Mass Spectrometric Characterisation of Protein Assemblies from *Bitis arietans* Snake Venom

Emily Rose Bubner

Bachelor of Science (Advanced)

Supervisors:

Assoc. Prof. Tara Pukala

Dr Blagojce Jovcevski

A thesis submitted for the degree of Master of Philosophy

THE UNIVERSITY of ADELAIDE

2021

School of Physical Sciences

The University of Adelaide

Contents

Contents i
Statement of Originality iii
Acknowledgements iv
Abstract vi
Abbreviations viii
List of Figures x
List of Schemes xi
List of Tables xi
Chapter 1: Introduction 1
1.1 Snake Venom Composition 1
1.2 Interest in Venom Characterisation 3
1.3 Biomolecular Mass Spectrometry 6
1.3.1 Instrumentation 10
1.4 Cross-Linking Mass Spectrometry 13
1.5 Project Aims 16
Chapter 2: The Synthesis and Application of Novel Cross-Linking Reagents for Cross-Linking Mass Spectrometry 17
2.1 Introduction 17
2.2 Development of a Modular Synthetic Protocol 23
2.3 Synthesis of Homobifunctional Linkers 25
2.3.1 Synthesis of the Affinity Tag Moiety and Spacer Arm 27
2.3.2 Reactive Group Incorporation 28
2.3.3 Formation of the Cleavable Sulphonium Ion 35
2.4 Heterobifunctional Linker Synthesis 37
2.4.1 Synthesis of the Affinity Tag Moiety and Spacer Arm 39
2.4.2 Specific Reactive Group Incorporation 40
2.5 Optimisation of the Cross-Linking Reaction 41
2.5.1 Reactive Group Activation 41
2.5.2 Protein Cross-Linking 44
2.6 Affinity Tag Activation 47
2.7 Localisation of Cross-Linking Sites 49
2.8 Chapter Conclusions and Future Work 50
2.9 Materials and Methods 51

2.9.1 Materials 51
2.9.2 General Synthetic Methods 51
2.9.3 NMR Analysis 52
2.9.4 MS Analysis of Cross-linking Reagents 52
2.9.5 General Biological Methods 53
2.9.6 General Synthetic Procedures 55
2.9.7 Synthesis of Homobifunctional Linkers 57
2.9.8 Synthesis of Heterobifunctional Linker 66
2.9.9 Cross-Linking Methods 70
Chapter 3: Mass Spectrometric Interrogation of the Higher-Order Protein Structures in *Bitis arietans* venom 72
3.1 Introduction 72
3.2 Method for Venom Proteome Analysis 73
3.2.1 Venom Source and Purification 73
3.2.2 Intact MS Experiments 74
3.2.3 Native IMMS Analysis 75
3.2.4 Proteomic Analysis 75
3.2.5 XLMS 76
3.3 Results and Discussion 77
3.3.1 *B. arietans* venom purification by SEC and SDS-PAGE 77
3.3.2 MS Analysis of *B. arietans* Venom Components 80
3.3.3 Native IMMS 86
3.3.4 Proteomic Analysis of fractionated *B. arietans* venom 90
3.3.5 Cross-Linking of Venom 92
3.4 Chapter Conclusions and Future Work 96
3.5 Materials and Methods 98
3.5.1 Materials 98
3.5.2 General Methods 98
3.5.3 Methods 99
Chapter 4: Summary and Conclusions 104
Bibliography 107
Appendix A: Supplementary Proteomics Data for XL Lysozyme Samples 118
Appendix B: Supplementary Proteomics Data for Venom 119
Appendix C: Publication 1 147
Appendix D: Publication 2 156

Statement of Originality

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Emily Rose Bubner

Acknowledgements

Firstly, a massive thank you to my supervisor Tara Pukala for her time, patience, guidance and support throughout this whole process, and for accepting me into the group even though "I cannot math". Thank you to my co-supervisor Blagojce Jovcevski (BJ) for providing me with the most thorough drafting of my scientific babble, and for teaching me your ways to figure construction… I hope you enjoy the 'where's BJ' surprise hidden somewhere in this thesis! A big thank you to the rest of Pukala group (Henry, Lee, Kat, Alex, Ruth, Jack, River and Lewis) for making post grad student life so enjoyable. Thanks for filling the past two years with bad jokes, fruchocs, toilet paper, weirdness and the hub chip box challenge, and thank you for truly believing that food is the answer to everything (especially when science is not working!!) A special shoutout goes to my fellow snake enthusiast pal Ruth – probably the kindest human to walk this earth, all your endeavours to help others do not go unnoticed and I definitely could not have submitted this thesis without your support! Also a special mention to Kat for all her help and entertainment in the Badger building. Further thank you's go to PPR student Mariana for having so much enthusiasm for everything science, to Parul Mittal and Chris Cursaro from Adelaide Proteomics and to Phil Clements and Matt Bull from the Chemistry Department for all their help with experiments and instrument maintenance. Thank you to Flinders University Analytical Facility, The University of Adelaide Glycomics facility, and UniSA Future Industries Institute for the use of their instruments.

To my friends, I'll forever appreciate you being there when I needed to take a step away from the lab. A special mention to my original uni pals Chloe and Carla who taught me the importance of burgers and accepting your inner weird. Another shoutout to my undergrad chem pals (special mention to Jess, Angela, Alanah, Eb, Georgia and Andrew) for helping me find my love for chemistry and for providing *alkynes* of chem puns *periodically* (which always induced a *reaction* to keep me in my *element*). Thank you for keeping your *ion* me throughout undergrad and for all the *TLC* you provided throughout our studies (tender loving care not thin layer chromatography). Thanks for helping me find the *solutions* when we were all *triene* our best but *diene* inside. I'm very grateful for the *bonds* we have made and I think you are all very *chemicool*. An equally big thank you to the other uni pals I've collected on the way (special mention Jess and Dee), you're all amazing. To my fitness pals (dance, footy (both the

Flinders Park and Tanunda gals) and my gym/yoga motivator Carla), thank you for helping me work off all the potato I consumed during my candidature. To my housemates Emma and Jesh, thank you for laughing at me when you'd find me hiding in the pantry eating chocolate instead of writing my thesis, you're the best. To the friends that have been there forever (Amelia, Lydia, Rose), and all my other pals not mentioned here who have given me reasons to smile, you're amazing and I love you all.

Lastly, I owe everything I have achieved to my family (special mention Mum, Dad and Nicola) – you have provided me with so many opportunities and an endless amount of love and support in everything I do. Thank you for supporting my procrasti-baking needs and for telling me off when you heard the crackle of a chip packet in the kitchen when I was meant to be studying. You inspire me everyday and I wish I could put into words the love and gratitude I have for you all.

Fangs for reading my thes-hissssssss.

Abstract

Snake venom is a sophisticated lethal weapon system designed to immobilise and debilitate prey. It is comprised of various bioactive compounds including proteins which contribute to the high potency of venom due to their selectivity when interacting with biological targets. It is believed the interactions that stabilise higher-order protein structures may enhance their activity. A lack of suitable techniques to study dynamic higher-order protein structures, however, has left a significant knowledge gap in understanding how protein interactions and synergistic effects contribute to biological activity and outcomes of envenomation. Thus, an in-depth characterisation of the highly specific mechanisms behind envenomation is lacking at the molecular level.

Cross-linking mass spectrometry can be used to study transient and dynamic interactions involved in the function and malfunction of biological systems. Although a well-established analytical method, a lack of commercially available multifunctional cross-linking reagents hampers the diversity of its application. This thesis contributes to new developments in the cross-linking mass spectrometry workflow. A novel cross-linker library has been developed based on a modular cross-linker structure comprised of a reactive group, a linker arm, and an affinity enrichment tag. Derivatives were synthesised by substituting different commercially available hydrocarbons and amino acids. These units can be linked using a small number of simple chemical reactions such as esterification, amide coupling, alkylation and hydrolysis. By incorporating different spacer arm lengths, reactive groups and cleavable groups, derivatives with different functionalities can be synthesised and applied to different protein systems. The utility of this synthetic process was validated through the synthesis of a non-cleavable heterobifunctional linker, a non-cleavable homobifunctional linker and a sulphonium ion cleavable homobifunctional linker. The heterobifunctional derivative contained a non-specific diazirine group and a lysine-specific fluorinated phenyl reactive group along with an azide affinity tag for copper catalysed alkyne azide click enrichment. Both homobifunctional linkers contained two lysine-specific fluorinated phenyl reactive groups and an alkyne affinity tag for enrichment. Lysozyme, which is known to self-associate at neutral pH, was used to develop a cross-linking assay to compare commercially available cross-linker disuccinimidyl sulphoxide (DSSO) with the linkers designed and synthesised in this thesis. The cross-links identified using

the novel homobifunctional linkers were consistent with those detected using the commercially available DSSO, suggesting the newly synthesised linkers can provide valuable structural information on complicated protein systems.

This thesis also demonstrates the use of an integrated mass spectrometric approach to elucidate the structures of proteins found in the venom of the African Puff Adder, *Bitis arietans.* Following fractionation of the crude venom, a combination of intact, native and ion-mobility mass spectrometry experiments was performed to study the structure and stoichiometry of the higher-order protein structures found within the venom. Bottom-up proteomic analysis was used to characterise the primary sequence of the protein components and cross-linking mass spectrometry was used to further probe the protein structures proposed by intact and native mass spectrometry. Cross-linking mass spectrometry was also used to stabilise the oligomeric species allowing protein identification. Using this approach, a 120 kDa octameric C-type lectin, a 60 kDa tetrameric C-type lectin and a 30 kDa dimeric C-type lectin were identified as higher-order protein assemblies existing in the venom.

Together, the structural information afforded by this work demonstrates the ability to study protein structures in complex biological systems using an integrated mass spectrometry approach. Furthermore, it provides the potential to expand the methods currently available for studying structure-function relationships in other biological assemblies. Following structure elucidation of these venom proteins, functional studies can be performed to explore this relationship in the oligomeric protein species. Not only will this assist in understanding the mechanism of envenomation, but it may also guide the development of venom-based therapeutics and biotechnological tools.

Abbreviations

3FTx	3-Finger toxin
5'NUC	5'Nucleotidase
ACE	Angiotensin-converting enzyme
AChE	Acetylcholinesterase
ACN	Acetonitrile
ATDs	Arrival time distribution
BPP	Bradykinin peptide
BTTAA	2-(4-((Bis((1-(*tert*-butyl)-1*H*-1,2,3-triazol-4-yl)methyl)amino)methyl)-1*H*-1,2,3-triazol-1-yl)acetic acid
C'	Complement protein
CCS	Collisional cross section
CD_3CN	Deuterated acetonitrile
CD_3OH	Deuterated methanol
$CDCl_3$	Deuterated chloroform
CID	Collision induced dissociation
CO_2	Carbon dioxide
CRISP	Cysteine-rich secretory protein
CTL	C-type lectin
CuAAC	Copper catalysed alkyne azide click
$CuSO_4$	Copper sulphate
Cy3	Cyanine fluorescent azide group
CYS	Cystatin
Da	Dalton
DCC	N,N'-dicyclohexylcarbodiimide
DCM	Dichloromethane
DIPEA	N,N-diisopropylethylamine
DIS	Disintegrin
DMF	Dimethylformaide
DMSO	Dimethyl sulphoxide
DSC	N,N'-disuccinimidyl carbonate
DSP	Disodium phosphate
DSSO	Disuccinimidyl sulphoxide
DTT	Dithiothreitol
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
emPAI	Exponentially modified protein abundance index
ESI	Electrospray ionisation
EtOAc	Ethyl acetate
FA	Formic acid
Fmoc	Fluorenylmethyloxycarbonyl
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
HCD	Higher-energy collision dissociation
HCl	Hydrochloric acid
HRMS	High resolution mass spectrometry
HYAL	Hyaluronidase
IAA	Iodoacetamide
IMMS	Ion mobility mass spectrometry
K	Lysine
kDa	Kilodalton
KUN	Kunitz-type serine protease inhibitor
LAAO	L-amino acid oxidase
LC/MS	Liquid chromatography mass spectrometry

LiOH	Lithium hydroxide
Lys	Lysine
m/z	Mass to charge ratio
MeOH	Methanol
MGF	Mascot generic format file
mM	Millimolar
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NaCl	Sodium chloride
NaH	Sodium hydride
NanoESI	Nano-electrospray ionisation
NaOH	Sodium hydroxide
N-Boc	*N*-butyloxycarbonyl
NGF	Nerve growth factor
NH_4OAc	Ammonium acetate
NHS	N-hydroxysuccinimide
NMR	Nuclear magnetic resonance
NP	Natriuretic peptide
PBS	Phosphate buffered saline
PDE	Phosphodiesterase
PFP	Pentafluorophenyl
PI	Proteinase inhibitor
PLA_2	Phospholipase A_2
PLB	Phospholipase B
PSM	Peptide spectral match
PTM	Post-translational modification
Q-IM-TOF	Quadrupole ion-mobility time-of-flight
Q-TOF	Quadrupole time-of-flight
R*f*	Retardation factor
rpm	Rotations per min
RSD	Relative standard deviation
RSL	Rattle-snake venom lectin
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
$SOCl_2$	Thionyl chloride
SVMP	Snake venom metalloproteinase
SVSP	Snake venom serine protease
T_3P	Propanephosphonic acid anhydride
TCEP	Tris(2-carboxyethyl)phosphine)
TEA	Triethyl amine
TFP	Tetrafluorophenyl
TGS	Tris glycine SDS buffer
THF	Tetrahydrofuran
TLC	Thin layer chromatography
TOF	Time-of-flight
UV	Ultraviolet
VEGF	Vascular endothelial growth factor
VESP	Vespryn
VF	Venom factor
WAP	Waprin
x *g*	Times gravity
XL	Cross-linking
XLMS	Cross-linking mass spectrometry

List of Figures

Figure 1.1: The four fundamental levels of protein structure. 4
Figure 1.2: The process of electrospray ionisation. 7
Figure 1.3: Schematic representation of a quadrupole mass analyser. 8
Figure 1.4: Product ions from peptide MS/MS fragmentation. 9
Figure 1.5: Generic bottom-up proteomics workflow. 10
Figure 1.6: Schematic representation of the LTQ Orbitrap mass spectrometer. 11
Figure 1.7: Schematic representation of the Agilent 6560 DTIMS mass spectrometer. 12
Figure 1.8: XLMS workflow for functionalised cross-linkers 15
Figure 2.1: Comparison of reactive group reagent derivatives. 18
Figure 2.2 The modular cross-linker design and general synthetic route. 24
Figure 2.3: Non-cleavable (9bii) and cleavable (10) homobifunctional cross-linkers. 27
Figure 2.4: Partial COSY spectra for the characterisation of 9aii at 500 MHz in CD_3CN 33
Figure 2.5: MS/MS spectra of linker (10) reveals selective fragmentation by CID. 37
Figure 2.6: Heterobifunctional linker design. 38
Figure 2.7: Mass spectrum of N-Boc lysine modified specifically with linker (20). 42
Figure 2.8: X-ray crystal structure of Gallus gallus lysozyme (PDB ID: P00698).[107] 44
Figure 2.9: Optimisation of cross-linking reaction conditions of DSSO with lysozyme. 46
Figure 2.10: SDS-PAGE of cross-linked lysozyme. 47
Figure 2.11: SDS-PAGE of cross-linked lysozyme and CuAAC linked Cy3. 48
Figure 3.1: The African Puff Adder, Bitis arietans. 73
Figure 3.2: MS workflow for studying the higher-order structures of venom proteins 77
Figure 3.3: Size exclusion chromatography elution profile of B. arietans venom. 78
Figure 3.4: SDS-PAGE of purified B. arietans fractions. 79
Figure 3.5: ESI-MS analysis of B. arietans venom Fraction A following SEC fractionation. 81
Figure 3.6: ESI-MS analysis of B. arietans venom Fraction B following SEC fractionation. 82
Figure 3.7: ESI-MS analysis of B. arietans venom Fraction C following SEC fractionation. 83
Figure 3.8: ESI-MS analysis of B. arietans venom Fraction D following SEC fractionation. 84
Figure 3.9: ESI-MS analysis of B. arietans venom Fraction E following SEC fractionation. 85
Figure 3.10: IMMS analysis following SEC fractionation of B. arietans venom. 87
Figure 3.11: Proteomic composition of B. arietans venom fractions. 92
Figure 3.12: SDS-PAGE analysis of cross-linked B. arietans fractions. 94

Figure 3.13: X-ray crystal structure of rattle snake venom lectin (PDB ID: P21963).[107,140] 96
Figure 3.14: Proposed structure of the protein assemblies in B. arietans venom. 97

List of Schemes

Scheme 2.1: Generalised copper catalysed alkyne azide click reaction. 20
Scheme 2.2: MS/MS fragmentation of previously reported cleavable cross-linkers. 22
Scheme 2.3: Synthetic route to homobifunctional cross-linkers. 26
Scheme 2.4: Synthetic scheme for the methylation of the sulphur group in linker (9ai). 36
Scheme 2.5: Synthetic route to the heterobifunctional cross-linker. 39

List of Tables

Table 1.1: Protein families frequently identified in snake venoms. 2
Table 2.1: TFP and PFP esterification conditions applied to diacids 8a, 8b and lysine 11. 35
Table 2.2: Assays performed to determine reactivity of diazirine group. 43
Table 2.3: Cross-links identified in lysozyme using DSSO and novel linker 10. 49
Table 3.1: Calculated CCS values for venom proteins contained in SEC Fractions B and C. ...88
Table 3.2: Effective densities calculated from IMMS. 89

Chapter 1: Introduction

1.1 Snake Venom Composition

Over 4.5 million snakebites are reported globally every year, resulting in over 100,000 deaths. The majority of these fatalities occur in rural regions where afflicted individuals have no immediate access to healthcare facilities.[1] In sub-Saharan Africa alone, approximately 300,000 snakebite cases are reported annually. These events result in over 32,000 deaths and leave a large proportion of remaining victims with permanent local tissue damage and chronic disabilities.[2]

Designed to immobilise prey and defend against predators, snake venom is a complex lethal cocktail containing various bioactive compounds.[3–5] Following administration through a single bite, this predominantly proteinaceous mixture can disrupt key physiological systems in the victim, leading to debilitation and often death.[4–6] Proteins are believed to be the major lethal components of most snake venoms due to their selectivity for biological targets.[5,7,8] Venom proteins are stabilised by an abundance of disulphide bonds and have high potency at low dosage.[5,9,10]

Snake venom proteins can be partitioned into a number of protein families with distinct functions, which can be broadly categorised into enzymatic and non-enzymatic toxins (Table 1.1).[7,11-13] Generally, enzymatic toxins contribute to the lethal effects of the venom by interfering with cellular pathways involved in haemostasis, tissue necrosis and myotoxicity.[10,14] Comparatively, non-enzymatic toxins usually contribute to prey immobilisation by disrupting cardiovascular and neuromuscular systems.[5]

Table 1.1: Protein families frequently identified in snake venoms.

TOXIN CLASS	TOXIN FAMILY	ABBREVIATION
Enzyme	5'Nucleotidase	5'NUC
	Acetylcholinesterase	AChE
	Hyaluronidase	HYAL
	L-amino acid oxidase	LAAO
	Phosphodiesterase	PDE
	Phospholipase A_2	PLA_2
	Phospholipase B	PLB
	Snake venom metalloproteinase	SVMP
	Snake venom serine protease	SVSP
Non-enzyme	3-Finger toxin	3FTx
	Bradykinin peptide	BPP
	Complement protein	C'
	C-type lectin	CTL
	Cystatin	CYS
	Cysteine-rich secretory protein	CRISP
	Disintegrin	DIS
	Kunitz-type serine protease inhibitor	KUN
	Natriuretic peptide	NP
	Nerve growth factor	NGF
	Proteinase inhibitor	PI
	Vascular endothelial growth factor	VEGF
	Venom factor	VF
	Vespryn	VESP
	Waprin	WAP

Whilst high homology exists within each protein family, there can be large variability in protein abundance between snake species, contributing to unique symptoms of envenomation for individual species.[1,4,10,15] Toxin abundance can also vary between populations and individuals of the same species located in different geographic areas, and within the same specimen during different ontogenetic stages of life.[2,4] Venom production has a substantial metabolic cost, so it is reasonable to predict that diversification in venom protein expression may reflect an organism's response to ecological and developmental change.[16]

1.2 Interest in Venom Characterisation

The application of whole snake venoms in traditional medicine has been traced back to ancient Greece.[7,12] In the 19th century, endeavours were made to probe the toxic, biochemical and pharmaceutical properties of individual venom components, exposing the true potential of venom-based therapeutics.[12] Recent studies highlighting the selectivity of individual venom proteins towards cardiovascular, immune and nervous systems demonstrate their promising potential in the treatment of medically notable diseases.[3,8,17] For example, proteins exhibiting analgesic, antitumoral, antimicrobial, antiparasitic, antiplatelet, neuroprotective, anticoagulant and procoagulant activity have been reportedly derived from various snake species.[4,7] Additionally, the high specificity, low immunogenicity and notable potency of venom proteins make them attractive candidates for therapeutic and diagnostic development.[5,9]

There are currently nine snake venom derived drugs either commercially available or under clinical trial.[4] Captopril was the first commercially available venom-derived therapeutic agent which is based on the bradykinin potentiating peptide isolated from the venom of the Brazilian pit viper *Bothrops jararaca*. It is used worldwide to treat hypertension by competitively inhibiting the angiotensin-converting enzyme (ACE) responsible for vasoconstriction stimulation.[4,18] Its successful release on the market in 1981 initiated an influx of potential venom derived drug candidates including cobra 3FTx capable of eliciting antitumoral activity, and PLA_2, SVMP and LAAO proteins capable of eliciting antimicrobial activity.[19] In addition to disease treatment and diagnosis, venom proteins have been used for research and biotechnical purposes by targeting specific receptors in feedback loops. For example, alpha-neurotoxins of the 3FTx family isolated from *Elapidae* snakes have been used as biological tools to gain valuable information on the structure and mechanism of action of nicotinic acetyl choline receptors, found in the central and peripheral nervous system.[8]

Despite the therapeutic potential that these proteins present, their inability to be successfully carried through clinical trials remains an issue to overcome. Many of these drug candidates fail clinical tests, most frequently due to toxic side effects, low bioavailability, lack of efficacy and instability *in vivo*.[9,20] These issues are likely owed in part to a lack in pharmacological and biochemical characterisation of the individual protein components involved in envenomation.

This can be further traced back to the knowledge gap in the structure-function relationships between these venom components.[5] The central dogma of structural biology states that the overall three-dimensional protein structure is critical for its biological function, as it determines if and how the protein interacts with other molecules.[21] To understand how this structure is determined, it is important to consider the four fundamental levels of protein structure summarised in Figure 1.1.

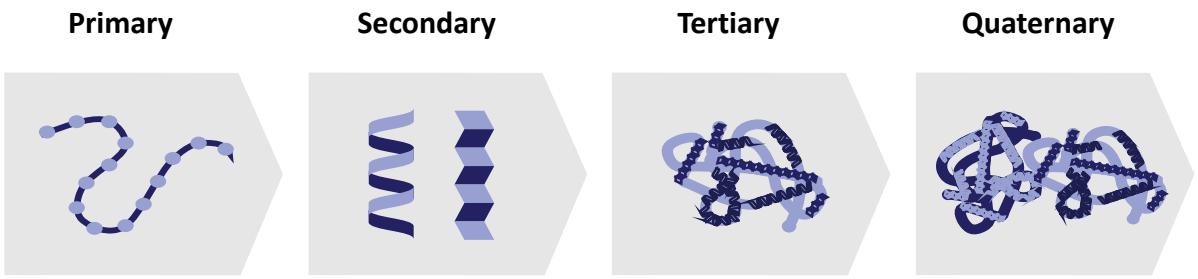


Figure 1.1: The four fundamental levels of protein structure.
The primary structure refers to the unique amino acid sequence, the secondary structure refers to repeating α-helical and β-sheet structures which are dependent on backbone hydrogen bonding networks, the tertiary structure describes the three-dimensional globular fold a monomeric protein adopts and the quaternary structure refers to the interaction between multiple monomeric subunits to form multi-subunit assemblies.

The primary protein structure refers to the unique amino acid sequence of the polypeptide chain. This determines how the protein folds at higher structural levels, therefore dictating the overall three-dimensional shape and hence, the overall function. This sequence can be further complicated by post-translational modifications (PTMs) on amino acid residues, including glycosylation, phosphorylation, and ubiquitination. These covalent modifications may occur during or after assembly of the globular protein, giving rise to further functional specificity.

The secondary structure level arises due to intramolecular hydrogen bonds forming between amino acid residues within the backbone of the polypeptide chain. This gives rise to repeating motifs like α-helices and β-pleated sheets. Amino acids have different propensities to form such structures which results in a unique combination of α-helices and β-sheets in every protein.

The tertiary level refers to the fully folded three-dimensional globular structure of a single polypeptide chain. This structure is stabilised by various covalent and non-covalent interactions which are dependent on amino acid hydrophobicity and solvent environment. In an aqueous solvent, hydrophobic residues are buried within folded protein structures, allowing charged and hydrophilic residues to remain on the surface for favourable solvent interactions.

Finally, quaternary structure refers to the higher-order association between two or more polypeptide chains to form multi-subunit protein assemblies, stabilised by non-covalent and covalent interactions. The prefixes 'homo' and 'hetero' are applied to determine whether the assembly is made up of identical or different monomeric polypeptide chains. Numeric suffixes such as 'dimer' and 'trimer' are used to detail the number of chains in a protein assembly.

Although toxins often exert bioactivity individually, many venom proteins are suspected to form higher-order quaternary assemblies to enhance the structural diversity of the venom.[5] The covalent and non-covalent interactions stabilising these structures are believed to augment the pharmacological activity of the venom. The large size of these assemblies increases the binding surface available for target molecules and the interactions stabilising these structures may expose critical amino acid residues usually buried within the monomeric counterparts, for recognition by new target molecules.[5,22] These effects may enhance venom potency, but very little work has involved characterisation of venom proteins at a quaternary structural level thus far.[5,23,24]

Elucidating the higher-order structure of venom protein assemblies is critical in understanding the role which structure-function relationships play in the pathophysiology of envenomation.[3–5,23] Previously, most work characterising the higher-order protein structures has involved high resolution techniques such as NMR spectroscopy and x-ray crystallography.[5,24,25] Such methods however, have difficulty capturing the potentially transient, dynamic and heterogeneous nature of the higher-order venom protein structures, particularly in a high throughput fashion.[5,26,27] With an endless range of venom proteins yet to be characterised, the development of new approaches for characterising quaternary protein structures is critical in advancing the pathophysiological understanding and functional applications of venom-derived proteins.[2,4,5,7,22]

1.3 Biomolecular Mass Spectrometry

Recent advances in mass spectrometry (MS) have permitted rapid and highly sensitive analysis of dynamic structures in complex protein samples.[4] Literature has widely demonstrated the use of MS-based proteomic sequencing and higher-order protein structural characterisation for many other protein systems, enabled by techniques such as native MS in conjunction with ion mobility-MS (IMMS) and shotgun proteomics.[28–30] Such high-throughput methods are well suited to interrogate the often heterogeneous nature of venom protein samples. Although bottom-up proteomics has been critical in cataloguing many venom proteins to date, native MS has been underutilised in the characterisation of venom protein assemblies.[11,17,31–33]

In a basic MS experiment, ions are generated, separated and detected in the gas phase, providing information on analyte mass and structure. Mass spectrometers are therefore comprised of three essential elements: an ionisation source for sample introduction, one or more mass analysers for ion separation and a detector for ion collection and recognition. The data is generated as a mass spectrum which displays the ion distributions for the components within the sample. This spectrum compares the *m/z* ratios of each of ion species along with their relative abundance. However, this measurement is not directly indicative of analyte concentration, as ion detection is also dependent on the relative ionisation and transmission efficiency of the species under analysis.

Electrospray ionisation (ESI) is a well-documented method for protein ionisation and is the method utilised in this thesis. It can be carried out under gentle conditions so has the ability to transfer native protein conformations from the solution phase into the gas phase based on the theory that the biomolecular phase transition occurs faster than molecular decomposition.[34,35] ESI can be performed offline whereby samples are introduced directly through a capillary needle, or online where the MS is coupled to a chromatographic system for preliminary sample purification. In the ESI process, an electric potential is applied to the needle which pulls the sample from the capillary towards the MS inlet as a fine mist of charged droplets. These droplets shrink in size as solvent is evaporated (often using heat and drying gas), forcing the charged analyte molecules closer together. Eventually, the surface tension holding the charged droplet intact is overcome by Coulombic repulsion, resulting in droplet

fission. Repeated fission events following further solvent evaporation result in analyte ion generation (Figure 1.2). Nanoelectrospray ionisation (nanoESI) is a variation of ESI often employed for native MS analysis. The nanoESI conductive capillary has a smaller aperture of approximately 1-3 µm to accommodate a lower sample volume and hence a lower flow rate. The smaller analyte-solvent droplets generated are capable of undergoing droplet fission under much milder conditions, so the high flow rates, pressures, temperatures, acidic conditions and organic solvents required for ionisation in regular ESI are no longer necessary. Instead, native-like buffers with high evaporation indices such as ammonium acetate can be used for preservation of the interactions stabilising native protein systems.[36–38]

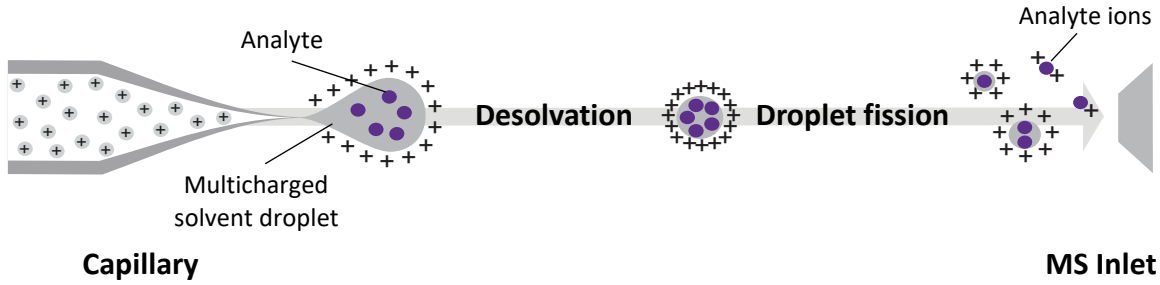


Figure 1.2: The process of electrospray ionisation.
Sample is drawn from the capillary towards the MS inlet cone. Desolvation occurs, shrinking the charged solvent droplet until the surface tension holding the charged droplet is overcome by Coulombic repulsion, resulting in droplet fission and generation of gas phase analyte ions.

Generated ions can be guided, filtered and separated using mass analysers such as time-of-flight (TOF) analysers and quadrupoles. In TOF mass analysers, ions are separated based on their travel time through a field-free flight tube of known length which directs ions to the detector in a passive manner.[39] The time taken for the ions to reach the detector is proportional to their mass to charge ratio (*m/z*), where ions with high *m/z* and lower kinetic energy will travel slower than smaller ions. Quadrupoles are designed to filter ions based on their *m/z* and consist of four metal rods of opposing charges positioned parallel to each other (Figure 1.3). A radio frequency (RF) voltage with a direct current (DC) offset voltage is applied to each rod pair, causing ions to oscillate between opposing electric fields as they pass through. Only ions of a specific *m/z* will have a stable trajectory at a given frequency, leaving other ions to collide with the parallel rods (Figure 1.3).[40]

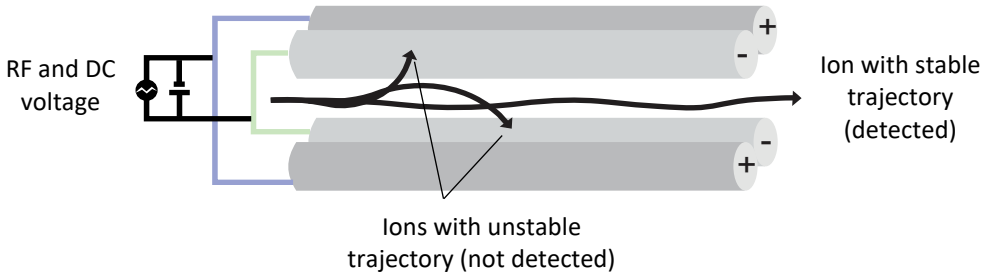


Figure 1.3: Schematic representation of a quadrupole mass analyser.
Ions pass through four parallel metal rods of opposing charges. Application of an RF voltage with a DC offset causes ions to oscillate. Ions with an unstable trajectory will crash into the metal rods, filtering ions with stable trajectories to pass through to the next sector of the instrument.

Altering the frequency of this voltage can broaden or narrow the *m/z* window of ions which successfully pass through the quadrupole to subsequent mass analysers. This is the foundation of tandem MS (MS/MS), a fundamental technique in proteomic analysis where precursor peptide ions are selected (MS^1), filtered and cleaved into smaller fragment ions for subsequent detection (MS^2). The corresponding fragmentation pattern is unique to the peptide under analysis, allowing for primary level sequencing. Many methods of fragmentation exist, but in this thesis, fragmentation was carried out by collision-induced dissociation (CID). In CID, the voltage within the collision cell is increased, triggering collisions between accelerated precursor ions and inert gas molecules such as nitrogen or argon.[41–43] This results in subsequent fragmentation primarily at the amide bonds along the peptide backbone of the precursor ion (Figure 1.4), forming a series of fragment ions differing in mass by one or more amino acids. If the charge is retained at the N-terminus, the fragment is termed a b ion, and if the charge is retained at the C-terminus, the fragment is termed a y ion. Subscripts indicate the number of amino acid residues on the fragment. Different methods of fragmentation favour cleavage of peptides at different positions along the backbone resulting in additional a, c, x and y fragment ions (Figure 1.4).[44] The unique fragmentation pattern allows the sequencing of peptides based on amino acid specific mass loss. This establishes the basis of MS-based proteomic analysis, a technique used to characterise the primary amino acid sequence of proteins from biological systems.[45] Not only does this provide information

on the sequence, instructing the three-dimensional assembly of the protein, but it allows for potential quantification within samples.[34,46–48]

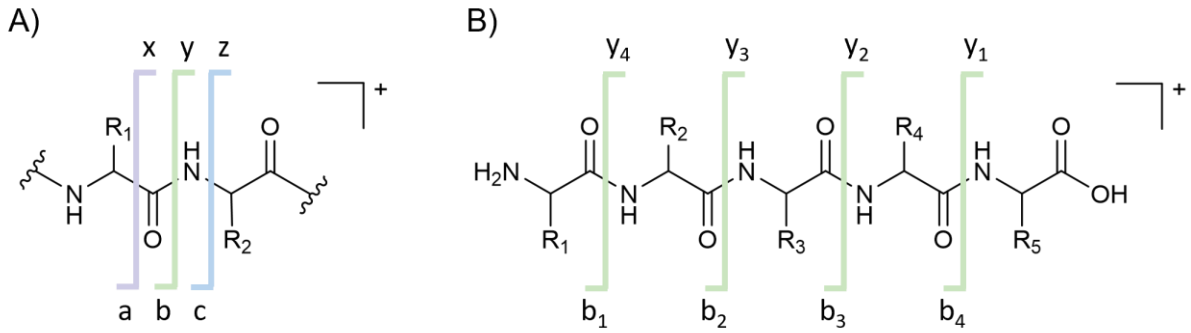


Figure 1.4: Product ions from peptide MS/MS fragmentation.
A) Fragment ions observed when the charge is retained at the N-terminus (a, b, c) and when the charge is retained at the C-terminus (x, y, z). **B**) The major fragment product ions (b, y) formed upon CID. The subscript indicates the number of amino acid residues (indicated by R groups) on the fragment.

Proteomic analysis can be carried out in a top-down approach, where proteins are sequenced in their intact state, or using a bottom-up approach, where proteins are proteolytically digested for peptide sequencing (Figure 1.5). This thesis principally employs the latter approach. Following isolation from a biological source, proteins are denatured and disulphide bonds are reduced before the sample is digested with a proteolytic enzyme such as trypsin. Trypsin cleaves the peptide backbone at the C-terminal end of arginine and lysine residues. The resulting peptides are typically separated by chromatographic methods such as liquid chromatography (LC) which can separate molecules based on polarity, charge or size, depending on the column or medium used. Upon elution from the column, samples are ionised, and ions are subjected to MS/MS analysis. Proteins within the sample are identified using an *in silico* workflow. This involves comparison of the experimentally acquired spectra to mass and theoretical fragmentation patterns of known peptides within a pre-existing protein sequence database.[49] Peptide sequences matching precursor ions within a given mass tolerance are reported along with a statistical evaluation and match score (Figure 1.5).

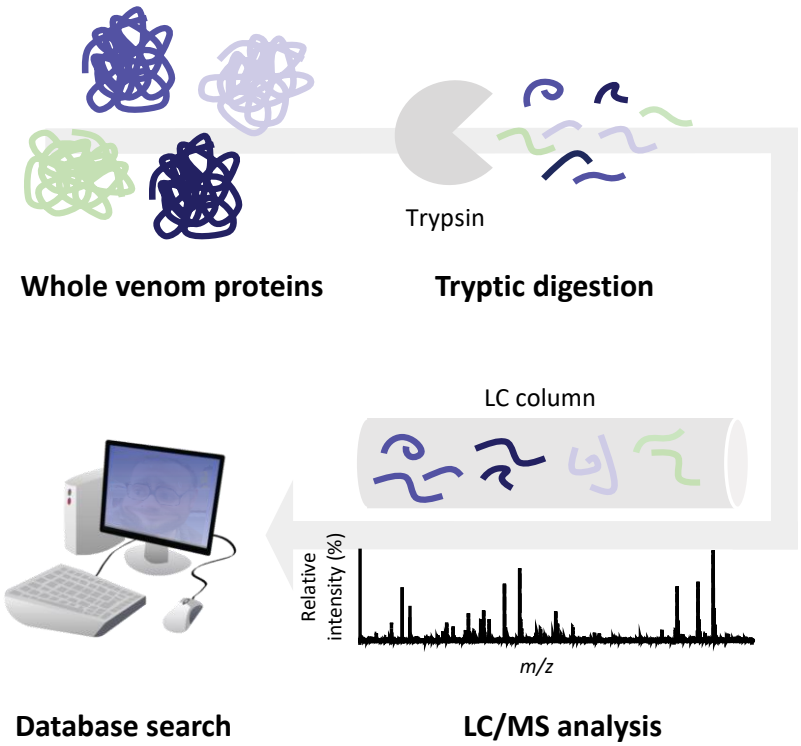


Figure 1.5: Generic bottom-up proteomics workflow.
Protein samples are digested with a proteolytic enzyme such as trypsin, before undergoing LC/MS analysis. Resulting spectra are compared to a library of pre-existing protein fragmentation patterns for identification and possible quantification analyses of complex samples.

1.3.1 Instrumentation

The configuration of MS instruments can vary substantially across applications, platforms and vendors, allowing for a variety of experiments to be performed. In this thesis, the two most frequently used instrument platforms are the linear trap quadrupole (LTQ) Orbitrap and the Agilent 6560 ion mobility quadrupole time-of-flight (Q-TOF) spectrometers. The LTQ Orbitrap (Figure 1.6) is used in this work to characterise the primary sequence of proteins by bottom-up proteomic analysis. The high resolution, sensitivity and mass accuracy of this instrument makes it an ideal candidate for the application of MS-based proteomics.[48] Following ESI, ions enter the linear trap quadrupole (LTQ) where they are refocussed and accumulated by the trap component and filtered by *m/z* by the quadrupole component. Selected ions are fragmented in either the trap by collision induced dissociation (CID) or the higher-energy collisional dissociation (HCD) cell before undergoing analysis in the orbitrap mass analyser. Here, ions are trapped in a process referred to as 'electrodynamic squeezing' where they

rotate around an inner spindle-like electrode and between two outer bell-shaped electrodes. The frequency at which different ions oscillate is proportional to their *m/z,* allowing for this property to be determined through Fourier-transformation for presentation as a mass spectrum (Figure 1.6).

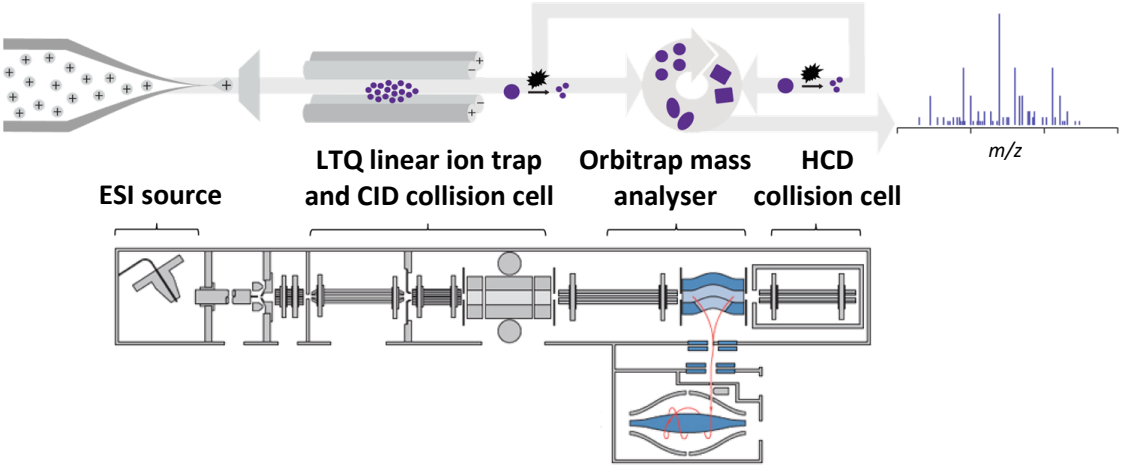


Figure 1.6: Schematic representation of the LTQ Orbitrap mass spectrometer.
The sample is ionised through ESI. Ions are trapped and filtered in the LTQ component based on their mass to charge ratio. Ions with a stable trajectory pass through the quadrupole component and are either fragmented in the CID or HCD collision cells before entering the orbitrap mass analyser. Here they rotate around an inner electrode in a process called electrodynamic squeezing which further separates based on *m/z*.

The Agilent 6560 Ion Mobility Q-TOF (6560) (Figure 1.7), was used in this work for studying higher-order protein structures by native MS (Figure 1.7). This instrument is a hybrid mass spectrometer with both a quadrupole and a TOF mass analyser. Furthermore, it has the added capability of ion mobility (IM) separation, allowing ions with identical *m/z* but different 3D shape to be distinguished from one another.[28,50] The instrument separates ions through drift tube IM (DTIMS). Following ESI, charged species are trapped using a trapping funnel and trapping gate, where ions are released in small packets into the linear drift tube which is filled with a stationary buffer gas. An electric field is applied to the linear drift tube for IM separation. The time taken for ions to drift through the length of the tube is measured as the arrival time distribution, which can be converted to a collisional cross section (Ω or CCS) allowing for interrogation of the 3D structure of different protein isoforms.[51–53] Larger ions

collide more frequently with stationary buffer gas due to their large surface area, which retards their movement through the drift tube and hence lengthens their drift time. More highly charged molecules experience a greater force from the constant electric field and therefore move faster through the buffer gas than singly charged ions. Ions are consequently separated based on Ω/*z* ratio. Ions are refocused through the rear funnel following IM separation, where they are filtered or selected in the quadrupole mass analyser. Subsequently, ions can undergo optional CID if MS/MS experiments are desired. Alternatively, ions are pushed directly through to the TOF sector before being reflected towards the detector to determine ion *m/z* for spectrum generation.[42,54,55]

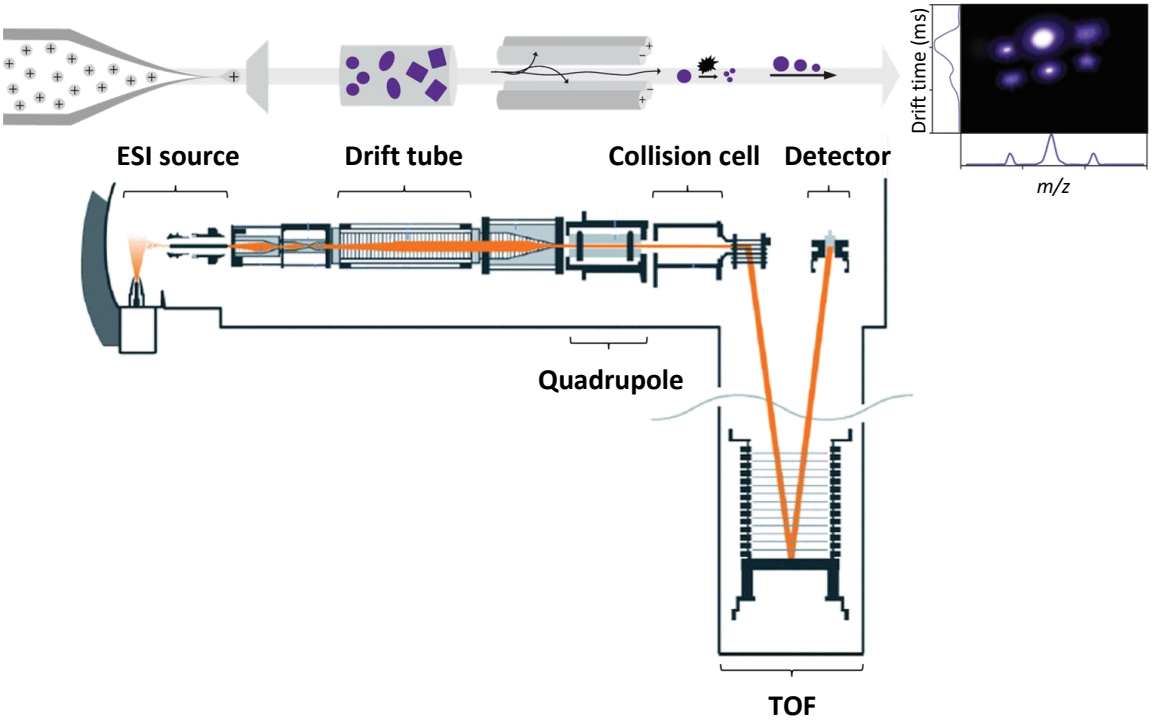


Figure 1.7: Schematic representation of the Agilent 6560 DTIMS mass spectrometer.
Following ionisation by ESI, samples are separated by DTIMS in the drift tube based on their Ω/*z*. Ions are filtered by *m/z* in the quadrupole and separated by *m/z* in the TOF component. Ions are detected giving a 3D spectrum relating *m/z* ratio to drift time. Relative abundance is indicated by colour intensity in the contour plot.

1.4 Cross-Linking Mass Spectrometry

Many of the interactions stabilising higher-order protein structures are non-covalent and may be disrupted during routine proteomic analysis. Chemical cross-linkers are frequently applied in structural biology experiments to stabilise interactions formed within protein assemblies under physiological conditions.[56,57] Advances in MS over recent decades have facilitated the development of cross-linking mass spectrometry (XLMS) as an approach for studying these interactions.[58–60] A cross-linking reagent is covalently reacted with two amino acid residues within a protein system. Cross-linked samples can either be analysed as whole proteins or as peptides using a bottom-up proteomic workflow which allows identification of cross-linked residues. Bottom-up XLMS is a rapid and powerful technique requiring minimal sample for analysis. It has the ability to analyse very large proteins systems as MS analysis is performed at the peptide level.[58,61–64] The 15 subunit 670 kDa protein complex of RNA polymerase II is an example of one of the first large protein systems analysed using XLMS.[59,65,66] Additionally, the ability to carry out the cross-linking reaction under physiological conditions enables an accurate representation of the dynamic native protein conformation to be captured.[61,63–65] It is therefore critical that the cross-linker is used in excess to form sufficient cross-links, but not such that excessive cross-linking distorts the native conformation of the proteins.[59,65]

The bottom-up XLMS workflow (Figure 1.8) is adapted from the workflow described for bottom-up proteomic analysis in Figure 1.5. Proteins of interest are cross-linked and digested using a proteolytic enzyme such as trypsin. The digestion may be performed in solution, or on a protein gel band purified by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). SDS-PAGE separates proteins within a sample by molecular weight. Samples are denatured by heating with SDS, an ionic detergent which coats proteins with a net negative charge. Samples may also be treated with a reducing agent such as dithiothreitol (DTT) to reduce any covalent disulphide interactions which may be stabilising higher-order protein structures. Using an applied electric field, the negatively charged proteins migrate through the gel matrix towards a positive electrode. Their migration rate is proportional to their overall charge density, where smaller more compact species experience less resistance from the gel and therefore migrate faster than larger species. Performing a digest on a purified gel

band can provide higher efficiency in cross-linked residue identification due to fewer background peptides being present.

Following digestion, an enrichment step may be introduced to concentrate cross-linked peptides and improve the signal detection in subsequent MS analyses.[67–69] Enriched cross-linked peptides can then be analysed by LC/MS, which may involve tandem MS experiments. Tandem MS or MS/MS can be performed when multiple mass analysers are coupled together, allowing various stages of *m/z* separation. Once ions are generated and separated at the first mass analyser, select precursor ions can be split into fragment ions under different fragmentation conditions at a second mass analyser, where they can be further separated by *m/z*. When gas-labile linkers are used in XLMS, cross-linked peptides can be split at a second mass analyser (MS^2) under mild fragmentation conditions before undergoing amino acid sequencing in a third round of *m/z* analysis (MS^3) under harsher cleavage conditions. Hence, rather than sequencing both of the cross-linked peptides simultaneously, they can be sequenced individually, simplifying the spectra and analysis of linkage sites.[69–72] Subsequently, the peptide cross-links can be identified through fragmentation patterns using cross-linking analysis software such as MeroX or XLinkX.[59,61,65] This provides information on the distance constraints between amino acid residues within the native conformation of a protein, defined by the length of the crosslinking reagent, and hence provides topological information regarding multiunit protein assemblies within a sample.[64,73]

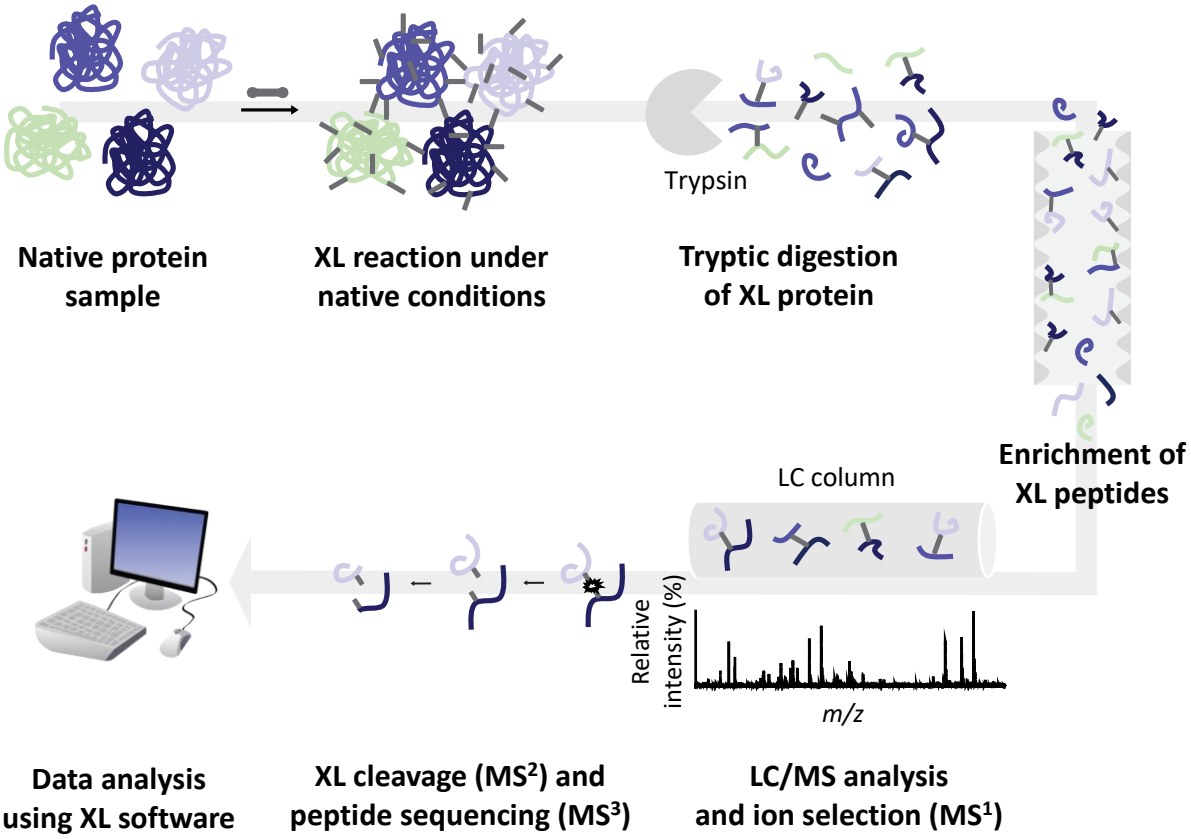


Figure 1.8: XLMS workflow for functionalised cross-linkers.
Following the cross-linking reaction (where grey bars represent cross-linking reagents), proteins are digested with an enzyme such as trypsin and resulting peptides are enriched through affinity purification before undergoing LC/MS analysis. Initially, ions are detected and selected for at MS^1 before cross-linkers are cleaved through MS^2 fragmentation and finally the peptide is sequenced through MS^3 fragmentation. Resulting data is analysed using a cross-linking software capable of analysing cleavable cross-linking data, such as MeroX or XLinkX.

The lack of commercially available multifunctional cross-linkers limits the broader application of XLMS. Inflexibility in the design of many commercial cross-linkers means they cannot be readily modified to suit different experimental conditions. Hence, the development of new cross-linking reagents with additional functionalities is ongoing to improve the efficiency, sensitivity and speed of cross-linking workflows.[22,56,62]

1.5 Project Aims

In this project, we aim to design effective methods for the synthesis of novel bifunctional chemical cross-linkers. These linkers will exhibit various functionalities to optimise the XLMS workflow, expanding the range of biological systems amenable to this technique. This will assist the development of high-throughput methods available to proficiently investigate the structure-function relationships of protein assemblies. These linkers will then be applied to characterise the protein structures found in medically significant snake venom. In parallel, we aim to apply an integrated MS-based workflow to characterise the higher-order protein structures of medically significant snake venom. This workflow will present a combination of proteomic, native, and intact MS experiments as an effective approach for gaining information on the structure and stoichiometry of protein complexes found within the venom.

Chapter 2: The Synthesis and Application of Novel Cross-Linking Reagents for Cross-Linking Mass Spectrometry

2.1 Introduction

Cross-linking mass spectrometry (XLMS) is a well-established yet still rapidly developing technique used to probe the higher-order structures of protein complexes. It allows structural information to be obtained on a protein complex which may not be amenable to high-resolution protein structural techniques alone.[59–61] Despite that XLMS is a frequently utilised method, a lack in commercially available multifunctional cross-linkers hampers the diversity of its application. Thus, development of new cross-linking reagents which incorporate additional functionalities is ongoing.

At a minimum, cross-linking reagents are comprised of two reactive groups linked by a spacer arm. The reactive groups allow a covalent bond to form between two amino acids of a protein system within a confined three-dimensional space. Traditionally, cross-linking reagents are homobifunctional, meaning they possess two identical reactive groups. Heterobifunctional reagents, which possess two different reactive groups, can introduce another element of flexibility over cross-link formation.[74] Different reactive groups can be used to target different amino acids either specifically or non-specifically. Chemoselective *N*-hydroxysuccinimide (NHS) esters, which target the primary amine in lysine residues and free N-termini of the peptide, are the most frequently used reactive groups in chemical cross-linking reagents to date.[58] However, their highly labile characteristics mean they experience instability in long-term storage and can react unexpectedly with the hydroxyl groups in serine, threonine and tyrosine side chains which complicate cross-link site identification during MS analysis.[65,75–77] Targeting lysine residues is advantageous in that they are relatively abundant and are generally located on solvent accessible regions of the protein.[5] Fluorophenyl esters are an alternative lysine specific functional group possessing better stability and improved chemoselectivity for lysine residues.[75] Examples of such compounds include

tetrafluorophenyl (TFP) and pentafluorophenyl (PFP) derivatives, which possess optimal leaving groups stabilised by resonance and inductive effects.[78] These groups can be incorporated into the linker structure by esterification with a free acid group on the linker arm. Traditionally, this esterification reaction is performed using a hydroxyl derivative of the NHS precursor, however, recent protocols have used the carbonate precursor N,N'-disuccinimidyl carbonate (DSC).[79] Carbonate precursors release carbon dioxide during esterification, driving the reaction forward in the absence of the additional coupling reagents required when using hydroxyl derivatives. A summary of alcohol and carbonate reagent derivatives used to introduce NHS and fluorophenyl esters into an acid is depicted in Figure 2.1.

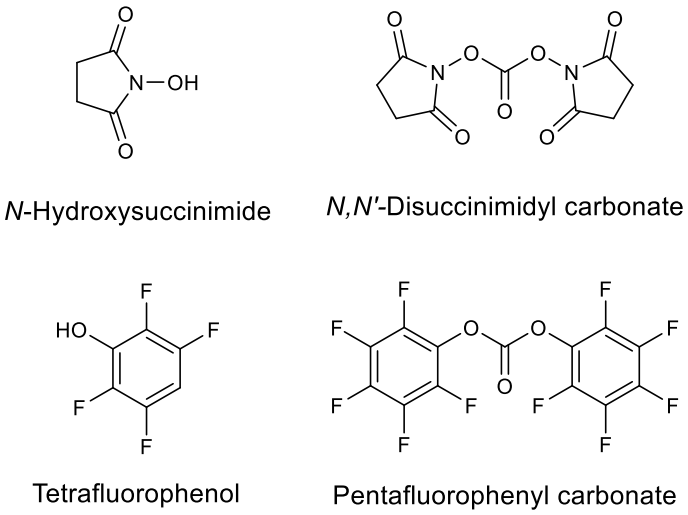


Figure 2.1: Comparison of reactive group reagent derivatives.
NHS groups may be introduced into a chemical cross-linker by *N*-hydroxysuccinimide (NHS) or *N,N'*-disuccinimidyl carbonate (DSC). Fluorophenyl groups may be introduced using tetrafluorophenol (TFP) or pentafluorophenyl carbonate (PFP). Both reactive groups target lysine groups specifically, but the electron-withdrawing fluorine stabilised aromatic ring in the fluorophenyl ester enhances the stability and hence, reduces chances of cross-reactivity with other amino acids. Hydroxyl groups react with the acid group directly, whereas the carbonate reagents undergo a preliminary reaction with base to release carbon dioxide and a deprotonated equivalent of the reactive group which thus reacts with the targeted acid.

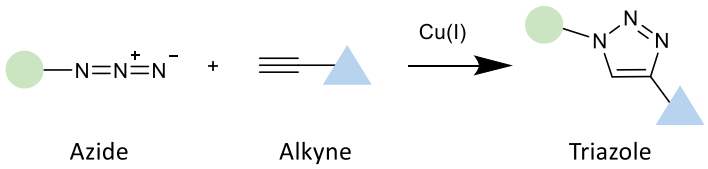
Although ambiguity in identifying cross-linked residues can be reduced using chemoselective reactive groups, their application is limited to proteins containing enough of the targeted

residues within close three-dimensional space. Alternatively, photoreactive groups such as diazirines, which can insert non-specifically into chemical bonds upon exposure to UV light, can be used to increase the number of cross-links formed.[58] A potential drawback of a homobifunctional photoreactive linker would be the diversity of products arising from a single experiment which may complicate identification of cross-linked residues amongst the complicated spectra.[58] Having a heterobifunctional linker with a chemoselective group on one end and a photoreactive group on the other end allows for cross-linking in a more controlled manner.[58] In this case, initial residue specific cross-linking is carried out in darkness before the sample is exposed to UV light for photo-induced non-specific cross-linking with adjacent residues.[77,80] This increases the number of cross-links formed as the second reactive group can insert into any amino acid within the distance constraint of the linker.[58]

Unfortunately, the cross-linking reaction itself can be quite inefficient and, following digestion of the protein, will usually result in a very small percentage of cross-linked products among an excess of unmodified peptides.[71] Furthermore, the number of potential cross-linked sites are limited by the proximity of the reactive sites on the protein. Nevertheless, within the small number of modified products, four types of modifications can exist: (1) dead-end cross-links, which are formed when one of the reactive cross-linker groups modifies a peptide, but the other end does not react with another residue, typically due to hydrolysis; (2) loop-links, which are formed when both ends of the linker react with different residues on the same peptide; (3) intra-cross-links, which are formed between different peptides that have originated from the same protein; and (4) inter-cross-links, which are formed between peptides that have originated from different proteins.[22,62,74,81] More than one type of modification can occur on a given peptide, further complicating the range of cross-linked products that may form. Difficulty in identifying the useful intra- and inter-cross-links hidden amongst this complex mixture of peptides hampers the sensitivity of the MS experiment.

The length of the spacer arm establishes an upper distance constraint between the two reacting residues, as a cross-link will only form if the two reactive sites can be reached in three-dimensional space.[58,59,64,74] Modifying the number of atoms between the two reactive groups changes the upper limit on the constrained distance between cross-linked residues. Furthermore, additional modifications to the spacer arm can change the functionality of the cross-linking reagent, assisting in analytical simplification.

Steps for isolation of cross-linked peptides from an excess of unreacted peptides can be implemented to increase MS sensitivity prior to analysis during the enrichment step of the XLMS workflow (Figure 1.8). This has been achieved previously using electrophoretic or chromatographic methods, but more selective separation is obtained from the incorporation of affinity tags into the structure of cross-linkers which allows for enrichment.[62,65,67–70,80] Previously, biotin tags have been exploited for this purpose using avidin-biotin affinity purification. The biotin-avidin interaction is the strongest non-covalent interaction known in nature, so adding biotin to a cross-linker allows cross-linked peptides to be purified, for example, using streptavidin beads.[82] However, the increased steric bulk from biotin moieties can hinder cross-links from forming, and may distort native structures.[22,80,81,83,84] Copper-catalysed azide-alkyne cycloaddition (CuAAC) reactions are an efficient approach describing the high yielding reaction between the small and biologically inert azide and alkyne groups, and has been successfully applied in the enrichment of cross-linked peptides.[62,85] During this reaction, copper (I) catalyses a highly selective bio-orthogonal 1,3 dipolar cycloaddition reaction between an azide and alkyne group, resulting in a stable triazole product, as detailed in Scheme 2.1.[86] Incorporating such groups within the spacer arm of a linker allows post-linkage modifications to be carried out using CuAAC. For example, enrichment of cross-linked peptides may be carried out using a solid support presenting the orthogonal group.[62,81,87] Cleavage of the tags from the resin allows release of the cross-linked peptides for increased sensitivity in downstream analysis.

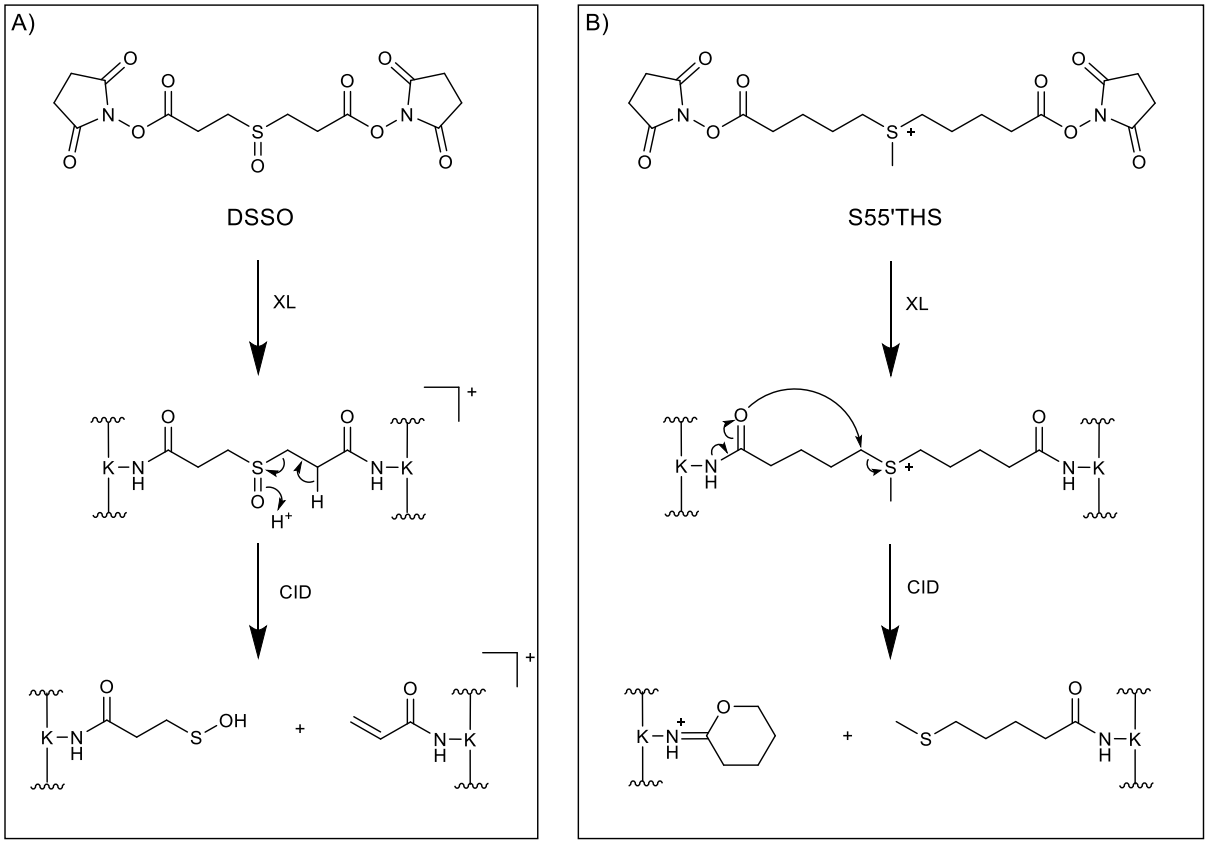


Scheme 2.1: Generalised copper catalysed alkyne azide click reaction.

Most XLMS work to date has been performed using non-cleavable amine reactive cross-linkers.[59] This can produce complex datasets, as for each cross-linked product, two peptides are being sequenced simultaneously.[22,88] Interpretation of this data requires powerful tools to consider every possible combination of peptide pairs which match the mass of the cross-

linked product.[22,59] Introducing a cleavable group into the spacer arm can assist in simplification of the spectra obtained in the peptide sequencing step of the XLMS workflow (Figure 1.8). Such groups allow separation of cross-linked peptides prior to sequencing, yet the cross-linked residues remain labelled, permitting unambiguous peptide identification with site-specific linkage information.[59,62] Depending on the cleavable group, linkers may be cleaved chemically or preferentially through CID within the mass spectrometer which reduces sample handling.[58,62] Analysis of cleavable cross-linked peptides occurs at three stages; initially, peptides are detected at the MS^1 level before being selected for MS^2 analysis and undergoing CID fragmentation to separate cross-linked peptide constituents using the cleavable bond. Subsequently, peptides are sequenced through fragmentation at the MS^3 level along the peptide backbone for sequence identification.[58,59,62,71] These cleavage products possess distinct fragmentation patterns which assist in peptide identification.

Although data is simplified using cleavable linkers, dedicated software tools such as XLinkX and MeroX must be used to assist in the analysis.[59] Disuccinimidyl sulphoxide (DSSO) is an example of a commercially available cleavable cross-linking reagent which has been utilised in literature recently.[72] It is a homobifunctional linker comprised of two reactive NHS esters, a moderate length carbon chain and a sulphoxide group which introduces two cleavable sites into the linker as detailed in Scheme 2.2 A. Alternatively, the homobifunctional cross-linker *S*-methyl 5,5'-thiodipentanaylhydroxylsuccinimide (S55'THS) synthesised by Lu et al[85] possesses a cleavable sulphonium ion group, presented in Scheme 2.2 B.[85,89] Upon CID, this linker produces a 6-membered iminotetrahydropyral ion with a mass of +83 Da, and a *S*-methylthiopentanoyl group with a mass of +130 Da.[85]



Scheme 2.2: MS/MS fragmentation of previously reported cleavable cross-linkers.
Cleavage products of the sulphoxide group in DSSO (**A**) include the corresponding sulphenic acid and alkene moieties. Cleavage products of the sulphonium ion in S-methyl 5,5'-thiodipenanaylhydroxylsuccinimide (**B**) include the corresponding 6-membered iminotetrahydropyral ion and a sulphide moiety.

This chapter investigates the design and synthesis of a generic modular cross-linker structure that can selectively incorporate features based on the desired application. The features incorporated within this design enable enrichment and cleavage steps described in the bottom-up XLMS workflow (Figure 1.8). Using well-characterised chemistry, this can generate a library of derivatives with different functionalities which can be applied to various biological systems. Furthermore, this chapter also describes work aimed at optimising the cross-linking reactions for the newly synthesised linker library. Initially, the reactivity of both specific and non-specific reactive groups was assessed by reacting the linkers with N^{α}-Boc-L-lysine. Once an understanding of the XL reactivity was obtained, successful linkers were applied to a simple protein system, lysozyme, to assess their reactivity *in vitro*. The XL reaction was also carried out using DSSO (Scheme 2.2 A) to enable comparison to a commercially available linker. Once the conditions required for XLMS have been optimised, these linkers can be applied to

increasingly complex venom systems to derive new information regarding the higher-order structures of protein components.

2.2 Development of a Modular Synthetic Protocol

A modular synthetic protocol (Figure 2.2 A) was developed to synthesise a library of linkers incorporating three major components: a linker arm (**X**) dictating the distance between two cross-linked residues and which may incorporate additional features such as cleavable groups; a core unit which in this library is an affinity tag moiety (**Y**) used for enrichment of cross-linked peptides prior to analysis; and finally two reactive groups (**Z**) which covalently link the target amino acid residues together. The synthetic protocol is based on simple hydrocarbon chains and modified amino acid building blocks which can easily be coupled using well-established chemistry. Incorporating different variations of these building blocks allows inclusion or exclusion of different linker functionalities depending on the desired cross-linker application, enabling access to a diverse linker library through accessible chemical synthesis.[89]

In this general strategy, following synthesis and deprotection of the functionalised building blocks, the spacer arm (**X**) is linked to the amino acid base unit (**Y**) through peptide coupling. Free acid groups can then be exposed by methyl ester hydrolysis before undergoing coupling to the reactive groups (**Z**) through esterification. In this thesis, the utility of this method will be demonstrated by synthesis of a library of homobifunctional and heterobifunctional linker derivatives with both specific and non-specific reactive groups and both cleavable and non-cleavable spacer arms, all of which are summarised by modular linker **1** (Figure 2.2 A). Examples of the building blocks utilised to create this library are summarised in Figure 2.2 B.

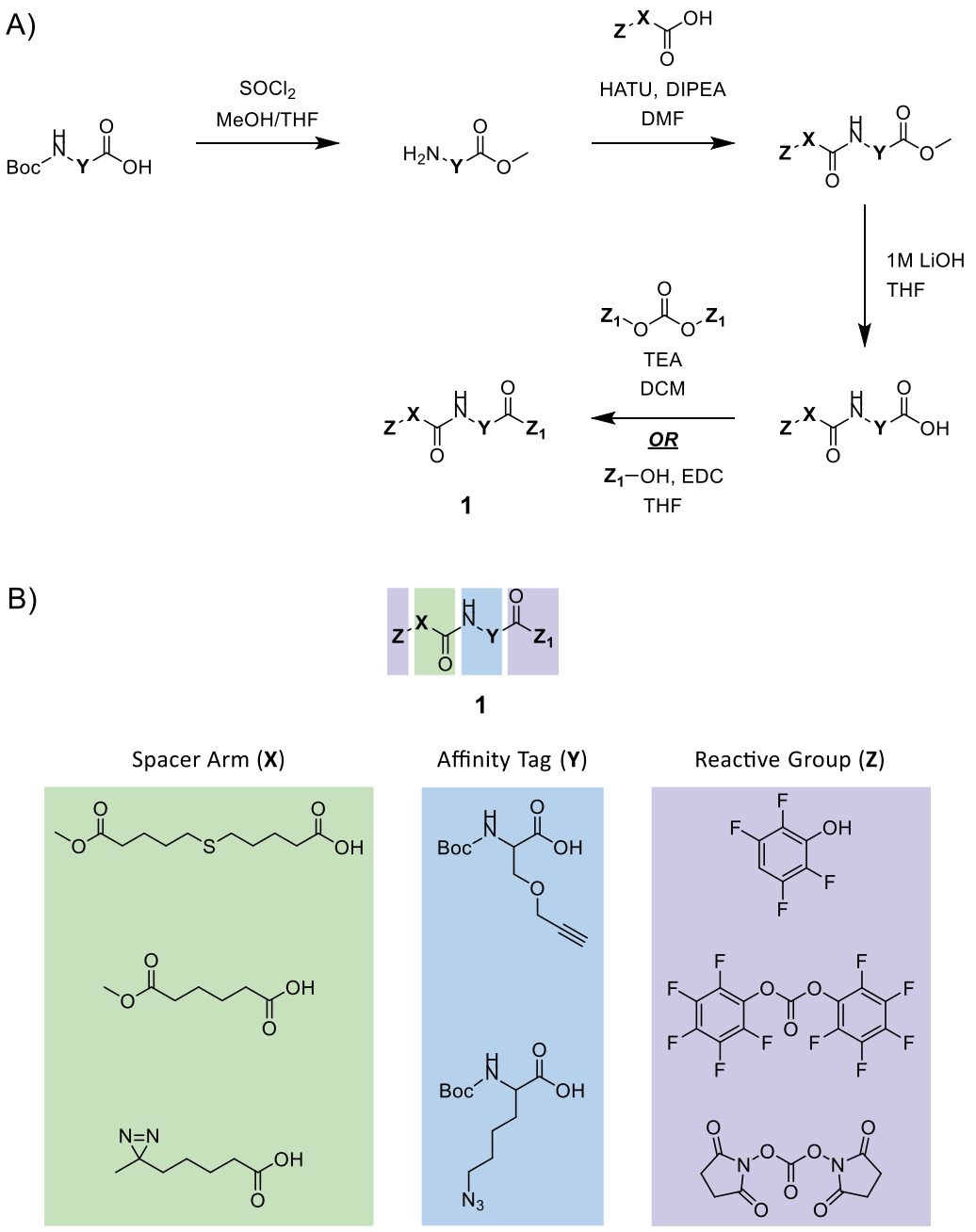
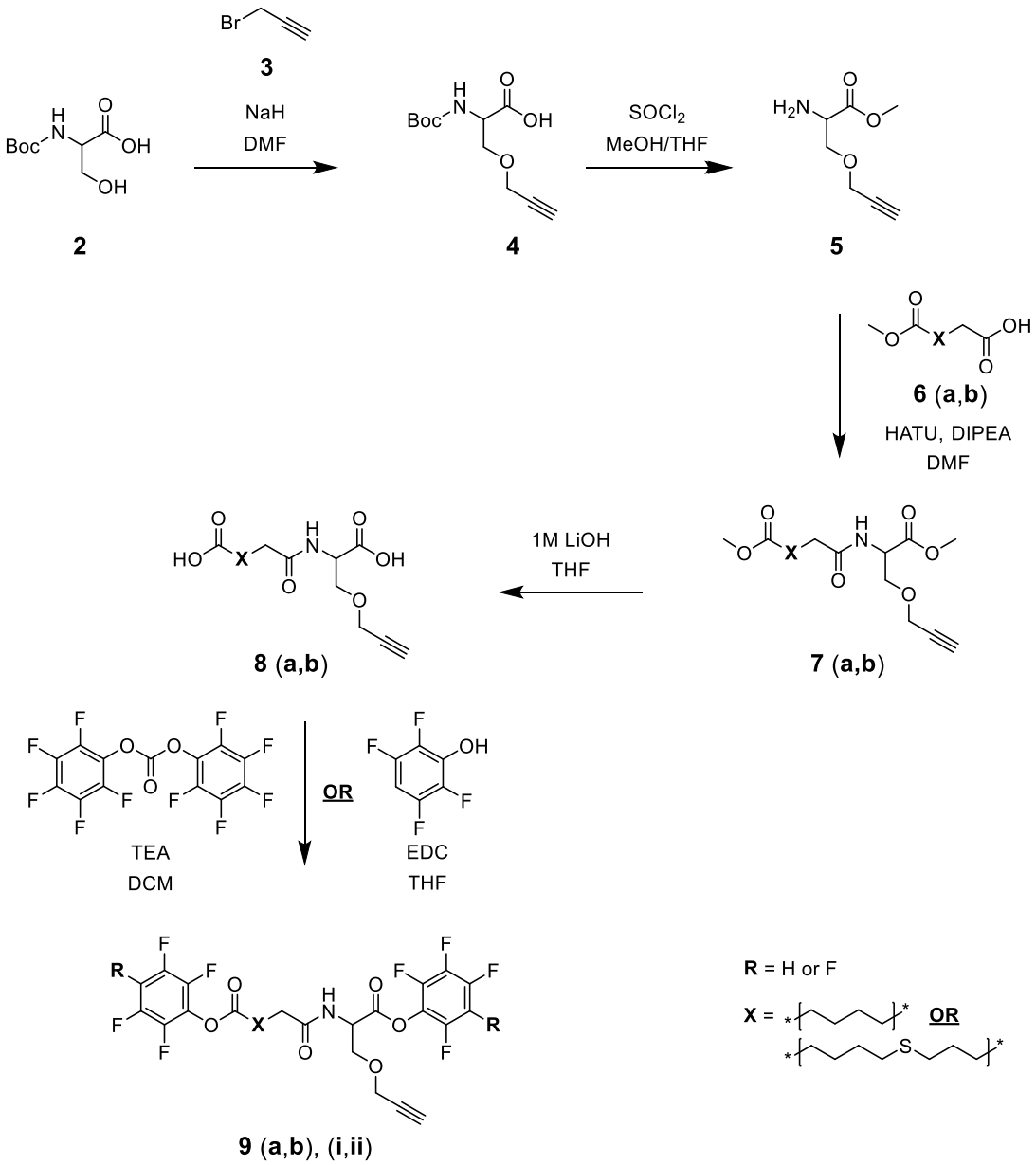


Figure 2.2 The modular cross-linker design and general synthetic route.

The modular synthetic protocol designed to synthesise the linker library in this thesis (**A**). This library is summarised by cross-linker **1**. Here, **X** represents different spacer arm functionalities including variable length and cleavable groups, **Z** and **Z_1**, represents the reactive groups incorporated within the linker and **Y** represents the affinity tag incorporated using a modified amino acid. Examples of the building blocks used to synthesise derivatives of modular linker **1** are detailed in (**B**). Three linker arm (**X**) moieties include S-methyl 5,5'-thiodipentanoic acid, monomethyl adipate and 5-(3-methyl-diazarin-3-yl) pentanoic acid. The two affinity tags (**Y**) are alkyne-modified N-Boc serine and azide-modified N-Boc lysine. The three reactive group reagents (**Z**) are TFP, PFP and DSC. It should be noted that **Z** groups within linkers may be incorporated by esterification (as shown by **Z_1**) or may be pre-incorporated within the spacer arm (**Z**).

2.3 Synthesis of Homobifunctional Linkers

The homobifunctional linker derivatives were synthesised first, using the synthetic strategy described in Scheme 2.3 below. Here, both a cleavable and non-cleavable homobifunctional linker derivative were designed and synthesised based on the modular structure in Figure 2.2. At each step, the cleavable derivatives are labelled **a,** whereas the non-cleavable derivatives are labelled **b**. Both a non-cleavable (**9bii**) and a CID cleavable (**10**) derivative were designed, by incorporation of the relevant structures within the spacer arm (highlighted in green). Both crosslinking reagents contain an alkyne affinity tag moiety linked to the spacer arm highlighted in blue. The synthesis and linkage of the spacer arm and affinity tag moieties are discussed in section 2.3.1 below. These linkers also incorporate two identical lysine-specific fluorinated phenyl reactive groups, either TFP or PFP, highlighted in purple. In the general scheme below, TFP esters are labelled **i**, whereas PFP esters are labelled **ii**. The synthesis of these protein coupling groups are discussed in section 2.3.2 below.



Scheme 2.3: Synthetic route to homobifunctional cross-linkers.

In the cleavable derivative, **X** equals $(CH_2)_4S(CH_2)_3$ derived from 5-[(5-methoxy-5-oxopentyl)sulphanyl]pentanoic acid **6a**, whereas in the non-cleavable derivative, **X** equals $(CH_2)_3$ derived from monomethyl adipate **6b**. When the reactive group is TFP, **R** equals H. When the reactive group is PFP, **R** equals F. The cleavable derivatives are labelled '**a**' and the non-cleavable derivatives are labelled as '**b**'. Derivatives synthesised using TFP are labelled as '**i**' whilst derivatives synthesised with PFP are labelled as '**ii**'.

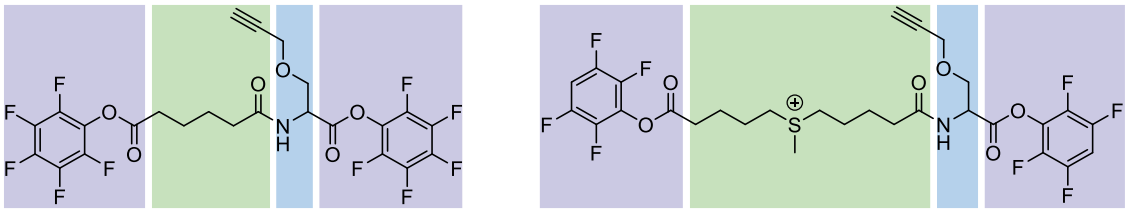


Figure 2.3: Non-cleavable (9bii) and cleavable (10) homobifunctional cross-linkers.
These cross-linking reagents are designed based on the modular structure in Figure 2.2 and are comprised of a spacer arm (green), and affinity tag moiety (blue) and two reactive protein coupling groups (purple).

Aside from using different reactive groups, the synthesis of both derivatives were analogous using different commercially available spacer arms. The non-cleavable linker was synthesised from monomethyl adipate, a simple short hydrocarbon chain possessing no cleavable functionalities, whereas the cleavable linker was synthesised using S-methyl 5,5'-thiodipentanoiec acid. The sulphur in this reagent can be converted into a cleavable sulphonium ion within the long spacer arm, allowing CID cleavage via the mechanism described for S-methyl 5,5'-thiodipenanaylhydroxylsuccinimide in Scheme 2.2. Following synthesis of the affinity tag moiety and coupling to the spacer arm, the protective ester groups were hydrolysed to reveal a diacid structure. Previous projects within the Pukala group attempting to synthesise similar linker derivatives meant than an excess of cleavable diacid **8a**, where **X** equals $(CH_2)_4S(CH_2)_3$, was already synthesised. Hence, the synthesis of the affinity tag moiety and coupling to the spacer arm discussed below describes that for the non-cleavable diacid **8b** derivative only, however the synthesis of **8a** was carried out in a similar manner.

2.3.1 Synthesis of the Affinity Tag Moiety and Spacer Arm

Firstly, the affinity tag moiety was synthesised. In the homobifunctional linker, this was achieved through *O*-alkylation of the C_β hydroxyl group of commercially available *N*-Boc-L-Serine (**2**) with propargyl bromide. Sodium hydride was used to deprotonate the free alcohol on **2** so that the resulting anion could react with propargyl bromide (**3**) to yield the desired alkyne (**4**) in an 89 % yield without further purification. The ^{1}H NMR spectrum for **4** showed a new triplet at δ 2.45 ppm identified as the propargyl alkyne group signal, as well as a new doublet at δ 4.15 ppm as assigned to the propargyl CH_2 group. The rest of the ^{1}H NMR data

was consistent with literature spectra.[89] Furthermore, HRMS analysis identified the sodium adduct of the linker, confirming successful synthesis.

Thionyl chloride was subsequently added to alkyne **4** to deprotect the *N*-terminus by cleaving the *N*-Boc group and convert the carboxylic acid into an acid chloride in a one pot reaction. The resulting acid chloride was then methylated to form the desired methyl ester **5** in a yield of 67 % without further purification. Absence of the *N*-Boc signal at δ 1.43 ppm and the presence of a new downfield singlet identified as the methyl ester signal at δ 3.48 ppm in the ^{1}H NMR spectrum suggested the desired product had formed.

The affinity tag and spacer arm moieties were linked using an amide coupling reaction between methyl ester **5** and monomethyl adipate (**6b**) to give dimethyl ester **7b** in a yield of 26 %. This reaction was carried out in dimethylformamide (DMF) using 1-Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium-3-oxide hexafluorophosphate (HATU) as the coupling reagent and N,N-Diisopropylethylamine (DIPEA) as the base. Protecting the non-participating acid and amine groups prevented the formation of undesirable side products and permitted purification by silica chromatography. Comparison of the ^{1}H NMR spectrum for **7b** with that for monomethyl adipate **6b** and the serine methyl ester **5** identified the additional serine methyl ester signal at δ 3.52 ppm, along with monomethyl adipate hydrocarbon multiplets observed between δ 2.23-2.13 ppm and δ 1.58-1.49 ppm. This, along with observation of the product ion by HRMS analysis indicated successful coupling.

Following isolation of **7b**, the protective methyl groups were removed by hydrolysis with lithium hydroxide in a yield of 84 % with formation of the corresponding diacid **8b** indicated by an absence of the methyl ester peaks in the ^{1}H NMR. This free diacid was then capable of undergoing esterification to incorporate the reactive groups.

2.3.2 Reactive Group Incorporation

In the proposed synthetic strategy, the reactive groups were to be integrated into the linker structure by esterification of free acids. Previous projects within the Pukala group attempted to synthesise a cleavable sulphonium ion linker using NHS as the reactive group. Unfortunately, difficulties were reported in the purification of the NHS esterified derivative, potentially due to hydrolysis of this labile group. Given that an excess of diacid **8a**, where **X**

equals $(CH_2)_4S(CH_2)_3$, was already prepared from these projects, esterification of the cleavable homobifunctional linker derivative was attempted first.

Initially, TFP was used as the reactive group as it possesses a distinct downfield multiplet observable in ^{1}H NMR analysis, simplifying detection of the desired product during synthesis. The TFP esterification conditions were first optimised on the free acid group of *N*$^{\alpha}$-Fmoc-*N*$^{\varepsilon}$-Boc-L-lysine (**11**) since it is easily accessible, affordable, and mimics the steric bulk at the acid group present in the final linker.

The first esterification of *N*$^{\alpha}$-Fmoc-*N*$^{\varepsilon}$-Boc-L-lysine **11** with TFP was attempted using a one-pot halogenation and subsequent esterification reaction based on the protocols of Davis and Linxis.[90,91] The acid group of lysine was converted into an acyl chloride intermediate using thionyl chloride to increase the susceptibility of the carboxyl group to nucleophilic attack by the deprotonated TFP hydroxyl group. The crude mixture was purified via silica chromatography but did not yield the desired product.

In a second attempt, an alternative coupling reagent dicyclohexylcarbodiimide (DCC) was reacted with *N*$^{\alpha}$-Fmoc-*N*$^{\varepsilon}$-Boc-L-lysine **11** and TFP overnight. The mildly basic character of DCC deprotonates the carboxyl group of *N*$^{\alpha}$-Fmoc-*N*$^{\varepsilon}$-Boc-L-lysine **11** for reaction with the carbodiimide to form an *O*-acylisourea. This group is lost upon nucleophilic attack of the acid group by TFP to give the desired ester.[92] The crude residue was purified by silica chromatography to give the desired lysine ester **12i** in a 20 % yield. It should be noted that some co-elution was observed, yet pure fractions were collected exclusively. Product **12i** was identified using ^{1}H and ^{13}C NMR spectra where a new TFP multiplet at δ 7.06-6.97 ppm was observed.[93,94]

With the conditions established for successful esterification on *N*$^{\alpha}$-Fmoc-*N*$^{\varepsilon}$-Boc-L-lysine **11**, the same reaction was carried out on **8a** in attempt to generate the diester linker, but with 3 molar equivalents of both TFP and DCC used to compensate for the presence of two acid groups. The crude ^{1}H NMR spectrum showed 2 TFP peaks, one slightly downfield at δ 7.31 ppm and the other slightly upfield at δ 6.57 ppm (1H) to free TFP, suggesting the formation of an excess of diester and a small amount of monoester. The crude mixture was purified by silica chromatography but the diester could not be identified in the eluted fractions by NMR or HRMS.

The reaction was next attempted using the alternative coupling reagent propanephosphonic anhydride (T3P), traditionally used in peptide synthesis.[95] Diacid **8a** was dissolved in anhydrous ethyl acetate in the presence of TEA, T3P and TFP. Deprotonation of the carboxylic acid groups of **8a** under these basic conditions resulted in attack of the phosphorous in T3P, forming an electrophilic moiety. Reaction with the nucleophile of interest, TFP, ideally forms the desired ester whilst releasing the water-soluble T3P by-product. The crude ^{1}H NMR saw three TFP signals, two slightly downfield (δ 7.34 and 7.03 ppm) and one slightly upfield (δ 6.63) ppm of free TFP usually observed at δ 6.72 ppm, potentially indicating formation of the diacid. To confirm this, attempts were made to purify components of the crude mixture, however, the desired TFP diester was not recovered. Only trace quantities of the TFP monoester and starting material were detected by ^{1}H NMR and HRMS analysis.

It is possible that the difficulty in esterification may be due to an incompatibility of the linker with these conditions. Whilst the TFP group was expected to modify the free linker arm, the bulky azide tag may have prevented the addition of the TFP group at the serine end of the linker. Additionally, it is possible that the long linker arm containing the sulphur group may be susceptible to degradation or may be interfering with the TFP esterification. Hence, esterification was next attempted on the shorter, non-cleavable derivative **8b** to determine whether this was the case.

Literature protocols demonstrating success in NHS esterification using the carbonate precursor DSC highlight the potential of using a carbonated fluorinated phenol group to introduce the reactive group.[79] Unfortunately, TFP was not available as a carbonate, so PFP carbonate was purchased. This group has an additional fluorine in place of the proton in TFP meaning it cannot be detected in ^{1}H NMR analysis but should possess similar chemoselectivity and reactivity. The additional fluorine may even further stabilise the PFP group through resonance and inductive effects which may be advantageous during linker storage.

Originally, esterification of **8b** was attempted using PFP carbonate in the presence of DIPEA.[96,97] During this reaction, a colour change from colourless to pink to dark red was observed, and two new spots appeared in TLC analysis. These products were isolated by silica chromatography, but ^{1}H NMR, ^{13}C NMR and HRMS analyses revealed that only starting material and free PFP-OH were recovered. An excess of free hydrolysed PFP molecules may swamp the detection of trace quantities of desired product. Hence, the reaction was repeated

by reducing the equivalents of PFP carbonate and increasing the equivalents of DIPEA to deprotonate the acid **8b**. The crude product was purified on TEA deactivated silica to reduce product degradation but NMR and HRMS analysis revealed that this did not afford the desired PFP-ester **8b**.

Although some protocols leave this reaction to proceed for between 3-6 h, the deep colour change observed within the first 30 min suggested the reaction may be occurring much earlier than expected, leading to product degradation if in solution for too long. Using N^{α}-Fmoc-N^{ε}-Boc-L-lysine **11** as a model system, the reaction was repeated over a shorter time-period of 30 min to determine whether this early colour change was indicative of the desired product. Reaction of this model system produced a colour change from colourless to yellow over a similar timeframe observed for **8b**. ^{1}H NMR, HRMS and TLC analysis of **11ii** confirmed the reaction was almost complete within the first 30 min, suggesting the colour change observed is indicative of product formation.

The reaction was therefore repeated on **8b** under the same conditions and the reaction mixture was analysed following observation of the colour change within the first 30 min. Notably, the desired product was again not detected by ^{1}H NMR and HRMS analysis. Despite that these literature methods have been successful for PFP esterification of other monoacids, these conditions were determined to be unsuitable for this diacid.

As a final esterification attempt using PFP carbonate, the diacid precursor **8a** was used. Although this reaction is generally carried out over several hours, one literature protocol suggested extended reaction duration may be appropriate.[97] This may increase the opportunity for the PFP group to join at both ends, even in the presence of the bulky alkyne tag. Hence, the reaction was left overnight, and TEA was used as an alternative base to DIPEA. It was difficult to detect the desired product upon ^{1}H NMR and MS analyses of the crude residue, so a small sample was purified on a microcolumn using phosphate buffered silica. Deactivated silica, which is treated with phosphoric acid to deactivate the polar silica groups and then neutralised using disodium hydrogen phosphate, allows enough interactions to separate components whilst reducing degradation of the desired product during purification.

Comparing the ^{1}H NMR spectra of the starting material **8a** and the crude **9aii** spectra assisted in the proton signal assignment for the purified product **9aii**, however it was difficult to specifically assign all signals. The consistency between the downfield chemical shifts and

signal integrations with that expected for the desired product made proton assignment of these signals straightforward. However, additional signals between δ 2.8-1.5 ppm with large integrations suggested that the product was not pure, making it difficult to assign the upfield protons in the hydrocarbon chain using 1D data alone. Therefore, homonuclear correlation spectroscopy (COSY), which permits identification of neighbouring protons up to four bonds away, was performed on the crude and purified sample (Figure 2.4 A and B). This assisted in identifying signals corresponding to impurities which may have been partially removed or incorporated during the purification process. Signals A1 and B1 between 2.84-2.76 ppm have been identified as the alkyne signal which is predicted to exist in this range. Given that this atom is more than 4 bonds away from any other upfield protons in the product structure, the small correlation with B2 is unlikely to be from the desired product. In the purified product, both B2 and B3 only correlate with B5, suggesting these three signals make up the carbon chain of the cleavable arm. A2 and B2 is likely to be the carbon adjacent to the carbonyl group due to its triplet character, however specific assignment of the B3 and B5 and the A3 and A5 signals was difficult. A5 and B4 both show correlations with A6 and B5, however also show a correlation with A1 and B1. Given that A1 and B1 have been assigned as the alkyne group which is not localised near the hydrocarbon chain, it is predicted that this correlation is from contaminants.

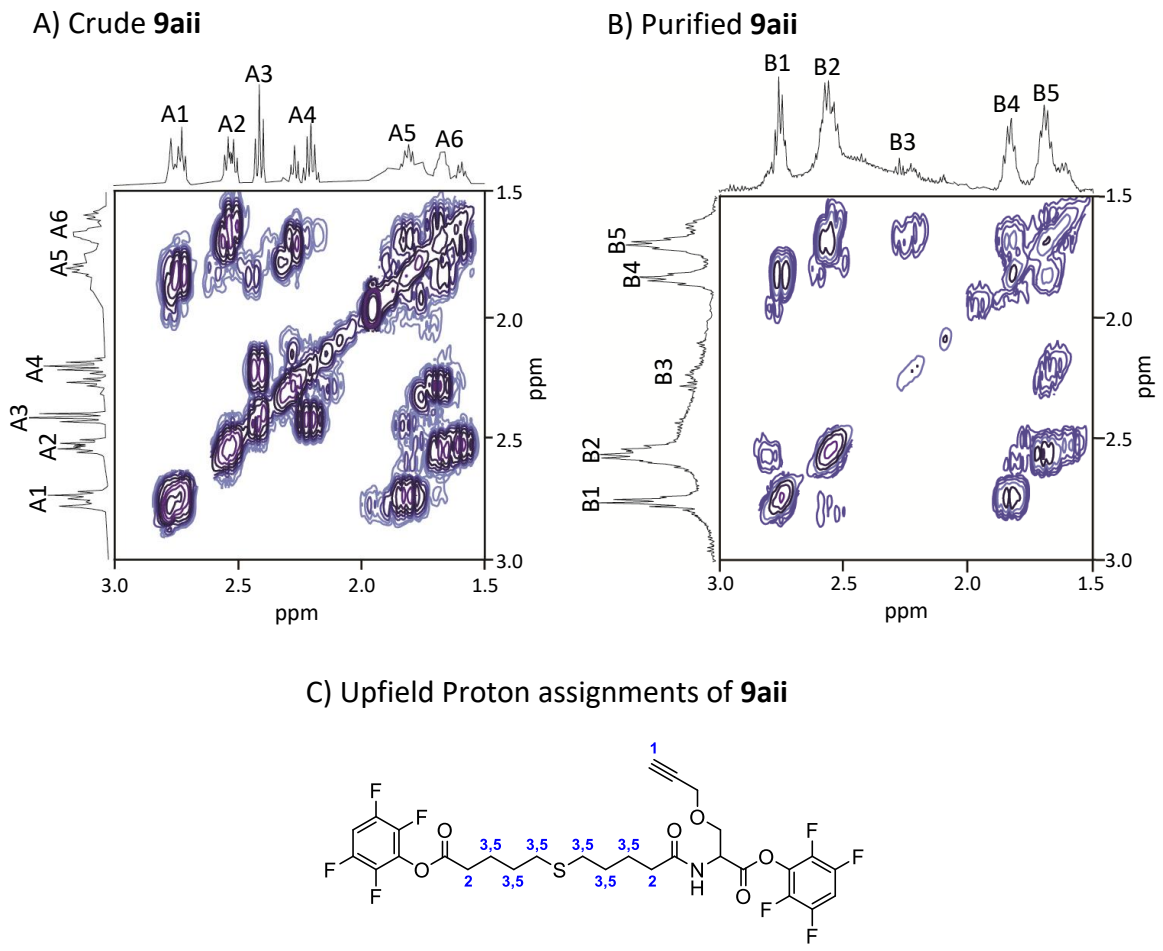


Figure 2.4: Partial COSY spectra for the characterisation of 9aii at 500 MHz in CD_3CN.
Comparison between the crude product and the semi purified product, with major peaks labelled A1-6 and B1-5 for clarity.

It must be noted that a larger number of peaks were observed in the ^{13}C NMR than expected based on the unique number of carbons in the proposed structure. It is likely this has occurred due to carbon-fluorine coupling in the PFP group.[98]

The esterification methods attempted initially in this project involved adding excess base. It is possible that the basic conditions may be causing hydrolysis of the desired product, preventing isolation. A study by McLaughlin *et. al.* detailed the synthesis of aryl-esters through a base-free approach using coupling reagent 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC).[99] The mildly basic character of this coupling reagent means no other base is required, reducing the chance of base-induced-hydrolysis. Hence, EDC was next used in attempts to couple the free TFP alcohol to diacid **8a** to form **9ai**. TLC analysis of the crude reaction mixture was complex making it difficult to identify exactly

which spot was the desired product. However, given that the expected product ion was observed via HRMS and that the ^{1}H NMR and ^{13}C NMR spectra suggested some of the desired product had formed, a small amount of the crude product was set aside and the rest was purified by silica chromatography using phosphate buffered silica. Again, the cross-linker could not be purified completely from the starting material making assignment of the NMR spectra difficult. However, signals consistent with the TFP proton at δ 7.36-7.28 ppm, the affinity tag between δ 7.06-3.88 ppm, and the cleavable hydrocarbon chain between δ 2.81-1.52 ppm were observed. The latter signals were consistent with peaks B1, B2, B4 and B5 from the ^{1}H NMR spectra of **9aii** in Figure 2.4 and the starting material spectra, suggesting that product **9ai** had been synthesised.

Given the success of the EDC reaction, these conditions were applied to diacid **8b** on a small scale to afford trace quantities of crude **9bi**. Despite that the product ion was observed by HRMS analysis and a single spot was observed through TLC analysis, the very small scale that this reaction was performed on meant that not enough product was formed for confirmation of structure by ^{1}H NMR. Furthermore, the product was not purified due to risk of losing the small quantity of product.

Due to time constraints and limited resources, no further attempts at synthesis of the homobifunctional crosslinkers were made. Although limitations existed regarding the purification of the final linkers, the crude mixture of **9bi** and the semi-purified linkers **9ai** and **9aii** were applied to protein systems directly without further purification.

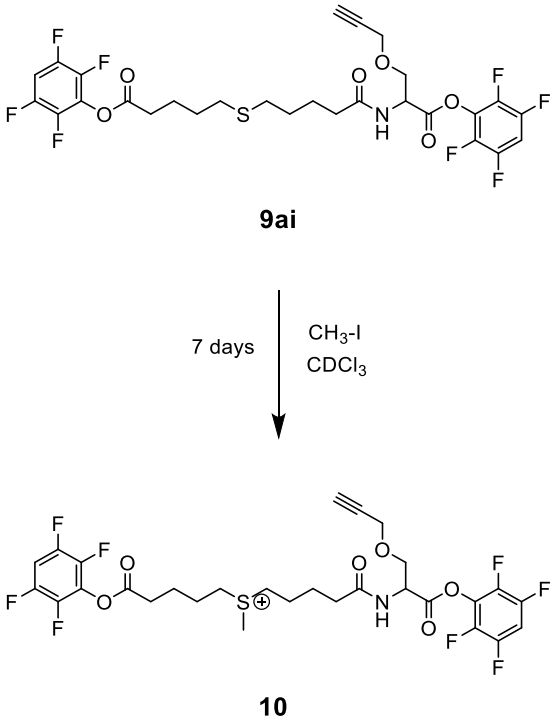
Each of the esterification conditions applied to diacids **8a**, **8b** and lysine **11** are summarised in Table 2.1. Future experimentation may involve investigating different methods of purification which will not degrade the moderately labile ester group, experimenting with other coupling reagents, and carrying out the PFP carbonate reaction with less equivalents of base in attempt to increase the yield and the purity. Given that yield optimisation was not the focus of the project and enough of the novel linker was synthesised to apply to a biological system, this was not explored here, but could be investigated in further studies.

Table 2.1: TFP and PFP esterification conditions applied to diacids 8a, 8b and lysine 11.
Lysine was N^{α}-Fmoc-N^{ε}-Boc protected. Starting material, reagents and time (o/n: overnight) of reaction are specified, along with assessment of the success of the reaction (synthesis of the crude product) and the purification (isolation of the crude product and percent yield recovered) (success: ✓, no reaction: ✗, not attempted: -).

STARTING MATERIAL	ADDITIONAL REAGENTS	REACTION TIME	CRUDE PRODUCT	PURIFICATION OF PRODUCT
Lysine **11**	$SOCl_2$, TFP, DIPEA	o/n	✗	✗
Lysine **11**	DCC, TFP	o/n	✓	✓(20 %)
Diacid **8a**	DCC, TFP	o/n	✗	✗
Diacid **8a**	T3P, TFP, TEA	o/n	✗	✗
Diacid **8b**	PFP carbonate, DIPEA	4 h	✗	✗
Diacid **8b**	PFP carbonate, DIPEA	5 h	✗	✗
Lysine **11**	PFP carbonate, DIPEA	30 min	✓	-
Diacid **8b**	PFP carbonate, DIPEA, TEA	30 min	✗	✗
Diacid **8a**	PFP carbonate, TEA	o/n	✓	✓(10 %)
Diacid **8b**	EDC, TFP	o/n	✓	-
Diacid **8a**	EDC, TFP	o/n	✓	✓(4 %)

2.3.3 Formation of the Cleavable Sulphonium Ion

The final step in synthesising the cleavable linker was methylation of the sulphur atom in **9a**, as detailed in Scheme 2.4. This reaction was only carried out using **9ai** due to the ease in purification and higher yield (recovery of 8 mg). Diester **9ai** was successfully methylated by reaction with methyl iodide for one week.[85]



Scheme 2.4: Synthetic scheme for the methylation of the sulphur group in linker (9ai).

As this cleavable linker was designed to fragment selectively by cyclisation under increased collision energies, observation of the cleavage patterns in the MS/MS analysis of the purified linker further confirms successful synthesis. Figure 2.5 shows that at low collision energy, the parent ion mass for the linker can be principally detected (670 *m/z*). However, as the collision energy is increased, a signal representing cleavage and cyclisation at the modified serine end of the linker can be observed. Notably, although the linker has the potential to cyclise on either side of the sulphonium ion upon CID, only the cleavage product where the affinity tag is incorporated within the cyclic ring is observed. This suggests the alternative cyclisation product ion energetically less favourable and the alternative mechanism dominates, as observed in Figure 2.5.

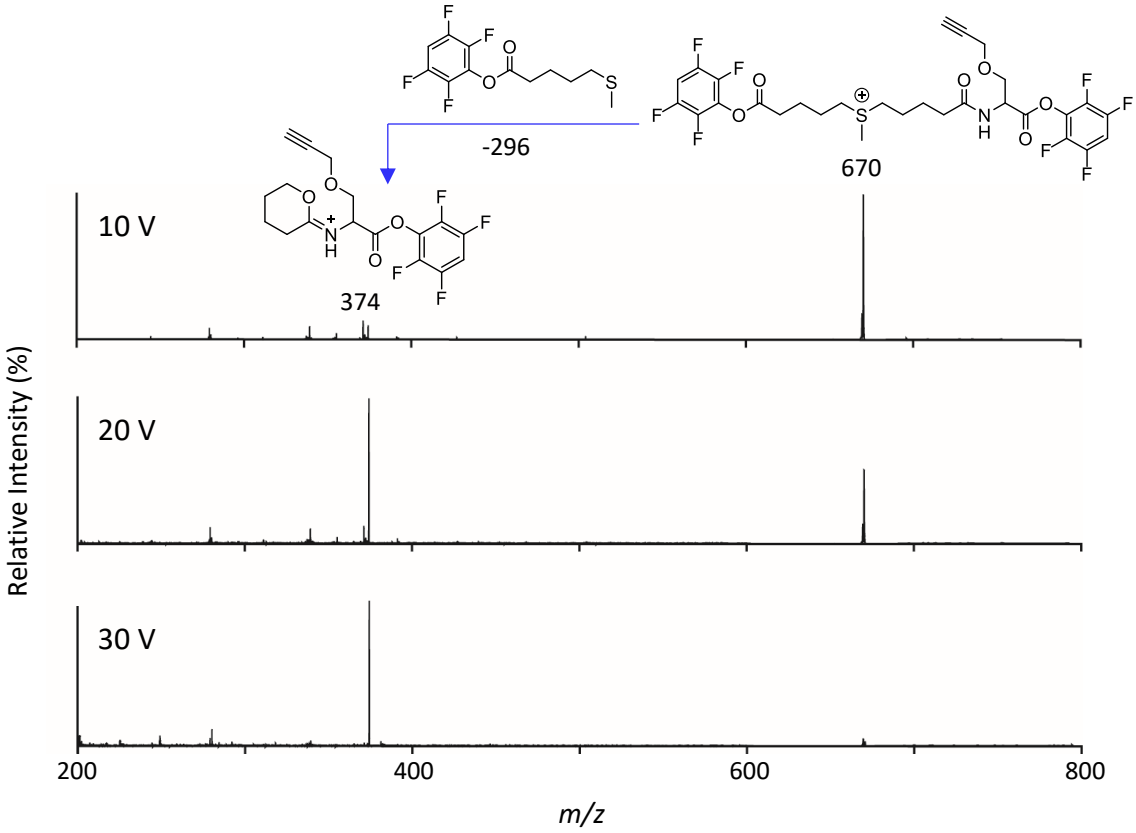
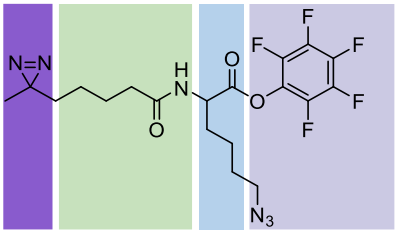


Figure 2.5: MS/MS spectra of linker (10) reveals selective fragmentation by CID.
Collision energy was increased from 10 to 30 V to observe cleavage of linker **10** ($[M+H]^+$ = 670 *m/z*) into the diagnostic cleavage product ($[M+H]^+$ = 374 *m/z*), following neutral mass loss of 296 *m/z.* Analysis was performed in 50 % ACN.

2.4 Heterobifunctional Linker Synthesis

Following successful synthesis of the homobifunctional linkers, a non-cleavable heterobifunctional linker was designed as a target for synthesis based on the modular synthetic strategy. This structure is detailed in Figure 2.6 which shows a short non-cleavable arm highlighted in green and an azide affinity tag moiety highlighted in blue. The synthesis of these moieties is discussed in section 2.4.1. This linker incorporates one lysine specific group as seen in linkers **10** and **9bi** (highlighted in light purple) and a non-specific diazirine reactive group highlighted in dark purple. The diazirine is activated under UV light and hence, following formation of this group, subsequent synthesis was carried out in darkness. Incorporation of the non-specific diazirine group is discussed in section 2.4.1 with the

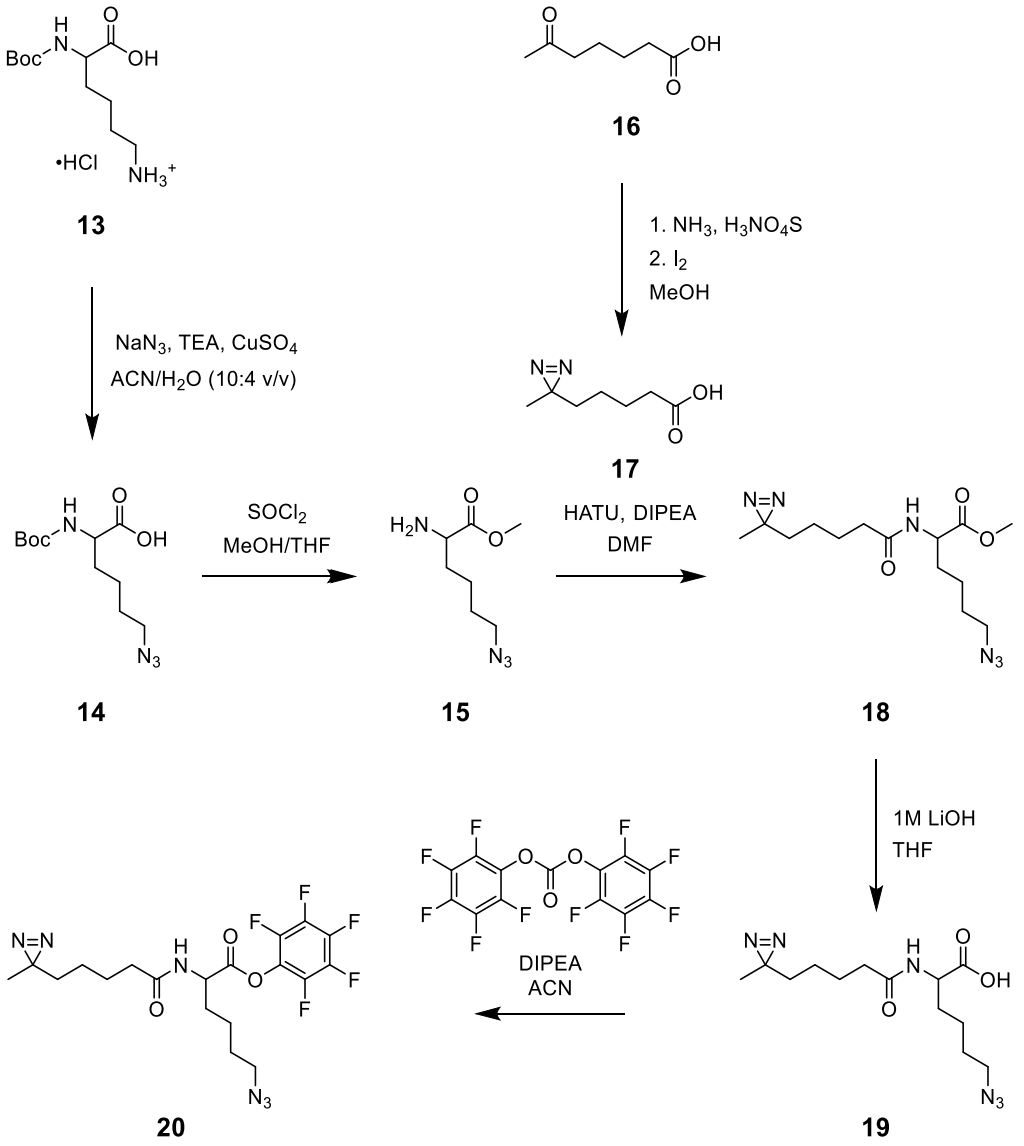
synthesis of the spacer arm and incorporation of the specific ester group discussed in section 2.4.2.



20

Figure 2.6: Heterobifunctional linker design.
This cross-linking reagent is designed based on the modular structure in Figure 2.2. This linker is comprised of a non-specific diazirine reactive group (dark purple), a specific PFP reactive group (light purple), a spacer arm (green), and an affinity tag moiety (blue).

The specific synthetic route used to generate this linker is given in Scheme 2.5. The azide tag was incorporated into a lysine moiety and the diazirine group was incorporated into a short ketone. The spacer arm and affinity tag moieties were connected through an amide coupling reaction before the protective ester group was hydrolysed to reveal the acid group for PFP esterification.



Scheme 2.5: Synthetic route to the heterobifunctional cross-linker.
Synthesis is based on the general linker synthetic route detailed in Scheme 2.1. Rather than introducing both reactive groups by esterification, the non-specific diazirine group was incorporated within the spacer arm in early synthetic steps.

2.4.1 Synthesis of the Affinity Tag Moiety and Spacer Arm

Previous projects within the Pukala group had prepared an excess of the azide *N*-Boc-L-lysine moiety (**14**). This was generated through an azide transfer reaction on the primary amine of a *N*-Boc-L-lysine (**13**) side chain in the presence of sodium azide, copper sulphate and TEA.[100] The carboxyl group of the modified lysine **14** was protected by methylation using thionyl chloride and methanol. This also cleaved the *N*-Boc group, freeing the amine to react in

subsequent steps. The resulting modified lysine **15** was collected in a yield of 84 % without further purification. Disappearance of the *N*-Boc signal previously visible at δ 2.15 ppm and addition of the methyl ester singlet at δ 3.85 ppm in the ^{1}H NMR spectrum confirmed that the desired product was formed.

The nonspecific diazirine functional group was incorporated into the linker arm by modifying the ketone of 5-acetovaleric acid (**16**). A diaziridine group was first formed by reacting 5-acetovaleric acid with ammonia solution and hydroxylamine-o-sulphonic acid (HAOSA). Subsequently, the diaziridine was oxidised using iodine to form the desired diazirine **17** in a yield of 14 %. This product was detected by HRMS in negative mode and confirmed by ^{1}H NMR analysis. Substantial upfield shifts in the methyl group (from δ 2.15 ppm to δ 1.02 ppm) and methylene group (from δ 2.38 ppm to δ 1.62 ppm) proton NMR signals were observed. Such shifts are consistent with previous studies which have reported successful synthesis of diazirines from carbonyl precursors.[57] This indicates successful formation of the diazirine product. Furthermore, the ^{13}C NMR spectrum showed a loss of the carbonyl signal (δ 211.6 ppm) and gain of a signal consistent with a carbon adjacent to a diazirine group (δ 36.6 ppm).

The affinity tag and spacer arm moieties were linked using an amide coupling reaction between the free amine group of **15** and the free acid group of **17**. This was performed in THF using the coupling reagent HATU and base DIPEA. The crude residue was purified by silica chromatography to give the desired product in a yield of 24 %. Protection of the acid group in **15** meant that the desired product **18** could be purified by silica chromatography and that **15** would not self-react. Following isolation of **18**, the protective methyl ester groups were removed by hydrolysis with lithium hydroxide to give the desired acid **19** in a yield of 93 %. ^{1}H NMR analysis showed the successful removal of the methyl group (δ 3.62 ppm) and the presence of a new broad acid peak at δ 9.95 ppm.

2.4.2 Specific Reactive Group Incorporation

Given that many cross-linkers in the literature use NHS as the reactive group, an NHS esterification reaction was first carried out using DSC and DIPEA in ACN. The reaction was monitored by TLC but reaction completion was difficult to determine due to streaking along the plate. Hence, the reaction was left overnight at room temperature. Initially, the crude

residue was purified by silica chromatography using 40 % ethyl acetate in hexane. Very little product was recovered so fractions were combined and purified on a smaller scale using 10 % ethyl acetate in petroleum benzene. Unfortunately, ^{1}H NMR analysis did not reveal any fractions containing signals for both the linker backbone and the NHS ester. This is consistent with claims that NHS esters are labile, complicating purification.[101]

The esterification reaction was repeated using a fluorinated benzene group. During the synthesis of the homobifunctional linker PFP, esterification of two acid groups took much longer to complete than the literature suggested, even with an increase in equivalents. Given that many papers have reported success in PFP esterification with PFP-carbonate on monoacids, and that *N*-Boc-L-lysine was easily esterified using PFP-carbonate earlier, the esterification of monoacid **19** was next carried out using PFP-carbonate. This reaction was performed over 5 h using PFP-carbonate and DIPEA in ACN, and the crude residue was purified by silica chromatography using 30 % ethyl acetate in hexane to give **20** in a 41 % yield. The product was detected by HRMS analysis, and ^{1}H NMR showed loss of the broad OH signal (δ 9.95 ppm), a slight upfield shift of the NH proton (from δ 6.49 ppm to δ 6.15 ppm) and a slight downfield shift of all other CH signals, likely due to the inductive effects of the PFP aromatic group. Given that the PFP group could not be detected by ^{1}H NMR, ^{19}F NMR was performed on the purified fractions. Six distinctive fluorine peaks were observed, three of which are predicted to be free PFP. Although this indicates mild degradation of the product, it is evident that the desired linker was successfully synthesised.

2.5 Optimisation of the Cross-Linking Reaction

2.5.1 Reactive Group Activation

XLMS aims to probe and map protein interactions involved in multiprotein networks. Therefore, it is critical that cross-linking reagents successfully form covalent linkages between associated protein species. Although the synthesis of novel crosslinking reagents was successful, optimisation of the XL reaction is necessary to consider chemical and physical compatibility with protein systems.

Activation of Fluorophenyl Groups

To experimentally demonstrate the efficient bio-conjugation of fluorophenyl groups to the primary amine of lysine residues, linker **20** was reacted with N^{α}-Boc-L-lysine in a linker to lysine ratio of 4:1. The MS spectrum detailed in Figure 2.7 indicates successful linkage, revealing both proton and sodium adducts of the linker-lysine adduct at 539 and 561 *m/z* respectively. The efficiency of this reaction was further supported by the lack of signals indicative of starting material N^{α}-Boc-L-lysine at 247 *m/z* and free linker **20** at 478 *m/z*.

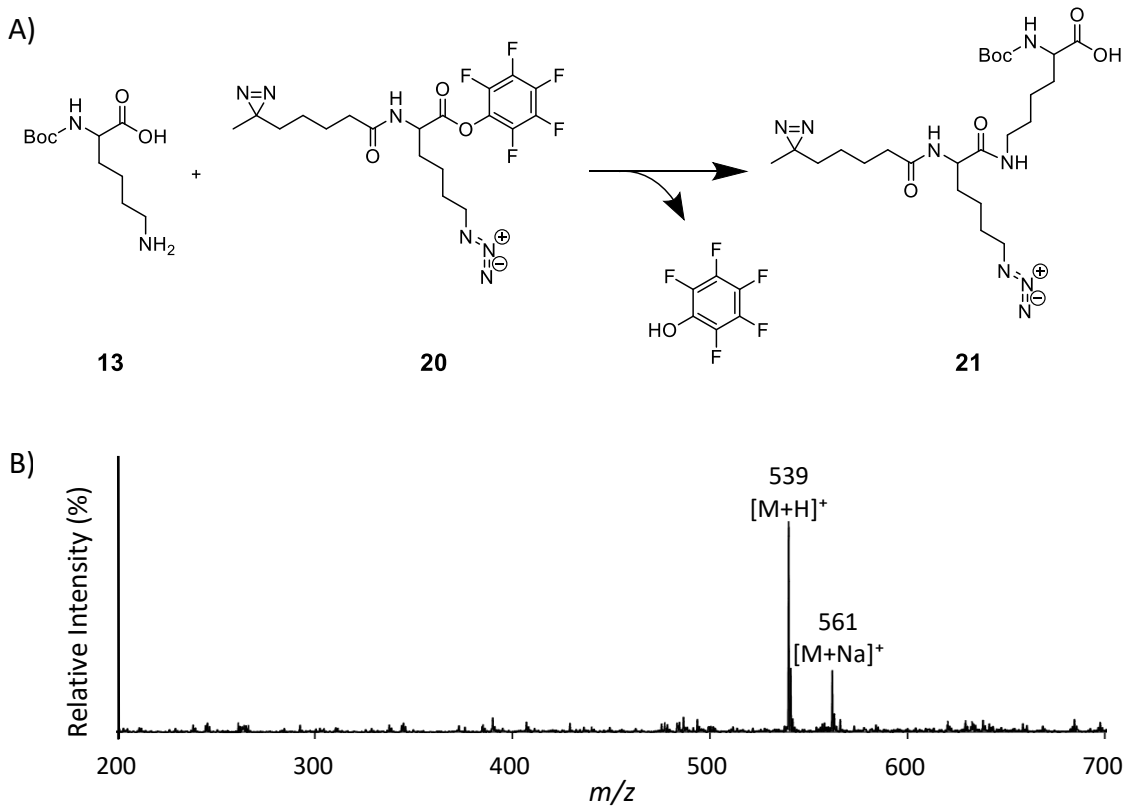


Figure 2.7: Mass spectrum of *N*-Boc lysine modified specifically with linker (20).
A) The reaction carried out between lysine and linker to form the modified lysine residue. This reaction was carried out in darkness to prevent the diazirine group from reacting. **B**) The mass spectrum of the reaction products reveals proton and sodium adducts of the modified lysine ions exclusively. Analysis was performed in 50 % ACN at collision energy 10 V.

Activation of the Diazirine Group

To optimise the conditions required for diazirine activation in linker **20**, various ratios of linker **20** and *N*$^{\alpha}$-Boc-L-lysine were exposed to UV light under a range of conditions including varying wavelength, distance from light source and exposure time (Table 2.2). Unfortunately, no linker-linker or linker-lysine product ions were observed upon MS analysis. In attempt to confirm the diazirine group had not been disrupted during synthesis, the diazirine precursor **17** was reacted with different combinations of *N*$^{\alpha}$-Boc-L-lysine **13** and exposed to a similar range of UV light conditions. Unfortunately, only starting material was observed, suggesting that although the preliminary MS and NMR analysis indicated successful formation of the diazirine group, it is not as reactive as desired. All conditions trialled are summarised in Table 2.2.

Table 2.2: Assays performed to determine reactivity of diazirine group.
Lysine:linker ratio refers to the molar ratio of cross-linking reagent in DMSO to-Boc-L-lysine **13** in PBS in the sample. UV wavelengths of 365 nm were generated using an 8 W UV lamp and wavelengths of 470 nm were generated using a photo-illuminator. Distance refers to the distance between the sample and the light source.

LINKER	LYSINE:LINKER MOLAR RATIO	UV WAVELENGTH, TIME, DISTANCE
17	1:0	365 nm, 15 min, 1 cm
17	1:1	365 nm, 15 min, 1 cm
N/A	0:1	365 nm, 15 min, 1 cm
17	5:1	365 nm, 15 min, 1 cm
17	1:0	365 nm, 15 min, 1 cm
17	1:1	365 nm, 15 min, 1 cm
N/A	0:1	365 nm, 15 min, 1 cm
17	5:1	365 nm, 15 min, 1 cm
20	5:1	365 nm, 30 min, 10 cm
20	5:1	365 nm, 30 min, 5 cm
20	5:1	365 nm, 1 h, 5 cm
20	10:1	365 nm, 40 min, 1 cm
20	10:1	365 nm, 2 h, 1 cm
N/A	0:1	365 nm, 2 h, 1 cm
20	10:1	470 nm, 30 min, 1 cm
20	10:1	470 nm, 2 h, 1 cm
N/A	0:1	470 nm, 2 h, 1 cm

2.5.2 Protein Cross-Linking

Before eliminating linker **20** from further analysis, it was applied to a protein system to enable comparison of linkage results with a commercial crosslinking reagent. These experiments were carried out on the standard protein lysozyme from the egg white of chickens, *Gallus gallus*, which is a widely used model protein due to its commercial availability, affordability and extensive characterisation.[102–104] The monomer is comprised of a single amino acid chain of 129 residues with a molecular mass of 14.3 kDa.[102] It contains 6 lysine resides (Lys1, Lys13, Lys33, Lys96, Lys97 and Lys116) (Figure 2.7), and is known to self-associate into homodimers and higher oligomeric states in aqueous solution at a neutral pH.[103–106] Hence, it is an ideal candidate for the development of a XL assay to probe the XL functionality of the designed lysine-specific linkers.

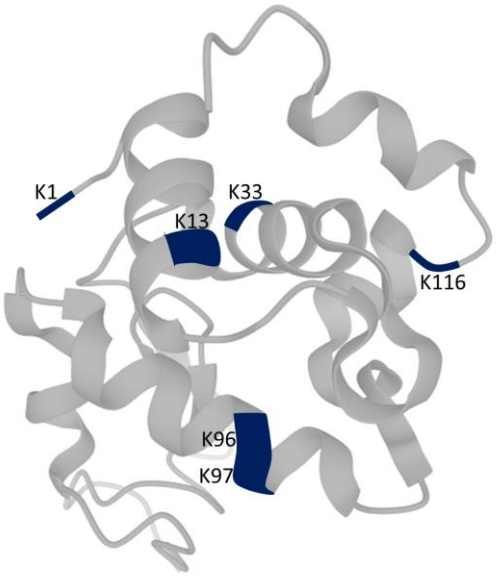


Figure 2.8: X-ray crystal structure of *Gallus gallus* lysozyme (PDB ID: P00698).[107]
The six lysine residues (K1, K13, K33, K96, K97 and K116) within lysozyme targeted for novel cross-linker validation are highlighted in blue.

Following the cross-linking reaction, the cross-linked product can be separated using SDS-PAGE. When carried out under reducing and denaturing conditions, any non-cross-linked dimer will dissociate. This allows isolation of higher-order structures only retained under

cross-linking conditions. These oligomeric species can then be selectively digested for proteomic analysis as described in Chapter 1.[59,85]

As the cross-linking reaction is not typically efficient, it was critical to identify a concentration of linker which would sustain formation of higher-order protein structures. This was achieved first using the standard cross-linker DSSO, which is a well-established reagent used in protein cross-linking.[72] The DSSO cross-linking reaction was carried out on lysozyme with varying concentrations of linker as a function of time, and SDS-PAGE was performed to determine the optimal conditions required for cross-linking. The cross-linked samples were compared to a negative control where no linker was added. As depicted in Figure 2.9 A-D, the dimer band is of reasonable intensity at all linker concentrations. Hence, it was determined that using an excess of 100x is not required for cross-linking. Comparatively, there is no major variation in the cross-linking products with an increase in reaction duration (Figure 2.7). Furthermore, increased incubation time does not appear to affect the stability of the cross-linked protein assembly.

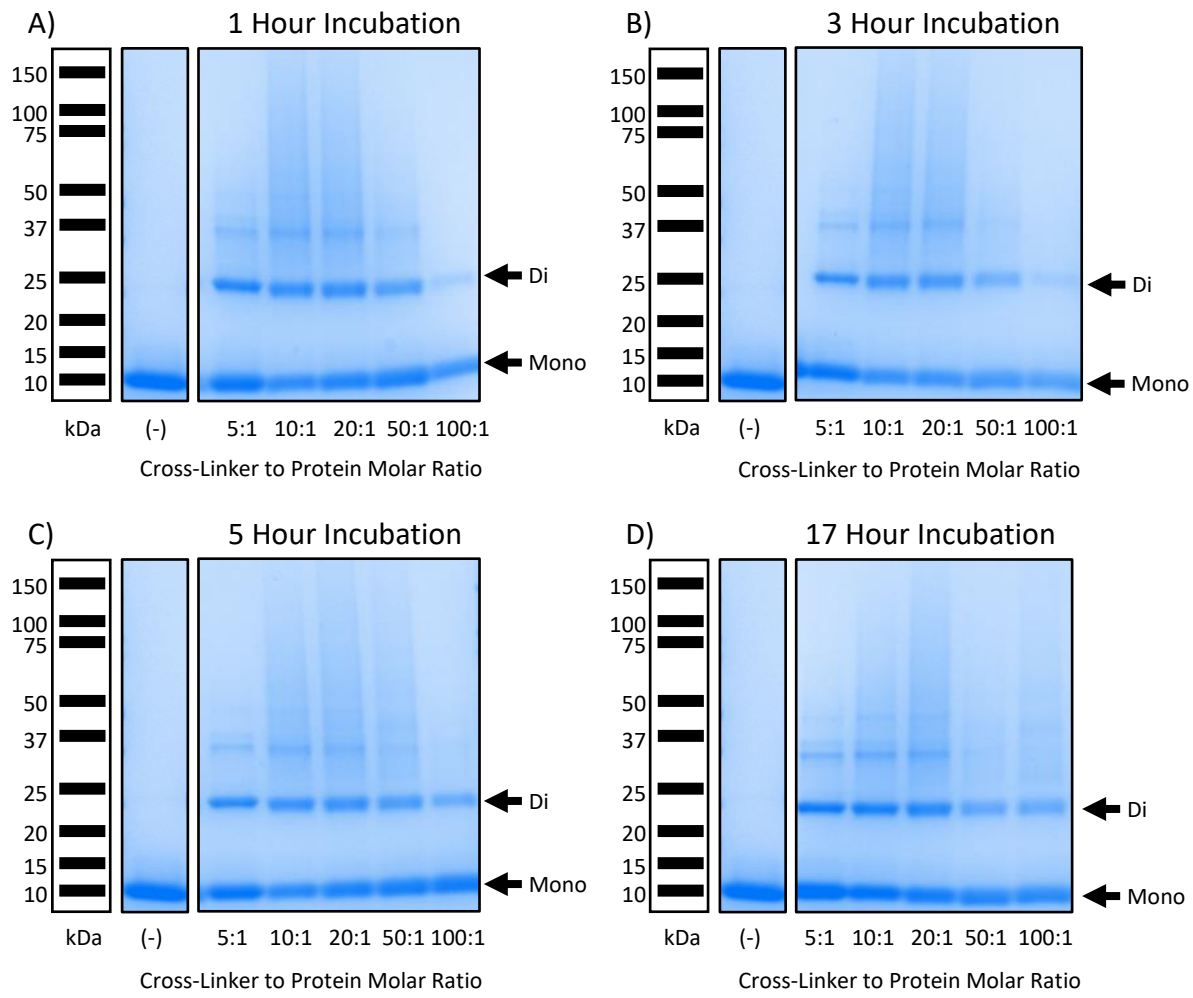


Figure 2.9: Optimisation of cross-linking reaction conditions of DSSO with lysozyme.
DSSO was applied in molar excess (5x, 10x, 20x, 50x and 100x) to lysozyme (1 mg/mL in PBS) over 1 h (**A**), 3 h (**B**), 5 h (**C**) and 17 h (**D**) to confirm the formation of cross-linked lysozyme by SDS-PAGE under reducing conditions. Bands corresponding to lysozyme dimers are labelled 'Di' and bands corresponding to lysozyme monomers are labelled 'Mono'.

These results can be used as a guide for the novel cross-linkers **9bii**, **10** and **20** with less reactive fluorinated phenyl reactive groups. Given that this group is unlikely to react as readily as the NHS group in DSSO, 20 equivalents of linkers **9bii**, **10** and **20** were used in subsequent cross-linking experiments with lysozyme. Similarly, a time between 3-5 h was chosen as the optimal cross-linking reaction for future cross-linking experiments presented in this chapter.

Cross-linking reactions were carried out on lysozyme with novel linkers **9bii**, **10** and **20** and DSSO. Following the lysine specific cross-linking conditions described above, samples containing linker **20** were additionally irradiated at 365 nm for 45 minutes to activate the non-

specific diazirine group. Subsequent SDS-PAGE analysis (Figure 2.10) demonstrates that the DSSO linked dimer was more abundant than dimers formed upon treatment with fluorophenyl linkers. The lack of dimer upon incubation with linker **20** confirms that the diazirine group is not reactive under these conditions, resulting in inefficient cross-linking. Out of the linkers synthesised in this project, linker **10** appeared to be most effective based on the intensity of the dimer band in the SDS-PAGE results.

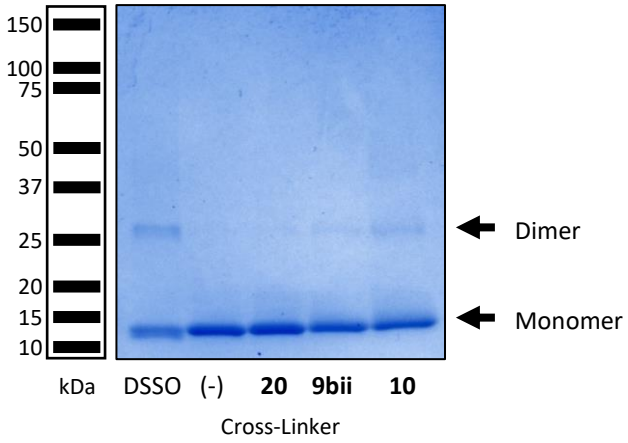


Figure 2.10: SDS-PAGE of cross-linked lysozyme.
Commercial lysine cross-linker (DSSO) was compared with the novel heterobifunctional linker **20** and the homobifunctional linkers **9bii** and **10,** demonstrating successful lysozyme cross-linking.

2.6 Affinity Tag Activation

To assess the feasibility of using CuAAC derivatisation post-linkage, a cyanine fluorescent azide group (Cy3) was linked onto the alkyne affinity tag of the cross-linked lysozyme species (Figure 2.11 A). Cy3 can mimic the steric bulk of the solid support used in peptide enrichment. Cy3 fluoresces at 532 nm, meaning cross-linked proteins successfully modified by Cy3 will be visualised under UV light, demonstrating the success of the CuAAC reaction on cross-linked proteins. The mechanistic basis for this reaction is discussed in section 2.1, Scheme 2.1. Following Cy3 linkage, the cross-linked samples were again analysed using SDS-PAGE. The gel was stained with Coomassie stain to visualise all protein bands (Figure 2.11 B) and then illuminated at 532 nm to visualise any Cy3 modified protein bands (Figure 2.11 C). SDS-PAGE shows dimer formation in the cross-linked samples in the absence of Cy3 under Coomassie

conditions (Figure 2.11 B), and in the presence of Cy3 under fluorescent conditions (Figure 2.11 C). The monomers under all conditions are clearly the most predominant species, followed by the dimers linked by **9bii** and the dimer linked by **10**. The broad monomeric band in the Cy3 lanes under fluorescent conditions reveal the large number of dead-end linkages occurring on the monomeric lysozyme species. Furthermore, the mass shift in the monomeric bands observed as the Cy3 group is attached under Coomassie conditions shows that there are many Cy3 modified linkers in these samples. These results demonstrate the feasibility of carrying out CuAAC chemistry on cross-linked samples, and further suggests that this reaction could be carried out using a click-resin column for peptide enrichment following protein digestion.

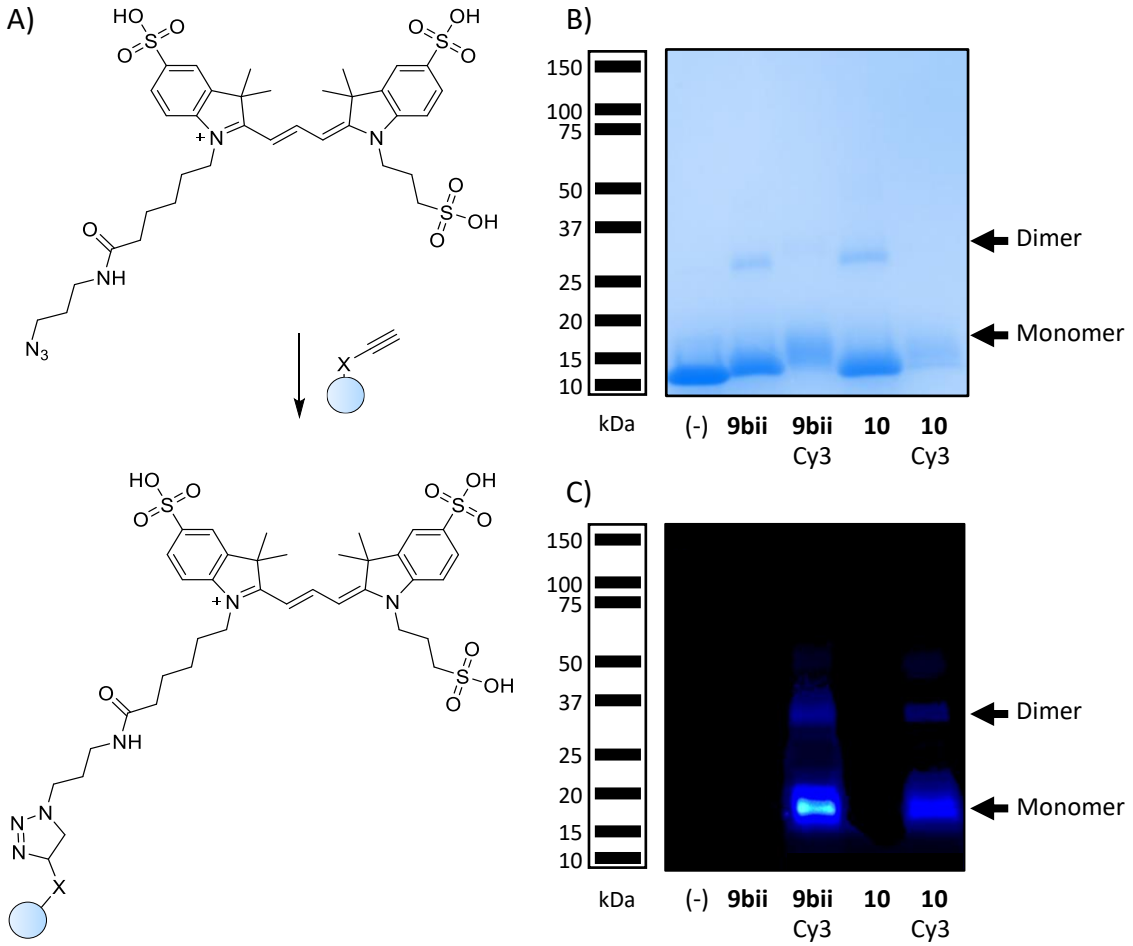


Figure 2.11: SDS-PAGE of cross-linked lysozyme and CuAAC linked Cy3.
Lysozyme samples (1 mg/mL) were cross-linked with linkers **9bii** and **10** using a 20x molar excess. Cy3 was conjugated onto the affinity handle of cross-linked lysozyme samples via CuAAC. Both monomeric and dimeric bands are indicated by arrows. **A**) The reaction between azide-modified Cy3 and a

cross-linked protein with an alkyne affinity tag. Blue circle represents the cross-linked protein moiety. **B**) SDS-PAGE of unmodified (-) and cross-linked lysozyme (with linkers **9bii** and **10**) in the presence and absence of Cy3 following Coomassie staining. **C**) Illumination of SDS-PAGE gel B) at 532 nm, specific to Cy3.

2.7 Localisation of Cross-Linking Sites

To further validate application of the novel cross-linking reagents synthesised here, the cross-linked dimer bands were subsequently digested using an in-gel digest protocol before being subjected to proteomic analysis. To simplify initial analysis, the proteomic data was acquired as MS^2 data, so linker cleavage data was not obtained at a separate level to proteomic sequencing. Hence, samples linked with the cleavable DSSO and linker **10** were treated as non-cleavable linkers. Subsequently, Thermo Proteome Discoverer 2.4.1.15 (ThermoFisher) was used to predict and identify cross-linked peptides using XlinkX nodes.[108] This was first performed using lysozyme samples cross-linked with DSSO and linker **10**, exhibiting 50 % and 62 % sequence coverage respectively. Both the DSSO linked sample and that linked with **10** identified a single analogous cross-link between lysine residues K13 and K116, with slightly varied peptide cleavage between samples (Table 2.3). This suggests that linker **10** functions as efficiently as commercially available DSSO. Furthermore, the large distance between K13 and K116 in three-dimensional space based on the x-ray crystal structure of lysozyme (PDB ID: P00698) suggests that this cross-link must be intermolecular.

Table 2.3: Cross-links identified in lysozyme using DSSO and novel linker 10.

LINKER	XL PEPTIDE FRAGMENTS	LINKED RESIDUES	XLINKX SCORE	MH^+ (Da)	*m/z*	OBSERVED IONS	RT (min)
DSSO	C**K**GTDVQAWIR-CELAAAM**K**R	K13 K116	112.06	2557.18	639.80	+4	16.66
10	NRC**K**GTDVQAWIR-VFGRCELAAAM**K**R	K13 K116	56.67	3112.61	778.66	+4	39.63

2.8 Chapter Conclusions and Future Work

Through this work, we have developed a generic workflow to synthesise a library of protein cross-linking reagents with readily interchangeable functionalities. These linkers have been applied to a protein system for identification of cross-linked residues and demonstrate promising potential in the design of other linker derivatives with further functionalities.

Overall, this work demonstrates the potential for using a modular cross-linker structure for the design and synthesis of a linker library incorporating varying functionalities. Similar chemistry can be used to include different moieties within linker structures to design unique derivatives. The synthesis of both a cleavable and non-cleavable homobifunctional linker and a non-cleavable heterobifunctional linker was successful. Purification of the homobifunctional linkers, however, requires improvement. Additionally, using fluorinated aromatic groups did not notably simplify the synthesis of the linker compared to NHS esters. The synthesis of the heterobifunctional linker was much simpler, and the product was easily isolated using silica chromatography. Given the success in the synthesis of this non-cleavable derivative, future work should be directed towards the synthesis of a cleavable derivative.

Cross-linking reactions carried out on lysozyme identified dimeric bands by SDS-PAGE which are only detected under cross-linking conditions. The limited number of lysine residues available for lysine-specific linkers to target meant that only one cross-link was identified using both DSSO and novel linker **10**. Given that cross-linking lysozyme with both DSSO and novel linker **10** identified the same cross-link by proteomic analysis, it can be determined that novel linker **10** is as effective as commercially available DSSO. The same could not be said for linker **20** however, as there was difficulty activating the diazirine group. It is possible that activation under a more intense light source or under different wavelengths may induce reactivity. Future work may investigate the application of these linkers to other protein systems, as well as further optimisation of the cross-linking reaction to identify other cross-linked residues. Furthermore, digested cross-linked protein samples may be enriched using CuAAC through a solid support to increase cross-link detection.

2.9 Materials and Methods

2.9.1 Materials

All reagents, solvents and starting materials were purchased from AK Scientific (California, USA), Chem Supply (South Australia, Australia), Fluorochem (Derbyshire, UK), Merck (New Jersey, USA), RCI Labscan (Bangkok, Thailand), Scharlau (Barcelona, Spain), Sigma Aldrich (New South Wales, Australia), or ThermoFisher Scientific (Massachusetts, USA) unless otherwise specified. DSSO was purchased from Sapphire Bioscience (NSW, Australia). Precast gels (4-15 % Mini-Protean® TGX™ polyacrylamide precast gel) and protein markers (Precision Plus Dual Colour protein standards) were purchased from Bio-Rad (California, USA).

2.9.2 General Synthetic Methods

Reactions carried out under an inert nitrogen atmosphere were performed with oven dried glassware and sealed with a septum injected with a syringe attached to an inflated nitrogen balloon. Reactions over ice were placed in an ice-bath and left to warm to room temperature over the duration specified.

All reactions were monitored by TLC using either 10 % methanol in DCM for polar stable compounds, and 40 % ethyl acetate in hexane for compounds susceptible to nucleophilic attack. TLC was performed using Merck aluminium sheets with silica gel 60 F_{254}. Spots were either visualised under a Vilber Lourmat UV lamp at 254 nm or using potassium permanganate stain (potassium permanganate (3 g), potassium carbonate (20 g); sodium hydroxide (5 %, 5 mL) and water (300 mL)) or iodine vapour. All Rf values were recorded to the nearest 0.01 using the solvent front carefully marked as soon as the TLC plate was removed from the jar.

Organic extracts were dried over anhydrous sodium sulphate or magnesium sulphate and then filtered through glass wool by gravity filtration.

Removal of solvent '*in vacuo*' and 'under reduced pressure' refers to removal of a large quantity of solvent through rotary evaporation. This was followed by drying under a high vacuum pump for a minimum of 1 h.

All yields are reported as isolated yields.

Silica chromatography was performed under positive nitrogen pressure using glass wool to prevent silica from flowing through the column. Silica deactivated with TEA was prepared by treating silica (100 g) with TEA (3 drops) in the solvent (100 mL), drying by vacuum filtration and then washing three times with solvent. Phosphate buffered silica was prepared by treating silica (100 g) with aqueous disodium hydrogen phosphate (0.2 M, 1L), and then adjusting the pH to 7 with phosphoric acid. The silica was then filtered and dried overnight.

2.9.3 NMR Analysis

NMR spectra were recorded using an Agilent DD2 spectrometer operating at 500 MHz for ^{1}H NMR, 125 MHz for ^{13}C NMR and 470 MHz for ^{19}F NMR (Agilent Technologies, Santa Clara, CA, USA). The main solvents used to obtain spectra consisted of deuterated chloroform, acetonitrile and methanol, depending on sample solubility. Spectra were analysed using MestReNova version 14.0.0 from Mestrelab Research S.L. Chemical shifts are reported on the δ-scale relative to TMS (δ_H 0.00) in ppm and are rounded to the nearest 0.01 ppm. Spin multiplicities are reported as (br s) broad singlet, (s) singlet, (d) doublet, (t) triplet, (q) quartet, and (m) multiplet. For all doublets, triplets, quartets and combinations, *J* values are reported to the nearest 0.1 Hz. The chemical shifts of multiplets are reported as a range.

2.9.4 MS Analysis of Cross-linking Reagents

High resolution mass spectra were recorded using an Agilent Technologies 6230 Accurate-Mass TOF LC/MS system equipped with an Infinity 1260 LC System (Agilent Technologies, California, USA). Samples (5 µL) in ACN (50 %) were directly injected at a flow rate of 0.3 mL/min in 0.1 % formic acid in 99.9 % ACN without chromatographic separation. The instrument conditions were set as follows: *m/z* range from 100-1600 *m/z*; polarity positive; capillary voltage 3.5 kV; nozzle voltage 2 kV; gas temperature 200 °C; gas flow 8 L/min. Spectra was acquired using Agilent MassHunter Workstation Data Acquisition (vB.08.00, Agilent Tehcnologies) and was processed using Mass Hunter Qualitative Analysis software (B.06.00). Masses are reported to 0.0001 *m/z*.

Offline MS/MS spectra were obtained using a Waters Micromass Q-TOF 2 mass spectrometer (Massachusetts, USA). Samples were ionised through nano-spray ESI using platinum coated capillary needles made in house. Capillary voltage ranged from 1.5-1.8 kV and cone voltage ranged from 20-30 V. Collision energy was increased incrementally from 10 to 50 V as

necessary to observe fragment ions. Spectra was acquired and processed using MassLynx (v4.1).

2.9.5 General Biological Methods

UV light sources for diazirine activation were an Ultra-Violet Products Inc. Chromato-VUE® transilluminator (model 6-62, 470 nm, 220 volts, 50 Hz, 0.75 AMPs) or a Spectroline long wave UV lamp (model EA-180/FE, 365 nm, 230 volt, 50 Hz, 0.17 AMPs).

SDS-PAGE

All protein samples (10 µL) were combined with 3x reducing gel loading buffer (10 µL) and denatured at 96 °C for 15 min. Samples were separated on a 4-15 % Mini-Protean® TGX™ polyacrylamide precast gel (Bio-Rad) with Precision Plus Dual Colour protein standards (Bio-Rad) as molecular weight markers. Electrophoresis was performed at 140 V, 400 mA in 1x SDS TGS running buffer for 1 h. SDS-PAGE gels were stained at room temperature with Coomassie Brilliant Blue stain prior to de-staining with Coomassie destain. Gels were imaged on the Bio-Rad ChemiDoc MP imaging system (Bio-Rad) using the Coomassie or Cy3 method as required.

Coomassie Brilliant Blue staining solution was made up of 0.2 % (w/v) Coomassie Brilliant Blue R250, 40 % ethanol (v/v) and 10 % (v/v) glacial acetic acid. Coomassie de-stain consisted of 40 % (v/v) methanol, 10 % (v/v) acetic acid.

SDS-PAGE loading buffer (3x, reducing) consisted of 150 mM Tris-HCl, 300 mM DTT, 6 % SDS, 30 % glycerol and 0.3 % (w/v) bromophenol blue, pH 6.8. 1x SDS-tris-glycine running buffer was diluted from 10x running buffer (25 mM tris, 192 mM glycine, 0.1 % SDS, pH 8.5).

In-Gel Digestion

Protein bands were excised and washed with NH_4HCO_3 (500 µL, 50 mM), and the Coomassie stain was removed through three 15 min sonicating incubations with NH_4HCO_3 (400 µL, 50 mM in 30 % ACN). Following solvent removal, gel bands were incubated with ACN (200 µL) for 15 min before removal *in vacuo* and resuspension in DTT (50 µL, 10 mM in 100 mM NH_4HCO_3) at 56 °C for 45 min. Following solvent removal, gel pieces were incubated with ACN (200 µL) for 15 min, solvent was removed *in vacuo* and resuspended in IAA (50 µL, 55 mM in 100 mM NH_4HCO_3) and incubated in the dark for 30 min. Following solvent removal, samples were washed with NH_4HCO_3, (100 µL, 5 mM) before being resuspended in ACN (200 µL) for

15 min and incubated with trypsin (10 µL, 10 ng/µL in 5 mM NH_4HCO_3) for 15 min. Additional NH_4HCO_3 (10 µL, 5 mM in 20 % ACN) was added before the sample was incubated at 37 °C overnight. Peptides were extracted from gel pieces by subsequent 15min incubations and extractions with formic acid (20 µL, 1 % in water), (50 µL, 1 % in 50 % ACN) and ACN (100 µL) before solvent was removed *in vacuo*. Samples were resuspended in ACN (5 µL, 3 % in 0.1 % FA) and analysed accordingly.

Digested samples were analysed using an Ultimate 3000 nano-flow system (ThermoFisher Scientific) coupled to an LTQ XL Orbitrap ETD mass spectrometer (ThermoFisher Scientific), controlled by Xcalibur (Version 2.1, ThermoFisher Scientific) in a data-dependent mode. Using a C_{18} trapping column (Acclaim PepMap 100 C_{18} 75 µm x 20 mm, ThermoFisher Scientific), approximately 2 µg of each peptide sample was pre-concentrated over 10 min, at a flow rate of 5 µL/min using 0.1 % trifluoroacetic acid in 2 % ACN. Peptides were then separated using a 75 µm ID C_{18} column (Acclaim PepMap100 C_{18} 75 µm x 50 cm, ThermoFisher Scientific) at a flow rate of 0.3 µL/min, where a linear gradient of 5 % to 45 % solvent B (0.1 % formic acid in 80 % ACN) in solvent A (0.1 % formic acid in 2 % ACN) was applied over 60 min. Subsequently, the line was washed with 90 % solvent B over 5 min before the system was equilibrated with 5 % solvent B over 15 min. Mass spectra were acquired over the mass range of 300-2000 *m/z* in positive mode at a resolution of 60,000 in FT mode. CID fragmentation was performed on the 10 most intense precursor ions using a dynamic exclusion of 5 s where the dynamic exclusion criteria included a minimum relative signal intensity of 1000 and $\geqslant$2 positive charge state. The isolation width used was 3.0 *m/z* and a normalised collision energy of 35 was applied.

Cross-Linking MS Data Analysis

Cross-linking data was analysed using XlinkX on Thermo Proteome Discoverer 2.4.1.15. The Heck Lab non-cleavable analysis template was used in the analysis of both the cleavable and non-cleavable data.[108] In the consensus workflow (Figure 2.11 A), the 'MSF File' and 'Peptide Validator' nodes required updating, but no changes were made to the settings of other nodes. In the processing workflow (Figure 2.11 B), the 'Spectrum Files' and 'Spectrum Selector' nodes were used as loaded. In the 'XlinkX Detect' node, the cross-linker was entered, 'enable protein N terminus' was selected as 'true', and the 'non-cleavable acquisition' strategy was chosen. The 'XlinkX Filter' nodes remained unchanged, but in the 'Sequest HT' node, the mass

tolerance was increased to 20 ppm for the Orbitrap instrument used to obtain this data. Additionally, dynamic modifications including oxidation, hydrolysed linker, cleaved linker were added. The 'Perculator' node was replaced with 'Target Decoy PSM Validator' but no other settings were changed. In the 'XlinkX Search' node, the static modification 'Carbamidomethyl' was added to identify any IAA capping, the 'Perculator Mass Tolerance' was changed to 20, and the 'FTMS' to 30.

2.9.6 General Synthetic Procedures

General Procedure 2A: TFP esterification with DCC

The respective acid was dissolved in a 1:1 DCM/ACN (200 mL/1 g of acid) over ice. A solution of DCC (1.5 eq) and TFP (2 eq) in 1:1 DCM/ACN (5 mL) was added to the reaction which was warmed to room temperature overnight under a nitrogen atmosphere. The solvent was removed *in vacuo*. The crude residue was purified by silica chromatography to yield the desired ester.

General procedure 2B: TFP esterification with EDC

The relevant acid was dissolved in anhydrous THF (100 mL/1 g of acid). Subsequently, EDC (3 eq) was added and the reaction was stirred at room temperature for 10 mins. A solution of TFP (6 eq) in THF (2 mL) was added and the reaction was stirred at room temperature under an atmosphere of nitrogen overnight. The solvent was removed *in vacuo* and the resulting residue was partitioned between DCM and water. The aqueous layer was extracted with DCM (3x) and the combined organic extracts were dried over magnesium sulphate and filtered. The solvent was removed *in vacuo* before the crude product was purified by silica chromatography.

General procedure 2C: One-pot *N*-Boc deprotection and methyl esterification

The *N*-boc protected amino acid of interest was dissolved in methanol (40 mL/1 g of acid) over ice. Thionyl chloride (2 eq) was added dropwise (1 drop/sec) over ice and the reaction was stirred under an atmosphere of nitrogen at room temperature overnight. The solvent was removed *in vacuo* and the resulting oil was resuspended in methanol and concentrated *in vacuo* three times to remove residual thionyl chloride.

General procedure 2D: Amide coupling

A solution of the relevant amine in DMF (5 mL/ 1 g of amine) was added dropwise (1 drop/sec) to a stirring solution of HATU (1.2 eq), DIPEA (4.5 eq) and the relevant acid (1 eq) in DMF (5 mL/1 g of acid). The reaction was stirred at room temperature under a nitrogen atmosphere overnight before the mixture was diluted with ethyl acetate (50 mL) and washed with water (50 mL). The aqueous layer was re-extracted with ethyl acetate (50 mL) and the combined organic extracts were washed with 1M HCl (50 mL), 1M NaOH (2x 50 mL) and saturated NaCl solution (50 mL). The combined organic extracts were dried over magnesium sulphate and filtered before the solvent was removed under reduced pressure and the crude residue was purified via silica chromatography.

General procedure 2E: Hydrolysis of C-terminal methyl ester

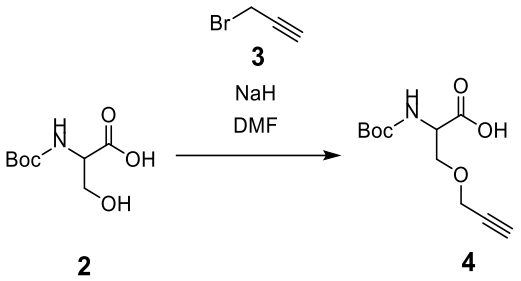
The appropriate diester was diluted in THF (35 mL/1 g of diester). A 1 M solution of lithium hydroxide (35 mL/1 g of diester) was added dropwise (1 drop/sec) over ice and the reaction was stirred for 3 h. The mixture was concentrated under reduced pressure before being diluted with water (20 mL) and acidified with 1 M HCl. The solution was extracted with ethyl acetate (3x 30 mL) and the combined organic extracts were washed with saturated sodium chloride solution (30 mL). The organic extracts were dried over magnesium sulphate, filtered and the solvent was removed *in vacuo.*

General procedure 2F: PFP carbonate esterification

The respective acid was dissolved in anhydrous ACN before adding DIPEA (4 eq per reactive acid group) dropwise (1 drop/sec). A solution of PFP carbonate (2.2 eq) in ACN (5 mL) was added and the reaction was stirred at room temperature for 0.5-5 h under a nitrogen atmosphere. The solvent was removed *in vacuo* and the residue was purified by silica chromatography.

2.9.7 Synthesis of Homobifunctional Linkers

2-[(tert-butoxycarbonyl)amino]-3-(prop-2-yn-1-yloxy)propanoic acid **(4)**



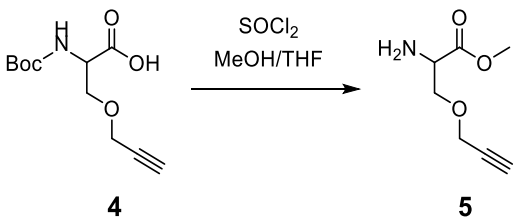
N-Boc-L-Serine **2** (4.34 g, 21 mmol) was dissolved in anhydrous DMF (150 mL) over ice. A 60 % wt/wt dispersion of sodium hydride (2.5 eq) was gradually added to the solution which was stirred over ice for 45 min. An 80 % v/v solution of propargyl bromide **3** in toluene (1 eq) was added dropwise (1 drop/sec) to the reaction which was left to warm to room temperature overnight. The reaction was quenched with water (30 mL) and stirred for another 20 min before being washed with diethyl ether (2x). The combined aqueous layers were acidified with 15 % potassium hydrogen sulphate solution and then extracted with ethyl acetate (3x). The combined organic extracts were washed with water (3x) and saturated NaCl solution before being dried over magnesium sulphate. Once filtered, the solution was concentrated *in vacuo* to give **4** as a dark yellow oil without further purification (4.55 g, 89 %). R*f* (10 % methanol in DCM): 0.33.

^{1}H NMR (500 MHz, $CDCl_3$): δ 10.67 (1H, br s, O**H**), 5.41 (1H, d, *J* = 8.6 Hz, N**H**), 4.52-4.43 (1H, m, **H**αC), 4.15 (2H, d, *J* = 2.3 Hz, OC**H$_2$**CCH), 3.97 (1H, dd, *J* = 9.4, 3.2 Hz, HαCC**H**HO), 3.79 (1H, dd, *J* = 9.4, 3.4 Hz, HαCCH**H**O), 2.45 (1H, t, *J* = 2.3 Hz, OCH_2CC**H**), 1.43 (9H, s, N-Boc) ppm.

HRMS (ESI) *m/z*: $[M+H]^+$ calc'd for $C_{11}H_{17}NO_5$, 244.1181; found 244.1181.

NMR data was consistent with literature.[89,96]

Methyl 2-amino-3-(prop-2-yn-1-yloxy)propanoate **(5)**



Modified serine **4** (5.92 g, 24 mmol) underwent *N*-Boc deprotection and esterification according to General Procedure 2C to yield **5** as a brown oil without further purification (4.16 g, 67 %).

R*f* (40 % ethyl acetate in hexane): 0.48.

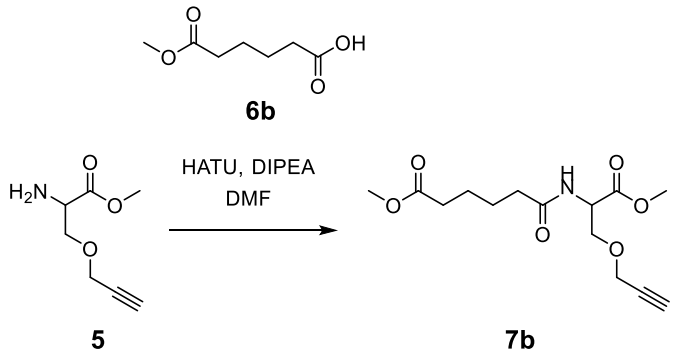
^{1}H NMR (500 MHz, $CDCl_3$): δ 8.70 (2H, Br s, N**H2**), 4.55-4.42 (1H, m, **H**αC), 4.28 (2H, ABqd, J_{AB} = 16.3, 2.3 Hz, OC**H2**CCH), 4.15 (2H, d, *J* = 3.3 Hz, HαCC**H2**O), 3.48 (3H, s, OC**H3**), 2.50 (1H, t, *J* = 2.3 Hz, OCH2CC**H**) ppm.

^{13}C NMR: (125 MHz, chloroform): δ 170.5, 81.3, 78.4, 69.2, 61.4, 56.2, 56.1 ppm.

HRMS (ESI) *m/z*: $[M+H]^+$ calc'd for $C_7H_{11}NO_3$, 158.0812; found 158.0811.

NMR data was consistent with literature.[1,2]

Methyl 5-{[1-methoxy-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl]carbamoyl} pentanoate (**7b**)[89]



Modified serine **5** (1.77 g, 11.26 mmol) was amide coupled to monomethyl adipate **6b** as described in General Procedure 2D. The crude residue was purified via silica chromatography (30-70 % ethyl acetate in petroleum ether) to obtain **7b** as a yellow oil (0.85 g, 26 %).

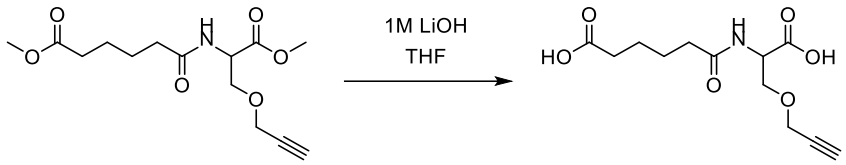
R*f* (50 % ethyl acetate in petroleum ether): 0.22.

^{1}H NMR (500 MHz, $CDCl_3$): δ 6.29 (1H, d, *J* = 8.2 Hz, N**H**), 4.77 (1H, dt, *J* = 8.2, 3.1 Hz, **H**αC), 4.02-3.99 (2H, m, OC**H$_2$**CCH), 3.82 (1H, dd, *J* = 9.4, 3.1 Hz, HαCC**H**HO), 3.65-3.60 (4H, m, HαCCH**H**O and OC**H$_3$**), 3.52 (3H, s, OC**H$_3$**), 2.39 (1H, t, *J* = 2.4 Hz, OCH_2CC**H**), 2.23-2.13 (4H, m, COC**H$_2$**CH_2CH_2C**H$_2$**CO), 1.58-1.49 (4H, m, $COCH_2$C**H$_2$**C**H$_2$**CH_2CO) ppm.

HRMS (ESI) *m/z*: $[M+H]^+$ calc'd for $C_{14}H_{21}NO_6$, 300.1442; found 300.1444.

NMR data was consistent with literature.[89]

5-{[1-carboxy-2-(prop-2-yn-1-yloxy)ethyl]carbamoyl}pentanoic acid **(8b)**[89]



7b **8b**

Diester **7b** was deprotected according to General Procedure 2E to give **8b** as a pale-yellow oil (0.57 g, 84 %).

R*f* (10 % methanol in dichloromethane): 0.06.

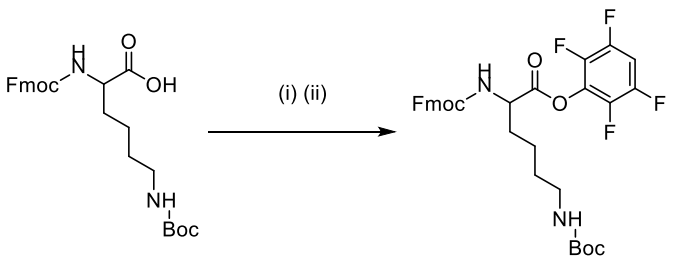
^{1}H NMR (500 MHz, CD_3CN): δ 6.91 (1H, d, *J* = 7.7 Hz, N**H**), 4.60 (1H, ddd, *J* = 7.7, 3.6, 4.4 Hz, **H**αC), 4.18 (2H, m, OC**H$_2$**CCH), 3.91 (1H, dd, *J* = 9.7, 4.4 Hz, HαCC**H**HO), 3.72 (1H, dd, *J* = 9.7, 3.7 Hz, HαCCH**H**O), 2.76 (1H, t, *J* = 2.4 Hz, OCH_2CC**H**), 2.37-2.33 (4H, m, COC**H$_2$**CH_2CH_2C**H$_2$**CO), 1.61 (4H, m, $COCH_2$C**H$_2$**C**H$_2$**CH_2CO) ppm.

^{13}C NMR: (125 mHz, CD_3CN): δ 177.5, 176.5, 173.6, 81.9, 78.0, 71.8, 60.8, 55.1, 37.7, 35.7, 27.5, 26.7 ppm.

HRMS (ESI) *m/z*: $[M+H]^+$ calc'd for $C_{12}H_{17}NO_6$, 272.1129; found 272.1128.

NMR data was consistent with literature.[89]

2,3,5,6-tetrafluorophenyl 6-[(tert-butoxycarbonyl)amino]-2-{[(9H-fluoren-9-ylmethoxy)carbonyl]amino}hexanoate (**12i**)



11 **12i**

(i) Via acyl chloride intermediate

A solution of N^{α}-Fmoc-N^{ε}-Boc-L-lysine **11** (0.12 g, 0.26 mmol) was dissolved in anhydrous DCM (4mL) over ice. Thionyl chloride (27 eq) was added and the reaction was stirred under an atmosphere of nitrogen for 1 h. A solution of TFP (3.8 eq) and DIPEA (3.4 eq) in anhydrous DCM (3 mL) was added dropwise (1 drop/sec) and the reaction was warmed to room temperature overnight under an atmosphere of nitrogen. The solvent was removed *in vacuo* to afford a brown oil. The crude residue was purified by silica chromatography (40 % ethyl acetate in hexane) but did not yield the desired TFP ester **12i**.

(ii) Using DCC

N^{α}-Fmoc-N^{ε}-Boc-L-lysine **11** (0.11 g, 0.23 mmol) was esterified with TFP using DCC according to General Procedure 2A. The residue was purified by silica chromatography (20-40 % ethyl acetate in hexane) to give the TFP ester **12i** as a green waxy solid (0.03 g, 20 %).

R*f*: (40 % ethyl acetate in hexane): 0.54.

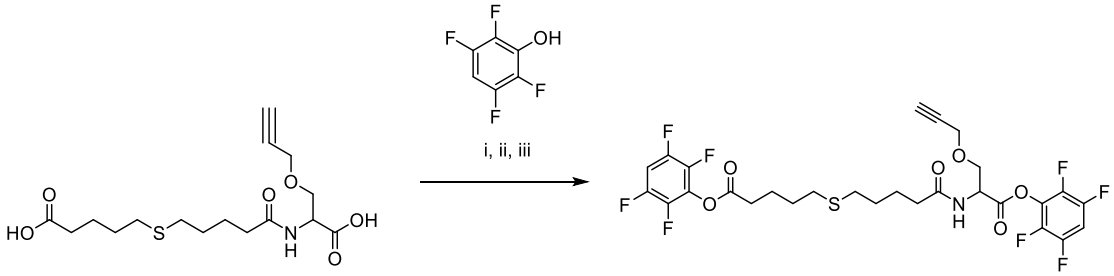
^{1}H NMR (500 MHz, $CDCl_3$): δ 7.76 (2H, d, *J* = 7.6 Hz, Fmoc-**arom**), 7.61 (2H, d, *J* = 7.6 Hz Fmoc-**arom**), 7.40 (2H, t, *J* = 7.5 Hz, Fmoc-**arom**), 7.32 (2H, t, *J* = 7.5 Hz, Fmoc-**arom**), 7.06-6.97 (1H, m, TFP-**H**), 5.53 (1H, d, *J* = 8.0 Hz, N**H**), 4.72 (1H, q, *J* = 7.5 Hz, **H**αC), 4.59 (1H, s br, N**H**), 4.51-4.38 (2H, m, Fmoc-C**H_2**), 4.24 (1H, t *J* = 7.0 Hz, Fmoc-C**H**), 3.16 (2H, d *J* = 7.2 Hz, NHHαCCHHCH_2CH_2C**H_2**NHBoc), 2.11-2.00 (1H, m, NHHαCC**H**HCH_2CH_2CH_2NHBoc), 1.99-1.86 (1H, m, NHHαCCH**H**CH_2CH_2CH_2NHBoc), 1.63-1.38 (13H, m, NHHαCCHHC**H_2**C**H_2**CH_2NH**Boc**) ppm.

^{13}C NMR: (125 MHz, $CDCl_3$): δ171.4, 158.6, 149.6, 147.6, 146.4, 146.2, 144.0, 142.01, 130.4, 129.7, 127.6, 122.6, 69.9, 56.4, 49.8, 42.4, 34.3, 32.3, 31.1, 24.8 ppm.

HRMS (ESI) *m/z*: $[M+H]^+$ calc'd for $C_{32}H_{32}F_4N_2O_6$, 617.2270; found 617.2263.

NMR data was consistent with literature.[93,94,109]

2,3,5,6-tetrafluorophenyl 5-[(4-{[1-oxo-3-(prop-2-yn-1-yloxy)-1-(2,3,5,6-tetrafluorophenoxy)propan-2-yl]carbamoyl}butyl)sulphanyl]pentanoate **(9ai)**



8a **9ai**

(i) Using DCC

8a (0.10 g, 0.29 mmol) was esterified with TFP (6 eq) using DCC (5 eq) according to General Procedure 2A. The residue was purified by silica chromatography (10-40 % ethyl acetate in hexane) but the desired TFP ester was not recovered. The fractions were recombined, the solvent was removed *in vacuo* and the crude oil was repurified by silica chromatography (20 % ethyl acetate in hexane) but the desired product **9ai** was still not recovered.

(ii) Using T3P

8a (0.09 g, 0.24 mmol) was dissolved in anhydrous ethyl acetate (15 mL) over ice under a nitrogen atmosphere. TEA (12.6 eq) was added, followed by T3P in a 50 % weight solution in ethyl acetate 5.6 eq). A solution of TFP (2.7 eq) was added to the reaction which warmed to room temperature overnight under a nitrogen atmosphere. The resulting residue was concentrated *in vacuo* and purified by silica chromatography (20 % ethyl acetate in hexane) but did not yield the desired ester **9ai**.

(iii) Using TFP and EDC[99,110]

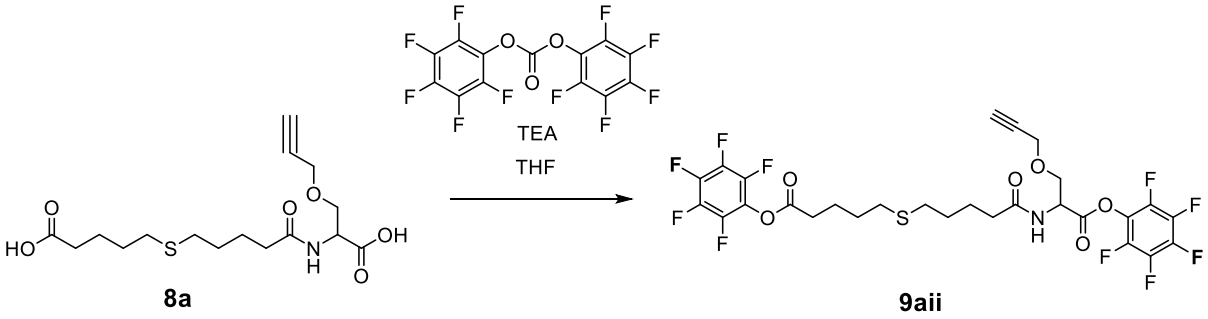
8a (0.10 g, 0.28 mmol) was esterified with TFP using EDC (3.4 eq) according to General Procedure 2B. The residue was purified via silica chromatography (10 % ethyl acetate in hexane) to obtain **9ai** as a pale yellow oil (0.01 g, 4 %).

^{1}H NMR (500 MHz, CD_3CN): δ 7.36-7.28 (2H, m, TFP-**H**), 7.06 (1H, m, N**H**), 4.99 (1H, dt, *J* = 7.7, 4.0 Hz, **H**αC), 4.26 (2H, d, *J* = 2.3 Hz, OC**H$_2$**CCH), 4.10 (1H, dd, *J* = 9.7, 4.2 Hz, HαCC**H**HO), 3.88 (1H, dd, *J* = 9.8, 3.6 Hz, HαCCH**H**O), 2.81-2.73 (5H, m, OCH_2CC**H** and $CO(CH_2)_4S(CH_2)_4CO$), 2.60-2.52 (4H, m, $CO(CH_2)_4S(CH_2)_4CO$), 1.92-1.52 (8H, m, $CO(CH_2)_4S(CH_2)_4NH$) ppm.

^{13}C NMR (125 MHz, $CDCl_3$): δ 180.7, 175.6, 172.1, 150.0, 148.1, 141.8, 139.8, 139.3, 81.6, 78.0, 71.3, 61.0, 55.2, 38.1, 35.0, 31.4, 31.0, 27.1, 26.2 ppm.

HRMS (ESI) *m/z*: $[M+H]^+$ calc'd for $C_{28}H_{25}F_8NO_6S$, 656.1343; found 656.1348.

2,3,4,5,6-pentafluorophenyl 5-[(4-{[1-oxo-1-(2,3,4,5,6 pentafluorophenoxy)-3-(prop-2-yn-1-yloxy)propan-2-yl]carbamoyl}butyl)sulphanyl]pentanoate **(9aii)**



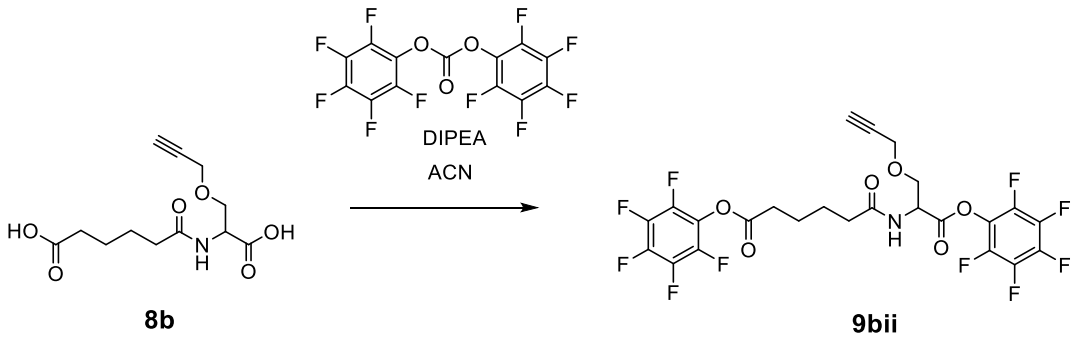
Diacid **8a** (0.10 g, 0.28 mmol) was dissolved in anhydrous THF (10 mL). An excess of TEA (10.3 eq) was added and the reaction was concentrated *in vacuo*. The yellow oil was redissolved in anhydrous THF (10 mL) over ice before the PFP-carbonate was added (0.45 g, 1.77 mmol, 6.3 eq). The reaction was stirred at room temperature overnight under an atmosphere of nitrogen before the solvent was removed *in vacuo* to afford the crude **9aii** as an orange oil. The crude residue was purified via silica chromatography (10 % ethyl acetate in hexane) to obtain **9aii** as a pale-yellow oil.

^{1}H NMR (500 MHz, CD_3CN): δ 7.16 (1H, d, *J* = 7.5 Hz, N**H**), 4.99 (1H, dt, *J* = 7.7, 3.9 Hz, **H**αC), 4.27-4.24 (2H, m, OC**H$_2$**CCH), 4.11 (1H, dd, *J* = 9.8, 4.2 Hz, HαCC**H**HO), 3.88 (1H, dd, *J* = 9.7, 3.6 Hz, HαCCH**H**O), 2.84-2.76 (1H, m, OCH$_2$CC**H**), 2.58-2.50 (4H, m, CO(C**H$_2$**)$_4$S(C**H$_2$**)$_4$CO), 2.26-2.17 (4H, m, CO(C**H$_2$**)$_4$S(C**H$_2$**)$_4$CO), 1.76-1.56 (8H, m, CO(C**H$_2$**)$_4$S(C**H$_2$**)$_4$NH) ppm.

^{13}C NMR (125 MHz, CD_3CN): δ 180.7, 175.9, 172.3, 142.4, 141.8, 140.5, 139.7, 137.7, 135.9, 81.5, 71.2, 61.0, 55.2, 37.3, 35.0, 33.6, 33.4, 31.4, 31.0, 30.0, 26.2, 24.6 ppm.

HRMS (ESI) *m/z*: $[M+H]^+$ calc'd for $C_{28}H_{23}F_{10}NO_6S$, 692.1159; found 692.1162.

2,3,4,5,6-pentafluorophenyl 5-{[1-oxo-1-(2,3,4,5,6-pentafluorophenoxy)-3-(prop-2-yn-1-yloxy)propan-2-yl]carbamoyl}pentanoate **(9bii)**[96,97]



8b **9bii**

(i) PFP carbonate and DIPEA

Esterification of diacid **8b** (0.35 g, 1.11 mmol) with PFP carbonate (3.3 eq) using DIPEA (6.6 eq) was attempted using General Procedure 2F. The reaction was left for 4 h before the crude residue was purified by silica chromatography (10 % ethyl acetate in hexane), but did not yield the desired **9bii**.

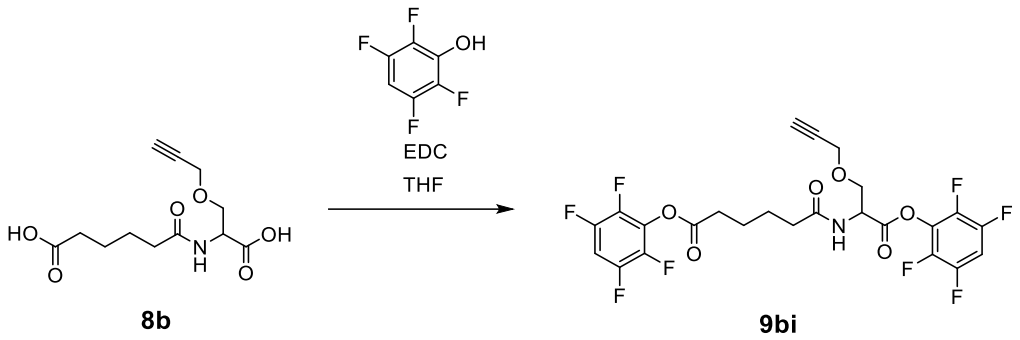
(ii) Alteration of equivalents

Esterification of diacid **8b** (0.07 g, 0.26 mmol) with PFP carbonate was attempted using General Procedure 2F. The reaction was left for 5 h before the crude residue was purified by silica chromatography (0-10 % ethyl acetate in hexane) which did not yield the desired **9bii**.

(iii) Without isolation

Esterification of diacid **8b** (0.25 g, 0.94 mmol) with PFP carbonate was attempted using General Procedure 2F. The reaction was left for 30 min before workup. The reaction was stored at -20°C for 3 weeks before it was resuspended in THF (20 mL). Additional PFP carbonate (0.5 eq) and TEA (0.5 eq) was added and the reaction was stirred overnight. The solvent was removed under reduced pressure but did not yield the desired **9bii**.

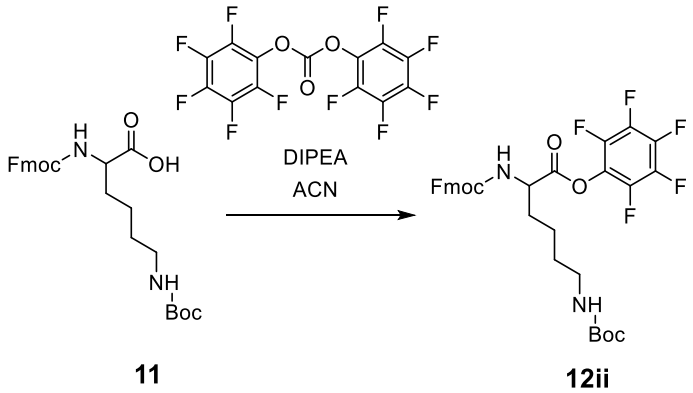
2,3,5,6-tetrafluorophenyl 5-{[1-oxo-3-(prop-2-yn-1-yloxy)-1-(2,3,5,6-tetrafluorophenoxy)propan-2-yl]carbamoyl}pentanoate **(9bi)**[99,110]



A solution of **8b** in ethyl acetate was washed with 1M HCl (3x) and concentrated *in vacuo* to afford **8b** (0.01 g) which was resuspended in anhydrous THF (3 mL). Subsequently, the diacid **8b** was esterified with TFP using EDC as described in General Procedure 2B to afford trace quantities of crude **9bi** as a pale-yellow oil.

HRMS (ESI) *m/z*: $[M+H]^+$ calc'd for $C_{24}H_{17}F_8NO_6$, 568.1034; found 568.1035.

2,3,4,5,6-pentafluorophenyl 6-[(tert-butoxycarbonyl)amino]-2-{[(9H-fluoren-9-ylmethoxy)carbonyl]amino}hexanoate **(11ii)**



N$^{\alpha}$-Fmoc-*N*$^{\varepsilon}$-Boc-L-lysine **10** (0.22 g, 0.47 mmol) was esterified using PFP carbonate as described in General Procedure 2F. The reaction was stirred for 30 min before work up to afford crude ester **11ii** as a brown oil.

R*f* (40 % ethyl acetate in hexane): 0.6.

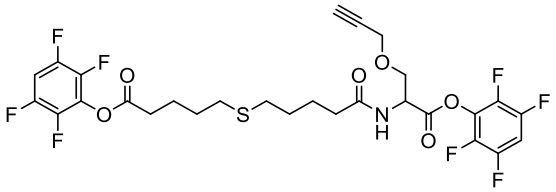
^{1}H NMR (500 MHz, $CDCl_3$): δ 7.76 (2H, d, *J* = 7.6 Hz, Fmoc-**arom**), 7.60 (2H, d, *J* = 7.5 Hz Fmoc-**arom**), 7.39 (2H, t, *J* = 7.6 Hz, Fmoc-**arom**), 7.30 (2H, t, *J* = 7.5 Hz, Fmoc-**arom**), 5.57 (1H, d, *J*

= 7.7 Hz, N**H**), 4.69 (1H, q, *J* = 7.0 Hz, **H**αC), 4.61 (1H, br, N**H**), 4.51-4.39 (2H, m, Fmoc-C**H$_2$**), 4.24 (1H, t *J* = 7.0 Hz, Fmoc-C**H**), 3.16 (2H, d *J* = 7.0 Hz, NHHαCCHHCH$_2$CH$_2$C**H$_2$**NHBoc), 2.10-1.99 (1H, m, NHHαCC**H**HCH$_2$CH$_2$CH$_2$NHBoc), 1.98-1.86 (1H, m, NHHαCCH**H**CH$_2$CH$_2$CH$_2$NHBoc), 1.64-1.38 (13H, m, NHHαCCHHC**H$_2$**C**H$_2$**CH$_2$NH**Boc**) ppm.

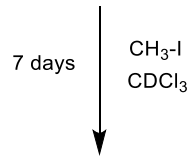
HRMS (ESI) *m/z*: $[M+Na]^+$ calc'd for $C_{32}H_{31}F_5N_2O_6$, 657.2012; found 657.2027.

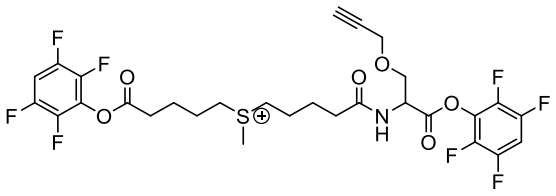
NMR data was consistent with literature.[109]

Methyl(4-{[1-oxo-3-(prop-2-yn-1-yloxy)-1-(2,3,5,6-tetrafluorophenoxy)propan-2-yl]carbamoyl}butyl)[5-oxo-5-(2,3,5,6-tetrafluorophenoxy)pentyl]sulphanium **(12)**



9ai





10

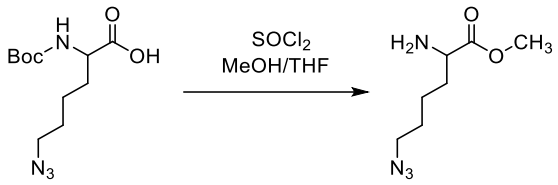
To a solution of **9ai** (0.008 g, 0.012 mmol) in chloroform (4 mL), methyl iodide (6 eq) was added. The reaction was stirred at room temperature for 7 days. The solvent was removed *in vacuo* to give **12** as a pale yellow waxy solid (0.006 g, 74 %).

R*f* (40 % ethyl acetate in hexane): 0.00.

HRMS (ESI) *m/z*: $[M]^+$ calc'd for $C_{29}H_{28}F_8NO_6S$, 670.1490; found 670.1501.

2.9.8 Synthesis of Heterobifunctional Linker

Methyl 2-amino-6-azidohexanoate **(15)**



14 **15**

Modified lysine **14** (2.09 g, 7.67 mmol) was *N*-Boc cleaved and esterified according to General Procedure 2C to yield **15** as a pale yellow oil (1.64 g, 84 %).

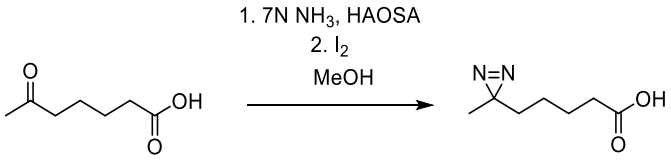
R*f*: (10 % methanol in DCM): 0.00.

^{1}H NMR (500 MHz, CD_3OH): δ 4.07 (1H, t, *J* = 6.3 Hz, **H**αC), 3.85 (3H, s, COOC**H$_3$**), 3.35 (2H m, HαC(C**H$_2$**)$_4$), 1.94 (2H, m, HαC(C**H$_2$**)$_4$), 1.65 (2H, m, HαC(C**H$_2$**)$_4$), 1.57 (1H, m, HαC(C**H**$_2$)$_4$N$_3$), 1.33 (1H, m, HαC(C**H**$_2$)$_4$) ppm.

^{13}C NMR: (125 MHz, $CDCl_3$): δ 165.4, 55.8, 53.2, 34.1, 32.6, 30.8, 24.9 ppm.

HRMS (ESI) *m/z*: $[M+H]^+$ calc'd for $C_7H_{14}N_4O_2$, 186.1117; found 188.1220.

5-(3-methyldiazirin-3-yl)pentanoic acid **(17)**



14 **15**

A solution of 7N ammonia in methanol (45 mL) and 5-acetovaleric acid **16**(4.35 g, 37.2 mmol) was stirred on ice under an atmosphere of nitrogen for 3 h. Hydroxyl-O-sulphonic acid (0.9 eq) in methanol (30 mL) was added dropwise (1 drop/sec). The reaction was warmed to room temperature overnight under an atmosphere of nitrogen. Nitrogen gas was bubbled through the resulting suspension for 1.5 h to remove excess ammonia before the pale-yellow solution was isolated via vacuum filtration and concentrated *in vacuo* to give a viscous yellow oil. The residue was redissolved in methanol (30 mL) and stirred over ice for 15 min, before adding TEA (1.5 eq) and stirring for another 10 min. Iodine chips (1.3 eq) were added gradually until a permanent brown colour was apparent. The reaction was diluted in ethyl acetate (30 mL)

and washed with 1M HCl (30 mL), 10 % sodium thiosulphate (3 x 30 mL) and saturated NaCl solution (30 mL). The organic layer was dried over anhydrous magnesium sulphate and filtered before the solvent was removed under reduced pressure to give **17** as a yellow-brown oil (4.04 g, 69 %).

R*f*: (10 % methanol in DCM): 0.22.

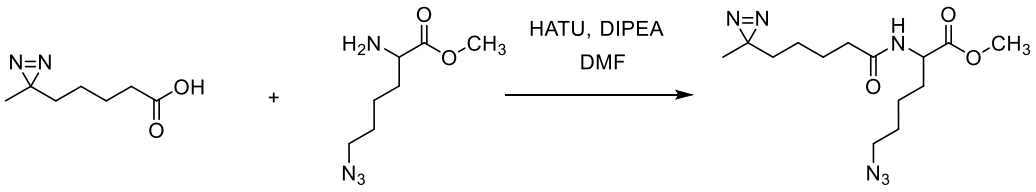
^{1}H NMR (500 MHz, $CDCl_3$): δ 10.63 (1H, br s, COO**H**), 2.35 (2H, t, J = 7.4 Hz, $CH_2C\mathbf{H_2}COOH$), 1.62 (2H, dt, J = 15.3, 7.5 Hz, $C\mathbf{H_2}CH_2COOH$), 1.42-1.36 (2H, m, $N{=}NCH_2C\mathbf{H_2}CH_2CH_2COOH$), 1.28-1.20 (2H, m, $N{=}NC\mathbf{H_2}CH_2$), 1.02 (3H, s, $\mathbf{H_3}CCN{=}N$) ppm.

^{13}C NMR (125 MHz, $CDCl_3$): δ 182.2, 45.8, 36.6, 36.4, 26.8, 26.1, 22.4 ppm.

HRMS (ESI) *m/z*: $[M\text{-}H]^-$ calc'd for $C_7H_{12}N_2O_2$, 155.0826; found 155.0827.

NMR data was consistent with literature.[57,111]

Methyl 6-azido-2-[5-(3-methyldiazirin-3-yl)pentanamido]hexanoate **(18)**



17 **15** **18**

Acid **17** (1.25 g, 8.02 mmol) was amide coupled to modified lysine **15** (1 eq) described in General Procedure 2D in darkness. The mixture was extracted using diethyl ester rather than ethyl acetate. The crude residue was purified via silica chromatography (2 % methanol in DCM) to obtain **18** as a dark yellow oil (0.31 g, 16 %).

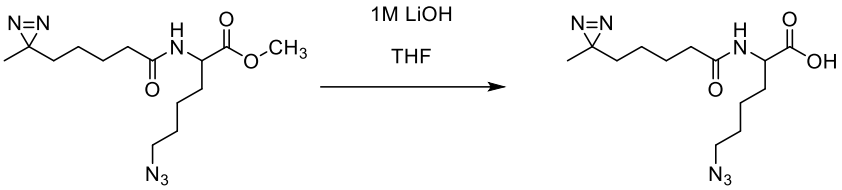
R*f*: (10 % methanol in DCM): 0.78.

^{1}H NMR (500 MHz, $CDCl_3$): δ 6.39 (1H, d, *J* = 8.0 Hz, **N**H), 4.46 (1H, dt, *J* = 8.0, 5.4 Hz, **H**αC), 3.62 (3H, s, $OC\mathbf{H_3}$), 3.16 (2H, dt, *J* = 6.7, 1.2 Hz, $H\alpha CCH_2CH_2CH_2C\mathbf{H_2}N_3$), 2.08 (2H, t, *J* = 7.5 Hz, $H_3CN{=}NCH_2CH_2CH_2C\mathbf{H_2}CO$), 1.78-1.69 (1H, m, $H\alpha CC\mathbf{H_2}CH_2CH_2CH_2N_3$), 1.65-1.51 (2H, m, $H\alpha CC\mathbf{H_2}CH_2CH_2CH_2N_3$), 1.48 (4H, m, $H\alpha CCH_2CH_2C\mathbf{H_2}CH_2N_3$ and $H_3CN{=}NCH_2CH_2C\mathbf{H_2}CH_2CO$), 1.36-1.26 (2H, m, $H\alpha CCH_2C\mathbf{H_2}CH_2CH_2N_3$), 1.26-1.21 (2H, m, $H_3CN{=}NCH_2C\mathbf{H_2}CH_2CH_2CO$), 1.09-1.02 (2H, m, $H_3CN{=}NC\mathbf{H_2}CH_2CH_2CH_2CO$), 0.87 (3H, s, $\mathbf{H_3}CCN{=}N$) ppm.

^{13}C NMR: (125 MHz, $CDCl_3$): δ 175.6, 174.9, 55.1, 54.4, 53.7, 38.8, 36.6, 34.7, 31.0, 28.3, 27.6, 26.3, 25.0, 22.5 ppm.

HRMS (ESI) *m/z*: $[M+H]^+$ calc'd for $C_{14}H_{24}N_6O_3$, 324.1910; found 325.1983.

6-azido-2-[5-(3-methyldiazirin-3-yl)pentanamido]hexanoic acid **(19)**



18 **19**

Diester **18** was deprotected according to General Procedure 2E using ethyl acetate as the extraction solvent to give **19** as a pale yellow solid (0.20 g, 93 %).

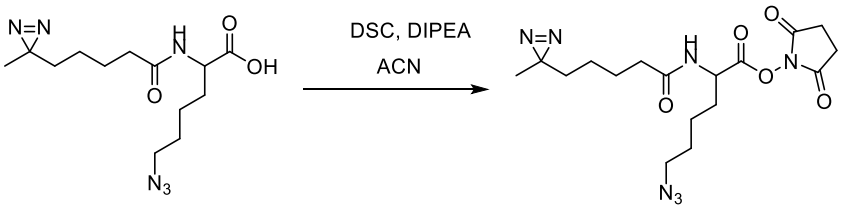
R*f*: (10 % methanol in DCM): 0.17.

^{1}H NMR (500 MHz, $CDCl_3$): δ 9.95 (1H, br s, O**H**), 6.49 (1H, d, *J* = 7.4 Hz, N**H**), 4.54 (1H, m, **H**αC), 3.24 (2H, t, *J* = 6.8 Hz, HαCCH$_2$CH$_2$CH$_2$C**H$_2$**N$_3$), 2.19 (2H, t, *J* = 7.5 Hz, H$_3$CN=NCH$_2$CH$_2$CH$_2$C**H$_2$**CO), 1.90-1.84 (1H, m, HαCC**H$_2$**CH$_2$CH$_2$CH$_2$N$_3$), 1.72-1.69 (1H, m, HαCC**H$_2$**CH$_2$CH$_2$CH$_2$N$_3$), 1.62-1.54 (4H, m, HαCCH$_2$CH$_2$C**H$_2$**CH$_2$N$_3$ and H$_3$CN=NCH$_2$CH$_2$C**H$_2$**CH$_2$CO), 1.43-1.38 (2H, m, HαCCH$_2$C**H$_2$**CH$_2$CH$_2$N$_3$), 1.33-1.30 (2H, m, H$_3$CN=NCH$_2$C**H$_2$**CH$_2$CH$_2$CO), 1.16-1.10 (2H, m, H$_3$CN=NC**H$_2$**CH$_2$CH$_2$CH$_2$CO), 0.94 (3H, s, **H$_3$**CCN=N) ppm.

^{13}C NMR: (125 MHz, $CDCl_3$): δ 177.8, 176.5, 54.7, 53.7, 38.7, 36.5, 34.2, 31.0, 28.4, 27.7, 26.1, 25.0, 22.4 ppm.

HRMS (ESI) *m/z*: $[M+H]^+$ calc'd for $C_{13}H_{22}N_6O_3$, 310.1753; found 311.1829.

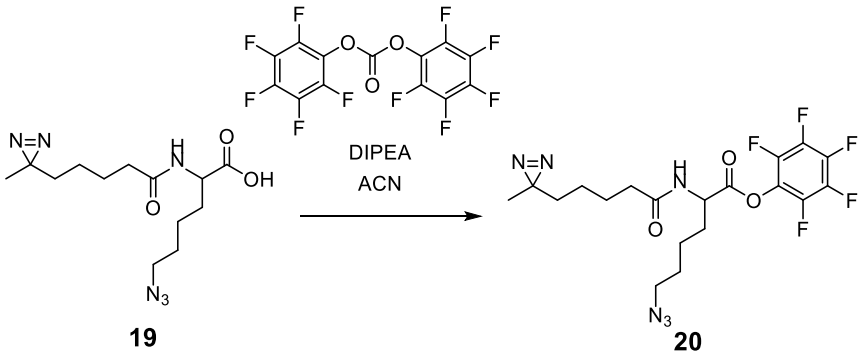
2,5-dioxopyrrolidin-1-yl 6-azido-2-[5-(3-methyldiazirin-3-yl)pentanamido] hexanoate **(21)**



19 **20**

To a solution of **19** (0.11 g, 0.36 mmol) in anhydrous ACN (20 mL), DIPEA (4 eq) and DSC (1.1 eq) were added and the reaction was stirred at room temperature overnight. The solvent was removed *in vacuo* and the crude yellow residue was purified by silica chromatography (40 % ethyl acetate in hexane) which did not yield the desired NHS ester **21**. Fractions were recombined and purified by silica chromatography (10 % ethyl acetate in petroleum benzene) but the desired NHS ester was not recovered.

2,3,4,5,6-pentafluorophenyl 6-azido-2-[5-(3-methyldiazirin-3-yl)pentanamido] hexanoate **(20)**



Acid **19** (0.23 g, 0.74 mmol) was esterified with PFP carbonate (1.5 eq) according to General Procedure 2F over 5 h. The crude residue was purified by silica chromatography (30 % ethyl acetate in hexane) to afford PFP-ester **20** as a dark brown oil (0.14 g, 41 %).

R*f*: (30 % ethyl acetate in hexane) 0.49.

^{1}H NMR (500 MHz, $CDCl_3$): δ 6.15 (1H, d, *J* = 7.8 Hz, N**H**), 4.97 (1H, td, *J* = 7.7, 5.4 Hz, **H**αC), 3.33 (2H, t, *J* = 6.6 Hz, HαCCH$_2$CH$_2$CH$_2$C**H$_2$**N$_3$), 2.28 (2H, t, *J* = 7.5 Hz, H$_3$CN=NCH$_2$CH$_2$CH$_2$C**H$_2$**CO), 2.12-2.04 (1H, m, HαCC**H**$_2$CH$_2$CH$_2$CH$_2$N$_3$), 1.97-1.86 (1H, m, HαCC**H**$_2$CH$_2$CH$_2$CH$_2$N$_3$), 1.79-1.60 (4H, m, HαCCH$_2$CH$_2$C**H$_2$**CH$_2$N$_3$ and H$_3$CN=NCH$_2$CH$_2$C**H$_2$**CH$_2$CO), 1.59-1.51 (2H, m, HαCCH$_2$C**H$_2$**CH$_2$CH$_2$N$_3$), 1.44-1.15 (4H, m, H$_3$CN=NC**H$_2$**C**H$_2$**CH$_2$CH$_2$CO), 0.98 (3H, s, **H$_3$**CCN=N) ppm.

^{13}C NMR: (125 MHz, $CDCl_3$): δ 176.4, 171.1, 144.7, 142.6, 141.7, 139.6, 136.5, 134.3, 54.6, 53.6, 38.6, 36.5, 34.3, 30.9, 28.3, 27.6, 26.1, 25.0, 22.3 ppm.

^{19}F NMR: (470 MHz, $CDCl_3$): δ -152.5, -156.7, -161.5 ppm.[75]

HRMS (ESI) *m/z*: $[M+H]^+$ calc'd for $C_{19}H_{21}F_5N_6O_3$, 476.1595; found 477.1667.

2.9.9 Cross-Linking Methods

Activation of Fluorinated Phenyl Groups

A solution of N^{α}-Boc-L-lysine (3 mg, 0.011 mmol) and linker **20** (4 eq) in DMSO (80 µL) was incubated at room temperature with shaking (130 rpm) for 30 min. The sample was diluted in 1:1 ACN and water and analysed by MS.

Activation of Diazirine Group

Samples of N^{α}-Boc-L-lysine **13** (1 mg/mL in PBS) with linker **20** (in DMSO, linker to lysine ratios of 1:0, 1:1, 0:1, 5:1, 10:1) or precursor **17** (in DMSO, linker to lysine ratios of 1:0, 1:1, 0:1, 5:1, 10:1) were exposed to UV light (365 nm or 470 nm, 1-10 cm from light source) for 15 min – 2 h. The sample was diluted in 1:1 ACN and water and analysed by MS.

Cross-linking of Lysozyme

Lysozyme samples (1 mg/mL in PBS) were cross-linked with DSSO and novel linkers **9bii**, **10**, **20** in a 1-100x molar excess (linker stocks in DMSO for final volume of 5 % DMSO).

Samples were incubated at room temperature with shaking (100 rpm) for 1-17 h, before samples containing linker **20** were illuminated at 365 nm, 1 cm from the light source for 0.5-2 h. Cross-linked samples were then identified by SDS-PAGE under reducing conditions. A non-cross-linked lysozyme sample was compared to the cross-linked sample as a negative control.

Affinity Tag Activation

To the cross-linked protein samples (50 µL, 1 mg/mL in PBS), Cy3 (20 µL, 2.5 mM in DMSO) and PBS (90 µL) were added. Sequentially, BTTAA (10 µL, 100 mM in PBS), $CuSO_4$ (10 µL, 20 mM in water) and sodium ascorbate (10 µL, 300 mM in milliQ water) were added before the mixture was incubated in darkness for 30 min. Excess reagents were removed by washing with PBS through Amicon Ultra-0.5 mL centrifugal filter units (Merck Millipore) with a 3 kDa molecular weight cut off at 14,000 *xg*. When flow-through became colourless, cross-linked-Cy3-linked samples were removed from spin filters and subjected to SDS-PAGE, as described in section 2.9.5. Gels were first imaged using a ChemiDoc MP imaging system (Bio-Rad) using the inbuilt transilluminating Cy3 method at 532 nm, before being stained with Coomassie and imaged using the Coomassie method.

Chapter 3: Mass Spectrometric Interrogation of the Higher-Order Protein Structures in *Bitis arietans* venom

3.1 Introduction

The complex proteinaceous mixture of snake venom is designed to immobilise and debilitate prey in a highly specific manner. Synergistic interactions arising from higher-order protein structures are believed to enhance the potency, yet an in-depth characterisation of these structures is lacking. Recent advances in MS have demonstrated promising results in characterising higher-order protein assemblies in other biological systems. Such advances may therefore contribute to the characterisation of venom proteins.

The *Bitis* genus, belonging to the venomous viper family *Viperidae*, is responsible for many of the snakebites reported in Sub-Saharan Africa.[2] Of particular significance is the African Puff Adder *Bitis arietans* (Figure 3.1) which is considered to have the most toxic venoms of the *Viperidae* family based on LD_{50} studies carried out on mice.[2,112] Given it's venom potency and wide distribution across Africa, the Puff Adder is the second most deadly snake in the continent. Envenomation can lead to bleeding, swelling, necrosis and death. Despite the severity and frequency of *B. arietans* snakebites, the mechanism behind envenomation remains poorly understood and the proteinaceous components from the venom are not well characterised. Previous work has touched on proteomic analysis to classify the individual proteins found within the venom, with other work investigating the biochemical activities of a select number of proteins including the PLA_2 Bitanarin and the CTL Bitisectin.[8,11,17,32,33,113,114] This work however is not overly detailed in terms of structural analysis, and very little work has been performed to investigate the role of quaternary structure in envenomation. Thus, the protein composition of *B. arietans* venom is an ideal model candidate for demonstrating the application of MS in studying higher-order venom protein structures.

Figure 3.1: The African Puff Adder, *Bitis arietans.*
Image attributions: "Puff adder *Bitis arietans*" by jonsson.nick is licenced under CC BY-SA 2.0.

This chapter aims to characterise the protein components within the venom of *B. arietans* using proteomic sequencing in conjunction with intact and native MS methods. Not only will this structural understanding inform mechanistic insights of envenomation, but it may assist in applying these isolated proteins as therapeutic agents and biotechnical tools. Furthermore, this chapter aims to demonstrate the use of a MS-based workflow for the characterisation of higher-order venom protein systems.

3.2 Method for Venom Proteome Analysis

The highly complex proteinaceous composition of snake venom has prevented a thorough characterisation of venom protein structures and identities. Furthermore, currently established methods for protein studies such as NMR and x-ray crystallography have difficulties in capturing the potentially dynamic interactions and heterogeneous nature of these protein systems in a high throughput manner. The MS-based workflow proposed in this section attempts to overcome such challenges, and the steps are explained below.

3.2.1 Venom Source and Purification

The crude whole *B. arietans* venom used in this project is extracted through a process called milking and is lyophilised for long-term storage. Crude whole venom contains a highly complex mixture of proteins and small molecules which can create difficulties in the analysis of the venom. When analysing highly heterogeneous samples by MS, less abundant protein

species are often undetected amongst the dominant signals of highly abundant species. Additionally, the ionisation energy applied to crude samples must be distributed across a broad range of species and may not be enough to ionise all species within the sample, especially those in lower abundance or with lower ionisation efficiency. Due to the complex proteinaceous nature of this venom, purification is mandatory for an in-depth characterisation using MS.[115]

Size exclusion chromatography (SEC) is an efficient and reproducible technique that can be used to separate protein components within a complex mixture based on movement through a porous gel. Larger components flow through the SEC column uninterrupted, whereas smaller components diffuse readily into the gel's pores, decreasing their net flow rate. Proteins eluting from the column are detected by UV absorbance at 280 nm. When a sample is eluted in a physiologically relevant buffer such as ammonium acetate the non-covalent protein assemblies are retained, allowing for downstream native MS analysis and investigation of higher-order structures.[116] It must be noted that although these fractions will contain proteins of a similar size, they will likely contain a mixture of isoforms and a range of globular proteins which may coelute together. SDS-PAGE can be used to simply profile the complexity of components in the SEC fractions by further separating proteins by molecular weight. This can be performed under denaturing and/or reducing conditions, providing an initial understanding of the higher-order interactions which may be stabilising larger protein assemblies.

3.2.2 Intact MS Experiments

Intact MS experiments can be performed under both denaturing and reducing conditions, allowing accurate mass measurement of protein subunits, and some insight into quaternary level characterisation of the protein units within a sample. Exposure to denaturing organic solvents such as ACN and acidic conditions can expose previously hidden hydrophobic residues and conceal previously exposed hydrophilic residues within the core of the protein. This causes disruption of the non-covalent intermolecular forces stabilising higher-order structures of protein assemblies, releasing non-covalently linked monomeric species. For example, an oligomeric structure of 100 kDa in native conditions may be revealed as a non-covalent stabilised dimer of two 50 kDa species under denaturing conditions. Reducing agents

such as DTT and tris(2-carboxyethyl)phosphine (TCEP) can be used to disrupt covalent disulphide bonds stabilising higher-order structures. Exposure to reducing and denaturing conditions together ensures disulphide interactions usually buried within solvent-inaccessible regions of the protein can also be reduced. Reducing agents are usually incompatible with MS analysis so must be used at low concentrations or removed before analysis. Comparison between denaturing and reducing experiments provides an insight into the presence of covalent interactions such as disulphide bonds used to stabilise quaternary structures. The structures revealed under denaturing and reducing conditions can then be compared to those identified in native MS experiments.

3.2.3 Native IMMS Analysis

Under native conditions, non-covalent interactions are maintained, allowing identification of quaternary interactions between monomeric protein units expected in physiological conditions. Mass changes between native and intact MS experiments can therefore indicate dissociation of protein assemblies under given conditions. Furthermore, IMMS experiments can be conducted on purified fractions to characterise the three-dimensional protein structure under physiological conditions. As described in Chapter 1.3.1, IMMS can separate proteins by the additional dimension of mobility, allowing identification of isomers with different conformations. Very few venom proteins have been characterised by this method thus far, but potential has been shown in the characterisation of phospholipase A_2s from the eastern brown snake *Pseudonaja textilis.*[117]

3.2.4 Proteomic Analysis

Proteomic analysis can be used to identify the protein species present within a sample. To reveal the protein profiles in the venom samples, the bottom-up proteomic strategy described in Chapter 1.3.1 can be implemented. This method allows primary level sequencing of each digested protein within a given sample and can be performed on both whole venom and purified venom fractions. This technique complements the MS-based higher-order structural methods discussed above by revealing the nature of different toxins found within each sample. Although proteins within the same toxin family possess some characteristic sequences, differences in PTMs and the overall protein sequence can cause substantial variation in the mass of individual proteins, particularly if from a different genus. Hence,

uncharacterised proteins can typically be classified into a protein family, but their exact identity cannot typically be revealed using bottom-up proteomics alone unless a genomic database is available for high fidelity sequence matching.

A variety of techniques are available to quantify the relative abundance of protein species in samples from proteomic analysis. Traditionally, many of these methods utilise isotopic labels but these reagents are expensive and require comparisons between standard samples which can complicate large studies, particularly if the proteins under analysis are unknown.[118] Advances in MS technologies have introduced a variety of label-free protein quantification methods, however these are only able to give a relative abundance as ion detection is dependent on ionisation efficiency. Studies have demonstrated a correlation between peak area intensity with protein abundance in complex samples.[119–121] Alternatively, another label-free method, termed spectral counting, compares the total number of MS/MS spectra assigned to a given protein, called a peptide spectrum match (PSM).[122,123] It must be noted that using PSM values only provides an approximation for relative abundance, as not only do all peptides have different ionisation efficiencies, but larger proteins will naturally have a larger number of tryptic peptides than smaller proteins. Despite these limitations, spectral counting has been shown to correlate reasonably well with known protein concentrations and many research groups now use spectral counting as a semi-quantitative approach to determine relative protein abundance.[122–129] In this thesis, PSMs are used to describe the relative abundance of proteins identified.

3.2.5 XLMS

XLMS can be applied to the venom samples to confirm the higher-order protein structures proposed by MS analysis. Cross-linking samples and then performing reducing and denaturing SDS-PAGE allows visual identification of higher-order protein assemblies normally stabilised by non-covalent and covalent disulphide interactions. Bottom-up proteomics can also be performed to identify the binding interfaces associated with higher-order oligomer formation. The workflow applied to *B. arietans* venom in this thesis is summarised in Figure 3.2 below and utilises the techniques discussed above. Following successful application to this venom system, it may be broadly applied to other complicated protein systems in biology.

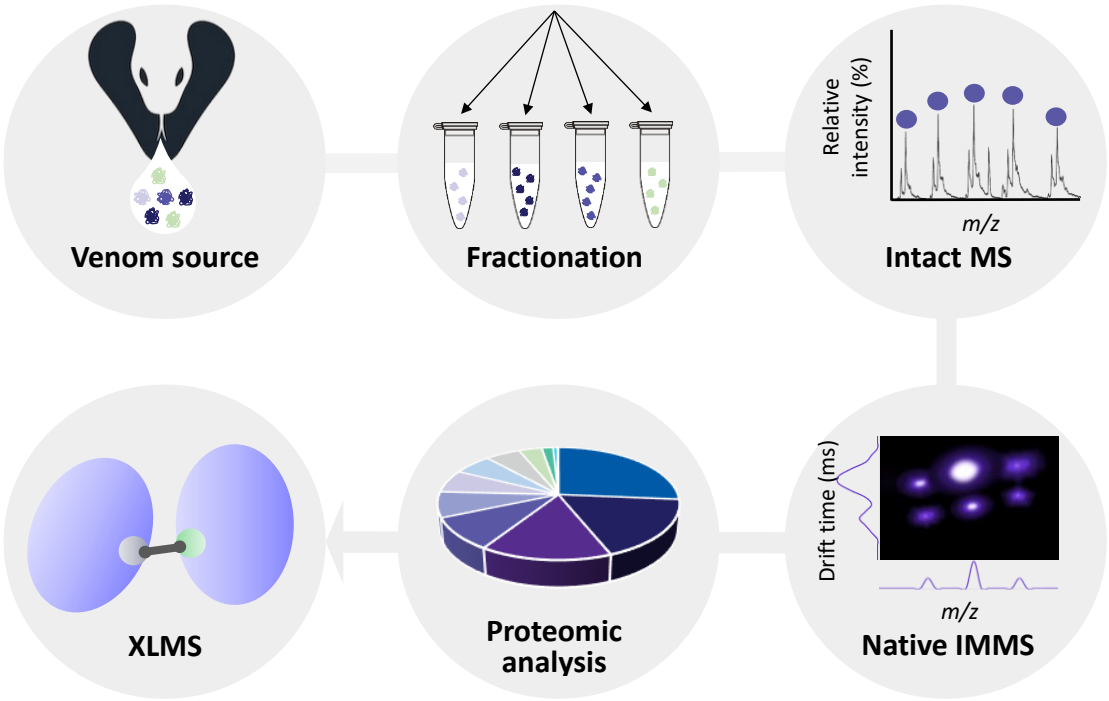


Figure 3.2: MS workflow for studying the higher-order structures of venom proteins.
Following isolation from the venom source, crude venom is fractionated prior to intact MS analysis under denaturing, reducing and native conditions. Proteomic analysis is used to identify the proteins present in the venom. IMMS is performed to understand the three-dimensional structure of the monomeric protein subunits. Finally, cross-linking MS is used to identify the nature of interactions within the higher-order assembly. Once the structure of these protein systems is characterised, functional studies can be carried out.

3.3 Results and Discussion

3.3.1 *B. arietans* venom purification by SEC and SDS-PAGE

Fractionation of the crude *B. arietans* venom was performed using SEC (Figure 3.3). Seven major peaks were observed representing a diverse suite of protein sizes. Peaks are labelled by fraction alphabetically along the elution profile, where Fractions C and G are the most abundant, followed by Fractions B, E and F and finally Fractions A and D. The samples were eluted in ammonium acetate for conservation of higher-order protein structures. The poor baseline resolution suggests there is overlap in the proteins present under each peak. It is most likely that any higher-order protein assemblies exist in Fractions A, B, C and D due to their early elution volume. Hence, purified fractions (A-E) were selected for further structural

characterisation. The late elution times for Fraction F and G suggest they only contain peptides and/or small molecules, and thus were excluded from further studies.

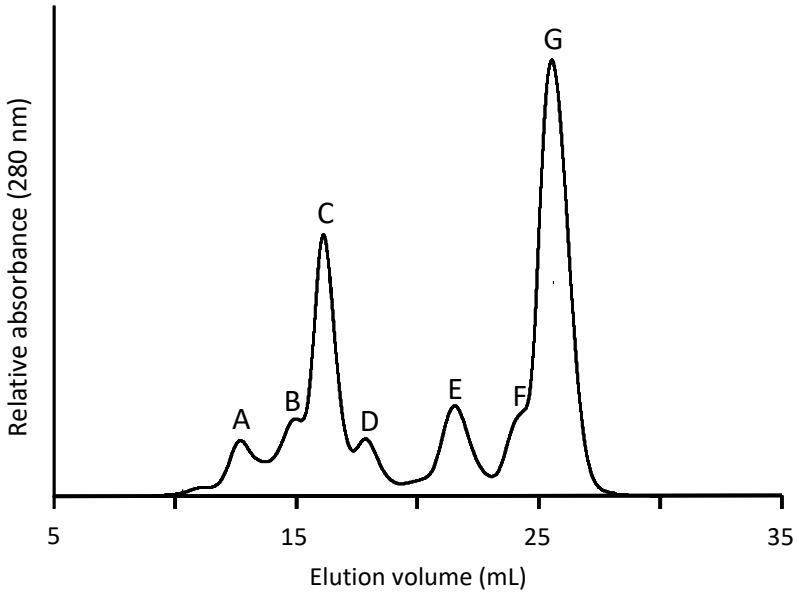


Figure 3.3: Size exclusion chromatography elution profile of *B. arietans* venom.
Separation was performed using a Superdex 200 GL column in 200 mM NH_4OAc (pH 6.8) and Fractions A-E were collected for further higher-order structural analysis.

It is likely the fractions contain multiple protein species of similar molecular weight and multiple isoforms of the same protein with slightly different modifications. Hence, SDS-PAGE was carried out on each fraction to reveal the protein complexity before performing MS analysis. It must be noted that, although a ladder of reference proteins is used to quantify the mass of the species observed, very compact proteins which are not fully denatured or exposed to SDS may appear smaller than their actual mass due to experiencing less resistance whilst migrating through the gel matrix. This is an important consideration in the analysis of venom proteins which often contain a high number of disulphide bonds, contributing to their compact nature. Hence, molecular weights identified by SDS-PAGE must only be used as a rough guide. However, SDS-PAGE still provides an overall picture of the protein species present within a sample.

Comparison of the SDS-PAGE gels under non-reducing and reducing conditions for SEC Fractions A-E gives the first indication that protein complexes exist within the venom

(Figure 3.4). Under non-reducing conditions, Fraction A shows an approximately 60 kDa species which appears to dissociate into a smaller 15 kDa species under reducing conditions. Similarly, Fraction B and C contain abundant species at approximately 15 kDa under reducing conditions which were not observed under non-reducing conditions. The presence of these small protein species under reducing conditions in fractions predicted to contain larger proteins suggest that smaller covalently linked species are comprising larger structures. Furthermore, it is evident that a variety of different protein species exist within each fraction at varying concentrations which may comprise some non-covalently linked higher-order structures. The structure of these assemblies requires further characterisation using a technique with higher precision, such as MS.

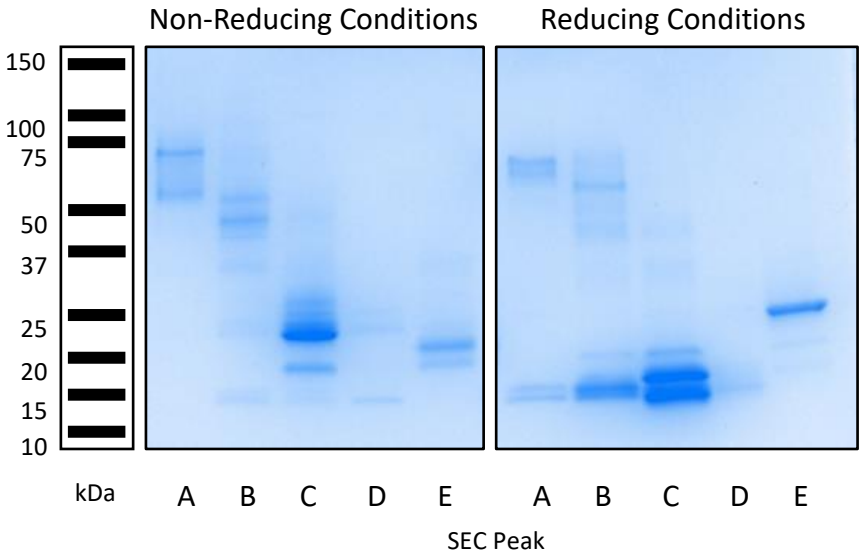


Figure 3.4: SDS-PAGE of purified *B. arietans* fractions.
SDS-PAGE was performed under both non-reducing and reducing (DTT) denaturing conditions using pooled fractions corresponding to SEC Fractions A-E from Figure 3.3. Protein samples in ammonium acetate (10 µL) were denatured with reducing (left) or non-reducing (right) running buffer (10 µL) at 96 °C for 15 min before being loaded onto the gel. Gels were stained with Coomassie before imaging.

3.3.2 MS Analysis of *B. arietans* Venom Components

Comparison of the protein species in each fraction under denaturing, reducing and native conditions was performed to provide further insight into the presence of higher-order protein assemblies in the *B. arietans* venom. Denaturing intact MS was first performed to identify the covalent structures of the proteins present in each SEC fraction. Following sample concentration, proteins were denatured with ACN (50 %) and formic acid (FA) (0.1 %).[130] Exposure to denaturing solvents causes protein unfolding by interfering with the non-covalent interactions usually stabilising the protein. This exposes amino acid residues usually hidden within the protein's core, allowing more highly charged states to be populated upon ionisation. Hence, successful denaturation was indicated in each sample by a series of highly charged ion species corresponding to each protein. To identify the presence of covalent disulphide bonds stabilising quaternary protein structures, intact MS analysis was also performed under reducing conditions. For these experiments, samples were treated with 3 mM TCEP for 5 h in the presence ACN (50 % v/v) and formic acid (FA) (0.1 % v/v). Finally, the fractions were also analysed by native MS to determine how the monomeric protein units are assembled under physiological-like conditions. The subsequent MS spectra were interpreted by deconvoluting the charge states from the *m/z* to obtain the protein's molecular weight. Changes in molecular masses under these different conditions permitted the identification of higher-order assemblies and gave insight into the bonding interactions through which they arise.

The denaturing MS spectra for Fraction A (Figure 3.5 A) shows a broad Gaussian distribution of poorly resolved signals. This series corresponds to a species of approximately 60 kDa with 5 identified isoforms of 59.2, 59.4, 59.5 59.7 and 60.1 kDa. The broadness of these signals is likely attributed to a variety of PTMs or slight differences in amino acid sequences. Upon reduction, the 60 kDa species is no longer present (Figure 3.5 B). Instead, a series of modified 14-15 kDa species is observed, with identified isoforms of 14.2, 14.3, 14.8, 15.1, 15.3, 15.7 and 15.8 kDa. This data suggests that four 15 kDa species make up the 60 kDa tetramer which is observed under denaturing conditions, and that these are linked by covalent disulfide bonds. The native spectra of Fraction A (Figure 3.5 C) revealed a predominant 120 kDa species, likely to be made up of two non-covalently linked 60 kDa species observed under denaturing conditions.

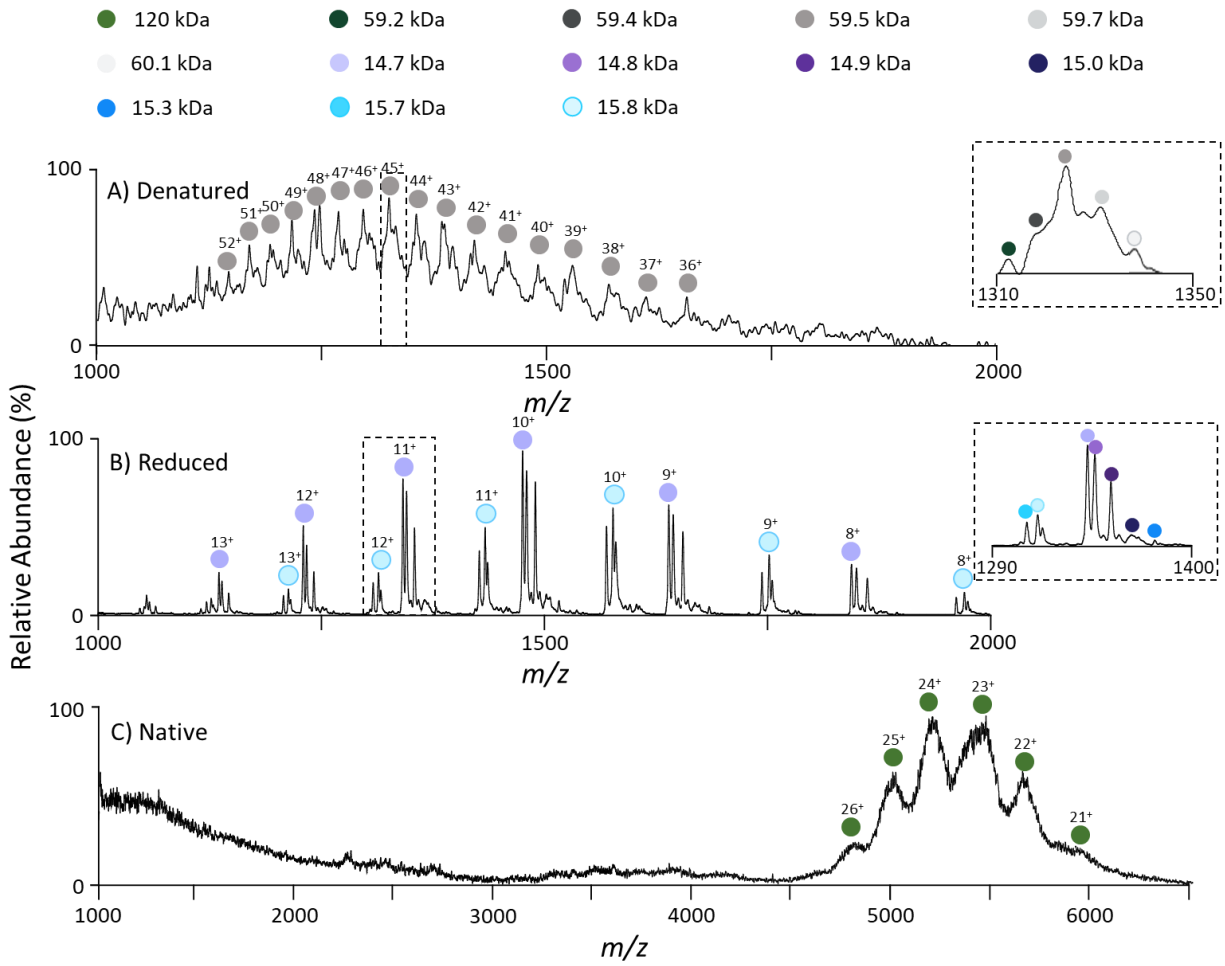


Figure 3.5: ESI-MS analysis of *B. arietans* venom Fraction A following SEC fractionation.
A) Denatured MS was performed in 50 % ACN, 50 % 200 mM NH_4OAc pH 7.0. **B**) Reduced MS was performed in 50 % ACN, 50 % 200 mM NH_4OAc. Samples were treated with 3 mM TCEP and denatured with 50 % ACN for 5 h before analysis. **C**) Native MS was performed in 200 mM NH_4OAc pH 7.0 by nanoESI. Identified protein masses are labelled with coloured circles and charge states are denoted above annotated fractions. Side panel insets show expanded *m/z* regions for selected dominant charge states, revealing detail of isoforms present.

Under denaturing conditions (Figure 3.6 A), Fraction B appears to contain a wide charge-state distribution of five 30 kDa isoforms with the 20+ charge state being most abundant. A 30.0 kDa species is most predominant, closely followed by a 30.5 kDa species. The 29.9, 30.1 and 30.6 kDa isoforms are in lower abundance. These species appear to dissociate into approximately 14-15 kDa monomers upon reduction with TCEP (Figure 3.6 B). Under native conditions (Figure 3.9 C), a 60.7 kDa species appears to be most predominant, with an abundant 18+ charge state. Given that this species is only observed under native conditions,

it is likely a dimer of ~30 kDa proteins. An additional low abundance 9 kDa species is detected in the native mass spectrum which may be hidden by the more abundant signals observed under denaturing and reducing conditions, reiterating the importance of fractionation in identifying the many proteins in venoms.

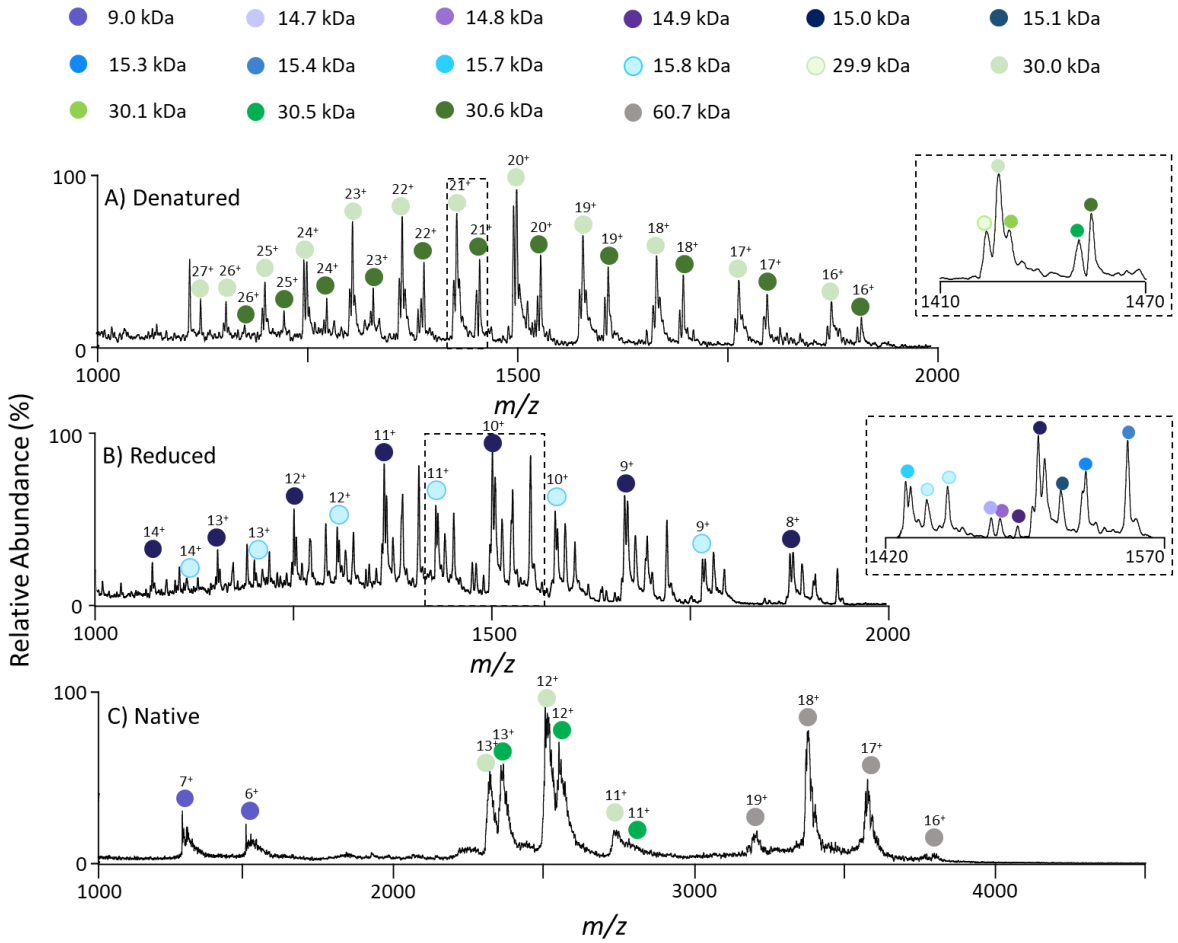


Figure 3.6: ESI-MS analysis of *B. arietans* venom Fraction B following SEC fractionation.
A) Denatured MS was performed in 50 % ACN, 50 % 200 mM NH_4OAc pH 7.0. **B**) Reduced MS was performed in 50 % ACN, 50 % 200 mM NH_4OAc. Samples were treated with 3 mM TCEP and denatured with 50 % ACN for 5 h before analysis. **C**) Native MS was performed in 200 mM NH_4OAc pH 7.0 by nanoESI. Identified protein masses are labelled with coloured circles and charge states are denoted above annotated fractions. Side panel insets show expanded *m/z* regions for selected dominant charge states, revealing detail of isoforms present.

Both the 30 kDa and 60.7 kDa assemblies identified in Fraction B were also identified in Fraction C (Figure 3.7), and the species observed under each condition support the

conclusions made in Fraction B. Given the overlap between Peaks B and C in Figure 3.3, the venom protein characterisation may be improved by using more extensive fractionation techniques such as ion exchange chromatography, to further isolate protein components.

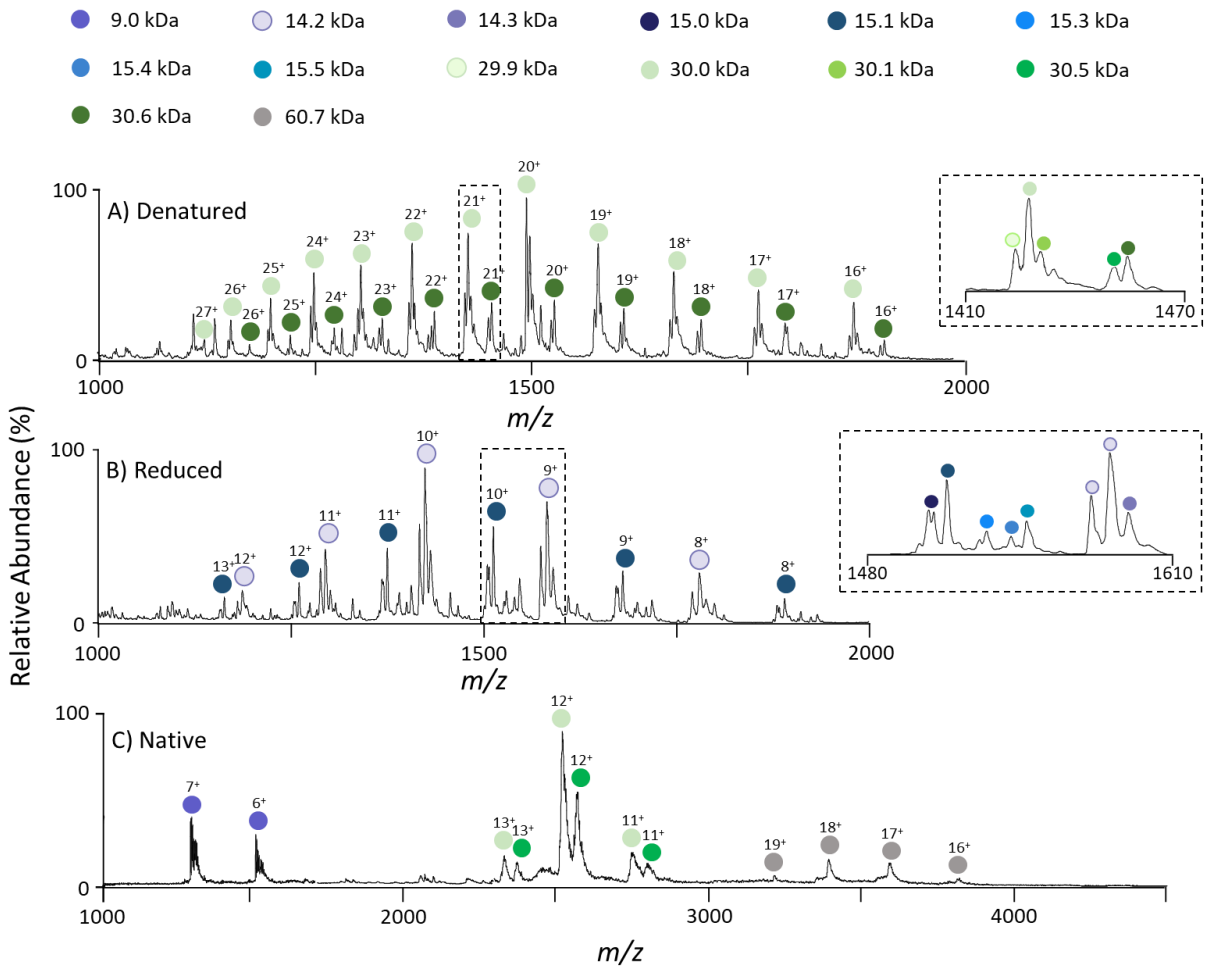


Figure 3.7: ESI-MS analysis of *B. arietans* venom Fraction C following SEC fractionation.
A) Denatured MS was performed in 50 % ACN, 50 % 200 mM NH_4OAc pH 7.0. **B**) Reduced MS was performed in 50 % ACN, 50 % 200 mM NH_4OAc. Samples were treated with 3 mM TCEP and denatured with 50 % ACN for 5 h before analysis. **C**) Native MS was performed in 200 mM NH_4OAc pH 7.0 by nanoESI. Identified protein masses are labelled with coloured circles and charge states are denoted above annotated fractions. Side panel insets show expanded *m/z* regions for selected dominant charge states, revealing detail of isoforms present.

The denatured spectrum for Fraction D showed a predominant species at 23.5 kDa with isoforms at 23.3 and 23.6 kDa, and a 9.0 kDa species with isoforms at 8.9 and 9.1 kDa (Figure 3.8 A). As was the case with Fraction B and C, the most abundant charge state for this 9.0 kDa species was 6+ and it was only distributed across three identified charge states. The 23 kDa

species had a much broader range of highly resolved charge states, with the most abundant being 19+. Under reducing conditions (Figure 3.8 B), no change was observed to the 23 kDa or the 9 kDa isoforms. However, some low abundant signals corresponding to 14.2 and 14.3 kDa species were observed, potentially from unresolved separation of the 30 kDa species found predominantly in Fractions B and C. Under native conditions (Figure 3.8 C), the 9 kDa species and the 23 kDa species were still observed, indicating that these species did not participate in any higher-order structures. Two low abundant charge states at 30.0 and 30.5 kDa were observed and identified as a dimer of 14 kDa species detected in the reduced spectra.

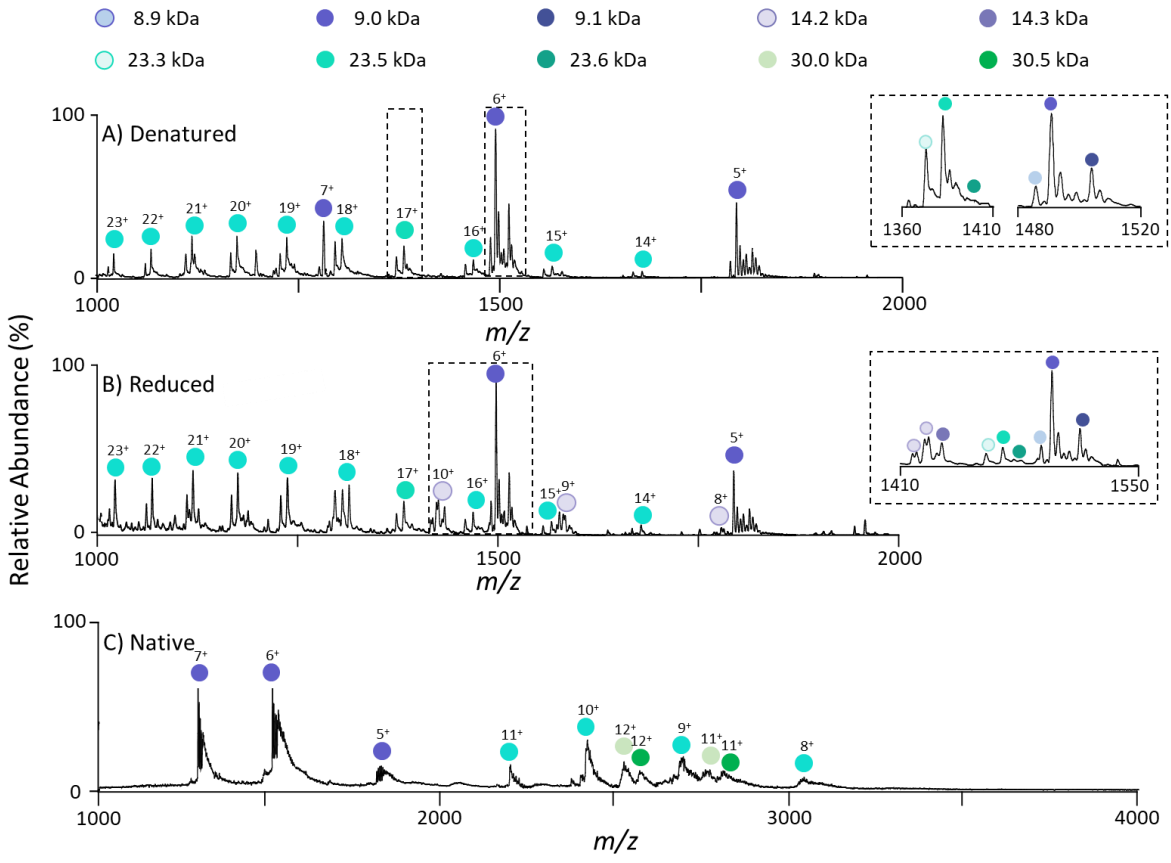


Figure 3.8: ESI-MS analysis of *B. arietans* venom Fraction D following SEC fractionation.
A) Denatured MS was performed in 50 % ACN, 50 % 200 mM NH_4OAc pH 7.0. **B**) Reduced MS was performed in 50 % ACN, 50 % 200 mM NH_4OAc. Samples were treated with 3 mM TCEP and denatured with 50 % ACN for 5 h before analysis. **C**) Native MS was performed in 200 mM NH_4OAc pH 7.0 by nanoESI. Identified protein masses are labelled with coloured circles and charge states are denoted above annotated fractions. Side panel insets show expanded *m/z* regions for selected dominant charge states, revealing detail of isoforms present.

Similarly, Fraction E contained the same 23 kDa isoforms identified in Fraction D (Figure 3.9 A-C). The consistent mass identified under all three conditions confirms that this protein is not involved in any higher-order protein assemblies, in agreement with the denaturing and reducing SDS-PAGE analysis in Figure 3.3.

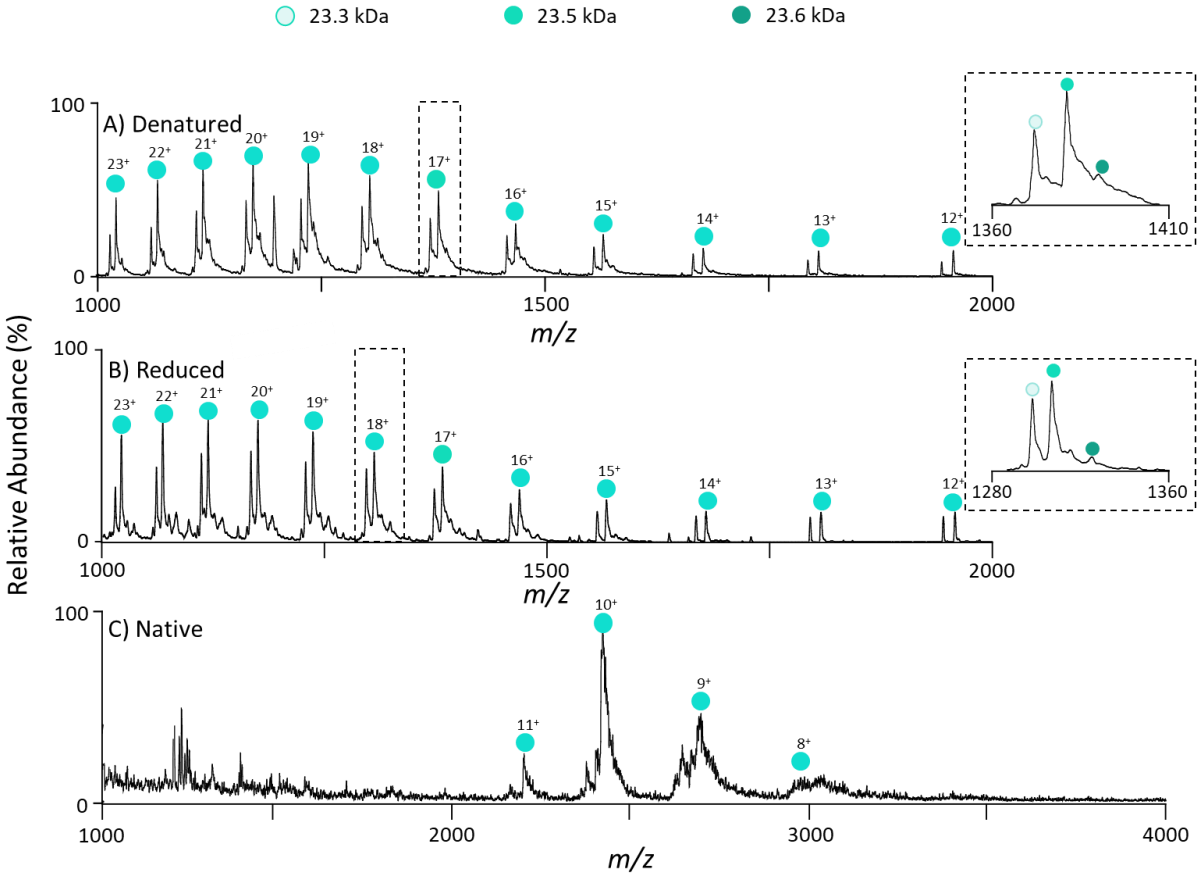


Figure 3.9: ESI-MS analysis of *B. arietans* venom Fraction E following SEC fractionation.
A) Denatured MS was performed in 50 % ACN, 50 % 200 mM NH_4OAc pH 7.0. **B**) Reduced MS was performed in 50 % ACN, 50 % 200 mM NH_4OAc. Samples were treated with 3 mM TCEP and denatured with 50 % ACN for 5 h before analysis. **C**) Native MS was performed in 200 mM NH_4OAc pH 7.0 by nanoESI. Identified protein masses are labelled with coloured circles and charge states are denoted above annotated fractions. Side panel insets show expanded *m/z* regions for selected dominant charge states, revealing detail of isoforms present.

3.3.3 Native IMMS

It must be noted that intact protein MS analysis is limited as it detects only signals of the most readily ionised species in the sample, which can cause suppression of low abundance species. Additionally, the ionisation efficiency which dictates how readily a protein ionises, may differ between protein species and between proteins and peptides. This may be influenced by the presence of PTMs such as glycosylation, the overall size and stability of the molecule, the amino acid composition and environmental factors such as buffer pH.[125,131,132]

To increase the SEC fraction capacity and investigate the presence of protein isomers, native IMMS was employed. By separating samples using the additional dimension of ion mobility, an investigation into the three-dimensional globular protein arrangement can be performed, and any otherwise undetected isoforms can ideally be revealed. Fraction B and C were selected as representative examples for further discussion in this thesis as they were of highest protein abundance and provided the highest quality data. The corresponding IMMS plots are detailed in Figure 3.10.

Generally, the protein arrival time distributions (ATDs), described by the features in Figure 3.10, show a narrow time distribution. This indicates that the proteins are not conformationally diverse. In Fraction B (Figure 3.10 A), protein species of 9.0 kDa, 30.0 kDa, 30.5 kDa and 60.7 kDa were identified, consistent with data in Figure 3.6. Given that protein isoforms of 30.0 and 30.5 kDa share identical ATDs, it is reasonable to assume they share similar conformation, despite a minor deviation in mass. In Fraction C (Figure 3.10 B), two isomers for the 30.0 kDa and 30.5 kDa species were identified by observation of two resolved ATDs. Additionally, a 29.2 kDa species was detected which was not identified using native MS alone, highlighting the value of enhanced peak capacity afforded by IM separation.

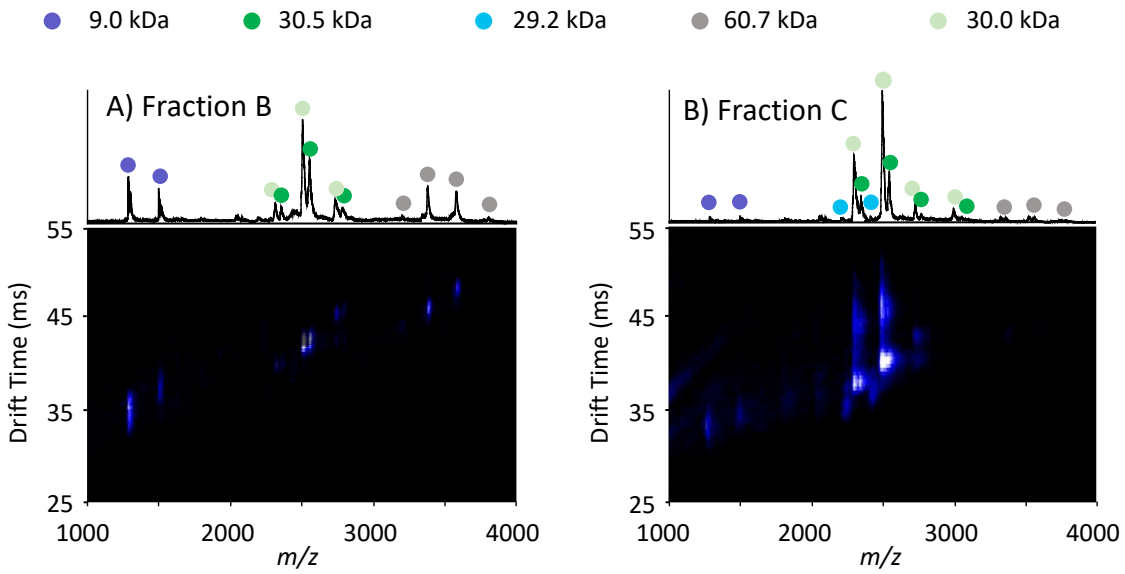


Figure 3.10: IMMS analysis following SEC fractionation of *B. arietans* venom.
Fraction B (**A**) and Fraction C (**B**) are presented as representative examples for the application of native IMMS to venom protein analysis. Native IMMS was performed in 200 mM NH_4OAc with an ion-mobility drift tube voltage of 1200 V. Identified protein masses are labelled with coloured circles.

Collision cross section (CCS) values were calculated for each ion species using the multifield CCS calculation feature on the Agilent IMMS browser B.07.01 (Agilent).By acquiring observed drift times for an ion under various drift field strengths, this calculation uses Mason-Schamp equation to calculate the CCS value for the corresponding ion.[133] Uncertainty of this calculation typically ranges between 0-2 %RSD.[134,135] These values are presented in Table 3.1 and provide information on the rotationally averaged area of the protein as it traverses the drift cell. As expected, smaller CCS values are associated with smaller protein species. The isomeric species observed for both 30.0 kDa and 30.5 kDa proteins in Figure 3.10 B possess notably different CCS values for ions of the same mass. For clarity, the isomers with shorter drift times are identified as (i) and those with longer drift times are identified as (ii). The CCS values of these more extended proteins isomers (ii) vary more substantially between ion states compared to the other isomer species (i). This is consistent with the typical behaviour of an unfolded protein, supporting the statement that isomers (ii) are less compact.[136]

Table 3.1: Calculated CCS values for venom proteins contained in SEC Fractions B and C.

FRACTION	MASS (kDa)	CHARGE	CCS (Å^2)
B	9.0	+7	1379
		+6	1546
B	30.0	+13	2663
		+12	2790
		+11	2663
B	30.5	+13	2827
		+12	2824
		+11	2783
B	60.7	+19	4762
		+18	4613
		+17	4506
C	9.0	+7	1325
		+6	1259
C	29.2	+13	2632
		+12	2629
C	30.0 (i)	+13	2818
		+12	2782
		+11	2748
C	30.5 (i)	+13	2799
		+12	2792
		+11	2740
C	30.0 (ii)	+13	3253
		+12	3143
		+11	3023
C	30.5 (ii)	+13	3331
		+12	3126

To relate the CCS value to protein compactness, the effective density (D_{eff}) was calculated for each protein species using the method described by Bush *et al.*[137] In this method it is assumed that proteins adopts a predominantly spherical shape so that insight can be obtained on the degree of compactness adopted by the structure. The calculations proposed by Bush *et al*. assume the CCS values were recorded in helium. Since CCS values were recorded here in N_2, a correction factor was applied by determining the average ratio between CCS values measured in nitrogen and helium, using the CCS database developed by Bush et al.[137] The average CCS value measured experimentally in nitrogen was divided by this value as described in Equation 1:

$$\Omega_{He} = \frac{\Omega_{N_2\,(Measured)}}{1.160} \qquad (1)$$

Subsequently, the effective protein radius (r_{eff}) was calculated from the average CCS values in helium for all observed charge states (Ω_{He}), as described in Equation 2:

$$r_{eff} = \sqrt{\frac{\Omega}{\pi}} - r_{He} \quad \text{(where } r_{He} = 1\text{)} \qquad (2)$$

The effective protein volume (V_{eff}) was then calculated using Equation 3:

$$V_{eff} = \frac{4}{3}\pi r_{eff}^3 \qquad (3)$$

Finally, the effective density (D_{eff}) was calculated from Equation 3 using the molecular weight (MW) of the protein and Avogadro's number (N_0) as described in Equation 4:

$$D_{eff} = \frac{MW}{N_0} \times \frac{1}{V_{eff}} \qquad (4)$$

D_{eff} values were converted to g/cm^3 for comparison to literature to determine the level of compactness of the protein (Table 3.2). Previous studies have reported native-like proteins to have a D_{eff} value of 0.6 g/cm^3.[137] The D_{eff} values calculated in this experiment are within close proximity to this literature value. Hence, it is reasonable to presume that these protein assemblies lack significant extended structural components, consistent with the high number of disulphide bonds understood to stabilise venom protein structures.

Table 3.2: Effective densities calculated from IMMS.

FRACTION	PROTEIN MASS (kDa)	D_{eff} (g/cm^3)
B	9.0	0.52
B	30.0	0.66
B	30.5	0.63
B	60.7	0.58
C	9.0	0.63
C	29.2	0.63
C	30.0 (i)	0.52
C	30.5 (i)	0.64
C	30.0 (ii)	0.51
C	30.5 (ii)	0.66

Overall, this data demonstrates the application of IMMS analysis for identification of species with identical *m/z* and for studying the level of protein compactness. Furthermore, these results validate the use of this technique for future high-throughput structural characterisation and protein modelling.

3.3.4 Proteomic Analysis of fractionated *B. arietans* venom

Bottom-up proteomic analysis was performed on each major SEC fraction to identify the types of toxins present within each sample.[138] Results were searched against the *Serpentes* database encompassing all known snake proteins, and only those exhibiting toxic function were selected for analysis. The relative abundance of each protein family was determined by summing the number of PSMs for peptides within each protein group. This data is presented in Figure 3.11, and the protein family acronyms are summarised in Table 1.1. Further details corresponding to the protein homologies, protein masses, peptide *m/*z and peptide sequences are detailed in Appendix B, Tables B1-B6.

Sixteen toxin families were identified across the five fractions, with CTL being the most abundant in Fractions A and B, and DIS being the most abundant in Fractions C, D and E. These two protein families have been identified as major components in *Bitis arietans* previously, and participate in the disruption of homeostasis, prevention of platelet aggregation and ultimately haemorrhage, all of which are characteristic symptoms of viperid envenomation.[2,11,113] Previous studies however have also found a higher abundance of SVMP and SVSP in *B arietans* venom.[2,113] As stated earlier, variation in venom protein composition can occur within a species based on its geographical location and its ontogenetic stage of life.[2,4] Hence, it is important that protein species within the venom are identified based on the proteome of the exact venom under analysis, rather than a related proteome database found in literature.

In total, eight toxin families were identified in Fraction A with CTL being most abundant followed by SVMP. Ten toxin families were identified in SEC Fraction B where CTL and SVSP were the most abundant protein families. Eleven families were identified in Fraction C, where DIS comprised over 50 % of the population, with CTL being the next most abundant. Fraction D and E were also predominantly comprised of DIS, with a total of ten and seven protein families identified in the sample respectively. Generally, the masses of the protein hits

identified in Appendix B are within a similar mass range to the species identified in the intact MS experiments discussed in the sections above. To summarise the two most abundant species, proteins within the CTL family ranged between 12.7-19.3 kDa, and proteins within the DIS family generally ranged between 5.8-13.8 kDa. It is expected that proteins within a toxin family will vary slightly in mass but will exist in a similar mass range and hence, it is likely that the species identified earlier in this chapter can be aligned with this proteomic data.

Notably, the SDS-PAGE analysis presented in section 3.3.1 and MS spectra presented in section 3.4.2 do not suggest the same venom diversity as the proteomic data in Figure 3.11. As mentioned previously, MS analysis is limited by the ionisation efficiency and complexity of the sample. Proteomic analysis separates digested peptides by LC/MS and hence is less subjective to concentration dependent signal suppression, identifying a larger variety of protein species. Similarly, given the high level of homology between toxin proteins, it is possible that some peptide hits may overlap between proteins within a family, and hence the PSM may be overestimated for some protein families. Therefore, although this proteomic data is useful in identifying possible protein families present within the venom sample, it cannot be used to identify the individual species detected by MS alone.

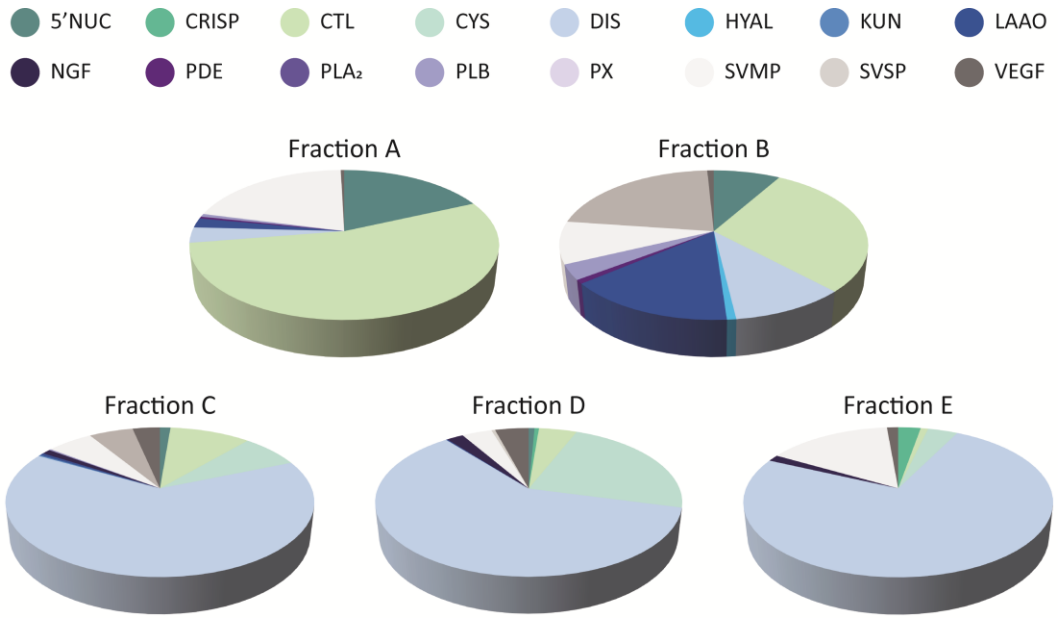


Figure 3.11: Proteomic composition of *B. arietans* venom fractions.
Bottom-up proteomic analysis was performed on venom SEC fractions to identify toxin families present. The total peptide spectrum matches (PSMs) for each toxin family were summed and presented to provide a semi-quantitative measure of relative abundance. The individual proteins, along with their masses, peptide sequences and *m/z* are summarised in Appendix B. Protein family acronyms are described in Table 1.1.

3.3.5 Cross-Linking of Venom

To further investigate the higher-order structures stabilised under native conditions, cross-linking reagents were applied to the venom samples prior to SDS-PAGE analysis (Figure 3.12). This was carried out under denaturing and reducing conditions, where higher-order assemblies should only be visible after cross-linker stabilisation. Initially, commercially available DSSO was applied to Fractions A-E in a 20-molar excess to total protein. Cross-linking reactions were carried out in native like conditions using PBS rather than ammonium acetate which would have quenched the amine reactive linkers.

Under reducing conditions alone, Fraction A presented two major species at approximately 60 kDa and 15 kDa, which is consistent with MS results. When cross-linked, an additional band around 120-140 kDa is observed which is in good agreement with the 120 kDa species detected by native MS. Furthermore, the intensity of the 60 kDa and 15 kDa bands reduces, supporting the likelihood of higher-order structure formation. The inefficiency of cross-linking reactions justifies the observation of the 60 kDa monomeric species under cross-linking

conditions. In Fractions B and C, bands at around 15 kDa are observed under reducing conditions. Upon cross-linking, an additional band at approximately 30 kDa is observed and the intensity of the 15 kDa band is reduced. This also supports the claim that a 30 kDa species is composed of two 15 kDa monomers, however, does not support the claim that this species dimerises to form a 60.7 kDa species. It is possible that the cross-linker used is not of optimum length for stabilizing the two 30 kDa species in the 60.7 kDa dimer. No major changes were observed under cross-linking conditions for Fractions D or E, supporting previous claims that the species found in these fractions are not involved in any multi-subunit protein association.

The cross-linkers **9bii**, **10** and **20** synthesised in Chapter 2 were applied to the protein species in Fraction B and compared to commercially available DSSO and DSP (Figure 3.12 B). The band observed at approximately 30 kDa indicates that DSSO was the most efficient cross-linker, closely followed by linkers **9bii** and **10**. Linker **20** and DSP are comparable to the negative control, suggesting they were ineffective in forming cross-links. Although DSSO was identified as the most effective cross-linker, Figure 3.12 B demonstrates that linkers **9bii** and **10** have potential in characterising venom protein systems. However, given the limited quantities of novel linkers available, DSSO was used for structural characterisation in the rest of this thesis.

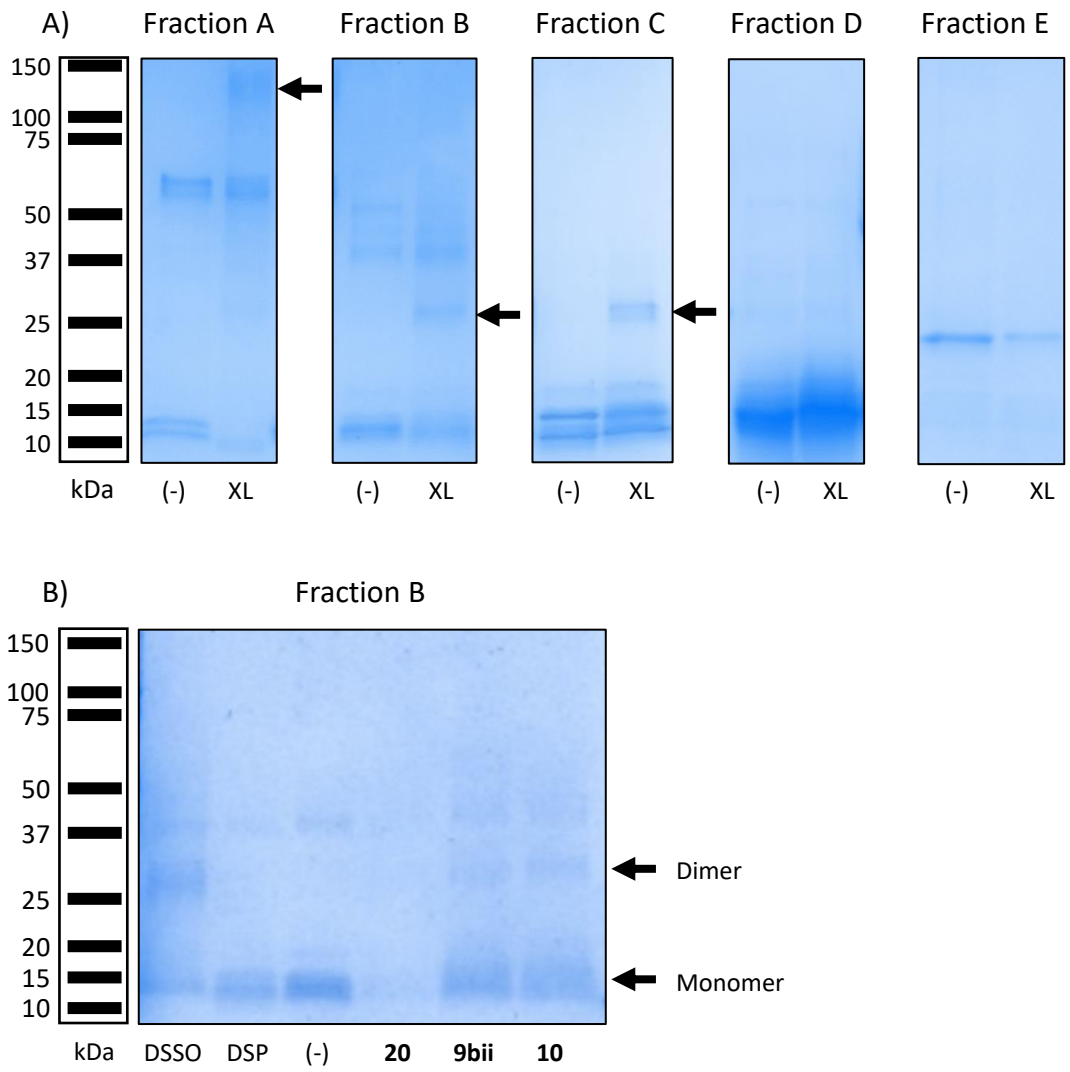


Figure 3.12: SDS-PAGE analysis of cross-linked *B. arietans* fractions.
A) SEC Fractions A-E were cross-linked with 20 molar equivalents of DSSO to protein. SDS-PAGE was performed under reducing and denaturing conditions in the absence (-) and the presence (XL) of the linker. **B**) Linkers **9bii**, **10** and **20** from Chapter 2 were applied to fractions corresponding to Fraction B and compared to commercial linkers DSSO an DSP. SDS-PAGE was performed under reducing and denaturing conditions in the absence (-) and the presence of the linker.

To identify the monomeric counterparts of the oligomeric protein species detected in the venom, the cross-linked stabilised oligomers from the SDS-PAGE in Figure 3.12 A were digested in-gel and subjected to proteomic analysis. As predicted, the list of identified protein hits in the individual dimer band was more concise than that for the whole fraction and is summarised in Appendix B, Tables B6-8. The approximate masses of the protein hits were compared with the masses identified in the above analysis. It must be noted that the protein hits in the proteomic analysis are based on overlapping sequences with pre-characterised

protein structures. The actual proteins within this venom are likely to contain similar moieties and sequences, however, are likely to be uncharacterised proteins with additional unique features and PTMs giving them different masses. Hence, the molecular weights for protein hits were only used as a rough guide in identification of the species under analysis.

Upon digestion of the 120 kDa band observed under cross-linking conditions in Fraction B (Figure 3.12 A), 18.1 kDa CTL, 96.1 kDa PDE and a 68.1 kDa SVMP candidates were identified by bottom-up proteomics (Appendix B, Table B6). Initially, it was predicted that this 120 kDa species was composed of two 68.1 kDa SVMP monomers. However, given that further dissociation of this structure was observed under reducing conditions in both SDS-PAGE (Figure 3.12) and MS analysis (Figure 3.6), it is more likely that this species is composed of a CTL oligomer. Given the masses observed in the MS analysis in Figure 3.6, it is understood this 120 kDa structure is an octamer composed of two 60 kDa non-covalently linked tetramers. The 15 kDa monomeric components of this tetramer are likely disulphide stabilised as they were only detected under reducing conditions. Although PDE and SVMP are different toxin families to CTL, it is possible their identification in this dimer band may be due to an overlap in sequence similarity.[2,11,113] The 30 kDa species observed under cross-linking conditions in both Fraction B and C (Figure 3.12 A) is also believed to be a higher-order structure comprised of CTL monomers based on the large number of CTL hits identified from digestion of this band in both fractions. The CTL matches identified in this fraction (Appendix B, Tables B7 and 8) were of mass ranges between 14.4 and 18.1 kDa, suggesting two CTL monomers have self-associated and are stabilised by disulphide bonds to form a 30 kDa dimer.

Given that the proteomic data in Figure 3.11 identifies CTL as the most abundant species in Fraction A and B, and the second most abundant species in Fraction C, the likelihood of these species being CTL is high. Furthermore, these claims are consistent with literature which has reported oligomerisation of CTLs on other systems previously.[2,5,139] Walker *et al.* reported the x-ray structure of the CTL rattle-snake venom lectin (RSL) isolated from *Crotalus atrox* venom.[140] This structure revealed that RSL is an oligomeric protein composed of five disulphide linked dimers, forming an overall decameric protein (Figure 3.13). RSL is from a family of CTL's isolated from other viperid and elapid venoms previously and hence, is likely related to the CTL oligomeric structures identified from *B. arietans* in this study. Furthermore, the cysteine-rich character of CTLs means these proteins can form disulphide bonds to

stabilise such higher-order structures. This supports the fact that the monomeric counterparts were only observed under reducing conditions in both the MS and gel analysis.[2,5,139] Thus, the oligomeric CTL structures proposed in this thesis are feasible.

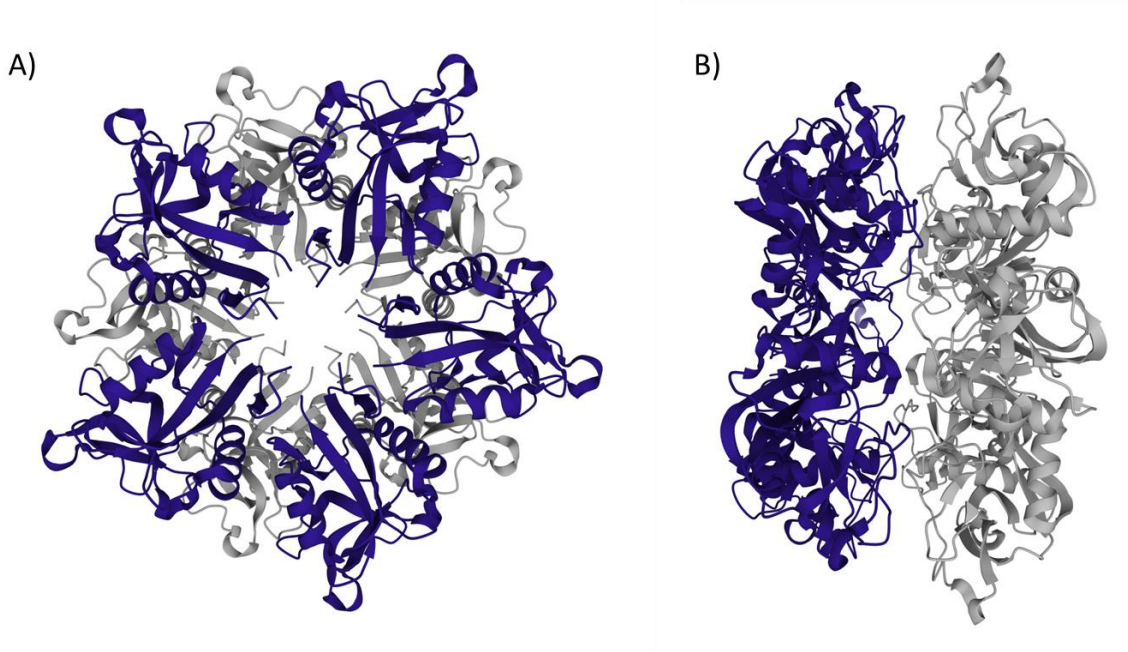


Figure 3.13: X-ray crystal structure of rattle snake venom lectin (PDB ID: P21963).[107,140]

Ribbon diagrams of the decamer structure of RSV, where one monomer from each pentameric dimer is labelled in purple and the other in grey. **A**) Top-down view where pentameric dimers can be visualised in each colour. **B**) Side on view showing the division between each monomer.

3.4 Chapter Conclusions and Future Work

The MS-based workflow described in section 3.2 was used to identify several higher-order protein assemblies within the venom. Fraction A contained a 120 kDa dimer, comprised of two 60 kDa species held together non-covalently. These 60 kDa species appeared to be comprised of four 15 kDa CTL species linked covalently (Figure 3.14 A). The MS data indicates that Fraction B contained a 60.7 kDa homodimer comprised of two non-covalently linked 30 kDa species. These monomers were each comprised of two 15 kDa species linked covalently (Figure 3.14 B). Unfortunately, this assembly could not be stabilised by XL SDS-PAGE in Figure 3.13, potentially requiring a different length XL reagent. Both Fraction B and C contained 30 kDa dimers, comprised of two covalently linked 15 kDa species also identified

as CTL. Self-association of CTL has been reported previously, supporting the identification of these oligomeric species.[5,140] Fractions B, C and D contained a monomeric species of around 9 kDa which did not appear to participate in any quaternary structures. Fractions D and E also contained a monomeric species of 23 kDa which again, did not appear to participate in any higher-order structures. It must be noted that the SDS-PAGE analysis suggests these fractions contain other species undetected by the MS analysis alone. Hence, further investigation may be possible with additional sample fractionation.

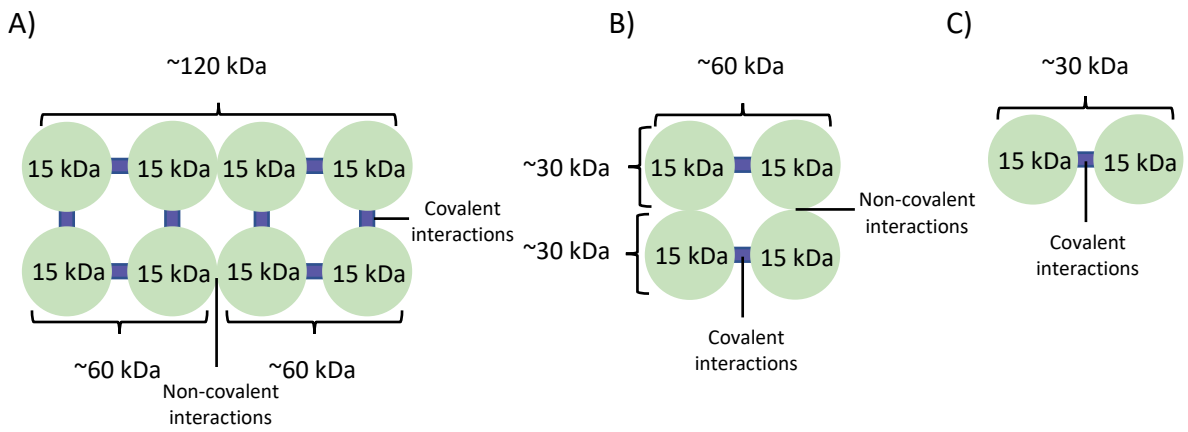


Figure 3.14: Proposed structure of the protein assemblies in *B. arietans* venom.
A) Four covalently linked 15 kDa monomers forming a 60 kDa tetramer. Two of these 60 kDa 'dimeric' assemblies associate non-covalently to form a 120 kDa dimer, based on the intact and native MS analysis described above to essentially form a 'dimer of dimers'. **B**) Proposed structure of the 60.7 kDa tetramer from Fraction B with four 15 kDa monomers associated to form a 60.7 tetramer. Two of these monomers are covalently linked forming two 30 kDa dimers. **C**) Proposed structure of the 30 kDa dimer from Fraction B and C with two 15 kDa monomers associated to form a 30 kDa dimer linked covalently.

Overall, this MS-based approach appears to characterise the dominant species in the sample. Utilising a range of complementary structure determination approaches including SDS-PAGE, denatured and reducing intact MS, native MS and proteomic analysis clearly demonstrates the sample complexity and the presence of higher-order protein structures in *B. arietans* venom. Furthermore, establishing this MS-based structural determination workflow provides promising potential in other biological systems. Future work may implement top-down proteomic analysis where sequencing is performed on intact proteins rather than peptides,

and selected MS signals can be targeted directly for identification. This minimises any ambiguity in identifying proteins in complex samples. Additionally, IMMS could be combined with CID experiments using an instrument capable of selecting proteins with large *m/z* to observe the controlled dissociation of the non-covalent species by MS. Again, this will assist in minimising ambiguity in the dissociation of individual species. Furthermore, XLMS should be carried out on all identified structures to determine the residues involved in stabilising the higher-order structures. This may assist in identifying surface areas of the protein involved in crucial interactions for envenomation. Sequence coverage and protein identification may be improved on XL samples using top-down proteomics as described above. Additionally, more definitive protein identification may also be carried out using transcriptomics. Here, the total ribonucleic acid (RNA) molecules present within a sample is studied to determine which genes are expressed within an organism at a given time. This may assist in understanding what toxins can be produced and under what conditions, and to more definitively assign protein identities from bottom-up proteomic analyses.

Overall, several higher-order protein structures have been identified in *B. arietans* venom using this MS-based workflow and the types of toxin proteins which may be participating in these structures have been identified.

3.5 Materials and Methods

3.5.1 Materials

All reagents were purchased from Sigma Aldrich (New South Wales, Australia) or ThermoFisher Scientific (Massachusetts, USA) unless otherwise specified. Precast gels (4-15 % MiniProtean® TGX™ polyacrylamide precast gel) and protein markers (Precision Plus Dual Colour protein standards) were purchased from Bio-Rad (California, USA). Whole lyophilised *B. arietans* venom was purchased from Venom Supplies Pty. Ltd. (South Australia, Australia), and was stored at -20 °C until required.

3.5.2 General Methods

200 mM ammonium acetate (NH_4OAc) buffer (pH 7.0) was filtered using a Nalgene Rapid-Flow bottle top filter unit (Thermo Fisher Scientific, Massachusetts, USA) and de-aerated with an ultrasonic cleaner (Soniclean, South Australia, Australia) before SEC and MS analysis.

SDS-PAGE

Protein samples (10 µL) were combined with 3x gel loading buffer (10 µL) and denatured at 96 °C for 15 min. Samples were separated on a 4-15 % Miniprotean® TGX™ polyacrylamide precast gel (Bio-Rad) with Precision Plus Dual Colour protein standards (Bio-Rad) as molecular weight markers. Electrophoresis was performed at 140 V in 1x SDS TGS running buffer for 45 min. The resulting gel was stained at with Coomassie Brilliant Blue stain at room temperature for 3-17 h, before de-staining with Coomassie destain for 1-3 days. Gels were imaged on the Bio-Rad ChemiDoc MP imaging system (Bio-Rad) using the Coomassie method.

Coomassie Brilliant Blue staining solution was made up of 0.2 % (w/v) Coomassie Brilliant Blue R250, 40 % ethanol (v/v) and 10 % (v/v) glacial acetic acid. Coomassie de-stain consisted of 40 % (v/v) methanol, 10 % (v/v) acetic acid.

SDS-PAGE loading buffer (3x reducing) consisted of 150 mM Tris-HCl, 300 mM DTT, 6 % SDS, 30 % glycerol and 0.3 % (w/v) bromophenol blue, pH 6.8. SDS-PAGE loading buffer (3x non-reducing) consisted of 150 mM Tris-HCl, 6 % SDS, 30 % glycerol and 0.3 % (w/v) bromophenol blue, pH 6.8.

1x SDS-tris-glycine running buffer was diluted from 10x running buffer (25 mM tris, 192 mM glycine, 0.1 % SDS, pH 8.5).

Gels were imaged using a ChemiDoc MP imaging system (Bio-Rad) using the Coomassie method.

3.5.3 Methods

Size Exclusion Chromatography (SEC)

Lyophilised whole *B. arietans* venom (10 mg) was resuspended in 200 mM NH_4OAc (pH 6.8) (1 mL) and loaded onto a Superdex200 10/300 size exclusion column (GE Healthcare, Illinois, USA) coupled to an AKTA Prime FPLC system (Amersham, UK), following column equilibration with 200 mM NH_4OAc (pH 6.8) prior to loading. Fractions of 400 µL were collected at a flow rate of 0.4 mL/min with 200 mM NH_4OAc (pH 6.8) as the eluent over a volume of 36 mL. Samples were stored at -20 °C until required.

Denatured MS

Denatured protein mass spectra were obtained using a 1260 LC system coupled to a 6230 TOF mass spectrometer (Agilent Technologies) using ESI. Samples (2 µL) in ACN (50 %, 0.1 % FA) were directly injected at a flow rate of 0.5 mL/min in 50 % solvent A, 50 % solvent B without chromatographic separation. The instrument conditions were set as follows: *m/z* range 500-3200; polarity positive; capillary voltage 3.5 kV; nozzle voltage, 2 kV; gas temperature 325 °C; gas flow 8 L/min. Mass spectra was acquired using Agilent MassHunter Workstation Data Acquisition (vB.08.00, Agilent Technologies) and data was analysed using Agilent MassHunter Qualitative Analysis Software (vB.07.00).

Solvent A was 0.1 % (v/v) FA and solvent B was 99.9 % (v/v) ACN 0.1 % (v/v) FA.

Reduced and Denatured MS

Samples were prepared by incubated with TCEP (3 mM) and ACN (50 %) for 5 h prior to MS analysis. Samples were analysed using the 1260 LC system coupled to a 6230 TOF mass spectrometer (Agilent Technologies) employing ESI and were separated on an Aeris Widepore C4 LC column (150 x 4.6 mm). Samples (10 µL) in ACN (50 %, 0.1 % FA) were injected at a flow rate of 0.8 mL/min starting at 90 % solvent A. For the first 3 min, sample was directed to waste to eliminate free TCEP. Subsequently, solvent A was dropped to 10 %, increasing to 90 % solvent A over 45 min. The instrument conditions were set as follows: *m/z* range 500-3200; polarity positive; capillary voltage 4 kV; nozzle voltage, 2 kV; gas temperature 325 °C; gas flow 13 L/min. Agilent MassHunter Workstation Data Acquisition (vB.08.00, Agilent Technologies) was used to acquire mass spectra, and Agilent MassHunter Qualitative Analysis Software (vB.07.00) was used for data analysis.

Solvent A was 0.1 % (v/v) FA and solvent B was 99.9 % (v/v) ACN 0.1 % (v/v) FA.

Native and Ion Mobility MS

All native MS spectra were obtained using an Agilent 6560 IMMS (Agilent Technologies) instrument. Samples (4 µL) were introduced into the instrument by nano-ESI using platinum-coated borosilicate capillaries prepared in-house. The instrument conditions were set to preserve non-covalent interactions as follows: *m/z* range 500-10000; polarity positive; collision energy 10; capillary voltage 1.3 kV; gas temperature 50 °C; gas flow 10 L/min. IMMS

conditions were as described above, in addition to IM drift time range (0-100); trap fill time (20,000 µs); trap release time (4000 µs) and the IM drift tube voltage was set to increase incrementally by 100 V from 1200 V to 1700 V every 30 sec to allow for multi-field cross-section calculation. MS data acquisition was controlled using Agilent MassHunter Workstation Data Acquisition software (vB.08.00, Agilent Technologies). Native MS data analysis was performed using Agilent MassHunter Qualitative Analysis software (vB.07.00) and IMMS data analysis was performed using Agilent IM-MS Browser B.07.01. CCS values were determined using the Agilent CCS multifield calculation function.

Filter-Aided In-Solution Digestion

Protein samples (~0.1 mg) were denatured with 7 M urea/100 mM NH_4HCO_3 (200 µL) and concentrated by centrifugation using Amicon Ultra-0.5 mL centrifugal filter units (MerckMillipore, Darmstadt, Germany) with a 10 kDa molecular weight cut-off (14,000 x g, 15 min). Samples were incubated with DTT (100 µL, 50 mM in 7 M urea/100 mM NH_4HCO_3) for 1 h which was removed by centrifugation (14,000 x *g,* 15 min). Subsequently, samples were incubated with IAA (100 µL, 55 mM in 7 M urea/100 mM NH_4HCO_3) for 20 min in darkness before removal by centrifugation (14,000 x g, 15 min). To ensure complete removal of excess DTT and IAA, samples were washed with 7 M urea/100 mM NH_4HCO_3, 50 mM NH_4HCO_3 and 10 mM NH_4HCO_3 by centrifugation (14,000 x g, 15 min). The concentrated proteins were resuspended in trypsin solution (100 ng/µL in 5 mM NH_4HCO_3) with a mass ratio of 1:50 (enzyme:protein). Samples were incubated overnight at 37 °C before digested peptides were eluted from the spin-filter using 1 % FA (50 µL). Samples were dried through vacuum centrifugation before being resuspended in 2 % ACN 0.1 FA (100 µL). Peptide concentrations were verified before analysis using a NanoDrop 2000/2000c UV-Vis spectrophotometer (Thermo Scientific, Massachusetts, USA) at a wavelength of 205 nm and ε_{205} of 31 mL mg^{-1} cm^{-1} as per the manufacturer's instructions, before analysis.

Proteomic MS Data Analysis

Digested samples were analysed using an Ultimate 3000 RSLCnano system coupled to an Orbitrap Exploris 480 mass spectrometer (ThermoFisher Scientific, Bremen, Germany). Approximately 300 ng of each peptide sample was loaded onto a 25 cm fused silica column with an internal diameter of 75 µm, packed with 1.9 µm C_{18} particles, and heated to 50 °C.

Peptide separation was performed over a 10 min linear gradient (3 to 28 % ACN in 0.1 % FA) at a flow rate of 300 nL/min. Ionised peptides were filtered into the mass spectrometer using compensation voltages (-50 and -70 V) generated by a FAIMS Pro interface (Thermo Scientific). Mass spectra were acquired over the mass range of 300-1500 *m/z* in positive mode at a resolution of 60,000 (200 *m/z*) in FT mode. Data-dependent MS/MS was performed on multiply charged precursors with an intensity greater than 10^6 at a resolution of 15,000 following the application of 27.5 % HCD collision energy. A dynamic exclusion period of 10 sec was specified.

The MS/MS data obtained in RAW file format was analysed in Proteome Discoverer 2.4.1.15 (Thermo Fisher). Qualitative protein identification was carried out using the CWF Basic Annotation consensus workflow. In the processing workflow, spectra were searched against the *Serpentes* protein database with search parameters as follows: Spectrum Files RC compared Trypsin (Full) data against the Serpentes FASTA file downloaded from UniProt Spectrum selector used MS^1 precursor selection; Precursor Detector used a S/N Threshold of 1.5; Sequest HT allowed tryptic peptides with a maximum of 2 missed cleavages, with a min. peptide length of 6 and a max peptide length of 144. The peptide mass tolerance was 10 ppm and the fragment mass tolerance was 0.02 Da. Dynamic modifications of methionine oxidation, asparagine and glutamine deamination, Acetylation of the N termini and Methionine loss of the N termini were included. Static carbamidomethylation modifications of cysteines were also included. Data was selected using the Perculator node. Non-toxin protein hits were eliminated from the resulting data, and toxins corresponding to the same protein family were grouped.

Cross-linking of venom with DSSO and novel cross-linkers

Venom fractions corresponding to SEC Fractions A-E were lyophilised to remove any ammonium acetate which would quench the cross-linking reaction. Samples were resuspended in PBS (50 µL), and were cross-linked with linkers **9bii**, **10**, **20**, DSSO and DSP (5.26 µL, 70 mM in DMSO) in a linker to protein ratio of 20:1, with a final volume of 5 % DMSO. Samples were incubated at room temperature with shaking (100 rpm) for 3 h before the sample containing linker **20** was illuminated at 365 nm, 1 cm from the light source for 45 min. Each cross-linked sample was separated by SDS-PAGE under reducing conditions along with an unmodified sample for each fraction.

Following excision of the gel band, gel pieces were washed with NH_4HCO_3 (500 µL, 50 mM), and the Coomassie stain was removed through three 15 min sonicating incubations with NH_4HCO_3 (400 µL, 50 mM in 30 % ACN). Following solvent removal, gel pieces were incubated with ACN (200 µL) for 15 min before removal *in vacuo* and were then resuspended in DTT (50 µL, 10 mM in 100 mM NH_4HCO_3) at 56 °C for 45 min. Following solvent removal, gel pieces were again incubated with ACN (200 µL) for 15 min before removal *in vacuo* and resuspension in IAA (50 µL, 55 mM in 100 mM NH_4HCO_3) and incubation in darkness for 30 min. Following solvent removal, gel pieces were washed with NH_4HCO_3, (100 µL, 5 mM) before being resuspended in ACN (200 µL) for 15 min. Following solvent removal, gel pieces were resuspended in a solution of trypsin (10 µL, 10 ng/µL in 5 mM NH_4HCO_3) and incubated for 15 min. Additional NH_4HCO_3 (10 µL, 5 mM in 20 % ACN) was added before the sample was incubated at 37 °C overnight. Peptides were extracted from gel pieces by 15-min incubations and extractions with formic acid (20 µL, 1 % in water), (50 µL, 1 % in 50 % ACN) and ACN (100 µL) before solvent was removed *in vacuo*. Samples were resuspended in ACN (5 µL, 3 % in 0.1 % FA) and analysed by MS according to the details described above.

Chapter 4: Summary and Conclusions

The work described in this thesis demonstrates the use of an integrative MS approach in studying complex protein systems. Using this approach, the protein components in *B. arietans* venom have been characterised, with a focus on higher-order protein assemblies.

Chapter 2 presents the design and synthesis of a novel modular cross-linking library with varying reactive groups, linker arms and affinity tags. These three components can be adapted for cross-linker functionalisation, overcoming many of the challenges associated with the cross-linking MS workflow, including protein system compatibility, linker stability, linking efficiency and data complexity. The linkers synthesised in this thesis are designed to incorporate easily accessible building groups like simple and commercially available amino acids and hydrocarbons, using efficient chemistry such as esterification and amide coupling. The homobifunctional non-cleavable **9bii** and cleavable **10** linkers were successfully synthesised according to the designed synthetic route in Figure 2.2 to incorporate lysine specific fluorophenyl esters as the reactive groups, an alkyne affinity tag for enrichment of cross-linked peptides and a cleavable sulphonium ion for simplification of spectra. The heterobifunctional non-cleavable linker **20** was also successfully synthesised according to the designed synthetic route in Figure 2.2 to incorporate a fluorophenyl ester and a non-specific diazirine reactive group, and an azide tag for peptide enrichment. Purification of linker **20** was achieved, but difficulties were experienced in the purification of **9bii** and **10**. Future work should include method optimisation for purification of these homobifunctional structures and incorporation of different building blocks to broaden the functionalities and modularity of the linker library.

These linkers were successfully applied to the model system lysozyme, where homodimers were only visible by SDS-PAGE analysis upon cross-linker stabilisation using both modular linkers **9bii** and **10** and commercially available linker DSSO. Difficulty was experienced in diazirine activation, meaning linker **20** was not successfully applied in these studies. Proteomic analysis found that a cross-link between lysine 13 and lysine 116 were stabilising this dimer using both linker **10** and DSSO. The affinity tag on linker **10** was successfully modified with Cy3 through CuAAC chemistry, suggesting that the affinity tag could be used for enrichment to an affinity column using CuAAC. Future work may involve optimising the

diazirine activation reaction or using an alternative non-specific group in the linker design. Additionally, the CuAAC reaction carried out on the affinity tag with Cy3 could be optimised for application to affinity column enrichment of cross-linked peptides for analytical data simplification.

Chapter 3 provides insight into the complexity of snake venom systems using *B. arietans* venom as a chosen example. Higher-order protein systems within snake venom samples have received minimal attention to date, particularly using MS-based techniques. Reducing and denaturing SDS-PAGE was deemed useful in gaining an overall understanding of protein structures within a sample, but intact and native mass spectrometry can more accurately quantify the mass of the protein species. Furthermore, IMMS can be used to characterise the dynamic movement and the three-dimensional protein structures and can be used to identify isomers with identical *m/z* through ion mobility separation, which is useful in complex samples. Cross-linkers were successful in stabilising the native higher-order protein structures, allowing for their isolated identification by bottom-up proteomics. This thesis has identified never-before characterised oligomeric protein species within *B. arietans* venom, including a 120 kDa dimer of tetramers, a 60 kDa dimer of dimers and a 30 kDa dimer. These systems were all identified as being part of the CTL toxin family through bottom-up proteomic analysis.

The MS-based workflow utilised has not only revealed the complexity of *B. arietans* venom, but also provides promising potential for the characterisation of higher-order protein structures in other complex biological systems. Future work may involve confirming the identity of these protein structures by sequencing the intact protein top-down proteomic analysis. Furthermore, performing CID experiments in conjunction with IMMS may visualise dissociation of monomeric species stabilised by non-covalent interactions, confirming the presence of these protein structures. Identifying the cross-linked residues by proteomic analysis may assist in discovering surface areas of the protein involved in crucial interactions for envenomation.

Together, this thesis provides an in-depth characterisation on the structure of the protein species present in *B. arietans* venom and the design of cross-linking reagents. This information may be enhanced by performing functional studies on these protein species to provide insights into how the higher-order protein structure enhances toxin function. Long-

term, this may assist in better understanding the process of envenomation, development of antivenoms and development of other pharmaceuticals by harvesting the pharmaceutical power of these highly specific proteins.

Bibliography

(1) O'Brien, J.; Lee, S.-H.; Onogi, S.; Shea, K. J. Engineering the Protein Corona of a Synthetic Polymer Nanoparticle for Broad-Spectrum Sequestration and Neutralization of Venomous Biomacromolecules. *J. Am. Chem. Soc.* **2016**, *138* (51), 16604–16607. https://doi.org/10.1021/jacs.6b10950.

(2) Paixão-Cavalcante, D.; Kuniyoshi, A. K.; Portaro, F. C. V.; da Silva, W. D.; Tambourgi, D. V. African Adders: Partial Characterization of Snake Venoms from Three Bitis Species of Medical Importance and Their Neutralization by Experimental Equine Antivenoms. *PLoS Negl. Trop. Dis.* **2015**, *9* (2), e0003419. https://doi.org/10.1371/journal.pntd.0003419.

(3) Andrade-Silva, D.; Zelanis, A.; Kitano, E. S.; Junqueira-de-Azevedo, I. L. M.; Reis, M. S.; Lopes, A. S.; Serrano, S. M. T. Proteomic and Glycoproteomic Profilings Reveal That Post-Translational Modifications of Toxins Contribute to Venom Phenotype in Snakes. *J Proteome Res* **2016**, *15* (8), 2658–2675. https://doi.org/10.1021/acs.jproteome.6b00217.

(4) Boldrini-França, J.; Cologna, C. T.; Pucca, M. B.; Bordon, K. de C. F.; Amorim, F. G.; Anjolette, F. A. P.; Cordeiro, F. A.; Wiezel, G. A.; Cerni, F. A.; Pinheiro-Junior, E. L.; Shibao, P. Y. T.; Ferreira, I. G.; de Oliveira, I. S.; Cardoso, I. A.; Arantes, E. C. Minor Snake Venom Proteins: Structure, Function and Potential Applications. *BBA-Gen. Subj.* **2017**, *1861* (4), 824–838. https://doi.org/10.1016/j.bbagen.2016.12.022.

(5) Doley, R.; Kini, R. M. Protein Complexes in Snake Venom. *Cell. Mol. Life Sci.* **2009**, *66* (17), 2851–2871. https://doi.org/10.1007/s00018-009-0050-2.

(6) Snijder, J.; Heck, A. J. R. Analytical Approaches for Size and Mass Analysis of Large Protein Assemblies. *Annu. Rev. Anal. Chem.* **2014**, *7* (1), 43–64. https://doi.org/10.1146/annurev-anchem-071213-020015.

(7) Akef, H. M. Snake Venom: Kill and Cure. *Toxin Rev* **2019**, *38* (1), 21–40. https://doi.org/10.1080/15569543.2017.1399278.

(8) Vulfius, C. A.; Spirova, E. N.; Serebryakova, M. V.; Shelukhina, I. V.; Kudryavtsev, D. S.; Kryukova, E. V.; Starkov, V. G.; Kopylova, N. V.; Zhmak, M. N.; Ivanov, I. A.; Kudryashova, K. S.; Andreeva, T. V.; Tsetlin, V. I.; Utkin, Y. N. Peptides from Puff Adder Bitis Arietans Venom, Novel Inhibitors of Nicotinic Acetylcholine Receptors. *Toxicon* **2016**, *121*, 70–76. https://doi.org/10.1016/j.toxicon.2016.08.020.

(9) Chan, Y. S.; Cheung, R. C. F.; Xia, L.; Wong, J. H.; Ng, T. B.; Chan, W. Y. Snake Venom Toxins: Toxicity and Medicinal Applications. *Appl. Microbiol. Biotechnol.* **2016**, *100* (14), 6165–6181. https://doi.org/10.1007/s00253-016-7610-9.

(10) Kang, T. S.; Georgieva, D.; Genov, N.; Murakami, M. T.; Sinha, M.; Kumar, R. P.; Kaur, P.; Kumar, S.; Dey, S.; Sharma, S.; Vrielink, A.; Betzel, C.; Takeda, S.; Arni, R. K.; Singh, T. P.; Kini, R. M. Enzymatic Toxins from Snake Venom: Structural Characterization and Mechanism of Catalysis: Enzymatic Toxins from Snake Venom. *FEBS J.* **2011**, *278* (23), 4544–4576. https://doi.org/10.1111/j.1742-4658.2011.08115.x.

(11) Tasoulis, T.; Isbister, G. A Review and Database of Snake Venom Proteomes. *Toxins* **2017**, *9* (9), 290–313. https://doi.org/10.3390/toxins9090290.

(12) Utkin, Y. N. Animal Venom Studies: Current Benefits and Future Developments. *World J. Biol. Chem.* **2015**, *6* (2), 28–33. https://doi.org/10.4331/wjbc.v6.i2.28.

(13) Xiong, S.; Huang, C. Synergistic Strategies of Predominant Toxins in Snake Venoms. *Toxicol. Lett.* **2018**, *287*, 142–154. https://doi.org/10.1016/j.toxlet.2018.02.004.
(14) McCleary, R. J. R.; Kini, R. M. Non-Enzymatic Proteins from Snake Venoms: A Gold Mine of Pharmacological Tools and Drug Leads. *Toxicon* **2013**, *62*, 56–74. https://doi.org/10.1016/j.toxicon.2012.09.008.
(15) Kularatne, S. A. M.; Senanayake, N. Venomous Snake Bites, Scorpions, and Spiders. In *Handbook of Clinical Neurology*; Elsevier, 2014; Vol. 120, pp 987–1001. https://doi.org/10.1016/B978-0-7020-4087-0.00066-8.
(16) Lin, C.-W.; Chen, J.-M.; Wang, Y.-M.; Wu, S.-W.; Tsai, I.-H.; Khoo, K.-H. Terminal Disialylated Multiantennary Complex-Type N-Glycans Carried on Acutobin Define the Glycosylation Characteristics of the Deinagkistrodon Acutus Venom. *Glycobiology* **2011**, *21* (4), 530–542. https://doi.org/10.1093/glycob/cwq195.
(17) Harrison, J. A.; Aquilina, J. A. Insights into the Subunit Arrangement and Diversity of Paradoxin and Taipoxin. *Toxicon* **2016**, *112*, 45–50. https://doi.org/10.1016/j.toxicon.2016.01.054.
(18) Opie, L. H.; Kowolik, H. The Discovery of Captopril: From Large Animals to Small Molecules. *Cardiovasc. Res.* **1995**, *30* (1), 18–25. https://doi.org/10.1016/s0008-6363(95)00006-2.
(19) Gasanov, S. E.; Dagda, R. K.; Rael, E. D. Snake Venom Cytotoxins, Phospholipase A2s, and Zn2+-Dependent Metalloproteinases: Mechanisms of Action and Pharmacological Relevance. *J. Clin. Toxicol.* **2014**, *4* (1), 1000181. https://doi.org/10.4172/2161-0495.1000181.
(20) Harvey, A. L. Toxins and Drug Discovery. *Toxicon* **2014**, *92*, 193–200. https://doi.org/10.1016/j.toxicon.2014.10.020.
(21) Singh, D. B.; Tripathi, T. *Frontiers in Protein Structure, Function, and Dynamics*; Springer: Singapore, 2020.
(22) Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E. Mass Spectrometry Identifiable Cross-Linking Strategy for Studying Protein−Protein Interactions. *Anal. Chem.* **2005**, *77* (1), 311–318. https://doi.org/10.1021/ac0488762.
(23) Laustsen, A. H. Toxin Synergism in Snake Venoms. *Toxin Rev.* **2016**, *35* (3–4), 165–170. https://doi.org/10.1080/15569543.2016.1220397.
(24) Matsui, T.; Hamako, J.; Titani, K. Structure and Function of Snake Venom Proteins Affecting Platelet Plug Formation. *Toxins* **2009**, *2* (1), 10–23. https://doi.org/10.3390/toxins2010010.
(25) Oswald, R. E.; Sutcliffe, M. J.; Bamberger, M.; Loring, R. H.; Braswell, E.; Dobson, C. M. Solution Structure of Neuronal Bungarotoxin Determined by Two-Dimensional NMR Spectroscopy: Sequence-Specific Assignments, Secondary Structure, and Dimer Formation. *Biochemistry* **1991**, *30* (20), 4901–4909. https://doi.org/10.1021/bi00234a010.
(26) Loo, J. A. Studying Noncovalent Protein Complexes by Electrospray Ionization Mass Spectrometry. *Mass Spectrom. Rev.* **1997**, *16* (1), 1–23. https://doi.org/10.1002/(SICI)1098-2787(1997)16:1.
(27) Osipov, A. V.; Rucktooa, P.; Kasheverov, I. E.; Filkin, S. Yu.; Starkov, V. G.; Andreeva, T. V.; Sixma, T. K.; Bertrand, D.; Utkin, Y. N.; Tsetlin, V. I. Dimeric α-Cobratoxin X-Ray Structure: Localization of Intermolecular Disulfides and Possible Mode of Binding to Nicotinic Acetylcholine Receptors. *J. Biol. Chem.* **2012**, *287* (9), 6725–6734. https://doi.org/10.1074/jbc.M111.322313.

(28) Ben-Nissan, G.; Sharon, M. The Application of Ion-Mobility Mass Spectrometry for Structure/Function Investigation of Protein Complexes. *Curr. Opin. Chem. Biol.* **2018**, *42*, 25–33. https://doi.org/10.1016/j.cbpa.2017.10.026.
(29) Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. R. Ion Mobility Mass Spectrometry of Proteins and Proteinassemblies. *Chem Soc Rev* **2010**, *39* (5), 1633–1655. https://doi.org/10.1039/B914002F.
(30) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V. Ion Mobility–Mass Spectrometry Analysis of Large Protein Complexes. *Nat. Protoc.* **2008**, *3* (7), 1139–1152. https://doi.org/10.1038/nprot.2008.78.
(31) Hart, A. J.; Isbister, G. K.; O'Donnell, P.; Williamson, N. A.; Hodgson, W. C. Species Differences in the Neuromuscular Activity of Post-Synaptic Neurotoxins from Two Australian Black Snakes (Pseudechis Porphyriacus and Pseudechis Colletti). *Toxicol. Lett.* **2013**, *219* (3), 262–268. https://doi.org/10.1016/j.toxlet.2013.03.026.
(32) Juárez, P.; Wagstaff, S. C.; Oliver, J.; Sanz, L.; Harrison, R. A.; Calvete, J. J. Molecular Cloning of Disintegrin-like Transcript BA-5A from a Bitis Arietans Venom Gland CDNA Library: A Putative Intermediate in the Evolution of the Long-Chain Disintegrin Bitistatin. *J. Mol. Evol.* **2006**, *63* (1), 142–152. https://doi.org/10.1007/s00239-005-0268-z.
(33) Fasoli, E.; Sanz, L.; Wagstaff, S.; Harrison, R. A.; Righetti, P. G.; Calvete, J. J. Exploring the Venom Proteome of the African Puff Adder, Bitis Arietans, Using a Combinatorial Peptide Ligand Library Approach at Different PHs. *J. Proteomics* **2010**, *73* (5), 932–942. https://doi.org/10.1016/j.jprot.2009.12.006.
(34) Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003**, *422* (6928), 198–207. https://doi.org/10.1038/nature01511.
(35) Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246* (4926), 64–71. https://doi.org/10.1126/science.2675315.
(36) Ho, C. S.; Lam, C. W. K.; Chan, M. H. M.; Cheung, R. C. K.; Law, L. K.; Lit, L. C. W.; Ng, K. F.; Suen, M. W. M.; Tai, H. L. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin. Biochem. Rev.* **2003**, *24* (1), 3–12.
(37) Banerjee, S.; Mazumdar, S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int J Anal Chem* **2012**, *2012*, 1–40. https://doi.org/10.1155/2012/282574.
(38) Dyson, P. J.; Khalaila, I.; Luettgen, S.; McIndoe, J. S.; Zhao, D. Direct Probe Electrospray (and Nanospray) Ionization Mass Spectrometry of Neat Ionic Liquids. *Chem. Commun.* **2004**, No. 19, 2204–2205. https://doi.org/10.1039/b407217k.
(39) Gross, J. H. *Mass Spectrometry: A Textbook*, 3rd ed.; Springer International Publishing: Cham, 2017. https://doi.org/10.1007/978-3-319-54398-7.
(40) Douglas, D. J. Linear Quadrupoles in Mass Spectrometry. *Mass Spectrom. Rev.* **2009**, *28* (6), 937–960. https://doi.org/10.1002/mas.20249.
(41) Hua, Y.; Wainhaus, S. B.; Yang, Y.; Shen, L.; Xiong, Y.; Xu, X.; Zhang, F.; Bolton, J. L.; Breemen, R. B. Comparison of Negative and Positive Ion Electrospray Tandem Mass Spectrometry for the Liquid Chromatography Tandem Mass Spectrometry Analysis of Oxidized Deoxynucleosides. *J. Am. Soc. Mass Spectrom.* **2001**, *12* (1), 80–87. https://doi.org/10.1016/S1044-0305(00)00191-4.

(42) Zinnel, N. F.; Russell, D. H. Size-to-Charge Dispersion of Collision-Induced Dissociation Product Ions for Enhancement of Structural Information and Product Ion Identification. *Anal. Chem.* **2014**, *86* (10), 4791–4798. https://doi.org/10.1021/ac403929u.
(43) Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G. Mass Spectrometry of Peptides and Proteins. *Methods* **2005**, *35* (3), 211–222. https://doi.org/10.1016/j.ymeth.2004.08.013.
(44) Macias, L. A.; Santos, I. C.; Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2020**, *92* (1), 227–251. https://doi.org/10.1021/acs.analchem.9b04859.
(45) Medzihradszky, K. F. Peptide Sequence Analysis. In *Methods in Enzymology*; Elsevier, 2005; Vol. 402, pp 209–244. https://doi.org/10.1016/S0076-6879(05)02007-0.
(46) Macklin, A.; Khan, S.; Kislinger, T. Recent Advances in Mass Spectrometry Based Clinical Proteomics: Applications to Cancer Research. *Clin. Proteomics* **2020**, *17* (1), 17. https://doi.org/10.1186/s12014-020-09283-w.
(47) *Mass Spectrometry Data Analysis in Proteomics*, Third edition.; Matthiesen, R., Ed.; Methods in molecular biology; Humana Press: New York, NY, 2020.
(48) Han, X.; Aslanian, A.; Yates, J. R. Mass Spectrometry for Proteomics. *Curr. Opin. Chem. Biol.* **2008**, *12* (5), 483–490. https://doi.org/10.1016/j.cbpa.2008.07.024.
(49) Chait, B. T. Mass Spectrometry: Bottom-Up or Top-Down? *Science* **2006**, *314* (5796), 65–66. https://doi.org/10.1126/science.1133987.
(50) Bohrer, B. C.; Merenbloom, S. I.; Koeniger, S. L.; Hilderbrand, A. E.; Clemmer, D. E. Biomolecule Analysis by Ion Mobility Spectrometry. *Annu. Rev. Anal. Chem.* **2008**, *1* (1), 293–327. https://doi.org/10.1146/annurev.anchem.1.031207.113001.
(51) Jurneczko, E.; Kalapothakis, J.; Campuzano, I. D. G.; Morris, M.; Barran, P. E. Effects of Drift Gas on Collision Cross Sections of a Protein Standard in Linear Drift Tube and Traveling Wave Ion Mobility Mass Spectrometry. *Anal. Chem.* **2012**, *84* (20), 8524–8531. https://doi.org/10.1021/ac301260d.
(52) May, J. C.; McLean, J. A. Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Anal. Chem.* **2015**, *87* (3), 1422–1436. https://doi.org/10.1021/ac504720m.
(53) Pukala, T. Importance of Collision Cross Section Measurements by Ion Mobility Mass Spectrometry in Structural Biology. *Rapid Commun. Mass Spectrom.* **2019**, *33* (S3), 72–82. https://doi.org/10.1002/rcm.8294.
(54) Agilent 6200 Series TOF and 6500 Series Q-TOF-LC/MS System: Concepts Guide. Agilent Technologies: California 2014.
(55) Ruwan Kurulugama; Ken Imantani; Lester Taylor. The Agilent Ion Mobility Q-TOF Mass Spectrometer System: Technical Overview. Agilent Technologies: California 2013.
(56) Yang, Y.; Song, H.; He, D.; Zhang, S.; Dai, S.; Lin, S.; Meng, R.; Wang, C.; Chen, P. R. Genetically Encoded Protein Photocrosslinker with a Transferable Mass Spectrometry-Identifiable Label. *Nat. Commun.* **2016**, *7* (1), 12299. https://doi.org/10.1038/ncomms12299.
(57) Bond, M. R.; Zhang, H.; Vu, P. D.; Kohler, J. J. Photocrosslinking of Glycoconjugates Using Metabolically Incorporated Diazirine-Containing Sugars. *Nat. Protoc.* **2009**, *4* (7), 1044–1063. https://doi.org/10.1038/nprot.2009.85.
(58) Sinz, A. The Advancement of Chemical Cross-Linking and Mass Spectrometry for Structural Proteomics: From Single Proteins to Protein Interaction Networks. *Expert*

Rev. Proteomics **2014**, *11* (6), 733–743. https://doi.org/10.1586/14789450.2014.960852.
(59) Iacobucci, C.; Götze, M.; Ihling, C. H.; Piotrowski, C.; Arlt, C.; Schäfer, M.; Hage, C.; Schmidt, R.; Sinz, A. A Cross-Linking/Mass Spectrometry Workflow Based on MS-Cleavable Cross-Linkers and the MeroX Software for Studying Protein Structures and Protein–Protein Interactions. *Nat. Protoc.* **2018**, *13* (12), 2864–2889. https://doi.org/10.1038/s41596-018-0068-8.
(60) Politis, A.; Park, A. Y.; Hyung, S.-J.; Barsky, D.; Ruotolo, B. T.; Robinson, C. V. Integrating Ion Mobility Mass Spectrometry with Molecular Modelling to Determine the Architecture of Multiprotein Complexes. *PLoS ONE* **2010**, *5* (8), e12080. https://doi.org/10.1371/journal.pone.0012080.
(61) Götze, M.; Iacobucci, C.; Ihling, C. H.; Sinz, A. A Simple Cross-Linking/Mass Spectrometry Workflow for Studying System-Wide Protein Interactions. *Anal. Chem.* **2019**, *91* (15), 10236–10244. https://doi.org/10.1021/acs.analchem.9b02372.
(62) Burke, A. M.; Kandur, W.; Novitsky, E. J.; Kaake, R. M.; Yu, C.; Kao, A.; Vellucci, D.; Huang, L.; Rychnovsky, S. D. Synthesis of Two New Enrichable and MS-Cleavable Cross-Linkers to Define Protein–Protein Interactions by Mass Spectrometry. *Org. Biomol. Chem.* **2015**, *13* (17), 5030–5037. https://doi.org/10.1039/C5OB00488H.
(63) Schmidt, C.; Zhou, M.; Marriott, H.; Morgner, N.; Politis, A.; Robinson, C. V. Comparative Cross-Linking and Mass Spectrometry of an Intact F-Type ATPase Suggest a Role for Phosphorylation. *Nat. Commun.* **2013**, *4* (1), 1985. https://doi.org/10.1038/ncomms2985.
(64) 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. https://doi.org/10.1038/s41594-018-0147-0.
(65) Sinz, A. The Advancement of Chemical Cross-Linking and Mass Spectrometry for Structural Proteomics: From Single Proteins to Protein Interaction Networks. *Expert Rev. Proteomics* **2014**, *11* (6), 733–743. https://doi.org/10.1586/14789450.2014.960852.
(66) Chen, Z. A.; Jawhari, A.; Fischer, L.; Buchen, C.; Tahir, S.; Kamenski, T.; Rasmussen, M.; Lariviere, L.; Bukowski-Wills, J.-C.; Nilges, M.; Cramer, P.; Rappsilber, J. Architecture of the RNA Polymerase II–TFIIF Complex Revealed by Cross-Linking and Mass Spectrometry. *EMBO J.* **2010**, *29* (4), 717–726. https://doi.org/10.1038/emboj.2009.401.
(67) 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. https://doi.org/10.1021/acscentsci.9b00416.
(68) Vellucci, D.; Kao, A.; Kaake, R. M.; Rychnovsky, S. D.; Huang, L. Selective Enrichment and Identification of Azide-Tagged Cross-Linked Peptides Using Chemical Ligation and Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (8), 1432–1445. https://doi.org/10.1016/j.jasms.2010.04.004.
(69) Sibbersen, C.; Lykke, L.; Gregersen, N.; Jørgensen, K. A.; Johannsen, M. A Cleavable Azide Resin for Direct Click Chemistry Mediated Enrichment of Alkyne-Labeled Proteins. *Chem Commun* **2014**, *50* (81), 12098–12100. https://doi.org/10.1039/C4CC05246C.

(70) Petrotchenko, E. V.; Serpa, J. J.; Borchers, C. H. An Isotopically Coded CID-Cleavable Biotinylated Cross-Linker for Structural Proteomics. *Mol. Cell. Proteomics* **2011**, *10* (2), M110.001420. https://doi.org/10.1074/mcp.M110.001420.
(71) Müller, M. Q.; Dreiocker, F.; Ihling, C. H.; Schäfer, M.; Sinz, A. Fragmentation Behavior of a Thiourea-Based Reagent for Protein Structure Analysis by Collision-Induced Dissociative Chemical Cross-Linking. *J. Mass Spectrom.* **2010**, *45* (8), 880–891. https://doi.org/10.1002/jms.1775.
(72) Kao, A.; Chiu, C.; Vellucci, D.; Yang, Y.; Patel, V. R.; Guan, S.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L. Development of a Novel Cross-Linking Strategy for Fast and Accurate Identification of Cross-Linked Peptides of Protein Complexes. *Mol. Cell. Proteomics* **2011**, *10* (1), M110.002212. https://doi.org/10.1074/mcp.M110.002212.
(73) Leitner, A.; Faini, M.; Stengel, F.; Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem. Sci.* **2016**, *41* (1), 20–32. https://doi.org/10.1016/j.tibs.2015.10.008.
(74) Tran, B. Q.; Goodlett, D. R.; Goo, Y. A. Advances in Protein Complex Analysis by Chemical Cross-Linking Coupled with Mass Spectrometry (CXMS) and Bioinformatics. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2016**, *1864* (1), 123–129. https://doi.org/10.1016/j.bbapap.2015.05.015.
(75) Lockett, M. R.; Phillips, M. F.; Jarecki, J. L.; Peelen, D.; Smith, L. M. A Tetrafluorophenyl Activated Ester Self-Assembled Monolayer for the Immobilization of Amine-Modified Oligonucleotides. *Langmuir* **2008**, *24* (1), 69–75. https://doi.org/10.1021/la702493u.
(76) Burke, A. M.; Kandur, W.; Novitsky, E. J.; Kaake, R. M.; Yu, C.; Kao, A.; Vellucci, D.; Huang, L.; Rychnovsky, S. D. Synthesis of Two New Enrichable and MS-Cleavable Cross-Linkers to Define Protein–Protein Interactions by Mass Spectrometry. *Org. Biomol. Chem.* **2015**, *13* (17), 5030–5037. https://doi.org/10.1039/C5OB00488H.
(77) Sinz, A. Chemical Cross-Linking and Mass Spectrometry to Map Three-Dimensional Protein Structures and Protein–Protein Interactions. *Mass Spectrom. Rev.* **2006**, *25* (4), 663–682. https://doi.org/10.1002/mas.20082.
(78) Rilatt, I. Preparation of Humanized Anti-Human IGF-1R, HER2 and Protein Axl Antibodies Conjugated with Drug for Treatment of Cancer. World Intellectual Property Organisation 2015.
(79) Katsuhiko, H.; Takao, Y.; Kazuhiko, I.; Hirofumi, H. Chemistry of Succinimido Esters. IV. A Facile Preparation of N-Succinimidyl Carboxylates Using N,N'-Disuccinimidyl Carbonate. *Yukagaku* **1987**, *36* (1), 16–20.
(80) Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R. Probing Native Protein Structures by Chemical Cross-Linking, Mass Spectrometry, and Bioinformatics. *Mol. Cell. Proteomics* **2010**, *9* (8), 1634–1649. https://doi.org/10.1074/mcp.R000001-MCP201.
(81) Vellucci, D.; Kao, A.; Kaake, R. M.; Rychnovsky, S. D.; Huang, L. Selective Enrichment and Identification of Azide-Tagged Cross-Linked Peptides Using Chemical Ligation and Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (8), 1432–1445. https://doi.org/10.1016/j.jasms.2010.04.004.
(82) He, Y.-F.; Bao, H.-M.; Xiao, X.-F.; Zuo, S.; Du, R.-Y.; Tang, S.-W.; Yang, P.-Y.; Chen, X. Biotin Tagging Coupled with Amino Acid-Coded Mass Tagging for Efficient and Precise Screening of Interaction Proteome in Mammalian Cells. *Proteomics* **2009**, *9* (24), 5414–5424. https://doi.org/10.1002/pmic.200800864.

(83) Schmidt, R.; Sinz, A. Improved Single-Step Enrichment Methods of Cross-Linked Products for Protein Structure Analysis and Protein Interaction Mapping. *Anal. Bioanal. Chem.* **2017**, *409* (9), 2393–2400. https://doi.org/10.1007/s00216-017-0185-1.
(84) Back, J. W.; de Jong, L.; Muijsers, A. O.; de Koster, C. G. Chemical Cross-Linking and Mass Spectrometry for Protein Structural Modeling. *J. Mol. Biol.* **2003**, *331* (2), 303–313. https://doi.org/10.1016/S0022-2836(03)00721-6.
(85) Lu, Y.; Tanasova, M.; Borhan, B.; Reid, G. E. Ionic Reagent for Controlling the Gas-Phase Fragmentation Reactions of Cross-Linked Peptides. *Anal. Chem.* **2008**, *80* (23), 9279–9287. https://doi.org/10.1021/ac801625e.
(86) MacKinnon, A. L.; Taunton, J. Target Identification by Diazirine Photo-Cross-Linking and Click Chemistry. *Curr. Protoc. Chem. Biol.* **2009**, *1* (1), 55–73. https://doi.org/10.1002/9780470559277.ch090167.
(87) Chowdhury, S. M.; Du, X.; Tolić, N.; Wu, S.; Moore, R. J.; Mayer, M. U.; Smith, R. D.; Adkins, J. N. Identification of Cross-Linked Peptides after Click-Based Enrichment Using Sequential Collision-Induced Dissociation and Electron Transfer Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2009**, *81* (13), 5524–5532. https://doi.org/10.1021/ac900853k.
(88) Yu, C.; Huang, L. Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal. Chem.* **2018**, *90* (1), 144–165. https://doi.org/10.1021/acs.analchem.7b04431.
(89) Downey, K. M. Design and Synthesis of Protein Chemical Crosslinkers: A Modular Approach, The University of Adelaide, 2017.
(90) C. Davis, M. Tricarbamate of 1,3,5-Triaminobenzene via Curtius Rearrangement of Trimesic Acid and Subsequent Nitration. *Synth. Commun.* **2007**, *37* (9), 1457–1462. https://doi.org/10.1080/00397910701227044.
(91) Linxis, B. V.; Merkul, E.; Sijbrandi, N. J.; Muns, J. A.; Van Dongen, A. A. M. S.; Steverink, P. J. G. M.; Houthoff, H. J. Platinum-Based Functional Moieties for Preparing Cell Targeting Conjugates. A61K47/68 (2017.01).
(92) Petrou, C.; Sarigiannis, Y. Peptide Synthesis. In *Peptide Applications in Biomedicine, Biotechnology and Bioengineering*; Elsevier, 2018; pp 1–21. https://doi.org/10.1016/B978-0-08-100736-5.00001-6.
(93) Bonke, G.; Vedel, L.; Witt, M.; Jaroszewski, J.; Olsen, C.; Franzyk, H. Dimeric Building Blocks for Solid-Phase Synthesis of α-Peptide-β-Peptoid Chimeras. *Synthesis* **2008**, No. 15, 2381–2390. https://doi.org/10.1055/s-2008-1067171.
(94) Merck. Fmoc Chloride. Sigma Aldrich.
(95) Augustine, J. K.; Vairaperumal, V.; Narasimhan, S.; Alagarsamy, P.; Radhakrishnan, A. Propylphosphonic Anhydride (T3P®): An Efficient Reagent for the One-Pot Synthesis of 1,2,4-Oxadiazoles, 1,3,4-Oxadiazoles, and 1,3,4-Thiadiazoles. *Tetrahedron* **2009**, *65* (48), 9989–9996. https://doi.org/10.1016/j.tet.2009.09.114.
(96) Morphological Transition Triggered by Mannose Conjugation to a Cyclic Hexapeptide. *Arkivoc* **2012**, *2013* (2), 82. https://doi.org/10.3998/ark.5550190.0014.208.
(97) Nguyen, H. P.; Stewart, S.; Kukwikila, M. N.; Jones, S. F.; Offenbartl-Stiegert, D.; Mao, S.; Balasubramanian, S.; Beck, S.; Howorka, S. A Photo-responsive Small-Molecule Approach for the Opto-epigenetic Modulation of DNA Methylation. *Angew. Chem. Int. Ed.* **2019**, *58* (20), 6620–6624. https://doi.org/10.1002/anie.201901139.

(98) Muller, N.; Carr, D. T. CARBON-13 SPLITTINGS IN FLUORINE NUCLEAR MAGNETIC RESONANCE SPECTRA ^{1}. *J. Phys. Chem.* **1963**, *67* (1), 112–115. https://doi.org/10.1021/j100795a026.
(99) McLaughlin, C.; Slawin, A. M. Z.; Smith, A. D. Base-free Enantioselective C(1)-Ammonium Enolate Catalysis Exploiting Aryloxides: A Synthetic and Mechanistic Study. *Angew. Chem. Int. Ed.* **2019**, *58* (42), 15111–15119. https://doi.org/10.1002/anie.201908627.
(100) Jain, D. R.; Ganesh, K. N. Clickable C^{γ}-Azido(Methylene/Butylene) Peptide Nucleic Acids and Their Clicked Fluorescent Derivatives: Synthesis, DNA Hybridization Properties, and Cell Penetration Studies. *J. Org. Chem.* **2014**, *79* (14), 6708–6714. https://doi.org/10.1021/jo500834u.
(101) Mädler, S.; Bich, C.; Touboul, D.; Zenobi, R. Chemical Cross-Linking with NHS Esters: A Systematic Study on Amino Acid Reactivities. *J. Mass Spectrom.* **2009**, *44* (5), 694–706. https://doi.org/10.1002/jms.1544.
(102) Masuda, T.; Ide, N.; Kitabatake, N. Structure–Sweetness Relationship in Egg White Lysozyme: Role of Lysine and Arginine Residues on the Elicitation of Lysozyme Sweetness. *Chem. Senses* **2005**, *30* (8), 667–681. https://doi.org/10.1093/chemse/bji060.
(103) Hampe, O. G. Conformation of Lysozyme in Aqueous Solution. Effect of Ionic Strength and Protein Concentration. *Eur. J. Biochem.* **1972**, *31* (1), 32–37.
(104) Thomas, B.; Vekilov, P.; Rosenberger, F. Heterogeneity Determination and Purification of Commercial Hen Egg-White Lysozyme. *Acta Crystallogr.* **1996**, *D52*, 776–784.
(105) Onuma, K.; Inaka, K. Lysozyme Dimer Association: Similarities and Differences Compared with Lysozyme Monomer Association. *J. Cryst. Growth* **2008**, *310* (6), 1174–1181. https://doi.org/10.1016/j.jcrysgro.2007.12.029.
(106) Maroufi, B.; Ranjbar, B.; Khajeh, K.; Naderi-Manesh, H.; Yaghoubi, H. Structural Studies of Hen Egg-White Lysozyme Dimer: Comparison with Monomer. *Biochim. Biophys. Acta* **2008**, *1784* (7–8), 1043–1049. https://doi.org/10.1016/j.bbapap.2008.03.010.
(107) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242. https://doi.org/10.1093/nar/28.1.235.
(108) Heck, A. J. R. XlinkX for Proteome Discoverer https://www.hecklab.com/software/xlinkx/ (accessed 2020 -11 -20).
(109) Suaifan, G. A. R. Y.; Arafat, T.; Threadgill, M. D. Synthetic Approaches to Peptides Containing the L-Gln-l-Val-D(S)-Dmt Motif. *Bioorg. Med. Chem.* **2007**, *15* (10), 3474–3488. https://doi.org/10.1016/j.bmc.2007.03.005.
(110) Makris, G.; Papagiannopoulou, D. Synthesis, Characterization, and Biological Evaluation of New Biotinylated Tc/Re-Tricarbonyl Complexes. *J. Label. Compd. Radiopharm.* **2016**, *59* (3), 95–102. https://doi.org/10.1002/jlcr.3372.
(111) Henze, M.; Kreye, O.; Brauch, S.; Nitsche, C.; Naumann, K.; Wessjohann, L.; Westermann, B. Photoaffinity-Labeled Peptoids and Depsipeptides by Multicomponent Reactions. *Synthesis* **2010**, No. 17, 2997–3003. https://doi.org/10.1055/s-0030-1258182.
(112) Laing, G. D.; Renjifo, J. M.; Ruiz, F.; Harrison, R. A.; Nasidi, A.; Gutierrez, J.-M.; Rowley, P. D.; Warrell, D. A.; Theakston, R. D. G. A New Pan African Polyspecific

Antivenom Developed in Response to the Antivenom Crisis in Africa. *Toxicon* **2003**, *42* (1), 35–41. https://doi.org/10.1016/S0041-0101(03)00098-9.
(113) Currier, R. B.; Harrison, R. A.; Rowley, P. D.; Laing, G. D.; Wagstaff, S. C. Intra-Specific Variation in Venom of the African Puff Adder (Bitis Arietans): Differential Expression and Activity of Snake Venom Metalloproteinases (SVMPs). *Toxicon* **2010**, *55* (4), 864–873. https://doi.org/10.1016/j.toxicon.2009.12.009.
(114) Hamako, J.; Matsui, T.; Suzuki, M.; Ito, M.; Makita, K.; Fujimura, Y.; Ozeki, Y.; Titani, K. Purification and Characterization of Bitiscetin, a Novel von Willebrand Factor Modulator Protein from Bitis ArietansSnake Venom. *Biochem. Biophys. Res. Commun.* **1996**, *226* (1), 273–279. https://doi.org/10.1006/bbrc.1996.1345.
(115) Munekiyo, S. M.; Mackessy, S. P. Effects of Temperature and Storage Conditions on the Electrophoretic, Toxic and Enzymatic Stability of Venom Components. *Comp Biochem Physiol* **1998**, *119B* (1), 119–127.
(116) Hong, P.; Koza, S.; Bouvier, E. S. P. A Review Size-Exclusion Chromatography for the Anaysis of Protein Biotherapeutics and Their Aggregates. *J. Liq. Chromatogr. Relat. Technol.* **2012**, *35* (20), 2923–2950. https://doi.org/10.1080/10826076.2012.743724.
(117) Aquilina, J. A. The Major Toxin from the Australian Common Brown Snake Is a Hexamer with Unusual Gas-Phase Dissociation Properties. *Proteins Struct. Funct. Bioinforma.* **2009**, *75* (2), 478–485. https://doi.org/10.1002/prot.22259.
(118) Liu, H.; Sadygov, R. G.; Yates, J. R. A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Anal. Chem.* **2004**, *76* (14), 4193–4201. https://doi.org/10.1021/ac0498563.
(119) Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H. Quantification of Proteins and Metabolites by Mass Spectrometry without Isotopic Labeling or Spiked Standards. *Anal. Chem.* **2003**, *75* (18), 4818–4826. https://doi.org/10.1021/ac026468x.
(120) Bondarenko, P. V.; Chelius, D.; Shaler, T. A. Identification and Relative Quantitation of Protein Mixtures by Enzymatic Digestion Followed by Capillary Reversed-Phase Liquid Chromatography−Tandem Mass Spectrometry. *Anal. Chem.* **2002**, *74* (18), 4741–4749. https://doi.org/10.1021/ac0256991.
(121) Chelius, D.; Bondarenko, P. V. Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry. *J. Proteome Res.* **2002**, *1* (4), 317–323. https://doi.org/10.1021/pr025517j.
(122) Lu, B.; Xu, T.; Park, S. K.; Yates, J. R. Shotgun Protein Identification and Quantification by Mass Spectrometry. In *Proteomics*; Reinders, J., Sickmann, A., Eds.; Methods in Molecular Biology™; Humana Press: New Jersey, 2009; Vol. 564, pp 261–288. https://doi.org/10.1007/978-1-60761-157-8_15.
(123) Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of Label-Free Methods for Quantifying Human Proteins by Shotgun Proteomics. *Mol. Cell. Proteomics* **2005**, *4* (10), 1487–1502. https://doi.org/10.1074/mcp.M500084-MCP200.
(124) Pang, J. X.; Ginanni, N.; Dongre, A. R.; Hefta, S. A.; Opiteck, G. J. Biomarker Discovery in Urine by Proteomics. *J. Proteome Res.* **2002**, *1* (2), 161–169. https://doi.org/10.1021/pr015518w.
(125) Gao, Y.; Wang, Y. A Method to Determine the Ionization Efficiency Change of Peptides Caused by Phosphorylation. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (11), 1973–1976. https://doi.org/10.1016/j.jasms.2007.08.010.

(126) Washburn, M. P.; Wolters, D.; Yates, J. R. Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology. *Nat. Biotechnol.* **2001**, *19* (3), 242–247. https://doi.org/10.1038/85686.
(127) Griss, J.; Stanek, F.; Hudecz, O.; Dürnberger, G.; Perez-Riverol, Y.; Vizcaíno, J. A.; Mechtler, K. Spectral Clustering Improves Label-Free Quantification of Low-Abundant Proteins. *J. Proteome Res.* **2019**, *18* (4), 1477–1485. https://doi.org/10.1021/acs.jproteome.8b00377.
(128) Choi, H.; Kim, S.; Fermin, D.; Tsou, C.-C.; Nesvizhskii, A. I. QPROT: Statistical Method for Testing Differential Expression Using Protein-Level Intensity Data in Label-Free Quantitative Proteomics. *J. Proteomics* **2015**, *129*, 121–126. https://doi.org/10.1016/j.jprot.2015.07.036.
(129) Ning, K.; Fermin, D.; Nesvizhskii, A. I. Comparative Analysis of Different Label-Free Mass Spectrometry Based Protein Abundance Estimates and Their Correlation with RNA-Seq Gene Expression Data. *J. Proteome Res.* **2012**, *11* (4), 2261–2271. https://doi.org/10.1021/pr201052x.
(130) Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B.; Lantz, C.; Jung, W.; Johnson, K. R.; Koller, A.; Wolff, J. J.; Campuzano, I. D. G.; Auclair, J. R.; Ivanov, A. R.; Whitelegge, J. P.; Paša-Tolić, L.; Chamot-Rooke, J.; Danis, P. O.; Smith, L. M.; Tsybin, Y. O.; Loo, J. A.; Ge, Y.; Kelleher, N. L.; Agar, J. N. Best Practices and Benchmarks for Intact Protein Analysis for Top-down Mass Spectrometry. *Nat. Methods* **2019**, *16* (7), 587–594. https://doi.org/10.1038/s41592-019-0457-0.
(131) Liigand, P.; Kaupmees, K.; Kruve, A. Influence of the Amino Acid Composition on the Ionization Efficiencies of Small Peptides. *J. Mass Spectrom.* **2019**, *54* (6), 481–487. https://doi.org/10.1002/jms.4348.
(132) Liigand, J.; Laaniste, A.; Kruve, A. PH Effects on Electrospray Ionization Efficiency. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (3), 461–469. https://doi.org/10.1007/s13361-016-1563-1.
(133) MassHunter IM-MS Browser Software Guide. Agilent Technologies 2015.
(134) Lee, J.-Y.; Bilbao, A.; Conant, C. R.; Bloodsworth, K. J.; Orton, D. J.; Zhou, M.; Wilson, J. W.; Zheng, X.; Webb, I. K.; Li, A.; Hixson, K. K.; Fjeldsted, J. C.; Ibrahim, Y. M.; Payne, S. H.; Jansson, C.; Smith, R. D.; Metz, T. O. AutoCCS: Automated Collision Cross-Section Calculation Software for Ion Mobility Spectrometry–Mass Spectrometry. *Bioinformatics* **2021**, btab429. https://doi.org/10.1093/bioinformatics/btab429.
(135) Harrison, J. A.; Kelso, C.; Pukala, T. L.; Beck, J. L. Conditions for Analysis of Native Protein Structures Using Uniform Field Drift Tube Ion Mobility Mass Spectrometry and Characterization of Stable Calibrants for TWIM-MS. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (2), 256–267. https://doi.org/10.1007/s13361-018-2074-z.
(136) Chen, S.-H.; Russell, D. H. How Closely Related Are Conformations of Protein Ions Sampled by IM-MS to Native Solution Structures? *J. Am. Soc. Mass Spectrom.* **2015**, *26* (9), 1433–1443. https://doi.org/10.1007/s13361-015-1191-1.
(137) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Anal. Chem.* **2010**, *82* (22), 9557–9565. https://doi.org/10.1021/ac1022953.
(138) Wang, C. R.; Bubner, E. R.; Jovcevski, B.; Mittal, P.; Pukala, T. L. Interrogating the Higher Order Structures of Snake Venom Proteins Using an Integrated Mass

Spectrometric Approach. *J. Proteomics* **2020**, *216*, 103680. https://doi.org/10.1016/j.jprot.2020.103680.
(139) Clemetson, K. J. Snaclecs (Snake C-Type Lectins) That Inhibit or Activate Platelets by Binding to Receptors. *Toxicon* **2010**, *56* (7), 1236–1246. https://doi.org/10.1016/j.toxicon.2010.03.011.
(140) Walker, J. R.; Nagar, B.; Young, N. M.; Hirama, T.; Rini, J. M. X-Ray Crystal Structure of a Galactose-Specific C-Type Lectin Possessing a Novel Decameric Quaternary Structure. *Biochemistry* **2004**, *43* (13), 3783–3792. https://doi.org/10.1021/bi035871a.

Appendix A: Supplementary Proteomics Data for XL Lysozyme Samples

Table A1: Peptide hits identified in DSSO cross-linked lysozyme sample

SEQUENCE	MODIFICATIONS	# PSMS	MH^+ (Da)
NTDGSTDYGILQINSR		39	1753.84
IVSDGNGMNAWVAWR	1×Oxidation [M8]	5	1691.80
KIVSDGNGMNAWVAWR	1×Oxidation [M9]	3	1819.89
KIVSDGNGMNAWVAWR	1×Oxidation [M9]; 1×DSSO [K1]	3	1977.89
CKGTDVQAWIR	1×Carboxymethyl [C1]; 1×DSSO [K2]	4	1492.66
HGLDNYR		20	874.42
NTDGSTDYGILQINSRWWCNDGR	1×Carboxymethyl [C19]	1	2729.20
GTDVQAWIR		1	1045.54
RHGLDNYR		1	1030.52
IVSDGNGMNAWVAWRNRCK	1×Carboxymethyl [C18]; 1×DSSO [K19]	2	2393.06
SLLILVLCFLPLAALGK	1×Carboxymethyl [C8]	1	1842.12
CKGTDVQAWIRGCRL	2×Carboxymethyl [C1; C13]; 1×DSSO [K2]	1	1979.88

Table A2: Peptide hits identified in linker 10 cross-linked lysozyme sample

SEQUENCE	MODIFICATIONS	# PSMS	MH^+ (Da)
NLCNIPCSALLSSDITASVNCAK	3xCarboxymethyl [C3; C7; C21]	1	2511.14
FESNFNTQATNRNTDGSTDYGILQINSR		1	3163.47
NTDGSTDYGILQINSR		25	1753.84
KIVSDGNGMNAWVAWR	1×Oxidation [M9]	11	1819.89
KIVSDGNGMNAWVAWR	1×Oxidation [M9]; 1xLinker **10**-OH [K1]	3	2177.05
IVSDGNGMNAWVAWR	1×Oxidation [M8]	15	1691.80
CKGTDVQAWIR	1×Carboxymethyl [C1];	1	1334.65
FESNFNTQATNR		4	1428.65
GTDVQAWIR		5	1045.54
HGLDNYR		1	874.42
CKGTDVQAWIRGCR	2×Carboxymethyl [C1; C13]	4	1708.79
VFGRCELAAAMK	1xCarboxymethyl [C5]; 1xOxidation [M11]; 1xLinker **10**-OH [K12]	1	1709.82

Appendix B: Supplementary Proteomics Data for Venom

Table B1: Fraction A peptide hits

FAMILY	ACCESSION CODE	HOMOLOGY	COVERAGE (%)	MW (kDa)	SEQUENCE	MODIFICATIONS	# PSMS	*Z*	m/z (Da)	RT (min)
5'NUC	T1E3Y5_CROHD	*Crotalus horridus*	20	64.8	CTGQDCYGGVAR	2×Carbamidomethyl [C1; C6]	4	2	672.28	31.18
					ETPVLSNPGPYLEFR		3	2	859.94	28.80
					GDSSNHSSGNLDISIVGDYIK	1×Deamidated [N10]	2	3	727.01	31.18
					IIALGHSGFSEDQR		5	2	765.39	28.03
					IINVGSEK		4	2	430.25	29.08
					LTTLGVNK		2	2	423.26	31.18
					QAFEHSVHR		4	3	370.85	28.80
					QVPVVQAYAFGK		3	2	653.86	28.80
					TIVYLNGTTQACR	1×Carbamidomethyl [C12]; 1×Deamidated [N6]	2	2	749.37	31.18
					VFPAVEGR		5	2	437.74	28.80
	V5NTD_GLOBR	*Gloydius brevicaudus*	23	64.4	CTGQDCYGGVAR	2×Carbamidomethyl [C1; C6]	4	2	672.28	34.97
					ETPVLSNPGPYLEFR		3	2	859.94	33.47
					GDSSNHSSGNLDISIVGDYIK	1×Deamidated [N10]	2	3	727.01	34.97
					IINVGSEK		4	2	430.25	28.03
					LTTLGVNK		2	2	423.26	33.47
					QAFEHSVHR		4	3	370.85	33.47
					QVPVVQAYAFGK		3	2	653.86	33.47
					TIVYLNGTTQACR	1×Carbamidomethyl [C12]; 1×Deamidated [N6]	2	2	749.37	34.97
					VFPAVEGR		5	2	437.74	33.47

					VVSLNVLCTECR	2×Carbamidomethyl [C8; C11]	5	2	725.36	28.03
	W8EFS0_MACLB	*Macrovipera lebetina*	22	45.0	TIVYLNGTTQACR	1×Carbamidomethyl [C12]; 1×Deamidated [N6]	2	2	749.37	33.49
					IQLQNYYSQEIGK		6	2	792.41	29.08
					ASGNPILLNK		4	2	513.80	33.49
					LTTLGVNK		2	2	423.26	33.85
					ETPVLSNPGPYLEFR		3	2	859.94	33.85
					QVPVVQAYAFGK		3	2	653.86	30.21
					QAFEHSVHR		4	3	370.85	33.36
	A0A0F7YZM6_MICFL	*Micrurus fulvius*	15	63.0	VPIYVPLQMEK	1×Deamidated [Q8]	6	2	659.36	32.59
					VPIYVPLQMEK	1×Deamidated [Q8]; 1×Oxidation [M9]	2	2	667.36	32.67
					VVSLNVLCTECR	2×Carbamidomethyl [C8; C11]	5	2	725.36	34.95
					RVVSLNVLCTECR	2×Carbamidomethyl [C9; C12]	2	3	535.94	33.25
					ASGNPILLNK		4	2	513.80	34.46
					LTTLGVNK		2	2	423.26	31.08
					ETPVLSNPGPYLEFR		3	2	859.94	31.08
					QVPVVQAYAFGK		3	2	653.86	31.08
					QAFEHSVHR		4	3	370.85	31.08
					IINVGSEK		4	2	430.25	31.28
					ASGNPILLNK		4	2	513.80	31.80
					PVVQAYAFGK		2	2	540.30	29.84
					QAFEHSVHR		4	3	370.85	31.80
	A0A6I9Y3T1_9SAUR	*Thamnophis sirtalis*	19	33.0	TIVYLNGTTQACR	1×Carbamidomethyl [C12]; 1×Deamidated [N6]	2	2	749.37	31.80
					VFPAVEGR		5	2	437.74	29.84
					VFPAVEGR		5	2	437.74	29.84
					VVYDLSQK		2	2	476.26	29.84

CTL	A0A6G5ZVN1_9SAUR	*Vipera anatolica*	24	32.5	ASGNPILLNK		4	2	513.80	28.62
					ETPVLSNPGPYLEFR		3	2	859.94	34.82
					IINVGSEK		4	2	430.25	31.80
					NANFPILSANIRPK		2	3	518.96	28.62
					QVPVVQAYAFGK		3	2	653.86	34.82
					TIVYLNGTTQACR	1×Carbamidomethyl [C12]; 1×Deamidated [N6]	2	2	749.37	29.14
					AFDEPKR		6	2	431.72	28.10
					EQYCNKK	1×Carbamidomethyl [C4]	23	2	485.23	20.86
					KEQYCNK	1×Carbamidomethyl [C5]	9	2	485.23	17.63
					KEQYCNKK	1×Carbamidomethyl [C5]	5	3	366.52	26.04
					KNCFGLEK	1×Carbamidomethyl [C3]	7	2	498.25	28.95
					KNCFGLEK	1×Carbamidomethyl [C3]; 1×Deamidated [N2]	2	2	498.75	29.40
					KWTDGSSVIYQNMVER	1×Oxidation [M13]	4	3	643.64	31.57
	A0A140DC06_BITAR	*Bitis arietans*	37	17.9	KWTDGSSVIYQNMVER		12	2	956.96	33.19
					NCFGLEK	1×Carbamidomethyl [C2]	40	2	434.21	32.24
					NCFGLEKESGYR	1×Carbamidomethyl [C2]	2	2	730.34	30.16
					SSPDYVWIGLWNQR		2	2	860.93	35.70
					WTDGSSVIYQNMVER		22	2	892.92	35.03
					WTDGSSVIYQNMVER	1×Deamidated [N/Q]	7	2	893.41	33.09
					WTDGSSVIYQNMVER	1×Oxidation [M12]	5	2	900.91	32.52
					WTDGSSVIYQNMVER	1×Deamidated [N/Q]; 1×Oxidation [M12]	3	2	901.41	32.15
	A0A1B3AXS3_BITAR	*Bitis arietans*	64	17.9	EEADFVAQLVSENVK		4	2	839.42	35.51
					ESGYRTWLNLR		3	3	465.58	33.52
					SVDAEKFCVEQAGHLASIESK	1×Carbamidomethyl [C8]	11	3	769.04	32.07
					EEADFVAQLVSENVK	1×Deamidated [N13]	2	2	839.91	35.45
					CGDDYPFVCKFPPRC	3×Carbamidomethyl [C1; C9; C15]	5	3	639.94	33.30

					CGDDYPFVCK	2×Carbamidomethyl [C1; C9]	25	2	630.75	32.11
					FCVEQAGHLASIESK	1×Carbamidomethyl [C2]	3	2	838.41	31.86
					KEQYCNK	1×Carbamidomethyl [C5]	9	2	485.23	17.63
					EQYCNKK	1×Carbamidomethyl [C4]	23	2	485.23	20.86
					KNCFGLEK	1×Carbamidomethyl [C3]	7	2	498.25	28.95
					KNCFGLEK	1×Carbamidomethyl [C3]; 1×Deamidated [N2]	2	2	498.75	29.40
					SSPDYVWIGLWNQR		2	2	860.93	35.70
					NCFGLEKESGYR	1×Carbamidomethyl [C2]	2	2	730.34	30.16
					KEQYCNKK	1×Carbamidomethyl [C5]	5	3	366.52	26.04
					AFDEPKR		6	2	431.72	28.10
					NCFGLEK	1×Carbamidomethyl [C2]	40	2	434.21	32.24
	SLRB_BITRH	*Bitis rhinoceros*	10	17.7	KTWADAEK		3	2	474.74	27.84
	I7JA31_BITRH	*Bitis rhinoceros*	12	17.7	WEWSDNAK		2	2	518.23	31.65
	I7JX23_BITRH	*Bitis rhinoceros*	19	18.1	AFDEPKR		6	2	431.72	28.10
					EEADFVAQLISDNIK		8	2	846.43	35.98
					EEADFVAQLISDNIK	1×Deamidated [Q/N]	4	2	846.92	35.55
					NCFGLEK	1×Carbamidomethyl [C2]	40	2	434.21	32.24
	SLA_ECHCS	*Echis carinatus*	5	15.8	TWDEAEK		2	2	439.70	26.64
	SL5_BITAR	*Bitis arietans*	5	17.2	SWAEAEK		12	2	410.70	29.45
	SLB_BITAR	*Bitis arietans*	10	14.8	DEGCLPDWSSYK	1×Carbamidomethyl [C4]	2	2	728.81	33.32
	SLB1_DEIAC	*Deinagkistrodon acutus*	10	16.7	FCTQQHK	1×Carbamidomethyl [C2]	104	3	316.82	25.37
					TWADAEKFCTQQHK	1×Carbamidomethyl [C9]	7	4	438.21	30.81
	SLCB_DEIAC	*Deinagkistrodon acutus*	5	16.7	TWAEAQK	1×Deamidated [Q6]	10	2	417.70	29.58
	R4FJL5_DENDV	*Denisonia devisi*	4	19.2	HPFVCK	1×Carbamidomethyl [C5]	2	2	394.20	27.76
DIS	DID_ATHCH	*Atheris chlorechis*	19	12.2	FINSGTICKK	1×Carbamidomethyl [C8]	2	3	389.88	29.22
					GEHCISGPCCR	3×Carbamidomethyl [C4; C9; C10]	2	3	444.85	27.65

	VM2_BITAR	*Bitis arietans*	55	9.0	GDWNDDYCTGK	1×Carbamidomethyl [C8]	2	2	665.75	30.00
					SSDCPWNH	1×Carbamidomethyl [C4]	3	2	501.69	29.77
	DID8A_CERCE	*Cerastes cerastes*	42	7.1	GEHCISGPCCR	3×Carbamidomethyl [C4; C9; C10]	2	3	444.85	27.65
					MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]	2	3	634.61	28.05
					MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]; 1×Deamidated [N2]; 1×Met-loss [N-Term]	2	3	591.26	27.97
					MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]; 1×Met-loss [N-Term]	2	3	590.93	27.68
	DIDB_CERVI	*Cerastes vipera*	41	7.0	GEHCISGPCCR	3×Carbamidomethyl [C4; C9; C10]	2	3	444.85	27.65
					NSAHPCCDPVTCKPK	3×Carbamidomethyl [C6; C7; C12]	2	3	590.93	27.50
	VM2D3_BITAR	*Bitis arietans*	55	9.1	GDWNDDYCTGK	1×Carbamidomethyl [C8]	2	2	665.75	30.00
					SSDCPWNH	1×Carbamidomethyl [C4]	3	2	501.69	29.77
	VM2D2_BITAR	*Bitis arietans*	43	9.1	GDWNDDYCTGK	1×Carbamidomethyl [C8]	2	2	665.75	30.00
					SSDCPWNH	1×Carbamidomethyl [C4]	3	2	501.69	29.77
	A0A0A1WDN5_ECHCO	*Echis coloratus*	24	35.0	EADYEEFLEIAR		3	2	742.85	35.41
					VTVLEASER		3	2	502.27	30.13
LAAO					YPVKPSEAGK		4	2	538.29	27.72
	OXLA_ECHOC	*Echis coloratus*	12	56.5	YPVKPSEAGK		4	2	538.29	27.72
	OXLAB_CERCE	*Cerastes cerastes*	12	19.9	ADDKNPLEECFR	1×Carbamidomethyl [C10]	2	2	747.34	31.56
					VTVLEASER		3	2	502.27	30.13
PDE	PDE_MACLB	*Macrovipera lebetina*	9	96.1	AATYFWPGSEVK		2	2	678.34	33.78
					YCLLHQAK	1×Carbamidomethyl [C2]	2	2	516.77	29.40
PLB	A0A346CLY2_9SAUR	*Ahaetulla prasina*	13	64.0	FTAYAINGPPVEK	1×Deamidated [N7]	2	2	704.36	32.52
					IANMMADSGK	1×Oxidation [M]	4	2	527.24	27.73

SVMP	VM3E1_ECHOC	*Echis ocellatus*	3	68.7	NPCCNAATCK	3×Carbamidomethyl [C3; C4; C9]; 1×Deamidated [N5]	2	2	598.73	27.07
	A0A0B4SX85_9SAUR	*Philodryas chamissonis*	1	68.4	AAGTECR	1×Carbamidomethyl [C6]	61	2	382.67	13.89
	A0A098LX17_PANGU	*Pantherophis guttatus*	5	21.5	YTGNLTAIR	1×Deamidated [N4]	3	2	505.27	31.33
	A0A0B4U9L8_VIPAA	*Vipera ammodytes*	4	68.7	LKPGAECGDGVCCYQCR	4×Carbamidomethyl [C7; C12; C13; C16]	4	2	1015.42	29.23
	A0A1I9KNR6_VIPAA	*Vipera ammodytes*	2	68.3	KENDVPIPCAAEDVK	1×Carbamidomethyl [C9]	2	2	842.92	30.52
	E9JG16_ECHCO	*Echis coloratus*	5	25.4	TLCAGVLEGGK	1×Carbamidomethyl [C3]	2	2	552.79	31.23
	E9JG31_ECHCS	*Echis carinatus*	5	39.0	AESYFYCR	1×Carbamidomethyl [C7]	2	2	548.23	30.86
	F8S112_CROAD	*Crotalus adamanteus*	3	61.0	CILNEPLR	1×Carbamidomethyl [C1]	2	2	507.77	32.45
					TAGTECR	1×Carbamidomethyl [C6]	68	2	397.68	12.77
	VM3M1_NAJMO	*Naja mossambica*	2	68.1	DPNYGMVAPGTK		30	2	625.30	31.62
					DPNYGMVAPGTK	1×Oxidation [M6]	14	2	633.29	29.85
VEGF	TXVE_BITAR	*Bitis arietans*	38	16.6	TVELQVMQVTPK	1×Oxidation [M7]	2	2	694.88	31.77
					TVELQVMQVTPK		2	2	686.88	33.29

Table B2: Fraction B peptide hits

FAMILY	ACCESSION CODE	HOMOLOGY	COVERAGE (%)	MW (kDa)	SEQUENCE	MODIFICATIONS	# PSMS	*Z*	m/z (Da)	RT (min)
5'NUC	V5NTD_CROAD	*Crotalus adamanteus*	17	64.6	CTGQDCYGGVAR	2×Carbamidomethyl [C1; C6]	2	2	672.28	28.61
					ETPVLSNPGPYLEFR		2	2	859.94	34.97
					IIALGHSGFSEDQR		2	2	765.39	31.32
					IINVGSEK		3	2	430.25	29.15
					LTTLGVNK		2	2	423.26	29.87
					QAFEHSVHR		3	3	370.85	28.14
					QVPVVQAYAFGK		2	2	653.86	33.33
					TIVYLNGTTQACR	1×Carbamidomethyl [C12]; 1×Deamidated [N6]	2	2	749.37	31.82
					VFPAVEGR		3	2	437.74	30.82
	V5NTD_NAJAT	*Naja atra*	18	58.2	ASGNPILLNK		3	2	513.80	31.09
					ETPVLSNPGPYLEFR		2	2	859.94	34.97
					IINVGSEK		3	2	430.25	29.15
					LTTLGVNK		2	2	423.26	29.87
					QAFEHSVHR		3	3	370.85	28.14
					QVPVVQAYAFGK		2	2	653.86	33.33
					RVVSLNVLCTECR	2×Carbamidomethyl [C9; C12]	3	3	535.95	32.26
					VVSLNVLCTECR	2×Carbamidomethyl [C8; C11]	2	2	725.36	33.46
CTL	SL2_BITGA	*Bitis gabonica*	22	18.1	EEADFVAQLISDNIK		2	2	846.43	35.64
					EGESQMCQALTK	1×Carbamidomethyl [C7]	5	2	691.31	30.08
					EGESQMCQALTK	1×Carbamidomethyl [C7]; 1×Oxidation [M6]	8	2	699.30	28.19
					EQQCSSEWNDGSK	1×Carbamidomethyl [C4]	7	2	777.81	28.58
					NCFGLEK	1×Carbamidomethyl [C2]; 1×Deamidated [N1]	3	2	434.70	31.71

DIS	SL5_BITAR	*Bitis arietans*	25	17.2	FCMEQANDGHLVSIQSIK	1×Carbamidomethyl [C2]; 1×Oxidation [M3]	3	3	698.33	31.88
					VTYVNWR		3	2	469.25	35.61
					VTYVNWR	1×Deamidated [N5]	2	2	469.74	31.78
	SL7_DABSI	*Daboia siamensis*	15	18.1	FCNEQVNGGYLVSFR	1×Carbamidomethyl [C2]; 1×Deamidated [N]	2	2	895.92	32.25
					KTWEDAEK		4	2	503.75	29.42
					TWEDAEK		11	2	439.70	29.17
	SLA_BITAR	*Bitis arietans*	18	14.9	LASQTLTK		10	2	431.26	28.27
			18	14.8	DEGCLPDWSSYK	1×Carbamidomethyl [C4]	6	2	728.81	33.64
	SLB_ECHCS	*Echis carinatus*	5	14.9	AWDNER		3	2	395.67	27.67
	SLB1_DEIAC	*Deinagkistrodon acutus*	10	16.7	FCTQQHK	1×Carbamidomethyl [C2]	15	3	316.82	25.55
	SLED_CALRH	*Calloselasma rhodostoma*	12	14.8	LASIHSR		2	2	392.23	27.20
					WEWSDDAK		3	2	518.72	31.99
					KTWADAEK		18	2	474.74	27.64
	SLRB_BITRH	*Bitis rhinoceros*	12	17.7	QQYFVCK	1×Carbamidomethyl [C6]; 1×Deamidated [Q]	19	2	487.23	31.37
					QQYFVCK	1×Carbamidomethyl [C6]	4	2	486.73	29.97
					TWADAEK		13	2	410.70	30.42
					TWADAEKFCK	1×Carbamidomethyl [C9]	2	3	419.20	30.65
	DID_ATHCH	*Atheris chlorechis*	19	12.2	FINSGTICK	1×Carbamidomethyl [C8]	3	2	520.27	30.46
					FINSGTICK	1×Carbamidomethyl [C8]	3	2	520.27	30.46
					FINSGTICK	1×Carbamidomethyl [C8]; 1×Deamidated [N3]	2	2	520.76	30.93
					FINSGTICKK	1×Carbamidomethyl [C8]	2	3	389.88	29.21
	DID5_CERCE	*Cerastes cerastes*	43	7.2	MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]	4	3	634.61	28.04
					MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]; 1×Deamidated [N2]; 1×Met-loss [N-Term]	4	3	591.26	28.16
					MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]; 1×Met-loss [N-Term]	5	3	590.93	27.54

	VM2_BITAR	*Bitis arietans*	86	9.0	RGEHCISGPCCR	3×Carbamidomethyl [C5; C10; C11]	2	4	372.91	29.00
					CCNAATCK	3×Carbamidomethyl [C1; C2; C7]	2	2	492.69	25.98
					GDWNDDYCTGK	1×Carbamidomethyl [C8]	6	2	665.75	30.37
					ILEQGEDCDCGSPANCQDR	3×Carbamidomethyl [C8; C10; C16]	5	2	1112.44	29.33
					LTPGSQCNYGECCDQCR	4×Carbamidomethyl [C7; C12; C13; C16]	4	3	702.27	29.38
					SPPVCGNK	1×Carbamidomethyl [C5]	4	2	429.71	26.23
					SSDCPWNH	1×Carbamidomethyl [C4]	3	2	501.69	29.79
					SSDCPWNH	1×Carbamidomethyl [C4]; 1×Deamidated [N7]	2	2	502.18	29.92
	VM2D2_BITAR	*Bitis arietans*	53	9.1	LTPGSQCNYGECCDQCK	4×Carbamidomethyl [C7; C12; C13; C16]	2	3	692.93	29.10
HYAL	HYAL_ECHOC	*Echis ocellatus*	5	52.5	ICSHALCR	2×Carbamidomethyl [C2; C7]	2	3	339.50	27.66
					YMNGPLGR	1×Deamidated [N3]	2	2	454.72	30.74
	OXLA_BOTCO	*Bothrops cotiara*	80	1.8	ADDRNPLEECFR	1×Carbamidomethyl [C10]	3	2	761.34	33.03
	OXLA_BUNFA	*Bungarus fasciatus*	5	58.7	EADYEEFLEIAR		3	2	742.85	35.40
LAAO	OXLA_DABRR	*Daboia russelii*	20	56.9	IFLTCTK	1×Carbamidomethyl [C5]	3	2	441.74	31.34
					IFLTCTKK	1×Carbamidomethyl [C5]	2	3	337.53	29.67
					FDEIVGGMDQLPTSMYR	1×Oxidation [M]	2	2	987.95	34.03
					YDTYSTK		5	2	439.20	27.82
					STTDLPSR		2	2	438.73	28.35
					YPVKPSEAGK		5	3	359.20	28.02
					AVEELKR		2	2	422.75	27.29
					IQQNAEK		12	2	415.72	23.61
					KDPGLLK		2	2	385.74	28.60
					KDPGLLKYPVKPSEAGK		2	4	457.52	31.25
					HDDIFAYEKR		2	3	431.88	31.00
					HDDIFAYEK		5	2	569.26	30.99
	OXLA_GLOBL	*Gloydius blomhoffii*	11	57.1	NDKEGWYANLGPMR	1×Deamidated [N1]	2	3	551.26	33.72

	OXLAB_CERCE	*Cerastes cerastes*	12	19.9	ADDKNPLEECFR	1×Carbamidomethyl [C10]	10	2	747.34	31.88
					VTVLEASER		12	2	502.27	30.33
PDE	PDE_MACLB	*Macrovipera lebetina*	14	96.1	YCLLHQAK	1×Carbamidomethyl [C2]	2	2	516.77	29.47
					TLGMLMEGLK	1×Oxidation [M]	2	2	554.79	33.93
	PLB_CROAD	*Crotalus adamanteus*	16	64.0	NVIIEQK		2	2	422.25	29.21
	PLB_DRYCN	*Drysdalia coronoides*	13	64.1	NNKDDPFWR	1×Deamidated [N2]	2	3	398.18	32.22
					FTAYAINGPPVEK	1×Deamidated [N7]	2	2	704.36	32.72
PLB					IANMMADSGK	1×Oxidation [M]	3	2	527.24	27.70
					HNPCNTICCR	3×Carbamidomethyl [C4; C8; C9]	3	2	666.27	27.58
					VADINMAAK		2	2	466.75	29.82
					IANMMADSGK		2	2	519.24	29.64
	VM1N_BOTPA	*Bothrops pauloensis*	6	22.5	AYLGSMCNPK	1×Carbamidomethyl [C7]; 1×Deamidated [N8]	5	2	571.25	30.62
					AYLGSMCNPKR	1×Carbamidomethyl [C7]; 1×Deamidated [N8]	3	3	433.21	29.43
					AYLGSMCNPK	1×Carbamidomethyl [C7]; 1×Deamidated [N8]; 1×Oxidation [M6]	7	2	579.25	29.55
	VM2H1_BOTLA	*Bothriechis lateralis*	4	54.3	AAGTVCR	1×Carbamidomethyl [C6]	3	2	367.68	25.99
					VNGEPVVLHLEK	1×Deamidated [N2]	3	2	667.87	32.17
SVMP	VM33_CROAD	*Crotalus adamanteus*	3	68.8	DYSETHYSPDGR		2	2	713.80	28.83
	VM3CX_MACLB	*Macrovipera lebetina*	3	68.7	TAGTVCR	1×Carbamidomethyl [C6]	4	2	382.69	26.32
	VM3E1_ECHOC	*Echis ocellatus*	6	68.7	LYCFDNLPEHK	1×Carbamidomethyl [C3]	4	3	479.23	31.86
	VM3M1_NAJMO	*Naja mossambica*	2	68.1	DPNYGMVAPGTK		3	2	625.30	30.81
					DPNYGMVAPGTK	1×Oxidation [M6]	7	2	633.30	29.33
	VM3VA_MACLB	*Macrovipera lebetina*	10	68.7	LTPGSQCADGECCDQCK	4×Carbamidomethyl [C7; C12; C13; C16]	2	2	993.38	28.31
					IPCAPQDVK	1×Carbamidomethyl [C3]	2	2	514.27	29.56

	VASP1_VIPAA	*Vipera ammodytes*	8	22.1	TLCAGILQGGIDSCK	2×Carbamidomethyl [C3; C14]; 1×Deamidated [Q8]	4	2	797.39	33.47
	VSP_BOTIN	*Bothrops insularis*	4	28.3	TLCAGVLQGGK	1×Carbamidomethyl [C3]; 1×Deamidated [Q8]	38	2	552.79	31.39
	VSP05_TRIST	*Trimeresurus stejnegeri*	3	28.5	WNKDIMLIK	1×Deamidated [N2]	17	2	581.32	33.61
					WNKDIMLIK	1×Deamidated [N2]; 1×Oxidation [M6]	11	2	589.32	32.67
SVSP	VSP1_MACLB	*Macrovipera lebetina*	3	28.7	EMFFCLSNK	1×Carbamidomethyl [C5]	2	2	588.27	34.24
					EMFFCLSNK	1×Carbamidomethyl [C5]; 1×Oxidation [M2]	3	2	596.26	33.26
	VSP13_TRIST	*Trimeresurus stejnegeri*	9	28.4	NHTQWNKDIMLIR	3×Deamidated [N1; Q4; N6]; 1×Oxidation [M10]	3	3	563.28	35.04
	VSPB_GLOBL	*Gloydius blomhoffii*	9	25.7	VIGGDECNINEHR	1×Carbamidomethyl [C7]; 1×Deamidated [N8]	3	2	757.34	29.04
					YFCLSSR	1×Carbamidomethyl [C3]	24	2	466.72	33.23
	VSPP_CERCE	*Cerastes cerastes*	4	28.0	VFDYTDWIR		3	2	607.80	35.55
VEGF	TXVE_BITAR	*Bitis arietans*	27	16.6	TVELQVMQVTPK		2	2	686.88	33.29
					TVELQVMQVTPK	1×Oxidation [M7]	2	2	694.88	31.79

Table B3: Fraction C peptide hits

FAMILY	ACCESSION CODE	HOMOLOGY	COVERAGE (%)	MW (kDa)	SEQUENCE	MODIFICATIONS	# PSMS	*Z*	M/Z (Da)	RT (min)
5'NUC	V5NTD_CROAD	*Crotalus adamanteus*	7	64.6	CTGQDCYGGVAR	2×Carbamidomethyl [C1; C6]	2	2	672.28	28.66
					IINVGSEK		2	2	430.25	29.20
					IINVGSEK		2	2	430.25	29.20
					QVPVVQAYAFGK		2	2	653.86	33.32
					QVPVVQAYAFGK		2	2	653.86	33.32
CTL	SL2_BITGA	*Bitis gabonica*	14	18.1	EEADFVAQLISDNIK		2	2	846.43	35.65
	SL3_BITGA	*Bitis gabonica*	26	18.1	EGESQMCQALTK	1×Carbamidomethyl [C7]	2	2	691.31	30.11
					EQQCSSEWNDGSK	1×Carbamidomethyl [C4]; 1×Deamidated [Q/N]	2	2	778.31	28.54
					EQQCSSEWNDGSK	1×Carbamidomethyl [C4]	4	2	777.81	28.18
	SL5_BITAR	*Bitis arietans*	26	17.2	TWEDAEK		6	2	439.70	28.18
					EANFVAK	1×Deamidated [N3]	3	2	390.20	29.52
					SWAEAEK		2	2	410.70	28.44
					VTYVNWR	1×Deamidated [N5]	2	2	469.74	31.92
					VTYVNWR		7	2	469.25	34.35
	SL7_DABSI	*Daboia siamensis*	15	18.1	KTWEDAEK		2	2	503.75	27.70
					TWEDAEK		6	2	439.70	28.18
	SLA_BITAR	*Bitis arietans*	13	14.9	LASQTLTK		5	2	431.26	28.26
					VGTWEDAEK		3	2	517.74	29.27
			15	14.8	DEGCLPDWSSYK	1×Carbamidomethyl [C4]	4	2	728.81	33.26
					TWADAEK		3	2	410.70	31.01
	SLB_ECHCS	*Echis carinatus*	5	14.9	AWDNER		3	2	395.67	27.77
	SLB1_DEIAC	*Deinagkistrodon acutus*	10	16.7	FCTQQHK	1×Carbamidomethyl [C2]	2	3	316.82	26.29
					TWADAEK		3	2	410.70	31.01

	SLED_CALRH	*Calloselasma rhodostoma*	23	14.8	LASIHSR		2	2	392.23	27.22
					WEWSDDAK		2	2	518.72	32.08
					KTWADAEK		2	2	474.74	27.72
	SLRB_BITRH	*Bitis rhinoceros*	10	17.7	QQYFVCK	1×Carbamidomethyl [C6]; 1×Deamidated [Q2]	4	2	487.23	30.36
					QQYFVCK	1×Carbamidomethyl [C6]	3	2	486.73	30.04
					TWADAEK		3	2	410.70	31.01
					DVTDPDVQEAAAFAVEK		5	2	902.94	34.23
					IPGGLSPR		18	2	398.74	31.00
CYS	CYT_BITAR	*Bitis arietans*	57	12.7	NDYYFK	1×Deamidated [N1]	4	2	425.69	31.76
					RVVEAQSQVVSGVK		9	2	743.42	29.23
					VVEAQSQVVSGVK	1×Deamidated [Q5]	2	2	665.87	30.24
					VVEAQSQVVSGVK		18	2	665.37	29.84
					FINSGTICK	1×Carbamidomethyl [C8]	22	2	520.27	33.64
	DID_ATHCH	*Atheris chlorechis*	19	12.2	FINSGTICK	1×Carbamidomethyl [C8]; 1×Deamidated [N3]	6	2	520.76	30.94
					FINSGTICKK	1×Carbamidomethyl [C8]	2	3	389.88	29.17
					GEHCISGPCCR	3×Carbamidomethyl [C4; C9; C10]	4	2	666.77	27.63
DIS					GEHCISGPCCR	3×Carbamidomethyl [C4; C9; C10]	4	2	666.77	27.63
	DID5_CERCE	*Cerastes cerastes*	43	7.2	MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]; 1×Deamidated [N2]; 1×Met-loss [N-Term]	7	3	591.26	27.85
					MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]; 1×Met-loss [N-Term]	7	3	590.93	27.67

				MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]; 1×Deamidated [N2]	3	4	476.45	28.27
				MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]; 1×Oxidation [M1]	8	3	639.94	27.65
				MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]; 1×Deamidated [N2]; 1×Oxidation [M1]	2	4	480.45	27.97
				MNSAHPCCDPVTCKPK	1×Acetyl [N-Term]; 3×Carbamidomethyl [C7; C8; C13]; 1×Deamidated [N2]	3	3	648.94	28.80
				MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]	6	2	951.41	27.99
				RGEHCISGPCCR	3×Carbamidomethyl [C5; C10; C11]	21	3	496.88	27.78
VM2_BITAR	*Bitis arietans*	94	9.0	CCNAATCK	3×Carbamidomethyl [C1; C2; C7]; 1×Deamidated [N3]	4	2	493.18	25.85
				CCNAATCK	3×Carbamidomethyl [C1; C2; C7]	65	2	492.69	19.15
				GDWNDDYCTGK	1×Carbamidomethyl [C8]; 1×Deamidated [N4]	2	2	666.25	30.35
				GDWNDDYCTGK	1×Carbamidomethyl [C8]	24	2	665.75	30.05
				GDWNDDYCTGKSSDCPWNH	2×Carbamidomethyl [C8; C15]	3	3	771.96	31.87
				ILEQGEDCDCGSPANCQDR	3×Carbamidomethyl [C8; C10; C16]	7	2	1112.44	29.26
				KAGTVCR	1×Carbamidomethyl [C6]	2	2	396.21	15.54
				SPPVCGNK	1×Carbamidomethyl [C5]	15	2	429.71	27.05

VM2D2_BITAR	*Bitis arietans*	53	9.1	SSDCPWNH	1×Carbamidomethyl [C4]; 1×Deamidated [N7]	3	2	502.18	29.88
				SSDCPWNH	1×Carbamidomethyl [C4]	20	2	501.69	31.49
				CCNAATCK	3×Carbamidomethyl [C1; C2; C7]; 1×Deamidated [N3]	4	2	493.18	25.85
				CCNAATCK	3×Carbamidomethyl [C1; C2; C7]	65	2	492.69	19.15
				GDWNDDYCTGK	1×Carbamidomethyl [C8]; 1×Deamidated [N4]	2	2	666.25	30.35
				GDWNDDYCTGK	1×Carbamidomethyl [C8]; 1×Deamidated [N4]	2	2	666.25	30.35
				GDWNDDYCTGK	1×Carbamidomethyl [C8]	24	2	665.75	30.05
				GDWNDDYCTGKSSDCPWNH	2×Carbamidomethyl [C8; C15]	3	3	771.96	31.87
				LTPGSQCNYGECCDQCK	4×Carbamidomethyl [C7; C12; C13; C16]; 2×Deamidated [Q6; N8]	2	2	1039.89	27.48
				SSDCPWNH	1×Carbamidomethyl [C4]; 1×Deamidated [N7]	3	2	502.18	29.88
				SSDCPWNH	1×Carbamidomethyl [C4]	20	2	501.69	31.49
VM2D3_BITAR	*Bitis arietans*	86	9.1	CCNAATCK	3×Carbamidomethyl [C1; C2; C7]; 1×Deamidated [N3]	4	2	493.18	25.85
				CCNAATCK	3×Carbamidomethyl [C1; C2; C7]	65	2	492.69	19.15
				GDWNDDYCTGK	1×Carbamidomethyl [C8]	24	2	665.75	30.05

					GDWNDDYCTGKSSDCPWNH	2×Carbamidomethyl [C8; C15]	3	3	771.96	31.87
					SPPVCGNELLEEGEECDCGSPANCQDR	4×Carbamidomethyl [C5; C16; C18; C24]	5	3	1027.08	32.30
					SSDCPWNH	1×Carbamidomethyl [C4]; 1×Deamidated [N7]	3	2	502.18	29.88
					SSDCPWNH	1×Carbamidomethyl [C4]	20	2	501.69	31.49
KUN	VKT2_BITGA	*Bitis gabonica*	9	10.0	FYYDSASK		2	2	490.72	29.77
LAAO	OXLA_DABRR	*Daboia russelii*	7	56.9	IFLTCTK	1×Carbamidomethyl [C5]	2	2	441.74	31.21
					YPVKPSEAGK		2	3	359.20	27.73
NGF	NGFV_NAJAT	*Naja atra*	10	13.1	ALTMEGNQASWR	1×Oxidation [M4]	5	2	690.32	30.13
					ALTMEGNQASWR		3	2	682.33	31.52
PLB	PLB_CROAD	*Crotalus adamanteus*	2	64.0	FTAYAINGPPVEK		2	2	703.87	32.45
SVMP	VM1N_BOTPA	*Bothrops pauloensis*	6	22.5	AYLGSMCNPK	1×Carbamidomethyl [C7]; 1×Deamidated [N8]	6	2	571.26	30.82
					AYLGSMCNPK	1×Carbamidomethyl [C7]; 1×Deamidated [N8]; 1×Oxidation [M6]	10	2	579.25	30.14
					AYLGSMCNPKR	1×Carbamidomethyl [C7]; 1×Deamidated [N8]	2	3	433.21	29.39
	VM2DI_GLOHA	*Gloydius halys*	8	53.1	QGAQCAEGLCCDQCRFMK	4×Carbamidomethyl [C5; C10; C11; C14]; 1×Deamidated [Q13]	2	4	555.72	34.27
	VM33_CROAD	*Crotalus adamanteus*	4	68.8	DYSETHYSPDGR		3	2	713.80	28.81
	VM3CX_MACLB	*Macrovipera lebetina*	5	68.7	TAGTVCR	1×Carbamidomethyl [C6]	2	2	382.69	26.21
	VM3M1_NAJMO	*Naja mossambica*	2	68.1	DPNYGMVAPGTK		4	2	625.30	30.63
					DPNYGMVAPGTK	1×Oxidation [M6]	8	2	633.30	29.97

	VM3VA_MACLB	*Macrovipera lebetina*	11	68.7	IPCAPQDVK	1×Carbamidomethyl [C3]	3	2	514.27	29.70
					LTPGSQCADGECCDQCK	4×Carbamidomethyl [C7; C12; C13; C16]	2	2	993.38	28.29
					NPCQIYYTPSDENKGMVDPGTK	1×Carbamidomethyl [C3]; 1×Deamidated [N13]; 1×Oxidation [M16]	2	3	844.37	30.61
SVSP	VASP1_VIPAA	*Vipera ammodytes*	12	22.1	TLCAGILQGGIDSCK	2×Carbamidomethyl [C3; C14]; 1×Deamidated [Q8]	3	2	797.39	33.48
					WDKDIMLIR		4	2	595.32	33.42
	VSP_BOTIN	*Bothrops insularis*	4	28.3	TLCAGVLQGGK	1×Carbamidomethyl [C3]; 1×Deamidated [Q8]	8	2	552.79	31.24
	VSP05_TRIST	*Trimeresurus stejnegeri*	3	28.5	WNKDIMLIK	1×Deamidated [N2]; 1×Oxidation [M6]	2	2	589.32	32.21
					WNKDIMLIK	1×Deamidated [N2]	6	2	581.32	33.18
	VSP13_TRIST	*Trimeresurus stejnegeri*	5	28.4	NHTQWNKDIMLIR	3×Deamidated [N1; Q4; N6]; 1×Oxidation [M10]	6	3	563.28	34.32
					NHTQWNKDIMLIR	3×Deamidated [N1; Q4; N6]	5	2	836.42	35.28
	VSPB_GLOBL	*Gloydius blomhoffii*	9	25.7	YFCLSSR	1×Carbamidomethyl [C3]	5	2	466.72	34.60
	VSPP_CERCE	*Cerastes cerastes*	4	28.0	VFDYTDWIR		3	2	607.80	35.36
VEGF	TXVE_BITAR	*Bitis arietans*	46	16.6	CGGCCSDESLTCTSVGER	4×Carbamidomethyl [C1; C4; C5; C12]	4	2	1017.88	29.97
					EHTACECRPR	2×Carbamidomethyl [C5; C7]	2	3	439.19	27.03
					ETLVSILEEYPDK		2	3	512.60	35.50
					FREHTACECRPR	2×Carbamidomethyl [C7; C9]	2	4	405.44	30.04
					IFRPSCVAVLR	1×Carbamidomethyl [C6]	3	3	439.92	32.02
					TVELQVMQVTPK	1×Oxidation [M7]	13	2	694.88	32.32

Table B4: Fraction D peptide hits

Family	Accession Code	*Homology*	Coverage (%)	Mw (kDa)	Sequence	Modifications	# PSMs	*z*	m/z (Da)	RT (min)
5'NUC	A0A6B2FAP6_9SAUR	*Bothriechis nubestris*	7	64.7	CTGQDCYGGVAR	2×Carbamidomethyl [C1; C6]	3	2	672.28	28.63
					VVSLNVLCTECR	2×Carbamidomethyl [C8; C11]	2	2	725.36	33.49
	A0A6G5ZVN1_9SAUR	*Vipera anatolica*	15	32.5	IINVGSEK		3	2	430.25	29.16
CRISP	CRVP_LATSE	*Laticauda semifasciata*	18	26.4	CPASCFCR	3×Carbamidomethyl [C1; C5; C7]	3	2	529.20	28.94
	R4G2Q3_9SAUR	*Brachyurophis roperi*	17	27.7	CPASCFCR	3×Carbamidomethyl [C1; C5; C7]	3	2	529.20	28.94
	A0A1B3AXS3_BITAR	*Bitis arietans*	21	17.9	CGDDYPFVCK	2×Carbamidomethyl [C1; C9]	2	2	630.76	31.74
	A0A5A4WN20_BITAR	*Bitis arietans*	10	17.4	TWADAEK		3	2	410.69	28.62
	I7JA31_BITRH	*Bitis rhinoceros*	12	17.7	WEWSDNAK		2	2	518.23	31.61
					YVSFVCK	1×Carbamidomethyl [C6]	2	2	451.73	30.96
					YVSFVCKFPA	1×Carbamidomethyl [C6]	2	2	609.31	33.98
	I7JX23_BITRH	*Bitis rhinoceros*	14	18.1	EEADFVAQLISDNIK		2	2	846.43	35.63
CTL	SL3_BITGA	*Bitis gabonica*	20	18.1	EGESQMCQALTK	1×Carbamidomethyl [C7]	4	2	691.31	30.09
					EQQCSSEWNDGSK	1×Carbamidomethyl [C4]; 1×Deamidated [N/Q]	2	2	778.31	28.47
					EQQCSSEWNDGSK	1×Carbamidomethyl [C4]	2	2	777.81	28.47
	SL5_BITAR	*Bitis arietans*	21	17.2	TWEDAEK		3	2	439.70	28.85
					VTYVNWR	1×Deamidated [N5]	2	2	469.74	32.00
					VTYVNWR		3	2	469.25	31.55
	SL9B_ECHCA	*Echis carinatus*	14	14.4	TWEDAEK		3	2	439.70	28.85
	SLA_BITAR	*Bitis arietans*	13	14.9	LASQTLTK		2	2	431.26	28.48

CYS	SLB_BITAR	*Bitis arietans*	15	14.8	DEGCLPDWSSYK	1×Carbamidomethyl [C4]	2	2	728.81	33.29
					TWADAEK		3	2	410.69	28.62
	SLCB_DEIAC	*Deinagkistrodon acutus*	5	16.7	TWAEAQK	1×Deamidated [Q6]	3	2	417.70	31.12
	SLMB_MACLB	*Macrovipera lebetina*	6	15.0	KTWEDAEK		4	2	503.75	27.82
					TWEDAEK		3	2	439.70	28.85
	SLRB_BITRH	*Bitis rhinoceros*	18	17.7	TWADAEK		3	2	410.69	28.62
	A0A077LA61_PROEL	*Protobothrops elegans*	17	26.9	MEWYPEAAANAER	1×Oxidation [M1]	8	2	777.34	31.83
					MEWYPEAAANAER		8	3	513.23	32.79
					RSVNPTASNMLK		2	3	439.90	29.39
					SVNPTASNMLK	1×Deamidated [N8]	2	2	581.79	30.87
					SVNPTASNMLK		4	2	581.30	30.53
					SVNPTASNMLK	1×Oxidation [M9]	4	2	589.30	28.84
					KPEIQNQIVDLHNSLR	2×Deamidated [N13; N/Q]	2	3	636.01	32.34
					KPEIQNQIVDLHNSLR	1×Deamidated [Q/N]	6	3	635.68	33.05
	A0A194AMN4_9SAUR	*Agkistrodon piscivorus*	17	26.7	MEWYPEAAANAER	1×Oxidation [M1]	8	2	777.34	31.83
					MEWYPEAAANAER		8	3	513.23	32.79
					RSVNPTASNMLK		2	3	439.90	29.39
					SVNPTASNMLK	1×Deamidated [N8]	2	2	581.79	30.87
					SVNPTASNMLK		4	2	581.30	30.53
					SVNPTASNMLK	1×Oxidation [M9]	4	2	589.30	28.84
	CRVP_ECHCO	*Echis coloratus*	10	24.7	MEWYPEAAANAER	1×Oxidation [M1]	8	2	777.34	31.83
					MEWYPEAAANAER		8	3	513.23	32.79
					DVTDPDVQEAAAFAVEK		8	2	902.93	34.38
	CYT_BITAR	*Bitis arietans*	35	12.7	IPGGLSPR		47	2	398.74	29.99
					RVVEAQSQVVSGVK		8	2	743.42	29.08
					VVEAQSQVVSGVK		23	2	665.37	29.81
					KPEIQNEIVDLHNSLR		3	3	635.68	32.86
	F2Q6E5_CROHD	*Crotalus horridus*	23	24.7	MEWYPEAAANAER	1×Oxidation [M1]	8	2	777.34	31.83
					MEWYPEAAANAER		8	3	513.23	32.79
					RSVNPTASNMLK		2	3	439.90	29.39
					SVDFDSESPR		4	2	569.75	30.03

	F2Q6F2_DABRR	*Daboia russelii*	12	26.7	SVNPTASNMLK	1×Deamidated [N8]	2	2	581.79	30.87
					SVNPTASNMLK		4	2	581.30	30.53
					SVNPTASNMLK	1×Oxidation [M9]	4	2	589.30	28.84
					MEWYPEAAANAER	1×Oxidation [M1]	8	2	777.34	31.83
					MEWYPEAAANAER		8	3	513.23	32.79
					MEWYPEAAANAER	1×Oxidation [M1]	8	2	777.34	31.83
					MEWYPEAAANAER		8	3	513.23	32.79
	F2Q6F3_DABRR	*Daboia russelii*	19	25.0	RSVNPTASNMLK		2	3	439.90	29.39
					SVDFDSESPR		4	2	569.75	30.03
					SVNPTASNMLK	1×Deamidated [N8]	2	2	581.79	30.87
					SVNPTASNMLK		4	2	581.30	30.53
					SVNPTASNMLK	1×Oxidation [M9]	4	2	589.30	28.84
DIS	DID2_BITGA	*Bitis gabonica*	25	13.8	RGEHCISGPCCR	3×Carbamidomethyl [C5; C10; C11]	4	4	372.91	28.53
					TMLDGLNDYCTGVTPDCPR	2×Carbamidomethyl [C10; C17]	3	3	728.98	33.11
	DID5A_ECHOC	*Echis ocellatus*	9	12.4	FINSGTICK	1×Carbamidomethyl [C8]; 1×Deamidated [N3]	4	2	520.76	30.91
					FINSGTICKK	1×Carbamidomethyl [C8]	4	3	389.88	29.16
					FLNSGTICK	1×Carbamidomethyl [C8]	24	2	520.27	30.41
	DID8A_CERCE	*Cerastes cerastes*	43	7.1	MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]; 1×Deamidated [N2]; 1×Met-loss [N-Term]	3	3	591.26	28.10
					MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]	3	3	634.61	28.00
					MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]; 1×Oxidation [M1]	3	4	480.21	27.73
					MNSAHPCCDPVTCKPK	1×Acetyl [N-Term]; 3×Carbamidomethyl	2	3	648.93	28.78

						[C7; C8; C13]; 1×Deamidated [N2]				
					RGEHCISGPCCR	3×Carbamidomethyl [C5; C10; C11]	4	4	372.91	28.53
	DIDB_CERVI	*Cerastes vipera*	42	7.0	NSAHPCCDPVTCKPK	3×Carbamidomethyl [C6; C7; C12]	7	3	590.93	27.56
					RGEHCISGPCCR	3×Carbamidomethyl [C5; C10; C11]	4	4	372.91	28.53
	Q4JCR9_BITAR	*Bitis arietans*	70	11.9	AGTVCRIARGDWNDDYCTGK	2×Carbamidomethyl [C5; C17]	2	4	579.51	31.45
					GDWNDDYCTGK	1×Carbamidomethyl [C8]	25	2	665.75	34.34
					GDWNDDYCTGKSSDCPWNH	2×Carbamidomethyl [C8; C15]	7	3	771.96	32.89
					ILEQGEDCDCGSPANCQDR	3×Carbamidomethyl [C8; C10; C16]	27	2	1112.44	29.25
					ILEQGEDCDCGSPANCQDR	3×Carbamidomethyl [C8; C10; C16]; 1×Deamidated [N15]	2	2	1112.94	29.44
					SSDCPWNH	1×Carbamidomethyl [C4]; 1×Deamidated [N7]	4	2	502.18	30.00
					SSDCPWNH	1×Carbamidomethyl [C4]	58	2	501.69	31.39
					TVIVSPPVCGNK	1×Carbamidomethyl [C9]	3	2	635.85	31.17
					AGTVCRIARGDWNDDYCTGK	2×Carbamidomethyl [C5; C17]	2	4	579.51	31.45
	VM2_BITAR	*Bitis arietans*	96	9.0	CCNAATCK	3×Carbamidomethyl [C1; C2; C7]	66	2	492.69	17.09
					CCNAATCK	3×Carbamidomethyl [C1; C2; C7]; 1×Deamidated [N3]	2	2	493.18	26.14
					GDWNDDYCTGK	1×Carbamidomethyl [C8]	25	2	665.75	34.34

					GDWNDDYCTGKSSDCPWNH	2×Carbamidomethyl [C8; C15]	7	3	771.96	32.89
					ILEQGEDCDCGSPANCQDR	3×Carbamidomethyl [C8; C10; C16]	27	2	1112.44	29.25
					ILEQGEDCDCGSPANCQDR	3×Carbamidomethyl [C8; C10; C16]; 1×Deamidated [N15]	2	2	1112.94	29.44
					SPPVCGNK	1×Carbamidomethyl [C5]	26	2	429.71	26.68
					SSDCPWNH	1×Carbamidomethyl [C4]; 1×Deamidated [N7]	4	2	502.18	30.00
					SSDCPWNH	1×Carbamidomethyl [C4]	58	2	501.69	31.39
					CCNAATCK	3×Carbamidomethyl [C1; C2; C7]	66	2	492.69	17.09
					CCNAATCK	3×Carbamidomethyl [C1; C2; C7]; 1×Deamidated [N3]	2	2	493.18	26.14
					GDWNDDYCTGK	1×Carbamidomethyl [C8]	25	2	665.75	34.34
					GDWNDDYCTGKSSDCPWNH	2×Carbamidomethyl [C8; C15]	7	3	771.96	32.89
	VM2D3_BITAR	*Bitis arietans*	92	9.1	LTPGSQCSYGECCDQCK	4×Carbamidomethyl [C7; C12; C13; C16]	2	3	683.93	27.25
					SPPVCGNELLEEGEECDCGSPANCQDR	4×Carbamidomethyl [C5; C16; C18; C24]	20	3	1027.08	32.79
					SPPVCGNELLEEGEECDCGSPANCQDR	4×Carbamidomethyl [C5; C16; C18; C24]; 1×Deamidated [N7]	3	3	1027.40	32.04
					SSDCPWNH	1×Carbamidomethyl [C4]; 1×Deamidated [N7]	4	2	502.18	30.00
					SSDCPWNH	1×Carbamidomethyl [C4]	58	2	501.69	31.39

	VM2TI_ERIMA	*Eristicophis macmahoni*	22	5.8	GDWNNDYCTGK	1×Carbamidomethyl [C8]; 2×Deamidated [N4; N5]	4	1	1331.50	30.00
					GDWNNDYCTGK	1×Carbamidomethyl [C8]; 1×Deamidated [N]	49	1	1330.50	29.87
LAAO	A0A0A1WCY6_ECHCO	*Echis coloratus*	7	56.7	IFLTCTK	1×Carbamidomethyl [C5]	2	2	441.74	31.18
NGF	A0A670YIL7_PSETE	*Pseudonaja textilis*	12	18.0	QYFFETK		2	2	481.73	32.19
	Q9PS29_BITAR	*Bitis arietans*	52	2.5	EVRPFMEVYQR		15	3	485.25	32.08
					EVRPFMEVYQR	1×Oxidation [M6]	6	3	490.58	31.05
SVMP	A0A194AQG5_9SAUR	*Agkistrodon piscivorus*	2	67.9	TDIVSPPVCGNK	1×Carbamidomethyl [C9]	3	2	643.82	31.06
	A0A1W7RJU5_AGKCO	*Agkistrodon contortrix*	1	68.1	CCDAATCK	3×Carbamidomethyl [C1; C2; C7]	2	2	493.18	30.30
	A0A2H4Z2X4_DABSI	*Daboia siamensis*	8	69.1	IPCAPQDVK	1×Carbamidomethyl [C3]	2	2	514.27	29.55
	A0A6B7FNN4_VIPAA	*Vipera ammodytes*	7	68.3	LTPGSQCADGECCDQCK	4×Carbamidomethyl [C7; C12; C13; C16]	2	2	993.38	28.27
					IPCAPQDVK	1×Carbamidomethyl [C3]	2	2	514.27	29.55
	E9JG31_ECHCS	*Echis carinatus sochureki*	5	39.0	AESYFYCR	1×Carbamidomethyl [C7]	5	2	548.23	30.84
					KAESYFYCR	1×Carbamidomethyl [C8]	4	2	612.28	29.72
					LFCELIK	1×Carbamidomethyl [C3]	4	2	461.76	33.53
	F8S112_CROAD	*Crotalus adamanteus*	4	61.0	DYSETHYSPDGR		3	2	713.80	28.77
	V8NKR8_OPHHA	*Ophiophagus hannah*	5	48.1	DYSETHYSPDGR		3	2	713.80	28.77
	VM1N_BOTPA	*Bothrops pauloensis*	6	22.5	AYLGSMCNPK	1×Carbamidomethyl [C7]; 1×Deamidated [N8]	2	2	571.26	30.26
	VM3VA_MACLB	*Macrovipera lebetina*	10	68.7	IPCAPQDVK	1×Carbamidomethyl [C3]	2	2	514.27	29.55

					LTPGSQCADGECCDQCK	4×Carbamidomethyl [C7; C12; C13; C16]	2	2	993.38	28.27
					NPCQIYYTPSDENKGMVDPGTK	1×Carbamidomethyl [C3]; 1×Deamidated [N13]	3	3	839.04	31.36
					NPCQIYYTPSDENKGMVDPGTK	1×Carbamidomethyl [C3]; 1×Deamidated [N13]; 1×Oxidation [M16]	2	3	844.37	30.62
SVSP	D5KRX9_ECHOC	*Echis ocellatus*	8	29.0	DPCAYSFSPGMYTK	1×Carbamidomethyl [C3]; 1×Oxidation [M11]	2	2	820.34	32.55
	E9JG21_ECHCO	*Echis coloratus*	3	26.4	LYNYSVCR	1×Carbamidomethyl [C7]; 1×Deamidated [N3]	3	2	538.25	30.55
VEGF	TXVE_BITAR	*Bitis arietans*	38	16.6	CGGCCSDESLTCTSVGER	4×Carbamidomethyl [C1; C4; C5; C12]	6	2	1017.89	29.93
					ETLVSILEEYPDK		2	2	768.39	35.48
					IFRPSCVAVLR	1×Carbamidomethyl [C6]	9	3	439.92	32.07
					TVELQVMQVTPK	1×Oxidation [M7]	13	2	694.88	33.19
					TVELQVMQVTPK		16	2	686.88	33.28

Table B5: Fraction E peptide hits

FAMILY	ACCESSION CODE	HOMOLOGY	COVERAGE (%)	MW (kDa)	SEQUENCE	MODIFICATIONS	# PSMS	*Z*	m/z (Da)	RT (min)
CRISP	F2Q6E5_CROHD	*Crotalus horridus*	21	24.7	MEWYPEAAANAER		3	2	769.34	31.15
	F2Q6F2_DABRR	*Daboia russelii*	12	26.7	MEWYPEAAANAER		3	2	769.34	31.15
CTL	A0A2D4F193_MICCO	*Micrurus corallinus*	5	15.9	YNGHCYK	1×Carbamidomethyl [C5]; 1×Deamidated [N2]	2	3	314.80	25.69
CYS	CYT_BITAR	*Bitis arietans*	35	12.7	IPGGLSPR		3	2	398.74	28.64
					VVEAQSQVVSGVK	1×Deamidated [Q]	2	2	665.87	28.71
					VVEAQSQVVSGVK		3	2	665.37	28.37
DIS	DID2_BITGA	*Bitis gabonica*	24	13.8	RGEHCISGPCCR	3×Carbamidomethyl [C5; C10; C11]	2	2	744.82	26.19
	DID5A_ECHOC	*Echis ocellatus*	9	12.4	FLNSGTICK	1×Carbamidomethyl [C8]; 1×Deamidated [N3]	2	2	520.76	29.53
	DID8A_CERCE	*Cerastes cerastes*	43	7.1	MNSAHPCCDPVTCKPK	3×Carbamidomethyl [C7; C8; C13]; 1×Oxidation [M1]	2	3	639.94	26.56
					RGEHCISGPCCR	3×Carbamidomethyl [C5; C10; C11]	2	2	744.82	26.19
	DIDB_CERVI	*Cerastes vipera*	42	7.0	NSAHPCCDPVTCKPK	3×Carbamidomethyl [C6; C7; C12]	2	4	443.45	26.44
					RGEHCISGPCCR	3×Carbamidomethyl [C5; C10; C11]	2	2	744.82	26.19
					ILEQGEDCDCGSPANCQDR	3×Carbamidomethyl [C8; C10; C16]	3	2	1112.44	27.92
	Q4JCR9_BITAR	*Bitis arietans*	46	11.9	SSDCPWNH	1×Carbamidomethyl [C4]	8	2	501.69	29.72
					SSDCPWNH	1×Carbamidomethyl [C4]; 1×Deamidated [N7]	3	2	502.18	28.61
	VM2_BITAR	*Bitis arietans*	65	9.0	CCNAATCK	3×Carbamidomethyl [C1; C2; C7]	18	2	492.69	11.59

					CCNAATCK	3×Carbamidomethyl [C1; C2; C7]; 1×Deamidated [N3]	29	2	493.18	25.24
					ILEQGEDCDCGSPANCQDR	3×Carbamidomethyl [C8; C10; C16]	3	2	1112.44	27.92
					SPPVCGNK	1×Carbamidomethyl [C5]	7	2	429.71	25.98
					SPPVCGNK	1×Carbamidomethyl [C5]; 1×Deamidated [N7]	2	2	430.20	25.81
					SSDCPWNH	1×Carbamidomethyl [C4]	8	2	501.69	29.72
					SSDCPWNH	1×Carbamidomethyl [C4]; 1×Deamidated [N7]	3	2	502.18	28.61
	VM2D3_BITAR	*Bitis arietans*	86	9.1	CCNAATCK	3×Carbamidomethyl [C1; C2; C7]	18	2	492.69	11.59
					CCNAATCK	3×Carbamidomethyl [C1; C2; C7]; 1×Deamidated [N3]	29	2	493.18	25.24
					SPPVCGNELLEEGEECDCGSPANCQDR	4×Carbamidomethyl [C5; C16; C18; C24]	3	3	1027.08	30.68
					SSDCPWNH	1×Carbamidomethyl [C4]	8	2	501.69	29.72
					SSDCPWNH	1×Carbamidomethyl [C4]; 1×Deamidated [N7]	3	2	502.18	28.61
	VM2TI_ERIMA	*Eristicophis macmahoni*	22	5.8	GDWNNDYCTGK	1×Carbamidomethyl [C8]; 1×Deamidated [N5]	2	2	665.76	28.76
NGF	Q9PS29_BITAR	*Bitis arietans*	52	2.5	EVRPFMEVYQR		3	3	485.25	30.41
SVMP	A0A0B4U9L8_VIPAA	*Vipera ammodytes*	4	68.7	NPCCDAATCK	3×Carbamidomethyl [C3; C4; C9]	2	2	598.73	25.79
	A0A1W7RJU5_AGKCO	*Agkistrodon contortrix*	1	68.1	CCDAATCK	3×Carbamidomethyl [C1; C2; C7]	18	2	493.18	19.65
	Q4QT06_BITAR	*Bitis arietans*	3	57.4	FLVDHKPK		2	3	328.53	27.04

	VM3E1_ECHOC	*Echis ocellatus*	3	68.7	LYCFDNLPEHK	1×Carbamidomethyl [C3]	4	3	479.23	30.17
	VM3VA_MACLB	*Macrovipera lebetina*	7	68.7	IPCAPQDVK	1×Carbamidomethyl [C3]	3	2	514.27	28.11
					LYCFDNLPEHK	1×Carbamidomethyl [C3]	4	3	479.23	30.17
VEGF	TXVE_BITAR	*Bitis arietans*	46	16.6	IFRPSCVAVLR	1×Carbamidomethyl [C6]	3	3	439.92	30.37

Table B6: Protein hits in Fraction A cross-linked 120 kDa dimer

FAMILY	ACCESSION CODE	HOMOLOGY	COVERAGE (%)	MW (kDa)	SEQUENCE	MODIFICATIONS	# PSMS	*Z*	m/z (Da)	RT (min)
SVMP	VM3M1_NAJMO	*Naja mossambica*	2	68.1	DPNYGMVAPGTK	1×Oxidation [M6]	2	3	422.53	27.84
					DPNYGMVAPGTK	1×Oxidation [M6]; 1×DSSO Amidated [K12]	1	2	622.29	27.81
					DPNYGMVAPGTK	1×Deamidated [N3]; 1×Oxidation [M6]	1	2	633.79	28.25
CTL	SL2_BITGA	*Bitis gabonica*	10	18.1	EEADFVAQLISDNIK		2	3	564.62	33.87
PDE	PDE_MACLB	*Macrovipera lebetina*	1	96.1	LNLNNQAK		1	2	457.76	27.33

Table B7: Protein hits in Fraction B cross-linked 30 kDa dimer

FAMILY	ACCESSION CODE	HOMOLOGY	COVERAGE (%)	MW (kDa)	SEQUENCE	MODIFICATIONS	# PSMS	*Z*	m/z (Da)	RT (min)
CTL	SLA_BITAR	*Bitis arietans*	12	14.9	EEADFVTKLASQTLTK		1	2	890.95	35.42
	SL1_BITAR	*Bitis arietans*	5	16.6	TWADAEK		1	2	410.70	28.63
	SLRB_BITRH	*Bitis rhinoceros*	5	17.7	TWADAEK		1	2	410.70	28.63
	SLB_BITAR	*Bitis arietans*	6	14.8	TWADAEK		1	2	410.70	28.63

Table B8: Protein hits in Fraction C cross-linked 30 kDa dimer

FAMILY	ACCESSION CODE	HOMOLOGY	COVERAGE (%)	MW (kDa)	SEQUENCE	MODIFICATIONS	# PSMS	*Z*	m/z (Da)	RT (min)
CTL	SL1_BITAR	*Bitis arietans*	5	16.6	TWADAEK		1	2	410.69	27.30
	SL3_BITGA	*Bitis gabonica*	4	18.1	TWEDAEK		1	2	410.69	27.30
	SLA_BITAR	*Bitis arietans*	6	14.9	LASQTLTK		1	2	431.25	27.07
	SLA1_MACLB	*Macrovipera lebetina*	4	17.8	TWEDAEK		1	2	410.69	27.30
	SLB_BITAR	*Bitis arietans*	6	14.8	TWADAEK		1	2	410.69	27.30
	SLRB_BITRH	*Bitis rhinoceros*	5	17.7	TWADAEK		1	2	410.69	27.30
CYS	CYT_BITAR	*Bitis arietans*	12	12.7	VVEAQSQVVSGVK		1	2	665.37	28.20
SVMP	VM3M1_NAJMO	*Naja mossambica*	2	68.1	DPNYGMVAPGTK	1×Oxidation [M6]	1	2	633.29	27.69
	VM1N_BOTPA	*Bothrops pauloensis*	5	22.5	AYLGSMCNPK	1×Carboxymethyl [C7]; 1×Oxidation [M6]	1	2	579.25	27.41
VEGF	TXVE_BITAR	*Bitis arietans*	8	16.6	TVELQVMQVTPK	1×Oxidation [M7]	1	2	694.88	30.04

Appendix C: Publication 1

Journal of Proteomics 216 (2020) 103680

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Interrogating the higher order structures of snake venom proteins using an integrated mass spectrometric approach

C. Ruth Wang[a], Emily R. Bubner[a], Blagojce Jovcevski[a], Parul Mittal[b], Tara L. Pukala[a,*]

[a] Department of Chemistry, School of Physical Sciences, University of Adelaide, Adelaide 5005, Australia
[b] Adelaide Proteomics Centre, University of Adelaide, Adelaide 5005, Australia

ARTICLE INFO

Keywords:
Snake venom
Native mass spectrometry
Proteomics
Protein complexes

ABSTRACT

Snake venoms contain complex mixtures of proteins vital for the survival of venomous snakes. Aligned with their diverse pharmacological activities, the protein compositions of snake venoms are highly variable, and efforts to characterise the primary structures of such proteins are ongoing. Additionally, a significant knowledge gap exists in terms of the higher-order protein structures which modulate venom potency, posing a challenge for successful therapeutic applications. Here we use a multifaceted mass spectrometry approach to characterise proteins from venoms of Collett's snake *Pseudechis colletti* and the puff adder *Bitis arietans*. Following chromatographic fractionation and bottom-up proteomics analysis, native mass spectrometry identified, among other components, a non-covalent L-amino acid oxidase dimer in the *P. colletti* venom and a C-type lectin tetramer in the *B. arietans* venom. Furthermore, a covalently-linked phospholipase A_2 (PLA_2) dimer was identified in *P. colletti* venom, from which the PLA_2 species were shown to adopt compact geometries using ion mobility measurements. Interestingly, we show that the dimeric PLA_2 possesses greater bioactivity than the monomeric PLA_2s. This work contributes to ongoing efforts cataloguing components of snake venoms, and notably, emphasises the importance of understanding higher-order venom protein interactions and the utility of a combined mass spectrometric approach for this task.

Significance: The protein constituents of snake venoms represent a sophisticated cocktail of biologically active molecules ideally suited for further exploration in drug design and development. Despite ongoing efforts to characterise the diverse protein components of such venoms there is still much work required in this area, particularly in moving from simply describing the protein primary sequence to providing an understanding of quaternary structure. The combined proteomic and native mass spectrometry workflow utilised here gives new insights into higher order protein structures in selected snake venoms, and can underpin further investigation into the protein interactions which govern snake venom specificity and potency.

1. Introduction

Snake venoms contain a complex and widely underexplored variety of biologically active proteins that serve as the animal's primary hunting tool, facilitating the immobilisation, killing and digestion of prey [1–7]. The remarkable pharmacological effects of these proteins translate well to a therapeutic context, with analgesic, antimicrobial, antitumor, anti- and pro-coagulant bioactivities, among others, being extensively described for selected snake venom proteins [1,8]. In addition, the notable potency, high specificity and poor immunogenicity of venom components are attractive traits for drug design [1,5]. All these appealing properties suggest snake venoms to be promising sources for therapeutic development.

Snake venom proteins can be generally categorised into two classes: enzymatic toxins and non-enzymatic toxins. Enzymatic constituents often include toxin superfamilies such as phospholipase A_2s (PLA_2s),

Abbreviations: 3FTx, 3-finger toxin; 5′NUC, 5′ nucleotidase; AChE, acetylcholinesterase; BA, *Bitis arietans*; BPP, bradykinin-potentiating peptide; CCS, collision cross-section; CRISP, cysteine-rich secretory protein; CTL, C-type lectin; CYS, cystatin; DIS, disintegrin; FASP, filter aided sample preparation; IM, ion mobility; KUN, kunitz-type serine protease inhibitors; LAAO, L-amino acid oxidases; LPC, lyso-phosphatidylcholine; MS, mass spectrometry; NGF, nerve growth factor; NP, natriuretic peptide; PC, *Pseudechis colletti*; PDE, phosphodiesterase; PI, proteinase inhibitor; PLA_2, phospholipase A_2; PLB, phospholipase B; POPC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; PTM, post-translational modification; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; SEC, size exclusion chromatography; SVMP, snake venom metalloproteinase; SVSP, snake venom serine proteases; VEGF, vascular endothelial growth factor

* Corresponding author.
E-mail address: tara.pukala@adelaide.edu.au (T.L. Pukala).

https://doi.org/10.1016/j.jprot.2020.103680
Received 29 November 2019; Received in revised form 20 January 2020; Accepted 2 February 2020
Available online 03 February 2020
1874-3919/

snake venom serine proteases (SVSPs), metalloproteinases (SVMPs), L-amino acid oxidases (LAAOs), acetylcholinesterases (AChE), and various nucleotidases [2]. They are generally known to disrupt cellular pathways involved in haemostasis, tissue necrosis, and myotoxicity, and are often responsible for the lethal effects of venom [5,7]. In contrast, non-enzymatic toxins are thought to be primarily responsible for immobilisation of prey by interference with cardiovascular and neuromuscular systems [6,9], and include a diverse range of prominent superfamilies such as 3-finger toxins (3FTxs), C-type lectins (CTLs), proteinase inhibitors (PIs), nerve growth factors (NGFs), natriuretic peptides (NPs), bradykinin-potentiating peptides (BPPs), cysteine-rich secretory proteins (CRISPs), vascular endothelial growth factors (VEGFs), and disintegrins (DIS).

The advent of 'omics' technologies in recent years has enabled high-throughput identification of various proteins to afford a richer understanding of venom proteomes [1,10]. There remains, nonetheless, an immeasurable array of proteins yet to be characterised along with their wide spectrum of post-translational modifications (PTMs), to enable better understanding of venom diversity among the many venomous species of snakes, coupled to the possible ecological factors which drive variability. Furthermore, while many snake toxins exhibit pharmacological activity on their own, some are known to form higher order assemblies whereby the synergistic interactions between proteins may enhance the potency of the venom [5,11]. Therefore, characterisation of the quaternary structures of venom proteins is essential to establish a thorough understanding of their interactions and potential mechanism of bioactivity. While studies have demonstrated successful characterisation of venom protein complexes using NMR spectroscopy and X-ray crystallography, challenges remain in capturing these interactions in a high-throughput manner, given the heterogeneous nature of many of these complexes and the expansive catalogue of snake venom proteins [5,12,13].

Mass spectrometry (MS) has emerged as powerful analytical tool enabling both proteomic sequencing and higher-order protein structure determination. A growing body of literature has demonstrated the ability of native MS, particularly in conjunction with ion mobility (IM) measurements, to reveal details of the composition and topology of protein assemblies [14–16]. Importantly, these MS-based methods are rapid, highly sensitive, and capable of interrogating heterogeneous protein samples, advantages well suited for the analysis of snake venoms. While bottom-up proteomics has been critical in cataloguing the vast array of proteins in venoms, native MS is comparably underutilised, with only a few venom protein assemblies characterised so far. Textilotoxin from *Pseudonaja textilis*, Paradoxin from *Oxyuranus microlepidotus* and taipoxin from *Oxyuranus scutellatus* are prominent examples [17,18].

Here we describe the characterisation of various venom proteins from two geographically and phylogenetically variable species that are of medical significance; an Australian elapid representative, *Pseudechis colletti*, and the African viperid *Bitis arietans*. The *Pseudechis* genus (black snakes) is considered to be one of the most venomous Australian snake genera, and despite the placid disposition of *P. colletti* (Collett's snake) resulting in fewer reported cases of envenomation, *P. colletti* venom is highly toxic and arguably one of the most underexplored of the *Pseudechis* genus [19]. Native to widespread regions of sub-Saharan Africa, *B. arietans* (puff adder) is also regarded as an extremely venomous species, and is responsible for the majority of snakebite fatalities in Africa [20].

For both species, endeavours to catalogue the venom proteome have commenced using proteomic techniques [21–24]. There have also been various biochemical and immunological studies to characterise the activities of individual venom proteins [20,25–27], and the three-dimensional structures corresponding to some of these protein families have been described [28,29]. However, the formation and structure of any higher order protein assemblies still remain largely unexplored. The multi-faceted MS based approach applied here not only identifies new protein complexes in these venoms for the first time, but provides a strong foundation for further understanding of venom components from a structural and functional perspective.

2. Material and methods

2.1. Materials

All reagents were purchased from Sigma Aldrich unless specified otherwise.

Whole lyophilised *P. colletti* and *B. arietans* venoms were purchased from Venom Supplies Pty. Ltd. (Tanunda, Australia) and were stored at −20 °C until required.

2.2. Chromatographic fractionation

Venom proteins were first fractionated by size exclusion chromatography (SEC). Lyophilised venom was reconstituted in 200 mM ammonium acetate (NH_4OAc, pH 7.0) at a concentration of 10 mg/mL and loaded onto a Superdex200 10/300 size-exclusion column (GE Healthcare) coupled to an ÄKTA Pure FPLC system (GE Healthcare). The column was equilibrated with 200 mM NH_4OAc (pH 7.0) prior to sample loading (400 μL). Fractions (400 μL) were collected at a flow rate of 0.4 mL/min with 200 mM NH_4OAc as the eluent over a volume of 36 mL. For further fractionation by ion exchange chromatography (IEC), selected SEC fractions of interest were pooled, freeze-dried and reconstituted in Buffer A (20 mM Tris-HCl, 1 mM EDTA, 0.02% NaN_3 in milliQ water). The sample was loaded onto a MonoQ 5/50 GL anion-exchange column (GE Healthcare) coupled to an ÄKTA Pure FPLC system (GE Healthcare). The column was equilibrated with Buffer A at a flow rate of 1 mL/min and a linear gradient to 50% Buffer B (Buffer A containing 2 M NaCl) was applied over 20 mL, followed by a 15 mL elution with Buffer B. Fractions (2 mL) were collected and buffer-exchanged into 200 mM NH_4OAc (pH 7.0) using Vivaspin 2 centrifugation units (GE Healthcare) with a molecular weight cut-off of 3 kDa. Samples were stored at −20 °C until required for analysis.

2.3. 1D SDS-polyacrylamide gel electrophoresis

Venom fractions of interest were added in a 1:1 *v*/v ratio to 3× reducing sample buffer (150 mM Tris-HCl, 300 mM DTT, 6% SDS, 30% glycerol, 0.3% bromophenol blue, pH 6.8) and denatured at 95 °C for 15 min prior to loading onto a 4–15% Mini-Protean TGX Tris-HCl polyacrylamide gel (Bio-Rad). Gel electrophoresis was performed at 140 V and 400 mA for 1 h in 1× SDS tris-glycine running buffer. Precision Plus Protein Dual Colour standards (Bio-Rad) were used as molecular weight markers. SDS-PAGE gels were visualised by Coomassie Brilliant Blue staining (Coomassie Brilliant Blue R250 dye, 10% (*v*/v) glacial acetic acid, 40% (v/v) methanol) prior to imaging using an Imagescanner densitometer (Amersham Biosciences).

2.4. Filter-aided, in-solution tryptic digestion

Venoms were digested using a Filter Aided Sample Preparation (FASP) method [30] in Amicon Ultra-0.5 mL centrifugal filter units (Merck Millipore) with a 10 kDa molecular weight cut-off. Venom (approximately 0.1 mg) in 200 μL of 7 M urea/100 mM ammonium bicarbonate was incubated with 50 mM dithiothreitol for 1 h at room temperature, and further incubated with 55 mM iodoacetamide for 20 min in darkness. Promega MS grade trypsin (ThermoFisher Scientific), resuspended at 100 ng/μL in 10 mM ammonium bicarbonate, was added to the sample at a mass ratio of 1:50 (enzyme:protein), and the sample was incubated at 37 °C overnight. Digested peptides were then eluted through the spin-filter, collected and dried using vacuum centrifugation, before being reconstituted in 100 μL of 2% (*v*/v) acetonitrile 0.1% (v/v) formic acid. The sample was then purified with a C_{18}

Biospin column (ThermoFisher Scientific), and concentrations were verified on a NanoDrop 2000/2000c UV–Vis spectrophotometer at a wavelength of 205 nm using an extinction coefficient, ε_{205}, of 31 mL mg^{-1} cm^{-1} [31]. Samples were stored at −20 °C until required for LC-MS/MS analysis.

2.5. LC-MS/MS analysis of venom samples

Digested venom samples were analysed using an Ultimate 3000 nano-flow system (ThermoFisher Scientific) coupled to a LTQ XL Orbitrap ETD mass spectrometer (ThermoFisher Scientific). Approximately 2 μg of each peptide sample was pre-concentrated on a C_{18} trapping column (Acclaim PepMap 100 C_{18} 75 μm × 20 mm, Thermo-Fisher Scientific) at a flow rate of 5 μL/min using 2% (*v*/v) acetonitrile 0.1% (v/v) trifluoroacetic acid over 10 min. Peptides were then separated using a 75 μm ID C_{18} column (Acclaim PepMap100 C_{18} 75 μm × 50 cm, Thermo-Fisher Scientific) at a flow rate of 0.3 μL/min, where a linear gradient of 5% to 45% solvent B was applied over 60 min. This was followed by a 5 min wash with 90% B, and then a 15 min equilibration process with 5% B. (Solvent A: 2% (*v*/v) acetonitrile 0.1% (v/v) formic acid. Solvent B: 80% (v/v) acetonitrile 0.1% (v/v) formic acid). LC-MS/MS acquisitions were controlled by Xcalibur version 2.1 (ThermoFisher Scientific), and the mass spectrometer was operated in data-dependent mode. Spectra were acquired in positive mode over the mass range of 300–2000 *m/z* at a resolution of 60,000 in FT mode. The 10 most intense precursor ions were selected for CID fragmentation using a dynamic exclusion of 5 s where the dynamic exclusion criteria included a minimum relative signal intensity of 1000 and ≥2 positive charge state. The isolation width used was 3.0 *m/z* and a normalised collision energy of 35 was applied. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [32] with the dataset identifier PXD016503.

2.6. MASCOT analysis

MS/MS data was converted to MGF file format and submitted for qualitative protein identification on the in-house Mascot server (v2.3.01, Matrix Science). The data was searched against all *Chordata* entries present in the Swiss-Prot database. Parameters for the performed search were as follows: tryptic peptides with a maximum of 2 missed cleavages were allowed, peptide mass tolerance of 30 ppm, fragment mass tolerance of 0.8 Da, cysteine carbamidomethylation set as fixed modification and methionine oxidation, acetylation of the protein N-terminus, and deamidation of glutamine and asparagine set as variable modifications. The significance threshold was set as *P*-value < 0.05 and false positives were identified using a False Discovery Rate (FDR) of 0.05%. Any false positives, contaminants and proteins that did not possess a toxic function were eliminated from further analysis. Relative abundance values of the toxin families in a given venom mixture were calculated from the sum of the emPAI scores generated by Mascot for the given toxin family.

2.7. Native ion mobility-mass spectrometry

All native MS and ion mobility (IM) data were obtained using a Synapt G1 HDMS quadrupole ion mobility time-of-flight mass spectrometer (Waters). Protein samples were buffer-exchanged into 200 mM NH_4OAc buffer prior to MS analysis. Sample (4 μL) was introduced into the instrument by nano-ESI using platinum-coated borosilicate capillaries that were prepared in-house. The instrument conditions were set to preserve non-covalent interactions as follows: *m/z* range, 500–6000; polarity, positive; capillary voltage, 1.5 kV; sample cone voltage, 50 kV; extraction cone, 3 kV; source temperature, 50 °C; desolvation temperature; 180 °C; trap collision energy, 30 V; transfer collision energy, 30 V; IMS wave velocity, 300 m/s; IMS wave height, 7 V; IMS gas flow, 28 mL/min; backing pressure, 4.07 mbar. All native mass spectra were analysed using MassLynx (v4.1, Waters) and IM data was analysed using Driftscope (v2.1, Waters). Drift times for the identified charge states of interest were extracted from IM heatmaps and collision cross-section (CCS) values were determined as described previously [15], using equine cytochrome C and equine myoglobin as calibrants.

2.8. Intact protein mass spectrometry

Denatured protein mass spectra were obtained using a 1260 LC system coupled to a 6230 TOF mass spectrometer (Agilent Technologies). Sample (2 μL) was directly injected in 50% (*v*/v) aqueous acetonitrile containing 0.1% (v/v) formic acid and introduced into the instrument by ESI-MS at a flow rate of 0.2 mL/min without chromatographic separation. The instrument conditions were set as follows: *m/z* range, 500–3200; polarity, positive; capillary voltage, 3.5 kV; nozzle voltage, 2 kV; gas temperature; 325 °C. Mass spectra acquisition and data analysis were performed using MassHunter software (vB.08.00, Agilent Technologies). For MS experiments under both reducing and denatured conditions, the sample was first incubated with 1 mM DTT in 200 mM NH_4OAc for 1 h prior to MS analysis by the workflow described above.

2.9. PLA_2 enzymatic assay

This assay was adapted from [33] with minor modifications. *P. colletti* dimeric and monomeric PLA_2 samples were concentrated using Vivaspin 2 centrifugation units (GE Healthcare) with a molecular weight cut-off of 3 kDa, and protein concentration was determined using a bicinchoninic acid (BCA) assay (ThermoFisher Scientific). PLA_2 samples were prepared to 1 μM total protein in 200 mM NH_4OAc (pH 7.0). A 9 μM stock solution of lipid substrate (POPC 16:0/18:1) (Avanti Polar Lipids) was prepared in 5 mM NH_4OAc (pH 7.0) in methanol. PLA_2 (30 μL) was added to the lipid substrate (30 μL), followed by mixing by pipetting and a 3 min incubation at room temperature.

All samples were analysed using an Agilent 1260 LC system coupled to an Agilent 6230 TOF mass spectrometer (Agilent Technologies) tuned over a 1700 *m/z* mass range. 5 μL of sample was directly injected into the instrument via the LC auto sampler and eluted at a flow rate of 1 mL/min without chromatographic separation. A linear gradient of 70% to 100% Solvent B was applied over 1 min and maintained for 2 min to ensure elution of both POPC substrate and LPC products; Solvent A (0.1% (*v*/v) FA in water) and Solvent B (99.9% (v/v) ACN 0.1% (v/v) FA). The reaction was monitored every 3 min for a 15 min duration. MS instrument conditions were set as follows: *m/z* range, 100–1700; polarity, positive; capillary voltage, 1.2 kV; nozzle voltage, 1 kV; gas temperature; 325 °C. Mass spectra acquisition and data analysis were performed using MassHunter software (vB.08.00, Agilent Technologies). Spectra were summed over the time period of sample elution and the error tolerance was set to 60 ppm. The extracted ion count intensities of intact POPC substrate (760.6 *m/z*), LPC 18:1 (522.3 *m/z*), and LPC 16:0 (496.3 *m/z*) product ions along with the corresponding adducted ions were summed at a given time-point and the assay was performed in technical triplicate.

3. Results and discussion

3.1. Fractionation of P. colletti and B. arietans whole venoms

Fractionation of crude whole venom from *P. colletti* and *B. arietans* by size-exclusion chromatography (SEC) was first performed in order to obtain a broad view of the respective venom complexities (Fig. 1A). The proteins were eluted in ammonium acetate buffer which is suitable to maintain the proteins and any potential non-covalent complexes in their native state, and is compatible with downstream MS analysis [34,35]. SEC peaks are labelled numerically along the elution profiles

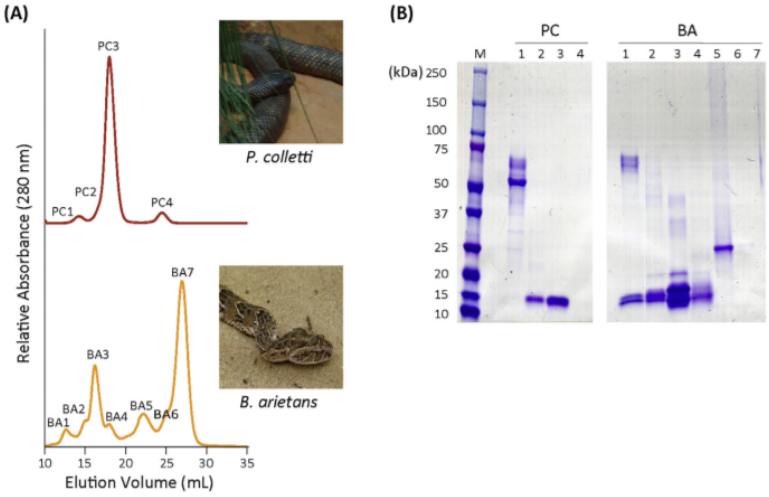


Fig. 1. (A) Size exclusion chromatography elution profiles of whole venoms from *P. colletti* and *B. arietans*. **(B)** SDS-PAGE of *P. colletti* and *B. arietans* venom SEC fractions. Image attributions: *"Pseudechis colletti"* by Taipan198, licensed under CC BY-SA 3.0. *"Young Puff Adder (Bitis arietans)"* by Bernard Dupont, licensed under CC BY-SA 2.0.

from *P. colletti* (PC) and *B. arietans* (BA) venom for ease of reference. Based on evaluation of the SEC profile, *P. colletti* venom appears as the simplest venom with only three main peaks corresponding to proteins that were eluted in the high, intermediate, and low mass ranges, with an abundance of intermediate-sized proteins. In contrast, the SEC trace from *B. arietans* venom is significantly more complex with numerous peaks across the elution range, inferring a diverse suite of proteins of various sizes. Of note, large and very small (potentially peptidic) species appeared particularly abundant (fractions BA3 and BA7 respectively), the former of which in the high mass range fractions may also encompass assemblies constituted from smaller proteins.

In order to further separate the protein components in the venom fractions as well as begin probing potential non-covalent complexes present, reducing SDS-PAGE was conducted for the various venom SEC fractions across the entire elution range (Fig. 1B). Again, *P. colletti* venom appears relatively simple, with only larger (50–70 kDa) and smaller (15 kDa) proteins identified. Similarly, consistent with the SEC analysis, *B. arietans* venom contains a greater diversity of proteins across the range from approximately 15–70 kDa. Notably, many of the smaller protein species around 15 kDa were found in high mass range SEC fractions (such as fractions BA1- BA4) giving a first indication of protein complexes in the venom that would be disrupted to monomers by the reducing and denaturing conditions of the SDS-PAGE analysis.

3.2. Proteomic identification of venom components

To catalogue the proteomic composition of the *P. colletti* and *B. arietans* venoms, whole venoms were analysed by LC-MS/MS using a bottom-up workflow. Identification was performed by searching against the *Chordata* database, and proteins that did not possess a toxic function were eliminated from further analysis. Identified proteins are given in supplementary data, Tables S1 and S2. The relative abundance of the different toxin families within a whole venom mixture was determined from the Exponentially Modified Protein Abundance Index (emPAI) scores generated by Mascot, which is a label-free quantitative estimation for relative protein abundance within the sample based on spectral counting [36]. The sum of the emPAI scores for protein hits from the same family, for instance the PLA_2 superfamily, were then used to construct the relative abundance plots shown in Fig. 2.

As shown in Fig. 2, significant proteomic diversity based on the identification of distinct toxin superfamilies was observed across the two different venoms. *P. colletti* venom appears as the least diverse with only four protein families identified, and demonstrates a strikingly high proportion of PLA_2 proteins. *B. arietans* venom however displays a substantially more varied proteome with 12 distinct protein families identified, including highly abundant protein families such as DIS, CTL, 3FTx, and VEGF. The intrinsic proteomic profiles may be a good reflection of the phylogenetic diversity of these species of snakes, and are largely in agreement with what is known in literature about the types of proteins in these venoms [22,23] and from symptomatic effects of envenomation [19,20,25,27]. However, the relatively strong presence of 3FTx in *B. arietans* venom is interesting as previous studies have not observed this, but rather higher abundance of SVMPs and SVSPs [20,27], reiterating the need for ongoing proteomic characterisation efforts given the intrinsic intra-species variation in venom proteomes. It is also noteworthy that while some proteomic studies have been performed for these species in the past, to our knowledge such a high-throughput shotgun proteomics approach is reported here for the first time for these venoms.

SEC fractions corresponding to the 14 mL and 18 mL elution volumes were selected from both venoms as an example of high and intermediate molecular weight fractions for further characterisation of potential protein complexes. Prior to interrogation by native MS however, it is necessary to have confidence in the identity of proteins found specifically in these fractions. Therefore, these high and intermediate molecular weight fractions were selectively analysed by bottom-up proteomics, and proteins were identified and their relative abundances determined by Mascot-generated emPAI scores in the same manner as those in the proteomic analysis of the whole venoms (Fig. S1, Tables S3–S6). From this data, a variety of protein families were identified across the four venom fractions. Proteomic analysis of the *P. colletti* venom fractions revealed LAAO is the predominant protein family in fraction PC1 and PLA_2 is the most abundant family noted in fraction PC3. In terms of the predominant protein families in the fractions of *B. arietans* venom, CTL and DIS are the most abundant families in Fraction BA2 whereas DIS appear to be the most abundant species in Fraction BA4.

While some protein families of low abundance are noted here that were not observed in the whole venoms, this is most likely due to the fact that further fractionation of the whole venoms has increased analytical capacity, allowing identification of more proteins. Nonetheless, the number of protein groups and protein hits identified in this work appear comparable to those reported in other snake venom proteomic studies that have utilised comparable venom protein concentrations and a bottom-up proteomic approach for similar species of snakes [23,24]. Furthermore, despite the diversity of snake venom proteins,

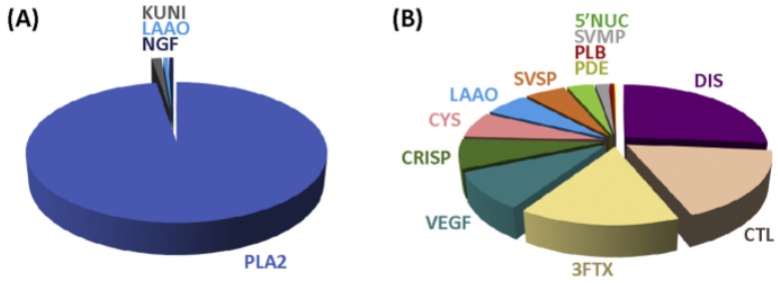


Fig. 2. Proteomic composition of whole venoms from **(A)** *P. colletti* and **(B)** *B. arietans* based on the estimated relative abundances of the toxin families within each venom. Abbreviations for the toxin families are as follows: phospholipase A_2 (PLA_2), 3-finger toxin (3FTx), snake venom metalloproteinase (SVMP), cysteine-rich secretory protein (CRISP), L-amino acid oxidase (LAAO), 5'nucleotidase (5'NUC), phosphodiesterase (PDE), kunitz-type serine protease inhibitors (KUN), snake venom serine protease (SVSP), phospholipase B (PLB), nerve growth factor (NGF), cystatin (CYS), C-type lectin (CTL), vascular endothelial growth factor (VEGF) and disintegrin (DIS).

this analysis is consistent with reports that more than 90% of elapid and viper venoms are composed of just ten protein families, and that the abundant protein families typically constitute greater than 90% of the total venom composition with 3–4 orders of magnitude in dynamic range between rare and abundant proteins [37]. This does imply some limitations in the shotgun proteomics approach and suggests fractionation may be necessary for future studies where wide proteome coverage is required. Importantly however, the results from the proteomic analysis of these SEC fractions can supplement the assignment of protein identities for higher-order structural analysis.

3.3. Native MS analysis of SEC fractions

Native MS was used to interrogate the potential higher-order structural interactions present in the four selected venom fractions, whereby the fractionated protein samples were infused directly into the mass spectrometer under gentle ionization conditions known to preserve intact protein assemblies [34,35]. Fig. 3 shows the resulting native mass spectra for the four fractions analysed.

Interrogation of the venom fractions that were eluted in the intermediate mass range (SEC elution volumes of 18 mL) by native MS showed that the proteins being identified are principally monomeric (Fig. 3A and B), and in agreement with the SDS-PAGE results that showed protein bands corresponding to masses of 15 kDa or lower (Fig. 1B). For *P. colletti* venom, fraction PC3 showed predominant protein species on the order of 13 kDa (Fig. 4A) is assigned to PLA_2 enzymes based on the proteomic analysis of the fraction (discussed further below). The predominant 9.0 kDa protein species identified in the *B. arietans* venom fraction BA4 (Fig. 4B) correspond to monomeric DIS components which are in line with the proteomic results for this fraction where DIS is the most abundant constituent.

Native mass spectra for the two high mass range fractions eluting at 14 mL by SEC (Fig. 3C and D) reveal distinct protein populations, all of which appear to display some form of PTM, most likely glycosylation, as suggested by the broadness of the peaks [17,18]. Two main protein species can be identified in the native mass spectrum of fraction PC1 from *P. colletti* venom, with molecular masses of 64 kDa and 117 kDa. Based on the measured molecular weights, the lack of a protein band corresponding to 117 kDa in the SDS-PAGE analysis and the predominance of LAAO identified in this fraction by shotgun proteomics, these species are assigned as monomeric and dimeric LAAOs respectively. Similarly, a 60 kDa protein species is observed in the *B. arietans*

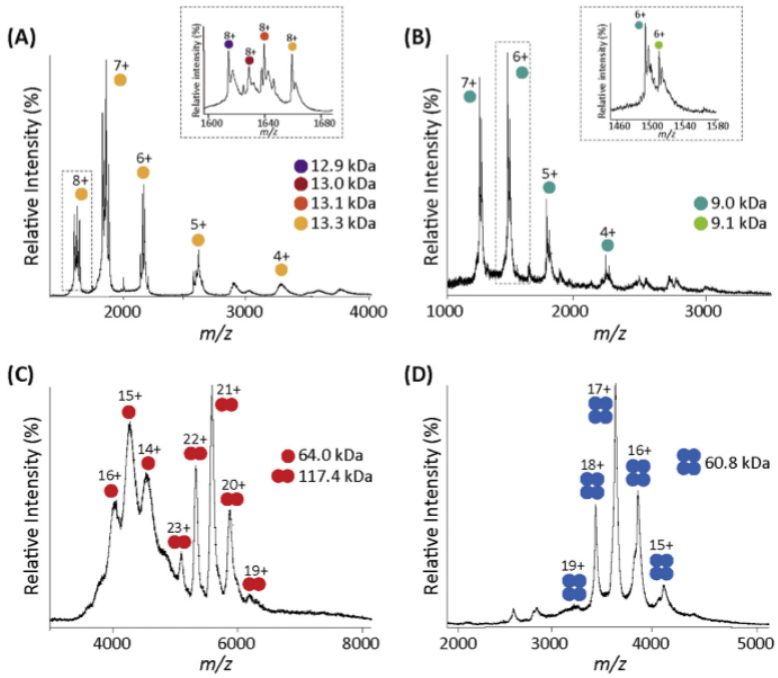


Fig. 3. Native mass spectra of selected *P. colletti* and *B. arietans* venom fractions. SEC fractions **(A)** PC3 (*P. colletti*) and **(B)** BA4 (*B. arietans*) were obtained at 18 mL elution volume. SEC fractions **(C)** Fraction PC1 (*P. colletti*) and **(D)** BA2 (*B. arietans*) were obtained at 14 mL elution volume. Proteins were maintained in 200 mM ammonium acetate (pH 7.0) for nanoESI-MS analysis. Different protein species identified with their various charge states and oligomeric states are labelled with coloured circles, and insets show zoomed in regions of the spectra as indicated by dashed boxes.

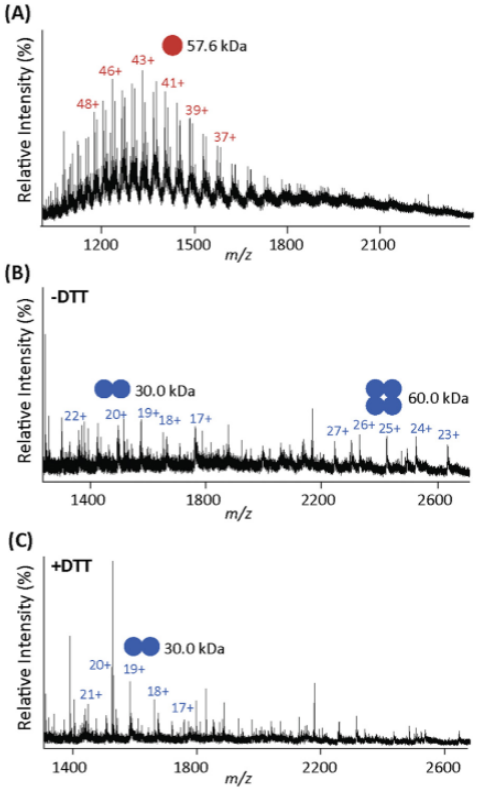


Fig. 4. Denatured MS analysis of the high mass range SEC venom fractions: **(A)** Fraction PC1 (*P. colletti*), **(B)** Fraction BA2 (*B. arietans*), and **(C)** Fraction BA2 (*B. arietans*) treated with 1 mM dithiothreitol. Proteins were diluted with 50% ACN and 0.1% FA prior to ESI-MS analysis, and the different protein species identified with their various charge states and oligomeric states are labelled with coloured circles.

venom Fraction BA2. Given the single 15 kDa protein band observed by SDS-PAGE and the corresponding proteomic analysis for this fraction, this 60 kDa protein species assigned as a tetramer of a 15 kDa monomeric CTL, which are known to be capable of oligomerisation [5,27,38].

MS experiments were also performed under denaturing conditions for the venom fractions containing higher order assemblies in order to further probe the nature of the interactions between the protein subunits (Fig. 4). A 57.6 kDa protein species is predominant in the denatured mass spectrum for *P. colletti* venom fraction PC1 (Fig. 4A) along with various isoforms of similar mass. Importantly, the absence of the larger dimeric LAAO species in the denatured MS spectrum is a good indication that the species being observed is a non-covalent complex. Furthermore, homodimeric LAAOs are known to exist around the 110–150 kDa mass range, and that both covalent and non-covalent interactions between the subunits are possible [5]. 60 kDa and 30 kDa species were observed in the denatured MS spectrum of the *B. arietans* venom fraction BA2 (Fig. 4B).

Considering the 60 kDa CTL tetramer assigned in the corresponding native MS spectrum and the 15 kDa monomeric subunits noted in the reducing SDS-PAGE analysis, the species observed under these denaturing conditions in the gas-phase appear to be a covalently-linked CTL tetramer and a dissociated dimer. Further treatment of the 60 kDa tetramer with the reducing agent dithiothreitol and subsequent denatured MS analysis confirmed disulphide interactions between the protein subunits as only 30 kDa CTL dimers were observed, indicating dissociation of the tetramer (Fig. 5C). As monomeric CTL masses were not noted, stronger and irreversible reducing conditions may be considered in future experiments to further investigate the inter-subunit disulphide bonding. The stability of these CTL proteins is also inferred from observation of the average charge states under denaturing (and reducing conditions). Comparison of monomeric LLAO (57.6 kDa, Fig. 4A) and tetrameric CTL (60.0 kDa, Fig. 4B) species reveal vastly different charge state distributions centred around 43+ and 25+ charges respectively, despite similar molecular weights. The narrower charge state distribution with low average charge states observed for the CTL assemblies is consistent with a largely folded structure remaining [39], even in denaturing conditions, suggesting extensive disufide bonding stabilises the higher-order structure of these protein complexes.

3.4. Further characterisation of *P. colletti* PLA_2 proteins

It can be observed that the native MS spectra presented in Fig. 3 appear simpler in terms of the number of protein species observed in comparison to the various protein constituents identified in the proteomic analysis of the corresponding venom fractions, and consequently, the native MS analysis highlights only the most prominent protein species. Signal suppression of lower abundant protein ions may be contributing to the absence of certain protein families that were noted in the proteomic analyses. In addition, the intrinsic ionization efficiency of different proteins is also a considerable factor determining which proteins ions can be observed. We have therefore extended this analysis to a more focused investigation of the *P. colletti* PLA_2 proteins by further fractionation of the venom using anion-exchange chromatography.

As described above, *P. colletti* venom is highly abundant in PLA_2 proteins which elute in the primarily in the major fraction at approximately 18 mL during SEC (fraction PC3, Fig. 1A). Fractions corresponding to the PC3 peak from SEC separation of *P. colletti* whole venom were then pooled, and subsequent separation by anion exchange chromatography resulted in the elution of four peaks (designated PC3-A - PC3-D), corresponding principally to distinct PLA_2 isoforms (Supplementary Fig. S2). This data demonstrates the structural complexity of PLA_2 proteins in this seemingly simple venom mixture.

The quaternary structure of isolated PLA_2 proteins from ion exchange chromatography were interrogated by native IM-MS (Fig. 5 and Fig. S4). The majority of the PLA_2 species across the four fractions examined are monomeric at approximately 13 kDa, consistent with the native MS data presented in Fig. 3. The observed variation in masses is indicative of numerous PLA_2 isoforms which differ by amino acid variants or additionally due to PTMs. Previous studies have indicated glycosylation to be a common modification in snake venoms, with highly variable and often atypical glycosylation patterns, including sialic acid-containing glycans that fall within the reported mass range [40,41]. Interestingly, a 27.7 kDa species was also observed in fraction PC3-C which is indicative of dimeric PLA_2 based on its charge state distribution and molecular mass (Fig. 5C). The presence of this dimeric species is novel as there is no structural data reporting the existence of PLA_2 dimers in *P. colletti* venom, although the presence of such dimers has been briefly speculated in a previous study based on SDS-PAGE analysis [23]. To date therefore, this is the first observation of dimeric PLA_2 in a venom thought to contain monomeric PLA_2 and indicates that quaternary structure and dynamics may be at play in affecting PLA_2 function. Denaturing MS experiments reveal that the PLA_2 dimer is covalent in nature, with a 27.7 KDa species observed to dissociate to the monomeric 13.8 KDa subunits only upon reduction with dithiothreitol (Fig. S3).

While the native MS analysis show the various populations of PLA_2 isoforms, further structural insight into the overall conformation of

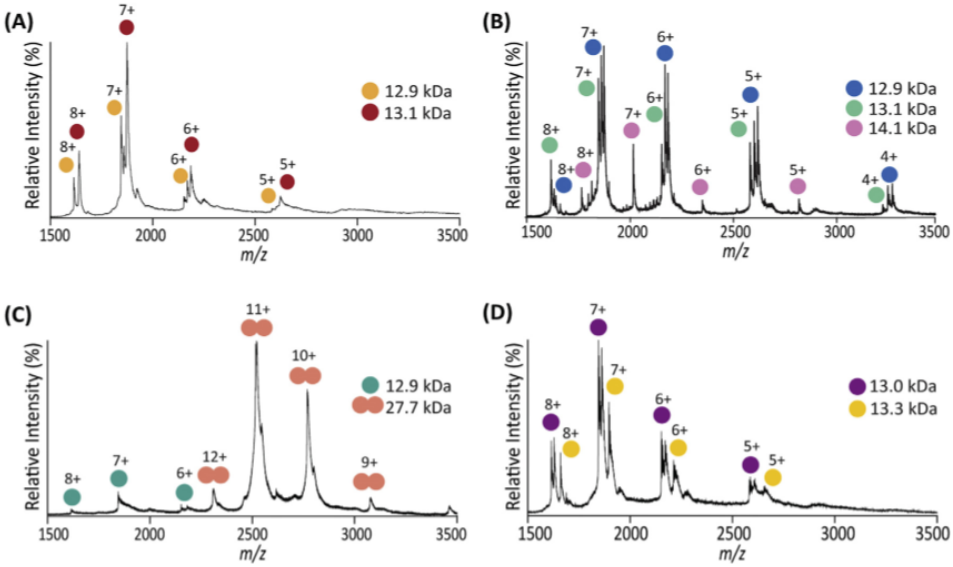


Fig. 5. Native MS reveals monomeric and dimeric *P. colletti* PLA_2s. Native mass spectra of PLA_2 isoforms (10 μM) from anion exchange chromatography **(A)** fraction PC3-A, **(B)** fraction PC3-B **(C)** fraction PC3-C, **(B)** fraction PC3-D. The various monomeric and dimeric PLA_2 species and their corresponding charge states are indicated.

Table 1
Collison cross section values of PLA_2 ions isolated from *P. colletti* venom. CCS measurements were made for all observed charge states for the proteins isolated from two selected anion exchange fractions shown in Fig. 5 (PC3-B and PC3-C).

Protein ID (Da)	*m/z* (charge)	Collision cross-section (Å^2)	Average effective protein density (D_{eff}) (g cm^{-1})
PC3-B PLA_2 monomer (14,097)	2821 (5)	987	0.88
	2350 (6)	1154	
	1874 (7)	1280	
	1763 (8)	1350	
PC3-B PLA_2 monomer (13,093)	3274 (4)	783	0.90
	2620 (5)	976	
	2169 (6)	1136	
	1871 (7)	1228	
	1637 (8)	1395	
PC3-B PLA_2 monomer (12,907)	3228 (4)	773	0.89
	2582 (5)	982	
	2152 (6)	1120	
	1844 (7)	1249	
	1614 (8)	1387	
PC3-C PLA_2 dimer (27,741)	3078 (9)	1629	0.85
	2771 (10)	1605	
	2521 (11)	1859	
	2310 (12)	2294	
PC3-C PLA_2 monomer (12,910)	2153 (6)	1120	0.72
	1845 (7)	1249	
	1617 (8)	1420	

these protein species can be gained by IM derived CCS measurements. We selected fractions PC3-B and PC3-C for further IM analysis in order to compare monomer and dimer rich fractions, and CCS values (derived from data presented in Fig. S4) were found to range from 773 to 2294 Å^2 for these ions (Table 1). Measured CCS values were subsequently used to determine the average effective density (D_{eff}) for each of the protein species, which infers a preliminary view of the structural geometry and the degree of sphericity for these proteins. CCS values in helium were approximated from the measured CCS values for the PLA_2 proteins (Table 1) using the method in [42], in order to determine the effective density of PLA_2 species (as detailed further in supporting information).

The calculated D_{eff} for the various PLA_2 species were relatively consistent (Table 1), ranging from 0.72 to 0.90 g cm^{-3}. These density values correspond well to previously reported native-like proteins [42] and also implies that these proteins adopt a spherical geometry based on preliminary coarse-grain sphere fitting [42]. The general sphericity of these PLA_2s demonstrate the lack of significant extended or unfolded structural components and further implies the degree of compactness observed in these *P. colletti* PLA_2 ions may be correlated to the cysteine-rich structure known for PLA_2 and the high number of disulphide bonds that ensure the protein's structural integrity is maintained. The determined D_{eff} values and corresponding implied compact spherical structures for these PLA_2s is an interesting prelude to more refined structural characterisation and modelling, and demonstrates the power of IM-MS analysis in high-throughput structural characterisation.

3.5. Functional characterisation of dimeric and monomeric *P. colletti* PLA_2

A particular interest in this study was to investigate how the differences in oligomeric state of the venom proteins may affect biological activity. To probe the functional effects of oligomerisation on PLA_2 from *P. colletti* venom, an MS-based enzyme assay was used to compare catalytic efficiencies of representative monomeric and dimeric PLA_2s in the semi-purified *P. colletti* venom fractions from anion exchange chromatography (PC3-B and PC3-C respectively), by monitoring substrate depletion and concomitant product formation simultaneously for multiple substrates.

PLA_2s hydrolyse phospholipids specifically at the *sn*-2 position on the glycerol backbone to liberate a fatty acid from the lyso-phosphatidylcholine (LPC). Here, 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) was selected as the phospholipid substrate as the asymmetry of the phospholipid in its fatty acid chains enables discrimination of substrate hydrolysis by PLA_2, as opposed to PLA_1 activity which hydrolyses at the *sn*-1 position. POPC was chosen as a simple mimic of numerous cell membranes [33,43] which is a known site of PLA_2 enzyme interaction [44,45]. In this instance, a pair of regioisomers, POPC

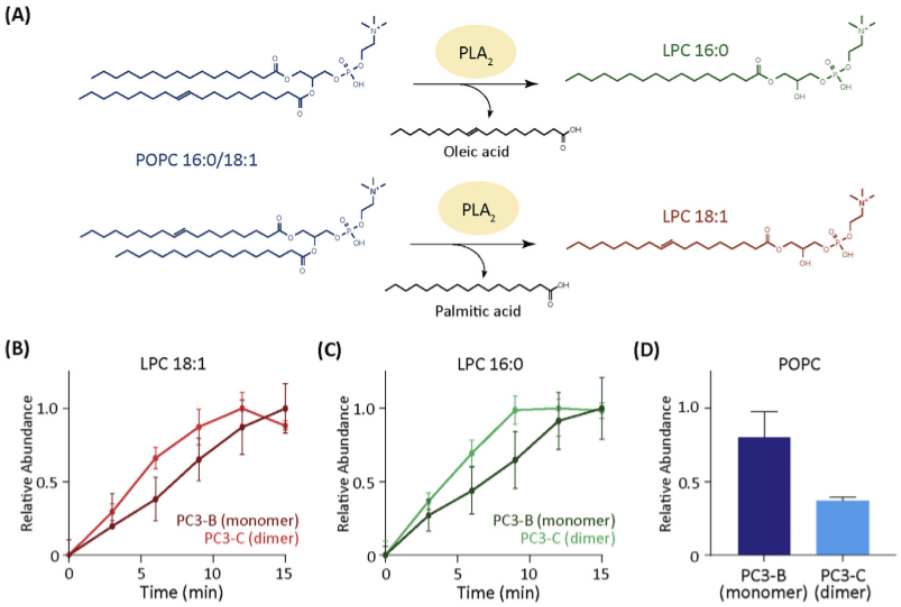


Fig. 6. MS-based PLA_2 enzyme assays. **(A)** PLA_2-mediated hydrolysis of a pair of regioisomeric phosphatidylcholines (POPC) 16:0/18:1 and 18:1/16:0. The liberated *lyso*-phosphatidylcholine (LPC) products 18:1 and 16:0 as monitored by MS for the representative monomeric and dimeric PLA_2s in the semi-purified *P. colletti* venom fractions (PC3-B and PC3-C respectively) are shown in **(B)** and **(C)** respectively. **(D)** Relative abundance of POPC at the conclusion of the MS-based enzyme assay (15 min) upon treatment with fractions PC3-B and PC3-C ($n = 3$).

16:0/18:1 and POPC 18:1/16:0 (number of carbons: number of unsaturated bonds along the fatty acid chain) were utilised which produce different LPC major products (LPC 16:0 and LPC 18:1, Fig. 6A).

Relative enzymatic activity was determined for the representative monomeric and dimeric PLA_2 rich fractions (PC3-B and PC3-C respectively) by monitoring ion abundance of the intact substrate and LPC products (as well as sodium adducted ions) over a 15 min period, whereby the ion intensity was taken as a measure of relative abundance of the species present (Fig. 6B–D). It should be noted that, while there is a mixed population of PLA_2s in the fractions, the same total protein concentration was used in each case. Both fractions exhibited significant PLA_2 activity, whereby the temporal depletion of POPC and the evolution of LPC products were observed during the assay. Notably, the rate of LPC formation was substantially greater for the dimer rich fraction compared with the monomers (Fig. 6B and C). The abundance of POPC substrate was also observed to be significantly lower for the dimeric PLA_2 fraction than it was for monomeric PLA_2s at the end of the assays (Fig. 6D), confirming the enhanced enzymatic activity of the enriched dimer fraction. Despite being a preliminary assessment of PLA_2 enzyme activity of crude venom fractions, this difference in enzymatic activity further highlights the importance of recognising higher-order protein interactions in the scheme of characterising venom protein function and activity.

4. Conclusions

Here we have demonstrated the use of an integrated MS-based approach to catalogue the identity and explore the higher-order structures of various venom proteins from the phylogenetically diverse venoms of *P. colletti* and *B. arietans*. Shotgun proteomics revealed the diversity in the venom proteomes where *P. colletti* venom is simple yet highly abundant in PLA_2, whereas *B. arietans* venom reflects the proteomic characteristics of viperid venom where DIS and CTL are some of the more abundant constituents. Importantly, supplemented by the proteomic findings, new higher order protein complexes were identified by native MS analysis in various venom fractions, including both non-covalently and covalently-linked homo-oligomers ranging from dimer to tetramer.

The finer structural details that enable the formation of these observed protein complexes are not yet known. Whether it is specific PTMs or structural moieties that facilitate oligomerisation are interesting aspects to explore in future experiments. Nevertheless, the interesting observation that the oligomeric state may influence enzymatic activity, at least for the PLA_2 enzymes, is an important step towards better understanding potency and specificity of venom components and enabling better treatments towards envenomation as well as development towards new therapeutic agents.

The sheer complexity of protein structures in snake venoms, considering the highly diverse proteoforms, various post-translational modifications, intrinsic intra- and inter-species variability and finally the possible oligomeric states means significant, ongoing efforts are required to fully explore these rich bioactive cocktails. While this study only highlights protein species from selected snake venom fractions, the established MS based workflow combining proteomics with native MS interrogation is nonetheless a good foundation to build additional structural analyses upon, as well as to explore a wider variety of venoms.

Acknowledgements

CRW and ERB were supported by an Australian Government Research Training Program Scholarship from the University of Adelaide. We thank the Adelaide Proteomics Centre and Flinders Analytical for access to mass spectrometry instrumentation. We also thank Katherine Stevens from the University of Adelaide for her support with the PLA_2 assay method and instrument optimization, and staff at Venom Supplies for support and helpful discussions regarding the snake venoms.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jprot.2020.103680.

References

[1] Y.S. Chan, R.C. Cheung, L. Xia, J.H. Wong, T.B. Ng, W.Y. Chan, Snake venom toxins: toxicity and medicinal applications, Appl. Microbiol. Biotechnol. 100 (14) (2016) 6165–6181.

[2] T.S. Kang, D. Georgieva, N. Genov, M.T. Murakami, M. Sinha, R.P. Kumar, P. Kaur, S. Kumar, S. Dey, S. Sharma, A. Vrielink, C. Betzel, S. Takeda, R.K. Arni, T.P. Singh, R.M. Kini, Enzymatic toxins from snake venom: structural characterization and mechanism of catalysis, FEBS J. 278 (23) (2011) 4544–4576.

[3] S. Xiong, C. Huang, Synergistic strategies of predominant toxins in snake venoms, Toxicol. Lett. 287 (2018) 142–154.

[4] C. Lister, K. Arbuckle, T.N.W. Jackson, J. Debono, C.N. Zdenek, D. Dashevsky, N. Dunstan, L. Allen, C. Hay, B. Bush, A. Gillett, B.G. Fry, Catch a tiger snake by its tail: differential toxicity, co-factor dependence and antivenom efficacy in a procoagulant clade of Australian venomous snakes, Comp. Biochem. Physiol. C Toxicol. Pharmacol. 202 (2017) 39–54.

[5] R. Doley, R.M. Kini, Protein complexes in snake venom, Cell. Mol. Life Sci. 66 (17) (2009) 2851–2871.

[6] R.J. McCleary, R.M. Kini, Non-enzymatic proteins from snake venoms: a gold mine of pharmacological tools and drug leads, Toxicon 62 (2013) 56–74.

[7] S.A. Kularatne, N. Senanayake, Venomous snake bites, scorpions, and spiders, Handb. Clin. Neurol. 120 (2014) 987–1001.

[8] S. Peigneur, J. Tytgat, Toxins in drug discovery and pharmacology, Toxins 10 (3) (2018) 126–130.

[9] Y.N. Utkin, Animal venom studies: current benefits and future developments, World J. Biol. Chem. 6 (2) (2015) 28–33.

[10] J.J. Calvete, Snake venomics: from the inventory of toxins to biology, Toxicon 75 (2013) 44–62.

[11] A.H. Laustsen, Toxin synergism in snake venoms, Toxin Rev. 35 (3–4) (2016) 165–170.

[12] J.R. Walker, B. Nagar, N.M. Young, T. Hirama, J.M. Rini, X-ray crystal structure of a galactose-specific C-type lectin possessing a novel decameric quaternary structure, Biochem. 43 (13) (2004) 3783–3792.

[13] A.V. Osipov, P. Rucktooa, I.E. Kasheverov, S.Y. Filkin, V.G. Starkov, T.V. Andreeva, T.K. Sixma, D. Bertrand, Y.N. Utkin, V.I. Tsetlin, Dimeric alpha-cobratoxin X-ray structure: localization of intermolecular disulfides and possible mode of binding to nicotinic acetylcholine receptors, J. Biol. Chem. 287 (9) (2012) 6725–6734.

[14] C. Uetrecht, R.J. Rose, E. van Duijn, K. Lorenzen, A.J. Heck, Ion mobility mass spectrometry of proteins and protein assemblies, Chem. Soc. Rev. 39 (5) (2010) 1633–1655.

[15] B.T. Ruotolo, J.L. Benesch, A.M. Sandercock, S.J. Hyung, C.V. Robinson, Ion mobility-mass spectrometry analysis of large protein complexes, Nat. Protoc. 3 (7) (2008) 1139–1152.

[16] G. Ben-Nissan, M. Sharon, The application of ion-mobility mass spectrometry for structure/function investigation of protein complexes, Curr. Opin. Chem. Biol. 42 (2018) 25–33.

[17] J.A. Harrison, J.A. Aquilina, Insights into the subunit arrangement and diversity of paradoxin and taipoxin, Toxicon 112 (2016) 45–50.

[18] J.A. Aquilina, The major toxin from the Australian common brown snake is a hexamer with unusual gas-phase dissociation properties, Proteins 75 (2) (2009) 478–485.

[19] G.K. Isbister, M.R. Hooper, R. Dowsett, G. Maw, L. Murray, J. White, Collett's snake (*Pseudechis colletti*) envenoming in snake handlers, QJM 99 (2) (2006) 109–115.

[20] R.B. Currier, R.A. Harrison, P.D. Rowley, G.D. Laing, S.C. Wagstaff, Intra-specific variation in venom of the African puff adder (*Bitis arietans*): differential expression and activity of snake venom metalloproteinases (SVMPs), Toxicon 55 (4) (2010) 864–873.

[21] P. Juarez, S.C. Wagstaff, J. Oliver, L. Sanz, R.A. Harrison, J.J. Calvete, Molecular cloning of disintegrin-like transcript BA-5A from a Bitis arietans venom gland cDNA library: a putative intermediate in the evolution of the long-chain disintegrin bitistatin, J. Mol. Evol. 63 (1) (2006) 142–152.

[22] E. Fasoli, L. Sanz, S. Wagstaff, R.A. Harrison, P.G. Righetti, J.J. Calvete, Exploring the venom proteome of the African puff adder, *Bitis arietans*, using a combinatorial peptide ligand library approach at different pHs, J. Proteome 73 (5) (2010) 932–942.

[23] D. Georgieva, D. Hildebrand, R. Simas, M.A. Coronado, M. Kwiatkowski, H. Schluter, R. Arni, P. Spencer, C. Betzel, Protein profile analysis of two Australian snake venoms by one-dimensional gel electrophoresis and MS/MS experiments, Curr. Med. Chem. 24 (17) (2017) 1892–1908.

[24] N.R. Casewell, S.C. Wagstaff, W. Wüster, D.A.N. Cook, F.M.S. Bolton, S.I. King, D. Pla, L. Sanz, J.J. Calvete, R.A. Harrison, Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms, PNAS 111 (25) (2014) 9205–9210.

[25] A.J. Hart, W.C. Hodgson, T. Scott-Davey, J.B. Harris, Neuromuscular toxicology of the venom of Collett's snake (*Pseudechis colletti*): a histopathological study, Muscle Nerve 43 (4) (2011) 552–559.

[26] A.J. Hart, G.K. Isbister, P. O'Donnell, N.A. Williamson, W.C. Hodgson, Species differences in the neuromuscular activity of post-synaptic neurotoxins from two Australian black snakes (*Pseudechis porphyriacus* and *Pseudechis colletti*), Toxicol. Lett. 219 (3) (2013) 262–268.

[27] D. Paixao-Cavalcante, A.K. Kuniyoshi, F.C. Portaro, W.D. da Silva, D.V. Tambourgi, African adders: partial characterization of snake venoms from three *Bitis* species of medical importance and their neutralization by experimental equine antivenoms, PLoS Negl. Trop. Dis. 9 (2) (2015) 1–18.

[28] T. Matsui, J. Hamako, K. Titani, Structure and function of snake venom proteins affecting platelet plug formation, Toxins 2 (1) (2010) 10–23.

[29] Y. Yamazaki, Y. Matsunaga, Y. Tokunaga, S. Obayashi, M. Saito, T. Morita, Snake venom Vascular Endothelial Growth Factors (VEGF-Fs) exclusively vary their structures and functions among species, J. Biol. Chem. 284 (15) (2009) 9885–9891.

[30] J.R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, Nat. Methods 6 (5) (2009) 359–362.

[31] N.J. Anthis, G.M. Clore, Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm, Protein Sci. 22 (6) (2013) 851–858.

[32] J.A. Vizcaíno, A. Csordas, N. del-Toro, J.A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q.-W. Xu, R. Wang, H. Hermjakob, 2016 update of the PRIDE database and its related tools, Nucleic Acids Res. 44 (D1) (2015) D447–D456.

[33] B.R. Hamilton, D.L. Marshall, N.R. Casewell, R.A. Harrison, S.J. Blanksby, E.A.B. Undheim, Mapping enzyme activity on tissue by functional-mass spectrometry imaging, Angew. Chem. Int. Ed. (2019), https://doi.org/10.1002/anie.201911390 in press.

[34] J.A. Loo, Studying noncovalent protein complexes by electrospray ionization mass spectrometry, Mass Spectrom. Rev. 16 (1998) 1–23.

[35] G.R. Hilton, J.L. Benesch, Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry, J. R. Soc. Interface 9 (70) (2012) 801–816.

[36] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, M. Mann, Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein, MCP 4 (9) (2005) 1265–1272.

[37] T. Tasoulis, G.K. Isbister, A review and database of snake venom proteomes, Toxins 9 (9) (2017).

[38] K.J. Clemetson, Snaclecs (snake C-type lectins) that inhibit or activate platelets by binding to receptors, Toxicon 56 (7) (2010) 1236–1246.

[39] J. Li, C. Santambrogio, S. Brocca, G. Rossetti, P. Carloni, R. Grandori, Conformational effects in protein electrospray-ionization mass spectrometry, Mass Spectrom. Rev. 35 (1) (2016) 111–122.

[40] D. Andrade-Silva, A. Zelanis, E.S. Kitano, I.L. Junqueira-de-Azevedo, M.S. Reis, A.S. Lopes, S.M. Serrano, Proteomic and glycoproteomic profilings reveal that post-translational modifications of toxins contribute to venom phenotype in snakes, J. Proteome Res. 15 (8) (2016) 2658–2675.

[41] M. Degueldre, J. Echterbille, N. Smargiasso, C. Damblon, C. Gouin, G. Mourier, N. Gilles, E. De Pauw, L. Quinton, In-depth glyco-peptidomics approach reveals unexpected diversity of glycosylated peptides and atypical post-translational modifications in *Dendroaspis angusticeps* snake venom, Int. J. Mol. Sci. 18 (11) (2017) 2453–2466.

[42] M.F. Bush, Z. Hall, K. Giles, J. Hoyes, C.V. Robinson, B.T. Ruotolo, Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology, Anal. Chem. 82 (22) (2010) 9557–9565.

[43] M.A. Sani, J.D. Gehman, F. Separovic, Lipid matrix plays a role in Abeta fibril kinetics and morphology, FEBS Lett. 585 (5) (2011) 749–754.

[44] C. Betzel, N. Genov, K.R. Rajashankar, T.P. Singh, Modulation of phospholipase A_2 activity generated by molecular evolution, Cell. Mol. Life Sci. 56 (1999) 384–397.

[45] L. Dominguez, L. Foster, J.E. Straub, D. Thirumalai, Impact of membrane lipid composition on the structure and stability of the transmembrane domain of amyloid precursor protein, PNAS 113 (36) (2016) E5281–E5287.

Appendix D: Publication 2

(In Preparation for submission to Analytical Chemistry using data from Chapter 2)

Development of a modular synthetic route for protein chemical cross-linking reagents

Henry M. Sanders[1], Emily R. Bubner[1], Katherine G. Stevens[1], Kayla M. Downey[1], Andrew D. Abell[1], Tara L. Pukala[1*]

[1] School of Physical Sciences, The University of Adelaide, Adelaide, SA 5005, Australia

*Correspondence: Tara L. Pukala: School of Physical Sciences, The University of Adelaide, Adelaide, SA 5005, Australia; tara.pukala@adelaide.edu.au; Tel. +61 8 8313 5497

Abstract

Cross-linking mass spectrometry is a rapidly emerging technique that gives information on protein structure and subunit architecture. Typically in such an experiment, chemical reagents covalently link protein sites together, which, when followed by proteolytic digestion and detection of linkage sites by mass spectrometry analysis, gives information on inter- and intra-protein contacts. While many cross-linking reagents are commercially available, the specific functionalities may not always satisfy the diverse chemical and analytical requirements best suited for a particular system or experiment of interest. Here we describe a modular synthetic protocol allowing customisation of reactive groups and spacer-arms, giving flexibility in linker design to enable the vast variety of potential protein cross-linking experiments. Implementing the general synthetic method allowed for the production of 8 unique homo- and heterobifunctional cross-linkers, including 4 different reactive groups, positive- and negative-mode mass spectrometry cleavable spacer-arms and sites for post-linkage derivatisation, for example with enrichment and identification tags. Optimisation and application of these reagents to model systems exemplifies the potential for this modular protocol to address some of the analytical challenges associated with protein structure determination by cross-linking mass spectrometry.

Introduction

The function of proteins is inherently tied to their three-dimensional structure and how they interact with other molecules. Consequently, the ongoing development of analytical methods to probe protein organisation is critical for advancements in structural biology. Many protein systems are not amenable to structure determination by traditional high-resolution techniques such as x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, due to inherent limitations of the methods. As a result, alternative low-resolution approaches such as chemical cross-linking combined with mass spectrometry (XL-MS) have emerged as powerful alternatives. XL-MS is a highly versatile tool to interrogate protein structure and map protein-protein binding interfaces.[1] Unlike these traditional methods, XL-MS does not require a homogenous analyte structure and can give information on highly dynamic proteins. Only micrograms of sample is required and it can be performed both *in vitro* and *in vivo,*[2,3] with either targeted or proteome wide applications.[4–6]

In a standard XL-MS workflow, proteins are chemically ligated in their native state using small molecule cross-linkers and then enzymatically digested. The resultant peptide mixture is sequenced by liquid chromatography tandem MS (LC-MS/MS) to locate linkage sites. The simplest cross-linker design includes two reactive units separated by a spacer arm which allow the compound to covalently link amino acids, commonly at basic or sulfhydryl residues (**Fig. 1**). For example N-hydroxysuccinimide (NHS), the most common reactive group, targets lysine or N-terminal residues, while maleimide reactive groups target cysteines.[3,7] In the case of certain proteins, amino acids of the same type are scarce, leading to the development of heterobifunctional linkers such as N-(α-maleimidoacetoxy) succinimide ester (AMAS), a linker that combines NHS and maleimide reactive groups. Building upon the concept of heterobifunctional reagents, non-specific linkers combine a 'standard reactive group' with a photoreactive group, typically aryl azides or diazirines, capable of linking indiscriminately to greatly increase potential linkage opportunities.

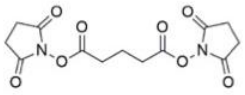


Figure 1. An example of a simple, commercially available cross-linker, disuccinimidyl glutarate (DSG), comprised of a non-cleavable 5 carbon chain between two lysine reactive NHS residues.

Following covalent cross-linking, as the length of the spacer arm is known, inter- and intra-protein upper distance constraints can be defined through site specific localisation of the cross-links.[3] These distance constraints allow for the interrogation of protein binding interfaces, intramolecular contacts and stoichiometry.[8] Furthermore, while techniques

like crystallography require rigid structures for accurate models to be constructed, XL-MS can also capture regions that exhibit structural differences between conformational states, and can give information on the rearrangements and transient interactions of proteins which are often critical in controlling cellular processes.[2,3]

Despite great advances in the field of XL-MS, analytical limitations remain. For example, due to low reaction efficiencies, unlinked peptides and 'dead-end' modifications, which have linked on one end and are hydrolysed on the other, significantly outnumber useful intra- and inter-protein cross-links. This has led to an increased interest in the introduction of enrichment tags capable of extracting modified peptides through a variety of common methods such as avidin-affinity or immobilised metal affinity chromatography.[9–12] In addition, peptides joined with non-cleavable linkers are required to be analysed as covalent pairs, giving rise to more complex MS/MS data as fragmentation occurs simultaneously from both chains. This has spurred the development of cleavable spacer-arms capable of fragmenting under low energy gas phase activation regimes (such as collision induced dissociation, CID) during MS.[13,14] Here, linked peptides initially fragment at the labile spacer-arm yielding a characteristic fragmentation pattern, distinguishing them from unlinked and dead-end linked peptides, and enabling the peptide pairs to be sequenced independently by MS^3 analysis to more easily identify the residue specific site of modification.[15] Finally, alternative strategies such as use of isotopically labelled linkers or linkers with reporter tag motifs which show characteristic mass differences during MS analysis can improve linked peptide identification.[16]

The structural diversity of proteins also creates unique challenges for each individual protein system. For example, only having access to relatively few reactive amino acids can conceal important linkage information. Differences in the amounts and types of available amino acids necessitates different linker reactivity, while larger distances between reactive groups may require different length spacer-arms. This has driven the development of new reactive groups like aromatic glyoxal cross-linkers (ArGOs), dihydrazide sulfoxide and 1,1'- carbonyldiimidazole, capable of reacting with arginine, acidic and hydroxyl amino acids, respectively.[17–19] While the wealth of new cross-linker options is beneficial to the field, in most applications, these are typically confined to commercially available compounds that lack flexibility in terms of spacer-arm length and the analytical features described above.

The ability to easily access a diverse range of cross-linking reagents with different reactivities, structures and properties would greatly expand the capabilities of XL-MS. This study describes the development of a modular synthetic strategy which allows for the straight forward production of cross-linkers that can incorporate various functionalities, including enrichment and fluorescent tags, varied length or CID-cleavable spacer-arms and hetero- or homobifunctional reactive groups, to enable researchers to readily adapt linker selection to the specific needs of the experiment or system of interest. We exemplify the synthetic utility of this strategy through construction of a small linker library and demonstrate application of the linkers to example peptide and protein systems.

Results and Discussion

Development of a modular synthetic protocol

To address the lack of diversity in commercially available cross-linker options, a modular synthetic protocol was developed based around amino acid building blocks to ensure the components are readily available, low cost and easily adaptable for functional group tuneability (**Fig. 2**). The synthesis was designed to allow for reactive groups and spacer arms that can be exchanged depending on the demands of the experiment, while also including a motif to which enrichment or luminescent tags (or other functionalities) can be attached after cross-linking.

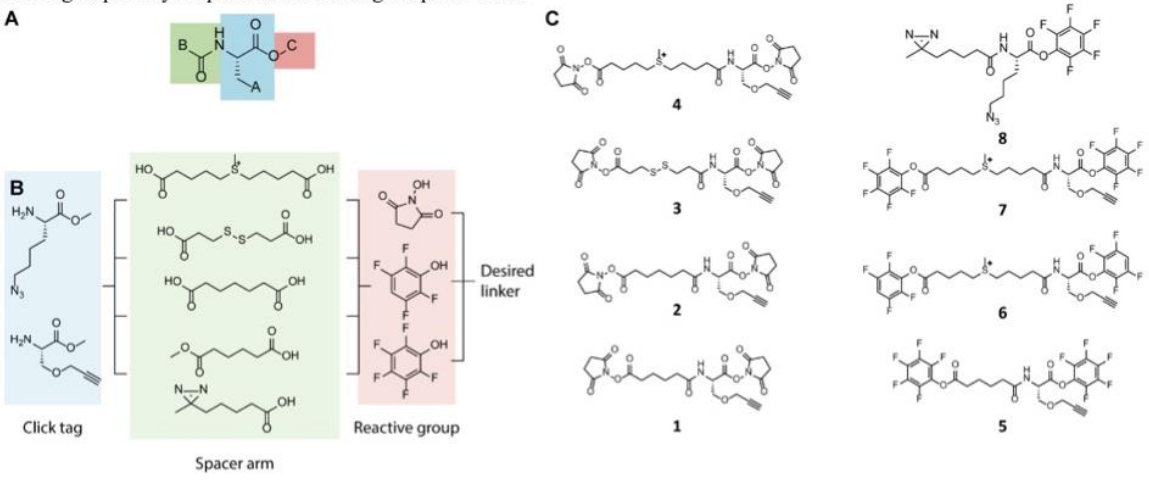


Figure 2. (**A**) The general structure of the modular linker includes 3 components; the 'click' modified amino acid core (**blue**), the spacer arm (**green**) and the reactive group (**red**), which can be synthesised in a combinatorial fashion (**B**). (**C**) Structures contained in the cross-linker library synthesised using the modular protocol.

At the core of the synthetic strategy is a base unit that includes a 'click' tag capable of undergoing derivatisation under biocompatible conditions. This click chemistry is mediated by an alkyne or azide tag which allows for the inclusion of post-linkage modifications via a copper catalysed Huisgen cycloaddition. Currently, there are numerous examples of alkyne/azide reagents that are commercially available, including modifications such as dyes, FLAG tags and agarose or magnetic beads. Two base units were proposed with an alkyne and an azide click tag (**Fig. 2B**), which allow for the same sample to be aliquoted and derivatised with more than one modification, which would be impossible if modifications were introduced to the linker prior to cross-linking. The alkyne base unit (**12**) was synthesised by O-alkylation of commercially available tert-butyloxycarbonyl-protected (N-Boc) serine with propargyl bromide, followed by a one pot N-Boc cleavage and methyl esterification (**Supp. Scheme 1**). The azide base unit (**10**) was synthesized by a diazo transfer reaction with an N-Boc protected lysine, followed by the same one pot N-Boc cleavage and methyl esterification reaction (**Supp. Scheme 1**).

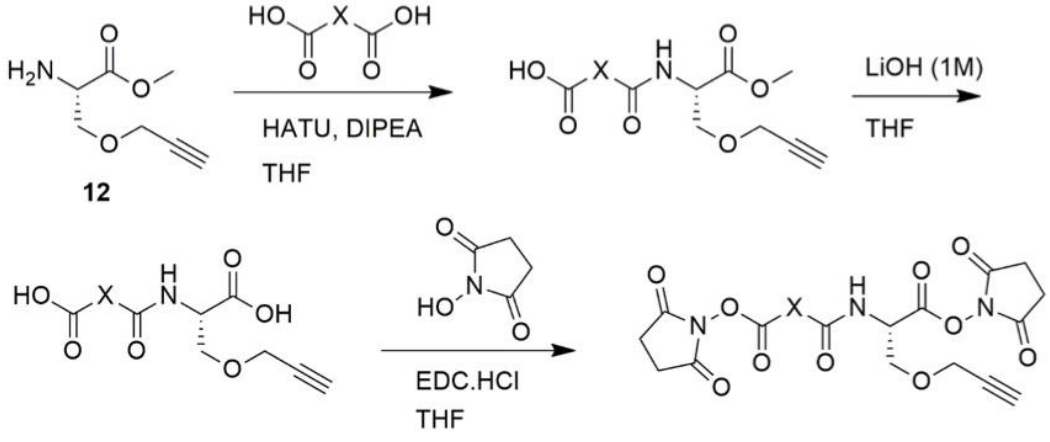
Following synthesis of the two base units, all subsequent reactions followed the same general reaction scheme wherein in the spacer-arm is coupled to the amino acid base unit by a peptide coupling, followed by a methyl ester hydrolysis and finally coupling of the reactive groups to the exposed acids (**Scheme 1**). For the easy incorporation of different spacer arms, the synthesis was designed around a hexafluorophosphate azabenzotriazole tetramethyl uronium (HATU) mediated peptide coupling so that any spacer arm with a carboxylic acid could be incorporated into the final compound (**Fig. 2B**). This allows different spacer arm lengths to be chosen by selecting the appropriate length diacid for coupling. In this study two examples were chosen, monomethyl adipate and heptanedioic acid, to showcase the simplicity by which different carbon chains lengths can be included.

To demonstrate the incorporation of gas-phase labile bonds into the cross-linker spacer arm, two additional diacids were chosen for coupling; S-methyl 5,5′-thiodipentanoic acid, capable of fragmentation under low energy CID conditions in the positive ion mode[20] and 3,3'-dithiopropionic acid, capable of facile fragmentation under negative ion CID conditions.[21]. It is possible to perform XL-MS analysis under both positive and negative ion MS conditions, and therefore it is advantageous to incorporate moieties that undergo selective fragmentation in both of these modes. For linkers **4**, **6** and **7** an additional step was necessary wherein the penultimate NHS-coupled compound undergoes methylation with methyl iodide to form the fixed charge sulfonium ion required for positive ion-mode fragmentation.

Following peptide coupling of the spacer arm, alkaline hydrolysis of the methyl ester yields two free carboxylic acids with which a variety of reactive groups are available for EDC coupling to generate the final reactive linker. Further customisation is possible to introduce UV-reactive heterobifunctional cross-linking through the coupling of 3-methyl-diazirine-3-propanoic acid, instead of a diacid, to yield linker **8**. Subsequent steps follow the same protocol, with the exception of reducing the equivalents of NHS coupling reagents to reflect that only one NHS is coupled to the final linker.

Based on this general synthetic protocol, a small library of eight linkers was synthesised with differing spacer arm lengths, fragmentation moieties and reactive groups, all of which included an alkyne or azide tag available for further modification post-linkage (**Fig. 2C**).

Scheme 1. General synthetic protocol highlighting the modularity of the approach. Introducing a diacid and reactive group of choice allows for synthesis of a custom linker.



Cross-linking and fragmentation characterisation

To demonstrate the reactivity of the synthesised library, a selection of amine reactive linkers was applied to a test peptide, acetylated AAKA (AcAAKA), which is a simple model system with a single primary amine available for cross-linking. Incubation of cross-linkers with AcAAKA at a high concentration allows for the simulation of the peptide linkages that could be expected in a proteolytic digest. Initially, compound **1** was used to form a cross-link between AcAAKA peptides and MS/MS analysis was performed on the linker peptide complex, the structure of which was

confirmed by the expected fragmentation products (**Supp. Fig. 1**). Following this, the facile fragmentation of CID cleavable linkers was validated using a negative and positive ion mode cleavable linker (compounds **3** and **4** respectively). Both compounds were reacted with AcAAKA and MS/MS analysis revealed spectra consistent with expected dissociation patterns (**Fig. 2**). Compound **4** formed a 6-membered oxazoline ring under low energy CID, preferencing the non-serine side, yielding a 125 *m/z* mass difference between fragmentation products (**Fig. 2A**).[20] Capable instruments may then scan for this characteristic 125 *m/z* doublet during MS/MS under lower fragmentation energies and isolate each ion for higher energy MS sequencing experiments, allowing each linked peptide to be more easily identified and then sequenced, greatly improving site-specific localisation of the modification.[22] Similarly, the negative ion mode linker cleaved as expected forming a characteristic quartet fragmentation pattern, also suited to subsequent MS^3 sequencing experiments.[23] The facile fragmentation about the disufide bond in the linker chain is effected by either an enolate anion or by an anion situated directly adjacent to the disulfide.[24] There are therefore four potential products possible as fragmentation can take place on the serine or non-serine side of the disulphide bond, although the relative abundance of these signals in the MS/MS spectra of the linked AcAAKA peptides did not reveal significant preference for either side in this example (**Fig. 2B**).

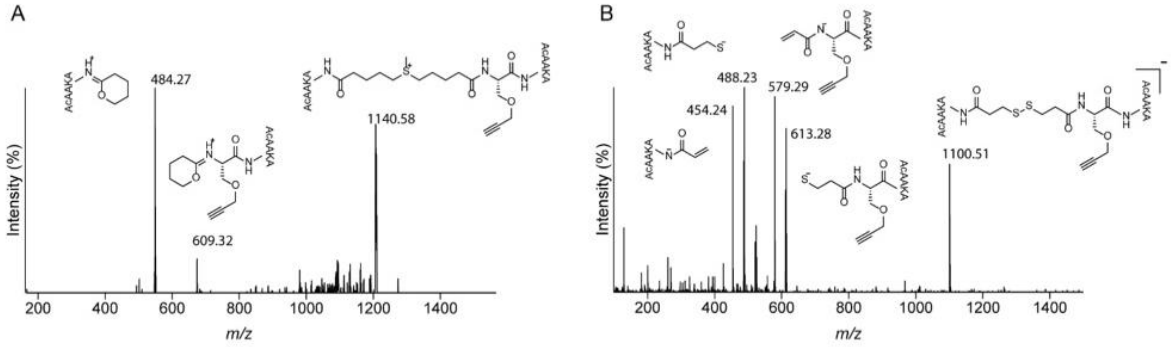


Figure 2. (**A**) Low CID energy of linker **3** cross-linked to AcAAKA yield two major daughter ions corresponding to 6-member oxazoline rings on the non-serine (484.27 *m/z*) or serine side (609.32 *m/z*) from the parent ion (1140.58 *m/z*). (**B**) The linker **4** parent ion (1100.51 *m/z*) yielded 4 daughter ions upon CID at a low energy: an enol product on the non-serine (454.24 *m/z*) and serine (579.29 *m/z*) side and a thiol product on the non-serine (488.23 *m/z*) and serine (613.28 *m/z*).

Overall, these results confirm that the alkyne modified serine is a suitable base unit to be combined with diacid spacer arms in order to successfully form covalent links between lysine residues, and the resultant linkers are able to effectively fragment under CID conditions.

Post cross-link derivatisation of the modular linker

A feature of the modular linker design is the included alkyne/azide tag which can be derivatised subsequent to the reaction with proteins. This is advantageous since it avoids the inclusion of bulky groups such as biotin in the reactive linker, which can potentially restrict access to reactive sites on the protein during the cross-linking reaction. It also allows for flexibility in the choice of 'clickable' motif to include a range of functionalization. Alkyne and azide groups on biotin, fluorescent tags, agarose UV-cleavable beads, peptides and oligos, amongst many other examples, are becoming increasingly available commercially, and the flexibility to apply several tags to a single cross-linked sample is attractive. After validation of covalent linkage, the ability for post-cross-linking derivatisation of the linkers through a copper catalysed Huisgen cycloaddition reaction was demonstrated with both biotin for enrichment and Cy3 for fluorescent visualisation.

Following covalent linkage between compound **1** and AcAAKA, a simple copper catalysed click reaction facilitated the addition of biotin azide. Evidence for the successful derivatisation was provided by MS, with MS/MS analysis of the expected $(M+H)^+$ ion of the product at 1294 *m/z* giving rise to expected fragmentation ions (**Fig. 3A**). Biotin affinity purification following this reaction allows for enrichment in a manner which is familiar and available to many biochemistry laboratories.

The flexibility offered by post-cross-linking derivatisation was further showcased using two different linkers, compound **5** and **6**. Both compounds and a commercially available control, disuccinimidyl sulfoxide (DSSO), were used to cross-link hen egg lysozyme and the resulting products were separated by SDS-PAGE and visualised using Coomassie Brilliant Blue stain (**Fig 3B**). Modification of the lysozyme monomer by cross-linkers was confirmed by MS (**Supp. Fig. 2**). Despite differences in linker length that may affect lysine accessibilities, both custom linkers showed comparable cross-linking efficiency with the DSSO control, and dimer band densities were reflective of the monomer to dimer ratio in buffered solution.[25] Before PAGE, a sample of protein linked with compounds **5** and **6** was further covalently modified with cy3-azide by copper catalysed cycloaddition, and modifications were evidenced by an upward band shift in the gel due to the added mass of the cy3 tag. The gels were then visualised by cy3 fluorescence emission at 556 nm, following excitation at 553 nm. To emphasize how fluorescence tags could be used to increase visibility of hidden covalently modified protein complexes,

smaller amounts of cy3-modified proteins were loaded on the gel to reduce the dimer intensity below limits of detection of the Coomassie stain. (**Fig 3B**). Notably, the band width of lysozyme:cy3-linker complexes increased, giving an indication of heterogenous mass changes due to attached cy3 tags. Upon irradiation and fluorescence detection, the dimeric species that were undetectable using Coomassie staining alone were revealed by the cy3 dye.

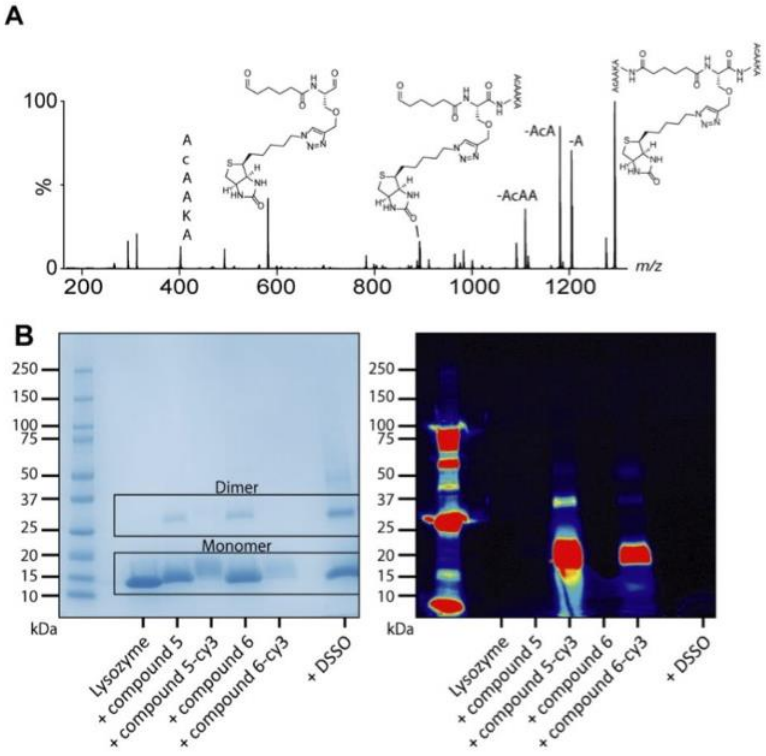


Figure 3. (**A**) CID was used to characterise cross-linked AcAAKA using linker **1** followed by post-linkage modification with biotin-azide using a Cu catalysed click cycloaddition reaction. The parent ion was identified (1293 *m/z*) and MS/MS analysis revealed characteristic fragment ions at 892 and 581 *m/z* that represent the parent ion minus one and two AcAAKA peptides, respectively. (**B, left**) SDS-PAGE revealed successful covalent linkage of lysozyme dimers using compounds **5** and **6** by comparison to a commercially available linker (DSSO), as well as post-linkage modification with cy3 (**B, right**). Fluorescence imaging highlights the ability to uncover hidden oligomers through dye derivatisation of linked samples.

The application of XL-MS has in recent years gone beyond *in vitro* samples and small numbers of interacting proteins, with cellular and tissue samples now subject to XL-MS for proteome wide interactomic analysis. The avenue of *in vivo* cross-linking and the increased complexity of these samples has led to a greater dependence on enrichment tags and cleavable spacer-arms. For example, cross-linking of mammalian cells with azide-A-DSBSO, a cross-linker capable of undergoing biotin click chemistry, was able to enrich and sequence 136 intrasubunit and 104 intersubunit protein cross-links.[26] Following this, new enrichment techniques allow for covalent linkage directly to alkyne or azide containing beads, skipping biotin-streptavidin enrichment techniques and improving enrichment efficiency by up to 5 times.[27] While impressive, the structural information obtained could be expanded upon further by introducing different spacer-arm lengths and reactive motifs using the modular approach presented. While this study explores 4 potential reactive groups, cysteine-reactive bismalemide and acid-reactive dihydrazide groups offer alternative linking chemistry that can be easily derived in one step from NHS-containing linkers as shown by *Gutierrez, et al.*[7,18] Furthermore, the wide availability of isotopically labelled amino acids would allow for the simple inclusion of a heavy serine/lysine base unit for synthesis, potentially yielding greater cross-link identification and even relative quantitation via the detection of characteristic isotope patterns during MS.[16]

Conclusion

This study presents a modular synthetic protocol for the synthesis of protein cross-linking reagents capable of tackling the individual challenges associated with protein structure determination by XL-MS. The ability to exchange reactive groups and spacer arms as needed was showcased

through the synthesis of a library containing 8 unique cross-linkers. The reagents analysed here displayed covalent linkages with model peptides and lysozyme with a comparable efficiency to that of commercially available cross-linkers. The introduction of cleavable spacer arms allows for improved identification strategies through diagnostic fragmentation peaks. In addition, centring the synthesis around an alkyne-/azide-modified amino acid allows for introduction of additional functionality through copper click chemistry, as we demonstrated through post-linkage modification of biotin for avidin-affinity purification or cy3 for improved visualisation through fluorescence. Overall, this synthetic protocol aims to reduce the complexity of XL-MS by offering a scaffold for greatly improving the diversity of potential cross-linking reagents, intentionally designed for the highly variable requirements of the field.

Materials and methods

Materials and reagents

Chemicals were purchased from Merck (Kenilworth, NJ, U.S.A.) or AK Scientific (Union City, CA, U.S.A.). The peptide AcAAKA was synthesised in-house using standard Fmoc solid-phase methods on 2-chlorotrityl chloride resin (GL Biochem, Shanghai, China) and purified by high-performance liquid chromatography (HPLC) to greater than 95 % purity as described previously.[24]

Synthesis of cross-linkers

All linkers were synthesised from a general synthetic procedure beginning with amide coupling between the desired diacid spacer arm (this includes coupling with 3-methyl-diazirine-3-propanoic acid) and either the alkyne base unit (**12**) or the azide base unit (**10**), followed by deprotection by methyl ester hydrolysis. Finally, esterification with N-hydroxysuccinimide (NHS), tetrafluorophnenol (TFP) or pentafluorophenol (PFP) and EDC-HCl yielded the final linkers (**Fig. 1C**). 500 MHz ^{1}H and ^{13}C nuclear magnetic resonance (NMR) spectra were obtained using an Agilent 500/54 Premium Shielded NMR spectrometer (Agilent Technologies, Santa Clara, CA, U.S.A.). High-resolution mass spectrometry (HRMS) data were obtained using an Agilent 6230 TOF LC/MS equipped with an Infinity 1260 LC system (Agilent Technologies). MSMS data were obtained using a Micromass QTOF2 mass spectrometer (Milford, MA, U.S.A.).

Synthesis of compound **10**

Triflic anhydride (25 mL, 146 mmol) was added to a solution of sodium azide (26.6 g, 410 mmol) in water (100 mL) while stirring at 0 °C. The reaction was allowed to warm to room temperature while stirring for 2 h, then extracted with dichloromethane (2 x 50 mL) and the combined organic extracts washed with saturated Na_2CO_3 (100 mL). The organic extract was added dropwise to a stirred solution of *N*-α-Boc-*L*-lysine (10.1 g, 41.1 mmol), $CuSO_4.5H_2O$ (1.11 g, 4.45 mmol) and K_2CO_3 (8.68 g, 62.8 mmol) in 2:1 (v/v) methanol/water (450 mL) at 0 °C. The solvent was removed under reduced pressure and the resulting residue diluted with water (75 mL) and acidified by dropwise addition of 6 N HCl until the formation of a precipitate. The suspension was then diluted with potassium phosphate buffer (pH 6.2, 150 mL of 1:1 (v/v) 0.25 M K_2HPO_4/ 0.25 M KH_2PO_4) before further acidifying to pH 3 with 6 N HCl. The mixture was then extracted with ethyl acetate (2 x 150 mL). The combined organic extract was washed with water (100 mL) and brine (100 mL), then dried ($NaSO_4$) and the solvent removed under reduced pressure. The residue was dissolved chloroform (100 mL), filtered by suction filtration and the solvent removed under reduced pressure to obtain a yellow oil (10.4 g, 93 %).

^{1}H NMR (500 MHz, CD_3OD, δ): 4.11-4.08 (q, J = 0.1 Hz, 1H), 3.33-3.30 (m, 2H), 1.88-1.81 (m, 1H), 1.72-1.64 (m, 1H), 1.65-1.60 (q, J = 0.2 Hz, 2H), 1.54-1.47 (m, 2H), 1.46 (s, 9H). ^{13}C NMR (125 MHz, CD_3OD, δ): 176.2, 158.1, 80.5, 79.4, 54.7, 52.2, 32.4, 29.4, 28.7, 24.1. HRMS (ESI) *m/z*: $[M+H]^+$ calculated for $C_7H_{14}N_4O_2$, 186.1117; found 188.1220.

Synthesis of compound **12**

Compound **12** was synthesised by dissolving Boc-L-Serine (2.01 g, 9.8 mmol) in DMF (100 mL) and stirring on ice for 15 min. A 60 % wt/wt dispersion of sodium hydride (995 mg, 24.8 mmol) was added and the solution was stirred on ice for 45 min. The reaction mixture was allowed to warm to room temperature and then stir overnight under a nitrogen atmosphere. Water (15 mL) was added and stirred for 5 minutes. The solution was washed with diethyl ether, acidified to <3 pH with 1 M $KHSO_4$ and then extracted with ethyl acetate (3x 30 mL). The combined extracts were washed with water and dried over Na_2SO_4. The solvent was removed *in vacuo* to yield **11** as a yellow oil which was used without further purification (2.1 g, 91 %).

^{1}H NMR (500 MHz, DMSO-d_6, δ): 5.38 (d, J = 8.2 Hz, 1H), 4.49 (d, J = 8.1 Hz, 1H), 4.23-4.14 (m, 2H), 4.00 (dd, J = 8.8, 2.2 Hz, 1H), 3.81 (dd, J = 9.4, 3.5 Hz, 1H), 2.47 (t, J = 2.3 Hz, 1H), 1.46 (s, 9H). ^{13}C NMR (125 MHz, $CDCl_3$, δ): 175.1, 155.9, 80.6, 78.8, 75.4, 69.6, 58.8, 53.8, 28.41. All NMR data is consistent with literature.[28]

The acid (**11**) (2.00 g, 8.22 mmol) was dissolved in methanol (100 mL) and stirred on ice for 15 min. Thionyl chloride was then added dropwise (1.20 mL, 16.4 mmol) under a nitrogen atmosphere. The solvent was removed *in vacuo* and the residue was co-evaporated with methanol three times to yield compound **12** (1.64 g, 97 %) as an orange oil.

^{1}H NMR (500 MHz, DMSO-d_6, δ): 8.64 (br s, 3H), 4.37 (t, J = 3.3 Hz, 1H), 4.24 (dd, J = 16.1, 2.2 Hz, 1H), 4.20 (dd, J = 16.1, 2.2 Hz, 1H), 3.91 (dd, J = 10.5, 4.1 Hz, 1H), 3.85 (dd, J = 10.6, 3.0 Hz, 1H), 3.76 (s, 3H) 3.56 (t, J = 2.1 Hz, 1H). ^{13}C NMR (125 MHz, CD_3OD, δ): 168.7, 79.4, 77.2, 67.6, 59.4, 54.3, 53.9. All NMR data is consistent with literature.[29]

HRMS (ESI+) calculated for $C_7H_{11}NO_3$ $[M+H]^+$ = 157.0739, found $[M+H]^+$: 158.0359

General modular synthetic protocol

10 or **12** was dissolved in dry tetrahydrofuran (THF) (5 mL per 100 mg) with diisopropylethylamine (4.0 mol eq) while

stirring. The desired spacer arm diacid (1.0 mol eq) was added to the solution followed by HATU (1.2 mol eq). The mixture was stirred overnight at room temperature under a nitrogen atmosphere. 1 M HCl (5 mL per 100 mg) was added and the solution was extracted with ethyl acetate (1.5 mL per 100 mg, 3×). The organic phase was washed with water (3×) and brine then dried over Na_2SO_4 and the solvent was removed *in vacuo*. The product was then purified by flash chromatography on normal phase silica to yield the desired amide.

To hydrolyse the methyl ester, the amide was dissolved in THF (1 mL per 100 mg) and stirred on ice for 15 minutes. 1 M LiOH (4 mol eq.) was added and the reaction was stirred until completion as determined by thin layer chromatography. The solvent was removed *in vacuo* and the residue was partitioned between DCM and water. The organic layer was isolated and the aqueous layer was extracted with DCM (2×). The combined organic extracts were washed with brine and dried over Na_2SO_4. The solvent was removed *in vacuo* to yield the desired diacid as a yellow oil.

The diacid was dissolved in anhydrous THF (10 mL per 100 mg) with EDC-HCl (3.4 mol eq) and stirred at room temperature for 10 minutes. NHS, TFP or PFP (6 eq.) were added and the reaction was stirred at room temperature under a nitrogen atmosphere overnight. The solvent was removed *in vacuo* and the resulting residue was partitioned between DCM and water. The aqueous layer was extracted with DCM (3x) and the combined organic extracts were dried over $MgSO_4$ and filtered. The solvent was removed from the filtrate *in vacuo* to afford the ester, which was used for cross-linking reactions without further purification.

Synthesis of compound **4**, **6** and **7**

Synthesis of positive mode cleavable linkers were based on work by *Lu, et al.*[20] Following coupling with 5,5'-thiodipentanoic acid and the corresponding reactive group, the thioether was dissolved with methyl iodide (2.2 eq.) in dry acetonitrile (1.5 mL) and allowed to react while stirring at room temperature for 4 days. The solvent was then removed *in vacuo* yielding a yellow oil which was used for cross-linking reactions without further purification.

Cross-linking of AcAAKA

Each linker was dissolved in DMSO to 10 mM with AcAAKA in a 1:1 linker:peptide ratio with 1 eq. of diisopropylethylamine. The mixture was incubated at room temperature for 1 hour. The mixture was then diluted to 10 μM in 50:50 water:acetonitrile for analysis by negative ion mode tandem MS and with 0.1 % aqueous formic acid for analysis by positive ion mode tandem MS.

Cross-linking of hen egg lysozyme

Lysozyme was dissolved in PBS to a concentration of 1 mg/mL. The relevant cross-linker was dissolved in DMSO (10 mM) and added to lysozyme solution at a ratio of 20:1 linker:lysozyme ratio. Protein was used for SDS-PAGE immediately or was buffer exchanged into ammonium acetate (100 mM) using a 10 kDa MWCO Amicon Ultra-0.5 Centrifugal Filter Device (Millipore, Jaffrey, NH, U.S.A.).

Copper click reaction

The general procedure for CuAAC reactions was performed using a method adapted from a protocol available from Broadpharm.[30] Peptide/protein conjugated with azide/alkyne linker (1mg/mL, 20 μL) was diluted with PBS (90 μL), followed by addition of aqueous solutions of 2.5 mM biotin-azide (20 μL) or 2.5 mM Cy3-azide, 100 mM tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (BTTAA, 10 μL) (Click Chemistry Tools, Scottsdale, U.S.A.), 20 mM copper sulfate (10 μL) and 300 mM sodium ascorbate (10 μL).The reaction was incubated at room temperature for 30 min, then purified and concentrated into PBS (100 μL) using a 10 kDA MWCO Amicon Ultra-0.5 Centrifugal Filter Device.

SDS–polyacrylamide gel electrophoresis

Lysozyme was cross-linked and examined by SDS–PAGE (12 % gels) (Bio-rad, Hercules, CA, U.S.A.). Bands were visualised using Coomassie Brilliant Blue stain (ThermoFisher, Waltham, MA, U.S.A.) according to the manufacturer's instructions. Fluorescence imaging was performed on a ChemiDoc gel imagine system (Bio-rad, Hercules, CA, U.S.A.) using an excitation wavelength of 553 nm and emission was measured at 556 nm.

MALDI analysis

Matrix-assisted laser desorption/ionization (MALDI) spectra were acquired using an ultraFlextreme MALDI-TOF/TOF mass spectrometer in reflector positive mode, controlled by FlexControl softwarev3.4 (Bruker Daltonics). Protein samples were prepared for MALDI-MS analysis by mixing with an equal volume of 10 mg/mL α-cyano-4-hydroxycinnamic acid (HCCA) matrix in 70 % acetonitrile plus 1 % trifluoroacetic acid (TFA). This solution was deposited on a ground steel 384 target plate for analysis of intact proteins. Spectra were converted to mzXML file format using FlexAnalysis v3.4 (Bruker Daltonics, Billerica, MA U.S.A.), then processed, including smoothing, cropping and baseline subtraction, using mMassv5.5.0 (http://www.mmass.org/).[31]

Acknowledgements

H.M.S. and K.G.S. were supported by a Faculty of Sciences Divisional Scholarship from the University of Adelaide. This work was supported in part by a Discovery Project Grant awarded to T.L.P. by the Australian Research Council (DP170102033).

Competing Interests

The authors declare that they have no competing interests associated with the contents of this manuscript.

Author Contributions

H.M.S., K.G.S., A.D.A, and T.L.P. designed the research; H.M.S., E.R.B., K.G.S. and K.M.D. performed the experiments; H.M.S., K.G.S. and T.L.P. analysed experimental data; H.M.S., A.D.A. and T.L.P. prepared the manuscript.

References

(1) Politis, A.; Stengel, F.; Hall, Z.; Hernández, H.; Leitner, A.; Walzthoeni, T.; Robinson, C. V.; Aebersold, R. A Mass Spectrometry–Based Hybrid Method for Structural Modeling of Protein Complexes. *Nat. Methods* **2014**, *11* (4), 403–406. https://doi.org/10.1038/nmeth.2841.

(2) Schmidt, C.; Zhou, M.; Marriott, H.; Morgner, N.; Politis, A.; Robinson, C. V. Comparative Cross-Linking and Mass Spectrometry of an Intact F-Type ATPase Suggest a Role for Phosphorylation. *Nat. Commun.* **2013**, *4*. https://doi.org/10.1038/ncomms2985.

(3) 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. https://doi.org/10.1038/s41594-018-0147-0.

(4) Nguyen, V. Q.; Ranjan, A.; Stengel, F.; Wei, D.; Aebersold, R.; Wu, C.; Leschziner, A. E. Molecular Architecture of the ATP-Dependent Chromatin-Remodeling Complex SWR1. *Cell* **2013**, *154* (6), 1220–1231. https://doi.org/10.1016/j.cell.2013.08.018.

(5) Liu, F.; Rijkers, D. T. S.; Post, H.; Heck, A. J. R. Proteome-Wide Profiling of Protein Assemblies by Cross-Linking Mass Spectrometry. *Nat. Methods* **2015**, *12* (12), 1179–1184. https://doi.org/10.1038/nmeth.3603.

(6) Chavez, J. D.; Weisbrod, C. R.; Zheng, C.; Eng, J. K.; Bruce, J. E. Protein Interactions, Post-Translational Modifications and Topologies in Human Cells. *Mol. Cell. Proteomics* **2013**, *12* (5), 1451–1467. https://doi.org/10.1074/mcp.M112.024497.

(7) Gutierrez, C. B.; Block, S. A.; Yu, C.; Soohoo, S. M.; Huszagh, A. S.; Rychnovsky, S. D.; Huang, L. Development of a Novel Sulfoxide-Containing MS-Cleavable Homobifunctional Cysteine-Reactive Cross-Linker for Studying Protein-Protein Interactions. *Anal. Chem.* **2018**, *90* (12), 7600–7607. https://doi.org/10.1021/acs.analchem.8b01287.

(8) Leitner, A.; Faini, M.; Stengel, F.; Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem. Sci.* **2016**, *41* (1), 20–32. https://doi.org/10.1016/j.tibs.2015.10.008.

(9) 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. https://doi.org/10.1021/acscentsci.9b00416.

(10) Petrotchenko, E. V.; Serpa, J. J.; Borchers, C. H. An Isotopically Coded CID-Cleavable Biotinylated Cross-Linker for Structural Proteomics. *Mol. Cell. Proteomics* **2011**, *10* (2), 1–8. https://doi.org/10.1074/mcp.M110.001420.

(11) Vellucci, D.; Kao, A.; Kaake, R. M.; Rychnovsky, S. D.; Huang, L. Selective Enrichment and Identification of Azide-Tagged Cross-Linked Peptides Using Chemical Ligation and Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (8), 1432–1445. https://doi.org/10.1016/j.jasms.2010.04.004.

(12) Sibbersen, C.; Lykke, L.; Gregersen, N.; Jørgensen, K. A.; Johannsen, M. A Cleavable Azide Resin for Direct Click Chemistry Mediated Enrichment of Alkyne-Labeled Proteins. *Chem. Commun.* **2014**, *50* (81), 12098–12100. https://doi.org/10.1039/c4cc05246c.

(13) Müller, M. Q.; Dreiocker, F.; Ihling, C. H.; Schäfer, M.; Sinz, A. Cleavable Cross-Linker for Protein Structure Analysis: Reliable Identification of Cross-Linking Products by Tandem MS. *Anal. Chem.* **2010**, *82* (16), 6958–6968. https://doi.org/10.1021/ac101241t.

(14) Kao, A.; Chiu, C.; Vellucci, D.; Yang, Y.; Patel, V. R.; Guan, S.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L. Development of a Novel Cross-Linking Strategy for Fast and Accurate Identification of Cross-Linked Peptides of Protein Complexes. *Mol. Cell. Proteomics* **2011**, *10* (1), M110.002212. https://doi.org/10.1074/mcp.m110.002212.

(15) Iacobucci, C.; Götze, M.; Ihling, C. H.; Piotrowski, C.; Arlt, C.; Schäfer, M.; Hage, C.; Schmidt, R.; Sinz, A. A Cross-Linking/Mass Spectrometry Workflow Based on MS-Cleavable Cross-Linkers and the MeroX Software for Studying Protein Structures and Protein–Protein Interactions. *Nat. Protoc.* **2018**, *13* (12), 2864–2889. https://doi.org/10.1038/s41596-018-0068-8.

(16) Ihling, C. H.; Springorum, P.; Iacobucci, C.; Hage, C.; Götze, M.; Schäfer, M.; Sinz, A. The Isotope-Labeled, MS-Cleavable Cross-Linker Disuccinimidyl Dibutyric Urea for Improved Cross-Linking/Mass Spectrometry Studies. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (2), 183–189. https://doi.org/10.1021/jasms.9b00008.

(17) Jones, A. X.; Cao, Y.; Tang, Y. L.; Wang, J. H.; Ding, Y. H.; Tan, H.; Chen, Z. L.; Fang, R. Q.; Yin, J.; Chen, R. C.; et al. Improving Mass Spectrometry Analysis of Protein Structures with Arginine-Selective Chemical Cross-Linkers. *Nat. Commun.* **2019**, *10* (1), 3911. https://doi.org/10.1038/s41467-019-11917-z.

(18) Gutierrez, C. B.; Yu, C.; Novitsky, E. J.; Huszagh, A. S.; Rychnovsky, S. D.; Huang, L. Developing an Acidic Residue Reactive and Sulfoxide-Containing MS-Cleavable Homobifunctional Cross-Linker for Probing Protein-Protein Interactions. *Anal. Chem.* **2016**, *88* (16), 8315–8322. https://doi.org/10.1021/acs.analchem.6b02240.

(19) Hage, C.; Iacobucci, C.; Rehkamp, A.; Arlt, C.; Sinz, A. The First Zero-Length Mass Spectrometry-Cleavable Cross-Linker for Protein Structure Analysis. *Angew. Chemie Int. Ed.* **2017**, *56* (46), 14551–14555. https://doi.org/10.1002/anie.201708273.

(20) Lu, Y.; Tanasova, M.; Borhan, B.; Reid, G. E. Ionic Reagent for Controlling the Gas-Phase Fragmentation Reactions of Cross-Linked Peptides. *Anal. Chem.* **2008**, *80* (23), 9279–9287. https://doi.org/10.1021/ac801625e.

(21) Lomant, A. J.; Fairbanks, G. Chemical Probes of Extended Biological Structures: Synthesis and Properties of the Cleavable Protein Cross-Linking Reagent [35S]Dithiobis(Succinimidyl Propionate). *J. Mol. Biol.* **1976**, *104* (1), 243–261. https://doi.org/10.1016/0022-2836(76)90011-5.

(22) Liu, F.; Lössl, P.; Scheltema, R.; Viner, R.; Heck, A. J. R. Optimized Fragmentation Schemes and Data Analysis Strategies for Proteome-Wide Cross-Link Identification. *Nat. Commun.* **2017**, *8* (1), 15473. https://doi.org/10.1038/ncomms15473.

(23) Calabrese, A. N.; Wang, T.; Bowie, J. H.; Pukala, T. L. Negative Ion Fragmentations of Disulfide-Containing Crosslinking Reagents Are Competitive with Aspartic Acid Side-Chain-Induced Cleavages. *Rapid Commun. Mass Spectrom.* **2013**, *27* (1), 238–248. https://doi.org/10.1002/rcm.6445.

(24) Calabrese, A. N.; Good, N. J.; Wang, T.; He, J.; Bowie, J. H.; Pukala, T. L. A Negative Ion Mass Spectrometry Approach to Identify Cross-Linked Peptides Utilizing Characteristic Disulfide Fragmentations. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (8), 1364–1375. https://doi.org/10.1007/s13361-012-0407-x.

(25) Maroufi, B.; Ranjbar, B.; Khajeh, K.; Naderi-Manesh, H.; Yaghoubi, H. Structural Studies of Hen Egg-White Lysozyme Dimer: Comparison with Monomer. *Biochim. Biophys. Acta - Proteins Proteom* **2008**, *1784* (7–8), 1043–1049. https://doi.org/10.1016/j.bbapap.2008.03.010.

(26) 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. https://doi.org/10.1074/mcp.M114.042630.

(27) 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. https://doi.org/10.1021/acs.jproteome.0c00003.

(28) Pehere, A. D.; Sumby, C. J.; Abell, A. D. New Cylindrical Peptide Assemblies Defined by Extended Parallel β-Sheets. *Org. Biomol. Chem.* **2013**, *11* (3), 425–429. https://doi.org/10.1039/C2OB26637G.

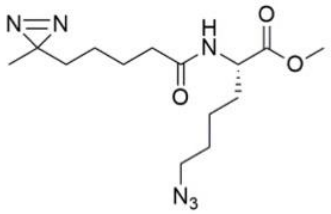
(29) Barman, A. K.; Gour, N.; Verma, S. Morphological Transition Triggered by Mannose Conjugation to a Cyclic Hexapeptide. *Arkivoc* **2012**, *2013* (2), 82–99. https://doi.org/10.3998/ark.5550190.0014.208.

(30) Broadpharm. Click Chemsitry Protocols https://broadpharm.com/public/uploads/protocol_files/Click Chemistry Protocol.pdf (accessed Jun 25, 2020).

(31) Strohalm, M.; Hassman, M.; Košata, B.; Kodíček, M. MMass Data Miner: An Open Source Alternative for Mass Spectrometric Data Analysis. *Rapid Communications in Mass Spectrometry*. 2008. https://doi.org/10.1002/rcm.3444.

Characterisation Data

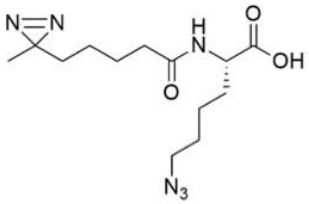
(S)-methyl 6-azido-2-(5-(3-methyl-3H-diazirin-3-yl)pentanamido)hexanoate



^{1}H NMR (500 MHz, $CDCl_3$): δ 6.39 (d, *J* = 8.0 Hz, 1H), 4.46 (dt, *J* = 8.0, 5.4 Hz, 1H), 3.62 (s, 3H), 3.16 (dt, *J* = 6.7, 1.2 Hz, 2H), 2.08 (t, *J* = 7.5 Hz, 2H), 1.78-1.69 (m, 1H), 1.65-1.51 (m, 2H), 1.48 (m, 4H), 1.36-1.26 (m, 2H), 1.26-1.21 (m, 2H), 1.09-1.02 (m, 2H), 0.87 (s, 3H) ppm.

^{13}C NMR: (125 MHz, $CDCl_3$): δ 175.6, 174.9, 55.1, 54.4, 53.7, 38.8, 36.6, 34.7, 31.0, 28.3, 27.6, 26.3, 25.0, 22.5 ppm.

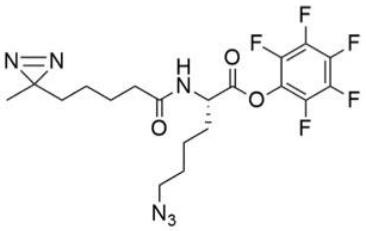
(S)-6-azido-2-(5-(3-methyl-3H-diazirin-3-yl)pentanamido)hexanoic acid



^{1}H NMR (500 MHz, $CDCl_3$): δ 9.95 (br s, 1H), 6.49 (d, *J* = 7.4 Hz, 1H), 4.54 (m, 1H), 3.24 (t, *J* = 6.8 Hz, 2H), 2.19 (t, *J* = 7.5 Hz, 2H), 1.90-1.84 (m, 1H), 1.72-1.69 (m, 1H), 1.62-1.54 (m, 4H), 1.43-1.38 (m, 2H), 1.33-1.30 (m, 2H), 1.16-1.10 (m, 2H), 0.94 (s, 3H) ppm.

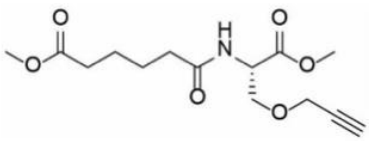
^{13}C NMR: (125 MHz, $CDCl_3$): δ 177.8, 176.5, 54.7, 53.7, 38.7, 36.5, 34.2, 31.0, 28.4, 27.7, 26.1, 25.0, 22.4 ppm.

(S)-perfluorophenyl 6-azido-2-(5-(3-methyl-3H-diazirin-3-yl)pentanamido)hexanoate (8)



HRMS (ESI+) calculated for $C_{19}H_{21}F_5N_6O_3$ $[M+H]^+$ = 476.1595, found $[M+H]^+$: 477.1667.

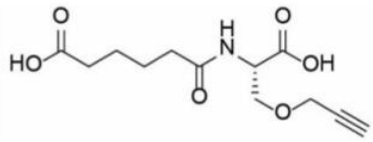
(S)-methyl 6-((1-methoxy-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-6-oxohexanoate



^{1}H NMR (500 MHz, $CDCl_3$, δ): 6.29 (d, *J* = 7.9 Hz, 1H), 4.79 (dt, *J* = 8.1, 3.1 Hz, 1H), 4.15 (dd, 2H), 3.97 (dd, *J* = 9.4, 3.1 Hz, 1H), 3.82 – 3.74 (m, 4H), 3.67 (s, 3H), 2.46 (t, *J* = 2.4 Hz, 1H), 2.34 (t, *J* = 6.9 Hz, 2H), 2.28 (t, *J* = 6.4 Hz, 2H), 1.75 – 1.61 (m, 4H).

^{13}C NMR (125 MHz, $CDCl_3$, δ): 174.0, 172.5, 170.7, 79.0, 75.3, 69.6, 58.7, 52.8, 52.4, 51.7, 36.1, 33.9, 25.1, 24.5.

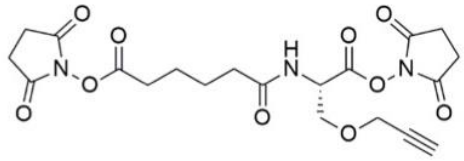
(S)-6-((1-carboxy-2-(prop-2-yn-1-yloxy)ethyl)amino)-6-oxohexanoic acid



^{1}H NMR (500 MHz, CD_3OD, δ): 4.63 (t, *J* = 4.2 Hz, 1H), 4.18 (d, *J* = 2.4 Hz, 2H), 3.91 (dd, *J* = 9.6, 5.0 Hz, 1H), 3.79 (dd, *J* = 9.6, 3.6 Hz, 1H), 2.86 (t, *J* = 2.4 Hz, 1H), 2.31 (q, *J* = 7.0 Hz, 4H), 1.71 – 1.59 (m, 4H).

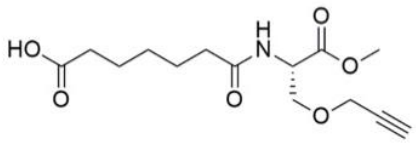
^{13}C NMR (125 MHz, $CDCl_3$, δ): 174.0, 172.5, 170.7, 79.0, 75.3, 69.6, 58.7, 52.8, 52.4, 51.7, 36.1, 33.9.

(S)-2,5-dioxopyrrolidin-1-yl 6-((1-((2,5-dioxopyrrolidin-1-yl)oxy)-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-6-oxohexanoate (1)



HRMS (ESI+) calculated for $C_{20}H_{23}N_3O_{10}$ $[M+Na]^+$ = 488.1261, found $[M+Na]^+$: 488.1271.

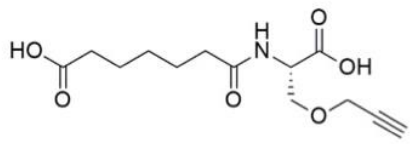
(S)-7-((1-methoxy-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-7-oxoheptanoic acid



^{1}H NMR (500 MHz, CD_3OD, δ): 4.71 (dt, J = 5.1, 3.7 Hz, 1H), 4.22 (dd, J = 2.4, 0.8 Hz, 2H), 3.95 (dd, J = 9.7, 5.1 Hz, 1H), 3.81 (dd, J = 9.7, 3.8 Hz, 1H), 3.78 (s, 3H), 2.92 (t, J = 2.4 Hz, 1H), 2.33 (q, J = 7.4 Hz, 4H), 1.74 – 1.62 (m, 4H), 1.48 – 1.35 (m, 2H)

^{13}C NMR (125 MHz, CD3OD, δ): 178.7, 177.5, 173.1, 81.3, 77.6, 71.3, 60.4, 55.2, 54.1, 37.6, 36.0, 30.9, 27.8, 27.0.

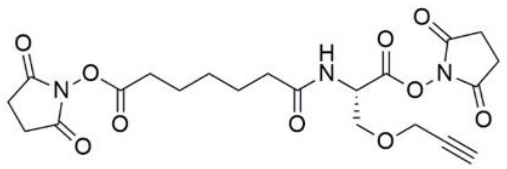
(S)-7-((1-carboxy-2-(prop-2-yn-1-yloxy)ethyl)amino)-7-oxoheptanoic acid



^{1}H NMR (500 MHz, CD_3OD, δ): 4.64 (dt, *J* = 5.0, 3.6 Hz, 1H), 4.19 (d, *J* = 2.4 Hz, 2H), 3.92 (dd, *J* = 9.6, 5.0 Hz, 1H), 3.80 (dd, *J* = 9.6, 3.7 Hz, 1H), 2.87 (t, *J* = 2.4 Hz, 1H), 2.30 (td, *J* = 7.5, 3.7 Hz, 4H), 1.92 – 1.83 (m, 4H), 1.64 (dp, *J* = 11.8, 7.5 Hz, 2H)

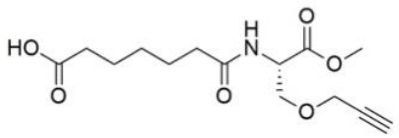
^{13}C NMR (125 MHz, CD_3OD, δ): 178.8, 177.4, 174.3, 81.4, 77.6, 71.5, 60.5, 55.1, 37.7, 36.0, 31.0, 27.8, 27.0.

(S)-2,5-dioxopyrrolidin-1-yl 7-((1-((2,5-dioxopyrrolidin-1-yl)oxy)-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-7-oxoheptanoate (2)



HRMS (ESI+) calculated for $C_{21}H_{25}N_3O_{10}$ $[M+Na]^+$ = 502.1438, found $[M+Na]^+$: 502.1421.

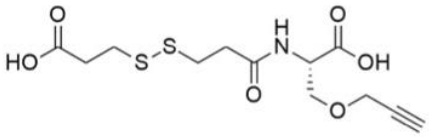
(S)-3-((3-((1-methoxy-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-3-oxopropyl)disulfanyl)propanoic acid



^{1}H NMR (500 MHz, CD_3OD, δ): 4.71 (dt, J = 5.1, 3.7 Hz, 1H), 4.22 (dd, J = 2.4, 0.8 Hz, 2H), 3.95 (dd, J = 9.7, 5.1 Hz, 1H), 3.81 (dd, J = 9.7, 3.8 Hz, 1H), 3.78 (s, 3H), 2.92 (t, J = 2.4 Hz, 1H), 2.33 (q, J = 7.4 Hz, 4H), 1.74 – 1.62 (m, 4H), 1.48 – 1.35 (m, 2H)

^{13}C NMR (125 MHz, CD3OD, δ): 178.7, 177.5, 173.1, 81.3, 77.6, 71.3, 60.4, 55.2, 54.1, 37.6, 36.0, 30.9, 27.8, 27.0.

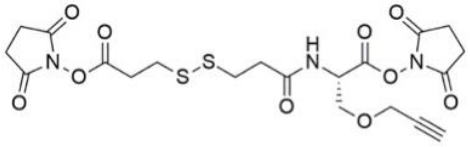
(S)-2-(3-((2-carboxyethyl)disulfanyl)propanamido)-3-(prop-2-yn-1-yloxy)propanoic acid



^{1}H NMR (500 MHz, CD_3OD, δ): 4.66 (t, J = 4.2, 4.2 Hz, 1H), 4.19 (d, J = 2.4 Hz, 2H), 3.94 (dd, J = 9.7, 4.8 Hz, 1H), 3.80 (dt, J = 9.8, 2.8 Hz, 1H), 3.01 – 2.90 (m, 4H), 2.72 (tt, J = 7.0, 3.2 Hz, 4H).

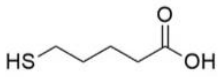
^{13}C NMR (125 MHz, CD_3OD, δ): 176.7, 176.7, 175.1, 81.4, 77.7, 71.5, 60.5, 55.2, 37.5, 36.2, 36.1, 35.6.

(S)-2,5-dioxopyrrolidin-1-yl 2-(3-((3-((2,5-dioxopyrrolidin-1-yl)oxy)-3-oxopropyl)disulfanyl)propanamido)-3-(prop-2-yn-1-yloxy)propanoate (3)



HRMS (ESI+) calculated for $C_{20}H_{23}N_3O_{10}S_2$ $[M+Na]^+$ = 552.0723, found $[M+Na]^+$: 552.0727

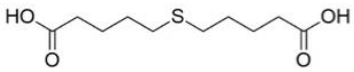
5-mercaptopentanoic acid



^{1}H NMR (500 MHz, $CDCl_3$, δ): 11.29 (s br, 1H), 2.59 – 2.51 (m, 2H), 2.38 (t, *J* = 7.1 Hz, 2H), 1.80 – 1.62 (m, 4H), 1.36 (td, *J* = 7.9, 1.1 Hz, 1H).

As per NMR data is consistent with literature.[20]

5,5'-thiodipentanoic acid

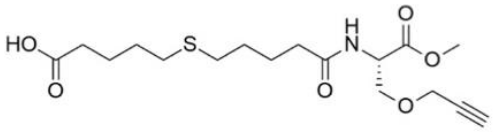


^{1}H NMR (500 MHz, $CDCl_3$, δ): 2.54 (t, *J* = 7.2 Hz, 4H), 2.40 (t, J = 7.2 Hz, 4H), 1.80 – 1.61 (m, 8H).

^{13}C NMR (125 MHz, $CDCl_3$, δ): 177.46, 36.02, 30.88, 27.00, 22.09.

As NMR data as per literature[20]

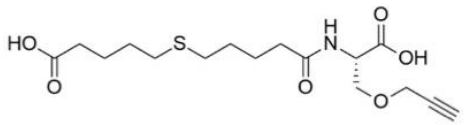
(S)-5-((5-((1-methoxy-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-5-oxopentyl)thio)pentanoic acid



^{1}H NMR (500 MHz, CD_3OD, δ): 4.67 (dt, *J* = 3.60, 1.37 Hz, 1H), 4.18 (d, *J* = 2.4 Hz, 2H), 3.92 (dd, *J* = 9.7, 5.0 Hz, 1H), 3.77 (dd, *J* = 9.7, 3.8 Hz, 1H), 3.74 (s, 3H), 2.89 (t, *J* = 2.4 Hz, 1H), 2.59 – 2.51 (m, 4H), 2.31 (q, *J* = 7.3 Hz, 4H), 1.78 – 1.57 (m, 8H).

^{13}C NMR (125 MHz, CD_3OD, δ): 178.3, 176.3, 175.9, 83.1, 81.5, 72.6, 61.7, 55.9, 55.9, 38.2, 37.1, 34.6, 34.6, 32.5, 32.4, 28.3, 27.7.

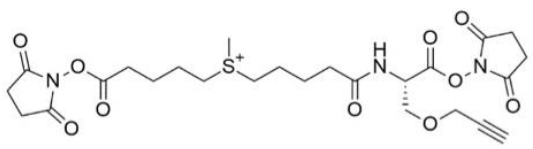
(S)-5-((5-((1-carboxy-2-(prop-2-yn-1-yloxy)ethyl)amino)-5-oxopentyl)thio)pentanoic acid



^{1}H NMR (500 MHz, CD_3OD, δ): 4.63 (t, *J* = 4.3 Hz, 1H), 4.19 (d, *J* = 2.4 Hz, 2H), 3.91 (dd, J = 9.6, 5.0 Hz, 1H), 3.80 (dd, J = 9.6, 3.6 Hz, 1H), 2.87 (t, J = 2.5 Hz, 1H), 2.54 (t, J = 7.1 Hz, 4H), 2.31 (q, *J* = 6.8 Hz, 4H), 1.78 – 1.66 (m, 4H), 1.63 (ddt, *J* = 12.1, 7.7, 5.0 Hz, 4H).

^{13}C NMR (125 MHz, CD_3OD, δ): 178.3, 176.1, 175.4, 83.8, 81.5, 73.0, 61.7, 55.9, 38.3, 37.1, 34.6, 34.6, 32.5, 32.5, 28.4, 27.6.

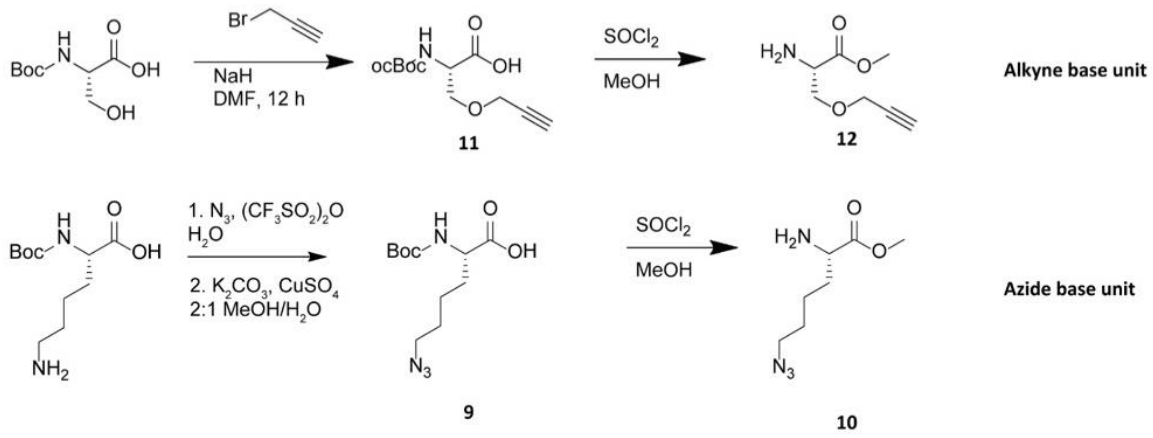
(5-(((S)-1-((2,5-dioxopyrrolidin-1-yl)oxy)-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-5-oxopentyl)(5-((2,5-dioxopyrrolidin-1-yl)oxy)-5-oxopentyl)(methyl)sulfonium iodide (4)



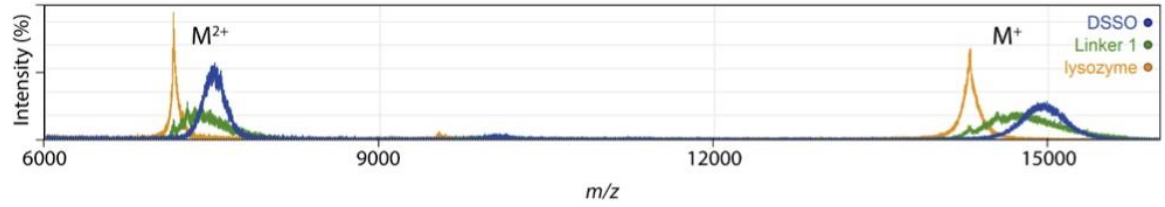
HRMS (ES+) calculated for $C_{25}H_{34}N_3O_{10}S$ $[M]^+$ = 568.1959, found $[M]^+$: 568.2037

Supplementary Information for Publication 2

Supplementary Scheme 1



Supplemetary Figure 1: MS/MS analysis of AcAAKA peptide cross-linked by compound **1** (*m/z* $[M+H]^+ = 1038$). Expected fragmentation patterns were used to confirm the presence of the peptide-linker complex.



Supplementary Figure 2. MALDI spectrum shows doubly charged (approximately 7100 *m/z*) and singly charged (14,200 *m/z*) lysozyme. The protein is modified, increasing its mass, upon the covalent linkage with multiple cross-linkers.