



WICHITA STATE  
UNIVERSITY

UNIVERSITY LIBRARIES

## Boronic acid: A versatile handle for the enrichment of cross-linked peptides

Item Type	Thesis
Authors	Thole, Mitali
Publisher	Wichita State University
Rights	© Copyright 2022 by Mitali Thole All Rights Reserved
Download date	2026-05-17 23:33:43
Link to Item	<a href="https://soar.wichita.edu/handle/10057/24965">https://soar.wichita.edu/handle/10057/24965</a>

BORONIC ACID- A VERSATILE HANDLE FOR THE ENRICHMENT OF CROSS-LINKED  
PEPTIDES

A Thesis by

Mitali Thole

Bachelor of Science, Savitribai Phule Pune University, 2019

Submitted to the Department of Chemistry and Biochemistry  
and the faculty of the Graduate School of  
Wichita State University  
in partial fulfillment of  
the requirements for the degree of  
Master of Science

December 2022

© Copyright 2022 by Mitali Thole

All Rights Reserved

BORONIC ACID- A VERSATILE HANDLE FOR THE ENRICHMENT OF CROSS-LINKED  
PEPTIDES

The following faculty members have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfilment of the requirement for the degree of Master of Science, with a major in Chemistry.

---

Haifan Wu, Committee Chair

---

Moriah Beck, Committee Member

---

Coleen Pugh, Committee Member

---

Maojun Gong, Committee Member

---

George Bousfield, Committee Member

## DEDICATION

To my family and my country, INDIA.

## ACKNOWLEDGEMENTS

I would like to acknowledge my supervisor, Dr. Haifan Wu, who made this work possible. His guidance and advice improved me through all the stages of this project. I would also like to thank my committee members, for their time, comments, and for pushing me intellectually. I was grateful to have had helpful lab mates- Emmanuel A. Gayani S. for all her help during my initial time at WSU and Hyma M. for keeping up with me as a flatmate and a lab mate, too. I thank Dr. Kevin L., for his assistance on various instruments in the chemistry department, specially NMR. I will also thank stockroom members Mary C. and Susan M. for their assistance during lab teachings and for ordering chemicals as required.

I would also like to thank Dr. Abegel F., for reviewing my work and sharing valuable tips and recommendations while I was writing my thesis. I couldn't have made this far without her help. Thanks to Maya E. and Rahin K. for making my experience as a graduate student more entertaining. My friends, Prithvi M., Madhuri D., Dhruv S., Shashank S., Talha K., and Pragna K., always supported me throughout my Master's degree. Twinkle T., Aarya K., Sakshi B., Greesha M., Nidhi D., Amruta S., Rishi C., and others always encouraged me to do better. I would also like to thank my sister Niyati T. for always being there for me.

Last, but not least, I would like to thank my grandparents, and my entire family in India. From the time I decided to go to graduate school, my parents have always supported my journey and have provided encouragement when I needed it most. My mom and dad always believed in me, listened and gave me valuable advice at my lowest points. I could not have accomplished this degree without their support and blessings. My gratitude goes toward the Wichita State University Chemistry and Biochemistry Department faculty, staff, and students for their support.

## ABSTRACT

Virtually all cellular processes in living organisms rely on protein machinery of complex three-dimensional structures and protein-protein interactions (PPIs). Therefore, significant effort has been devoted for elucidating protein structures and mapping PPI networks to understand relevant biological processes. Cross-linking mass spectrometry (XL-MS) is a robust technique that can provide conformational restraints within proteins and protein complexes for modelling protein structures and analysing PPI networks on a whole proteome scale both *in vitro* and *in vivo*. In XL-MS, protein samples are chemically cross-linked, digested by proteases, and analysed by LC-MS/MS to identify cross-linked peptides. However, one challenge associated with XL-MS is the sample complexity. Because of the low abundance of cross-linked peptides in the mixture, it is challenging to identify cross-linked peptides. The aim of this thesis is to develop a robust affinity purification (AP) protocol to enrich cross-linked peptides prior to LC-MS/MS analysis. Boronic acids (BAs) can form reversible covalent bonds with diols and they are highly versatile as building blocks of many useful chemical transformations. In addition, they have low toxicity as several BA containing drugs have been approved by FDA. Therefore, we investigated the use of boronic acid as a handle for enriching cross-linked peptides. We have synthesized two BA-containing cross-linkers and a diol-containing resin for AP. Using model proteins, protein complexes and cell lysates, we found that BA-based enrichment is highly sensitive and specific. The BA cross-linkers can also be used for live-cell cross-linking due to its cell permeability. Therefore, we believe this new enrichment handle will further expand the power of XL-MS in the analysis of complex samples.

## TABLE OF CONTENTS

Chapter	Page
CHAPTER 1.....	1
1.1 Importance of Proteins.....	1
1.1 Protein-Protein Interactions (PPIs).....	4
1.2 Mapping PPIs.....	7
1.2.1 Far Western Blotting (WB).....	8
1.2.2 Yeast two-Hybrid Assay (Y2H).....	9
1.2.3 Fluorescence-based assays: fluorescence resonance energy transfer (FRET), and bioluminescence resonance energy transfer (BRET).....	9
1.2.4 X-ray Crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy.....	10
1.2.5 Affinity Purification-Mass Spectrometry (AP-MS).....	11
1.2.6 Proximity Labelling (PL).....	12
1.3 Overview of Cross-linking Mass Spectrometry (XL-MS).....	14
1.3.1 Protein/Peptide Cross-linking.....	15
1.3.2 Anatomy of a Cross-linker.....	19
1.3.3 XL-MS Work-flow.....	29
1.3.4 Applications of XL-MS.....	41
1.4 Boronic acid (BA) as a Bio-orthogonal handle.....	45
1.4.1 Chemical and Physical properties of BAs.....	48
1.4.2 Stability of Boronic Acids.....	51
1.4.3 Applications of Boronic Acids.....	54
1.5 Current Cross-linkers.....	61
1.6 Conclusion.....	63
CHAPTER 2.....	65
2.1 Introduction.....	65
2.2 Materials.....	65
2.3 Instrumentation.....	67
2.4 Synthetic Schemes.....	69
2.4.1 Synthesis of Shorter Cross-linker (SMT).....	69
2.4.2 Synthesis of Longer Cross-linker (LMT).....	70
2.4.3 Synthesis of Diol Resin.....	71
2.5 Results and Chemical Synthesis and Analytical Data.....	72
2.5.1 S1: Synthesis of (3,5-bis(((2,5-dioxopyrrolidin-1-yl)oxy)carbonyl)phenyl)boronic acid.....	72
2.5.2 S2: Total Synthesis of (3,5-bis(2-((2,5-dioxopyrrolidin-1-yl)oxy)-2-oxoethyl)phenyl)boronic acid and corresponding intermediates.....	75

## TABLE OF CONTENTS (continued)

Chapter	Page
2.5.3 S3: Total Synthesis of (1R,2S,3S,5R)-6,6-dimethyl-2-((4-methyl-1H-1,2,3-triazol-1-yl)methyl)bicyclo[3.1.1]heptane-2,3-diol .....	83
2.6 Automated Peptide Synthesis .....	91
2.6.1 N-terminal modifications of model peptides .....	91
2.6.2 Synthesis of Cross-linked Peptide .....	93
2.7 Results and Discussion .....	94
2.8 Conclusion .....	95
 CHAPTER 3.....	 97
3.1 Introduction.....	97
3.2 Peptide-level Enrichment.....	97
3.2.1 Enrichment on Model Linear Peptides .....	98
3.2.2 Enrichment on model Cross-linked Peptides .....	99
3.3 Optimization of the Enrichment Protocol .....	100
3.3.1 Effect of Wash .....	101
3.3.2 H <sub>2</sub> O <sub>2</sub> Oxidation for Elution.....	102
3.3.3 Methionine Oxidation .....	104
3.3.4 Amount of resin .....	106
3.4 Affinity Purification.....	108
3.4.1 Protocol for Enrichment.....	108
3.5 Protein-level Enrichment .....	109
3.5.1 Cross-linking on a Model Protein.....	110
3.5.2 Enrichment on Model proteins.....	111
3.5.3 Cross-linking and Enrichment of HeLa Cell Lysate.....	112
3.5.4 Live-cell cross-linking using SMT in live Hela cells .....	113
3.6 Results and Discussion .....	114
3.7 Conclusion .....	115
3.8 Future Directions .....	116
 REFERENCES.....	 118

## LIST OF FIGURES

Figure	Page
1.1: Major functions of proteins in the human body. ....	1
1.2: The four levels of protein structure. ....	2
1.3: Illustration of protein-protein interactions as a strategy towards characterization of unknown proteins, predicting inter-domain interactions and potential drug targets. ....	4
1.4: Different types of direct protein interactions based on differences in interaction interfaces, stability and binding affinity. ....	5
1.5: Identifying novel protein interactions: Atomic resolution, Mass spectrometry, Bio-physical and Computational methods. ....	8
1.6: Protein bioconjugation reaction using a chemical cross-linking reagent. ....	17
1.7: Transglutaminase mediated formation of intermolecular isopeptide bond via Enzymatic cross-linking reaction. ....	18
1.8: Photo-induced covalent cross-linking via benzophenone based crosslinkers for the analysis of PPIs. ....	19
1.9: Anatomy of a cross-linker. Different reactive groups and spacer scaffolds, and multiple different cleavable, releasable and enrichable groups can be used. ....	20
1.10: -NHS and Imido-ester cross-linking reaction scheme for chemical conjugation to a primary amine. ....	23
1.11: EDC (Carbodiimide) cross-linking reaction scheme for chemical conjugation to a carboxylic acid. ....	24
1.12: Maleimide cross-linking reaction scheme for chemical conjugation to a sulfhydryl. ....	25
1.13: Examples of Cross-linkers having varying solubilities due to different end reactive groups. ....	26
1.14: Examples of Hetero-bifunctional and Homo-bifunctional cross-linkers. ....	27
1.15: Disulphide bridge in DTBP allows for easy cleavage of a protein conjugate using common reducing agents. ....	28

## LIST OF FIGURES (continued)

Figure	Page
1.16: General work-flow strategy for XL-MS without enrichment: The overall workflow is divided into five major steps as displayed. ....	31
1.17: Modified XL-MS work-flow with enrichment step prior to LC-MS analysis. ....	32
1.18: Steps involved in the work-flow of Affinity Purification involves: equilibration, binding washing and elution of protein of interest from the solid support. ....	36
1.19: Typical steps followed in bottom-up and top-down proteomics. ....	38
1.20: Annotation of peptide fragmentation ions b-and y-, a-and x-, as well as c-and z-, ions generated most commonly upon collision-induced dissociation and electron capture dissociation of peptide ions. ....	39
1.21: General scheme for a bioconjugation reaction using a payload forming a bioconjugate with enhanced physical and chemical properties. ....	47
1.22: Application of BAs as carbohydrate sensors, due to their complexation with polyols to form neutral boronate esters. ....	47
1.23: Common oxygen-containing Organoboron compounds and their related applications. ....	48
1.24: Structure and bonding of BAs. ....	49
1.25: Ionization of BA in water, when $\text{pH} > \text{pKa}$ forms tetrahedral anionic ester. ....	50
1.26: BAs behave as Lewis acids due to the vacant p-orbital activating it towards nucleophilic attack by nucleophiles. ....	51
1.27: Oxidative hydroxylation of aryl BAs to form phenols at rt in presence of oxidising agents or air. ....	53
1.28: Acid and Base catalysed Protodeboronation of BAs. ....	54
1.29: BAs for sensing and other general applications. ....	55
1.30: BAs are commonly employed as serine proteases due to their reaction with serine side chain residues forming anionic boronate ester. ....	56
1.31: BA backbone in pharmacologically relevant compounds. ....	57
1.32: Mechanism of BA and diol complexation in aqueous solutions. ....	60
1.33: Deprotection of BEs. 1: Oxidative cleavage using $\text{H}_2\text{O}_2$ . 2. Transesterification. ....	61
1.34: Examples of cross-linkers with common Enrichment handles. ....	63

LIST OF FIGURES (continued)

Figure	Page
2.1: Shorter cross-linker (SMT) .....	72
2.2: <sup>1</sup> H-NMR for SMT .....	74
2.3: <sup>13</sup> C-NMR for SMT .....	74
2.4: Longer cross-linker (LMT) .....	75
2.5: 1-bromo-3,5-bis(bromomethyl)benzene .....	75
2.6: 2,2'-(5-bromo-1,3-phenylene)diacetonitrile .....	76
2.7: 2,2'-(5-bromo-1,3-phenylene)diacetic acid .....	77
2.8: dimethyl 2,2'-(5-bromo-1,3-phenylene)diacetate .....	78
2.9: dimethyl 2,2'-(5-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1,3-phenylene)diacetate .....	79
2.10: Potassium salt of dimethyl 2,2'-(5-(trifluoro-λ4-boranyl)-1,3-phenylene)diacetate .....	80
2.11: 2,2'-(5-borono-1,3-phenylene)diacetic acid .....	81
2.12: (3,5-bis(2-((2,5-dioxocyclopentyl)oxy)-2-oxoethyl)phenyl)boronic acid .....	82
2.13: Diol resin .....	83
2.14: 4-nitrophenyl pent-4-ynoate .....	84
2.15: Activation of agarose alkyne beads .....	85
2.16: (-)-nopol-azide-diol .....	87
2.17: (-)-nopol-tosylate .....	87
2.18: (-)-nopol-azide .....	88
2.19: (-)-nopol-azide-diol .....	89
2.20: 1R,2S,3S,5R)-6,6-dimethyl-2-((4-methyl-1H-1,2,3-triazol-1-yl)methyl)bicyclo[3.1.1]heptane-2,3-diol .....	90
2.21: N-terminal acetylated model peptide .....	92
2.22: MS confirmed, N-terminal acetylated peptide, model peptide, MT-A-122 .....	92
2.23: MS confirmed, N-terminal BA containing, model peptide, MT-A-122 .....	93

## LIST OF FIGURES (continued)

Figure	Page
2.24: MS confirmed cross-linked peptide, MT-A-130.....	94
2.25: Enrichment of BA cross-linked peptides using diol containing resin followed by oxidative cleavage using H <sub>2</sub> O <sub>2</sub> Binding scheme. ....	96
3.1: Control run, 50μM MT-A-121 and MT-A-122 .....	98
3.2: After Binding Flowthrough.....	99
3.3: After Elution Flowthrough.....	99
3.4: Control run, Peptide mixture.....	100
3.5: After Binding Flowthrough.....	100
3.6: After Elution Flowthrough.....	100
3.7: Control run, 50μM MT-A-130.....	102
3.8: Buffer wash.....	102
3.9: Urea Wash.....	102
3.10: Control run, 50μM MT-A-130.....	103
3.11: Elution using 5mM H <sub>2</sub> O <sub>2</sub> .....	103
3.12: Elution using 100mM H <sub>2</sub> O <sub>2</sub> .....	104
3.13: Control run, 50μM of Ac-AAAMAARF .....	105
3.14: Elution using 5mM H <sub>2</sub> O <sub>2</sub> , 10 min shaking.....	105
3.15: Elution using 5mM H <sub>2</sub> O <sub>2</sub> , 15 min shaking.....	105
3.16: Elution using 100mM H <sub>2</sub> O <sub>2</sub> , 10 min shaking.....	105
3.17: Control run, 50μM MT-A-130.....	106
3.18: 100μL resin .....	107
3.19: 50μL resin .....	107
3.20: 25μL resin .....	107

## LIST OF FIGURES (continued)

Figure	Page
3.21: 10 $\mu$ L resin .....	108
3.22: Steps involved in Cross-linked Peptide Enrichment.....	109
3.23: Cross-linking using SMT on model protein 14-3-3 .....	110
3.24: Cross-linking using LMT on model protein 14-3-3.....	111
3.25: a) Enrichment of SMT and LMT on 14-3-3 and BSA .....	112
3.26: Live-cell cross-linking of SMT in Hela cells.....	114

## LIST OF SCHEMES

Scheme	Page
2.1: Synthesis of (3,5-bis(((2,5-dioxopyrrolidin-1-yl)oxy)carbonyl)phenyl)boronic acid .....	69
2.2: Synthesis of (3,5-bis(2-((2,5-dioxopyrrolidin-1-yl)oxy)-2-oxoethyl)phenyl)boronic acid. ....	70
2.3: Synthesis of (1R,2S,3S,5R)-6,6-dimethyl-2-((4-methyl-1H-1,2,3-triazol-1-yl)methyl)bicyclo[3.1.1]heptane-2,3-diol. ....	72
2.4: S3 Total synthesis scheme for (-)-nopol-azide-diol .....	87

## LIST OF TABLES

Table	Page
1: Classification of Proteins ( <sup>1</sup> ) .....	3
2: Popular cross-linker functionalities for protein conjugation ( <sup>40</sup> ) .....	21
3: List of Common Algorithms for Cross-Link Data Analysis ( <sup>41</sup> ).....	41
4: Identification of peptide pairs under 5% FDR control.....	113

# CHAPTER 1

## INTRODUCTION

### 1.1 Importance of Proteins

Proteins are large macromolecules making up nearly 60% of the cell's dry weight.<sup>1</sup> The high percentage of these biomolecules is attributed to their involvement in virtually all biological processes in all the major living organisms, as shown in Figure 1.1.<sup>2,3</sup> Proteins are regarded as the structural and functional units of life; they not only provide structural support to all the body's tissues but also maintain and repair them.<sup>1</sup>

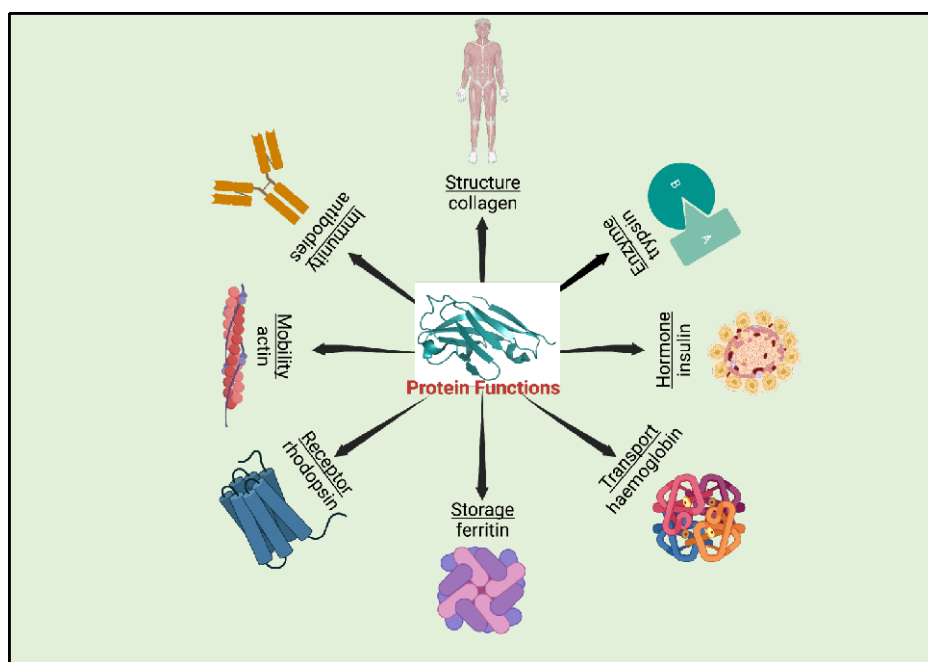


Figure 1.1: Major functions of proteins in the human body.  
Adapted from:<sup>3</sup>

Proteins are known to function as enzymes, hormones, signal transducers, etc.<sup>1</sup> The dynamic mobility of proteins determines their capability to carry out such critical physiological tasks. The chemical activities of a protein molecule are governed by specifically incorporated moving parts of proteins

whose mechanical actions are complementary to these chemical events.<sup>1</sup> This dynamic nature of proteins is, in turn, controlled by the three-dimensional array of structures which forms the core of molecule<sup>4</sup>.

Protein complexes consist of a straightforward primary sequence of amino acids which eventually folds into a well-defined three-dimensional shape.<sup>1</sup> Figure 1.2 describes the various classes of protein structures: primary, secondary, tertiary and quaternary. Apart from this, Table 1 describes the classification of proteins based on their biological function. The amino acid residues are linked to each other in long chains known as polypeptide chains through peptide bonds which are formed by a simple condensation reaction between the amino group and the carboxy terminus forming the backbone of a protein molecule.<sup>2</sup> Each protein has its unique sequence of amino acids, which describes the innate diversity in chemical properties among the different classes of proteins.<sup>2</sup>

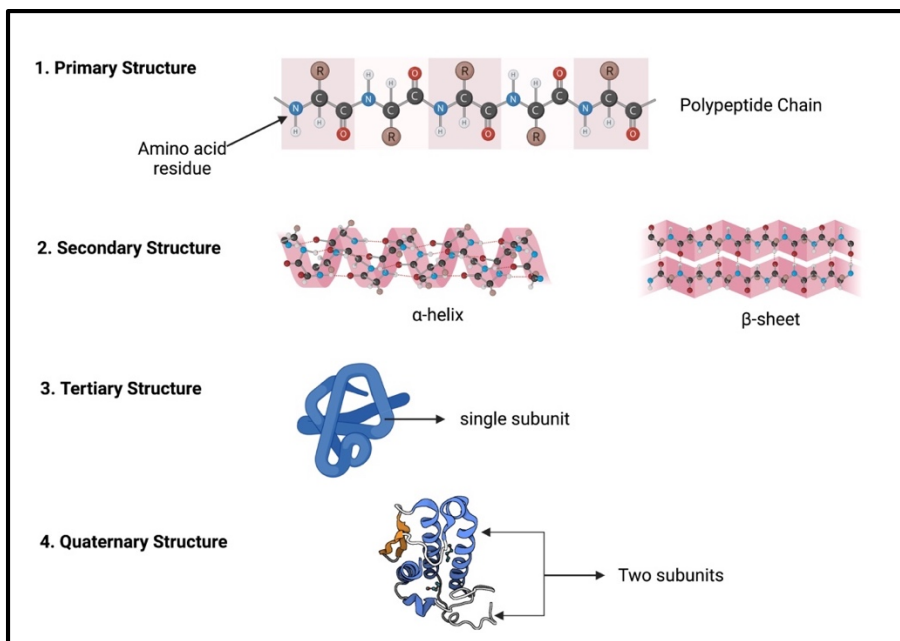


Figure 1.2: The four levels of protein structure.  
Adapted from: “Protein Structure”, by BioRender.com (2022).

Amongst the twenty amino acid residues that make up proteins in the human bodies, several side chains are either positively or negatively charged, while some may have a polar or neutral nonpolar character.<sup>5</sup> Oppositely charged positive and negative residues can form strong ionic bonds, whereas the polar and neutral non-polar side chain interactions give rise to rigid hydrogen bonds.<sup>2</sup> These non-covalent interactions induce specific polarities in the protein molecule, making it either hydrophobic or hydrophilic, thus increasing its susceptibility to bind with other biomolecules to carry out its essential functions.<sup>5</sup> The range of different amino acid sequences, sizes, three-dimensional shapes, and chemical characteristics reflect protein function in practically all cellular operations.<sup>2</sup>

Generally described as the workhorses of the body, proteins can perform significant functions single-handedly. Still certain biological activities rely on binding of proteins with different biomolecules.<sup>1</sup>

Table 1: Classification of Proteins ( <sup>1</sup> )

<b>Type</b>	<b>Examples</b>	<b>Biological Function</b>
1. Structural	Actin, keratin, myosin etc.	Provides strength and structure to body tissues.
2. Enzyme	Pepsin, lactase, amylase etc.	Digestion of monomers by hydrolysis of macronutrients for absorption.
3. Hormones	Insulin, glucagon etc.	Chemical messengers that initiate specific responses.
4. Transport	Haemoglobin, albumin, carrier proteins etc.	Transport of essential molecules across cellular membranes.
5. Storage	Ovalbumin, casein etc.	Storage of amino acids for later use.
6. Defence	Collagen, antibodies etc.	Recognize and respond to foreign pathogens.
7. Receptor Proteins	$\beta$ -adrenergic receptor	Receive and respond to chemical signals.

## 1.1 Protein-Protein Interactions (PPIs)

While many proteins function independently, most of them are known to interact with other proteins to regulate biological processes at both cellular and system levels.<sup>6</sup> Several analyses revealed that nearly 80% of proteins involved in similar biochemical processes are repeatedly found to be interacting with each other.<sup>7</sup> Protein interfaces serve as the primary mediators of protein-protein interactions.<sup>6</sup>

The physical contact between two or more different protein molecules are protein-protein interactions (PPIs). These networks originate from simple non-covalent linkages, like electrostatic forces, hydrogen bonds, or hydrophobic interactions, arising due to the different side-chain residues present on the surface of proteins, thus making PPIs highly specific.<sup>5</sup> As shown in Figure 1.3, known proteins can be used as bait proteins to determine the chemical functions of an unknown protein or a previously identified protein. These non-covalent interactions are proven to govern protein assembly and protein function.<sup>8</sup>

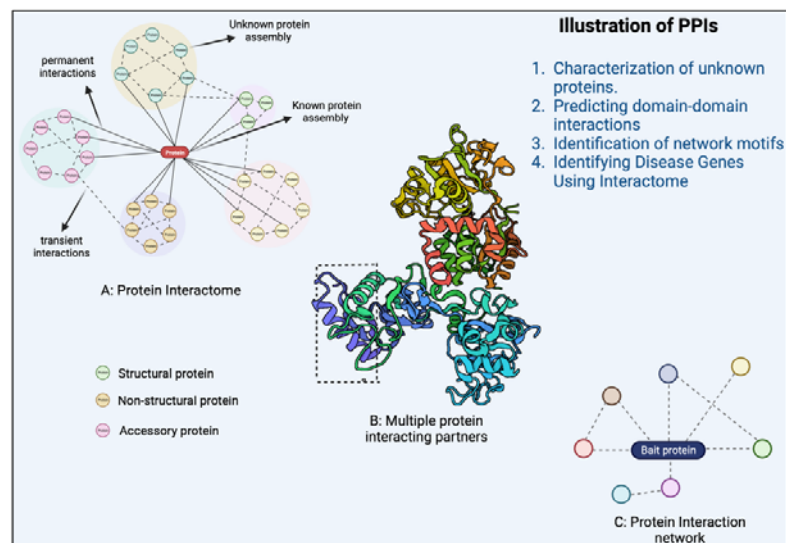


Figure 1.3: Illustration of protein-protein interactions as a strategy towards characterization of unknown proteins, predicting inter-domain interactions and potential drug targets. Adapted from: “Protein-protein Interactions”, by BioRender.com (2022).

Proteins must explicitly recognize a wide variety of distinct interaction partners to carry out nearly all cellular functions. Therefore, several protein interactions with comparable characteristics can all be classified into different categories based on contrasting structural and functional properties.<sup>7</sup> The most fundamental classification depends on the composition of the interacting partners; homo-oligomers are networks comprised of identical proteins, whereas hetero-oligomers are complexes formed of diverse protein partners.<sup>7,9</sup> Another differentiation can be made depending upon the stability, either into an obligate or non-obligate type of interaction can be seen in Figure 1.4. Non-obligate interactions are unstable *in vivo*, while the obligate protomers can exist as stable structures. The PPIs may either be temporary or permanent depending on the binding affinity of the protein partners.<sup>5</sup>

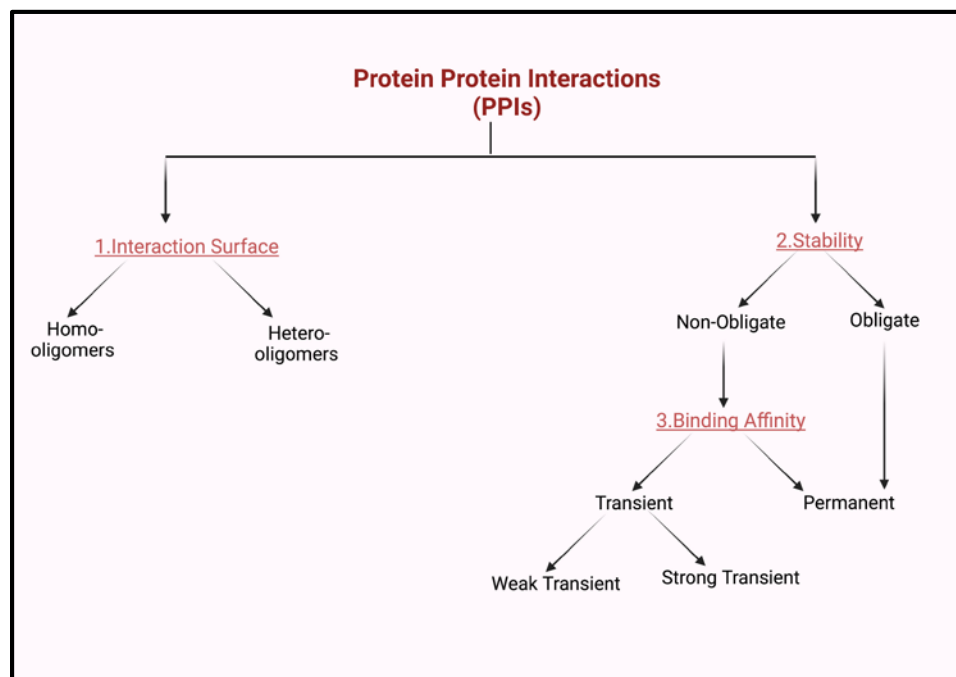


Figure 1.4: Different types of direct protein interactions based on differences in interaction interfaces, stability and binding affinity. Adapted from:<sup>7</sup>

As shown in Figure 1.4, all obligate PPIs are permanent, but not all interactions are obligate.<sup>7</sup> In comparison, the non-obligate interactions can either be transient or permanent.<sup>7</sup> Additionally, the

protein interactions that switch from an unbound or weakly bound state to a highly bound state are often triggered by an effector molecule. This interaction falls under the strong transient category.<sup>5,9</sup>

Protein-protein interactions modulate the scaffolding of protein complexes and are fundamental in carrying out critical functions mandatory in cellular processes in all living organisms. To comprehend the cell's biochemistry, it will be helpful to resolving and characterize the relationship between proteins in a specific proteome.<sup>10</sup> Significant time has been devoted to determining these physical contacts between PPIs and predicting the mechanism underlying complex molecular relationships.<sup>8</sup> A detailed study of PPIs can help infer information for the following:<sup>5,10</sup>

1. Predicting the functionality of unidentified proteins based on their interactions with a protein whose function is already known
2. PPIs help predict and identify potential drug targets on the surface of proteins
3. PPIs are known to alter the kinetics of enzymes leading to elusive changes in substrate binding site and allosteric effects
4. Act as a general mechanism to facilitate substrate channelling, allowing the diffusion of substrates between different protein domains and their related subunits
5. PPIs also form specific binding sites for small effector molecules.
6. PPIs can alter a protein's substrate specificity through interactions with various binding partners and,
7. PPIs can have a regulating role in an upstream or downstream event.

A variety of human disorders can result from defective PPIs because they can have severe repercussions on biochemical processes necessary for cell growth, proliferation, and homeostasis.<sup>8</sup> Evidence confirms that these aberrant PPIs are also associated with several neurodegenerative diseases like Alzheimer's disease.<sup>11</sup> Consequently, PPIs have subsequently been acknowledged as a new

paradigm for discovering novel therapeutics.<sup>8</sup> To fully comprehend protein function and regulation, and to describe the molecular mechanisms underlying human pathologies to find prospective targets for better therapeutic approaches, it is essential to map PPIs and their interaction interfaces in living cells.<sup>8</sup>

## 1.2 Mapping PPIs

The function and molecular properties of PPIs have been the subject of extensive research in cell, molecular and structural biology. Decades of research have given remarkable insights into these molecular networks that operate intricate physiochemical processes in living cells.<sup>1</sup> A critical step is to map these protein-to-protein physical interactions. The term "interactome", as shown in Figure 1.3, refers to a complete map of all possible protein interactions in a living organism.<sup>6</sup> The ongoing analysis of the protein interactome networks presents one of the significant challenges in contemporary biomedicine.<sup>2</sup> Numerous *in vivo* and *in vitro* approaches for analysing and characterizing these protein interactomes exist. These methods must be capable enough to:<sup>12</sup>

1. Demonstrate how to identify protein interaction partners
2. Evaluate PPIs quantitatively and qualitatively
3. Monitor these interactions as they occur in living cells, and
4. Determine interaction interfaces.

Conventionally, the techniques used are often differentiated into high-throughput and low-throughput approaches;<sup>7</sup> the former allows for screening many interactions, whilst the latter does not. Different bioassays have varying sensitivities, and while they can all typically identify permanent PPIs, only some can detect transient interactions.<sup>13</sup> Additionally, not all methodologies scale up in the same manner. Some can have substantial benefits on a smaller scale but become considerably more expensive and time-consuming as the number of interactions being examined increases.<sup>13,7</sup> The four

characterization techniques that can be used to study PPIs are included in Figure 1.5. These include atomic resolution techniques, mass spectrometry techniques, biophysical techniques, and computational techniques.

Some standard *in vivo* techniques for identifying and mapping PPIs include western blotting (WB), yeast two-hybrid (Y2H), fluorescence resonance energy transfer (FRET), and bioluminescence resonance energy transfer (BRET).<sup>14</sup> Some *in vitro* approaches include x-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, affinity purification-mass spectrometry (AP-MS), and protein labelling (PL) techniques.<sup>7,14</sup>

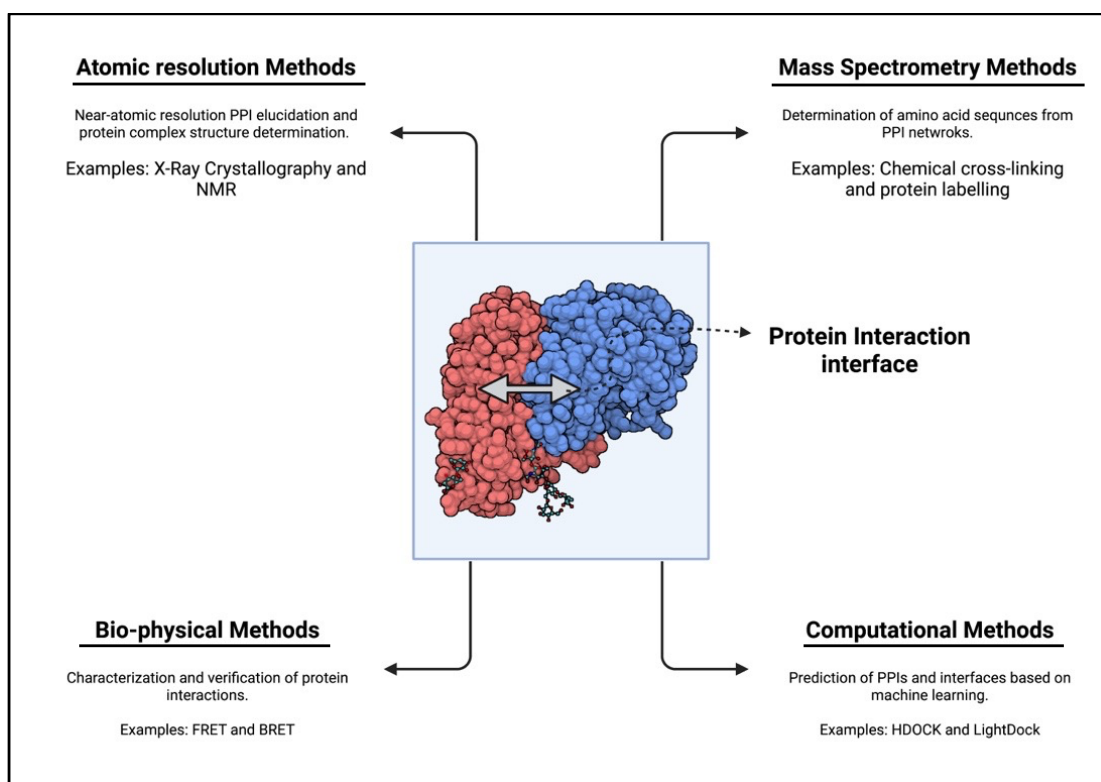


Figure 1.5: Identifying novel protein interactions: Atomic resolution, Mass spectrometry, Bio-physical and Computational methods.

Adapted from:<sup>15</sup>

### 1.2.1 Far Western Blotting (WB)

One of the earliest methods for identifying PPIs is WB, which relies on the specific binding of an unknown prey protein to an already purified known bait protein, forming a complex.<sup>14</sup> The WB approach helps detect direct links between these interacting protein pairs found on the surface of a membrane.<sup>14</sup> But inevitably, this arduous technique necessitates the isolation of at least one binding protein partner.<sup>14</sup>

### 1.2.2 Yeast two-Hybrid Assay (Y2H)

The Y2H assay developed almost 25 years ago,<sup>13</sup> is a robust technique used widely to research *in vivo* PPIs both in prokaryotes and eukaryotes.<sup>15</sup> Two functional components of a yeast transcription factor, a DNA-binding domain (BD) and a transcriptional activation domain (AD) are physically separated during a typical Y2H process. These two domains are necessarily required in a standard Y2H assay. The interaction between the two domains, fuses them together to form potential interacting proteins.<sup>15</sup> Then, the transcription factors activity is restored by bringing the two domains in close physical proximity. If the two proteins interact, it can generate a discernible output. Without interaction, the domains remain independent, preventing any detectable output.<sup>13</sup>

This strategy, however, has several drawbacks, including the overexpression of some proteins that results in non-specific interactions and the use of yeast as the host cell, which lacks the required cofactors and post-translational modifications. Therefore, this technique requires thorough analysis to identify biologically relevant PPIs due to the high false discovery rate.<sup>15</sup>

### 1.2.3 Fluorescence-based assays: fluorescence resonance energy transfer (FRET), and bioluminescence resonance energy transfer (BRET).

Proteomics extensively uses fluorescence-based techniques, such as FRET and BRET<sup>16</sup> in a high-throughput approach to characterize instantaneous PPIs in real-time, enabling the measurement of transient interactions with short half-lives.<sup>13</sup> The premise of the above-mentioned techniques is the non-radiative energy transfer from an excited donor fluorophore to an adjacent acceptor molecule.<sup>7</sup> The choice of the donor and acceptor is made so that their respective fluorophores' absorption and emission spectra overlap.<sup>16</sup>

The protein of interest (POI) must be fused to small fluorophore molecules serving as donors, and the other protein must be connected to any acceptor to perform FRET. While in BRET, the protein of interest is fused to a bioluminescent molecule, such as the Renilla Luciferase enzyme, which serves as the donor and the other interacting partner is linked to a fluorescent protein.<sup>13</sup> When the donor and acceptor fluorophores are brought close to each other due to the interaction of their fusion partners, there is an energy transfer from the excited donor to the surrounding acceptor resulting in the emission of a fluorescent signal. This emission signal is quenched, leading to an increase in the intensity of the acceptor signal, effectively determining PPIs.<sup>13</sup>

Though BRET has higher sensitivity than FRET, both techniques have limitations.<sup>16</sup> Not only do both assays require the fluorophores to be in close spatial proximity for effective energy transfer, but strong background signals tend to interfere with the output due to the auto-illumination of samples. Hence, control experiments are necessary to quantify the results obtained by the two techniques to avoid false positives.<sup>16</sup> Nevertheless, high-throughput methods can investigate many potential PPIs, but they come with their own set of challenges.

#### 1.2.4 X-ray Crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy

Prevailing low-throughput methods for assessing PPIs are x-ray crystallography and NMR spectroscopy.<sup>7</sup> These techniques aim to identify molecular fragments, verified by fragment libraries.<sup>17,11</sup> Due to the low molecular weight of these fragments, x-ray crystallography and NMR spectroscopy can provide valuable structural data by creating distinct atomic models for the interacting protein partners.<sup>11</sup> However, both techniques are associated with extensive analysis.<sup>17</sup> X-ray crystallography is restricted to the proteins that can be crystallized, while NMR spectroscopy can be applied to proteins that do not crystallize, though it requires the protein to be highly pure.<sup>17</sup>

#### 1.2.5 Affinity Purification-Mass Spectrometry (AP-MS)

Affinity purification (AP) coupled with mass spectrometry (MS) has evolved into a powerful and popular technique to detect PPIs globally due to the ease of sample preparation and new developments in bioinformatics.<sup>8</sup> Immobilizing a bait protein with a distinct functional group (FG) on a solid support, generally agarose beads or magnetic resin is the premise of this technique.<sup>15</sup> The loaded bait can explicitly sequester the intended POI from a mixture of soluble proteins.<sup>13</sup> Once isolated, the POI is subjected to protease digestion into small peptide fragments, followed by ionization and detection by liquid chromatography-mass spectrometry (LC-MS).<sup>13</sup> The appropriate choice of bait protein depends on the protein under investigation.

Not only are AP-MS experiments capable of profiling PPIs in a native cellular environment, thereby preserving the original protein structure, but it also allows the discovery of PPI networks with increased sensitivity, accuracy, adaptability, and speed.<sup>8</sup> So far, AP-MS has been identified as a successful technique capable of working with proteins in their natural endogenous origin like antibodies and also allowing interrogation of protein baits tagged with specific epitopes, for example, His-tagged

proteins.<sup>13</sup> Despite these advantages, AP-MS has several limitations demanding further investigation and improved strategies to overcome certain inefficacies.<sup>8</sup>

One limitation is that PPIs can reorganize due to structural manipulation following harsh biochemical cell lysis. Therefore, weak/temporary contacts can disappear during AP, making it impossible to characterize native PPIs fully.<sup>18</sup> Also, AP-MS has the disadvantage of being prone to picking up on indirect interactions along with direct physical interactions. Because many of the co-purified prey proteins are actually indirect interaction partners of the bait protein, they interact with the bait protein implicitly through a series of physical interactions involving other proteins in the complex leading to many false positives.<sup>19</sup>

#### 1.2.6 Proximity Labelling (PL)

Integrating different approaches yields the maximum level of precision.<sup>7</sup> Therefore, several cofractionating-based MS approaches, including cell engineering, PL, and others, have recently gained much attention in profiling global PPIs with better throughputs and overall versatility.<sup>18</sup>

PL approaches have contributed the most to interactome research out of all the newly described methods. In PL, the proteins are initially biotin-tagged since this co-enzyme has a high affinity for glycoproteins like avidin and streptavidin.<sup>20</sup> These biotinylated proteins are then subjected to conventional antibody-based APs to record and detect PPIs.<sup>21</sup> The foundation of this method is based on the strong affinity between biotin and the target protein that draws them together and enables the detection of weak and sporadic interactions as well.<sup>21</sup> PL can be classified into the following categories depending on the species of the labelling enzyme used to biotinylate the adjacent biomolecules.

##### 1.2.6.1 Proximity-dependent biotin identification coupled to mass spectrometry (BioID-MS)

As mentioned in the previous section, proteomic mapping has dramatically benefited from integrating of AP-MS with proximity-dependent biotin identification (BioID).<sup>21</sup> BioID enhances AP-MS by generating complementary information by labelling and identifying proximate proteins, whereas AP-MS can only provide information on PPIs in a stable protein complex.<sup>22</sup> The *E. coli* biotin ligase (BirA\*), having a particular site mutation (R11G), was used in the 2012 discovery, BioID approach.<sup>20</sup> This mutation destabilizes the ligase and releases biotin, which is sequestered by lysine side chains on proteins.<sup>21</sup> Now, this biotinylated protein permits its selective isolation utilizing a biotin affinity capture method based on avidin/streptavidin glycoproteins following MS identification to substantiate the molecular context of the protein.<sup>13</sup> Although PPIs can be examined in their native cellular environment,<sup>13</sup> this method is unreliable due to false negatives and requires protein purification due to low expression levels of PPI partners.<sup>22</sup>

#### 1.2.6.2 Engineered ascorbate peroxidase (APEX)-based PL

Another PL method, introduced right after BioID,<sup>21</sup> exhibits APEX as a genetic tag, allowing for systematic, temporally and spatially resolved proteome mapping.<sup>23</sup> Engineered APEX oxidizes biotin-phenols into biotin-phenoxy radicals in the presence of H<sub>2</sub>O<sub>2</sub>. These transient radicals biotinylate electron-rich amino acids like tyrosine on the surrounding proteins around a few nanometers radii.<sup>23</sup> APEX has the virtue of being faster than BioID, marking nearby proteins in a matter of minutes as compared to hours.<sup>24</sup> Scientists have determined the proteome map of the human mitochondrial matrix owing to APEX. APEX's relatively low cellular activity and sensitivity are one of its main drawbacks. This may be due to its poor heme binding, substandard folding/stability, or a combination of these elements.<sup>25</sup>

The techniques described above work well for characterizing and identifying potential interactors, but they have a several shortcomings.<sup>26</sup> The natural cellular environment is significantly harmed by cell lysis during the AP process because it impacts protein localization, interaction, and function.<sup>13,27</sup> Additionally, non-specific binding frequently generates high background levels, leading to false negatives in AP. Due to the necessity for reciprocal purification and the repetitive tagging of several baits to deconstruct proteome-wide PPI networks, the analysis throughput is constrained in AP methods.<sup>18</sup>

The worldwide characterization of transcription factors from cell lysates has been made possible by recent advancements in MS methods and complex-centric data analysis, which have boosted accuracy and efficiency.<sup>18</sup> Despite the effectiveness of this auspicious approach, certain AP experiments can cause biochemical manipulation under native purification conditions.<sup>8</sup> To circumvent this issue, cross-linking mass spectrometry (XL-MS) and AP have been combined to progress PPI investigations to make it easier to collect and identify protein complex interactions. *In vitro* chemical cross-linking has effectively stabilized protein interactions in native cells or tissues before subjecting them to cell lysis.<sup>8</sup>

### 1.3 Overview of Cross-linking Mass Spectrometry (XL-MS)

Cross-linking mass spectrometry, also referred to as XL-MS, CX-MS, or CL-MS, has recently been recognized as a productive method to supplement the limitations of the existing techniques by characterizing protein topologies, quantifying conformational states, and mapping dynamic PPIs on system-wide basis more thoroughly.<sup>28</sup> The uniqueness of XL-MS is based on its strikingly simple mechanistic work-flow which utilizes a chemical reagent to covalently cross-link proximate amino acid residues between two proteins in a protein complex, enabling identification of PPIs and multiple contact

sites in the same native chemical environment.<sup>18,29</sup> The use of XL-MS methods has been on the ascent since it is rapid and demands less sample purity than conventional methods.<sup>30</sup> The older approaches, like x-ray crystallography and NMR spectroscopy, are renowned for providing high-resolution 3D data on protein structure. However, recent techniques that offer low to medium-resolution information about PPIs can also provide necessary subsidiary information at different structural levels allowing XL-MS to gain more prominence over current approaches.<sup>29</sup> Subsidiary information includes information about protein architecture and folds.<sup>28</sup> Moreover, the combination of XL-MS with other structure elucidation techniques, such as cryogenic electron microscopy (cryo-EM), hydrogen-deuterium exchange mass spectrometry (HDX-MS), and others, have led the way for further investigation of structural information by identifying not only specific conformational changes but also to determine tertiary structures in cell lysates.<sup>28,29</sup>

Irrespective of the above advantages, certain aspects limit its use.<sup>30</sup> Complex samples containing different proteins typically yield non-cross-linked, linear peptides in relatively overwhelming numbers compared to the desired cross-linked peptides, which occur in sub-stoichiometric concentrations after protein digestion.<sup>28,30</sup> Therefore, routine detection of cross-linked peptides is difficult due to these sub-stoichiometric reaction efficiencies.<sup>31</sup> The following sections describe the highlights of the XL-MS approach.

### 1.3.1 Protein/Peptide Cross-linking

Chemically cross-linking proteins or biomolecules by covalently joining FGs on proximally close amino acids is the best approach to yield conjugates with superior functionalities in terms of biological and physical activity.<sup>32</sup> The non-static nature of proteins limits understanding their spatial organization, which is the key to resolving and understanding specific individual roles of proteins.

Furthermore, it is also essential to comprehend, integrate, predict, and reconstruct the purpose of the entire protein complex.<sup>28,33</sup> Therefore, it is difficult to iterate how protein complexes operate mechanistically. However, the ability to alter protein structure by introducing site-specific modifications opens the prospect of exploring protein activities at the molecular level. It makes it possible to identify and quantify flexible and dynamic regions of the proteins in a heterogenous complex.<sup>28,34</sup> Cross-linking is thus employed for a variety of reasons, such as:<sup>32</sup>

1. Stabilizing innate protein structures
2. Screening and identifying previously undiscovered protein interactors and their associated domains
3. Fusing a tag or an enzyme to purify antibodies
4. Immobilizing desired proteins from a protein complex to use in assays and AP analysis and,
5. Peptide binding to carrier proteins for handling and storage.

Protein chemistry utilises small cross-linking reagents for bioconjugation to study intermolecular interactions and develop novel conjugates. The cross-linking reaction can be chemical, enzymatic or based on the use of photo-activable species. The details of the three cross-linking techniques are described in the following sections:

#### 1.3.1.1 Chemical Cross-linking

Chemical cross-linking approaches take advantage of the fact that small bifunctional molecules containing a variety of reactive ends can attach to specific FGs found on proteins or other molecules, successfully modifying nucleic acids, drugs, and other polymers, as shown in Figure 1.6.<sup>35</sup> The availability of different FGs on the surface of proteins, due to the diversity of amino acid side-chain residues, renders them as potential and easy targets for protein bioconjugation. Characterization of PPIs

is challenging. Hence by adding selective stable chemical linkages, permanent interactions can be established which would otherwise be temporary. Over the years, protein studies and engineering have substantially benefited from the simplicity and dependability of utilizing cross-linking techniques.<sup>32</sup> The interactions can be developed between an amine to amine or thiol to thiol, etc. Chemical crosslinkers with homo-bifunctional or hetero-bifunctional functionality, (as seen in Figure 1.14) include two reactive groups to target proteins at the same or distinct functional groups, respectively.<sup>35</sup>

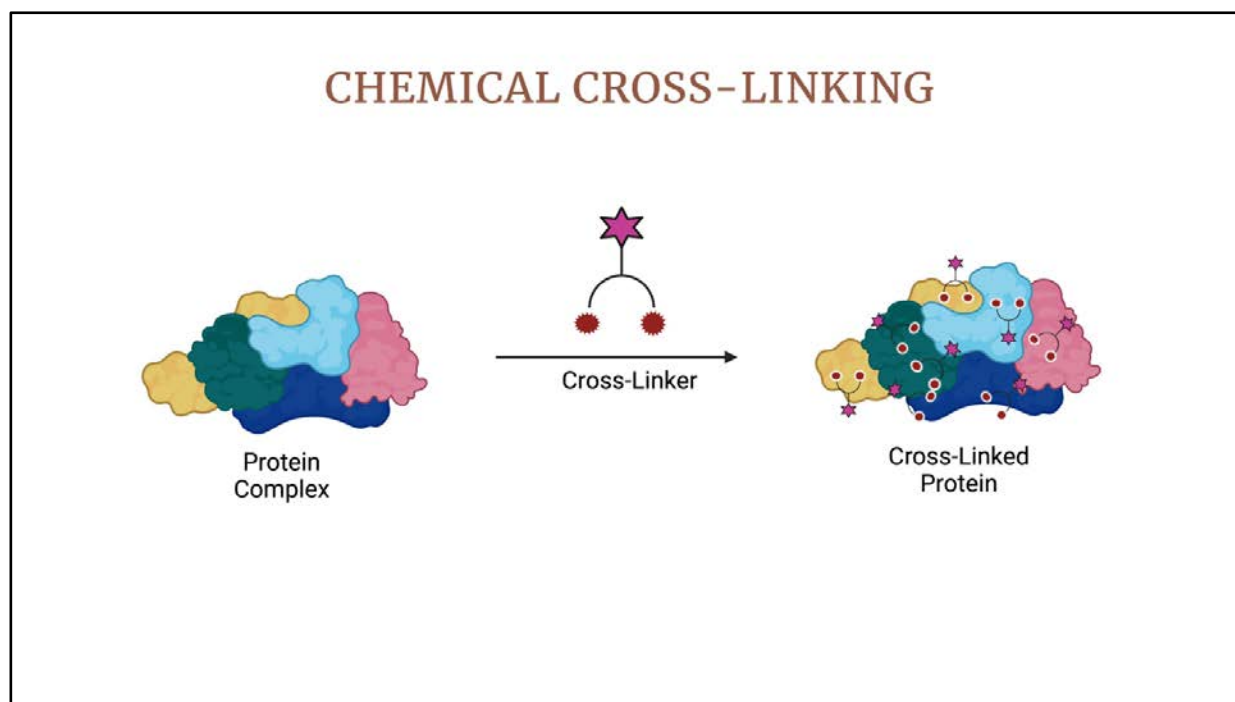


Figure 1.6: Protein bioconjugation reaction using a chemical cross-linking reagent.

#### 1.3.1.2 Enzymatic Cross-linking

Enzyme-based cross-linking strategies work better when chemical cross-linkers are not feasible for coupling site-restrained proteins like ubiquitin,<sup>35</sup> having various properties to resolve polymeric protein networks.<sup>35</sup> Several enzymes, including hydrolases, transferases, and others, serve as agitators to introduce post-translational modifications such as glycosylation, lipidation, etc., to further

enhance the proteome functionality.<sup>35</sup> Figure 1.7 depicts how transglutaminase (Tgase) can catalyse cross-linking of lysine and glutamine residues enzymatically by forming a peptide bond between the amino group and acyl group. For manufacturing and pharmaceutical purposes, where residual organic compounds may not be tolerated to comply with safety issues, selecting enzyme-based cross-linking strategies is preferable. Enzymatic cross-linking approaches have unique advantages over chemical cross-linking reagents in several ways:

1. Highly selective, site-specific cross-linking
2. Fast reaction kinetics
3. Does not require the use of harsh organic solvents or unusual pH conditions, works well in optimal conditions and,
4. Maintains and preserves the structural integrity of the proteins.

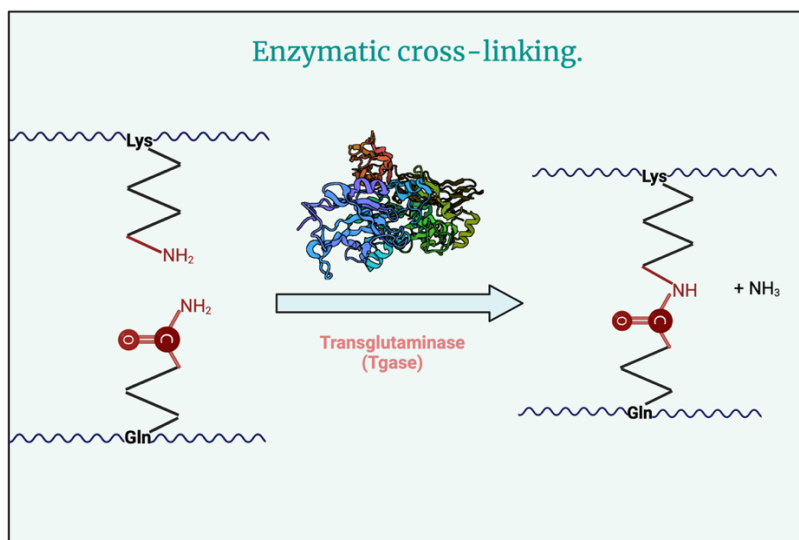


Figure 1.7: Transglutaminase mediated formation of intermolecular isopeptide bond via Enzymatic cross-linking reaction.  
Adapted from: <sup>35</sup>

### 1.3.1.3 Photo Cross-linking

Protein studies frequently employ photochemistry to initiate cross-linking reactions due to its compatibility with native cellular settings to observe intracellular interactions with great convenience for *in vivo* applications.<sup>36</sup> The conventional methods to study protein behaviour *in vitro* in a non-native cellular environment. Hence, the interactions identified using these techniques must be validated and examined in a native context.<sup>36</sup> Therefore, benzophenone derivatives like p-benzoyl-l-phenylalanine (pBpa) are popular due to their orthogonality with various biomolecules, as evident in Figure 1.8, and ease of integration with analytical methods to analyse PPIs on a proteome level.<sup>36</sup> Photo cross-linking is an efficient tool to introduce covalent bonds forming stable complexes, as seen in Figure 1.8.<sup>36</sup> The repertoire for solving specific challenges in protein studies has been greatly improved by using photo crosslinkers.

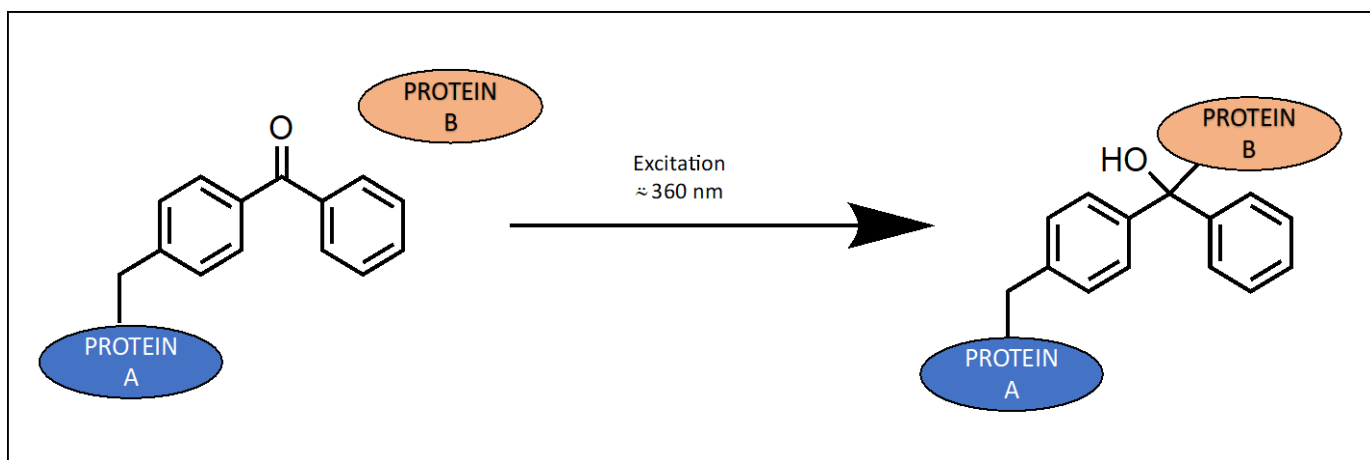


Figure 1.8: Photo-induced covalent cross-linking via benzophenone based crosslinkers for the analysis of PPIs.

Adapted from: <sup>36</sup>

### 1.3.2 Anatomy of a Cross-linker

XL-MS employs a compound called the crosslinker, which has two arms on either end and each is incorporated with two reactive groups that are separated by a spacer region of defined length.<sup>37</sup> As

stated before, amines, thiols and carboxylic acids are commonly found on the backbone of proteins, rendering these FGs potential targets for cross-linking reactions. XL-MS has become the core technology to explore protein networks in cell lysates and can be extended to structural biology.<sup>38</sup> The architecture of a cross-linker, has a crucial impact on the effectiveness of a bioconjugation reaction.<sup>39</sup> The technology depends heavily on crosslinking reagents. The anatomy of crosslinkers can be modular, sometimes comprising combinations of functional groups. While designing a cross-linker, several parameters are taken into consideration: enhancing the information density through reaction selectivity, enrich-ability for better detection, cleavability for accurate identification, and isotope-labelling for quantification.<sup>38,39</sup> Figure 1.9 shows the requisite parameters necessary for designing a successful cross-linker.

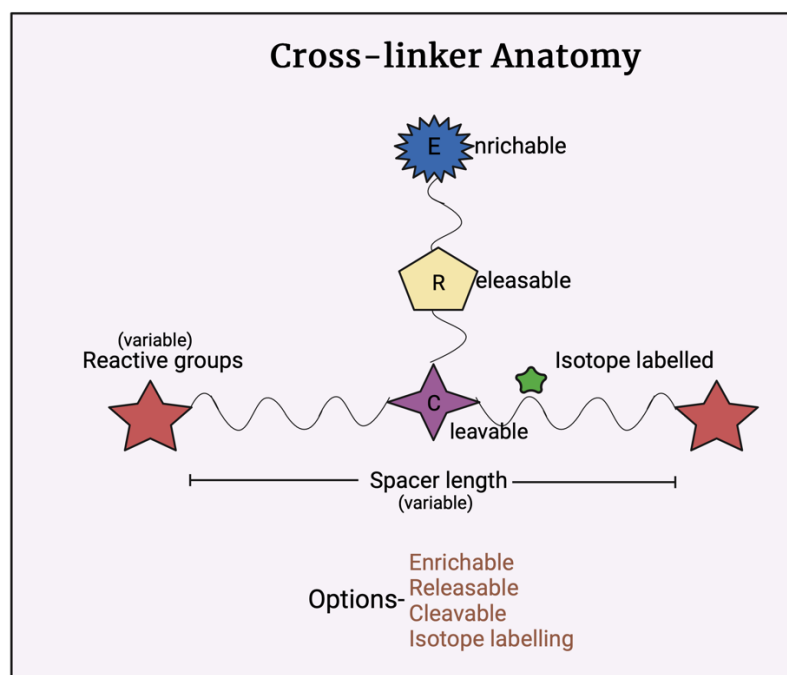


Figure 1.9: Anatomy of a cross-linker. Different reactive groups and spacer scaffolds, and multiple different cleavable, releasable and enrichable groups can be used. Adapted from:<sup>38</sup>

### 1.3.2.1 Chemical Specificity of Reactive Groups

The spatial information that can be retrieved depends on the cross-linkers reactive group and is directed towards the target amino acid residues.<sup>28</sup> Despite the complexity of the protein structure, just a few FGs can be targeted by cross-linking methods.<sup>39</sup> Table 2 lists the most common FGs found on proteins. The limited number of possible combinations of cross-linked residues generated by cross-linkers that target specific FGs limits the amount of structural data that can be obtained from the cross-linking process while avoiding non-specific interactions.<sup>28,39</sup>

Table 2: Popular cross-linker functionalities for protein conjugation (<sup>40</sup>)

<b>Reactivity Class</b>	<b>Target Functional Group on Proteins</b>	<b>Common Cross-linker functionalities</b>
1. Amine-reactive	-NH <sub>2</sub>	N-hydroxy succinimide esters (-NHS esters) Imido-ester
2. Carboxyl-to-amine reactive	-COOH	Carbodiimide (e.g., EDC)
3. Sulfhydryl-reactive	-SH	Maleimide
4. Hydroxyl-reactive	-OH	Isocyanate
5. Aldehyde reactive	-CHO	Hydrazide Alkoxyamine

#### 1.3.2.1.1 Amine Reactive-chemical Groups

The most prevalent FG on a protein surface are primary amines (-NH<sub>2</sub>), which can either be present at the N-terminus of a polypeptide chain or in side chain residues of amino acids, such as lysine, arginine and histidine.<sup>41</sup> Due to its positive charge at physiological, pH most amine groups are outwardly facing from the polypeptide chains due to the hydrophilicity, rendering them accessible to effective

cross-linking without any hydrolysis.<sup>32</sup> The cross-links can also have a positive charge at physiological pH following cross-linking reaction because imido esters are susceptible to reacting with primary amines and producing protonated amidine bonds, as shown in Figure 1.10. Therefore, imido esters can be used to explore associations and interactions between membrane proteins and can also be used to immobilize proteins on a solid support at the isoelectric pH without impairing the protein's natural structure. However, due to several shortcomings of the imido ester chemistry, it was quickly replaced with the reliable and effective -NHS ester chemistry.<sup>37</sup>

-NHS esters are the most frequent reagent used to target primary amines. Figure 1.10 illustrates the reaction where cross-linkers with activated -NHS esters react with the primary amines on peptides around pH 7.2-8.5 to produce persistent amide linkages with the loss of a free -NHS molecule which can be removed by desalting procedures.<sup>37</sup> Because lysine residues are uniformly distributed and are abundant on the surface of proteins, increasing the popularity of -NHS esters. This facilitates data analysis for calculating and quantifying lysine-lysine cross-links.<sup>41</sup> Due to their chemical reactivity, -NHS esters are also attracted to several different nucleophiles, including serine and threonine, which contain alcohol (-OH) groups. As a result, the hydrolysis of the -NHS ester forming carboxylic acid competes with the primary amine reaction. To attain the highest cross-linking efficiency, such side reactions can primarily be controlled by optimizing pH conditions and buffer concentrations.<sup>41</sup> Sulphonate groups are added to the rings of -NHS cross-linkers to increase their water solubility. These sulfo-NHS esters are used for some cell-surface cross-linking since the charged sulfo groups are cell impervious.<sup>40</sup> Figure 1.10 displays the general reaction for sulfo-NHS esters.

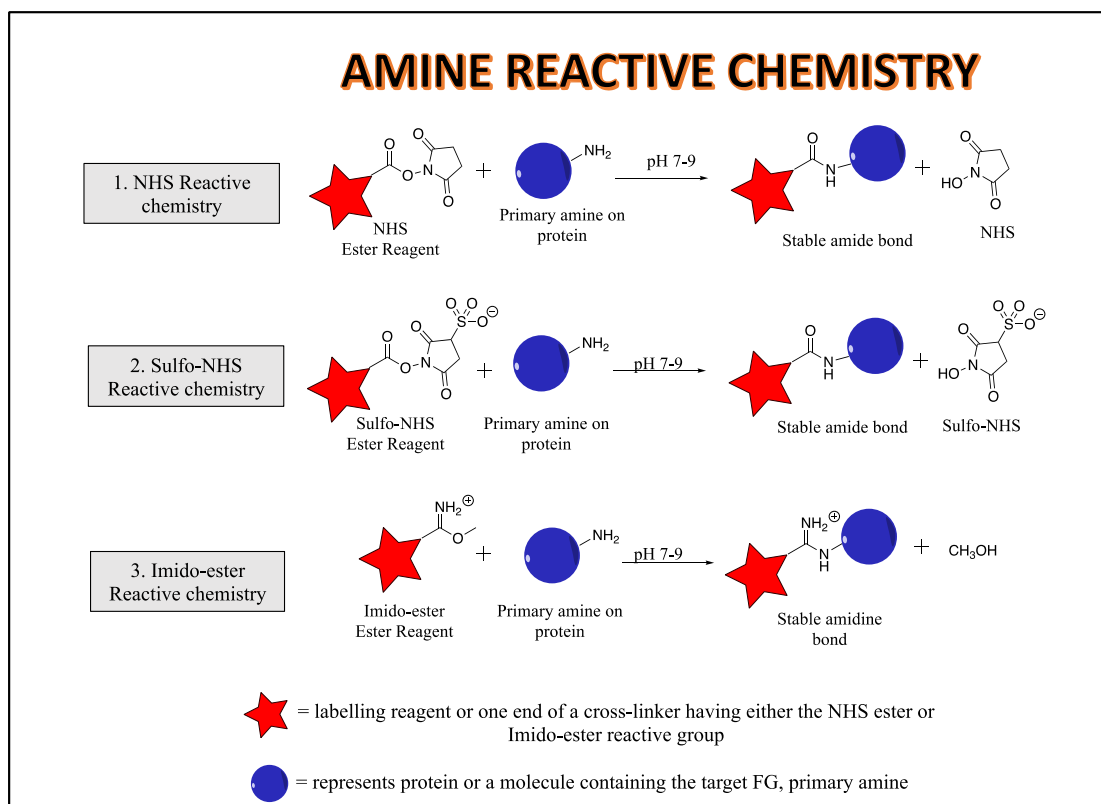


Figure 1.10: -NHS and Imido-ester cross-linking reaction scheme for chemical conjugation to a primary amine.  
Adapted from: <sup>40</sup>

### 1.3.2.1.2 Carboxylic Acid-reactive Groups

Following primary amines, carboxylic acids ( $-\text{COOH}$ ) are the other most frequent FG to be identified on the surface of proteins. Carboxylic acid found in the side chains for glutamic acid and aspartic acid are found on the C-terminus of polypeptide chains.<sup>37</sup> Carbodiimide (EDC) reaction-mediated cross-linking chemistry is the most popular process for conjugating of a primary amine with a  $\text{COOH}$  group without having EDC as a part of the final cross-linked amide bond, as illustrated in Figure 1.11.<sup>42</sup> EDC and its intermediate, iso-urea, are both soluble in aqueous solvents, resulting in hassle-free dialysis or ultrafiltration to remove these by-products. After coming into contact with EDC in a slightly acidic medium, carboxylic acids produce an intermediate called O-acylisourea. This intermediate spontaneously gets displaced by amines, forming an amide bond and an EDC by-

product known as iso-urea which is released.<sup>40</sup> In aqueous solutions, the intermediate O-acylisourea is unstable. Without an amine, the intermediate hydrolyses, the carboxylic acid regenerates, and N-unsubstituted urea is released. The actual reaction mechanism can be seen as shown in Figure 1.11. In conjunction with EDC, NHS or sulfo-NHS is often used for better coupling efficiency.<sup>42</sup>

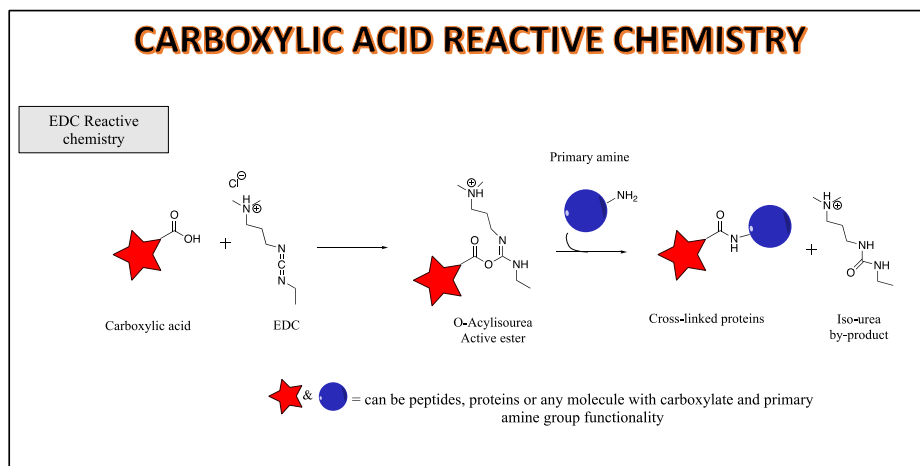


Figure 1.11: EDC (Carbodiimide) cross-linking reaction scheme for chemical conjugation to a carboxylic acid.  
Adapted from: <sup>40</sup>

### 1.3.2.1.3 Sulfhydryl-reactive Chemical Groups

Sulfhydryl groups (-SH), often known as thiols, are frequent protein conjugation and labelling targets. They typically occur as disulphide bonds (-S-S-) between polypeptide chains or as free sulfhydryl groups in cysteine side chains. Typically, the -S-S- bond must be reduced to -SH to make it sensitive to cross-linking reagents. Since thiol groups are less abundant than primary amines and carboxylic acids, several classes of cross-linkers are highly specific to them, allowing for more precise and accurate cross-linking.<sup>43</sup> Maleimides react with thiols in a slightly alkaline medium to irreversibly form stable thioether linkages as shown in Figure 1.12. At higher pH levels, maleimides favor primary amines over thiols, increasing the hydrolysis rate of maleimide. Typically, the buffers are made with materials not containing thiols to avoid competing for reactions but use some chelating agents such as

ethylenediaminetetraacetic acid (EDTA). Chelating agents are known to control the over-oxidation of thiol groups. Amino acid residues like tyrosine, and histidine, are unreactive towards maleimide.<sup>43</sup>

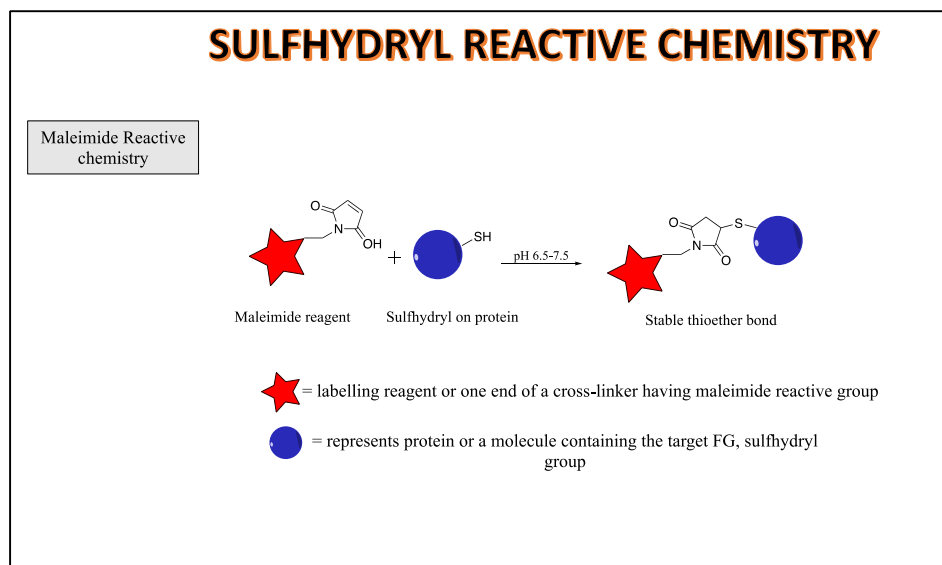


Figure 1.12: Maleimide cross-linking reaction scheme for chemical conjugation to a sulfhydryl. Adapted from: <sup>40</sup>

### 1.3.2.2 Length of the Spacer Arm

The spacer region of the cross-linker refers to the distance between the two residues once they are conjugated, as seen in Figure 1.9.<sup>39</sup> The spacer arm not only determines the workflow but also dictates the method for MS acquisition for data interpretation in terms of spatial resolution and data density.<sup>28</sup> Only the residues that are close in space can be chemically cross-linked. The arm length range can be classified as either short ( $< 10 \text{ \AA}$ ), medium ( $10 - 30 \text{ \AA}$ ), or long ( $> 30 \text{ \AA}$ ).<sup>39</sup> If the main objective is to detect proteins nearby, a longer spacer might be advantageous due to higher flexibility since it allows more residue combinations to fall into a cross-linkers range by generating more intermolecular contacts. Though, a longer arm is also more prone to non-specific binding.<sup>28</sup> Cross-linkers with short to medium spacer arms function best when intramolecular constraints are the main focus of attention.

### 1.3.2.3 Solubility of Cross-linker

Usually, organic solvents, like DMF or DMSO, are preferred for dissolving cross-linkers, but some FGs make them soluble in aqueous buffers as well.<sup>38</sup> Figure 1.13 shows cross-linkers soluble in organic solvents and aqueous buffers, respectively. Solubility indicates the importance of strong conjugation by allowing it to permeate the cell barrier. It also determines favorable contacts between hydrophobic residues within protein membranes for effective bioconjugation to occur.<sup>39</sup> The composition of the spacer arm is also crucial in influencing the solubility of cross-linking reagents, cross-linkers with hydrocarbon or polyethylene glycol (PEG) chains spacer arms perform well in intercellular cross-linking experiments due to hydrophobicity and ease of permeability within cell membranes. These cross-linkers often dissolve in organic solvents like DMSO or DMF but are insoluble in water. Charged sulfonate groups are applied to the linker, which assists in cell surface binding experiments by increasing the water solubility of the cross-linker.

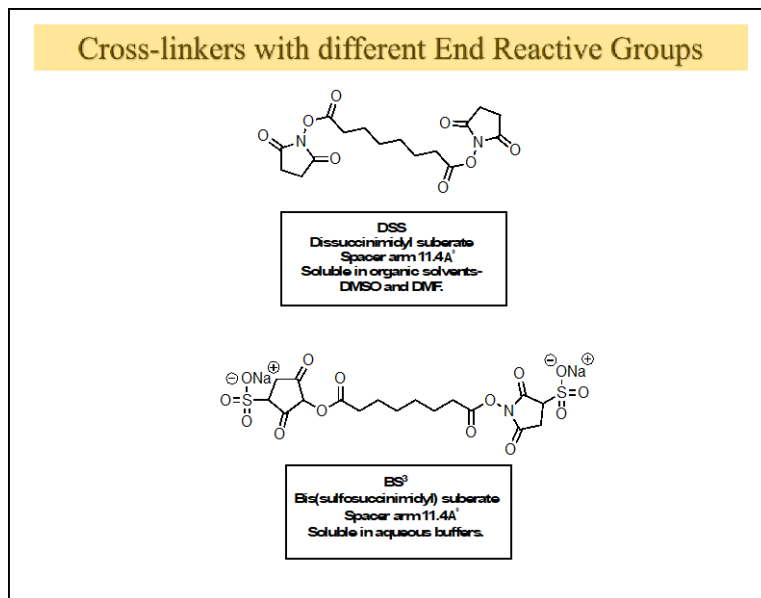


Figure 1.13: Examples of Cross-linkers having varying solubilities due to different end reactive groups.

Adapted from: <sup>44</sup>

### 1.3.2.4 Chemical Specificity

Chemical cross-linkers with two identical reactive ends are homo-bifunctional. In contrast, a cross-linker with different reactive ends is known as a hetero-bifunctionalized cross-linking reagent.<sup>35</sup> Figure 1.14 shows examples of hetero-bifunctional and homo-bifunctional cross-linkers. Both functionalities are unique in terms of data output. Homo-bifunctional cross-linkers are generally employed with simple proteins, where intramolecular cross-links are sufficient to generate information about protein organization and function. While, if a complex protein mixture has to be effectively cross-linked, hetero-bifunctional cross-linker generate intermolecular cross-links, which play a significant role in developing information by defining polymeric networks and avoiding undesirable polymerization side-reactions as seen in the case of homo-bifunctional cross-linkers.<sup>39</sup>

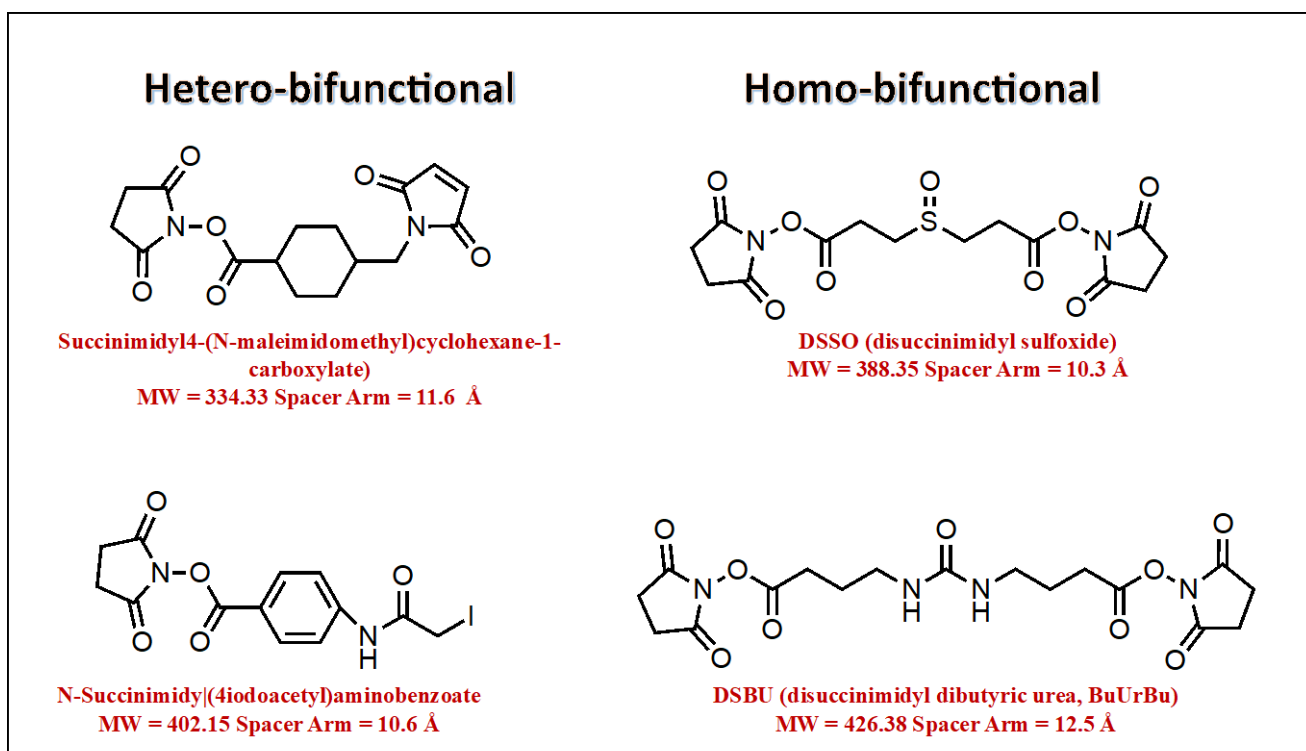


Figure 1.14: Examples of Hetero-bifunctional and Homo-bifunctional cross-linkers.  
Adapted from:<sup>43</sup>

### 1.3.2.5 Cleavable Cross-linkers

Standard non-cleavable cross-linkers and MS-cleavable cross-linkers can both accurately map PPIs. The cleavability of cross-linkers is one of the essential parameters taken into account while designing cross-linkers. However, MS-cleavable cross-linkers have several benefits. By facilitating the simple release of the cross-linked peptide in its native state, MS cleavable cross-linkers can produce distinctive ion pairs producing a unique fragmentation generated by gas phase ions.<sup>39</sup> Switching between different levels of tandem and various fragmentation techniques allows for improved identification of cross-linked peptides during MS analysis.<sup>38</sup> The spacer arm can easily integrate cleavable sites in a cross-linker reagent. The most frequently employed cleavage site is a disulphide bridge (Figure 1.15) which can easily break intact covalent bonds and recover individual components when reduced using substances like beta-mercaptoethanol or tris(2-carboxyethyl)phosphine (TCEP).

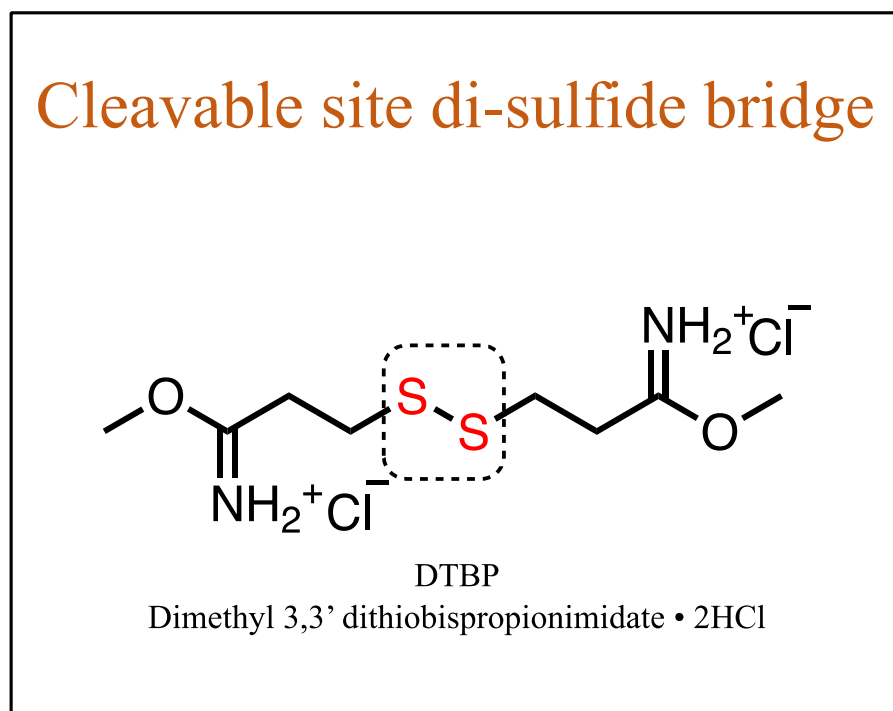


Figure 1.15: Disulphide bridge in DTBP allows for easy cleavage of a protein conjugate using common reducing agents.

### 1.3.2.6 Labelling Conditions

Labelling refers to the attachment of an enrichment handle on the surface of the cross-linkers, as illustrated in Figure 1.9, to make them susceptible towards different analytical techniques used to quantify crosslinked peptides from a mixture of mono-linked and linear peptides. The cross-linking reaction is only stable at physiological pH with mild buffers to preserve the innate chemical structure. Hence, cross-linking experiments are usually performed using near-physiologic conditions. The molar ratio of the cross-linker to protein must be determined empirically and is crucial for efficient cross-linking reactions to occur.<sup>39</sup> The molar ratio is also influenced by the abundance of FGs on the protein or peptide surface. A higher cross-linker to protein ratio is typically helpful in capturing PPIs if fewer target sites are available on the protein surface. In contrast, a lower cross-linker:protein ratio is employed if the target protein surface has abundant FGs on the surface.<sup>40</sup> The choice of the appropriate buffer system is also crucial for effective conjugation and avoiding side reactions. -NHS ester cross-linking reactions are carried out in phosphate buffers, with reaction times varying from 0.5 hours to 4 hours, with an optimal temperature being 4°C. Primary amine buffers, such as Tris-buffered saline (TBS), are incompatible because they compete for reaction. EDC crosslinking reactions must be carried out in an environment free of extraneous amines and carboxyl. The most efficient 4-morpholino-ethanesulfonic acid (MES) buffer is acidic (pH 4.5 to 5.5). However, phosphate buffers with pH values lower than 7.2 are also acceptable to the reaction chemistry. It is best to perform imido ester reactions under amine-free, alkaline (pH 10) conditions, such as using borate buffer, to ensure specificity for primary amines.<sup>40</sup>

### 1.3.3 XL-MS Work-flow

The evolution of XL-MS has been on the rise since early 2000s. Due to advancements in the development of analytical tools, cross-linking agents, and better database modelling software, XL-MS has recently gained appeal, as suggested by the booming applications with improved sensitivity in cross-linking detection, offering medium-resolution structural data.<sup>30</sup> Currently, the analysis can be performed in solution, making it possible to analyse the protein conformations in their native physiological conditions. Although, the sample heterogeneity makes it time-consuming, it does not hamper the data acquisition. XL-MS is not only limited to whole bacterial cell lysates but can also be used to explore and resolve structural motifs such as coiled coils or partially folded structures that are challenging to analyse otherwise.<sup>45</sup> A generic XL-MS workflow strategy has a colloquial premise that can be divided into four sequential steps without enrichment, a concept that will be introduced shortly.<sup>18</sup> Figure 1.16 is a visual representation of the four steps. Protein samples are chemically cross-linked, digested by proteases, and analysed by LC-MS/MS to identify cross-linked peptides, which are then subjected to docking analysis for modelling cross-linked networks and interactions by using current bioinformatics softwares.<sup>45</sup> However, it was reported that the cross-linking reaction efficiency is within the range of 1-5 %.<sup>31</sup> This is because, after protein digestion, cross-linked peptides are in sub-stoichiometric ratios compared to their linear, mono-linked counterparts.<sup>28</sup> Due to the low abundance of cross-linked peptides in the mixture, their identification remains challenging. Moreover, cross-linking reagents having -NHS esters as reactive ends further lowers the fraction of those residues which were close to be detected, as the ester is partially hydrolysed while diffusing aqueous environments and into the cell membranes to reach its target protein.<sup>41</sup> This situation particularly hinders *in vivo* applications.

# XL-MS Work-flow without Enrichment

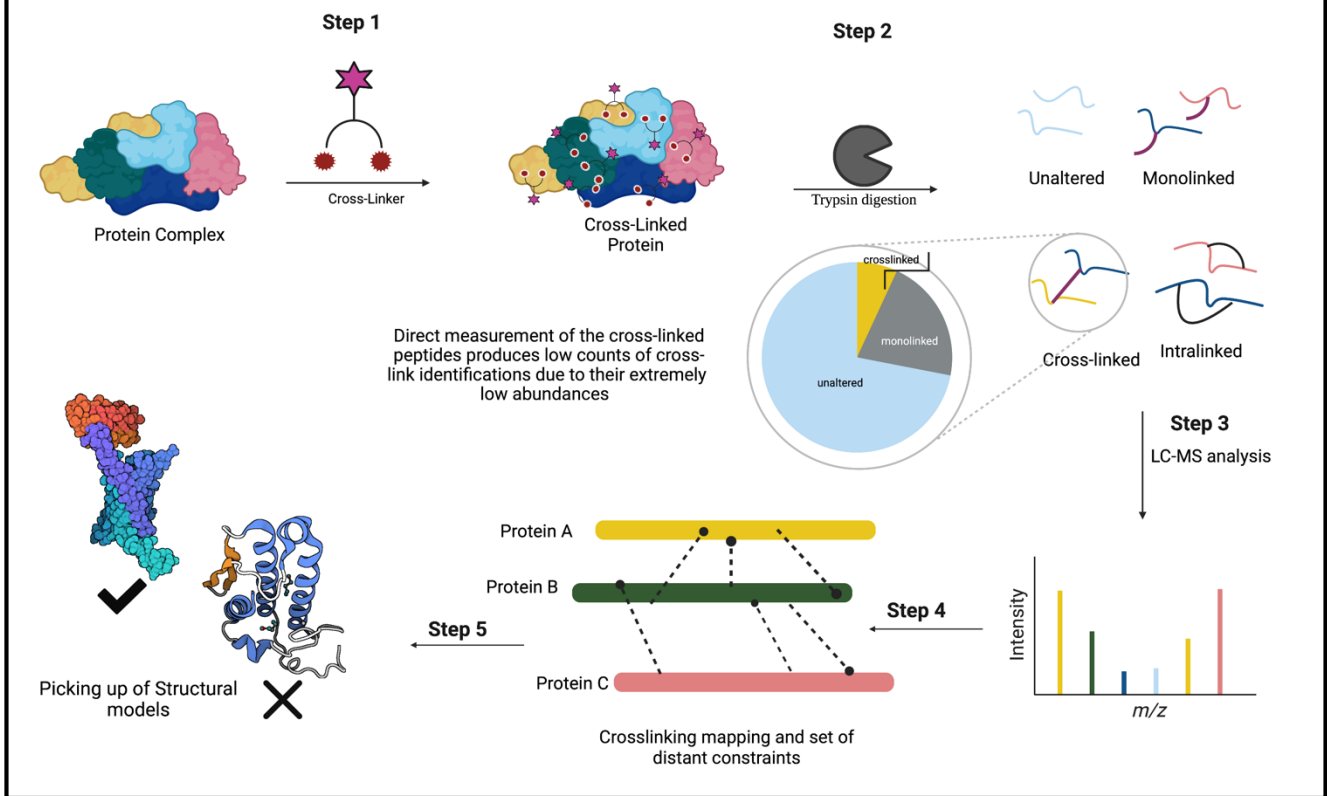


Figure 1.16: General work-flow strategy for XL-MS without enrichment: The overall workflow is divided into five major steps as displayed.

Adapted from: <sup>46</sup>

Enrichment of cross-linked peptides before LC-MS/MS measurement is one way to separate the bulk of cross-linked peptides from a mixture of linear, mono-linked, and intra-linked peptides. Attempts have been made to install an enrichment handle directly onto the crosslinker, as seen in Figure 1.9, resulting in a trifunctional molecule which can exclusively sequester cross-linked peptides from a mixture of peptides, as evident in Figure 1.18.<sup>31</sup> The most prominent enrichment method currently used is the streptavidin-biotin system. The second, less common technique, uses small molecules known as tags that are engineered to the crosslinker and are then subjected to specific analytical purification techniques

as illustrated in Figure 1.18. This process results in some bioorthogonal transformations that separate the cross-linked peptides from the peptide mixture.<sup>31</sup> Developments in these enrichment techniques have resulted in inspiring breakthroughs reaching levels of >95% enrichment specificity and sensitivity.<sup>31</sup> AP to enrich cross-linked peptides before LC-MS/MS analysis, is a common enrichment technique. Hence, for a successful XL-MS analysis the modified workflow can be divided into protein cross-linking reaction, digestion of cross-linked proteins to peptides, sample preparation followed by enrichment, data acquisition by MS analysis of cross-linked peptides, and data analysis by identification of cross-linked peptides. Each step is described in detail in the following sections.

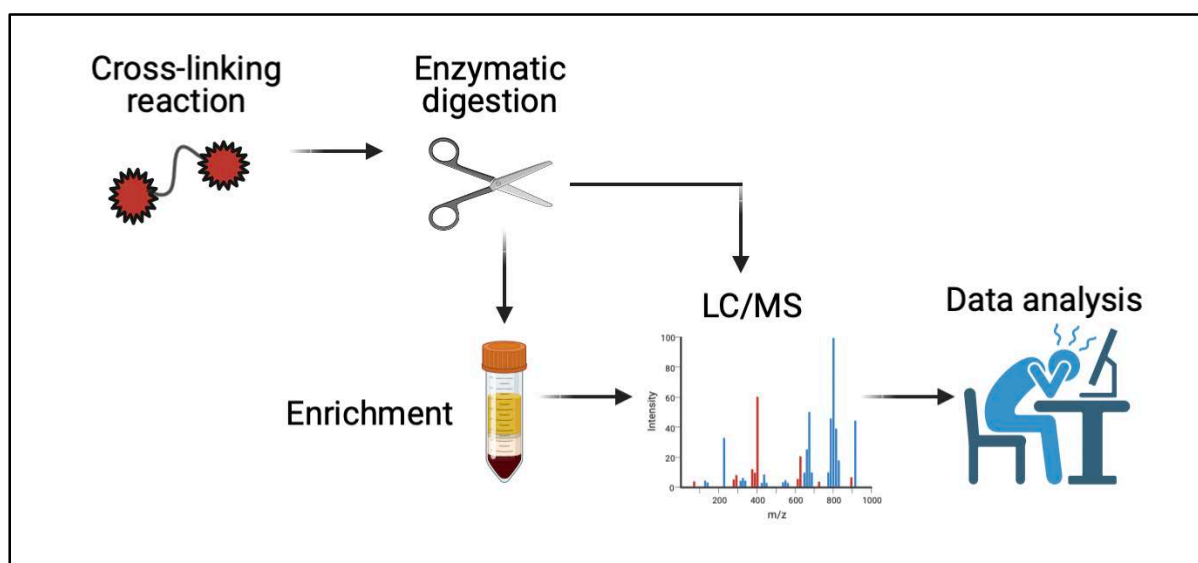


Figure 1.17: Modified XL-MS work-flow with enrichment step prior to LC-MS analysis.

### 1.3.3.1 Protein Cross-linking Reaction

Small bifunctional cross-linkers are easily used in a facile conjugation reaction to cross-link proteins in a single step to yield short, cross-linked peptide sequences, as seen in the first step of Figure 1.16. The cross-linking reaction can either be chemical Figure 1.6, enzymatic Figure 1.7 or photo cross-linking Figure 1.7. Primary amines are the easiest to bioconjugate using -NHS esters in near physiologic pH conditions, as seen in Figure 1.10. Due to the high occurrence of amino groups on proteins, they are

the most favorable and susceptible targets to undergo cross-linking with reagents containing -NHS esters as their reactive groups. Figure 1.13 shows cross-linkers with -NHS ester and sulfo-NHS ester functionality. Thiols, carboxylic acids, and other typical FGs can also be conjugated. The reaction mechanism is shown in Figure 1.12 and Figure 1.11, respectively. Only residues that are spatially proximate to one another can be cross-linked successfully. The target FG forms the basis for the design of the cross-linker to ensure specificity and avoid non-specific interactions. The distance constraints are computed from the cross-linkers spacer length, flanked by two peptide residues after a cross-linking reaction, either within the same protein or between two distinct proteins, shown in Figure 1.9.<sup>41,45</sup>

### 1.3.3.2 Digestion of Cross-linked proteins to Peptides

At the peptide level, data acquisition is convenient in a bottom-up fashion. Cross-linked proteins are first digested using enzymes before conducting an MS analysis, as shown in step two of Figure 1.16. Proteins are generated rapidly in Femto- to attomole levels through enzyme breakdown, which simplifies analysis because the mass of protein complexes is potentially limitless, making their analysis difficult compared to peptides.<sup>29</sup> Following cross-linking, the proteins are subsequently subjected to alkylation, reduction, and digestion by enzyme proteases, typically trypsin, into short peptide sequences.<sup>41</sup> Trypsin is the most common protease used for this purpose and automatically cleaves the peptide at the location of basic amino acid residues like lysine or arginine to produce peptide sequences with an amino-terminal group and bears an amino group of either lysine or the arginine residue.<sup>29,41</sup> However, using amine-reactive cross-linkers can lead to mis-cleavage, resulting in non-homogeneous long peptides after digestion.<sup>41</sup> Standard proteomics is competent in quantifying and identifying linear peptides with specific modifications. Given the various species that can be identified after digesting a cross-linked peptide, peptide cross-linking is undoubtedly more informative than

protein cross-linking. There are several different types of cross-linked peptides that can be detected, including mono-linked, intra-linked (internally bridged), and higher-order cross-linked peptides, shown in step two of Figure 1.16. Cross-linked proteins complicate regular analysis by adding to sample heterogeneity.<sup>45</sup> However, cross-linking a protein allows for mapping various structurally constrained peptides and protein complexes with intramolecular linkages in addition to intermolecular contacts. A significant amount of information can be obtained from all the types of peptides, but cross-linked peptides are the focus since they provide information on long-distance networks, whereas mono-linked and loop-linked peptides help predict the local structural motifs like alpha helices but not necessarily on high-order cross-linking information.<sup>45</sup> The downside is that linear and mono-linked peptides are in high abundance compared to cross-linked peptides, as observed in Figure 1.16. Hence the mixture of peptides is usually subjected to various enrichment techniques before LC/MS analysis to yield a sample with perfect cross-linked peptides.<sup>41</sup>

### 1.3.3.3 Sample Preparation and Enrichment

As already mentioned, cross-linking reaction efficiency is typically low, ranging from 1-5%, with overwhelming numbers of linear peptides in contrast to cross-linked peptides, which are relatively few.<sup>31</sup> Multiple cross-links in a single product and some incomplete cross-links exacerbate this problem by limiting the signals due to cross-linked peptides.<sup>45</sup> There are a couple of chromatographic techniques that demand the cross-linked peptides to be subjected to a series of fractionating steps. However, using these techniques still yields a sample with a high background of linear peptides.<sup>31</sup> Chromatographic techniques, like size-exclusion chromatography (SEC) or strong-cation exchange chromatography (SCX), rely on physical properties, such as size or charge, to separate molecules. They are not explicitly designed to provide specific information restricted only to cross-linked peptides.<sup>47</sup> Recently, affinity-

based enrichment strategies have been on the rise due to their ease of sample preparation and the fact that these techniques preserve the native chemical structure of the peptides after enrichment. These methods involve integrating an enrichment handle onto the cross-linkers surface, producing a reagent adept at isolating cross-linked peptides from a peptide mixture.

#### 1.3.3.3.1 Affinity Purification (AP)

AP is a technique highly selective towards small enrichable tags that are engineered on the cross-linker and makes use of the differences in binding interactions between the tagged peptides and other un-tagged peptides with the ligand packed onto the column as solid support for effective separation.<sup>47</sup> AP discreetly targets and sequesters cross-linked peptides from a mixture of linear and mono-linked peptides. Figure 1.18 describes the mechanism of a standard AP strategy, which involves the loading of solid support on the column, along with an immobilized ligand, that is highly specific towards the tag.<sup>41</sup> As affinity matrices, various solid supports, including agarose, cellulose or porous gel supports, or resin beads, can also be used. Due to the porosity and robust packing onto the column, various types of resin beads are typically favored over other solid supports. Different resins are available, each with unique chemical and physical characteristics, including loading capacity, bead size, and coupling capacity. They are applied onto the column as a wet slurry, shown in step one of Figure 1.18. Hence, after a complex sample is passed onto the column, only the tagged cross-linked peptides with a specific binding affinity towards the ligand remain on the column. The other peptides, such as linear, mono-linked, and other unwanted contaminants from the matrix, are washed away sequentially using stringent washing buffers, as seen in step three of Figure 1.18. The interaction between the solid support and the peptide of interest is modified after the impurities are eliminated, either by using a mixture of buffers at a certain pH or by selective oxidizing or reducing agents. The peptides are

recovered in the native original state. The elution step, described in Figure 1.18, can be optimized based on the choice of cross-linker and other reagents.<sup>45,41</sup>

The drawback of the AP approaches is that non-specific interactions and other side reactions cannot always be controlled. As mono-links are significantly more abundant than cross-links, they can interfere with routine analysis.<sup>41</sup> In summary, it can be stated that no technique can separate pure cross-linked peptides from a complex sample. However, if parameters like sample recovery, sample preparation, and others are tuned to minimize loss, combining orthogonal enrichment techniques can yield better results.

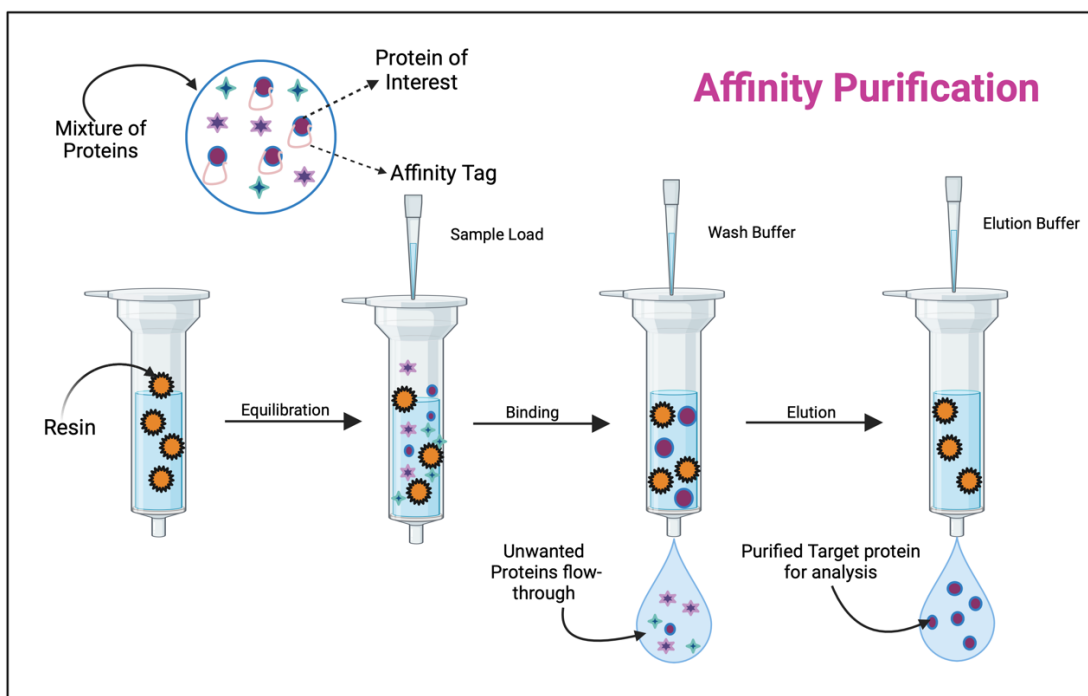


Figure 1.18: Steps involved in the work-flow of Affinity Purification involves: equilibration, binding washing and elution of protein of interest from the solid support. Adapted from:<sup>48</sup>

#### 1.3.3.4 Data Acquisition by MS Analysis of Cross-linked Peptides

The scope of data acquisition has been broadened in terms of better signal resolution and increased sensitivity over time with the development of new mass spectrometers, and other innovative reagents and software tools.<sup>45</sup> Some ground-breaking achievements have been delivered in the field of XL-MS, allowing system-wide investigation of protein topologies and networks with LC coupled to MS.<sup>29</sup> MS data acquisition is the main component of the XL-MS workflow, as shown in step four of Figure 1.16, where parameters such as sample, cross-linking reagents, and spectrometer determine the type of settings used for the acquisition.<sup>41</sup> Typically, electrospray ionization (ESI) is used to ionize the sample. Then the resulting ions are subjected to fragmentation tools to generate fragmentation spectra for further identification through database searches with online resources. Therefore, MS acquisition is essentially a two-stage procedure, where the first step entails the identification of cross-linked peptides in the form of fragments. Then, using various bioinformatics tools, this data is compared to protein databases readily available online.<sup>29</sup> The approximate number of sites or linkages between the cross-linked peptides can be extrapolated from fragmentation spectra, which also provides insights into the precise cross-links. High-resolution mass spectrometers are becoming more popular since the quality of the spectrum also plays a crucial role in acquiring these fragments.<sup>45</sup> Collision-induced dissociation (CID), one of the fragmentation techniques, enables fragments to break primarily along the backbone at the peptide bond, simplifying analysis. The bottom-up approach, as shown in Figure 1.19, can then subject the mass spectra to software tools to automatically detect cross-links.<sup>29,45</sup> Top-down approaches are less common since they work better with small proteins, as seen in Figure 1.19. The primary mechanism of CID is backbone cleavage at the peptide bond, which typically results in the formation of the two main fragment types, b-ions and y-ions,<sup>49</sup> as seen in Figure 1.20.

Another alternative approach to CID is electron transfer dissociation (ETD), which fragments ions at a different site than the peptide backbone as in CID and yields c- and z-ions as a result.<sup>49</sup> Figure

1.20 shows the standard nomenclature for sequence ions in the mass spectra of peptides. To perform this analysis, MS2-MS3 strategies are frequently used. In these techniques, the ions are first fragmented at lower collision energies, at the MS2 level, to discover the masses of the peptides. Once the masses are known and confirmed, they are subjected to fragmentation at higher energies, at the MS3 level, to assess the peptide sequence.<sup>41</sup> These statistics are crucial as they reveal the number of peptide residues in the spectra and the charge of each ion.<sup>45</sup> These fragment ion numbers determine the charge of the ion by labelling the precise isotopic peaks, and, the mass of the fragment can be computed. This is because LC-MS analysis measures the charge-to-mass ratio ( $m/z$ ). To evaluate the peptide sequences, rigorous post-fragmentation analysis is required because the spectra will inevitably contain peaks from fragments from both cross-linked peptides.<sup>45</sup> CID and ETD's key benefits enable cross-link mirroring in the native chemical environment, which is beneficial for membrane proteins.<sup>29</sup>

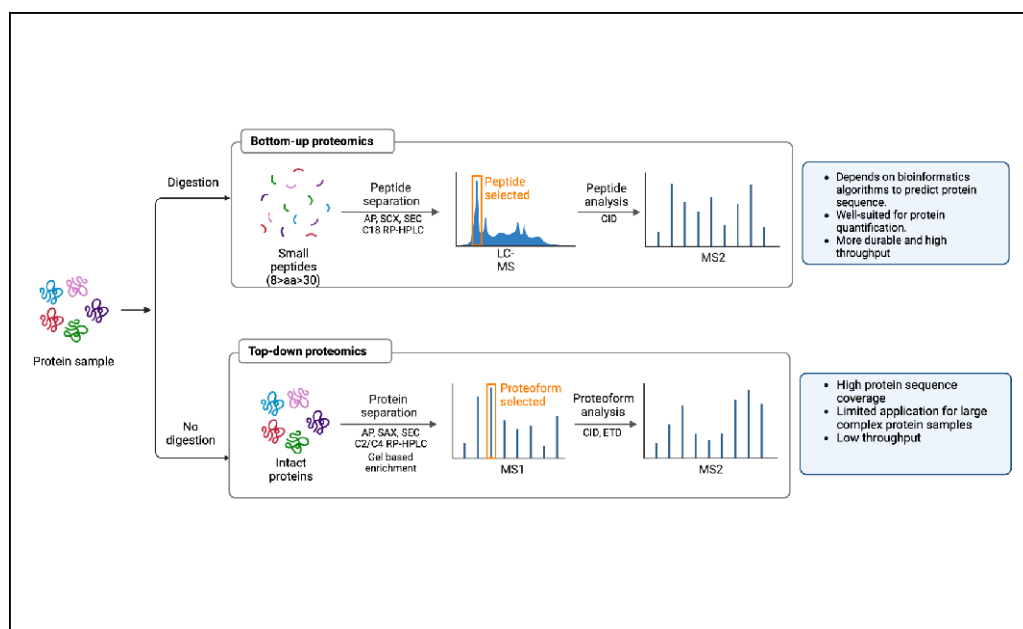


Figure 1.19: Typical steps followed in bottom-up and top-down proteomics. Adapted from: “Types of proteomics workflows”, by BioRender.com (2022) SCX: strong cation exchange; SAX: strong anion exchange; RP-HPLC: reversed-phase high performance liquid chromatography; CID: collisionally-activated dissociation; ETD: electron- transfer dissociation.

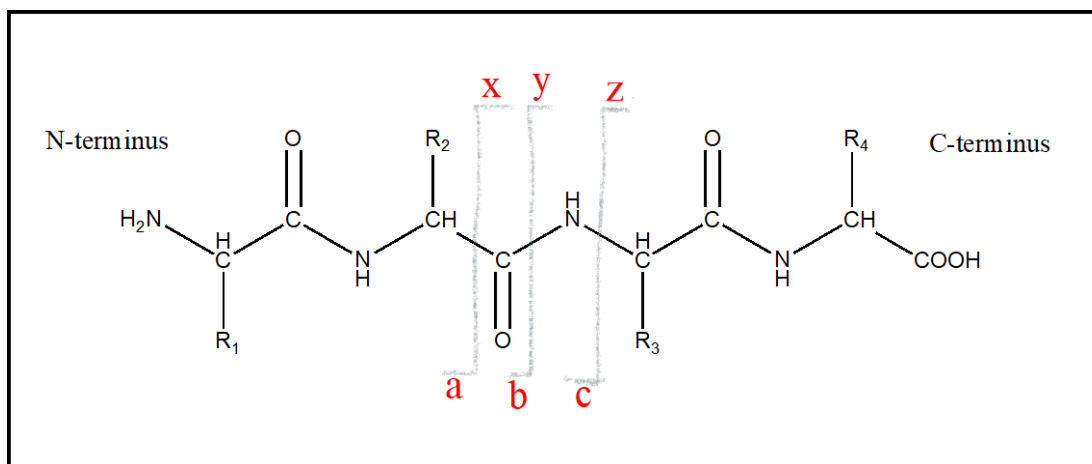


Figure 1.20: Annotation of peptide fragmentation ions b-and y-, a-and x-, as well as c-and z-, ions generated most commonly upon collision-induced dissociation and electron capture dissociation of peptide ions.

Adapted from: <sup>50</sup>

### 1.3.3.5 Data Analysis by Identification of Cross-linked Peptides

With the increasing multiformity of protein samples, there is a need to develop automated software to analyze the precise crosslinks to accurately match the fragmentation patterns. Hence, continuous development in machine learning and bioinformatics tools is a must.<sup>29</sup> Data analysis of the fragmentation patterns and the results validation is subsequently the most crucial of the workflow for XL-MS.<sup>41</sup> Several steps are involved in data processing after obtaining MS fragmentation patterns. Using online, free-accessible tools, the raw data files are first converted to software-compatible file formats like mzML. After receiving the appropriate file format, it is processed using conventional algorithms <sup>41</sup>.

Linear peptides and cross-linked peptides are subject to slightly different analysis procedures. In the case of linear peptides, the peptide mass selects the appropriate candidates from the online database, and the fragmentation pattern validates which of all the previously chosen candidates is the best match.<sup>45</sup> However, unlike linear peptides, crosslinked peptide analysis is more complex.<sup>45</sup> Before

determining the actual sequence in the case of cross-linked peptides, all probably digested peptides must be considered. The number of combinations can be calculated using the formula  $(n^2 + n)/2$ , where ‘n’ is the # of peptides from possible cross-links, considering all potential pairs. MeroX and XlinkX are popular programs used to analyze cross-linked peptides and they are free and provide an intuitive graphical user interface for beginners.<sup>41</sup>

xQuest is also a popular software for identifying cross-links from an extensive database in the case of isotopically labelled cross-linked peptides apart from MeroX and XlinkX.<sup>29</sup> Currently, about 20 algorithms are available to carry out this analysis and the most commonly used are described in Table 3.<sup>41</sup> While working with intricate samples that contain non-cleavable cross-linkers, the issues caused by false identification are exacerbated. Plink2 adopts a stepwise fragmentation index to identify alpha-helical residues first and then moves on to beta peptides, compared against a known peptide index for confirmation.<sup>41</sup> Other than pLink2, XlinkX also allows for MS2-MS3 compatibility and supports functions to export visual data tools. Cross-linked peptides frequently result in random matches and false positives since there are more pairs to evaluate in the database than for linear peptides.<sup>45</sup> However, rigorous and extensive analysis can reduce these false positives by excluding the matches that do not match the parent protein sequence and applying filters to further remove the unwanted sequences.<sup>45</sup> Additional software is also available to quantify potential cross-links, identify various conformational changes in complex mixtures, and more. Estimating the false-discovery rate (FDR) from algorithms should be considered.<sup>41</sup> Data visualization of the predicted data from the algorithms can be carried out by search engines like protein data bank (PDB) and others. Finally, for determining the confidence of a match, data processing from multiple algorithms might be beneficial to obtain complementary validation of results for increasing the confidence rate.

Table 3: List of Common Algorithms for Cross-Link Data Analysis <sup>(41)</sup>

<b>Name</b>	<b>MS-cleavable XL</b>	<b>Proteome wide</b>	<b>Quantitative XL data</b>
1. MassAI	NO	NO	NO
2. MeroX	YES	YES	NO
3. pLink2	NO	YES	YES
4. XlinkX	YES	YES	NO
5. xQuest	NO	NO	YES

#### 1.3.4 Applications of XL-MS

Undoubtedly, XL-MS has achieved notable milestones in structural biology quickly. Recently, XL-MS has gained recognition as a robust technology for examining protein topologies, protein assemblies, different protein conformations, and determining protein tertiary structure. The rise in the number of publications demonstrates the exponential growth of the field. The major break-through cannot solely be attributed to the development of improved high-resolution mass spectrometers but also the advancements made in the XL-MS strategies, including the discovery of new algorithms and other bioinformatics tools and development of novel reagents, which have proven to be successful for the XL-MS approach.<sup>29</sup> Because of the results, XL-MS can also be applied in molecular biology and the emerging fields of interactomics in an integrative approach with the combination of other low to medium-resolution structural elucidation techniques. Some typical applications of XL-MS include the investigation of protein topologies, high-density cross-linking mass spectrometry for the determination of 3D protein structures and quantitative XL-MS approach for benchmarking studies, which are described in detail in the following sections.

#### 1.3.4.1 Investigation of Protein Topologies

XL-MS has made significant contributions in elucidating protein assembly and its related sub-domains. It does not require highly pure samples, and efficient analysis can be carried out with less amount of samples, which makes it more widespread. In contrast to other structure-elucidation methodologies, XL-MS performs cross-linking close to physiological pH conditions, maintaining the integrity of native chemical connections.<sup>28</sup> The combination of XL-MS with different modelling approaches, in an integrative fashion, has proved to be quite beneficial in structural biology to complement the missing information obtained from other experimental procedures. To analyse protein structural domains and assess surface-accessible areas, modelling densities for cryo-EM maps of the distinct subunits is essential.<sup>28</sup> For instance, XL-MS successfully addressed the correct arrangement of a 1-MDa hetero-oligomer, TRiC/CCT chaperonin consisting of 16 subunits initially mis-predicted by other experimental methods.<sup>51</sup> In addition, XL-MS has significant advantages when combined with powerful techniques, such high-resolution electron microscopy, since it can be used to establish the position and organization of subunits in large protein assemblies, such as nuclear-pore complexes, yeast spliceosomes, and others.<sup>52</sup> The delineation between inter- and intra-molecular cross-links is challenging in homo-multimeric complexes, which involve complex protein assembly and identical subunits. A practical method, in this case, is mixed-isotope cross-linking, which demands mixing unlabelled <sup>14</sup>N protein and <sup>15</sup>N isotopes in equimolar ratios to produce cross-linked peptides with either mixed labels, no labels, or complete isotope labelling. The presence of intermolecular interactions is confirmed by the cross-links with full isotope labelling.<sup>29</sup>

Last but not least, XL-MS can support the validation of *in vivo* cross-links based on knowledge of complicated sub-units obtained from *in vitro* studies. Targeted MS approaches can complement *in situ* assays that identify cross-links by confirming their presence. Low-density cross-links and high-

resolution techniques were used to get the information in the above-mentioned applications. Albeit this may have needed to be more accurate in the case of dynamic protein assemblies which can adopt a variety of conformations.<sup>28</sup>

#### 1.3.4.2 Modelling Protein Structures using High-density Cross-linking Mass Spectrometry (HD-XLMS).

The type of information generated through XL-MS mostly depends upon the cross-linker used. Mapping of restraint-driven protein complexes, in which the cross-links are dispersed across a large protein surface, places limitations on potential subunit organizational studies since it is difficult to analyse information that is not restricted to any subdomain.<sup>29</sup> Therefore, employing cross-linkers with chemical groups tailored to a particular reactive end can narrow the data to a specific sub-unit or domain, providing more accurate models. In these circumstances, hetero-bifunctional cross-linkers typically have UV photoactivable groups, like azido or diazirine functionality on one reactive end and, -NHS ester or sulfo-NHS ester groups on the other reactive end, are preferable.<sup>28</sup> The -NHS ester side is selective towards basic amino acid residues such as lysine and threonine.

In contrast, the opposite side can anchor to any residue after photoactivation, thus making data analysis simple. When using this kind of cross-linker, the cross-linking is typically referred to as high-density cross-linking (HD-XLMS). However, this type of cross-linkers is only efficient when working with a single protein and not a large protein assembly must be addressed.<sup>28</sup> The type of information generated from HD-XLMS predicts much information about the native folds by confining the overall structural domains, loops and other residues in a proximate space. One of the earliest studies using a sulfo-NHS-diazirine (sulfosuccinimidyl 4,4'-azipentanoate) (sulfo-SDA) that was probed with a 66-kDa protein, human serum albumin (HSA), was capable of effectively yielding 1,495 cross-links that were

specific enough to model the HSA structure, which was further validated by the existing crystal structure of HSA.<sup>53</sup> These constraints are crucial in mapping the tertiary structure of the intended protein. HD-XLMS is currently a part of the community for the Critical Assessment of Protein Structure Prediction (CASP), which strives to assess the reliability of existing modelling tools and other algorithm structure predictions. The first occurrences of HD-XLMS were in CASP11 and CASP12. It served as a platform for further refining already-existing approaches in the structural proteomics toolbox and establishing the infrastructure for designing and resolving de novo structures, a niche area of study.<sup>28</sup>

#### 1.3.4.3 Quantitative XL-MS (Q-XLMS) approach for Benchmarking Studies

Proteins exhibit a great deal of conformational heterogeneity by existing as an ensemble of different conformations with dynamic and flexible regions in solution. By conducting comparative experimental studies under other conditions, XL-MS can assist in quantifying protein dynamics by establishing cross-links that cannot be explained by the known crystal structures.<sup>28</sup> Quantitative XL-MS (Q-XLMS) has recently developed as a benchmarking tool to avoid discrepancies by generating high-confidence cross-links in solutions for alternative conformations which cannot be validated by crystal structures.<sup>29</sup> However, such cross-linking data can reflect a multitude of conformations and therefore, algorithms to process such quantitative information in an automated fashion is crucial.<sup>29</sup> A major breakthrough was achieved when Q-XLMS could provide the conformational topologies of G Protein-Coupled Receptor Kinase 5 (GRK5) in interaction with the  $\beta_2$ -adrenergic receptor by accurately measuring long-distance contacts to differentiate between its multiple states.<sup>28,54</sup> Today, several approaches that facilitate comparison studies can be used to study conformational changes caused by cross-linking. Techniques involving isotope-labelled cross-linkers, label-free data-dependent acquisition, and others can generate quantitative information regarding the abundance of cross-links

allowing for direct comparison among variable conformational states.<sup>28,29</sup> The isotope-labelled technique, as described in the previous section, uses a combination of the same cross-linker, substituting one with its corresponding stable heavy isotopes. After undergoing enzymatic digestion, the protein samples are next cross-linked with the cross-linkers individually before being mixed in an equimolar ratio; MS signals can be used to quantify the intensity of the cross-links. However, due to a lack of automated bio-informatics tools that can adequately extrapolate quantitative data for both MS cleavable and MS non-cleavable cross-linkers, the cross-links must be manually validated.<sup>29</sup> Applications to automate this data are continuously being designed; as a result, programs like xTract and MaxQuant can now be integrated to conduct these evaluations.<sup>28</sup>

Q-XLMS was also used to study a large protein assembly using an MS-cleavable cross-linker, which involved using isotope-labelled protein interaction reporters (PIRs) to cross-link two conformations of *E. coli in vivo*. Based on the total number of cross-links, a 78% quantification rate was achieved.<sup>29,55</sup> The analysis of complex conformational rearrangements seen in cullin-Ring ubiquitin ligases has also been facilitated by Q-XLMS.<sup>28,56</sup> Stable isotope-labelling using amino acids in cell culture (SILAC) is another isotope-labelled strategy that recently gained prominence when it was successfully used to analyse the fluctuation of PPIs of drug-sensitive and drug-resistant cancer cells.<sup>29,57</sup> The fundamentals of an isotope-labelling method are also the foundation of SILAC where the only difference is that the amino acids are isotopically labelled and not cross-linked. The most common amino acid, in this case, is lysine due to its higher abundance. Novel algorithmic software that work with well-planned workflows is still under development to increase the accuracy and dependency of Q-XLMS.<sup>28,29</sup>

#### 1.4 Boronic acid (BA) as a Bio-orthogonal handle

In the field of conjugation chemistry, there is a considerable demand for biocompatible conjugates that can be applied in the field of medicine not only study but also to determine the ability of different biomolecules to develop innovative therapies.<sup>58</sup> Bioconjugation requires the incorporation of payloads that have specific functionality on biomolecules, such as proteins, nucleic acids, and even vitamins, to yield multi-functional constructs. Figure 1.21 is a general representation of a simple bioconjugation reaction. These bioconjugates are utilised in the design of novel therapeutics. These ground-breaking discoveries have spurred the search for additional biocompatible reactions.<sup>59</sup> Despite the remarkable progress, the present conjugation chemistries have some inefficiencies, such as slow reaction rates and off-target labelling.<sup>58</sup> The majority of bioconjugates were designed to produce stable frameworks tolerant to harsh physiological conditions. BAs, are a versatile class of reagents that can be used as building blocks to generate new C-C bonds with ease.<sup>59</sup> These simple reactants can undergo remarkable reversible conjugation chemistries, yielding biocompatible conjugates that have biological applications that can be investigated to find new therapies for biological uses.<sup>58,59</sup> BAs ability to form reversible covalent linkages with diols (Figure 1.22) elevates its applicability in the health industry.<sup>60</sup> They are currently employed as carbohydrate sensors and applicable in the design of enzyme inhibitors and can behave as Lewis acid catalysts.<sup>58,59,61</sup> Boron neutron capture therapy (BNCT) is one of the clinical applications for BA reagents in medicinal chemistry.<sup>62</sup> Based on the selective accumulation of molecules containing  $^{10}\text{B}$  into malignant cells, BNCT is a non-invasive method for eliminating cancer cells. Velcade, a cancer treatment drug, was the first BA-containing drug to hit the market.<sup>60</sup>

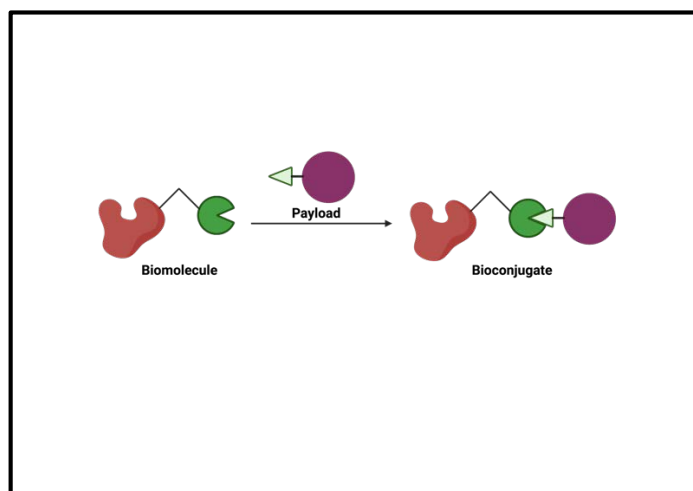


Figure 1.21: General scheme for a bioconjugation reaction using a payload forming a bioconjugate with enhanced physical and chemical properties.  
Adapted from: <sup>63</sup>

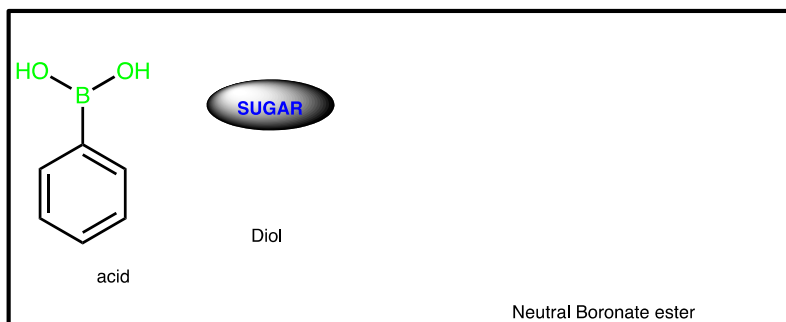


Figure 1.22: Application of BAs as carbohydrate sensors, due to their complexation with polyols to form neutral boronate esters.  
Adapted from: <sup>64</sup>

BAs gained recognition as a prime class of intermediates for several C-C bond-forming chemical processes after the Nobel prize-winning Suzuki Miyaura Borylation reaction. They can generate nucleophiles *in situ*, which are precursors for cross-coupling and substitution reactions. Common oxygen-containing boron compounds are shown in Figure 1.23. With the use of BA and other organoboranes as synthetic intermediates over the past decade, significant progress has been made in various overlooked sectors, such as materials science and catalysis. Due to their unique chemical and

physical characteristics, BAs are used to synthesize therapeutically beneficial bioconjugates. The following sections describe the chemical and physical properties of BAs in detail.

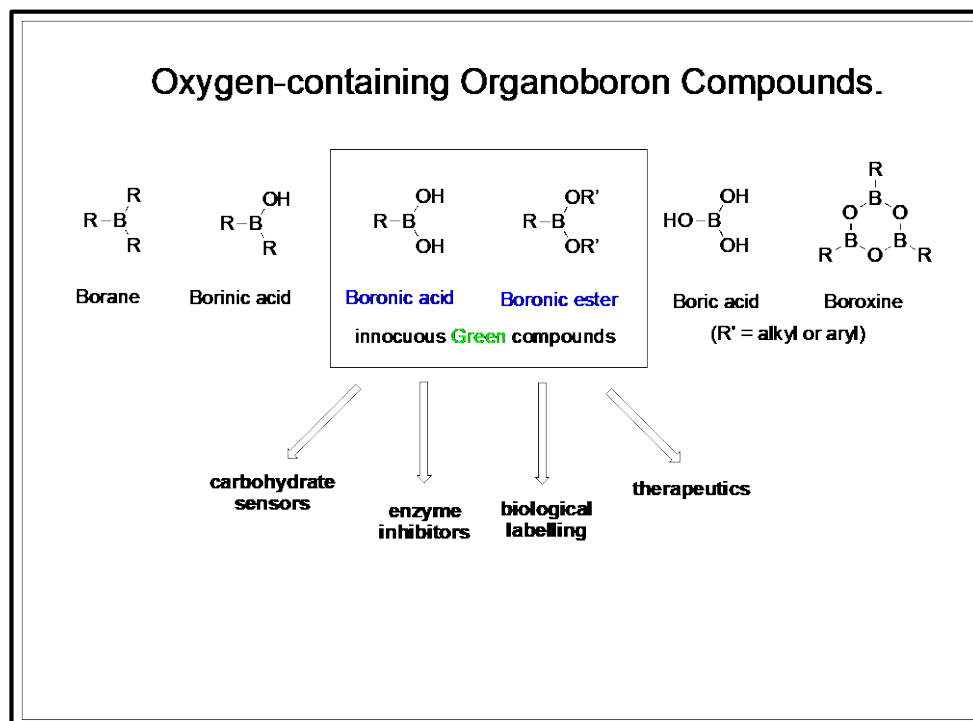


Figure 1.23: Common oxygen-containing Organoboron compounds and their related applications. Adapted from: <sup>60</sup>

#### 1.4.1 Chemical and Physical properties of BAs

To fully appreciate its scope of applicability, one must have a better understanding of the factors that influence the stability of BAs. The next section describes a few physical and chemical properties of BAs and their derivatives.

##### 1.4.1.1 Structure and Bonding

BAs are trivalent organic compounds containing carbon as one substituent and two hydroxyl groups attached to the boron directly, making it  $sp^2$  hybridized. The boron atom is deficient by two electrons, leaving the p-orbital vacant, as evident in Figure 1.24. Due to the vacant p-orbitals on the

trivalent boron atom, the nature of the bonding surrounding the boron atom in both boronic and borinic acids is distinct. The nature of the boron-oxygen bonding is heavily influenced by donor-acceptor interactions involving vacant p-orbitals and the lone-pair orbitals on the hydroxyl oxygen atoms. The low-energy p-orbital, is orthogonal to the other three orbitals; thus, the molecule aligns itself in a trigonal planar configuration. BAs take on a configuration that is isoelectronic to that of a neutral  $sp^3$ -hybridized carbon after coordinating with basic molecules, as seen in Figure 1.25.<sup>59,60</sup>

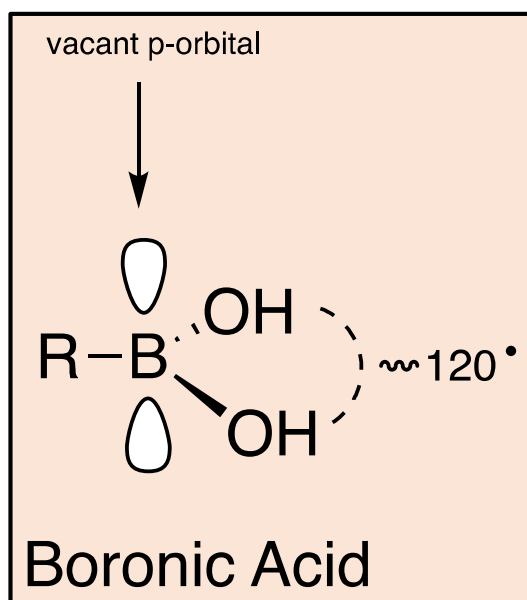


Figure 1.24: Structure and bonding of BAs.  
Adapted from: <sup>65</sup>

#### 1.4.1.2 Storage and handling

BAs occur as crystalline solids having long shelf lives in an inert atmosphere at ambient temperatures. Due to their bench-top stability, they do not require special precautions, unlike the other boron homologues like alkyl-substituted BAs that tend to oxidize in the air. BAs also display excellent thermal stability, even at high temperatures, and can be stored under a high vacuum to avoid auto-oxidation altogether.<sup>59</sup> However, at such elevated temperatures, BAs spontaneously convert to their

anhydrides; while they show similar reactivity as BAs for coupling reactions, they are more susceptible to undergoing auto-oxidation. BAs are stored under slight moisture conditions for long-term storage purposes.

#### 1.4.1.3 Acidic Character

The vacant p-orbital contributes to the slightly acidic nature of BAs making them behave as mild Lewis acids. They reversibly coordinate with many nucleophiles. Lewis bases give them a tetrahedral structure, as seen in Figure 1.25 and Figure 1.26. The uncharged trigonal planar structure is transformed into an anionic tetravalent borate species by the vacant p-orbitals, which attract oxygen and nitrogen, as observed in Figure 1.26, due to its acidity. Due to its Lewis acidic nature and hydrogen bond donor capability, BAs display polar character. The polar nature can be mitigated by attaching hydrophobic tails to the hydroxyl groups, making them amphipathic. If amphiphilic, boronic esters have partial solubility issues in water and organic solvents, complicating isolation. BAs have increased solubility in water in their ionized form when in high-pH aqueous solutions. They can be extracted more readily into organic solvents from, aqueous solutions of low pH.<sup>60</sup> The neutral trivalent specie is in equilibrium with its anionic tetrahedral form as seen in Figure 1.25

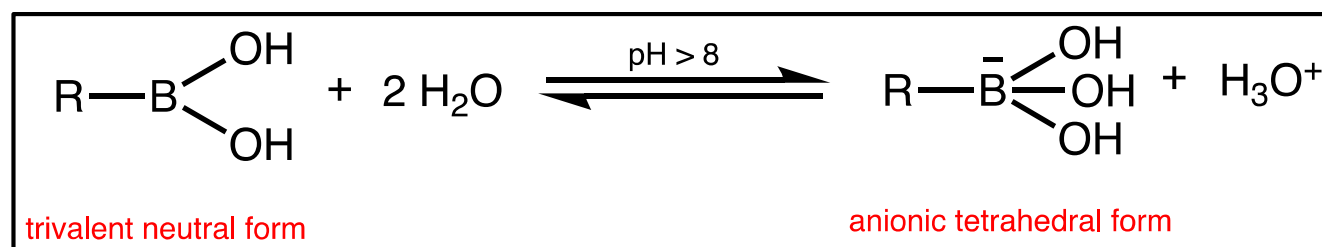


Figure 1.25: Ionization of BA in water, when pH > pKa forms tetrahedral anionic ester. Adapted from: <sup>60</sup>

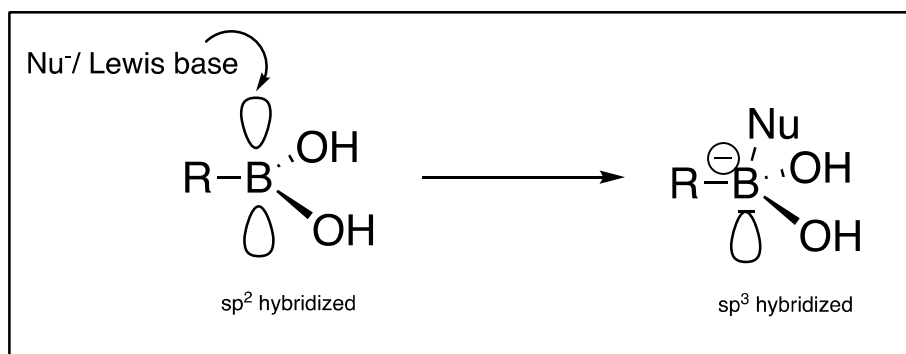


Figure 1.26: BAs behave as Lewis acids due to the vacant p-orbital activating it towards nucleophilic attack by nucleophiles.

Adapted from: <sup>60, 65</sup>

#### 1.4.1.4 Safety Handling

Considering their benign nature, BAs are frequently referred to as "green" chemicals that are non-toxic to the environment and humans.<sup>60</sup> Compared to other organic compounds, BAs have gained recognition in medicine due to their ease of handling and low cytotoxicity.<sup>59</sup> Recently, several drugs for treating cancer patients with a BA backbone have been approved by Food and Drug Administration (FDA) to enter the marketplace. Compared to other hydrocarbons, BA is a superior contender because it eventually degrades to its non-toxic equivalent boric acid, as seen in Figure 1.23. According to reports, consuming tiny amounts of BA has no adverse health impacts on individuals unless they also have other underlying issues.<sup>60</sup>

#### 1.4.2 Stability of Boronic Acids

Unlike carboxylic acids, BAs must be synthesized and cannot be isolated from natural sources.<sup>60</sup> Due to their instability in ambient oxygen, boranes undergo oxidation to produce borinic acid, which then undergoes additional oxidation to yield BA. All the successive intermediates are shown in Figure 1.23.<sup>59,60</sup>

#### 1.4.2.1 Auto-oxidation/Air-oxidation

In the presence of water, BAs are vulnerable to air oxidation. Despite being a slow process kinetically, the cleavage of the B-C bond is advantageous thermodynamically because of the different bond energies between the two corresponding atoms, B-C and B-O.<sup>60</sup> Typically, aryl BAs produce phenols during oxidative cleavage,<sup>60</sup> as seen below in Figure 1.27. Although phenols are commonly thought of as by-products, this method can rapidly synthesize functionalized phenols.<sup>66</sup> However, no actual procedure for synthesising phenols employing an oxidative hydroxylation strategy and aryl BAs is outlined. Typically, phenols can be utilized as flexible intermediates in various cross-coupling processes.<sup>66</sup> The kinetics of the reaction is accelerated in the presence of a ligand giving the Cu-based catalyst a surface to coordinate on during the oxidative hydroxylation process at room temperature (rt) in the presence of water. Thus, without much input, this technique can produce a range of ortho and para-substituted phenols in high yields via this technique.<sup>66</sup>

Compared to other species, BA derivatives with potent electron-withdrawing substituents tend to be more resistant to oxidation. In addition, the process becomes even slower in an aqueous basic medium.<sup>60</sup> Under air, BA and its derivatives, including 2-heterocyclic, vinyl, and cyclopropyl BAs, quickly degrade. External stimuli, such as heat, a base, or a catalyst, accelerate the degradation and make it more competitive with other cross-coupling reactions. Strong oxidants like peroxides can quickly oxidize BA esters in addition to BA. The development of air-stable surrogates, such as trialkoxy or trihydroxy borate salts, diethanolamine adducts, sterically bulky boronic esters, and boroxines, has led to the development of several ideas to address this problem.<sup>67</sup>

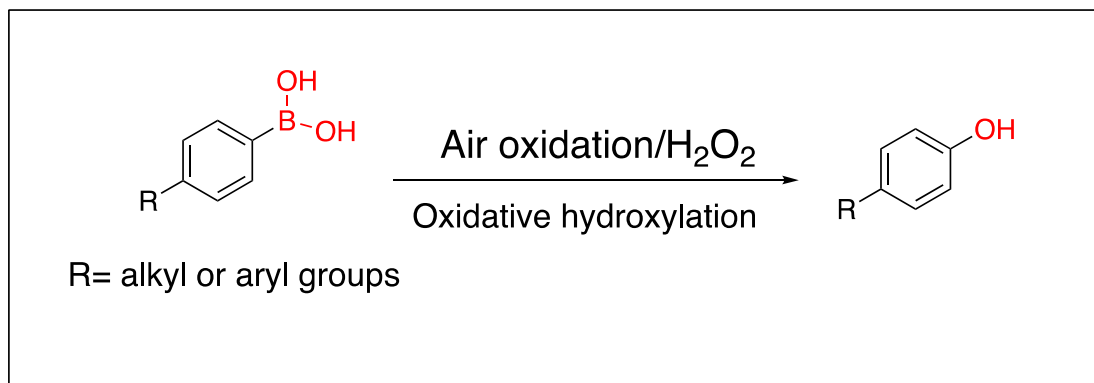


Figure 1.27: Oxidative hydroxylation of aryl BAs to form phenols at rt in presence of oxidising agents or air.

Adapted from: <sup>66</sup>

#### 1.4.2.2 Protodeboronation

The proteolysis of the B-C bond is another decomposition pathway that aryl BAs undergo. This undesired reaction turns the  $\text{Ar-B(OH)}_2$  into  $\text{Ar-H}$ .<sup>61</sup> Although most aryl BAs are often resistant to protodeboronation, the reaction cannot be inhibited at high temperatures.<sup>60</sup> Eventhough protodeboronation reduces reaction yields and gives rise to side-products, it can be helpful in some cases, such as removing  $\text{B(OH)}_2$  is required.<sup>61</sup> The pH of the medium significantly impacts the deboronation of aryl BAs; in a basic medium, reactions are substantially expedited, whereas, in neutral and acidic media, reaction kinetics are slow, and electronic factors not being considered.<sup>68</sup> Deboronation activities can impede reactions that use BAs as reagents because the Suzuki-Miyaura cross-coupling reaction necessitates a high temperature in basic solutions.<sup>61</sup> Elect electronic factors greatly influence the reaction rates, with electron-deficient substituents deboronating more rapidly than their electron-rich counterparts. In such cases, aryl BAs tend to deboronate more quickly in acidic media.<sup>60,61</sup> This has currently increased interest in protodeboronation studies. Numerous articles discuss the kinetics of protodeboronation under various conditions, including the influence of acids, temperature changes, and aryl BA ring substitution.<sup>61,68,69</sup> It was noted that the protodeboronation reaction is complex in pH-

dependent media with at least two proposed mechanisms, seen in Figure 1.28.<sup>69</sup> Additionally, all varieties of aryl BAs can be deboronated in the presence of a metal catalyst.<sup>70</sup>

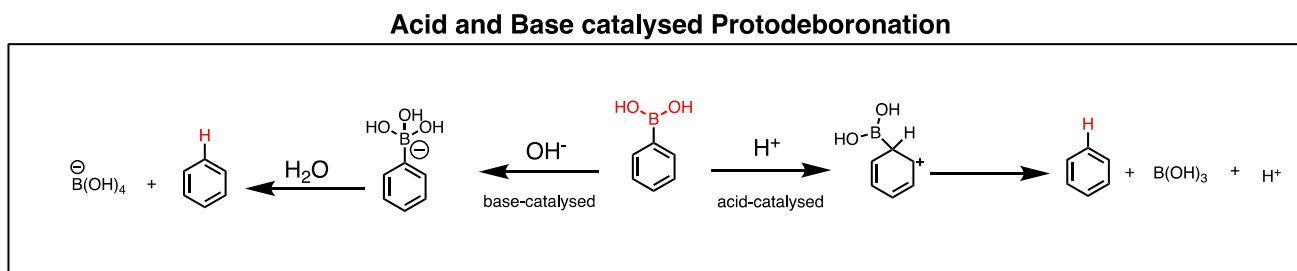


Figure 1.28: Acid and Base catalysed Protodeboronation of BAs.  
Adapted from:<sup>70</sup>

### 1.4.3 Applications of Boronic Acids

As mentioned in earlier sections, the applications in bioconjugation engineering have seen exponential growth by designing and constructing multifunctional adducts endowed with the characteristics of the payload used, as also seen in Figure 1.21. The payload used defines the biomolecule's principal chemistry, evoking its pharmacological purpose.<sup>71</sup> The payload induces the therapeutic characteristics required for the construct is a critical component of bioconjugates. The improved molecule can carry out chemical processes and is more functional in terms of fluorescence, improved biological activity, and affinity toward a specific functional group depending on its payload.<sup>72</sup> However, as earlier mentioned, these bioconjugation reactions are subject to low reaction efficiencies and slow kinetics.<sup>59</sup> The solution to this problem is to connect the payload to the necessary biomolecule via chemical spacers.<sup>59</sup> BAs are ideal candidates to perform this task since they can function as warheads, conjugate linkers, and payloads themselves.<sup>73</sup> BAs as payloads and warheads will be discussed further in the proceeding sections.

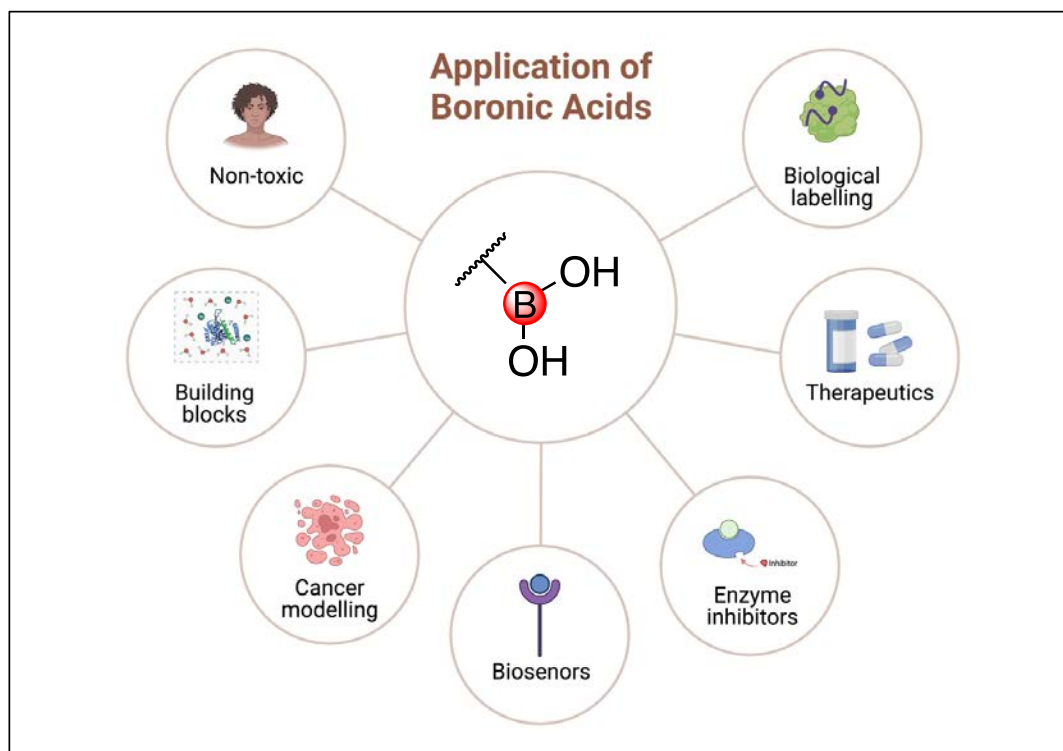


Figure 1.29: BAs for sensing and other general applications

#### 1.4.3.1 BAs as Payloads

Due to the simplicity with which BAs can be incorporated into organic compounds through cross-coupling reactions, such as the Suzuki Miyaura Borylation technique, they have been frequently utilized as payloads in the production of bioconjugates. It is possible to construct bioconjugates using a range of BAs moieties as payloads. These bioconjugates can be exploited therapeutically for developing anti-microbial drugs, as well as in the treatment of cancer. It has numerous uses, both as prodrugs and drugs.<sup>59</sup>

##### 1.4.3.1.1 Application of Boronic Acids as Drugs

With a significant focus on drugs that can be used as protein inhibitors,<sup>74</sup> BAs have a wide range of applications as bioactive components. Only a limited fraction of the more than 300 protein

inhibitor designs could survive the clinical trials.<sup>75</sup> In comparison to other nucleophiles, it has been found that BAs preferentially inhibit serine proteases.<sup>59</sup> The mechanism of action is the formation of a tetracoordinate boronate complex with the hydroxyl side chain of a serine residue. The reaction is shown in Figure 1.30. Figure 1.31 shows some common examples of BAs and their derivatives in pharmacologically relevant compounds. The anti-fungal agents Tavaborole and Crisaborole for treating eczema are just a few of the medications to reach the market.<sup>76</sup> Ixazomib is the first BA medicine used orally, while Bortezomib is another proteasome inhibitor used to diagnose symptoms like multiple myeloma.<sup>77</sup> The primary issue with designing medications with BA moieties is their poor pharmacokinetic qualities due to their sensitivity to numerous nucleophiles. Recently, this issue has been overcome by incorporating nitrogen-containing heterocycles in BAs, which can control its stability and avoid off-target drug delivery.<sup>59,78</sup>

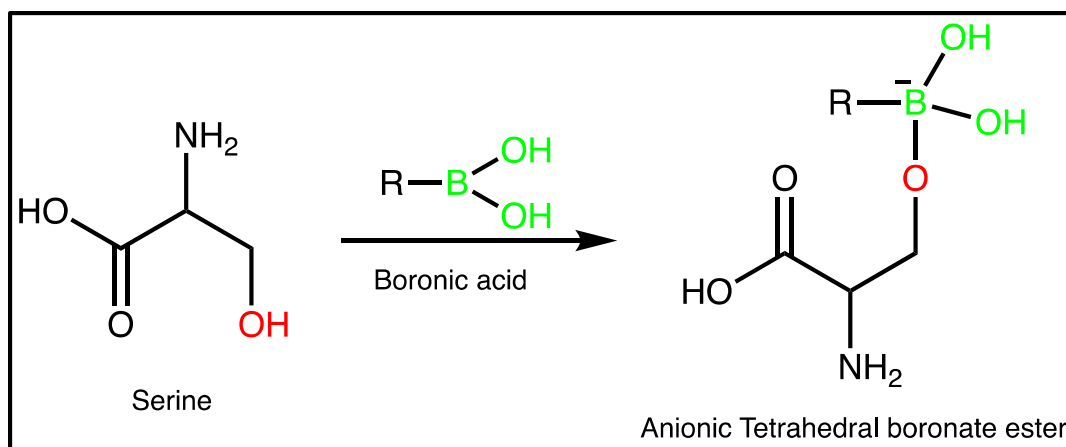


Figure 1.30: BAs are commonly employed as serine proteases due to their reaction with serine side chain residues forming anionic boronate ester.  
Adapted from:<sup>79</sup>

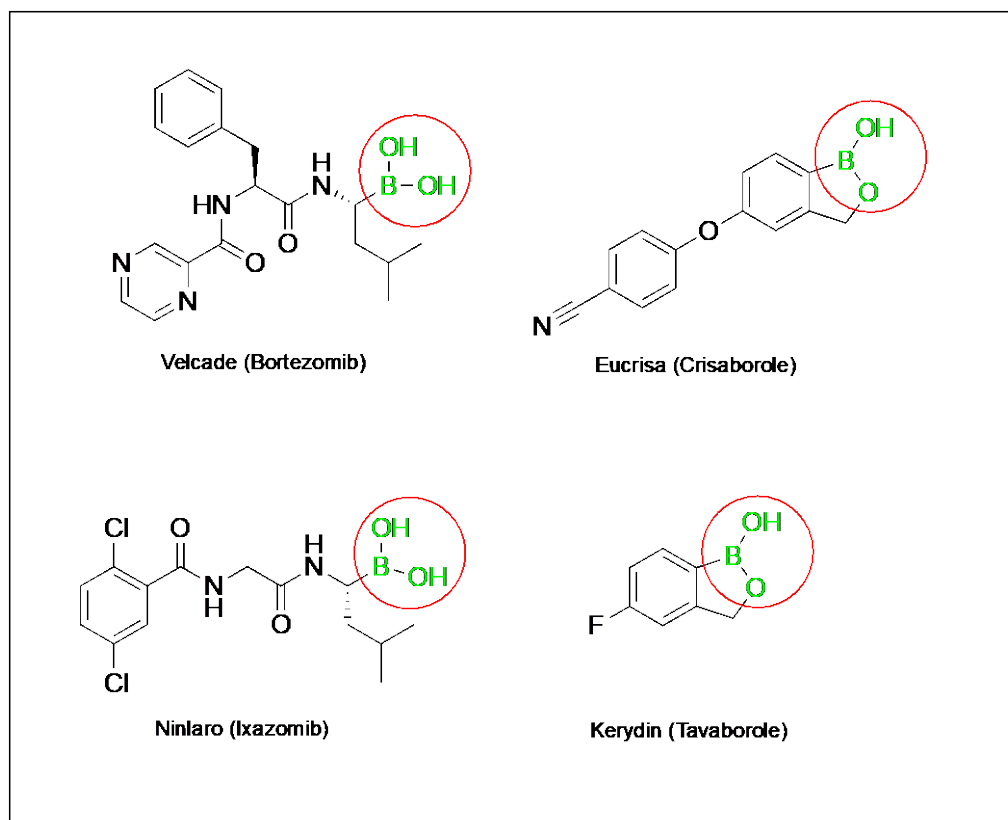


Figure 1.31: BA backbone in pharmacologically relevant compounds.  
Adapted from: <sup>59</sup>

#### 1.4.3.1.2 Application of boronic acids as prodrugs

As mentioned, aryl BAs readily oxidize to phenols in the presence of oxidants such as  $\text{H}_2\text{O}_2$  or other reactive oxygen species (ROS). This property of phenyl BAs have been exclusively exploited to design prodrugs having the phenol group from the original structure replaced by a BA structure forming the prodrug. The elevated ROS concentration in a tumour cell can stimulate payload release and convert the BA to a phenol.<sup>80</sup> This technique has proven to give better results *in vivo* as compared to the original drug itself. The bio-active prodrug for FDA-approved Irinotecan has achieved better *in vivo* outcomes in mice models.<sup>59</sup> Instead of phenol, the prodrug SN-38 contains a borylated BA, which can be actively converted to the desired drug from the prodrug at  $\text{H}_2\text{O}_2$  concentrations as low as 0.5

mM.<sup>81</sup> The original molecule loses its fluorescence after borylation. Still, when oxidative species are present, it is changed to its phenol counterpart, which revives its fluorescence properties, allowing it to be exploited as a fluorescent probe as well.<sup>82</sup> The underlying molecular mechanisms driving ROS signalling for targeted drug delivery can be explored further.<sup>59</sup>

#### 1.4.3.2 BAs as Warheads

Bioconjugates are often designed to be able to withstand the complexity of the biological environment. In addition to serving as payloads, BAs can be used to construct bioconjugates that rely on reversible conjugation chemistries known as warheads. Covalent bonds can be produced by BAs reversibly ligating various nucleophiles. These conjugates are capable of generating responses to a stimulus.

##### 1.4.3.2.1 Boronate Esters (BEs)

BAs can react efficiently with various nucleophiles, such as carbohydrates, reversibly replacing the hydroxyl groups with alkoxy or aryloxy groups forming boronate esters (BEs). The reaction is feasible and has fast reaction rates, as the boron centre easily coordinates with vicinal diols to yield the less polar esters.<sup>59</sup> Combining BAs with esters also acts as a protective measure for the simple handling and storage of BAs.<sup>60</sup> The reversible reaction, causing hydrolysis, can be regulated using factors such as pH of the medium and the diol's structure. Rigid and restricted cis diols are favourable for stability if the correct pH is maintained. Azeotropic distillation using water is a typical method to push the reaction forward in cases where rigid or cyclic diols are not present.<sup>60</sup> Figure 1.32 demonstrates that higher pH facilitates the reaction by deprotonating the BA, which enables easy

complexation of the anion with the diol.<sup>59</sup> The hydrolysis reaction is fast in an acidic medium and therefore arouses concerns regarding the stability of the ester.

Additionally, the solubility of the product also plays a major role in isolating the product, as soluble products are challenging to isolate compared to insoluble products. The factors mentioned above affect the general stability of BEs, although the factors that influence this process are still subject to debate.<sup>83</sup> BAs are widely utilized as carbohydrate sensors that can be used to quantify carbohydrates due to the simplicity of the reaction.<sup>84</sup> Additionally, BAs and carbohydrate affinity can be fully utilized for targeted drug delivery because unwanted glycosylation causes malignant cells to multiply, producing cancerous tumour cells because of overexpressed antigens.<sup>85</sup> Therefore, BEs ability to conjugate with organic residues has a variety of biomedical applications.<sup>59</sup> The strong carbohydrate affinity of phenylboronic acid (PBA), as seen in Figure 1.22, is a valuable application of this technology. In PEG polymer chains, PBA is typically included as a warhead to act as a sugar-binding motif and prevent cell adherence to cell surfaces.<sup>59</sup> Using this technology, transplants of artificial prosthetic organs successfully prevent antibodies from adhering to red blood cells.<sup>86</sup>

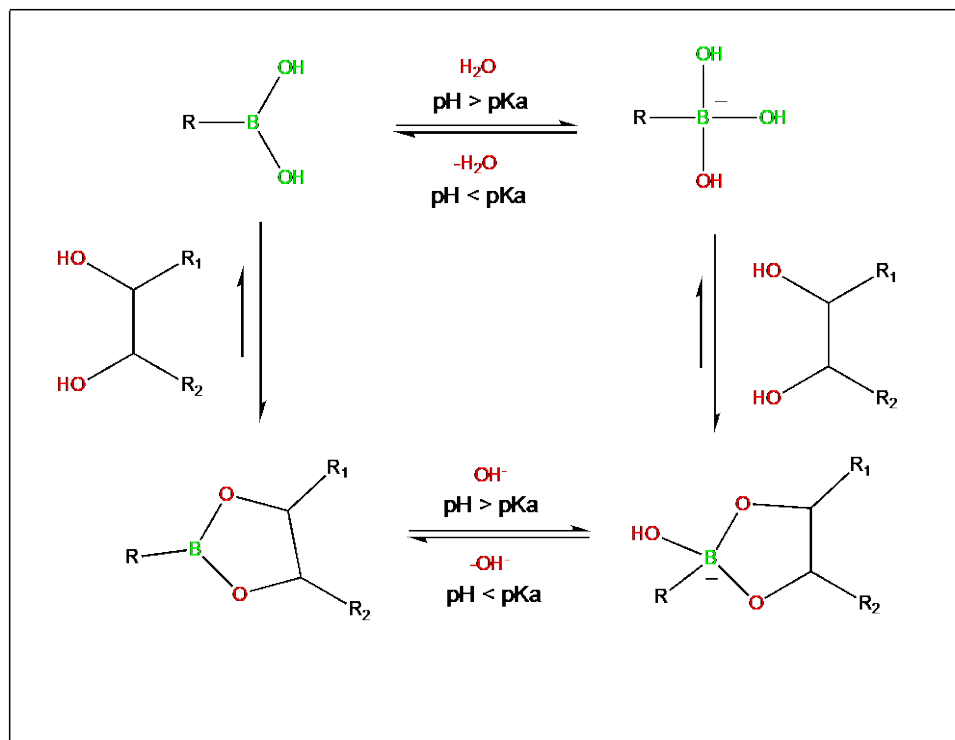


Figure 1.32: Mechanism of BA and diol complexation in aqueous solutions.  
Adapted from: <sup>83</sup>

BE deprotection can be problematic in some cases. This is because the reaction equilibrium favors the starting BE.<sup>87</sup> However, two methods can be applied for deprotection, described in Figure 1.33. Deprotection can be done in the presence of oxidizing agents to afford the product as aryl BAs or phenols.<sup>87</sup> One limitation whilst working with nucleophilic oxidants is that the reaction can be destructive. The alternative technique is a standard trans-esterification reaction, which typically uses aromatic BAs, such as PBA, in a biphasic solvent system at a low pH and can then be further separated using ion exchange chromatography.<sup>88</sup> When boronates differ in solubility, the second approach is commonly used.<sup>87</sup>

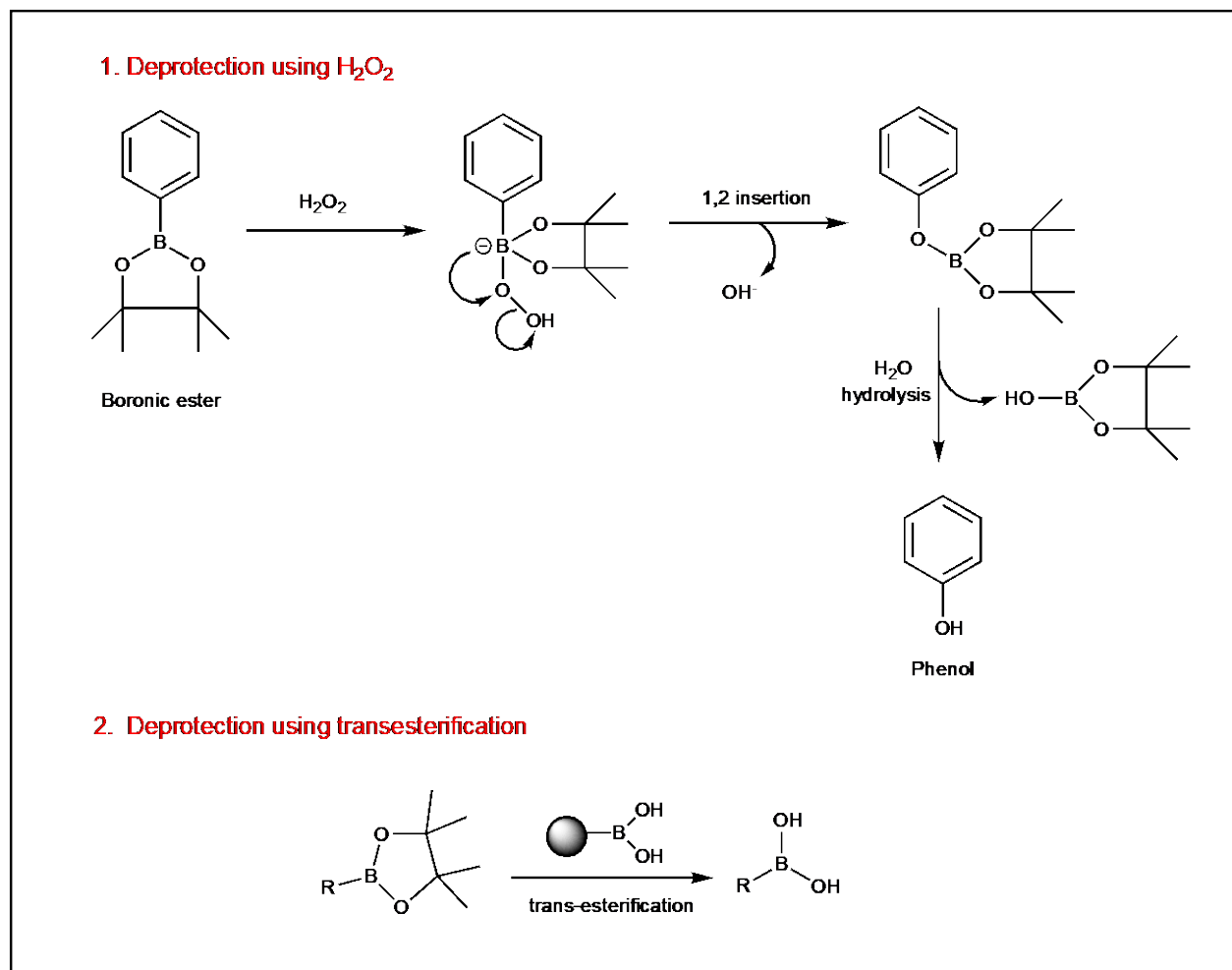


Figure 1.33: Deprotection of BEs. 1: Oxidative cleavage using  $H_2O_2$ . 2. Transesterification. Adapted from: <sup>89</sup>

## 1.5 Current Cross-linkers

Protein samples can currently be cross-linked using a variety of cross-linking agents. Typically, the cross-linkers contain an enrichment handle for enriching cross-linked peptides, and biotin is the most commonly utilized enrichment reagent. Streptavidin and avidin are two glycoproteins for which biotin has a great affinity. Biotin serves as the enrichment handle in the cross-linker biotinylated Azo-Leiker 2 and can be subjected to AP using streptavidin beads. However, biotin is exceedingly bulky

and, if not cleaved during AP, can obstruct subsequent MS analysis. The large handle also inhibits cell permeability.

Additionally, because biotin is already present in cells, it may interact in non-specific ways, resulting in indirect binding.<sup>30</sup> The other frequently employed cross-linkers utilize azide-alkyne cycloadditions for AP, for example, Azide-A-DSBSO. Usually, digestion is followed by adding the azide tag to the surface of the cross-linker, and then the peptides are exposed to purification. However, because of the low affinity, azide-alkyne click chemistry typically necessitates the introduction of biotin addition, adding extra steps to the purification process.<sup>90</sup> Recently, a cross-linker called PhoX that can be enriched via immobilized metal affinity chromatography (IMAC) was also designed. PhoX has a sturdy phosphonic handle. However, PhoX cannot be utilized for *in vitro* cross-linking studies because its negative charge prohibits it from passing through cellular membranes. Although the negative charge can be concealed, -NHS esters are vulnerable to hydrolysis as they pass through the membranes.<sup>31</sup> To overcome these problems, a new version of PhoX was developed, where the bulky tert-butyl group protects the phosphonic handle. Although tBu-PhoX solves the cell permeability issues, phosphate groups are post-translation modification (PTM) on proteins and hence can be quickly phosphorylated to endogenous phosphopeptides. Therefore, extra steps would be required to remove these endogenous peptides from the mixture of cross-linked peptides.<sup>91</sup>

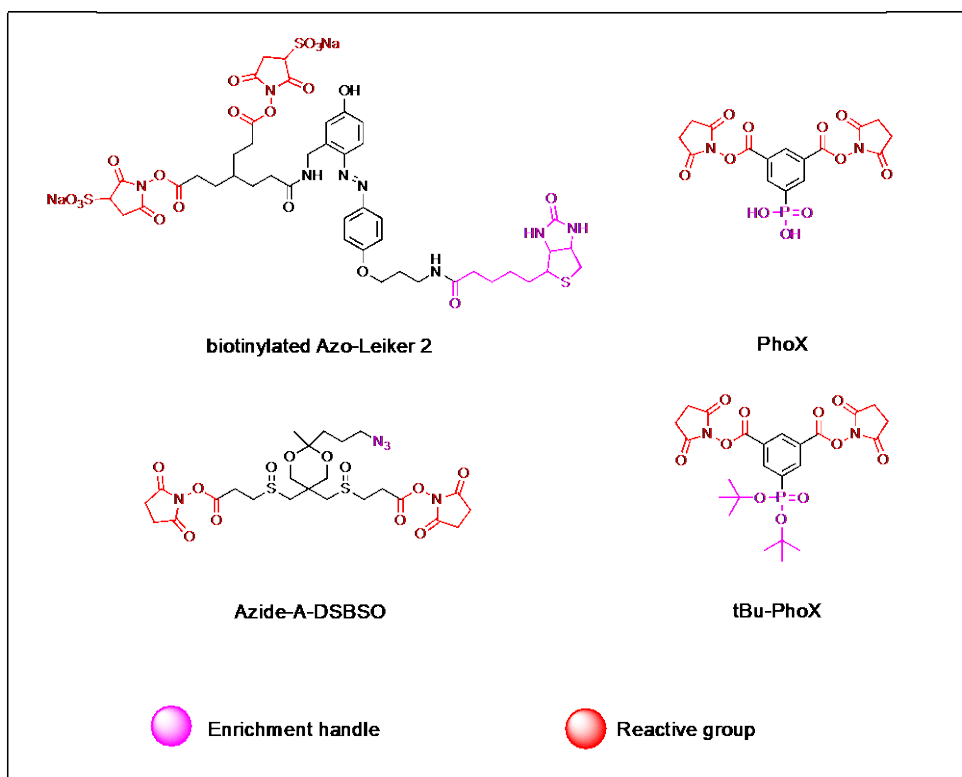


Figure 1.34: Examples of cross-linkers with common Enrichment handles

## 1.6 Conclusion

Despite having a high enrichment efficiency, the existing cross-linkers have some drawbacks. If the enrichment handle is not cleaved during AP, its enormous size may lead to non-specific interactions impeding the subsequent MS analysis. In such circumstances, cell permeability is also problematic. Handles having a weak binding affinity with their other binding partner require extra steps to add more efficient handles for adequate enrichment. Additionally, the cell-impermeability of cross-linkers due to negative charges limits their utility for live-cell cross-linking.

Lastly, cross-linkers that occur as frequent PTMs are linked to issues with co-enrichment of the endogenous peptides and actual peptides, necessitating more separation and lengthening the analysis. Therefore, there is a need for a non-bulky enrichment handle having a condensed spacer arm, which is orthogonal and has high specificity for its other binding partner for robust enrichment. The cross-linker

should be compatible with easy single-step enrichment strategies for a rapid workflow. Novel cross-linkers that can be applied for live-cell cross-linking and that have enrichment handles that are not found as PTMs can sabotage the limitations of the currently used cross-linkers.

## CHAPTER 2

### SYNTHESIS OF CROSS-LINKERS AND DIOL RESIN

#### 2.1 Introduction

A low abundance of cross-linked peptides after XL-MS renders routine analysis problematic. Therefore, to overcome this challenge, one solution is to enrich cross-linked peptides prior to LC/MS analysis. Current cross-linkers have high cross-linking efficiency however are associated with problems such as bulky size leading to non-specific interactions, negative charge leading to cell permeability issues, extra steps to perform purification during enrichment, containing phosphate groups as handles which commonly occur as PTM, and non-cell permeable. Hence, the primary focus of this research was on the design and synthesis of two novel cross-linkers decorated with a stable BA enrichment handle. The following section describes the synthesis and the related chemical data for two BA-containing cross-linkers, SMT and LMT, respectively as well as the synthesis of the diol-containing-resin.

#### 2.2 Materials

(1R)-(-)-Nopol (Sigma Aldrich, 98%), sodium sulphate anhydrous (Fisher Chemical, Certified ACS 100.3%), potassium carbonate anhydrous (Fisher Chemical, Certified ACS 99.94%), carboxylink coupling gel (thermo Scientific, loading-16 umoles/ml gel), MES free acid monohydrate (Fisher bioagents, 98% min.), potassium phosphate monobasic (Fisher Chemical, Certified ACS 99.0%), sodium phosphate monobasic (Fisher Chemical, Certified ACS 99.0%), 4-carboxybenzeneboronic acid (Boron Molecular, >97%), O-(1H-6-Chlorobenzotriazole-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (Chem-impex Int'l Inc., 99.9%), 3,5-Dicarboxyphenylboronic acid (AOB Chem, 95%), 1,3-Dicyclohexylcarbodiimide (Chem-impex Int'l Inc., 99.42%), N-Hydroxysuccinimide (TCI

Chemicals, 98%) (HOSu), urea, crystallized (VWR Chemicals, ACS), guanidine hydrochloride (Alfa Aesar, 98%), boric Acid (Fisher Chemical, Certified ACS crystalline, 99.9%), EDTA (Fisher Chemicals, Certified ACS 99.0 to 101.0%), HEPES (Alfa Aesar, 99%), 2,2'-Azobis(2-methylpropionitrile) (AIBN, Aldrich Chemistry, 98%), lithium hydroxide monohydrate (Alfa Aesar, 56.6% min.), potassium hydrogen fluoride (Alfa Aesar, 98%), 4-Methylpiperidine (Acros Organics, 99%), silica gel (Acros Organics) and 1-Hydroxybenzotriazole hydrate (Chem-impex Int'l Inc., 99.9%) were used without further purification.

Dimethyl sulfoxide (Sigma Aldrich, anhydrous, 99.9%), hydrochloric acid (Fisher Chemical, Certified ACS plus, 21.1 Normal), ethyl acetate (Fisher Chemical, Certified ACS 99.5%), ethyl ether anhydrous (Fisher Chemical, BHT stabilized, Certified ACS 99%), hexanes (Fisher Chemical, Certified ACS 98.5%), dichloromethane (Fisher Chemical, Certified ACS Stabilized 99.5%), dimethylformamide (J.T.Baker, ACS Reagent 99.8%), N,N-Dimethylformamide anhydrous (Acros Organics, 99.8%, Extra Dry), methanol (Fisher Chemical, Certified ACS 99.8%), hydrogen peroxide (Fisher Chemical, Certified ACS 30%), trifluoroacetic acid (Chem-impex Int'l Inc., 99.9%), acetonitrile (VWR Chemicals, HPLC grade), ethanol (Decon Labs, Inc., 200 Proof Anhydrous), sulfuric Acid (Fisher Chemical, Certified ACS Plus 95.0 to 98.0 %), acetic anhydride (Fisher Chemical, Certified ACS ) and acetone (Fisher Chemical, Certified ACS 99.5%) were used without further purification.

4-Toluenesulfonyl chloride, pyridine (ACS reagent), sodium azide, 4-Methylmorpholine 4-oxide (NMO), potassium dioxidodioxosmium dihydrate, 4-Pentynoic acid, 2-Chloro-1-methylpyridinium iodide, 4-Nitrophenol, 2,6-Dimethylpyridine, 2,2'-Dipyridyl (ACS grade), copper(I) bromide (Trace metals grade), (+)-sodium L-ascorbate crystalline, N,N-Diisopropylethylamine (DIEA), N-Bromosuccinimide (NBS), 5-Bromo-m-xylene, sodium cyanide, Bis(pinacolato) diboron (Bpin), potassium acetate and 1,1'-Bis(Diphenylphosphino)ferrocenepalladium (II) were purchased from

Oakwood Chemicals. All other reagents and solvents were commercially available and used as received.

All the amino acids used were Fmoc protected and were purchased from Chem-impex Int'l Inc., 99.88% assay. Rink amide-MBHA resin (Chem-impex Int'l Inc., 0.3-0.6 mmol/g, 100-200 mesh). Chloroform-d for NMR, (Acros Organics, 99.8+ atom % D, contains 0.03 v/v% TMS) and methyl sulfoxide-d<sub>6</sub>, for NMR (Acros Organics, 99.9 atom % D) were used without further purification

### 2.3 Instrumentation

Unless otherwise stated, all the reactions were performed under a nitrogen atmosphere from Schlenk lines using acetone-dried glassware. If some reactions required dry solvents they were obtained from the Innovating Technology Pure Solv MD-7 solvent purification system (SPS) equipped with Schlenk manifold. Thin layer chromatography (TLC) was performed on 60A R24D aluminium backed TLC plates (Oakwood Chemicals) and was visualized with a 254 nm Bio Surplus UV-detector or KMnO<sub>4</sub> stain. UV-Vis absorbances were measured on a Cary Bio 100 UV-Vis Spectrometer using the necessary solvent as a reference in a cuvette (Beckman USA) for completing the baseline scan. Pure Prep Chorus peptide synthesizer was used for automated solid-phase peptide synthesis (SPPS). Samples were concentrated on a rotary evaporator Hei-VAP core purchased from Heidolph equipped with a chemical resistant vacuum pump purchased from Rocker Scientific Co., Ltd.

Some small-scale reactions were monitored for completion on a Varian Prostar analytical high performance liquid chromatography (HPLC) system using solvents like ultrapure water (Gen Pure UV-TOC/UF system) and acetonitrile with 0.1% trifluoroacetic acid (TFA, HPLC grade) as solvent A and B respectively. The sample was introduced into the column using direct injection technique. Samples were prepared in solvent B with maximum loading capacity of 10 ug in the reverse phase column

purchased from Agilent, (Zorbax, SB-C18 5  $\mu$ M, 4.6 x 150 mm). The column was equilibrated with the same standard before each run for at least 7 minutes. Before injection some samples were centrifuged on a micro-centrifuge purchased from Eppendorf and the vortex shaker from Fisher brand. Varian Prostar preparative high performance liquid chromatography (HPLC) equipped with a Varian Prostar UV-Vis detector was used for purification of short peptide sequences and small organic molecules. With the same solvent system as mentioned above, the sample was dissolved in either of the two solvents before manual syringe filtration using PTFE 0.45  $\mu$ M syringe filters. Direct injection was used to inject the sample onto the column purchased from Higgins Analytical Inc., (Proto 300 C4 10  $\mu$ m, 250 x 20 mm). Maximum loading capacity for the column is 5-6 mg/ml.

After HPLC purification, the collected solutions were lyophilized under high vacuum in a 115 Volts Labconco Corporation freeze dryer for 2 days or until dry white powder is observed in the VWR centrifuge tubes. ESI-MS spectra were recorded on a Agilent Technologies 6230 TOF Mass spectrometer; all spectra were measured in positive reflector mode. Samples were prepared to a final concentration of 1  $\mu$ g/mL in 1:1 MeOH:H<sub>2</sub>O (v/v). The sample was introduced to the electrospray source via direct injection at a flow rate of 250  $\mu$ l/h. A mass range of 200-800 m/z was scanned. Under these conditions, ([M+1]<sup>+</sup>) were observed.

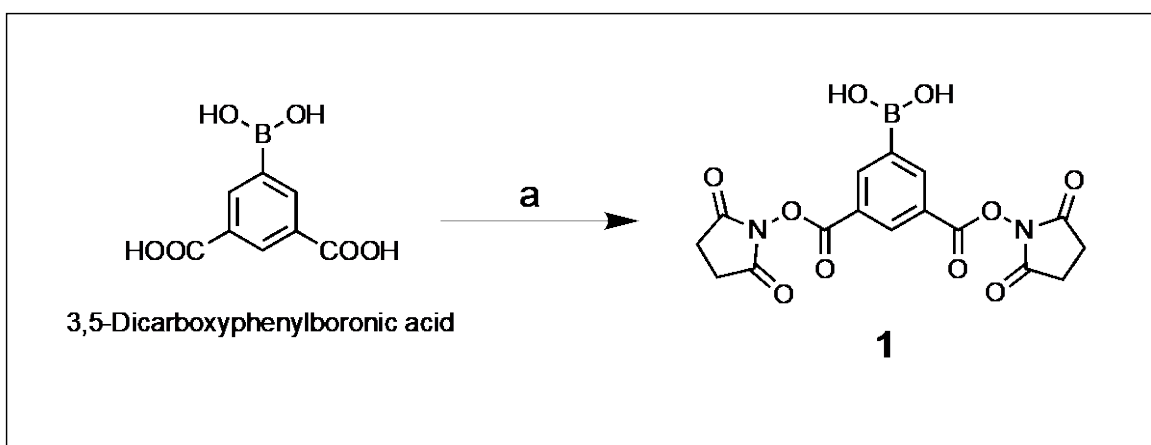
Affinity purification experiments were conducted in 2.0 ml non-sterile Spin-X Centrifuge Tube filter purchased from Corning Incorporated, 0.45  $\mu$ m cellulose acetate. Low retention pipette tips were bought from Fisher brand Sure One.

<sup>1</sup>H NMR ( $\delta$ , ppm) spectra were recorded on a Unity Inova 400 (400 MHz) NMR spectrometer. All spectra were recorded in d-CDCl<sub>3</sub> or d-DMSO and resonances were referenced to tetramethyl silane

(TMS) (0.00 ppm). The residual solvent protons ( $^1\text{H}$ ) of  $\text{CDCl}_3$  (7.26 ppm),  $\text{ACN-d}_3$  (1.94 ppm),  $\text{DMSO-d}_6$  (2.50 ppm),  $\text{acetone-d}_6$  (2.05 ppm),  $\text{CD}_3\text{OD}$  (3.31 ppm) were used as internal standards.

## 2.4 Synthetic Schemes

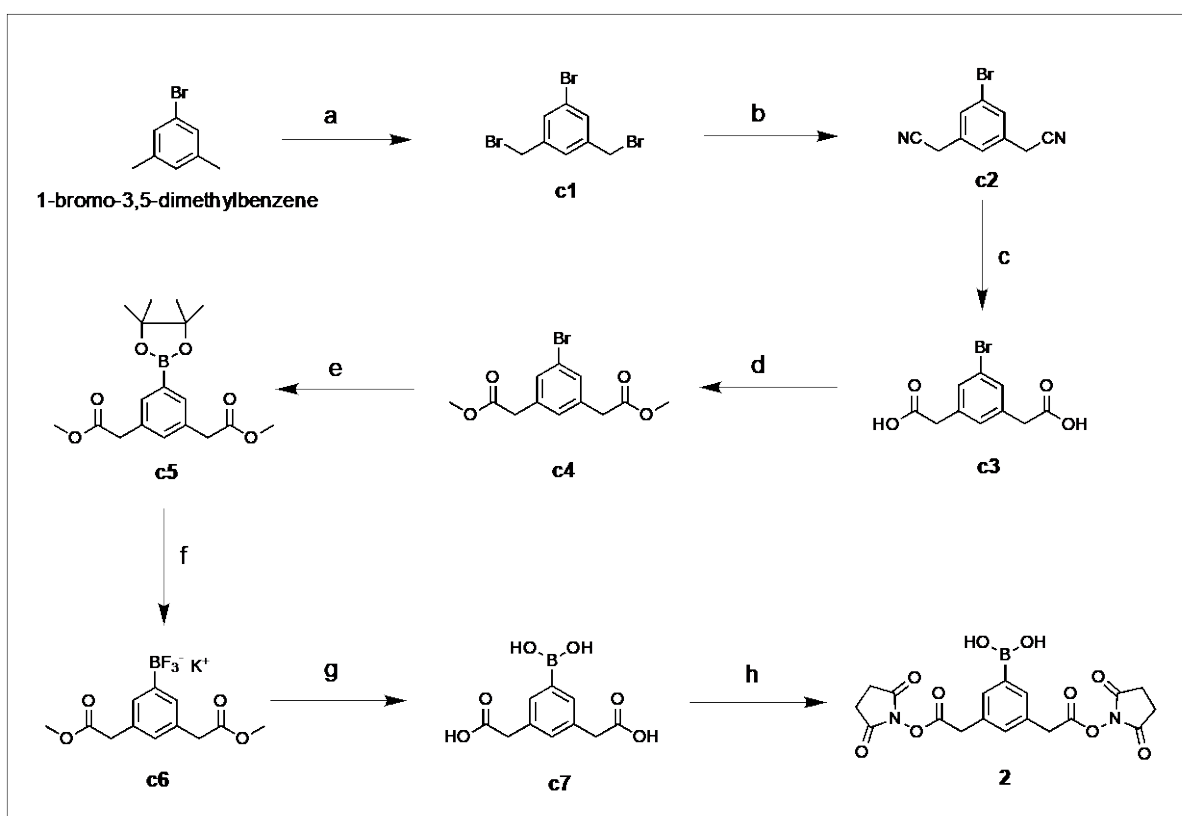
### 2.4.1 Synthesis of Shorter Cross-linker (SMT)



Scheme 2.1: Synthesis of (3,5-bis(((2,5-dioxopyrrolidin-1-yl)oxy)carbonyl)phenyl)boronic acid

Reaction conditions: (a) DCC, HOSu, dry DMF, rt, 24 h.

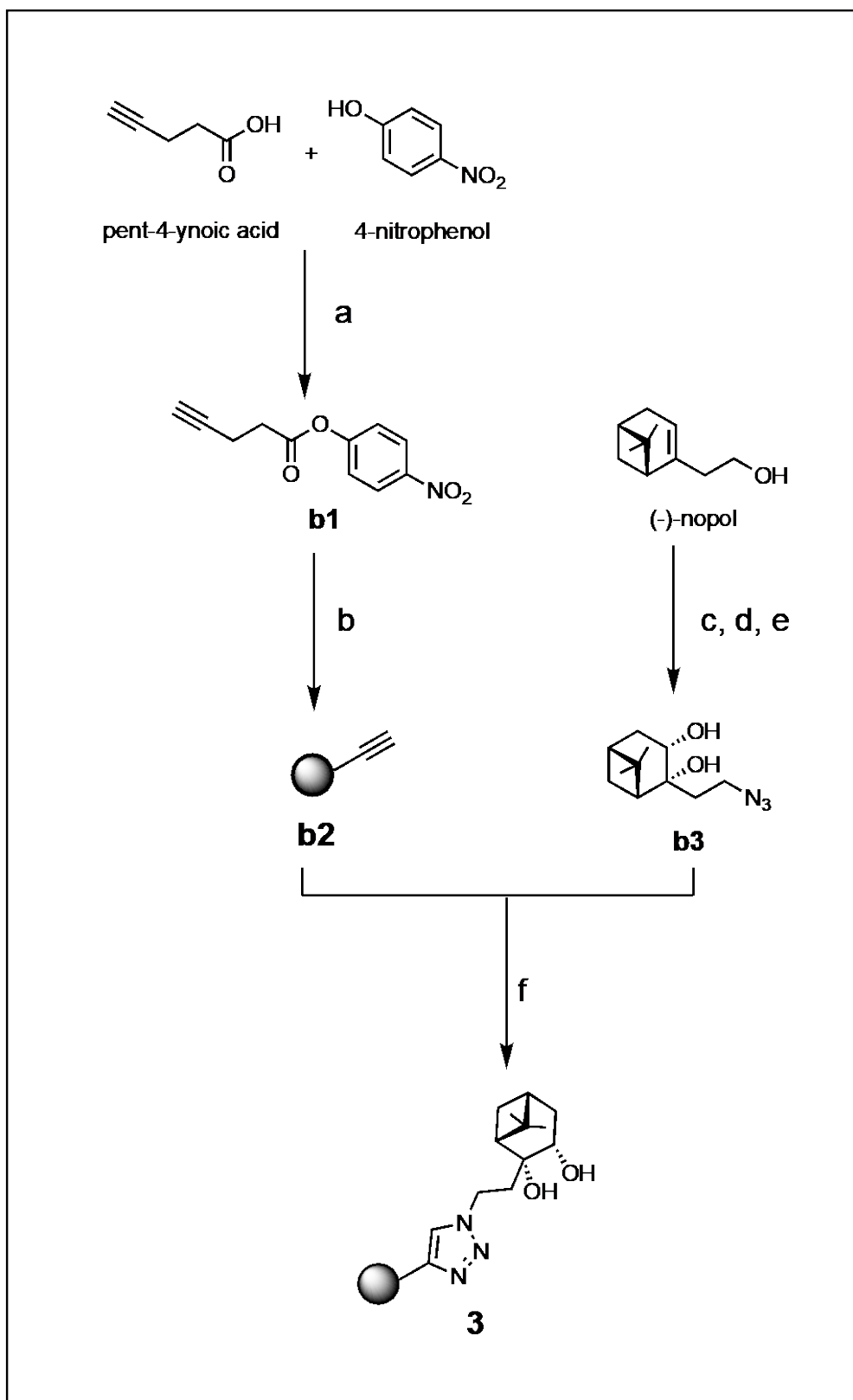
## 2.4.2 Synthesis of Longer Cross-linker (LMT)



Scheme 2.2: Synthesis of (3,5-bis(2-((2,5-dioxopyrrolidin-1-yl)oxy)-2-oxoethyl)phenyl)boronic acid.

Reaction conditions: **(c1)** *1-bromo-3,5-bis(bromomethyl)benzene* **(a)** NBS, AIBN, CH<sub>3</sub>CN, 80 °C, 3 h. **(c2)** *2,2'-(5-bromo-1,3-phenylene)diacetonitrile* **(b)** NaCN, EtOH, H<sub>2</sub>O, rt, 30 mins. **(c3)** *2,2'-(5-bromo-1,3-phenylene)diacetic acid* **(c)** conc. HCl, 130 °C, 24 h. **(c4)** *dimethyl 2,2'-(5-bromo-1,3-phenylene)diacetate* **(d)** dry MeOH, trace H<sub>2</sub>SO<sub>4</sub>, 140 °C, 24 h. **(c5)** *dimethyl 2,2'-(5-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1,3-phenylene)diacetate* **(e)** B<sub>2</sub>pin<sub>2</sub>, KOAc, PdCl<sub>2</sub>(dppf), dry DMF, 90 °C, 24 h. **(c6)** Potassium salt of *dimethyl 2,2'-(5-(trifluoro-λ<sup>4</sup>-boraneyl)-1,3-phenylene)diacetate* **(f)** KHF<sub>2</sub>, MeOH, H<sub>2</sub>O, rt, 30 min. **(c7)** *2,2'-(5-borono-1,3-phenylene)diacetic acid* **(g)** LiOH, ACN, H<sub>2</sub>O, rt, 24 h. **(2)** *(3,5-bis(2-((2,5-dioxocyclopentyl)oxy)-2-oxoethyl)phenyl)boronic acid* **(h)** DCC, HOSu, dry DMF, rt, 24 h.

### 2.4.3 Synthesis of Diol Resin



Scheme 2.3: Synthesis of (1R,2S,3S,5R)-6,6-dimethyl-2-((4-methyl-1H-1,2,3-triazol-1-yl)methyl)bicyclo[3.1.1]heptane-2,3-diol.

Reaction conditions: **(b1)** 4-nitrophenyl pent-4-ynoate **(a)** 2-chloropyridinium iodide,  $K_2CO_3$ , DCM, rt, 4 h. **(b2)** agarose alkyne beads **(b)** carboxy link gel, 5% MES Buffer in DMF, rt, 48 h. **(b3)** (-)-nopol-azide-diol **(c)** Ts-Cl, pyridine, 0 °C, 4 h **(d)**  $NaN_3$ , DMSO, rt, 16 h **(e)**  $K_2OsO_4 \cdot 2H_2O$ , pyridine, NMO (50%), acetone-water, 65 °C, 24 h. **(3)** (1R,2S,3S,5R)-6,6-dimethyl-2-((4-methyl-1H-1,2,3-triazol-1-yl)methyl)bicyclo[3.1.1]heptane-2,3-diol **(f)** 2,6-Dimethylpyridine, 2,2'-Dipyridyl, CuBr, sodium ascorbate, DMF, rt, 48 h.

## 2.5 Results and Chemical Synthesis and Analytical Data

The following section describes the detailed procedure for the chemical synthesis of all the intermediate steps required for the synthesis of the cross-linkers, SMT and LMT. Also, it covers the necessary steps involved in the synthesis of the diol-containing resin.

### 2.5.1 S1: Synthesis of (3,5-bis(((2,5-dioxopyrrolidin-1-yl)oxy)carbonyl)phenyl)boronic acid

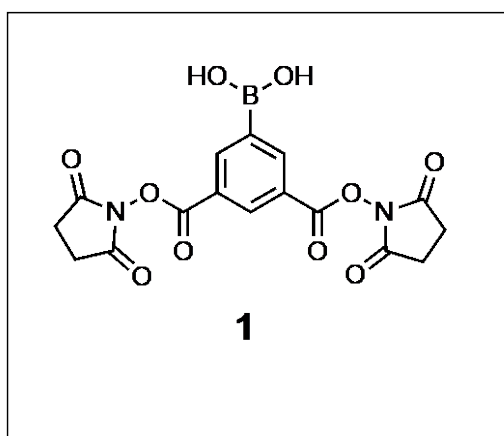


Figure 2.1: Shorter cross-linker (SMT)

A 50 ml, single-necked, round-bottomed flask containing a Teflon coated magnetic stir rod was charged with the dry ingredients: 3,5-dicarboxyphenylboronic acid (300 mg, 1.42 mmol, 1.0

equiv.), followed by DCC (737.2 mg, 3.57 mmol, 2.5 equiv.) and HOSu (411.23 mg, 3.57 mmol, 2.5 equiv.). This dry mixture was degassed with three vacuum/nitrogen purge cycles. Extra dry DMF (7.5 ml) was obtained via a syringe while the bottle was purging under N<sub>2</sub> and is added through the rubber septum. Again, the reaction vessel was degassed with three vacuum/nitrogen purge cycles. The resulting reaction mixture was left for stirring overnight at room temperature under N<sub>2</sub>, during which the reaction mixture turns cloudy due to the formation of insoluble DCU. After, the reaction mixture was mini-centrifuged to remove the excess precipitate of DCU, followed by manual syringe filtration using minimum amount of dry DMF (4 ml). The mixture was then further concentrated by rotary evaporation (42 °C, 6-10 mbar) to give a white solid. The solid was dissolved in 1:1 HPLC solvents A and B before injecting onto preparative HPLC (15-70 method, 254 nm detection) column, for final purification of the crude. The solution was then freeze-dried under high vacuum for 2 days to afford the product (**1**) as a white bench stable powder. <sup>1</sup>H-NMR (400 MHz; CDCl<sub>3</sub>): δ 2.87 (s, CH<sub>2</sub>, 4H), 8.61 (s, 1 aromatic H para to -B(OH)<sub>2</sub>), 8.83 (s, 2 aromatic H ortho to -B(OH)<sub>2</sub>). <sup>13</sup>C-NMR (100 MHz; CDCl<sub>3</sub>): δ 25.58 (CH<sub>2</sub>, 4C), 125.17 (CCO<sub>2</sub>, 2C), 132.16 (CB(OH)<sub>2</sub>, 1C), 141.44 (CHC(B(OH)<sub>2</sub>), 2C), 160.99 (CO<sub>2</sub>, 2C), 170.12 (CH<sub>2</sub>CON, 4C).

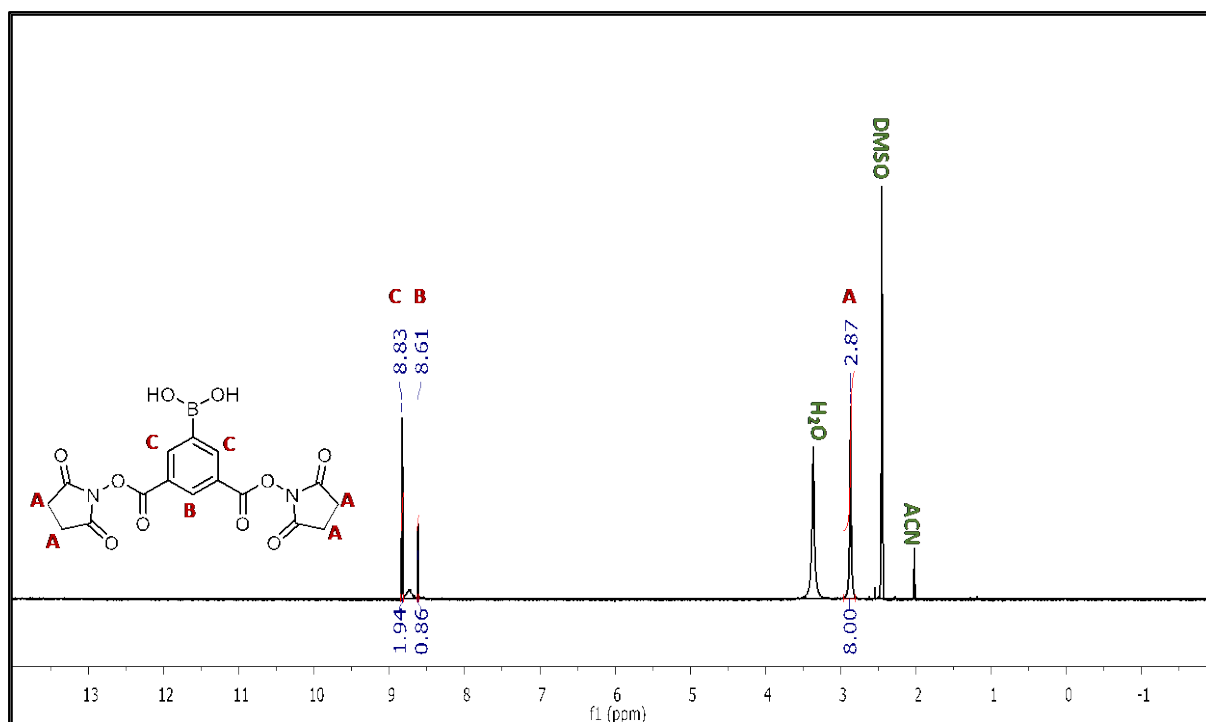


Figure 2.2:  $^1\text{H}$ -NMR for SMT

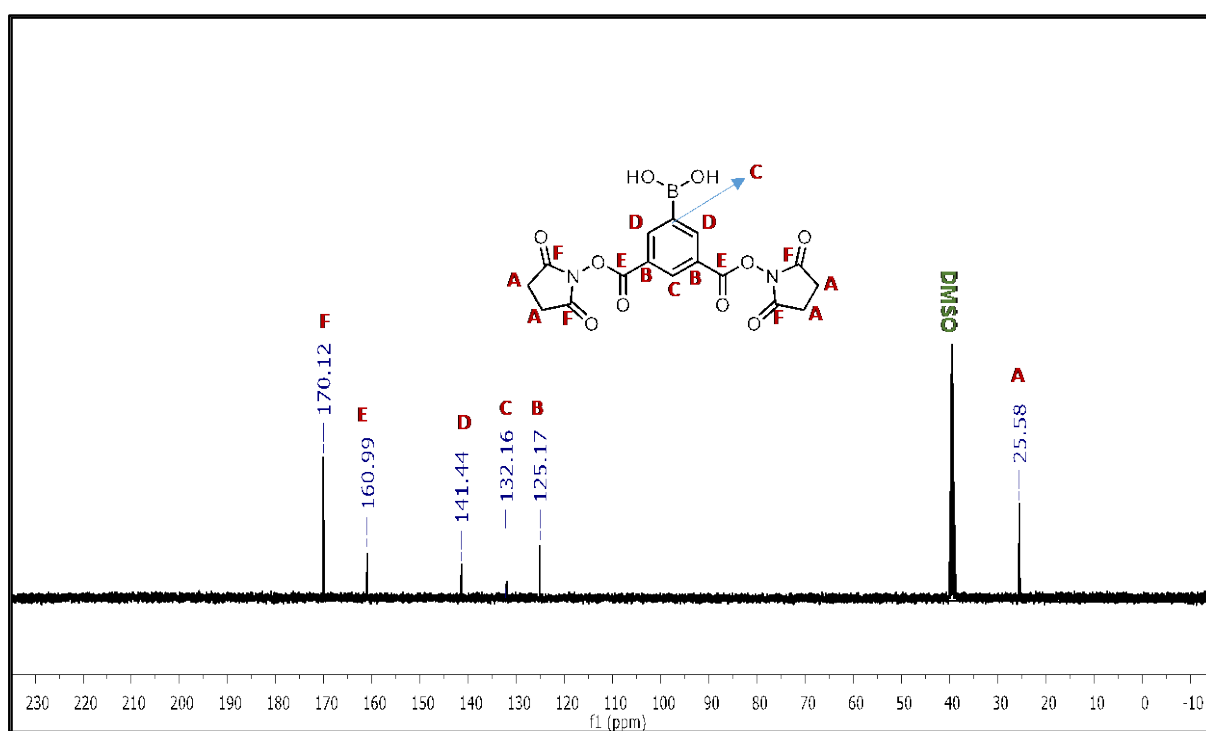


Figure 2.3:  $^{13}\text{C}$ -NMR for SMT

2.5.2 S2: Total Synthesis of (3,5-bis(2-((2,5-dioxopyrrolidin-1-yl)oxy)-2-oxoethyl)phenyl)boronic acid and corresponding intermediates

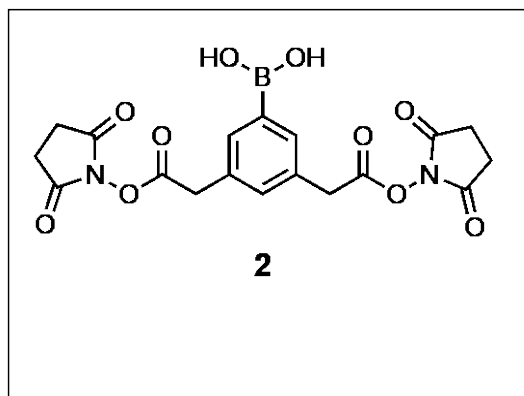


Figure 2.4: Longer cross-linker (LMT)

2.5.2.1 S2 (c1) 1-bromo-3,5-bis(bromomethyl)benzene:

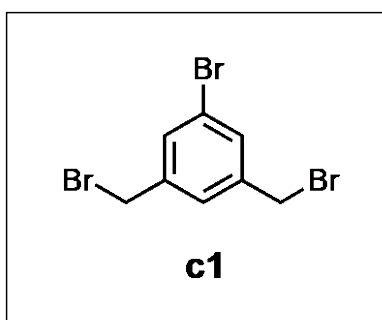


Figure 2.5: 1-bromo-3,5-bis(bromomethyl)benzene

Commercially available 1-bromo-3,5-dimethylbenzene (5 g, 27.01 mmol, 1.0 equiv.) was mixed with NBS (9.13 g, 51.3 mmol, 1.9 equiv.) and catalytic quantity of AIBN (200 mg, 1.21mmol, 4mol%) in a 200 ml single-necked round-bottomed flask. A graduated cylinder is used to measure CH<sub>3</sub>CN (75 ml) and poured into the same reaction vessel. The flask was then fitted with a 24/40 reflux condenser – the top of the reflux condenser is connected to a N<sub>2</sub> Schlenk line. With water running through the condenser, the flask was lowered into a silicone oil bath pre-heated to 140 °C for 3 hours. The completion of reaction was monitored with TLC.

The flask was removed from the oil bath and was allowed to cool in the air until it reached room temperature. The reaction mixture was then subjected to the highest vacuum on rotary evaporation (42 °C, 200-230 mbar) to remove CH<sub>3</sub>CN completely. Then the corresponding solid was heated for 10 minutes in CCl<sub>4</sub> (70 ml) till it is dissolved. The reaction mixture was filtered through a Büchner funnel over a pad of Celite (10 g). The cake is washed with CCl<sub>4</sub> (2 x 10 ml). The filtrate was transferred into a 250 ml round-bottomed flask and concentrated (42 °C, 250-260 mbar) under reduced pressure to remove the residual solvent. The crude residue was purified using flash chromatography using hexane (1000 ml) as eluent to afford the pure product (**c1**) (5.88 g, 64% yield) as a bench stable white flaky solid.

#### 2.5.2.2 S2 (c2) 2,2'-(5-bromo-1,3-phenylene)diacetonitrile

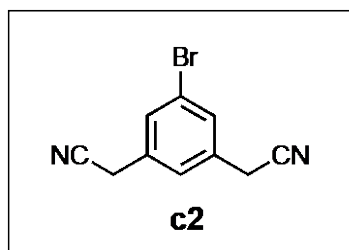


Figure 2.6: 2,2'-(5-bromo-1,3-phenylene)diacetonitrile

2,2'-(5-bromo-1,3-phenylene)diacetonitrile was synthesized in 84% yield. In a typical reaction, (**c1**) (1.5 g, 4.41 mmol, 1.0 equiv.) was treated with EtOH (7 ml) in a 50 ml single-necked round-bottomed flask. NaCN (432.7 mg, 8.82 mmol, 2.0 equiv.) was dissolved in H<sub>2</sub>O (0.75 ml) in a clean vial and then transferred carefully using a plastic pipette to the same flask. The flask was then fitted with a 24/40 reflux condenser – the top of the reflux condenser is kept open to air. With water running through the condenser, the flask was lowered into a silicone oil bath pre-heated to 140 °C for for 30 mins. The completion of reaction was monitored with TLC. The flask was removed from the oil bath and was allowed to cool in the air until it reached

room temperature. The magnetic stir bar was removed and the solution is poured in a 250 ml separatory funnel. The layers were separated and washed sequentially. The aqueous layer was extracted using ethyl acetate (3 x 30 ml) and the combined organic layers were washed with deionized water (2 x 30 ml) and saturated NaCl solution (2 x 20ml) and dried over Na<sub>2</sub>SO<sub>4</sub> (10 g). After, the crude was decanted in a 500 ml single-necked round-bottomed flask and concentrated (42 °C, 180-200 mbar), followed by dry loading using silica gel (2 g) and DCM (20 ml) for flash chromatography. The product eluted with 50% ethyl acetate (250 ml) and after concentration (42 °C, 180-200 mbar) the pure product (**c2**) (0.86 g, 84% yield) appeared as a bench stable white solid.

#### 2.5.2.3 S2 (c3) 2,2'-(5-bromo-1,3-phenylene)diacetic acid

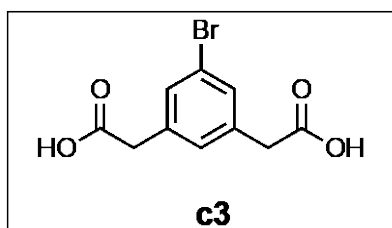


Figure 2.7: 2,2'-(5-bromo-1,3-phenylene)diacetic acid

A 25 ml single-necked round-bottomed flask fitted with a magnetic stirrer was charged with (**c2**) (0.86 g, 3.67 mmol) obtained from the step described above. Next, concentrated HCl (5 ml) was poured carefully using a 10 ml graduate beaker. The flask was then fitted with a 24/40 reflux condenser – the top of the reflux condenser is kept open to air. With water running through the condenser, the flask was lowered into a silicone oil bath pre-heated to 165 °C for 24 hours. The completion of reaction was monitored with TLC. The flask was removed from the oil bath, and was allowed to cool in the air until it reached room temperature. The magnetic stir bar was removed and the solution is poured in a 250 ml separatory funnel. The layers were separated and washed sequentially. The aqueous layer was extracted using ethyl acetate (2 x

30 ml). The combined organic phase were extracted with deionized water (2 x 30 ml) and then saturated NaCl solution (1 x 20ml). The organic phase was then transferred into a 250 ml Erlenmeyer flask, and dried by adding Na<sub>2</sub>SO<sub>4</sub> (10 g). After, the sodium sulphate is removed using decantation and the resulting solution was concentrated by rotary evaporation (42 °C, 180-200 mbar) to give the crude. It was further dried under high vacuo in an oil pump for 30 minutes to give (**c3**) (0.95 g) as a white solid. Without any further purification (**c3**) was completely used up for the next step.

#### 2.5.2.4 S2 (c4) dimethyl 2,2'-(5-bromo-1,3-phenylene)diacetate

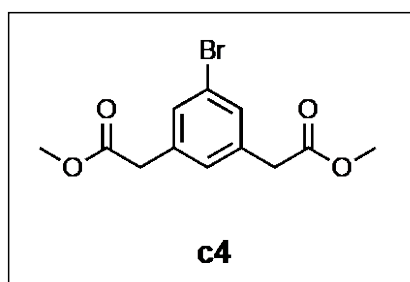


Figure 2.8: dimethyl 2,2'-(5-bromo-1,3-phenylene)diacetate

In a 50 ml centrifuge tube, dry MeOH (45 ml) was obtained from SPS solvent system prior to use. To the same centrifuge tube, concentrated H<sub>2</sub>SO<sub>4</sub> (1 drop) was added using a plastic pipette and was swirled for 30 seconds. This solution was then transferred to a 100 ml single-necked round-bottomed flask containing a magnetic stirrer along with (**c3**) (0.95 g, 3.49 mmol). The flask was then fitted with a 24/40 reflux condenser – the top of the reflux condenser is kept open to air. With water running through the condenser, the flask was lowered into a silicone oil bath pre-heated to 140 °C for 24 hours. The completion of reaction was monitored with TLC. The flask was removed from the oil bath, and was allowed to cool in the air until it reached room temperature. The magnetic stir bar is removed and the solution was poured in a 250 ml separatory funnel. The layers were separated and washed sequentially. The aqueous

layer was extracted using ethyl acetate (2 x 30 ml) and the combined organic layers were washed with deionized water (1 x 30 ml) and saturated NaCl solution (1 x 20ml). The organic phase was then transferred into a 250 ml Erlenmeyer flask, and dried by adding Na<sub>2</sub>SO<sub>4</sub> (10 g). After, the sodium sulphate was removed using decantation and the resulting solution was concentrated by rotary evaporation (42 °C, 180-200 mbar) to afford the crude product (**c4**) (0.99 g) as a fine brown powder. Without any further purification (**c4**) was completely used up for the next step.

2.5.2.4.1 S2 (**c5**) dimethyl 2,2'-(5-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1,3-phenylene)diacetate

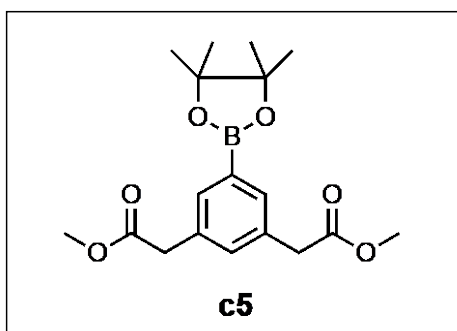


Figure 2.9: dimethyl 2,2'-(5-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1,3-phenylene)diacetate

In a 50 ml single-necked round-bottomed flask, dry ingredients such as (**c4**) (990 mg, 3.64 mmol, 1.0 equiv.), BPin (1016.8 mg, 4.00 mmol, 1.1 equiv.), PdCl<sub>2</sub>(dppf) (59.4 mg, 0.081 mmol, 6 mol%) and KOAc (1071.7 mg, 10.92 mmol, 3.0 equiv.) were added along with a magnetic stirrer. Next, the round-bottomed flask was fitted with a rubber septum and the system was degassed with three vacuum/nitrogen purge cycles. Now extra dry DMF (25 ml) was added to the flask via a syringe, the DMF bottle was degassed and purged with N<sub>2</sub> before usage. Again, the reaction mixture was degassed with three vacuum/nitrogen purge cycles. The flask was lowered into a silicone oil bath pre-heated to 140 °C and was left for stirring overnight

under N<sub>2</sub>, The completion of reaction was monitored with TLC. The flask was removed from the oil bath, and was allowed to cool in the air until it reached room temperature. The magnetic stir bar was removed and the reaction mixture was transferred to a 250 ml separatory funnel using ethyl acetate (1 x 30 ml) to transfer all material out of the flask. The layers were separated and washed sequentially. The aqueous layer was extracted using ethyl acetate (3 x 30 ml). The organic phase was extracted with deionized water (1 x 30 ml) and saturated NaCl solution (1 x 20ml). The organic layers were then transferred into a 250 ml Erlenmeyer flask, and dried by adding Na<sub>2</sub>SO<sub>4</sub> (10 g). After, the sodium sulphate was removed using decantation and the resulting solution is concentrated by rotary evaporation (42 °C, 180-200 mbar) to give the crude which was further dried under high vacuo in an oil pump for 30 minutes to give **(c5)** (1.46 g) as a brown solid. Without any further purification crude **(c5)** was completely used up for the next step.

2.5.2.4.2 S2 (c6) Potassium salt of dimethyl 2,2'-(5-(trifluoro-λ<sup>4</sup>-boraneyl)-1,3-phenylene)diacetate

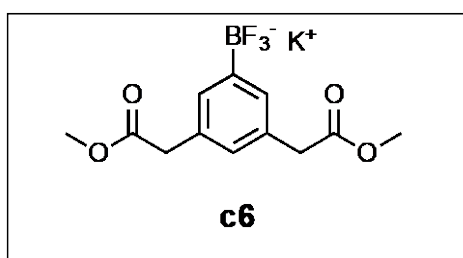


Figure 2.10: Potassium salt of dimethyl 2,2'-(5-(trifluoro-λ<sup>4</sup>-boraneyl)-1,3-phenylene)diacetate

A 50 ml single-necked round-bottomed flask containing a magnetic stirrer was charged with **(c5)** (1.45 g, 4.52 mmol, 1.0 equiv.) and KHF<sub>2</sub> (2122.3 mg, 27.17 mmol, 6.0 equiv.). A graduated cylinder is used to measure MeOH (20 ml) and H<sub>2</sub>O (10 ml), which was then poured into the round-bottomed flask to give a cloudy brown suspension. The flask was then fitted

with a rubber septum and left for room temperature stirring for 45 minutes. After, the round-bottomed flask was subjected to rotary evaporation (42 °C, 300→50 mbar) to remove the methanol-water mixture completely. Next, to the same flask acetone (20 ml) is added and was subjected to heating for 5 minutes at 55 °C. The filtrate was collected into a tared, 50 ml, round-bottomed flask after a hot gravity filtration using a Buchner funnel. The collected acetone was concentrated to around 0.5 ml by rotary evaporation (42 °C, 500-520 mbar) to give a dense brown solution. The final product was afforded as a brown coarse powder by ether precipitation of the above concentrated solution in diethyl ether (35 ml). The solid was further dried under N<sub>2</sub> from Schlenk line to remove any residual solvent. Without any further purification crude (**c6**) was completely used up for the next step.

#### 2.5.2.5 S2 (**c7**) 2,2'-(5-borono-1,3-phenylene)diacetic acid

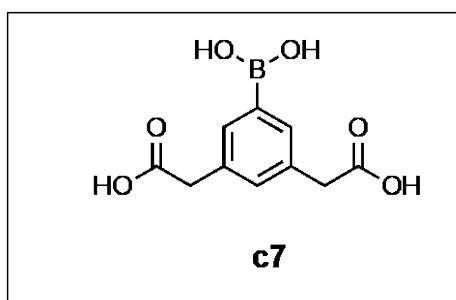


Figure 2.11: 2,2'-(5-borono-1,3-phenylene)diacetic acid

In a typical reaction, dry ingredients (**c6**) (1.46 g, 4.52 mmol, 1.0 equiv.) and LiOH (1082.54 mg, 45.2 mmol, 10.0 equiv.) were charged in a 50 ml single-necked round-bottomed flask also containing a magnetic stirrer. Graduated cylinders were used to measure out acetonitrile (15 ml) and H<sub>2</sub>O (12 ml) and poured into the same round-bottomed flask. The flask was then fitted with a rubber septum and left for room temperature stirring, overnight. The completion of reaction was monitored with Analytical HPLC. Next, the solution was concentrated (42 °C, 500→35 mbar) to remove the acetonitrile-water mixture completely. Hydriion pH paper was

used to check the pH of the leftover suspension. Concentrated HCl (5 drops) were used to adjust to final pH 3 from initial pH 9. The magnetic stir bar was removed and the reaction mixture was transferred to a 250 ml separatory funnel using ethyl acetate (1 x 30 ml) to transfer all material out of the flask. The layers were separated and washed sequentially. The aqueous layer was extracted using ethyl acetate (3 x 30 ml). The organic phase was extracted with deionized water (2 x 30 ml) and saturated NaCl solution (1 x 20ml). The organic layers were then transferred into a 250 ml Erlenmeyer flask, and dried by adding Na<sub>2</sub>SO<sub>4</sub> (10 g). After, the sodium sulphate was removed using decantation and the resulting solution was concentrated by rotary evaporation (42 °C, 180-200 mbar) to give the crude (**c7**) (0.17g) as a light brown solid. Without any further purification crude (**c7**) was completely used up for the next step.

#### 2.5.2.6 S2 (2) (3,5-bis(2-((2,5-dioxocyclopentyl)oxy)-2-oxoethyl)phenyl)boronic acid

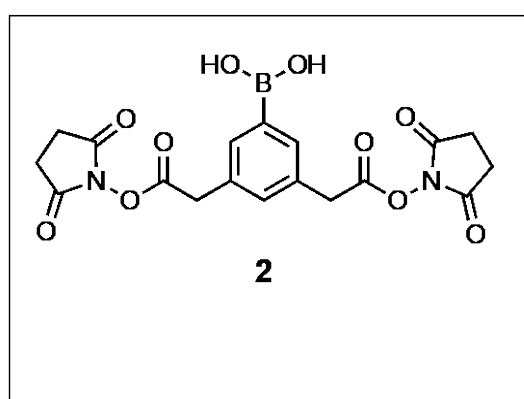


Figure 2.12: (3,5-bis(2-((2,5-dioxocyclopentyl)oxy)-2-oxoethyl)phenyl)boronic acid

A 25 ml, single-necked, round-bottomed flask containing a Teflon coated magnetic stir rod was charged with dry ingredients (**c7**) (170 mg, 0.714 mmol, 1.0 equiv.), followed by DCC (368.35 mg, 1.785 mmol, 2.5 equiv.) and HOSu (205.4 mg, 1.785 mmol, 2.5 equiv.). This dry mixture was degassed with three vacuum/nitrogen purge cycles. Next, extra dry DMF (4.5 ml) was added to the flask via a syringe, the DMF bottle was degassed and purged with nitrogen before usage. Again, the reaction vessel was degassed with three vacuum/nitrogen purge cycles. The

resulting reaction mixture was left for stirring overnight at room temperature under N<sub>2</sub> from Schlenk line, during which the reaction mixture turns cloudy due to the formation of insoluble DCU. After, the reaction mixture was mini-centrifuged to remove the excess precipitate of DCU, followed by manual syringe filtration using minimum amount of dry DMF (4 ml). The mixture was then further concentrated by rotary evaporation (42 °C, 6-10 mbar) to give a white solid. The solid was dissolved in 1:1 HPLC solvents A and B before injecting onto Preparative HPLC (15-70 method, 254 nm detection) column, for final purification of the crude. The product was then freeze-dried under high vacuum for 2 days to get **(2)** (46 mg, 15% yield).

### 2.5.3 S3: Total Synthesis of (1R,2S,3S,5R)-6,6-dimethyl-2-((4-methyl-1H-1,2,3-triazol-1-yl)methyl)bicyclo[3.1.1]heptane-2,3-diol

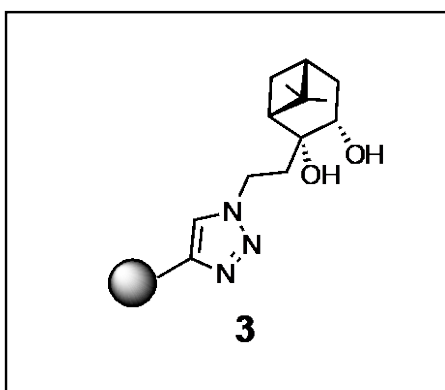


Figure 2.13: Diol resin

#### 2.5.3.1 S3 (b1) 4-nitrophenyl pent-4-ynoate

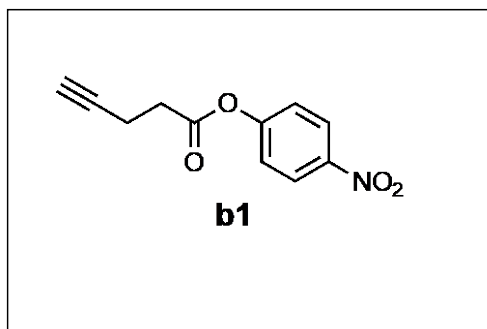


Figure 2.14: 4-nitrophenyl pent-4-ynoate

A acetone dried 100 ml single-necked round-bottomed flask was fitted with a magnetic stirrer and charged with dry ingredients such as 4-pentynoic acid (500 mg, 5.096 mmol, 1.1 equiv.), 2-chloropyridinium iodide (1432.4 mg, 5.606 mmol, 1.1 equiv.), 4-nitrophenol (709.02 mg, 5.096 mmol, 1 equiv.) and  $K_2CO_3$  (1549.75 mg, 11.21 mmol, 2.2 equiv.) as the starting materials. A graduated cylinder was used to measure DCM (40 ml), which was then poured into the round-bottomed flask to give yellow cloudy suspension. The flask was then fitted with a rubber septum and left for room temperature stirring for 4 hours. The completion of reaction was monitored with TLC. The reaction mixture was then subjected to gravity filtration using a Buchner funnel and the filtrate was collected into a tared, 250 ml, round-bottomed flask. Next, dry loading using silica gel (2 g) and DCM (10 ml) for flash chromatography. The product eluted with 30% ethyl acetate in hexane (400 ml) and after concentration (42 °C, 180-200 mbar) the pure product was further dried under high vacuo in an oil pump for 30 minutes to give (**b1**) (0.58 g, 52% yield) as bench stable white crystals.

#### 2.5.3.2 S3 (b2) agarose alkyne beads

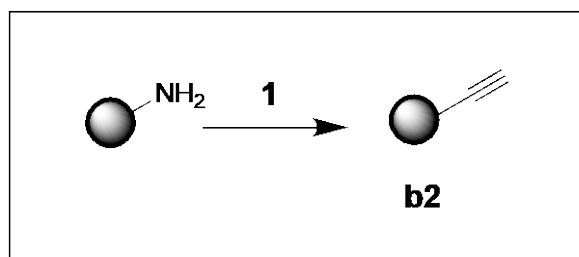


Figure 2.15: Activation of agarose alkyne beads

In a typical chemical reaction, carboxy link gel (8 ml, maximum of 64  $\mu\text{mol}$  of amine groups) was transferred to a disposable frit, allowed to settle for 30 min, and equilibrated with 5 column volumes of pH 4.7 MES buffer (0.1 M). To the gel was added *p*-nitrophenyl ester (**b1**) (143.4 mg, 64 mmol, 16  $\mu\text{mol}/\text{ml}$ ) dissolved in DMF (8 ml) with 5% of the MES buffer (0.4 ml), and the capped fritted tube was rotated gently at room temperature for 2 days. Phosphate buffer (600  $\mu\text{l}$ , pH 7, 1.5 M) was added to the reaction mixture, which was then drained into a clean scintillation vial and the resin was washed with DMF (3 x 5 ml) prior to absorbance measurement. The amount of attached azide was assumed to be equivalent to the amount of released *p*-nitrophenol, which was determined with the aid of an extinction coefficient ( $\epsilon = 432 \text{ M}^{-1} \text{ cm}^{-1}$ ) as reported by the paper. The absorbance was measured on a Cary Bio 100 UV-Vis Spectrometer using DCM as the reference in a cuvette for completing a baseline scan, then followed by the sample collected in the vial. The UV data collected is as follows

## UV-VIS DATA:

### 1. Without Dilution:

Max Wavelength: 429 nm

Absorbance: 3.442, Total volume = 9 ml

Molar extinction Coefficient : 432 M<sup>-1</sup> cm<sup>-1</sup>

On calculating, Concentration= 0.0078 Molar

**Moles= 0.071 mmol / 71 umoles.**

## **2. With Dilution: 10-fold dilution:**

Max Wavelength: 397 nm

Absorbance: 2.391, Total volume = 9 ml

Molar extinction Coefficient : 432

On calculating, Concentration= 0.0498 M

Concentration before dilution= 0.498 M

**Moles= 4.482 umoles.**

2.5.3.3 S3 (b3) (-)-nopol-azide-diol

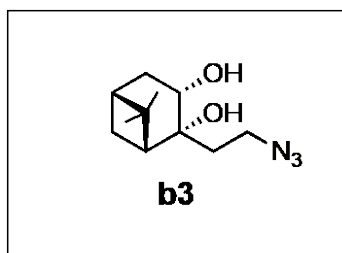
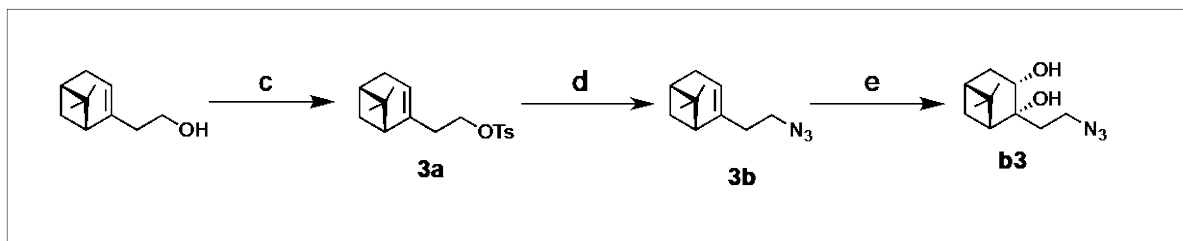


Figure 2.16: (-)-nopol-azide-diol



Scheme 2.4: S3 Total synthesis scheme for (-)-nopol-azide-diol

#### 2.5.3.3.1 S3 (3a) (-)-nopol-tosylate

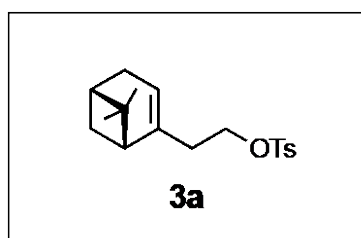


Figure 2.17: (-)-nopol-tosylate

Commercially available (-)-nopol (2.24 g, 13.5 mmol, 1.0 equiv.) was dissolved in dry pyridine (4 ml) in a 25 ml single-necked round-bottomed flask with a magnetic stirrer. The reaction mixture was cooled to 0°C on an ice bath, and to this solution was added p-toluene sulfonyl chloride (2.95 g, 15.5 mmol, 1.1 equiv.) to give a white precipitate. The flask was then fitted with a rubber septum and left for room temperature stirring for 4 hours. The completion of reaction was monitored with TLC. The resulting mixture was placed in a refrigerator for 20 h. Then, the reaction mixture containing a white precipitate was treated with aqueous solution of HCl (10 ml) (prepared from 7.5 mL of 36% aqueous solution of HCl and 15 mL of water). The

magnetic stir bar was removed and the reaction mixture was transferred to a 250 ml separatory funnel using diethyl ether (1 x 50 ml) to transfer all material out of the flask. The layers were separated and washed sequentially. The aqueous layer was extracted using diethyl ether (2 x 50 ml). Combined ethereal layers were then transferred into a 250 ml Erlenmeyer flask, and dried by adding Na<sub>2</sub>SO<sub>4</sub> (10 g). After, the sodium sulphate was removed using decantation and the resulting solution was concentrated by rotary evaporation (42 °C, 572 mbar) to afford (**3a**) (3.95 g, 91% yield) as a yellow oil. Without any further purification crude (**3a**) was used for the next step.

#### 2.5.3.3.2 S3 (**3b**) (-)-nopol-azide

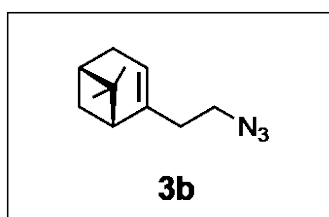


Figure 2.18: (-)-nopol-azide

(-)-nopol-tosyl (**3a**) was immediately used to synthesize (-)-nopol-azide (**3b**). A 100 ml single-necked round-bottomed flask was charged with (**3a**) (2 g, 6.24 mmol, 1.0 equiv.) in DMSO (7.3 ml), and the mixture was stirred using a magnetic stirrer for 10 minutes at room temperature, till sodium azide (0.81 g, 12.48 mmol, 2.0 equiv.) was added. The flask was then fitted with a rubber septum and left for room temperature stirring for 16 hours. The completion of reaction was monitored with TLC. The magnetic stir bar is removed and the reaction mixture was transferred to a 250 ml separatory funnel using deionized water (1 x 50 ml) to transfer all material out of the flask. The layers were separated and washed sequentially. The aqueous layer was extracted using ethyl acetate (3 x 30 ml) and the combined organic layers were washed with deionized water (2 x 50 ml) to remove any leftover DMSO. The organic phase was then

transferred into a 250 ml Erlenmeyer flask, and dried by adding Na<sub>2</sub>SO<sub>4</sub> (10 g). After, the sodium sulphate was removed using decantation and the resulting solution was concentrated by rotary evaporation (42 °C, 250-200 mbar) to afford the crude product (**3b**) (0.99 g) as a transparent oily liquid. Without any further purification (**3b**) was used for the next step.

#### 2.5.3.3.3 S3 (b3) (-)-nopol-azide-diol

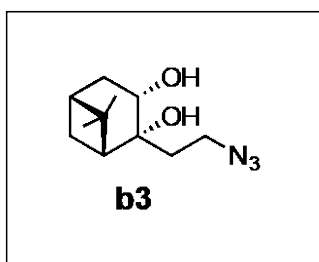


Figure 2.19: (-)-nopol-azide-diol

A 250 ml single-necked round-bottomed flask, containing a magnetic stirrer was charged with NMO 50 wt% in H<sub>2</sub>O (1.27 g, 10.92 mmol, 1.3 equiv.) and pyridine (0.79 ml, 10.08 mmol, 1.2 equiv.) were added to the solution of (**3b**) (1.61 g, 8.4 mmol, 1.0 equiv.) in acetone: water (12:0.9 ml). Lastly, K<sub>2</sub>OsO<sub>4</sub>•2H<sub>2</sub>O (30.9 mg, 0.08 mmol, 2 mol%) was added forming a white oily solution. The flask was then fitted with a 24/40 reflux condenser – the top of the reflux condenser was kept open to air. With water running through the condenser, the flask was lowered into a silicone oil bath pre-heated to 140 °C for 24 hours. The completion of reaction was monitored with TLC. The flask was removed from the oil bath, and was allowed to cool in the air until it reached room temperature. Then, the reaction mixture was concentrated in vacuo and mixed with EtOAc (50 ml), the magnetic stir bar is removed and the solution was poured in a 250 ml separatory funnel. The layers were separated and washed sequentially. The organic phase was washed with HCl (1 × 20 ml, 1 M), distilled water (1 × 20 ml) and brine (1 × 20 ml). The organic layers were then transferred into a 250 ml Erlenmeyer flask, and dried by

adding Na<sub>2</sub>SO<sub>4</sub> (10 g). After, the crude was decanted in a 500 ml single-necked round-bottomed flask and concentrated (42 °C, 180-200 mbar), followed by dry loading using silica gel (2 g) and ethyl acetate (20 ml) for flash chromatography. The product eluted with 30% ethyl acetate (200 ml) and after concentration (42 °C, 180-200 mbar) the pure product (**b3**) (1.08 g, 57% yield) appears as a brown oil.

2.5.3.3.4 S3 (1R,2S,3S,5R)-6,6-dimethyl-2-((4-methyl-1H-1,2,3-triazol-1-yl)methyl)bicyclo[3.1.1]heptane-2,3-diol

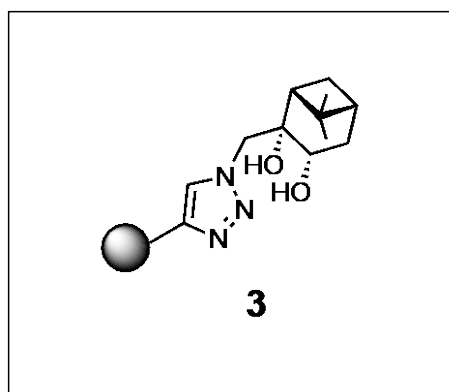


Figure 2.20: 1R,2S,3S,5R)-6,6-dimethyl-2-((4-methyl-1H-1,2,3-triazol-1-yl)methyl)bicyclo[3.1.1]heptane-2,3-diol

In a typical Cu mediated azide-alkyne based click chemistry reaction, a slurry of agarose alkyne beads (**b2**) (8 ml, 0.071 mmol, 1.0 equiv.) was transferred to a disposable frit, allowed to settle for 30 min, and equilibrated with 5 column volumes of DMF. Next, the disposable frit was treated with (**b3**) (128  $\mu$ l, 0.284 mmol, 4.0 equiv.), followed by 2,6-lutidine (121.6 mg, 0.568 mmol, 8.0 equiv.), 2,2'-bipyridine (177.42 mg, 0.568 mmol, 8.0 equiv.), cuprous bromide (81.4 mg, 0.284 mmol, 4.0 equiv.), and finally sodium ascorbate (225 mg, 0.568 mmol, 8.0 equiv.) in DMF (10 ml). The resulting suspension was then capped in a fritted tube and was rotated gently at room temperature for 2 days. The reaction mixture was drained and washed sequentially with approximately 5 column volumes each of DMF, H<sub>2</sub>O, MeOH, 0.1 M aq EDTA, H<sub>2</sub>O, and DMF. The resin was stored in 10mM NaPhos buffer at 4°C.

## 2.6 Automated Peptide Synthesis

Automated peptide synthesis was performed on a Pure Prep Chorus peptide synthesizer. Peptides were synthesized following the general protocol using DMF as solvent, deprotected for 10 min (x 2) in 20 % piperidine/DMF and 0.1 M HOBT. Amino acids were coupled for 2 h at room temperature using excess amino acids (4.0 equiv.), DIEA (8.0 equiv.) and HCTU (4.0 equiv.) as coupling reagent. Rink amide resin was used for the solid phase synthesis. Peptide is precipitated in an ethereal medium using a cleavage cocktail. Cocktail used: 95% TFA + 5% H<sub>2</sub>O + 5% TIPS (5 ml). Peptides after synthesis were purified on preparative HPLC.

### 2.6.1 N-terminal modifications of model peptides

Peptides, after chemical synthesis, have a free amino terminus. N-terminal acetylation is a common N-terminal modification typically done so that the peptide closely mimics the charge state as found in native proteins. Along with N-terminal modification, acylation of the N-terminus using BAs is also described in the following sections.

#### 2.6.1.1 N-terminal Acetylation

Manual coupling of acetic anhydride after the synthesis of the peptide by SPPS. In a 15 ml centrifuge tube, acetic anhydride (47.26  $\mu$ l, 0.1 mmol, 5.0 equiv.), and DIEA (174.19  $\mu$ l, 0.1 mmol, 10.0 equiv.) were taken along with DMF (1.5 ml). Next, it is added to the resin and was subjected to shaking at room temperature for 2 hours. The modified peptide was purified using preparative HPLC (5-45 method,  $t_R$  = 24 min, 220 nm detection).

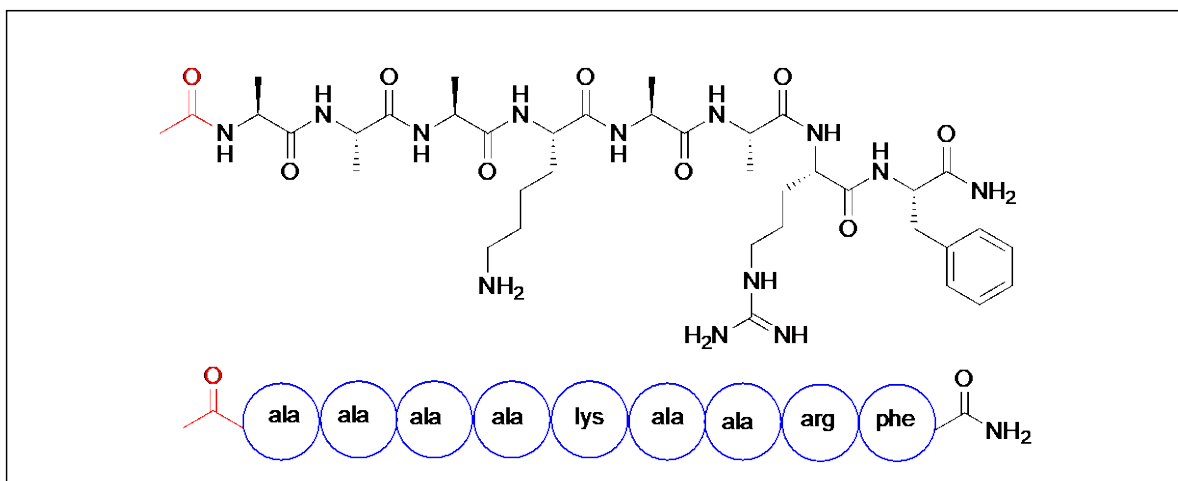


Figure 2.21: N-terminal acetylated model peptide

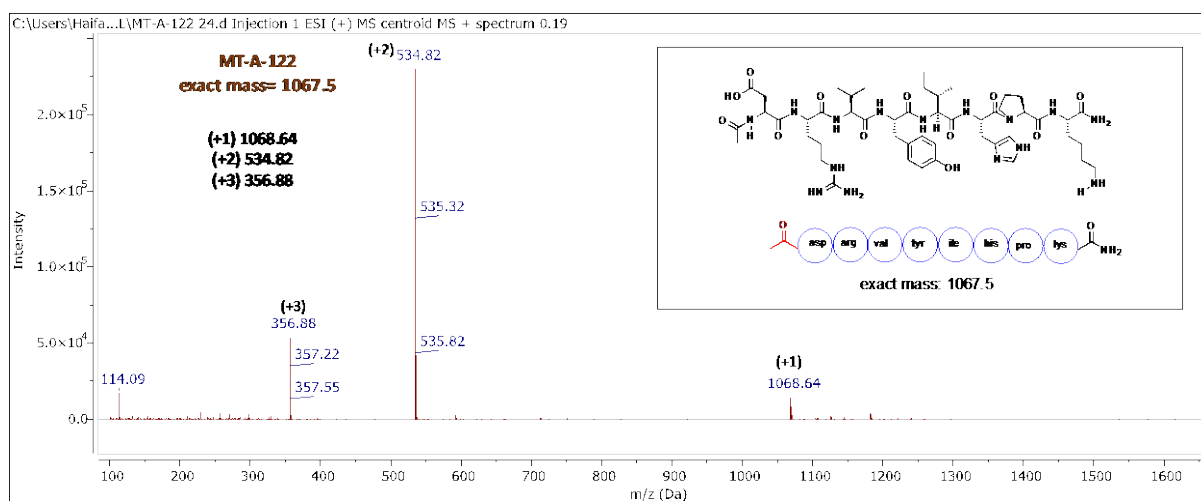


Figure 2.22: MS confirmed, N-terminal acetylated peptide, model peptide, MT-A-122

### 2.6.1.2 Selective N-terminal acylation of peptides using BA

Manual coupling of 4-carboxyphenyl boronic acid is performed after the synthesis of the peptide by SPPS. In a 15 ml centrifuge tube, HCTU (82.7 mg, 0.1 mmol, 2.0 equiv.), 4-carboxyphenyl boronic acid (33.1 mg, 0.1 mmol, 2.0 equiv.) were taken along with DMF (1.5 ml). After, DIEA (70  $\mu$ l, 0.1 mmol, 4.0 equiv.) is added to the resin the mixture was subjected to shaking at room temperature for 2 hours. The modified peptide was purified using Preparative HPLC (5-45 method,  $t_R$  = 26 min, 220 nm detection).

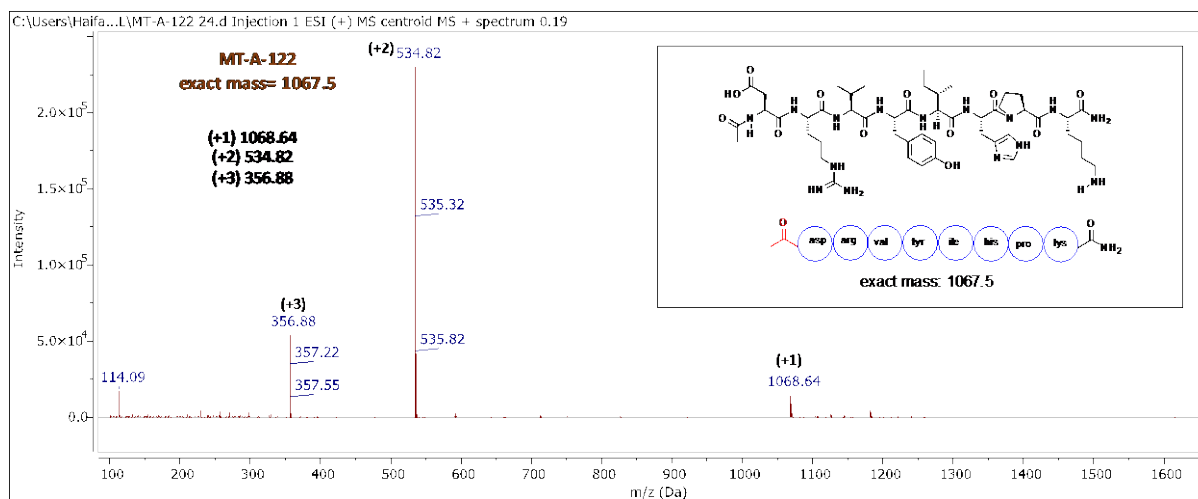


Figure 2.23: MS confirmed, N-terminal BA containing, model peptide, MT-A-122

## 2.6.2 Synthesis of Cross-linked Peptide

In two different 1.5 ml microcentrifuge tube (Fisher brand premium), the peptide (A-A-A-K-A-A-R-F) (5.1 mg, 0.0048 mmol, 2.0 equiv.) and the SMT cross-linker (0.7 mg, 0.0017 mmol, 1.0 equiv.) were weighed out respectively. A 1:1 (100 mM NaPhos:acetonitrile) (1 ml) was made in a 1.5 ml microcentrifuge tube prior to usage. Next, the above solvent system is added to the tube containing the peptide, and subjected to vortex mixing for 30 seconds, before sonication for 2 mins. After, the above reaction mixture is mixed with the tube containing cross-linker, and the solution is incubated for 1 hour at room temperature. The completion of reaction was monitored with analytical HPLC. After, 1 hour of incubation, the sample was diluted to 2 ml using 10mM ammonium bicarbonate buffer, prior to manual syringe filtration. Direct injection was used to inject the sample in semi-prep HPLC (15-70 method, 254 nm detection) C-18 column for the purification of the cross-linked peptide. The product was then freeze-dried under high vacuum for 2 days followed by subsequent analysis using LC/MS.

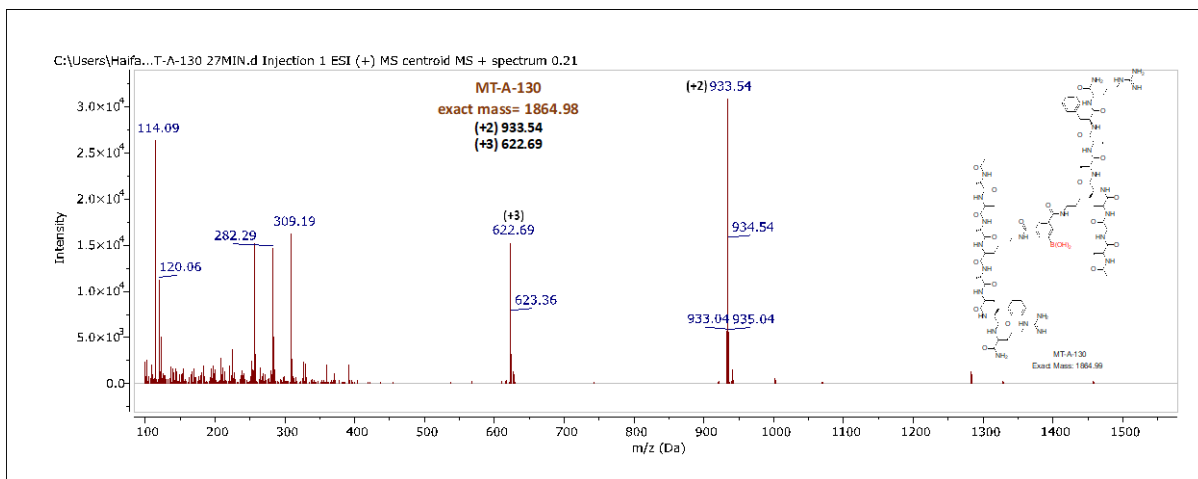


Figure 2.24: MS confirmed cross-linked peptide, MT-A-130

## 2.7 Results and Discussion

The synthesis of SMT was a straightforward 1-step synthesis using commercially available starting materials. The final crude product was purified on preparative HPLC through a reverse phase column. The final yield of the compound was in the range of 50-60%. The synthesis of LMT was an 8-step synthesis using commercially available materials. The synthetic route was optimized several times until the best protocol was achieved for each intermediate step. Most of the intermediates were used as crude without any further purification for the next step. The final compound was HPLC purified. The yield for the final compound was 15%. The intermediates were obtained in good to moderate yields. After HPLC purification, the collected solutions were lyophilized under a high vacuum in a freeze dryer for two days or until dry white powder was observed in the centrifuge tubes. One explanation for low yields for both the final compounds can be attributed to HPLC purification, assuming  $\mu\text{g}$  of compounds getting stuck to the column itself. The other probable reason might be that the intermediates were not purified before conducting the next step, hence decreasing the overall yield.

The synthesis of the diol-containing resin is divided into three parts. The first part involves synthesising the ester, then activating the carboxy link resin and finally, a click

chemistry reaction between the two species to form the rigid diol. The ester was purified on a flash column, and the final yield was 52%. The activation of the link resin continues for 48 hours and involves the conversion of the amine to the alkyne. The solution obtained after draining the frits was analysed for absorbance measurements to determine the conversion in  $\mu$ moles using Beer-Lambert's law. The same amount determined the calculations for the following click chemistry reaction. The click-chemistry reaction between azide and alkyne undergoes a 1,3 cycloaddition mechanism to form the 1,2,3-triazole ring in 48 hours. The synthesized rigid diol resin was stored in 20% ethanol at 4°C as a storage buffer for long-term purposes.

## 2.8 Conclusion

The preceding section described the synthesis of two small non-sterically hindered BA-containing cross-linkers, SMT and LMT, which can efficiently cross-link amino side chain residues in close proximity and a diol-containing resin for AP, for isolating the cross-linked peptides from a mixture of peptides. The diol resin is a rigid cis diol having sub-micromolar dissociation constant values hence, assuring tight covalent bond formation between the BA cross-linkers and the diol present on the resin. Displays the binding scheme for the BA and diol-containing resin, followed by oxidative cleavage using  $H_2O_2$ .

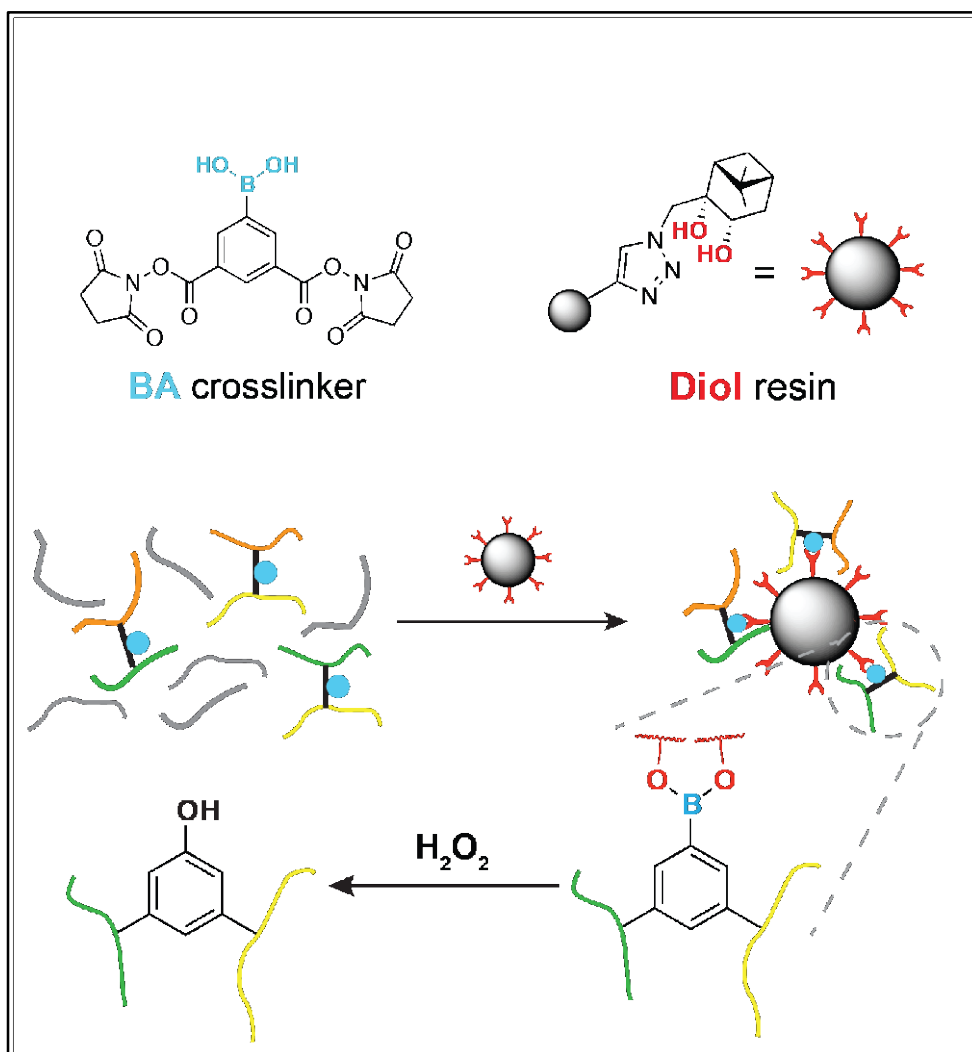


Figure 2.25: Enrichment of BA cross-linked peptides using diol containing resin followed by oxidative cleavage using  $H_2O_2$  Binding scheme.

## CHAPTER 3

# ENRICHMENT OF CROSS-LINKED PEPTIDES WITH MODEL PEPTIDES AND PROTEINS

### 3.1 Introduction

This chapter outlines the experiments to determine whether the diol resin is specific to BA. Before analysing complicated protein samples and cell lysates, the primary goal of this experiment was to assess the enrichment of the diol resin on a set of model linear peptides and model cross-linked peptides. To imitate the actual standard cross-linked peptides that are produced following the proteolytic cleavage of cross-linked proteins, SMT and LMT are used to synthesize model cross-linked peptides. Before examining model proteins, the cross-linking effectiveness is first determined on model peptides. The enrichment of linear peptides, cross-linked peptides, complex protein samples, and cell lysates are briefly discussed in the following subsections.

### 3.2 Peptide-level Enrichment

Automated SPPS was used to synthesize the model linear peptides MT-A-121 and MT-A-122, and the model cross-linked peptide MT-A-130. The N-terminal modified BA-containing peptide, MT-A-121, is used to verify the specificity of the synthesized diol resin, with the N-terminal acetylated linear peptide, MT-A-122, serving as a negative control. MT-A-130 was cross-linked utilizing the SMT and is used as a standard to imitate the real cross-linked peptides produced after the digestion of a cross-linked protein. Next, MT-A-130 and a mixture of different peptides were used to test the resin's enrichment capabilities. All of the

experiments were examined using analytical HPLC, and the protocol for enrichment is described in the preceding section.

### 3.2.1 Enrichment on Model Linear Peptides

The retention times for the BA linear peptide and the peptide serving as the negative control are shown in Figure 3.1 HPLC trace. The resin and the 50 $\mu$ M peptide mixture were enriched for 1 hour. Only the BA peptide is sequestered by the resin after consecutive washing, as shown in Figure 3.2. The negative control peptide is still present in the flowthrough collected after binding. This demonstrates the selectivity of the diol resin for BA over other peptides. Finally, the BA peptide was eluted from the resin beads by oxidative cleavage using H<sub>2</sub>O<sub>2</sub>. As demonstrated in Figure 3.3, the elution is clean with moderate peptide intensity. A change in retention time is due to the difference in the peptide after elution. The boronate ester is cleaved to a phenol during elution.

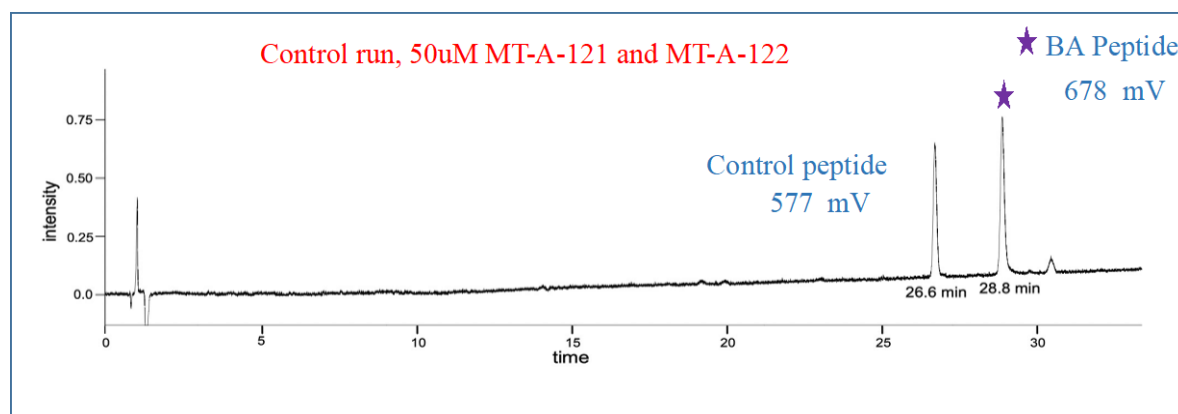


Figure 3.1: Control run, 50 $\mu$ M MT-A-121 and MT-A-122

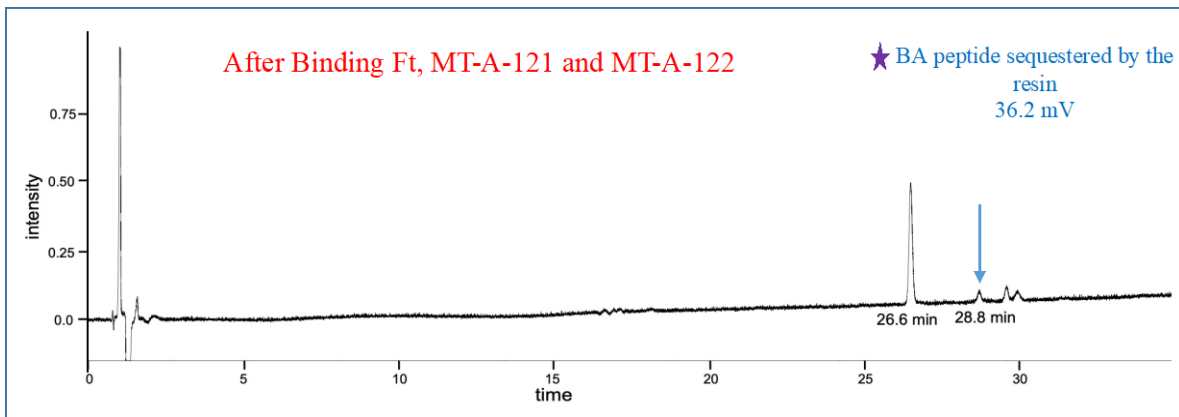


Figure 3.2: After Binding Flowthrough

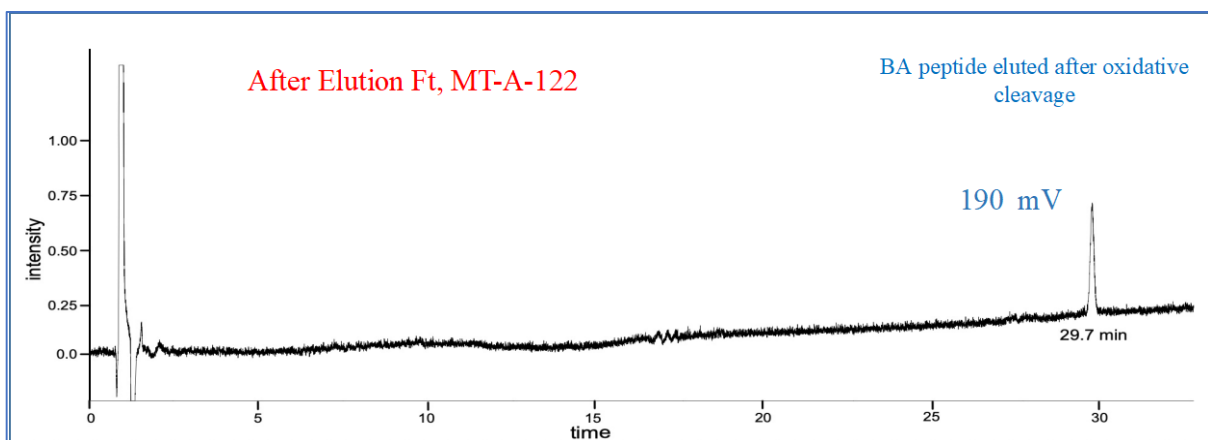


Figure 3.3: After Elution Flowthrough

### 3.2.2 Enrichment on model Cross-linked Peptides

Following the testing of model linear peptides, enrichment experiments were conducted using the cross-linked peptide MT-A-130, which replicates the typical peptides produced after proteolysis. As observed in the control run in Figure 3.4, a mixture of different peptides was added to the cross-linked peptide (MT-A-130) to test the selectivity of the diol resin. The retention time for MT-A-130 is 19.5 mins, and it is apparent from the after-binding flowthrough, (as shown in Figure 3.5) that only one peptide—out of the many present in the mixture—is bound to the resin. All the other peptides were washed out using appropriate buffer solutions and then oxidative cleavage, using  $H_2O_2$ , eluting the BA peptide from the resin, as

seen in Figure 3.6. The results of this experiment show that the resin, even in a peptide mixture, is highly specific for BA.

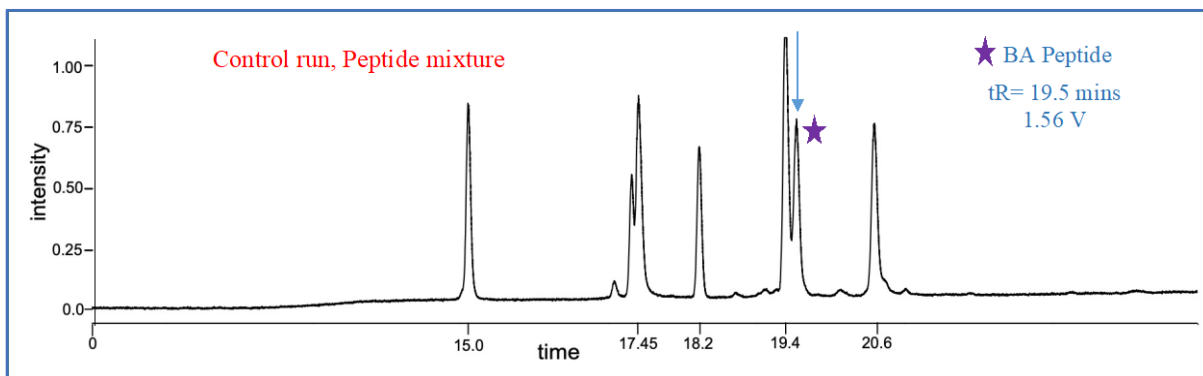


Figure 3.4: Control run, Peptide mixture

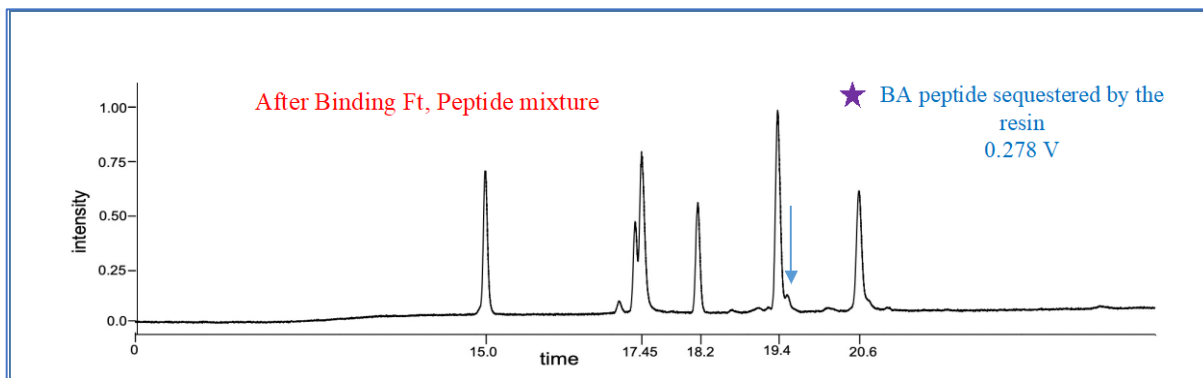


Figure 3.5: After Binding Flowthrough

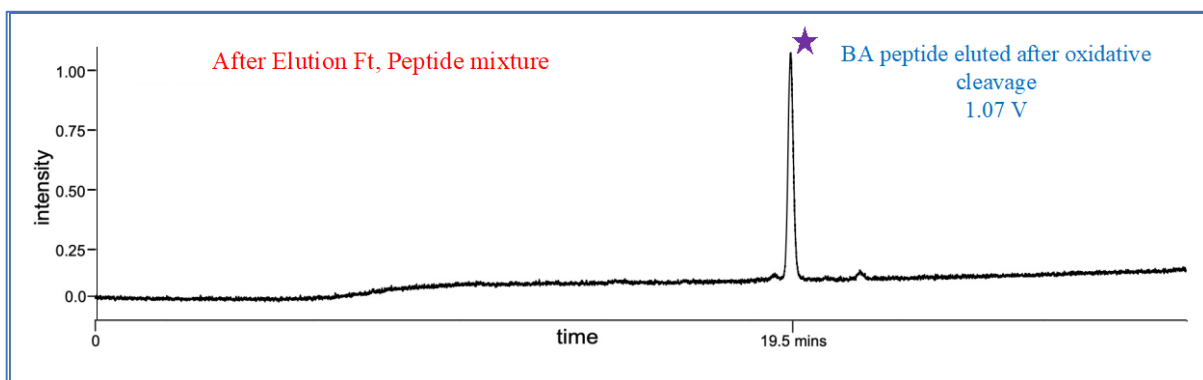


Figure 3.6: After Elution Flowthrough

### 3.3 Optimization of the Enrichment Protocol

The diol resin successfully binds the BA-containing peptides from a mixture of different peptides. No likely side reactions were seen in the HPLC traces during the BA peptide's clean and feasible elution from the resin beads. After it was established that the enrichment process works and is specific towards BA, it was further refined to achieve the best results. The effect of washing the resin with gentle and harsh buffers, different H<sub>2</sub>O<sub>2</sub> concentrations for oxidative cleavage, the amount of methionine oxidation caused by H<sub>2</sub>O<sub>2</sub>, and finally, the amount of resin to carry out enrichment are some of the various parameters that required optimization. In this sense, it was determined that all experiments should follow a single, reliable enrichment methodology. The following sections describe these experiments in detail.

### 3.3.1 Effect of Wash

Following binding, various washing conditions and washing procedures were tested. After the binding stage, the goal of the washing is to eliminate any undesirable linear, mono-linked, and non-specific interactions. Typically, the wash buffers can be gentle washing solutions like HEPES or NaPhos wash, or they can even be harsh washing agents like urea. This experiment was carried out to check the stability of the standard mimic cross-linked mimic peptide MT-A-130 in vigorous washing conditions. Typically, while analysing small peptides, harsh washing conditions can be avoided. However, while working with complex proteins, harsh elution conditions like urea wash are mandatory to remove the undesired linear, mono-linked, and non-specific types of interactions. It can be seen from Figure 3.8 and Figure 3.9 that there is not much of a change in the intensity levels or the peptide's retention time following a mild or vigorous wash when compared to the control run, Figure 3.7 MT-A-130, is resistant to abrasive washing and therefore, can be utilized as a standard mimic for enrichment studies.

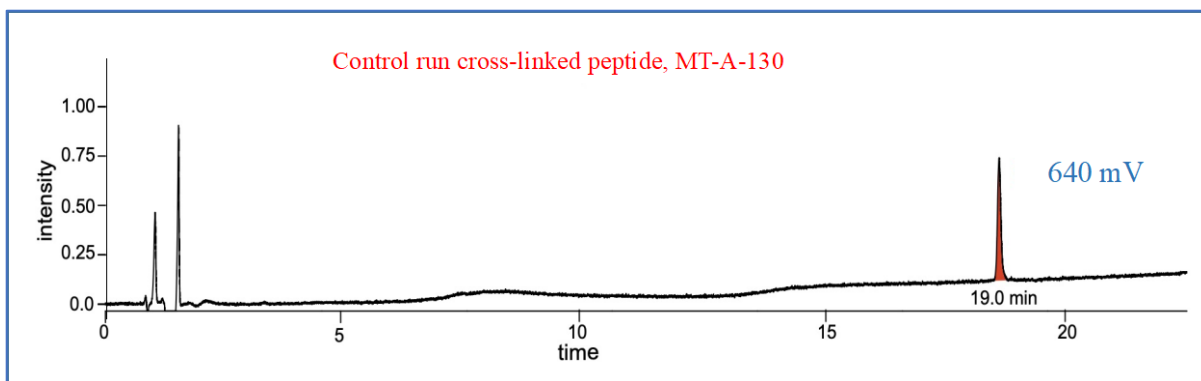


Figure 3.7: Control run, 50 $\mu$ M MT-A-130

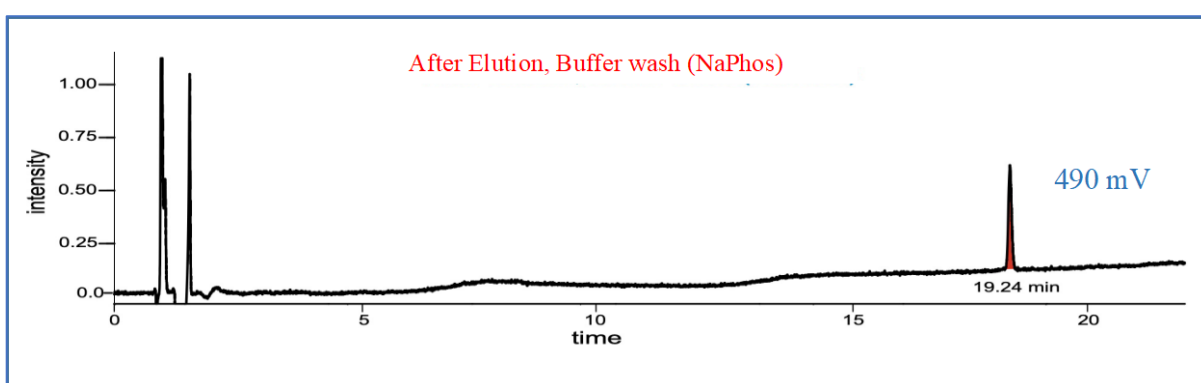


Figure 3.8: Buffer wash

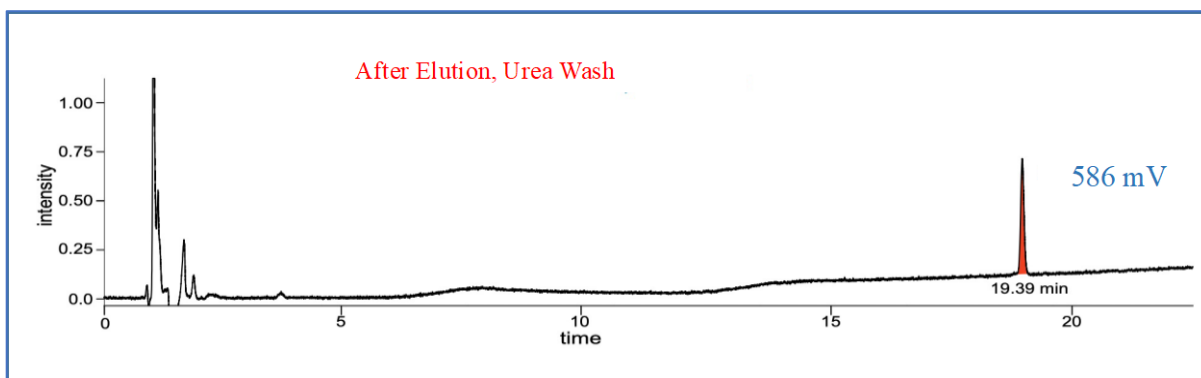


Figure 3.9: Urea Wash

### 3.3.2 H<sub>2</sub>O<sub>2</sub> Oxidation for Elution

BAs form reversible covalent bonds with the diol resin to form boronate ester. The undesired peptides are then removed using the appropriate buffers, and the desired peptide is subsequently released from the resin beads using H<sub>2</sub>O<sub>2</sub>'s oxidative cleavage reaction. The

oxidation goes by a mechanism that modifies the ester bond, resulting in a peptide with a hydroxyl group. Several experiments were conducted to determine the ideal  $\text{H}_2\text{O}_2$  concentration. Figure 3.11 shows that  $\text{H}_2\text{O}_2$  concentrations as low as 5mM prepared in suitable buffers can produce satisfactory results.

In contrast, Figure 3.12 shows that 100mM is typically of high concentration and can substantially decrease intensity. As a result, 10 minutes of shaking is adequate to elute the peptide with 5mM  $\text{H}_2\text{O}_2$ . High quantities of  $\text{H}_2\text{O}_2$  can cause methionine to oxidize into methionine sulfoxide species. Therefore, the subsequent series of experiments were run to monitor the oxidation of methionine in the presence of  $\text{H}_2\text{O}_2$ .

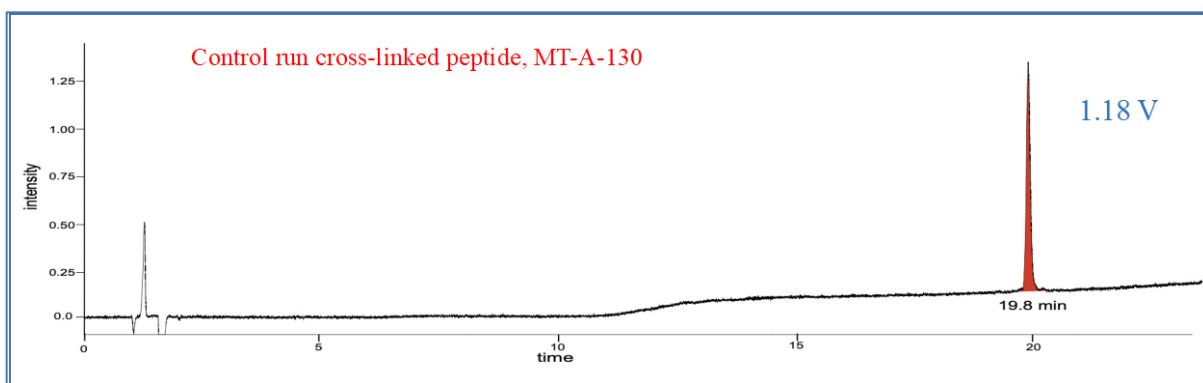


Figure 3.10: Control run, 50 $\mu\text{M}$  MT-A-130

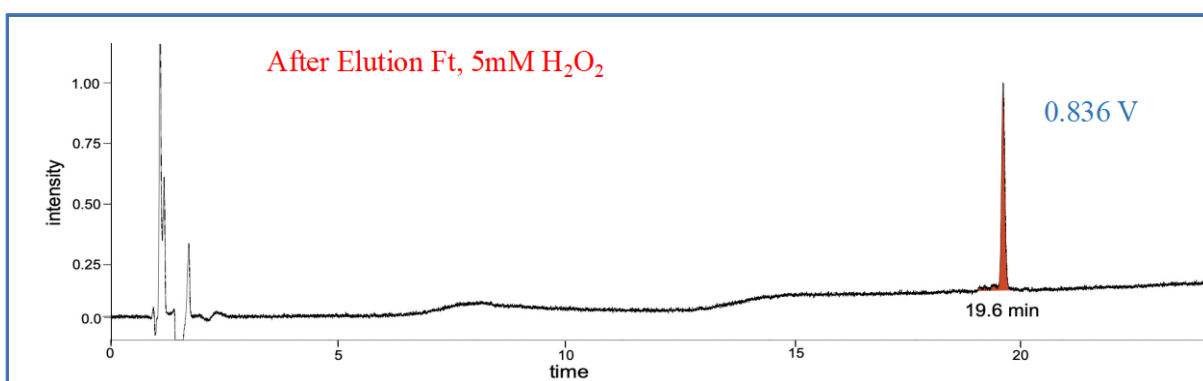


Figure 3.11: Elution using 5mM  $\text{H}_2\text{O}_2$

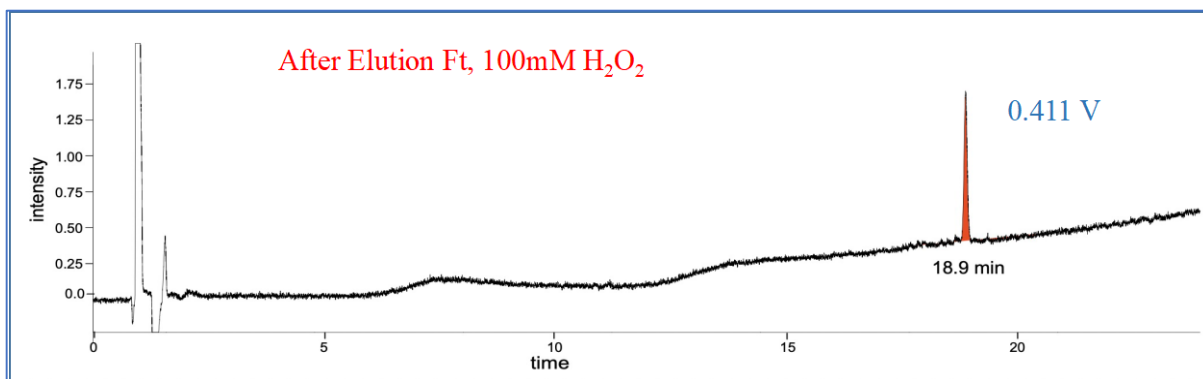


Figure 3.12: Elution using 100mM H<sub>2</sub>O<sub>2</sub>

### 3.3.3 Methionine Oxidation

Typically methionine or cysteine residues get oxidized in the presence of strong oxidising agents such as H<sub>2</sub>O<sub>2</sub>. However, using a low concentration of H<sub>2</sub>O<sub>2</sub> has a negligible effect on the methionine residues. To evaluate the effects of oxidation, a model peptide containing methionine residue was synthesized Figure 3.13, and it was then treated with various concentrations of H<sub>2</sub>O<sub>2</sub> for different periods. Figure 3.14 shows the results of elution with 5 mM H<sub>2</sub>O<sub>2</sub> for 10 minutes and the minimal degree of oxidation that occurred. The amount of met oxidation becomes significant as shaking time increases, and this also causes a loss in peptide intensity, as illustrated in Figure 3.15. As shown in Figure 3.16, even 10 minutes with a high concentration of H<sub>2</sub>O<sub>2</sub>, (100 mM) can cause substantial oxidation. Hence, 5mM H<sub>2</sub>O<sub>2</sub> while shaking for 10 mins can result in oxidative cleavage of the peptide with negligible methionine oxidation.

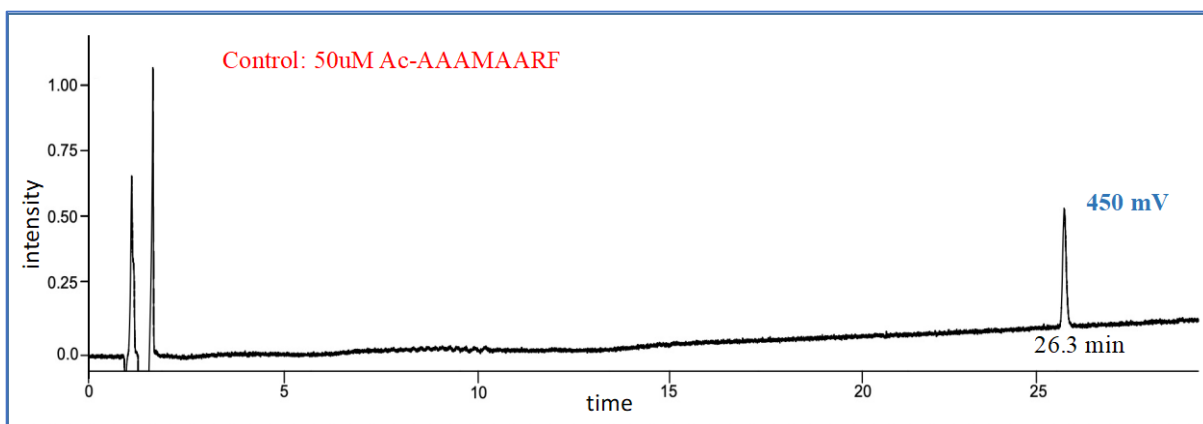


Figure 3.13: Control run, 50 $\mu$ M of Ac-AAAMAARF

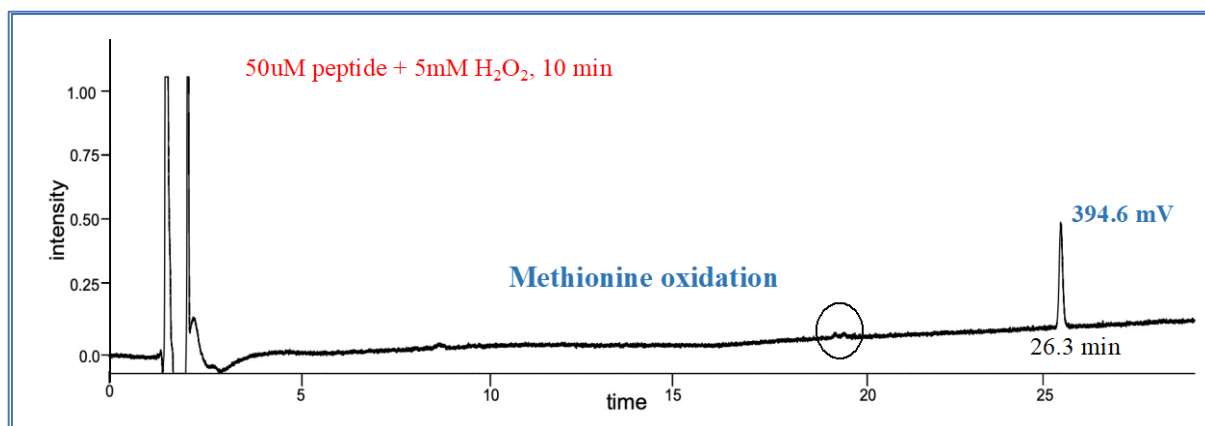


Figure 3.14: Elution using 5mM H<sub>2</sub>O<sub>2</sub>, 10 min shaking

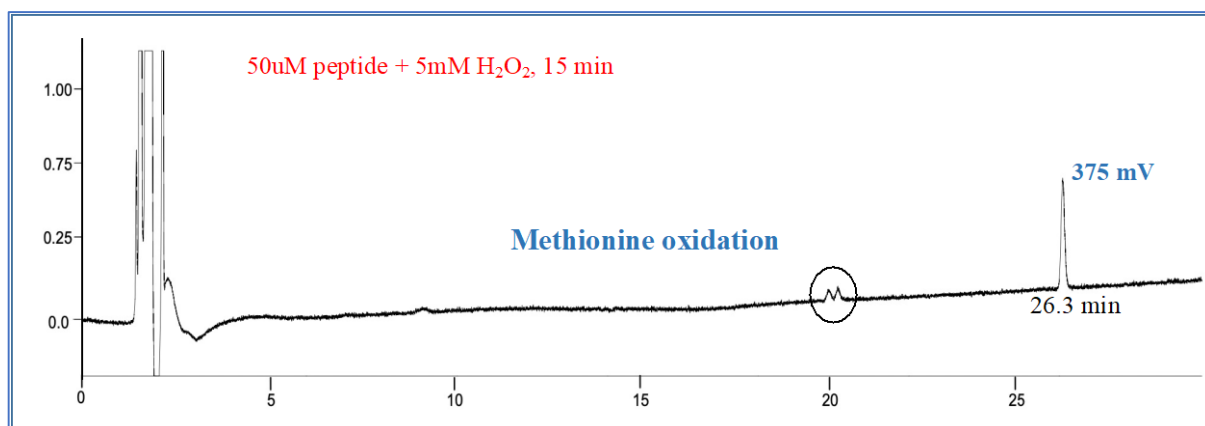


Figure 3.15: Elution using 5mM H<sub>2</sub>O<sub>2</sub>, 15 min shaking

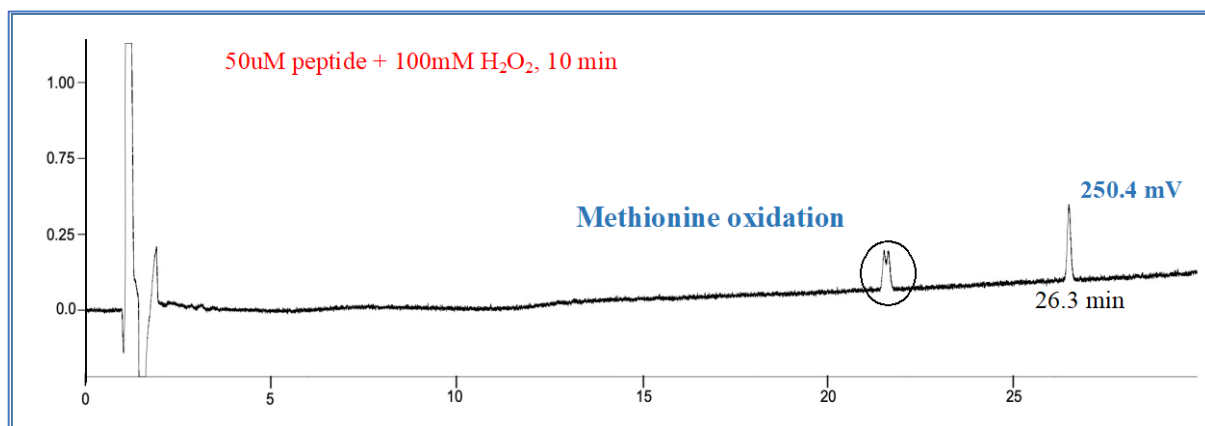


Figure 3.16: Elution using 100mM H<sub>2</sub>O<sub>2</sub>, 10 min shaking

### 3.3.4 Amount of resin

Several experiments were conducted with different amounts of resin under the same conditions to find the optimal amount for enrichment. The resin was used in quantities of 100, 50, 25, and 10 $\mu$ L, respectively with 50 $\mu$ M of MT-A-130 was used in the enrichment experiment. As can be observed in Figure 3.17, the intensity for the control run is 784 milli-volts (mV). In contrast to 50 $\mu$ L and 25 $\mu$ L resin, which still retains some peptide after binding flowthroughs as seen in Figure 3.19 and Figure 3.20, 100 $\mu$ L of resin can bind the complete peptide yielding 0 mV of intensity in the flowthrough Figure 3.18. While Figure 3.21 shows that 10 $\mu$ L of resin is insufficient to bind the peptide since the flowthrough has an intensity of about 158 mV. According to the HPLC traces, more resin results in better cross-linked peptide binding. 25 $\mu$ L resin is practical and ideal for conducting binding tests for analysing small protein/peptide samples.

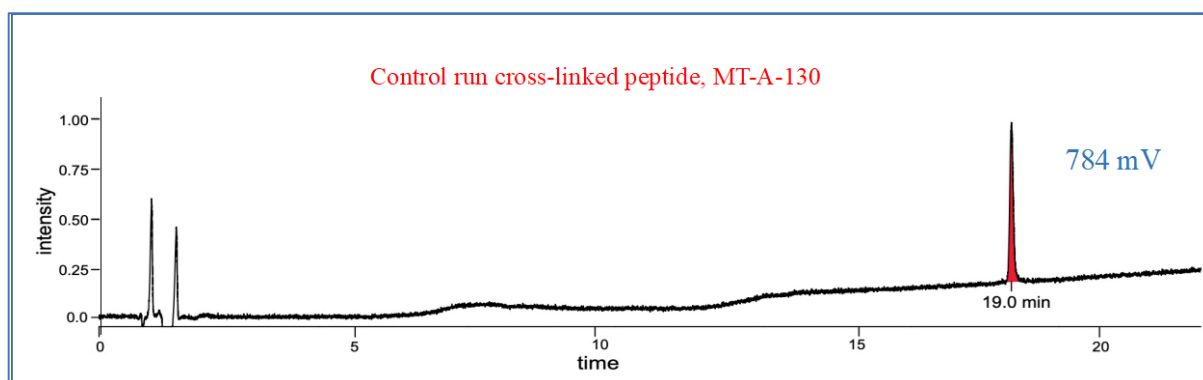


Figure 3.17: Control run, 50 $\mu$ M MT-A-130

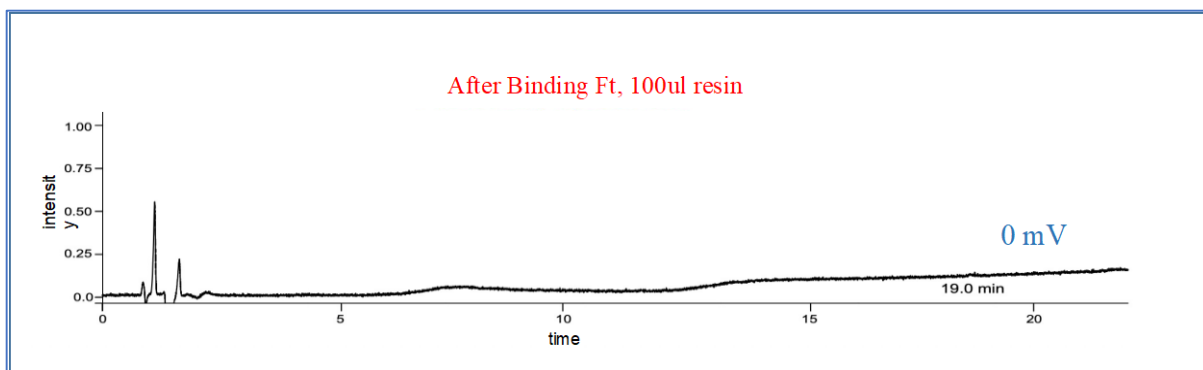


Figure 3.18: 100 $\mu$ L resin

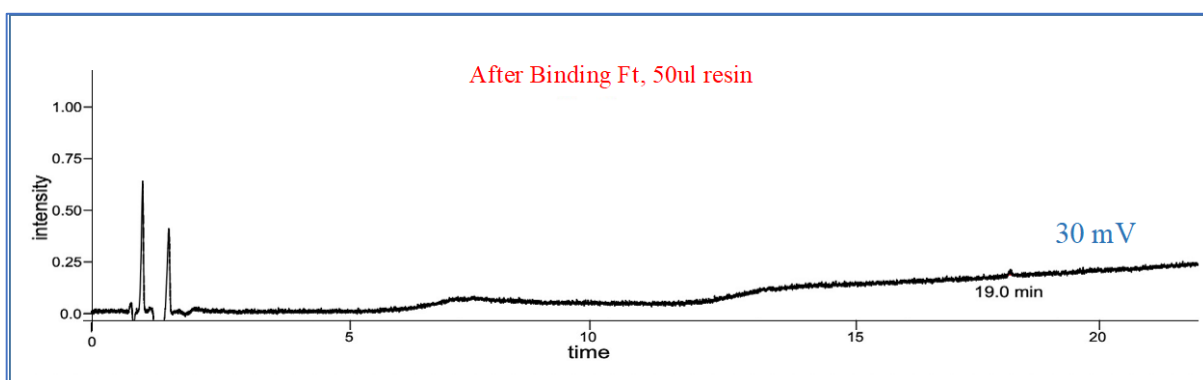


Figure 3.19: 50 $\mu$ L resin

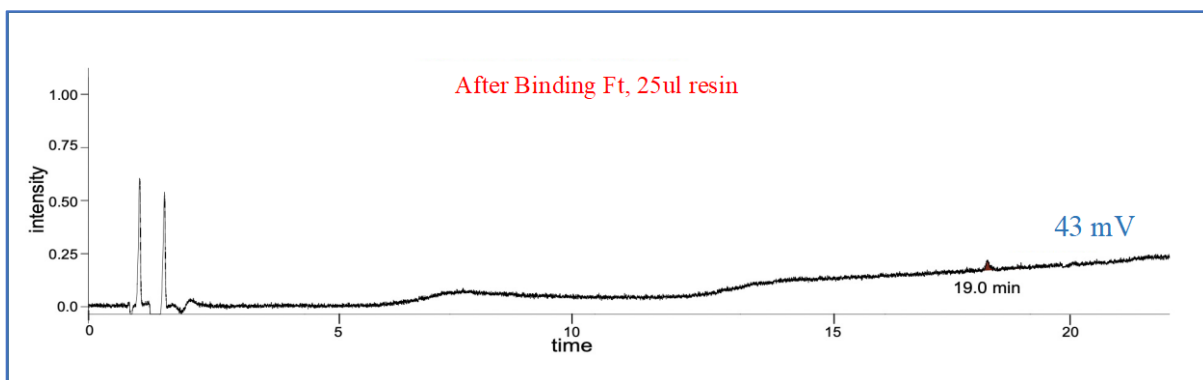


Figure 3.20: 25 $\mu$ L resin

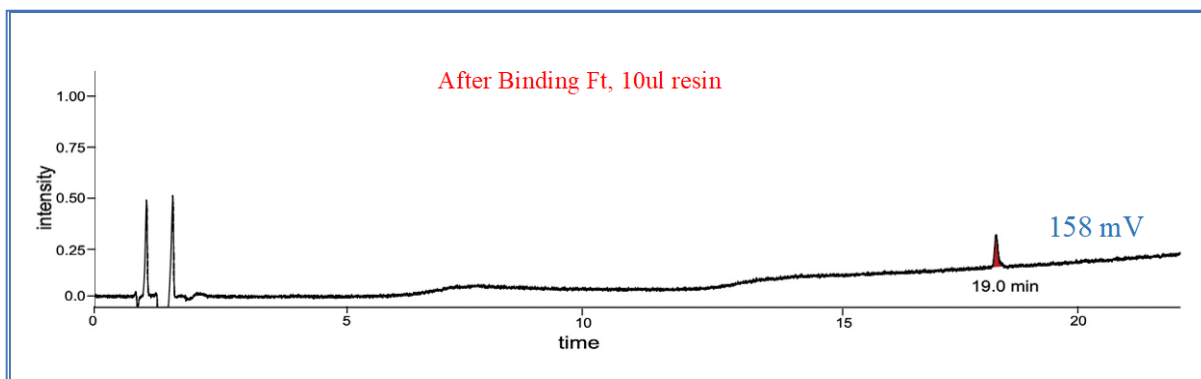


Figure 3.21: 10µL resin

### 3.4 Affinity Purification

This section describes the final optimized protocol for enriching cross-linked peptides before LC/MS-MS analysis. After optimization of the parameters, the following protocol gives maximum efficiency. A summary of the four significant steps involved in enrichment are described in Figure 3.22. The next section describes the protocol in detail.

#### 3.4.1 Protocol for Enrichment

Diol resin (suspended in equal volume of 20% ethanol) was resuspended in the storage buffer using a pipette tip. Once resuspended 1:1, 50 µl of diol resin was obtained in a 2.0 ml spin-X centrifuge tube filter and subjected to spin on the micro-centrifuge at  $3000 \times g$  for 30 seconds to remove storage solution. Next, 200 µl of equilibration buffer (10mM NaPhos pH 7.4), was added to the filter tube, shaken on the vortex shaker at 400 rpm at room temperature for 5 mins. After spinning the tube at  $3000 \times g$  for 30 seconds the washing solution is removed. This step was repeated thrice. Next, 100 µl of binding solution (peptide mixture in 10 mM NaPhos pH 7.4), is added to the equilibrated resin and the tube is shaken at 600 rpm at room temperature for 1 hour. Again, the samples was spun at  $3000 \times g$  for 60 seconds and the flowthrough was collected. Next, 200 µl of wash buffer (10mM NaPhos pH 7.4), was added to the filter tube, and then spun at  $3000 \times g$  for 60 seconds, to remove the wash buffer which can

be discarded. This step is repeated thrice. Finally for oxidative cleavage, elution buffer (5 mM H<sub>2</sub>O<sub>2</sub> in 10 mM NaPhos pH 7.4) was freshly prepared and 100 µl is added to the resin and shaken at 800 rpm at room temperature for 10 mins. Spin at 3000× g for 60 seconds. Collect the eluate. The viability of the resin was monitored using an analytical HPLC. Three traces are usually obtained for any particular experiment performed. Firstly, the control was used to determine the exact retention time for the peptide binding onto the resin. Second, the flowthrough after binding was used to ascertain whether that the resin was able to successfully sequesters the BA-containing peptide. Finally, the flowthrough after oxidative cleavage was used to demonstrates that the peptide can be recovered after the experiment.

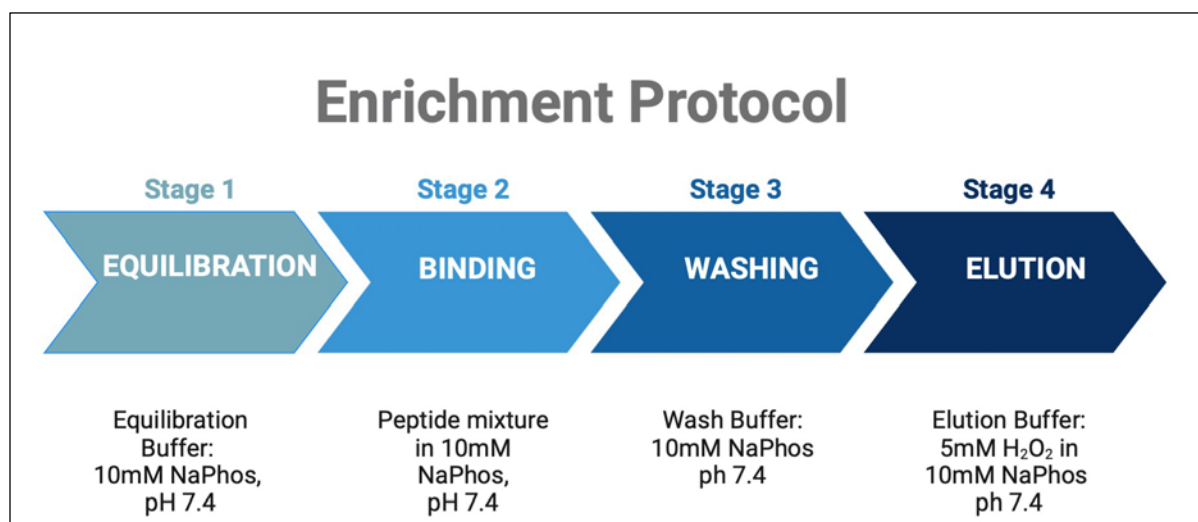


Figure 3.22: Steps involved in Cross-linked Peptide Enrichment.

### 3.5 Protein-level Enrichment

SMT and LMT's cross-linking efficiency was evaluated using the model protein 14-3-3. To observe the cross-linking between the protein and the cross-linkers, SDS-PAGE gels are utilized. Enrichment of the SMT and LMT cross-linked peptides was tested on BSA and 14-3-3, respectively. To verify the cell permeability of SMT, *in vitro* cross-linking was done on HeLa cell lysates.

### 3.5.1 Cross-linking on a Model Protein

Model protein 14-3-3 is used to tailor the optimum concentration and the time required for cross-linking SMT and LMT cross-linkers *in vitro*. Figure 3.23 and Figure 3.24 illustrate that monomeric bands at 25 kilodaltons (kDa) disappear and dimeric covalent cross-linked bands appear at 70 kDa, corresponding to effective cross-linking with SMT and LMT, respectively. The negative control, which implies no cross-linking, exhibits no presence at the higher molecular weight. With 1mM of SMT, sufficient cross-linking was seen after an hour. While in the case of LMT, a similar degree of cross-linking was obtained with 2mM of LMT in 1 hour. Higher molecular weight oligomeric bands, around 150 kDa, appear with LMT if the cross-linking time is increased. Therefore, 2mM of LMT and 1mM of SMT can express significant dimeric cross-links within 1 hour. The optimized concentration and time for cross-linking are used for further studies to identify cross-linked peptides by MS analysis.

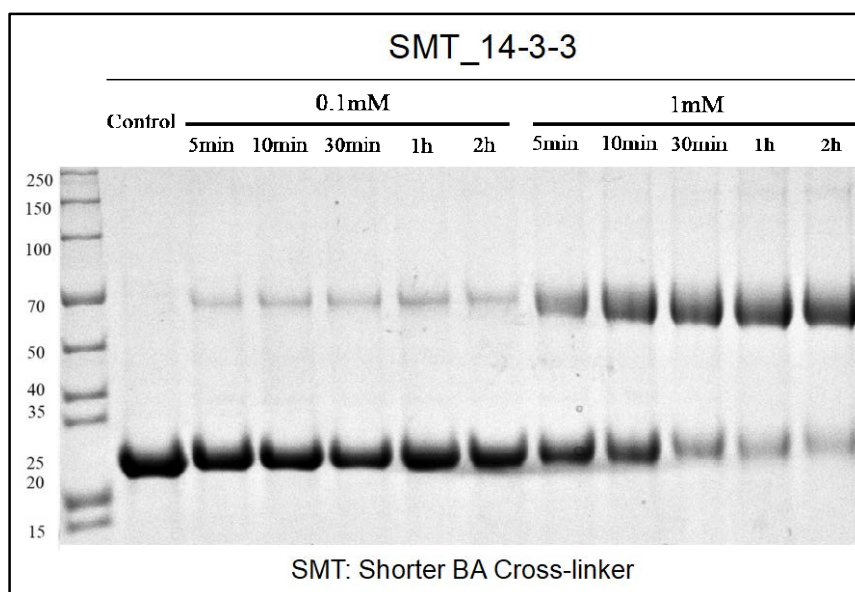


Figure 3.23: Cross-linking using SMT on model protein 14-3-3

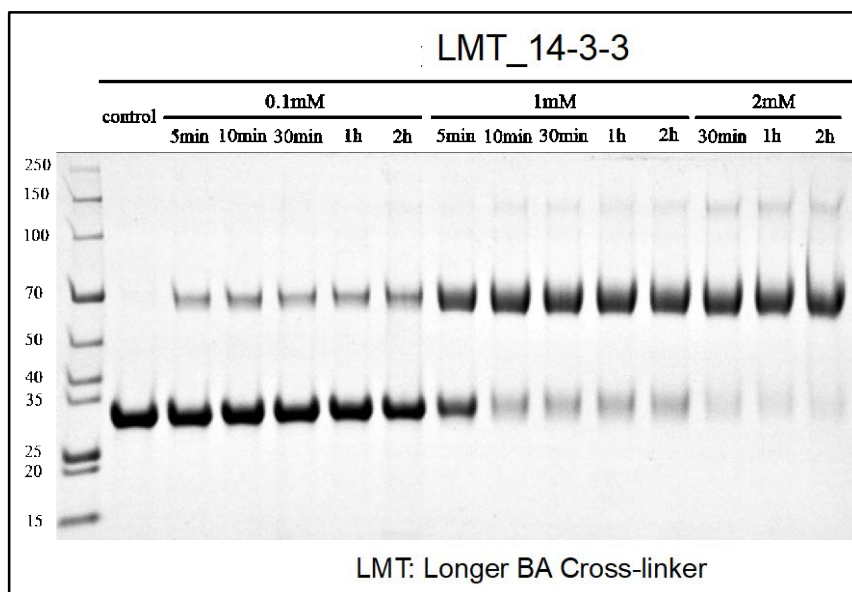


Figure 3.24: Cross-linking using LMT on model protein 14-3-3

### 3.5.2 Enrichment on Model proteins

The enrichment of cross-linked peptides was studied using the model proteins 14-3-3 and BSA. 50 $\mu$ g of both model proteins were cross-linked using the conditions above. Proteins were then reduced, alkylated, and finally digested by trypsin. The protein is digested into a mixture of cross-linked, mono-linked, linear, and other peptides. This mixture is intended to enhance the less prevalent cross-linked peptides. Following enrichment, samples were analysed by LC/MS-MS analysis, and Plink2 analysed the MS raw data to identify cross-linked peptides. After enriching SMT cross-linked peptides, we observed an 80–90% fold increase in the number of unique cross-links for both 14-3-3 and BSA model proteins (Figure 3.25a). Similarly, there is a considerable increase in the number of LMT cross-links after enrichment, indicating high enrichment efficiency for the model proteins 14-3-3 and BSA, as shown in Figure 3.25a.

The next objective was to evaluate the enrichment approach using complex mixtures, such as *E.coli* tryptic peptides. To mimic a complex environment, 10 $\mu$ g of 14-3-3 cross-linked peptides were spiked into 1 milli-gram (mg) of *E.coli* tryptic peptides (100-fold excess).

Typically, various peptide mixtures make it challenging to identify cross-links in complex systems. After enriching SMT and LMT cross-linked peptides, there is a significant increase in cross-links, as seen in Figure 3.25b. After enrichment with SMT, the cross-links notably rose from two to twenty two. Sixteen unique cross-links were observed after LMT cross-linked peptides were enriched in the complex mixture, in contrast to no cross-links before enrichment. The number of cross-linked peptides after enrichment increased significantly in a single HPLC analysis suggesting high enrichment efficiency.

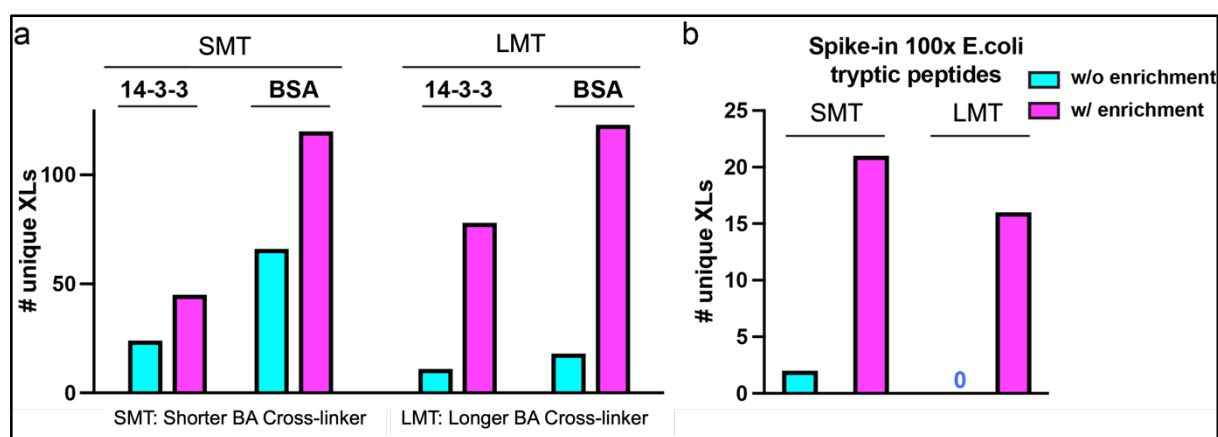


Figure 3.25: a) Enrichment of SMT and LMT on 14-3-3 and BSA  
b) Spiking cross-linked 14-3-3 peptides in *E. coli* tryptic peptides

### 3.5.3 Cross-linking and Enrichment of HeLa Cell Lysate

Cross-linking of SMT was tried on HeLa cell lysate using the optimized concentrations and time. Next, the cross-linked cell was subjected to lysis using a mild buffer in presence of some detergent before reduction, alkylation and trypsin digestion. The mixture of peptides obtained after digestion were then enriched using the diol-containing resin and only 14.2% of the total identified peptides are regular linear peptides, as seen in Table 4 . Good recovery of cross-linked peptides in a single LC-MS run without further fractionation, proved efficient enrichment. Only 3% of the cross-links are inter-protein cross-links. Several PPIs were identified but the top hits include Hsp90, ribosomal complexes and eF2.

Table 4: Identification of peptide pairs under 5% FDR control

Types	Counts	Percentage
1. Cross-linked peptides	2081	12.3%
2. Loop-linked peptides	3589	21.2%
3. Mono-linked peptides	8841	52.2%
4. Regular peptides	2410	14.2%

### 3.5.4 Live-cell cross-linking using SMT in live HeLa cells

To test the cell permeability of SMT, live HeLa cells were treated with SMT. Figure 3.26 shows the cell permeability of DSS, which is employed in this experiment as a positive control. After cross-linking, low molecular weight bands disappear, and expression is visible at higher molecular weights. The sharp bands at low molecular weight sharp bands become a smear following cross-linking and may also form oligomers, which is why they occur at higher kDa levels. Another likely explanation could be intracellular or exogenous cross-linking of DSS with other interacting proteins. Different concentrations of SMT were used to cross-link. SMT's cell permeability is demonstrated by the absence of bands at low kDa values and strong bands that indicate cross-linking. As a result, SMT can be used to successfully to conduct *in vivo* cell cross-linking assays successfully

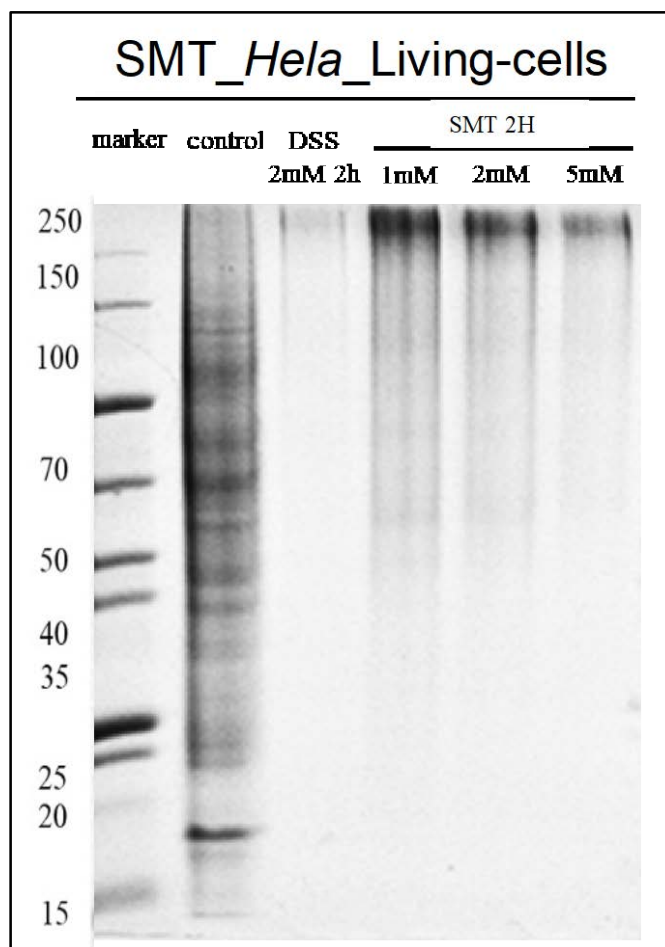


Figure 3.26: Live-cell cross-linking of SMT in HeLa cells

### 3.6 Results and Discussion

Several experiments were carried out on model peptides, complex proteins, and cell lysates, and it was found that the BA-based enrichment is quite specific and efficient. Analysis carried out using model peptides demonstrated the feasibility of the enrichment protocol. It also confirmed that the diol resin is particular in binding the BA-containing peptides, even in a mixture. Experiments on MT-A-130 confirmed that enrichment is possible on cross-linked peptides. Negligible methionine oxidation was observed while using 5mM H<sub>2</sub>O<sub>2</sub> as compared to higher concentrations of 100mM. After oxidative cleavage using H<sub>2</sub>O<sub>2</sub>, the boronate ester is cleaved to phenols after the oxidative cleavage, reflecting changes in the retention times of

the peak. The phenols are expected to elute later compared to the boronate ester counterpart. The peak intensity is measured in terms of volts or milli-volts, and the retention time is reported in seconds.

### 3.7 Conclusion

One of the primary objectives of proteomics today is to comprehend how protein-protein interactions (PPIs) are structured as a complex network. Much PPI data has recently become available due to high-throughput experimental techniques, offering general a idea of how proteins interact in biological systems. XL-MS has become a robust technique for analysing protein-protein interactions due to its ability to simultaneously capture PPI from their natural environment and reveal their physical interaction contacts. XL-MS is exceptional among MS-based techniques and facilitates the determination of both the identity and connectivity of protein-protein interactions in cells. Here we presented the design and synthetic schemes for two novel cross-linkers, SMT and LMT, which are functionalized with a non-sterically hindered boronic acid handle on the spacer arm. The small and non-bulky design avoids non-specific interactions and enables cell permeability. High cross-linking efficiency is achieved with 1mM SMT and 2mM LMT within one hour of incubation time with complex protein samples such as 14-3-3. Moreover, the positively charged handle makes the cross-linkers cell-permeable, promoting live cell cross-linking on whole cell lysates such as HeLa cells for *in vivo* assays.

The new cross-linkers have a BA backbone, which is amenable to one-step affinity purification for the enrichment of cross-linked peptides before LC-MS analysis. BAs can form reversible solid covalent bonds with rigid diols; therefore, we also designed and synthesized a diol-containing resin for affinity purification. The diol can form tight covalent bonds with BA. The aryl BA and the diol resin form a tetracyclic adduct. We have also described an optimized

protocol for feasible and robust enrichment. SMT and LMT were cross-linked with complex protein samples, and then the digested cross-linked peptides were enriched, and a 90% increase in the number of unique cross-links was observed. After initial screening, it was noted that the diol resin was highly specific towards BA-containing peptides. The elution of the cross-linked peptides from the resin beads requires low concentrations of H<sub>2</sub>O<sub>2</sub> for the oxidative cleavage-yielding, unmodified peptides. Methionine oxidation is not noticeable at such low H<sub>2</sub>O<sub>2</sub> concentrations. However, the current software has options enabling docking analysis that can automatically take methionine oxidation to methionine sulfoxide into account. Thus, it can be concluded that BA enrichment is extremely sensitive and specific.

### 3.8 Future Directions

There is always space for improvement. MS cleavage is not possible with SMT or LMT cross-linkers, but more work is being done to develop future iterations with an MS cleavable handle. The MS-cleavable handles in the future versions will facilitate MS2 and MS3 fragmentation for more detailed analysis and simple cleavage. More advancements and research are being pursued to understand live cell cross-linking for *in vivo* assays fully. Although the current cross-linking efficiency is remarkably high, amino acid side chains show interlinking amongst themselves, leading to mono-linked and other products. Likewise, the enrichment methodology is being further optimized for a straightforward and quick workflow that leverages standardized approaches for the enrichment of cross-linked peptides.

## REFERENCES

## REFERENCES

- (1) Powell, T.; Callahan, A.; Leonard, H.; *Nutrition: Science and Everyday Application*, 2<sup>nd</sup> ed.; Pressbooks, 2020.
- (2) O'Connor, C.; Jill U, A.; M. A., J. U. *Essentials of Cell Biology*, 1<sup>st</sup> ed.; NPG Education: MA: 2010.
- (3) Sleator, R. D. Prediction of Protein Functions. Kaufmann, M., Klinger, C. (eds) *Functional Genomics. Methods in Molecular Biology*,. **2012**, vol 815, 15-24, online.
- (4) Bruce Alberts, A. J., Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. Molecular Biology of the Cell. *Biochemistry and Molecular Biology Education*. **2006**, 4th edition, 212 - 214.
- (5) Rao, V. S.; Srinivas, K.; Sujini, G. N.; Kumar, G. N. Protein-protein interaction detection: methods and analysis. *Int J Proteomics* **2014**, 2014, 147648.
- (6) De Las Rivas, J.; Fontanillo, C. Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput Biol* **2010**, 6 (6), e1000807.
- (7) Goebels, F. Classification of protein protein interactions. Technical University of Munich, Weihenstephan science center for nutrition, land use and the environment, 2014.
- (8) Kaake, R. M.; Wang, X.; Burke, A.; Yu, C.; Kandur, W.; Yang, Y.; Novtisky, E. J.; Second, T.; Duan, J.; Kao, A.; et al. A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells. *Mol Cell Proteomics* **2014**, 13 (12), 3533-3543.
- (9) blog, C. p. Brief Introduction of Protein-Protein Interaction (PPI). creative proteomics blog: online, 2018; Vol. 2018.
- (10) Acuner Ozbabacan, S. E.; Engin, H. B.; Gursoy, A.; Keskin, O. Transient protein-protein interactions. *Protein Eng Des Sel* **2011**, 24 (9), 635-648.
- (11) Lu, H.; Zhou, Q.; He, J.; Jiang, Z.; Peng, C.; Tong, R.; Shi, J. Recent advances in the development of protein-protein interactions modulators: mechanisms and clinical trials. *Signal Transduct Target Ther* **2020**, 5 (1), 213.
- (12) Wiederschain, G. Y. Protein-Protein Interactions. Methods and Applications (Fu, H. (ed.) in *Methods in Molecular Biology* (Walker, J., ed.), Vol. 261, Humana Press, Humana, 2004, 532 p., \$125). *Biochemistry (Moscow)* **2005**, 70 (5), 608-608.
- (13) Snider, J.; Kotlyar, M.; Saraon, P.; Yao, Z.; Jurisica, I.; Stagljar, I. Fundamentals of protein interaction network mapping. *Mol Syst Biol* **2015**, 11 (12), 848.
- (14) Wu, Y.; Li, Q.; Chen, X. Z. Detecting protein-protein interactions by Far western blotting. *Nat Protoc* **2007**, 2 (12), 3278-3284.

- (15) Kaiser, G. 7.26A: Mapping Protein-Protein Interactions. In *BOOK: MICROBIOLOGY*, Kaiser, G. Ed.; LibreTexts, 2021.
- (16) Yan, Y.; Marriott, G. Analysis of protein interactions using fluorescence technologies. *Curr Opin Chem Biol* **2003**, *7* (5), 635-640.
- (17) McHaourab, H. S.; Lin, Y. L.; Spiller, B. W. Crystal structure of an activated variant of small heat shock protein Hsp16.5. *Biochemistry* **2012**, *51* (25), 5105-5112.
- (18) Wheat, A.; Yu, C.; Wang, X.; Burke, A. M.; Chemmama, I. E.; Kaake, R. M.; Baker, P.; Rychnovsky, S. D.; Yang, J.; Huang, L. Protein interaction landscapes revealed by advanced in vivo cross-linking-mass spectrometry. *Proc Natl Acad Sci U S A* **2021**, *118* (32).
- (19) Kim, E. D.; Sabharwal, A.; Vetta, A. R.; Blanchette, M. Predicting direct protein interactions from affinity purification mass spectrometry data. *Algorithms Mol Biol* **2010**, *5*, 34.
- (20) Roux, K. J.; Kim, D. I.; Raida, M.; Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **2012**, *196* (6), 801-810.
- (21) Branon, T. C.; Bosch, J. A.; Sanchez, A. D.; Udeshi, N. D.; Svinkina, T.; Carr, S. A.; Feldman, J. L.; Perrimon, N.; Ting, A. Y. Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **2018**, *36* (9), 880-887.
- (22) Liu, X.; Salokas, K.; Weldatsadik, R. G.; Gawriyski, L.; Varjosalo, M. Combined proximity labeling and affinity purification-mass spectrometry workflow for mapping and visualizing protein interaction networks. *Nat Protoc* **2020**, *15* (10), 3182-3211.
- (23) Nguyen, T. M. T.; Kim, J.; Doan, T. T.; Lee, M. W.; Lee, M. APEX Proximity Labeling as a Versatile Tool for Biological Research. *Biochemistry* **2020**, *59* (3), 260-269.
- (24) Rhee, H. W.; Zou, P.; Udeshi, N. D.; Martell, J. D.; Mootha, V. K.; Carr, S. A.; Ting, A. Y. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **2013**, *339* (6125), 1328-1331.
- (25) Xu, Y.; Fan, X.; Hu, Y. In vivo interactome profiling by enzyme-catalyzed proximity labeling. *Cell Biosci* **2021**, *11* (1), 27.
- (26) Richards, A. L.; Eckhardt, M.; Krogan, N. J. Mass spectrometry-based protein-protein interaction networks for the study of human diseases. *Mol Syst Biol* **2021**, *17* (1), e8792.
- (27) Zuzow, N.; Ghosh, A.; Leonard, M.; Liao, J.; Yang, B.; Bennett, E. J. Mapping the mammalian ribosome quality control complex interactome using proximity labeling approaches. *Mol Biol Cell* **2018**, *29* (10), 1258-1269.
- (28) O'Reilly, F. J.; Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat Struct Mol Biol* **2018**, *25* (11), 1000-1008.

- (29) Piersimoni, L.; Kastritis, P. L.; Arlt, C.; Sinz, A. Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein-Protein Interactions horizontal line A Method for All Seasons. *Chem Rev* **2022**, *122* (8), 7500-7531.
- (30) Tan, D.; Li, Q.; Zhang, M. J.; Liu, C.; Ma, C.; Zhang, P.; Ding, Y. H.; Fan, S. B.; Tao, L.; Yang, B.; et al. Trifunctional cross-linker for mapping protein-protein interaction networks and comparing protein conformational states. *Elife* **2016**, *5*.
- (31) Steigenberger, B.; Pieters, R. J.; Heck, A. J. R.; Scheltema, R. A. PhoX: An IMAC-Enrichable Cross-Linking Reagent. *ACS Cent Sci* **2019**, *5* (9), 1514-1522.
- (32) Bode, M.; Woellhaf, M. W.; Bohnert, M.; van der Laan, M.; Sommer, F.; Jung, M.; Zimmermann, R.; Schroda, M.; Herrmann, J. M. Redox-regulated dynamic interplay between Cox19 and the copper-binding protein Cox11 in the intermembrane space of mitochondria facilitates biogenesis of cytochrome c oxidase. *Mol Biol Cell* **2015**, *26* (13), 2385-2401.
- (33) Rappsilber, J.; Siniosoglou, S.; Hurt, E. C.; Mann, M. A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal Chem* **2000**, *72* (2), 267-275.
- (34) Profacgen. Protein/Peptide Crosslinking. Profacgen: online.
- (35) Heck, T.; Faccio, G.; Richter, M.; Thony-Meyer, L. Enzyme-catalyzed protein crosslinking. *Appl Microbiol Biotechnol* **2013**, *97* (2), 461-475.
- (36) Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S. Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat Methods* **2005**, *2* (3), 201-206.
- (37) Kilkenny, M. L.; De Piccoli, G.; Perera, R. L.; Labib, K.; Pellegrini, L. A conserved motif in the C-terminal tail of DNA polymerase alpha tethers primase to the eukaryotic replisome. *J Biol Chem* **2012**, *287* (28), 23740-23747.
- (38) Belsom, A.; Rappsilber, J. Anatomy of a crosslinker. *Curr Opin Chem Biol* **2021**, *60*, 39-46.
- (39) Man, T. P. 8 Factors to Consider when Selecting a Protein Cross-linker. In *A Discussion of Protein Research*, blog, T. p. m. s., Ed.; G-Biosciences: online, 2019; Vol. 2019.
- (40) Scientific, T. *Crosslinking Technical Handbook*; Thermo Fisher Scientific
- (41) Matzinger, M.; Mechtler, K. Cleavable Cross-Linkers and Mass Spectrometry for the Ultimate Task of Profiling Protein-Protein Interaction Networks in Vivo. *J Proteome Res* **2021**, *20* (1), 78-93.
- (42) Man, T. p. High Efficiency & Stability Protein CrossLinking with EDC & NHS. In *A Discussion of Protein Research*, |, T. P. M. s. B., Ed.; G-Biosciences online, 2017; Vol. 2017.
- (43) Chen, F.; Nielsen, S.; Zenobi, R. Understanding chemical reactivity for homo- and heterobifunctional protein cross-linking agents. *J Mass Spectrom* **2013**, *48* (7), 807-812.

- (44) Scientific, T. *Overview of Crosslinking and Protein Modification*. ThermoFisher Scientific (accessed).
- (45) Rappsilber, J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol* **2011**, *173* (3), 530-540.
- (46) Yu, C.; Huang, L. Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal Chem* **2018**, *90* (1), 144-165.
- (47) Scientific, T. *Overview of Affinity Purification*. ThermoFisher Scientific, (accessed).
- (48) *Protein Affinity Chromatography*. IBA Lifesciences GmbH, (accessed).
- (49) Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **1984**, *11* (11), 601.
- (50) Steen, H.; Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **2004**, *5* (9), 699-711.
- (51) Kalisman, N.; Adams, C. M.; Levitt, M. Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. *Proc Natl Acad Sci U S A* **2012**, *109* (8), 2884-2889.
- (52) Kim, S. J.; Fernandez-Martinez, J.; Nudelman, I.; Shi, Y.; Zhang, W.; Raveh, B.; Herricks, T.; Slaughter, B. D.; Hogan, J. A.; Upla, P.; et al. Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **2018**, *555* (7697), 475-482.
- Yan, C.; Hang, J.; Wan, R.; Huang, M.; Wong, C. C.; Shi, Y. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **2015**, *349* (6253), 1182-1191.
- (53) Belsom, A.; Schneider, M.; Fischer, L.; Brock, O.; Rappsilber, J. Serum Albumin Domain Structures in Human Blood Serum by Mass Spectrometry and Computational Biology. *Mol Cell Proteomics* **2016**, *15* (3), 1105-1116.
- (54) Komolov, K. E.; Du, Y.; Duc, N. M.; Betz, R. M.; Rodrigues, J.; Leib, R. D.; Patra, D.; Skiniotis, G.; Adams, C. M.; Dror, R. O.; et al. Structural and Functional Analysis of a beta2-Adrenergic Receptor Complex with GRK5. *Cell* **2017**, *169* (3), 407-421 e416.
- (55) Zhong, X.; Navare, A. T.; Chavez, J. D.; Eng, J. K.; Schweppe, D. K.; Bruce, J. E. Large-Scale and Targeted Quantitative Cross-Linking MS Using Isotope-Labeled Protein Interaction Reporter (PIR) Cross-Linkers. *J Proteome Res* **2017**, *16* (2), 720-727.
- (56) Yu, C.; Mao, H.; Novitsky, E. J.; Tang, X.; Rychnovsky, S. D.; Zheng, N.; Huang, L. Gln40 deamidation blocks structural reconfiguration and activation of SCF ubiquitin ligase complex by Nedd8. *Nat Commun* **2015**, *6*, 10053.
- (57) Chavez, J. D.; Schweppe, D. K.; Eng, J. K.; Zheng, C.; Taipale, A.; Zhang, Y.; Takara, K.; Bruce, J. E. Quantitative interactome analysis reveals a chemoresistant edgotype. *Nat Commun* **2015**, *6*, 7928.

- (58) Cambray, S.; Gao, J. Versatile Bioconjugation Chemistries of ortho-Boronyl Aryl Ketones and Aldehydes. *Acc Chem Res* **2018**, *51* (9), 2198-2206.
- (59) Antonio, J. P. M.; Russo, R.; Carvalho, C. P.; Cal, P.; Gois, P. M. P. Boronic acids as building blocks for the construction of therapeutically useful bioconjugates. *Chem Soc Rev* **2019**, *48* (13), 3513-3536.
- (60) Structure, Properties, and Preparation of Boronic Acid Derivatives Overview of Their Reactions and Applications. 2011.
- (61) Cox, P. A.; Reid, M.; Leach, A. G.; Campbell, A. D.; King, E. J.; Lloyd-Jones, G. C. Base-Catalyzed Aryl-B(OH)<sub>2</sub> Protodeboronation Revisited: From Concerted Proton Transfer to Liberation of a Transient Aryl Anion. *J Am Chem Soc* **2017**, *139* (37), 13156-13165.
- (62) Cesare Achilli, A. C., Maurizio Fagnoni, Cesare Balduini & Giampaolo Mi. Susceptibility to hydrolysis of phenylboronic pinacol esters at physiological pH. *Central European Journal of Chemistry* **2012**.
- (63) Akgun, B.; Hall, D. G. Fast and Tight Boronate Formation for Click Bioorthogonal Conjugation. *Angew Chem Int Ed Engl* **2016**, *55* (12), 3909-3913.
- (64) Lacina, K.; Skladal, P.; James, T. D. Boronic acids for sensing and other applications - a mini-review of papers published in 2013. *Chem Cent J* **2014**, *8* (1), 60.
- (65) Hall, D. G. Boronic acid catalysis. *Chemical Society Reviews* **2019**.
- (66) Xu, J.; Wang, X.; Shao, C.; Su, D.; Cheng, G.; Hu, Y. Highly efficient synthesis of phenols by copper-catalyzed oxidative hydroxylation of arylboronic acids at room temperature in water. *Org Lett* **2010**, *12* (9), 1964-1967.
- (67) Knapp, D. M.; Gillis, E. P.; Burke, M. D. A general solution for unstable boronic acids: slow-release cross-coupling from air-stable MIDA boronates. *J Am Chem Soc* **2009**, *131* (20), 6961-6963.
- (68) Henry G. Kuivila, J. F. R., and John A. Mangravite. Metal Ion Catalysis in the Protodeboronation of Areneboronic Acids. **1964**.
- (69) Cox, P. A.; Leach, A. G.; Campbell, A. D.; Lloyd-Jones, G. C. Protodeboronation of Heteroaromatic, Vinyl, and Cyclopropyl Boronic Acids: pH-Rate Profiles, Autocatalysis, and Disproportionation. *J Am Chem Soc* **2016**, *138* (29), 9145-9157.
- (70) contributors, W. Protodeboronation. *Wikipedia contributors* **27 October 2022**, 1118578590.
- (71) Jing, C.; Cornish, V. W. Chemical tags for labeling proteins inside living cells. *Acc Chem Res* **2011**, *44* (9), 784-792.
- (72) Renault, K.; Fredy, J. W.; Renard, P. Y.; Sabot, C. Covalent Modification of Biomolecules through Maleimide-Based Labeling Strategies. *Bioconjug Chem* **2018**, *29* (8), 2497-2513.

- (73) Davies, S.; Stenton, B. J.; Bernardes, G. J. L. Bioorthogonal Decaging Reactions for Targeted Drug Activation. *Chimia (Aarau)* **2018**, *72* (11), 771-776.
- (74) Yang, W.; Gao, X.; Wang, B. Boronic acid compounds as potential pharmaceutical agents. *Med Res Rev* **2003**, *23* (3), 346-368.
- (75) Prlic, A.; Kalro, T.; Bhattacharya, R.; Christie, C.; Burley, S. K.; Rose, P. W. Integrating genomic information with protein sequence and 3D atomic level structure at the RCSB protein data bank. *Bioinformatics* **2016**, *32* (24), 3833-3835.
- (76) Elewski, B. E.; Aly, R.; Baldwin, S. L.; Gonzalez Soto, R. F.; Rich, P.; Weisfeld, M.; Wiltz, H.; Zane, L. T.; Pollak, R. Efficacy and safety of tavaborole topical solution, 5%, a novel boron-based antifungal agent, for the treatment of toenail onychomycosis: Results from 2 randomized phase-III studies. *J Am Acad Dermatol* **2015**, *73* (1), 62-69.
- (77) Akama, T.; Baker, S. J.; Zhang, Y. K.; Hernandez, V.; Zhou, H.; Sanders, V.; Freund, Y.; Kimura, R.; Maples, K. R.; Plattner, J. J. Discovery and structure-activity study of a novel benzoxaborole anti-inflammatory agent (AN2728) for the potential topical treatment of psoriasis and atopic dermatitis. *Bioorg Med Chem Lett* **2009**, *19* (8), 2129-2132.
- (78) Antonio, J. P. M.; Goncalves, L. M.; Guedes, R. C.; Moreira, R.; Gois, P. M. P. Diazaborines as New Inhibitors of Human Neutrophil Elastase. *ACS Omega* **2018**, *3* (7), 7418-7423.
- (79) Ahmed, V.; Liu, Y.; Silvestro, C.; Taylor, S. D. Boronic acids as inhibitors of steroid sulfatase. *Bioorg Med Chem* **2006**, *14* (24), 8564-8573.
- (80) Aykin-Burns, N.; Ahmad, I. M.; Zhu, Y.; Oberley, L. W.; Spitz, D. R. Increased levels of superoxide and H<sub>2</sub>O<sub>2</sub> mediate the differential susceptibility of cancer cells versus normal cells to glucose deprivation. *Biochem J* **2009**, *418* (1), 29-37.
- (81) Wang, L.; Xie, S.; Ma, L.; Chen, Y.; Lu, W. 10-Boronic acid substituted camptothecin as prodrug of SN-38. *Eur J Med Chem* **2016**, *116*, 84-89.
- (82) Du, L.; Li, M.; Zheng, S.; Wang, B. Rational Design of a Fluorescent Hydrogen Peroxide Probe Based on the Umbelliferone Fluorophore. *Tetrahedron Lett* **2008**, *49* (19), 3045-3048.
- (83) Wu, X.; Li, Z.; Chen, X. X.; Fossey, J. S.; James, T. D.; Jiang, Y. B. Selective sensing of saccharides using simple boronic acids and their aggregates. *Chem Soc Rev* **2013**, *42* (20), 8032-8048.
- (84) Ramsay, W. J.; Bayley, H. Single-Molecule Determination of the Isomers of d-Glucose and d-Fructose that Bind to Boronic Acids. *Angew Chem Int Ed Engl* **2018**, *57* (11), 2841-2845.
- (85) Munkley, J. The glycosylation landscape of pancreatic cancer. *Oncol Lett* **2019**, *17* (3), 2569-2575.

- (86) Winblade, N. D.; Nikolic, I. D.; Hoffman, A. S.; Hubbell, J. A. Blocking adhesion to cell and tissue surfaces by the chemisorption of a poly-L-lysine-graft-(poly(ethylene glycol); phenylboronic acid) copolymer. *Biomacromolecules* **2000**, *1* (4), 523-533.
- (87) SJ Coutts, J. A., D Krolikowski, RJ Snow. Two efficient methods for the cleavage of pinanediol boronate esters yielding the free boronic acids. *Tetrahedron letters* **1994**.
- (88) Kinder, D. H.; Katzenellenbogen, J. A. Acylamino boronic acids and difluoroborane analogues of amino acids: potent inhibitors of chymotrypsin and elastase. *J Med Chem* **1985**, *28* (12), 1917-1925. Roger J. Snow, W. W. B., Randall W. Barton, Scot J. Campbell, Simon J. Coutts, Dorothy M. Freeman, William G. Gutheil, Terence A. Kelly, Charles A. Kennedy, and *JACS* **1994**.
- (89) Sandford, C.; Aggarwal, V. K. Stereospecific functionalizations and transformations of secondary and tertiary boronic esters. *Chem Commun (Camb)* **2017**, *53* (40), 5481-5494.
- (90) Matzinger, M.; Kandioller, W.; Doppler, P.; Heiss, E. H.; Mechtler, K. Fast and Highly Efficient Affinity Enrichment of Azide-A-DSBSO Cross-Linked Peptides. *J Proteome Res* **2020**, *19* (5), 2071-2079.
- (91) Jiang, P. L.; Wang, C.; Diehl, A.; Viner, R.; Etienne, C.; Nandhikonda, P.; Foster, L.; Bomgarden, R. D.; Liu, F. A Membrane-Permeable and Immobilized Metal Affinity Chromatography (IMAC) Enrichable Cross-Linking Reagent to Advance In Vivo Cross-Linking Mass Spectrometry. *Angew Chem Int Ed Engl* **2022**, *61* (12), e202113937.