

Comparison of Formaldehyde with Established Crosslinkers to Characterize Transient Protein Structures Using Mass Spectrometry

by

Savita Srinivasa

B.Sc., The University of Pittsburgh, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2017

© Savita Srinivasa 2017

Abstract

Chemical cross-linking along with mass spectrometry can elucidate protein geometry by introducing stabilizing covalent linkages as distance constraints. Formaldehyde's small size allows it to quickly permeate the cellular membrane without external manipulation and preserve close-proximity and transient protein interactions under physiological conditions. Despite its established uses in biology and compatibility with mass spectrometry, formaldehyde has not yet been applied to structural proteomics, which other cross-linkers have already accomplished. In this thesis, formaldehyde along with four other established cross-linkers (three N-hydroxysuccinimide ester cross-linkers and one zero-length cross-linker), varying in size and reactivity, were shown to capture Ribonuclease S and the Ca^{2+} -free calmodulin-melittin, which are two weak protein complexes. It was demonstrated that the yield of close-proximity crosslinking from zero-length and formaldehyde cross-linkers reflected the dissociation constants of both transient complexes. A comparison between the identification of formaldehyde and established cross-linked species via first stage mass spectrometry (MS) and tandem mass spectrometry (MS/MS) provided insight into what evidence is sufficient to confirm formaldehyde cross-linked species. Cross-linked species from all cross-linkers were identified via MS/MS in the Ca^{2+} -free calmodulin-melittin. These were used to impose different distance constraints to examine the unknown binding orientation of Ca^{2+} -free calmodulin to melittin. The relatively straightforward discovery of N-hydroxysuccinimide ester cross-linking was offset by its large size and ambiguous distance constraints that may not be suitable for small proteins. Although zero-length cross-linkers create close proximity linkages, the high abundance of its reactive sites in calmodulin-melittin produced diversified products, complicating mass spectrometric detection. The increased complexity in identifying formaldehyde reaction products via mass spectrometry was due to its reactivity

Abstract

with several amino acids. This work represents the first report of formaldehyde cross-links identified between non-covalently associated protein components, supporting formaldehyde's ability to stabilize weak interactions. Four formaldehyde crosslinking sites were localized in calmodulin-melittin, and the mechanisms of the formation of these cross-links were revealed using *in vivo*-like conditions. The uniformity of formaldehyde crosslink localization reflected the uniform binding structure of calmodulin. Furthermore, the binding orientation of calmodulin and melittin captured by formaldehyde was shown to be most consistent with recent literature compared to the other cross-linkers.

Lay Summary

Biological processes are governed by the interactions of proteins in cells. Covalent linkages can be introduced via chemical cross-linking reagents to connect and freeze interacting proteins, providing a snapshot of the configurations of proteins in cells. Mass spectrometry can identify protein interacting partners and localize specific interacting regions by measuring the mass-to-charge ratio of species in a sample. Formaldehyde has been extensively used with biological material such as preserving clinically diagnosed tissues and identifying protein interacting partners. Surprisingly, unlike other established cross-linking reagents, formaldehyde has yet to be applied to examine protein geometry in biologically relevant systems. In this work, cross-linkers of various lengths and reactivity were applied to obtain a comprehensive picture of two different protein interactions. Furthermore, formaldehyde was compared to established cross-linkers to reveal its potential to characterize structures of weak protein interactions using mass spectrometry for the first time.

Preface

This thesis project was proposed by my supervisor, Professor Juergen Kast. I was responsible for the experimental work, data analysis and literature searches. Dr. Nikolay Stoyanov and Jason Rogalski performed the mass spectrometry.

Chapter 1, the Introduction, was adapted from the following publication:

Srinivasa, Savita, Xuan Ding, and Juergen Kast. "Formaldehyde cross-linking and structural proteomics: Bridging the gap." *Methods* 89 (2015): 91-98.

Table of Contents

| | |
|---|---------|
| Abstract | ii |
| Lay Summary | iv |
| Preface | v |
| Table of Contents | vi |
| List of Tables | x |
| List of Figures | xv |
| List of Abbreviations | .xxviii |
| Acknowledgements | .xxxiii |
| Dedication | .xxxv |
| 1 Introduction | 1 |
| 1.1 Structural Proteomics | 1 |
| 1.2 Mass Spectrometry in Structural Proteomics | 2 |
| 1.3 Mass Spectrometry and Chemical Cross-linking | 6 |
| 1.4 Tandem Mass Spectrometric Fragmentation and Nomenclature of Cross-linked Species | 20 |
| 1.5 Importance of Formaldehyde Cross-linking: Common Applica- tions and Key Features | 22 |
| 1.6 Formaldehyde Cross-linking <i>in vivo</i> | 24 |
| 1.7 Formaldehyde Cross-linking in Model Proteins | 26 |

Table of Contents

| | | |
|----------|---|-----------|
| 1.8 | Non-covalent Protein Complex Model Systems | 28 |
| 1.9 | Thesis Aims | 33 |
| 2 | Methods | 36 |
| 2.1 | Materials | 36 |
| 2.2 | Chemical Cross-linking Reactions | 37 |
| 2.3 | Tris-Tricine Sodium Dodecyl Sulfate Polyacrylamide Gel Elec- trophoresis Separation | 38 |
| 2.4 | Trypsin Digestion | 40 |
| 2.5 | Reverse Phase High Performance Liquid Chromatography Tandem Mass Spectrometric Analysis | 41 |
| 2.6 | Mass Spectrometric Data Analysis | 42 |
| 2.7 | Analysis Based on Relative Abundance Calculations in the Calmodulin-Melittin System | 49 |
| 2.8 | Crystal Structure Distance Constraints for the Calmodulin-Melittin System | 55 |
| 3 | General Data Analysis of Cross-linked Calmodulin-Melittin and Ri- bonuclease S | 57 |
| 3.1 | Cross-linking of Calmodulin-Melittin and Ribonuclease-S Com- plexes with Various Cross-linkers | 57 |
| 3.2 | Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Sep- aration | 59 |
| 3.3 | Mass Spectrometric Analysis | 74 |
| 3.4 | Moving toward the MS/MS Verification of Cross-Linked Species | 94 |
| 4 | Tandem Mass Spectrometric Fragmentation of Calmodulin-Melittin Cross-linked Species | 96 |
| 4.1 | Tandem Mass Spectrometric Verification and Fragmentation Rules for Cross-linked Species | 96 |
| 4.2 | Tandem Mass Spectrometric Fragmentation of Other Cross-linkers | 99 |
| 4.3 | Tandem Mass Spectrometric Fragmentation of Formaldehyde . . | 137 |
| 4.4 | Formaldehyde versus other Cross-linker Fragmentation | 156 |

Table of Contents

| | | |
|----------|---|------------|
| 4.5 | General Criteria for Evaluating Tandem Mass Spectrometric Patterns of Cross-linked Species | 157 |
| 4.6 | A Second Look at Trypsin Digestion of Cross-linked Residues . . . | 158 |
| 4.7 | MS/MS Analysis of Formaldehyde Cross-linked Ribonuclease-S | 159 |
| 5 | Structural Characterization of Calmodulin-Melittin and Ribonuclease S Cross-linked Species | 160 |
| 5.1 | Trypsin Cleavage and Accessibility of Residues | 160 |
| 5.2 | Relative Abundance of Formaldehyde Cross-linking | 168 |
| 5.3 | Cross-linked Product Classification and Abundance in the Calmodulin-Melittin System | 176 |
| 5.4 | Crystal Structure Distance Constraints | 178 |
| 6 | Arising Limitations of Mass Spectrometric Data Analysis of Formaldehyde Cross-linked Species | 189 |
| 6.1 | Identification of Limitations Arising in Workflow | 189 |
| 6.2 | Mass Spectrometer Comparison | 190 |
| 6.3 | Assignment of Monoisotopic Masses | 201 |
| 6.4 | Complexity of Cross-linked Candidates Confirmed by Mass Spectrometry | 205 |
| 6.5 | Manual versus Software Identification of Calmodulin-Melittin Cross-links | 208 |
| 6.6 | Limitations in the MS/MS Analysis of Formaldehyde Cross-linked Ribonuclease-S | 233 |
| 7 | Conclusion and Future Outlook | 235 |
| 7.1 | Clarification of Formaldehyde Cross-linking Reaction Chemistry | 235 |
| 7.2 | Trypsin Digestion Efficiency of Formaldehyde Reaction Products | 238 |
| 7.3 | Establishing Tandem Mass Spectrometric Fragmentation Rules for Formaldehyde Cross-link Identification | 239 |
| 7.4 | The Structural Characterization of Ca ²⁺ -free calmodulin-melittin via Comprehensive Cross-linking | 240 |
| 7.5 | Revealing Complexity of Cross-linking Reaction Mixtures | 242 |
| 7.6 | Comparing Software Versus Manual Cross-link Identification . . . | 245 |

Table of Contents

| | |
|---|-----|
| 7.7 Moving toward Cellular, <i>in vivo</i> Systems | 246 |
| Bibliography | 248 |
| Appendices | |
| A First Appendix | 263 |
| A.1 Confirming the Calcium-Free Calmodulin-Melittin System | 263 |
| A.2 Data Analysis Codes | 270 |
| A.3 Bruker Impact II Tandem Mass Spectrometric Analysis Method Details | 273 |
| A.4 Calmodulin-Melittin Cross-linked Candidates From First-Stage Mass Spectrometry | 274 |
| A.5 Ribonuclease S Cross-linked Candidates | 282 |

List of Tables

| | | |
|-----|---|----|
| 2.1 | Maximum cross-linking distances for every combination of possible reactive sites for each cross-linker are listed | 56 |
| 3.1 | A list of MS/MS confirmed peptides in the control calmodulin-melittin sample. The m/z , experimental mass, calculated mass, mass accuracy, normalized peak area, molecular weight of the protein gel band origin, sequence and number of missed cleavages for each peptide are listed left to right. | 76 |
| 3.2 | A list of MS/MS confirmed calmodulin-melittin peptides in PFA treated sample without PFA modifications. The m/z , experimental mass, calculated mass, mass accuracy, normalized peak area, molecular weight of the protein gel band origin, sequence and number of missed cleavages for each peptide are listed left to right. | 80 |
| 3.3 | A list of MS/MS confirmed calmodulin-melittin peptides in PFA treated sample with PFA modifications. The m/z , experimental mass, calculated mass, mass accuracy, normalized peak area, molecular weight of the protein gel band origin, sequence and number of missed cleavages for each peptide are listed left to right. (+12) and (+30) denotes a Schiff Base/Intrapeptide cross-link and methylol, respectively, localized on the residue before it. | 81 |
| 3.4 | Cross-linking and modification sites in calmodulin and melittin for each cross-linker | 85 |

List of Tables

| | | |
|-----|---|-----|
| 3.5 | List of S-peptide, S-protein and RNaseA peptides in the control RNaseS sample. The m/z , experimental mass, calculated mass, mass accuracy, normalized peak area, molecular weight of the protein gel band origin, sequence and number of missed cleavages for each peptide are listed left to right. | 89 |
| 3.6 | List of S-peptide, S-protein and RNaseA peptides in the PFA treated RNaseS sample. The m/z , experimental mass, calculated mass, mass accuracy, normalized peak area, molecular weight of the protein gel band origin, sequence and number of missed cleavages for each peptide are listed left to right. (+12) and (+30) denotes a Schiff Base/Intrapeptide cross-link and methylol, respectively, localized on the residue before it. | 92 |
| 3.7 | Cross-linking and modification sites in RNaseS for each cross-linker. | 93 |
| 4.1 | EDC Calmodulin-Calmodulin interpeptide cross-linked species, in which cross-linking sites are highlighted in red. For species appearing with two different charge states, annotated MS/MS spectra is shown for the m/z marked with an “*”. | 101 |
| 4.2 | EDC calmodulin-melittin interpeptide cross-linked species are listed and classified as capturing antiparallel (shaded in blue) or parallel (white) binding. Reactive residues/possible cross-linking sites are highlighted in red. For species appearing with two different charge states, annotated MS/MS spectra is shown for the m/z marked with an “*”. | 106 |
| 4.3 | sulfoDST calmodulin-melittin interpeptide cross-linked species are listed and classified as capturing antiparallel (shaded in blue) or parallel (white) binding. Cross-linking sites are highlighted in red. | 111 |
| 4.4 | BS ³⁺ calmodulin interpeptide cross-linked species, in which cross-linking sites are highlighted in red. For species appearing with two different charge states, annotated MS/MS spectra is shown for the m/z marked with an “*”. | 115 |

List of Tables

| | | |
|-----|--|-----|
| 4.5 | BS ³⁺ calmodulin-melittin interpeptide cross-linked species are listed and classified as capturing antiparallel (shaded in blue) or parallel (white) binding. Cross-linking sites are highlighted in red. For species appearing with two different charge states, annotated MS/MS spectra is shown for the <i>m/z</i> marked with an “*”. | 120 |
| 4.6 | SulfoEGS calmodulin interpeptide cross-linked species, in which cross-linking sites are highlighted in red. For species appearing with two different charge states, annotated MS/MS spectra is shown for the <i>m/z</i> marked with an “*”. | 129 |
| 4.7 | SulfoEGS calmodulin-melittin interpeptide cross-linked species are listed and classified as capturing antiparallel (shaded in blue) or parallel (white) binding. Cross-linking sites are highlighted in red. | 132 |
| 4.8 | PFA calmodulin interpeptide cross-linked species, in which cross-linking sites are highlighted in red. | 139 |
| 4.9 | PFA calmodulin-melittin interpeptide cross-linked species are listed and classified as capturing antiparallel (shaded in blue) or parallel (white) binding. Cross-linking sites are highlighted in red. | 147 |
| 5.1 | Percent abundances of cleaved trypsin cleavage sites observed in the control and PFA treated (in 14-19, 19-33 and > 33 kDa proteins) calmodulin-melittin samples | 165 |
| 5.2 | Percent abundances of cleaved trypsin cleavage sites observed in the control and PFA treated (> 12 kDa proteins) RNaseS samples | 168 |
| 5.3 | The percent abundance of PFA cross-linking sites in the unmodified, modified and cross-linked forms in PFA treated calmodulin-melittin; Note: additional decimal places are reported to clarify that values are > 0% or < 100%, as described in section 2.7 | 170 |
| 5.4 | The calculated equilibrium constants for each cross-linking reaction step for each identified PFA cross-linking site in PFA treated calmodulin-melittin | 172 |

List of Tables

| | | |
|------|--|-----|
| 5.5 | The percent abundance of PFA modification sites in the unmodified and modified forms in PFA treated calmodulin-melittin; Note: additional decimal places are reported to clarify that values are > 0% or < 100%, as described in section 2.7 | 174 |
| 5.6 | The calculated equilibrium constants for the modification reaction for each identified PFA modification site in PFA treated calmodulin-melittin | 174 |
| 5.7 | The percent abundance of PFA modification sites in the unmodified and modified forms and the calculated equilibrium constants for the modification reaction for each identified PFA modification site in PFA treated RNaseS | 175 |
| 5.8 | Relative abundance of calmodulin-calmodulin and calmodulin-melittin cross-linked peptides | 176 |
| 5.9 | Relative abundance of calmodulin-melittin interpeptide cross-links supporting the parallel and antiparallel binding orientation | 177 |
| 5.10 | The maximum distances between all MS/MS identified cross-linking sites and the respective binding orientation it supports for each cross-linker; Parallel orientations are shaded in white and antiparallel orientations are shaded in blue. | 184 |
| 6.1 | The number of identified calmodulin-melittin cross-linked candidates identified via MS and confirmed via MS/MS using the QStar and Impact II | 191 |
| 6.2 | Mascot MS/MS search results for the Impact II analyzed calmodulin sample with the highest scoring match for each peptide identified listed. The sequence position (starting and ending residue), observed m/z , experimental monoisotopic mass, theoretical monoisotopic mass, mass accuracy, number of missed cleavages, Mascot score, and sequence (trypsin cleavage site displayed in the beginning and end of the sequence as “R.” or “K.”) are listed left to right. | 200 |

List of Tables

| | | |
|------|--|-----|
| 6.3 | Mascot MS/MS search results for the QStar analyzed calmodulin sample with the highest scoring match for each peptide identified listed. The sequence position (starting and ending residue), observed m/z , experimental monoisotopic mass, theoretical monoisotopic mass, mass accuracy, number of missed cleavages, Mascot score, and sequence (trypsin cleavage site displayed in the beginning and end of the sequence as “R.” or “K.”) are listed left to right. | 201 |
| 6.4 | The percent of the total number of MS candidate cross-linked species with incorrectly assigned monoisotopic peaks by the software for each cross-linker | 205 |
| 6.5 | The percent of the total number of MS confirmed candidate cross-linked masses that correspond to modified peptides, undetermined species, species with insufficient MS/MS and cross-linked species for each cross-linker. | 205 |
| A.1 | MS Candidate Cross-linked Species for EDC | 274 |
| A.2 | MS Candidate Cross-linked Species for PFA | 275 |
| A.3 | MS Candidate Cross-linked Species for PFA | 276 |
| A.4 | MS Candidate Cross-linked Species for PFA | 277 |
| A.5 | MS Candidate Cross-linked Species for sulfoDST | 278 |
| A.6 | MS Candidate Cross-linked Species for BS ³ | 279 |
| A.7 | MS Candidate Cross-linked Species for sulfoEGS | 280 |
| A.8 | MS Candidate Cross-linked Species for sulfoEGS | 281 |
| A.9 | Candidate Cross-linked Species for EDC | 282 |
| A.10 | Candidate Cross-linked Species for PFA | 283 |
| A.11 | Candidate Cross-linked Species for PFA | 284 |
| A.12 | Candidate Cross-linked Species for sulfoDST | 285 |
| A.13 | Candidate Cross-linked Species for BS ³ | 286 |
| A.14 | Candidate Cross-linked Species for BS ³ | 287 |
| A.15 | Candidate Cross-linked Species for sulfoEGS | 288 |
| A.16 | Candidate Cross-linked Species for sulfoEGS | 289 |

List of Figures

- 1.1 A summary of the bottom-up proteomic strategy, in which proteins are purified, separated via SDS-PAGE and enzymatically digested into peptides, which are separated through HPLC and eluted into the mass spectrometer. The schematic diagram of a Bruker Impact II QqTOF mass spectrometer is shown on the top right (adapted from reference [14] with permission, <https://creativecommons.org/licenses/by/4.0/>). At each time point an MS spectrum of the ions eluted is plotted. For example at time = T_1 , a species with a m/z of 301.00 ($z = 2$) eluted and the signal appears as isotopic cluster such that the charge is the inverse of the difference between each isotopic peak i.e. 1 divided by 0.5 . The MS/MS of precursor ion produces fragment ion signals (b1,b2...y1,y2...) which are typically the result of the cleavage of peptide bonds in CID. MS/MS spectra and MS spectra are matched to theoretical databases for identifying and sequencing peptides. . . 3
- 1.2 In general, cross-linkers introduce covalent bonds in proteins via a two-step reaction: modification of protein site 1 and cross-link formation between protein sites 1 and 2. Upon enzymatic digestion, a complex mixture of different types of peptides is produced. . . . 7

List of Figures

- 1.3 In the sulfoDST cross-linking reaction scheme, protein site 1 (mass = m_1) reacts with the sulfoDST to form a modification which immediately reacts with protein site 2 (mass = m_2) to form a cross-linking bridge (highlighted in red). A competing hydrolysis reaction product ($m_1 + \text{bridge} + \text{H}_2\text{O}$) can also occur instead of the cross-linked product ($M = m_1 + m_2 + \text{bridge}$). R_1 , R_2 and R_4 are defined in the dotted box. 10
- 1.4 In the BS^3 cross-linking reaction scheme, protein site 1 (mass = m_1) reacts with the sulfoDST to form a modification which immediately reacts with protein site 2 (mass = m_2) to form a cross-linking bridge (highlighted in red). A competing hydrolysis reaction product ($m_1 + \text{bridge} + \text{H}_2\text{O}$) can also occur instead of the cross-linked product ($M = m_1 + m_2 + \text{bridge}$). R_1 , R_2 and R_4 are defined in the dotted box. 11
- 1.5 In the sulfoEGS cross-linking reaction scheme, protein site 1 (mass = m_1) reacts with the sulfoDST to form a modification which immediately reacts with protein site 2 (mass = m_2) to form a cross-linking bridge (highlighted in red). A competing hydrolysis reaction product ($m_1 + \text{bridge} + \text{H}_2\text{O}$) can also occur instead of the cross-linked product ($M = m_1 + m_2 + \text{bridge}$). R_1 , R_2 and R_4 are defined in the dotted box. 12
- 1.6 In the EDC/sulfoNHS cross-linking reaction scheme, protein site 1 (mass = m_1) reacts with the cross-linker. The intermediate can react with sulfoNHS to form an amine reactive sulfoNHS ester, which forms a cross-link with protein site 2 (mass = m_2). The intermediate can also produce a stable N-acylisourea ($m_1 + 155$ Da). Cross-linker bridges are highlighted in red. R_1 and R_2 represent two protein sites and specific reactive amino acids are defined in dashed line boxes. Mass of each species are provided in solid line boxes. 14

List of Figures

| | | |
|------|---|----|
| 1.7 | PFA reacts with protein site 1 (mass = m_1), which forms a methylol modification (mass = $m_1 + 30$ Da). This can dehydrate into a Schiff Base (mass = $m_1 + 12$ Da), which can continue to react with protein site 2 (mass = m_2) to form a methylene bridge. This produces a cross-linked species (mass = $m_1 + 12$ Da + m_2). Cross-linker bridges are highlighted in red. R_1 and R_2 represent two protein sites and specific reactive amino acids are defined in dashed line boxes. Mass of each species are provided in solid line boxes. | 16 |
| 1.8 | The PFA modification reaction generic mechanism (a) and for each reactive residue (b) (Adapted from reference [37], with permission) | 17 |
| 1.9 | The PFA cross-linking reaction generic mechanism (a) and for each reactive residue (b) (Adapted from reference [37], with permission) | 18 |
| 1.10 | (a) Cross-linked peptides (denoted as I and II) can fragment exclusively at the cross-linker (type 1), exclusively at the peptide backbone (type 2) and at both the cross-linker and peptide backbone (type 3). (b) Diagnostic ions/specific fragmentation for NHS Esters (c) Specific fragment ions produced from sulfoEGS fragmentations. (a-c) Cross-linker bridges and cross-linked lysines from each protein site are highlighted in red. | 21 |
| 1.11 | (a) Amino acid sequence of the calmodulin N-terminal (teal) and C-terminal (blue) domains connected by a flexible linker (black) and of melittin (purple); All possible trypsin cleavage sites are denoted with red vertical bars. (b) Calmodulin binds to Ca^{2+} , which induces the formation of a dumbbell-shaped conformation; Upon binding to melittin, a similar conformational change occurs for both Ca^{2+} -saturated and Ca^{2+} -free calmodulin. Melittin competitively binds to calmodulin, inhibiting calmodulin's activity. | 31 |
| 1.12 | (a) Amino acid sequences of RNaseS components S-protein (blue) and the S-peptide (green). The S-protein contains 8 cysteines, which were reduced and alkylated in this model system. The S-peptide to S-protein binding sites are underlined. (b) The crystal structure (1RNU) [108] of RNaseS is shown with the S-protein (blue) to S-peptide (green) binding site highlighted in red. | 33 |

List of Figures

| | | |
|-----|--|----|
| 2.1 | Data Analysis workflow for cross-link identification, where items in parentheses are values that have been eliminated. All elimination and matching of monoisotopic masses using Mathematica and Excel was performed using a mass accuracy of ± 0.2 Da. | 47 |
| 2.2 | PFA cross-linking equilibrium reaction steps, where K_1 , K_2 , and K_3 are the respective equilibrium constants for the formation of a methylol, Schiff Base and methylene bridge, respectively. The notation for each reactant and product is defined. R_1 and R_2 represent protein sites 1 and 2, respectively. | 52 |
| 2.3 | The PFA modification equilibrium reaction step defines K_{1+2} in the case where a methylol modification was not identified. R_1 and R_2 represent protein sites 1 and 2, respectively. | 52 |
| 3.1 | SDS-PAGE of calmodulin-melittin cross-linking reaction mixtures with the protein marker (lane 1); control samples with EDC buffer conditions (lane 2), and control samples with all other cross-linker buffer conditions (lane 3); cross-linked samples EDC/sulfoNHS (lane 4), PFA (lane 5), sulfoDST (lane 6), BS ³ (lane 7) and sulfoEGS (lane 8); Four approximate molecular weight categories of each protein/protein complex band are labelled with the type of crosslinking (if any) indicated. | 61 |
| 3.2 | SDS-PAGE of RNaseS cross-linking reaction mixtures with protein marker (lane 1); control samples with EDC buffer conditions (lane 2), and control samples with all other cross-linker buffer conditions (lane 3); cross-linked samples EDC/sulfoNHS (lane 4), PFA (lane 5), sulfoDST (lane 6), BS ³ (lane 7) and sulfoEGS (lane 8); Four approximate molecular weight categories of each protein/protein complex band are labelled with the type of crosslinking (if any) indicated. | 62 |

List of Figures

| | | |
|-----|---|-----|
| 3.3 | Literature study that compares the SDS-PAGE of calmodulin in the presence of EDTA, without Ca ²⁺ (lane 5 and 6) and in the presence of Ca ²⁺ , without EDTA (lane 2 and 3). Lanes 1 and 4 are protein markers. The amounts of calmodulin used were 6µg (lanes 2 and 5) and 12µg (lanes 3 and 6). The concentrations of CaCl ₂ and EDTA were both 5 mM (adapted from reference [139], with permission). | 63 |
| 3.4 | Relative yield of cross-linked species (blue) versus non-cross linked species (red) in the Calmodulin-Melittin complex measured via SDS-PAGE | 70 |
| 3.5 | Relative yield of cross-linked species (blue) versus non-cross linked species (red) in the RNaseS complex measured via SDS-PAGE | 71 |
| 3.6 | In the dotted boxes, the structures of proteins with a lysine side chain modified four ways as indicated with their respective pKa values are shown. Cross-linked bridges are highlighted in red. . . | 83 |
| 3.7 | Trypsin's catalytic triad consists of aspartic acid (D102), histidine (H57) and serine (S195). Aspartic acid and histidine increase the nucleophilicity of serine, which attacks the partially positive carbonyl carbon of the protein. The positively charged amino group on lysine increases the electrophilicity of the carbonyl carbon. The peptide bond is cleaved and the trypsin catalyst is regenerated. (adapted from reference [153]) | 84 |
| 4.1 | The nomenclature used to annotate MS/MS spectra of cross-linked species for each type of fragment ion | 98 |
| 4.2 | Interpeptide EDC calmodulin cross-link at <i>m/z</i> 1100.54 (<i>z</i> = 3) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red; | 102 |
| 4.3 | Interpeptide EDC calmodulin cross-link at <i>m/z</i> 716.96 (<i>z</i> = 5) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom) ; Cross-linker bridges are indicated in red. . . . | 103 |

List of Figures

| | | |
|------|---|-----|
| 4.4 | Interpeptide EDC calmodulin cross-link at m/z 875.76 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 104 |
| 4.5 | Interpeptide EDC calmodulin-melittin cross-link m/z 621.58 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom) ; Cross-linker bridges are indicated in red. | 107 |
| 4.6 | Interpeptide EDC calmodulin-melittin cross-link m/z 616.66 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom) are shown | 108 |
| 4.7 | Interpeptide EDC calmodulin-melittin cross-link m/z 1101.26 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 109 |
| 4.8 | Interpeptide sulfoDST calmodulin-melittin cross-link m/z 842.77 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 112 |
| 4.9 | Interpeptide sulfoDST calmodulin-melittin cross-link m/z 493.92 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. Note: Fragmentation indicated on the backbone of the peptide corresponds to type 3 ions only. | 113 |
| 4.10 | Interpeptide BS^{3+} calmodulin cross-link m/z 1085.55 ($z = 2$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 116 |
| 4.11 | Interpeptide BS^{3+} calmodulin cross-link m/z 1079.51 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 117 |
| 4.12 | Interpeptide BS^{3+} calmodulin cross-link m/z 588.96 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom) ; Cross-linker bridges are indicated in red. | 118 |
| 4.13 | Interpeptide BS^{3+} calmodulin-melittin cross-link m/z 850.46 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 121 |

List of Figures

| | | |
|------|---|-----|
| 4.14 | Interpeptide BS ³⁺ calmodulin-melittin cross-link m/z 732.39 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 122 |
| 4.15 | Interpeptide BS ³⁺ calmodulin-melittin cross-link m/z 699.36 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 123 |
| 4.16 | Interpeptide BS ³⁺ calmodulin-melittin cross-link m/z 733.35 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 124 |
| 4.17 | Interpeptide BS ³⁺ calmodulin-melittin cross-link m/z 481.00 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 125 |
| 4.18 | Interpeptide BS ³⁺ calmodulin-melittin cross-link m/z 520.02 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 126 |
| 4.19 | Interpeptide BS ³⁺ calmodulin-melittin cross-link m/z 448.97 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 127 |
| 4.20 | Interpeptide sulfoEGS calmodulin cross-link m/z 956.96 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 130 |
| 4.21 | Interpeptide sulfoEGS calmodulin cross-link m/z 934.46 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. MS/MS spectra is annotated such that I = II. | 131 |
| 4.22 | Interpeptide sulfoEGS calmodulin-melittin cross-link m/z 571.54 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 133 |
| 4.23 | Interpeptide sulfoEGS calmodulin-melittin cross-link m/z 635.81 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 134 |

List of Figures

| | | |
|------|---|-----|
| 4.24 | Interpeptide sulfoEGS calmodulin-melittin cross-link m/z 791.72 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 135 |
| 4.25 | Interpeptide PFA calmodulin cross-link m/z 666.33 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. Note: Fragmentation indicated on the backbone of the peptide corresponds to type 3 ions only | 140 |
| 4.26 | Interpeptide PFA calmodulin cross-link m/z 736.59 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. Note: Fragmentation indicated on the backbone of the peptide corresponds to type 3 ions only. | 141 |
| 4.27 | Interpeptide PFA calmodulin cross-link m/z 1192.23 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. | 142 |
| 4.28 | Degree of Modification: Bar graph depicting the DOM of each b ion and y ion for $^1\text{GIGAVLK}^7$ in cross-linked species m/z 744.73 ($z = 3$) | 144 |
| 4.29 | Degree of Modification: Bar graph depicting the DOM of each b ion and y ion for $^1\text{GIGAVLK}^7$ in cross-linked species m/z 484.46 ($z = 5$) | 145 |
| 4.30 | Degree of Modification: Bar graph depicting the DOM of each b ion and y ion for $^1\text{GIGAVLK}^7$ in cross-linked species m/z 588.52 ($z = 5$) | 146 |
| 4.31 | Interpeptide PFA calmodulin-melittin cross-link m/z 744.73 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. Note: Fragmentation indicated on the backbone of the peptide corresponds to type 3 ions only | 148 |

List of Figures

| | | |
|------|---|-----|
| 4.32 | Interpeptide PFA calmodulin-melittin cross-link m/z 484.46 ($z = 5$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. Note: Fragmentation indicated on the backbone of the peptide corresponds to type 3 ions only | 149 |
| 4.33 | Interpeptide PFA calmodulin-melittin cross-link m/z 588.52 ($z = 5$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. Note: Fragmentation indicated on the backbone of the peptide corresponds to type 3 ions only | 150 |
| 4.34 | Interpeptide PFA calmodulin-melittin cross-link m/z 730.65 ($z = 4$) proposed structures with fragment ion evidence; Cross-linker bridges are indicated in red. | 151 |
| 4.35 | Interpeptide PFA calmodulin-melittin cross-link m/z 730.65 ($z = 4$) MS/MS spectra; Cross-linker bridges are indicated in red. | 152 |
| 4.36 | Reaction mechanisms of the PFA modification (i) and cross-linking formation (ii) of melittin R24 to calmodulin R126 (a), calmodulin K77 to Q3 (b), melittin G1 to calmodulin Y99 (c) and melittin G1 to calmodulin Q8 (d); Reactive regions are highlighted in red. R_1 and R_2 , and R_3 and R_4 , represent arbitrary sections of the modified and cross-linked proteins, respectively. | 155 |
| 5.1 | Identified cross-links were mapped on the Ca^{2+} -free unbound (a) and bound-state (b) calmodulin conformation. Orange and grey lines represent inter-residue distances that do and do not agree with maximum cross-linker distances, respectively. Cross-linking sites are highlighted in red and calmodulin C and N terminal domains are colored in blue and teal, respectively. | 182 |

List of Figures

| | | |
|-----|--|-----|
| 5.2 | Calmodulin-melittin binding structures (two views of the same structure) proposed by cross-linking distance constraints that supported (a) parallel (yellow lines) and (b) antiparallel (orange lines) binding;. Orange/yellow and grey lines represent inter-residue distances that do and do not agree with maximum cross-linker distances, respectively. Cross-linking sites are highlighted in red, melittin is shown in purple, calmodulin C and N terminal domains are shown in blue and teal, respectively. | 186 |
| 5.3 | Calmodulin-melittin binding structures proposed by EDC (a) and PFA (b) distance constraints; W19 on melittin is highlighted in yellow. Orange and yellow lines represent inter-residue distances that support antiparallel and parallel binding, respectively. Cross-linking sites are highlighted in red, melittin is shown in purple, calmodulin C and N terminal domains are shown in blue and teal, respectively. | 187 |
| 6.1 | QStar acquired MS (middle) and MS/MS (bottom) spectrum of interpeptide BS ³⁺ calmodulin-melittin cross-link m/z 896.93 ($z = 2$) | 193 |
| 6.2 | Impact II acquired MS(middle) and MS/MS (bottom) spectrum of interpeptide BS ³⁺ calmodulin-melittin cross-link m/z 896.93 ($z = 2$); Proposed structure with the sequence fragment ion evidence indicated on the backbone of the peptide that corresponds to type 2 ions only (top). | 194 |
| 6.3 | QStar acquired MS (middle) and MS/MS (bottom) of interpeptide PFA calmodulin cross-link m/z 603.24 ($z = 3$); Proposed structure with the sequence fragment ion evidence indicated on the backbone of the peptide that corresponds to type 3 ions only (top). | 196 |
| 6.4 | Impact II acquired MS (middle) and MS/MS (bottom) of interpeptide PFA calmodulin cross-link m/z 666.33 ($z = 3$); Proposed structure with the sequence fragment ion evidence indicated on the backbone of the peptide that corresponds to type 3 ions only (top). | 198 |

List of Figures

| | | |
|------|---|-----|
| 6.5 | (a) Monoisotopic peak assigned for MS Signal m/z 539.07 ($z = 5$) by DeconMSn, SNAP, SumPeak and Apex peak picking methods as indicated by the blue, red, purple and teal arrows, respectively. (b) Summary of peak picking methods for a PFA treated calmodulin-melittin sample | 203 |
| 6.6 | (a) Venn Diagram of MS/MS verified EDC cross-linked species identified by each software (StavroX and pLink) and manual method, two methods (two region overlap), all methods (center region overlap); (b) The calculated monoisotopic mass, cross-link peptide sequence (cross-linked residues highlighted in red), m/z , experimental monoisotopic mass, and mass accuracy for species identified by StavroX and pLink. | 211 |
| 6.7 | An example of a EDC cross-linked species ($m/z = 580.84$, $z = 2$) and its annotated MS/MS spectrum from StavroX | 212 |
| 6.8 | (a) Venn Diagram of MS/MS verified sulfoDST cross-linked species identified by each software (StavroX and pLink) and manual method, two methods (two region overlap), all methods (center region overlap); (b) The calculated monoisotopic mass, cross-link peptide sequence (cross-linked residues highlighted in red), m/z , experimental monoisotopic mass, and mass accuracy for species identified by StavroX and pLink. Cross-links that agreed with the manual detection are highlighted in purple. | 214 |
| 6.9 | An example of a sulfoDST cross-linked species ($m/z = 567.53$, $z = 4$) and its annotated MS/MS spectrum from StavroX | 215 |
| 6.10 | (a) Venn Diagram of MS/MS verified BS ³ cross-linked species identified by each software (StavroX and pLink) and manual method, two methods (two region overlap), all methods (center region overlap); (b) The calculated monoisotopic mass, cross-link peptide sequence (cross-linked residues highlighted in red), m/z , experimental monoisotopic mass, and mass accuracy for species identified by StavroX and pLink. Cross-links that agreed with the manual detection are highlighted in purple. | 218 |

List of Figures

| | | |
|------|---|-----|
| 6.11 | An example of a BS ³ cross-linked species ($m/z = 890.75$, $z = 4$) and its annotated MS/MS spectrum from StavroX; In the cross-linked structure shown on top, the “m” in the peptide sequence refers to M ^(ox) , i.e. an oxidized M residue. | 219 |
| 6.12 | (a) Venn Diagram of MS/MS verified sulfoEGS cross-linked species identified by each software (StavroX and pLink) and manual method, two methods (two region overlap), all methods (center region overlap); (b) The calculated monoisotopic mass, cross-link peptide sequence (cross-linked residues highlighted in red), m/z , experimental monoisotopic mass, and mass accuracy for species identified by StavroX and pLink. Cross-links that agreed with the manual detection are highlighted in purple. | 221 |
| 6.13 | An example of a sulfoEGS cross-linked species ($m/z = 936.46$, $z = 4$) and its annotated MS/MS spectrum from StavroX | 222 |
| 6.14 | (a) Venn Diagram of MS/MS verified PFA cross-linked species identified by each software (MeroX/StavroX and pLink) and manual method, two methods (two region overlap), all methods (center region overlap); (b) The calculated monoisotopic mass, cross-link peptide sequence (cross-linked residues highlighted in red), m/z , experimental monoisotopic mass, and mass accuracy for species identified by MeroX/StavroX and pLink. Cross-links that agreed with the manual detection are highlighted in purple. | 227 |
| 6.15 | The proposed structure of PFA cross-linked species ($m/z = 701.01$, $z = 6$) with insufficient MS/MS evidence from StavroX; In the cross-linked structure sequence, R ⁽⁺¹²⁾ , K ⁽⁺¹²⁾ , and M ^(ox) are denoted as “&”, “\$” and “m”, respectively. | 228 |
| 6.16 | The MS/MS spectra of PFA cross-linked species ($m/z = 701.01$, $z = 6$) with insufficient MS/MS evidence from StavroX; In the cross-linked structure sequence, R ⁽⁺¹²⁾ , K ⁽⁺¹²⁾ , and M ^(ox) are denoted as “&”, “\$” and “m”, respectively. | 229 |

List of Figures

| | | |
|------|--|-----|
| 6.17 | An example of a confirmed PFA cross-linked species ($m/z = 606.58$, $z = 4$) and its annotated MS/MS spectrum from MeroX. In the dotted box, the fragment ion evidence of the same cross-linked structure ($m/z = 484.46$, $z = 5$) | 230 |
| 6.18 | An example of a confirmed PFA cross-linked species ($m/z = 588.52$, $z = 5$) and its annotated MS/MS spectrum from MeroX; In the dotted box, the fragment ion evidence of the same cross-linked structure ($m/z = 588.52$, $z = 5$) is shown. In the cross-linked structure sequence, $R^{(+12)}$ is denoted as “&”. | 231 |
| A.1 | Excel Code for Elimination | 270 |
| A.2 | Mathematica Code for Possible cross-linked Species | 271 |
| A.3 | Mathematica Code for Candidate Cross-linked Species | 272 |
| A.4 | Collision Energy Table for Bruker Impact II LC-MS/MS | 273 |

List of Abbreviations

General Abbreviations

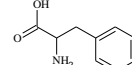
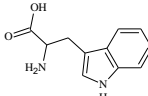
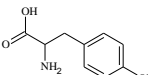

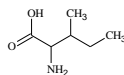
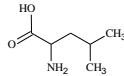
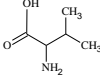


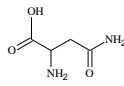
| | |
|-----------------|--|
| PFA | formaldehyde |
| NMR | nuclear magnetic resonance |
| SDS-PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| HPLC or LC | high performance reverse-phase chromatography |
| MS | mass spectrometry |
| MS/MS | tandem mass spectrometry |
| MS ⁿ | multistage mass spectrometry |
| CID | collision-induced dissociation |
| ECD | electron capture dissociation |
| ETD | electron transfer dissociation |
| ESI | nanoelectrospray ionization |
| MALDI | matrix-assisted laser desorption/ionization |
| QqTOF | quadrupole time of flight mass spectrometers |
| FT-ICR | Fourier transform ion cyclotron resonance |
| TOF | time of flight mass spectrometer |
| QStar | ABI QStar XL quadrupole time of flight mass spectrometer |
| Impact II | Bruker Impact II quadrupole time of flight mass spectrometer |
| <i>m/z</i> | mass to charge ratio |
| [M] | monoisotopic mass |

General Abbreviations

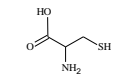
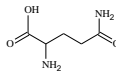
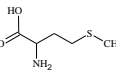
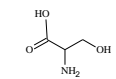
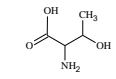
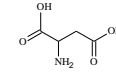
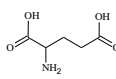
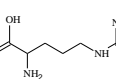
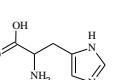
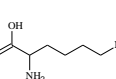
| | |
|-----------------|---|
| EDTA | Ethylenediaminetetraacetic acid |
| EDC | 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride |
| sulfoNHS | N-hydroxysulfosuccinimide |
| sulfoDST | disulfosuccinimidyl tartrate |
| BS ³ | bis(sulfosuccinimidyl) suberate |
| sulfoEGS | ethylene glycol bis(succinimidyl succinate) |
| RNaseA | Ribonuclease A |
| RNaseS | Ribonuclease S |
| SNAP | sophisticated numerical annotation procedure |
| S/N | signal to noise ratio |
| DOM | degree of modification |
| SCX | strong cation exchange |
| SEC | size exclusion chromatography |
| IMS | ion mobility spectrometry |

Amino Acids[1]

Amino Acids[1]

| Name | Letter Symbol | Molecular Weight | pKa (Side chain) | Structure |
|---------------|---------------|------------------|------------------|---|
| Phenylalanine | F | 165.19 | - |  |
| Tryptophan | W | 204.23 | - |  |
| Tyrosine | Y | 181.19 | 10.10 |  |
| Alanine | A | 89.09 | - |  |
| Isoleucine | I | 131.17 | - |  |
| Leucine | L | 131.17 | - |  |
| Valine | V | 117.15 | - |  |
| Glycine | G | 75.07 | - |  |
| Proline | P | 115.13 | - |  |
| Asparagine | N | 132.12 | - |  |

Amino Acids[1]

| Name | Letter Symbol | Molecular Weight | pKa (Side chain) | Structure |
|---------------|---------------|------------------|------------------|---|
| Cysteine | C | 121.16 | 8.14 |  |
| Glutamine | Q | 146.15 | - |  |
| Methionine | M | 149.21 | - |  |
| Serine | S | 105.09 | - |  |
| Threonine | T | 119.12 | - |  |
| Aspartic Acid | D | 133.1 | 3.71 |  |
| Glutamic Acid | E | 147.13 | 4.15 |  |
| Arginine | R | 174.2 | 12.10 |  |
| Histidine | H | 155.16 | 6.04 |  |
| Lysine | K | 146.19 | 10.67 |  |

Typical Protein Modifications

| Symbol | Modification | Mass Shift (Da) |
|----------------------|---------------------------|-----------------|
| (ac) | Acetylation | + 42.01 |
| (ox) | Oxidation | +15.99 |
| (am) | Amidation | -0.98 |
| (dm) | Deamidation | +0.98 |
| (tm) | Trimethylation | +42.06 |
| (cm) | Carbamidomethylation | +57.02 |
| -H ₂ O | Water loss | - 18.02 |
| -NH ₃ | Ammonia loss | -17.02 |
| -CH ₃ SOH | Methanesulfenic Acid loss | -64.11 |

Acknowledgements

I would like to thank my supervisor, Professor Juergen Kast for providing me with this project. I really appreciate how accommodating and flexible Dr. Kast has been throughout my Ph.D education, allowing me to think independently and boldly as a researcher. I would like to acknowledge my supervisory committee for always being very encouraging, supportive and extremely understanding. I am indebted to the University of British Columbia for awarding me with a Four-Year Doctoral Fellowship, International partial tuition award and Gladys Estella Laird Research Fellowship to fund my Ph.D. There are numerous individuals that have gone out of their way to guide me. Dr. Nikolay Stoynov not only patiently provided me with the bulk of mass spectrometric data that had pushed my Ph.D forward but always tried his very best to go out of the way to accommodate my several requests even if it meant transferring large data files remotely. Finally, he spent numerous days setting up and running strong cation exchange and size exclusion chromatography experiments for me solely out of his kindness. I also cannot thank Jason Rogalski enough for always going out of the way to accommodate sample requests for mass spectrometric analysis no matter how challenging. Jason has also helped meticulously proofread my proposal and other writing early in my Ph.D. Finally, whether a part of our lab or not he has always provided me with support, such as taking the time out to answer my numerous questions, and his encouragement has really helped me to reach the final stages of my Ph.D. I would also like to thank Shujun Lin who has provided me with training in mass spectrometry instrumentation, sample preparation and has run numerous samples for strong cation exchange chromatography and mass spectrometric analysis. Another individual that I am extremely grateful to is Ru Li, a fellow Ph.D student in our lab. She taught me skills in running SDS-PAGEs, sample preparation etc that have been extremely vital for moving my project forward. Her hardworking and dedicated, yet extremely

Acknowledgements

humble and patient nature has always inspired me. Having her support through several obstacles faced throughout the Ph.D has truly helped me. Several past lab members have also made an impact in the pursuit of my Ph.D. A fellow lab member, Xuan Ding showed me the basics of the lab techniques/protocols for the experiment and helped me transition into taking over the project. Jane O'Hara and Cordula Klockenbusch, former Post-Doctorates, have helped me proofread and revise various written materials throughout the early years of my Ph.D. Another fellow lab member, Davin Carter, had always enthusiastically helped me revise and format presentations and written materials for conferences. Iris Egner, a former Post-Doc, patiently helped me refine my thesis outline, offered advice for presentations, and taught me several crucial strategies for improving SDS-PAGE. Finally, I would like to thank the rest of my fellow lab members past and present for their support and encouragement. My interest in chemistry was initiated by my high school chemistry teacher, Mr. Timothy Lattanzio who provided an unforgettable foundation in chemistry that I have deeply relied on throughout my undergraduate and graduate schooling. I also will never forget the kindness of my physical chemistry professor, Professor Geoffrey Hutchison, who not only gifted his students with a strong foundation in chemistry but went out of the way to direct me to my undergraduate research supervisor and supported my graduate school ambitions with reference letters and guidance. My undergraduate research supervisor, Professor Seth Horne, was one of the main reasons I was able to pursue a Ph.D. He not only meticulously provided me with my initial training in chemistry research, but guided and supported me through the transition from my final undergraduate year to Ph.D by providing reference letters, advice and encouragement. Finally, I only hope that one day I can do even a little bit of justice to the God-sent family that I am extremely blessed to have. My parents and my sister's undying positive energy, faith and meticulous guidance fueled my Ph.D. Their strong passion for science and experiences completing Ph.Ds in physics and chemistry had inspired me to embark on my own Ph.D journey. This Ph.D education has not only taught me skills to study a specialized field but has provided me with a new perspective and training for life.

Dedication

To My Parents and Sister, Vanita

Chapter 1

Introduction

1

1.1 Structural Proteomics

Proteins are dynamic entities constantly altering their structures based on their functions inside and outside the cell. The fundamental aim of structural proteomics is defining three dimensional protein structures. Protein conformation and reactivity is dictated by its sequence. Protein function is governed by the interactions of protein structures in stable multi-subunit complexes as well as more transient assemblies. Such interactions are responsible for carrying out a multitude of tasks in cells from binding to small molecules during storage, transport and cellular signaling to serving as molecular switches, structural supports and catalysts [3]. Irregularities in protein structures at every level caused by mutating amino acids, denaturing or aggregating structures, or non-specific binding can indicate the presence of disease [4–6]. A large collection of high resolution three dimensional protein structures have been solved by X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and electron microscopy, many of which are housed in the Protein Data Bank (PDB). No other technique has been able to offer the same precision and detail of protein structural determination by X-ray crystallography. However, X-ray crystallography is limited to proteins for which crystals can be obtained. At body temperature, there is sufficient energy to cause protein motions that cannot be measured by X-ray crystallography, which represents most protein structures as static, average structures. A major advantage of NMR is that it can measure protein motions that occur down to the ps range and provide information

¹The Introduction chapter was adapted from the author's previous publication[2]

regarding protein kinetics, although it requires a large amount of protein. Nonetheless, these methods are unsuitable to characterize protein complexes or proteins with low solubility, due their large size, or reduced stability. [7–10]. MS has an advantage over X-ray crystallography and NMR with its attomolar sensitivity, rapid measurement and ability to retain physiological conditions of proteins even though it cannot match the high resolution structural data of these techniques[11]. Therefore, combining the high resolution capabilities of X-ray crystallography and NMR with the sensitivity, speed and gentle nature of MS can serve as a powerful tool to examine proteins and protein complexes.

1.2 Mass Spectrometry in Structural Proteomics

MS measures the mass to charge ratio (m/z) of an ionized molecular species in a sample. The analyte is ionized via an ionization source producing ions that are accelerated through an electric field and organized based on their m/z in a mass analyzer. Finally, the intensity of each m/z is recorded by a detector. The mass spectrum is a plot of the intensity versus m/z . MS provides the means to study the structure of proteins as dynamic entities, which is comparable to that observed in a physiological environment in solution, making it the primary technology implemented in structural proteomics.

Bottom-up and top-down methodologies based on the MS analysis of peptide digests and intact proteins, respectively, have both been employed although bottom-up is more widely implemented[12].

In a typical bottom-up proteomics experiment (see Figure 1.1), proteins are first isolated from cell lysates or biological species. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is routinely utilized for separating complex protein mixtures. SDS unfolds proteins and adds negatively charged SDS sulfate groups approximately proportional to its molecular weight. Samples are loaded into a polyacrylamide gel matrix and a high voltage is applied. The protein molecules migrate to a distance inversely proportional to their size [13].

1.2. Mass Spectrometry in Structural Proteomics

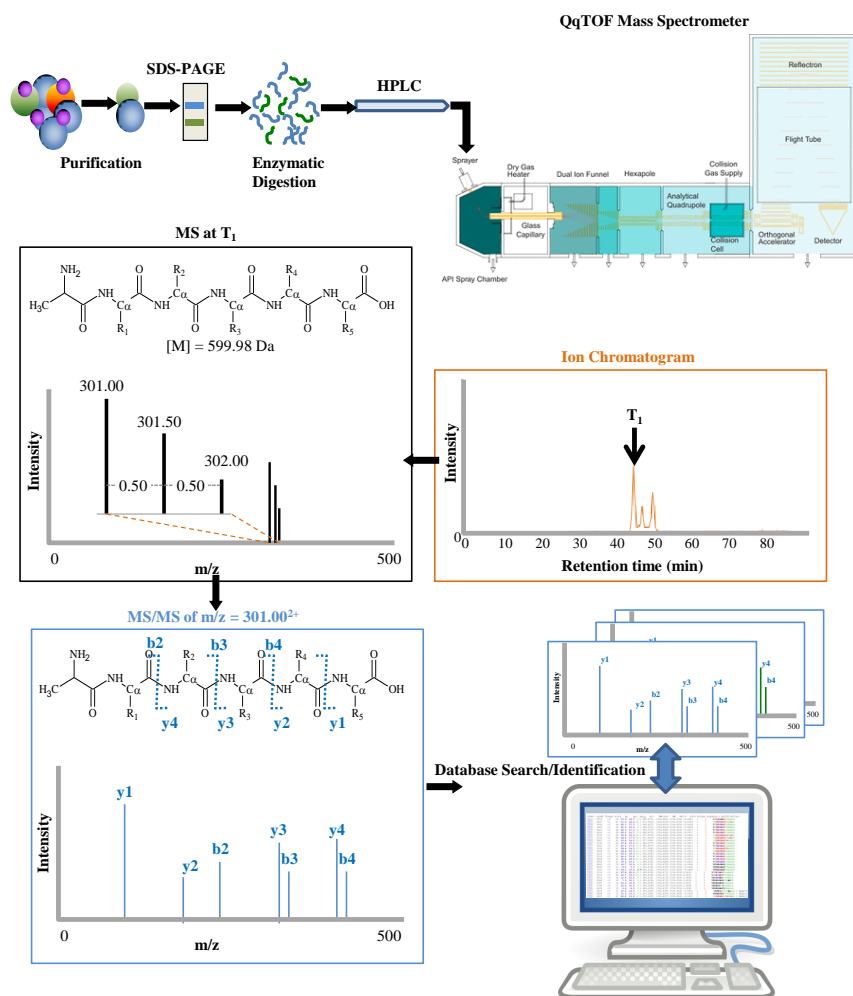


Figure 1.1: A summary of the bottom-up proteomic strategy, in which proteins are purified, separated via SDS-PAGE and enzymatically digested into peptides, which are separated through HPLC and eluted into the mass spectrometer. The schematic diagram of a Bruker Impact II QqTOF mass spectrometer is shown on the top right (adapted from reference [14] with permission, <https://creativecommons.org/licenses/by/4.0/>). At each time point an MS spectrum of the ions eluted is plotted. For example at time = T_1 , a species with a m/z of 301.00 ($z = 2$) eluted and the signal appears as isotopic cluster such that the charge is the inverse of the difference between each isotopic peak i.e. 1 divided by 0.5. The MS/MS of precursor ion produces fragment ion signals ($b_1, b_2, \dots, y_1, y_2, \dots$) which are typically the result of the cleavage of peptide bonds in CID. MS/MS spectra and MS spectra are matched to theoretical databases for identifying and sequencing peptides.

Trypsin is a common choice of enzyme due to its high specificity in cleaving mainly at the carboxyl side of lysine or arginine residues (except when either is followed by a proline), which are abundant in proteins. Following the enzymatic digestion of proteins in gel, the resulting peptides are extracted and separated through reversed-phase high performance liquid chromatography (HPLC or also referred to as LC). Peptides elute in order of their hydrophobicity into the mass spectrometer and MS signals or m/z of precursor ions are recorded for each elution time point. The number of ions (intensity) is plotted versus retention time (time each component elutes from LC column) to construct ion chromatograms. Soft ionization methods such as nanoelectrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) can ionize biomolecules without producing significant damage. Ionized tryptic peptides are normally protonated at the terminal lysine or arginine residue and the peptide N-terminus, giving peptide ions a +2 charge. Additional basic residues within the peptide such as histidine or missed cleaved lysines or arginines may result in higher charge states of the peptide. Peptide signals appear as an isotopic cluster and the charge can be calculated by determining the inverse difference between each isotopic peak. The monoisotopic masses ($[M]$) of each species can be derived from the m/z and charge. For example, a doubly charged species at m/z 301.00 would correspond to $[M+2H]^{2+}$ such that $[M+2(1.01\text{Da})]/2 = 301.00$, which gives a $[M] = 599.98$ Da [15][16]. Therefore, a resolution that can clearly depict isotopic patterns for determining accurate monoisotopic masses is crucial. This makes time of flight (TOF) mass spectrometers particularly favorable for such analyses with their high resolution capabilities ($\sim 100,000$) [14]. In TOF-MS, ions are accelerated through a flight tube and flight times are measured, which are proportional to the square root of m/z . Ions of different kinetic energies are corrected by a reflectron such that ions with equal m/z arrive at the detector at the same time. Fourier transform ion cyclotron resonance (FT-ICR) mass analyzers also provide high resolution measurements ($\sim 100,000$). In FT-ICR MS, a magnetic field and an orthogonal oscillating electric field accelerate ions in a circular motion. The time stable superconducting magnetic field allows for the high mass accuracy. The frequency of the ions is measured to derive

1.2. Mass Spectrometry in Structural Proteomics

the m/z , which is independent of ion speed and m/z values can be measured at the same time, allowing for high S/N [17]. Orbitrap mass analyzers operate similarly to FT-ICR except that the absence of the magnetic field has allowed for increased mass resolution ($\sim 300,000$). However, unlike TOF MS, Orbitrap's resolving power is compromised for speed in the MS/MS acquisition mode [18, 19].

Out of all the precursor ion signals recorded, abundant signals usually based on user defined parameters are selected within a chosen isolation window of m/z values for a second stage of MS (tandem MS or MS/MS). In quadrupole time of flight mass spectrometers (QqTOF), ions are selected in the first quadrupole (Q1), subjected to MS/MS in the second quadrupole or collision cell (q2) and separated in the TOF based on m/z . Quadrupoles are composed of four rods with a DC and RF voltage applied across each pair of rods producing a varying electric field. Each electric field generated only allows ions of specific m/z values that have stable oscillations to be transmitted. In the collision cell, collision induced dissociation (CID) occurs in which precursor ions hit a stationary inert gas and break into smaller fragment ions[20]. When collision energy is equal to bond energy, the bond breaks. The most commonly observed peptide CID fragmentation occurs at the peptide bond such that the charge can be retained on the N-terminal or C-terminal of the peptide, producing b ions or y ions, respectively. Fragmentation occurring between the alpha carbon and carbonyl carbon or the nitrogen and alpha carbon is less likely under CID. Additional fragmentation of fragment ions via multiple stages of MS/MS (MS^n) can provide more precise details about sequences and modified amino acids. Also, with every subsequent MS, the signal intensities decrease and MS^n relies on strong precursor ion signals. In-source fragmentation techniques have been combined with triple quadrupole mass spectrometry for pseudo MS^3 [21].

Other fragmentation mechanisms utilized include electron capture dissociation (ECD) and electron transfer dissociation (ETD). ECD was designed for Fourier Transform where as ETD was designed for Quadrupole-based mass analyzers. ECD and ETD differ from CID in that fragmentation occurs at low energy producing c and z ions from the fragmentation of the C_α -N bond. Advantages of ECD and ETD over CID include the preservation of post-translational modifications and the fragmentation of disulfide bonds. However, ECD and ETD lack efficiency when

fragmenting ions with low charge states (+1, +2). Recent combinations of CID and ECD/ETD have demonstrated to be a powerful tool to obtain complementary MS/MS fragment evidence [22]

Peptides can be identified by matching their monoisotopic masses to MS signals and confirming their sequence by matching theoretical peptide fragment ions to MS/MS signals. Large databases of proteins and MS or MS/MS based software such as GPM [23], Mascot [24] and SEQUEST [25] are available for automatic protein/peptide identification.

1.3 Mass Spectrometry and Chemical Cross-linking

MS has fostered an immense growth in proteomics research. LC-MS coupled with affinity purification analyses identifies components of protein complexes and networks [26]. Cutting-edge instrumentation in MS/MS maps proteins and their modifications at the amino acid residue level for high resolution geometry [27]. Native ion mobility MS characterizes protein conformation by measuring its cross-sectional area [28, 29]. Imaging MALDI-MS can be used to investigate the spatial arrangement of protein structures in intact tissues. Approaches combining these MS technologies with chemical methods such as limited proteolysis, chemical surface modification and hydrogen-deuterium exchange, monitor the solvent accessibility of regions of a protein to observe conformational changes [12, 30, 31]. The enzymatic cleavage, modification, or deuterium exchange at particular amino acids directly correlates to the exposure of that respective region to the solvent and these can be used to determine structural changes [30, 31]. Chemical cross-linkers form covalent bonds in proteins, which preserve their cellular context and introduce distance constraints to map their structure.

Cross-linking occurs in two steps: first, the modification reaction and second, the cross-linking reaction. Upon cross-linking and enzymatic digestion of a protein complex, a mixture of unmodified, modified, intra-cross linked (cross-link of two residues within one peptide) and inter-cross linked (cross-link of two residues on two different peptides) peptides is produced (see Figure 1.2).

1.3. Mass Spectrometry and Chemical Cross-linking

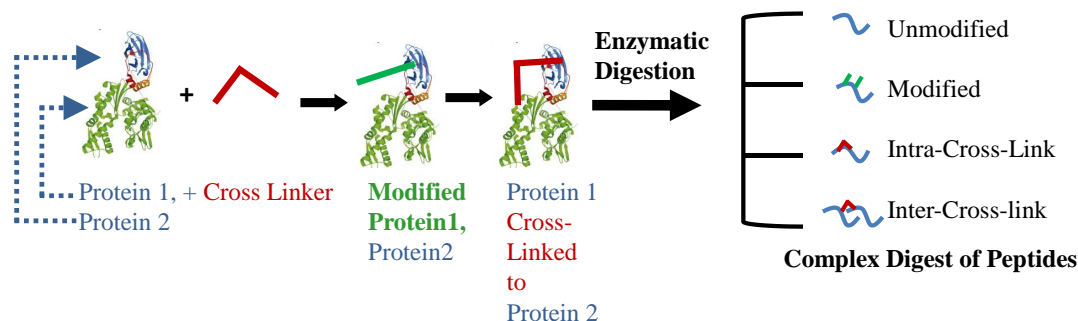


Figure 1.2: In general, cross-linkers introduce covalent bonds in proteins via a two-step reaction: modification of protein site 1 and cross-link formation between protein sites 1 and 2. Upon enzymatic digestion, a complex mixture of different types of peptides is produced.

Unmodified peptides can be used to identify the protein components participating in the quaternary interaction holding the complex together. Measuring distances between both intra and inter cross-linking sites on peptides can offer a low resolution picture of secondary and tertiary protein structures and monitor changes under different conditions. Localizing the cross-linking sites to specific amino acids can depict precise geometries and reaction interfaces of proteins at the primary structural level. Modified peptides can be used to observe fluctuations in conformation via solvent accessibility by tracking the variation in the degree of modification as a function of projected external factors. The mass of a cross-linked species should equal the sum of the masses of each component peptide, bridge, and any additional modification. Over 100 different cross-linkers are commercially available. Different cross-linkers of various sizes, solubility, lengths and reactivity may be suitable depending on the location of the complex, the reactive sites and intermolecular distances of protein components, or the information desired, such as structural or mechanistic properties. For proteomic research, amine-reactive cross-linkers are most widely used due to the abundance of primary amine (primarily N-terminal and lysines) that are also more solvent and reagent accessible than the hydrophobic sites that are buried within the protein structure. Non-specific cross-

linking technologies such as photoactivatable cross-linkers that form linkages that are mostly independent of the type of amino acid upon light irradiation can be applied to various proteins regardless of amino acid composition [12]. Below, the cross-linking chemistries explored in this study are described.

1.3.1 Types of Chemical Cross-linkers

1.3.1.1 N-hydroxy Succinimide Ester Cross-linkers

N-hydroxy succinimide (NHS) esters are common homobifunctional cross-linkers, or cross-linkers with the same reactive groups on either end. They form cross-links specifically between primary amino groups present in lysine (K) amino acid side chains and the N-terminus. Since lysine residues are abundant, and are accessible at the surface in proteins, NHS esters are widely used for protein cross-linking. NHS ester cross-linkers are generally cellular membrane permeable, water insoluble and are routinely used to stabilize intracellular protein complexes. With the addition of sulfonyl groups, these cross-linkers become water soluble and membrane impermeable and are used to characterize cellular surface proteins. Various lengths of NHS ester cross-linkers are commercially available. Examples of water soluble, cellular impermeable cross-linkers include disulfosuccinimidyltartrate (sulfoDST), bis(sulfosuccinimidyl) suberate (BS³), and ethylglycol bis(sulfosuccinimidylsuccinate) (sulfoEGS) with cross-link bridge lengths of 6, 12 and 16Å, respectively. Figure 1.3, 1.4, and 1.5 depicts the cross-linking reactions for sulfoDST, BS³, and sulfoEGS, respectively, which share similar cross-linking mechanisms. One major drawback of NHS ester cross-linkers is that they rapidly hydrolyze under cross-linking reaction conditions (pH > 7, 25-37 °C) with a half life on the scale of tens of minutes. This restricts cross-linking to short reaction times, making it difficult to increase product yield with larger reaction times in dilute protein solutions. Also, since NHS esters react with only basic sites on the protein, the overall positive charge is reduced, which can induce conformational changes, hinder trypsin cleavage, and reduce the ionization efficiency in MS. Finally, the longer the cross-linker bridge, the higher the probability of it existing within the distance of two residues if its structurally flexible. Lysine's flexible long side chain makes it likely to randomly move within the cross-linking bridge dis-

1.3. Mass Spectrometry and Chemical Cross-linking

tance of these larger cross-linkers. Therefore, distinguishing between specific and non-specific cross-linking with long cross-linker bridges can be challenging [32].

1.3. Mass Spectrometry and Chemical Cross-linking

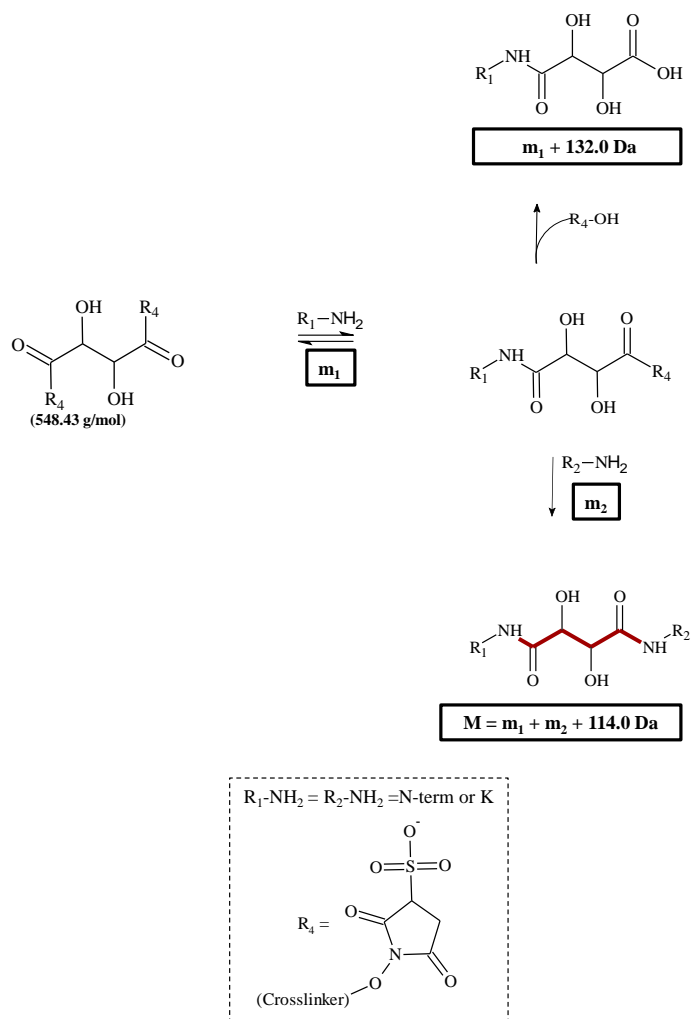


Figure 1.3: In the sulfoDST cross-linking reaction scheme, protein site 1 (mass = m_1) reacts with the sulfoDST to form a modification which immediately reacts with protein site 2 (mass = m_2) to form a cross-linking bridge (highlighted in red). A competing hydrolysis reaction product ($m_1 + \text{bridge} + H_2O$) can also occur instead of the cross-linked product ($M = m_1 + m_2 + \text{bridge}$). R_1 , R_2 and R_4 are defined in the dotted box.

1.3. Mass Spectrometry and Chemical Cross-linking

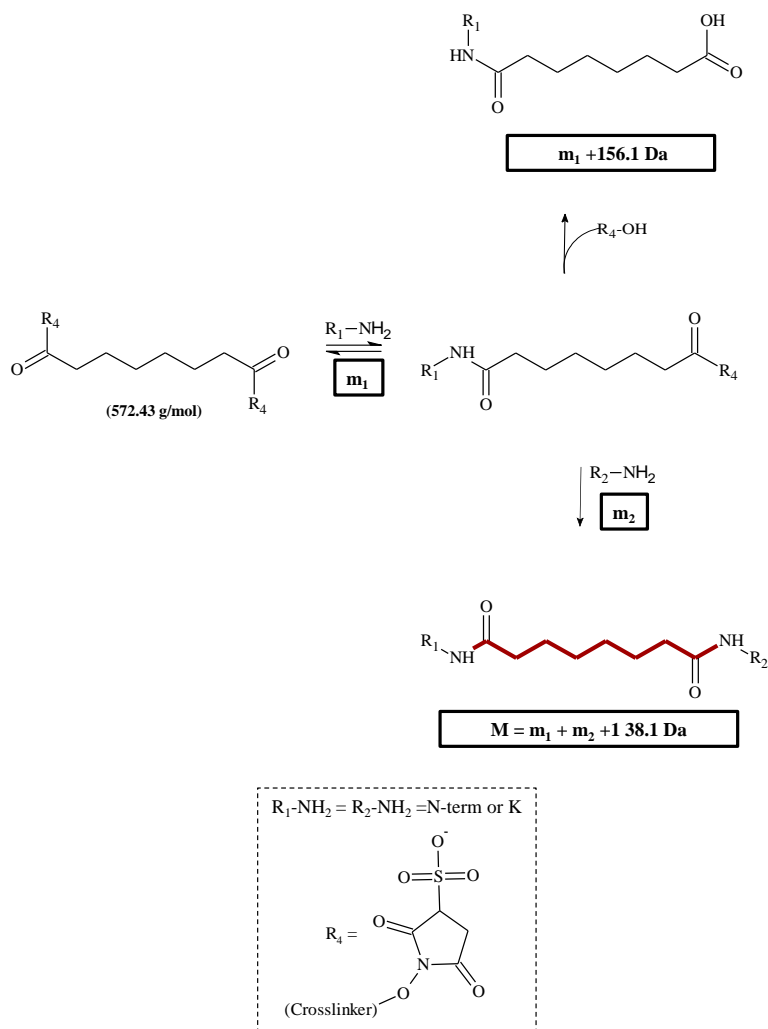


Figure 1.4: In the BS³ cross-linking reaction scheme, protein site 1 (mass = m_1) reacts with the sulfoDST to form a modification which immediately reacts with protein site 2 (mass = m_2) to form a cross-linking bridge (highlighted in red). A competing hydrolysis reaction product (m_1 +bridge + H₂O) can also occur instead of the cross-linked product ($M = m_1 + m_2 +$ bridge). R₁, R₂ and R₄ are defined in the dotted box.

1.3. Mass Spectrometry and Chemical Cross-linking

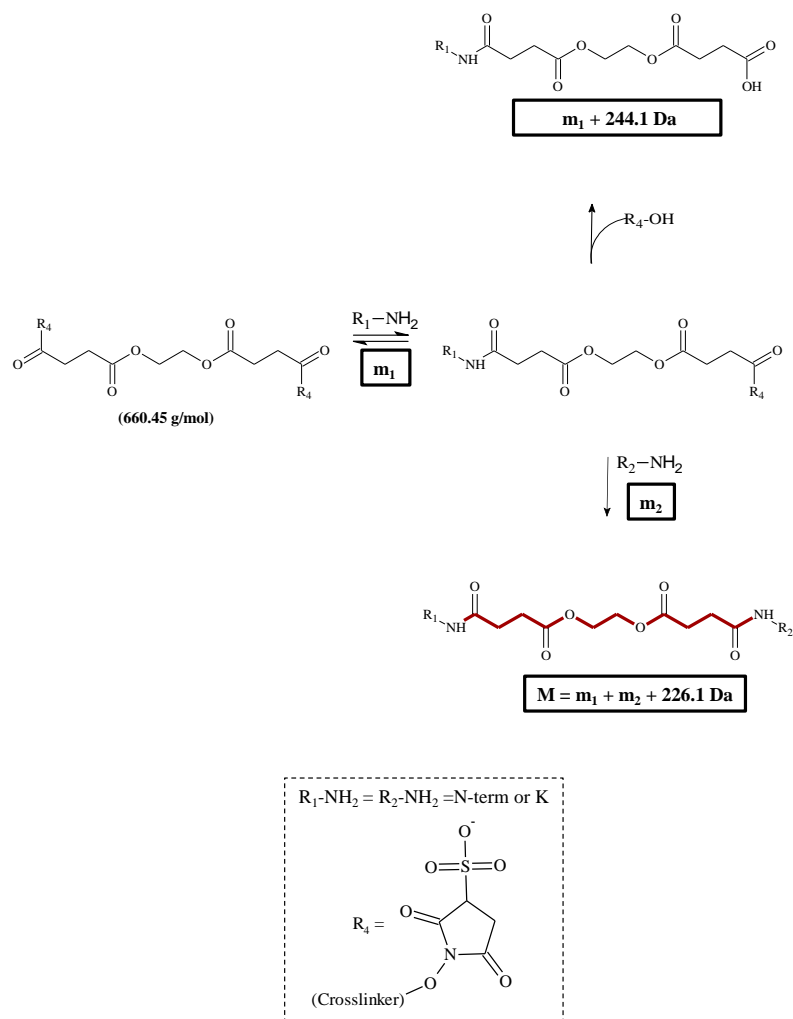


Figure 1.5: In the sulfoEGS cross-linking reaction scheme, protein site 1 (mass = m_1) reacts with the sulfoDST to form a modification which immediately reacts with protein site 2 (mass = m_2) to form a cross-linking bridge (highlighted in red). A competing hydrolysis reaction product ($m_1 + \text{bridge} + \text{H}_2\text{O}$) can also occur instead of the cross-linked product ($M = m_1 + m_2 + \text{bridge}$). R_1 , R_2 and R_4 are defined in the dotted box.

1.3.1.2 Zero-length Cross-linkers

Zero-length cross-linkers add a cross-linker bridge the length of a single bond and thus join very close proximity protein sites together, despite the bulky size of the cross-linker reagent itself. A common zero-length cross-linking strategy uses 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) in conjunction with N-hydroxysulfosuccinimide (sulfoNHS). As Figure 1.6 shows, these heterobifunctional (i.e. cross-linkers with two different reactive groups) cross-linkers form peptide bonds between primary amino and carboxylic acid groups (glutamic acids (E) aspartic acids (D) and C-terminus), inducing an overall mass shift of -18.02 Da i.e. form cross-links via a condensation reaction. EDC first is attacked by the carboxylate oxygen in the a carboxylic group on the first protein site (m_1) to form an unstable intermediate (O-acylisourea). This intermediate reacts with sulfoNHS to form an amine-reactive sulfoNHS ester, This intermediate then reacts with a primary amino group on the second protein site (m_2) to form a peptide bond. The formation of the cross-link thus should not change the overall net charge since both a negatively charged and positively charged group are neutralized. The O-acylisourea can also rearrange to form a N-acylisourea ($m_1 + 155$ Da) shown in the top of Figure 1.6. In this stable product, a positively charged modification replaces the negatively charged carboxylic group, which alters the overall net charge [33]. Due to the lack of an actual cross-link bridge, locating the cross-linking sites of zero-length cross-linkers is difficult via MS. These cross-linkers are water soluble, but are cellular membrane impermeable [34]. Cross-linking is performed at pH ~ 6.5 for the best results and a major drawback is the inefficiency of EDC/sulfoNHS² cross-linking under physiological pH conditions[33, 35].

²Note: Throughout the text the combined EDC and sulfoNHS cross-linking chemistry is referred to as “EDC” cross-linking.

1.3. Mass Spectrometry and Chemical Cross-linking

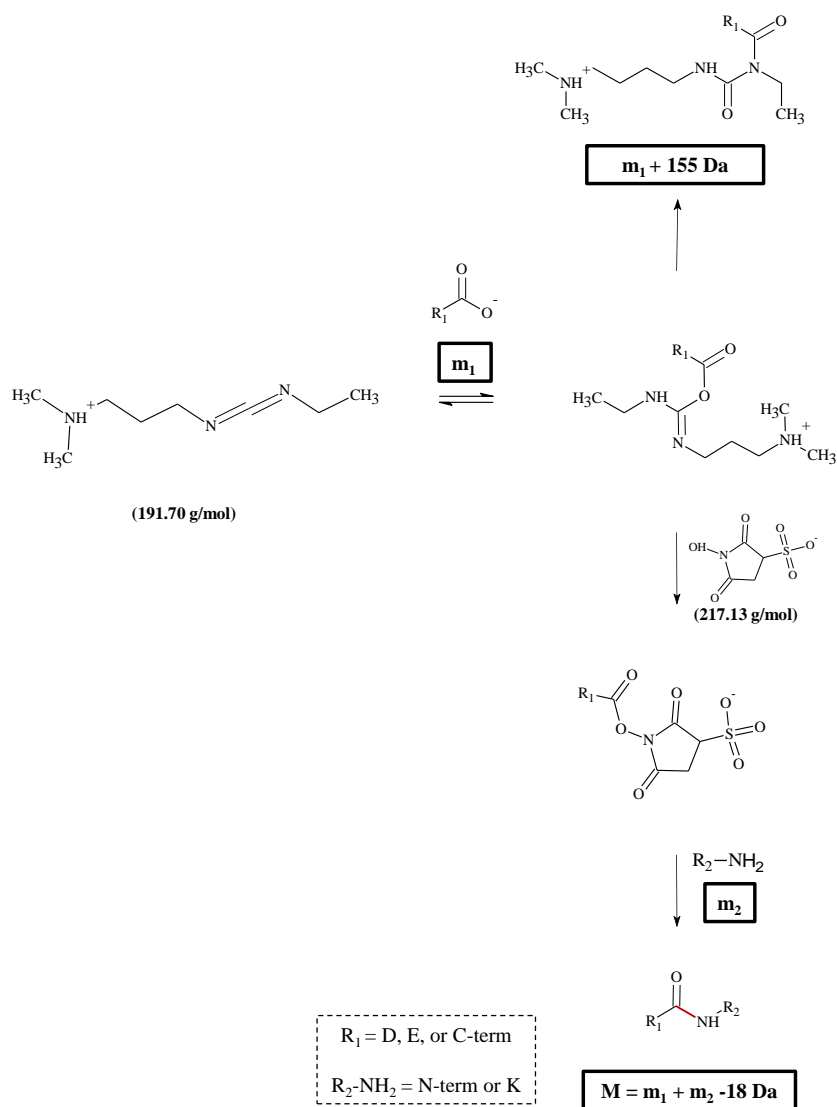


Figure 1.6: In the EDC/sulfoNHS cross-linking reaction scheme, protein site 1 (mass = m_1) reacts with the cross-linker. The intermediate can react with sulfoNHS to form an amine reactive sulfoNHS ester, which forms a cross-link with protein site 2 (mass = m_2). The intermediate can also produce a stable N-acylisourea ($m_1 + 155 \text{ Da}$). Cross-linker bridges are highlighted in red. R_1 and R_2 represent two protein sites and specific reactive amino acids are defined in dashed line boxes. Mass of each species are provided in solid line boxes.

1.3. Mass Spectrometry and Chemical Cross-linking

Formaldehyde (PFA)³ can be thought of as a pseudo zero length cross-linker due to its a relatively small size (2.3 Å). Figure 1.7 shows the general PFA reaction scheme. Figures 1.8 and 1.9 show the mechanisms and reaction schemes for potential PFA modification and cross-linking sites, respectively. Primary amino groups on the first protein site (mass = m_1) nucleophilically attack PFA's carbonyl carbon to form a methylol intermediate, adding +30 Da to the mass of the protein. Unlike those of NHS esters and EDC cross-linkers, PFA modification intermediates are stable in solution and have been detected via MS and MS/MS to probe accessibility in proteins [36]. The methylol intermediate can dehydrate into a stable Schiff base modified protein (mass = $m_2 + 12$ Da). Without water, the equilibrium between the methylol and Schiff base intermediates would shift toward the right. However, in solution, the equilibrium is unaffected by the amount of water. In proteins, this equilibrium would be affected by the accessibility of water, which is dictated by its structure. In addition, the equilibrium may be driven to the right if the Schiff base is involved in a subsequent reaction. The Schiff base intermediate reacts with the second protein site (mass = m_2), forming a methylene bridge between the protein sites and an overall mass shift of +12 Da is produced. PFA is both membrane permeable and water soluble and can form cross-links rapidly under physiological conditions, making it particularly useful for examining weakly associated protein complexes in their native environment. In contrast to other cross-linkers, PFA's semi-specific reactivity allows cross-link formation among a wide range of amino acids [37].

³Formaldehyde exists as paraformaldehyde, a polymerized form of formaldehyde, as a solid. Paraformaldehyde is dissolved in solution to obtain the formaldehyde reagent used for cross-linking. The non-polymerized formaldehyde is referred to as "PFA" throughout the text.

1.3. Mass Spectrometry and Chemical Cross-linking

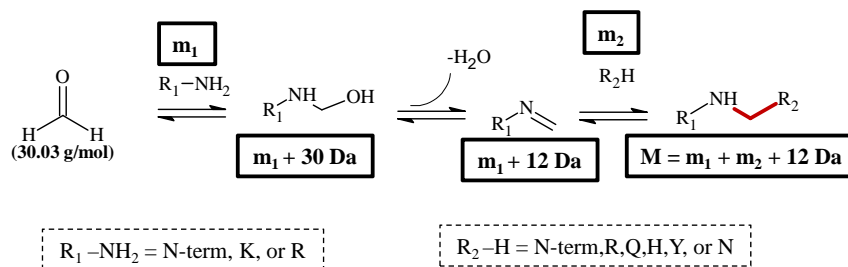
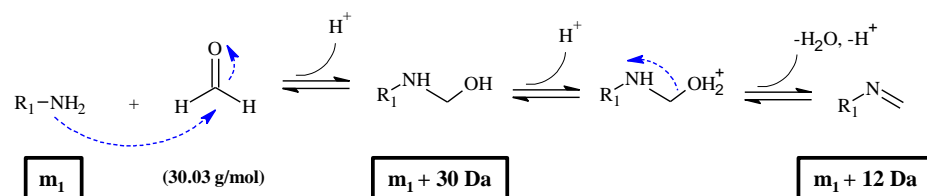


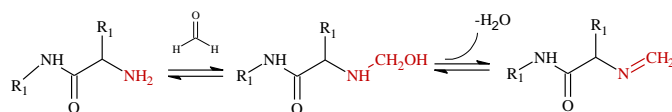
Figure 1.7: PFA reacts with protein site 1 (mass = m₁), which forms a methylol modification (mass = m₁ + 30 Da). This can dehydrate into a Schiff Base (mass = m₁ + 12 Da), which can continue to react with protein site 2 (mass = m₂) to form a methylene bridge. This produces a cross-linked species (mass = m₁ + 12 Da + m₂). Cross-linker bridges are highlighted in red. R₁ and R₂ represent two protein sites and specific reactive amino acids are defined in dashed line boxes. Mass of each species are provided in solid line boxes.

1.3. Mass Spectrometry and Chemical Cross-linking

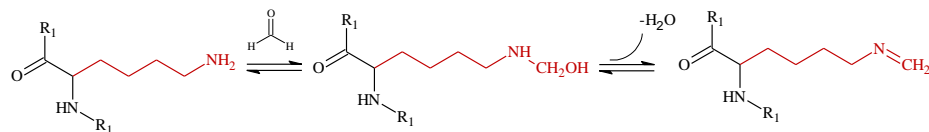
(a)



(b) $R_1 = \text{N-term}$



$R_1 = \text{K}$



$R_1 = \text{R}$

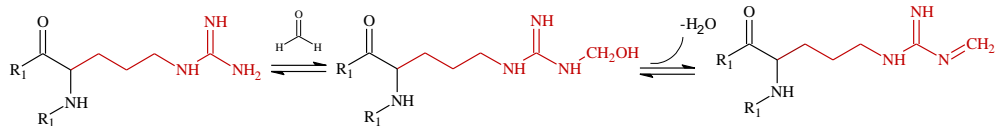
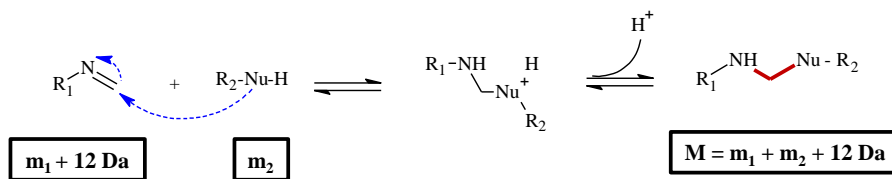


Figure 1.8: The PFA modification reaction generic mechanism (a) and for each reactive residue (b) (Adapted from reference [37], with permission)

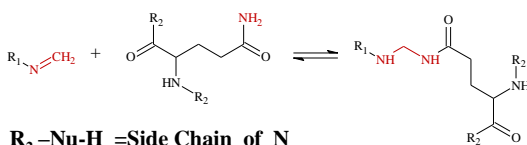
1.3. Mass Spectrometry and Chemical Cross-linking

(a)

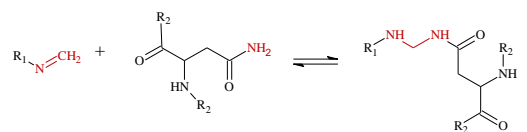


Nu = nucleophile in protein R₂

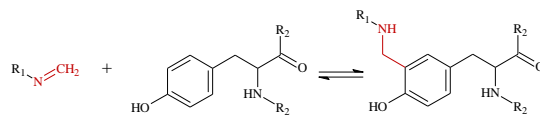
(b) R₂-Nu-H = Side Chain of Q



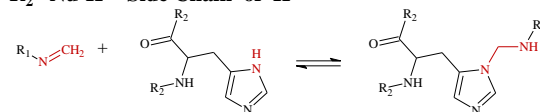
R₂-Nu-H = Side Chain of N



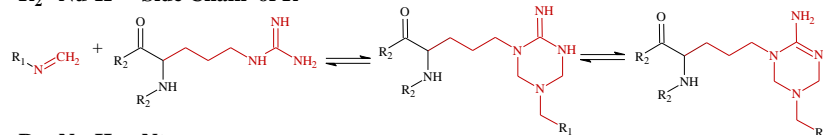
R₂-Nu-H = Side Chain of Y



R₂-Nu-H = Side Chain of H



R₂-Nu-H = Side Chain of R



R₂-Nu-H = Nterm

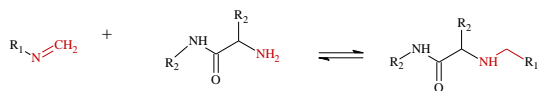


Figure 1.9: The PFA cross-linking reaction generic mechanism (a) and for each reactive residue (b) (Adapted from reference [37], with permission)

1.3.2 Identification of Cross-linking and Structure Elucidation

The identification of cross-linked species generally follows the bottom-up proteomics method. However, to aid the MS detection of the low abundant cross-linked species in the reaction mixture, isotopic labeling, affinity or reporter groups have been incorporated. Cross-linker bridges synthesized with bonds susceptible to cleavage either chemically or via CID to release component peptides for subsequent MSⁿ have aided in determining cross-links. Unfortunately, cross-linkers such as EDC and PFA are unable to exploit such technologies due to their small size[38].

A variety of cross-link analyses software exists with capabilities to identify cross-linked species, localize cross-linking sites, sequence peptide components, distinguish intra versus intermolecular cross-linking and evaluate cross-links using protein geometries[32, 34, 39–41]. These bioinformatics tools offer the versatility to customize searches based on characteristics of standard cross-linkers and identification features such as cleavable cross-linker bonds, isotopic labeling, or affinity tags used in experiments.

With recent advancements in cross-linking technologies, large-scale *in vivo* topology mapping via cross-linking site localization has now been performed. For example, the protein interaction reporter (PIR) is a cross-linker constructed with an affinity tag and cleavable bond in MS/MS to assist in cross-link identification. Similar to PFA, this membrane-permeable cross-linker is applied to cells prior to lysis. PIR technology analyzed various key membrane proteins in E.coli[42, 43] and even human cells[44]. These studies represent one of the largest identifications of cross-linked species. Although, PIR technologies have successfully identified numerous cross-linked species and relevant interaction sites, its large size and bulky substituents may hinder in capturing exclusively true interactions, interfere with reactivity or affect physiological structure of proteins [45]. On the other hand, PFA's short cross-linker bridge may not be suitable for protein interactions with intermolecular cross-linking sites that have large distances. Therefore, using multiple cross-linkers varying in reactivity and cross-linker bridge lengths in conjunction with PFA would prove to be a powerful tool for a comprehensive picture of biological systems and to further understand what interactions PFA can

1.4. *Tandem Mass Spectrometric Fragmentation and Nomenclature of Cross-linked Species*

specifically capture with respect to other cross-linkers.

Over the last decade, cross-linking along with MS has complemented molecular modeling experiments to verify protein topologies and contribute information about the dynamics of proteins[32, 37, 39, 46–49]. However, PFA has yet to match conventional cross-linkers engineered to produce straightforward MS analyses of three dimensional protein structures. The question remains, why has PFA's potential for structural proteomics not been fully unleashed despite its compatibility with MS and the successful MS discovery of protein interactions *in vivo*?

1.4 Tandem Mass Spectrometric Fragmentation and Nomenclature of Cross-linked Species

Upon cross-linking and the trypsin digestion of proteins, the resultant cross-linked peptides produce unique MS/MS fragmentation patterns in contrast to single peptides under CID. Since cross-linked peptides contain two peptide backbones and an additional bridge, three possible types of fragmentation can occur as illustrated in Figure 1.10a: at the bridge (type 1), at the peptide backbone (type 2) and at both the bridge and peptide backbone (type 3). Cross-linked component peptides are denoted as “I” and “II” and the cross-linked bridge is highlighted in red. Fragmentation at the cross-linker bridge generates whole peptide component fragment ions that can aid in the identification of each peptide. The sequence of each peptide component can be verified by the type 3 ions arising from the fragmentation of both the cross-link bridge and peptide backbone. Type 3 along with type 2 fragment ions, in which the cross-linker bridge remains intact and the peptide backbone is fragmented, can localize the cross-linking sites [50–52][50–52].

1.4. Tandem Mass Spectrometric Fragmentation and Nomenclature of Cross-linked Species

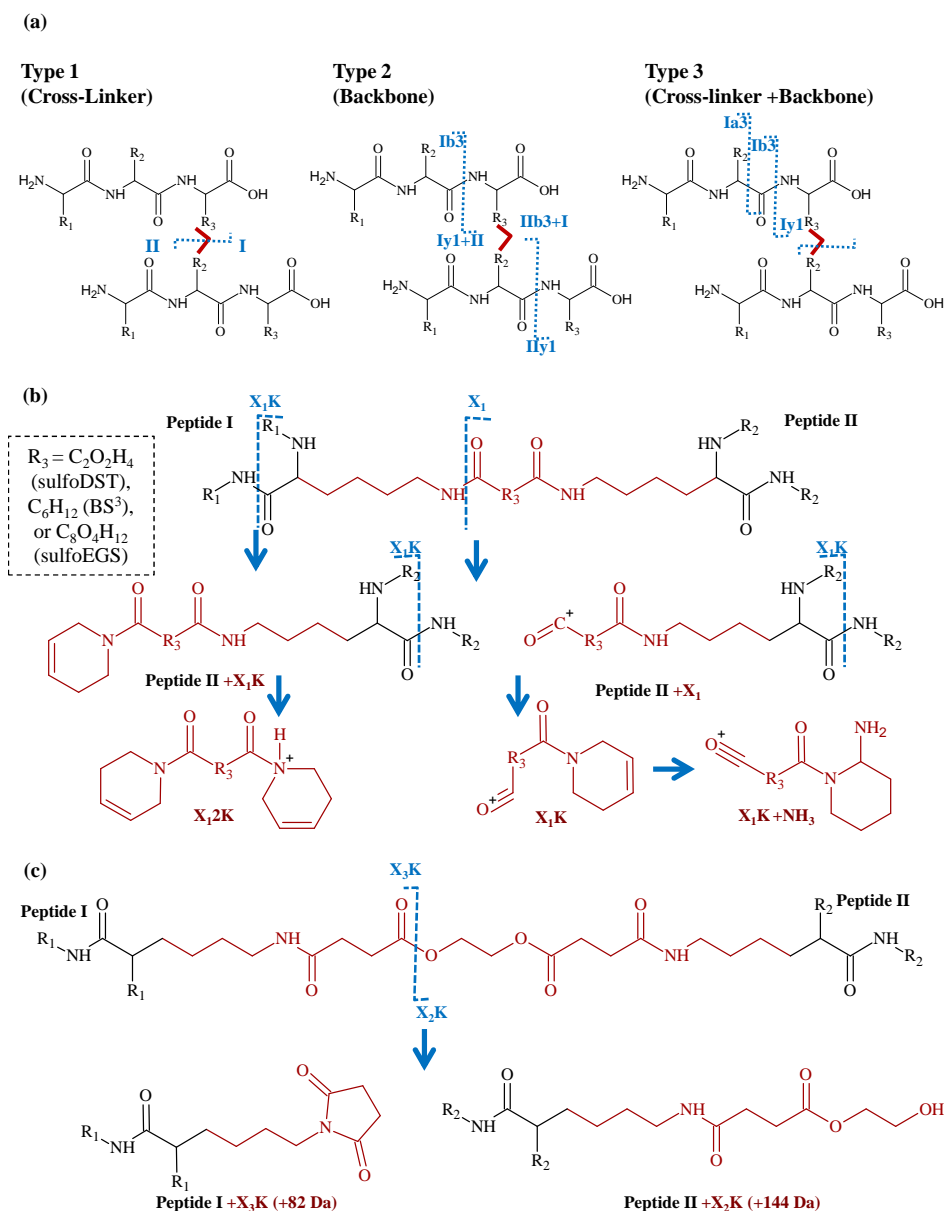


Figure 1.10: (a) Cross-linked peptides (denoted as I and II) can fragment exclusively at the cross-linker (type 1), exclusively at the peptide backbone (type 2) and at both the cross-linker and peptide backbone (type 3). (b) Diagnostic ions/specific fragmentation for NHS Esters (c) Specific fragment ions produced from sulfoEGS fragmentations. (a-c) Cross-linker bridges and cross-linked lysines from each protein site are highlighted in red.

1.5. Importance of Formaldehyde Cross-linking: Common Applications and Key Features

In addition, NHS ester cross-linked peptides can fragment either at the bond between the cross-linker carbonyl carbon and lysine nitrogen or at the peptide backbone bond, producing ions with modifications denoted as $+X_1$ or $+X_1K$, respectively (See Figure 1.10b) [53]. A double X_1K type fragmentation at both cross-linked lysine peptide backbone bonds produces X_12K , X_1K and X_1K+NH_3 ions, which can serve as diagnostic ions to confirm the presence of the cross-linker [52]. Figure 1.10c illustrates the diagnostic ions specific to the sulfoEGS crosslinker where fragment ions with $+X_3K$ (+82 Da) and $+X_2K$ (+144 Da) modifications can be generated from the fragmentation at the bond between the carbonyl carbon and oxygen within the sulfoEGS crosslinker bridge. However, PFA and EDC, being pseudo zero and zero-length cross-linkers, are unlikely to produce diagnostic ions.

1.5 Importance of Formaldehyde Cross-linking: Common Applications and Key Features

PFA is one of the oldest and widely applied cross-linkers in biological studies. Over a hundred years ago, PFA was discovered to be a suitable tissue fixative. Its widespread use has generated thousands, if not millions of tissue samples with their structural and functional properties preserved with PFA. Histopathological studies performed on these formalin-fixed tissues has unlocked valuable information to characterize and diagnose diseases [54]. PFA treatment inactivates enzymes and destroys bacteria responsible for tissue degradation, allowing tissues to be sustained for a long period of time [55]. In the early 1920s, Ramon recognized PFA's utility to develop vaccines because it inactivates toxins and viral proteins without destroying and possibly stabilizing antigens under mild PFA reaction conditions [56–58]. Essentially, this demonstrated that PFA can suspend the function of proteins without permanently damaging sites it modifies. PFA's minimal effect on antigens has enabled its conjunction with immunoprecipitation methods relying on accessible antigen sites for antibody binding to facilitate the investigation of protein–protein interactions. The first application of PFA cross-linking to capture specific interactions between biological species in their native environment was examining protein–DNA binding [59, 60]. Unlike tissue fixation, which utilizes long

1.5. Importance of Formaldehyde Cross-linking: Common Applications and Key Features

reaction times and high concentrations of PFA to extensively cross-link and maintain tissues long-term, these experiments call for selective cross-linking using short reaction times (10–30 min) and low concentrations (1%) of PFA. In chromatin immunoprecipitation, PFA is directly introduced to cells to cross-link and maintain the spatial context of protein–DNA complexes, which are then isolated with an appropriate antibody-antigen interaction to obtain genomic binding sites [61]. These successful experiments promoted the expansion of the PFA cross-linking approach to enhance the detection of true, specific protein–protein interactions by maintaining their cellular environment and spatial constraints, using similar experimental conditions [37, 62]. Importantly, protein interaction analyses demonstrate the utilization of mass spectrometry (MS) to identify interacting proteins preserved by PFA [62]. Indeed, MS is a critical component in many current structural proteomics studies. Notably, the development of a polyacrylamide gel silver-staining procedure designed for subsequent MS analyses involved the incorporation of PFA, suggesting PFA-treated material can be subjected to MS [37, 63]. Hence, MS is compatible with PFA-induced chemical changes, albeit not requiring their actual detection. Dimethyl labeling, a routine MS-based technology, exploits short reaction times and low PFA concentrations to produce Schiff-base modifications. These are immediately reduced to dimethyl substituents with NaCNBH_3 instead of forming methylene cross-link bridges [64]. Stable isotope dimethyl labeling has successfully quantified protein expression levels in cells by being applied to both peptide digests and intact proteins [65]. Minimal side reactions and the conservation of charge states in PFA modified peptides and proteins were reported, indicating that these modifications should not significantly disrupt chemical or physical properties of proteins. Dimethyl substituents were observed almost exclusively on lysines, illustrating the high specificity of the PFA modification in the first step of the reaction at low PFA concentrations and short incubation times [65, 66]. As described above, PFA cross-linking is suitable for *in vivo* biological applications and has been effectively utilized with MS, which are essential qualities for structural proteomics.

1.6 Formaldehyde Cross-linking *in vivo*: Mass Spectrometric Analyses of Formalin Fixed Tissues and Protein–Protein Interactions

PFA cross-linking in biological applications relies on one of two distinct strategies. The long-term preservation of protein structures in tissues employs high concentrations (>4%) of PFA and long incubation times (several hours to days) to form non-specific cross-links with a higher yield. However, to access proteins for functional, and possibly structural, information, the yield of cross-linking should be reduced to be compatible with MS analysis [67]. In contrast, less extensive PFA cross-linking conditions (0.05–1% PFA and 5–20 min incubation) have already routinely been implemented to capture relevant interactions through *in vivo* PFA cross-linking of protein complexes [37, 62, 68]. Such analysis is not dependent on detecting cross-linked species and relies on utilizing unmodified peptides to identify interacting proteins. Thus, maximizing PFA cross-linking yield to detect cross-linked species with MS without sacrificing specificity has yet to be achieved *in vivo*.

1.6.1 Formalin-fixed tissue: extensive cross-linking using long incubation times and high concentrations of Formaldehyde

An exhaustive range of formalin-fixed paraffin-embedded tissues stored in archives over the years has triggered the recent enthusiasm toward investigating specific protein structural and functional changes in these tissues with predetermined disease states. This structural analysis relies on MS-based detection of a sufficient amount of protein. It has been shown that heating can reverse cross-links and restore antigen reactivity for immunodetection to enable effective protein extraction and MS-analysis [55, 67, 69–71]. Notably, spectroscopy experiments with PFA cross-linked ribonuclease A had proven that PFA preserves tertiary structure even when applying high PFA concentrations and long incubation times similar to tissue fixation cross-linking [71]. This suggests that protein structure may be recovered post PFA cross-link reversal in fixed tissues for subsequent MS investigations. The LC–MS/MS analysis of hundreds to even thousands of proteins in an expansive

collection of formalin-fixed tissue samples has been achieved[67]. With the advancement of MS imaging, antigen retrieval and enzymatic digestion can directly be applied to intact formalin-fixed, paraffin-embedded tissue slides without disturbing the native environment of quaternary structures. This newly developed technology has defined targets for the early detection of cancer in various tissues [67, 69]. Taken together, the proven ability of PFA cross-linking of tissue in preserving quaternary structures and its compatibility with MS are key prerequisites for its use for comprehensive structural analysis, although the cross-links themselves have never been observed in these experiments.

1.6.2 Protein-Protein Interactions: specific, controlled cross-linking using short incubation times and low concentrations of Formaldehyde

For protein interaction studies, PFA is directly introduced to cells to freeze protein structures in their physiological environment prior to lysing cells. The mild PFA cross-linking conditions utilized in these experiments generate a low yield of cross-linked complexes. Following enzymatic digestion, LC-MS/MS analyses identifies proteins via the large abundance of unmodified peptides present [68]. This has been successful in determining protein interacting partners in various species [37, 62, 68, 72–84]. For example, PFA's capacity to quickly diffuse through and freeze protein geometries in living biological material has triggered large-scale cross-linking of even whole organisms through time-controlled transcardiac perfusion cross-linking [74, 76, 79]. It has also enabled the localization of transient and dynamic proteasome complexes in human leukemic cells for the first time. The stabilization induced by PFA cross-links prevented inter-compartment leakage and retained the location of proteasome complexes in cells [84]. Although, PFA has examined protein interactions in tissues and cells originating from a vast number of organisms, structural analysis derived from direct evidence of binding via reaction interface mapping is limited by the inability to detect and identify cross-linked species *in vivo* using MS.

1.7 Formaldehyde Cross-linking in Model Proteins

The apparent failure to observe PFA cross-linked species via MS in cells and tissues is remarkable. In part, this can be explained by the incomplete understanding of PFA chemistry and MS behavior. Consequently, simpler protein model system investigations are needed to clarify these aspects.

1.7.1 Formaldehyde Modification: Reactive sites and potential for exploiting solvent accessibility

In model peptides and proteins treated with PFA for several days, similar to the extensive treatment of PFA in fixed tissues, modifications on the N-terminus and side chains of lysine (K), arginine (R), histidine (H), cysteine (C), tyrosine (Y), tryptophan (W), serine (S), threonine (T) and phenylalanine (F) were discovered [85–87]. In contrast, under reaction conditions derived for specific *in vivo* cross-linking in protein interaction studies (physiological pH buffer, 1% PFA concentration and 5–20 min incubation time), only the N-terminus and cysteine, lysine, and to a lesser extent, arginine side chains, were primary reactive sites in the PFA modification reaction [36, 85]. In model peptides, in which varying site accessibility due to protein folding would not be a concern, N-termini were more reactive than lysine side chains and cysteines exhibited the highest reactivity using comparable *in vivo* protein interaction cross-linking reaction times. Reactivity of PFA with cysteines in proteins has yet to be studied under these reaction conditions [85]. If observed, PFA could possibly be used to also investigate cysteine-containing proteins, in which availability of free cysteines to react would depend on the protein's cellular location and functions [88, 89]. The increase in the quantity of reactive residues with exposure to higher concentrations of PFA for longer times exemplifies that the specificity of the modification reaction is reduced under these reaction conditions. This is consistent with dimethyl labeling methods that utilize mild PFA reaction conditions for specificity, as stated earlier in section 1.5 [66]. In general, these studies demonstrate the localization of modification sites, suggest that microenvironments of amino acids play a key role in PFA reactivity, and validate the specificity of the modification reaction [85].

At the protein level, PFA modification is also governed by protein folding,

which affects the accessibility of reactive sites. MS studies conducted with myoglobin and lysozyme, using *in vivo* protein interaction reaction conditions, examined the effect of PFA modification on structure. With only one N-terminal residue and no cysteines, the primary reactive sites in the myoglobin model system were lysines and, to a lesser extent, arginines. When PFA treated and untreated myoglobin samples were unfolded by decreasing the pH, both samples experienced similar increases in charge states, i.e. unfolding, irrespective of PFA treatment [36]. This agrees with the previously mentioned dimethyl labeling experiments reporting that PFA modification maintained the charge state of proteins and peptides [66]. A drastic rise in PFA modification resulted at the time points denaturant was added. The degree of modification increased as the protein unfolded due to the rise in accessibility of reactive amino acids that were previously buried, demonstrating that PFA reactivity is dictated by a protein's spatial constraints from folding. Furthermore, the low degree of modification of PFA treated myoglobin in the absence of denaturant validates that PFA itself does not induce protein unfolding, confirming its reliability for examining protein structural changes [36].

1.7.2 Formaldehyde Cross-linking: Reactive sites and formaldehyde cross-linked species in model systems.

Studies with single amino acids, model peptides and proteins using long incubation times similar to tissue fixation conditions for extensive cross-linking claimed that N-terminal groups, and the side chains of R, Y, H, W, asparagine (N) and glutamine (Q) residues are potential reactive sites in the second, cross-linking step, suggesting PFA's capacity to probe reaction interfaces containing any of these residues [86, 87, 90–92]. However, whether the number of potential PFA cross-linking sites would decrease using protein interaction, *in vivo* PFA cross-linking protocols that have been optimized for specific, less extensive cross-linking must be clarified [68]. Deducing cross-linking reactive sites in proteins is more suitable than in peptides. The yield of non-specific, intermolecular cross-linking in solution is low since it relies on the random contact of two peptide or protein molecules. In peptide cross-linking, only non specific or intermolecular cross-linking can occur due to the lack of tertiary structure. In solution cross-linking that occurs within a

protein or complex depends on its specific geometry and therefore should generate a higher yield of cross-linking than with peptides [85]. However, unlike with peptides, examining relative reactivity in proteins is a challenge due to the varying accessibility of sites from folding. Furthermore, protein systems can introduce complexities such as the increased number of possible reactive sites, which are partially occupied, giving rise to heterogeneous products in the case of PFA [36].

A bovine insulin model system (51 amino acids, 5.7 kDa) was treated with PFA under mild reaction conditions similar to *in vivo* protein interaction studies and digested with endoproteinase Glu-C under reducing and non-reducing conditions. In this model system, interpeptide PFA cross-linked species and cross-linking of the N-terminus to N and Y, and between K and Y were identified for the first time. The presence of fragment ions with and without the cross-link bridge intact exposed the semicleavable nature of PFA cross-link bridges subjected to CID in this study, which facilitated validation of cross-linked species and the regions of cross-linking. Along with fragmentation patterns, these studies also revealed that mild reaction conditions similar to *in vivo* protein interaction PFA cross-linking provide sufficient yield and specificity to detect PFA cross-linked species. Although, this proved that detecting PFA cross-linked species via MS/MS is possible in very simple, small proteins, exploring a wide spectrum of different protein structures varying in complexity is required to eventually reach the complexity of detecting PFA cross-linking in cells and tissues[50].

1.8 Non-covalent Protein Complex Model Systems

1.8.1 Calmodulin-Melittin Complex

It has been illustrated that PFA cross-linking can stabilize transient protein interactions in cells and tissues, however the detection of these cross-links for structural analysis has yet to be achieved. Weak transient interactions bind with a dissociation constant (K_d) in the μM range whereas those classified as strong transient interactions bind with a K_d in the nM range[93]. This present work aims to address the potential of PFA to capture weak transient interactions by examining the Ca^{2+} -free calmodulin-melittin complex.

1.8. Non-covalent Protein Complex Model Systems

Calmodulin is a Ca^{2+} binding protein that functions as calcium signal transducer in many key cellular pathways. It is a 16,779.78 Da protein composed of 148 amino acids (see Figure 1.11a) with a N-terminal domain (residues 1-76) and C-terminal domain (residues 80-148) connected by a flexible linker (77-79). Modifications on calmodulin include the removal of the N-terminal methionine (M), acetylation of the N-terminus (+42 Da mass modification), and trimethylated K116 (+42 Da mass modification). With four EF hands that have a signature helix-loop-helix structure, calmodulin can bind up to four Ca^{2+} . The *apo* calmodulin (Ca^{2+} unsaturated) adopts a dumbbell conformation upon Ca^{2+} binding, as illustrated in Figure 1.11b. Ca^{2+} binds sequentially to the C-terminal domain and then the N-terminal domain due to its higher affinity for the negatively charged C-terminal domain. In Ca^{2+} saturated calmodulin, buried hydrophobic pockets are exposed and this triggers complex formation. Calmodulin is also known to bind to proteins in the cell in the absence of Ca^{2+} such as actin-binding proteins, cytoskeletal and membrane proteins, enzymes, and receptors [94]. Typically, calmodulin forms a globular compact structure that surrounds its binding partner, wrapping the flexible loop around the target [95]. Melittin, the main constituent in bee venom, is a 2844.73 Da, 26 amino acid peptide (see Figure 1.11a) that can competitively bind to and inhibit calmodulin. It exists in solution in a tetrameric form. Each melittin molecule is composed of mostly hydrophobic amino acids (1-20) with positively charged C-terminal residues (20-24). The C-terminus of melittin is amidated (-1 Da mass modification). Upon binding, melittin becomes a bent rod structure: two alpha helices connected by the central proline residue [96] [97].

Melittin binds to Ca^{2+} -saturated calmodulin with K_d of 3nM. However, in the absence of Ca^{2+} , Melittin can still form a weaker complex with calmodulin with a K_d of 10 μ M, as depicted in Figure 1.11b. NMR studies have shown that both these complexes exist with the same conformation [98]. Presently, complete X-ray crystal structural data of the calmodulin-melittin complex are not available [95]. Models have been constructed from distance constraints using various cross-linkers applied to Ca^{2+} -saturated calmodulin bound to melittin. However, these studies confirmed cross-linked species only at the MS level, which may not provide sufficient verification and localization of cross-linking [99]. cross-linking with disuccinimidyl suberate (DSS) combined with MS/MS verification has also been ap-

1.8. Non-covalent Protein Complex Model Systems

plied to a Ca^{2+} -saturated calmodulin-melittin model system [100]. neither of these studies have examined the weaker complex formed between Ca^{2+} -free calmodulin and melittin. Two possible binding orientations between these components are possible: N-terminal domains are aligned with the C-terminal domains (antiparallel) or C-terminal domains are aligned with C-terminal domains and vice versa (parallel). Since the calmodulin C-terminal domain is more negatively charged than the N-terminal domain and melittin's C-terminus is positively charged, it is more electrostatically favorable to assume a parallel orientation[101]. Previous MS-based cross-linking studies have observed both types of orientations[99]. MS -based EDC cross-linking and limited proteolysis experiments supported antiparallel binding [102]. However, recent NMR and MS/MS based DSS cross-linking studies have shown that calmodulin and melittin predominantly bind in a parallel orientation[103] [100]. It has been shown that the C-terminal domain of Ca^{2+} -free calmodulin primarily interacts with the C-terminus of melittin and the N-terminal calmodulin domain is left free[95]. In addition, melittin induces a conformational change in Ca^{2+} -free calmodulin upon binding to its C-terminal domain [95]. Specifically, spectroscopic experiments have shown conformational fluctuations involving Y99 and Y138 of calmodulin when binding to melittin and melittin is oriented perpendicular to calmodulin, with its W19 becoming inaccessible upon binding [101].

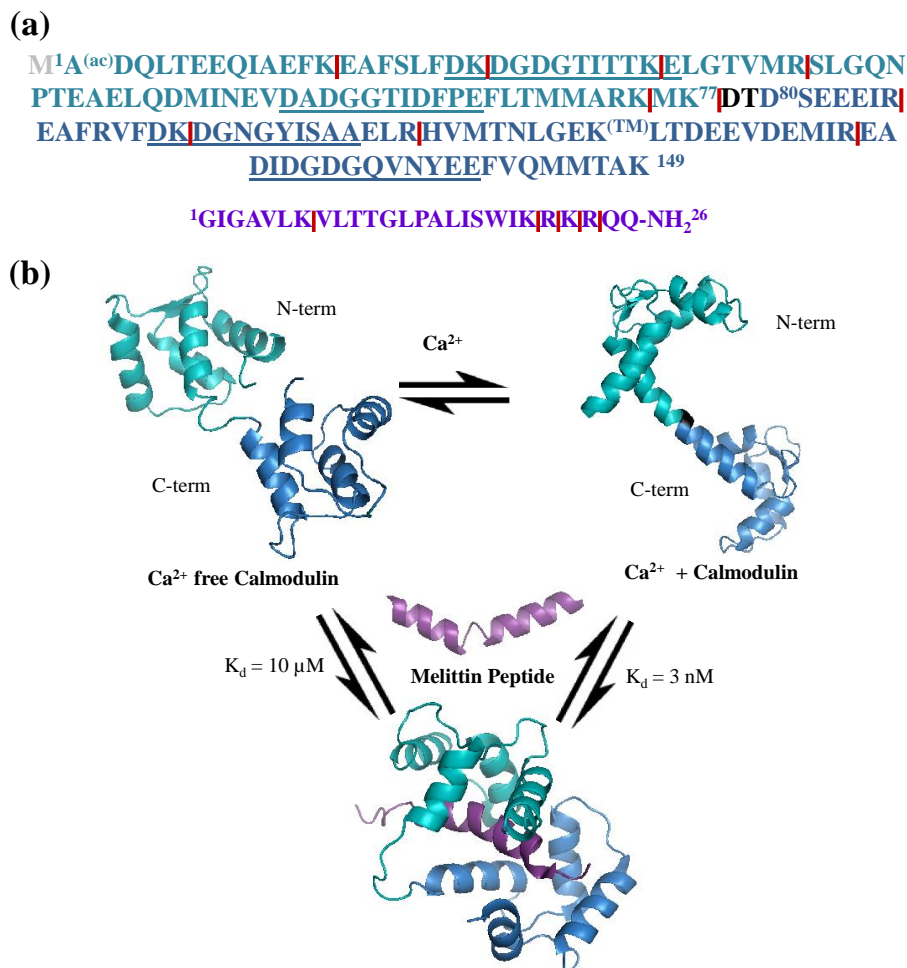


Figure 1.11: (a) Amino acid sequence of the calmodulin N-terminal (teal) and C-terminal (blue) domains connected by a flexible linker (black) and of melittin (purple); All possible trypsin cleavage sites are denoted with red vertical bars. (b) Calmodulin binds to Ca^{2+} , which induces the formation of a dumbbell-shaped conformation; Upon binding to melittin, a similar conformational change occurs for both Ca^{2+} -saturated and Ca^{2+} -free calmodulin. Melittin competitively binds to calmodulin, inhibiting calmodulin's activity.

1.8.2 Ribonuclease S Complex

Another transient protein complex examined in this project is the Ribonuclease S (RNaseS) complex. Ribonucleases are responsible for catalyzing the degradation of RNA molecules. Ribonuclease A (RNaseA) is cleaved by subtilisin to form RNaseS. RNaseS is composed of an S-peptide and S-protein, non-covalently associated, which retains the enzymatic activity of RNaseA when both components are present. Subtilisin is a non-specific enzyme that works like a serine protease, using a catalytic triad D-H-S to cleave such that the N-terminal primary amine is retained on the S-protein[104]. The cleavage of RNaseA can occur anywhere between residues 16 – 21, resulting in a mixture of S-proteins and S-peptides with various sequences[105, 106]. Figure 1.12 shows the masses and sequences of RNaseS components resulting from a cleavage after RNaseA residue 20. In Figure 1.12, the 11530.30 Da S-protein (residues 21-124) has eight cysteine residues engaged in disulfide bonding, which are reduced and stabilized with carbamidomethyl groups in the model system examined in this work. The 2165.01 Da S-peptide is composed of 20 amino acids (1-20). S-peptide and S-protein binding involves S-peptide residues 11-14 and S-protein residues 44-48. The binding affinity ($K_d = 1\mu\text{M}$)[107] of S-peptide to S-protein is significantly less than nucleotide binding, which involves residues H12, H119, K41 and Q11. It is also known that R10, R33, F8 and M13 are involved in the stabilization of the complex[108]. Various cross-linkers have previously been applied to RNaseS[105, 109, 110]. Interestingly, the distance between the S-protein and S-peptide binding sites is about 3 Å, suggesting that PFA possesses a suitable cross-linker bridge length (2.3 Å). It remains to be seen whether PFA can cross-link to preserve the weak interaction between RNaseS components.

(a)

S-peptide: 2165.01 Da (residues 1-20)

¹KETAAAKFERQHMDSSTSAA²⁰

S-protein : 11530.30 Da (residues 21-124)

²¹SSSNYCNQMMKSRNLTKDRCKPVNTEFVHESLADVQ
AVCSQKNVACKNGQTNCYQSYSTMSITDCRETGSSK
YPNCAYKTTQANKHIIVACEGPNPYVPVHFDASV¹²⁴

(b)

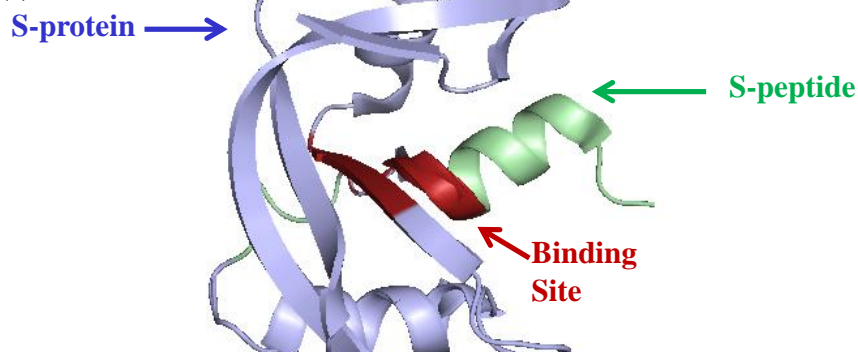


Figure 1.12: (a) Amino acid sequences of RNaseS components S-protein (blue) and the S-peptide (green). The S-protein contains 8 cysteines, which were reduced and alkylated in this model system. The S-peptide to S-protein binding sites are underlined. (b) The crystal structure (1RNU) [108] of RNaseS is shown with the S-protein (blue) to S-peptide (green) binding site highlighted in red.

1.9 Thesis Aims

Based on the recent promises of structural proteomics and the unique potential of PFA as a cross-linker that can capture close-proximity transient interactions, this thesis work has the following aims:

- (1) Identify and localize PFA cross-linking in protein systems larger than synthetic peptides, amino acids or insulin.

1.9. Thesis Aims

Both the Ca^{2+} -free calmodulin bound to melittin and RNaseS non-covalent protein complexes are chosen to examine PFA's capabilities to stabilize transient protein interactions (under mild reaction conditions similar to that of *in vivo* protein interaction studies). Identifying cross-linking and modification sites can confirm the PFA cross-linking mechanism and reaction chemistry under mild, *in vivo*-like conditions, which has thus far been only revealed in small peptides and insulin. Confirming PFA cross-linking in these more biologically relevant systems can bring the field closer to exploring and understanding PFA cross-linking in cellular and *in vivo* environments.

(2) Examine the reaction chemistry, MS analysis and MS/MS fragmentation patterns of PFA cross-linking with respect to other established cross-linking chemistries.

In addition to PFA, other commercially available cross-linkers of different reactivity and lengths are chosen to be applied in parallel. The zero-length cross-linker, EDC, is selected due to its comparable small size to PFA and heterobifunctional reactivity. NHS ester cross-linkers of various lengths (sulfoDST, BS³ and sulfoEGS) can be used to explore a variety of distance constraints and the effect of cross-linker bridge, relying on its predictable reactive sites. SDS-PAGE can be used to estimate and compare the cross-linking yield across intact cross-linked proteins for each cross-linker. Upon in-gel trypsin digestion and LC-MS/MS analysis, the complexity of analyzing cross-linked samples via MS and MS/MS can be examined for each cross-linker. The MS/MS fragmentation patterns that are established for EDC and the NHS ester cross-linkers can be used as guidelines to confirm PFA cross-links and derive MS/MS fragment ion evidence criteria to confirm cross-linking. Unlike these other cross-linkers, PFA cross-linking has yet to be identified with current cross-linking identification software. Therefore, the manual versus software identification of PFA cross-linking and of other established cross-linking can be compared in this study to highlight attributes for future software specifically tailored for PFA cross-link identification. The MS and MS/MS identification of cross-linking depends on chemical properties such as the trypsin cleavage efficiency of modified/cross-linked residues and formation of cross-linker-specific modifications vs cross-links, and also on instrument/software factors such as the sensitivity, accuracy and resolution of the mass spectrometer and MS peak picking

1.9. Thesis Aims

algorithms. Thus, these were also chosen to be explored. All together, these aspects shall place PFA in perspective with other established cross-linkers that have already achieved what PFA cross-linking has yet to accomplish i.e. cross-linking localization in biologically relevant systems and capturing three-dimensional protein structures.

(3) Apply cross-linking to explore the transient interaction between Ca^{2+} -free calmodulin and melittin, a protein complex with an unknown binding orientation, and introduce distance constraints for mapping protein structure.

Although the Ca^{2+} saturated calmodulin-melittin system has been examined with established cross-linkers, the more transient Ca^{2+} -free calmodulin-melittin complex has yet to be examined via cross-linking to my knowledge. Therefore, using the distance constraints imposed by cross-linkers of various reactivity and length can potentially provide a comprehensive structural analysis of the calmodulin-melittin complex, an unresolved structure. Furthermore, this would reveal whether PFA cross-linking can map three-dimensional protein structure and match the capabilities of established cross-linkers.

Chapter 2

Methods

2.1 Materials

Purified calmodulin ($\geq 95\%$ purity by SDS-PAGE), from bovine brain, lyophilized from a 400 μL buffer containing 150 mM sodium chloride (NaCl), 50 mM Tris-Hydrochloride (Tris-HCl), 2 mM Ethylenediaminetetraacetic acid (EDTA), and melittin ($>97\%$ purity by HPLC), from *Apis mellifica*, were purchased from MilliporeSigma (Darmstadt, Germany). The 16% formaldehyde solution (PFA) ampules were purchased from Thermo scientific (Waltham, MA). 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC), N-hydroxysulfosuccinimide (sulfoNHS), disulfosuccinimidyl tartrate (sulfoDST), bis(sulfosuccinimidyl) suberate (BS^3), and ethylene glycol bis(sulfosuccinimidyl succinate) (sulfoEGS) were bought from CovaChem (Loves Park, IL). Purified Ribonuclease S ($> 70\%$ purity by UV), Trizma Base (Tris), 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES), 2-(N-morpholino) ethanesulfonic acid (MES), sodium dodecyl sulfate (SDS), ammonium bicarbonate (NH_4HCO_3), glycerol, sodium hydroxide, tetramethylethylenediamine (TEMED), tricine, sodium chloride, potassium chloride and iodoacetamide were obtained from Sigma (St. Louis, MO). Sequencing grade modified trypsin, from porcine, was acquired from Promega (Madison, WI). Acetonitrile (ACN, HPLC grade), formic acid (FA, 88%), acetic acid (glacial) were obtained from Fisher. Acrylamide, ammonium persulfate (APS), bromophenol blue, Coomassie brilliant blue R250, gel casting and running systems were purchased from Biorad (Hercules, CA). Dithiothreitol was purchased from BDH Chemicals (London, United Kingdom). Deionized water was obtained using a Nanopure Ultrapure Water System Barnstead (Dubuque, IA).

2.2 Chemical Cross-linking Reactions

Calmodulin (at a final concentration of 60 μM and containing EDTA at a final concentration of 807 μM , see Appendix A.1.1 for the calculation of EDTA concentration) and melittin (at a final concentration of 60 μM) were incubated for 20 minutes at 37 °C in the respective cross-linking buffer (100 mM MES, pH = 6.5 for EDC samples and 20 mM HEPES, pH = 7.4 for other cross-linker samples) prior to the addition of each cross-linker reagent. The purified RNaseS sample was purchased with the components of the complex already present together so incubation prior to addition of cross-linker reagent was not required. RNaseS cross-linking was performed at a final concentration of 50 μM in 100 mM MES buffer (pH = 6.5) for EDC samples, and in 10x PBS buffer (pH = 7.4) for other cross-linker samples.

All cross-linkers were dissolved in cross-linking buffers just before their addition to the protein complexes. EDC was added to the protein mixture which was immediately followed by the addition of sulfoNHS with final concentrations of 60 and 30 mM, respectively. A EDC control sample (referred to as “-EDC”) was prepared by replacing the combined EDC-sulfoNHS volume with MES buffer. For each NHS ester (sulfoDST, BS³ and sulfoEGS) and PFA cross-linked sample, a 3 mM and 330 mM final cross-linker concentration was utilized, respectively. The control sample (denoted as “-others”) for NHS ester and PFA cross-linkers was prepared by replacing the cross-linker volume with HEPES and 10x PBS buffer for calmodulin-melittin and RNaseS samples, respectively. EDC and NHS ester cross-linking reactions were performed for 1 hour at 37 °C. PFA cross-linking was carried out for 6 hours at 37 °C.

EDC cross-linking reactions were quenched with dithiothreitol (DTT), using a final concentration of 40 mM. NHS ester and PFA cross-linking reactions were quenched with Tris buffer, using final concentrations of 60 and 500 mM, respectively.

2.3 Tris-Tricine Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

2.3.1 Casting Gels

Tris-Tricine SDS Gels were cast using a 16% running and 4 % stacking gel. The following protocol was used to cast two gels at a time. A Tris-HCL/SDS buffer with 3M Tris-HCl and 0.3% SDS, and pH= 8.45, was prepared. The 16% running gel was prepared by first mixing 30% acrylamide/0.8% bisacrylamide (7.5 mL) , 3M Tris-HCl/SDS (5 mL, pH = 8.45), glycerol (1.5 mL) and H₂O (1 mL) together. APS (0.2 mL) followed by TEMED (0.01 mL) was added and the running gel was immediately poured between gel casting glass plates (1mm), followed by the addition of isopropyl alcohol to cover the top of the gel. The gel was left to polymerize for approximately 30 minutes. The 4% stacking gel was prepared by first mixing 30% acrylamide/0.8%bisacrylamide (0.8 mL), 3M Tris-HCl/SDS (1.5 mL, pH = 8.45), and H₂O (3.7 mL) together. After removing the isopropyl alcohol and washing the top of the gels with H₂O, the stacking gel was poured between the glass plates immediately after the addition of APS (0.08 mL) and TEMED (0.01 mL) to the stacking gel solution. A 10-well green comb was inserted between the glass plates. After 20 minutes at room temperature, gels were stored at 4 °C overnight before use.

2.3.2 One Dimensional Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

The Tris-Tricine gel running chamber was prepared by filling the inner chamber between the glass plates holding the polymerized gels with a Tris-Tricine cathode buffer (0.1 M Tris, 0.1 M Tricine, 0.1% SDS) and the outer chamber with a Tris-Tricine anode buffer (0.1 M Tris, pH = 8.9). Reaction mixtures were mixed with 4X sample buffer (200 mM Tris-HCl, 2% SDS, 40% glycerol, 0.04% Coomassie brilliant blue R250) and incubated for 5 minutes at 65°C. Each sample was cooled and a prestained protein marker (10 kDa- 250 kDa), -EDC, -others, +EDC, PFA, sulfoDST, BS³ and sulfoEGS samples were loaded in each gel well from left to right. Electrophoresis was conducted at 25 mA until the gel bands passed through the

2.3. *Tris-Tricine Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation*

stacking gel, then the current was increased to 40 mA until the gel bands reached the end of the plate. Gels were visualized with coomassie brilliant blue R250 and gel images were obtained using an Odyssey infrared imaging system (LI-COR Biosciences, Lincoln, NE).

2.3.3 Analysis of Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

Molecular weights of gel bands were estimated using ImageJ, an image processing program designed for scientific research (University of Wisconsin) [111]. Migration distances from the dye front to the midpoint of each gel band appearing on the SDS-PAGE was measured using ImageJ. A standard curve was prepared using the protein marker by plotting the log of the known molecular weights (kDa) of each band versus its migration distance (pixels) in Microsoft Excel 2007. The equation of the best fitting line for the calmodulin-melittin SDS-PAGE was $y = 50.911x^{-0.662}$ ($R^2 = 0.9899$), where x is the migration distance and y is the log of the molecular weight. For the RNaseS SDS-PAGE, this equation was $y = 73.13x^{-0.723}$ ($R^2 = 0.9910$). For the literature SDS-PAGE of Ca^{2+} -free versus Ca^{2+} loaded calmodulin, this equation was $y = (3 \times 10^{-7})x^2 - 0.0037x + 1.9277$ ($R^2 = 0.9909$). These equations were used to estimate the molecular weights of gel bands for each sample lane.

Intensities of each gel band were analyzed using ImageJ. Each lane in the SDS gel image was selected in ImageJ, and a density plot for each lane was obtained such that the intensity of bands from the top to bottom (highest to lowest molecular weight) of the gel were plotted left to right. The peak area was proportional to the intensity of each gel band in each lane. Cross-linked protein bands were distinguished from non-cross-linked protein bands using these density plots. The peaks corresponding to the unmodified protein in the control sample was used to determine the unmodified protein peak in the experimental samples. The relative yield of cross-linking was calculated by summing the intensities of the cross-linked protein bands divided by the total intensity of all the bands in each cross-linked sample lane. The relative yield of non-cross linked species was calculated by summing the intensities of the non-cross linked protein bands divided by the total intensity of all

2.4. Trypsin Digestion

the bands in each cross-linked sample lane.

The isoelectric point of each protein component was determined using the ExPASy (Swiss Institute of Bioinformatics) ProtParam tool[112].

2.3.4 Excision and Washing of Gel Bands

Molecular weight categories to group relevant gel bands were determined based on the position and size of each gel band. Control (-EDC and -others) and sulfoDST cross-linked calmodulin-melittin samples, gel bands in each lane were excised and approximately grouped based on the following molecular weight categories: < 10 kDa, 10 - 20 kDa, and 20 - 40 kDa. For EDC, PFA, BS³ and sulfoEGS cross-linked calmodulin-melittin samples, where gel bands shifted 3.2, 4.2, 3.5, and 4.0 kDa below the actual molecular weight, respectively, the gel bands were approximately grouped based on the following molecular weight categories: < 14 kDa, 14-19 kDa, 19-33 kDa and > 33 kDa. For all RNaseS samples, gel bands were approximately grouped based on the following molecular weight categories: < 12 kDa, 12-20 kDa, and > 20 kDa. It is important to note that these molecular weight categories are approximate and variation between extremities in each group most likely exist.

Protein gel bands were washed with alternating cycles of acetonitrile and 100 mM NH₄HCO₃. For RNaseS, gel bands were reduced with 10 mM DTT at 56°C for 30 minutes in the dark and alkylated with 55 mM iodoacetamide (IAA) at 25°C for 1 hour in the dark.

2.4 Trypsin Digestion

An in-gel trypsin digestion was performed by incubating gel pieces with trypsin. Approximately 50 and 8 μ g of calmodulin and melittin, respectively, were present in the reaction mixtures loaded onto the SDS gel. For the RNaseS complex, approximately 58 and 11 μ g of the S-protein and S-peptide, respectively, were present in the loaded reaction mixtures. Gel bands were resuspended in 50 mM NH₄HCO₃ and trypsin was added to each gel band such that the protein-to-trypsin ratio was approximately 50 weight/weight (w/w). The digestion was performed overnight at 37 °C and was quenched with 5% formic acid. Peptides were extracted with ace-

2.5. Reverse Phase High Performance Liquid Chromatography Tandem Mass Spectrometric Analysis

tonitrile, purified with C18 stage tips, and resuspended in 0.035 mL of 5% formic acid.

2.5 Reverse Phase High Performance Liquid Chromatography Tandem Mass Spectrometric Analysis

2.5.1 Bruker Impact II Quadrupole Time-of-Flight Analysis For Cross-linked Calmodulin-Melittin and RNaseS Peptides

Purified trypsin digests were separated using a nano HPLC column, and an injection volume of 5 μ L was used. The pre-column was 4 cm long, with a 100 μ m diameter, and was packed with 5 μ m Aqua C18 material (Phenomenex). The nano column was 40 cm long with a 75 μ m diameter, and packed with 3 μ m Reprosil-Pur C18-AQ material (Dr. Maisch GmbH, Ammerbuch, Germany). The mobile phase buffers used were Buffer A (0.1 % formic acid) and Buffer B (80% acetonitrile and 0.1% formic acid). A flow rate of 250 nL/min was used. A 90 minute run was used with a gradient as follows:

t = 0 mins: 95% Buffer A, 5% Buffer B

t = 30 mins: 83% Buffer A, 17% Buffer B

t = 73 mins: 65% Buffer A, 35% Buffer B

t = 75 mins: 0 % Buffer A, 100% Buffer B

Samples were analyzed by the Bruker Impact II QqTOF (Bruker Daltonics Billerica, MA). MS spectra were collected for a m/z range of 200 - 2200. An isolation window was varied based on the m/z as follows: 2 Th for m/z = 300 and 400, 3 Th over a m/z = 500 - 800, 5 Th for m/z > 900 and interpolated for values falling between these ranges. The collision gas was nitrogen and the collision energy varied between 23-65 eV as a function of m/z and charge. Precursor ions were excluded for 0.4 min after being fragmented. MS/MS spectra were collected using data dependent acquisition over a m/z range of 50 - 2200. Appendix A.4 lists the collision energy used for each m/z and charge (intermediate values were interpolated).

2.5.2 ABI QStar XL Quadrupole Time of Flight Analysis For Mass Spectrometer Comparison

The nanoHPLC separation of the samples was performed using a lab-made nano-column that was 15 cm long, 75 μm diameter and packed with 3 μm Reprosil-Pur C18-AQ material (Dr. Maisch GmbH, Ammerbuch, Germany). A 100 min gradient was used with mobile phase Buffer A (0.1 % formic acid) and Buffer B (80% acetonitrile and 0.1% formic acid). Samples run on the ABI QStar XL QqTOF (Applied Biosystems, Foster City, CA). MS/MS spectra were collected using an information dependent acquisition (IDA). The collision gas was nitrogen and the collision energy was varied based as function of m/z and charge.

2.6 Mass Spectrometric Data Analysis

2.6.1 MaxQuant Verification of Unmodified/Modified Peptides via Tandem Mass Spectrometry

The raw LC-MS/MS Bruker Impact II QqTOF data files for the calmodulin - melittin and RNaseS samples (control and PFA treated) were directly loaded into MaxQuant (a proteomics software distributed by the Max Planck Society) [113]. The search was configured by adding custom protein fasta files and PFA-specific modifications. Fasta files of calmodulin-melittin and RNaseS were prepared by inputting the protein sequences into Format Converter (HIV Sequence Database)[114]. For RNaseS, fasta files were prepared with two S-peptide sequences (RNaseA residues 1-19 and 1-20), two S-protein sequences (RNaseA residues 20-124 and 21-124), and RNaseA sequence (RNaseA 1-124). For the calmodulin-melittin searches, the variable modifications used were the following: amidation of the C-terminus, acetylation of the N-terminus, deamidation of N, oxidation of M and trimethylation of K. For the PFA-cross-linked sample, Schiff base and methylol modifications on the N-terminus, K residues and R residues were added as variable modifications. For RNaseS searches, the following variable modifications were used: deamidation of N, oxidation of M and Carbamidomethyl on C. A minimum score cut off of 40 and a mass accuracy cut off of 40 ppm was used.

2.6.2 Manual Verification of Cross-linked Species via Tandem Mass Spectrometry

Figure 2.1a outlines the data analysis procedure used to identify and analyze cross-linked species.

For all samples analyzed with the Bruker Impact II QqTOF, the raw LC-MS/MS data was prepared as follows. Monoisotopic mass lists were acquired by using the sophisticated numerical annotation procedure (SNAP) algorithm to pick peaks with a signal to noise ratio (S/N) > 2 and the deconvolution feature in the Bruker Daltonics Compass Data Analysis 4.2 software. For samples analyzed with the ABI QStar XL QqTOF, monoisotopic mass lists were obtained using the Analyst QS 1.1 software and signals that had a S/N below 5 were omitted. For each cross-linked sample, m/z values < 400 and signals appearing in control samples were excluded to remove background signals. Cross-linked species were assumed to have at least a +2 charge since they are composed of two tryptic peptides, each with at least a +1 charge from the terminal K or R residues. Using Microsoft Office Excel 2007, masses corresponding to signals in control samples were eliminated from the mass lists from cross-linked species using a + 0.2 Da window. For EDC cross-linked samples, the “-EDC” control sample was used and for PFA, sulfoDST, BS³, and sulfoEGS the “-others” control sample was used.

A theoretical list of possible cross-linked masses specific for each cross-linker was derived assuming that a cross-linked mass was equal to the mass of each component peptide plus the cross-linker bridge mass plus any additional modification mass. ExPASy Peptide mass[115] was used to perform a theoretical trypsin digestion considering a maximum of four missed cleavages to produce a list of possible unmodified peptides. For calmodulin-melittin, the following fixed modifications: amidation of the C-terminus, acetylation of the N-terminus, and trimethylation of K were considered; and the following variable modifications were considered: deamidation of N and oxidation of M. The N-terminus of calmodulin is blocked by an acetyl modification and K115 is blocked with a trimethyl modification, so these sites were disregarded as modification and/or cross-linking sites for all cross-linkers. For RNaseS the following variable modifications were considered: deamidation of N, oxidation of M and Carbamidomethyl on C. The number of

2.6. Mass Spectrometric Data Analysis

cross-linking and modification sites in each peptide was determined for each cross-linker. In the case of PFA modified species, two modifications are possible at each site (N-terminal, R and K residues): Schiff Base (+12 Da mass shift) and methylol (+30 Da). Also, intra and interpeptide PFA cross-links produce a +12 Da mass shift and can form on N-terminal, R, H, Q, N and Y residues. For EDC, a -18.02 Da mass shift for each cross-linking site (E/D to N-terminal/K residues) and a +155.00 Da mass shift for each modification site (D or E residues) was considered. For the NHS ester cross-linkers, a 114.00, 138.07, and 226.05 Da mass shift for sulfoDST, BS³ and sulfoEGS, respectively, on both N-terminal and K residues were considered. In addition, modification after the hydrolysis of sulfoDST, BS³ and sulfoEGS cross-linkers with mass shifts of 132.01, 156.08, and 244.06 Da, respectively, were also considered for additional N-terminal or K residues on each peptide. The mass of every possible modified peptide involved in a cross-link was determined assuming each peptide could possess zero up to one minus the maximum possible modifications. A list of possible unmodified and modified peptide masses for each cross-linker was created and Mathematica 10 (Wolfram Research Inc., Champaign, IL)[116] was used to find every possible combination of these masses to produce a theoretical cross-link mass list specific to each cross-linker. Every possible combination of unmodified peptides was also derived using Mathematica to produce a list of impossible cross-link masses. Using Microsoft Excel, unmodified/modified peptide and impossible cross-link masses were eliminated from the monoisotopic mass lists for each cross-linker to obtain a final mass list. Microsoft Excel and Mathematica codes can be found in the Appendix A.2

Final mass lists from each cross-linked sample were matched to theoretical lists for each cross-linker with Mathematica using a +0.2 Da window to obtain the candidate list. For each candidate cross-linked species, the MS and MS/MS signals were inspected manually using the Bruker Daltonics Compass Data Analysis 4.2 or Analyst QS 1.1 software to visualize the LC-MS/MS data sets produced by the Bruker Impact II or ABI QStar XL, respectively. The extracted ion chromatogram (EIC) was obtained for each candidate to confirm the elution timepoint, and the MS spectrum at that timepoint was examined to verify the presence of the signal. Signals that did not exist, that were at the noise level of the spectrum, or were mixed with other signals were rejected to produce a MS Candidate List. For each

2.6. Mass Spectrometric Data Analysis

MS candidate signal, the MS/MS spectrum was obtained. In the Bruker Daltonics Compass Data Analysis 4.2 software, a list of all MS/MS spectra for each MS spectrum is provided and the MS/MS spectrum corresponding to the MS signal of the candidate was searched and selected manually. If a species with the same m/z and charge eluted at different timepoints, the MS spectrum of the timepoint at which the software was able to obtain a MS/MS spectrum was selected. In the Analyst QS 1.1 software, the elution time point of the MS candidate was selected in the total ion chromatogram (TIC) of the MS/MS scans to obtain all of the MS/MS spectra collected for each species, which were summed together.

Theoretical fragment ions for the backbone fragmentation of unmodified peptides from calmodulin and melittin were obtained using Protein Prospector (University of California, San Francisco) MS product [117]. These were used to manually prepare theoretical cross-linked and/or modified peptide fragment ion databases specific to each cross-linker chemistry. In each peptide, for each fragment that contained a cross-linking or modification site, modified fragment masses were obtained by adding the mass of each cross-linker-specific modification. In the case of PFA, for fragment ions containing modification sites, a series of these fragment ion masses with multiples of +12 Da, multiples of +30 Da and every combination of +12 Da and +30 Da mass shifts were prepared depending on the number of modification sites. For PFA fragment ions containing cross-linking sites, a series of fragment ion masses with multiples of +12 Da mass shifts were derived depending on the number of cross-linking sites. For EDC fragment ion containing modification sites, a series of fragment ion masses with multiples of + 155.00 Da mass shifts was prepared. For sulfoDST, BS³ and sulfoEGS cross-linkers, a series of fragment ion masses with mass shifts of +132.01, 156.08, and 244.06 Da, respectively, for each modification site was derived. Each MS/MS spectrum of candidate cross-linked species was inspected manually for the presence of unmodified and/or modified backbone fragment ions. If the MS/MS signals matched the theoretical fragment ion signals of one peptide part of the database, the mass of the precursor ion was used to deduce the mass of the possible second component peptide. The mass of the potential second peptide is equal to the mass of precursor ion minus the mass of the first peptide minus the mass of the cross-linker bridge pertaining to the sample. In addition, cross-links between adjacent peptides were not considered

2.6. Mass Spectrometric Data Analysis

as cross-linked species since they also matched single peptides with missed cleavages. If this mass did not correspond to any peptide in the database but matched a potential modification (i.e. oxidation, deamidation, or cross-linker-specific modification) then this species was considered to be a single peptide and not a cross-link. If the mass matched a another peptide then the expected backbone fragment ions of the second peptide were searched for manually in the MS/MS spectrum. Fragment ions corresponding to intact unmodified or modified peptides were searched for in the MS/MS spectrum. In addition, a list of fragment ion masses equal to the backbone fragment ion mass of one peptide plus the cross-linker bridge plus the mass of the other peptide component was prepared and compared to the MS/MS spectrum of the potential cross-linked species. All fragment ions with a maximum of one minus the charge state of the precursor ion were considered. If MS/MS signals did not match peptides in the database or matched one peptide with an unexplainable modification mass, these species were labelled as “undetermined.”

Cross-linking sites were localized by considering the residue reactivity of each cross-linker and examining modified type 3 and type 2 fragment ion evidence. If cross-linking sites were ambiguous, the degree of modification (DOM) was calculated using the following formula:

$$DOM = \frac{PA_0}{PA_0+PA_1} + \frac{PA_1}{PA_0+PA_1}, \text{ where}$$

PA_0 = Relative peak area of unmodified fragment ion and PA_1 = Relative peak area of singly modified fragment ion

Peak area values were normalized to the total peak area and were equated to the relative peak areas of each signal.

2.6. Mass Spectrometric Data Analysis

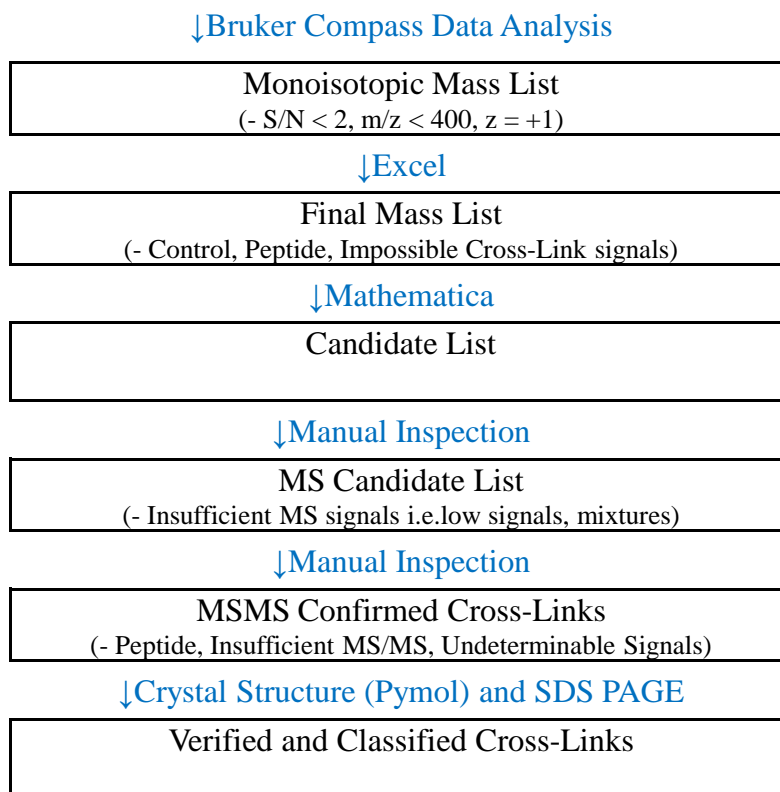


Figure 2.1: Data Analysis workflow for cross-link identification, where items in parentheses are values that have been eliminated. All elimination and matching of monoisotopic masses using Mathematica and Excel was performed using a mass accuracy of ± 0.2 Da.

2.6.3 Mascot Analysis of Unmodified Calmodulin

The MS/MS data from a control calmodulin sample produced by the QStar and Bruker Impact II were analyzed using the Mascot MS/MS peptide search (Matrix Science, Boston, MA)[118]. The Mascot generic format (mgf) files were generated using the Bruker Daltonics Compass Data Analysis 4.2 or Analyst QS 1.1 software. The MS/MS Ion Search was performed using a peptide mass tolerance and fragment mass tolerance of 0.2 Da, a significant threshold $p < 0.05$ and a score

cut off of 20. Variable modifications were set as follows: Trimethyl (K), Oxidation (M), Acetyl (N-terminus), Deamidated (NQ) and Acetyl (Protein N-terminus). The database was set to UP_cow and the instrument was set to ESI-QUAD-TOF.

2.6.4 Automated Cross-link Identification

StavroX/MeroX (University of Halle-Wittenberg)[119, 120] and pLink (Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China)[121] were used to compare manual versus software identified cross-linked species.

Parameters similar to the manual MS/MS search for cross-linked species were implemented into software searches. For all the software, an MS/MS search was performed. The mgf files for each LC-MS/MS sample data set was prepared using the Bruker Daltonics Compass Data Analysis 4.2 software. Fasta files of calmodulin-melittin and RNaseS were prepared by inputting the protein sequences into a Format Converter (HIV Sequence Database)[114]. Minimum mass accuracies of 60 ppm for MS and 10 ppm for MS/MS signals, respectively were used, based on the mass accuracy of manually identified cross-linked species. A maximum of four missed cleavages was accounted for and trypsin was set to cleave after cross-linked K and R residues to maximize the number of cross-linked species identified. For the calmodulin-melittin searches, the following fixed modifications were used: amidation of the C-terminus, acetylation of the N-terminus, and trimethylation of K; and the following variable modifications were used: deamidation of N and oxidation of M. For the PFA-cross-linked sample, Schiff base and methylol modifications on the N-terminus, K residues and R residues were added as variable modifications. For sulfoDST, BS³ and sulfoEGS cross-linked samples, modifications due to the hydrolysis of each cross-linker were set as variable modifications. For EDC cross-linked samples, variable modifications from the formation of N-acylurea (+ 155.00 Da) was considered. The PFA cross-linker had to be added to the cross-linker database for both MeroX and pLink and was set to form cross-links between N-terminal, K and R, residues to N-terminal, H, Y, Q, R and N residues with a bridge mass of 12 Da. For StavroX, a score cut off of 50 was used, which corresponds to a false discovery rate (FDR) < 5% [120]. In pLink, only cross-linked species that fall below a 5% FDR are reported [122]. The minimum

2.7. *Analysis Based on Relative Abundance Calculations in the Calmodulin-Melittin System*

MS/MS fragment ion evidence established to confirm a cross-linked species the manually was used to assess whether cross-links identified by the software existed (see section 4.5 for specific criteria). The m/z and charge of the software identified cross-linked species were examined in the MS spectrum using Bruker Daltonics Compass Data Analysis 4.2 software. If the MS signals of the cross-linked species were not present, they were classified as having insufficient MS evidence for confirmation. StavroX produces an annotated spectrum for each cross-linked species identified and this was used to evaluate each cross-link identification. The pLink software did not provide an annotated MS/MS spectra and therefore signals were checked in the raw MS and MS/MS spectra manually using the Bruker Daltonics Compass Data Analysis 4.2 software.

2.7 Analysis Based on Relative Abundance Calculations in the Calmodulin-Melittin System

For all unmodified, modified and cross-linked species that were identified, abundances were equated to the normalized peak area of their MS signal. The peak area of each peptide was normalized by dividing it by the total peak area of all peptides in each sample. Normalized peak areas were summed across all runs for each unique peptide. In order to make conclusions based on the abundance of species in each sample, a universal assumption was made in this study that no peptides were lost in the sample preparation, MS-detection or identification processes since it is not feasible to account for this loss. It is also assumed that there is uniformity in the MS-response of all peptides in the sample and that all peptides were ionized with similar efficiency. Percent abundance values ($> 1\%$) are reported with zero decimal places to reflect the accuracy of the MS peak area measurements. For percent abundance values $< 1\%$, two or three decimal places are given to prevent reporting a 0% percent abundance. In addition, in order for the total percent abundance to equal 100% for a particular site, the percent abundance values $> 99\%$ are reported to a number of decimal places that matches the corresponding values $< 1\%$.

2.7. *Analysis Based on Relative Abundance Calculations in the Calmodulin-Melittin System*

2.7.1 Percent Trypsin Cleavage and Accessibility

The percent trypsin cleavage for the control and PFA treated samples in both calmodulin-melittin and RNaseS systems was derived for all observed trypsin cleavage sites in these samples. A list of all trypsin cleavage sites and missed cleavage sites were prepared from the MaxQuant identified control, PFA unmodified and PFA modified peptide lists. A list of all trypsin cleavage sites and missed cleavage sites were also prepared from the PFA cross-linked peptides identified manually. For each peptide, N-terminal and C-terminal K or R residues (with the exception of the calmodulin C-terminal residue K148 and melittin N-terminal residue G1) were considered as trypsin cleavage sites. Internal K or R residues were considered missed cleavage sites (with the exception of the calmodulin trimethylated K115). In addition missed cleaved residues with either +12 or +30 Da PFA modification were excluded as trypsin missed cleavages sites to only examine trypsin cleavage as a function of structural accessibility. In the case where the position of the PFA modification was ambiguous, the number of trypsin missed cleavage sites was subtracted by the number of modifications per peptide to obtain the number of trypsin cleavage sites that were not a result of a PFA modification.

The percent trypsin cleavage for calmodulin and melittin in control vs PFA treated samples was calculated for each cleavage site by dividing the normalized abundance of peptides that supported cleavage at a particular site by the total normalized abundance of all peptides containing the particular site (with or without a missed cleavage). This was performed for unmodified, modified and cross-linked peptides identified in the PFA treated sample and the unmodified peptides identified in the control sample.

2.7.2 Relative Abundances and Equilibrium of Formaldehyde Unmodified, Modified and Cross-linked Species

2.7.2.1 Percent Relative Abundance of Cross-linking and Modification

For each cross-linking site identified in the PFA treated calmodulin-melittin samples, the relative abundances of all the peptides that were unmodified, modified with a +12 Da modification, modified with a +30 Da modification and cross-

2.7. *Analysis Based on Relative Abundance Calculations in the Calmodulin-Melittin System*

linked at that particular cross-linking site were summed for each category. This determined the total unmodified abundance, modified abundance, and cross-linked abundance for each cross-linking site. To determine the total percent of unmodified, modified and cross-linked protein, the abundances of each cross-linking site in its unmodified, modified and cross-linked form were each summed and divided by the total sum of abundances for all forms. It was assumed that all +12 Da modifications were Schiff Base modifications since previous PFA studies have confirmed that the formation of Schiff Base modifications is more likely than intrapeptide cross-links in proteins when using short reaction times similar to this current study[36, 85] and to account for all potential Schiff Base modifications.

2.7.2.2 **Derivation of Equilibrium Constants for Formaldehyde Cross-linking Reaction**

The equilibrium constant for the PFA cross-linking that occurred between the identified cross-linking sites for the calmodulin-melittin system were determined. For cross-linking sites that involved residues that were identified with both methylol (+30 Da) and Schiff Base (+12 Da) modified peptides, the equilibrium constants K_1 , K_2 and K_3 are defined in the reaction scheme depicted in Figure 2.2. However, for those cross-linking sites that involved residues that were identified with only +12 modified peptides, the equilibrium constant K_{1+2} is defined in the reaction scheme depicted in Figure 2.3.

2.7. Analysis Based on Relative Abundance Calculations in the Calmodulin-Melittin System

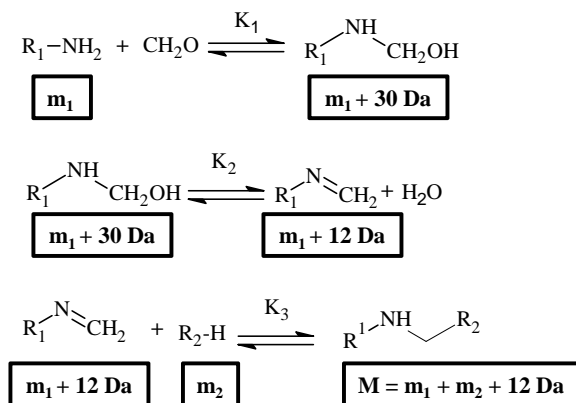


Figure 2.2: PFA cross-linking equilibrium reaction steps, where K_1 , K_2 , and K_3 are the respective equilibrium constants for the formation of a methylol, Schiff Base and methylene bridge, respectively. The notation for each reactant and product is defined. R_1 and R_2 represent protein sites 1 and 2, respectively.

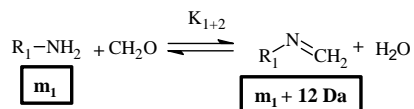


Figure 2.3: The PFA modification equilibrium reaction step defines K_{1+2} in the case where a methylol modification was not identified. R_1 and R_2 represent protein sites 1 and 2, respectively.

The expressions for the equilibrium constants were derived. The equilibrium constant is defined as the concentration of products over the concentration of reactants (in dilute solutions):

$$K = \frac{[\text{Products}]}{[\text{Reactants}]}$$

The total equilibrium constant is the product of the equilibrium constants (K_1 , K_2 , and K_3) for each intermediate step of the reaction:

$$K_1 = \frac{[\text{R}_1\text{CH}_2\text{OH}]}{[\text{R}_1\text{NH}_2][\text{CH}_2\text{O}]}$$

2.7. Analysis Based on Relative Abundance Calculations in the Calmodulin-Melittin System

$$K_2 = \frac{[R_1NCH_2][H_2O]}{[R_1CH_2OH]}$$

$$K_3 = \frac{[[R_1NHCH_2R_2]]}{[R_1NCH_2][R_2H]}$$

The amount of PFA was added in much higher excess than protein (5500X) and the amount of PFA that reacted with the protein is negligible. Therefore, it is assumed that PFA is a constant and was excluded from the equilibrium expression. Also, since cross-linking and subsequent analyses were performed in solution, water is also assumed to not affect the equilibrium and was thus excluded from the equilibrium expression. K_1' and K_2' are defined below, assuming PFA and water are constants in the equilibrium expressions:

$$K_1' = \frac{[R_1CH_2OH]}{[R_1NH_2]}$$

$$K_2' = \frac{[R_1NCH_2]}{[R_1CH_2OH]}$$

“A” is defined as the normalized abundance (i.e. the normalized MS peak area) of each species, where n is the identity of the species:

$$A_n = \text{Abundance of species } n$$

The concentration of a species involved in the equilibrium reaction is proportional to its relative abundance. The relative abundance was calculated by dividing the normalized abundance of each product or reactant by the total normalized abundance of all products and reactants. This ensures that equilibrium constant remains unitless in all three equilibrium constant equations. The relative abundance was substituted in the equilibrium constant expression and the definition of the equilibrium constant based on abundance, K_{MS} , is given below:

$$K_{MS} = \frac{\frac{A_{product}}{\sum A}}{\frac{A_{reactant}}{\sum A}}$$

2.7. Analysis Based on Relative Abundance Calculations in the Calmodulin-Melittin System

For the PFA cross-linking reaction shown in Figure 2.2 and 2.3, the total normalized abundance of all products and reactants for each step of the reaction are defined as follows:

$$\sum A_1 = A_{m_1} + A_{m_1+30}$$

$$\sum A_2 = A_{m_1+12} + A_{m_1+30}$$

$$\sum A_3 = A_{m_2} + A_{m_1+12} + A_M$$

The relative abundances were substituted in the equilibrium constant expressions to derive the following equations:

$$K'_{1MS} = \frac{\frac{A_{m_1+30}}{\sum A_1}}{\frac{A_{m_1}}{\sum A_1}} = \frac{A_{m_1+30}}{A_{m_1}}$$

$$K'_{2MS} = \frac{\frac{A_{m_1+12}}{\sum A_2}}{\frac{A_{m_1+30}}{\sum A_2}} = \frac{A_{m_1+12}}{A_{m_1+30}}$$

$$K'_{(1+2)MS} = K'_{1MS} K'_{2MS} = \frac{A_{m_1+12}}{A_{m_1}}$$

$$K_{3MS} = \frac{\frac{A_M}{\sum A_3}}{\left(\frac{A_{m_1+12}}{\sum A_3}\right)\left(\frac{A_{m_2}}{\sum A_3}\right)} = \frac{A_M}{(A_{m_1+12})(A_{m_2})} (\sum A_3)$$

2.7.3 Relative Abundance of Different Types of Cross-linking

The relative abundance of calmodulin-calmodulin versus calmodulin-melittin and antiparallel vs parallel calmodulin-melittin cross-linking was calculated using the identified EDC, PFA, sulfoDST, BS³ and sulfoEGS cross-linked species. Cross-links were classified as calmodulin-calmodulin cross-linking if both cross-linking sites were from the calmodulin molecule. Each calmodulin-melittin cross-link was classified as “parallel” if the cross-linking occurred between residues from the N-terminal domain of both calmodulin and melittin. Each calmodulin-melittin cross-link was classified as “antiparallel” if the cross-linking occurred between residues from opposite domains of calmodulin and melittin (i.e. the N-terminal domain of one component and the C-terminal domain of the other component). For each cross-linker, the normalized peak area values of all identified cross-linked species for each type of cross-linked species (calmodulin-calmodulin, calmodulin-melittin, parallel calmodulin-melittin or antiparallel calmodulin-melittin) was divided by the sum of normalized peak area values of all identified cross-linked species.

2.8 Crystal Structure Distance Constraints for the Calmodulin-Melittin System

Crystal structures for unbound Ca²⁺ saturated calmodulin (3CLN)[123], unbound Ca²⁺ -free calmodulin (1CFC) [124], calmodulin (1PRW) [125] in its bound state and melittin(2MLT)[126] were obtained from Research Collaboratory for Structural Bioinformatics: Protein Data Bank (RCSB PDB)[127]. All distance measurements on each crystal structure were performed using Schrödinger PyMol, a molecular visualizations software [128] (Schrödinger, New York, NY). The C α is defined as the carbon atom on the peptide backbone connected to the side chain. For intramolecular calmodulin cross-links, distances between C α of each cross-linked amino acid side chain (C α -C α distance) were measured on unbound calmodulin and bound calmodulin crystal structures. The maximum cross-linking distance is defined as the largest distance that two amino acid reactive sites can exist in the protein structure to form a cross-link. Table 2.1 lists the maximum cross-linking distances for each cross-linker for every possible combination of reactive sites. It

2.8. Crystal Structure Distance Constraints for the Calmodulin-Melittin System

was assumed that side chains can freely rotate about the $C\alpha$ with a radius equal to the length of the side chain. Side chains that can come into contact with the cross-linker within this radius were considered cross-linked. This accounts for the side chain flexibility. To account for backbone flexibility, an additional 6 Å was added to the maximum cross-linking distance. All together, maximum cross-linking distances were calculated by summing the length of each amino acid side chain involved in the cross-link (measured from $C\alpha$ to cross-linked atom), the cross-linker spacer arm length (bridge length), and 6 Å for backbone flexibility. This calculation was based on recent molecular dynamic simulations [129]. As shown in Table 2.1, the maximum cross-linking distance possible for EDC, PFA, sulfoDST, BS3, and sulfoEGS cross-linkers is 17.3, 22.1, 25.2, 30.2 and 34.9 Å, respectively

Table 2.1: Maximum cross-linking distances for every combination of possible reactive sites for each cross-linker are listed

| Cross-Linker | Spacer Arm Length (Å) | Cross-Linking Residues | Modified Residues | | |
|-----------------|-----------------------|------------------------|-------------------|----------|----------|
| EDC | 0 | | D | E | |
| | | N-term | 9.7 Å | 10.9 Å | |
| | | K | 16.1 Å | 17.3 Å | |
| PFA | 2.3 | | N-term | K | R |
| | | N-term | 8.3 Å | 14.7 Å | 15.2 Å |
| | | Q | 12.2 Å | 18.6 Å | 19.1 Å |
| | | N | 10.8 Å | 17.2 Å | 17.7 Å |
| | | R | 15.2 Å | 21.6 Å | 22.1 Å |
| | | H | 13 Å | 19.4 Å | 19.9 Å |
| | | Y | 12.9 Å | 19.3 Å | 19.8 Å |
| | | | N-term | K | |
| SulfoDST | 6.4 | N-term | 12.4 Å | 18.8 Å | |
| | | K | 18.8 Å | 25.2 Å | |
| BS ³ | 11.4 | N-term | 17.4 Å | 23.8 Å | |
| | | K | 23.8 Å | 30.2 Å | |
| SulfoEGS | 16.1 | N-term | 22.1 Å | 28.5 Å | |
| | | K | 28.5 Å | 34.9 Å | |

The $C\alpha$ - $C\alpha$ distance were compared to the maximum cross-linking distance to evaluate intra calmodulin cross-linked species. Calmodulin-melittin structures were derived using maximum cross-linking distances between all MS/MS verified cross-linked species.

Chapter 3

General Data Analysis of Cross-linked Calmodulin-Melittin and Ribonuclease S

3.1 Cross-linking of Calmodulin-Melittin and Ribonuclease-S Complexes with Various Cross-linkers

PFA along with various established cross-linkers (EDC/sulfoNHS, sulfoDST BS³, and sulfoEGS) were applied to two transient protein complexes: Ca²⁺-free calmodulin-melittin and RNaseS (S-peptide and S-protein).

EDC/sulfoNHS, sulfoDST, BS³ and sulfoEGS cross-linking of this complex was carried out using previously optimized protocols [99] using a one hour incubation time at 37 °C. All cross-linking was performed at physiological pH (pH = 7.4) except EDC, which was performed at a pH = 6.5. The activation step of EDC and sulfoNHS has been shown to be most efficient at this pH [33, 35]. Mild PFA cross-linking reaction conditions (physiological pH, 1% PFA, and 6 hour reaction time) were utilized to mimic *in vivo* protein interaction studies [50, 68]. The two control samples (one for EDC and one for other cross-linkers) were prepared by replacing the cross-linker reagent with an equal volume of buffer and served as a negative controls. EDC's zero-length cross-linker bridge and heterobifunctionality [33] served as suitable comparison to PFA's pseudo zero-length bridge and

3.1. Cross-linking of Calmodulin-Melittin and Ribonuclease-S Complexes with Various Cross-linkers

differing residue reactivity between the first and second step of the cross-linking reaction. On the other hand, sulfoNHS ester cross-linkers (sulfoDST, BS³ and sulfoEGS) are reactive with lysines and N-terminal residues in both steps of the cross-linking reaction[32], making their reactive sites much more predictable than PFA. With straightforward reactivity, the NHS ester cross-linkers of various lengths were chosen to examine the effect of cross-linker size on specificity and capturing protein structure.

3.1.1 Confirming Calcium-Free Calmodulin-Melittin

In this study, the Ca²⁺-free calmodulin-melittin system was analyzed. In order to confirm that the calmodulin-melittin system remained Ca²⁺-free regardless of potential Ca²⁺ contamination in the solution, the expected ratio of Ca²⁺-loaded calmodulin to Ca²⁺-free calmodulin was calculated. The level of Ca²⁺ reported in unpurified tap water in the University of British Columbia (the location of the lab) was < 2.5 mg/L (6.2×10⁻⁵ M) [130]. However, the water purification system (Nanopure Ultrapure Water System Barnstead) utilized in this experiment was shown to produce deionized water that contained < 1 ng/L (2.5×10⁻¹¹ M) of Ca²⁺ [131]. The presence of EDTA, a calcium chelating reagent, in the calmodulin sample can maintain a Ca²⁺-free complex by binding to any residual Ca²⁺ from water in the sample, as demonstrated in previous literature [132, 133]. EDTA and calmodulin were present in sample at a final concentration of 8.1 ×10⁻⁴ M and 6.0 ×10⁻⁵ M in this experiment, respectively. To determine whether the concentration of EDTA present in the sample was sufficient to prevent Ca²⁺-loaded calmodulin from forming to a significant extent, the equilibria of Ca²⁺ binding to EDTA and to calmodulin was simultaneously considered. EDTA binds to Ca²⁺ with a higher affinity than calmodulin and the concentration of EDTA is 13 times greater than calmodulin [134]. The pH-dependent binding affinity of EDTA to Ca²⁺ was taken into account for the two different buffer pH conditions utilized in this experiment, i.e. pH = 6.5 for EDC cross-linking and pH = 7.4 for other (PFA, sulfoDST, BS³, and sulfo EGS) cross-linking[135]. Previous studies have shown that upon binding to melittin, the binding affinity of calmodulin for Ca²⁺ increases 300 times [136]. The total concentration of Ca²⁺-loaded calmodulin-melittin was calculated

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

to be $7.9 \times 10^{-13} M$ and $8.2 \times 10^{-12} M$ in the presence of EDTA under the EDC and other cross-linking buffer conditions, respectively. The ratio of Ca^{2+} -loaded calmodulin-melittin to Ca^{2+} -free calmodulin-melittin was calculated to be only 1.3×10^{-8} and 1.4×10^{-7} , in the EDC and other cross-linking reaction mixtures, respectively, (the details of the calculation are shown in Appendix A.1.3), which are negligible.

These ratios were calculated based on the calcium content in the ultrapure, deionized water utilized in this experiment. However, if calcium contamination had occurred, it would require an initial Ca^{2+} concentration of at least $7.8 \times 10^{-5} M$ (3.3 mg/L) or $6.2 \times 10^{-5} M$ (2.5 mg/L) in EDC and other cross-linking reaction mixtures, respectively, in this experiment for even 1% of the calmodulin-melittin to be Ca^{2+} -loaded (calculation shown in Appendix A.1.3). These Ca^{2+} concentrations are within the concentration of Ca^{2+} in regular, unpurified tap water. This suggests that if an amount of tap water had been introduced into the reaction mixture, Ca^{2+} -loaded calmodulin-melittin would not form to a significant extent in the samples. All together, this supports that the majority of the calmodulin-melittin complexes should be Ca^{2+} -free in this study.

3.2 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

3.2.1 Running Conditions and Gel Band Appearance

Reaction mixtures were separated by one dimensional SDS-PAGE and visualized using Coomassie Blue to qualitatively assess cross-linking products and presence of protein components. A 16% acrylamide gel, Tris-Tricine buffer system (mass range of 1-100 kDa) was used since it is suitable for small proteins < 30 kDa such as calmodulin (~16.9 kDa) and the S-protein (~11.5 kDa) of RNaseS and also for peptides such as melittin (~2.9 kDa) and the S-peptide (~2.2 kDa) in RNaseS. In this buffer system, both the stacking and running gels have a pH = 8.45, which allows for proteins to move slower for an improved separation of proteins with a narrow range of molecular weights in comparison to glycine-based, buffer systems that decrease the pH to 6.8 for the stacking gel. Furthermore, a higher percent of

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

acrylamide was used to further enhance the resolution of small protein components by allowing them to migrate at a slower rate [137].

Since cross-linked species form at a relatively low yield, a higher protein concentration must be used in comparison to other protein experiments in order to be detected by LC-MS/MS post in-gel digestion and purification procedures[32]. Approximately 50 and 8 μg of calmodulin and melittin, respectively, were present in the loaded reaction mixtures. For the RNaseS complex, approximately 58 and 11 μg of the S-protein and S-peptide, respectively were present in the loaded reaction mixtures. Loading a high protein concentration in the SDS-PAGE can cause poor resolution and streaking [138]. For the calmodulin-melittin complex (Figure 3.1), a poor resolution of bands was observed especially for PFA (lane 5), BS³ (lane 7) and sulfoEGS (lane 8) cross-linked samples. Both streaking and poor resolution were more prominent in the RNaseS SDS-PAGE (Figure 3.2) for the other cross-linker control sample (lane 3) and the PFA (lane 5) and BS³ (lane 7) cross-linked samples. Achieving an excellent resolution with cross-linked samples is challenging since unmodified, intra-cross-linked and modified proteins do not differ significantly in mass and also possess similar molecular shapes and amino acid sequences, thus appearing as overlapping bands.

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

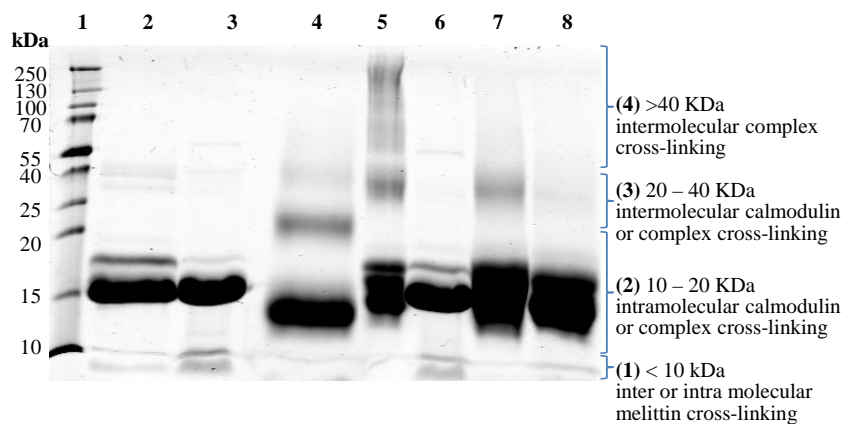


Figure 3.1: SDS-PAGE of calmodulin-melittin cross-linking reaction mixtures with the protein marker (lane 1); control samples with EDC buffer conditions (lane 2), and control samples with all other cross-linker buffer conditions (lane 3); cross-linked samples EDC/sulfoNHS (lane 4), PFA (lane 5), sulfoDST (lane 6), BS³ (lane 7) and sulfoEGS (lane 8); Four approximate molecular weight categories of each protein/protein complex band are labelled with the type of crosslinking (if any) indicated.

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

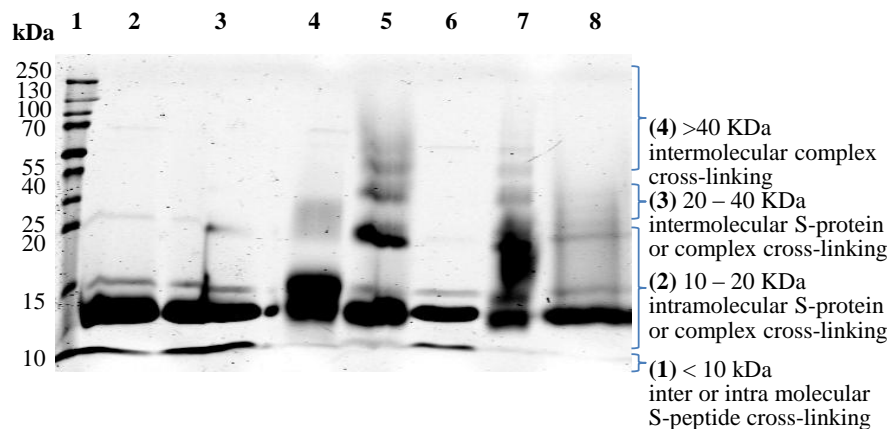


Figure 3.2: SDS-PAGE of RNaseS cross-linking reaction mixtures with protein marker (lane 1); control samples with EDC buffer conditions (lane 2), and control samples with all other cross-linker buffer conditions (lane 3); cross-linked samples EDC/sulfoNHS (lane 4), PFA (lane 5), sulfoDST (lane 6), BS³ (lane 7) and sulfoEGS (lane 8); Four approximate molecular weight categories of each protein/protein complex band are labelled with the type of crosslinking (if any) indicated.

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

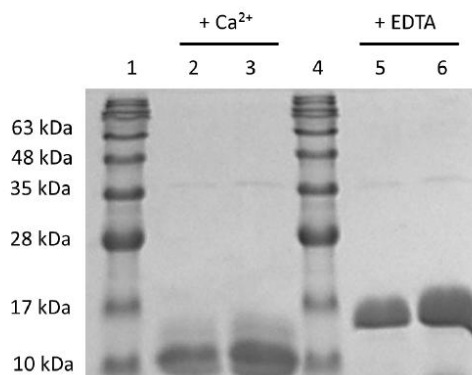


Figure 3.3: Literature study that compares the SDS-PAGE of calmodulin in the presence of EDTA, without Ca²⁺ (lane 5 and 6) and in the presence of Ca²⁺, without EDTA (lane 2 and 3). Lanes 1 and 4 are protein markers. The amounts of calmodulin used were 6 μg (lanes 2 and 5) and 12 μg (lanes 3 and 6). The concentrations of CaCl₂ and EDTA were both 5 mM (adapted from reference [139], with permission).

3.2.2 Unmodified Protein Gel Bands

First, the control samples (lane 2 and 3) were examined to ensure the presence of the protein components in the reaction mixture and to observe the behavior of the unmodified protein complex under the conditions utilized in the SDS-PAGE. In Figure 3.1, control samples (lanes 2 and 3) contained a strong band for calmodulin (~ 17 kDa) and faint bands for melittin (< 10 kDa). An additional band in the control sample, more prominent in the EDC control (lane 2), appeared between 17 - 20 kDa. This band was also observed in literature [99, 140, 141] for calmodulin, suggesting that it is most likely an impurity associated with calmodulin. In this current study, the purchased calmodulin was 95% pure by SDS-PAGE, so it is possible that impurities are present. The buffer pH used for the EDC control (lane 2) and the control for other cross-linkers (lane 3) was 6.5 and 7.4, respectively. Higher pH promotes oligomerization of melittin due to its alpha amino group being deprotonated [142]. This may explain the bands for melittin oligomers (< 10 kDa

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

bands) appearing in only the control for other cross-linkers.

It was determined whether the appearance of the unmodified calmodulin SDS-PAGE gel bands in this experiment matched that of the SDS-PAGE gel bands produced by Ca^{2+} -free calmodulin in literature. Upon binding to Ca^{2+} , calmodulin adopts a more compact structure with a smaller Stokes radius than its Ca^{2+} -free structure [143]. Therefore Ca^{2+} -loaded calmodulin possesses an increased electrophoretic mobility during SDS-PAGE than Ca^{2+} -free calmodulin [144]. It was observed previously that Ca^{2+} -loaded calmodulin migrates further down the SDS-PAGE gel than Ca^{2+} -free calmodulin, even under denaturing conditions in which the protein should be unfolded [145]. Figure 3.3 depicts a 17% SDS-PAGE gel from a previous study that examined the electrophoretic mobility of calmodulin in the presence of Ca^{2+} or EDTA [139]. As shown in Figure 3.3, the bands for Ca^{2+} -free calmodulin (treated with EDTA) appeared at 14-18 kDa whereas Ca^{2+} -loaded calmodulin (treated with CaCl_2) bands appeared at 10-14 kDa, illustrating the faster migration of the more compact structure of Ca^{2+} -loaded calmodulin. It is expected that Ca^{2+} -calmodulin would produce a series of bands corresponding to calmodulin loaded with 1,2,3 or 4 Ca^{2+} ions and Ca^{2+} -free calmodulin would produce a single band for calmodulin. For both samples, the bands spread over a 4 kDa range. However, in the Ca^{2+} -loaded calmodulin sample, more distinct bands corresponding to multiple forms of calmodulin loaded with different numbers of Ca^{2+} ions were observed in contrast to the uniformly spread band in the Ca^{2+} -free calmodulin sample. In this study, the bands for unmodified calmodulin were observed at ~17 kDa, as shown in Figure 3.1, which is similar to the Ca^{2+} -free calmodulin sample in Figure 3.3. In addition, the spreading of the calmodulin band in this study resembled the spreading that occurred with the Ca^{2+} -free calmodulin sample more than the Ca^{2+} -loaded calmodulin sample in the literature study shown in Figure 3.3. Therefore, both these attributes support that the calmodulin in this experiment was Ca^{2+} -free when comparing the SDS-PAGE with what was observed in literature [139]. This observation is also consistent with section 3.1.1, in which the presence of EDTA was shown to quench the majority of the Ca^{2+} in the sample, resulting in a nearly zero concentration of Ca^{2+} -loaded calmodulin in the sample.

Figure 3.2 shows the SDS-PAGE of the cross-linked RNaseS samples. Since

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

the cleavage site of RNaseA can be between residues 16 – 21, the resulting heterogeneous mixture of S-protein and S-peptide with various sequences may explain the appearance of a series of overlapping bands. The control samples (lane 2 and 3) contained a strong band for the S-protein (11.4 – 11.6 kDa) and a faint band for RNaseA (13.7 kDa). A thin band appearing around 10 kDa could possibly be an impurity associated with the sample since the purchased RNaseS was > 70% pure. Although the Tris-Tricine buffer system was chosen specifically to capture low molecular weight proteins/peptides, the S-peptide (2.0 - 2.2 kDa) was still not observed. Other reports of RNaseS cross-linking have also been unable to observe the S-peptide on a gel due to its small size, which makes it difficult to remain in the gel [105]. The S-peptide (17-20 residues) is only slightly smaller than melittin (26 residues), which was observed as faint bands in the calmodulin-melittin SDS-PAGE. However, the S-peptide has an almost neutral isoelectric point (~6.8) in comparison to melittin (12.0) and may have migrated faster toward the end of the gel (toward the positive electrode), increasing the probability of escaping from the gel.

3.2.3 Estimation of Molecular Weight for Cross-linked Protein Gel Bands

Intermolecular cross-linking between two different calmodulin/S-protein (~33 kDa/ ~23 kDa) or calmodulin-melittin/RNaseS complex

(> ~39 kDa/~27 kDa) molecules are expected to appear with at least double the mass of unmodified calmodulin/S-protein or complex molecules. Intramolecular cross-linking within calmodulin-melittin or RNaseS would appear with a mass approximately equal to the mass of the complex (~20 kDa and 14 kDa, respectively) since the mass change produced by the cross-linker bridges and/or modifications would be too small to observe on an SDS-PAGE gel. However, intramolecular cross-linking within calmodulin/S-protein would produce a negligible mass difference in comparison to unmodified calmodulin/S-protein. This negligible mass increase would be based on only on the number of cross-links and/or modifications formed and cross-linker bridge mass.

Using the protein marker bands of known molecular weights, a standard curve

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

for the molecular weight vs distance migrated from the top of the running gel was plotted and this was used to estimate the molecular weights of all gel bands for the calmodulin-melittin and RNaseS SDS-PAGE. Interestingly, cross-linked bands appeared below the expected mass, especially with EDC and PFA treated calmodulin-melittin. This may be due to the stabilizing nature of cross-linking, which resists the denaturation process of SDS-PAGE and maintains the compact protein structures. SDS-PAGE operates by unfolding proteins and adding negatively charged SDS sulfate groups approximately proportional to each protein's molecular weight. As proteins without defined three-dimensional structure, they migrate down the gel based on their molecular weight. However, if the structures of these proteins are preserved with cross-linking, then they would travel down the gel as function of their size/structure and molecular weight. Since more compact or smaller size structures travel further down the gel, their bands correspond to a lower molecular weight than the unfolded, larger size conformation. This observation is consistent with previous studies showing that non-reduced lysozyme migrated further than reduced lysozyme due to its disulfide bonds retaining its folded structure. Thus, the non-reduced lysozyme appeared with a molecular weight 14% lower than expected [146]. Furthermore, other studies have shown that lysine acetylation, which also neutralizes these basic sites similar to lysine cross-linking, blocks SDS binding and increases protein migration in SDS-PAGE [147]. The effect of cross-linking on SDS-induced protein unfolding may also be more of an issue in this present experiment due to the milder denaturing conditions used, i.e. incubating samples with SDS at 65°C rather than 90-99 °C utilized by other studies [99] prior to electrophoresis. Nevertheless, increasing the temperature to ensure proper denaturation may reverse cross-links [68] and was therefore avoided. Thus the location of the cross-linked gel bands was not solely based on the molecular weight, but also a function of size or compactness of the protein, which depended on the extent of cross-linking and/or cross-linker.

This phenomenon was more of an issue with calmodulin-melittin than with RNaseS. This may be due to calmodulin being more acidic (pI ~ 4) than RNaseS (pI ~ 9) resulting in more repulsion with the negatively charged SDS and thus increased resistance to unfold the protein [148]. The displacement of cross-linked gel bands from the position corresponding to their actual molecular weight was

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

estimated by measuring the difference in molecular weight between the non-cross-linked calmodulin or S-protein bands in each cross-linked sample lane and the calmodulin or S-protein band in the control sample lanes. It was determined that PFA, EDC, sulfoDST, BS³, and sulfoEGS cross-linked gel bands appeared approximately 3.2, 4.2, 0.1, 3.5, and 4.0 kDa, respectively, lower than their actual molecular weight. SulfoDST exhibited the smallest displacement in comparison to other cross-linkers, suggesting that its cross-linking did not significantly block the SDS-induced denaturation of the three-dimensional protein structure. For RNaseS, the calculated displacement in the molecular weight of protein gel bands was negligible for PFA, EDC, sulfoDST, BS³, and sulfoEGS (+0.2, +0.1, +0.4, -0.2, and -0.1 kDa, respectively).

3.2.4 Cross-linking Evidence

The series of overlapping bands reflects the range of different cross-linked products produced [149], which was most prominent with PFA treated calmodulin-melittin samples due to PFA's diverse reactivity. As mentioned in section 3.2.3, cross-linked gel band positions appeared much lower than the actual molecular weight. In the EDC treated calmodulin-melittin sample (lane 4), the occurrence of intermolecular cross-linking is supported by bands appearing at ~30-40 kDa (corrected to ~34-44 kDa based on the calculated molecular weight displacement from section 3.2.3). It is possible that intramolecular calmodulin EDC cross-linking also occurred, supported by the strong band below 14 kDa (corrected to ~ 18 kDa). The 20-25 kDa band (corrected to 24-29 kDa) suggests that EDC cross-linking either occurred between calmodulin and melittin tetramers, or involved the impurities that were observed at 17 - 20 in the EDC control sample. For the PFA treated sample (lane 5), the series of bands between 13-18 kDa (corrected to ~ 16-21 kDa) indicate the formation of intramolecular cross-linked species. Higher order complexes formed from extensive intermolecular cross-linking appeared as low intensity bands at higher molecular weights (> 30 kDa, corrected to > 33 kDa). These complexes are most likely artifacts generated by PFA cross-linking that do not reflect true interactions. This observation is consistent with PFA cross-linked protein complexes in cellular studies, in which such artifacts formed using PFA concentrations above

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

0.4% [73]. This consistency supports the use of this current non-covalent protein complex model system to reveal PFA cross-linking characteristics in future cellular studies. The appearance of unreacted melittin oligomers (bands < 10 kDa) and calmodulin (17 kDa) in the sulfoDST (lane 6) treated sample similar to the control sample (lane 3) suggests that cross-linking between calmodulin and melittin did not occur as extensively as with other cross-linkers. This is consistent with previous observations in literature [99]. For the BS³-crosslinked sample (lane 7), bands appearing between ~13 -16 kDa (corresponds to ~ 17-22 kDa) reflect intramolecular cross-linking. A faint band appearing at approximately 30 kDa (corresponds to ~34 kDa) suggests the formation of intermolecular cross-linked species. For the sulfoEGS cross-linked sample (lane 8), strong bands appearing between ~13-16 kDa (corresponds to ~ 17-22 kDa) indicate the presence of intramolecular cross-linking. A very low intensity band appearing at about 30 kDa (corresponds to ~34 kDa) suggests that intramolecular cross-linking may have occurred.

Strong bands for intramolecular cross-linked RNaseS appeared for PFA, BS³ and EDC cross-linked samples at around ~14 kDa. Intermolecular cross-linking between multiple S-protein and complex molecules are marked by bands appearing above ~20 kDa. Faint bands between 20-30 kDa appeared in the EDC cross-linked sample, which corresponds to these intermolecular cross-linked complexes. Similar to the calmodulin-melittin model, PFA contained such high order cross-linked complexes marked by a series of bands between 25 - 100 kDa, which are also likely to be non-specific complexes formed by PFA cross-linking. Cross-linking evidence for sulfoDST was weaker than with the other cross-linkers, marked by strong bands for unreacted protein at below 12 kDa, similar to the bands appearing in the control sample in lane 3. For the BS³ cross-linked sample, a series of bands between 12- 25 kDa suggested that intramolecular cross-linking occurred and a series of faint bands above 20 kDa suggested that some intermolecular cross-linking also occurred. Very faint bands between 12-20 and above 20 kDa suggests that some intramolecular and intermolecular cross-linking, respectively, did occur with sulfoEGS and RNaseS.

To determine the relative yield of cross-linking for each cross-linker based on SDS-PAGE, the relative densities of the cross-linked protein bands were summed and compared to the relative density of the non-cross linked protein band for each

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

cross-linker. The bar graphs illustrating these values are shown in Figure 3.4 for the calmodulin-melittin system and Figure 3.5 for the RNaseS system. For the calmodulin-melittin model system, PFA displayed the highest cross-linking yield (77%), which is what is expected based on the increased number of cross-linking sites in comparison to other cross-linkers. The lowest yield of cross-linking (9%) occurred in the sulfoDST sample, which is consistent with the observation of unreacted protein bands in the SDS-PAGE and literature [99]. Being a Ca^{2+} -binding protein, calmodulin contains many negatively charged D and E residues which also are major reactive sites for EDC. Since the calmodulin-melittin system has a high content of EDC-specific reactive sites, the relative yield was higher (73%). Although sulfoDST, BS³ and sulfoEGS contain the same number of reactive sites (K and N-terminal residues), which are fewer than EDC and PFA, BS³ exhibited a higher cross-linking yield (67%) than sulfoEGS (44%) and sulfoDST (9%). Since the cross-linker bridge is the main variable across these three NHS ester cross-linkers, the higher cross-linking yield for BS³ suggests that many of the K and N-terminal residues exist within the distance of the cross-linker bridge. Possible explanations for a lower cross-linking yield with sulfoEGS may be that as a large cross-linker with a longer bridge, it formed more interprotein cross-linked complexes that precipitated out and thus did not appear on the gel, or that it could not permeate through the protein complex as easily as the smaller BS³ cross-linker.

For the RNaseS model system, similar to the calmodulin-melittin system, both EDC and PFA cross-linked samples exhibited similar cross-linking yields of 55% and 51%, respectively. This may have occurred since both cross-linkers form close proximity cross-links with multiple types of residues. The observation of unreacted protein bands in the sulfoDST sample lane supported sulfoDST's relatively low cross-linking yield of 7%. BS³ cross-linked samples displayed the highest cross-linking yield of 80%. The sulfoEGS cross-linked samples possessed a cross-linking yield of 45%. Even though the NHS ester cross-linkers have the same residue reactivity, the varying cross-linking yields illustrate that the cross-linker length plays a crucial role in forming cross-links specific to each protein structure, as observed with the calmodulin-melittin system. In both RNaseS and calmodulin-melittin, BS³ displayed the highest cross-linking yield out of these NHS ester cross-linkers, suggesting that K and N-terminal residues have a highest

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

probability of existing within the length of the BS³ bridge and/or its bridge has the optimal balance of length and flexibility to form such cross-links in these protein complexes. The higher cross-linking yield of PFA in the calmodulin-melittin system than in the RNaseS system suggests that there are more PFA-reactive residues in the calmodulin-melittin that are close in proximity since PFA possesses a small cross-linker bridge. The lower cross-linking yield in RNaseS for PFA samples may be due to PFA reactive residues being inaccessible or not in close proximity. Likewise, the lower cross-linking yield in RNaseS versus calmodulin-melittin for EDC samples suggests the reduced number of the reactive sites accessible to EDC or not within a zero-length proximity with each other. Nevertheless, these hypotheses must be confirmed via MS and MS/MS analysis of the cross-linked samples.

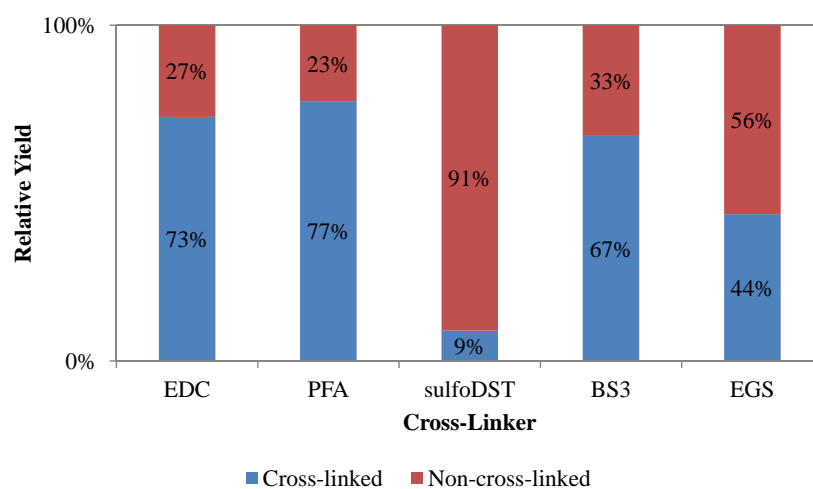


Figure 3.4: Relative yield of cross-linked species (blue) versus non-cross linked species (red) in the Calmodulin-Melittin complex measured via SDS-PAGE

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

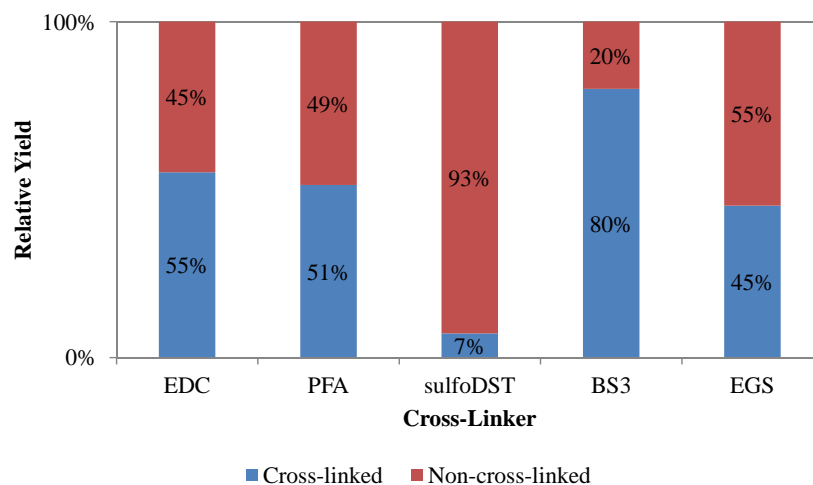


Figure 3.5: Relative yield of cross-linked species (blue) versus non-cross linked species (red) in the RNaseS complex measured via SDS-PAGE

3.2.5 Cross-linking Yield and Protein Complex Dissociation Constant

The percent cross-linked and non-cross-linked protein complex is hypothesized to reflect the the amount of bound calmodulin-melittin ([CM]) and unbound calmodulin [C] and melittin [M], respectively, which was calculated using the known dissociation constant (K_d) of $10\mu\text{M}$ [98]. The dissociation constant is defined as follows:

$$K_d = \frac{[C][M]}{[CM]} = 10\mu\text{M},$$

where [C], [M], and [CM] are the equilibrium concentrations of calmodulin, melittin and the calmodulin-melittin complex, respectively. The initial concentration of calmodulin ($[C]_0$) and melittin ($[M]_0$) used was $60\mu\text{M}$ in the reaction mixture. However, this is the concentration of the calmodulin and melittin samples that were 95% and 97% pure, respectively. A more accurate estimate of the initial concentrations of calmodulin and melittin present in the reaction mixture would be

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

95% and 97% of 60 μM , respectively. Therefore, the equilibrium concentrations are defined as follows:

$$[C] = [C]_0 - [CM]$$

$$[M] = [M]_0 - [CM]$$

$$[C] = (95\%)(60\mu\text{M}) - [CM]$$

$$[M] = (97\%)(60\mu\text{M}) - [CM]$$

$$[C] = 57\mu\text{M} - [CM]$$

$$[M] = 58\mu\text{M} - [CM]$$

Substituting $58\mu\text{M} - [CM]$ and $57\mu\text{M} - [CM]$ for the concentration of melittin and calmodulin, respectively, in the K_d expression above and solving for $[CM]$ gives the following:

$$K_d = \frac{(57\mu\text{M} - [CM])(58\mu\text{M} - [CM])}{[CM]} = 10\mu\text{M}$$

$$[CM] = 38\mu\text{M}$$

$$[C] = 57\mu\text{M} - [CM] = 57\mu\text{M} - 38\mu\text{M} = 19\mu\text{M}$$

$$[M] = 58\mu\text{M} - [CM] = 58\mu\text{M} - 38\mu\text{M} = 20\mu\text{M}$$

The total concentration of the sample was set to 60 μM

$$\% \text{BoundCalmodulinMelittin} = \frac{38\mu\text{M}}{60\mu\text{M}} = 63\%$$

$$\% \text{UnboundCalmodulin} = \frac{19\mu\text{M}}{60\mu\text{M}} = 32\%$$

$$\% \text{UnboundMelittin} = \frac{20\mu\text{M}}{60\mu\text{M}} = 33\%$$

3.2. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Separation

Therefore, the theoretical percent of bound calmodulin-melittin is 63% and unbound calmodulin and melittin is 32% and 33%, respectively in the 95% and 97% pure calmodulin and melittin samples, respectively. This is fairly close to the percent yield of cross-linking observed for PFA, EDC and BS³ cross-linkers (77%, 73% and 67%, respectively). Thus, this suggests that cross-linking with these cross-linkers can reflect the amount of bound calmodulin-melittin expected and preserve this interaction.

This analysis was also performed for the RNaseS system (with a $K_d = 1\mu\text{M}$ [107]), in which K_d is defined as follows:

$$K_d = \frac{[S_{pep}][S_{pro}]}{[RNaseS]} = 1\mu\text{M},$$

where $[S_{pep}]$, $[S_{pro}]$, and $[RNaseS]$ are the equilibrium concentrations of S-peptide, S-protein and the RNaseS complex, respectively. Unlike the calmodulin and melittin samples, that were 95% pure, the purchased RNaseS sample was only 70% pure. Therefore, it is assumed that only 70% of the initial concentration correspond to RNaseS components. The initial concentration of the RNaseS sample used was 50 μM and the equilibrium concentrations are defined as follows:

$$[S_{pro}] = [S_{pro}]_0 - [RNaseS]$$

$$[S_{pep}] = [S_{pep}]_0 - [RNaseS]$$

$$[S_{pro}] = [S_{pep}] = (70\%)(50\mu\text{M}) - [RNaseS]$$

$$[S_{pro}] = [S_{pep}] = 35\mu\text{M} - [RNaseS]$$

Substituting $35\mu\text{M} - [RNaseS]$ for the concentration of S-peptide and S-protein in the K_d expression above and solving for $[RNaseS]$ gives the following:

$$K_d = \frac{(35\mu\text{M} - [RNaseS])^2}{[RNaseS]} = 1\mu\text{M}$$

$$[RNaseS] = 23\mu\text{M}$$

$$[S_{pro}] = [S_{pep}] = 35\mu\text{M} - 23\mu\text{M} = 12\mu\text{M}$$

3.3. Mass Spectrometric Analysis

$$\%BoundRNaseS = \frac{23\mu M}{50\mu M} = 49\%$$

$$\%UnboundPeptide = \%UnboundSProtein = \frac{12\mu M}{50\mu M} = 24\%$$

This gives a theoretical percent of the RNaseS complex and unbound S-protein and S-peptide of 49 % and 24 %, respectively, in the 70% pure sample. Interestingly, the percentage of bound RNaseS is fairly close to percent yield of cross-linking for EDC, PFA, and sulfoEGS (55%, 51% and 45 %, respectively). This suggests that these cross-linkers are suitable to capture the RNaseS interaction and also provide evidence that the amount of cross-linking observed relates to the expected amount of protein complex formed. EDC and PFA crosslinking yields supported this relationship for both calmodulin-melittin and RNaseS complexes.

3.3 Mass Spectrometric Analysis

3.3.1 Calmodulin-Melittin System

3.3.1.1 Control Peptide Analysis via MaxQuant

The MS/MS is automatically acquired on the most intense MS signals. MaxQuant can be used to perform an MS/MS search of the control calmodulin-melittin sample (-others) to confirm the presence of the expected tryptic peptides without cross-linker-specific modifications and cross-links. Table 3.1 lists the 34 calmodulin and melittin peptides identified in the control sample with their m/z , experimental monoisotopic mass ($[M]_{exp}$), calculated monoisotopic mass ($[M]_{calc}$), mass accuracy, sum of the normalized peak areas of all their signals (Norm. Peak Area), molecular weight of the protein gel band origin, sequence (with position in protein and modifications shown) and number of trypsin missed cleavages. A 100% sequence coverage of calmodulin and melittin was obtained in control samples within a mass accuracy of 20 ppm for the peptide matches. The average number of missed cleavages for the identified peptides was one, suggesting that the trypsin digestion was efficient. The following expected modifications were confirmed: trimethylated K115 (+42.06 Da), acetylated calmodulin N-terminus (+42.01 Da) and amidated melittin C terminus (-0.98 Da), which are denoted as (tm), (ac),

3.3. Mass Spectrometric Analysis

and (am), respectively. A subset of calmodulin peptides containing oxidized M residues (at positions 36, 51, 71, 72, 76, 109, 124, 144,145) and deamidated asparagines (at positions 97 and 111) were observed producing a mass shift of +15.99 and +0.98 Da and denoted as (ox) and (dm), respectively. Peptides with both unmodified and modified methionines and asparagines were observed. Oxidation of methionines can occur during sample preparation such as the extensive vacuum drying of samples[150] and the deamidation of glutamine/asparagine is common during trypsin digestion [151]. Overall, it was confirmed that experimental processing itself did not result in significant protein loss and conveyed what peptides and modifications are expected to be present in cross-linked samples. Based on these findings, cross-linking or modification on K115, the acetylated calmodulin N-terminal residue, and the amidated melittin C-terminal residue were not considered.

The position of each SDS-PAGE gel band was used to estimate the molecular weight of the protein in which each peptide was identified, post in-gel trypsin digestion and MS analysis. The molecular weight of the proteins in which all melittin and calmodulin peptides originated from were < 10 kDa and 10 - 20 kDa, respectively. This agrees with expected molecular weight of melittin tetramers (~6 kDa) and a single calmodulin molecule (~17 kDa). In addition, this suggests that calmodulin and melittin exist as unbound species in the control sample.

3.3. Mass Spectrometric Analysis

Table 3.1: A list of MS/MS confirmed peptides in the control calmodulin-melittin sample. The m/z , experimental mass, calculated mass, mass accuracy, normalized peak area, molecular weight of the protein gel band origin, sequence and number of missed cleavages for each peptide are listed left to right.

| m/z | z | [M] _{exp} (Da) | [M] _{calc} (Da) | Mass Accuracy (ppm) | Norm. Peak Area | Protein MW Origin (kDa) | Peptide Sequence | Missed Cleavages |
|----------------------------|-----|----------------------------|-----------------------------|------------------------|-------------------------|----------------------------|--|------------------|
| Calmodulin Peptides | | | | | | | | |
| 782.37 | 2 | 1562.75 | 1562.74 | 8.3 | 4.02% | 10 - 20 | ¹ A(ac)DQLTEEQIAEFK ¹³ | 0 |
| 478.74 | 2 | 955.48 | 955.47 | 13.7 | 0.03% | 10 - 20 | ¹⁴ EAFSLFDK ²¹ | 0 |
| 855.42 | 2 | 1708.84 | 1708.84 | 4.0 | 2.94x10 ⁻³ % | 10 - 20 | ²² DGDGTTITTKELGTVM ^(ox) R ³⁷ | 1 |
| 411.21 | 2 | 820.41 | 820.42 | 10.6 | 6.96% | 10 - 20 | ³¹ ELGTVM ^(ox) R ³⁷ | 0 |
| 403.22 | 2 | 804.42 | 804.42 | 4.5 | 3.50% | 10 - 20 | ³¹ ELGTVMR ³⁷ | 0 |
| 1030.21 | 4 | 4116.82 | 4116.84 | 3.9 | 0.30% | 10 - 20 | ³⁸ SLGQNPTEAELQDM ^(ox) INEVDADGNGTIDFPEFLTM ^(ox) M ^(ox) AR ⁷⁴ | 0 |
| 1058.24 | 4 | 4228.92 | 4228.93 | 2.3 | 0.01% | 10 - 20 | ³⁸ SLGQNPTEAELQDM ^(ox) INEVDADGNGTIDFPEFLTM ^(ox) MARK ⁷⁵ | 1 |
| 1062.24 | 4 | 4244.92 | 4244.93 | 3.5 | 6.43x10 ⁻³ % | 10 - 20 | ³⁸ SLGQNPTEAELQDM ^(ox) INEVDADGNGTIDFPEFLTM ^(ox) M ^(ox) ARK ⁷⁵ | 1 |
| 901.82 | 5 | 4504.05 | 4504.07 | 3.3 | 0.01% | 10 - 20 | ³⁸ SLGQNPTEAELQDMINEVDADGNGTIDFPEFLTM ^(ox) M ^(ox) ARKM ^(ox) K ⁷⁷ | 2 |
| 1131.2 | 4 | 4520.05 | 4520.07 | 4.4 | 0.01% | 10 - 20 | ³⁸ SLGQNPTEAELQDM ^(ox) INEVDADGNGTIDFPEFLTM ^(ox) M ^(ox) ARKM ^(ox) K ⁷⁷ | 2 |
| 992.48 | 2 | 1982.94 | 1982.94 | 1.6 | 0.14% | 10 - 20 | ⁷⁵ KMKDTSDEEEIREAFR ⁹⁰ | 3 |
| 451.54 | 3 | 1351.59 | 1351.59 | 1.9 | 17.38% | 10 - 20 | ⁷⁶ MKDTDSEEEIR ⁸⁶ | 1 |
| 684.8 | 2 | 1367.59 | 1367.59 | 1.8 | 1.70% | 10 - 20 | ⁷⁶ M ^(ox) KDTSDEEEIR ⁸⁶ | 1 |
| 619.29 | 3 | 1854.84 | 1854.84 | 1.0 | 1.57% | 10 - 20 | ⁷⁶ MKDTDSEEEIREAFR ⁹⁰ | 2 |
| 464.72 | 4 | | | | 1.74% | | | |
| 624.62 | 3 | 1870.84 | 1870.84 | 1.7 | 6.19% | 10 - 20 | ⁷⁶ M ^(ox) KDTSDEEEIREAFR ⁹⁰ | 2 |
| 547.24 | 2 | 1092.46 | 1092.46 | 2.6 | 11.22% | 10 - 20 | ⁷⁸ DTDSEEEIR ⁸⁶ | 0 |
| 798.86 | 2 | 1595.71 | 1595.71 | 2.3 | 1.27% | 10 - 20 | ⁷⁸ DTDSEEEIREAFR ⁹⁰ | 1 |
| 532.91 | 3 | | | | 1.48% | 10 - 20 | | |
| 506.27 | 2 | 1010.52 | 1010.52 | 0.0 | 4.78x10 ⁻³ % | 10 - 20 | ⁸⁷ EAFRVFDK ⁹⁴ | 1 |
| 877.94 | 2 | 1753.86 | 1753.86 | 2.0 | 1.88% | 10 - 20 | ⁹¹ VFDKDGNGYISAAELR ¹⁰⁶ | 1 |
| 585.63 | 3 | | | | 8.46% | 10 - 20 | | |
| 878.43 | 2 | 1754.85 | 1754.86 | 7.1 | 0.23% | 10 - 20 | ⁹¹ VFDKDGNG ^(dm) GYISAAELR ¹⁰⁶ | 1 |
| 585.96 | 3 | | | | 6.23% | | | |
| 633.31 | 2 | 1264.60 | 1264.60 | 0.1 | 1.04% | 10 - 20 | ⁹⁵ DGNGYISAAELR ¹⁰⁶ | 0 |
| 633.8 | 2 | 1265.59 | 1265.60 | 12.6 | 0.10% | 10 - 20 | ⁹⁵ DGN ^(dm) GYISAAELR ¹⁰⁶ | 0 |
| 601.05 | 4 | 2400.17 | 2400.12 | 19.6 | 0.52% | 10 - 20 | ¹⁰⁷ HVMTNLGEK ^(tm) LTDEEVDEMIR ¹²⁶ | 1 |
| 801.39 | 3 | 2401.15 | 2401.12 | 12.9 | 0.17% | 10 - 20 | ¹⁰⁷ HVMTN ^(dm) LGEK ^(tm) LTDEEVDEMIR ¹²⁶ | 1 |
| 601.29 | 4 | | | | 0.02% | | | |
| 1209.9 | 2 | 2416.16 | 2416.12 | 17.4 | 0.53% | 10 - 20 | ¹⁰⁷ HVM ^(ox) TNLGEK ^(tm) LTDEEVDEMIR ¹²⁶ | 1 |
| 605.05 | 4 | | | | 3.50% | | | |
| 806.72 | 3 | 2417.15 | 2417.12 | 10.8 | 0.01% | 10 - 20 | ¹⁰⁷ HVM ^(ox) TN ^(dm) LGEK ^(tm) LTDEEVDEMIR ¹²⁶ | 1 |
| 605.29 | 4 | | | | 0.39% | | | |
| 811.73 | 3 | 2432.16 | 2432.12 | 15.2 | 1.72% | 10 - 20 | ¹⁰⁷ HVM ^(ox) TNLGEK ^(tm) LTDEEVDEM ^(ox) IR ¹²⁶ | 1 |
| 609.05 | 4 | | | | 7.91% | | | |
| 1245.54 | 2 | 2489.08 | 2489.07 | 2.8 | 0.22% | 10 - 20 | ¹²⁷ EADIDGDGQVNYEEFVQMMTAK ¹⁴⁸ | 0 |
| 836.02 | 3 | 2505.07 | 2505.07 | 0.6 | 2.55% | 10 - 20 | ¹²⁷ EADIDGDGQVNYEEFVQM ^(ox) MTAK ¹⁴⁸ | 0 |
| 841.36 | 3 | 2521.07 | 2521.07 | 1.0 | 1.49% | 10 - 20 | ¹²⁷ EADIDGDGQVNYEEFVQM ^(ox) M ^(ox) TAK ¹⁴⁸ | 0 |
| Melittin Peptides | | | | | | | | |
| 657.43 | 1 | 656.43 | 656.42 | 14.8 | 0.41% | < 10 | ¹ GIGAVLK ⁷ | 0 |
| 756.46 | 2 | 1510.91 | 1510.91 | 1.5 | 0.01% | < 10 | ⁸ VLTTLGLPALISWIK ²¹ | 0 |
| 504.64 | 3 | | | | 4.82% | | | |
| 556.68 | 3 | 1667.01 | 1667.01 | 2.0 | 0.04% | < 10 | ⁸ VLTTLGLPALISWIKR ²² | 1 |
| 442.27 | 5 | 2206.37 | 2206.35 | 10.7 | 6.80x10 ⁻³ % | < 10 | ⁸ VLTTLGLPALISWIKRKRQQ ^(am) ²⁶ | 3 |

3.3.1.2 PFA Modified Peptide Analysis via MaxQuant

A MaxQuant MS/MS search was used to determine the modified and unmodified peptides in the PFA treated sample. This was used to shed light on reaction products of the first, i.e. the modification, step of PFA cross-linking (see Figure 1.6) and aid in clarifying cross-linking mechanisms. Unlike established cross-linkers, the modification step in PFA cross-linking produces two stable intermediates: methylol (+30 Da mass shift), which dehydrates into a Schiff Base (+12 Da mass shift). In the calmodulin-melittin system, R, K and N-terminal melittin residues can potentially form PFA modifications under the mild PFA reaction conditions utilized (physiological pH, 1% PFA, and 6 hour reaction time). Using MaxQuant, a total of 41 and 30 unmodified (Table 3.2) and PFA-modified (Table 3.3), respectively unique peptides were identified in the PFA treated calmodulin-melittin sample with a mass accuracy within 30 ppm. Similar to the control sample, the PFA peptides produced a 100% sequence coverage of calmodulin and melittin with the following modifications: trimethylated K115, acetylated calmodulin N-terminus, amidated melittin C terminus, oxidation of M and deamidation of N. In the modified peptides, the following calmodulin residues appeared with a +12 Da mass shift: K75, K77, K94, K148, R106, R74, R86, and R90. A +12 Da mass shift indicates that either a Schiff Base formed and did not proceed in the second step of the cross-linking reaction or an intrapeptide cross-link formed.

The following calmodulin residues appeared with a +30 Da mass shift: K30, K75, K77, and K94. A +30 Da mass shift indicates that a methylol modification formed, which means that these residues were modified in the first step of the cross-linking reaction and did not proceed to form a cross-link in the second step. Interestingly, no R residues were observed with a +30 Da modification suggesting that R is less reactive in the first, modification step of the reaction. This is consistent with previous PFA studies in which R was shown to be less reactive than K residues in a myoglobin model protein system[36] and in model peptides[85]. If the +12 Da modified R residues are Schiff Bases and not intrapeptide cross-links, another possibility is that methylol modified R is significantly less stable than Schiff Base

3.3. Mass Spectrometric Analysis

modified R residues, favoring the immediate dehydration of the methylol modified R residue to a Schiff Base. For melittin, A +12 Da mass shift was localized on G1 in the peptides ¹GIGAVLKVLTTGLPALISWIK²¹ and ¹GIGAVLK⁷ indicating that either a Schiff Base or intrapeptide cross-link formed in this peptide. The melittin peptide ⁸VLTGTPALISWIKRKRQQ^(am)²⁶ appeared with a +24 and +36 Da modification. Since the C-terminus of the peptide contains four consecutive potential PFA modification sites (²¹KRKR²⁴), it was not possible to localize the modifications and/or intrapeptide cross-links to individual residues.

For the PFA treated sample, the molecular weight range of the proteins in which the peptides originated from are listed in Tables 3.2 and 3.3. This was estimated based on the position of SDS-PAGE gel band in which each peptide was identified in post in-gel trypsin digestion and MS analysis. In addition, the molecular weight correction factor (+ 3.2 kDa) calculated in section 3.2.3 was applied. No peptides from proteins < 14 kDa were identified suggesting that no isolated melittin tetramers or unbound melittin peptides were identified, which is consistent with the lack of a gel band at this molecular weight position observed in the SDS-PAGE. There were 30 out of 71 peptides (14 with PFA modifications) originating from 14- 19 kDa proteins, which could correspond to either unmodified or modified calmodulin, intramolecular cross-linked calmodulin or intermolecular cross-linked melittin tetramers. A total of 45 out of 71 peptides (19 with PFA modifications) from 19- 33 kDa proteins. which were most likely from intramolecular cross-linked calmodulin-melittin complexes (~ 20 kDa) or intermolecular cross-linked calmodulin molecules (~34 kDa). Finally, 41 out 71 peptides (15 with PFA modifications) came from proteins > 33 kDa, which were most likely from protein complexes with intermolecular cross-links between multiple calmodulin or calmodulin-melittin complex molecules. Therefore, peptides with the same identity appeared across proteins with different molecular weights i.e. different types of cross-linking. In addition, both unmodified and modified calmodulin peptides were identified in the digest of 14-19, 19-33 and > 33 kDa protein samples. Although unmodified melittin peptides were identified in the digest of 14-19, 19-33 and > 33 kDa protein samples, modified melittin peptides were only identified in the digest of 14-19 and 19-33 kDa protein samples. Overall, this supports the identification of calmodulin-melittin peptides from crosslinked species via MS/MS, providing

3.3. *Mass Spectrometric Analysis*

assurance for the subsequent MS/MS identification of cross-linked peptides.

3.3. Mass Spectrometric Analysis

Table 3.2: A list of MS/MS confirmed calmodulin-melittin peptides in PFA treated sample without PFA modifications. The m/z , experimental mass, calculated mass, mass accuracy, normalized peak area, molecular weight of the protein gel band origin, sequence and number of missed cleavages for each peptide are listed left to right.

| m/z | z | [M] _{exp} | [M] _{calc} | Mass Accuracy (ppm) | Norm. Peak Area | Protein MW Origin (kDa) | Sequence | Missed cleavages |
|----------------------------|-----|--------------------|---------------------|---------------------|-----------------|-------------------------|--|------------------|
| Calmodulin Peptides | | | | | | | | |
| 782.38 | 2 | 1562.76 | 1562.76 | 0.0 | 11.51% | 14-33 | ¹ A ^(ac) DQLTEEQIAEFK ¹³ | 0 |
| 478.74 | 2 | 955.48 | 955.47 | 13.7 | 7.10% | 14-19 | ¹⁴ EAFSLFDK ²¹ | 0 |
| 570.62 | 3 | 1708.85 | 1708.83 | 12.8 | 0.17% | > 33 | ²² DGDGTITTKELGTVM ^(ox) R ³⁷ | 1 |
| 403.22 | 2 | 804.43 | 804.42 | 18.1 | 28.47% | < 19, >33 | ³¹ ELGTVMR ³⁷ | 0 |
| 411.21 | 2 | 820.43 | 820.41 | 17.7 | 45.90% | 14-19, >33 | ³¹ ELGTVM ^(ox) R ³⁷ | 0 |
| 1026.21 | 4 | 4100.86 | 4100.83 | 7.1 | 0.23% | 14-19 | ³⁸ SLGQNPTAEALQDMINEVDADGNGTIDFPEFLTM ^(ox) M ^(ox) AR ⁷⁴ | 0 |
| 1030.21 | 4 | 4116.85 | 4116.82 | 7.1 | 5.47% | > 19 | ³⁸ SLGQNPTAEALQDM ^(ox) INEVDADGNGTIDFPEFLTM ^(ox) M ^(ox) AR ⁷⁴ | 0 |
| 1030.46 | 4 | 4117.84 | 4117.81 | 7.1 | 0.05% | 14-33 | ³⁸ SLGQNPTAEALQDM ^(ox) INEVDADGN ^(dm) GTIDFPEFLTM ^(ox) M ^(ox) AR ⁷⁴ | 0 |
| 1062.24 | 4 | 4244.95 | 4244.92 | 6.9 | 0.92% | 19-33 | ³⁸ SLGQNPTAEALQDM ^(ox) INEVDADGNGTIDFPEFLTM ^(ox) M ^(ox) ARK ⁷⁵ | 1 |
| 901.82 | 5 | 4504.09 | 4504.05 | 8.1 | 0.13% | 14-19 | ³⁸ SLGQNPTAEALQDMINEVDADGNGTIDFPEFLTM ^(ox) M ^(ox) ARKM ^(ox) K ⁷⁷ | 2 |
| 905.02 | 5 | 4520.09 | 4520.05 | 8.0 | 0.77% | > 14 | ³⁸ SLGQNPTAEALQDM ^(ox) INEVDADGNGTIDFPEFLTM ^(ox) M ^(ox) ARKM ^(ox) K ⁷⁷ | 2 |
| 494.24 | 3 | 1479.71 | 1479.69 | 14.8 | 0.24% | 14-33 | ⁷⁵ KMKDTSDEEIR ⁸⁶ | 2 |
| 499.57 | 3 | 1495.70 | 1495.68 | 14.6 | 0.16% | 14-19 | ⁷⁵ KM ^(ox) KDTSDEEIR ⁸⁶ | 2 |
| 496.74 | 4 | 1982.97 | 1982.94 | 14.7 | 1.32% | > 33 | ⁷⁵ KMKDTSDEEIREAFR ⁹⁰ | 3 |
| 500.74 | 4 | 1998.96 | 1998.93 | 14.6 | 18.77% | >14 | ⁷⁵ KM ^(ox) KDTSDEEIREAFR ⁹⁰ | 3 |
| 676.8 | 2 | 1351.61 | 1351.59 | 10.8 | 2.54% | > 19 | ⁷⁶ MKDTSDSEEIR ⁸⁶ | 1 |
| 684.8 | 2 | 1367.60 | 1367.59 | 10.6 | 3.41% | > 33 | ⁷⁶ M ^(ox) KDTSDEEIR ⁸⁶ | 1 |
| 619.29 | 3 | 1854.86 | 1854.84 | 11.8 | 1.15% | > 19 | ⁷⁶ MKDTSDSEEIREAFR ⁹⁰ | 2 |
| 936.43 | 2 | 1870.85 | 1870.84 | 7.8 | 19.40% | > 19 | ⁷⁶ M ^(ox) KDTSDEEIREAFR ⁹⁰ | 2 |
| 722.35 | 5 | 3606.73 | 3606.69 | 10.1 | 0.43% | > 14 | ⁷⁶ M ^(ox) KDTSDEEIREAFRVFDKDGNGYISAAELR ¹⁰⁶ | 4 |
| 547.24 | 2 | 1092.47 | 1092.46 | 13.3 | 10.65% | 14-33 | ⁷⁸ DTSDSEEIR ⁸⁶ | 0 |
| 798.86 | 2 | 1595.72 | 1595.71 | 9.1 | 12.38% | > 33 | ⁷⁸ DTSDSEEIREAFR ⁹⁰ | 1 |
| 696 | 3 | 2084.99 | 2084.97 | 10.5 | 0.11% | > 19 | ⁷⁸ DTSDSEEIREAFRVFDK ⁹⁴ | 2 |
| 833.9 | 4 | 3331.59 | 3331.56 | 8.7 | 3.80% | > 14 | ⁷⁸ DTSDSEEIREAFRVFDKDGNGYISAAELR ¹⁰⁶ | 3 |
| 753.38 | 3 | 2257.13 | 2257.11 | 9.7 | 2.41% | > 14 | ⁸⁷ EAFRVFDKDGNGYISAAELR ¹⁰⁶ | 2 |
| 753.71 | 3 | 2258.12 | 2258.10 | 9.7 | 5.46% | 14-19, >33 | ⁸⁷ EAFRVFDKDGNGYISAAELR ¹⁰⁶ | 2 |
| 877.94 | 2 | 1753.88 | 1753.86 | 8.3 | 101.38% | > 14 | ⁹¹ VFDKDGNGYISAAELR ¹⁰⁶ | 1 |
| 878.43 | 2 | 1754.86 | 1754.85 | 8.3 | 7.53% | > 14 | ⁹¹ VFDKDGNGYISAAELR ¹⁰⁶ | 1 |
| 633.31 | 2 | 1264.62 | 1264.60 | 11.5 | 19.10% | > 33 | ⁹⁵ DGNGYISAAELR ¹⁰⁶ | 0 |
| 633.8 | 2 | 1265.60 | 1265.59 | 11.5 | 1.94% | < 33 | ⁹⁵ DGN ^(dm) GYISAAELR ¹⁰⁶ | 0 |
| 601.05 | 4 | 2400.20 | 2400.17 | 12.1 | 3.70% | 14-19 | ¹⁰⁷ HVMTNLGEK ^(tm) LTDEEVDEMIR ¹²⁶ | 1 |
| 601.29 | 4 | 2401.18 | 2401.15 | 12.1 | 0.21% | > 19 | ¹⁰⁷ HVMTN ^(dm) LGEK ^(tm) LTDEEVDEMIR ¹²⁶ | 1 |
| 605.05 | 4 | 2416.19 | 2416.16 | 12.0 | 52.08% | 14-33 | ¹⁰⁷ HVM ^(ox) TNLGEK ^(tm) LTDEEVDEMIR ¹²⁶ | 1 |
| 605.29 | 4 | 2417.17 | 2417.15 | 12.0 | 0.39% | >14 | ¹⁰⁷ HVM ^(ox) TN ^(dm) LGEK ^(tm) LTDEEVDEMIR ¹²⁶ | 1 |
| 609.05 | 4 | 2432.19 | 2432.16 | 12.0 | 49.32% | 19-33 | ¹⁰⁷ HVM ^(ox) TNLGEK ^(tm) LTDEEVDEM ^(ox) IR ¹²⁶ | 1 |
| 609.29 | 4 | 2433.17 | 2433.14 | 12.0 | 13.26% | >14 | ¹⁰⁷ HVM ^(ox) TN ^(dm) LGEK ^(tm) LTDEEVDEM ^(ox) IR ¹²⁶ | 1 |
| 839.36 | 3 | 2515.07 | 2515.05 | 8.7 | 0.40% | >14 | ¹²⁷ EADIDGDGQVNYEEFVQM ^(ox) MTAK ¹⁴⁸ | 0 |
| Melittin Peptides | | | | | | | | |
| 657.43 | 1 | 656.43 | 656.42 | 12.8 | 3.29% | 14-33 | ¹ GIGAVLK ⁷ | 0 |
| 756.46 | 2 | 1510.93 | 1510.91 | 9.6 | 1.56% | > 19 | ⁸ VLTGTPALISWIK ²¹ | 0 |
| 834.51 | 2 | 1667.03 | 1667.01 | 8.7 | 8.36% | 14-19, >33 | ⁸ VLTGTPALISWIKR ²² | 1 |
| 599.38 | 3 | 1795.13 | 1795.11 | 12.2 | 0.15% | 14-33 | ⁸ VLTGTPALISWIKR ²³ | 2 |

3.3. Mass Spectrometric Analysis

Table 3.3: A list of MS/MS confirmed calmodulin-melittin peptides in PFA treated sample with PFA modifications. The m/z , experimental mass, calculated mass, mass accuracy, normalized peak area, molecular weight of the protein gel band origin, sequence and number of missed cleavages for each peptide are listed left to right. (+12) and (+30) denotes a Schiff Base/Intrapeptide cross-link and methylol, respectively, localized on the residue before it.

| m/z | z | $[M]_{exp}$ | $[M]_{calc}$ | Mass Accuracy (ppm) | Norm. Peak Area | Protein MW Origin (kDa) | Sequence | Missed cleavages |
|----------------------------|-----|-------------|--------------|---------------------|-------------------------|-------------------------|---|------------------|
| Calmodulin Peptides | | | | | | | | |
| 575.29 | 3 | 1722.87 | 1722.85 | 11.6 | 0.09% | > 33 | ²² DGDGTITTK(+30)ELGTVMR ³⁷ | 1 |
| 1133.02 | 4 | 4528.08 | 4528.05 | 6.6 | 0.53% | 14-33 | ³⁸ SLGQNPTEAELQDMINEVDADGNGTIDFPEFLTMMAR(+12)K(+12)MK ⁷⁷ | 0 |
| 754.85 | 2 | 1507.70 | 1507.68 | 13.3 | 1.06% | >14 | ⁷⁵ K(+12)M(ox)KDTDSEEEIR ⁸⁶ | 1 |
| 508.24 | 3 | 1521.72 | 1521.70 | 13.1 | 0.04% | 14-19 | ⁷⁵ K(+30)MK(+12)DTDSEEEIR ⁸⁶ | 1 |
| 513.57 | 3 | 1537.71 | 1537.69 | 13.0 | 1.18% | >14 | ⁷⁵ K(+30)M(ox)K(+12)DTDSEEEIR ⁸⁶ | 1 |
| 499.74 | 4 | 1994.96 | 1994.94 | 10.0 | 4.01x10 ⁻³ % | 14-19 | ⁷⁵ K(+12)MKDTDSEEEIREAFR ⁹⁰ | 2 |
| 669.99 | 3 | 2006.97 | 2006.94 | 14.9 | 0.06% | 14-19 | ⁷⁵ K(+12)MK(+12)DTDSEEEIREAFR ⁹⁰ | 2 |
| 503.74 | 4 | 2010.96 | 2010.93 | 14.9 | 0.92% | 19-33 | ⁷⁵ K(+12)M(ox)KDTDSEEEIREAFR ⁹⁰ | 2 |
| 403.6 | 5 | 2013.00 | 2012.95 | 24.8 | 7.86% | > 33 | ⁷⁵ KM(ox)K(+30)DTDSEEEIREAFR ⁹⁰ | 2 |
| 506.74 | 4 | 2022.96 | 2022.93 | 14.8 | 2.74% | 14-33 | ⁷⁵ K(+12)M(ox)K(+12)DTDSEEEIREAFR ⁹⁰ | 2 |
| 511.24 | 4 | 2040.96 | 2040.94 | 9.8 | 0.96% | > 33 | ⁷⁵ K(+30)MK(+12)DTDSEEEIREAFR ⁹⁰ | 3 |
| 455.54 | 3 | 1363.62 | 1363.59 | 22.0 | 0.07% | 14-33 | ⁷⁶ MK(+12)DTDSEEEIR ⁸⁶ | 3 |
| 461.54 | 3 | 1381.62 | 1381.60 | 14.5 | 3.19% | >14 | ⁷⁶ MK(+30)DTDSEEEIR ⁸⁶ | 3 |
| 627.29 | 3 | 1878.87 | 1878.84 | 16.0 | 0.16% | >14 | ⁷⁶ MK(+12)DTDSEEEIR(+12)EAFR ⁹⁰ | 3 |
| 628.62 | 3 | 1882.86 | 1882.84 | 10.6 | 0.13% | 14-19, >33 | ⁷⁶ M(ox)K(+12)DTDSEEEIREAFR ⁹⁰ | 3 |
| 472.22 | 4 | 1884.88 | 1884.85 | 15.9 | 3.00% | 14-19 | ⁷⁶ MK(+30)DTDSEEEIREAFR ⁹⁰ | 3 |
| 632.62 | 3 | 1894.86 | 1894.84 | 10.6 | 2.06% | 14-33 | ⁷⁶ M(ox)K(+12)DTDSEEEIR(+12)EAFR ⁹⁰ | 3 |
| 634.62 | 3 | 1900.86 | 1900.85 | 5.3 | 0.71% | > 19 | ⁷⁶ M(ox)K(+30)DTDSEEEIREAFR ⁹⁰ | 1 |
| 727.15 | 5 | 3630.73 | 3630.69 | 9.9 | 0.44% | > 19 | ⁷⁶ M(ox)KDTDSEEEIREAFR(+12)VFDK(+12)DGNGYISAAELR ¹⁰⁶ | 1 |
| 839.9 | 4 | 3355.60 | 3355.56 | 11.9 | 0.64% | >14 | ⁷⁸ DTDSEEEIR(+12)EAFRVFDK(+12)DGNGYISAAELR ¹⁰⁶ | 4 |
| 761.38 | 3 | 2281.14 | 2281.11 | 13.2 | 5.29% | 14-33 | ⁸⁷ EAFR(+12)VFDK(+12)DGNGYISAAELR ¹⁰⁶ | 2 |
| 589.63 | 3 | 1765.89 | 1765.86 | 17.0 | 5.37% | > 19 | ⁹¹ VFDK(+12)DGNGYISAAELR ¹⁰⁶ | 2 |
| 892.94 | 2 | 1783.88 | 1783.87 | 5.6 | 12.48% | >14 | ⁹¹ VFDK(+30)DGNGYISAAELR ¹⁰⁶ | 0 |
| 839.41 | 5 | 4192.05 | 4192.01 | 9.5 | 0.45% | >14 | ⁹¹ VFDK(+12)DGNGYISAAELR(+12)HVM(ox)TNLGEK(tm)LTDEEVDEM(ox)IR ¹²⁶ | 4 |
| 837.7 | 3 | 2510.10 | 2510.07 | 12.0 | 0.11% | >14 | ¹²⁷ EADIDGDGQVNYEEFVQMMTAK(+12) ¹⁴⁸ | 2 |
| Melittin Peptides | | | | | | | | |
| 669.43 | 1 | 668.43 | 668.41 | 30.1 | 2.04% | 19-33 | ¹ G(+12)IGAVLK ⁷ | 0 |
| 721.45 | 3 | 2161.35 | 2161.32 | 13.9 | 8.95x10 ⁻⁴ % | 14-19 | ¹ G(+12)IGAVLKVLTTGLPALISWIK ²¹ | 0 |
| 659.41 | 3 | 1975.23 | 1975.21 | 10.1 | 1.04% | 14-19 | ⁸ VLTGLPALISWI[KRKR](+24) ²⁴ | 4 |
| 744.44 | 3 | 2230.34 | 2230.33 | 4.5 | 0.18% | 14-33 | ⁸ VLTGLPALISWI[KRKR](+24)QQ(am) ²⁶ | 3 |
| 748.44 | 3 | 2242.34 | 2242.33 | 4.5 | 0.02% | 14-19 | ⁸ VLTGLPALISWI[KRKR](+36)QQ(am) ²⁶ | 2 |

3.3.1.3 Trypsin Cleavage After Modified and Cross-linked Sites

All cross-linked samples were digested with trypsin, which specifically cleaves after residues that also are potential modification and cross-linking sites for EDC, PFA, sulfoDST,BS³,and sulfoEGS .i.e. K and R residues. Previous cross-linking experiments with NHS ester cross-linkers and EDC have claimed that trypsin rarely cleaves modified and/or cross-linked K residues, resulting in missed cleavages [34, 38, 152]. However, whether this applies to PFA cross-linked species must be examined to effectively predict products formed by cross-linking.

The cleavage of trypsin is triggered by the electrophilic center induced by the positively charged primary amino group on K and R residues under physiological pH conditions, which favors the nucleophilic attack by serine in the active site of trypsin (see Figure 3.7)[153]. Figure 3.6 displays four structures of K upon modification or cross-linking and their associated pKa values. The formation of PFA induced Schiff bases (pKa ~ 4 [154]) reduces the electrophilic nature of K residues, which diminishes serine's tendency to attack and cleave its peptide bond. This is supported by the lack of Schiff Base and/or methylol modifications on terminal K and R residues in peptides (with the exception of the calmodulin C-terminus, K148) observed in this study (see Table 3.3), suggesting that trypsin did not cleave after modified R and K residues, as expected. Furthermore, the average number of missed cleavages increased from 1.2 for the non-PFA modified peptides (Table 3.2) in the PFA-treated samples to 2.1 for the PFA modified peptides, suggesting that the overall trypsin cleavage efficiency is reduced at PFA modified sites. It is important to note that trypsin cleavage is not expected to occur at calmodulin K115 since it is trimethylated.

However, trypsin cleavage after PFA cross-linked K residues may still be possible since trypsin cleavage occurs after monomethylated K residues, both of which are similar to a secondary amine functional group[155]. This hypothesis is supported by the negligible difference in pKa values of primary amines and secondary amines (both with a pKa ~ 10.7), which suggests that these groups are protonated under trypsin digestion buffer conditions (ammonium bicarbonate buffer, pH = 7.8). Furthermore, it has been observed that PFA does not change the charge state of proteins treated with PFA, suggesting that the electrophilicity of K residues re-

3.3. Mass Spectrometric Analysis

quired to activate trypsin cleavage should not be significantly affected [36]. In contrast, when NHS esters react to form cross-links with K residues, an amide replaces the primary amine group, which has a much lower pKa value (-0.5). This results in a drastic reduction of the electrophilicity of lysine, hampering trypsin cleavage [156]. However, another factor to consider is that cross-linked multimers occupy a larger surface area than unmodified monomers and in order for trypsin to cleave, the cleavage site must be able to fit in the active site of trypsin. It remains to be seen whether PFA cross-links are potentially small enough to fit in the active site of trypsin. Therefore, the possibility of terminal K or R residues in tryptic peptides forming PFA cross-links was not ruled out in this study. However, based on the list of PFA modified peptides containing only internal modifications, it was assumed that trypsin does not cleave after Schiff Base or methylol modifications.

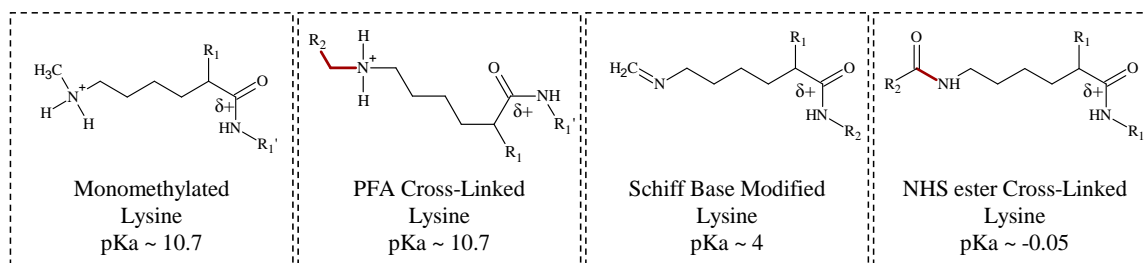


Figure 3.6: In the dotted boxes, the structures of proteins with a lysine side chain modified four ways as indicated with their respective pKa values are shown. Cross-linked bridges are highlighted in red.

3.3. Mass Spectrometric Analysis

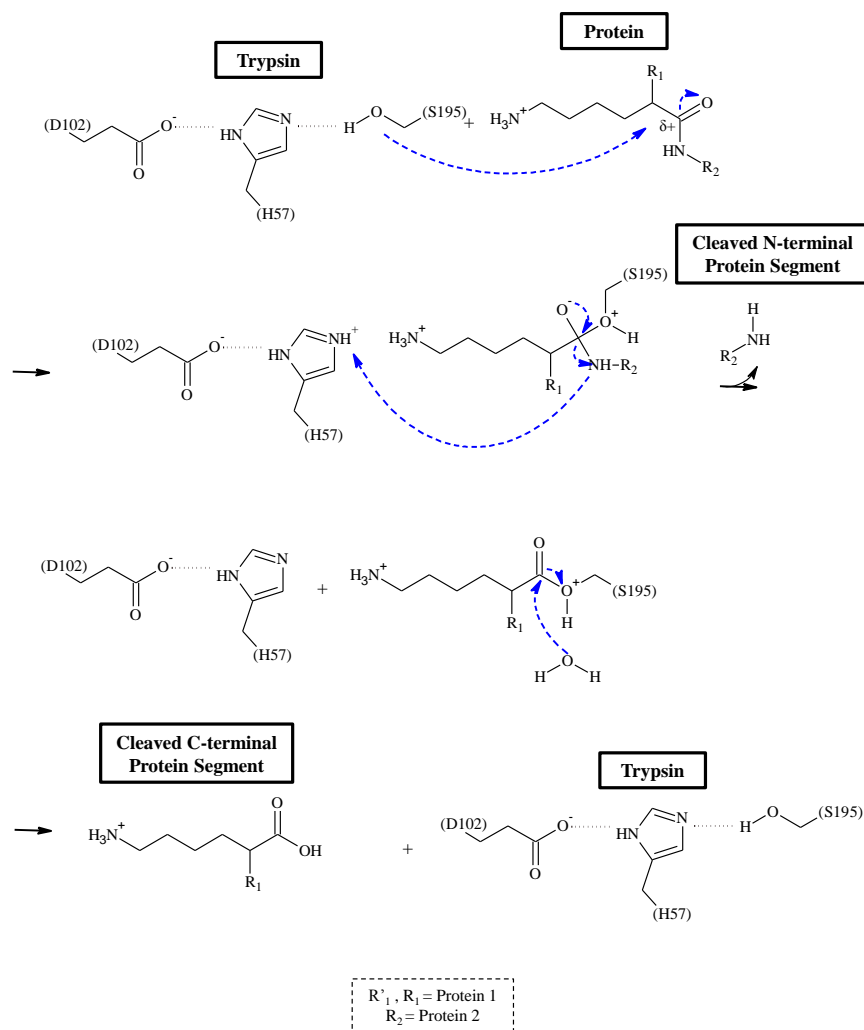


Figure 3.7: Trypsin's catalytic triad consists of aspartic acid (D102), histidine (H57) and serine (S195). Aspartic acid and histidine increase the nucleophilicity of serine, which attacks the partially positive carbonyl carbon of the protein. The positively charged amino group on lysine increases the electrophilicity of the carbonyl carbon. The peptide bond is cleaved and the trypsin catalyst is regenerated. (adapted from reference [153])

3.3. Mass Spectrometric Analysis

3.3.1.4 Cross-link Identification via Manual Data Analysis

In addition to unmodified and modified peptides, the trypsin digestion of cross-linked proteins also produces cross-linked peptides, which were evaluated manually via MS and MS/MS. Based on the identified unmodified/modified peptides, the amidation of the C-terminus, acetylation of the N-terminus, and trimethylation of K were set as fixed modifications and the deamidation of N and oxidation of M were considered as variable modifications. Four missed cleavages were accounted for since peptides with up to four missed cleavages were observed. A theoretical digestion using ExPASy Peptide mass[115] produced a total of list of 50 calmodulin and 17 melittin peptides without cross-linker-specific modifications or cross-links. Table 3.4 lists each cross-linker's potential modification and cross-linking sites specific to calmodulin and melittin.

Cross-linked masses were prepared assuming that the mass of a cross-linked species is equal to the mass of each peptide component plus the mass of the bridge and any extra modification.

Table 3.4: Cross-linking and modification sites in calmodulin and melittin for each cross-linker

| Cross-Linker | Melittin | | Calmodulin | | Peptide Components | Possible Cross-linked Species |
|-----------------|------------------------|------------------------|--------------------|------------------------|--------------------|-------------------------------|
| | Modification Sites | Cross-Linking Sites | Modification Sites | Cross-Linking Sites | | |
| EDC | 0 | 4 (3K + 1 Nterm) | 38 (17D + 21E) | 7 (K) | 353 | 123684 |
| PFA | 6 (3K + 1 Nterm + 2R) | 5 (1 Nterm + 2R + 1 Q) | 13 (6K + 6R) | 20 (6R + 6Q + 6N + 2Y) | 439 | 191796 |
| SulfoDST | 4 (3K + 1 Nterm) | 4 (3K + 1 Nterm) | 6 (K) | 6 (K) | 160 | 24675 |
| BS ³ | 4 (3K + 1 Nterm) | 4 (3K + 1 Nterm) | 6 (K) | 6 (K) | 160 | 24675 |
| SulfoEGS | 4 (3K + 1 Nterm) | 4 (3K + 1 Nterm) | 6 (K) | 6 (K) | 160 | 24675 |

In melittin, there are a total of five and six potential PFA modification and cross-linking reactive sites, respectively. In calmodulin, there are a total of 13 and 20 potential PFA modification and cross-linking reactive sites, respectively. To

3.3. Mass Spectrometric Analysis

illustrate the strategy used to derive a list of PFA modified peptides for preparing a list of theoretical cross-linked masses, peptide VFDKDGNGYISAAELR is discussed. For the peptide VFDKDGNGYISAAELR, K and R are potential modification sites and N, Y and R are potential cross-linking sites. In other words, K and R can each contain a +12 Da or +30 Da mass shift and N, Y and R can contain a +12 Da mass shift. Since trypsin is expected to not cleave after PFA modified residues, the terminal R residue is only considered as a potential cross-linking site in this case. PFA cross-linked masses must contain at least one +12 Da mass shift corresponding to the cross-linker bridge and when searching for interpeptide cross-linking, it is assumed that one of the reactive sites on each peptide is cross-linked to another peptide. Therefore, in the case of peptide VFDKDGNGYISAAELR, only three of the four sites can be modified or be occupied in intrapeptide cross-linking. When preparing the list of possible cross-linked masses, only the total modification mass between both peptide components is considered. Since possible modifications are also considered for the second peptide component, one less than the total reactive sites are considered for each peptide. Therefore seven total forms of peptide VFDKDGNGYISAAELR are included, i.e. the mass of the peptide plus 0, 12, 24, 30, 36, 42 and 54 Da. The presence of a methylol indicates that the residue is modified and not cross-linked. Thus, cross-linked masses that correspond to the mass of two peptides with only a +30 Da total mass shift are not possible.

In melittin, there are a total of zero and four potential EDC modification and cross-linking reactive sites, respectively. In calmodulin, there are a total of 38 and seven potential EDC modification and cross-linking reactive sites, respectively. Therefore, calmodulin contained the highest number of possible EDC modification sites. Nevertheless, the number of EDC cross-linking sites was equal to that of NHS esters. EDC can potentially also form +155 Da modifications on D or E residues. Although this EDC modification is rare under short EDC incubation times (1 hr), it was still considered [33, 34]. There are total of four and seven cross-linking/modification sites in melittin and calmodulin, respectively, for NHS ester cross-linkers. Hence, the number of possible cross-linked masses for EDC and NHS ester cross-linkers was significantly lower than for PFA.

Applying a similar approach to each peptide for calmodulin and melittin, pos-

3.3. Mass Spectrometric Analysis

sible modified peptide masses were derived. This was performed for PFA, EDC, and the NHS ester cross-linkers (sulfoDST, BS³, and sulfoEGS) to generate a list of 439, 353, and 160 peptide components, respectively, to make up cross-linked peptides. Mathematica was used to find every combination of all possible modified and unmodified peptide masses to create lists of possible cross-linked masses of 191,796, 123,684 and 24,675 species for PFA, EDC, and the NHS ester cross-linkers (sulfoDST, BS³, and sulfoEGS), respectively.

The total number of calmodulin-melittin cross-linked candidates identified at the MS level for EDC, PFA, sulfoDST, BS³ and sulfoEGS were 160, 335, 62, 77, and 158, respectively (listed in AppendixA.4). These findings are consistent with the relative cross-linking yield observed in the SDS-PAGE. PFA and EDC demonstrated the highest cross-linking yield and also the highest number of MS candidate cross-links. In addition, NHS ester cross-linkers demonstrated a lower cross-linking yield and lower number of candidates that increased proportional to their length.

3.3.2 Ribonuclease-S System

3.3.2.1 Control Peptide Analysis via MaxQuant

To determine the composition of a tryptic digest of RNaseS without cross-linking, the RNaseS control sample peptides were analyzed using MaxQuant. Although cleavage of RNaseA can occur between residues 16 – 21 to form RNaseS, the protein system would be highly complex if five forms of the S-peptide and five forms of the S-protein were considered. Therefore, the two major forms of each protein component based on literature findings were considered, which are produced from the cleavage after residue 19 and 20 of RNaseA [105, 106, 157, 158]. In addition, a band corresponding to RNaseA was observed in the SDS-PAGE of RNaseS and therefore, RNaseA peptides were also considered. A maximum of four missed cleavages were accounted for and the following variable modifications were searched: deamidation of N, oxidation of M and Carbamidomethyl on C.

Table 3.5 lists the RNaseS peptides identified by MaxQuant. A 92% sequence coverage for the RNaseS complex was obtained within a mass accuracy of 20 ppm. Two small S-protein peptides ⁶²NVAC^(cm)K⁶⁶ and ⁸⁶ETGSSK⁹¹ were miss-

3.3. Mass Spectrometric Analysis

ing from identified peptides, giving a sequence coverage less than 100 %. The MS signals corresponding to these peptides were also not identified manually. Therefore, it is hypothesized that these small peptides eluted too early from the HPLC column and/or did not ionize efficiently due to their short length and low charge state, and thus were not detected by MS. An average of one and three missed cleavages were observed for the S-protein and S-peptide, respectively. Non-reduced RNaseS possesses disulfide bonds between the following cysteine (C) residues: C26 to C84, C58 to C110, C40 to C95, and C65 to C72 [159]. Carbamidomethyl modifications from the reduction and alkylation of disulfide bonds were identified on C26, C72, C84, and C110. In addition, non-alkylated (reduced, without carbamidomethyl modifications) C26, C95, C58, and C40 residues were also identified. This suggests that the disulfide bonds in RNaseS were reduced, however, the alkylation was inefficient for C40, C58, and C95.

MaxQuant identified 22 unique peptides out of which five corresponded to RNaseA, two were products of the RNaseA cleavage after residue 19 (one S-peptide and one S-protein tryptic peptide), and 15 were products of the RNaseA cleavage after residue 20 (two S-peptides and 13 S-protein tryptic peptides). Therefore, the majority of the identified RNaseS peptides supported cleavage after RNaseA residue 20.

The position of the SDS PAGE gel band in which each peptide was identified was used to estimate the molecular weight of the protein from which each peptide originated. S-peptide and S-protein tryptic peptides were identified in < 12 kDa proteins, suggesting that the S-peptide (~ 2.2 kDa) and S-protein (~11.5 kDa) remained unbound in the control samples. The peptides would have been identified in ~ 13.7 kDa proteins if the S-peptide and S-protein were in their bound, RNaseS complex form. The RNaseA peptides were identified in 12- 15 kDa, as expected based on the molecular weight of RNaseA (~13.7 kDa).

3.3. Mass Spectrometric Analysis

Table 3.5: List of S-peptide, S-protein and RNaseA peptides in the control RNaseS sample. The m/z , experimental mass, calculated mass, mass accuracy, normalized peak area, molecular weight of the protein gel band origin, sequence and number of missed cleavages for each peptide are listed left to right.

| m/z | z | $[M]_{exp}$ | $[M]_{calc}$ | Mass Accuracy (ppm) | Norm. Peak Area | Protein MW Origin (kDa) | Sequence | Missed cleavages |
|---------------------------|-----|-------------|--------------|---------------------|-----------------|-------------------------|--|------------------|
| S-peptide Peptides | | | | | | | | |
| 722.68 | 3 | 2165.04 | 2165.02 | 10.1 | 0.16% | < 12 | ¹ KETAAAKFERQHMSSTSA ²⁰ | 3 |
| 1019.47 | 2 | 2036.94 | 2036.92 | 7.1 | 0.28% | < 12 | ² ETAAAKFERQHMSSTSA ²⁰ | 2 |
| 699.00 | 3 | 2094.00 | 2093.98 | 10.4 | 0.54% | < 12 | ¹ KETAAAKFERQHMSSTSA ¹⁹ *From S-peptide 1-19 | 3 |
| S-protein Peptides | | | | | | | | |
| 646.76 | 2 | 1291.51 | 1291.50 | 11.3 | 1.51% | < 12 | ²¹ SSSNYCNQMMK ³¹ | 0 |
| 662.75 | 2 | 1323.50 | 1323.49 | 11.0 | 1.17% | < 12 | ²¹ SSSNYCNQM ^(ox) M ^(ox) K ³¹ | 0 |
| 691.26 | 2 | 1380.53 | 1380.51 | 10.5 | 0.27% | < 12 | ²¹ SSSNYCNQMMKSRNLT ^(CM) K ³¹ | 0 |
| 755.02 | 3 | 2262.05 | 2262.03 | 9.7 | 0.82% | < 12 | ²¹ SSSNYCNQMMKSRNLT ³⁹ | 3 |
| 832.05 | 3 | 2493.16 | 2493.13 | 8.8 | 0.38% | < 12 | ²¹ SSSNYCNQMMKSRNLT ⁴¹ | 4 |
| 675.55 | 5 | 3372.74 | 3372.70 | 10.8 | 0.45% | < 12 | ³² SRNLT ³⁹ KDRCKPVNTFVHESLADVQAVCSQK ⁶¹ | 4 |
| 373.71 | 2 | 745.42 | 745.41 | 19.5 | 0.08% | < 12 | ²⁴ NLT ³⁹ KDR | 1 |
| 535.67 | 5 | 2673.34 | 2673.30 | 13.6 | 0.79% | < 12 | ³⁸ DRCKPVNTFVHESLADVQAVCSQK ⁶¹ | 2 |
| 801.73 | 3 | 2402.19 | 2402.17 | 9.1 | 6.87% | < 12 | ⁴⁰ CKPVNTFVHESLADVQAVCSQK ⁶¹ | 1 |
| 481.44 | 5 | 2402.21 | 2402.17 | 15.1 | | | | 1 |
| 601.55 | 4 | 2402.20 | 2402.17 | 12.1 | | | | 1 |
| 1151.46 | 2 | 2300.92 | 2300.91 | 6.3 | 8.33% | < 12 | ⁶⁷ NGQTNC ^(CM) YQSYSTM ^(ox) SITDC ^(CM) R ⁸⁵ | 0 |
| 429.69 | 2 | 857.39 | 857.37 | 17.0 | 11.88% | < 12 | ⁹² YPNCAYK ⁹⁸ | 0 |
| 937.47 | 3 | 2809.41 | 2809.39 | 7.8 | 0.99% | < 12 | ⁹⁹ TTQANKHIIVACEGNPYVPVHFDASV ¹²⁴ | 1 |
| 742.03 | 3 | 2223.10 | 2223.08 | 9.8 | 66.01% | < 12 | ¹⁰⁵ HIIVAC ^(CM) EGNPYVPVHFDASV ¹²⁴ | 0 |
| 855.73 | 3 | 2564.19 | 2564.17 | 8.5 | 0.17% | < 12 | ²⁰ ASSSNYCNQMMKSRNLT ⁴¹ *From S-protein 20-124 | 4 |
| RNaseA Peptides | | | | | | | | |
| 1337.61 | 3 | 4009.84 | 4009.81 | 5.4 | 0.06% | 12 – 20 | ² ETAAAKFERQHMSSTSAASSSNYCNQMMKSRNLT ³⁷ | 4 |
| 1237.56 | 3 | 3709.67 | 3709.65 | 5.9 | 0.31% | 12 – 20 | ⁸ FERQHMSSTSAASSSNYCNQMMKSRNLT ³⁹ | 4 |
| 1228.22 | 3 | 3681.66 | 3681.64 | 5.9 | 0.69% | 12 – 20 | ¹ KETAAAKFERQHMSSTSAASSSNYCNQMMKSR ³³ | 4 |
| 769.97 | 3 | 2306.92 | 2306.90 | 9.5 | 0.90% | 12 – 20 | ¹¹ QHMSSTSAASSSNYCNQMMK ³¹ | 0 |
| 851.02 | 2 | 2550.06 | 2550.04 | 8.6 | 0.10% | 12 – 20 | ¹ QHMSSTSAASSSNYCNQMMKSR ³³ | 1 |

3.3.2.2 PFA-Treated Peptide Analysis via MaxQuant

A MaxQuant MS/MS search was used to determine the modified and unmodified peptides in the PFA treated sample, to understand the modification reaction products formed from the PFA cross-linking of RNaseS. A total of eight unique peptides without PFA modifications and four unique peptides with PFA modifications were identified by MaxQuant within a mass accuracy of 15 ppm and are listed in Table 3.6. A 95% sequence coverage was obtained for the S-protein. Similar to the control sample, S-protein peptide $^{62}\text{NVACK}^{66}$ was missing from the identification. Among the identified peptides corresponding to the S-protein, the peptide $^{21}\text{SSSNYC}^{(CM)}\text{NQM}^{(ox)}\text{M}^{(ox)}\text{K}^{(+12)}\text{SR}^{(+12)}\text{NLTK}^{37}$ supported cleavage after RNaseA residue 20 and no peptides supported cleavage of RNaseA after residue 19. Carbamidomethyl modifications from the reduction and alkylation of RNaseS were identified on C26, C40, C58, C72, C84, C95, and C110, i.e. all expected C residues except C65. RNaseS peptides containing C65 were not identified. Only C95 also appeared in the reduced form without alkylation.

The average number of missed cleavages for the S-protein was one, which was consistent with that observed in the control samples. Consistent with the PFA modified calmodulin-melittin peptides, trypsin cleavage was not observed after PFA modified residues in the RNaseS system. A +12 Da modification was localized to S-protein K31, R33, K91, and K98. A +30 Da modification was only localized on S-protein K98. This confirms that K98, K31, and K91 are modified in the first step of the cross-linking reaction of RNaseS. The absence of +30 modified R residues identified in RNaseS is consistent with the observations in the calmodulin-melittin system. Since +12 Da could correspond to either a Schiff base or intrapeptide cross-link, it is ambiguous whether R33 (both a potential cross-linking and modification site) was modified or cross-linked.

No RNaseA peptides or S-peptide peptides were identified by MaxQuant. This is consistent with the SDS-PAGE of PFA treated RNaseS, where no band for RNaseA and for the S-peptide was observed. A band corresponding to unmodified S-peptide did not appear in the PFA treated sample but appeared in the control sample. However, a band for the RNaseS complex at ~14 kDa was observed for the PFA treated sample. This suggested that the S-peptide molecules are engaged

3.3. Mass Spectrometric Analysis

in intramolecular PFA cross-links with the S-protein. Previous kinetics experiments that examined trypsin cleavage in RNaseS, demonstrated that the S-peptide is not accessible for trypsin cleavage while bound to the S-protein in the complex. If the S-peptide is cleaved, then the interaction between the S-peptide and S-protein was diminished[160]. Thus, if cross-linking truly preserved the RNaseS complex and partially preserved the complex structure through SDS denaturation, then trypsin digestion products from the S-peptide would be expected to be rare and only cross-linked S-peptides would be expected. This may explain the absence of the S-peptide in the MaxQuant identified peptides from the PFA treated sample.

Table 3.6 lists the molecular weight of the protein each peptide originated from, which was estimated by SDS-PAGE. No peptides from < 12 kDa proteins were identified, supporting the absence of individual S-protein molecules or S-peptide molecules. Nine out of the twelve (two of them being PFA modified peptides) S-Protein peptides were from 12-20 kDa proteins, which are likely from intramolecular cross-linked RNaseS (~14 kDa). Seven out of the twelve (two of them being PFA modified peptides) S-Protein peptides were from 20-30 kDa proteins, which are likely from intermolecular cross-linked S-protein (~23 kDa) or RNaseS (~28 kDa) molecules. However, these hypotheses must be verified by identifying PFA cross-linked RNaseS peptides via MS/MS.

3.3. Mass Spectrometric Analysis

Table 3.6: List of S-peptide, S-protein and RNaseA peptides in the PFA treated RNaseS sample. The m/z , experimental mass, calculated mass, mass accuracy, normalized peak area, molecular weight of the protein gel band origin, sequence and number of missed cleavages for each peptide are listed left to right. (+12) and (+30) denotes a Schiff Base/Intra-peptide cross-link and methylol, respectively, localized on the residue before it.

| m/z | z | $[M]_{\text{exp}}$ | $[M]_{\text{calc}}$ | Mass Accuracy (ppm) | Norm. Peak Area | Protein MW Origin (kDa) | Sequence | Missed cleavages |
|--|-----|--------------------|---------------------|---------------------|-----------------|-------------------------|--|------------------|
| S-protein Peptides | | | | | | | | |
| 558.48 | 5 | 2787.38 | 2787.34 | 13.1 | 1.98% | 12 – 30 | ³⁸ DRC ^(CM) KPVNTFVHESLADVQAVC ^(CM) SQK ⁶¹ | 1 |
| 767.98 | 3 | 2300.93 | 2300.91 | 9.5 | 68.89% | 12-20, > 30 | ⁶⁷ NGQTNC ^(CM) YQSYSTM ^(ox) SITDC ^(CM) R ⁸⁵ | 0 |
| 964.40 | 3 | 2890.20 | 2890.18 | 7.6 | 2.24% | 12-20 | ⁶⁷ NGQTNC ^(CM) YQSYSTM ^(ox) SITDC ^(CM) RETGSSK ⁹¹ | 1 |
| 502.23 | 3 | 1503.69 | 1503.67 | 14.5 | 20.68% | > 12 | ⁸⁶ ETGSSKYPNC ^(CM) AYK ¹⁰⁴ | 1 |
| 537.76 | 4 | 2147.02 | 2147.00 | 13.6 | 8.01% | 20-30 | ⁸⁶ ETGSSKYPNC ^(CM) AYKTTQANK ¹⁰⁴ | 2 |
| 520.25 | 3 | 1557.75 | 1557.72 | 14.0 | 37.34% | 12 – 30 | ⁹² YPNC ^(CM) AYKTTQANK ¹⁰⁴ | 1 |
| 717.61 | 4 | 2866.44 | 2866.41 | 10.2 | 2.22% | 12-20 | ⁹⁹ TTQANKHIIIVAC ^(CM) EGNPYVPVHFDASV ¹²⁴ | 1 |
| 742.03 | 3 | 2223.10 | 2223.08 | 9.8 | 26.45% | > 12 | ¹⁰⁵ HIIIVAC ^(CM) EGNPYVPVHFDASV ¹²⁴ | 0 |
| S-protein Peptides with PFA Modifications | | | | | | | | |
| 702.31 | 3 | 2103.94 | 2103.91 | 10.4 | 6.44% | 12-20 | ²¹ SSSNYC ^(CM) NQM ^(ox) M ^(ox) K ⁽⁺¹²⁾ SR ⁽⁺¹²⁾ NLTK ³⁷ | 2 |
| 526.50 | 4 | 2102.00 | 2101.97 | 13.8 | 0.51% | 20-30 | ⁸⁶ ETGSSK ⁽⁺¹²⁾ YPNCA YKTTQANK ¹⁰⁴ | 2 |
| 524.25 | 3 | 1569.75 | 1569.72 | 13.9 | 9.64% | 20-30 | ⁹² YPNC ^(CM) AYK ⁽⁺¹²⁾ TTQANK ¹⁰⁴ | 1 |
| 530.25 | 3 | 1587.76 | 1587.74 | 13.8 | 115.59% | > 12 | ⁹² YPNC ^(CM) AYK ⁽⁺³⁰⁾ TTQANK ¹⁰⁴ | 1 |

3.3.2.3 Cross-link Identification via Manual Data Analysis

To determine RNaseS cross-linked species, the trypsin digest of cross-linked RNaseS samples were analyzed manually via MS using a similar data analysis procedure that was used for the calmodulin-melittin system. To simplify the theoretical cross-link search space, peptides from S-proteins and S-peptides formed from the cleavage of RNaseA after residue 20 were only considered, since this represents the majority of the peptides identified by MaxQuant. ExPASy Peptide mass [115] was used to perform a theoretical trypsin digestion to produce 80 S-protein

3.3. Mass Spectrometric Analysis

and 10 S-peptide peptides. Since control peptides with missed cleavages up to four were observed in the MaxQuant search results, four missed cleavages were considered. The following variable modifications were considered: deamidation of N, oxidation of M and Carbamidomethyl on C. Carbamidomethyl on C was set as a variable modification because peptides with both non-alkylated and alkylated cysteines were identified by MaxQuant. The number of possible unmodified peptides in the RNaseS system was 23 more than in the calmodulin-melittin system, making it a slightly more complex model. Table 3.7 lists each cross-linker's modification and cross-linking sites specific to RNaseS.

Table 3.7: Cross-linking and modification sites in RNaseS for each cross-linker.

| Cross-Linker | S-peptide | | S-protein | | Peptide Components | Possible Cross-linked Species |
|-----------------|----------------------|---------------------|--------------------|-----------------------------|--------------------|-------------------------------|
| | Modification Sites | Cross-Linking Sites | Modification Sites | Cross-Linking Sites | | |
| EDC | 3(1D + 2E) | 3 (2K+1Nterm) | 7(4D + 3E) | 9 (8K+1Nterm) | 356 | 125811 |
| PFA | 4 (2K +1Nterm + 1R) | 3 (1R +1Q+1Nterm) | 12(8K+1Nterm +3R) | 26(3R + 6Q+1Nterm +10N +6Y) | 911 | 828996 |
| SulfoDST | 3 (K+1Nterm) | 3 (K+1Nterm) | 9(8K+1Nterm) | 9(8K+1Nterm) | 354 | 124391 |
| BS ³ | 3 (K+1Nterm) | 3 (K+1Nterm) | 9(8K+1Nterm) | 9(8K+1Nterm) | 354 | 124391 |
| SulfoEGS | 3 (K+1Nterm) | 3 (K+1Nterm) | 9(8K+1Nterm) | 9(8K+1Nterm) | 354 | 124391 |

Masses of all possible modified and unmodified forms of each peptide were calculated to derive a list of possible unmodified/modified peptide components that could be cross-linked. This was performed for PFA, EDC, and the NHS ester cross-linkers (sulfoDST, BS³, and sulfoEGS) to generate a list of 911, 356, and 354 peptide components, respectively. In the S-peptide, there are four and three PFA modification and cross-linking sites, respectively. In the S-protein, there are 12 and 26 PFA modification and cross-linking sites, respectively. The total number of PFA reactive sites in the RNaseS system is eight more than in the calmodulin system. Although RNaseS is only slightly more complex than the calmodulin-melittin system, the number of theoretical PFA cross-linked species increased from

3.4. Moving toward the MS/MS Verification of Cross-Linked Species

191,796 to 828,996. However, with the other cross-linkers, the increase in possible cross-linked species was not as high in magnitude since they have fewer possible reactive sites. For EDC, four and three potential modification and cross-linking sites, respectively are present in the S-peptide. In the S-protein, there are seven and nine EDC modification and cross-linking sites, respectively. The total number of possible EDC cross-linked species was 125,811, which is 2127 more than in the calmodulin-melittin system. For the NHS ester cross-linkers, there are three and nine cross-linking/modification sites in the S-peptide and S-protein, respectively. The total number of possible sulfoDST, BS3 and sulfoEGS cross-linked species was 124,391, which is 99,716 more than in the calmodulin-melittin system. The total number of RNaseS candidates for EDC, PFA, sulfoDST, BS3 and sulfoEGS cross-links was 220, 607, 208, 471 and 463, respectively (listed in Appendix A.5). This represents the number of candidate cross-linked species in the RNaseS system upon which further manual inspection of MS signals subsequent MS/MS spectra are required to verify cross-linked species. Similar to the calmodulin-melittin system, PFA generated the highest number of candidates in the RNaseS system. However, since EDC reactive sites are less abundant in the RNaseS system, EDC did not produce the second highest number of candidates. Finally, among the NHS ester cross-linkers, BS³ produced the highest number of candidates in the RNaseS model system, which is consistent with the relatively high cross-linking yield observed in the SDS-PAGE.

3.4 Moving toward the MS/MS Verification of Cross-Linked Species

In this chapter, a preliminary evaluation of PFA, EDC, sulfoDST, BS³, and sulfoEGS cross-linking in Ca²⁺-free calmodulin-melittin and RNaseS was performed in order to facilitate the identification and localization of cross-linked species via MS/MS. First, the intact protein complexes produced by cross-linking were examined via SDS-PAGE to determine what type of cross-linking occurred. This supported the formation of both intra and inter-PFA, EDC, BS³, and sulfoEGS cross-linked species for both Ca²⁺-free calmodulin-melittin and RNaseS. The rel-

3.4. Moving toward the MS/MS Verification of Cross-Linked Species

ative cross-linking yields for PFA, EDC and BS³ species were comparable to the theoretical yields of each Ca²⁺-free calmodulin-melittin and RNaseS complex derived from their respective dissociation constants, suggesting that these cross-linkers captured relevant non-covalent interactions. For intact sulfoDST cross-linked species, only intra-cross-links were observed in the SDS-PAGE for both protein model systems. In addition, sulfoDST produced a lower relative yield than the other cross-linkers. Second, PFA-modified and unmodified peptide analysis confirmed the MS-detection of peptides from cross-linked species for Ca²⁺-free calmodulin-melittin and RNaseS, although less peptides were identified in the RNaseS model system. PFA-modified peptides revealed modification sites to aid in cross-link localization. PFA modifications were observed primarily on N-terminal and K, and to lesser extent, R residues, which confirmed the specificity of the modification reaction under the *in vivo*-like, mild reaction conditions utilized, similar to PFA cellular cross-linking. Furthermore, the observation of only internal PFA modifications on peptides demonstrated that trypsin cleavage after PFA modifications (Schiff Base or methylol) is unlikely. However, the mechanism of trypsin digestion suggests that trypsin can potentially cleave after cross-linked residues produced by PFA unlike those produced with established cross-linkers. Therefore, PFA cross-linking at terminal residues is considered a possibility. Finally, a list of MS-confirmed candidate cross-linked species was obtained. Consistent with the relative yield measured in the SDS-PAGE, PFA generated the most number of MS-confirmed candidate cross-linked candidates in both protein model systems out of all the cross-linkers due to its ability to react with several amino acids and potential to produce multiple modifications and cross-links. EDC reactive sites are abundant in the Ca²⁺-free calmodulin-melittin, and therefore produced the second highest number of MS-confirmed candidate cross-linked species. The MS-confirmed candidate cross-linked species produced by NHS esters cross-linkers of different lengths suggested that cross-linker bridge length is proportional to the number of candidates, due to its increased probability of being within the distance of two reactive sites. However, MS/MS verification of these candidates is required to confirm cross-linked structures and localize cross-linking sites, which shall be explored in the following chapter.

Chapter 4

Tandem Mass Spectrometric Fragmentation of Calmodulin-Melittin Cross-linked Species

4.1 Tandem Mass Spectrometric Verification and Fragmentation Rules for Cross-linked Species

In chapter 3, a total number of 160, 335, 62, 77, and 158 calmodulin-melittin cross-linked candidates for EDC, PFA, sulfoDST, BS³ and sulfoEGS, respectively were identified at the MS level. The MS/MS spectra for each calmodulin-melittin MS candidate were analyzed to verify cross-linking. Theoretical fragment ions for the backbone fragmentation of unmodified peptides from calmodulin and melittin were obtained using protein prospector MS product[117]. These were used to manually prepare theoretical cross-linked and/or modified peptide fragment ion databases specific to each cross-linker chemistry. In each peptide, for each fragment that contained a cross-linking or modification site, modified fragment masses were obtained by adding the mass of each cross-linker-specific modification (i.e. +12 or +30, +155.00, +132.01, +156.08, and +244.06 Da, for PFA, EDC, sulfoDST, BS³, and sulfoEGS, respectively).

The MS/MS spectrum of each MS cross-linked candidate was inspected manually for the three types of fragment ions that can be generated when a cross-linked species undergoes CID (illustrated in Figure 1.10): whole peptide component ions (type 1), peptide backbone ions with the cross-linker intact (type 2) and peptide

4.1. Tandem Mass Spectrometric Verification and Fragmentation Rules for Cross-linked Species

backbone ions with cross-link bridge fragmented (type 3). In addition, residue reactivity (Figures 1.3,1.4,1.5,1.6, and1.7) was also considered in evaluating cross-linked species. Cross-linked peptides are represented in the text as (peptide I) ^ (peptide II), where “^” symbolizes the cross-linker bridge. Figure 4.1 illustrates the nomenclature used to annotate the fragment ion peaks in the MS/MS spectra of each cross-linked species where “ABCDE” represents a generic sequence of each peptide component. The peaks labelled as “[M]” in these spectra correspond to the precursor ion of the intact cross-linked species.

4.1. Tandem Mass Spectrometric Verification and Fragmentation Rules for Cross-linked Species

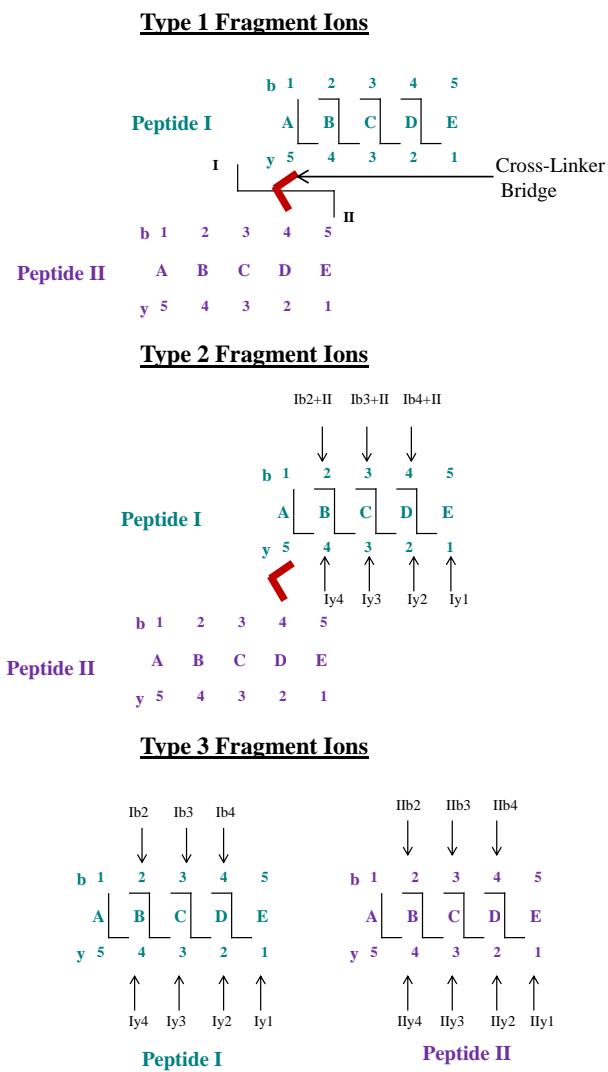


Figure 4.1: The nomenclature used to annotate MS/MS spectra of cross-linked species for each type of fragment ion

EDC, sulfoDST, BS³ and sulfoEGS are classified as “other cross-linkers” and

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

their established fragmentation patterns are compared to PFA cross-links to clarify acceptable fragment ion evidence to confirm PFA cross-linked species. Cross-links formed by other cross-linkers were compared to literature however these studies were based on Ca^{2+} -bound calmodulin complexed with melittin unlike the Ca^{2+} -free calmodulin-melittin complex used in this study. According to NMR and fluorescence experiments in literature, both complexes exhibit similar conformations and thus cross-linking sites should be comparable [95]. Nevertheless, it remains to be seen whether this is supported by the MS/MS verified cross-linking identified in this current study. It is hypothesized that Ca^{2+} binding that occurs with negatively charged carboxyl groups will create variations in cross-linking sites, especially with the carboxyl group reactive EDC cross-linking.

Cross-linking between two calmodulin peptides and between calmodulin and melittin peptides was observed for all cross-linkers except the sulfoDST cross-linker, for which only cross-linking between calmodulin and melittin was identified. Previous studies have shown that calmodulin could form two different structures (antiparallel or parallel binding modes) with melittin, supporting that cross-linked species signals could be representing multiple conformations [99, 100]. However, in both modes of binding, the conformation of calmodulin itself should be uniform, and consequently it is expected that cross-linking within calmodulin should be consistent in major cross-linked products. This was considered while deducing cross-linking sites via MS/MS.

4.2 Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

4.2.1 EDC

4.2.1.1 Calmodulin-Calmodulin Cross-linked Peptides

The MS/MS fragmentation patterns of EDC cross-linked species were examined. Three calmodulin interpeptide cross-links were discovered, which are listed in Table 4.1 that conveys the m/z , charge, monoisotopic mass (experimental and calculated), mass accuracy (ppm), the mass and sequence of each component peptide

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

(highlighted residues correspond to cross-linked sites).

The cross-linked structure $^{91}\text{VFDKDGNGYISAAELR}^{106}\wedge^1\text{A}^{(ac)}\text{DQLTEEQIAEFK}^{13}$ appeared as a triply charged species at m/z 1100.54 and with a charge of four at m/z 825.66. with two different charge states (Figure 4.2). Type 1 ions of each whole peptide component confirmed the cross-linked structure composition. A series of Ily1 to Ily7 and Ib3 to Ib5 ions for $^1\text{A}^{(ac)}\text{DQLTEEQIAEFK}^{13}$, and Iy1 to Iy3, Iy5 to Iy7, and Iy10 to Iy12 ions for $^{91}\text{VFDKDGNGYISAAELR}^{106}$ matched the unmodified sequences for each peptide. A series of type 2 ions Ily9-I to Ily12-I, with $^{91}\text{VFDKDGNGYISAAELR}^{106}$ attached to $^2\text{DQLTEEQIAEFK}^{13}$, $^3\text{QLTEEQIAEFK}^{13}$, $^4\text{LTEEQIAEFK}^{13}$ and $^5\text{TEEQIAEFK}^{13}$ along with the unmodified Ily1 to Ily7 ions localized the cross-link to $^5\text{TE}^6$. In this segment, E6 is the only residue with a carboxylic group for cross-linking. Since EDC forms cross-links between carboxylic groups and primary amino groups, the cross-link formed between E6 and K94, the only lysine residue in $^{91}\text{VFDKDGNGYISAAELR}^{106}$. The N-terminal residue A1 is acetylated and thus was excluded as a potential cross-linking site.

The species at m/z 716.96, with a charge of five, corresponded to $^{14}\text{EAFSLFDKDGDTITTK}^{30}\wedge^{91}\text{VFDKDGNGYISAAELR}^{106}$ (Figure 4.3). Ily1 to Ily7, Iib2 and Iib3 ions confirmed the part of the unmodified sequence of $^{91}\text{VFDKDGNGYISAAELR}^{106}$. Iy1 to Iy16 ions along with a type 1 ion confirmed the unmodified sequence of $^{14}\text{EAFSLFDKDGDTITTK}^{30}$. Type 2 ions Ib3+II, Ib6+II to Ib8+II suggested that E14 was modified and cross-linked to K94, the only lysine residue in $^{91}\text{VFDKDGNGYISAAELR}^{106}$. Similar to this structure, $^{91}\text{VFDKDGNGYISAAELR}^{106}\wedge^{22}\text{DGDGTITTK}^{30}$ was detected as a triply charged species at m/z 875.76 (4.4). Iy1 to Iy9, Ib2 and Ib3 ions confirmed the unmodified sequence of $^{91}\text{VFDKDGNGYISAAELR}^{106}$. Type 2 ions Ib4-II to Ib10-II localized the cross-link to $^{94}\text{KDGNGYI}^{101}$. The $^{22}\text{DGDGTITTK}^{30}$ peptide was identified exclusively by its mass since its respective type 1 or backbone fragment ions did not appear in the MS/MS spectra. The mass of the cross-linked species and the type 2 ions Ib4-II to Ib10-II suggests that two cross links formed, each resulting in a 18.02 Da decrease. This indicates that K30 and either D22 or D24 should have formed an intrapeptide cross-link. With K30 occupied in the intrapeptide cross-link, $^{22}\text{DGDGTITTK}^{30}$ must have been modified first at either

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

D22 or D24 to form the interpeptide cross-link to K94, the only lysine residue in $^{94}\text{KDGNGYI}^{101}$. It was noted in section 3.3.1.3 that trypsin is unlikely to cleave after EDC cross-links, which is contradicted by the formation of the intrapeptide cross-link with terminal K30. Without peptide backbone fragment ions from $^{22}\text{DGDGTITTK}^{30}$, it is not possible to determine whether the extra -18.02 Da shift corresponds to the proposed intrapeptide cross-link or the loss of H_2O from an amino acid side chain such as threonine. It is hypothesized that the intrapeptide bond indeed formed and sterically hindered the peptide backbone fragmentation of $^{22}\text{DGDGTITTK}^{30}$. Nevertheless, both species at m/z 716.96 and 875.76 with a charge of five and three, respectively, supported cross-linking between similar regions of calmodulin. However, the different cross-linking sites indicate that they represent two different reaction products. The MS/MS identification of EDC cross-link formation between calmodulin peptides $^{14}\text{EAFSLFDKDGDTITTK}^{30}$ and $^{91}\text{VFDKDGNGYISAAELR}^{106}$ was discovered for the first time. Furthermore, this is the first report of MS/MS verified EDC cross-linking between E6 and K94.

Table 4.1: EDC Calmodulin-Calmodulin interpeptide cross-linked species, in which cross-linking sites are highlighted in red. For species appearing with two different charge states, annotated MS/MS spectra is shown for the m/z marked with an “*”.

| Cross-Linked Species | | | | | Calmodulin Peptide 1 | | Calmodulin Peptide 2 | |
|----------------------|-----|-----------------------------------|------------------------------------|------------------------|----------------------|--|----------------------|--------------------------------------|
| m/z | z | $[\text{M}]_{\text{exp}}$ (Da) | $[\text{M}]_{\text{calc}}$ (Da) | Mass Accuracy (ppm) | $[\text{M}]$ (Da) | Sequence | $[\text{M}]$ (Da) | Sequence |
| 875.76 | 3 | 2624.28 | 2624.27 | 3 | 906.43 | $^{22}\text{DGDGTITTK}^{30}$ | 1753.86 | $^{91}\text{VFDKDGNGYISAAELR}^{106}$ |
| 825.66 | 4 | 3298.62 | 3298.60 | 6 | 1562.75 | $^1\text{A}^{(\text{ac})}\text{DQLTEEQIAEFK}^{13}$ | 1753.86 | $^{91}\text{VFDKDGNGYISAAELR}^{106}$ |
| *1100.54 | 3 | | | | | | | |
| 716.96 | 5 | 3579.8 | 3579.74 | 18 | 1753.86 | $^{91}\text{VFDKDGNGYISAAELR}^{106}$ | 1843.88 | $^{14}\text{EAFSLFDKDGDTITTK}^{30}$ |

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

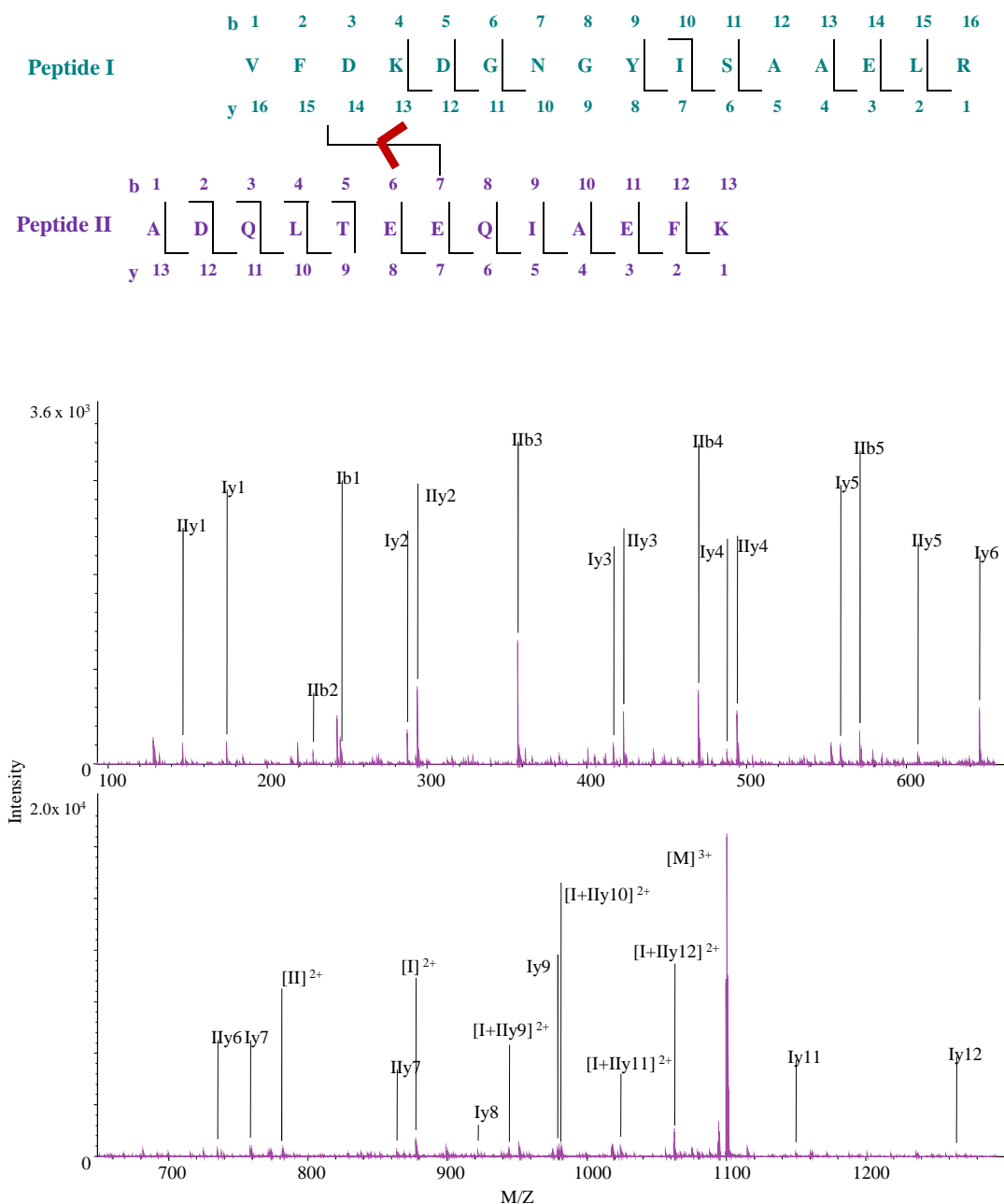


Figure 4.2: Interpeptide EDC calmodulin cross-link at m/z 1100.54 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red; 102

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

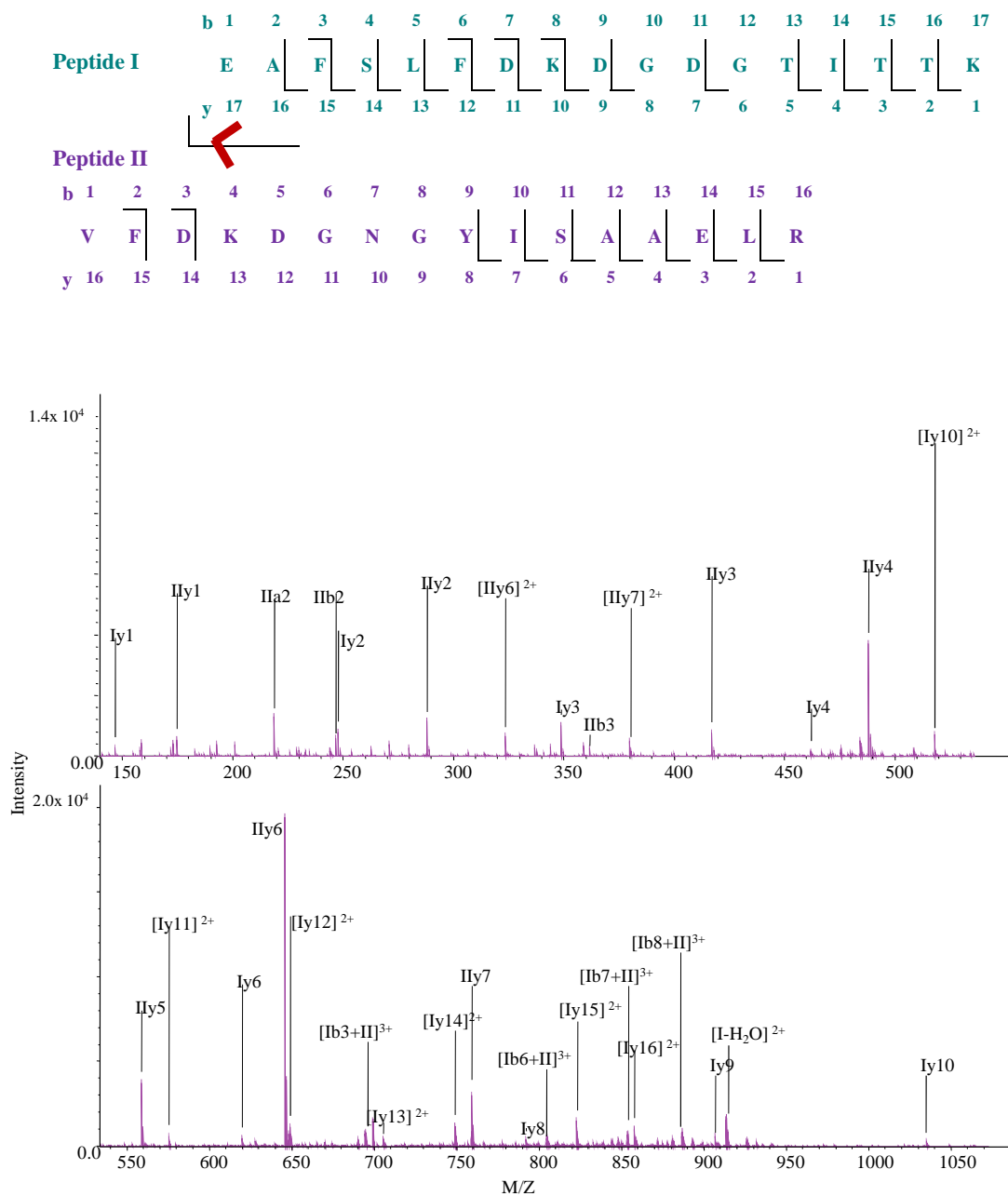


Figure 4.3: Interpeptide EDC calmodulin cross-link at m/z 716.96 ($z = 5$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom) ; Cross-linker bridges are indicated in red. 103

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

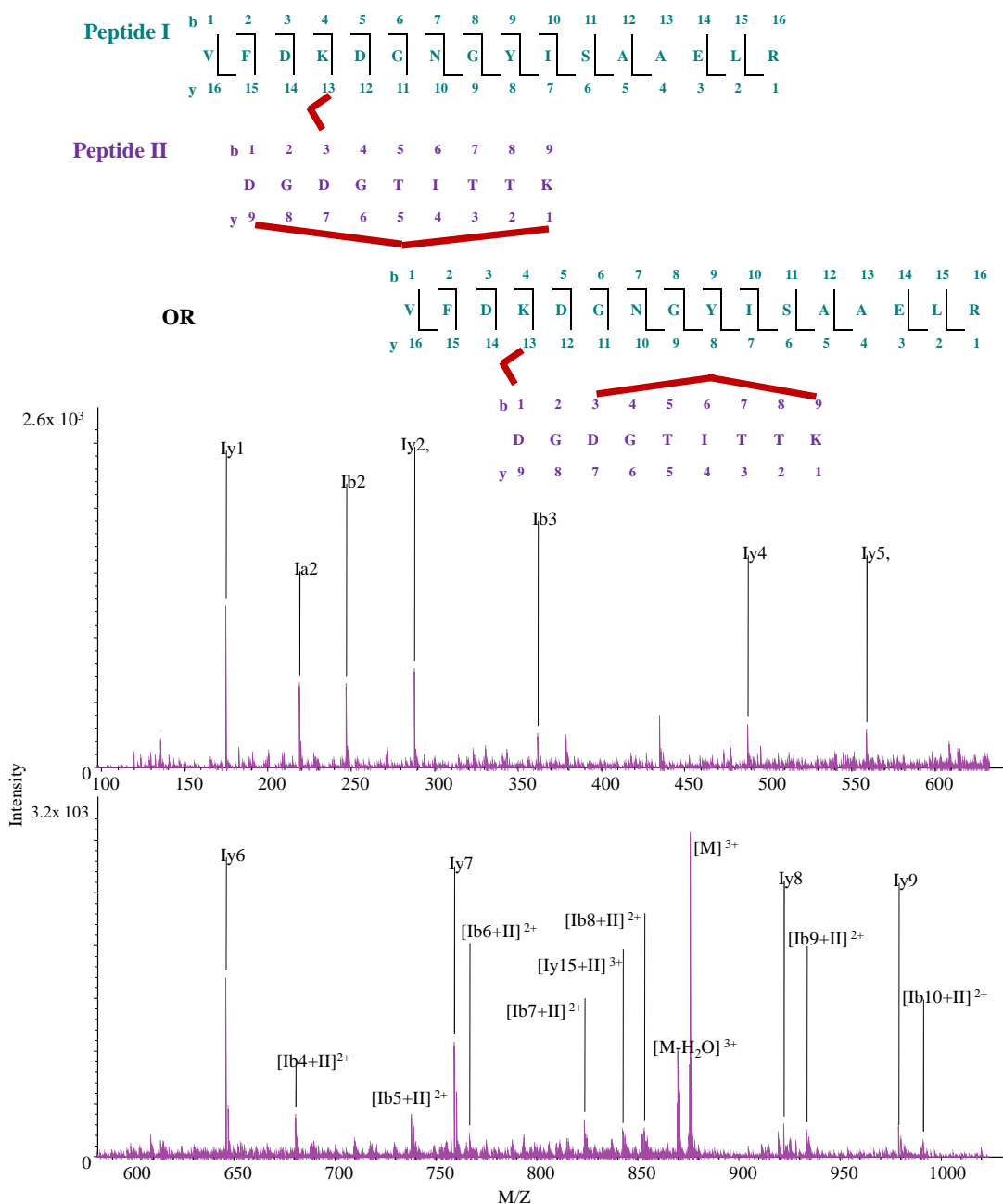


Figure 4.4: Interpeptide EDC calmodulin cross-link at m/z 875.76 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

4.2.1.2 Calmodulin-Melittin Cross-linked Peptides

Table 4.2 lists the identified EDC cross-linked species composed of calmodulin and melittin peptides. MS/MS spectra of all of these cross-linked peptides produced type 1 ions for the melittin peptide, confirming its presence. For $^{14}\text{EAFSLFDKDGDTITTK}^{30}\wedge^1\text{GIGAVLK}^7$, appearing at m/z 621.58 with a charge of four (Figure 4.5), type 2 ions Ib2+II to Ib4+II localized the cross-link to E14. Therefore the calmodulin peptide was modified and cross-linked to a primary amino group (G1 or K7) in the melittin peptide. An almost complete unmodified y ion series for $^1\text{GIGAVLK}^7$ and $^{14}\text{EAFSLFDKDGDTITTK}^{30}$ confirmed the sequence of each component peptide and the absence of unmodified b ions suggested that G1 was cross-linked. Since K7 is a terminal lysine residue, it is more likely that the N-terminus of G1 formed the cross-link.

The triply charged species at m/z 616.66 corresponded to the structure $^1\text{A}^{(ac)}\text{DQLTEEQIAEFK}^{13}\wedge^{23}\text{KR}^{24}$ (Figure 4.6). Unmodified Ib2 to Ib5 and Iy1 to Iy2 ions verified the sequence of the calmodulin peptide and type 2 ions Iy3+II to Iy12+II ions localized the cross-link to E11. A type 1 ion and Iy2 ion verified the presence of the melittin peptide. Therefore, calmodulin E11 was cross-linked to K23, the only reactive residue in the melittin peptide.

The cross-linked species³⁸ $\text{SLGQNPTEAELQDMINEVDADGNGTIDFPEFLT M}^{(ox)}\text{M}^{(ox)}\text{AR}^{74}\wedge^{23}\text{KR}^{24}$ appeared with a signal at m/z 1097.26, with a charge of four, and also with an additional oxidized M residue (+15.99 Da mass shift) producing signals at m/z 881.21, with a charge of five and m/z 1101.26, with a charge of four (Figure 4.7). Unmodified Iy1 to Iy12 and Ib2 to Ib5 ions verified the sequence of the calmodulin peptide and a series of type 2 ions (Ib18+II to Ib32+II, Iy21+II to Iy24+II and Iy27+II) localized the cross-linking site to E54. A type 1 ion corresponding to the melittin peptide confirmed its presence. Thus, calmodulin E54 was cross-linked to melittin K23, the only EDC reactive residue in the melittin peptide. Similarly, for the unoxidized form of this cross-linked species ($m/z = 1097.26$, with a charge of four), unmodified fragment ions Ib2 to Ib5 and Iy1 to Iy11 confirmed the sequence of the calmodulin peptide and the unmodified type 1 ion verified the presence of the melittin peptide. Type 2 ions (Ib23+II to Ib25+II, Ib27+II, Ib28+II, Iy22+II, and Iy23+II) supported cross-linking at calmodulin E54

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

and thus provided evidence for cross-linking between melittin K23 and calmodulin E54.

The calmodulin-melittin EDC cross-linking sites E14 to G1 and E11 to K23 are consistent with previous studies that identified EDC cross-linking using a combination of MS and Edman degradation [102]. The EDC cross-linking was identified in Ca²⁺-saturated calmodulin-melittin in previous literature. Ca²⁺ binding occurs with negatively charged carboxyl groups that are also EDC reactive. However, the consistency of cross-linking sites between Ca²⁺-free (used in this study) and Ca²⁺-saturated calmodulin-melittin proposes that Ca²⁺ binding does not significantly affect EDC cross-linking. Overall, this present study allowed for further validation and localization of specific cross-linking sites with the use of MS/MS for the first time. This is also the first report of EDC cross-linking between calmodulin E54 and melittin K23.

Table 4.2: EDC calmodulin-melittin interpeptide cross-linked species are listed and classified as capturing antiparallel (shaded in blue) or parallel (white) binding. Reactive residues/possible cross-linking sites are highlighted in red. For species appearing with two different charge states, annotated MS/MS spectra is shown for the m/z marked with an “*”.

| Cross-Linked Species | | | | | Melittin Peptide | | Calmodulin Peptide | | |
|----------------------|--------|----------------------------|-----------------------------|------------------------|------------------|-----------------------------------|--------------------|--|--|
| m/z | z | [M] _{exp} (Da) | [M] _{Calc} (Da) | Mass Accuracy (ppm) | [M] (Da) | Sequence | [M] (Da) | Sequence | |
| 621.58 | 4 | 2482.33 | 2482.30 | 15 | 656.42 | ¹ GIGAVLK ⁷ | 1562.75 | ¹⁴ EAFSLFDKDGDTITTK ³⁰ | |
| 616.66 | 3 | 1846.97 | 1846.94 | 13 | 302.21 | ²³ KR ²⁴ | 1562.75 | ¹ A ^(ac) DQLTEEQIAEFK ¹³ | |
| 1097.26 | 4 | 4385.04 | 4385.04 | 1 | 302.21 | ²³ KR ²⁴ | 4100.84 | ³⁸ SLGQNPTEAELQDMINEVDADGNGTIDFPEFLTM ^(ox) M ^(ox) AR ⁷⁴ | |
| *1101.26 881.21 | 4 5 | 4401.05 4401.05 | 4401.05 | 1 | 302.21 | ²³ KR ²⁴ | 4116.85 | ³⁸ SLGQNPTEAELQDM ^(ox) INEVDADGNGTIDFPEFLTM ^(ox) M ^(ox) AR ⁷⁴ | |

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

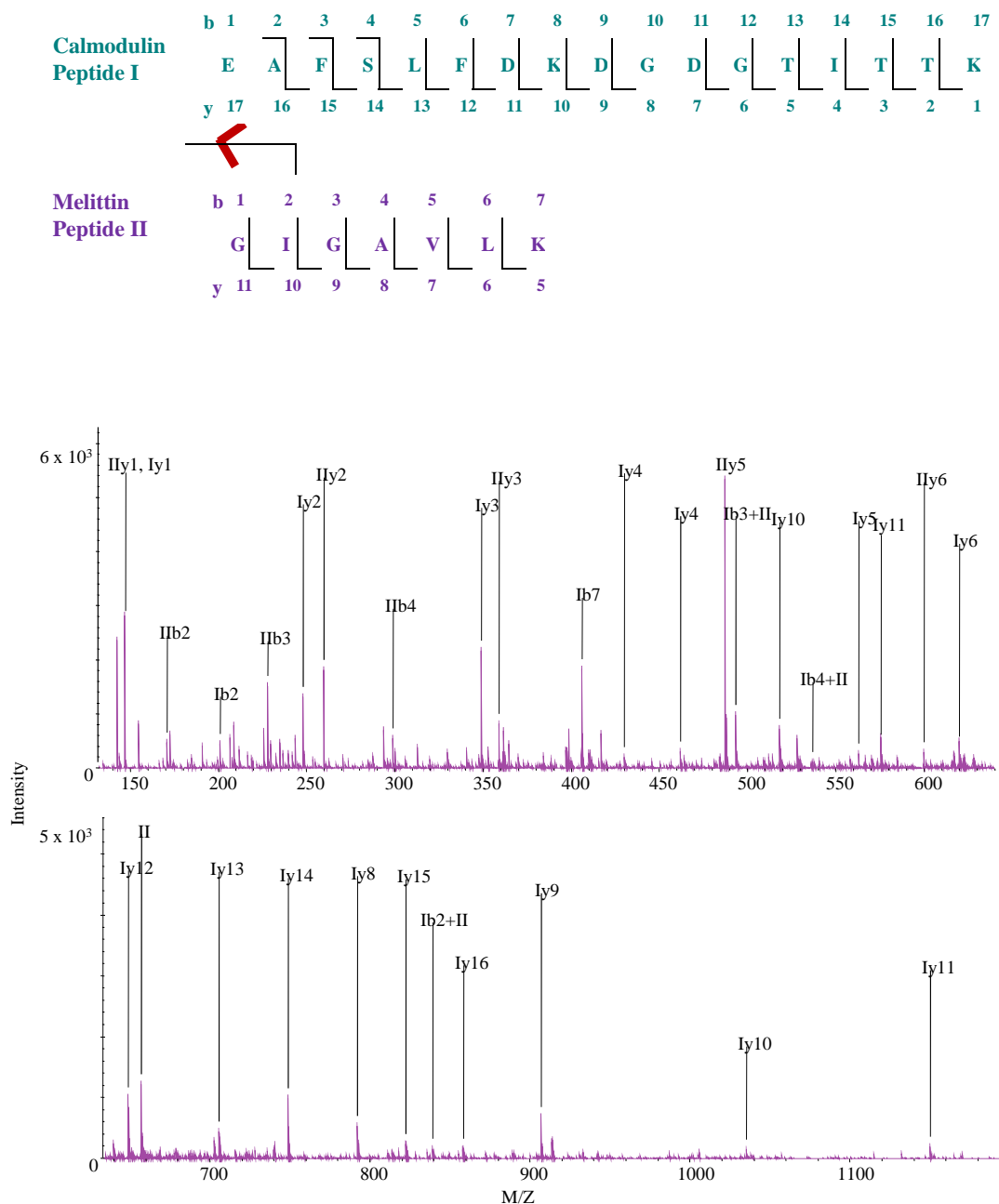


Figure 4.5: Interpeptide EDC calmodulin-melittin cross-link m/z 621.58 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom) ; Cross-linker bridges are indicated in red. 107

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

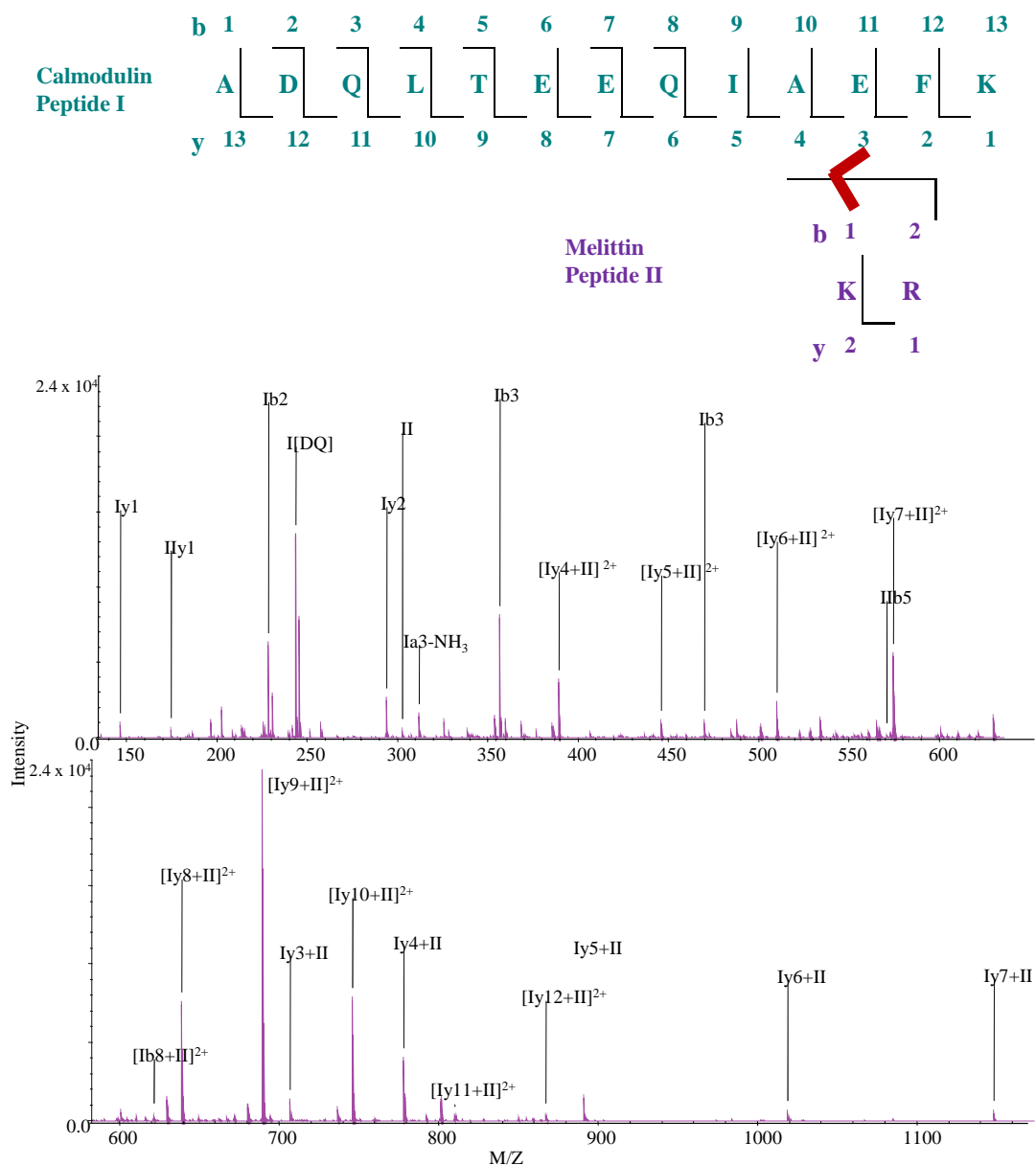


Figure 4.6: Interpeptide EDC calmodulin-melittin cross-link m/z 616.66 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom) are shown

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

Calmodulin Peptide I

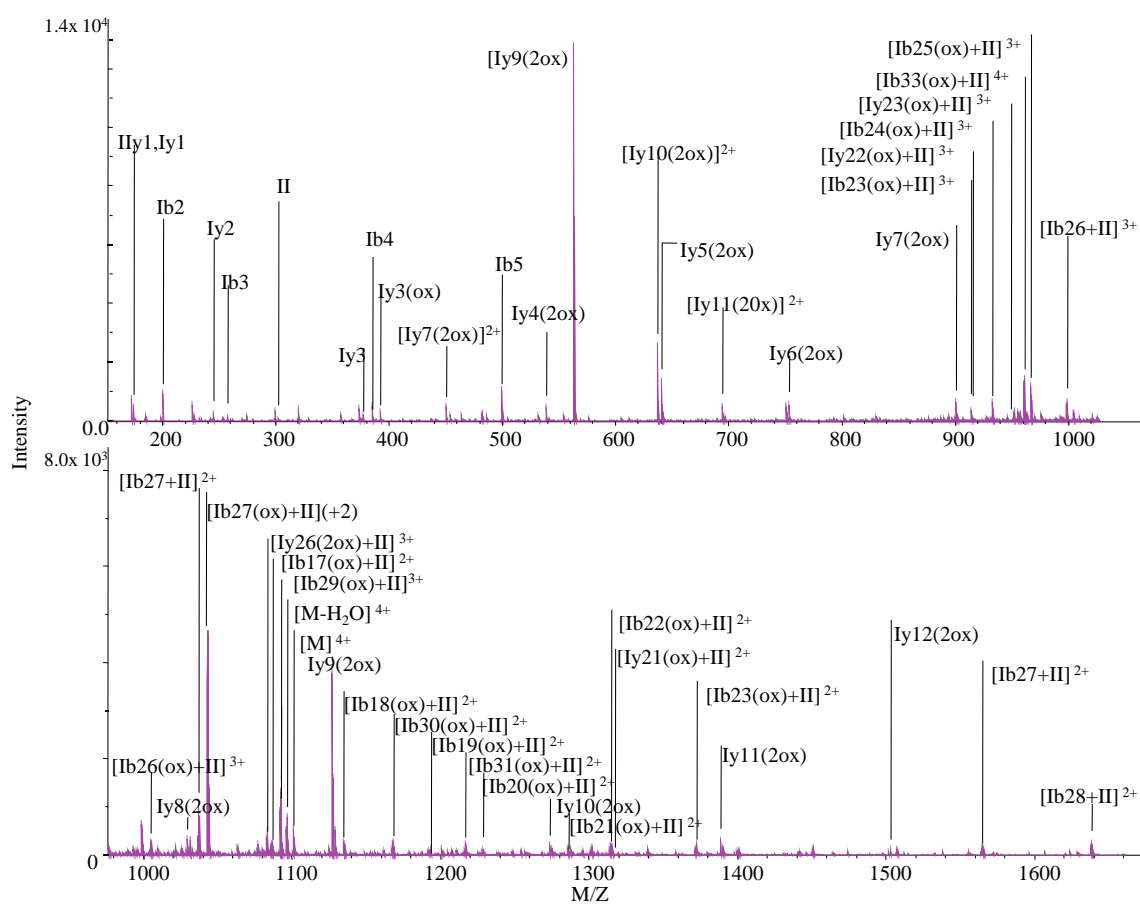
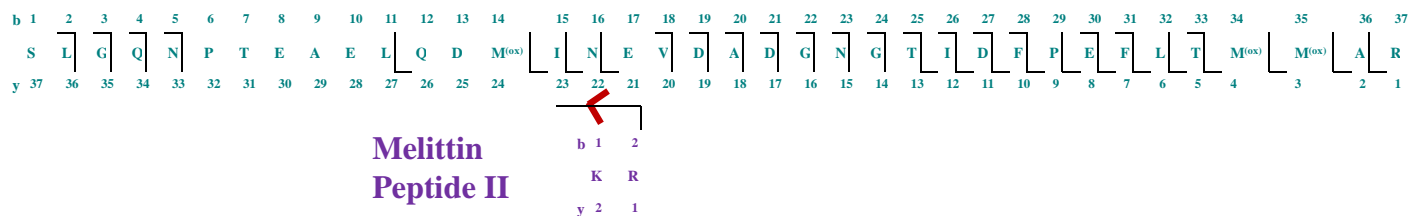


Figure 4.7: Interpeptide EDC calmodulin-melittin cross-link m/z 1101.26 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

No simultaneous fragmentation at the bridge and backbone i.e. type 3 fragmentation was observed in EDC cross-linked species. A possible reason is that since the EDC cross-linked bridge is a peptide bond, it has an equal bond energy to the backbone peptide bonds and would thus require the same collision energy to fragment. In addition, fragmentation of the backbone of the smaller component peptide was not as extensive as with the larger component peptide, which is common phenomenon [51].

4.2.2 SulfoDST

4.2.2.1 Calmodulin-Melittin Cross-linked Peptides

Cross-linking with sulfoDST involving only calmodulin was not detected in this study. Two calmodulin-melittin sulfoDST cross-linked species were identified and listed in Table 4.3, which lists the m/z , charge, monoisotopic mass (experimental and calculated), mass accuracy (ppm), the mass and sequence of each component peptide (highlighted residues correspond to cross-linked sites). Cross-linked species $^{91}\text{VFDKD}^{(dm)}\text{NGYISAAELR}^{106\wedge 1}\text{GIGAVLK}^7$ appeared as a triply charged species at m/z 842.77 (Figure 4.8). A full y ion series of $^1\text{GIGAVLK}^7$ (Ily1 to Ily6) and of $^{91}\text{VFDKDG}^{(dm)}\text{NYISAAELR}^{106}$ (Iy1 to Iy12 and Ib2 to Ib3) confirmed the sequences of the cross-linked peptides. The deamidation of N97 was marked by the + 0.98 Da mass shift on Iy10 to Iy12. This deamidated N97 was also previously observed in other calmodulin studies [161] and also in control samples (see Table 3.1). Type 2 ions Ib4+II to Ib6+II, Iy13+II to Iy15+II and Iib6+I localized the cross-linking between G1 and K94.

The cross-linked candidate at m/z 493.92, with a charge of three, corresponded to the following structure $^{22}\text{DGDGTITTK}^{30\wedge 22}\text{RKR}^{24}$ (Figure 4.9). A type 1 ion, type 3 ions Iy1 to Iy4 and a Ib2 ion confirmed the sequence of $^{22}\text{DGDGTITTK}^{30}$. Only a Iib2 ion appeared from $^{22}\text{RKR}^{24}$, however poor fragmentation of the smaller peptide in a cross-linked species is typical [51]. Although no type 2 ions were present, with only one reactive residue in each peptide, cross-linking must have occurred between K30 from calmodulin and K23 from melittin. The unmodified Iy1 to Iy4 ions suggests that the cross-linker bridge and peptide backbone were simultaneously fragmented. SulfoDST cross-linking between calmod-

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

ulin K30 and melittin K23 is consistent with previous MS-based identification of sulfoDST cross-linking between calmodulin segment 1- 37 to melittin K23 [99]. However, the sulfoDST cross-link formation at K30 contradicts previous claims that trypsin cleavage at cross-linked K is unlikely, as discussed in section 3.3.1.3.

Table 4.3: sulfoDST calmodulin-melittin interpeptide cross-linked species are listed and classified as capturing antiparallel (shaded in blue) or parallel (white) binding. Cross-linking sites are highlighted in red.

| Cross-Linked Species | | | | | Melittin Peptide | | Calmoulin Peptide | | |
|----------------------|----------|----------------------------|-----------------------------|---------------------------|------------------|-----------------------------------|-------------------|---|--|
| <i>m/z</i> | <i>z</i> | [M] _{exp} (Da) | [M] _{Calc} (Da) | Mass Accuracy (ppm) | [M] (Da) | Sequence | [M] (Da) | Sequence | |
| 842.77 | 3 | 2525.30 | 2525.28 | 7 | 656.42 | ¹ GIGAVLK ⁷ | 1754.86 | ⁹¹ VFDK ^{dm} DGN ^(dm) GYISAAELR ¹⁰⁶ | |
| 493.92 | 3 | 1478.75 | 1478.73 | 12 | 458.31 | ²² RKR ²⁴ | 906.43 | ²² DGDGTITTK ³⁰ | |

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

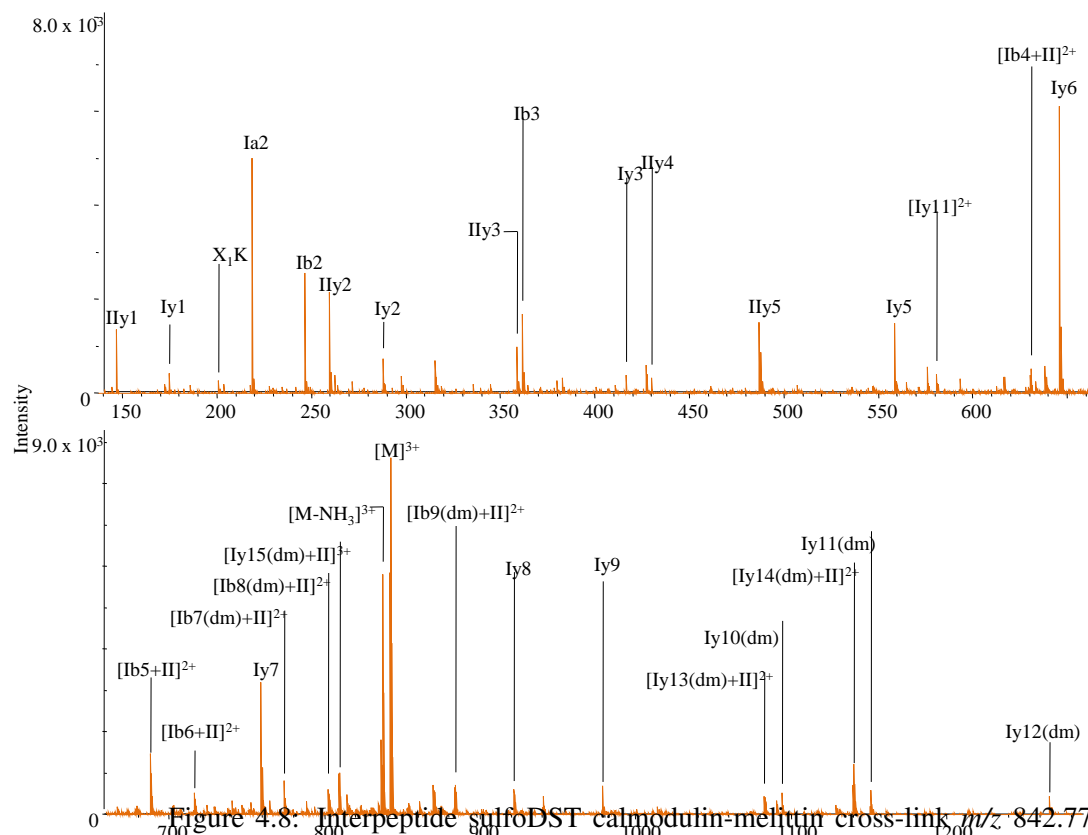
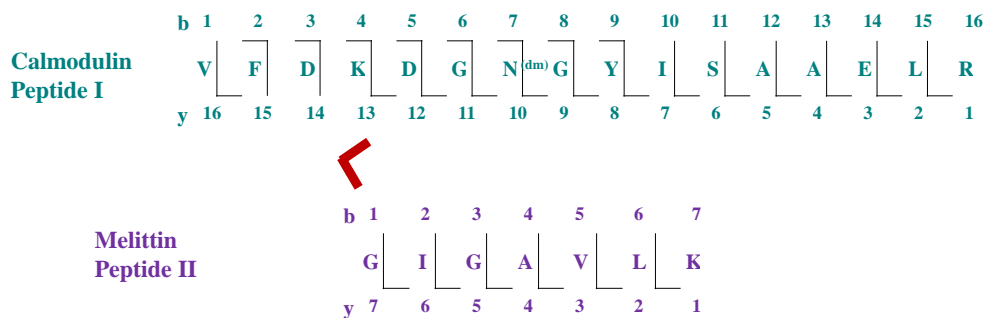
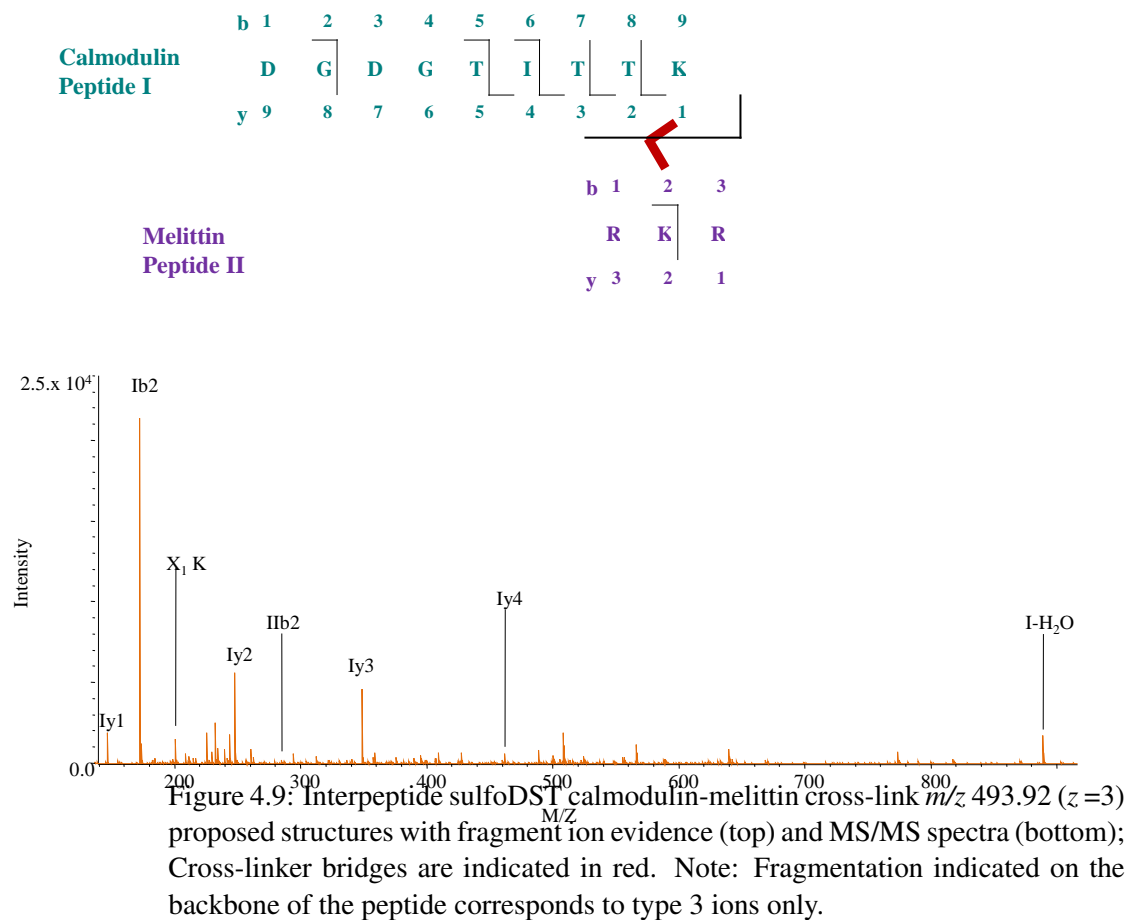


Figure 4.8: Interpeptide sulfoDST calmodulin-melittin cross-link m/z 842.77 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers



A diagnostic ion X_1K ($m/z = 200$) appeared in the MS/MS spectrum of $^{91}\text{VFDKDGNGYISAAELR}^{106}\wedge^{1}\text{GIGAVLK}^7$, which verified the presence of the cross-linker and indicated that fragmentation occurred at the bond connecting the cross-linker bridge to melittin G1 and at the peptide backbone bond of calmodulin K94. In addition, the production of type 3 ions from $^{22}\text{DGDGTITTK}^{30}\wedge^{22}\text{RKR}^{24}$ signifies simultaneous fragmentation of the bridge and peptide backbone.

4.2.3 BS³

4.2.3.1 Calmodulin-Calmodulin Cross-linked Peptides

Three BS³ interpeptide calmodulin cross-linked species were discovered and are listed in Table 4.4, which lists the m/z , charge, monoisotopic mass (experimental and calculated), mass accuracy (ppm), the mass and sequence of each component peptide (highlighted residues correspond to cross-linked sites). The BS³ cross-linked structure ⁹¹VFDKDGNGYISAAELR¹⁰⁶⁷⁶MK⁷⁷ appeared as a doubly charged species at m/z 1085.55 (Figure 4.10) and as a triply charged species at m/z 724.04 i.e. with two different charge states. A full unmodified y ion series (Iy1 to Iy12) and Ib2 to Ib3 ions for ⁹¹VFDKDGNGYISAAELR¹⁰⁶ confirmed its sequence. Type 2 ions (Ib4+II, Ib5+II and Iy13+II to Iy15+II) verified that cross-linking occurred between K77 and K94. The ⁷⁶MK⁷⁷ was confirmed by the appearance of its type 1 ion.

Candidate species appearing as a triply charged species at m/z 1079.51 matched the mass of a BS³ cross-link between two identical peptides: ⁷⁵KMKDTDSEEEIR⁸⁶⁷⁵KMKDTDSEEEIR⁸⁶ with two cross-linked bridges (Figure 4.11). Unmodified type 3 ions (Iy1 to Iy4, Iy6, Iy8, Iy9) confirmed the sequence of the peptide components. Although, type 2 ions were absent, type 3 b ions attached to the BS³ bridge (Ib3+X₁ to Ib7+X₁) localized one cross-linking site to K77. With only two K residues in each peptide, either two K77 to K75 interpeptide cross-links or one K77 to K77 and one K75 to K75 interpeptide cross-link formed.

The triply charged species at m/z 588.96 and doubly charged species at m/z 882.94 corresponded to K cross-linked ⁷⁵KMKDTDSEEEIR⁸⁶ (Figure 4.12). It was not possible to determine whether the K was either K75 or K77 from calmodulin, or K23 from melittin. Both ⁹¹VFDKDGNGYISAAELR¹⁰⁶⁷⁶MK⁷⁷ and ⁷⁵KMKDTDSEEEIR⁸⁶^K suggest that trypsin cleaved a cross-linked lysine, which is not expected, as explained in section 3.3.1.3.

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

Table 4.4: BS³⁺ calmodulin interpeptide cross-linked species, in which cross-linking sites are highlighted in red. For species appearing with two different charge

| Cross-Linked Species | | | | | Calmodulin Peptide 1 | | Calmodulin Peptide 2 | |
|----------------------|----------|----------------------------|-----------------------------|---------------------------|----------------------|---|----------------------|---|
| <i>m/z</i> | <i>z</i> | [M] _{exp} (Da) | [M] _{calc} (Da) | Mass Accuracy (ppm) | [M] (Da) | Sequence | [M] (Da) | Sequence |
| *1085.55 724.04 | 2 3 | 2169.10 | 2169.08 | 9 | 277.15 | ⁷⁶ M ^K ⁷⁷ | 1753.86 | ⁹¹ VFD ^K DGNGYISAAELR ¹⁰⁶ |
| 1079.51 | 3 | 3235.54 | 3235.51 | 10 | 1479.69 | ⁷⁵ K ^M ^K DTDSEEEIR ⁸⁶ | 1479.69 | ⁷⁵ K ^M ^K DTDSEEEIR ⁸⁶ |
| *588.96 882.94 | 3 2 | 1763.88 | 1763.86 | 10 | 146.11 | ⁷⁵ K or ²³ K (melittin) | 1479.69 | ⁷⁵ K ^M ^K DTDSEEEIR ⁸⁶ |

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

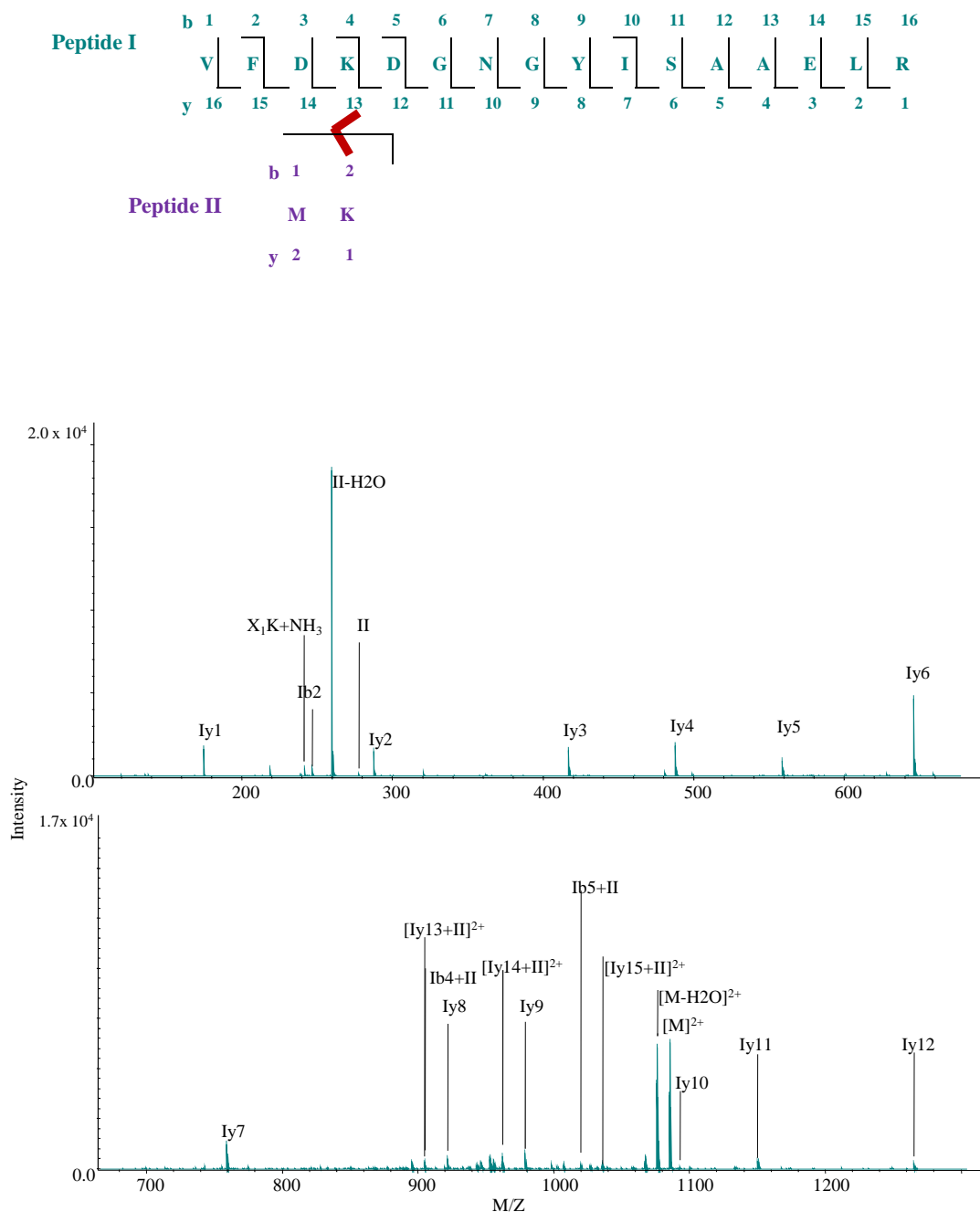


Figure 4.10: Interpeptide BS³⁺ calmodulin cross-link m/z 1085.55 ($z = 2$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. 116

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

Type 1

| | | | | | | | | | | | | | |
|------------|---|--------------|----|--------------|---|---|---|---|---|---|----|----|----|
| | b | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Peptide I | | K | M | K | D | T | D | S | E | E | E | I | R |
| | y | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| | | [Red bridge] | | [Red bridge] | | | | | | | | | |
| | b | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Peptide II | | K | M | K | D | T | D | S | E | E | E | I | R |
| | y | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

Type 3

| | | | | | | | | | | | | | |
|--|---|----|----|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | b | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| | | K | M | K | D | T | D | S | E | E | E | I | R |
| | y | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| | | | | [Red bridge] | [Red bridge] | [Red bridge] | [Red bridge] | [Red bridge] | [Red bridge] | [Red bridge] | [Red bridge] | [Red bridge] | [Red bridge] |

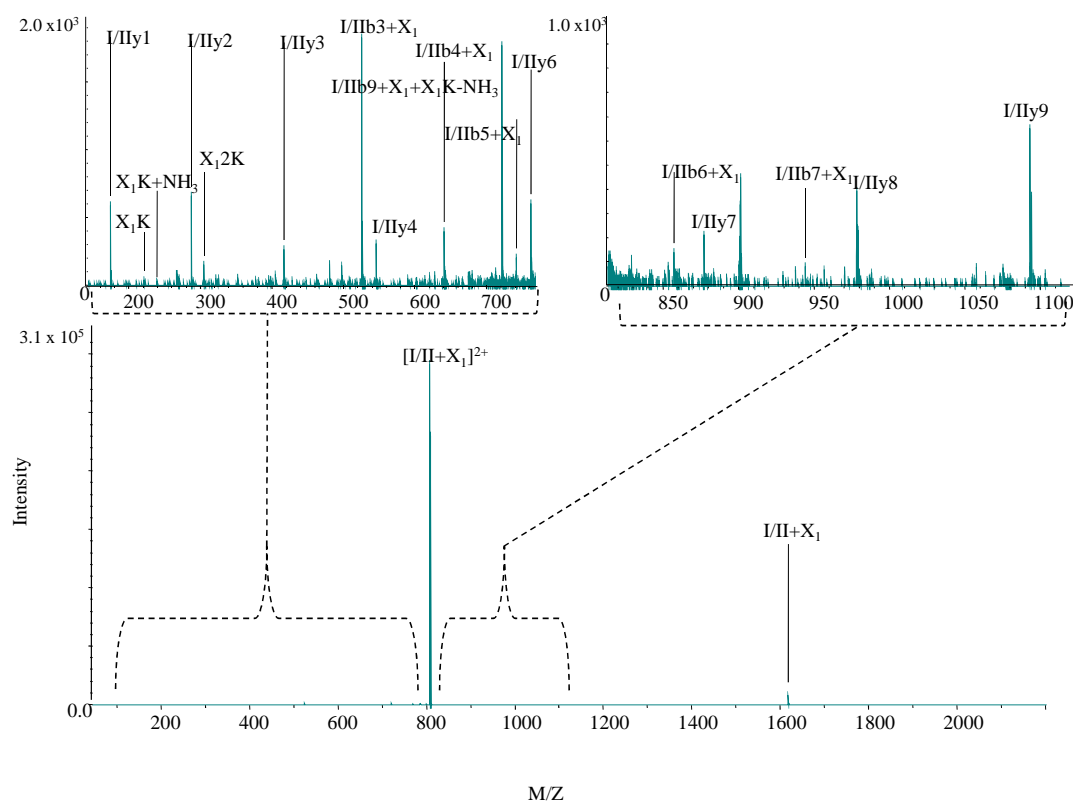


Figure 4.11: Interpeptide BS³⁺ calmodulin cross-link m/z 1079.51 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

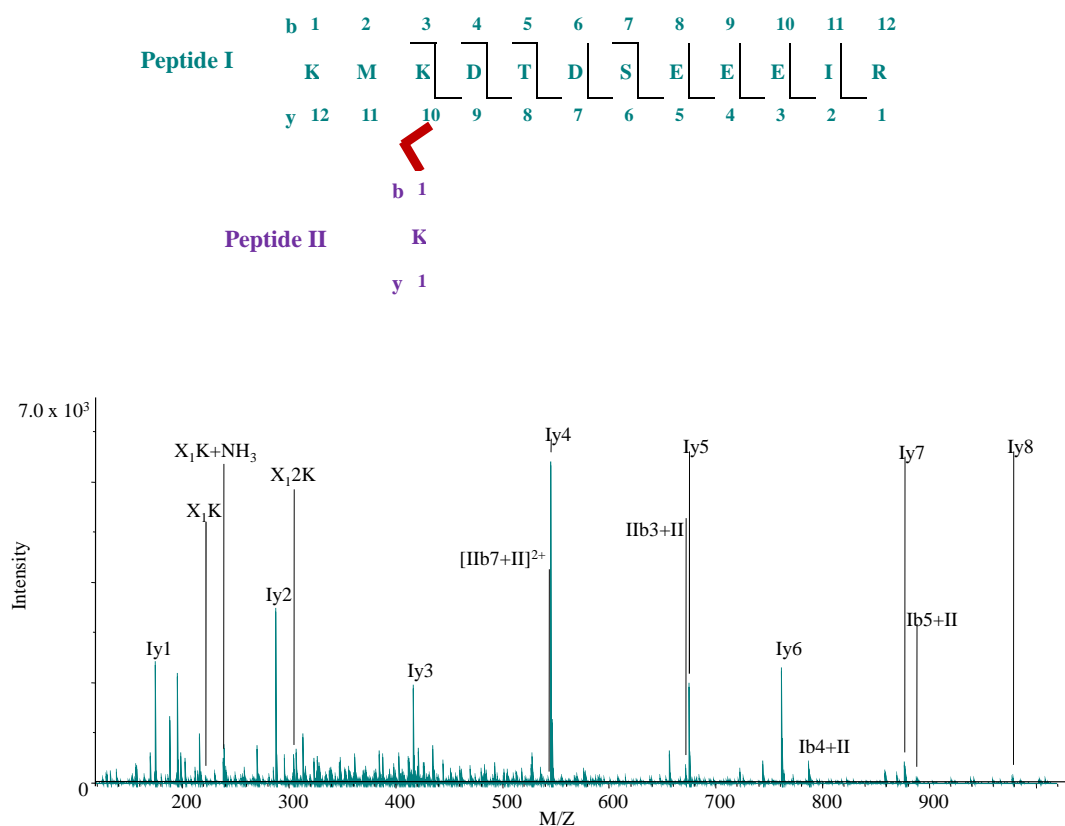


Figure 4.12: Interpeptide BS³⁺ calmodulin cross-link m/z 588.96 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom) ; Cross-linker bridges are indicated in red.

4.2.3.2 Calmodulin-Melittin Cross-linked Peptides

Eight calmodulin-melittin BS³ cross-linked peptides were identified, which are listed in Table 4.5. The BS³ cross-linked structure ⁹¹VFDKDGNGYISAAELR¹⁰⁶^1GIGAVLK⁷ appeared as a triply charged species at m/z 850.46 and a doubly charged species at m/z 1275.19 (Figure 4.13). A complete series of unmodified Iy1 to Iy12 and Ib2 to Ib3 ions confirmed the calmodulin peptide component. Type 2 ions Ib4+II to Ib12+II and Iy13+II to Iy15+II localized the cross-link to K94. A full unmodified y ion series for the

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

melittin peptide verified its sequence and type 2 ion Ib2+I localized cross-linking to G1.⁹¹VFDKDGNGYISAAELR¹⁰⁶∧²³KR²⁴, appearing as a triply charged species at a m/z 732.39, was identified (Figure 4.14). A full y ion series and type 1 ion for the calmodulin peptide and a modified melittin peptide type 1 ion verified the components of the cross-linked structure. Consecutive type 2 ions Ib4+II to Ib12+II and Iy13+II to Iy15+II localized the cross-linking site to K94, which was cross-linked to K23, the only lysine in the melittin peptide.

The species at m/z 699.36, with a charge of four, corresponded to the cross-linked structure: ⁷⁵KM^(ox)KDTDSEEEIREAFR⁹⁰∧¹GIGAVLK⁷ (Figure 4.15). A complete unmodified y ion series of both component peptides and a type 1 ion of the melittin peptide confirmed the cross-linked species sequence. Iy15(ox) confirmed M76 oxidation, a modification also observed in control samples (see Table 3.1). Type 2 ions Ib2+I to Ib6+I localized the cross-linking to G1. Ib3+II, Ib4+II and Ib14+II localized the cross-link to ⁷⁵KM^(ox)K⁷⁷. Since Iy15 and Iy14 appeared unmodified by the cross-linker, calmodulin K75 likely formed the cross-link with melittin G1.

The species with a charge of four appearing at m/z 733.35 (Figure 4.16) and at m/z 741.35 corresponded to ¹²⁷EADIDGDGQVNYEEFVQMMTAK¹⁴⁸∧²³KR²⁴ where the latter differed due to oxidized methionines. Unmodified Ib2 to Ib16 ions confirmed the sequence of the calmodulin peptide and type 2 ions Iy1+II to Iy11+II localized the cross-linking to calmodulin K148 and melittin K23. A type 1 ion (II+X₁) for the melittin peptide confirmed its presence. Likewise, for the oxidized species with a charge of four at m/z 741.35, unmodified Ib2 to Ib11 ions confirmed the calmodulin peptide and type 1 ion (II+X₁) verified the melittin peptide. Type 2 ions Iy1+II to Iy12+II localized the cross-linking to calmodulin K148 and melittin K23.

Cross-linking between calmodulin K77 and melittin K23 was supported by three cross-linked species appearing with a charge of four and at m/z 481.00, 520.02, and 448.97 with structures ⁷⁵KMKDTDSEEEIR⁹⁰∧²³KR²⁴, ⁷⁵KMKDTDSEEEIR⁹⁰∧²²RKR²⁴ and ⁷⁶MKDTDSEEEIR⁹⁰∧²³KR²⁴, respectively (Figures 4.17, 4.18, and 4.19). In all three species, an unmodified type 1 ion of the melittin peptide, a series of unmodified y ions for the calmodulin peptide, and type 2 b ions of the calmodulin peptide connected to the whole melittin peptide verified

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

both peptide components and confirmed that calmodulin K77 was cross-linked to melittin K23.

Table 4.5: BS³⁺ calmodulin-melittin interpeptide cross-linked species are listed and classified as capturing antiparallel (shaded in blue) or parallel (white) binding. Cross-linking sites are highlighted in red. For species appearing with two different charge states, annotated MS/MS spectra is shown for the *m/z* marked with an “*”.

| Cross-Linked Species | | | | Melittin Peptide | | Calmodulin Peptide | | |
|----------------------|----------|----------------------------|-----------------------------|---------------------------|-------------|-----------------------------------|-------------|---|
| <i>m/z</i> | <i>z</i> | [M] _{exp} (Da) | [M] _{Calc} (Da) | Mass Accuracy (ppm) | [M] (Da) | Sequence | [M] (Da) | Sequence |
| *732.39 | 3 | 2194.16 | 2194.14 | 10 | 302.21 | ²³ KR ²⁴ | 1753.86 | ⁹¹ VFDK K DGNGYISAAELR ¹⁰⁶ |
| 549.54 | 4 | | | | | | | |
| 699.36 | 4 | 2793.46 | 2793.43 | 11 | 656.42 | ¹ GIGAVLK ⁷ | 1998.94 | ⁷⁵ KM ^(ox) KDTDSEEEIREAFR ⁹⁰ |
| *733.35 | 4 | 2929.39 | 2929.35 | 14 | 302.21 | ²³ KR ²⁴ | 2489.07 | ¹²⁷ EADIDGDGQVNYEEFVQM TAK ¹⁴⁸ |
| 977.46 | 3 | | | | | | | |
| 741.35 | 4 | 2961.38 | 2961.35 | 12 | 302.21 | ²³ KR ²⁴ | 2521.07 | ¹²⁷ EADIDGDGQVNYEEFVQM ^(ox) M ^(ox) TAK ¹⁴⁸ |
| 481.00 | 4 | 1919.99 | 1919.96 | 15 | 302.21 | ²³ KR ²⁴ | 1479.69 | ⁷⁵ KMKDTDSEEEIR ⁸⁶ |
| 520.02 | 4 | 2076.07 | 2076.06 | 6 | 458.31 | ²² RKR ²⁴ | 1479.69 | ⁷⁵ KMKDTDSEEEIR ⁸⁶ |
| 448.97 | 4 | 1791.90 | 1791.87 | 17 | 302.21 | ²³ KR ²⁴ | 1351.59 | ⁷⁶ MKDTDSEEEIR ⁸⁶ |
| *850.46 | 3 | 2548.38 | 2548.35 | 9 | 656.42 | ¹ GIGAVLK ⁷ | 1753.86 | ⁹¹ VFDK K DGNGYISAAELR ¹⁰⁶ |
| 1275.19 | 2 | | | | | | | |

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

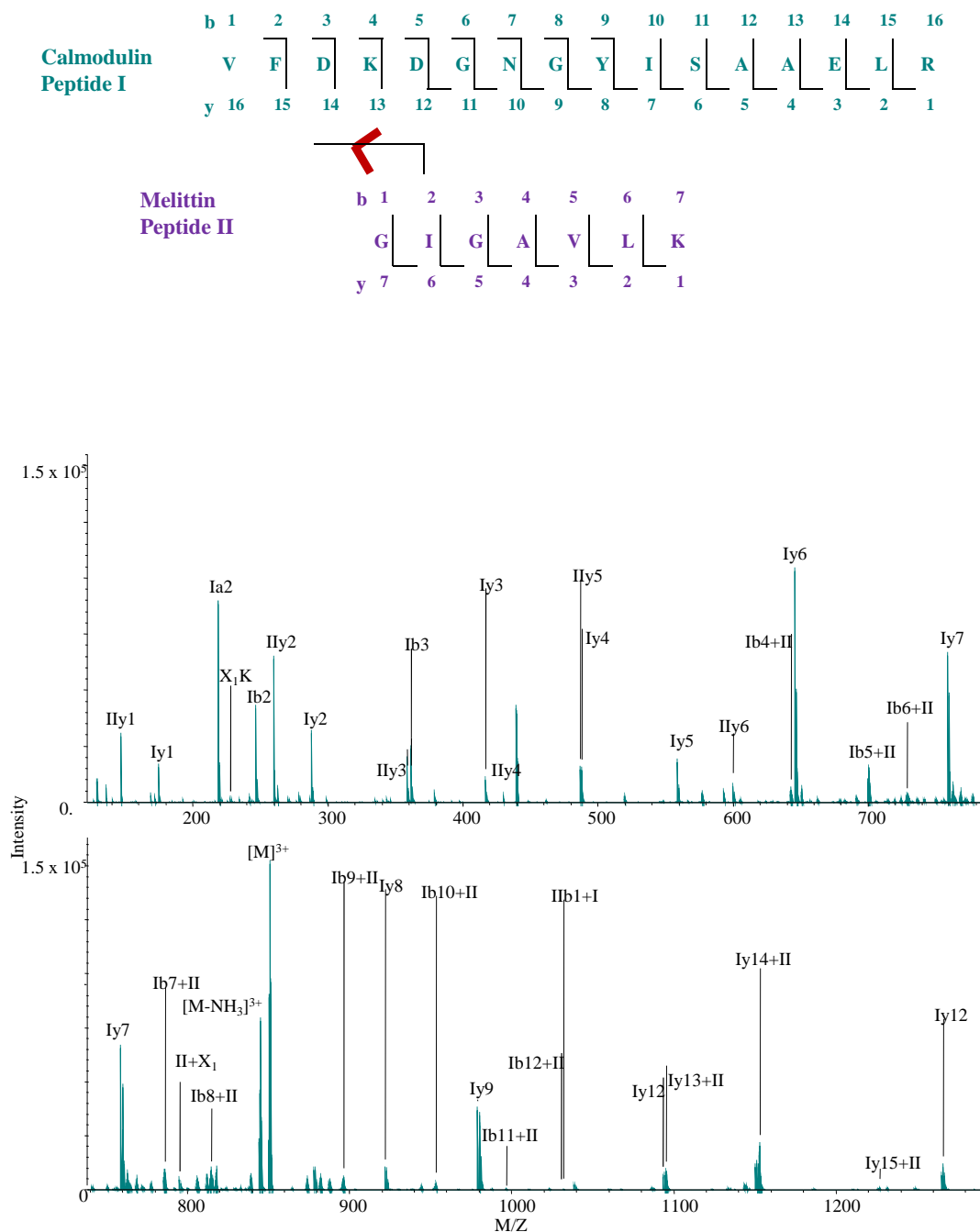


Figure 4.13: Interpeptide BS³⁺ calmodulin-melittin cross-link m/z 850.46 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. 121

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

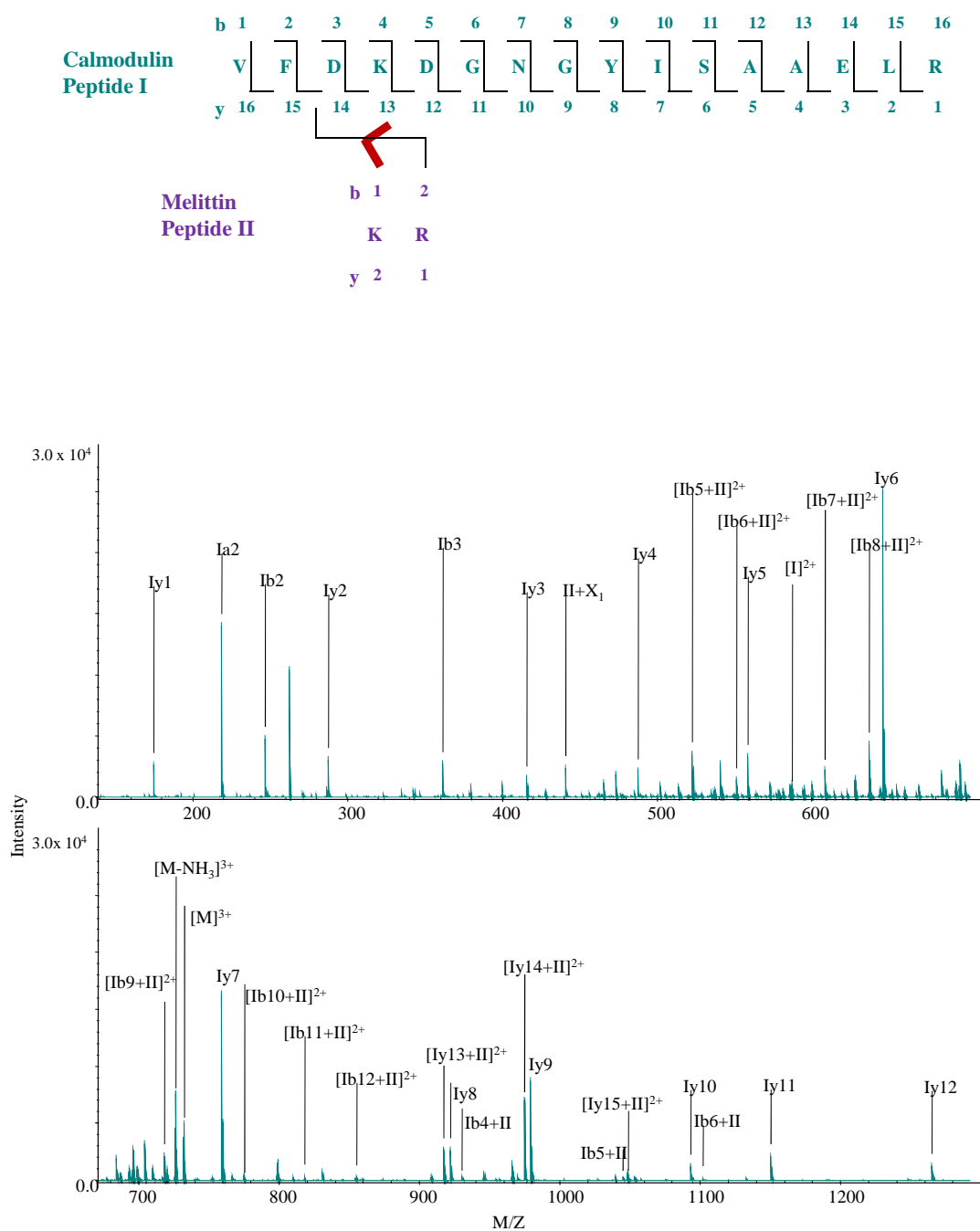


Figure 4.14: Interpeptide BS^{3+} calmodulin-melittin cross-link m/z 732.39 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

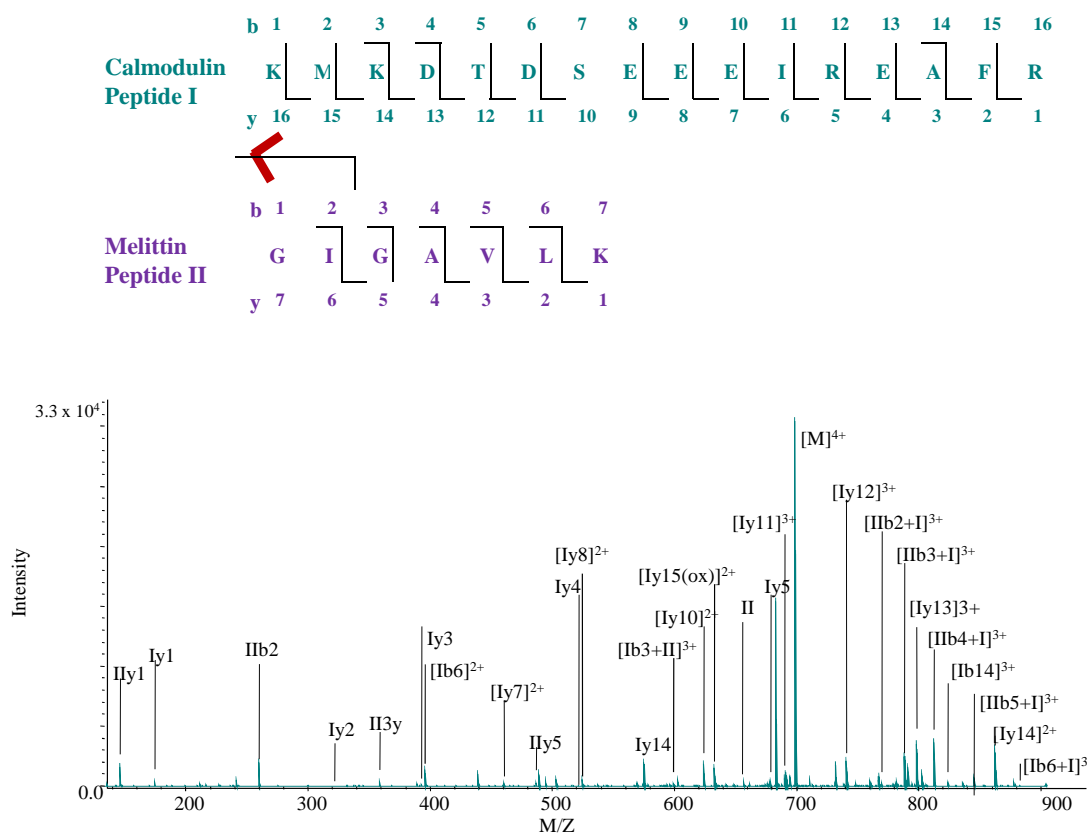


Figure 4.15: Interpeptide BS³⁺ calmodulin-melittin cross-link m/z 699.36 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

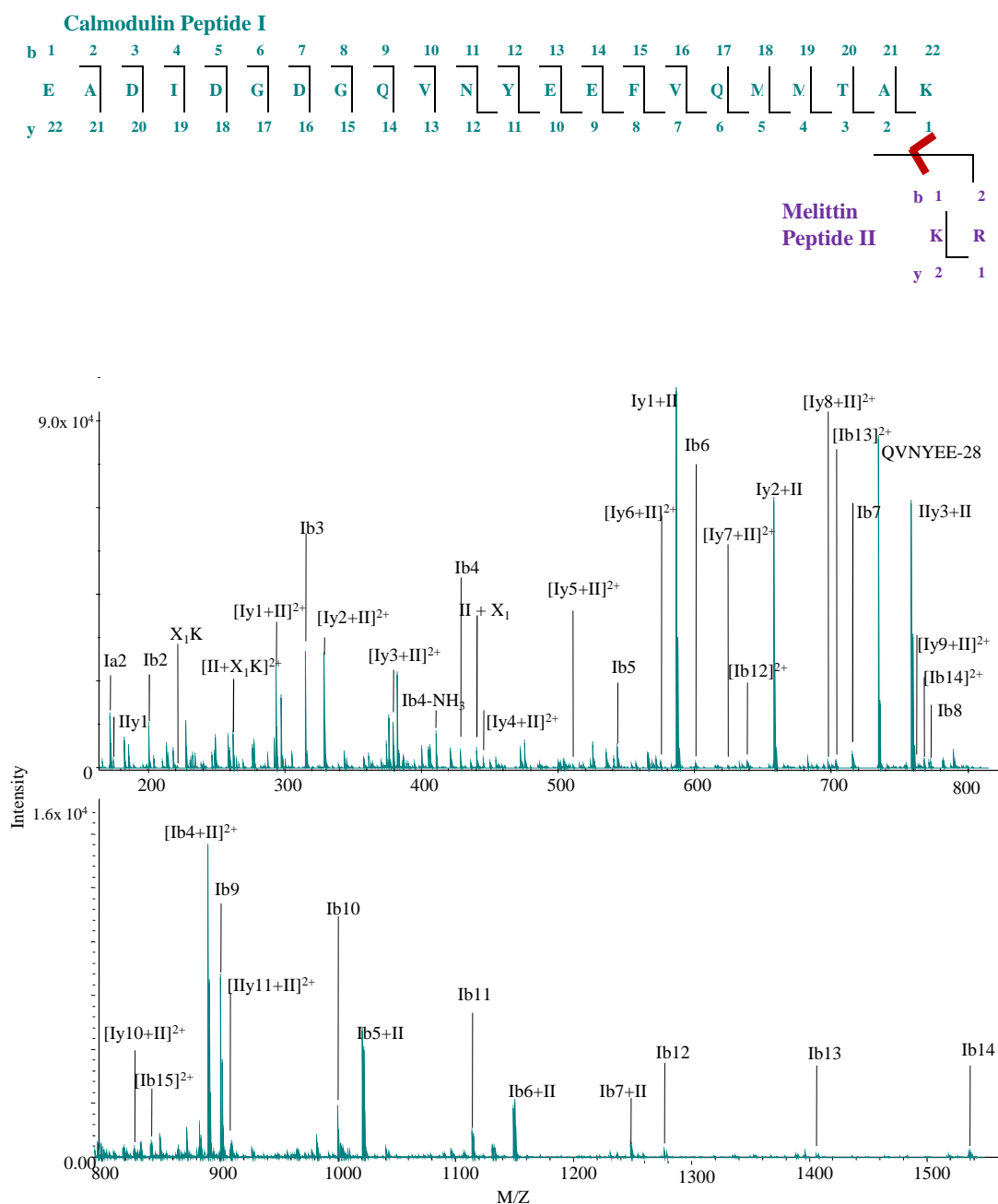


Figure 4.16: Interpeptide BS³⁺ calmodulin-melittin cross-link m/z 733.35 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

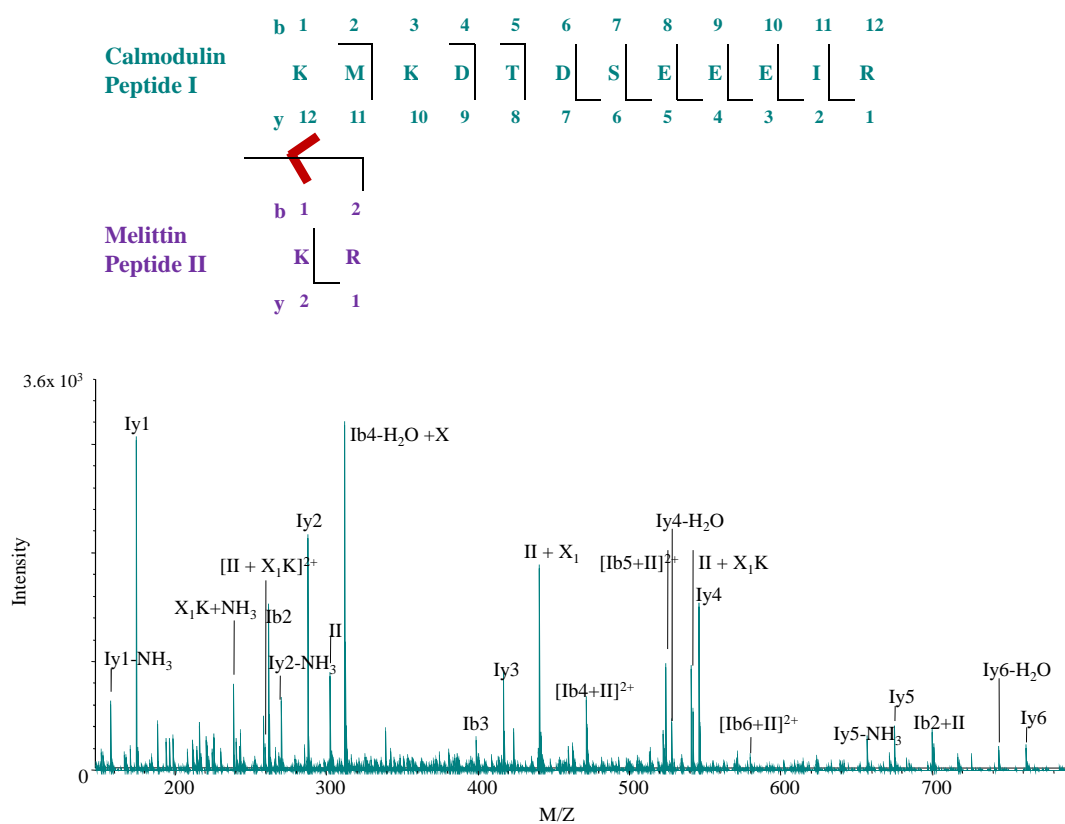


Figure 4.17: Interpeptide BS³⁺ calmodulin-melittin cross-link m/z 481.00 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

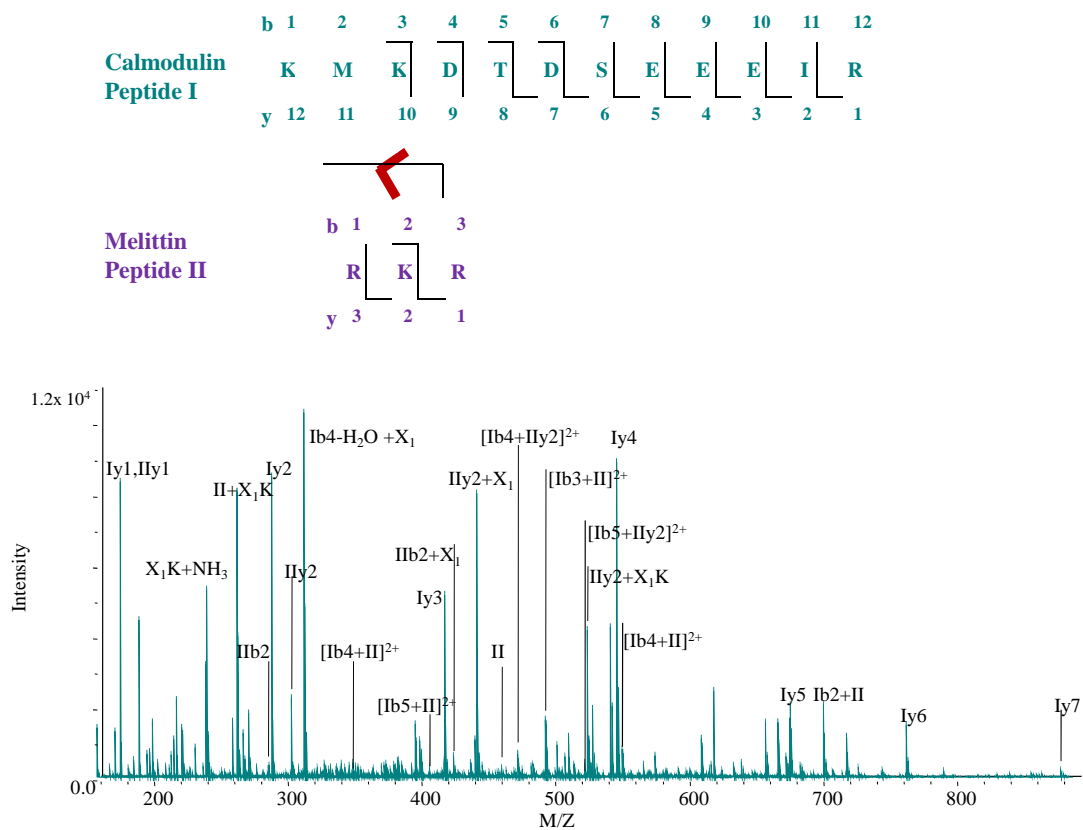


Figure 4.18: Interpeptide BS³⁺ calmodulin-melittin cross-link m/z 520.02 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

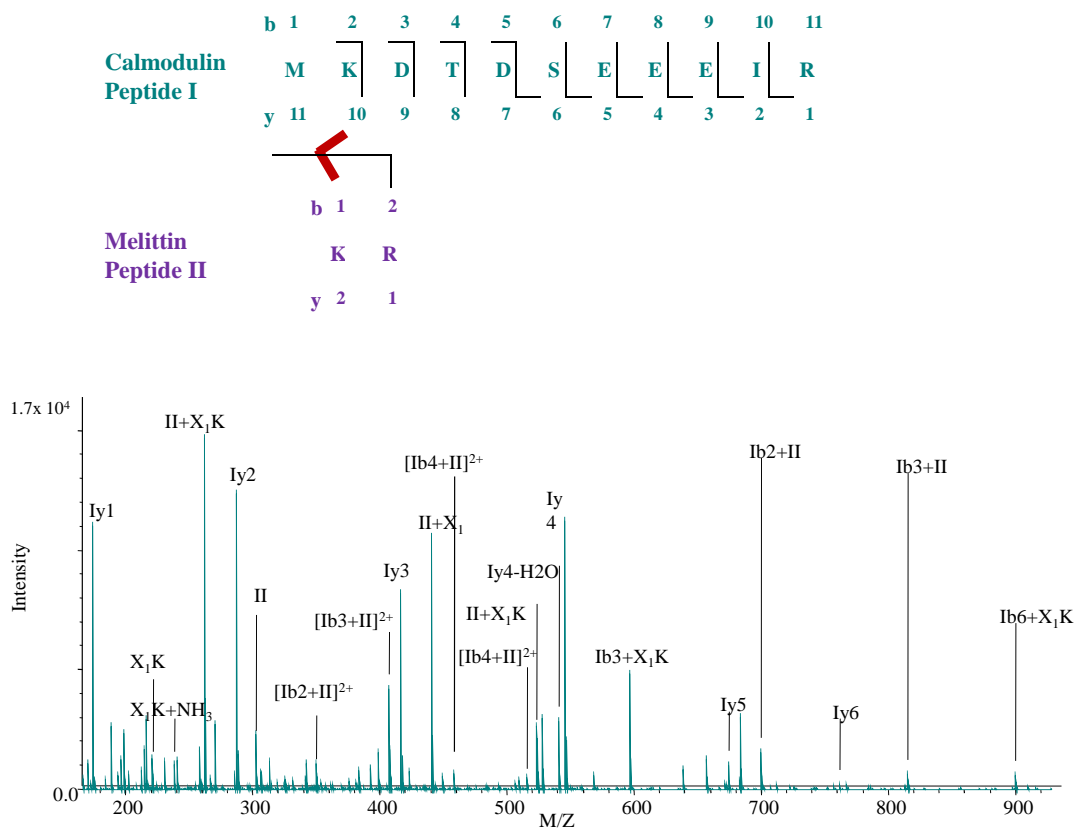


Figure 4.19: Interpeptide BS³⁺ calmodulin-melittin cross-link m/z 448.97 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

Previous MS [99] and MS/MS [100] experiments confirmed BS³ cross-linking of calmodulin K77 to melittin G1. Similar regions were shown to cross-link in this present study, which localized cross-linking between calmodulin K75 to melittin G1. In addition, MS/MS verified cross-linking of calmodulin K94 to melittin G1 and K23, calmodulin K77 to melittin K23, and calmodulin K148 to melittin K23 were discovered for the first time.

Diagnostic ions (See Figure 1.10b) X₁K, X₁K+NH₃, and X₁2K ($m/z = 222.14$, 239.17 and 305.22 for BS³, respectively), which have

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

previously been reported [52], appeared in the MS/MS spectra of $^{75}\text{KMKD}\text{T}\text{D}\text{S}\text{E}\text{E}\text{E}\text{I}\text{R}^{86\wedge 75}\text{KMKD}\text{T}\text{D}\text{S}\text{E}\text{E}\text{E}\text{I}\text{R}^{86}$ and $\text{K}^{75}\text{KMKD}\text{T}\text{D}\text{S}\text{E}\text{E}\text{E}\text{I}\text{R}^{86}$. The appearance of at least one diagnostic ion indicates the presence of the BS^3 cross-linker. Only one diagnostic ion, $\text{X}_1\text{K}+\text{NH}_3$, appeared in the MS/MS spectrum of $^{76}\text{MK}^{77\wedge 91}\text{VFDKDGNGYISAAELR}^{106}$. The MS/MS spectra of all three species $^{75}\text{KMKD}\text{T}\text{D}\text{S}\text{E}\text{E}\text{E}\text{I}\text{R}^{90\wedge 23}\text{KR}^{24}$, $^{75}\text{KMKD}\text{T}\text{D}\text{S}\text{E}\text{E}\text{E}\text{I}\text{R}^{90\wedge 22}\text{RKR}^{24}$ and $^{76}\text{MKD}\text{T}\text{D}\text{S}\text{E}\text{E}\text{E}\text{I}\text{R}^{90\wedge 23}\text{KR}^{24}$ contained diagnostic ion $\text{X}_1\text{K}+\text{NH}_3$ and a type 1 ion corresponding to the melittin peptide modified with a (+ X_1K), indicating that fragmentation occurred at the peptide bond of K77. The unmodified type 1 ion of the melittin peptide in all three cross-linked species suggests that fragmentation also occurred at the amide bond between K23 and the BS^3 cross-linker bridge. $^{91}\text{VFDKDGNGYISAAELR}^{106\wedge 1}\text{GIGAVLK}^7$ contained both type 1 ion $\text{II}+\text{X}_1$ and diagnostic ion X_1K , indicating that the cross-linker bridge was fragmented at the amide bond connecting calmodulin K94 to the cross-linker bridge, and at the calmodulin backbone peptide bond of K94. Diagnostic ion X_1K and type 1 ion $\text{II}+\text{X}_1\text{K}$ in the MS/MS spectrum of $^{127}\text{EADIDGDGQVNYEEFVQMMTAK}^{148\wedge 23}\text{KR}^{24}$ suggests that fragmentation occurred at the peptide bond of K148. However, diagnostic ions in both $^{91}\text{VFDKDGNGYISAAELR}^{106\wedge 23}\text{KR}^{24}$ and $^{75}\text{KM}^{(ox)}\text{KDT}\text{D}\text{S}\text{E}\text{E}\text{E}\text{I}\text{R}\text{EAFR}^{90\wedge 1}\text{GIGAVLK}^7$ MS/MS spectra were not detected. Generally, simultaneous fragmentation of the cross-linker bridge and peptide backbone (type 3 ions) were not observed and this is consistent with literature [51] However, an exception is $^{75}\text{KMKD}\text{T}\text{D}\text{S}\text{E}\text{E}\text{E}\text{I}\text{R}^{86\wedge 75}\text{KMKD}\text{T}\text{D}\text{S}\text{E}\text{E}\text{E}\text{I}\text{R}^{86}$. It remains to be seen whether this is a characteristic behavior of cross-linked identical peptides under CID, due to their equal bond energies.

4.2.4 SulfoEGS

4.2.4.1 Calmodulin-Calmodulin Cross-linked Peptides

Two calmodulin interpeptide cross-links were identified in sulfoEGS treated samples as shown in Table 4.6, which lists the m/z , charge, monoisotopic mass (experimental and calculated), mass accuracy (ppm), the mass and sequence of each component peptide (highlighted residues correspond to cross-linked sites). The cross-

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

linked structure $^{14}\text{EAFSLFDKDGDTITTK}^{30\wedge 91}\text{VFDKDGNGYISAAELR}^{106}$ appeared with a charge of four at m/z 956.96 and with a charge of five at m/z 765.77 (Figure 4.20). Fragment ions Iy1 to Iy9 and Ib2 to Ib7, and Ily1 to Ily12, Iib2 and Iib3 confirmed the sequence of each component peptide, respectively. The following type 2 ions localized the cross-linking sites to K21 and K94: Ib13+II to Ib15+II, Iy12+II, Iy14+II, Iy15+II, Iib5+I to Iib10+I, Ily13+I and Ily14+I. The cross-linked species with a charge of four at m/z 934.46 corresponded to the following structure composed of identical peptides: $^{91}\text{VFDKDGNGYISAAELR}^{106\wedge 91}\text{VFDKDGNGYISAAELR}^{106}$ (Figure 4.21). Fragment ions I/Ily1 to I/Ily12, I/Iib2 and I/Iib3 confirmed the peptide sequence. Type 2 ion I/Iib8+II/I verified cross-linking. Since K94 is the only reactive residue in each peptide, cross-linking must have occurred at this site. The presence of type 2 ions and lack of type 1 ions indicated that the cross-linker bridge remained intact unlike the BS³ cross-link between identical peptides.

Table 4.6: SulfoEGS calmodulin interpeptide cross-linked species, in which cross-linking sites are highlighted in red. For species appearing with two different charge states, annotated MS/MS spectra is shown for the m/z marked with an “*”.

| Cross-Linked Species | | | | | Calmodulin Peptide 1 | | Calmodulin Peptide 2 | |
|----------------------|--------|----------------------------|-----------------------------|------------------------|----------------------|--------------------------------------|----------------------|--------------------------------------|
| m/z | z | $[M]_{\text{exp}}$ (Da) | $[M]_{\text{calc}}$ (Da) | Mass Accuracy (ppm) | $[M]$ (Da) | Sequence | $[M]$ (Da) | Sequence |
| *956.96 765.77 | 4 5 | 3823.85 | 3823.80 | 13 | 1753.86 | $^{91}\text{VFDKDGNGYISAAELR}^{106}$ | 1753.86 | $^{14}\text{EAFSLFDKDGDTITTK}^{30}$ |
| 934.46 | 4 | 3733.83 | 3733.77 | 14 | 1753.86 | $^{91}\text{VFDKDGNGYISAAELR}^{106}$ | 1479.69 | $^{91}\text{VFDKDGNGYISAAELR}^{106}$ |

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

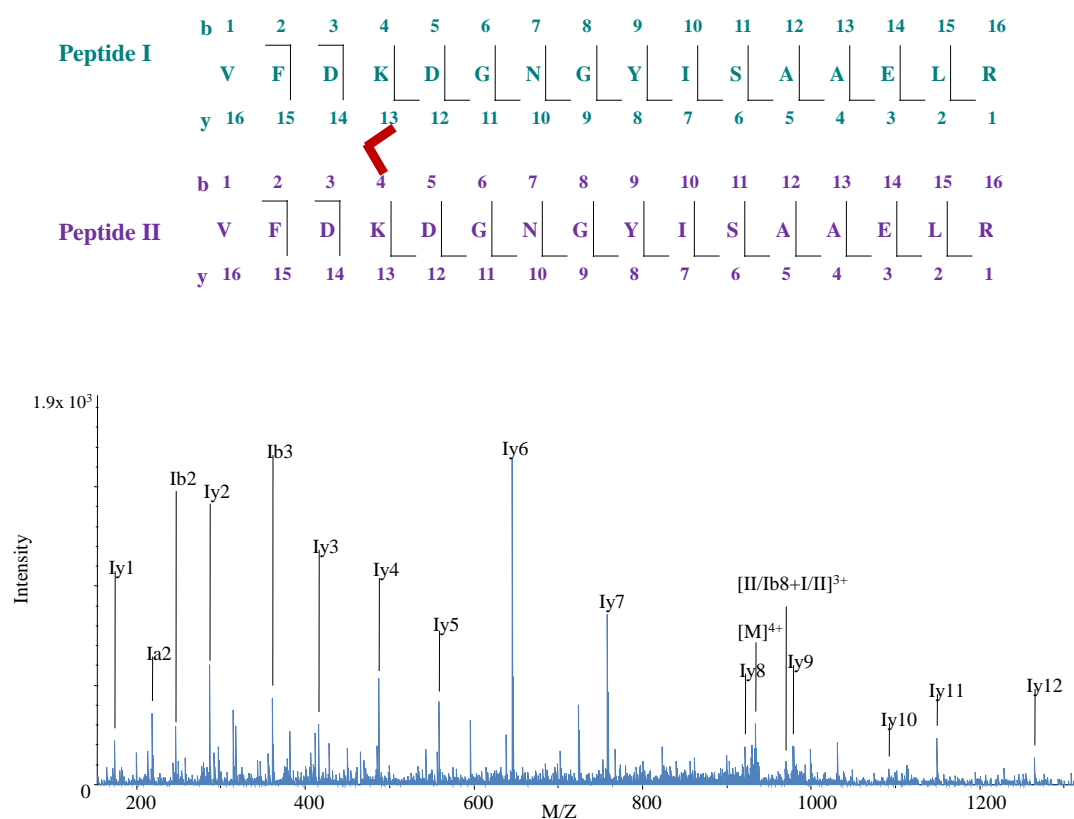


Figure 4.20: Interpeptide sulfoEGS calmodulin cross-link m/z 956.96 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

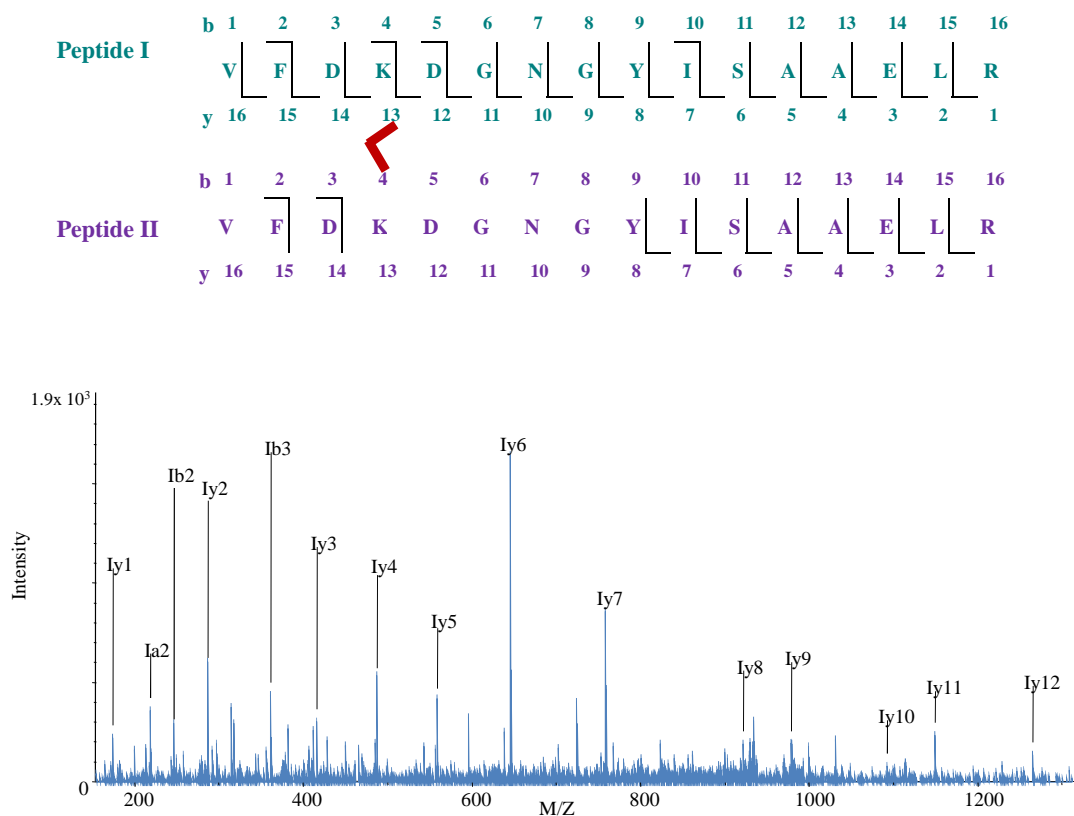


Figure 4.21: Interpeptide sulfoEGS calmodulin cross-link m/z 934.46 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. MS/MS spectra is annotated such that I = II.

4.2.4.2 Calmodulin-Melittin Cross-linked Peptides

Three calmodulin-melittin sulfoEGS cross-linked species were discovered and are listed in Table 4.7. The species with a charge of four appearing at m/z 571.54 corresponded to the structure $^{91}\text{VFDKDGNGYISAAELR}^{106\wedge 23}\text{KR}^{24}$ (Figure 4.22). This structure and sequence was confirmed with a type 1 ion of the melittin peptide and a series of unmodified y ions (Iy1 to Iy8), a Ib2 ion and a Ib3 ion of the calmodulin peptide. With only one lysine in each peptide, cross-linking must

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

have occurred between calmodulin K94 and melittin K23. Type 2 ions Iy15+II, Ib8+II and Ib11+II were also consistent with this conclusion. The cross-linked species with a charge of four at m/z 635.81 matched the following structure: $^{76}\text{MKD}\text{TDSEEEI}\text{REAFR}^{90}\wedge^{22}\text{RKR}^{24}$ (Figure 4.23). An unmodified type 1 ion for the melittin peptide, a modified type 1 ion (I+X₁) of the calmodulin peptide, and an almost complete y ion series (Iy1, Iy3, Iy5, Iy6 to Iy13) of the calmodulin peptide confirmed the cross-linked structure. SulfoEGS cross-linking was determined to occur between calmodulin K77 and melittin K23, the only possible reactive sites, and type 2 ions Ib11+II and Iy15+II also supported this finding. $^{14}\text{EAFSLFDKDG}\text{DGTITTK}^{30}\wedge^{23}\text{KR}^{24}$ appeared as a triply charged species at m/z 791.72 (Figure 4.24). Unmodified ions Iy1 to Iy9 and Ib2 to Ib7 confirmed the calmodulin peptide sequence and a type 1 ion verified the presence of the melittin peptide. Type 2 ions Iy10+II to Iy16+II, Ib8+II, Ib9+II, Ib11+II and Ib12+II localized cross-linking between K30 to K23. This is the first report of MS/MS verified sulfoEGS cross-linking in the calmodulin-melittin complex.

Table 4.7: SulfoEGS calmodulin-melittin interpeptide cross-linked species are listed and classified as capturing antiparallel (shaded in blue) or parallel (white) binding. Cross-linking sites are highlighted in red.

| Cross-Linked Species | | | | | Melittin Peptide | | Calmoulin Peptide | |
|----------------------|-----|----------------------------|-----------------------------|------------------------|------------------|---------------------------------|-------------------|--|
| m/z | z | [M] _{exp} (Da) | [M] _{Calc} (Da) | Mass Accuracy (ppm) | [M] (Da) | Sequence | [M] (Da) | Sequence |
| 571.54 | 4 | 2282.15 | 2282.12 | 14 | 302.21 | ²³ KR ²⁴ | 1753.86 | ⁹¹ VFDKDGNGYISAAELR ¹⁰⁶ |
| 635.81 | 4 | 2539.25 | 2539.20 | 20 | 458.31 | ²² RKR ²⁴ | 1854.84 | ⁷⁶ MKD ^{TDSEEEI} REAFR ⁹⁰ |
| 791.72 | 3 | 2372.17 | 2372.14 | 13 | 302.21 | ²³ KR ²⁴ | 1843.884 | ¹⁴ EAFSLFDK ^{DG} DGTITTK ³⁰ |

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

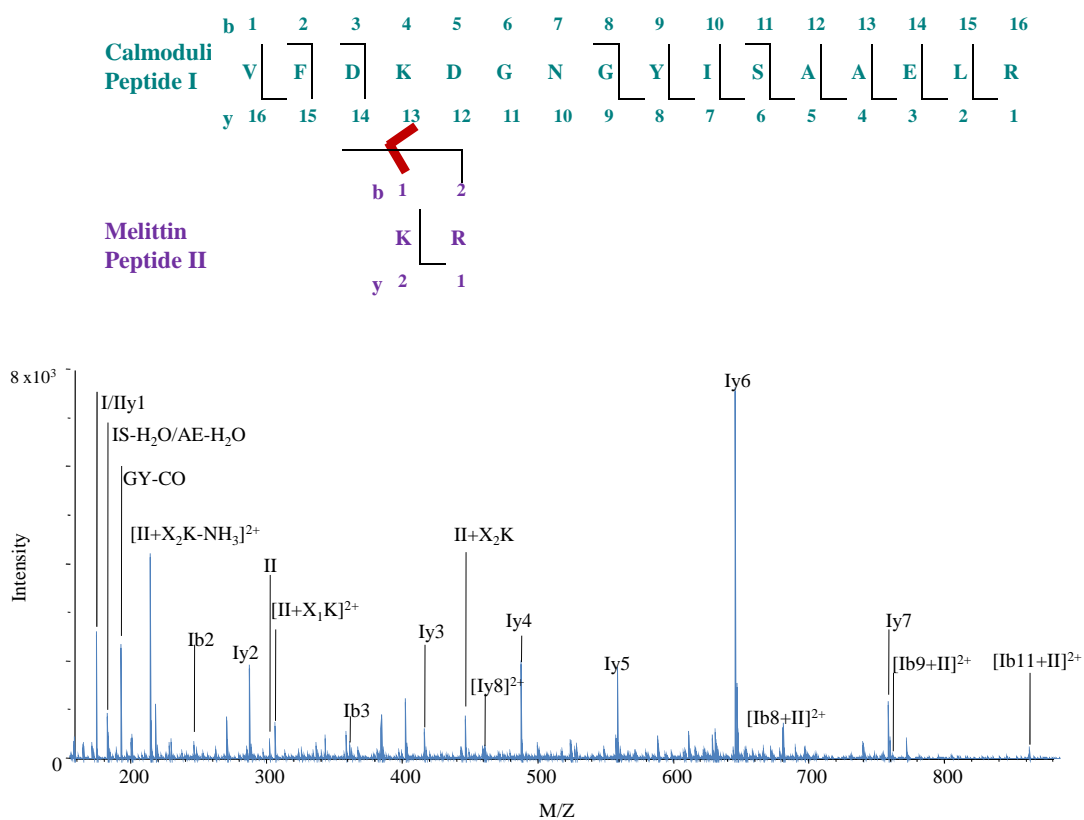


Figure 4.22: Interpeptide sulfoEGS calmodulin-melittin cross-link m/z 571.54 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

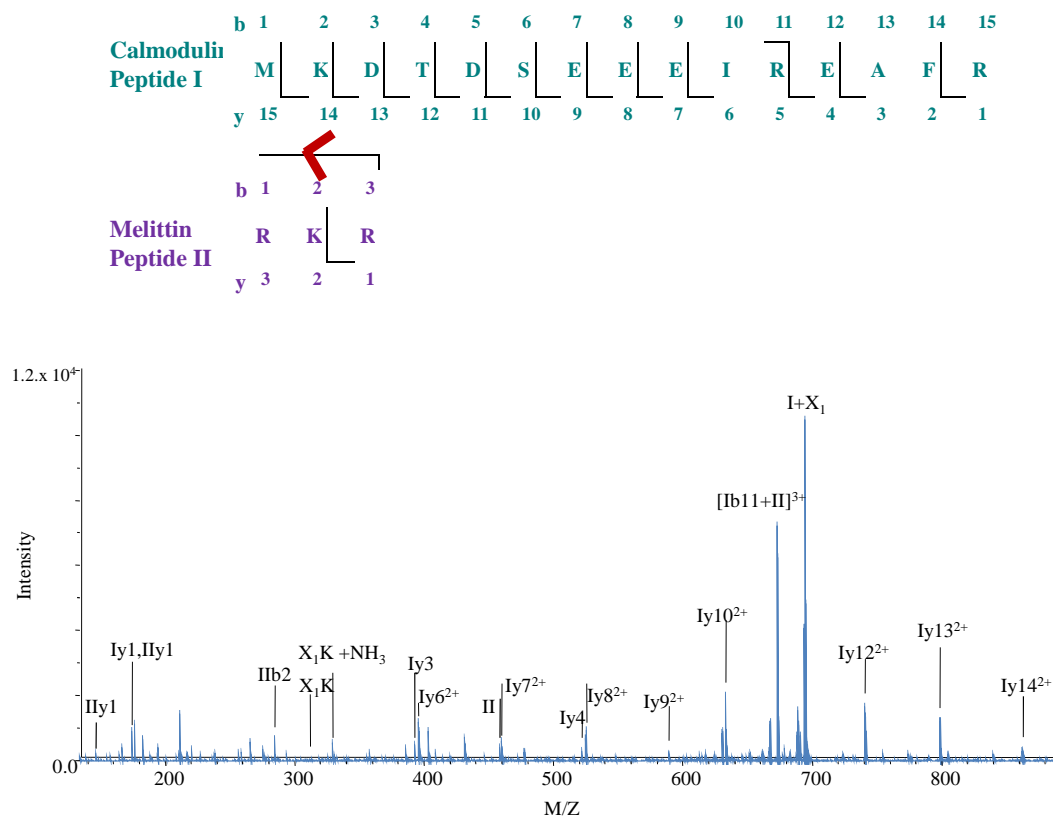


Figure 4.23: Interpeptide sulfoEGS calmodulin-melittin cross-link m/z 635.81 ($z = 4$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

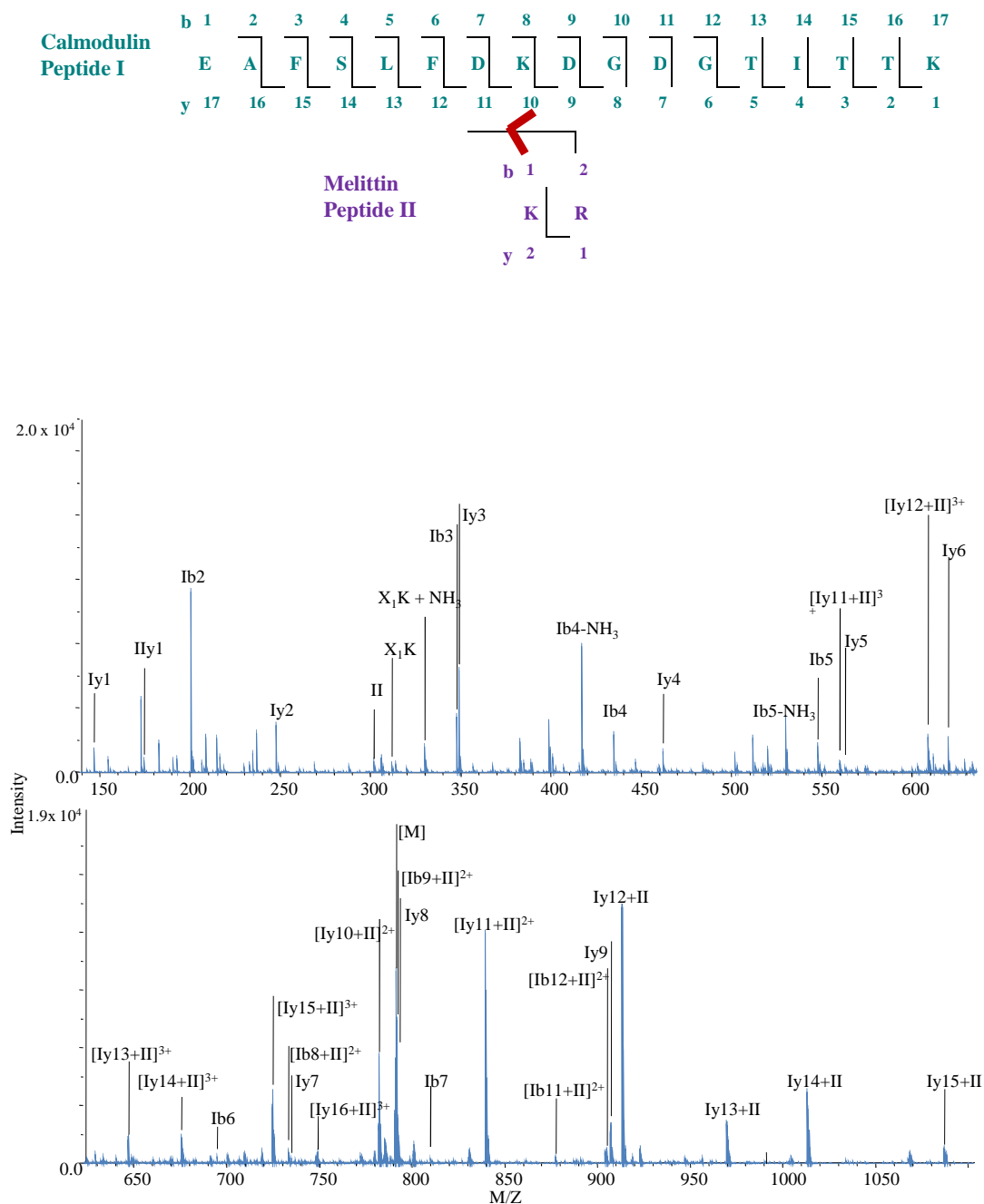


Figure 4.24: Interpeptide sulfoEGS calmodulin-melittin cross-link m/z 791.72 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. 135

4.2. Tandem Mass Spectrometric Fragmentation of Other Cross-linkers

Diagnostic ions (See Figure 1.10b) X_1K and X_1K+NH_3 at $m/z = 312$ and 330 , respectively were observed and confirmed the presence of the cross-linker. Figure 1.10c summarizes the fragmentation patterns that have been previously observed for sulfoEGS cross-linked species. Fragmentation occurred at the peptide bond of the cross-linked lysine residue on $^{23}KR^{24}$ giving rise to $II+X_1K$. Fragmentation also occurred within the cross-linker connected to $^{23}KR^{24}$ at the C-O bond, giving rise to $II+X_2K$, which is commonly observed with sulfoEGS cross-linking [162]. The MS/MS spectra of $^{76}MKDSTDSEEEIREAFR^{90\wedge 22}RKR^{24}$ and $^{14}EAFSLFDKDGDTITTK^{30\wedge 23}KR^{24}$ contained diagnostic ion X_1K . Fragmentation of $^{91}VFDKDGNGYISAAELR^{106\wedge 23}KR^{24}$ produced diagnostic ions $II+X_2K-NH_3$ and $II+X_2K$. Both calmodulin cross-linked species did not produce type 1 or type 3 ions under CID. However, all calmodulin-melittin cross-linked species generated a type 1 ion for the melittin peptide and $^{76}MKDSTDSEEEIREAFR^{90\wedge 22}RKR^{24}$ produced type 1 ions for both peptide components under CID.

4.2.5 Summary of Other Cross-linker Fragmentation Patterns

Common fragmentation patterns were observed across most EDC, sulfoDST, BS^3 and sulfoEGS cross-linked peptides. In general, type 1 ions of at least the smaller component peptide and unmodified backbone fragment ions of at least one (usually y ions of the larger peptide) is expected. Type 2 ions with the cross-linker attached to a whole component is of less importance in NHS ester cross-linking to localize cross-linking since cross-linking almost exclusively occurs between K residues. Nevertheless, most cross-linked species produced extensive type 2 ions in their MS/MS spectra, which further confirmed cross-linking. Type 3 ions were rarely observed for all cross-linkers. The observation of more extensive backbone fragments from the larger peptide and intact smaller peptide ions in the MS/MS of NHS ester cross-linking is consistent with literature [52]. NHS ester cross-link bridges allow for additional possible fragment ions generated by dissociation of bonds within the bridge. Both sulfoEGS and BS^3 produced X_1K fragmentation at the peptide backbone bond of melittin K23 and both sulfoDST and BS^3 pro-

duced X₁K fragmentation at the peptide backbone bond of calmodulin K94, suggesting that this type of fragmentation may not depend on cross-linker length but the specific cross-linked lysine environment. Production of diagnostic ions reveal the presence of the cross-linker bridge, however, diagnostic ions were not detected in every identified cross-linked species, which was also previously observed in literature [52, 53].

4.3 Tandem Mass Spectrometric Fragmentation of Formaldehyde

4.3.1 Calmodulin-Calmodulin Cross-linked Peptides

The MS/MS spectra of PFA cross-linked calmodulin-melittin was examined. A total of three interpeptide cross-linked species involving calmodulin were identified and confirmed via MS/MS as shown in Table 4.8, which lists the *m/z*, charge, monoisotopic mass (experimental and calculated), mass accuracy (ppm), the mass and sequence of each component peptide (highlighted residues correspond to cross-linked/modified sites).

The mass of the triply charged species at *m/z* 666.33 corresponded to the mass of calmodulin peptides ⁷⁵KM^(ox)K⁷⁷ and ¹A^(ac)DQLTEEQIAEFK¹³ plus the 12 Da bridge (Figure 4.25). Type 3 (Iy1 to Iy10 and Ib2 to Ib6) and type 1(I-NH₃) ions confirmed ¹A^(ac)DQLTEEQIAEFK¹³. Type 3 (Iib2(ox), Iiy1, Iiy2(ox)), and type 1(II(ox)+12) ions verified peptide ⁷⁵KM^(ox)K⁷⁷. All together the cross-linked species structure was confirmed. The acetylated N-terminus of ¹A^(ac)DQLTEEQIAEFK¹³ prevented Schiff base, methylol and cross-link formation at this site. In this cross-linked structure, the residues that can form cross-links in the second step, i.e. Q3 and Q8, are only present in ¹A^(ac)DQLTEEQIAEFK¹³. Therefore, it can be concluded that ⁷⁵KM^(ox)K⁷⁷ was modified first at either K75 or K77 and subsequently reacted with Q3 or Q8 in ¹A^(ac)DQLTEEQIAEFK¹³. Modified type 3 ions Ib3+12, Ib5+12 and Ib6+12 suggested that Q3 was involved the cross-link and the type 3 ion, Iiy2(ox) +12, suggested that K77 was modified, supporting cross-linking between Q3 and K77. However, the absence of type 2 ions prevented further validation of cross-linking sites.

4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

The species with a charge of four at m/z 735.59 species corresponded to $^1A^{(ac)}DQLTEEQIAEFK^{13}\wedge^{76}M^{(ox)}KDT-DSEEEIR^{86}$ (Figure 4.26). In the MS/MS spectra, unmodified type 3 (Iy1 to Iy10, Ib2 to Ib7 and Ily1 to Ily9) ions for both component peptides confirmed the sequence of the cross-linked peptides. Type 1 ions for both peptides were present. The type 1 ion ,II(ox)+12- CH_3SOH , resulted from the loss of methanesulfenic acid (-64 Da) observed in M residue containing peptides under CID[163]. Fragmentation occurred such that the +12 Da methylene bridge remained attached to peptide $^{76}M^{(ox)}KDTDSEEEIR^{86}$, as indicated by the II(ox) + 12 type 1 ion. Interestingly, +12 Da modified type 3 ions (Iib2(ox)+12, Iib3(ox)+12, Iib5(ox)+12, and Iib6(ox)+12) only appeared for the peptide containing the K77 as well and localized this mass shift to K77. Therefore, K77 was modified and cross-linked to either Q3 or Q8, similar to the previous cross-linked species discussed, but the lack of type 2 ions prevented further validation of cross-linking sites.

The mass of the triply charged species at m/z 1192.23 matched the following structure: $^{75}KM^{(ox)}KDTDSEEEIREAFR^{90}\wedge^1A^{(ac)}DQLTEEQIAEFK^{13}$ (Figure 4.27). The MS/MS spectra contained only signals pertaining to the whole, component peptide components due to exclusive cross-link bridge fragmentation. The lack of type 3 and type 2 ions prevented peptide backbone sequencing and cross-link site localization. Similar to the previous two cross-linked species, type 1 ion I(ox) +12 indicated that the +12 Da bridge remained attached to the peptide containing K77.

These three cross-linked species propose that K77 was modified in the first step of the reaction and retained the PFA cross-link bridge during type 1 fragmentation. All three PFA cross-linked calmodulin species involve the same region of the protein and only differ in missed cleavage sites at K75 and K77. From the MaxQuant unmodified peptide analysis in chapter 3 (Table 3.1), a missed cleavage after K75 and K77 was observed suggesting that the missed cleavage at these residues can occur even without being modified with PFA. Interestingly, the signal intensity of $^{75}KM^{(ox)}K^{77}$ and $^1A^{(ac)}DQLTEEQIAEFK^{13}$ was 16 times stronger than for $^{75}KM^{(ox)}KDTDSEEEIREAFR^{90}\wedge^1A^{(ac)}DQLTEEQIAEFK^{13}$, indicating that the latter formed in a much smaller quantity. It is not clear whether all cross-linked peptides represented one cross-link formation between the same two sites

4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

or a mixture of all combinations of proposed cross-linking sites. Under these mild cross-linking conditions, it is likely that only major cross-link reaction products are detected. Therefore, evidence supports cross-linking between K77 and Q3 as the major species. In chapter 3, a +12 and +30 Da modification was localized on K77 in modified peptides identified via MaxQuant. This further provides evidence that K77 was modified in the first step of the reaction and formed a cross-link with Q3 in the second step.

Table 4.8: PFA calmodulin interpeptide cross-linked species, in which cross-linking sites are highlighted in red.

| Cross-Linked Species | | | | | Calmodulin Peptide 1 | | Calmodulin Peptide 2 | | Total Modification Mass |
|----------------------|-----|-------------------------|--------------------------|---------------------|----------------------|--|----------------------|---|-------------------------|
| m/z | z | $[M]_{\text{exp}}$ (Da) | $[M]_{\text{calc}}$ (Da) | Mass Accuracy (ppm) | $[M]$ (Da) | Sequence | $[M]$ (Da) | Sequence | |
| 666.33 | 3 | 1996.00 | 1995.99 | 7 | 421.24 | ⁷⁵ KM ^(ox) K ⁷⁷ | 1562.75 | ¹ A ^(ac) DQLTEEQIAEFK ¹³ | 12 |
| 736.59 | 4 | 2942.37 | 2942.34 | 9 | 1367.59 | ⁷⁶ M ^(ox) KD ⁷⁷ TDSEEEIR ⁸⁶ | 1367.59 | ¹ A ^(ac) DQLTEEQIAEFK ¹³ | 12 |
| 1192.23 | 3 | 3573.69 | 3573.68 | 3 | 1998.97 | ⁷⁵ KM ^(ox) KD ⁷⁷ TDSEEEIREAFR ⁹⁰ | 1562.75 | ¹ A ^(ac) DQLTEEQIAEFK ¹³ | 12 |

4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

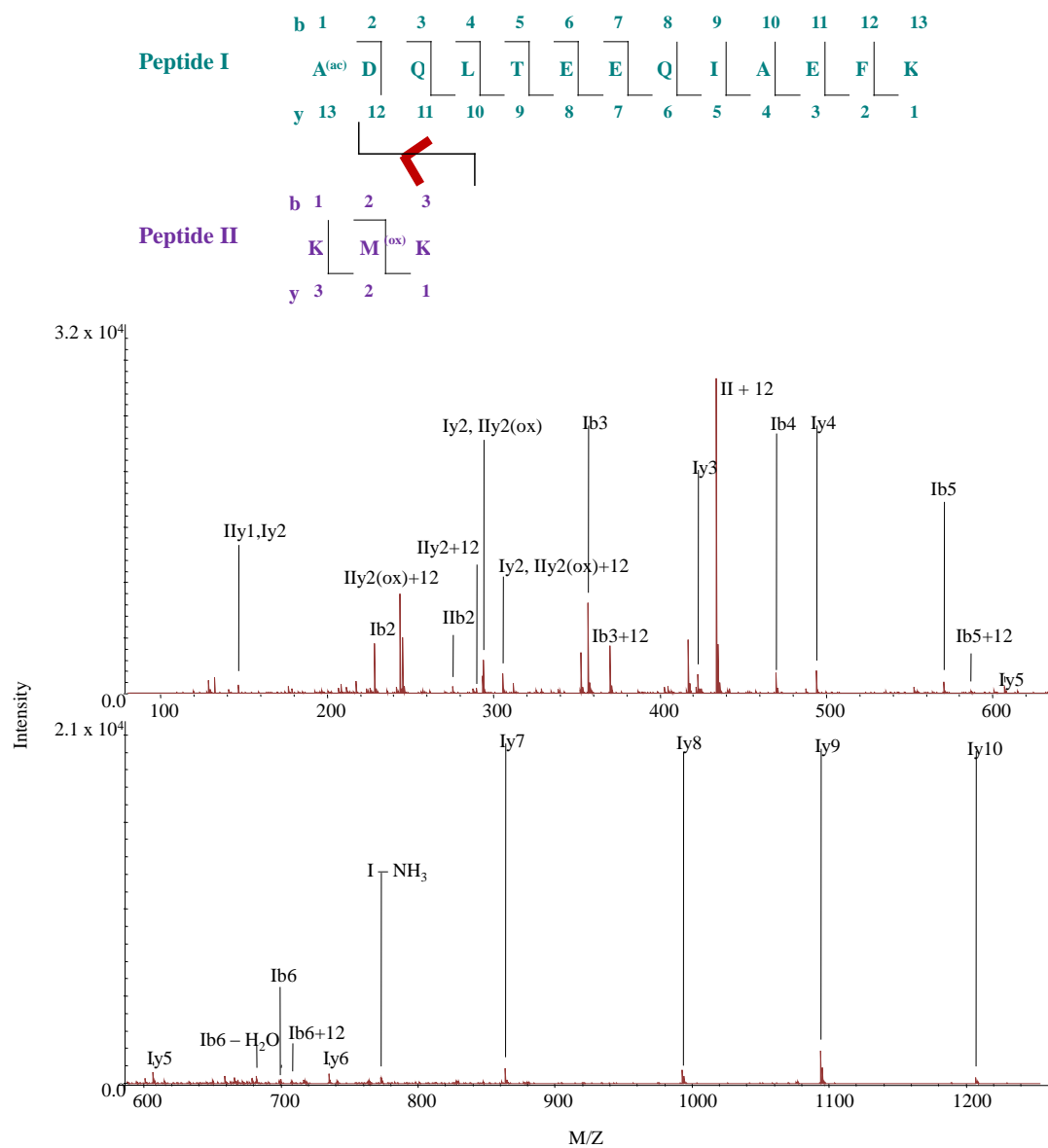


Figure 4.25: Interpeptide PFA calmodulin cross-link m/z 666.33 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. Note: Fragmentation indicated on the backbone of the peptide corresponds to type 3 ions only

4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

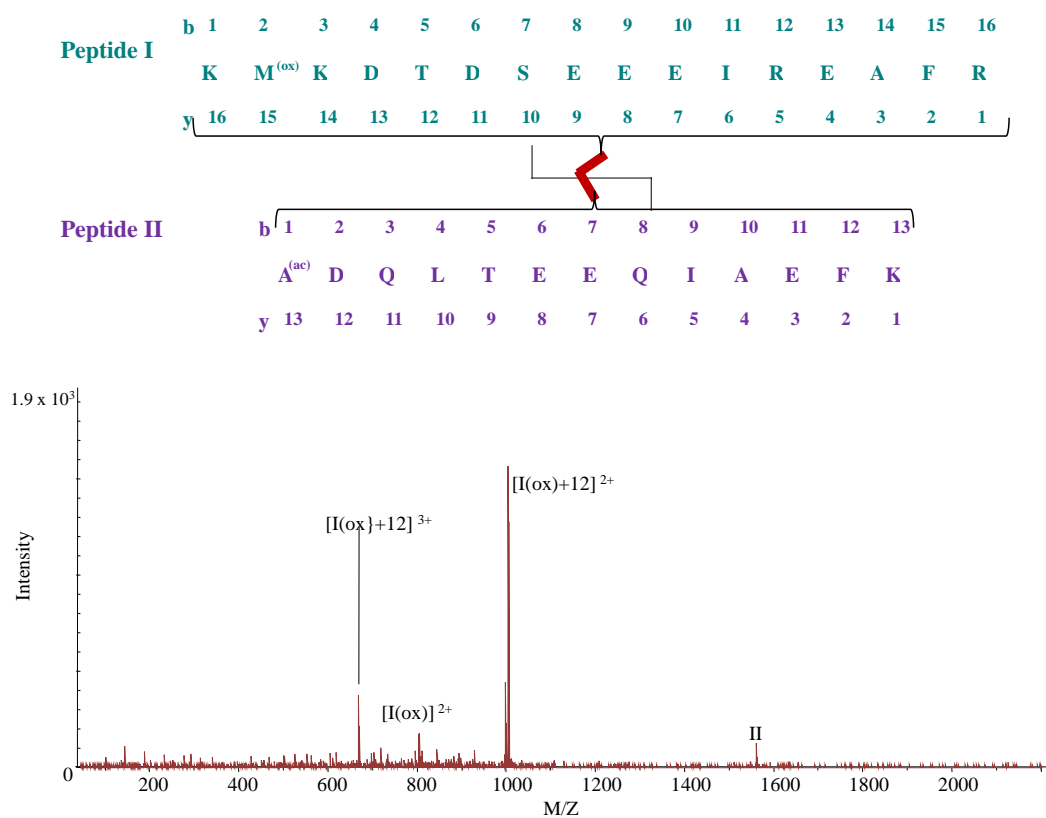


Figure 4.27: Interpeptide PFA calmodulin cross-link m/z 1192.23 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red.

4.3.2 Calmodulin-Melittin Cross-linked Peptides

Four PFA cross-links between calmodulin and melittin were identified as shown in Table 4.9. The candidate with a triple charge at m/z 744.73 matched the mass of melittin peptide $^1\text{GIGAVLK}^7$ cross-linked to calmodulin peptide $^1\text{A}^{(ac)}\text{DQLTEEQIAEFK}^{13}$ (Figure 4.31). Type 3 ions (Iy1 to Iy10 and Ib2 to Ib10) confirmed the sequences of $^1\text{A}^{(ac)}\text{DQLTEEQIAEFK}^{13}$. A full series of +12 Da modified type 3 b and y ions and a type 1 ion confirmed $^1\text{GIGAVLK}^7$ with a modification either at G1 or K7. The lack of MS/MS evidence for modified calmodulin K13 suggested that melittin G1 or K7 was modified and cross-linked to calmodulin Q3 or Q8. It is possible that both reactions occurred, producing structural isomers with different cross-linking sites. If cross-linking within calmodulin is assumed to be uniform across major reaction products, it is unlikely that calmodulin residues occupied in cross-links to other calmodulin residues will form cross-links with melittin residues. Cross-linking between G1 or K7 and Q8 is most likely since Q3 formed a cross-link to K77 within calmodulin. However, hypothesis must be further validated. Nevertheless, type 3 ion Iib8+12 provides evidence that Q8 was involved in the cross-link formation. Stronger signals for modified G1 in comparison to K7 suggested that cross-linking between melittin G1 and calmodulin Q8 represented the major product. To determine whether G1 or K7 represented the major modification site, the degree of modification (DOM) was calculated, as described in section 2.6.2.

The plot of the DOM of $^1\text{GIGAVLK}^7$ (Figure 4.28) revealed a drastic increase in modification for Iib2 and a negligible rise in DOM for the y ions, supporting that G1 was the main site for the modification. This is consistent with the observation of +12 Da modified melittin G1 in the PFA modified peptides identified in chapter 3.

4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

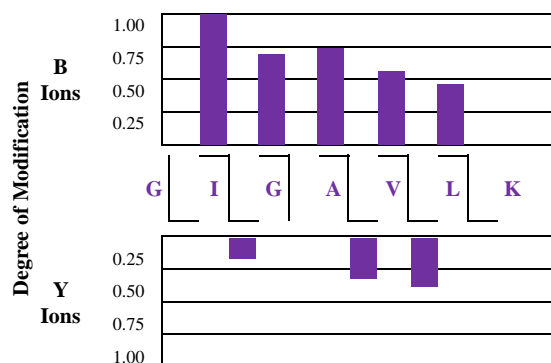


Figure 4.28: Degree of Modification: Bar graph depicting the DOM of each b ion and y ion for $^1\text{GIGAVLK}^7$ in cross-linked species m/z 744.73 ($z = 3$)

Proposed structures of the species with a charge of five and m/z at 484.46 and 588.52 are $^{91}\text{VFDKDGNGYISAAELR}^{106}\wedge^1\text{GIGAVLK}^7$ (Figure 4.32) and $^{87}\text{EAFR}^{(+12)}\text{VFDKDGNGYISAAELR}^{106}\wedge^1\text{GIGAVLK}^7$ plus an additional Schiff base modification (Figure 4.33), respectively. For the species with a charge of five and m/z at 484.46, the calmodulin peptide, $^{91}\text{VFDKDGNGYISAAELR}^{106}$, was confirmed with the type 3 ions Ib2, Ib4, Ib5, Ib8 and Ib9 and the melittin peptide $^1\text{GIGAVLK}^7$ was confirmed with a +12 Da modified type 1 ion and type 3 ions Iib2 to Iib5, Iiy1 to Iiy3, and Iiy5. For the species with a charge of five and m/z at 588.52, the calmodulin peptide $^{87}\text{EAFR}^{(+12)}\text{VFDKDGNGYISAAELR}^{106}$ was confirmed with type 3 ions Ib2 to Ib10 and Ib13 and the melittin peptide $^1\text{GIGAVLK}^7$ was confirmed with a +12 Da modified type 1 ion and type 3 ions Iib2 to Iib6, Iiy1 to Iiy3, and Iiy5. The species with m/z at 588.52 species differs from the species with m/z at 484.46 only in that the calmodulin peptide has an additional missed cleavage site and there is an extra modification. Therefore it is likely that the modification occurred at the missed cleavage site, R90. This is supported by the series of Ib4 to Ib12 ions (with only Ib11 + 12 missing) modified with a +12 Da mass in the MS/MS spectra of the species with a charge of five and m/z at 588.52. Possible cross-linking mechanisms include: (1) melittin G1 or K7 was modified and cross-linked to calmodulin N97, Y99 or R106; (2) calmodulin

4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

K94 or R106 was modified and cross-linked to melittin G1. The presence of a modified type 1 ion for the melittin peptide supports that $^1\text{GIGAVLK}^7$ was modified first and cross-linked to calmodulin $^{91}\text{VFDDKDGNGYISAAELR}^{106}$. Modified fragment ions (Iib2+12 to Iib5+12) support a modified G1. Weaker signals for Ily1+12 and Ily2+12 fragment ions support a modified K7. Similar to the triply charged species at m/z 744.73 discussed above, DOM plots localized the +12 Da modification to G1 for both cross-linked species at m/z 484.46 and 588.52 (see Figure 4.29 and 4.30, respectively).

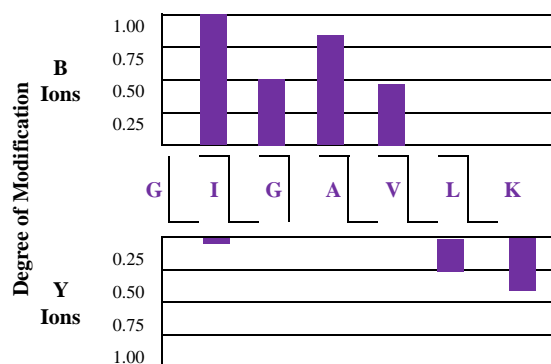


Figure 4.29: Degree of Modification: Bar graph depicting the DOM of each b ion and y ion for $^1\text{GIGAVLK}^7$ in cross-linked species m/z 484.46 ($z = 5$)

4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

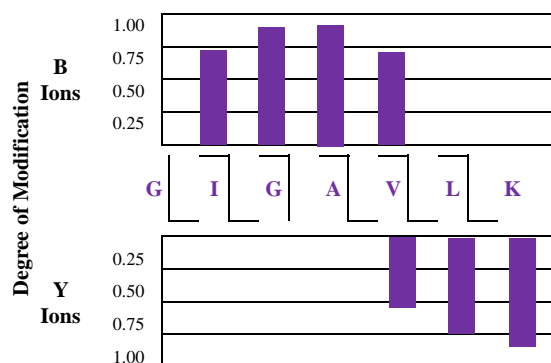


Figure 4.30: Degree of Modification: Bar graph depicting the DOM of each b ion and y ion for $^1\text{GIGAVLK}^7$ in cross-linked species m/z 588.52 ($z = 5$)

In the MS/MS spectra of the species at m/z 484.46, Iy8 +12 and Iy9+12 ions suggest that Y99 is likely to be involved in a cross-link. Similar to the MS/MS spectrum of the species at m/z 588.52, Iib2+12 to Iib5+12 support a modified G1 and weaker signals for Iiy1+12 and Iiy2+12 fragment ions support a modified K7. Assuming that cross-linking sites are consistent among most cross-linked complex molecules involving similar regions, pooling information from the MS/MS spectra of both species at m/z 484.46 and 588.52 indicates that cross-linking occurred primarily between G1 and Y99 with an extra modification on R90. This is consistent with modified peptides identified in chapter 3 with +12 Da modifications localized on melittin G1 and calmodulin R90.

Species with a charge of four at m/z 730.65 corresponded to the mass of the following structure: $^{107}\text{HVM}^{(ox)}\text{TNLGEK}^{(tm)}\text{LTDEEVDEM}^{(ox)}\text{IR}^{126}\wedge^{22}\text{RKR}^{24}$ with an additional mass of +48 Da. Figure 4.34 displays the proposed structure with the fragment ion evidence and Figure 4.35 shows the MS/MS spectrum. Only type 1 ions corresponding to each peptide component with very few type 3 ions matching $^{107}\text{HVM}^{(ox)}\text{TNLGEK}^{(tm)}\text{LTDEEVDEM}^{(ox)}\text{IR}^{126}$ were identified in the MS/MS spectrum. Type 1 ion I+60 suggests that the +12 Da bridge along with an additional +48 Da modification was retained on peptide $^{107}\text{HVM}^{(ox)}\text{TNLGEK}^{(tm)}\text{LTDEEVDEM}^{(ox)}\text{IR}^{126}$. Type 1 ion I(ox)-H₂O and

4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

Type 3 (Iy3(ox) to Iy6(ox) and Ib4(ox)+12) ions suggest that M residues were oxidized. The +48 Da corresponds to the mass of three oxygens which can be explained by one methionine oxidized to methionine sulfoxide (+16 Da) and the other oxidized to methionine sulfone (+32 Da) [164]. Since K115 was blocked in the calmodulin peptide, cross-linking could have occurred with the following possible mechanisms: (1) melittin R22, K23 or R24 was modified and cross-linked to calmodulin H107, N111 or R126; (2) calmodulin R126 was modified and cross-linked to melittin R22 or R24. Type 2 ion Iy8(ox)+II+12 suggests that R126 is cross-linked and Iy14(2ox)+IIy1+12 suggests that R126 is connected to R24. This is consistent with with +36 Da (3 Schiff Bases) localized on melittin ²¹KRKR²⁴ in modified peptides identified in chapter 3.

Table 4.9: PFA calmodulin-melittin interpeptide cross-linked species are listed and classified as capturing antiparallel (shaded in blue) or parallel (white) binding. Cross-linking sites are highlighted in red.

| Cross-Linked Species | | | | | Melittin Peptide | | Calmodulin Peptide | | Total PFA Modification n Mass |
|----------------------|---|----------------------------|-----------------------------|------------------------|------------------|-----------------------------------|--------------------|--|-------------------------------------|
| m/z | z | [M] _{exp} (Da) | [M] _{Calc} (Da) | Mass Accuracy (ppm) | [M] (Da) | Sequence | [M] (Da) | Sequence | |
| 744.73 | 3 | 2231.19 | 2231.17 | 9 | 656.42 | ¹ GIGAVLK ⁷ | 1562.75 | ¹ A(ac)DQLTEEQIAEFK ¹³ | 12 |
| 730.65 | 4 | 2918.60 | 2918.43 | 59 | 458.31 | ²² RKR ²⁴ | 2400.12 | ¹⁰⁷ HVM(ox) TNLGEK(tm)LTDEEVDEM(ox) IR ¹²⁶ | 12 |
| 484.46 | 5 | 2422.30 | 2422.29 | 6 | 656.42 | ¹ GIGAVLK ⁷ | 1753.86 | ⁹¹ VFDKDGNGYISAAELR ¹⁰⁶ | 12 |
| 588.52 | 5 | 2937.60 | 2937.53 | 22 | 656.42 | ¹ GIGAVLK ⁷ | 2257.11 | ⁸⁷ EAFRVFDKDGNGYISAAELR ¹⁰⁶ | 24 |

4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

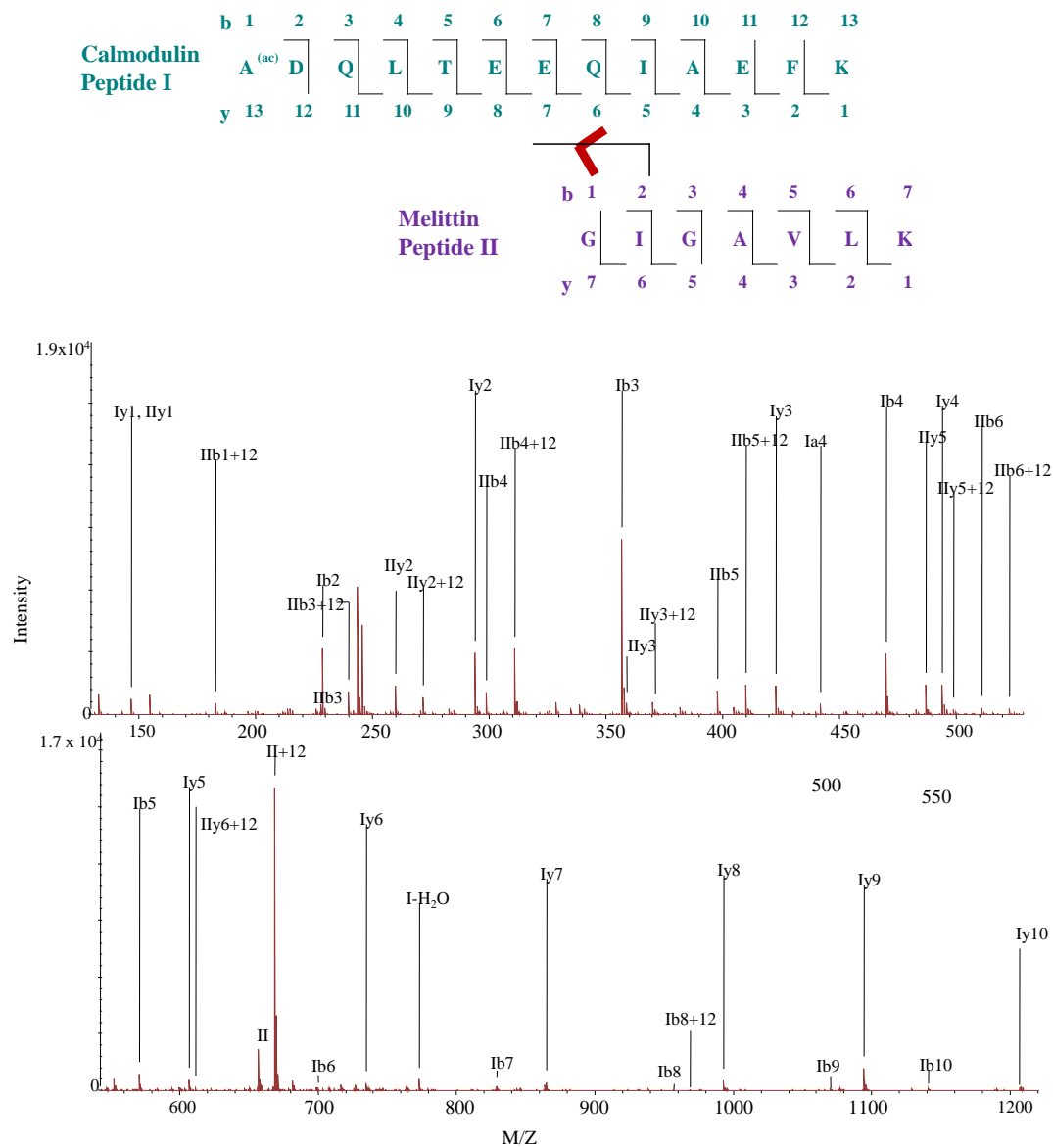


Figure 4.31: Interpeptide PFA calmodulin-melittin cross-link m/z 744.73 ($z = 3$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. Note: Fragmentation indicated on the backbone of the peptide corresponds to type 3 ions only

4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

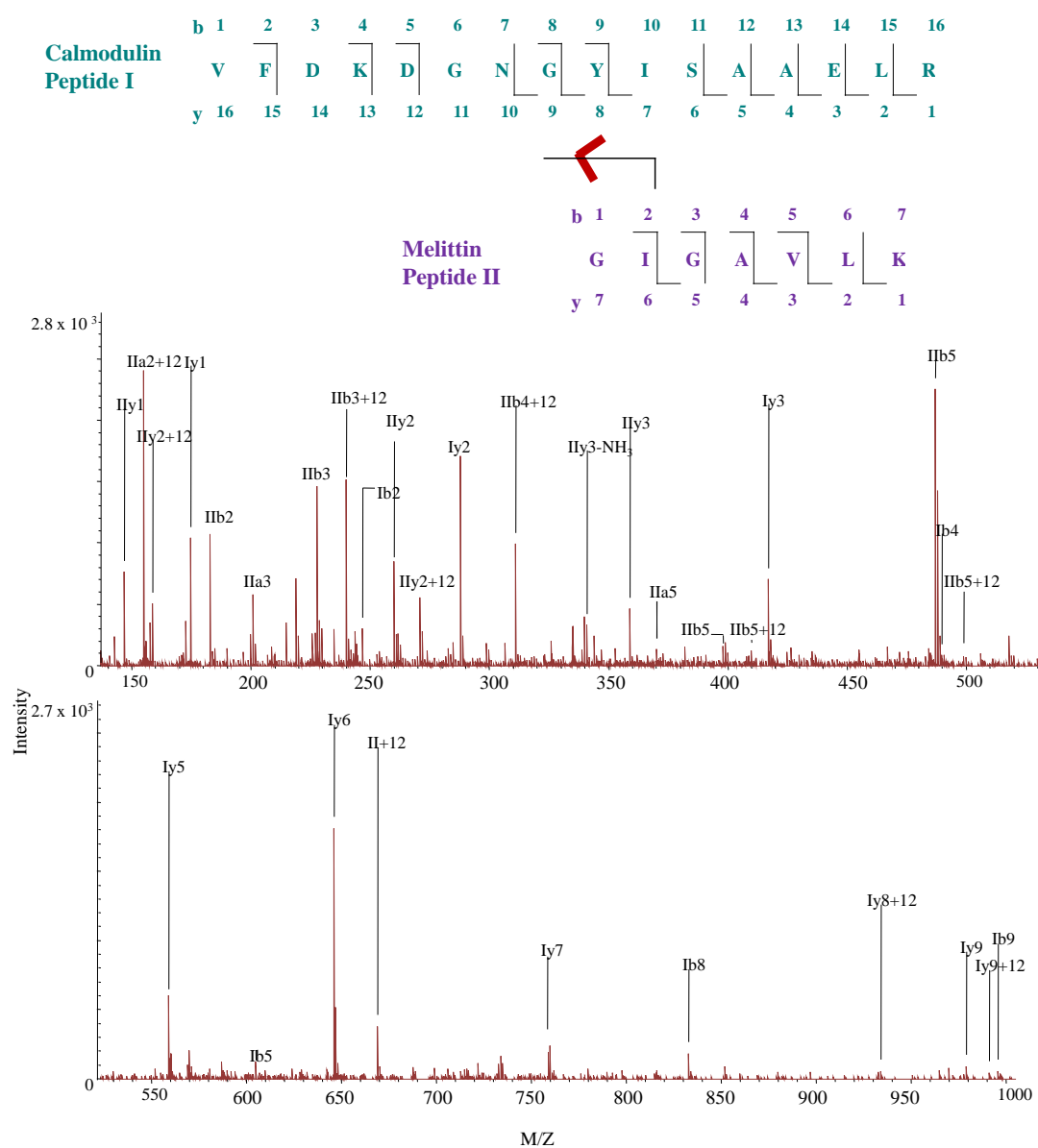


Figure 4.32: Interpeptide PFA calmodulin-melittin cross-link m/z 484.46 ($z = 5$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. Note: Fragmentation indicated on the backbone of the peptide corresponds to type 3 ions only

4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

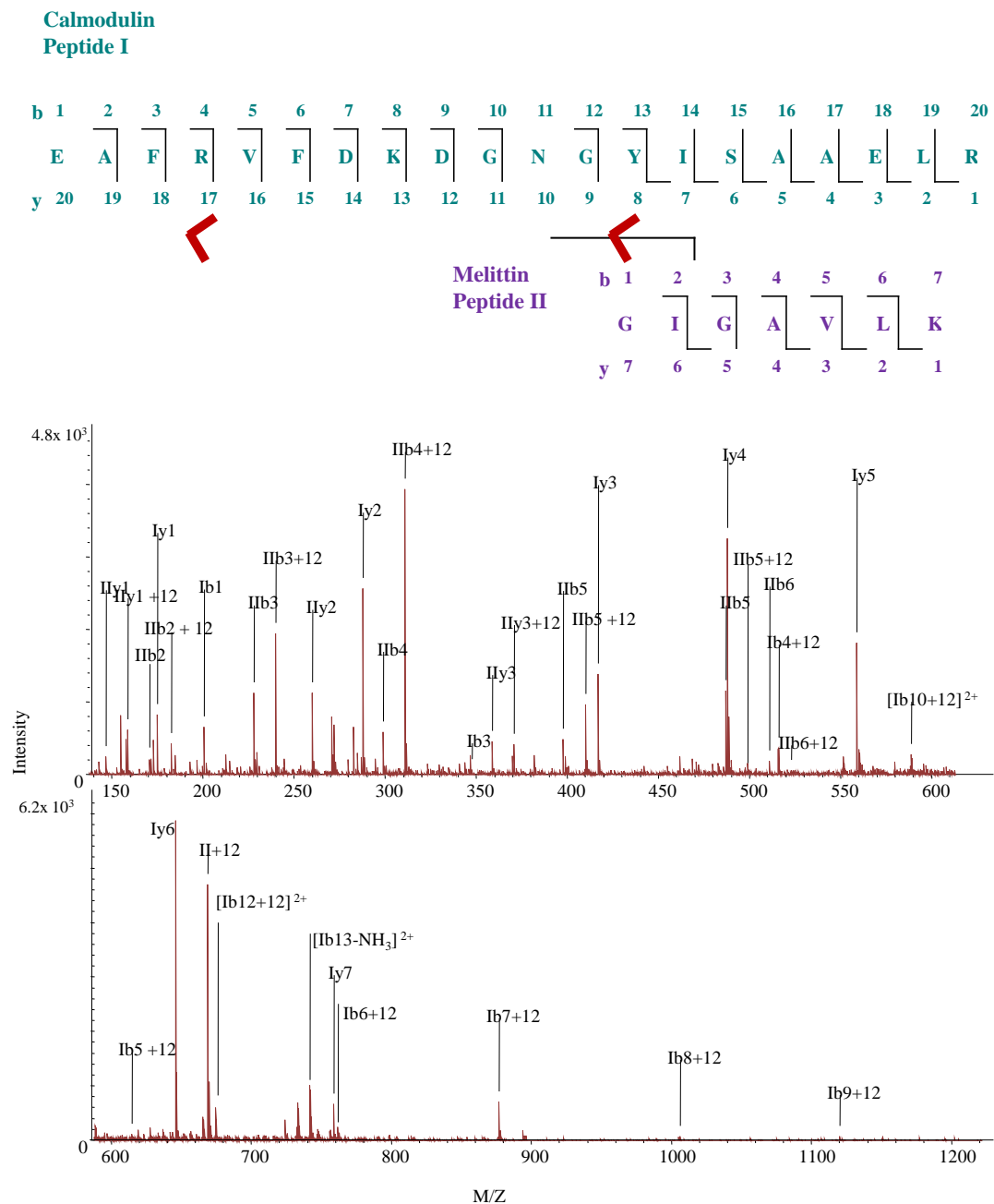
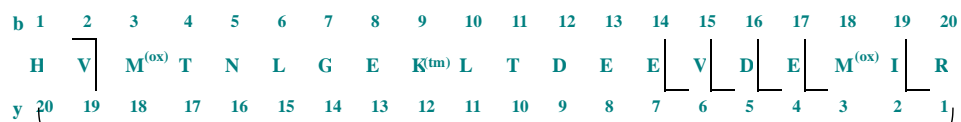


Figure 4.33: Interpeptide PFA calmodulin-melittin cross-link m/z 588.52 ($z = 5$) proposed structures with fragment ion evidence (top) and MS/MS spectra (bottom); Cross-linker bridges are indicated in red. Note: Fragmentation indicated on the backbone of the peptide corresponds to type 3 ions only

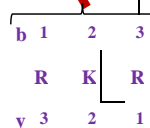
4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

Type 1 and 3 ions

Calmodulin Peptide I



Melittin Peptide II



Type 2 ions

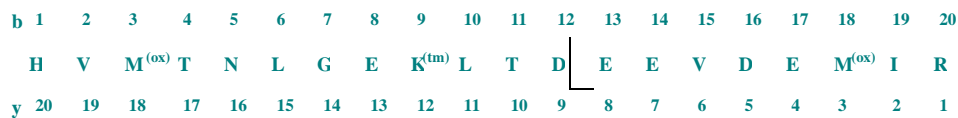
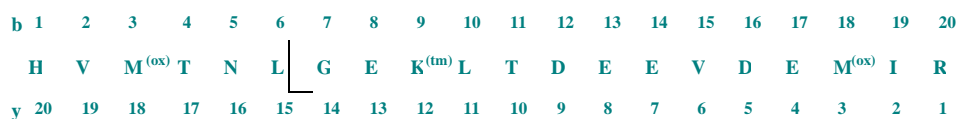


Figure 4.34: Interpeptide PFA calmodulin-melittin cross-link m/z 730.65 ($z = 4$) proposed structures with fragment ion evidence; Cross-linker bridges are indicated in red.

4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

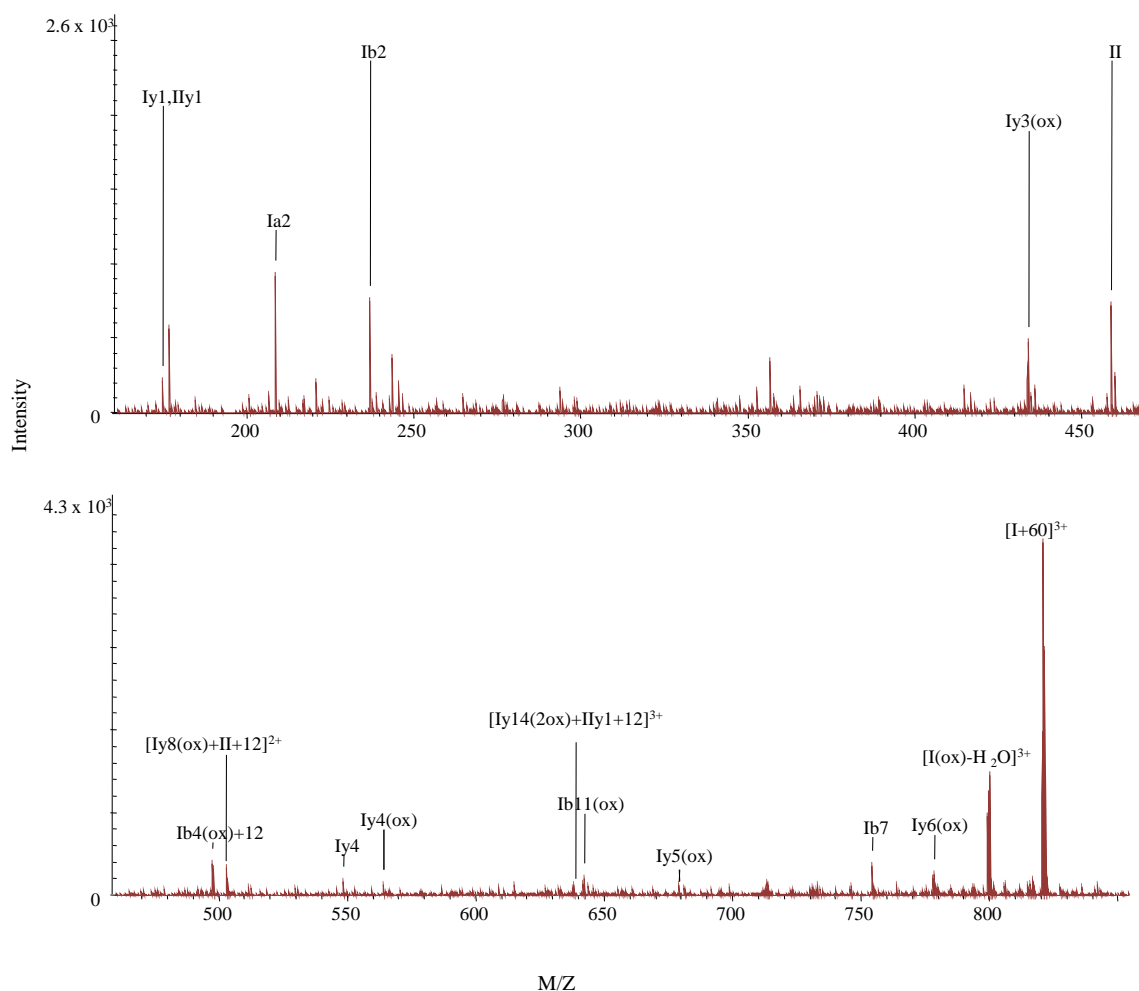


Figure 4.35: Interpeptide PFA calmodulin-melittin cross-link m/z 730.65 ($z = 4$) MS/MS spectra; Cross-linker bridges are indicated in red.

The labile nature of PFA's cross-linker bond was illustrated in MS/MS of the cross-linked species $^{107}\text{HVM}^{(\text{ox})}\text{TNLGEK}^{(\text{tm})}\text{LTDEEVDEM}^{(\text{ox})}\text{IR}^{126\wedge 22}\text{RKR}^{24}$ and $^{75}\text{KM}^{(\text{ox})}\text{KDTDSEEEIREAFR}^{90\wedge 1}\text{A}^{(\text{ac})}\text{DQLTEEQIAEFK}^{13}$, in which exclusive fragmentation at the cross-linker bridge generated fragment ions corresponding to intact peptide components. Although both type 3 and type 2 ion evi-

4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

dence was not relatively abundant, high intensity signals pertaining to component peptides and modified component peptides in the MS/MS spectra were deemed as sufficient to confirm cross-linking.

Type 2 ions were observed only for cross-linked species $^{107}\text{HVM}^{(ox)}\text{TNLGEK}^{(tm)}\text{LTDEEVDEM}^{(ox)}\text{IR}^{126\wedge 22}\text{RKR}^{24}$. It is hypothesized that this species required a higher CID energy to fragment both the cross-link bridge and backbone. Thus this cross-linked species produced type 2 ions and mainly type 1 ions at the lower CID energy utilized.

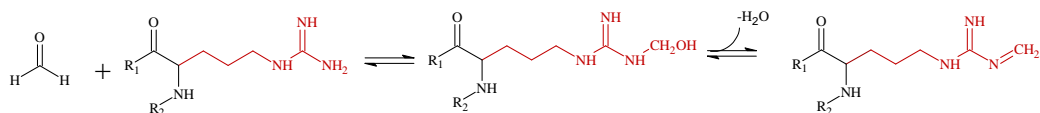
The PFA bridge fragmented such that the peptide that was modified in the first step of the reaction retained the +12 Da mass shift on its subsequent fragment ions. Since the specific subset of amino acids that can form modifications is known, this could potentially clarify the cross-linking mechanism for the elucidation of cross-linking sites. The exception was the cross-linking observed between melittin R24 and calmodulin R126, suggesting that this trend is also a function of the identity of the amino acids involved in the cross-link that would dictate the bond energy of the cross-linking bridge. Therefore, MS/MS spectra of cross-linked species from more protein complexes with different cross-linking sites are required to explore this trend.

The cross-linking sites localized are consistent with the modified peptides identified in chapter 3. Peptides with a +12 Da modification localized on calmodulin K77 and melittin G1, and with +36 Da (3 Schiff Bases) localized on melittin $^{21}\text{KRKR}^{24}$ confirm that these residues/regions were modified in the first step of the reaction. In addition, peptides with a +12 Da modification on R90 was also observed and this is consistent with the extra modification localized on this site in the cross-link $^{87}\text{EAFR}^{(+12)}\text{VFDKDGNGYISAAELR}^{106\wedge 1}\text{GIGAVLK}^7$. Figure 4.36 depicts the proposed reaction mechanisms for the formation of cross-linking of melittin R24 to calmodulin R126 (a), calmodulin K77 to Q3 (b), melittin G1 to Y99 (c) and melittin G1 to calmodulin Q8 (d).

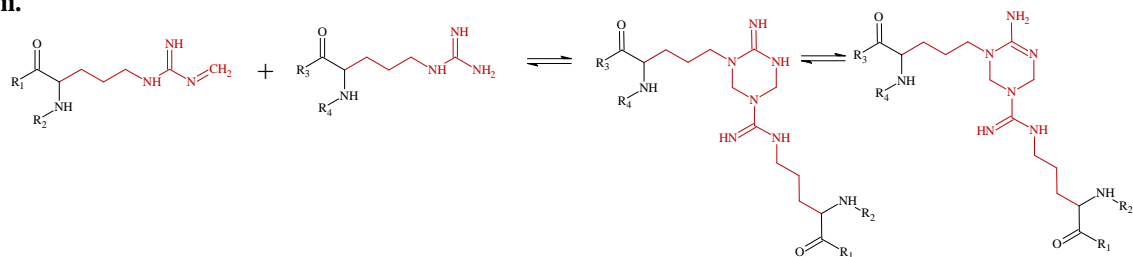
4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

(a)

i.

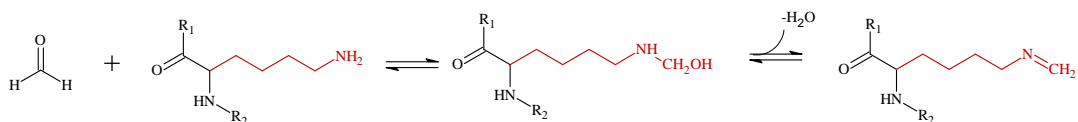


ii.

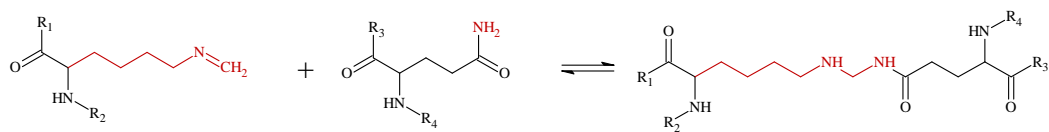


(b)

i.

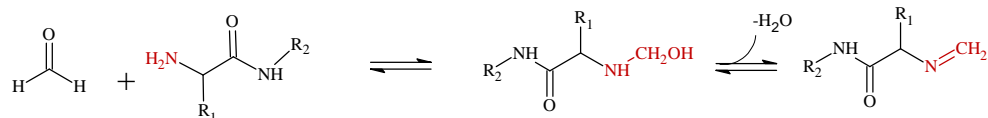


ii.

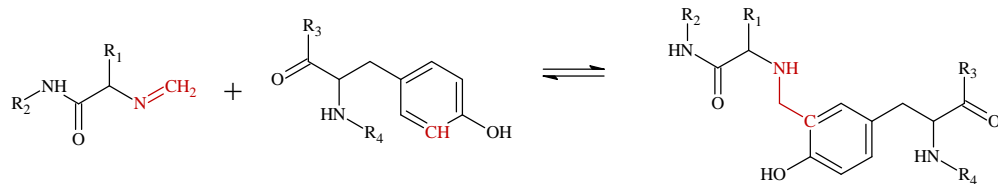


(c)

i.



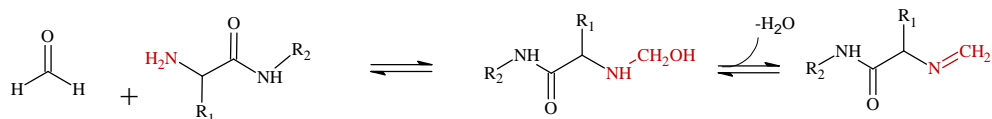
ii.



4.3. Tandem Mass Spectrometric Fragmentation of Formaldehyde

(d)

i.



ii.

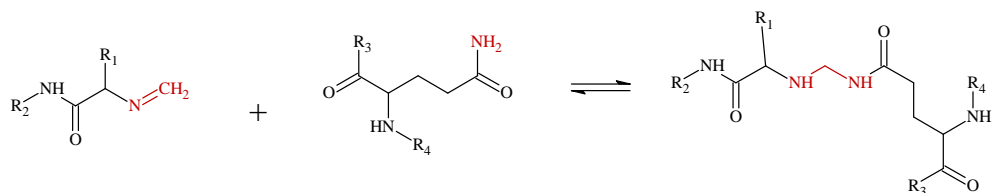


Figure 4.36: Reaction mechanisms of the PFA modification (i) and cross-linking formation (ii) of melittin R24 to calmodulin R126 (a), calmodulin K77 to Q3 (b), melittin G1 to calmodulin Y99 (c) and melittin G1 to calmodulin Q8 (d); Reactive regions are highlighted in red. R_1 and R_2 , and R_3 and R_4 , represent arbitrary sections of the modified and cross-linked proteins, respectively.

PFA cross-linking of calmodulin K77 to Q3 and of melittin G1 to calmodulin Q8 represent the first interpeptide PFA cross-links identified in proteins involving glutamine, and PFA cross-linking between melittin R24 and calmodulin R126 represent the first interpeptide PFA cross-linking identified between two R residues in proteins under mild *in vivo*-like reaction conditions. These were previously only observed with long and extensive PFA cross-linking conditions using model peptides and small proteins [86, 87, 90–92]. PFA cross-linking between melittin G1 and calmodulin Y99 is consistent with previous studies that observed cross-linking between these amino acids in small model proteins under mild *in vivo*-like reaction conditions [50]. In general, this is the first study that identified PFA cross-linking within a non-covalent protein complex.

4.4 Formaldehyde versus other Cross-linker Fragmentation

Placing the observed fragmentation patterns of PFA cross-linked species in context to those of other established cross-linkers in this study aided in establishing fragmentation rules to confirm the presence of PFA cross-linked species. With other cross-linkers, type 1 fragment ions (i.e. fragment ions corresponding to the intact peptide component) for the smaller peptide component was most often detected and type 1 evidence of none or both components was unusual, which was also observed with PFA cross-linking. Type 3 ions (i.e. fragment ions from the fragmentation of the crosslinker bridge and peptide backbone) were rarely observed and type 2 ions (i.e. fragment ions from the fragmentation of the peptide backbone) with the cross-linker intact facilitated the majority of the cross-linking identification with other cross-linkers. In contrast, type 2 ions were uncommon for PFA cross-linked species, and type 3 ions were used instead to verify cross-linked structures. Therefore, type 1 ion evidence of at least one component peptide along with type 3 ion evidence of the second component peptide was accepted as the minimum cross-linking evidence required to confirm a candidate PFA cross-linked species detected in the MS spectrum. This is based on studies of fragmentation patterns of other cross-linkers, which established that it is common to observe extensive backbone fragmentation of only the larger peptide component in a cross-linked species [51, 53]. The lack of type 2 evidence was justified by PFA's labile cross-linker bond under CID fragmentation that has been previously reported to fragment simultaneously with the peptide backbone [50]. Diagnostic ions were not observed and PFA's cross-linker bridge is too small to generate fragments within the linker unlike the large NHS ester cross-linkers.

In general, the evidence for localizing PFA cross-linking is much more ambiguous than with other cross-linkers, especially since multiple structural isomers are possible. This was particularly evident in cross-links involving ¹GIGAVLK⁷, where evidence supported cross-linking at both G1 and K7. However, DOM calculations consistently proved that G1 was the major cross-linking site across all these PFA calmodulin-melittin cross-links. Furthermore, all PFA cross-links within calmodulin involved the same regions. Despite the diverse reactivity of PFA, the

4.5. *General Criteria for Evaluating Tandem Mass Spectrometric Patterns of Cross-linked Species*

consistency in cross-link site localization observed across different cross-linked peptides illustrates that PFA has the potential to capture specific interactions.

4.5 General Criteria for Evaluating Tandem Mass Spectrometric Patterns of Cross-linked Species

Typically, cross-link identification software derives scores for cross-linked species based on the percent of expected backbone fragment ions present (i.e. the number of expected y and b ions divided by the total length of the peptide)[120, 165, 166]. Therefore a similar approach was used to summarize the general criteria for confirming cross-linking manually. It was observed that the larger peptide extensively fragmented while the shorter peptide tended to remain intact under CID fragmentation, which is consistent with previous observations of cross-link fragmentation in literature[51, 53]. Nevertheless, the appearance of two intact peptide ion MS/MS signals from the exclusive cross-link bridge fragmentation proves that two peptides exist and along with their combined mass plus cross-linker bridge mass matching the mass of the cross-linked species, this was also considered sufficient evidence to confirm cross-linking. Therefore, candidates that did not produce fragment ion evidence that met any one of the requirements listed below were not considered as crosslinked species:

(1) > 20% of the expected y and b ions from the backbone fragmentation (type 2) or backbone plus cross-linker fragmentation (type 3) of each peptide component;

(2) > 20% of the expected y and b ions from the backbone fragmentation (type 2) or backbone plus cross-linker fragmentation (type 3) of one peptide component plus type 1 ions of the other peptide component from the exclusive cross-linker bridge fragmentation;

(3) type 1 ions of both peptide components from the exclusive cross-linker bridge fragmentation;

In addition, cross-linking between adjacent peptides were classified as single peptides with missed cleavages and not crosslinked species. Overall, this provides objective guidelines to evaluate cross-linking when interpreting the MS/MS of candidate cross-linked species manually.

4.6 A Second Look at Trypsin Digestion of Cross-linked Residues

In chapter 3, it was shown that trypsin cleavage did not occur after PFA modified residues due to the lack of modifications localized on terminal K and R residues i.e trypsin cleavage sites. Theoretically, since PFA cross-linking does not affect the electrophilicity of K and R residues, and forms small cross-link bridges that could potentially fit in the active site of trypsin, the possibility of trypsin cleaving after PFA cross-links was not excluded. Using MS/MS, cross-linking was localized on terminal calmodulin K77 and R126 and melittin R24 in the following PFA cross-linked species: $^{75}\text{KM}^{(ox)}\text{K}^{77}$ and $^1\text{A}(\text{ac})\text{DQLTEEQIAEFK}^{13}$ and $^{107}\text{HVM}^{(ox)}\text{TNLGEK}^{(tm)}\text{LTDEEVDEM}^{(ox)}\text{IR}^{126\wedge 22}\text{RKR}^{24}$. These findings suggest that trypsin can potentially cleave after PFA cross-linked residues.

Based on the peptide bond hydrolysis mechanism via trypsin and previous literature [34, 38, 152], it was concluded that cross-linked K residues formed by EDC/sulfoNHS, sulfoDST, BS³, sulfoEGS would not be cleaved. However, trypsin cleavage was observed after BS³ cross-linked K77. It is hypothesized that the adjacent oxidized methionine's electron withdrawing nature increased the partial positive charge on the carbonyl carbon connected to the cross-linked lysine to favor trypsin cleavage. Cleavage after sulfoDST and EDC cross-linked K30 was also observed. Whether the lack of a missed cleavage is sufficient enough to eliminate these cross-linked species when MS/MS provides evidence of cross-linking is unknown. To clarify this, it may be useful to explore trypsin cleavage efficiency after cross-linked residues with various microenvironments.

4.7 MS/MS Analysis of Formaldehyde Cross-linked Ribonuclease-S

Although the MS/MS verification of PFA cross-linked species was successful in the camodulin-melittin system, similar analysis could not be performed for the RNaseS system. The limitations associated with the PFA cross-linked RNaseS system are discussed in chapter 6. Nevertheless, insight into the structure and stabilization of the PFA treated RNaseS complex is examined in the next chapter.

Chapter 5

Structural Characterization of Calmodulin-Melittin and Ribonuclease S Cross-linked Species

5.1 Trypsin Cleavage and Accessibility of Residues

5.1.1 Comparing Cross-linker Reagent versus Literature Trypsin Accessibility in Calmodulin-Melittin

Examining trypsin cleavage sites in calmodulin-melittin can provide insight into the conformation of calmodulin and melittin, since the extent of trypsin cleavage at a particular site conveys the extent of the site's exposure to the trypsin reagent, which is dictated by the conformation of the protein. Upon binding, both calmodulin and melittin undergo a change in conformation and would thus differ in reagent accessibilities in their bound versus unbound state. In this study, the Ca^{2+} -free calmodulin-melittin structure is analyzed via cross-linking. NMR and fluorescence experiments in literature, have shown that both Ca^{2+} -free calmodulin-melittin and Ca^{2+} -saturated calmodulin-melittin complexes exhibit similar conformations [95]. Therefore structural characteristics are comparable between both complexes.

A previous limited proteolysis experiment examining a Ca^{2+} -saturated calmodulin-melittin system revealed that major trypsin cleavage sites were R37, R74, K75, K77 and R106 in unbound calmodulin. Upon complex formation, only K77 and to a smaller extent K75 and R74 were trypsin cleavage sites in calmodulin [102]. This suggests that K77 is accessible in the complex conformation of

5.1. Trypsin Cleavage and Accessibility of Residues

calmodulin, allowing access to cross-linking reagents. This is consistent with the observation in this study of cross-linking at K77 with PFA, BS³ and sulfoEGS, i.e. cross-linkers that vary in size and membrane permeation ability, supporting that K77 was accessible also in these cross-linked structures. Similarly, the limited proteolysis experiment in literature showed that for melittin, cleavage sites K7, K21, R22, K23 and R24 in the unbound molecule became inaccessible in melittin upon binding to calmodulin except for R24. PFA cross-linking was observed at R24, which is adjacent to K23, a major cross-linking site for other cross-linkers in this present study, which also supports the accessibility of this region in the cross-linked structures. Furthermore, the limited proteolysis experiment in literature concluded that terminal amino acids in melittin and the flexible loop (specifically, residues 76-78 and 87-92) in calmodulin were most accessible in the complex conformation. This may explain why only terminal segments of melittin (1-7 and 22-24) were observed in all identified cross-linked species and further supports extensive cross-linking localized within this flexible loop (K75 and K77) of calmodulin in this study. Finally, it was demonstrated that calmodulin E54 interacts with the melittin C-terminus in the limited proteolysis experiment in literature, which is consistent with EDC cross-linking observed between calmodulin E54 and melittin K23 in this study [102]. All together, the accessibility of the calmodulin-melittin complex conformation observed in the limited proteolysis experiment in literature agrees with the cross-linking structures observed in this study. This supports that cross-linking potentially captured the binding of calmodulin to melittin.

5.1.2 Comparing Trypsin Accessibility in Control versus PFA Treated Calmodulin-Melittin

It is hypothesized that in the absence of stabilizing cross-linking, the transient calmodulin-melittin interaction would not be preserved. In other words, it is expected that control samples would contain calmodulin and melittin molecules in the unbound state and cross-linked samples would contain calmodulin-melittin molecules in the bound state conformation. It is important to note that in this experiment, trypsin digestion was performed in gel overnight after denaturing proteins with SDS whereas the trypsin digestion in the limited proteolysis experiment

5.1. Trypsin Cleavage and Accessibility of Residues

in literature was performed in solution without denaturing the proteins over a 15 min to four hour time course [102]. Therefore trypsin cleavage sites observed in the literature experiment may not be comparable to trypsin cleavage sites in this experiment. Nevertheless, for cross-linked samples, it was observed that the SDS was inefficient in denaturing cross-linked proteins. This was supported by the proteins migrating further down the gel than expected, suggesting that the protein was not fully folded. However, control protein gel bands migrated to expected positions, suggesting that the protein was efficiently denatured by SDS. Therefore, it is hypothesized that trypsin cleavage would vary between control and cross-linked samples based on differing accessibility of the structures captured in these samples.

The trypsin cleavage sites observed in the control vs PFA cross-linked sample was compared. In the list of identified unmodified peptides originating from control samples (see Tables 3.1), trypsin cleavage was observed for calmodulin at K13, K21, K30, R37, R74, K75, K77, R86, R90, K94, R106, and R126 and for melittin at K7, K21, and R22 suggesting that these residues were accessible. Interestingly, all residues that were shown to be accessible in the unbound conformation of calmodulin and melittin in literature [102] were also trypsin cleavage sites in the control sample, supporting that the transient interaction between calmodulin and melittin was not preserved in the control sample after SDS denaturing. This is consistent with section 3.3.1.1, where control peptides were identified in a single calmodulin molecule (~17 kDa) or melittin tetramers (~ 6 kDa), suggesting that calmodulin and melittin exist as separate species in the control sample.

Assessing the accessibility in cross-linked calmodulin-melittin samples was more complicated since variations in the extent and type of cross-linking can occur that may preserve the calmodulin-melittin interaction differently. Upon PFA cross-linking of calmodulin-melittin, SDS-PAGE separation, and the subsequent in-gel trypsin digestion, all intramolecular calmodulin cross-links and the intramolecular calmodulin-melittin cross-link at m/z 730.65 ($z = 4$) were identified in 14-19 kDa proteins and the intramolecular calmodulin-melittin cross-links at m/z 744.73 ($z = 3$), 484.46 ($z = 5$) and 588.52 ($z = 5$) were identified in 19-33 kDa proteins via MS/MS (see chapter 4, Tables 4.8 and 4.9). Although the SDS-PAGE provided evidence for intermolecular cross-linking between multiple calmodulin or calmodulin-melittin complex molecules (> 33 kDa), no cross-linked peptides

5.1. Trypsin Cleavage and Accessibility of Residues

were confirmed via MS/MS in their respective gel bands. Regardless, PFA unmodified and modified peptides were identified in proteins from all three molecular weight categories via MaxQuant (see section 3.3.1). It is hypothesized that the intramolecular cross-linked proteins (14-33 kDa) would be stabilized and retain or partially retain the calmodulin-melittin binding structure. Since no cross-linked peptides were identified in proteins > 33 kDa, it is unknown whether only intermolecular cross-links formed or whether intramolecular cross-links preserving the calmodulin-melittin interaction also formed. In addition, it is possible that the 14-19 kDa molecular weight protein gel band also contained unmodified calmodulin (~17 kDa) or calmodulin with only modifications and no cross-links. These proteins are hypothesized to not retain structural properties of calmodulin-melittin, similar to the control sample. Nevertheless, 14-19 kDa may also contain calmodulin-melittin crosslinked species, which would preserve their binding structure, since distinguishing between the 19 kDa cut-off and 19.6 kDa (i.e. the mass of a calmodulin-melittin complex) is difficult on the SDS-PAGE gel.

To gain insight into whether cross-linking preserved or partially preserved cross-linked structures in all three molecular weight categories, the accessibility was examined via trypsin cleavage. Using the PFA unmodified and modified peptides identified in chapter 3 (from Tables 3.2, and 3.3) and the PFA cross-linked peptides identified in chapter 4 (from Tables 4.8 and 4.9), the percent trypsin cleavage of PFA treated calmodulin-melittin samples was compared to the control calmodulin-melittin sample.

The percent trypsin cleavage for calmodulin and melittin in control vs PFA treated samples was calculated for each cleavage site by dividing the abundance (i.e. normalized MS signal peak areas) of peptides that supported cleavage at a particular site by the abundance of all peptides containing the particular site (with or without a missed cleavage). This was performed for unmodified, modified and cross-linked peptides identified in the PFA treated sample and the unmodified peptides identified in the control sample. It was assumed that no significant loss of peptides occurred prior or during the MS detection and identification, and that normalized peak areas of identified peptides accounted for the variation in MS-reponse. Since a modification at a trypsin cleavage site may hinder cleavage, missed cleavage sites that contained a PFA modification or a trimethyl group (i.e. calmodulin

5.1. Trypsin Cleavage and Accessibility of Residues

K115) were disregarded to observe trypsin cleavage only as a function of protein conformation and not modification. In melittin, the position of the PFA modifications were ambiguous for the segment ²¹KRKR²⁴, which appeared with +24 and +36 Da modifications. Since the +24 (2 Schiff Base modifications) and +36 (3 Schiff Base modifications) Da modified peptides exhibited three and four missed cleavages, respectively, one of the missed cleavages in each of the peptide was not a result of the PFA modification. Thus, one missed cleavage site was accounted for this peptide segment.

As shown in Table 5.1, there was a decrease in percent trypsin cleavage at calmodulin K75, K77, R86, R90, and K94, and melittin ²¹KRKR²⁴ in all peptides identified in the PFA treated samples versus the control sample. This is consistent with the reduced trypsin cleavage at these sites observed in bound calmodulin in literature. However, bound calmodulin is also expected to experience reduced trypsin cleavage at calmodulin R37, R106, and to a smaller extent, R74, and melittin K7 [102]. In contrast, a 100% trypsin cleavage was observed at these sites in all three PFA treated samples in this study. All together, it suggests that PFA cross-linking partially preserved the structure of calmodulin-melittin post the SDS-PAGE denaturation process. In the gel bands with 19-33 kDa proteins, only calmodulin-melittin cross-links were identified. Thus, it was expected that this sample would provide the most accurate picture of whether PFA cross-linking between calmodulin-melittin preserved its binding structure even after SDS-PAGE denaturation. Since these trends in percent trypsin cleavage were fairly consistent across peptides originating from proteins of all three molecular weight categories, it suggests that similar structures were preserved with different types of PFA cross-linking. Nevertheless, this also supports that PFA intermolecular cross-linked species (> 33 kDa proteins) that likely formed may have been lost in the sample processing or detection and thus were not identified. This may have caused a slight discrepancy in the percent trypsin cleavage at calmodulin K77, where > 33 kDa proteins did not exhibit as great decrease in trypsin cleavage as 14-33 kDa proteins treated with PFA. The peptides originating from 14 - 19 kDa proteins in the PFA treated samples matched the percent trypsin cleavage patterns observed in the 19-33 kDa proteins more than in the control samples. This suggested that PFA cross-linked proteins were the majority of the 14-19 kDa proteins and not PFA-

5.1. Trypsin Cleavage and Accessibility of Residues

unmodified proteins or proteins with only PFA modifications. This was further supported in section 3.2.4, where a high yield of PFA cross-linked versus non-cross linked proteins was observed. The > 19 kDa proteins did not display as great decrease in trypsin cleavage as 14-19 kDa proteins at calmodulin R90. It is possible that the 14-19 kDa proteins contained mostly intramolecular calmodulin cross-linked proteins and thus would differ from > 19 kDa proteins that most likely contained calmodulin-melittin crosslinked proteins as the major species. In conclusion, this supports that PFA cross-linking partially preserved the calmodulin-melittin structure even after SDS-PAGE denaturation, which was not retained in the absence of cross-linking in the control samples.

Table 5.1: Percent abundances of cleaved trypsin cleavage sites observed in the control and PFA treated (in 14-19, 19-33 and > 33 kDa proteins) calmodulin-melittin samples

| Trypsin Cleavage Site | Percent Abundance of Cleaved Residue | | | |
|----------------------------|--------------------------------------|---------------------|----------|---------|
| | Control Sample | PFA Treated Samples | | |
| | | 14-19kDa | 19-33kDa | >33 kDa |
| Calmodulin Residues | | | | |
| K13 | 100% | 100% | 100% | N/A |
| K21 | 100% | 100% | N/A | N/A |
| K30 | 100% | 100% | N/A | 100% |
| R37 | 100% | 100% | 100% | 100% |
| R74 | 95% | 100% | 100% | 100% |
| K75 | 50% | 23% | 44% | 40% |
| K77 | 99% | 9% | 8% | 29% |
| R86 | 71% | 33% | 23% | 43% |
| R90 | 100% | 39% | 98% | 87% |
| K94 | 6% | 0% | 1% | 0% |
| R106 | 100% | 100% | 100% | 100% |
| R126 | 100% | 100% | 100% | 100% |
| Melittin Residues | | | | |
| K7 | 100% | 100% | 100% | N/A |
| K21 to R24 | 99% | 35% | 49% | 54% |

5.1.3 Comparing Trypsin Accessibility in Control versus PFA Treated Ribonuclease S

In order gain insight into whether PFA preserved or partially preserved the RNaseS complex, even after SDS denaturation procedures prior to trypsin digestion, the percent trypsin cleavage was examined in both control and PFA treated RNaseS samples. Similar to the calmodulin-melittin analysis in section 5.1.2, the percent cleavage was calculated for each cleavage site using the peptides identified by MaxQuant in control (Table 3.5) and PFA treated (Table 3.6) RNaseS samples. The analysis was performed using only S-protein and S-peptides produced from the cleavage of RNaseA at residue 20 for consistency. Also, in order to observe trypsin cleavage as a function of accessibility and not the chemical modification of a cleavage site, sites with PFA modifications were excluded as missed cleavage sites, similar to the analysis of calmodulin-melittin. It is assumed that the majority of the peptides identified in the PFA treated RNaseS samples were digestion products of cross-linked proteins since all peptides originated from proteins > 12 kDa (i.e. greater than the molecular weight of a single S-protein or S-peptide) according to the molecular weight measured by SDS-PAGE. It is important to note that trypsin cleavage was performed in-gel after proteins were denatured with SDS and are therefore expected to be unfolded. However, whether the RNaseS interaction was diminished in the control sample and partially preserved by in PFA treated samples, as observed in the calmodulin-melittin system, was examined in the RNaseS system.

Table 5.2 lists the percent trypsin cleavage of each site observed in the control and PFA treated RNaseS samples. In control RNaseS, trypsin cleavage was observed at S-protein R39, K61, K66, R85, K91, K98, and K104, and to smaller extent K31 and R33. Almost no trypsin cleavage (7% cleaved) observed at K41 was due to the adjacent proline residue and minimal cleavage (~ 5%) observed at K37 suggested that this residue was not accessible. No trypsin cleavage was observed within the S-peptide. For the PFA treated sample, a significant decrease in percent trypsin cleavage was observed at K31, R33, and K98. Previous limited MS-based oxidation studies have shown that S-protein residues 96 -100 are blocked when bound to the S-peptide while S-protein residues 39, 85-95 and 101-

5.1. Trypsin Cleavage and Accessibility of Residues

104 remain accessible for trypsin cleavage [167]. This is consistent with the lack of trypsin cleavage observed at K98, while trypsin cleavage was observed at R39, R85 and K104 in the S-protein for PFA treated RNaseS. Therefore, this suggests that PFA cross-linking may have preserved the S-protein to S-peptide binding interaction. Other studies have shown that S-protein R33 plays a key role in stabilizing RNaseS via a salt bridge to S-peptide D14 [168], which also agrees with the reduced trypsin cleavage at R33 observed in PFA treated samples and supports the stabilization of RNaseS via PFA cross-linking. As mentioned in section 3.3.2.2, previous kinetics experiments that examined trypsin cleavage in RNaseS, demonstrated that the S-peptide is not accessible for trypsin cleavage while bound to the S-protein [160]. Thus, the absence of single S-peptide peptides in PFA treated RNaseS samples observed in this study supports the preservation of the RNaseS complex. The identification of S-peptides in the control RNaseS sample, suggests that the S-peptide was accessible and unbound to S-protein. This was also supported in section 3.3.2.1, where S-peptide and S-protein peptides were identified in < 12 kDa proteins, suggesting that they were not bound in a RNaseS complex (~13.7 kDa). Overall, these findings support that PFA preserved the RNaseS interaction, which was lost in the control sample due to the absence of cross-links.

5.2. Relative Abundance of Formaldehyde Cross-linking

Table 5.2: Percent abundances of cleaved trypsin cleavage sites observed in the control and PFA treated (> 12 kDa proteins) RNaseS samples

| Trypsin Cleavage Site | Percent Abundance of Cleaved Residue | |
|-----------------------|--------------------------------------|--------------------|
| | Control Sample | PFA Treated Sample |
| S-Protein Residues | | |
| K31 | 40% | 0% |
| R33 | 34% | 0% |
| K37 | 5% | 0% |
| R39 | 99% | 100% |
| K41 | 7% | N/A |
| K61 | 100% | N/A |
| K66 | 100% | 100% |
| R85 | 100% | 85% |
| K91 | 100% | 93% |
| K98 | 100% | 14% |
| K104 | 100% | 99% |
| S-Peptide Residues | | |
| K1 | 0% | N/A |
| K7 | 0% | N/A |
| R8 | 0% | N/A |

5.2 Relative Abundance of Formaldehyde Cross-linking

5.2.1 Percent Abundance and Equilibrium of Formaldehyde Cross-linking Sites in Calmodulin-Melittin

PFA cross-linking in the calmodulin-melittin system was localized to the following cross-linking sites: calmodulin K77 to calmodulin Q3, melittin G1 to calmodulin Y99, melittin G1 to calmodulin Q8, and melittin R24 to calmodulin R126. All of these cross-linked peptides appeared in proteins that were 14-33 kDa, suggesting that these represented intramolecular cross-links within one calmodulin molecule or calmodulin-melittin complex. The abundance of each amino acid involved in the cross-link in its unmodified, modified (+12 or +30 Da modification) and cross-linked form was determined. Unmodified and modified peptides identified by MaxQuant in the PFA treated sample, listed in Tables 3.2 and 3.3, and

5.2. Relative Abundance of Formaldehyde Cross-linking

cross-linked species identified manually, listed in Tables 4.8 and 4.9, were used. As mentioned in section 2.7, the abundance of each unique peptide was equated to the sum of normalized MS peak areas of all identical peptides. In addition, it was assumed that species were not lost during the sample preparation, MS detection or identification and that the MS-response of these species were uniform.

The percent of each cross-linking site in its unmodified, modified and cross-linked forms are listed in Table 5.3. Melittin G1 (36 %) had the highest percent in the modified form, followed by calmodulin K77 (32 %). Only 14 % of melittin R24 was identified its modified form. The cross-linked residues Q3, Y99, Q8 and R126 were not identified with a modification, which is consistent with proposed mechanism in which these residues formed cross-links in the second step of the reaction. For the calmodulin intramolecular cross-link, 2 % of K77 was cross-linked to Q3 and 2 % of Q3 was cross-linked to K77. About 5 % of melittin G1 was cross-linked to Y99 and 0.2 % of Y99 was cross-linked to G1. Only 0.5 % of G1 was cross-linked to Q8 and 0.05% of Q8 was cross-linked to G1. Finally, 2 % of R24 was cross-linked to R126 and only 0.02% of R126 was cross-linked to R24. Overall, this displays that the majority of the residues involved in the identified cross-links remained in their unmodified form, with only a small percent engaged in cross-linking. The percent abundance in the cross-linked form was similar for K77 and Q3, which reflects the uniformity predicted for intramolecular calmodulin cross-linked species. In contrast, the percent abundance of cross-linked forms between melittin and calmodulin residues in each identified cross-linking site varied significantly, which suggests that multiple conformations may exist that would disperse the abundance among different conformations.

5.2. Relative Abundance of Formaldehyde Cross-linking

Table 5.3: The percent abundance of PFA cross-linking sites in the unmodified, modified and cross-linked forms in PFA treated calmodulin-melittin; Note: additional decimal places are reported to clarify that values are $> 0\%$ or $< 100\%$, as described in section 2.7

| Modification Site (R_1, NH_2) | | | Cross-Linking Site (R_2, H) | | | | |
|-----------------------------------|--------------------|------------------|---------------------------------|-----------------|--------------------|------------------|---------------------|
| | Percent Unmodified | Percent Modified | Percent Crosslinked | | Percent Unmodified | Percent Modified | Percent Crosslinked |
| Calmodulin K77 | 67 % | 32 % | 2 % | Calmodulin Q3 | 98 % | 0 % | 2 % |
| Melittin G1 | 59 % | 36% | 5 % | Calmodulin Y99 | 99.8 % | 0 % | 0.2% |
| Melittin G1 | | | 0.5% | Calmodulin Q8 | 99.95 % | 0 % | 0.05% |
| Melittin R24 | 84 % | 14 % | 2 % | Calmodulin R126 | 99.98 % | 0 % | 0.02% |

The equilibrium constants for the formation of methylol and/or Schiff Base modified and cross-linked species were examined for each identified PFA cross-linking site. Figures 2.2 and 2.3 depict the equilibrium reaction for PFA cross-linking. The equilibrium constant expressions based on the abundance of each species, measured by the MS peak area, were derived in section 2.7.2.2. In these expressions, $K_{1_{MS}}$, $K_{2_{MS}}$ and $K_{3_{MS}}$ are the equilibrium constants for the formation of methylol modifications, Schiff Base modifications, and cross-links, respectively. For the formation of cross-links at sites in which a methylol intermediate was not identified, $K_{(1+2)_{MS}}$ was denoted as the equilibrium constant for the formation of Schiff Base modifications.

Out of all the PFA cross-linking sites, only K77 was identified with both a methylol (+30 Da) and a Schiff Base (+12 Da) intermediate. In this current analysis, distinguishing between Schiff Base and intrapeptide cross-links, both of which produce a +12 Da mass shift, is not possible. However, in proteins where cross-linking is a function of its structural restraints, it is less likely that cross-linking will be localized between residues existing on the same tryptic peptide, especially using short reaction times that promote more specific cross-linking. Previous PFA studies have confirmed that the formation of Schiff Base modifications is more likely than intrapeptide cross-links in proteins when using short reaction times similar to this current study [36]. Furthermore, previous studies have shown that although a +30 Da modified form is typically identified along with a +12 Da modified form at each modification site, it is possible to identify +12 Da modified form without a

5.2. Relative Abundance of Formaldehyde Cross-linking

+30 Da counterpart [85]. Thus, to account for the maximum number of modified forms for each cross-linking site, it is assumed that all +12 Da modifications were Schiff Base modifications.

Table 5.4 lists the equilibrium constants for each step of the PFA cross-linking reaction for each identified cross-linking site. Interestingly, all calculated equilibrium constants were less than one, suggesting that the equilibrium lies toward the unmodified form for each modification site and toward the modified over cross-linked form for each site. Theoretically, the equilibrium should favor the formation of the more stable methylene bridge cross-link structure in comparison to the Schiff Base structure [86]. Although the unmodified amino group is more stable than a Schiff Base modified protein amino group, the formation of the stable methylene bridge is expected to drive the modification reaction forward. However, the equilibrium constants calculated in this study are not only a reflection of the stability of each reaction product but are governed by other factors introduced by the protein complex and its structure. It was shown in section 3.2.5, that only 63% of calmodulin and melittin are expected to be bound together, i.e. be cross-linked together, based on the dissociation constant of the complex. Therefore, the total percent of reactive sites in their cross-linked form should be less than 63 %. Another factor to consider, is the possibility of multiple structural isomers produced from PFA cross-linking that would disperse the MS intensity across different products. This was apparent in the two different cross-links that were identified at melittin G1, i.e. one cross-link to calmodulin Y99 and the other to calmodulin Q8, indicating the presence of two structural isomers. Consequently, the percent of Y99 and Q8 in their cross-linked form and their cross-link formation equilibrium constant was relatively lower than with other cross-linking sites. Finally, the variation in the equilibrium constants across different cross-linking sites may be indicative of the accessibility of each cross-linking site, which is dictated by the structural restraints of the calmodulin-melittin complex. All together, this demonstrates that equilibrium is not only dependent on the reactivity and individual amino acids, but also the specific protein structure and attributes.

5.2. Relative Abundance of Formaldehyde Cross-linking

Table 5.4: The calculated equilibrium constants for each cross-linking reaction step for each identified PFA cross-linking site in PFA treated calmodulin-melittin

| Modification Site | Cross-linking Site | K_{1MS} | K_{2MS} | $K_{(1+2)MS}$ | K_{3MS} |
|-----------------------------------|--------------------|-----------|-----------|---------------|-----------|
| Calmodulin-Calmodulin Cross-links | | | | | |
| K77 | Q3 | 0.3 | 0.6 | | 0.1 |
| Calmodulin-Melittin Cross-links | | | | | |
| G1 | Y99 | | | 0.6 | 0.1 |
| G1 | Q8 | | | 0.6 | 0.01 |
| R24 | R126 | | | 0.2 | 0.1 |

5.2.2 Percent Abundance and Equilibrium of Formaldehyde Modification Sites in Calmodulin-Melittin

In the previous section, the percent abundance and equilibrium of PFA cross-linking sites in their unmodified, modified and cross-linked forms were examined. For calmodulin K77, it was shown that the unmodified form is favored over the methylol form and the methylol form is favored over the Schiff Base form. In addition, for melittin R24 and G1, it was demonstrated that the unmodified form is favored over the Schiff Base form. It was examined whether this trend is also true for sites where only modification and no cross-linking was observed. As shown in section 3.3.1.2, both methylol and Schiff Base modifications were localized on calmodulin K75 and K94 via MaxQuant. Only Schiff Base modifications were identified on calmodulin K148, R106, R74, R86, and R90, and only a methylol modification was localized on calmodulin K30 via MaxQuant. Percent abundance and equilibrium constant calculations were performed analogous to the previous section. These calculations could not be performed for the modified melittin segment ²¹KRK²³ since the two Schiff Base modifications identified on this segment via MaxQuant could not be localized to specific residues, as mentioned in section 3.3.1.2. As shown in Table 5.5, the percent of all modification sites in their modified form were $\leq 33\%$, suggesting that the majority of these residues remained unmodified. The percent abundances of all R residues in their modified form ($\leq 8\%$) were significantly lower than for K residues, further supporting that R residues are less reactive than K residues in the PFA modification reaction, as mentioned in section 3.3.1.2 and in literature [36, 85].

As displayed in Table 5.6, all equilibrium constants were less than one. This

5.2. Relative Abundance of Formaldehyde Cross-linking

indicates that the equilibrium lies toward the unmodified form over the modified form for all identified PFA modification sites, similar to the equilibrium constants of the PFA cross-linking sites shown in the previous section. The equilibrium constants for the formation of a methylol (K_{1MS}) and Schiff Base (K_{2MS}) at K75 were 0.1 and 2, respectively. This suggests that the Schiff Base modification formation is favored over the methylol intermediate and the methylol formation is not favored over the unmodified form for K75. This is consistent with the majority of modification sites identified by MaxQuant, i.e. K148, R106, R74, R86, and R90, containing only Schiff Base and no methylol modifications. Nevertheless, exceptions to the assumption that all +12 Da modifications correspond to Schiff Bases may exist. It is hypothesized that the large K_{2MS} value may be due to K75 existing in an intrapeptide cross-link, which would drive the reaction toward the formation of a stable cross-link bridge. For K94, equilibrium constants K_{1MS} and K_{2MS} were 0.1 and 0.5, respectively, suggesting that the unmodified form is favored over the methylol form and the methylol form is favored over the Schiff Base form. This is consistent with the trend observed at the cross-linking site K77. Cross-linking sites (calmodulin K77, melittin G1, melittin R24) modified in the first step of the reaction exhibited a higher average percent in their modified form (27% versus 15%) and larger average $K_{1MS} / K_{(1+2)MS}$ values (0.4 versus 0.1) than the modification sites at which cross-links were not identified (calmodulin K75, K94, K30, K148, R106, R74, R86, and R90). This is hypothesized to reflect the stability of the PFA cross-linking bridge, which drives the modification reaction forward when a cross-link formation follows.

Overall, nine out of the 13 potential PFA modifications sites in calmodulin and approximately four out of the five potential PFA modification sites in melittin were actually identified with PFA modifications. Calmodulin K13, K21, and R37 and potentially one of the melittin residues in $^{21}KRK^{23}$ were not identified with PFA modifications, suggesting that these were not accessible to PFA in the calmodulin-melittin structure. Although R126 was a potential modification site, no modifications and only cross-linking was identified at this site. This suggests that R126 was still accessible to PFA. In contrast to the modification sites, only four of the 20 potential PFA cross-linking sites and none of the potential melittin cross-linking sites were observed to have formed cross-links in the second step of the PFA cross-

5.2. Relative Abundance of Formaldehyde Cross-linking

linking reaction. The small subset of PFA reactive sites that were identified with cross-links demonstrates the specificity of PFA cross-linking, which is a function of close-proximity interactions within the calmodulin-melittin structure.

Table 5.5: The percent abundance of PFA modification sites in the unmodified and modified forms in PFA treated calmodulin-melittin; Note: additional decimal places are reported to clarify that values are $> 0\%$ or $< 100\%$, as described in section 2.7

| Modification Site (Calmodulin Residue) | Percent Unmodified | Percent Modified |
|--|--------------------|------------------|
| K30 | 67% | 33% |
| R74 | 93% | 7% |
| K75 | 80% | 20% |
| R86 | 92% | 8% |
| R90 | 94% | 6% |
| K94 | 86% | 14% |
| R106 | 99.7% | 0.3 % |
| K148 | 79% | 21% |

Table 5.6: The calculated equilibrium constants for the modification reaction for each identified PFA modification site in PFA treated calmodulin-melittin

| Modification Site (Calmodulin Residue) | K_{1MS} | K_{2MS} | $K_{(1+2)MS}$ |
|--|-----------|-----------|---------------|
| K30 | 0.5 | | |
| R74 | | | 0.1 |
| K75 | 0.1 | 2 | |
| R86 | | | 0.1 |
| R90 | | | 0.1 |
| K94 | 0.1 | 0.5 | |
| R106 | | | 0.003 |
| K148 | | | 0.3 |

5.2.3 Percent Abundance and Equilibrium of Formaldehyde Modification Sites in Ribonuclease-S

The percent PFA modification of the S-protein in the RNaseS system was examined, using a similar approach as section 5.2.2. Four PFA modification sites were

5.2. Relative Abundance of Formaldehyde Cross-linking

identified via MaxQuant in the RNaseS system: S-protein R33, K31, K91 and K98. Since unmodified forms of R33 and K31 were not identified, either a 100% modification of these sites occurred or most likely, not enough PFA peptides were identified to comment on the extent of their modification. Therefore, the analysis was continued with K91 and K98. As listed in Table 5.7, the percent of K91 and K98 in their modified forms was 28 % and 26 %, respectively, which suggests that the majority of the identified PFA modification sites remained unmodified. Nevertheless, the similar percent modification at both sites demonstrates consistency in reactivity across these PFA modification sites. Similar to the calmodulin-melittin system, the equilibrium constants for the formation of a modification at K91 and K98 was determined and are displayed in Table 5.7. For K98, both methylol and Schiff Base forms were identified with K_{1MS} and K_{2MS} values of 0.2 and 0.7, respectively. With no identified methylol form of K91, the equilibrium constant, $K_{(1+2)MS}$ for the Schiff Base formation on K91 was determined to be 0.4. Similar to the calmodulin-melittin system, all equilibrium constants for the formation of a PFA modification in the RNaseS system were below one, suggesting that the equilibrium favored the unmodified forms of K91 and K98. Furthermore, even though RNaseS contains 12 potential PFA modification sites, PFA modifications were only localized on four sites. All together, this suggests a reduced accessibility of modification sites to PFA due to the structural constraints of RNaseS. Nevertheless, the lack of peptides with sites modified by PFA identified in the RNaseS system does not facilitate a proper analysis of accessibility in RNaseS.

Table 5.7: The percent abundance of PFA modification sites in the unmodified and modified forms and the calculated equilibrium constants for the modification reaction for each identified PFA modification site in PFA treated RNaseS

| Modification Site (S-Protein Residue) | Percent Unmodified | Percent Modified | K_{1MS} | K_{2MS} | $K_{(1+2)MS}$ |
|---------------------------------------|--------------------|------------------|-----------|-----------|---------------|
| K91 | 72% | 28% | | | 0.4 |
| K98 | 74% | 26% | 0.2 | 0.7 | |

5.3 Cross-linked Product Classification and Abundance in the Calmodulin-Melittin System

In chapter 4, it was stated that uniform cross-linking is expected within calmodulin but not between calmodulin and melittin, due to the proposed existence of two different conformations of the complex. To validate this claim, the relative abundances of conformations captured by EDC, PFA, sulfoDST, BS³ and sulfoEGS cross-linked species were examined. MS/MS spectra confirmed different types of cross-linked products: interpeptide calmodulin-calmodulin and calmodulin-melittin cross-linked species, in which antiparallel (N-terminal domains aligned) or parallel (C-terminal and N-terminal domains aligned) binding conformations occurred. In chapter 4, Tables 4.2, 4.3, 4.5, 4.7, and 4.9 highlight EDC, sulfoDST, BS³, sulfoEGS and PFA cross-linked species, respectively, that support an antiparallel (blue) and parallel (white) binding orientation of calmodulin to melittin. Interpeptide calmodulin-calmodulin cross-links for EDC, BS³, sulfoEGS and PFA are displayed in Tables 4.1, 4.4, 4.6, and 4.8, respectively. Relative abundances were calculated by dividing the normalized peak area values of all identified cross-linked species for each type of cross-linked species (calmodulin-calmodulin, calmodulin-melittin, parallel calmodulin-melittin or antiparallel calmodulin-melittin) by the sum of normalized peak area values of all identified cross-linked species. These values are depicted in Tables 5.8 and 5.9.

Table 5.8: Relative abundance of calmodulin-calmodulin and calmodulin-melittin cross-linked peptides

| Interaction Captured | Cross-Linker | | | | |
|-----------------------|--------------|-----|----------|-----------------|----------|
| | EDC | PFA | SulfoDST | BS ³ | SulfoEGS |
| Calmodulin-Calmodulin | 5% | 94% | 0% | 6% | 51% |
| Calmodulin-Melittin | 95% | 6% | 100% | 94% | 49% |

5.3. Cross-linked Product Classification and Abundance in the Calmodulin-Melittin System

Table 5.9: Relative abundance of calmodulin-melittin interpeptide cross-links supporting the parallel and antiparallel binding orientation

| Calmodulin-Melittin Orientation Supported | Cross-Linker | | | | |
|---|--------------|-----|----------|-----------------|----------|
| | EDC | PFA | SulfoDST | BS ³ | SulfoEGS |
| Antiparallel | 98% | 3% | 88% | 72% | 64% |
| Parallel | 2% | 97% | 12% | 26% | 36% |

For EDC, sulfoDST, BS³, and sulfoEGS calmodulin-melittin cross-links (95, 100, 94, and 49% abundance, respectively) were greater or almost equal in abundance to calmodulin-calmodulin cross-links, out of which most captured the antiparallel orientation (98, 88, 72, and 64% abundance, respectively). However, PFA cross-linked species supported the parallel orientation over antiparallel (97 and 3% abundance, respectively). The PFA calmodulin-calmodulin cross-links, in which one major reaction product is expected, were observed to have a higher relative abundance (94%) than calmodulin-melittin cross-links (6%), in contrast to other cross-linkers. A higher relative abundance is a reflection of the formation of more of the same product, i.e. more calmodulin-melittin complex molecules containing the same calmodulin-calmodulin cross-links. This supports the expected uniformity in calmodulin's conformation and the reliability of PFA cross-linking to capture this uniformity. Interestingly, sulfoEGS demonstrated the least amount of uniformity in capturing calmodulin-calmodulin vs calmodulin-melittin cross-linking and in the orientation of calmodulin-melittin binding, which can be supported by the reduced specificity of its long cross-linker bridge.

Previous studies using Ca²⁺ saturated calmodulin-melittin have supported that calmodulin and melittin bind in both orientations [99, 102] and more recent reports have shown that the parallel binding is the major binding mode, which supports the observed PFA cross-linked structures [100, 101, 103]. However, the specific geometry of binding is unknown. It is also possible that calmodulin binds to melittin in only one orientation that fits the distance constraints of both antiparallel and parallel binding. This theory is based on the ambiguous definition of antiparallel and parallel binding i.e. based on the alignment of domains of calmodulin and melittin which span a range of about 75 and 13 amino acids, respectively. For example, the

difference between an “antiparallel” or “parallel” classification could potentially be between two adjacent residues between the N-terminal and C-terminal domain or opposite ends of the domains. Therefore, a more precise evaluation of distance constraints imposed by each cross-linker on an amino acid level may be required.

5.4 Crystal Structure Distance Constraints

5.4.1 Measuring and Applying Cross-linking Distances on Crystal Structures

The cross-linking distances were measured on the known calmodulin crystal structure. Cross-linking distances were also used to derive possible orientations of calmodulin and melittin. On each protein component, it was assumed that side chains can freely rotate about the alpha carbon with a radius equal to the length of the side chain. Side chains that can come into contact with the cross-linker within this radius were considered cross-linked to account for side chain flexibility. To account for backbone flexibility, an additional 6 Å was added to the maximum cross-linking distance. This calculation was based on recent molecular dynamic simulations [129]. Table 2.1 lists the maximum cross-linking distances for each cross-linker for every possible combination of reactive sites, which consider the backbone and amino acid side chain flexibility of each site. Cross-linking sites are listed from top to bottom and modification sites are listed left to right. For example, the maximum cross-linking distance between residues D and K for the EDC cross-linker is 16.1 Å.

5.4.2 Calmodulin Structure: Correlating Identified Cross-links to known crystal structure

Mapping cross-links identified within calmodulin on its known crystal structure can validate cross-linkers for their application to the ambiguous crystal structure of the calmodulin-melittin complex. Since calmodulin adopts a comparable structure when binding melittin regardless of the presence of Ca²⁺, the Ca²⁺-saturated, binding crystal structure of calmodulin was used for this purpose [95, 125]. Iden-

5.4. Crystal Structure Distance Constraints

tified distances between calmodulin-calmodulin cross-linked residues were measured using PyMol.

The calmodulin PFA cross-link was evaluated using the bound-state crystal structure of calmodulin. The MS/MS evidence discussed in section 4.3, supported cross-linking of K77 to Q3. As Table 2.1 indicates, PFA can cross-link K to Q up to 18.6 Å apart. The $C\alpha$ - $C\alpha$ distance between K77 and Q3 was measured to be 11.8 Å, falling within the maximum cross-linking distance of PFA. Since this cross-link formed within the distance constraints of the bound crystal structure of calmodulin, it is likely that this cross-link represents intramolecular cross-linking within one calmodulin molecule. This is also conveyed by the position of its SDS-PAGE band (14 - 19 kDa) in Figure 3.1.

For EDC cross-linked species $^{91}\text{VFDKDGNGYISAAELR}^{106}$ $^{22}\text{DGDGTITTK}^{30}$, the intrapeptide cross-link could have formed between K30 and either D22 or D24. Inter-residue $C\alpha$ - $C\alpha$ distances measured on the bound calmodulin crystal structure of 11.5 and 13.6 Å, respectively support cross-link formation between either of these sites. The interpeptide cross-link in this structure between K94 and either D22 or D24 had $C\alpha$ - $C\alpha$ distances of 26.3 and 30 Å, respectively, which were greater than the maximum cross-linking distance of EDC cross-links between K and D (16.1 Å). The $C\alpha$ - $C\alpha$ distances between interpeptide cross-linking sites E6 to K94 and E14 to K94 were significantly larger (28.9 and 20.4 Å) than the maximum cross-linking distance of EDC (17.3 Å). These interpeptide cross-links suggest that cross-linking occurred between two different calmodulin molecules. This contradicts the molecular weight of ~16 – 20 kDa measured by the location of SDS-PAGE gel band in which these cross-linked species originated (see Figure 3.1). However, the displaced gel bands for EDC cross-linked species, appearing significantly lower than what was expected, suggests that the structure of calmodulin may have been modified upon cross-linking. This change in protein surface area would affect the migration of the band. Two possible scenarios could have occurred: EDC formed cross-links between two separate calmodulin molecules and the cross-linking caused the protein to migrate further than expected in SDS-PAGE or EDC cross-linking occurred within a calmodulin molecule, altering it to a structure with inter-residues distances that do not match established calmodulin structures. The first scenario

5.4. Crystal Structure Distance Constraints

is supported by the study observing faster migration of proteins stabilized by disulfide bonds in comparison to its reduced form [146], which was also explained as a common phenomenon across all cross-linkers in this present study in section 3.2. The second scenario is supported by the fact that EDC is known to affect protein conformation [33]. Carboxylic groups are very abundant in calmodulin and are also only reactive with EDC, setting it apart from other cross-linker chemistry utilized in this present experiment. EDC forms cross-links with negatively charged carboxyl groups and positively charged amino groups, replacing them with neutral peptide bonds. This decrease in hydrophilic groups and increase in hydrophobic sites could induce a conformational change that decreases the protein's surface area thus making distances measured on the native calmodulin structures irrelevant. Residues may be closer in distance than represented in these structures, allowing for EDC cross-linking to occur. Previous reports of forming more compact protein structures upon the neutralization of charges support this theory [169, 170]. It is hypothesized that this is due to the decrease in electrostatic repulsion from adjacent negatively charged groups, promoting hydrogen bonding and forcing a more compact conformation.

The $C\alpha$ - $C\alpha$ distance between BS³ cross-linked residues K77 and K94 of 26.0 Å and sulfoEGS cross-linked residues K21 to K94 of 22.8 Å were within the maximum cross-linking distances of 30.2 and 34.9 Å, respectively. This supports cross-link formation within one calmodulin molecule, which is consistent with the molecular weight (14-19 kDa) measured by the SDS-PAGE analysis (Figure 3.1). Cross-linked species between identical peptides that were only observed with BS³ and sulfoEGS indicates that cross-linking captured the dimeric interaction between two different calmodulin molecules, that was observed in previously in literature via FTICR-MS [171].

All The $C\alpha$ - $C\alpha$ distances were measured on the crystal structure of bound calmodulin (Figure 5.1b). However, one possibility to be considered is that calmodulin existed in its unbound conformation (Figure 5.1a), thus making the identified calmodulin-melittin cross-linked species a result of random contact between uncomplexed melittin and calmodulin molecules in solution. Therefore, distances between verified cross-linking sites for each cross-linker were measured on the crystal structure of unbound, Ca²⁺-free calmodulin. Intramolecular cross-links

5.4. Crystal Structure Distance Constraints

were considered i.e. cross-linking of Q3 to K77 (PFA), K77 to K94 (BS³) and K21 to K94 (sulfoEGS) were compared to the maximum cross-linking distances of each cross-linker. The distances between K77 and K94 and between Q3 and K77 in the unbound and bound state of calmodulin were both within the maximum cross-linking distance of BS³ and PFA. However, the distance between K21 and K94 (~52 Å) was significantly greater than the maximum length of the sulfoEGS in the unbound calmodulin structure. Also, C α -C α distances between interpeptide EDC cross-linking sites were even larger in the unbound versus bound calmodulin crystal structure. It has been shown previously that Ca²⁺-free and Ca²⁺-loaded calmodulin-melittin share similar structures[95]. However in the absence of melittin, Ca²⁺-loaded calmodulin possesses a more compact, dumbbell structure in contrast to Ca²⁺-free calmodulin. One possibility to also consider is that the calmodulin remained unbound to melittin and had bound to the trace amount of Ca²⁺ ions in the deionized water. On the unbound, Ca²⁺-loaded calmodulin structure, both PFA and BS³ calmodulin cross-linking sites were within the maximum cross-linking distances of each cross-linker. However, the distance between calmodulin K21 and K94 (46.3 Å) was larger than the maximum cross-linking distance of sulfo-EGS. In addition, the distances between EDC calmodulin cross-linking sites E6 to K94 and E14 to K94 were 42.3 and 39.9 Å, respectively. These were also much larger than maximum cross-linking distance of EDC and even larger than the distances measured on melittin-bound calmodulin structure. This supports the absence of unbound, Ca²⁺-loaded calmodulin.

Overall, this suggests that the calmodulin likely existed in its bound state, supporting the complex formation between melittin and calmodulin. This is consistent with percent of expected bound calmodulin-melittin complex (63%) being higher than the percent of unbound calmodulin and melittin (32% and 33%, respectively). Cross-linked species identified with PFA, BS³ and sulfoEGS captured the tertiary structure of calmodulin. BS³ and sulfoEGS also captured the dimeric interaction between two calmodulin molecules. On the other hand, EDC cross-linked species did not coincide with inter-residue distances of either bound or unbound calmodulin structures.

5.4. Crystal Structure Distance Constraints

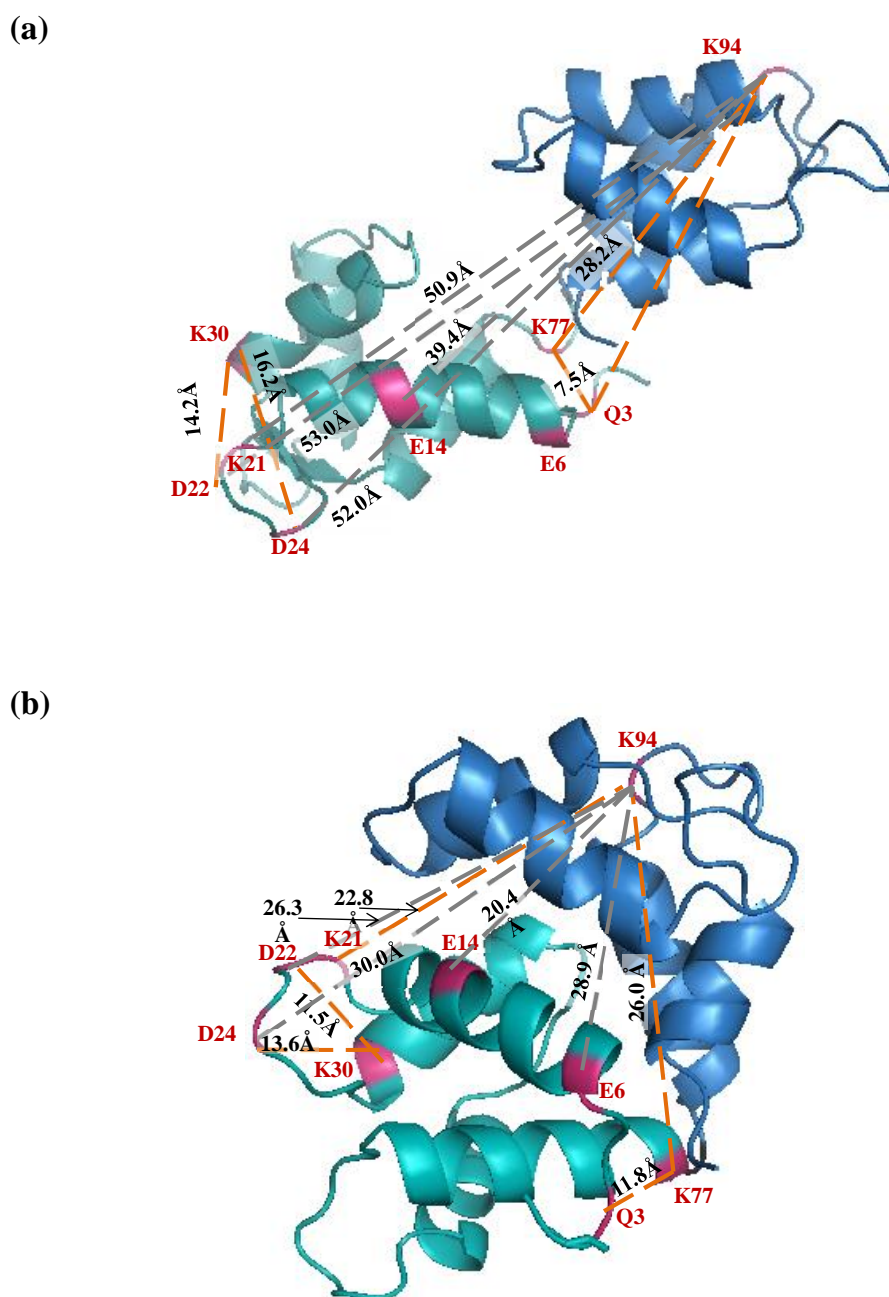


Figure 5.1: Identified cross-links were mapped on the Ca²⁺-free unbound (a) and bound-state (b) calmodulin conformation. Orange and grey lines represent inter-residue distances that do and do not agree with maximum cross-linker distances, respectively. Cross-linking sites are highlighted in red and calmodulin C and N terminal domains are colored in blue and teal, respectively.

5.4.3 Calmodulin-Melittin Complex Structure: Applying distance constraints from Cross-linking to Propose Binding Orientation of Calmodulin-Melittin

In order to examine the unknown binding orientation of calmodulin to melittin, crystal structures of the bound conformation of melittin and of the bound conformation calmodulin were oriented to fit the distance constraints imposed by the identified cross-linked species using PyMol. The distance between melittin and calmodulin residues in the identified cross-links were equated to the maximum cross-linking distances for each cross-linker. Table 5.10 summarizes the cross-linking sites and maximum distances used to derive structures. A minimum cross-linking distance constraint was imposed such that the $C\alpha$ - $C\alpha$ distance plus the backbone flexibility distance and side chain lengths (to account for flexibility of side chains) must be at least the length of the cross-linker bridge[172]. However, the scenario where inter-residue distances were rejected for being smaller than expected only arose with N-terminus to lysine residues (minimum cross-linking distance of 12.4 Å) cross-linked by the sulfoEGS cross-linker due to its long bridge. N-terminus to N-terminus (minimum cross-linking distance of 6 Å) cross-links are not possible in this study due to the acetylated N-terminus of calmodulin and the lysine to lysine minimum cross-linking distance, 18.8 Å, is longer than all cross-linker bridges used in this study.

5.4. Crystal Structure Distance Constraints

Table 5.10: The maximum distances between all MS/MS identified cross-linking sites and the respective binding orientation it supports for each cross-linker; Parallel orientations are shaded in white and antiparallel orientations are shaded in blue.

| Melittin Residue | Calmodulin Residue | Cross-Linker | Maximum Inter-residue Distance | Binding Orientation |
|------------------|--------------------|---------------------------|--------------------------------|---------------------|
| G1 | E14 | EDC | 10.9 Å | Parallel |
| K23 | E11 | EDC | 16.1 Å | Anti-Parallel |
| K23 | E54 | EDC | 16.1 Å | Anti-Parallel |
| R24 | R126 | PFA | 22.1 Å | Parallel |
| G1 | Q8 | PFA | 12.2 Å | Parallel |
| G1 | Y99 | PFA | 12.9 Å | Anti-Parallel |
| K23 | K30 | sulfoDST | 25.2 Å | Parallel |
| G1 | K94 | sulfoDST, BS ³ | 18.8 Å | Anti-Parallel |
| G1 | K75 | BS ³ | 23.8 Å | Anti-Parallel |
| K23 | K148 | BS ³ | 30.2 Å | Parallel |
| K23 | K94 | BS ³ ,sulfoEGS | 30.2 Å | Anti-Parallel |
| K23 | K77 | BS ³ ,sulfoEGS | 30.2 Å | Anti-Parallel |
| K23 | K21 | sulfoEGS | 34.9 Å | Parallel |

Structures proposed by cross-links supporting parallel binding (Figure 5.2a), cross-links supporting antiparallel binding (Figure 5.2b), EDC cross-links (Figure 5.3a), and PFA cross-links (Figure 5.3b), were examined. Parallel binding was classified as the binding of calmodulin's C-terminal and N-terminal domain to melittin's C-terminus and N-terminus, respectively. Antiparallel binding was defined as the binding of calmodulin's N-terminal and C-terminal domain to melittin's C-terminus and N-terminus, respectively.

It is important to note that other structures with slight variations in distances between melittin and calmodulin are always possible, but these variations did not significantly affect the general orientation of the two components. Furthermore, it is impossible to accurately depict the constant fluctuations of molecules in solution with one rigid structure.

Four out of the six NHS ester cross-linked species involved melittin binding to K residues near the flexible linker of calmodulin (K94 and K77), which is known

5.4. Crystal Structure Distance Constraints

to bind to target peptides [98]. Although this is consistent with literature findings, one may have to be cautious when applying these large cross-linkers to the small calmodulin-melittin complex. The maximum sulfoEGS cross-linking distance is 29 - 35 Å, which is almost the end to end distance of a melittin molecule (35.87 Å [173]). Therefore whether sulfoEGS captured a parallel or antiparallel orientation is ambiguous. On the melittin structure, the C α -C α distance between G1 and K23 is 30.1 Å and between K7 and K23 is 24.3 Å. The maximum BS³ cross-linking distance is 24-30 Å, so it is possible for BS³ to form a cross-link between an N-terminal lysine of calmodulin and a C-terminal lysine of melittin even if these molecules exist in a parallel orientation. In fact, when generating parallel and antiparallel binding structures that only satisfied EDC and PFA distance restraints specific to each orientation, all structures satisfied sulfoDST, BS³ and sulfoEGS distance restraints regardless of the orientation. The only exception was one antiparallel sulfoDST cross-link (the smallest NHS ester bridge of the three) that did not fit the parallel conformation. Interestingly, it was not possible to satisfy both EDC and PFA distance restraints simultaneously.

The structure proposed by PFA cross-links agrees with the characteristics of previous NMR and spectroscopic experiments. For example, W19 in melittin is known to be surrounded and blocked by the calmodulin C-terminal domain, which is shown in the structure. Also Y99 is known to be crucial in binding to melittin, which is a PFA cross-linking site. Cross-linking between calmodulin Q8 and melittin G1 is in compliance with the observation that residues 1 - 36 of calmodulin lie closest to the melittin's helix [101]. Furthermore, as recent NMR studies have shown, melittin primarily binds to the C-terminal domain of calmodulin, which is also supported by the structure derived by PFA cross-links[95]. Distance constraints imposed by EDC cross-linking was used to derive a binding conformation for the calmodulin-melittin complex. Cross-linking supported an antiparallel-like structure with the melittin N-terminus pointed toward the C-terminal calmodulin domain. Other than W19 in melittin being inaccessible upon binding to calmodulin and cross-linking between calmodulin E14 and melittin G1, EDC derived structural attributes were inconsistent with recent NMR and spectroscopy experimental findings in contrast to that of PFA [95, 101].

5.4. Crystal Structure Distance Constraints

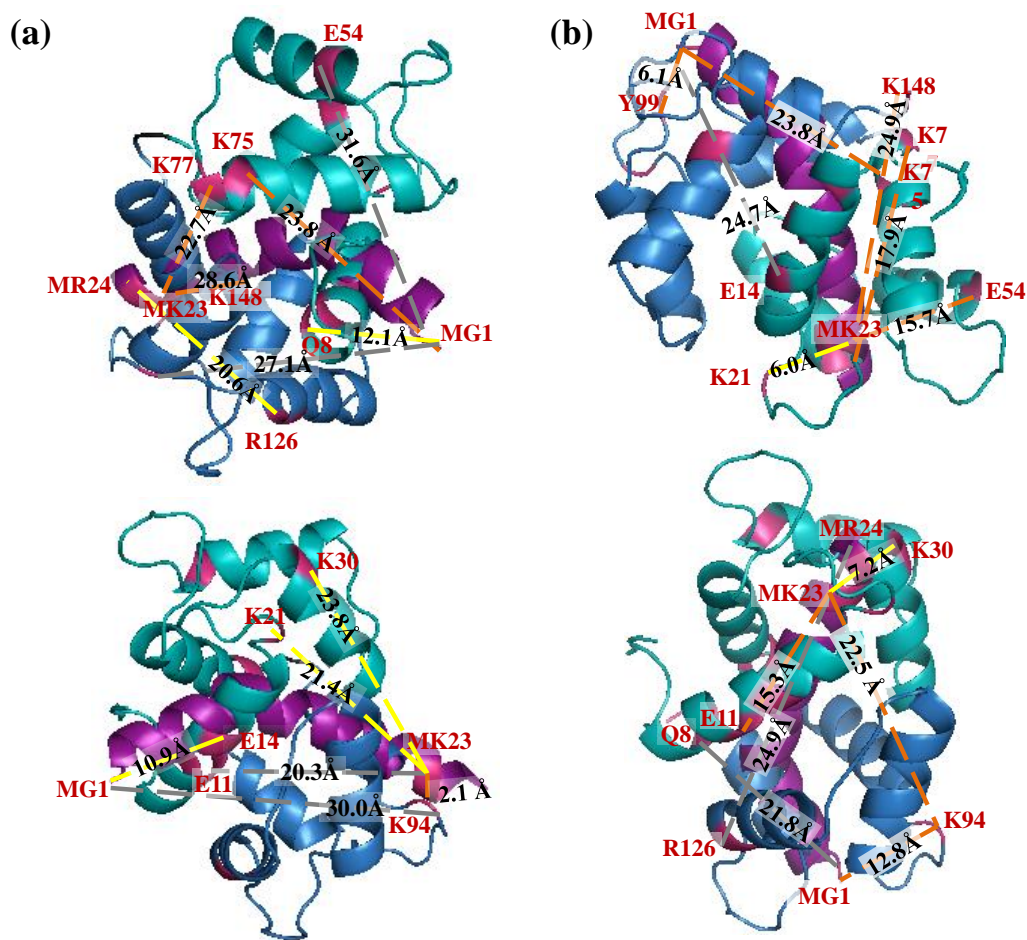


Figure 5.2: Calmodulin-melittin binding structures (two views of the same structure) proposed by cross-linking distance constraints that supported (a) parallel (yellow lines) and (b) antiparallel (orange lines) binding;. Orange/yellow and grey lines represent inter-residue distances that do and do not agree with maximum cross-linker distances, respectively. Cross-linking sites are highlighted in red, melittin is shown in purple, calmodulin C and N terminal domains are shown in blue and teal, respectively.

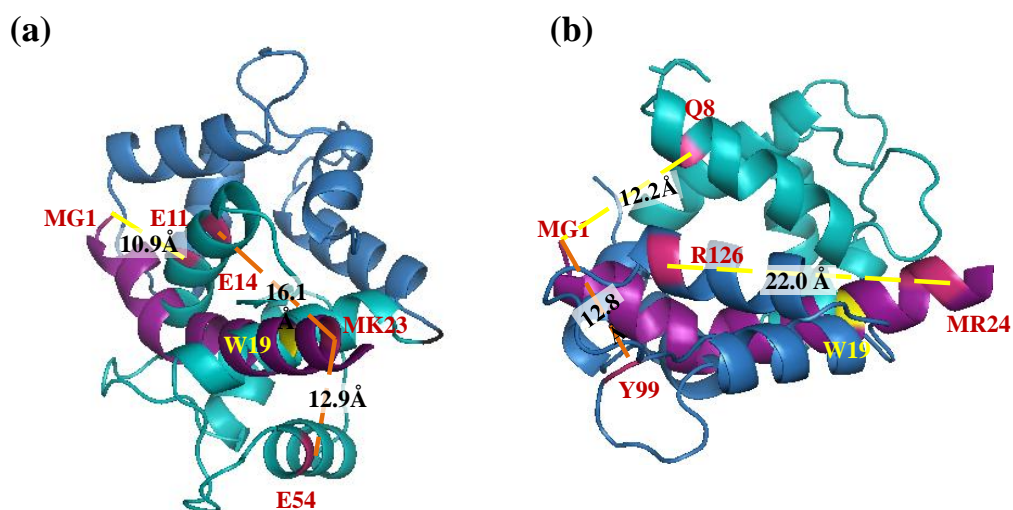


Figure 5.3: Calmodulin-melittin binding structures proposed by EDC (a) and PFA (b) distance constraints; W19 on melittin is highlighted in yellow. Orange and yellow lines represent inter-residue distances that support antiparallel and parallel binding, respectively. Cross-linking sites are highlighted in red, melittin is shown in purple, calmodulin C and N terminal domains are shown in blue and teal, respectively.

Overall, PFA cross-linked species imposed distance constraints that were most consistent with recently established structural attributes of calmodulin-melittin in

5.4. *Crystal Structure Distance Constraints*

literature [95, 101]. Also, this is the first time Ca^{2+} -free calmodulin bound to melittin, i.e. a transient complex, was stabilized by various cross-linkers and low-resolution structural information was deduced using MS/MS confirmed cross-linked species. This also represents the first study of PFA cross-link identification between non-covalently bound protein components, which demonstrated its potential for protein structure mapping. Although, more cross-links identified with higher confidence are required to verify structures, how each protein complex may require different cross-linkers was revealed. For example, small complexes such as calmodulin-melittin may not be accurately defined by large cross-linker bridges similar to the NHS esters. Also, implementing EDC cross-linking may not be favorable with proteins like calmodulin, which contain a high content of carboxyl group side chains. Moreover, extensive EDC cross-linking may affect protein structure.

Chapter 6

Arising Limitations of Mass Spectrometric Data Analysis of Formaldehyde Cross-linked Species

6.1 Identification of Limitations Arising in Workflow

Upon cross-linking, SDS-PAGE separation and trypsin digestion of proteins, there are several factors that affect the subsequent MS-detection and identification of resulting cross-linked peptides. First, the sensitivity, accuracy and resolution capabilities of the MS instrumentation to detect cross-linked species is vital. Second, since cross-linked species are identified and matched based on their monoisotopic mass, a reliable software is crucial for selecting the correct monoisotopic signal and calculating the monoisotopic mass. Third, in the manual identification of cross-linked species, it is important to examine the number of MS candidate cross-linked species that represent true MS/MS confirmed cross-linked species. This provides insight into the complexity of the reaction mixtures produced from cross-linking as a function of cross-linker type (i.e. EDC, PFA, sulfoDST, BS³ and sulfoEGS) to improve the identification of crosslinked species. Finally, present cross-link identification software programs have been successfully applied to established cross-linkers. However, understanding whether these software programs are comparable to the manual identification of cross-linked species would be useful, especially for PFA cross-linked species, which have yet to be identified by such software. Therefore, these aspects are examined to gain insight into the limiting factors associated

with this thesis work.

6.2 Mass Spectrometer Comparison

6.2.1 ABI QStar Versus Bruker Impact II QqTOF LC-MS/MS Analysis

Our lab was equipped with the ABI QStar XL QqTOF (QStar) [174] mass spectrometer for the major duration of this project and the LC-MS/MS analysis of various cross-linkers applied to the calmodulin-melittin and RNaseS protein systems was performed. The Bruker Impact II QqTOF (Impact II) [14] was available for use in November, 2014 for this current study. The mass accuracy and resolution of the QStar is < 10 ppm and 15,000, respectively, with a femtomolar detection limit. For the Impact II the mass accuracy and resolution is < 2 ppm and 40,000, respectively, with an attomolar detection limit and a dynamic range 5 orders of magnitude. The impact of the introduction of the new generation Impact II versus the outdated, 15 year old QStar for this cross-linking project was examined using the calmodulin-melittin protein model system.

6.2.2 Mass Spectrometric Data Analysis of Cross-linked Samples

Samples were prepared for the QStar analysis similar to the samples prepared for the Impact II analysis, as described in chapter 3. PFA along with various established cross-linkers (EDC/sulfoNHS, BS³, and sulfoEGS) were applied to the Ca²⁺-free calmodulin-melittin. Reaction mixtures were separated by SDS-PAGE, which provided cross-linking evidence similar to the SDS-PAGE shown in section 3.2. The LC-MS/MS data from the QStar was examined using the Analyst 1.1 QS Software (Analyst) to prepare monoisotopic mass lists and visualize MS and MS/MS spectra. The monoisotopic mass lists were processed and analyzed using the exact same procedure described in section 3.3. Control, peptide and impossible cross-linked signals were filtered from mass lists using Microsoft Excel and remaining MS signals were matched to theoretical cross-linked species using Mathematica all using a $+ 0.2$ Da window. The MS candidates were verified by manually examining the presence of their MS signals using Analyst and match-

6.2. Mass Spectrometer Comparison

ing their respective MS/MS spectra to expected theoretical fragment ions specific for each cross-linked species. The total number of MS candidates for EDC, PFA, BS3, and sulfoEGS cross-linked calmodulin-melittin was five, 18, 15 and seven, respectively out of which only one BS³ cross-linked candidate possessed a sufficient amount of MS/MS evidence for confirmation. In contrast, the total number of MS candidates identified in the Impact II analysis for EDC, PFA, sulfoDST, BS³, and sulfoEGS cross-linked calmodulin-melittin was 160, 335, 62, 77, and 158, respectively, out of which seven, seven, two, 11 and five, respectively, were interpeptide cross-links confirmed via MS/MS. This is displayed in Table 6.1.

Table 6.1: The number of identified calmodulin-melittin cross-linked candidates identified via MS and confirmed via MS/MS using the QStar and Impact II

| Cross Linker | MS Candidate Crosslinks | | MS/MS Confirmed Crosslinks | |
|--------------|-------------------------|-----------|----------------------------|-----------|
| | QStar | Impact II | QStar | Impact II |
| EDC | 5 | 160 | 0 | 7 |
| PFA | 18 | 335 | 0 | 7 |
| sulfoDST | N/A | 62 | N/A | 2 |
| BS3 | 15 | 77 | 1 | 11 |
| sulfoEGS | 7 | 158 | 0 | 5 |

The MS spectrum corresponding to a BS³ cross-linked structure, ⁷⁶MKDTDSEEEIR⁹⁰∧²³KR²⁴ ($m/z = 896.93$, $z = 2$), acquired on the QStar (Figure 6.1) and Impact II (Figure 6.2) is shown. The mass accuracy of this cross-linked species using the QStar and Impact II was 18 ppm and 2 ppm, respectively. The intensity of the signal detected by the Impact II was approximately three orders of magnitude higher than the signal detected by the QStar. Also, the narrow, defined peaks produced by the Impact II in contrast to the overlapping, broader peaks generated by the QStar analysis illustrates the superior resolving capabilities of the Impact II. Figure 6.1 shows the MS/MS spectra of the confirmed BS³ cross-linked species ⁷⁶MKDTDSEEEIR⁹⁰∧²³KR²⁴ from the QStar acquisition. A series of unmodified y ions (Iy1 to Iy9) for peptide I confirmed its sequence and two type 2 ions localized the cross-link between calmodulin K77 and melittin K23.

6.2. Mass Spectrometer Comparison

Since peptide II is only two amino acids in length, only one backbone ion, Ily1 corresponding to a terminal R residue, was present. However, since both peptide I and peptide II contain a terminal R residue, it is not clear whether the Ily1 was produced from the fragmentation of peptide I or II. No type 1 ions were present to further confirm the cross-link structure. The same cross-linking structure was identified by the Bruker Impact II, as seen in section 4.2.3 and in Figure 6.2 as a doubly charged species at m/z 896.93. The MS/MS signals acquired with the Impact II were almost 2 orders of magnitude higher in intensity. A series of unmodified y ions (Iy1 to Iy9) confirmed the sequence of peptide I and an unmodified type 1 ion for the melittin peptide confirmed its presence. Seven type 2 b ions (Ib2-II to Ib5-II and Ib8-II to Ib10+II) localized the cross-link to calmodulin K77 and melittin K23. Overall, with higher intensity signals and a more extensive fragment ion sequence coverage, the quality of the MS/MS spectrum acquired by the Impact II was shown to be enhanced in comparison to the QStar acquired MS/MS spectrum for this species.

6.2. Mass Spectrometer Comparison

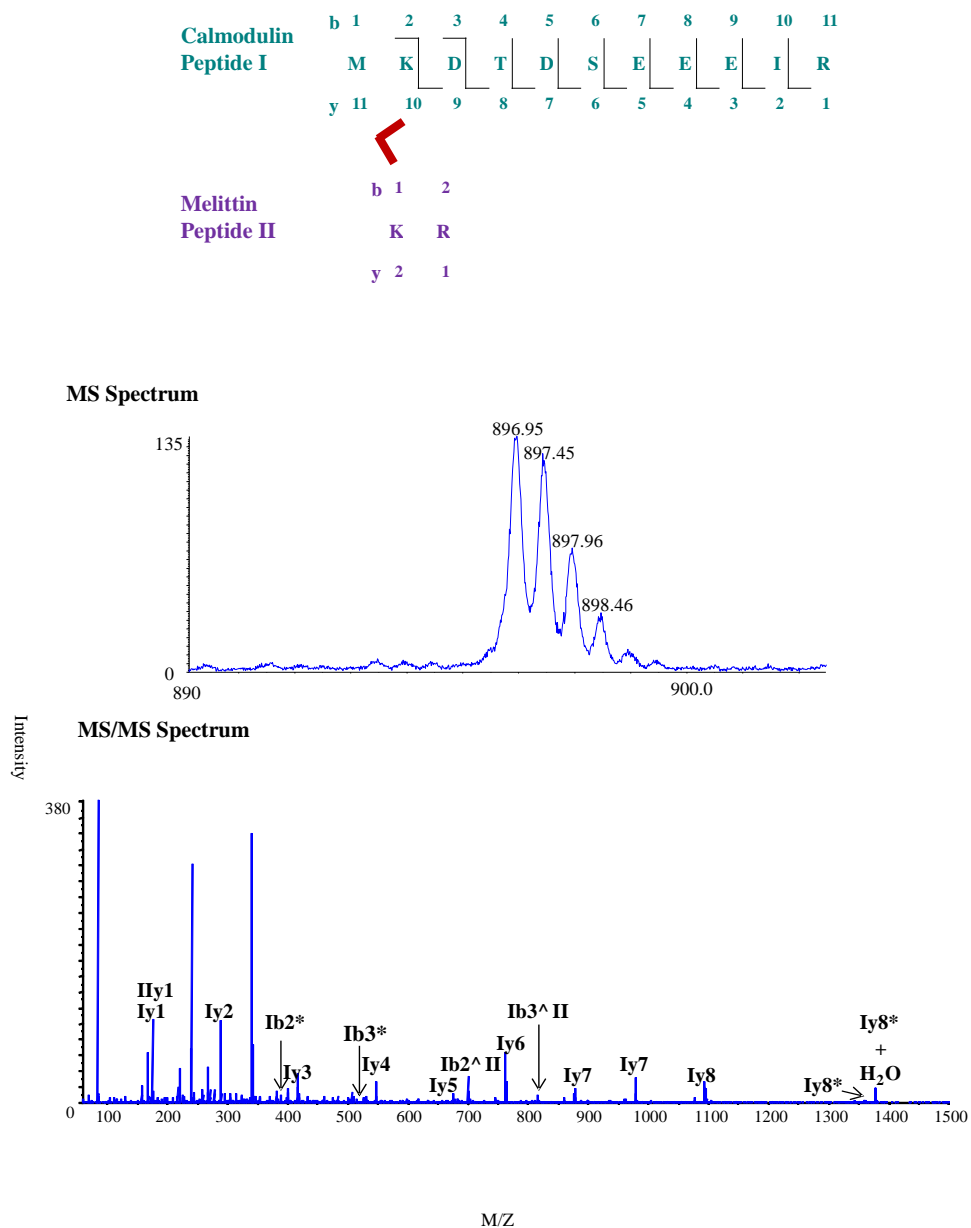


Figure 6.1: QStar acquired MS (middle) and MS/MS (bottom) spectrum of inter-peptide BS^{3+} calmodulin-melittin cross-link m/z 896.93 ($z = 2$)

6.2. Mass Spectrometer Comparison

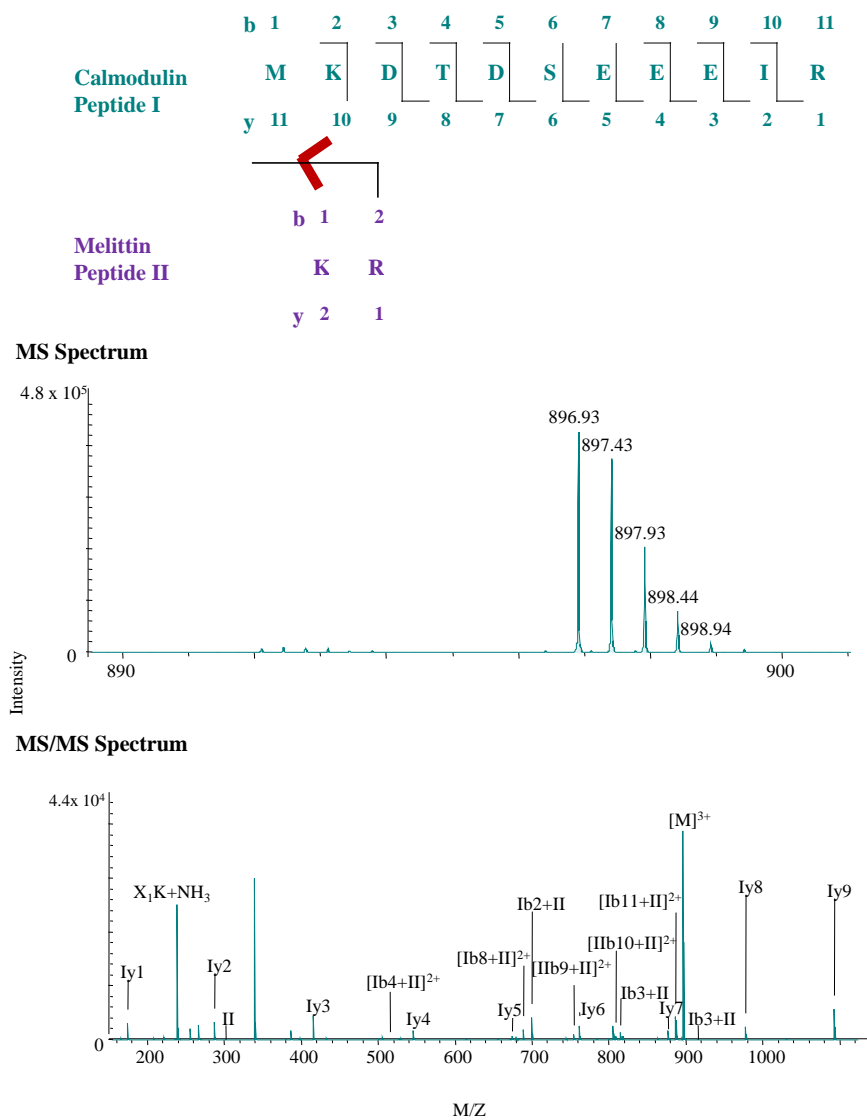


Figure 6.2: Impact II acquired MS(middle) and MS/MS (bottom) spectrum of interpeptide BS^{3+} calmodulin-melittin cross-link m/z 896.93 ($z = 2$); Proposed structure with the sequence fragment ion evidence indicated on the backbone of the peptide that corresponds to type 2 ions only (top).

6.2. Mass Spectrometer Comparison

Out of the 18 MS candidates for PFA cross-linked species, the QStar acquired an MS/MS spectrum (Figure 6.3) for only one with a signal, a triply charged species at m/z 603.24. This signal corresponded to the mass of the proposed structure $^{95}\text{DGNGYISAAELR}^{106\wedge 22}\text{RKR}^{24}$ with a total additional mass modification of +72 Da (two methylol and one Schiff Base modifications). As shown in Figure 6.3, the MS signal appeared at almost the noise level (intensity ~ 100) of the spectrum. In the MS/MS spectrum, only unmodified type 3 y-ions corresponding to peptide I (Iy1 to Iy6) were present and without any b-ions it is ambiguous whether this represents a cross-linked species or a modified missed cleaved peptide with the same terminal sequence. Only one type 3 y-ion and a modified type 1 ion (II+42) for peptide II was present. Again, since both peptide I and peptide II contain a terminal R residue, it is not clear whether the Ily1 was produced from the fragmentation of peptide I or II. Finally, several signals that did not correspond to the proposed PFA cross-linked structure or any modified/unmodified calmodulin or melittin peptide were present, suggesting a mixture with a contaminant. Overall, the MS/MS evidence was not sufficient to confirm the cross-linked species. In contrast, seven interpeptide PFA cross-linked species were confirmed via the MS/MS acquired with the Impact II, as shown in chapter 4.

6.2. Mass Spectrometer Comparison

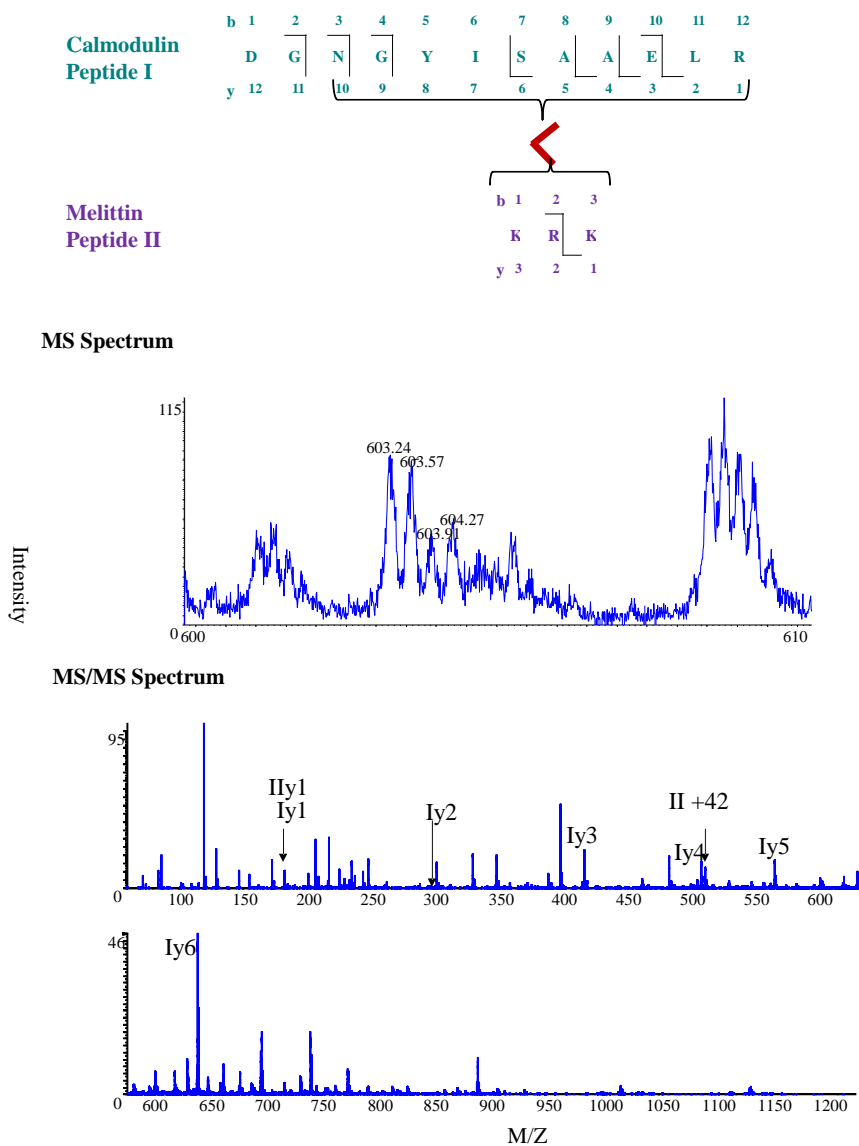


Figure 6.3: QStar acquired MS (middle) and MS/MS (bottom) of interpeptide PFA calmodulin cross-link m/z 603.24 ($z = 3$); Proposed structure with the sequence fragment ion evidence indicated on the backbone of the peptide that corresponds to type 3 ions only (top).

6.2. Mass Spectrometer Comparison

Interestingly, the MS signal the triply charged species at m/z 603.24 was detected in the control sample by the Impact II and was thus eliminated as a potential cross-linked species. Therefore, Figure 6.4 shows an example of the MS and MS/MS spectrum acquired by the Impact II for a different PFA cross-linked species at m/z 666.33 and a triple charge to illustrate the evidence required to confirm a PFA cross-linked species. This signal corresponded to the mass of calmodulin peptides $^{75}\text{KM}^{(\text{ox})}\text{K}^{77}$ and $^1\text{A}(\text{ac})\text{DQLTEEQIAEFK}^{13}$ plus the 12 Da bridge. A much more extensive and selective MS/MS spectra was produced, allowing for the confirmation of this cross-linked species. Type 3 (Iy1 to Iy10 and Ib2 to Ib6) and type 1(I-NH₃) ions confirmed $^1\text{A}(\text{ac})\text{DQLTEEQIAEFK}^{13}$. Type 3 (Iib2(ox), Iiy1, Iiy2(ox)), and type 1(II(ox)+12 and) ions verified $^{75}\text{KM}^{(\text{ox})}\text{K}^{77}$. The intensities of the MS and MS/MS signals from the Impact II of a confirmed PFA cross-linked species are 5 and 4 orders of magnitude higher than the MS and MS/MS signals from the QStar, respectively, due to its enhanced ion extraction and detection. The clearly defined peaks comprising the MS signal of the cross-linked species exemplifies the higher resolution capabilities of the Impact II versus the QStar. Also, the majority of the signals appearing in the MS/MS spectrum from the Impact II were assigned to the cross-linking structure, illustrating the increased selectivity of the Impact II acquisition and/or HPLC separation.

6.2. Mass Spectrometer Comparison

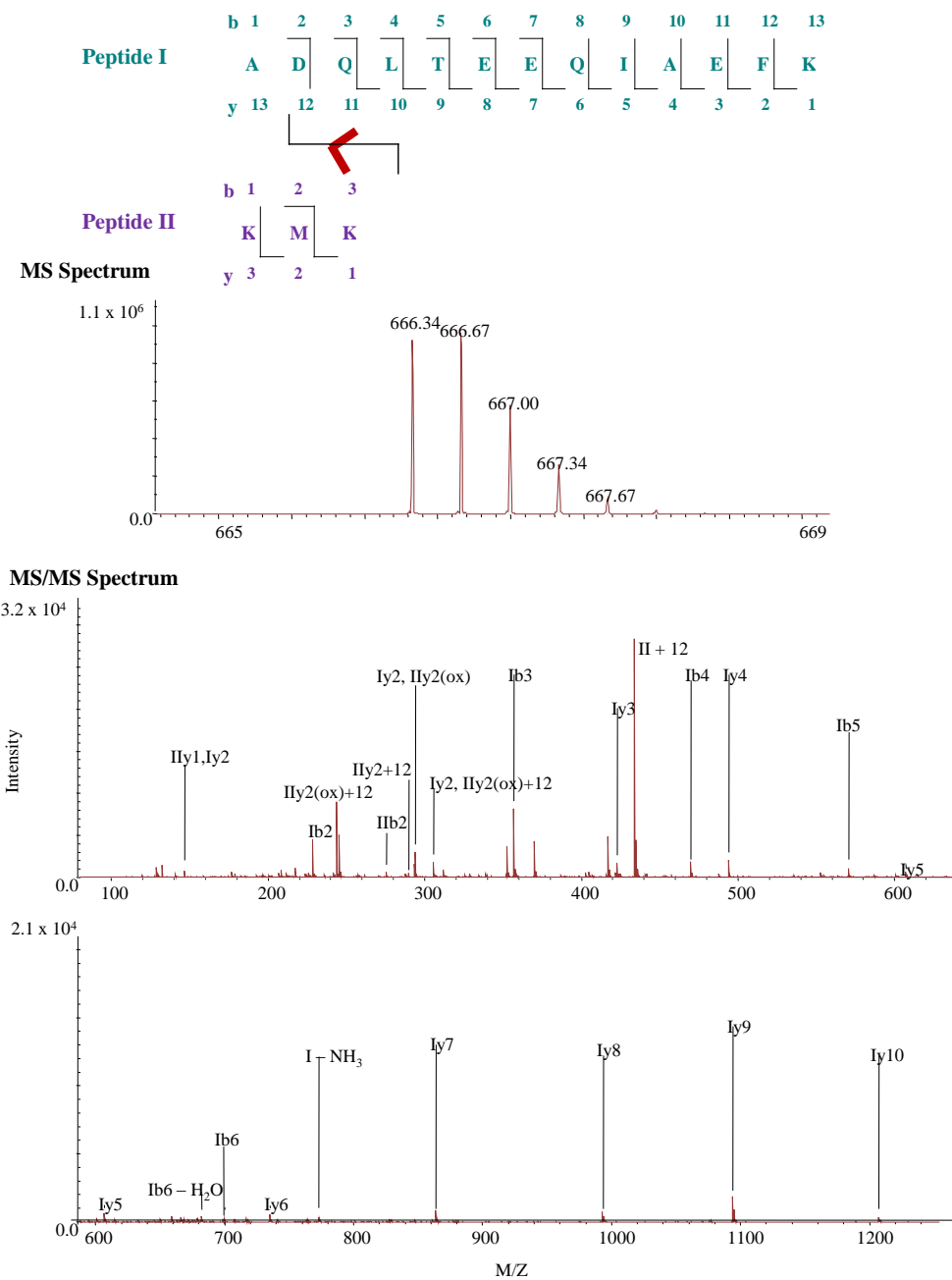


Figure 6.4: Impact II acquired MS (middle) and MS/MS (bottom) of interpeptide PFA calmodulin cross-link m/z 666.33 ($z = 3$); Proposed structure with the sequence fragment ion evidence indicated on the backbone of the peptide that corresponds to type 3 ions only (top).

As Table 6.1 shows, the cross-link identification at the MS and MS/MS level was drastically improved by the use of the Impact II. Furthermore, the quality of MS/MS spectra required to confirm cross-linking in the calmodulin-melittin system exceeds the capacity of the QStar and was possible only with the new generation, Impact II. Overall, this revealed the crucial role of a sensitive mass spectrometer in this study of cross-linked species.

6.2.3 Mass Spectrometric Analysis of Unmodified Calmodulin Peptides

To verify that the instrument performance itself limited the quality of LC-MS/MS data produced, the Impact II and QStar performance was evaluated using an unmodified calmodulin sample. The MS/MS data produced by the QStar and Impact II for the calmodulin sample was analyzed using the Mascot MS/MS peptide search[118]. The MS/MS Ion Search was performed using a peptide mass tolerance and fragment mass tolerance of 0.2 Da, a significant threshold $p < 0.05$ and a score cut off of 20. Variable modifications were set to as follows: Trimethyl (K), Oxidation (M), Acetyl (N-term), Deamidated (NQ) and Acetyl (Protein N-term).

Mascot identified 59 unique calmodulin peptides in the Impact II data with a sequence coverage and average mass accuracy of the peptides of 100% and 2.2 ppm, respectively. In contrast, Mascot only identified 11 unique calmodulin peptides in the QStar data with a sequence coverage and average mass accuracy of 67% and 26.7 ppm, respectively. Tables 6.2 and 6.3 list the highest scoring match for each peptide identified in the Impact II and QStar data, respectively. This illustrates that even in a simple, unmodified protein sample, the QStar was unable to produce sufficient quality of data and further supported the limitations of using the QStar for this project.

6.2. Mass Spectrometer Comparison

Table 6.2: Mascot MS/MS search results for the Impact II analyzed calmodulin sample with the highest scoring match for each peptide identified listed. The sequence position (starting and ending residue), observed m/z , experimental monoisotopic mass, theoretical monoisotopic mass, mass accuracy, number of missed cleavages, Mascot score, and sequence (trypsin cleavage site displayed in the beginning and end of the sequence as “R.” or “K.”) are listed left to right.

| Start – End | Observed | Mr(expt) | Mr(cal) | Mass Accuracy (ppm) | Missed Cleavage | Score | Peptide |
|-------------|----------|----------|---------|---------------------|-----------------|-------|--|
| 1 – 13 | 761.37 | 1520.73 | 1520.74 | 0.6 | 0 | 84 | K.ADQLTEEQIAEFK.E |
| 1 – 13 | 782.38 | 1562.74 | 1562.75 | 1.9 | 0 | 98 | K.ADQLTEEQIAEFK.E + Acetyl (N-term) |
| 1 – 13 | 782.87 | 1563.73 | 1563.73 | 2.0 | 0 | 99 | K.ADQLTEEQIAEFK.E + Acetyl (N-term); Deamidated (NQ) |
| 1 – 13 | 1130.55 | 3388.62 | 3388.62 | 1.1 | 2 | 67 | K.ADQLTEEQIAEFK.EAFSLFDKDGDTITTK.E + Acetyl (N-term) |
| 1 – 30 | 478.74 | 955.46 | 955.47 | 0.8 | 0 | 42 | K.EAFSLFDK.D |
| 14 – 21 | 922.95 | 1843.88 | 1843.88 | 0.7 | 1 | 125 | K.EAFSLFDKDGDTITTK.E |
| 14 – 30 | 943.95 | 1885.89 | 1885.89 | 0.7 | 1 | 110 | K.EAFSLFDKDGDTITTK.E + Trimethyl (K) |
| 14 – 30 | 643.65 | 1927.94 | 1927.98 | 18.7 | 1 | 21 | K.EAFSLFDKDGDTITTK.E + 2 Trimethyl (K) |
| 14 – 30 | 658.58 | 2630.29 | 2630.29 | 0.4 | 2 | 48 | K.EAFSLFDKDGDTITTKELGTVMR.S |
| 14 – 37 | 662.58 | 2646.29 | 2646.28 | 1.5 | 2 | 39 | K.EAFSLFDKDGDTITTKELGTVMR.S + Oxidation (M) |
| 14 – 37 | 570.61 | 1708.82 | 1708.83 | 4.6 | 1 | 21 | K.DGDTITTKELGTVMR.S + Oxidation (M) |
| 22 – 37 | 403.21 | 804.41 | 804.42 | 2.7 | 0 | 54 | K.ELGTVMR.S |
| 31 – 37 | 411.21 | 820.41 | 820.41 | 1.5 | 0 | 50 | K.ELGTVMR.S + Oxidation (M) |
| 31 – 37 | 1367.95 | 4100.82 | 4100.83 | 2.0 | 0 | 122 | R.SLQGNPTEAELQDMINEVDADNGTIDFPFLTMAR.K + 2 Oxidation (M) |
| 38 – 74 | 1368.28 | 4101.81 | 4101.81 | 1.5 | 0 | 95 | R.SLQGNPTEAELQDMINEVDADNGTIDFPFLTMAR.K + 2 Oxidation (M); Deamidated (NQ) |
| 38 – 74 | 1373.28 | 4116.82 | 4116.82 | 1.6 | 0 | 107 | R.SLQGNPTEAELQDMINEVDADNGTIDFPFLTMAR.K + 3 Oxidation (M) |
| 38 – 74 | 1373.61 | 4117.80 | 4117.81 | 1.4 | 0 | 129 | R.SLQGNPTEAELQDMINEVDADNGTIDFPFLTMAR.K + 3 Oxidation (M); Deamidated (NQ) |
| 38 – 74 | 1373.94 | 4118.78 | 4118.79 | 1.8 | 0 | 92 | R.SLQGNPTEAELQDMINEVDADNGTIDFPFLTMAR.K + 3 Oxidation (M); 2 Deamidated (NQ) |
| 38 – 74 | 1062.24 | 4244.91 | 4244.92 | 1.1 | 1 | 59 | R.SLQGNPTEAELQDMINEVDADNGTIDFPFLTMAR.M + 3 Oxidation (M) |
| 38 – 75 | 898.62 | 4488.06 | 4488.06 | 0.1 | 2 | 24 | R.SLQGNPTEAELQDMINEVDADNGTIDFPFLTMAR.MK.D + 2 Oxidation (M) |
| 38 – 77 | 901.81 | 4504.03 | 4504.05 | 4.4 | 2 | 42 | R.SLQGNPTEAELQDMINEVDADNGTIDFPFLTMAR.MK.D + 3 Oxidation (M) |
| 38 – 77 | 1131.02 | 4520.04 | 4520.05 | 2.1 | 2 | 35 | R.SLQGNPTEAELQDMINEVDADNGTIDFPFLTMAR.MK.D + 4 Oxidation (M) |
| 38 – 77 | 905.42 | 4522.05 | 4522.02 | 7.3 | 2 | 36 | R.SLQGNPTEAELQDMINEVDADNGTIDFPFLTMAR.MK.D + 4 Oxidation (M); 2 Deamidated (NQ) |
| 38 – 77 | 494.24 | 1479.69 | 1479.69 | 0.5 | 2 | 84 | R.KMKDTSSEEIFR.E |
| 75 – 86 | 499.57 | 1495.68 | 1495.68 | 0.8 | 2 | 21 | R.KMKDTSSEEIFR.E + Oxidation (M) |
| 75 – 86 | 496.74 | 1982.93 | 1982.94 | 1.6 | 3 | 75 | R.KMKDTSSEEIFR.AFR.V |
| 75 – 90 | 500.74 | 1998.93 | 1998.93 | 1.6 | 3 | 73 | R.KMKDTSSEEIFR.AFR.V + Oxidation (M) |
| 75 – 90 | 451.54 | 1351.59 | 1351.59 | 1.3 | 1 | 93 | K.MKDTSEEIFR.E |
| 76 – 86 | 676.80 | 1351.59 | 1351.59 | 0.4 | 1 | 94 | K.MKDTSEEIFR.E |
| 76 – 86 | 456.87 | 1367.59 | 1367.59 | 0.7 | 1 | 87 | K.MKDTSEEIFR.E + Oxidation (M) |
| 76 – 86 | 464.72 | 1854.84 | 1854.84 | 1.1 | 2 | 62 | K.MKDTSEEIFR.AFR.V |
| 76 – 90 | 936.42 | 1870.83 | 1870.84 | 1.9 | 2 | 85 | K.MKDTSEEIFR.AFR.V + Oxidation (M) |
| 76 – 90 | 547.24 | 1092.46 | 1092.46 | 0.4 | 0 | 54 | K.DTSEEIFR.E |
| 78 – 86 | 798.86 | 1595.71 | 1595.71 | 0.8 | 1 | 80 | K.DTSEEIFR.AFR.V |
| 78 – 90 | 546.91 | 1637.72 | 1637.72 | 0.7 | 1 | 45 | K.DTSEEIFR.AFR.V + Acetyl (N-term) |
| 78 – 90 | 667.32 | 3331.56 | 3331.56 | 0.9 | 3 | 51 | K.DTSEEIFR.AFR.VFDKDGNGYISAAELR.H |
| 78 – 106 | 1129.58 | 2257.14 | 2257.11 | 13.5 | 2 | 23 | R.EAFRVFDKDGNGYISAAELR.H |
| 87 – 106 | 877.94 | 1753.86 | 1753.86 | 1.4 | 1 | 145 | R.VFDKDGNGYISAAELR.H |
| 91 – 106 | 878.43 | 1754.85 | 1754.85 | 1.1 | 1 | 136 | R.VFDKDGNGYISAAELR.H + Deamidated (NQ) |
| 91 – 106 | 633.37 | 1264.72 | 1264.60 | 93.9 | 0 | 22 | K.DGNGYISAAELR.H |
| 95 – 106 | 633.31 | 1264.60 | 1264.60 | 0.6 | 0 | 74 | K.DGNGYISAAELR.H |
| 95 – 106 | 633.80 | 1265.59 | 1265.59 | 0.7 | 0 | 76 | K.DGNGYISAAELR.H + Deamidated (NQ) |
| 95 – 106 | 801.06 | 2400.16 | 2400.17 | 0.9 | 1 | 86 | R.HVMTNLGEKLTDEEVDIMIR.E + Trimethyl (K) |
| 107 – 126 | 601.30 | 2401.15 | 2401.15 | 0.6 | 1 | 38 | R.HVMTNLGEKLTDEEVDIMIR.E + Trimethyl (K); Deamidated (NQ) |
| 107 – 126 | 806.39 | 2416.16 | 2416.16 | 1.5 | 1 | 71 | R.HVMTNLGEKLTDEEVDIMIR.E + Trimethyl (K); Oxidation (M) |
| 107 – 126 | 605.29 | 2417.15 | 2417.15 | 1.4 | 1 | 40 | R.HVMTNLGEKLTDEEVDIMIR.E + Trimethyl (K); Oxidation (M); Deamidated (NQ) |
| 107 – 126 | 811.73 | 2432.16 | 2432.16 | 0.3 | 1 | 59 | R.HVMTNLGEKLTDEEVDIMIR.E + Trimethyl (K); 2 Oxidation (M) |
| 107 – 126 | 609.29 | 2433.15 | 2433.14 | 4.0 | 1 | 34 | R.HVMTNLGEKLTDEEVDIMIR.E + Trimethyl (K); 2 Oxidation (M); Deamidated (NQ) |
| 107 – 126 | 975.25 | 4871.23 | 4871.23 | 0.1 | 2 | 22 | R.HVMTNLGEKLTDEEVDIMIREADIDGQVNYEEFVQMMTAK.- + Trimethyl (K) |
| 107 – 148 | 978.45 | 4887.22 | 4887.22 | 0.7 | 2 | 37 | R.HVMTNLGEKLTDEEVDIMIREADIDGQVNYEEFVQMMTAK.- + Trimethyl (K); Oxidation (M) |
| 107 – 148 | 981.65 | 4903.21 | 4903.22 | 1.2 | 2 | 38 | R.HVMTNLGEKLTDEEVDIMIREADIDGQVNYEEFVQMMTAK.- + Trimethyl (K); 2 Oxidation (M) |
| 107 – 148 | 984.85 | 4919.21 | 4919.21 | 0.9 | 2 | 39 | R.HVMTNLGEKLTDEEVDIMIREADIDGQVNYEEFVQMMTAK.- + Trimethyl (K); 3 Oxidation (M) |
| 107 – 148 | 988.05 | 4935.20 | 4935.21 | 1.2 | 2 | 23 | R.HVMTNLGEKLTDEEVDIMIREADIDGQVNYEEFVQMMTAK.- + Trimethyl (K); 4 Oxidation (M) |
| 107 – 148 | 675.32 | 1348.62 | 1348.62 | 0.3 | 0 | 69 | K.LTDEEVDIMIR.E |
| 116 – 126 | 683.31 | 1364.61 | 1364.61 | 1.0 | 0 | 71 | K.LTDEEVDIMIR.E + Oxidation (M) |
| 116 – 126 | 1245.54 | 2489.06 | 2489.07 | 3.3 | 0 | 183 | R.EADIDGQVNYEEFVQMMTAK.- |
| 127 – 148 | 1246.04 | 2490.07 | 2490.06 | 3.9 | 0 | 103 | R.EADIDGQVNYEEFVQMMTAK.- + Deamidated (NQ) |
| 127 – 148 | 1253.54 | 2505.06 | 2505.07 | 2.8 | 0 | 188 | R.EADIDGQVNYEEFVQMMTAK.- + Oxidation (M) |
| 127 – 148 | 1261.54 | 2521.06 | 2521.06 | 2.5 | 0 | 160 | R.EADIDGQVNYEEFVQMMTAK.- + 2 Oxidation (M) |

6.3. Assignment of Monoisotopic Masses

Table 6.3: Mascot MS/MS search results for the QStar analyzed calmodulin sample with the highest scoring match for each peptide identified listed. The sequence position (starting and ending residue), observed m/z , experimental monoisotopic mass, theoretical monoisotopic mass, mass accuracy, number of missed cleavages, Mascot score, and sequence (trypsin cleavage site displayed in the beginning and end of the sequence as “R.” or “K.”) are listed left to right.

| Start – End | Observed | Mr(expt) | Mr(calc) | Mass Accuracy (ppm) | Missed Cleavage | Score | Peptide |
|-------------|----------|----------|----------|---------------------|-----------------|-------|---|
| 1 – 13 | 782.37 | 1562.72 | 1562.75 | 18.4 | 0 | 55 | K.ADQLTEEQIAEFK.E + Acetyl (N-term) |
| 1 – 13 | 783.39 | 1564.77 | 1564.71 | 34.2 | 0 | 45 | K.ADQLTEEQIAEFK.E + Acetyl (N-term); 2 Deamidated (NQ) |
| 31 – 37 | 403.21 | 804.41 | 804.42 | 12.9 | 0 | 32 | K.ELGTVMR.S |
| 75 – 86 | 451.56 | 1351.65 | 1351.59 | 39.4 | 1 | 38 | K.MKDTDSEEEIR.E |
| 75 – 90 | 496.76 | 1983.02 | 1982.94 | 41.3 | 3 | 40 | R.KMKDTDSEEEIREAFR.V |
| 75 – 96 | 500.75 | 1998.97 | 1998.93 | 19.6 | 3 | 45 | R.KMKDTDSEEEIREAFR.V + Oxidation (M) |
| 78 – 96 | 532.92 | 1595.72 | 1595.71 | 10.5 | 1 | 24 | K.DTDSEEEIREAFR.V |
| 95 – 106 | 633.31 | 1264.61 | 1264.60 | 4.3 | 0 | 49 | K.DGNGYISAAELR.H |
| 107 – 126 | 605.32 | 2417.23 | 2417.15 | 36.8 | 1 | 23 | R.HVMTNLGKLTDEEVDE |
| 127 – 148 | 1246.09 | 2490.16 | 2490.06 | 42.4 | 0 | 37 | R.EADIDGDGQVNYEEFVQMMTAK.- + Deamidated (NQ) |
| 127 – 148 | 1255.05 | 2508.08 | 2508.02 | 23.4 | 0 | 28 | R.EADIDGDGQVNYEEFVQMMTAK.- + Oxidation (M); 3 Deamidated (NQ) |

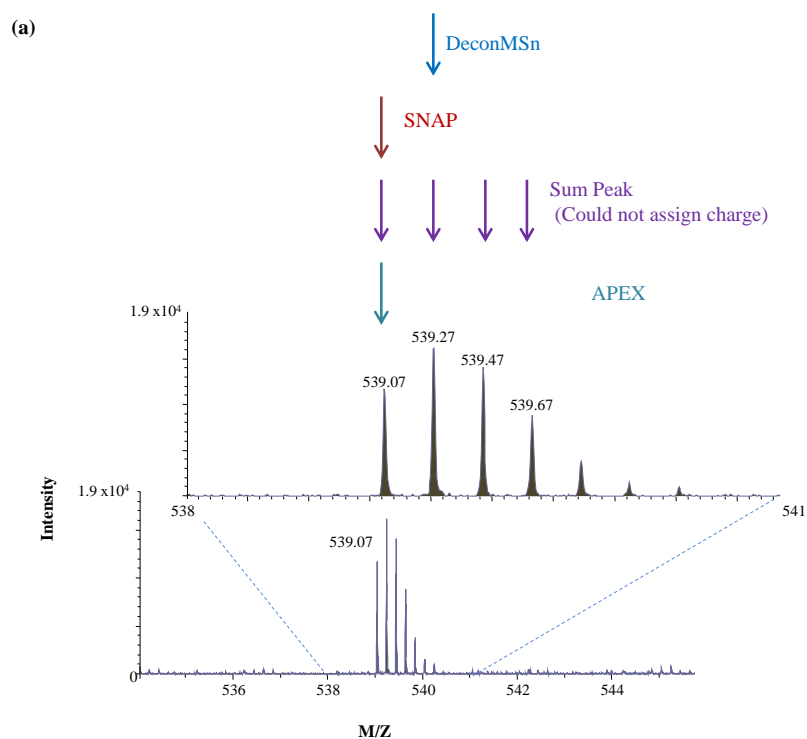
6.3 Assignment of Monoisotopic Masses

Deriving possible cross-linked masses is dependent on obtaining an exclusive list of accurate monoisotopic masses from the raw MS spectra. The Bruker Daltonics Compass Data Analysis 4.2 software offers three peak picking algorithms: Apex, SNAP and SumPeak. Apex picks peaks by calculating the derivatives of the MS signal intensities such that the peak maximum would have a first derivative equal to zero and a negative number for the second derivative. This algorithm works well for isotope-resolved peaks. SNAP takes into account molecular features to calculate the isotopic distribution for determining the monoisotopic mass, which is favorable for polymers such as proteins/peptides. Sum Peak uses a “pseudo slope” instead of calculating the derivative as with the APEX algorithm [175]. There are also other peak picking software available such as DeconMSn [176], which calculates monoisotopic peaks by determining the isotopic distribution by overlapping theoretical isotopic patterns with observed patterns. The three Bruker peak picking algorithms and DeconMSn were tested using a S/N cut off of 2 in a PFA cross-linked calmodulin-melittin sample, the most complex sample in the calmodulin-melittin MS data set. Figure 6.5a illustrates the accuracy of each peak

6.3. Assignment of Monoisotopic Masses

picking method using a species with a charge of 5 at m/z 539.07. SNAP and Apex accurately determined the monoisotopic signal, SumPeak selected all isotopic signals and DeconMSn selected the highest intensity peak as the monoisotopic signal instead of the first peak. As shown in Figure 6.5b, DeconMSn produced a list of 151840 m/z values, treating every isotopic peak as a separate signal and calculating its respective mass. To manually distinguish between the isotopic and monoisotopic peaks in this large data set would be highly tedious and therefore, DeconMSn was not used. The Bruker peak picking algorithms, Apex, SumPeak and SNAP generated a list of 1631, 2418 and 1701 m/z values. Although all algorithms calculated the monoisotopic peak correctly, Apex and SumPeak outputted isotopic peaks in addition to the monoisotopic peak for each mass. Apex also missed 133 monoisotopic signals that SumPeak had detected. Finally both Apex and SumPeak were not able to calculate the monoisotopic mass of 480 and 569 signals, respectively, since both algorithms failed to interpret the charge of these signals. SNAP, on the other hand, was able to determine the charge of every signal and calculated all monoisotopic masses. To prevent the tedious task of manually interpreting the charges of signals using the APEX and SumPeak algorithms, SNAP was chosen as the peak picking algorithm for this study.

6.3. Assignment of Monoisotopic Masses



(b)

| Peak Picking Method | DeconMSn | Bruker Algorithms | | |
|---------------------------------------|---|---|---|--|
| | | APEX | SUMPEAK | SNAP |
| Number of MS Signals | 151840 | 1631 | 2418 | 1701 |
| Accurate Assignment of Peak | Assigned highest intensity and treated every isotopic peak after as a separate peak | Calculated correct mass but outputted all isotopic m/z signals; | Calculated correct mass but outputted all isotopic m/z signals; | Calculated correct Monoisotopic mass and only outputted m/z of monoisotopic peak |
| Missing charge/ Failed to Deconvolute | 0 | 480 | 569 | 0 |

Figure 6.5: (a) Monoisotopic peak assigned for MS Signal m/z 539.07 ($z = 5$) by DeconMSn, SNAP, SumPeak and Apex peak picking methods as indicated by the blue, red, purple and teal arrows, respectively. (b) Summary of peak picking methods for a PFA treated calmodulin-melittin sample

6.3. Assignment of Monoisotopic Masses

However, in highly complex cross-linked peptide mixtures, even SNAP did not always consistently pick the correct monoisotopic peak and the deconvoluted final mass lists contained values that were + 1 Da off the actual mass. Assigning the wrong monoisotopic peak in proteomic MS spectra is a commonly observed issue. Furthermore, cross-linked peptides tend to be highly charged, larger species. The larger the species, the more isotopic forms exist, producing more isotopic peaks. Therefore the relative intensity of the monoisotopic peak decreases and often appears at a lower intensity than the adjacent isotopic peaks and the software tends to pick the most intense peak [177, 178]. The list of incorrectly assigned m/z for each calmodulin-melittin cross-linked sample can be found in Appendix A.4, which supports this trend since most actual monoisotopic masses were ~ 1 Da lower than the assigned monoisotopic masses. Table 6.4 displays the percent of the total MS candidate cross-linked species that had incorrectly assigned monoisotopic masses for each cross-linker. The percent of incorrectly assigned monoisotopic masses is similar for EDC, PFA and sulfoEGS (31%, 27% and 31%, respectively). EDC and PFA produced the most number of reaction products due to a larger number of their reactive sites in calmodulin-melittin. Thus, with more species in the reaction mixture, the signal intensity is dispersed over more products and the monoisotopic peak becomes more difficult to distinguish. For the NHS ester cross-linkers, which have the same cross-linking site specificity, the percent of incorrectly assigned monoisotopic masses increased with larger cross-linker bridge lengths. As the length of the cross-linker bridge increases, the chance of two reactive residues existing within the distance of the cross-linker bridge increases, thus increasing the number of cross-linked products. An increase in the number of reaction products would decrease the relative monoisotopic signal intensity, making it more ambiguous to the software to pick the correct monoisotopic peak.

6.4. Complexity of Cross-linked Candidates Confirmed by Mass Spectrometry

Table 6.4: The percent of the total number of MS candidate cross-linked species with incorrectly assigned monoisotopic peaks by the software for each cross-linker

| Cross-Linker | Percent of Incorrectly Assigned M/Z |
|-----------------|-------------------------------------|
| EDC | 31 % |
| PFA | 27 % |
| SulfoDST | 10 % |
| BS ³ | 21 % |
| SulfoEGS | 31 % |

6.4 Complexity of Cross-linked Candidates Confirmed by Mass Spectrometry

The total number of MS candidates for EDC, PFA, sulfoDST, BS₃, and sulfoEGS cross-linked calmodulin-melittin was 160, 335, 62, 77, and 158 respectively (*m/z* lists are shown in Appendix A.4). Upon the manual inspection of each candidate's MS/MS spectrum, the distribution of MS candidates was determined and is shown in Table 6.5.

Table 6.5: The percent of the total number of MS confirmed candidate cross-linked masses that correspond to modified peptides, undetermined species, species with insufficient MS/MS and cross-linked species for each cross-linker.

| Candidate Classification | Cross-Linker | | | | |
|---------------------------|--------------|-----|----------|-----------------|----------|
| | EDC | PFA | SulfoDST | BS ³ | SulfoEGS |
| Peptides | 38% | 47% | 34% | 29% | 39% |
| Undetermined Species | 8% | 18% | 11% | 19% | 8% |
| Insufficient MSMS Spectra | 49% | 33% | 52% | 38% | 50% |
| Cross-Link | 4% | 2% | 2% | 15% | 4% |

MS candidates were classified as “modified peptides” if their MS/MS signals matched the sequence of only one peptide component and the mass of the candidate

6.4. Complexity of Cross-linked Candidates Confirmed by Mass Spectrometry

was equal to the mass of the matching peptide and a modification (i.e. either cross-linker or protein-specific modification). The relative occurrence of these peptides that share the same mass as a possible cross-linked species was 38%, 47%, 34%, 29% , and 39% of the total number of MS candidates for EDC, PFA, sulfoDST, BS³ and sulfoEGS, respectively. “Undetermined Species” were MS candidates with MS/MS signals that do not match the sequence of any calmodulin or melittin peptide and were 8% ,18%, 11%, 19%, and 8% of the total number of MS candidates for EDC, PFA, sulfoDST, BS³ and sulfoEGS, respectively. MS candidates with “Insufficient MSMS Spectra” are those with precursor signals in which an MS/MS spectrum was not generated or contained only a few signals above the noise level of the spectrum. These species were 49%, 33%, 52%, 38%, and 50% of the total number of MS candidates for EDC, PFA, sulfoDST, BS³ and sulfoEGS, respectively. Out of the total number of species that possessed sufficient MS/MS spectra, 46%, 64%, 45%, 48%, and 47% were identified as non-cross linked (peptides or undetermined) species for EDC, PFA, sulfoDST, BS³ and sulfoEGS, respectively.

The substantially larger number of MS candidates for PFA is due to its reactivity with several amino acids and numerous possible modifications and cross-links. Similar to PFA, EDC forms close proximity cross-links due to its zero-length bridge and has several reactive sites present in this model system, supporting its relatively higher number of MS candidates. For the NHS ester cross-linkers, the number of MS candidates increased as a function of the cross-linker length. This illustrates the higher probability of two cross-linking sites existing within the distance of a cross-linker’s bridge for longer cross-linkers. This may reduce its specificity for only capturing cross-linking sites that are close enough to interact or are structurally relevant [32]. Regardless, the high specificity of NHS ester cross-linkers comes from its reactivity with only N-terminal and K residues, which is supported by the production of relatively fewer reaction products in comparison to PFA and EDC observed in this study. For PFA and sulfoDST, only 2% of the total candidates were interpeptide cross-linked species. Out of the total EDC and sulfoEGS MS candidates, only 4% were interpeptide cross-linked species. Finally, BS³ produced the highest percent of cross-linked species with 15% of its MS candidates being actual interpeptide cross-links. In general, this revealed that MS/MS

6.4. Complexity of Cross-linked Candidates Confirmed by Mass Spectrometry

confirmed interpeptide cross-linked species comprise of a very minute subset of the MS candidates for all cross-linkers. PFA produced the highest percentage of peptides and other species with masses that were equal to possible cross-linked species, supporting the complexity of its reaction products. This illustrates the crucial role of MS/MS to distinguish between actual cross-linked species and such reaction products especially with PFA. The acquisition of MS/MS spectra in this experiment was dependent on a sufficient precursor ion signal intensity. Forming a few reaction products with a high yield would increase the signal intensity of each precursor ion and increase the chance of a better quality MS/MS spectrum. Interestingly, although PFA produced the most number of different reaction products, it exhibited a relatively lower percent of insufficient MS/MS spectra for these products. This suggests that the yield of these products was high enough to compensate precursor ion signal dispersion among several products. However, out of the PFA reaction products that possessed a sufficient MS/MS spectrum, the majority of them corresponded to non-cross linked species supporting that non-cross linked species matching the mass of candidate cross-linked species are abundant in the PFA reaction products. In all other cross-linkers besides PFA, the percent of non-cross linked species that matched the mass of their MS candidate cross-linked species was lower illustrating their reduced complexity and more straightforward identification.

In conclusion, NHS esters have high specificity that comes with a higher chance of false positive identification. Although smaller and less specific cross-linkers such as EDC and PFA do not introduce additional degrees of freedom from a linker bridge, they produce more reaction products. Nevertheless, a combination of all these types of cross-linkers can build a more accurate MS/MS analysis of protein complexes. With the MS/MS verified cross-linked species representing a small pool of the MS candidates, many of which corresponded to unknown structures or modified peptides, this study also emphasized that MS/MS is even more crucial in the case of PFA to confirm the presence of cross-links.

6.5 Manual versus Software Identification of Calmodulin-Melittin Cross-links

6.5.1 Cross-linking Software Search Parameters and Evaluation Criteria

A variety of software have successfully been applied to identify EDC, sulfoDST, BS³ and sulfoEGS cross-links, however, these automated methods have yet to be applied to PFA cross-links [32, 34, 39–41]. Therefore, the reliability of software to find PFA calmodulin-melittin cross-linked species at the MS/MS level was tested. Since no cross-linking software specifically designed for PFA exists, StavroX[120] and pLink[121] were selected since users can define custom cross-linkers with multiple cross-linking sites, protein databases, and multiple cross-linker specific variable modifications. These software programs were also tested on EDC, sulfoDST, BS³ and sulfoEGS calmodulin-melittin cross-linked sample data for a comparison. Both pLink and StavroX have previously been utilized to identify EDC cross-linked species [34] and NHS-ester cross-linked species [120, 121]. In both StavroX and pLink software programs, the cross-linker bridge is assumed to not be cleavable. A variation of StavroX, MeroX[119], assumes the cross-linker bridge fully fragments. Although PFA cross-links can theoretically contain both intact and fragmented cross-link bridges under CID [50], the majority of the PFA cross-linked species examined in chapter 4 did not possess an intact cross-linker bridge. Regardless, both StavroX and MeroX was tested on PFA cross-linked sample data. StavroX was used for the established cross-linked samples since NHS ester and EDC cross-link bridges are not expected to significantly fragment under CID [51, 52]. In all software program searches, the possibility of trypsin cleaving after a cross-linked residue was not excluded since cleavage after PFA cross-linked residues was observed manually in chapter 4. The general criteria for verifying cross-linked species utilized in section 4.5 was used to filter cross-links identified by the software. This criteria is similar to the scoring methods utilized in StavroX/MeroX and pLink, which are based on the percent of MS/MS fragment ion evidence (i.e. the number of expected y and b ions divided by the total length of the peptide)[120, 165, 166]. For StavroX and MeroX, the annotated spectra

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

provided by these software programs were inspected. In this spectra, the notation is used such that the smaller and larger peptide components are referred to as “ β ” and “ α ” (i.e. peptide I and II). The peptide N-terminus and C-terminus are denoted as “[” and “]”, respectively and the protein N-terminus and C-terminus are represented by “{” and “}”, respectively. However, for consistency, the notation used in chapter 4 is used in the text (i.e. peptide I and II for “ β ” and “ α ”). For pLink, a corresponding annotated MS/MS spectrum for identified cross-linked species was not generated and thus these were verified manually using the Bruker Daltonics Compass Data Analysis 4.2 software.

6.5.2 Established Cross-linkers

6.5.2.1 EDC

Figure 6.6 lists the identified cross-linked species and illustrates the overlap between each StavroX, pLink and manual cross-link identification method. StavroX identified a total of 19 unique EDC cross-linked species, out of which 15 had insufficient MS/MS evidence to confirm its presence, two corresponded to single peptides with missed cleavage sites, and two were actual cross-linked species. There was no overlap between the manual and StavroX cross-link species identification. Figure 6.7 shows an example of an annotated MS/MS spectrum of a EDC cross-linked species identified by StavroX. Six out of the 13 (46%) and two out of the five (40%) expected backbone fragment ions for the melittin and calmodulin peptide was present, respectively. With sufficient y and b fragment ions for both component peptides with the cross-linker bridge intact, this species was accepted as a cross-link. pLink identified only three unique EDC cross-linked species, out of which one corresponded to a single peptide with a missed cleavage site and two were actual cross-linked species. All of these species were also identified with StavroX and there was also no overlap between the manual and pLink cross-linked species identification.

Out of two cross-linked species identified by the software, the doubly charged species at m/z 580.84 corresponded to a signal that also appeared in the EDC control sample and the triply charged species at m/z 577.96 was not selected by the peak peaking software, because its isotopic peaks overlapped with a signal of a

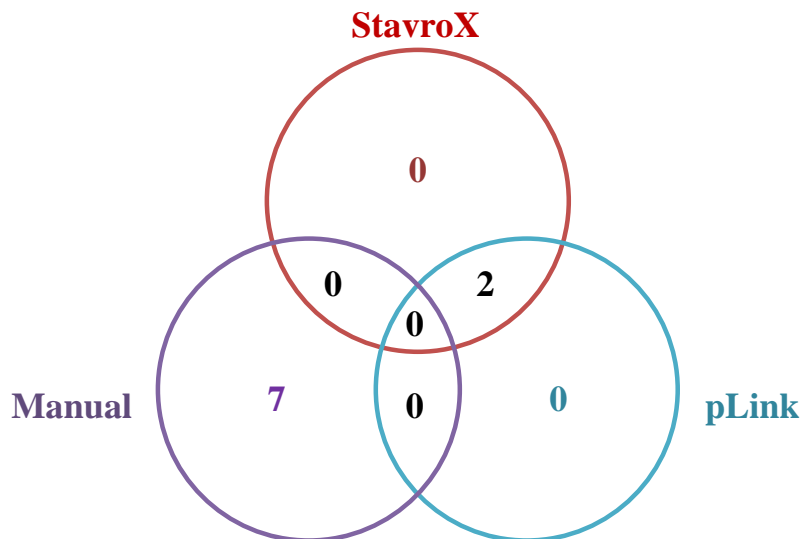
6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

species with a similar m/z and elution time. Therefore, these species were not identified manually.

As shown in Figure 6.6b (as indicated by residues highlighted in red), the StavroX and pLink identified cross-linked species demonstrated cross-linking between melittin G1 and calmodulin E87 and E84. Although melittin G1 was revealed as a cross-linking site manually, calmodulin E84 and E87 were not found to be cross-linking sites via manual identification.

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

(a)



(b)

| [M] _{calc} (Da) | Cross-Link Structure | | StavroX | | | | pLink | | | |
|---|-----------------------------------|--------------------------------------|------------|----------|----------------------------|---------------------------|------------|----------|----------------------------|---------------------------|
| | Peptide 1 | Peptide 2 | <i>m/z</i> | <i>z</i> | [M] _{exp} (Da) | Mass Accuracy (ppm) | <i>m/z</i> | <i>z</i> | [M] _{exp} (Da) | Mass Accuracy (ppm) |
| Calmodulin-Melittin Cross-Linked Species | | | | | | | | | | |
| 1159.68 | ¹ GIGAVLK ⁷ | ⁸⁷ EAFR ⁹⁰ | 580.84 | 2 | 1159.68 | 1.7 | 580.84 | 2 | 1159.68 | 8.3 |
| 1730.88 | ¹ GIGAVLK ⁷ | ⁷⁸ DTDSEEER ⁸⁶ | 577.96 | 3 | 1730.88 | 1.2 | 577.96 | 3 | 1730.88 | 9.8 |

Figure 6.6: (a) Venn Diagram of MS/MS verified EDC cross-linked species identified by each software (StavroX and pLink) and manual method, two methods (two region overlap), all methods (center region overlap); (b) The calculated monoisotopic mass, cross-link peptide sequence (cross-linked residues highlighted in red), *m/z*, experimental monoisotopic mass, and mass accuracy for species identified by StavroX and pLink.

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

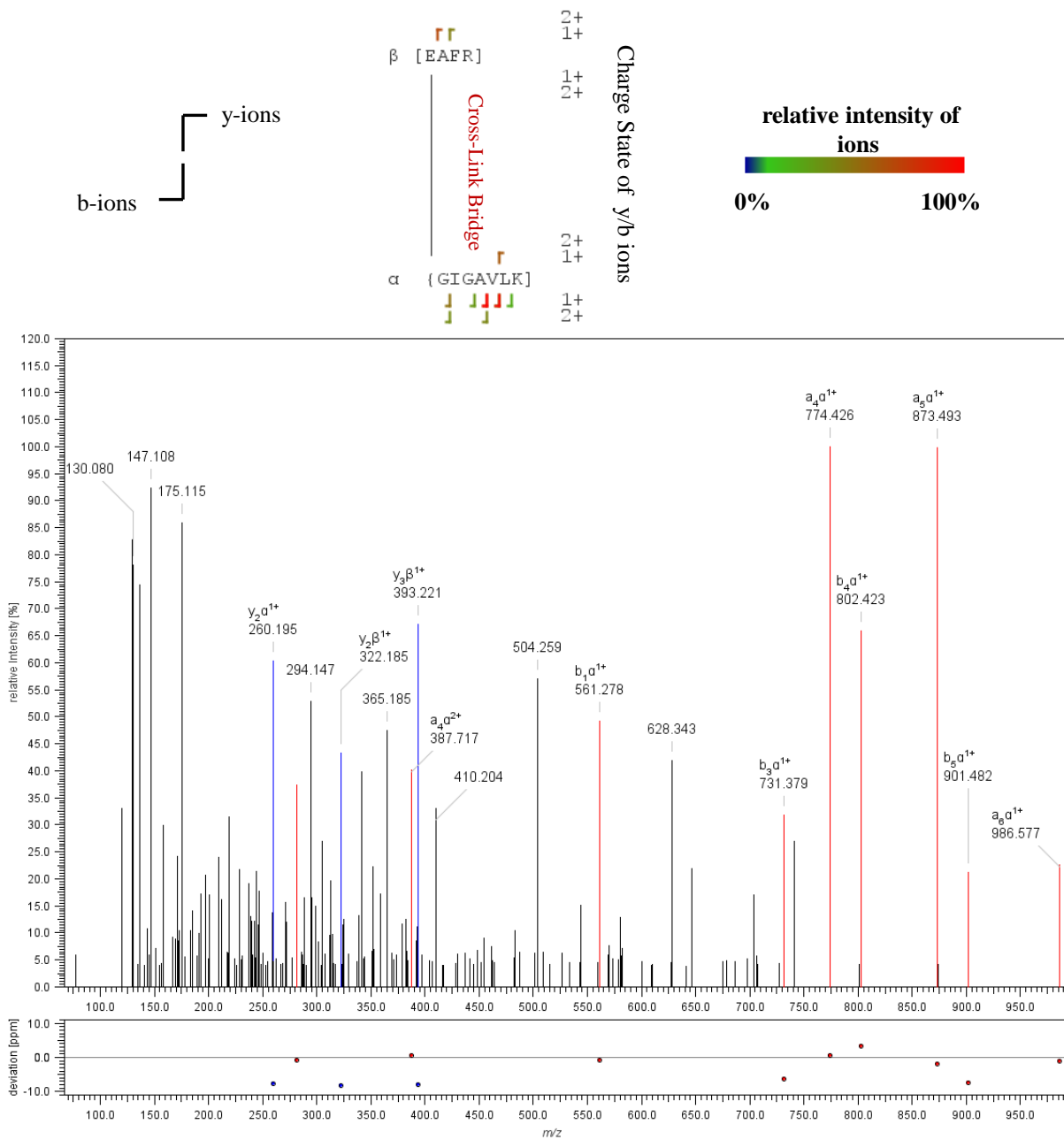


Figure 6.7: An example of an EDC cross-linked species ($m/z = 580.84$, $z = 2$) and its annotated MS/MS spectrum from StavroX

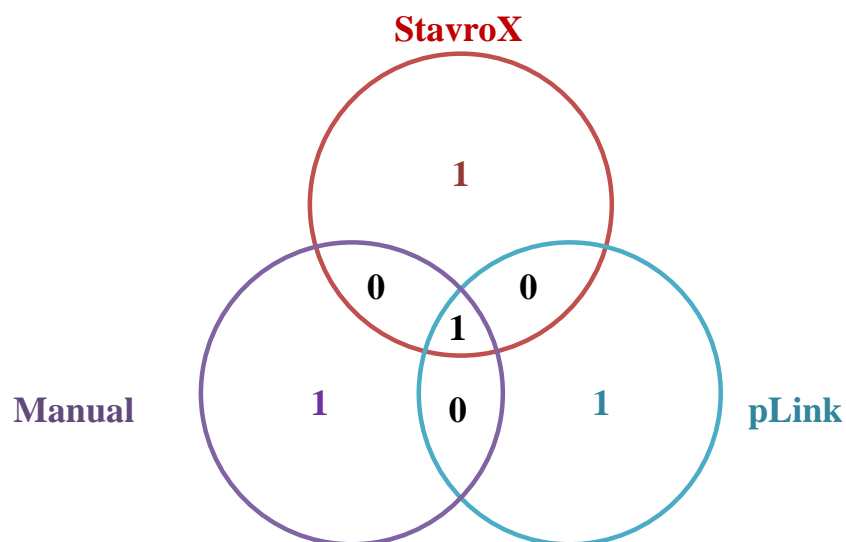
6.5.2.2 SulfoDST

Figure 6.8 lists the identified cross-linked species and illustrates the overlap between each StavroX, pLink and manual cross-link identification method. StavroX identified a total of six unique sulfoDST cross-linked species out of which three did not meet the acceptance criteria due to insufficient MS/MS evidence, and three were actual cross-linked species. Out of the three StavroX identified and confirmed cross-linked species, one of them was also identified manually. pLink identified four unique sulfoDST cross-linked species out of which two could not be confirmed due to insufficient MS/MS evidence, and two were actual cross-linked species. Out of two verified pLink identified cross-linked species, one overlapped with the manual identification and one was the unoxidized form of the cross-link identified by StavroX. Figure 6.9 shows the annotated MS/MS provided by StavroX for the cross-linked species with a charge of four at m/z 567.53 that was identified by both StavroX and pLink. Four out of the 13 (31%) and eight out of the 21 (38%) expected backbone fragment ions for the melittin and calmodulin peptide was present, respectively. With sufficient y and b fragment ions for both component peptides with the cross-linker bridge intact, this species was accepted as a cross-link.

The cross-linked species identified only by pLink, the species at m/z 563.54, corresponded to a signal that also appeared in the control sample. The cross-linked species only identified by StavroX possessed two m/z values that were missed by the peak picking software, one of which, a species with a charge of four at m/z 693.35 was mixed with signals of species with a similar m/z and elution time. Thus, these were not identified manually.

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

(a)



(b)

| [M] _{calc} (Da) | Cross-Link Structure | | StavroX | | | | pLink | | | |
|---|-----------------------------------|--|------------|----------|----------------------------|------------------------|------------|----------|----------------------------|------------------------|
| | Peptide 1 | Peptide 2 | <i>m/z</i> | <i>z</i> | [M] _{exp} (Da) | Mass Accuracy (ppm) | <i>m/z</i> | <i>z</i> | [M] _{exp} (Da) | Mass Accuracy (ppm) |
| Calmodulin-Melittin Cross-Linked Species | | | | | | | | | | |
| 2250.11 | ¹ GIGAVLK ⁷ | ⁷⁵ KMKD K DTDSEEEIR ⁸⁶ | | | | | 563.54 | 4 | 2250.15 | 18.0 |
| 2769.36 | ¹ GIGAVLK ⁷ | ⁷⁵ KM ^(ox) KD K DTDSEEEIR ⁸⁶ | 554.88 | 5 | 2769.37 | 4.3 | | | | |
| | | | 693.35 | 4 | 2769.39 | 10.8 | | | | |
| 2524.29 | ¹ GIGAVLK ⁷ | ⁹¹ VFD K DGNGYISAAELR ¹⁰⁶ | 842.44 | 3 | 2524.30 | 4.0 | 842.44 | 3 | 2524.31 | 10.4 |

Figure 6.8: (a) Venn Diagram of MS/MS verified sulfoDST cross-linked species identified by each software (StavroX and pLink) and manual method, two methods (two region overlap), all methods (center region overlap); (b) The calculated monoisotopic mass, cross-link peptide sequence (cross-linked residues highlighted in red), *m/z*, experimental monoisotopic mass, and mass accuracy for species identified by StavroX and pLink. Cross-links that agreed with the manual detection are highlighted in purple.

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

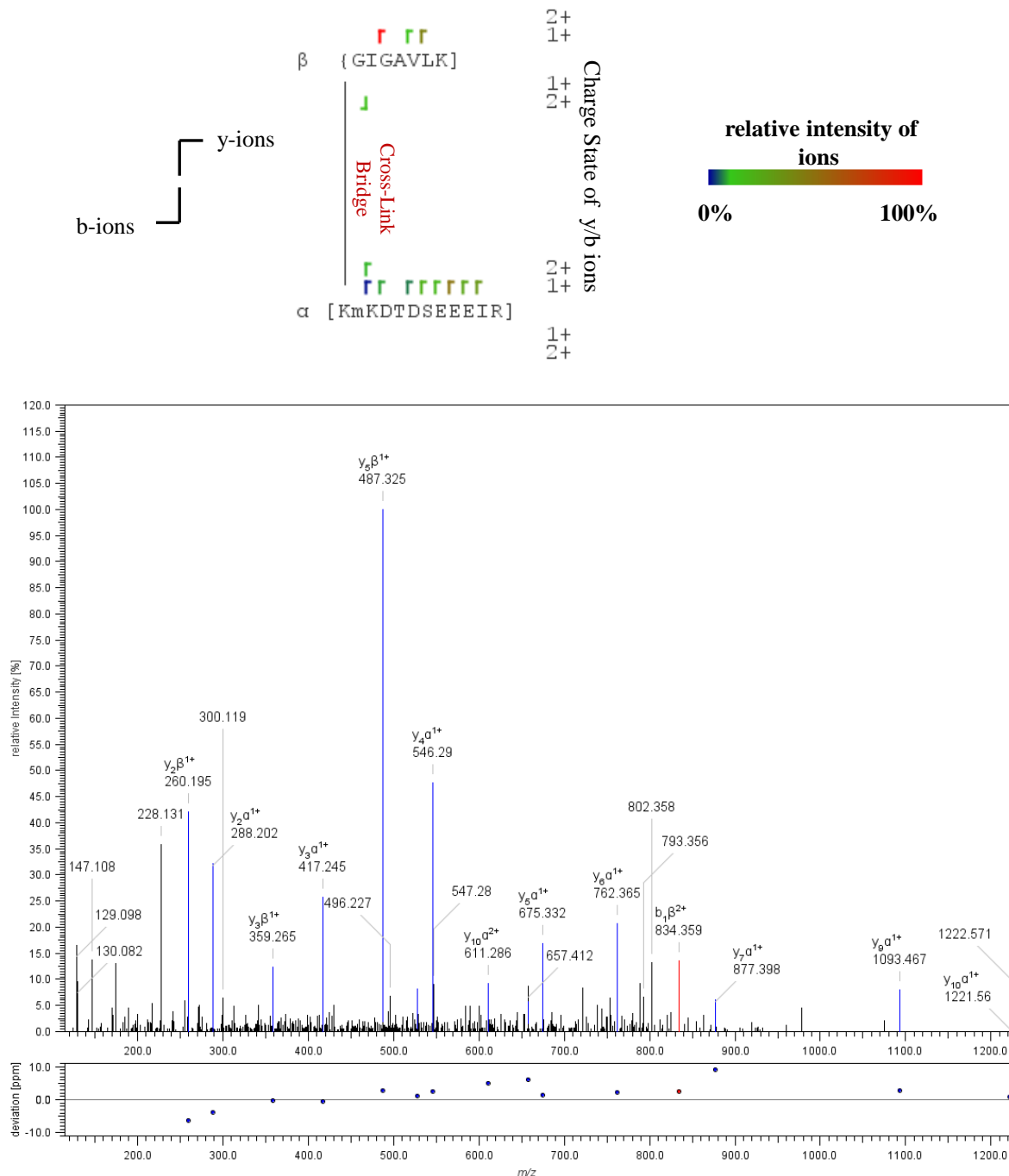


Figure 6.9: An example of a sulfodST cross-linked species ($m/z = 567.53$, $z = 4$) and its annotated MS/MS spectrum from StavroX

6.5.2.3 BS³

Figure 6.10 lists the identified cross-linked species and illustrates the overlap between each StavroX, pLink and manual cross-link identification method. A total of 79 unique BS³ cross-linked species were identified with StavroX out of which 58 did not satisfy the acceptance criteria due to insufficient MS/MS evidence, three had insufficient MS evidence to confirm their presence, one was a peptide with a missed cleavage, and 17 were actual cross-linked species. Out of the 17 StavroX identified and confirmed cross-linked species, three of them were also identified manually. A total of nine unique BS³ cross-linked species were identified with pLink out of which one did not possess sufficient MS/MS evidence for its confirmation, four did not possess corresponding MS signals, and four were actual cross-linked species. Out of the four pLink identified and confirmed cross-linked species, two of them were also identified manually and all four were also identified by StavroX.

Figure 6.11 shows an example of an annotated MS/MS spectrum of a cross-linked species identified by StavroX. Ten out of the 27 (31%) and six out of the 29 (22%) expected backbone fragment ions for the melittin and calmodulin peptide were present, respectively. With sufficient y and b fragment ions for both component peptides with the cross-linker bridge intact, this species was accepted as a cross-link.

There were 14 cross-linked species identified by the software that were not found manually, which are listed in Figure 6.10b (shaded in white). Six out of these 14 matched signals appeared in the control sample ($m/z = 600.88$, $z = 2$), corresponded to a single peptide ($m/z = 569.56$, $z = 4$) or matched an impossible cross-linked species ($m/z = 406.25$, $z = 3$; $m/z = 716.37$, $z = 3$; $m/z = 579.30$, $z = 4$; $m/z = 556.50$, $z = 5$). The remaining eight out of 14 cross-linked species were not found manually due to the absence of these signals in the peak lists generated by SNAP. Out of these eight signals, five of them appeared as incorrectly assigned masses ($m/z = 721.70$, $z = 3$; $m/z = 573.55$, $z = 4$; $m/z = 773.73$, $z = 3$; $m/z = 649.83$, $z = 4$; $m/z = 890.74$, $z = 4$) and it is unknown why SNAP missed the remaining three signals (706.36 , $z = 4$; 678.53 , $z = 5$; and 575.30 , $z = 4$).

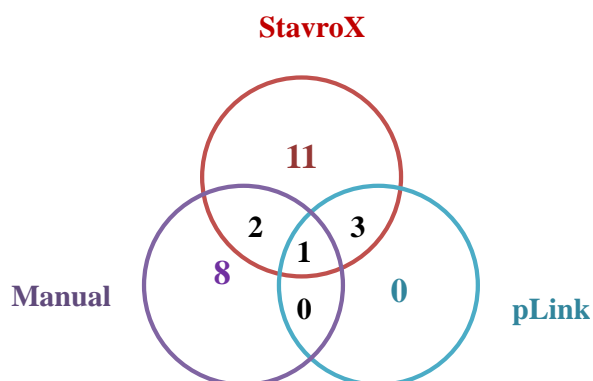
StavroX failed to find eight of the manually identified cross-linked species and

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

pLink only identified one manually identified BS³ cross-linked species. All of the cross-linked species detected by the software supported the same or similar cross-linking sites as with manual detection: melittin G1 and K23 to calmodulin K75/K77, and calmodulin K75 to K94 (indicated by residues highlighted in red in Figure 6.10b) . Cross-linking between melittin K21 and calmodulin K94 was only detected by the software, however this is a similar region to the cross-linking between melittin K23 and calmodulin K94, detected manually. Since similar cross-linking sites were identified by both manual and automated methods, the existence of these cross-linking sites was further validated. Furthermore, it suggests that both methods identified true cross-links.

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

(a)



(b)

| [M] _{calc} (Da) | Cross-Link Structure | | StavroX | | | | pLink | | | |
|---|---|---|---------|---|----------------------------|------------------------|--------|---|----------------------------|------------------------|
| | Peptide 1 | Peptide 2 | m/z | z | [M] _{exp} (Da) | Mass Accuracy (ppm) | m/z | z | [M] _{exp} (Da) | Mass Accuracy (ppm) |
| Calmodulin-Melittin Cross-Linked Species | | | | | | | | | | |
| 1199.74 | ¹ GIGAVLK ⁷ | ⁷⁵ KMK ⁷⁷ | 600.88 | 2 | 1199.75 | 8.3 | | | | |
| 1215.74 | ¹ GIGAVLK ⁷ | ⁷⁵ KM ^(ox) K ⁷⁷ | 406.25 | 3 | 1215.75 | 11.5 | | | | |
| 2076.06 | ²² RKR ²⁴ | ⁷⁵ KMKDSEEEIR ⁸⁶ | 693.03 | 3 | 2076.08 | 7.7 | | | | |
| 2146.10 | ¹ GIGAVLK ⁷ | ⁷⁶ MKDSEEEIR ⁸⁶ | 716.37 | 3 | 2146.11 | 6.5 | | | | |
| 2162.09 | ¹ GIGAVLK ⁷ | ⁷⁶ M ^(ox) KDSEEEIR ⁸⁶ | 721.70 | 3 | 2162.11 | 6.9 | | | | |
| 2274.20 | ¹ GIGAVLK ⁷ | ⁷⁵ KMKDSEEEIR ⁸⁶ | 569.56 | 4 | 2274.22 | 10.1 | 759.07 | 3 | 2274.21 | 4.4 |
| 2290.19 | ¹ GIGAVLK ⁷ | ⁷⁵ KM ^(ox) KDSEEEIR ⁸⁶ | 573.55 | 4 | 2290.21 | 7.9 | | | | |
| 2318.13 | ¹ GIGAVLK ⁷ | ⁷⁵ KM ^(ox) KDSEEEIR ⁸⁶ | 773.73 | 3 | 2318.20 | 6.9 | | | | |
| 2548.37 | ¹ GIGAVLK ⁷ | ⁹¹ VFDKDGNGYISAAELR ¹⁰⁶ | 850.46 | 3 | 2548.38 | 4.3 | 840.46 | 3 | 2518.38 | 11.9 |
| 2595.31 | ²² RKR ²⁴ | ⁷⁵ KM ^(ox) KDSEEEIREAFR ⁹⁰ | 649.83 | 4 | 2595.32 | 6.9 | | | | |
| 2777.45 | ¹ GIGAVLK ⁷ | ⁷⁵ KMKDSEEEIREAFR ⁹⁰ | 556.50 | 5 | 2777.48 | 9.7 | 556.49 | 5 | 2777.45 | 0 |
| 2793.44 | ¹ GIGAVLK ⁷ | ⁷⁵ KM ^(ox) KDSEEEIREAFR ⁹⁰ | 559.69 | 5 | 2793.47 | 7.5 | | | | |
| 2821.44 | ¹ GIGAVLK ⁷ | ⁷⁵ KM ^(ox) KDSEEEIREAFR ⁹⁰ | 706.36 | 4 | 2821.45 | 4.3 | | | | |
| 3558.97 | ⁸ VLTTGLPALISWIKR ²² | ⁹¹ VFDKDGNGYISAAELR ¹⁰⁶ | 890.75 | 4 | 3558.99 | 4.2 | 890.75 | 4 | 3559.00 | 8.4 |
| Calmodulin-Calmodulin Cross-Linked Species | | | | | | | | | | |
| 2297.19 | ⁷⁵ KMK ⁷⁷ | ⁹¹ VFDKDGNGYISAAELR ¹⁰⁶ | 575.30 | 4 | 2297.21 | 10.0 | | | | |
| 2313.18 | ⁷⁵ KM ^(ox) K ⁷⁷ | ⁹¹ VFDKDGNGYISAAELR ¹⁰⁶ | 579.30 | 4 | 2313.20 | 12.5 | | | | |
| 3387.63 | ⁷⁵ KM ^(ox) DTSEEEIR ⁸⁶ | ⁹¹ VFDKDGNGYISAAELR ¹⁰⁶ | 678.53 | 5 | 3387.66 | 8.0 | | | | |

Figure 6.10: (a) Venn Diagram of MS/MS verified BS³ cross-linked species identified by each software (StavroX and pLink) and manual method, two methods (two region overlap), all methods (center region overlap); (b) The calculated monoisotopic mass, cross-link peptide sequence (cross-linked residues highlighted in red), m/z, experimental monoisotopic mass, and mass accuracy for species identified by StavroX and pLink. Cross-links that agreed with the manual detection are highlighted in purple.

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

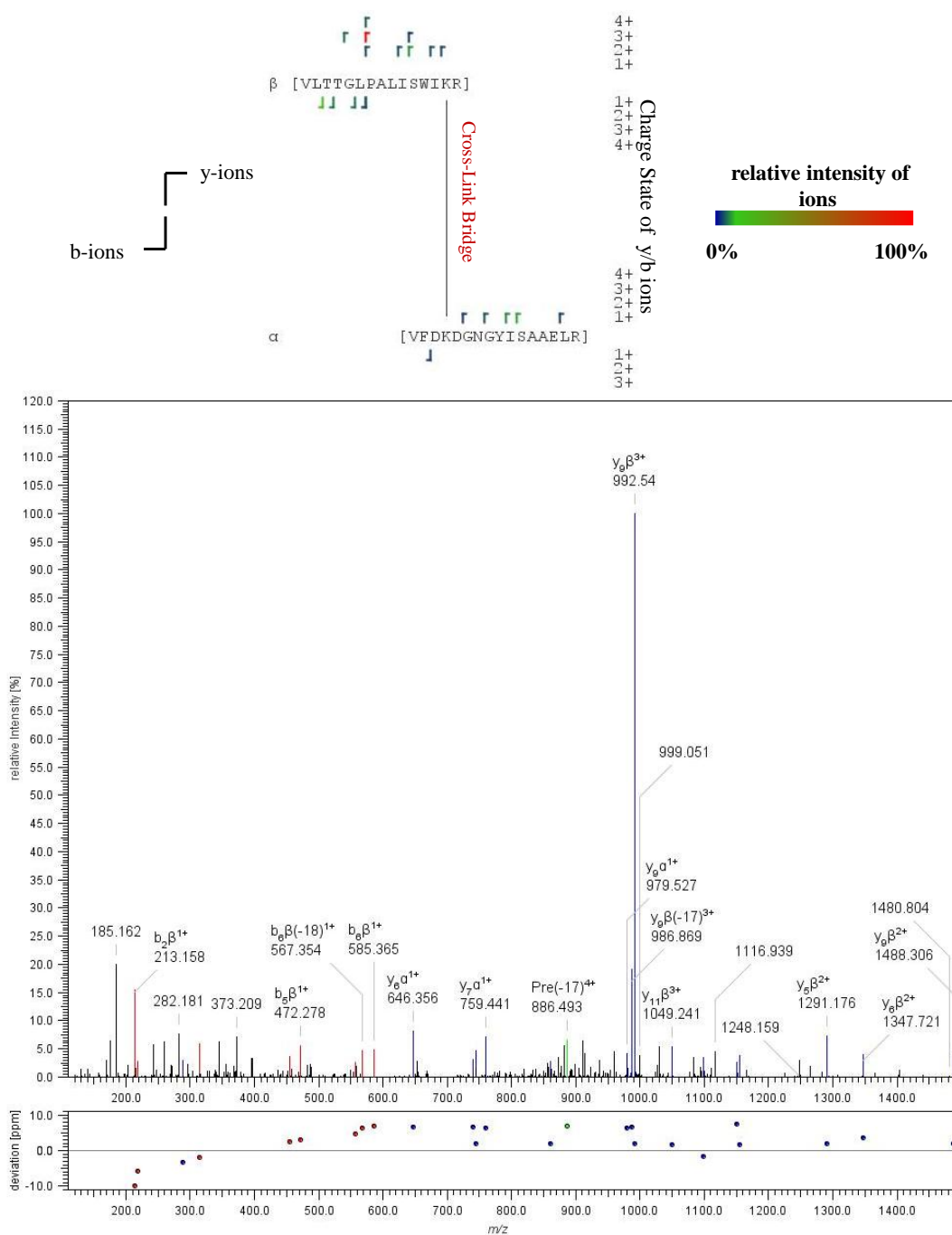


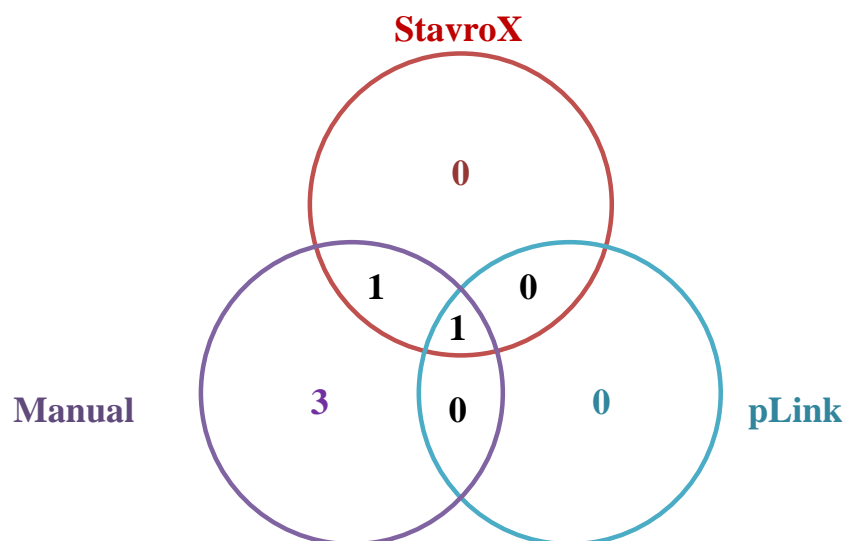
Figure 6.11: An example of a BS³ cross-linked species ($m/z = 890.75$, $z = 4$) and its annotated MS/MS spectrum from StavroX; In the cross-linked structure shown on top, the “m” in the peptide sequence refers to M^(ox), i.e. an oxidized M residue.

6.5.2.4 SulfoEGS

Figure 6.12 lists the identified cross-linked species and illustrates the overlap between each StavroX, pLink and manual cross-link identification method. StavroX identified a total of 12 unique sulfoEGS cross-linked species out of which ten had insufficient MS/MS evidence to confirm their presence, and two were actual cross-linked species. pLink identified five unique sulfoEGS cross-linked species out of which two did not possess sufficient MS/MS evidence for confirmation, two did not produce sufficient MS signals and one was an actual cross-linked species. The pLink identified sulfoEGS cross-linked species was also identified by StavroX and manually. All StavroX identified sulfoEGS cross-linked species were also identified manually. Figure 6.13 shows an example of an annotated MS/MS spectrum generated by StavroX for a cross-linked species identified by StavroX, pLink and manually. Six out the 29 (21%) of the expected backbone fragment ions for each calmodulin peptide was present, respectively. With sufficient y and b fragment ions for both component peptides with the cross-linker bridge intact, this species was accepted as a cross-link.

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

(a)



(b)

| [M] _{calc} (Da) | Cross-Link Structure | | StavroX | | | | pLink | | | |
|---|---|--|---------|---|----------------------------|-------------------------------|--------|---|----------------------------|-------------------------------|
| | Peptide 1 | Peptide 2 | m/z | z | [M] _{exp} (Da) | Mass Accurac y (ppm) | m/z | z | [M] _{exp} (Da) | Mass Accurac y (ppm) |
| Calmodulin-Melittin Cross-Linked Species | | | | | | | | | | |
| 2539.21 | ²² RKR ²⁴ | ⁷⁶ M ^(ox) KDTDSEEEIREAFR ⁹⁰ | 635.81 | 4 | 2539.23 | 8.7 | | | | |
| Calmodulin-Calmodulin Cross-Linked Species | | | | | | | | | | |
| 3733.78 | ⁹¹ VFDKDGNGYISAAELR ¹⁰⁶ | ⁹¹ VFDKDGNGYISAAELR ¹⁰⁶ | 934.46 | 4 | 3733.80 | 5.4 | 934.46 | 4 | 3733.82 | 11.2 |

Figure 6.12: (a) Venn Diagram of MS/MS verified sulfoEGS cross-linked species identified by each software (StavroX and pLink) and manual method, two methods (two region overlap), all methods (center region overlap); (b) The calculated monoisotopic mass, cross-link peptide sequence (cross-linked residues highlighted in red), m/z , experimental monoisotopic mass, and mass accuracy for species identified by StavroX and pLink. Cross-links that agreed with the manual detection are highlighted in purple.

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

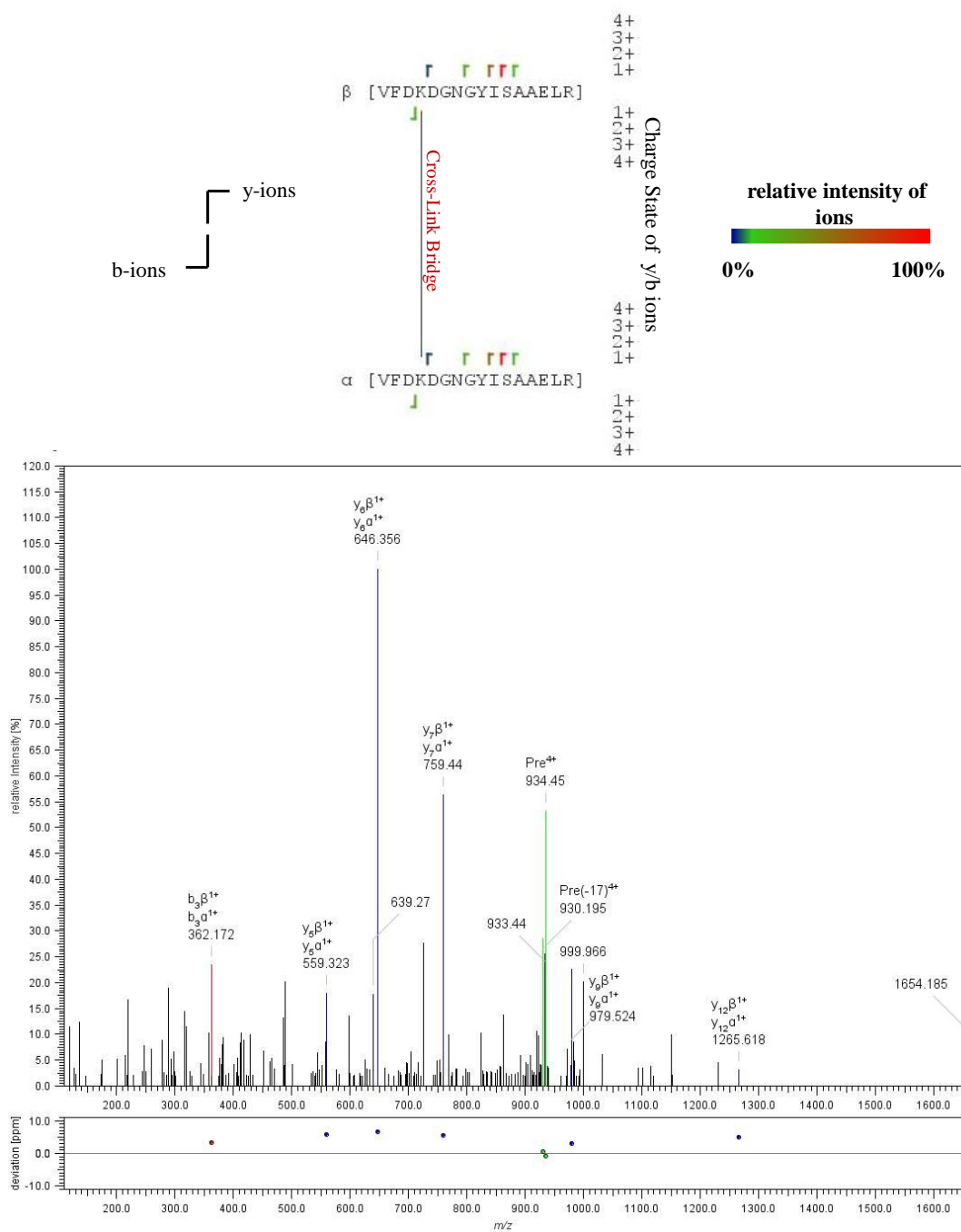


Figure 6.13: An example of a sulfoEGS cross-linked species ($m/z = 936.46$, $z = 4$) and its annotated MS/MS spectrum from StavroX

6.5.3 Formaldehyde

StavroX and pLink identified cross-linked species were shown to identify cross-linked species from established cross-linkers, as expected since these software programs were designed for such cross-linking chemistry. These software programs were used to search PFA cross-linked samples by inputting PFA-specific reactive sites, bridge composition and modifications to test whether they can be applied to PFA as well.

Figure 6.14 lists the identified cross-linked species and illustrates the overlap between each MeroX, StavroX, pLink and manual cross-link identification method. MeroX identified a total of nine unique PFA cross-linked species out of which seven did not satisfy the acceptance criteria due to insufficient MS/MS evidence, and two were true cross-linked species. StavroX identified a total of 11 unique PFA cross-linked species out of which seven corresponded to single peptides, two were not confirmed due to insufficient MS/MS evidence, and two were true cross-linked species. All MeroX and StavroX identified and confirmed PFA cross-linked species were also identified manually. pLink identified two unique PFA cross-linked species out of which one had insufficient MS/MS evidence for its verification, and one was an actual cross-linked species that was also identified manually and by MeroX.

To illustrate PFA cross-linked species identified by the software that possessed insufficient MS/MS evidence for confirmation, the proposed structure and MS/MS spectrum of the highest scoring example, which was identified by StavroX, is shown in Figures 6.15 and 6.16, respectively. For the PFA cross-linked species with a charge of six at m/z 701.01 (Score = 101), the following structure was proposed by StavroX: $^{75}\text{KMK}^{77\wedge 75}\text{KM}^{(ox)}\text{K}^{(+12)}\text{DTDSEEEIR}^{(+12)}\text{EAFR}^{(+12)}\text{VFDDKDGNGYISAAELR}^{126}$. For this calmodulin-calmodulin cross-linked species, StavroX identified only eight type 3 ions (Iy2 to Iy7, Ib7 and Ib18) for $^{75}\text{KM}^{(ox)}\text{K}^{(+12)}\text{DTDSEEEIR}^{(+12)}\text{EAFR}^{(+12)}\text{VFDDKDGNGYISAAELR}^{126}$ and one type 2 ion (Iy17+II), which is only 9 % of the total expected number of fragment ions. In addition, there was no fragment ion evidence for $^{75}\text{KMK}^{77}$. Therefore, this cross-linked species was classified as having insufficient MS/MS

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

evidence to confirm its presence.

The MS/MS fragment ion evidence assignment of the software vs manual identification was compared. Figure 6.17 shows an example of an annotated MS/MS spectrum generated by MeroX for a cross-linked species identified by MeroX ($m/z = 606.58$, $z = 4$), pLink ($m/z = 808.44$, $z = 3$) and manually ($m/z = 484.46$, $z = 5$), see Figure 4.32 for MS/MS spectrum) with the following structure: $^{91}\text{VFDKDGNGYISAAELR}^{106}\wedge^1\text{GIGAVLK}^7$. In this case, no type 2 ions were identified by MeroX and manual analysis of this cross-linked species. A series of unmodified type 3 for the calmodulin peptide $^{91}\text{VFDKDGNGYISAAELR}^{106}$ and modified type 3 ions for the melittin peptide $^1\text{GIGAVLK}^7$ along with a type 1 ion for the +12 Da modified melittin peptide were identified by both the MeroX and manual analysis to confirm the cross-linked species. In both analysis methods, cross-linking was localized to melittin G1 and calmodulin Y99. MeroX identified 34% and 55% of the expected fragment ions for $^{91}\text{VFDKDGNGYISAAELR}^{106}$ and $^1\text{GIGAVLK}^7$, respectively, which is sufficient MS/MS fragment ion evidence to confirm this cross-linked species. However, MeroX missed the type 3 ions Iy1, Iiy1, Iib2, and Iib5 that were identified manually and the manual analysis did not identify the type 3 ion Iy6 that the MeroX had found. Nevertheless, this illustrated the reliability of both methods to confirm a PFA cross-linked species and the respective cross-linking sites. Figure 6.18 shows the second cross-linked species identified by MeroX, StavroX and manually (see Figure 4.33 for MS/MS spectrum), a species with a charge of 5 at m/z 588.52, with the following structure: $^{87}\text{EAFR}^{(+12)}\text{VFDKDGNGYISAAELR}^{106}\wedge^1\text{GIGAVLK}^7$. For this cross-linked species, both type 2 and type 1 ions were not identified by MeroX, StavroX or manually. MeroX identified 22% and 45% of the fragment ion evidence of $^{87}\text{EAFR}^{(+12)}\text{VFDKDGNGYISAAELR}^{106}$ and $^1\text{GIGAVLK}^7$, respectively, which is sufficient fragment ion evidence to confirm this cross-linked species. Although all MS/MS fragment ions identified by the MeroX were also identified manually, MeroX missed the following fragment type 3 ions that were identified manually: Iy1, Ib2, Ib3, Ib5 – Ib10 + 12, Ib13 +12 Iiy1, Iiy3, Iib2+12 and Iib6+12. However, MeroX was still able to localize the extra modification on calmodulin R90 and provide evidence for a cross-link between calmodulin Y99 and melittin G1, which is consistent with the manual analysis.

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

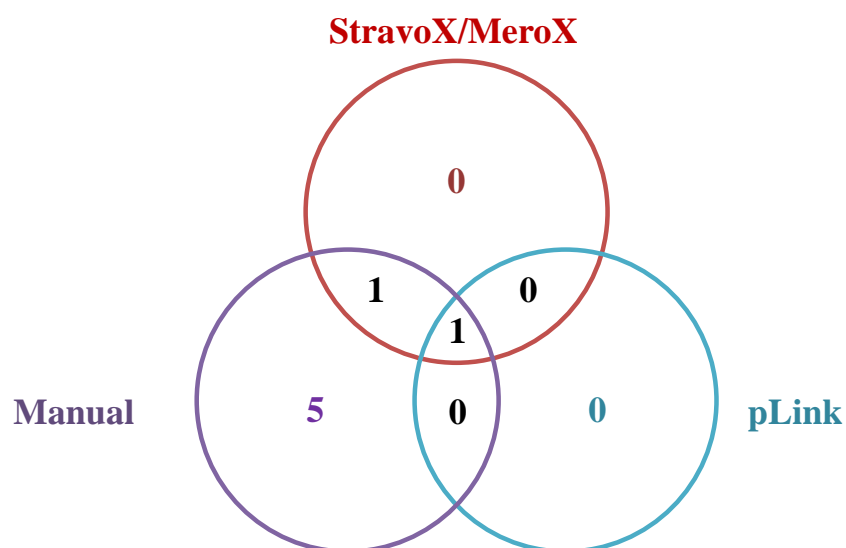
As shown in Figure 6.14, only two out of the seven manually detected cross-linked species were identified by MeroX and StavroX, and only one cross-linked species was identified by pLink. Therefore, manual detection appeared to provide a more reliable method to discover a more exhaustive list of PFA cross-linked species. There were also several disadvantages when using MeroX/StavroX and pLink for PFA cross-linked sample data. First, other established cross-linked species searches lasted on the scale of minutes whereas PFA cross-linked species searches lasted on the scale of hours to a day for both pLink and MeroX/StavroX, which significantly slowed down the process of analysis. Furthermore, for larger data sets and protein complexes producing more PFA cross-linked candidates, it is expected that the length of search time will increase. Another disadvantage is being unable to account for the possibility of methylol modifications on cross-linked species using MeroX/StavroX. Although PFA could produce both Schiff Base and methylol modifications, when these were considered, the software would run for an entire day and then shut down before completion. Therefore, only Schiff base modifications could be considered to obtain any results. For pLink searches, only Schiff base modifications were also considered because it also could not complete the search if methylol modifications were also considered. For the calmodulin-melittin system, none of the manually identified cross-linked species contained methylol modifications, so the comparison to software identification could still provide some insight into the whether software can identify PFA cross-linking. However, for more complex protein systems that may have more reactive sites and therefore be more extensively modified, it would be important to consider both types of modifications. As mentioned earlier, the scenario of both fragmented (i.e. MeroX searches) and intact (i.e StavroX searches) cross-linker bridges under CID could not be considered simultaneously to obtain sufficient fragment ions for the identification of PFA cross-linked species. This demonstrates that the current software can identify PFA cross-linked species in simple protein systems such as the calmodulin-melittin complex, however, it cannot replace manual analysis since the majority of the cross-linked species were missed by the software. It would be beneficial for future automated methods tailored specifically for PFA to simultaneously search for fragment ions from both intact and fragmented cross-linker bridges and handle the large number of theoretical candidates produced from all possible PFA

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

modifications. These would be required to improve PFA cross-link identification and analyze larger protein systems.

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

(a)



(b)

| [M] _{calc} (Da) | Cross-Link Structure | | MeroX/StavroX | | | pLink | | |
|-----------------------------|-----------------------------------|--|----------------------|----------------------------|---------------------------|----------------------|----------------------------|---------------------------|
| | Melittin Peptide | Calmodulin Peptide | m/z | [M] _{exp} (Da) | Mass Accuracy (ppm) | m/z | [M] _{exp} (Da) | Mass Accuracy (ppm) |
| 2422.29 | ¹ GIGAVLK ⁷ | ⁹¹ VFDKDGNGYISAAELR ¹⁰⁶ | 606.58 ⁴⁺ | 2422.32 | 12.4 | 808.44 ³⁺ | 2422.32 | 12.4 |
| 2937.53 | ¹ GIGAVLK ⁷ | ⁸⁷ EAF ^R VFDKDGNGYISAAELR ¹⁰⁶ | 588.52 ⁵⁺ | 2937.6 | 23.8 | | | |

Figure 6.14: (a) Venn Diagram of MS/MS verified PFA cross-linked species identified by each software (MeroX/StavroX and pLink) and manual method, two methods (two region overlap), all methods (center region overlap); (b) The calculated monoisotopic mass, cross-link peptide sequence (cross-linked residues highlighted in red), m/z , experimental monoisotopic mass, and mass accuracy for species identified by MeroX/StavroX and pLink. Cross-links that agreed with the manual detection are highlighted in purple.

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

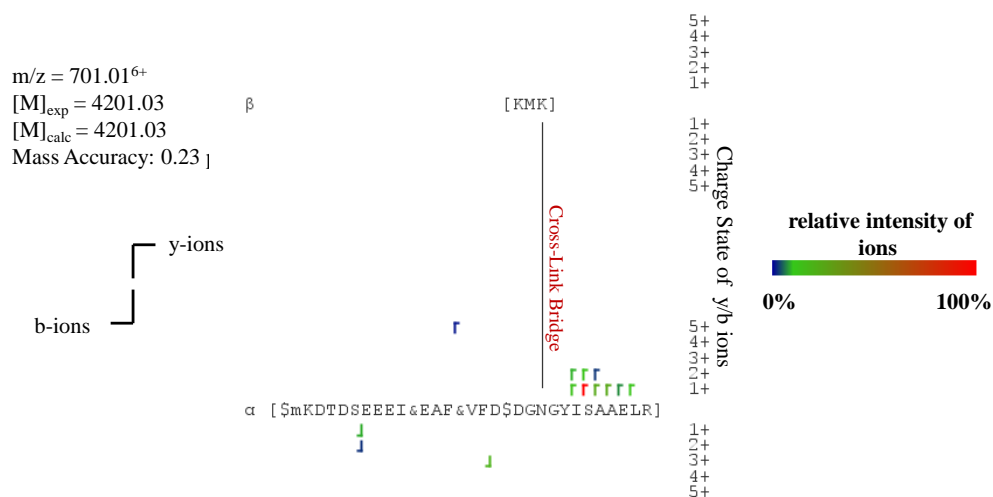


Figure 6.15: The proposed structure of PFA cross-linked species ($m/z = 701.01$, $z = 6$) with insufficient MS/MS evidence from StavroX; In the cross-linked structure sequence, $R^{(+12)}$, $K^{(+12)}$, and $M^{(ox)}$ are denoted as “&”, “\$” and “m”, respectively.

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

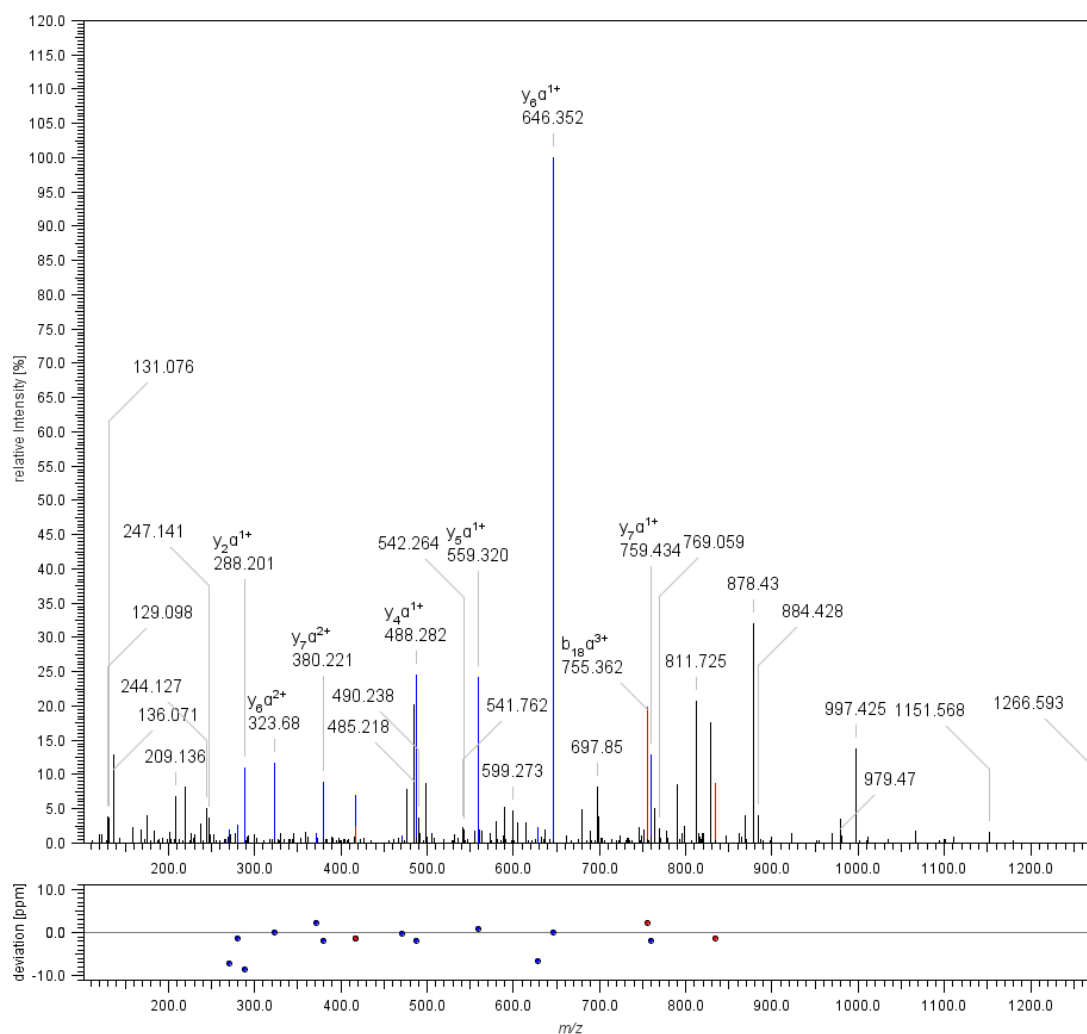


Figure 6.16: The MS/MS spectra of PFA cross-linked species ($m/z = 701.01$, $z = 6$) with insufficient MS/MS evidence from StavroX; In the cross-linked structure sequence, $R^{(+12)}$, $K^{(+12)}$, and $M^{(ox)}$ are denoted as “&”, “\$” and “m”, respectively.

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

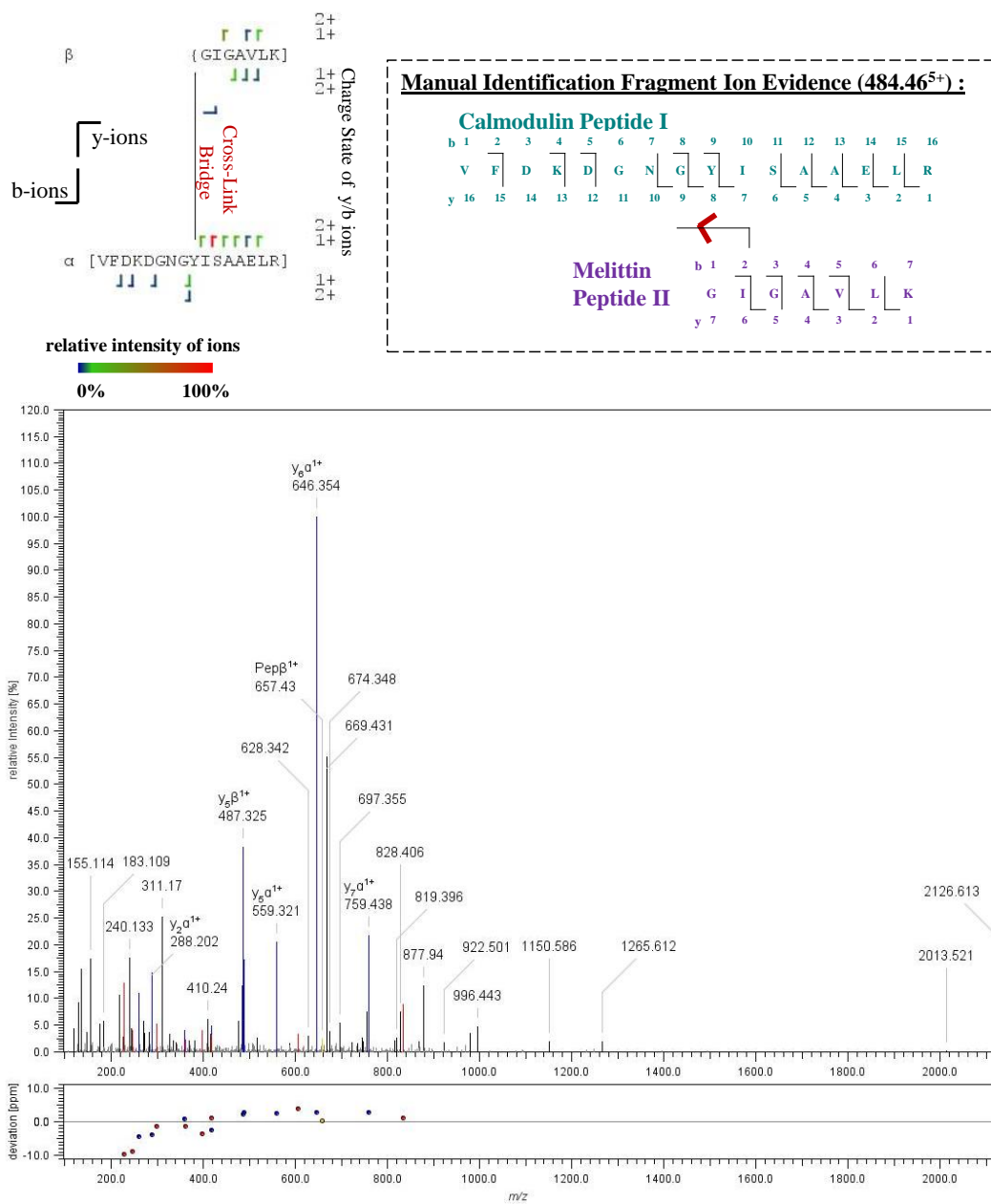


Figure 6.17: An example of a confirmed PFA cross-linked species ($m/z = 606.58$, $z = 4$) and its annotated MS/MS spectrum from MeroX. In the dotted box, the fragment ion evidence of the same cross-linked structure ($m/z = 484.46$, $z = 5$)

6.5. Manual versus Software Identification of Calmodulin-Melittin Cross-links

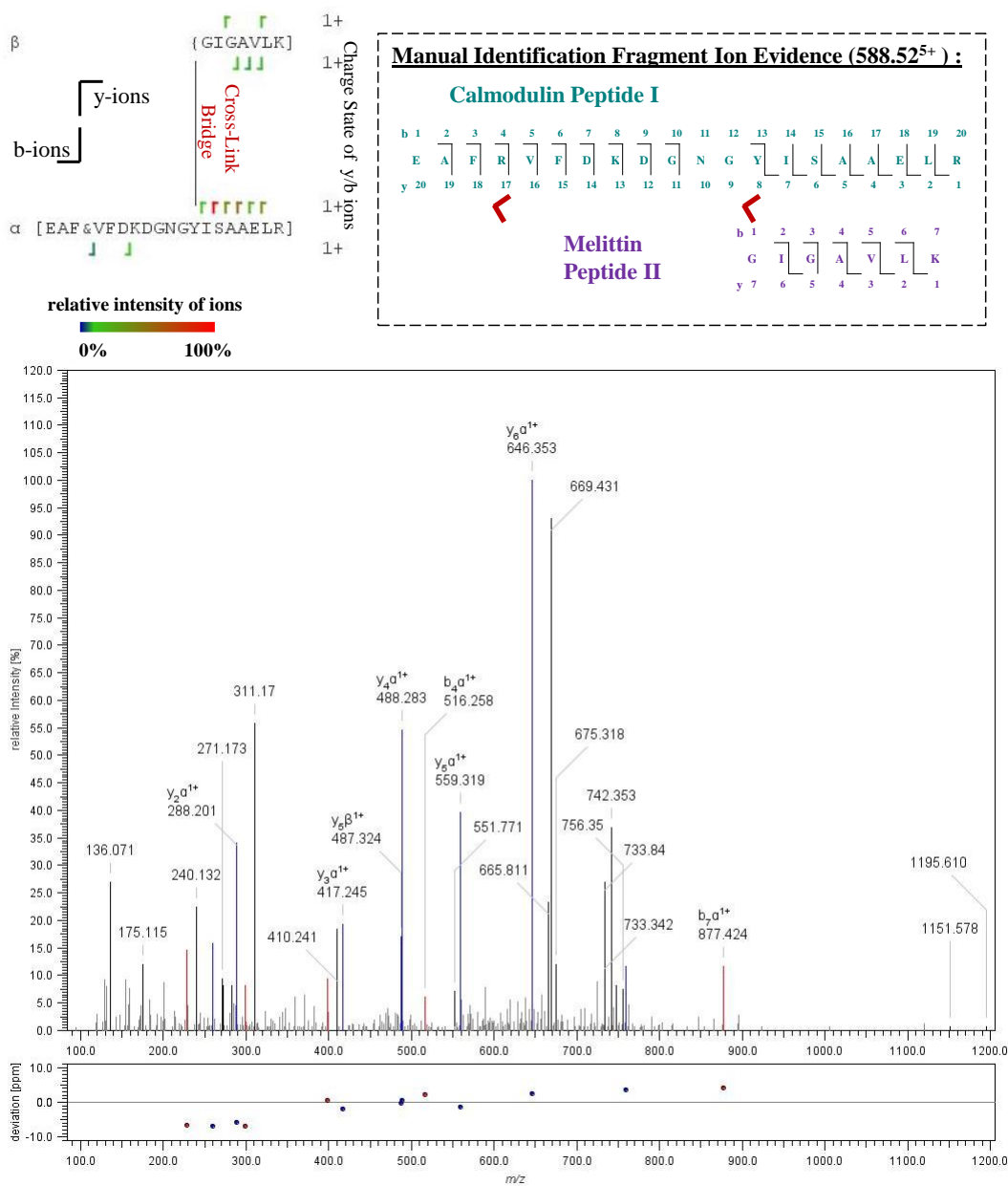


Figure 6.18: An example of a confirmed PFA cross-linked species ($m/z = 588.52$, $z = 5$) and its annotated MS/MS spectrum from MeroX; In the dotted box, the fragment ion evidence of the same cross-linked structure ($m/z = 588.52$, $z = 5$) is shown. In the cross-linked structure sequence, $R^{(+12)}$ is denoted as “&”.

6.5.4 Inconsistencies between Manual vs. Software Cross-link identification

There were two, two, and 14 cross-linked species identified exclusively by the software and not manually for EDC, sulfoDST, and BS³ cross-linked samples, respectively. One reason is that signals of the software identified cross-links also appeared in the control sample, corresponded to single peptides or matched impossible cross-linked species. These had to be filtered to distinguish between cross-linked species and non-cross linked species at the MS level during manual analysis. This was performed to eliminate the probability of identifying false positives and to reduce the search space to candidate cross-links that had the highest probability of existing before the lengthy process of manually inspecting their MS/MS spectra. Secondly, signals were absent in the peak lists generated by SNAP. For sulfoDST and EDC, this was mainly due to overlapping signals at the MS level in which the monoisotopic signal was too ambiguous for SNAP to identify it. For BS³, five out of the eight absent signals were incorrectly assigned by SNAP. Although there were errors in the SNAP peak picking algorithm used for the manual analysis, the software was unable to identify a significant number of manually identified species i.e. seven, one, eight, three, and five out of seven, two, 11, five, and seven total manually identified cross-linked species for EDC, sulfoDST, BS³, sulfoEGS and PFA, respectively. This study suggested that manual analysis was able to identify more cross-links for EDC, PFA and sulfoEGS, cross-linkers which also produced the most number of reaction products and MS candidates. For sulfoDST, which produced very few cross-links in general, it is ambiguous whether software or manual analysis is more suitable. For BS³, each method of analysis identified several cross-links with a very minimal overlap, suggesting that a combination of both software and manual detection would maximize cross-link identification. Overall, the cross-linking software was shown to handle more established cross-linking chemistry and that it would be beneficial for future cross-linking software to be designed to tackle more complicated cross-linking reactions that occur with EDC and to a larger extent, PFA.

6.6 Limitations in the MS/MS Analysis of Formaldehyde Cross-linked Ribonuclease-S

In the RNaseS system, the SDS-PAGE (section 3.2.5) analysis of PFA cross-linked RNaseS provided evidence that the yield of PFA cross-linking was consistent with the expected yield of the bound RNaseS complex structure derived from its K_d . Furthermore, percent trypsin cleavage trends (section 5.2) observed in the PFA treated RNaseS peptides agreed with the accessibility and trypsin cleavage observed in the bound RNaseS complex in literature [160, 167, 168]. Therefore, this supported PFA's ability to preserve the non-covalent interaction in RNaseS. However, complications in the analysis of the RNaseS system did not facilitate the MS/MS localization and identification of its PFA cross-linked species even though PFA crosslinked species in calmodulin-melittin system were successfully verified via MS/MS manually. As shown in chapter 3, MaxQuant identified unmodified calmodulin-melittin peptides that provided a 100% sequence coverage in both control and PFA treated samples, and identified a sufficient list of PFA modified calmodulin-melittin peptides (30 unique peptides). This provided assurance to continue with subsequent cross-link peptide analysis. Although there were more PFA reactive sites in the RNaseS system than in the calmodulin-melittin system, a lack of PFA modified peptides was observed in the RNaseS system. Only a small subset of PFA modified peptides (four unique peptides) were identified via MaxQuant in the PFA treated RNaseS sample. Furthermore only 12 unique peptides (27 peptides total) were identified in the PFA treated RNaseS sample, whereas 71 unique peptides (250 total peptides) were identified in the PFA treated calmodulin-melittin sample. This reduced amount of data supporting the identified RNaseS peptides species results in a lower confidence in the subsequent analyses based on these peptides. In addition, the S-peptide was not identified in the PFA treated RNaseS sample, and therefore, information regarding S-peptide modification sites for verifying cross-link localization or understanding the accessibility of its sites in the RNaseS complex structure was unavailable.

Whether the lack of identified peptides was due to the structural restraints of RNaseS or the MS-detection is not clear. The average length and charge-state of peptides detected by MS and identified by MaxQuant in both samples of PFA-

6.6. *Limitations in the MS/MS Analysis of Formaldehyde Cross-linked Ribonuclease-S*

treated RNaseS and calmodulin-melittin system were similar, i.e. ~18 residues and +3 peptides, respectively. However, since RNaseS (pI ~ 9) is more basic than calmodulin-melittin (pI~4), it is possible that longer, more basic peptides formed, which may not have been efficiently detected by MS. The MS-detection of longer peptides is difficult since they contain more bonds to fragment and would thus experience a reduced fragmentation efficiency. In addition, the higher charge-states of long, basic peptides would decrease their monoisotopic signal intensity in MS. As shown in sections 5.2 and 3.3.2.2, trypsin cleavage is hindered by both structural restraints and chemical modification of cleavage sites in PFA treated RNaseS samples. Therefore, the production of longer, missed cleaved peptides is more of an issue in cross-linked samples. Crosslinks between these long peptides would result in species that would be even more difficult to detect via MS and thus whether long, basic RNaseS peptides tend to form must be tested before further cross-link peptide analyses. It is also hypothesized that since the sample contained RNaseA and RNaseS complexes, produced from subtilisin cleavage at different RNaseA residues, that the resulting peptide mixture was highly complex. Thus the MS intensities of all peptide species were dispersed among many species, reducing the signal intensities of all peptides. The complexity and insufficient number of peptides identified in the RNaseS system prevented subsequent manual identification of cross-linked peptides. In addition, the cross-link identification software failed to identify PFA cross-linked RNaseS species. The PFA treated RNaseS sample data was inputted into MeroX and pLink software. MeroX ran for several days before freezing and pLink was unable to identify any PFA cross-linked species. This occurred when considering both methylol and Schiff Base, only Schiff Base and no modifications as variable modifications. The inability to analyze the PFA treated RNaseS samples via these software programs further supports the complexity of these samples. Overall, PFA was able to capture the RNaseS transient interaction and an approach to tackle the complexity of the RNaseS sample can facilitate future cross-link identification.

Chapter 7

Conclusion and Future Outlook

For the first time, MS/MS confirmed PFA cross-linking was identified and localized in a biologically relevant, non-covalent protein system under mild *in vivo*-like conditions. This has bridged the gap between previous model systems and *in vivo* protein-protein interaction analyses with PFA. Furthermore, PFA cross-linking was placed in context of other established cross-linkers that have already accomplished cross-link localization in biologically relevant systems. In addition, this thesis reveals the promise in examining protein complexes through different perspectives from cross-linkers of different lengths and reactivity.

7.1 Clarification of Formaldehyde Cross-linking Reaction Chemistry

This work provided insight into the reactivity, mechanism and equilibrium of PFA cross-linking in the calmodulin-melittin system under mild *in vivo*-like PFA reaction conditions (1% PFA, 6 hr incubation and physiological pH). It was confirmed that PFA is primarily reactive with N-terminal and K residues, and to a lesser extent R residues in the first modification step of the PFA cross-linking reaction, thus demonstrating the specificity of the PFA modification reaction under mild *in vivo*-like reaction conditions. This is consistent with previous studies that examined protein modification [36]. For the first time, PFA cross-linking localized between K and Q, between N-terminal and Q residues, and between two R residues, was observed in protein systems under mild *in vivo*-like cross-linking conditions. PFA cross-linking localized between N-terminal and Y residues under mild *in vivo*-like cross-linking conditions is consistent with previous PFA protein cross-linking experiments [50]. Combining the modifications localized in PFA modified pep-

7.1. Clarification of Formaldehyde Cross-linking Reaction Chemistry

tides from chapter 3 and the cross-links localized in chapter 4, the mechanism of calmodulin-melittin cross-linking was clarified such that calmodulin K77 and R90, and melittin G1 and R24 were modified in the first step of the reaction and formed cross-links with calmodulin Q3, Q8, Y99 and R126 in the second step of the reaction. The equilibrium for the formation of methylol, Schiff Base and cross-linked products was measured at localized PFA cross-linking sites using the PFA modified, unmodified and cross-linked peptides, respectively identified in this study. The relative abundance of sites with Schiff Base, methylol and PFA cross-links demonstrated that the formation of Schiff Base modifications was favored over the formation of PFA cross-linking. This was consistent with the low relative percent abundance (< 10%) of PFA cross-linking observed at the identified PFA cross-linking sites. It is hypothesized that this is a reflection of the specificity of PFA cross-linking, i.e. capturing only relevant interactions that are within close proximity specific to the calmodulin-melittin system. Exploring more protein complexes can confirm whether PFA cross-linking yield is directly a function of the protein interaction it is capturing.

In this current study, it was assumed that all +12 Da modifications corresponded to Schiff Base modifications. Even though intrapeptide cross-linking would produce the same +12 Da mass shift in MS and MS/MS spectra, this has been only identified in proteins and peptides that have undergone long PFA reaction times (days to weeks) so far[85, 87]. However, to distinguish between Schiff Base and intrapeptide cross-linking, NaCNBH₃ could be utilized. NaCNBH₃ selectively reduces Schiff bases without affecting the carbonyls on the protein or aldehydes at the physiological pH used in these formaldehyde-based studies. Also, it is resistant to hydrolysis at pH values greater than 3, making it a favorable reducing reagent. In dimethyl labeling, NaCNBH₃ is successfully used to reduce Schiff Base modifications to dimethyl substituents which produces an overall mass shift of +28 Da [65, 66]. Therefore, by introducing NaCNBH₃ to the reaction mixture after cross-linking proteins with PFA, Schiff Base modifications should theoretically be reduced to +28 Da mass modifications while cross-links should appear with +12 Da mass shift. This technique was used previously to distinguish intrapeptide and Schiff Base modifications in model peptides using long PFA incubation times [86] and has yet to be applied to protein systems under mild *in*

7.1. Clarification of Formaldehyde Cross-linking Reaction Chemistry

vivo-like PFA cross-linking conditions.

In the calmodulin-melittin system, there were no cysteine residues, which are potentially highly reactive with PFA. It was observed in previous studies that PFA treated model peptides, in which lysine remained unmodified, formed methylol modifications on cysteine residues within 10 minutes[85]. Even though cysteines exhibited the highest reactivity in model peptides using comparable cellular cross-linking reaction times, reactivity of these side chains in most model proteins were not observed since they are usually occupied in disulfide bonds. However, the majority of intracellular proteins intrinsically exist with reduced, free cysteines due to the redox environment inside the cells. Additionally, many functions are governed by the reduction and oxidation of cysteines, which is referred to as a “disulfide bond switch” [179, 180]. Thus, confirming formaldehyde reactivity with cysteines in proteins could unlock a powerful *in vivo* application of formaldehyde to examine these cysteine-involved cellular protein interactions.

Another aspect of PFA chemistry to examine in the future is the reversal mechanism of PFA cross-linking. In PFA treated tissues and cells, heating is used to reverse cross-links and restore antigen reactivity for immunodetection to enable effective protein extraction and MS-analysis [55, 67, 69–71]. Also, previous spectroscopy experiments with PFA cross-linked RNase A suggested that protein structure is recovered after the reversal of PFA cross-linking [71]. Being able to recover protein structures for subsequent analysis is especially vital for unlocking information stored in clinical tissues extensively cross-linked with PFA. However, the mechanism of reversal remains unknown to support these findings. Furthermore, understanding the reversal of PFA cross-linking could clarify the possibility of whether PFA cross-linking is reversed prior to MS analysis during sample preparation. This is crucial to assess whether PFA cross-links are lost through this processing.

7.2 Trypsin Digestion Efficiency of Formaldehyde Reaction Products

The trypsin cleavage efficiency at PFA modified and cross-linked residues was revealed, which is vital to effectively predict products formed from PFA cross-linking. In this study, it was demonstrated that trypsin will most likely not cleave after PFA modified residues, which was supported by the observation of Schiff Base and methylol modifications localized only on internal K and R residues. The formation of PFA induced Schiff Bases and methylol groups reduce the electrophilic nature of K residues, which explains why trypsin is less likely to cleave after PFA modified K residues. Two PFA cross-linked species were identified with cross-links on terminal K and R residues, supporting that trypsin can cleave after PFA cross-linking. This may have occurred since PFA cross-linking does not affect the electrophilicity of K and R residues and forms small cross-link bridges that could potentially fit in the active site of trypsin.

Nevertheless, more PFA cross-links must be identified to confirm this. It also may be useful to explore trypsin cleavage efficiency after cross-linked residues as a function of different neighboring amino acids. Furthermore, determining the rate of trypsin cleavage versus PFA modification and cross-link formation on K and R residues would better define the efficiency of trypsin cleavage at these sites. It is hypothesized that incubating PFA treated proteins with trypsin for an extended time may result in trypsin itself reacting with Schiff Base modifications. Histidine, which is a key component of trypsin's catalytic triad in its active site, has been observed to form cross-links in small peptides and between single amino acids under long PFA incubation times (week). However, histidine cross-linking has yet to be observed in protein systems under mild PFA cross-linking conditions [86, 87]. This would be another aspect to consider which may help explain several unknown products detected in the PFA cross-linked samples.

7.3 Establishing Tandem Mass Spectrometric Fragmentation Rules for Formaldehyde Cross-link Identification

This present study has moved the field closer to establishing fragmentation rules to confirm the presence of PFA cross-linked species by manually comparing its MS/MS spectra to those of other cross-linkers (EDC, sulfoDST, BS³,and sulfoEGS) with established fragmentation patterns. Common fragmentation patterns were revealed across most EDC, sulfoDST, BS³,sulfoEGS cross-linked species. In general, these cross-linked species produced type 1 ions of the smaller peptide, type 2 ions and usually no type 3 ions. With these cross-linkers, the backbone fragmentation of the larger peptide was shown to be more extensive than in the smaller peptide. The consistency of these observations with literature [52] validated the manual inspection of the MS/MS spectra of cross-linked species. It was demonstrated that the labile PFA cross-linker bridge produced unique fragment ion evidence unlike other cross-linkers. PFA cross-linked species were shown to mainly produce type 3 ions and type 1 of the smaller peptide. Although, diagnostic fragment ions for NHS ester cross-linkers were observed due to the additional sites of fragmentation introduced by their cross-linker bridge, these were not observed for PFA and EDC cross-linkers. A list of guidelines was prepared to summarize the manual assessment of the MS/MS signals of each cross-linked species to confirm cross-linked species based on the scoring methods of established cross-link identification software[120, 165, 166].

It was concluded that the evidence for localizing PFA cross-linking is much more ambiguous than with other cross-linkers, especially since multiple structural isomers from multiple possible cross-linking and modification sites and thus, requiring DOM calculations to identify the highest probable cross-linking sites. Also, both N-terminal and R residues can form either a PFA modification or cross-link, introducing additional ambiguity in the PFA cross-linking mechanism. However, for PFA cross-linked species, it was observed that the peptide that was modified in the first step of the reaction retained the +12 Da bridge post CID fragmentation for cross-linking of K to Q, N-terminal to Q and N-terminal to Y residues.

7.4. *The Structural Characterization of Ca²⁺-free calmodulin-melittin via Comprehensive Cross-linking*

If verified by exploring an increased number of cross-linking sites in different sequence environments, this could potentially serve as a useful tool to distinguish between modification and cross-linking sites and clarify cross-linking mechanisms. In general, cross-linking localization is more easily clarified with the presence of type 2 ions, where the cross-linker bridge remains intact. However, in this study, type 2 ions were rarely observed upon the fragmentation of PFA cross-linked species, in contrast to other established cross-linker species. This may be due to the PFA cross-link bridge bonds being more labile than other cross-linker bridge bonds between the protein and cross-linker. Therefore, it is hypothesized that using a low energy fragmentation method such as ETD/ECD may generate supplementary fragment ion evidence such as these type 2 ions to clarify vague cross-linking sites and provide a more confident confirmation of cross-linking. Another method that could be used to improve cross-link identification would be to adjust the CID energy to produce different degrees of fragmentation at the backbone and cross-linker to produce more extensive fragment ion evidence. However, both these fragmentation methods must be optimized based on the the cross-linked species sequence composition and length. Many standard cross-linkers have also been synthesized with MS/MS cleavable bonds to efficiently detect cross-linked species via MSⁿ of component peptides [181]. This approach could be exploited with PFA, given its intrinsic semi-cleavable bridge in CID. For example, the MSⁿ of type 1 ions, arising from exclusive fragmentation at the cross-linker bridge, could verify sequences of component peptides. MSⁿ of type 2 ions, with intact cross-linked bridges, and type 3 ions, with both the bridge and backbone fragmented, could potentially aid in the localization of cross-linking and modification sites.

7.4 The Structural Characterization of Ca²⁺-free calmodulin-melittin via Comprehensive Cross-linking

For the first time, cross-links generated by PFA, along with the four established cross-linkers (EDC, sulfoDST, BS³, and sulfoEGS) were identified via MS/MS to capture the weak interaction between Ca²⁺-free calmodulin and melittin. It was

7.4. *The Structural Characterization of Ca²⁺-free calmodulin-melittin via Comprehensive Cross-linking*

shown that distance constraints imposed by PFA, sulfoEGS, and BS³ cross-linked species supported the existence of calmodulin in its binding conformation. Furthermore, it was revealed that PFA captured the uniformity of the bound calmodulin conformation in the complex. This was supported by the consistency of cross-link site localization and the high relative abundance of PFA intramolecular calmodulin cross-links from producing more of the same product. In this study, the binding orientation of calmodulin and melittin captured by each cross-linker illustrated that PFA produced the most favorable distance constraints for this particular protein complex. NHS ester cross-linking was shown to produce long and ambiguous distance constraints that may not be suitable for such small protein complexes. EDC cross-links did not support distance constraints within the known structure of calmodulin. Calmodulin-melittin conformations derived by EDC cross-linking were inconsistent with PFA cross-linking despite their comparable size. PFA cross-linked species were identified in a transient protein complex for the first time, supporting PFA's ability to stabilize weak interactions. Distance constraints imposed by PFA agreed with recent NMR and other spectroscopy experiments, demonstrating its potential for the structural characterization of proteins.

The percent of intact, bound RNaseS and calmodulin-melittin calculated from their dissociation constants agreed with the percent cross-linking yield of PFA, EDC, and sulfoEGS cross-linkers, and PFA, EDC and BS³ cross-linkers, respectively, determined in this study. This correlation supports that cross-linking can capture true biologically relevant weak interactions. Even after SDS-PAGE denaturation, this work revealed that cross-linking partially preserves these non-covalent interactions, which are lost in the absence of cross-linking.

Although the percent trypsin cleavage was examined to use accessibility to assess protein structure after SDS-PAGE denaturation, it may be useful to examine trypsin cleavage in solution without the SDS denaturation of cross-linked intact proteins. A limited proteolysis experiment can be applied to study the accessibility of cross-linked versus non-cross linked structures. This can verify that the RNaseS and calmodulin-melittin interactions were preserved by cross-linking and were lost without cross-linking, as observed in this study with SDS-PAGE and MS-analysis. Furthermore, quantifying the yield of cross-linking in intact proteins using top-down proteomics and MS can provide a more precise derivation of the dissociation

constant. The ratio of the total peak area of cross-linked protein complex versus non-cross linked protein components can be derived from the MS spectrum to determine the dissociation constant, as done previously [182]. Overall, connecting this intact protein analysis to the results obtained from the MS/MS at the peptide level can build a more confident and comprehensive picture of these non-covalent protein complexes.

With the potential of PFA cross-linking to characterize protein structures exemplified in this study, the next step would be to explore a wide variety of different protein structures of increasing size and complexity. In addition, more cross-linking sites need to be localized for imposing distance constraints to increase the confidence of derived structures. Adding the structural information conveyed from PFA cross-linking to the structural analysis from other cross-linkers for different protein complexes can expand the field of structural proteomics. Achieving this with the combination of NMR and molecular docking can serve as a powerful tool for the analysis of protein structure and dynamics.

7.5 Revealing Complexity of Cross-linking Reaction Mixtures

The complexity associated with PFA versus established cross-linkers, (EDC, sulfoDST, BS³, and sulfoEGS) which vary in size and reactivity, to examine the calmodulin-melittin and RNaseS via MS was illustrated in this study. It was shown that PFA produced the most number of MS-candidate cross-linked species, followed by EDC and then, NHS ester cross-linkers. Although NHS cross-linkers produced fewer MS-candidate cross-linked species due to their specific reactivity, this work revealed that the cross-linker bridge length was proportional to the number of reaction products produced.

This study has illustrated that MS/MS confirmed cross-linked species represent less than 15% of MS candidate cross-linked species, i.e a small subset of the cross-linking reaction products, especially PFA. In fact, cross-link identification in proteins was shown to strongly depend on the high resolution, mass accuracy and dynamic range of current generation mass spectrometers such as the Bruker Impact

7.5. *Revealing Complexity of Cross-linking Reaction Mixtures*

II QqTOF that was available for this study November, 2014. The ABI QStar XL, an outdated mass spectrometer built in the early 2000s, was used prior to the introduction of the Bruker Impact II and it failed to produce sufficient MS/MS evidence to confirm cross-linking in the calmodulin-melittin model system. In fact, the QStar could not match even the unmodified peptide identification capabilities of the Bruker Impact II. Nevertheless, sensitive detection of cross-linked species must also be combined with a reliable software to accurately determine their monoisotopic masses in complex, cross-linked peptide mixtures. Although the Bruker's SNAP peak picking algorithm was observed to be the most reliable, even SNAP did not consistently assign correct monoisotopic masses across all cross-linked samples and this issue was more prominent for PFA, EDC and sulfoEGS cross-linkers that produced more reaction products. With the manual identification process, it was possible to correct these incorrect assignments, by examining MS signals that also corresponded to +1 Da of the candidate cross-linked monoisotopic mass. However, this is not a feasible solution especially for more complex systems since even for the calmodulin-melittin system, there were 49, 90, 6, 16, and 49 incorrect assignments for EDC, PFA, sulfoDST, BS³, and sulfoEGS cross-linkers, respectively.

It is hypothesized that reducing the complexity of the reaction mixtures through enrichment techniques may reduce the chances of the software misinterpreting MS signals. For example, strong cation exchange chromatography (SCX) has been successfully implemented to enrich and separate cross-linked species [183–186]. Since trypsin cleaves after basic residues (i.e K and R), each tryptic peptide has at least one basic site at its N-terminus and one basic site at its C-terminus. Cross-linked peptides should have double the number of basic sites that single peptides have. Using an acidic buffer environment such that all ionizable sites are protonated, there will be least a +2 charge on tryptic peptides and +4 charge on cross-linked peptides. Therefore, cross-linked peptides can be separated from single peptides by increasing the salt concentration of the elution buffer to allow species to elute in order of increasing charge. A complication that may occur would be that missed cleaved peptides could potentially also contain the same number of basic sites as cross-linked species and separating these peptides from cross-linked species would be challenging. To further enrich and separate such mixtures, tech-

7.5. *Revealing Complexity of Cross-linking Reaction Mixtures*

niques such as size exclusion chromatography (SEC) or MS-based techniques such as ion mobility MS (IMS) could be used in conjunction with SCX to separate cross-linked peptides from single peptides due to their differences in molecular size and shape. Both SEC and IMS have been successfully utilized previously to enrich and separate cross-links[187, 188].

The RNaseS model system presented a higher level of complexity than the calmodulin-melittin system. The analysis of control RNaseS peptides revealed a mixture of RNaseA, two forms of S-peptides and two forms of S-protein. Even when the presence of only one form of S-protein and S-peptide was considered, the number of possible cross-linked species and candidate cross-linked species was significantly higher than with the calmodulin-melittin system, especially for PFA. The complexity of the RNaseS protein mixture may have caused a lack of peptides identified in PFA treated sample by MS, which prevented further cross-link peptide identification. Therefore, a possible solution would be to use SEC on intact cross-linked RNaseS prior to trypsin digestion and subsequent MS-analysis to separate and enrich cross-linked RNaseS proteins from RNaseA, unmodified RNaseS, and impurities. This would improve the MS-detection of peptides from RNaseS cross-linked species. It was also hypothesized that long, basic peptides from the insufficient trypsin cleavage of PFA treated RNaseS were not efficiently detected by MS. In order to test this hypothesis, multiple proteases that cleave at different sites or chemical cleavage reagents such as cyanogen bromide, which cleave at M residues, can be applied to these samples in parallel with trypsin [189, 190]. If shorter peptides, with lower charge states are produced and the protein sequence coverage is increased, then this hypothesis would be supported. Furthermore, increasing the sequence coverage of longer species using this technique would be important for examining larger, PFA crosslinked RNaseS peptides via MS/MS in the future.

7.6 Comparing Software Versus Manual Cross-link Identification

The manual data analysis procedure, even though extremely tedious, was shown to successfully identify EDC, PFA, sulfoDST, BS3 and sulfoEGS cross-linking in the calmodulin-melittin system. Even with the slightly more complicated RNaseS model system, a drastic increase in the number of MS signals of candidate cross-linked species to be examined manually was revealed. This suggested that a more suitable, automated method is required. However, current automated cross-link identification have not been optimized to identify PFA cross-links and these limitations were highlighted in this work. In the calmodulin-melittin system, the MeroX and pLink cross-link identification software failed to identify the majority of PFA cross-links (five out of seven). In fact, these software searches could not handle the complexity of PFA cross-linked RNaseS samples and thus, a successful search was not even possible with this model system. Even with the calmodulin-melittin system, when the search was set to include both Schiff Base and methylol modifications as variable modifications, the software could not complete the search. Therefore, only Schiff Base modifications could be considered. For more complex and larger proteins with more modification sites, the probability of both methylol and Schiff Base modifications existing increases and cross-linked peptides with methylol modifications would be lost using current cross-linking software. In addition, the scenario of both fragmented and intact cross-linker bridges under CID could not be considered simultaneously to obtain sufficient fragment ions for the identification of PFA cross-linked species. Thus, PFA cross-linked species that produce a mixture of both type 3 and type 2 ions may be lost if neither type of evidence by itself is sufficient to confirm the cross-linked species. Finally, the search time even for the calmodulin-melittin system was on the scale of several hours to a day. For more complex, larger protein systems with more theoretical cross-linked species and larger MS/MS data, it is hypothesized that this would increase drastically.

All together, this revealed that the development of cross-link identification software that can handle large, complex MS/MS data from PFA cross-linked samples, account for PFA's semi CID cleavable cross-linker bridge, and all PFA specific

modifications is necessary. This study provided further insight into the MS/MS fragmentation patterns of PFA and reaction chemistry which can aid in developing software specifically designed for PFA cross-linked species. These can be confirmed and PFA cross-linking reaction chemistry and sequence specific fragmentation rules can be further defined by obtaining MS/MS spectra of an exhaustive selection of PFA cross-linked products from a wide variety of proteins.

7.7 Moving toward Cellular, *in vivo* Systems

This study is the first to identify PFA cross-linking in biologically relevant, transient interactions using *in vivo*-like reaction conditions. However, with complications in the MS and MS/MS analysis of even model proteins, identifying PFA cross-links in cellular proteins still remains a challenge. The digestion of cross-linked cellular proteins generates a highly complex mixture comprised of mostly non-cross linked species in comparison to cross-linked species. Increasing the cross-linker-to-protein ratio for improving the yield is not a favorable option because increasing PFA concentrations will induce the formation of non-soluble complexes that will result in protein loss due to precipitation. In contrast to cross-linking single purified proteins with uniform reactive site concentrations, the yield of *in vivo* cross-linking is dictated by the varying abundance levels of target cross-linked protein complexes and reactive interface concentrations across the cell [191]. Therefore, for cellular model systems the development of suitable enrichment and separation methods is needed to improve the yield and detection of PFA cross-linked species is even more crucial than with model systems. Even though the chemical simplicity of PFA excludes the incorporation of affinity groups for enrichment, as mentioned earlier, various chromatography enrichment techniques such as SCX [183–186] and SEC [187] or MS-based techniques such as IMS[188] that have been utilized with other cross-linkers could also potentially enrich PFA cross-links. Unlike other larger cross-linkers, however, PFA's small cross-linker bridge and mass prevents the use of isotopic labeling or reporter groups to aid in MS detection of its cross-linked peptides. Nevertheless, the information regarding PFA reaction chemistry, MS and MS/MS characteristics of PFA cross-links dis-

7.7. Moving toward Cellular, *in vivo* Systems

covered in this study and expanding on this by exploring an exhaustive selection of protein complexes can be used to develop software for identifying *in vivo* PFA cross-linked species. This study has also demonstrated PFA's potential for characterizing protein structure. Therefore overcoming existing limitations in the MS detection of *in vivo* cross-linked species would transform PFA cross-linking into a useful approach for structural proteomics.

Bibliography

- [1] Haynes, W. *CRC Handbook of Chemistry and Physics, 96th Edition*; CRC Press, 2015.
- [2] Srinivasa, S.; Ding, X.; Kast, J. *Methods* **2015**, *89*, 91–98.
- [3] G.A. Petsko, D. R. , *Protein Structure and Function*,; New Science Press, 2004.
- [4] Freed-Pastor, W. A.; Prives, C. *Genes & development* **2012**, *26*, 1268–1286.
- [5] Wright, J.; Lim, C. *Protein engineering* **2001**, *14*, 479–486.
- [6] Berger, M.; Stahl, N.; Del Sal, G.; Haupt, Y. *Molecular and cellular biology* **2005**, *25*, 5380–5388.
- [7] Terwilliger, T. C.; Bricogne, G. *Acta Crystallographica Section D* **2014**, *70*, 2533–2543.
- [8] Petsko, G. A.; Ringe, D. *Protein structure and function*; New Science Press, 2004.
- [9] Sinz, A.; Arlt, C.; Chorev, D.; Sharon, M. *Protein Science* **2015**, *24*, 1193–1209.
- [10] Fisette, O.; Lagüe, P.; Gagné, S.; Morin, S. *BioMed Research International* **2012**, *2012*.
- [11] Sinz, A. *Journal of Mass Spectrometry* **2003**, *38*, 1225–1237.
- [12] Paramelle D., M. G. S. G.; Martinez, J. *Proteomics* **2013**, *13*, 438–456.
- [13] Chevalier, F. *Proteome Science* **2010**, *8*, 23.

Bibliography

- [14] Beck, S.; Michalski, A.; Raether, O.; Lubeck, M.; Kaspar, S.; Goedecke, N.; Baessmann, C.; Hornburg, D.; Meier, F.; Paron, I. *Molecular & Cellular Proteomics* **2015**, *14*, 2014–2029.
- [15] Angel, T. E.; Aryal, U. K.; Hengel, S. M.; Baker, E. S.; Kelly, R. T.; Robinson, E. W.; Smith, R. D. *Chemical Society Reviews* **2012**, *41*, 3912–3928.
- [16] Steen, H.; Mann, M. *Nature reviews Molecular cell biology* **2004**, *5*, 699–711.
- [17] Shi, S. D.-H.; Drader, J. J.; Freitas, M. A.; Hendrickson, C. L.; Marshall, A. G. *International Journal of Mass Spectrometry* **2000**, *195*, 591–598.
- [18] Nagornov, K. O.; Gorshkov, M. V.; Kozhinov, A. N.; Tsybin, Y. O. *Analytical chemistry* **2014**, *86*, 9020–9028.
- [19] Zubarev, R. A.; Makarov, A. *Analytical chemistry* **2013**, *85*, 5288–5296.
- [20] Lacorte, S.; Fernandez-Alba, A. R. *Mass Spectrometry Reviews* **2006**, *25*, 866–880.
- [21] Abdelhameed, A. S.; Kadi, A. A.; Abdel-Aziz, H. A.; Angawi, R. F.; Attwa, M. W.; Al-Rashood, K. A. *The Scientific World Journal* **2014**, *2014*.
- [22] Campbell, J. L.; Hager, J. W.; Le Blanc, J. Y. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 1672–1683.
- [23] Fenyö, D.; Eriksson, J.; Beavis, R. *Computational Biology*; Springer, 2010; pp 189–202.
- [24] Chamrad, D. C.; Körting, G.; Stühler, K.; Meyer, H. E.; Klose, J.; Blüggel, M. *Proteomics* **2004**, *4*, 619–628.
- [25] Qian, W.-J.; Liu, T.; Monroe, M. E.; Strittmatter, E. F.; Jacobs, J. M.; Kangas, L. J.; Petritis, K.; Camp, D. G.; Smith, R. D. *Journal of proteome research* **2005**, *4*, 53–62.
- [26] Bian, S.; Chowdhury, S. *Austin J Biomed Eng* **2014**, *1*, 1017.

Bibliography

- [27] Belov, M. E.; Damoc, E.; Denisov, E.; Compton, P. D.; Horning, S.; Makarov, A. A.; Kelleher, N. L. *Analytical chemistry* **2013**, *85*, 11163–11173.
- [28] Van Riper, S. K.; de Jong, E. P.; Carlis, J. V.; Griffin, T. J. *Radiation Proteomics*; Springer, 2013; pp 1–35.
- [29] Konijnenberg, A.; Butterer, A.; Sobott, F. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2013**, *1834*, 1239–1256.
- [30] Serpa, J. J.; Makepeace, K. A.; Borchers, T. H.; Wishart, D. S.; Petrotchenko, E. V.; Borchers, C. H. *Journal of proteomics* **2014**, *100*, 160–166.
- [31] Serpa, J. J.; Patterson, A. P.; Pan, J.; Han, J.; Wishart, D. S.; Petrotchenko, E. V.; Borchers, C. H. *Journal of proteomics* **2013**, *81*, 31–42.
- [32] Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R. *Molecular & Cellular Proteomics* **2010**, *9*, 1634–1649.
- [33] López-Alonso, e. a., Jorge P. *Bioconjugate chemistry* **2009**, *20*, 1459–1473.
- [34] Sriswasdi, S.; Harper, S. L.; Tang, H.-Y.; Speicher, D. W. *Journal of proteome research* **2014**, *13*, 898–914.
- [35] Slusarewicz, P.; Zhu, K.; Hedman, T. *Journal of Materials Science: Materials in Medicine* **2010**, *21*, 1175–1181.
- [36] Toews, J.; Rogalski, J. C.; Kast, J. *Analytica chimica acta* **2010**, *676*, 60–67.
- [37] Sutherland, B. W.; Toews, J.; Kast, J. *Journal of Mass Spectrometry* **2008**, *43*, 699–715.
- [38] Holding, A. N. *Methods* **2015**,
- [39] Petrotchenko, E. V.; Borchers, C. H. *Advances in protein chemistry and structural biology* **2014**, *95*, 193.

Bibliography

- [40] Bandeira, N. *BioTechniques* **2007**, *42*, 687–695.
- [41] Mayne, S. L.; Patterton, H.-G. *Briefings in bioinformatics* **2011**, *12*, 660–671.
- [42] Zheng, C.; Yang, L.; Hoopmann, M. R.; Eng, J. K.; Tang, X.; Weisbrod, C. R.; Bruce, J. E. *Molecular & Cellular Proteomics* **2011**, *10*, M110–006841.
- [43] Weisbrod, C. R.; Chavez, J. D.; Eng, J. K.; Yang, L.; Zheng, C.; Bruce, J. E. *Journal of proteome research* **2013**, *12*, 1569–1579.
- [44] Chavez, J. D.; Weisbrod, C. R.; Zheng, C.; Eng, J. K.; Bruce, J. E. *Molecular & Cellular Proteomics* **2013**, *12*, 1451–1467.
- [45] Bian, S.; Chowdhury, S. *Austin J Biomed Eng* **2014**, *1*, 1017.
- [46] Hyung, S.-J.; Ruotolo, B. T. *Proteomics* **2012**, *12*, 1547–1564.
- [47] Rappsilber, J. *Journal of structural biology* **2011**, *173*, 530–540.
- [48] Walzthoeni, T.; Leitner, A.; Stengel, F.; Aebersold, R. *Current opinion in structural biology* **2013**, *23*, 252–260.
- [49] Sinz, A. *Expert review of proteomics* **2014**, *11*, 733–743.
- [50] Ding, X. A method to characterize formaldehyde cross-linking in proteins by mass spectrometry. Ph.D. thesis, University of British Columbia, 2011.
- [51] Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M. *Journal of the American Society for Mass Spectrometry* **2003**, *14*, 834–850.
- [52] Santos, L. F.; Iglesias, A. H.; Gozzo, F. C. *Journal of Mass Spectrometry* **2011**, *46*, 742–750.
- [53] Iglesias, A. H.; Santos, L. F.; Gozzo, F. C. *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 557–566.
- [54] Shi, S.-R.; Cote, R. J.; Taylor, C. R. *Journal of Histochemistry & Cytochemistry* **1997**, *45*, 327–343.

Bibliography

- [55] O'Leary, T.; Fowler, C.; Evers, D.; Mason, J. *Biotechnic & Histochemistry* **2009**, *84*, 217–221.
- [56] Aggerbeck, H.; Heron, I. *Biologicals* **1992**, *20*, 109–115.
- [57] Petre, J.; Pizza, M.; Nencioni, L.; Podda, A.; De Magistris, M.; Rappuoli, R. *Developments in biological standardization* **1995**, *87*, 125–134.
- [58] Rappuoli, R. *Vaccine* **1994**, *12*, 579–581.
- [59] Orlando, V.; Strutt, H.; Paro, R. *Methods* **1997**, *11*, 205–214.
- [60] Varshavsky, A. J.; Sundin, O.; Bohn, M. *Cell* **1979**, *16*, 453–466.
- [61] Kim, T. H.; Ren, B. *Annu. Rev. Genomics Hum. Genet.* **2006**, *7*, 81–102.
- [62] Klockenbusch, O. J. E., Cordula; Kast, J. *Analytical and bioanalytical chemistry* **2012**, *404*, 1057–1067.
- [63] Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Analytical chemistry* **1996**, *68*, 850–858.
- [64] Hsu, J.-L.; Huang, S.-Y.; Chow, N.-H.; Chen, S.-H. *Analytical chemistry* **2003**, *75*, 6843–6852.
- [65] She, Y.-M.; Rosu-Myles, M.; Walrond, L.; Cyr, T. D. *Proteomics* **2012**, *12*, 369–379.
- [66] Ji, C.; Li, L. *Journal of proteome research* **2005**, *4*, 734–742.
- [67] Maes, E.; Broeckx, V.; Mertens, I.; Sagaert, X.; Prenen, H.; Landuyt, B.; Schoofs, L. *Amino Acids* **2013**, *45*, 205–218.
- [68] Klockenbusch, C.; Kast, J. *BioMed Research International* **2010**, *2010*.
- [69] Steiner, C.; Ducret, A.; Tille, J.-C.; Thomas, M.; McKee, T. A.; Rubbia-Brandt, L.; Scherl, A.; Lescuyer, P.; Cutler, P. *Proteomics* **2014**, *14*, 441–451.
- [70] Ralton, L. D.; Murray, G. I. *Journal of clinical pathology* **2011**, jcp–2010.

Bibliography

- [71] Rait, V. K.; O’Leary, T. J.; Mason, J. T. *Laboratory investigation* **2004**, *84*, 292–299.
- [72] Layh-Schmitt, G.; Podtelejnikov, A.; Mann, M. *Microbiology* **2000**, *146*, 741–747.
- [73] Vasilescu, J.; Guo, X.; Kast, J. *Proteomics* **2004**, *4*, 3845–3854.
- [74] Bai, Y.; Markham, K.; Chen, F.; Weerasekera, R.; Watts, J.; Horne, P.; Wakutani, Y.; Bagshaw, R.; Mathews, P. M.; Fraser, P. E. *Molecular cellular proteomics* **2008**, *7*, 15–34.
- [75] Müller, V. S.; Jungblut, P. R.; Meyer, T. F.; Hunke, S. *Proteomics* **2011**, *11*, 2124–2128.
- [76] Jeon, A. H. W.; Schmitt-Ulms, G. *Chemical Proteomics*; Springer, 2012; pp 231–246.
- [77] Herzberg, C.; Weidinger, L. A. F.; Dörrbecker, B.; Hübner, S.; Stülke, J.; Commichau, F. M. *Proteomics* **2007**, *7*, 4032–4035.
- [78] Müller, V. S.; Tschauer, K.; Hunke, S. *JoVE (Journal of Visualized Experiments)* **2013**, e50810–e50810.
- [79] Watts, J. C.; Huo, H.; Bai, Y.; Ehsani, S.; Jeon, A.; Shi, T.; Daude, N.; Lau, A.; Young, R.; Xu, L. *PLoS Pathog* **2009**, *5*, e1000608.
- [80] Lougheed, K. E.; Bennett, M. H.; Williams, H. D. *Journal of microbiological methods* **2014**, *105*, 67–71.
- [81] Gao, M.; McCluskey, P.; Loganathan, S. N.; Arkov, A. L. *JoVE (Journal of Visualized Experiments)* **2014**, e51387–e51387.
- [82] Zhu, L.; Li, M.; Wei, L.; Liu, X.; Yin, J.; Gao, Y. *Proteome science* **2014**, *12*, 6.
- [83] Kaake, R. M.; Milenkovic, T.; Przulj, N.; Kaiser, P.; Huang, L. *Journal of proteome research* **2010**, *9*, 2016–2029.

Bibliography

- [84] Fabre, B.; Lambour, T.; Delobel, J.; Amalric, F.; Monsarrat, B.; Burlet-Schiltz, O.; Bousquet-Dubouch, M.-P. *Molecular & Cellular Proteomics* **2013**, *12*, 687–699.
- [85] Toews, J.; Rogalski, J. C.; Clark, T. J.; Kast, J. *Analytica chimica acta* **2008**, *618*, 168–183.
- [86] Metz, B.; Kersten, G. F.; Baart, G. J.; de Jong, A.; Meiring, H.; ten Hove, J.; van Steenberg, M. J.; Hennink, W. E.; Crommelin, D. J.; Jiskoot, W. *Bioconjugate chemistry* **2006**, *17*, 815–822.
- [87] Metz, B.; Kersten, G. F.; Hoogerhout, P.; Brugghe, H. F.; Timmermans, H. A.; De Jong, A.; Meiring, H.; ten Hove, J.; Hennink, W. E.; Crommelin, D. J. *Journal of Biological Chemistry* **2004**, *279*, 6235–6243.
- [88] Hogg, P. J. *Trends in biochemical sciences* **2003**, *28*, 210–214.
- [89] Linke, K.; Jakob, U. *Antioxidants and Redox Signaling* **2003**, *5*, 425–434.
- [90] Fraenkel-Conrat, H.; Olcott, H. S. *Journal of the American Chemical Society* **1948**, *70*, 2673–2684.
- [91] Fraenkel-Conrat, H.; Cooper, M.; Olcott, H. S. *Journal of the American Chemical Society* **1945**, *67*, 950–954.
- [92] Fraenkel-Conrat, H.; Olcott, H. S. *Journal of the American Chemical Society* **1946**, *68*, 34–37.
- [93] Perkins, J. R.; Diboun, I.; Dessailly, B. H.; Lees, J. G.; Orengo, C. *Structure* **2010**, *18*, 1233–1243.
- [94] Jurado, L. A.; Chockalingam, P. S.; Jarrett, H. W. *Physiological reviews* **1999**, *79*, 661–682.
- [95] Newman, R. A.; Van Scyoc, W. S.; Sorensen, B. R.; Jaren, O. R.; Shea, M. A. *Proteins: Structure, Function, and Bioinformatics* **2008**, *71*, 1792–1812.

Bibliography

- [96] Terwilliger, T. C.; Eisenberg, D. *Journal of Biological Chemistry* **1982**, *257*, 6010–6015.
- [97] Maulet, Y.; Cox, J. A. *Biochemistry* **1983**, *22*, 5680–5686.
- [98] Pan, J.; Konermann, L. *Biochemistry* **2010**, *49*, 3477–3486.
- [99] Schulz, D. M.; Ihling, C.; Clore, G. M.; Sinz, A. *Biochemistry* **2004**, *43*, 4703–4715.
- [100] Chavez, J. D.; Liu, N. L.; Bruce, J. E. *Journal of proteome research* **2011**, *10*, 1528–1537.
- [101] Wong, J. W.; Maleknia, S. D.; Downard, K. M. *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 225–233.
- [102] Scaloni, A.; Miraglia, N.; Orrù, S.; Amodeo, P.; Motta, A.; Marino, G.; Pucci, P. *Journal of molecular biology* **1998**, *277*, 945–958.
- [103] Seeholzer, S. H.; Cohn, M.; Wand, A. J.; Crespi, H. L.; Putkey, J. A.; Means, A. R. *Calcium-Binding Proteins in Health and Disease* **2012**, 360.
- [104] Richards, F. M. *Proceedings of the National Academy of Sciences* **1958**, *44*, 162–166.
- [105] Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E. *Analytical chemistry* **2005**, *77*, 311–318.
- [106] Méndez, T. J.; Johnson, J. V.; Richardson, D. E. *Analytical biochemistry* **2000**, *279*, 114–118.
- [107] Gawronski, T. H.; Wold, F. *Biochemistry* **1972**, *11*, 442–448.
- [108] Kim, J.-S.; Raines, R. T. *Protein Science* **1993**, *2*, 348–356.
- [109] Chen, F.; Nielsen, S.; Zenobi, R. *Journal of Mass Spectrometry* **2013**, *48*, 807–812.
- [110] Petrotchenko, E. V.; Olkhovik, V. K.; Borchers, C. H. *Molecular & Cellular Proteomics* **2005**, *4*, 1167–1179.

Bibliography

- [111] Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. *Nat methods* **2012**, *9*, 671–675.
- [112] Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; De Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E. *Nucleic acids research* **2012**, gks400.
- [113] Schaab, C.; Geiger, T.; Stoehr, G.; Cox, J.; Mann, M. *Molecular & Cellular Proteomics* **2012**, *11*, M111–014068.
- [114] Calef, C.; Kuiken, C.; Szinger, J.; Gaschen, B.; Abfalterer, W.; Zhang, M.; Tao, N.; Funkhouser, R.; Yusim, K.; Flynn, M. *HIV Sequence Compendium 2005* **2005**, 49–79.
- [115] Wilkins, M. R.; Lindskog, I.; Gasteiger, E.; Bairoch, A.; Sanchez, J.-C.; Hochstrasser, D. F.; Appel, R. D. *Electrophoresis* **1997**, *18*, 403–408.
- [116] Wolfram Research, I. *Mathematica 10*; Wolfram Research, Inc: Champaign, Illinois, 2014.
- [117] ProteinProspector MS Product, v 5.17.1. 2016; <http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct>.
- [118] Cottrell, J. S.; London, U. *Electrophoresis* **1999**, *20*, 3551–3567.
- [119] Götze, M.; Pettelkau, J.; Fritzsche, R.; Ihling, C. H.; Schäfer, M.; Sinz, A. *Journal of The American Society for Mass Spectrometry* **2015**, *26*, 83–97.
- [120] Götze, M.; Pettelkau, J.; Schaks, S.; Bosse, K.; Ihling, C. H.; Krauth, F.; Fritzsche, R.; Kühn, U.; Sinz, A. *Journal of the American Society for Mass Spectrometry* **2012**, *23*, 76–87.
- [121] Fan, S.-B.; Meng, J.-M.; Lu, S.; Zhang, K.; Yang, H.; Chi, H.; Sun, R.-X.; Dong, M.-Q.; He, S.-M. *Current Protocols in Bioinformatics* **2015**, 8–21.
- [122] Yang, B.; Wu, Y.-J.; Zhu, M.; Fan, S.-B.; Lin, J.; Zhang, K.; Li, S.; Chi, H.; Li, Y.-X.; Chen, H.-F. *Nature methods* **2012**, *9*, 904–906.

Bibliography

- [123] Babu, Y. S.; Bugg, C. E.; Cook, W. J. *Journal of molecular biology* **1988**, *204*, 191–204.
- [124] Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. *Nature structural biology* **1995**, *2*, 768–776.
- [125] Fallon, J. L.; Quijoco, F. A. *Structure* **2003**, *11*, 1303–1307.
- [126] Terwilliger, T. C.; Eisenberg, D. *Journal of Biological Chemistry* **1982**, *257*, 6016–6022.
- [127] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic acids research* **2000**, *28*, 235–242.
- [128] Schrödinger, LLC,
- [129] Merkley, E. D.; Rysavy, S.; Kahraman, A.; Hafen, R. P.; Daggett, V.; Adkins, J. N. *Protein Science* **2014**, *23*, 747–759.
- [130] *UBC Water Quality Results*; University of British Columbia, 2016.
- [131] Phillips, D.; Julie Foster, T. F. S. A. N., Ph.D. Trace Metal Analysis of Water Produced From a Thermo Scientific Barnstead GenPure UV/UF-TOC Water Purification System. 2014.
- [132] Tomasselli, A. G.; Howe, W. J.; Hui, J. O.; Sawyer, T. K.; Reardon, I. M.; DeCamp, D. L.; Craik, C. S.; Heinrikson, R. L. *Proteins: Structure, Function, and Bioinformatics* **1991**, *10*, 1–9.
- [133] Stigler, J.; Rief, M. *Proceedings of the National Academy of Sciences* **2012**, *109*, 17814–17819.
- [134] Kuchumov, A. R.; Loo, J. A.; Vinogradov, S. N. *Journal of protein chemistry* **2000**, *19*, 139–149.
- [135] Harris, D. C. *Quantitative Chemical Analysis, 7th edition*; W.H. Freeman and Company, 2007.

Bibliography

- [136] Sperry, J. B.; Huang, R. Y.; Zhu, M. M.; Rempel, D. L.; Gross, M. L. *International journal of mass spectrometry* **2011**, *302*, 85–92.
- [137] Schägger, H. *Nature Protocols* **2006**, *1*, 16–22.
- [138] Grabski, A. C.; Novagen, R. R. B. **2010**,
- [139] Biovision, Calmodulin, human recombinant. 2015; <http://www.biovision.com/manuals/7838.pdf>.
- [140] Rauh, J. J.; Nelson, D. L. *The Journal of cell biology* **1981**, *91*, 860–865.
- [141] Tatsumi, R.; Wuollet, A. L.; Tabata, K.; Nishimura, S.; Tabata, S.; Mizunoya, W.; Ikeuchi, Y.; Allen, R. E. *American Journal of Physiology-Cell Physiology* **2009**, *296*, C922–C929.
- [142] Bello, J.; Bello, H. R.; Granados, E. *Biochemistry* **1982**, *21*, 461–465.
- [143] Sorensen, B. R.; Shea, M. A. *Biophysical journal* **1996**, *71*, 3407–3420.
- [144] Bahloul, A.; Chevreux, G.; Wells, A. L.; Martin, D.; Nolt, J.; Yang, Z.; Chen, L.-Q.; Potier, N.; Van Dorsselaer, A.; Rosenfeld, S. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 4787–4792.
- [145] Klee, C.; Crouch, T.; Krinks, M. *Proceedings of the National Academy of Sciences* **1979**, *76*, 6270–6273.
- [146] Dunker AK, K. A. *Biochemical Journal*. **1976**;, *153*, 191–197.
- [147] Shi, Y.; Mowery, R. A.; Ashley, J.; Hentz, M.; Ramirez, A. J.; Bilgicer, B.; Slunt-Brown, H.; Borchelt, D. R.; Shaw, B. F. *Protein Science* **2012**, *21*, 1197–1209.
- [148] Shirai, A.; Matsuyama, A.; Yashiroda, Y.; Hashimoto, A.; Kawamura, Y.; Arai, R.; Komatsu, Y.; Horinouchi, S.; Yoshida, M. *Journal of Biological Chemistry* **2008**, *283*, 10745–10752.

Bibliography

- [149] Rath, A.; Glibowicka, M.; Nadeau, V. G.; Chen, G.; Deber, C. M. *Proceedings of the National Academy of Sciences* **2009**, *106*, 1760–1765.
- [150] Linde, S.; Nielsen, J.; Hansen, B.; Welinder, B. *Journal of Chromatography B: Biomedical Sciences and Applications* **1990**, *530*, 29–37.
- [151] Ren, D.; Pipes, G. D.; Liu, D.; Shih, L.-Y.; Nichols, A. C.; Treuheit, M. J.; Brems, D. N.; Bondarenko, P. V. *Analytical biochemistry* **2009**, *392*, 12–21.
- [152] Snijders, A. P.; Hung, M.-L.; Wilson, S. A.; Dickman, M. J. *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 88–96.
- [153] Berg, J. L. T., Jeremy Mark; Stryer., L. *Biochemistry 7th ed*; Basingstoke : W.H. Freeman, 2012.
- [154] Remington, J. P.; Troy, D. B.; Beringer, P. *Remington: The science and practice of pharmacy*; Lippincott Williams & Wilkins, 2006; Vol. 1.
- [155] Poncz, L.; Dearborn, D. *Journal of Biological Chemistry* **1983**, *258*, 1844–1850.
- [156] Howard, G. C.; Bethell, D. R. *Basic methods in antibody production and characterization*; CRC Press, 2000.
- [157] Loo, R. R. O.; Goodlett, D. R.; Smith, R. D.; Loo, J. A. *Journal of the American Chemical Society* **1993**, *115*, 4391–4392.
- [158] Yin, S.; Xie, Y.; Loo, J. A. *Journal of the American Society for Mass Spectrometry* **2008**, *19*, 1199–1208.
- [159] Watkins, R. W.; Arnold, U.; Raines, R. T. *Chemical Communications* **2011**, *47*, 973–975.
- [160] Allende, J. E.; Richards, F. M. *Biochemistry* **1962**, *1*, 295–304.
- [161] Krauth, F.; Ihling, C. H.; Ruettinger, H. H.; Sinz, A. *Rapid Communications in Mass Spectrometry* **2009**, *23*, 2811–2818.

Bibliography

- [162] Dreiling, A.; Hanneken, M.; König, S. *Rapid Communications in Mass Spectrometry* **2014**, *28*, 2385–2388.
- [163] Guan, Z.; Yates, N. A.; Bakhtiar, R. *Journal of the American Society for Mass Spectrometry* **2003**, *14*, 605–613.
- [164] Korfmacher, W. A. *Mass spectrometry for drug discovery and drug development*; John Wiley & Sons, 2012; Vol. 46.
- [165] Jaiswal, M.; Crabtree, N. M.; Bauer, M. A.; Hall, R.; Raney, K. D.; Zybailov, B. L. *BMC bioinformatics* **2014**, *15*, 1.
- [166] Trnka, M. J.; Baker, P. R.; Robinson, P. J.; Burlingame, A.; Chalkley, R. J. *Molecular & Cellular Proteomics* **2014**, *13*, 420–434.
- [167] Wong, J. W.; Maleknia, S. D.; Downard, K. M. *Analytical chemistry* **2003**, *75*, 1557–1563.
- [168] Patthy, L.; Smith, E. L. *Journal of Biological Chemistry* **1975**, *250*, 565–569.
- [169] McClendon, S.; Rospigliosi, C. C.; Eliezer, D. *Protein Science* **2009**, *18*, 1531–1540.
- [170] Shirota, M.; Ishida, T.; Kinoshita, K. *Protein Science* **2008**, *17*, 1596–1602.
- [171] Lafitte, D.; Heck, A. J.; Hill, T. J.; Jumel, K.; Harding, S. E.; Derrick, P. J. *European Journal of Biochemistry* **1999**, *261*, 337–344.
- [172] Leavell, M. D.; Novak, P.; Behrens, C. R.; Schoeniger, J. S.; Kruppa, G. H. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 1604–1611.
- [173] Terwilliger, T. C.; Weissman, L.; Eisenberg, D. *Biophysical journal* **1982**, *37*, 353.
- [174] Elias, J. E.; Haas, W.; Faherty, B. K.; Gygi, S. P. *Nature methods* **2005**, *2*, 667–675.

Bibliography

- [175] Bruker Impact II Compass Data Analysis 4.2. Bruker Daltonik GmbH: Billerica, MA · USA, 2014.
- [176] Mayampurath, A. M.; Jaitly, N.; Purvine, S. O.; Monroe, M. E.; Auberry, K. J.; Adkins, J. N.; Smith, R. D. *Bioinformatics* **2008**, *24*, 1021–1023.
- [177] Chu, F.; Baker, P. R.; Burlingame, A. L.; Chalkley, R. J. *Molecular & Cellular Proteomics* **2010**, *9*, 25–31.
- [178] Senko, M. W.; Beu, S. C.; McLafferty, F. W. *Journal of the American Society for Mass Spectrometry* **1995**, *6*, 229–233.
- [179] Hogg, P. J. *Trends in biochemical sciences* **2003**, *28*, 210–214.
- [180] Linke, K.; Jakob, U. *Antioxidants and Redox Signaling* **2003**, *5*, 425–434.
- [181] Liu, F.; Wu, C.; Sweedler, J. V.; Goshe, M. B. *Proteomics* **2012**, *12*, 401–405.
- [182] Mädler, S.; Seitz, M.; Robinson, J.; Zenobi, R. *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 1775–1783.
- [183] Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R. *Nature methods* **2008**, *5*, 315–318.
- [184] Chen, Z. A.; Jawhari, A.; Fischer, L.; Buchen, C.; Tahir, S.; Kamenski, T.; Rasmussen, M.; Lariviere, L.; Bukowski-Wills, J.-C.; Nilges, M. *The EMBO journal* **2010**, *29*, 717–726.
- [185] Lauber, M. A.; Reilly, J. P. *Journal of proteome research* **2011**, *10*, 3604–3616.
- [186] Fritzsche, R.; Ihling, C. H.; Götze, M.; Sinz, A. *Rapid Communications in Mass Spectrometry* **2012**, *26*, 653–658.
- [187] Leitner, A.; Reischl, R.; Walzthoeni, T.; Herzog, F.; Bohn, S.; Förster, F.; Aebersold, R. *Molecular & Cellular Proteomics* **2012**, *11*, M111–014126.

- [188] Lee, Y. J. *Molecular BioSystems* **2008**, *4*, 816–823.
- [189] Leitner, A.; Reischl, R.; Walzthoeni, T.; Herzog, F.; Bohn, S.; Förster, F.; Aebersold, R. *Molecular & Cellular Proteomics* **2012**, *11*, M111–014126.
- [190] van Montfort, B. A.; Doeven, M. K.; Canas, B.; Veenhoff, L. M.; Poolman, B.; Robillard, G. T. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **2002**, *1555*, 111–115.
- [191] Bruce, J. E. *Proteomics* **2012**, *12*, 1565–1575.
- [192] MilliporeSigma,, Calmodulin, bovine brain 208690-1MG, D00129241. 2017.
- [193] Christensen, T.; Gooden, D. M.; Kung, J. E.; Toone, E. J. *Journal of the American Chemical Society* **2003**, *125*, 7357–7366.
- [194] Nemirovskiy, O.; Giblin, D. E.; Gross, M. L. *Journal of the American Society for Mass Spectrometry* **1999**, *10*, 711–718.

Appendix A

First Appendix

A.1 Confirming the Calcium-Free Calmodulin-Melittin System

A.1.1 Concentration of EDTA in Calmodulin-Melittin Cross-linking Reaction Mixture

The purchased calmodulin sample (molecular weight = 16799.78 g/mol) was lyophilized from a 400 μ L buffer containing 2 mM EDTA. No calcium was added or present in the calmodulin sample prior to or post lyophilization [192]. In the 1 mg of lyophilized calmodulin, the amount of EDTA (molecular weight 292.24 g/mol) was calculated to be 0.23 mg, as shown below:

$$\frac{2\text{mmol}}{1\text{L}} \times \frac{0.0004\text{L}}{1} \times \frac{292.24\text{mg}}{1\text{mmol}} = 0.234\text{mg}$$

A 1 mM stock solution of calmodulin was prepared by adding 59.5 μ L of deionized water (calculation shown below) to 1 mg of calmodulin.

$$\frac{1\text{L}}{1\text{mmol}} \times \frac{1\text{mmol}}{16799.78\text{mg}} \times \frac{1\text{mg}}{1} = 59.5\mu\text{L}$$

The concentration of EDTA ($[EDTA]_{Stock}$) in the 1 mM calmodulin stock solution was 13 mM (calculation shown below).

$$[EDTA]_{Stock} = \frac{2\text{mmol}}{1\text{L}} \times \frac{400\mu\text{L}}{1} \times \frac{1}{59.5\mu\text{L}} = 13\text{mM}$$

Cross-linking was performed by the addition of 6 μ L of the 1mM calmodulin stock solution to each cross-linking buffer. The total volume of each cross-linking reaction solution (containing the protein, buffer and cross-linker reagent solutions) was 100 μ L. Therefore, the final concentration of calmodulin ($[Calmodulin]_{Rxn}$)

A.1. Confirming the Calcium-Free Calmodulin-Melittin System

and EDTA ($[EDTA]_{Rxn}$) in the cross-linking reaction mixture was $60\mu\text{M}$ and $807\mu\text{M}$, respectively, as shown below:

$$[\text{Calmodulin}]_{Rxn} = \frac{1\text{mmol}}{1\text{L}} \times \frac{6\mu\text{L}}{1} \times \frac{1}{100\mu\text{L}} = 60\mu\text{M}$$

$$[EDTA]_{Rxn} = \frac{13\text{mmol}}{1\text{L}} \times \frac{6\mu\text{L}}{1} \times \frac{1}{100\mu\text{L}} = 807\mu\text{M}$$

A.1.2 Conditional Formation Constant for EDTA-Calcium binding

EDTA is a hexaprotic acid that forms four bonds with one Ca^{2+} ion, producing a very stable complex [193]. The equilibrium expression for the binding of Ca^{2+} to EDTA and the definition of its association constant (K_{EDTA}) are shown below [134]:



However the amount of EDTA in its EDTA^{4-} form is pH-dependent and therefore, the association constant must be adjusted for the pH of the reaction buffer such that:

$$K'_{EDTA} = \alpha_{[\text{EDTA}^{4-}]} K_{EDTA},$$

where $\alpha_{[\text{EDTA}^{4-}]}$ is the fraction of EDTA in its EDTA^{4-} form and K'_{EDTA} is the pH-corrected association constant. In this experiment two buffers were used: HEPES at a pH of 7.4 for PFA and NHS-ester cross-linking (referred to as “other” cross-linking), and MES at a pH of 6.5 for EDC cross-linking. The corresponding $\alpha_{[\text{EDTA}^{4-}]}$ values in solutions at a pH of 7.4 and 6.5 are 0.0026 and 0.00017, respectively [135]. Therefore, the K'_{EDTA} for each buffer condition was determined as shown below:

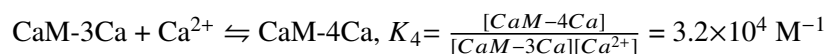
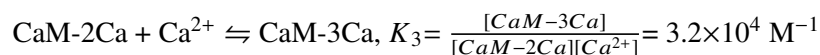
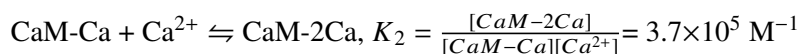
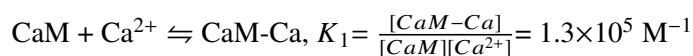
$$K'_{EDTA_{pH=7.4}} = (0.0020)(4.5 \times 10^{10} \text{ M}^{-1}) = 9.0 \times 10^7 \text{ M}^{-1}$$

$$K'_{EDTA_{pH=6.5}} = (0.00013)(4.5 \times 10^{10} \text{ M}^{-1}) = 5.9 \times 10^6 \text{ M}^{-1}$$

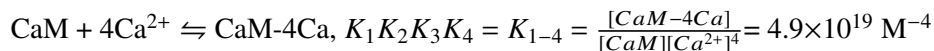
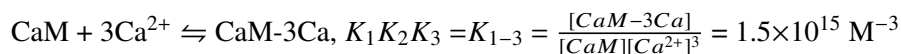
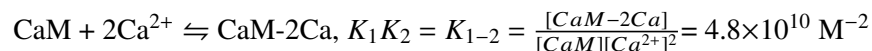
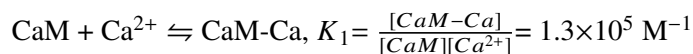
A.1.3 Concentration of Calcium-Free Versus Calcium-Loaded Calmodulin-Melittin

The ratio of Ca²⁺-loaded calmodulin to Ca²⁺-free calmodulin in the presence of EDTA in this study was calculated. EDTA and calmodulin were present in sample at a final concentration of 8.1×10^{-4} M and 6.0×10^{-5} M in this experiment, respectively. The equilibria of Ca²⁺ binding to EDTA and to calmodulin was simultaneously considered.

Ca²⁺-free calmodulin (CaM) cooperatively binds up to four Ca²⁺ ions, as depicted by the equilibrium expressions below. Ca²⁺-loaded calmodulin species with 1, 2, 3 and 4 Ca²⁺ ions are denoted as CaM-Ca, CaM-2Ca, CaM-3Ca, and CaM-4Ca, respectively. The association constants for the sequential binding each of Ca²⁺ ion to calmodulin are denoted as K₁, K₂, K₃ and K₄ and are defined below [194]:



Each equilibrium expression can be combined with the preceding expression to define each association constant in terms of the equilibrium Ca²⁺-free calmodulin ([CaM]) and Ca²⁺ ([Ca²⁺]) concentrations, as shown below:



Using the expressions for the association constants defined above, and an expression for the concentration of each form of Ca²⁺-loaded calmodulin was derived:

A.1. *Confirming the Calcium-Free Calmodulin-Melittin System*

$$[CaM - Ca] = K_1[CaM][Ca^{2+}]$$

$$[CaM - 2Ca] = K_{1-2}[CaM][Ca^{2+}]^2$$

$$[CaM - 3Ca] = K_{1-3}[CaM][Ca^{2+}]^3$$

$$[CaM - 4Ca] = K_{1-4}[CaM][Ca^{2+}]^4$$

To consider the increased Ca^{2+} binding affinity of the calmodulin-melittin complex, the association constants K_1 , K_2 , K_3 and K_4 were multiplied by 300 and the association constants accounting for the presence of melittin are defined below:

$$K_{M1} = (300)K_1 = 3.9 \times 10^7 M^{-1}$$

$$K_{M1-2} = (300)^2 K_{1-2} = 4.3 \times 10^{15} M^{-1}$$

$$K_{M1-3} = (300)^3 K_{1-2} = 4.2 \times 10^{22} M^{-1}$$

$$K_{M1-4} = (300)^4 K_{1-2} = 4.0 \times 10^{29} M^{-1}$$

A mass balance expression was derived assuming that the total number of moles of calmodulin is conserved. Therefore, the initial concentration of calmodulin ($[CaM]_0$) was equal to the equilibrium concentrations of Ca^{2+} -free calmodulin and each form of Ca^{2+} -loaded calmodulin ($[CaM - nCa]$, $n = 1, 2, 3$ or 4) as shown below:

$$[CaM]_0 = [CaM] + [CaM - Ca] + [CaM - 2Ca] + [CaM - 3Ca] + [CaM - 4Ca] = 6.0 \times 10^{-5} M$$

$$[CaM]_0 = [CaM](1 + K_{M1}[Ca^{2+}] + K_{M1-2}[Ca^{2+}]^2 + K_{M1-3}[Ca^{2+}]^3 + K_{M1-4}[Ca^{2+}]^4)$$

$$[CaM] = \frac{[CaM]_0}{(1 + K_{M1}[Ca^{2+}] + K_{M1-2}[Ca^{2+}]^2 + K_{M1-3}[Ca^{2+}]^3 + K_{M1-4}[Ca^{2+}]^4)}$$

A mass balance expression was derived assuming that the initial concentration of EDTA ($[EDTA^{4-}]_0$) is equal to the sum of the equilibrium concentrations of EDTA in its Ca^{2+} -free ($[EDTA^{4-}]$) and Ca^{2+} -loaded ($[EDTA - Ca^{2+}]^{2-}$) forms:

A.1. Confirming the Calcium-Free Calmodulin-Melittin System

$$[EDTA^{4-}]_0 = [EDTA^{4-}] + \{[EDTA - Ca^{2+}]^{2-}\} = 8.1 \times 10^{-4} M$$

$$[EDTA^{4-}] = [EDTA^{4-}]_0 - \{[EDTA - Ca^{2+}]^{2-}\}$$

By combining the mass balance and equilibrium constant expressions for EDTA, $\{[EDTA - Ca^{2+}]^{2-}\}$ can be defined in terms of the equilibrium concentration of Ca^{2+} , as shown below:

$$K'_{EDTA} = \frac{\{[EDTA - Ca^{2+}]^{2-}\}}{([EDTA^{4-}]_0 - \{[EDTA - Ca^{2+}]^{2-}\})[Ca^{2+}]}$$

$$\{[EDTA - Ca^{2+}]^{2-}\} = \frac{K'_{EDTA}[EDTA^{4-}]_0[Ca^{2+}]}{1 + K'_{EDTA}[Ca^{2+}]}$$

In this experiment, deionized, ultrapure water was used and therefore the initial concentration of Ca^{2+} was set to 2.5×10^{-11} M [131]. A mass balance equation for the total Ca^{2+} concentration, which is equal to the initial Ca^{2+} concentration ($[Ca^{2+}]_0$), is equal to the sum of the equilibrium concentrations of all Ca^{2+} containing species:

$$[Ca^{2+}]_0 = [Ca^{2+}] + [CaM - Ca] + 2[CaM - 2Ca] + 3[CaM - 3Ca] + 4[CaM - 4Ca] + \{[EDTA - Ca^{2+}]^{2-}\} = 1.13 \times 10^{-4} M$$

$$[Ca^{2+}]_0 = [Ca^{2+}] + [CaM](1 + K_{M1}[Ca^{2+}] + 2K_{M1-2}[Ca^{2+}]^2 + 3K_{M1-3}[Ca^{2+}]^3 + 4K_{M1-4}[Ca^{2+}]^4) + \frac{K'_{EDTA}[EDTA^{4-}]_0[Ca^{2+}]}{1 + K'_{EDTA}[Ca^{2+}]}$$

$$[Ca^{2+}]_0 = [Ca^{2+}] + \frac{[CaM]_0(1 + K_{M1}[Ca^{2+}] + 2K_{M1-2}[Ca^{2+}]^2 + 3K_{M1-3}[Ca^{2+}]^3 + 4K_{M1-4}[Ca^{2+}]^4) + \frac{K'_{EDTA}[EDTA^{4-}]_0[Ca^{2+}]}{1 + K'_{EDTA}[Ca^{2+}]}}{(1 + K_{M1}[Ca^{2+}] + K_{M1-2}[Ca^{2+}]^2 + K_{M1-3}[Ca^{2+}]^3 + K_{M1-4}[Ca^{2+}]^4)}$$

Solving for the equilibrium concentration of free Ca^{2+} gives the following:

$$[Ca^{2+}]_{pH=7.4} = 1.81 \times 10^{-9} M$$

$$[Ca^{2+}]_{pH=6.5} = 2.73 \times 10^{-8} M$$

Inserting the equilibrium concentration of free Ca^{2+} into the expression for Ca^{2+} -free calmodulin yields the following:

A.1. Confirming the Calcium-Free Calmodulin-Melittin System

$$[CaM] = \frac{[CaM]_0}{(1+K_1[Ca^{2+}]+K_{1-2}[Ca^{2+}]^2+K_{1-3}[Ca^{2+}]^3+K_{1-4}[Ca^{2+}]^4)}$$

$$[CaM]_{pH=7.4} = 6.00 \times 10^{-5} M$$

$$[CaM]_{pH=6.5} = 6.00 \times 10^{-5} M$$

The total Ca^{2+} -loaded calmodulin concentrations were calculated as shown:

$$\sum_{n=1}^{n=4} [CaM - nCa] = [CaM - Ca] + [CaM - 2Ca] + [CaM - 3Ca] + [CaM - 4Ca] = [CaM]_0 - [CaM]$$

$$\sum_{n=1}^{n=4} [CaM - nCa]_{pH=7.4} = 7.9 \times 10^{-13} M$$

$$\sum_{n=1}^{n=4} [CaM - nCa]_{pH=6.5} = 8.2 \times 10^{-12} M$$

Therefore, the ratios of Ca^{2+} -loaded calmodulin to Ca^{2+} -free calmodulin in the presence of EDTA were calculated to be:

$$\frac{\sum_{n=1}^{n=4} [CaM - nCa]_{pH=7.4}}{[CaM]_{pH=7.4}} = \frac{7.9 \times 10^{-13} M}{6.0 \times 10^{-5} M} = 1.3 \times 10^{-8}$$

$$\frac{\sum_{n=1}^{n=4} [CaM - nCa]_{pH=6.5}}{[CaM]_{pH=6.5}} = \frac{8.2 \times 10^{-12} M}{6.0 \times 10^{-5} M} = 1.4 \times 10^{-7}$$

The initial Ca^{2+} concentrations in each sample required to convert 1% of the calmodulin-melittin complexes into their Ca^{2+} -loaded state in the presence of EDTA were determined. The equilibrium concentration of Ca^{2+} in this case was calculated as shown below:

$$\sum_{n=1}^{n=4} [CaM - nCa] = [CaM]_0 - [CaM] = 0.01[CaM]_0$$

$$[CaM] = 0.99[CaM]_0$$

$$[CaM] = 0.99[CaM]_0 = \frac{[CaM]_0}{(1+K_{M1}[Ca^{2+}]+K_{M1-2}[Ca^{2+}]^2+K_{M1-3}[Ca^{2+}]^3+K_{M1-4}[Ca^{2+}]^4)}$$

$$[Ca^{2+}] = 2.6 \times 10^{-10} M$$

Using the equilibrium concentration of Ca^{2+} above, the equilibrium concentrations of Ca^{2+} - loaded EDTA were calculated:

A.1. *Confirming the Calcium-Free Calmodulin-Melittin System*

$$[{\text{EDTA}} - \text{Ca}^{2+}]^{2-} = \frac{K'_{\text{EDTA}}[\text{EDTA}^{4-}]_0[\text{Ca}^{2+}]}{1 + K'_{\text{EDTA}}[\text{Ca}^{2+}]}$$

$$[{\text{EDTA}} - \text{Ca}^{2+}]^{2-}_{\text{pH}=7.4} = 1.8 \times 10^{-5} \text{ M}$$

$$[{\text{EDTA}} - \text{Ca}^{2+}]^{2-}_{\text{pH}=6.5} = 1.2 \times 10^{-5} \text{ M}$$

Using the mass balance expression derived above, the initial Ca^{2+} concentrations were calculated for each reaction buffer.

$$[\text{Ca}^{2+}]_0 = [\text{Ca}^{2+}] + [\text{CaM} - \text{Ca}] + 2[\text{CaM} - 2\text{Ca}] + 3[\text{CaM} - 3\text{Ca}] + 4[\text{CaM} - 4\text{Ca}] + [{\text{EDTA}} - \text{Ca}^{2+}]^{2-}$$

$$[\text{Ca}^{2+}]_0 = [\text{Ca}^{2+}] + [\text{CaM}](1 + K_{M1}[\text{Ca}^{2+}] + 2K_{M1-2}[\text{Ca}^{2+}]^2 + 3K_{M1-3}[\text{Ca}^{2+}]^3 + 4K_{M1-4}[\text{Ca}^{2+}]^4) + [{\text{EDTA}} - \text{Ca}^{2+}]^{2-}$$

$$[\text{Ca}^{2+}]_{0,\text{pH}=7.4} = 7.8 \times 10^{-5} \text{ M}$$

$$[\text{Ca}^{2+}]_{0,\text{pH}=6.5} = 6.2 \times 10^{-5} \text{ M}$$

A.2 Data Analysis Codes

Microsoft Excel Code

Eliminating Control Values in Experimental List
i = number of experimental value
j = number of control value

```
Sub CalculateResultValues()  
  
For i = 2 To (Add total number of experimental values)  
  Cells(i, 3) = Cells(i, 2)  
  For j = 2 To (Add total number of control values)  
  
    If Abs(Cells(i, 2) - Cells(j, 1)) < 0.2 Then  
      Cells(i, 4) = ""  
      Exit For  
    Else  
      Cells(i, 4) = Cells(i, 2)  
    End If  
  Next j  
Next i  
  
End Sub
```

Figure A.1: Excel Code for Elimination

Mathematica Example Codes: Possible Cross-linked Species

```
SetDirectory["/home/kast/savi848/"]
/home/kast/savi848
list = Import["PFA peptides.txt", "List"];
Length[list]
439
sumlist = Table[{i, j, list[[i]] + list[[j]]}, {i, Length[list]}, {j, Length[list]}];
Dimensions[sumlist]
{439, 439, 3}
CrossLinks = Flatten[sumlist, {1, 2}];
CrossLinks >> "Possible Cross Linked Species CaM PFA";
CrossLinks = << "Possible Cross Linked Species CaM PFA";
Dimensions[CrossLinks]
{192721, 3}
Export["Possible Cross Linked Species CaM PFA.txt", CrossLinks];
```

Figure A.2: Mathematica Code for Possible cross-linked Species

A.2. Data Analysis Codes

Mathematica Example Codes: Candidate Cross-Linked Species

```
SetDirectory["/home/kant/savi848/"]
/home/kant/savi848

CrossLinks = << "Possible Cross Linked Species CaM PFA";
Dimensions[CrossLinks]
{192721, 3}

CaMPPFA1 = Import["CaM PFA1-CPI.txt", "List"];
CaMPPFA1[[2]]
92.5992

CaMPPFA1[[2]]
92.5992

Max[CrossLinks[[All, 3]]]
13578.2

Min[CrossLinks[[All, 3]]]
292.211

CaMPPFA1NewCrossLinks = Cases[CrossLinks, {_, _, x_} /; 390 ≤ x ≤ 5900];
Dimensions[CaMPPFA1NewCrossLinks]
{89034, 3}

CaMPPFA1NewCrossLinks[[1 ;; 10]]
{{1, 9, 420.233}, {1, 10, 423.252}, {1, 11, 432.233}, {1, 12, 435.252}, {1, 13, 444.233},
CaMPPFA1NewCrossLinks[[1 ;; 10]]
{{1, 9, 420.233}, {1, 10, 423.252}, {1, 11, 432.233}, {1, 12, 435.252}, {1, 13, 444.233},
CaMPPFA1CandidateList = Table[i, {i, 10000}];
i = 1;
Date[]
Do[Do[
  If[Abs[CaMPPFA1[[j]] - CaMPPFA1NewCrossLinks[[k, 3]]] ≤ 0.2, (CaMPPFA1CandidateList[[i] = CaMPPFA1NewCrossLinks[[k]]];
  i = i + 1)], {k, Length[CaMPPFA1NewCrossLinks]}], {j, Length[CaMPPFA1]}]
Date[]
{2015, 2, 25, 0, 19, 27.545972}

CaMPPFA1CandidateList >> "CaMPPFA1 Candidate Cross Linked Peptides";
CaMPPFA1CandidateList = << "CaMPPFA1 Candidate Cross Linked Peptides";
Dimensions[CaMPPFA1CandidateList]
{10000}
{10000}
{10000}

CaMPPFA1CandidateList[[1 ;; 10]]
{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}

CaMPPFA1CandidateList[[1 ;; 10]]
{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}

CaMPPFA1CandidateList[[1, 3]]
CaMPPFA1CandidateList[[1, 3]]

Cases[CaMPPFA1CandidateList, {_, _, _}] // Dimensions
{0}
{62784, 3}
{62784, 3}

CaMPPFA1NewCandidateList = Cases[CaMPPFA1CandidateList, {_, _, _}];
Round[Cases[CaMPPFA1CandidateList, { . . . }], 0.01][[1 ;; 10]]
Part::take: Cannot take positions 1 through 10 in { }.>
{}[[1 ;; 10]]

Export["CaMPPFA1 Candidate Cross Linked Species.txt", CaMPPFA1NewCandidateList, "Text"]
CaMPPFA1 Candidate Cross Linked Species.txt

"CaMPPFA1 Candidate Cross Linked Species.txt"
CaMPPFA1 Candidate Cross Linked Species.txt

CaMPPFA1ShortCandidateList = Union[Round[Cases[CaMPPFA1CandidateList, {_, _, _}][[All
Export["CaMPPFA1 Short Candidate Cross Linked Species.txt", CaMPPFA1ShortCandidateList,
CaMPPFA1 Short Candidate Cross Linked Species.txt
```

Figure A.3: Mathematica Code for Candidate Cross-linked Species

A.3 Bruker Impact II Tandem Mass Spectrometric Analysis Method Details

| m/z | width | Collision Energy (eV) | Charge |
|------|-------|-----------------------|--------|
| 300 | 2 | 23 | 1 |
| 300 | 2 | 23 | 2 |
| 300 | 2 | 23 | 3 |
| 400 | 2 | 23 | 1 |
| 400 | 2 | 23 | 2 |
| 400 | 2 | 23 | 3 |
| 500 | 2 | 27 | 1 |
| 500 | 2 | 27 | 2 |
| 500 | 2 | 25 | 3 |
| 600 | 2 | 33 | 1 |
| 600 | 2 | 33 | 2 |
| 600 | 2 | 27 | 3 |
| 700 | 2 | 33 | 1 |
| 700 | 2 | 33 | 2 |
| 700 | 2 | 27 | 3 |
| 800 | 3 | 45 | 1 |
| 800 | 3 | 45 | 2 |
| 800 | 3 | 40 | 3 |
| 900 | 3 | 50 | 1 |
| 900 | 3 | 50 | 2 |
| 900 | 3 | 45 | 3 |
| 1000 | 3 | 55 | 1 |
| 1000 | 3 | 55 | 2 |
| 1000 | 3 | 50 | 3 |
| 1100 | 3 | 65 | 1 |
| 1100 | 3 | 65 | 2 |
| 1100 | 3 | 55 | 3 |
| 1300 | 3 | 65 | 1 |
| 1300 | 3 | 65 | 2 |
| 1300 | 3 | 55 | 3 |

Figure A.4: Collision Energy Table for Bruker Impact II LC-MS/MS

A.4. Calmodulin-Melittin Cross-linked Candidates From First-Stage Mass Spectrometry

A.4 Calmodulin-Melittin Cross-linked Candidates From First-Stage Mass Spectrometry

Table A.1: MS Candidate Cross-linked Species for EDC

| Peptides | | | Insufficient MSMS | | | Undetermined | | | Incorrect Assignments | | | | | |
|----------|---------|---|-------------------|---------|---|--------------|---------|---|-----------------------|---------|---|---------------|------------|----------|
| [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | Actual [M]exp | Actual m/z | Actual z |
| 1696.75 | 849.37 | 2 | 1397.61 | 699.81 | 2 | 1028.60 | 515.30 | 2 | 1145.56 | 573.78 | 2 | 1144.58 | 573.29 | |
| 1807.79 | 603.60 | 3 | 1494.73 | 748.36 | 2 | 1138.56 | 570.28 | 2 | 1376.68 | 689.34 | 2 | 1375.71 | 688.86 | |
| 1902.89 | 952.44 | 2 | 1558.76 | 520.59 | 3 | 1145.56 | 573.78 | 2 | 1397.61 | 699.81 | 2 | 1396.82 | 699.41 | |
| 1962.94 | 655.31 | 3 | 1749.68 | 584.23 | 3 | 1342.76 | 448.59 | 3 | 1466.59 | 734.29 | 2 | 1465.59 | 733.79 | |
| 1994.89 | 998.45 | 2 | 1807.81 | 452.95 | 4 | 1376.68 | 689.34 | 2 | 1476.67 | 739.34 | 2 | 1474.79 | 738.40 | |
| 2067.03 | 690.01 | 3 | 1840.88 | 921.44 | 2 | 1466.59 | 734.29 | 2 | 1558.76 | 520.59 | 3 | 1558.95 | 520.65 | |
| 2136.05 | 713.02 | 3 | 1896.83 | 475.21 | 4 | 1476.67 | 739.34 | 2 | 1749.68 | 584.23 | 3 | 1748.68 | 583.89 | |
| 2150.07 | 717.69 | 3 | 1926.86 | 964.43 | 2 | 1668.83 | 557.28 | 3 | 1791.74 | 896.87 | 2 | 1790.74 | 896.37 | |
| 2304.03 | 1153.01 | 2 | 1957.86 | 979.93 | 2 | 1791.74 | 896.87 | 2 | 1926.86 | 964.43 | 2 | 1925.82 | 963.91 | |
| 2430.18 | 487.04 | 5 | 1959.87 | 654.29 | 3 | 1807.87 | 904.93 | 2 | 1994.00 | 499.50 | 4 | 1993.00 | 499.25 | |
| 2430.20 | 608.55 | 4 | 1980.95 | 496.24 | 4 | 1930.01 | 483.50 | 4 | 2081.02 | 694.67 | 3 | 2082.02 | 695.01 | |
| 2454.19 | 819.06 | 3 | 2014.01 | 672.34 | 3 | 1951.99 | 651.66 | 3 | 2136.05 | 713.02 | 3 | 2137.02 | 713.34 | |
| 2510.07 | 837.69 | 3 | 2144.06 | 715.69 | 3 | 1963.06 | 655.35 | 3 | 2142.04 | 715.01 | 3 | 2142.97 | 715.32 | |
| 2513.05 | 629.26 | 4 | 2202.08 | 735.03 | 3 | 1980.96 | 661.32 | 3 | 2184.08 | 729.03 | 3 | 2185.15 | 729.38 | |
| 2533.26 | 634.31 | 4 | 2211.09 | 369.52 | 6 | 2067.03 | 690.01 | 3 | 2202.08 | 735.03 | 3 | 2201.18 | 734.73 | |
| 2542.01 | 848.34 | 3 | 2255.10 | 752.70 | 3 | 2081.02 | 694.67 | 3 | 2211.12 | 738.04 | 3 | 2211.11 | 738.04 | |
| 2550.18 | 638.55 | 4 | 2326.04 | 1164.02 | 2 | 2151.14 | 538.78 | 4 | 2220.07 | 741.02 | 3 | 2218.21 | 740.40 | |
| 2554.30 | 852.43 | 3 | 2457.03 | 1229.52 | 2 | 2184.08 | 729.03 | 3 | 2225.14 | 742.71 | 3 | 2224.13 | 742.38 | |
| 2568.17 | 857.06 | 3 | 2486.07 | 1244.03 | 2 | 2202.08 | 735.03 | 3 | 2241.97 | 1121.99 | 2 | 2239.96 | 1120.98 | |
| 2573.00 | 515.60 | 5 | 2515.26 | 629.81 | 4 | 2211.12 | 738.04 | 3 | 2272.08 | 758.36 | 3 | 2271.06 | 758.02 | |
| 2573.08 | 1287.54 | 2 | 2542.02 | 636.51 | 4 | 2220.07 | 741.02 | 3 | 2286.13 | 763.04 | 3 | 2285.14 | 762.71 | |
| 2622.19 | 525.44 | 5 | 2555.34 | 512.07 | 5 | 2225.14 | 742.71 | 3 | 2430.18 | 487.04 | 5 | 2432.19 | 487.44 | |
| 2688.32 | 897.11 | 3 | 2556.27 | 853.09 | 3 | 2272.08 | 758.36 | 3 | 2454.19 | 819.06 | 3 | 2453.20 | 818.73 | |
| 2689.33 | 673.33 | 4 | 2557.03 | 853.34 | 3 | 2286.13 | 763.04 | 3 | 2457.03 | 1229.52 | 2 | 2457.00 | 1229.50 | |
| 4250.86 | 1063.71 | 4 | 2557.09 | 853.36 | 3 | 2300.13 | 767.71 | 3 | 2510.07 | 837.69 | 3 | 2509.15 | 837.38 | |
| 4271.98 | 1069.00 | 4 | 2588.33 | 648.08 | 4 | 2304.04 | 769.01 | 3 | 2533.26 | 634.31 | 4 | 2531.26 | 633.82 | |
| 4401.05 | 1101.26 | 4 | 2601.13 | 868.04 | 3 | 2326.08 | 776.36 | 3 | 2542.01 | 848.34 | 3 | 2539.99 | 847.66 | |
| | | | 2682.23 | 895.08 | 3 | 2486.08 | 829.69 | 3 | 2542.02 | 636.51 | 4 | 2540.00 | 636.00 | |
| | | | 2700.39 | 541.08 | 5 | 2798.29 | 933.76 | 3 | 2558.02 | 853.67 | 3 | 2555.99 | 853.00 | |
| | | | 2700.40 | 451.07 | 6 | 2804.93 | 702.23 | 4 | 2568.17 | 857.06 | 3 | 2567.18 | 856.73 | |
| | | | 2717.31 | 906.77 | 3 | 3176.63 | 795.16 | 4 | 2573.00 | 515.60 | 5 | 2571.34 | 515.27 | |
| | | | 2728.29 | 910.43 | 3 | 3446.68 | 575.45 | 6 | 2588.33 | 648.08 | 4 | 2587.30 | 647.83 | |
| | | | 2754.29 | 689.57 | 4 | 3464.65 | 1155.88 | 3 | 2622.19 | 525.44 | 5 | 2621.18 | 525.24 | |
| | | | 2771.28 | 924.76 | 3 | 3478.60 | 1160.53 | 3 | 2635.31 | 659.83 | 4 | 2634.33 | 659.58 | |
| | | | 2771.29 | 693.82 | 4 | | | | 2689.33 | 673.33 | 4 | 2688.33 | 673.08 | |
| | | | 2796.34 | 700.08 | 4 | | | | 2717.31 | 906.77 | 3 | 2716.20 | 906.40 | |
| | | | 2806.30 | 702.58 | 4 | | | | 2771.28 | 924.76 | 3 | 2770.29 | 924.43 | |
| | | | 2814.27 | 1408.13 | 2 | | | | 2792.33 | 559.47 | 5 | 2791.33 | 559.27 | |
| | | | 2820.27 | 941.09 | 3 | | | | 2796.34 | 700.08 | 4 | 2797.31 | 700.33 | |
| | | | 2837.39 | 568.48 | 5 | | | | 2837.39 | 568.48 | 5 | 2835.35 | 568.07 | |
| | | | 2840.23 | 569.05 | 5 | | | | 2852.23 | 714.06 | 4 | 2853.23 | 714.31 | |
| | | | 2934.30 | 979.10 | 3 | | | | 2934.30 | 979.10 | 3 | 2931.53 | 978.18 | |
| | | | 2939.27 | 980.76 | 3 | | | | 4221.94 | 1056.49 | 4 | 4219.93 | 1055.98 | |
| | | | 3202.45 | 534.74 | 6 | | | | 4267.86 | 1067.96 | 4 | 4266.84 | 1067.71 | |
| | | | 3509.72 | 1170.91 | 3 | | | | | | | | | |
| | | | 3539.64 | 1180.88 | 3 | | | | | | | | | |
| | | | 4026.86 | 1343.29 | 3 | | | | | | | | | |
| | | | 4026.87 | 1007.72 | 4 | | | | | | | | | |
| | | | 4065.83 | 1356.28 | 3 | | | | | | | | | |
| | | | 4083.84 | 1362.28 | 3 | | | | | | | | | |
| | | | 4083.91 | 1021.98 | 4 | | | | | | | | | |
| | | | 4221.94 | 1056.49 | 4 | | | | | | | | | |
| | | | 4250.84 | 1417.95 | 3 | | | | | | | | | |
| | | | 4267.84 | 1423.61 | 3 | | | | | | | | | |
| | | | 4267.86 | 1067.96 | 4 | | | | | | | | | |
| | | | 4503.06 | 1126.77 | 4 | | | | | | | | | |
| | | | 4505.07 | 902.01 | 5 | | | | | | | | | |
| | | | 4518.07 | 1130.52 | 4 | | | | | | | | | |
| | | | 4536.04 | 1135.01 | 4 | | | | | | | | | |
| | | | 5365.51 | 1342.38 | 4 | | | | | | | | | |
| | | | 5572.89 | 1115.58 | 5 | | | | | | | | | |
| | | | 5572.90 | 929.82 | 6 | | | | | | | | | |
| | | | 5572.91 | 797.13 | 7 | | | | | | | | | |
| | | | 5572.93 | 697.62 | 8 | | | | | | | | | |
| | | | 6250.02 | 1563.50 | 4 | | | | | | | | | |

A.4. Calmodulin-Melittin Cross-linked Candidates From First-Stage Mass Spectrometry

Table A.3: MS Candidate Cross-linked Species for PFA

| Peptides | | | Insufficient MSMS | | | Undetermined | | | Incorrect Assignments | | | | | |
|----------|---------|---|-------------------|---------|----|--------------|-----|---|-----------------------|---------|---|---------------|------------|----------|
| [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | Actual [M]exp | Actual m/z | Actual z |
| 2076.95 | 1039.48 | 2 | 2528.18 | 843.73 | 3 | | | | 2456.16 | 819.72 | 3 | 2454.28 | 819.09 | |
| 2100.04 | 701.01 | 3 | 2535.09 | 1268.55 | 2 | | | | 2463.51 | 616.88 | 4 | 2462.52 | 616.63 | |
| 2106.05 | 703.02 | 3 | 2565.27 | 642.32 | 4 | | | | 2475.17 | 826.06 | 3 | 2472.18 | 825.06 | |
| 2114.75 | 1058.38 | 2 | 2623.24 | 1312.62 | 2 | | | | 2475.51 | 619.88 | 4 | 2475.20 | 619.80 | |
| 2127.04 | 710.01 | 3 | 2623.27 | 875.42 | 3 | | | | 2491.21 | 623.80 | 4 | 2489.67 | 623.42 | |
| 2144.06 | 715.69 | 3 | 2626.54 | 657.63 | 4 | | | | 2569.54 | 643.39 | 4 | 2565.26 | 642.31 | |
| 2145.05 | 716.02 | 3 | 2627.28 | 657.82 | 4 | | | | 2575.02 | 644.75 | 4 | 2573.38 | 644.35 | |
| 2154.09 | 1078.04 | 2 | 2652.30 | 885.10 | 3 | | | | 2603.00 | 651.75 | 4 | 2602.28 | 651.57 | |
| 2154.10 | 719.03 | 3 | 2655.29 | 886.10 | 3 | | | | 2626.54 | 657.63 | 4 | 2625.30 | 657.32 | |
| 2158.06 | 720.35 | 3 | 2668.28 | 890.43 | 3 | | | | 2628.28 | 877.09 | 3 | 2627.28 | 876.76 | |
| 2162.07 | 721.69 | 3 | 2688.31 | 1345.15 | 2 | | | | 2701.31 | 676.33 | 4 | | | 3 |
| 2183.02 | 1092.51 | 2 | 2706.31 | 542.26 | 5 | | | | 2728.28 | 683.07 | 4 | 2726.27 | 682.57 | |
| 2188.06 | 730.35 | 3 | 2728.28 | 683.07 | 4 | | | | 2813.29 | 938.76 | 3 | | | 6 |
| 2196.06 | 733.02 | 3 | 2734.31 | 547.86 | 5 | | | | 2842.41 | 711.60 | 4 | | | 7 |
| 2205.99 | 1103.99 | 2 | 2792.35 | 699.09 | 4 | | | | 2845.20 | 712.30 | 4 | 2846.44 | 712.61 | |
| 2206.08 | 736.36 | 3 | 3003.41 | 601.68 | 5 | | | | 2851.30 | 951.43 | 3 | 2846.43 | 949.81 | |
| 2210.12 | 737.71 | 3 | 3003.42 | 751.85 | 4 | | | | 2905.76 | 727.44 | 4 | 2903.41 | 726.85 | |
| 2218.36 | 1110.18 | 2 | 3051.65 | 763.91 | 4 | | | | 2938.56 | 735.64 | 4 | 2938.37 | 735.59 | |
| 2218.37 | 740.46 | 3 | 3053.37 | 1527.68 | 2 | | | | 2974.41 | 744.60 | 4 | 2975.40 | 744.85 | |
| 2218.37 | 555.59 | 4 | 3147.48 | 1050.16 | 3 | | | | 3069.47 | 768.37 | 4 | 3070.44 | 768.61 | |
| 2230.36 | 744.45 | 3 | 3352.78 | 839.19 | 4 | | | | 3147.47 | 787.87 | 4 | 3147.47 | 787.87 | |
| 2230.37 | 558.59 | 4 | 3377.57 | 1126.86 | 3 | | | | 3191.85 | 1064.95 | 3 | 3194.43 | 1065.81 | |
| 2230.78 | 1116.39 | 2 | 3402.65 | 1135.22 | 3 | | | | 3389.64 | 1130.88 | 3 | 3385.72 | 1129.57 | |
| 2238.03 | 1120.02 | 2 | 3889.70 | 1297.57 | 3 | | | | 4773.37 | 796.56 | 6 | 4776.60 | 797.10 | |
| 2238.15 | 747.05 | 3 | 3889.72 | 1297.57 | 3 | | | | 4923.28 | 985.66 | 5 | 806.41 | 984.85 | |
| 2242.37 | 748.46 | 3 | 4198.07 | 600.72 | 7 | | | | 4964.24 | 1242.06 | 4 | 4963.24 | 1241.81 | |
| 2242.37 | 561.59 | 4 | 4520.08 | 1131.02 | 4 | | | | 5211.35 | 1303.84 | 4 | 5207.64 | 1302.91 | |
| 2247.38 | 450.48 | 5 | 4520.14 | 754.36 | 6 | | | | 6774.22 | 1130.04 | 6 | 6769.14 | 1129.19 | |
| 2248.38 | 563.09 | 4 | 4687.25 | 1563.42 | 3 | | | | | | | | | |
| 2248.38 | 1125.19 | 2 | 4773.37 | 796.56 | 6 | | | | | | | | | |
| 2255.06 | 1128.53 | 2 | 4947.27 | 825.54 | 6 | | | | | | | | | |
| 2260.37 | 754.46 | 3 | 4948.24 | 1238.06 | 4 | | | | | | | | | |
| 2260.38 | 566.10 | 4 | 4949.25 | 1238.31 | 4 | | | | | | | | | |
| 2341.16 | 781.39 | 3 | 4965.26 | 994.05 | 5 | | | | | | | | | |
| 2368.19 | 474.64 | 5 | 5211.35 | 1303.84 | 4 | | | | | | | | | |
| 2381.97 | 794.99 | 3 | 5261.60 | 1316.40 | 4 | | | | | | | | | |
| 2388.11 | 1195.06 | 2 | 5573.89 | 929.98 | 6 | | | | | | | | | |
| 2395.20 | 799.40 | 3 | 5573.89 | 1115.78 | 5 | | | | | | | | | |
| 2416.93 | 605.23 | 4 | 5573.91 | 929.98 | 6 | | | | | | | | | |
| 2422.32 | 606.58 | 4 | 5573.92 | 797.27 | 7 | | | | | | | | | |
| 2427.17 | 607.79 | 4 | 5573.93 | 697.74 | 8 | | | | | | | | | |
| 2429.18 | 608.29 | 4 | 6080.82 | 1014.47 | 6 | | | | | | | | | |
| 2432.93 | 609.23 | 4 | 6096.80 | 1017.13 | 6 | | | | | | | | | |
| 2443.18 | 815.39 | 3 | 6110.81 | 1019.47 | 6 | | | | | | | | | |
| 2449.17 | 817.39 | 3 | 6122.81 | 1021.47 | 6 | | | | | | | | | |
| 2462.20 | 616.55 | 4 | 6745.12 | 1350.02 | 5 | | | | | | | | | |
| 2464.32 | 1233.16 | 2 | 6745.13 | 1125.19 | 6 | | | | | | | | | |
| 2474.21 | 495.84 | 5 | 6769.12 | 1354.82 | 5 | | | | | | | | | |
| 2475.17 | 826.06 | 3 | 6769.13 | 1129.19 | 6 | | | | | | | | | |
| 2475.51 | 619.88 | 4 | 8636.07 | 786.10 | 11 | | | | | | | | | |
| 2475.63 | 826.21 | 3 | | | | | | | | | | | | |
| 2480.64 | 827.88 | 3 | | | | | | | | | | | | |
| 2486.09 | 1244.04 | 2 | | | | | | | | | | | | |
| 2486.11 | 829.70 | 3 | | | | | | | | | | | | |
| 2488.10 | 830.37 | 3 | | | | | | | | | | | | |
| 2491.21 | 623.80 | 4 | | | | | | | | | | | | |
| 2503.08 | 835.36 | 3 | | | | | | | | | | | | |
| 2517.09 | 840.03 | 3 | | | | | | | | | | | | |
| 2532.18 | 845.06 | 3 | | | | | | | | | | | | |
| 2532.23 | 634.06 | 4 | | | | | | | | | | | | |
| 2535.09 | 1268.54 | 2 | | | | | | | | | | | | |
| 2554.21 | 852.40 | 3 | | | | | | | | | | | | |
| 2554.30 | 852.43 | 3 | | | | | | | | | | | | |
| 2554.41 | 855.70 | 2 | | | | | | | | | | | | |

A.4. Calmodulin-Melittin Cross-linked Candidates From First-Stage Mass Spectrometry

Table A.4: MS Candidate Cross-linked Species for PFA

| Peptides | | | Insufficient MSMS | | | Undetermined | | | Incorrect Assignments | | | | | |
|----------|---------|---|-------------------|-----|---|--------------|-----|---|-----------------------|-----|---|--------|-----|----------|
| [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | Actual z |
| 2564.11 | 855.70 | 3 | | | | | | | | | | | | |
| 2566.10 | 1284.05 | 2 | | | | | | | | | | | | |
| 2569.54 | 643.39 | 4 | | | | | | | | | | | | |
| 2574.01 | 859.00 | 3 | | | | | | | | | | | | |
| 2575.02 | 644.75 | 4 | | | | | | | | | | | | |
| 2585.25 | 862.75 | 3 | | | | | | | | | | | | |
| 2590.27 | 648.57 | 4 | | | | | | | | | | | | |
| 2591.20 | 864.73 | 3 | | | | | | | | | | | | |
| 2617.22 | 873.41 | 3 | | | | | | | | | | | | |
| 2618.29 | 655.57 | 4 | | | | | | | | | | | | |
| 2628.28 | 877.09 | 3 | | | | | | | | | | | | |
| 2631.29 | 658.82 | 4 | | | | | | | | | | | | |
| 2655.30 | 886.10 | 3 | | | | | | | | | | | | |
| 2661.32 | 666.33 | 4 | | | | | | | | | | | | |
| 2671.31 | 668.83 | 4 | | | | | | | | | | | | |
| 2672.31 | 891.77 | 3 | | | | | | | | | | | | |
| 2688.31 | 1345.15 | 2 | | | | | | | | | | | | |
| 2688.32 | 897.11 | 3 | | | | | | | | | | | | |
| 2688.32 | 673.08 | 4 | | | | | | | | | | | | |
| 2689.43 | 897.48 | 3 | | | | | | | | | | | | |
| 2706.33 | 677.58 | 4 | | | | | | | | | | | | |
| 2718.32 | 907.11 | 3 | | | | | | | | | | | | |
| 2807.35 | 702.84 | 4 | | | | | | | | | | | | |
| 2974.41 | 744.60 | 4 | | | | | | | | | | | | |
| 3051.66 | 763.91 | 4 | | | | | | | | | | | | |
| 3147.47 | 787.87 | 4 | | | | | | | | | | | | |
| 3191.85 | 1064.95 | 3 | | | | | | | | | | | | |
| 3200.49 | 1067.83 | 3 | | | | | | | | | | | | |
| 3402.64 | 1135.21 | 3 | | | | | | | | | | | | |
| 4544.08 | 909.82 | 5 | | | | | | | | | | | | |
| 4545.08 | 1137.27 | 4 | | | | | | | | | | | | |
| 4964.24 | 1242.06 | 4 | | | | | | | | | | | | |

A.4. Calmodulin-Melittin Cross-linked Candidates From First-Stage Mass Spectrometry

Table A.5: MS Candidate Cross-linked Species for sulfoDST

| Peptides | | | Insufficient MSMS | | | Undetermined | | | Incorrect Assignments | | | | |
|----------|---------|---|-------------------|---------|---|--------------|--------|---|-----------------------|-----|--------|---------------|------------|
| [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | Actual [M]exp | Actual m/z |
| 859.48 | 430.74 | 2 | 788.44 | 395.22 | 2 | 1220.72 | 611.36 | 2 | 1855.88 | 3 | 619.63 | 1854.87 | 619.29 |
| 891.51 | 446.75 | 2 | 937.47 | 469.74 | 2 | 1243.62 | 622.81 | 2 | 2175.08 | 3 | 726.03 | 2174.09 | 725.70 |
| 1228.60 | 615.30 | 2 | 1072.65 | 358.55 | 3 | 1357.71 | 679.86 | 2 | 2369.19 | 4 | 593.30 | 2368.20 | 593.05 |
| 1577.72 | 526.91 | 3 | 1144.57 | 573.29 | 2 | 1525.92 | 509.64 | 3 | 2520.19 | 3 | 841.06 | 2521.09 | 841.36 |
| 1714.77 | 572.59 | 3 | 1176.69 | 589.35 | 2 | 1526.00 | 509.67 | 3 | 2847.44 | 4 | 712.86 | 2846.45 | 712.61 |
| 1742.88 | 581.96 | 3 | 1195.68 | 598.84 | 2 | 2175.08 | 726.03 | 3 | 2807.36 | 4 | 702.84 | 2806.36 | 702.59 |
| 1855.88 | 619.63 | 3 | 1228.60 | 615.30 | 2 | 2444.20 | 612.05 | 4 | | | | | |
| 1857.95 | 620.32 | 3 | 1352.79 | 677.40 | 2 | | | | | | | | |
| 1868.87 | 935.43 | 2 | 1352.80 | 677.40 | 2 | | | | | | | | |
| 2017.89 | 1009.95 | 2 | 1365.66 | 683.83 | 2 | | | | | | | | |
| 2154.11 | 719.04 | 3 | 1525.92 | 509.64 | 3 | | | | | | | | |
| 2172.06 | 725.02 | 3 | 1558.92 | 520.64 | 3 | | | | | | | | |
| 2369.19 | 593.30 | 4 | 1634.77 | 818.39 | 2 | | | | | | | | |
| 2429.20 | 608.30 | 4 | 1723.83 | 431.96 | 4 | | | | | | | | |
| 2444.19 | 815.73 | 3 | 1939.95 | 647.65 | 3 | | | | | | | | |
| 2520.19 | 841.06 | 3 | 2128.04 | 710.35 | 3 | | | | | | | | |
| 2520.25 | 631.06 | 4 | 2131.06 | 711.35 | 3 | | | | | | | | |
| 2847.44 | 712.86 | 4 | 2145.06 | 716.02 | 3 | | | | | | | | |
| 4116.85 | 1373.28 | 3 | 2210.13 | 737.71 | 3 | | | | | | | | |
| 4116.86 | 1030.22 | 4 | 2233.15 | 559.29 | 4 | | | | | | | | |
| 4116.86 | 1030.22 | 4 | 2234.04 | 559.51 | 4 | | | | | | | | |
| | | | 2369.19 | 593.30 | 4 | | | | | | | | |
| | | | 2444.20 | 612.05 | 4 | | | | | | | | |
| | | | 2501.25 | 834.75 | 3 | | | | | | | | |
| | | | 2525.30 | 842.77 | 3 | | | | | | | | |
| | | | 2787.46 | 930.15 | 3 | | | | | | | | |
| | | | 2911.43 | 728.86 | 4 | | | | | | | | |
| | | | 3191.44 | 1064.81 | 3 | | | | | | | | |
| | | | 3353.77 | 839.44 | 4 | | | | | | | | |
| | | | 4116.85 | 1373.28 | 3 | | | | | | | | |
| | | | 4116.86 | 1030.22 | 4 | | | | | | | | |
| | | | 5262.61 | 1316.65 | 4 | | | | | | | | |

A.4. Calmodulin-Melittin Cross-linked Candidates From First-Stage Mass Spectrometry

Table A.6: MS Candidate Cross-linked Species for BS³

| Peptides | | | Insufficient MSMS | | | Undetermined | | | Incorrect Assignments | | | | |
|----------|---------|---|-------------------|---------|---|--------------|---------|---|-----------------------|--------|---|---------------|------------|
| [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | Actual [M]exp | Actual m/z |
| 1602.78 | 535.26 | 3 | 898.52 | 450.26 | 2 | 898.46 | 450.23 | 2 | 842.48 | 422.24 | 2 | 841.58 | 421.79 |
| 1651.78 | 826.89 | 2 | 1074.66 | 538.33 | 2 | 922.52 | 462.26 | 2 | 1239.67 | 414.22 | 3 | 1237.65 | 413.55 |
| 1651.79 | 551.60 | 3 | 1154.55 | 578.27 | 2 | 1074.47 | 1075.47 | 1 | 1388.69 | 695.34 | 2 | 1387.68 | 694.84 |
| 1686.70 | 563.23 | 3 | 1217.74 | 406.91 | 3 | 1154.60 | 578.30 | 2 | 1602.78 | 535.26 | 3 | 1603.79 | 535.60 |
| 1704.85 | 569.28 | 3 | 1220.71 | 611.36 | 2 | 1217.73 | 609.86 | 2 | 1721.85 | 574.95 | 3 | 1720.83 | 574.61 |
| 1738.84 | 870.42 | 2 | 1254.63 | 628.32 | 2 | 1244.75 | 623.37 | 2 | 1925.97 | 642.99 | 3 | 1923.98 | 642.33 |
| 1738.86 | 435.72 | 4 | 1347.73 | 450.24 | 3 | 1503.63 | 502.21 | 3 | 1926.19 | 643.06 | 3 | 1924.97 | 642.66 |
| 1925.97 | 642.99 | 3 | 1388.69 | 695.34 | 2 | 1679.78 | 840.89 | 2 | 1963.88 | 655.63 | 3 | 1963.07 | 655.36 |
| 1963.88 | 655.63 | 3 | 1400.67 | 701.33 | 2 | 2107.32 | 527.83 | 4 | 2108.04 | 528.01 | 4 | 2109.04 | 528.26 |
| 2108.04 | 528.01 | 4 | 1763.87 | 882.93 | 2 | 2117.11 | 706.70 | 3 | 2194.17 | 549.54 | 4 | 2195.16 | 549.79 |
| 2179.08 | 1090.54 | 2 | 1770.88 | 886.44 | 2 | 2210.16 | 553.54 | 4 | 2210.16 | 553.54 | 4 | 2207.80 | 552.95 |
| 2210.16 | 553.54 | 4 | 1791.86 | 896.93 | 2 | 2225.10 | 742.70 | 3 | 2225.07 | 557.27 | 4 | 2223.16 | 556.79 |
| 2225.07 | 557.27 | 4 | 1926.19 | 482.55 | 4 | 2604.34 | 869.11 | 3 | 2346.23 | 587.56 | 4 | 2345.23 | 587.31 |
| 2346.23 | 587.56 | 4 | 2107.32 | 703.44 | 3 | 2730.33 | 911.11 | 3 | 2730.33 | 911.11 | 3 | 2730.08 | 911.03 |
| 2433.06 | 609.27 | 4 | 2117.11 | 1059.56 | 2 | 3267.54 | 1090.18 | 3 | 2794.46 | 699.61 | 4 | 2793.46 | 699.36 |
| 2501.30 | 834.77 | 3 | 2117.11 | 706.70 | 3 | | | | 2800.39 | 701.10 | 4 | 2798.37 | 700.59 |
| 2503.25 | 626.81 | 4 | 2194.15 | 1098.08 | 2 | | | | | | | | |
| 2784.37 | 1393.19 | 2 | 2194.16 | 732.39 | 3 | | | | | | | | |
| 2784.39 | 697.10 | 4 | 2210.13 | 737.71 | 3 | | | | | | | | |
| 2800.39 | 701.10 | 4 | 2225.10 | 1113.55 | 2 | | | | | | | | |
| 2933.54 | 734.38 | 4 | 2295.13 | 766.04 | 3 | | | | | | | | |
| 2961.38 | 741.35 | 4 | 2414.17 | 1208.09 | 2 | | | | | | | | |
| | | | 2472.09 | 1237.04 | 2 | | | | | | | | |
| | | | 2548.37 | 1275.19 | 2 | | | | | | | | |
| | | | 2649.36 | 663.34 | 4 | | | | | | | | |
| | | | 3052.60 | 764.15 | 4 | | | | | | | | |
| | | | 3191.45 | 1064.82 | 3 | | | | | | | | |
| | | | 3267.53 | 1634.76 | 2 | | | | | | | | |
| | | | 3472.72 | 1158.57 | 3 | | | | | | | | |

A.4. Calmodulin-Melittin Cross-linked Candidates From First-Stage Mass Spectrometry

Table A.7: MS Candidate Cross-linked Species for sulfoEGS

| Peptides | | | Insufficient MSMS | | | Undetermined | | | Incorrect Assignments | | | | | |
|----------|-----|---------|-------------------|-----|---------|--------------|-----|---------|-----------------------|-----|---------|---------------|------------|----------|
| [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | Actual [M]exp | Actual m/z | Actual z |
| 1089.54 | 2 | 545.77 | 744.31 | 2 | 373.16 | 859.52 | 2 | 430.76 | 1046.51 | 2 | 524.25 | 1044.57 | 523.29 | |
| 1298.73 | 3 | 433.91 | 744.41 | 2 | 373.21 | 1105.54 | 2 | 553.77 | 1202.67 | 3 | 401.89 | 1201.66 | 401.55 | |
| 1467.62 | 2 | 734.81 | 780.42 | 2 | 391.21 | 1503.78 | 2 | 752.89 | 1340.69 | 3 | 447.90 | 1339.69 | 447.56 | |
| 1537.73 | 3 | 513.58 | 859.45 | 2 | 430.73 | 1671.84 | 3 | 558.28 | 1340.69 | 2 | 671.35 | 1339.71 | 670.85 | |
| 1549.82 | 2 | 775.91 | 859.52 | 2 | 430.76 | 2026.02 | 3 | 676.34 | 1347.61 | 2 | 674.81 | 1346.65 | 674.32 | |
| 1567.73 | 3 | 523.58 | 1074.66 | 2 | 538.33 | 2216.05 | 3 | 739.68 | 1452.72 | 3 | 485.24 | 1453.71 | 485.57 | |
| 1639.75 | 3 | 547.58 | 1107.63 | 3 | 370.21 | 2488.23 | 4 | 623.06 | 1567.73 | 3 | 523.58 | 1565.72 | 522.91 | |
| 1639.79 | 2 | 820.89 | 1187.52 | 2 | 594.76 | 2557.21 | 4 | 640.30 | 1584.77 | 2 | 793.39 | 1581.96 | 791.98 | |
| 1659.78 | 2 | 830.89 | 1202.65 | 2 | 602.33 | 2584.28 | 4 | 647.07 | 1615.78 | 2 | 808.89 | 1614.67 | 808.34 | |
| 1665.82 | 2 | 833.91 | 1202.67 | 3 | 401.89 | 2676.32 | 5 | 536.26 | 1641.98 | 2 | 821.99 | 1640.80 | 821.40 | |
| 1667.77 | 3 | 556.92 | 1206.61 | 2 | 604.31 | 2734.32 | 5 | 547.86 | 1667.77 | 3 | 556.92 | 1667.04 | 556.68 | |
| 1739.77 | 2 | 870.88 | 1206.74 | 2 | 604.37 | 3195.67 | 5 | 640.13 | 1684.04 | 4 | 422.01 | 1682.82 | 421.71 | |
| 1739.77 | 3 | 580.92 | 1223.69 | 4 | 306.92 | 6032.82 | 6 | 1006.47 | 1834.80 | 2 | 918.40 | 1835.89 | 918.95 | |
| 1739.83 | 2 | 870.91 | 1287.88 | 7 | 184.98 | | | | 1926.90 | 4 | 482.72 | 1927.88 | 482.97 | |
| 1764.82 | 3 | 589.27 | 1345.71 | 3 | 449.57 | | | | 1926.99 | 2 | 964.49 | 1925.91 | 963.96 | |
| 1859.91 | 3 | 620.97 | 1345.71 | 3 | 449.57 | | | | 1944.94 | 3 | 649.31 | 1943.93 | 648.98 | |
| 1865.86 | 4 | 467.46 | 1386.61 | 2 | 694.30 | | | | 2015.93 | 2 | 1008.97 | 2012.92 | 1007.46 | |
| 1926.91 | 3 | 643.30 | 1428.82 | 2 | 715.41 | | | | 2025.97 | 4 | 507.49 | 2024.98 | 507.24 | |
| 1926.99 | 2 | 964.49 | 1439.70 | 5 | 288.94 | | | | 2095.97 | 4 | 524.99 | 2094.95 | 524.74 | |
| 1952.87 | 3 | 651.96 | 1453.71 | 3 | 485.57 | | | | 2114.99 | 4 | 529.75 | 2113.00 | 529.25 | |
| 1952.91 | 2 | 977.45 | 1538.80 | 3 | 513.93 | | | | 2118.99 | 2 | 1060.50 | 2117.97 | 1059.98 | |
| 1956.97 | 3 | 653.32 | 1539.01 | 9 | 172.00 | | | | 2185.34 | 4 | 547.33 | 2184.05 | 547.01 | |
| 1987.94 | 2 | 994.97 | 1589.61 | 5 | 318.92 | | | | 2188.06 | 3 | 730.35 | 2185.01 | 729.34 | |
| 1987.95 | 3 | 663.65 | 1604.76 | 2 | 803.38 | | | | 2283.14 | 4 | 571.79 | 2282.15 | 571.54 | |
| 2015.93 | 2 | 1008.97 | 1644.61 | 5 | 329.92 | | | | 2292.13 | 4 | 574.03 | 2291.13 | 573.78 | |
| 2025.97 | 4 | 507.49 | 1651.94 | 4 | 413.98 | | | | 2298.11 | 3 | 767.04 | 2297.64 | 766.88 | |
| 2046.99 | 2 | 1024.49 | 1684.04 | 4 | 422.01 | | | | 2298.18 | 4 | 575.54 | 2295.13 | 574.78 | |
| 2064.96 | 2 | 1033.48 | 1684.04 | 4 | 422.01 | | | | 2301.09 | 2 | 1151.55 | 2298.07 | 1150.04 | |
| 2064.97 | 3 | 689.32 | 1684.80 | 3 | 562.60 | | | | 2310.07 | 3 | 771.02 | 2309.03 | 770.68 | |
| 2065.00 | 5 | 414.00 | 1695.03 | 3 | 566.01 | | | | 2370.20 | 5 | 475.04 | 2376.04 | 476.21 | |
| 2095.97 | 4 | 524.99 | 1851.86 | 4 | 463.97 | | | | 2383.18 | 4 | 596.80 | 2384.18 | 597.04 | |
| 2108.99 | 3 | 704.00 | 1885.93 | 5 | 378.19 | | | | 2385.15 | 2 | 1193.58 | 2384.17 | 1193.09 | |
| 2127.01 | 3 | 710.00 | 1903.93 | 3 | 635.64 | | | | 2429.18 | 3 | 810.73 | 2428.20 | 810.40 | |
| 2142.99 | 2 | 1072.49 | 1926.90 | 4 | 482.72 | | | | 2429.19 | 4 | 608.30 | 2428.20 | 608.05 | |
| 2143.01 | 4 | 536.75 | 1944.94 | 3 | 649.31 | | | | 2456.11 | 3 | 819.70 | 2455.08 | 819.36 | |
| 2269.12 | 4 | 568.28 | 1951.97 | 5 | 391.39 | | | | 2470.11 | 3 | 824.37 | 2469.12 | 824.04 | |
| 2269.12 | 3 | 757.37 | 1956.96 | 2 | 979.48 | | | | 2486.11 | 3 | 829.70 | 2485.10 | 829.37 | |
| 2310.07 | 3 | 771.02 | 2014.91 | 3 | 672.64 | | | | 2528.24 | 4 | 633.06 | 2527.06 | 632.76 | |
| 2352.19 | 4 | 589.05 | 2050.96 | 4 | 513.74 | | | | 2584.28 | 4 | 647.07 | 2582.51 | 646.63 | |
| 2383.16 | 5 | 477.63 | 2114.95 | 3 | 705.98 | | | | 2689.32 | 4 | 673.33 | 2688.34 | 673.09 | |
| 2383.18 | 4 | 596.80 | 2114.99 | 4 | 529.75 | | | | 2720.33 | 4 | 681.08 | 2719.32 | 680.83 | |
| 2429.18 | 3 | 810.73 | 2127.00 | 2 | 1064.50 | | | | 2857.35 | 3 | 953.45 | 2856.37 | 953.12 | |
| 2429.19 | 4 | 608.30 | 2143.00 | 3 | 715.33 | | | | 2873.37 | 3 | 958.79 | 2872.36 | 958.45 | |
| 2444.18 | 3 | 815.73 | 2209.01 | 2 | 1105.51 | | | | 2952.43 | 4 | 739.11 | 2954.38 | 739.59 | |
| 2470.11 | 3 | 824.37 | 2209.02 | 3 | 737.34 | | | | 3019.42 | 3 | 1007.47 | 3018.43 | 1007.14 | |
| 2689.32 | 4 | 673.33 | 2216.03 | 4 | 555.01 | | | | 3400.63 | 3 | 1134.54 | 3399.64 | 1134.21 | |
| 2712.33 | 3 | 905.11 | 2301.09 | 2 | 1151.55 | | | | 3743.68 | 3 | 1248.89 | 3742.67 | 1248.56 | |
| 2712.34 | 4 | 679.08 | 2346.11 | 4 | 587.53 | | | | 3822.83 | 4 | 956.71 | 3823.85 | 956.96 | |
| 2728.31 | 2 | 1365.15 | 2369.19 | 4 | 593.30 | | | | 3824.85 | 5 | 765.97 | 3823.84 | 765.77 | |
| 2728.32 | 3 | 910.44 | 2372.18 | 4 | 594.05 | | | | | | | | | |
| 2728.33 | 4 | 683.08 | 2382.97 | 2 | 1192.48 | | | | | | | | | |
| 2746.33 | 3 | 916.44 | 2382.98 | 3 | 795.33 | | | | | | | | | |
| 2746.34 | 4 | 687.58 | 2385.15 | 2 | 1193.58 | | | | | | | | | |
| 2872.35 | 2 | 1437.18 | 2486.11 | 3 | 829.70 | | | | | | | | | |
| 2872.38 | 4 | 719.09 | 2528.24 | 4 | 633.06 | | | | | | | | | |
| 2873.37 | 3 | 958.79 | 2603.09 | 2 | 1302.55 | | | | | | | | | |
| 2890.36 | 2 | 1446.18 | 2665.14 | 2 | 1333.57 | | | | | | | | | |
| 2890.38 | 4 | 723.60 | 2676.31 | 5 | 536.26 | | | | | | | | | |
| 3019.42 | 3 | 1007.47 | 2716.39 | 5 | 544.28 | | | | | | | | | |
| 3034.42 | 3 | 1012.47 | 2765.16 | 3 | 922.72 | | | | | | | | | |
| 3034.43 | 4 | 759.61 | 2790.36 | 3 | 931.12 | | | | | | | | | |
| | | | 2790.36 | 4 | 698.59 | | | | | | | | | |

A.4. Calmodulin-Melittin Cross-linked Candidates From First-Stage Mass Spectrometry

Table A.8: MS Candidate Cross-linked Species for sulfoEGS

| Peptides | | | Insufficient MSMS | | | Undetermined | | | Incorrect Assignments | | | | |
|----------|-----|---|-------------------|-----|---------|--------------|-----|---|-----------------------|-----|---|------------------|---------------|
| [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | Actual [M]exp | Actual m/z |
| | | | 2790.37 | 4 | 698.59 | | | | | | | | |
| | | | 3016.45 | 4 | 755.11 | | | | | | | | |
| | | | 3155.48 | 3 | 1052.83 | | | | | | | | |
| | | | 3191.45 | 3 | 1064.82 | | | | | | | | |
| | | | 3400.63 | 3 | 1134.54 | | | | | | | | |
| | | | 3456.69 | 6 | 577.11 | | | | | | | | |
| | | | 3456.70 | 7 | 494.81 | | | | | | | | |
| | | | 3543.67 | 3 | 1182.22 | | | | | | | | |
| | | | 3561.68 | 3 | 1188.23 | | | | | | | | |
| | | | 3646.96 | 4 | 912.74 | | | | | | | | |
| | | | 3795.84 | 3 | 1266.28 | | | | | | | | |
| | | | 3834.81 | 3 | 1279.27 | | | | | | | | |
| | | | 3995.88 | 3 | 1332.96 | | | | | | | | |
| | | | 4099.87 | 4 | 1025.97 | | | | | | | | |
| | | | 4952.27 | 5 | 991.45 | | | | | | | | |
| | | | 5507.64 | 4 | 1377.91 | | | | | | | | |
| | | | 8702.11 | 10 | 871.21 | | | | | | | | |

A.5. Ribonuclease S Cross-linked Candidates

Table A.12: Candidate Cross-linked Species for sulfoDST

| [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z |
|---------|--------|---|---------|---------|---|---------|---------|----|----------|---------|----|
| 620.39 | 311.19 | 2 | 1872.10 | 625.03 | 3 | 3705.67 | 1853.83 | 2 | 4705.12 | 1569.37 | 3 |
| 620.40 | 311.20 | 2 | 1888.08 | 630.36 | 3 | 3707.63 | 412.96 | 9 | 4730.15 | 1577.72 | 3 |
| 620.40 | 311.20 | 2 | 1952.95 | 977.48 | 2 | 3707.68 | 1236.89 | 3 | 4759.18 | 1587.39 | 3 |
| 679.39 | 340.70 | 2 | 2009.05 | 1005.52 | 2 | 3726.67 | 1864.33 | 2 | 4776.19 | 1593.06 | 3 |
| 711.50 | 356.75 | 2 | 2032.14 | 1017.07 | 2 | 3726.67 | 1243.22 | 3 | 4848.19 | 1617.06 | 3 |
| 711.50 | 356.75 | 2 | 2035.08 | 1018.54 | 2 | 3726.68 | 1864.34 | 2 | 4930.25 | 987.05 | 5 |
| 851.38 | 426.69 | 2 | 2042.02 | 681.67 | 3 | 3726.68 | 1243.23 | 3 | 4930.28 | 1233.57 | 4 |
| 863.49 | 432.74 | 2 | 2067.99 | 414.60 | 5 | 3726.72 | 932.68 | 4 | 4934.22 | 1645.74 | 3 |
| 868.46 | 435.23 | 2 | 2089.16 | 1045.58 | 2 | 3735.69 | 1868.84 | 2 | 4940.24 | 1647.75 | 3 |
| 992.55 | 497.28 | 2 | 2148.07 | 717.02 | 3 | 3740.68 | 1247.89 | 3 | 4960.25 | 1654.42 | 3 |
| 1010.29 | 506.14 | 2 | 2218.02 | 1110.01 | 2 | 3750.71 | 1251.24 | 3 | 4964.27 | 1242.07 | 4 |
| 1023.48 | 512.74 | 2 | 2237.14 | 1119.57 | 2 | 3756.83 | 627.14 | 6 | 4969.99 | 1243.50 | 4 |
| 1081.56 | 541.78 | 2 | 2246.23 | 749.74 | 3 | 3771.68 | 1886.84 | 2 | 4971.02 | 995.20 | 5 |
| 1111.59 | 556.79 | 2 | 2262.74 | 1132.37 | 2 | 3771.69 | 1258.23 | 3 | 4987.26 | 1663.42 | 3 |
| 1118.69 | 560.35 | 2 | 2280.15 | 761.05 | 3 | 3782.94 | 344.90 | 11 | 4997.30 | 1666.77 | 3 |
| 1123.58 | 562.79 | 2 | 2370.19 | 593.55 | 4 | 3788.70 | 1895.35 | 2 | 5236.27 | 1310.07 | 4 |
| 1134.64 | 379.21 | 3 | 2380.20 | 477.04 | 5 | 3809.68 | 1905.84 | 2 | 5259.64 | 1754.21 | 3 |
| 1175.69 | 392.90 | 3 | 2454.12 | 1228.06 | 2 | 3809.70 | 1270.90 | 3 | 5278.71 | 1320.68 | 4 |
| 1191.67 | 398.22 | 3 | 2472.12 | 1237.06 | 2 | 3819.61 | 1910.81 | 2 | 5280.51 | 881.09 | 6 |
| 1196.73 | 599.36 | 2 | 2487.09 | 1244.55 | 2 | 3849.70 | 1925.85 | 2 | 5387.79 | 1347.95 | 4 |
| 1197.67 | 400.22 | 3 | 2498.14 | 625.53 | 4 | 3855.43 | 1286.14 | 3 | 5387.80 | 1078.56 | 5 |
| 1213.66 | 607.83 | 2 | 2506.21 | 836.40 | 3 | 3899.74 | 1300.91 | 3 | 5387.81 | 898.97 | 6 |
| 1236.63 | 619.31 | 2 | 2544.11 | 1273.05 | 2 | 3909.56 | 1955.78 | 2 | 5391.78 | 1079.36 | 5 |
| 1241.65 | 621.83 | 2 | 2544.12 | 849.04 | 3 | 3981.76 | 996.44 | 4 | 5531.70 | 1383.93 | 4 |
| 1249.67 | 625.83 | 2 | 2546.12 | 849.71 | 3 | 4051.47 | 1351.49 | 3 | 5547.50 | 1850.17 | 3 |
| 1249.67 | 625.84 | 2 | 2574.14 | 1288.07 | 2 | 4053.87 | 1352.29 | 3 | 5591.35 | 1119.27 | 5 |
| 1250.77 | 313.69 | 4 | 2623.27 | 1312.63 | 2 | 4084.86 | 1362.62 | 3 | 5594.49 | 1865.83 | 3 |
| 1266.62 | 423.21 | 3 | 2640.28 | 881.09 | 3 | 4084.87 | 1362.62 | 3 | 5651.34 | 1413.83 | 4 |
| 1310.65 | 656.33 | 2 | 2644.10 | 1323.05 | 2 | 4084.88 | 1022.22 | 4 | 5684.53 | 1422.13 | 4 |
| 1399.80 | 350.95 | 4 | 2655.30 | 886.10 | 3 | 4100.85 | 1367.95 | 3 | 5811.04 | 969.51 | 6 |
| 1407.74 | 352.93 | 4 | 2678.40 | 1340.20 | 2 | 4100.86 | 1367.95 | 3 | 5834.61 | 973.44 | 6 |
| 1431.83 | 358.96 | 4 | 2735.32 | 548.06 | 5 | 4119.05 | 687.51 | 6 | 5848.74 | 488.40 | 12 |
| 1447.87 | 362.97 | 4 | 2756.25 | 1379.12 | 2 | 4122.87 | 1031.72 | 4 | 5855.61 | 1952.87 | 3 |
| 1467.72 | 367.93 | 4 | 2778.21 | 1390.11 | 2 | 4126.88 | 1376.63 | 3 | 5884.56 | 1962.52 | 3 |
| 1467.86 | 490.29 | 3 | 2808.36 | 562.67 | 5 | 4138.88 | 1035.72 | 4 | 6204.85 | 1035.14 | 6 |
| 1507.74 | 503.58 | 3 | 2812.25 | 1407.12 | 2 | 4151.83 | 1038.96 | 4 | 6320.73 | 791.09 | 8 |
| 1559.93 | 520.98 | 3 | 2827.59 | 1414.80 | 2 | 4161.60 | 1041.40 | 4 | 6998.22 | 1167.37 | 6 |
| 1574.72 | 315.94 | 5 | 2830.21 | 1416.10 | 2 | 4184.93 | 1395.98 | 3 | 7627.40 | 1272.23 | 6 |
| 1574.75 | 788.38 | 2 | 2830.22 | 1416.11 | 2 | 4186.90 | 1047.72 | 4 | 7962.30 | 1328.05 | 6 |
| 1574.92 | 394.73 | 4 | 2849.24 | 1425.62 | 2 | 4214.15 | 422.42 | 10 | 11873.93 | 1485.24 | 8 |
| 1592.73 | 531.91 | 3 | 2876.30 | 959.77 | 3 | 4254.96 | 1064.74 | 4 | | | |
| 1610.79 | 806.40 | 2 | 2894.25 | 1448.12 | 2 | 4277.61 | 1426.87 | 3 | | | |
| 1623.00 | 406.75 | 4 | 2911.30 | 1456.65 | 2 | 4277.61 | 1070.40 | 4 | | | |
| 1623.97 | 812.99 | 2 | 2934.30 | 979.10 | 3 | 4294.61 | 1432.54 | 3 | | | |
| 1640.60 | 329.12 | 5 | 2934.31 | 1468.16 | 2 | 4322.62 | 1441.87 | 3 | | | |
| 1649.97 | 330.99 | 5 | 2962.37 | 1482.19 | 2 | 4352.96 | 1451.99 | 3 | | | |
| 1656.78 | 829.39 | 2 | 2991.35 | 998.12 | 3 | 4381.09 | 731.18 | 6 | | | |
| 1692.02 | 565.01 | 3 | 3002.38 | 1502.19 | 2 | 4401.06 | 1101.26 | 4 | | | |
| 1692.89 | 565.30 | 3 | 3024.36 | 1009.12 | 3 | 4404.01 | 1469.00 | 3 | | | |
| 1703.80 | 426.95 | 4 | 3048.38 | 1017.13 | 3 | 4428.02 | 1477.01 | 3 | | | |
| 1703.81 | 852.91 | 2 | 3078.63 | 1027.21 | 3 | 4433.04 | 1478.68 | 3 | | | |
| 1726.87 | 576.62 | 3 | 3146.41 | 1574.20 | 2 | 4458.02 | 1487.01 | 3 | | | |
| 1733.92 | 578.97 | 3 | 3210.44 | 1606.22 | 2 | 4489.78 | 562.22 | 8 | | | |
| 1739.03 | 870.52 | 2 | 3244.43 | 1623.22 | 2 | 4490.02 | 1123.51 | 4 | | | |
| 1764.81 | 883.40 | 2 | 3293.43 | 1647.71 | 2 | 4499.05 | 1500.68 | 3 | | | |
| 1764.85 | 883.42 | 2 | 3314.62 | 553.44 | 6 | 4502.09 | 1126.52 | 4 | | | |
| 1764.87 | 589.29 | 3 | 3315.48 | 1658.74 | 2 | 4507.09 | 1127.77 | 4 | | | |
| 1764.87 | 589.29 | 3 | 3381.55 | 1128.18 | 3 | 4522.12 | 1131.53 | 4 | | | |
| 1764.88 | 589.29 | 3 | 3399.54 | 1700.77 | 2 | 4567.10 | 1142.77 | 4 | | | |
| 1789.83 | 895.92 | 2 | 3505.73 | 877.43 | 4 | 4641.10 | 1161.28 | 4 | | | |
| 1789.85 | 597.62 | 3 | 3543.60 | 1772.80 | 2 | 4646.10 | 1549.70 | 3 | | | |
| 1798.85 | 900.43 | 2 | 3591.48 | 1796.74 | 2 | 4646.11 | 1549.70 | 3 | | | |
| 1810.88 | 906.44 | 2 | 3593.65 | 1198.88 | 3 | 4673.09 | 1558.70 | 3 | | | |
| 1811.08 | 604.69 | 3 | 3661.65 | 1831.83 | 2 | 4677.13 | 1560.04 | 3 | | | |
| 1866.91 | 467.73 | 4 | 3661.67 | 1831.83 | 2 | 4682.09 | 1561.70 | 3 | | | |
| 1867.10 | 623.37 | 3 | 3699.68 | 1850.84 | 2 | 4704.15 | 1569.05 | 3 | | | |

A.5. Ribonuclease S Cross-linked Candidates

Table A.13: Candidate Cross-linked Species for BS³

| [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z |
|---------|--------|---|---------|---------|---|---------|---------|---|---------|---------|----|
| 604.34 | 303.17 | 2 | 1674.80 | 559.27 | 3 | 2116.14 | 706.38 | 3 | 2771.32 | 308.92 | 9 |
| 752.47 | 377.24 | 2 | 1674.81 | 559.27 | 3 | 2123.31 | 531.83 | 4 | 2775.32 | 926.11 | 3 |
| 752.47 | 377.24 | 2 | 1678.80 | 840.40 | 2 | 2134.12 | 1068.06 | 2 | 2784.37 | 1393.19 | 2 |
| 752.52 | 377.26 | 2 | 1683.70 | 562.23 | 3 | 2146.10 | 716.37 | 3 | 2784.39 | 929.13 | 3 |
| 752.53 | 377.26 | 2 | 1683.82 | 562.27 | 3 | 2155.05 | 539.76 | 4 | 2784.39 | 697.10 | 4 |
| 835.42 | 418.71 | 2 | 1684.86 | 422.21 | 4 | 2156.05 | 540.01 | 4 | 2784.40 | 697.10 | 4 |
| 874.51 | 438.25 | 2 | 1692.81 | 565.27 | 3 | 2156.32 | 1079.16 | 2 | 2789.24 | 1395.62 | 2 |
| 932.45 | 467.23 | 2 | 1692.82 | 424.21 | 4 | 2175.95 | 1088.98 | 2 | 2794.32 | 559.86 | 5 |
| 960.48 | 481.24 | 2 | 1702.80 | 852.40 | 2 | 2178.06 | 545.51 | 4 | 2799.39 | 934.13 | 3 |
| 960.49 | 481.25 | 2 | 1708.81 | 570.60 | 3 | 2189.97 | 1095.98 | 2 | 2816.25 | 1409.12 | 2 |
| 960.55 | 321.18 | 3 | 1713.03 | 572.01 | 3 | 2189.99 | 731.00 | 3 | 2816.26 | 1409.13 | 2 |
| 989.57 | 330.86 | 3 | 1716.08 | 859.04 | 2 | 2190.32 | 1096.16 | 2 | 2820.26 | 1411.13 | 2 |
| 989.62 | 495.81 | 2 | 1718.85 | 860.43 | 2 | 2197.16 | 550.29 | 4 | 2822.33 | 565.47 | 5 |
| 1017.53 | 509.76 | 2 | 1759.97 | 587.66 | 3 | 2206.98 | 1104.49 | 2 | 2834.24 | 1418.12 | 2 |
| 1019.48 | 510.74 | 2 | 1770.89 | 591.30 | 3 | 2222.15 | 1112.08 | 2 | 2840.40 | 711.10 | 4 |
| 1047.69 | 524.84 | 2 | 1775.82 | 888.91 | 2 | 2251.14 | 751.38 | 3 | 2845.26 | 1423.63 | 2 |
| 1065.53 | 533.76 | 2 | 1775.84 | 888.92 | 2 | 2271.06 | 758.02 | 3 | 2845.28 | 949.43 | 3 |
| 1088.63 | 545.31 | 2 | 1807.86 | 904.93 | 2 | 2281.16 | 761.39 | 3 | 2856.24 | 1429.12 | 2 |
| 1088.63 | 545.32 | 2 | 1807.88 | 603.63 | 3 | 2290.20 | 573.55 | 4 | 2876.04 | 576.21 | 5 |
| 1101.59 | 551.80 | 2 | 1807.88 | 362.58 | 5 | 2348.19 | 588.05 | 4 | 2889.26 | 964.09 | 3 |
| 1116.77 | 559.39 | 2 | 1815.85 | 908.93 | 2 | 2350.19 | 471.04 | 5 | 2889.27 | 1445.64 | 2 |
| 1141.67 | 571.83 | 2 | 1815.99 | 606.33 | 3 | 2362.22 | 591.56 | 4 | 2897.32 | 966.77 | 3 |
| 1141.73 | 381.58 | 3 | 1816.86 | 909.43 | 2 | 2375.14 | 1188.57 | 2 | 2900.30 | 1451.15 | 2 |
| 1141.73 | 381.58 | 3 | 1830.86 | 611.29 | 3 | 2410.17 | 804.39 | 3 | 2907.38 | 1454.69 | 2 |
| 1204.59 | 603.29 | 2 | 1836.88 | 919.44 | 2 | 2437.22 | 1219.61 | 2 | 2913.29 | 1457.64 | 2 |
| 1222.64 | 612.32 | 2 | 1836.97 | 613.32 | 3 | 2451.14 | 351.16 | 7 | 2914.29 | 1458.15 | 2 |
| 1256.61 | 629.31 | 2 | 1838.90 | 920.45 | 2 | 2451.18 | 613.79 | 4 | 2942.47 | 981.82 | 3 |
| 1272.65 | 425.22 | 3 | 1869.90 | 624.30 | 3 | 2488.22 | 498.64 | 5 | 2961.38 | 741.35 | 4 |
| 1272.77 | 425.26 | 3 | 1870.88 | 468.72 | 4 | 2531.36 | 844.79 | 3 | 2963.34 | 1482.67 | 2 |
| 1279.62 | 640.81 | 2 | 1870.89 | 624.63 | 3 | 2543.11 | 848.70 | 3 | 2973.36 | 1487.68 | 2 |
| 1284.63 | 643.31 | 2 | 1870.90 | 936.45 | 2 | 2575.11 | 859.37 | 3 | 2984.35 | 1493.18 | 2 |
| 1300.66 | 651.33 | 2 | 1906.94 | 954.47 | 2 | 2585.13 | 862.71 | 3 | 2998.34 | 1500.17 | 2 |
| 1316.73 | 330.18 | 4 | 1906.95 | 636.65 | 3 | 2585.14 | 647.29 | 4 | 3032.36 | 1517.18 | 2 |
| 1316.74 | 330.18 | 4 | 1915.05 | 639.35 | 3 | 2592.12 | 865.04 | 3 | 3046.53 | 1016.51 | 3 |
| 1329.73 | 665.87 | 2 | 1922.94 | 962.47 | 2 | 2605.15 | 869.38 | 3 | 3066.42 | 1023.14 | 3 |
| 1332.72 | 445.24 | 3 | 1925.94 | 963.97 | 2 | 2612.30 | 523.46 | 5 | 3085.68 | 343.85 | 9 |
| 1332.74 | 334.19 | 4 | 1943.89 | 972.95 | 2 | 2618.29 | 1310.14 | 2 | 3164.66 | 1055.89 | 3 |
| 1342.69 | 336.67 | 4 | 1948.95 | 975.47 | 2 | 2618.29 | 1310.14 | 2 | 3175.54 | 1588.77 | 2 |
| 1342.74 | 448.58 | 3 | 1951.71 | 976.86 | 2 | 2643.14 | 1322.57 | 2 | 3175.63 | 794.91 | 4 |
| 1358.72 | 453.91 | 3 | 1961.87 | 981.94 | 2 | 2643.16 | 882.05 | 3 | 3194.47 | 1598.24 | 2 |
| 1451.71 | 726.86 | 2 | 1981.97 | 991.98 | 2 | 2648.26 | 1325.13 | 2 | 3207.56 | 535.59 | 6 |
| 1454.72 | 728.36 | 2 | 1981.98 | 661.66 | 3 | 2661.16 | 1331.58 | 2 | 3225.45 | 1076.15 | 3 |
| 1460.70 | 487.90 | 3 | 1997.98 | 500.50 | 4 | 2661.18 | 888.06 | 3 | 3225.66 | 1076.22 | 3 |
| 1473.73 | 737.87 | 2 | 1997.99 | 667.00 | 3 | 2675.16 | 1338.58 | 2 | 3241.67 | 811.42 | 4 |
| 1479.75 | 370.94 | 4 | 1999.98 | 401.00 | 5 | 2675.30 | 536.06 | 5 | 3339.76 | 835.94 | 4 |
| 1479.93 | 494.31 | 3 | 1999.98 | 1000.99 | 2 | 2690.27 | 1346.13 | 2 | 3339.76 | 835.94 | 4 |
| 1488.83 | 497.28 | 3 | 1999.98 | 1000.99 | 2 | 2690.33 | 539.07 | 5 | 3339.77 | 668.95 | 5 |
| 1502.73 | 752.37 | 2 | 1999.99 | 667.66 | 3 | 2694.17 | 1348.08 | 2 | 3341.58 | 1114.86 | 3 |
| 1513.73 | 757.87 | 2 | 2001.71 | 1001.86 | 2 | 2694.28 | 1348.14 | 2 | 3349.88 | 335.99 | 10 |
| 1513.94 | 505.65 | 3 | 2004.94 | 1003.47 | 2 | 2718.19 | 1360.10 | 2 | 3353.77 | 1677.88 | 2 |
| 1545.75 | 773.87 | 2 | 2014.02 | 1008.01 | 2 | 2718.19 | 1360.10 | 2 | 3353.77 | 839.44 | 4 |
| 1582.82 | 317.56 | 5 | 2014.03 | 672.34 | 3 | 2718.20 | 907.07 | 3 | 3353.78 | 671.76 | 5 |
| 1582.83 | 317.57 | 5 | 2014.03 | 1008.01 | 2 | 2718.20 | 907.07 | 3 | 3353.79 | 559.96 | 6 |
| 1605.98 | 536.33 | 3 | 2017.92 | 1009.96 | 2 | 2718.22 | 1360.11 | 2 | 3353.79 | 839.45 | 4 |
| 1607.79 | 536.93 | 3 | 2033.05 | 509.26 | 4 | 2718.31 | 544.66 | 5 | 3366.55 | 1123.18 | 3 |
| 1607.80 | 804.90 | 2 | 2041.00 | 1021.50 | 2 | 2724.24 | 1363.12 | 2 | 3393.80 | 849.45 | 4 |
| 1633.77 | 817.88 | 2 | 2084.01 | 695.67 | 3 | 2724.26 | 1363.13 | 2 | 3399.40 | 1700.70 | 2 |
| 1633.77 | 817.88 | 2 | 2088.94 | 1045.47 | 2 | 2729.32 | 683.33 | 4 | 3402.54 | 1702.27 | 2 |
| 1634.78 | 545.93 | 3 | 2092.05 | 419.41 | 5 | 2729.47 | 455.91 | 6 | 3427.64 | 1143.55 | 3 |
| 1643.81 | 548.94 | 3 | 2099.12 | 1050.56 | 2 | 2730.41 | 547.08 | 5 | 3442.55 | 1722.27 | 2 |
| 1647.95 | 824.97 | 2 | 2100.12 | 1051.06 | 2 | 2734.31 | 547.86 | 5 | 3444.67 | 862.17 | 4 |
| 1651.79 | 551.60 | 3 | 2100.12 | 701.04 | 3 | 2762.22 | 921.74 | 3 | 3471.53 | 868.88 | 4 |
| 1651.79 | 826.89 | 2 | 2115.91 | 1058.96 | 2 | 2762.24 | 1382.12 | 2 | 3476.72 | 580.45 | 6 |
| 1651.79 | 551.60 | 3 | 2116.14 | 706.38 | 3 | 2764.33 | 553.87 | 5 | 3486.52 | 499.07 | 7 |
| 1666.82 | 556.61 | 3 | 2116.14 | 1059.07 | 2 | 2768.37 | 1385.18 | 2 | 3520.44 | 321.04 | 11 |
| 1674.79 | 838.40 | 2 | 2116.14 | 1059.07 | 2 | 2768.38 | 923.79 | 3 | 3525.75 | 588.62 | 6 |

A.5. Ribonuclease S Cross-linked Candidates

Table A.14: Candidate Cross-linked Species for BS³

| [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z |
|---------|---------|----|---------|---------|----|---------|---------|----|----------|---------|----|
| 3560.58 | 1781.29 | 2 | 4144.87 | 1382.62 | 3 | 4968.27 | 1657.09 | 3 | 8422.78 | 1404.80 | 6 |
| 3581.53 | 896.38 | 4 | 4152.05 | 347.00 | 12 | 4970.42 | 829.40 | 6 | 8512.47 | 1703.49 | 5 |
| 3616.65 | 1809.32 | 2 | 4165.90 | 1042.47 | 4 | 4970.96 | 1243.74 | 4 | 8523.57 | 1705.71 | 5 |
| 3646.84 | 912.71 | 4 | 4166.87 | 1389.96 | 3 | 5001.26 | 1668.09 | 3 | 8730.22 | 1092.28 | 8 |
| 3676.67 | 1839.33 | 2 | 4171.88 | 1043.97 | 4 | 5006.28 | 1669.76 | 3 | 8981.10 | 1497.85 | 6 |
| 3679.67 | 1840.83 | 2 | 4174.89 | 1392.63 | 3 | 5019.27 | 1674.09 | 3 | 9014.13 | 1503.36 | 6 |
| 3679.67 | 1840.83 | 2 | 4185.19 | 1396.06 | 3 | 5026.28 | 1676.43 | 3 | 9376.57 | 1876.31 | 5 |
| 3679.68 | 1840.84 | 2 | 4196.92 | 1050.23 | 4 | 5031.21 | 1678.07 | 3 | 9433.60 | 944.36 | 10 |
| 3686.82 | 615.47 | 6 | 4196.93 | 1050.23 | 4 | 5062.51 | 507.25 | 10 | 10176.75 | 1697.12 | 6 |
| 3697.67 | 1849.84 | 2 | 4196.96 | 700.49 | 6 | 5073.29 | 1692.10 | 3 | 10799.68 | 1350.96 | 8 |
| 3697.70 | 1849.85 | 2 | 4215.91 | 1054.98 | 4 | 5083.24 | 1695.41 | 3 | | | |
| 3710.68 | 1856.34 | 2 | 4227.88 | 1057.97 | 4 | 5108.30 | 1703.77 | 3 | | | |
| 3710.68 | 1856.34 | 2 | 4227.90 | 1410.30 | 3 | 5109.42 | 568.71 | 9 | | | |
| 3710.69 | 1856.34 | 2 | 4232.96 | 706.49 | 6 | 5122.26 | 1708.42 | 3 | | | |
| 3710.69 | 1237.90 | 3 | 4234.89 | 1412.63 | 3 | 5123.32 | 1708.77 | 3 | | | |
| 3724.68 | 1863.34 | 2 | 4257.83 | 1420.28 | 3 | 5127.36 | 1710.12 | 3 | | | |
| 3735.86 | 934.97 | 4 | 4258.87 | 1420.62 | 3 | 5134.43 | 1712.48 | 3 | | | |
| 3739.67 | 1870.84 | 2 | 4288.93 | 1430.64 | 3 | 5142.36 | 1715.12 | 3 | | | |
| 3763.69 | 1882.85 | 2 | 4289.62 | 1430.87 | 3 | 5180.23 | 1296.06 | 4 | | | |
| 3764.68 | 1883.34 | 2 | 4356.23 | 397.02 | 11 | 5314.78 | 1063.96 | 5 | | | |
| 3767.68 | 1884.84 | 2 | 4358.97 | 1090.74 | 4 | 5316.29 | 1330.07 | 4 | | | |
| 3767.68 | 1884.84 | 2 | 4411.24 | 552.41 | 8 | 5408.43 | 902.40 | 6 | | | |
| 3767.69 | 1256.90 | 3 | 4413.05 | 1472.02 | 3 | 5504.68 | 918.45 | 6 | | | |
| 3767.72 | 942.93 | 4 | 4448.09 | 1483.70 | 3 | 5546.87 | 925.48 | 6 | | | |
| 3768.69 | 1885.34 | 2 | 4468.34 | 1490.45 | 3 | 5600.91 | 934.49 | 6 | | | |
| 3768.70 | 1257.23 | 3 | 4469.02 | 1118.25 | 4 | 5600.93 | 801.13 | 7 | | | |
| 3783.68 | 1892.84 | 2 | 4473.05 | 1119.26 | 4 | 5614.94 | 702.87 | 8 | | | |
| 3783.70 | 1262.23 | 3 | 4486.05 | 1496.35 | 3 | 5647.77 | 942.30 | 6 | | | |
| 3799.68 | 1900.84 | 2 | 4500.09 | 1126.02 | 4 | 5654.78 | 943.46 | 6 | | | |
| 3811.67 | 1906.84 | 2 | 4503.07 | 1502.02 | 3 | 5713.83 | 953.31 | 6 | | | |
| 3821.66 | 1274.89 | 3 | 4515.05 | 1506.02 | 3 | 5724.60 | 573.46 | 10 | | | |
| 3826.70 | 1914.35 | 2 | 4515.09 | 1129.77 | 4 | 5726.54 | 573.65 | 10 | | | |
| 3828.81 | 349.07 | 11 | 4515.20 | 452.52 | 10 | 5840.65 | 974.44 | 6 | | | |
| 3831.15 | 639.52 | 6 | 4518.33 | 904.67 | 5 | 5949.13 | 992.52 | 6 | | | |
| 3843.68 | 1922.84 | 2 | 4528.05 | 1510.35 | 3 | 5965.74 | 995.29 | 6 | | | |
| 3851.58 | 351.14 | 11 | 4529.73 | 1510.91 | 3 | 5978.61 | 997.44 | 6 | | | |
| 3857.43 | 1929.71 | 2 | 4533.06 | 1512.02 | 3 | 6024.17 | 1005.03 | 6 | | | |
| 3873.72 | 1292.24 | 3 | 4538.06 | 1513.69 | 3 | 6122.10 | 1021.35 | 6 | | | |
| 3914.72 | 1958.36 | 2 | 4543.07 | 1515.36 | 3 | 6251.03 | 1563.76 | 4 | | | |
| 3914.73 | 1305.91 | 3 | 4548.07 | 1517.02 | 3 | 6271.83 | 1568.96 | 4 | | | |
| 3914.73 | 1305.91 | 3 | 4551.09 | 1138.77 | 4 | 6345.79 | 1587.45 | 4 | | | |
| 3915.77 | 1306.26 | 3 | 4556.12 | 1140.03 | 4 | 6467.22 | 924.89 | 7 | | | |
| 3933.68 | 1312.23 | 3 | 4561.09 | 1521.36 | 3 | 7063.01 | 1178.17 | 6 | | | |
| 3972.77 | 1325.26 | 3 | 4561.13 | 1141.28 | 4 | 7180.55 | 599.38 | 12 | | | |
| 3972.79 | 994.20 | 4 | 4569.47 | 572.18 | 8 | 7191.55 | 1199.59 | 6 | | | |
| 3975.45 | 1326.15 | 3 | 4591.09 | 1531.36 | 3 | 7352.33 | 1839.08 | 4 | | | |
| 3983.74 | 996.93 | 4 | 4608.06 | 1537.02 | 3 | 7385.31 | 1847.33 | 4 | | | |
| 3996.91 | 400.69 | 10 | 4626.08 | 1543.03 | 3 | 7407.29 | 1852.82 | 4 | | | |
| 4009.87 | 1003.47 | 4 | 4642.08 | 1548.36 | 3 | 7407.37 | 1852.84 | 4 | | | |
| 4026.86 | 1343.29 | 3 | 4674.28 | 1169.57 | 4 | 7413.31 | 1854.33 | 4 | | | |
| 4032.80 | 1345.27 | 3 | 4694.34 | 1174.58 | 4 | 7457.35 | 622.45 | 12 | | | |
| 4032.88 | 1009.22 | 4 | 4694.34 | 939.87 | 5 | 7519.30 | 1880.82 | 4 | | | |
| 4037.85 | 1346.95 | 3 | 4715.11 | 1572.70 | 3 | 7562.35 | 946.29 | 8 | | | |
| 4052.82 | 1351.94 | 3 | 4736.17 | 1579.72 | 3 | 7617.85 | 1524.57 | 5 | | | |
| 4083.85 | 1362.28 | 3 | 4750.17 | 1584.39 | 3 | 7631.37 | 848.93 | 9 | | | |
| 4088.86 | 1363.95 | 3 | 4785.17 | 1596.06 | 3 | 7653.38 | 1914.34 | 4 | | | |
| 4101.86 | 1368.29 | 3 | 4786.17 | 1596.39 | 3 | 7747.86 | 969.48 | 8 | | | |
| 4110.87 | 1371.29 | 3 | 4795.38 | 480.54 | 10 | 7802.38 | 1301.40 | 6 | | | |
| 4120.86 | 516.11 | 8 | 4814.31 | 963.86 | 5 | 7812.29 | 1563.46 | 5 | | | |
| 4125.87 | 1376.29 | 3 | 4834.19 | 1612.40 | 3 | 7833.00 | 871.33 | 9 | | | |
| 4125.88 | 1376.29 | 3 | 4840.20 | 1614.40 | 3 | 7905.60 | 1582.12 | 5 | | | |
| 4139.87 | 1380.96 | 3 | 4871.93 | 975.39 | 5 | 7925.62 | 1586.12 | 5 | | | |
| 4141.88 | 1381.63 | 3 | 4913.23 | 1638.74 | 3 | 8194.74 | 1025.34 | 8 | | | |
| 4141.88 | 1036.47 | 4 | 4927.22 | 1643.41 | 3 | 8220.75 | 1371.13 | 6 | | | |
| 4143.87 | 1382.29 | 3 | 4929.22 | 1644.07 | 3 | 8276.71 | 1380.45 | 6 | | | |
| 4143.88 | 829.78 | 5 | 4937.26 | 988.45 | 5 | 8382.75 | 1398.12 | 6 | | | |

A.5. Ribonuclease S Cross-linked Candidates

Table A.15: Candidate Cross-linked Species for sulfoEGS

| [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z |
|---------|--------|---|---------|---------|---|---------|---------|---|---------|---------|---|
| 866.32 | 434.16 | 2 | 1782.08 | 595.03 | 3 | 2226.05 | 1114.02 | 2 | 2728.33 | 683.08 | 4 |
| 895.42 | 448.71 | 2 | 1791.84 | 896.92 | 2 | 2244.93 | 1123.47 | 2 | 2728.33 | 683.08 | 4 |
| 977.45 | 489.72 | 2 | 1794.92 | 449.73 | 4 | 2259.07 | 1130.54 | 2 | 2728.40 | 546.68 | 5 |
| 977.45 | 489.72 | 2 | 1803.67 | 902.83 | 2 | 2294.12 | 1148.06 | 2 | 2745.33 | 916.11 | 3 |
| 992.51 | 497.25 | 2 | 1806.88 | 452.72 | 4 | 2294.15 | 1148.07 | 2 | 2745.38 | 550.08 | 5 |
| 1010.55 | 337.85 | 3 | 1806.89 | 362.38 | 5 | 2294.16 | 765.72 | 3 | 2762.32 | 921.77 | 3 |
| 1154.57 | 578.28 | 2 | 1851.88 | 926.94 | 2 | 2301.11 | 1151.56 | 2 | 2764.34 | 553.87 | 5 |
| 1171.54 | 586.77 | 2 | 1851.89 | 926.94 | 2 | 2315.11 | 1158.55 | 2 | 2768.38 | 693.10 | 4 |
| 1174.58 | 588.29 | 2 | 1864.88 | 933.44 | 2 | 2322.11 | 1162.05 | 2 | 2774.35 | 925.78 | 3 |
| 1177.60 | 393.53 | 3 | 1866.94 | 623.31 | 3 | 2324.08 | 1163.04 | 2 | 2779.34 | 556.87 | 5 |
| 1177.70 | 589.85 | 2 | 1869.90 | 624.30 | 3 | 2332.08 | 1167.04 | 2 | 2788.26 | 1395.13 | 2 |
| 1225.60 | 613.80 | 2 | 1870.88 | 468.72 | 4 | 2332.21 | 467.44 | 5 | 2791.29 | 1396.64 | 2 |
| 1225.65 | 409.55 | 3 | 1870.88 | 624.63 | 3 | 2335.20 | 468.04 | 5 | 2791.34 | 698.84 | 4 |
| 1237.65 | 413.55 | 3 | 1870.89 | 936.45 | 2 | 2336.16 | 585.04 | 4 | 2794.33 | 1398.16 | 2 |
| 1241.68 | 414.89 | 3 | 1875.97 | 469.99 | 4 | 2341.05 | 1171.52 | 2 | 2798.31 | 1400.16 | 2 |
| 1241.69 | 414.90 | 3 | 1892.90 | 947.45 | 2 | 2343.05 | 1172.52 | 2 | 2802.36 | 561.47 | 5 |
| 1281.63 | 641.82 | 2 | 1913.91 | 638.97 | 3 | 2356.11 | 1179.05 | 2 | 2810.32 | 1406.16 | 2 |
| 1281.82 | 428.27 | 3 | 1936.88 | 646.63 | 3 | 2357.11 | 1179.56 | 2 | 2810.33 | 1406.16 | 2 |
| 1288.63 | 645.32 | 2 | 1936.99 | 388.40 | 5 | 2359.06 | 1180.53 | 2 | 2811.35 | 938.12 | 3 |
| 1309.59 | 655.79 | 2 | 1954.94 | 978.47 | 2 | 2363.24 | 394.87 | 6 | 2820.35 | 565.07 | 5 |
| 1309.78 | 655.89 | 2 | 1967.93 | 984.97 | 2 | 2378.72 | 1190.36 | 2 | 2822.29 | 1412.14 | 2 |
| 1312.64 | 657.32 | 2 | 1969.93 | 985.96 | 2 | 2388.13 | 1195.06 | 2 | 2824.27 | 1413.14 | 2 |
| 1312.65 | 657.32 | 2 | 1980.92 | 991.46 | 2 | 2410.25 | 402.71 | 6 | 2824.34 | 942.45 | 3 |
| 1323.69 | 442.23 | 3 | 1985.94 | 993.97 | 2 | 2413.17 | 805.39 | 3 | 2828.33 | 1415.16 | 2 |
| 1329.74 | 665.87 | 2 | 1986.02 | 497.50 | 4 | 2413.20 | 604.30 | 4 | 2828.34 | 943.78 | 3 |
| 1330.60 | 666.30 | 2 | 1988.96 | 663.99 | 3 | 2429.20 | 810.73 | 3 | 2828.34 | 1415.17 | 2 |
| 1333.73 | 445.58 | 3 | 1997.94 | 999.97 | 2 | 2435.14 | 1218.57 | 2 | 2828.35 | 943.78 | 3 |
| 1383.70 | 692.85 | 2 | 1997.94 | 999.97 | 2 | 2444.19 | 1223.09 | 2 | 2831.23 | 1416.62 | 2 |
| 1399.75 | 467.58 | 3 | 1997.95 | 666.98 | 3 | 2444.19 | 815.73 | 3 | 2845.33 | 1423.66 | 2 |
| 1436.69 | 719.34 | 2 | 1997.96 | 666.99 | 3 | 2444.21 | 612.05 | 4 | 2845.35 | 712.34 | 4 |
| 1442.71 | 722.36 | 2 | 1999.75 | 1000.87 | 2 | 2444.21 | 612.05 | 4 | 2846.31 | 1424.15 | 2 |
| 1445.71 | 723.85 | 2 | 1999.94 | 1000.97 | 2 | 2475.08 | 496.02 | 5 | 2855.30 | 1428.65 | 2 |
| 1449.61 | 725.81 | 2 | 2003.91 | 668.97 | 3 | 2477.11 | 1239.55 | 2 | 2864.27 | 1433.13 | 2 |
| 1449.61 | 725.81 | 2 | 2010.88 | 1006.44 | 2 | 2482.10 | 621.52 | 4 | 2864.34 | 717.08 | 4 |
| 1449.72 | 725.86 | 2 | 2013.03 | 672.01 | 3 | 2485.23 | 622.31 | 4 | 2877.26 | 1439.63 | 2 |
| 1467.62 | 734.81 | 2 | 2035.00 | 408.00 | 5 | 2485.24 | 498.05 | 5 | 2880.28 | 1441.14 | 2 |
| 1467.63 | 734.81 | 2 | 2060.94 | 516.24 | 4 | 2486.08 | 1244.04 | 2 | 2880.32 | 961.11 | 3 |
| 1523.68 | 762.84 | 2 | 2061.04 | 413.21 | 5 | 2501.11 | 1251.55 | 2 | 2886.34 | 1444.17 | 2 |
| 1523.70 | 762.85 | 2 | 2065.00 | 517.25 | 4 | 2501.21 | 834.74 | 3 | 2886.37 | 963.12 | 3 |
| 1523.74 | 762.87 | 2 | 2073.87 | 1037.94 | 2 | 2501.21 | 626.30 | 4 | 2890.37 | 1446.18 | 2 |
| 1541.75 | 771.87 | 2 | 2073.89 | 692.30 | 3 | 2501.22 | 626.30 | 4 | 2890.40 | 723.60 | 4 |
| 1550.73 | 776.36 | 2 | 2080.98 | 521.24 | 4 | 2501.23 | 501.25 | 5 | 2897.51 | 322.95 | 9 |
| 1553.75 | 777.87 | 2 | 2080.98 | 521.24 | 4 | 2503.07 | 1252.54 | 2 | 2901.35 | 968.12 | 3 |
| 1559.71 | 780.85 | 2 | 2080.99 | 694.66 | 3 | 2503.09 | 1252.54 | 2 | 2921.40 | 585.28 | 5 |
| 1571.70 | 524.90 | 3 | 2084.01 | 695.67 | 3 | 2510.25 | 1256.13 | 2 | 2922.33 | 975.11 | 3 |
| 1571.72 | 786.86 | 2 | 2087.96 | 1044.98 | 2 | 2529.24 | 633.31 | 4 | 2932.33 | 1467.16 | 2 |
| 1608.76 | 537.25 | 3 | 2087.97 | 1044.99 | 2 | 2529.25 | 506.85 | 5 | 2944.45 | 737.11 | 4 |
| 1624.86 | 542.62 | 3 | 2087.99 | 697.00 | 3 | 2536.09 | 1269.05 | 2 | 2949.40 | 984.13 | 3 |
| 1634.75 | 545.92 | 3 | 2098.97 | 1050.49 | 2 | 2543.11 | 848.70 | 3 | 2956.30 | 1479.15 | 2 |
| 1638.78 | 820.39 | 2 | 2098.99 | 525.75 | 4 | 2549.09 | 1275.54 | 2 | 2967.32 | 1484.66 | 2 |
| 1638.78 | 820.39 | 2 | 2099.00 | 700.67 | 3 | 2554.17 | 639.54 | 4 | 2972.34 | 1487.17 | 2 |
| 1639.75 | 547.58 | 3 | 2112.72 | 353.12 | 6 | 2582.12 | 1292.06 | 2 | 2972.37 | 1487.19 | 2 |
| 1639.75 | 547.58 | 3 | 2122.01 | 425.40 | 5 | 2594.13 | 1298.06 | 2 | 2974.40 | 992.47 | 3 |
| 1653.81 | 552.27 | 3 | 2127.87 | 304.98 | 7 | 2600.09 | 1301.05 | 2 | 2979.38 | 994.13 | 3 |
| 1662.80 | 832.40 | 2 | 2155.95 | 1078.97 | 2 | 2620.12 | 1311.06 | 2 | 3006.51 | 335.06 | 9 |
| 1662.98 | 832.49 | 2 | 2156.03 | 540.01 | 4 | 2620.13 | 874.38 | 3 | 3011.34 | 1506.67 | 2 |
| 1674.81 | 559.27 | 3 | 2178.06 | 545.52 | 4 | 2620.18 | 525.04 | 5 | 3017.41 | 1006.80 | 3 |
| 1684.82 | 562.61 | 3 | 2186.01 | 1094.00 | 2 | 2643.12 | 1322.56 | 2 | 3093.43 | 1032.14 | 3 |
| 1685.79 | 843.89 | 2 | 2186.08 | 1094.04 | 2 | 2657.31 | 665.33 | 4 | 3138.38 | 1570.19 | 2 |
| 1710.04 | 571.01 | 3 | 2189.00 | 1095.50 | 2 | 2672.28 | 1337.14 | 2 | 3150.45 | 351.05 | 9 |
| 1718.76 | 573.92 | 3 | 2204.93 | 1103.46 | 2 | 2682.19 | 1342.10 | 2 | 3151.45 | 1576.72 | 2 |
| 1759.64 | 880.82 | 2 | 2205.12 | 1103.56 | 2 | 2710.30 | 1356.15 | 2 | 3151.50 | 788.88 | 4 |
| 1768.82 | 885.41 | 2 | 2206.99 | 736.66 | 3 | 2718.17 | 1360.09 | 2 | 3154.49 | 1578.25 | 2 |
| 1773.86 | 592.29 | 3 | 2207.93 | 368.99 | 6 | 2718.19 | 1360.10 | 2 | 3225.49 | 1076.16 | 3 |
| 1776.93 | 889.47 | 2 | 2225.04 | 1113.52 | 2 | 2727.29 | 1364.65 | 2 | 3227.46 | 1614.73 | 2 |
| 1781.82 | 594.94 | 3 | 2226.02 | 743.01 | 3 | 2728.30 | 1365.15 | 2 | 3279.47 | 1640.74 | 2 |

A.5. Ribonuclease S Cross-linked Candidates

Table A.16: Candidate Cross-linked Species for sulfoEGS

| [M]exp | m/z | z | [M]exp | m/z | z | [M]exp | m/z | z |
|---------|---------|----|---------|---------|----|---------|---------|----|
| 3309.52 | 1655.76 | 2 | 3855.46 | 964.87 | 4 | 4513.13 | 1129.28 | 4 |
| 3353.78 | 839.44 | 4 | 3856.42 | 1929.21 | 2 | 4519.34 | 904.87 | 5 |
| 3377.30 | 1689.65 | 2 | 3857.45 | 1286.82 | 3 | 4521.13 | 1131.28 | 4 |
| 3378.30 | 1690.15 | 2 | 3858.74 | 1287.25 | 3 | 4526.05 | 1509.68 | 3 |
| 3384.55 | 1693.28 | 2 | 3867.72 | 1290.24 | 3 | 4533.07 | 1512.02 | 3 |
| 3391.52 | 1696.76 | 2 | 3869.45 | 1290.82 | 3 | 4536.16 | 1135.04 | 4 |
| 3397.58 | 1699.79 | 2 | 3872.44 | 1937.22 | 2 | 4551.08 | 1518.03 | 3 |
| 3448.68 | 1150.56 | 3 | 3873.44 | 1292.15 | 3 | 4555.09 | 1519.36 | 3 |
| 3453.58 | 314.96 | 11 | 3874.43 | 1292.48 | 3 | 4557.15 | 1140.29 | 4 |
| 3466.60 | 1734.30 | 2 | 3884.45 | 1295.82 | 3 | 4560.79 | 571.10 | 8 |
| 3472.68 | 1158.56 | 3 | 3885.72 | 1296.24 | 3 | 4634.16 | 1545.72 | 3 |
| 3482.60 | 1742.30 | 2 | 3899.70 | 1300.90 | 3 | 4649.87 | 1550.96 | 3 |
| 3486.54 | 499.08 | 7 | 3899.74 | 1300.91 | 3 | 4675.88 | 1559.63 | 3 |
| 3517.60 | 1759.80 | 2 | 3900.43 | 1951.22 | 2 | 4720.17 | 1574.39 | 3 |
| 3546.67 | 1774.33 | 2 | 3900.43 | 1301.14 | 3 | 4735.20 | 1579.40 | 3 |
| 3546.69 | 1183.23 | 3 | 3900.44 | 1301.15 | 3 | 4754.23 | 1585.74 | 3 |
| 3561.69 | 1188.23 | 3 | 3902.41 | 1952.21 | 2 | 4760.19 | 1587.73 | 3 |
| 3570.66 | 1786.33 | 2 | 3903.41 | 1952.70 | 2 | 4762.13 | 1588.38 | 3 |
| 3584.67 | 1793.34 | 2 | 3903.43 | 1952.72 | 2 | 4764.97 | 1192.24 | 4 |
| 3615.73 | 452.97 | 8 | 3903.44 | 1302.15 | 3 | 4765.18 | 1589.39 | 3 |
| 3632.72 | 1211.91 | 3 | 3903.45 | 1302.15 | 3 | 4785.18 | 1596.06 | 3 |
| 3642.64 | 1215.21 | 3 | 3903.46 | 976.87 | 4 | 4785.20 | 1596.07 | 3 |
| 3656.83 | 610.47 | 6 | 3904.94 | 781.99 | 5 | 4789.18 | 1597.39 | 3 |
| 3657.82 | 610.64 | 6 | 3914.45 | 1958.22 | 2 | 4797.29 | 1200.32 | 4 |
| 3659.65 | 1830.83 | 2 | 3914.75 | 1305.92 | 3 | 4803.30 | 687.19 | 7 |
| 3661.67 | 1831.84 | 2 | 3914.76 | 1958.38 | 2 | 4807.19 | 1603.40 | 3 |
| 3661.68 | 1221.56 | 3 | 3927.46 | 1310.15 | 3 | 4815.20 | 1606.07 | 3 |
| 3672.75 | 1837.38 | 2 | 3933.44 | 1312.15 | 3 | 4862.14 | 1621.71 | 3 |
| 3672.77 | 1225.26 | 3 | 3942.45 | 1315.15 | 3 | 4870.29 | 1218.57 | 4 |
| 3685.67 | 1229.56 | 3 | 3965.42 | 1983.71 | 2 | 4897.26 | 1225.32 | 4 |
| 3698.68 | 1233.89 | 3 | 3965.45 | 1322.82 | 3 | 4910.23 | 1637.74 | 3 |
| 3702.69 | 1235.23 | 3 | 4008.89 | 1003.22 | 4 | 4924.22 | 1642.41 | 3 |
| 3709.70 | 1237.57 | 3 | 4009.88 | 1337.63 | 3 | 4937.24 | 1235.31 | 4 |
| 3709.71 | 1855.86 | 2 | 4026.89 | 1343.30 | 3 | 4943.89 | 412.99 | 12 |
| 3710.68 | 1856.34 | 2 | 4056.71 | 580.53 | 7 | 4944.19 | 413.02 | 12 |
| 3710.70 | 1237.90 | 3 | 4067.47 | 1356.82 | 3 | 4984.25 | 1662.42 | 3 |
| 3735.70 | 1246.23 | 3 | 4072.85 | 1358.62 | 3 | 5024.29 | 1675.76 | 3 |
| 3740.37 | 341.03 | 11 | 4083.87 | 1362.29 | 3 | 5032.29 | 1678.43 | 3 |
| 3740.68 | 1871.34 | 2 | 4103.87 | 1368.96 | 3 | 5037.16 | 840.53 | 6 |
| 3740.69 | 1871.34 | 2 | 4105.43 | 411.54 | 10 | 5041.22 | 1681.41 | 3 |
| 3740.69 | 1247.90 | 3 | 4106.86 | 1369.95 | 3 | 5067.37 | 1690.12 | 3 |
| 3752.56 | 1877.28 | 2 | 4125.89 | 1376.30 | 3 | 5139.98 | 1285.99 | 4 |
| 3752.56 | 1877.28 | 2 | 4125.90 | 1032.48 | 4 | 5187.13 | 742.02 | 7 |
| 3758.66 | 1253.89 | 3 | 4138.89 | 1035.72 | 4 | 5347.45 | 1783.48 | 3 |
| 3764.69 | 1255.90 | 3 | 4141.89 | 1381.63 | 3 | 5443.66 | 545.37 | 10 |
| 3764.70 | 1883.35 | 2 | 4141.90 | 829.38 | 5 | 5736.81 | 718.10 | 8 |
| 3767.70 | 1884.85 | 2 | 4141.90 | 1036.48 | 4 | 5794.54 | 1449.63 | 4 |
| 3767.71 | 1256.90 | 3 | 4146.85 | 1383.28 | 3 | 5866.82 | 734.35 | 8 |
| 3768.69 | 1885.34 | 2 | 4166.87 | 1389.96 | 3 | 6033.97 | 604.40 | 10 |
| 3768.70 | 1257.23 | 3 | 4174.90 | 1392.63 | 3 | 6068.86 | 1012.48 | 6 |
| 3772.68 | 1258.56 | 3 | 4227.91 | 1410.30 | 3 | 6299.49 | 573.68 | 11 |
| 3772.78 | 472.60 | 8 | 4256.50 | 609.07 | 7 | 6485.94 | 1081.99 | 6 |
| 3783.70 | 1262.23 | 3 | 4284.08 | 715.01 | 6 | 6497.70 | 929.24 | 7 |
| 3801.68 | 1901.84 | 2 | 4293.59 | 1432.20 | 3 | 6533.04 | 1634.26 | 4 |
| 3809.68 | 1905.84 | 2 | 4293.61 | 1074.40 | 4 | 6596.20 | 1100.37 | 6 |
| 3813.72 | 1907.86 | 2 | 4293.61 | 1432.20 | 3 | 6642.67 | 665.27 | 10 |
| 3813.74 | 1907.87 | 2 | 4293.62 | 1074.41 | 4 | 6785.24 | 1131.87 | 6 |
| 3821.66 | 1911.83 | 2 | 4299.91 | 1434.30 | 3 | 6851.12 | 1713.78 | 4 |
| 3821.67 | 1911.83 | 2 | 4319.64 | 1080.91 | 4 | 7367.27 | 1842.82 | 4 |
| 3826.71 | 1276.57 | 3 | 4334.86 | 867.97 | 5 | 7488.09 | 577.01 | 13 |
| 3826.71 | 1914.35 | 2 | 4362.00 | 1455.00 | 3 | 7562.35 | 1891.59 | 4 |
| 3839.45 | 1280.82 | 3 | 4385.06 | 1097.27 | 4 | 7805.46 | 651.46 | 12 |
| 3840.70 | 1281.23 | 3 | 4392.97 | 1465.32 | 3 | 7973.68 | 886.96 | 9 |
| 3841.70 | 1921.85 | 2 | 4431.07 | 1108.77 | 4 | 8099.66 | 675.97 | 12 |
| 3846.73 | 1924.36 | 2 | 4459.12 | 1115.78 | 4 | 8233.73 | 1647.75 | 5 |
| 3847.72 | 1283.57 | 3 | 4481.06 | 1494.69 | 3 | 8337.73 | 1192.10 | 7 |
| | | | | | | 8575.90 | 1430.32 | 6 |