D- and L- epimers. If the fragmentation abundances are exactly identical for the two peptide epimers, $R_{Chiral} = 1$, indicating no chiral discrimination. If $R_{Chiral} > 1$, a larger number reflects a higher degree of chiral recognition.

$$R_{Chiral} = \frac{R_D}{R_L} \quad R_{Chiral} = R_D/R_L \tag{2.1}$$

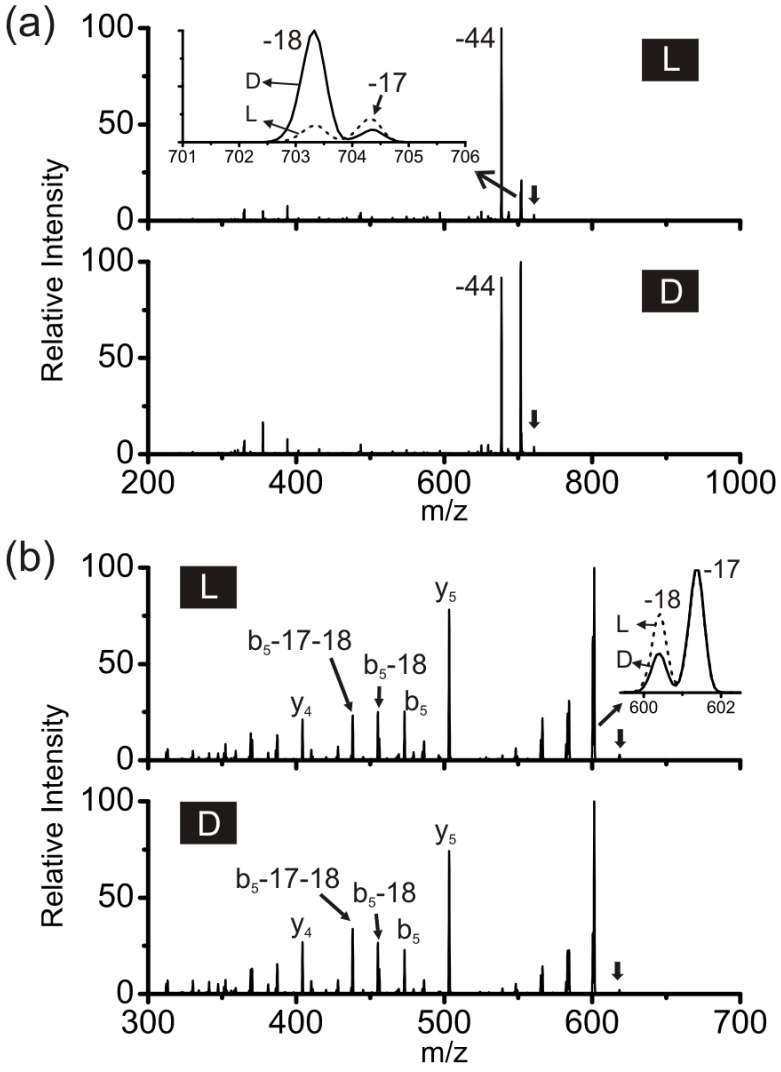


Figure 2.2 (a) CID of re-isolated radical peptide $[(^1$DVGSNK-$NH_2)\bullet + H]^+$ formed in Figure 2.1a. The inset highlights differences in the losses of water and NH_3. (b) Traditional CID of protonated DVGSNK-NH_2.

The R_{Chiral} value for PD based on the side chain losses (-30/-44) is 31±2, and for PD-CID (-18/-17) is 13±4. For comparison, the CID spectra of the two protonated unmodified DVG<u>S</u>NK-NH_2 peptides under the same collision energy are shown in Figure 2.2b. The loss of ammonia and water, and a few backbone fragments are observed for both peptides. Relative intensities of some backbone fragments (y_4^+, b_5^+, y_5^+) and -17/-18Da losses vary slightly between the two spectra. The R_{Chiral} value based on CID (y_4^+/-18) is 2.6±0.3. It is clear that for DVG<u>S</u>NK-NH_2, PD is the best way to differentiate the two serine epimers, and that both PD and PD-CID have significantly better chiral discrimination than CID.

Two other D-ser peptides were analyzed by RDD following modification by **1**, Octopus cardioactive peptide[26] (G<u>S</u>WD, Figure 2.3) and a portion of ω-Aga-IVB/IVC peptide[27] (GL<u>S</u>FA-NH_2, Figure 2.4). Both peptides exhibit a notable difference in the relative intensity of some fragments between the two epimers (supporting information). Photodissociation of these two peptides does not lead to any significant side chain loss or backbone cleavage. This further supports the idea that the unique -30 serine side chain loss is specific for the DVG<u>S</u>NK-NH_2 peptide. The R_{Chiral} value for the PD-CID fragmentation of ^{1}G<u>S</u>WD is 12±1 (peak 294.17/b_3•) and for ^{1}GL<u>S</u>FA-NH_2 is 32±4 (b_4•/c_3). For comparison, the R_{Chiral} values from CID for these two peptides are 6.9±0.2 (b_4-18/-17) and 1.85±0.05 (b_3-18/y_2), respectively. In summary, for all three D-serine peptides studied here, either PD or PD-CID is more effective at discriminating chiral epimers than normal CID.

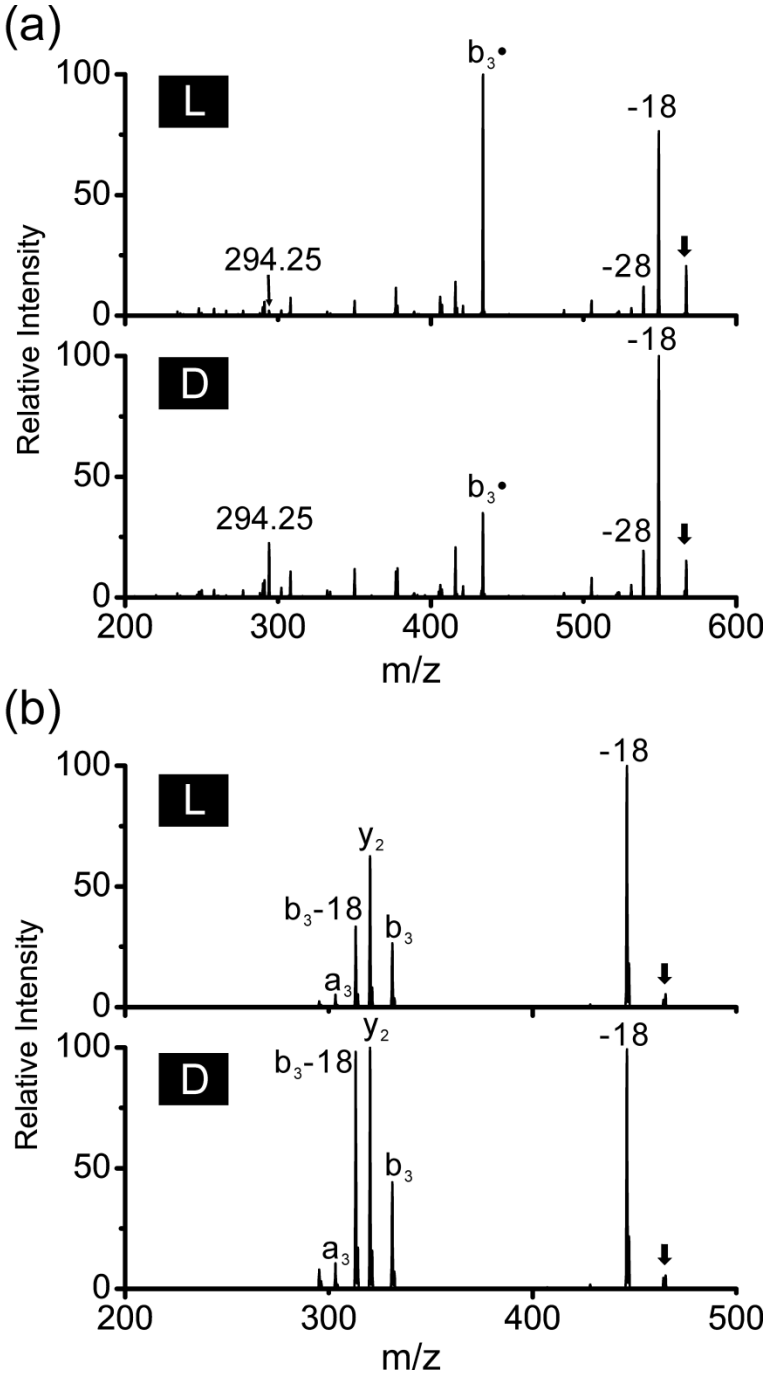


Figure 2.3 (a) PD-CID of $[^{1}G\underline{S}WD + H]^{+}$ and (b) CID of $[G\underline{S}WD + H]^{+}$. RDD shows a better chiral discrimination ability with an R_{Chiral} value of 12±1 (peak 294.17/b_3•) while the number for CID is 1.85±0.05 (b_3-18/y_2).

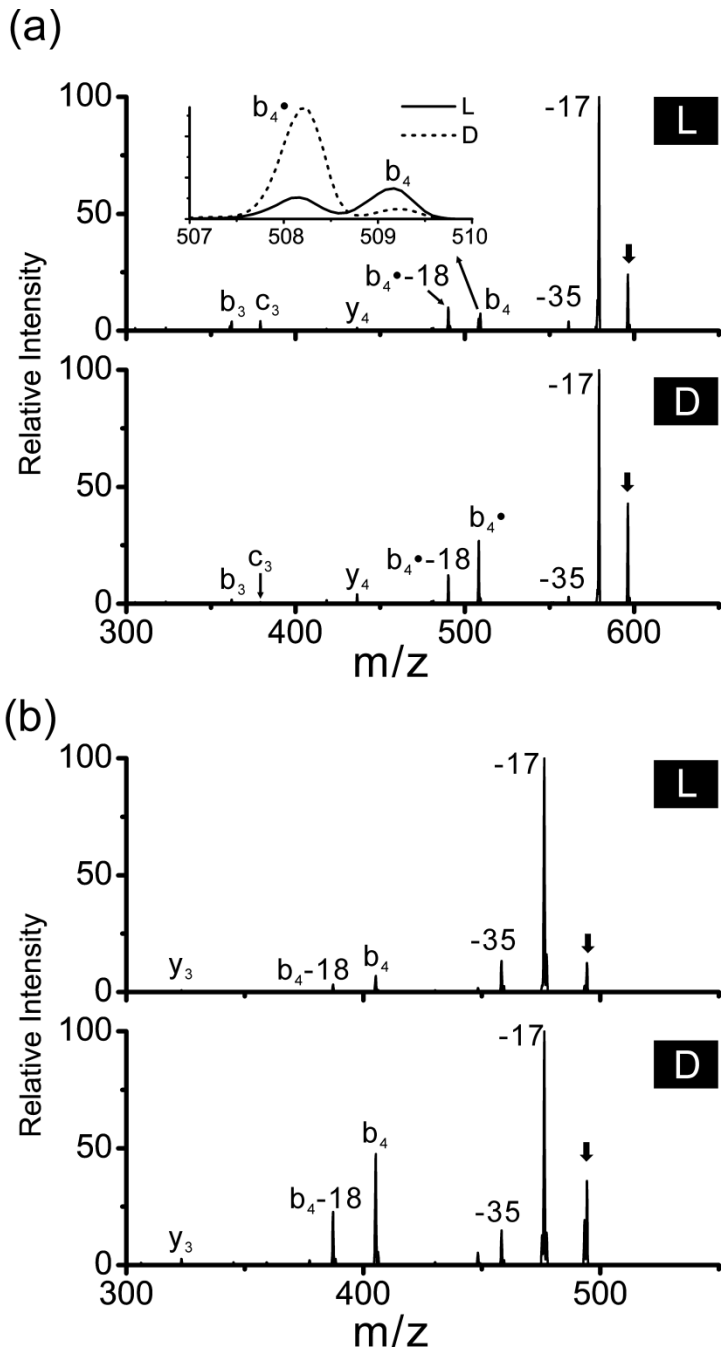


Figure 2.4 (a) PD-CID of $[^{1}GL\underline{S}FA\text{-}NH_2 + H]^+$ and (b) CID of $[GL\underline{S}FA\text{-}NH_2 + H]^+$. Relative intensities of b_4• are significantly different between the two epimers. The R_{Chiral} value is 32±4 (b_4•/c_3) for RDD and 6.9±0.2 (b_4-18/-17) for CID.

2.3.2 D-Asp Peptides.

D-Aspartic acid residues are commonly found in aged animal peptides and proteins extracted from brain, lens, and tooth samples.[28-31] Herein, two tryptic peptides, DAEFR from β-amyloid and IQTGLDATHAER[32] from human lens α-crystallin are examined. Figure 2.5 shows the PD spectra of the singly (a) and doubly (b) charged ^{1}IQTGLDATHAER and its D-Asp epimer. In addition to containing D-Asp, this peptide is also significantly longer than the three serine epimers examined above and can retain multiple charges following ESI. The difference between the spectra of the two epimers is obvious for both charge states, i.e., the −56 Da loss from the Leu side chain is prominent for the D-epimer but minuscule for the L-epimer. The R_{chiral} values are 8.8± 0.8 and 4.5 ± 0.1 for the +1 and +2 charge states, respectively. It is clear that epimers with longer sequences and multiple charges do not impede excellent discrimination by RDD. On the other hand, barely any difference is observed between the CID spectra of the nonradical epimers (Figure 2.6). Among all the three charge states, the best R_{chiral} value obtained by CID is 1.66 ± 0.07. Similar results were obtained by examination of DAEFR as shown in Figure 2.7, where epimerization occurs at the N-terminus. The R_{chiral} value obtained by PD-CID for ^{1}DAEFR is 6.2 ± 0.3, which is again significantly better than the value obtained by CID (2.33 ± 0.07). RDD can efficiently discriminate peptide epimers containing serine or aspartic acid residues; however, both of these amino acids have side chains which are capable of forming strong interactions via hydrogen bonding.

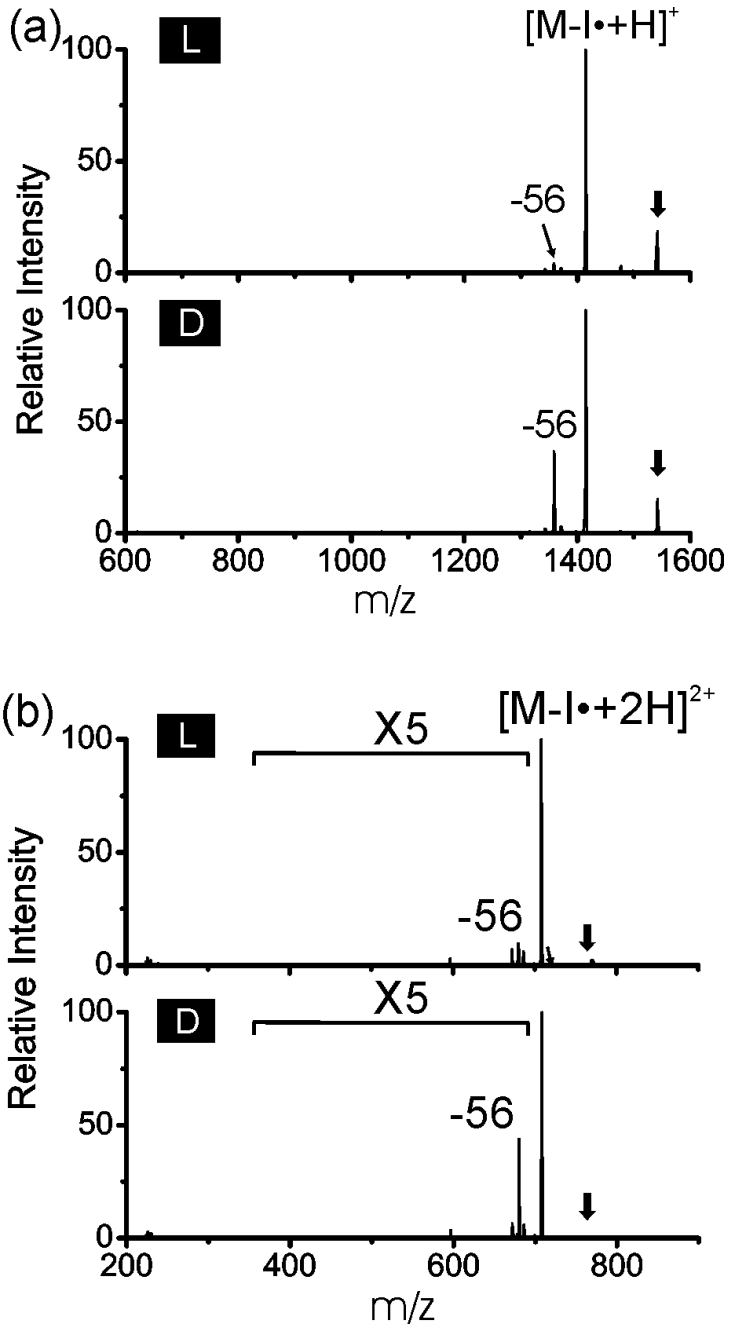


Figure 2.5 (a) PD of [^{1}IQTGLDATHAER+H]$^{+}$ epimers. (b) PD of [^{1}IQTGLDATHAER+2H]$^{2+}$ epimers. The D-form shows abundant −56 Da loss from Leu for both charge states. Bold down arrow indicates precursor ion.

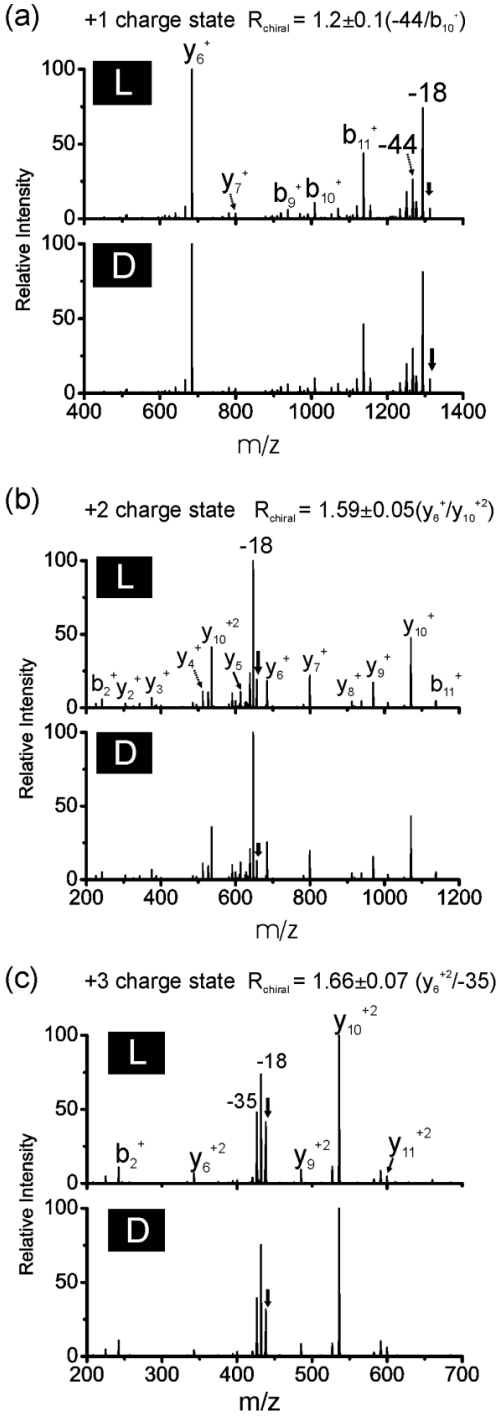


Figure 2.6 (a) CID of $[IQTGL\underline{D}ATHAER+H]^{+}$ epimers. (b) CID of $[IQTGL\underline{D}ATHAER+2H]^{2+}$ epimers. (c) CID of $[IQTGL\underline{D}ATHAER+3H]^{3+}$ epimers. This demonstrated that RDD provides better chiral discrimination than CID in different charge states.

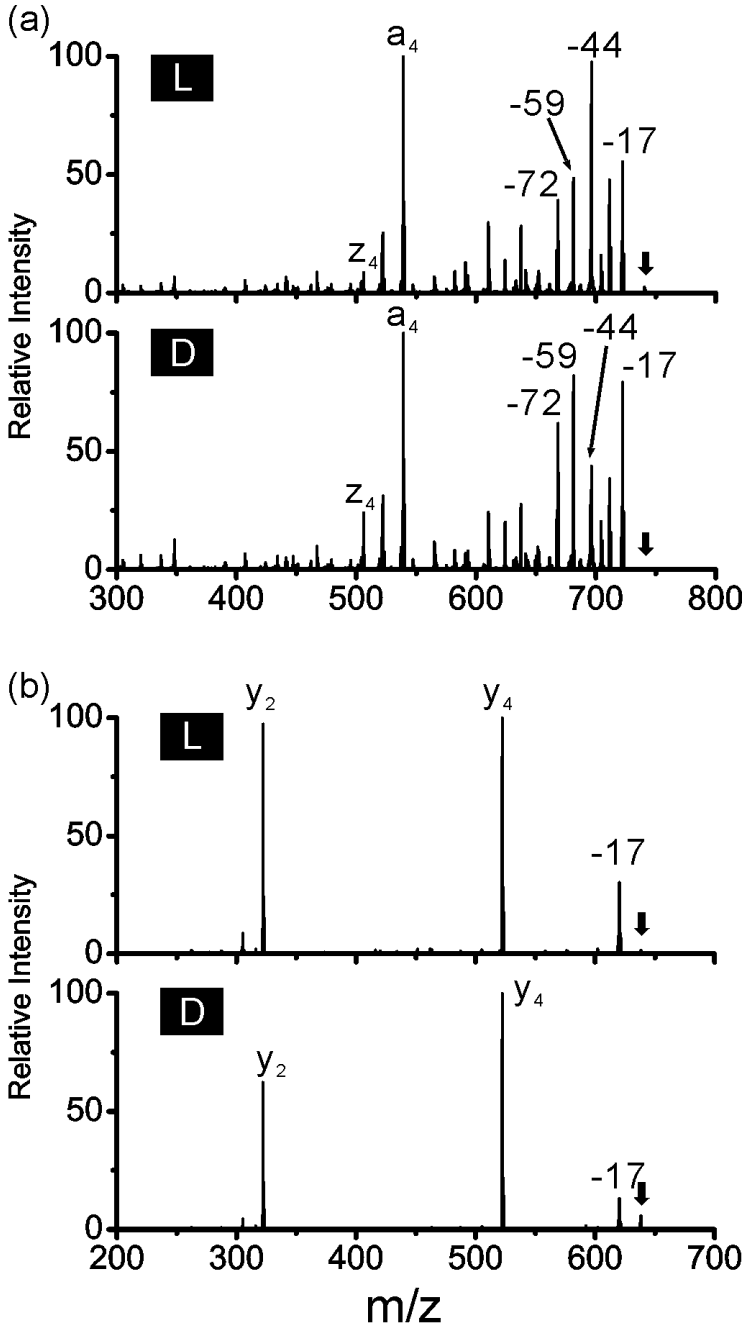


Figure 2.7 (a) PD-CID of [^{1}DAEFR + H]$^{+}$ and (b) CID of [DAEFR + H]$^{+}$ epimers. RDD shows a better chiral discrimination ability with a R_{Chiral} value of 6.2±0.3 (peak z_4^+/-44) while the number for CID is 2.33±0.07 (y_4^+/-17).

2.3.3 D-Ala Peptides.

Alanine is the smallest amino acid that has a chiral carbon. Alternating the chirality of Ala in a peptide should have the least impact on the conformation, and should make discrimination of alanine epimers the most difficult. In this section, some D-Ala peptides will be examined by RDD with different photolabile radical precursors.

Figure 2.8a shows the PD-CID spectra of singly deprotonated dermorphin4 (^{1}YAFGYPS-NH_2) radical and its D-Ala epimer. The fragmentation of the L-epimer is dominated by the loss of 106 Da from tyrosine side chain, while a loss of 31 Da is prevalent for the D-epimer. The R_{Chiral} value (-31/-106) from this experiment is 15±3. This demonstrates that excellent discrimination of peptide epimers can be achieved with RDD even when alanine is the site of chiral inversion. It should be noted that in this case, analysis of the peptide in negative ion mode yielded the best results. Fortunately, previous experiments have demonstrated that for RDD the chemistry which dictates radical migration and fragmentation is largely independent of charge state or charge polarity[33,34] (although differences in structure can still influence the resulting spectra). This ability to interrogate numerous charge states is an important advantage for this peptide, although discrimination was still achieved in positive ion mode (5.1±0.8 for z_6/y_5). CID of the protonated unmodified peptides is shown in Figure 2.8b for comparison. The major difference between the two spectra is the y_5 peak, but with fairly low abundance, and the best R_{Chiral} value is 2.5±0.2 (a_5/y_5).

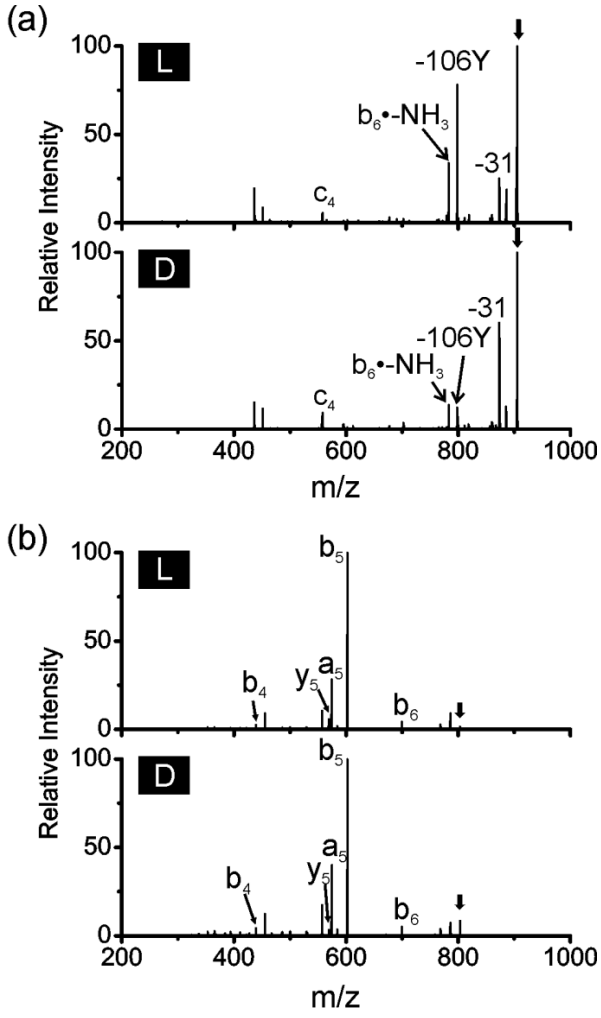


Figure 2.8 (a) PD-CID of $[(^{1}\text{Y}\underline{\text{A}}\text{FGYPS-NH}_2)\bullet - \text{H}]^-$ epimers. (b) CID of $[\text{Y}\underline{\text{A}}\text{FGYPS-NH}_2 + \text{H}]^+$ epimers show similar spectra.

It is also possible to deliver radicals to peptides via noncovalent attachment of a suitable radical precursor such as **3**. In these experiments, **3** is simply added to the peptide solution prior to electrospray. Isolation of the complex $[\text{Y}\underline{\text{A}}\text{FGYPS-NH}_2+\mathbf{3}+\text{H}]^+$ followed by photoactivation at 266nm leads to nearly complete loss of $\text{I}^\bullet$ with significant preservation of the noncovalent interaction, as shown in Figure 2.9a. Some spontaneous dissociation of the noncovalent complex is also observed, releasing the protonated

peptide. However, closer inspection of the peak (see inset in Figure 2.9a) reveals that a distribution of protonated peptide and peptide radical is actually generated. This is not unexpected. As the radical 3$^{\bullet}$ dissociates from the peptide, it may abstract a hydrogen from the peptide or depart as a radical (creating either radical or canonical peptide). Interestingly, the relative abundances of the peptide radicals for the two epimers are different. Figure 2.9b shows the MS^3 CID spectra of the noncovalent radical complexes. The primary product again corresponds to loss of the adduct, although a few peptide fragments are also observed. Once again, the percentage of peptide radical varies between the two epimers as shown in the magnified spectra, indicating that the hydrogen abstraction efficiency is limited in the D-Ala peptide. The R_{Chiral} based on the radical/nonradical pair is 3.7±0.1. Although this number is not significantly higher than the value obtained by CID (2.5±0.2), it is obtained without detailed analysis of a complex fragmentation spectrum. The m/z shifts that occur following loss of **3** are easily predicted, meaning that the radical transfer efficiency could in theory be examined in an automated fashion. Similar results are obtained for other D-Ala containing peptides Deltorphin-I (Y<u>A</u>FDVVG-NH_2 Figure 2.10), and <u>A</u>STTTNYT-NH_2 (Figure 2.11).[35] In both cases, the D-epimer exhibits less radical yield. For the three alanine containing peptide epimers studied herein, chiral inversion can be detected by intermolecular radical transfer. Unfortunately, none of the previously examined serine peptides and only one of the aspartic acid peptides displayed any differences in radical transfer efficiency. There is a likely explanation for these results. In all three alanine peptides, the chiral inversion occurs near the N-terminus, which also happens to be the most likely 18C6 binding site

for all three peptides (due to lack of arginine or lysine residues). Although one of the serine peptides also fits these criteria, alanine is an unlikely site for radical abstraction, while serine is not. In any case, the potential utility of using radical transfer efficiency to detect chiral inversion will be limited to some subset of all epimers.

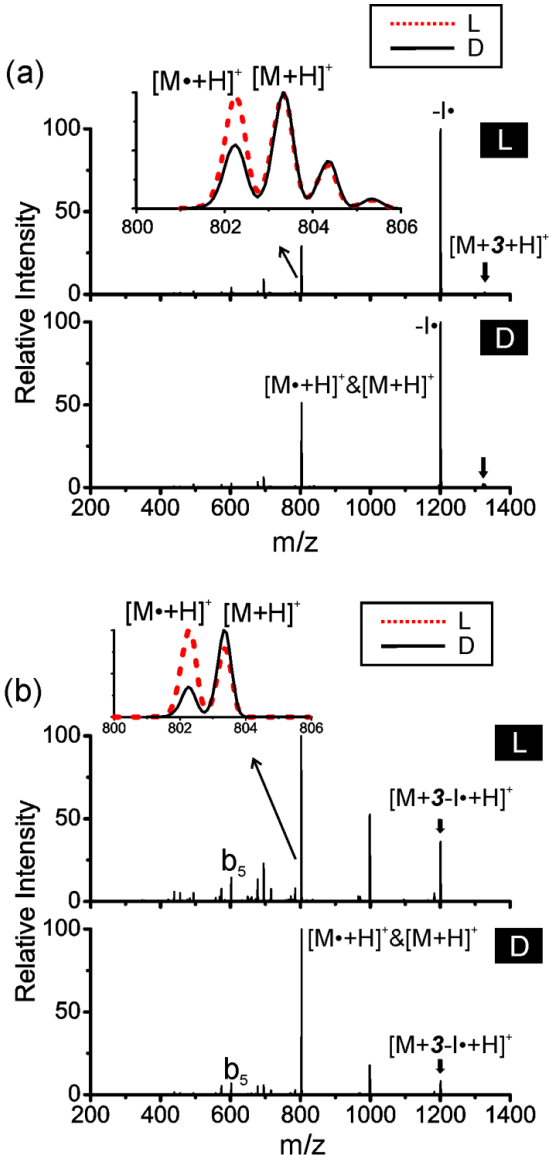


Figure 2.9 (a) PD of [Y<u>A</u>FGYPS-NH_2 + ***3*** + H]$^+$ epimers. (b) CID of the radical complex generated by loss of iodine in panel a. The insets highlight differences in the ratio of radical/canonical peptide generated by loss of the crown adduct. Bold down arrow indicates precursor ion.

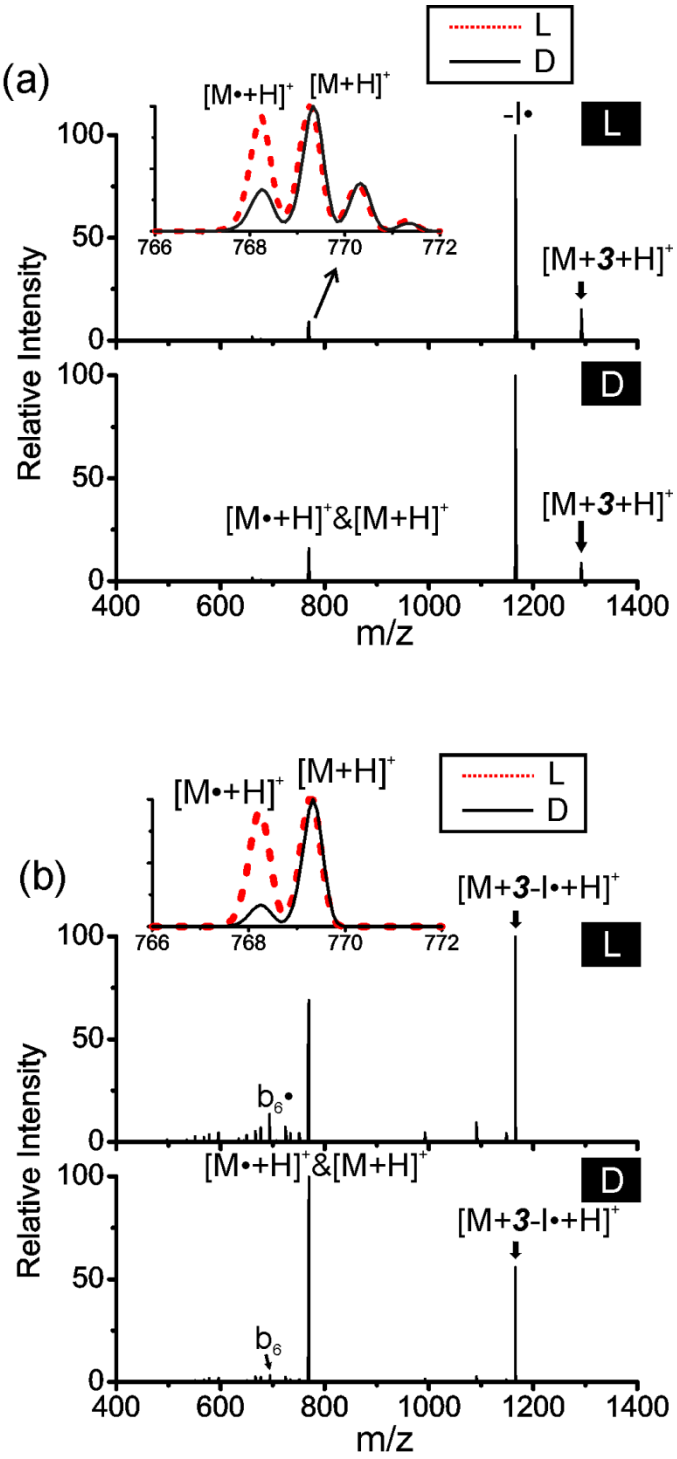


Figure 2.10 (a) PD of [Y<u>A</u>FDVVG-NH_2 + ***3*** + H]$^+$ and (b) PD-CID of [Y<u>A</u>FDVVG-NH_2+ ***3*** +H]$^+$ epimers. The hydrogen abstraction efficiency is limited in the D-Ala peptide.

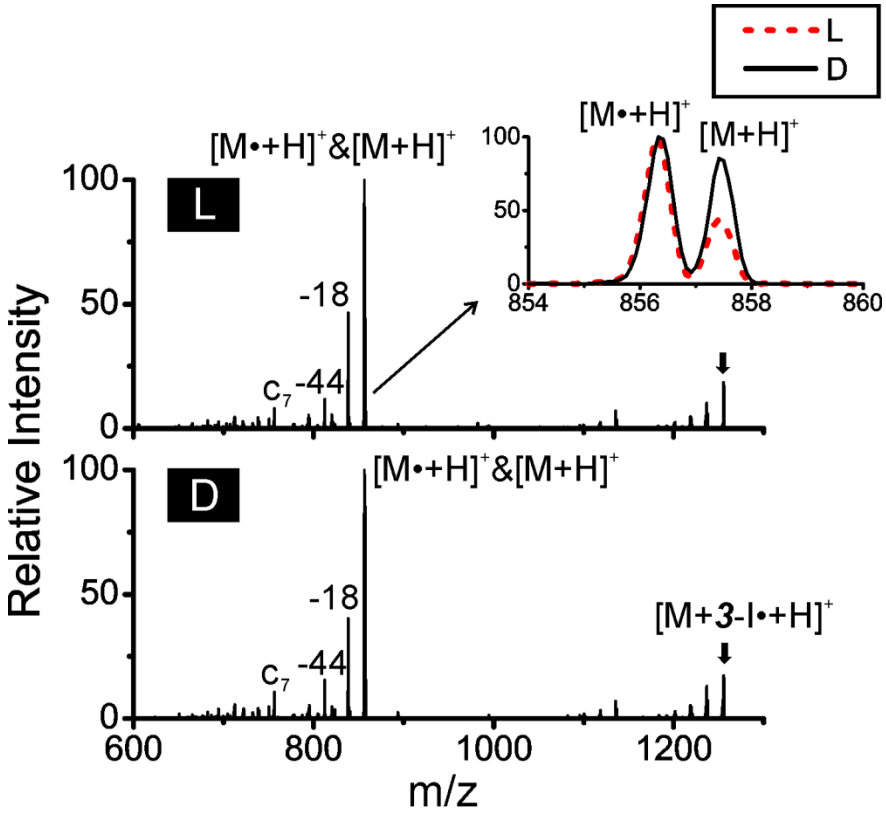


Figure 2.11 PD-CID of [ASTTTNYT-NH_2+3+H]$^+$ epimers. The radical transfer efficiency difference is very easy to identify.

2.3.4 Discrimination Abilities

A summary of R_{Chiral} values from RDD and CID for eight D-amino acid containing peptides is shown in Table 2.1. The RDD experiments include both covalent and noncovalent radical delivery methods. It is clear that RDD provides better chiral discrimination than CID for most of the peptides, except Deltorphin C (YAFDVVG-NH_2). In order to achieve chiral discrimination by comparative dissociation, two conflicting goals must be achieved simultaneously. One, the entire peptide structure must be subject to examination (otherwise, the chiral center in question may not be probed). Two, the pathway that leads to any given dissociation should ideally be quite specific. In

RDD, a radical is generated at one specific site, which reduces the number of redundant pathways that could lead to a particular dissociation. This is a significant advantage over proton initiated dissociation, where multiple protons or equivalent protonation sites frequently exist. Following radical initiation in RDD, the number of sites where radical migration can occur exceeds the number of sites where fragmentation will be favored, which increases the probability that the entire peptide structure will be probed prior to fragmentation. Importantly, radical migration is known to be kinetically constrained by structure, meaning that redundant pathways leading to the same fragmentation are disfavored. Thus, RDD appears to naturally have the right mixture of specificity and inclusivity to enable identification of single amino acid chiral inversions.

In addition to being useful for identification, the abundances of the fragment ions in RDD are also reproducible and quantifiable. Figure 2.12 shows the quantitative analysis of mixtures of peptides epimers using both covalent and noncovalent methods. Figure 2.12a shows the quantitative dependence of the relative abundance of -30 Ser side chain loss from PD of $^{\mathbf{1}}$DVG<u>S</u>NK-NH_2 as a function of D-serine epimer percentage. In Figure 2.12b, the abundance ratio of radical and nonradical peptide after PD of [Y<u>A</u>FDVVG-NH_2+**3**+H]$^+$ is plotted verses the percentage of L-Ala epimer. Both plots provide linear calibration responses with $r > 0.99$. Based on the error bars and slope, it is estimated that the detection limit of the measurement is around 1-2%. Furthermore, larger error bars are present with decreasing percentage of the L-form DVG<u>S</u>NK-NH_2 peptide; but in practice, samples containing smaller amounts of the D-epimer would likely be examined, where the calibration curve has relatively small error bars.

Table 2.1 Comparison of chiral recognition factors among different dissociation methods. [a,b,c]

Dissociation Methods / Peptides	RDD			CID
	Covalently Modification		IBA-18C6 PD-CID	
	PD	PD-CID		
DVG<u>S</u>NK-NH_2	31±2	13±4	2.5±0.2[g]	2.6±0.3
GL<u>S</u>FA-NH_2	--[d]	32±4	9±1[g]	6.9±0.2
G<u>S</u>WD	--[d]	12±1	1.6±0.2[g]	1.85±0.05
Y<u>A</u>FDVVG-NH_2	--[d]	10±2[e]	5.3±0.1[h]	18±1
Y<u>A</u>FGYPS-NH_2	5.7±0.3	15±3[f]	5.1±0.3[g]	2.5±0.2
<u>A</u>STTTNYT-NH_2	5.7±0.2	7±2	1.9±0.1[h]	3.2±0.5
<u>D</u>AEFR	--[d]	6.2±0.3	1.62±0.04[h]	2.33±0.07
IQTGL<u>D</u>ATHAER[i]	8.8±0.8	5.0±0.4[i]	8±1[g,j]	1.66±0.07[j]

[a]Single letter codes are used for the amino acids. The underlined residue corresponds to the site of epimerization. [b]Errors represent the standard deviation of the mean. [c]The relative intensity of the chirality determining ions have to be higher than 3% of the base peak in at least one of the spectra, either D or L. All the values are calculated from the +1 charge state unless otherwise noted. [d]Only loss of iodine was observed. [e]Radical introduced by iodo-tyrosine. [f]The value is obtained under negative mode. [g]Calculated by the intensity variations of backbone fragmentations. [h]Calculated by radical transfer efficiency difference. [i]Calculated from the +2 charge state. [j]Calculated from the +3 charge state.

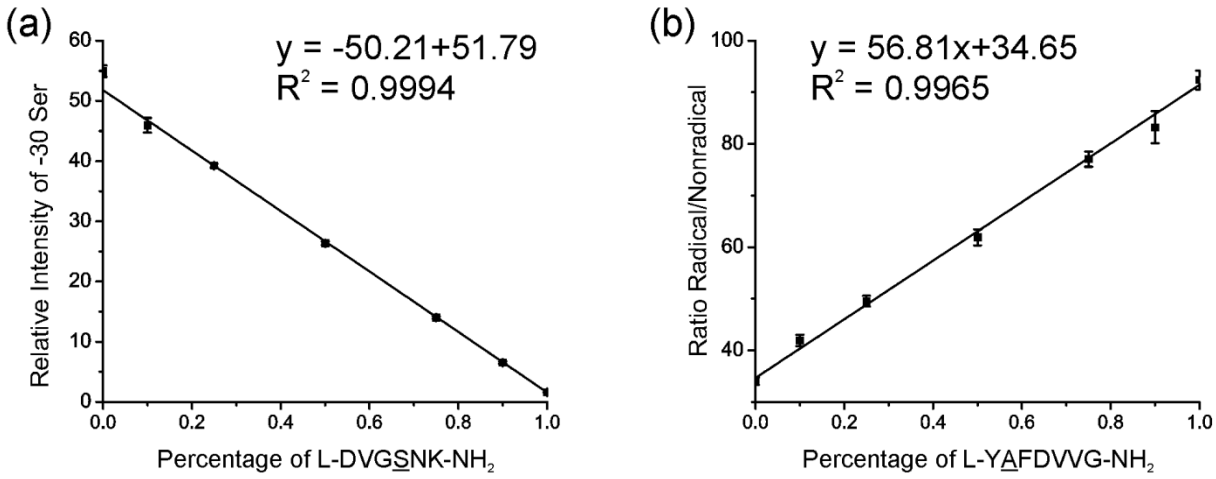


Figure 2.12 Calibration curve for PD of ^{1}DVG<u>S</u>NK-NH_2 (a) and PD of Y<u>A</u>FDVVG-NH_2 + ***3*** complex (b).

2.4 Conclusions

Radical directed dissociation (RDD) is demonstrated to successfully discriminate D-amino acid containing peptide epimers and quantitatively measure the relative content of one epimer in a diastereomeric mixture. Different methods including noncovalent and covalent labeling were applied for introducing photolabile precursors onto the peptides. For most of the peptides that were examined, RDD provides a significantly better chiral recognition than normal CID. The diversity of the radical delivery methods and the flexibility of examining peptides in various different charge states is a significant advantage. These promising results suggest that RDD might be a viable method for examining peptide epimers from complex samples via coupling with liquid chromatography.

References

[1]Iida, T.; Santa, T.; Toriba, A.; Imai, K., Amino acid sequence and D/L-configuration determination methods for D-amino acid-containing peptides in living organisms. *Biomed. Chromatogr*. **2001,** *15*, 319-327.

[2] Kreil, G., D-amino acids in animal peptides. *Annu. Rev. Biochem*. **1997,** *66*, 337-345.

[3]Erspamer, V.; Melchiorri, P.; Falconierierspamer, G.; Negri, L.; Corsi, R.; Severini, C.; Barra, D.; Simmaco, M.; Kreil, G. Deltorphins – Deltorphins: A family of naturally occurring peptides with high affinity and selectivity for δ opioid binding sites. *Proc. Natl. Acad. Sci.* **1989,** *86*, 5188-5192.

[4] Broccardo, M.; Erspamer, V.; Falconierierspamer, G.; Improta, G.; Linari, G.; Melchiorri, P.; Montecucchi, P. C. Pharmacological data on dermorphins, a new class of potent opioid-peptides from amphibian skin. *Br. J. Pharmacol.* **1981,** *73*, 625-631.

[5] Shapira, R.; Chou, C. H. J. Differential racemization of aspartate and serine in human myelin basic protein. *Biochem. Biophys. Res. Commun.* **1987,** *146*, 1342-1349.

[6]Kaneko, I.; Yamada, N.; Sakuraba, Y.; Kamenosono, M.; Tutumi, S. Suppression of mitochondrial succinate dehydrogenase, a primary target of β-Amyloid, and its derivative racemized at Ser residue. *J. Neurochem.***1995**, *65*, 2585-2593.

[7] Ilisz, I.; Berkecz, R.; Peter, A. Application of chiral derivatizing agents in the high-performance liquid chromatographic separation of amino acid enantiomers: A review. *J. Pharm. Biomed. Anal.* **2008**, *47*, 1-15.

[8] Takeuchi, Y.; Marchand, A. P. Applications of NMR spectroscopy to problems in stereochemistry and conformational analysis. VCH Publishers: Deerfield Beach, FL, 1986.

[9] Ewing, M. A.; Wang, J.; Sheeley, S. A.; Sweedler, J. V. Detecting D-Amino acid-containing neuropeptides using selective enzymatic digestion. *Anal. Chem.* **2008,** *80*, 2874-2880.

[10] Tao, W. A.; Zhang, D. X.; Nikolaev, E. N.; Cooks, R. G. Copper(II)-assisted enantiomeric analysis of D,L-amino acids using the kinetic method: Chiral recognition and quantification in the gas phase. *J. Am. Chem. Soc.* **2000,** *122*, 10598-10609.

[11] Huang, L.; Lu, X.; Gough, P. C.; De Felippis, M. R. Identification of Racemization Sites Using Deuterium Labeling and Tandem Mass Spectrometry. *Anal. Chem.* **2010,** *82*, 6363-6369.

[12] Serafin, S. V.; Maranan, R.; Zhang, K. L.; Morton, T. H. Mass spectrometric differentiation of linear peptides composed of L-amino acids from isomers containing one D-amino acid residue. *Anal. Chem.* **2005,** *77*, 5480-5487.

[13] Sachon, E.; Clodic, G.; Galanth, C.; Amiche, M.; Ollivaux, C.; Soyez, D.; Bolbach, G. ., D-Amino Acid Detection in Peptides by MALDI-TOF-TOF. *Anal. Chem.* **2009**, *81*, 4389-4396.

[14]Bai, L.; Romanova, E. V.; Sweedler, J. V. Distinguishing Endogenous D-Amino Acid-Containing Neuropeptides in Individual Neurons Using Tandem Mass Spectrometry. *Anal. Chem.* **2011**, *83*, 2794-2800.

[15] Adams, C. M.; Kjeldsen, F.; Zubarev, R. A.; Budnik, B. A.; Haselmann, K. F. Electron capture dissociation distinguishes a single D-amino acid in a protein and probes the tertiary structure. *J. Am. Soc. Mass. Spectrom.* **2004,** *15*, 1087-1098.

[16] Adams, C. M.; Zubarev, R. A. Distinguishing and quantifying peptides and proteins containing D-amino acids by tandem mass spectrometry. *Anal. Chem.* **2005,** *77*, 4571-4580.

[17] Ly, T.; Julian, R. R. Elucidating the Tertiary Structure of Protein Ions in Vacuo with Site Specific Photoinitiated Radical Reactions. *J. Am. Chem. Soc.* **2010,** *132,* 8602-8609.

[18] Diedrich, J. K.; Julian, R. R. Site-specific radical directed dissociation of peptides at phosphorylated residues. *J. Am. Chem. Soc.* **2008,** *130*, 12212-12213.

[19] Diedrich, J. K.; Julian, R. R. Facile Identification of Phosphorylation Sites in Peptides by Radical Directed Dissociation. *Anal. Chem.* **2011,** *83*, 6818-6826.

[20] Moore, B. N.; Blanksby, S. J.; Julian, R. R. Ion-molecule reactions reveal facile radical migration in peptides. *Chem. Commun.* **2009,** 5015-5017.

[21]Ly, T.; Zhang, X.; Sun, Q. Y.; Moore, B.; Tao, Y. Q.; Julian, R. R. Rapid, quantitative, and site specific synthesis of biomolecular radicals from a simple photocaged precursor. *Chem. Commun.* **2011,** *47*, 2835-2837.

[22]Julian, R. R.; Beauchamp, J. L. Site specific sequestering and stabilization of charge in peptides by supramolecular adduct formation with 18-crown-6 ether by way of electrospray ionization. *Int. J. Mass spectrom.* **2001,** *210*, 613-623.

[23] Julian, R. R.; May, J. A.; Stoltz, B. M.; Beauchamp, J. L. Biomimetic approaches to gas phase peptide chemistry: combining selective binding motifs with reactive carbene precursors to form molecular mousetraps. *Int. J. Mass spectrom.* **2003,** *228*, 851-864.

[24] Chan, W. C.; White, P. D. *Fmoc Solid Phase Peptide Synthesis,* 1st ed.; Oxford University Press: New York, 2000.

[25] Ly, T.; Julian, R. R. Residue-specific radical-directed dissociation of whole proteins in the gas phase. *J. Am. Chem. Soc.* **2008,** *130*, 351-358.

[26] Iwakoshi, E.; Hisada, M.; Minakata, H. Cardioactive peptides isolated from the brain of a Japanese octopus, Octopus minor. *Peptides,* **2000**, *21*, 623-630.

[27] Heck, S. D.; Siok, C. J.; Krapcho, K. J.; Kelbaugh, P. R.; Thadeio, P. F.; Welch, M. J.; Williams, R. D.; Ganong, A. H.; Kelly, M. E.; Lanzetti, A. J.; Gray, W. R.; Phillips, D.; Parks, T. N.; Jackson, H.; Ahlijanian, M. K.; Saccomano, N. A.; Volkmann, R. A. Functional consequences of posttranslational isomerization of Ser(46) in a calcium channel toxin. *Science*, **1994,** *266*, 1065-1068.

[28] Masters, P. M.; Bada, J. L.; Zigler, J. S. Aspartic-acid racemization in human lens during aging and in cataract formation. *Nature,* **1977,** *268,* 71-73.

[29] D'Aniello, A. D-Aspartic acid: An endogenous amino acid with an important neuroendocrine role. *Brain Res. Rev*, **2007,** *53*, 215-234.

[30] Roher, A. E.; Lowenson, J. D.; Clarke, S.; Wolkow, C.; Wang, R.; Cotter, R. J.; Reardon, I. M.; Zurcherneely, H. A.; Heinrikson, R. L.; Ball, M. J.; Greenberg, B. D. Structural alterations in the peptide backbone of beta-amyloid core protein may account for its deposition and stability in Alzheimer's disease. *J. Biol. Chem.* **1993,** *268*, 3072-3083.

[31] Fujii, N. D-amino acid in elderly tissues. *Biol. Pharm. Bull.* **2005,** *28*, 1585-1589.

[32] Fujii, N.; Satoh, K.; Harada, K.; Ishibashi, Y. Simultaneous stereoinversion and isomerization at specific aspartic acid residues in αA-crystallin from human lens. *J. Biochem.* **1994**, *116*, 663−669.

[33] Sun, Q. Y.; Nelson, H.; Ly, T.; Stoltz, B. M.; Julian, R. R. Side Chain Chemistry Mediates Backbone Fragmentation in Hydrogen Deficient Peptide Radicals. *J. Proteome Res.* **2009,** *8*, 958-966.

[34] Moore, B.; Sun, Q.; Hsu, J. C.; Lee, A. H.; Yoo, G. C.; Ly, T. ., Dissociation chemistry of hydrogen-deficient radical Peptide anions.*J. Am. Soc. Mass. Spectrom.* **2011,** *23,* 460-8.

[35] Raychaudhuri, S. K.; Raychaudhuri, S. P.; Farber, E. M. Anti-chemotactic activities of peptide-T: a possible mechanism of actions for its therapeutic effects on psoriasis. *Int. J. Immunopharmacol.* **1998,** *20,* 661-667.

Chapter 3

ISOMER PROTEOMICS: A CASE STUDY OF EPIMERIZATION AND ISOMERIZATION IN CRYSTALLINS

3.1 Introduction

The eye lens is a very peculiar and interesting construct.[1,2] It is composed of 90% crystallin proteins.[3] Among the crystallins, Alpha crystallins are most abundant and serve as structural elements and as chaperones.[4,5] In order to achieve the desired index of refraction required for an optic, the concentration of proteins within the lens must be high.[6] At the same time, aggregation of proteins into particles capable of scattering light must also be avoided. The alpha crystallins assemble into large oligomeric species of between 20-40 monomers which are highly dynamic in nature. Numerous studies have indicated a preference for even numbered oligomers, suggesting that the assemblies are comprised of dimer building blocks.[7] Additionally, within each monomer several structural regions have been identified that each serve distinct roles. The central alpha crystallin domain folds into a well-defined structure that is conserved across many of the small heat shock proteins. In contrast, the N-terminal region and the C-terminal extension do not form well-defined structures at the monomer level; however, these regions help regulate higher order oligomer assembly and are necessary for chaperone activity.

As fiber cells in the eye lens mature, all organelles are ejected. The mature fiber cells perform very few metabolic functions, have low oxygen and energy demands, and are

avascular. Perhaps most interestingly, the crystallin proteins that were present in the lens of a person when born are still present when they die. In other words, there is no turnover of proteins within lens fiber cells. Given these properties, it is not surprising that many studies have focused on examination of the changes that occur to lens proteins upon aging. It has been established that numerous post-translational modifications occur to crystallin proteins as a function of aging. Deamidation, truncation, glycation, phosphorylation, disulfide bond formation, oxidation, acetylation, and methylation are among the most commonly studied modifications.[8-12] However, there is also evidence that epimerization, which occurs when a single amino acid acid undergoes stereoinversion, is also an important modification that occurs as a function of aging.[13-15] Epimerization leads to no change in mass and it is significantly more difficult to detect than the other post-translational modifications (PTMs) listed above, which may explain why epimerization has received significantly less attention.

Aspartic acid is a special amino acid in the context of PTMs that are not accompanied by changes in mass. Backbone attack of the aspartic acid side chain yields a succinimide ring that can lead to the formation of several isomeric forms. Simple ring opening will yield a mixture of aspartic acid and isoaspartic acid, where the side chain has essentially inserted into the peptide backbone.[16] In addition, the chiral alpha hydrogen atom in the succinimide ring can undergo stereoinversion, which can lead to the formation of D-aspartic acid and D-isoaspartic acid. Since these two pathways can occur in conjunction, the end result is that aspartic acid is frequently converted into four isomeric states, none of which are distinguished by a shift in mass. Both isomerization and epimerization

significantly perturb the local structure of the molecule at the affected residue. For example, it has been demonstrated that substitution of D-residues can significantly reduce the propensity of peptides to adopt alpha helical structures.[17] Similar structural changes may also explain why elevated epimerization is also associated unfavorably with numerous diseases. For example, racemization of serine residues in β-Amyloid increases the rate of aggregation and accelerates degeneration of neuronal cells, which may be connected to the cause of Alzheimer's disease.[18] Epimerized residues are also more frequently detected in crystallin proteins from cataract sufferers than in age controlled healthy individuals.[19,20] It is clear from these initial findings that characterization of isomeric PTMs, though difficult, is warranted.

The majority of work in this area involving mass spectrometry has focused on examination of isolated molecules. For example, it has been demonstrated that differences in MS^2 spectra can be used to distinguish aspartic acid from isoaspartic acid.[21] Similarly, differences in MS^2 spectra can be used to distinguish peptide epimers.[22-26] Quantitative analysis is typically carried out by calculation of an R value[27] that corresponds to the degree of difference between the two fragments that change the most in the MS^2 spectra obtained from each isomer (additional details about R values are provided below). R values of 1 correspond to no difference between the spectra, and larger values reflect bigger differences between the spectra being compared. Importantly, this method requires that both the all L peptide and the epimer with a single D residue be independently evaluated. Furthermore, not all fragmentation methods are equivalent in epimer disambiguation. For example, although collision induced dissociation (CID) can

yield acceptable results, it generally offers less structural sensitivity than electron or radical based dissociation methods. Recent work has demonstrated that radical directed dissociation (RDD) yields the highest R values for epimer detection and has the advantage of the greatest flexibility in terms of charge state selection.[28] For identification of isoaspartic acid, electron capture dissociation (ECD) is advantageous because it yields a characteristic fragment that can facilitate identification.[29,30]

Implementation of mass spectrometry in conjunction with liquid chromatography for the analysis of more complicated isomer containing samples presents additional opportunities and challenges. Fortunately, analysis of biological samples within the context of aging simplifies the experiment in one important way− the original isomers will always be present. All processes by which spontaneous epimerization/isomerization occurs are incomplete, therefore some of the original peptide or protein will always remain. Given this information, the challenge can be broken down into two components: separation and characterization. Separation is typically carried out with standard liquid chromatography (LC), which is capable of baseline separating many isomers using typical C18 columns (i.e. columns packed with chiral media are not required).[31] Therefore, in terms of the goals and approaches for separations, traditional proteomics and isomer proteomics are very similar. However, for characterization the experimental protocol for examining epimers or other isobaric isomers is fundamentally different in at least one important way. In a typical proteomics experiment, the goal is usually to characterize each unique m/z or peptide once and only once. On the other hand, if the experiment is to identify post-translational isomerization, then it is not only desirable, but

also required that the same m/z or peptide be examined multiple times. Multiple spectra must be acquired for comparison to confirm epimerization and for calculation of relevant R values. The need to acquire multiple spectra for the same m/z values must be then balanced against the traditional goal of also simultaneously examining as many unique peptides as possible. These requirements place limits on the complexity of samples that can be evaluated in an online fashion for isomer proteomics.

In the present work, I characterized the isomeric PTMs in sheep crystallins extracted from the eye lens. The results from both RDD and CID on LC separated peptides were combined to improve isomer identification. A short exclusion time and a target peptide mass list were used to ensure that each peptide was examined multiple times to allow for comparison of the relevant tandem mass spectra. Three crystallin proteins (αA-, αB-, and βB3-crystallin) were identified from the ovine database with excellent sequence coverage. Several additional β- and γ- crystallins were also identified from the bovine database (the ovine database is incomplete) due to high sequence homology. The results illustrate that the greatest degree of isomerization and epimerization occurs in the disordered N-terminal and C-terminal regions of αA- and αB-crystallin, which are abundant and important proteins in the lens that function as chaperones and also serve as structural elements.

3.2 Experimental Methods

3.2.1 Materials

Organic solvents and reagents were purchased from Sigma-Aldrich (St. Louis, MO) or Acros Organics (Geel, Belgium) and used without further purification. Water was

purified to 18.2 MΩ by a Millipore 147 (Billerica, MA) Direct-Q system. Amino acids and resin were purchased from Ana Spec (Fremont, CA). Trypsin was purchased from Sigma-Aldrich (St. Louis, MO).

3.2.2 Peptide and Radical Precursor Synthesis

All synthetic peptides were synthesized manually using standard fmoc procedures with Rink Amide Resin or Wang Resin for the solid support. [32] N-hydroxysuccinimide (NHS) activated iodo-benzoyl esters were synthesized by a previously reported procedure.[33]

3.2.3 Protein Extraction and Digestion

Sheep eyes were obtained from discarded tissue Corona Cattle Inc. (Corona, CA). The lenses were separated, washed with distilled water, and then homogenized in 50mM Tris/HCl, pH 7.8 buffer. The homogenate was centrifuged at 10,000 rpm for 20min at 4°C. The supernatant was purified by dialysis against water and lyophilized. The lyophilized protein was dissolved in 50mM NH_4HCO_3 buffer, pH 7.8, and the disulfide bonds were reduced in 100mM DTT at 95°C for 5min. After returning to room temperature, 100mM iodoacetamide solution was added and the mixture was incubated in the dark for 20min. Then proteins were digested with trypsin overnight at 37°C, with the protein-enzyme ratio at 50:1. For the iodo-benzoic modification, the digestion mixture was first purified with a peptide trap (Michrom Bioresource Inc.). Approximately 5nmol protein digestion mixture, 15μL of 15mM NHS ester dioxane solution, and 5μL borate buffer (pH 8.6) were combined and incubated for 1h at 40°C. Important: note that dimethyl sulfoxide (DMSO) should not be used for this step because it can easily cause

aspartic acid isomerization. The modification side products at arginine and tyrosine side chain were removed by incubating the reaction mixture in 1M hydroxylamine (pH 8.5 adjusted by NaOH) for 4h. The exact same procedure was performed on control peptides. Since the deamidation and racemization of asparaginyl and aspartyl residues are non-enzymatic spontaneous reactions that can occur under physiological conditions,[34] control experiments with a synthetic peptide (TVLDSGISEVR) were performed to ensure that the sample preparation procedure will not induce any isomerization. After reduction by DTT in 95°C and incubation with trypsin overnight, no isomerization was detected by LC-MS (data not shown).

3.2.4 Mass Spectrometry and Radical Directed Dissociation

Solutions were analyzed by an LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) with a standard electrospray ionization (ESI) source. The back plate of the mass spectrometer was modified with a quartz window to transmit fourth harmonic (266 nm) laser pulses from a Nd:YAG laser (Continuum, Santa Clara,CA). Photodissociation of the labeled peptide homolytically cleaves the C-I bond in the chromophore and produces a radical peptide. Further MS^3 experiments were performed by re-isolation and collision-induced dissociation (CID) of the radical species.

3.2.5 LC-MS Data Acquisition and Analyisis

An Agilent 1100 series HPLC system (Agilent, Santa Clara, CA) with a BetaBasic-18 column (150×2.1 mm, particle size 5μm) was coupled to the LTQ mass spectrometer. Peptides were separated using 0.1% formic acid in water (mobile phase A) and 0.1%

formic acid in acetonitrile (mobile phase B) with a flow rate of 0.2mL/min. The digestion mixture was separated by the following gradient: 5%B to 20%B over 25min, 20%B to 30%B over the next 35min, 30% to 50%B over the next 15min, 50%B to 95%B over the final 10min. The MS instrument was operated in the data-dependent mode. In a CID-only LC-MS run, the first scan event is full MS from m/z 300 to 2000 Da, followed by ultrazoom (scan event 2) and CID-MS^2 (scan event 3). In a RDD LC-MS run, the laser pulses were triggered during the MS^2 (scan event 3) and the CID was performed as a pseudo-MS^3 step (scan event 4).[35] Since the photodissociation of 4-iodo-benzoic labeled peptide will always produce the radical peptide as the major product, the precursor ion of CID in the MS^3 step is the radical species rather than the original peptide. The exclusion time was 60 sec for the identification of peptides and 16 sec for the isomer discrimination.

MS data were acquired with Xcalibur software (Thermo Fisher Scientific). The raw files were converted to mgf files by MM File Conversion. The mgf files were searched with X! Tandem (version 2013.02.01.1) against the *ovis aries* database UniProt 2014 06, 26,849 entries). The cleavage sites were set as lysine and arginine (semi cleavage was turned on), allowing up to two missed cleavages and one point mutation. Carbamidomethylation (+57.02 Da at Cys) was set as fixed modification, and N-acetylation was considered a variable modification. For the modified digestion mixture, the 4-iodobenzoic acid modification (+230.01 Da) was considered a variable modification at either the N-terminus or lysine side chain. The parent monoisotopic mass error was set to ±1 Da and the fragment monoisotopic mass error was set to ±0.4 Da. The minimum parent ion mass was set to 400 Da. The criteria used for accepting peptides

identification is e < 0.005 for peptides. The false discovery rate is 1%, calculated by searching the data against the reversed database. Given that the content of the eye lens is ~90% crystallin proteins, the data was also searched against a smaller database that contains primarily crystallin proteins. A few additional peptides were identified in this fashion and their identities were confirmed by manual assignment of the MS/MS data.

3.2.6 Calculation of R Values

To quantify the isomer discrimination sensitivity, an R value originally reported by Tao et al.[27] for chiral selectivity is used. In the present work, the R is calculated by equation 1. R_A and R_B represent ratios of the relative intensities of a pair of fragment ions which varies the most between two isomers. Therefore, $R_{isomer} = 1$ indicates that the two tandem MS spectra are exactly identical and no isomerization occurs. If $R_{isomer} > 1$, a larger number reflects a higher probability of isomerization. The statistical significance of the results is addressed in the Results and Discussion. In addition, we use an S value calculated in the same fashion to provide a quantitative measure of the similarity of experimental spectra to those obtained from synthetic standards. In the case of the S value, the number should be smaller than the threshold. Although the same formula is used in both cases, since the value should be higher in one case and lower in the other, we have given them different designations to avoid confusion.

$$S_{CID/RDD} = R_{isomer} = R_A/R_B \qquad (3.1)$$

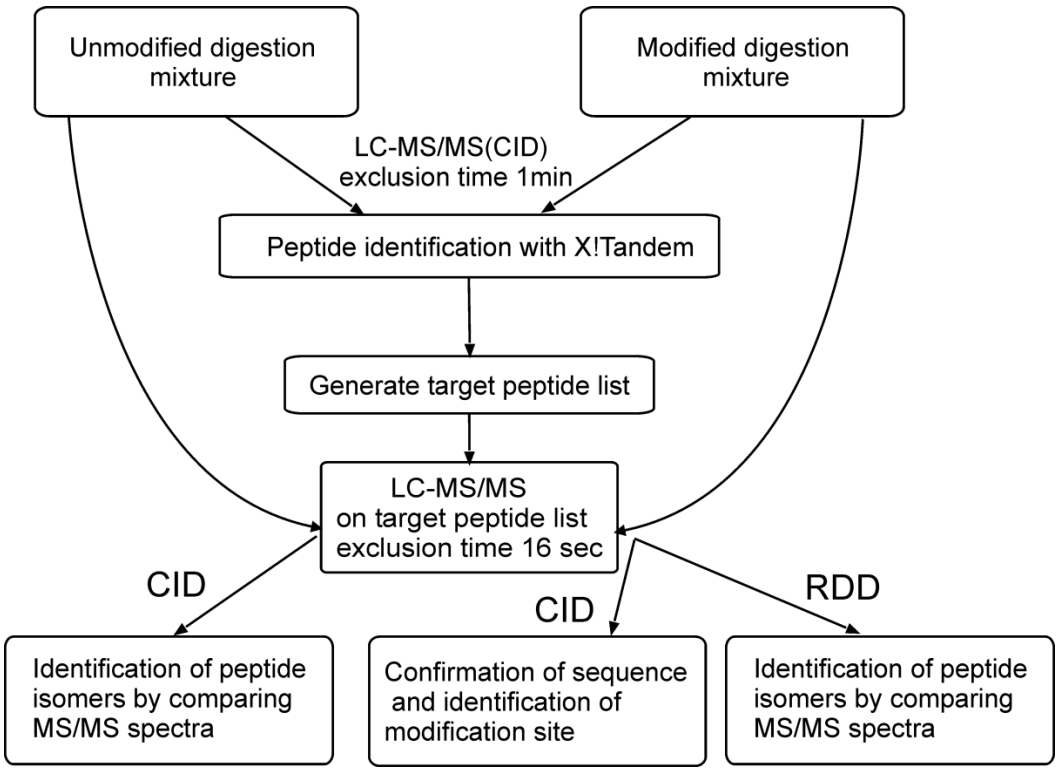
3.3 Results and Discussion

3.3.1 General Approach

Our general experimental procedure for identifying peptide epimers in a mixture of proteins is shown in Scheme 3.1. Following protein isolation and digestion with trypsin, the sample is split into two pools and half of it is covalently modified. There are two advantages to covalent modification. First, the covalent modification allows for incorporation of a chromophore suitable for converting the peptides into radical species and analysis by RDD. Previous work has demonstrated that RDD is the most sensitive method for epimer discrimination.[28] Second, the covalent modification frequently enhances the separation of epimers as is described in greater detail below. There are also obvious disadvantages to covalent modification, including loss of sample and additional experimental complexity. Therefore, the unmodified samples are also analyzed using standard CID. Although CID provides less ability to distinguish epimers, more peptides and particularly those in low abundance are able to be analyzed. Given that distinguishing epimers is a difficult task, it is also beneficial to carry out the analysis using two independent methods. Both the modified and unmodified samples are then subjected to a typical proteomics LC-MS run using CID for the purpose of peptide identification.

Using this information, a target peptide list is generated and all charge states of these peptides are then examined in a second LC-MS run by both CID and RDD with a 16 second exclusion window. Multiple charge states are examined because R-values frequently vary significantly for different charge states. The 16 second exclusion window

is used because it enables the examination and re-examination of up to four co-eluting species within typical LC peak widths. Therefore, every peptide on the target list will be examined multiple times, even if it is only at the leading and trailing edges of a single LC peak (see discussion below for the significance of re-examining the same LC peak). Although this strategy will limit the complexity of sample that can be analyzed in a single run, it should be possible to mitigate this problem by carrying out additional runs (if needed).



Scheme 3.1. Workflow to identify peptide isomers in a protein digestion mixture by LC-MS/MS

3.3.2 Isomer Separation

Comparison of potentially distinguishing MS/MS spectra first requires that the isomers of interest be evaluated independently of one another. Although separations can be carried out with chiral media,[36] isomers (including epimers) will also separate on traditional columns, which are more frequently used.[31] Incomplete separations complicate analysis and make quantitation significantly more challenging. Fortunately, the covalent modification that we use to install labile bonds for RDD also changes the chromatographic behavior of the modified molecules. An example of this is shown in Figure 3.1 for the peptide DAEFR, which is a small tryptic peptide from β-amyloid. All of the four forms of aspartic acid (L-Asp,D-Asp, L-isoAsp, D-isoAsp) have been detected in the human brain and the isomerization may be related to the pathology of Alzheimer's disease.[37] Complete separation of these four isomers is difficult to achieve due to their structural similarity.[38] Figure 3.1a shows the LC-MS/MS results of mixture of DAEFR, DAEFR, **D**AEFR, **D**AEFR (underlined residues corresponds to D- amino acid, and bold residues correspond to isoaspartic acid). Only three LC peaks are observed because the L-isoAsp and D-Asp containing peptides coelute. The following MS/MS spectra are used to identify each peptide. 4-iodobenzoic acid (structure shown in the inset of Figure 3.1b) is one of the chromophores that can be used for generating radicals for RDD (^{I}X will be used to represent the 4-iodobenzoic acid modification of X, where X is any amino acid).[33] Figure 3.1b shows the LC-MS/MS results of the same four peptide isomers after covalent modification by 4-iodobenzoic acid. All four DAEFR isomers are nearly baseline separated in the LC chromatogram and the following MS/MS spectra are different among

the four peptides. In addition, the elution times are shifted and the elution order is changed. It is clear that the addition of the hydrophobic tag significantly impacts the elution properties of DAEFR, leading to improved separation.

Some peptide isomers do not separate even after modification with 4-iodobenzoic acid. For example, in Figure 3.2a a single peak is detected for the peptide DVGSNK-NH_2 despite the fact that two epimers (both D- and L-Ser) are present in the solution. Fortunately in this case the two fragmentation spectra for the epimers are sufficiently distinct that the presence of two species can still be detected. This is achieved by examination of the leading and trailing edge of the LC peak, which yields the corresponding MS/MS spectra shown in Figure 3.2b and an R value of 3.1. Analysis of these same epimers when injected individually into the instrument yields an R value of 13.[28] It is clear from this data that the epimers are still partially resolved in the LC even though a single peak is apparent on the chromatogram. For comparison, the LC chromatogram and MS/MS spectra for the leading and trailing edge of a peak containing a single synthetic peptide that was added to the sample are shown in Figures 3.2c and 3.2d. In this case, the two spectra are virtually identical as expected.

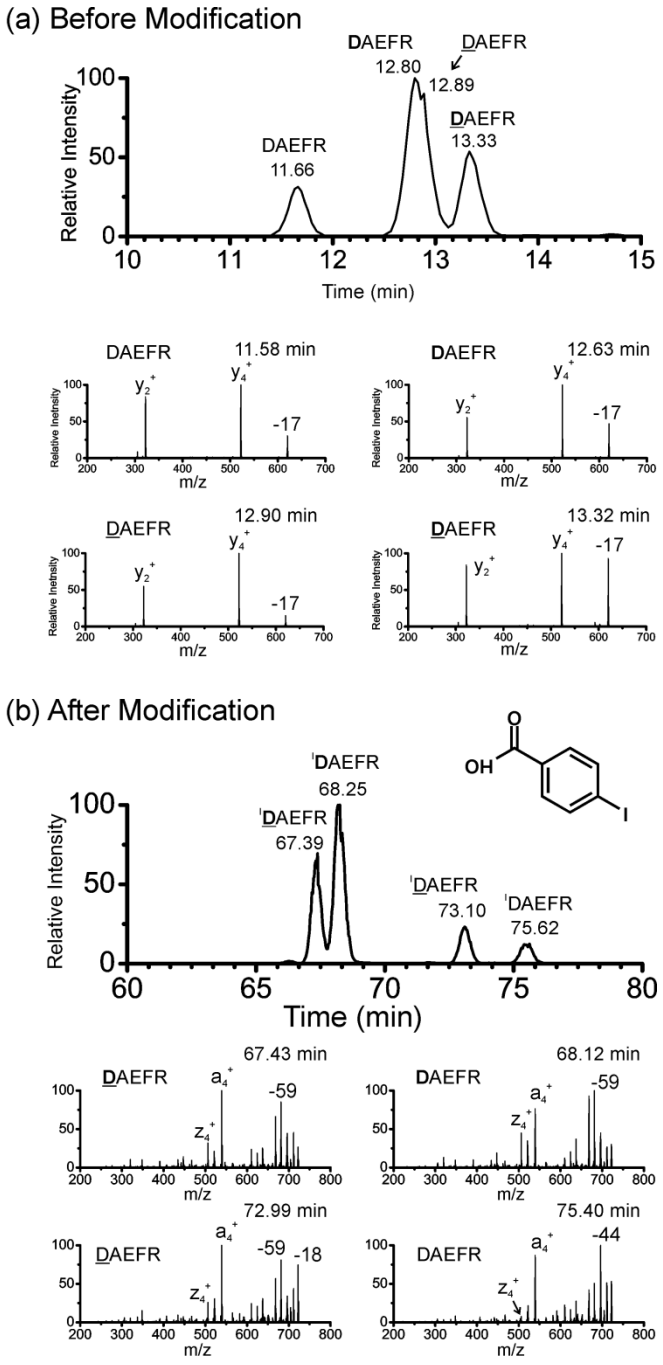


Figure 3.1 LC-MS/MS results for a mixture of the four isomers of synthetic peptide DAEFR (a) before (b) after modification. ^{I}X represents the 4-iodo-benzoic acid modification of X where X is any amino acid. Underlined residues correspond to D-amino acid, and bold residues correspond to isoaspartic acid. The inset of (b) shows the structure of 4-iodo-benzoic acid. The four isomers can be easily separated after covalent modification by 4-iodo-benzoic acid.

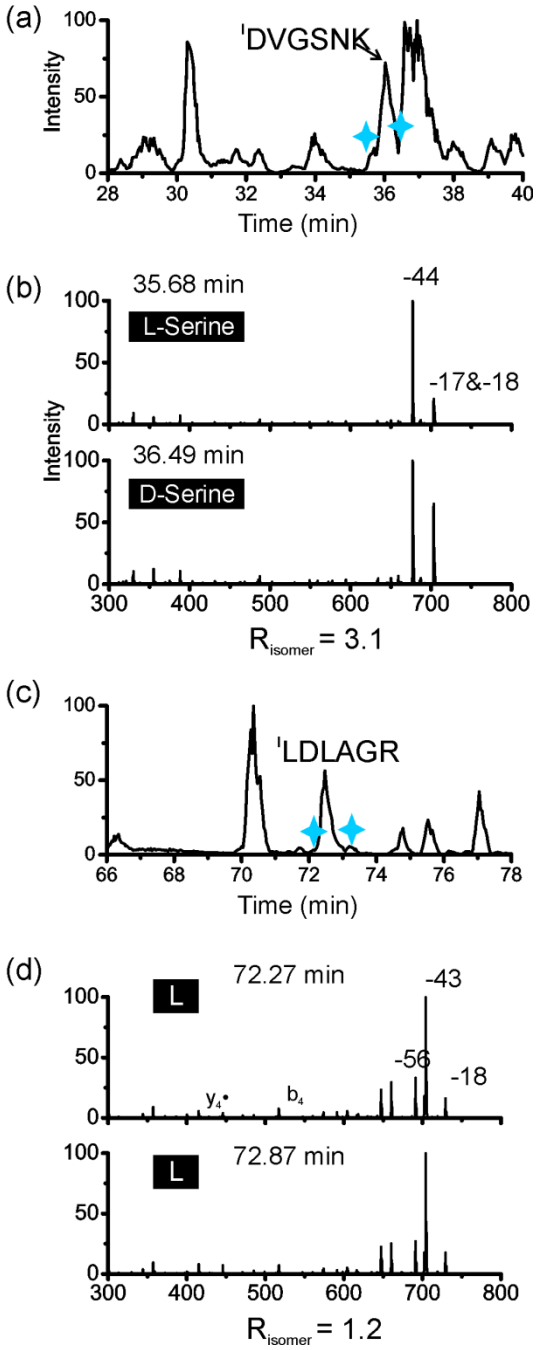


Figure 3.2 (a) LC chromatogram of a peptide mixture containing several synthetic peptides including ^{I}DVGSNK-NH_2, ^{I}DVG<u>S</u>NK-NH_2 and ^{I}LDLAGR. The ^{I}DVGSNK epimers cannot be completely separated. (b) MS^3 (RDD) spectra at 35.68min and 36.49min (the blue asterisks) of (a). The two spectra are significantly different, indicating the LC peak contains two peptides. (c) LC chromatogram of the same run as (a) during a later elution time. (d) MS^3 (RDD) spectra at 72.27min and 72.87min (the blue asterisks) of (c). The two spectra are almost identical.

3.3.3 Data Analysis

When attempting to identify potentially isomerized peptides in unknown samples, candidate peptides from different LC peaks are selected if they have the same mass and exhibit similar fragmentation patterns. In addition, the leading and trailing edge of each individual LC peak is examined for differing MS/MS spectra. In each situation, the R value is calculated from the relevant tandem MS spectra. Theoretically, any R_{isomer} value which is bigger than 1 should indicate the presence of isomers. However, in reality, MS/MS spectra are not perfectly reproducible (especially on the limited LC-MS timescale) and the relative fragmentation abundances of the same peptide in two different spectra acquired at different times are always slightly different. Previous studies have reported that values of $R > 1.2$ are sufficient to indicate statistically significant differences in spectra, based on the reproducibility of ion intensity measurements from direct injection experiments.[24] However, in an LC-MS run, the error is higher because only a few scans are averaged to obtain the MS/MS spectra for one peptide. In contrast, spectra are usually averaged over 50-100 scans in direct infusion experiments, which significantly reduces variation of the mean intensities. To establish the relevant threshold for R values reported herein, a standard LC-MS run was performed on a mixture of synthetic peptides without any isomers using both CID and RDD under identical conditions to those employed on actual samples. The R values for each peptide were calculated from two dissociation spectra at the beginning and the end of each LC peak. Six R values were obtained this way for CID: 1.13, 1.20, 1.88, 1.55, 1.27, and 1.35. A standard t-test was performed and the 99% confidence interval corresponds to R values

from 0.94 to 1.85. Based on these numbers, we have set the threshold to identify peptide isomers from CID fragmentation in this work to R values > 1.9. For RDD, the ion intensities vary slightly more and the threshold to identify peptide isomers corresponds to R values > 2.4 (See supporting information for detailed values). Finally, to eliminate interference from chemical noise, the relative intensities of the fragment pairs for the calculation of R values must be higher than 10% of the base peak in at least one of the two spectra. Furthermore, peaks assigned as sequential fragments and ^{13}C isotopes tend to have higher errors and are not used to determine R values.

Localization of the isomerized residue in a peptide requires comparison of data obtained from biological sources with standard peptides. In these comparisons, the data for sample peptides and standard peptides are acquired in separate LC-MS runs. Hence, more error arises from different ionization efficiency or other random uncertainties between different LC-MS runs. Similar metrics to those used to distinguish dissimilar peptides from each other can be used to identify whether two peptides are likely to be the same. Although the same equation as that for an R value is used, we will refer to similarity scores as S values to avoid confusion. The criteria in this case will be that the S value should be *lower* than the threshold, which will indicate that the two peptides cannot be distinguished from each other and are likely identical. Furthermore, while comparing the tandem mass spectra between synthetic standards and experimental peptides, the experiments for unknown peptides and the standard peptides have to be conducted in two separate LC-MS runs. More errors arise from the different ionization efficiency or other random uncertainties between different LC-MS runs. To establish the S value threshold

for positive identification of peptide isomers by comparing the MS/MS spectra with standard peptides, another set of standard LC-MS runs were performed. Six S values were obtained by comparing the MS/MS spectra acquired during different LC-MS runs of the same peptides. The threshold for positive identification is 1.9 for CID and 3.2 for RDD, i.e. S values below these numbers will indicate that the peptides are likely the same (See supporting information for detailed values).

In addition to the types of amino acid isomerization discussed up to this point, there are a few other modifications that do not result in any mass change that are worth mentioning. For example, cis/trans- proline isomerization can have significant structural implications. However, this type of isomerization is typically dynamic and it is unlikely that such isomers can be chromatographically separated at room temperature.[39] Furthermore, our results do not suggest a bias towards identification of isomerization for peptides that contain proline, as proline is present in both peptides that have and have not undergone isomerization. Another type of modification that could occur in some rare cases would be when two amino acids inverted in sequence,[40] which could lead to separation of the isomers by liquid chromatography. It is not anticipated that such isomers will occur frequently enough to significantly impact our results, and it is possible that these isomers could be identified by analysis of the MS/MS data if fragmentation between the relevant residues is observed. Above all, if any of these alternate possibilities or any other type of isomerization does occur, the isomers could be potentially identified by the method described herein. If the isomerized peptide were determined to be

important, subsequent scrutiny via comparison with authentic standards would reveal the underlying modification responsible for the isomerization.

3.3.4 Sheep Crystallins

We applied the protocol described above to examine crystallin proteins extracted from the eye lens of an approximately one year old sheep. Three variants of crystallin were identified from the ovine database (αA-crystallin, αB-crystallin, βB3- crystallin). A comprehensive list of identified peptide isomers from these three proteins is provided in Table 3.1. The sequence coverage, degree of isomerization, and epimerization for these proteins are summarized in graphical format in Figure 3.3. The sequence coverage (the top bar in Figure 3.3) is excellent for all three proteins. Areas that are missing are primarily due to very short peptide fragments. The degree of isomerization is shown for each protein in the middle bar of Figure 3.3. This bar represents the presence of any isomer that was detected (presumably mostly isoaspartic acid and aspartic acid or serine epimerization). The degree of isomerization is highest for αA-crystallin (67%), indicating that the most abundant protein is also subject to the greatest amount of modification. αB-crystallin and βB3- crystallin exhibit a similar degree of isomerization, just below 50% of the total sequence. In the lower bar in Figure 3.3, the extent of epimerization is shown. We consider a peptide to be epimerized if the number of detected isomers is >4 for peptides containing two aspartic acids, >2 for peptides containing one aspartic acid, or >1 for peptides that lack aspartic acid. The degree of suspected epimerization is again greatest for αA-crystallin. Although it is possible that some of the isomers identified in the middle but not lower bars of Figure 3.3 could also be epimers, this possibility is

probably less likely given that aspartic acid is the most likely residue to isomerize and the rate of isoaspartic acid formation is greater than that of epimerization. Nevertheless, it is possible for exceptions to exist. Additionally, based on the sequence homology to bovine crystallins, several other ovine β- and γ- crystallins were identified by searching the data against bovine database (Table 3.2). [41]

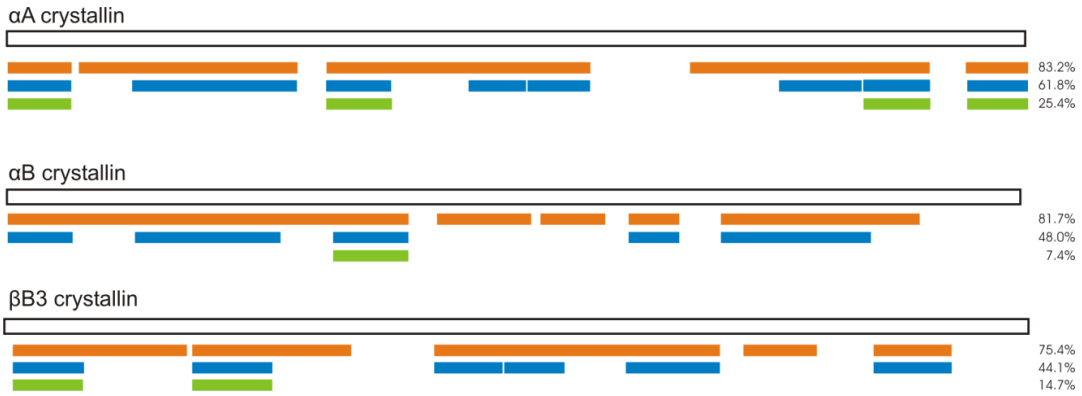


Figure 3.3 Sequence coverage (orange), degree of isomerization (blue) and degree of epimerization (green) for αA-crystallin, αB-crystallin, and βB3- crystallin. The white bar represents the full protein sequence.

Table 3.1 Identified peptide isomers from sheep eye lens digest in ovine database. [a]

Peptide sequence	Crystallin	Number of peaks in LC separation	Relative Abundance (%) [f]	Number of Peptide Isomers Confirmed by MS/MS [b]
Ac-[1]M**D**IAIQHPWF K[11] [c]	αA	4	3.7%, 1.2%, 93.9%, 1.2%	4
[22]LF**D**QFFGEGL FEY**D**LLPFLS STISPYYR[49]	αA	3	7.8%, 32.4%, 59.8%	3

[55]TVL**D**SGISEV R[65]	αA	3	47.4%, 50.8%, 1.8%	4 *g,i*
[79]HFSPE**D**LTVK[88]	αA	2	10.4%, 89.6%	2
[89]VQE**D**FVEIHG K[99]	αA	1	---	2
[104]Q**DD**HGYISR[112] *d,e*	αA	5	4.7%, 21.1%, 15.1%, 3.1%, 56.0%	4
[132]SLSA**D**GMLTF SGPK[145]	αA	2	13.7%, 86.3%	2
[146]VPSGV**D**AGHS ER[157]	αA	3	8.7%, 53.3%, 38.0%	3
[164]EEKPSSAPSS[173]*e*	αA	3	11.8%, 82.3%, 5.9%	3
Ac-[1]M**D**IAIHHPWI R[11]	αB	2	9.1%, 91.9%	2
[23]LF**D**QFFGEHL LES**D**LFPAST SLSPF[47]	αB	2	9.3%, 90.7%	2
[57]APSWI**D**TGLS E MR[69]	αB	6	1.6%, 34.0%, 1.0%, 1.8%, 59.7%, 1.9%	4
[108]Q**D**EHGFISR[116] *d,e*	αB	2	4.2%, 95.8%	2
[124]IPA**D**V**D**PLTI TSSLSS**D**GVL TVNGPR[149]	αB	1	---	2
Ac-[2]AEQHSAPEQA AAGK[15] *c*	βB3	1	---	2 *i*
[39]C**ELTAEC**PNL TESLLEK[55] *h*	βB3	2	14.2%, 85.8%	2
[89]W**D**AWSNSHHS **D**SLL[102]	βB3	1	---	2
[103]SLRPLHI**D**GP **D**HK[115]	βB3	1	---	2
[129]MEIV**DDD**VPS LWAHGFQ**D**R[147]	βB3	1	---	2
[180]HWNEW**D**ANQP QLQSVR[195]	βB3	1	---	2

a Single letter codes are used for the amino acids. The underlined residues correspond to most likely sites of epimerization. Bold residues are likely sites of isomerization. The UniprotKB Accession Numbers for the identified proteins are: αA crystallin, Q5ENZ0,

αB crystallin, W5Q0R4, and βB3 crystallin, Q52NW3 [b] For those peptides that are separated by multiple peaks in HPLC, MS/MS spectra are compared carefully to confirm the number of peptide isomers. [c] Ac- represents N-terminal acetylation. [d] N-terminal glutamine cyclization is observed.[42] [e] These peptides were identified by searching the data against a smaller database as detailed in the experimental section above. [f] Calculated from area of peak in the extracted ion chromatogram for the target peptides from unmodified peptide mixture. [g] Determined from the combination of RDD and CID results. [h] The double star represents iodoacetamide modified cysteine (+57Da). [i] Isomers were confirmed by comparing with synthetic peptides.

Table 3.2 Identified ovine crystallin proteins from bovine database.

Protein Name	UniprotKB Acession Number	Sequence Coverage %	Unique Peptides Detected	Expectation Value
β-crystallin A3	P11843	80%	11	-462.3
β-crystallin A4	Q6DTZ8	86%	9	-432.3
β-crystallin B2	P02522	71%	9	-360.9
β-crystallin B1	P07318	57%	10	-302.0
β-crystallin S	P06504	54%	6	-201.6
zeta-crystallin	O97764	49%	9	-189.9
γ-crystallin F	P23005	59%	6	-177.7
γ-crystallin B	P02526	42%	4	-154.6
γ-crystallin D	P08209	32%	3	-99.0
γ-crystallin C	Q28088	17%	2	-53.3
γ-crystallin A	P02527	14%	1	-46.0
β-crystallin A2	P26444	24%	3	-40.1

Identification of peptide isomers as outlined in Table 3.1 and Figure 3.3 does not reveal the actual sites of isomerization. Previous work has revealed that aspartic acid is the most likely site for both isomerization and epimerization due to aging.[14] Therefore, peptides containing aspartic acid are likely modified at this residue. Serine is the second most likely site to undergo spontaneous epimerization.[19] All peptides that we have identified contain either aspartic acid or serine or both. It is possible to positively identify the site of isomerization by synthesizing synthetic standards and comparing the respective MS/MS spectra, as detailed below. Recent work has also demonstrated that ion mobility may be able to pinpoint (or at least narrow down) sites of epimerization, which may simplify site identification in future experiments on crystallins.[43] For the present study, we synthesized a small number of authentic standards to compare with our results. TVLDSGISEVR is a tryptic peptide from α-crystallin which separates into 3 peaks by LC (Figure 3.4). By comparing the MS/MS CID spectra with those obtained from synthetic peptides where all four different forms of aspartic acid were incorporated, two of the three peaks are determined to contain L-isoAsp and D-Asp. The R values for each of these spectra relative to the experimental spectra are 1.3 and 1.7, indicating positive identification (<1.9). The peptide eluting at 26.59 minutes does not match any of the isomers (with R values of 2.7, 3.5, 13.2, 3.9 for L-Asp, D- Asp, L-isoAsp, and D-isoAsp containing synthetic peptides, respectively); indicating that potentially the serine[19] residue is also epimerized or that the remaining two Asp isomers are co-eluting. After iodo-benzoic acid modification, three peaks are again detected by LC (Figure 3.5); however, comparison with the synthetic peptides by RDD reveals that the three peaks

correspond to L-isoAsp, D-isoAsp and L-Asp with corresponding R values of 2.6, 2.4 and 3.1 (< 3.2). Taken together, the combination of RDD and CID confirm the presence of all four Asp isomers, which suggests that serine epimerization may not be the explanation for the unassigned peak in Figure 3.4.

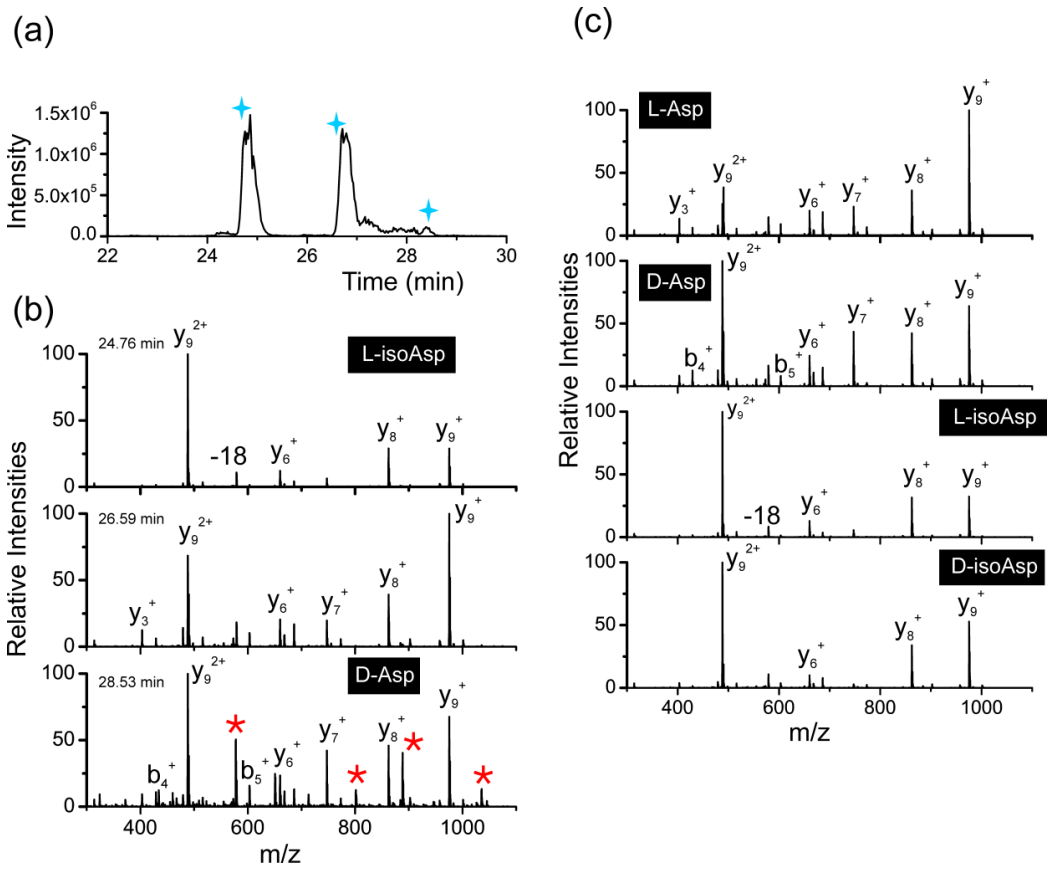


Figure 3.4 (a) LC chromatogram for peptide TVLDSGISEVR in sheep eye crystallin digestion mixture. (b) CID of $[TVLDSGISEVR + 2H]^{2+}$ at 24.76min, 26.59min, 28.53min separately. The spectra at 24.76min and 28.53min are identified as L-isoAsp and D-Asp containing peptide by comparing with (c). The R values for each of these spectra relative to the synthetic peptides spectra are 1.3 and 1.7. The red stars represent fragments from a co-eluting peptide HFSPEDLTVK. The peptide eluting at 26.59 minutes does not match any of the isomers (with R values of 2.7, 3.5, 13.2, 3.9 for L-Asp, D- Asp, L-isoAsp, and D-isoAsp containing synthetic peptides, respectively); indicating that potentially the serine residue is also epimerized or that the remaining two

Asp isomers are co-eluting. (c) CID spectra of four synthetic peptides L-Asp, D- Asp, L-isoAsp, and D-isoAsp TVLDSGISEVR.

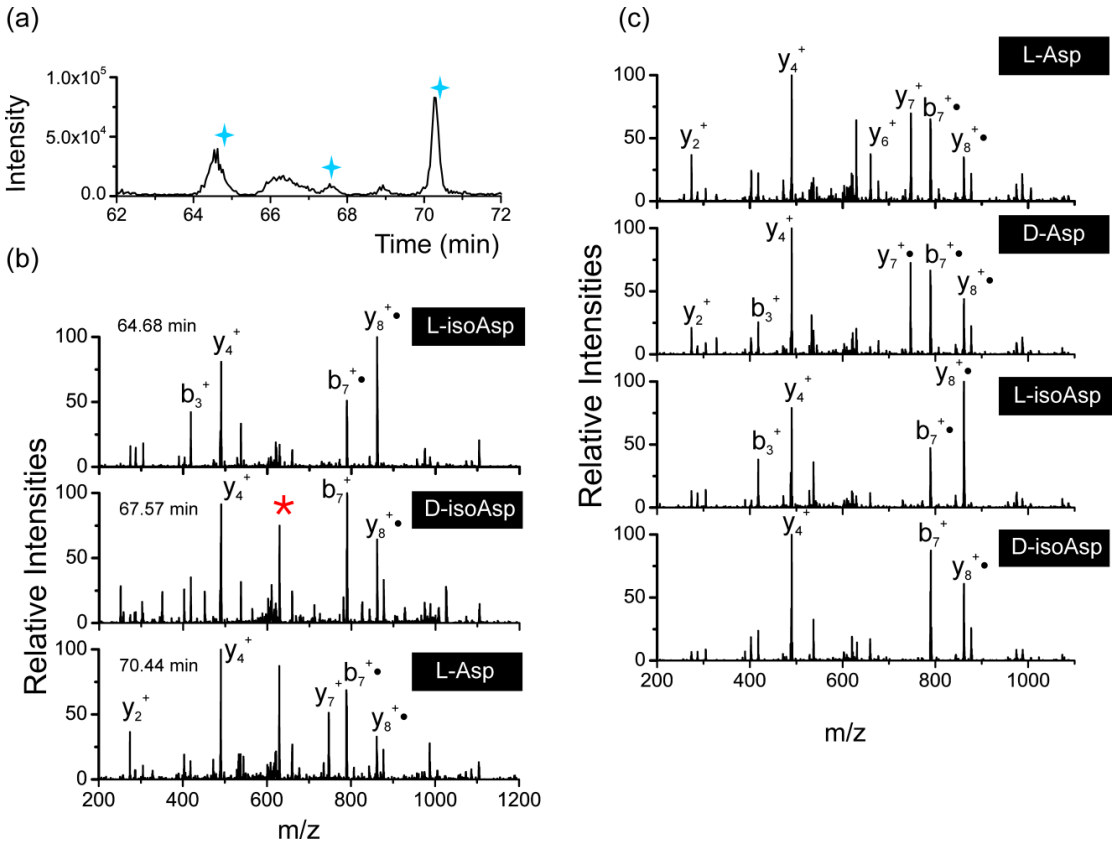


Figure 3.5 (a) LC chromatogram for peptide ^{I}TVLDSGISEVR in sheep eye crystallin digestion mixture after modification. (b) RDD spectra of $[^{I}\text{TVLDSGISEVR} + 2H]^{2+}$ at 64.68min, 67.57min, 70.44min separately. The three spectra are identified as L-isoAsp, D-isoAsp, and L- Asp containing peptide by comparing with (c). The R values for each of these spectra relative to the synthetic peptides spectra are of 2.6, 2.4 and 3.1. The red stars represent fragments from a co-eluting peptide. (c) RDD spectra of four synthetic peptides L-Asp, D- Asp, L-isoAsp, and D-isoAsp ^{I}TVLDSGISEVR.

3.3.5 Discrimination Abilities

In this chapter, the most important strategy to enhance the isomer discrimination ability is to use mass list and short exclusion time in LC-MS/MS run. Specifically, after peptides were identified by database searching of the LC-MS/MS dataset with normal exclusion time (60 second), a mass list of identified peptides is created and the m/z of every possible charge state are included. This is to eliminate all the other contaminations or random non-crystallin peptides to be selected as parent ions in the isomer-identification LC-MS/MS run. Also because the fragmentation pattern and the isomer discrimination ability can vary between different charge states, this strategy greatly enhances the possibility of every peptide isomer to be distinctly identified. For example, an isomerized peptide, HFSPEDLTVK from αA-crystallin, was successfully identified with this approach using CID. The peptide was shown as two separated peaks in LC (Figure 3.6a), and the +2 charge state is prevalent in the full MS spectra. The R_{isomer} value is 1.5 for +2 charge state (Figure 3.6b), smaller than 1.9, which is difficult to differentiate between the situation of isomerization and co-elution. However, after an LC-MS/MS run with the identified peptide mass list including every possible charge state, the +1 charge state of HFSPEDLTVK is selected as the parent ion although the relative intensity is small. The CID spectra of +1 charge state from the two LC peaks show an R_{isomer} value of 7.3 (Figure 3.6c), suggesting the occurrence of isomerization. Results from RDD further confirm the identification; with a significant higher R_{isomer} value of 25.3 (Figure 3.6d and 3.6e).

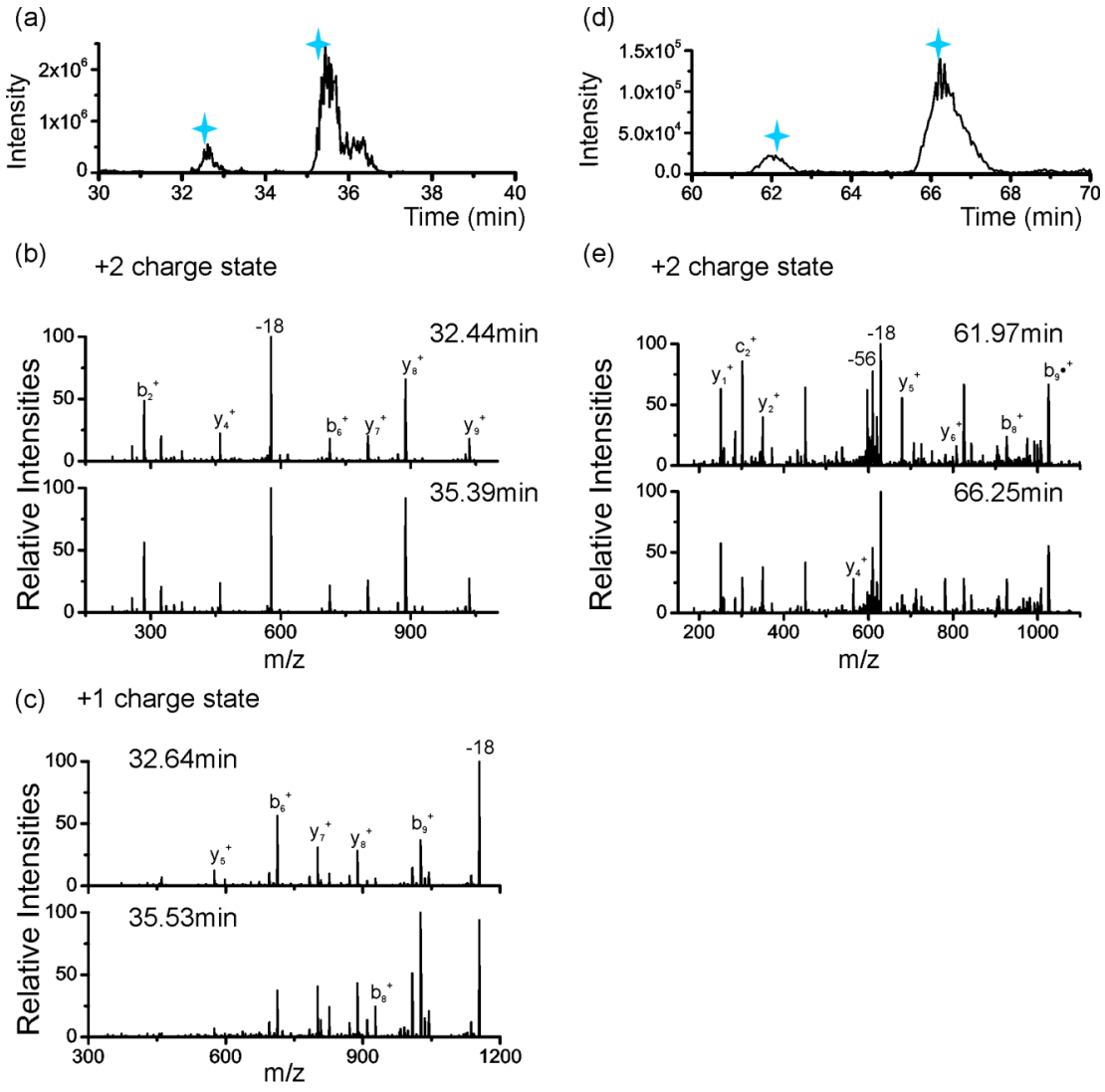


Figure 3.6 LC chromatogram for peptide HFSPEDLTVK in sheep eye crystallin digestion mixture. (b) CID spectra of $[HFSPEDLTVK + 2H]^{2+}$ at 32.44min and 35.39min, separately. The R_{isomer} value is 1.5 (y_9^+ & -18). (c) CID spectra of $[HFSPEDLTVK + H]^+$ at 32.64 and 35.53min. The R_{isomer} value is 7.3 (b_8^+ & y_5^+). The isomer discrimination is achieved at +1 charge state. (d) LC chromatogram for peptide $HFSPEDLTVK^I$ in sheep eye crystallin digestion mixture after modification. (e) RDD spectra of $[HFSPEDLTVK^I + 2H]^{2+}$ at 61.97 and 66.25min. The R_{isomer} value is 25.3 (y_6^+& y_4^+).

Besides the similarity of MS/MS spectra among the peptide isomers, another difficulty in the peptide isomer identification is that not every peptide isomers can be completely separated by LC. Therefore, it is necessary that for every LC peak that was identified as a candidate peptide, the MS/MS spectra of the front and the back of the peak are collected and analyzed carefully. This requires a relatively short exclusion time (shorter than a typical LC peak width). Not surprisingly, in an LC-MS run with 16 second dynamic exclusion time, several peptides show different MS/MS fragmentation abundances within one LC peak, suggesting that these single LC peak may contain multiple partially separated peptide isomers. For example, Figure 3.7 shows the LC-MS/MS results obtained from the sheep eye lens crystallin digestion mixture for peptide Ac-AEQHSAPEQAAAGK (from βB3 crystallin). Only a single LC peak was observed for this peptide as shown in Figure 3.7a. However, using the shorter exclusion time and the peptide target list reveals that the MS/MS spectra at the beginning and the end of the LC peak are different, which indicates that at least two peptide isomers are present. By comparing the MS/MS CID spectra with those obtained from synthetic peptides with serine as the epimerization site (Figure 3.7c), the two spectra in Figure 3.7b are confirmed to contain L-serine and D-serine. However, in some other cases, since there could be multiple amino acid epimerization and isomerization sites in one peptide, it is not possible to identify the number of peptide isomers exist in one LC peak, only based on the MS^2 results without any standard peptides. Nevertheless, this result demonstrated that even when separations are incomplete, the isomer identification can still be accomplished if the resulting fragmentation chemistry is sufficiently discriminating. It is

also clear that relying solely on separations for isomer identification is a flawed strategy since it is likely that some isomer will not be chromatographically resolved. To conclude, by using short dynamic exclusion time and target peptide mass list with every possible charge state, the number of identified peptide isomers is significantly increased.

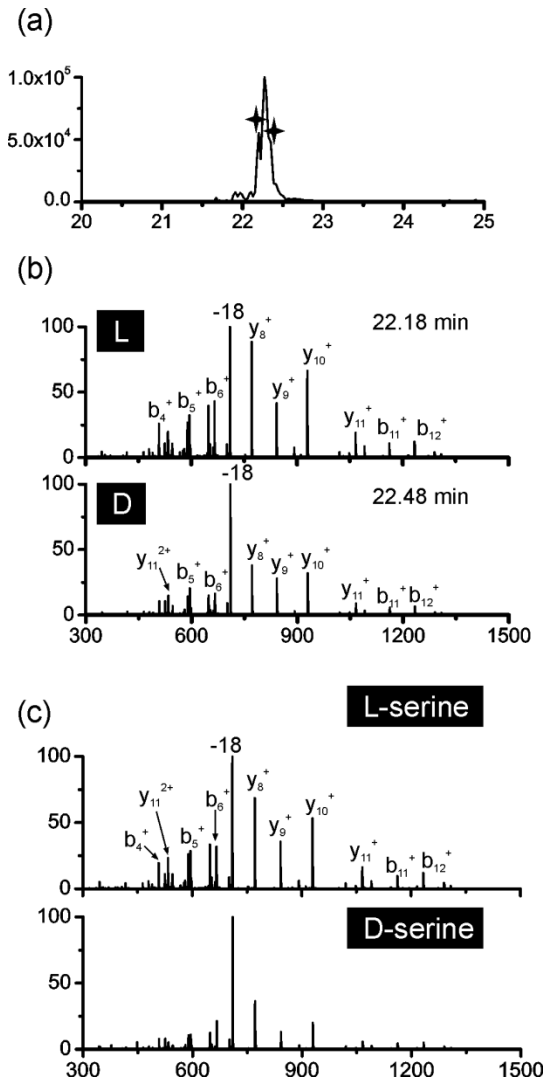


Figure 3.7 (a) LC chromatogram for peptide Ac-AEQHSAPEQAAAGK in sheep eye crystallin digestion mixture (b) CID spectra of $[Ac\text{-}AEQHSAPEQAAAGK + 2H]^{2+}$ at 22.18min and 22.48min, separately. The two spectra are identified as L-serine and D-serine containing peptide by comparing with (c). (c) CID spectra of two synthetic peptides L-serine and D-serine Ac-AEQHSAPEQAAAGK.

Instances of partial separation serve to highlight the importance of using a dissociation method capable of generating high R values, since the partial resolution of the epimers or isomers will reduce discrimination capability. Among all the dissociation techniques of mass spectrometry that have been applied for isomer discrimination, RDD provides the highest R value. Some of the previously reported R values for CID of peptide epimers are smaller than 1.9, which is within the error described before and will bias the isomerization identification in a large-scale proteomics analysis. In contrast, the R value for RDD reported so far are relative high enough (>2.4). Table 3.3 shows the R_{isomer} values for peptide isomers that identified from the ovine database in the sheep eye lens crystallin digestion mixture. For those peptides listed in Table 3.1 but not in Table 3.3, the modified versions were not observed. It is reasonable because the isomerized peptides are usually less abundant and can be lost during the modification procedure. This tends to be the major drawback of RDD. Nevertheless, for most of the peptides that identified by RDD, the R_{isomer} values are higher than CID. There are a few peptides that have comparable R_{isomer} values between RDD and CID; but no significantly better value was observed for CID. This suggests the advantages of using RDD; in the case where complete separation cannot be achieved by LC (as shown in Figure 3.7), a method that capable of generating high R values will provide a better chance for detecting isomers. Also in the case of peptide TVLDSGISEVR described before, both RDD and CID are important for the identification of four Asp isomer peptides. Above all, since the detection of peptide isomers is an extremely hard work, it is always good to have multiple

methods. The combination of RDD and CID is necessary in order to achieve convincing isomer identification and maximum discrimination ability.

Table 3.3 R_{isomer} value of the identified isomer peptides for both RDD and CID. [a,b]

Peptide sequence	CID	RDD
Ac-^{1}M**D**IAIQHPWFK^{11}	11.7	17.9
^{55}TVL**DS**GI**S**EVR^{65}	20.4	47.2
^{79}HF**S**PE**D**LTVK^{88}	7.3	25.3
^{89}VQE**D**FVEIHGK^{99}	5.8	3.5
^{146}VP**S**GV**D**AGH**S**ER^{157}	4.0	3.3
^{57}AP**S**WI**D**TGL**S**EMR^{69} [c]	10.9	18.8
Ac-^{2}AEQH**S**APEQAAAGK^{15}	2.7	2.8

[a] If multiple peaks were observed, only the highest R_{isomer} value is reported here. [b]The table does not contain peptides isomers that can only be identified by CID because no modification product was observed.

Another interesting aspect worth careful examination is that both the LC separation and MS/MS fragmentation are related to peptide structures. Especially in RDD, the radical migration is very structurally sensitive and can be used to probe peptide/protein tertiary structures.[44] Figure 3.8 shows the relationship between the retention time difference and the R_{isomer} value for peptide isomers listed in Table 3.1 and Table 3.3. For

CID, no clear linear relationship is observed; however, there is a trend of higher R value accompanied with better separation. The RDD results show a relatively good linear relationship between the retention time difference and the R value; but the number of RDD data shown here is not substantial enough to draw a firm conclusion. Nevertheless, there could be a correlation between the isomer separation ability of HPLC (retention time difference) and the isomer discrimination ability of mass spectrometry (R value). Additional experiments or a larger dataset are required for further investigation.

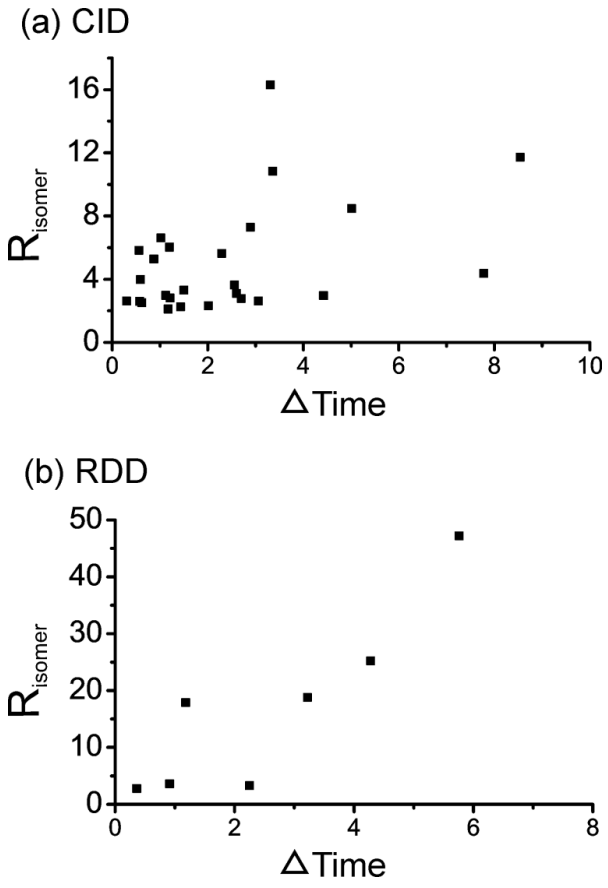


Figure 3.8 Relationship between retention time difference in HPLC and the R_{isomer} value of peptide isomers listed in Table 3.1, for CID (a) and RDD (b).

3.3.6 Isomerization and Functionality

The degree of isomerization for both αA-crystallin and αB-crystallin is shown in Figure 3.9 as a function of the structural region of the protein (i.e. N-terminal region, alpha crystallin domain, or C-terminal extension). In Figure 3.9, the sequence of the protein is represented by the color coded bar with aspartic acid and serine residues indicated by black and white asterisks, respectively. Below each bar, the degree of isomerization detected within each peptide is shown (as determined by the total fractional abundance of isomers). It is clear that αA-crystallin undergoes the greatest extent of isomerization, with significant amounts of isomerization being detected in both the N-terminal domain and the C-terminal extension. The alpha crystallin domain, which corresponds to the globular part of the structure, is isomerized to a lesser extent although some modification is noted.

Although some level of isomerization is found throughout the crystallins, as shown in Figure 3.3, the degree of isomerization is highest in the disordered N-terminal and C-terminal domains, as shown in Figure 3.9. Since these disordered regions are largely responsible for regulating assembly of the higher order oligomers, it is certainly possible that epimerization or formation of isoaspartic acid could significantly perturb the delicate structural interactions that regulate the overall oligomerization state of the crystallins and significantly impact their ability to function properly. Furthermore, it is known that a site of isomerization identified in Figure 3.9 that occurs within the central domain, ^{79}HFSPEDLTVK^{88}, is an important chaperone binding interaction site.[45] The accumulation of isomerization over time in these critical structural regions may

contribute to cataract formation via loss of chaperone functionality, as has been suggested previously.

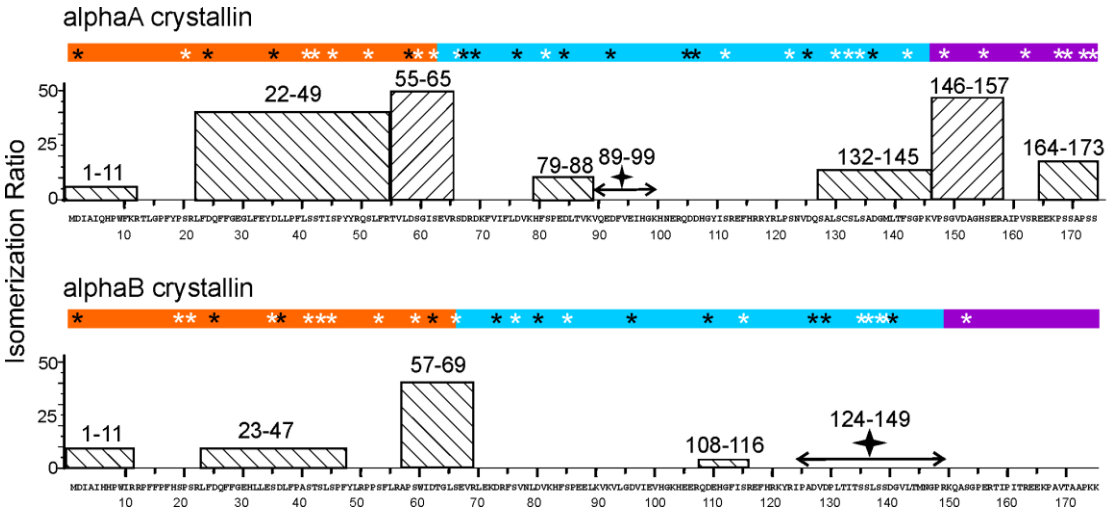


Figure 3.9 Isomerization ratio of αA and αB crystallins. Different colors indicate the three structural regions of crystallin, with the N-terminal region in orange, the alpha crystallin domain in blue, and the C-terminal extension in purple.[46] The black asterisks represent aspartic acid residues and the white asterisks represent serine residues. The black stars (peptide 89-99 in αA and peptide 124-149 in αB) indicate regions where isomerization was detected but not quantified due to incomplete separation by HPLC.

3.4 Conclusions

Although the detection of isomeric post-translational modifications in proteomics experiments poses additional challenges relative to traditional proteomics, the importance of these modifications justifies the development of methods for examining them. Herein I have demonstrated that LC-MS can successfully identify isomers with excellent overall sequence coverage. Importantly, isomer proteomics differs from traditional proteomics in several important ways. First, while re-examination of a particular peptide is typically avoided in traditional proteomics, it is required in isomer proteomics because isomers can only be revealed by comparison of their respective MS/MS spectra. Second, the use of

multiple fragmentation methods and chemical derivatization is highly beneficial because the chances for separating and distinguishing isomers are significantly increased when the analysis is carried out with several orthogonal methods. Overall, these differences will place constraints on the complexity of samples that can be examined; however, we have demonstrated that over a dozen target proteins from a biological sample can be examined.

Results obtained on eye lens sheep crystallins reveal that a significant amount of isomerization can be observed even for a young animal. Both epimerization of aspartic acid and serine containing peptides was observed. The most abundant protein, αA-crystallin was isomerized and epimerized to the greatest extent. Examination of the isomerization in relation to the structural regions of αA-crystallin reveals that modification is more prevalent in regions of the protein that are not structurally well ordered. These modifications may impact in the functionality of αA-crystallin and may be one of the causes of age-related cataract. The greater isomerization of unstructured regions may also suggest that natively disordered proteins are more susceptible to isomerization, although this idea will have to be examined in future studies.

References

[1] Sharma, K. K.; Santhoshkumar, P. Lens aging: effects of crystallins. *Biochim. Biophys. Acta* **2009**, *1790*, 1095–1108.

[2] Bloemendal, H.; de Jong, W.; Jaenicke, R.; Lubsen, N. H.; Slingsby, C.; Tardieu, A. Ageing and vision: structure, stability and function of lens crystallins. *Prog. Biophys. Mol. Biol.* **2004**, *86*, 407–485.

[3] Bloemendal, H., The vertebrate eye lens. *Science* **1977**, *197*, 127-138.

[4] Delbecq, S. P.; Klevit, R. E. One size does not fit all: the oligomeric states of αB crystallin. *FEBS Lett.* **2013**, *587*, 1073–1080.

[5] Horwitz, J. Alpha crystallin: the quest for a homogeneous quaternary structure. *Exp. Eye Res.* **2009**, *88*, 190–194.

[6] Clark, A. R.; Lubsen, N. H.; Slingsby, C. sHSP in the eye lens: crystallin mutations, cataract and proteostasis. *Int. J. Biochem. Cell Biol.* **2012**, *44*, 1687–1697.

[7] Baldwin, A. J.; Lioe, H.; Robinson, C. V; Kay, L. E.; Benesch, J. L. P. αB-crystallin polydispersity is a consequence of unbiased quaternary dynamics. *J. Mol. Biol.* **2011**, *413*, 297–309.

[8] Hains, P. G.; Truscott, R. J. W., Post-translational modifications in the nuclear region of young, aged, and cataract human lenses. *J. Proteome Res.* **2007**, *6*, 3935-3943.

[9] Hoehenwarter, W.; Klose, J.; Jungblut, P. R., Eye lens proteomics. *Amino Acids* **2006**, *30*, 369-389.

[10] Spector, A.; Roy, D., Disulfide-linked high molecular-weight protein associated with human cataract. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*, 3244-3248.

[11] Perry, R. E.; Swamy, M. S.; Abraham, E. C., Progressive changes in lens crystallin glycation and high-molecular-weight aggregate formation leading to cataract development in streptozotocin-diabetic rats. *Exp. Eye Res.* **1987**, *44*, 269-282.

[12] Wilmarth, P. A.; Tanner, S.; Dasari, S.; Nagalla, S. R.; Riviere, M. A.; Bafna, V.; Pevzner, P. A.; David, L. L., Age-related changes in human crystallins determined from comparative analysis of post-translational modifications in young and aged lens: Does deamidation contribute to crystallin insolubility? *J. Proteome Res.* **2006**, *5*, 2554-2566.

[13] Ritz-Timme, S.; Collins, M. J., Racemization of aspartic acid in human proteins. *Ageing Res. Rev.* **2002**, *1*, 43-59.

[14] Fujii, N., D-amino acid in elderly tissues. *Biol. Pharm. Bull.* **2005**, *28*, 1585-1589.

[15] Cloos, P. A. C.; Fledelius, C., Collagen fragments in urine derived from bone resorption are highly racemized and isomerized: a biological clock of protein aging with clinical potential. *Biochemical J.* **2000**, *345*, 473-480.

[16] Geiger, T.; Clarke, S., Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides – succinimide-linked reactions that contribute to protein-degradation. *J. Bio. Chem.* **1987**, *262* , 785-794.

[17] Krause, E.; Beyermann, M.; Dathe, M.; Rothemund, S.; Bienert, M. Location of an amphipathic alpha-helix in peptides using reversed-phase HPLC retention behavior of D-amino acid analogs. *Anal. Chem.* **1995**, *67*, 252–258.

[18] Kaneko, I.; Yamada, N.; Sakuraba, Y.; Kamenosono, M.; Tutumi, S., Suppression of mitochondrial succinate-dehydrogenase, a primary target of beta-amyloid, and its derivative racemized at ser residue. *J. Neurochem.* **1995**, *65*, 2585-2593.

[19] Hooi, M. Y. S.; Raftery, M. J.; Truscott, R. J. W., Age-dependent racemization of serine residues in a human chaperone protein. *Protein Sci.* **2013**, *22*, 93-100.

[20] Hooi, M. Y. S.; Truscott, R. J. W., Racemisation and human cataract. d-Ser, d-Asp/Asn and d-Thr are higher in the lifelong proteins of cataract lenses than in age-matched normal lenses. *Age* **2010**, *33*, 131-141.

[21] Sargaeva, N. P.; Lin, C.; O'Connor, P. B., Identification of Aspartic and Isoaspartic Acid Residues in Amyloid beta Peptides, Including A beta 1-42, Using Electron-Ion Reactions. *Anal. Chem.* **2009**, *81*, 9778-9786.

[22] Bai, L.; Romanova, E. V.; Sweedler, J. V., Distinguishing Endogenous D-Amino Acid-Containing Neuropeptides in Individual Neurons Using Tandem Mass Spectrometry. *Anal. Chem.* **2011**, *83*, 2794-2800.

[23] Adams, C. M.; Zubarev, R. A., Distinguishing and quantifying peptides and proteins containing D-amino acids by tandem mass spectrometry. *Anal. Chem.* **2005**, *77*, 4571-4580.

[24] Serafin, S. V.; Maranan, R.; Zhang, K. L.; Morton, T. H., Mass spectrometric differentiation of linear peptides composed of L-amino acids from isomers containing one D-amino acid residue. *Anal. Chem.* **2005**, *77*, 5480-5487.

[25] Sachon, E.; Clodic, G.; Galanth, C.; Amiche, M.; Ollivaux, C.; Soyez, D.; Bolbach, G., D-Amino Acid Detection in Peptides by MALDI-TOF-TOF. *Anal. Chem.* **2009**, *81*, 4389-4396.

[26] Gonzalez, L. J.; Shimizu, T.; Satomi, Y.; Betancourt, L.; Besada, V.; Padron, G.; Orlando, R.; Shirasawa, T.; Shimonishi, Y.; Takao, T., Differentiating alpha- and beta-aspartic acids by electrospray ionization and low-energy tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2092-2102.

[27] Tao, W. A.; Zhang, D. X.; Nikolaev, E. N.; Cooks, R. G., Copper(II)-assisted enantiomeric analysis of D,L-amino acids using the kinetic method: Chiral recognition and quantification in the gas phase. *J. Am. Chem. Soc.* **2000**, *122*, 10598-10609.

[28] Tao, Y.; Quebbemann, N. R.; Julian, R. R., Discriminating D-Amino Acid-Containing Peptide Epimers by Radical-Directed Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* , 6814-6820.

[29] Hurtado, P. P.; O'Connor, P. B., Differentiation of isomeric amino acid residues in proteins and peptides using mass spectrometry. *Mass Spectrom. Rev.* **2012**, *31*, 609-625.

[30] Yang, H.; Fung, E. Y. M.; Zubarev, A. R.; Zubarev, R. A., Toward Proteome-Scale Identification and Quantification of Isoaspartyl Residues in Biological Samples. *J. Proteome Res.* **2009**, *8* , 4615-4621.

[31] Sargaeva, N. P.; Goloborodko, A. A.; O'Connor, P. B.; Moskovets, E.; Gorshkov, M. V., Sequence-specific predictive chromatography to assist mass spectrometric analysis of asparagine deamidation and aspartate isomerization in peptides. *Electrophoresis* **2011**, *32*, 1962-1969.

[32] Chan W. C.; White P. D. *Fmoc Solid Phase Peptide Synthesis.* 1st ed.; Oxford University Press: New York, 2000.

[33] Ly, T.; Zhang, X.; Sun, Q. Y.; Moore, B.; Tao, Y. Q.; Julian, R. R., Rapid, quantitative, and site specific synthesis of biomolecular radicals from a simple photocaged precursor. *Chem. Commun.* **2011**, *47*, 2835-2837.

[34] McKerrow, J. H., Non-enzymatic, post-translational, amino-acid modifications in aging – brief review. *Mech. Ageing Dev.* **1979**, *10*, 371-377.

[35] Diedrich, J. K.; Julian, R. R., Facile identification of photocleavable reactive metabolites and oxidative stress biomarkers in proteins via mass spectrometry. *Anal. Bioanal. Chem.* **2012**, *403*, 2269-2277.

[36] Bai, L.; Sheeley, S.; Sweedler, J. V., Analysis of endogenous D-amino acid-containing peptides in Metazoa. *Bioanal. Rev.* **2009**, *1*, 7-24.

[37] Fukuda, H.; Shimizu, T.; Nakajima, M.; Mori, H.; Shirasawa, T., Synthesis, aggregation, and neurotoxicity of the Alzheimer's A beta 1-42 amyloid peptide and its isoaspartyl isomers. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 953-956.

[38] Roher, A. E.; Lowenson, J. D.; Clarke, S.; Wolkow, C.; Wang, R.; Cotter, R. J.; Reardon, I. M.; Zurcherneely, H. A.; Heinrikson, R. L.; Ball, M. J.; Greenberg, B. D., Structural alterations in the peptide backbone of beta-amyloid core protein may account for its deposition and stability in alzheimers-disease. *J. Biol. Chem.* **1993**, *268*, 3072-3083.

[39]Kalman, A.; Thunecke, F.; Schmidt, R.; Schiller, P. W.; Horvath, C., Isolation and identification of peptide conformers by reversed-phase high-performance liquid chromatography and NMR at low temperature. *J. Chromatogr., A* **1996**, *729*, 155-171.

[40] Brüeckner, C.; Bunz, S-C.; Neusüß, C.; Scriba, G. K. E., CE-MS Identification of Amino Acid Sequence Inversion as a New Degradation Pathway of an Aspartyl Model Tripeptide. *Chromatographia* **2012**, *75*, 1205-1210.

[41] Robertson, L. J. G.; David, L. L.; Riviere, M. A.; Wilmarth, P. A.; Muir, M. S.; Morton, J. D., Susceptibility of ovine lens crystallins to proteolytic cleavage during formation of hereditary cataract. *Invest. Ophthalmol. Vis. Sci.* **2008**, *49*, 1016-1022.

[42] Khandke, K. M.; Fairwell, T.; Chait, B. T.; Manjula, B. N. *Int. J. Pept. Protein Res.* **1989**, *34*, 118-123.

[43] Jia, C.; Lietz, C. B.; Yu, Q.; Li, L., Site-specific characterization of D-amino acid containing peptide epimers by ion mobility spectrometry. *Anal. Chem.* **2014**, *86*, 2972-2981

[44] Ly, T.; Julian, R. R., Elucidating the tertiary structure of protein ions *in Vacuo* with site specific photoinitiated radical reactions. *J. Am. Chem. Soc.* **2010,** *132*, 8602-8609.

[45] Sharma, K. K. Identification of 1,1'-Bi(4-anilino)naphthalene-5,5'-disulfonic Acid Binding Sequences in alpha –Crystallin. *J. Biol. Chem.* **1998**, *273*(25), 15474–15478.

[46] Kim, K. K.; Kim, R.; Kim, S. H., Crystal structure of a small heat-shock protein. *Nature* **1998**, *394* (6693), 595-599.

Supporting Information

Table S 3.1 Source ions for R and S values

R Values for CID

Peptide	Fragment 1	Fragment 2	R value
RGYALG	a_5^+	-17	1.13
MEHFRW	b_3^+	-17	1.20
LVFFAEDVGSNK	b_{10}^+	b_5^+	1.88
RPPGFSPFR	b_6^+	-17	1.55
PHCKRM	b_5^+	y_4^+	1.27
RYLPT	b_2^+	-17	1.35

R values for RDD

Peptide	Fragment 1	Fragment 2	R value
LDLAGR	-56L	-43L	1.22
IQTGLDATHAER	-72E	-18	1.70
DAEFR	y_4^+	-59E	1.18
HGPLGPL	-56L	-43L	1.21
NGPLQAGQPGER	$y_{11}\bullet^{2+}$	-18	2.01
PSKYEPFV	$b_5\bullet^+$	$y_7\bullet^+$	2..31

S values for CID

Peptide	Fragment 1	Fragment 2	R value
TVLDSGISEVR (L-Asp)	b_2^+	y_6^+	1.63
TVLDSGISEVR (D-isoAsp)	y_4^+	y_8^+	1.13
TVLDSGISEVR (D-Asp)	y_4^+	y_5^+	1.58
TVLDSGISEVR (isoAsp)	y_6^+	-18	1.62
TVLDSGISEVR (L-Asp)	b_2^+	y_9^{2+}	1.72
TVLDSGISEVR (isoAsp)	y_8^+	-18	1.22

S values for RDD

Peptide	Fragment 1	Fragment 2	R value
TVLDSGISEVR (isoAsp)	$y_9\bullet^+$	$y_6\bullet^+$	2.38
TVLDSGISEVR (L-Asp)	b_5^+	$y_3\bullet^+$	2.02
TVLDSGISEVR (D-isoAsp)	b_5^+	$y_9\bullet^+$	2.22
TVLDSGISEVR (D-Asp)	$y_9\bullet^+$	$y_9\bullet^{2+}$	3.21
TVLDSGISEVR (iso-Asp)	$b_3\bullet^+$	y_2^+	2.07
TVLDSGISEVR (L-Asp)	b_6^+	$b_4\bullet^+$	2.70

Table S3.2 Detailed list for all the peptides identified in sheep eye crystallin from Ovine database.[a,b]

Peptide Sequence	Mass Deviation (Da)	Peptide E-value
αA Crystallin Ovine Q5ENZ0		
^{1}MDIAIQHPWFK^{11}R [c]	-0.274	3.4e-009
R^{13}TLGPFYPSR^{21}L	0.491	2.8e-003
R^{22}LFDQFFGEGLFEYDLLPFLSSTISPYYR^{49}Q	0.720	1.5e-011
F^{28}GEGLFEYDLLPFLSSTISPYYR^{49}Q	0.542	8.2e-012
R^{55}TVLDSGISEVR^{65}S	0.146	1.5e-006
R^{66}SDRDKFVIFLDVK^{78}H	1.352	1.3e-003
R^{69}DKFVIFLDVK^{78}H	0.609	2.1e-004
K^{79}HFSPEDLTVK^{88}V	-0.042	1.1e-006
K^{89}VQEDFVEIHGK^{99}H	0.400	7.6e-006
R^{118}YRLPSNVDQSALSCSLSADGMLTFSGPK^{145}V	0.478	1.7e-010
R^{120}LPSNVDQSALSCSLSADGMLTFSGPK^{145}V	0.352	3.0e-013
D^{126}QSALSCSLSADGMLTFSGPK^{145}V	-0.934	2.1e-005
S^{128}ALSCSLSADGMLTFSGPK^{145}V	-0.043	8.0e-006
C^{132}SLSADGMLTFSGPK^{145}V	0.579	1.4e-009
K^{146}VPSGVDAGHSER^{157}A	0.312	4.4e-008

R[164]EEKPSSAPSS[173]	0.424	8.0e-003
αB Crystallin Ovine Q5ENY9		
[1]MDIAIHHPWIR[11]R [c]	0.477	4.1e-007
R[12]RPFFPFHSPSR[22]L	-0.114	1.5e-005
R[23]LFDQFFGEHLLESDLFPASTSLSPF[47]Y	0.639	6.5e-011
F[48]YLRPPSFLR[56]A	0.015	2.2e-003
R[57]APSWIDTGLSEM*R[69]L [d]	0.473	5.7e-009
K[73]DRFSVNLDVK[82]	0.481	1.4e-004
R[75]FSVNLDVKHFSPEELK[90]V	0.775	1.8e-005
K[83]HFSPEELK[90]V	0.519	5.9e-003
K[93]VLGDVIEVHGK[103]H	0.315	6.2e-005
R[108]QDEHGFISR[116]E [e]	1.107	8.0e-003
K[122]YRIPADVDPLTITSSLSSDGVLTV*NGPR[149]K [f]	0.413	1.2e-010
R[124]IPADVDPLTITSSLSSDGVLTMNGPR[149]K	0.828	2.4e-012
[151]**K**QASGPER[158]T	0.124	9.8e-003
βB3 Crystallin Ovine Q52NW3		
M[2]AEQHSAPEQAAAGK[15]S [c]	0.336	1.3e-009
K[16]SHGGLGGSYK[25]V	0.424	1.7e-005
K[26]VIVYEMENFQGK[37]R	0.279	1.5e-009
R[39]CELTAECPNLTESLLEK[55]V	0.315	3.2e-012
K[56]VGSIQVESGPWLAFER[71]R	0.400	5.4e-012

$R^{89}WDAWSNSHHSDSLL^{102}S$	0.408	1.1e-004
$L^{103}SLRPLHIDGPDHK^{115}L$	0.496	2.7e-005
$K^{116}LHLFENPAFGGR^{127}K$	0.472	1.0e-009
$R^{128}KMEIVDDDVPSLWAHGFQDR^{147}V$	0.657	1.2e-010
$K^{129}MEIVDDDVPSLWAHGFQDR^{147}V$	0.269	1.3e-011
$^{153}AINGTWVGYEFPGYR^{167}G$	0.652	1.8e-008
$R^{180}HWNEWDANQPQLQSVR^{195}R$	0.531	1.3e-012

[a] Bold residues represent iodo-benzoic acid modified amino acids. [b] Cysteine are observed as iodoacetamide modified (+57Da). [c] N-terminal acetylated peptide. [d] The star represents a point mutation from Val to Met. [e]The N-terminal glutamine cyclization peptide is observed. [f] The star represents a point mutation from Met to Val.

Chapter 4

FACTORS THAT INFLUENCE COMPETITIVE INTERMOLECULAR SOLVATION OF PROTONATED GROUPS IN PEPTIDES AND PROTEINS IN THE GAS PHASE

4.1 Introduction

Although the first examples were synthesized in the 1930's, crown ethers only came to the forefront following recognition of their cation binding properties by Pederson in the late 1960's.[1] Crown ethers are cyclic oligomers, most typically constituted of ethylene oxide unit repeats. The nomenclature of crown ethers refers to the total number of atoms comprising the ring followed by the number of heteroatoms which are present. For example, 18-crown-6 (18C6) is an 18 atom ring with 6 oxygen atoms. The cyclic arrangement of crown ethers creates electronegative cavities which are well suited for binding to cations. Crown ethers are freely soluble in both polar and nonpolar solvents, making them well-suited for phase transfer catalysis.[2] Crown ethers have also been used for cation recognition,[3] in separations,[4] and for gas phase experiments.[5-7]

18C6 in particular has found use in mass spectrometry (MS) based experiments due to its ability to bind protonated primary amines via the formation of 3 specific hydrogen bonds between every other oxygen atom.[8] Attachment of 18C6 to protonated groups has been used to: investigate mechanistic aspects of hydrogen/deuterium exchange,[9] examine protein structure,[10] generate radical peptides,[11] in conjunction with ion mobility,[12] for charge stripping,[13] and for spectroscopy.[14] These experiments are typically conducted with electrospray ionization, where 18C6 is added to the solution to be electrosprayed.

Under sufficiently gentle ionization conditions, noncovalent 18C6 adducts are formed and can be observed in the final mass spectra. It is important to point out that, with 18C6, observation of a noncovalent adduct in the gas phase does not necessarily imply that the complex was present in solution. For example, the dissociation constant for 18C6 with protonated primary amines in water is quite low (in the high millimolar range),[15] yet abundant adducts can be observed from aqueous solutions of peptides or proteins which contain 18C6. It has been proposed that complexation occurs during the electrospray process, where the effective concentration of 18C6 increases dramatically during droplet evaporation and drives complex formation.[16]

Once in the gas phase, the fundamental interaction between 18C6 and several biologically relevant protonated groups is quite strong.[17-19] For example, the bond dissociation energy for protonated butyl amine (a lysine mimic) and 18C6 is 223 kJ/mol, which represents a significant fraction of the energy required to break a typical covalent bond.[18] However, the bond dissociation energies of small molecules cannot be straightforwardly used to estimate the binding of 18C6 to similar functional groups in more complex molecules such as peptides or proteins. For these larger molecules, intramolecular binding sites that can competitively interact with charged groups are frequently available and can significantly reduce the effective binding energy of the crown. Entropy also becomes potentially more important when intramolecular charge solvation is possible because the noncovalent bonding of two molecules is always entropically unfavorable. The stability of the 18C6 adducts in larger molecules therefore depends on competition between intramolecular and intermolecular solvation of charged

sites, keeping in mind that optimal configurations for each may be separated by significant kinetic barriers.

In this chapter, I examine the potential of 18C6 to act as "solvent" for charged residues in the gas phase. It is found that for systems with significant structural flexibility and available hydrogen binding sites, 18C6 adducts are kinetically trapped unstable complexes. Very mild excitation leads to prompt loss of the 18C6 adduct and rearrangement of peptide structure to accommodate the charged group. In more rigidly constrained systems where optimal charge solvation is either not feasible or protected by large kinetic barriers, 18C6 adduct stability is significantly enhanced. Molecular mechanics calculations on several small glycine oligomers are used to examine the structures of specific examples of unstable and stable 18C6 adducts. Collisional activation of multi-adduct proteins reveals that adduct stability increases significantly with increasing charge state, which may reflect the degree of structural reorganization that has taken place in transit into the gas phase.

4.2 Experimental Methods

4.2.1 Materials

All the organic solvents were purchased from Sigma-Aldrich (St. Louis, MO) and used without further purification. All amino acids and resin were purchased from Ana Spec (Fremont, CA). Cytochrome c (cytc) and myoglobin were purchased from Sigma-Aldrich (St. Louis, MO). 18-crown-6 was purchased from Alfa Aesar (Pelham, NH). Water was purified to 18.2 MΩ using a Millipore 147 (Billerca, MA) Direct-Q system. Peptides used

in the radical directed dissociation experiments were purchased from American Peptide (Sunnyvale, CA).

4.2.2 Peptide Synthesis and Iodination

The GG, KG, RG peptide series were synthesized manually using standard fmoc procedures with Wang Resin as the solid support.[20] Amino acids with protected side chains were used when needed. The N-terminus of KG and RG peptide series were acetylated by acetic acid before the cleavage from beads (indicated by Ac- prior to the peptide sequence).

Peptide EGVYVHPV were iodinated using a previously published method.[21] Sodium iodide, chloramine-T and peptide were mixed together in water with 1:1:1 molar ratio. After 1 min reaction, sodium metabisulfite with 2 times of the peptide concentration was added to the solution to quench the excess I_2. The reaction mixture was purified via peptide trap (Michrom Bioresource. Inc.).

4.2.3 Mass Spectrometry

Mass spectra were recorded with a Finnigan LTQ ion trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) with a standard ESI source. The concentrations for all peptides and proteins were 10μM. All samples were prepared in 50/50 water/methanol except for holo myoglobin. 18C6 at 5~10 times the peptide/protein concentration was added to the sample before electrospraying. The isolation window width was set to 10 Da for the peptide-crown adduct peak and 10 Da for the protein-adduct peak. This wide isolation window is needed to avoid collisional heating of the ions during isolation. The width is also 10Da for ultrazoom scan mode (lower scan speed). The activation voltage

(V) was was converted from normalized collisional energy (NCE) using the following equation

$$V = \text{NCE} \times (\text{parent mass} \times \text{tick amp slope} + \text{tick amp intercept})$$

Where tick amp slope and tick amp intercept are instrument-specific parameters, equal to 0.000026 and 0.010236. The degrees of freedom (DOF) for an N-atom molecule can be obtained with: DOF = 3N-6.

For radical directed dissociation, the posterior plate of mass spectrometer was modified to transit fourth harmonic pulses (266nm) from a flash-pumped Nd:YAG laser (Continuum, Santa Clara, CA) through a quartz window into the ion trap. Photodissociation (PD) of the iodinated peptide or peptide-crown complex always produces the loss of iodine as the major fragment. Further MS^n experiments described in detailed below were performed by re-isolation of the radical peptide or radical peptide-crown complex followed by collision induced dissociation (CID). NCE is 10 to dissociate the peptide and crown complex and 13 for peptide fragmentation.

4.2.4 Molecular Mechanics Calculations

Maestro and MacroModel (Schrodinger, Inc., San Diego, CA) were used to build models and perform conformational searches. For all the calculations, the OPLS atomic force field was used, with no solvent. The initial peptide structures were alpha helical with the positive charge on the N-terminus for GGGG and the lysine side chain for Ac-KGGG. For the GGGG complex with 18C6, the 18C6 molecule was manually placed next to the N-terminus. For Ac-KGGG-18C6 complex, the lysine side chain was rotated away from the peptide backbone into a position where 18C6 could interact. A torsional

sampling conformational search algorithm (MCMM) was used to search conformational space and identify low energy structures. 50000 structures were sampled in each conformational search and 1000 low energy structures were saved. The lowest energy structures obtained in each case were used to calculate binding energies (ΔH between isolated and complexed, minimized structures).

4.3 Results and Discussion

4.3.1 Mass Shift of Crown Complex in the Ion Trap

While performing experiments that utilizing the attachment of 18C6 to generate noncovalent complexes in the gas phase in the ion trap instruments, I have noticed that on some occasions that the masses of the complexes deviate significantly (>0.2 Da at normal scan speed) from the expected masses (Figure 4.1a and Figure 4.1b). In all cases, collisional activation of the peak results in loss of a molecule corresponding nominally to the mass of 18C6 and generation of the peptide with the correct mass (Figure 4.1c), which suggests that the complexes are correctly assigned but appear at the wrong m/z for some reason. It has been well documented that fragile ions can fragment during resonant excitation in ion traps, leading to peak broadening and mass shifts.[22-24] The LTQ is a forward scanning instrument, which means that it ejects ions of lower mass first. Fragile ions ejected prematurely during this process will therefore broaden and mass shift towards the direction of lower m/z. Evaluation of identical samples with time-of-flight (TOF) MS, which does not activate ions during detection (Figure 4.1d), yields masses and isotope patterns that agree well with predicted values.

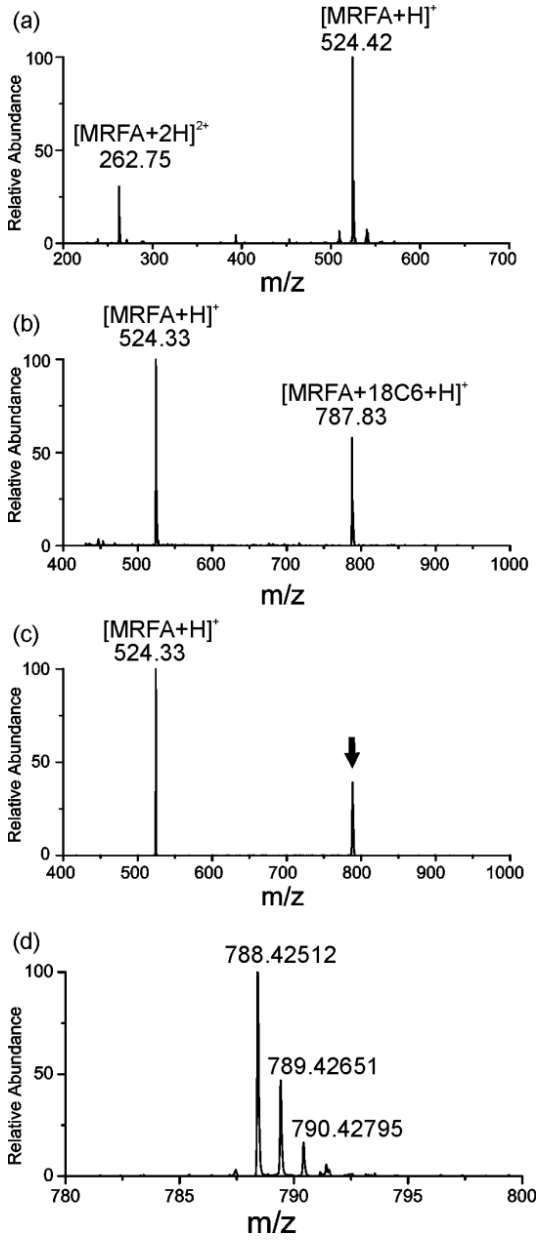


Figure 4.1 (a) Full MS spectrum of peptide MRFA. The exact mass is 524.27 Da for $[MRFA+H]^+$ (b) Full MS spectrum of peptide MRFA and 18C6 acquired on a different day under different operating conditions. The exact mass is 788.42Da for $[MRFA+18C6+H]^+$. The complex exhibits significant mass deviation (0.59 Da) while the peptide peak is observed at correct mass (deviation within 0.2 Da). (c) CID of $[MRFA+18C6+H]^+$. The arrow represents the precursor ion. The peptide mass is correct after CID. (d) Full mass spectrum of MRFA and 18C6 in time-of-flight instrument (zoom in between 780Da and 800Da). The mass and isotope patterns agree well with predicted values.

4.3.2 Standard Peptides

Figure 4.2 illustrates isolation windows for a series of Ac-KG_x (x = 0 to 5) peptides that have been complexed with 18C6. The predicted masses and relative intensities for isotopic peaks are shown for each complex as red lines. The agreement between the expected and observed mass is very good for the 18C6 complex with Ac-K; however, there is a clear trend of increasingly large mass shifts towards lower m/z that is accompanied by peak broadening as the number of glycine residues increases. These results are consistent with decreasing complex stability as more glycine residues are added to the peptide. This assumption is further supported by results from a TOF mass spectrometer, data shown in Figure 4.3. The sample contains five peptides (Ac-$KG_{x,}$ x = 1 to 5) and 18C6. For crown complexes Ac-KG-18C6, Ac-KGG-18C6, Ac-KGGG-18C6, and Ac-KGGGG-18C6 masses and isotope distributions agree well with predicted values. However, the relative abundance for the Ac-KGGGG-18C6 complex is low (the inset of Figure 4.3). No Ac-KGGGGG-18C6 complex is observed, while the bare peptide Ac-KGGGGG without crown complex is present in the spectrum (not shown). These results confirmed that as the number of Glycine residues increases, the crown complex stability is decreasing. The polyglycine part of the backbone is flexible and contains numerous hydrogen bonding sites which can easily interact with the flexible lysine side chain.

The results in Figure 4.2 and 4.3 strongly suggest that the 18C6 adducts represent kinetically trapped structures which rapidly dissociate upon very minimal activation, leading to the observed mass shifts and broadening. In Figure 4.4a, the precursor ion survival yields are shown as a function of activation voltage normalized by the number of

vibrational degrees of freedom.[25,19] Figure 4.4b shows that the trends in peptide complex stability are nearly identical to those observed by mass shifting in Figure 4.2. Here the degree of dissociation (α) is calculated using the following equation:

$$\alpha = \frac{\text{Relative abundance of precursor ion}}{\text{Relative abundance of precursor ion} + \text{Relative abundance of peptide ion}}$$

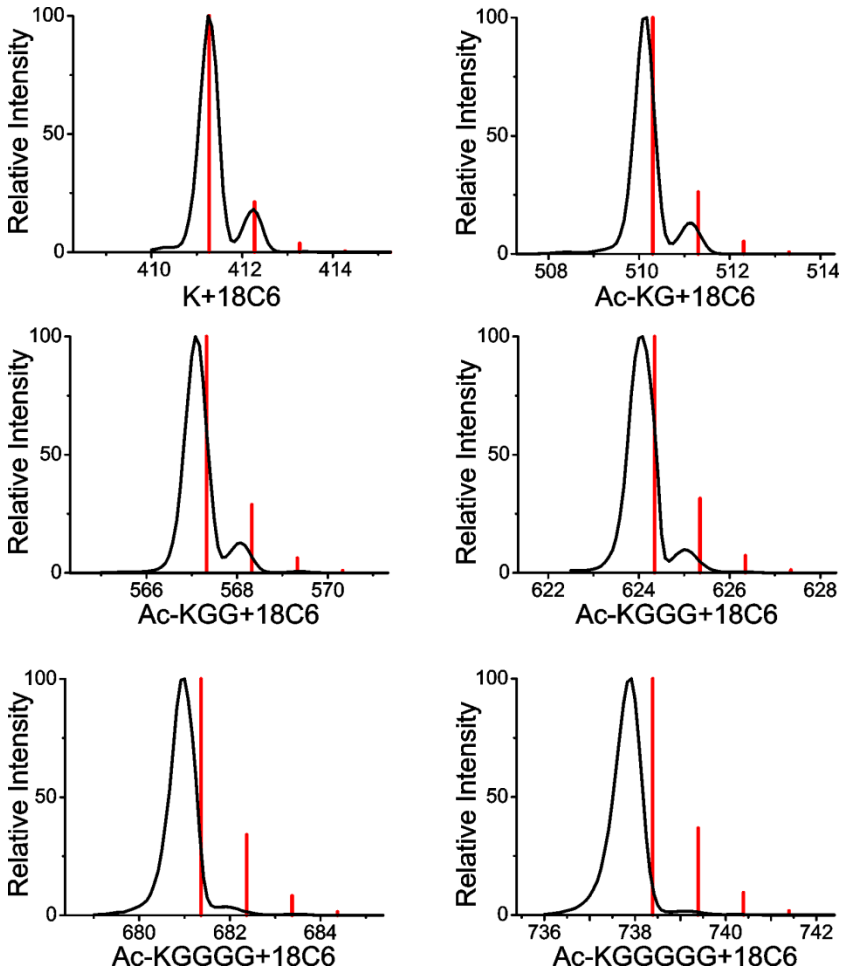


Figure 4.2 Isolation windows for a series of Ac-KG$_x$ (x = 0 to 5) peptides that have been complexed with 18C6. The predicted masses and isotope distributions are shown as red lines. The mass shifts increase as more glycine residues are added to the peptide.

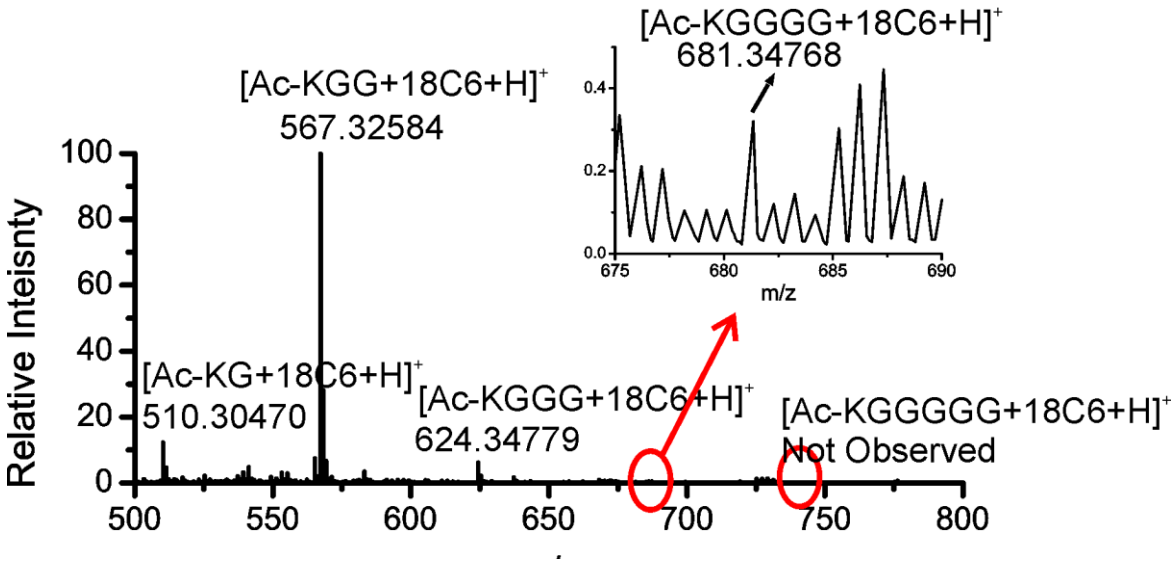


Figure 4.3 TOF mass spectrum of the mixture of 18C6 and Ac-KG_x (x = 1 to 5) peptides. The masses and isotope distributions agree well with predicted values for the 18C6 complex with Ac-KG_x (x = 1 to 3). The 18C6 complex with Ac-KGGGG has very low relative abundance while the Ac-KGGGGG-crown complex cannot be observed.

A similar set of experiments were conducted with Ac-RG_x (x = 0 to 5) and G_x (x = 2 to 5) peptides and the results are shown in Figure 4.5 and Figure 4.6. Arginine is the most basic residue and is therefore also a potential target for complexation by 18C6 at the protonated guanidinyl side chain. Previous results have suggested that arginine complexation with 18C6 is less favorable than with lysine.[8,17] The results in Figure 4.5 are consistent with this finding. A significant mass deviation is observed even for the complex with acetylated arginine itself. Overall, the trend is similar to that observed for the Ac-KG_x peptides, although the degree of mass shifting and peak broadening is greater for arginine than it is for lysine. Given that the inherent binding energy of arginine to 18C6 is weaker than it is for lysine, it is logical that competitive intramolecular binding is able to out-compete retention of the 18C6 adduct more easily.

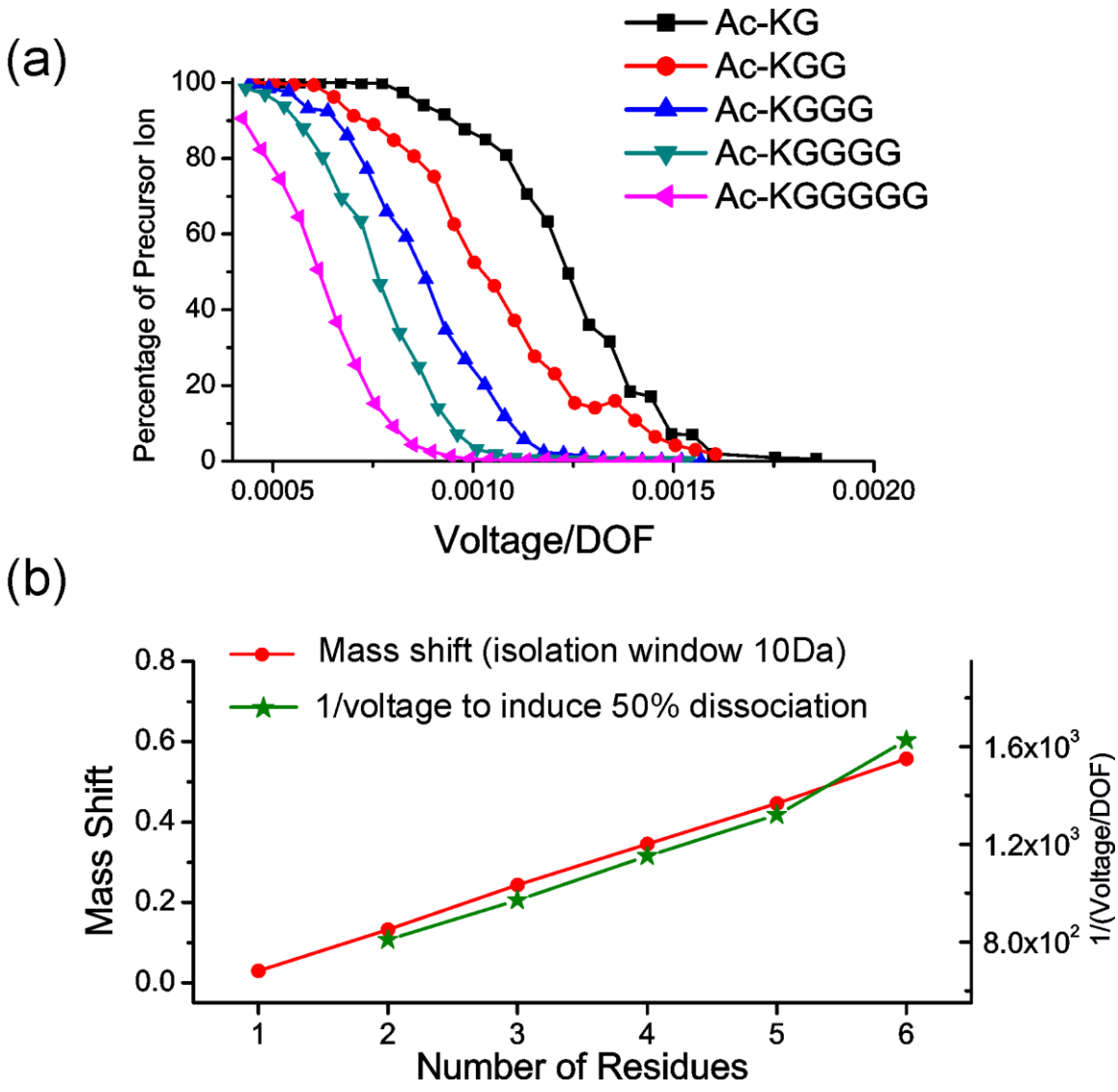


Figure 4.4 (a) Precursor ion survival as a function of excitation voltage/degrees of freedom. Again additional glycine residues lead to decreased stability. (b) The trends in peptide complex stability determined independently by mass shifting and collisional activation for Ac-KGx peptides. The y axis on the right represents the reciprocal of the activation voltage (normalized by the number of degrees of freedom) to induce 50% dissociation of the crown complex by CID.

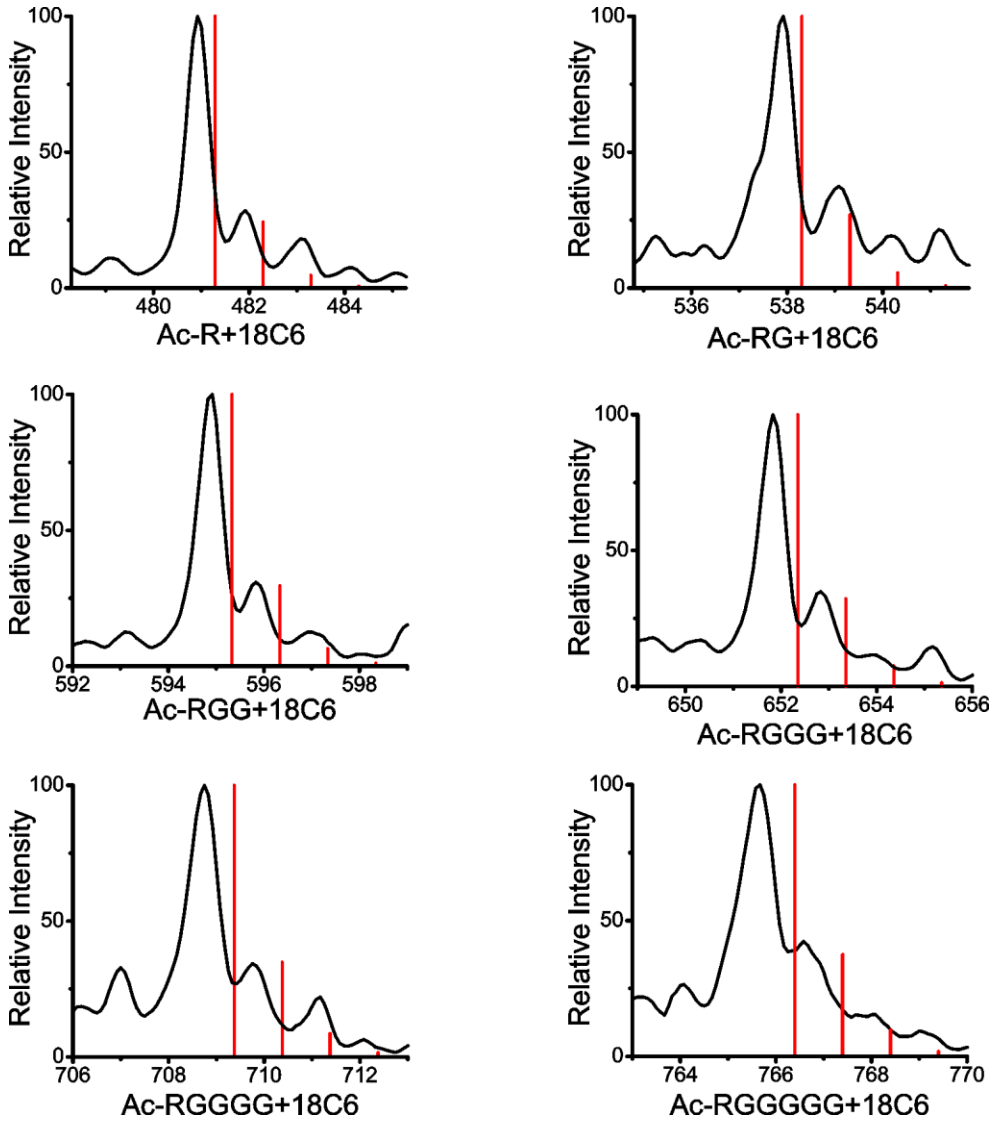


Figure 4.5 Isolation windows (10Da) for a series of Ac-RG_x (x = 0 to 5) peptides that have been complexed with 18C6. The predicted masses and isotope distributions are shown as red lines. The mass shifts increase as more glycine residues are added to the peptide.

For polyglycine, 18C6 attachment most likely occurs at the protonated N-terminus. Interestingly, very minimal mass shifts or peak broadening (Figure 4.6) are observed for most of the polyglycine peptides. Even for GGGGG, only minimal mass shifting is

noticed. The fundamental binding energy for 18C6 to the N-terminus (223 kJ/mol)[17] is comparable to that for the side chain of lysine (based on n-butyl amine, 223 kJ/mol),[18] meaning that there must be another explanation for the enhanced stability. Although the polyglycine backbone is fairly flexible, it is not sufficiently long to effectively solvate a charge at the protonated N-terminus via multidentate interactions as are present with 18C6. It appears that these steric constraints inhibit competitive complexation of the charge by the peptide backbone which leads to enhanced retention of the 18C6 adduct. It has been demonstrated previously that changes in the polarizability of small molecules may influence crown binding strengths,[17,18] though we anticipate such effects to be secondary to hydrogen bonding capacity for these larger systems.

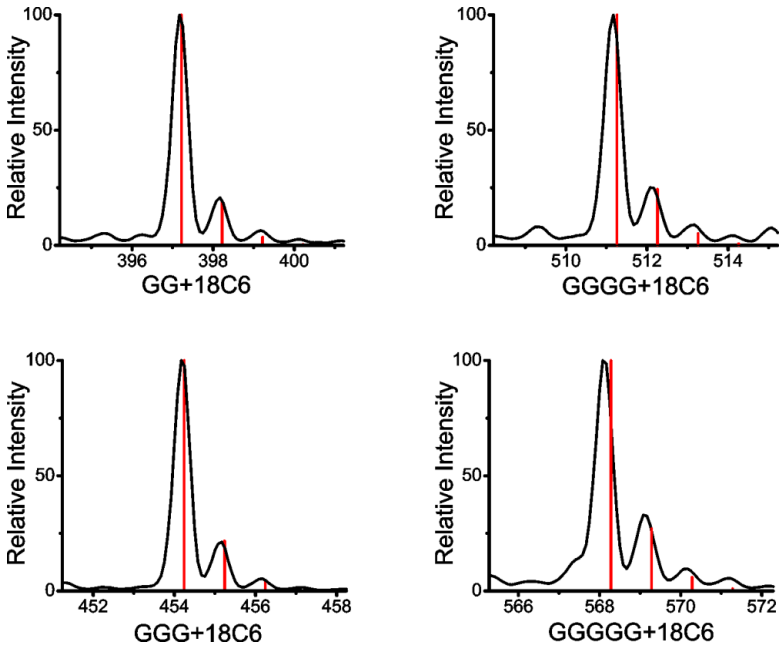


Figure 4.6 Isolation windows (10 Da) for a series of G_x (x = 2 to 5) peptides that have been complexed with 18C6. The predicted masses and isotope distributions are shown as red lines. The mass shifts and peak widths are reduced compared to Ac-RG_x and Ac-KG_x peptides.

The results from Figure 4.2, Figure 4.5 and Figure 4.6 are summarized for two different scan speeds in Figure 4.7. The lower scan speed will lead to excitation of the ions closer to the true m/z, and is observed to reduce the mass shifting significantly though the same trends are observed in both Figures 4.7a and 4.7b. At the slower scan speed in Figure 4.7b, the mass shifts are negligible for the polyglycines.

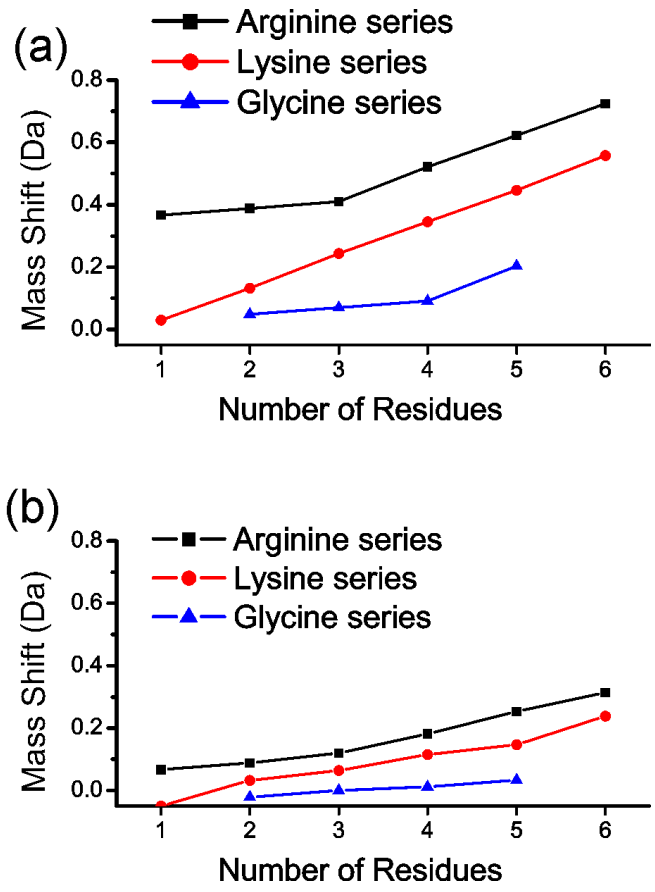


Figure 4.7 (a) Mass shifts for the primary isotopic peaks obtained from isolation windows of 18C6 adducts with Ac-RG$_x$, Ac-KG$_x$, and G$_x$ peptides. Reduced intramolecular charge solvation in polyglycine leads to greater 18C6 adduct stability and smaller mass shifts. (b) The magnitude of the mass shifting is reduced if the scan speed of the instrument is decreased.

4.3.3 Molecular Dynamics Simulations

Molecular dynamics calculations were performed to explore the structures and energetics of 18C6 complexes with Ac-KGGG and GGGG. The lowest energy structures that we were able to obtain for each peptide alone and complexed with 18C6 are shown in Figure 4.8. When attached to 18C6, the optimal structures have the protonated amines in both peptides forming 3 hydrogen bonds with 18C6, indicating a strong interaction between the charged site and the crown. However, solvation of the charged group in the absence of 18C6 is much less comparable between the two peptides. For Ac-KGGG, the peptide backbone and flexible side chain are able to rearrange in such a way that three relatively strong hydrogen bonds again stabilize the charged sited. For GGGG, only a single strong hydrogen bond is formed with an additional hydrogen bond that is weakened by larger separation and an unfavorable OHN interaction angle. The calculated binding energies for the two complexes are -176 kJ/mol for Ac-KGGG and -221 kJ/mol for GGGG. It is unlikely that molecular dynamics can be relied upon to accurately quantify these binding energies; however, the magnitude of the difference suggests that the trend would likely hold up even with higher level calculations. The trend in binding energy obtained by the calculations is also in agreement with experiment.

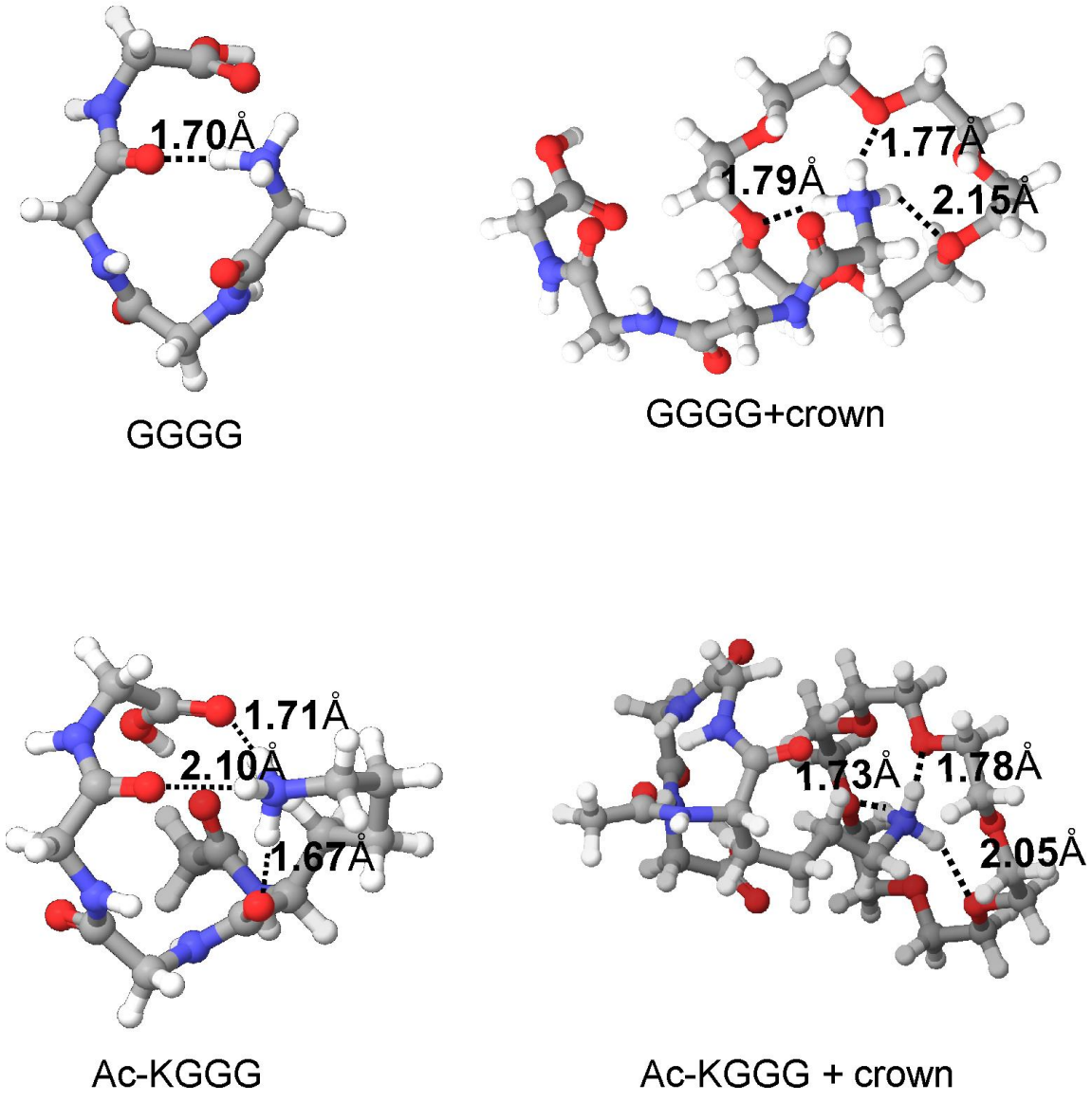


Figure 4.8 Lowest energy structures for GGGG, GGGG–18C6, Ac-KGGG, and Ac-KGGG–18C6 from molecular dynamics conformational searches. The hydrogen bond distances between H and O atoms are shown for each structure. The peptide structures change significantly when 18C6 is attached. The interactions between peptide backbone and side chain are stronger for Ac-KGGG than for GGGG.

4.3.4 Protein 18C6 Adduct Stability

The results in Figures 4.1-4.8 suggest that 18C6 adduct stability is a sensitive probe of local conformational flexibility and hydrogen bonding capacity. We next examined 18C6 adduct stability in proteins, which frequently attach multiple 18C6 adducts. In Figure 4.9 the stability of 18C6 adducts on several whole proteins (cytc, ubiquitin, and apomyoglobin) is examined as a function of charge state. In these experiments, a protein peak containing four 18C6 adducts was isolated and subjected to collisional activation. The amount of activation was kept constant between different charge states and proteins as defined by the percent activation parameter in the LTQ software. As a control, we examined the sensitivity of the observed distributions to small changes in activation energy and found very little dependence for small variations (±3%) in activation.

Results for cytc in the +9 through +15 charge states are shown in Figure 4.9a. Interestingly, for lower charge states the 18C6 adduct stability is very low. For example, collisional activation of the +9 charge state yields the bare protein as the primary product and higher order adducts are only retained in small abundance. At the other extreme, excitation of the +15 charge state favors loss of only a single 18C6, leading to formation of an abundant 3 adduct peak. There is essentially a smooth transition in behavior for all of the charge states between these two extrema. For ubiquitin, similar (though not identical) results are obtained. The lowest charge state exhibits the least 18C6 adduct stability and the highest charge states again retain the most 18C6 adducts; however, there is a significant discontinuity in the distributions between the +7 and +8 charge states. It is also worth noting that the distributions themselves are dissimilar between cytc and

ubiquitin, even if distributions at similar m/z are compared. The uniqueness of the distributions suggests that they are not the result of some simple statistical process. In Figure 4.9c, distributions obtained by examining apomyoglobin are shown. The general trends are again similar, but only two 18C6 adducts are retained in the highest charge states, compared with three for cytc and ubiquitin.

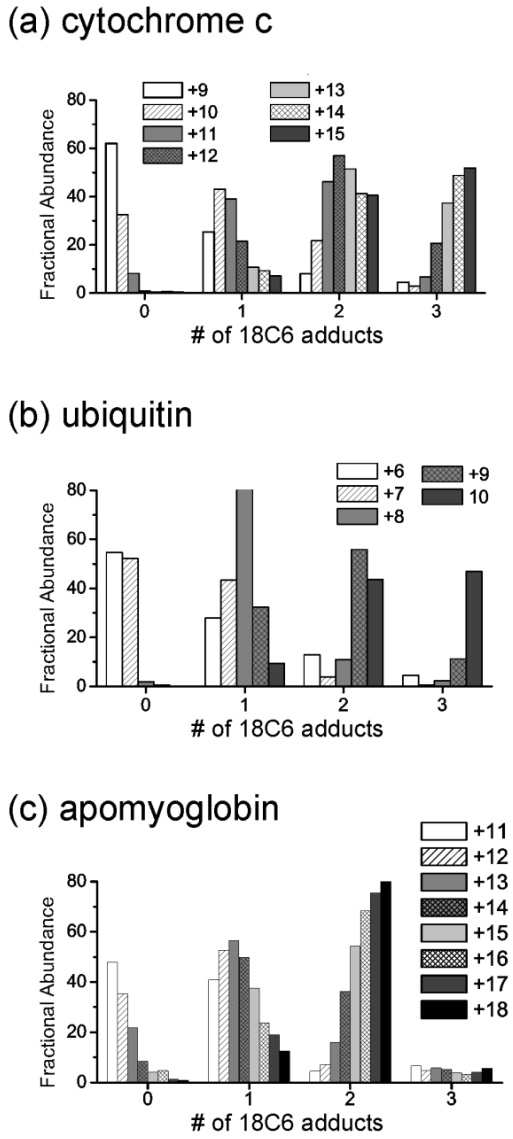


Figure 4.9 Distributions of remaining 18C6 adducts following collisional activation of the four adduct peak from different proteins in 50/50 water/methanol. (a) Cytochrome c from charge state +9 to +15. (b) Ubiquitin from charge state +6 to +10. (c) apomyoglobin from charge state +11 to +18. In general higher charge states retain more 18C6 adducts.

There are several conclusions that can be drawn from these results. First, 18C6 adduct stability is reduced for proteins in lower charge states, as is described in greater detail below in Figure 4.10 and 4.12. A second conclusion is that the degree of 18C6 retention

observed for each protein is unique. It is likely that the nature of the distributions generated upon loss of 18C6 adducts from various proteins is related to differences in sequence and three dimensional structure. For example, such structural features unique to each protein likely lead cytc to retain three 18C6 adducts in higher charge states under the same conditions where apomyoglobin retains only two.

In Figure 4.10, spectra obtained by collisional activation of four-adduct peaks at various excitation energies are shown for the +9 charge states of cytc. A single step of activation is sufficient to produce the bare protein. In fact, once sufficient energy is used to completely deplete the precursor ion, the bare protein is by far the most dominant product. This is a very interesting result that requires careful consideration. In an ion trap, activation is achieved by multiple, low-energy collisions. As a result, the lowest energy dissociation pathways are typically observed. Furthermore, excitation is resonant, meaning that once a single 18C6 adduct is lost no further energy will be pumped into the ion. Generally, this will lead to loss of a single 18C6 adduct from a multiple adduct complex following collisional activation because this represents the lowest energy dissociation pathway. For example, excitation of $[KKKKK+4(18C6)+4H]^{4+}$ yields dominant loss of a single 18C6 (Figure 4.11). Therefore, in order for all four 18C6 adducts to be lost simultaneously, the complex must acquire sufficient energy to lose all four adducts prior to the loss of a single 18C6, or the acquisition of energy must somehow result in significantly lower binding strength of the adducts to the protein. The second scenario is plausible if activation leads to significant structural perturbation of the protein, leading to weakening of 18C6 binding for all four adducts. In this case, the

rearrangement would presumably be from a kinetically trapped solution phase-like structure to the preferred gas phase structure. Alternatively, proteins are very large molecules and can potentially store sufficient energy to eventually lose four adducts prior to the loss of the first adduct. Given the rapid cooling that occurs in ion traps (as illustrated by difficulty in carrying out ion activation by infrared radiation),[26,27] we find this possibility to be less likely.

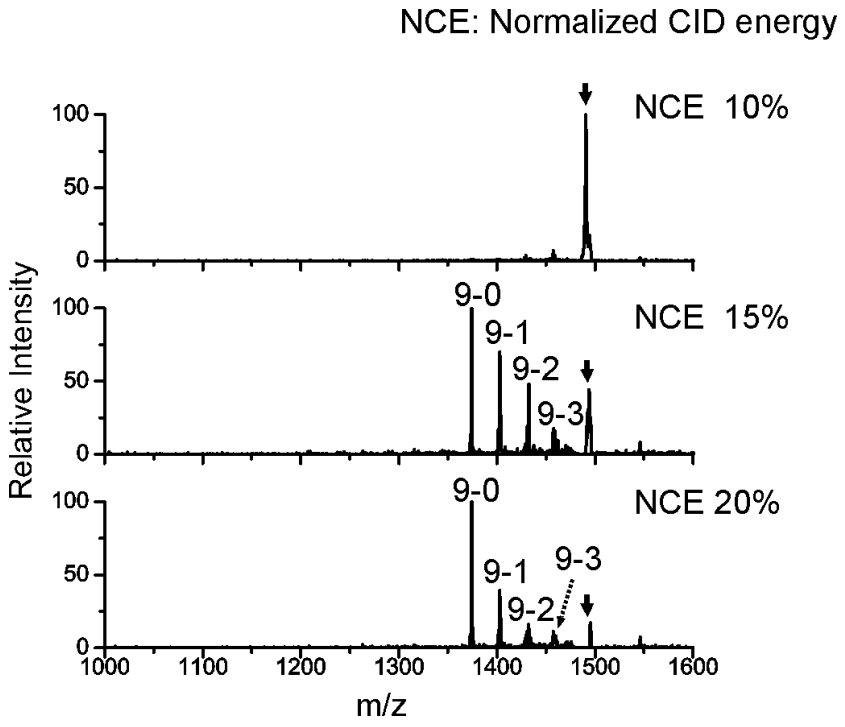


Figure 4.10 CID spectra of cytochrome c at +9 charge state with four-adduct peaks at various excitation energies. The black arrows represent precursor ions. Peaks are labeled by "charge state-number of crown adducts". 15% NCE is sufficient to produce the bare protein.

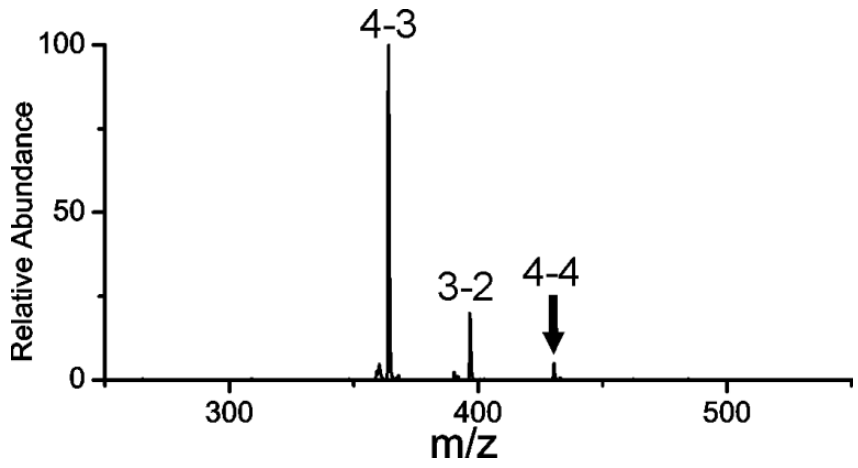


Figure 4.11 CID of $[KKKKK+4(18C6)+4H]^{4+}$. The black arrow represents the precursor ion. Peaks are labeled by "charge state - number of 18C6 adducts". The loss of one crown from +4 charge state is observed yields the primary product.

In contrast, for the +14 charge state activation leads primarily to the loss of one or two 18C6 adducts regardless of activation energy as shown in Figure 4.12. In this case, it is difficult to distinguish between a situation where modest structural rearrangement could occur and facilitate loss of two 18C6 adducts or the case where a large protein might store sufficient energy to accommodate loss of two crowns. Nevertheless, activation of the three adduct or two adduct peaks (following loss of one or two crowns) does not lead to preferential formation of the naked protein (Figure 4.13), suggesting that significant structural rearrangement is unlikely and that a different mechanism is responsible for 18C6 adduct loss in the higher charge state. Since higher charges states consist of elongated structures[28] that likely bear little resemblance to solution phase structures, it is less likely that these structures would be kinetically trapped and prone to undergo significant structural reorganization following collisional activation. This may

account for the observation that 18C6 adducts are more stable on higher charge state proteins. Furthermore, as additional charges are added to a protein, the capacity of the protein for self-solvation will decrease, which should also lead to increased 18C6 adduct stability.

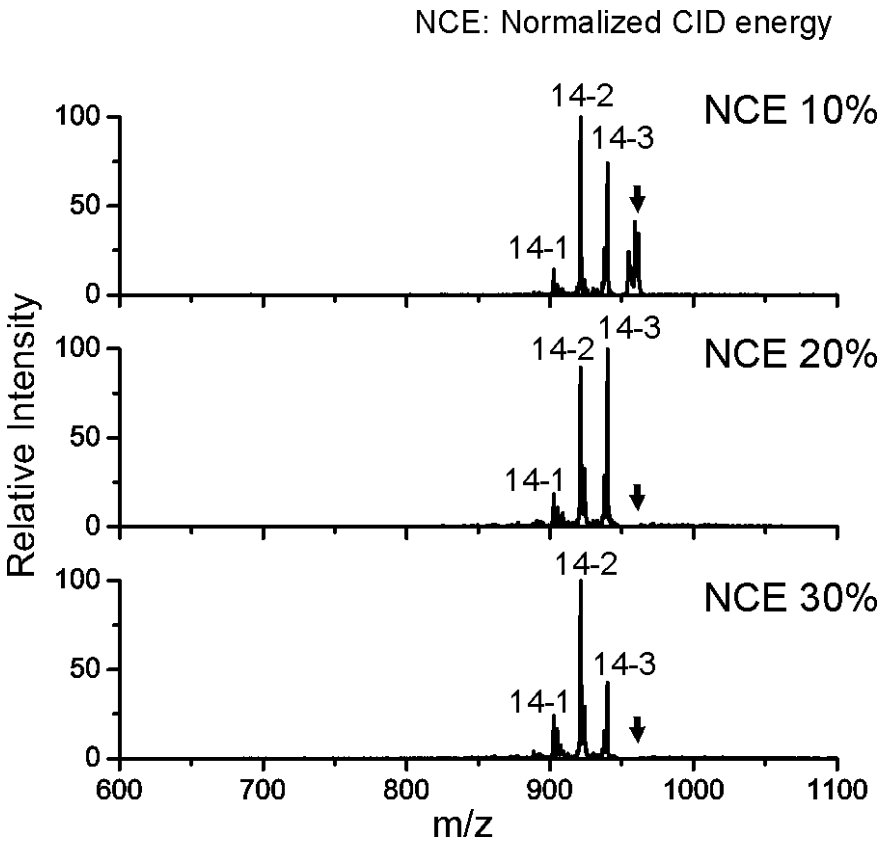


Figure 4.12 CID spectra of cytochrome c at +14 charge state with four-adduct peaks at various excitation energies. The black arrows represent precursor ions. Peaks are labeled by "charge state-number of crown adducts". The activation of +14 charge state mainly leads to the loss of two crowns even at higher activation energies.

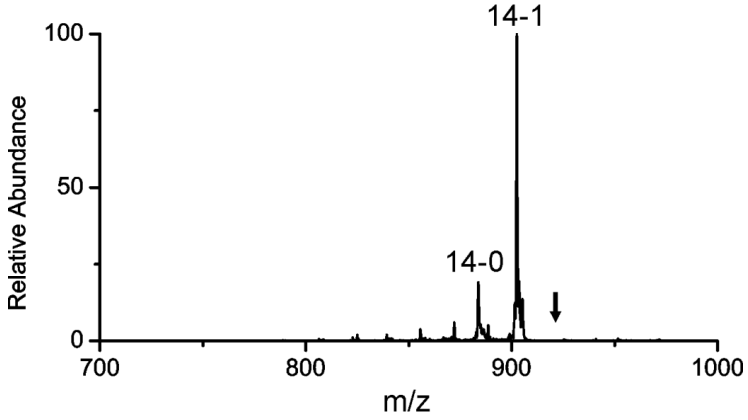


Figure 4.13 MS^3 spectrum showing collisional activation of the 14-2 peak of cytochrome c. Naked protein is not the preferred product. Peaks are labeled by "charge state - number of 18C6 adducts". The precursor ion was generated by collisional activation of the 14-4 peak.

An interesting possibility arises from consideration of these results, namely, the degree of loss of 18C6 adducts may reflect the degree to which a protein has undergone structural reorganization to accommodate the gas phase environment. Greater loss of 18C6 would then presumably indicate less reorganization occurred during transit into the gas phase and therefore greater similarity with solution phase structure (at least prior to loss of the crowns). One possible benefit that could occur from addition of 18C6 is that it may serve as solvent replacement during the transition of the protein from solution into the gas phase, allowing the protein to retain a greater degree of structural resemblance to the solution phase structure. If so, then 18C6 adduct formation may represent a method for essentially preserving solution phase structures via pseudo-solvation. Indeed recent work examining collision cross sections from ion mobility experiments concluded that 18C6 can micro-solvate charge sites and help to preserve solution phase structure for certain charge states.[29]

The possibility of native structure retention is examined further in Figure 4.14 where the results from collisional activation of holomyoglobin in the +10 and +11 charge states with four 18C6 adducts are shown. Holomyoglobin has a noncovalently attached heme group that is known to be labile and easily lost in MS experiments.[30] Interestingly, the results in Figure 4.14 show that heme loss generally occurs only after 18C6 is lost. For example in Figure 4.14a, the 10-0 (where 10 indicates charge state and 0 indicates the number of 18C6 adducts) apo peak is not accompanied by any 10-1 or higher order 18C6 adduct peaks, indicating that it originated only from the 10-0 holo peak. Similarly, in Figure 4.14b, the 10-0 apo peak is not accompanied by any 18C6 adducts. The 11-0 apo peak does have an accompanying 11-1 adduct peak, but the intensity of this peak is significantly lower than the 11-1 holo peak. All of these results are consistent with preferential loss of 18C6 over loss of heme (the key indicator for loss of native-like structure). This is further support that 18C6 can behave in a protective fashion, solvating charged side chains and preserving native structures. For both the +10 and +11 data, loss of all four adducts is observed in a single activation step, which is also consistent with our analysis of the data in Figure 4.10 and Figure 4.12 and with significant structural reorganization and possible retention of solution`n phase structure prior to activation.

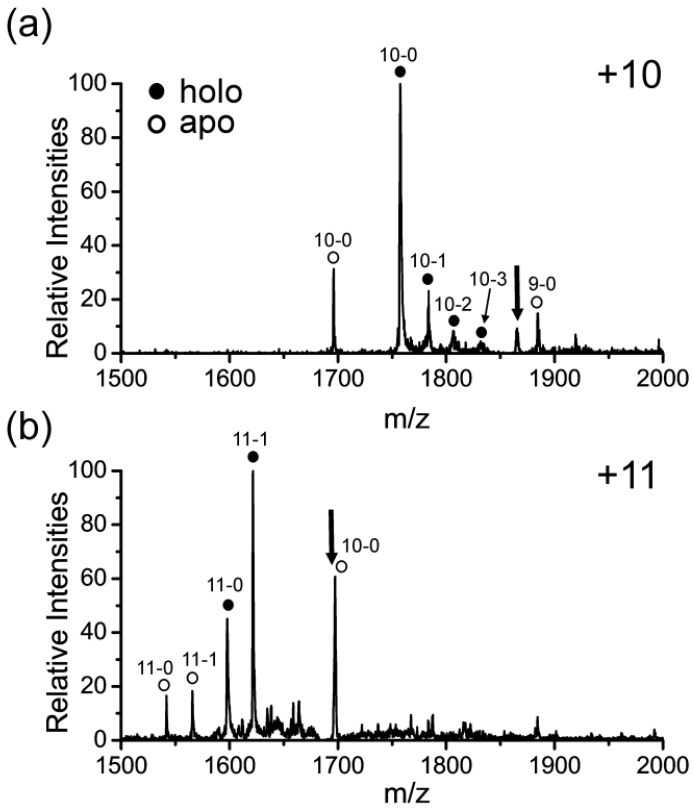


Figure 4.14 CID of holomyoglobin with four crown adducts for the (a) +10 and (b) +11 charge states. Heme loss occurs primarily after 18C6 is lost, suggesting a protective function. Numbers refer to charge state and number of 18C6 adducts, i.e. 11-1 is the +11 charge state with a single 18C6 adduct. The black arrows represent the precursor ions.

4.4 Conclusions

Although the fundamental interaction between 18C6 and protonated amines in the gas phase is quite strong, it is clear that in complex molecules this interaction can be significantly weakened by competitive intramolecular charge solvation. The structures of small model peptides which have been examined in detail undergo significant rearrangement following loss of 18C6 in the gas phase. This observation is logical, given

that hydrogen bonds with charged groups are among the strongest noncovalent forces present in the gas phase. Examination of proteins reveals that 18C6 adducts on lower charge states are weakly bound and easily lost upon collisional activation. In contrast, higher charge states exhibit greater 18C6 adduct retention. These results are consistent with the idea that proteins in lower charge states have greater resemblance to solution phase structures and therefore undergo more structural rearrangement that facilitates loss of 18C6. If this is the case, then 18C6 may serve the function of solvating side chains and protecting solution phase structure. Experiments with holomyoglobin support this idea because the labile heme group is observed to be lost primarily after 18C6 adducts. These results suggest that the potential of 18C6 as a solvent replacement should be investigated further.

References

[1] Gokel, G. W.; Leevy, W. M.; Weber, M. E., Crown ethers: Sensors for ions and molecular scaffolds for materials and biological models. *Chem. Rev.* **2004,** *104*, 2723-2750.

[2] Landini, D.; Montanar.F; Pirisi, F. M., Crown ethers as phase-transfer catalysts in 2-phase reactions. *J. Chem. Soc., Chem. Commun.* **1974,** *21*, 879-880.

[3] De Silva, A. P.; Desilva, S. A., Fluorescent signaling crown ethers-switching on of fluorescence by alkali-metal ion recognition and binding insitu. *J. Chem. Soc., Chem. Commun.* **1986,** *23*, 1709-1710.

[4] Kuhn, R.; Stoecklin, F.; Erni, F., Chiral separations by host-guest complexation with cyclodextrin and crown-ether in capillary zone electrophoresis. *Chromatographia* **1992,** *33*, 32-36.

[5] Maleknia, S.; Brodbelt, J., Gas-phase selectivities of crown ethers for alkali-metal ion complexation. *J. Am. Chem. Soc.* **1992,** *114*, 4295-4298.

[6] Chu, I. H.; Zhang, H.; Dearden, D. V., Macrocyclic chemistry in the gas-phase – intrinsic cation affinities and complexation rates for alkali-metal cation complexes of crown-ethers and glymes. *J. Am. Chem. Soc.* **1993,** *115*, 5736-5744.

[7] More, M. B.; Ray, D.; Armentrout, P. B., Intrinsic affinities of alkali cations for 15-crown-5 and 18-crown-6: Bond dissociation energies of gas-phase M+-crown ether complexes *J. Am. Chem. Soc.* **1999,** *121*, 417-423.

[8] Julian, R. R.; Beauchamp, J. L., Site specific sequestering and stabilization of charge in peptides by supramolecular adduct formation with 18-crown-6 ether by way of electrospray ionization *Int. J. Mass Spectrom.* **2001,** *210*, 613-623.

[9] Lee, S. W.; Lee, H. N.; Kim, H. S.; Beauchamp, J. L., Selective binding of crown ethers to protonated peptides can be used to probe mechanisms of H/D exchange and collision-induced dissociation reactions in the gas phase. *J. Am. Chem. Soc.* **1998,** *120*, 5800-5805.

[10] Ly, T.; Julian, R. R., Using ESI-MS to probe protein structure by site-specific noncovalent attachment of 18-crown-6. *J. Am. Soc. Mass Spectrom.* **2006,** *17*, 1209-1215.

[11] Sun, Q. Y.; Nelson, H.; Ly, T.; Stoltz, B. M.; Julian, R. R., Side Chain Chemistry Mediates Backbone Fragmentation in Hydrogen Deficient Peptide Radicals. *J. Proteome Res.* **2009,** *8*, 958-966.

[12] Bohrer, B. C.; Clemmer, D. E., Shift Reagents for Multidimensional Ion Mobility Spectrometry-Mass Spectrometry Analysis of Complex Peptide Mixtures: Evaluation of 18-Crown-6 Ether Complexes. *Anal. Chem.* **2011,** *83*, 5377-5385.

[13] Pagel, K.; Hyung, S.-J.; Ruotolo, B. T.; Robinson, C. V., Alternate Dissociation Pathways Identified in Charge-Reduced Protein Complex Ions. *Anal. Chem.* **2010,** *82*, 5363-5372.

[14] Kupser, P.; Pagel, K.; Oomens, J.; Polfer, N.; Koksch, B.; Meijer, G.; von Helden, G., Amide-I and -II Vibrations of the Cyclic beta-Sheet Model Peptide Gramicidin S in the Gas Phase. *J. Am. Chem. Soc.* **2010,** *132*, 2085-2093.

[15] Buschmann, H. J.; Schollmeyer, E.; Mutihac, L., The complexation of the ammonium ion by 18-crown-6 in different solvents and by noncyclic ligands, crown ethers and cryptands in methanol. *Supramol. Sci.* **1998,** *5*, 139-142.

[16] Hamdy, O. M.; Julian, R. R., Reflections on Charge State Distributions, Protein Structure, and the Mystical Mechanism of Electrospray Ionization. *J. Am. Soc. Mass Spectrom.* **2012,** *23*, 1-6.

[17] Chen, Y.; Rodgers, M. T., Structural and Energetic Effects in the Molecular Recognition of Amino Acids by 18-Crown-6. *J. Am. Chem. Soc.* **2012,** *134*, 5863-5875.

[18] Chen, Y.; Rodgers, M. T., Structural and Energetic Effects in the Molecular Recognition of Protonated Peptidomimetic Bases by 18-Crown-6. *J. Am. Chem. Soc.* **2012,** *134*, 2313-2324.

[19] David, W. M.; Brodbelt, J. S., Threshold dissociation energies of protonated amine/polyether complexes in a quadrupole ion trap *J. Am. Soc. Mass Spectrom.* **2003,** *14*, 383-392.

[20] Chan, W.C.; White, P.D. *Fmoc Solid Phase Peptide Synthesis*, Oxford University Press: New York, 2000.

[21] Ly, T.; Julian, R. R. Residue-specific radical-directed dissociation of whole proteins in the gas phase. *J. Am. Chem. Soc.* **2008,** *130*, 351-358.

[22] McClellan, J. E.; Murphy, J. P.; Mulholland, J. J.; Yost, R. A., Effects of fragile ions on mass resolution and on isolation for tandem mass spectrometry in the quadrupole ion trap mass spectrometer. *Anal. Chem.* **2002,** *74*, 402-412.

[23] Murphy, J. P.; Yost, R. A., Origin of mass shifts in the quadrupole ion trap: dissociation of fragile ions observed with a hybrid ion trap/mass filter instrument. *Rapid Commun. Mass Spectrom.* **2000,** *14*, 270-273.

[24] Cowan, D. A.; Kiman, A. T.; Kubli-Garfias, C.; Welchman, H. J., Ion trap MS/MS of intact testosterone and epitestosterone conjugates - Adducts, fragile ions and the advantages of derivatisation. *Steroids* **2008,** *73*, 621-628.

[25] Huffman, C. L.; Williams, M. L.; Benoist, D. M.; Overstreet, R. E.; Jellen-McCullough, E. E., Dependence of collision-induced dissociation energy on molecular degrees of freedom as a means to assess relative binding affinity in multivalent complexes. *Rapid Commun. Mass Spectrom.* **2011,** *25*, 2299-2306.

[26] Brodbelt, J. S., Shedding Light on the Frontier of Photodissociation. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 197-206.

[27] Newsome, G. A.; Glish, G. L., Improving IRMPD in a quadrupole ion trap. *J. Am. Soc. Mass Spectrom.* **2009,** *20*, 1127-1131.

[28] Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. R., Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* **2009,** *39*, 1633-1655.

[29] Warnke, S.; von Helden, G.; Pagel, K., Protein structure in the gas phase: the influence of side-chain microsolvation. *J. Am. Chem. Soc.* **2013,** *135*, 1177-1180.

[30] Babu, K. R.; Douglas, D. J., Methanol-induced conformations of myoglobin at pH 4.0. *Biochemistry*, **2000,** *39*, 14702-14710.

Chapter 5

EXAMINING PROTEIN SURFACE STRUCTURE IN HIGHLY CONSERVED SEQUENCE VARIANTS WITH MASS SPECTROMETRY

5.1 Introduction

Sequence variation in proteins occurs frequently in nature and has significant importance in various biological contexts. Certain proteins have similar amino acid sequences but exhibit very different behaviors which are related to structure. For example, the three proteins from the synuclein family (α-, β-, γ-synuclein) share a high level of sequence homology, but only α -synuclein is linked with fibril formation and the pathology of Parkinson's disease.[1] Furthermore, single point mutations in the α -synuclein gene are known to greatly increase the rate of protein aggregation.[2-4] Sickle cell anemia is another well-known disease caused by a point mutation.[5] The glutamic acid residue in the sixth position of the β chain of normal hemoglobin is replaced by valine in sickle cell hemoglobin.[6] This mutation shifts the hydrophobicity of the protein surface and dramatically reduces the solubility of the deoxygenated hemoglobin in the blood, which is ultimately responsible for the disease. It is clear from these examples that small changes in the primary sequence can dramatically alter protein structure and function. It is also well known that not all amino acid substitutions lead to significant structural perturbations in terms of three-dimensional backbone structure. For example, the native fold of ubiquitin can accommodate a large number of amino acid substitutions without significant perturbation.[7,8] Point mutations are also regularly introduced into proteins

intentionally by researchers for a variety of reasons,[9] frequently with the implicit hope that the structure and behavior of the protein will not be affected.

It is possible to examine the effects of point mutations with traditional protein structure determination methods such as x-ray crystallography[10] and NMR.[11] In some cases this may be warranted, but frequently the time and sample consumption required for these methods will preclude their use. In contrast, mass spectrometry (MS) is well suited to examine proteins both quickly and with excellent sensitivity. Although the three dimensional structures obtained by x-ray and NMR cannot be derived from MS-based experiments, there are aspects of protein structure which are most easily examined by MS. The development of soft ionization methods, which has enabled protein analysis by MS, also led to the emergence of methods that provide information about protein structure. It was recognized early on that the mere process of electrospraying a protein reveals some information about structure. For example, the charge state distribution observed for a protein that has been electrosprayed can be used to coarsely determine folding state. [12 - 14] Subsequently, more sophisticated experiments utilizing hydrogen/deuterium exchange, irreversible covalent labeling, or crosslinking have been used to reveal more detailed information about protein conformation.[15-17] These methods are the preferred manner to probe solvent accessibility and protein-protein interactions. Another MS based technique known as SNAPP (selective noncovalent adduct protein probing) has been developed to examine protein solution phase structure with noncovalent probes. [18-20] In this method, 18-crown-6 ether (18C6) is used as a recognition molecule which can noncovalently attach to basic sites (Lys/Arg side chains, N-terminus)

in proteins. Regardless of whether a particular site is available for binding by 18C6 is determined by the presence and abundance of competitive intramolecular interactions, including salt bridges and hydrogen bonds. Salt bridges with acidic groups are most effective at interfering with 18C6 binding;[19] therefore, SNAPP is sensitive to the arrangement of basic, acidic, and polar groups at the protein surface (which is ultimately a function of the overall three-dimensional protein structure). 18C6 adducts do not simply count the number of basic sites. Importantly, because of the relative solution and gas phase binding properties, 18C6 does not interact significantly with proteins until after droplet formation during ESI.[21,22] SNAPP is unique in that the focus is exclusively on side chain interactions at the protein surface and structure probing occurs in the transition between solution and the gas phase. The raw mass spectra provide a distribution of protein-18C6 complexes (SNAPP distribution) for each charge state. The shape and relative intensity of the distributions are very sensitive to protein surface structure. If the protein is modified by ligand binding, denaturation, or point mutations which modulate the chemical environment around 18C6 binding sites, shifts in the SNAPP distributions will be observed. Therefore, SNAPP is primarily a comparison method which can determine if changes to a protein or its environment lead to structural shifts, as well. It is important to note that because SNAPP provides an electrostatic map of protein surface structure, changes which do not result in rearrangement of the tertiary fold can still yield different SNAPP distributions if the surface structure has been perturbed. There could be important consequences for this type of situation, e.g. the catalytic properties of two proteins with similar backbone structure might well be quite comparable, but protein-

protein recognition could be significantly different due to changes at the surface where recognition occurs.

Herein several orthologous proteins from multiple species that share a common function and varying degrees of sequence homology are examined by the SNAPP-MS method. In each case where structures are known, these proteins have been determined to have nearly identical native folds by x-ray crystallography; however, the effects of sequence variation on surface structures have not been previously experimentally probed. For three variants of insulin, sequence mutations are very minor and SNAPP distributions are all very similar, suggesting no disturbance to the electrostatic surface structure. Cytochrome c (cytc) variants from four species were examined; three have known tertiary structures. Interestingly, yeast yields a distinct SNAPP distribution relative to horse, bovine and pigeon cytc. Yeast cytc contains the largest number of mutations in acidic/basic residues and also exhibits reduced stability, leading to denaturation under milder conditions than those required for the other variants. Two forms of lysozyme with significant sequence mutations, but identical backbone structures, were found to yield SNAPP distributions that are quite different. The relative importance of various types of mutations on the observed SNAPP distributions and protein surface structures is discussed.

5.2 Experimental Methods

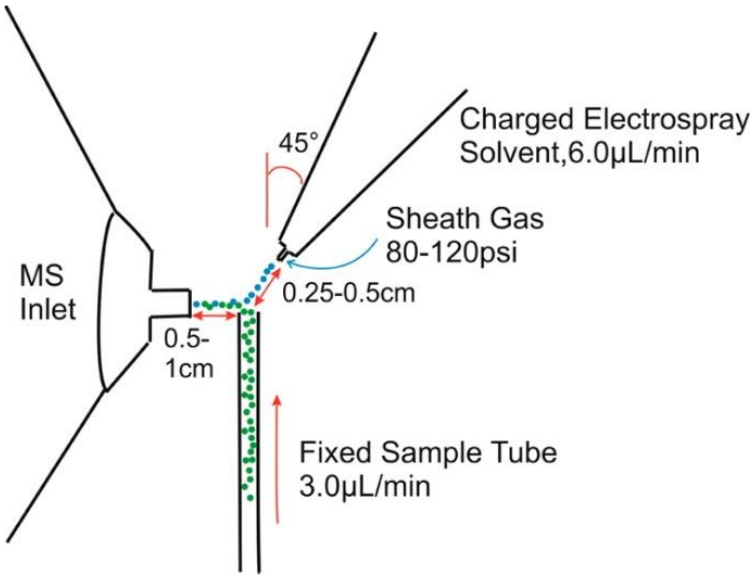
5.2.1 Protein Samples and Purification

Recombinant cytc iso-1 from Yeast was purchased from abcam (Cambridge, MA). All other proteins used in this work were purchased from Sigma-Aldrich (St. Louis, MO). Cytc from yeast and lysozyme from human was further purified by dialysis against water and lyophilized. Methanol (Sigma Aldrich, St. Louis, MO) and acidic acid (Mallinckrodt Baker Inc. Phillipsburg, NJ) were of analytical grade and used without further purification. All protein samples were prepared using a Millipore (Millipore, Billerica, MA) Direct-Q purified water without any acid or buffer, unless otherwise noted. The concentrations for all proteins were kept in the 7~10μM range. 18C6 (Alfa Aesar, Pelham, NH) with 10 times of protein concentration was added to the sample solution prior to electrospray. For example, for a final cytc concentration of 10 μM, the concentration of 18C6 would be 100 μM. All samples were of neutral pH as determined by litmus paper.

5.2.2Mass Spectrometry

Mass spectra were recorded with a Finnigan LCQ 3D ion trap mass spectrometer equipped with a modified liquid desorption electrospray ionization (DESI) source. SNAPP experiments are very sensitive to the exact source conditions that are employed, and therefore, the instrument was calibrated against a standard immediately prior to each experimental run to verify similar source conditions. For these experiments, cytc from horse in a 50/50 water/methanol mixture was used as the standard, which was verified to yield a reproducible spectrum. Although standard ESI can be used for SNAPP, we have

found that the liquid DESI arrangement provides for easier reproducibility. The complete details of these findings will be the subject of a future publication. The DESI source was implemented by removal of the original electrospray nozzle from the source mount. The nozzle was then oriented as shown in Scheme 5.1 with a ring stand and clamp. The gas to the nozzle was provided directly from a gas cylinder rather than being passed through the LCQ. The liquid sample was pumped from a fixed tube placed ~90° to the mass spectrometer inlet, with a distance of 0.5−1 cm as shown in Scheme 5.1. The sample was ionized through interaction with the charged solvent droplets generated by the electrospray nozzle. The sample flow rate was 3.0 μL/min, and the spray flow rate was 6.00 μL/min. The sheath gas pressure was increased to >80 psi. The solvent solution for DESI source was a 50/50 water/methanol mixture with 1% acetic acid. Typical settings were as follows: capillary voltage of 100 V, capillary temperature of 215 °C, tube lens offset of −65 V, and spray voltage of 5 kV. Once optimized, the instrument parameters remained unchanged for all experiments.



Scheme 5.1 Diagram of the Liquid DESI Source

5.2.3 Protein Structures

The following Protein Data Bank (PDB) entries for the crystal structures discussed here were used: porcine insulin, 2EFA; bovine insulin, 9INS; human insulin, 3I3Z; horse cytc, 1HRC; bovine cytc, 2B4Z; yeast cytc, 1YCC; human lysozyme, 1LZ1; hen lysozyme, 2LYZ. The surface electrostatic maps were generated from the Maestro computing program (Schrodinger, Inc., San Diego, CA) by choosing the "molecular surface" option, and the color is based on the residue charge (positive, blue; negative, red).

5.3 Results and Discussion

5.3.1 Insulin

Aligned backbone structures in ribbon form obtained from the crystal structures of human, porcine, and bovine insulin are shown in Figure 5.1a.[23-25] The three structures are virtually identical. Sequence alignment for the same set of proteins is shown in Figure 5.1b. Only one residue is different between human and porcine insulin, while bovine

insulin has two additional mutations (all highlighted in off-white). None of the mutations involve basic or acidic residues, which are highlighted in blue and red, respectively. Figure 5.1c shows the mass spectrum obtained from a solution containing bovine, porcine, human insulin and 18C6. The three proteins were examined simultaneously, which eliminates any potential effects from variations in source or solution conditions. Reassuringly, the results are almost identical to those obtained from each protein examined separately, when care is taken to ensure similar source conditions. Comparable distributions both in terms of shape and relative intensities are obtained for all three proteins. The highly aligned backbone structures and extremely similar sequences of these variants suggest that they should have virtually identical surface structures as well, which is reflected in the observed SNAPP distributions.

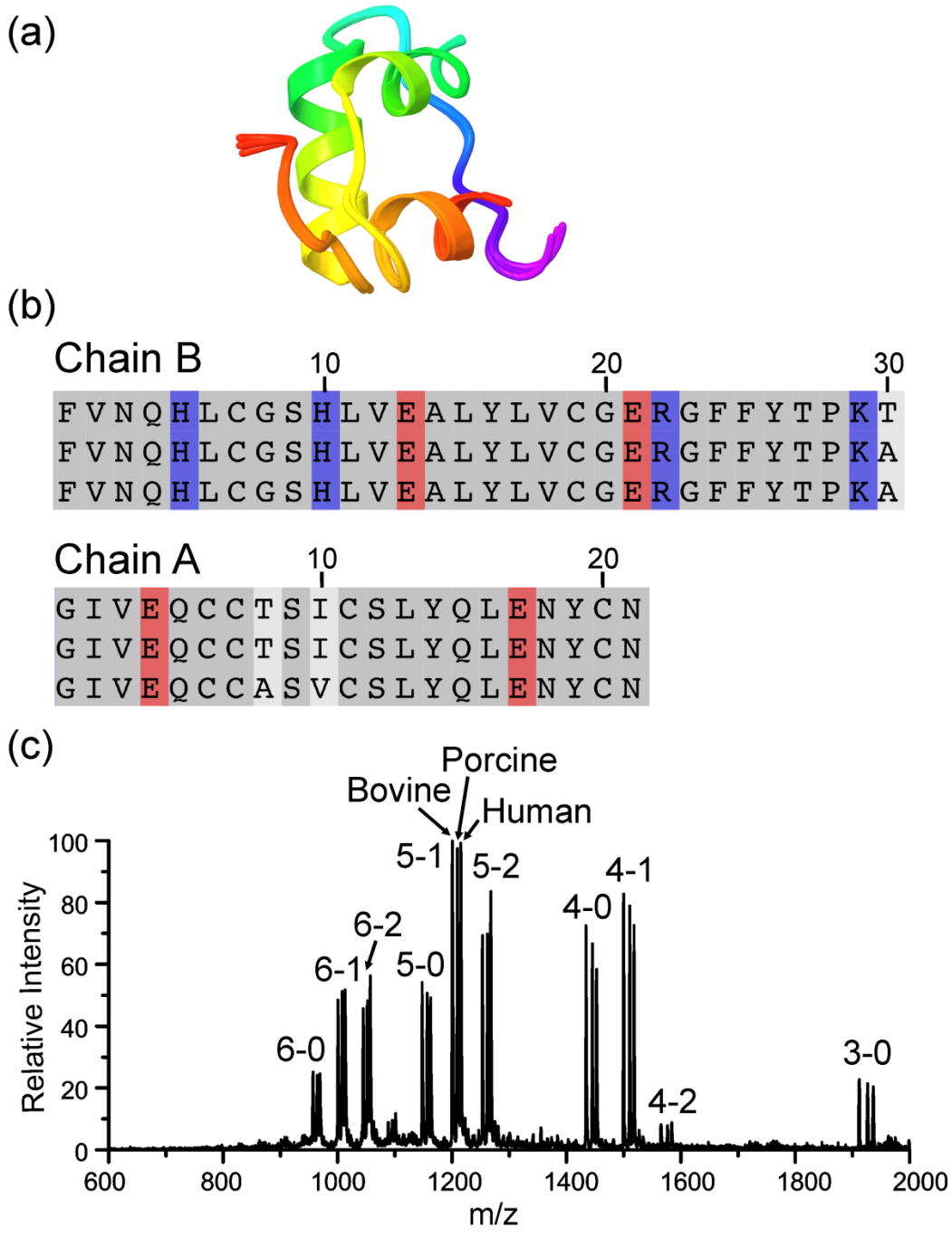


Figure 5.1 (a) Backbone structural alignment for the three variants of insulin. (b) Sequence alignment for human, porcine and bovine insulin from top to bottom, respectively. Sequence variation (white), basic residues (blue) and acidic residues (red) are highlighted. (c) ESI-MS spectrum of insulin from bovine, porcine, human and 18C6 in water. Peaks are labeled by (charge state) – (number of 18C6 adducts). The results are similar for all three proteins.

5.3.2 Cytochrome c (cytc)

Figure 5.2 shows the sequence alignment for horse, bovine, pigeon and yeast cytc. Several mutations are observed among the four proteins. Horse and bovine cytc have the most similar sequences with only one basic/acidic amino acid variation, K60G. Pigeon cytc is also very similar with a total of five basic/acidic amino acid variations. In contrast, yeast cytc has numerous basic/acidic amino acid variations. It also contains five additional N-terminal residues (including one lysine) and is missing one C-terminal residue. Overall, the yeast variant has one fewer basic and one fewer acidic residue than the horse variant. In Figure 5.3a, the backbone crystal structures for horse, bovine, and yeast cytc are shown from identical perspectives. Space filled atoms represent the side chains of basic and acidic residue mutations relative to horse cytc. The three proteins adopt very similar backbone structures, despite a fair number of point mutations. Electrostatic surface surfaces derived from the crystal structures are shown in Figure 5.3b from the same perspective shown in Figure 5.3a (red and blue areas are negatively and positively charged, respectively).

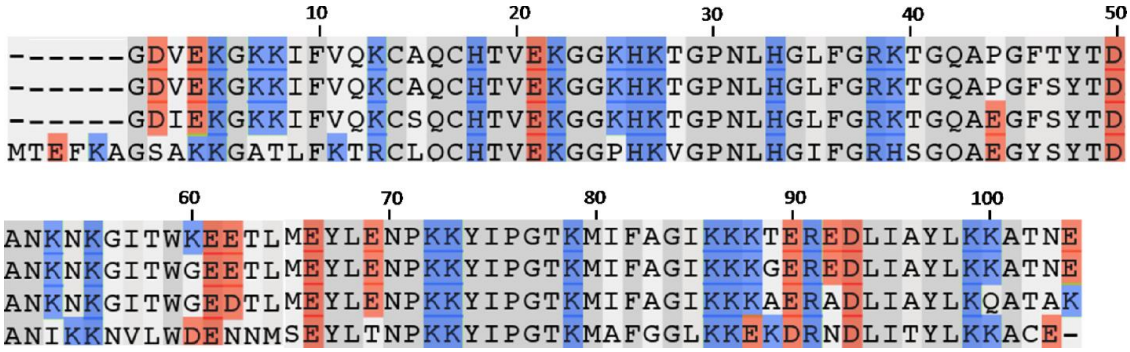


Figure 5.2 Sequence alignment for horse, bovine, pigeon, and yeast cytc from top to bottom, respectively.

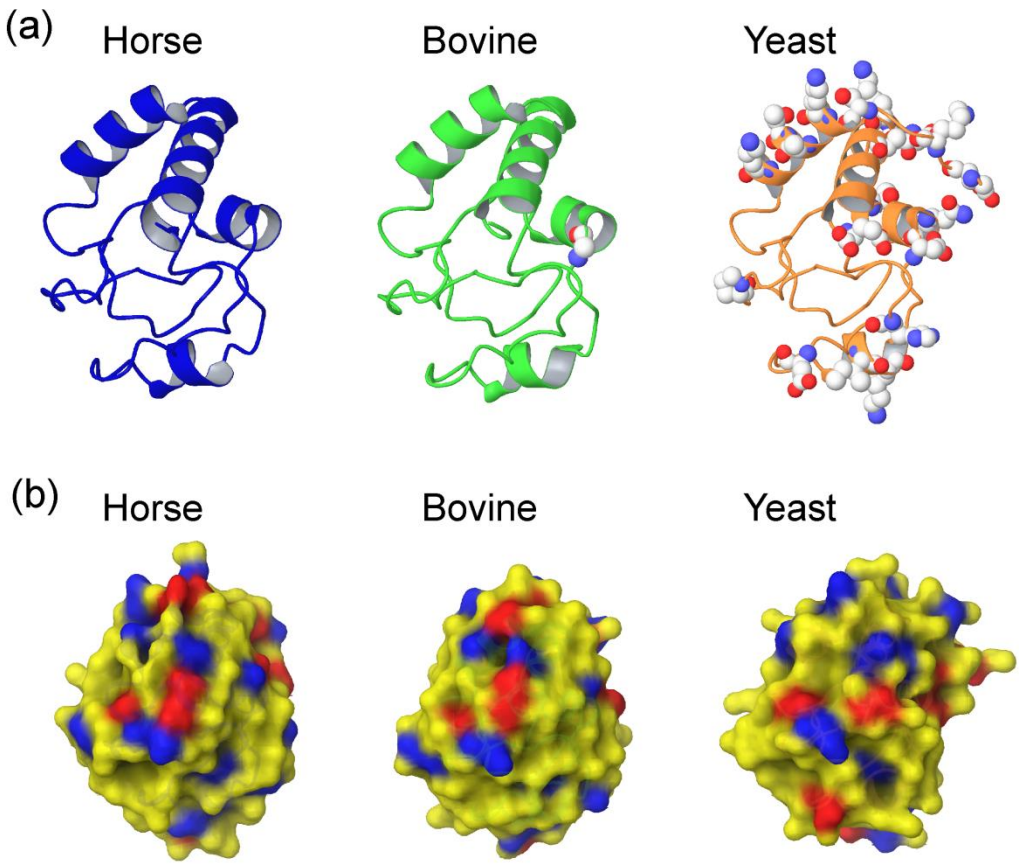


Figure 5.3 (a) Backbone structures of horse, bovine, and yeast cytc. Displayed atoms represent basic and acidic sequence variations relative to horse cytc. (b) Surface electrostatic distributions derived from crystal structures. The colors represent charge polarity (positive charge, blue; negative charge, red).

Figure 5.4 shows the extracted SNAPP distributions for the four variants of cytc for the +9 and +10 charge states. The intensities are normalized to the non-adduct protein peak for comparison, and the error bars represent the standard deviation of the mean. Although bovine cytc has one fewer lysine residue than the horse variant, they have nearly identical SNAPP distributions, suggesting that the additional lysine 60 in horse cytc does not contribute to the SNAPP distribution. Comparison of the two electrostatic surfaces shown

in Figure 5.3b does not reveal significant differences in the region where Lys60 is located, in agreement with the SNAPP results. Inspection of the NMR structure[26] for horse cytc also reveals a potential salt bridge interaction between Lys60 and Glu62, which would interfere with 18C6 attachment. These results confirm previous findings[19] indicating that SNAPP distributions do not simply count the number of basic residues present in a protein, but rather provide information about the surface accessibility of basic residues in relation to the surrounding chemical environment. The crystal structure of pigeon cytc has not been reported yet, but based on the SNAPP distributions, it is likely similar to the known horse and bovine structures. Although SNAPP only directly probes surface structures, it is unlikely that two proteins with high sequence homology and similar surface structures would be able to simultaneously adopt two highly dissimilar backbone structures. Therefore, for homologous proteins, similar SNAPP distributions will likely imply similar backbone structures, although dissimilar SNAPP distributions may not imply dissimilar backbone structures.

This is illustrated in the SNAPP distributions for yeast cytc, which are somewhat different from the remaining proteins despite the fact that the backbone structures are the same. In Figure 5.4, it is clear that more 18C6 attaches to yeast cytc than the remaining variants. It is tempting to suggest that the additional N-terminal lysine residue in yeast might account for the increased 18C6 attachment, but the comparison above between the horse and bovine variants reveals that one specific lysine residue does not necessarily affect the SNAPP distribution. In this case, the lysine residue in the additional n-terminal segment is also accompanied by an additional glutamic acid, which may mediate any

increased 18C6 attachment. Other potentially important mutations involving change of polarity (basic and acidic residues), such as K7A, K8T, V11K, K25P, P44E, K53I, N54K, K60D, E62N, E69T, K88E, T89K, E92N, N103E, are also probable to contributors to the observed shift. It is most likely that the sum of these mutations leads to perturbation of the overall surface structure and a small shift in the observed SNAPP distributions. Comparison of the electrostatic surfaces in Figure 5.3b also confirms that yeast, while similar, has distinct features relative to the other variants, again in agreement with the data obtained by SNAPP.

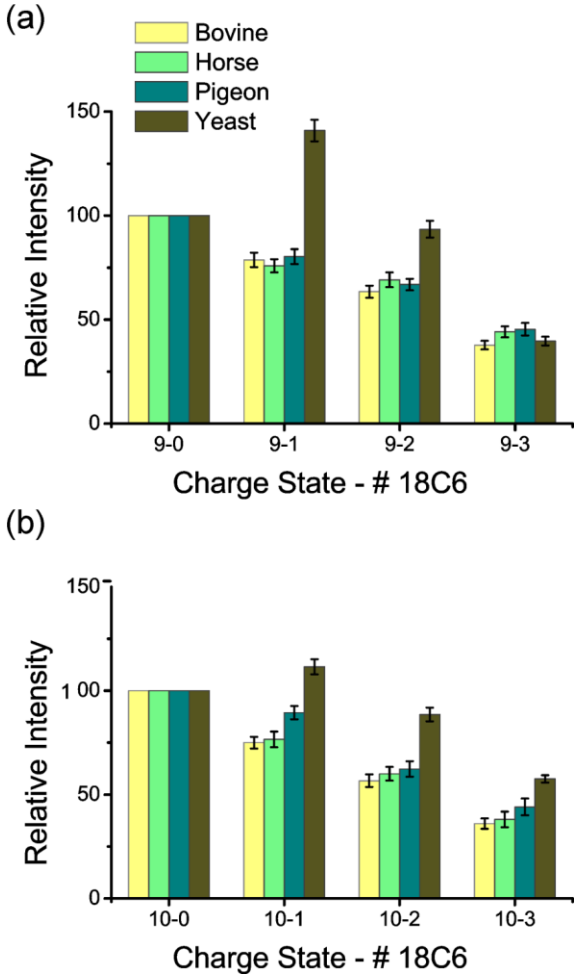


Figure 5.4 SNAPP distributions for four variants of cytc for the (a) +9 and (b) +10 charge states. Bovine, horse, and pigeon cytc exhibit similar results, while yeast cytc has a distinct distribution.

5.3.3 Lysozyme

The crystal structures of the hen and human variants of Lysozyme are well aligned as shown by the ribbon representations in Figure 5.5. However, the predicted charge distribution on the surface of these proteins is quite different, as shown in Figure 5.6. The origin of the dissimilarity can be seen in the sequence alignment for the two proteins as shown in Figure 5.7. The mutated acidic residues are highlighted with red stars and the mutated basic residues are denoted by blue asterisks. In contrast to cytc, where many

mutations correspond to minor shifting in location of the same residue, in lysozyme many mutations occur in regions that contain no charged residues in the other variant.

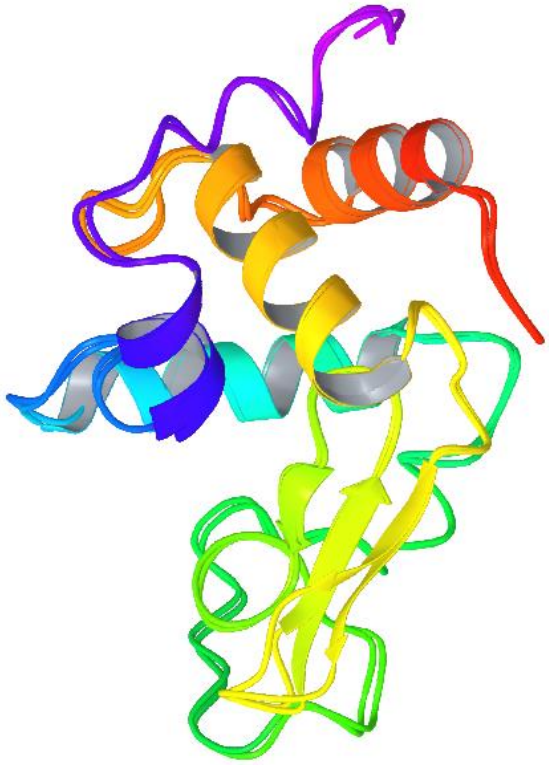


Figure 5.5 Backbone structural alignment of hen and human lysozyme.

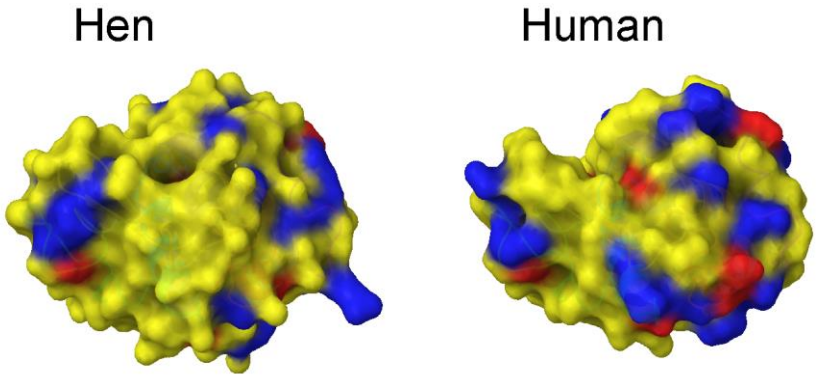


Figure 5.6 Surface electrostatic distributions derived from crystal structures. The colors represent charge polarity (positive charge, blue; negative charge, red).

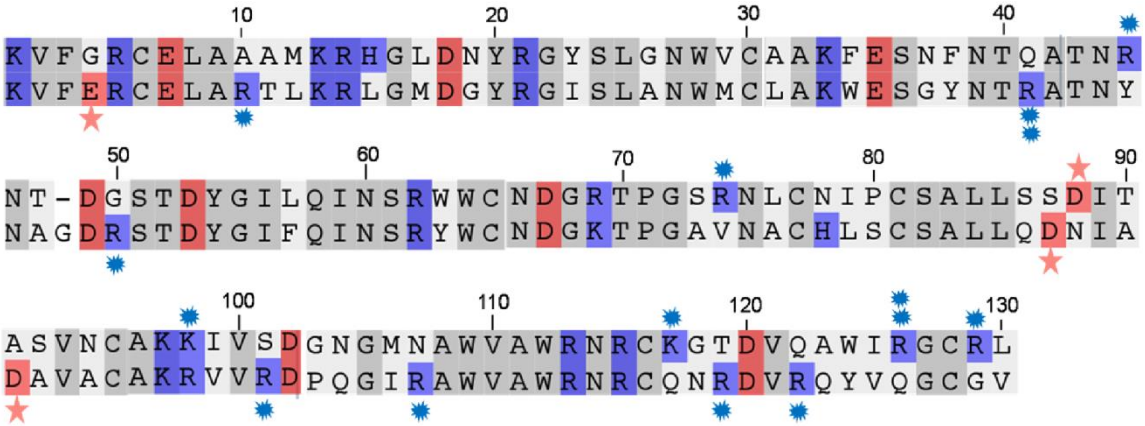


Figure 5.7 Sequence alignment of hen and human lysozyme (top and bottom, respectively). The blue stars indicate basic and acidic sequence variations.

It is not surprising to find that the SNAPP distributions for hen and human lysozyme are not very similar, as shown in Figure 5.8 for charge states +11 and +12. Neither the relative ntensities nor the shapes of the distributions are similar for the two variants. Significantly more 18C6 attaches to human lysozyme, suggesting greater overall availability of the basic residues. The SNAPP results confirm that human and hen lysozyme have distinct surface structures, which is consistent with the picture predicted by x-ray crystallography in Figure 5.6.

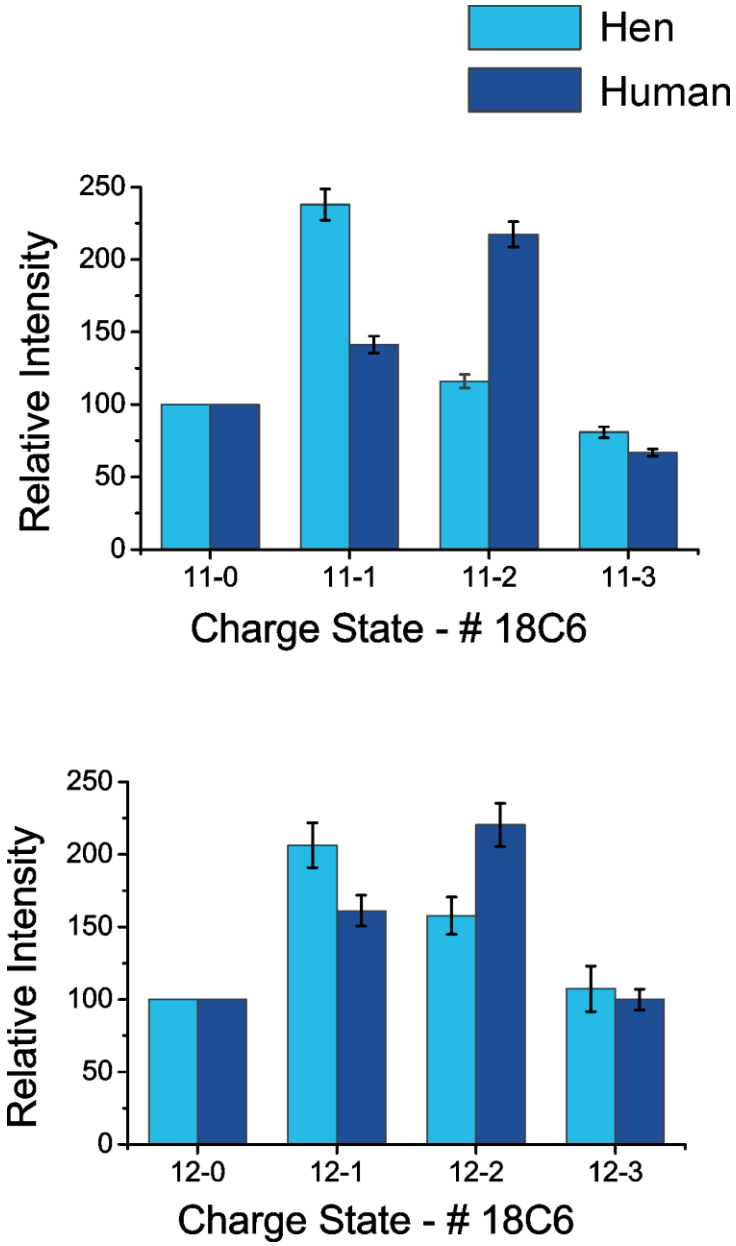


Figure 5.8 SNAPP distributions for hen and human lysozyme, for the +11 and +12 charge states.

5.3.4 Comparative Analysis

The relative changes in SNAPP distributions for insulin, cytc, and lysozyme are summarized in Table 5.1 in comparison with various changes in sequence. The ΔSNAPP column shows the percent change in the SNAPP distribution for each protein (P) as a

function of charge state relative to the first variant (P_{ref}) according to eq 5.1. The calculated differences reflect values between the error bars. Although it is difficult to summarize the complexity of a SNAPP distribution in a single number, the results in Table 5.1 do offer some interesting insight. For example, with insulin, none of the ΔSNAPP scores are particularly high, but bovine insulin exhibits greater scores than porcine insulin, which does correlate with the overall sequence variation. In the case of cytc, the yeast variant clearly stands out as the most distinct protein. Interestingly, the observed shift does not correlate well with changes to the sequence because yeast has more 18C6 molecules attached but fewer basic residues. Many of the basic residues in yeast are shifted in location (see Figure 5.2), which may be the more important factor. For lysozyme, the two proteins can clearly be distinguished by ΔSNAPP values and have significant sequence variation. Overall, the results from this limited data set indicate that a ΔSNAPP score >10 indicates a high probability of surface structure variation. Furthermore, the structural shifts between hen and human lysozyme and between yeast cytc and the remaining cytc proteins are comparable in magnitude. The distinct surface features of proteins with similar backbone structures led us to investigate if other properties, such as denaturation, were dissimilar, as well.

$$\Delta SNAPP = \sum \left| \frac{P}{\sum P} - \frac{P_{ref}}{\sum P_{ref}} \right| \times 100 \qquad (5.1)$$

Table 5.1 Comparison of Relative Changes in SNAPP versus Sequence

		ΔSNAPP[a]			ΔSequence[b]						
Charge State / Protein Name		+8	+9	+10	R	K	D,E	R,K,D,E		All	
Cytc	Horse	0	0	0	2	19	12	33		104	
					ΔR	ΔK	Δ(D,E)	|Δ(R,K,D,E)|		|ΔAll|	
	Bovine	1.7	1.3	0.9	0	-1	0	1	3.03 %	3	2.88 %
	Pigeon	6.7	0.4	2.6	0	-1	-1	2	6.06 %	11	10.6 %
	Yeast	20.2	20.0	10.5	+1	-3	-1	5	15.2 %	45	43.3 %
Insulin		+4	+5	+6	R	K	D,E	R,K,D,E		All	
	Human	0	0	0	1	1	4	6		51	
					ΔR	ΔK	Δ(D,E)	|Δ(R,K,D,E)|		|ΔAll|	
	Bovine	2.9	3.6	1.6	0	0	0	0	0%	3	5.88 %
	Porcine	2.1	1.5	0.3	0	0	0	0	0%	1	1.96 %
Lysozyme		+ 10	+11	+12	R	K	D,E	R,K,D,E		All	
	Hen	0	0	0	11	6	9	26		129	
					ΔR	ΔK	Δ(D,E)	|Δ(R,K,D,E)|		|ΔAll|	
	Human	26.1	33.1	9.7	+3	-1	+2	6	23.1 %	52	40.3 %

[a] The relative changes in SNAPP distributions as defined in eq 5.1. [b]Relative changes in sequence; summed values represent the absolute change. Single letter codes are used for the amino acids.

5.3.5 Protein Denaturation

The stability of native protein structures in atypical solvent systems can vary substantially and is a property which can be easily examined with ESI-MS.[27] As shown above in Figure 5.3a, the tertiary structure of cytc contains three major α-helices and no β-sheets. There is also a structurally relevant heme group covalently linked to the protein, and there are no disulfide bonds to inhibit unfolding.[28] When sufficient methanol is present, cytc adopts a compact state with native-like secondary structures but ill-defined tertiary structure.[29] This state is easily detected by SNAPP,[18] but is not obvious by scrutiny of charge state distributions. Lowering the pH in the presence of methanol unfolds the protein further and a distinct shift in the charge state distribution can be observed by ESI-MS (Figure 5.9).[30]

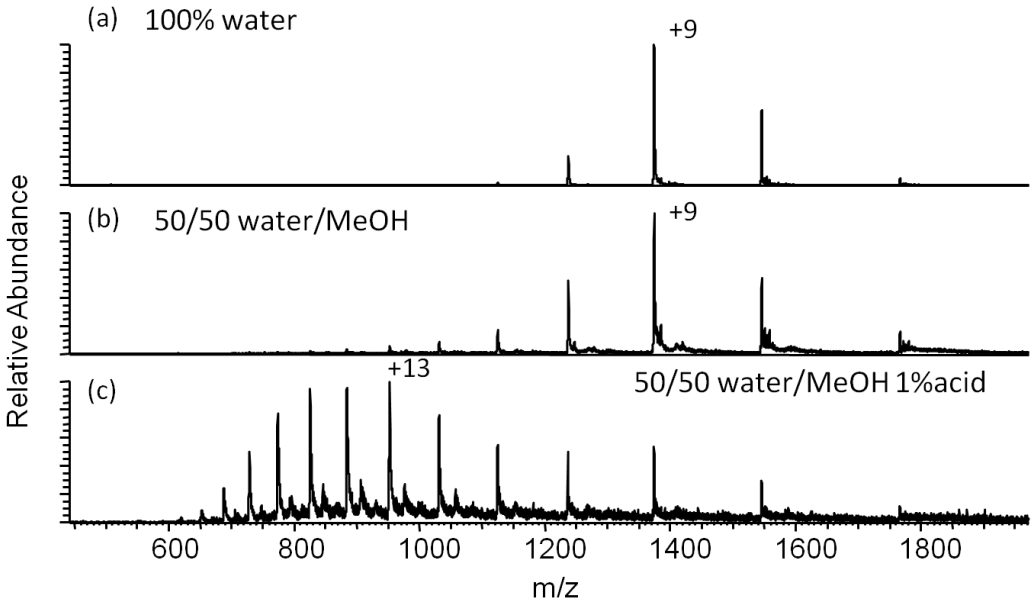


Figure 5.9 ESI-MS spectra of horse cytc in (a) 100% water (b) 50/50 water/methanol and (c) 50/50 water/MeOH with 1% acid. No charge state distribution shift was observed for 50/50 water/MeOH, while a significant shift was observed with 0.1% acid.

Figure 5.10 shows the DESI mass spectra of horse, bovine, pigeon and yeast cytc in 100% water and 50/50 water/methanol at neutral pH. The first three variants exhibit charge state distributions similar to those observed in water, with +9 being the most intense peak. However, for yeast, the introduction of methanol alone is sufficient to unfold the protein, which can be observed by the dramatic shift in the charge state distribution. The results indicate that the tertiary structure of yeast cytc is less stable than horse, bovine, or pigeon despite the fact that the backbone structures are virtually identical. Structural stability is determined by a variety of factors, including packing of the hydrophobic core, hydrogen bonding, salt bridges, etc.[31] Comparison of the cytc sequences in Figure 5.2 reveals that yeast has several mutations to the hydrophobic core in addition to the mutations that lead to a distinct surface structure. The SNAPP data in Figure 5.4 indicates more 18C6 attachment to yeast compared with the other variants, which is consistent with fewer favorable intramolecular ionic or hydrogen bonding interactions and reduced stability. It is likely that some combination of both surface and core mutations leads to the reduced structural stability of yeast cytc.

In Figure 5.11, the SNAPP distributions of +9 charge state for horse, bovine, and pigeon cytc in 50/50 water/MeOH are shown in comparison to those obtained in water. The SNAPP distributions clearly indicate a structural shift has occurred for each of these proteins as well, even though complete denaturation is not observed as in the case of yeast cytc. The SNAPP distributions further suggest that the structure of pigeon cytc has

changed more than horse or bovine (this is particularly evident by comparison of the higher crown adduct peaks). The native structures of all three variants were very similar by SNAPP (Figure 5.4); however, pigeon cytc contains a P44E mutation in a turn region that may enable pigeon to adopt a different partially denatured state. Following the addition of acid to yield complete denaturation, the SNAPP distributions for the denatured states of all four proteins are comparable. Furthermore, the addition of acid does not significantly change the yeast SNAPP distribution relative to that obtained with just the addition of methanol.

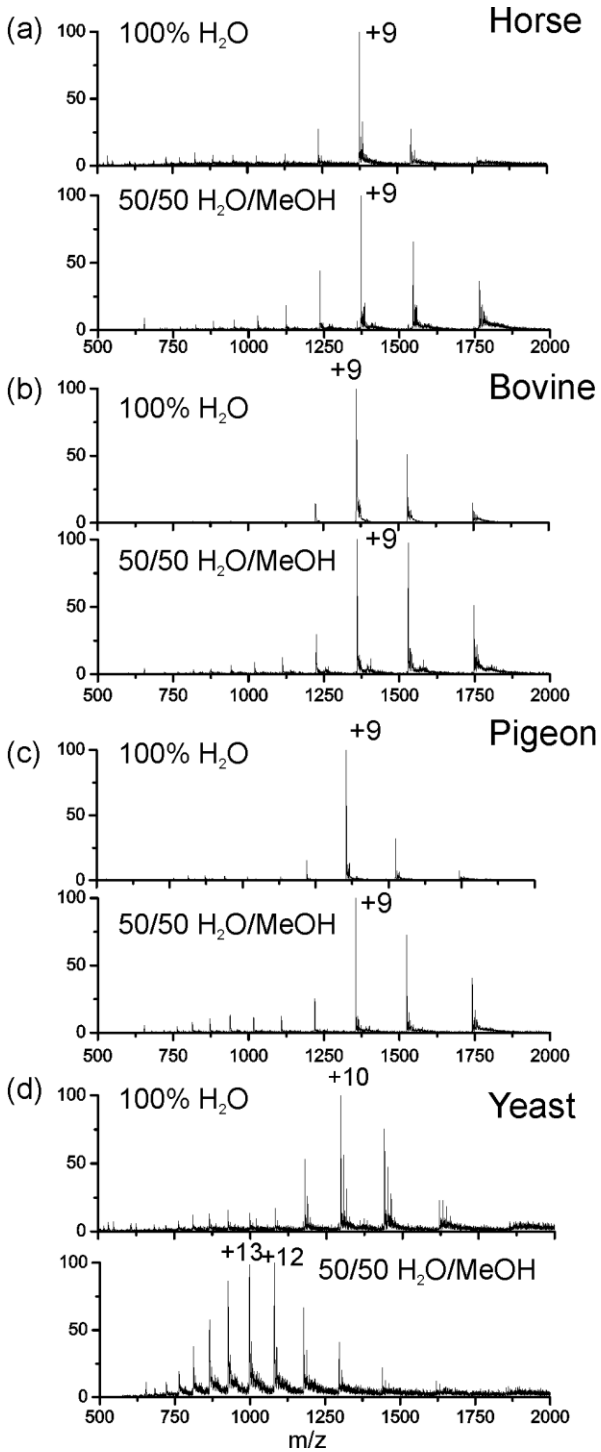


Figure 5.10 DESI mass spectra of (a) horse, (b) bovine, (c) pigeon, and (d) yeast cytc in 100% water and 50/50 water/methanol mixture. The yeast variant is significantly unfolded in 50% methanol.

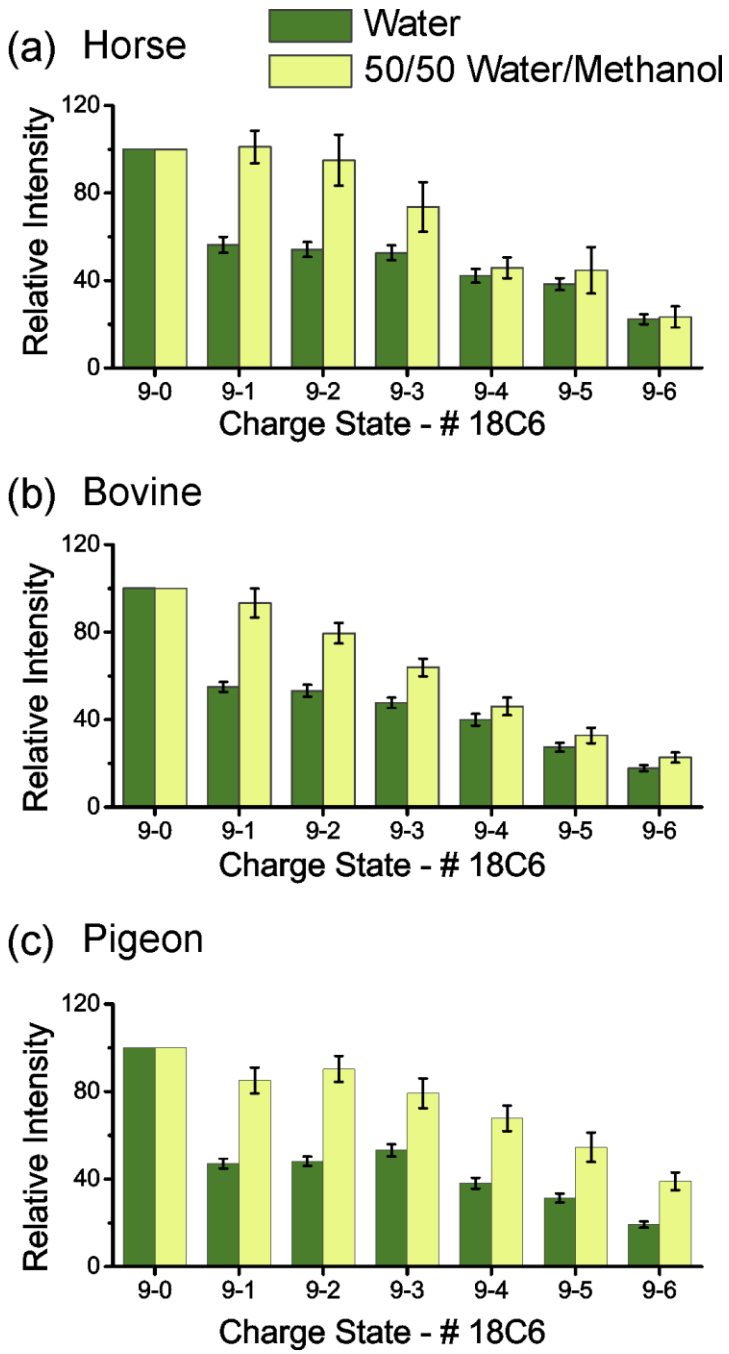


Figure 5.11 SNAPP distributions of +9 charge state for horse, bovine, and pigeon cytc in water and 50/50 water/MeOH. Partial denaturation in the presence of methanol when compared to that in water.

5.4 Conclusions

The SNAPP−MS method is shown to be a useful probe of surface structure for proteins with highly homologous sequences and nearly identical three-dimensional backbone structures. For insulin, very minor sequence variation leads to very similar SNAPP distributions for three variants. In the case of cytc, yeast exhibits the greatest change in surface structure and also denatures more easily in the presence of an organic solvent. These changes do not correlate well solely with differences in the number of potential 18C6 binding sites, suggesting that sequence shifts that alter the surface environment of basic residues are also important. This hypothesis is further supported by results with lysozyme where significant sequence shifting yields quite disparate SNAPP distributions. These findings are consistent with previous results in which SNAPP distributions have been shown to be sensitive to the availability of charged basic side chains. Changes to the arrangement of charged groups on the protein surface lead to changes in the observed SNAPP distributions. These results clearly demonstrate that proteins that adopt nearly identical tertiary structure may have substantially different electrostatic surfaces that could easily modulate protein−protein or small molecule recognition responses. The SNAPP−MS method is easy to conduct and requires minimal sample consumption, which should make it ideal for structure validation for proteins that have been subjected to site-directed mutagenesis. Furthermore, SNAPP can assess variations in structurally ill-defined or highly dynamic states, such as proteins that have been partially or fully denatured. The effects of mutations can therefore be tracked from completely folded to completely unfolded structures by the same method.

References

[1] Sung, Y. H.; Eliezer, D., Residual structure, backbone dynamics, and interactions within the synuclein family. *J. Mol. Biol.* **2007**, *372*, 689-707.

[2] Kruger, R.; Kuhn, W.; Muller, T.; Woitalla, D.; Graeber, M.; Kosel, S.; Przuntek, H.; Epplen, J. T.; Schols, L.; Riess, O., Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nature Genet.* **1998**, *18*, 106-108.

[3] Polymeropoulos, M. H.; Lavedan, C.; Leroy, E.; Ide, S. E.; Dehejia, A.; Dutra, A.; Pike, B.; Root, H.; Rubenstein, J.; Boyer, R.; Stenroos, E. S.; Chandrasekharappa, S.; Athanassiadou, A.; Papapetropoulos, T.; Johnson, W. G.; Lazzarini, A. M.; Duvoisin, R. C.; DiIorio, G.; Golbe, L. I.; Nussbaum, R. L., Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science*, **1997**, *276*, 2045-2047.

[4] Zarranz, J. J.; Alegre, J.; Gomez-Esteban, J. C.; Lezcano, E.; Ros, R.; Ampuero, I.; Vidal, L.; Hoenicka, J.; Rodriguez, O.; Atares, B.; Llorens, V.; Tortosa, E. G.; del Ser, T.; Munoz, D. G.; de Yebenes, J. G., The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann. Neurol.* **2004**, *55*, 164-173.

[5] Ingram, V. M., Gene mutations in human haemoglobin- chemical difference between normal and sickle cell haemoglobin. *Nature*, **1957**, *180*, 326-328.

[6] Harrington, D. J.; Adachi, K.; Royer, W. E., Crystal structure of deoxy-human hemoglobin beta 6 Glu -> Trp - Implications for the structure and formation of the sickle cell fiber. *J. Biol. Chem.* **1998**, *273*, 32690-32696.

[7] Rao-Naik, C.; delaCruz, W.; Laplaza, J. M.; Tan, S.; Callis, J.; Fisher, A. J., The rub family of ubiquitin-like proteins - Crystal structure of arabidopsis Rub1 and expression of multiple rubs in arabidopsis. *J. Biol. Chem.* **1998**, *273*, 34976-34982.

[8] Finley, D.; Varshavsky, A., The ubiquitin system – functions and mechanisms. *Trends Biochem. Sci.* **1985**, *10*, 343-347.

[9] Taylor, S. V.; Kast, P.; Hilvert, D., Investigating and engineering enzymes by genetic selection. *Angew. Chem.-Int. Edit.* **2001**, *40*, 3310-3335.

[10] Mirkin, N.; Jaconcic, J.; Stojanoff, V.; Moreno, A., High resolution X-ray crystallographic structure of bovine heart cytochrome c and its application to the design of an electron transfer biosensor. *Proteins*, **2008**, *70*, 83-92.

[11] Ferentz, A. E.; Wagner, G., NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q. Rev. Biophys.* **2000,** *33*, 29-65.

[12] Chowdhury, S. K.; Katta, V.; Chait, B. T., Probing Conformational-Changes in Proteins by Mass-Spectrometry. *J. Am. Chem. Soc.* **1990,** *112*, 9012-9013.

[13] Grandori, R., Origin of the conformation dependence of protein charge-state distributions in electrospray ionization mass spectrometry. *J. Mass Spectrom.* **2003**, *38*, 11-15.

[14] Kaltashov, I. A.; Mohimen, A., Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. *Anal. Chem.* **2005,** *77*, 5370-5379.

[15] Wales, T. E.; Engen, J. R., Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom. Rev.* **2006**, *25*, 158-170.

[16] Mendoza, V. L.; Vachet, R. W., Protein surface mapping using diethylpyrocarbonate with mass spectrometric detection. *Anal. Chem.* **2008**, *80*, 2895-2904.

[17] Sinz, A., Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom. Rev.* **2006**, *25*, 663-682.

[18] Ly, T.; Julian, R. R., Using ESI-MS to probe protein structure by site-specific noncovalent attachment of 18-crown-6. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1209-1215.

[19] Liu, Z. J.; Cheng, S. J.; Gailie, D. R.; Julian, R. R., Exploring the mechanism of selective noncovalent adduct protein probing mass spectrometry utilizing site-directed mutagenesis to examine ubiquitin. *Anal. Chem.* **2008**,*80*, 3846-3852.

[20] Ly, T.; Julian, R. R., Protein-metal interactions of calmodulin and alpha-Synuclein monitored by selective noncovalent adduct protein probing mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1663-1672.

[21] Sun, Q. Y.; Tyler, R. C.; Volkman, B. F.; Julian, R. R., Dynamic interchanging native states of lymphotactin examined by SNAPP-MS. *J. Am. Soc. Mass Spectrom.* **2011,** *22*, 399-407.

[22] Hamdy, O.; Julian, R. R., Reflections on charge state distributions, protein structure, and the mystical mechanism of electrospray ionization. *J. Am. Soc. Mass. Spectrom.* **2012**, *23*, 1-6.

[23] Timofeev, V. I.; Chuprov-Netochin, R. N.; Samigina, V. R.; Bezuglov, V. V.; Miroshnikov, K. A.; Kuranova, I. P., X-ray investigation of gene-engineered human insulin crystallized from a solution containing polysialic acid. *Acta Crystallogr. F-Struct. Biol. Cryst. Commun.* **2010**, *66*, 259-263.

[24] Ishikawa, T.; Chatake, T.; Ohnishi, Y.; Tanaka, I.; Kurihara, K.; Kuroki, R.; Niimura, N., A neutron crystallographic analysis of a cubic porcine insulin at pD 6.6. *Chem. Phys.* **2008**, *345*, 152-158.

[25] Gursky, O.; Li, Y. L.; Badger, J.; Caspar, D. L. D., Monovalent cation binding to cubic insulin crystals. *Biophys. J.* **1992**, *61*, 604-611.

[26] Banci, L.; Bertini, I.; Huber, J. G.; Spyroulias, G. A.; Turano, P., Solution structure of reduced horse heart cytochrome c. *J. Biol. Inorg. Chem.* **1999**, *4*, 21-31.

[27] Kaltashov, I. A.; Eyles, S. J., Studies of biomolecular conformations and conformational dynamics by mass spectrometry. *Mass Spectrom. Rev.* **2002**, *21*, 37-71.

[28] Fisher, W. R.; Taniuchi, H.; Anfinsen, C. B., Role of heme in formation of structure of cytochrome-c. *J. Biol. Chem.* **1973**, *248*, 3188-3195.

[29] Kamatari, Y. O.; Konno, T.; Kataoka, M.; Akasaka, K., The methanol-induced globular and expanded denatured states of cytochrome c: A study by CD fluorescence, NMR and small-angle X-ray scattering. *J. Mol. Biol.* **1996,** *259*, 512-523.

[30] Konermann, L.; Douglas, D. J., Acid-induced unfolding of cytochrome c at different methanol concentrations: Electrospray ionization mass spectrometry specifically monitors changes in the tertiary structure. *Biochemistry*, **1997,** *36*, 12296-12302.

[31] Dyson, H. J.; Wright, P. E.; Scheraga, H. A., The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc. Natl. Acad. Sci.* **2006**, *103*, 13057-13061.

Chapter 6

CONCLUDING REMARKS

The goal of my PhD study during the last five years is to develop new mass spectrometry based techniques to study large biomolecules. The method described herein would benefit the identification of subtle post translational modification (PTM) and the characterization of protein solution phase structure.

While many PTMs are catalyzed by enzymes immediately after the protein are synthesized in cell, epimerization and isomerization are usually spontaneous reactions. The spontaneous isomerization processes are associated with age-related protein dysfunctions and pathogenesis of diseases. To date, these PTMs have not received sufficient attentions due to the difficulty in the detection. The work in Chapter 2 and Chapter 3 leads to the comprehensive study of isomerization in biological systems. Based on the LC-MS/MS method proposed in Chapter 3, future direction is the application of this methodology to explore isomerization in a broad variety of proteins from different organisms. The comparison of isomerization ratio between healthy and damaged cells can potentially facilitate the discovery of biomarkers for certain diseases.

One difficulty in mass-spectrometry based isomerization detection lies in the identification of the specific isomerized amino acid in a peptide. Although tandem mass spectra can be used to distinguish two peptide isomers, the location of the isomerization

cannot be directly determined without comparing with synthetic standard peptides. Another future direction could be the analysis of a tandem mass spectral database from a vast number of standard peptide epimers, trying to unravel possible correlation between the isomerization site and the dissociation pattern differences. Since RDD fragmentation is highly structurally sensitive and the radical migration pathway is dependent on the peptide local structure, it is very likely that side chain losses or radical induced backbone fragmentations from RDD can reflect the location of isomerized amino acid. Still, a large fragmentation spectra dataset of peptide epimers is required to draw a firm conclusion.

Besides isomerization and epimerization, several other PTMs were identified in the crystallin sample in Chapter 3, including disulfide bond, oxidation, glutathionylation, etc. Most of these PTMs are not listed in the protein database and thus might occur during aging. These modifications can be responsible for the protein aggregation in the eye lens and cataract disease. A deep understanding of the age related PTMs is therefore needed. Additionally, the relative ratios of these PTMs between the water-soluble and water-insoluble proteins can reveal the effects of PTMs on protein aggregation.

For examining whole protein structure or large protein complex by mass spectrometry, it is crucial that the protein native structure is preserved during the electrospray (ESI) process. Results from this thesis demonstrated that 18C6 can protect protein solution phase structure by intramolecular charge solvation and thus be used to probe protein electrostatic surface structure and dynamic state. Analysis of the protein-18C6 complex stability provides another aspect of information about protein structure or protein

unfolding process during the transition to gas phase. This inspires us to further study protein structural change in different solvent conditions. Furthermore, results from the SNAPP experiment of orthologous proteins from various species demonstrated that although two proteins have highly aligned backbone structures, sequence mutations can significantly affect protein surface structures, which could easily modulate protein-protein interactions. The SNAPP technique allows further investigation of the influence of single point mutations on protein dynamic structure and function.

In summary, this dissertation described a new aspect of proteomics research, isomer proteomics, for large-scale analysis of peptide isomerization in complex biological samples. Extending this newly developed proteomics method opens a door to a full investigation of age related PTMs and potentially a better understanding of disease pathology. This work also demonstrated the utility of 18C6 adducts for protecting protein solution phase structure and probing protein surface structure.