

FOURIER TRANSFORM ION CYCLOTRON RESONANCE MASS
SPECTROMETRY METHODOLOGIES FOR THE ANALYSIS OF BIOMOLECULES

by

MELISSA JOYCE STOUDEMAYER

(Under the Direction of I. Jonathan Amster)

ABSTRACT

Fourier transform ion cyclotron mass spectrometry (FT-ICR) is a very powerful technique that is well suited for the analysis of many types of biomolecules. This dissertation presents four experiments utilizing the strengths of FT-ICR. The first experiment presented evaluates the use of accurate mass measurement and nitrogen stoichiometry to characterize the proteome of *Acinetobacter baylyi* ADP1. Then, accurate mass measurements, paired with endogenous labeling, are used to investigate the metabolic pathway of dimethylsulphoniopropionate in *Ruegeria pomeroyi* DSS-3. The next analysis investigates the robustness and effectiveness of using chemical crosslinking and mass defect labeling as a means of detecting and evaluating reaction products in individual proteins and peptides. The last topic presented examines the various factors responsible for the generation of an uncommon ion, $[M-H]^+$ in a set of synthesized curcumin analogues.

INDEX WORDS: Fourier transform ion cyclotron mass spectrometry (FT-ICR), Mass spectrometry, Electrospray ionization (ESI), Matrix assisted laser desorption ionization (MALDI), Shotgun proteomics, Accurate mass

measurement, Nitrogen stoichiometry, Mass defect labeling, Chemical crosslinking, Curcumin analogues

FOURIER TRANSFORM ION CYCLOTRON RESONANCE MASS
SPECTROMETRY METHODOLOGIES FOR THE ANALYSIS OF BIOMOLECULES

by

MELISSA JOYCE STOUDEMAYER

B.S., University of South Carolina Aiken, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2014

© 2014

Melissa Joyce Stoudemayer

All Rights Reserved

FOURIER TRANSFORM ION CYCLOTRON RESONANCE MASS
SPECTROMETRY METHODOLOGIES FOR THE ANALYSIS OF BIOMOLECULES

by

MELISSA JOYCE STOUDEMAYER

Major Professor: I. Jonathan Amster

Committee: Lance Wells
Geoffrey Smith

Electronic Version Accepted:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
August 2014

DEDICATION

This dissertation is dedicated to my parents, Richard and Patricia Warren, and my husband, Christopher, for their unconditional love, support, and encouragement. To my beautiful daughter Avianna, who has filled me with joy and happiness, I never knew how perfect my life could be until the moment you were born.

ACKNOWLEDGEMENTS

I would like to thank Dr. I. Jonathan Amster for all of his support and direction during my research at UGA. His patience has allowed me to grow not only scientifically, but also personally. He opened my eyes to the amazing world of mass spectrometry, and for that I cannot thank him enough. I would like to thank my committee members, Dr. Geoff Smith and Dr. Lance Wells, for their assistance and advice as I progressed through my graduate degree.

I have been very lucky to be surrounded by so many supportive and helpful people during my time at UGA. Some of them, Dr. Dennis Phillips and Dr. Greg Wylie, taught me ways to troubleshoot the frustrating mechanical limitations and breakdowns that are bound to occur in any research project. Finally I would like to thank those friends and group members that made this journey rewarding on so many levels: Lindsay Renbaum-Wolff, Jeremy Wolff, Rachel Parker, Patience Wright, Jiana Duan, Joshua Driver, and Franklin Leach.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER	
1 INTRODUCTION AND LITERATURE REVIEW	1
2 EXPERIMENTAL	34
3 SHOTGUN PROTEOMIC ANALYSIS AND ASSIGNMENT OF <i>ACINETOBACTER BAYLYI</i> ADP1 VIA ACCURATE MASS MEASUREMENT	52
4 METABOLIC PATHWAY OF DIMETHYLSULPHONIOPROPIONATE IN <i>RUEGERIA POMEROYI</i> DSS-3	68
5 CHEMICAL CROSSLINKING USING A NOVEL MASS DEFECT LABELED CROSSLINKER	80
6 CURCUMIN ANALOGUE ANALYSIS AS AN INSIGHT INTO MALDI MECHANISMS	114
7 CONCLUSIONS	136
APPENDICES	140
A Supplemental Data for Chapter 5	141
B Supplemental Data for Chapter 6	147

LIST OF TABLES

Table 1.1	Mass defect contributions of the elements typically found in peptides	25
Table 2.1	Gradients used of the course of the chemical crosslinking experiments	38
Table 2.2	Sample numbers, monoisotopic masses, and structures of the curcumin analogues analyzed	50
Table 3.1	Summary of peptide and protein assignments from the global digest of the <i>Acinetobacter baylyi</i> ADP1 proteome	59
Table 5.1	Mass defects of elements that comprise a peptide of averagine at m/z 1000	81
Table 5.2	Observed crosslinker reaction products of the neurotensin and bradykinin mixture	92
Table 5.3	Distances between lysines of interest on RNase S protein and S-peptide	97
Table 6.1	The tested curcumin analogues, their structures and monoisotopic masses	121
Table 6.2	Theoretical and measured monoisotopic masses for the $[M+H]^+$ and $[M-H]^+$ peaks from the MALDI FT-ICR analyses	124
Table 6.3	Theoretical and measured monoisotopic mass for the $[M-H]^+$ peaks observed during the LD FT-ICR analyses	124
Table 6.4	Summary of the various factors tested in the experiments	128

LIST OF FIGURES

Figure 1.1	Comparison of generalized bottom-up and top-down proteomic workflows	4
Figure 1.2	General schematic of the electrospray ionization process	7
Figure 1.3	General mechanism for MALDI desorption and ionization	9
Figure 1.4	Schematic of the ion motion through a FAIMS device, as well as the applied waveform for separation	12
Figure 1.5	Schematic of a cubic analyzer cell	14
Figure 1.6	The cyclotron motion of an ion	15
Figure 1.7	Ion trajectory illustrating the combination of cyclotron and magnetron motion on an ion	17
Figure 1.8	Time-domain and frequency-domain excitation waveforms	19
Figure 1.9	Production of the image current of the ions in the analyzer cell during the detection stage of the ICR experiment	20
Figure 1.10	Schematic of the Orbitrap analyzer cell	20
Figure 1.11	Some of the possible crosslinker reaction products expected during a traditional experiment	22
Figure 1.12	Schematic comparing conventional peptide and protein workflows to those used in crosslinking experiments	24

Figure 1.13	Histogram of the relative mass defect versus frequency observed in tryptic peptides	26
Figure 1.14	Histogram of peptides within a 2 amu range	27
Figure 1.15	Simplified algorithm that can be used to determine whether an observed peak in a peptide mixture has been modified with a mass defect label	27
Figure 2.1	Sample chromatogram showing the elution gradient used	35
Figure 2.2	Liquid chromatography gradient used during a 180 minute experiment	38
Figure 2.3	Synthesis reaction for the mass defect labeled crosslinker DiBBSIAS	47
Figure 3.1	Experimental workflow used for the shotgun proteome analysis	56
Figure 3.2	HPLC separation of the proteome digest of <i>Acinetobacter baylyi</i> ADP1	57
Figure 3.3	Results from the global digest of <i>A. baylyi</i> ADP1 labeled by endogenous ¹⁵ N labeling	59
Figure 4.1	HPLC separations of the extracted CoA mixtures	70
Figure 4.2	MALDI FT-ICR mass spectrum of MMPA-CoA extracted from the whole cells after incubation of the cells with MMPA	71
Figure 4.3	MALDI FT-ICR mass spectrum of MMPA-CoA extracted from the whole cells incubated with [1,2,3- ¹³ C] DMSP	72
Figure 4.4	MALDI FT-ICR mass spectrum of MMPA-CoA extracted from the cell-free extracts with MMPA-CoA and hexafluorophosphate	73

Figure 4.5	^1H NMR spectrum of DmdD catalyzed reaction products	74
Figure 4.6	^{13}C NMR spectrum of $[1-^{13}\text{C}]$ MTA-CoA	75
Figure 4.7	Pathway of DMSP demethylation in <i>R. pomeroyi</i> DSS-3	76
Figure 5.1	Relationship between nominal mass and exact mass of peptides	82
Figure 5.2	Histogram of peptide within a 2 amu range	83
Figure 5.3	Isotopic envelope of mass defect labeled peptides	84
Figure 5.4	Chemical structure of the novel chemical crosslinker used	85
Figure 5.5	Data workflow used to determine the crosslinker reaction products	86
Figure 5.6	Chemical crosslinking reaction scheme used for the test peptide samples	87
Figure 5.7	MALDI-TOF analysis of the test peptide and crosslinker digested solution	88
Figure 5.8	FT-ICR analysis of the test peptide and crosslinker digested solution	89
Figure 5.9	Isotopic envelope for crosslinker reaction product from the test peptide.	90
Figure 5.10	FT-ICR analysis of the crosslinked test peptide	90
Figure 5.11	Direct infusion mass spectrum of the neurotensin and bradykinin mixture reacted with a 20 times molar excess of crosslinker	91
Figure 5.12	FAIMS analysis of the neurotensin and bradykinin crosslinked solution	93
Figure 5.13	Total and extracted ion chromatograms from the analysis of the neurotensin and bradykinin crosslinked reaction solution	94

Figure 5.14	Isotopic distribution for m/z 722.20 ²⁺	95
Figure 5.15	MS ² fragmentation spectrum of the m/z 722.20 ²⁺ peak (CAD 20V)	95
Figure 5.16	MS ² fragmentation spectrum of m/z 1050 ²⁺	96
Figure 5.17	Illustration of RNase S complex	97
Figure 5.18	MALDI FT-ICR spectrum of the crosslinked RNase S complex tryptic digest	98
Figure 5.19	Plot of Δ mass defect values versus m/z for all the peaks found in a MALDI mass spectrum of crosslinked RNase S	99
Figure 5.20	Fragmentation spectra sequencing S-peptide	100
Figure 5.21	Preliminary compensation voltage scan obtained from scanning 1-30 positive volts	101
Figure 5.22	Extracted spectral information from the preliminary voltage scan	101
Figure 5.23	Extracted ion chromatogram (EIC) of selected peaks from the spectrum in Figure 5.21	102
Figure 5.24	Fragmentation spectrum of m/z 1131.243 ⁺	104
Figure 5.25	The total ion chromatogram for the bovine hemoglobin crosslinked digestion solution collected on the nano-LC-Elite Orbitrap instrument	106
Figure 5.26	The CAD fragmentation spectrum of the m/z 984.64 ⁴⁺ precursor peak at 25V of normalized collision energy	107
Figure 5.27	The total ion chromatogram for the bovine hemoglobin crosslinked digestion solution on the 9.4T LC-FT-ICR instrument	107

Figure 6.1	General structure of the curcumin analogues analyzed (-3,5-bis(benzylidene)-4-piperidones)	115
Figure 6.2	Possible chemical structures for the two most intense peaks observed in positive mode MALDI analysis of the curcumin analogues	116
Figure 6.3	General schematic of the major processes proposed in the lucky survivor model	118
Figure 6.4	FT-ICR MALDI-MS mass spectrum of curcumin analogue sample NC 2311	123
Figure 6.5	FT-ICR LD-MS mass spectrum of curcumin analogue sample NC 2311	123
Figure 6.6	MALDI-TOF analysis of NC 2144 under various conditions	125
Figure 6.7	LD-TOF analysis of NC 2144 under various conditions	126
Figure 6.8	Comparison of observed species in the positive mode analysis of NC 2144 using LD and MALDI with the addition of a radical scavenger	127
Figure 6.9	Proposed formation of $[M-H]^+$ from the $[M+H]^+$ by loss of a hydrogen molecule	130
Figure 6.10	Proposed formation of $[M-H]^+$ from the unstable radical form $[M]^{\bullet+}$	131
Figure 6.11	Proposed formation of $[M-H]^+$ by the direct loss of a negative hydrogen ion from a neutral curcumin analogue molecule through a redox reaction	131
Figure A.1	Proton NMR of the mass defect labeled crosslinker	142

Figure A.2	Mass spectrum of m/z 1543.460 ⁺ from RNase S reacted with 100 times molar excess DiBBSIAS	143
Figure A.3	Mass spectrum of m/z 1561.462 ⁺ from RNase S reacted with 100 times molar excess DiBBSIAS	144
Figure A.4	Mass spectrum of m/z 781.231 ²⁺ from RNase S reacted with 100 times molar excess DiBBSIAS	145
Figure A.5	CAD fragmentation spectra sequencing S-peptide	146
Figure B.1	Mass spectra of curcumin analogue sample NC 2311	148
Figure B.2	Mass spectra of curcumin analogue sample NC 2453	149
Figure B.3	Mass spectra of curcumin analogue sample NC 2144	150
Figure B.4	Mass spectra of curcumin analogue sample NC 2138	151
Figure B.5	Mass spectra of curcumin analogue sample NC 2094	152
Figure B.6	Mass spectra of curcumin analogue sample NC 2138	153
Figure B.7	Mass spectra of curcumin analogue sample NC 2311	154
Figure B.8	Mass spectra of curcumin analogue sample NC 2453	155
Figure B.9	Mass spectra of curcumin analogue sample NC 2144	156
Figure B.10	Mass spectra of curcumin analogue sample NC 2094	157
Figure B.11	MALDI and LD analysis of NC 2311	158
Figure B.12	MALDI and LD analysis of NC 2311	159
Figure B.13	MALDI and LD analysis of NC 2094	160
Figure B.14	MALDI and LD analysis of NC 2094	161
Figure B.15	MALDI and LD analysis of NC 2144	162
Figure B.16	MALDI and LD analysis of NC 2144	163

Figure B.17	MALDI and LD analysis of NC 2138	164
Figure B.18	MALDI and LD analysis of NC 2138	165

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

PROTEOMICS

Proteomics is a very diverse and varied field of study, which focuses on the proteins observed under specific cellular conditions in cells, tissues, or organisms.¹ The term was coined in 1994 to be the protein complement to the genome.² The central dogma in biology describes how the genome of an organism is transcribed to produce messenger RNA (mRNA), which is then translated to the protein amino acid sequence.³ While the genome of an organism remains relatively static, with primarily mutations and transcription errors responsible for variations, the proteome of an organism is highly dynamic in response to internal and external stimuli.⁴ There is up and down regulation in the expression of genes, and these changes do not necessarily cause a linear change in the quantities of the proteins produced, and typically produces a dynamic range of more than 10^6 fold in protein abundance.⁵ Once the protein is built, it must still be modified to produce a functional form. This includes ensuring that the protein is properly folded and any post-translational modifications are performed. There are thousands of potential modifications to proteins including acetylation, methylation, and glycosylation.⁶ With all of the variances in observable proteins and their quantities, proteomics has developed many techniques to measure and quantify these systems.

Classical proteomics

The traditional approach to proteomic analysis combined 2-dimensional gel electrophoresis (2DE), which allows for protein visualization, quantification, and separation, with mass spectrometry for protein assignment or identification.⁷ Multidimensional gel electrophoresis separates proteins orthogonally by differing characteristics, such as isoelectric point (pI) and molecular weight. The proteins are then stained so that they can be observed, excised from the gel, and extracted from the gel matrix for analysis. The extracted samples then typically undergo enzymatic digestion to produce a mixture of peptides, which are then measured on the mass spectrometer. The resulting mass lists are then used to search a database that has been constructed from the theoretically observable peptides that result from *in-silico* digestion of the amino acid sequences encoded by the genetic sequence of multiple organisms.

There are several potential pitfalls with a classical proteomics style experiment.^{5,8} These experiments require the gel separation to be successful, but there are many classes of proteins that have extreme pI ranges, poor solubility, or high and low molecular weights, which make gel separations poor. The very large dynamic range of proteins observed from a cellular extract also makes 2DE non-ideal, without pre-gel enrichment steps. The very labor intensive nature of these experiments, as well as the inability to automate the steps, make them relatively low throughput. The incorporation of liquid chromatography would increase the throughput and ability to automate these experiments dramatically, as well as coupling the experiment more easily to mass spectrometry analysis.

The field of proteomics has expanded greatly over the last couple decades as mass spectrometry has become the primary tool for analysis. The introduction of soft ionization techniques, such as electrospray ionization (ESI)⁹ and matrix assisted laser desorption ionization (MALDI)^{10,11} has allowed for the mass analysis of large thermally labile biomolecules with minimal fragmentation. The integration of mass spectrometry into proteomic analyses has allowed for higher throughput experiments identifying and quantifying proteins, characterizing post-translational modifications (PTMs), and investigating protein-protein interactions. There are two general approaches to proteomics using MS-based techniques: bottom-up and top-down, as shown in Figure 1.1.

Bottom-up vs Top-down

Bottom-up proteomics describes the approach where one or more proteins are proteolytically digested into peptides prior to analysis.^{12,13} These analyses generally start with the protein, or mixture of proteins, digested with a proteolytic enzyme to produce a mixture of peptides. These can then be analyzed directly by the mass spectrometer, or can undergo a chromatographic separation prior to mass analysis. The protein identification is then made by taking the observed mass lists and comparing them to a database populated with *in-silico* digested proteins. The peptides can also be fragmented under tandem mass spectrometric conditions and the resulting spectra are compared to sequence databases.

As bottom-up proteomics is well established experimentally and widely used, there have been many optimizations and techniques developed to be used in tandem with these methods. However, there are still limitations, such as difficulty in obtaining the

complete protein sequence, which requires many experiments to be performed in hopes of observing the majority of the proteins in a sample. This limitation decreases the throughput of these experiments, which is further slowed by, each replicate requiring the use of liquid chromatography paired with tandem mass spectrometry (LC-MS/MS) to obtain full validation of peptide sequence necessary for unambiguous protein identification.

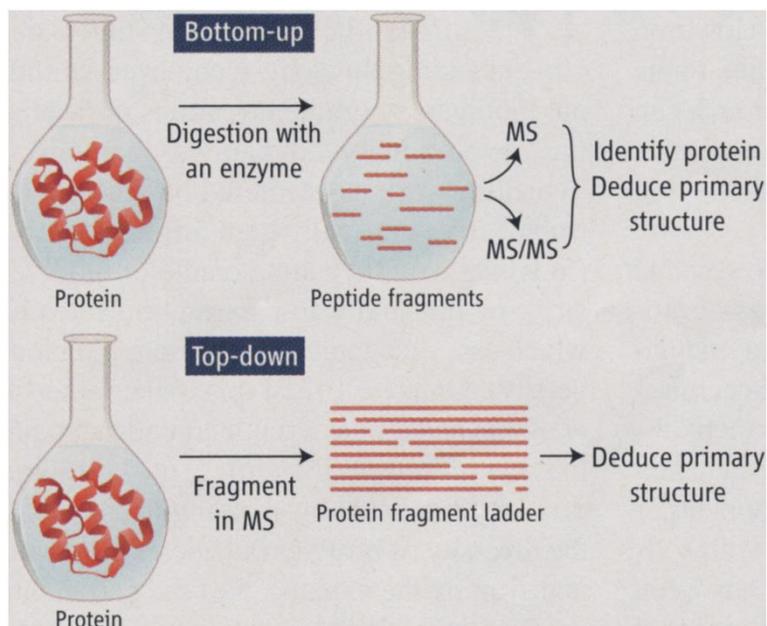


Figure 1.1 Comparison of generalized bottom-up and top-down proteomic workflows.¹⁴

This is in stark contrast to the top-down approach, where ionization of the intact protein, or protein complex, is followed by tandem mass spectrometric analysis.^{13,15} Top-down experiments are typically performed using electrospray ionization to generate multiply charged ions that are introduced into a Fourier transform ion cyclotron mass spectrometer. Fragmentation can be performed via collisionally activated dissociation (CAD), but the most commonly used methods are electron capture dissociation (ECD) and electron transfer dissociation (ETD).¹⁶

Top-down analysis has two major advantages: complete protein sequence is available from the beginning and the ability to locate and characterize post-translational modifications.¹⁷ There are also several limitations to this technique as well. The spectra that are produced from this type of analysis are highly complex, with multiple charged species for each component. This makes it quite difficult to interpret. Another disadvantage is that these analyses must be performed on very high resolution and high mass accuracy instruments. There is also increased difficulty in successful analysis as the protein gets larger in molecular weight and as the number of proteins in the mixture increases.¹⁸

Shotgun proteomics

Shotgun proteomics is a gel-less method that uses liquid chromatography to separate the complex mixture of peptides generated by the enzymatic digestion of all the proteins extracted, typically from a whole cell lysate.^{19,20} The fractions from the separation are then analyzed by mass spectrometry. The peptides observed are then correlated back to identify which proteins were present in the original mixture. This can be done using tandem mass spectrometry, to produce fragmentation data which can then be searched in a database, or from the accurate mass measurements.

Shotgun proteomics has several advantages over classical proteomic experiments. The first benefit being that these analyses can be performed in days instead of weeks. The second is the amount of sample required for analysis. Traditional proteomics experiments require much more protein, as the excision of proteins in the gel requires

enough of each protein to be present after separation for the stain to be visible, whereas shotgun analyses are dependent on only the limit of detection on the mass spectrometer.

MASS SPECTROMETRY IN PROTEOMICS

Mass spectrometry (MS) is an analytical technique that measures the mass to charge ratio (m/z) of gas phase ions with very high sensitivity. The increased use of mass spectrometry as a technique in the analysis of biomolecules over the last few decades is due almost exclusively to the advent of soft ionization techniques, primarily electrospray ionization (ESI)⁹ and matrix assisted laser desorption ionization (MALDI).^{10,11} These methods allow large, nonvolatile, thermally labile analytes, such as peptides and proteins, to be ionized without inducing significant fragmentation.

Electrospray ionization (ESI)

Electrospray ionization (ESI) was first coupled to mass spectrometry by John Fenn in 1989 for the analysis of large biomolecules.⁹ Electrospray has several advantages in addition to being a soft ionization method. The ions produced can contain multiple charges, which can allow for tandem mass spectrometry fragmentation to be performed and elucidation of information regarding the chemical structure. ESI is performed under atmospheric pressure conditions, yielding high ionization efficiency. The reaction solution is continually introduced allowing for easy interfacing of liquid separation techniques, such as high performance liquid chromatography (HPLC) or capillary electrophoresis (CE).

The exact mechanism for ion production via electrospray is not fully understood, but research is continuously being published helping to clarify the finer details. Electrospray ionization produces a fine spray of charged droplets by passing a solution through a strong electric field via a capillary tube, that are introduced into the mass spectrometer.²¹ The potential difference between the capillary and the counter electrode to charge the ions is 2-4kV. In the presence of this field, the solution becomes a fine mist of charged droplets. The solvent is evaporated from the droplets, usually with heat or dry gas, causing a buildup of charges on the droplet surface. This increase in charge density continues until coulombic repulsion becomes too great, and the ions are ejected from the droplet or the droplet explodes, releasing the ions.²²

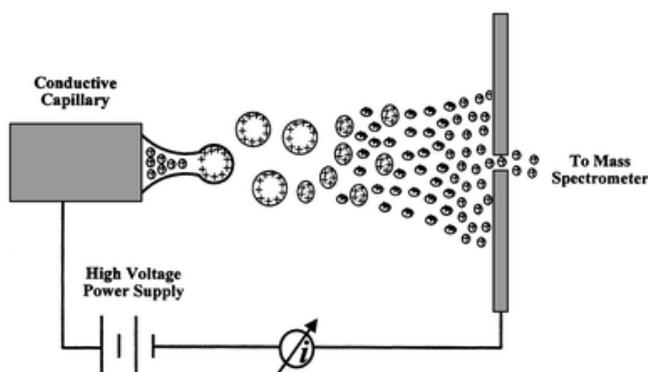


Figure 1.2 General schematic of the electrospray ionization process.²³

Electrospray is well suited for the analysis of larger biomolecules such as peptides and intact proteins, as the ions generated usually carry multiple charges.²⁴ As the mass spectrometer measures mass to charge, and not molecular weight, multiple charges on a molecule allow for high mass molecules to be observed at much lower mass to charge ratios. Multiple charging also allows for tandem mass spectrometry, where fragmentation is performed. Higher charge states on precursor ions yield more fragments and higher fragmentation efficiencies.²⁵ With the ability of producing multiple charges

on each component in a mixture, it can become difficult to analyze the spectra of mixtures, as each component will yield several spectral peaks corresponding to the analyte with varying numbers of protons.

One disadvantage of electrospray ionization is observed during the analysis of complex mixtures, wherein the most polar analyte, or analytes if the concentration is high, undergoes ion suppression. Ion suppression can be mitigated by dilution and using separation techniques to reduce the number of components in the spray at any given time. The presence of salts, detergents, or other small molecules typically left over from sample preparation severely reduce sensitivity, as these molecules tend to ionize far more efficiently, and outcompete for the finite amount of charge on the droplets. The sensitivity can be improved by using lower flow rates, as is seen in nanospray. Nanospray is able to pre-concentrate the sample, as well as desalt the sample prior to introduction into the mass spectrometer, with the additional bonus of longer analysis times. The longer analysis time allows data to be collected for a much longer period of time on a significantly smaller analyte sample size.

Matrix assisted laser desorption ionization (MALDI)

Matrix assisted laser desorption ionization (MALDI) was introduced for analysis of biomolecules in 1988 by Karas, Hillenkamp, and coworkers¹¹ by ionizing amino acids with a pulsed laser. In the same year Tanaka demonstrated that a protein can be ionized in the presence of a photon-absorbing matrix, paving the way to the analysis of large biomolecules with this method.¹⁰ MALDI desorbs and ionizes the analyte mixed with a matrix with a pulsed laser. The matrix strongly absorbs energy from the laser and rapidly

desorb into the gas phase, carrying the co-crystallized analyte molecules. For peptide analysis, a 337 nm N₂ laser is typically used along with either 2,5-dihydroxybenzoic acid (DHB) or α -cyano-4-hydroxycinnamic acid (CHCA) as the matrix. Since the matrix absorbs most of the laser energy, little is transferred into the analyte molecules, allowing for ionization without fragmentation, yielding primarily singly charged ions. The exact mechanism for ionization is not understood, but experiments, such as those discussed in Chapter 5, continue to tease out the chemical and physical interactions of the analytes and matrices.²⁶

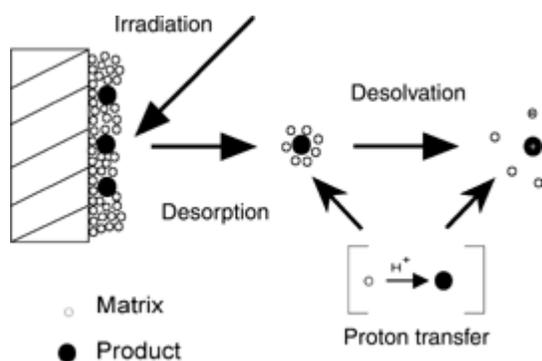


Figure 1.3 General mechanism for MALDI desorption and ionization.²⁷

There are two widely accepted mechanisms that have been proposed. In the first model, the matrix molecules are photoionized forming matrix ions, which then transfer the charge from the matrix ions to the neutral analyte molecules in the gas phase producing the analyte ions.^{28,29} The second model proposes that the analytes are incorporated into the matrix as charged molecules with its corresponding counter ions (based on the solvent used). The matrix crystals are desorbed by the laser beam and quantitative charge neutralization between analyte and counter ions produce the singly charged analyte ions (lucky survivors).³⁰

MALDI has several advantages over ESI in proteomics experiments. MALDI is far less sensitive to the presence of salts in the solution, allowing for far less sample purification and preparation prior to analysis. The spectra produced from MALDI experiments primarily only contain singly charged ions, which renders them simpler to interpret, although the presence of the matrix in the analyte mixture can significantly increase the background signal in the lower mass range, making it difficult to analyze components below 600 Da. This also limits the analysis of high mass components with high resolution instruments.

Separation technologies

The first step in most proteomic analyses is a separation of the proteins and peptides. These can be performed by gel technologies, liquid chromatography, and increasingly gas phase separation techniques, such as high field asymmetric ion mobility spectrometry (FAIMS). The separation of sample components prior to analysis has multiple benefits to the sample analysis. Separations can be used to purify samples by removing salts and buffers that may act as ion suppressors during electrospray ionization. Complex samples that may lead to very difficult to interpret spectra can benefit by having their complexity reduced by using a separation techniques that spreads the sample out over a period of time, such as liquid chromatography. This allows for a smaller number of components to be interrogated at one time, leading to simpler spectra. It also allows for more useful tandem mass spectrometry to be performed, as it is less likely that there will be ions co-isolated during the fragmentation.

Separations can also provide additional information about the samples as there are many potential characteristics that proteins can be separated by, such as molecular weight and isoelectric point in the case of gel separations, and hydrophobicity in the case of liquid chromatography. One new advancement in separation technology has been the development of gas phase separation prior to the introduction of the ions into the mass spectrometer. The separation is based on the differences in the mobility of the ions in an electric field, and is called differential mobility spectroscopy (DMS). One specific type of DMS is high-field asymmetric waveform ion mobility spectrometry (FAIMS) which has been demonstrated to be applicable to highly complex mixtures.

The ions created at atmospheric pressure are transported by a carrier gas between the planar electrodes. The mobility of an ion differs in relation to the electric field it is experiencing. An asymmetric waveform is applied, where the ions experience a high electric field for a short period of time, and then a low electric field for a relatively longer period of time, as shown in Figure 4. The dispersion voltage (DV), sometimes also called the separation voltage (SV)³¹, is the maximum value of this waveform and is static through the experiment. The ions are selectively filtered by varying the compensation voltage (CV) that is applied to the plates. This voltage is used to offset the migration of the ion caused the applied electric fields, and therefore can be scanned to provide a survey of the molecules in solution. It can also be made static so as to act as an ion filter prior to the mass spectrometer inlet, allowing for increased detectability of lower abundance ions.³²

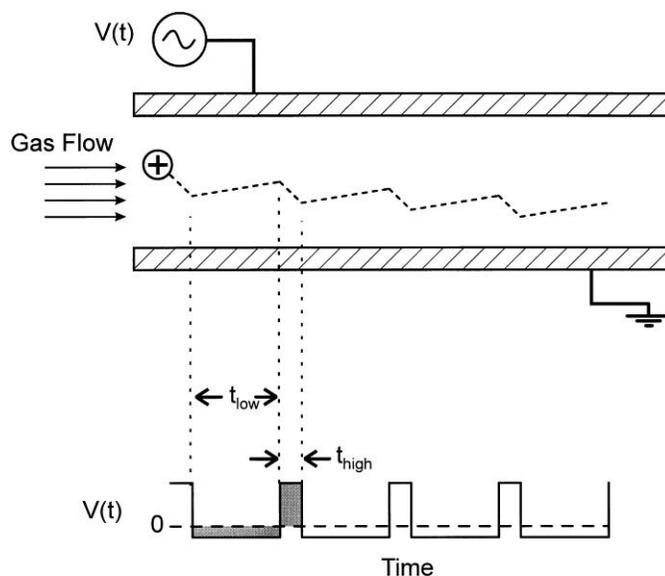


Figure 1.4 Schematic of the ion motion through a FAIMS device, as well as the applied waveform for separation.³²

The ability to scan the voltages or remain static provides many benefits to proteomic analyses. The initial survey scan allow for a quick survey of the peptides present prior to tandem experiments, allowing for fragmentation data to only be acquired for the specific peaks of interest. The ability to hold the voltage static allows for ion accumulation of low abundance ions, increasing not only their detectability, but also their fragmentation efficiency due to higher ion abundances. FAIMS suffers from the lack of ability to predict at what the compensation voltage the peptides will be observed, unlike in liquid chromatography, where the elution orders can be roughly estimated. There is also a decrease in ion sensitivity, although development of new designs and experimental methods are combating this problem.

FOURIER TRANSFORM MASS SPECTROMETRY

Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

Fourier transform ion cyclotron resonance mass spectroscopy (FT-ICR-MS) is a very power analytical technique, providing high mass accuracy, resolution, and sensitivity, which are provided by the fundamental mechanism by which these instruments work. The ion cyclotron motion in a plane perpendicular to a uniform magnetic field is measured to determine the frequency of the ion, which is dependent on the mass to charge (m/z) of the ion. Ion cyclotron motion was first described in the 1930s by Lawrence *et al*³³, and was first incorporated into a mass spectrometer in the 1950s by Sommer *et al*.³⁴ The first coupling of Fourier transform methods to ICR culminated in the first FTMS instrument, built by Comisarow and Marshall in 1974.³⁵ FTMS instrumentation did not become wide spread until after the introduction of external ion sources, robust pumping systems that could produce and maintain ultra-high vacuum, and computers with large enough data storage capacities and processing speeds.

FT-ICR instruments are composed of a magnet, an analyzer cell, an ultra-high vacuum system, and sophisticated data systems. The magnet is crucial as ion cyclotron motion is induced in a uniform magnetic field. Their typical magnetic strength ranges from 3 to 12 Tesla (T), but there is current research in the development of magnets up to 21 T.³⁶ Inside of the magnet, in a region of homogeneous magnetic field strength, the analyzer cell is located. The analyzer cell is responsible for trapping and detecting the ions, and can be shaped in a multitude of geometries, such as cubic or as an open ended cylinder. The cell is comprised of three pairs of electrodes, each placed opposite to each

other within the cell: two detection plates, two excitation plates, and two trapping plates, as shown below in Figure 5.

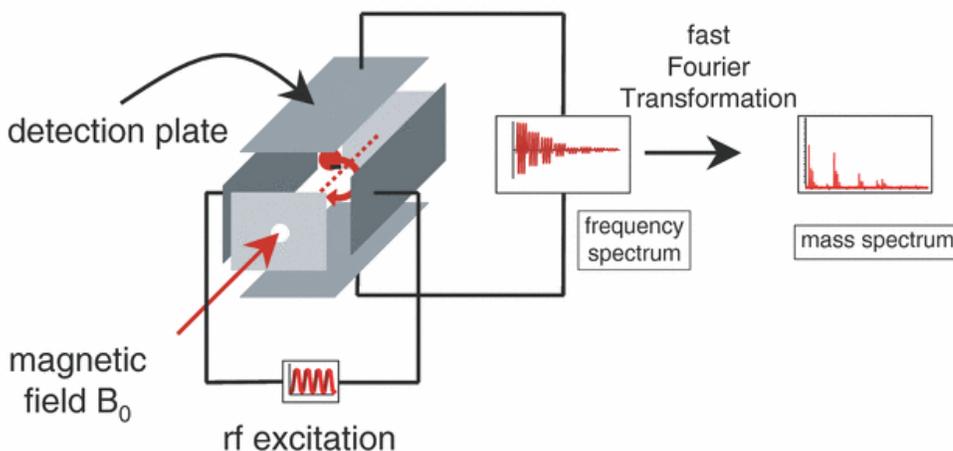


Figure 1.5 Schematic of a cubic analyzer cell. The detection plates (top and bottom), excitation plates (left and right), and trapping electrodes (front and back) are shown.³⁷

In order to achieve high mass resolution, ultra-high vacuum (10^9 - 10^{10} torr) must be obtained in the analyzer region of the instrument, as it decreases the number of possible collisions with neutral molecules and extends the lifetime of the ion within the analyzer cell. This is typically achieved by using turbomolecular pumps. Finally, to coordinate all of the electronics and data acquisition software, a highly sophisticated data system is required for the acquisition, processing, and analysis of the data.

Ion motion

During acquisition of FT-ICR data, the ions are subjected to three types of motion: cyclotron, trapping, and magnetron. When ions enter a magnetic field, they undergo cyclotron motion (Figure 1.6), which is the result of the Lorentz force and the centrifugal force both acting on an ion in opposite directions.

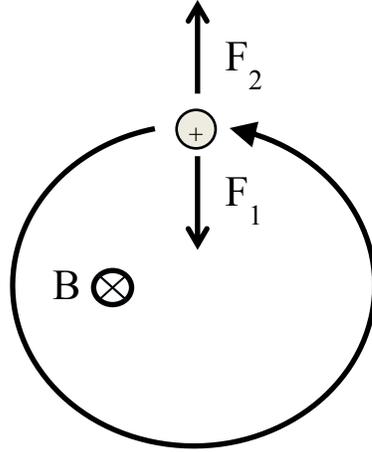


Figure 1.6 The cyclotron motion of an ion. This results from the balance between the Lorentz force ($F_1 = qvB$) and the centrifugal force ($F_2 = (mv^2/r)$), where q is the charge of the ion, v the velocity of the ion, r the radius of cyclotron motion, m the mass of the ion, and B the magnetic field strength. Modified from Amster 1996.³⁸

Since the magnitudes of both forces are equal, the ion cyclotron frequency, typically on the order of 5 kHz – 5 MHz,³⁸ can be calculated using the equation below, where f_c is the unperturbed cyclotron frequency, B the magnetic field strength, q the charge of the ion, and m the mass of the ion. The magnetic field strength remains constant during the course of the experiment, leaving the q/m ratio to be accurately determined.

$$f_c = \frac{qB}{2\pi m} \quad (1)$$

Since the ions are confined axially by the magnetic field, trapping motion is induced by the introduction of the electric fields used to confine the ions inside of the analyzer cell. The electric fields are generated by placing a small, typically symmetric voltage on the front and back trapping plates, which are perpendicular to the magnetic field. The ions then move along the magnetic field axis in a simple harmonic oscillation, which can be determined by equation 2, where ω_T is the trapping frequency, a is the

distance between the trapping plates, α is a constant related to cell geometry, V_T is the trapping potential, and q and m are the charge and mass of the ion respectively.

$$\omega_T = \frac{2}{a} \sqrt{\frac{\alpha q V_T}{m}} \quad (2)$$

The final mode of motion that ions undergo during FT-ICR experiments is magnetron motion. During magnetron motion, the center of the cyclotron orbit moves around the center axis; this motion is created by the combination of the electric and magnetic fields. This motion is also observed at a significantly lower frequency (1 – 100 Hz). Unlike cyclotron frequencies, magnetron frequencies are mass independent. The magnetron frequency of an ion in a perfectly quadrupolar static trapping field can be determined by equation 3, where α is a constant based on cell geometry, V_T is the trapping potential, a is the diameter of the cell, and B the magnetic field strength.³⁸

$$\omega_m = \frac{2\alpha V_T}{a^2 B} \quad (3)$$

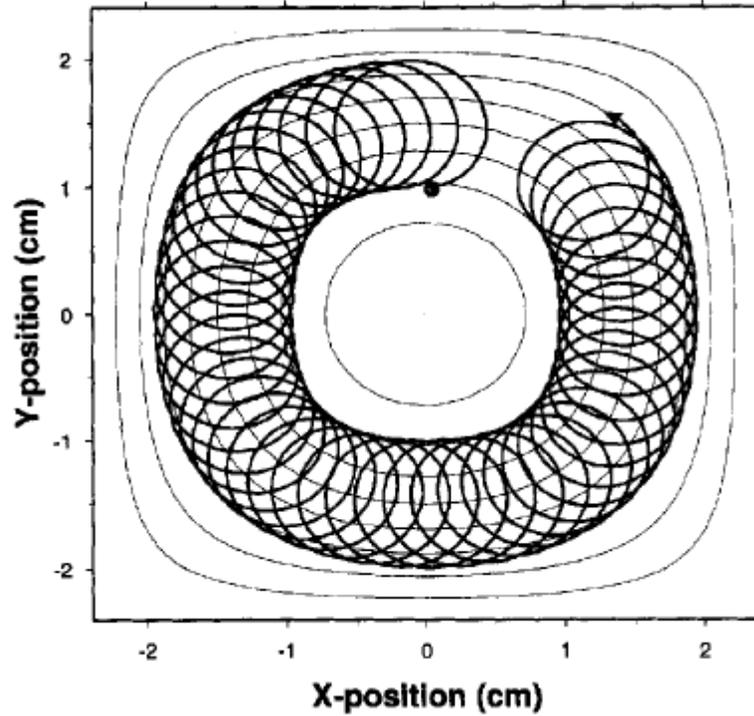


Figure 1.7 Ion trajectory illustrating the combination of cyclotron and magnetron motion on an ion. The magnetron motion is illustrated by the slow progression about the center axis of the analyzer cell and the cyclotron motion by the higher frequency rotation along the magnetron path.³⁸

With all of these motions affecting the ions during the FT-ICR measurement, the measured frequencies are shifted from the theoretical unperturbed cyclotron frequencies by a small fraction. Another significant variable that can affect the observed frequencies is presence and magnitude of space charge.³⁹ The observed cyclotron frequency (ω_o) is a result of the unperturbed cyclotron frequency (ω_c), the magnetron frequency (ω_m), and space-charge component(δ ⁴⁰), shown respectively in equations 4 and 5.⁴¹

$$\omega_o = \omega_c - \omega_m - \delta \quad (4)$$

$$\omega_o = \frac{qB}{m} - \frac{2\alpha V_T}{a^2 B} - \frac{q\rho G_i}{\epsilon_0 B} \quad (5)$$

Space-charge is the perturbation of the electric fields experienced by the ions, due to the interactions with ions of differing mass-to-charges and is represented as the third term in

equations 5, where q is the charge, ρ the ion density, G_i the ion cloud geometry and ϵ_0 the permeability of free space.³⁹ Corrections for space-charge in mass spectral data can be performed by attempting to match the ion density in the analyzer cell during mass calibration.^{42,43}

Ion excitation and Detection

During the course of a FT-ICR experiment, there is a series of events that occurs: ion generation, ion excitation, and ion detection. The ions can be created either external or internal to the analyzer cell. External ion sources, such as ESI or MALDI, must then have the ions transferred to the analyzer region via a series of electrostatic or RF lenses. Ions can also be generated within the confines of the analyzer cell by electron impact ionization (EI). Once the ions are trapped in the analyzer cell, they begin to undergo cyclotron motion, but with a very small radius. A RF pulse is then applied to the excitation plates. All ions of the same m/z will be excited to larger radii when the frequency of the RF is in resonance with the ion's cyclotron frequency and will form a coherent cloud.

There are several types of excitation waveforms that can be used, most of which are comprised of multiple frequencies to ensure that all of the ions are excited to the higher cell radius. The measurement of multiple populations varying in m/z is called broadband detection. The most common waveforms used during these types of measurements is a "chirp". This rapid frequency sweep provides excitation over a relatively broad range in a short time, although it can provide non-uniform excitation across the frequency range and is limited by the start and end frequencies. These

disadvantages were addressed by the development of an optimal excitation waveform called stored waveform inverse Fourier transform (SWIFT). This waveform is constructed by specifying the characteristics of the excitation profile in the mass domain, then converting it into the frequency domain, and performing an inverse Fourier transform to produce the time domain excitation waveform.⁴⁴

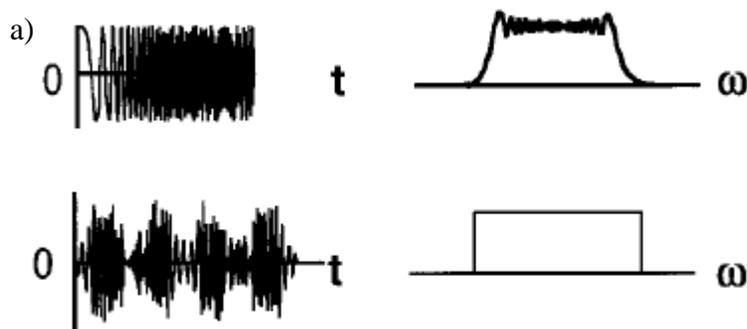


Figure 1.8 Time-domain (left) and frequency-domain (right) excitation waveforms. a) Chirp excitation, b) SWIFT excitation. Adapted from Marshall *et al.*⁴⁵

As the ion cloud passes by the detection plates at the higher cell radius, a current is produced (Figure 1.8). When all of the ions in the cell are simultaneously excited, the image current (transient) produced is a combination of different frequencies. The individual frequencies for each ion population can be obtained by performing a fast Fourier transform (FFT) to the time domain signal. The mass spectrum can be obtained by applying a calibration formula to the frequency spectrum obtained from the FFT. FT-ICR detection is nondestructive, unlike other types of mass spectrometry where the ions are detected by collisions with an electron multiplier, and allows for high mass resolution to be obtained. Since the resolution is proportional to the transient ($R = f_c T/2$), longer transients should be obtained in order to gain the highest mass resolution.

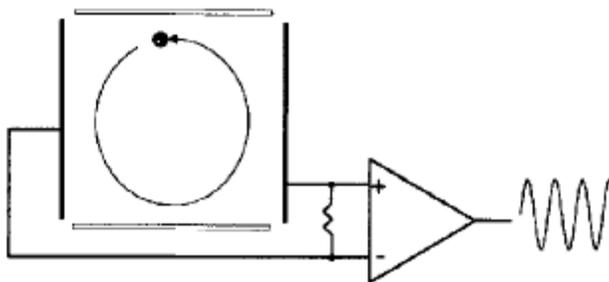


Figure 1.9 Production of the image current of the ions in the analyzer cell during the detection stage of the ICR experiment. Modified from Amster 1996.³⁸

Electrostatic Axially Harmonic Orbital Trapping: Orbitrap

The Orbitrap analyzer was first described in 2000 by Makarov⁴⁶ and shares several features with FT-ICR mass spectrometers, including high mass accuracy and high resolving power. Its method of detection, however, is quite different with the most obvious difference is the lack of magnetic field.

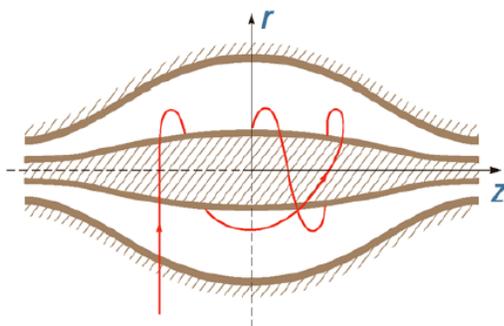


Figure 1.10: Schematic of an Orbitrap analyzer cell.⁴⁷

The analyzer is comprised of a curved outer electrode, split in half, surrounding a center spindle electrode that is held at high voltage. Ions are injected into the space between the electrodes where they rotate around the center electrode while oscillating along the spindle axis. The oscillations induce a current on the space on the outer

electrode. This frequency undergoes a Fourier transform to extract all of the constituent frequencies, which can be converted to m/z spectra.⁴⁶

CHEMICAL CROSSLINKING

Traditional approaches

Chemical crosslinking reactions have been used to study noncovalent molecular complexes between various biomolecules by probing specific interactions.⁴⁸ The use of mass spectrometry for analysis of chemical crosslinking experiments has several advantages.⁴⁹ The high sensitivity of the instrumentation allows for small amount of sample and crosslinker to be used. The ability to perform tandem mass spectrometry experiments also allows for sequence conformation to be performed. High resolution instrumentation, such as Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR/MS), typically allows for higher resolving power and better mass accuracy, leading to much higher confidence in crosslinking assignments.^{50,51} Chemical crosslinking paired with mass spectrometry can be a powerful tool in determining structural information of biomolecules, including peptides, proteins, oligonucleotides, and their interactions with each other and small molecules.⁵²

The complexity of the samples during chemical crosslinking experiments can make their analysis difficult for several reasons. First is that the peaks that yield the most useful structural information are not going to be the most abundant products in the reaction mixture. From any chemical crosslinking experiment, there are several possible reaction products (Figure 1.11): unreacted peptides, peptides that have reacted with the a crosslinker that hydrolyzed before reacting with another peptide (dead-ended), intra-

peptide (cyclic) crosslinks, two peptides that are crosslinked to each other, and reaction products linking more than two peptides together (higher ordered).⁵³ On top of this, the complexity from standard proteomic analyses is still present, including post translational modifications and missed proteolytic cleavages.

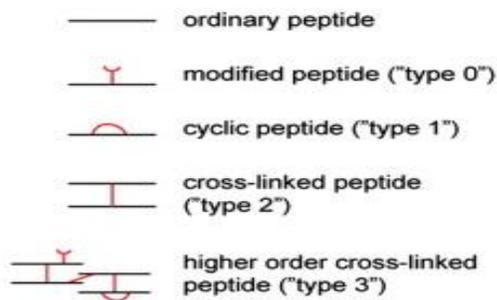


Figure 1.11 Some of the possible crosslinker reaction products expected during a traditional experiment. Type 2 links being considered the most structurally useful. Modified from Rappsilber 2011.⁵³

Another type of limitation often seen in chemical crosslinking experiments is due to the effect the crosslinking reagent itself has on the mass spectrometry analysis of the mixture. When the crosslinker reacts with a peptide, typically at lysine residues, there is an increase in the number of missed cleavages during the proteolytic digestion due to the now modified lysine not being cleaved during the enzymatic digestion, typically performed by trypsin. Reaction of the crosslinker with lysine has a secondary effect of decreasing the overall charge and ionization efficiency of the peptide, leading to decreased fragmentation efficiency as well. As peptides get larger, assignments based on accurate mass are less confident, even with high mass accuracy instrumentation. This leads to requiring fragmentation, which has lower efficiency on lowly charged peptides, for assignment.

The last major limitation in traditional chemical crosslinking experiments is the computational requirements for prediction and assignment of spectra. As the number of crosslinker reactive sites in a peptide or protein increases, the number of binary combinations of possible linked peptides increases factorially.⁴⁹ Traditional analysis is typically performed by having a stage of liquid chromatography prior to injection of the sample into the mass spectrometer. (Figure 1.12) In order to have a preprogrammed list of peaks for automated tandem analysis, a list of all possible reaction products would need to be generated. This list will be incomplete unless all possible participating proteins are already known. This is not beneficial if trying to probe complexes to find participating proteins. Once the peaks are isolated and fragmented, there is still the task of sequencing the fragments to obtain the identities of the participating peptides. The ease of assignment of fragmentation spectra would be dependent on the lability of the crosslinker compared to the peptide bonds. If the peptide bonds are more labile than the bonds in the crosslinker, the spectra may require manual interpretation as the fragments observed may not be predictable. If the most labile bonds reside in the crosslinker, MS^3/MS^n would be required to then sequence the peptides, which would require additional LC runs, with additional preference lists.

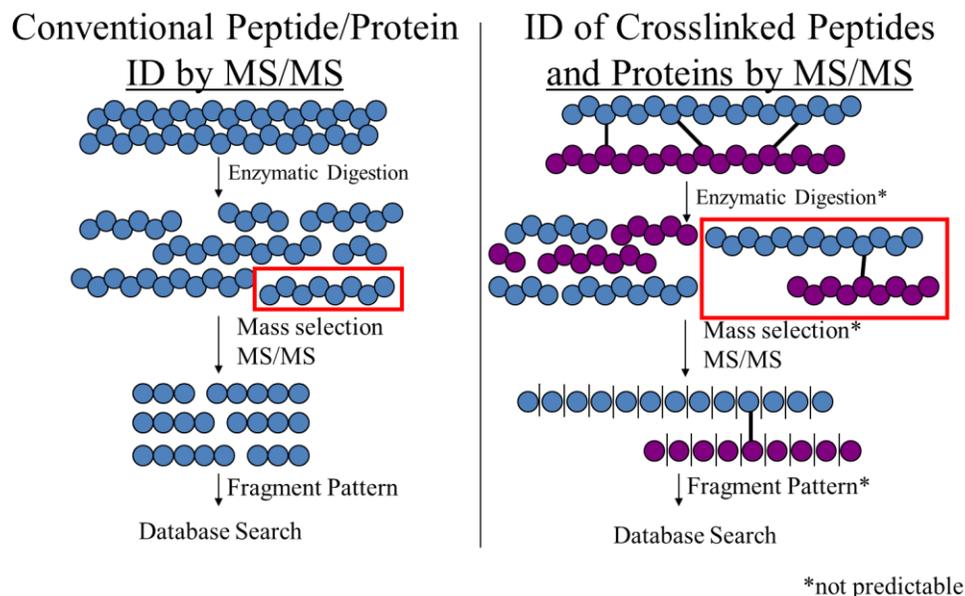


Figure 1.12: Schematic comparing conventional peptide and protein workflows to those used in crosslinking experiments. During crosslinking experiments, each step increases the difficulty in predicting the expected products in the next step.

Mass Defect and Labeling

Mass defect is the difference between the exact mass of a compound and the nominal molecular weight calculated using integer atomic masses. The mass defect in peptides is primarily influenced by the amount of hydrogen present due to the fact that it has the most significant mass defect out of the elements typically found in peptides (C, 0.0 mmu; H, 7.8 mmu; O, -5.1 mmu; N, 3.1mmu; and S), as well as the relative number of hydrogens present. A peptide of averagine, the average elemental ratio for an amino acid residue($C_{4.9384}H_{7.7583}N_{1.3577}O_{1.4773}S_{0.0417}$)⁵⁴, at nominal mass m/z 1000 (exact mass 999.7136 amu) has 96% of its mass defect due to hydrogen.

Table 1.1 Mass defect contributions of the elements typically found in the peptides. The average composition is based on a molecule of average with a nominal mass of m/z 1000, corresponding to an exact mass of 999.7136 amu.

<u>Element</u>	<u>Exact Mass</u>	<u>Nominal Mass</u>	<u>Mass Defect</u>	<u>Composition Average</u>	<u>Mass Defect Contribution</u>
C	12.0000	12	0.0000	44	0
H	1.0078	1	0.0078	95	0.74337
N	14.0031	14	0.0031	12	0.03684
O	15.9959	16	-0.0041	13	-0.05317
S	31.9721	32	-0.0279	0	0

Although the mass defect is sizable in peptides due to the large number of atoms found in them, the distribution of the relative mass defects is quite narrow, allowing the use of labeling agents containing significant mass defects to be used as a means of identifying peptides with specific features, such as cysteines.⁵⁵ (Figure 1.13) This narrow distribution characteristic is not unique to peptides. The distinctive mass defects of perfluoroalkanes⁵⁶ have allowed them to be quite successful calibrants, as they do not overlap with most organic compounds.

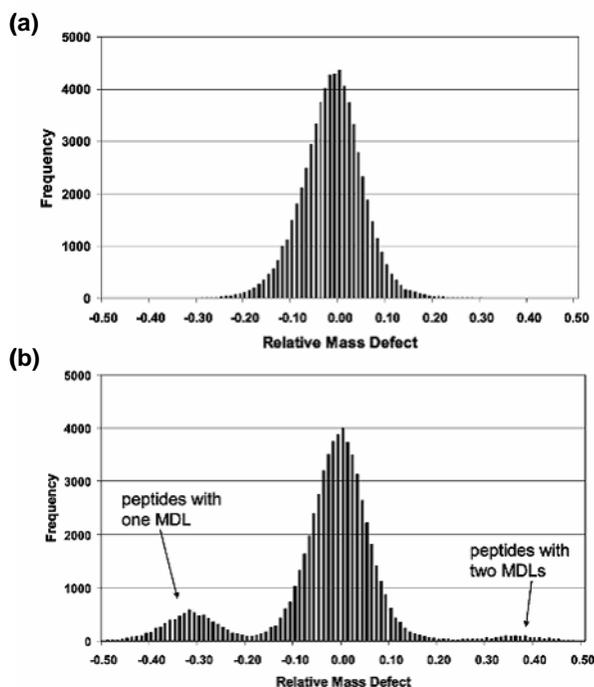


Figure 1.13 Histogram of the relative mass defect versus frequency observed in tryptic peptides. a) Relative mass defect, the difference between the observed mass defect of a peptide and the average mass defect for all peptides of the same nominal mass, observed in 66,000 tryptic peptides derived from *Methanococcus maripaludis* ranging from 700-3000 amu. b) Relative mass defects observed when all peptides containing cysteine are labeled with a mass defect label. Peptides exhibiting a relative mass defect of approximately -0.3 amu contain one labeled cysteine, and those at +0.4 amu contain two labeled cysteines. Adapted from Heranadez *et al.*⁵⁵

Complex mixtures, such as petroleum or dissolved organic matter, have benefited from the use of Kendrick plots, which plot the mass defect versus the nominal mass as a means of classifying the compositions present.^{57,58,59,60,61} Due to their small mass defects, the peptide molecular weights occupy only one-third of a unit mass in any given nominal mass, causing masses to overlap and consequently reducing the specificity of identification, and the complexity in making assignments.

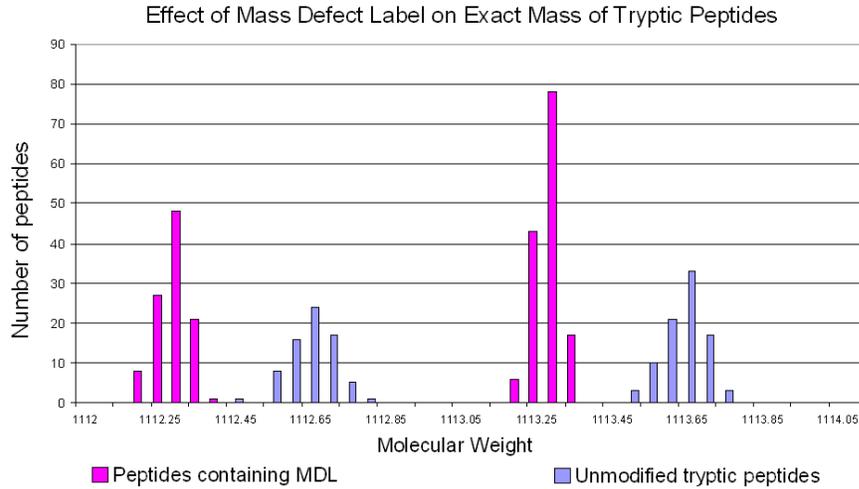


Figure 1.14 Histogram of peptides within a 2 amu range. The blue bars (right-most of each pair) represent standard unmodified tryptic peptides and the magenta bars (left-most of each pair) represent the positions of peptides containing the mass defect label. The mass defect allows for easy assignment of peaks as reaction products as the distributions of modified and unmodified peptides do not occupy the same region within each amu span.

As the mass defect of a peptide at a particular mass to charge is predictable, reaction products can be easily marked by comparing the measured mass defect of an observed peak to the expected mass defect. The difference between these values will be relatively small (± 0.2 Da) for peptides not containing the mass defect label, as shown in Figure 1.14. A peak list can be run through an algorithm, shown in Figure 1.15, to quickly assess the nature of the observed peaks.

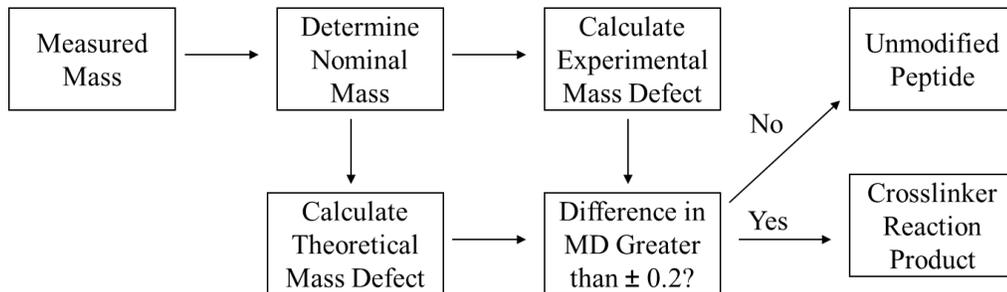


Figure 1.15 Simplified algorithm that can be used to determine whether an observed peak in a peptide mixture has been modified with a mass defect label.

OBJECTIVE

This dissertation explores the use of Fourier transform mass spectrometry, primarily Fourier transform ion cyclotron resonance mass spectrometry, in the analysis of biomolecules. The experimental techniques, including sample preparation and instrumentation parameters, used to analyze the samples in each of the experiments are discussed in Chapter 2. The first experiment, examined in Chapter 3, describes the shotgun proteomic analysis and assignment via accurate mass measurement of an endogenously labeled proteome sample from *Methanococcus maripaludis*. Accurate mass measurement is also the central theme of Chapter 4, where metabolites and enzymes were isolated from *Ruegeria pomeroyi* DSS-3, and analyzed to elucidate the metabolic pathway of dimethylsulphonioproionate. Chapter 5 assesses the robustness and practicality of chemical crosslinking using a novel mass defect labeled crosslinker, using ESI-FAIMS-FT-ICR/MSⁿ, LC-ESI-FT-ICR/MS and nanoLC-ESI-Orbi-FTMS. In Chapter 6, possible mechanisms involved during the ion formation of curcumin analogues in intermediate pressure MALDI are interrogated. The final chapter of this dissertation addresses conclusions regarding the role and benefits of Fourier transform mass spectrometry in the analysis of biomolecules.

REFERENCES

1. Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J.-C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humphery-Smith, I., From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Nat. Biotechnol.* **1996**, *14* (1), 61-65.
2. Wasinger, V. C.; Cordwell, S. J.; Cerpa-Poljak, A.; Yan, J. X.; Gooley, A. A.; Wilkins, M. R.; Duncan, M. W.; Harris, R.; Williams, K. L.; Humphery-Smith, I., Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* **1995**, *16* (1), 1090-1094.
3. Crick, F., Central dogma of molecular biology. *Nature* **1970**, *227* (5258), 561-563.
4. Dove, A., Proteomics: translating genomics into products? *Nat. Biotechnol.* **1999**, *17* (3), 233-236.
5. Lane, C., Mass spectrometry-based proteomics in the life sciences. *Cell. Mol. Life Sci.* **2005**, *62* (7-8), 848-869.
6. Beevers, L., Post-translational modifications. In *Nucleic Acids and Proteins in Plants I*, Springer: 1982; pp 136-168.
7. Peng, J.; Gygi, S. P., Proteomics: the move to mixtures. *J. Mass Spectrom.* **2001**, *36* (10), 1083-1091.
8. Yates III, J. R., Mass spectrometry: from genomics to proteomics. *Trends Genet.* **2000**, *16* (1), 5-8.
9. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246* (4926), 64-71.
10. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T., Protein and polymer analyses up to m/z 100,000 by laser ionization time of flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **1988**, *2* (8), 151-153.
11. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **1988**, *60* (20), 2299-2301.
12. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003**, *422* (6928), 198-207.
13. Bogdanov, B.; Smith, R. D., Proteomics by FT-ICR mass spectrometry: top down and bottom up. *Mass Spectrom. Rev.* **2005**, *24* (2), 168-200.

14. Chait, B. T., Mass Spectrometry: Bottom-Up or Top-Down? *Science* **2006**, *314* (5796), 65-66.
15. Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W., Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proc. Natl. Acad. Sci.* **2002**, *99* (4), 1774-1779.
16. Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J.-H.; Begley, T. P.; McLafferty, F. W., Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc.* **2002**, *124* (4), 672-678.
17. Siuti, N.; Kelleher, N. L., Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* **2007**, *4* (10), 817-821.
18. Han, X.; Jin, M.; Breuker, K.; McLafferty, F. W., Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science* **2006**, *314* (5796), 109-112.
19. Wolters, D. A.; Washburn, M. P.; Yates, J. R., An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001**, *73* (23), 5683-5690.
20. Washburn, M. P.; Wolters, D.; Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19* (3), 242-247.
21. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization—principles and practice. *Mass Spectrom. Rev.* **1990**, *9* (1), 37-70.
22. Fernandez De la Mora, J., Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta* **2000**, *406* (1), 93-104.
23. Sanz-Medel, A., Analytical atomic spectrometry going into the next millennium: photons or ions, atoms or molecules? Presented at SAC 99, Dublin, Ireland, July 25–30, 1999. *Analyst* **2000**, *125* (1), 35-43.
24. Smith, R. D.; Loo, J. A.; Loo, R. R. O.; Busman, M.; Udseth, H. R., Principles and practice of electrospray ionization—mass spectrometry for large polypeptides and proteins. *Mass Spectrom. Rev.* **1991**, *10* (5), 359-452.
25. Siuzdak, G., *The expanding role of mass spectrometry in biotechnology*. Mcc Press San Diego:: 2003.

26. Jaskolla, T. W.; Karas, M., Compelling evidence for lucky survivor and gas phase protonation: the unified MALDI analyte protonation mechanism. *J. Am. Soc. Mass. Spectrom.* **2011**, 22 (6), 976-988.
27. Hoffmann, E., *Mass spectrometry*. Wiley Online Library: 1996.
28. Zenobi, R.; Knochenmuss, R., Ion formation in MALDI mass spectrometry. *Mass Spectrom. Rev.* **1998**, 17 (5), 337-366.
29. Karas, M.; Bachmann, D.; Bahr, U. e.; Hillenkamp, F., Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int. J. Mass Spectrom. Ion Processes* **1987**, 78, 53-68.
30. Karas, M.; Glückmann, M.; Schäfer, J., Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J. Mass Spectrom.* **2000**, 35 (1), 1-12.
31. Schneider, B. B.; Covey, T. R.; Coy, S. L.; Krylov, E. V.; Nazarov, E. G., Planar differential mobility spectrometer as a pre-filter for atmospheric pressure ionization mass spectrometry. *Int. J. Mass Spectrom.* **2010**, 298 (1), 45-54.
32. Purves, R. W.; Guevremont, R., Electrospray ionization high-field asymmetric waveform ion mobility spectrometry-mass spectrometry. *Anal. Chem.* **1999**, 71 (13), 2346-2357.
33. Lawrence, E. O.; Edlefsen, N. E., The National Academy of Sciences. *Science* **1930**, 72 (1867), 372-378.
34. Hipple, J.; Sommer, H.; Thomas, H. A., A precise method of determining the Faraday by magnetic resonance. *Phys. rev.* **1949**, 76 (12), 1877.
35. Comisarow, M. B.; Marshall, A. G., Fourier transform ion cyclotron resonance spectroscopy. *Chem. Phys. Lett.* **1974**, 25 (2), 282-283.
36. Marshall, A.; Hendrickson, C.; Emmett, M.; Rodgers, R.; Blakney, G.; Nilsson, C., Fourier transform ion cyclotron resonance: state of the art. *Eur. J. Mass Spectrom.* **2007**, 13 (1), 57-60.
37. Schrader, W.; Klein, H.-W., Liquid chromatography/Fourier transform ion cyclotron resonance mass spectrometry (LC-FT-ICR MS): an early overview. *Anal. Bioanal. Chem.* **2004**, 379 (7-8), 1013-1024.
38. Amster, I. J., Fourier transform mass spectrometry. *J. Mass Spectrom.* **1996**, 31 (12), 1325-1337.
39. Jeffries, J.; Barlow, S.; Dunn, G., Theory of space-charge shift of ion cyclotron resonance frequencies. *Int. J. Mass Spectrom. Ion Processes* **1983**, 54 (1), 169-187.

40. Beauchamp, J. L.; Armstrong, J. T., An Ion Ejection Technique for the Study of Ion-Molecule Reactions with Ion Cyclotron Resonance Spectroscopy. *Rev. Sci. Instrum.* **1969**, *40* (1), 123-128.
41. Francl, T. J.; Sherman, M. G.; Hunter, R. L.; Locke, M. J.; Bowers, W. D.; McIver Jr, R. T., Experimental determination of the effects of space charge on ion cyclotron resonance frequencies. *Int. J. Mass Spectrom. Ion Processes* **1983**, *54* (1), 189-199.
42. Ledford Jr, E. B.; Rempel, D. L.; Gross, M., Space charge effects in Fourier transform mass spectrometry. II. Mass calibration. *Anal. Chem.* **1984**, *56* (14), 2744-2748.
43. Easterling, M. L.; Mize, T. H.; Amster, I. J., Routine part-per-million mass accuracy for high-mass ions: Space-charge effects in MALDI FT-ICR. *Anal. Chem.* **1999**, *71* (3), 624-632.
44. Marshall, A. G.; Wang, T. C. L.; Ricca, T. L., Tailored excitation for Fourier transform ion cyclotron mass spectrometry. *J. Am. Chem. Soc.* **1985**, *107* (26), 7893-7897.
45. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **1998**, *17* (1), 1-35.
46. Makarov, A., Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal. Chem.* **2000**, *72* (6), 1156-1162.
47. Thermo Fisher Scientific, I., *Thermo Fisher Scientific Orbitrap Elite Hardware Manual*. Thermo Fisher Scientific: Bremen, Germany, 2011.
48. Back, J. W.; de Jong, L.; Muijsers, A. O.; de Koster, C. G., Chemical cross-linking and mass spectrometry for protein structural modeling. *J. Mol. Biol.* **2003**, *331* (2), 303-313.
49. Petrotchenko, E. V.; Borchers, C. H., Crosslinking combined with mass spectrometry for structural proteomics. *Mass Spectrom. Rev.* **2010**, *29* (6), 862-876.
50. Sinz, A., Chemical cross-linking and FT-ICR mass spectrometry for protein structure characterization. *Anal. Bioanal. Chem.* **2005**, *381* (1), 44.
51. Ihling, C.; Schmidt, A.; Kalkhof, S.; Schulz, D. M.; Stingl, C.; Mechtler, K.; Haack, M.; Beck-Sickinger, A. G.; Cooper, D. M. F.; Sinz, A., Isotope-labeled cross-linkers and Fourier transform ion cyclotron resonance mass spectrometry for structural analysis of a protein/peptide complex. *J. Am. Soc. Mass. Spectrom.* **2006**, *17* (8), 1100.

52. Sinz, A., Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom. Rev.* **2006**, *25* (4), 663.
53. Rappsilber, J., The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **2011**, *173* (3), 530-540.
54. Senko, M. W.; Beu, S. C.; McLafferty, F. W., Automated Assignment of Charge States from Resolved Isotopic Peaks for Multiply Charged Ions. *Journal of the American Society for Mass Spectrometry* **1995**, *6* (1), 52-56.
55. Hernandez, H.; Niehauser, S.; Boltz, S. A.; Gawandi, V.; Phillips, R. S.; Amster, I. J., Mass defect labeling of cysteine for improving peptide assignment in shotgun proteomic analyses. *Anal. Chem.* **2006**, *78* (10), 3417-3423.
56. Beynon, J. H.; Willaims, A. E., *Mass and Abundance Tables for Use in Mass Spectrometry*. 2nd ed.; Elsevier: Amsterdam, 1963.
57. Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K., Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem.* **2001**, *73* (19), 4676-4681.
58. Wu, Z. G.; Rodgers, R. P.; Marshall, A. G., Two- and three-dimensional van Krevelen diagrams: A graphical analysis complementary to the Kendrick mass plot for sorting elemental compositions of complex organic mixtures based on ultrahigh-resolution broadband Fourier transform ion cyclotron resonance mass measurements. *Anal. Chem.* **2004**, *76* (9), 2511-2516.
59. Stenson, A. C.; Marshall, A. G.; Cooper, W. T., Exact masses and chemical formulas of individual Suwannee River fulvic acids from ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectra. *Anal. Chem.* **2003**, *75* (6), 1275-1284.
60. Marshall, A. G.; Rodgers, R. P., Petroleomics: The next grand challenge for chemical analysis. *Acc. Chem. Res.* **2004**, *37* (1), 53-59.
61. Kendrick, E., A mass scale based on $\text{CH}_2 = 14.0000$ for high resolution mass spectrometry of organic compounds. *Anal. Chem.* **1963**, *35*, 2146-2154.

CHAPTER 2

EXPERIMENTAL

This chapter describes the instrumentation, sample preparation techniques, and materials used during the shotgun proteomic analysis (Chapter 3), metabolic pathway determination (Chapter 4), chemical crosslinking experiments (Chapter 5), and the analysis of curcumin analogues (Chapter 6). All of these experiments Fourier transform mass spectrometry (FTMS), specifically Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR/MS) due to its high sensitivity, mass accuracy, and resolving power, which can be utilized with many other techniques and technologies to analyze a variety of biological molecules.

INSTRUMENTATION

High Performance Liquid Chromatography (HPLC)

Shotgun proteomics analysis

The complex peptide mixtures examined in Chapter 3 were separated (Figure 2.1) via reverse phase liquid chromatography using a Dionex LC Packings Ultimate Plus system (Sunnyvale, CA) controlled by Chromeleon software (Version 6.60, Build 1428). The separation was performed offline from the mass spectrometer. The FAMOS

microautosampler injected (full loop injection) the peptide mixture at a flow rate of 30 $\mu\text{L}/\text{min}$ using a 20 μL injection loop onto a reverse phase nano HPLC column (75 μm I.D. x 15 cm) packed with C18 stationary phase (Dionex LC Packings, Sunnyvale, CA). The mobile phase is filtered and degassed with helium prior to eluting the peptides from the column using a shallow gradient of Solvent A (95/5/0.1% water/acetonitrile/trifluoroacetic acid, by volume) decreasing in relation to Solvent B (20/80/0.1% water/acetonitrile/trifluoroacetic acid, by volume) over 130 minutes.



Figure 2.1 Sample chromatogram showing the elution gradient used. The x-axis is time (min) and the y-axis is the measured absorbance in arbitrary units. The sample shown here is the mixture of peptides from a proteome sample derived from a whole cell lysate from *Methanococcus maripaludis* Δlrp .

The eluted peptides were detected via the UV detector monitoring at 214 nm, and spotted onto the MALDI targets using the Probot Micro Fraction collector, using the $\mu\text{Carrier}$ software (version 1.68). The robotic fraction collector is designed to take fractions, eluants from specific time spans, from the chromatographic separation and place them onto individual spots on the MALDI target for analysis. The distance between the UV detector and the spotter was minimized to decrease the mixing of components post detection. Once the collected fractions dried on the target, 0.4 μL of 1M 2,5-dihydroxybenzoic acid (DHB) is added on top of each spot and allowed to dry prior

to analysis. The DHB solution was prepared by dissolving purified and recrystallized DHB into 50/50/0.1% water/acetonitrile/trifluoroacetic acid (by volume).

Metabolite analysis

The separations of the Coenzyme A metabolites from the *R. pomeroyi* DSS-3 cellular extracts were performed using several HPLC gradients and columns. For the initial extraction of the acyl-CoA mixture, an oligo purification cartridge (OPC, Applied Biosystems) was used. It was pre-washed and the separation was performed using 70: 30 (ACN: KH₂PO₄) for both steps. The products that remained on the cartridge were analyzed via HPLC using a 4.6 x 250 mm, 5 µm particle size, Aquasil column (ThermoFisher). At a flow rate of 1 mL/min, a gradient of 2-20% ACN in 50 mM ammonium acetate (pH 6.0) was used for the 40 minute run, where the products were detected using a 260 nm UV absorbance detector.

Purification of the enzymes responsible for catalyzing the metabolism was performed using many high performance sepharose columns (GE Healthcare). The cellular extracts were loaded onto a Q-Sepharose HP anion exchange column, using 50mM Tris-HCl (pH 8.0) to equilibrate the column. With a flow rate of 2 mL/min, the protein samples were eluted using a 0-1M NaCl gradient. The fractions collected from this separation were then loaded onto a phenyl-Superose HR hydrophobic interaction column. Using a flow rate of 1 mL/min, the proteins were eluted using a gradient of 0-1.7M (NH₄)SO₄ in 50 mM Tris-HCl. The fractions from this separation that showed enzymatic activity were pooled and concentrated.

The concentrated protein samples were then separated using a HiTrap Blue chromatography column (GE Healthcare), which was equilibrated with potassium

phosphate (pH 7.5). The samples were then eluted using a 0-2M KCl gradient. The fraction with the highest activity was concentrated and applied to a Sephacryl S200 gel filtration column (1.6 x 25 cm) that had been equilibrated with 50mM Tris-HCl (pH 7.5) and 150mM NaCl. The elution was performed with the buffer solution at a flow rate of 1 mL/min, and the active fractions collected were analyzed on a SDS-PAGE gel and stained with GelCode Blue. The excised gel bands were analyzed via mass spectrometry.

Chemical Crosslinking

The reaction mixtures discussed in Chapter 4 were separated via reverse phase liquid chromatography using an Agilent 1100 Series system (Santa Clara, CA), composed of a Agilent G1312A binary solvent delivery system and Agilent G1389A autosampler. The samples were injected by the autosampler and loaded onto the C18 column (PepMap C18, 300 μ m x 15 cm, 3 μ m, 100 Å, LC Packing). The flow rate was set at 0.400 mL/min prior to the adjustable flow splitter (Analytical Scientific Instruments, Richmond, CA). The splitter was adjusted so that the final flow rate post column was approximately 8-10 μ L/min. Solvent A (99.9% H₂O, 0.1% formic acid) and Solvent B (99.9% acetonitrile, 0.1% formic acid) were used to create a gradient that was optimized for each mixture. Sample gradients used are listed in Table 2.1 below.

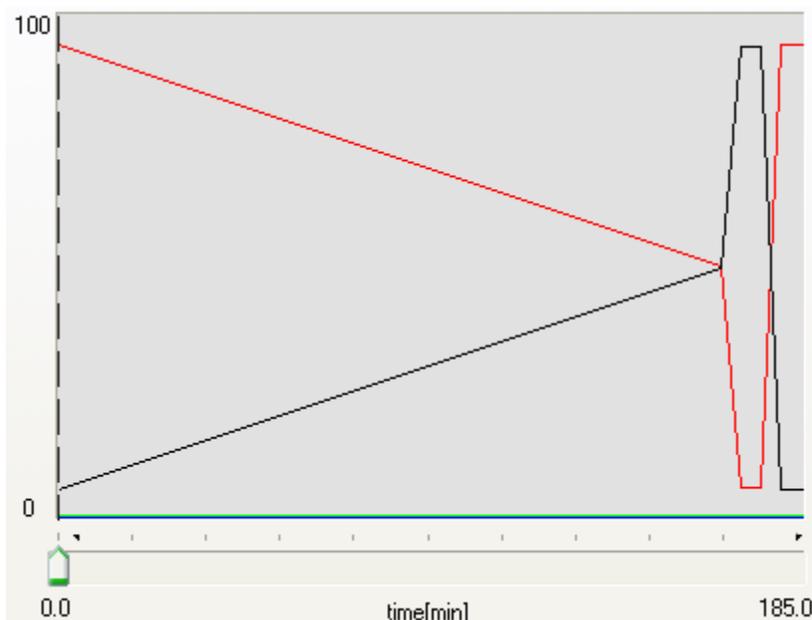


Figure 2.2 Liquid chromatography gradient used during a 180 minute experiment. The red line is mobile phase A, and the black line is mobile phase B. The x-axis measures percent volume of mobile phase and the y-axis is time measured in minutes.

Table 2.1 Gradients used over the course of the chemical crosslinking experiments. Solvent A (99.9% H₂O, 0.1% formic acid) and Solvent B (99.9% acetonitrile, 0.1% formic acid) were used in each gradient.

<u>90 Minute Gradient</u>		<u>120 Minute Gradient</u>		<u>180 Minute Gradient</u>	
<u>B%</u>	<u>Time (min)</u>	<u>B%</u>	<u>Time (min)</u>	<u>B%</u>	<u>Time (min)</u>
5% to 50%	0 to 75	5% to 50%	0 to 105	5% to 50%	0 to 165
50% to 95%	75 to 80	50% to 95%	105 to 110	50% to 95%	165 to 170
95% to 95%	80 to 85	95% to 95%	110 to 115	95% to 95%	170 to 175
95% to 5%	85 to 90	95% to 5%	115 to 120	95% v 5%	175 to 180

All LC parameters (autosampler, binary pump, and gradient) were controlled through the Hystar Version 3.4 software (Bruker Daltonics, Billerica, MA). This software also coordinates the LC with the mass spectrometer data acquisition software, ApexControl (Bruker Daltonics, Billerica, MA). For the mass spectrometric analysis of these samples, the liquid chromatograph was coupled to the instrument by using the post column capillary as the spray line into the entrance of the mass spectrometer.

Intermediate Pressure Matrix Assisted Laser Desorption Ionization Fourier Transform Ion Cyclotron Resonance (IP-MALDI-FT-ICR) 7T

Shotgun proteomic analysis, metabolite analysis, and curcumin analogue analysis were performed using matrix assisted laser desorption ionization (MALDI) as the ionization technique on the 7T BioApex FT-ICR mass spectrometer with an intermediate pressure Scout100 (IP-MALDI) source (Bruker Daltonics, Billerica, MA). The samples were spotted onto the stainless steel MALDI target and the matrix solution is added and allowed to dry under atmospheric conditions. The target was inserted into the source and ionization was performed using a 337 nm UV laser, and argon at 10^{-3} torr for collisional cooling of the ions, which decreases the number of metastable ions. The ions are then passed through the transfer optics and enter the analyzer cell. The Infinity Analyzer Cell (Bruker Daltonics, Billerica, MA) allows for mass spectra to be obtained with high mass accuracy.

Data acquisition is performed using either a chirp waveform, produced by the instrument console, or stored waveform inverse Fourier transform (SWIFT) waveform, generated by a PXI chassis with a PXI-8184 embedded controller and a PXI-5412 arbitrary waveform generator (National Instruments, Austin, TX). The SWIFT waveform was generated to produce uniform excitation power across the m/z range (500 to 10,000 Da, corresponding to 215,283 to 10,752 Hz). Spectra were collected using 512k data points in each transient, which was then processed. Spectral processing includes the addition of a zero-fill, apodization with a sinebell function, and fast Fourier transform and magnitude calculation.

Calibration of the mass data was performed in a combination of three ways: external, step-wise external, and internal. External calibration occurred using Bruker XMASS 7.0.8 software (Bruker Daltonics, Billerica, MA) using a tryptic digest of standard proteins (ovalbumin or bovine serum albumin) and peptides such as insulin chain B and angiotensin II. For the curcumin samples, cesium iodide was also used as a low mass calibrant. In the software, the cal2 equation¹ (equation 1) was used on the chosen peaks, yielding calibration constants A and B. These constants were used to convert the m/z values into frequencies, which are needed for the step-wise external calibration.²

$$\frac{m}{z} = \frac{A}{f+B} \quad (1)$$

Step-wise external calibration was performed by recording a mass spectrum at a relatively lower cell trapping potential (0.6V-0.8V). At this voltage the ion capacity in the cell is high enough to observe the ions with decent signal-to-noise, but not high enough to observe significant frequency shifts due to space charge, allowing for higher mass accuracy of the ions observed, but lower sensitivity. The trapping voltage is then increased (1.0V-1.2V) and a second mass spectrum is recorded for the same sample. At high trapping voltages, the cell ion capacity is quite large, allowing for good signal-to-noise of peptides, even those of low abundance. At this higher sensitivity, the trade-off is more interaction between the ion clouds in the cell, decreasing the mass accuracy of the measured peaks.

The peaks observed in the higher mass accuracy experiments, at lower trapping potentials, are used as internal calibrants in the higher trapping potential spectrum data set. Using the equation below (equation 2), the calibration constants A, B, and C are

determined and then used to fit the measured frequencies and intensities of the peaks in both spectra. A minimum of four peaks with signal-to-noise of at least five is needed for each mass spectrum.

$$\left(\frac{m}{z}\right)_i = \frac{A}{f_i + B + C \cdot I_i} \quad (2)$$

Matrix Assisted Laser Desorption Ionization Time of Flight (MALDI-TOF)

Curcumin analogue analysis was performed using a Bruker AutoFlex MALDI-TOF (Bruker Daltonics, Billerica, MA), equipped with a 337 nm nitrogen laser, delayed extraction optics, and reflectron mode, and controlled using the flexControl Software (Version 3.4, Build 119). Data viewing and mass list extraction was performed using flexAnalysis Software (version 3.4, Build 70). External calibration was performed using tryptically digested bovine serum albumin.

Electrospray Ionization Fourier Transform Ion Cyclotron Resonance (ESI-Q-FT-ICR) 9.4T

Chemical crosslinking experiments were performed using a 9.4T Apex-Qe FT-ICR (Bruker Daltonics, Billerica, MA) equipped with the Apollo II dual ESI and MALDI source, a 25 W CO₂ laser (Synrad model J48-2, Mukilteo, WA) for IRMPD, and an indirectly heated hollow cathode for generating electrons inside of the analyzer cell for electron fragmentation methods. ApexControl Expert (Version 3.0.0, Build 72) was used to control data acquisition, including the automated MS/MS during the LC analyses. The source can also be coupled with the Agilent 1100 Series HPLC to allow for analysis of the fractions, including tandem mass spectrometric measurements. Direct infusion

experiments were performed with all sample solutions introduced at a rate of 120 $\mu\text{L}/\text{h}$. Calibration was performed using 0.01 mg/mL sodium trifluoroacetate (NaTFA) in 50/50/0.1 water/methanol/formic acid.

High-Field Asymmetric Ion Mobility Spectrometry (FAIMS) Assembly

The separation of the simple peptide mixture of bradykinin and neurotensin in Chapter 4 was performed using a planar FAIMS device (Bruker Daltonics, Billerica, MA). The gap between the two planar electrodes was 0.5 mm. This device was interfaced with the electrospray capillary inlet of the 9.4T FT-ICR/MS instrument by way of a cylindrical peek fitting.

The dispersion voltage (DV) applied for the peptide separations was 1600 volts peak-to-peak. The compensation voltages (CV) scanned during the analysis ranged from -30V to 0V and 0V to 30V. The scans were initially performed to find the regions of interest with 2 spectra acquisitions at each step, with the steps occurring every one volt. Once peaks of interest were observed, the compensation scan was repeated with more acquisitions at each step of the narrower CV range and smaller voltage steps through the scan range. The FAIMS compensation voltages, both scanning and static, were controlled by the CV Scanning Main.vi program (Labview 2009, Version 9.0).

Electrostatic Axially Harmonic Orbital Trapping

For the analysis of the chemically crosslinked bovine hemoglobin in Chapter 4, an LTQ Orbitrap Elite mass spectrometer (Thermo Electron, Bremen, Germany) equipped with a nanospray ion source (Proxeon Biosystems, Odense, Denmark) coupled to an Easy

nLC II nanoliquid chromatography system (Proxeon Biosystems, Odense, Denmark) was employed. Spectra were collected using both the low resolution and high resolution mode in the Thermo Xcalibur (2.07 Build) software.

Data Analysis Software

All spectra collected on the FT-ICR instruments were analyzed using DataAnalysis Version 4.0 Build 281 (Bruker Daltonics, Billerica, MA). The SNAP function (version 2.0) allowed for monoisotopic peaks to be chosen for standard peptides, as well as for peaks containing a mass defect label, and assign a m/z value. Spectra collected using the Orbitap Elite were viewed and mass lists extracted from the QualBrowser section of Xcalibur. Mass lists were manually de-isotoped, and the monoisotopic lists were analyzed with crosslinking software discussed below.

Several pieces of software were used in the analysis of the chemical crosslinking data acquired. Software was used in predicting and assigning the chemical crosslinking reaction products. CLMP and DXSMS were developed by Creative Molecules, Inc. (Victoria, British Columbia, Canada), and were used in the manual assignment of spectral data. For higher throughput of the LC acquired data from the bovine hemoglobin crosslinking experiment, MS-Bridge (University of California, San Francisco, CA) was used to batch process the spectral data.

CLMP provided lists of predicted singly crosslinked peptides from an input of the protein amino acid sequences, intrapeptide mass of the crosslinker, crosslinker reaction sites, enzyme digestion sites, and number of missed cleavages. These lists were used for quick peptide assignment of simple reaction mixtures. By assigning the peaks by

accurate mass, tandem MS was not required for every peptide that exhibited the mass defect label, and more time could be used to investigate the unpredicted, but mass defect labeled peaks. This assisted in ensuring that the crosslinking reactions occurred and produced the expected reaction products.

DXMSMS was used to predict the fragments masses observed from the crosslinked peptides from tandem MS experiments. The reacted protein's amino acid sequence is input, along with digestion parameters, and the characteristics of the crosslinker (M_{ip} , $M_{clrest1}$, and $M_{clrest2}$). The mass observed from a reaction of the crosslinker with two peptides is composed of the masses of the two reacted peptides, $[M_x + H]^+$, plus the intrapeptide mass of the crosslinker, M_{ip} , shown in equation 1.

$$[M_{1,2} + H]^+ = [M_1 + H]^+ + [M_2 + H]^+ + M_{ip} \quad (1)$$

During the fragmentation of this peak, the crosslinker can break in many potential ways, each yielding a different mass. The crosslinker can also not fragment and instead fragmentation is observed along the peptide backbone. In order to create predictions for each scenario, the $M_{clrest1}$, and $M_{clrest2}$ parameters are used to denote the mass of the crosslinker attached to the peptide(s) based on the cleavage site of the crosslinker, as shown in equation 2, where $[M_1CL + H]^+$ is the mass of the protonated crosslinked peptide, $[M_1 + H]^+$ is the mass of the protonated unmodified peptide from the protein sequence, and $M_{clrest1}$, and $M_{clrest2}$ are the masses of the crosslinker that is attached to the peptides post-fragmentation. If the bond at the site of the crosslinker attachment breaks, $M_{clrest1}$ will be equal to the mass of the crosslinker attached to the peptide, and $M_{clrest2}$ will be equal to zero.

$$[M_1CL + H]^+ = [M_1 + H]^+ + M_{clrest1} + M_{clrest2} \quad (2)$$

The MS-Bridge software allowed for input of multiple proteins and data from first stage MS scans, as well as personalization of the crosslinker behavior. The personalization required an input of the elemental composition of the bridging molecule and the amino acids to be linked, as well as the compositions and reactions sites of other possible reaction products (deadends). The output is a list of possible matching crosslinked peptides based on accurate mass.

SAMPLE PREPARATION

All protein samples underwent proteolytic digestion to produce peptides for analysis. The protein samples were dissolved in either 20 mM ammonium bicarbonate buffer, in the case of the extracted cellular proteins in the shotgun proteomic analysis, or in either 100 mM triethylammonium acetate buffer or phosphate buffer (pH 6.8) for the chemical crosslinking experiments.

Once the protein samples were dissolved in the buffer, tris-(2-carboxyethyl)phosphine hydrochloride (TCEP HCl) or dithiothreitol (DTT) was added to the solution in stoichiometry amounts relative to the number of cysteines present in the protein and the samples were heated at 97°C for 10 minutes. The heat denatures the proteins, exposing the disulfide linkages to the solution. TCEP and DTT then reduce these linkages, allowing the protein to fully unfold, as well as leave the reduced sulfides. Alkylation of the sulfides, using iodoacetamide in stoichiometry amounts, was then performed to eliminate the opportunity for the disulfide linkages to reform, either prior to enzymatic digestion or prior to peptide analysis.

After the protein samples were reduced and alkylated, the protease was added. The trypsin added was modified, where the lysines are dimethylated (Promega, Madison, WI) to eliminate enzymatic cleavage of the protease during protein digestion. The trypsin was resuspended in the buffer and added to the protein mixture. A protease to protein ratio for standard proteins is 1:100, however for the analysis of the proteomes, a ratio of 1:50 was used. The solution was incubated overnight at 37°C. After digestion the samples were stored at -80°C to prevent sample degradation.

For the chemical crosslinking experiments, the protein samples underwent exposure to the chemical crosslinking reagent prior to enzymatic digestion. The protein or peptide samples were dissolved in triethylammonium acetate buffer. The mass defect labeled crosslinker was dissolved in dry acetone and stored under an argon atmosphere until the reaction was performed. The protein sample was mixed with the crosslinker solution, at varying molar ratios optimized for each mixture, and vortexed at room temperature for 30 minutes to allow the reaction to reach completion.

The mass defect labeled (MDL) crosslinker, 2,4-dibromobenzenesulfonamide, N,N-di-2-ethanoic acid N-hydroxysuccinimide ester (DiBBSIAS), was synthesized following the reaction shown in Figure 2.3. Dibromobenzenesulfonylchloride and iminoacetic acid were reacted in equimolar concentrations (0.1 mmole) with 0.3 mmole of NaOH in 100µL of water for 24 hours at 75°C. The reaction product, dibromobenzenesulfoneiminodiacetic acid was dissolved in dimethylsulfoxide (DMSO) and activated with N-hydroxysuccinimide and dicyclohexylcarbodiimide overnight at STP. The MDL crosslinker was precipitated from the reaction solution, dried *in vacuo*, and stored at -80°C.

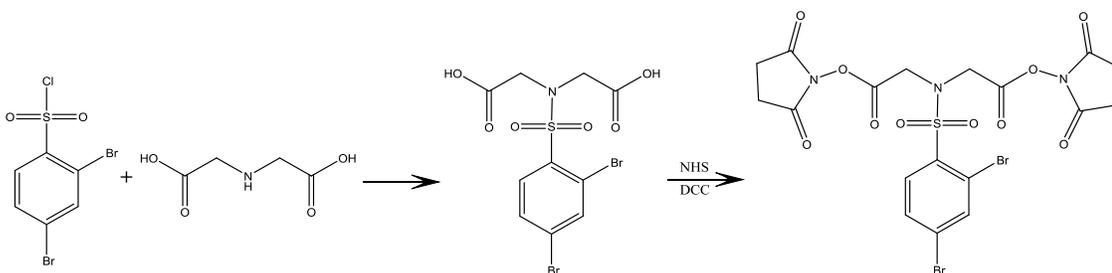


Figure 2.3 Synthesis reaction for the mass defect labeled crosslinker DiBBSIAS, 2,4-dibromobenzenesulfonamide, N,N-di-2-ethanoic acid N-hydroxysuccinimide ester.

Preparation for the curcumin samples entailed dissolving trace amounts of the analytes into acetonitrile to be spotted on the MALDI plate. For the MALDI analysis, 2,5-dihydroxybenzoic acid or 3-hydroxy-4-nitrobenzoic acid was dissolved in acetonitrile at a concentration of 10 mg/mL, and the matrix was spotted atop the dried samples. For the laser desorption analysis, no matrix was added to the MALDI plate prior to sample introduction into the instruments.

Samples were also analyzed via a solvent-free preparation method. For solvent-free sample prep, the analyte crystals were crushed directly onto the MALDI plate with a spatula, and the excess was knocked off prior to the insertion of the plate into the instrument.

MATERIALS

Chapter 3: Shotgun proteome analysis

Ammonium bicarbonate, ovalbumin, insulin chain B, bovine serum albumin, and dithiothreitol were purchased from Sigma (St. Louis, MO). Sequencing grade trypsin was purchased from Promega (Madison, WI). 2,5-Dihydroxybenzoic acid (DHB) was purchased from Lancaster Synthesis (Pelham, NH). Acetonitrile and trifluoroacetic acid

were purchased from Fisher Scientific (Fair Lawn, NJ). All water used was purified using a NanopureInfinity ultrapure water system (Barnstead/Thermolyne, Dubuque, IA).

Chapter 4: Metabolite analysis

Ammonium acetate, sodium phosphate, ammonium sulfate, sodium chloride, potassium chloride, and potassium phosphate were purchased from Sigma (St. Louis, MO). 2,5-Dihydroxybenzoic acid (DHB) was purchased from Lancaster Synthesis (Pelham, NH). Acetonitrile, acetic acid, and trifluoroacetic acid were purchased from Fisher Scientific (Fair Lawn, NJ). Tris-HCl was purchased from Bio-Rad, Inc. (Hercules, CA). All water used was purified using a NanopureInfinity ultrapure water system (Barnstead /Thermolyne, Dubuque, IA). All of the chemicals used during the biological isolations and gene and enzyme amplification are described in detail in the original transcript (Reisch, C. R.; Stoudemayer, M. J.; Varaljay, V. A.; Amster, I. J.; Moran, M. A.; Whitman, W. B., Novel pathway for assimilation of dimethylsulphoniopropionate widespread in marine bacteria. *Nature* **2011**, *473* (7346), 208-211).

Chapter 5: Mass Defect Labeled Crosslinker

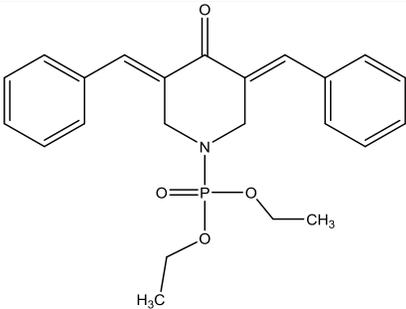
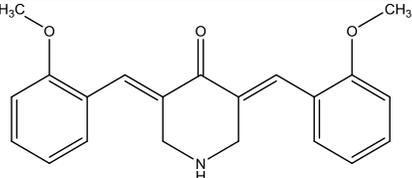
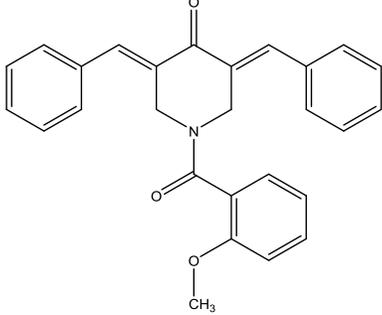
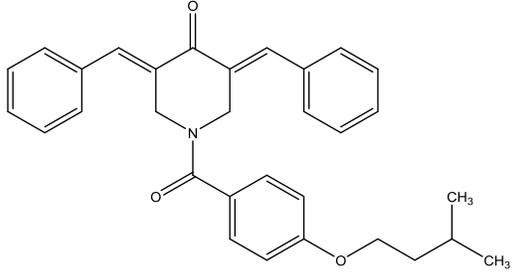
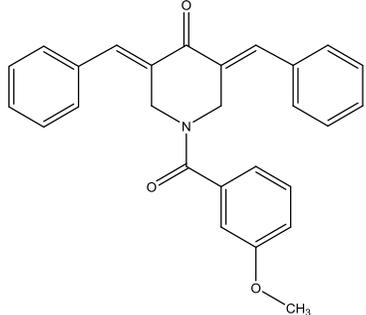
Triethylammonium acetate, potassium phosphate, RNase S, neurotensin, bradykinin, and bovine hemoglobin were purchased from Sigma (St. Louis, MO). Sequencing grade trypsin was purchased from Promega (Madison, WI). 2,5-Dihydroxybenzoic acid (DHB) was purchased from Lancaster Synthesis (Pelham, NH). Water, acetonitrile, and trifluoroacetic acid were purchased from Fisher Scientific (Fair Lawn, NJ). The test peptide was purchased from Creative Molecules, Inc. (Victoria,

British Columbia, Canada). The mass defect labeled crosslinker was synthesized by collaborators: Dr. Evgeniy Petrotchenko and Dr. Christoph Borchers at the University of Victoria – Genome British Columbia Protein Centre (Victoria, British Columbia, Canada).

Chapter 6: Curcumin Analogues Analysis

The curcumin analogues tested (NC 2311, NC 2453, NC 2144, and NC 2094, structures shown below in Table 2.2) were synthesized by collaborator Dr. Jonathan Dimmock and his research lab.³ 2,5-Dihydroxybenzoic acid (DHB) was purchased from Lancaster Synthesis (Pelham, NH), 3-hydroxy-4-nitrobenzoic acid (3H4NBA) was purchased from Fluka Analytical (St. Gallen, Switzerland), and 3,5-di-tert-butyl-4-hydroxytoluene (butylated hydroxytoluene, BHT) was purchased from Supelco (Bellefonte, PA). Cesium iodide was purchased from Sigma (St. Louis, MO). Water, acetonitrile, and trifluoroacetic acid were purchased from Fisher Scientific (Fair Lawn, NJ).

Table 2.2 Sample numbers, monoisotopic masses, and structures of the curcumin analogues analyzed.

<u>Sample Number</u>	<u>Monoisotopic Mass</u>	<u>Structure</u>
NC 2311	411.1599 Da	
NC 2453	335.1521 Da	
NC 2144	409.1678 Da	
NC 2094	465.2304 Da	
NC2138	409.1678	

REFERENCES

1. Francl, T. J.; Sherman, M. G.; Hunter, R. L.; Locke, M. J.; Bowers, W. D.; McIver Jr, R. T., Experimental determination of the effects of space charge on ion cyclotron resonance frequencies. *Int. J. Mass Spectrom. Ion Processes* **1983**, *54* (1), 189-199.
2. Wong, R. L.; Amster, I. J., Sub part-per-million mass accuracy by using stepwise-external calibration in Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **2006**, *17* (12), 1681-1691.
3. Das, S.; Das, U.; Selvakumar, P.; Sharma, R. K.; Balzarini, J.; De Clercq, E.; Molnár, J.; Serly, J.; Baráth, Z.; Schatte, G., 3, 5-Bis (benzylidene)-4-oxo-1-phosphonopiperidines and Related Diethyl Esters: Potent Cytotoxins with Multi-Drug-Resistance Reverting Properties. *ChemMedChem* **2009**, *4* (11), 1831-1840.

CHAPTER 3

SHOTGUN PROTEOMIC ANALYSIS AND ASSIGNMENT OF *ACINETOBACTER BAYLYI* ADP1 VIA ACCURATE MASS MEASUREMENT

Accurate mass measurement as a means of protein identification is based on the ability to correlate measured peptide mass to those generated by an *in-silico* digestion of all possible proteins that can be expressed from the open reading frames of the genome.^{1,2,3} Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) is well suited for this analysis due to its ability to collect high mass accuracy and high resolution mass spectra.⁴ By using external calibration, mass accuracy of only a few parts per million (ppm) can readily be achieved,^{5,6} and when coupled with internal calibration, sub ppm can be reached.⁷ With high enough mass accuracy, peptide assignments can be correlated uniquely to the originating protein.^{4,8}

Accurate mass measurement as a means of peptide assignment has several advantages over the traditional technique of tandem MS for protein identification.^{8,9} Tandem MS is limited by several characteristics that are due to the nature of the technique.¹⁰ This technique can be low throughput, as the analysis generates fragmentation spectra for each peptide.¹¹ When a proteome digest can contain many thousands of components, it is not practical to fragment each peptide, which leads to another limitation of tandem MS, which is undersampling peaks for analysis.¹² Since a limited number of peptides can be chosen for fragmentation during data-dependent

tandem experiments, only a fraction of the observed peaks are selected from each precursor scan. Peak selection tends to favor higher intensity and higher charge state peptides, biasing the analysis to highly abundant proteins, or highly abundant protein isoforms. Tandem MS also leads to very large datasets, as each peak can be fragmented to generate a sequence spectrum. These spectra require significant computational resources to process and analyze.¹⁰

In order to unambiguously identify a protein from the accurate mass measurement of its peptides, typically three to six peptides are required.¹³ In contrast, the sequence obtained from a tandem MS fragmentation spectrum of a single peptide can be used to make the protein identification.^{13,1} To improve the success of identifications from accurate mass measurement experiments, the ability to observe peptides that are unique to each protein is beneficial.¹⁴ Whether a peptide can be used to uniquely identify a protein is based on the characteristics of the peptide, such as mass, based on those from the rest of the proteome.¹⁵

It is possible to have two or more peptides with the same measured mass, within even a small mass window, but originate from different proteins.¹⁶ These peptides cannot be used to make a definitive identification, if mass is the only searching constraint used.¹⁷ There are many other peptide characteristics that can be used as additional searching constraints, such as retention time or isoelectric point, both of which are derived from separation techniques, and presence of specific amino acids, typically observed via tagging technologies.^{18,17,13} The additional searching constraints improve the likelihood of protein identification, by increasing the number of peptides that can be uniquely assigned to a protein.^{19,20}

For proteomes that are smaller and with minimal post-translational modifications, such as archaea and some other prokaryotes, SILAC experiments can be performed that take advantage of differential metabolic isotopic labelling.²¹ By having a sample composed of peptides containing only light or heavy isotopes, mass difference between these peptides can be indicative of elemental composition information, such as nitrogen stoichiometry.²²

Nitrogen stoichiometry was calculated by finding peak pairs created during the endogenous nitrogen labeling of the proteins within the organism.^{23,24} The samples consist of a mixture of ^{14}N and ^{15}N labeled proteins, created by growing the cultures with the sole nitrogen source being isotopically controlled. The extracted proteins from each of these cell populations were proteolytically digested by trypsin, creating a mixture consisting of both the light and heavy tryptic peptides. Upon mass analysis, the spectra consisted of pairs of peaks, corresponding to the labeled peptides. The mass difference between the monoisotopic peaks for each peak pair was measured. This number was then divided by the mass difference of ^{14}N and ^{15}N (0.997 amu), providing the number of nitrogens found in the peptide.²⁵ By using both the accurate measured mass and the nitrogen stoichiometry as searching constraints during the database query, the number of peptides that can uniquely assigned to one protein increases.²⁴

Accurate mass measurement of proteolytic peptides using FT-ICR has been used to quickly and accurately determine the proteins present in *Methanococcus maripaludis*.²⁴ Previous research has shown that HPLC-MALDI-FT-ICR is able to assign a large number of peptides by accurate mass measurement and use of nitrogen stoichiometry.²³ The proteome *A. baylyi* ADP1 has 3325 ORFs, roughly twice as large as that of *M.*

maripaludis.^{26,27} This experiment will test the robustness of our accurate mass measurement shotgun proteomic analysis approach on a more complex proteome.

Acinetobacter are non-motile, gram-negative gamma-proteobacteria, and are capable of using a very diverse range of compounds as sources of carbon and energy.²⁸ Most *Acinetobacter* species are considered to be nonpathogenic to healthy individuals, but there are several species that persist in hospital environments, causing severe infections for immunologically compromised patients.²⁹ The efficient means of horizontal gene transfer increases the pathogenic potential of *Acinetobacter*.³⁰ A possible mechanism for this has been observed in *Acinetobacter baylyi*, a species that is found in soil.^{31,32}

EXPERIMENTAL

The proteome sample was obtained from an *Acinetobacter baylyi* ADP1 whole cell lysate, grown with ammonium sulfate as the sole source of nitrogen. The cells were grown using ammonium sulfate with naturally occurring isotopic composition (99.6% ¹⁴N) and with 99% ¹⁵N-enrichment. The cells were collected by centrifugation at 10,000 *g* for 10 minutes, resulting in a cell pellet, which was resuspended in 20mM ammonium bicarbonate buffer. The cells were lysed using sonication. Protein lysis was prevented by the addition of phenylmethanesulfonyl fluoride (PMSF), and DNase was used to digest the DNA. The samples were then centrifuged to remove cellular debris. The final protein concentration was established using a bicinchoninic acid protein assay kit, and measuring the absorbance at 562 nm.

Protein extracts were then mixed in equal concentrations prior to trypsinolysis. The peptide mixtures were filtered using a 3,000 MWCO centrifugal filter to remove the small molecules added during the sample preparation. The resulting digest was analyzed using HPLC-MALDI-FT-ICR/MS.

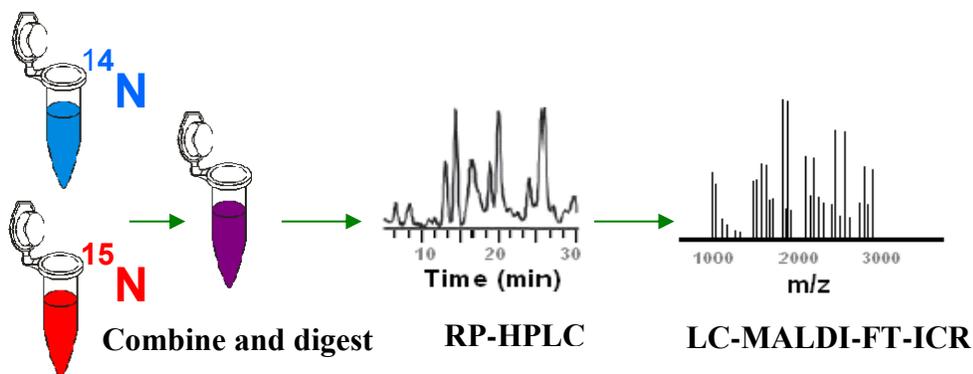


Figure 3.1 Experimental workflow used for the shotgun proteome analysis. Protein extracts from *A. baylyi* grown in natural abundance nitrogen growth medium and in 99% ^{15}N enriched growth medium are combined 1:1 (equal concentrations). The mixture is digested with trypsin, separated on a reverse phase C18 column, and spotted onto a MALDI target for mass analysis.

The HPLC analysis was performed using the UltimatePlus system equipped with a $75\mu\text{m}$ i.d. x 15 cm C18 column with $3\mu\text{m}$ particles and 300\AA pore size. The elution gradient used mobile phase A (99/1/0.1) and B (20/80/0.1), water/acetonitrile/trifluoroacetic acid (by volume), at a flow rate of $300\text{nL}/\text{min}$ for 130 minutes. Mobile phase B was started at 0% and was ramped to 15% by 85 minutes and to 100% by 110 minutes; it was held at 100% for 10 minutes. As the fractions eluted, they were collected onto stainless steel MALDI target plates. The fractions were spotted every 90s using the Probot Micro Fraction Collector. The samples then air dried and

were manually spotted with 0.5 μ L of DHB matrix solution, prepared by dissolving 10 mg of DHB in 50 μ L of 50/50/0.1 water/acetonitrile/trifluoroacetic acid (by volume).



Figure 3.2 HPLC Separation of the proteome digest of *Acinetobacter baylyi* ADP1. The mobile phase B gradient is shown as a dotted line.

The fractions were analyzed using a 7T Bio-Apex Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer. Ionization was performed with an intermediate pressure Scout 100 source, using a 337 nm pulsed nitrogen laser. Calibration was performed using a combination of external and step-wise external calibration. External calibration was performed using a mixture of tryptic peptides from ovalbumin, bovine serum albumin, and insulin chain B. Stepwise external calibration was performed as well, using a low cell trapping potential of 0.6V and a high cell trapping potential of 1.10V.³³

The peptides were assigned by comparing the experimental measured molecular mass and nitrogen stoichiometry to the theoretical values derived from the *A. baylyi* database built in-house. The *A. baylyi* database was generated by translating the genome into the open reading frames (proteins). These proteins were cleaved *in-silico* to create a file containing all tryptic peptides. The elemental composition was determined from the

amino acid sequence of each peptide, so that nitrogen stoichiometry could be used as a searching constraint.

RESULTS AND DISCUSSION

Results

From the *in-silico* tryptic digestions of the *A. baylyi* ADP1 sequence database yielded 143,595 peptides with up to one missed cleavage, and with a mass of 700 Da or higher. The statistical analysis of the *A. baylyi* ADP1 tryptic peptide table indicated that 35% of the proteolytic peptides can be assigned at a search tolerance of 10 ppm. Proteins are assigned by the type of peptides that are observed: non-unique peptides and unique peptides. Nonunique peptides are those that can be assigned to multiple proteins within a specific mass error window. Unique peptides can only be attributed to a single protein. As the number of peptides assigned to a protein increase, the confidence that the protein is being expressed in the proteome sample also increases. Proteins are only considered present in this analysis if there is at least one unique peptide assigned from them.

Mass spectra containing peptides were collected for 45 of the fractions collected from the HPLC. Each fraction was analyzed by MALDI-FT-ICR to determine the monoisotopic mass and the number of nitrogens in each peptide. A summary of peptide and protein assignments is included as Table 3.1. 138 proteins were identified at 5 ppm mass tolerance and 112 proteins were identified at 10 ppm mass tolerance. These proteins observed at 5 ppm are shown in the proteome map in Figure 3.3. In the protein map, black boxes represent proteins for which no peptide could be attributed, green/

yellow boxes indicate proteins were observed by the presence of non-unique peptides, and red/ purple boxes indicate the protein was observed by one or more unique peptides.

Table 3.1: Summary of peptide and protein assignments from the global digest of the *A. baylyi* ADP1 proteome.

	<u>5 ppm</u>	<u>10 ppm</u>
Total Peptides	1017	1017
Unique Peptides	217	163
Non-unique Peptides	364	556
Unassigned Peptides	436	298
Assigned Proteins	138	112

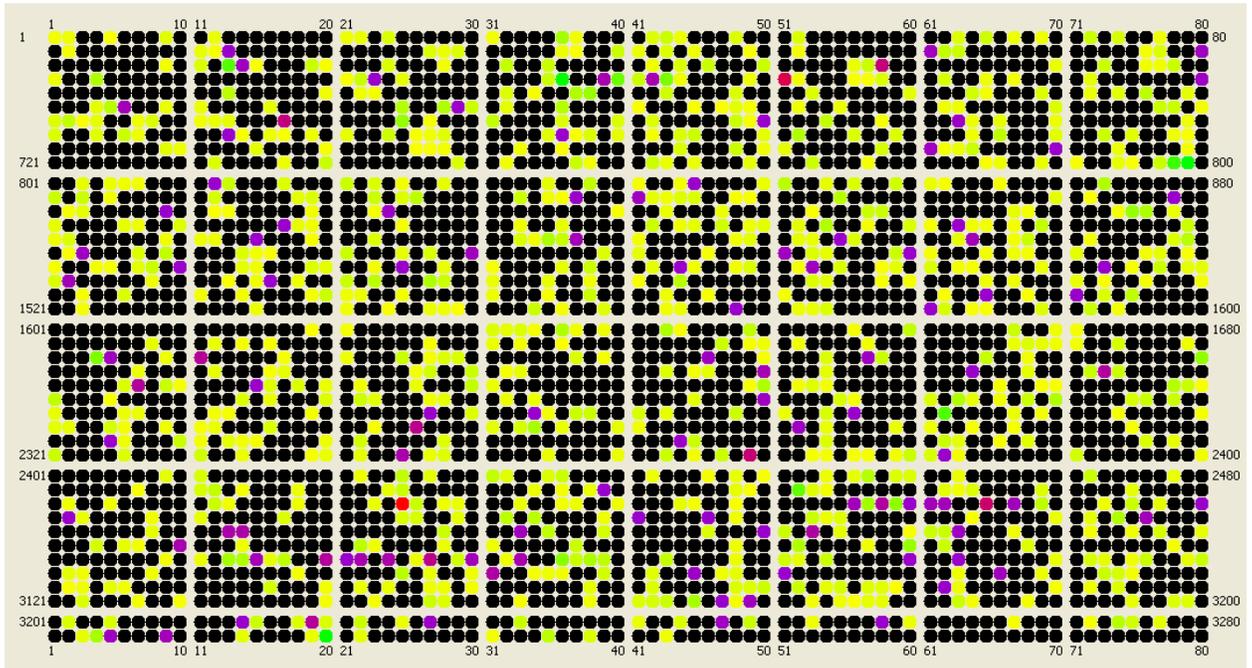


Figure 3.3 Results from global digest of *A. baylyi* ADP1 labeled by endogenous ^{15}N labeling. These assignments are based on the search performed at 5ppm. Each box represents an open reading frame that should code for a protein.

Conclusions

Although it was theoretically possible to assign 50,258 peptides using a search tolerance of 10 ppm, significantly fewer than this many peptides were even observed in the experiment. One reason is due to the fact that not all of the genes are turned on at any given time, therefore the proteins encoded from the genes would not exist. There are several other factors that led to this diminished peptide yield, including low peptide abundance in the fractions, protein selection biases, and database construction.

The minimum absorbance needed for peptides to be observed in the mass spectrometer is typically around 80 mAU. Enough signal is needed from each pair of labeled peaks to ensure that the monoisotopic peak can be confidently be assigned a mass, which is used to determine the nitrogen stoichiometry. If both monoisotopic peaks from the labeled pair are not correctly chosen, the peptides will not be found in the database. In the chromatographic separation seen in Figure 3.2, only a small fraction of the separated peptides are observed above 80 mAU. This seemed to be the case for all *A. baylyi* ADP1 samples that were tested during the course of the experiment.

During the digestion of the proteome sample, protein precipitation was observed. This was not unexpected as the digestion of the previous proteomes (*Methanococcus maripaludis*) always exhibited some precipitation at the beginning of the digestion, but there was no residual protein left after the digestion. This was not the case at the end of the *A. baylyi* ADP1 digestions. There were large amounts of undigested protein left, indicating that the digestion protocol needed to be modified.

Changes were made to the standard digestion protocol to try to increase the number of peptides observed in the UV trace, as well as in the mass spectra. Increased

amounts of trypsin, longer digestion times, incorporation of organic solvents into the reduction and denaturation stages, and alkylation of the sample were all tested but with no significant change to the number of assigned peptides and proteins. There was also minimal change to the amount of undigested protein left at the end of the digestion.

This undigested protein is likely due to the inability to keep very large proteins in solution, especially at the concentrations needed for the proteome analysis, which is typically close to 1.5 mg/mL of total protein. Large proteins and membrane proteins require different solution conditions to remain soluble. If the proteins are not in solution, they interact with the trypsin minimally, decreasing the abundance of peptides created and observed.

To remove the large fraction of proteins from the sample for later analysis, a 30K MWCO filter was used. Although there was no undigested protein observed in the digests that were filter, there was also no increase in proteome coverage. One cause for this may be that the large proteins when filtered out, smaller proteins become entangled within them. Even with thorough rinsing, the smaller proteins are unable to escape and are lost from the analysis. There is also the problem that proteins close to 30kDa will become stuck inside of the pores in the filter.

There are also problems encountered during sonication and protein extraction from the cells. When the cells are lysed, the cell membranes are broken, and the membrane proteins that are embedded into the cell membranes are centrifuged into a pellet that is discarded. Membrane proteins can account for a large portion of the encoded proteins in a genome. It is believed that up to 30% of all genes are responsible for encoding membrane proteins in most genomes.³⁴ Previous analysis of *M. maripaludis*

did not suffer as greatly from difficulties in cell lysis, as the cells easily became hypotonic in the presence of pure water, and underwent cell lysis spontaneously.

The last source of reduced peptide and protein assignment is likely due to an insufficient database. The database included the tryptic digest population including up to one missed enzymatic cleavage. There was not inclusion of post-translation modifications (PTMs), as this would increase the database size exponentially. By including many of the most common PTMs, the number of increased assignments would dwarf the number of those lost due to decreased confidence from the larger database population. Any change to the database that decreases the likelihood of unique protein assignments weakens this technique.

Although this technique has excelled in the assignment of proteins, under qualitative and quantitative conditions, in past experiments using *Methanococcus maripaludis*, the limitations reduced the effectiveness in the analysis of the *Acinetobacter baylyi* ADP1 proteome. As the number of open reading frames in a genome increases, the ability of using measured accurate mass as a primary source of protein information decreases dramatically. Including the nitrogen stoichiometry does help increase the confidence in assignments, as well as increase the number of unique peptides that can be used to assign proteins in the mixture.

Improving the mass accuracy of the experiments would increase identification numbers, but not without significant improvement, such as sub ppm mass accuracy on experimental samples. More information regarding the post-translation modification would also improve the database entries, increasing identification and assignments. Even with the difficulties experienced during the sample analysis, 138 proteins were identified

at 5 ppm mass tolerance and 112 proteins were identified at 10 ppm mass tolerance, when using accurate mass measurement and nitrogen stoichiometry were used.

REFERENCES

1. Yates, J. R., Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **1998**, *33* (1), 1-19.
2. Washburn, M. P.; Wolters, D.; Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19* (3), 242-247.
3. Wolters, D. A.; Washburn, M. P.; Yates, J. R., An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001**, *73* (23), 5683-5690.
4. Clauser, K. R.; Baker, P.; Burlingame, A. L., Role of accurate mass measurement (± 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* **1999**, *71* (14), 2871-2882.
5. Easterling, M. L.; Mize, T. H.; Amster, I. J., Routine part-per-million mass accuracy for high-mass ions: Space-charge effects in MALDI FT-ICR. *Anal. Chem.* **1999**, *71* (3), 624-632.
6. Bruce, J. E.; Anderson, G. A.; Brands, M. D.; Pasa-Tolic, L.; Smith, R. D., Obtaining more accurate Fourier transform ion cyclotron resonance mass measurements without internal standards using multiply charged ions. *J. Am. Soc. Mass. Spectrom.* **2000**, *11* (5), 416-421.
7. Burton, R. D.; Matuszak, K. P.; Watson, C. H.; Eyler, J. R., Exact mass measurements using a 7 tesla fourier transform ion cyclotron resonance mass spectrometer in a good laboratory practices-regulated environment. *J. Am. Soc. Mass. Spectrom.* **1999**, *10* (12), 1291-1297.
8. He, F.; Emmett, M. R.; Håkansson, K.; Hendrickson, C. L.; Marshall, A. G., Theoretical and experimental prospects for protein identification based solely on accurate mass measurement. *J. Proteome Res.* **2004**, *3* (1), 61-67.
9. Haas, W.; Faherty, B. K.; Gerber, S. A.; Elias, J. E.; Beausoleil, S. A.; Bakalarski, C. E.; Li, X.; Villén, J.; Gygi, S. P., Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol. Cell. Proteomics* **2006**, *5* (7), 1326-1337.
10. Tabb, D. L.; MacCoss, M. J.; Wu, C. C.; Anderson, S. D.; Yates, J. R., Similarity among tandem mass spectra from proteomic experiments: detection, significance, and utility. *Anal. Chem.* **2003**, *75* (10), 2470-2477.
11. Dowell, J. A.; Frost, D. C.; Zhang, J.; Li, L., Comparison of two-dimensional fractionation techniques for shotgun proteomics. *Anal. Chem.* **2008**, *80* (17), 6715-6723.

12. Du, P.; Kibbe, W. A.; Lin, S. M., Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* **2006**, *22* (17), 2059-2065.
13. Conrads, T. P.; Anderson, G. A.; Veenstra, T. D.; Paša-Tolic, L.; Smith, R. D., Utility of accurate mass tags for proteome-wide protein identification. *Anal. Chem.* **2000**, *72* (14), 3349-3354.
14. Higgs, R. E.; Knierman, M. D.; Bonner Freeman, A.; Gelbert, L. M.; Patil, S. T.; Hale, J. E., Estimating the statistical significance of peptide identifications from shotgun proteomics experiments. *J. Proteome Res.* **2007**, *6* (5), 1758-1767.
15. Nesvizhskii, A. I.; Aebersold, R., Interpretation of shotgun proteomic data the protein inference problem. *Mol. Cell. Proteomics* **2005**, *4* (10), 1419-1440.
16. Mann, M.; Kelleher, N. L., Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci.* **2008**, *105* (47), 18132-18138.
17. Goodlett, D. R.; Bruce, J. E.; Anderson, G. A.; Rist, B.; Pasa-Tolic, L.; Fiehn, O.; Smith, R. D.; Aebersold, R., Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. *Anal. Chem.* **2000**, *72* (6), 1112-1118.
18. Hernandez, H.; Niehauser, S.; Boltz, S. A.; Gawandi, V.; Phillips, R. S.; Amster, I. J., Mass defect labeling of cysteine for improving peptide assignment in shotgun proteomic analyses. *Anal. Chem.* **2006**, *78* (10), 3417-3423.
19. Nesvizhskii, A. I., A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J. Proteomics* **2010**, *73* (11), 2092-2123.
20. Baczek, T., Improvement of peptides identification in proteomics with the use of new analytical and bioinformatic strategies. *Current Pharm. Anal.* **2005**, *1* (1), 31-40.
21. Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, *1* (5), 376-386.
22. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003**, *422* (6928), 198-207.
23. Jing, L.; Li, C.; Wong, R. L.; Parks, B. A.; Liu, Y.; Whitman, W. B. New Techniques for Proteomic Analysis with FTICR-MS: Instrumentation, Automation, and Application. Doctorate, University of Georgia, 2009.

24. Jing, L.; Amster, I. J., Rapid and automated processing of MALDI-FTICR/MS data for ¹⁵N-metabolic labeling in a shotgun proteomics analysis. *International journal of mass spectrometry* **2009**, *287* (1), 27-31.
25. Jing, L.; Amster, I., An improved calibration method for the matrix-assisted laser desorption/ionization-Fourier transform ion cyclotron resonance analysis of ¹⁵N-metabolically-labeled proteome digests using a mass difference approach. *Eur. J. Mass Spectrom.* **2012**, *18* (3), 269.
26. Iacono, M.; Villa, L.; Fortini, D.; Bordoni, R.; Imperi, F.; Bonnal, R. J.; Sicheritz-Ponten, T.; De Bellis, G.; Visca, P.; Cassone, A., Whole-genome pyrosequencing of an epidemic multidrug-resistant *Acinetobacter baumannii* strain belonging to the European clone II group. *Antimicrob. Agents Chemother.* **2008**, *52* (7), 2616-2625.
27. Hendrickson, E.; Kaul, R.; Zhou, Y.; Bovee, D.; Chapman, P.; Chung, J.; de Macario, E. C.; Dodsworth, J.; Gillett, W.; Graham, D., Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*. *J. Bacteriol.* **2004**, *186* (20), 6956-6969.
28. de Berardinis, V.; Vallenet, D.; Castelli, V.; Besnard, M.; Pinet, A.; Cruaud, C.; Samair, S.; Lechaplais, C.; Gyapay, G.; Richez, C., A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* **2008**, *4* (1).
29. Chen, T.-L.; Siu, L.-K.; Lee, Y.-T.; Chen, C.-P.; Huang, L.-Y.; Wu, R. C.-C.; Cho, W.-L.; Fung, C.-P., *Acinetobacter baylyi* as a pathogen for opportunistic infection. *J. Clin. Microbiol.* **2008**, *46* (9), 2938-2944.
30. Valenzuela, J. K.; Thomas, L.; Partridge, S. R.; van der Reijden, T.; Dijkshoorn, L.; Iredell, J., Horizontal gene transfer in a polyclonal outbreak of carbapenem-resistant *Acinetobacter baumannii*. *J. Clin. Microbiol.* **2007**, *45* (2), 453-460.
31. Nielsen, K.; Gebhard, F.; Smalla, K.; Bones, A.; Van Elsas, J., Evaluation of possible horizontal gene transfer from transgenic plants to the soil bacterium *Acinetobacter calcoaceticus* BD413. *Theor. Appl. Genet.* **1997**, *95* (5-6), 815-821.
32. Nielsen, K. M.; van Elsas, J. D.; Smalla, K., Transformation of *Acinetobacter* sp. strain BD413 (pFG4ΔnptII) with transgenic plant DNA in soil microcosms and effects of kanamycin on selection of transformants. *Appl. Environ. Microbiol.* **2000**, *66* (3), 1237-1242.
33. Wong, R. L.; Amster, I. J., Sub part-per-million mass accuracy by using stepwise-external calibration in Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **2006**, *17* (12), 1681-1691.

34. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L., Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **2001**, *305* (3), 567-580.

CHAPTER 4

METABOLIC PATHWAY OF DIMETHYLSULPHONIOPROPIONATE IN *RUEGERIA POMEROYI* DSS-3

Dimethylsulphoniopropionate (DMSP) is a marine phytoplankton metabolite that is present ubiquitously in marine surface waters. It accounts for up to 10% of the carbon fixed by marine phytoplankton in ocean surface waters^{1,2}, and one of the most abundant low molecular weight sources of carbon and sulfur in these ecosystems.^{3,4} To become a source of sulfur, the DMSP is processed by the marine bacteria via the demethylation/demethiolation pathway⁵, which releases methanethiol (MeSH) for incorporation, or by the cleavage pathway to produce the climatically active gas dimethylsulfide (DMS).⁶ DMS is a climatically active gas emitted by phytoplankton that can induce the formation of aerosols, which can act as cloud condensation nuclei and promote of solar backscatter.⁷

Although the existence of the differing pathways had been known, the biochemical pathway and the enzymes responsible for the demethylation/demethiolation pathway remained unknown. To elucidate the metabolic pathway for the production of (DMSP), *Ruegeria pomeroyi* DSS-3, a marine roseobacter, was studied.^{8,9} There are four genes that have been identified to catalyze the release of DMS^{10,11,12,13}, and one gene (*dmdA*) that produces the initial demethylase in the MeSH pathway.¹⁴ The initial demethylation produces methylmercaptopropionate (MMPA), which is further

catabolized into MeSH.¹⁵ The catabolism occurs via a series of Coenzyme A (CoA) mediated reactions, comparable to fatty acid β -oxidation.

EXPERIMENTAL

In order to determine the metabolic pathway, acyl-CoA intermediates were studied using various analytical techniques, including high performance liquid chromatography (HPLC), nuclear magnetic resonance (NMR), and Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS). The acyl-CoA mixtures were extracted from the *R. pomeroyi* DSS-3 cell cultures by lysing the cells with trichloroacetic acid. The cellular debris was removed by centrifuging the samples, and the supernate was then passed through an oligo purification cartridge. The cartridge was rinsed to eliminate nonbonded constituents, and the retained acyl-CoA products were analyzed via HPLC. Analysis of the enzymes and genes responsible for the metabolism were also studied. These required the cells be lysed using a French press, and the samples were then centrifuged. With the cellular debris removed, the samples could then be studied for enzyme activity.

The nuclear magnetic resonance analysis was performed using a Varian Unity Inova 500 MHz spectrometer. For the ¹H NMR analysis, 1,000 scans were collected, with a relaxation delay of 2s. The ¹³C NMR analysis was performed at 125 MHz, where 2,000 scans were collected with a relaxation delay of 5s. The mass spectrometric analysis was performed using a 7 tesla BioApex Fourier transform ion cyclotron resonance mass spectrometer. Samples were analyzed using 2,5-dihydroxybenzoic acid as the matrix.

RESULTS

In order to elucidate the DMSP demethylation pathway, the Coenzyme A intermediates were investigated. A mutant was created that was deficient in the gene *dmdA*, making it incapable of DMSP demethylation. It was then supplied with DMSP. The wild type was given MMPA. The CoA was extracted and separated using HPLC, as shown in Figure 4.1.

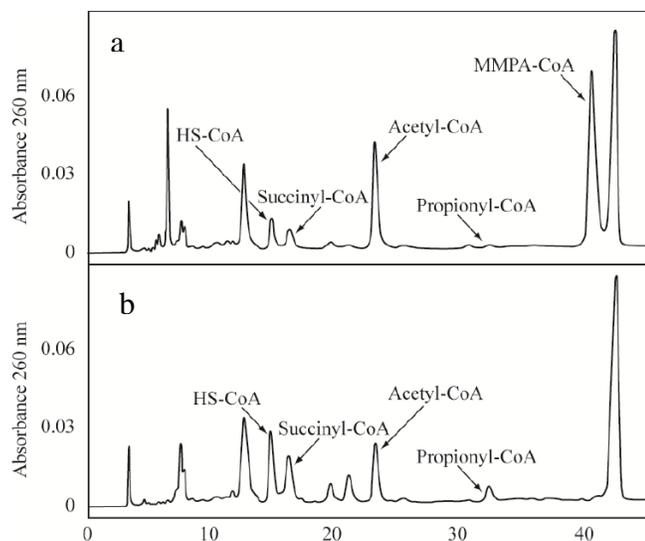


Figure 4.1 HPLC separations of the extracted CoA mixtures. (a) Wild type *R. pomeroyi* DSS-3 extraction separation. The intermediates observed are attributed to the demethylation pathway. (b) *DmdA*⁻ mutant *R. pomeroyi* DSS-3 extraction separation. The observed intermediates are attributed to the cleavage pathway. The shared intermediates are from downstream metabolic pathways. The peaks assigned based on the co-elution of standards are indicated. Figure adapted from Reisch 2011.⁸

An unknown CoA intermediate was observed in the sample supplied with MMPA. The fraction containing the unknown intermediate was collected and was analyzed via matrix assisted laser desorption ionization Fourier transform ion cyclotron resonance mass spectrometry (MALDI-FT-ICR-MS), shown in Figure 4.2. The unknown intermediate was observed at m/z 870.137. This mass was consistent with the theoretical monoisotopic mass of a CoA-MMPA thioester (MMPA-CoA). A peak at m/z 768.122 was observed and assigned as CoA-SH, produced by the hydrolysis of the thioester bond, as it was found in all of the analyzed CoA samples.

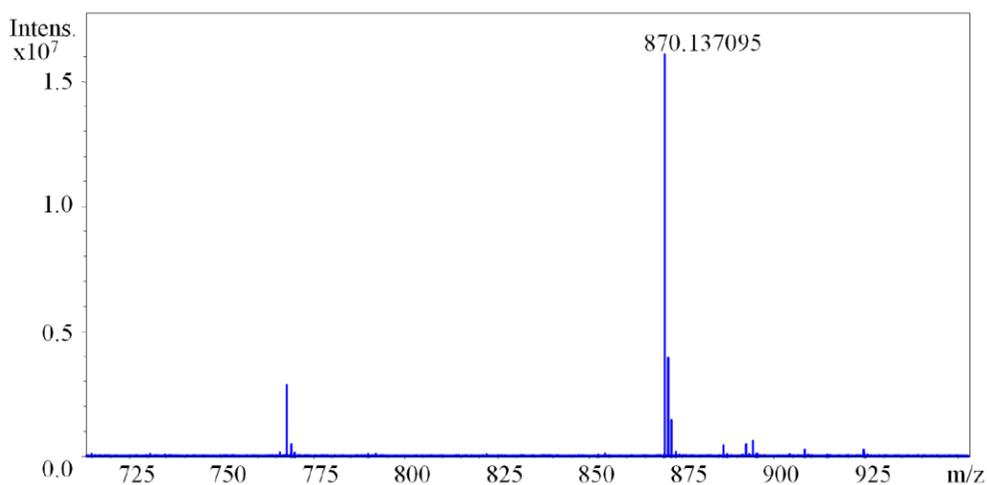


Figure 4.2 MALDI FT-ICR mass spectrum of MMPA-CoA (m/z 870.137) extracted from the whole cells after incubation of the cells with MMPA. The peak observed at m/z 768 is due to the spontaneous hydrolysis of the thioester bond to produce HS-CoA. Figure adapted from Reisch 2011.⁸

MMPA-CoA was then synthesized and analyzed. It was observed to have the same retention time and molecular weight as the unknown intermediate. Next, the mutant cells were incubated in [1,2,3-¹³C] DMSP as their carbon source, and the fractions analyzed via MALDI-FT-ICR-MS, shown in Figure 4.3. The MMPA-CoA peak at m/z 870.137 was still observed, along with an additional peak at m/z 873.148. The mass

difference between the peaks is 3.011 Da, which corresponds very closely to the mass of three ^{13}C atoms. The assignment of this peak based on accurate mass measurement has a mass error of 1.09 ppm.

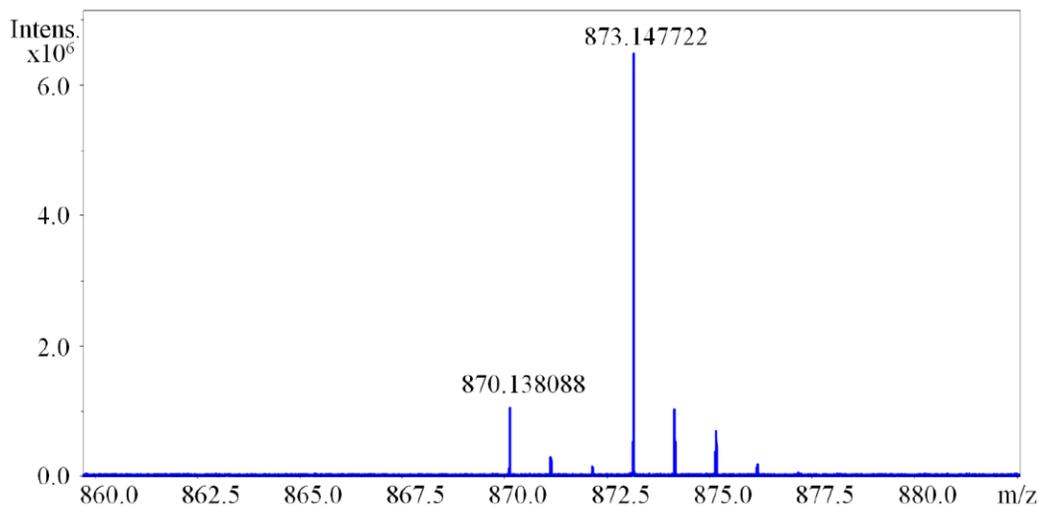


Figure 4.3 MALDI-FT-ICR mass spectrum of MMPA-CoA extracted from whole cells incubated with [1,2,3- ^{13}C] DMSP. Peaks are observed for MMPA-CoA (m/z 870.138) and ^{13}C labeled MMPA-CoA (m/z 873.148). Figure adapted from Reisch 2011.⁸

With the identity of the first intermediate assigned, the enzyme catalyzing the formation of the MMPA-CoA needed to be determined. The native enzyme was purified from *R. pomeroyi* DSS-3 cell extracts, and run on a SDS-PAGE. Four proteins were observed during the silver staining. MALDI-TOF analysis identified one of the observed proteins as a medium chain fatty-acid CoA ligase. The protein was designated as 3-methylmercaptopropionyl-CoA ligase (DmdB), and determined to be encoded by gene SPO2045. Confirmation of the gene-protein assignment was performed by creating a recombinant *Escherichia coli* strain containing a cloned in SPO2045, and testing the MMPA-CoA ligase activity.

The next step of the pathway needed to be investigated. By incubating crude cell extracts with MMPA-CoA and various redox factors, the MMPA-CoA-consuming

activity could be probed. When phenazine methosulfate or ferrocenium hexafluorophosphate was used, MMPA-CoA was consumed, and an unknown intermediate was produced. The intermediate was analyzed via MALDI-FT-ICR-MS, and m/z 868.121 was observed, as shown in Figure 4.4. This mass is consistent with the theoretical monoisotopic mass of methylthioacryloyl-CoA (MTA-CoA).

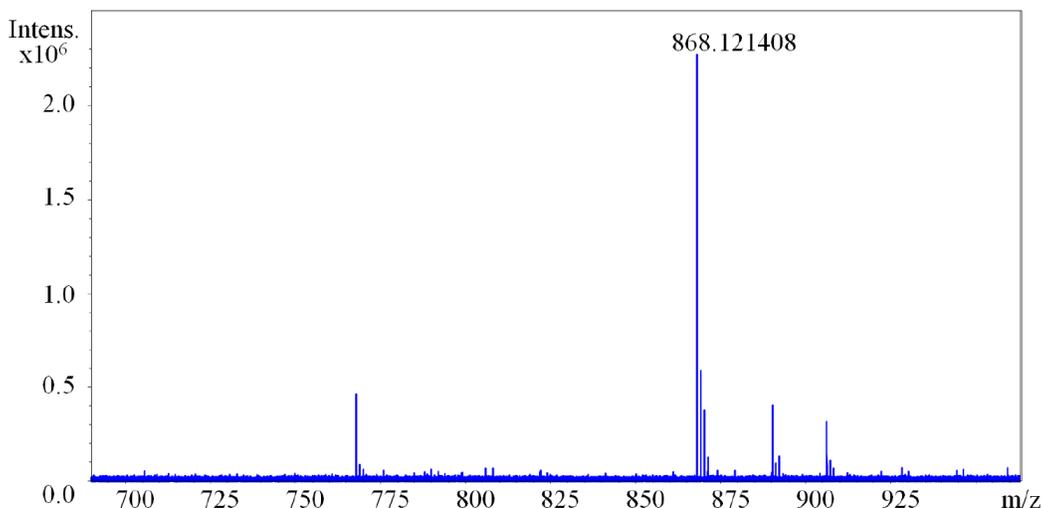


Figure 4.4 MALDI-FT-ICR mass spectrum of MTA-CoA extracted from cell-free extracts with MMPA-CoA and ferrocenium hexafluorophosphate. The peak observed at m/z 768 is due to the spontaneous hydrolysis of the thioester bond to produce HS-CoA. Figure adapted from Reisch 2011.⁸

The addition of MTA-CoA to a *R. pomeroyi* DSS-3 crude cell extract yielded MeSH and free CoA. The enzyme catalyzing this reaction was purified and identified as SPO3805, an enoyl-CoA hydratase. This gene (SPO3805) and the one upstream from it (SPO3804) were cloned and expressed in *E. coli*. These strains exhibited activity for both MMPA-CoA dehydrogenase (SPO3804) and MTA-CoA hydratase (SPO3805), and the genes are designated *dmdC* and *dmdD*, respectively.

The reaction catalyzed by DmdD, produced MeSH and CoA in stoichiometric amounts, and an unknown product, which was analyzed by ¹H and ¹³C NMR

spectroscopy, shown in Figure 4.5. It was determined that the unknown reaction product was acetaldehyde, leaving the thioester bonded carbon of the MTA-CoA unaccounted for. It was believed that this carbon is lost as CO_2 during the reaction, supported by Figure 4.6.

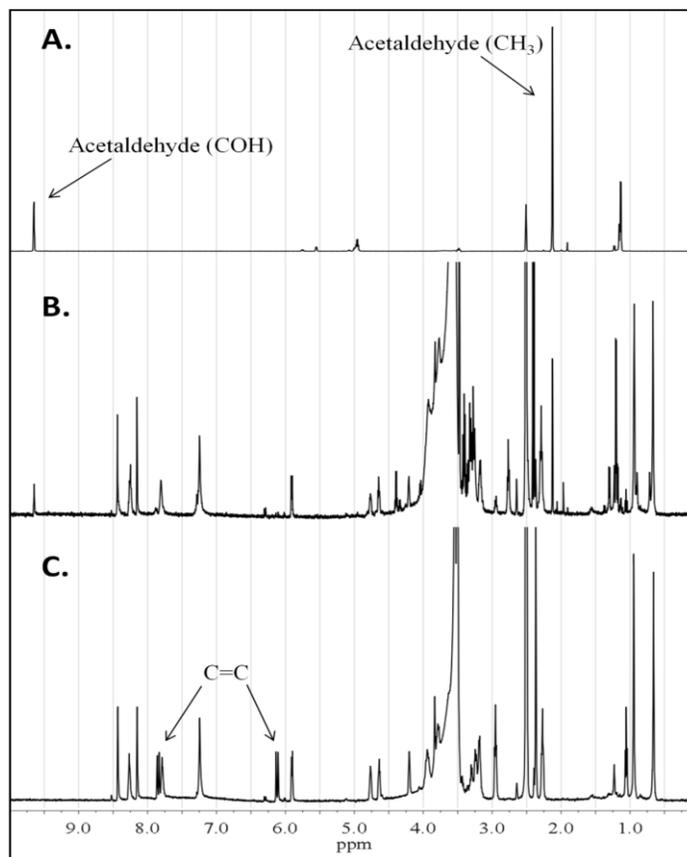


Figure 4.5 ^1H NMR spectrum of DmdD catalyzed reaction products. (a) Pure aqueous acetaldehyde. (b) MTA-CoA after the addition of purified DmdD. (c) MTA-CoA with no DmdD addition. The peaks assigned C=C (due to the double bonded carbons in MTA-CoA) disappear upon reaction with DmdD, and the signals attributed to acetaldehyde appear. All other peaks in (b) and (c) are due to Coenzyme A and other reaction mixture components. Figure adapted from Reisch 2011.⁸

A.

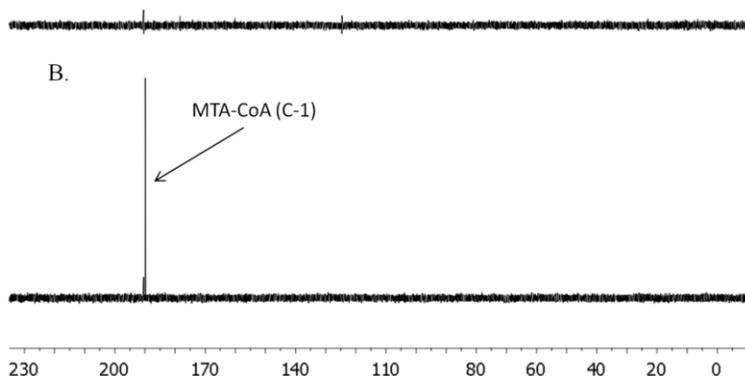


Figure 4.6 ¹³C NMR spectrum of [1-¹³C] MTA-CoA (a) before reaction with DmdD and (b) after reaction with DmdD. Upon reaction with DmdD, the enriched ¹³C is completely lost, indicating decarboxylation, though it is not known whether the decarboxylation is enzymatically catalyzed or due to loss of molecular stability during the reaction. Figure adapted from Reisch 2011.⁸

CONCLUSIONS

From the experiments discussed here, as well as in the papers of Reisch *et al.*^{8,9}, a metabolic pathway using DMSP as the starting point and ending with acetaldehyde was ascertained in *R. pomeroyi* DSS-3. The DMSP is demethylated to form MMPA by DmdA and tetrahydrofolate. MMPA then reacts with Coenzyme A, catalyzed by DmdB, to form the MMPA-CoA thioester. MTA-CoA is then formed by the dehydrogenation of MMPA-CoA, mediated by DmdC. Hydration of the MTA-CoA yields free CoA, MeSH1, and acetaldehyde, which can then be oxidized by alcohol dehydrogenase to form acetic acid. A summary of the demethylation pathway is shown in Figure 4.7.

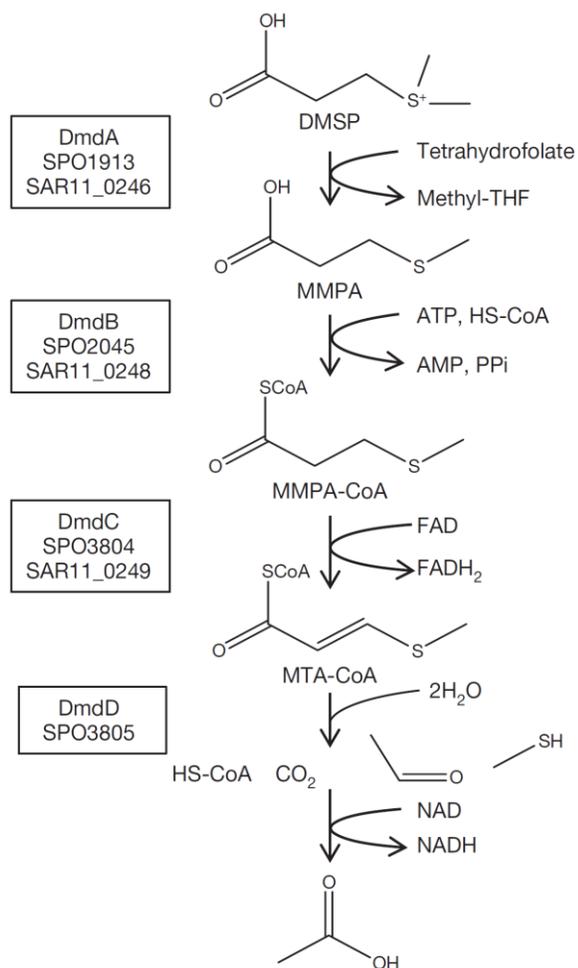


Figure 4.7 Pathway of DMSP demethylation in *R. pomeroyi* DSS-3. The genes responsible are identified in the boxes to the left of each step. The DMSP is demethylated forming MMPA, via a *DmdA* catalyzed tetrahydrofolate-dependent reaction.¹⁶ MMPA-CoA is then formed by *DmdB*, and is dehydrogenated to form MTA-CoA. Hydration of the MTA-CoA yields CO₂, free CoA, acetaldehyde, and MeSH. The acetaldehyde is oxidized to form acetic acid. Figure adapted from Reisch 2011.⁸

Although the pathway presented was elucidated in *R. pomeroyi* DSS-3, the genes that participate in it have many homologues found in diverse groups of bacteria and the initial source of MMPA would be inconsequential in these bacteria, as they would be able to utilize the MMPA-CoA pathway. Two ubiquitous families of marine bacteria, *Rhodobacteraceae* and *Pelagibacteraceae*, have been shown to possess the MMPA-CoA

pathway,⁸ and may account for up to 70% of the surface ocean bacterioplankton during the summer.¹⁷ The use of this pathway may be even more widespread throughout the ocean ecosystems, leading to a significant impact to climate, as it forms MeSH and acetate, while preventing the formation of DMS.

REFERENCES

1. Archer, S. D.; Widdicombe, C. E.; Tarran, G. A.; Rees, A. P.; Burkill, P. H., Production and turnover of particulate dimethylsulphoniopropionate during a coccolithophore bloom in the northern North Sea. *Aquat. Microb. Ecol.* **2001**, *24* (3), 225-241.
2. Simo, R.; Archer, S. D.; Gilpin, L.; Stelfox-Widdicombe, C. E., Coupled dynamics of dimethylsulfonylpropionate and dimethylsulfide cycling and the microbial food web in surface waters of the North Atlantic. *Limnol. Oceanogr.* **2002**, *47* (1), 53-61.
3. Curson, A. R.; Todd, J. D.; Sullivan, M. J.; Johnston, A. W., Catabolism of dimethylsulphonylpropionate: microorganisms, enzymes and genes. *Nature Reviews Microbiology* **2011**, *9* (12), 849-859.
4. Moran, M. A.; Reisch, C. R.; Kiene, R. P.; Whitman, W. B., Genomic insights into bacterial DMSP transformations. *Annual review of marine science* **2012**, *4*, 523-542.
5. Kiene, R. P.; Linn, L. J.; Bruton, J. A., New and important roles for DMSP in marine microbial communities. *J. Sea Res.* **2000**, *43* (3), 209-224.
6. Stefels, J.; Steinke, M.; Turner, S.; Malin, G.; Belviso, S., Environmental constraints on the production and removal of the climatically active gas dimethylsulphide (DMS) and implications for ecosystem modelling. *Biogeochemistry* **2007**, *83* (1-3), 245-275.
7. Charlson, R. J.; Lovelock, J. E.; Andreae, M. O.; Warren, S. G., Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature* **1987**, *326* (6114), 655-661.
8. Reisch, C. R.; Stoudemayer, M. J.; Varaljay, V. A.; Amster, I. J.; Moran, M. A.; Whitman, W. B., Novel pathway for assimilation of dimethylsulphonylpropionate widespread in marine bacteria. *Nature* **2011**, *473* (7346), 208-211.
9. Reisch, C. R.; Crabb, W. M.; Gifford, S. M.; Teng, Q.; Stoudemayer, M. J.; Moran, M. A.; Whitman, W. B., Metabolism of dimethylsulphonylpropionate by *Ruegeria pomeroyi* DSS-3. *Mol. Microbiol.* **2013**, *89* (4), 774-791.
10. Todd, J. D.; Rogers, R.; Li, Y. G.; Wexler, M.; Bond, P. L.; Sun, L.; Curson, A. R.; Malin, G.; Steinke, M.; Johnston, A. W., Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. *Science* **2007**, *315* (5812), 666-669.
11. Todd, J.; Curson, A.; Dupont, C.; Nicholson, P.; Johnston, A., The dddP gene, encoding a novel enzyme that converts dimethylsulfonylpropionate into dimethyl

sulfide, is widespread in ocean metagenomes and marine bacteria and also occurs in some Ascomycete fungi. *Environ. Microbiol.* **2009**, *11* (6), 1376-1385.

12. Todd, J. D.; Curson, A. R.; Kirkwood, M.; Sullivan, M. J.; Green, R. T.; Johnston, A. W., DddQ, a novel, cupin-containing, dimethylsulfoniopropionate lyase in marine roseobacters and in uncultured marine bacteria. *Environ. Microbiol.* **2011**, *13* (2), 427-438.
13. Curson, A.; Rogers, R.; Todd, J.; Brearley, C.; Johnston, A., Molecular genetic analysis of a dimethylsulfoniopropionate lyase that liberates the climate-changing gas dimethylsulfide in several marine α -proteobacteria and Rhodobacter sphaeroides. *Environ. Microbiol.* **2008**, *10* (3), 757-767.
14. Howard, E. C.; Sun, S.; Biers, E. J.; Moran, M. A., Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environ. Microbiol.* **2008**, *10* (9), 2397-2410.
15. Kiene, R. P., Production of methanethiol from dimethylsulfoniopropionate in marine surface waters. *Mar. Chem.* **1996**, *54* (1), 69-83.
16. Reisch, C. R.; Moran, M. A.; Whitman, W. B., Dimethylsulfoniopropionate-dependent demethylase (DmdA) from Pelagibacter ubique and Silicibacter pomeroyi. *J. Bacteriol.* **2008**, *190* (24), 8018-8024.
17. Morris, R. M.; Rappé, M. S.; Connon, S. A.; Vergin, K. L.; Siebold, W. A.; Carlson, C. A.; Giovannoni, S. J., SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **2002**, *420* (6917), 806-810.

CHAPTER 5

CHEMICAL CROSSLINKING USING A NOVEL MASS DEFECT LABELED CROSSLINKER

Chemical crosslinking paired with mass spectrometry can be a powerful tool in determining structural information of biomolecules, including peptides, proteins, oligonucleotides, and their interactions with each other and small molecules.^{1,2,3,4,5,6,7} Advances in mass spectrometry instrumentation⁸, novel crosslinker designs^{9,10,11,12,13,14}, and improved software tools^{15,16,17} are being used to overcome a few of the challenges in traditional chemical crosslinking proteomic experiments, such as product assignment and validation due to the high complexity of crosslinked samples.^{18,19,20,21,22}

We have approached these current challenges by the introduction of a mass defect label on our chemical crosslinker, as well as utilizing the relationship between naturally occurring mass defects in peptides with those produced by the presence of our crosslinker.²³ This addition allows for the easy identification of reaction products and allows for a more focused study of those peaks likely to yield structurally important information. We have also developed methods using high-field asymmetric ion mobility spectroscopy as a way to increase the detectability of lower abundance species of interest, many of which will include the crosslinked peptides.²⁴ The goals of this project are to develop methodologies utilizing a novel crosslinker to help overcome some of the

experimental limitations encountered during traditional chemical crosslinking experiments. The methodologies will include developing a crosslinker that uses mass defect technologies, investigating new instrumentation techniques, including the use of high field asymmetric ion mobility spectrometry (FAIMS)²⁴, and developing data analysis techniques to aid in the assignment of reaction products.

Mass defect is the difference between the exact mass of a compound and the nominal molecular weight calculated using integer atomic masses.²⁵ The mass defect in peptides is primarily influenced by the amount of hydrogen present, as shown in Table 5.1. This is due to the mass defects of the hydrogen not being cancelled out, like in the case of nitrogen and oxygen, as well as the sheer number of hydrogens typically found in a peptide.²⁶ A peptide of averagine at nominal mass m/z 1000 (exact mass 999.7136 amu) has 96% of its mass defect due to hydrogen.

Table 5.1 Mass defects of elements that comprise a peptide of averagine at m/z 1000.

<u>Element</u>	<u>Exact Mass</u>	<u>Nominal Mass</u>	<u>Mass Defect</u>	<u>Composition Averagine</u>	<u>MD Contribution</u>
C	12.0000	12	0.0000	44	0
H	1.0078	1	0.0078	95	0.74337
N	14.0031	14	0.0031	12	0.03684
O	15.9959	16	-0.0041	13	-0.05317

By plotting the nominal mass vs the exact mass, the relationship between these values can be determined, as shown in Figure 5.1. The range of mass defect values (MDVs) for unmodified peptides is determined and a comparison of expected and observed MDVs shows whether the observed MDVs are outside the natural variance of a

peptide of a specific mass to charge. Peptides outside of the accepted range are assigned as containing the mass defect label.²³

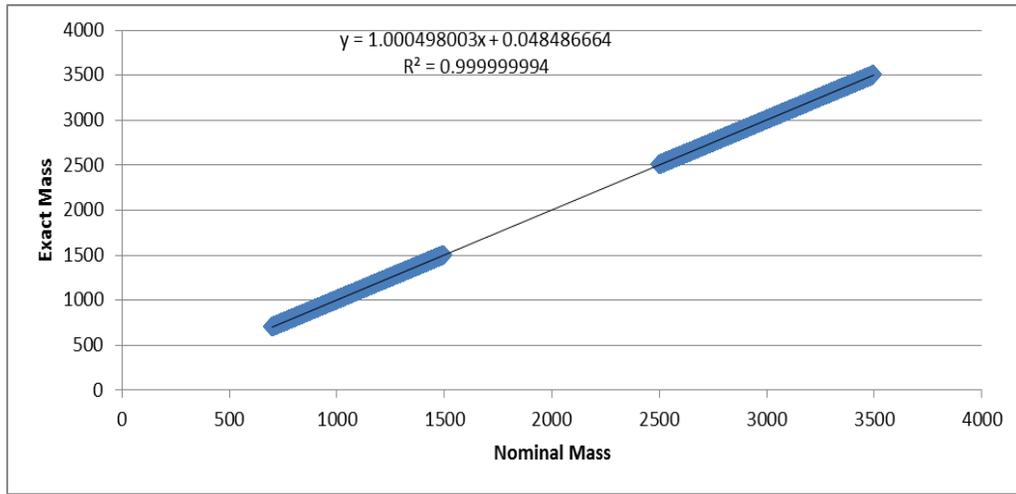


Figure 5.1 Relationship between nominal mass and exact mass of peptides. Plot of the nominal mass versus exact mass of all calculated tryptic peptides from in silico digestion of the *Acinetobacter baylyi* ADP1 proteome, including up to two missed cleavages and containing 109,599 peptides.

By using a crosslinker that contains a mass defect label, crosslinker reaction products are easily flagged, as their observed mass will be shifted into a region of the mass scale that is not occupied by unmodified peptides, decreasing the complexity of the spectrum and improving peak identification, illustrated in Figure 5.2.

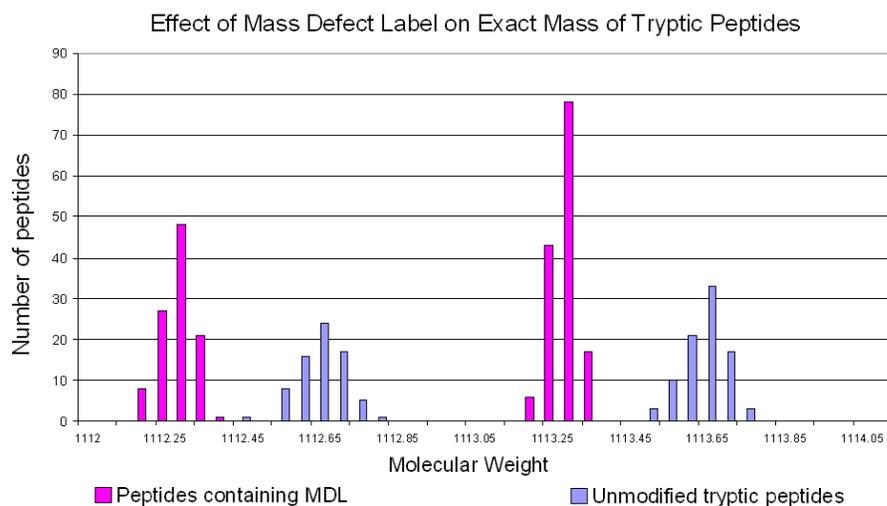


Figure 5.2 Histogram of peptides within a 2 amu range. The blue bars (right-most of each pair) represent standard unmodified tryptic peptides and the magenta bars (left-most of each pair) represent the positions of peptides containing the mass defect label.

Bromine is an ideal mass defect element because of its large mass defect (-82.0 mmu) and thus will be incorporated into our novel crosslinker. It also has an added advantage of changing the isotopic envelope distribution when compared to the standard peptide envelope pattern, as illustrated below. (Figure 5.3).

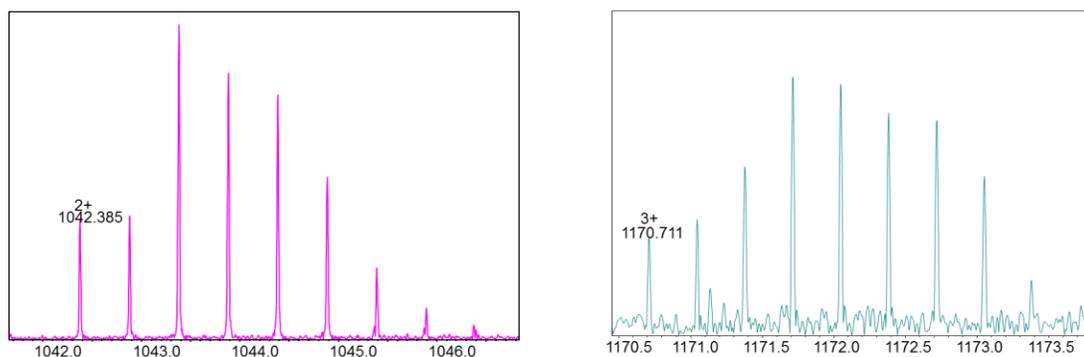
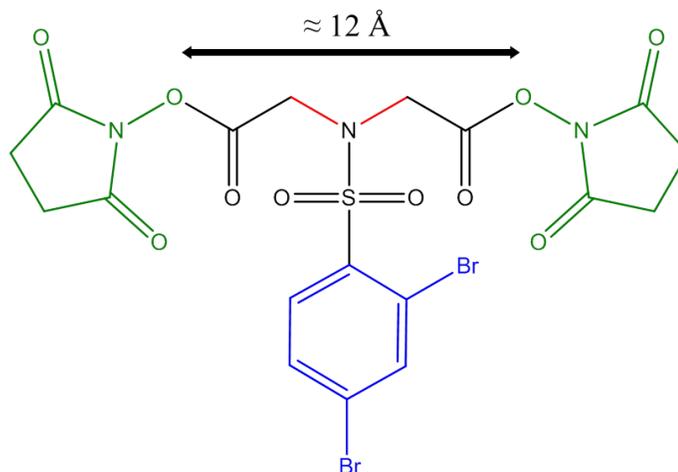


Figure 5.3 Isotopic envelope of mass defect labeled peptides. In addition to the shift in the measured mass, the isotopic envelope also exhibits a significant change, when compared to unmodified peptides. Both spectra were collected from a reaction mixture of bradykinin peptide and neurotensin peptide in the presence of the MDL crosslinker.

The crosslinker developed (Figure 5.4) contains succinimide ester reactive endgroups to allow for chemical specificity to free amines, which includes the protein N-terminus and amino acid lysine. The mass defect label allows for easy detection of reaction products during the data analysis. The tertiary amine and sulfone group increase the lability of the crosslinker during CAD experiments.¹⁸



bis(2,5-dioxopyrrolidin-1-yl) 2,2'-((2,4-dibromophenyl)sulfonyl)azanediyl diacetate

Chemical Formula: C₁₈H₁₅Br₂N₃O₁₀S

Exact Mass: 622.88449

Reactive End Group

Mass Defect Label

CID-Cleavable

Figure 5.4 Chemical structure of the novel chemical crosslinker used. The succinimide esters allow for chemical specificity to free amines, which includes the protein N-terminus and amino acid lysine. The mass defect label allows for easy detection of reaction products during the data analysis. The tertiary amine and sulfone group increase the lability of the crosslinker during CAD experiments.

As the mass defect of a peptide at a particular mass to charge is predictable, reaction products can be easily marked by comparing the measured mass defect of an observed peak to the expected mass defect. The difference between these values will be relatively small (± 0.2 Da) for peptides not containing the mass defect label. A peak list

can be run through an algorithm, shown in Figure 5.5, to quickly assess the nature of the observed peaks. By developing a crosslinker with a labile bond, the fragmentation patterns observed from reaction products should be easier to predict, facilitating the searching and sequencing of MSⁿ spectra.

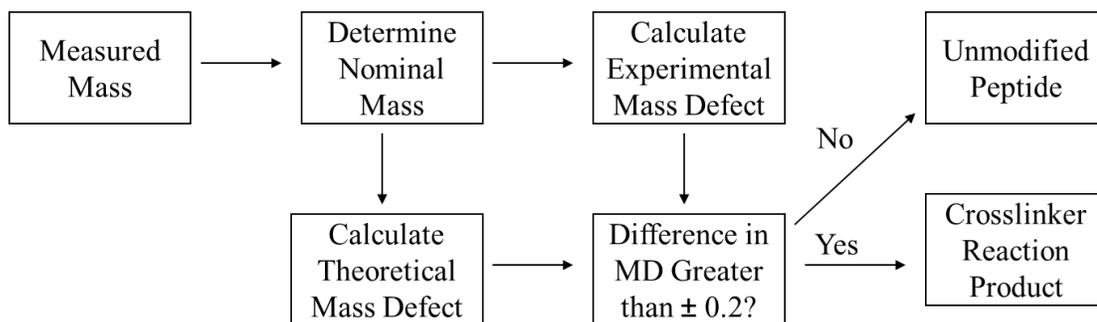


Figure 5.5 Data workflow used to determine the crosslinker reaction products.

Experimental Approach

In order to test the efficacy and robustness of the mass defect labeled crosslinker in chemical crosslinking methodologies, several experiments were performed. The first being a reaction of the crosslinker with a test peptide to assess the reaction products present from a simple crosslinking reaction, and to evaluate the usefulness of the mass defect label in data analysis. The second experiment was a high resolution field asymmetric ion mobility spectrometry (FAIMS) analysis of a simple mixture of peptides. This method allows for the analysis of low intensity peaks of interest, such as crosslinker reaction products, without the time constraints associated with traditional chromatography techniques.²⁴ The final experiment performed was a high resolution analysis of bovine hemoglobin under more traditional experimental conditions, including analysis performed on the liquid chromatography time-scale and much larger data loads.

RESULTS AND DISCUSSION

Test Peptide reactions with mass defect crosslinker

In order to demonstrate that the mass defect labeled crosslinker will allow for easier identification of reaction product species, the MDL crosslinker was reacted with the test peptide, which had been dissolved in triethylammonium acetate buffer. The reaction mixture was then subjected to standard proteolysis, using trypsin as the digestion enzyme. (Figure 5.6) The trypsin will not enzymatically cleave the modified lysine residues.

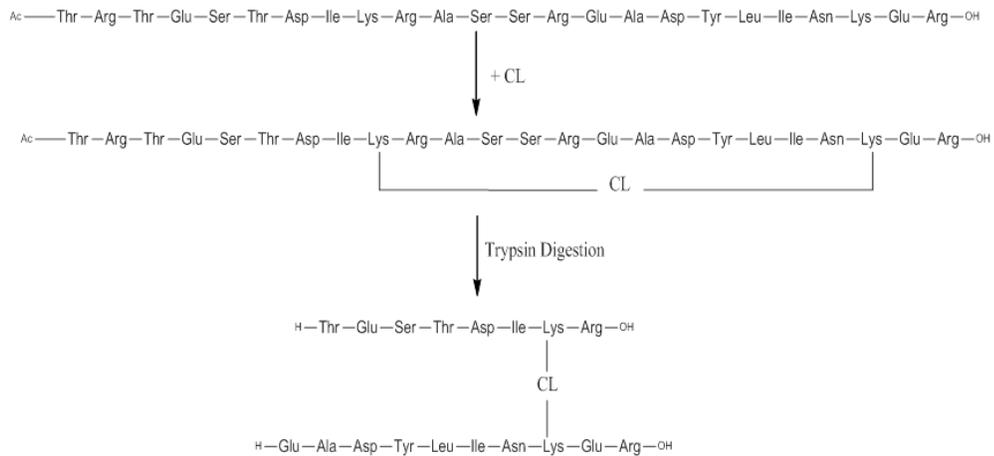


Figure 5.6 Chemical crosslinking reaction scheme used for the test peptide samples.

The resulting mixture was spotted onto the MALDI target and analyzed. (Figure 5.7) During this lower resolution analysis of the sample, all of the expected reaction products were observed: The unreacted tryptic peptides (TESTDIKR and EADYLINKER), the dead-ended reaction products of each lysine containing peptide, and the crosslinker bound to each peptide.

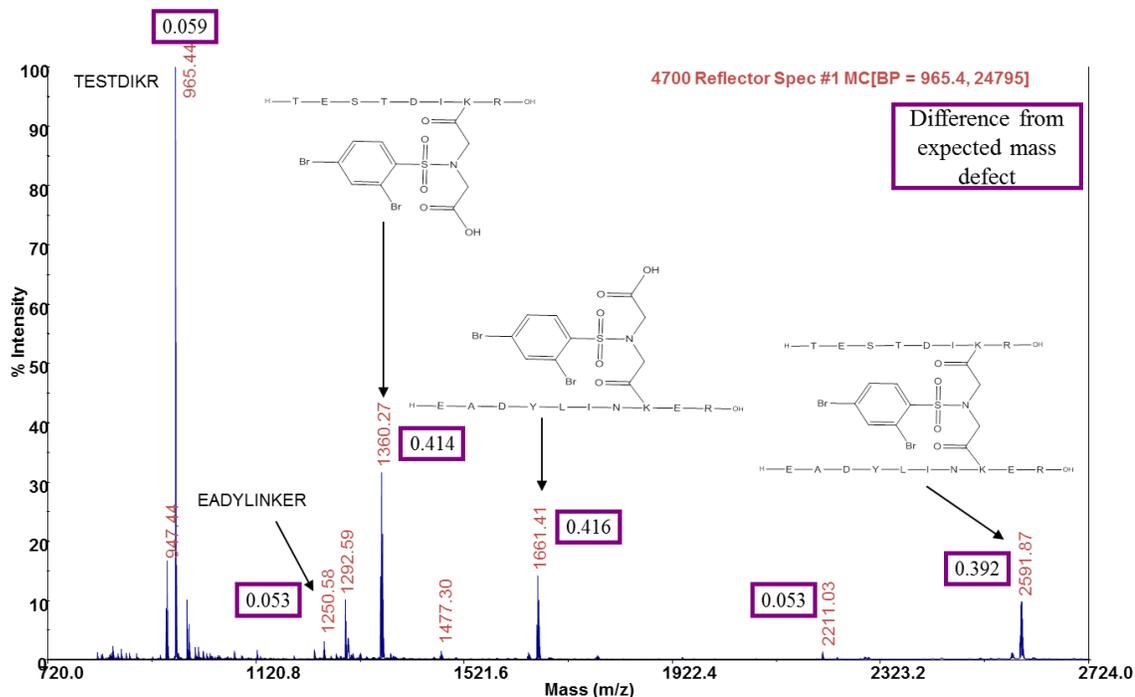


Figure 5.7 MALDI-TOF analysis of the test peptide and crosslinker digested solution. Peaks are observed for each of the unreacted peptides, hydrolyzed dead-ended crosslinked peptides, and the target product containing the two peptides linked.

For each observed mass, the mass defect was calculated compared to the expected mass defect at that mass-to-charge. In Figure 5.7, this value has been boxed next to the product peak. It should be noted that the peaks corresponding to the peptides containing the crosslinker have mass defects significantly different that would be expected for a peak at that mass-to-charge, and therefore should be further investigated.

Although the peak at m/z 2591.87 could be assigned based on accurate mass measurements, the sample was further analyzed via CAD to ensure that the fragmentation of the peak yielded the expected peptides. The sample was analyzed via direct infusion on a 9.4T ESI-FT-ICR, shown in Figure 5.8.

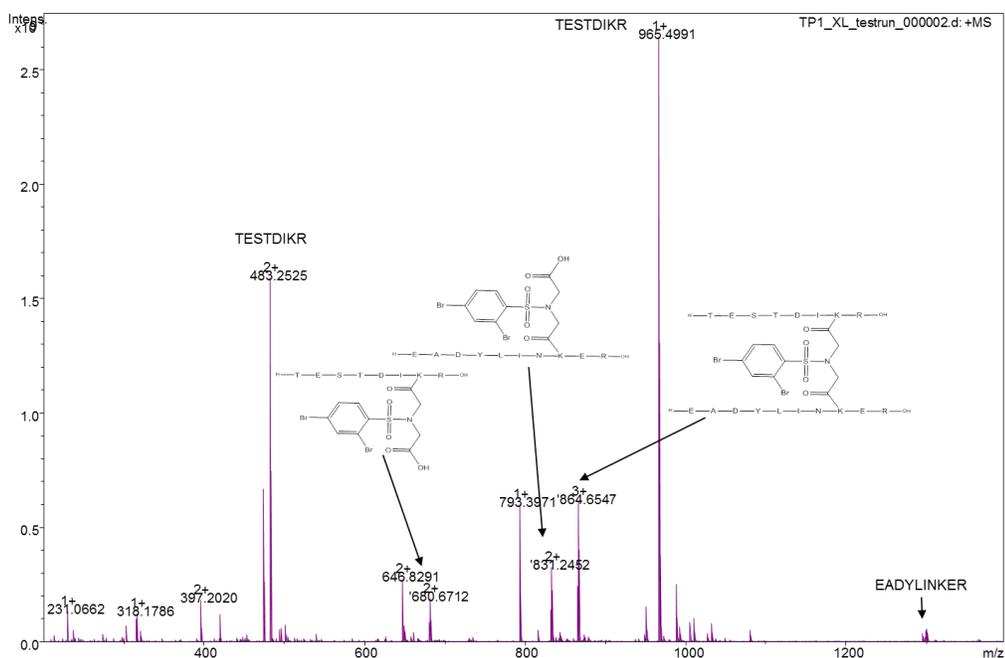


Figure 5.8 FT-ICR analysis of the test peptide and crosslinker digested solution. Peaks are observed for each of the unreacted peptides, hydrolyzed dead-ended crosslinked peptides, and the target product containing the two peptides linked. The peak at m/z 864.65³⁺ was chosen for isolation and CAD fragmentation.

The peak of interest, m/z 864.65³⁺, was isolated and upon isolation, the experimental isotopic distribution was compared to the theoretical isotopic distribution (Figure 5.9), where it was observed that the two matched very well. The ions were then subjected to low collisional energy CAD (10V), and some fragmentation was observed. (Figure 5.10) As the voltage was increased, the fragmentation product intensities and the number of fragments observed increased. It should be noted that the most desired fragments from this experiment would have been the cleavage of the crosslinker from each of the peptides at the reacted lysines, but this was not observed. The fragments that contained the crosslinker still contained the mass defect shift allowing for easier assignment of the fragmentation products.

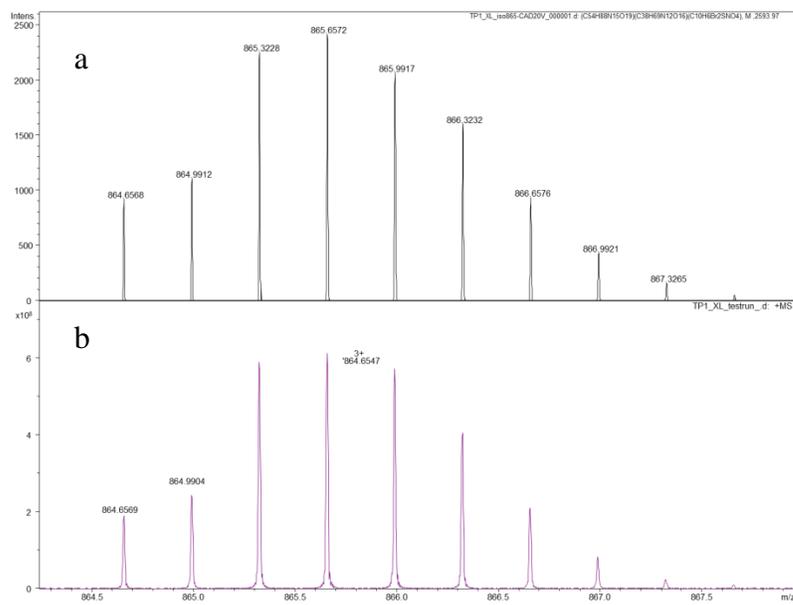


Figure 5.9 Isotopic envelope for crosslinker reaction product from the test peptide. (a) Theoretical isotopic distribution for the target product containing the two peptides linked from the test peptide and crosslinker reaction. (b) Isotopic distribution for m/z 864.65³⁺ precursor ion isolated from the FT-ICR analysis of the test peptide and crosslinker reaction mixture. The isotopic patterns match well, as do the isotopic masses.

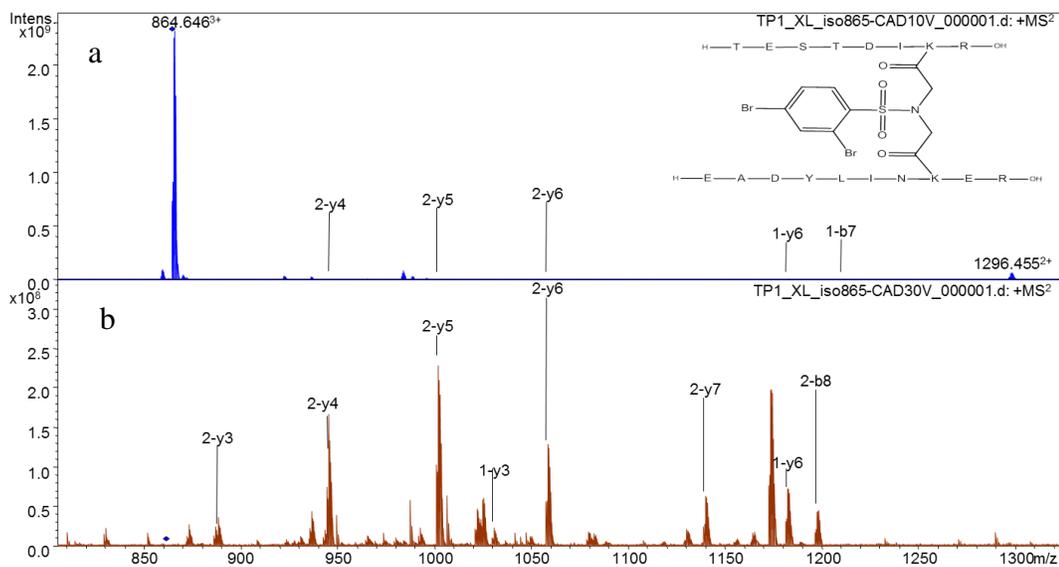


Figure 5.10 FT-ICR analysis of the crosslinked test peptide. The precursor at m/z 864.65³⁺ was fragmented using (a) lower energy CAD (5V) and (b) higher energy CAD (30V).

This set of experiments allowed for the quick assessment of reaction products created by the crosslinking reaction. The use of the mass defect labeled crosslinker allowed for the easy identification of reaction products and fragments that contained the crosslinker.

High Resolution FAIMS analysis of simple peptide mixtures

One of the limitations in chemical crosslinking experiments mentioned earlier was that the most structurally informative peaks tend to be lower in abundance in the sample mixture, which typically also leads to them not being observed with high intensity in the mass spectra.²⁷ High-field asymmetric ion mobility spectrometry has been demonstrated by others to be applicable to highly complex mixtures, and can increase the detectability of lower abundance species.^{24,8} This should be beneficial to the analysis of crosslinked peptides, which are present as minor components after proteolytic digestion of protein complexes that have undergone crosslinking reactions.²⁸ The use of FAIMS as a separation technique yields several advantages over the more conventional method of analysis, separation by high performance liquid chromatography.²⁹ The advantages of FAIMS include speed and flexibility in peak acquisition, longer accumulation times, and multiple scans for fragmentation of low abundance species.

To test the efficacy, robustness, and flexibility of a FAIMS-FT-ICR-MS method, two types of biological samples were analyzed: simple peptide mixture (neurotensin and bradykinin), and a small protein-peptide complex (ribonuclease S and S-peptide). Prior to analysis of protein complexes using FAIMS, neurotensin and bradykinin were dissolved in triethylammonium acetate buffer, and the crosslinker was dissolved in dry acetone, at concentrations of 10 mM each. The solutions were mixed in varying amounts

to determine the optimum molar ratio for the crosslinking reaction, which were then incubated while vortexing for one hour. This reaction mixture was then analyzed using direct infusion (Figure 5.11) to ensure the reaction conditions yielded crosslinker reaction products. The assignable crosslinker reaction products and their relative abundances are listed in Table 5.2. The proposed structures and isotopic envelope analysis are included in Appendix A.

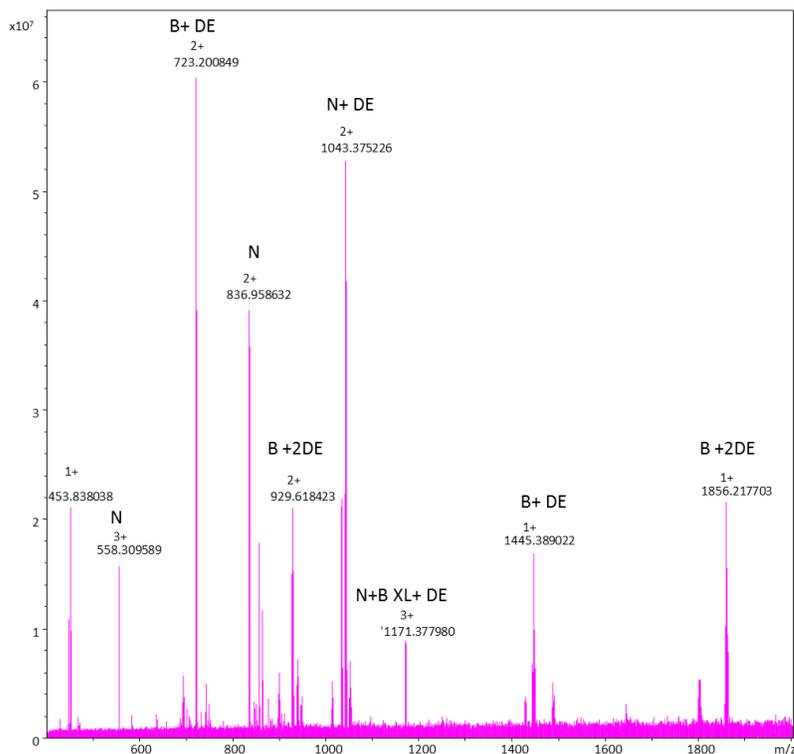


Figure 5.11 Direct infusion mass spectrum for the neurotensin and bradykinin mixture reacted with a 20 times molar excess of crosslinker. The reaction products are labeled: (N= neurotensin, B= bradykinin, DE = hydrolyzed crosslinker (dead-end) attached to the peptide, XL = interpeptide crosslink).

Table 5.2 Observed crosslinker reaction products of the neurotensin and bradykinin mixture.

<u>Reaction Product</u>	<u>Ion Intensity (%)</u>
Unmodified neurotensin	4.95
Unmodified bradykinin	0.00
Neurotensin with 1 DE	15.49
Bradykinin with 1 DE	8.78
Bradykinin with 2 DE	20.62
Bradykinin linked neurotensin with DE	2.96
Bradykinin linked bradykinin with 2 DE	0.98

* DE = hydrolyzed crosslinker (dead-end)

Once these conditions were determined (20 molar excess crosslinker), the sample was analyzed via FAIMS-ESI-FT-ICR/MSⁿ. For the FAIMS analysis, a compensation voltage (CV) scan (1-30V) was collected. Figure 5.12a shows the CV scan acquired, from 1-8V. The mass spectra were then extracted (Figure 5.12b). Peaks containing the mass defect label (m/z 455.84, m/z 530.98, m/z 636.09, m/z 1050²⁺, and m/z 1445.37) were flagged for further analysis. The lower mass peaks (m/z 455.84, m/z 530.98, m/z 636.09) could not be peptide crosslinks, as the crosslinker weighs at least 422 Da. These peaks were considered to be crosslinker salts, and side reaction products. The peaks at m/z 836.95²⁺ and m/z 1032.55 correspond to the masses of the unreacted neurotensin and bradykinin respectively.

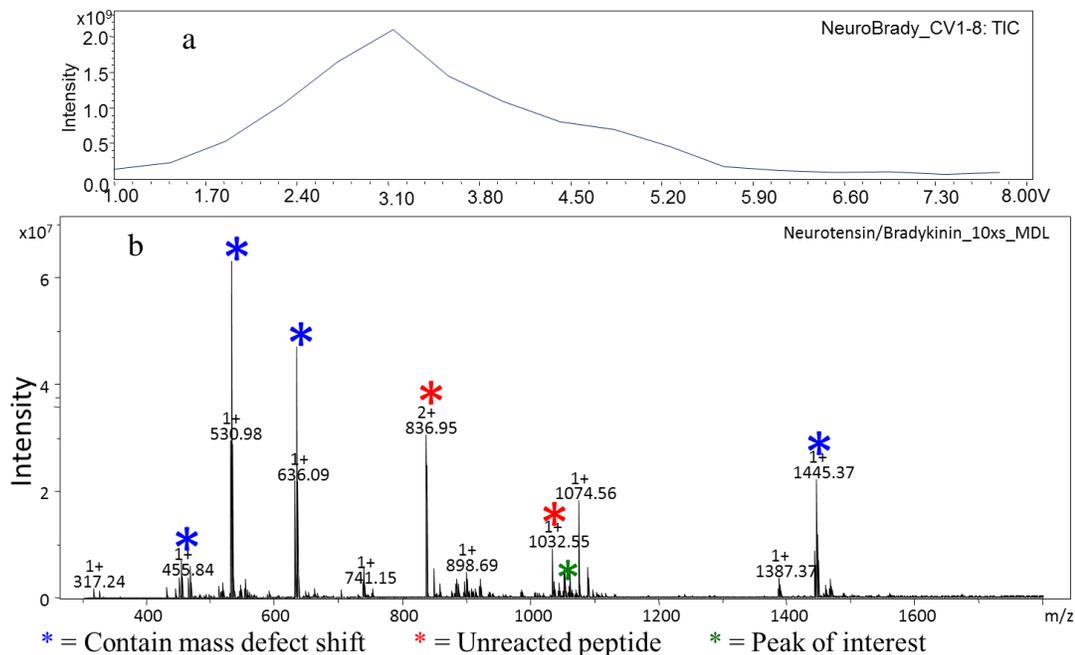


Figure 5.12 FAIMS analysis of the neurotensin and bradykinin crosslinked solution. (a) The total ion chromatogram (TIC) for a mixture of neurotensin, bradykinin, and crosslinker over a CV range of 1 to 8 volts. (b) Extracted mass spectrum from the TIC. Peaks investigated further are m/z 1445.37 and m/z 1050²⁺.

The two peaks chosen for further analysis (m/z 1050²⁺, and m/z 1445.37) had their ion chromatograms extracted from the data to determine the optimal CV voltage for high signal intensity and minimal overlap with other components (Figure 5.13). The optimal CV voltage for each, 2.4V for m/z 1445.37, and 4.5V for m/z 1051²⁺, was used for isolation and fragmentation experiments.

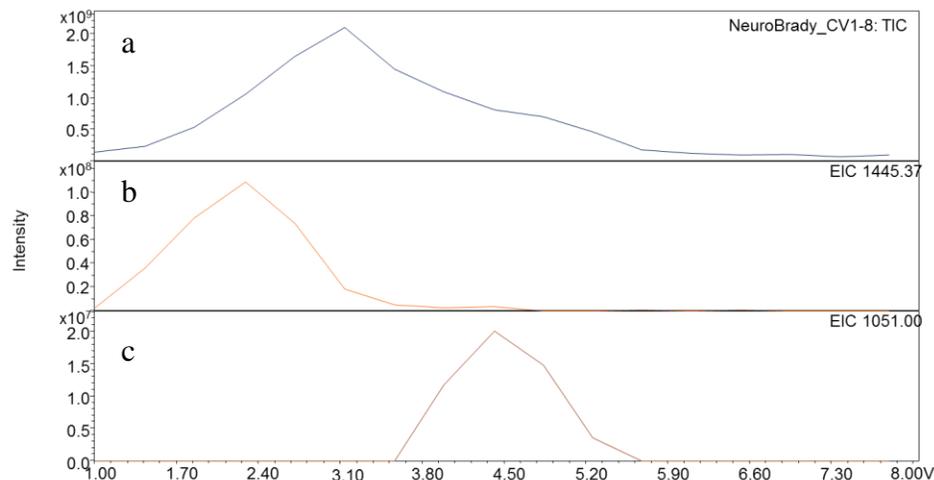


Figure 5.13 Total and extracted ion chromatograms from the analysis of the neurotensin and bradykinin crosslinked reaction solution. (a)The total ion chromatogram for a mixture of neurotensin, bradykinin, and crosslinker over a CV range of 1 to 8 volts. The extracted ion chromatograms for the moieties at m/z 1445.37 (b) and m/z 1051.00²⁺ (c) are shown. The optimal CV voltage for each, 2.4V and 4.5V respectively, were used during the isolation and fragmentation portions of the experiment.

Upon acquiring data at a CV of 2.4V, peaks were observed at m/z 1445.37 and m/z 722.20²⁺, the corresponding doubly charged ion, with the doubly charged ion selected for further isolation in the quadrupole collision cell. The isotopic distribution was observed and matched the theoretical distribution well, in both envelope shape and mass accuracy. (Figure 5.14) The ion was subjected to 20V of collision energy, producing a fragmentation spectrum that confirmed the structure to be a hydrolyzed crosslinker reacted to the N-terminal lysine of the bradykinin peptide. (Figure 5.15)

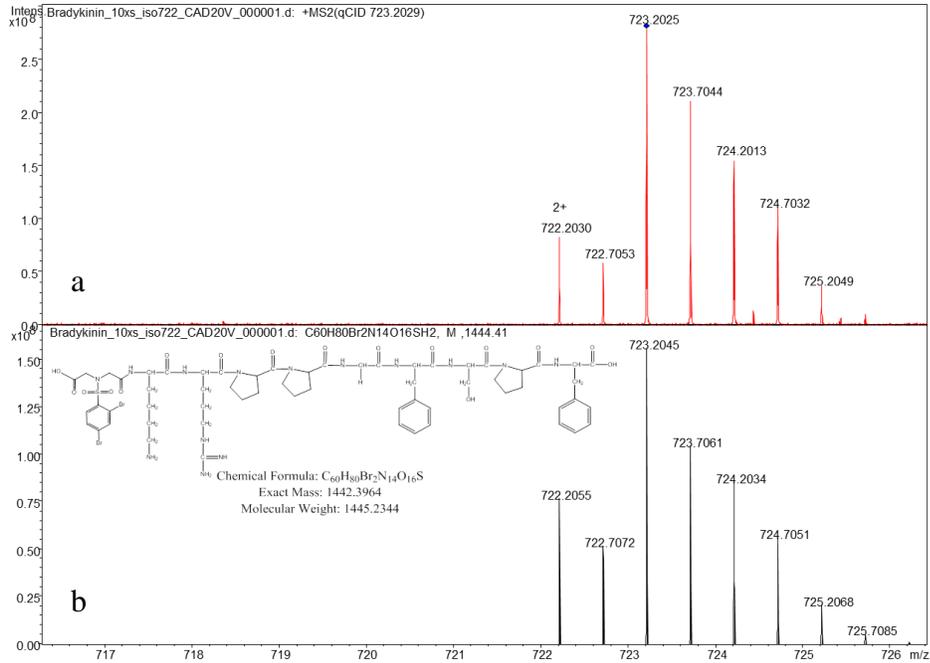


Figure 5.14 Isotopic distribution for m/z 722.20²⁺. (a) Experimental isotopic distribution acquired during isolation of the peak prior to fragmentation. (b) Theoretical isotopic distribution of bradykinin peptide with a hydrolyzed crosslinker attached.

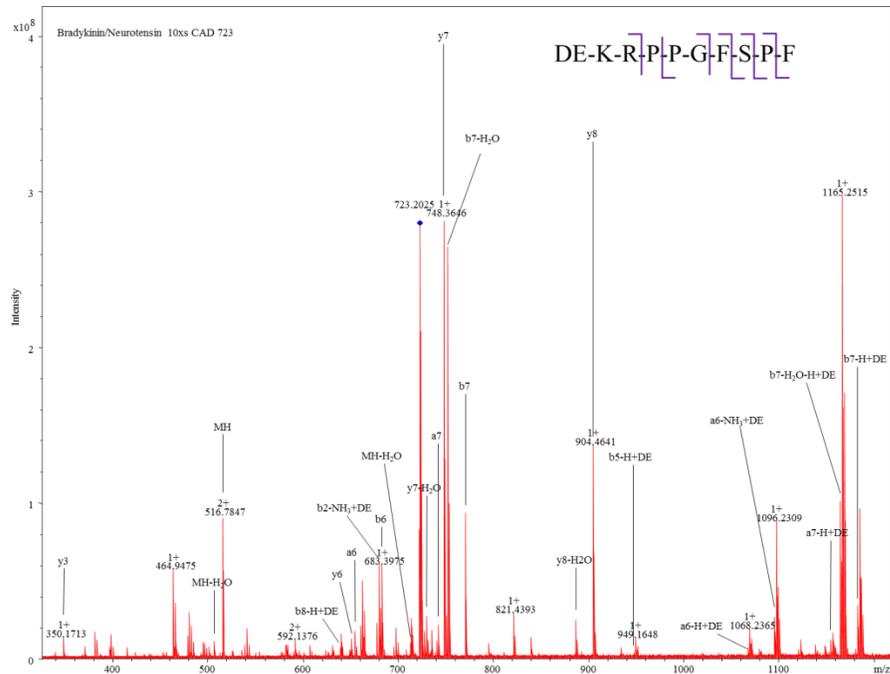


Figure 5.15 MS² fragmentation spectrum of the m/z 722.20²⁺ peak (CAD 20V). This confirms the structure to be a hydrolyzed crosslinker reacted to the N-terminal lysine of the bradykinin peptide.

When acquiring data at a CV of 4.5V, m/z 1050²⁺ was observed. The peak was isolated in the collision cell and activated with 5V of CAD. This led to a mass loss corresponding to the mass of the crosslinker. The m/z 836.95²⁺ peak that resulted from this loss was subjected to infrared multiphoton dissociation (IRMPD), yielding a fragmentation spectrum corresponding to the sequence of the neurotensin peptide. (Figure 5.16)

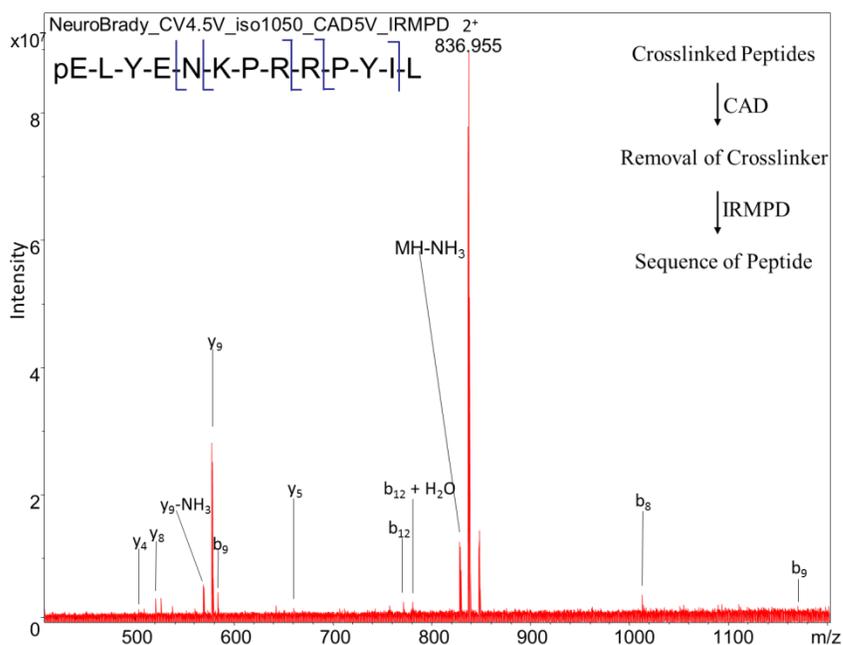


Figure 5.16 MS³ fragmentation spectrum of m/z 1050²⁺. The MDL peak was isolated using a CV of 4.5V, activated with 5V of CAD energy to break the crosslinker moiety off, and then irradiated to yield sequence ions.

Although the ultimate goal from this experiment was to observe a crosslink between the neurotensin and bradykinin peptide, and to then perform a MS³ analysis (use FAIMS to reduce sample complexity and find the crosslinked peak, isolate and activate with low energy CAD to break the crosslinker from the two peptides, then sequence each peptide to confirm identification), the experiments performed did illustrate several advantages of the mass defect labeled crosslinker and the use of FAIMS separation for

crosslinked peptide samples. With the method validated in principle on a very simple mixture, the next step was to move to a slightly more complex system. For this, the protein-peptide complex of ribonuclease S (RNase S) and S-peptide was chosen.

The RNase S complex³⁰ (13.7 kDa), shown in Figure 5.17 is composed of S peptide (amino acid residues 1-20) and S protein (amino acid residues 21-124) and is the hydrolysis product of ribonuclease A (RNase A) reacted with subtilisin, cleaving RNase A between amino acids Ala20 and Ser21.^{31,32} The lysines in the region of interest for crosslinking between S peptide and S protein are Lys1 and Lys7 on S peptide and Lys31 and Lys37 on S protein.^{33,34} The distances between these lysines are listed in Table 5.3.

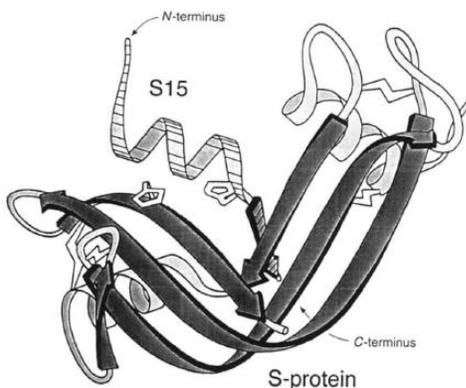


Figure 5.17 Illustration of RNase S complex. The complex is composed of RNase S protein and S-peptide. Adapted from Kim *et al.*³¹

Table 5.3 Distances between lysines of interest on RNase S protein and S-peptide.³⁵

<u>Lysines Linking</u>	<u>Distance (Å)</u>
Lys1 → Lys7	9.516
Lys1 → Lys31	18.79
Lys1 → Lys37	17.18
Lys7 → Lys31	14.66
Lys7 → Lys37	15.22

The RNase S complex was dissolved in triethylammonium acetate buffer, and reacted with the crosslinker, dissolved in dry acetone, with a ten-fold molar excess of crosslinker. The mixture was briefly analyzed via MALDI FT-ICR to ensure the presence of mass defect labeled peaks prior to ESI and FAIMS analysis, as shown in Figure 5.18.

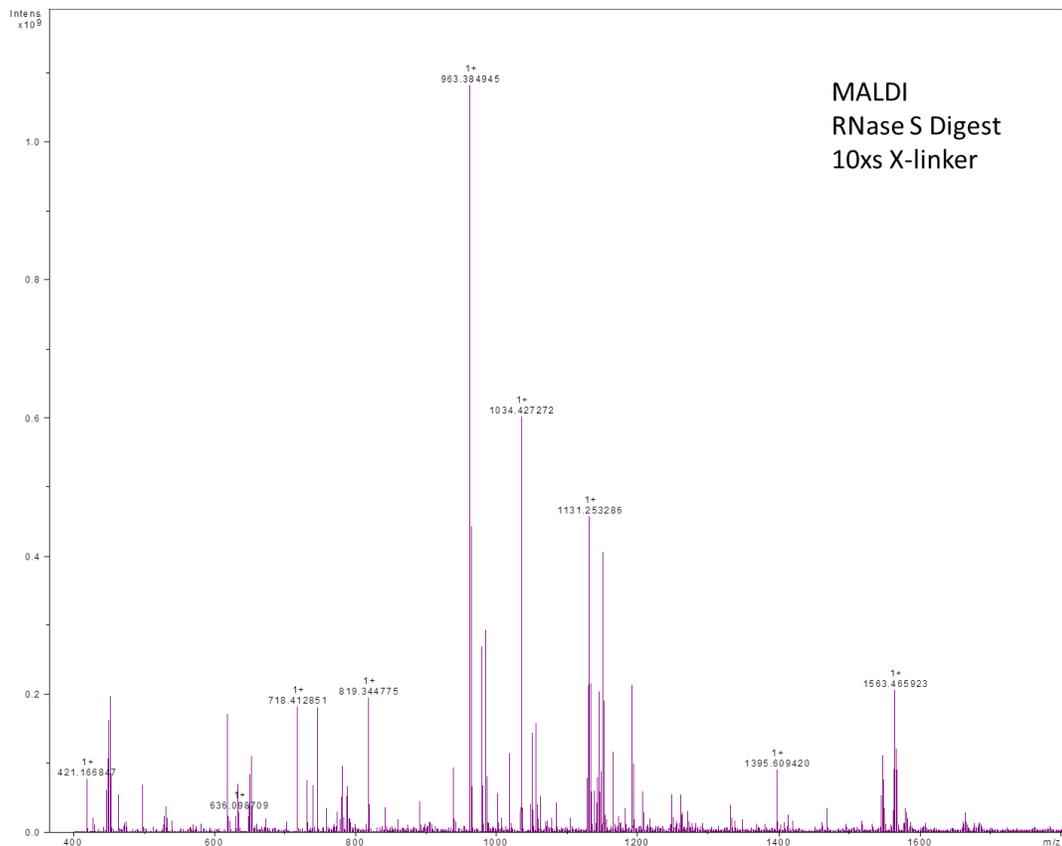


Figure 5.18 MALDI FT-ICR spectrum of the crosslinked RNase S complex tryptic digest.

The experimental mass defects for all of the monoisotopic peaks were calculated, compared to the expected mass defects, and the difference in these values (Δ mass defect) was plotted versus the measured m/z, shown in Figure 5.19. A set of upper and lower

bounds were marked at ± 0.35 Δ mass defect. Any peptides that fell outside of these boundaries indicated labeled peptides.

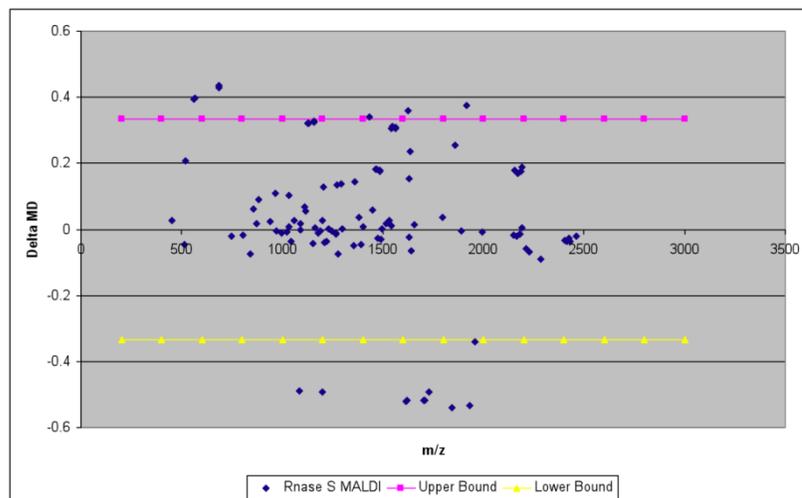


Figure 5.19. Plot of Δ mass defect values versus m/z for all the peaks found in a MALDI mass spectrum of crosslinked RNase S. The points lying outside the upper and lower bounds, shown in pink and yellow respectively, were investigated for evidence of crosslinking.

From the mass spectrum and scatterplot, m/z 1563.459 and m/z 1976.297 were chosen for further analysis. Peaks below m/z 1000 that indicate mass defect labeling are highly unlikely to be due to crosslinked peptides, as the crosslinker itself weighs over 400 Da. Even if the peaks are due to crosslinked peptides, the sequence lengths of the peptide are unlikely to provide any structural information. The crosslinked digest mixture was introduced via direct infusion into the ESI FT-ICR for tandem analysis.

The fragmentation spectra produced yielded very high sequence coverage for a crosslinker modified S-peptide, as shown in Figure 5.20. For m/z 782.233²⁺ (Figure 5.20a) and m/z 988.652²⁺ (Figure 5.20b), the y -ion series from y^4 through y^9 are observed. This information validates the assignment of these peaks based on accurate mass measurement, with m/z 782.233²⁺ corresponding to with one hydrolyzed crosslinker

and m/z 988.652²⁺ corresponding to two hydrolyzed crosslinkers on the N-terminal lysine.

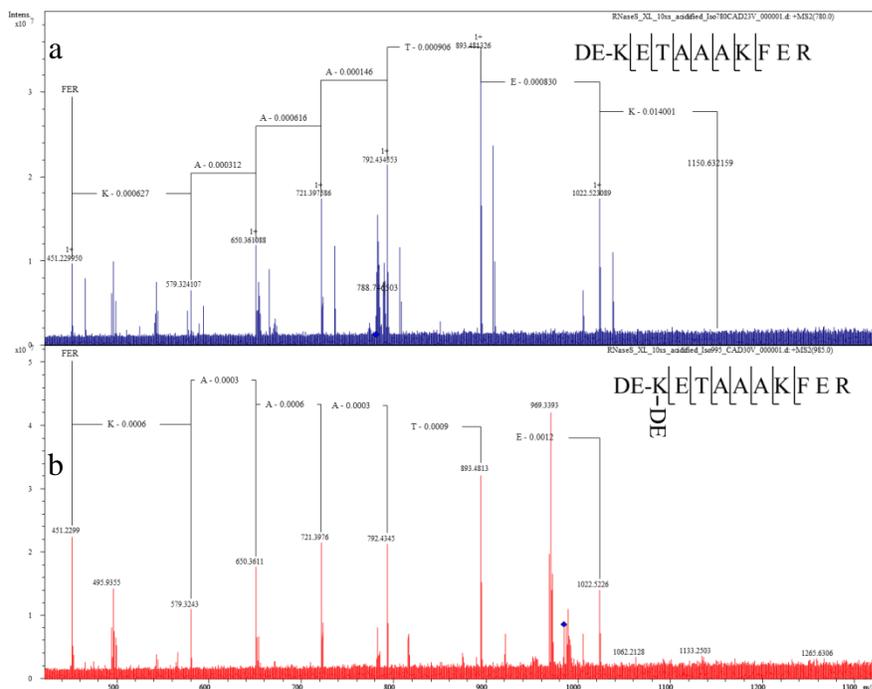


Figure 5.20 Fragmentation spectra sequencing S-peptide. (a) m/z 782.233²⁺ with one hydrolyzed crosslinker and (b) m/z 988.652²⁺ two hydrolyzed crosslinkers on the N-terminal lysine.

Many of the peaks that exhibited the mass defect label were unable to be cleanly isolated during the direct infusion experiment. This necessitated the use of the FAIMS methodology developed to decrease the peak overlap, and allowed for clean isolation and accumulation of crosslinker reaction products. For the FAIMS analysis, the preliminary CV scan (1-30V) was performed (Figure 5.21), the mass spectrum was extracted (Figure 5.22), and the mass list was analyzed to find peaks exhibiting mass defects indicating labeling occurred. Once some candidate peaks were found, an extracted ion chromatogram was performed on the CV chromatogram to determine the optimal CV for

each ion. As shown in Figure 5.23, the optimal CV for the chosen m/z 1131.24 is 4.5V, since there is no overlap with the other peaks that are being investigated as well.

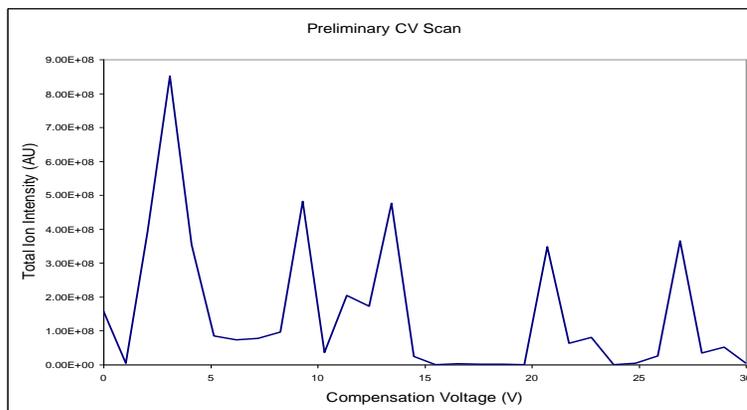


Figure 5.21 Preliminary compensation voltage scan obtained from scanning 1-30 positive volts.

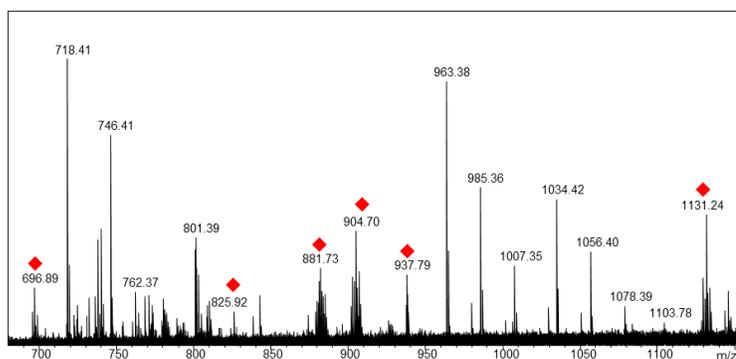


Figure 5.22 Extracted spectral information from the preliminary compensation voltage scan. Red diamonds indicate peaks that contain the mass defect label.

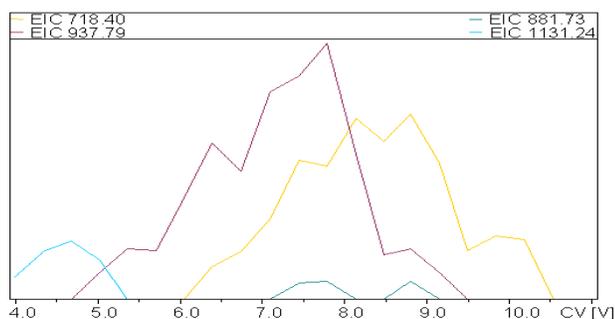


Figure 5.23 Extracted ion chromatograms (EIC) of selected peaks from the spectrum in Figure 5.21.

The compensation voltage was set at 4.5V and held static during the ion accumulation and fragmentation, seen in Figure 5.24. The m/z 1131.243 ion was isolated in the quadrupole, while the CV remained static. A small amount (5V) of collisional energy was imparted to the molecule, inducing fragmentation of the most labile bond, which corresponded to a hydrolyzed crosslinker moiety. The resulting mass at m/z 718.408 is from the first seven amino acids from S-peptide (KETAAAK). From the amino acid sequence, it can be inferred that the hydrolyzed crosslinker was attached at the N-terminus, as the trypsin does not cleave at modified lysine residues.

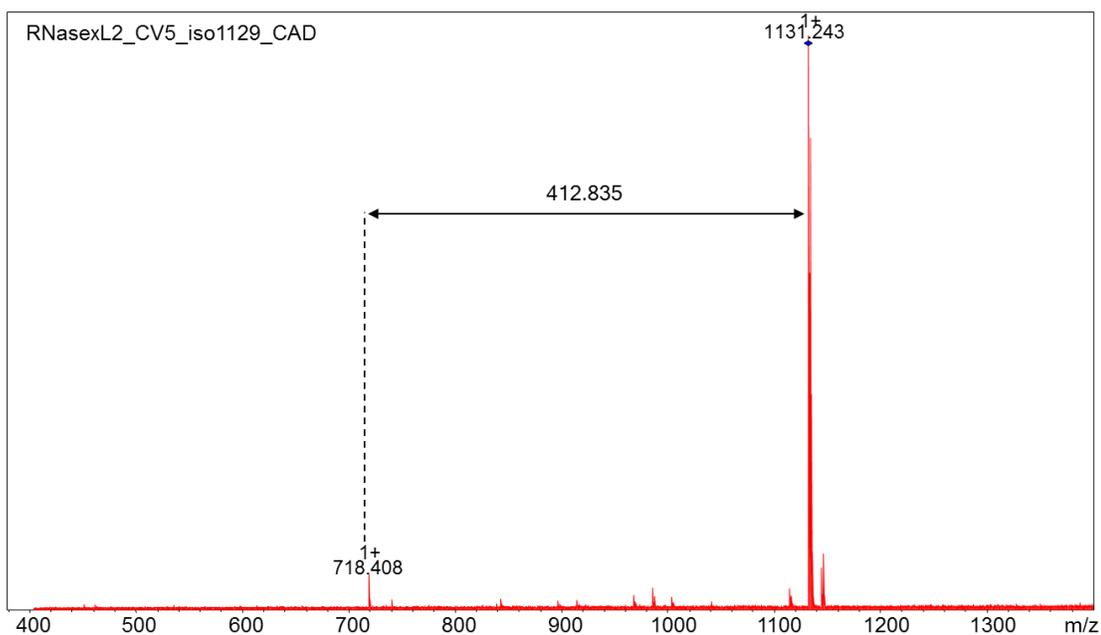


Figure 5.24 Fragmentation spectrum of m/z 1131.243⁺. This peak corresponds to the first seven amino acids from S-peptide (m/z 718.408⁺) and the loss of a mass corresponding to a hydrolyzed crosslinker (412.835 Da).

Although no crosslinks were observed between RNase S protein and S-peptide, this method allowed for characterization of the peaks chosen for analysis, validating the general methodology presented. The experiments demonstrated several of the advantages of using FAIMS as a separation technique. The nature of this technique allows for

alterations to be made to the experimental parameters during the course of the sample analysis, such as method of fragmentation and energies applied, as well as the peaks chosen for analysis, which is one of the advantages of this method compared to other complex mixture analysis techniques. Determination of the optimum compensation voltage of a particular ion allows for long accumulation times, similar to peak parking in standard liquid chromatography, increasing the ability to analyze low abundance ions. This method, paired with the novel mass defect label, can be used to quickly identify reaction products from chemical crosslinking experiments.

High Resolution Bovine Hemoglobin Crosslinking

The crosslinker methodologies proved to be successful in workflows that were not constrained by the standard HPLC chromatographic time-scale. In order to determine the robustness of the mass defect labeled crosslinker, it was tested under conditions that are more consistent with standard proteomic analysis methods, primarily using reverse phase HPLC separation and a more complex sample.

The bovine hemoglobin was dissolved in potassium phosphate buffer, and reacted with the MDL crosslinker which had been dissolved in minimal quantities of acetone, at a ratio of 100 to 1. The reaction solution was quenched using 100 mM ammonium bicarbonate, digested using trypsin, and was analyzed using during the LC-ESI-FT-ICR-MS/MS and nano-LC-ESI-Orbitrap-MS/MS. After the initial chromatograms were collected, the peaks lists were extracted. The mass defect calculation was used to find crosslinked reaction products. The assignments of the peaks were based on sequencing, when available, and the accurate mass measured.

The total ion chromatogram for the bovine hemoglobin crosslinked digestion solution collected on the nano-LC-Elite Orbitrap instrument is shown in Figure 5.25. The LC gradient lasted for 85 minutes, and over 35,000 precursor and fragmentation spectra were collected throughout the LC run.

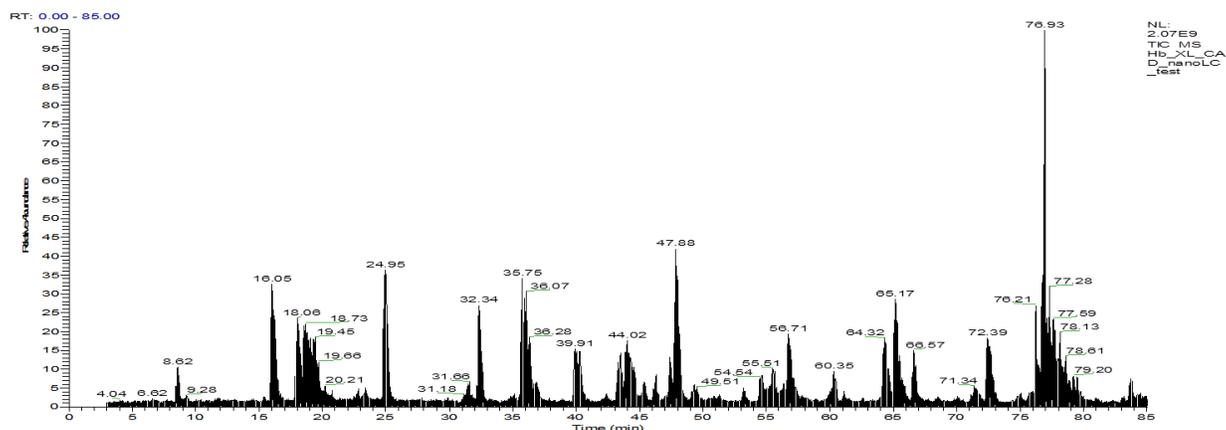


Figure 5.25 The total ion chromatogram for the bovine hemoglobin crosslinked digestion solution collected on the nano-LC-Elite Orbitrap instrument.

The analysis of the data encountered a few problems that had not been seen in the previous experiments. The major difficulty was due to the assignment of the monoisotopic mass for each isotopic envelope observed. The mass defect labeled peaks do not exhibit the same shape isotopic envelope compared to standard peptides, due to the presence of the two bromines in the crosslinker. As most instrumental software assigns monoisotopic peaks based on fitting the observed envelope against a theoretical model, all monoisotopic peaks that contained the crosslinker were improperly assigned. Although all peaks in the isotopic envelope of a mass defect labeled moiety will exhibit the distinctive shift, the correct monoisotopic mass is required in order to make assignments based on accurate mass. This necessitated that each spectrum that contained any isotopes exhibiting the mass defect shift to be manually checked to determine the

monoisotopic masses of the peaks. Although this is not a difficult task in and of itself, it does remove any high-throughput aspect that may have been achieved by introducing the mass defect label.

The peaks that were analyzed and found to contain crosslinked peptides tended to have very high fragmentation efficiency, when they were selected, as seen in Figure 5.26. Using the DXMSMS software¹⁰, sequencing information was manually assigned to each fragmentation spectrum. Again, the preferred bonds to be broken would have been those connecting the crosslinker to the lysine residues, but we did not observe any such fragments. The fragmentation pattern observed was sufficient enough to give us the identity of the two participating peptides, which corresponded to amino acids 40-60 and 65-75 on hemoglobin β subunit. Although there were several crosslinked peptides observed, the sheer volume of the data and the manual assigning of monoisotopic peaks, as well as fragmentation spectra, made the analysis of the entire LC run unfeasible.

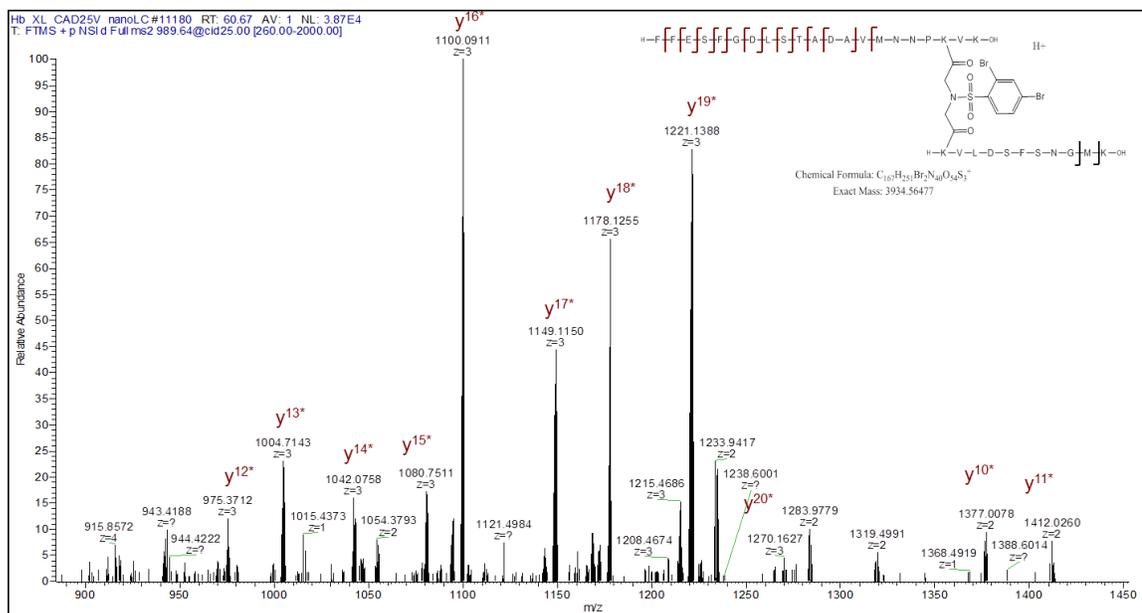


Figure 5.26 The CAD fragmentation spectrum of the m/z 984.64 4^+ precursor peak at 25V of normalized collision energy. The two peptides crosslinked are from hemoglobin β subunit, amino acids 40-60 and 65-75.

The total ion chromatogram for the bovine hemoglobin crosslinked digestion solution collected on the 9.4T LC-FT-ICR instrument is shown in Figure 5.27. The LC gradient lasted for 90 minutes and 3000 spectra were collected during the LC runs. The extracted mass lists were much simpler to obtain from the Bruker Daltonics Data Analysis® software, as it allows changes to be made to the expected elemental composition of the theoretical peaks used to monoisotopic mass assignment. By changing the theoretical composition to also include bromine, the monoisotopic peaks were properly assigned in all spectra automatically. This dramatically improved the speed of the data analysis.

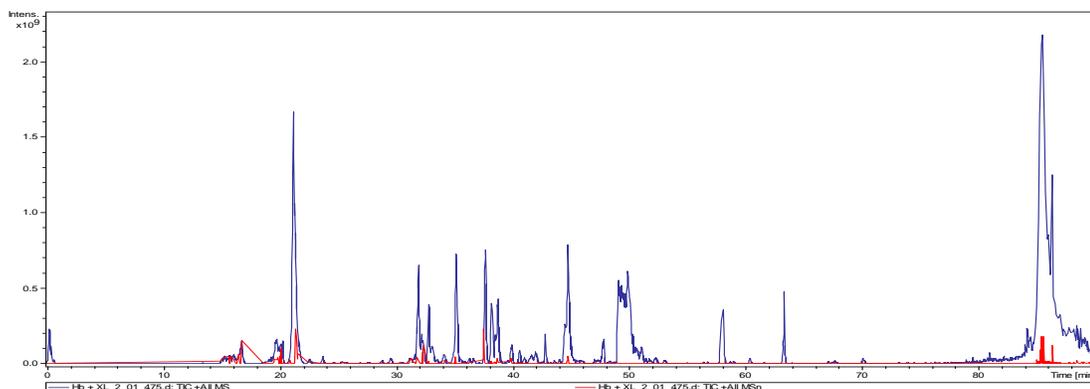


Figure 5.27 The total ion chromatogram for the bovine hemoglobin crosslinked digestion solution collected on the 9.4T LC-FT-ICR instrument. The blue line corresponds to the total ion current observed, and the red line corresponds to the ion intensities observed during the MS/MS analysis.

There were a few problems observed in the FT-ICR analysis though; the first is the ICR duty cycle and the second was the fragmentation efficiency observed. FT-ICR data acquisition takes significantly longer than other mass spectrometry techniques³⁶, and while this is typically not an issue, it can decrease the number of components in a mixture that can be analyzed during a single LC run, as the optimal LC parameters may not allow enough time for all the peaks observed to be isolated, fragments, and analyzed before the next fraction is being introduced to the mass spectrometer. Coupled with this constraint was that very few of the fragmentation spectra yielded more than the precursor peak. Without the fragmentation data, assignments from the FT-ICR experiments were based on accuracy mass measurements.^{37,38} The assignments were performed using MS-Bridge software.³⁹

From the LC-ESI-FT-ICR experiment, the peaks were analyzed for mass defect and compared to the predicted mass lists within a search tolerance of 20 ppm, resulting in 735 unique masses detected with 99 of these masses exhibiting a mass defect outside the expected natural variance. From these, we are able to assign 12 interpeptide crosslinks

and 2 intrapeptide crosslinks. Heterogeneity and impurities within the sample increase the difficulty in unambiguous assignments, even under high mass accuracy conditions.

CONCLUSIONS

From the experiments performed, this chemical crosslinker allows for easy determination of chemical crosslinking reaction products due to the presence of the significant mass defect present. The mass defect crosslinker increases the speed of analysis of crosslinked peaks by focusing peak analysis on the peaks demonstrating the characteristic mass shift. By pairing the mass defect labeled crosslinker with FAIMS, the decreased sample complexity allowed easier isolation of peaks of interest that had been identified as reaction products. The increase in the intensity of low abundance ions of interest would lead to higher efficiency in MS³ analysis of reaction products.

By developing an instrumental method that allows for preferential selection of precursor ions exhibiting the mass defect label, the low intensity crosslinking reaction products will be more likely to be analyzed without the need for crosslink reaction product prediction prior to experimentation. Although there were a few shortcomings of the crosslinker when placed in a traditional proteomic analysis workflow, many of these are due to lacking developments in software for both peak prediction and peak detection during the data acquisition.⁷ If the instrumental software was able to incorporate the ability to identify mass defect shifts during the precursor scan into the peak selection algorithms used for fragmentation, the need for preference lists and multiple LC runs would be minimized, allowing for higher throughput experiments. Finally, increasing the lability of the crosslinker at the site of reaction will allow for a MS³ experiment to be

developed, wherein the crosslinked peaks detected in the MS scan can be fragmented under lower energy CAD conditions to produce the two peptides, which can be further fragmented via higher energy methods to provide the amino acid sequence on reaction products.

REFERENCES

1. Sinz, A., Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J. Mass Spectrom.* **2003**, *38* (12), 1225.
2. Ahrends, R.; Kosinski, J.; Kirsch, D.; Manelyte, L.; Giron-Monzon, L.; Hummerich, L.; Schulz, O.; Spengler, B.; Friedhoff, P., Identifying an interaction site between MutH and the C-terminal domain of MutL by crosslinking, affinity purification, chemical coding and mass spectrometry. *Nucleic Acids Res.* **2006**, *34* (10), 3169.
3. Back, J. W.; Notenboom, V.; de Koning, L. J.; Muijsers, A. O.; Sixma, T. K.; de Koster, C. G.; de Jong, L. Z., Identification of cross-linked peptides for protein interaction studies using mass spectrometry and O-18 labeling. *Anal. Chem.* **2002**, *74* (17), 4417-4422.
4. Alley, S. C.; Trakselis, M. A.; Mayer, M. U.; Ishmael, F. T.; Jones, A. D.; Benkovic, S. J., Building a replisome solution structure by elucidation of protein-protein interactions in the bacteriophage T4 DNA polymerase holoenzyme. *J. Biol. Chem.* **2001**, *276* (42), 39340-39349.
5. Back, J. W.; Sanz, M. A.; De Jong, L.; De Koning, L. J.; Nijtmans, L. G. J.; De Koster, C. G.; Grivell, L. A.; Van der Spek, H.; Muijsers, A. O., A structure for the yeast prohibitin complex: Structure prediction and evidence from chemical crosslinking and mass spectrometry. *Protein Sci.* **2002**, *11* (10), 2471-2478.
6. Dihazi, G. H.; Sinz, A., Mapping low-resolution three-dimensional protein structures using chemical cross-linking and Fourier transform ion-cyclotron resonance mass spectrometry. *Rapid Communications in Mass Spectrometry* **2003**, *17* (17), 2005.
7. Novak, P.; Haskins, W. E.; Ayson, M. J.; Jacobsen, R. B.; Schoeniger, J. S.; Leavell, M. D.; Young, M. M.; Kruppa, G. H., Unambiguous assignment of intramolecular chemical cross-links in modified mammalian membrane proteins by Fourier transform-tandem mass spectrometry. *Anal. Chem.* **2005**, *77* (16), 5101.
8. Barnett, D. A.; Ells, B.; Guevremont, R.; Purves, R. W., Application of ESI-FAIMS-MS to the analysis of tryptic peptides. *J. Am. Soc. Mass Spectrom.* **2002**, *13* (11), 1282-1291.
9. Brinkley, M., A brief survey of methods for preparing protein conjugates with dyes, haptens and crosslinking reagents. *Bioconjugate Chem.* **1992**, *3* (1), 2-13.

10. Back, J. W.; Hartog, A. F.; Dekker, H. L.; Muijsers, A. O.; de Koning, L. J.; de Jong, L., A new crosslinker for mass spectrometric analysis of the quaternary structure of protein complexes. *J. Am. Soc. Mass Spectrom.* **2001**, *12* (2), 222.
11. Senter, P. D.; Tansey, M. J.; Lambert, J. M.; Blattler, W. A., Novel Photocleavable Protein Crosslinking Reagents and Their Use in Preparation of Antibody-Toxin Conjugates. *Photoch. Photobiol.* **1985**, *42* (3), 231-237.
12. Blattler, W. A.; Kuenzi, B. S.; Lambert, J. M.; Senter, P. D., New heterobifunctional protein crosslinking reagent that forms an acid-labile link. *Biochemistry* **1985**, *24* (6), 1517-1524.
13. Bruce, J. E., In vivo protein complex topologies: Sights through a cross-linking lens. *Proteomics* **2012**, *12* (10), 1565-1575.
14. Petrotchenko, E. V.; Olkhovik, V. K.; Borchers, C. H., Isotopically coded cleavable cross-linker for studying protein-protein interaction and protein complexes. *Mol. Cell. Proteomics* **2005**, *4* (8), 1167.
15. Gao, Q.; Xue, S.; Doneanu, C. E.; Shaffer, S. A.; Goodlett, D. R.; Nelson, S. D., Pro-CrossLink. Software tool for protein cross-linking and mass spectrometry. *Anal. Chem.* **2006**, *78* (7), 2145-2149.
16. Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R., Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell. Proteomics* **2010**, *9* (8), 1634-1649.
17. Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M., MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J. Am. Soc. Mass Spectrom.* **2003**, *14* (8), 834-850.
18. Petrotchenko, E. V.; Borchers, C. H., Crosslinking combined with mass spectrometry for structural proteomics. *Mass Spectrom. Rev.* **2010**, *29* (6), 862-876.
19. Petrotchenko, E. V.; Borchers, C. H., ICC-CLASS: isotopically-coded cleavable crosslinking analysis software suite. *BMC Bioinformatics* **2010**, *11* (1), 64.
20. Taverner, T.; Hall, N. E.; O'Hair, R. A. J.; Simpson, R. J., Characterization of an antagonist interleukin-6 dimer by stable isotope labeling, cross-linking, and mass spectrometry. *J. Biol. Chem.* **2002**, *277* (48), 46487-46492.
21. Seebacher, J.; Mallick, P.; Zhang, N.; Eddes, J. S.; Aebersold, R.; Gelb, M. H., Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing. *J. Proteome Res.* **2006**, *5* (9), 2270.

22. Trester-Zedlitz, M.; Kamada, K.; Burley, S. K.; Fenyo, D.; Chait, B. T.; Muir, T. W., A modular cross-linking approach for exploring protein interactions. *J. Am. Chem. Soc.* **2003**, *125* (9), 2416-2425.
23. Hernandez, H.; Niehauser, S.; Boltz, S. A.; Gawandi, V.; Phillips, R. S.; Amster, I. J., Mass defect labeling of cysteine for improving peptide assignment in shotgun proteomic analyses. *Anal. Chem.* **2006**, *78* (10), 3417-3423.
24. Guevremont, R., High-field asymmetric waveform ion mobility spectrometry (FAIMS). *Can. J. Anal. Sci. Spectrosc.* **2004**, *49* (3), 105-113.
25. Gamow, G., Mass defect curve and nuclear constitution. *Proc. R. Soc. A.* **1930**, *126* (803), 632-644.
26. Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K., Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem.* **2001**, *73* (19), 4676-4681.
27. Rappsilber, J., The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Bio.* **2011**, *173* (3), 530-540.
28. Sinz, A., Investigation of protein–protein interactions in living cells by chemical crosslinking and mass spectrometry. *Anal. Bioanal. Chem.* **2010**, *397* (8), 3433-3440.
29. Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J., Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* **2010**, *39* (5), 1633-1655.
30. Smyth, D. G.; Stein, W. H.; Moore, S., The sequence of amino acid residues in bovine pancreatic ribonuclease: revisions and confirmations. *J. Biol. Chem.* **1963**, *238* (1), 227-234.
31. Kim, J. S.; Raines, R. T., Ribonuclease S-peptide as a carrier in fusion proteins. *Protein Sci.* **1993**, *2* (3), 348-356.
32. Hirs, C.; Moore, S.; Stein, W. H., Peptides obtained by tryptic hydrolysis of performic acid-oxidized ribonuclease. *J. Biol. Chem.* **1956**, *219* (2), 623-642.
33. Dill, K. A.; Bromberg, S.; Yue, K.; Chan, H. S.; Ftebig, K. M.; Yee, D. P.; Thomas, P. D., Principles of protein folding—a perspective from simple exact models. *Protein Sci.* **1995**, *4* (4), 561-602.
34. Tirado-Rives, J.; Jorgensen, W. L., Molecular dynamics simulations of the unfolding of an. alpha.-helical analog of ribonuclease A S-peptide in water. *Biochemistry* **1991**, *30* (16), 3864-3871.

35. Lin, S. H.; Konishi, Y.; Denton, M. E.; Scheraga, H. A., Influence of an extrinsic crosslink on the folding pathway of ribonuclease A. Conformational and thermodynamic analysis of crosslinked (7-lysine, 41-lysine)-ribonuclease A. *Biochemistry* **1984**, *23* (23), 5504-5512.
36. Comisarow, M. B.; Marshall, A. G., Fourier transform ion cyclotron resonance spectroscopy. *Chem. Phys. Lett.* **1974**, *25* (2), 282-283.
37. Zhang, L. K.; Rempel, D.; Pramanik, B. N.; Gross, M. L., Accurate mass measurements by Fourier transform mass spectrometry. *Mass Spectrom. Rev.* **2005**, *24* (2), 286-309.
38. Sinz, A., Chemical cross-linking and FTICR mass spectrometry for protein structure characterization. *Anal. Bioanal. Chem.* **2005**, *381* (1), 44.
39. Chu, F.; Baker, P. R.; Burlingame, A. L.; Chalkley, R. J., Finding chimeras: a bioinformatics strategy for identification of cross-linked peptides. *Mol. Cell. Proteomics* **2010**, *9* (1), 25-31.

CHAPTER 6

CURCUMIN ANALOGUE ANALYSIS AS AN INSIGHT INTO MALDI MECHANISMS

Accurate mass measurements of intact molecular ions, especially those of thermally labile molecules, require soft ionization methods, such as matrix assisted laser desorption ionization (MALDI). The MALDI principle was discovered in 1984¹, used to measure an oligopeptide in 1986², and used to measure proteins with masses exceeding 10kDa and 100kDa in 1988.³ Tanaka and coworkers developed the use of MALDI concurrently in 1987, utilizing a more convenient physical matrix applied to the samples.⁴ During a MALDI experiment, the sample is mixed with a photon-absorbing organic acid matrix, which is subsequently ablated and desorbed, from the surface the sample is deposited onto, with the use of a pulsed laser.

The analysis of biologically active pharmaceuticals often employs mass spectrometry techniques. The accurate mass measurement of these molecules ensures that the samples not only have the elemental composition expected, but further fragmentation of the isolated peaks can yield structural information. Mass spectrometric analysis also can be used to determine the purity of single component mixtures after

synthesis. In order for these techniques to be of use, the observed peaks produced should be easily assigned, usually as a protonated or deprotonated molecular ion, depending on whether the analysis is performed in positive or negative ion mode. When peaks are observed that do not fall into the expected molecular ion form, the sample is usually questioned, but in the following experiments the ionization mechanism is investigated to gain insight into the chemistry that is occurring, to produce these uncommon ionization products.

The analysis of several antineoplastic agents yielded an uncommon ionization peak ($[M-H]^+$) typically not observed in positive mode matrix assisted laser desorption ionization experiments. These molecules were designed by structural modification of curcumin molecules to enhance bioavailability and anticancer characteristics.⁵ The curcumin analogues are 3,5-bis(benzylidene)-4-piperidone compounds, with 1,5-diaryl-3-oxo-1,4-pentadienyl pharmacophore (Figure 6.1) which is responsible for the antitumor effect of these compounds.⁶

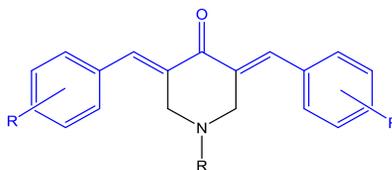


Figure 6.1 General structure of the curcumin analogues analyzed (-3,5-bis(benzylidene)-4-piperidones).⁶ The 1,5-diaryl-3-oxo-1,4-pentadienyl pharmacophore is highlighted in blue.

Positive mode ionization of the curcumin analogues should produce a protonated molecular ion, as the basic site (nitrogen) will happily accept the proton, producing the

$[M+H]^+$ ion. In addition to this ion, a mass coinciding with the $[M-H]^+$ ion was observed.

The proposed structures for these ions are shown in Figure 6.2.

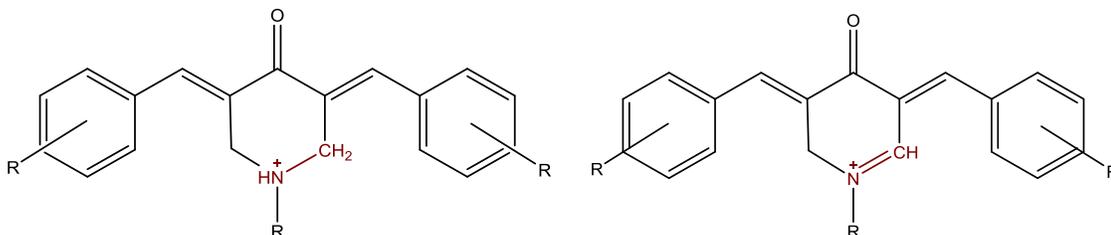


Figure 6.2 Possible chemical structures for the two most intense peaks observed in positive mode MALDI analysis of the curcumin analogues. The $[M+H]^+$ ion and $[M-H]^+$ ion are shown on the left and right respectively.

The $[M-H]^+$ ion has been observed in many other classes of molecules and with many ionization methods, including chemical ionization (CI)^{7,8}, atmospheric pressure chemical ionization (APCI)⁹, electrospray ionization (ESI)^{10,11}, MALDI^{12,13,14,15,16}, atmospheric pressure photoionization (APPI)¹⁷, desorption APPI (DAPPI)¹⁸, desorption ESI (DESI)¹⁸, and direct analysis in real time (DART).¹⁹ The current explanations regarding the formation of the $[M-H]^+$ ion can be categorized into three general mechanisms: loss of a hydrogen molecule from the typical protonated ion ($[M+H]^+$) (reaction 1)¹⁵, hydrogen transfer from the analyte radical cation (reaction 2)^{14,15}, and hydride abstraction from the neutral analyte molecule (reaction 3).^{12,13} These mechanisms will be the focus of the experiments performed, but other general mechanisms related to general MALDI ion formation will also be considered.





There are two major models used to explain the ion generation mechanism observed during MALDI analysis: gas phase protonation and lucky survivor. It is also possible that some of the ions produced are a result of photoionization directly via the ultraviolet laser. There have also been proposed models that involve a combination of both of the two major models.^{20,21} Many other models have been proposed, including excited state proton transfer²², polar fluid model^{23,24}, and expanding bubble²⁵ model.²⁶ For the scope of the experiment discussed, the focus will be limited to the aspects related to the two major models.

The gas phase protonation mechanism proposes matrix molecules are photoionized forming matrix ions, followed by charge transfer from matrix ions to neutral analyte molecules in the gas phase producing the analyte ions.^{27,28,29} The photon interaction with the neutral matrix molecules leads to the formation of both negative and positive ions.²⁸ This is supported by the presence of many types of radical species that are generated during ionization.³⁰ There are a few shortcomings of this model, as the required amount of energy needed for matrix ionization to occur would require multiple photons, so new models have been developed implementing energy concentration, also called “pooling”.²⁸

In the lucky survivor model, the analytes are incorporated into the matrix as charged molecules with its corresponding counter ions (based on the solvent used). Desorption of matrix crystals by laser beam is followed by quantitative charge neutralization between analyte and counter ions producing the singly charged analyte

ions (lucky survivors).²¹ Figure 6.3 illustrates the major mechanisms involved during desorption and ionization following the lucky survivor model.

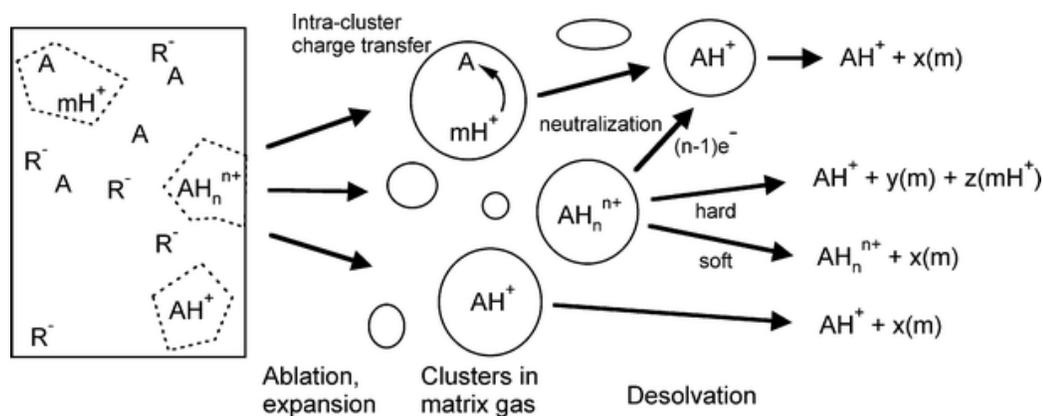


Figure 6.3 General schematic of the major processes proposed in the lucky survivor model. Figure adapted from Knochenmuss.²⁶

The species illustrated are the matrix (m), analyte (A), and generic counter-ions (R). Preformed ions, contained in clusters, are ablated from the target surface, yielding clusters of matrix and analyte in the gas phase. These clusters are desolvated and yield free ions.

To determine the extent to which the various possible ionization mechanisms participated in the formation of the $[M-H]^+$ ion, the experiments investigated various factors related to the solvent conditions of the sample and matrix, gas phase environment and, instrumentation. Experiments investigated whether the solvent conditions of the sample and the matrix used affected the ions observed. To test the sample solvent conditions, the samples were dissolved in an aprotic (acetonitrile) and protic (50: 50: 0.1 water: acetonitrile: trifluoroacetic acid (TFA)) solvents prior to deposition onto the

MALDI plate. The matrix used (2,5-dihydroxybenzoic acid) was dissolved in the same solvent for the analysis. The influence of TFA was also questioned, so samples were prepared with the protic solvent minus TFA.

The environment the ions experience directly after ablation can have a large effect on the ions generated. To better understand the influence of a few controllable factors, the laser power, matrix presence, and the presence of a radical scavenger (butylated hydroxytoluene, BHT) were tested. The matrix was also interrogated regarding its influence on the presence of the $[M-H]^+$ ion. Two matrices were tested, DHB and 3-hydroxy-4-nitrobenzoic acid, as well as performing control experiments with no matrix utilized. The experiments without matrix deposition are called laser desorption (LD) as the ablation, desorption, and ionization of the ions are unaided by the matrix.

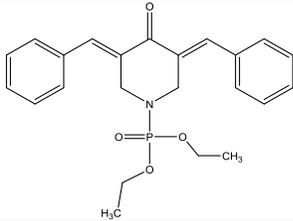
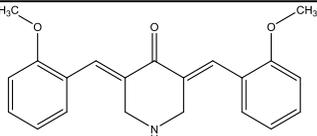
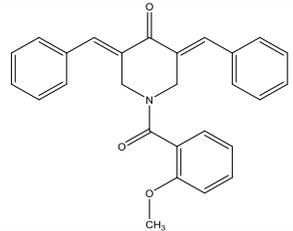
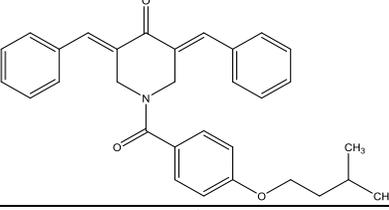
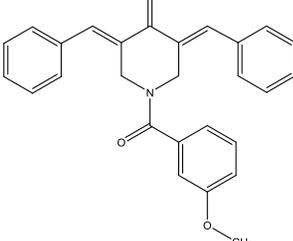
The impact of the instrumental parameters used during ion generation and detection were also probed. The experiments looked specifically at the effect of the laser power and the type of mass analyzer. Samples were analyzed using two laser powers: high (lowest attenuation available) and threshold (minimal power necessary to create observable ions). The analyses were performed on two types of MALDI instruments: a 7T MALDI-FT-ICR and a MALDI TOF. Both instruments used the same laser wavelength (337 nm) and similar vacuum pressures during ion generation ($\sim 10^{-6}$ to 10^{-7} torr). The ion generation and detection of the ions from these two techniques are significantly different though. This can be beneficial in allowing the ions generated to be measured at two time points, as the TOF analysis occurs within roughly 100 μ s of ion generation, and ICR detection is on the 1s timescale.

EXPERIMENTAL

The samples (Table 6.1) were synthesized by the Dr. J. Dimmock⁶ research group and were analyzed using MALDI and laser desorption (LD) experiments performed on two MALDI instruments: a Bruker Daltonics 7T MALDI-FT-ICR-MS equipped with 337 nm nitrogen laser, operated in the positive ion mode, and a Bruker Daltonics AutoFlex MALDI-TOF equipped with a 337 nm nitrogen laser in positive and negative mode. The instruments were calibrated with a tryptic digest of bovine serum albumin. The samples were prepared by using trace amounts of curcumin analogue dissolved in various solvents and spotted, adding a matrix in a ratio of 50:1 matrix: analyte. The LD analysis was performed in the exact same manner without the matrix.

Samples were also prepared without solvents. For these solvent-free samples, the dry sample crystals were crushed onto the surface of the MALDI target using a spatula. Matrix and butylated hydroxytoluene, if used, were added to the target and crushed along with the sample. The targets were gently tapped on the bench top to remove an excess powdered sample.

Table 6.1 The tested curcumin analogues, their structures and monoisotopic masses.⁶

<u>Sample</u>	<u>Monoisotopic Mass</u>	<u>Structure</u>
NC 2311	411.1599 Da	
NC 2453	335.1521 Da	
NC 2144	409.1678 Da	
NC 2094	465.2304 Da	
NC 2138	409.1678 Da	

RESULTS AND CONCLUSIONS

Results

Under the initial experimental MALDI conditions, the sample and DHB were dissolved in 50:50:0.1 (water: acetonitrile: TFA) and analyzed on the FT-ICR. The LD experiment was prepared the in the same manner as the MALDI analysis without the matrix. The spectra acquired from the analyses of the samples are included in Appendix B, and the accurate mass measurements are summarized in Table 5.2 and Table 5.3. In each MALDI spectrum acquired, there were peaks observed corresponding to the traditional $[M+H]^+$ ion, as well as the less common $[M-H]^+$ ion (Figure 6.4). The LD analysis of each sample yielded the $[M-H]^+$ as the primary assignable peak. As shown in Figure 6.5, $[M+H]^+$ may be present, but the abundance is too low to discern from the isotope series peaks, as the peak at m/z 412.167 can correspond to $[M+H]^+$ or $[M+H]^+$ with two heavy isotopes (naturally occurring abundances). In the MALDI FT-ICR analyses, both $[M+H]^+$ and $[M-H]^+$ are observed under all sample and matrix solvent conditions, as well as under all laser powers tested. However, during the LD analysis, the $[M-H]^+$ peak was observed and the $[M+H]^+$ peak was absent.

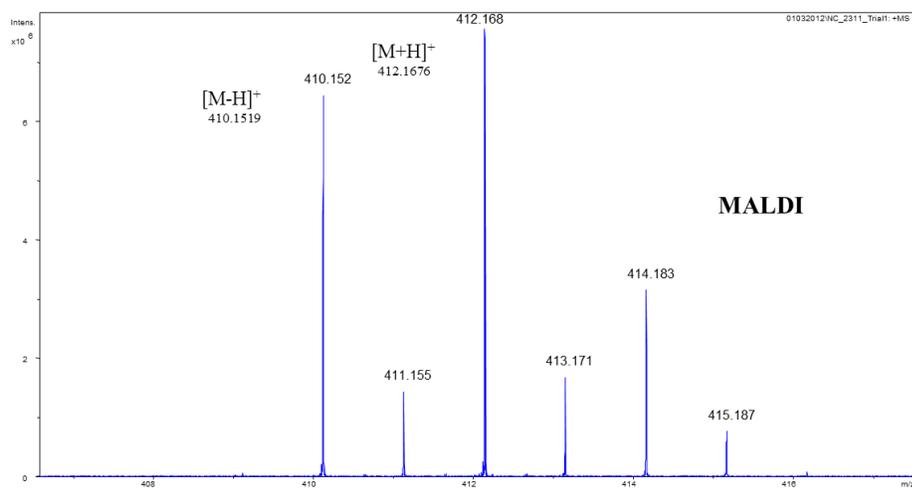


Figure 6.4: FT-ICR MALDI-MS mass spectrum of curcumin analogue sample NC 2311. The peaks of interest are $[M+H]^+$ (m/z 412.1676) and $[M-H]^+$ (m/z 410.1519).

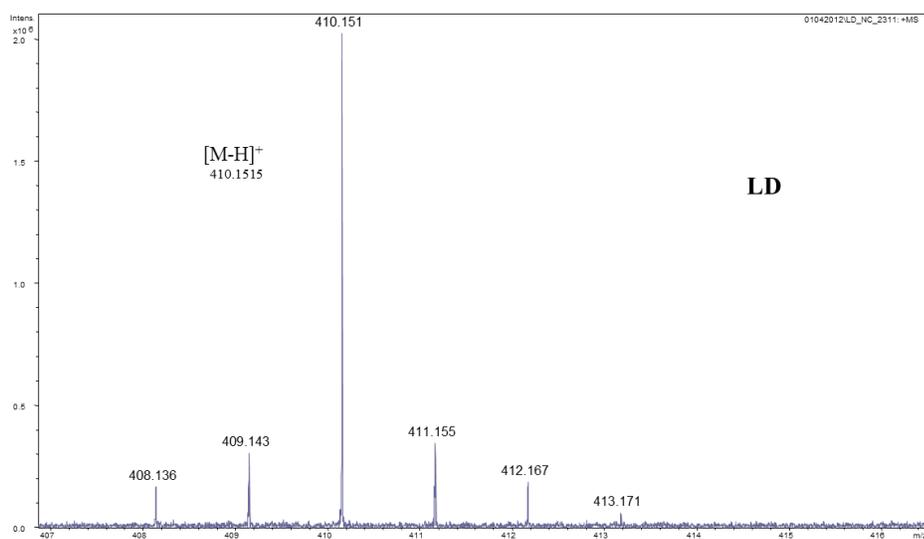


Figure 6.5: FT-ICR LD-MS mass spectrum of curcumin analogue sample NC 2311. The peak of interest is $[M-H]^+$ (m/z 410.1519).

Table 6.2 Theoretical and measured monoisotopic masses for the [M+H]⁺ and [M-H]⁺ peaks from the MALDI FT-ICR analyses.

<u>Compound</u>	<u>Theoretical m/z</u>		<u>Observed m/z</u>		<u>Mass Accuracy (ppm)</u>	
	<u>[M+H]⁺</u>	<u>[M-H]⁺</u>	<u>[M+H]⁺</u>	<u>[M-H]⁺</u>	<u>[M+H]⁺</u>	<u>[M-H]⁺</u>
NC 2311	412.1672	410.1516	412.1676	410.1519	-0.878	-0.789
NC 2453	336.1594	334.1438	336.1597	334.1440	-0.730	-0.928
NC 2144	410.1751	408.1594	410.1748	408.1592	0.638	0.568
NC 2094	466.2377	464.2220	466.2375	464.2219	0.314	0.429
NC 2138	410.1751	408.1594	410.1746	408.1590	0.997	1.23

Table 6.3 Theoretical and measured monoisotopic mass for the [M-H]⁺ peaks observed during the LD FT-ICR analyses. Masses calculated for NC 2453 are for the dimer observed.

<u>Compound</u>	<u>Theoretical m/z</u>	<u>Observed m/z</u>	<u>Mass Accuracy (ppm)</u>
NC 2311	410.1516	410.1515	0.173
NC 2453	667.2802	667.2800	0.373
NC 2144	408.1594	408.1590	1.06
NC 2094	464.2220	464.2214	1.24
NC 2138	408.1594	408.1591	0.669

The samples were also analyzed on the MALDI-TOF. In addition to spectra collected in the same manner as the FT-ICR analysis, spectra were also collected in negative mode. Samples were also analyzed after the addition of BHT to the samples. BHT is a radical scavenger in solution, but it is unknown whether the radical scavenging occurs in the gas phase at the same efficiency. The addition of the BHT to the samples should allow for radical pathways to be shut down, eliminating them as potential sources of [M-H]⁺ ion generation. The data is shown in Figure 6.6, Figure 6.7, and Figure 6.8.

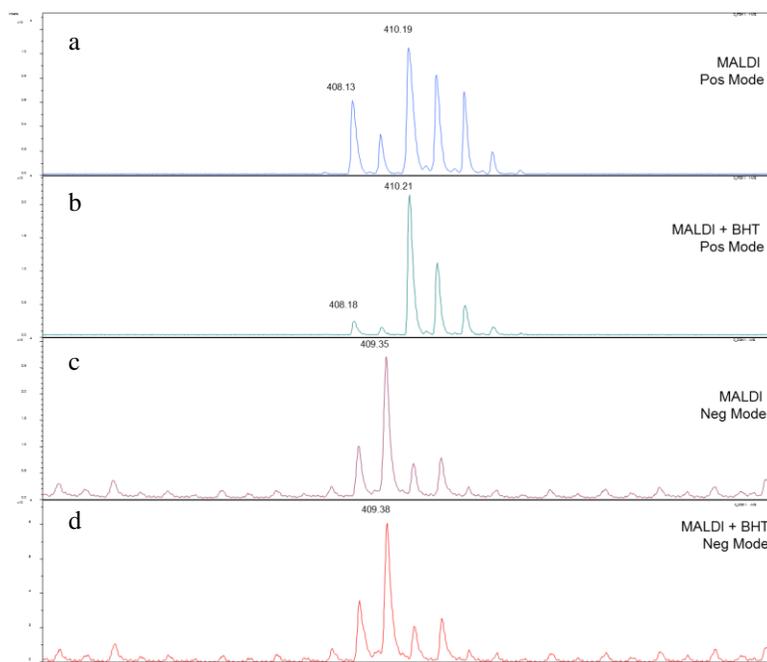


Figure 6.6: MALDI-TOF analysis of NC 2144 under various conditions. (a) Positive mode with no radical scavenger, (b) positive mode with radical scavenger, (c) negative mode with no radical scavenger, and (d) negative mode with radical scavenger. The peaks of interest are $[M-H]^+$ (m/z 408.1), $[M+H]^+$ (m/z 410.2) and $[M]^-$ (m/z 409.4).

$[M-H]^+$ (m/z 408.1) and $[M+H]^+$ (m/z 410.2) peaks are observed in the MALDI-TOF data obtained in positive mode, and $[M]^-$ (m/z 409.4) is in negative mode. There was no observed change to the ions observed in the negative mode MALDI experiments when the radical scavenger was added. The addition of the radical scavenger significantly decreased the presence of $[M-H]^+$ in the MALDI positive mode data.

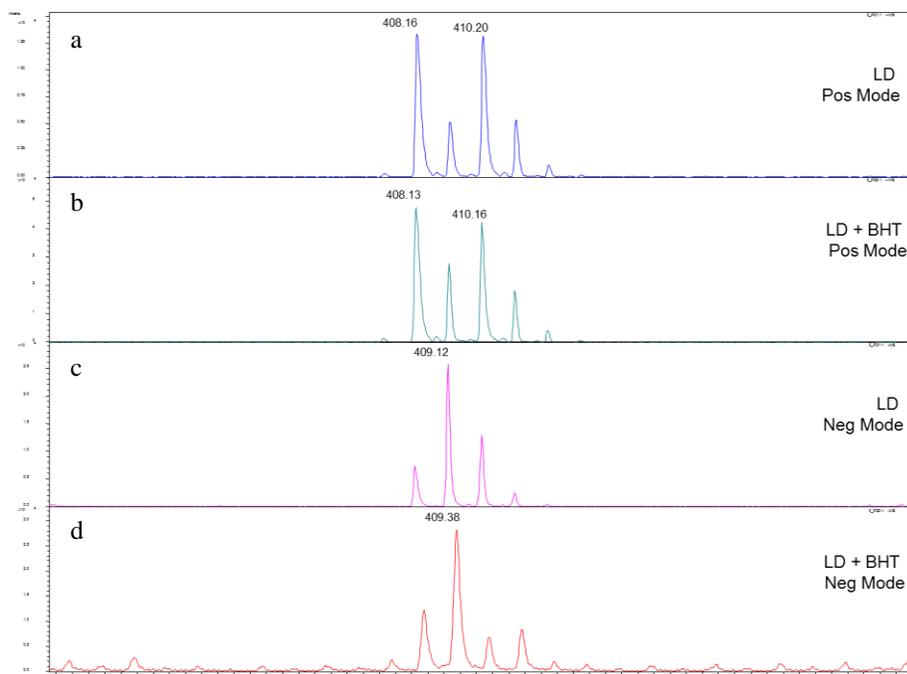


Figure 6.7: LD-TOF analysis of NC 2144 under various conditions: (a) Positive mode with no radical scavenger, (b) positive mode with radical scavenger, (c) negative mode with no radical scavenger, and (d) negative mode with radical scavenger. The peaks of interest are $[M-H]^+$ (m/z 408.1), $[M+H]^+$ (m/z 410.2) and $[M]^-$ (m/z 409.4).

$[M-H]^+$ (m/z 408.1) and $[M+H]^+$ (m/z 410.2) peaks are observed in the LD data obtained in positive mode, contrary to what is observed in the FT-ICR experiments and the $[M]^-$ (m/z 409.4) is found in negative mode. There was no observed change to the ions observed in the negative or positive mode LD experiments when the radical scavenger was added.

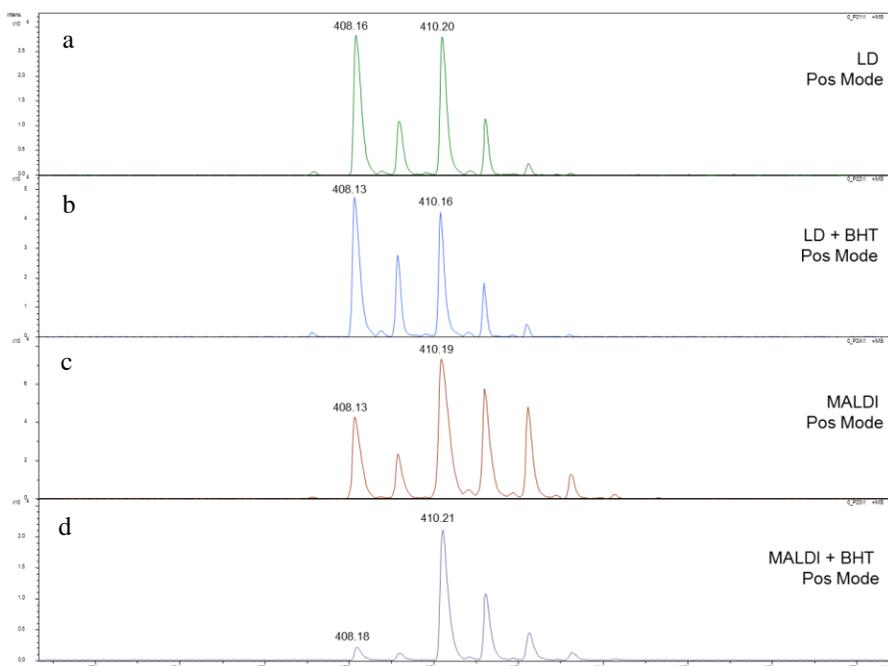


Figure 6.8: Comparison of observed species in the positive mode analysis of NC 2144 using LD and MALDI with the addition of a radical scavenger. : (a) LD positive mode with no radical scavenger, (b) LD positive mode with radical scavenger, (c) MALDI positive mode with no radical scavenger, and (d) MALDI positive mode with radical scavenger The peaks of interest are $[M-H]^+$ (m/z 408.1) and $[M+H]^+$ (m/z 410.2).

The LD spectra, with and without the radical scavenger, and the MALDI spectra, without the radical scavenger, observed the $[M-H]^+$ (m/z 408.1) and $[M+H]^+$ (m/z 410.2) ions. Interestingly the MALDI analysis with the addition of the BHT saw a significantly diminished $[M-H]^+$ peak, which was not observed in the LD analysis under the same conditions.

Many experiments were performed testing the various factors that may contribute to the $[M+H]^+$ ion and $[M-H]^+$ ion generation. Table 6.4 lists a summary of the experiments performed.

Table 6.4 Summary of the various factors tested in the experiments.

<u>Experiment</u>	<u>Observed Ions</u>		
	$[M+H]^+$	$[M-H]^+$	$[M]^-$
<i>Laser Power</i>			
High Laser Intensity	x	x	n/a
Low Laser Intensity	x	x	n/a
<i>Mass Analyzer</i>			
FT-ICR	x	x	n/a
TOF	x	x	x
<i>Matrix</i>			
2,5-Dihydroxybenzoic acid	x	x	n/a
2-Hydroxy-4-Nitrobenzoic acid	x	x	n/a
None (solvent free)	-	x	n/a
<i>Radical Scavenger (TOF)</i>			
Positive mode	x	x	n/a
Negative mode	n/a	n/a	x
<i>Sample and Matrix Solvent</i>			
Protic	x	x	n/a
Aprotic	x	x	n/a
<i>Trifluoroacetic acid in Solvent</i>			
TFA Present	x	x	n/a
TFA Absent	x	x	n/a

Discussion

Direct photoionization is responsible for a significant portion of the ion formation of $[M-H]^+$, as it was formed in the absence of ionization enhancers, such as matrix and protic solvents during LD analysis. It may also be responsible for enhancing the formation of $[M+H]^+$, but the experiments indicate that the presence of matrix molecules open other ionization pathways that may be more favorable. Photoionization requires the compounds of low ionization energy to absorb photon energy, as shown in the equation

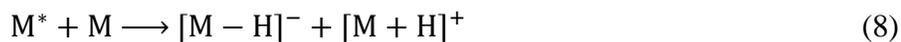
below in reaction 4, where M is the analyte molecule, $h\nu$ the photon, and $M^{+\cdot}$ the radical cation product:



In the presence of solvent, the $M^{+\cdot}$ can undergo hydrogen atom transfer with the solvent molecule (S) (reaction 5), or in the absence of solvent, another neutral analyte molecule (reaction 6):



All of these reactions point to the importance of radicals in the formation of ions via photoionization, but these species were not observed in the MALDI and LD analyses of the curcumin analogues. This absence could be explained by several situations. The first is that they were formed, but the resolving power and mass accuracy of the instrument did not allow these peaks to be distinguished from the isotopic envelope peaks. It is also possible that radical species were converted to more stable reaction products, such as $[M+H]^{+}$. Another likely scenario would be that the laser photon energy was not sufficient to generate the radicals in the first place, yielding instead the excited state analyte molecule (M^{*}) (reaction 7), which could undergo photon transfer to form the protonated ($[M+H]^{+}$) and deprotonated ($[M-H]^{-}$) molecular ions (reaction 8).



The direct photoionization of the curcumin molecules is similar to the direct ionization of matrix molecules, since both share a general core structure, being composed of highly aromatic substituted benzyl rings. Matrices are directly ionized by the laser producing radicals or excited state matrix molecules, which go on to ionize the analyte in the gas phase.²⁹ It is not unreasonable to propose that the curcumin molecules are behaving the same way, based on the LD experiments performed.

Prior to the completion of all of the experiments, there were three general mechanisms proposed for the formation of the $[M-H]^+$ ions observed: loss of molecular hydrogen from the $[M+H]^+$ ion, hydrogen atom transfer from the $M^{+\bullet}$ created during photoionization, and hydride abstraction from a neutral analyte molecule. The evidence or lack thereof will be discussed in more detail for each of these below.

Formation of $[M-H]^+$ from the $[M+H]^+$

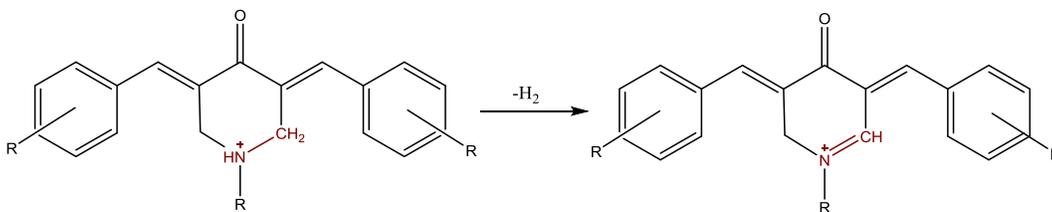


Figure 6.9 Proposed formation of $[M-H]^+$ from the $[M+H]^+$ by loss of a hydrogen molecule.

The $[M+H]^+$ ion is observed by LD-TOF analysis, but not by LD-FT-ICR, while the $[M-H]^+$ is observed by both methods. This suggests that $[M-H]^+$ is formed from $[M+H]^+$, with a reaction time comparable to the TOF analysis (100 μ s), but much shorter than the time required for FT-ICR analysis (1 s). Furthermore, MALDI using the cool

matrix 2,5-DHB, produces $[M+H]^+$ that is less susceptible to decomposition to $[M-H]^+$, resulting in $[M+H]^+$ present in the MALDI-FT-ICR and MALDI-TOF mass spectra.

Formation of $[M-H]^+$ from the unstable radical form $[M]^{+\bullet}$

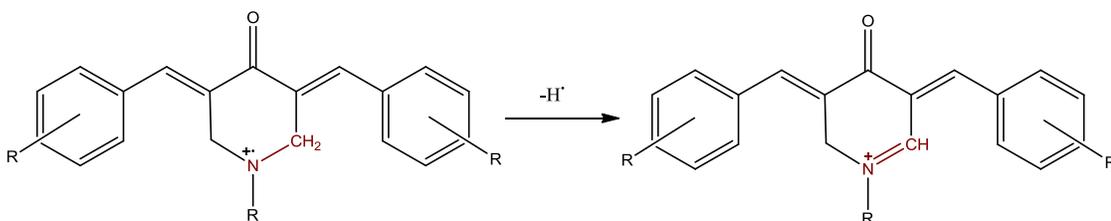


Figure 6.10 Proposed formation of $[M-H]^+$ from the unstable radical form $[M]^{+\bullet}$.

From the MALDI TOF experiments, a significant loss in signal intensity of the $[M-H]^+$ was observed when the radical scavenger was added. This suggests that a radical species may contribute as one origin of the $[M-H]^+$ species.

Formation of $[M-H]^+$ by direct loss of negative hydrogen ion from a neutral curcumin analogue molecule through a redox reaction.

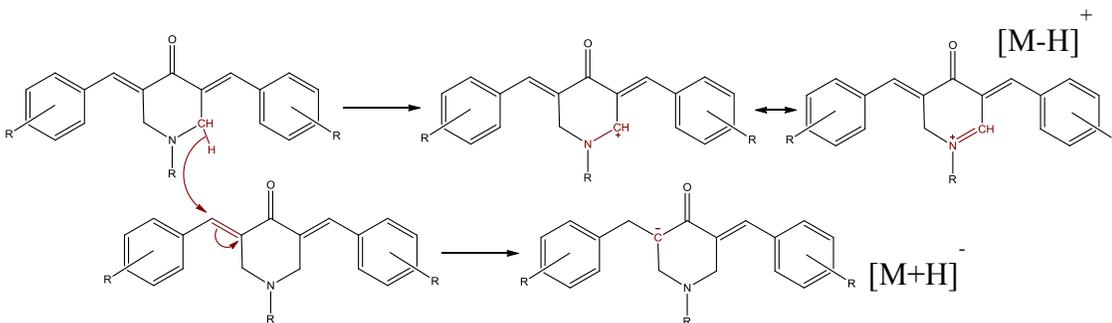


Figure 6.11 Proposed formation of $[M-H]^+$ by direct loss of negative hydrogen ion from a neutral curcumin analogue molecule through a redox reaction.

This mechanism does not seem to be supported by the experiments shown here. There is not a significant or observable $[M+H]^-$ peak observed in the LD spectra from either the TOF or FT-ICR analyses.

Conclusions

From all experiments shown, a general mechanism for the formation of the $[M-H]^+$ in curcumin analogues can be proposed. With LD, M^+ is formed by photoionization, which undergoes hydrogen transfer from another analyte molecule to form $[M+H]^+$, and the subsequent loss of H_2 yields the $[M-H]^+$. During MALDI, there is a proton transfer from the matrix to the analyte, yielding the $[M+H]^+$, and the subsequent loss of H_2 yields the $[M-H]^+$. The $[M+H]^+$ ion is more stable in the MALDI FT-ICR experiments than the LD FT-ICR experiments. This suggests that ion transfer via the matrix is less exothermic than the LD process described above. The $[M]^-$ ion is very stable, as it is observed in both the LD and MALDI negative mode analyses. The addition of a radical scavenger did not deplete this peak.

Analysis of more curcumin analogues would be beneficial, since the ones studied here belong to various structural families: phosphoramidates (NC2311), secondary amines (NC 2453), and amides (NC 2144, NC 2138, NC 2094). Although it is possible that the same mechanism(s) is responsible for the formation of $[M-H]^+$ in these molecules, each of the structural classes may be more susceptible to ion generation via a specific mechanism based on their individual chemistries.

REFERENCES

1. Karas, M.; Bachmann, D.; Hillenkamp, F., Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Anal. Chem.* **1985**, *57* (14), 2935-2939.
2. Hillenkamp, F.; Benninghoven, A.; Colton, R.; Simons, D.; Werner, H., Secondary Ion Mass Spectrometry SIMS V. Springer, Berlin: 1986.
3. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **1988**, *60* (20), 2299-2301.
4. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T., Protein and polymer analyses up to m/z 100 000 by laser ionization time of flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **1988**, *2* (8), 151-153.
5. Strimpakos, A. S.; Sharma, R. A., Curcumin: preventive and therapeutic properties in laboratory studies and clinical trials. *Antioxid. Redox Signaling* **2008**, *10* (3), 511-546.
6. Das, S.; Das, U.; Selvakumar, P.; Sharma, R. K.; Balzarini, J.; De Clercq, E.; Molnár, J.; Serly, J.; Baráth, Z.; Schatte, G., 3, 5-Bis (benzylidene)-4-oxo-1-phosphonopiperidines and Related Diethyl Esters: Potent Cytotoxins with Multi-Drug-Resistance Reverting Properties. *ChemMedChem* **2009**, *4* (11), 1831-1840.
7. Field, F. H., Chemical ionization mass spectrometry. *Acc. Chem. Res.* **1968**, *1* (2), 42-49.
8. Munson, M.; Field, F.-H., Chemical ionization mass spectrometry. I. General introduction. *JACS* **1966**, *88* (12), 2621-2630.
9. Carroll, D.; Dzidic, I.; Stillwell, R.; Haegele, K.; Horning, E., Atmospheric pressure ionization mass spectrometry. Corona discharge ion source for use in a liquid chromatograph-mass spectrometer-computer analytical system. *Anal. Chem.* **1975**, *47* (14), 2369-2373.
10. Chai, Y.; Sun, H.; Wan, J.; Pan, Y.; Sun, C., Hydride abstraction in positive-ion electrospray interface: oxidation of 1, 4-dihydropyridines in electrospray ionization mass spectrometry. *Analyst* **2011**, *136* (22), 4667-4669.
11. Orelli, L. R.; García, M. B.; Perillo, I. A.; Tonidandel, L.; Traldi, P., A comparison of the electron ionization and electrospray behaviour of some N, N'-disubstituted hexahydropyrimidines. *Rapid Commun. Mass Spectrom.* **2006**, *20* (5), 823-828.
12. Nuutinen, J. M.; Purmonen, M.; Ratilainen, J.; Rissanen, K.; Vainiotalo, P., Mass spectrometric studies on pyridine-piperazine-containing ligands and their

- complexes with transition metals formed in solution. *Rapid Commun. Mass Spectrom.* **2001**, *15* (15), 1374-1381.
13. Yang, H.-J.; Lee, A.; Lee, M.-K.; Kim, W.; Kim, J., Detection of small neutral carbohydrates using various supporting materials in laser desorption/ionization mass spectrometry. *Bull. Korean Chem. Soc* **2010**, *31* (1), 35.
 14. Kang, C.; Zhou, Y.; Du, Z.; Bian, Z.; Wang, J.; Qiu, X.; Gao, L.; Sun, Y., Dehydrogenation and dehalogenation of amines in MALDI-TOF MS investigated by isotopic labeling. *J. Mass Spectrom.* **2013**, *48* (12), 1318-1324.
 15. Lou, X.; Spiering, A.; de Waal, B. F.; van Dongen, J. L.; Vekemans, J. A.; Meijer, E., Dehydrogenation of tertiary amines in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom.* **2008**, *43* (8), 1110-1122.
 16. Cox, F. J.; Johnston, M. V.; Dasgupta, A., Characterization and relative ionization efficiencies of end-functionalized polystyrenes by matrix-assisted laser desorption/ionization mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **2003**, *14* (6), 648-657.
 17. Kauppila, T.; Nikkola, T.; Ketola, R.; Kostianen, R., Atmospheric pressure photoionization-mass spectrometry and atmospheric pressure chemical ionization-mass spectrometry of neurotransmitters. *J. Mass Spectrom.* **2006**, *41* (6), 781-789.
 18. Suni, N. M.; Aalto, H.; Kauppila, T. J.; Kotiaho, T.; Kostianen, R., Analysis of lipids with desorption atmospheric pressure photoionization-mass spectrometry (DAPPI-MS) and desorption electrospray ionization-mass spectrometry (DESI-MS). *J. Mass Spectrom.* **2012**, *47* (5), 611-619.
 19. Cody, R. B., Observation of molecular ions and analysis of nonpolar compounds with the direct analysis in real time ion source. *Anal. Chem.* **2008**, *81* (3), 1101-1107.
 20. Jaskolla, T. W.; Karas, M., Compelling evidence for lucky survivor and gas phase protonation: the unified MALDI analyte protonation mechanism. *J. Am. Soc. Mass. Spectrom.* **2011**, *22* (6), 976-988.
 21. Karas, M.; Glückmann, M.; Schäfer, J., Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J. Mass Spectrom.* **2000**, *35* (1), 1-12.
 22. Karbach, V.; Knochenmuss, R., Do single matrix molecules generate primary ions in ultraviolet matrix-assisted laser desorption/ionization. *Rapid Commun. Mass Spectrom.* **1998**, *12* (14), 968-974.

23. Niu, S.; Zhang, W.; Chait, B. T., Direct comparison of infrared and ultraviolet wavelength matrix-assisted laser desorption/ionization mass spectrometry of proteins. *J. Am. Soc. Mass. Spectrom.* **1998**, *9* (1), 1-7.
24. Chen, X.; Carroll, J. A.; Beavis, R. C., Near-ultraviolet-induced matrix-assisted laser desorption/ionization as a function of wavelength. *J. Am. Soc. Mass. Spectrom.* **1998**, *9* (9), 885-891.
25. Talroze, V.; Jacob, R.; Burlingame, A.; Baldwin, M., Insight into the MALDI mechanism: matrix decomposition and pneumatic assistance in plume formation *Adv. Mass Spectrom* **2001**, *15*, 481-482.
26. Knochenmuss, R., Ion formation mechanisms in UV-MALDI. *Analyst* **2006**, *131* (9), 966-986.
27. Karas, M.; Bachmann, D.; Bahr, U. e.; Hillenkamp, F., Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int. J. Mass Spectrom. Ion Processes* **1987**, *78*, 53-68.
28. Ehring, H.; Karas, M.; Hillenkamp, F., Role of photoionization and photochemistry in ionization processes of organic molecules and relevance for matrix-assisted laser desorption ionization mass spectrometry. *Org. Mass Spectrom* **1992**, *27* (4), 472-480.
29. Zenobi, R.; Knochenmuss, R., Ion formation in MALDI mass spectrometry. *Mass Spectrom. Rev.* **1998**, *17* (5), 337-366.
30. Hillenkamp, F.; Peter-Katalinic, J., *MALDI MS*. John Wiley & Sons: 2007.

CHAPTER 7

CONCLUSIONS

Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR) is a very powerful analytical technique that is well suited for the analysis of many types of biomolecules. The increased use of mass spectrometry as a technique in the analysis of biomolecules over the last few decades is due almost exclusively to the advent of soft ionization techniques, primarily electrospray ionization (ESI)¹ and matrix assisted laser desorption ionization (MALDI).^{2,3} These methods allow large, nonvolatile, thermally labile analytes, such as peptides and proteins, to be ionized without inducing significant fragmentation.

FT-ICR has the ability to accurately measure the mass to charge ratio, with high sensitivity, and with the ability to be interfaced with separation techniques, such as high performance liquid chromatography (HPLC) and high field asymmetric ion mobility spectrometry (FAIMS).⁴ This dissertation presented four experimental methods using the strengths of FT-ICR.

The first experiment, presented in Chapter 3, evaluates the use of accurate mass measurement and nitrogen stoichiometry to characterize the proteome of *Acinetobacter baylyi* ADP1 through a shotgun proteomic approach. The high sensitivity, high resolving power, and ability to measure mass to charge with very high accuracy allowed FT-ICR to

identify 138 proteins at 5 ppm mass tolerance and 112 proteins at 10 ppm mass tolerance, when using accurate mass measurement and nitrogen stoichiometry were used.

Accurate mass measurement is not only useful in the identification of proteins, but also for small organic molecules, such as metabolites, as presented in Chapter 4. The use of FT-ICR mass spectrometry was beneficial as the elemental composition of analytes can be determined, which allowed for identification of the metabolic acyl-CoA intermediates present during the metabolism of dimethylsulphoniopropionate in marine bacteria. By comparing the accurate measured masses of the unlabeled and isotopically labeled metabolites, the identity of the intermediate at each step of the pathway was discerned.

The next analysis, described in Chapter 5, investigates the robustness and effectiveness of using chemical crosslinking and mass defect labeling as a means of detecting and evaluating reaction products in individual proteins and peptides. The mass defect labeled chemical crosslinker allows for easy determination of chemical crosslinking reaction products due to the presence of the significant mass defect present. The mass defect crosslinker increases the speed of analysis of crosslinked peaks by focusing peak analysis on the peaks demonstrating the characteristic mass shift. By pairing the mass defect labeled crosslinker with FAIMS, the decreased sample complexity allowed easier isolation of peaks of interest that had been identified as reaction products. The increase in the intensity of low abundance ions of interest would lead to higher efficiency in MS³ analysis of reaction products.

The final topic, presented in Chapter 6, examines the various factors responsible for the generation of an uncommon ion, [M+H]⁺ in a set of synthesized curcumin

analogues, and how these variables can help us elucidate the mechanisms at play. Although there are likely many mechanisms involved at varying extents, using a combination of MALDI and LD analysis, a general mechanism for the formation of the $[M-H]^+$ in curcumin analogues can be proposed. With LD, M^+ is formed by photoionization, which undergoes hydrogen transfer from another analyte molecule to form $[M+H]^+$, and the subsequent loss of H_2 yields the $[M-H]^+$. During MALDI, there is a proton transfer from the matrix to the analyte, yielding the $[M+H]^+$, and the subsequent loss of H_2 yields the $[M-H]^+$. Further studies regarding ionization mechanisms are continuously ongoing within the larger research field, and one day, the exact mechanism(s) will be fully revealed.

REFERENCES

1. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246* (4926), 64-71.
2. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T., Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **1988**, *2* (8), 151-153.
3. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry* **1988**, *60* (20), 2299-2301.
4. Guevremont, R., High-field asymmetric waveform ion mobility spectrometry (FAIMS). *Canadian journal of analytical sciences and spectroscopy* **2004**, *49* (3), 105-113.

APPENDICES

APPENDIX A

SUPPLEMENTAL DATA FOR CHAPTER 5

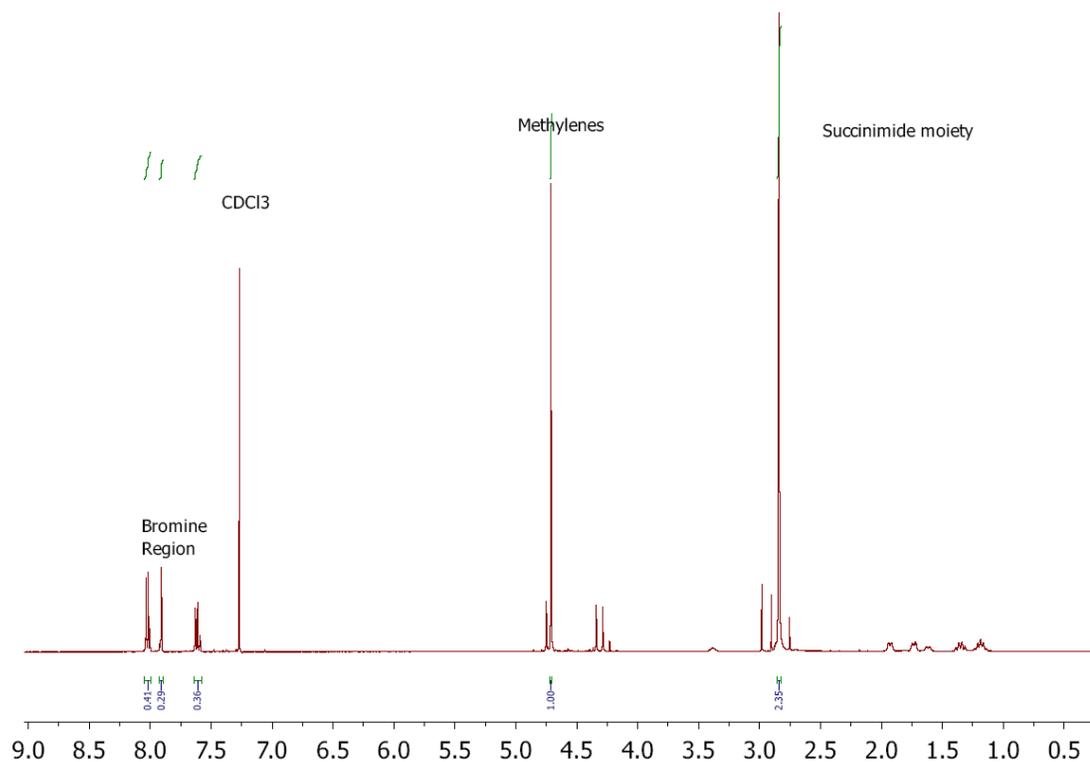


Figure A.1 Proton NMR of the mass defect labeled crosslinker bis(2,5-dioxopyrrolidin-1-yl) 2,2'-(((2,4-dibromophenyl)sulfonyl)azanediyl) diacetate (DiBBAIS). The lack of peaks above 9 ppm indicates that the crosslinker is not hydrolyzed, and is fully functional, with both N-succinimide moieties intact.

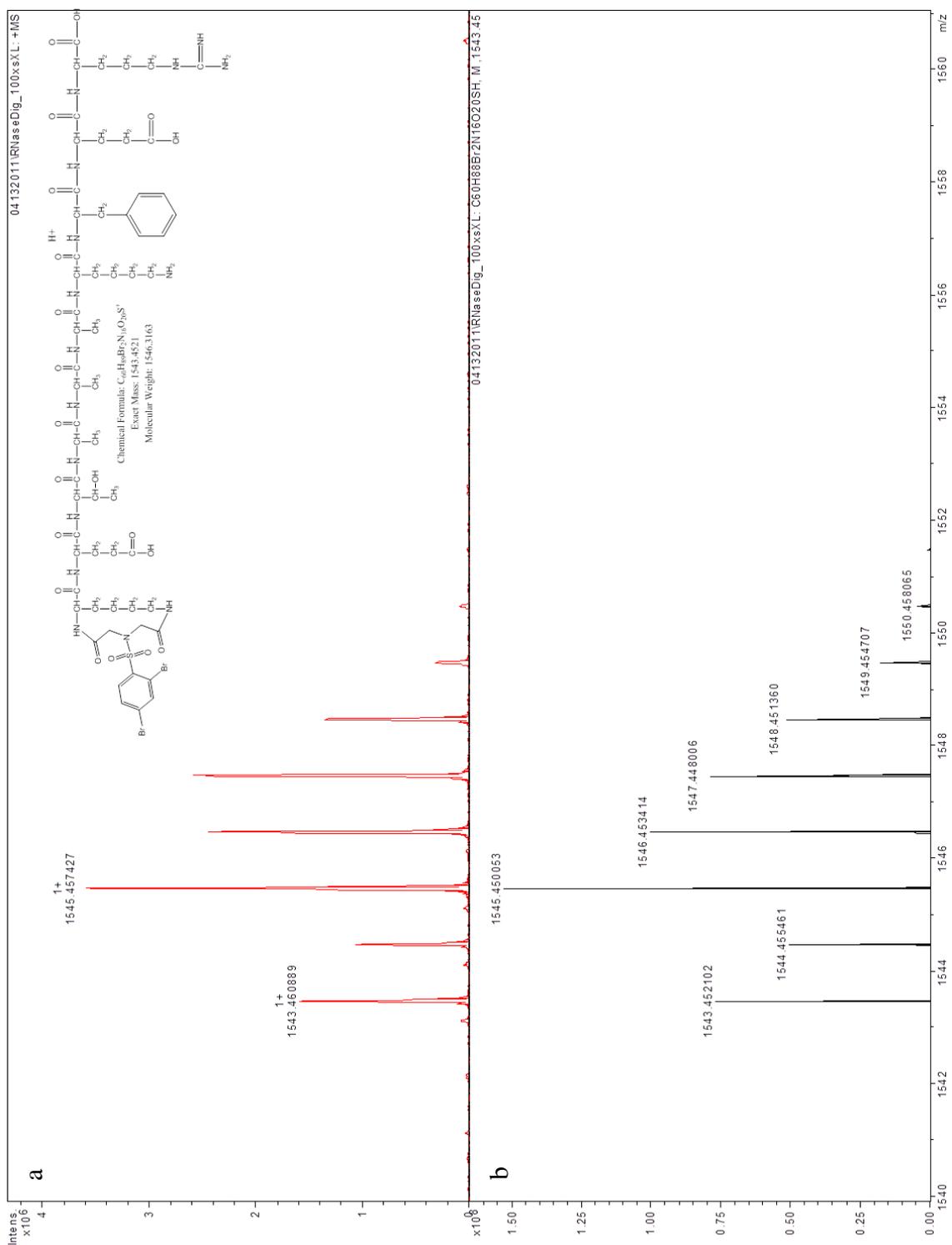


Figure A.2 Mass spectrum of m/z 1543.460⁺ from RNase S reacted with 100 times molar excess DiBBSIAS. The proposed chemical structure corresponding to the observed mass is also shown. (a) Experimental isotopic distribution observed. (b) Theoretical isotopic distribution.

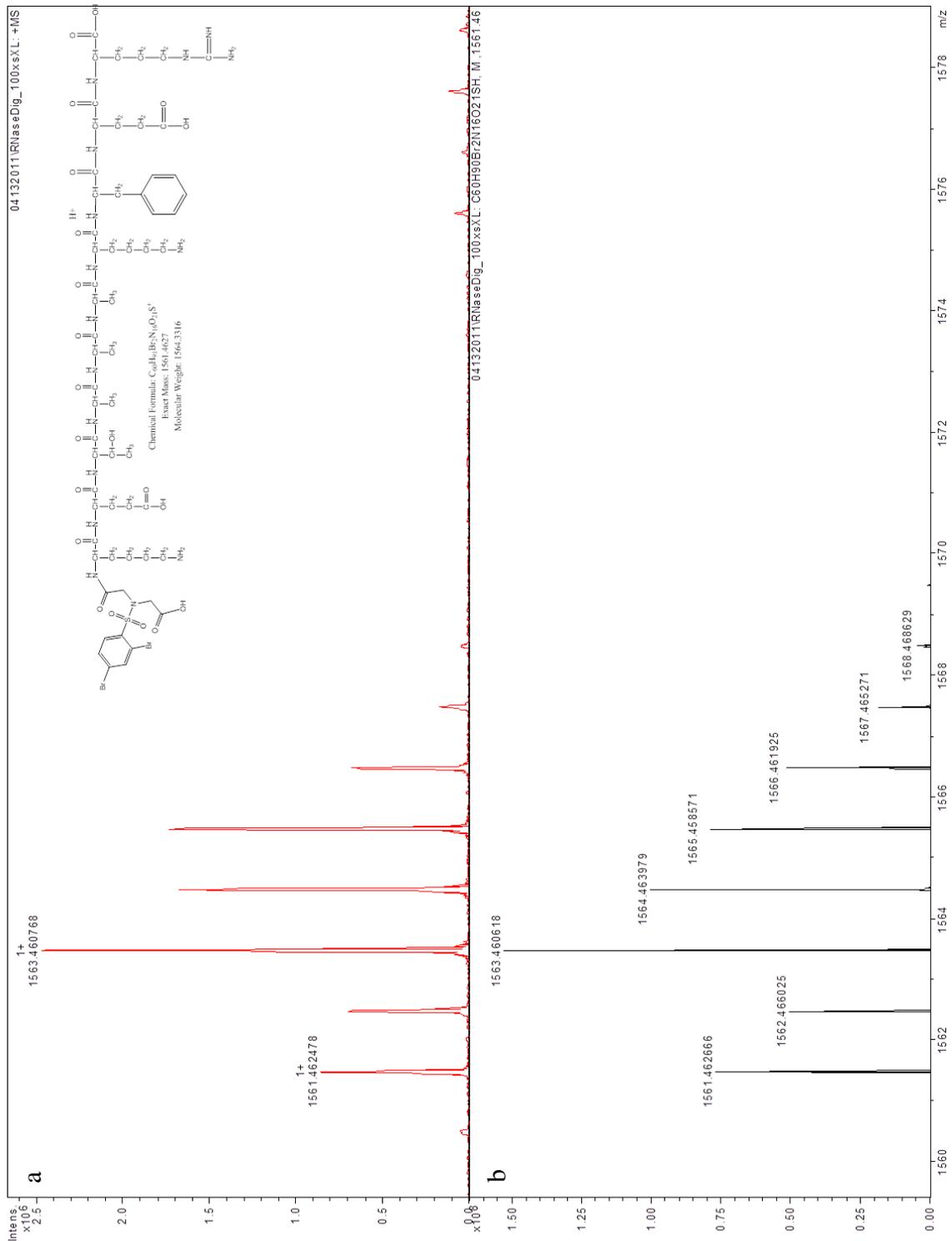


Figure A.3 Mass spectrum of m/z 781.231²⁺ from RNase S reacted with 100 times molar excess DiBBSIAS. The proposed chemical structure corresponding to the observed mass is also shown. (a) Experimental isotopic distribution observed. (b) Theoretical isotopic distribution.

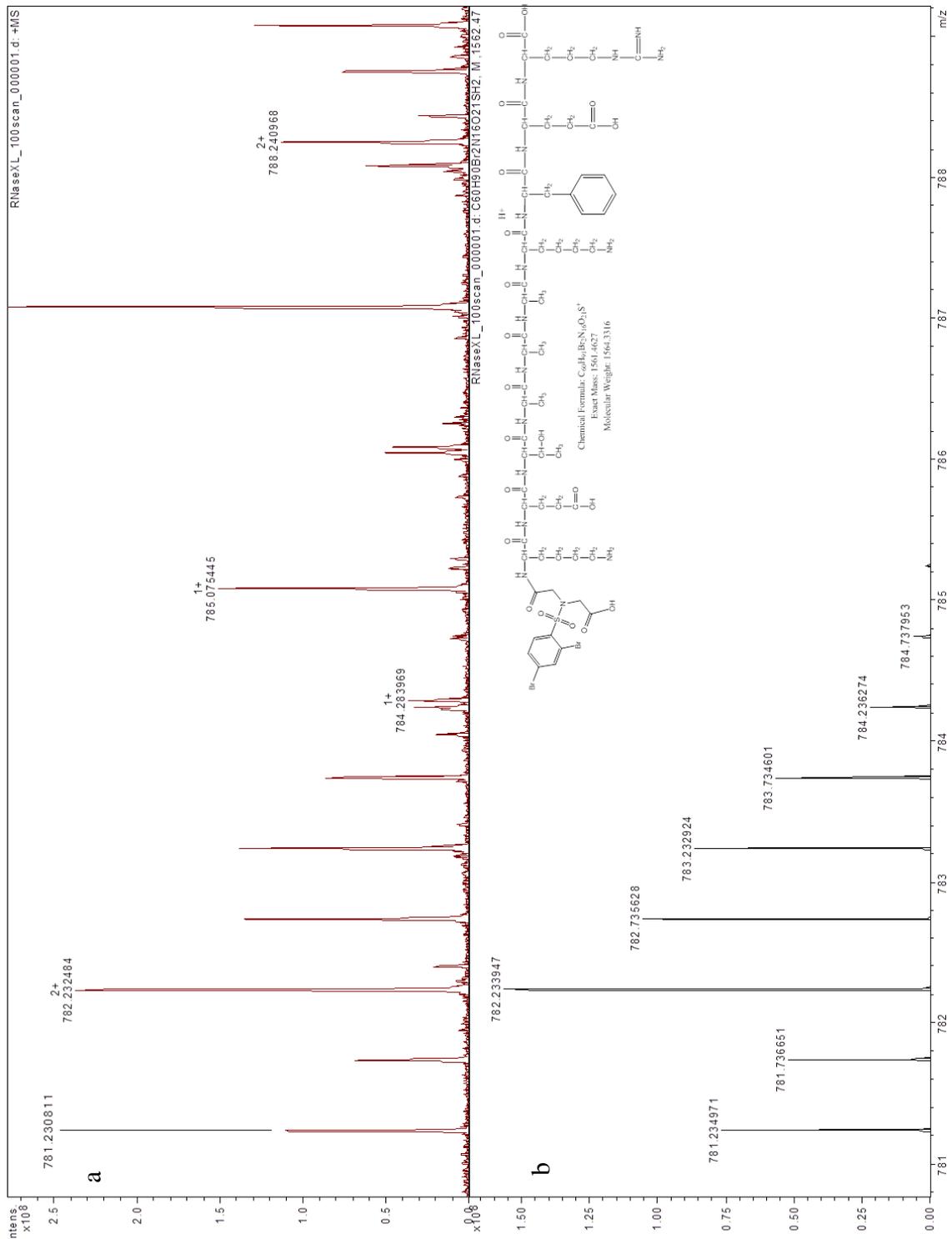


Figure A.4 Mass spectrum of m/z 781.231²⁺ from RNase S reacted with 100 times molar excess DiBBSIAS. The proposed chemical structure corresponding to the observed mass is also shown. (a) Experimental isotopic distribution observed. (b) Theoretical isotopic distribution.

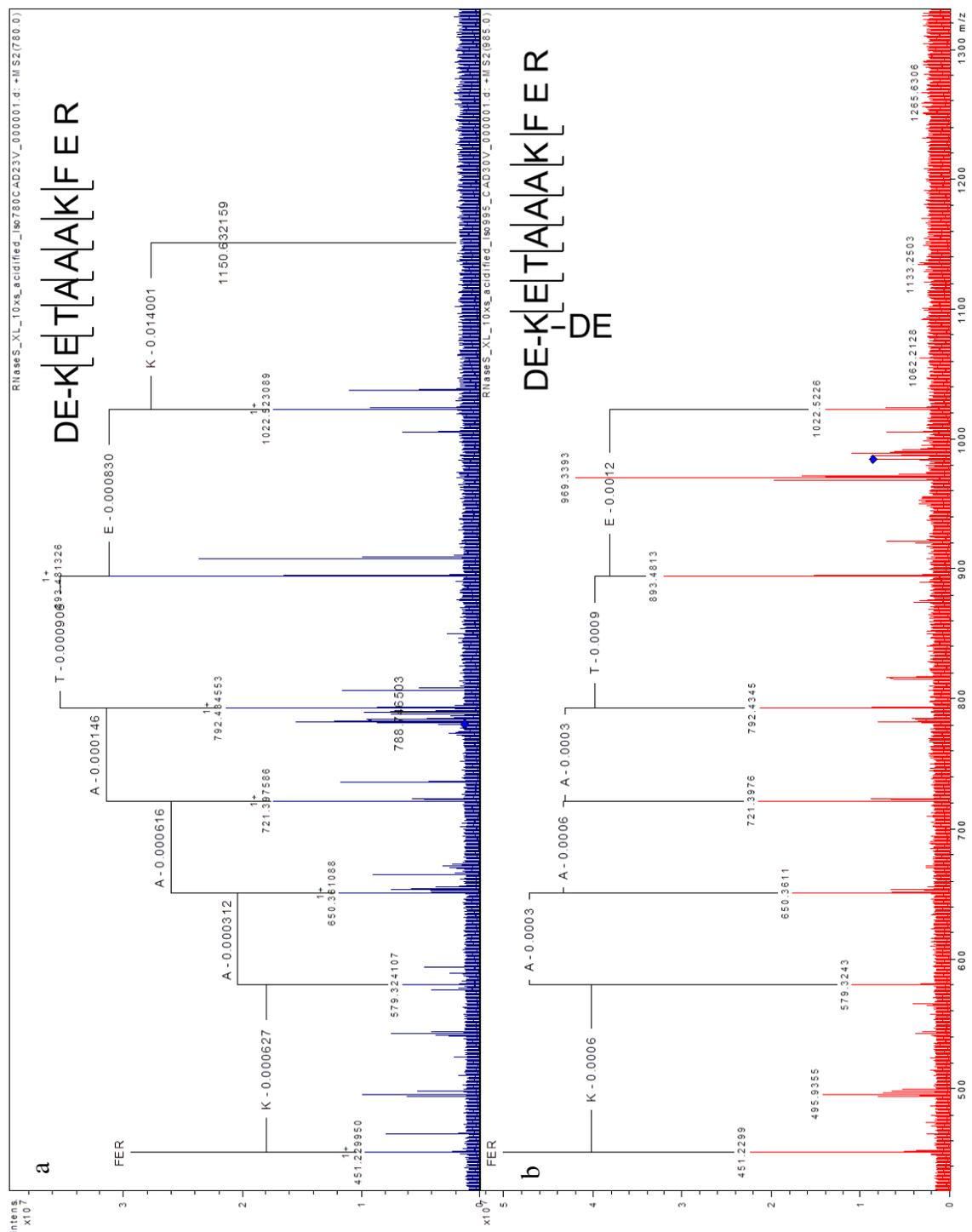


Figure A.5 CAD fragmentation spectra sequencing S-peptide. (a) m/z 782.233²⁺ with one hydrolyzed crosslinker and (b) m/z 988.652²⁺ two hydrolyzed crosslinkers on the N-terminal lysine.

APPENDIX B

SUPPLEMENTAL DATA FOR CHAPTER 6

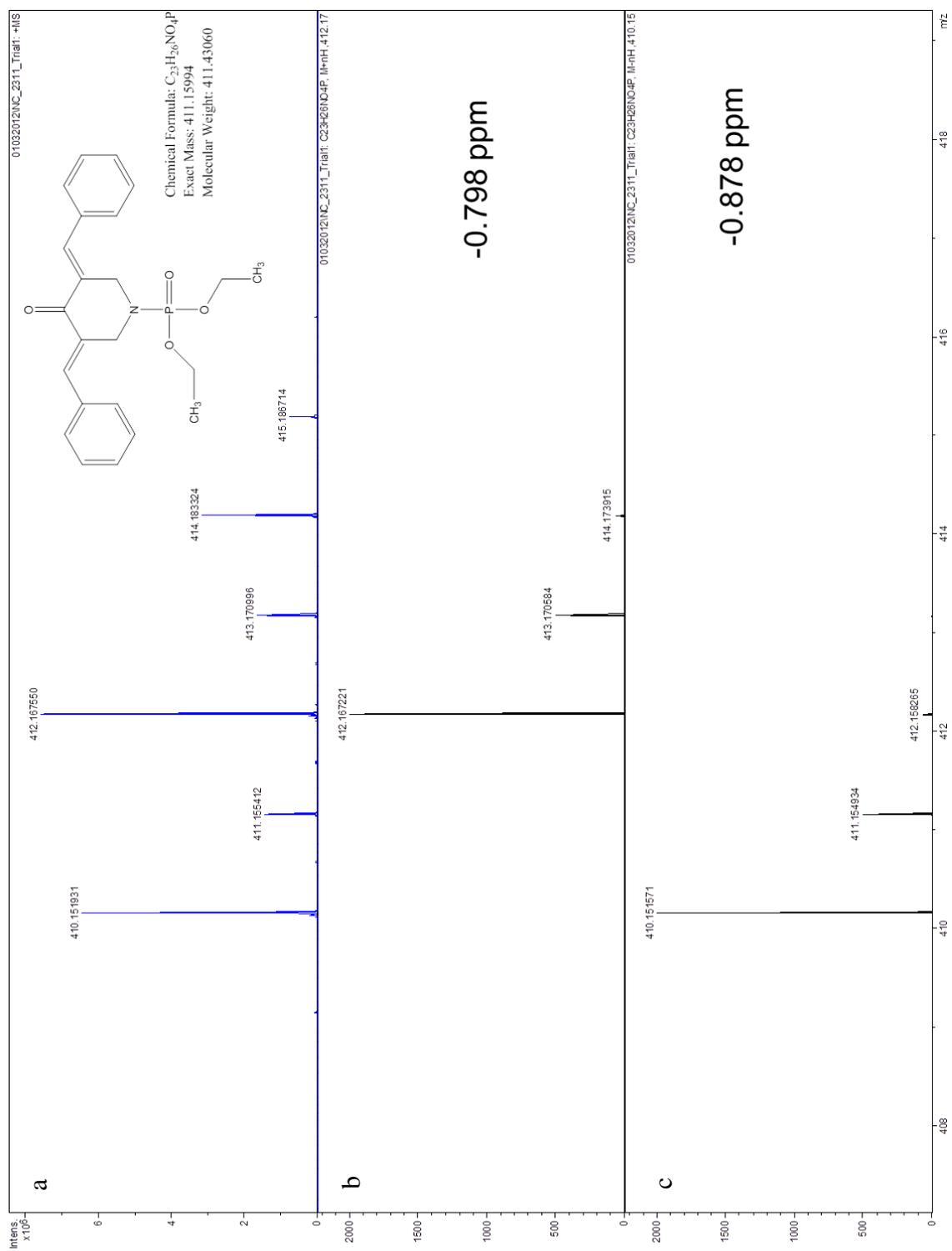


Figure B.1 Mass spectra of curcumin analogue sample NC 2311. The mass accuracy is listed in b and c for each ion. (a) MALDI-FT-ICR spectrum containing $[M+H]^+$ (m/z 412.1676) and $[M-H]^+$ (m/z 410.1519). (b) Theoretical mass spectrum for the $[M-H]^+$ ion. (c) Theoretical mass spectrum for the $[M+H]^+$ ion.

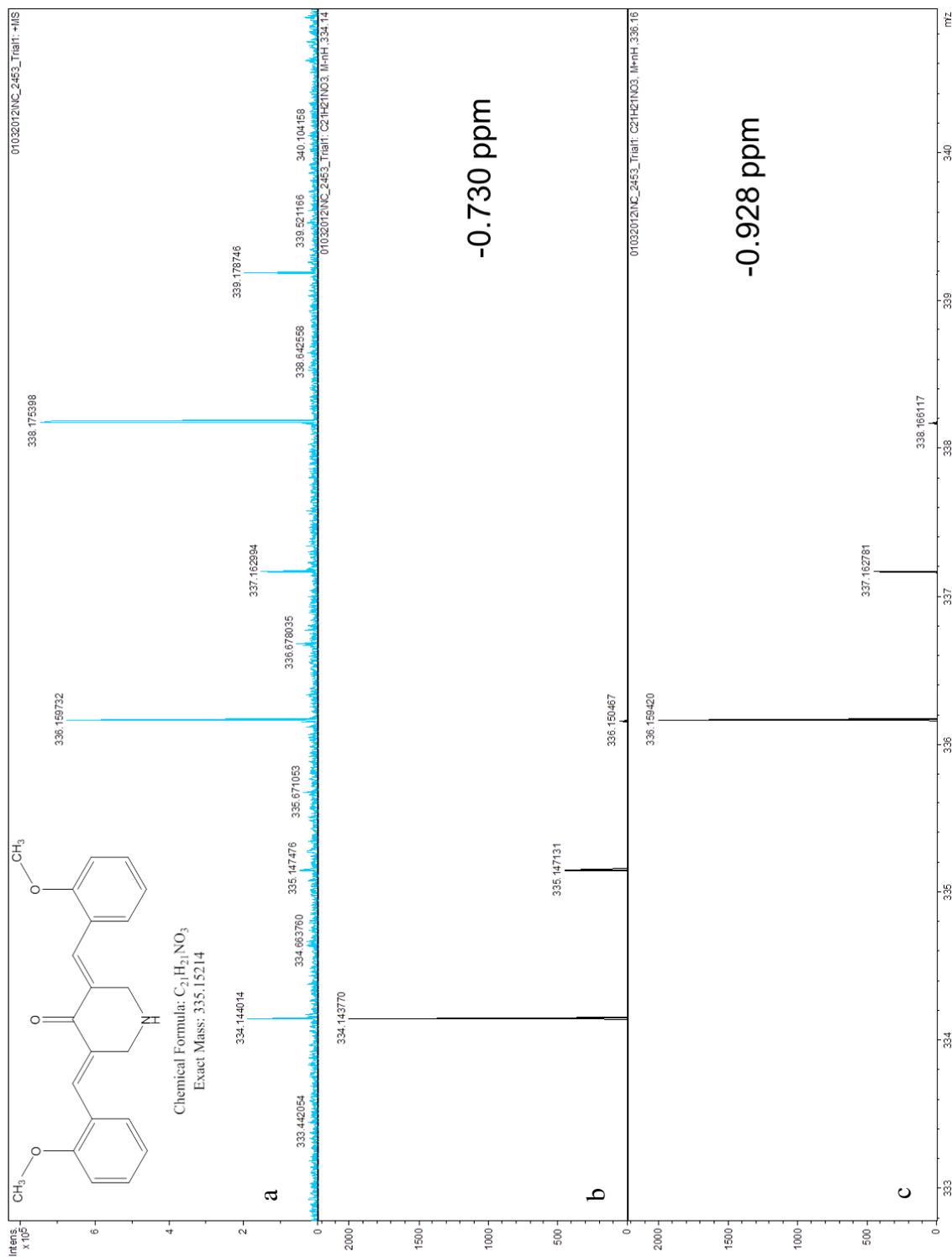


Figure B.2 Mass spectra of curcumin analogue sample NC 2453. The mass accuracy is listed in b and c for each ion. (a) MALDI-FT-ICR spectrum containing $[M+H]^+$ (m/z 336.1597) and $[M-H]^+$ (m/z 334.1440). (b) Theoretical mass spectrum for the $[M-H]^+$ ion. (c) Theoretical mass spectrum for the $[M+H]^+$ ion.

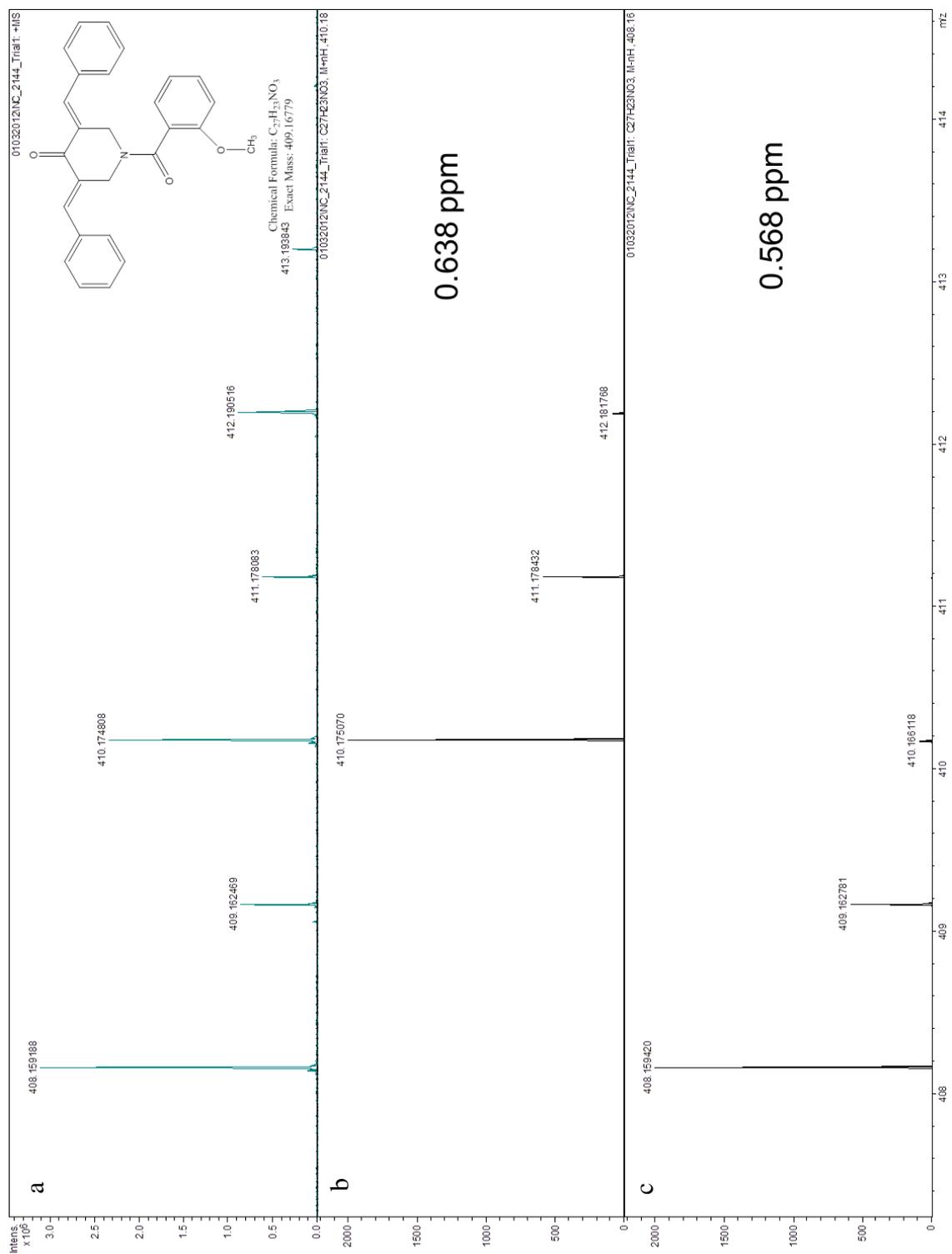


Figure B.3 Mass spectra of curcumin analogue sample NC 2144. The mass accuracy is listed in b and c for each ion. (a) MALDI-FT-ICR spectrum containing $[M+H]^+$ (m/z 410.1748) and $[M-H]^+$ (m/z 408.1592). (b) Theoretical mass spectrum for the $[M-H]^+$ ion. (c) Theoretical mass spectrum for the $[M+H]^+$ ion.

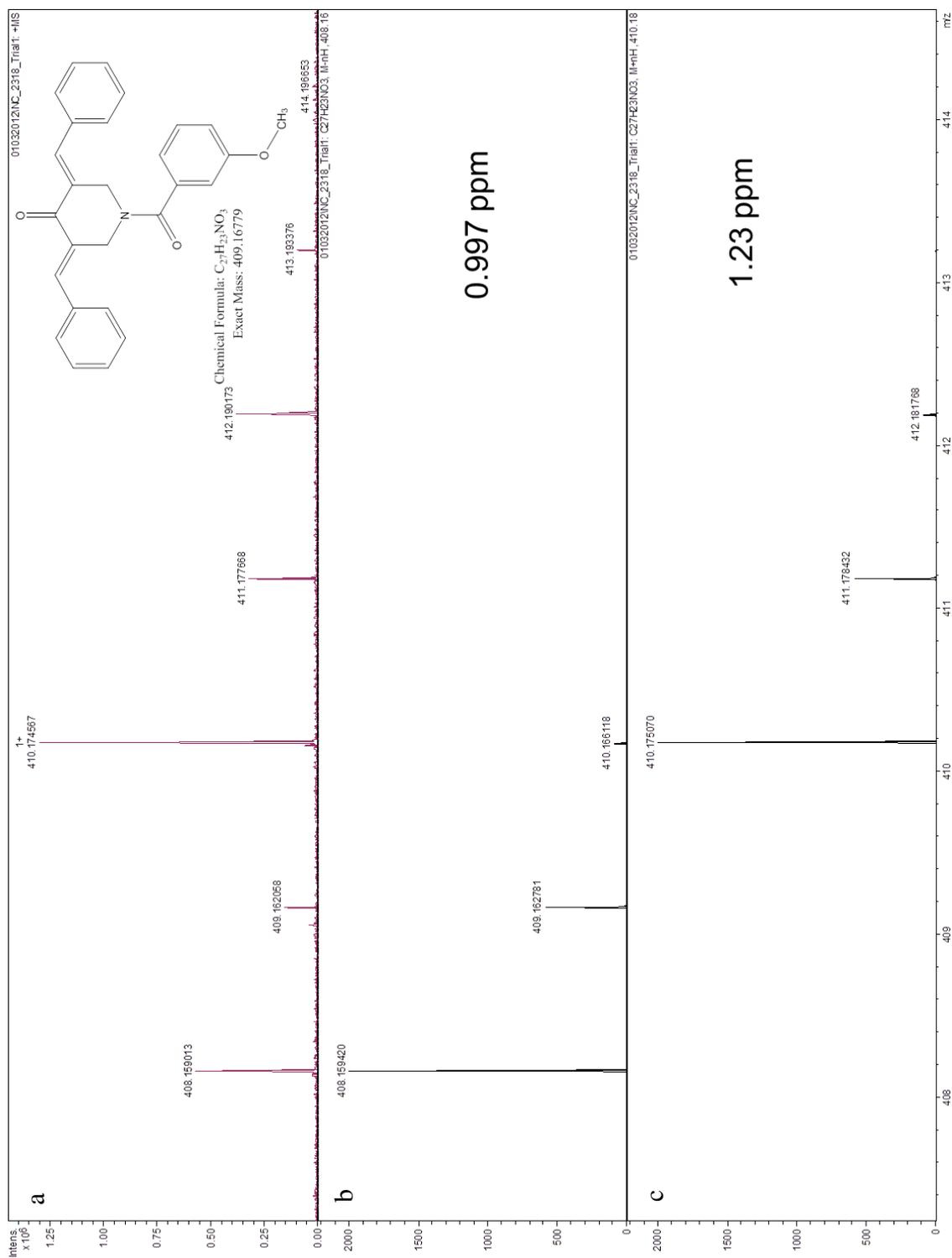


Figure B.4 Mass spectra of curcumin analogue sample NC 2138. The mass accuracy is listed in b and c for each ion. (a) MALDI-FT-ICR spectrum containing $[M+H]^+$ (m/z 410.1746) and $[M-H]^+$ (m/z 408.1591). (b) Theoretical mass spectrum for the $[M-H]^+$ ion. (c) Theoretical mass spectrum for the $[M+H]^+$ ion.

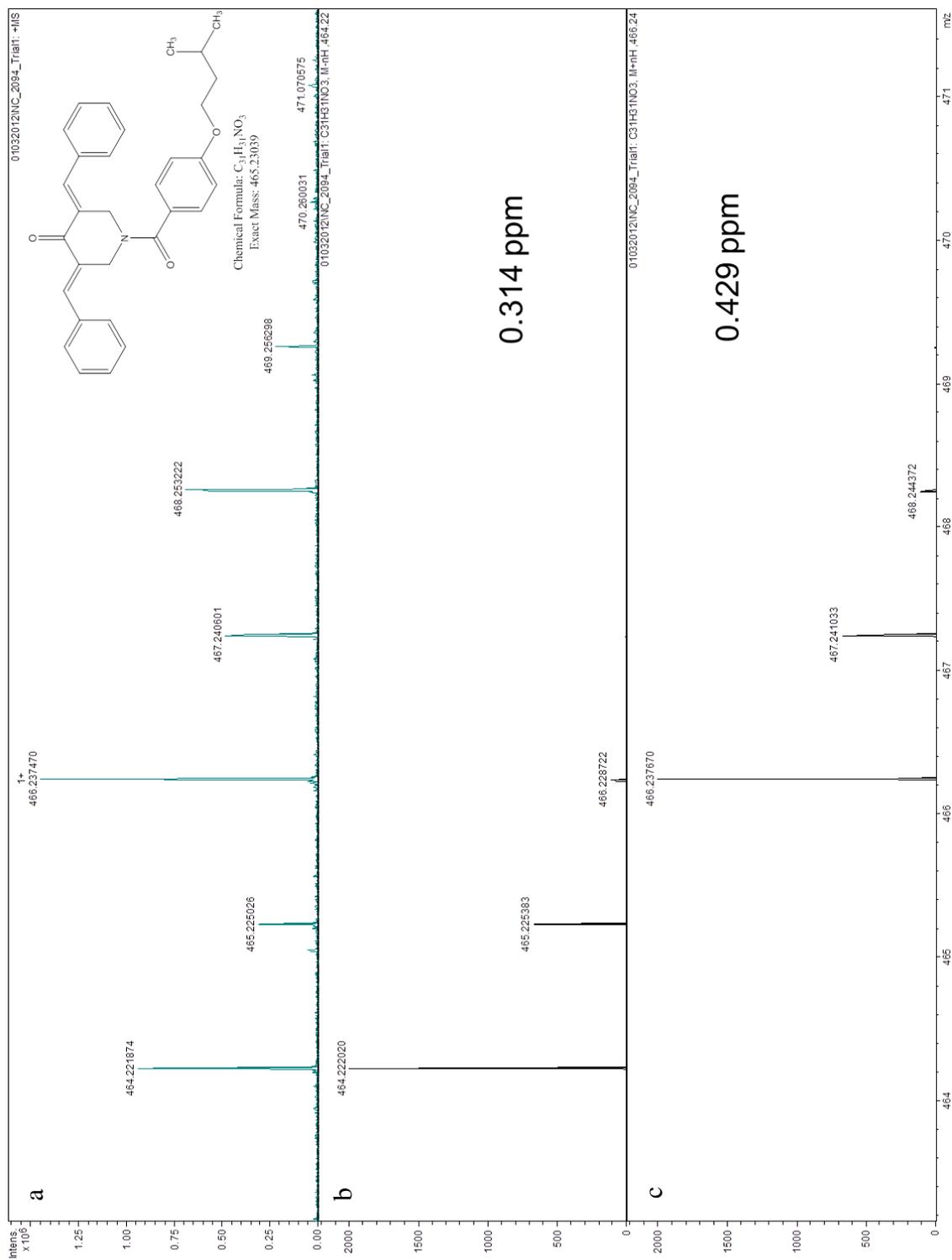


Figure B.5 Mass spectra of curcumin analogue sample NC 2094. The mass accuracy is listed in b and c for each ion. (a) MALDI-FT-ICR spectrum containing $[M+H]^+$ (m/z 466.2375) and $[M-H]^+$ (m/z 464.2219). (b) Theoretical mass spectrum for the $[M-H]^+$ ion. (c) Theoretical mass spectrum for the $[M+H]^+$ ion.

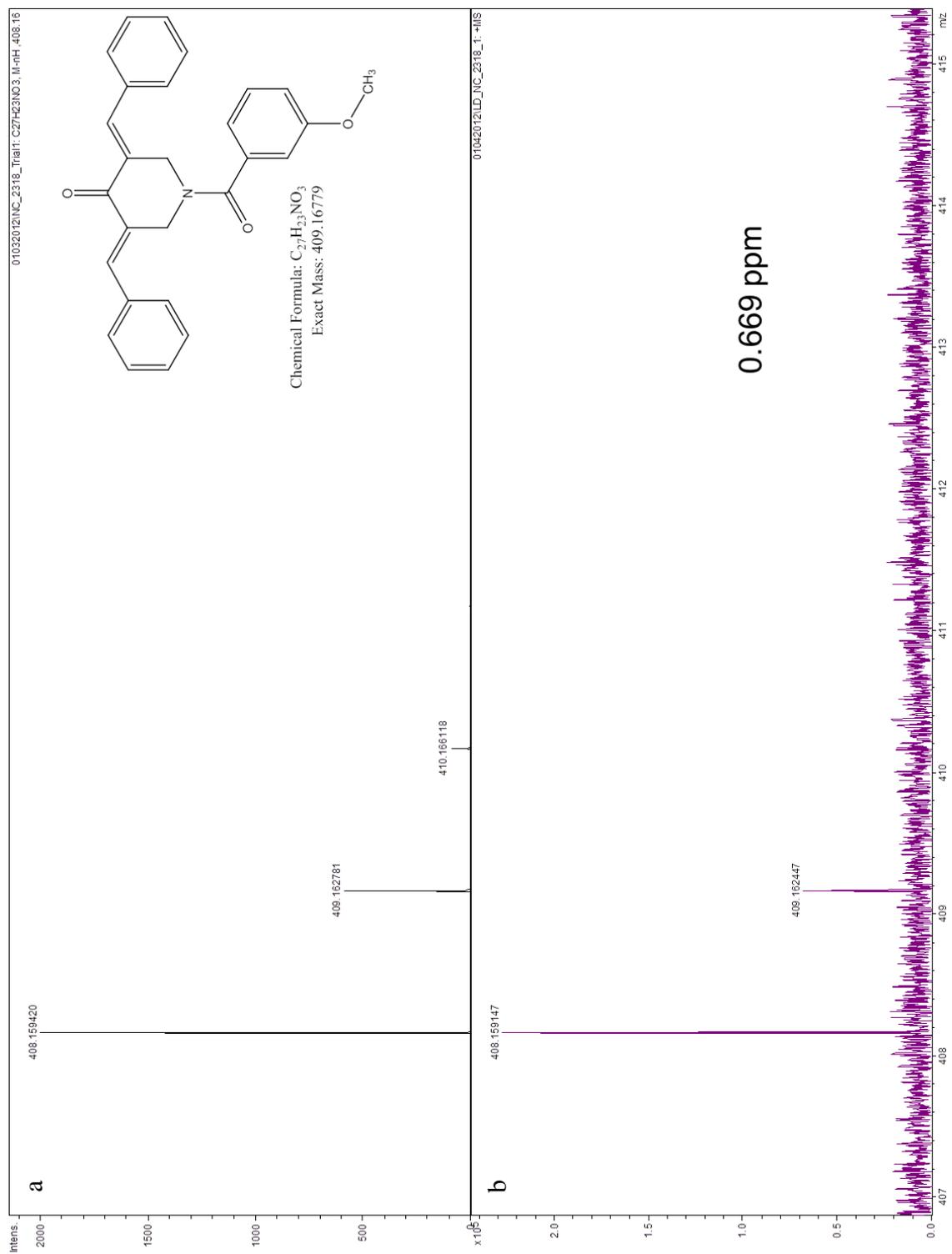


Figure B.6 Mass spectra of curcumin analogue sample NC 2138. (a) Theoretical mass spectrum for the $[M-H]^+$ ion. (b) LD-FT-ICR spectrum containing the $[M-H]^+$ (m/z 408.1594) ion and the mass error observed (0.669 ppm).

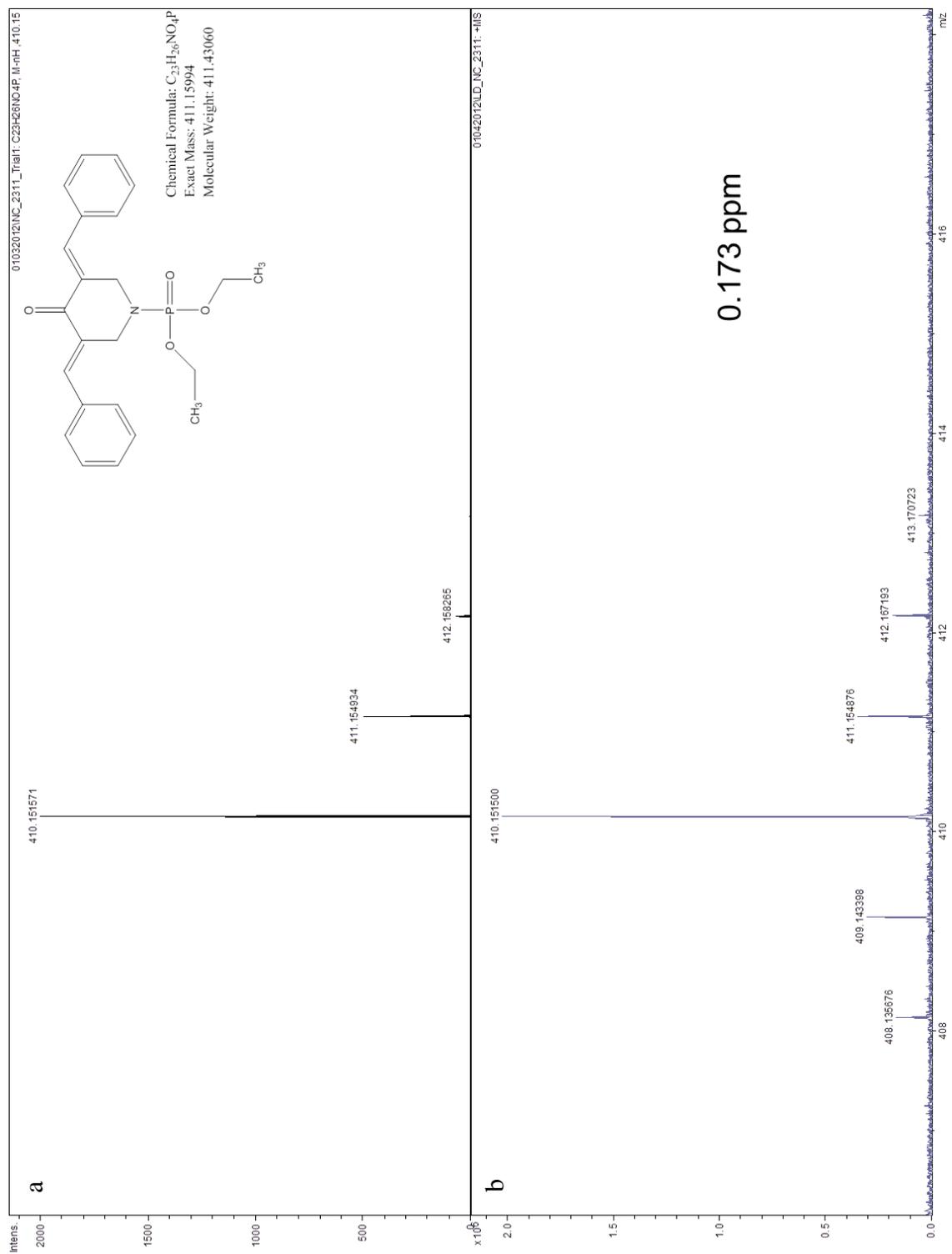


Figure B.7 Mass spectra of curcumin analogue sample NC 2311. (a) Theoretical mass spectrum for the $[M-H]^+$ ion. (b) LD-FT-ICR spectrum containing the $[M-H]^+$ (m/z 410.1516) ion and the mass error observed (0.173 ppm).

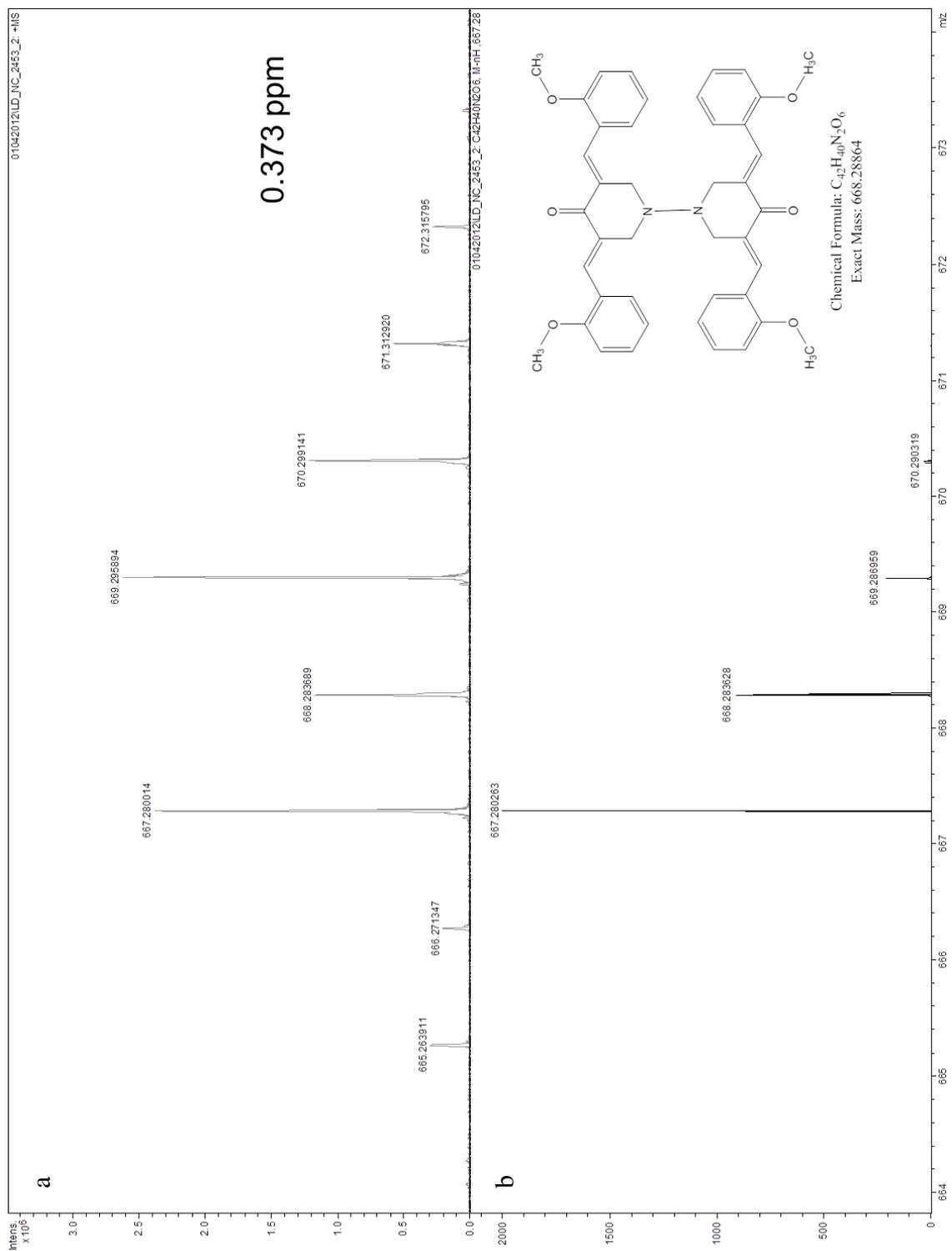


Figure B.8 Mass spectra of curcumin analogue sample NC 2453. (a) LD-FT-ICR spectrum containing the [M-H]⁺ (m/z 667.2800) ion and the mass error observed (0.373 ppm). (b) Theoretical mass spectrum for the [M-H]⁺ ion.

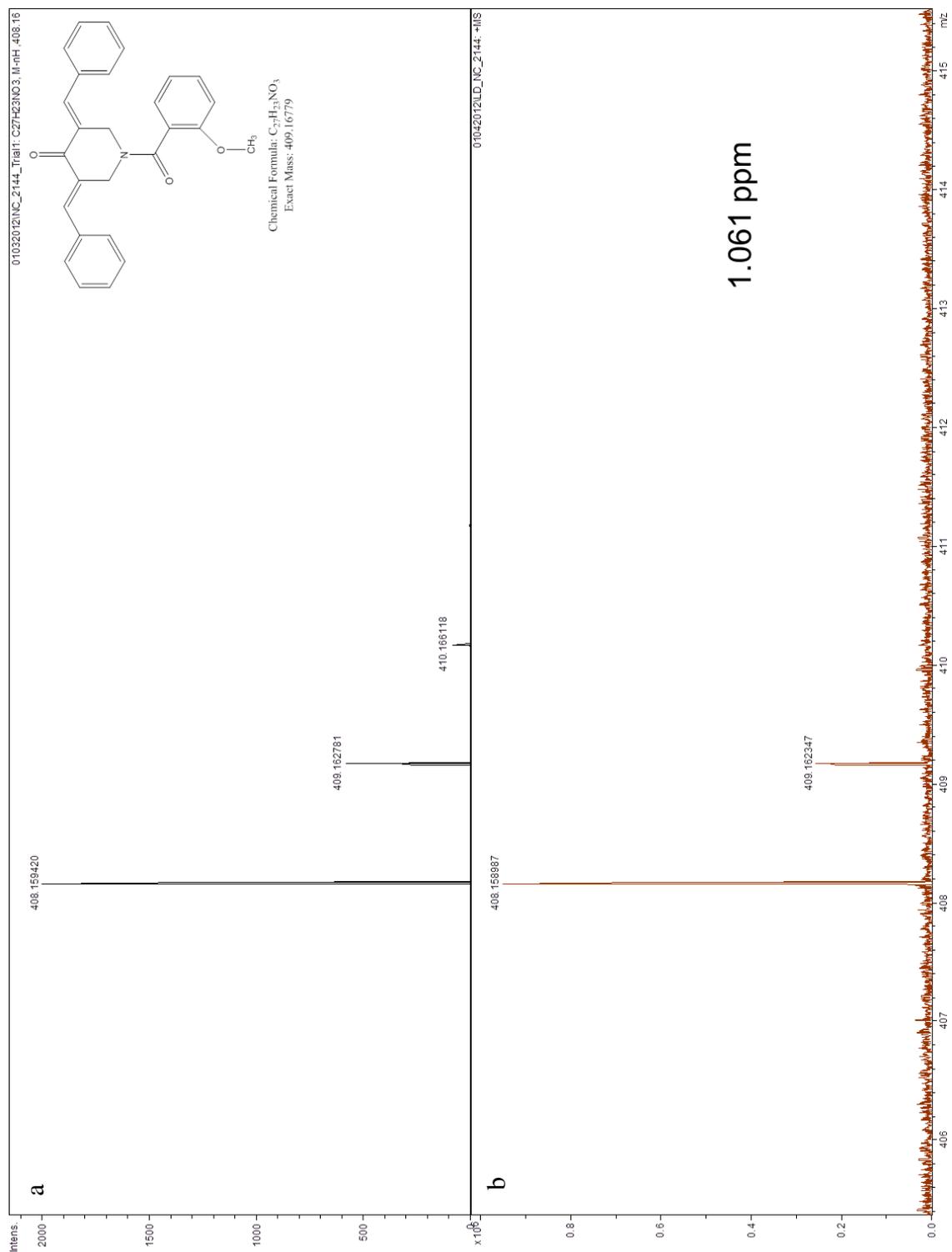


Figure B.9 Mass spectra of curcumin analogue sample NC 2144. (a) Theoretical mass spectrum for the $[M-H]^+$ ion. (b) LD-FT-ICR spectrum containing the $[M-H]^+$ (m/z 408.1594) ion and the mass error observed (1.061 ppm).

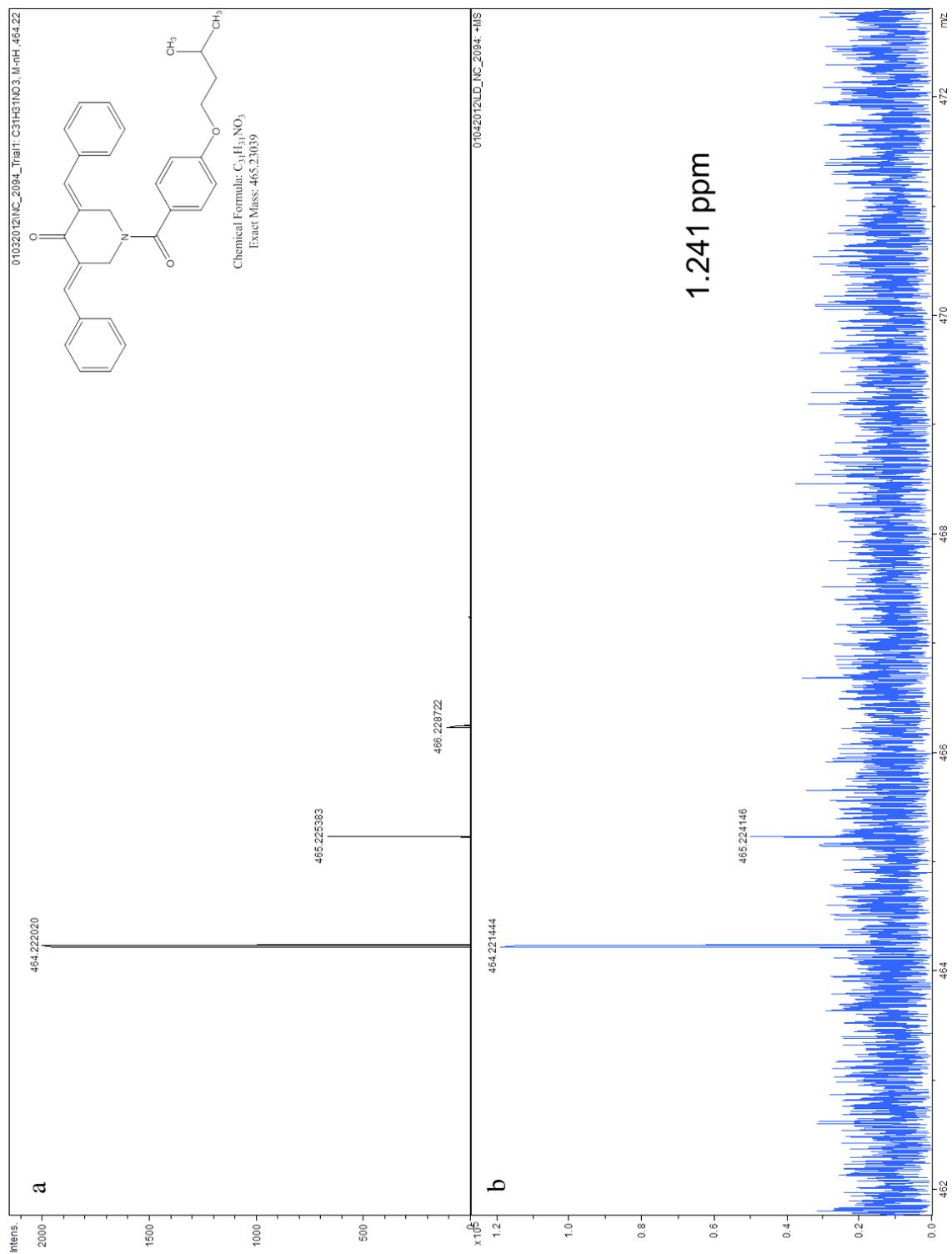


Figure B.10 Mass spectra of curcumin analogue sample NC 2094. (a) Theoretical mass spectrum for the $[M-H]^+$ ion. (b) LD-FT-ICR spectrum containing the $[M-H]^+$ (m/z 464.2220) ion and the mass error observed (1.241 ppm).

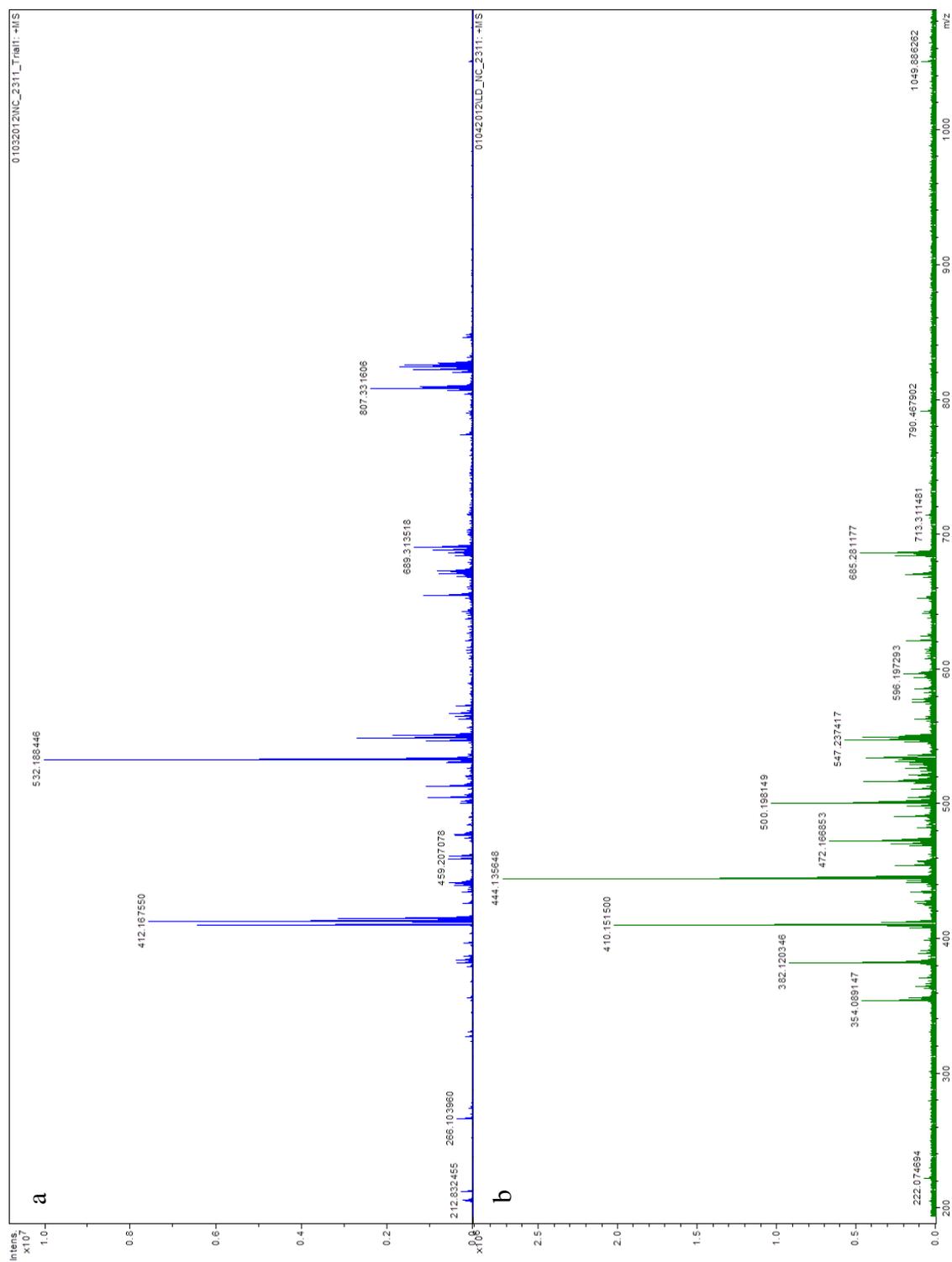


Figure B.11 MALDI and LD analysis of NC 2311. Comparison of the mass spectra showing the entire m/z range collected for curcumin analogue sample NC 2311 under (a) MALDI conditions and (b) LD conditions.

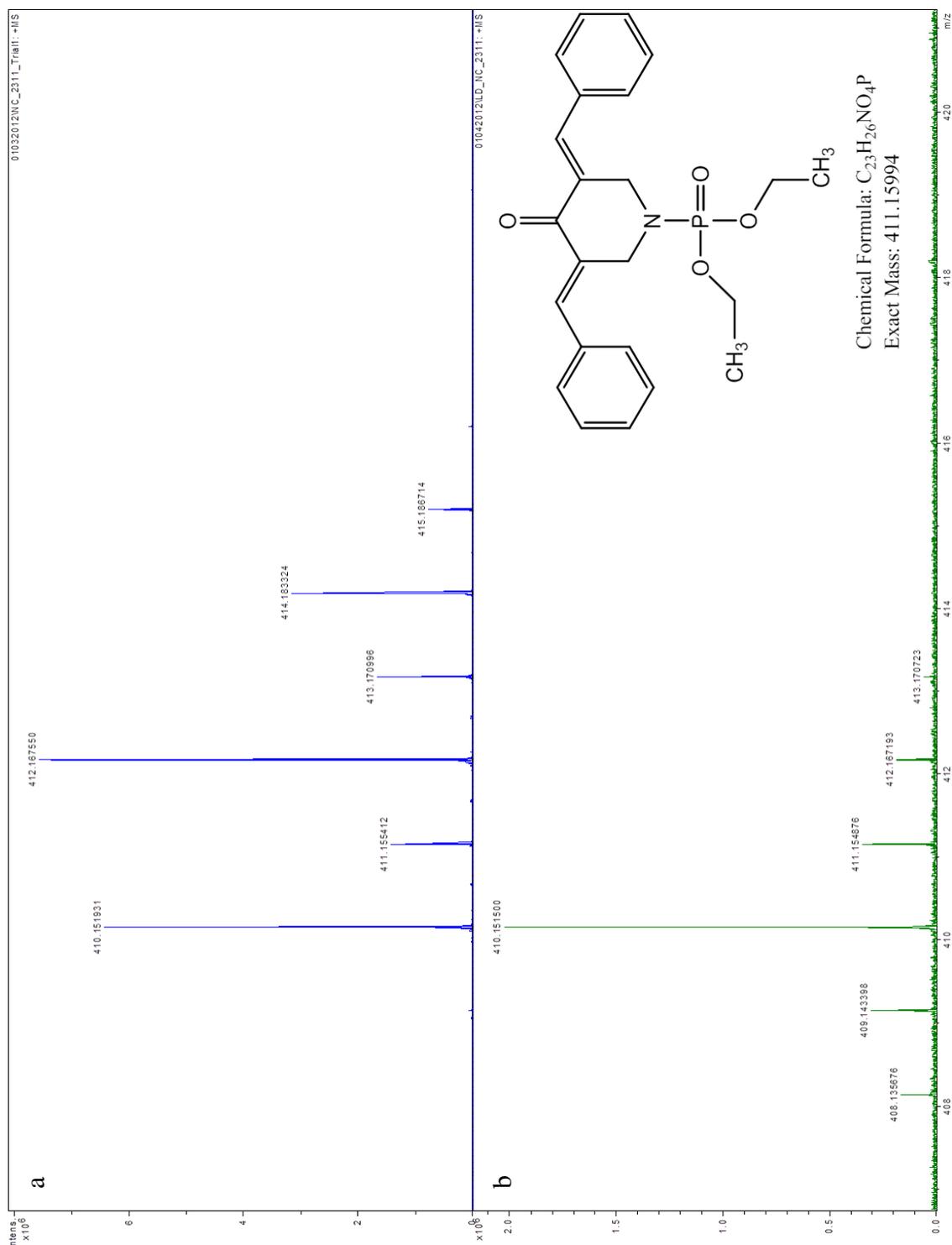


Figure B.12 MALDI and LD analysis of NC 2311. Comparison of the mass spectra, showing the m/z range showing the $[M+H]^+$ (m/z 412.1676) and $[M-H]^+$ (m/z 410.1519) ions, collected for curcumin analogue sample NC 2311 under (a) MALDI conditions and (b) LD conditions.

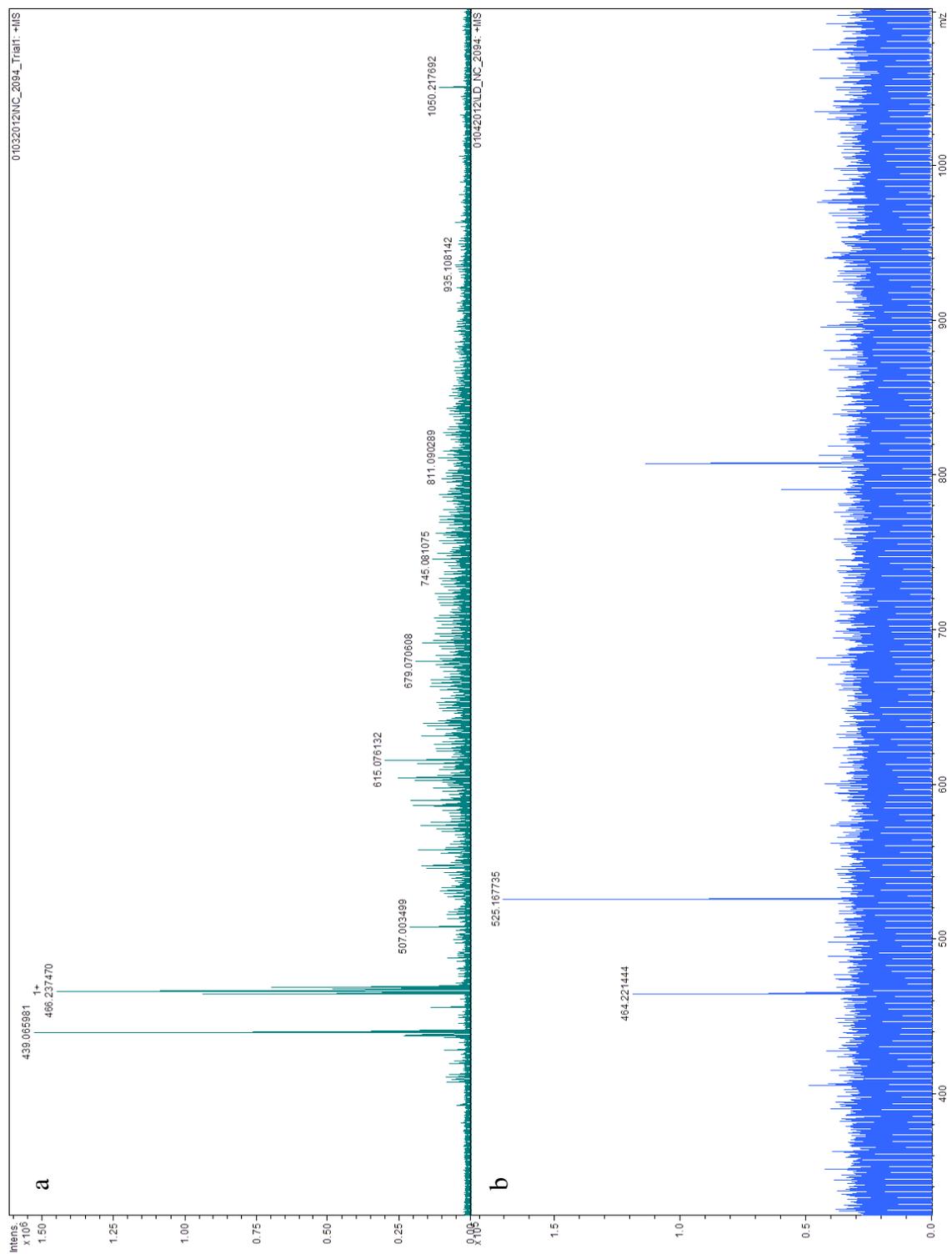


Figure B.13 MALDI and LD analysis of NC 2094. Comparison of the mass spectra showing the entire m/z range collected for curcumin analogue sample NC 2094 under (a) MALDI conditions and (b) LD conditions.

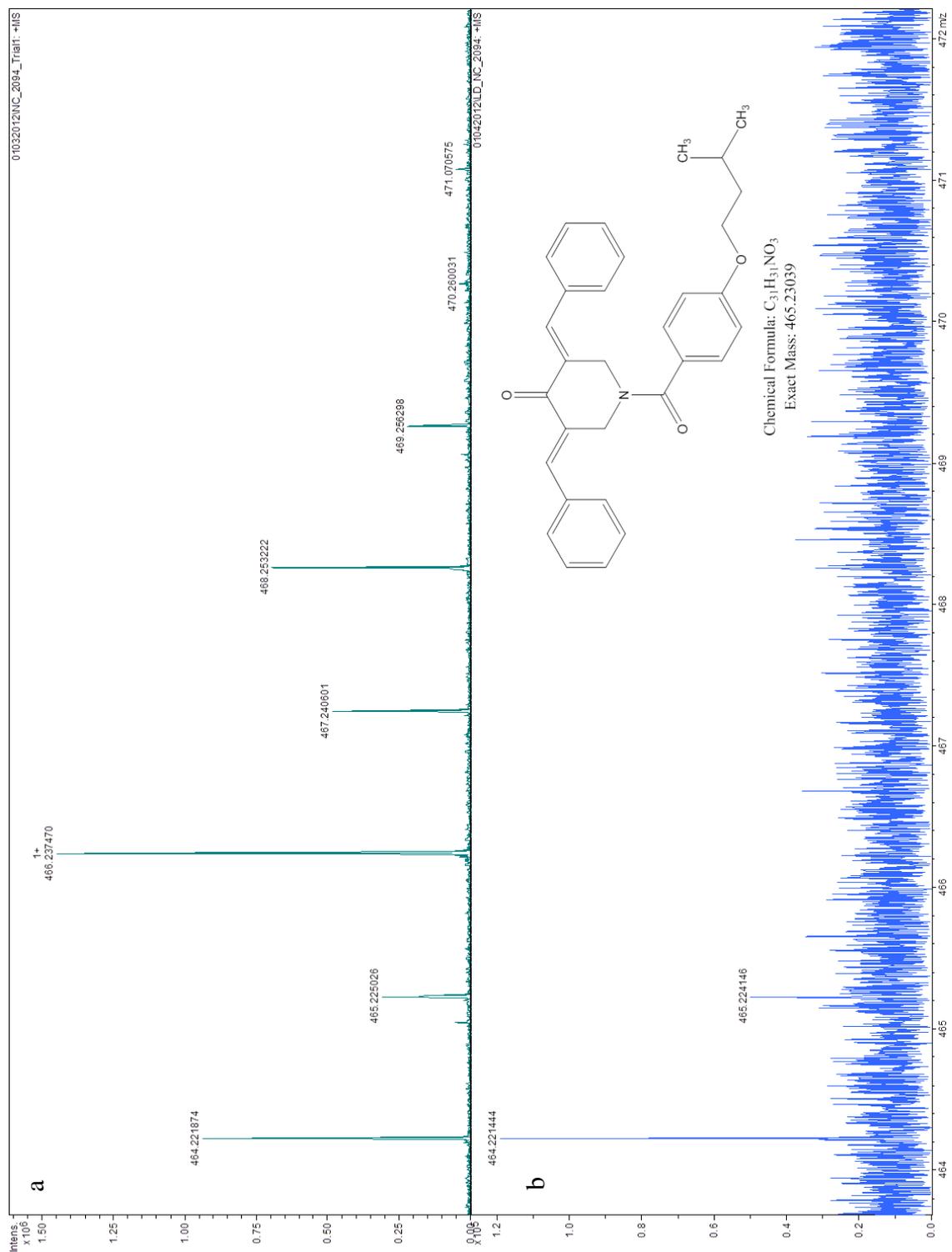


Figure B.14 MALDI and LD analysis of NC 2094. Comparison of the mass spectra, showing the m/z range showing the $[M+H]^+$ (m/z 466.2374) and $[M-H]^+$ (m/z 464.2219) ions, collected for curcumin analogue sample NC 2094 under (a) MALDI conditions and (b) LD conditions.

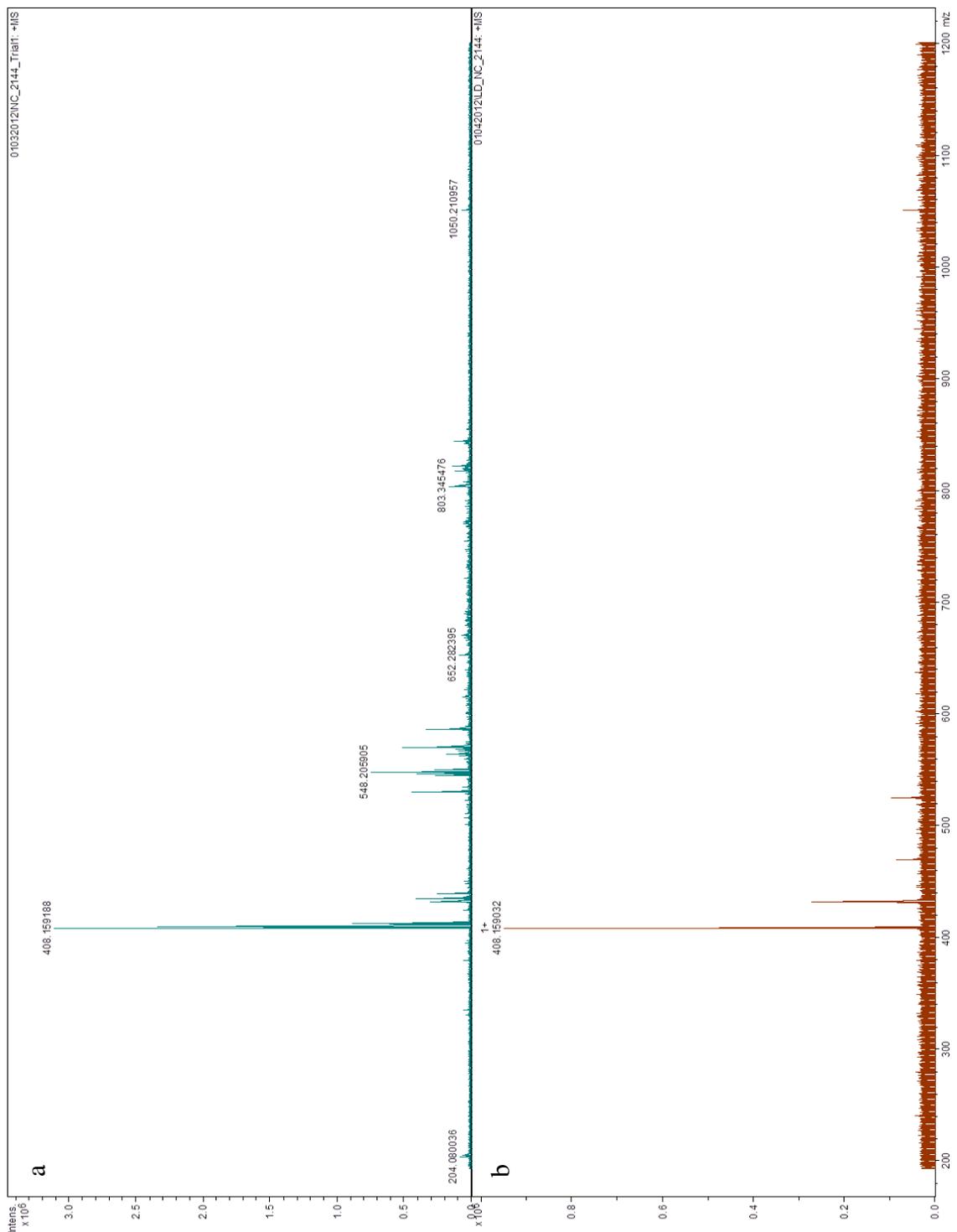


Figure B.15 MALDI and LD analysis of NC 2144. Comparison of the mass spectra showing the entire m/z range collected for curcumin analogue sample NC 2144 under (a) MALDI conditions and (b) LD conditions.

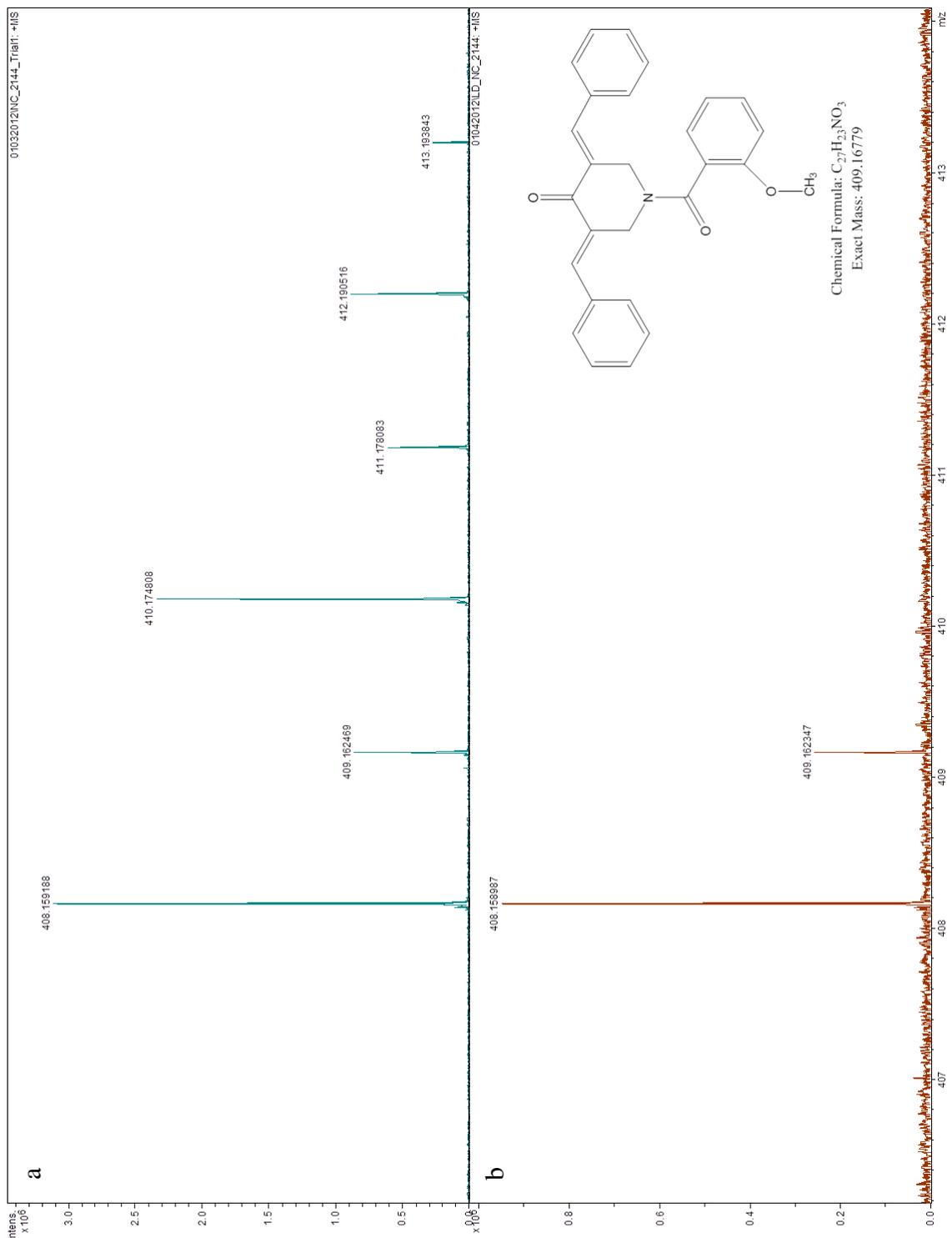


Figure B.16 MALDI and LD analysis of NC 2144. Comparison of the mass spectra, showing the m/z range showing the $[M+H]^+$ (m/z 410.1748) and $[M-H]^+$ (m/z 408.1592) ions, collected for curcumin analogue sample NC 2144 under (a) MALDI conditions and (b) LD conditions.

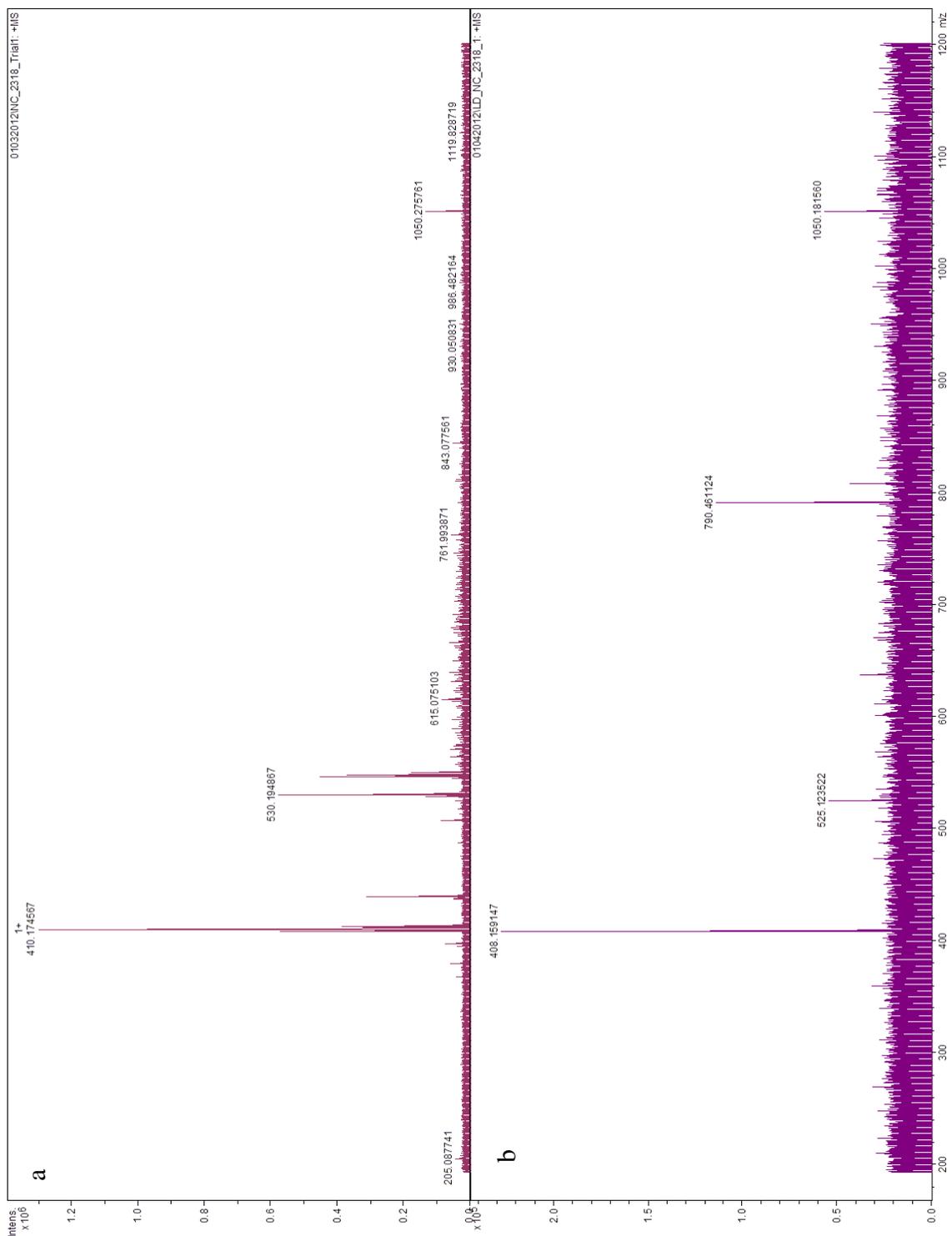


Figure B.17 MALDI and LD analysis of NC 2138. Comparison of the mass spectra showing the entire m/z range collected for curcumin analogue sample NC 2138 under (a) MALDI conditions and (b) LD conditions.

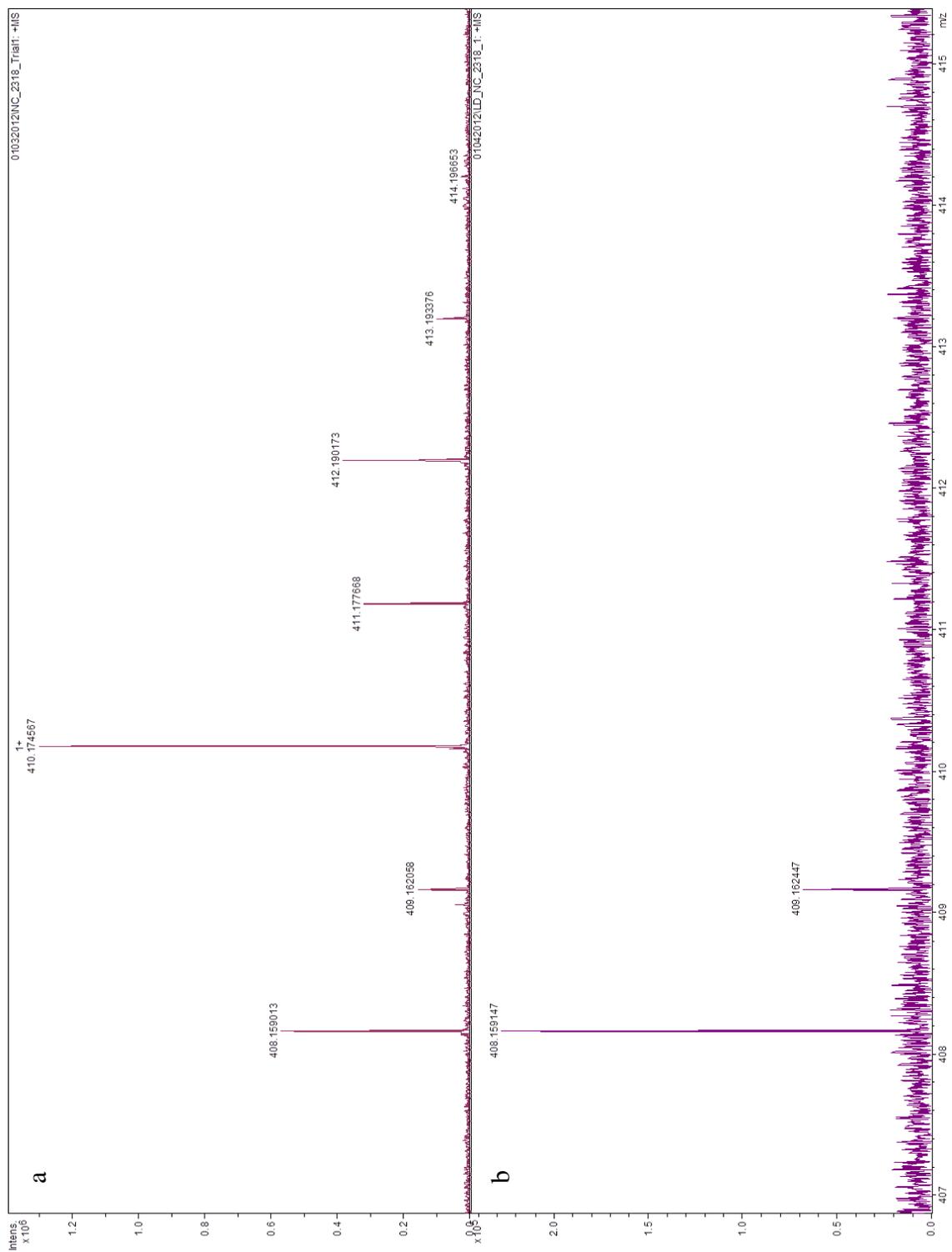


Figure B.18 MALDI and LD analysis of NC 2138. Comparison of the mass spectra, showing the m/z range showing the $[M+H]^+$ (m/z 410.1746) and $[M-H]^+$ (m/z 408.1590) ions, collected for curcumin analogue sample NC 2138 under (a) MALDI conditions and (b) LD conditions.