UNIVERSITY OF CALIFORNIA
RIVERSIDE

Mass Spectrometry Based Method Development for Elucidation of Protein Sequences, Structures and Post Translational Modifications

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

Doctor of Philosophy

in

Chemistry

by

Qingyu Sun

August 2011

Dissertation Committee:
Dr. Ryan R. Julian, Chairperson
Dr. Cynthia K. Larive
Dr. Yinsheng Wang

The Dissertation of Qingyu Sun is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENT

The completion of this dissertation would not have been possible without support from numerous people and funding agencies that I would like to acknowledge here. My advisor Dr. Ryan Julian deserves the most credit for helping me carrying out my research projects and developing problem solving skills. I want to thank Ryan for introducing me to the amazing world of mass spectrometry and training me strictly to be an independent scientist. Ryan always gave me inspiring suggestions to help me achieve progresses in research. Nevertheless, most of time, he encouraged me to solve problems on my own so that I can build up the capability of doing independent research. In addition, Ryan also taught me presentation and paper-writing skills, showed me how to communicate in a big team, and even gave me several lessons on how to be a good person. Under his supervision, I indulged in my interest in the field of mass spectrometry and have learned how to be a good scientist as well as a good person.

I thank those great colleagues in my lab, current and former for helping me with experiments and carrying out useful discussions (Dr. Tony Ly, Dr. Zhenjiu Liu, Jolene Diedrich, Ben Moore, Eric Knudsen, Yuanqi Tao, Vic Zhang, Julie Tsu, Emily Spenser and Geoff Yeh). I also thank the undergraduate Gene Yoo, who worked hard with me during the summer of 2010 and helped me finish my last project in graduate school.

I thank the following institutions for funding: UC Riverside (RRJ), the National Science Foundation CAREER Award (RRJ, CHE-0747481), the National Institutes of Health (RRJ, 1R01GM084106-01A1), and the UCR Graduate Division. I thank Professor

Brian Stoltz and Hosea Nelson in Caltech, collaborators of the research project described in Chapter 2, who kindly provided me the derivatized 18C6 molecule so that I could carry out the entire research in Chapter 2. I also want to express my gratefulness to Professor Joseph Loo and Dr. Sheng Yin in UCLA, collaborators for the work introduced in Chapter 3, who gave me the access to the LTQ-ICR mass spectrometer in UCLA. This collaboration work was funded by the National Institutes of Health (JAL, RR 20004) and a National Institutes of Health/National Center for Research Resources High-End Instrumentation Award (JAL, S10 RR023045). I also want to thank Professor Brian Volkman and Dr. Robert Tyler, collaborators of the research project discussed in Chapter 5. All the protein samples used in Chapter 5 were expressed and purified by Dr. Robert Tyler. The collaboration work was supported by National Institutes of Health (BFV, R01 AI063325).

Finally, I thank Dr. Yinsheng Wang, Dr. Cindy Larive, Dr. Quan (Jason) Chen and Dr. Wenwan Zhong for teaching me advanced analytical chemistry knowledge in graduate school. I thank my friends who always encouraged me when I had troubles in the past five years. I especially thank my mom and my boy friend Dr. Zijun Fang, who have always been my solid support and constantly given me encouragement and hope to make me continuously grow in my personal life and professional career.

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

Chapter 2: Sun, Q.; Nelson, H.; Ly, T.; Stoltz, B. M.; Julian, R. R. Side chain chemistry mediates backbone fragmentation in hydrogen deficient peptide radicals. *Journal of Proteome Research*, **2009**, 8, 958-966.

Chapter 3: Sun, Q.; Yin, S.; Loo, J. A.; Julian, R. R. Radical directed dissociation for facile identification of iodotyrosine residues using electrospray ionization mass spectrometry. *Analytical Chemistry*, **2010**, 82, 3826-3833.

Chapter 4: Sun, Q.; Julian, R. R. Probing sites of histidine phosphorylation with iodination and tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, **2011**, in press.

Chapter 5: Sun, Q.; Tyler, R. C.; Volkman, B. F.; Julian, R. R. Dynamic Interchanging Native States of Lymphotactin Examined by SNAPP-MS. *Journal of the American Society for Mass Spectrometry*, **2011**, 22, 399-407.

To my grandpa,

who is my first mentor and showed me how to live positively and happily

my entire life.

ABSTRACT OF THE DISSERTATION

Mass Spectrometry Based Method Development for Elucidation of Protein Sequences, Structures and Post Translational Modifications

by

Qingyu Sun

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, August 2011
Dr. Ryan R. Julian, Chairperson

The advanced development of mass spectrometry (MS) makes MS a powerful technique for proteomics study. The increasing demands for proteomics study stimulate creation of more applicable MS-based methods. This dissertation focuses on development of novel MS-based methods to characterize three different aspects of proteins: primary sequence, post-translational modifications (PTMs) and three-dimensional (3D) structure.

The first part (Chapter 2) is focused on primary sequence characterization. A novel technique radical directed dissociation (RDD) is developed for peptide gas phase fragmentation. RDD is recognized as a "charge-remote" dissociation in which RDD backbone fragments are mediated by βC-H bond dissociation energies (BDEs) of all 20 amino acids. Therefore, RDD fragmentation of peptides with known sequences is predictable. This discovery indicates RDD is promising for proteomics study in terms of significantly improving the confidence level of peptide identification.

The second part covers MS-based method development for PTM characterization. Chapter 3 describes using RDD to rapidly map iodinated tyrosines in intact proteins. The iodinated tyrosines are identified in the portions of a protein where RDD fragments are frequently located. The only limitation is that when multiple tyrosines are located in the identified iodinated region, it could be challenging to differentiate the iodination states between adjacent tyrosines.

In Chapter 4, iodination chemistry combined with tandem MS is used to localize acid-labile histidine phosphorylation sites in peptides. Phosphorylated histidines are immune of iodine-labeling whereas unmodified histidines are not. After iodination, labile histidine phosphates can be removed by acid treatment to yield free histidines, which can be easily identified by MS/MS. This new method provides a pathway forward for analyzing histidine phosphorylation in complex systems.

The last part is aimed at protein conformational study using selective noncovalent adduct protein probing-mass spectrometry (SNAPP-MS). SNAPP-MS employs 18-crown-6 ether (18C6) as a probe of lysines and the numbers of 18C6s attached to proteins reveal protein solution phase structural information. In Chapter 5, SNAPP-MS is used to study the dynamic structural interchanging in Lymphotactin (Ltn). SNAPP successfully encodes the interchanging transition state of Ltn, which provides valuable information for interpreting Ltn biological functions.

TABLE OF CONTENTS

LIST OF FIGURES

List of Schemes

CHAPTER 1

PROTEIN CHARACTERIZATION USING MASS SPECTROMETRY

1.1 Introduction

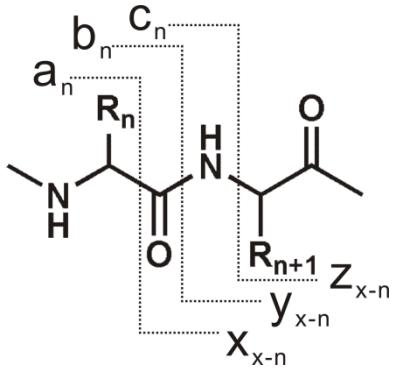
Proteins are one of the most abundant and complicated biomolecules in biological systems. They are linear polymer chains of amino acids covalently linked by amide bonds (peptide bonds). During protein biosynthesis, 20 standard amino acids are translated by genetic codes to form unique polypeptide sequences. Post-translational modifications (PTMs) occur at certain amino acid residues after formation of linear polypeptide chains and are independent of translation of polypeptide chains, therefore add additional complexities to proteins. Besides chemical composition of proteins in terms of sequence and PTMs, proteins also adopt specific three-dimensional (3D) structures in a biological way. Proteins have various biological functions including catalyzing biochemical reactions, maintaining cellular structure and mechanic functions, mediating cell signaling, etc. Any change of the sequence, PTMs or 3D structure of a protein could cause a fluctuation of protein biological functions and subsequently lead to malfunctions which could induce various diseases. Therefore, protein characterization is always a key interest, which includes but not limited to protein sequences, PTMs, domain structures, interactions, localization and system activities.[1]

Several analytical techniques have been developed to examine proteins, such as Edman sequencing for protein sequence elucidation;[2] fluorescence based cell imaging

techniques, different array and chip experiments for study of protein structures, localization etc.[3-5] With rapid development of mass spectrometry (MS) in the last two decades, MS has increasingly become another powerful tool for characterization of proteins.[6] Introduced in 1919 by JJ Thompson and Aston, MS is defined as an analytical technique that measures the mass to charge (m/z) ratio of analytes. A mass spectrometer is composed of an ion source to convert sample analytes into charged ions, a mass analyzer to obtain m/z ratios of analyte ions and a detector to record the number of ions at each m/z value. The original applications of mass spectrometry were mainly focused on element and small organic compound analysis until two soft ionization techniques: electrospray ionization (ESI) [7]and matrix-assisted laser desorption ionization (MALDI)[8] were introduced in the 1980s. Soft ionization sources gently ionize large molecules and retain the integrity of analyte ions for m/z analysis, which revolutionize protein and peptide analysis using mass spectrometry. Meanwhile, the rapid development of mass analyzers from 1980s to 2000s significantly improves the sensitivity, resolution and dynamic ranges of mass spectrometers for biomolecule characterization. In addition, the rapid development of separation technology, such as liquid chromatography (LC), simultaneously allows large loads of proteins with high orders of complexity and greatly facilities biomolecule analysis.

The rapid growth of MS as well as separation techniques makes MS-based technology rises as a promising tool for proteomics study, which adopts many advantages over classic biochemistry methods. However, it is still highly desired to improve MS performance in proteomics study in terms of increasing biomolecule identification

confidence level, simplifying MS data analysis, extending MS application to highly challenging analytes, such as extremely-labile PTMs, etc. Therefore, development of more useful MS-based methods is still in high demands. The objective of this dissertation is aimed at development of novel MS-based analytical methods for characterization of proteins with focus on three important aspects: primary sequences, PTMs and 3D structures.



Scheme 1.1 Nomenclature of peptide fragments by tandem mass spectrometry

1.2 Primary sequence elucidation

Sequence information of peptides and proteins can be conveniently obtained by gas phase dissociation of the selected ion molecules in the mass analyzer, the entire process of which is called tandem mass spectrometry (tandem MS). The nomenclature of peptide fragments by tandem MS is shown in Scheme 1.1. [9] Tandem MS has been an indispensable technique for elucidation of primary sequences of biomolecules. Several tandem MS methods have been invented for peptide/protein gas phase sequencing, among which collision induced dissociation (CID) is the most widely used dissociation

technique in proteomics study.[10] By exciting peptide/protein ions via multiple collisions with gas molecules, CID of peptides or proteins leads to cleavage of amide bonds in peptide backbones and generates b- and y-type fragment ions. CID is considered as a low energy dissociation which tends to break weak bonds in peptides and usually results in random backbone fragments with relatively low sequence coverage when compared to recently emerged electron capture/transfer dissociation (ECD/ETD).[11,12] Both ECD and ETD utilize electrons to facilitate backbone dissociation and subsequently generate primarily c- and z-type fragments. In addition to generating fragments with good sequence coverage, ECD and ETD are known to preserve labile PTMs in peptides and proteins,[13,14] such as phosphorylation. Photoactivation based dissociations, such as infrared multiphoton dissociation (IRMPD)[15] and ultra-violet photodissociation (UVPD),[16] are also available for characterization of peptides and proteins.

The fast growth of computational technologies also facilitates high throughput identification of peptides and proteins using MS. Several databases have been established which contains a large pool of "predicted" tandem mass spectrum of proteome with known sequences using different algorithms.[17] Unknown peptides can be rapidly identified by comparing experimentally generated tandem mass spectrum to the established proteome MS databases for matching hits. It seems logical that the identification confidence level and database searching rate will be significantly improved if "algorithm predicted" MS/MS spectrum are substantially overlapped with the experimental tandem mass spectrum. In other words, an ideal tandem MS method with predictable fragmentation is desired, which should necessarily improve the confidence

level of data base searching. It has been reported that mobile proton regulates CID dissociation pathway, which means the variation of charge location and numbers largely impacts peptide CID fragmentation. As a result, CID fragmentation of a known peptide is very difficult to predict.[18] Mechanisms of ECD and ETD are still under debate, but it is clear that both of them are also "charge driven" dissociation, which therefore makes prediction of ECD/ETD fragmentation patterns of known peptides still challenging. Although prediction of peptide fragments by CID, ECD and ETD has been reported; it is usually based on many empirical models and several artificial parameters have been mandatorily involved in the prediction algorithms. [19-21] The uncertainty of mobile protons in peptide backbones is always the main problem for prediction of peptide tandem mass spectrum. Therefore, this motivation subsequently stimulates development of alternative charge-independent dissociation methods.

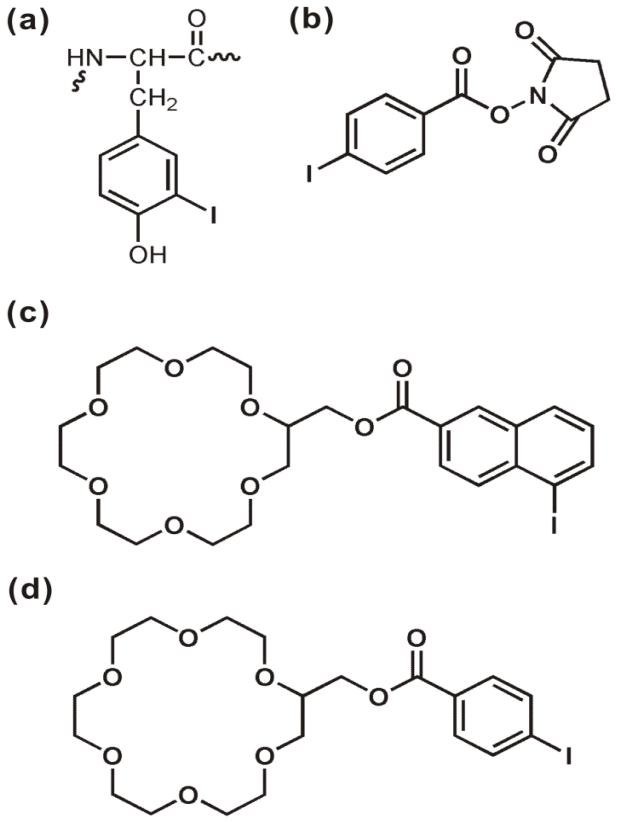


Figure 1.1 Currently available radical initiators used in RDD experiments. (a) Iodo-tyrosine residue (b) 5-Iodo-benzoyl-N-hydroxy-succinimide ester (c) 5-Iodo-2-(2-hydroxymethyl-18-crown-6)-naphthoate (d) 2-(hydroxyl-methyl-iodobenzoylester)-18C6.

Radical directed dissociation (RDD) is a recently developed radical chemistry based tandem MS method by Julian and coworkers,[22] which employs radicals to "direct" backbone fragments in peptides and proteins. In a RDD experiment, the photolysis of C-I bond is employed to initiate radicals on peptide or protein ions. C-I bond, known as a photolabile bond, can absorb photon energy to be activated to the excited state and homolytically breaks into two radicals. Several ways are available to introduce a C-I bond to peptides and proteins. Iodination of tyrosines is considered as one of the easiest

modifications of peptides and proteins. The side chains of tyrosines can be easily iodinated up to 100% using NaI and Chloramine T. The chemical structures of generated iodinated tyrosines are shown in Fig.1.1a. Ultra-violet photodissociation (UVPD) of iodinated biomolecules at 266 nm breaks the C-I bond among thousands of chemical bonds in a highly selective way and generates peptide/protein radicals in a decent yield (>40%), which can be further isolated for collision activation. The presence of hydrogen deficient radicals will direct peptide ions to fragment in a specific way which finally generates unique RDD fragmentation. In addition to iodinated tyrosines, 3 alternative radical initiators, as shown in Fig. 1.1b-d,[23,24] with each containing an iodine-labeled chromophore, have been synthesized for C-I bond introduction. These radical initiators can be conveniently coupled to peptides or proteins via formation of covalent or non-covalent chemical bonds. 18-crown-6 ether (18C6) is employed as a molecule recognition probe to form non-covalent bonds with peptides/proteins because 18C6 has good affinities towards the side chain of lysine and arginine as well as prononated N-terminus. During ESI process, 18C6 molecules rapidly attach to peptides or proteins and subsequently form stable gas phase complexes.[25,26] It is advantageous to use 18C6-based radical initiators because sample preparation can be maximally minimized by simply mixing peptide solution with 18C6-based initiators prior to ESI-MS analysis. Additionally, 18C6 does not absorb in near-ultraviolet and therefore will not interfere with photoactivation step for peptide/protein radical generation.

Distinct from CID, ECD or ETD, a-, c- and z-type ions associated with neutral side-chain losses are considered as typical RDD fragments. A comprehensive study of peptide

RDD fragmentation using the 18C6-based iodo-naphthoate radical initiator (Fig. 1.1c) is described in Chapter 2. Interestingly, our research discovered that peptide side chain chemistry instead of protons drives RDD fragmentation. In Chapter 2, more than 20 peptides have been subjected to fragmentation by RDD. The statistic analysis of RDD fragments of all the peptides shows that RDD backbone fragmentation is actually predictable, which is mediated by bond dissociation energies (BDEs) of βC-H bonds in the side chain of peptide residues. For example, abundant a-, c- and z-ions have been frequently observed at aromatic residues, such as tryptophan, tyrosine, histidine, because of the relatively low βC-H BDE values. Therefore, RDD is classified as "charge-remote" gas phase dissociation, which makes it a promising tandem MS method. RDD provides the foundation to establish predictable proteome tandem MS database to improve the confidence level of peptide identification. In addition to predictable backbone fragments, 11 amino acids out of 20 generate unique side-chain losses, which would be useful when combined with predictable RDD backbone fragments for rapid identification of peptides.

1.3 Mapping post-translational modification sites in peptides and whole proteins

Genes encode basic sequences of proteins,[27] whereas most proteins are not considered as mature states until they undergo the "decoration" by post-translational modifications. PTMs, as the most important post-chemistry after translation, play important roles in regulation of protein functions. More than 200 different varieties of PTMs have been reported in biological system so far[28] and new PTMs are still under investigations. Common PTMs include methylation, acetylation, phosphorylation, glycosylation etc. Each PTM can modify one or a few specific residues in proteins, which is responsible for mediating at least one regulatory function.[29] It has been discovered that multiple PTMs in proteins can substantially function together to mediate biological reactions and to adjust overall protein functionality and stability. Therefore, PTMs of proteins have been attracting enormous attentions from biological and biomedical research communities.

Recently, study of PTMs in proteins increasingly becomes one of the key research interests of MS-based proteomics. In fact, MS is an ideal tool to characterize PTMs in proteins because the addition of chemical moieties with specific mass shifts can be accurately monitored by MS and Tandem MS can even clearly elucidate which residues are specifically modified. Two independent MS based strategies, identified as "bottom-up" and "top-down" approaches, both have been frequently employed for PTM identification.[30] Bottom-up method analyzes small peptides from protein proteolysis, followed by tandem MS to collect as much sequence information as possible. Finally, the global picture of protein sequences associated with PTM site information is obtained by data base searching against the collected MS/MS spectrum. In contrast, top-down

approach dissects intact proteins directly in the gas phase, which is known as “gas phase proteolysis”. In top-down proteomics experiments, protein mixtures are either subjected to off-line or on-line LC separation, followed by gas phase dissociation of purified proteins one at a time and mass analysis of product ions in a high resolution mass analyzer for sequence and PTM site information. Bottom-up strategy for PTM site mapping allows high through-put analysis of highly complicated protein mixtures, however, whole sequence coverage cannot be always achieved which usually leads to incomplete PTM site mapping. This issue could be avoided by carrying out top-down proteomics approach since the entire protein is placed under investigation. The key to top-down mapping of PTMs is protein fragmentation efficiency. As protein size increases, the overall complexity of protein tertiary structure in the gas phase would correspondingly rises, which could subsequently decrease top-down fragmentation efficiency. As a result, backbone fragmentation might not necessarily occur on both sides of the post-translational modified residues, which also causes loss of PTM site information unfortunately.

To map one specific PTM site in a peptide or protein using tandem MS, an ideal situation is that two cleavages are generated at the modified residue and the adjacent residue on either side. The site-specific cleavages significantly simplify data analysis because the shift in mass from modified to unmodified fragments easily encodes the PTM site. It has been reported that phosphates at serines and threonines in peptides can be selectively replaced by chromophores. Photodissocation (PD) of chromophore-labeled peptides leads to selective cleavages at chromophore-labeled serines and threonines,

which represent the originally phosphorylated sites.[31] In fact, selective cleavages at modified sites have been achieved in top-down dissociation of whole proteins as well.[22] In Chapter 3, RDD has been found to yield selective backbone fragments around iodinated tyrosines in intact proteins, which is promising for rapid localization of naturally iodinated tyrosine sites in proteins. Our study discovers that RDD of an intact iodinated protein produces highly concentrated backbone fragments in localized narrowed portions (within 10 residues from the iodinated center) where iodinated tyrosine residues are always included. This discovery solidates the potential applications of RDD for mapping natural iodinated tyrosines in intact proteins (such as thyroglobulin), which will significantly simplify the experimental and data analysis work required in regular bottom-up and top-down approaches.

Additionally, this discovery also inspires us with the idea that iodination chemistry combined with RDD would be an ideal technique to map surface-available tyrosines in proteins, which provides useful information for protein conformational study. It has been known that iodination chemistry is sensitive to protein tertiary structures: only tyrosines with good solvent accessibilities will be labeled by iodines.[32] RDD can easily decipher the iodinated tyrosine sites which represent the accessible tyrosines on the surface. In Chapter 3, myoglobins with different initial structures were iodinated and investigated by RDD respectively. The results have shown that different tyrosines were iodinated after the native structure of myoglobin was altered, which indicates the surface accessibility of tyrosines has been changed.

Other than development of new techniques (such as site-specific dissociation) to solve the challenging data analysis issue in all proteomics experiments, scientists also focus on invention of novel methods to improve identification of low abundant and highly labile PTMs. Most PTMs are present at sub-stoichiometric level in proteins, which means only a small fraction at given sites are modified by post-translational chemistry. The low abundance of PTMs challenges advanced MS-based techniques for PTM identification, especially when the high labile property is additionally involved. Phosphorylation at histidine residues is one of the examples. Phospho-histidine is critical for mediating two-component signal transduction in eubacteria, archaea, plants, and eukaryotic organisms.[33,34] It is identified as a highly labile PTM due to the intrinsic biological function that the labile N-P bond is leveraged for phosphate transfer from phosphodonor to phosphoacceptor moieties. Phospho-histidines are super unstable under acidic environment (half-life < 30 min at pH 3),[35] obstructing the employment of standard LC technique (optimal pH <1.5) for separation of analytes containing phospho-histidines. As a result, it is impossible to exert the advantages of MS for phospho-histidine site identification without a good separation.

Several methods have been developed to minimize the loss of phosphates on histidines, but it is achieved at the cost of separation resolution or elongated separation time. Most of time, phosphates at histidines are still unavoidably lost using a fast or a neutral pH separation, with the hope that a small amount can be eventually preserved. To improve identification of phospho-histidines using LC-MS techniques, an iodine-labeling based subtractive method is developed, as described in details in Chapter 4. This new

subtractive method makes use of iodination chemistry to label all free histidine residues in the sample whereas all phosphorylated histidine residues are immune of modification. Therefore, iodine-labeling locks down the site information of histidine phosphorylation. After removal of all labile phosphates in acid, the iodine-labeled peptides can be easily separated by LC, followed by MS analysis for search of free histidine sites, which represent the original phosphorylated sites. This new methodology has been applied on simple peptide systems as well as complicated protein proteolysis mixtures, and all phosphorylated histidines are successfully identified and even semi-quantified. However, this method does adopt certain limitations, which is also discussed in Chapter 4. Overall, this new method should provide a pathway forward for analyzing histidine phosphorylation in complex systems.

1.4 Examination of protein tertiary structures

The three-dimensional structures of proteins are considered equally important as the primary sequences and PTMs of proteins in regulating protein biological functions. A protein with only primary sequence and PTMs information is like an essay with numerous random sentences, but without assembling the sentences/sequences into nicely ordered paragraphs/structures, it is impossible to read out the correct information the essay/protein wants to deliver. Interpreting how the higher order structures of proteins are related to protein biological functions is the key of structural biology. In the last century, enormous efforts have been contributed to establishing advanced methods for protein structure analysis, of which X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy emerge as two most prolific techniques for examination of protein 3D structures.[36] Both techniques yield high fidelity protein structures in high resolution and accuracy, although X-ray crystallography is considered more as the "gold standard", which constitutes 86% published structures of the protein structure database.[37] However, these two techniques usually require load of large amount protein samples, extremely long experimental time, and highly complicated structure analysis, and therefore leading to intensive labor work.

Beneficial from invention of ESI and MALDI, mass spectrometry has gradually become another important tool for dissecting protein structures.[38] While an intact protein is being ionized by ESI, different numbers of charges will be picked up by protein surface. Finally a charge state distribution (CSD) of the protein is generated, which has been used to interpret the different folding states of proteins.[39,40] MS-based hydrogen-

deuterium (H/D) exchange is another widely used technique with higher resolution for protein structure analysis.[41] The H/D exchange rate in proteins reflects the information of protein structure and dynamics: the protons on the surface of proteins usually exchange with deuterium more rapidly than the ones buried inside. Therefore, it is able to differentiate the protons buried within the hydrophobic core of a protein from the solvent-exposed protons by measuring H/D exchange rate. However, H/D exchange method is only applicable to proteins with relatively slow structure dynamics. In other words, it requires the analyzed proteins adopt stable, well-defined structures in solvents; otherwise fixing protein structure information using H/D exchange is problematic.

MS-based H/D exchange technique is categorized as a covalent labeling method for protein structure characterization. In fact, noncovalent chemistry can also be used to monitor conformational changes in whole proteins. As mentioned in *1.2*, 18C6 can selectively target lysine residues to form stable complexes during ESI process. Actually, 18C6 has been identified as an ideal noncovalent probe for characterization of solution-phase 3D structures of proteins, the technique of which is identified as selective noncovalent adduct protein probing-mass spectrometry (SNAPP-MS).[42] SNAPP-MS monitors the numbers and intensities of 18C6 molecules bound to protein ions at each charge state, which are defined as SNAPP distributions and can be used to interpret protein structures. The generation of various SNAPP distributions at different charge states is due to that a protein samples diverse cross sections at different charge states; therefore SNAPP distributions represent conformations of a protein separated in charge states. When the conformation of a protein is altered by outside triggers, such as addition

of organic contents, ligands, etc., the availability of lysines will be correspondingly changed and result in a variation of SNAPP distributions. This change could also be detected by MS-based H/D exchange; however, SNAPP-MS only requires minimum sample preparation and data analysis.

Another distinct characteristic of SNAPP-MS is that it allows sampling protein conformational distribution in equilibrium since binding of 18C6 to lysines is highly reversible in solution and information is only encoded during the later stages of ESI when the protein is transitioning into the gas phase. This property, as the most competitive feature of SNAPP that neither NMR or other MS-based structural analysis techniques adopts, enables SNAPP with the capability of monitoring conformations of natively disordered proteins, such as α-synuclein. It has been reported that SNAPP is a highly sensitive tool to detect conformation changes of α-synuclein upon metal-ion induced aggregation.[43] SNAPP is even able to differentiate the distinct aggregation behaviors caused by addition of different metal ions (Al^{3+} and Cu^{2+} respectively).

The great application of SNAPP for characterization of structures of natively disordered proteins could be even extended to interrogate the structure dynamics of certain unique proteins which adopt interchanging structures in solutions. Human Lymphotactin (Ltn) is one of the examples. Ltn belongs to the chemokine family and is involved in mediating the regular functions of immune system. Ltn adopts two equilibrium natively folded structures under physiological conditions. The two structures, named as Ltn10 and Ltn40 respectively, are significantly distinct from each other and responsible for independent biological functions.[44] Both structures can interchange with

each other and are present in equilibrium in the native state Ltn, although it is possible to drive one structure completely to the other by adjusting the ionic strength and temperature parameters of Ltn solutions.[45] NMR studies have been carried out to distinguish the two distinct conformations successfully. However, extreme conditions had to be applied in order to achieve one pure structure in solution because NMR is not powerful in resolving mixtures of distinct protein structures in equilibrium. This limitation of NMR makes revealing the interchanging transition state between Ltn10 and Ltn40 nearly impossible, which unfortunately is the most interesting area that structure biologists want to explore. In Chapter 5, SNAPP-MS is employed to characterize the conformations of native Ltn and two Ltn mutants (each mutant has a fixed Ltn10 and Ltn40 structure, respectively). Not surprisingly, SNAPP-MS easily differentiates the distinct structures of the two Ltn mutants, since the two mutants adopt highly dissimilar SNAPP distributions. The Ltn structure dynamics of the interchanging transition state has also been evaluated by SNAPP-MS, which provides valuable information for interpreting Ltn biological functions.

Primary sequence, PTMs and 3D structures constitute three indispensable characteristics of proteins. Interpreting these components in proteins is essential for exploration of the mysteries of biological systems. Mass spectrometry emerges as a powerful tool in encoding different aspects of proteins in combination with other techniques. The continuing development of advanced mass spectrometers as well as MS-based analytical techniques ensures mass spectrometry a promising long-term technology for protein investigation. It is our goal to develop more useful MS-based methods to

contribute to the field of protein study so that the complicated biological systems can be maximally understood.

---

[1] Aebersold, R.; Mann, M. *Nature*, ***2003***, *422*, 198-207.

[2] Edman, P. *Acta Chem. Scand.* ***1950***, *4*, *283-293.*

[3] Shoemaker, D. D.; Linsley, P. S. *Curr. Opin. Microbiol.* ***2002***, *5*, 334–337.

[4] Giaever, G. et al. *Nature* ***2002***, *418*, 387–391.

[5] Giepmans, B. N. G.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y. *Science* ***2006***, *312*, 217-224.

[6] Domon, B.; Aebersold, R. *Science* 2***006***, *312*, 212-217.

[7] Fenn, J.B.; Mann, M.; Meng, C.K.; Wong, S. F. *Science* ***1988***, *246*, 64-71.

[8] Karas, M.; Hillenkamp, F. *Anal. Chem.* ***1988***, *60*, 2299-2301.

[9] Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* ***1984***, *11*, 601.

[10] Shukla, A. K.; Futrell, J. H. *J Mass Spectrom.* ***2000***, *35*, 1069-1090.

[11] Zubarev, R. A.; Kelleher, N. L.; Mclafferty, F. W. *J. Am. Chem. Soc.* ***1998***, *120*, 3265-3266.

[12] Syka, J. E P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J. Hunt, D. F. *Proc. Natl. Acad. Sci. USA, **2004**, 101*, 9528-9533.

[13] Shi, D. H.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; Mclafferty, F. W. *Anal Chem. **2001**, 73*, 19-22.

[14] Chi, A.; Huttenhower, C.; Geer, L.Y.; Coon, J. J.; Syka, J. E. P.; Bai, D. L.; Shabanowitz, J. Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. *Proc, Natl. Acad. Sci. USA **2007**, 104*, 2193-2198..

[15] Brodbelt, J. S.; Wilson, J. *J. Mass Spectrom. Rev. **2009**, 28*, 390-424.

[16] Ly, T.; Julian, R. R. *Angew. Chem. Int. Ed. **2009**, 48*, 7130-7137.

[17] Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom. **1994**, 5*, 976-989.

[18] Wysocki, V. H.; Tsapriailis, G.; Smith, L. L.; Breci, L. A. *J. Mass Spectrom. **2000**, 35*, 1399-1406.

[19] Zhang, Z. *Anal. Chem. **2004**, 76*, 3908-3922.

[20] Zhang, Z. *Anal. Chem. **2005**, 77*, 6364-6377.

[21] Zhang, Z. *Anal. Chem. **2010**, 82*, 1990-2005.

[22] Ly, T.; Julian, R. R. *J. Am. Soc. Chem. **2008**, 130*, 351-358.

[23] Ly, T.; Zhang, X.; Sun, Q.; Moore, B.; Tao, Y.; Julian, R. R. *Chem. Comm.* ***2011****, 47*, 2835-2837.

[24] Sun, Q.; Nelson, H.; Ly, T.; Stoltz, B. M.; Julian, R. R. *J. Proteome Res.* ***2009****, 8*, 958-966.

[25] Julian, R. R.; Beauchamp, J. L. *Int. J. Mass Spectrom.* ***2001****, 210,* 613-623.

[26] Julian, R. R.; Akin, M.; May, J. A.; Stoltz, B. M.; Beauchamp, J. L. *Int. J. Mass Spectrom.* ***2002****, 220*, 87–96.

[27] Wilkins, M. R. *et al. Biotechnology* ***1996****,* 14**,** 61–65.

[28] Krishna, R. G.; Wold, F. *Adv. Enzymol. Relat. Areas Mol. Biol*. ***1993****, 67*, 265-298.

[29] Jensen, O. N. *Nature Rev. Mol. Cell Biol.* ***2006****, 7*, 391-403.

[30] Han, X.; Aslanian, A.; Yates, J. R. *Curr. Opin. Chem. Bio.* ***2008****, 12*, 483-490.

[31] Diedrich, J. K.; Julian, R. R. *J. Am. Chem. Soc.* ***2008****, 130*, 12212-12213.

[32] Santrucek, J.; Strohalm, M.; Kadlcik, V.; Hynek, R.; Kodicek, M. *Biochem. Biophys. Res. Commun.* ***2004****, 323*, 1151-1156.

[33] J. A. Hoch, T. J. Silhavy. (Eds) *Two-component signal transduction.* Washington, DC: ASM Press, 1995.

[34] H. Saito. *Chem. Rev.* ***2001****, 101*, 2497-2509.

[35] A. Sickman, H. E. Meyer. *Proteomics **2001**, 1*, 200-206.

[36] Sali, A.; Glaeser, R.; Earnest, T.; Baumeister, W. *Nature **2003**, 422*, 216-225.

[37] Westbrook, J. et al. *Nucleic Acids Res. **2002**, 30*, 245-248.

[38] Winston, R. L.; Fitzgerald, M. C. *Mass Spectrom. Rev. **1997**, 16*, 165-179.

[39] Chowdhury, S. K.; Katta, V.; Chait, R. T. *J. Am. Chem. Soc, **1990**, 112*, 9012-9013.

[40] Loo, J. A.; Edmonds, C. G.; Udseth, H. R.; Smith, R. D. *Anal. Chem. **1990**, 62*, 693-698.

[41] Wales, T. E.; Engen, J. R. *Mass Spectrom. Rev. **2006**, 25*, 158-170.

[42] Ly, T.; Julian, R. R. *J. Am. Soc. Mass Spectrom. **2006**, 17*, 1209-1215.

[43] Ly, T.; Julian, R. R. *J. Am. Soc. Mass Spectrom. **2008**, 19*, 1663-1672.

[44] Tuinstra, R.L.; Peterson, F.C.; Kutlesa, S.; Elgin, E.S.; Kron, M.A.; Volkman, B.F. *Proc. Natl. Acad. Sci. USA, **2008**, 105*, 5057-5062.

[45] Volkman, B.F.; Liu, T.Y.; Peterson, F.C. *Methods Enzymol. **2009**, 461*, 51-69.

CHAPTER 2

SIDE CHAIN CHEMISTRY MEDIATES BACKBONE FRAGMENTATION IN HYDROGEN DEFICIENT PEPTIDE RADICALS

2.1 Introduction

The development of new methods for fragmenting peptides in the gas phase continues to be a major area of interest in order to extend the utility of mass spectrometry (MS). In particular, radical chemistry has received renewed interest following the discovery of electron capture dissociation and electron transfer dissociation (ECD [1] and ETD [2], respectively), which generate radicals following the addition of an electron to an isolated positively charged even electron ion. The fragmentation obtained by ECD/ETD is desirable, stimulating the investigation of alternative methods for generating radicals on peptides. The most commonly utilized method to date exploits the facile association of many peptides with metal ligands.[3-7] Following collisional activation, the metal departs from the peptide to yield a radical species. This method is promising for several reasons: 1) sample preparation is simple, 2) the radical peptide does not need to be covalently modified, and 3) the observed fragmentation of these peptide radicals differs from collision induced dissociation (CID) experiments. On the other hand, this method can be limited by failure of some peptides to complex with the metal ligand. In addition, activation of the radical requires substantial heating of the molecule. Alternatively, radical precursors can be covalently installed and then activated by CID. [8 - 11] Peroxycarbamate, azo, and nitroso functional groups have been successfully employed in this type of experiment, where MS^3 is required to examine fragmentation of the peptide

radical. Again, the results can differ substantially from those observed by CID; however, prior chemical modification complicates both sample preparation and analysis of the results, and again activation by CID leads to heating of the molecule prior to generation of the radical.

Recently, we have been developing another approach for generating radicals by photolysis of highly labile radical precursors. For example, iodinated tyrosine residues lose I· exclusively upon absorption of a 266nm photon, yielding a tyrosyl radical by homolytic cleavage.[12] This chemistry occurs due to the presence of a low lying dissociative excited electronic state, and results in only minimal heating of the remaining molecule. In this manner a radical can be selectively generated at a specific site in a large molecule, without perturbing the remainder of the molecule. In the case of iodotyrosine, subsequent collisional heating of the radical results in fragmentation at tyrosine and other nearby residues. Similar chemistry has been observed following modification of phosphorylated serine and threonine, which utilizes direct dissociation of a carbon-sulfur bond.[13] However, this methodology has not previously been explored with noncovalently attached radical precursors, which is the subject of the present report.

There are a variety of potential scaffolds for delivering chemical functional groups to peptides via noncovalent interactions.[14-17] Arguably, 18-crown-6 (18C6) is the preferred solution due to its ability to recognize protonated primary amines.[18] Any lysine containing peptide is therefore an excellent host for 18C6, although a protonated N-terminus or even arginine will suffice under mild sampling conditions (particularly if lysine is absent). Therefore 18C6 can potentially attach to the vast majority of peptides

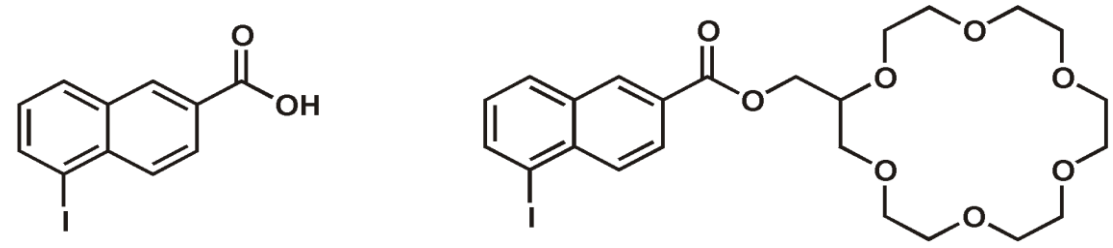
that might be produced by enzymatic digestion in a typical proteomics experiment. In addition, the 18C6 scaffold is amenable to derivatization, yielding lariat crown ethers where a large number of possible functional groups can be attached to the crown. 18C6 based lariat crowns have already been successfully employed to deliver functional groups via noncovalent attachment in previous gas phase experiments with peptides. [14,19] For the present work, it is additionally advantageous that 18C6 does not absorb in the near ultraviolet and will therefore not interfere with photochemistry carried out in this region.

We have combined the recognition capabilities of 18C6 with a photolabile iodonaphthyl radical precursor in the form of a lariat crown ether. Sample preparation involves simple addition of the lariat crown to the solution containing the peptide. A noncovalent complex then forms which can be efficiently transferred to the gas phase by electrospray ionization. Photoexcitation at 266nm leads to loss of I·, generating a naphthyl radical which readily abstracts a hydrogen atom from the peptide. Subsequent collisional activation leads to loss of the crown and fragmentation of the radical peptide. It is revealed that dissociation of the peptide is dominated by radical chemistry and surprisingly occurs through strongly favored dissociation channels. In some cases, highly preferential cleavage at aromatic residues is observed. Ab initio calculations suggest that selectivity is controlled by the relative β-hydrogen bond dissociation energies for each amino acid. In addition, prominent side chain losses are observed. Mechanisms for the production of backbone and side chain fragments are proposed, and the utility of side chain losses in peptide identification is explored.

2.2 Experimental Methods

2.2.1 Synthesis of radical initiators

5-I*odo*-2-*naphthoic acid* (***1***)—[20] To a round bottom flask charged with water (1 mL) and silver sulphate (500 mg, 1.6 mmol, 1.3 equiv) was added sulfuric acid (20 mL). The mixture was allowed to stir and subsequently cooled to room temperature. 2-Naphthoic acid (430 mg, 2.5 mmol, 1 equiv) was added, followed by the addition of iodine (750 mg, 3.0 mmol, 1.15 equiv). The reaction was stirred vigorously at room temperature for 2 hours. Carbon tetrachloride (20 mL) was added and the reaction was allowed to stir for an additional hour. The crude reaction mix was slowly added to water (150 mL), and the yellow precipitate was collected by vacuum filtration. The precipitate was taken up in a 10% potassium hydroxide solution (100 mL) and filtered through activated carbon. The resulting colorless solution was acidified to pH 2 with 1 N HCl. The precipitate was recrystalized form EtOAc to yield 200 mg of white crystals.



. **1** **2**

Scheme 2.1 (1) 5-Iodo-2-naphthoic acid (2) 5-Iodo-2-(2-hydroxymethyl-18-crown-6)-naphthoate

*5-Iodo-2-(2-hydroxymethyl-18-crown-6)-naphthoate (**2**)* — To a flame dried 1 dram vial charged with dichloromethane (1.2 mL) was added 5-iodo-2-napthoic acid (74 mg, 0.25 mmol, 1 equiv) and oxalyl chloride (26.4 μL, 0.30 mmol, 1.2 equiv). The mixture

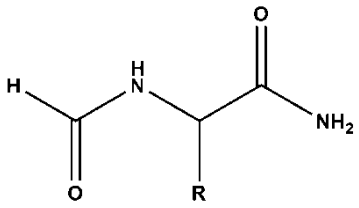
was stirred for 30 minutes at ambient temperature. A catalytic quantity of DMF (1 μL) was added, and the reaction mixture was stirred for an additional 3 hours. The solvent was removed by rotary evaporation to yield the crude acid chloride as a yellow solid. To the crude yellow solid was added dichloromethane (1.2 mL), and 2-hydroxymethyl-18-crown-6 (73.5 mg, 0.25 mmol, 1 equiv) and the reaction mixture was stirred for 30 minutes at room temperature. Triethylamine (70 μL, 0.30 mmol, 1.2 equiv) was added to the reaction mixture before it was heated to 40 ℃ for 4 hours. The reaction was cooled to room temperature, and condensed by rotary evaporation. Diethyl ether (5 mL) was added, and the resulting suspension was filtered through a celite column. After condensation by rotary evaporation, 115 mg of yellow oil was isolated.

2.2.2 Sample Preparation

Peptides (Ac-AKAKAKAY-OH, AAAYGGFL, AEAEYEK, RGYALG, DRVYIHPF, Ac-ERERERER-NH2, RYLPT, RYLGYL, RPPGFSPFR, AAGMGFFGAR, TRSAW, MEHFRWG, GFQEAYRRFYGPV, RRPWIL, DLWQK, YGGFLRK, KWDNQ, GGYR, YAFEVVG, KKPYIL, SLRRSSCFGGR, AAAKAAA, AAAAKAAAK, EMPFPK, VLPVPQK, AVPYPQR) were purchased from American Peptide Company (Sunnyvale, CA), Quality Controlled Biochemicals (Hopkinton, MA), or Sigma-Aldrich (St. Louis, MO). All chemicals and reagents in this work were used directly without purification unless noted. Sample solutions for electrospray were made by mixing one equivalent of each peptide stock solution with four equivalents of iodo-naphthyl crown in 98% acetonitrile to give final concentrations of 10μM and 40μM, respectively.

2.2.3 Mass Spectrometry

A flashlamp-pumped Nd:YAG laser (Continuum Minilite, Santa Clara, CA) was interfaced with the back of an LTQ linear ion trap mass spectrometer (Thermo Electron, San Jose, CA). Fourth-harmonic laser pulses (266 nm) were introduced to the linear ion trap through a quartz window at the posterior of the ion trap. Peptide-crown solutions were infused into a standard electrospray source and transmitted into the linear ion trap. MS^n type experiments (where the first CID step was replaced by photodissociation) were performed on noncovalent peptide-crown complex ions. Laser pulses were synchronized by feeding a TTL trigger signal from the mass spectrometer to the laser via a digital delay generator (Berkeley Nucleonics, San Rafael, CA). All CID steps were employed by applying an excitation voltage on mass-isolated ions using default instrument parameters. Peptide fragments were assigned with the aid of UCSF Protein prospector (http://prospector.ucsf.edu/).



Scheme 2.2 Model Structure for Calculations (R=all amino acid side chains)

2.2.4 Calculation

All *ab initio* calculations were performed at the B3LYP/6-31G(d) level as implemented in Gaussian 03 Version 6.1 Revision D.01. Model peptides (as shown above) were built using Gauss View 3.0. C-H bond dissociation energies (BDEs) for the β hydrogens of 19 amino acids were obtained by the use of isodesmic reactions, which

have been used previously to evaluate α C-H BDEs.[21,22] The reference molecule in each isodesmic reaction was as follows: H-CH_2OH (Ser, Thr), $C_6H_5CH_2$-H (Phe, Trp, His, Tyr), CH_3CH_2-H (Ala), $(CH_3)_2CH$-H (all others). Trans geometry was assumed for all peptides, except for proline, where both conformers were considered. Restricted and unrestricted methods (RB3LYP and UB3LYP) were used on closed-shell and open-shell systems respectively. Spin contamination was found to be minimal for all systems. The results are in good agreement with previous calculations.[23] The relevant computational numbers for the reference molecules were obtained from the NIST computational standards database.[24]

2.3 Results and Discussion

Reagent ***2*** is designed to complex with peptides in solution via the 18C6 portion of the molecule. The preferred target is lysine; however, under appropriately gentle electrospray conditions ***2*** will complex with virtually any protonated peptide. The iodonaphthyl portion of ***2*** is designed to yield a naphthyl radical following photodissociation of the carbon-iodine bond. Such bonds are known to fragment directly following electronic excitation by ultraviolet radiation.[25] Loss of I· from ***2*** generates a highly reactive radical in close proximity to a complexed peptide, which can lead to hydrogen abstraction and the generation of a hydrogen deficient peptide radical. To clarify, by hydrogen deficient we mean a peptide missing a hydrogen relative to the mass of a fully protonated, even electron ion (with no implication for the structure).[26] The hydrogen count distinguishes these radicals from those produced by ECD or ETD, which contain an additional

hydrogen relative to a protonated ion. Conveniently, complexes of ***2*** and the target peptide are created by simply mixing the two reagents and then electrospraying the solution. The addition of crown ethers also has beneficial effects on total ion counts as described previously.[18]

The full mass spectrum generated by electrospray ionization of a solution of ***2*** and RGYALG is shown in Figure 2.1a. Abundant adduct formation is observed, in agreement with previous observations. It should be noted that this peptide does not contain lysine, confirming that non-lysine peptides can interact with 18C6 under proper source conditions. Isolation of the peak corresponding to $[RGYALG+\boldsymbol{2}+H]^+$ followed by photoactivation at 266nm is shown in Figure 2.1b. There are three products observed. Direct dissociation of the carbon iodine bond leads to loss of I· with retention of the noncovalent complex. Simple dissociation of the noncovalent complex is also observed; however upon closer inspection, it is revealed in the zoomed in spectrum that a fraction of the protonated peptide has undergone hydrogen abstraction *and* fracture of the noncovalent complex. CID experiments on the peptide radical can be performed by re-isolating the directly generated peptide radical shown in Figure 2.1b; however, it is often simpler to generate the radical peptide by subjecting the noncovalent radical complex to CID as shown in Figure 2.1c. This pathway typically yields the peptide radical as the primary product. The photodissociation step in these experiments can be conducted on the nanosecond timescale, and the remaining steps are CID experiments, meaning that interrogation of radical peptides could be performed in tandem with CID experiments on the nonradical form of the peptide, if desired.

Sufficient radical peptide is produced in these experiments to carry out multiple subsequent steps of collisional activation; however, the overall efficiency of the process has not been maximized. The usable radical ion count is frequently only 10-20% of the precursor peptide intensity. It is likely that hydrogen abstraction efficiency is currently limited by three unproductive channels. First, internal conversion of the photon energy can occur and will lead to simple heating of the complex without generating any radicals. Second, any luminescence will similarly lead to unproductive deactivation. It may be possible to reduce these channels through optimization of the chromophore. Third, generation of the radical may be followed by fracture of the noncovalent bond holding the complex together without prior abstraction of hydrogen from the peptide. It is possible that further optimization of the radical delivery agent could reduce this phenomenon; however, the current reagent is well suited for experiments which are not high throughput and will suffice to explore the relevant chemistry for potentially interesting fragmentation.

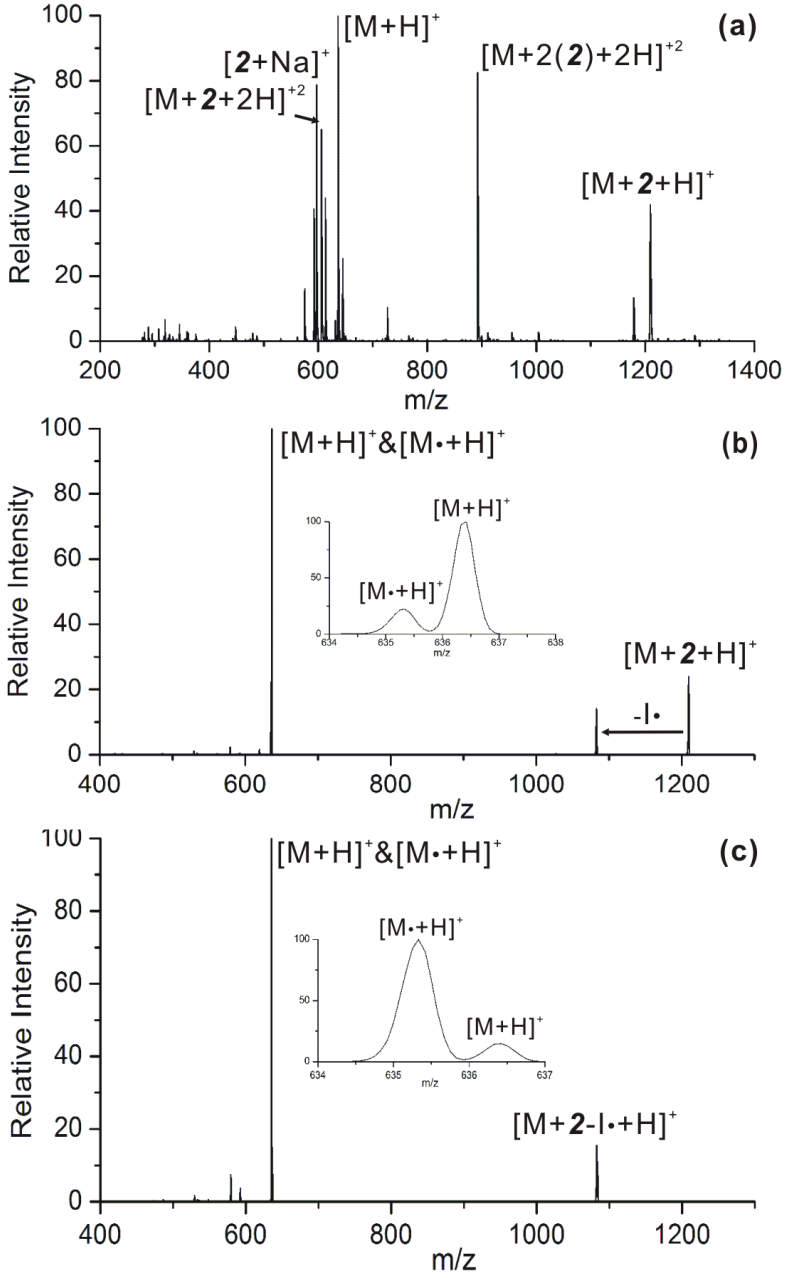


Figure 2.1 (a) Full mass spectrum of ***2*** and RGYALG, demonstrating abundant complex formation. (b) Photoactivation at 266nm leads to generation of a radical complex, and both radical and nonradical peptides. (c) Collisional activation of the radical complex yields primarily radical peptide.

The results obtained by fragmenting $[RGYALG\cdot+H]^+$ are shown in Figure 2.2a. There is only one major cleavage along the peptide backbone, yielding an a_3 ion as the most abundant peak in the spectrum. A minor a_5 fragment is also observed, in addition to several side chain losses. For comparison, the CID spectrum for the protonated peptide is shown in Figure 2.2b. Following CID, the loss of ammonia dominates and few backbone fragments are observed. No backbone dissociation at Tyr3 is detected. In fact, there is very little overlap between any of the fragments observed in the two spectra. This suggests that the radical is an active (and perhaps dominant) participant in the chemistry yielding the fragments in Figure 2.2a. The observation of a-type ions in this spectrum is consistent with previous experiments involving radicals.[12 27] The loss of 106 Da corresponds to elimination of the tyrosine side chain, which has also been observed previously.[3,28] In addition, there are two side chain losses from leucine, corresponding to the loss of 43 and 56 Da. The importance of these losses will be discussed further below.

Following the observation of preferential cleavage at tyrosine, it was hypothesized that aromatic residues might facilitate backbone fragmentation. This possibility was explored further with the peptides RYLGYL and RRPWIL. Fragmentation of $[RYLGYL\cdot+H]^+$ is shown in Figure 2.2c. For this peptide, abundant backbone fragmentation occurs at both tyrosine residues yielding dominant a_2 and a_5 fragments, confirming that tyrosine does facilitate backbone fragmentation. In addition, modest backbone dissociation at leucine yielding an a_3 ion is observed (this fragment is reminiscent of the a_5 fragment at leucine in Figure 2.2a). Importantly, no other backbone fragmentations are observed. Side chain losses similar to those in Figure 2.2a are noted, although the relative intensities are

substantially different. $[RRPWIL\cdot+H]^+$ which contains tryptophan, was fragmented as shown in Figure 2.2d. Again, a single dominant backbone fragment corresponding to cleavage at the tryptophan residue to yield an a_4 ion is observed. The only other backbone cleavage is minor and occurs at isoleucine to yield an a_5 fragment. Interestingly, arginine side chain loss (-86Da) is prevalent for this peptide, which was not observed previously despite the presence of arginine in the previous two peptides (RGYALG and RYLGYL). In addition, there are minor peaks corresponding to the loss of side chains from leucine, isoleucine, and tryptophan.

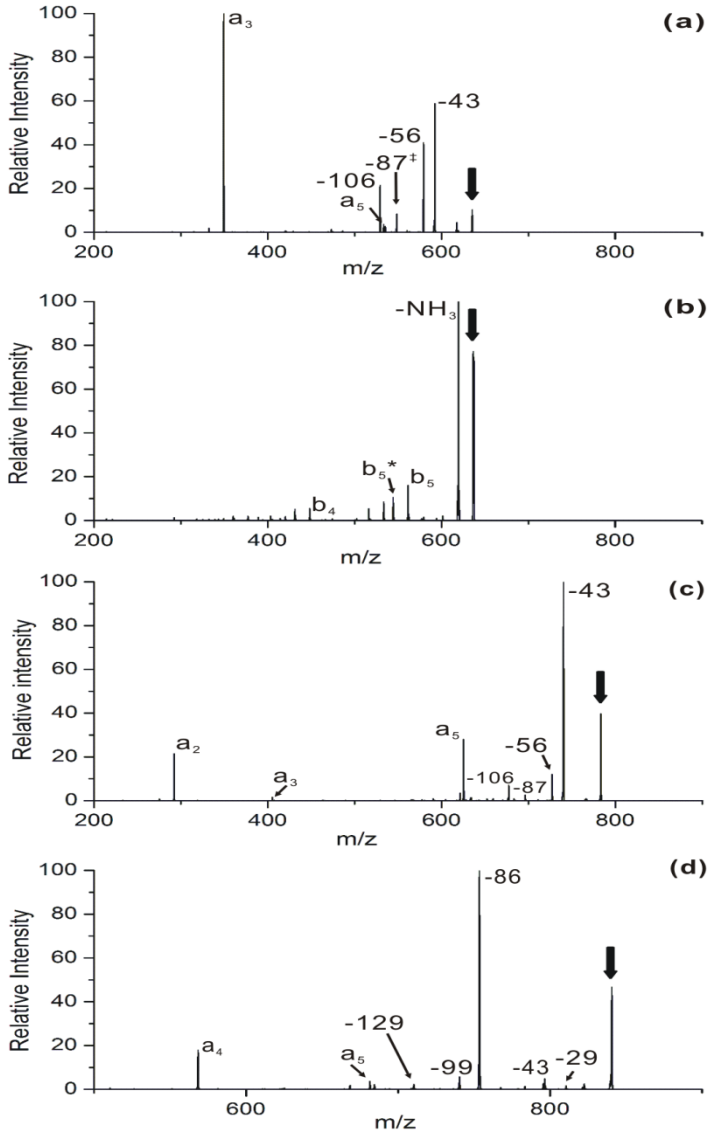


Figure 2.2 (a) Fragmentation of $[RGYALG\cdot+H]^+$ results in backbone cleavage at tyrosine. (b) By comparison, CID is dominated by loss of NH_3. (c) Fragmentation of $[RYLGYL\cdot+H]^+$ also results in backbone fragmentation at tyrosine. (d) Fragmentation of $[RRPWIL\cdot+H]^+$ yields backbone dissociation at tryptophan. Bold down arrows indicate ions subject to activation. ‡ This peak results from consecutive losses of 43 and 44. * loss of NH_3.

The fragmentation spectrum for bradykinin $[RPPGFSPFR\cdot+2H]^{2+}$, which contains two phenylalanines, is shown in Figure 2.3a. There are two dominant a-type fragments, a_5 and a_8, which correspond to cleavage at the phenylalanine residues. This confirms that aromatic residues facilitate fragmentation of the peptide backbone in general. Complete examination of the spectrum reveals that minor a-type fragments are also observed at Pro2 and Ser6. There is also a novel cleavage between phenylalanine and serine, yielding c_5 and z_4 ions. This observation relates to the presence of serine as is discussed further below. In addition, side chain losses from arginine are prominent, and several low intensity y-type ions are observed. Some of the y ions are radicals. The results from bradykinin suggest that increased sequence diversity can lead to a larger number of fragmentation pathways; however, comparison with the CID spectrum for the nonradical peptide[29] reveals small overlap suggesting that most of the fragments are generated by radical chemistry.

The spectrum shown in Figure 2.3b illustrates fragmentation of radical angiotensin II, $[DRVYIHPF\cdot+2H]^{2+}$. In this case an extensive series of a-type fragments is observed. Comparison with CID of $[DRVYIHPF+2H]^{2+}$ reveals that a-type ions are not generated in abundance by collisional activation in the absence of a radical.[29] Fragmentation at Tyr4 to yield a_4 is favorable and yields the largest backbone fragment. Abundant a_3 and a_6 ions, corresponding to cleavage at Val3 and His6 are also observed. Minor a_5 and a_7 ions are produced as well. Overall, the dominant fragmentation channels for this peptide are side chain losses from tyrosine (-106) and aspartic acid (-44). The loss of CO_2 could also theoretically occur from the C-terminus, but CID of the -44 fragment confirms that this

loss occurs primarily from aspartic acid.[29] The loss of CO_2 has been observed almost exclusively in previous experiments utilizing photoactivation of benzophenone.[30]

The results for fragmenting $[MEHFRWG\cdot+2H]^{2+}$ are shown in Figure 2.3c. This peptide contains several aromatic residues and methionine. As expected due to the presence of sulfur, fragmentation of the methionine side chain is favorable, yielding intense losses of 61 and 74. Atypically, there is also an abundant y_5 ion. This fragment is also dominant in the CID spectrum of the non-radical peptide indicating that cleavage of EH peptide bond is facile and not initiated by a radical. Significant a-type fragments are generated at Phe4 and Trp6, in agreement with previous results. In addition, fragmentation at His3 yields an abundant a_3 ion. There are also several z-type ions, which are discussed in greater detail below. In general, the fragmentation of this peptide follows similar trends to those observed in previous examples above.

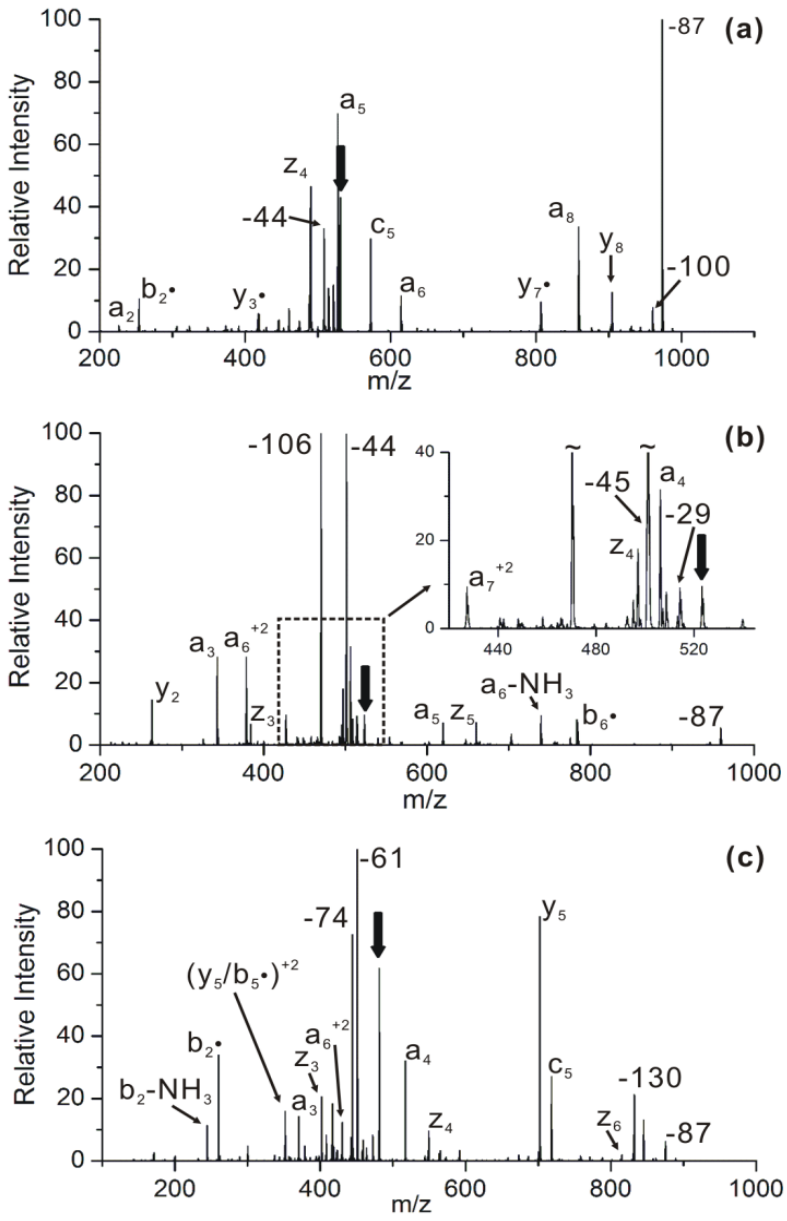
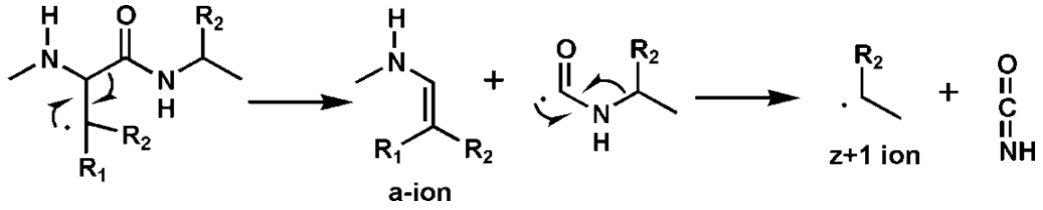


Figure 2.3 (a) Fragmentation of $[RPPGFSPFR\cdot+2H]^{2+}$ yields several fragments, including abundant fragmentation at both phenylalanine residues. (b) Fragmentation of $[DRVYIHPF\cdot+2H]^{2+}$. A nearly complete series of a-type ions is observed. (c) Fragmentation of $[MEHFRWG\cdot+2H]^{2+}$. Abundant side chain losses at methionine and several backbone fragments are observed.

2.3.1 Backbone Fragmentation

Backbone dissociation to yield a-type fragments occurs primarily following abstraction of a β-hydrogen as shown in Scheme 2.3. Consequently, a-type ions are never generated at glycine. Furthermore, the charge does not participate or affect the reaction. A similar pathway can also lead to the generation c and z-type ions, but this pathway is usually kinetically disfavored by several orders of magnitude according to theory.[31] Experimentally, we have determined that fragmentation at serine and threonine are the only exceptions. These residues generate c and z ions.[29] Importantly, regardless of whether c and z or a and x ions are produced, backbone fragmentation is mediated by initial abstraction of the β-hydrogen for all residues (except glycine).

The nature of the z-ions observed in these experiments requires further clarification. Most of the z ions assigned herein are actually z+1 ions (the same z-ion typically observed in ECD); however, true z ions would be generated by β-scission.[8] The presumed origin for most of the z+1 ions relates to the x-type fragments from Scheme 2.3. The x-type ions shown in Scheme 2.3 are unstable radical species which are never directly observed. However, it is possible for the x-ion to be stabilized by loss of isocyanic acid, yielding a z+1 ion. Indeed, there are several z+1 ions that are the complement ions to the abundant a-type fragments present in Figure 2.3. Nevertheless, z+1 ions are not always observed and are typically found in small abundance. One explanation for this is that an alternate pathway proceeding by loss of CO followed by loss of the proximate residue can yield an x_{n-1} radical fragment. This route can be repeated indefinitely and possibly lead to degradation of the entire C-terminal radical fragment.



Scheme 2.3 Mechanism of a- and z+1 type ions generated by CID of peptide radicals

Close examination of the results presented thus far reveals that backbone fragmentation appears to be favored at certain residues, as is most clearly illustrated in the case of aromatic side chains. Is there a rationale which can be used to account for this behavior? Selective fragmentation can only occur if some reaction routes are more favorable than others. For backbone fragmentation, selectivity should result if the initial β-radical in Scheme 2.3 is preferentially formed at particular amino acids. In order to understand how this occurs, both the kinetics and dynamics of the reactions leading to the formation of these radicals must be examined. The kinetics are relatively straightforward, where barriers to hydrogen abstraction in amino acids have been calculated, they are small or nonexistent in the absence of conformational barriers.[32 - 34] Conformational restraint can introduce barriers.[35] Therefore, if a reaction is thermodynamically favorable and the reactants can be brought together without steric hindrance, it should proceed spontaneously. Comparison of the thermodynamics is easily afforded by calculating the BDEs for removing hydrogen from each of the relevant species. The BDE for the α-naphthyl hydrogen is approximately 475 kJ/mol,[36] which is quite large and should enable the naphthyl radical to abstract hydrogen from virtually every site in a peptide. Indiscriminate reactivity is usually not a recipe for selectivity!

The key to understanding the results is that following the initial hydrogen abstraction, the next thermodynamically favored reaction can also proceed without barrier. In this manner, the radical site can "flow" downhill to a thermodynamically favored location, assuming that conformational flexibility of the peptide allows interaction between the relevant reaction sites. The calculated BDE values for abstracting the α and β hydrogens from all of the amino acids are shown in Figure 2.4 along with estimated uncertainties. The α hydrogens are always easier to abstract due to captodative stabilization; however, the BDEs for α hydrogens and the magnitude of the captodative effect are very sensitive to peptide structure.[37] If the nascent radical cannot adopt a planar configuration or favorable dihedral angles, the BDE can increase by as much as 40 kj/mol.[21] Structural effects can therefore cause the α and β BDEs to overlap for the lower β BDEs. Indeed, the aromatic residues which have the lowest β BDEs and should be among the most competitive sites for β-radical formation are experimentally observed to be the most favored sites for backbone dissociation. In contrast, amino acids with high β BDEs , such as methionine or alanine, never yield abundant a-type ions. In simplified terms, the amino acids on the left side of Figure 2.4 will yield both side chain and backbone fragments while the amino acids on the right side will primarily produce side chain losses. The relative intensity of backbone fragments is also frequently correlated with position in Figure 2.4, with those amino acids more to the left yielding more abundant backbone fragments. For example, the relative intensities of the a-type ions in Figure 2.2 follow the predicted trends, with fragments at tyrosine or tryptophan being favored over those at leucine or isoleucine.

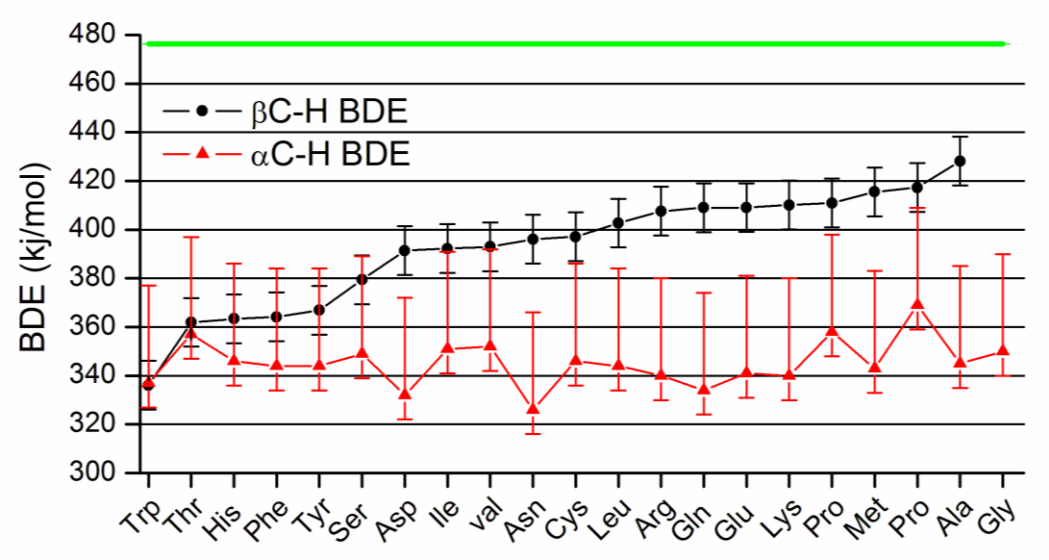


Figure 2.4 The calculated α and β BDEs for each amino acid are shown. The horizontal green line indicates the BDE for naphthalene. The trend for lower BDE in aromatic residues correlates well with the observation of more intense backbone fragmentation. See text for discussion. [a] α BDEs taken from ref [21] [c] cis [t] trans

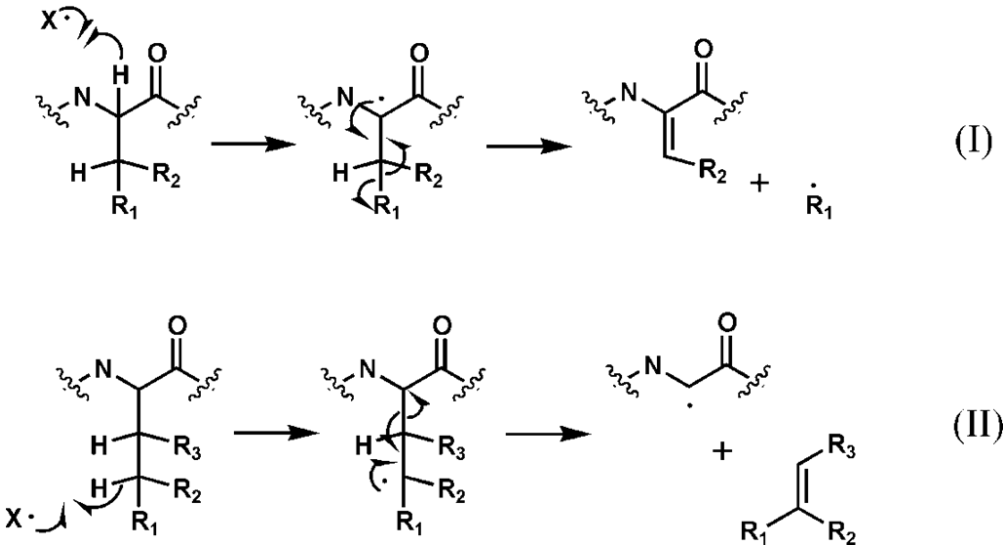
Very importantly, the magnitudes of the β BDEs themselves are dictated by side chain composition, meaning that fragmentation of the peptide backbone in these experiments is controlled primarily by side chain chemistry. This contrasts sharply with other currently available dissociation methods. CID fragmentation is controlled largely by the presence or absence of mobile protons.[38] The influence of side chains is highly variable in CID fragmentation. For example, the hydrocarbon side chains (alanine, valine, leucine, isoleucine, phenylalanine) have little influence on CID, yet fragmentation at these residues can occur. On the other hand, aspartic acid and proline actually facilitate backbone cleavage, but fragmentation does not always occur at these residues for all peptides.[39, 40] In the final analysis, CID can only be predicted with bioinformatic

approaches, and even then prediction is difficult and will be dependent on the overlap between training and data sets.[41] Fragmentation of the backbone by ECD and ETD is well known to occur without regard for side chains or even post-translational modifications.[42] The fact that radical mediated fragmentation yields primarily a-type ions where the abundance of each backbone cleavage is correlated with side chain composition is unique and should enable greater predictability of the results.

2.3.2 Side Chain Losses

It is clear from the experimental data that side chain losses play an important role in the chemistry of radical peptide dissociation. Previous experiments have demonstrated that similar side chains losses can occur in ECD experiments.[43] In contrast to backbone fragmentation, side chain losses are initiated by a diverse set of pathways.[44] Perhaps the two most common pathways are shown in Scheme 2.4. Mechanism (I) illustrates abstraction of the α-hydrogen to generate dehydroalanine, which leads to loss of a radical side chain fragment. Abstraction of the γ-hydrogen, as shown in mechanism (II), leaves the radical on the peptide in the α-position following loss of the entire side chain. Other mechanisms can mimic the results produced by (II). For example, abstraction of the tyrosine side chain -OH hydrogen followed by electronic rearrangement and loss of p-quinomethide yields the same peptide product as would be generated by (II). The predicted and observed side chain losses for all amino acids are shown below in Table 1. Data interpretation is somewhat simplified by the fact that not all amino acids will exhibit side chain losses. For example, glycine has no side chain. Alanine and proline cannot lose a side chain by mechanisms (I) or (II) and are not predicted to lose a side chain by any

other energetically competitive route. Phenylalanine and valine side chain loss by either mechanism is unlikely because aromatic or primary carbon radicals are involved. In fact, side chains losses are not observed for glycine, alanine, proline, phenylalanine, or valine to any significant extent. Thus, one quarter of amino acids are typically not subject to side chain losses.



Scheme 2.4 Two common side chain loss pathways in radical directed dissociation: hydrogen abstraction from (I) Cα-Hs of peptide backbones; (II) Cγ-Hs of peptide side chains.

The propensity for observing each side chain loss from a collection of peptides is shown along with the average values in Figure 2.5. The propensity for a loss is defined as the abundance of a loss relative to the largest side chain loss divided by the number of times that amino acid occurs in the peptide. If no side chain loss for an amino acid in a peptide which contains that amino acid is observed, the propensity is zero. The maximum value would be 100 for the most abundant side chain loss observed in a peptide with only

one occurrence of that amino acid. For several amino acids, the propensity for side chain loss ranges from 0 to 100, covering the entire range of possible values. The propensities are fairly scattered in general. This indicates that neither the occurrence nor relative intensity of side chain losses are easily predictable for a given peptide of unknown sequence. This may relate to the variety of pathways which lead to side chain losses. Consideration of the average propensities (shown in dark triangles) is more revealing. For each amino acid, the average loss from mechanism (I) occurs in greater abundance than the average loss from mechanism (II). This trend is in perfect agreement with the results that would be expected based on the relative BDEs shown in Figure 2.4. Furthermore, it is clear that (on average) some side chains are more reactive than others. One striking example is the greater fragmentation propensity of leucine compared to isoleucine.

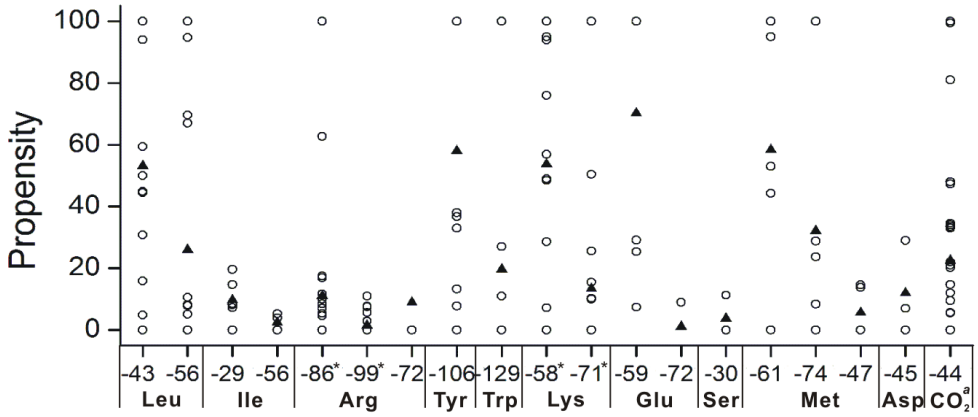


Figure 2.5 Propensities for side chain losses at each amino acid are given by mass. The propensities do not follow obvious trends and frequently fill the entire scale. Average values for each column are indicated by black triangles. [a] loss from Glu, Asp, and the C-terminus [*] protonated or neutral losses.

2.3.3 Utility of Side Chains Losses

Despite the fact that side chain losses are unpredictable, they are easily identified and can still yield important information. For example, leucine and isoleucine can be distinguished by side chain losses because unique mass fragments are generated by mechanism (II).[45] Furthermore, any unique mass side chain loss positively identifies one amino acid in the peptide (although its location is obviously not revealed). There are 11 amino acids which can be uniquely identified by side chain loss. The remaining 4 amino acids can be narrowed down to 2 or 3 possibilities. Due to the infrequency for which complete series backbone fragments are generated by CID, the identities of many amino acids in a peptide are not directly available. Of the amino acids in the peptides studied presently, side chain losses led to the positive identification of amino acids for 10/15 of the peptides for which the identity was not apparent from backbone dissociations. When submitting MS/MS information to a database, any hits not containing these amino acids could be eliminated.

Table 2.1 Side Chain Loss Masses by Mechanism.

Amino acids	Side chain losses			
	I	**II**	**III** [a]	**IV** [b]
Ala[c]	--	--		
Arg	86, 87[d]	99, 100[d]		72
Asn	44	--		
Asp	45	--		44
Cys	33	46		
Glu	59	72		45, 44
Gln	58	71		
Gly	--	--		
His	67	--	80	
Ile	29	56		
Leu	43	56		
Lys	58, 59[d]	71, 72[d]		
Met	61	74		47
Phe	*77*	--		
Pro	--	--		
Ser	17[e]	30		
Thr	15[e]	44		
Trp	116[e]	--	129, 130[d]	
Tyr	93[e]	--	106	
Val	*15*	*42*		

a. diverse mechanisms which yield the same peptide fragment as in (II).

b. various other experimentally observed side chain losses.

c. amino acids in italics do not exhibit any substantial side chain losses

d. these losses are protonated side chains

e. never observed

2.3.4 Influence of Charge State

Charge state does not typically play a dominant role in the fragmentation observed for the radical peptides herein. For example, doubly protonated $[RRPWIL\cdot+2H]^{2+}$ yields the fragmentation spectrum shown in Figure 2.6a. For the most part, similar fragments are generated as those observed in Figure 2.2d for the singly protonated peptide. The primary difference between the spectra is that the relative intensities of several fragments change significantly. Interestingly, the a_4 ion becomes the base peak and arginine side chain loss is suppressed. Also interesting is the increased abundance of the loss of leucine side chain (-43). The relative intensities of the a-type fragments with respect to each other have not changed, which is consistent with our description of backbone fragmentation above. Overall, these results suggest again that fragmentation is primarily controlled by radical rather than charge directed processes, in agreement with previous findings examining energetics explicitly.[27]

One situation where charge state can play a significant role is when a single charge is held near the C-terminal portion of the peptide. This is demonstrated by the fragmentation of $[MEHFRWG\cdot+H]^{+}$,[29] which generates fewer backbone fragments than $[MEHFRWG\cdot+2H]^{2+}$. This is most likely due to the fact that the singly protonated molecule will be preferentially charged at arginine, preventing observation of a-type fragments smaller than a_5. In order to avoid this situation, higher charge states should be interrogated when multiple options exist.

2.3.5 Influence of Peptide Structure

For the majority of peptides examined, the three dimensional structure (although presumed to be present) does not significantly influence backbone dissociation. However, we did encounter one striking example to the contrary. The peptide $[KKPYIL\cdot+2H]^{2+}$ yields a typical dissociation spectrum when generated from the +3 noncovalent complex as shown in Figure 2.6b. In this case, the crown departs with a charge following hydrogen abstraction, yielding the doubly protonated peptide radical. The results contrast sharply with what happens when the +2 radical peptide is generated from the doubly charged noncovalent complex as shown in Figure 2.4c. In this case, b and y ions dominate the spectrum. Radical initiated fragments are also present, but the relative intensities have changed dramatically between Figure 2.6b and 2.6c for fragmentation of the same ion! The most likely explanation in this case is that the +2 complex is situated such that the initial radical site on the peptide is already stable and the does not migrate to other locations. This leads to abundant fragments that are similar to CID of the nonradical.[29] In fact, assignment of these peaks localizes the radical to the first two residues, as the fragments containing these amino acids are radicals. The +3 complex will undoubtedly have a different and probably more extended structure, making it unlikely that the original site of abstraction is the same. This is the only peptide which is observed to behave this way. The preparation of the ion did not greatly affect the results for the other peptides (not all can be prepared by multiple paths). Briefly, structure can affect the results, but most of the time this will be manifested in variations in relative intensity rather than in the appearance or absence of fragments.

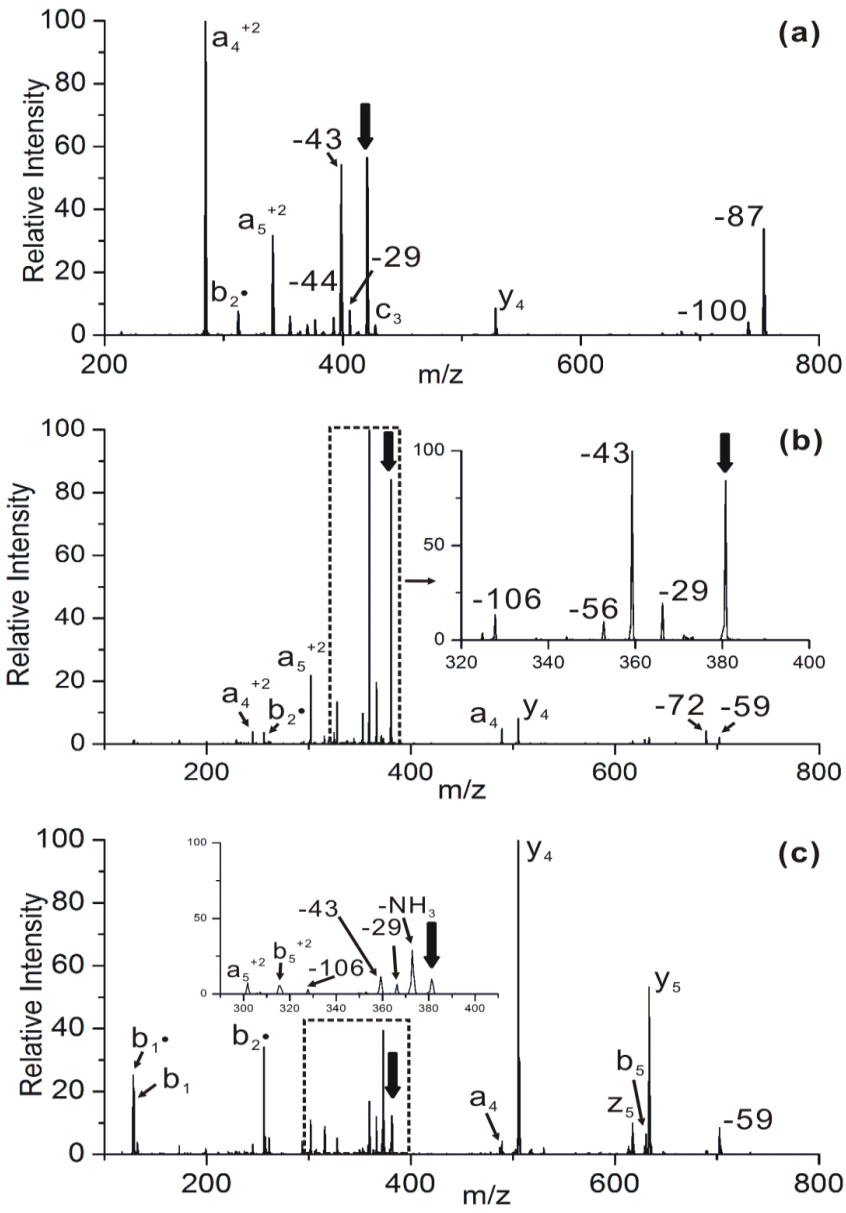


Figure 2.6 (a) Fragmentation of $[RRPWIL\cdot+2H]^{2+}$ yields a spectrum similar to the singly protonated radical. (b) Fragmentation of $[KKPYIL\cdot+2H]^{2+}$ prepared from $[KKPYIL+\boldsymbol{2}+2H]^{2+}$ revealing a typical spectrum. (c) Fragmentation of $[KKPYIL\cdot+2H]^{2+}$ prepared from $[KKPYIL+\boldsymbol{2}+3H]^{3+}$ yields a very dissimilar spectrum, revealing the influence of structural effects. This peptide is unique in this behavior.

2.3.6 Radical Transfer

Side chain losses can also be used to explore the possibility for radical transfer or migration.[35] The spectra shown in Figure 2.7 illustrate what happens when several of the side chain losses in Figure 2.2a were re-isolated and subjected to further collisional activation. In Figure 2.7a, the product resulting from tyrosine side chain loss was subjected to CID. Loss of the tyrosine side chain leaves the radical initially on the α-carbon. As shown in Figure 2.7a, activation of this radical leads to dominant side chain losses from leucine. Minor loss of CO_2 is also observed in conjunction with other losses. Fragmentation of the backbone to yield the a_5-106 peak represents a minor dissociation channel. In the absence of more competitive dissociation pathways, fragmentation at Ala4 to yield the a_4-106 ion is even observed in very minor abundance. These results suggest that radical movement is primarily in the thermodynamically favored direction, as discussed above. In Figure 2.7b the peak corresponding to the -56 Da leucine side chain loss, which will again generate a radical on the α-carbon, is subjected to further collisional activation. Interestingly, the a_3 ion is by far the most abundant fragment that is generated, followed by the loss of CO_2 and tyrosine side chain. This confirms that following radical migration to an α position, abstraction of a β-hydrogen from an aromatic residue is still a facile process. Furthermore, no appreciable fragmentation at Ala4 is observed, as was the case in Figure 2.7a. Abstraction of the beta hydrogen from alanine is simply not competitive in the presence of tyrosine, again in agreement with the predicted chemistry. In Figure 2.7c, further activation of the -43 Da leucine side chain loss is shown. In this situation, a radical fragment is lost, meaning that the peptide is not a

radical. Although the spectrum in Figure 2.7c does contain some a-type fragments not observed in the CID spectrum for the protonated peptide (see Figure 2.2b), in general fragmentation is similar in both spectra.

These results demonstrate that radical migration is possible. For example, if the initially formed naphthyl radical abstracts the α hydrogen from leucine in RGYALG, then it is clear that this radical can in turn abstract hydrogen from the β position of tyrosine to yield the a_3 fragment. Similarly, it is likely that abstraction of the hydrogens from either glycine residue could lead to a similar outcome. The probability for subsequent abstractions relates to the relevant barriers. For example, abstraction of the leucine α hydrogen could also be followed by side chain loss; however, if this channel is guarded by a larger barrier, then abstraction of the β hydrogen of tyrosine may prevail. The results also suggest that peptides are sufficiently flexible to minimize any conformational barriers which might prevent radical migration.

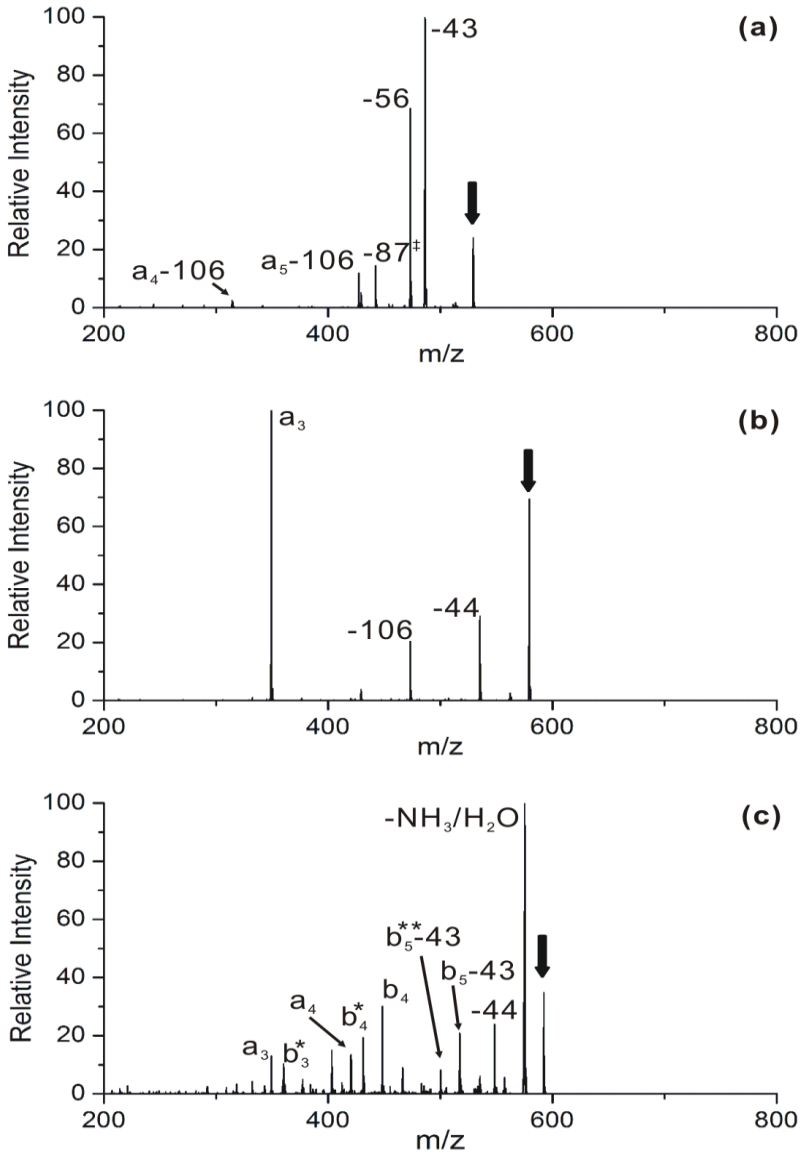


Figure 2.7 The side chain losses observed in Figure 2.2 are re-isolated and subjected to further collisional activation. Radical migration can be monitored. The relative energetics of different mechanistic pathways can be compared from a known starting point. a) CID spectrum for $[RGYALG\cdot\text{-}106+H]^+$. b) CID spectrum for $[RGYALG\cdot\text{-}56+H]^+$. c) CID spectrum for $[RGYALG\cdot\text{-}43+H]^+$. * loss of NH_3. ** loss of H_2O. ‡ consecutive loss of 44 and 43.

2.4 Conclusions

We have demonstrated that hydrogen deficient peptide radicals can be easily generated by the noncovalent attachment of a suitable photolabile hydrogen abstraction reagent. Collisional activation of these hydrogen deficient peptide radicals yields fragmentation patterns which are dissimilar to those produced by CID of nonradical peptides, ECD, ETD, or related techniques. It is found that radical chemistry controls the preferred fragmentation pathways, and that all other factors such as mobile protons or peptide structure play a secondary role. Interestingly, backbone fragmentation produces primarily a-type ions, and the probability for dissociation is dictated by the nature of the side chains to a greater extent than is observed with any other fragmentation method. This should translate into greater predictability of the results, and subsequently greater confidence in data analysis tasks such as peptide identification. Verification of this hypothesis will require analysis of a larger set of peptides, which is currently limited by the need for a more efficient method of preparing the peptide radicals. Fortunately, there are many available remaining options that have not been fully explored.

In addition to interesting backbone fragmentations, numerous side chain losses are observed. It is demonstrated that a large number of these losses can be used to uniquely identify amino acids in a peptide. Although the location of the amino acid is not revealed, this information can be used to narrow down candidate peptides during a database search. All peptides not containing the amino acids identified by side chain losses are simply eliminated from the search. The orthogonality and potential utility of the data obtained by

fragmenting hydrogen deficient radical peptides relative to other fragmentation techniques merits further exploration of this methodology.

Relative to most charge directed pathways, radical chemistry is able to dominate the fragmentation of peptides in these experiments by lowering the barriers to dissociation. In this regard, photoactivation is important for radical generation because the entire molecule is not heated substantially in the process. Although collisional activation can be easily implemented for radical activation, the concomitant heating of the entire peptide can enable nonradical dissociation pathways or less favorable radical dissociation pathways. This situation will be particularly acute if the barrier to generate the radical is substantial and larger than subsequent fragmentation barriers (which would be predicted to be the case since radicals typically facilitate fragmentation). Furthermore, mobile protons can prevent or interfere with radical generation by collisional activation but are not observed to affect photoactivation. Although our results are largely similar to those produced previously with collisionally activated radicals, we observe less and more selective fragmentation which we attribute to initially cooler radicals.

2.5 Supporting Information

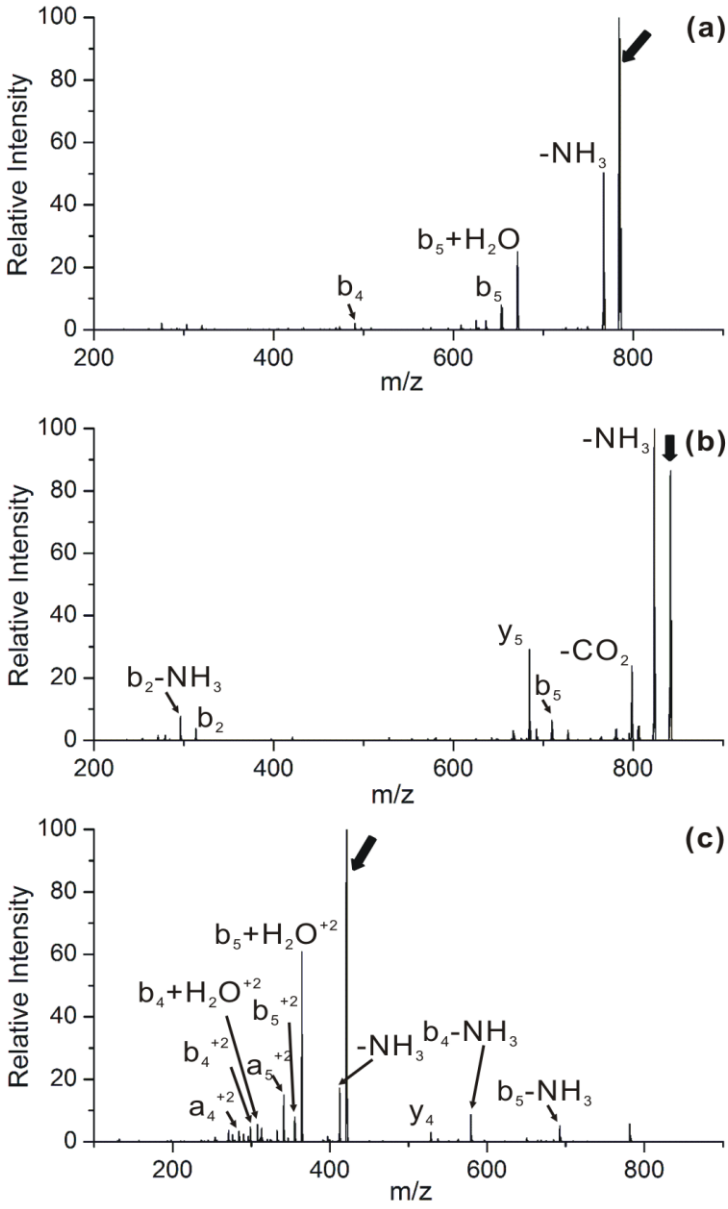


Figure 2.8 (a) CID of $[[RYLGYL+H]^{+}$ (b) CID of $[RRPWIL+H]^{+}$ (c) CID of $[[RRPWIL+2H]^{+2}$.

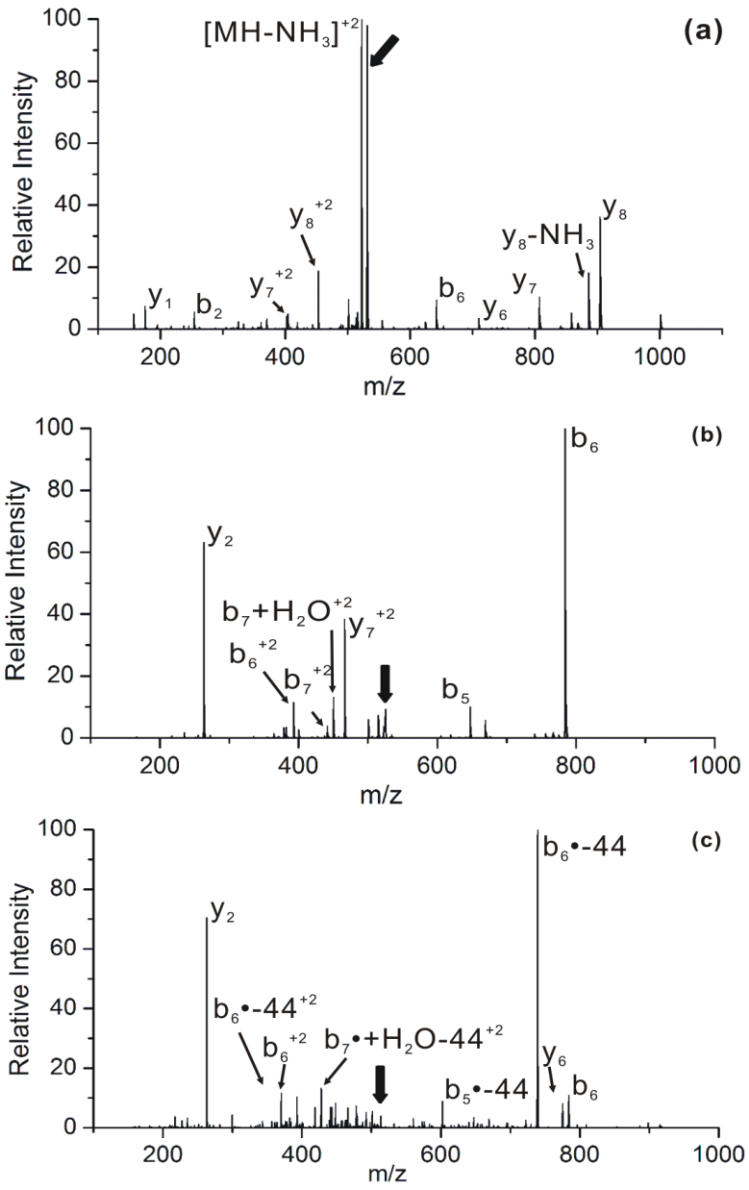


Figure 2.9 (a) CID of $[RPPGFSPFR+2H]^{+2}$ (b) CID of $[DRVYIHPF+2H]^{+2}$ (c) PD/CID of $[DRVYIHPF•-44+2H]^{+2}$

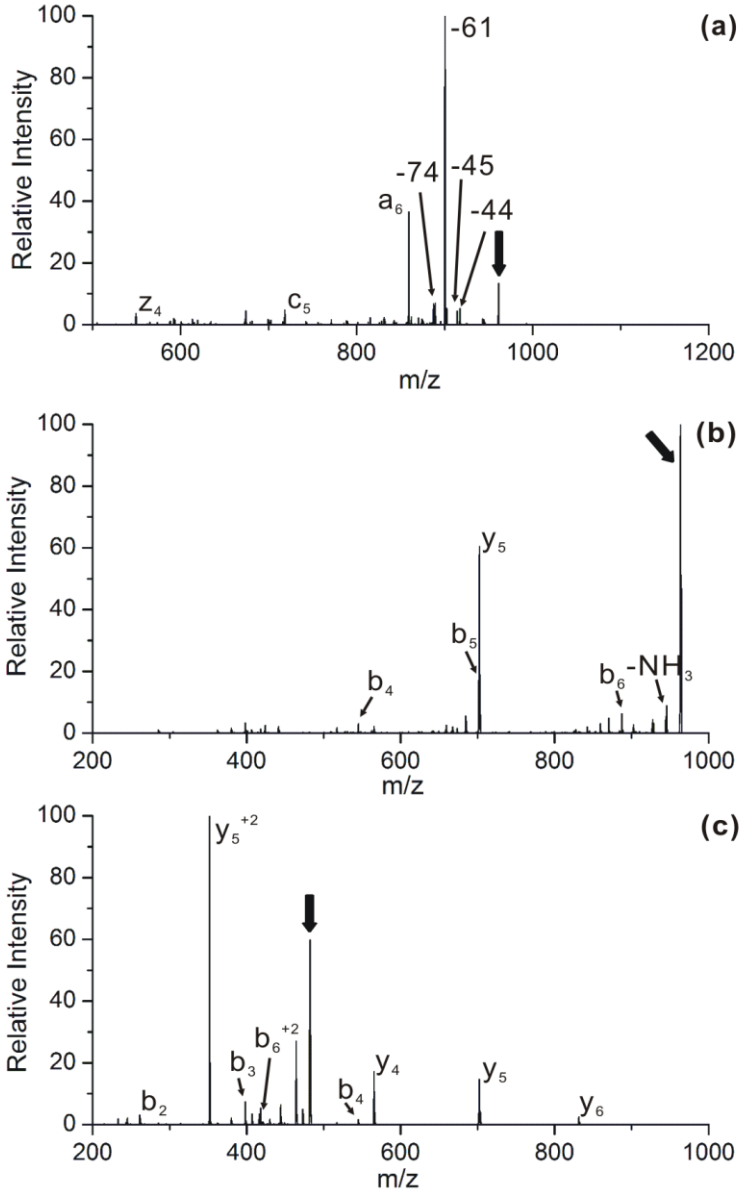


Figure 2.10 (a) PD/CID of $[MEHFRWG\bullet+H]^{+}$ (b) CID of $[MEHFRWG+H]^{+}$ (c) CID of $[MEHFRWG+2H]^{+2}$.

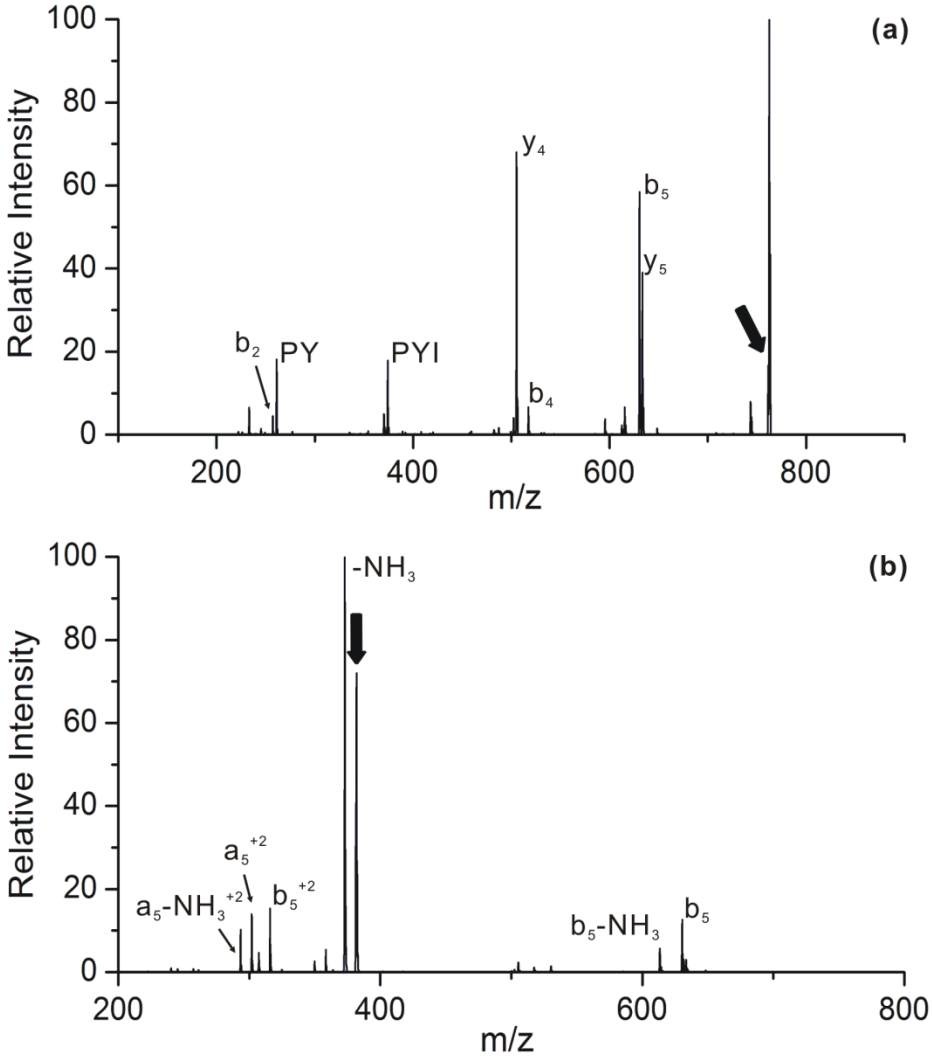


Figure 2.11 (a) CID of $[KKPYIL+H]^{+}$ (b) CID of$[KKPYIL+2H]^{+2}$.

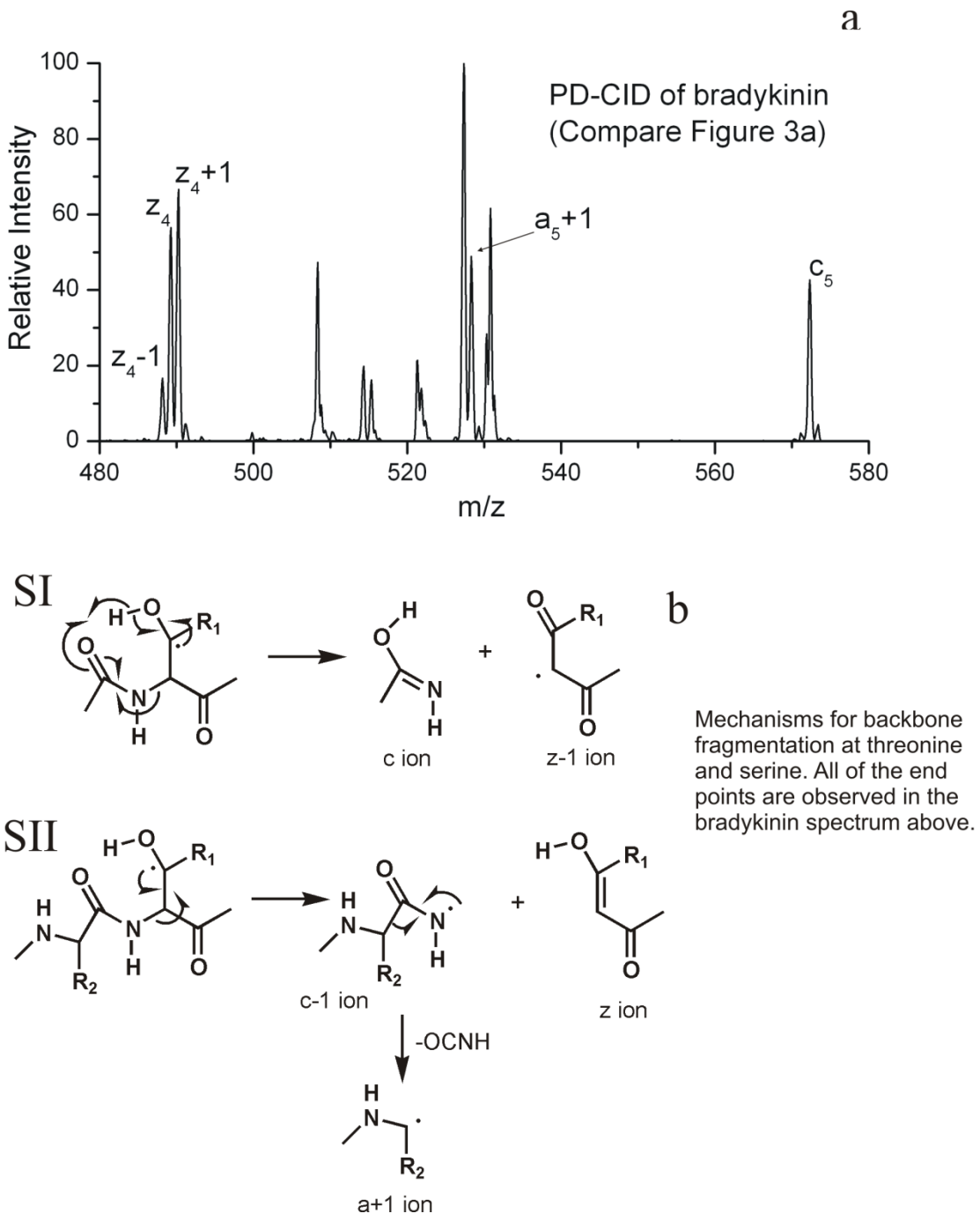
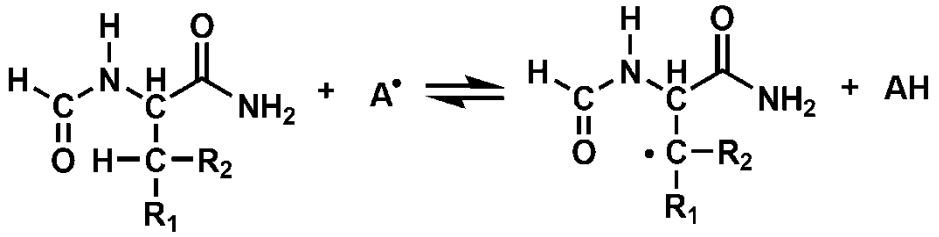


Figure 2.12 (a) CID of $[RPPGFSPFR\cdot+2H]^{2+}$ (b) Mechanism of c-, z-type ions generated on the N-terminus of serine and threorine residues.



Scheme 2.5 Isodesmic reaction used for all C-H BDE calculations

---

[1] Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 563-573.

[2] Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9528-9533.

[3] Chu, I. K.; Rodriquez, C. F.; Lau, T. C.; Hopkinson, A. C.; Siu, K. W. M. *J. Chem. Phys. B* **2000**, *104*, 3393-3397.

[4] Wee, S.; O'Hair, R. A. J.; McFadyen, W. D. *Int. J. Mass Spectrom.* **2004**, *234*, 101-122.

[5] Bagheri-Majdi, E.; Ke, Y. Y.; Orlova, G.; Chu, I. K.; Hopkinson, A. C.; Siu, K. W. M. *J. Chem. Phys. B* **2004**, *108*, 11170-11181.

[6] Barlow, C. K.; Wee, S.; McFadyen, W. D.; O'Hair, R. A. J. *Dalton Trans.* **2004**, 3199-3204.

[7] Laskin, J.; Yang, Z.; Chu, I. K. *J. Am. Chem. Soc.* **2008**, *130*, 3218-3230.

[8] Hodyss, R.; Cox, H. A.; Beauchamp, J. L. *J. Am. Chem. Soc.* **2005**, *127*, 12436-12437.

[9] Hao, G.; Gross, S. S. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1725-1730.

[10] Yin, H.; Chacon, A.; Porter, N. A.; Yin, H. Y.; Masterson, D. S. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 807-816.

[11] Masterson, D. S.; Yin, H. Y.; Chacon, A.; Hachey, D. L.; Norris, J. L.; Porter, N. A. *J. Am. Chem. Soc.* **2004**, *126*, 720-721.

[12] Ly, T.; Julian, R. R. *J. Am. Chem. Soc.* **2008**, *130*, 351-358.

[13] Diedrich, J. K.; Julian R. R. *submitted for publication.*

[14] Julian, R. R.; May, J. A.; Stoltz, B. M.; Beauchamp, J. L. *Ang. Chem. Int. Ed.* **2003**, *42*, 1012-1015.

[15] Julian, R. R.; Beauchamp, J. L. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 616-624.

[16] Friess, S. D.; Daniel, J. M.; Hartmann, R.; Zenobi, R. *Int. J. Mass Spectrom.* **2002**, *219*, 269-281.

[17] Julian, R. R.; Akin, M.; May, J. A.; Stoltz, B. M.; Beauchamp, J. L. *Int. J. Mass Spectrom.* **2002**, *220*, 87-96.

[18] Julian, R. R.; Beauchamp, J. L. *Int. J. Mass Spectrom.* **2001**, *210*, 613-623.

[19] Wilson, J. J.; Kirkovits, G. J.; Sessler, J. L.; Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 257-260.

[20] Jemmett, A. E.; Tucker, S. H.; Wellings, I.; *J. Chem. Soc.* **1958**, 2794-2797.

[21] Rauk, A.; Yu, D.; Taylor, J.; Shustov, G.V.; Block, D. A.; Armstrong, D. A. *Biochemistry **1999***, 38, 9089-9096.

[22] Armstrong, D. A.; Yu, D.; Rauk, A. *Can. J. Chem. **1996***, 74, 1192-1199.

[23] Rauk, A.; Yu, D.; Armstrong, D. A. *J. Am. Chem. Soc.* **1997**, *119*, 208-217.

[24] NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101 Release 14, Sept 2006, Editor: Russell D. Johnson III (http://srdata.nist.gov/cccbdb).

[25] Kavita, K.; Das, P. K. *J. Chem. Phys.* **2002**, *117*, 2038-2044.

[26] Nielsen, M. L.; Budnik, B. A.; Haselmann, K. F.; Olsen, J. V.; Zubarev, R. A. *Chem. Phys. Lett.* **2000,** 330, (5-6), 558-562.

[27] Laskin, J.; Yang, Z. B.; Lam, C.; Chu, I. K. *Anal. Chem.* **2007**, *79*, 6607-6614.

[28] Barlow, C. K.; McFadyen, W. D.; O'Hair, R. A. J. *J. Am. Chem. Soc.* **2005**, *127*, 6109-6115.

[29] See supporting information

[30] Bossio, R. E.; Hudgins, R. R.; Marshall, A. G. *J. Chem. Phys. B* **2003**, *107*, 3284-3289.

[31] Wood, G. P. F.; Easton, C. J.; Rauk, A.; Davies, M. J.; Radom, L. *J. Phys. Chem. A* **2006**, *110*, 10316-10323.

[32] Galano, A.; Alvarez-Idaboy, J. R.; Bravo-Perez, G.; Ruiz-Santoyo, M. E. *J. Mol. Struc. THEOCHEM* **2002**, *617*, 77-86.

[33] Galano, A.; Alvarez-Idaboy, J. R.; Agacino-Valdes, E.; Ruiz-Santoyo, M. E. *J. Mol. Struc. THEOCHEM* **2004**, *676*, 97-103.

[34] Cruz-Torres, A.; Galano, A.; Alvarez-Idaboy, J. R. *Phys. Chem. Chem. Phys.* **2006**, *8*, 285-292.

[35] Chu, I. K.; Zhao, J.; Xu, M.; Siu, S. O.; Hopkinson, A. C.; Siu, K. W. M. *J. Am. Chem. Soc.* **2008**, *130*, 7862-7872.

[36] Lardin, H. A.; Squires, R. R.; Wenthold, P. G. *J. Mass Spectrom.* **2001**, *36*, 607-615.

[37] Rauk, A.; Yu, D.; Armstrong, D. A. *J. Am. Chem. Soc.* **1997,** 119, (1), 208-217.

[38] Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. *J. Am. Chem. Soc.* **1996**, *118*, 8365-8374.

[39] Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Anal. Chem.* **1993**, *65*, 425-438.

[40] Tsaprailis, G.; Somogyi, A.; Nikolaev, E. N.; Wysocki, V. H. *Int. J. Mass Spectrom.* **2000**, *196*, 467-479.

[41] Zhang, Z. Q. *Anal. Chem.* **2004**, *76*, 3908-3922.

[42] Sweet, S. M. M.; Cooper, H. J. *Expert Rev. Proteomics* **2007**, *4*, 149-159.

[43] Fung, Y. M. E.; Chan, T. W. D. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1523-1535.

[44] Barlow, C. K.; McFadyen, W. D.; O'Hair, R. A. J. *J. Am. Chem. Soc.* **2005**, *127*, 6109-6115.

[45] Wee, S.; O'Hair, R. A. J.; McFadyen, W. D. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 884-890.

CHAPTER 3

RADICAL DIRECTED DISSOCIATION FOR FACILE IDENTIFICATION OF IODO-TYROSINE RESIDUES USING ESI-MS

3.1 Introduction

Iodination of proteins at tyrosine residues is arguably one of the simplest chemical modifications that can be carried out on such complex biological molecules. In fact, this modification was first reported in the 1800s, well before the true nature and structure of proteins was completely understood. Through the years tyrosine iodination has been employed for a variety of purposes. Iodination can be utilized to incorporate radioactive isotopes for tracing in biological samples[1] or for cancer treatments.[2] Iodine is a heavy element and can be used to assist phasing assignments in x-ray crystallography experiments.[3] The selective, accessibility based reactivity of iodine with tyrosine assists the reproducible labeling needed for ordered crystals in these experiments. To a first approximation, the rate of iodination at any given tyrosine residue is dependent on the degree to which that residue is exposed to solvent and the reactive iodinating reagent.[4] This structural based reactivity is increasingly being utilized for mapping surface structure in proteins, which has only become feasible with recent advances in mass spectrometry (MS) that can more easily identify modified residues.[5]

Soft ionization techniques coupled with MS have transformed protein characterization chemistry. The level of detail which can now be readily extracted from proteins and peptides far surpasses what was possible a few decades ago. For example, sequence assignment of unmodified peptides and proteins, which was once an onerous task, is now

routine if the genome is known. Post translational modifications (PTMs) introduce additional complexities and often require extra time or attention, but can normally still be successfully characterized in most instances. Iodination at tyrosine is both a biologically relevant[6] and chemically useful PTM.[1-5]

Site localization of iodination by MS would typically follow one of two pathways. In the first, a protein would be enzymatically digested into peptides, which would be individually analyzed for modifications. This bottom up approach is desirable because peptides are more amenable to liquid separations than proteins and sequencing peptides is easier than sequencing whole proteins. Unfortunately, the process of digestion frequently leads to the loss of many peptides, some of which may be important. The second typical method employed attempts direct analysis of the intact protein itself and is referred to as the top down approach. Top down MS is advantageous because no part of the protein is lost prior to examination. Nevertheless, full sequence analysis for whole proteins is difficult to achieve and is frequently complicated by incomplete fragmentation which can lead to ambiguous site assignments. Sensitivity can also be an issue because whole proteins are frequently more difficult to ionize, populate numerous charge states, and can fragment into hundreds of product ions.

One strategy for overcoming some of the difficulties associated with fragmenting whole proteins is to introduce chemistry which directs fragmentation to specific, targeted sites to yield the desired information. For example, if one desires to identify a PTM, chemistry would be introduced to cleave the protein at the specific PTM. This approach has already been demonstrated for peptide phosphorylation.[7] However, the advantages of

such an approach are even more compelling for whole proteins. For example, if a 100 residue protein with a single PTM at one residue were evaluated by electron capture dissociation (ECD), then complete sequence coverage would be desirable and with partial overlap of c and z ions in addition to side chain losses and b and y ions, it is reasonable to suggest that 150 product peaks could be generated. In the ideal case, at best four peaks among the 150 would contain information identifying the site of modification. (The relevant peaks would be the two on either side of a shift in mass from modified to unmodified fragments for either the c or z ion series.) In contrast, the ideal selective fragmentation scenario yields only two peaks, both of which are informative. By preventing dilution of the ion intensity into ~145 uninformative channels, substantial gains in sensitivity are expected. In addition, selectivity affords tremendous simplicity in data analysis because the desired information is directly revealed.

In the present work, we demonstrate that UV photodissociation which selectively targets iodinated tyrosine residues can be used to identify sites of iodination in whole proteins. Radical directed dissociation (RDD) occurs at and in the vicinity of the labeled tyrosine residues. High charge states are evaluated to limit radical transfer to sequence-remote portions of the protein. The method is compared with bottom up and top down approaches, which are both substantially more labor intensive. Where modification sites were able to be determined, agreement between all three methods was obtained. Frequently, one or more of the methods was unable to unambiguously identify the site of modification. The structural sensitivity of iodination was also explored using this facile

new site identification method. It is clear that iodination can be used to quickly gather course level information about changes in protein structure.

3.2 Experimental Methods

3.2.1 Materials

The proteins ubiquitin (bovine), myoglobin (horse heart), cytochrome c (horse heart), lysozyme (chicken egg), hemoglobin (human) and chemicals dithiolthreitol, iodoacetamide were purchased from Sigma Aldrich (St. Louis, MO). Trypsin (modified) was purchased from Roche (Palo Alto, CA). Water was purified to 18.2 MΩ resistivity by a Millipore Direct-Q (Millipore, Billerica, MA). Dialysis membrane (MWCO=3500Da) and clips were purchased from Spectra Por (Rancho Dominguez, CA). Sodium iodide, chloramine-T, sodium metabisulfite and all the other chemicals and solvents were purchased from Fisher Scientific (Fairlawn, NJ). All the chemicals and reagents were used without purification unless notified.

3.2.2 Preparation of iodinated proteins

Proteins were iodinated at room temperature using a previously published method.[8] Sodium iodide and chloramine-T were used as iodination reagents and the iodination reaction was quenched by addition of excess sodium metabisulfite. Stoichiometric quantities of iodination reagents were used to avoid protein oxidation and limit extent and heterogeneity of iodination (1:2:2:3 molar ratio of protein: sodium iodide: chloramine-T: sodium metabisulfite). Due to different three-dimensional structures of proteins, iodination reaction time was optimized for each protein to ensure at least 50% yield of

monoiodinated protein. Iodinated proteins were purified by removal of excess iodination reagents via membrane dialysis against water.

3.2.3 Iodination of denatured myoglobin

Myoglobin was incubated in 6M guanidine hydrochloride solution containing 50mM ammonium bicarbonate buffer at 50˚C for 3 hours.[9] Iodination of denatured myoglobin was not performed until the protein solution was completely cooled to room temperature. The same iodination conditions as used for iodination of folded myoglobin were used in the presence of 6M guanidine buffer solution. Excess reagents were removed from myoglobin solution by membrane dialysis against water overnight.

3.2.4 Enzymatic digestion of proteins

1nmol of each purified iodinated protein was dissolved in 50 μl of 50mM sodium bicarbonate buffer respectively (pH~8.5). Endopeptidase trypsin was added to each iodo-protein buffer solution. The ratio between protein and trypsin was 50:1 (by weight). Iodo-protein solutions containing trypsin were incubated at 37˚C overnight. Enzymatic digestion was quenched by addition of 1 μl trifluoroacetic acid (TFA). Enzymatic digested proteins were lyophilized and reconstituted in water with 0.1% TFA for further MALDI-MS analysis.

3.2.5 MALDI-MS/MS analysis of digested iodinated proteins

Tryptic digested iodinated proteins were analyzed with an Applied Biosystems (Foster City, CA) QSTAR XL quadrupole/time-of-flight mass spectrometer with an o-MALDI ion source (W. M. Keck Proteomics Laboratory/Center for Plant Cell Biology, Institute for Integrative Genome Biology, University of California, Riverside). Digested iodo-

proteins were firstly desalted with zip tips (Millipore, Billerica, MA) with C18 bedding, and then spotted onto a MALDI plate. α-Cyano-4-hydroxycinnamic acid (CHCA) was used as the matrix for MALDI-MS analysis. The iodinated tryptic peptides of each iodinated protein were sequenced by MALDI-MS/MS to elucidate the exact iodinated site. Collision energy used for MS/MS experiment varied from 50 to 75 eV depending on peptide sequences in order to achieve sufficient fragmentation.

3.2.6 Photodissociation/collision induced dissociation (PD/CID) of iodinated proteins

7μM iodinated protein solutions in 50%/50% acetonitrile/H_2O with 0.1% acetic acid were directly infused to a LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) with a standard ESI source. The posterior plate of the LTQ was modified with a quartz window to transit fourth harmonic (266nm) laser pulses from a flash lamp pumped Nd/YAG laser (Continuum, Santa Clara, CA). The pulses were synchronized to the end of the isolation step of a typical MS^2 experiment by sending a TTL trigger signal from the mass spectrometer to the laser by a digital delay generator (Berkeley Nucleonics, San Rafael, CA).[10] The isolation window width of MS^2 and MS^3 experiments was set to 5 Da.

3.2.7 Collision induced dissociation (CID) and electron capture dissociation (ECD) of iodinated proteins

CID and ECD of monoiodinated proteins were acquired using a 7T linear ion trap Fourier transform (LTQ-FT Ultra) mass spectrometer (Thermo Scientific, San Jose, CA) with ESI. The isolation window width for CID and ECD of iodinated proteins was set to 3 m/z-units. The excitation time period of ECD was optimized for each protein to

maximize ECD fragmentation. Mass accuracy of protein fragments by LTQ-FT MS was found to be no more than 4ppm.

3.2.8 Data analysis

Protein fragments were assigned with the aid of UCSF protein prospector v 5.3.0 and Fragmentor software (v.1.0.0.3; http://faculty.ucr.edu/ryanj/fragmentor.html). Nomenclature of protein fragments adopted previous literature.[11]

3.3 Results and Discussion

Previous experiments employing photodissociation coupled with collision induced dissociation demonstrated selective fragmentation in the vicinity of tyrosine residues for several monoiodinated proteins,[12] and these results suggest that RDD may be an attractive approach for the facile identification of sites of iodination in proteins. In Figure 3.1a, the full mass spectrum of iodinated ubiquitin (Ubi) is shown. The numbers in Figure 3.1a indicate charge state and number of attached iodines. For example, 10-1 represents the +10 charge state of monoiodinated ubiquitin, or ($[^{I}Ubi+10H]^{+10}$). Iodination is intentionally restricted to yield primarily the monoiodo-species in order to avoid protein oxidation and heterogeneity. Ubiquitin contains only a single tyrosine, which can be singly or doubly iodinated. PD of isolated $[^{I}Ubi+10H]^{+10}$ ions by exposure to 266nm ultra violet (UV) light is shown in Figure 3.1b. The only significant product is $[Ubi•+10H]^{+10}$, generated by a loss of 126 Da from the precursor ion. This loss is due to homolytic C-I bond cleavage as reported previously.[12] The intensity of the radical product produced by PD is ~50% of the precursor ion intensity. The high radical yield suggests iodination at

tyrosine rather than histidine, which typically produces no more than 10% radical relative to the precursor ion intensity.[13] This conclusion was verified by CID of $[^{I}Ubi+10H]^{+10}$, which confirms that Tyr59 is exclusively iodinated (see Supporting Information). In Figure 3.1c, the MS^3 CID spectrum for $[Ubi•+10H]^{+10}$ is shown. The most intense fragment in Figure 3.1c corresponds to a loss of 106 Da, which is due to side chain loss from Tyr59.[14] Backbone cleavage products are also present and primarily include a-, c- and z-type fragments which are signatures for RDD.[12,15] A few typical proton catalyzed fragments are present as well. The majority of the RDD associated peaks yield fragmentation of the ubiquitin backbone in close proximity to Tyr59 (*vide infra*).

The highly abundant loss of 106Da from tyrosine will initially create a radical at the α-position of Tyr59, which may lead to subsequent selectivity. This possibility was explored by collisionally activating the $[Ubi•+10H-106]^{+10}$ ion (MS^4) as shown in Figure 3.1d. In contrast to the results shown in Figure 3.1c, b- and y-type ions are the dominant fragments, suggesting that proton initiated fragmentation is more competitive in this case. Interestingly, a_{58}^{+7} and $a_{60}\text{-}106^{+7}$ are the only two abundant RDD type fragments observed, and they correspond to dissociation from both residues flanking Tyr59. The results suggest that the α-radical generated by tyrosine side chain loss is much less reactive than the radical initially generated by loss of iodine, in agreement with predicted reactivities.[15] In this situation, the reduced reactivity yields increased selectivity for the RDD fragments, and greatly facilitates identification of the initially iodinated tyrosine residue.

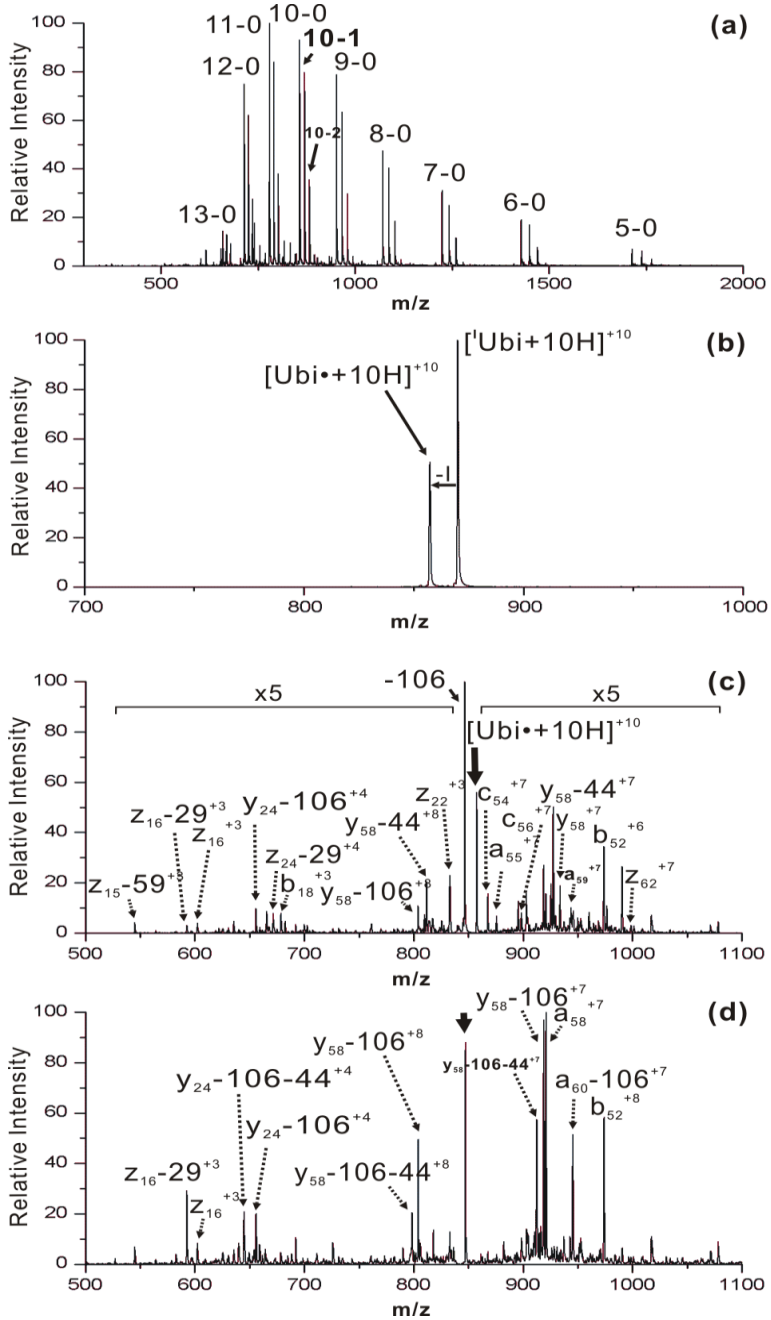


Figure 3.1 (a) Full ESI mass spectrum for iodinated ubiquitin. (b) Photodissociation of +10 charge state of monoiodo-ubiquitin. (c) CID spectrum of ubiquitin radical ($[Ubi•+10H]^{+10}$) generated by PD step, the -106 peak is due to tyrosine side chain loss. (d) CID spectrum of -106 fragment in (c).

To facilitate inspection of the results, a normalized histogram of unique RDD fragments is plotted as a function of sequence in Figures 3.2a and 3.2b for the precursor radical and tyrosine side chain loss radical, respectively. In Figure 3.2a, dissociation is concentrated between Glu51 and Glu64 and is only observed at residues which are in proximity to Tyr59. The histogram in Figure 3.2b for the Tyr-side chain loss radical shows fragmentation to be limited to residues adjacent to Tyr59. For comparison, the RDD histogram for $[Ubi\bullet+6H]^{+6}$ is shown in Figure 3.2c. Although fragmentation is still most abundant in the vicinity of Tyr59, there is also sequence remote fragmentation. It is well known that the gas phase structures of proteins are highly dependent on charge state, with higher charge states always yielding more extended structures.[16] The +6 charge state of ubiquitin has a more compact structure than the +10 charge state.[17] The compact structure facilitates through space radical transfer, which can in fact be utilized to examine protein structure, if desired.[18] In order to limit sequence remote radical migration, which is undesirable for the goals of the present study, high charge states will be exclusively examined for the proteins herein.

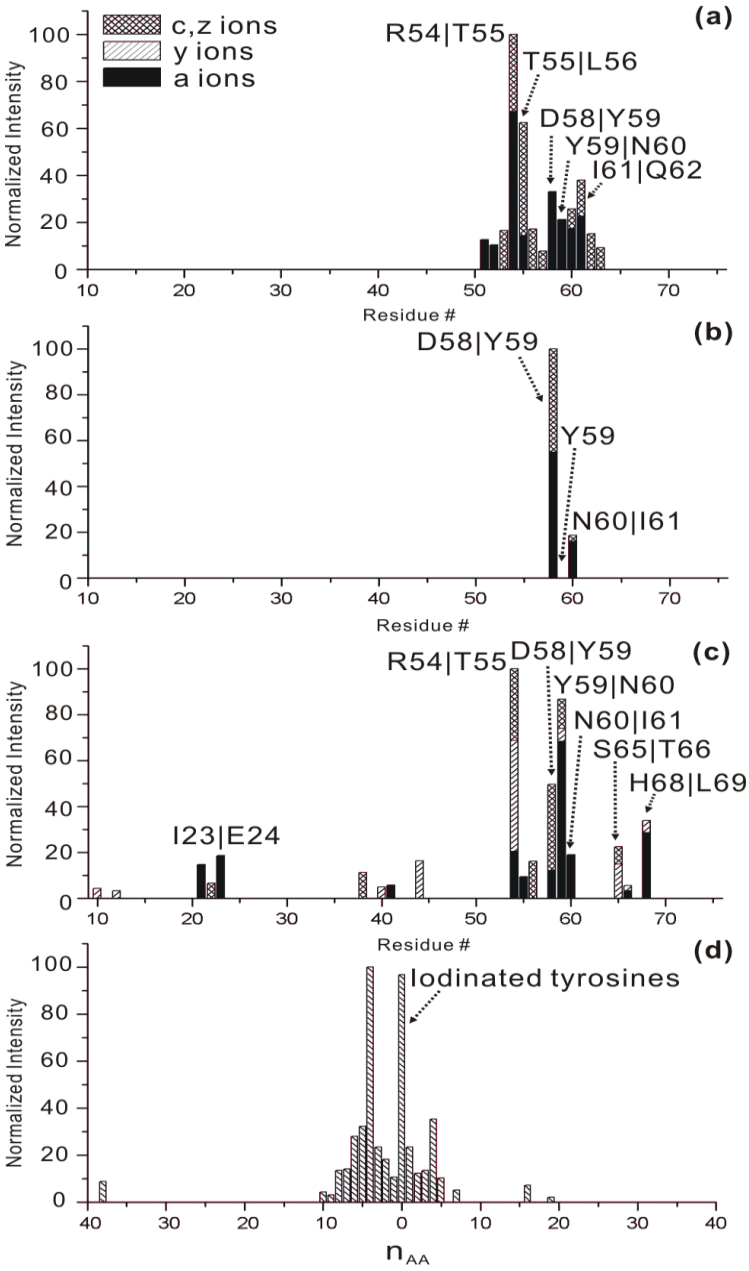
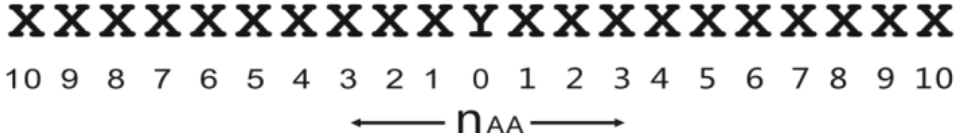


Figure 3.2 Histograms of normalized RDD fragments of ubiquitin in different charge states: (a) +10 charge state parent radical, (b) +10 charge state tyrosine side chain loss radical, (c) +6 charge state illustrating through space radical migration for more compact structures. (d) Histogram of dissociation proximity to iodinated tyrosine residues for all proteins investigated in the present work.

$$\text{Iodination score} = RI_Y + \sum_{n_{AA}=1}^{10}\left(\frac{1}{1+n_{AA}} \times RI_{AA}\right)$$



Scheme 3.1 Algorithm for iodination score calculation of each tyrosine residue within a single protein

3.3.1 Tyrosine iodination probability score

To quantitatively estimate the probability for iodination at different tyrosine residues within a protein, an iodination score for each tyrosine residue was calculated according to the equation illustrated in Scheme 3.1 In this equation, RI stands for relative intensity, n represents the amino acid position relative to tyrosine in either the N- or C-terminal direction, and the subscripts Y and AA represent tyrosine and any other amino acid, respectively. The iodination score then is simply the sum of the relative intensity of fragments observed at tyrosine and the ten adjacent amino acids in either direction. There is a weighting factor which reduces the relative importance of each amino acid going away from the central tyrosine. The selection of ten residues is based on the data shown in Figure 3.2d, illustrating that for all proteins examined herein fragmentation is typically observed within ten residues of tyrosine. In general, an iodination score over 60 will indicate a high likelihood for iodination at a tyrosine, with the exception occurring for multiple tyrosine residues that are in close sequence proximity (see below).

3.3.2 Myoglobin (Myo)

RDD results for $[Myo\bullet+15H]^{+15}$ are summarized in the upper histogram of Figure 3.3a. The iodination scores for each tyrosine residue are projected down at the appropriate sequence location in Figure 3.3a. All RDD fragments are localized within 10 residues from Tyr103 except for one low abundance fragment at His48. Tyr103 and Tyr146 are the only two tyrosines present in myoglobin, and the results indicate that Tyr103 is most likely iodinated (in agreement with previous literature[19]). To verify this conclusion, iodinated myoglobin was also examined with conventional mass spectrometry based methods. The full MALDI mass spectrum for trypsin digested myoglobin is shown in Figure 3.3b. Importantly, the digestion comes from a mixture of unmodified, singly, and doubly iodinated protein because separation of these species is exceedingly difficult. Therefore, the data in Figure 3.3b represent contributions from all three of these species, whereas the data in Figure 3.3a examine only the mono-iodinated protein. As seen in Figure 3.3b, the tryptic peptide Y_{103}-K_{118} exists in all three iodination states. By comparing the relative intensities of the iodination states for Y_{103}-K_{118} to the relative intensities of the iodination states for the whole protein (see Supporting Information), it is clear that very similar ratios are observed. This suggests that iodination occurs primarily on the Y_{103}-K_{118} portion of the protein. Further MS/MS experiments confirm Tyr103 as the exclusive site of iodination within this peptide. A tryptic peptide including Tyr146 was not observed by MALDI-MS, possibly due to the short anticipated peptide sequence (Y_{146}-K_{147}). Therefore, it is not possible to unambiguously exclude the possibility for iodination at this residue with this bottom up approach.

In order to interrogate iodination at Tyr146, top-down methods were employed. $[^{I}Myo+17H]^{+17}$ was isolated and subjected to ECD, and several key fragments are highlighted in Figures 3.3c-3e. The z_{18}^{+2} ion containing Tyr146 is detected as shown in the left part of Figure 3.3c. However, the iodinated version of z_{18}^{+2} is not observed as shown on the right side of Figure 3.3c, confirming that Tyr146 is not iodinated to any significant extent. Histidine is not easily iodinated if the pH is below 8.5 and therefore typically reacts 30-100 times slower than tyrosine.[20,21] Nevertheless, fragmentation in the vicinity of histidine at His48 and His97 suggests that further inspection is warranted in the case of myoglobin. In Figure 3.3d, the extent of iodination for the c_{98}^{+11} fragment is shown. This fragment includes eight histidine residues, but the ECD data suggests that no significant iodination occurs at His 48 or His97, which is also consistent with the high PD yield (50.3 % relative intensity). The abundant fragmentation at His97 is attributed to the fact that RDD is facile at aromatic residues, as established in previous work.[14,15] Finally, examination of the results shown in Figure 3.3e reveals that the c_{105}^{+11} fragment, which contains Tyr103, is fully iodinated as expected.

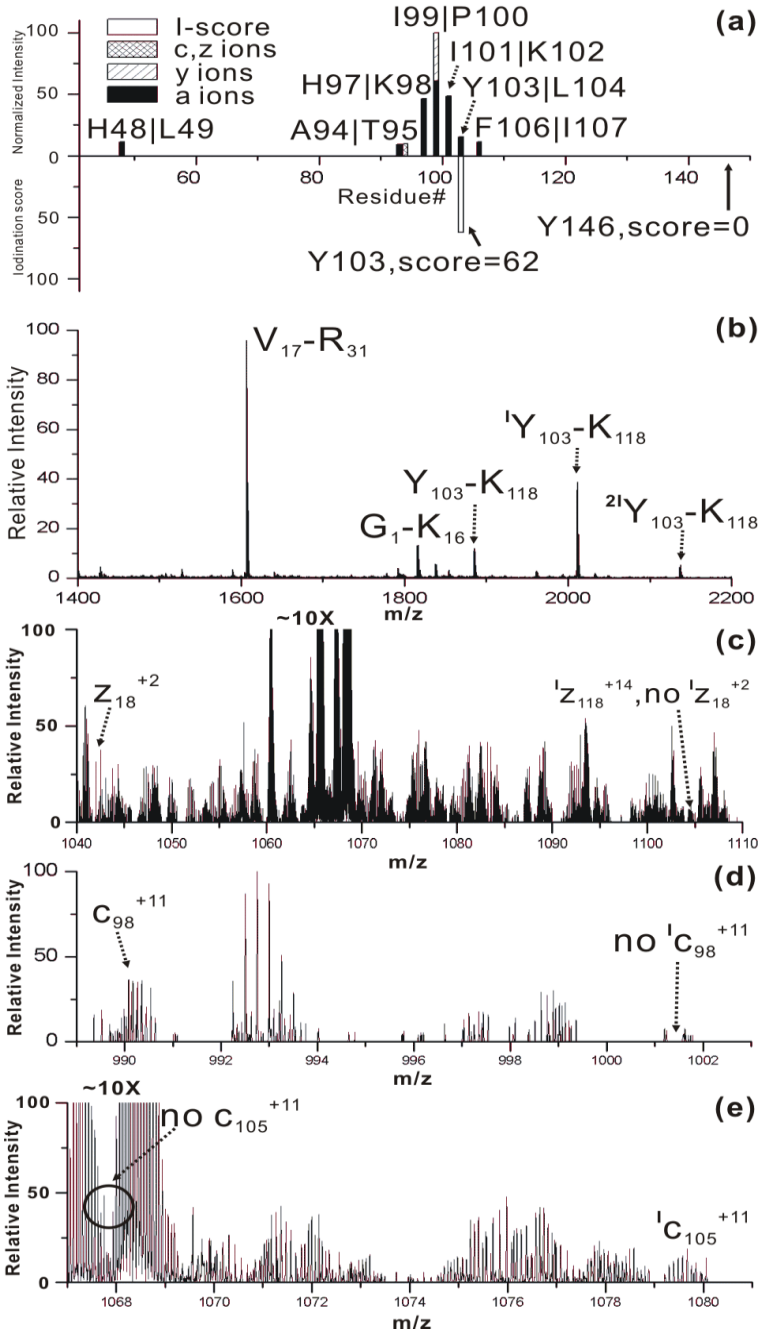


Figure 3.3 (a) Histogram of CID of $[Myo•+15H]^{+15}$ (top) and I-scores for Tyr103 and Tyr146 (bottom). (b) MALDI-MS analysis of tryptic digest of iodinated myoglobin. (c)-(e) zoomed-in ECD spectrum of $[^{I}Myo+17H]^{+17}$.

3.3.3 α-chain of Hemoglobin (Hba)

The α-chain of hemoglobin contains three tyrosine residues. In Figure 3.4a, the RDD results derived from $[Hba\bullet+12H]^{+12}$ are shown with the accompanying iodination scores. For Hba, dissociation is observed in two distinct regions centered on Tyr24 and Tyr140. An iodination score of 111 strongly suggests that Tyr140 is the primary site of iodination. The presence of several fragments in the vicinity of Tyr24 suggests that it may be a secondary site of iodination. The relative iodination scores (111/19), suggests that Tyr24 is not significantly iodinated.

In order to investigate the matter further, both bottom-up and top-down methods were utilized. In Figure 3.4b, the MALDI-MS spectrum of trypsin digested iodinated Hba is shown. Again, the digestion contains peptides from all states of iodination for Hba. Iodinated peptides containing Tyr24 and Tyr42 are both detected, although the extent of iodination for Tyr24 is much greater than Tyr42. For Tyr42, it is clear that very minimal iodination occurs. For Tyr24, the results are ambiguous. At most, ~28% of Tyr24 may be iodinated; however, since contributions from the doubly iodinated protein may contribute to this number, the actual percentage may be substantially lower. A peptide containing Tyr140 was not observed. The bottom up results are again incomplete, and so top down experiments were performed. In Figure 3.4c, results from ECD illustrate that the c_{35}^{+4} ion exists in both iodinated and unmodified forms. This confirms Tyr24 as a site of iodination and also indicates a secondary site. Unfortunately, the degree of iodination for different fragments containing Tyr24 are not consistent with each other, meaning that the data in Figure 3.4c cannot be used to quantify the degree of iodination. CID experiments

yield both y_{23}^{+2} and $^{I}y_{23}^{+2}$ fragments as shown in Figure 3.4d, confirming partial iodination of Tyr140. Unfortunately, CID does not provide quantitative information on the extent of iodination either (compare Figure 3.4d and 4e). The cause is most likely due to the low signal to noise ratios for the relevant fragment ions. It is also possible that iodine may be differentially lost during the excitation process. In this case, the RDD results are the least ambiguous and indicate that Tyr140 is the primary site for iodination, followed by Tyr24.

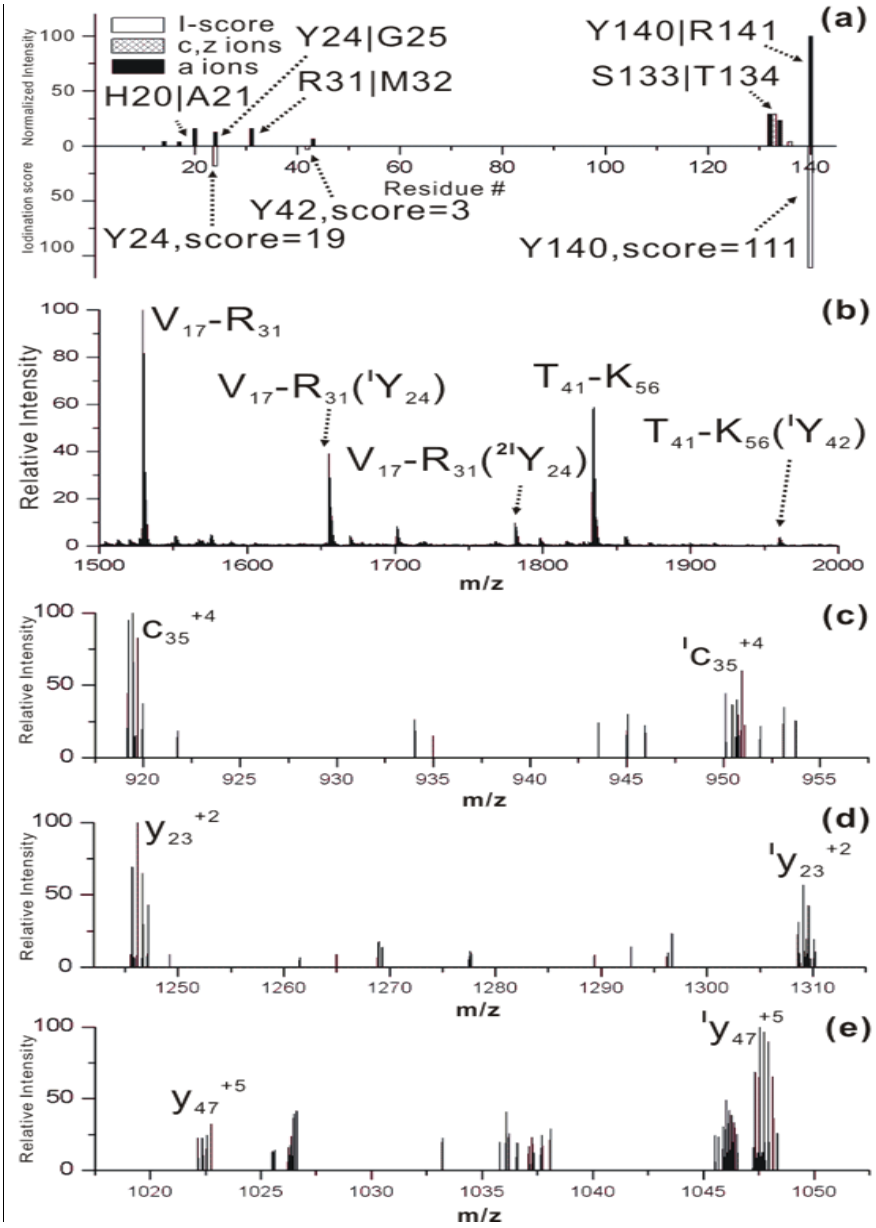


Figure 3.4 (a) Histogram of CID fragments of $[Hba\bullet+12H]^{+12}$ (top) and I-scores for Tyr24, Tyr42 and Tyr140 (bottom). (b) MALDI-MS of tryptic digest of iodinated Hba. (c) zoomed in portion of ECD spectrum for $[^{I}Hba+15H]^{+15}$. (d) zoomed in portion of CID spectrum for $[^{I}Hba+15H]^{+15}$ (e) another portion of the CID spectrum illustrating that iodination cannot be quantified from CID fragments.

3.3.4 Cytochrome c (Cytc)

Cytc has four tyrosine residues which can potentially be iodinated. RDD results and iodination scores for $[Cyt\ c+15H\bullet]^{+15}$ are illustrated in Figure 3.5a. Only five fragments are observed in total and four of them are within 5 residues of Tyr74. It is clear that Tyr74 and Tyr97 are the most likely and unlikely sites for iodination, respectively. An iodination score of 109 for Tyr74 suggests it is the primary iodination site, in agreement with previous investigations.[22-24] The score of zero for Tyr97 clearly indicates that it is not iodinated. A single fragment is detected in the vicinity of Tyr48, resulting in a iodination score of 9; therefore, it is unlikely that Tyr48 is significantly iodinated. The iodination propensity for Tyr67 is ambiguous because of its close proximity to Tyr74. When two tyrosine residues are within 10 amino acids, the fragments from one tyrosine can contribute to the iodination score of the other. This is an inherent limitation of the RDD approach; although, it should be mentioned that residues which are near in sequence present challenges for the standard bottom up and top down approaches as well.

Bottom up methods cannot provide useful information for Cytc unless the various iodination states can be separated from each other, which was not achieved. Therefore, $[^{I}Cyt\ c+13H]^{+13}$ was subjected directly to ECD and CID to probe sites of iodination and the results are shown in Figures 3.5b-5e. On the left side of Figure 3.5b, the c_{48}^{+5} ion which contains Tyr48 is detected only in the non-iodinated form, suggesting that Tyr48 is not iodinated in Cytc. Similarly, detection of the c_{67}^{+7} ion (Tyr67 included) and the y_{29}^{+3} ion (Tyr97 included), but not the corresponding iodinated ions in Figures 3.5c and Figure 3.5e clearly confirms that neither Tyr67 or Tyr97 is iodinated to a significant extent. In

contrast, Figure 3.5d illustrates detection of the iodinated fragment only for the c_{81}^{+9} ion, which implies (in conjunction with the other spectra) that Tyr74 is exclusively iodinated. Tyr74 has the highest solvent accessibility among all the tyrosines in Cytc which should result in the highest iodination reactivity.[25] Tyr97 is buried inside of Cytc and is not observed to be modified even under strongly iodinating conditions.[5] RDD again correctly identifies the primary iodination site; however, examination of Cytc has revealed that RDD may not be able to distinguish iodination states for tyrosine residues within 10 amino acids of each other. Nevertheless, the lack of dissociation at Tyr67 itself and the absence of dissociation on the N-terminal side of Tyr67 both suggest that substantial iodination at this residue may not occur.

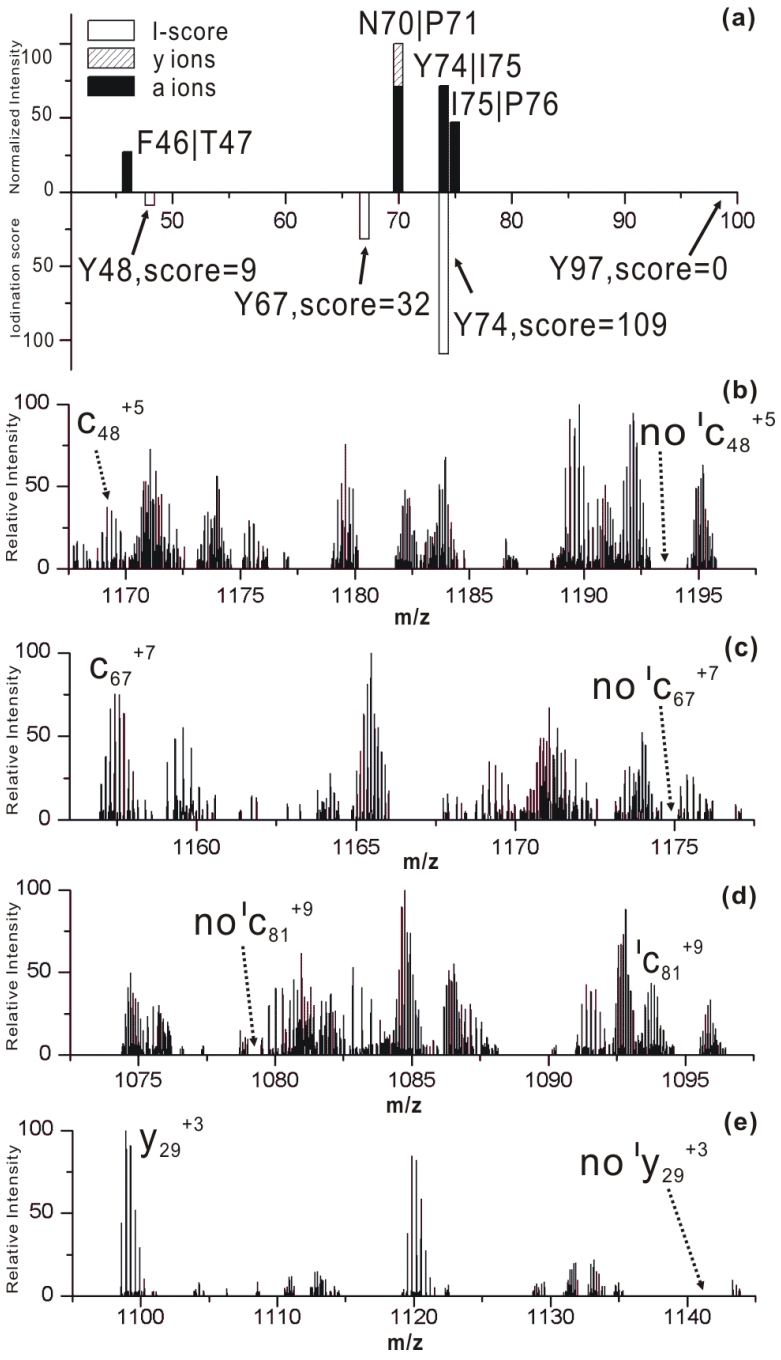


Figure 3.5 (a) Histogram of $[Cyt\ c\bullet+15H]^{+15}$ RDD fragments (top) and I-score values for Tyr48, Tyr67, Tyr74 and Tyr97 (bottom). (b)-(d) zoomed in ECD spectrum for $[^{I}Cyt\ c+13H]^{+13}$. (e) zoomed-in CID spectrum for $[^{I}Cyt\ c+13H]^{+13}$.

3.3.5 Lysozyme (Lyz)

Lysozyme contains three tyrosine residues, two of which are located within three residues of each other. In Figure 3.6a, RDD results are summarized for $[Lyz\bullet+14H]^{+14}$ and the corresponding iodination scores are given. Tyr53 has an iodination score of zero, suggesting that it is not iodinated. Almost all of the fragments observed are concentrated between Asn19 and Trp28, which contains both Tyr20 and Tyr23. Both residues have comparably high iodination scores, although Tyr23 scores slightly higher. These two residues are in very close proximity and RDD cannot be used to distinguish whether both or only one of them is iodinated.

In order to further investigate iodination for Lyz, traditional bottom up and top down methods were employed. The results from a tryptic digest of the iodinated protein mixture are shown in Figure 3.6b. The peptide H_{15}-R_{21} (containing Tyr20) is detected in low abundance and the monoiodinated species of this peptide is not observed. However, given the signal/noise ratio for the spectrum in the region of interest, it is not possible to conclude that Tyr20 is not iodinated to some extent. The mono- and di-iodo Tyr23 containing peptide (G_{22}-K_{32}) is detected with a 0.4:1:0.6 ratio between the unmodified, mono-, and di-iodo peptides. This ratio is consistent with the iodination extent ratio for the whole protein, suggesting that Tyr23 is the primary site for iodination. However, Tyr53 is slightly iodinated in the digested mixture, highlighting the difficulty with trying to quantitate results using bottom up methods.

CID fragmentation of the whole protein yields a y_{80}^{+8} fragment which contains a Tyr53. The mono-iodinated version of y_{80}^{+8} is not observed, confirming the results

obtained by RDD. Iodination for Tyr20 and Tyr23 is not revealed by either CID or ECD with the top-down MS method because fragmentation does not occur between these residues. In the case of Lyz, none of the three methods employed yields completely unambiguous results, highlighting the challenges which can be encountered in the determination of iodination or other post-translational modifications. However, it is also clear that examination by multiple techniques is important in such cases since the combination of information from each may indicate a clear conclusion. Examination of all of the results reveals that Tyr23 is the primary site of iodination, Tyr53 is not iodinated, and Tyr20 is iodinated at most to a very limited extent. It should be mentioned that the solvent accessibilities in the crystal structure for Tyr20 and Tyr23 are very similar.[26] The differential iodination must therefore be explained in terms of reactivity rather than accessibility. Thus, caution is suggested when attempting to attribute the degree of iodination to the relative solvent accessibility.

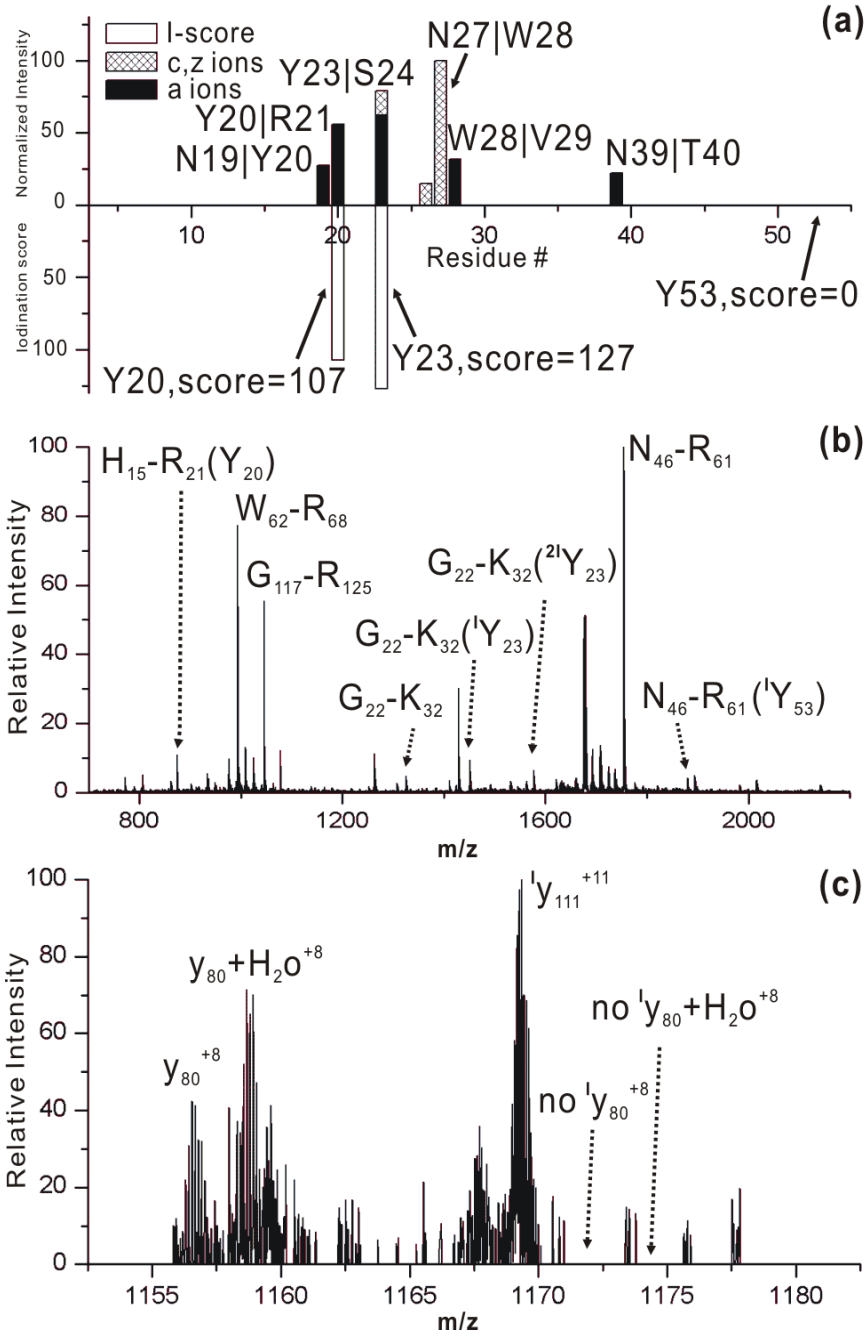


Figure 3.6 (a) Stack plot of $[Lyz\bullet+14H]^{+14}$ RDD fragments (top) and I-score information of Tyr20, Tyr23 and Tyr53 (bottom). (b) MALDI-MS analysis of tryptic digest of iodinated Lyz. (c) zoomed-in CID spectrum of $[Lyz\bullet+14H]^{+14}$.

3.3.6 Examination of Protein Structure

If indeed the extent of iodination is subject to protein structure, and the RDD data are representative of the extent of iodination, then changes to protein structure while iodination is carried out should be manifest in the RDD spectrum. To examine this possibility, myoglobin was iodinated under denaturing conditions. The same +15 charge state $[Myo•+15H]^{+15}$ as probed previously in Figure 3.3 was examined again and the results are shown in Figure 3.7. Comparison of the two spectra immediately reveals substantial differences. The iodination score for Tyr103 did not change substantially (62→75), suggesting that it is still a likely site for iodination. The iodination score for Tyr146 shifted substantially (0→126). Denaturation of the protein clearly exposes Tyr146 to solvent and enables iodination to occur. These results confirm that iodination is sensitive to protein structure and that RDD spectra are representative of the iodination state of a protein.

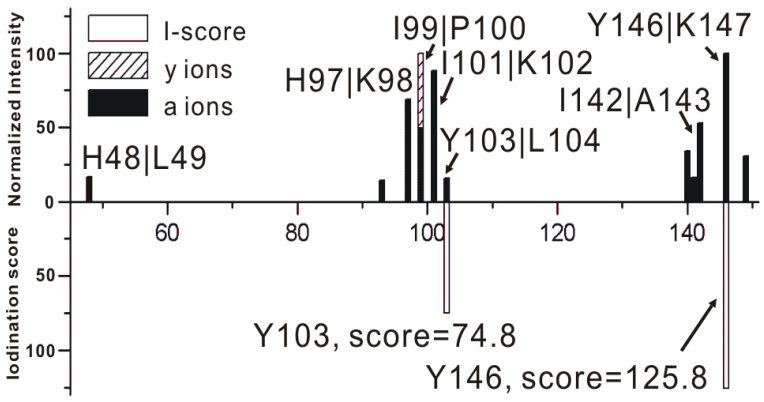


Figure 3.7 Fragment stack plot of RDD results for denatured Myoglobin, $[DMyo\bullet+15H]^{+15}$. Comparison with the data in Figure 3.3 reveals significant differences, confirming the structural sensitivity of iodination and the utility of RDD for iodination site identification.

3.4 Conclusion

Iodination of proteins is facile and has been used for a variety of purposes in numerous fields for over a century. Identification of the sites of iodination is essential for many of these applications, particularly where protein structure is important or where structural characterization is desired. Nevertheless, characterization is frequently nontrivial for these complex molecules. We have demonstrated that the site specificity of photodissociation can be used to generate radicals at iodinated tyrosine residues and that subsequent heating of these radicals leads to fragmentation primarily in the vicinity of the iodinated tyrosine. The highest charge states with reasonable ion intensity are examined to avoid complications arising from radical migration facilitated by protein tertiary structure. Radical initiated tyrosine side chain loss, when present in sufficient yield to allow further analysis, also yields very selective backbone fragmentation following

collisional activation and offers another means for iodination site identification in addition to direct backbone dissociation. In the majority of cases, the site of iodination can be easily identified with these RDD tools. One limitation occurs when two tyrosine residues are in close proximity, in which case iodination will be identified if either residue is modified, but discrimination of the degree of iodination for each residue is difficult. If neither tyrosine is modified, sequence proximity is not problematic. Importantly, a localized region (within approximately ten residues) where iodination occurs is still identified even if multiple tyrosines are present. The primary advantage of this method over bottom up approaches is that a specific protein iodination state is examined and the experiments are rapid and the results easily analyzed. Specificity is the main advantage over top down MS methods, which enhances sensitivity and ensures cleavage of bonds in the region of interest. The discovery of this method should greatly facilitate the use of iodination chemistry in several possible applications such as: 1) examination of course grain protein structure, 2) as a means for mapping out protein-protein interactions, 3) the facile identification and localization of naturally occurring PTMs such as iodotyrosine or bromotyrosine.

3.5 Supporting Information

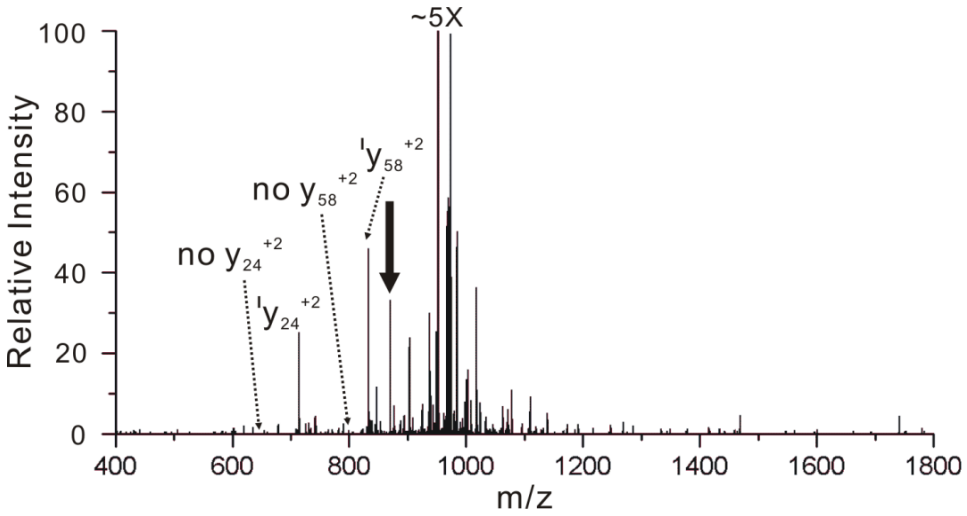


Figure 3.8 CID of $[^{I}\text{Ubi}+10\text{H}]^{+10}$

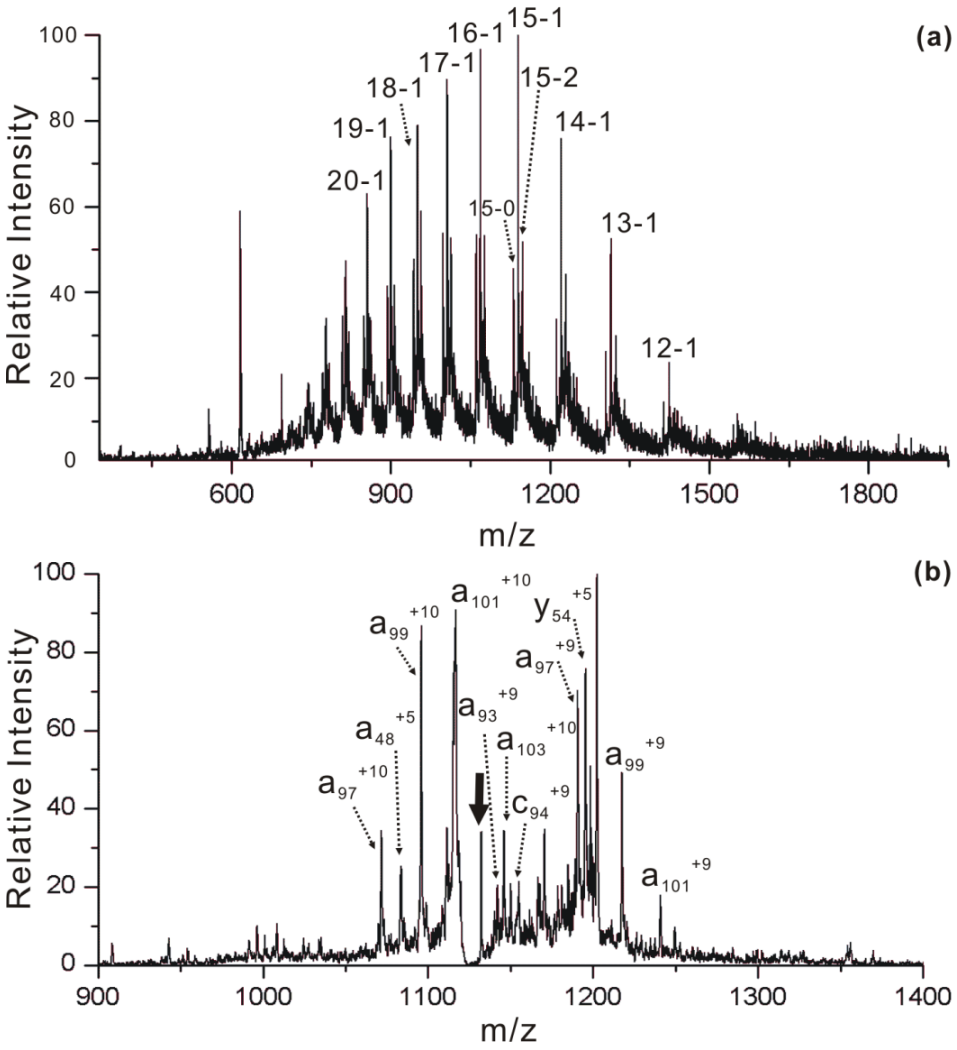


Figure 3.9 (a) Full MS of iodinated myoglobin (b) CID of $[Myo\bullet+15H]^{+15}$.

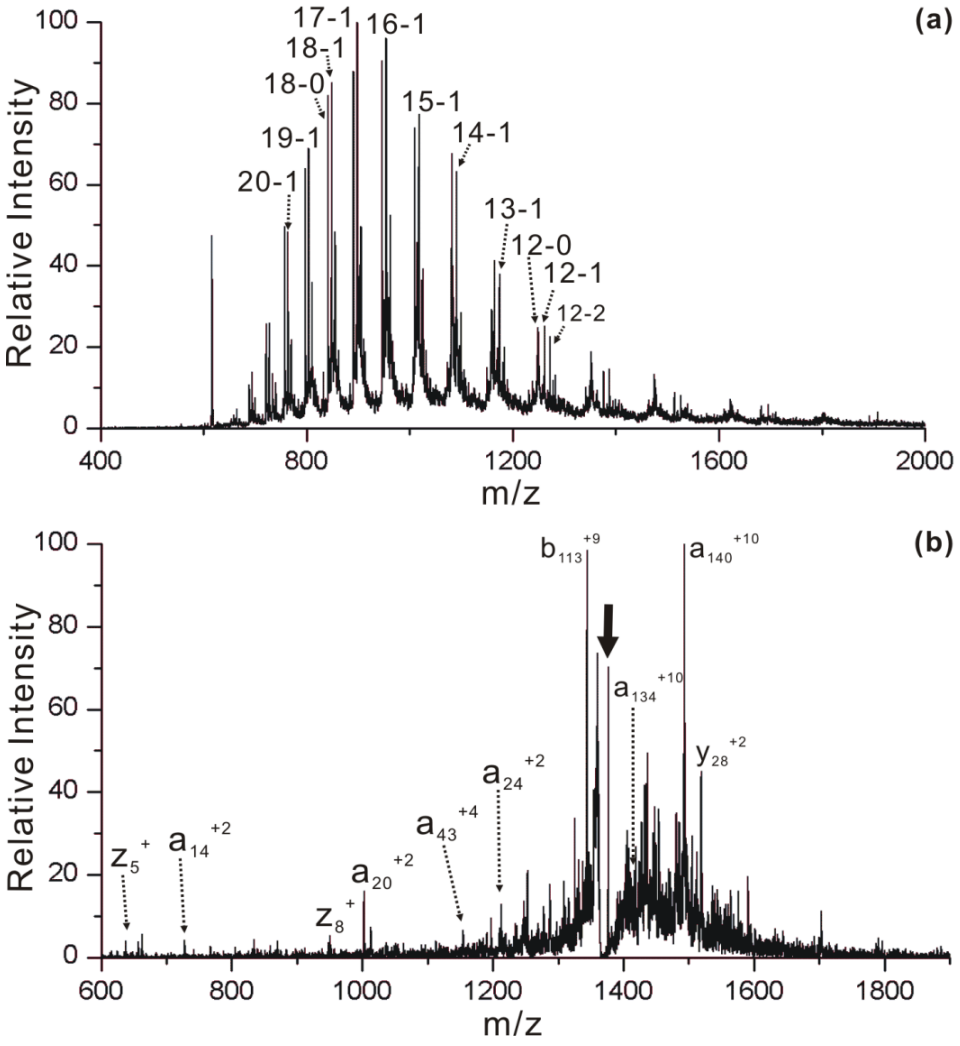


Figure 3.10 (a) Full MS of iodinated hemoglobin (b) CID of $[HEA\bullet+12H]^{+12}$.

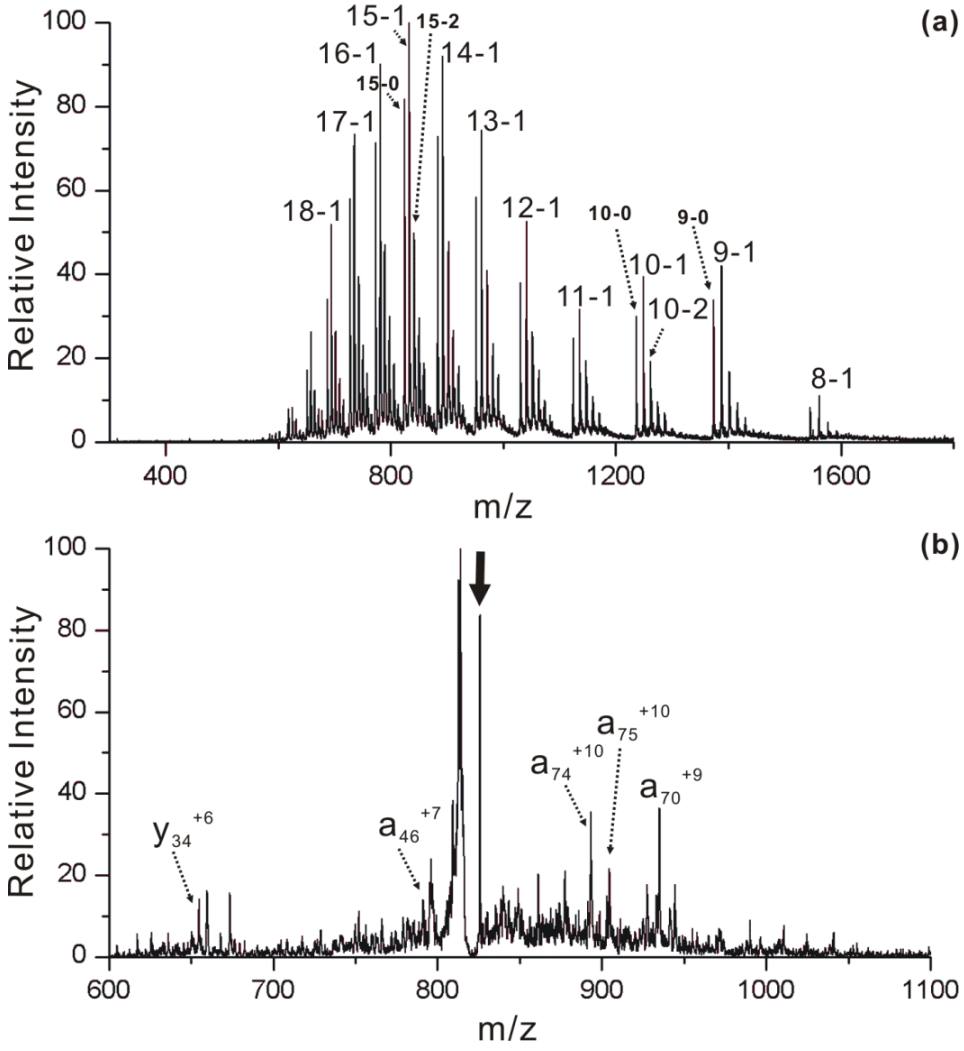


Figure 3.11 (a) Full MS of iodinated cytochrome c (b) CID of $[Cytc\bullet+15H]^{+15}$.

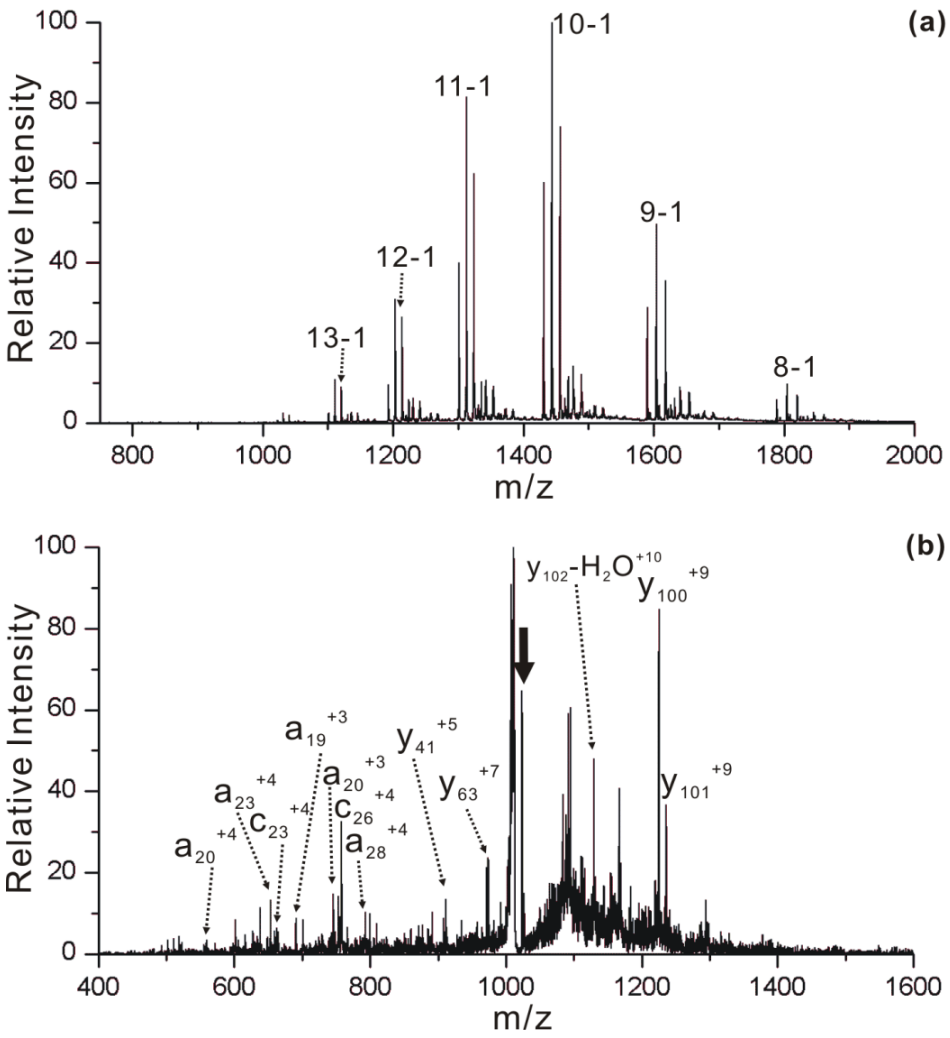


Figure 3.12 (a) Full MS of iodinated lysozyme (b) CID of $[Lyz\bullet+14H]^{+14}$.

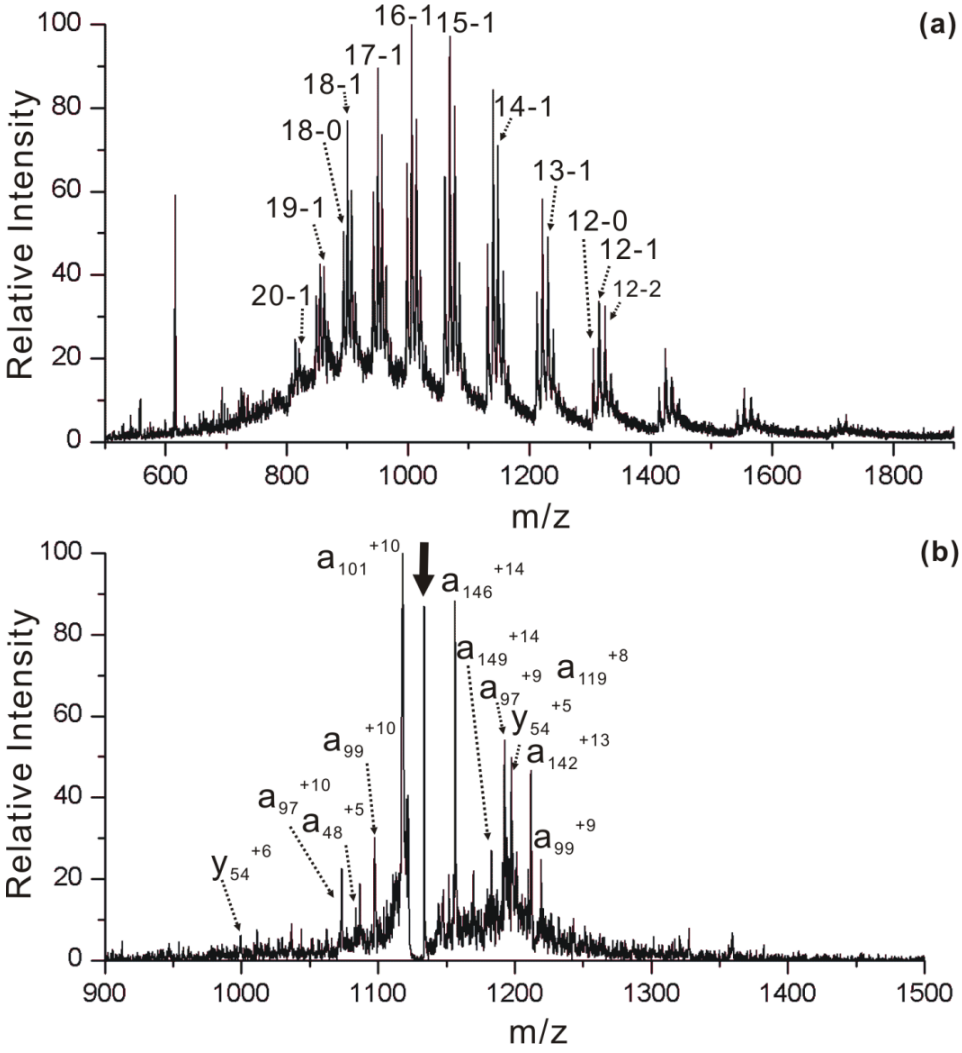


Figure 3.13 (a) Full MS of iodinated D-myoglobin (b) CID of $[D\text{-}Myo\bullet+15H]^{+15}$.

---

[1]Wright, P. J.; di Mayorca, G. *J. Virol.* ***1976****, 19,* 750-755.

[2] Becker, S.; Laffont, S.; Vitry, F.; Rolland, Y.; Lecloirec, J.; Boucher, E.; Raoul, J. L.; Herry, J. Y.; Bourguet, P.; Garin, E., *Nuc. Med. Commun.* **2008,** 29, 815-825.

[3] Dauter, Z. Dauter, M.; Rajashankar, K. R. *Acta. Crystallogr. D. Biol. Crystallogr.* ***2000***, 56, 232-237.

[4] Wolff. J.; Covelli, I. *Biochemistry*, ***1966***, 5, 867-871.

[5] Santrucek, J.; Strohalm, M.; Kadlcik, V.; Hynek, R.; Kodicek, M. *Biochem. Biophys. Res. Commun.* ***2004***, 323, 1151-1156.

[6] Seed. R. W.; Goldberg, I. H. *Science,* ***1965***, 149, 1380-1382.

[7] Diedrich, J. K.; Julian, R. R. *J. Am. Chem. Soc.* ***2008***, 130, 12212-12213.

[8] Regoeczi, E. *Iodine-Labeled Plasma Proteins*; CRC Press: Boca Raton, FL, 1984.

[9] Pace, C. Nick; Vanderburg, Keith E. *Biochemistry* ***1979*** 18, 288-292.

[10] Kim, T. Y.; Thompson, M. S.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* ***2005***, *19*, 1657-1665.

[11] Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* ***1984***, *11*, 601-601.

[12] Ly, T.; Julian, R.R. *J. Am. Chem. Soc.* ***2008***, 130, 352-358.

[13] Liu, Z.; Julian, R. R. *J. Am. Soc. Mass Spectrom.* ***2009***, 20, 965-971.

[14] Ly, T.; Julian, R. R. *J. Am. Soc. Mass Spectrom.* ***2009***, 20, 1148-1158.

[15] Sun, Q.; Nelson, H.; Ly, T.; Stoltz, B. M.; Julian, R. R. *J Proteome Res* ***2008***, 8, 958-966.

[16] Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* ***1997*** , 119, 2240-2248.

[17] Myung, S.; Badman, E.R.; Lee, Y.; Clemmer, D. E. J. Phys. Chem.A ***2002***, 106, 9976-9982.

[18] Ly, T.; Julian, R. R. *submitted for publication.*

[19] Iniesta, J.; Cooper, H. J.; Marshall, A. G.; Heptinstall, J.; Walton, D. J.; Peterson, I. R. *Arch. Biochem. Biophys.* ***2008***, 474, 1-7.

[20] Miyashita, M.; Yamashita, S. *J. Chromatogr.* ***1989***, *475*, 135-144.

[21] Ramachandran, L. K. *Chem. Re*V. ***1956***, *56*, 109-218.

[22] Amico, V.; Fotit, S.; Saletti, R.; Cambria, A.; Petrone, G. *Biomed Environ Mass Spectrom.* ***1988***, 16, 431-437.

[23] McGowan, E.B.; Stellwagen, E. *Biochemistry* ***1970***, 3047–3053.

[24] Osheroff, N.; Feinberg, B.A.; Margoliash, E. *J. Biol. Chem.* ***1977***, 252, 7743–7751.

[25] Sokolovsky, M.; Aviram, I.; Schejter, A. *Biochemistry* ***1970***, 9 5113–5118.

[26] Diamond, R. *J. Mol. Biol.* **1974**, *82*, 371.

CHAPTER 4

PROBING SITES OF HISTIDINE PHOSPHORYLATION WITH IODINATION AND TANDEM MASS SPECTROMETRY

4.1 Introduction

Phosphorylation is a common post translational modification (PTM) of proteins in both eukaryotic and prokaryotic cells where it plays essential roles in both signal transduction and pathway regulation.[1-3] Serine and threonine are the most common sites for phosphorylation, via the formation of phosphoester bonds.[4] Phosphorylation can also occur at nitrogen atoms in amino acids, yielding highly labile phosphoramidate modified residues. For example, the two nitrogen atoms in the imidazole ring of histidine can be phosphorylated to form 1- or 3-phosphohistidine.[5] In fact, phosphohistidine is more widespread in proteins than is generally believed. It has been estimated that up to 6% of protein phosphorylation in eukaryotes may occur at histidine,[6] which would make phosphohistidine significantly more abundant than phosphotyrosine. It is also known that phosphohistidine is critical for mediating two-component signal transduction in eubacteria, archaea, plants, and eukaryotic organisms,[7-10] where the lability of the N-P bond is leveraged for phosphate transfer from phosphodonor to phosphoacceptor moieties.[11,12] In addition, the instability of phosphohistidine makes it a good candidate for chemical signaling involving changes to pH, temperature, and other factors.[13,14]

In proteomics experiments minute quantities of phosphorylated peptides and proteins are located within highly complicated mixtures, motivating the use of both separation and

enrichment strategies prior to MS analysis.[15-17] Liquid chromatography (LC) has been used extensively in conjunction with mass spectrometry (MS) to study phosphorylation at serine, threonine, and tyrosine.[18] However, LC separations typically employ acidic conditions (pH~1.5), which are optimal for peak resolution and also compatible with phosphoester stability. Unfortunately, phosphorylated histidine residues are not acid stable (less than 30min half-life at pH $\leq$ 3),[19] and are frequently quite labile even at neutral pH. Therefore standard separation approaches are incompatible with phosphohistidine identification. In addition, phosphoramidate bonds are unstable in the gas phase and can readily decompose in transit to or upon mild activation within the mass spectrometer. This further complicates site specific analysis, particularly where collision induced dissociation (CID) is implemented. Although electron capture dissociation (ECD)[20, 21] and electron transfer dissociation (ETD)[22, 23] have greatly facilitated phosphorylation site identification, these methods are not universally available.

Several efforts have been made to develop specific techniques for examining phosphorylated histidine residues. For example, radioactive ^{32}P labeling combined with SDS-PAGE has been used to achieve large-scale rapid detection of phosphohistidine.[24] Unfortunately, this radioactive method does not typically provide information about the site of modification. Other efforts have focused on optimizing neutral or basic mobile phase LC in order to avoid the extensive dephosphorylation which occurs under acidic conditions. However even with a neutral mobile phase, significant loss of phosphohistidine can occur during the timescale needed for separation.[25] Under basic conditions, the stability of phosphoester bonds becomes questionable. Fast LC

separations (<7min) at pH~2 also lead to reduced phosphate loss, although at the cost of significantly degraded separation.[26] Therefore, it is difficult to resolve the issue of phosphohistidine identification by simply modifying the LC conditions, without incurring other undesirable side-effects.

Biomolecular labeling is alternative strategy for detecting PTM sites in peptides and proteins.[27-29] This involves chemical modification of the peptide to report, by presence or absence of the modifier, the sites of PTMs. Ideally the label also offers some advantages in identification, such as unique fragmentation, isotopic signatures, or chemical stability. Iodination is one of the easiest methods to modify proteins, with the modification typically occurring at tyrosine. This reaction proceeds with high yield (>90%) and has a variety of uses.[30-32] In addition to tyrosine, histidine residues can also be iodinated, although the rate is typically significantly slower. The disparity in iodination rate is considerably reduced above pH ~8.5.[33] Conveniently, phosphorylation inhibits iodination of both tyrosine and histidine which proceeds through deprotonated intermediates.

In the present work, iodination labeling and tandem mass spectrometry are combined to identify phosphorylated histidine sites in peptides. Iodination labels free histidine while leaving all phosphorylated histidine residues unmodified. After removal of all phosphoramidates with acid, originally phosphorylated histidines can be easily distinguished by tandem mass spectrometry. Iodinated histidine is stable at acidic pH and is not labile in the gas phase, enabling the use of standard LC-MS/MS methodology, including CID, for site specific identification. Semi-quantitative information regarding

the degree of phosphorylation is also easily obtained. This labeling strategy should provide an avenue for examining histidine phosphorylation in complicated biological samples.

4.2 Experimental Methods

4.2.1 Materials

Angiotensin (rat, DRVYIHP) and fibrinogen related peptide (GQQHHLGGAKQAGDV) were purchased from American Peptide Company (Sunnyvale, CA). Phosphoryl chloride, ammonium hydroxide, acetone, ethanol and ether, used for potassium phosphoramidate synthesis were obtained from Sigma Aldrich (St. Louis, MO). NaI, chloramine-T, $Na_2S_2O_5$ for iodination were purchased from Fisher Scientific (Fairlawn, NJ). All the solvents used in the current work, including acetonitrile (ACN), methanol, acetic acid, and trifluoroacetic acid (TFA) were purchased from Fisher Scientific (Fairlawn, NJ).

4.2.2 Phosphorylation of histidines in peptides

Peptides containing phosphorylated histidine are not commercially available; therefore, potassium phosphoramidate was synthesized [34] to phosphorylate histidine residues in peptides. 1mg purified potassium phosphoramidate was added to 50μL of 0.1mM peptide solution with 100mM Tris (pH ~9) and incubated at room temperature for 24 hours. Desalting of the synthesized phosphorylated peptides was carried out as explained below.

4.2.3 Iodination of peptides

To iodinate 10nmol Angiotensin (DRVYIHP) for example, I_2 as an active iodination reagent was synthesized first. 60nmol chloramine-T mixed with 200nmol NaI was diluted to 40μL in a 1.5ml Eppendorf vial and placed in the dark for 10 min. 20μL of I_2 solution was spiked into Angiotensin in Tris buffer (50mM, pH~9) to react for 20min, followed by 5 aliquots of 20μL added every 10min. The entire reaction time was set as 100min and finally 50nmol $Na_2S_2O_5$ was used to quench the left over I_2. For iodination of Angiotensin, the quantitative ratio of I_2 and the reactive sites in the peptide was 1.5, which was kept as the same for iodination of all the other peptides, including those with phosphorylated histidines.. This optimized iodination procedure can completely iodinate all the available tyrosines and histidines in peptides without causing extensive side reactions.

4.2.4 Desalting of phosphorylated and iodinated peptides

Both the phosphorylated and iodinated peptides were desalted using a peptide macrotrap (Michrom Bioresources, Inc. Auburn, CA). The peptide macrotrap was first equilibrated with 200μL 98/1.99/0.01 H_2O/ACN/TFA solution injected at a rate of 200μL/min, followed by peptide loading. 1mL of the 98/1.99/0.01 H_2O/ACN/TFA solution was loaded to remove the salts left in the peptide. Finally, the desalted peptide was eluted using 100μL of 10/89.9/0.01 H_2O/ACN/TFA solution and lyophilized to remove TFA for subsequent use. The entire desalting procedure only took about five minutes.

4.2.5 Protein enzymatic digest and iodination

1nmol of Cytochrome c (Cytc) and Ovalbumin (Oval) were reconstituted individually in 30μl 50mM Tris buffer (pH~8.5) containing 8M Urea. For Oval, DTT was added to make a final concentration of 10mM in order to reduce the disulfide bonds. The two protein buffers were incubated in a 54 ℃ H2O bath for 45min. Iodoacetamide in two times stoichimetric amount of DTT was added, followed by a 45min incubation in the dark under 54℃ to block all the reduced cysteins and left-over DTT in the Oval sample. After denaturation and cystein protection, both Cytc and protected Oval solution was diluted to 120μl and added with trypsin (1:50 ratio of proteins). Both proteins were incubated under 37℃ overnight.

After tryptic digest, protein digests were desalted using the peptide macrotrap before iodination reaction. The desalted Cytc and Oval digests were lyophilized and reconstituted in 1ml and 2ml 50mM Tris buffer (pH~9) respectively. Excess NaI was added to achieve a concentration 2mM. Chloramine-T in 2.5 times stoichimetric amount of all iodination sites in the samples was mixed with excess amount NaI for 10min to generate I_2. The I_2 solution was split into 6 aliquots and spiked into the prepared protein digests every 10 min. The total reaction time was 100min to drive the iodination to completion. $Na_2S_2O_5$ in 1.5 times of the I_2 amount was added to quench the excess I_2. The iodinated Cytc and Oval tryptic digests were concentrated and reconstituted in 20μl H_2O+0.1% TFA for LC/MS analysis.

4.2.6 Peptide MS analysis and LC/MS analysis

7μM Peptide solutions in 50/50 H_2O/ACN were directly infused to a Thermo LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) with a standard electrospray ion source (ESI). Ions of interest were isolated and subjected to CID for sequence or modification site investigation. The isolation window width was set to 3Da. The activation energy was set to diminish 70% of the original parent ion.

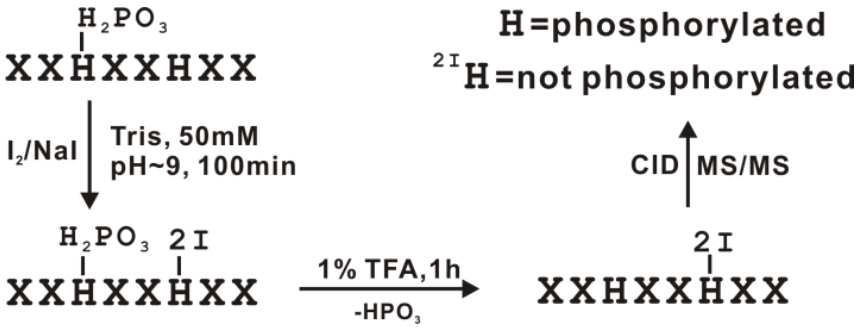
An Agilent 1100 series HPLC (Agilent, Santa Clara, CA) was coupled to the Thermo LTQ linear ion trap mass spectrometer. 20μl of 1nmol protein digests were injected onto a C18 column (150mm×2.1mm, particle size 5μm, Thermo Scientific San Jose, CA) respectively for LC/MS analysis. Mobile phase A was 0.1% TFA in water and mobile phase B was 0.1% formic acid in acetonitrile. The LC gradient is set as follows: flow rate at 200μl/min using 5%B to 40%B in 50min, then 40%B to 75%B in 15min. The MS instrument was operated in a data-dependent mode: the first survey MS (scan 1) from m/z 300 Da to 2000 Da followed by two consecutive independent steps: Ultra-Zoom (scan 2) and CID-MS^2 (scan 3), which are both dependent on ions from scan 1. The CID-MS^2 step employed the default instrument parameters: isolation window width 3.5 Da and activation energy 35%.

4.3 Results and discussion

4.3.1 Iodine-labeling strategy

Iodination can be used to lock in site specific information about phosphohistidine as shown in the example in Scheme 4.1. A model peptide $XX_pHXXHXX$ containing both

phosphorylated ($_pH$) and unmodified (H) histidine is used to demonstrate the general approach; X is any other amino acid residue. In order to map the phosphorylated site, $XX_pHXXHXX$ is subjected to iodination. The free histidine will be doubly iodinated, whereas the phosphorylated histidine will remain unchanged. Iodo-histidine is stable over a wide pH range, whereas the acid-labile phosphoramidate bond is completely hydrolyzed after one-hour acid incubation. Sites of histidine phosphorylation can then be deduced from the presence/absence of iodination, with iodinated residues corresponding to unmodified histidines and free histidines corresponding to previously phosphorylated residues. Tandem mass spectrometry can be utilized to localize the original PTM sites. Tyrosine is also frequently present in proteins and peptides and is more easily iodinated than histidine.[35] Thus, tyrosine residues which have not been phosphorylated or sulfated will also be iodinated, but this does not interfere with the subsequent analysis of histidine phosphorylation.



Scheme 4.1 Demonstration of the iodine-labeling based strategy for identification of phosphorylated histidine sites in peptides

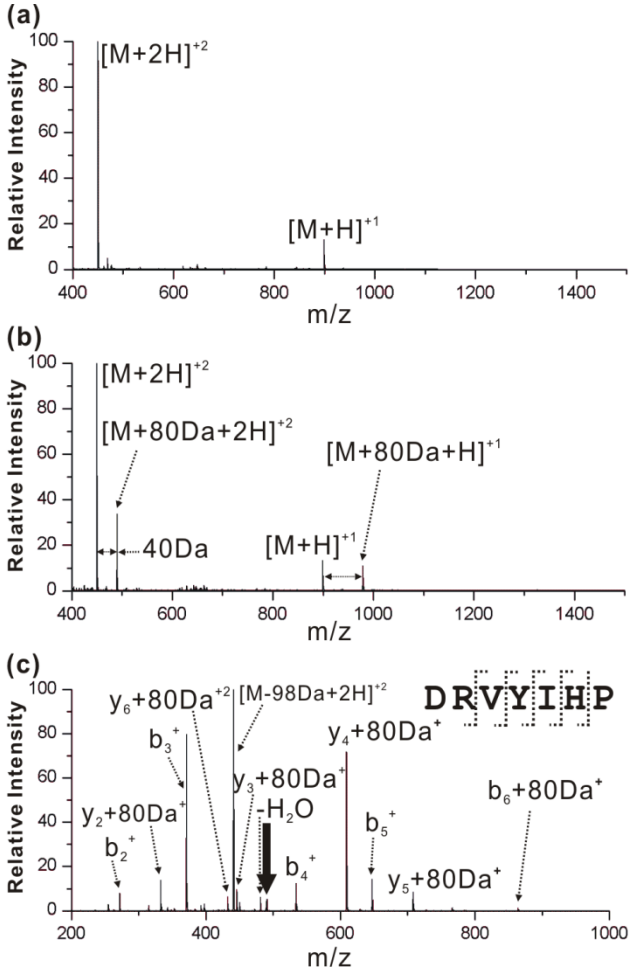


Figure 4.1 (a) Full MS of Angiotensin (DRVYIHP) in 50/50 H_2O/MeOH. (b) Full MS of purified phosphorylated DRVYIHP (c) CID spectrum of +2 charge state of phosphorylated Angiotensin ($[M+80Da+2H]^{+2}$), down arrow indicates the precursor ion.

4.3.2 Model peptides

The full mass spectrum of Angiotensin (DRVYIHP) is shown in Figure 4.1a. Application of potassium phosphoramidate yields the mass spectrum shown in Figure 4.1b. After rapid desalting, the phosphorylated peptide is partially recovered, corresponding to the ions with a mass shift of 80Da. The ratio of the phosphorylated

peptide over the unmodified peptide for the +1 charge state is approximately 1, which is obviously different from the ratio of ~0.3 for +2 charge state. This disparity suggests that phosphorylation may substantially influence electrospray ionization efficiencies in some peptides. It is likely that the actual abundance of the phosphorylated histidine is greater than it appears in Figure 4.1b due to several factors. Histidine is normally a basic residue and potential protonation site in positive ion electrospray. Phosphorylation not only prevents protonation, but also converts histidine into an acidic residue and likely site of deprotonation, which is known to cause ion suppression in phosphorylated residues.[36,37]

The $[M+80Da+2H]^{+2}$ ion was subjected to CID, as shown in Figure 4.1c. The most abundant fragment in Figure 4.1c is a 98Da neutral loss. Previous work suggests that this 98Da loss can result from consecutive losses of HPO_3 and H_2O.[38,39] Although phosphate is lost from the parent ion during CID, a fortuitous series of sequence informative b- and y-type fragments are also observed with phosphate retained. In Figure 4.1c, the presence of b_2^+ through b_5^+ indicates that none of the residues on the N-terminal side of the histidine are phosphorylated. In contrast, y_2^+ and y_3^+ are exclusively phosphorylated, suggesting that histidine is selectively phosphorylated.

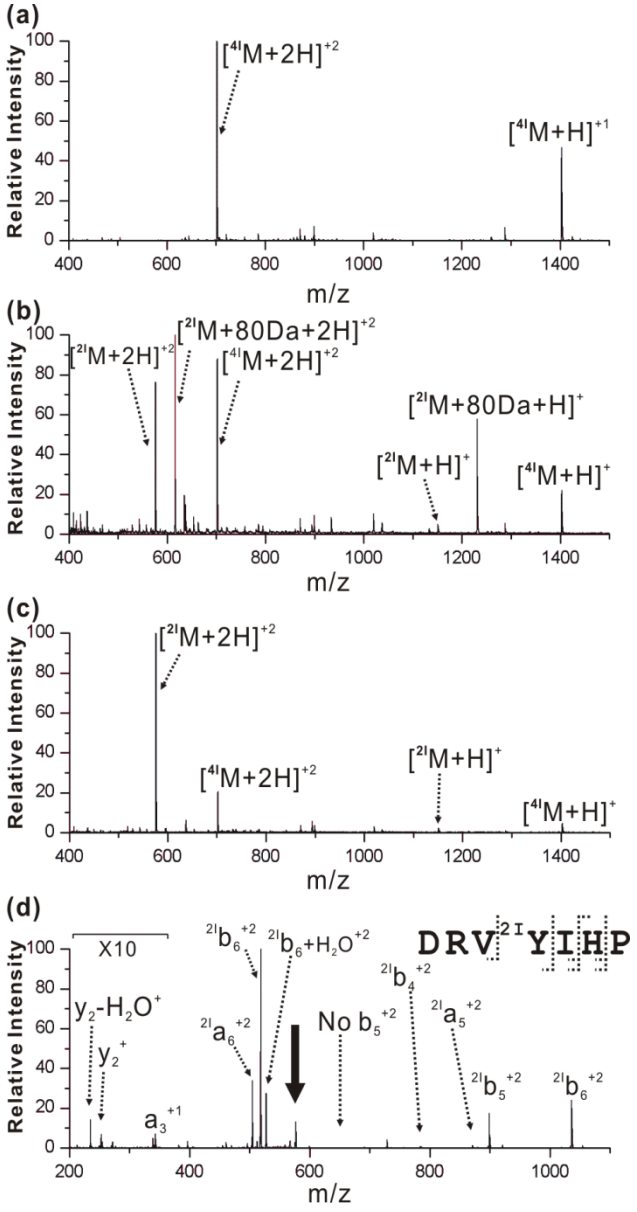


Figure 4.2 (a) Full MS of Angiotensin after iodination. (b) Full MS of phospho-Angiotensin after iodination. (c) Full MS of the iodine-labeled Angiotensin after dephosphorylation. (d) CID mass spectrum of the +2 iodine-labeled Angiotensin after dephosphorylation ($[DRV^{2I}YIHP+2H]^{+2}$), down arrow indicates the precursor ion.

After the phosphohistidine containing angiotensin was successfully synthesized, iodination chemistry was used to independently determine the phosphorylation site. Unmodified Angiotensin was first iodinated as a control; the mass spectrum is shown in Figure 4.2a. Using optimal iodination conditions, four iodines were coupled to Angiotensin and CID confirmed both tyrosine and histidine were doubly iodinated (data not shown). The recovered phosphorylated angiotensin was then treated with the same iodination conditions and subjected to MS analysis, as shown in Figure 4.2b. The resulting spectrum reveals that a significant amount of phosphorylation survives the iodination procedure. Four iodine atoms add to the unmodified angiotensin as observed before. However, only two iodine atoms add to the peptide which has retained the phosphate. A significant amount of doubly iodinated non-phosphorylated peptide is observed as well, which likely originates from the loss of phosphate during desalting. After one-hour acid incubation, the phosphate is completely removed and only iodinated peptides remain as seen in Figure 4.2c. The $[^{2I}M+2H]^{+2}$ ions were subjected to CID as shown in Figure 4.2d. The b_4^{+2} and b_5^{+2} ions include tyrosine but not histidine and are both exclusively doubly iodinated, indicating that tyrosine contains both iodines. The y_2^+ ion which is the complementary fragment of b_5, is observed without iodine modification, confirming that the histidine is not iodinated. The observation of free histidine reveals that the peptide was originally phosphorylated at histidine, in agreement with the result in Figure 4.1. Therefore, iodine labeling and tandem MS can successfully identify the original phosphorylation site in the phosphorylated angiotensin.

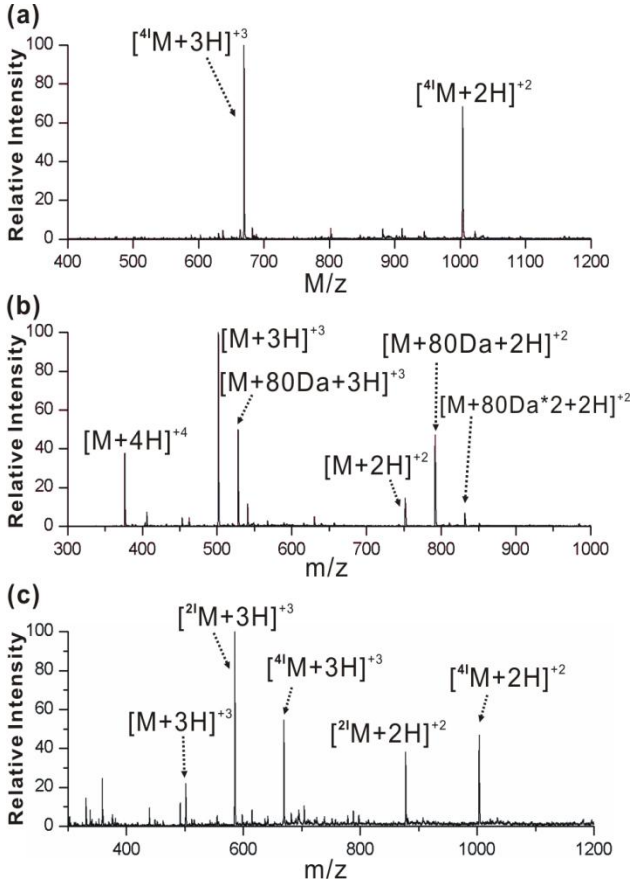


Figure 4.3 (a) Full MS of iodination of the unmodified GQQ as a control. (b) Full MS of the phosphorylated GQQ. (c) Full MS of iodine-labeled GQQ, with phosphate removed after iodination.

Phosphorylated fibrinogen related peptide (GQQHHLGGAKQAGDV, abbreviated as GQQ) which contains two histidines was also examined. The unmodified GQQ peptide can be completely iodinated as shown in Figure 4.3a. The full MS of the modified GQQ after reacting with the phosphorylation reagent is shown in Figure 4.3b. The peptide picks up one or two phosphates, leading to multiple 80Da mass shifts relative to the unmodified GQQ. The relative abundance of the modified GQQ peptides again varies as a function of charge state, indicating ion suppression of the negative phosphate

containing peptides. In order to elucidate the sites of phosphorylation, iodination was then applied to the phosphorylated GQQ mixture, followed by dephosphorylation. The results are shown in Figure 4.3c. GQQ modified with zero, two and four iodines are all observed, in contrast to the iodination control in Figure 4.3a, suggesting the presence of histidine phosphorylation. The ^{4I}GQQ species originates from the unmodified GQQ in the phosphorylated peptide mixture. The ^{2I}GQQ and GQQ ions in Figure 4.3c represent the singly and doubly phosphorylated peptides in which one or both histidines are shielded from iodination due to phosphorylation.

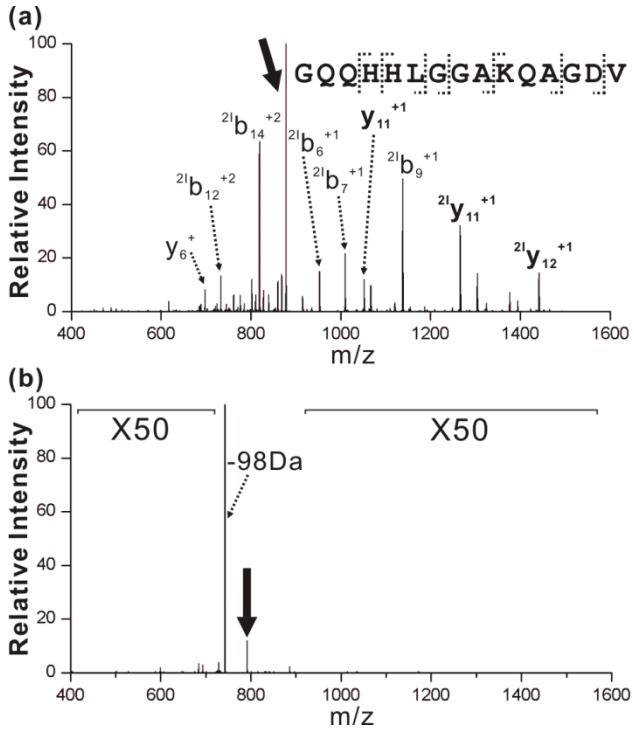


Figure 4.4 (a) CID of $[^{2I}GQQ+2H]^{+2}$ derivatized from $[^{2I}GQQ+80Da+2H]^{+2}$. (b) CID spectrum of $[GQQ+80Da+2H]^{+2}$.

Due to the presence of two histidines, the site of iodination in ^{2I}GQQ is ambiguous, which also makes the original phosphorylation site(s) unclear. CID was employed to fragment ^{2I}GQQ in order to obtain site specific information. The resulting mass spectrum is shown in Figure 4.4a. Several doubly iodinated b- and y-type fragments are produced by CID. The y_{11} ion corresponds to cleavage between His4 and His5 and is found in both the free and doubly iodinated forms. The ratio reveals that phosphorylation occurs to a larger extent on His5. The y_{12} ion is found only in the doubly iodinated state, confirming that iodination occurs only at the histidine residues. CID of mono-phosphorylated GQQ itself was also carried out in an attempt to obtain phosphorylation site information directly. Unfortunately, as seen in Figure 4.4b, the almost exclusive fragment is a 98Da neutral loss, which is more typical that the results in Figure 4.1c due to the lability of phosphoramidate bonds in the gas phase. Therefore in the case of GQQ, iodination is not only convenient for guarding against PTM lability in solution, but also needed if site specific identification by CID is desired.

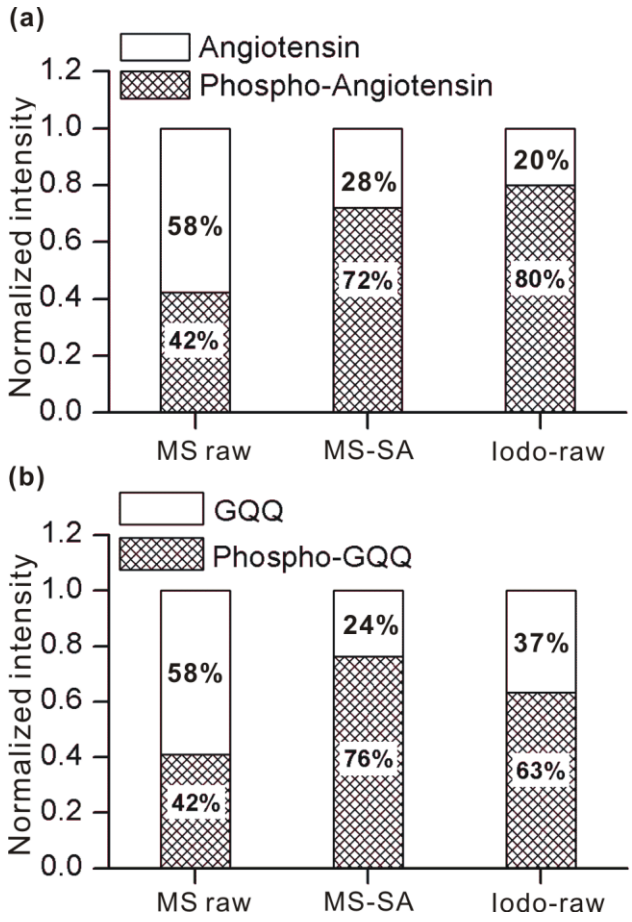


Figure 4.5 (a) Quantitation of phosphorylated Angiotensin; (b) Quantitation of phosphorylated GQQ. MS raw indicates relative ion intensities from mass spectrum; MS-SA refers to quantitation based on standard addition; Iodo-raw indicates abundance derived from intensities following iodination.

4.3.3 Quantitation

It is clear from the results above that iodination preserves phosphorylation site information in peptides; however quantitative aspects of this chemistry have not yet been explored. For phosphorylated angiotensin, the relative ratios of the modified and free peptides based on raw ESI ion intensities from Figure 4.1b are plotted in Figure 4.5a (MS-raw). Although 42% of the ion intensity corresponds to phosphorylated angiotensin, this value is likely inaccurate due to the ion suppression effects mentioned above. To compensate for ionization effects, a known amount of unmodified angiotensin was added to the phosphorylated sample, and the corresponding increase in relative ion count was recorded. Implementing the standard addition correction, the relative amount of the phosphopeptide is then found to be ~72% (MS-SA). Quantitation based on the raw ion counts following iodination yields a similar result of ~80% phosphopeptide (Iodo-raw). This does not take into account differences in the relative ionization efficiencies for the different iodinated peptides, but these differences are expected to be smaller in magnitude as there is no charge inversion associated with iodination. A similar comparison for the phospho-GQQ peptide is shown in Figure 4.5b. Agreement between methods is again reasonable, although the effect of iodination on ion count is in the opposite direction. In a complicated sample, such ionization effects would be a function of each individual sequence and therefore difficult to predict, suggesting that quantitation by this method is likely to be limited to providing information on the order of ±20% of the actual value. Given that no quantitative method can currently provide site specific information, this semi-quantitative information may still be useful.

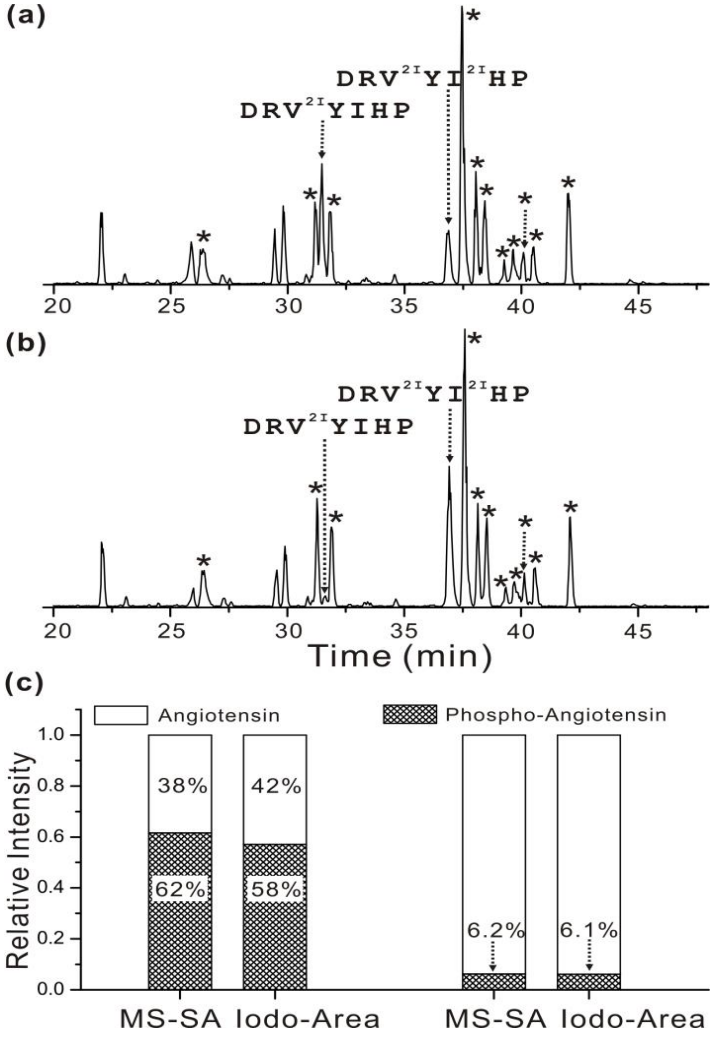


Figure 4.6 (a) LC/MS ion chromatography (IC) of iodination of 1nmol Cytc tryptic digests and 1nmol phosphorylated Angiotensin mixture. (b) LC/MS IC of iodination of 1nmol Cytc trypic digests with 0.1nmol phosphorylated Angiotensin mixture and 0.9nmol Angiotensin. (c) Quantitation of phosphorylated Angiotensin in the Cytc tryptic mixture based on the standard addition (SA) method and peak area integral of the two iodinated Angiotensin peaks in Figure 4.6a and Figure 4.6b.

4.3.4 Whole proteins

To further explore the applicability of this technique, several additional experiments were carried out with whole proteins. Cytochrome c and Ovalbumin were digested with trypsin and then iodinated. Among the 40 peptides that were observed for the two proteins, 17 out of 18 iodinateable peptides were fully iodinated. The single case of incomplete iodination is easily rationalized. This peptide (VHHANENIFYCPIAIMSALAMVYLGAK) has a molecular weight of nearly 3 kDa and comprises 27 residues, four of which can be iodinated. It is well known that iodination is influenced by the protein secondary and tertiary structure, causing less accessible tyrosine or histidine residues to be less easily iodinated.[40,41] Large digest fragments may retain elements of structure and interfere with complete iodination (for the peptide above, iodination states incorporating 4, 6 and 8 iodines were observed). To avoid possible false positives, the method described in this manuscript should only be applied to peptides which are smaller than ~20 residues. The sequence of the protein of interest should therefore be consulted to select the appropriate protease(s) to ensure suitable peptide size. The size limitation applies only to residues that contain histidine as false positives cannot otherwise be obtained.

Quantitative aspects were also explored in the context of a whole protein digest. In Figure 4.6a is the total ion chromatogram for a tryptic digest of Cytochrome c with an equimolar amount of angiotensin added in. Angiotensin is present in both phosphorylated and canonical form, and the entire mixture was subjected to iodination. The extent of iodination closely correlates with the degree of phosphorylation as seen in Figure 4.6c,

where the actual amount of phosphorylation as determined by standard addition (MS-SA) is compared with the amount determined by iodination (Iodo-area). In this case, quantitative results were extracted by integrating the area under the relevant peaks in the ion chromatogram. The experiment was repeated with $1/10^{th}$ the amount of phosphorylated peptide (corresponding to 6% phosphorylated peptide overall). This substoichiometric amount was easily detected. All nonphosphorylated peptides were fully iodinated except those which were not iodinateable. These experiments confirm that our method is applicable to more complex environments.

4.4 Conclusion

Histidine phosphorylation is an important PTM which is abundant in both prokaryotes and eukaryotes. The high instability of this modification significantly complicates the use of typical LC-MS/MS based proteomics techniques to identify histidine phosphorylation sites. We have demonstrated that iodination labeling is able to lock in site specific information about the locations of phosphohistidines in peptides. Standard LC methods employing acidic conditions and tandem mass spectrometry, including CID, can then be implemented without modification to identify sites of phosphorylation. The method is demonstrated to be semi-quantitative, enabling relative abundances to be determined within approximately ±20%. This method should provide an avenue for wide scale examination of histidine phosphorylation in complicated samples.

---

[1] Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002**, 298, 1912-1918.

[2] Blume-Jensen, P.; Hunter. T. *Nature* **2001**, 411, 355-365.

[3] Ubersax, J. A.; Ferrell J. E. Jr. *Nat. Rev. Mol. Cell Biol.* **2007**, 8, 530-541.

[4] Nita-Lazar, A.; Saito-Benz, H.; White, F. M. *Proteomics* **2008**, 8, 4433-4443.

[5] Pirrung, M. C. *Chem. Biol.* **1999**, 6, 166-173.

[6] Matthews, H. R. *Pharmac. Ther.* **1995**, 67, 323-350.

[7] Hoch, J. A.; Silhavy, T. J. (Eds) *Two-component signal transduction.* Washington, DC: ASM Press, 1995.

[8] Chang, C.; Stewart, R. C. *Plant Physiol.* **1998**, 117, 723-731.

[9] Saito, H. *Chem. Rev.* **2001**, 101, 2497-2509.

[10] Stock, A. M.; Robinson, V. L.; Goudreau, P. N. *Annu. Rev. Biochem.* **2000**, 69, 183-215.

[11] Alex, L.; Simon, M. I. *Trends Genet.* **1994**, 10, 133-138.

[12] West, A. H.; Stock, A. M. *Trends Biochem. Sci.* **2001**, 26, 369-376.

[13] Appleby, J. L.; Parkinson, J. S.; Bourrett, R. B. *Cell* **1996**, 86, 845-848.

[14] Perraud, A. L.; Weiss, V.; Gross, R. *Trends Microbiol.* **1999**, 7, 115-120.

[15] Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. *Nat. Biotechnol.* **2002**, 20, 301-305.

[16] Goshe, M. B.; Conrads, T. P.; Panisko, E. A.; Angell, N. H.; Veenstra, T. D.; Smith, R. D. *Anal. Chem.* **2001**, 73, 2578-2586.

[17] Oda, Y.; Nagasu, T.; Chait, B. T. *Nat. Biotechnol.* **2001**, 19, 379-382.

[18] Salih, E. *Mass Spectrom. Rev.* **2005**, 24, 828-846.

[19] Sickman, A.; Meyer, H. E. *Proteomics* **2001**, 1, 200-206.

[20] Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, 120, 3265-3266.

[21] Cooper, H. J.; Hakansson, K.; Marshall, A. G. *Mass Spec. Rev.* **2005**, 24, 201-222.

[22] Syka, J. E.P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U S A* **2004**, 101, 9528-9533.

[23] Swaney, D. L.; Wenger, C. D.; Thomson, J. A.; Coon, J. J. *Proc. Natl. Acad. Sci. U S A* **2009**, 106, 995-1000.

[24] Noiman, S.; Shaul, Y. *FEBS Lett.* **1995**, 364, 63-66.

[25] Ham, B. M.; Yang, F.; Purvine, S. O.; Zhao, R.; Smith, R. D.; Lipton. M. S. 57th ASMS conference on mass spectrometry and allied topics, Philadelphia, PA.

[26] Kleinnijenhuis, A. J.; Kjeldsen, F.; Kallipolitis, B.; Haselmann, K. F.; Jensen, O. N. *Anal. Chem.* **2007**, 79, 7450-7456.

[27] Diedrich, J. K.; Julian, R. R. *J. Am. Chem. Soc.* **2008**, 130, 12212-12213.

[28] Bonda, M. R.; Kohler, J. J. *Curr. Opin. Chem. Biol.* **2007**, 11, 52-58.

[29] Yu, Y.; Hoffhines, A. J.; Moore, K. L.; Leary. J. A. *Nature Methods* **2007**, 4, 583-588.

[30] Becker, S.; Laffont, S.; Vitry, F.; Rolland, Y.; Lecloirec, J.; Boucher, E.; Raoul, J. L.; Herry, J. Y.; Bourguet, P.; Garin, E. *Nuc. Med. Commun.* **2008**, 29, 815-825.

[31] Langen, K. J.; Roosen, N.; Coenen, H. H.; Kuikka, J. T.; Kuwert, T.; Herzog, H.; Stöcklin, G.; Feinendegen, L. F. *J. Nucl. Med.* **1991**, 32, 1225-1228.

[32] Jager, P. L.; Vaalburg, W.; Pruim, J.; de Vries, E. G.; Langen, K.; Piers, A. D. *J. Nucl. Med*. **2001**, 42, 432-445.

[33] Covelli, I.; Wolff, J. *J. Bio. Chem.* **1967**, 242, 881-886.

[34] Wei, Y. F.; Mattews, H. R. *Methods Enzymol.* **1991**, 200, 388-414.

[35] Ramachandran, L. K. *Chem. Rev.* **1956**, *56*, 109-218.

[36] Steen, H.; Jebanathirajah, J. A.; Rush, J.; Morrice, N.; Kirschner, M. W. *Mol. Cell Proteomics* **2006**, 5, 172-181.

[37] Craig, A. G.; Hoeger, C. A.; Miller, C. L.; Goedken, T.; Rivier, J. E.; Fischer, W. H. *Biol. Mass Spectrom.* **1994**, 23, 519-528.

[38] DeGnore. J. P.; Qin, J. *J. Am. Soc. Mass Spectrom.* **1998**, 9, 1175-1188.

[39] Tholey, A.; Reed, J.; Lehmann, W. D. *J. Mass Spectrom.* **1999**, 34, 117-123.

[40] Sun, Q. Y.; Yin, S.; Loo, J. A.; Julian, R. R. *Anal. Chem.* **2010,** *82*, 3826-3833.

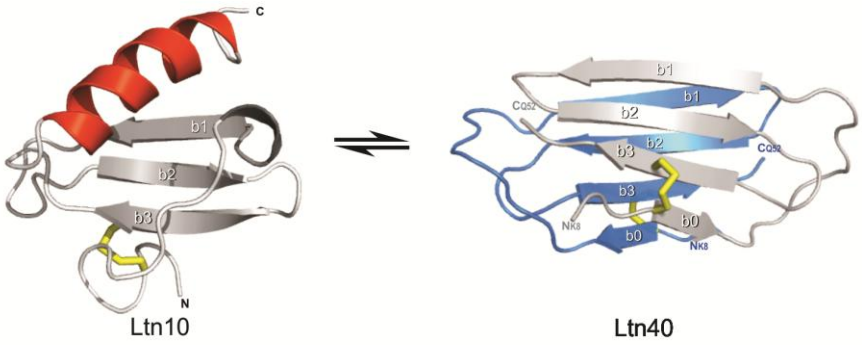
[41] Santrucek, J.; Strohalm, M.; Kadlcik, V.; Hynek, R.; Kodicek, M. *Biochem. Biophys. Res. Commun.* **2004**, 323, 1151-1156.

CHAPTER 5

DYNAMIC INTERCHANGING NATIVE STATES OF LYMPHOTACTIN EXAMINED BY SNAPP-MS

5.1 Introduction

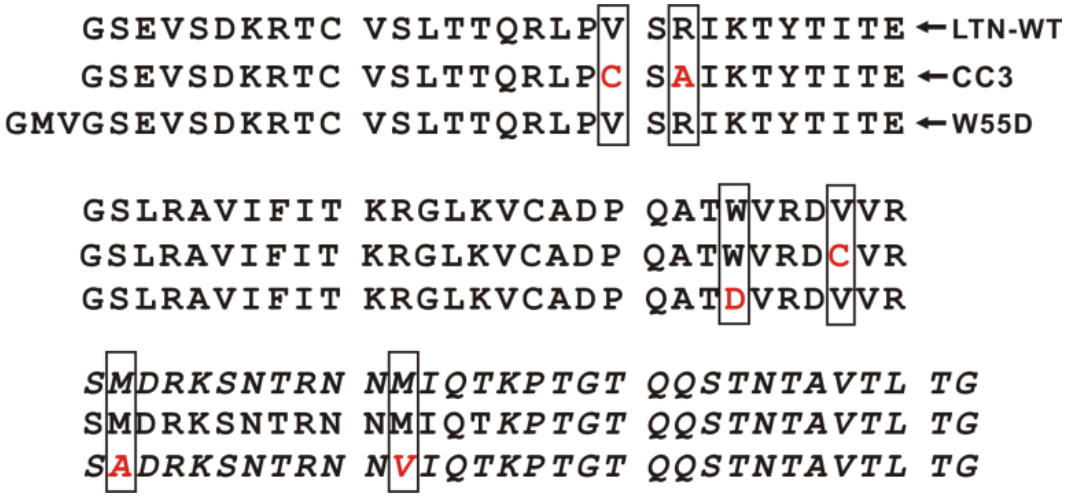
Lymphotactin is an unusual protein in many ways. It is a human chemokine which can adopt *two* natively folded states, which are both functionally relevant *in vivo*.[1] The two native state structures are very dissimilar as shown in Scheme 5.1.[2] One structure represents the typical chemokine fold, a single alpha helix adjacent to three beta-sheet strands, and is referred to as the Ltn10 structure. The other structure is known as Ltn40 and is characterized by dimerization of two proteins, with both adopting four beta-sheet strands oriented in a symmetrical head to tail fashion. Although both structures share beta-sheet motifs, the composition of the beta sheets is quite different between the two structural forms. Similarly, the hydrophobic cores and other tertiary and quaternary structural characteristics are essentially entirely dissimilar. Under typical biological conditions, both forms of the wild type protein exist in equilibrium and interconvert on a timescale of seconds.[2] In the Volkman laboratory, two mutants have been developed which adopt only one of the two structural motifs.[3] The CC3 mutant contains an additional disulfide bond which favors the typical Ltn10 structure. Disruption of the Ltn10 hydrophobic core in the W55D mutant pushes the structural equilibrium strongly towards the Ltn40 structure. These mutants can be utilized to examine aspects of each structure independently.



Scheme 5.1 The two structures of Lymphotactin.

Selective noncovalent adduct protein probing (SNAPP) is a recently developed method for examining protein structure in solution.[4] This method relies on specific noncovalent interactions to probe protein structure. Experiments are conducted by introducing a reagent which binds weakly in solution, such as 18-crown-6 ether (18C6), which becomes strongly attached to the protein in the gas phase. 18C6 attaches to lysine in the gas phase due to three specific hydrogen bonds between the protonated side-chain amine and alternating oxygen atoms in the crown ether. The binding energy for this attachment is ~54kcal/mol.[5] Less specific, although fairly strong noncovalent (42kcal/mol) associations between 18C6 and the protonated side chain of arginine can form as well. SNAPP experiments are conducted by subjecting an aqueous solution containing protein and 18C6 to electrospray ionization (ESI). Source conditions which favor both complete desolvation and retention of noncovalent adducts must be employed. During the process of ESI, binding interactions between 18C6 and protonated side chains transition from weak to strong as the protein is rapidly desolvated. Previous experiments

have determined that the local chemical environment surrounding each lysine residue strongly influences the probability for binding 18C6, with proximate salt bridge or hydrogen bonding interactions interfering the most.[6] Structural information about the protein is therefore obtained because the local chemical environment surrounding each residue is a function of protein structure.[4,7] In the end, a distribution of peaks (known as a SNAPP distribution) representing the number of 18C6s attached to the protein is observed in the mass spectrometer and is characteristic of a particular protein conformation. Changes to the protein structure will typically lead to changes in the SNAPP distribution; therefore the strength of SNAPP is in measuring structural differences for either static or dynamic systems.



Scheme 5.2 Sequences for each Lymphotactin variant.

SNAPP is closely related to other mass spectrometry (MS) based protein structure determination methods such as charge state distribution analysis,[8-10] hydrogen/deuterium (H/D) exchange,[9-11] and covalent labeling[12], but also differs in several important ways. H/D exchange monitors protein structure as a function of the exchange of amide hydrogens. The overall number, locations, and rates of exchanges can be used to interrogate structure.[13] Various covalent labeling methods can also be used to probe structure, although in this case reactions are ideally restricted to a single modification to avoid perturbation of the target structure.[14] The primary difference between SNAPP and both of these methods is that the reporting chemistry in SNAPP is highly reversible (if present at all) in solution and information is only encoded during the later stages of ESI when the protein is transitioning into the gas phase. This enables SNAPP to probe highly dynamic systems where structural changes may be occurring rapidly in solution, and on the same timescale as H/D exchange or covalent labeling. If, for example, a protein adopted an unfolded and folded state under given conditions, the SNAPP distributions for both would be determined independently during ESI and observable in different charge states (due to the different sizes) as long as the structural transition took longer than a few milliseconds (the time required to desolvate the protein). SNAPP is therefore an appropriate method to interrogate the structural features of Lymphotactin, which undergoes structural changes on a significantly longer timescale.[2]

Herein it is demonstrated that the Ltn10 and Ltn40 structures yield easily distinguishable SNAPP distributions. The wild type protein yields results that are most similar to the W55D variant. The addition of acid or organic denaturants does not have a

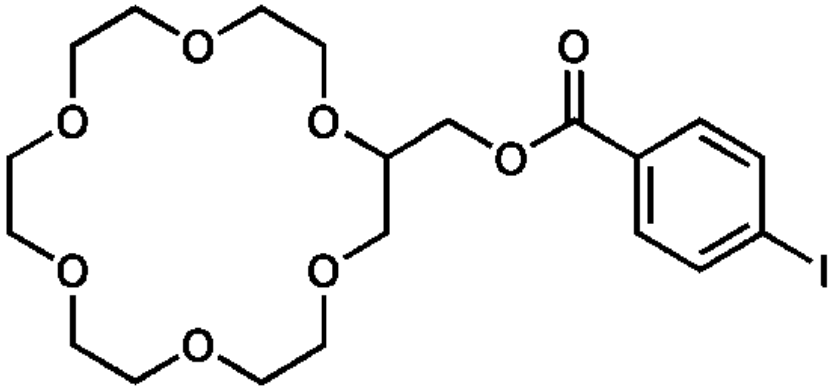
large impact on the SNAPP distributions for the wild type, CC3, or W55D proteins. Reduction and blocking of the disulfide bonds affects the structures more dramatically and yields shifts in the SNAPP distributions that are consistent with significant unfolding. Data obtained from NMR experiments supports this conclusion. The effects of cold temperature on all three proteins are explored. Finally, a combination of SNAPP and radical chemistry is utilized to compare the gas phase structures of the wild type, CC3, and W55D proteins.

5.2 Experimental methods

5.2.1 Materials

Recombinant human Lymphotactin proteins used in SNAPP study were expressed and purified as previously described.[15] Lymphotactin mutants CC3 and W55D were prepared by site-directed mutagenesis using the Stratagene QuickChangeTM kit following established protocols. Purified proteins were lyophilized and stored at -20˚C for subsequent study. 18-crown-6 ether used for SNAPP studies was purchased from Alfa Aesar (Ward Hill, MA). Dithiolthreitol (DTT) and iodo-acetamide (IAA) were purchased from Sigma Aldrich (St. Louis, MO). H_2O used for SNAPP experiments was purified to 18.2MΩ resistivity using Millipore Direct-Q (Billerica, MA). All the solvents were obtained from Fisher Scientific (Fairlawn, NJ) and used as received. 2-(hydroxymethyliodobenzoylester)-18C6 (IBA-18C6), employed in the radical directed dissociation experiments, was prepared as follows: 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 hour reaction period, a crystalline hair-like precipitate was observed. The precipitate was removed by filtration. The filtrate was then passed through celite and evaporated over nitrogen. The product was recovered as a white solid.



Scheme 5.3 2-(hydroxymethyliodobenzoylester)-18C6 (IBA-18C6)

5.2.2 S-S bond reduction and cysteine protection of Ltn proteins

To reduce disulfide bonds in the Lymphotactin proteins, 60µl solution containing 10nmol protein and excess amount of DTT (10mM) was incubated at 54˚C for 45min. Free cysteine and excess DTT were treated with a stoichiometric amount of IAA (1.22µmol for wild type Ltn and W55D, 1.24µmol for CC3) to block all free thiol groups. The blocked proteins were purified by a protein trap (Michrom Bioresources, Inc. Auburn, CA), lyophilized and stored at -20˚C for subsequent use.

5.2.3 SNAPP experiments

CC3, wild type Ltn, and W55D stock solutions were diluted to 7µM in H_2O respectively. The final concentration of 18C6 in SNAPP solutions was 84µM. Mass spectra were obtained using an LTQ linear ion trap mass spectrometer (Thermo Fisher

Scientific, San Jose, CA) equipped with a standard ESI-source. Protein samples mixed with 18C6 were directly infused into LTQ mass spectrometer. The electrospray parameters, such as spray voltage, sheath gas flow rates, capillary voltage, temperature, etc. were optimized and were similar to parameters described previously for SNAPP experiments.[4] Once optimized, all the parameters were maintained for all SNAPP experiments presented herein. The following are the optimized source parameters for all SNAPP experiments: spray voltage 4.8kV, sheath gas flow rates 11, tube lens 160V, capillary voltage 44V and capillary temperature 275˚C.

5.2.4 Radical directed dissociation of Ltn proteins

5μM Lymphotactin protein was mixed with 15μM IBA-18C6 in 50/50 H2O/MeOH solution and directly infused into the LTQ linear ion trap mass spectrometer. The posterior plate of the LTQ is modified with a quartz window to transit fourth harmonic (266nm) laser pulses from a flash lamp pumped Nd/YAG laser (Continuum, Santa Clara, CA). The laser pulses are synchronized to the end of the isolation step of a typical MS^2 experiment by a TTL trigger signal from the mass spectrometer to the laser vai a digital delay generator (Berkeley Nucleonics, San Rafael, CA). The isolation window width of MS^2-MS^4 experiments was set to 5 Da.

5.3 Results and Discussion

The stack plot in Figure 5.1a shows the mass spectra for Lymphotactin alone acquired in water (front), water/methanol (middle), and water/methanol/acetic acid (back). These solvent systems are increasingly denaturing for most proteins and will typically yield

substantial shifts in the charge state distribution towards higher charge states.[16,17] The increase in charge state is typically rationalized by the notion that denatured proteins are less compact and can therefore accommodate additional charges without increasing Coulombic repulsion. Interestingly, the typical shift towards higher charge states is not observed for Lymphotactin. In fact, the charge state distributions barely change at all. Therefore, charge state distributions alone yield virtually no information about the structure of Lymphotactin. As detailed in the introduction, it is known that Lymphotactin has two highly dissimilar native structures, which are under equilibrium. It is likely that transitioning from one structure to the other requires sampling of an unfolded state due to the large disparity between the two conformations. The data in Figure 5.1a is consistent with the notion that Lymphotactin accesses an unfolded state even under non-denaturing conditions, and therefore the addition of methanol and acid does not significantly shift the charge state distribution because unfolded structures are already being sampled. From this data alone, it is clear that Lymphotactin is an unusual protein.

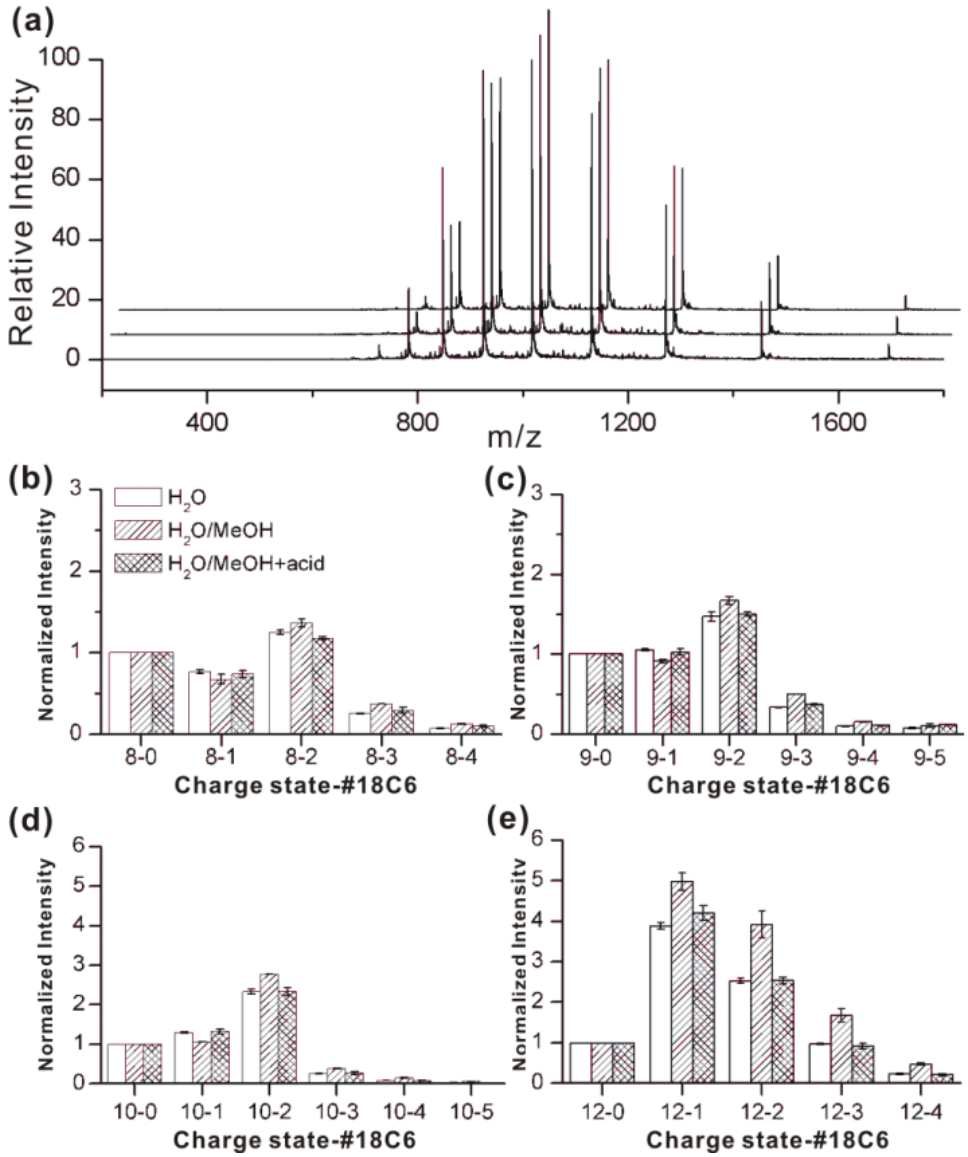


Figure 5.1 (a) ESI-MS spectra for wild type Lymphotactin acquired in water, 50/50 water/methanol, and 49/49/1 water/methanol/acetic acid from front to back, respectively. SNAPP distributions for wild type Lymphotactin in various charge states are shown in (b)+8, (c)+9, (d)+10, and (e)+12.

In Figures 5.1b-d, several representative SNAPP distributions for Lymphotactin in the same series of solvents are shown. These distributions have been extracted from the raw mass spectra and illustrate the number of 18C6 adducts which attach to Lymphotactin, separated by charge state. The addition of 18C6 reveals another layer of structural information, allowing identical charge states to be compared with each other. The SNAPP distributions for the +8 through +10 charge states are fairly similar to each other and are also fairly constant in the various solvents. These distributions most likely represent the more folded conformations of the protein. In the +11 and +12 charge states, the SNAPP distributions begin to shift towards an increasing number of adducts. The average number of 18C6 adducts goes from 1.3 for the +8 charge state to 1.6 for the +12 charge state. Some differences between the various solvent systems also become notable. Interestingly, the number of crown adducts increases with the addition of methanol and then decreases again with the addition of acid. For systems where structural changes are not dominant, it is known that the addition of acid leads to a decrease in crown binding due to competitive binding by the acid.[4] This confirms that acid does not significantly influence the structure of Lymphotactin. Given that the protein is highly basic (6 Lys, 8 Arg), with few acidic residues, this result is perhaps not surprising.

In Figure 5.2, the SNAPP distributions obtained in water for wild type Ltn, Ltn10 (CC3), and Ltn40 (W55D) are shown. There is a clear and reproducible difference between the SNAPP distributions for the CC3 and W55D mutants. Less 18C6 attachment to the W55D protein is observed in every case. SNAPP-MS is therefore easily able to distinguish the two structural variants of Lymphotactin from each other when examined

separately. As shown in Scheme 5.2, these two mutants vary in sequence by only six amino acids (W55D also contains 3 additional N-terminal residues, none of which are targets for 18C6, nor do they alter the Ltn40 conformation). Interestingly, one of the mutations, Ala in CC3 for Arg in W55D, creates an additional target residue for 18C6 binding. Despite this change, 18C6 attachment in W55D is less than that observed in CC3. Therefore, the differential response of SNAPP to these two proteins is clearly due to structural differences. Salt bridges are known to significantly interfere with 18C6 binding.[4] Both the Ltn10 and Ltn40 structures contain two salt bridges (Arg61-Asp64 and Arg35-Glu31 for Ltn10, Arg9-Asp50 and Glu31-Lys25(intermolecular) for Ltn40). However lysine is the preferred target for 18C6, and the transition for Lys25 from being freely available in Ltn10 to being incorporated into a salt bridge in Ltn40 likely accounts for some of the reduced 18C6 binding to the W55D mutant.

Although the wild type protein can adopt either the Ltn10 or Ltn40 structures, it is clear from the data in Figure 5.2 that the wild type Ltn SNAPP distributions are most similar to those from the W55D mutant. The similarity likely results from a combination of two factors. 1) It is known that wild type Ltn favors the Ltn40 structure at low ionic strength.[18] 2) The W55D mutant is able to freely access unfolded states which are expected to be present in wild type Ltn, but the CC3 mutant is significantly constrained by an additional disulfide bond which likely interferes with adopting similar unfolded states. Despite similarity to W55D, wild type Ltn does present some intermediate behavior as well (for example in the 8-2 and 12-1 distributions where the wild type Ltn peaks are more similar to the CC3 values), suggesting that the equilibrium is not entirely

shifted to the Ltn40 structure. It is worth noting that the dimeric association for Ltn40 known to dominate in solution is not observed in the gas phase. Presumably this is due to loss of the hydrophobic effect as the Lymphotactin dimer is desolvated, which is primarily driving the association in solution. Nevertheless, it is clear from the data in Figure 5.2 that the structural information encoded by attachment of 18C6 is locked in prior to the loss of the dimer during the transition to the gas phase.

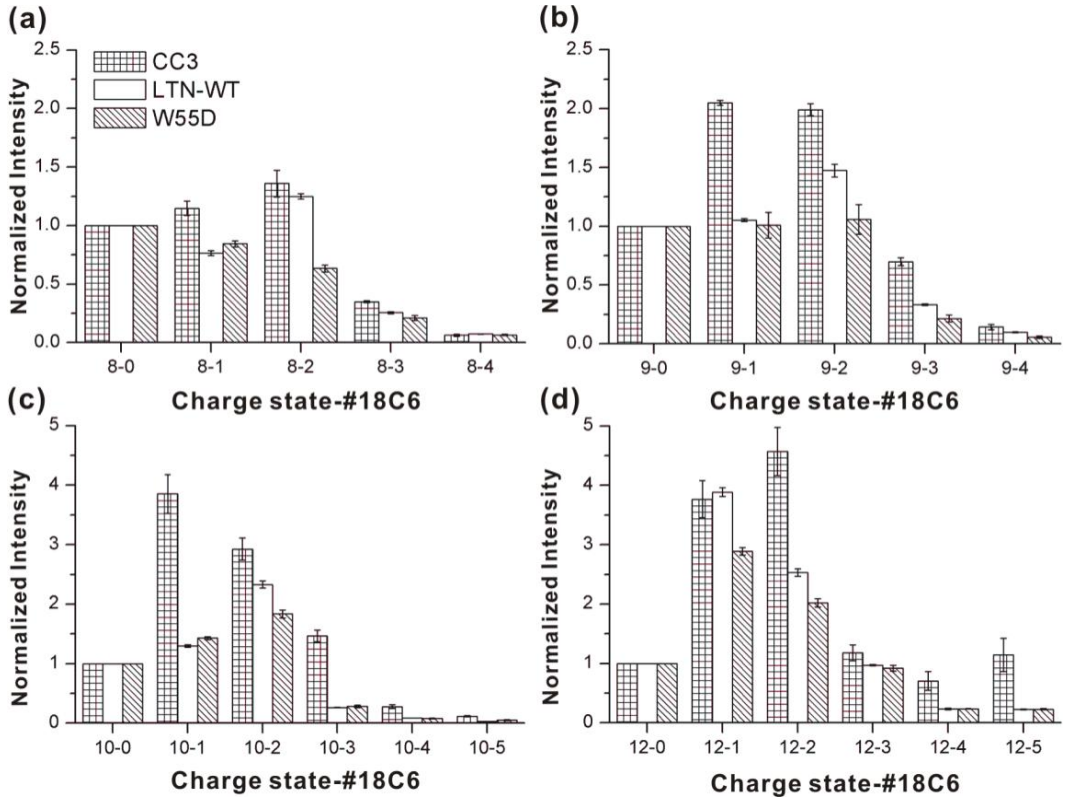


Figure 5.2 SNAPP distributions at different charge states (a) +8 (b) +9 (c) +10 (d) +12, for three variants of Lymphotactin sampled from water. The CC3 and W55D mutants clearly produce distinguishable distributions.

Characterization of the transition between the two forms of Lymphotactin is an active area of interest.[19] The ability of SNAPP to examine and potentially separate out highly dynamic structural features is an important advantage which should facilitate characterization of the transition structures. As described above, the higher charge states most likely represent structural states which are at least partially unfolded and therefore capable of transitioning between the two structural forms. One possibility for transition is that the unfolded protein becomes entirely disordered, and then subsequently refolds into either native state with probabilities determined by the nature of the solution. To investigate this theory, we obtained SNAPP distributions for proteins where the disulfide bonds had been reduced and blocked. Results from NMR experiments indicate that elimination of the disulfide bonds leads to highly disordered states (see supporting information). In Figure 5.3 the average number of 18C6 adducts for blocked and disulfide bound protein are plotted as a function of charge state. Direct comparison reveals reduced 18C6 complexation following disulfide reduction and capping for each protein.

These results can be rationalized by examining potential competitive intramolecular binding sites. Lymphotactin has few acidic residues, suggesting that the primary competition for lysine or arginine side chain binding should originate from hydrogen bonding with the backbone or side chains. When the protein is in the folded state, many backbone hydrogen binding sites will be unavailable due to involvement in secondary structure. For example, in an alpha helix all of the internal amides are hydrogen bound to each other. It has been demonstrated previously that loss of secondary structure to random coil structure can lead to significantly reduced crown attachment due to

competitive hydrogen bonding with the backbone.[4] In the case of Lymphotactin, the reduction in crown binding is consistent with adoption of a highly disordered state where disruption of backbone hydrogen bonding leads to competitive inhibition of 18C6 binding. Comparison of the higher charge states for the blocked and unblocked structures (i.e. +11 and +12 in Figure 5.3) reveals significantly different behavior, suggesting that the transition structures are likely not entirely disordered. Similarly, the average binding numbers for unblocked protein are consistent with results given above and suggest that the wild type Ltn and W55D proteins access similar transition state structures, while CC3 does not (Figure 5.3d). For the blocked proteins, all three have virtually identical average 18C6 binding in the +11 and +12 charge states (Figure 5.3d), suggesting absence of any structural differences for the unfolded, disordered states.

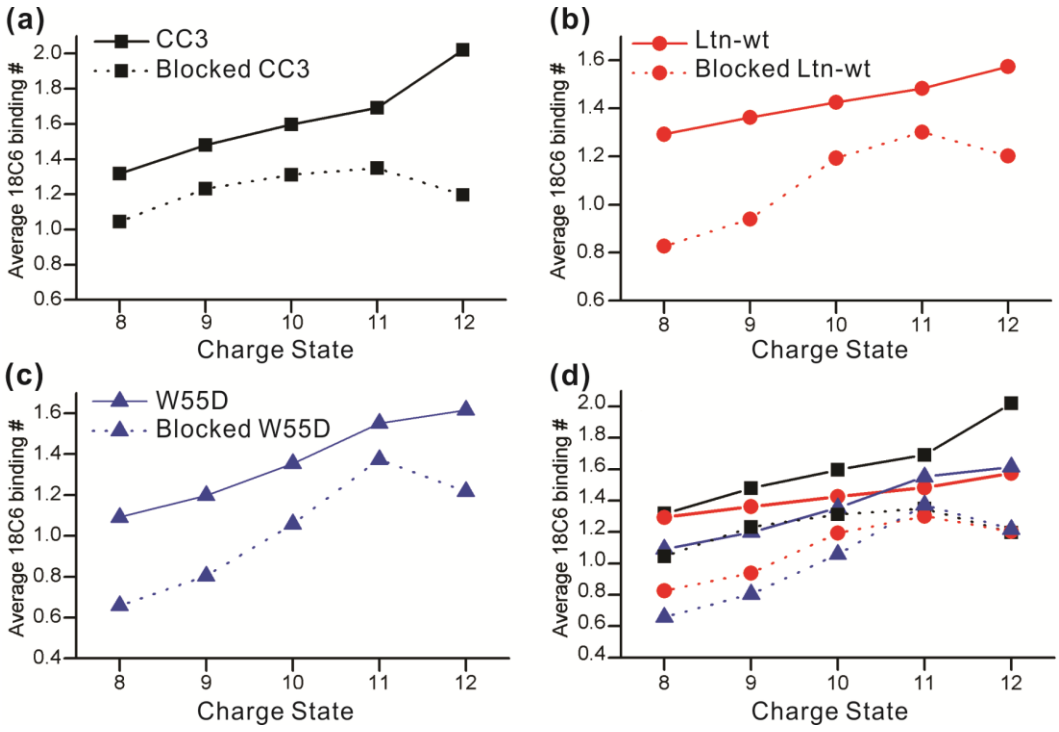


Figure 5.3 The average number of 18C6 adducts is shown as a function of charge state for (a) CC3, (b) wild type Ltn, and (c) W55D. The protected (disulfide reduced and capped) proteins are shown in dotted lines. (d) All proteins are shown together to demonstrate relative binding. The key is identical to a)-c).

Another interesting variable which can produce structural change and is known to influence Lymphotactin is temperature. To examine temperature effects, we simply gathered data on identical solutions at room temperature and with the addition of an ice pack to the syringe containing the protein solution. The temperature of the syringe is ~0 ℃, although the solution will warm slightly passing through the line to the electrospray tip. Results were acquired after 20 minutes equilibration time and are shown in Figure 5.4. Both the CC3 and W55D distributions do not undergo any compelling shifts in 18C6

adduction after application of the ice pack. Interestingly, wild type Ltn does exhibit a decrease in 18C6 binding at lower temperature. It is therefore clear that temperature induced structural shifts can be observed by SNAPP. Conversely, merely dropping the temperature does not necessarily influence SNAPP distributions since the CC3 and W55D distributions remain unchanged. The results suggest that, out of the three variants examined, wild type Ltn is most susceptible to temperature denaturation. The drop in number of 18C6 adducts at lower temperatures is similar to the trend observed in Figure 5.3 and suggests cold denaturation is responsible for the shift in the SNAPP distribution, rather than some other type of structural shift (for example skewing the equilibrium in favor of the CC3 structure, which would lead to an increase in 18C6 attachment). It should be remembered that wild type Ltn is the only form of the protein which is capable of simultaneously interchanging between the Ltn-10 and Ltn-40 structures. It is likely that this ability necessitates that any particular folded state be in a rather shallow well of stability, making susceptibility to denaturation more likely.

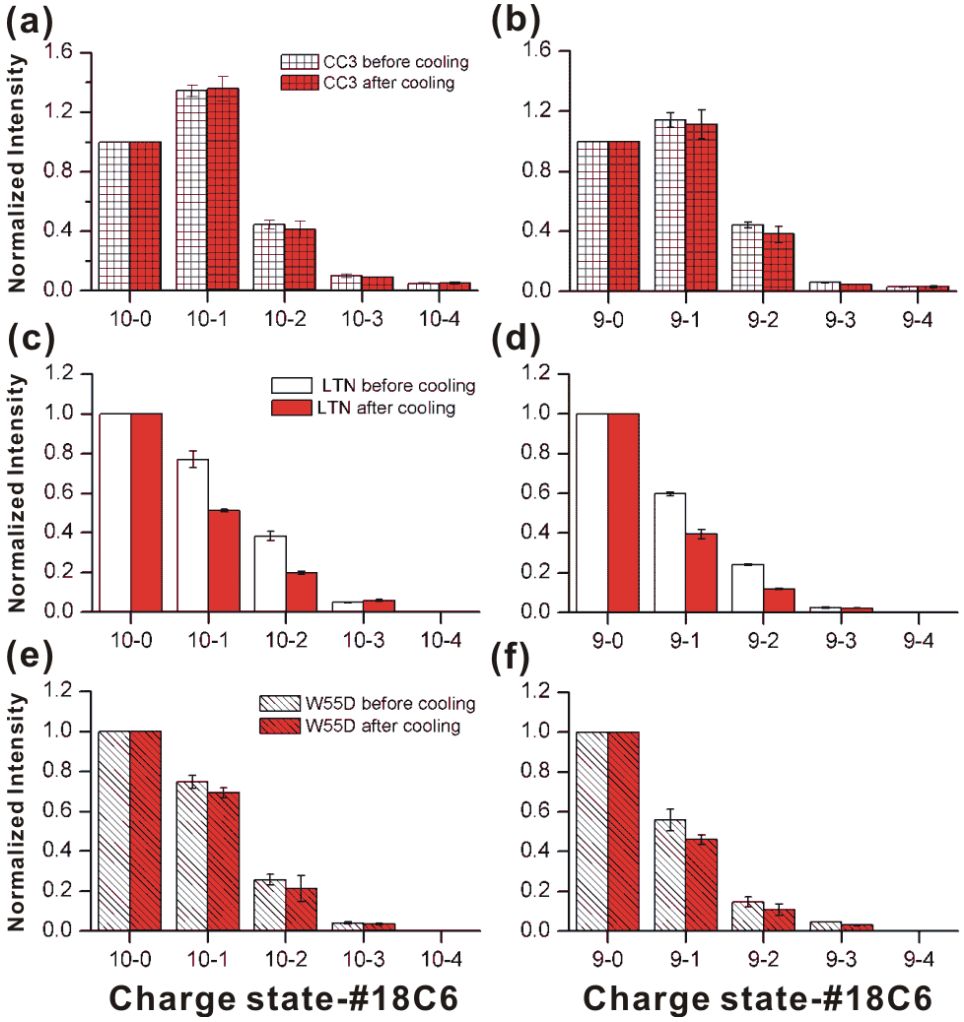


Figure 5.4 SNAPP distributions of Ltn mutants acquired in water at room temperature (plot a, c, e) and after the addition of an ice pack (plot b, d, f). Only the wild type Ltn distribution shifts to any significant extent.

The SNAPP experiments reported up to this point were designed to investigate the solution phase structure of Lymphotactin, even though ultimate detection took place in the gas phase. It is also possible that information about solution phase structures may be obtained by actually examining molecules in the gas phase.[2021] This would be possible in

at least two scenarios, 1) if the solution phase structure was retained in the gas phase, then direct analysis would apply, or 2) if differences in the solution phase structures led to differences in the gas phase structures, then it would be possible to monitor changes in structure. The second scenario would of course only reveal differences in structure rather than information about the structures themselves. It is beyond the scope of the present experiments to attempt to determine if the solution and gas phase structures for Lymphotactin are the same; however, we did examine the gas phase structures of the wild type Ltn, CC3, and W55D proteins to determine if they were distinguishable from each other. This was accomplished by using IBA-18C6, which is an 18C6 based molecule with a highly efficient radical precursor attached. Following photoactivation with 266nm light, a radical is generated by dissociation of the carbon-iodine bond. This radical is subsequently transferred to the protein. As we have demonstrated previously, radical migration and consequent dissociation in gaseous protein ions is largely determined by the overall protein structure.[22] Therefore, differences in protein structure would be expected to yield differences in dissociation. Interestingly with respect to SNAPP, this experiment also potentially offers information about preferred binding sites for 18C6.

The results from radical directed dissociation of all three proteins in the +7 charge state are shown in Figure 5.5. It is clear that all three dissociation spectra are distinct, although the wild type Ltn and W55D spectra share similarities. For CC3, there is an abundant cleavage between N67 and T68, as shown in Figure 5.5a. In addition, there is modest dissociation observed along the remainder of the c-terminal portion of the protein. Not surprisingly, there is no dissociation observed between the portions of the protein

which are linked by disulfide bonds. This does not mean that there is no dissociation occurring in the region, as such dissociation could not be observed unless two backbone bonds were broken. There is also a very modest amount of dissociation on the N-terminal side of the disulfide linked portion, suggesting that indeed some segments of the protein within the disulfide bonds are likely accessible to the radical. The wild type Ltn in Figure 5.5b yields a significantly different distribution. Dissociation at several residues in close proximity to Lys76 is abundant, suggesting that this may be a favorable site for attachment of IBA-18C6. In comparison with CC3, several peaks towards the C-terminal portion of the protein are missing. W55D yields results similar to those for wild type Ltn, but even less dissociation is observed. Importantly, even though there is an M to V substitution in the region where dissociation is observed for the W55D mutant (see Scheme 5.2), the mutation does not appear to affect dissociation, which is actually observed at that location in all three proteins.

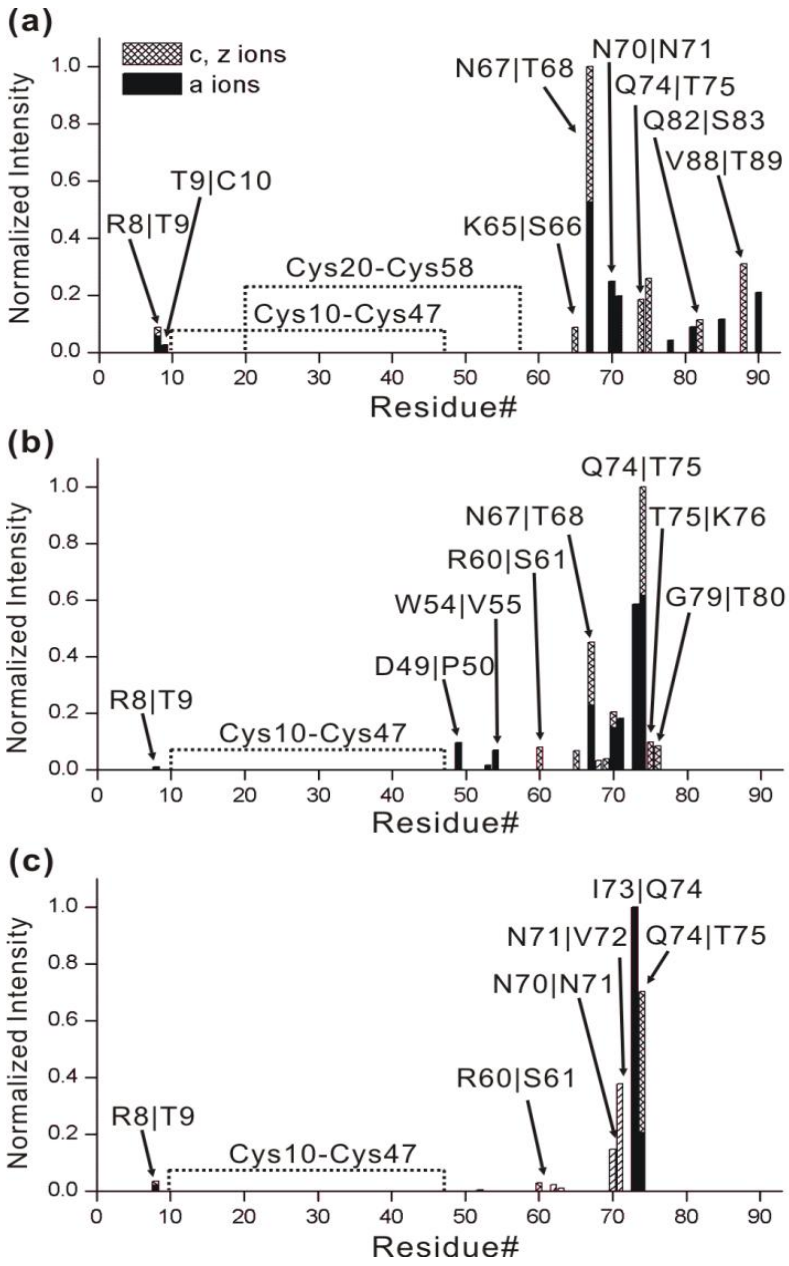


Figure 5.5 Dissociation points are shown as a function of sequence for experiments probing the gas phase structures of the +7 charge state for (a) CC3, (b) wild type Ltn, and (c) W55D (the numbering of residues for W55D has been shifted to be consistent with the other proteins, for actual numbering see Scheme 5.2). Differences in dissociation indicate differences in structure as described in the text.

Investigation of higher charge states, where tertiary structure is lost in the gas phase due to Coulombic repulsion, reveals that most of the dissociation is centered near Lys76. In the absence of tertiary structure, dissociation is expected to be most abundant in the vicinity of the starting location of the radical. This suggests that Lys76 is a preferred binding site for 18C6. Due to the cyclic nature of all three proteins which all contain one or more disulfide bonds, it cannot be determined whether there are other equally preferred binding sites. In any case, the differences between the dissociation spectra clearly indicate that the gas phase structures for the proteins are dissimilar. Since the sequences of these proteins are all very similar, the differences most likely result from either retention or memory of the solution phase structures.

For both the Ltn10 and Ltn40 structures, portions of the c-terminal tail of the protein are largely disordered as determined by NMR (see italicized residues in Scheme 5.2). These sections are therefore absent in the structures shown in Scheme 5.1, but they are still present and can contribute to the results which are obtained in SNAPP experiments. These regions contain several charged residues which are potential 18C6 targets and the results from RDD suggest that at least one lysine in this region is a favorable binding site. Fortunately, SNAPP is capable of extracting structural information even for proteins which are completely disordered.[23] In the case of Lymphotactin, some of the differences between the observed SNAPP results and those that would be predicted by the known structures may be attributable to structural shifts in these disordered regions.

5.4 Conclusions

The conformationally distinct Ltn10 and Ltn40 native states of Lymphotactin can clearly be distinguished and structurally characterized by SNAPP-MS. The results also suggest that the distribution of structures occupied by Lymphotactin is not significantly influenced by the addition of methanol or acid. In contrast, reduction and blocking of the disulfide bonds produces a denatured protein. Cold temperature experiments suggest that, out of three sequence variants examined, the wild type protein is the most susceptible to denaturation. With regards to the structural transition between the two native states of Lymphotactin, the data reveals that slightly increased 18C6 binding is observed for the transition structures relative to folded conformations. This observation is consistent with a small increase in intramolecular hydrogen bonding interactions that enhances the availability of lysine or arginine for 18C6 binding. These intrachain interactions may coordinate the transition between Ltn10 and Ltn40. In the gas phase, the sequence variants of Lymphotactin also adopt distinct structures, as revealed by radical directed dissociation. It is unclear whether the differences result from retention of solution phase structure in the gas phase or some other source. Overall, the results indicate that SNAPP is a useful method for probing structure in proteins that undergo slow global conformational transitions.

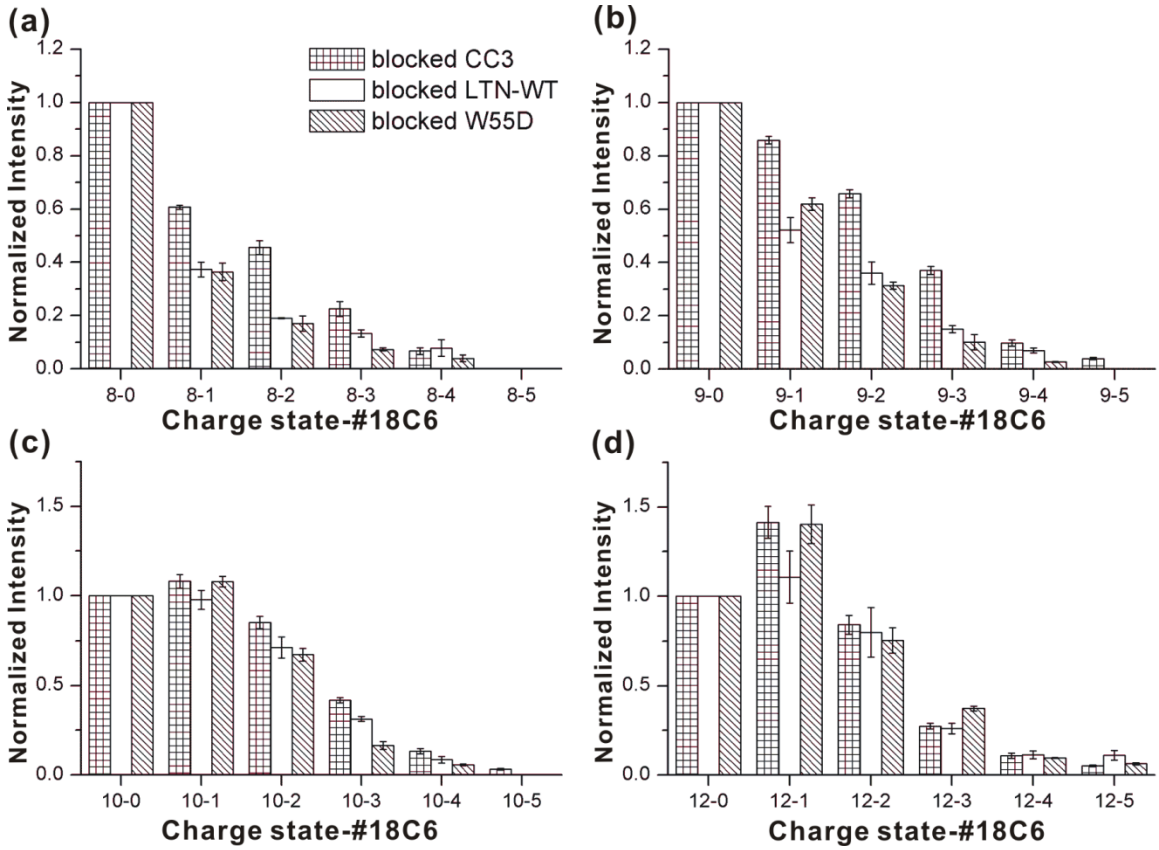


Figure 5.6 SNAPP distributions for three variants of blocked Lymphotactin sampled from water.

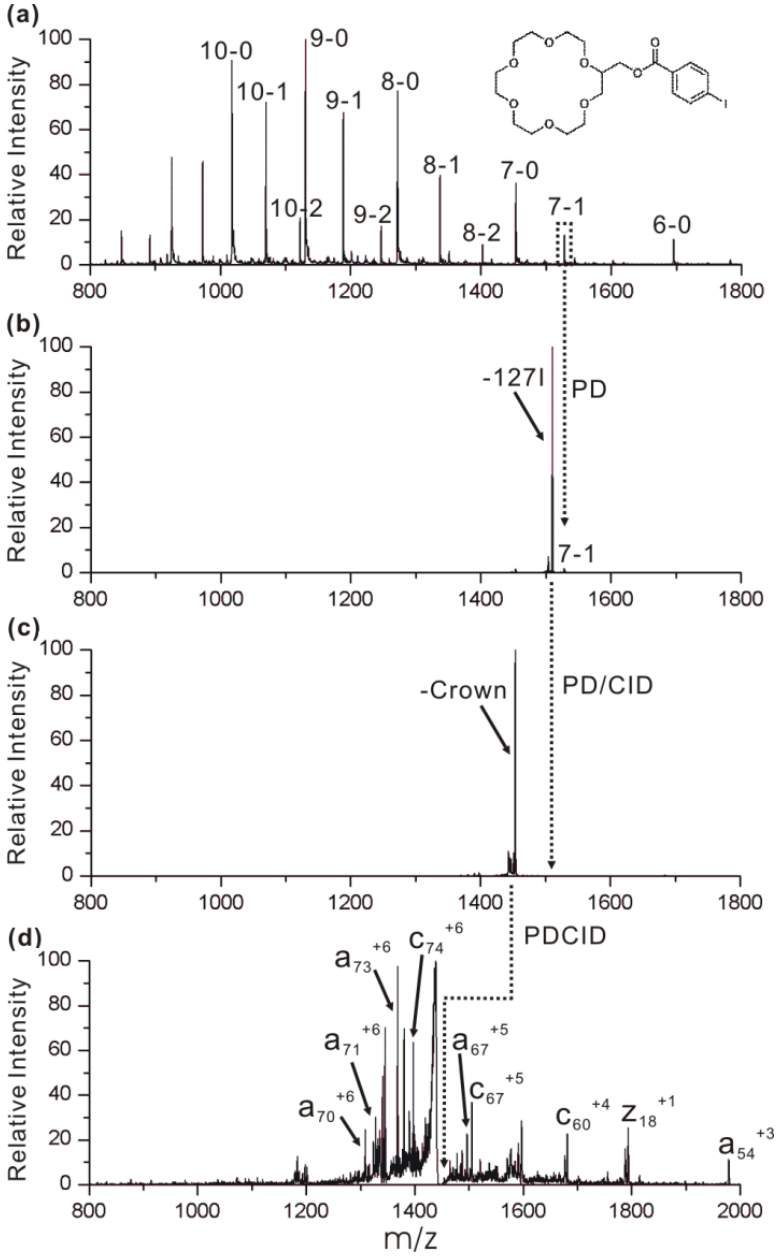


Figure 5.7 Radical directed dissociation procedures of high charge state of Ltn-wt protein in the gas phase. (a) Full MS distribution of Ltn-wt-IBA18C6 adduct (b) PD of $[\text{Ltn-wt+IBA18C6}]^{+7}$ (c) PD/CID of $[\text{Ltn-wt+IBA18C6-127I}\bullet]^{+7}$ (d) PD/CID of $[\text{Ltn-wt}\bullet]^{+7}$.

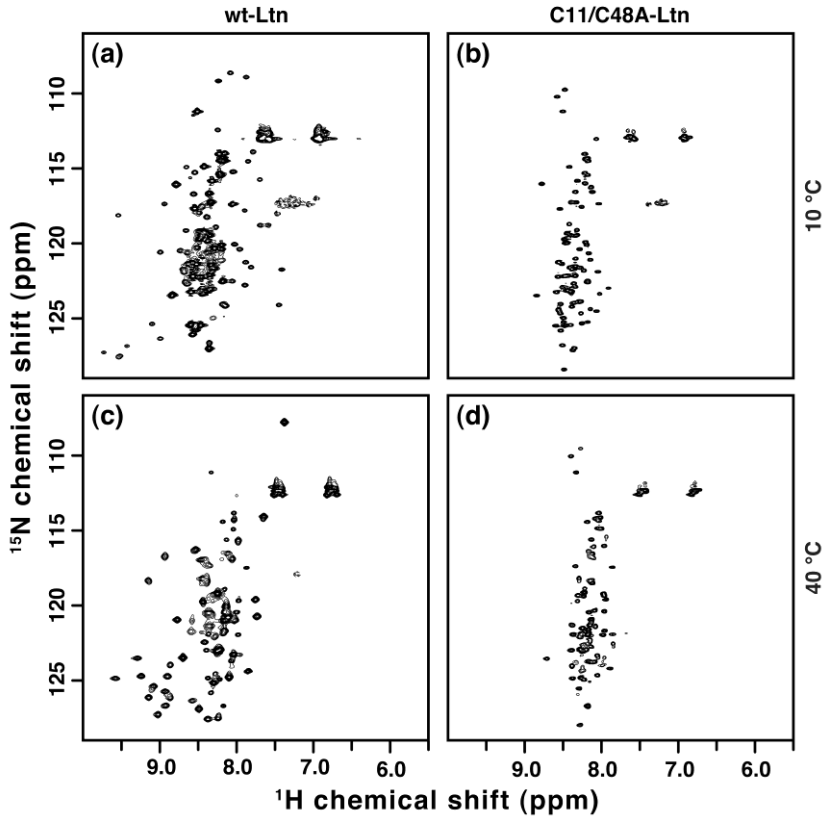


Figure 5.8 Alanine mutation of the Ltn disulfide bond unfolds Ltn. ^{1}H-^{15}N HSQC spectra of wild-type and C11A/C48A-Ltn acquired at 10 and 40 ˚C. All spectra were acquired in 20 mM sodium phosphate at pH 6.0.

---

[1] Tuinstra, R.L.; Peterson, F.C.; Kutlesa, S.; Elgin, E.S.; Kron, M.A.; Volkman, B.F. *PNAS*. **2008**, 105, 557-5062.

[2] Volkman, B.F.; Liu, T.Y.; Peterson, F.C. M*ethods Enzymol*. **2009**, 461, 51-69.

[3] Tuinstra, R.L.; Peterson, F.C.; Elgin, E.S.; Pelzek, A.J.; Volkman, B.F. *Biochemistry*. **2007**, 46, 2564-2573.

[4] Ly, T.; Julian, R. R. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1209-1215.

[5] Rodgers, M. T.; Chen, Y. 58th ASMS Conference on Mass Spectrometry and Allied Topics, Salt Lake City, Utah.

[6] Liu, Z.; Cheng, S.; Gallie, D. R.; Julian, R. R. *Anal. Chem.* **2008**, *80*, 3846-3852.

[7] Ly, T; Pujanauski, B. G.; Sarpong, R.; Julian, R. R. *Anal. Chem.* **2008**, *80*, 5059-5064

[8] Grandori, R. *J. Mass. Spectrom*. **2003**, *38*, 11-15

[9] Kaltashov, I. A.; Eyles, S. J. *Mass Spectrom. Rev*. **2002**, *21*, 37-71.

[10] Yan, X.; Watson, J.; Ho, P. S.; Deinzer, M. L. *Mol. Cell. Proteom*. **2003**, *3*, 10-23

[11] Jones, L. M.; Zhang, H.; Vidavsky, I.; Gross, M. L. *Anal. Chem.* **2010**, *82*, 1171-1174.

[12] Konermann, L.; Stocks, B. B.; Czarny, T. *Anal. Chem.* **2010**, *82*, 6667-6674.

[13] Wales, T. E.; Engen, J. R. *Mass Spectrom. Rev*. **2006**, *25*, 158-170

[14] Mendoza, V. L.; Vachet, R. W. *Mass Spectrom. Rev*. **2009**, *28*, 785-815.

[15] Peterson, F. C.; Elgin, E. S.; Nelson, T. J.; Zhang, F.; Hoeger, T. J.; Linhardt, R. J.; Volkman, B. F. *J Biol. Chem. 2004*, 279, 12598–12604.

[16] Grandori, R. *J. Mass. Spectrom*. **2003**, *38*, 11-15.

[17] Dobo, A.; Kaltashov, I. A. *Anal. Chem*. **2001**. *73*, 4763-4773

[18] Kuloglu, E.S.; McCaslin, D.R.; Markley, J.L.; Volkman, B.F. *J. Biol. Chem*. **2002**, 277, 17863-17870.

[19] Camilloni, C; Sutto, L. *J. Chem. Phys.* **2009**, 131, 245105-254105-6.

[20] Clemmer, D.E.; Jarrold, M. F. *J. Mass Spectrom.* **1997**, *32*, 577-592.

[21] Danell, A. S.; Parks, J. H. *Int. J. Mass Spectrom*. **2003**, *229*, 35-45.

[22] Ly, T.; Julian, R. R. *J. Am. Chem. Soc.* **2010**, *132*, 8602-8609.

[23] Ly, T: Julian, R. R. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1663-1672.

CHAPTER 6

CONCLUDING REMARKS

This dissertation introduces four of the projects I have done during my Ph.D study. The central theme that links all the projects together is elucidation of protein characteristics in three different categories using novel MS-based methods.

Development of RDD based methods for characterization of proteins constitutes a large portion in my Ph.D work. One of the most promising aspects of RDD is that RDD yields predictable peptide fragmentation. We have proposed that establishment of proteome MS-RDD data bases will eventually increase the confidence level of peptide and protein identification, and boost data base searching rate. To realize this goal, it is necessary to develop high throughput RDD for peptide fragmentation in bottom-up experiments. The problem to be solved is how to create hydrogen-deficient peptide radicals in a high throughput way when the analyte is a proteolytic mixture. It will be ideal if a commercialized mass spectrometer can be finally implemented with the feature to generate hydrogen-deficient peptide radicals without radical initiators. This requires enormous efforts from physicists, chemists as well as electronic engineers in order to develop such an amazing mass spectrometer, which probably won't become true in the short term. Alternatively, if radical initiators can be introduced in a high through-put way to every peptide in a proteolytic mixture, peptide radicals can be efficiently generated by subjecting the modified mixture to online LC-UVPD. By doing data dependent scans,

peptide radicals can be further collision activated to generate RDD fragments! In fact, one of our radical initiators, 5-Iodo-benzoyl-N-Hydroxysuccinimide ester (Fig. 1.1b), is capable of fulfilling this task. This iodinated NHS ester can react with the C-terminus residues (either lysine or arginine) of tryptic peptides in a high yield to introduce the C-I bond, which makes the labeled peptides ready for RDD investigation.[1] It has been reported that this iodo-NHS ester guarantees a high radical yield (~98%), which improves RDD fragment efficiency, therefore ensures the quality of high throughput RDD data. The establishment of high throughput RDD would make a big progress in both MS method development and proteomics study, and will greatly benefit researches that are in high demands of rapid identifications of peptides and proteins.

Other areas that RDD could have impacts on include but not limited to mapping iodination PTM sites in biological relevant proteins, localization of accessible tyrosines in protein-protein interaction interfaces etc. LC separation of protein mixtures prior to RDD top-down sequencing would always be essential in order to extend RDD to examine more complicated biological samples. Therefore, development of online LC/MS-RDD should be another addressed direction of RDD future development.

SNAPP-MS technique is an orthogonal method to RDD, which has had many applications for protein structural studies. SNAPP-MS could be employed to study more natively ordered proteins, compare structure differences between protein mutants and explore protein conformational changes upon protein-ligand, and protein-protein interactions.

In summary, this dissertation summarizes the work I have done in graduate school to use powerful mass spectrometry to solve some real problems in the field of proteomics study. Some of the described methods are still under development, which might require more creations and efforts for improvement. However, the diversity of MS-based applications strongly indicates that MS is a robust technique for exploration of the myriad physical and chemical properties of proteins to understand the global picture of the complicated biological systems.

---

[1] Ly, T.; Zhang, X.; Sun, Q.; Moore, B.; Tao, Y.; Julian, R. R. *Chem. Comm.* ***2011***, *47*, 2835-2837.