HIGH CONFIDENCE IDENTIFICATION OF CROSS-LINKED PEPTIDES BY MASS SPECTROMETRY CLEAVABLE CROSS-LINKING TECHNOLOGIES

by

JAYANTA K. CHAKRABARTY

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2019

Copyright by

Jayanta K. Chakrabarty

2019

All Rights Reserved

Acknowledgement

During my graduate life in UTA, I had the opportunities to meet and work with a lot of amazing people whose remarkable support, guidance and have made me successful up to the point. I have always had inadequate loquacity, so often I have not expressed my gratitude to them. Finally, I get the change to write a thanksgiving note to certain individuals for their support in every aspect of my PhD journey.

First and foremost, I would like to express the deepest appreciation to my advisor, Dr. Saiful M. Chowdhury for instilling in me the qualities of being a good scientist. I am grateful for his auspicious guidance and mentorship throughout my PhD journey. I sincerely appreciate his appropriate support at crucial moments in my work.

I would also like to convey my thankfulness to all my committee members, Dr. Kevin Schug, Dr. Daniel Armstrong and Dr. Jonguyn Heo for their valuable feedback, advice, and support to my research work. I express my sincere thanks to our collaborator Dr. Alejandro Bugarin group for synthesizing different compounds for my projects. I also owe sincere and earnest thankfulness to my past and present lab mates, specially Dr. Abu hena M Kamal, Dr. ADA Shahinuzzaman, Dr. Maheshika Wanigasekara, Dr. Ruchika Bhawal, and Zixiang Fang. I could never have reached or explored the depths without their help and support. I want to thank Dr. Enrique Barragan and Dr. Pawan Thapa for their guidance in solution phase synthesis.

I would like to thank Dr. William Cleaver and Dr. Cynthia Griffith for supporting my teaching duties. My deep appreciation goes to Dr. Roy McDougald and Dr. Brian Edwards for their time and supportive efforts to ensure the instrument's smooth operation. I hugely appreciate Debbie Cooke, Jill Howard, and Stephanie Henry for their kind help and tremendous support.

My humble gratitude also goes to my colleagues Dr. Mohammad Kabir Hossain, Dr. Sajib Barman, Dr. Md. Fakhrul Islam, Dr. Monira Obaid for their valuable suggestions and motivations. Dr. Mohammad Showkat Hossain and Dr. Pranab Sarker also helped me initially settling down in Texas. I remember them thankfully.

Finally, my heartiest gratitude is reserved for my parents who have instilled in me a dream. I am eternally thankful for my wife for his endless love, and support. Words cannot express how grateful I am for everything she has done for me. I'm grateful to have them all throughout my life.

Abstract

High Confidence Identification of Cross-linked Peptides by Mass Spectrometry Cleavable Cross-linking Technologies

Jayanta K. Chakrabarty, PhD

The University of Texas at Arlington, 2019

Supervising Professor: Dr. Saiful M. Chowdhury

Mass spectrometry (MS)-based chemical cross-linking technique is a potential tool for understanding the protein structure and protein-protein interactions. However, low abundances of cross-linked products and the data complexity are the major challenges in this field. New strategies are required to pinpoint cross-linked peptides with high confidence.

Our lab designed and synthesized a novel cross-linker with two differential MS-cleavable bonds, namely DUCCT. DUCCT-cross-linked peptide generates different diagnostic fragment ions in Collision Induced Dissociation (CID) and Electron Transfer Dissociation (ETD) (tandem mass spectrometric technique) which helps in the unambiguous identification of cross-linked peptide. We demonstrate the efficiency of DUCCT technology with model peptides and proteins. Moreover, we found that DUCCT showed better labeling efficiency compared to BS3 (commercial cross-linker) in an immune cell called macrophage.

The low abundance of inter-crosslinked peptides in a mixture of high abundance of unmodified and dead-end peptides hinders the data analysis. To address this limitation, our lab developed a next-generation DUCCT cross-linker with enrichment functionality (PC-DUCCT-Biotin). This technique can enrich the cross-linked products from the complex biological sample, and the

tandem mass spectrometry technique provides confidence in the identification of cross-linked peptides by generating diagnostic fragments.

DUCCT-treated cross-linked protein samples were automatically analyzed by the in-house developed bio-informatics tool, Cleave-XL. We also demonstrated the capability of Cleave-XL in searching the cross-linked products from protein complexes. Currently, the software can efficiently search the small or medium size protein complexes, containing 5 to 10 proteins. The improvement process of the Cleave-XL is underway, and we believe that it will be an efficient bio-informatics tool for protein structure and interactome study in the bio-medical field.

As a continuation of cross-linking development, we designed another compact MS-cleavable cross-linker, HI-ETD-XL. The signature fragmentation pattern of HI-ETD-XL during electron transfer dissociation (ETD) will contribute significantly to the structural proteomics field.

Currently, we are working with Toll-like Receptors (TLRs) signaling cascades using different generations of cleavable cross-linkers. We believe our newly developed cross-linking approaches will significantly contribute to protein interactions and structural proteomics research.

Table of Contents

	Acknowledgments	iii
	Abstract	v
	List of Illustrations	x
	List of Tables	xiv
	List of Abbreviations	xv
Chapter 1	**Introduction**	**1-12**
1.1	Proteomics	1
1.2	Structural Proteomics	2
1.3	Protein-Protein Interactions	2
1.4	Mass Spectrometry for Proteomics	3
1.5	Tandem Mass Spectrometry	4
1.6	Mass spectrometry (MS) based bottom-up proteomics	5
1.7	Chemical Cross-linking Techniques (XL-MS)	6
1.8	Cleavable Cross-linkers	8
1.9	Thesis organization	9
1.10	References	10
Chapter 2	**Differential tandem mass spectrometry-based cross-linker: a new approach for high confidence in identifying protein cross-linking**	**13-36**
2.1	Introduction	14
2.2	Experimental Section	17
2.2.1	Materials	17
2.2.2	Cross-Linking of Neurotensin (Peptide)	17
2.2.3	Cross-Linking of Ubiquitin and BSA	18
2.2.4	In-Solution Digestion	18
2.2.5	In-Gel Digestion	18
2.2.6	Instrumental Analysis	19
2.2.7	Synthesis of the Cross-Linker	20
2.2.8	LPS-Biotin Pulldown Studies	20
2.3	Results and Discussion	22
2.3.1	Cross-Linker Design and Synthesis	22
2.3.2	Cross-Linked Peptide Identification Strategies	23
2.3.3	Proof-of Concept Studies with DUCCT Cross-Linker	25
2.3.4	Confidence in Cross-Linked Peptide Identifications	31
2.3.5	Labeling Studies in Cell Lysate	32
2.4	Conclusions	33
2.5	References	34
Chapter 3	**High confidence identification of cross-linked peptides by an enrichment-based dual cleavable cross-linking technology (DUCCT) and data analysis tool Cleave-XL**	**37-61**
3.1	Introduction	38
3.2	Experimental	41
3.2.1	Materials	41

3.2.2	Synthesis of PC-DUCCT-Biotin	42
3.2.3	Cross-Linking of Model Peptides	42
3.2.4	Cross-Linking of Ubiquitin and BSA	43
3.2.5	Cross-linking of an immune-receptor protein complex (TLR4 protein complexes)	43
3.2.6	Avidin-Biotin affinity purification and Photo-cleavage	44
3.2.7	Nano HPLC/ Nano-ESI-LTQ Mass Spectrometry Analysis	44
3.3	Result and Discussions	45
3.3.1	Photo-cleavable DUCCT cross-linker concept and strategies	45
3.3.2	Cross-linking of standard peptides:	47
3.3.3	Software Description / Program Workflow	49
3.3.4	Testing the search efficiency of Cleave-XL with standard proteins	51
3.3.5	Testing Cleave-XL software in DUCCT cross-linked immune receptor signaling protein complex	55
3.4	Conclusion	58
3.5	References	59
Chapter 4	**Evaluating the Performance of an ETD-Cleavable Cross-Linking Strategy for Elucidating Protein Structures**	**62-85**
4.1	Introduction	63
4.2	Materials and Methods	66
4.2.1	Materials	66
4.2.2	Synthesis of Cross-linking Reagent	67
4.2.3	Cross-Linking of standard Peptides	68
4.2.4	Protein Cross-linking Experiments	68
4.2.5	Digestion of Cross-linked Proteins	69
4.2.6	Nano HPLC/ Nano-ESI-LTQ Analysis	69
4.2.7	Data Analysis	70
4.3	Result and discussions	71
4.3.1	Design and Synthesis of a Novel ETD-Cleavable Cross-Linker	71
4.3.2	Cross-Linked Peptide Identification Strategies	72
4.3.3	Characterization of HI-ETD-XL Cross-Linked Model Peptides by MS^n Analysis	74
4.3.4	Characterization of HI-ETD-XL Cross-Linked Model Proteins by MSn Analysis	76
4.3.5	HI-ETD-XL Cross-Linking Maps of Ubiquitin and Myoglobin	81
4.4	Conclusions	82
4.5	References	82

Chapter 5	**Concluding Summary**	**86-87**
Appendix A	Supporting Information for Chapter 2	88-99
Appendix B	Supporting Information for Chapter 3	100-117
Appendix C	Supporting Information for Chapter 4	118-134
	Biographical Information	135

List of Illustrations

Figure No.	Legend	Page No.
Figure 1-1	Major directions of Proteomics	1
Figure 1-2	Schematic of the main components of a mass spectrometer	4
Figure 1-3	Schematic diagram of Tandem mass spectrometry	5
Figure 1-4	Bottom-up proteomics approach for protein identifications	6
Figure 1-5	Simplified representation of cross-linking strategy	7
Figure 1-6	Mass Spectrometry-Cleavable Cross-linking Approach	8
Figure 2-1	The design and chemical structure of the DUCCT cross-linker. (A) The design of the dual CID and ETD cleavable cross-linker. (B) The chemical structure of the novel CID and ETD MS-cleavable cross-linker. The green and blue symbols indicate the CID (DP bond) and ETD (hydrazone) MS-cleavable sites, respectively. (C) The structure of DUCCT is shown with spacer chain length	22
Figure 2-2	The scheme to identify inter-cross-linked (top) and dead-end peptides (bottom) by CID and ETD tandem mass spectrometry. MS3 can be used for CID-only identification. Both CID and ETD will identify peptide sequence without further MS/MS	24
Figure 2-3	CID and ETD mass spectra of a cross-linked neurotensin peptide dimer. (A) CID and (B) ETD fragmentation methods show different mass spectra signatures for the same peptide. Please see the Supplementary Information for the added masses from the cross-linker due to differential cleavages	26
Figure 2-4	An example of CID and ETD mass spectra of an inter-cross-linked peptide derived from ubiquitin. The CID and ETD mass spectra show different signatures for the same cross-linked peptide	28
Figure 2-5	Examples of CID and ETD mass spectra of a dead-end cross-linked peptide derived from ubiquitin. The CID and ETD mass spectra show different signatures for the same cross-linked peptide	30
Figure 2-6	(Upper left) The SDS-PAGE gel of biotin-avidin pulldown experiments (M = marker, 1 = LPS biotin, 2 = LPS-biotin/DUCCT, and 3 = LPS-biotin/BS3). (Upper right) A Venn diagram of the number of proteins identified. (Bottom) A partial heat map of selected proteins identified exclusively in cross-linked samples after pulldown studies with avidin (see the full heat map in Figures S11, S12). The scale denotes the spectral counts. For the heat map, 1 = LPS biotin, 2 = LPS-biotin/BS3, and 3 = LPS-biotin/DUCCT.	33
Figure 3-1	The design and chemical structure of the affinity tagged photo-cleavable dual cleavable cross-linking technology (PC-DUCCT-biotin). (A) The design of the dual CID and ETD cleavable cross-linker with an enrichment functionality and a photo-cleavable moiety. The orange and blue diamond symbols indicate the CID (DP bond) and ETD (hydrazone) MS-cleavable sites, respectively. Red triangle indicates photo-cleavable sites. (B) The chemical structure of the PC-DUCCT-Biotin.	45
Figure 3-2	Scheme to identify cross-linked peptides by PC-DUCCT Biotin using CID and ETD tandem mass spectrometry. Complementary CID and ETD-MS/MS will identify peptide sequence without further MS/MS of the fragment ions.	46

Figure 3-3	Experimental CID and ETD mass spectra of a cross-linked neurotensin peptide dimer identified by PC-DUCCT-Biotin cross-linker. The CID and ETD mass spectra showed different signatures for the same cross-linked peptide.	48
Figure 3-4	A screen shot of our standalone software Cleave-XL for the confident identification of cross-linked peptides by sequential CID and ETD cleavable mass spectrometry.	50
Figure 3-5	Example of CID and ETD mass spectra of an inter cross-linked peptide, derived from BSA. CID and ETD signatures of the same cross-linked peptide provided high-confidence identification of this cross-linked peptide.	53
Figure 3-6	Example of an inter cross-linked peptide at m/z 622.48 (6+) identified from TLR4 complex by DUCCT. CID and ETD mass spectra generated different signatures for the same cross-linked peptides.	57
Figure 4-1	The design and chemical structure of the ETD-Cleavable cross-linker (ETXL). (A) The design of the ETD cleavable cross-linker. (B) The chemical structure of ETD MS-cleavable cross-linker. (C) The structure of ET-XL is shown with spacer chain length (13.965 A).	71
Figure 4-2	Structure of ET-XL with ETD cleaved signature fragments. The blue arrow points to the bond that is preferentially cleaved by ETD process. The workflow for cross-linking proteins is shown. After protein cross-linking, trypsin digest generates the cross-linked peptide for LC/MSn analysis. ETD-MS/MS leads to selective cleavage of the N-N bond.	73
Figure 4-3	ETD-MS/MS of identified Bradykinin (RPPGFSPFR) Dimer (m/z 474.63, 5+). Showing the selective cleavage in N-N bond and generating 2 doubly charged diagnostic ions (m/z 590.28 and 596.37).	75
Figure 4-4	MSn Analysis of identified inter cross-linked peptides derived from Myoglobin protein. (A) ETD-MS/MS of cross-linked peptide IPIKYLEFISDAIIHVLHSK-----LFTGHPETLEKFDK (m/z 850.66, 5+). (B) CID-MS3 of one of the diagnostic fragment ions derived from ETD product ions.	77
Figure 4-5	(A) Characteristic fragmentation pattern of ET-XL cross-linked peptide (ALKAWSVAR--------LAKEYEATLEECCAK) derived from BSA. 2 signature ions (m/z 973.34 and m/z 1121.65) are produced by ETD-MS/MS. (B) MS3 spectrum of one of the MS2 fragment ions (m/z 973.34) is demonstrated.	78
Figure 4-6	Examples of one of the inter cross-linked peptide identified from Ubiquitin. (A) ETD-MS/MS of cross-linked peptide (IQDKEGIPPDQQR--------TLSDYNIQKESTLHLVLR) (m/z 781.56, 5+).	80
Figure A-S1	Fragmentation of inter-cross-liked peptides by Collision Induced Dissociation (CID). CID cleaves in a specific site (DP bond) of cross-linker and produces two peptide ions (P1 and P2) with fragmented cross-linker residue (F1 and F2) (signature fragments).	88
Figure A-S2	Fragmentation of inter-cross-linked peptides by Electron Transfer Dissociation (ETD). ETD cleaves in different site (N-N bond) than CID and produces two peptide ions (P1 and P2) with added different cross-linker residues (F1 and F2) (signature fragments) and neutralized precursor ions	88
Figure A-S3	CID (top) and ETD (bottom) fragmentation patterns of dead-end peptides (one-end hydrolyzed). CID and ETD generate two complementary fragment signature ions of one peptide	89

Figure A-S4	ESI-IT-TOF-MS measurement of CID and ETD based cross-linker ([M+H]+ = 864.2421). The molecular mass of the cross-linker is 863.2358 Da. We observed the mass in the MS as the most abundant peak. Efficient fragmentation is shown in Asp-Pro peptide bond by CID. Cleavage mass, ([M+H]+ = 385.1343, mass accuracy 2.9 ppm). RG-reactive groups.	90
Figure A-S5	Structure of DUCCT with CID cleaved signature fragments. DUCCT has two reacting groups (NHS ester). Loss of NHS happened during the reaction with peptides. CID cleaves Asp-Pro peptide bond of DUCCT and generates 2 fragments.	92
Figure A-S6	SDS PAGE gel image of cross-linked Ubiquitin (left) and BSA (right). A distinct higher molecular weight band of cross-linked protein is the primary evidence of efficient labeling of DUCCT.	93
Figure A-S7	Structure of DUCCT with ETD cleaved signature fragments. ETD also generates 2 frag-ments that are different from CID cleaved fragments.	95
Figure A-S8	Examples of MS/MS/MS spectra of cross-linked fragments derived from A) Ubiquitin and B) Neurotensin. These fragmentation patterns unambiguously confirm the sequence of the peptides	96
Figure A-S9	CID and ETD mass spectra of dead-end peptides derived from A) Neurotensin and B) Ubiquitin. CID and ETD mass spectra generate different signature ions for the same dead-end peptide.	97
Figure A-S10	Crystal structure of ubiquitin (pdb-1V80) with the color-coded sites and distances. High confidence identification clearly pointed out the cross-linked lysine residues.	98
Figure A-S11	Full heat map of proteins which were observed in BioLPS (none), BS3, and DUCCT pulldown studies.	99
Figure A-S12	Partial heatmap of Figure 6 shown with all the protein identifications.	99
Figure B-S1	3-D structure of the novel enrichment based dual cleavable cross-linker (PC-DUCCT-Biotin). The length of the reactive spacer arms is calculated 12.063 Å using a chem draw software.	100
Figure B-S2	CID (top) and ETD (bottom) fragmentation patterns of a dead-end peptide (one-end hydrolyzed) derived from Neurotensin.	101
Figure B-S3	CID and ETD mass spectra of an inter cross-linked peptides derived from Bradykinin. CID and ETD mass spectra generate different signature ions for the same cross-linked peptide.	102
Figure B-S4	Workflow for the identification of inter-peptide cross-links using our Cleave-XL software.	103
Figure B-S5	X-ray crystallographic structure of Ubiquitin (pdb-1V80) with the color-coded cross-linking sites and distances.	104
Figure B-S6	Example of CID and ETD mass spectra of an inter cross-linked peptide, derived from ubiquitin. Lysine 6 (amino acids 1–11); of ubiquitin was found to be cross-linked with lysine 33 (amino acids 30–42).	105
Figure B-S7	Examples of MS/MS of cross-linked fragments derived from Ubiquitin by collision induced dissociation (CID). MS/MS confirmed the modified peptide sequences.	106
Figure B-S8	Inter-crosslinked peptide at m/z 477.25 (5+) (SLGKVGTR…. EKVLTSSAR) derived from BSA after cross-linking with PC-DUCCT-Biotin cross-linker	107

Figure B-S9	Examples of MS/MS of cross-linked fragments derived from BSA by collision induced dissociation (CID). MS/MS of the modified-peptide sequences provided the additional confidence of identification.	108
Figure B-S10	Examples of signature fragmentation patterns (CID-MS/MS and ETD-MS/MS) of dead-end peptide derived from BSA.	109
Figure B-S11	X-ray crystallographic structure of BSA (pdb- 4F5S) with the distance constrains between interacting residues.	110
Figure B-S12	Example of CID and ETD mass spectra of an inter cross-linked peptide, derived from TLR4 and LBP (TLR4 Complex).	111
Figure B-S13	Examples of MS/MS of cross-linked fragments derived from TLR4-Complex by collision induced dissociation (CID)	112
Figure C-S1	The scheme to identify inter-cross-linked and dead-end peptides by ETD tandem mass spectrometry. MS3 can be used for sequence confirmation of cross-linked modified fragments.	124
Figure C-S2	ESI-LTQ-MS measurement of ETD based cross-linker ([M+H] + = 480.52). The molecular mass of the cross-linker is 479.1077 Da.	124
Figure C-S3	ETD mass spectra of a Neurotensin dimer identified by ETD-XL.	125
Figure C-S4	ETD mass spectra of dead-end peptides derived from (A) BSA and (B) Myoglobin. ETD mass spectra generate one signature ion for the dead-end peptide.	126
Figure C-S5	CID-MS/MS of doubly charged signature ions IQDKEGIPPDQQR (A) and TLSDYNIQKESTLHLVLR (B) provide the sequence information of the modified peptides.	127
Figure C-S6	X-ray crystallographic structure of Ubiquitin (pdb- 1V80) with the distance constrains between interacting residues.	128
Figure C-S7	Crystal structure of Myoglobin (pdb-1dwr) with the color-coded sites and distances. High confidence identification clearly pointed out the cross-linked lysine residue.	129
Figure C-S8	Crystal structure of BSA showing the distance between the interacting peptides, identified by ET-XL (pdb :4fs)	130

List of Tables

Table No.	Legends	Page No.
Table A-S1	List of calculated masses and m/z values used to identify cross-linked reaction products. All the experimental m/z's are average due to the utilization of an LTQ Velos Pro MS	91
Table A-S2	List of calculated masses and m/z value used to assign cross-linked products (Ubiquitin)	94
Table A-S3	CID and ETD together yielded the list of inter cross-linked sites in ubiquitin	98
Table B-S1	List of calculated masses and m/z values used to identify cross-linked reaction products derived from neurotensin	113
Table B-S2	List of calculated masses and m/z values used to identify cross-linked reaction products derived from Bradykinin	114
Table B-S3	List of calculated masses and m/z values used to identify inter cross-linked peptides derived from BSA	115
Table B-S4	List of inter cross-linked peptides identified from BSA by Cleave-XL	115
Table B-S5	List of inter cross-linked peptides identified from TLR4 complex by Cleave-XL after comparing CID and ETD output file	116
Table B-S6	List of calculated masses and m/z value used to assign cross-linked products (TLR4 Dimer complex)	117
Table C-S1	List of calculated masses and m/z values used to identify cross-linked reaction products derived from Bradykinin.	131
Table C-S2	Summary of ET-XL-interlinked peptides of BSA identified by LC-MS/MS analysis	131
Table C-S3	Calculated masses and m/z values used to identify cross-linked reaction products derived from Myoglobin.	132
Table C-S4	List of calculated masses and m / z value for assigning cross-linking products (BSA)	132
Table C-S5	List of inter cross-linked peptides along with distance constrains identified from Ubiquitin	133
Table C-S6	List of calculated masses and m/z value used to assign cross-linked products (Ubiquitin)	133
Table C-S7	Table of Confident Identifications of Inter-Crosslinked Peptides from Myoglobin	134

List of Abbreviations

Abbreviations	Name
BSA	Bovine Serum Albumin
BS3	Bis(sulfosuccinimidyl)suberate
CD14	Cluster of differentiation 14
CID	Collision-induced dissociation
CL	Cross-Linker
DUCCT	Dual Cleavable Cross-linking Technology
DTT	Dithiothreitol
DMEM	Dulbecco's Modified Eagle Medium
ESI	Electrospray ionization
ETD	Electron-transfer dissociation
FDR	False Discovery Rate
HPLC	High performance liquid chromatography
LBP	LPS binding protein
LC-MS	Liquid chromatography mass spectrometry
LIT	Linear ion trap
LPS	Lipopolysaccharides
MALDI	matrix-assisted laser desorption/ionization
MD-2	Myeloid differentiation 2
MGF	Mascot generic file
MS/MS	Mass spectrometry in tandem
NHS	N-hydroxysuccinimide
PBS	Phosphate saline buffer
PTMs	Post -translational modifications
SDS	Sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
TNF-a	Tumor necrosis factor-alpha
TLR	Toll Like Receptor
TRIF	TIR-domain-containing adapter-inducing interferon-β
XL-MS	Cross-linking based Mass Spectrometry

Chapter 1

Introduction

1.1 Proteomics

The collection of proteins present in a living cell regulates the overall function of a biological process. The complete set of proteins found in each cell is known as the proteome. Proteomics is the comprehensive functional study of the proteome. Proteomics is complementary to genomics and very essential for understanding the biological function of a cell. The genome sequence itself is not enough to decode the gene function. Although the same set of genes is present in all cells of an organism, the proteins found in different tissues are unique and dependent on the expression of the gene [1, 2]. The genome therefore is fixed, but the proteome varies within an organism. Proteomics mainly deal with structural dynamics of protein, expression of protein, post-translational modifications (PTMs), the movement of proteins between subcellular compartments, and protein to protein interactions (Figure 1-1).

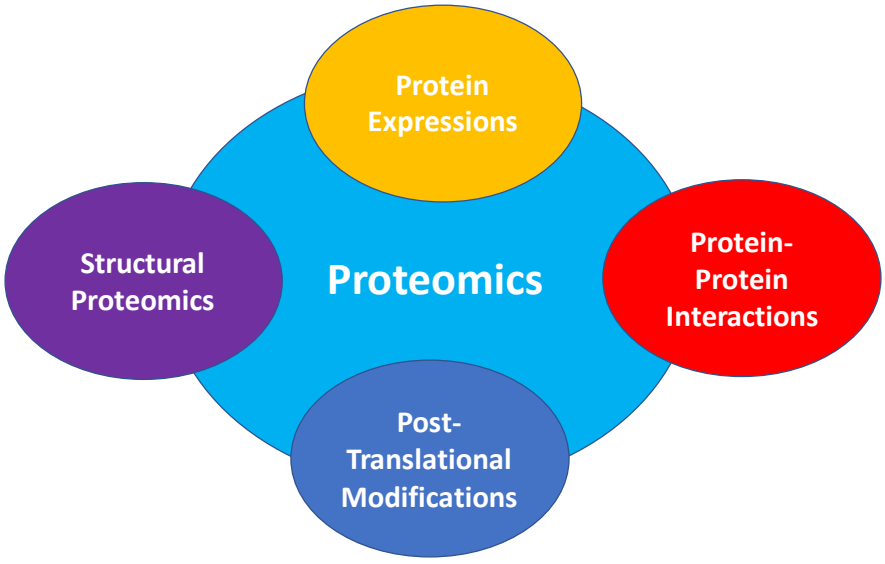


Figure 1-1. Major directions of Proteomics.

1.2 Structural Proteomics

Structural information of proteins is crucial to understand how proteins are functioning in cellular systems. Structural proteomics is the characterization of three-dimensional protein structures to understand the sequence and structure-activity relationship of the proteins [3, 4]. The structural proteomics provides the information at the amino acid residue level, including a distance constraint between functional groups, the degree of exposure of amino acid, and its role in different bonding. Such information is useful for molecular modeling of an unknown protein structure[5, 6].

Biophysical tools like X-ray crystallography and NMR spectroscopy are widely used approaches to characterize protein structures. However, applications of these techniques are limited for heterogeneous and dynamic protein complexes because of their difficulties in sample preparation. Hence, it warrants alternate structural elucidation tools. Recently, Mass spectrometry (MS)-based structural techniques are one of the valuable assets for resolving the low-resolution structure of protein complexes. MS-based strategies such as cross-linking mass spectrometry (XL-MS), covalent labeling/ foot printing, hydrogen−deuterium exchange, and native MS permit characterization of protein complexes with high throughput [7, 8].

1.3 Protein-Protein Interactions

Protein is the structural and functional unit of any biological species. Protein interacts with other molecules to conduct their functions. protein-protein interactions play a significant role in governing most of the biochemical processes, such as signal transduction, cellular metabolism, transportation across membranes, and muscle contractions [9-13]. Protein interactions are mainly classified as stable or transient. Stable interactions are commonly characterized by co-immunoprecipitation, pull-down, or far-western methods.

Most of the cellular processes are regulated by transient interactions. On the other hand, transient interaction is widely observed in different cellular and metabolic activities, including modification of proteins, cellular transportation, folding in structure, signal transduction, cell cycling, etc. Transient interactions are best studied by label transfer or chemical cross-linking techniques [14, 15].

1.4 Mass Spectrometry for Proteomics

Mass spectrometry is one of the most popular techniques over the past few decades for qualitative and quantitative analysis of proteins. It is a unique analytical technique that quantifies the mass of individual molecules and their subunits. The technique can be able to detect the trace amount of analyte and become one of the most sensitive techniques in the bio-analytical field. Nowadays, this technique is routinely used in different biomedical research fields to study complex biological systems. Mass spectrometric sequencing of proteins is one of the most important revolutions in the proteomics field. Mass spectrometry has mostly replaced the other classical techniques because of its sensitivity, versatility and offers higher throughput in proteome analysis. Mass spectrometry is usually coupled with chromatography for analyzing biological samples. The individual component of the samples is separated by chromatography and then detected by the mass spectrometer [16, 17].

Ion source, mass analyzer, and detection unit are three major components of the mass spectrometer (Figure 1-2). Initially, samples are inserted through ion source; the ions are segregated in a mass analyzer based on their mass to charge ratio and recorded by a detector. In the ionization step, the neutral sample molecules are converted into gas phase ionic species. In proteomics, electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) are widely used ionization methods [17-20].

Different kind of mass analyzers are also available, including, quadrupole, 3D ion trap, quadrupole linear ion trap (LTQ), time-of-flight, Fourier transform ion cyclotron resonance (FT-ICR) and, Orbitrap. These analyzers may be used individually or coupled with another one to obtain the tandem analysis. The detector system is composed of dynode and photomultiplier that quantify and amplify the ion current of individual ions. A data recording system assists in recording, processing, and displaying the data.

Figure 1-2. Schematic of the main components of a mass spectrometer.

1.5 Tandem Mass Spectrometry

Tandem Mass spectrometry (MS/MS or MS^n) is an excellent analytical tool for identifying or quantifying the compounds from complex samples. It has become necessary for the conduct of MS /MS experiments to obtain structural information of the compound. In MS/MS, additional energy is applied to fragment or break down the selected precursor ions. The energy can be provided by collisions, photons, or electrons which generates structural information of product ions from the targeted precursor ion. Multistage fragmentation (MS^n) experiments observed in the Ion Trap or FT-ICR type instrument, where repeated isolation and fragmentation of the precursor ions performed for better interpretation of the parent ions [21-23].

Figure 1-3 demonstrates the complete scheme of tandem mass spectrometry technique. Ion activation dissociation of the precursor is crucial to get the information about the structure-specific fragment ions in tandem mass spectrometry.

Collision-induced dissociation (CID) and electron-transfer dissociation (ETD) are two most common fragmentation techniques used for the identification of biomolecules (peptide). In CID, the activation and fragmentation of the precursor is obtained by collisions with an inert gas, such as helium or argon, and eventually cleaves the amide bond of the peptide and produces the structurally informative fragment ions (b and y ions). ETD is another fragmentation technique that allows electron from the reagent anion to be captured by positive precursor ions and subsequent dissociation of the peptide at N-Cα bonds. ETD leaves many CID labile sites intact and is relatively indifferent to peptide sequences. This feature makes ETD a common tool for characterizing post-translation modifications (PTMs), e.g., phosphorylation, glycosylation [24-27].

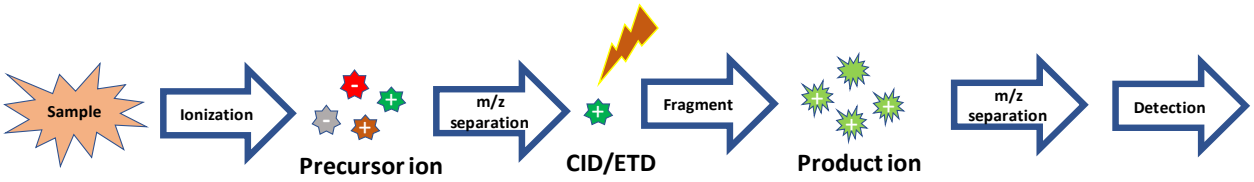


Figure 1-3. Schematic diagram of Tandem mass spectrometry.

1.6 Mass spectrometry (MS) based bottom-up proteomics

Bottom-up proteomics is the most commonly used method for analyzing the proteomic sample. Bottom-up proteomics, also known as shotgun proteomics, refers to identify and characterize the protein by analyzing the peptides. In this approach, protein or protein mixtures are digested with a proteolytic enzyme, followed by separation of the digested peptide by liquid chromatography and finally are subjected to mass spectrometry for identification (Figure 1-4). Peptide identification is accomplished by comparing the MS/MS spectra of the selected peptide with the theoretical database generated from in silico digestion of a protein [28].

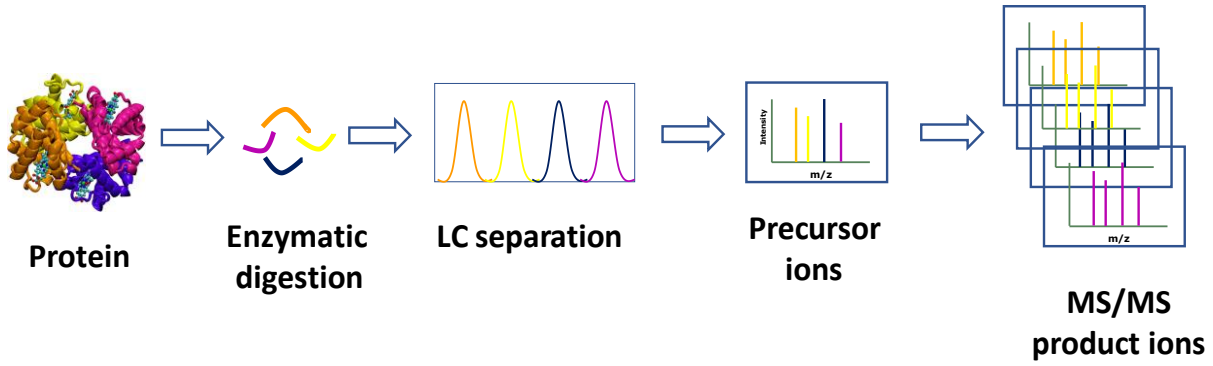


Figure 1-4. Bottom-up proteomics approach for protein identifications.

SEQUEST (Thermo Scientific) and Mascot (Matrix Science) are the two most commonly used tandem mass spectrometry-based data analysis algorithms for protein identification. Protein speculation is achieved by assigning the peptide sequences. The identification of at least two peptides from a specific protein is commonly considered to be a reliable identification of that protein [29].

1.7 Chemical Cross-linking Techniques (XL-MS)

Chemical cross-linking combined with mass spectrometric analysis (XL-MS) has emerged as a robust and reliable technique to study protein to protein interaction and protein structure. This technique eventually overcomes the limitation of conventional biophysical approaches in characterizing large heterogeneous protein complexes [30-34].

A cross-linker, core of the XL-MS, is a chemical reagent containing at least two reactive groups connected with a spacer arm of a defined length. The length of the spacer arm represents the distance between the two interacting residues.

Cross-linker can capture the nearby proteins or protein complexes by forming the covalent bond and hold them tightly so that they will not detach even after purification. Generally, the

functional groups of cross-linking agents are designed to target the nucleophiles such as amine and sulfhydryl groups that exist in the side chain of proteins. Nowadays, a significant number of cross-linkers utilize N-hydroxysuccinimide (NHS) esters and *N*-maleimide moieties that react with a primary amine and thiol groups, respectively.

Typically, chemical cross-linking studies are accomplished in a bottom-up manner, involving covalently linked proteins digestion followed by LC/ESI-MS/MS analysis (Figure 1-5). Usually, protein cross-linking generates three types of cross-link products, such as Type 0, Type 1and Type 2. Type 0 is a single peptide modified by one cross-linker function; Type 1 is an intra-peptide cross-link, and Type 2 is an inter-peptide cross-link**.** Among all, inter-peptide cross-link offers the structurally significant distance constraint information between a protein with its interacting partner [35].

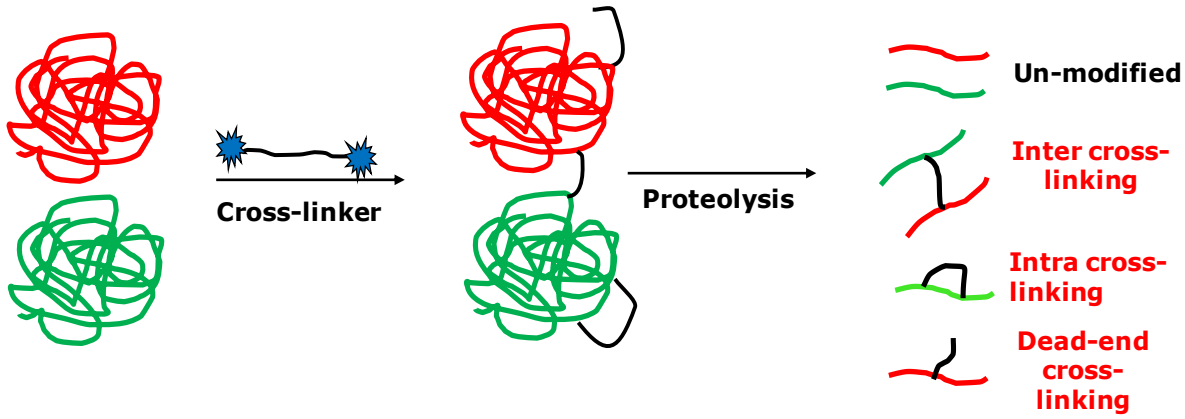


Figure 1-5. Simplified representation of cross-linking strategy.

1.8 Cleavable Cross-linking Agents

One of the major challenges of traditional cross-linking field is its ambiguity in determining the cross-linked peptides. Despite remarkable improvement in cross-linking technologies, database complexity, and variety of cross-linked products (dead-end, intra-, inter-) make this field more challenging for large-scale interactomes research. In native biological experiments, very low amount of inter-crosslinked peptides is generated, which are very difficult to locate from the enormous amount of mass spectrometry data. Also, MS/MS spectra of inter-crosslinked peptides are very complicated because of their fragmentation complexity. Moreover, this analysis becomes unmanageable when large-scale applications are planned. In recent years, different research groups developed several strategies to reduce the data complexity including isotopic labeling techniques, adding enrichment-based affinity tags, incorporating different cleavable bonds [7, 36-39].

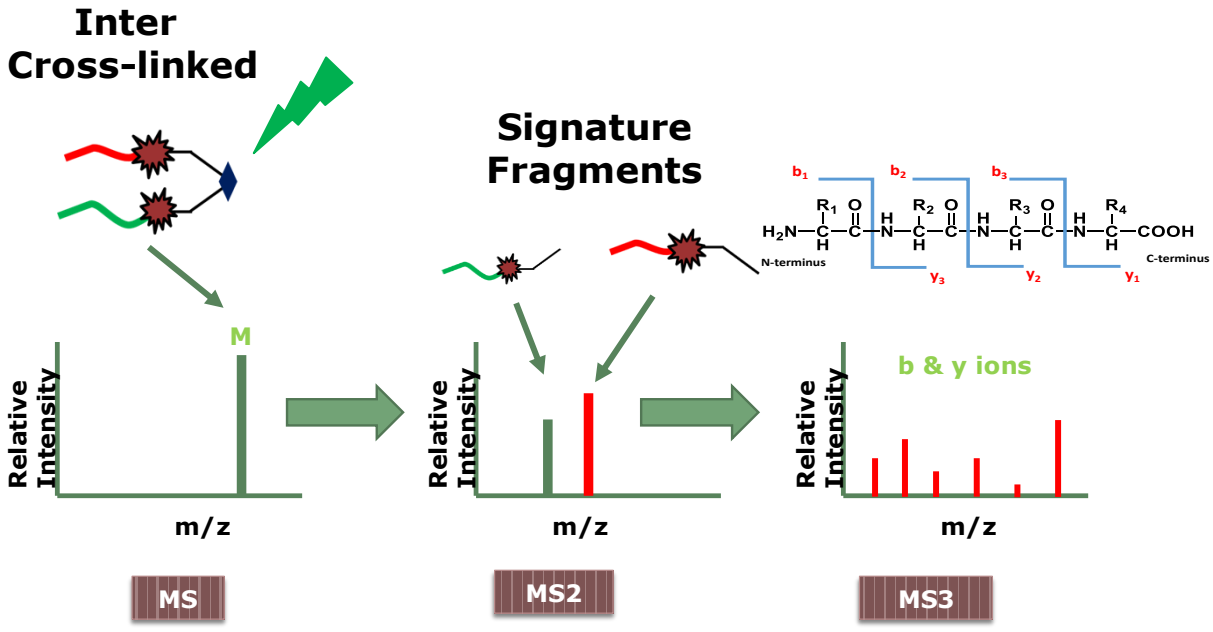


Figure 1-6. Mass Spectrometry-Cleavable Cross-linking Approach.

Cleavable cross-linkers contain one or more labile bonds in their structure and, based on their chemical properties, can preferentially be cleaved in different ways. For example, the cleavages of the labile bonds can be executed by photo, chemical, and tandem MS, which separate two cross-linked peptides before or during the MS analysis. Among all, MS-cleavable is the most attractive because of its unique feature that produces characteristic cross-link fragments in MS2 and subsequently identifies the cross-linked peptides by MS3rd sequencing (Figure 1-6) [7, 22, 40].

1.9 Thesis Organization

The dissertation focusses on the development of a series of novel MS-Cleavable cross-linkers and their application in proteins and protein complexes to improve the confidence in detecting cross-linked peptides. Chapter 2 describes the development of a novel MS-cleavable strategy, DUal Cleavable Cross-linking Technology (DUCCT), and demonstrates the efficiency of DUCCT in characterizing protein cross-linking. DUCCT-Cross-linker contains two gas-phase cleavable bonds in its structure and produces different diagnostic ions in two differential tandem mass spectrometric techniques (CID and ETD) separately. Two complementary characteristic fragment ions provide high confidence in cross-linked peptide identification. Chapter 3 focuses on the development of enrichment based next generation DUCCT cross-linker (PC-DUCCT-Biotin) and a novel bio-informatics software tool (CLEAVE-XL). PC-DUCCT-Biotin can enrich the cross-linked products from a complex biological sample, and the tandem mass spectrometry technique provides authentic identification of interacting peptides by generating signature fragments. Also, our specially designed software CLEAVE-XL can create a list of cross-linked peptides automatically from cross-linker treated mass-spectrometric datasets. Another compact

ETD-cleavable cross-linker (HI-ETD-XL) is described in Chapter 4. Here, we describe the HI-ETD-XL cross-linking experiments with different model peptides and proteins.

The characteristic fragmentation behavior of the HI-ETD-XL during electron transfer dissociation (ETD) pinpoints the cross-linked peptide identification. Chapter 5 presents a general summary of all the projects and future directions of these projects as well.

1.10 References

1. Blackstock, W.P., Weir, M.P.: Proteomics: quantitative and physical mapping of cellular proteins. Trends Biotechnol. 17, 121-127 (1999)
2. Vlahou, A., Fountoulakis, M.: Proteomic approaches in the search for disease biomarkers. J Chromatogr B Analyt Technol Biomed Life Sci. 814, 11-19 (2005)
3. Leitner, A., Joachimiak, L.A., Unverdorben, P., Walzthoeni, T., Frydman, J., Forster, F., Aebersold, R.: Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. Proc Natl Acad Sci U S A. 111, 9455-9460 (2014)
4. Petrotchenko, E.V., Borchers, C.H.: Modern mass spectrometry-based structural proteomics. Adv Protein Chem Struct Biol. 95, 193-213 (2014)
5. Sinz, A.: The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks. Expert Rev Proteomics. 11, 733-743 (2014)
6. Yee, A., Pardee, K., Christendat, D., Savchenko, A., Edwards, A.M., Arrowsmith, C.H.: Structural proteomics: toward high-throughput structural biology as a tool in functional genomics. Acc Chem Res. 36, 183-189 (2003)
7. Yu, C., Kandur, W., Kao, A., Rychnovsky, S., Huang, L.: Developing new isotope-coded mass spectrometry-cleavable cross-linkers for elucidating protein structures. Anal Chem. 86, 2099-2106 (2014)
8. Gutierrez, C.B., Block, S.A., Yu, C., Soohoo, S.M., Huszagh, A.S., Rychnovsky, S.D., Huang, L.: Development of a Novel Sulfoxide-Containing MS-Cleavable Homobifunctional Cysteine-Reactive Cross-Linker for Studying Protein-Protein Interactions. Anal Chem. 90, 7600-7607 (2018)
9. De Las Rivas, J., Fontanillo, C.: Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. Plos Comput Biol. 6, e1000807 (2010)
10. Jones, S., Thornton, J.M.: Principles of protein-protein interactions. Proc Natl Acad Sci U S A. 93, 13-20 (1996)
11. Keskin, O., Tuncbag, N., Gursoy, A.: Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. Chem Rev. 116, 4884-4909 (2016)
12. Miura, K.: An Overview of Current Methods to Confirm Protein-Protein Interactions. Protein Pept Lett. 25, 728-733 (2018)
13. Smits, A.H., Vermeulen, M.: Characterizing Protein-Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. Trends Biotechnol. 34, 825-834 (2016)
14. Westermarck, J., Ivaska, J., Corthals, G.L.: Identification of protein interactions involved in cellular signaling. Mol Cell Proteomics. 12, 1752-1763 (2013)

15. Acuner Ozbabacan, S.E., Engin, H.B., Gursoy, A., Keskin, O.: Transient protein–protein interactions. Protein Engineering, Design and Selection. 24, 635-648 (2011)
16. Finehout, E.J., Lee, K.H.: An introduction to mass spectrometry applications in biological research. Biochem Mol Biol Educ. 32, 93-100 (2004)
17. Liebisch, G., Scherer, M.: Quantification of bioactive sphingo- and glycerophospholipid species by electrospray ionization tandem mass spectrometry in blood. J Chromatogr B Analyt Technol Biomed Life Sci. 883-884, 141-146 (2012)
18. Fagerquist, C.K., Sultan, O.: A new calibrant for matrix-assisted laser desorption/ionization time-of-flight-time-of-flight post-source decay tandem mass spectrometry of non-digested proteins for top-down proteomic analysis. Rapid Commun Mass Spectrom. 26, 1241-1248 (2012)
19. Banerjee, S., Mazumdar, S.: Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. Int J Anal Chem. 2012, 282574 (2012)
20. Karas, M., Bahr, U., Dulcks, T.: Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. Fresenius J Anal Chem. 366, 669-676 (2000)
21. Chowdhury, S.M., Du, X., Tolic, N., Wu, S., Moore, R.J., Mayer, M.U., Smith, R.D., Adkins, J.N.: Identification of cross-linked peptides after click-based enrichment using sequential collision-induced dissociation and electron transfer dissociation tandem mass spectrometry. Anal Chem. 81, 5524-5532 (2009)
22. Soderblom, E.J., Goshe, M.B.: Collision-induced dissociative chemical cross-linking reagents and methodology: Applications to protein structural characterization using tandem mass spectrometry analysis. Anal Chem. 78, 8059-8068 (2006)
23. Sadygov, R.G., Cociorva, D., Yates, J.R., 3rd: Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. Nat Methods. 1, 195-202 (2004)
24. Bishop, A., Brodbelt, J.S.J.I.J.o.M.S.: Selective cleavage upon ETD of peptides containing disulfide or nitrogen–nitrogen bonds. 378, 127-133 (2015)
25. Chi, A., Huttenhower, C., Geer, L.Y., Coon, J.J., Syka, J.E., Bai, D.L., Shabanowitz, J., Burke, D.J., Troyanskaya, O.G., Hunt, D.F.: Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. Proc Natl Acad Sci U S A. 104, 2193-2198 (2007)
26. Frese, C.K., Altelaar, A.F., Hennrich, M.L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A.J., Mohammed, S.: Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J Proteome Res. 10, 2377-2388 (2011)
27. Zhao, B., Reilly, C.P., Reilly, J.P.: ETD-Cleavable Linker for Confident Cross-linked Peptide Identifications. J Am Soc Mass Spectrom. 30, 1631-1642 (2019)
28. Resing, K.A., Ahn, N.G.: Proteomics strategies for protein identification. FEBS Lett. 579, 885-889 (2005)
29. Pandey, A., Mann, M.: Proteomics to study genes and genomes. Nature. 405, 837-846 (2000)
30. Iacobucci, C., Piotrowski, C., Aebersold, R., Amaral, B.C., Andrews, P., Bernfur, K., Borchers, C., Brodie, N.I., Bruce, J.E., Cao, Y., Chaignepain, S., Chavez, J.D., Claverol, S., Cox, J., Davis, T., Degliesposti, G., Dong, M.Q., Edinger, N., Emanuelsson, C., Gay,

M., Gotze, M., Gomes-Neto, F., Gozzo, F.C., Gutierrez, C., Haupt, C., Heck, A.J.R., Herzog, F., Huang, L., Hoopmann, M.R., Kalisman, N., Klykov, O., Kukacka, Z., Liu, F., MacCoss, M.J., Mechtler, K., Mesika, R., Moritz, R.L., Nagaraj, N., Nesati, V., Neves-Ferreira, A.G.C., Ninnis, R., Novak, P., O'Reilly, F.J., Pelzing, M., Petrotchenko, E., Piersimoni, L., Plasencia, M., Pukala, T., Rand, K.D., Rappsilber, J., Reichmann, D., Sailer, C., Sarnowski, C.P., Scheltema, R.A., Schmidt, C., Schriemer, D.C., Shi, Y., Skehel, J.M., Slavin, M., Sobott, F., Solis-Mezarino, V., Stephanowitz, H., Stengel, F., Stieger, C.E., Trabjerg, E., Trnka, M., Vilaseca, M., Viner, R., Xiang, Y., Yilmaz, S., Zelter, A., Ziemianowicz, D., Leitner, A., Sinz, A.: First Community-Wide, Comparative Cross-Linking Mass Spectrometry Study. Anal Chem. 91, 6953-6961 (2019)

31. Tran, B.Q., Goodlett, D.R., Goo, Y.A.: Advances in protein complex analysis by chemical cross-linking coupled with mass spectrometry (CXMS) and bioinformatics. Biochim Biophys Acta. 1864, 123-129 (2016)
32. Du, X., Chowdhury, S.M., Manes, N.P., Wu, S., Mayer, M.U., Adkins, J.N., Anderson, G.A., Smith, R.D.: Xlink-identifier: an automated data analysis platform for confident identifications of chemically cross-linked peptides using tandem mass spectrometry. J Proteome Res. 10, 923-931 (2011)
33. Liu, F., Lossl, P., Scheltema, R., Viner, R., Heck, A.J.R.: Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. Nat Commun. 8, 15473 (2017)
34. Iacobucci, C., Götze, M., Ihling, C.H., Piotrowski, C., Arlt, C., Schäfer, M., Hage, C., Schmidt, R., Sinz, A.: A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein–protein interactions. Nat Protoc. 13, 2864 (2018)
35. Bian, S., Chowdhury, S.: Profiling Protein-Protein Interactions and Protein Structures Using Chemical Cross-linking and Mass Spectrometry. J Austin J Biomed Eng. 1, 1017 (2014)
36. Chakrabarty, J.K., Naik, A.G., Fessler, M.B., Munske, G.R., Chowdhury, S.M.: Differential Tandem Mass Spectrometry-Based Cross-Linker: A New Approach for High Confidence in Identifying Protein Cross-Linking. Anal Chem. 88, 10215-10222 (2016)
37. Chowdhury, S.M., Munske, G.R., Tang, X., Bruce, J.E.: Collisionally activated dissociation and electron capture dissociation of several mass spectrometry-identifiable chemical cross-linkers. Anal Chem. 78, 8183-8193 (2006)
38. Hage, C., Falvo, F., Schafer, M., Sinz, A.: Novel Concepts of MS-Cleavable Cross-linkers for Improved Peptide Structure Analysis. J Am Soc Mass Spectrom. 28, 2022-2038 (2017)
39. Iacobucci, C., Hage, C., Schafer, M., Sinz, A.: A Novel MS-Cleavable Azo Cross-Linker for Peptide Structure Analysis by Free Radical Initiated Peptide Sequencing (FRIPS). J Am Soc Mass Spectrom. 28, 2039-2053 (2017)
40. Yang, L., Tang, X., Weisbrod, C.R., Munske, G.R., Eng, J.K., von Haller, P.D., Kaiser, N.K., Bruce, J.E.: A photocleavable and mass spectrometry identifiable cross-linker for protein interaction studies. Anal Chem. 82, 3556-3566 (2010)

Chapter 2

Differential tandem mass spectrometry-based cross-linker: a new approach for high confidence in identifying protein cross-linking

Reproduced with permission from Chakrabarty, J. K., Naik, A. G., Fessler, M. B., Munske, G. R., & Chowdhury, S. M. (2016). Differential tandem mass spectrometry-based cross-linker: a new approach for high confidence in identifying protein cross-linking. *Analytical Chemistry*, 88(20), 10215-10222. Copyright © 2016, American Chemical Society. https://doi.org/10.1021/acs.analchem.6b02886

Abstract

Chemical cross-linking and mass spectrometry are now widely used to analyze large-scale protein-protein interactions. The major challenge in cross-linking approaches is the complexity of the mass spectrometric data. New approaches are required that can identify cross-linked peptides with high-confidence and establish a user-friendly analysis protocol for the biomedical scientific community. Here, we introduce a novel cross-linker that can be selectively cleaved in the gas phase using two differential tandem mass-spectrometric fragmentation methods, such as collision-induced or electron transfer dissociation (CID and ETD). This technique produces two signature mass spectra of the same cross-linked peptide, thereby producing high confidence in identifying the sites of interaction. Further tandem mass spectrometry can also give additional confidence on the peptide sequences. We demonstrate a proof-of-concept for this method using standard peptides and proteins. Peptides and proteins were cross-linked, and their fragmentation characteristics were analyzed using CID and ETD tandem mass spectrometry. Two sequential cleavages unambiguously identified cross-linked peptides. In addition, the labeling efficiency of the new cross-linker was evaluated in macrophage immune cells after stimulation with the microbial ligand lipopolysaccharide and subsequent pulldown experiments with biotin-avidin affinity chromatography. We believe this strategy will help advance insights into the structural biology and systems biology of cell signaling.

2.1 Introduction

Current biochemical methods are not very efficient at analyzing systems-level or large-scale protein interaction networks. Most studies utilize a technique called “co-immunoprecipitation,” in which a protein is isolated along with its interacting partners (i.e., the protein complex) using an antibody or by incorporating an affinity group in the protein, which can be used as a hook to selectively purify it [1]. This method is applicable for very strong and stable interactions, but most protein-protein interactions in cells are likely to be transient and weak. During the purification process these interactions may be lost completely. Moreover, co-immunoprecipitation is qualitative and generally provides little detailed information on the protein-to-protein interaction domains involved.

To solve this very important analytical shortcoming, a chemistry-based fixation method combined with mass spectrometry has come into the limelight, in which a reactive compound, referred to as a cross-linker, is utilized to stabilize a protein with its interaction partners by derivatizing certain side chains of the protein before cell lysis is performed [2]. A cross-linker can then fix adjacent proteins or protein complexes using a chemical reaction, holding them tightly so they will not detach during cell lysis or subsequent strict purification conditions. In addition, a cross-linker can only react within a limited distance; hence, protein reactive sites can be measured by calculating the distance between the cross-linked sites. This method has two advantages: 1) it can identify large-scale protein interactions; and 2) it can identify protein structures in their native biological conditions.

The widespread use of this technology is hindered due to several bottlenecks. Traditional cross-linking strategies generate an enormous amount of mass spectrometry data, which is extremely difficult to analyze with routine software tools. Finding these interactions in large datasets is equivalent to finding a needle in a haystack. In this regard, several researchers have contributed by designing new strategies that add either enrichment functionality or cleavable bonds in the cross-linkers [2-12]. To deal with the enormous complexity of data-analysis from traditional cross-linking, researchers have developed cleavable cross-linking approaches.

The first efficient cleavable cross-linker, called PIR, used reporter ions and was developed by the James Bruce group at Washington State University and the University of Washington [5,13].

Although there were some concerns about the reactive distances, this cross-linking approach has been successfully applied in several large-scale systems and is currently the most efficient cleavable cross-linker reported to date [14-17]. Another efficient cleavable cross-linker was developed by Lan Huang's group at UC Irvine. This DSSO cross-linker utilized the labile nature of the sulfoxide group and was also successfully used in several large-scale applications [16,17]. Goshe's group at NC State University utilized an Asp-Pro (DP) peptide bond in their cross-linkers, which are efficiently cleaved during low-energy CID-MS/MS.4,18 In addition, a significant number of cross-linkers have been reported by Petrotchenko et al [3,19]. The Heck group has also recently developed an integrated workflow for proteome-wide profiling of protein cross-linking using the CID cleavable cross-linker DSSO [20].

Although often advantageous, a cleavable cross-linker can also have several disadvantages compared with other traditional cross-linkers. Particularly after cleavage, it requires further tandem mass spectrometry of the cross-linked peptide for sequencing applications. Sometimes these fragmentations are very ambiguous due to the cleavage of the attached cross-linker parts.

All of these cross-linkers have contributed tremendously to the cross-linking field, but confident data analysis is still a major hurdle. Two advances will be critically important to make cleavable cross-linking technology widely amenable for analyzing large-scale protein interactions: 1) the design of effective cleavable chemical cross-linkers with innovative features, which will help reduce the complexity of mass-spectrometry data of large-scale protein interactions; and 2) the development of robust and user-friendly software tools [20-24].

To develop a cutting-edge cross-linking technology that will overcome the bottlenecks of existing cross-linking strategies, we have designed a DUal Cleavable Cross-linking Technology (DUCCT), which will improve the confidence in identifying cross-linked peptides by mass spectrometry. At the core of this method is the dual-mass-spectrometry-cleavable cross-linker, which can be fragmented by two differential tandem mass spectrometric techniques. These two differential tandem mass-spectrometric fragmentations will produce different signatures in the mass spectra for the same cross-linked peptide. The two complementary fragmentation signatures can help identify that cross-linked peptide with high confidence. Further MS/MS of the cross-linked peptide, if necessary, will produce additional confidence in identification.

We have synthesized the DUCCT cross-linker with two differential cleavable bonds. The individual cleavable groups have been tested by several groups, but due to the complexity of synthesis, no cross-linkers have been constructed using both of these bonds [4,25]. In this report, we show our primary analysis of the effectiveness of this novel cross-linker. The technical simplicity and the quick screening capabilities of DUCCT will provide high confidence in data analysis. We believe this will significantly enhance confidence in the unambiguous identification of protein interactions in large-scale protein-to-protein cross-linking experiments.

2.2 Experimental Section

2.2.1 Materials

Neurotensin (pyr-LYNKPRRPYIL) was purchased from Anaspec, Inc. (San Jose, CA). Bovine serum albumin (BSA), ubiquitin, Tris-HCl, dimethyl sulfoxide (DMSO), ammonium bicarbonate, and formic acid were obtained from Sigma-Aldrich (St. Louis, MO). Liquid chromatography-mass spectrometry (LC-MS) grade methanol and acetonitrile were obtained from VWR (Radnor, PA). A reducing agent, dithiothreitol (Biorad, CA), an alkylating agent, iodoacetamide (Sigma-Aldrich, MO), and a protease, trypsin (Promega, Madison, WI) were utilized to digest the proteins properly. For synthesis of the cross-linker, Fmoc amino acid starting materials were obtained from EMD Millipore (Billerica, MA), formylbenzoic acid from Sigma-Aldrich (St. Luis, MO), Fmoc hydralink 6-Fmoc-HNA from Advanced Automated Peptide Protein Technology (Louisville, KY), and Dicyclohexylcarbodiimide (DCC) from Life Technologies (Pittsburgh, PA). 18 Milli-Q water was used for all the studies and was obtained from a water filtration system purchased from Aries Filterworks (West Berlin, NJ).

2.2.2 Cross-Linking of Neurotensin (Peptide)

The studied cross-linking agent (DUCCT) was prepared in DMSO. Neurotensin was treated with DUCCT in a 1:10 molar ratio in PBS buffer (pH 7.2). The reaction was permitted to continue for 30 mins at ambient temperature, after which the reaction was quenched with 50 mM Tris-HCl buffer (PH 8.0). The samples were desalted by ZipTip (Thermo Scientific, Waltham, MA), dried by speed vacuum, and finally reconstituted in 0.1% formic acid solution in H_2O. Cross-linked peptide samples were analyzed with a Linear Ion Trap (LTQ) Velos Pro mass spectrometer (Thermo Scientific, USA).

2.2.3 Cross-Linking of Ubiquitin and BSA

The cross-linking reaction of two proteins, ubiquitin and BSA, with the DUCCT cross-linking agent was carried out in a 1:50 molar ratio in PBS buffer (pH 7.2). The reaction was allowed to proceed for 30 mins and then quenched by 50 mM Tris-HCl buffer. A protein concentrator (3 kDa molecular weight cut-off, Thermo Scientific, USA) was used to remove excess cross-linking agent. The protein concentration was determined using a BCA protein assay. Next, the cross-linked proteins were digested, using both in-gel and in-solution methods with trypsin.

2.2.4 In-Solution Digestion

The cross-linked protein was reduced with 10 mM dithiothreitol (Biorad, CA), alkylated with 55 mM iodoacetamide (Sigma-Aldrich, MO), and then digested by trypsin (Promega, Madison, WI). The protein-trypsin ratio was set at 50:1, and the sample was incubated overnight at 37 °C. The tryptic digestion was quenched using 0.1% Formic Acid (FA) in $_{H2O}$. The sample was dried using speed vacuum, desalted by ZipTip, and re-suspended in 0.1% FA solution in $_{H2O}$ and finally transferred to an LC vial for mass spectrometric analysis.

2.2.5 In-Gel Digestion

The cross-linked protein was prepared with Laemmli buffer (Bio-Rad, OH), and heated for 5 min at 95 °C for denaturation. Next, the protein sample was loaded on a 10% SDS-PAGE gel. Separation of the proteins was observed according to their molecular weights. The gel bands were excised and digested after reduction and alkylation. After tryptic digestion, the peptides were extracted by 50% acetonitrile, dried by speed vacuum, and reconstituted in a solution of 0.1% FA in H_2O.

2.2.6 Instrumental Analysis

For sample analysis, we utilized an LTQ Velos Pro mass spectrometer coupled with a UHPLC (UltiMate 3000, Dionix, USA). Cross-linked peptides were separated by reverse-phase chromatography, using a nano-viper analytical C18 column (Acclaim™ Pep Map™ 100 C18 LC Columns, Thermo Scientific). Separation was performed with a binary gradient system, in which the organic and aqueous mobile phases contained 95% acetonitrile and 98% water, respectively. The nano-column flow rate and injection volume were set at 300 nl/min and 5 µl (partial injection mode), respectively.

For ionization, the nano-electrospray ionization (ESI) source was utilized with a fixed spray voltage and heated capillary temperature of 2.0 V and 275 °C, respectively. Full scan spectra (AGC 3×04) were obtained from 350 to 2000 m/z. Data-dependent MS/MS spectra (AGC 1×104) were collected from the five most abundant precursor ions. The dynamic exclusion time was fixed at 30 ms for separating consecutive ions. Data acquisition was set for 90 mins. XCalibur software was utilized for data processing.

In CID fragmentation mode, the activation energy was set to 45%, along with an isolation width of 1.5 Da, activation Q of 0.25, and activation time of 10 ms. For ETD operation mode, we set the emission current, reagent ion electron energy, reagent ion source CI pressure, and reagent ion source temperature to 50 µA, −70 V, 20 psi, and 110 °C, respectively. The ETD reaction time was fixed at 80 ms, with an isolation width of 2 Da. Direct infusion was also conducted in the LTQ Velos Pro instrument to analyze the cross-linked peptides.

2.2.7 Synthesis of the Cross-Linker

The cross-linker was synthesized using Fmoc peptide synthesis reagents with an Applied Biosystems 431 Peptide Synthesizer at the 0.25 mmol scale. Fmoc-Gly was first coupled to the resin, followed by the super-sensitive form of Fmoc-Asp, Fmoc-Asp(O-2-PhiPr) or Fmoc-Gly-Wang resin was purchased (Anaspec Inc), which was followed by the super-sensitive Fmoc-Asp. This was then followed by coupling Pro and then the typical form of Asp using Fmoc Asp-OtBu. The free amine of this Asp residue was coupled to the reagent Hydralink 6-Fmoc-HNA. After Fmoc was released, we added 4-formylbenzoic acid, which formed the hydrazone link and a terminal carboxylic acid. The protection group of the super-sensitive Asp was removed by 2% Trifluroaceticacid (TFA) in dichloromethane to create the second carboxylic acid. These two carboxylic acids were activated by DCC, to which N-hydroxy succinimide (NHS) esters were then formed by adding 2 mmol N-hydroxysuccinimide (Sigma-Aldrich, St. Louis, MO) with the DCC. The final compound was cleaved from the resin by 95% TFA with 5% H_2O for 1 h. The product was precipitated with diethyl ether and centrifuged. The pellet was dissolved in 0.1% TFA in H_2O and quickly lyophilized.

2.2.8 LPS-Biotin Pulldown Studies

Lipopolysaccharide (LPS) was purchased from Invivogen (San Diego, CA). Dr. Michael B. Fessler at NIEHS, NIH gifted us RAW 264.7 macrophage cells. The cells were grown in Dulbecco's Modified Eagle Medium (DMEM) media (500 mL) with Penicillin-Streptomycin (Pen-Strep) and 10% Fetal Bovine Serum (FBS) at 37° C in a Forma incubator (Thermo Scientific). After 80% confluency, the cells were scraped and collected in a 15 mL centrifuge tube with DMEM media. 25 μg of LPS-biotin was added in the media (10 mL) and the cells were incubated at 37 °C for 15 min with mild stirring.

After 30 min, two quick 500 μL PBS washes were done to remove the LPS. 10 mL of PBS buffer was then added in the samples, followed by 1 mg of bis(sulfosuccinimidyl)suberate (BS3 croslinker to this solution. DUCCT cross-linker was added in the same amount (1 mg dissolved in 5 μL of DMSO). The solution was incubated for 30 min with mild stirring at room temp. After 30 minutes, 50 mM of 10 μL Tris-HCl was used to stop the reaction. Next, we washed the cells several times with PBS buffer. The cells were lysed with RIPA buffer (1h) and centrifuged to collect the proteins. After that, the proteins were incubated with 50 μL dynabeads streptavidin (Thermo Scientific). After several washes with PBS in a magnet stand, the proteins were eluted from the beads using 2X Laemmli buffer. SDS-PAGE gels were run and stained with SYPRO® ruby (Thermo Scientific, Pierce). The gel lanes were cut, digested (24 sections for each lane), and analyzed by LC-MS/MS in an Agilent XCT ETD mass spectrometer with CID and ETD-MS/MS fragmentation methods. A heat map using protein spectral counts was generated (Figure A-S11). All MS/MS samples were analyzed using Spectrum Mill (Agilent, Santa Clara, CA; version unknown). Spectrum Mill was set to search the NCBInr.rodent database (selected for All, unknown version, 14227560 entries) using the digestion enzyme trypsin. Spectrum Mill was searched with a fragment ion mass tolerance of 0.70 Da and a parent ion tolerance of 2.5 Da. Scaffold (version Scaffold_4.3.2, Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at a greater than 80.0% probability by the Peptide Prophet algorithm. Protein identifications were accepted if they could be established at a greater than 95.0% probability and contained at least one identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm [26].

2.3 Results and Discussion

2.3.1 Cross-Linker Design and Synthesis

In Figure 2-1A, we show the design of the novel cross-linker. The cross-linker contains two gas-phase-cleavable bonds, which can be selectively cleaved using collision-based or electron transfer-based dissociations. The cross-linker was synthesized using Fmoc peptide synthesis reagents (details are provided in the Experimental Section). The reactive groups used were NHS esters for conjugation of lysine residues.

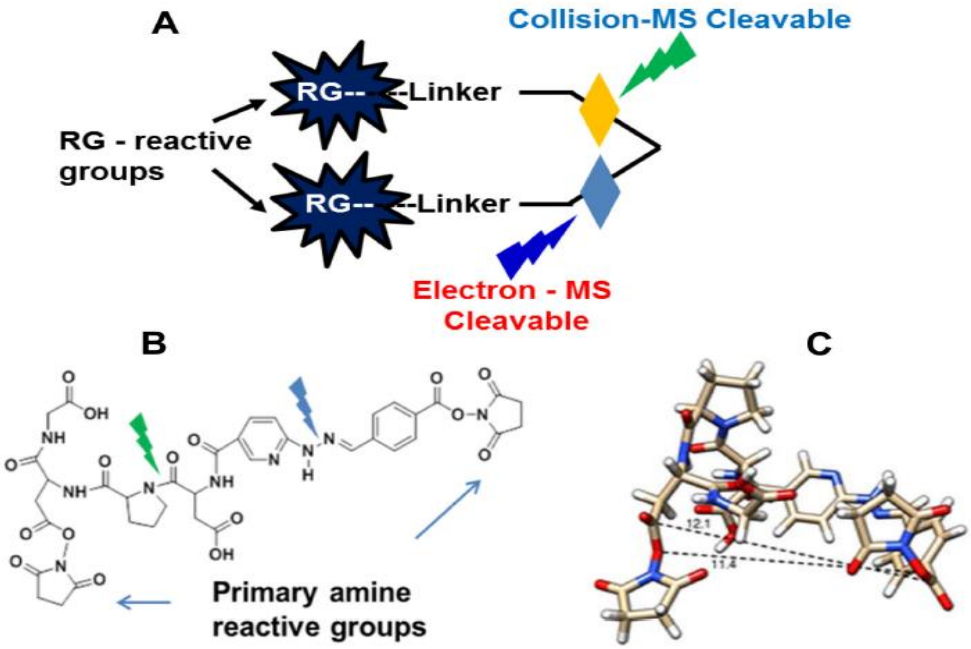


Figure 2-1. The design and chemical structure of the DUCCT cross-linker. (A) The design of the dual CID and ETD cleavable cross-linker. (B) The chemical structure of the novel CID and ETD MS-cleavable cross-linker. The green and blue symbols indicate the CID (DP bond) and ETD (hydrazone) MS-cleavable sites, respectively. (C) The structure of DUCCT is shown with spacer chain length.

Two gas-phase-cleavable bonds were added to the design. One is the Asp- Pro (DP) bond, which has been reported to be cleaved by low-energy CID.4 We added another bond (hydrazone), which researchers have reported can be cleaved using ETD.[25] Both of these bonds were sandwiched in between two reactive NHS ester groups, which altogether constituted a lysine-reactive cross-linker (Figure 2-1B).

The distance of the reactive group is a major concern for MS-cleavable cross-linkers. Most of the cleavable cross-linker reactive distances are very large, and this has been a major concern for non-specific labeling. CLIP, a clickable cross-linker, has reactive distances around 10Å, but has been reported to label across a distance as high as 22 Å [2]. However, for the DUCCT cross-linker, the distance between reactive groups was found to be ~12.1 Å (Figure 2-1C), which we believe even with extended conformation will be very suitable for cross-linking.

2.3.2 Cross-Linked Peptide Identification Strategies

In Figure 2-2, we illustrate the strategies for the identification of cross-linked peptides using the DUCCT cross-linker. After CID-MS/MS, an inter-cross-linked peptide will produce two peptide peaks attached with cross-linker pieces, whereas ETD will produce two peaks for the same cross-linked peptides with different cross-linker parts (Figure 2-2, and Figures A-S1-S2 in the Supporting Information). In addition, ETD will produce charge-reduced ions of the precursor masses. This will also help to identify the charge states of the precursor m/z, even if a low-resolution mass spectrometer is used. These two signature spectra will unambiguously identify cross-linked peptides. Furthermore, dead-end peptides will generate one peptide with residual cross-linker mass due to the hydrolysis of one reactive group in the cross-linker (Figure 2-2, bottom). Their CID and ETD mass spectra will generate two complementary fragmentation signatures of one cross-linked peptide (Figure A-S3).

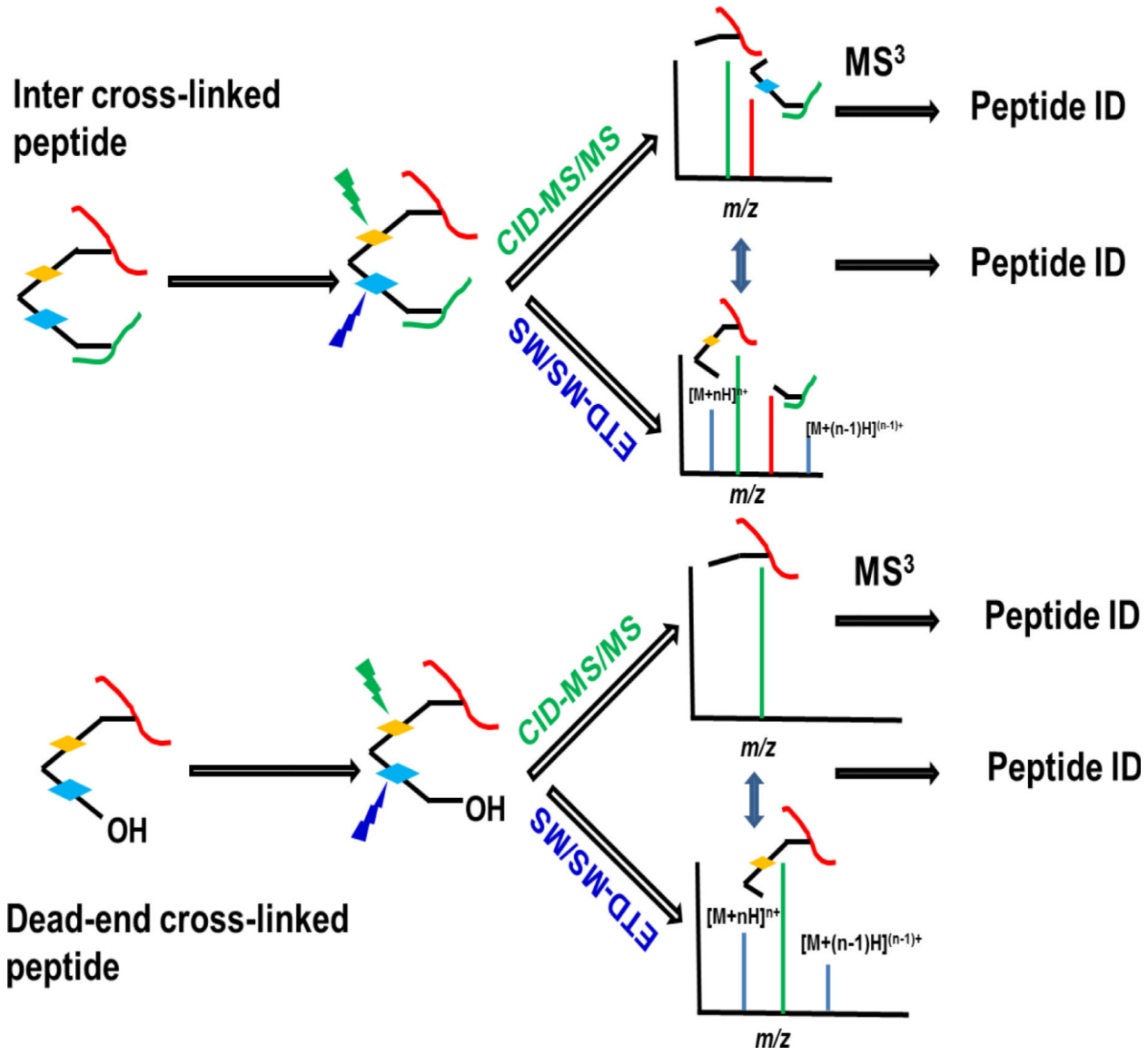


Figure 2-2. The scheme to identify inter-cross-linked (top) and dead-end peptides (bottom) by CID and ETD tandem mass spectrometry. MS3 can be used for CID-only identification. Both CID and ETD will identify peptide sequence without further MS/MS.

2.3.3 Proof-of Concept Studies with DUCCT Cross-Linker

Initial characterization of the cross-linker molecular weight (863.2358 Da) and the cleavage of the DP peptide bond is provided in the Supplementary Information (Figure A-S4). It is important to note that no HPLC purification was conducted for this cross-linker and that high-resolution MS and MS/MS were performed to confirm the structure. Another important issue to note is that the cross-linker will not be cleaved by ETD without conjugation with a peptide, since ETD induces charge-dependent fragmentation.

In validation tests, we next treated a standard peptide, neurotensin, with our cross-linker. Neurotensin has one lysine residue and its N-terminus is blocked with a pyro-Glu modification. The cross-linker labeled the peptide efficiently and produced inter and dead-end cross-linked peptides. Two inter-cross-linked peptides were identified at m/z's of 995.2577 (M+4H+) and 796.4076 (M+5H+), respectively. CID-MS/MS clearly showed cleavage at the DP peptide bond, and ETD-MS/MS showed cleavage at the N-N bonds. In CID, we should observe two peptide ions with added cross-linker residues, and in ETD for inter-cross-linking we should observe the same two peptide ions with different cross-linker residue masses. For neurotensin, we observed two fragment peaks at m/z 971.76 and 1019.25, respectively, which corresponds to the peptide mass and the corresponding CID-cleaved part of the cross-linker (Figure 2-3A). We also observed 3+ charge states of these peptides (m/z at 679.98 and 648.32). Complete calculations of cross-linked precursors and fragment ions are shown in the Supplementary Information (Table S1 and Figure A-S5). ETD clearly matched the calculated masses with the cross-linker residues. Charge-reduced precursor ions were also observed with ETD, which clearly provides proof of the charge states of the cross-linked peptides, even though a high-resolution mass spectrometer was not used for this study (Figure 2-3B).

This data clearly shows the CID and ETD signatures for an identically cross-linked peptide. These two mass spectral signatures of the same cross-linked peptide provide a means for the high-confidence identification of this material.

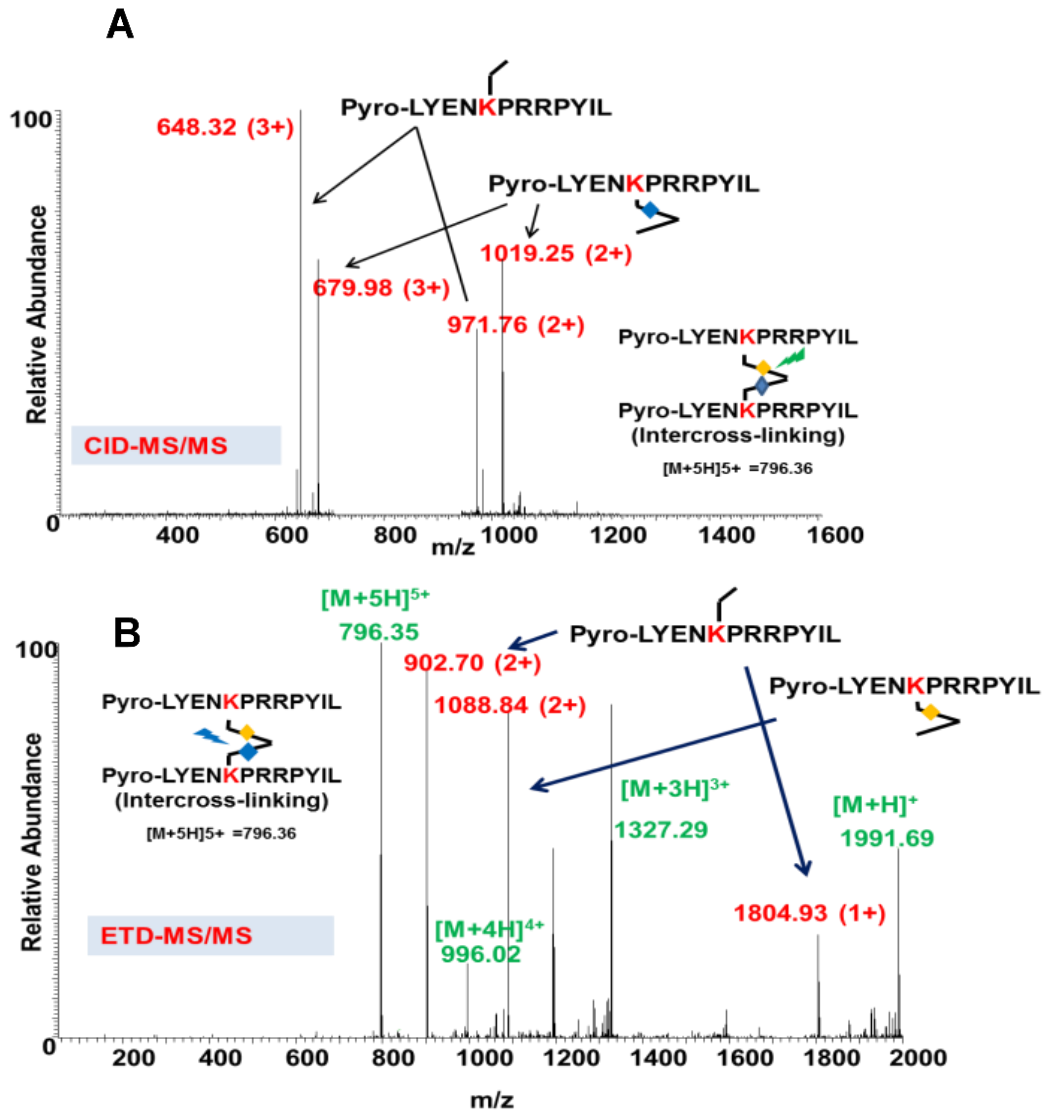


Figure 2-3. CID and ETD mass spectra of a cross-linked neurotensin peptide dimer. (A) CID and (B) ETD fragmentation methods show different mass spectra signatures for the same peptide. Please see the Supplementary Information for the added masses from the cross-linker due to differential cleavages.

Next, we tested the labeling efficiency of our cross-linker using a small protein, ubiquitin, and a large-protein, BSA. Both studies showed efficient labeling with higher molecular weight dimer bands in the SDS-PAGE gel (Figure A-S6). After in-solution digestion and subsequent identification, we have clearly identified the proteins' cross-linked sites in the CID and ETD MS/MS mass spectra. Ubiquitin is a small protein (~8500 Da), and is a very useful model protein for monitoring cross-linking, as lysine 48 of ubiquitin will form a cross-link with lysine 63, as has been reported by several groups [2,27]. After cross-linking with our DUCCT compound, we searched the mass spectral data with a newly developed software tool (the development of this software is being described in a separate manuscript in preparation). Even a small protein like ubiquitin can generate ~20,000 spectra in LC-MS/MS. We have identified several cross-linked peptides from ubiquitin, and using careful investigation, have clearly identified the CID and ETD spectra of cross-linked lysine 48 and 63 (Figure 2-4). Calculation of the fragment mass is provided in Table S2. It is clear that CID produces the expected fragments (peptide mass + cross-linked fragment mass-H) after selective cleavage, and ETD produces similar cleavages in specified sites (Figure A-S7). Charge-reduced precursor ions were also confirmed by the charge states of the precursor, even though a low-resolution mass spectrometer was used for this study. It is interesting to note that we have found that CID cleavages were efficient in both odd- and even-charge state precursor ions, but ETD-cleavages were observed more efficiently in odd-charge state precursor ions. The DUCCT cross-linker can also be applied using just the CID cleavage parts. Further MS/MS of the fragment masses can be utilized to sequence the cross-linked peptides (Figure A-S8A). The MS3rd of the fragment peaks also unambiguously confirmed the sequence of the peptides in ubiquitin and neurotensin (Figure A-S8A-B).

Dead-end cross-linking was identified very efficiently in the neurotensin cross-linked peptides. Dead-end peptides are very good indicators of the surface and reactive lysine residues.

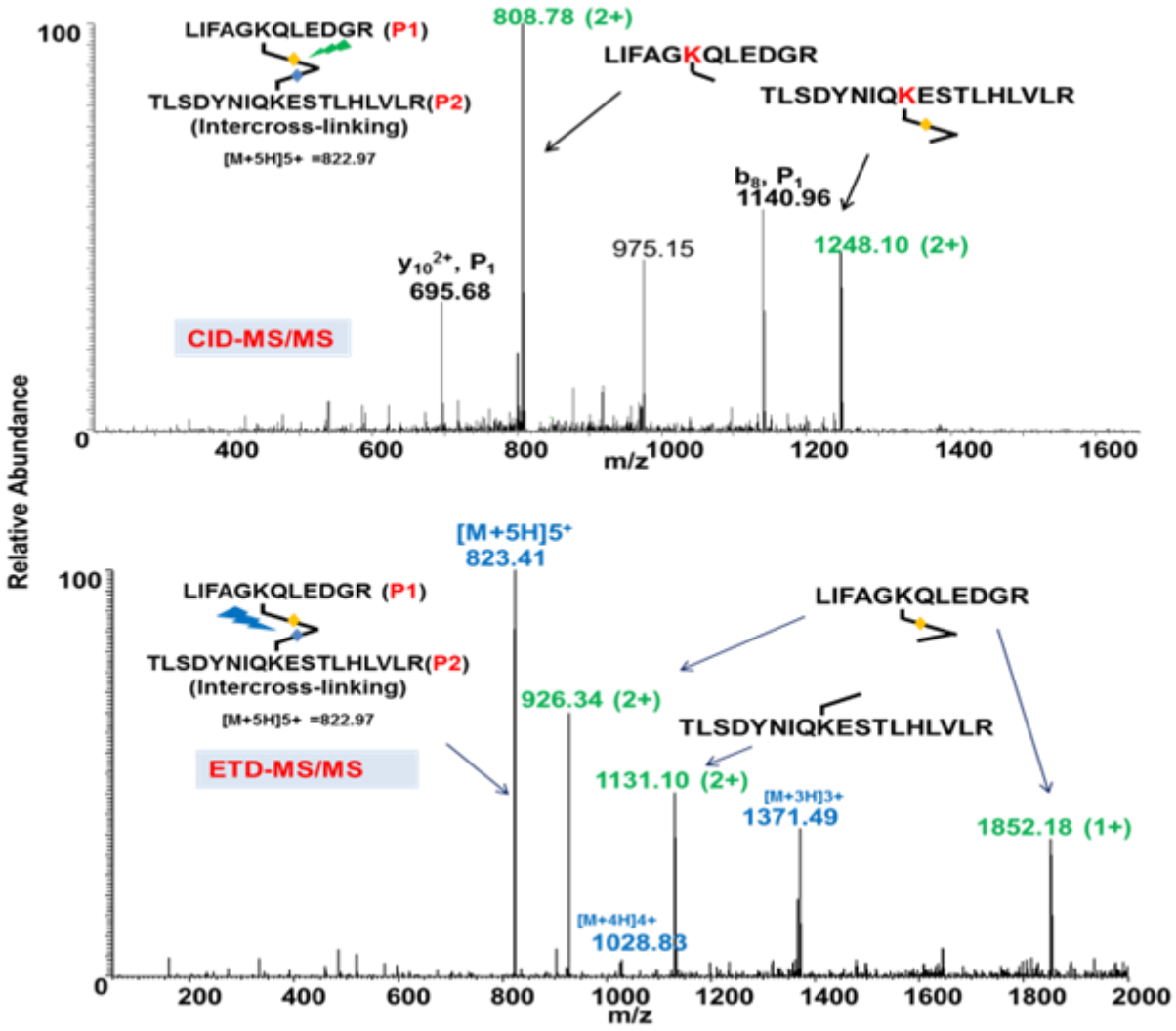


Figure 2-4. An example of CID and ETD mass spectra of an inter-cross-linked peptide derived from ubiquitin. The CID and ETD mass spectra show different signatures for the same cross-linked peptide.

Identifying dead-ends with the cleavable cross-linker requires further MS/MS of the cross-linked peptides, due to the generation of a single peptide ion in the mass spectrum. For an asymmetrical cleavable cross-linker, dead-end peptides can react in two different ways depending on the reactive ends. The majority of the mass spectra will be one peptide, but some of them can be a mixture of two peptides. Fortunately, the cross-linked precursor mass database of dead-end peptides will clearly filter these out during the first stage of searching the database.

Since dead-end cross-linked peptides can produce a peptide with added cross-linker masses, sometimes it can produce both peptides due to uneven fragmentation on both cleavage sites as a result of the labeling positions (Figure A-S3). This problem can easily be solved by the dual-cleavable properties of the DUCCT cross-linker. Both CID and ETD MS/MS will unambiguously identify dead-end peptides. A dead-end precursor peptide was calculated and found at m/z 775 (3+) for neurotensin. After CID-MS/MS, it was cleaved at the DP bond and generated a peptide with the added cross-linker masses at m/z 1019.01 (Figure A-S9). However, it produced one peptide with ETD cleavage masses (m/z = 1088.81, calculation on page S5 of the Supporting Information), as well as charge-reduced ions. In the ubiquitin cross-linking experiments, we observed several dead-end peptides. One of the dead-end peptides is shown in Figure 2-5A single peptide was identified in the mass spectrum after CID cleavage. This peptide requires further MS/MS/MS for confident identification. Due to the dual-cleavable properties, further ETD MS/MS produced the same peptides with added cross-linker parts. Charge-reduced precursor ions were also observed and provided the charge states of this precursor. It is clear that CID and ETD pinpointed the same cross-linked peptide with high confidence. However, the cleavable cross-linker is not suitable to analyze intra-peptide cross-links. Hence, we omitted identification of these peptides in our studies.

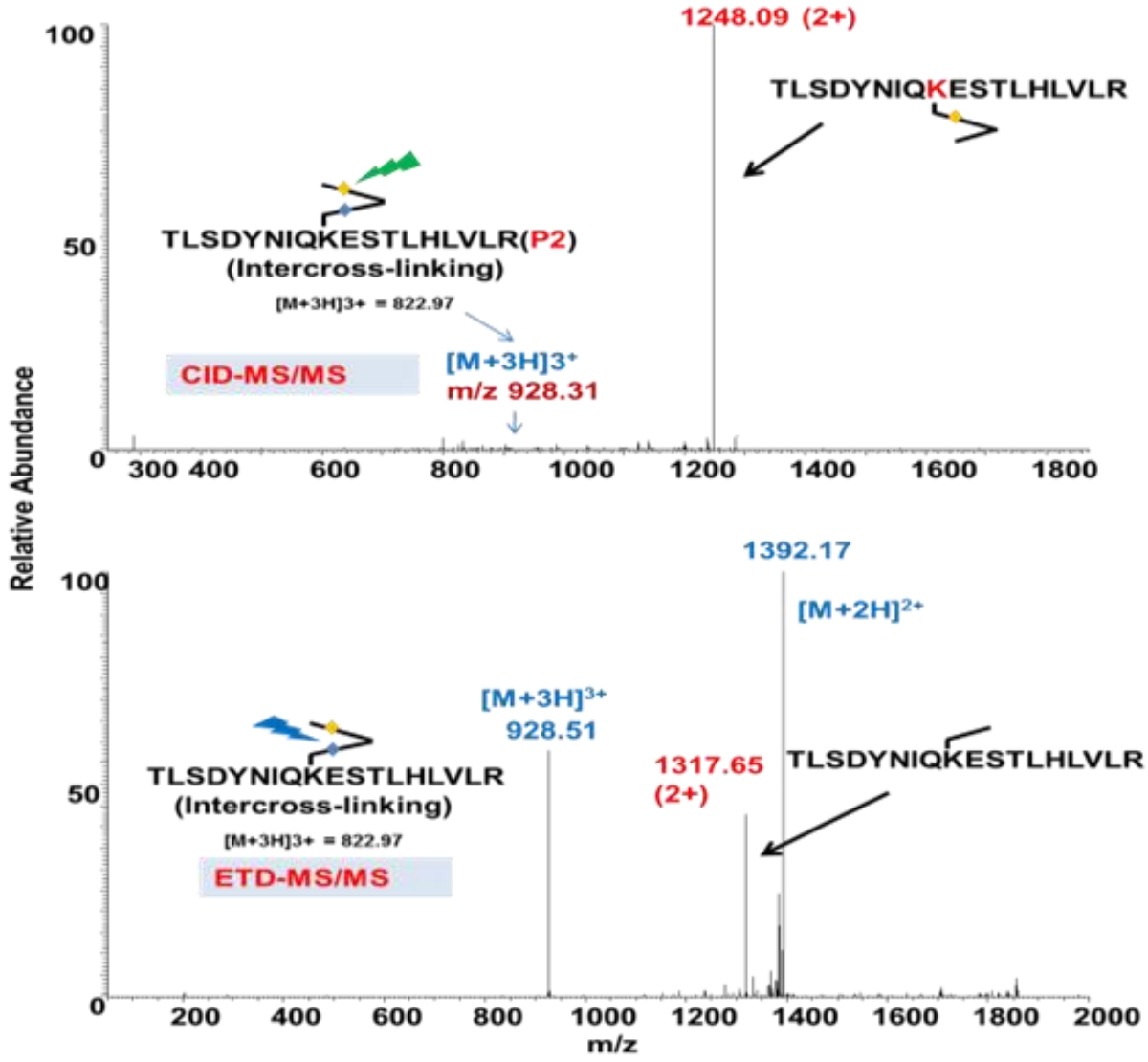


Figure 2-5. Examples of CID and ETD mass spectra of a dead-end cross-linked peptide derived from ubiquitin. The CID and ETD mass spectra show different signatures for the same cross-linked peptide.

2.3.4 Confidence in Cross-Linked Peptide Identifications

We developed an automated software tool to analyze this data. Even small proteins in a LC-MS/MS experiment generate ~20,000 mass spectra. To test how confidently we can identify cross-linked peptides by two differential cleavages, we created a theoretical database of cross-linked peptides using precursor and fragment masses. We then created an mgf file of an LC-MS/MS experiment of cross-linked ubiquitin, which contained an experimental list of precursors and fragment ion masses. We found approximately 30,500 scan ions in the mgf file. At first the software tool searched the precursor mass spectra database (2.5 Da), then matched spectra were further searched for fragment peaks within a defined mass threshold (0.6 Da). We chose these mass thresholds due to the use of a low-resolution mass spectrometer in this study. Cross searching our experimental (mgf) file with the theoretical cross-linked database, we found a number of hits in the CID search. It will require further MS/MS of both cross-linked peptides to establish full confidence in the identification of these materials (see the excel files in the Supporting Information). We also searched the mgf file for ETD-MS/MS of the cross-linked ubiquitin using the same process. The search found several hits from the theoretical ETD dataset. After comparing the common sequences identified in both datasets with precursor m/z restricted to ±0.5 Da, we confidently identified several peptide sequences from both datasets (Table S3, and Supporting Information excel files). We manually verified each result and found they efficiently matched our theoretical and experimental calculations. We are currently working on large datasets and developing scoring tools for analyzing the data with proper statistics. The software tool will be published in an appropriate journal soon. It is very clear that CID and ETD results pinpointed with confidence the cross-linked peptides in ubiquitin (see the crystal structure in Figure A-S10 for identified sites with distance) from these CID- and ETD-MS/MS experiments. We have also demonstrated MS/MS of these cleavage parts and confirmed some of

the peptide sequences with high confidence. These three steps: 1) CID-MS/MS of the cross-linked peptides; 2) ETD-MS/MS of the cross-linked peptides; and an 3) optional MS3rd of the cleaved cross-linked peptide parts, can simultaneously identify the cross-linked peptides without ambiguity. It is also important to note that a high-resolution mass spectrometer will significantly reduce the number of hits due to the search threshold of precursor and fragment masses.

2.3.5 Labeling Studies in Cell Lysate

In order to test the labeling efficiencies of DUCCT in in vivo pulldown samples, we tested our cross-linker in macrophages after stimulation with the Toll-Like Receptor 4 (TLR4) ligand LPS. RAW 264.7 macrophage cells were grown and treated for 1 h with LPS-biotin, followed by cross-linking with DUCCT or BS3, a commercial cross-linker. After SDS-PAGE separation of the samples, in-gel digestion was carried out for each gel lane (24 pieces), then we performed the LC-MS/MS experiments (see the Supporting Information for total protein identification). The DUCCT cross-linker-treated sample yielded more protein identifications compared to BS3 (Figure 2-6). We also confirmed some known protein interactions in cross-linked samples that had been treated with LPS-biotin, providing feasibilities of our method in large-scale cell signaling studies. CD14, a known LPS-binding protein that associates with the TLR4 receptor, was identified with high spectral counts using DUCCT compared to BS3 [28]. No CD14 peptides were identified in the control LPS-biotin treated sample (Figure 2-6, and Figures A-S11-S12). Our studies clearly showed that labeling efficiencies of our cross-linker are comparable to the widely used commercial cross-linker BS3. We are currently working on software that will allow us to search large-scale samples (> 500 proteins using our CID and ETD cleavage technique with appropriate statistics) with high confidence.

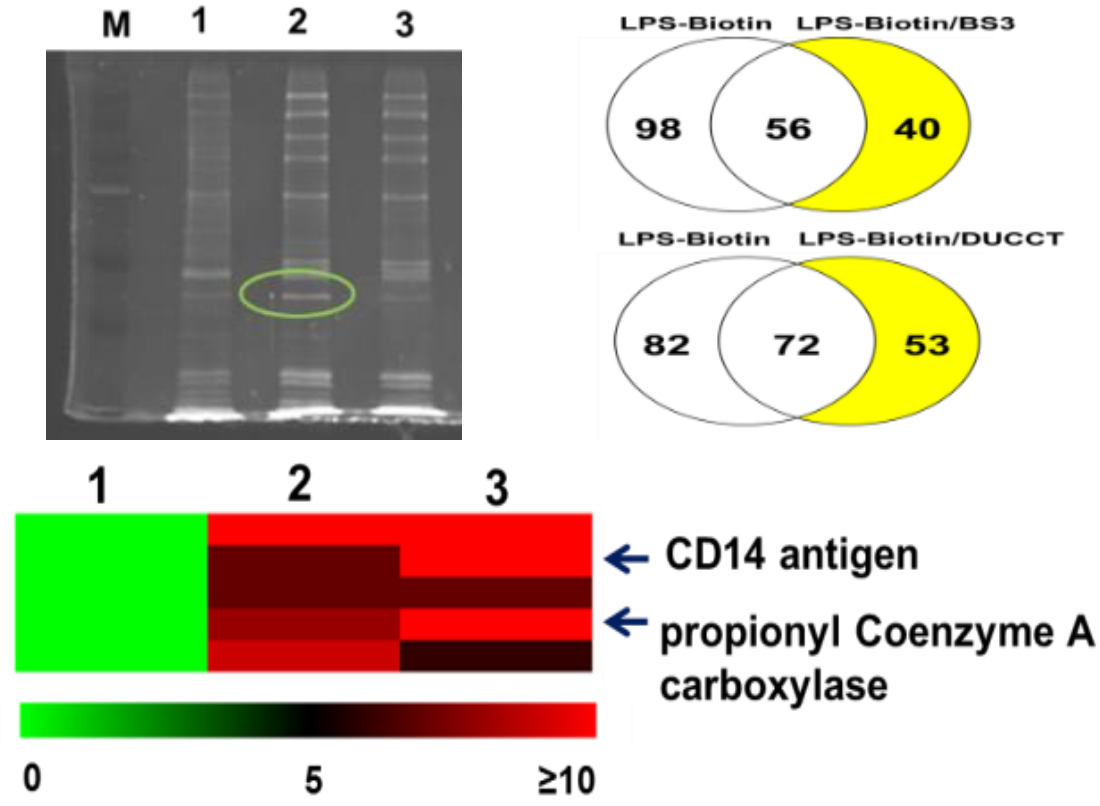


Figure 2-6. (Upper left) The SDS-PAGE gel of biotin-avidin pulldown experiments (M = marker, 1 = LPS biotin, 2 = LPS-biotin/DUCCT, and 3 = LPS-biotin/BS3). (Upper right) A Venn diagram of the number of proteins identified. (Bottom) A partial heat map of selected proteins identified exclusively in cross-linked samples after pulldown studies with avidin (see the full heat map in Figures S11, S12). The scale denotes the spectral counts. For the heat map, 1 = LPS biotin, 2 = LPS-biotin/BS3, and 3 = LPS-biotin/DUCCT.

2.4 Conclusions

We show here a novel cross-linker, DUCCT, which has two differential tandem mass spectrometry cleavage properties. Differential cleavage can pinpoint the same cross-linked peptides with high confidence. This cross-linker can also work as a CID-cleavable cross-linker and subsequent MS/MS/MS can confirm the peptide sequence. We have also developed a software tool and are currently improving it for large-scale data analysis.

Together with software capability and the DUCCT cross-linker's dual-gas-phase-cleavable properties, we demonstrate that unambiguous identification of protein interactions is feasible. We believe this cross-linking-based proteomics technology will tremendously advance the system-level identification of cell-signaling cascades. DUCCT cross-linking should allow investigators to quickly and efficiently screen interacting partners in discovery proteomics, providing lists of interacting proteins for further biochemical and/or molecular biological validation.

2.5 References

1. Gingras, A.-C., Gstaiger, M., Raught, B., Aebersold, R.: Analysis of protein complexes using mass spectrometry. Nat. Rev. Mol Cell Biol. 8, 645 (2007)
2. Chowdhury, S.M., Du, X., Tolic, N., Wu, S., Moore, R.J., Mayer, M.U., Smith, R.D., Adkins, J.N.: Identification of cross-linked peptides after click-based enrichment using sequential collision-induced dissociation and electron transfer dissociation tandem mass spectrometry. Anal Chem. 81, 5524-5532 (2009)
3. Petrotchenko, E.V., Serpa, J.J., Borchers, C.H.: An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics. Mol Cell Proteomics. 10, M110. 001420 (2011)
4. Soderblom, E.J., Goshe, M.B.: Collision-induced dissociative chemical cross-linking reagents and methodology: applications to protein structural characterization using tandem mass spectrometry analysis. Anal Chem. 78, 8059-8068 (2006)
5. Tang, X., Munske, G.R., Siems, W.F., Bruce, J.E.: Mass spectrometry identifiable cross-linking strategy for studying protein− protein interactions. Anal Chem. 77, 311-318 (2005)
6. Greber, B.J., Boehringer, D., Leibundgut, M., Bieri, P., Leitner, A., Schmitz, N., Aebersold, R., Ban, N.: The complete structure of the large subunit of the mammalian mitochondrial ribosome. Nature. 515, 283 (2014)
7. Tan, D., Li, Q., Zhang, M.-J., Liu, C., Ma, C., Zhang, P., Ding, Y.-H., Fan, S.-B., Tao, L., Yang, B.: Trifunctional cross-linker for mapping protein-protein interaction networks and comparing protein conformational states. Elife. 5, e12509 (2016)
8. Yu, C., Yang, Y., Wang, X., Guan, S., Fang, L., Liu, F., Walters, K.J., Kaiser, P., Huang, L.: Characterization of dynamic UbR-proteasome subcomplexes by in vivo cross-linking (X) assisted bimolecular tandem affinity purification (XBAP) and label-free quantitation. Mol Cell Proteomics. 15, 2279-2292 (2016)

9. Greber, B.J., Bieri, P., Leibundgut, M., Leitner, A., Aebersold, R., Boehringer, D., Ban, N.: The complete structure of the 55S mammalian mitochondrial ribosome. Science. 348, 303-308 (2015)
10. Lauber, M.A., Rappsilber, J., Reilly, J.P.: Dynamics of ribosomal protein S1 on a bacterial ribosome with cross-linking and mass spectrometry. Mol Cell Proteomics. 11, 1965-1976 (2012)
11. Arlt, C., Götze, M., Ihling, C.H., Hage, C., Schäfer, M., Sinz, A.: Integrated workflow for structural proteomics studies based on cross-linking/mass spectrometry with an MS/MS cleavable cross-linker. Anal Chem. 88, 7930-7937 (2016)
12.Gutierrez, C.B., Yu, C., Novitsky, E.J., Huszagh, A.S., Rychnovsky, S.D., Huang, L.: Developing an Acidic Residue Reactive and Sulfoxide-Containing MS-Cleavable Homobifunctional Cross-Linker for Probing Protein–Protein Interactions. Anal Chem. 88, 8315-8322 (2016)
13. Chowdhury, S.M., Munske, G.R., Tang, X.T., Bruce, J.E.: Collisionally activated dissociation and electron capture dissociation of several mass spectrometry-identifiable chemical cross-linkers. Anal Chem. 78, 8183-8193 (2006)
14. Chavez, J.D., Cilia, M., Weisbrod, C.R., Ju, H.-J., Eng, J.K., Gray, S.M., Bruce, J.E.: Cross-linking measurements of the Potato leafroll virus reveal protein interaction topologies required for virion stability, aphid transmission, and virus–plant interactions. J Proteome Res. 11, 2968-2981 (2012)
15. DeBlasio, S.L., Chavez, J.D., Alexander, M.M., Ramsey, J., Eng, J.K., Mahoney, J., Gray, S.M., Bruce, J.E., Cilia, M.: Visualization of host-polerovirus interaction topologies using protein interaction reporter technology. J Virol. 90, 1973-1987 (2016)
16. Yu, C., Yang, Y., Wang, X., Guan, S., Fang, L., Liu, F., Walters, K.J., Kaiser, P., Huang, L.: Characterization of dynamic UbR-proteasome subcomplexes by in vivo cross-linking (X) assisted bimolecular tandem affinity purification (XBAP) and label-free quantitation. Mol Cell Proteomics. 15, 2279-2292 (2016)
17. Kaake, R.M., Wang, X., Burke, A., Yu, C., Kandur, W., Yang, Y., Novtisky, E.J., Second, T., Duan, J., Kao, A.: A new in vivo cross-linking mass spectrometry platform to define protein–protein interactions in living cells. Mol Cell Proteomics. 13, 3533-3543 (2014)
18. Liu, F., Goshe, M.B.: Combinatorial electrostatic collision-induced dissociative chemical cross-linking reagents for probing protein surface topology. Anal Chem. 82, 6215-6223 (2010)
19. Makepeace, K.A., Serpa, J.J., Petrotchenko, E.V., Borchers, C.H.: Comprehensive identification of disulfide bonds using non-specific proteinase K digestion and CID-cleavable crosslinking analysis methodology for Orbitrap LC/ESI-MS/MS data. Methods. 89, 74-78 (2015)
20. Liu, F., Rijkers, D.T., Post, H., Heck, A.J.: Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. Nat Methods. 12, 1179 (2015)
21. Du, X., Chowdhury, S.M., Manes, N.P., Wu, S., Mayer, M.U., Adkins, J.N., Anderson, G.A., Smith, R.D.: Xlink-identifier: an automated data analysis platform for confident identifications of chemically cross-linked peptides using tandem mass spectrometry. J Proteome Res. 10, 923-931 (2011)

22.Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L.N., Beck, M., Schmidt, A., Mueller, M., Aebersold, R.: Identification of cross-linked peptides from large sequence databases. Nat Methods. 5, 315 (2008)
23. Yang, B., Wu, Y.-J., Zhu, M., Fan, S.-B., Lin, J., Zhang, K., Li, S., Chi, H., Li, Y.-X., Chen, H.-F.: Identification of cross-linked peptides from complex samples. Nat Methods. 9, 904 (2012)
24. Liu, F., Wu, C., Sweedler, J.V., Goshe, M.B.: An enhanced protein crosslink identification strategy using CID-cleavable chemical crosslinkers and LC/MSn analysis. Proteomics. 12, 401-405 (2012)
25. Gardner, M.W., Brodbelt, J.S.: Preferential Cleavage of N− N Hydrazone Bonds for Sequencing Bis-arylhydrazone Conjugated Peptides by Electron Transfer Dissociation. Anal Chem. 82, 5751-5759 (2010)
26. Nesvizhskii, A.I., Keller, A., Kolker, E., Aebersold, R.: A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. 75, 4646-4658 (2003)
27. Gardner, M.W., Vasicek, L.A., Shabbir, S., Anslyn, E.V., Brodbelt, J.S.: Chromogenic cross-linker for the characterization of protein structure by infrared multiphoton dissociation mass spectrometry. Anal Chem. 80, 4807-4819 (2008)
28. Rajaiah, R., Perkins, D.J., Ireland, D.D., Vogel, S.N.: CD14 dependence of TLR4 endocytosis and TRIF signaling displays ligand specificity and is dissociable in endotoxin tolerance. Proc Natl Acad Sci. U. S. A. 112, 8391-8396 (2015)

Chapter 3

High confidence identification of cross-linked peptides by an enrichment-based dual cleavable cross-linking technology (DUCCT) and data analysis tool Cleave-XL

Abstract

Cleavable cross-linking technology requires further MS/MS of the cleavable fragments for unambiguous identification of cross-linked peptides. These spectra are sometimes very ambiguous due to the sensitivity and complex fragmentation pattern of the peptides with the cross-linked residues. We recently reported a dual cleavable cross-linking technology (DUCCT), which can enhance the confidence in the identification of cross-linked peptides. The heart of this strategy is a novel dual mass spectrometry cleavable cross linker that can be cleaved preferentially by two differential tandem mass spectrometry methods, collision induced dissociation and electron transfer dissociation (CID and ETD). Different signature ions from two different mass spectra for the same cross-linked peptide helped identify the cross-linked peptides with high confidence. In this study, we developed an enrichment-based photocleavable DUCCT (PC-DUCCT-Biotin), where cross-linked products were enriched from biological samples using affinity purification and subsequently two sequential tandem (CID and ETD) mass spectrometry processes were utilized. Furthermore, we developed a prototype software called Cleave-XL to analyze cross-linked products generated by DUCCT. Photocleavable DUCCT was demonstrated in standard peptides and proteins. Efficiency of the software tools to search and compare CID and ETD data of photocleavable DUCCT biotin in standard peptides and proteins as well as regular DUCCT in protein complexes from immune cells were tested. The software is efficient in pinpointing cross-linked sites using CID and ETD cross-linking data.

We believe this new DUCCT and associated software tool Cleave-XL will advance high confidence identification of protein cross-linking sites and automated identification of low-resolution protein structures.

3.1 Introduction

The knowledge of protein to protein interactions and protein structures are important to understand how proteins are functioning in the cell [1]. Chemical cross-linking coupled with mass spectrometry is coming out as one of the robust strategies to characterize protein structures and their interactions [2-4]. Two major obstacles still make this technology more challenging for complex sample analysis [5]. First, relatively low quantity of most informative inter cross-linked products is generated compared to unmodified peptides, and the second one is computational challenges in cross-linked peptides data analysis [6]. The unambiguous confirmation of cross-linked products in proteins and protein complexes remains challenging. Innovative cross-linking strategies and user-friendly data analysis tools are required to overcome these hurdles.

The favorite technique to assist the identification of cross-linked products is the utilization of cleavable cross-linkers, where gas-phase cleavable-bonds are incorporated in the cross-linkers design for selective fragmentation. One of the most widely used cleavable cross-linker is Protein Interaction Reporters (PIRs), which was developed by Bruce and co-workers utilizing MS labile bonds that generates signature fragments after applying low energy collision induced dissociation (CID) [7, 8]. Goshe and co-workers designed MS-cleavable cross-linker utilizing Asp-Pro (DP) bond which efficiently cleaved by low energy in-source CID [9, 10]. CID-cleavable urea-based cross-linker was designed by Sinz and co-workers and the cross-linked peptides were identified by their MeroX software [11-13]. The Huang group designed a series of cross-linkers containing CID-cleavable sulfoxide bonds [3, 14, 15].

Although, CID- cleavable cross-linker has great advantages in this field, they have some limitations too. Unfortunately, straightforward use of the fragment ion masses is sometimes not enough for unambiguous detection of cross-linked peptides. Particularly, it requires further tandem mass spectrometry for the sequence identifications of fragment ions, which is not very sensitive and straightforward for all cross-linked peptides. Nevertheless, these are the most efficient techniques for analyzing cross-linked peptides by mass spectrometry.

Numerous data-analysis strategies have been developed based on bottom up proteomics and mass spectrometry using non-cleavable and cleavable cross-linkers. One of the common strategies is generating a modified database of cross-linked peptides utilizing protein sequence and cross-linker mass and identification can be accomplished by comparing the experimental spectra with the speculative data set of each candidate, and finally sorting out the candidates utilizing statistical confidence. Several data analysis software tools have been developed to analyze cross-linked data, which includes ASAP, X!Link, MS-Bridge, Virtual MSLab, X-Link, FindLink, StavroX, and pLink [16-21]. Another strategy to identify the cross-linked products is the use of stable isotope labeling. Isotope labeling generates a signature mass shift detected by mass spectrometry. Pro-cross-link, iXlink, doXlink, xQuest software tools are developed to analyze the resultant data [22-24]. The major limitation of this strategy is the fragmentation complexity. In addition, the combinatorial increase of the search space of cross-linked peptides makes the strategy more unfriendly for large-scale analysis. The number of possible inter peptide cross-links to be searched increases in proportion to n^2, where n is the number of peptides. In consequence, the complexity in data analysis is tremendous for protein complexes or whole cell lysate protein samples. Some limited data analysis capabilities are now available with XlinkX and MeroX software [25-27].

XlinkX and MeroX, both retrieve the precursor mass using the unique mass differences of MS2 spectra of cross-linkers, thus reducing the search space from n^2 to 2n. The software is very selective for a particular cross-linker reporting most of the identification observed for the HeLa cell lysate to be in the domain of high abundance proteins thereby emphasizing the need for affinity tagged cross-linkers [28].

To address the confidence in data analysis, our group recently developed a Dual Cleavable Cross-linking Technology (DUCCT). For the first time, we were able to incorporate two differential gas phase mass spectrometry labile bonds in one cross-linker. two tandem mass spectrometric techniques named as collision induced dissociation (CID) and electron transfer dissociation (ETD) were utilized to cleave those labile bonds and generating two different kinds of signatures from the same cross-linked precursors [2]. This provided high confidence and quick screening capability of protein cross-linking sites. Further data reduction is possible by incorporating an affinity group in the DUCCT cross-linkers.

In this paper, we demonstrated an enrichment-based DUCCT cross-linker and introduced a cross-linking software tool, called Cleave-XL. The software is specially designed to identify the cross-linked peptides automatically from the DUCCT cross-linker treated mass-spectrometric datasets. The next generation DUCCT cross-linker, PC-DUCCT-Biotin has two additional features, an affinity tag (biotin) and a photo cleavable group. The affinity tag enabled enriching of cross-linked peptides by avidin-biotin affinity chromatography [29]. In addition, the photo cleavable group enabled discharging cross-linked peptides from the avidin-biotin complex by UV light exposure. The software was designed to identify and compare cross-linked products from CID and ETD tandem mass spectrometry-cleavage data and provide a confident output list of cross-linked peptides.

The software is also capable to list cross-linked products from a cleavable cross-linker with one cleavage site. This software was evaluated with the data from the cross-linking experiments with DUCCT /PC-DUCCT-Biotin. Cleave-XL efficiency was demonstrated in standard peptides and proteins BSA and Ubiquitin. We also demonstrated its efficiency in a large-scale experiment by identifying protein complexes in immune cell macrophages using targeted database search. Currently the software is capable to search protein structures and protein complexes consist of ~10 to 15 proteins, which we believe is sufficient to study structure of protein and small to medium sized protein complexes. We believe with DUCCT cross-linking technology and automated software tool Cleave-XL, we can routinely screen high confidence cross-linked peptides from proteins and proteins complexes.

3.2 Experimental

3.2.1 Materials

The Dual Cleavable Cross-Linker (DUCCT) and PC-DUCCT-Biotin were synthesized in our laboratory (see the details in the synthesis sections). Two model peptides, Neurotensin and Bradykinin were purchased from Ana spec (San Jose, CA). Two proteins, Ubiquitin and Bovine Serum albumin (BSA) were purchased from Sigma-Aldrich (St. Louis, MO). Dimethyl sulfoxide (DMSO), and Tris.HCl were also obtained from the Sigma-Aldrich (St. Louis, MO). The reducing agent, Dithiothreitol (DTT); the alkylating agent, Iodoacetamide (IAM) and the protease Trypsin were purchased from Biorad (CA), Sigma-Aldrich (St. Luis, MO) and Promega (Medison, WI) respectively. LCMS grade Methanol and Acetonitrile were purchased from VWR (Randor, PA). Eighteen Milli-Q water (Aries Filter works, NJ) was utilized for all experimental studies. Monomeric Avidin was obtained from Themo Fisher scientific (Waltham, MA).

3.2.2 Synthesis of PC-DUCCT-Biotin

Prelude peptide synthesizer (gyros protein technologies, AZ) and Applied biosystems 431A Peptide synthesizer were utilized for synthesis of the PC-DUCCT-Biotin. The synthesis protocol is slightly modified from the previously synthesized DUCCT compound.[2] First, two Fmoc-Glu (Millipore Sigma) residues was coupled with Rink amide MBHA resin to produce the Glu-Glu sequence (Novabiochem) followed by the Fmoc-lysine-biotin (Anaspec.) and the photo-labile unit (Advanced Chem Tech). We add two glutamic acids to improve the solubility at neutral pH. Then the linker was followed by coupling of the super-sensitive form of Fmoc-Asp, Fmoc-Asp(O-2-PhiPr) and Fmoc-Proline and the typical form of Asp, Fmoc Asp-OtBu. Piperidine was used for releasing the Fmoc group and HBTU (Millipore Sigma) acted as a coupling reagent. Subsequent coupling of Hydralink 6-Fmoc-HNA with Asp residue was followed by the addition of 4-formylbenzoic acid, which generated the labile hydrazine bond and a free carboxylic acid terminal. Another terminal carboxylic acid was introduced by releasing the protecting group from the supersensitive Asp utilizing 2% Trifluoro acetic acid (TFA) in dichloromethane. These two carboxylic acids were then activated by forming the esters with N-hydroxysuccinimide (NHS). The final product was released from the resin with 95% TFA in water. The compound was reconstituted in 50% acetonitrile with 0.1% TFA after ether precipitation and then lyophilized.

3.2.3 Cross-Linking of Model Peptides

One microliter of 10 mM model peptide (10 nmol) (Neurotensin/Bradykinin) was treated with one microliter of 100 mM our novel cross-linker (10 nmol) (in 1:10 molar ratio) in PBS buffer (pH 7.2) and the reaction was allowed to be continued for 30 minutes at room temperature. After that, the reaction was quenched with 50 mM Tris-HCl buffer (pH 8.0).

3.2.4 Cross-Linking of Ubiquitin and BSA

The cross-linking reaction of protein with the novel cross-linker DUCCT/ PC-DUCCT-Biotin was performed in different molar ratio (1:50, 1:100, 1: 200, 1:500). PBS Buffer was utilized to maintain the pH. The reaction was incubated for 30 mins and then stopped by 50 mM Tris Buffer. To remove the excess cross-linker, a protein concentrator was utilized (3 KDa MW cut off, Thermo Scientific, USA). The cross-linked proteins were then reduced with DTT, alkylated with IAM and digested by trypsin [30].

3.2.5 Cross-Linking of An Immune-Receptor Protein Complex (TLR4 Protein Complexes)

For this study, we took a murine RAW 264.7 macrophage cell line. The cells were grown in supplemented DMEM (Dulbecco's Modified Essential Medium) at 37˚C under 5% CO_2. Cells were stimulated with LPS at 100 ng/ml and incubated for 4 hrs. at 37˚C. Then, cells were collected after several washes with PBS buffer and re-suspended in PBS Buffer (pH 7.4). Subsequently, the cells were treated with our cross-linker (DUCCT) at a final concentration of 1 mM and incubated for an hour. After that the reaction was washed with PBS buffer and lysed with RIPA buffer. The extracted proteins after lysis process was further purified with the methanol-chloroform precipitation method. After that, the purified protein pellet was resuspended in 50 mM NH_4HCO_3 [31, 32]. Finally, the proteins were reduced, alkylated and digested by the trypsin at a 1:50 concentration ratio for overnight at 37˚C. The digested peptide samples, desalted, and dried with speed vacuum and reconstituted in 0.1% formic acid. Samples were analyzed by nano-LC-MS/MS using a Velos Pro LTQ Mass Spectrometer coupled to an UltiMate 3000 ultra-high-performance liquid chromatography (UHPLC, Thermo Fisher Scientific, MA).

3.2.6 Avidin-Biotin Affinity Purification and Photo-Cleavage

After digestion, the solution was incubated with 20 µL ultra-link monomeric avidin bead suspension (Thermo Fisher Scientific) for 1/2 hours. After that, offline 365 nm UV light (UVGL-58 Handheld UV Lamp) was applied for photo-cleavage. The sample with the bead was incubated under UV irradiation (365 nm wavelength) for 2/4 hours and the cross-linked peptides were released from the bead and came to the solution. Then, the supernatant solution was taken and desalted by ZipTip (Thermo Scientific, Waltham, MA). Finally, the solvent was removed using speed vacuum and reconstituted in 0.1% formic acid. Linear Ion Trap (LTQ) Velos Pro mass spectrometer (Thermo Scientific, USA) was utilized to analyze the cross-linked peptide samples.

3.2.7 Nano HPLC/ Nano-ESI-LTQ Mass Spectrometry Analysis

The digested peptide mixtures were carried out into LTQ Velos pro mass spectrometry coupled with a UHPLC (UltiMate 3000, Dionex, USA). Cross-linked peptides were isolated by reverse–phase chromatography, utilizing a nano viper analytical C18 column. (Acclaim pepMap, 75 µm, 150 mm, 2 µm, 100 Å, Thermo Scientific, CA, USA). The cross-linked peptides were separated and eluted using gradients from 3 to 40% B (95% Acetonitrile, 5% Water and 0.1% FA). The gradient time varies from 60 to 90 minutes according to the complexity of the samples. The nano column flow rate and injection volume were set at 300 nl/min and 5 µl respectively. The nano HPLC system was directly connected with nano-ESI source with a fixed spray voltage and heated capillary temperature of 2-3 V and 275 °C. Data was obtained in a data dependent MS/MS mode: each high resolution full scan (m/z = 350 to 2000, R = 30,000) was followed by product ion scans in LTQ of the five most abundant signals in the full scan mass spectrum. Dynamic exclusion (exclusion duration time) was fixed 30 ms and data acquisition was controlled by X-Calibur 2.0 (Thermo Fisher Scientific).

In CID fragmentation mode, activation energy was set 45% along with isolation width 1.5 Da, activation Q 0.25, and activation time 10 ms. For ETD operation mode, source temperature of the reagent ion source, reagent ion source CI pressure, emission current, and reagent ion electron energy was set to 110°C, 20 psi, 50 UA and -70 v, respectively. The ETD reaction time was fixed 80 ms with isolation width of 2 Da.

3.3 Result and Discussions

3.3.1 Photo-Cleavable DUCCT Cross-Linker Concept and Strategies

We have developed a new cross-linker called PC-DUCCT-Biotin using our previously reported concept of dual cleavable cross-linking technology (DUCCT) [2].

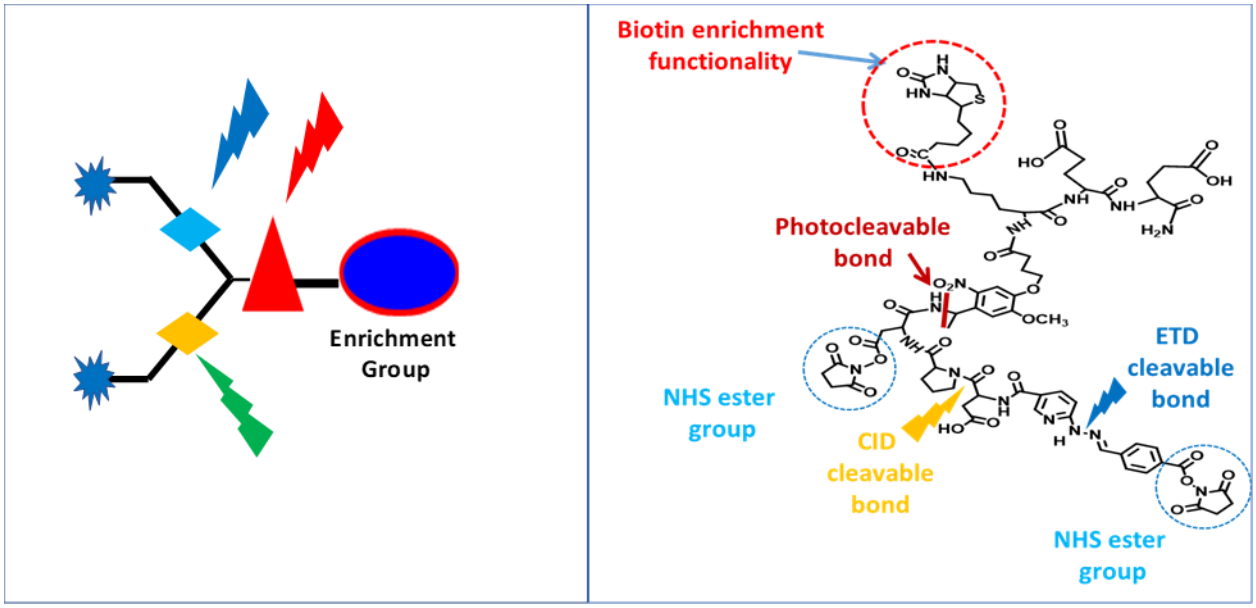


Figure 3-1. The design and chemical structure of the affinity tagged photo-cleavable dual cleavable cross-linking technology (PC-DUCCT-biotin). (A) The design of the dual CID and ETD cleavable cross-linker with an enrichment functionality and a photo-cleavable moiety. The orange and blue diamond symbols indicate the CID (DP bond) and ETD (hydrazone) MS-cleavable sites, respectively. Red triangle indicates photo-cleavable sites. (B) The chemical structure of the PC-DUCCT-Biotin.

This NHS ester based cross-linker covalently links with the primary amine side-chain of lysine residues or N-termini of proteins. It contains two gas phase cleavable bonds in its structure. One of them is aspartic acid-proline (DP) bond) which selectively cleaves by collision-based energy and another one is hydrazone (N-N) bond which is particularly cleaved by Electron transfer-based dissociations **(Figure 3-1).** In addition, it contains a biotin group with a photo-cleavable moiety. The spacer chain distance of this crosslinker is found ~12.0Å using a chem draw software **(Figure B-S1).**

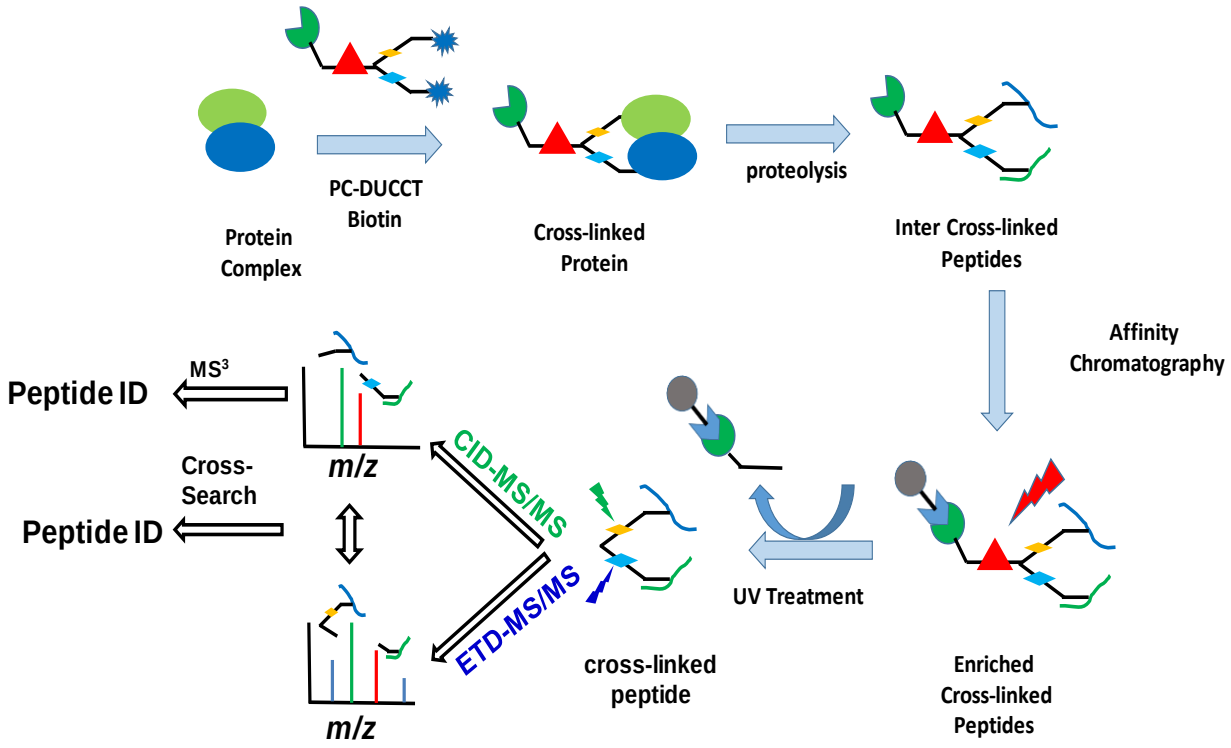


Figure 3-2. Scheme to identify cross-linked peptides by PC-DUCCT Biotin using CID and ETD tandem mass spectrometry. Complementary CID and ETD- MS/MS will identify peptide sequence without further MS/MS of the fragment ions.

In **Figure 3-2,** we demonstrate the methods for the identification of inter cross-linked peptides utilizing our enrichment based cross-linker. PC-DUCCT-Biotin contains one enrichment group in the design in addition to a photo-labile group. The enrichment tag (biotin) enabled us to purify cross-linked peptides by avidin-biotin affinity chromatography and photo-cleavable group allowed removal of the biotin group after purification of protein or protein complexes. This will reduce the fragmentation complexity of biotin group in the mass spectrometry data. After removal of the biotin group, the smaller cross-linked peptides can be cleaved by two mass spectrometry fragmentation techniques namely CID and ETD. These fragmentations techniques generated different signature fragments from the same cross-linked product and complementary signatures in differential fragmentation techniques provided high confidence in the identification.

3.3.2 Cross-Linking of Standard Peptides

Standard peptides, Neurotensin contains one lysine in its sequence and the N terminal is modified with pyroglutamate group and Bradykinin does not have any lysine but it has free N terminal. Our cross-linker PC-DUCCT-Biotin contains two NHS groups which selectively reacts with the primary amine of N terminal or the side-chain of lysines. The cross-linker efficiently labeled both peptides and generated dead-end and inter-cross-linking products.

For Neurotensin, inter cross-linked peptide was identified at m/z 784.81 (5+). CID-MS/MS generated two cleaved peaks with the highest intensity and added cross-linker residues [(m/z at 943.04 (2+) and 1019.40 (2+)] and ETD-MS/MS also produced two different fragment ions from same precursor (m/z at 902.86, 2+ and 1059.96, 2+)**.** We also observed several reduced charged precursors in ETD-MS/MS **(Figure 3-3)**. The complete calculations of cross-link precursors and reporter ions are illustrated in the supplementary section. **(Table B-S1).**

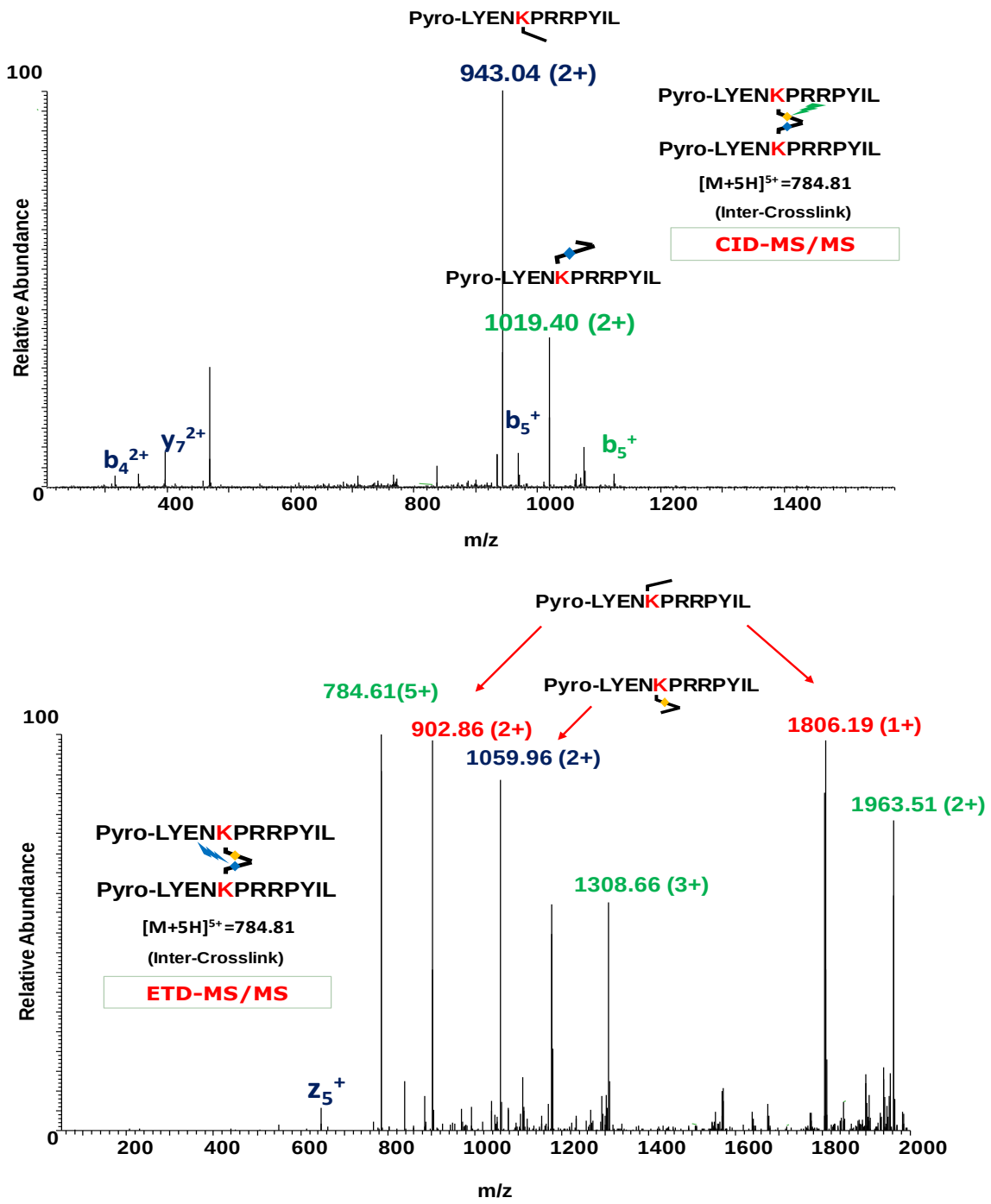


Figure 3-3. Experimental CID and ETD mass spectra of a cross-linked neurotensin peptide dimer identified by PC-DUCCT-Biotin cross-linker. The CID and ETD mass spectra showed different signatures for the same cross-linked peptide.

Dead-end peptides (partially hydrolyzed cross-linked product) gives the information on surface reactive lysine residues. A dead-end peptide derived from neurotensin was identified at m/z 756.36 (3+). Due to the asymmetric nature of our cross-linker, peptides can react with cross-linker in two different ways and consequently, after gas phase fragmentation, sometimes it can produce both peptide fragments in the same mass spectrum **(Figure B-S2).** In CID-MS/MS, the selective cleavage happened on DP bond site and generated the signature fragment ion (m/z = 942.93 (2+). We also observed 3+ charge state of peptide-fragment ions (m/z = 629.13 (3+). We observed the cleavage at the different site (N-N bond) in ETD-MS/MS and the generated peak matched with the theoretically calculated data. Charge-reduced precursor ions were also observed in ETD-MS/MS spectrum. Complementary CID and ETD MS/MS confidently identified dead-end peptides. In the same way, the cross-linked products from Bradykinin (both inter cross-linked and dead-end) were identified with high confidence **(Figure B-S3, Table B-S2).**

3.3.3 Software Description / Program Workflow

For analyzing protein structure and protein complexes using our DUCCT cross-linking technology, we have developed a prototype software tool so cross-linked peptides can be analyzed automatically.The Cleave- XL software was coded with java script. In the first step, the protein sequence is incorporated as a FASTA file input format from the Uniport database. Then the theoretical masses of peptides after trypsin digestion corresponding to the proteins are calculated. This is accomplished based on the setting of proteolysis and amino acid sequence provided by the FASTA file. After that, the cross-linked peptide (inter-cross-linked or dead-end) database was created according to our interest in application. These were generated by adding the cross-linker mass as well as fragment masses with generated peptide database mass.

Two different kinds of fragment database were created based on our applied fragmentation, collision induced dissociation (CID) and electron transfer dissociation (ETD).

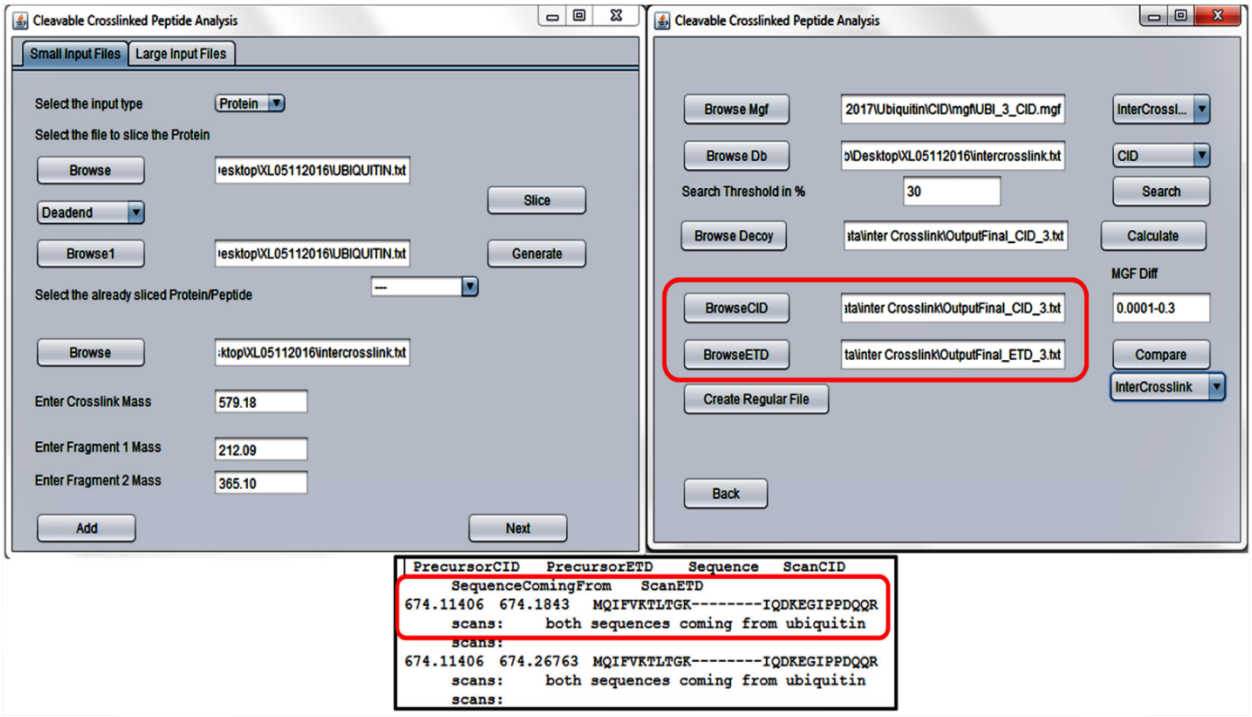


Figure 3-4. A screen shot of our standalone software Cleave-XL for the confident identification of cross-linked peptides by sequential CID and ETD cleavable mass spectrometry.

Experimental mass-spectrometric data was imported as MGF (Mascot Generic Format) file. First, we searched the CID-MS/MS experimental data (MGF file) with our created theoretical database and thus we produced a list of hits as CID output files. Another list of several hits was printed as ETD output files when we search the ETD-MS/MS mgf file against the theoretical data set. We also set an intensity threshold of the MS/MS fragment peaks to minimize the noise level and false identification. By cross matching experimental data with the theoretical database, we get a number of hits in the output file.

After comparing the common sequence of both output files, we identified the cross-linked peptides confidently, enlisted in the difference output file. The screen shot, and the schematic flow chart of the graphical user interface demonstrates the clear picture of the workflow of our developed software. **(Figure 3-4, S4)**

False-Discovery Rate: The software has the capability to calculate the false discovery rate. Regular search was performed by incorporating the correct parameters that produce the true positive value. Decoy search was executed with reverse protein sequence that gives the false positive value. By applying the following equation, **FDR = FP / (FP + TP),** FDR is calculated [33]. FDR calculation of small database is not ideal, but it will be very beneficial for manual screening of false positive ID's of cross-linked peptides.

3.3.4 Testing the Search Efficiency of Cleave-XL with Standard Proteins

After testing the cross-linked products of standard peptides, we decided to test the search efficiency of our software Cleave-XL in proteins Ubiquitin and Bovine Serum Albumin (BSA). In both cases, a series of inter cross-linked and dead-end products were identified. Ubiquitin is a small protein (mw ~ 8,500 Da) and a very good model for testing the formation of cross-linking products. First, we generated a theoretical database of cross-linked peptides utilizing protein database (fasta file) as well as precursor and fragment masses (both CID and ETD). Then we input the experimental data as a mgf file. We observed approximately 15,000 scans for ubiquitin experiment. Cleave-XL first searched the precursor mass spectra database against the theoretical database and then matched spectra were searched for fragment ions. CID search generates several hits after cross-searching of our experimental data with the generated theoretical database.

Only CID search did not provide the confidence in cross-linked peptide identification. It requires additional MS/MS of the fragments for confirmation. Our ETD search data can provide an additional step of validation. We searched our experimental ETD-MS/MS data with theoretical ETD dataset in the same way and found a list of hits. In the final step, we compared both CID and ETD output files and filtered the common sequences identified in both datasets; we were able to identify several cross-linked peptide sequences **(Figure 3-4).** Cleave-XL unambiguously identified several inter cross-linked and dead-end peptides derived from ubiquitin. Complementary signature fragments with high intensity in both CID-MS/MS and ETD-MS/MS spectra delivered the confidence in identification of cross-linked products. The list of inter-cross-linked and dead-end peptides are shown in **Figure B-S5**. Crystal structure helped to locate the site of cross-linking and distance between confidently identified cross-linked peptides of Ubiquitin. The distance constraint (11.6 Å) between cross-linked lysine residues also provided the confidence of the identification **(Figure B-S5).** An inter cross-linked peptide derived from Ubiquitin is demonstrated in **Figure B-S6**. Again, we manually evaluated fragment peaks from MS^{3rd} experiment for verification and observed that the spectrum confidently represents the sequence **(Figure B-S7).** After validation in a small protein, we tested our software in a large protein, BSA (~66 KDa). The software was able to identify the cross-linked peptides with high confidence. One of the identified inter peptide cross-linking product (**SLGKVGTR-------EKVLTSSAR**) with two different charge states at m/z 596.3095 ($M+4H^+$) and 477.2490 ($M+5H^+$) are shown in **Figure 3-5 and Figure B-S8**, respectively**.**

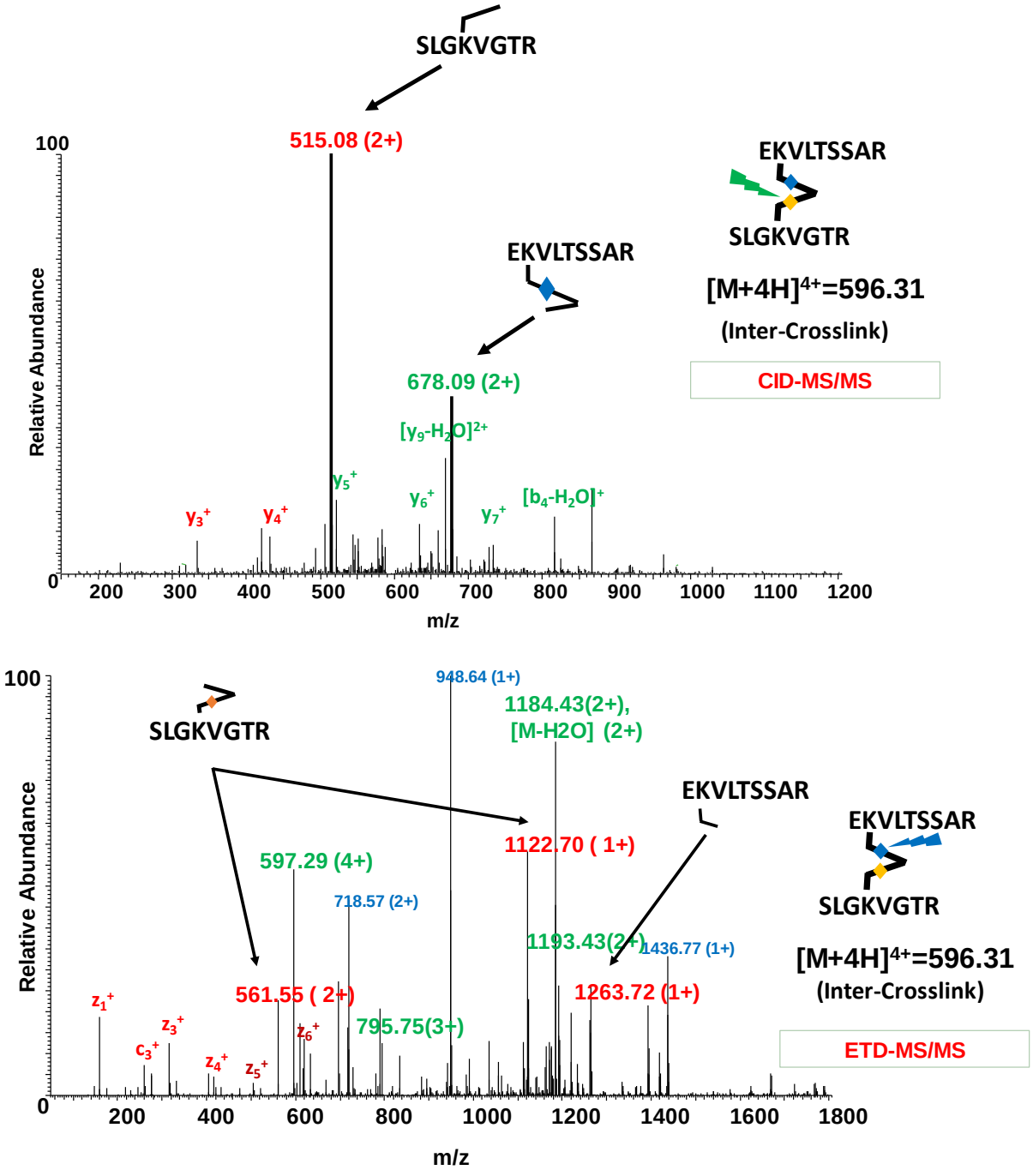


Figure 3-5. Example of CID and ETD mass spectra of an inter cross-linked peptide, derived from BSA. CID and ETD signatures of the same cross-linked peptide provided high-confidence identification of this cross-linked peptide.

After applying collision-based energy, cleavage happened in selective site and produced two signature ions, [(m/z = 515.08 (2+) and m/z = 678.09 (2+)]. Similar to CID, ETD also cleaved in the specific site and generated two signature fragments [(1122.70 (1+) and 1263.72 (1+). We also observed several fragment peaks (c and z ions) in the spectrum. PC-DUCCT-Biotin can also be utilized as a single CID-cleavable cross-linker. Further MS^{3rd} of the fragment masses can be used to confirm the sequence of the cross-linked peptides. MS/MS of the fragment peaks derived from the inter peptide cross-linking product (**SLGKVGTR-------EKVLTSSAR**) confidently sequenced the peptides **(Figure B-S9).**

Due to the asymmetrical behavior of cleavable cross-linkers, peptides can react in two different ways with the reactive ends of the cross-linkers and sometimes generate the mixture of two peptides in the mass spectra. The uneven fragmentation of both cleavage sites due to the different labeling positions, generate a peptide added with two different residual masses of the cross-linker**.** We have observed similar cross-linked products for inter-cross-linked peptide SLGKVGTR-------EKVLTSSAR **(Figure 3-5 and Figure B-S8)**. Our software Cleave-XL was capable of pinpointing these fragment ions with high confidence. Complete calculations of cross-linked precursors and fragment ions are demonstrated in **Table B-S3.** From these results, it is clear that DUCCT technology will provide an unambiguous identification of cross-linked products.

We also identified several dead-ends peptides derived from BSA. One of the dead ends is shown in **Figure B-S10.** Our software CLEAVE- XL automatically detected the fragments for both types of MS based fragmentations techniques. After comparing both datasets (CID output and ETD output) the CLEAVE- XL identified a list of cross-linked peptides from BSA with high confidence. A list of inter cross-linked peptides derived from BSA is shown in **Table B-S4.**

The distance constraints between cross-linked peptides were shown in the crystal structure that provided the information regarding spatial orientation and the connectivity in the overall protein structure. **(Figure B-S11).**

In our investigation, we identified 18 inter cross-linked peptides from BSA, which is much higher number compared to photo-cleavable PIR (pc-PIR) cross-linker reported by Bruce group. Whereas five inter cross-linked peptides identified by pcPIR from BSA experiment. It is clear PC-DUCCT-Biotin performed efficiently in proteins and, Cleave-XL search identified high confidence cross-linked peptides. We would also like to emphasis that this comparison is not ideal due to the difference in spacer chain length and also the design. The spacer chain distance of PC-DUCCT-Biotin is ~12.0 Å, which is much shorter compared to pcPIR (~40.0Å). DUCCT is developed to identify cross-linked peptide confidently and our focus is to identify the number of confident cross-linked peptides without ambiguity.

3.3.5 Testing Cleave-XL Software in DUCCT Cross-Linked Immune Receptor Signaling Protein Complex

Lipopolysaccharides (LPS) are mostly found in the outer membrane of the gram-negative bacteria [34]. The recognition mechanism for LPS was studied by several laboratories due to its potent immunostimulatory properties [35]. Toll Like Receptor 4 (TLR4) act as an LPS recognition receptor and helps to induce downstream signaling pathway which contributes in response against microbial infections. Based on the finding of different research groups, TLR4 requires three additional proteins for LPS recognition. These are LPS binding protein (LBP), CD14, and MD-2. LBP, a lipid transferase, helps to transfer the LPS from bacterial outer membrane to CD14 [36]. CD14 is a ~50-kDa glycoprotein, plays a crucial role in loading LPS to the TLR4/MD-2 receptor complex.

To test the efficiency of our DUCCT cross-linking technology and associated software tool Cleave-XL, we performed a cross-linking study in 264.7 mouse raw macrophage after LPS stimulation. Due to the limitation of searching large database by Cleave-XL, a targeted search is performed using these four proteins to see the efficiency of cross-linking and search process to find TLR4 protein complex.

We took the sequence of four protein (TLR4, MD-2, CD14, LBP) from uniport database for generating the theoretical database for cross-linked peptides. In this study, twenty-five inter-cross-linked peptides were identified from the TLR4 complex **(Table B-S5).** The list produced from both search (CID and ETD output) then compared to generate an output of a confident list of cross-linked products. One of the inter cross-linked peptides at m/z 622.48 (6+) (IMKTCLQNLAGLHVHR ……. FLQWVKEQK) was identified, which was derived from cross-linking within TLR4 **(Figure 3-6).** In CID-MS/MS, two reporter fragment masses at 565.01 (4+) and 491.56 (3+) were identified by Cleave-XL. Similarly, two signature ions at 1014.49 (2+) and 852.35 (2+) were found from ETD-MS/MS data of the same inter cross-linked product. Obviously, DUCCT cross-linking technology was able to identify this peptide with confidence. Complete calculations of precursors and fragment ions obtained from this inter cross-linked peptide is provided in the supplementary information **(Table B-S6).** Another inter-cross-linked peptide derived from cross-linking between TLR4 and LBP (IMKTCLQNLAGLHVHRLILGEFK….. GKWK) is shown in **Figure B-S12.** We also manually verified the fragment peaks from MS3rd experiment and conclude that the spectrum represents the sequence confidently **(Figure B-S13).**

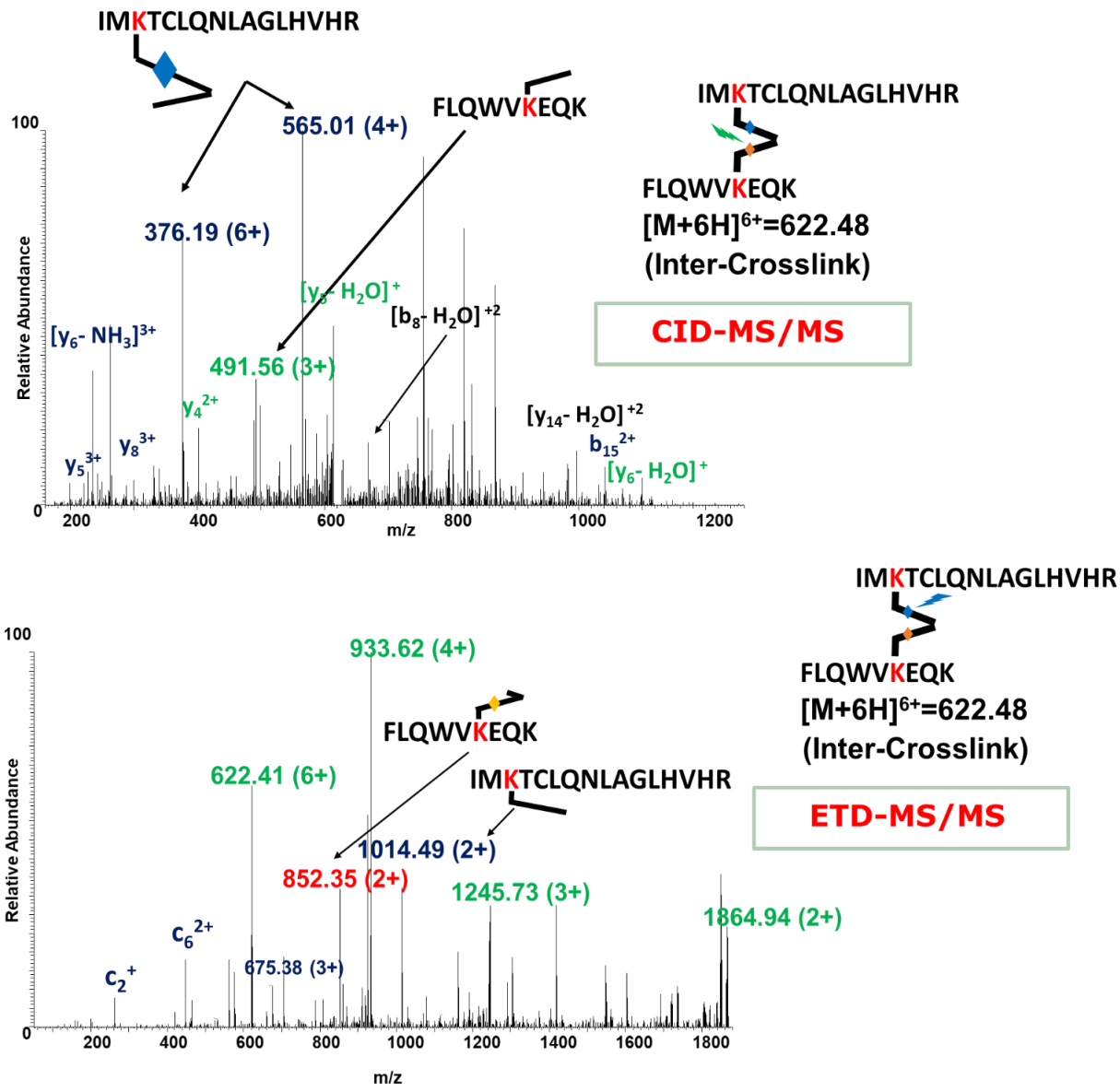


Figure 3-6: Example of an inter cross-linked peptide at m/z 622.48 (6+) identified from TLR4 complex by DUCCT. CID and ETD mass spectra generated different signatures for the same cross-linked peptides.

There is no in vivo study conducted on the structure of TLR4 complex. Several X-ray crystallography studies was conducted using human or mouse TLR4.

One study used ectodomain of human TLR4 (amino acids 27–631) and, the human MD-2 gene (amino acids 19–160) fused to the Protein A tag and co-expressed in High Five (Hi5) insect cells [37, 38]. The TLR4 then deglycosylated using PNGase F. Mouse TLR4/MD-2 complexes were prepared similar way for X-ray crystallography study [39]. TLR4 forms dimer upon recognition of LPS and forms TLR4-MD-2 complexes. Due to glycosylation this protein is also difficult to observe in regular bottom- up proteomic analysis. We provided a challenging study here to test the efficiency of DUCCT and Cleave-XL. We are planning a complete structural study of TLR4 complex using DUCCT technology.

3.4 Conclusions

In this report, we demonstrate a DUCCT cross-linking technology with an affinity tag called PC-DUCCT-Biotin. The cross-linker was demonstrated in standard peptides and proteins. Furthermore, we presented a prototype software Cleave-XL to analyze DUCCT cross-linked data. Cleave-XL proved to be successful in identifying the cross-linked products in standard peptides and proteins demonstrated with PC-DUCCT-Biotin and TLR4 protein complexes with a regular DUCCT with a targeted database search. Cleave-XL is designed to identify cross-linked peptides automatically from low complexity mass-spectrometric datasets utilizing the dual cleavable cross-linker, thus will be an excellent platform for studying protein structure and protein complexes. We believe DUCCT cross-linking technology and associated software tool Cleave-XL will provide quick screening of cross-linked products hence significantly advance structure biology research using cross-linking. Software development is always an evolving process; thus, further development of the Cleave-XL software tool is underway to screen interactome using DUCCT cross-linking from a large-scale experiment. The software is available to the scientific community upon request.

3.5 References

1. Sevimoglu, T., Arga, K.Y.: The role of protein interaction networks in systems biomedicine. Comput Struct Biotechnol J. 11, 22-27 (2014)
2. Chakrabarty, J.K., Naik, A.G., Fessler, M.B., Munske, G.R., Chowdhury, S.M.: Differential Tandem Mass Spectrometry-Based Cross-Linker: A New Approach for High Confidence in Identifying Protein Cross-Linking. Anal Chem. 88, 10215-10222 (2016)
3. Yu, C., Huang, L.: Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. Anal Chem. 90, 144-165 (2018)
4. Piotrowski, C., Sinz, A.: Structural Investigation of Proteins and Protein Complexes by Chemical Cross-Linking/Mass Spectrometry. Adv Exp Med Biol. 1105, 101-121 (2018)
5. Du, X., Chowdhury, S.M., Manes, N.P., Wu, S., Mayer, M.U., Adkins, J.N., Anderson, G.A., Smith, R.D.: Xlink-identifier: an automated data analysis platform for confident identifications of chemically cross-linked peptides using tandem mass spectrometry. J Proteome Res. 10, 923-931 (2011)
6. Barysz, H.M., Malmstrom, J.: Development of Large-scale Cross-linking Mass Spectrometry. Mol Cell Proteomics. 17, 1055-1066 (2018)
7. Tang, X., Bruce, J.E.: A new cross-linking strategy: protein interaction reporter (PIR) technology for protein-protein interaction studies. Mol BioSyst. 6, 939-947 (2010)
8. Chowdhury, S.M., Munske, G.R., Tang, X.T., Bruce, J.E.: Collisionally activated dissociation and electron capture dissociation of several mass spectrometry-identifiable chemical cross-linkers. Anal Chem. 78, 8183-8193 (2006)
9. Soderblom, E.J., Goshe, M.B.: Collision-induced dissociative chemical cross-linking reagents and methodology: Applications to protein structural characterization using tandem mass spectrometry analysis. Anal Chem. 78, 8059-8068 (2006)
10. Liu, F., Wu, C., Sweedler, J.V., Goshe, M.B.: An enhanced protein crosslink identification strategy using CID-cleavable chemical crosslinkers and LC/MSn analysis. Proteomics. 12, 401-405 (2012)
11. Arlt, C., Gotze, M., Ihling, C.H., Hage, C., Schafer, M., Sinz, A.: Integrated Workflow for Structural Proteomics Studies Based on Cross-Linking/Mass Spectrometry with an MS/MS Cleavable Cross-Linker. Anal Chem. 88, 7930-7937 (2016)
12. Gotze, M., Pettelkau, J., Fritzsche, R., Ihling, C.H., Schafer, M., Sinz, A.: Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis. J Am Soc Mass Spectrom. 26, 83-97 (2015)
13. Iacobucci, C., Gotze, M., Ihling, C.H., Piotrowski, C., Arlt, C., Schafer, M., Hage, C., Schmidt, R., Sinz, A.: A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions. Nat protoc. 13, 2864-2889 (2018)
14. Gutierrez, C.B., Yu, C., Novitsky, E.J., Huszagh, A.S., Rychnovsky, S.D., Huang, L.: Developing an Acidic Residue Reactive and Sulfoxide-Containing MS-Cleavable Homobifunctional Cross-Linker for Probing Protein-Protein Interactions. Anal Chem. 88, 8315-8322 (2016)
15. Kaake, R.M., Wang, X., Burke, A., Yu, C., Kandur, W., Yang, Y., Novtisky, E.J., Second, T., Duan, J., Kao, A., Guan, S., Vellucci, D., Rychnovsky, S.D., Huang, L.: A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells. Mol Cell Proteomics. 13, 3533-3543 (2014)

16. Young, M.M., Tang, N., Hempel, J.C., Oshiro, C.M., Taylor, E.W., Kuntz, I.D., Gibson, B.W., Dollinger, G.: High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. P Natl Acad Sci U S A. 97, 5802-5806 (2000)
17. Chu, F.X., Shan, S.O., Moustakas, D.T., Alber, F., Egea, P.F., Stroud, R.M., Walter, P., Burlingame, A.L.: Unraveling the interface of signal recognition particle and its receptor by using chemical cross-linking and tandem mass spectrometry. P Natl Acad Sci U S A. 101, 16454-16459 (2004)
18. de Koning, L.J., Kasper, P.T., Back, J.W., Nessen, M.A., Vanrobaeys, F., Van Beeumen, J., Gherardi, E., de Koster, C.G., de Jong, L.: Computer-assisted mass spectrometric analysis of naturally occurring and artificially introduced cross-links in proteins and protein complexes. Febs J. 273, 281-291 (2006)
19. Back, J.W., Notenboom, V., de Koning, L.J., Muijsers, A.O., Sixma, T.K., de Koster, C.G., de Jong, L.Z.: Identification of cross-linked peptides for protein interaction studies using mass spectrometry and O-18 labeling. Anal Chem. 74, 4417-4422 (2002)
20. Gotze, M., Pettelkau, J., Schaks, S., Bosse, K., Ihling, C.H., Krauth, F., Fritzsche, R., Kuhn, U., Sinz, A.: StavroX-A Software for Analyzing Crosslinked Products in Protein Interaction Studies. J Am Soc Mass Spectrom. 23, 76-87 (2012)
21. Yang, B., Wu, Y.J., Zhu, M., Fan, S.B., Lin, J.Z., Zhang, K., Li, S., Chi, H., Li, Y.X., Chen, H.F., Luo, S.K., Ding, Y.H., Wang, L.H., Hao, Z.Q., Xiu, L.Y., Chen, S., Ye, K.Q., He, S.M., Dong, M.Q.: Identification of cross-linked peptides from complex samples. Nat Methods. 9, 904 (2012)
22. Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L., Beck, M., Schmidt, A., Mueller, M., Aebersold, R.: Identification of cross-linked peptides from large sequence databases. Nat Methods. 5, 315-318 (2008)
23. Seebacher, J., Mallick, P., Zhang, N., Eddes, J.S., Aebersold, R., Gelb, M.H.: Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing. J Proteome Res. 5, 2270-2282 (2006)
24. Gao, Q.X., Xue, S., Doneanu, C.E., Shaffer, S.A., Goodlett, D.R., Nelson, S.D.: Pro-CrossLink. Software tool for protein cross-linking and mass spectrometry. Anal Chem. 78, 2145-2149 (2006)
25. Gotze, M., Pettelkau, J., Fritzsche, R., Ihling, C.H., Schafer, M., Sinz, A.: Automated Assignment of MS/MS Cleavable Cross-Links in Protein 3D-Structure Analysis. J Am Soc Mass Spectrom. 26, 83-97 (2015)
26. Liu, F., Lossl, P., Scheltema, R., Viner, R., Heck, A.J.R.: Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. Nat Commun. 8, 15473 (2017)
27. Liu, F., Rijkers, D.T.S., Post, H., Heck, A.J.R.: Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. Nat Methods. 12, 1179-+ (2015)
28. Chowdhury, S.M., Du, X., Tolic, N., Wu, S., Moore, R.J., Mayer, M.U., Smith, R.D., Adkins, J.N.: Identification of cross-linked peptides after click-based enrichment using sequential collision-induced dissociation and electron transfer dissociation tandem mass spectrometry. Anal Chem. 81, 5524-5532 (2009)
29. Yang, L., Tang, X., Weisbrod, C.R., Munske, G.R., Eng, J.K., von Haller, P.D., Kaiser, N.K., Bruce, J.E.: A photocleavable and mass spectrometry identifiable cross-linker for protein interaction studies. Anal Chem. 82, 3556-3566 (2010)

30. Fang, Z., Baghdady, Y.Z., Schug, K.A., Chowdhury, S.M.: Evaluation of Different Stationary Phases in the Separation of Inter-Cross-Linked Peptides. J Proteome Res. 18, 1916-1925 (2019)
31. Kamal, A.H.M., Chakrabarty, J.K., Udden, S.M.N., Zaki, M.H., Chowdhury, S.M.: Inflammatory Proteomic Network Analysis of Statin-treated and Lipopolysaccharide-activated Macrophages. Sci Rep. 8, 164 (2018)
32. Kamal, A.H.M., Fessler, M.B., Chowdhury, S.M.: Comparative and network-based proteomic analysis of low dose ethanol- and lipopolysaccharide-induced macrophages. PloS one. 13, e0193104 (2018)
33. Benjiamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc B. 57, 289-300 (1995)
34. Wang, X., Quinn, P.J., Yan, A.: Kdo2-lipid A: structural diversity and impact on immunopharmacology. Biol Rev. 90, 408-427 (2015)
35. Silipo, A., Molinaro, A.: The diversity of the core oligosaccharide in lipopolysaccharides. Subcell Biochem. 53, 69-99 (2010)
36. Miyake, K.: Invited review: Roles for accessory molecules in microbial recognition by Toll-like receptors. J Endotoxin Res. 12, 195-204 (2006)
37. Kim, H.M., Park, B.S., Kim, J.I., Kim, S.E., Lee, J., Oh, S.C., Enkhbayar, P., Matsushima, N., Lee, H., Yoo, O.J., Lee, J.O.: Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. Cell. 130, 906-917 (2007)
38. Park, B.S., Song, D.H., Kim, H.M., Choi, B.S., Lee, H., Lee, J.O.: The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature. 458, 1191-1195 (2009)
39. Ohto, U., Fukase, K., Miyake, K., Shimizu, T.: Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. P Natl Acad Sci U S A. 109, 7421-7426 (2012)

Chapter 4

Evaluating the performance of an ETD-Cleavable Cross-Linking strategy for Elucidating Protein Structures

Abstract

Chemical cross-linking is a powerful strategy for elucidating the structures of protein or protein complexes. The distance constraints obtained from cross-linked peptides represent the three-dimensional modeling of the protein complexes. Unfortunately structural analysis of cross-linking approach requires significant amount of data to elucidate protein structures. These requires different cleavable cross-linkers with different range of spacer chains distances so confident and large-number of data can be collected. ETD cleavable group hydrazone was reported. Its fragmentation with conjugated peptides showed promise for a new ETD cleavable cross-linker development. However no cross-linker was developed yet utilizing this group. For the first time we attempted to develop an ETD cleavable cross-linker utilizing hydrazone group. We overcome the pitfall for the synthesis of this cross-linker and an easy synthesis is reported. In this report, we evaluate the performance of this cross-linker called Hydrazone Incorporated ETD cleavable cross-linker (HI-ETD-XL) in model peptide and proteins. The characteristic fragmentation behavior of HI-ETD-XL during Electron transfer dissociation (ETD) and subsequent sequence identification of the fragment ions by CID-MS3 study allows identifying cross-linked peptides unambiguously. We believe availability of this ETD cleavable cross-linker will help providing additional data for protein structural predictions.

Significance

Many cellular processes rely on the structural dynamics of protein complexes. The detailed knowledge of the structure and dynamics of protein complexes is crucial for understanding their

biological functions and regulations. But most of the structure of these multiprotein entities remains uncharacterized, and sometimes it is very challenging to reveal with biophysical techniques alone. Chemical cross-linking combined with mass spectrometry (MS) has proven to be a dependable strategy in structural proteomics field. However, data complexity and false identifications are the significant hindrances for unambiguous identification of cross-linked peptides. Confidence identifications requires structural studies with cross-linkers of different spacer chain and properties. This ETD cleavable cross-linking workflow will provide additional confidence to overcome these drawbacks and allows us to pinpoint cross-linked peptides confidently.

Keywords: Cross-linking, Electron transfer dissociation, Mass spectrometry, 3D modeling.

4.1 Introduction

Chemical cross-linking coupled with mass spectrometric analysis (XL-MS), is an established alternative method for protein-structure studies. This strategy has the advantage over the traditional approaches to characterize the structure of protein or large protein complexes[1, 2].

The core of the XL-MS is a cross-linking reagent that covalently captures two proximal amino acid residues within protein or protein complexes. The cross-linker is composed of two reactive groups separated by a defined spacer length specifies the spatial orientation and distance information of the interacting proteins in multiprotein complex assembly. Usually, cross-linked products are identified using the bottom-up approach (LC-MS/MS) which utilizes enzymatic digestion to produce smaller peptides. The site of cross-linked amino acids along with the distance constraints between interacting residues predicts three-dimensional structural model of protein complexes[3-8].

Due to the challenge of identifying cross-linked peptide with high confidence, cleavable cross-linking approach was developed. These reagents contain one or more than one labile bond in their structures and can be preferentially cleaved in different ways based on their chemical properties. For example, cleavage of the labile bonds can be induced by photo, chemical, and gas-phase collision energy, which separates two inter cross-linked peptides before or during the MS analysis [5, 9-13]. Gas phase MS-cleavable approach is most attractive due to its unique power to produces two diagnostic cross-link fragments in MS2 and subsequent identification of the cross-linked peptides accurately by MS3rd sequencing [5, 14]. Currently, most of the MS-cleavable reagents are collision-induced dissociation (CID)-cleavable, for e.g. protein interaction reporters (PIR), disuccinimidyl sulfoxide (DSSO) and its derivatives, BuUrBu (DSBU), disuccinimidyl-succinamyl-aspartyl-proline (SuDP) and cyanurbiotindipropionylsuccinimide (CBDP) [10-12, 14-18]. Recently, our group developed a novel Dual Cleavable Cross-linking Technology (DUCCT), which produced high confidence identification of same cross-linked peptides by differential tandem mass spectrometry. Our cross-linking reagent preferentially cleaved by two tandem mass spectrometry fragmentation processes (CID and ETD), generated two different signature ion peaks from the same cross-linked peptide. We proved that this dual cleavage workflow identified cross-linked peptides accurately [5].

Although Electron transfer dissociation (ETD) is a widely used tandem mass spectrometry fragmentation techniques, however, considerably less attention was paid to design ETD-cleavable cross-linkers compared to CID. ETD is a radical driven process that keeps many CID labile bonds un-cleaved, which is very suitable to keep the modification on amino acid side chains during fragmentation [19-21]. This feature makes ETD a reliable tool for characterizing the post-translation modifications (PTMs), e.g., phosphorylation, glycosylation [22, 23].

The ETD appears to have a significant advantage in analyzing large, highly charged cross-linked peptides compared to collision-induced dissociation (CID). This process leads to comprehensive, uniformly distributed dissociation along peptide backbones, showing less impact on the peptide sequence than CID. Recently, the Heck Group has reported that ETD with the supplementary HCD (EThcD) activation generates the best sequence coverage for highly charged cross-linked species.[24, 25] One of the major concerns is the ETD fragmentation significantly depends on the precursor ion charge states. It is very crucial to increase the charge density of the selected precursor ions before activation. The charge density of the ions can be manipulated by covalent attachment of fixed charge sites and thus improve the fragmentation [26].

The first efficient ETD-cleavable cross-linker, DEB (1,3-diformyl-5-ethynylbenzene) was developed by the Burlingame group. DEB covalently link two interacting proteins or peptides through reductive animation and add two protonation sites to the cross-linked moiety that facilitate to separate the highly charged inter cross-linked peptides from the unmodified and dead-end modified peptides.[27, 28] Recently developed ETD-cleavable cross-linker, diethyl suberthioimidate (DEST) which contains a 6-carbon alkyl chain and, it breaks at the amidino groups generated after reaction with primary amines. DEST cross-linked peptides identified as diagnostic ion pairs, consisting of peptide- NH_2 and peptide+linker+NH_3 ions in ETD-MS/MS. However, DEST hydrolyzed very quickly and generated high abundant hydrolyzed dead-end spectra that limit its application as well [29, 30].

The mechanism of selective cleavage of the hydrazone bond was proposed by Jennifer Brodbelt group. First, they modified the peptides either with Succinimidyl 4-formylbenzoate (SFB) or succinimidyl 4-hydrazinonicotinate acetone hydrazone (SANH) and then the resultant peptides were conjugated together to form bis-arylhydrazone (BAH)-cross-linked peptides [31-33]. The

selective breakdown of the hydrazone bond was observed under ETD fragmentation. This functional group improved overall charge density of the cross-linked peptides, and selective cleavage facilitated to generate the diagnostic ions in ETD-MS2 spectra. Further, CID-MS3 experiment confirmed the sequence of the peptides. Unfortunately other than some fragmentation analysis no cross-linker was developed with hydrazone incorporated in the design. Additionally, their effectiveness in protein labeling was never demonstrated.

Structural analysis of an unknown protein by cross-linking is a challenging task. It requires significant experimental data from the treatment of several cross-linkers with different spacer chain lengths. This will identify different constraints in structural conformations of proteins. Cross-linkers with different spacer chains and properties will produce volume of confident experimental data so modeling study can be done effectively for proteins. Here, we introduce and evaluate an ETD cleavable cross-linker (ETD-XL), which will be very suitable for validation and benchmarking of protein structures. The cross-linker contains two reactive NHS ester groups for coupling with Lysine residue. We also incorporate a fixed charge containing hydrazone group, which can be preferentially cleaved using ETD. HI-ETD-XL labeling of proteins were demonstrated in different model peptides and proteins. We believe this HI-ETD-XL approach will be an additional structural tool kit for robust and quick structural analysis of proteins and protein complexes.

4.2 Materials and Methods

4.2.1 Materials

Two standard peptides, Neurotensin (pE-LYNKPRRPYIL) and Bradykinin (RPPGFSPFR) were ordered from Ana spec (San Jose, CA). Three proteins, Ubiquitin bovine erythrocytes, myoglobin of equine heart, and Bovine Serum albumin (BSA) were obtained from Sigma-

Aldrich (St. Louis, MO). The reducing agent, Dithiothreitol (DTT), was purchased from Bio-Rad (CA). Dimethyl sulfoxide (DMSO), Tris.HCl, Ammonium bicarbonate (ABC), and Formic acid (FA) were also obtained from the Sigma-Aldrich (St. Louis, MO). For proteolysis, sequencing-grade-modified trypsin ordered from Promega (Madison, WI, USA).3K MWCO protein concentrators (Pierce, IL) were utilized to remove excess cross-linking reagents and concentrate the protein after the cross-linking reaction. Pierce C18 Tips from Thermo Fisher Scientific (Rockford, IL, USA) were used to desalt the samples. Milli-Q-filtrated water (18 MΩ) (Aries Filter works, NJ) was used for all experimental studies.

The cross-linking reagent (ETD-XL) was synthesized in our laboratory. Starting materials, 6-hydrazinicotinic acid purchased from AK Scientific (Union City, CA) and, EDC from Chem Impex (Wood Dale, IL). 4-formyl benzoic acid and N-hydroxysuccinimide (NHS) purchased from Sigma Aldrich. SiliaFlash® P60 40-63 μm (230-400 mesh) 60 Å Irregular Silica Gel used for flash column chromatography.

4.2.2 Synthesis of Cross-linking Reagent

Detailed synthesis protocol was explained in the supplementary sections with characterization. Briefly, 4-formyl benzoic acid coupled with N-hydroxysuccinimide (NHS) (1.1 equiv.) in the presence of EDC and 2,5-dioxopyrrolidin-1-yl 4-formybenzoate formed as a white crystalline solid. In the second step, 2,5-dioxopyrrolidin-1-yl 6-(2-(propan-2-ylidene) hydrazinyl) nicotinate was synthesized after the coupling reaction of 6-hydrazinicotinic acid with N-hydroxysuccinimide (NHS) in the presence of triethylamine, acetone, and EDC. These two intermediate compounds then purified through silica-based column chromatography.
In final step, 2,5-dioxopyrrolidin-1-yl 4-formylbenzoate reacts with 2,5-dioxopyrrolidin-1-yl 6-(2-(propan-2-ylidene)hydrazinyl)nicotinate in presence of glacial acetic acid, resulting in a bright

yellow solution of final pure product 2,5-dioxopyrrolidin-1-yl (E)-4-(2-(4-(((2,5-dioxopyrrolidin-1-yl)oxy)carbonyl)benzylidene)hydrazinyl)-benzoate. Complete characterization and the NMR of the product was provided in the supplementary texts.

4.2.3 Cross-Linking of standard Peptides

The 100 mM stock solution of HI-ETD-XL cross-linker prepared in DMSO. Standard peptides (Neurotensin/Bradykinin) reacted with HI-ETD-XL in a 1:10 molar ratio in PBS buffer (pH 7.4). The cross-linking reaction was allowed to proceed for 30 mins at room temperature. Then, the reaction stopped by adding 50 mM Tris-HCl buffer (pH 8.5). After that, the samples were concentrated using a speed vacuum and desalted utilizing Pierce C18 zip tips. Desalted samples were dried and reconstituted with 49% Methanol/49% water/ 2% acetic acid (direct infusion solvent).

4.2.4 Protein Cross-Linking Experiments

The efficiency of the cross-linking reagent tested with three different proteins, Ubiquitin, Myoglobin, and BSA. 1 mM of standard protein was prepared and reacted with the cross-linking reagent in a different molar ratio (1:50, 1:100, or 1: 200) in PBS buffer (pH 7.4). The reaction performed at room temperature in thermomixer for 30 mins. After 30 mins, 50 mM Tris.HCL Buffer was utilized to quench the reaction. 3K MWCO protein concentrators were used to remove the hydrolyzed cross-linking reagent and exchange the buffer with 50 mM ammonium bicarbonate buffer (pH 8). The concentration of the protein was determined with the BCA protein assay before digestion.

4.2.5 Digestion of Cross-linked Proteins

Initially, the cross-linked proteins were reduced with 10 mM DDT and alkylated with 55 mM of iodoacetamide. Then sequence grade modified trypsin added at a 1:50 (w/w) ratio of trypsin to protein in 50 mM ammonium bicarbonate buffer. Enzymatic digestion was allowed to continue for 16 h at 37 °C with 500 rpm constant rotation. Then the digested samples were concentrated by a speed vacuum, desalted using ZipTip, and re-constitute in 0.1% FA solution in water and stored at -20 °C.[34-36]

4.2.6 Nano HPLC/ Nano-ESI-LTQ Analysis

Tryptic digests analyzed with LTQ Velos pro mass spectrometry combined with a UHPLC (UltiMate 3000, Dionex, USA). The digested peptides separated by a nano viper analytical C18 column. (Acclaim pepMap, 150 mm × 75 µm, 2 µm, 100 Å, Thermo Scientific, CA, USA). A 60 minutes gradient method performed to separate the digested peptides (0−3 min 4.0%B, 3−50 min 4.0−50.0% B, 50−50.1 min 50−90% B, 50.1−55 min 90% B, 55−55.1 min 90−4% B, 55.1−60 min 4% B; mobile phase A: 0.1% FA in water; mobile phase B: 0.1% FA in 95% acetonitrile, 5% water). 5 µl (partial injection mode) of digested samples injected at 300 nl/min flow.

The nano-electrospray ionization (ESI) source used with a fixed spray voltage at 2.4 kV and a heated capillary temperature at 275 °C. A full MS spectrum obtained in normal scan mode from 350 to 2000 m/z mass range. The data-dependent acquisition was performed using Electron transfer based dissociation to get the MS/MS spectra for the 5 most abundant ions. For ETD mode, reagent source temperature, reagent ion source CI pressure, emission current, and reagent ion-electron energy set to 110°C, 20 psi, 50 UA, and -70 v, respectively. The ETD reaction time maintained at 100 ms, with an isolation width of 2 Da. The charge state of the ions filtered to >2 for ETD-MS/MS mode.

4.2.7 Data Analysis

Data analysis of LC-MS^n spectra was carried out with our in-house developed software tool CLEAVE-XL. The detailed workflow of this Java coded CLEAVE-XL will be available soon as a published article. Our specially designed software can identify the cross-linked peptides automatically from the Dual cleavable (DUCCT/PC-DUCCT-Biotin) as well as single CID/ETD-cleavable cross-linkers. In the very beginning step, protein database inputted to the algorithm and theoretical database of the peptides were generated based on the setting of proteolysis and amino acid sequence provided by the input protein file. In next step is the creating of the cross-linked peptide database along with the fragment masses. This step accomplished by adding the cross-linker mass (HI-ETD-XL) as well as fragment residual masses with the generated peptide masses. The experimental ETD-MS/MS data incorporated as MGF (Mascot Generic Format) file. Then, ETD-MS/MS experimental data searched against the created theoretical database and a list of cross-linked peptides along with reduced charged precursor mass generated in ETD-output files. Intensity threshold of the MS/MS fragment peaks set to 30% to filter the unambiguous ID and reduce false-positive identification.

4.3 Result and discussions

4.3.1 Design and Synthesis of a Hydrazone Incorporated-ETD-Cleavable Cross-Linker (HI-ETD-XL)

In this study we demonstrated a reasonably compact ETD-cleavable cross-linker (spacer chain length, 14 Å) with different model peptides and proteins. The design of this cross-linker is shown in **Figure 4-1A.**

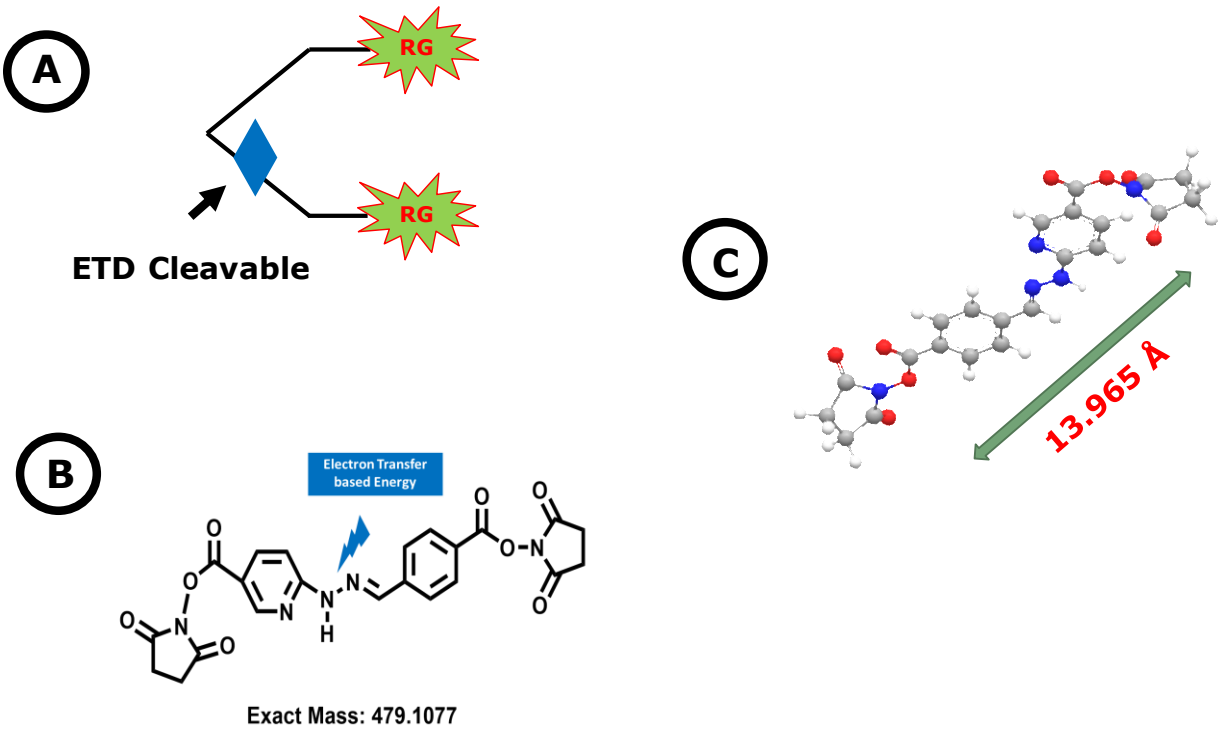


Figure 4- 1. The design and chemical structure of the ETD-Cleavable cross-linker (ETXL). (A) The design of the ETD cleavable cross-linker. (B) The chemical structure of ETD MS-cleavable cross-linker. (C) The structure of HI-ETD-XL is shown with spacer chain length (13.965 A). This NHS containing cross-linker reacts with the primary amine group of the N-terminal or Lysine residue. It contains one gas phase cleavable (N-N) bond, which preferentially cleaved by electron transfer-based dissociations before peptide backbone breakage. Selective cleavage of the Hydrazone Bond (N-N) by ETD was first reported by Jennifer Brodbelt group. Detail mechanism

of the cleavage described in Gardner et al [31]. Inspired by this study, we incorporated a hydrazine bond (N-N) bond in between two reactive NHS ester groups and synthesized a new gas-phase ETD-cleavable cross-linker, namely HI-ETD-XL **(Figure 4-1B).** HI-ETD-XL was synthesized in the solution phase synthesis process, as demonstrated in the supplementary section. The spacer length distance is quite reasonable (~14 Å) for this ETD cleavable cross-linker, making it suitable for interactome study as well structural studies of protein or protein complexes **(Figure 4-1C).** Considering the spacer length of this cross-linker and structural dynamics of the protein complexes, we estimated that a distance up to 35Å would be labeled.

4.3.2 Cross-Linked Peptide Identification Strategies

Similar to cleavable cross-linker, HI-ETD-XL cross-linking can also produce two types of cross-linked products: dead-end and inter cross-linked modified peptides. Inter cross-linked modified peptides provide essential information regarding spatial orientation and the distance constraints between the interacting residues. The protein identification workflow of the cross-linked peptides is illustrated in **Figure 4-2/Figure C-S1.** The cross-linker covalently connects the proteins with its interacting partner. Since HI-ETD-XL is an NHS-based cross-linker, it reacts with the nearby lysine residues of the protein.

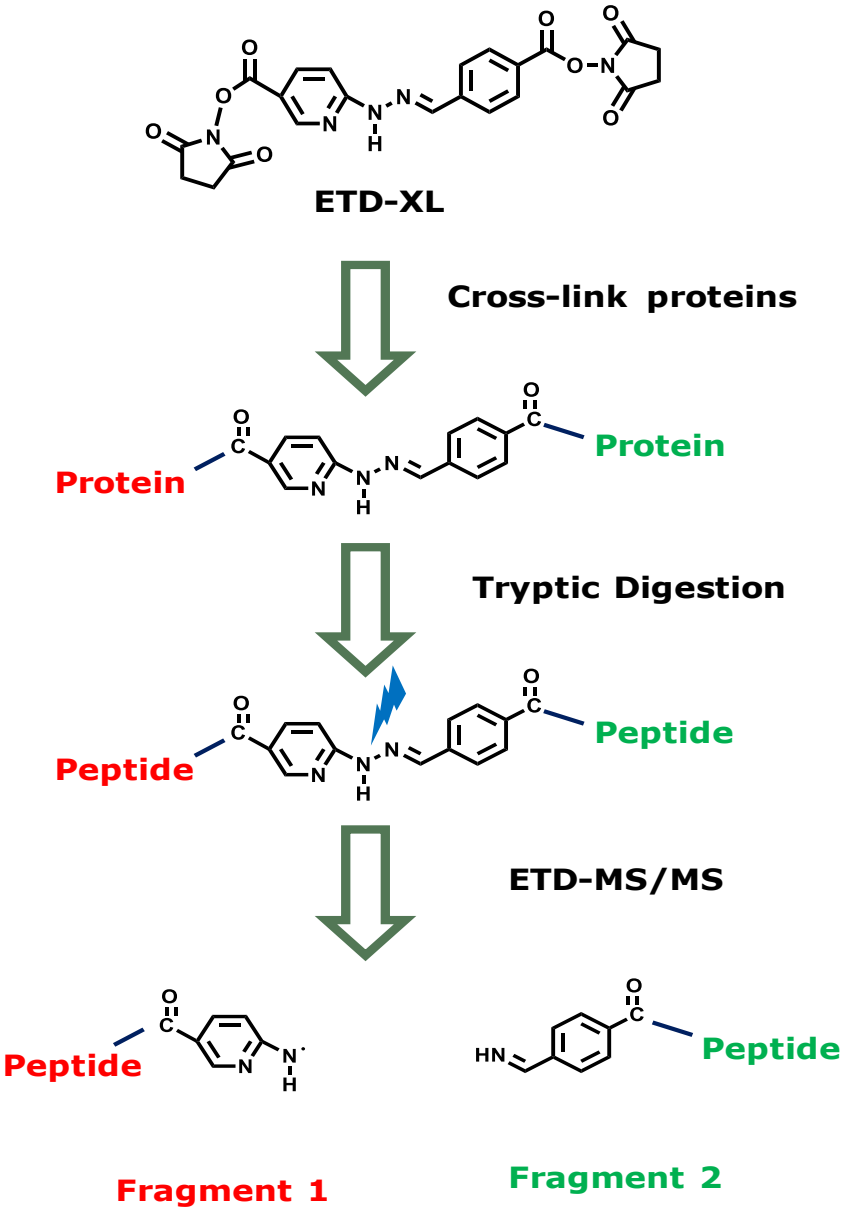


Figure 4-2. Structure of HI-ETD-XL with ETD cleaved signature fragments. The blue arrow points to the bond that is preferentially cleaved by ETD process. The workflow for cross-linking proteins is shown. After protein cross-linking, trypsin digest generates the cross-linked peptide for LC/MS^n analysis. ETD-MS/MS leads to selective cleavage of the N-N bond.

After tryptic digestion, cross-linked peptides is generated along with unmodified peptides. Finally, the cross-linked peptides are cleaved by ETD-MS/MS and further identified by CID-MS/MS tandem mass spectrometry. ETD preferentially cleaves the N-N bond (hydrazone) of the HI-ETD-XL cross-linked peptide and will generate two peptide peaks modified with cross-linker residual masses. Additionally, the generated charge-reduced ions will provide the charge states information of the precursor m/z if a low resolution mass spectrometer is used. Dead-end peptides usually produce one diagnostic ion peak in ETD-MS/MS, and further MS3rd experiment will confirm the sequence of the fragment ions **(Figure C-S1).**

4.3.3 Characterization of HI-ETD-XL Cross-Linked Model Peptides by MS^n Analysis

Initially, our synthesized cross-linker was characterized by NMR and Mass spectrum. The detailed description of the NMR spectra and Mass Spectrum provided in **Supporting Information**. ^{1}H & ^{13}C NMR Spectra confirm the structure of the cross-linker. Additionally, Full MS spectra provide the molecular weight (479.1077 Da) confirmation as well **(Figure C-S2).**

In Proof-of-Concept Studies, we tested the efficiency of cross-linker with two model peptides, namely Bradykinin (RPPGFSPFR) and Neurotensin (**pE-LYENKPRR(.** Bradykinin has a free N-terminal in its sequence but doesn't contain any Lysine residue. On the other hand, Neurotensin has one Lysine in its sequence, but its N terminal is blocked by a pyroglutamate group. We found that HI-ETD-XL efficiently reacted with both peptides and mostly formed inter-cross-linking products. The five charged states precursor of inter cross-linked peptide derived from the Bradykinin identified at m/z 474.63. After ETD-MS/MS selective cleavage was observed at N-N bond, and two diagnostic fragment peaks were produced at m/z 596.7 (2+) and 590.28 (2+). We also found the singly charge state of these modified peptides (m/z at 1180.65

and 1192.74) **(Figure 4-3).** Complete calculations of the cross-linked precursor and the fragment ion masses is provided in **Table C-S1.** Similar fragmentation pattern was observed for Neurotensin dimer. Signature fragment ions pinpointed the inter cross-linked dimer of the Neurotensin peptide **(Figure C-S3).**

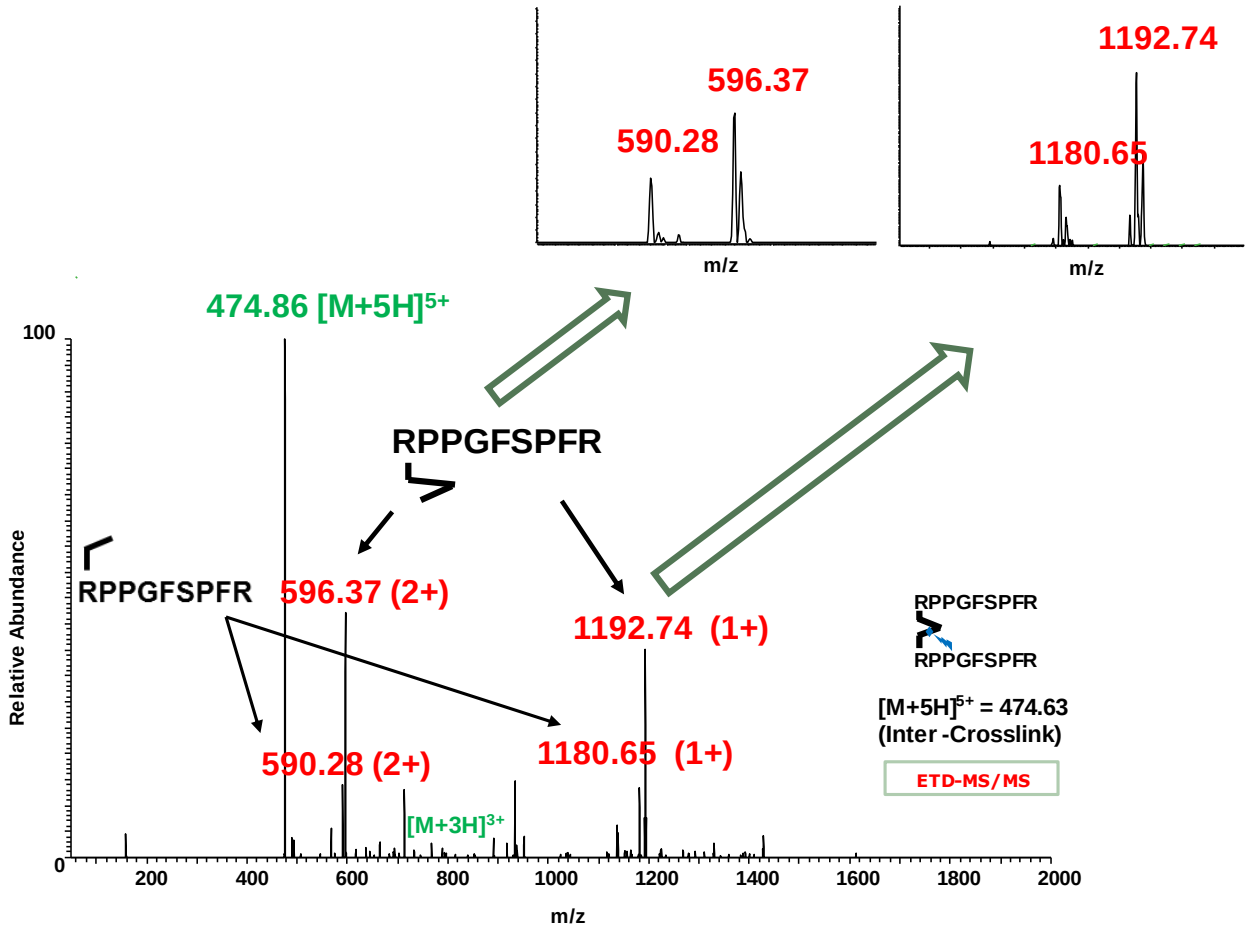


Figure 4-3. ETD-MS/MS of identified Bradykinin (RPPGFSPFR) Dimer (m/z 474.63, 5+). Showing the selective cleavage in N-N bond and generating 2 doubly charged diagnostic ions (m/z 590.28 and 596.37).

4.3.4 Characterization of HI-ETD-XL Cross-Linked Model Proteins by MS^n Analysis

To verify the efficiency of the HI-ETD-XL for protein cross-linking in vitro, we used three different model proteins for our experiment, namely bovine serum albumin (BSA), bovine ubiquitin, and equine heart myoglobin. In-solution digestion of the cross-linked proteins was performed and resulting cross-linked peptides were subjected to LC−MS^n analysis.

MS^n analysis of a representative inter-cross-linked Myoglobin peptide, IPIKYLEFISDAIIHVLHSK-----LFTGHPETLEKFDK (m/z 850.66, 5+) displayed in **Figure 4.** The ETD-MS/MS spectrum generated a pair of diagnostic ions after preferential cleavage of the hydrazone bond that revealed the confident identification of the cross-linked peptides. Further, MS/MS of the fragment ions m/z at 897.22 (2+) (LFTGHPETLEK*FDK) confirmed the sequence unambiguously **(Figure 4B).** Finally, MS2 diagnostic ions with MS3 sequencing confidently identified HI-ETD-XL inter cross-linked peptide (IPIKYLEFISDAIIHVLHSK…… LFTGHPETLEKFDK). Several dead-end peptides identified efficiently from myoglobin. An ETD MS2 spectrum of HI-ETD-XL dead-end from myoglobin displayed in **Figure C-S4B.** The triply charged precursor ion at 591.51 m/z confidently identified as HGTVVLTALGGILKK with a hydrolyzed dead end.

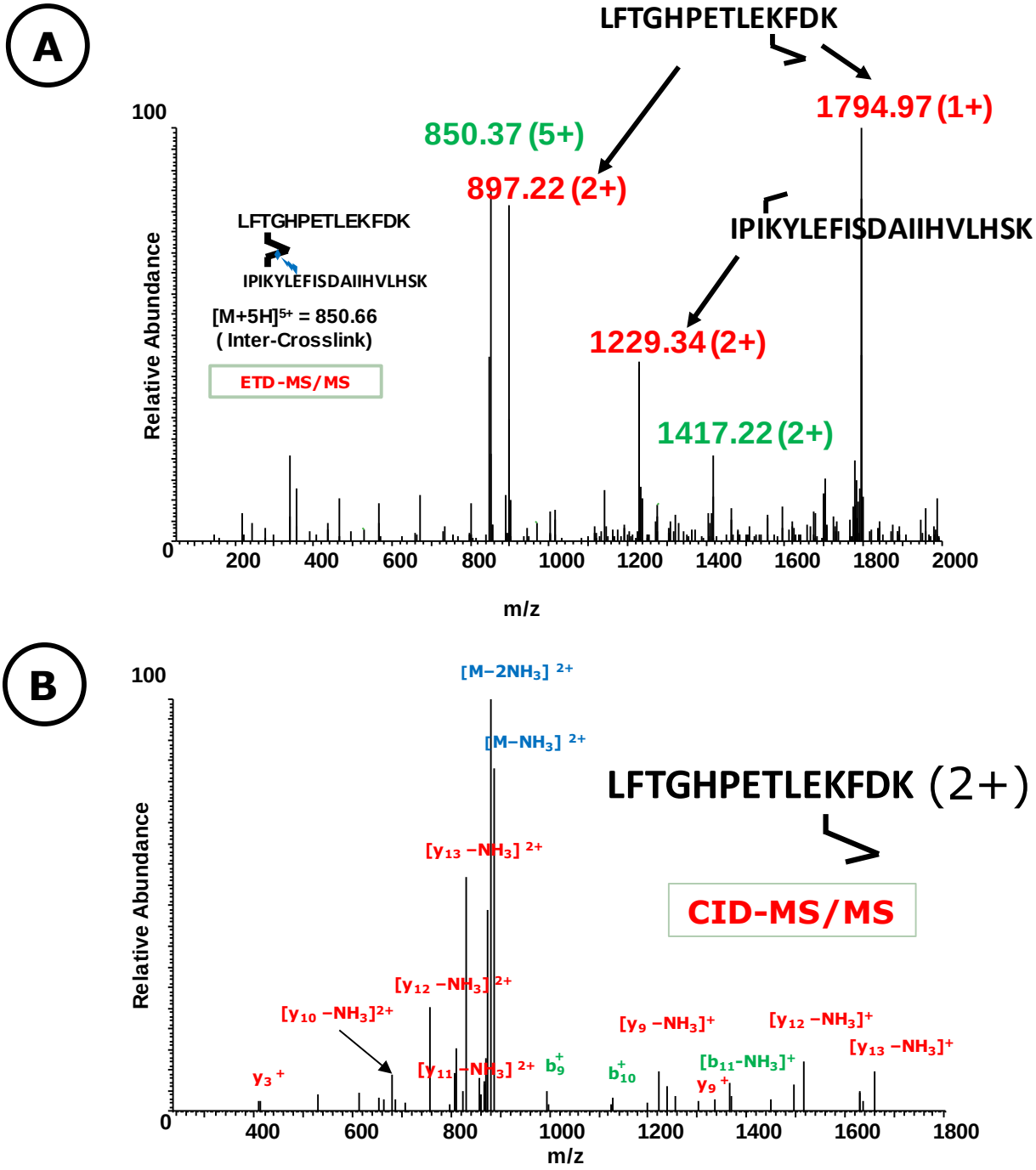


Figure 4-4. MS^n Analysis of identified inter cross-linked peptides derived from Myoglobin protein. (A) ETD-MS/MS of cross-linked peptide IPIKYLEFISDAIIHVLHSK----- LFTGHPETLEKFDK (m/z 850.66, 5+). (B) CID-MS^3 of one of the diagnostic fragment ions derived from ETD product ions.

After validation in a small protein, we evaluate the efficiency of HI-ETD-XL in a large protein, BSA (~66 KDa). Total of nine inter cross-linked peptides identified from BSA after HI-ETD-XL cross-linking experiment **(Table C-S2).** An example of four-charge state BSA cross-linked peptide (ALKAWSVAR--------LAKEYEATLEECCAK) is provided in **Figure 4-5.**

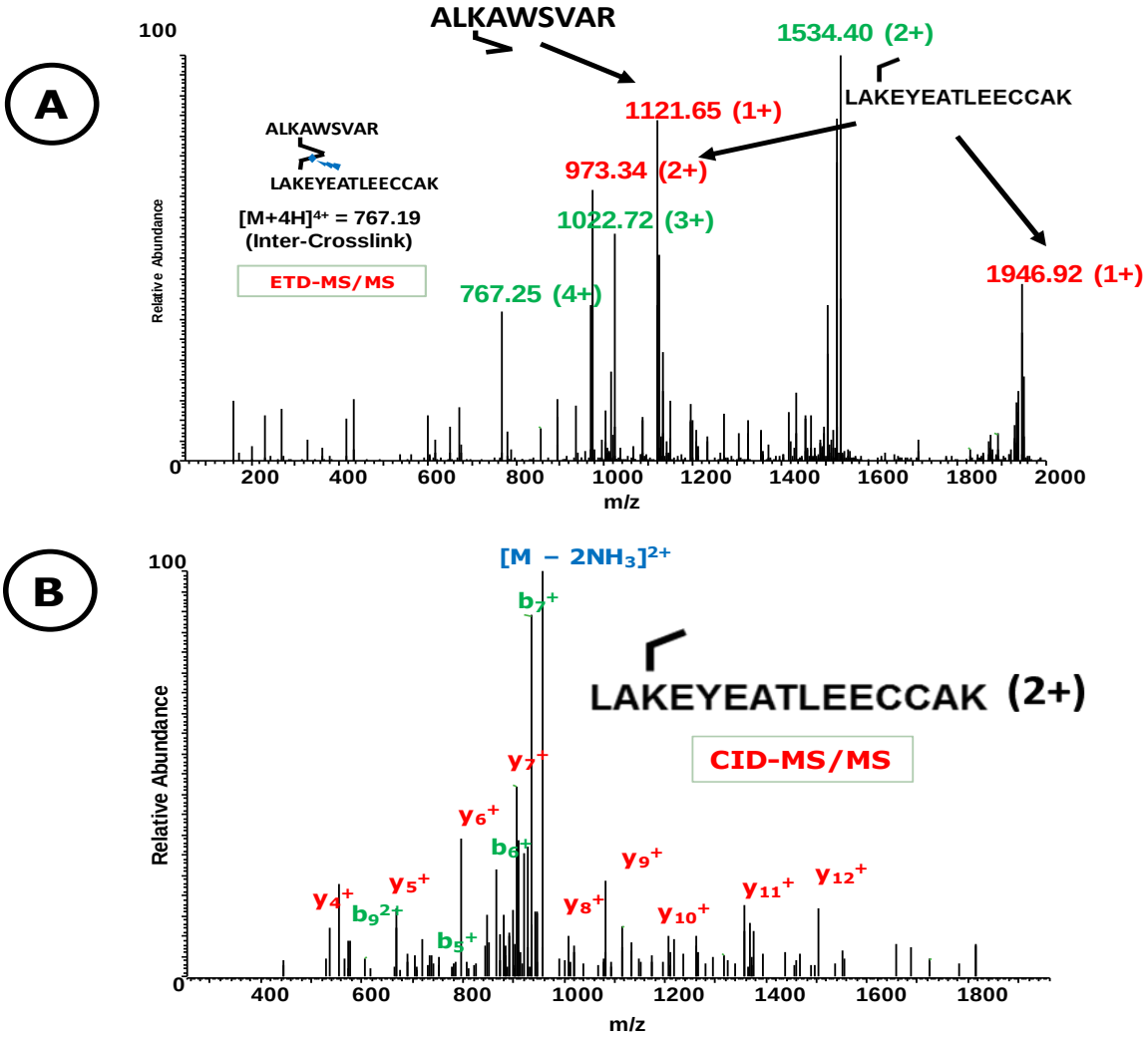


Figure 4-5. (A) Characteristic fragmentation pattern of HI-ETD-XL cross-linked peptide (ALKAWSVAR--------LAKEYEATLEECCAK) derived from BSA. 2 signature ions (m/z 973.34 and m/z 1121.65) are produced by ETD-MS/MS. (B) MS3 spectrum of one of the MS2 fragment ions (m/z 973.34) is demonstrated.

After applying Electron transfer based energy, the hydrazone bond was cleaved preferentially and produced two signature ions, [(m/z = 973.34 (2+) and m/z = 1121.65 (1+)]. Further CID-MS3 experiment was performed to confirm the cross-linked peptide sequence. One of the MS/MS of the signature fragment peaks (LAKEYEATLEECCAK, 2+) derived from the inter peptide cross-linking is presented in **Figure 5B.**

Again, one of the examples of identified dead-end peptides derived from BSA displayed in **Figure C-S4A.** High abundant diagnostic ion peaks along with reduced charged precursor ions unambiguously pinpointed the triply charged dead-end peptide K*QTALVELLK. We know that complete fragmentation of the precursor ions not commonly happened in ETD-MS/MS. However, we observed that more highly charged cross-linked peptide derived from the HI-ETD-XL showed efficient dissociation and high abundant signature mass pairs. Sometimes, precursor ion and its charge reduced ion masses are predominant in the MS2 spectra. Charge-reduced precursor ion is an excellent indicator of the charge states of the precursor ions. In **Figure 5A,** the HI-ETD-XL cross-linked peptide (ALKAWSVAR--------LAKEYEATLEECCAK) from BSA displayed high intense reduced charge precursors along with the diagnostic ion peaks. A complete calculated masses and m/z values used to assign cross-linked peptides derived from Myoglobin and BSA are provided in **Table C-S3** and **Table C-S4,** respectively.

Figure 6 presents an example of a five-charged state intercross-linked peptide (m/z at 781.56) identified from bovine ubiquitin. Two high abundant characteristic fragment pairs, IQDK*EGIPPDQQR (m/z at 1131.09, 2+) and TLSDYNIQK*ESTLHLVLR (m/z at 822.61, 2+), were generated from the ETD-MS/MS is displayed in **Figure 4-6**. The high intensities of these diagnostic peaks confirmed that ETD favors N-N (Hydrazone) cleavage. MS3 analysis of

diagnostic ions confirmed its sequence as IQDK*EGIPPDQQR and TLSDYNIQK*ESTLHLVLR, respectively **(Figure C-S5A and S5B)**.

Three inter cross-linked peptides were identified confidently using HI-ETD-XL from Ubiquitin. The list of inter-cross-linked peptides is shown in **Table C-S5.** The calculated masses of cross-linked precursors and fragment ions is provided in the Supplementary Information **(Table C-S6).**

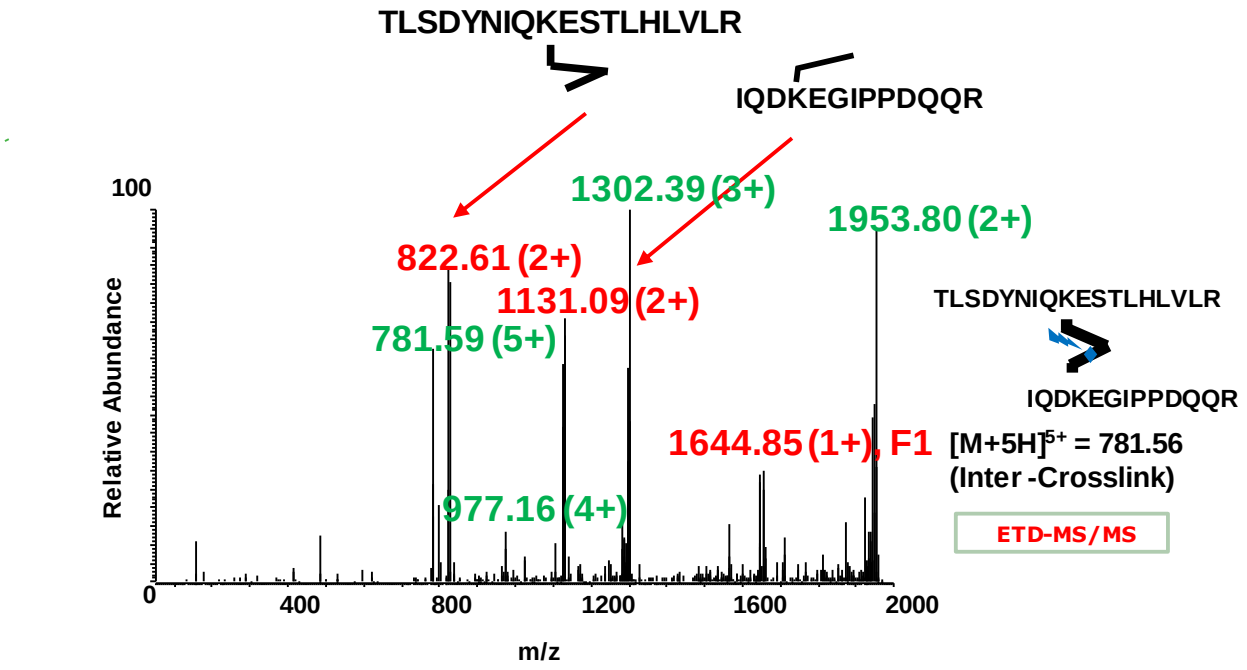


Figure 4-6: Example of one of the inter cross-linked peptides identified from Ubiquitin. ETD-MS/MS of crosslinked peptide (IQDKEGIPPDQQR--------TLSDYNIQKESTLHLVLR) (m/z 781.56, 5+).

4.3.5 HI-ETD-XL Cross-Linking Maps of Ubiquitin, Myoglobin and BSA

Inter cross-linked peptides identified by our ETD-XL strategy was used to generate protein cross-linking structural maps of three model proteins Ubiquitin, Myoglobin and BSA. The distances between the identified cross-linked Lysine residue were calculated and matched with the crystal structure of bovine Ubiquitin (PDB: 1V80) illustrated in **Figure C-S6.** Inter-linked Lysines identified within Ubiquitin had approximate distances in the range of 20-21 Å **(Table C-S5).** [LIFAGK$_{48}$QLEDGR ……. LIFAGK$_{48}$QLEDGR], is an identified inter cross-linked peptide from ubiquitin dimer. Lysine 48 located ideally in a hydrophobic patch function as a site for chain linkage for polyubiquitination and plays a vital role in protein interactions with other substrates [37].

Nine uniquely identified cross-linked peptides from equine Myoglobin listed in **Table C-S7.** Deposited crystal structure of myoglobin was used as a model to verify the cross-linked sites and their calculated alpha carbon distances (Cα-Cα distances) **(Figure C-S7).** We consider the spacer length of HI-ETD- XL (~14 Å), backbone flexibility of protein, and found that estimated upper limit for the Cα-Cα distances between HI-ETD-XL cross-linked lysine residues is equal to ~35 Å. Cα−Cα ranging from 8.1 to 34.5 Å have reported in all the interlinked peptides identified in Myoglobin. These distances reside well within our cross-linkers anticipated range (~35 Å). Similar cross-linked maps was generated for BSA and provided in **Figure C-S8**.

4.4 Conclusions

In this manuscript, we reported development and characterization of an ETD-based cross-linker HI-ETD-XL. HI-ETD-XL cross-linked peptides preferentially cleaved by electron transfer based dissociation and showed characteristic fragmentation pattern in MS2 spectra. The signature fragmentation pattern, along with MS3 sequencing identified the cross-linked peptides quickly and unambiguously. We showed the fragmentations in several model proteins and peptides. We believe this could be an effective supplemental structural tool kit for protein cross-linking and mass spectrometry field. Another interesting fact is that hydrazone was shown to preferentially cleave by 351 nm UV laser. Fragmentations of peptides both in UVPD and ETD will also make this an efficient dual cleavable cross-linkers. We believe this cross-linker can be used for benchmarking interactome studies and protein structural analysis along with the CID cleavable cross-linkers.

4.5 References:

[1] C. Iacobucci, C. Piotrowski, R. Aebersold, B.C. Amaral, P. Andrews, K. Bernfur, C. Borchers, N.I. Brodie, J.E. Bruce, Y. Cao, S. Chaignepain, J.D. Chavez, S. Claverol, J. Cox, T. Davis, G. Degliesposti, M.Q. Dong, N. Edinger, C. Emanuelsson, M. Gay, M. Gotze, F. Gomes-Neto, F.C. Gozzo, C. Gutierrez, C. Haupt, A.J.R. Heck, F. Herzog, L. Huang, M.R. Hoopmann, N. Kalisman, O. Klykov, Z. Kukacka, F. Liu, M.J. MacCoss, K. Mechtler, R. Mesika, R.L. Moritz, N. Nagaraj, V. Nesati, A.G.C. Neves-Ferreira, R. Ninnis, P. Novak, F.J. O'Reilly, M. Pelzing, E. Petrotchenko, L. Piersimoni, M. Plasencia, T. Pukala, K.D. Rand, J. Rappsilber, D. Reichmann, C. Sailer, C.P. Sarnowski, R.A. Scheltema, C. Schmidt, D.C. Schriemer, Y. Shi, J.M. Skehel, M. Slavin, F. Sobott, V. Solis-Mezarino, H. Stephanowitz, F. Stengel, C.E. Stieger, E. Trabjerg, M. Trnka, M. Vilaseca, R. Viner, Y. Xiang, S. Yilmaz, A. Zelter, D. Ziemianowicz, A. Leitner, A. Sinz, First Community-Wide, Comparative Cross-Linking Mass Spectrometry Study, Anal Chem 91(11) (2019) 6953-6961.
[2] A. Leitner, T. Walzthoeni, A. Kahraman, F. Herzog, O. Rinner, M. Beck, R. Aebersold, Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics, Mol Cell Proteomics 9(8) (2010) 1634-49.
[3] A. Leitner, L.A. Joachimiak, P. Unverdorben, T. Walzthoeni, J. Frydman, F. Forster, R. Aebersold, Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes, Proc Natl Acad Sci U S A 111(26) (2014) 9455-60.

[4] S. Bian, S. Chowdhury, Profiling Protein-Protein Interactions and Protein Structures Using Chemical Cross-linking and Mass Spectrometry, J Austin J Biomed Eng 1(4) (2014) 1017.
[5] J.K. Chakrabarty, A.G. Naik, M.B. Fessler, G.R. Munske, S.M. Chowdhury, Differential Tandem Mass Spectrometry-Based Cross-Linker: A New Approach for High Confidence in Identifying Protein Cross-Linking, Analytical Chemistry 88(20) (2016) 10215-10222.
[6] S.M. Chowdhury, X. Du, N. Tolic, S. Wu, R.J. Moore, M.U. Mayer, R.D. Smith, J.N. Adkins, Identification of cross-linked peptides after click-based enrichment using sequential collision-induced dissociation and electron transfer dissociation tandem mass spectrometry, Anal Chem 81(13) (2009) 5524-32.
[7] N.I. Brodie, K.I. Popov, E.V. Petrotchenko, N.V. Dokholyan, C.H. Borchers, Solving protein structures using short-distance cross-linking constraints as a guide for discrete molecular dynamics simulations, Sci Adv 3(7) (2017) e1700479-e1700479.
[8] D.B. Lima, J.T. Melchior, J. Morris, V.C. Barbosa, J. Chamot-Rooke, M. Fioramonte, T.A.C.B. Souza, J.S.G. Fischer, F.C. Gozzo, P.C. Carvalho, W.S. Davidson, Characterization of homodimer interfaces with cross-linking mass spectrometry and isotopically labeled proteins, Nat Protoc 13(3) (2018) 431-458.
[9] M.W. Gardner, J.S. Brodbelt, Ultraviolet photodissociation mass spectrometry of bis-aryl hydrazone conjugated peptides, Anal Chem 81(12) (2009) 4864-72.
[10] S.M. Chowdhury, G.R. Munske, X. Tang, J.E. Bruce, Collisionally activated dissociation and electron capture dissociation of several mass spectrometry-identifiable chemical cross-linkers, Anal Chem 78(24) (2006) 8183-93.
[11] A. Kao, C.L. Chiu, D. Vellucci, Y. Yang, V.R. Patel, S. Guan, A. Randall, P. Baldi, S.D. Rychnovsky, L. Huang, Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes, Mol Cell Proteomics 10(1) (2011) M110 002212.
[12] E.V. Petrotchenko, J.J. Serpa, C.H. Borchers, An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics, Mol Cell Proteomics 10(2) (2011) M110 001420.
[13] X. Tang, G.R. Munske, W.F. Siems, J.E. Bruce, Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions, Anal Chem 77(1) (2005) 311-8.
[14] E.J. Soderblom, M.B. Goshe, Collision-induced dissociative chemical cross-linking reagents and methodology: Applications to protein structural characterization using tandem mass spectrometry analysis, Analytical Chemistry 78(23) (2006) 8059-8068.
[15] F. Liu, M.B. Goshe, Combinatorial electrostatic collision-induced dissociative chemical cross-linking reagents for probing protein surface topology, Anal Chem 82(14) (2010) 6215-23.
[16] C.B. Gutierrez, S.A. Block, C. Yu, S.M. Soohoo, A.S. Huszagh, S.D. Rychnovsky, L. Huang, Development of a Novel Sulfoxide-Containing MS-Cleavable Homobifunctional Cysteine-Reactive Cross-Linker for Studying Protein-Protein Interactions, Anal Chem 90(12) (2018) 7600-7607.
[17] F. Liu, C. Wu, J.V. Sweedler, M.B. Goshe, An enhanced protein crosslink identification strategy using CID-cleavable chemical crosslinkers and LC/MSn analysis, Proteomics 12(3) (2012) 401-405.
[18] C. Hage, F. Falvo, M. Schafer, A. Sinz, Novel Concepts of MS-Cleavable Cross-linkers for Improved Peptide Structure Analysis, J Am Soc Mass Spectrom 28(10) (2017) 2022-2038.

[19] R.A. Zubarev, D.M. Horn, E.K. Fridriksson, N.L. Kelleher, N.A. Kruger, M.A. Lewis, B.K. Carpenter, F.W. McLafferty, Electron capture dissociation for structural characterization of multiply charged protein cations, Anal Chem 72(3) (2000) 563-73.
[20] J.E. Syka, J.J. Coon, M.J. Schroeder, J. Shabanowitz, D.F. Hunt, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, Proc Natl Acad Sci U S A 101(26) (2004) 9528-33.
[21] K.O. Zhurov, L. Fornelli, M.D. Wodrich, U.A. Laskay, Y.O. Tsybin, Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis, Chem Soc Rev 42(12) (2013) 5014-30.
[22] A. Chi, C. Huttenhower, L.Y. Geer, J.J. Coon, J.E. Syka, D.L. Bai, J. Shabanowitz, D.J. Burke, O.G. Troyanskaya, D.F. Hunt, Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry, Proc Natl Acad Sci U S A 104(7) (2007) 2193-8.
[23] N.M. Riley, J.J. Coon, The Role of Electron Transfer Dissociation in Modern Proteomics, Anal Chem 90(1) (2018) 40-64.
[24] C.K. Frese, A.F. Altelaar, M.L. Hennrich, D. Nolting, M. Zeller, J. Griep-Raming, A.J. Heck, S. Mohammed, Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos, J Proteome Res 10(5) (2011) 2377-88.
[25] F. Liu, P. Lossl, R. Scheltema, R. Viner, A.J.R. Heck, Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification, Nat Commun 8 (2017) 15473.
[26] H.H. Koolen, A.F. Gomes, N.V. Schwab, M.N. Eberlin, F.C. Gozzo, Imidate-based cross-linkers for structural proteomics: increased charge of protein and peptide ions and CID and ECD fragmentation studies, J Am Soc Mass Spectrom 25(7) (2014) 1181-91.
[27] M.J. Trnka, A.L. Burlingame, Topographic studies of the GroEL-GroES chaperonin complex by chemical cross-linking using diformyl ethynylbenzene: the power of high resolution electron transfer dissociation for determination of both peptide sequences and their attachment sites, Mol Cell Proteomics 9(10) (2010) 2306-17.
[28] F.X. Chu, S.O. Shan, D.T. Moustakas, F. Alber, P.F. Egea, R.M. Stroud, P. Walter, A.L. Burlingame, Unraveling the interface of signal recognition particle and its receptor by using chemical cross-linking and tandem mass spectrometry, P Natl Acad Sci USA 101(47) (2004) 16454-16459.
[29] B. Zhao, C.P. Reilly, J.P. Reilly, ETD-Cleavable Linker for Confident Cross-linked Peptide Identifications, J Am Soc Mass Spectrom 30(9) (2019) 1631-1642.
[30] M.A. Lauber, J.P. Reilly, Novel amidinating cross-linker for facilitating analyses of protein structures and interactions, Anal Chem 82(18) (2010) 7736-43.
[31] M.W. Gardner, J.S. Brodbelt, Preferential cleavage of N-N hydrazone bonds for sequencing bis-arylhydrazone conjugated peptides by electron transfer dissociation, Anal Chem 82(13) (2010) 5751-9.
[32] A. Bishop, J.S.J.I.J.o.M.S. Brodbelt, Selective cleavage upon ETD of peptides containing disulfide or nitrogen–nitrogen bonds, 378 (2015) 127-133.
[33] L. Vasicek, J.P. O'Brien, K.S. Browning, Z. Tao, H.W. Liu, J.S. Brodbelt, Mapping protein surface accessibility via an electron transfer dissociation selectively cleavable hydrazone probe, Mol Cell Proteomics 11(7) (2012) O111 015826.

[34] M.S.K. Wanigasekara, X. Huang, J.K. Chakrabarty, A. Bugarin, S.M. Chowdhury, Arginine-Selective Chemical Labeling Approach for Identification and Enrichment of Reactive Arginine Residues in Proteins, ACS Omega 3(10) (2018) 14229-14235.
[35] Z. Fang, Y.Z. Baghdady, K.A. Schug, S.M. Chowdhury, Evaluation of Different Stationary Phases in the Separation of Inter-Cross-Linked Peptides, J Proteome Res 18(4) (2019) 1916-1925.
[36] A.H.M. Kamal, J.K. Chakrabarty, S.N. Udden, M.H. Zaki, S.M. Chowdhury, Inflammatory Proteomic Network Analysis of Statin-treated and Lipopolysaccharide-activated Macrophages, Scientific reports 8(1) (2018) 164.
[37] A. Haririnia, M. D'Onofrio, D.J.J.o.m.b. Fushman, Mapping the interactions between Lys48 and Lys63-linked di-ubiquitins and a ubiquitin-interacting motif of S5a, 368(3) (2007) 753-766.

Chapter 5

Conclusions

Mass spectrometry-based chemical cross-linking approach is one of the widely used techniques to determine protein structure and protein-protein interactions. Although inter cross-linked peptides provide more structural information of proteins, their low abundance in complex mixtures and mass spectrometry data complexity restrict their confident identification. In this thesis, we demonstrated a series of MS-based cleavable cross-linking strategies that overcome those challenges and will contribute significantly in interactome and protein structural biology research.

In chapter 2, we demonstrate our developed dual cleavable cross-linking strategy (DUCCT). This cross-linker contains two differential cleavable bonds that can be cleaved by 2 different tandem mass spectrometry techniques. Cross-linked peptides are confidently identified based on the diagnostic ions generated in two different mass spectra (CID-MS/MS and ETD-MS/MS). In proof of concept, we test our strategy with standard peptides and proteins and immune cells and identify the cross-linked peptides unambiguously.

In Chapter 3, we describe the enrichment-based next generation DUCCT cross-linking technique, PC-DUCCT-Biotin, and introduce our in house developed cross-linking software tool Cleave-XL. PC-DUCCT-Biotin contains an affinity group (biotin) that assist in enriching cross-linked peptides using avidin-biotin affinity chromatography. It also contains a photolabile group that helps to discharge cross-linked peptides by UV light. Our proto-type software can automatically generate high confident cross-linked peptides list using a single or Dual cleavable cross-linker. Here, the efficiency of our developed software is demonstrated not only with model peptides but also, with a large-scale experiment using targeted database search.

In Chapter 4, we report a compact ETD-cleavable cross-linking strategy, HI-ETD-XL. ETD fragmentation mostly depends on the precursor ion charge density. HI-ETD-XL contributes an additional charge to precursor ions and eventually increase the charge states. Again, our HI-ETD-XL contains a labile hydrazone bond (N-N) bond in its structure that selectively cleaved by electron transfer based dissociation (ETD) and generate diagnostic ions in MS/MS spectra. We believe that this strategy will facilitate to illuminate the 3D structure of macromolecular protein complexes.

Several large-scale cellular level studies are underway using different generations of DUCCT cross-linkers. We perform a DUCCT-crosslinked study with RAW 264.7 cell line to test the capability of the cross-linker as well as in house developed software Cleave-XL. Cleave-XL is presently capable of searching for protein structures and protein complexes consisting of 5 to 10 proteins. Software development is under progress for large-scale interactome analysis. A complete structural analysis of TLR4 is currently underway using our novel strategies. We believe that our developed DUCCT cross-linking strategies will elucidate the signaling pathways of several TLRs.

Appendix A

Supporting Information for Chapter 2

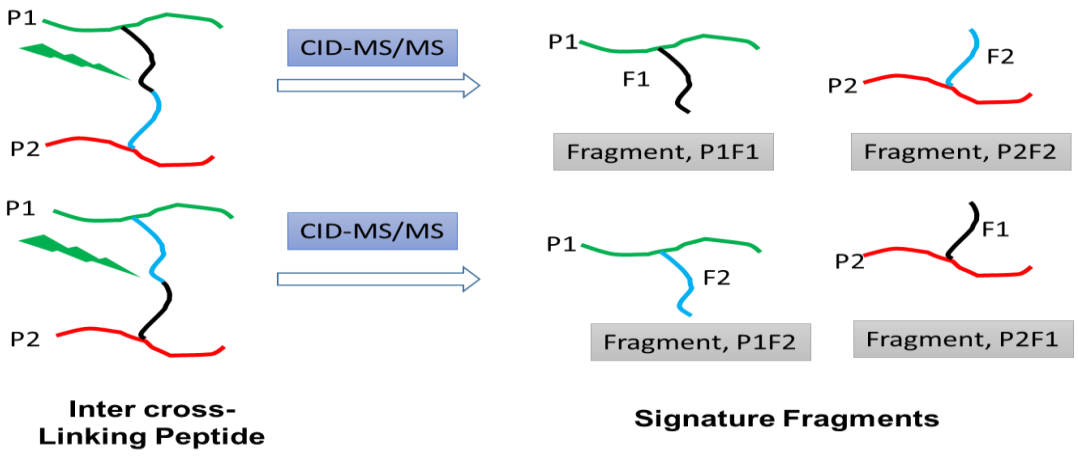


Figure A-S1: Fragmentation of inter-cross-liked peptides by Collision Induced Dissociation (CID). CID cleaves in a specific site (DP bond) of cross-linker and produces two peptide ions (P1 and P2) with fragmented cross-linker residue (F1 and F2) (signature fragments).

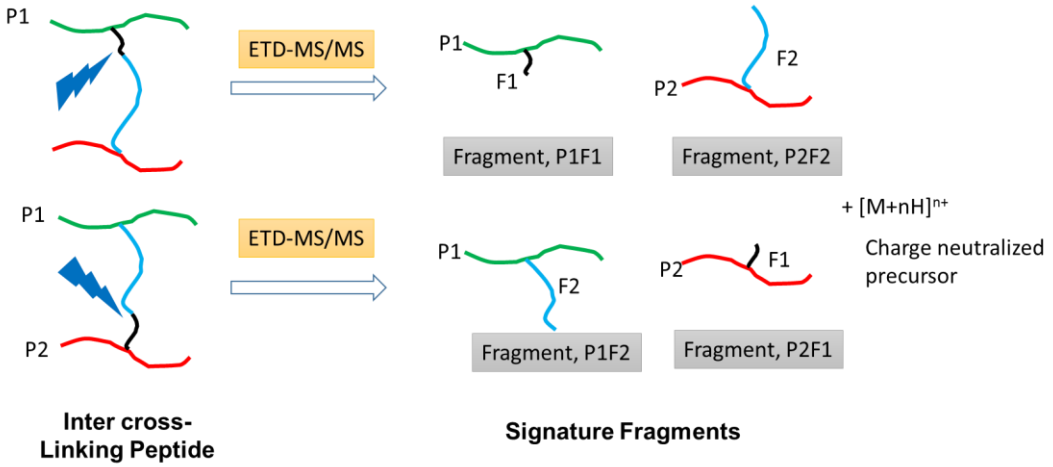


Figure A-S2: Fragmentation of inter-cross-linked peptides by Electron Transfer Dissociation (ETD). ETD cleaves in different site (N-N bond) than CID and produces two peptide ions (P1 and P2) with added different cross-linker residues (F1 and F2) (signature fragments) and neutralized precursor ions.

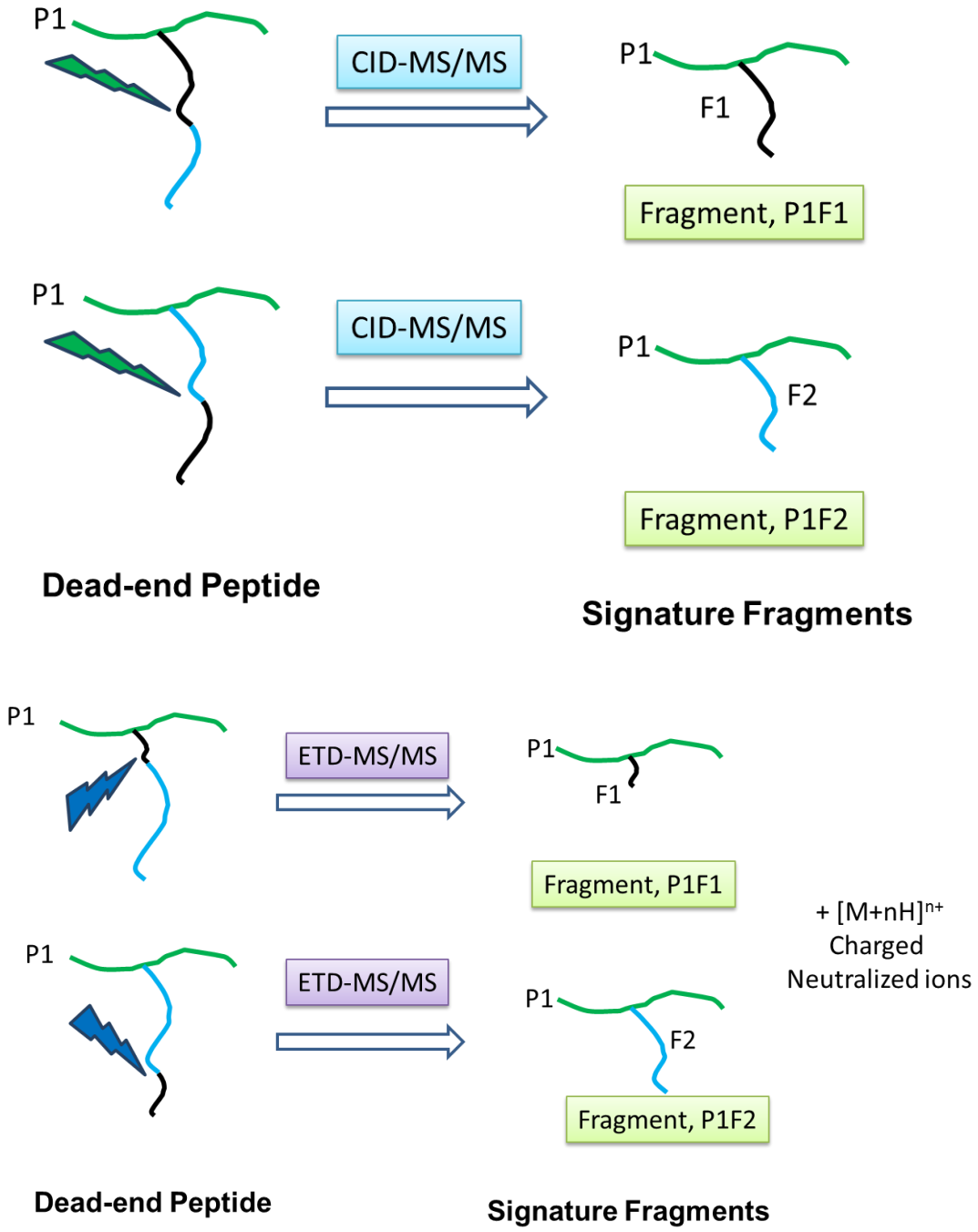
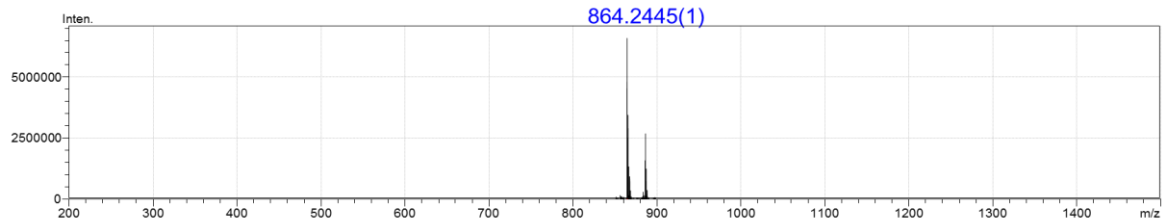


Figure A-S3: CID (top) and ETD (bottom) fragmentation patterns of dead-end peptides (one-end hydrolyzed). CID and ETD generate two complementary fragment signature ions of one peptide.



	Theoretical	Experimental	Mass Accuracy (ppm)
Cross-linker	864.2431	864.2445	1.6199

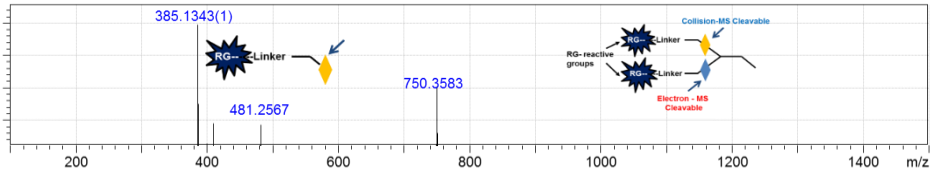


Figure A-S4: ESI-IT-TOF-MS measurement of CID and ETD based cross-linker ($[M+H]^+$ = 864.2421). The molecular mass of the cross-linker is 863.2358 Da. We observed the mass in the MS as the most abundant peak. Efficient fragmentation is shown in Asp-Pro peptide bond by CID. Cleavage mass, ($[M+H]^+$ = 385.1343, mass accuracy 2.9 ppm). RG-reactive groups.

Table A-S1: List of calculated masses and m/z values used to identify cross-linked reaction products. All the experimental m/z's are average due to the utilization of a LTQ Velos Pro MS.

	Exact Mass (2 decimal)	*m/z*
Cross-linker (DUCCT)	863.24	864.24 (1+)
CID Fragments	384.13 479.11	385.13 (1+) 480.11 (1+)
NHS (reactive group)	114.02	
Neurotensin + DUCCT (Inter crosslink)	(1671.92×2)(2 peptides) +863.24(Cross-linker)-(114.02×2) (2 leaving group)-2 (2 Hydrogen)=3977.04	1326.68 (3+) 995.26 (4+) 796.41 (5+)
Neurotensin + DUCCT (Dead-end)	1671.92 (peptide)+ 863.24(Cross-linker)-(114.02×2)(2 leaving group) + 17 (-OH) -1 (1 Hydrogen) = 2323.12	2324.12 (1+) 1162.56 (2+) 775.37 (3+)
Neurotensin + Fragment residue mass 1 (CID)	1671.92 (peptide)+ 270.11 (Fragment1)-1(1Hydrogen) =1940.03	1941.03 (1+) 971.02(2+) 647.68 (3+)
Neurotensin + Fragment residue mass 2 (CID)	1671.92 (peptide)+ 365.09 (Fragment 2)-1(1Hydrogen) =2036.01	2037.02 (1+) 1019.01 (2+) 679.67 (3+)
Neurotensin + Fragment residue mass 1 (ETD)	1671.92 (peptide)+ 504.16 (Fragment 1)-1(1Hydrogen) =2075.08	2176.08 (1+) 1088.54 (2+) 726.03 (3+)
Neurotensin + Fragment residue mass 2 (ETD)	1671.92 (peptide)+ 132.04 (Fragment 2)-1(1Hydrogen) =1802.96	1803.96(1+) 902.48(2+) 601.99(3+)

CID-MS/MS

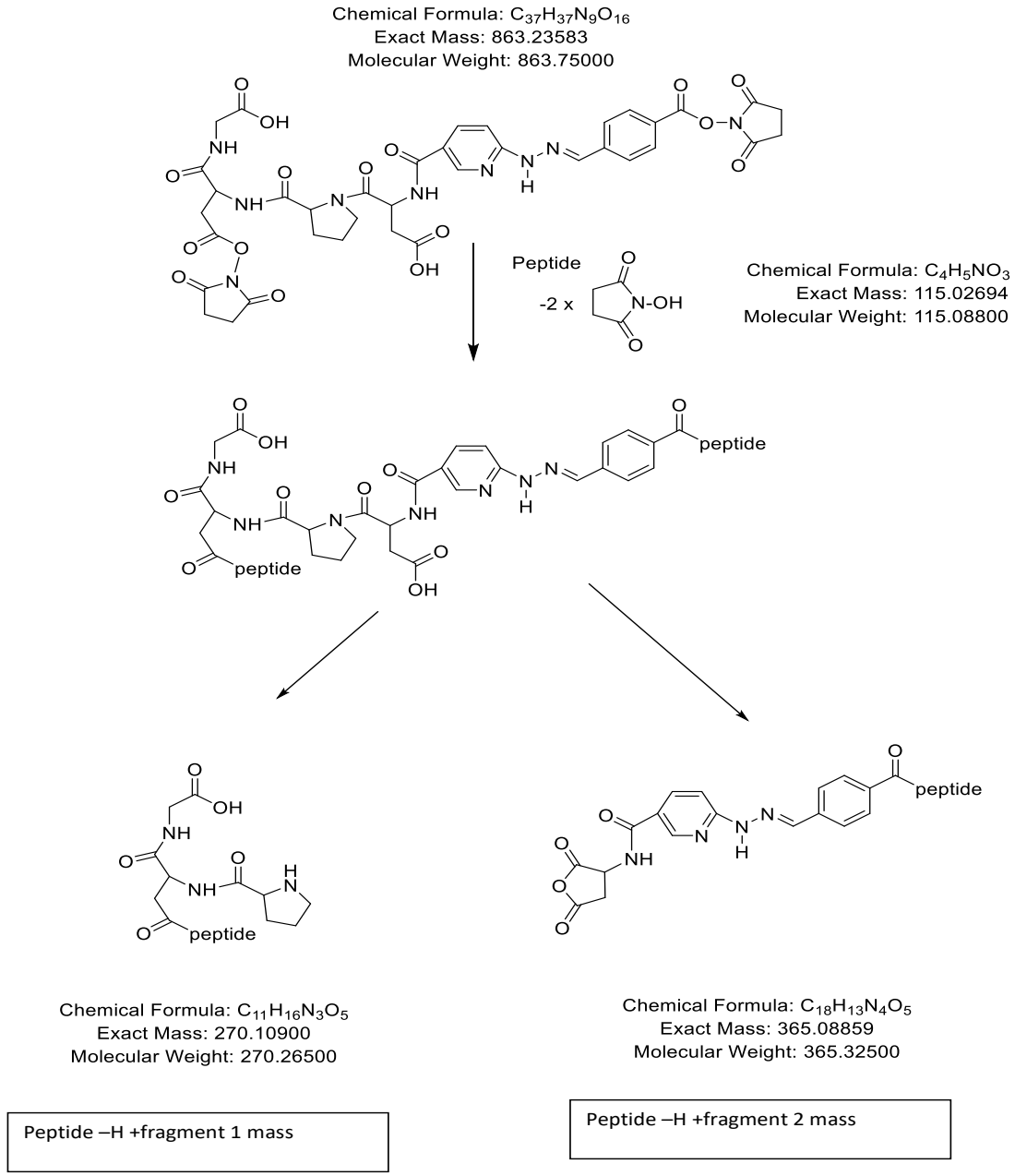


Figure A-S5: Structure of DUCCT with CID cleaved signature fragments. DUCCT has two reacting groups (NHS ester). Loss of NHS happened during the reaction with peptides. CID cleaves Asp-Pro peptide bond of DUCCT and generates 2 fragments.

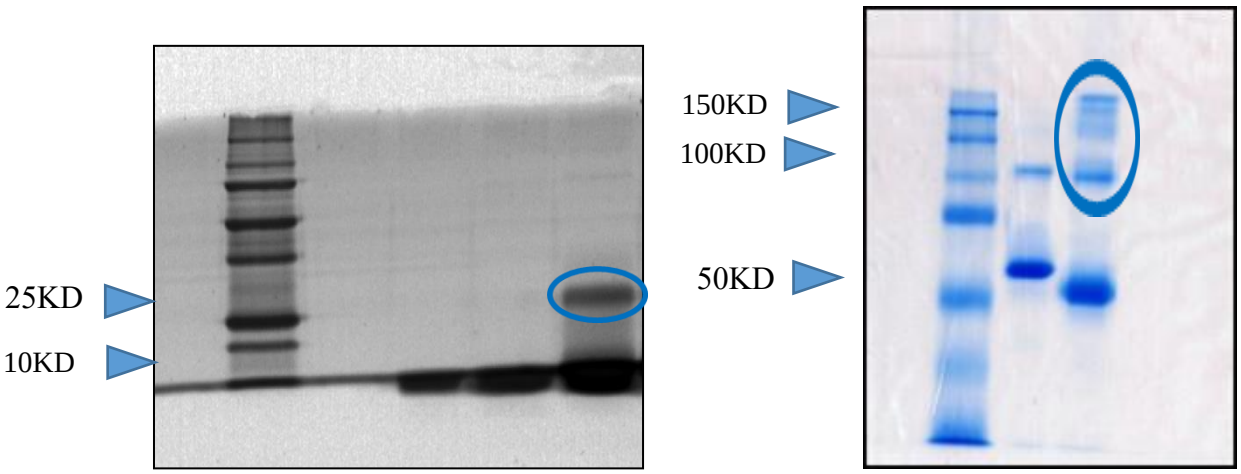


Figure A-S6: SDS PAGE gel image of cross-linked Ubiquitin (left) and BSA (right). A distinct higher molecular weight band of cross-linked protein is the primary evidence of efficient labeling of DUCCT.

Table A-S2: List of calculated masses and *m/z* value used to assign cross-linked products (Ubiquitin)

	Exact Mass (2 decimal)	m/z
LIFAGKQLEDGR------- TLSDYNIQKESTLHLVLR (inter-peptide cross-linked ubiquitin)	(1346.74+2130.15)(2 peptides)+863.24(Cross-linker)-(114.02×2) (2 leaving gr.)-2.00 (2 Hydrogen)=4110.09	4111.09 (1+) 823.02 (5+) 686.02 (6+)
LIFAGKQLEDGR (Peptide 1) + Fragment residue mass 1 (CID)	1346.74(peptide)+ 270.11 (Fragment1)-1(1Hydrogen) = 1615.85	1616.85 (1+) 808.93 (2+) 539.62 (3+) 404.96(4+)
TLSDYNIQKESTLHLVLR (Peptide 2) + Fragment residue mass 2 (CID)	2130.15 (peptide)+ 365.09 (Fragment 2)-1(1Hydrogen) =2494.24	2495.24 (1+) 1248.12 (2+) 832.41 (3+) 624.56 (4+)
LIFAGKQLEDGR (Peptide 1) + Fragment residue mass 1 (ETD)	1346.74 (peptide)+ 504.16 (Fragment 1)-1(1Hydrogen) =1849.90	1850.90 (1+) 925.95 (2+) 617.63 (3+) 463.48 (4+)
TLSDYNIQKESTLHLVLR (Peptide 2) + Fragment residue mass 2 (ETD)	2130.15 (peptide)+ 132.04 (Fragment 2)-1(1Hydrogen) = 2261.19	2262.19 (1+) 1131.60 (2+) 754.73 (3+) 566.30 (4+)

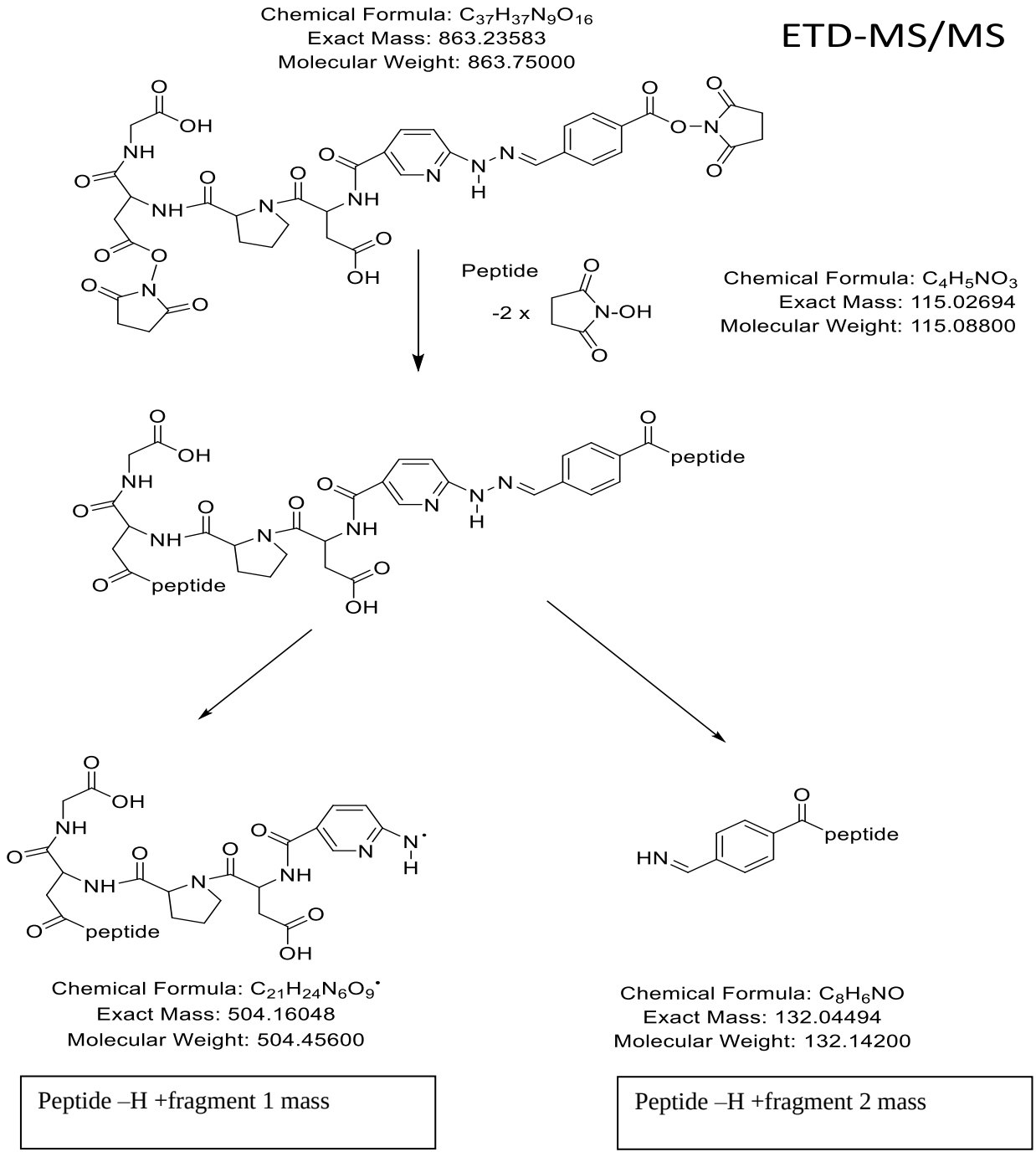
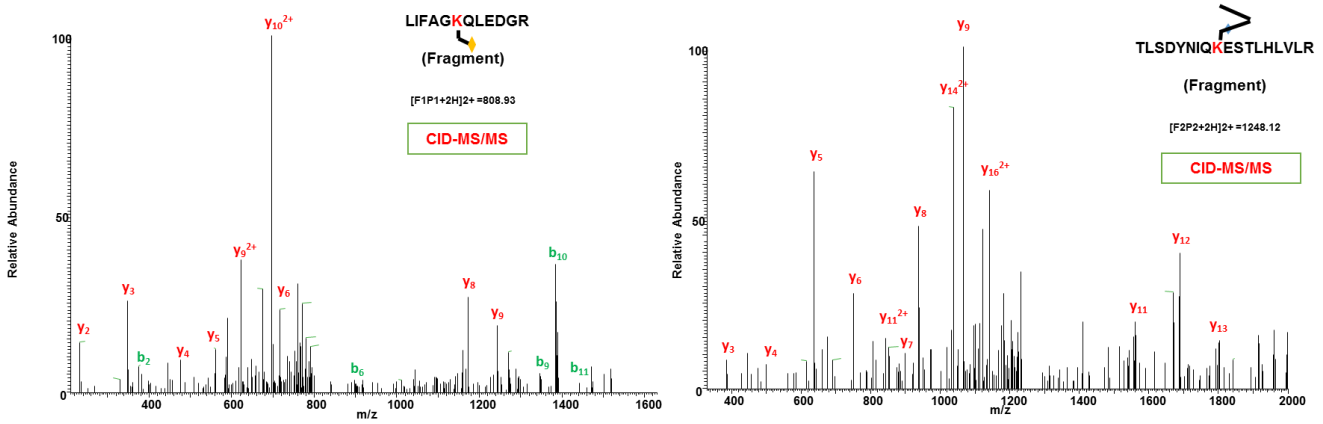


Figure A-S7: Structure of DUCCT with ETD cleaved signature fragments. ETD also generates 2 fragments that are different from CID cleaved fragments.

(A)



(B)

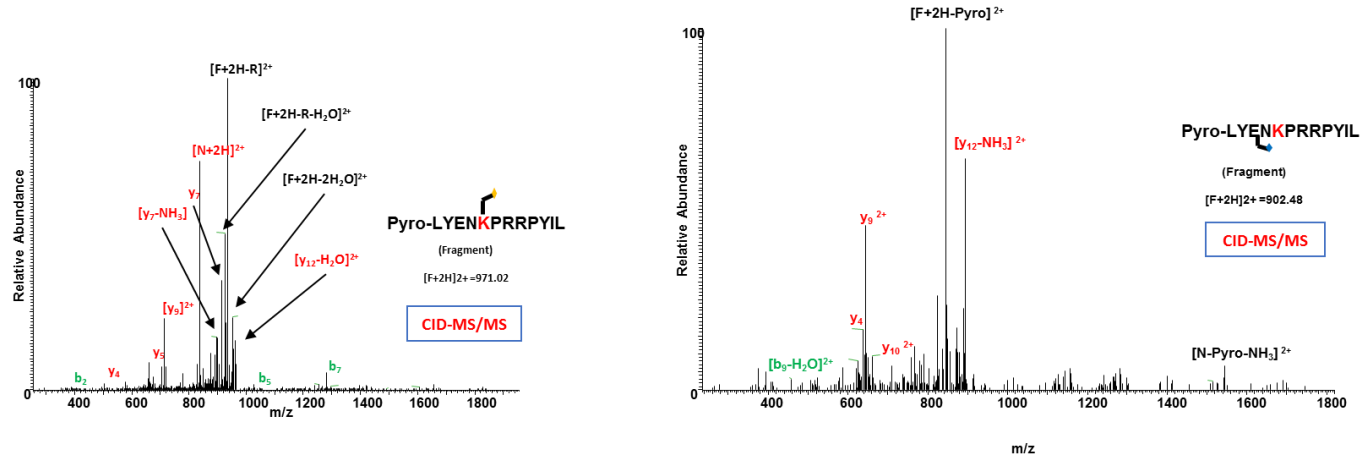
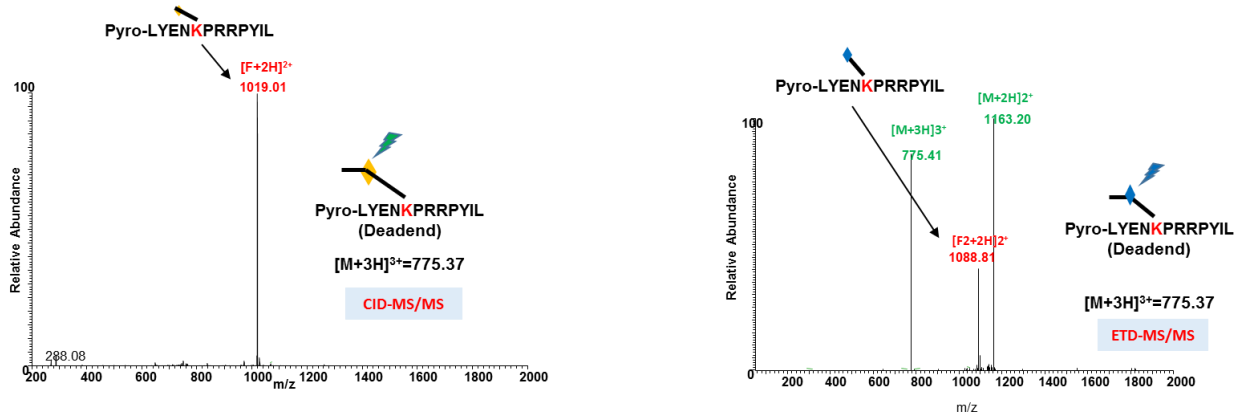


Figure A-S8: Examples of MS/MS/MS spectra of cross-linked fragments derived from A) Ubiquitin and B) Neurotensin. These fragmentation patterns unambiguously confirm the sequence of the peptides.

(A)



(B)

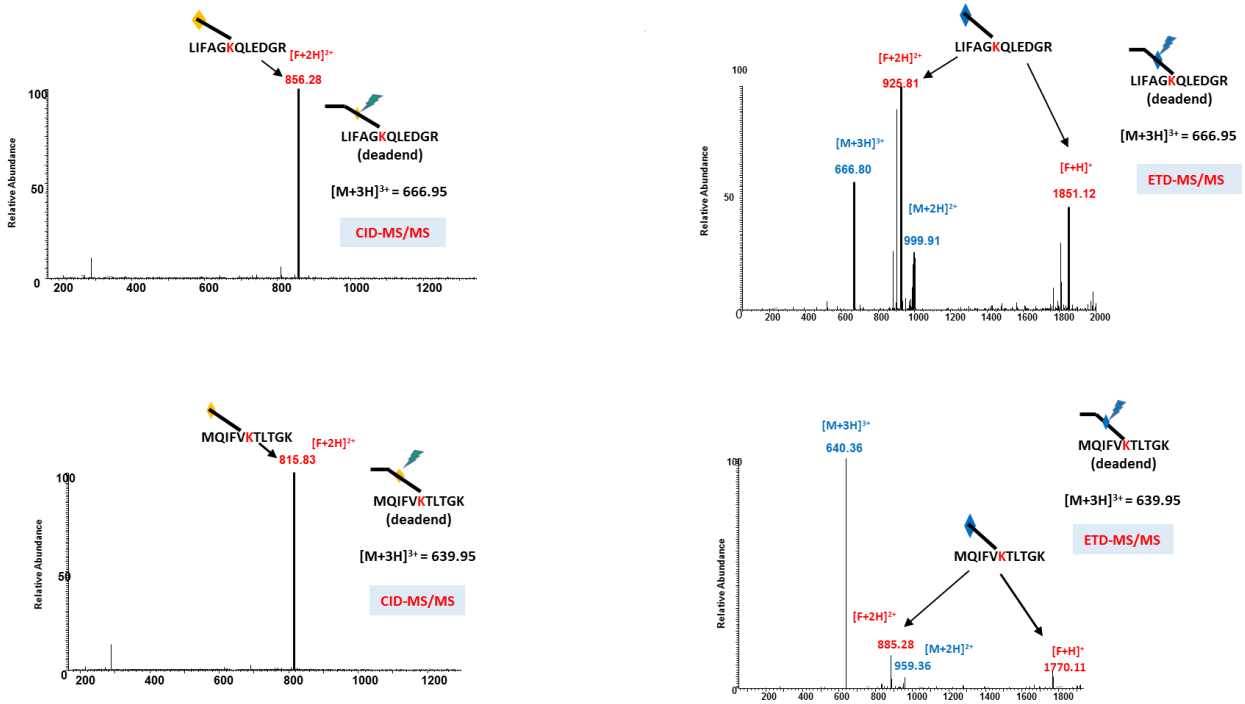


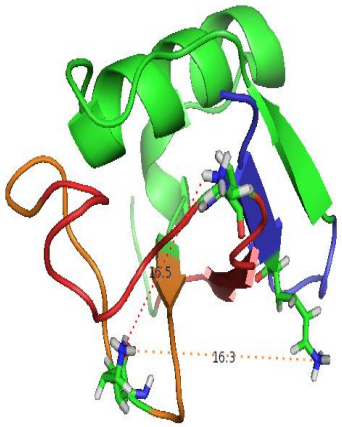
Figure A-S9: CID and ETD mass spectra of dead-end peptides derived from A) Neurotensin and B) Ubiquitin. CID and ETD mass spectra generate different signature ions for the same dead-end peptide.

Table A-S3: CID and ETD together yielded the list of inter cross-linked sites in ubiquitin.

PrecursorCID	PrecursorETD	Sequence	ScanCID	SequenceComingFrom	ScanETD
650.35175	650.26024	MQIFVKTLTGK-------- LIFAGKQLEDGR	scans:	both sequences coming from ubiquitin	scans:
823.19098	823.34947	LIFAGKQLEDGR-------- TLSDYNIQKESTLHLVLR	scans:	both sequences coming from ubiquitin	scans:

- **Ubiquitin sequence**

MQIFVK$_6$TLTGKTITLEVEPSDTIENVKAKI
QDKEGIPPDQQR**LIFAGK$_{48}$QLEDGRTLSD**
YNIQK$_{63}$ESTLHLVLRLRGG



- **Inter-crosslinked product**

  LIFAGKQLEDGR---TLSDYNIQKESTLHLR

  LIFAGKQLEDGR-----MQIFVKTLTGK

- **Dead-end product**

  LIFAGKQLEDGR

  MQIFVKTLTGK

  TLSDYNIQKESTLHLVLR

Figure A-S10: Crystal structure of ubiquitin (pdb-1V80) with the color-coded sites and distances. High confidence identification clearly pointed out the cross-linked lysine residues.

1. **Bio-LPS**
2. **BioLPS-BS3**
3. **BioLPS-DUCCT**

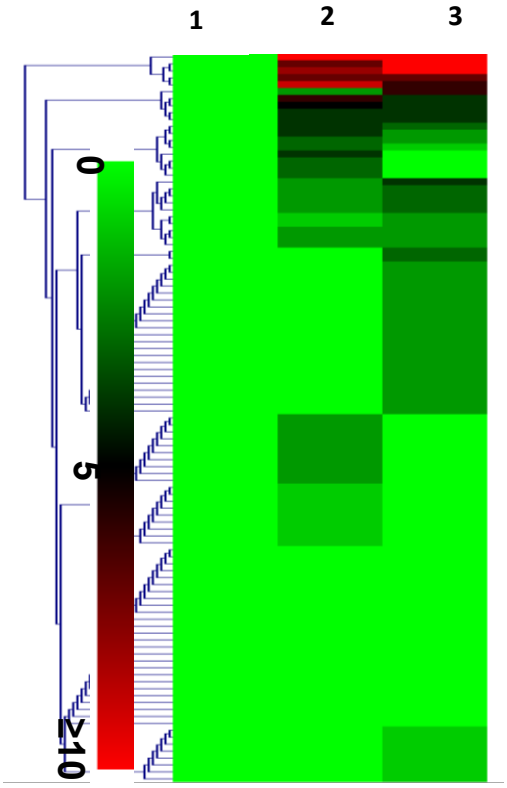


Figure A-S11: Full heat map of proteins which were observed in BioLPS (none), BS3, and DUCCT pulldown studies.

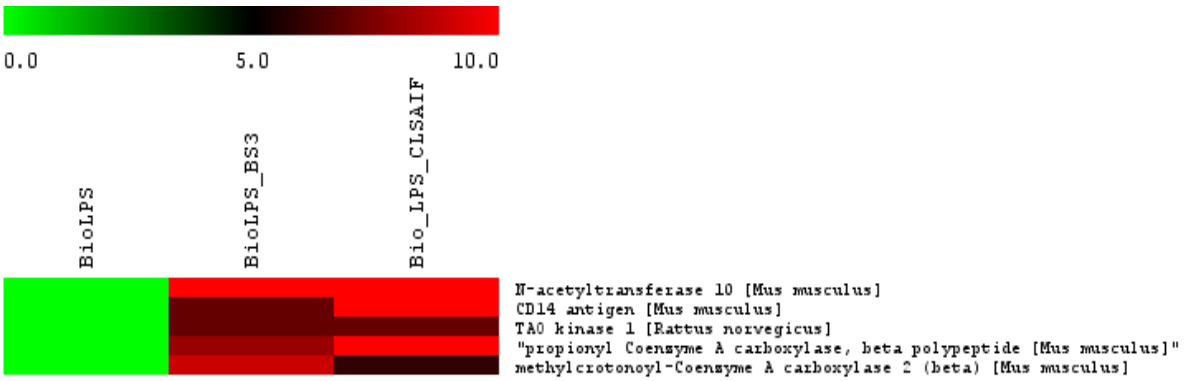


Figure A-S12: Partial heatmap of Figure 6 shown with all the protein identifications.

Appendix B

Supporting Information for Chapter 3

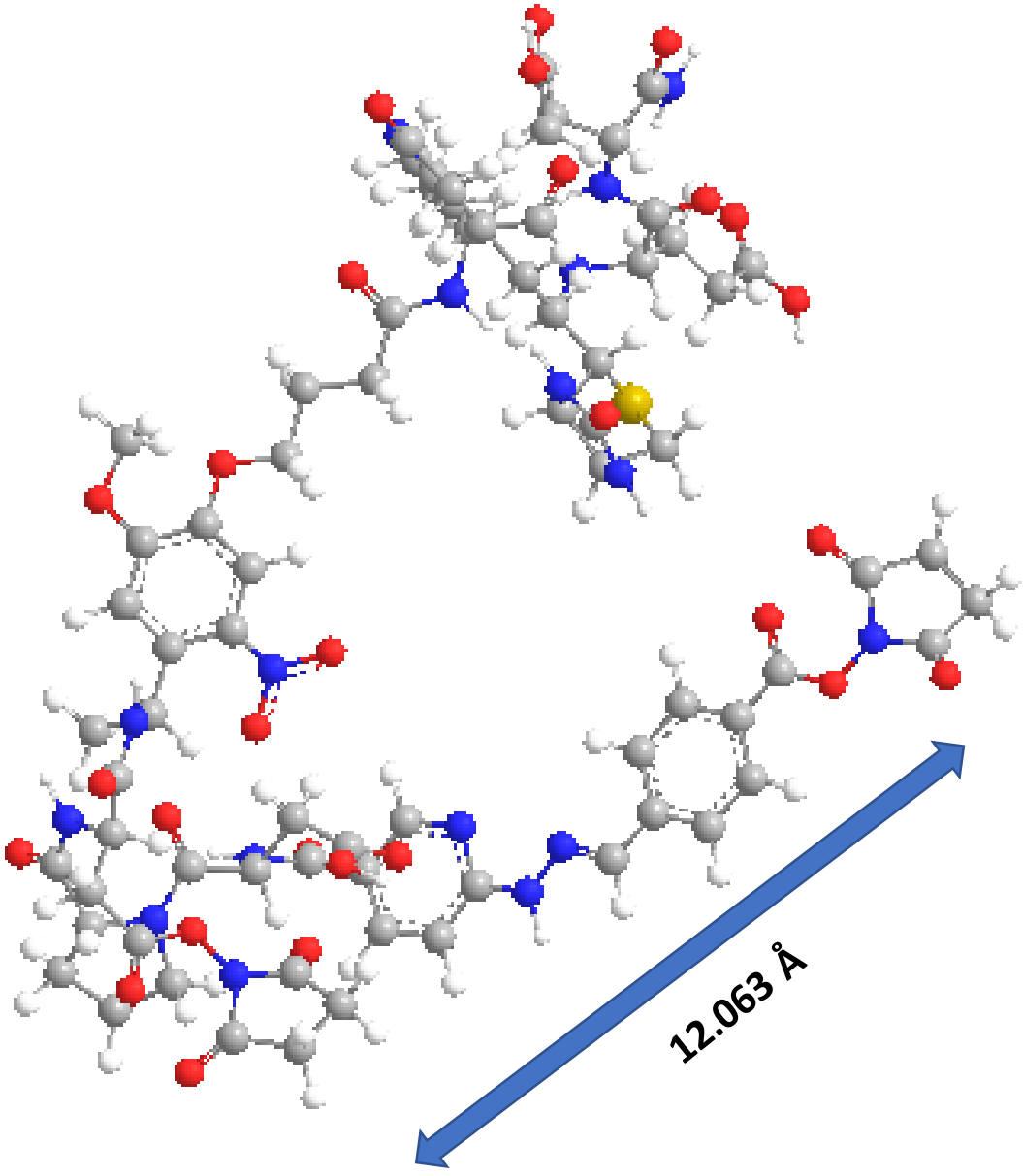


Figure B-S1: 3-D structure of the novel enrichment based dual cleavable cross-linker (PC-DUCCT-Biotin). The length of the reactive spacer arms is calculated 12.063 Å using a chem draw software.

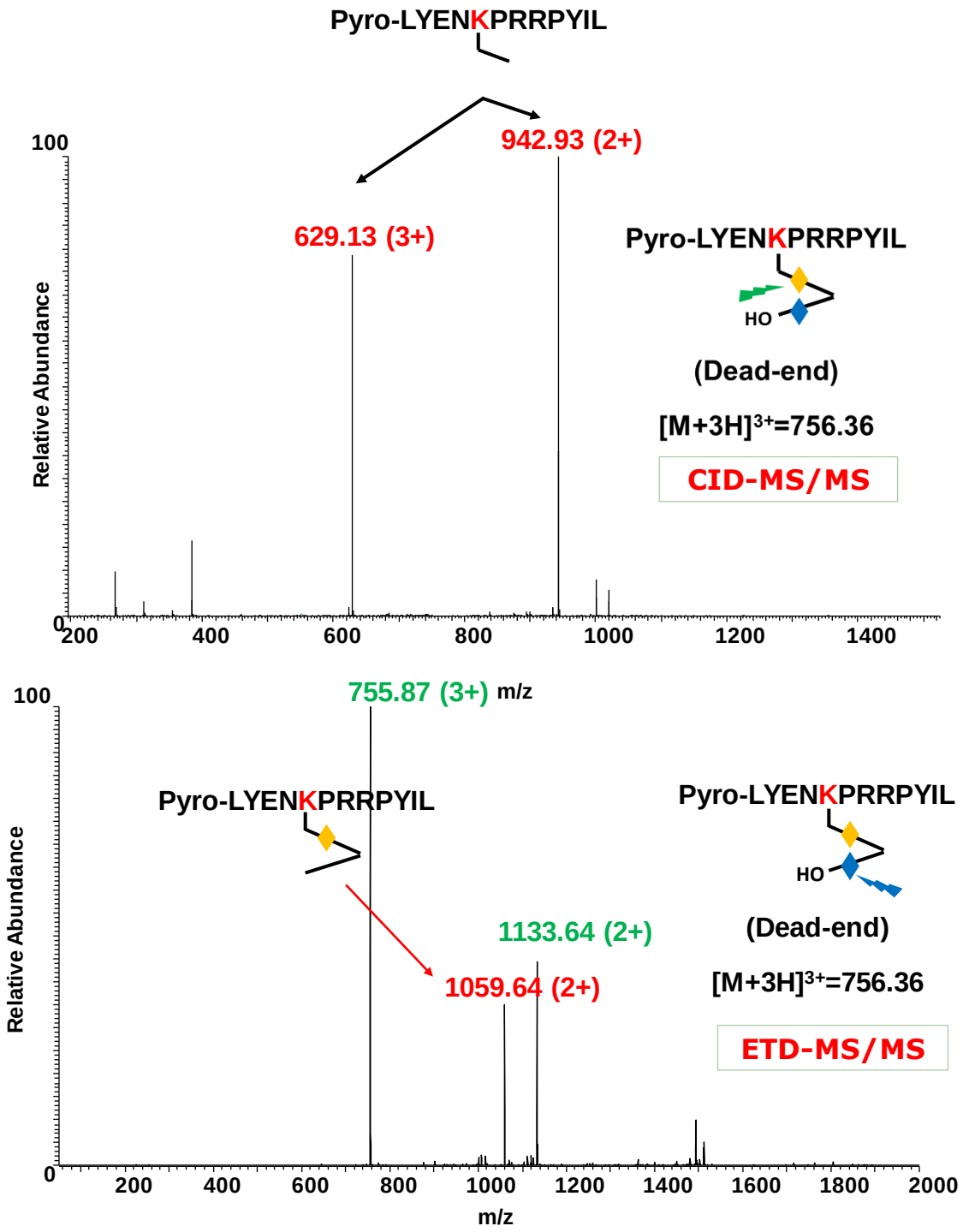


Figure B-S2. CID (top) and ETD (bottom) fragmentation patterns of a dead-end peptide (one-end hydrolyzed) derived from Neurotensin.

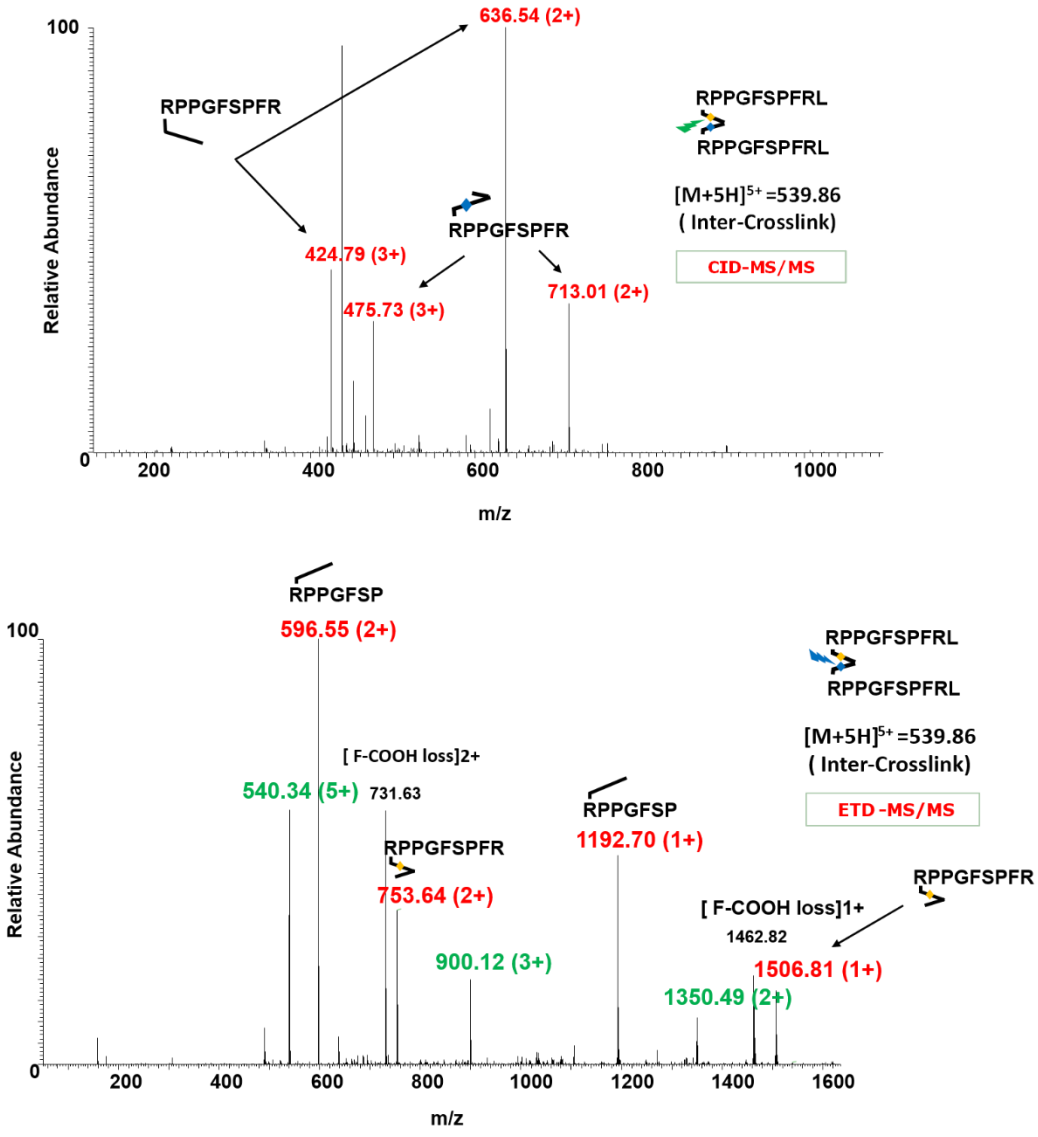


Figure B-S3. CID and ETD mass spectra of an inter cross-linked peptides derived from Bradykinin. CID and ETD mass spectra generate different signature ions for the same cross-linked peptide.

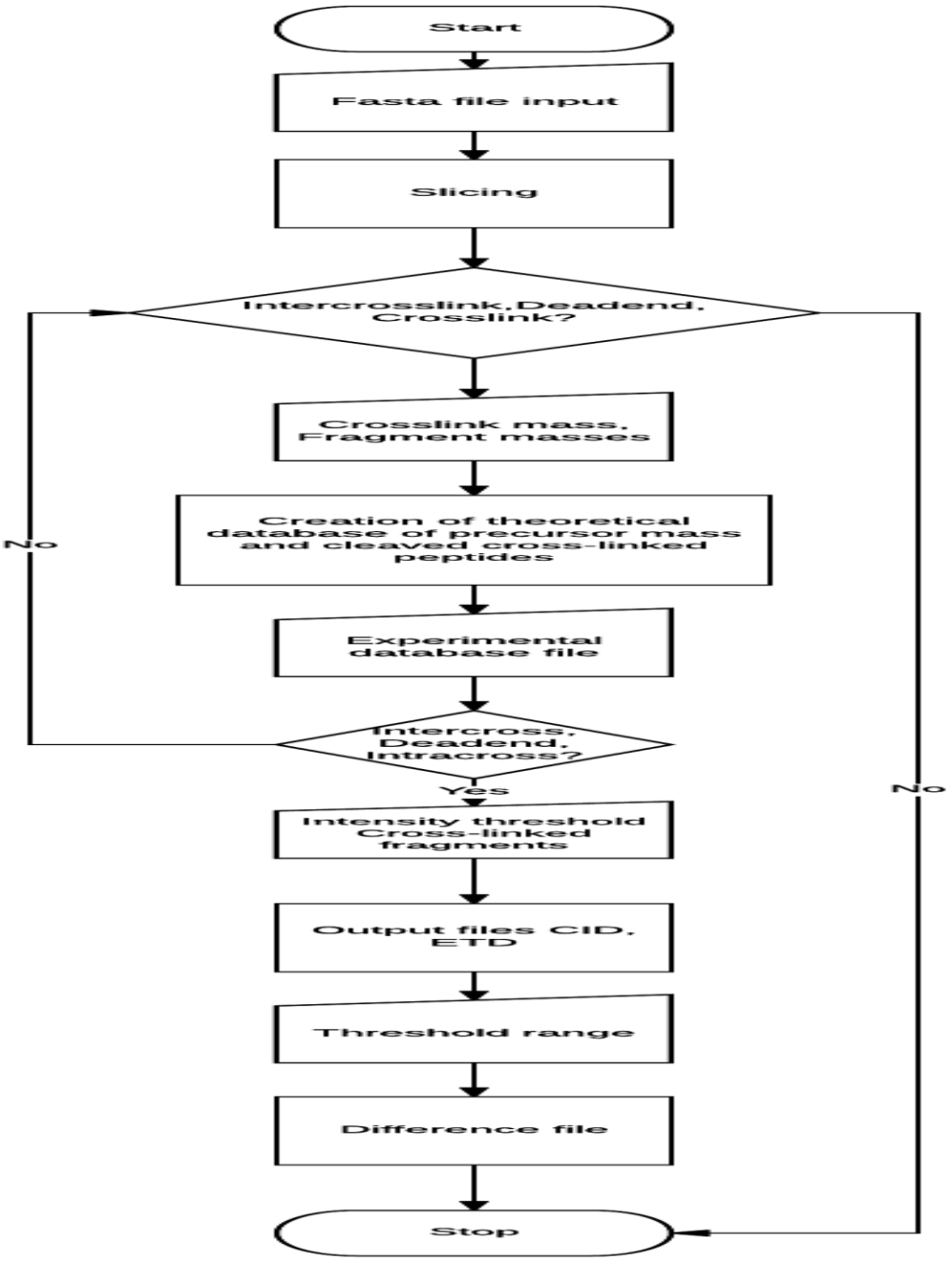


Figure B-S4. Workflow for the identification of inter-peptide cross-links using our Cleave-XL software.

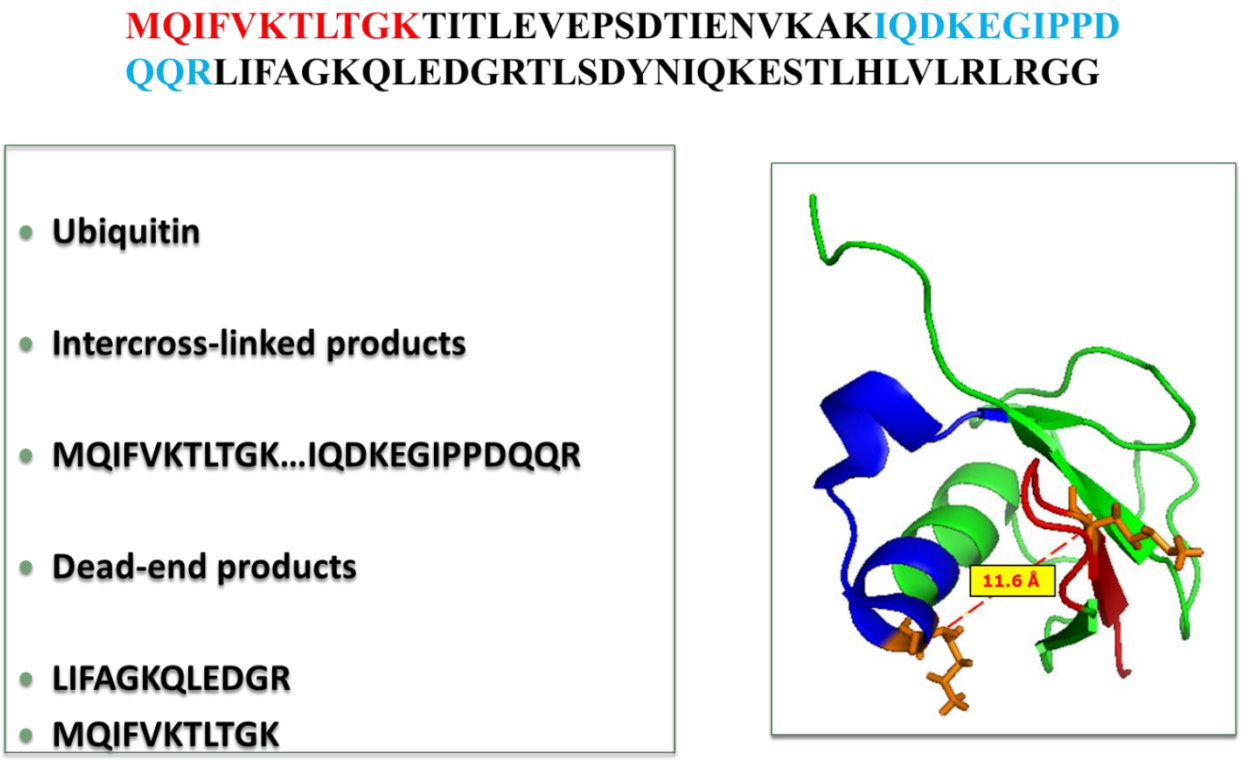


Figure B-S5. X-ray crystallographic structure of Ubiquitin (pdb-1V80) with the color-coded cross-linking sites and distances.

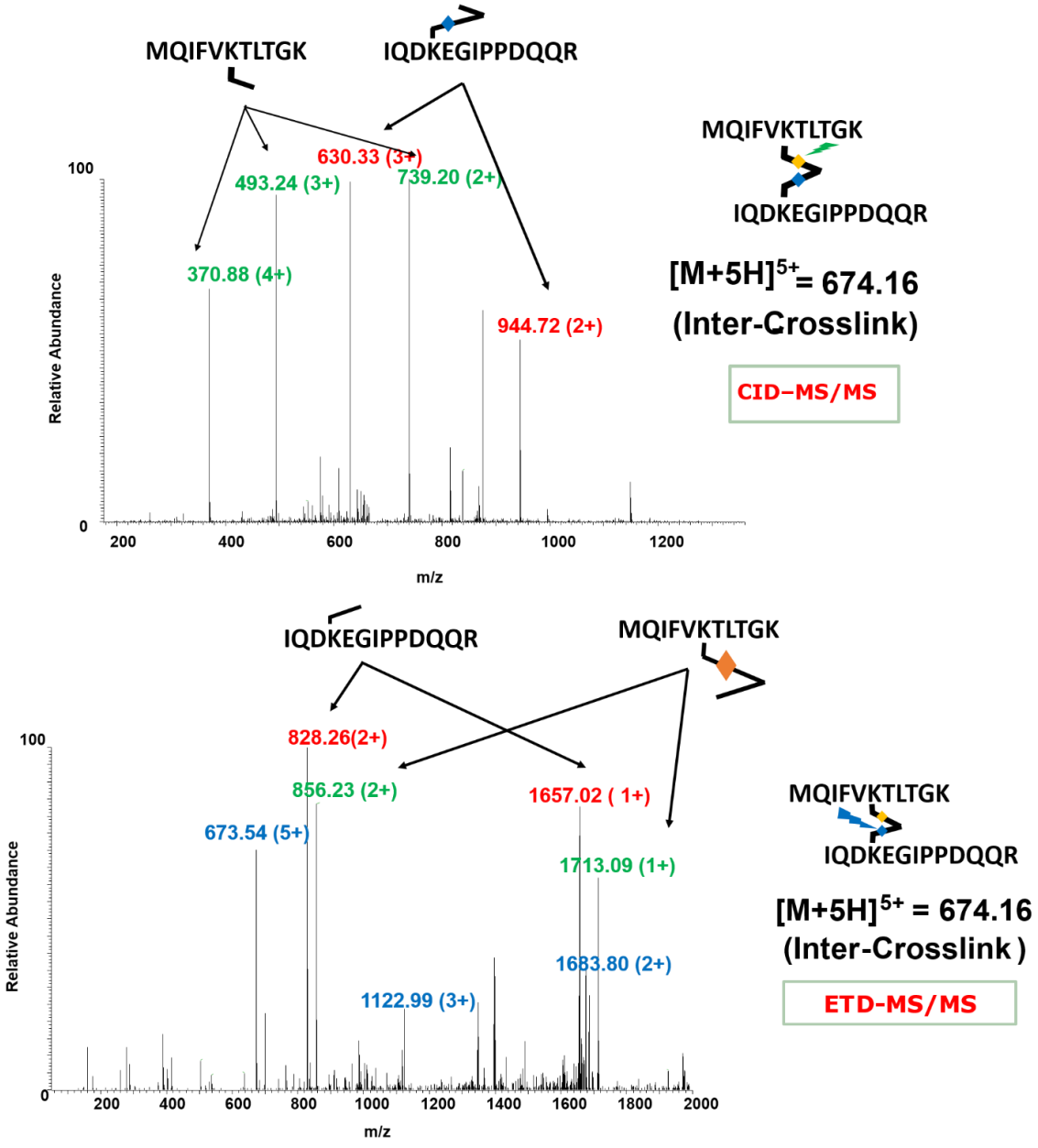


Figure B-S6. Example of CID and ETD mass spectra of an inter cross-linked peptide, derived from ubiquitin. Lysine 6 (amino acids 1–11); of ubiquitin was found to be cross-linked with lysine 33 (amino acids 30–42).

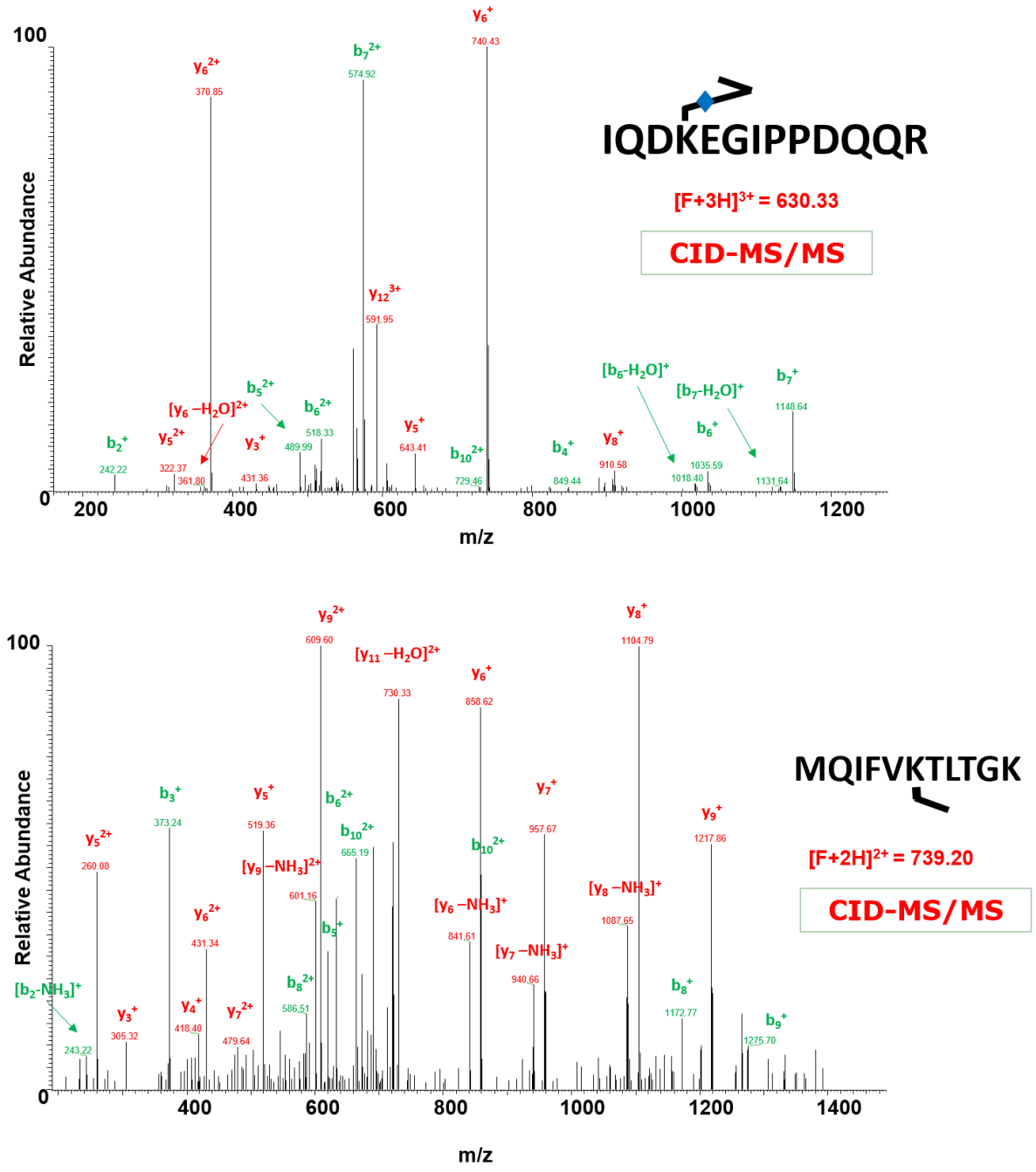


Figure B-S7. Examples of MS/MS of cross-linked fragments derived from Ubiquitin by collision induced dissociation (CID). MS/MS confirmed the modified peptide sequences.

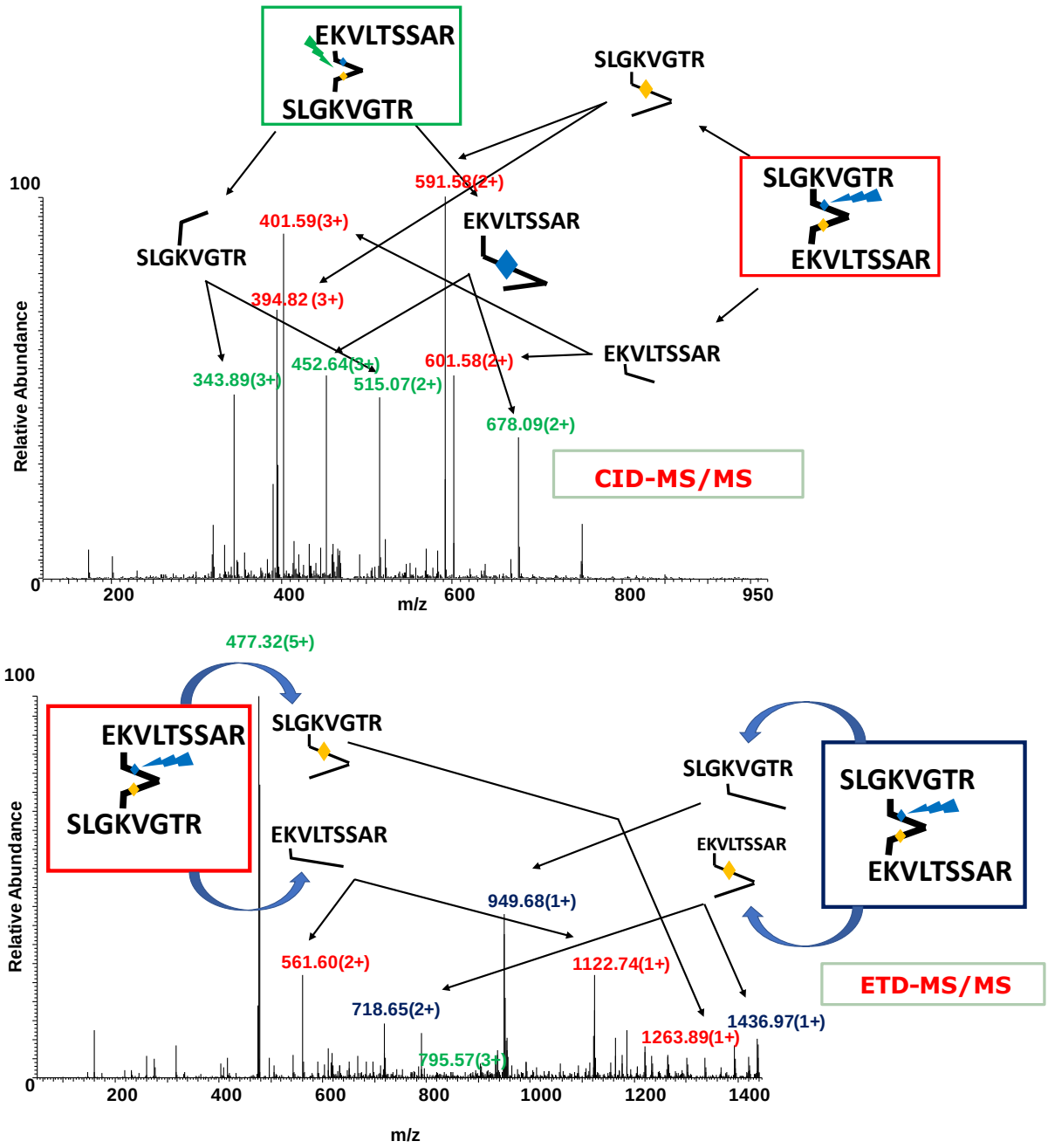


Figure B-S8. Inter-crosslinked peptide at m/z 477.25 (5+) **(SLGKVGTR…. EKVLTSSAR)** derived from BSA after cross-linking with PC-DUCCT-Biotin cross-linker.

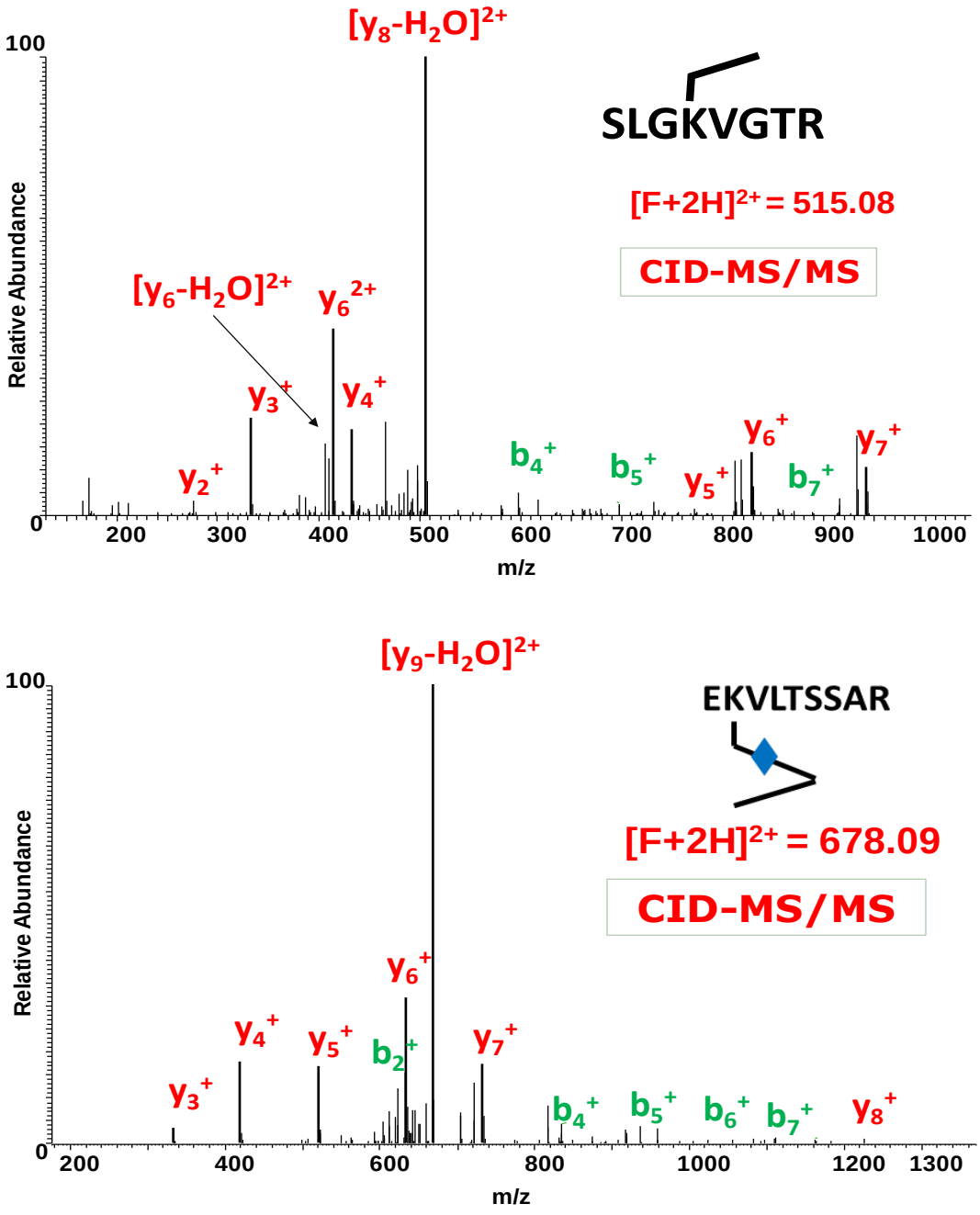


Figure B-S9. Examples of MS/MS of cross-linked fragments derived from BSA by collision induced dissociation (CID). MS/MS of the modified-peptide sequences provided the additional confidence of identification.

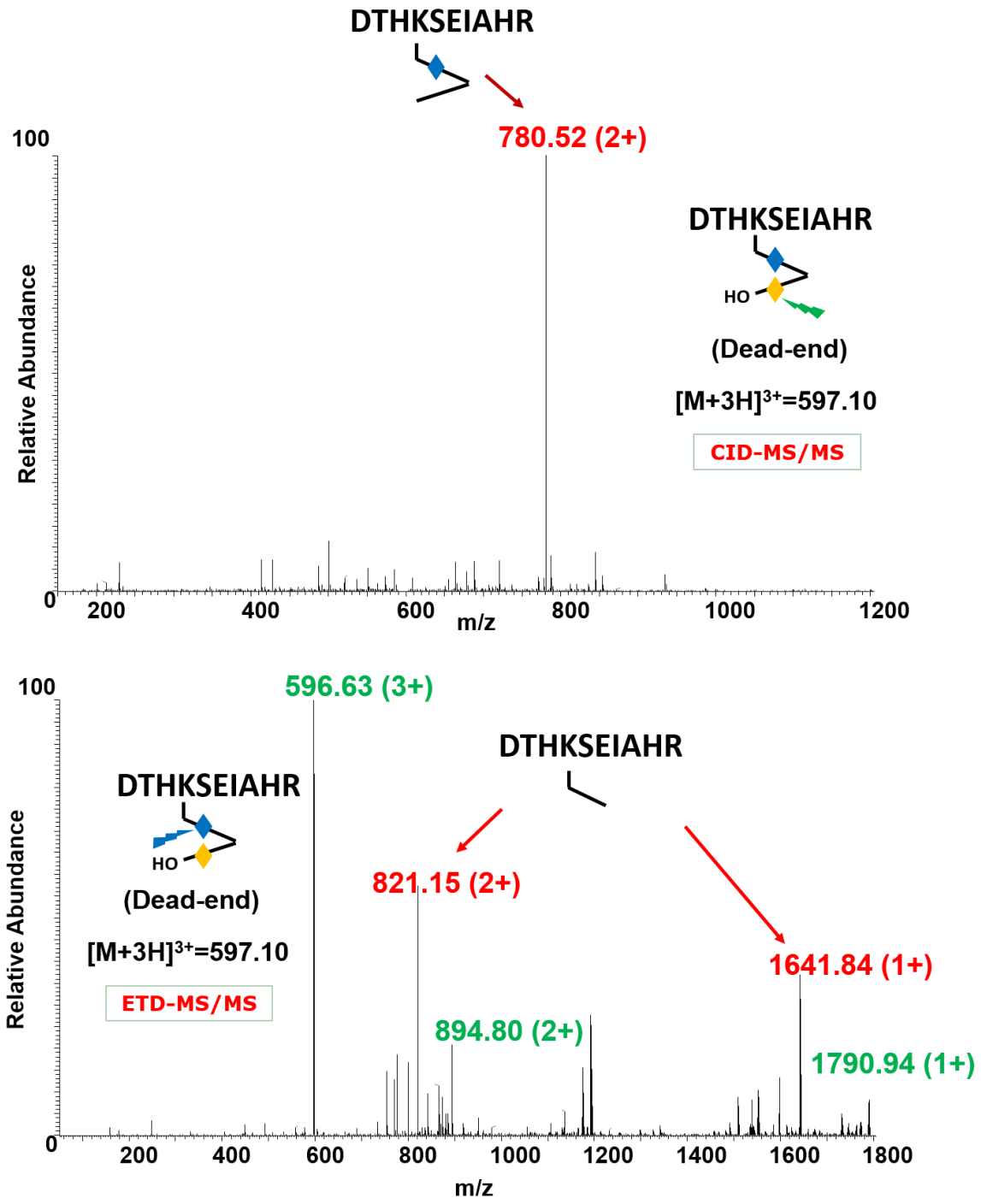


Figure B-S10. Examples of signature fragmentation patterns (CID-MS/MS and ETD-MS/MS) of dead-end peptide derived from BSA.

Figure B-S11. X-ray crystallographic structure of BSA (pdb- 4F5S) with the distance constrains between interacting residues.

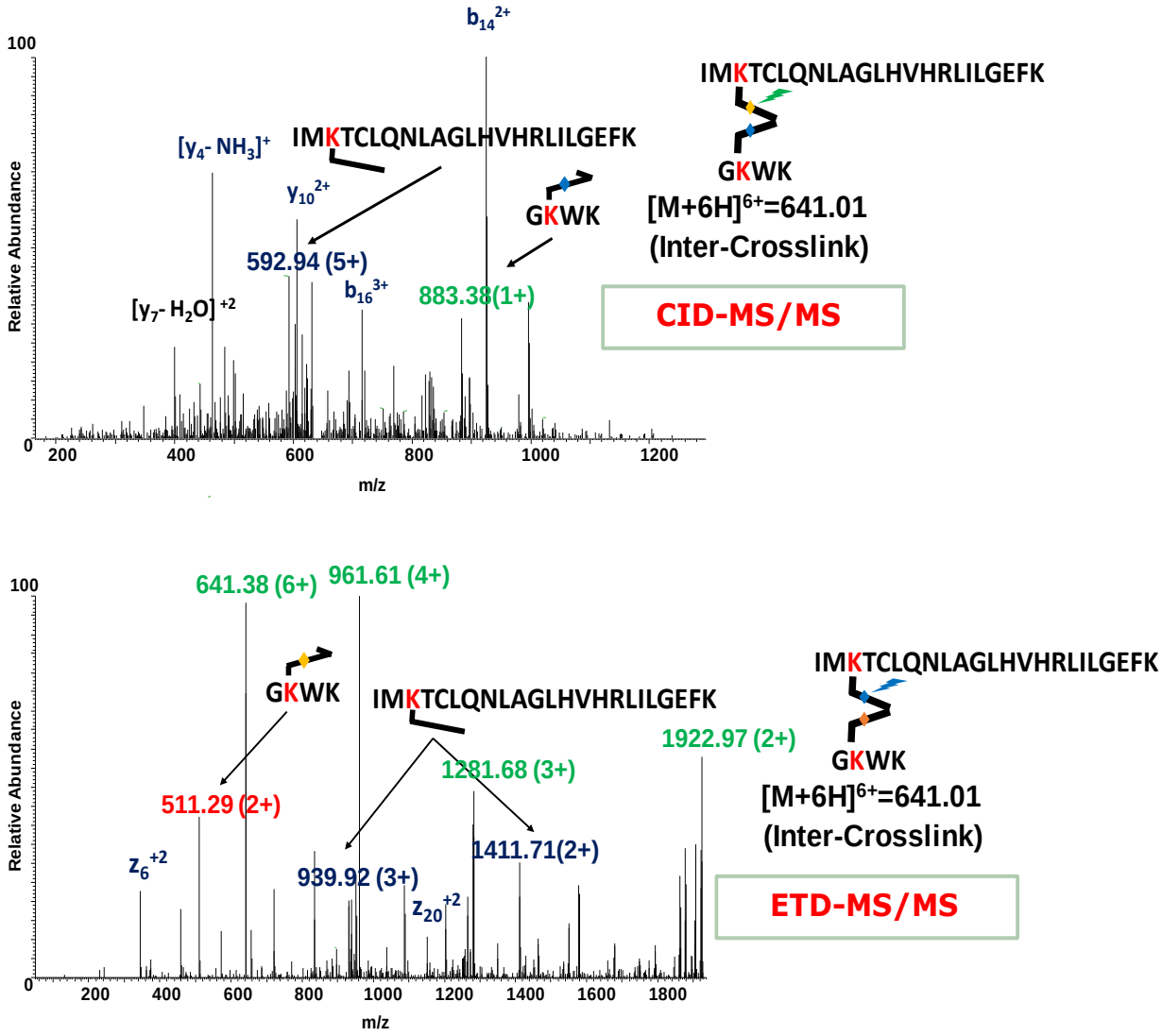


Figure B-S12: Example of CID and ETD mass spectra of an inter cross-linked peptide, derived from TLR4 and LBP (TLR4 Complex).

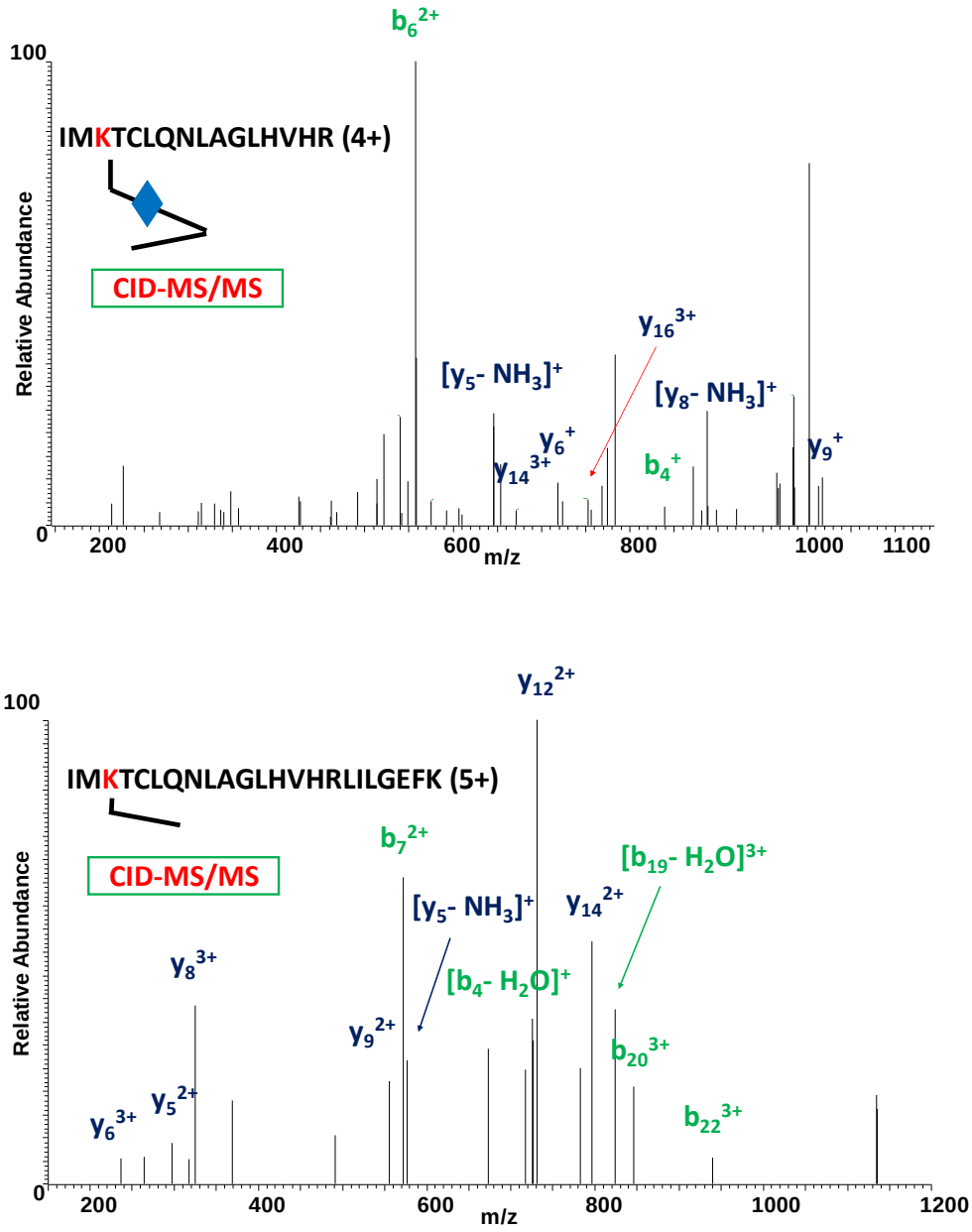


Figure B-S13: Examples of MS/MS of cross-linked fragments derived from TLR4-Complex by collision induced dissociation (CID).

Table B-S1: List of calculated masses and m/z values used to identify cross-linked reaction products derived from neurotensin.

	Exact Mass	**m/z**
Cross-linker (PC-DUCCT-Biotin)	1697.5940	1698.6013 (1+)
CID Fragments	1218.4863 479.1077	1219.4936(1+) 480.1150 (1+)
PC-DUCCT (after UV treatment)	805.2303	806.2376 (1+)
NHS (reactive group)	114.0191	
Neurotensin + PC-DUCCT (Inter cross-link)	(1671.9096×2) (2peptides) +805.2303 (Cross-linker) -(114.0191×2) (2 leaving group)-2.0156 (2 Hydrogen) = 3918.9957	980.7562 (4+) 784.8064 (5+)
Neurotensin + PC-DUCCT (Dead-end)	1671.9096 (peptide)+ 805.2303 (Cross-linker) -(114.0191×2) (2 leaving group) +17.0027 (-OH) -1.0078(hydrogen) = 2265.0966	2266.1039 (1+) 1133.5556 (2+) 756.0395 (3+)
Neurotensin + Fragment residue mass 1 (CID)	1671.9096 (Peptide)+ 212.1035 (Fragment 1)-1.0078 (1 Hydrogen) = 1883.0053	1884.0126 (1+) 942.5099 (2+) 628.6757 (3+)
Neurotensin + Fragment residue mass 2 (CID)	1671.9096 (Peptide)+ 365.0884 (Fragment 1)-1.0078 (1 Hydrogen) = 2035.9902	2036.9975 (1+) 1019.0024 (2+) 679.6707 (3+)
Neurotensin + Fragment residue mass 1 (ETD)	1671.9096 (Peptide)+ 132.0450 (Fragment 1)-1.0078 (1 Hydrogen) =1802.9468	1803.9541 (1+) 902.4807 (2+) 601.9849 (3+)
Neurotensin + Fragment residue mass 2 (ETD)	1671.9096 (Peptide)+ 446.1550 (Fragment 1)-1.0078 (1 Hydrogen) =2117.0568	2118.0641 (1+) 1059.5357 (2+) 706.6929 (3+)

Table B-S2: List of calculated masses and m/z values used to identify cross-linked reaction products derived from Bradykinin.

	Exact Mass	m/z
Bradykinin + PC-DUCCT (Inter-crosslink)	(1059.5614×2) (2peptides) +805.2303 (Cross-linker) - (114.0191×2) (2 leaving group)- 2.0156 (2 Hydrogen) = 2694.3131	899.1117 (3+) 674.5456 (4+) 539.8699 (5+)
Bradykinin + PC-DUCCT (Dead-end)	1059.5614 (peptide)+ 805.2303 (Cross-linker) -(114.02×2) (2 leaving group) + 17 (-OH) -1.0078 (Hydrogen) = 1652.7517	1653.7590 (1+) 827.3832 (2+) 551.9245 (3+)
Bradykinin + Fragment residue mass 1 (CID)	1059.5614+ 212.1035-1.0078 = 1270.6571	1272.6644 (1+) 636.3359 (2+) 424.5597 (3+)
Bradykinin + Fragment residue mass 2 (CID)	1059.5614+365.0884-1.0078= 1423.6420	1424.6493 (1+) 712.8283 (2+) 475.5546 (3+)
Bradykinin + Fragment residue mass 1 (ETD)	1059.5614+132.0450-1.0078=1190.5986	1191.6059 (1+) 596.3066 (2+) 397.8735 (3+)
Bradykinin + Fragment residue mass 2 (ETD)	1059.5614+446.1550-1.0078=1504.7086	1505.7159 (1+) 753.3616 (2+) 502.5768 (3+)

Table B-S3: List of calculated masses and m/z values used to identify inter cross-linked peptides derived from BSA.

	Exact Mass	m/z	m/z
SLGKVGTR....EKVLTSSAR + PC-DUCCT (Inter-crosslink)	(**816.4817+989.5505**) (2 peptides) +805.2303 (Cross-linker) - (114.0191×2) (2 leaving group)- 2.0156 (2 Hydrogen) = 2381.2087	2382.2160 (1+) 1191.61165 (2+) 794.7435 (3+) 596.3095 (4+) 477.2490 (5+)	
SLGKVGTR + Fragment residue mass 1 (CID)	**816.4817**+ 212.1035-1.0078 = 1027.5774 (small fragment)	1028.5847 (1+) 514.7960 (2+) 343.5331 (3+)	1181.5698 (1+) 591.2885 (2+) 394.5281(3+)
	816.4817+ 365.0884-1.0078 = 1180.5625 (large fragment)		
EKVLTSSAR + Fragment residue mass (CID)	**989.5505**+365.0884-1.0078= 1353.3832 (large fragment)	1354.3905 (1+) 677.6989 (2+) 452.1350 (3+)	1201.6535 (1+) 601.3304 (2+) 401.2227 (3+)
	989.5505+212.1035-1.0078= 1200.6462 (small fragment)		
SLGKVGTR + Fragment residue mass 1 (ETD)	**816.4810**+446.1550-1.0078= 1261.6282 (large fragment)	1262.6355 (1+) 631.8214 (2+) 421.5500 (3+)	948.5255 (1+) 474.7664 (2+) 316.8467 (3+)
	816.4810+132.0450-1.0078= 947.5182 (small fragment)		
EKVLTSSAR + Fragment residue mass 2 (ETD)	**989.5505**+ 132.0450-1.0078=1120.5877 (small fragment)	1121.5950 (1+) 561.30115 (2+) 374.5365 (3+)	1435.7050 (1+) 718.35615 (2+) 479.2398 (3+)
	989.5505+ 446.1550-1.0078=1434.6977 (large fragment)		

Table B-S4. List of inter cross-linked peptides identified from BSA by Cleave-XL.

List of Products	Distance (Å)	List of Products	Distance (Å)
•DTHKSEIAHR------KFWGK	**32.7**	•EKVLTSSARQR..... KFWGK	**22.6**
•KVPQVSTPTLVEVSR-------EKVLTSSAR	**29.9**	•EKVLTSSAR-------DDPHACYSTVFDKLK	**36.8**
•DTHKSEIAHR------LKECCDK	**36.8**	•EKVLTSSARQR..... PLLEKSHCIAEVEK	**17.1**
•DDSPDLPKLK------SLGKVGTR	**19.9**	•EKVLTSSAR-------SLGKVGTR	**13.7**
•AEFVEVTKLVTDLTK-------- NYQEAKDAFLGSFLYEYSR	**17.6**	•SLGKVGTR-------FKDLGEEHFK	**39.8**
•DDSPDLPKLK------TPVSEKVTKCCTESLVNR	**31.8**	•KQTALVELLK------LRCASIQKFGER	**34.8**
•DDSPDLPKLK------KQTALVELLK	**18.3**	•CCAADDKEACFAVEGPK-------QTALVELLKHKPK	**28.6**
•KVPQVSTPTLVEVSR-------TPVSEKVTK	**8.9**	•CASIQKFGER-------PCFSALTPDETYVPKAFDEK	**32.4**
•KVPQVSTPTLVEVSR-------KQTALVELLK	**25.9**	•CCTKPESER-------ADEKKFWGK	**40.5**

Table B-S5. List of inter cross-linked peptides identified from TLR4 complex by Cleave-XL after comparing CID and ETD output file.

Precursor CID	Precursor ETD	Sequence	Sequence Coming From
522.86295	523.09575	DFSGDFKIKAVGR--------SFNPLKILK	first sequence coming from: LBP, second sequence coming from: TLR4
598.53504	598.6845	GSISFKK--------GSISFKKVALPSLSYLDLSR	both sequences coming from TLR4
614.78621	614.93567	IQEGFHKSR--------HFKWQSLSIIRCQLK	both sequences coming from TLR4
622.6201	622.68623	IMKTCLQNLAGLHVHR--------FLQWVKEQK	both sequences coming from TLR4
641.45479	641.35423	IMKTCLQNLAGLHVHRLILGEFK--------GKWK	first sequence coming from: TLR4, second sequence coming from: LBP
647.78857	647.77136	ITDKGLAYAAK--------GLAYAAKEGLVALQRELYK	both sequences coming from LBP
674.20714	674.3566	LILGEFKDER--------GSISFKKVALPSLSYLDLSR	both sequences coming from TLR4
674.87385	674.77329	LILGEFKDER--------VDTEADLGQFTDIIKSLSLK	first sequence coming from: TLR4, second sequence coming from: CD14
678.29076	678.35689	RKEVLCHGHDDDYSFCR--------IQEGFHKSRK	both sequences coming from MD2
678.3741	678.44022	QFPTLDLPFLKSLTLTMNK--------LILGEFKDER	both sequences coming from TLR4
699.45895	699.27505	FHNQIESKLQKVLENK--------GCKKYSR	first sequence coming from: LBP, second sequence coming from: TLR4
703.87593	703.94205	YLEDVPKHFK--------SFLKLHGSFDLDVK	first sequence coming from: TLR4, second sequence coming from: LBP
741.46196	741.61142	GLAYAAKEGLVALQRELYK--------KDFLYLGANVQYMR	both sequences coming from LBP
753.04612	753.27893	FHNQIESKLQK--------PFTPQIYKKYPDMK	first sequence coming from: LBP, second sequence coming from: TLR4
761.21338	760.94614	LSKVPDDIPSSTK--------SLEYLLKR	first sequence coming from: TLR4, second sequence coming from: CD14
782.4649	782.36435	SFNPLKILK--------GLISALCPLKFPTLQVLALR	first sequence coming from: TLR4, second sequence coming from: CD14
849.13635	848.86912	KLNVAHNFIHSCK--------KFHCLANVSAMSLAGVSIK	both sequences coming from TLR4
854.05337	853.95281	GILQHFPKSLAFFNLTNNSVACICEHQK--------FLQWVKEQK	both sequences coming from TLR4
862.3873	862.45342	LILGEFKDER-------- SFLKLHGSFDLDVKGVTISVDLLLGMDPSGR	first sequence coming from: TLR4, second sequence coming from: LBP
868.55441	868.45385	SFLKLHGSFDLDVK--------GLAYAAKEGLVALQRELYK	both sequences coming from LBP
870.47121	870.37066	SFLKLHGSFDLDVK--------QVLDVMFKGEIFNRNHR	both sequences coming from LBP
901.47344	901.70624	GKWKVR--------VDTEADLGQFTDIIKSLSLK	first sequence coming from: LBP, second sequence coming from: CD14
901.55678	901.70624	GKWKVR--------RKEVLCHGHDDDYSFCR	first sequence coming from: LBP, second sequence coming from: MD2
909.64069	909.4568	FKDNTLSNVFANTTNLTFLDLSK-------- YLYFNLFISVNSIELPKR	first sequence coming from: TLR4, second sequence coming from: MD2
998.56374	998.62986	NELVKNLEEGVPR-------- GESIYDAFVIYSSQNEDWVRNELVKNLEEGVPR	both sequences coming from TLR4

Table B-S6: List of calculated masses and *m/z* value used to assign cross-linked products (TLR4 Dimer complex).

	Exact Mass	m/z
IMKTCLQNLAGLHVHR **…. FLQWVKEQK** DUCCT (Inter-crosslink)	(1890.0080+ 1205.6677) (2 peptides) +863.2358(Cross-linker) -(114.0191×2) (2 leaving gr.)-2.0156 (2 Hydrogen) = 3728.8577	3729.8650 (1+) 1865.4362 (2+) 1243.9599 (3+) 933.2217 (4+) 746.7788 (5+) 622.4836 (6+)
IMKTCLQNLAGLHVHR (Peptide 1) + Fragment residue mass 1 (CID)	1890.0080 (peptide)+ 365.0884(Fragment1)-1.0078(1Hydrogen) = 2254.0886	2255.0959 (1+) 1128.0516 (2+) 752.3702 (3+) 564.5295 (4+)
FLQWVKEQK (Peptide 2) + Fragment residue mass 2 (CID)	1205.6677 (peptide)+ 270.1090 (Fragment 2)-1.0078(1 Hydrogen) =1474.7689	1475.7762 (1+) 738.3918 (2+) 492.5969 (3+)
IMKTCLQNLAGLHVHR (Peptide 1) + Fragment residue mass 1 (ETD)	1890.0080 (peptide)+ 132.0450 (Fragment 1)-1.0078(1 Hydrogen) = 2021.0452	2022.0525 (1+) 1011.5299 (2+) 674.6890 (3+)
FLQWVKEQK (Peptide 2) + Fragment residue mass 2 (ETD)	1205.6677 (peptide)+ 504.1605 (Fragment 2)-1.0078(1Hydrogen) = 1708.8204	1709.8277 (1+) 855.4175 (2+) 570.6141 (3+)

Appendix C

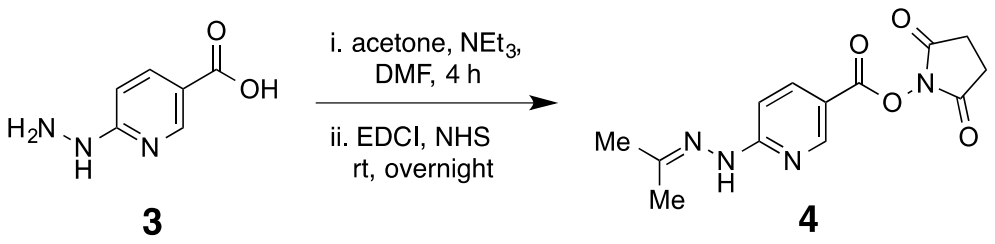
Supporting Information for Chapter 4

Detail Synthetic Scheme of HI-ETD-XL:

Step 1: Synthesis of 2,5-dioxopyrrolidin-1-yl 4-formylbenzoate1

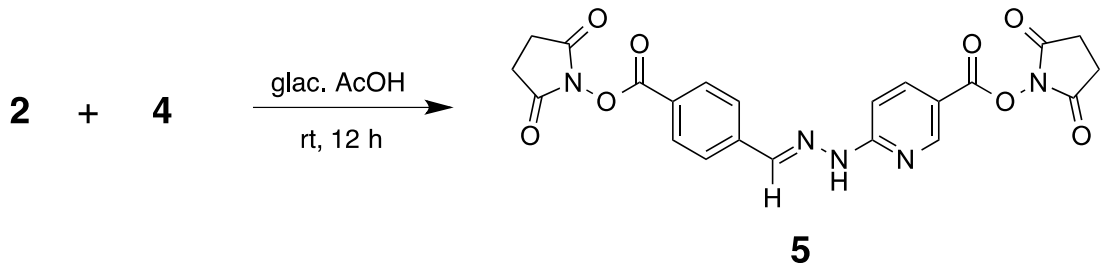
To a 50 mL roundbottom flask was added 4-formylbenzoic acid (1) (450 mg, 3.0 mmol) and 15 mL of DMF. 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (512 mg, 3.3 mmol, 1.1 equiv.) and N-hydroxysuccinimide (380 mg, 3.3 mmol, 1.1 equiv.) were added in one portion and the reaction stirred under argon atmosphere overnight. Chloroform (30 mL) was added and poured into 100 mL of 10% aqueous CuSO4. The layers were separated, and the aqueous layer separated 2x30 mL of chloroform. The combined organic layers were washed with 100 mL of 10% aqueous CuSO4 and 50 mL of brine before drying with MgSO4. The solvent was evaporated by rotary evaporation before purifying via silica gel flash chromatography (1:1 ethyl acetate/hexanes) to give the product with ~1 mL of DMF. Diethyl ether (10 mL) was added and the resulting solid filtered to yield 403 mg of 2,5-dioxopyrrolidin-1-yl 4-formybenzoate (2) (54%) as a white crystalline solid.

Step 2: Synthesis of 2,5-dioxopyrrolidin-1-yl 6-(2-(propan-2-ylidene)hydrazinyl)nicotinate1



To a 50 mL roundbottom flask was added 6-hydrazinicotinic acid (3) (542 mg, 3.54 mmol), triethylamine (1.3 mL, 9.3 mmol, 2.6 equiv.), acetone (0.50 mL, 6.8 mmol, 1.9 equiv.) and DMF (10 mL). After 4 h of stirring at room temperature, N-hydroxysuccinimide (449 mg, 3.9 mmol, 1.1 equiv.) and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (605 mg, 3.9 mmol, 1.1 equiv.) were added in one portion and the reaction was stirred under an argon atmosphere overnight. Chloroform (30 mL) was added and the solution poured into 100 mL of water. The layers were separated, and the aqueous layer was extracted 2x30 mL of chloroform. Combined organic layers were washed with 2x100 mL of water and dried (MgSO4), then the solvent was removed. The crude product was purified by column chromatography (5% MeOH in DCM) to give the pure product along with residual DMF from the reaction. Addition of diethyl ether (10 mL) and filtration yielded 148 mg of 2,5-dioxopyrrolidin-1-yl 6-(2-(propan-2-ylidene)hydrazinyl)nicotinate (4) as a white powder (14%).

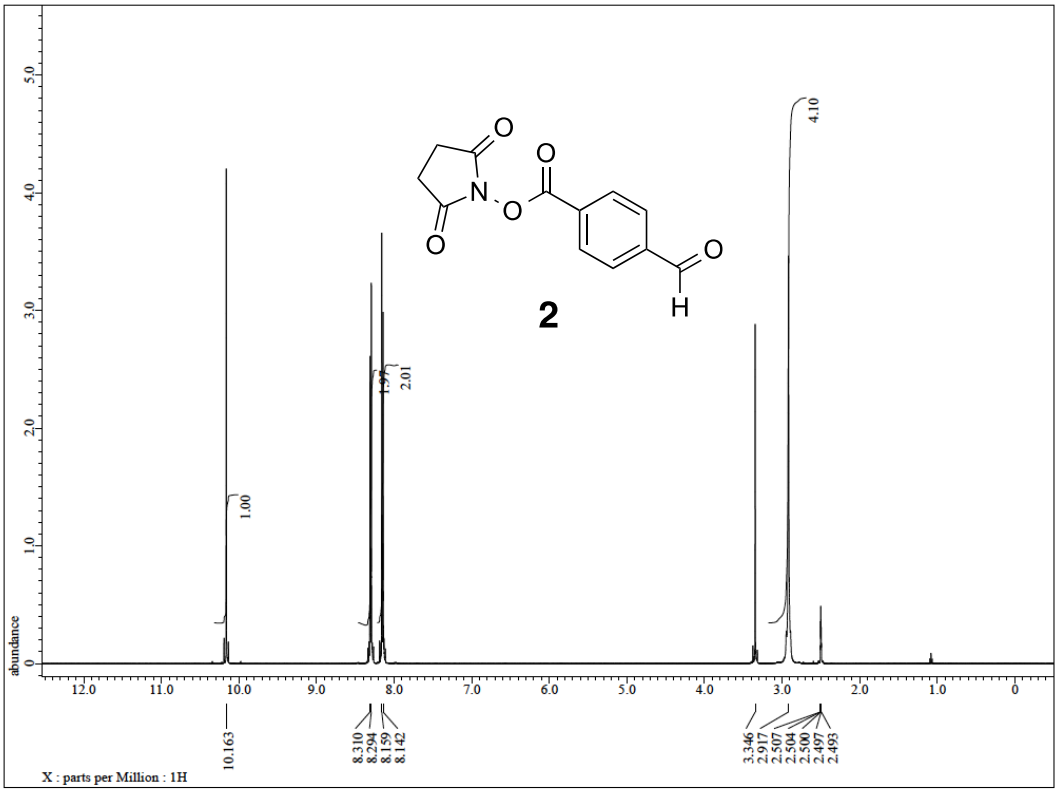
Step 3: Synthesis of 2,5-dioxopyrrolidin-1-yl (E)-4-(2-(4-(((2,5-dioxopyrrolidin-1-yl) oxy) carbonyl) benzylidene) hydrazinyl) benzoate



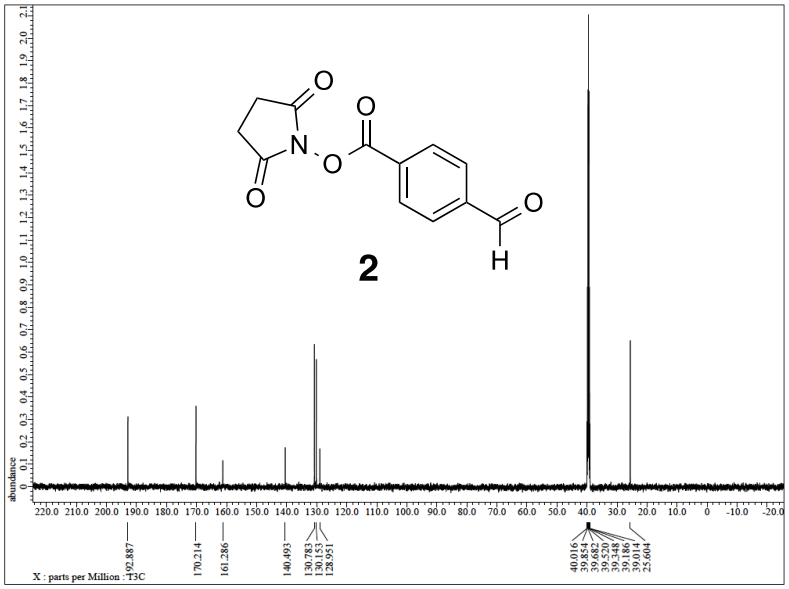
To a 5 mL roundbottom flask was added 2,5-dioxopyrrolidin-1-yl 4-formylbenzoate (2) (61.8 mg, 0.25 mmol), 2,5-dioxopyrrolidin-1-yl 6-(2-(propan-2-ylidene)hydrazinyl)nicotinate (4) (69.3 mg, 0.25 mmol) and 3.0 mL of glacial acetic acid. The reaction was stirred overnight, resulting in a bright yellow solution. Slow addition of 20 mL of diethyl ether caused a bright yellow precipitate to form in a clear solution. The bright yellow solid was washed several times with diethyl ether and dried under vacuum to yield 35.6 mg (30%) of pure 2,5-dioxopyrrolidin-1-yl (E)-4-(2-(4-(((2,5-dioxopyrrolidin-1-yl)oxy)carbonyl)benzylidene)hydrazinyl)-benzoate (5).

^{1}H & ^{13}C NMR Spectra

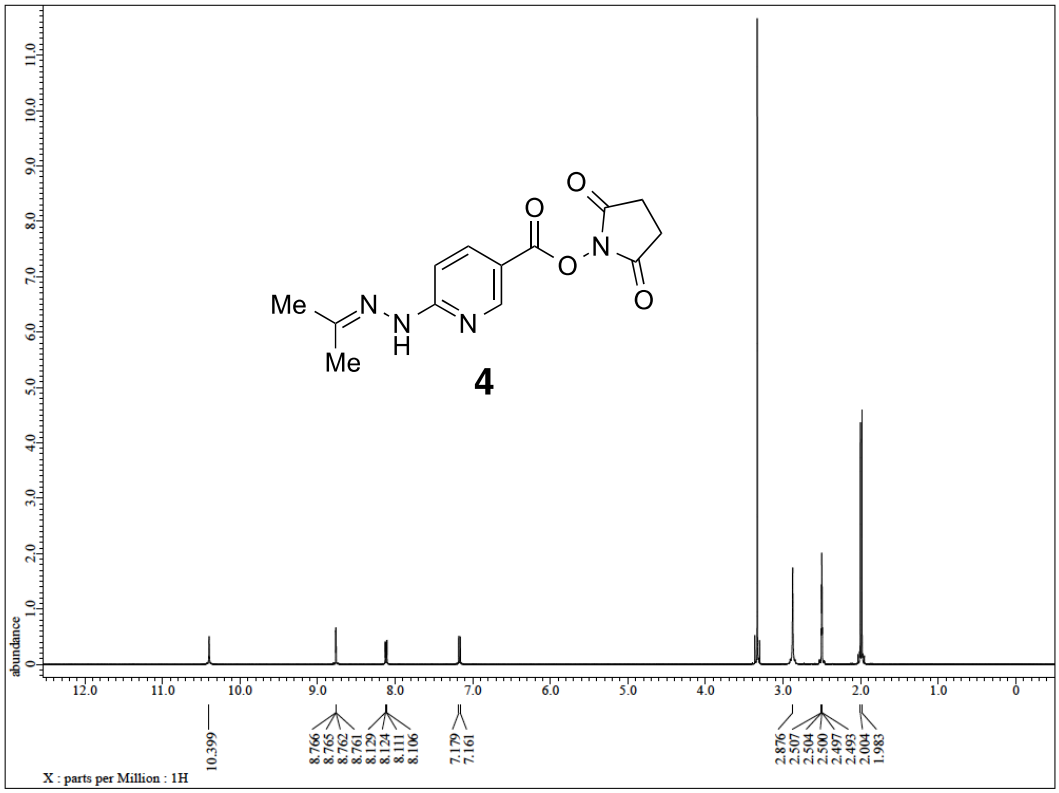
^{1}H NMR (500 MHz, DMSO-d_6)



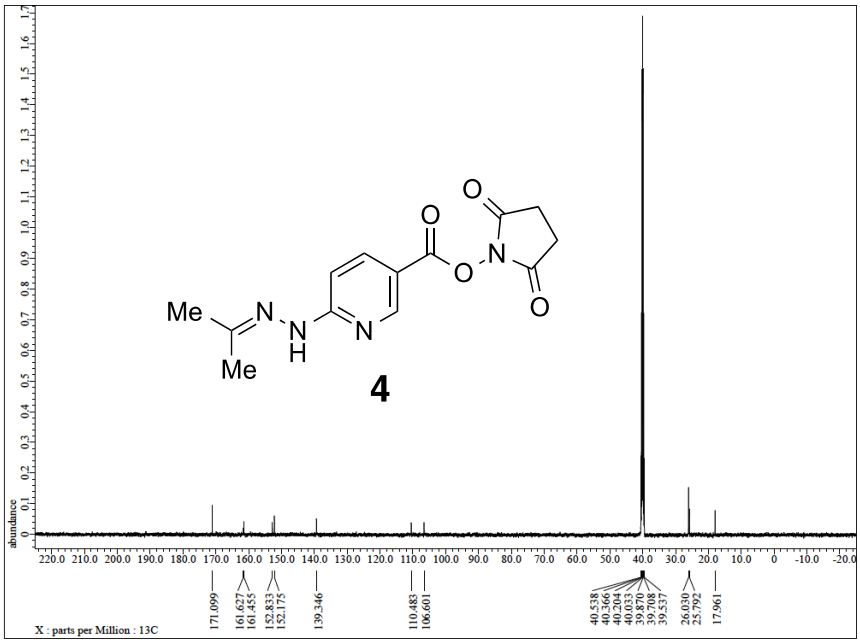
^{13}C NMR (125 MHz, DMSO-d_6)



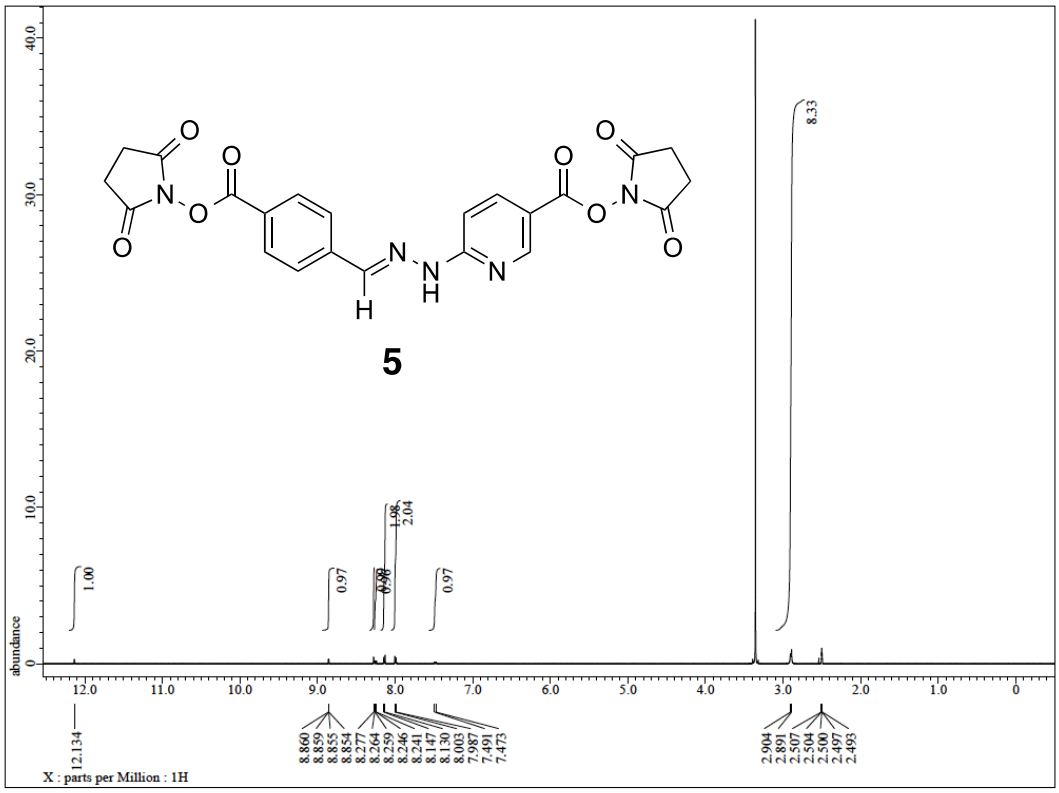
^{1}H NMR (500 MHz, DMSO-d_6)



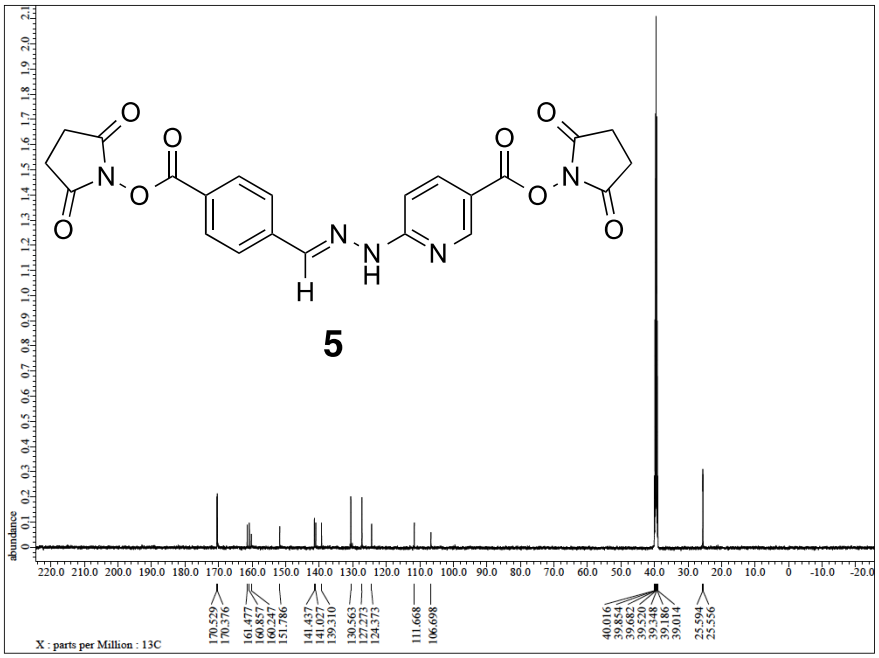
^{13}C NMR (125 MHz, DMSO-d_6)



^{1}H NMR (500 MHz, DMSO-d_6)



^{13}C NMR (125 MHz, DMSO-d_6)



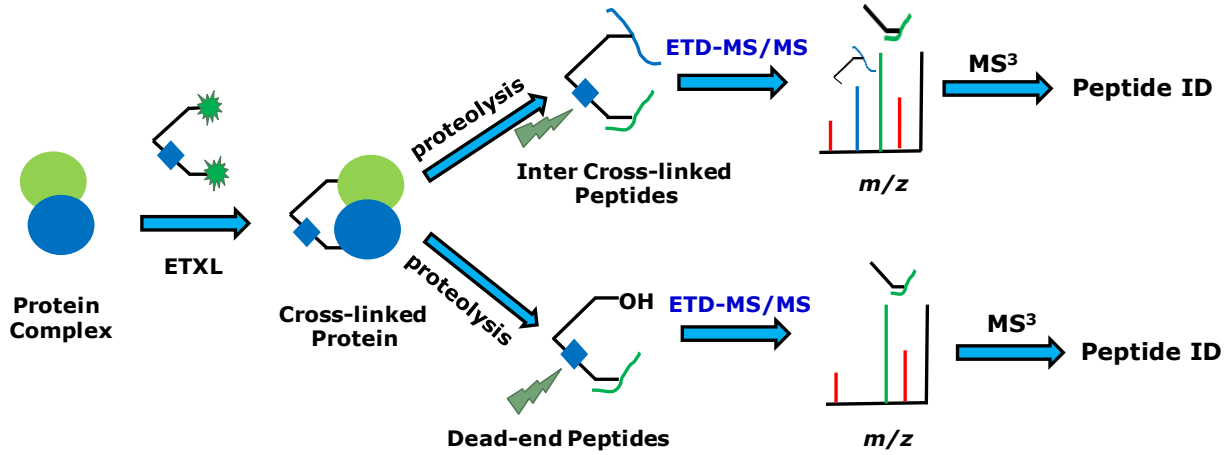


Figure C-S1: The scheme to identify inter-cross-linked and dead-end peptides by ETD tandem mass spectrometry. MS3 can be used for sequence confirmation of cross-linked modified fragments.

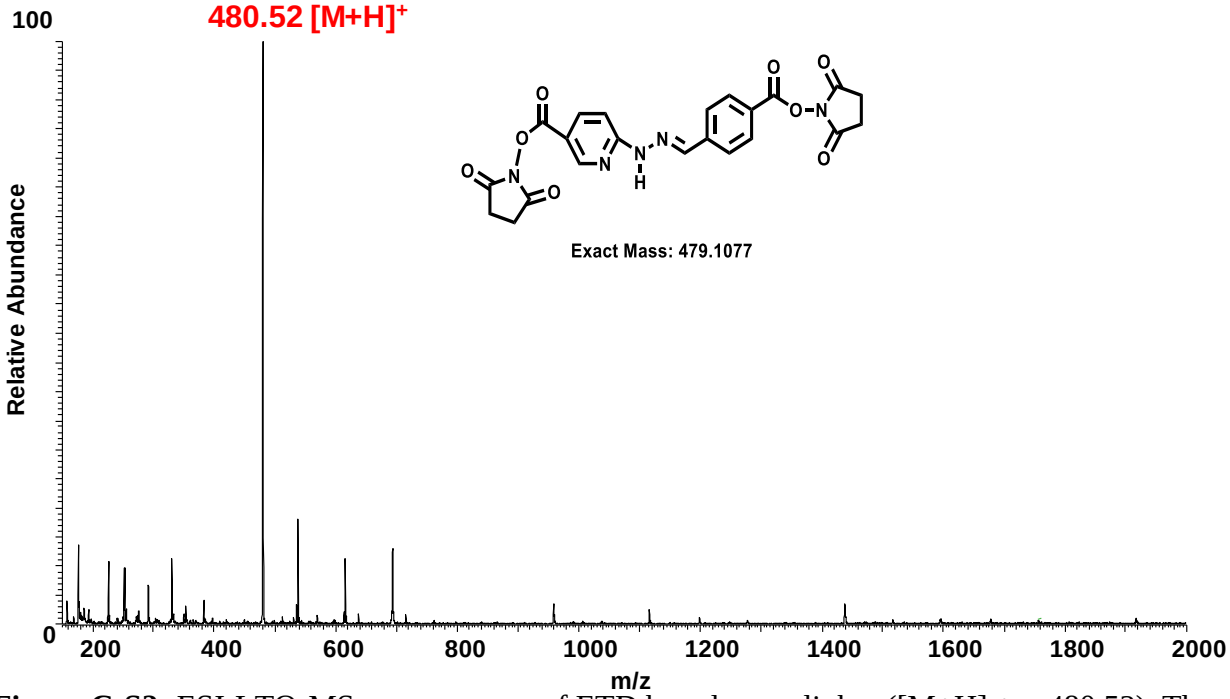


Figure C-S2: ESI-LTQ-MS measurement of ETD based cross-linker ([M+H] + = 480.52). The molecular mass of the cross-linker is 479.1077 Da.

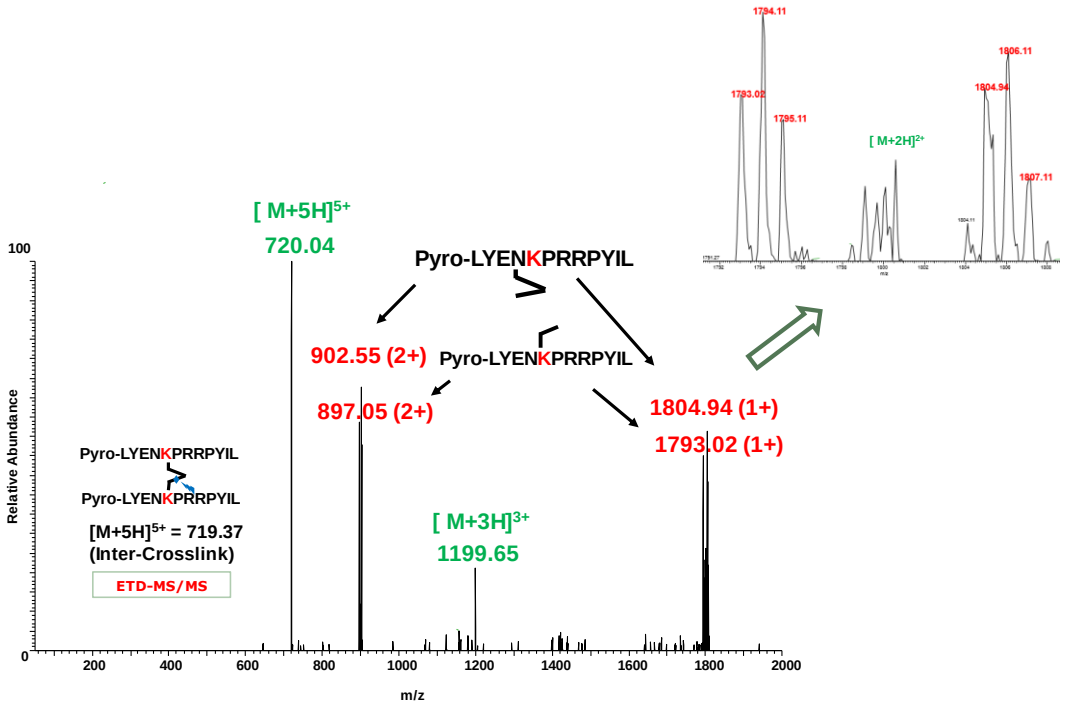


Figure C-S3: ETD mass spectra of a Neurotensin dimer identified by ETD-XL.

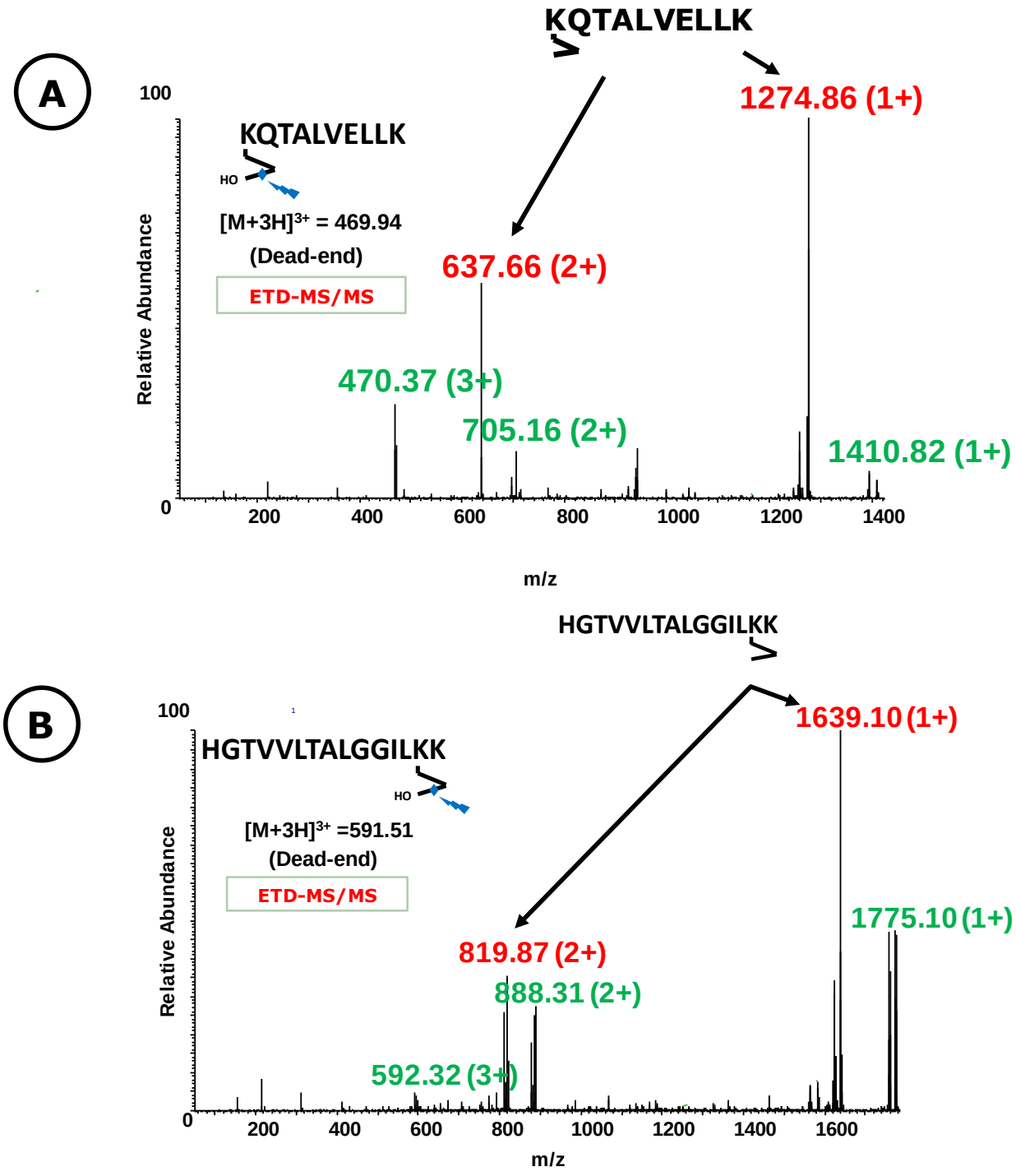


Figure C-S4: ETD mass spectra of dead-end peptides derived from (A) BSA and (B) Myoglobin. ETD mass spectra generate one signature ion for the dead-end peptide.

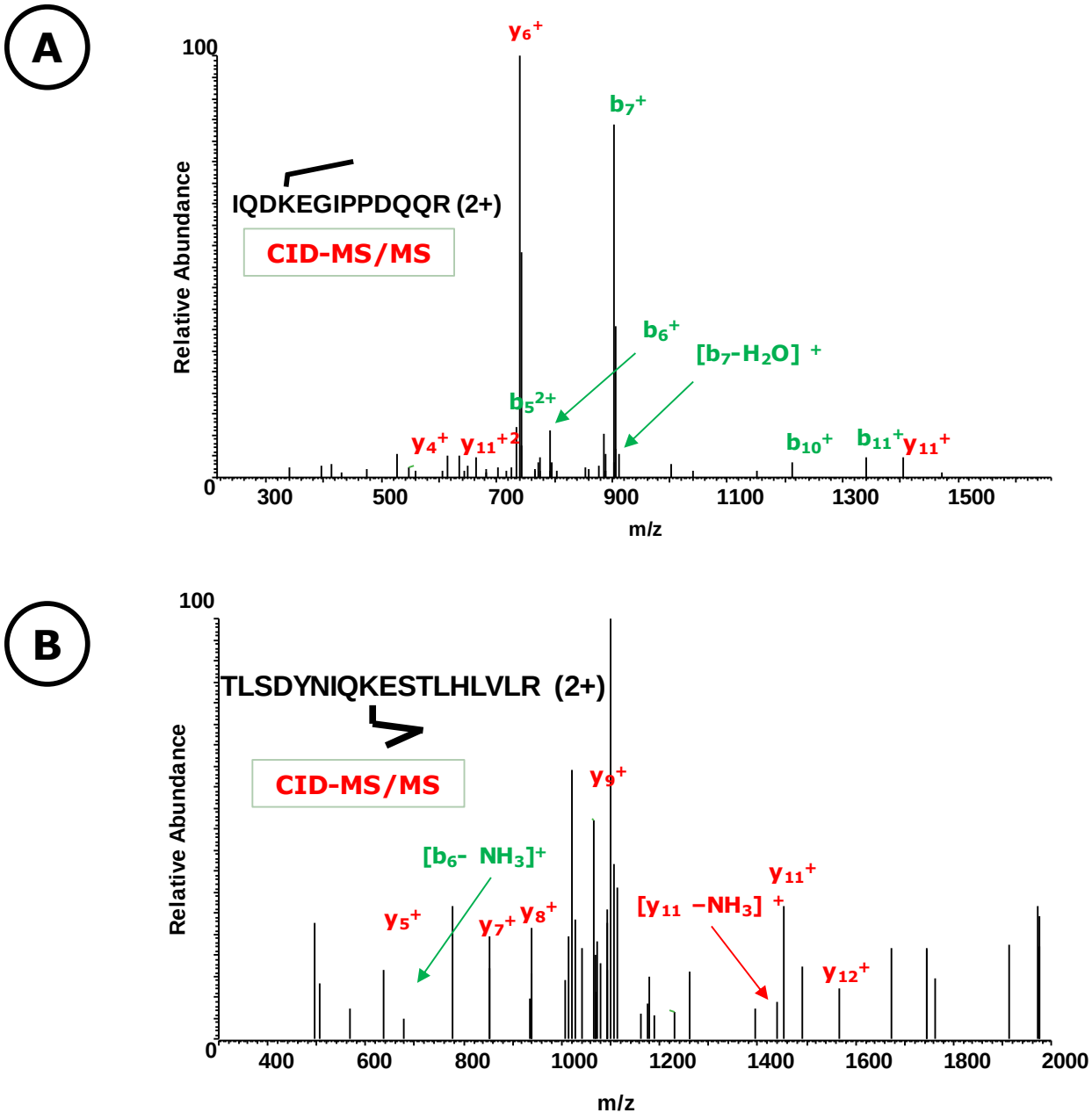


Figure C-S5: CID-MS/MS of doubly charged signature ions IQDKEGIPPDQQR (A) and TLSDYNIQKESTLHLVLR (B) provide the sequence information of the modified peptides.

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLE
DGRTLSDYNIQKESTLHLVLRLRGG

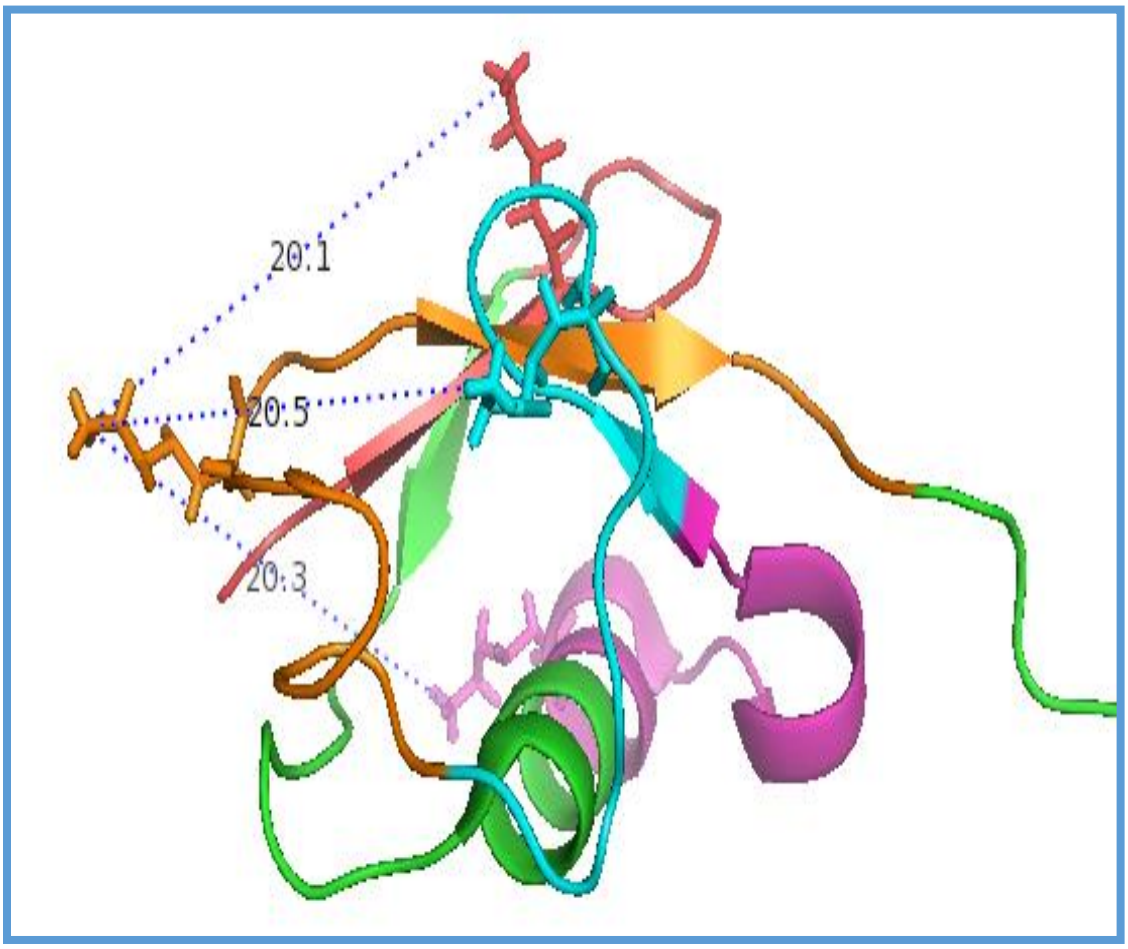


Figure C-S6: X-ray crystallographic structure of Ubiquitin (pdb- 1V80) with the distance constrains between interacting residues.

GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASED
LKKHGTVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFG
ADAQGAMTKALELFRNDIAAKYKELGFQG

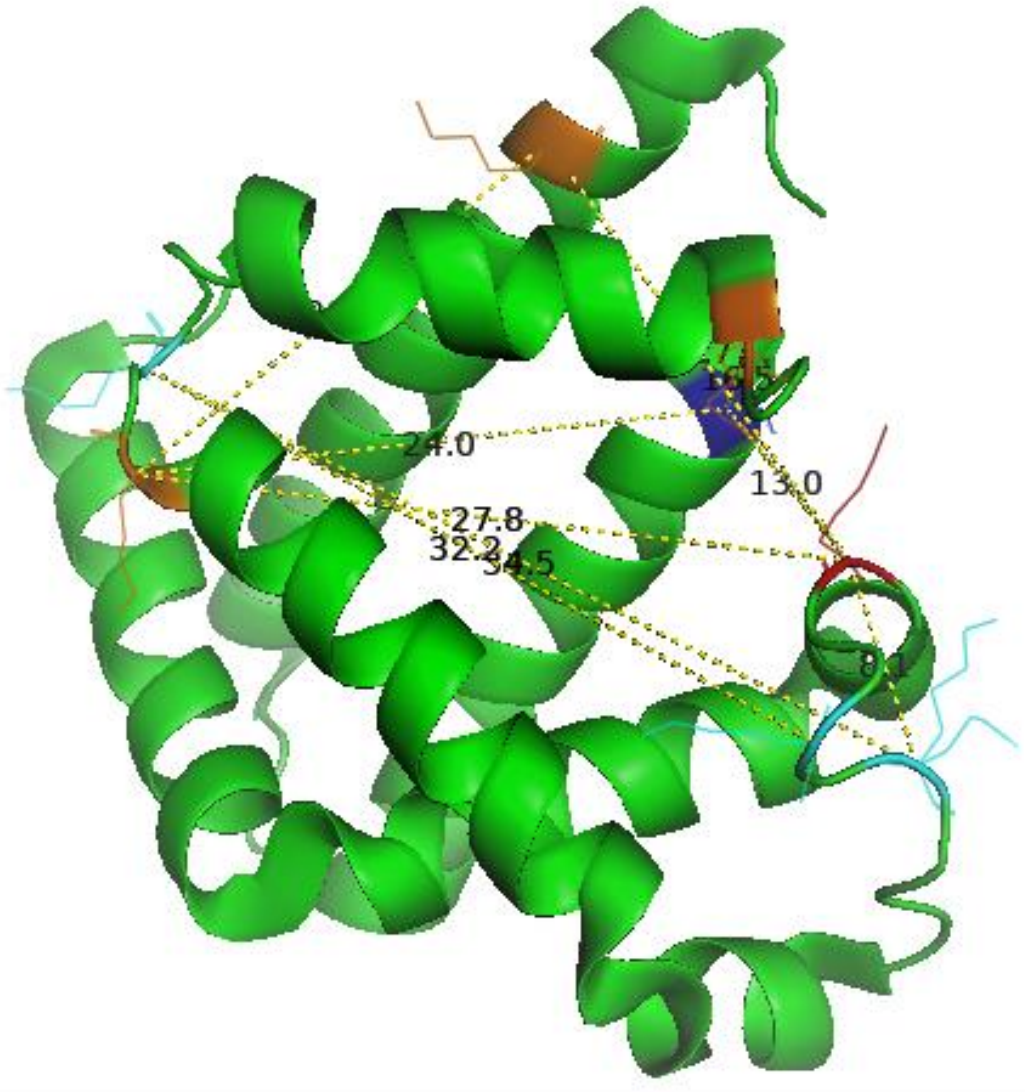


Figure C-S7: Crystal structure of Myoglobin (pdb-1dwr) with the color-coded sites and distances. High confidence identification clearly pointed out the cross-linked lysine residues.

DTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGC
EKSLHTLFGDELCKVASLRETYGDMADCCEKQEPERNECFLSHKDDSPDLPKLKPDPNTLCD
EFKADEKKFWGKYLYEIARRHPYFYAPELLYYANKYNGVFQECCQAEDKGACLLPKIETMRE
KVLTSSARQRLRCASIQKFGERALKAWSVARLSQKFPKAEFVEVTKLVTDLTKVHKECCHGD
LLECADDRADLAKYICDNQDTISSKLKECCDKPLLEKSHCIAEVEKDAIPENLPPLTADFAE
DKDVCKNYQEAKDAFLGSFLYEYSRRHPEYAVSVLLRLAKEYEATLEECCAKDDPHACYSTV
FDKLKHLVDEPQNLIKQNCDQFEKLGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGT
RCCTKPESERMPCTEDYLSLILNRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETY
VPKAFDEKLFTFHADICTLPDTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDKCC
AADDKEACFAVEGPKLVVSTQTALA

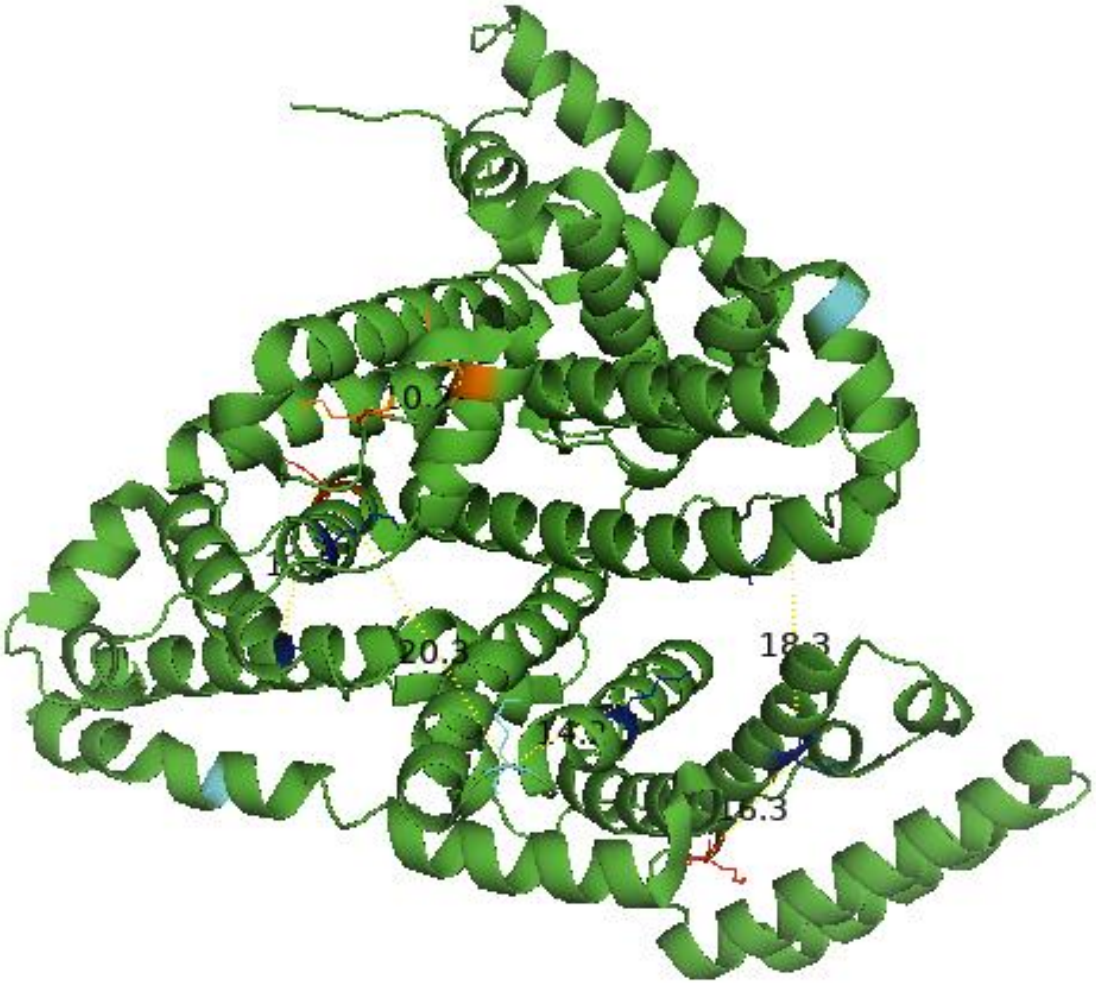


Figure C-S8: Crystal structure of BSA showing the distance between the interacting peptides, identified by HI-ETD-XL (pdb :4fs)

Table C-S1: List of calculated masses and m/z values used to identify cross-linked reaction products derived from Bradykinin.

	Exact Mass	**m/z**
Crosslinker (HI-ETD-XL)	479.1077	480.1150 (1+)
NHS (reactive group)	114.0191	
Bradykinin +HI-ETD-XL (Inter-crosslink)	(1059.5614×2) (2peptides) +479.1077 (Cross-linker) -(114.0191×2) (2 leaving group)-2.0156 (2 Hydrogen) = 2368.1767	593.0515 (4+) 474.6426 (5+)
Bradykinin + HI-ETD-XL (Dead-end)	1059.5614 (peptide) + 479.1077 (Cross-linker) -(114.0191×2) (2 leaving group) +17.0027 (-OH) -1.0078 (Hydrogen) = 1326.6258	1327.6331 (1+) 664.3202 (2+) 443.2159 (3+)
Bradykinin + Fragment residue mass 1 (ETD)	1059.5614+132.0449 -1.0078=1190.5985	1191.6058 (1+) 596.3065 (2+) 397.8735 (3+)
Bradykinin + Fragment residue mass 2 (ETD)	1059.5614+120.0324 -1.0078 =1178.5860	1179.5933 (1+) 590.3003 (2+) 393.8693 (3+)

Table C-S2. Summary of HI-ETD-XL-interlinked peptides of BSA identified by LC-MS/MS analysis

Sequence	Cross-linked Lysine residue	Distance (Å)
ALKAWSVAR--------LAKEYEATLEECCAK	K(211)…....K(350)	13.5
KFWGK--------ADEKKFWGK	N/A	N/A
KQTALVELLK--------HKPK	K(524)…....K(535)	16.3
HKPK--------QTALVELLKHKPK	N/A	N/A
LSQKFPK--------CCTKPESER	K(221)…....K(439)	20.3
LKECCDK--------PLLEKSHCIAEVEK	K(275)…....K(285)	10.2
PDPNTLCDEFKADEK--------DDPHACYSTVFDKLK	K(127)…....K(375)	47.4
DDSPDLPKLK--------KQTALVELLK	K(114)…....K(524)	18.3
SLGKVGTR--------CCTKPESER	K(431)…....K(439)	14.2

Table C-S3: Calculated masses and m/z values used to identify cross-linked reaction products derived from Myoglobin.

	Exact mass	**m/z**	**m/z**
LFTGHPETLEKFDK……. IPIKYLEFISDAIIHVLHSK + HI-ETD-XL (Inter-crosslink)	(**1660.8460… 2335.3303**) + 479.1077 (Cross-linker) -(114.0191×2) (2 leaving group)-2.0156 (2 Hydrogen) = 4245.2302	4246.2375(1+) 2123.6224(2+) 1416.0840 (3+) 1062.3148 (4+) 850.05334 (5+)	
LFTGHPETLEKFDK + Fragment residue mass 1 (ETD)	**1660.8460**+120.0324 - 1.0078= 1779.8706 **1660.8460** +132.0449 - 1.0078= 1791.8831	1780.8779(1+) 890.9426(2+) 594.2975(3+)	1792.8904(1+) 896.9488(2+) 598.3017 (3+)
IPIKYLEFISDAIIHVLHSK + Fragment residue mass 2 (ETD)	**2335.3303** + 132.0449 - 1.0078=2466.3674 **2335.3303** + 120.0324 - 1.0078=2454.3549	2467.3747(1+) 1234.1910(2+) 823.1297(3+)	2455.3622(1+) 1228.1847(2+) 819.1256(3+)

Table C-S4: List of calculated masses and m / z value for assigning cross-linking products (BSA)

	Exact mass	**m/z**	**m/z**
ALKAWSVAR…… LAKEYEATLEEC!C!AK + HI-ETD-XL (Inter-crosslink)	(**1000.5818… 1813.8226**) + 479.1077 (Cross-linker) -(114.0191×2) (2 leaving group)-2.0156 (2 Hydrogen) = 3063.4583	3064.4656 (1+) 1532.73645 (2+) 1022.1601(3+) 766.8719 (4+) 613.6990 (5+)	
ALKAWSVAR + Fragment residue mass 1 (ETD)	**1000.5818**+120.0324 - 1.0078= 1119.6064 **1000.5818**+132.0449 - 1.0078= 1131.6189	1120.6137(1+) 560.8105(2+) 374.2094(3+)	1132.6262(1+) 566.8167 (2+) 378.2136 (3+)
LAKEYEATLEEC!C!AK + Fragment residue mass 2 (ETD)	**1813.8226**+ 132.0449 - 1.0078=1944.8597 **1813.8226**+ 120.0324 - 1.0078=1932.8472	1945.8670(1+) 973.4372 (2+) 649.2939 (3+)	1933.8545(1+) 967.4309(2+) 645.2897 (3+)

Table C-S5. List of inter cross-linked peptides along with distance constrains identified from Ubiquitin

Sequence	Cross-linked Lysine residue	Distance (Å)
IQDKEGIPPDQQR-------- TLSDYNIQKESTLHLVLR	K(33)…….....K(63)	20.1
LIFAGKQLEDGR--------LIFAGKQLEDGR	N/A	N/A
LIFAGKQLEDGR-------- TLSDYNIQKESTLHLVLR	K(48)…….....K(63)	20.3
MQIFVKTLTGK-------- TLSDYNIQKESTLHLVLR	K(6)…….....K(63)	20.5

Table C-S6: List of calculated masses and m/z value used to assign cross-linked products

(Ubiquitin)

	Exact mass	**m/z**	**m/z**
TLSDYNIQKESTLHLVLR …. IQDKEGIPPDQQR + HI-ETD-XL (Inter-crosslink)	**(2129.1480… 1522.7739**) + 479.1077 (Cross-linker) -(114.0191×2) (2 leaving group)-2.0156 (2 Hydrogen) = 3900.9758	3901.9831(1+) 1951.4952(2+) 1301.3325(3+) 976.2512 (4+) 781.2025 (5+)	
TLSDYNIQKESTLHLVLR + Fragment residue mass 1 (ETD)	**2129.1480**+120.0324 - 1.0078= 2248.1726 **2129.1480**+132.0449 - 1.0078= 2260.1851	2249.1799(1+) 1125.0936(2+) 750.3982 (3+)	2261.1924(1+) 1131.0998(2+) 754.4023(3+)
IQDKEGIPPDQQR + Fragment residue mass 2 (ETD)	**1522.7739**+ 132.0449 - 1.0078=1653.811 **1522.7739**+ 120.0324 - 1.0078=1641.7985	1654.8183(1+) 827.9128(2+) 552.2776(3+)	1642.8058(1+) 821.9065(2+) 548.2735 (3+)

Table C-S7: Table of Confident Identifications of Inter-Crosslinked Peptides from Myoglobin

Sequence	Cross-linked Lysine residue	Distance (Å)
LFTGHPETLEKFDK--------FKHLK	K(42)…...K(47)	8.1
LFTGHPETLEKFDK--------NDIAAKYK	K(42)…...K(145)	19.7
NDIAAKYK--------HGTVVLTALGGILKK	K(145)…..K(77)	20.4
KGHHEAELK--------FDKFKHLK	K(79)…...K(45)	32.2
KGHHEAELK--------FKHLK	K(79)…...K(47)	34.5
HGTVVLTALGGILKK-------- LFTGHPETLEKFDK	K(145)…..K(42)	27.8
IPIKYLEFISDAIIHVLHSK-------- LFTGHPETLEKFDK	K(102)…..K(42)	13
PLAQSHATKHKIPIK-------- IPIKYLEFISDAIIHVLHSK	K(96)…..K(102)	16.5
IPIKYLEFISDAIIHVLHSK-------- HGTVVLTALGGILKK	K(102)…..K(77)	24

Biographical Information

Jayanta Kishor Chakrabarty was born and raised in Comilla, Bangladesh. He obtained his bachelor's and master's degree in pharmaceutical science from the University of Dhaka. His master's thesis mainly focused on Bioequivalence Studies of Omeprazole 20 mg Tablet in Healthy Bengali volunteers. After his graduation, he started his career as a lecturer at Northern University Bangladesh (NUB) in July 2012. After serving for one year, he moved to another renowned institution, BRAC University. His intellectual pursuits and desire for deeper understanding, eventually, brought him to the United States in 2014. His research mainly concentrated on the development of mass spectrometry based cleavable crosslinking approaches for deciphering protein structure and protein to protein interactions. He received his doctorate degree in Analytical Chemistry from the University of Texas at Arlington in December 2019. He also worked for Shimadzu Center for Advanced Analytical Chemistry (SCAAC) in Arlington as a research intern. He aspires to continue his research adventure either in academia or industry.