Design and Synthesis of Protein Chemical Crosslinkers: A Modular Approach

Kayla Monique Downey

Bachelor of Science (Molecular and Drug Design)

Supervisors:

Assoc. Prof. Tara Pukala

Prof. Andrew Abell

Thesis submitted for the degree of Master of Philosophy

THE UNIVERSITY *of* ADELAIDE

21st November 2017

School of Physical Sciences

The University of Adelaide

Contents

Abstract

The study of protein structure and interactions is pivotal in understanding the function and malfunction of complex biological systems. The structures of some proteins are unable to be determined using traditional high resolution biophysical techniques, requiring the development of amenable low resolution alternatives. Chemical Crosslinking Mass Spectrometry (CXMS) is one technique which can be used to probe protein structure through the formation of covalent linkages between protein residues. The formation of these links is facilitated by chemical crosslinking reagents.

Widespread use of the CXMS technique has been hampered primarily by analytical challenges pertaining to the detection and identification of crosslinked species using Mass Spectrometry (MS). Attempts to mitigate the challenges have been made by modifying the structure of chemical crosslinkers through the addition of functional groups such as affinity tags, isotope labels and cleavable bonds. Crosslinkers combining more than one type of functional group (combination crosslinkers) present the most promising targets for CXMS applications, combining the benefits of each functional group. However, combination crosslinkers are not commercially available, thus necessitating in-house synthesis. Incorporating more than one functionality also results in more complex molecular structures and synthetic processes, making the crosslinkers difficult to adapt to suit a particular experiment. Consequently, the use of combination crosslinkers has been limited to date to a small number of studies.

The research presented in this thesis describes the development of a modular chemical crosslinker design and corresponding synthetic protocol for the synthesis of combination crosslinkers. The modular crosslinker structure can be readily modified to include a range of functional groups using a small number of different reactions, including amide coupling and O-alkylation, and commercially available starting materials such as Boc-serine, from a minimum of five synthetic steps. The utility of the synthetic process was validated through the synthesis of a crosslinker containing an alkyne functional group, which can be used to attach a biotin affinity tag through alkyne-azide Huisgen Cyclisation.

Synthesis of two custom designed combination crosslinkers utilising alkyne tags and cleavable bonds is also described. The function of the cleavable bonds was established using collision induced dissociation processes within the mass spectrometer. Ensuring that a crosslinker is effective in probing quaternary structure and protein-protein interactions is essential as the investigation of these structures is a major goal of CXMS. Therefore, a crosslinking assay using *Staphylococcus aureus* biotin protein ligase, which forms homodimers when substrate bound, was also developed using the commercially available crosslinkers Disuccinimidyl Suberate (DSS) and Dithiobis(succinimidyl) Propionate (DSP), to enable the efficacy of crosslinkers synthesised using the modular synthetic protocol to be determined.

Declaration

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 consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the Universitys 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.

Kayla Monique Downey
21^{st} November 2017

Acknowledgements

I would like to thank my supervisors, Dr. Tara Pukala and Professor Andrew Abell, for their constant support and encouragement over the last two years. Thank you for being great guides, putting up with some tears, and for laughing with (or at) me. I would also like to thank the Pukala and Abell Groups (special mention to the ground floor Badger labs) and the people in chemistry with whom I have formed great friendships, for the company, laughs and guidance. You have all made these two years colourful and flavoursome, like a good soup.

I am grateful to the School of Physical Sciences Staff, and to the Technical Staff of the Badger and Johnson buildings for administrative and technical assistance throughout this journey.

I acknowledge and thank Steven Polyak and Louise Sternicki for their contribution to this thesis regarding the work with Biotin Protein Ligase. Your contributions are greatly appreciated.

This thesis would not exist without the tireless support of my family and friends. Thank you to my Mum and Dad, your endless love, support and nagging has guided me to this point. Thank you for everything. You have always believed in me, know that it is appreciated. To my Nanna, thank you for your love, encouragement and pride in all of my achievements. Your phone calls and text messages before every university event showed me how much you care, I will always be grateful for it. I am thankful everyday for my patient and understanding friends. To them I present the source of cancelled plans and unsociable behaviour, I promise now I will try to remember to reply to text messages.

Finally, thank you to my year 10 science teacher Clive Dobson and year 11 chemistry teacher Jane Nykke, for instilling a passion for science, chemistry and learning that will remain with me for life.

Abbreviations

Arom	Aromatic Chemical Shift
ATP	Adenosine Triphosphate
Boc	tert-Butyloxycarbonyl
BPL	Biotin Protein Ligase
CID	Collision Induced Dissociation
Cryo-EM	Cryo-electron Microscopy
CXMS	Cross Linking Mass Spectrometry
CXL	Chemical Cross Linking
Da	Daltons
kDa	kilo-Daltons
DC	Direct Current
DDA	Dithiodiglycolic Acid
DEA	Diethylamine
DCM	Dichloromethane
DIC	N'N-diisopropylcarbodiimide
DMF	Dimethylformamide
DNA	Deoxyribonucleic Acid
DSP	Dithiobis(succinimidyl) Propionate
DSS	Disuccinimidyl Suberate
EDC-HCl	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide Hydrochloride
ECD	Electron Capture Dissociation
EI	Electron Ionisation
eq	Equivalents
ESI	Electrospray Ionisation
FAB	Fast Atom Bombardment
Fmoc	Fluorenylmethyloxycarbonyl
h	hour(s)
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HDX	Hydrogen/Deuterium Exchange

HDX-MS	Hydrogen/Deuterium Exchange Coupled to Mass Spectrometry
HRMS	High Resolution Mass Spectrometry
HTH	Helix-Turn-Helix
HOBt	1-Hydroxybenzotriazole
HPLC	High Performance Liquid Chromatography
IMMS	Ion Mobility Mass Spectrometry
IR	Infrared
MALDI	Matrix Assisted Laser Desorption Ionisation
MCP	Microchannel Plate
MeO	Methyl Ester Protecting Group
MeOH	Methanol
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MS/MS/MS	3-Stage Tandem Mass Spectrometry
m/z	Mass to Charge Ratio
nESI	Nano-Electrospray Ionisation
NHS	N-hydroxy Succinimide
NMR	Nuclear Magnetic Resonance Spectroscopy
PBS	Phosphate-Buffered Saline
Q-ToF	Quadupole-Time of Flight
R_f	Retention Factor
RF	Radio Frequency
Rxn	Reaction
SaBPL	*Staphylococcus aureus* Biotin Protein Ligase
SAXS	Small Angle X-ray Scattering
$SOCl_2$	Thionyl Chloride
TFA-NHS	N-trifluoroacetoxy Succinimide
TFAA	Trifluoroacetic Anhydride
THF	Tetrahydrofuran
ToF	Time of Flight
TLC	Thin Layer Chromatography
WT	Wild Type

List of Figures

List of Schemes

Chapter 1

Introduction

1.1 Protein Structure

Proteins are functional macromolecules which perform a myriad of tasks within a biological system. They are comprised of polymeric chains of 20 different naturally occurring amino acids linked by peptide bonds. Proteins are involved in a plethora of biochemical pathways concerned with signalling, immunological function, synthesis and degradation, DNA transcription and translation, and homeostatic functions, regulating and catalysing many processes which are necessary for a system to function effectively. The structural and functional diversity of proteins and their complexes is the subject of ongoing study, as the function of a protein is underpinned by the three-dimensional structure that the polymeric chain adopts in solution. The three dimensional structure dictates the type, complexity and specificity of the reactions that can be catalysed and the type of molecular interactions which can be formed, generating the specificity required to allow many complex biological pathways to function simultaneously. The three-dimensional architecture of proteins consists of four levels of structure defined as primary, secondary, tertiary and quaternary structure (Figure 1.1).

The first or primary level of protein structure refers to the linear sequence of amino acids within the polypeptide chain. Each amino acid is linked by a 'peptide' or amide bond, forming the amide backbone. The ends of the polypeptide chain are referred to as the 'N-terminus' on the amino terminus and 'C-terminus' on the carboxyl terminus. The constituent amino acid residues are conventionally numbered from the N to the C terminus.

Secondary protein structure refers to regular structures formed within a domain of a polypeptide chain resulting from secondary interactions such as hydrogen bonding

between the functional groups of the amide backbone. Examples of these regular structures include β-sheets, β-strands, α-helices and turns.

Tertiary protein structure refers to the final three-dimensional structure formed by the folded polypeptide chain. The final structure consists of the native spatial arrangement of the primary and secondary structure motifs. Tertiary protein structure is a result of secondary interactions between amino acid side chains.

Finally, quaternary protein structure describes the three-dimensional structure of large multi-protein complexes formed by association of more than one polypeptide chain with tertiary structure.

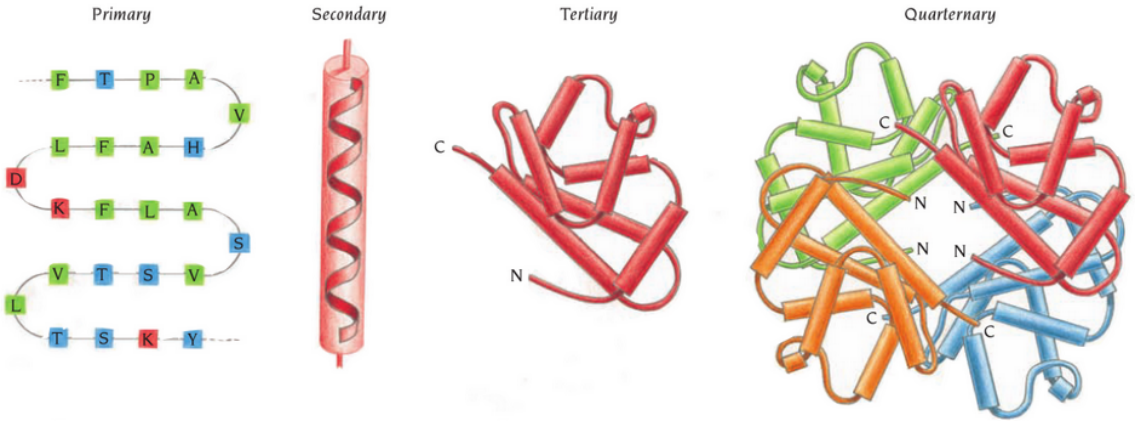


Figure 1.1: The four levels of protein structure, primary, secondary, tertiary and quaternary.[1]

1.2 Methods for Determining Protein Structure

As the function of a protein is predicated on the structure it adopts in solution, investigation of this structure can provide insight into its associated biological roles.[2,3] The roles of macromolecular complexes formed from intricate protein-protein interactions are also under scrutiny as more and more macromolecular assemblies taking part in major biological processes are discovered.[4,5] Detailed understanding of protein structure can also aid in the characterisation of disease phenotypes and development of novel treatment regimes using techniques such as structure based drug design.[6] Numerous techniques have been developed to explore protein structure at

varying levels of resolution. These techniques are generally divided into two main classes: high and low resolution.

High resolution techniques such as x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy provide atomic level detail, and hence provide the most structural information. However, experimental limitations inherent in these techniques have impeded the analysis of some biomolecules. Limitations pertaining to the solubility and molecular weight of the sample to be analysed have restricted NMR analysis of structures greater than tens of thousands of kDa with high resolution.[7,8,9] Conformational flexibility, hydrophobicity or intrinsically unstructured protein domains can affect the propensity of pure and heterogeneous samples to crystallise into an ordered 3-dimensional structure, hence some samples are intractable using x-ray crystallography.[10,11] Therefore, individual proteins and macromolecular complexes of high molecular weight, with inherently flexible structure and low solubility or inability to crystallise can be troublesome to characterise using high resolution methods.

Samples unable to be analysed by high resolution methods can be investigated using low resolution techniques such as cryo-electron microscopy (Cryo-EM),[12,13] small angle x-ray scattering (SAXS),[14,15] circular dichroism[16], dynamic light scattering[17] and mass spectrometry (MS)[18,19,20,21], among others. Combinations of high and low resolution techniques or more than one low resolution technique can give complementary information about a structure.[22,23,24,25,26] Low resolution techniques are also often used in association with computational modelling.[27,28]

Of the low resolution techniques, MS is of particular interest as: (1) the mass of the protein or protein complex to be analysed is theoretically unlimited, as it is the products of digestion which are analysed, (2) both stable and transient protein-protein associations can be probed yielding information about complex protein assemblies and interaction networks, (3) small amounts of sample are required for analysis (femtomole scales can theoretically be attained[29,30]), (4) obtaining data and subsequent analysis is generally fast and, (5) mixtures of post-translationally modified proteins or membrane proteins can be more suitable for analysis.

1.2.1 Mass Spectrometry and Protein Structure

MS was first developed in the early 20th century for the investigation of 'canal rays' (positive ions) by Joseph John Thomson. The earliest mass spectrometer was used to detect the first non-radioactive elemental isotope, ^{22}Ne by Thomson and Francis William Aston in 1913.[31] MS rapidly became a ubiquitous technique with the advancement of ionisation, detection and separation methods, swiftly evolving from the detection of atoms, to identification of molecular ions using their mass to charge ratio (m/z). Tandem mass spectrometry (MS/MS) was developed in the 1960s and allowed for the structural characterisation of a molecular ion based upon its fragmentation pattern.[32] Investigation of biological molecules using MS was impeded as traditional ionisation methods such as electron ionisation (EI) were not suitable for analysis of non-volatile biological compounds.[33] However, the establishment of novel 'soft' ionisation techniques such as electrospray ionisation (ESI)[34,35] and matrix-assisted laser desorption ionisation (MALDI)[36,37,38] in the 1980s by John Fenn, Franz Hillenkamp, Michael Karas and Koiochi Tanaka allowed for the ionisation of non-volatile biological samples. The development of these novel ionisation methods, and later, peptide MS/MS laid the foundation for structural analysis of protein samples using MS.

Several MS techniques have been developed to probe protein structure at all levels. Proteomics methods are used for the determination of primary protein structure using MS/MS approaches. The protein in question is typically subjected to proteolytic digestion, the constituent peptides separated chromatographically, the mass to charge ratio of the parent ion for each peptide is identified and then subjected to gas phase dissociation to obtain the individual sequences through characteristic peptide fragmentation patterns[39] (Section 1.4.2). The parent ion masses and individual sequences can then be compared to previously determined sequences in an appropriate database to identify the protein. Native MS and ion-mobility mass spectrometry (IMMS) are techniques through which intact proteins and protein complexes can be analysed.[19] Native MS allows for preservation of non-covalent interactions in the gas phase and is hence used to probe protein complex structure and interactions using the observed m/z ratio.[40] IMMS couples the measurement of collisional cross section (cross sectional area) and m/z ratio of a sample and is used to identify proteins from complex biological mixtures and to probe protein conformation in the gas phase.[41,42,43,44,45] Whilst IMMS is still in its infancy, evidence is mounting to suggest that solution based structures can be largely retained in the gas phase, further validating the utility of the technique.[19,46,47] Protein structure and dynamics can also

be investigated using hydrogen/deuterium exchange (HDX) coupled with mass spectrometry (HDX-MS), as the rate of HDX is influenced by the conformational features of a protein. Exposure of protein to deuterated water facilitates rapid exchange of the amide hydrogen to deuterium in regions of backbone surface accessibility and dynamic regions which lack stable hydrogen bonding, allowing for elements of protein architecture to be characterised.[48,49] The final prominent technique, chemical crosslinking mass spectrometry (CXMS), involves the coupling of chemical crosslinking and mass spectrometry used to yield information regarding secondary, tertiary and quaternary protein structure and often data pertaining to dynamic, or intrinsically unstructured protein domains. Chemical crosslinking (CXL) covalently traps proteins and protein complexes in their native conformational states using a reactive chemical agent, which can then be analysed using proteomics approaches.[50,29] CXMS is the protein structure determination technique which forms the basis of this thesis.

1.3 Chemical Crosslinking Mass Spectrometry

In some cases a protein may be intractable by traditional structure determination methods. In light of this, alternative methodologies such as CXL have been developed to allow characterisation of these systems. In its infancy, CXL was used in concert with gel electrophoretic methods to determine the identity of protein binding partners by their masses.[51] In 1974 Clegg *et al* used this method to describe the protein-protein interactions of the *Escherichia coli* ribosome.[52] However, the advent of peptide MS and *de novo* peptide sequencing allowed for the development of more advanced CXMS methods, in which the target proteins are digested by proteases following crosslinking, and analysed using MS techniques (Section 1.3.1). Structural studies pioneering the CXMS strategy began to be published in 2000. Young *et al* determined the fold structure of the bovine basic fibroblast growth factor FGF-2 by determining the distance constraints between lysine residues[53], Bennett *et al* described intermolecular contact regions between monomers of the homodimeric DNA binding protein ParR and interactions between glycoproteins CD28 and CD80[54] and Rappsilber *et al* investigated the spatial arrangement of the yeast nuclear pore complex Nup85p.[55] This pioneering work demonstrated the utility of CXMS as a method for investigating protein structure and associations, encouraging the wider development and application of the technique.

CXL applications still follow the same principles developed in the pioneering ex-

periments, in which a reactive chemical agent is added to a pure or mixed sample of protein to covalently trap a sample in its native conformational state. The site of crosslink formation is dependent upon the identity of the residue which is being targeted, and the distance between the targeted residues within the protein structure. The reactive chemical agent facilitating the linkage is known as a crosslinker (Figure 1.2). A crosslinker has two fundamental structural components termed the spacer arm and reactive group.

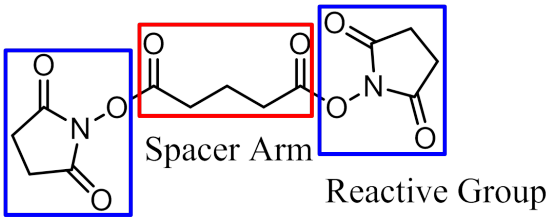


Figure 1.2: A simple crosslinker with two fundamental structural components, the spacer arm, highlighted in red, and the reactive groups, in blue.

The reactive group facilitates covalent bond formation between the crosslinker and protein(s). Crosslinkers are most commonly functionalised at both ends, defined as 'bifunctionalised', and in some cases the reactive groups at each end can be made to target different residues. Crosslinkers with the same reactive group at either end are defined as 'homo-bifunctionalised' and crosslinkers with two different reactive groups at each end are defined as 'hetero-bifunctionalised'. The nature of the reactive group can be modified to target specific protein functional groups such as amines (-NH_2), sulfhydryls (-SH) and carboxyl (-COOH) groups, or to non-specific photoreactive groups such as photoreactive diazrines. The reactive group specified in Figure 1.2 is an N-hydroxy succinimide (NHS) ester which is the most commonly used reactive group in crosslinking literature.[29] The NHS ester reacts with nucleophilic functional groups, typically the amine of the lysine side chain residue and occasionally the N terminus or the amide groups of the peptide backbone. NHS mediated crosslinking is advantageous as the reaction produces stable amide linkages at physiological pH (7.2-7.5).[30] Lysines also have a relatively high natural abundance in protein systems, especially on the surface topology, making these residues useful for probing quaternary interactions which take place primarily through association with surface residues.[50,30] Crosslinking in pH buffered solution is common as formation of amine specific crosslinks is strongly pH dependent. Acidic pH can result in the loss of crosslink specificity as NHS activated esters undergo side reactions with tyrosine, serine, threonine and histidine[56,57,58], which must be accounted for when determining and assigning crosslinks.

The spacer arm module typically acts as a defined distance constraint between the functionalised termini. However, the spacer arm also provides flexibility for the incorporation of functional moieties such as cleavable groups and isotopic labels (Section 1.3.3) in addition to customisation of crosslinker length, allowing for the probing of residues within a protein at varying distances.

A wide range of amino acid side chains and distances between residues can be probed with commercially available and literature reported crosslinkers that vary in length, reactive group identity and functionalisation, which are discussed in Section 1.3.3. Through the use of these crosslinkers the CXMS workflow is used to obtain structural information.

1.3.1 The CXMS Workflow

Two types of approaches to the CXMS workflow exist, namely, bottom-up and top-down. The top-down approach analyses intact crosslinked protein, utilising MS/MS fragmentation within the mass spectrometer without prior protein digestion.[59,60,61] The use of the top-down approach for analysis of larger proteins has been hindered due to limitations surrounding ionisation and fragmentation efficiency, which are closely related to the size and type of protein to be analysed.[29,30] The majority of CXMS experiments therefore follow a bottom-up approach in which the protein is enzymatically digested with a protease following crosslinking, with minor modifications depending on the crosslinker and protein to be analysed (Figure 1.3).[51,50] A protease commonly used in this workflow is trypsin, which cleaves the protein(s) on the C-terminal side of lysine and arginine residues.[62] Proteolytic digestion produces a complex mixture of modified (by the crosslinker) and unmodified peptides which are subject to MS analysis with the aim of identifying crosslinked species and site specific determination of crosslinked residues. These crosslinking sites can then be used to determine distance constraints between residues, which are defined by the spacer arm length of the crosslinker, to form a residue proximity map in which residues close in space, but not necessarily within the primary protein sequence, are associated within the three dimensional structure.[33]

Following digestion with a protease, several types of peptide species are produced, depending on the nature of the covalent interactions formed by the crosslinker (Figure 1.4). Different types of crosslinked peptides yield specific information regarding the structure under analysis and nomenclature for each type of crosslinked species

was coined by Schilling *et al* in 2003.[63] However, peptides that have not reacted with a crosslinker do not yield any higher order structural information regarding the structure in question. Four types of modified species are formed:

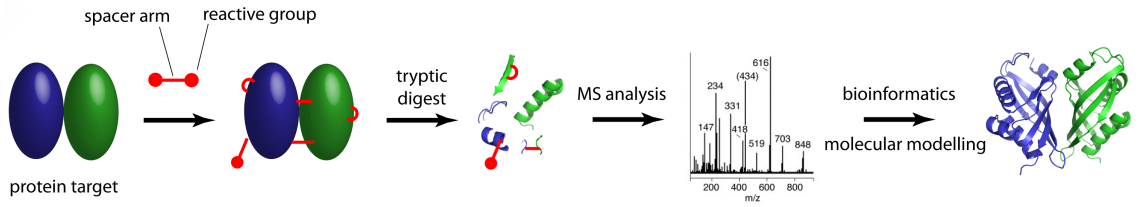


Figure 1.3: General bottom-up CXMS workflow for the analysis of an interacting protein system.[50]

Type 0 crosslinks arise when a crosslinker reacts with a peptide fragment at one of the functionalised termini while the other termini is hydrolysed (dead end). This type of crosslink does not provide information regarding distance constraints, but can be useful in determining solvent accessibility of the modified residue.

Type 1 crosslinks arise when both ends of a bifunctionalised crosslinker react with residues in the same peptide fragment (cyclic or intra-peptide link). Type 1 crosslinks can provide information about local secondary and tertiary protein structures.

Type 2 crosslinks are formed when both ends of a crosslinker react with residues from separate protein fragments (inter-peptide link). Type 2 crosslinks are arguably the most valuable crosslinked species, yielding distance information about long-range interactions within a single protein, or identifying protein-protein interactions as part of protein complexes.

Type 3 crosslinks also exist in which multiple crosslinks may have been formed to join several peptides together, amongst other complex linkages. Type 3 crosslinks are difficult to detect, identify and analyse and hence typically do not contribute to the structural information obtained from CXMS.

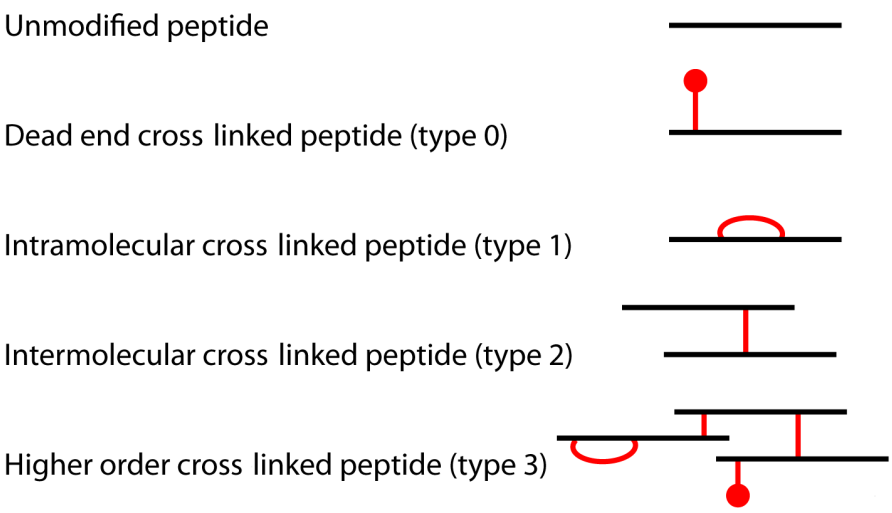


Figure 1.4: Peptide products that can be formed using a bifunctional crosslinker, after proteolytic digestion.[50]

Following crosslinking, peptide mixtures are usually separated using liquid chromatography (LC), ionised by electrospray ionisation (ESI) and analysed by MS/MS. Crosslinked peptides can be sequenced using *de novo* peptide sequencing experiments and the identity of unknown interacting species can be elucidated using protein data base searching. Sequencing using MS/MS can also generate information regarding the site of crosslink modification within a peptide. However, complex MS/MS behaviour, amongst other analytical challenges has hindered widespread use of the CXMS technique for structural analysis of proteins and protein complexes.

1.3.2 Advantages and Challenges of CXMS

Low resolution structure determination methods such as CXMS are advantageous for the characterisation of protein samples which otherwise are unable to be studied by high resolution methods (Section 1.2). Of the low resolution techniques, MS and CXMS afford some specific experimental advantages.

Using CXMS the mass of a protein sample to be analysed is theoretically unlimited, particularly using the bottom-up approach as the protein sample is digested with a protease after crosslinking (Figure 1.3). This allows for the analysis of large systems without the mass limitations imparted by high resolution techniques such as NMR. Some of the largest structures to be mapped using CXMS include the architecture

of the RNA polymerase II-TFIIF complex at 670 kDa,[64] the subunit order and architecture of the eukaryotic chaperonin TRiC/CCT at 1 MDa and the location of the PsbQ protein within the cyanobacterial photosystem II.[65,66,67]

Proteins which are generally insoluble in NMR amenable solvents, or are not able to crystallise effectively for x-ray crystallography can also be characterised using CXMS. Recently the architecture of full length retroviral integrase monomer and dimers from the avian sarcoma virus were studied using a combination of SAXS and CXMS, as the full length structure was not soluble enough to attempt structure determination using traditional biophysical techniques.[68] The study revealed differences between the full length structure and x-ray structures of individual domains pertaining to the interactions of core domains and stabilisation of the dimer structure by the C terminal domain.

Protein-protein interactions in macromolecular complexes, transient protein-protein interactions and the forms of conformationally dynamic/intrinsically disordered structures can be probed using CXMS.[69,50,70] The crosslinking process covalently traps a protein structure at a 'snapshot in time', establishing permanent linkages between transiently interacting species or short lived conformations in solution.[71] Identification of such interactions can allow for the elucidation of the identity of interacting species, as well as the characterisation of the binding interfaces.[21] Large scale mapping of protein interaction networks is also possible using this method, providing another alternative to conventional yeast two hybrid and co-immunoprecipitation assays. A notable array of protein-protein interactions within the protein phosphatase 2A network were recently characterised by Herzog *et al*, through which 176 interprotein (type 2) crosslinks and 570 intraprotein (type 1 and 2) crosslinks were identified, linking a large group of proteins to specific phosphatase 2A trimers which are postulated to control the trimers cellular functions.[27] Numerous other examples of protein interaction mapping using CXMS are also reported in literature.[72,73,74,75,76] The full length structure of the human p53 tumor suppressor protein monomers and tetramers were also recently studied, of which 37% of the structure is natively disordered.[77,24] Through the combination of SAXS and CXMS data it was determined that the N terminal domain of the p53 protein in tetrameric form was more compact than originally suggested by the SAXS structure.[24] The structural investigation of p53 using SAXS and CXMS is also a good example of the complementary structural information able to be obtained from the two techniques.

In contrast to the sizeable amount of protein sample required for structural analysis by x-ray crystallography and NMR, CXMS utilises a relatively small amount of

protein sample and obtaining data is generally fast. Femtomole scales for sample analysis can theoretically be attained,[29,30] which is facilitated by the advancements in sensitivity of MS instrumentation. As a result of the small amount of protein required for analysis, proteins which are not naturally abundant or difficult to isolate in large quantities can be studied. The study of post-translationally modified and membrane proteins are examples of species which due to the reduced sample requirements can, in some cases, be made amenable to CXMS analysis.

Most crosslinking reagents focus on the reactivity of easily accessible residues within the protein substrate. The high abundance and surface accessibility of lysine residues within proteins have posed a good target for crosslinking studies, however protein domains which consist mostly of hydrophobic residues may not be able to be mapped as efficiently using crosslinking experiments due to the lack of targetable residues.[51] This often results in the use of complementary structure determination techniques to supplement crosslinking data. The use of non-specific photoactivated crosslinkers can aid in the probing of hydrophobic structure.

The development of the CXMS technique has presented several challenges pertaining to the down stream analysis of peptide products. Proteolytic digestion produces a complex mixture of species, of which unmodified peptides are the most abundant. Unmodified peptides have not reacted with a crosslinker and hence do not yield higher order structural information about the protein under analysis (Figure 1.4). As a result of this, crosslinked species are generally of relatively low abundance and can be difficult to detect and identify directly from the output spectra.[50,29] Difficulties arise notably in the case of type 2 crosslinked peptides which yield valuable structural information. Type 2 crosslinks are the rarest species of all crosslink types and hence are the most challenging to detect and identify in complex mixtures. Information regarding protein-protein interactions and long range associations may be lost due to an inability to detect and identify these species.

The ability to identify the exact site of crosslink modification within a peptide allows for the determination of the most accurate distance constraints between residues and eliminates ambiguity when assigning crosslinks, which can arise when a peptide contains more than one residue which can be modified. The site specific location of crosslinked residues is more complex for type 2 crosslinked species (Figure 1.4) as once isolated, MS/MS analysis of type 2 crosslinks can result in incomplete or co-fragmentation of each peptide, complicating the MS/MS spectrum (Figure 1.5). Determining which peptide individual residues originated from can be onerous, and hence residue specific assignment of crosslink modification may not be possible.[50]

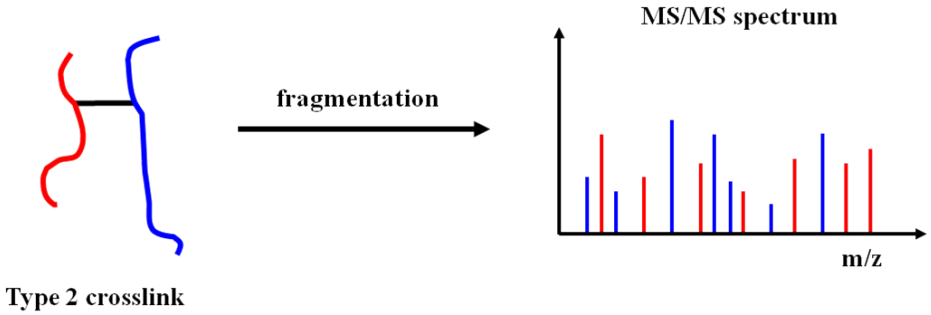


Figure 1.5: MS/MS fragmentation of a type 2 crosslinked species consisting of a red and blue peptide connected by a single crosslink. Co-fragmentation results in a complex MS/MS spectrum.

The complex nature of the peptide mixtures produced in proteolytic digestion and the volume of data generated by the crosslinking experiment has required the development of automated data analysis programs.[78] Calculation of possible peptide masses for all of the possible crosslink types is cumbersome, time consuming and complex, therefore software development has been imperative to improve the utility of the CXMS technique. The requirement for automated analysis becomes more prevalent as the protein sample becomes more complex and it is possible to form a greater number of crosslinks.

Potential solutions to the analytical conundrums facing CXMS are currently under investigation. Modern MS instruments have been developed with greater sensitivity, more fragmentation methods are under investigation and different software to aid in the analysis of crosslinking data has been developed and various specialised programs are more widely available including Pro-Crosslink,[79] StavroX/MeroX,[80,81], XlinkX,[82] X-Link Identifier,[83] MassMatrix[84] and many more.[53,85,86,87,88]

Diversification of crosslinker structure has also been a major focus of crosslinking literature in attempts to mitigate some of the current analytical challenges facing CXMS.[30]

1.3.3 Crosslinker Design

Various changes to the basic crosslinker design (Figure 1.2) have been implemented to address the analytical challenges which currently hinder the CXMS technique. The changes aim to reduce the complexity, and maximise the amount of information obtained from crosslinking experiments through the use of isotope labels, affinity tags and cleavable bonds.

Isotope Labelling

Crosslinkers can be labelled with stable isotopes, e.g, hydrogen (^{2}H) or carbon (^{13}C), at one or more sites within the crosslinker structure, as shown in Figure 1.6. Protein samples can then be crosslinked with a stoichiometric mixture such as a 1:1 ratio of the heavy and light versions of the crosslinker. The mass difference and stoichiometric ratio between the heavy and light versions of the crosslinker results in an isotopic signature with an intensity ratio corresponding to the stoichiometric ratio used and *m/z* difference to the mass difference between the heavy and light versions of the crosslinker. These attributes are useful for the identification of crosslinked species amongst an excess of unmodified peptides. Selected isotope labelled crosslinkers are commercially available, and can also be custom synthesised from isotopically labelled starting materials.[89,87]

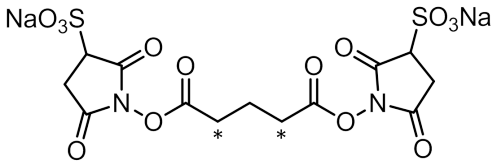


Figure 1.6: Crosslinker BSG-d_4 reported by Muller *et al*, labelled with 4 deuterium atoms at the sites indicated with an asterisk.[90]

Cleavable Bonds

Crosslinkers can be functionalised within the spacer arm to include cleavable bonds, as shown in Figure 1.7. The ability to cleave a crosslinker chemically, or *in vacuo* using gas phase dissociation methods, confers the capacity to separate and sequence constituent peptides using MS/MS methods. Crosslinkers containing cleavable bonds are beneficial for the identification of crosslinked species and the site

specific location of crosslinked residues. This advantage is highlighted notably in the case of type 2 or intermolecular crosslinked species, which can be particularly difficult to analyse due to incomplete or co-fragmentation of each linked peptide upon MS/MS analysis. Separation of the linked peptides enables the individual sequencing of each peptide, and hence the location of crosslink modification can be determined.

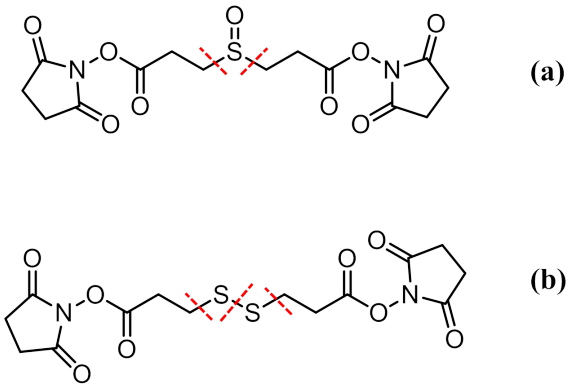


Figure 1.7: **(a)** Crosslinker DSSO reported by Kao *et al*[91] and **(b)** Crosslinker DSP reported by Lomant and Fairbanks.[92] **(a)** is cleavable using gas phase dissociation in positive mode, and **(b)** in negative mode along the dotted lines shown.[93]

The use of chemical cleavage steps such as disulfide reduction to cleave a bond within a crosslinker introduces extra steps into the CXMS workflow. The added complexity of such extra steps can therefore be reduced through the use of *in vacuo* bond cleavage methods in favour of chemical cleavage steps. *In vacuo* cleavable bonds are characterised by an ability to be dissociated within the mass spectrometer using methods such as low energy collision induced dissociation (CID) or electron capture dissociation (ECD) methods. Bond dissociation using low energy CID or ECD reduces the possible competitive fragmentation products from the constituent peptides, such as cleavage of the protein backbone through peptide bonds, which might later be exploited by MS^n approaches for site specific location of crosslinked residues. Many structural investigations have been published using cleavable linkers, including a recent investigation of the mammalian (55S) mitoribosome, which consists of the large (39S) and small (28S) subunits and is one of the largest structures to be investigated using this method.[25] Selected crosslinkers containing gas phase and chemically cleavable bonds are commercially available, and can also be custom synthesised from commercially available starting materials.[50,89,91,93,94,95,96,97]

Affinity Tags

Crosslinkers can be functionalised to include affinity tags, as shown in Figure 1.8. The inclusion of an affinity tag provides the ability to enrich crosslinked peptides from complex peptide mixtures produced by proteolytic digestion of the sample. The most common affinity tag used in crosslinking literature is biotin as it is stable and unreactive throughout the crosslinking process, and allows for the use of avidin-biotin affinity purification. Avidin purification exploits the natural affinity of the proteins avidin, from bird, reptile or amphibian egg whites or streptavidin, from *Streptomyces avidinii*, for biotin. Enriching crosslinked species simplifies the detection and subsequent identification of crosslinked species by removing unmodified peptides (Figure 1.4), increasing the amount of useful data able to be obtained from an experiment. An affinity tagged crosslinker reported by Ilver *et al* is commercially available,[98] however most reported affinity tagged crosslinkers have been custom designed and synthesised.[99,100]

Affinity tagged crosslinking reagents such as BCCL1 (Figure 1.8) are generally bulkier than traditional crosslinkers and some evidence exists suggesting that this may result in the formation of less crosslinks due to steric hindrance, a reduction in the accuracy of spatial constraints and at worst, disruption of native protein structure, resulting in the acquisition of erroneous crosslink data.[101,102] An alternative to affinity purification is the capture and extraction of crosslinked species on a solid support or resin. Capture methods generally require less bulky functional moieties than biotin such as azides or alkynes, which may aid in mitigating issues regarding crosslinker bulk. Capture can be mediated by chemistry such as the Huisgen azide-alkyne cycloaddition reaction[103,104,105,106,107] or capture of sulfhydryl (-SH) moieties using an iodoacetyl group.[108] Captured species can then be cleaved from the support by photoreactive or chemically labile bonds within the crosslinker. Solid supports for capture are commercially available, however the associated crosslinkers are not, requiring custom design and synthesis.

Crosslinkers have also been functionalised with fluorophore labels, which enable the monitoring and separation of crosslinked species by gel electrophoresis, and fluorescence based detection during separation of crosslinked peptides by high performance liquid chromatography (HPLC), amongst other techniques. Fluorophore labelled crosslinkers are commercially available, and can also be custom synthesised from commercially available starting materials.[109,110]

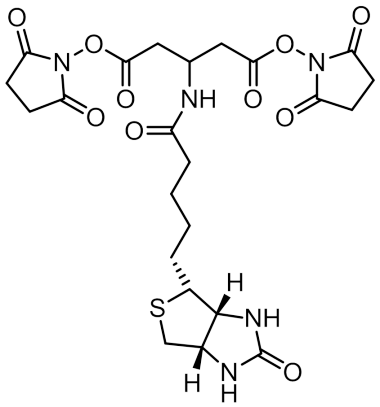


Figure 1.8: Crosslinker BCCL1 reported by Kang *et al*[99], containing a biotin affinity tag.

Combinations of functionality

Combinations of isotope labelling, cleavable groups and affinity tags within a single crosslinker design have been reported in the literature. Combination crosslinkers incorporate two[111,112,113] or in a few instances, all three of the functionalisation types,[114,115] as shown, for example, in Figure 1.9. Combination crosslinkers are advantageous as the benefits of the affinity tag, cleavable bond and/or isotope label are brought together and utilised within a single design, resulting in a crosslinker which is able to address more than one of the analytical challenges associated with CXMS. Combination crosslinkers are not commercially available, hence experiments endeavouring to use the linkers require synthesis in-house.

The structures of combination crosslinkers reported in literature are decidedly variable, and as a result of this, the procedures used to synthesise the linkers vary considerably, and changes to the fundamental structure to suit a particular experiment require new synthetic protocols. Combination crosslinkers also possess inherently complex structures in comparison to traditional reagents, as more than one functional moiety must be included, which can take many synthetic steps to assemble. The lack of commercial availability and the number and complexity of the synthetic steps required to make combination crosslinkers has restricted the use of the linkers to a small number of studies, despite the advantages of their use.

The research presented in this thesis aimed to define a modular crosslinker structure, able to be easily modified to suit different experiments, and to design and implement the corresponding synthetic protocol for the synthesis of purpose designed crosslinkers.

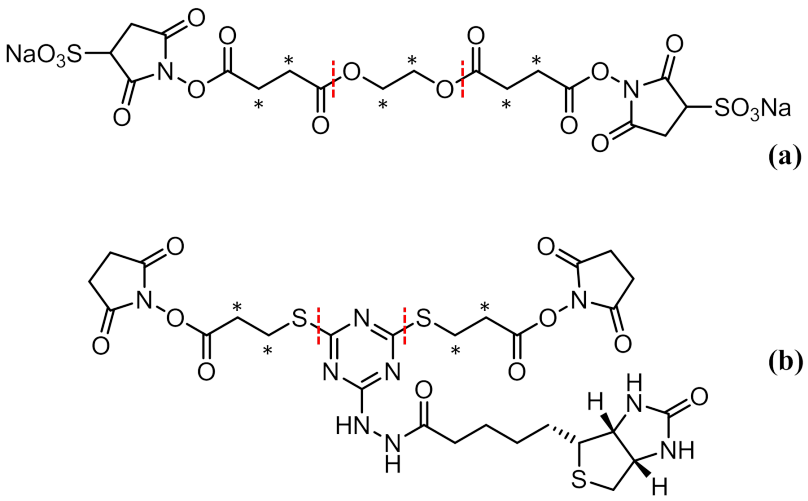


Figure 1.9: **(a)** Crosslinker EGS-d_{12} and **(b)** CBDPS reported by Petrochenko *et al.*[116,114] **(a)** is labelled with 12 deuterium atoms at the sites indicated with an asterisk, and is cleavable by MS in positive mode along the dotted lines shown. **(b)** is labelled with 8 deuterium atoms at the sites indicated with an asterisk, is cleavable by MS in positive mode along the dotted lines shown, and contains a biotin affinity tag.

Chapter two details the definition of a modular crosslinker structure, the design of the corresponding synthetic protocol and the synthesis of a simple crosslinker to establish the method. Chapter three describes the synthesis of two purpose designed combination crosslinkers containing affinity tags and cleavable bonds, and the establishment of a crosslinking assay using the *Staphylococcus aureus* biotin protein ligase enzyme to determine the crosslinking efficacy of any linkers synthesised from the modular method.

1.4 Mass Spectrometry

Fundamental understanding of the CXMS technique and the role of mass spectrometry is essential for the development of more effective chemical crosslinking reagents. The basic principles of mass spectrometry, tandem mass spectrometry, collision induced dissociation and the affiliated instrumentation provide insight into crosslinker design.

Mass spectrometry is an analytical technique in which samples are ionised, separated and detected based upon their mass to charge ratio within a mass spectrometer. It is used ubiquitously throughout analytical chemistry and biology in research and

commercial applications. Mass spectrometers traditionally consist of three major parts, the ion source, mass analyser and a detector. All tandem mass spectrometry data reported in this thesis was obtained using a Waters Q-ToF2 mass spectrometer (Figure 1.10).

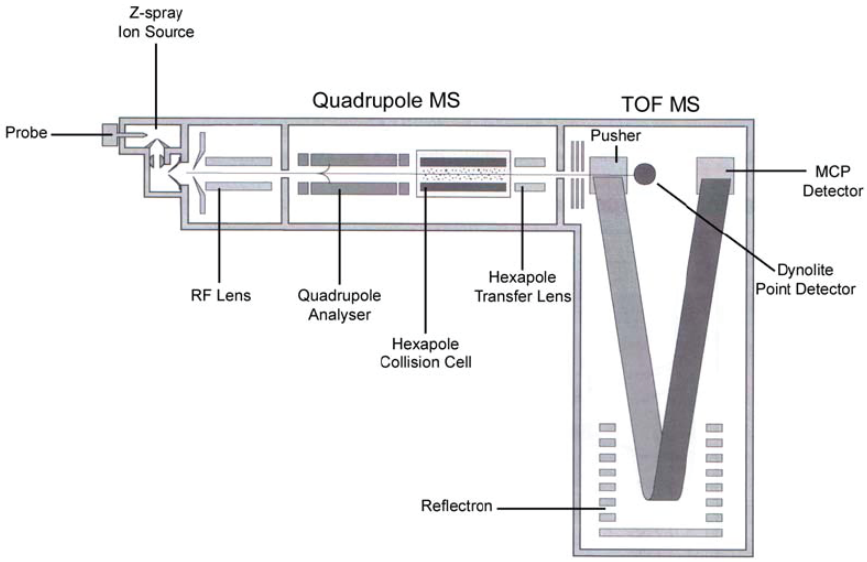


Figure 1.10: Schematic diagram of the Waters QToF2 tandem mass spectrometer, displaying the sequential order of the mass analysers and the hexapole collision cell.

1.4.1 The Ion Source

MS relies on the formation of gas phase ions and their subsequent separation and detection. The ion source is the component of the mass spectrometer responsible for the ionisation of analyte molecules into the gas phase. Many ionisation techniques exist, and are classified as hard or soft, depending on the degree of fragmentation of analyte molecules observed during the ionisation process. Hard ionisation techniques such as electron ionisation (EI) and fast atom bombardment (FAB) cause high levels of analyte fragmentation and produce more complex output spectra. However, FAB can be useful for the ionisation of less volatile compounds. Soft ionisation techniques such as Electrospray Ionisation (ESI) and Matrix Assisted Laser Desorption Ionisation (MALDI)[36,37,38] are suitable for the ionisation of less volatile biological samples and produce less sample fragmentation. The advent of ESI and MALDI have fostered the development of MS into a technique which can be used to study different aspects of macromolecular structures important in biology (Section 1.2.1).

Electrospray Ionisation

The ESI technique for MS was pioneered by John Fenn in the 1980s and is currently the most widely used ionisation method to study biological molecules.[34,35] ESI is used to ionise analyte samples from aqueous and organic solvents such water, methanol and acetonitrile.[117] Ionisation can be further facilitated through the formation of $(M+H)^+$ and $(M-H)^-$ ions in solution using compounds such as acetic acid and triethylamine respectively.

To effect solvent removal and ionisation, the sample solution is passed through a small capillary tube to which a strong positive or negative electric potential is applied (1-6 kV) (Figure 1.11). Upon reaching the end of the tube the strong electric field causes sample vapourisation into a spray of small charged droplets, forming a structure known as the Taylor cone.[18,118] Further solvent removal is effected through the flow of a drying gas such as N_2 past the end of the capillary tube, and/or heating of the sample, causing a reduction in droplet size. As the droplets reduce in size the surface charge density increases, resulting in strong coulombic repulsion between like charges. The charged droplets eventually reach a size at which the resultant coulombic repulsion overcomes the surface tension of the droplet, termed the Rayleigh limit, which causes the release of charged and neutral molecules. The charged molecules (ions) are sampled by the skimmer cone and accelerated into the mass spectrometer.[119]

ESI allows the sample to retain minimal internal energy, reducing the amount of analyte fragmentation in the ion source, and enables retention of non-covalent information from solution into the gas phase.[120] ESI is also unique as multiply charged ions can be produced, allowing the study of complexes of high molecular weight, effectively extending the mass range of the mass spectrometer from kDa to MDa.[119,117]

Nano-electrospray ionisation (nESI) is a variant of the ESI technique developed by Wilm and Mann in the 1990s.[121,122] A small glass capillary coated with a conductive material such as platinum or gold is used to load the sample, usually requiring only 1-5 µL of sample. nESI utilises low sample flow rates of approximately 20 to 50 nL/min in comparison to the µL/min flow rate of ESI. The reduced flow rates generate smaller droplet sizes upon initial vapourisation, improving ionisation efficiency. Increased salt tolerance, smaller sample volumes and concentrations, reduced flow rates and hence greater ionisation efficiency highlight nESI as a desirable ionisation technique for the study of biological molecules.[123]

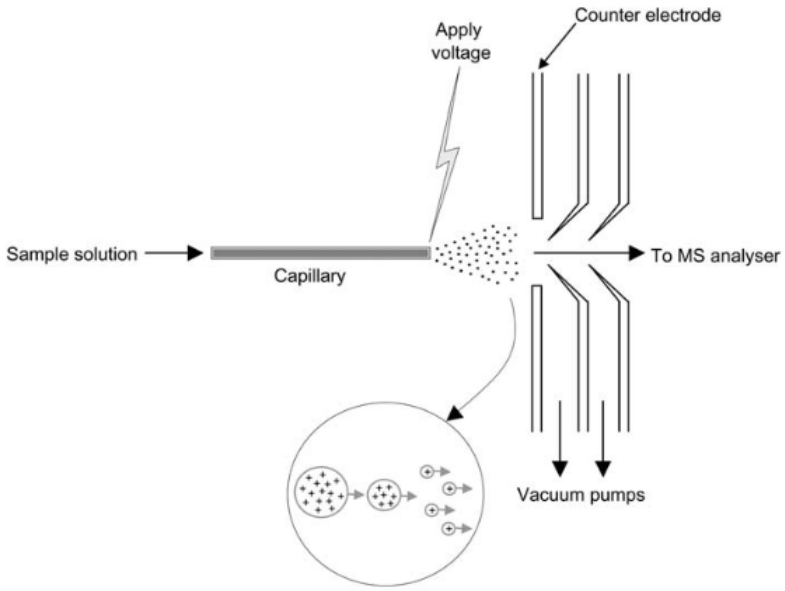


Figure 1.11: Diagrammatic representation of the ESI process.[118] Sample is vapourised by a strong electric field, upon reaching the Rayleigh limit, charged ions are released from the droplets, sampled by the skimmer cone and accelerated into the mass analyser.

1.4.2 Mass Analysers

Once samples have been ionised the mass analyser facilitates ion separation based upon mass to charge ratio for detection. Various mass analysers have been developed for this purpose including the quadrupole[124], time of flight (ToF)[125], ion cyclotron resonance, ion trap and orbitrap mass analysers. The Waters QToF2 instrument contains two mass analysers, the quadrupole and ToF, enabling MS/MS experiments.

Quadrupole

The quadrupole mass analyser developed by Paul and Steinwedel in 1953[124], also known as a quadrupole mass filter, is capable of filtering and selecting sample ions based upon mass to charge ratio. The quadrupole consists of four parallel cylindrical rods to which direct current (DC) and radio frequency (RF) alternating current voltages are applied (Figure 1.12). Each rod is paired electrically with a diagonal partner and an equal but opposite DC or RF voltage is applied to each pair. The resulting electric field causes the ions to travel through the quadrupole in an oscillating motion. The amplitude of the oscillation is related to the m/z ratio, which can be

controlled by changing the rod voltages. The rod voltages can be changed to ensure that the desired ion has a 'stable' oscillation and hence traverses the quadrupole, and that the oscillation of unwanted ions is destabilised, resulting in ejection from the quadrupole.[119] Other roles of the quadrupole include functioning as an ion guide, facilitating ion transfer to other sectors of the mass spectrometer, or as a collision cell using an inert collision gas such as argon (Ar).

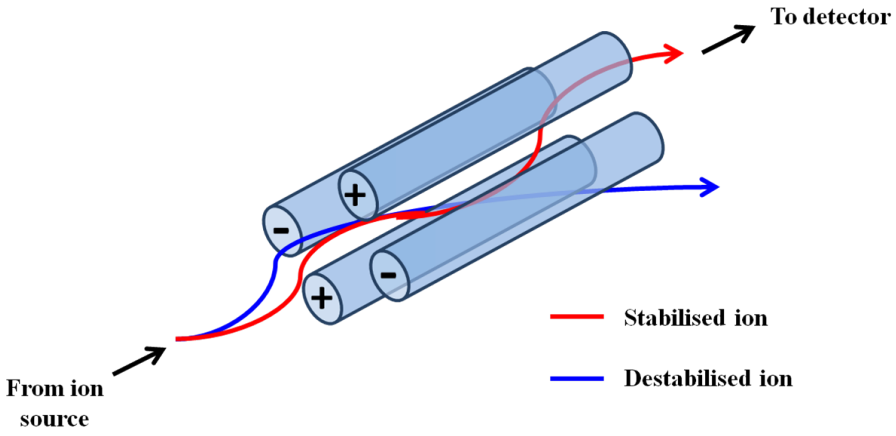


Figure 1.12: The quadrupole mass analyser. The oscillation of the ion in red has been stabilised by the electric fields, resulting in the ion traversing the quadrupole successfully. The oscillation of the blue ion has been destabilised by the electric fields, resulting in ejection from the quadrupole.

Time of flight (ToF)

The time of flight mass analyser connected to a pulse ion source was first developed by William. E. Stephens in 1946, and was designed to function without the need for a magnetic field.[125] Time of flight mass analysers separate ions based upon the time taken for an ion to traverse a known distance through a defined electric field.[117] Ions are accelerated using an electric field of defined strength, resulting in equal kinetic energy values for ions with the same charge. The velocity of the ion depends upon the mass to charge ratio, hence ions with a smaller m/z ratio will have greater velocity than ions with larger m/z ratios. As a result of this, the time taken for an ion to traverse a defined distance within the ToF analyser can be related to the m/z ratio. However, proximity to the accelerating electric field dictates how much energy an ion receives when an ion pulse enters the ToF, resulting in ions with the same m/z ratio possessing different kinetic energies, therefore traversing the ToF at different rates and lowering the instrument resolution. Introduction of the reflectron improved the resolution of ToF instruments by correcting the difference in

velocity of ions with the same m/z ratio (Figure 1.10).[126] Following separation in the ToF analyser, ions are detected using a micro-channel plate (MCP) detector. Each microchannel acts as an electron multiplier, the impact of an ion with the wall of a microchannel causes a cascade of electrons to propagate down the channel, causing the release of more electrons, magnifying the signal by up to 6 orders of magnitude. The signal can then be analysed by the time to digital converter to yield the m/z ratio.

Tandem Mass Spectrometry (MS/MS)

Tandem mass spectrometry involves the selection of an ion and sequential dissociation and separation of the constituent fragments in another mass analyser stage. The Waters QToF2 tandem mass spectrometer contains an array of quadrupole and ToF mass analysers separated by a hexapole collision cell, for analyte fragmentation (Figure 1.10). The desired ion is filtered using the quadrupole mass analyser, and subsequently fragmented using CID within the hexapole collision cell. The constituent fragments are then separated in the ToF mass analyser and detected. MS/MS experiments have been used extensively in structural biology to determine the linear sequence of amino acids within a protein structure (Section 1.2.1), and burgeoning techniques such as CXMS rely heavily on MS/MS analysis. Constituent peptides must be fragmented in the gas phase to enable *de novo* peptide sequencing. Fragmentation of analyte ions can be achieved through collision induced dissociation (CID), electron capture dissociation (ECD), electron transfer dissociation (ETD) and photodissociation processes amongst many other techniques.

Collision Induced Dissociation (CID)

Collision-induced dissociation is one of the most commonly used analyte dissociation techniques, and is employed by the Waters QToF2 mass spectrometer. Within the collision cell, the selected precursor ion is accelerated and allowed to collide with neutral gas molecules or collision gases such as argon, helium and nitrogen. The collision causes some of the kinetic energy which the ion possesses to be converted into internal energy, resulting in fragmentation of the precursor ion into smaller fragments which can yield structural information about the precursor ion.

Collisional activation of peptides produces characteristic dissociation around the amide bond between amino acid residues. In positive mode, if the charge is retained on the N terminal fragment the ions are referred to as a, b or c ions, and if the charge is retained on the C terminal fragment the ions are referred to as x, y or z ions using nomenclature developed by Roepstorff and Fohlman[127] and adjusted by Biemann.[128] The b and y ions are the most common fragmentation products observed by CID in positive mode and can be used to sequence fragmented peptides from the N and C termini (Figure 1.13).

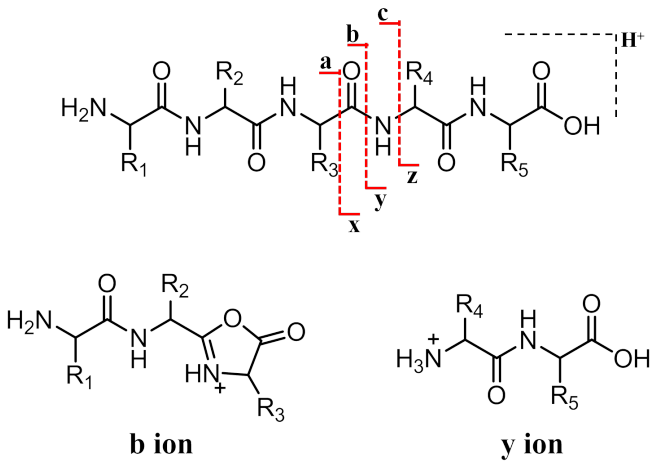


Figure 1.13: Fragment ions expected from a peptide dissociated in positive mode using CID.[129] The **b** and **y** ions are the most abundant backbone fragmentation ions observed upon peptide fragmentation using CID, where the **b** ions cyclise to form an oxazalone.

Peptide sequencing in negative ion mode has not been as widely utilised as sequencing in positive mode due to the greater complexity of the mass spectra obtained. In negative mode some of the most common fragmentation pathways correspond to side chain cleavages, rather than dissociation of the amide bond. However, dissociation of the amide bonds between amino acid residues in negative mode most commonly yields β and α ions in a very similar process as the formation of b and y ions in positive mode (Figure 1.14). The β and α ions can hence be used to determine the peptide sequence from the N and C termini. Analysis using negative ion mode does retain some advantages, namely the identification of post translational modifications, mapping of disulfide bonds, the characterisation of parts of a sequence not identified in positive mode and differentiation between aspartate and asparagine, glutamate and glutamine residues, and glutamine and lysine residues.[130,131,132,133]

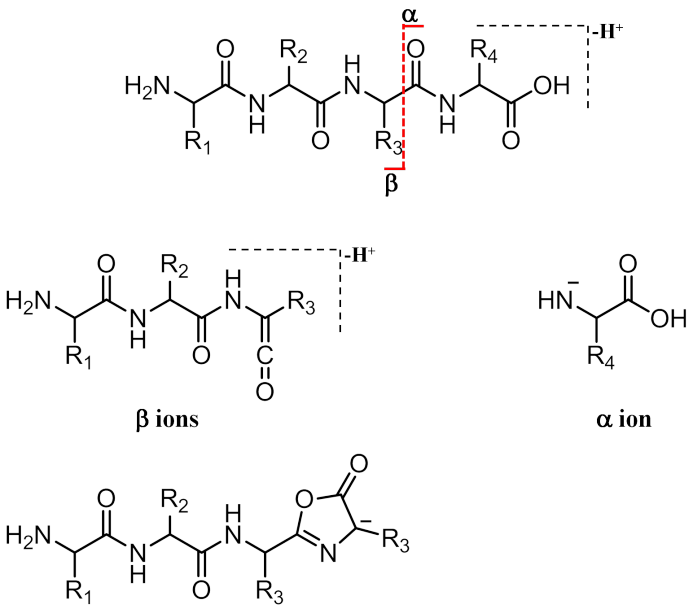


Figure 1.14: Fragment ions expected from a peptide dissociated in negative mode using CID.[130] The α and β ions provide analogous information to the **b** and **y** ions in positive mode.

Chapter 2

Design of a modular synthetic protocol for crosslinker diversification

2.1 Introduction

Widespread use of the CXMS technique has been hindered primarily by the analytical challenges summarised in Section 1.3.2.[50,29] Site specific location of crosslinked residues can be challenging due to complex MS fragmentation behaviour, and detecting and identifying crosslinked peptides within complex mixtures can also be problematic as crosslinked species are generally of low relative abundance.

Basic crosslinker design consisting of a spacer arm and reactive groups has been studied and improved by the addition of affinity tags, cleavable bonds and isotope labels in an attempt to mitigate some of the analytical challenges currently facing the CXMS technique. The use of these functionalities has been investigated as briefly summarised in Section 1.3.3. Crosslinkers such as EGS-d_{12},[116] which contains both isotope labels and cleavable bonds, provide dual functionality Figure 1.9(a). This then allows the crosslinker to address two analytical challenges within the same experiment, namely the identification of crosslinked species within complex mixtures using the isotopic signature generated by the use of deuterium labelling, and the site specific location of crosslinked residues through the utilisation of the cleavable bonds (Section 3.1.1, p. 48).

Crosslinkers containing different combinations of isotope labels, affinity tags and cleavable bonds have been reported in literature,[111,112,113,114,115] however samples are not commercially available. As a result, in-house synthesis is required. Reported structures and synthetic protocols for these crosslinkers are not able to be easily modified to suit varied experimental design. For example, the inclusion of an affinity tag such as biotin into the structure of EGS-d_{12}, would require significant structural and synthetic changes from the original design. Thus, the number of structural studies utilising crosslinkers containing more than one functionality has been limited by the requirement for in-house synthesis and inflexibility of reported designs. The development of a modular crosslinker structure and corresponding synthetic protocol able to be easily modified to suit experimental requirements would enable more widespread use of combination crosslinkers.

2.1.1 Aims

The synthesis of crosslinkers containing multiple functionalities for use in structural studies require a basic modular crosslinker structure. This must be easily modified to include isotope labels, affinity tags and cleavable bonds. This work aims to outline a suitable method for the synthesis of such crosslinkers. The viability of the synthetic method will be demonstrated through the synthesis of selected, purpose designed crosslinkers, including **1a** (Figure 2.8), to establish the modular design and corresponding synthesis as a suitable pathway for crosslinker modification.

2.2 Design of a modular synthetic protocol

Addressing the first aim of designing a synthetic protocol which allows assembly of different modules requires the definition of a basic crosslinker structure which can be modified to include affinity tags, isotope labels and cleavable bonds whilst retaining the ability to form covalent linkages between protein residues. The first two modules to be included in the design, reactive groups and a spacer arm, underpin essential crosslinker function. Incorporation of reactive groups at each end of the crosslinker is crucial, as they facilitate covalent bond formation between the crosslinker and functional groups of proteins. The spacer arm acts as a defined distance constraint or 'molecular ruler', establishing an upper limit on how far apart residues must be within a three dimensional protein structure to be crosslinked.

The opportunity to form covalent bonds with different protein residues can be afforded by using different reactive groups. Probing distance constraints between different protein residues can hence be permitted, providing a more comprehensive interaction map, and enabling the study of a more diverse range of protein structures not limited to those containing a high abundance of a single amino acid such as lysine. Reactive group identity could include N-hydroxy succinimide (NHS) or imido esters,[134,135] to react with amine side chains (-NH_2), maleimides to react with sulfhydryl groups (-SH),[136,137] and non-specific photoreactive diazirines,[138,139] amongst others (Figure 2.1).

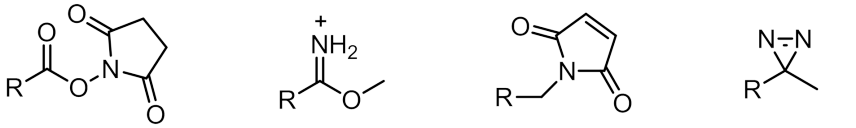
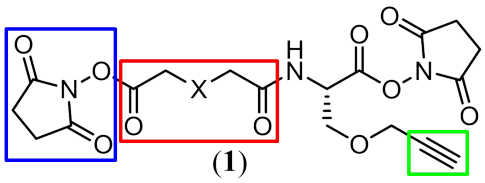


Figure 2.1: Several types of reactive group which can be used to crosslink protein functional groups. R indicates attachment to a crosslinker spacer arm.

The spacer arm provides an opportunity for the incorporation of an isotopic label and/or cleavable bond within the modular design whilst also allowing for variation of crosslinker length. The length of a crosslinker can be modified by changing the number of atoms constituting the chain between the two reactive groups. A range of starting materials are commercially available with appropriate functional groups to facilitate the inclusion of isotope labels and cleavable bonds within different chain lengths. For example, variants of different amino acids can be acquired commercially with deuterated or ^{1}3-C sites such as L-Serine-1-^{1}3-C and d3-iodomethane can also be utilised to introduce labelled components through methylation.

The affinity tag is difficult to incorporate into the spacer arm as few starting materials are commercially available containing a combination of an affinity tag such as biotin and an isotopic label or cleavable bond. As a result, the affinity tag should be attached via a functional group within the spacer arm, such as an amine, through peptide coupling. This is exemplified in crosslinker BCCL1 reported by Kang *et al*[99] (Figure 1.8) which utilised an amine within the spacer arm to facilitate biotin attachment. Difficulties in the detection and identification of crosslinked species highlight the importance of the ability to incorporate a tag into the crosslinker design, which conveys the ability to increase the concentration of crosslinked peptides in solution. Hence here an affinity tag is incorporated as a separate module such that it can be included or excluded from the overall synthesis as required within the crosslinker design.

Crosslinker **1** (Figure 2.2) represents the final general design upon which the modular synthesis will be based as a result of defining the reactive group, spacer arm and affinity tag modules.



(1)

Figure 2.2: Crosslinker **1**, including the reactive group(s) (blue), spacer arm (red) and affinity tag (green) modules, where X represents the site of spacer arm modification.

2.2.1 Retrosynthetic analysis of the target crosslinker

The avidin-biotin interaction has been successfully used to purify crosslinked peptides from complex mixtures,[107,99,114] and hence is utilised in the design of general crosslinker **1**, incorporating biotin as the affinity tag shown in Figure 2.3. However, work by Bobofchak *et al* and Paoli *et al* suggests that a large or bulky crosslinker can adversely affect the formation of covalent linkages during the crosslinking process (1.3.3, p. 15). Hence the incorporation of biotin into the crosslinker structure before protein crosslinking could result in the acquisition of erroneous crosslinking data. As a result, the procedure used to attach biotin to the crosslinker was devised such that it could be achieved after tryptic digestion of crosslinked protein within the CXMS workflow. This allows for flexibility within the crosslinker design for spacer arm modules of differing lengths and/or proportions to be included, without the added bulk of biotin possibly affecting the quality of crosslinking data. Tryptic digestion of crosslinked protein is typically done in an aqueous environment in the presence of buffers such as phosphate-buffered saline (PBS). To avoid buffer exchange of crosslinked peptides from an aqueous environment into an organic solvent to facilitate biotin attachment, the efficiency of the CXMS workflow would be maximised through the use of a chemical process able to proceed under aqueous conditions. Huisgen cycloaddition[103] of crosslinker **1** and biotin was one such technique which could facilitate bond formation in an aqueous environment, requiring azide and alkyne functional groups.

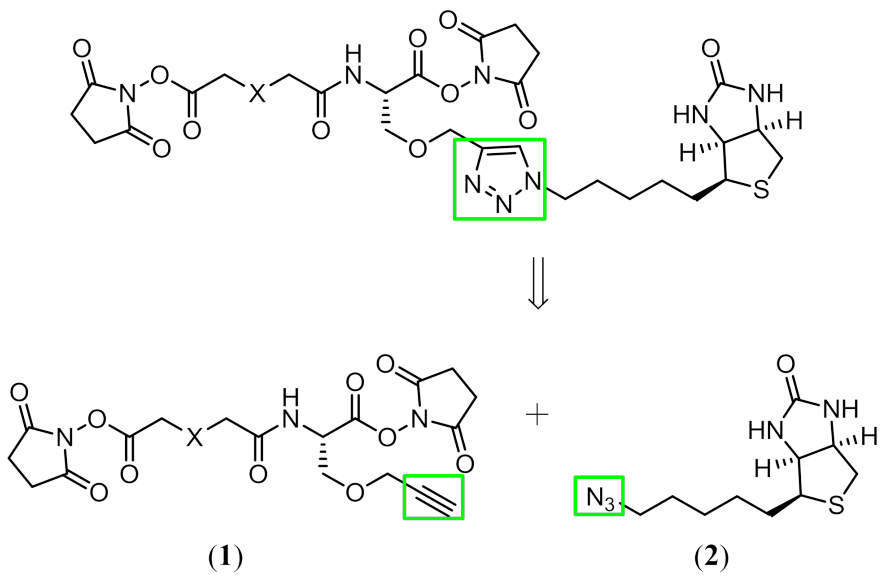


Figure 2.3: Retrosynthesis of the general biotinylated crosslinker containing the 1,4 triazole ring (green) facilitating the attachment of the affinity tag, into its constituent precursors crosslinker **1** containing the alkyne and biotin azide (2).

An advantage of Huisgen azide-alkyne cycloaddition is its biological compatibility.[140] Limited reactivity between the azide of biotin **2** and alkyne of crosslinker **1** and the functional groups present on the crosslinked peptides should be observed due to their inert nature and lack of compatible reaction partners.[141] The use of this reaction to attach biotin to crosslinked peptides in an aqueous environment has been reported.[107]

The azide of biotin was chosen as it can be readily synthesised using a reported method in five steps from commercially available biotin carboxylic acid (Figure 2.4).[142] Hence, the alkyne functional group was assigned to crosslinker **1**. However, it may also be possible to synthesise the azide of crosslinker **1** and alkyne of biotin using commercially available starting materials and similar methods.

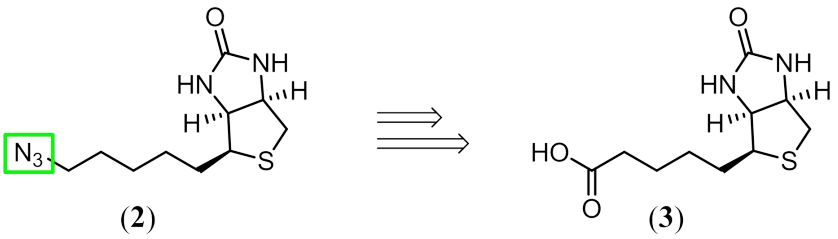


Figure 2.4: Retrosynthesis of biotin azide **2** (green) from commercially available biotin carboxylic acid (3).

Crosslinker **1** contains the spacer arm, reactive groups and the alkyne moiety. NHS esters are used most frequently in protein crosslinking studies and hence were chosen as the reactive group modules to be used in the design of **1** (Section 1.3).[29] Crosslinker **1** can be formed from diacid **4** as introduction of NHS can be readily achieved through activation of carboxylic acids. Diacid **4** can be synthesised from the corresponding dimethyl ester **5** (Figure 2.5). Dimethyl ester **5** provides a better alternative for purification on normal phase silica than diacid **4** as it is less polar.

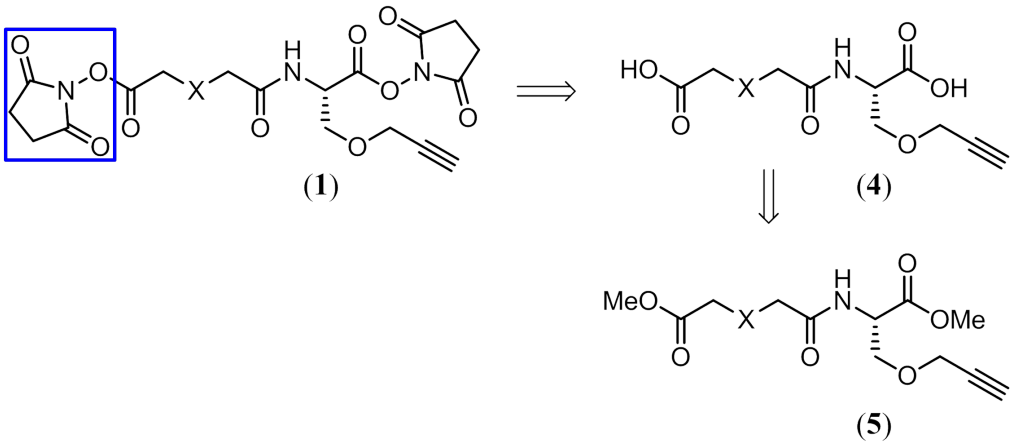


Figure 2.5: Retrosynthesis of crosslinker **1** from the precursor, dimethyl ester **5**, through diacid **4**. Reactive group(s) are shown (blue).

Diester **5** contains the spacer arm and alkyne functional group. The structure of the spacer arm is not specified at position 'X' to convey the ability to vary its structure to include a range of isotope labels and/or cleavable bonds. Connection of the spacer arm to the alkyne module requires the formation of a bond which is inert under the reaction conditions used during methyl ester hydrolysis, introduction of NHS and Huisgen cycloaddition to prevent decomposition. Attachment could be facilitated by the formation of a carbon-carbon bond using Wittig and Grignard reactions, a carbon-oxygen bond using a condensation reaction to produce an ester or an amide bond using peptide coupling methods. Amide bonds are formed from carboxylic acid and amine precursors, using a wide range of well characterised coupling reagents and reactions.[143,144] Amides do not require the synthesis of phosphorous ylides or organomagnesium reagents used by Wittig and Grignard reactions and are also more stable than esters due to resonance stabilisation. Hence **5** could be formed by amide coupling of hydrochloride salt **6** containing the alkyne to the spacer arm. The spacer arm then consists of a monomethyl ester which are commercially available in a wide range of structures, an ideal additional benefit arising from the spacer arm design. The use of a monomethyl ester is necessary to prevent the formation of amide bonds at both ends of the structure, termed a 'diaddition' product.

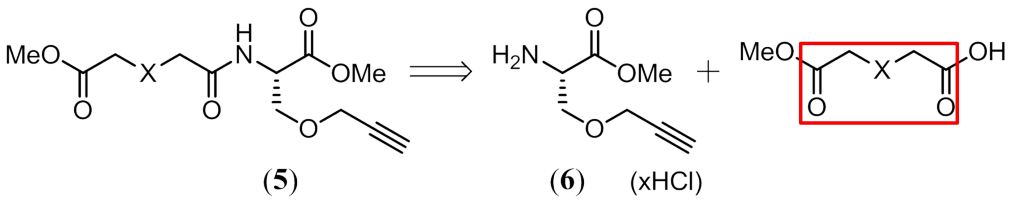


Figure 2.6: Retrosynthetic analysis of dimethyl ester **5**. The precursor to the spacer arm is shown (red).

Hydrochloride salt **6** is readily synthesised using a two step reported procedure from the commercially available amino acid serine, in N-Boc protected form **7**.[145] Serine provides the simplest component from which to build the affinity tag module due to the pre-existing functional group structure, ready availability, reasonable cost and well characterised reactivity.[146,147,148] **7** contains a protected amino (N) terminus to be used in amide coupling to the spacer arm, a carboxyl (C) terminus to be functionalised with NHS and an alcohol to facilitate the attachment of the alkyne tag using *O*-alkylation. The structure of the alkyne tag can be varied using commercially available alkyl halides to include a longer carbon chain or modified PEG linker and/or a cleavable bond, which could be used specifically in a solid resin support capture method as an alternative to the avidin-biotin affinity purification process (Section 1.3.3, p. 15). A different amino acid such as lysine, cysteine or tyrosine could also be used in place of serine, utilising the same peptide coupling methods, to suit design requirements.

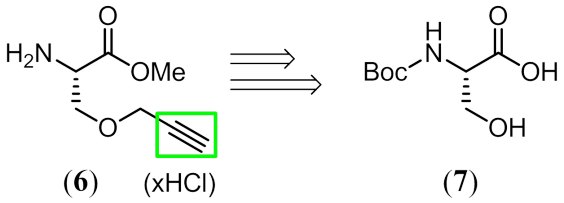


Figure 2.7: Retrosynthetic analysis of hydrochloride salt **6** with alkyne shown (green).

Overall, retrosynthetic analysis of crosslinker **1** suggested that it could be synthesised from the commercially available reagents serine **7**, the desired alkyl halide and monomethyl ester and NHS in a minimum of five synthetic steps. The biotin azide cycloaddition partner is also able to be synthesised from commercially available biotin carboxylic acid in five synthetic steps.

2.3 Crosslinker 1a

Crosslinker **1a** was initially targeted in order to demonstrate the viability of the synthetic protocol described in Section 2.2 (Figure 2.8). The structure includes all three modules, with a simple alkyl chain as the spacer arm. Crosslinker **1a** was the first design to be implemented using the synthetic protocol, hence the validation of each synthetic step was simplified by excluding cleavable bonds and isotope labels from the spacer arm to allow straightforward characterisation by NMR and to limit the possibility of side reactions or degradation processes occurring.

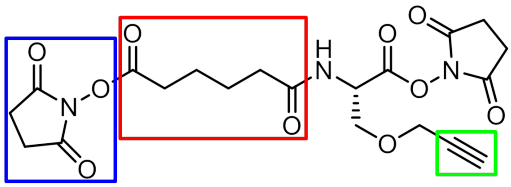


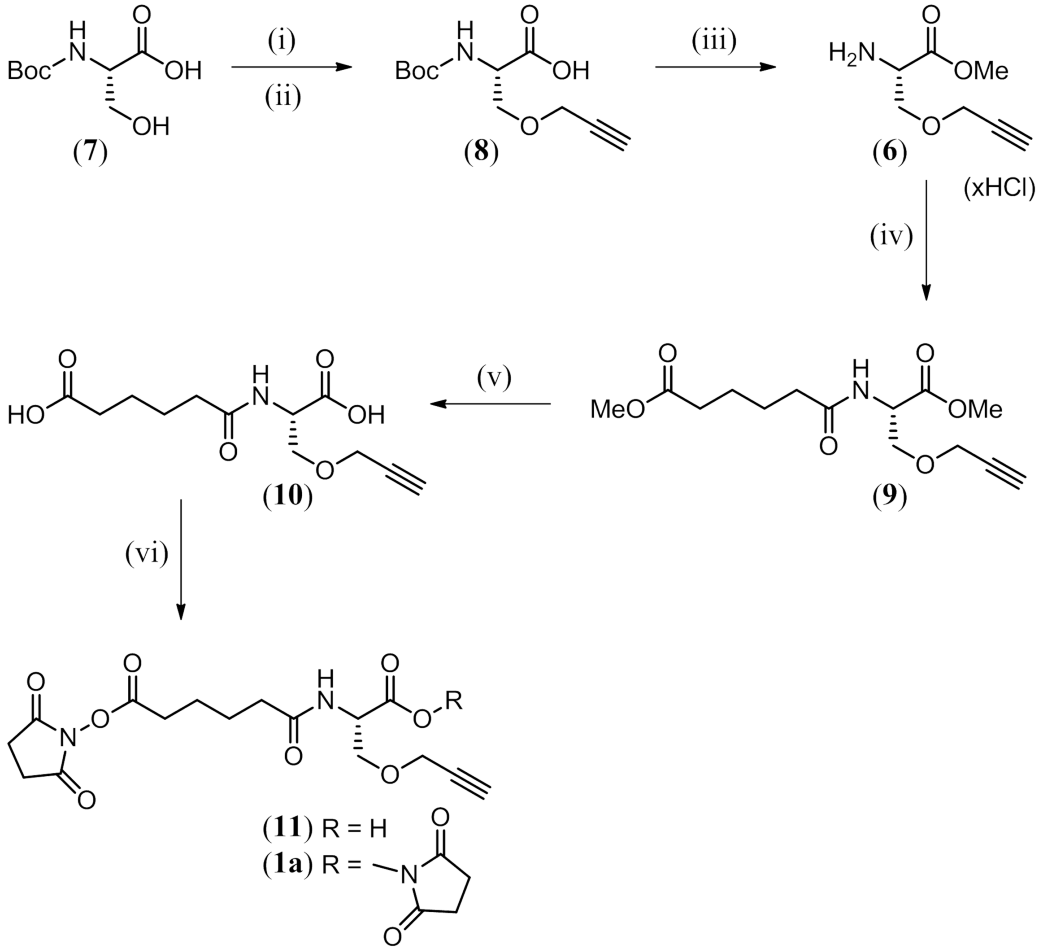
Figure 2.8: Crosslinker **1a**

Crosslinker **1a** addresses the current analytical challenges of CXMS pertaining to the detection and identification of crosslinked peptides through the inclusion of an affinity tag.

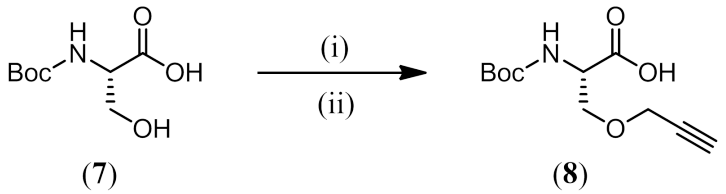
2.3.1 Synthesis of Crosslinker 1a

Crosslinker **1a** was prepared as shown in Scheme 2.1.

The synthesis of the hydrochloride salt **6** began with *O*-alkylation of commercially available Boc-L-Serine, **7**. The alcohol of **7** was deprotonated using sodium hydride and the resultant anion was alkylated with propargyl bromide, to give alkyne **8** in good yield of 91% without further purification (Scheme 2.2). The ^{1}H NMR spectrum of **8** showed a characteristic triplet corresponding to the alkyne hydrogen at 2.47 ppm with all ^{1}H NMR signals corresponding to data reported in the literature.[149]



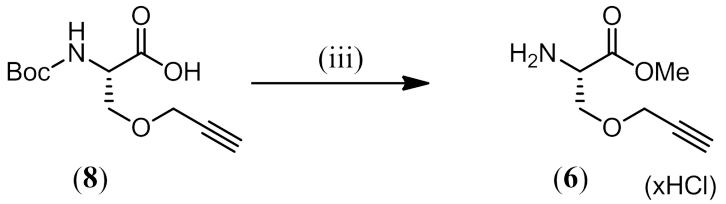
Scheme 2.1: Synthetic scheme for crosslinker **1a**. *Reagents and conditions*: (i) NaH, DMF, (ii) propargyl bromide, (iii) MeOH, $SOCl_2$, (iv) monomethyl adipate, anh. DMF, DIPEA, HATU, (v) 1.6 M NaOH, THF, MEOH, (vi) TFA-NHS, pyridine.



(**7**) (**8**)

Scheme 2.2: *O*-alkylation of Boc-L-Serine **2** with (i) NaH, DMF, (ii) propargyl bromide to give alkyne **8** (91%).

Alkyne **8** was then treated with thionyl chloride in methanol to remove the N-Boc protecting group and convert the acid into the acid chloride which reacts *in situ* with methanol to give hydrochloride salt **6** in a good yield of 92%, without further purification (Scheme 2.3). The ^{1}H NMR spectrum of **6** showed a characteristic singlet at 3.76 ppm corresponding to the methyl ester and the absence of the N-Boc hydrogen singlet at 1.46 ppm observed for **8**. All ^{1}H NMR signals corresponded to the data reported in literature.[145] The alkyne module was thus successfully completed to give hydrochloride salt **6**.

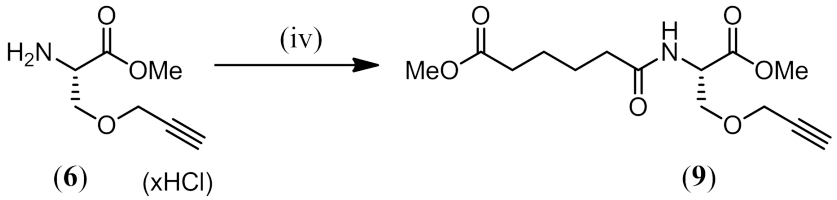


(**8**) (**6**) (xHCl)

Scheme 2.3: One pot N-Boc cleavage and esterification of **8** with (iii) MeOH, $SOCl_2$ to give the hydrochloride salt **6** (92%).

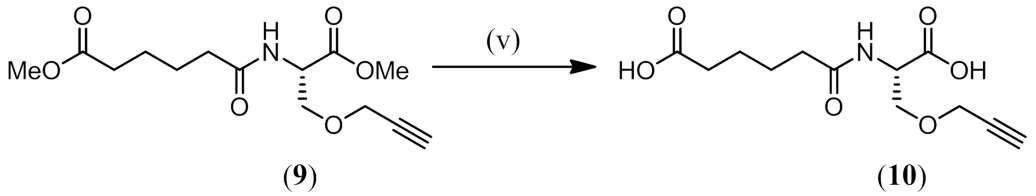
Commercially available monomethyl adipate was used as the spacer arm to couple to hydrochloride salt **6** using amide bond formation (Scheme 2.4). Coupling was attempted in the presence of the racemisation suppressant HOBt[150] in both DMF and DCM, however the low solubility of **6** in DCM gave starting material only, based upon ^{1}H NMR analysis. A similar coupling in DMF gave the desired diester **9**, however purification was problematic as the HOBt could not be completely removed through successive base washes using sodium bicarbonate in the work-up, and consistently co-eluted with **9** on a normal phase silica column in all solvent systems tested. The reaction was repeated in the absence of HOBt and this gave higher pure yields of **9** at 51% in comparison to 36% when HOBt was used. The difference in yield can most likely be attributed to the ability to purify **9** more efficiently without HOBt. Hence, dimethyl ester **9** was synthesised from hydrochloride salt **6** and monomethyl adipate in moderate yield of 51% using the coupling agent HATU and DIPEA in anhydrous DMF. The ^{1}H NMR spectrum of **9** showed the

two expected methyl ester singlets at 3.73 and 3.62 ppm, and the appearance of two triplets at 2.29 and 2.24 ppm and a multiplet at 1.64 ppm, corresponding to the four methylene groups.



Scheme 2.4: Coupling of alkyne **6** to the spacer arm using (iv) Monomethyl adipate, anh. DMF, DIPEA, HATU to give dimethyl ester **9** (51%).

Once purified, dimethyl ester **9** was then treated with sodium hydroxide in methanol and THF to give diacid **10** in a good yield of 72% (Scheme 2.5). ^{1}H NMR data showed a characteristic broad singlet at 12.29 ppm corresponding to the carboxylic acid resonances, and the absence of methyl ester singlet peaks at 3.73 and 3.62 ppm observed for **9**.



Scheme 2.5: Hydrolysis of dimethyl ester **9** using (v) 1.6 M NaOH, THF, MEOH to give diacid **10** (72%).

The introduction of NHS esters to diacid **10** was attempted using well characterised coupling conditions mediated by the carbodiimide based coupling reagents EDC and DIC (Table 2.1).[50,151,152] Activation of the diacid required an excess of 1.2 eq of coupling agent, and 2.0 eq of NHS per carboxylic acid. Introduction of NHS was initially attempted using EDC in DMF, however a complex mixture of starting material and by-products was observed by ^{1}H NMR, and **1a** could not be detected by HRMS. The solvent was then changed to anhydrous 1:1 acetone/DCM which was easier to remove *in vacuo*, enabling the analysis of crude ^{1}H NMR before the work up step was attempted. However the same result was obtained and **1a** was again not detected by HRMS. Following the use of EDC, the coupling reagent DIC was used in THF, as reported by Heng *et al.* However, HRMS and ^{1}H NMR analysis upon removal of THF gave the same mixture of starting material and by-products as the EDC mediated coupling.

An alternative method for the introduction of NHS esters was then attempted through the use of N-trifluoroacetoxy succinimide.[153,89] Unlike carbodiimide based coupling reagents which facilitate coupling through the formation of a reactive, electrophilic *O*-acylisourea intermediate,[154] N-trifluoroacetoxy succinimide is an activated form of NHS in which the trifluoroacetate substituent can as a good leaving group, facilitating nucleophilic attack on the NHS nitrogen from the lone pair of the alcohol oxygen of the carboxylic acid. The electrophile-nucleophile role reversal of the carboxylic acid functional group made this procedure a good alternative to carbodiimide mediated methods.

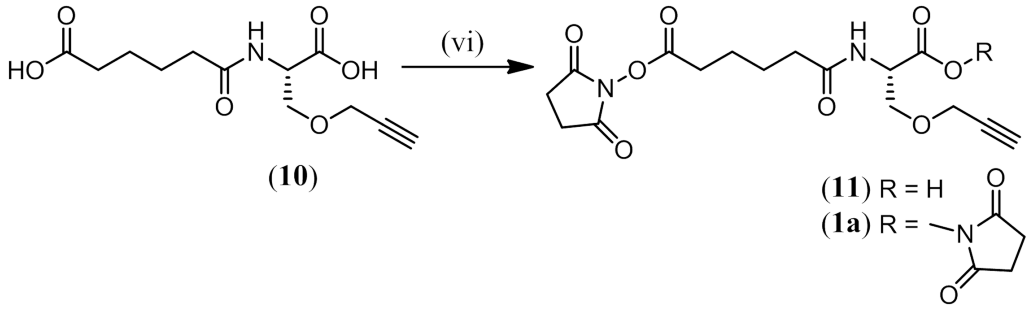
Table 2.1: Reaction conditions used and results obtained for the activation of diacid **10** with NHS.

Reagents	**Solvent, Rxn Time**	**Result**
NHS (4.0 eq), EDC-HCl (2.4 eq)	DMF, 6h	Mixture
NHS (4.0 eq), EDC-HCl (2.4 eq)	DMF, 18h	Mixture
NHS (4.0 eq), EDC-HCl (2.4 eq)	1:1 DCM:Acetone, 6h	Mixture
NHS (4.0 eq), EDC-HCl (2.4 eq)	1:1 DCM:Acetone, 18h	Mixture
NHS (4.0 eq), DIC (2.4 eq)	THF, 18h	Mixture
TFA-NHS (5.0 eq)	Pyridine, 18h	88% **11**, R = H, Trace quantities of **1a**, R = NHS

The original synthesis of N-trifluoroacetoxy succinimide was published in 1965 by Sakakibara *et al* from trifluoroacetic anhydride and NHS, and used benzene as the solvent.[155] Here, toluene was used as the solvent in the synthesis of N-trifluoroacetoxy succinimide to avoid the use of hazardous benzene. Sakakibara *et al* reported a quantitative yield for the activation of NHS in benzene, however, through the use of toluene N-trifluoroacetoxy succinimide was obtained in an excellent yield of 98%. The excellent yield and consistency of all ^{1}H, ^{19}F and ^{13}C NMR data with reported values[153] proved toluene to be a suitable solvent substitute in the preparation of N-trifluoroacetoxy succinimide.

Diacid **10** was reacted with N-trifluoroacetoxy succinimide in pyridine to yield a mixture of major products, monoester **11** and crosslinker **1a** (Scheme 2.6, Table 2.1). Monoester **11** was isolated by flash chromatography in 88% yield and crosslinker **1a** was synthesised in trace quantities, detected only by HRMS. Monoester **11** was fully characterised by NMR to determine which end of the molecule had become functionalised with NHS. The ^{1}H NMR showed a distinct downfield shift of a triplet resonance corresponding to a methylene group within the spacer arm, from 2.15

ppm in diacid **10** to 2.67 ppm in monoester **11** (Figure 2.9). This shift was also accompanied by a slight downfield shift in the multiplet at 1.50-1.46 in **10** to 1.66-1.54 in **9** corresponding to the two central methylene groups within the spacer arm. The second triplet corresponding to the methylene group at the amino acid end of the spacer arm had not shifted. Therefore, as the downfield shift of the three methylene groups within the spacer arm was not accompanied by a similar shift in the signal of the hydrogen attached to the αC of serine, it is unlikely that the functionalisation of diacid **10** occurred at the amino acid carboxyl terminus. It is also unlikely that the functionalisation of the amino acid terminus would cause downfield shifting of the central methylene groups of the spacer arm as these residues would be comparatively far from the site of modification. It was therefore determined that the site of NHS activation was the carboxyl terminus directly attached to the spacer arm to form monoester **11**.



Scheme 2.6: Functionalisation of **10** with NHS reactive groups using (vi) TFA-NHS, pyridine to give **11** and **1a** (88%, and trace quantities respectively).

Once isolated, monoester **11** was subject to further reaction with N-trifluoroacetoxy succinimide in pyridine in an attempt to force the reaction to completion with excess reagents. This reaction again gave trace quantities of crosslinker **1a** based on HRMS. Esterification proceeds at the sterically most accessible carboxylic acid to give **11** with minimal formation of crosslinker **1a**. More forceful conditions may therefore be required to esterify the amino acid carboxylic acid terminus with NHS.

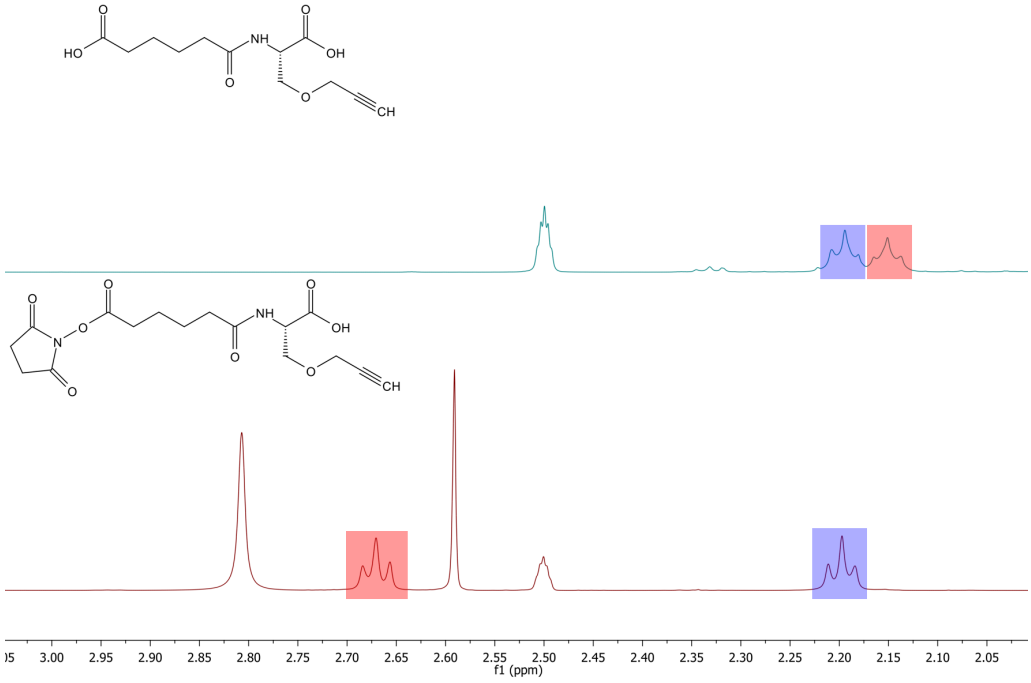


Figure 2.9: Comparison of the NMR spectra of diacid **10** (top) and monoester **11** (bottom) showing the characteristic downfield shift of the triplet peak from 2.15 ppm to 2.67 ppm upon NHS activation.

2.4 Conclusions and Future Work

A novel crosslinker design was proposed consisting of three basic modules, reactive groups, the spacer arm and an affinity tag, as seen in the structure of general crosslinker **1** (Figure 2.2). Supported by crosslinking literature, biotin was chosen as the affinity tag and NHS esters as the reactive groups. Work suggesting that bulky or large crosslinking reagents can disrupt native protein structure was taken into account, and hence it was decided to attach the biotin tag after tryptic digestion of crosslinked protein, facilitated by biologically compatible and robust Huisgen azide-alkyne cycloaddition. Retrosynthetic analysis yielded two key reagents, crosslinker **1** containing an alkyne, and biotin azide. Retrosynthesis of crosslinker **1** showed that it could be synthesised through a series of *O*-alkylation, amide coupling and NHS functionalisation steps from commercially available starting materials, while biotin azide could be synthesised from a documented procedure.[142]

Crosslinker **1a** was proposed (Figure 2.8) as a simple example to synthesise during the development of the synthetic protocol, whilst still addressing the analytical challenges of the detection and identification of crosslinked species. The utility of the *O*-alkylation, N-Boc deprotection and esterification, peptide coupling and ester

hydrolysis steps were shown through the successful synthesis of diacid **10**. Introduction of NHS esters to diacid **10** was successfully achieved to form trace quantities of crosslinker **1a** using N-trifluoroacetoxy succinimide. The major product, monoester **11**, requires further activation for complete transformation to the desired product.

Further investigation is required to determine the cause of the difficulties pertaining to the introduction of NHS to the amino acid terminus of crosslinker **1a**. Carbodiimide coupling reagents did not give any product, and the use of N-trifluoroacetoxy succinimide gave trace quantities. Activation could be investigated by undertaking model studies using simpler precursors such as alkyne **8**. This would give insight into the reactivity of the amino acid carboxyl terminus and hence its utility in future syntheses.

Concurrently, activation of diacid **10** could be probed through the use of later generation triazole based coupling reagents such as HATU and HBTU with different solvents such as chloroform or dioxane. Large excesses of N-trifluoroacetoxy succinimide and gentle heating may also aid in the formation of isolatable quantities of crosslinker **1a**. The use of a reagents such as N-bromosuccinimide (NBS) or N-hydroxysulfosuccinimide (Sulfo-NHS) in place of NHS may also pose reasonable alternatives.

Functionalisation of diacid **10** with reactive groups other than NHS could also be attempted. The preparation of maleimides, hydrazides, imido esters or diazirines could be achieved in this way (Figure 2.1).

The functionality of crosslinker **1a** must be determined through crosslinking experiments with simple peptide units, followed by larger and more complex protein systems. After crosslinking, methods for the Huisgen azide-alkyne cycloaddition between peptides crosslinked with crosslinker **1a** and biotin azide, and avidin affinity purification should be attempted and further investigated to determine their utility.

2.5 Experimental

2.5.1 General Methods

All starting materials were purchased from Sigma Aldrich, GL Biochem Tokyo Chemical Industry. Co (TCI) and Genetech unless otherwise stated, and were used as received. All NMR data was analysed using MestReNova 6.0 from Mestrelab Research S.L. All organic extracts were dried over anhydrous sodium sulfate. Thin layer chromatography was performed using Merck aluminium sheets with silica gel 60 F_{254}. Compounds were visualised using a UV lamp and/or with a potassium permanganate stain (3.0 g $KMnO_4$, 20.0 g K_2CO_3, 5 mL 10% NaOH, 300 mL H_2O) followed by heating. All R_f values were taken to the nearest 0.01. All reported yields were judged to be isolated yields, unless otherwise specified, by TLC and NMR. Flash chromatography was performed using Grace Davidson Discovery Sciences Davisil 40-63 grade silica gel, under positive nitrogen pressure. All solvents for chromatography were used as received. Infrared spectra were obtained using a Perkin Elmer BX FT-IR spectrometer. High field NMR spectra were recorded using an Agilent DD2 spectrometer operating at 500 MHz for ^{1}H, 125 MHz for ^{13}C and 470 MHz for ^{19}F. Solvents used to obtain spectra are specified. Chemical shifts are reported in ppm on the δ-scale relative to TMS (δ_H 0.00), $CDCl_3$ (δ_H 7.26, δ_C 77.16) or DMSO-d_6 (δ_H 2.50, δ_C 39.52). Spin multiplicities are reported as (br s) broad singlet, (s) singlet, (d) doublet, (t) triplet, (q) quartet, and (m) multiplet. All J values were rounded to the nearest 0.1 Hz and correspond to H-H coupling unless otherwise specified. All reactions done under an inert atmosphere (N_2) were carried out with oven dried glassware. Removal of solvent under reduced pressure refers to the removal of a large quantity of solvent by rotary evaporation followed by a high vacuum pump for a minimum of 1 hour. High resolution masses were measured using an Agilent Technologies 6230 Accurate-Mass TOF LC/MS to 0.0001 m/z.

General Procedure A: One Pot N-Boc Cleavage and Esterification
N-Boc carboxylic acid was dissolved in methanol (100 mL/3 g of amino acid) and cooled in an ice bath for 15 min. To the solution was added thionyl chloride (2.0 mol eq) dropwise under a nitrogen atmosphere and the mixture stirred at room temperature overnight. The solvent was removed *in vacuo* to yield an oil. This residue was re-dissolved in methanol and the solvent removed *in vacuo* three times to yield the desired methyl ester.

General Procedure B: HATU Mediated Peptide Coupling

The respective amine was dissolved in dry N'N dimethylformamide (3 mL/100 mg of amine) and DIPEA (4.0 mol eq) added whilst stirring. The relevant carboxylic acid (1.0 mol eq) was added to the stirred solution followed by HATU (1.2 mol eq) and the solution stirred at room temperature under a nitrogen atmosphere overnight. The solution was partitioned between ethyl acetate and 1 M HCl. The organic phase was then separated and washed with 1 M HCl (2x), water (2x) and brine, dried over Na_2SO_4 and the solvent removed *in vacuo*. Where applicable the crude product was purified by flash chromatography on normal phase silica to yield the desired amides.

General Procedure C: Esterification of Carboxylic Acids

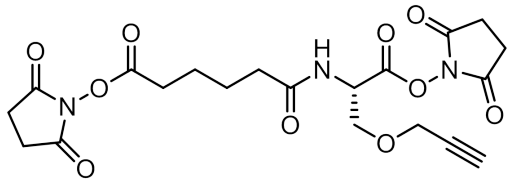
The relevant carboxylic acid was dissolved in methanol (100 mL/1 g of acid) and stirred on ice for 5 minutes. Thionyl chloride (2.0 mol eq) was added dropwise under a nitrogen atmosphere to the stirred solution. The mixture was stirred on ice for a further 15 min and then allowed to warm to room temperature and react for 6 h. The solvent was then removed *in vacuo*. The residue was re-dissolved in methanol and the solvent removed *in vacuo* three times to yield an oil. Where applicable the crude product was purified on normal phase silica to yield the desired esters.

General Procedure D: Hydrolysis of Methyl Esters

The respective methyl ester was dissolved in tetrahydrofuran (2.0 mL/100 mg of methyl ester), 1.6 M sodium hydroxide (1.0 mL/100 mg) and methanol (1.0 mL/100 mg) and the mixture stirred overnight. The solvent was removed *in vacuo* and the residue partitioned between ethyl acetate and 1M HCl. The organic layer was isolated and the aqueous layer extracted with ethyl acetate (4x). The combined organic extracts were washed with brine and dried over Na_2SO_4. The solvent was removed *in vacuo* to yield the desired free acids.

2.5.2 Synthesis and Characterisation

(*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 (1a)



Synthesis of **1a** was attempted using the reagents outlined in Table 2.1.

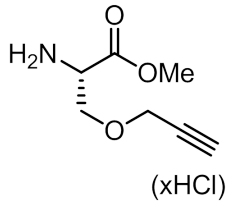
Using EDC mediated coupling in DMF:
Diacid **6** was dissolved in anhydrous DMF and stirred for 5 min. To this solution was added EDC-HCl (2.4 eq) and NHS (4.0 eq) and the reaction stirred for 6 or 18 h at room temperature under a nitrogen atmosphere. The reaction mixture was partitioned between ethyl acetate and 1 M HCl. The aqueous phase was extracted with ethyl acetate (2x) The combined organic extracts were washed with brine and dried over Na_2SO_4 and the solvent removed *in vacuo* to yield a brown oil. Flash chromatography purification did not yield **1a**.

Using EDC mediated coupling in acetone/DCM:
Diacid **6** (150 mg, 0.49 mmol) was dissolved in anhydrous 1:1 acetone/DCM and stirred for 5 min. To this solution was added EDC-HCl (2.4 eq) and NHS (4.0 eq) and the reaction stirred for 6 or 18 h at room temperature under a nitrogen atmosphere. The solvent was removed *in vacuo* to yield a brown oil. Flash chromatography purification did not yield **1a**.

Using DIC mediated coupling in THF:
Diacid **6** (150 mg, 0.49 mmol) was dissolved in anhydrous THF and stirred for 5 min. To this solution was added DIC (2.4 eq) and NHS (4.0 eq) and the reaction stirred for 18 h at room temperature under a nitrogen atmosphere. The solvent was removed *in vacuo* to yield a brown/yellow oil. Flash chromatography purification did not yield **1a**.

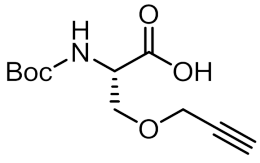
Using TFA-NHS in pyridine:
Diacid **6** (150 mg, 0.49 mmol) and N-trifluoroacetoxy succinimide (TFA-NHS) (4.0 eq) were dissolved in pyridine (6 mL) and stirred under a nitrogen atmosphere at room temperature for 8 h. The solution was partitioned between ethyl acetate and 1M HCl. The aqueous phase was extracted with ethyl acetate (2x). The combined organic extracts were washed with 1M HCl (2x), brine and dried over Na_2SO_4. The solvent was removed *in vacuo* to yield **11** (88%), and **1a** in trace quantities.
HRMS (ES+) calculated for $C_{20}H_{23}N_3O_{10}$ $[M+Na]^+$ = 488.1261, found $[M+Na]^+$: 488.1271 m/z.

(*S*)-methyl 2-amino-3-(prop-2-yn-1-yloxy)propanoate hydrochloride (6)



Alkyne **3** (2.16 g, 8.8 mmol) was N-Boc cleaved and esterified according to General Procedure A to yield **4** as an orange oil, without further purification (1.58 g, 92 %). ^{1}H NMR (500 MHz, DMSO-d_6, δ) 8.64 (3H, br s, N**H**$_3$), 4.37 (1H, t, J = 3.3 Hz, **H**αC), 4.24 (1H, dd, J = 16.1, 2.2 Hz, OC**H**HCCH), 4.20 (1H, dd, J = 16.1, 2.2 Hz, OCH**H**CCH), 3.91 (1H, dd, J = 10.5, 4.1 Hz, HαCC**H**HO), 3.85 (1H, dd, J = 10.6, 3.0 Hz, HαCCH**H**O), 3.76 (3H, s, OC**H**$_3$) 3.56 (1H, t, J = 2.1 Hz, CH_2CC**H**) All NMR data as per literature.[145]

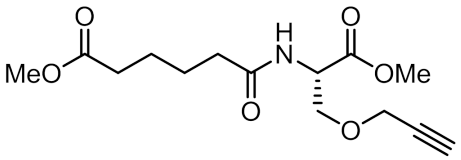
(*S*)-2-((tert-butoxycarbonyl)amino)-3-(prop-2-yn-1-yloxy)propanoic acid (8)



Boc-L-Serine (**2**) (2.01 g, 9.7 mmol) was dissolved in anhydrous DMF (100 mL) and stirred on ice for 15 min. To this solution was added a 60% wt/wt dispersion of sodium hydride (995 mg, 24.8 mmol) and the solution stirred on ice for a further 45 min. An 80% v/v solution of propargyl bromide (1.2 mL, 10.7 mmol) was added

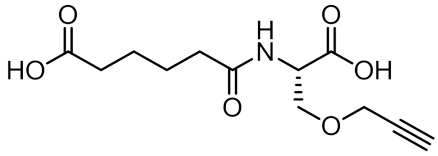
dropwise to the solution and allowed to stir on ice for a further 15 min. The solution was allowed to warm to room temperature overnight. Water (15 mL) was added and the solution stirred for 5 min. The solution was washed with diethyl ether and the aqueous phase acidified with $KHSO_4$ (95 mL of 1 M solution) and extracted with ethyl acetate (3x). The combined organic extracts were washed with water (3x) and brine and dried over Na_2SO_4. The solvent was removed *in vacuo* to yield **3** as a yellow oil, without further purification (2.16 g, 91 %).
^{1}H NMR (500 MHz, DMSO-d_6, δ) 5.38 (1H, d, J = 8.2 Hz, N**H**), 4.49 (1H, d, J = 8.1 Hz, **H**αC), 4.23-4.14 (2H, m, OC**H**$_2$CCH), 4.00 (1H, dd, J = 8.8, 2.2 Hz, HαCC**H**HO), 3.81 (1H, dd, J = 9.4, 3.5 Hz, HαCCH**H**O), 2.47 (1H, t, J = 2.3 Hz, CH_2CC**H**), 1.46 (9H, s, N-Boc). ^{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 as per literature.[149]

(*S*)-methyl 6-((1-methoxy-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-6-oxohexanoate (9)



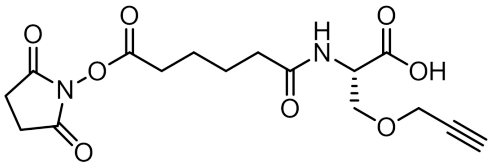
Hydrochloride salt **4** (1.0 g, 5.1 mmol) was coupled to monomethyl adipate according to General Procedure B. The residue was purified by flash chromatography on silica (ethyl acetate/petroleum ether, 1:1) to yield **5** as white solid (0.79 g, 51%). Mp: 78.9 - 79.5 °C, R_f: 0.24, IR (neat): 3310 cm^{-1} (N-H), 2947 cm^{-1} (C-H), 1731 cm^{-1} (C=O).
^{1}H NMR (500 MHz, $CDCl_3$, δ) 6.36 (1H, d, J = 7.9 Hz, N**H**), 4.74 (1H, ddd, J = 6.5, 3.0, 2.9 Hz, **H**αC), 4.14-4.06 (2H, m, OC**H**$_2$CCH), 3.92 (1H, dd, J = 9.4, 3.1 Hz, HαCCH**H**O), 3.75-3.71 (4H, m, HαCC**H**HO and OC**H**$_3$), 3.62 (3H, s, OC**H**$_3$), 2.44 (1H, t, J = 2.1 Hz, CH_2CC**H**), 2.29 (2H, t, J = 6.6 Hz, C**H**$_2$CONH), 2.24 (2H , t, J = 6.5 Hz, H_3CCOC**H**$_2$) 1.67-1.61 (4H, m, CH_2C**H**$_2$C**H**$_2$CH_2). ^{13}C NMR (125 MHz, $CDCl_3$, δ) 173.9, 172.5, 170.6, 78.9, 75.2, 69.5, 58.6, 52.7, 52.3, 51.6, 36.0, 33.7.
HRMS (ES+) calculated for $C_{14}H_{21}NO_6$ $[M+H]^+ = 300.1448$, found $[M+H]^+$: 300.1448 m/z.

(*S*)-6-((1-carboxy-2-(prop-2-yn-1-yloxy)ethyl)amino)-6-oxohexanoic acid (10)



Dimethyl ester **5** (400 mg, 1.2 mmol) was hydrolysed according to General Procedure D, to yield **6** as a colourless oil, without further purification (261 mg, 72 %). R_f: 0.11, IR (neat): 3284 cm^{-1} (O-H/N-H), 2939 cm^{-1} (C-H), 1709 cm^{-1} (C=O). ^{1}H NMR (500 MHz, DMSO-d_6, δ) 12.29 (2H, br s, O**H**), 8.12 (1H, d, J = 8.0 Hz, N**H**), 4.44 (1H, ddd, J = 8.1, 5.5, 4.3 Hz, **H**αC), 4.18-4.11 (2H, m, OCC**H**$_2$CCH), 3.73 (1H, dd, J = 9.6, 5.7 Hz, HαCC**H**HO), 3.63 (1H, dd, J = 9.6, 4.1 Hz, HαCCH**H**O), 3.44 (1H, t, J = 2.3 Hz, CH_2CC**H**), 2.19 (2H, t, J = 6.8 Hz, C**H**$_2$CONH), 2.15 (2H, t, J = 6.8 Hz, HOOCC**H**$_2$), 1.50-1.46 (4H, m, CH_2C**H**$_2$C**H**$_2$$CH_2$). ^{13}C NMR (125 MHz, DMSO-d_6, δ) 177.5, 175.3, 174.5, 82.9, 80.6, 72.1, 60.8, 55.1, 37.7, 36.5, 27.9, 27.2.
HRMS (ES+) calculated for $C_{12}H_{17}NO_6$ $[M+Na]^+$ = 294.0953, found $[M+Na]^+$: 294.0975 m/z.

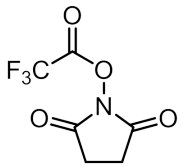
(*S*)-2-(6-((2,5-dioxopyrrolidin-1-yl)oxy)-6-oxohexanamido)-3-(prop-2-yn-1-yloxy)propanoic acid (11)



Diacid **6** (150 mg, 0.49 mmol) and N-trifluoroacetoxy succinimide (TFA-NHS) (4.0 eq) were dissolved in pyridine (6 mL) and stirred under a nitrogen atmosphere at room temperature for 8 h. The solution was partitioned between ethyl acetate and 1M HCl. The aqueous phase was extracted with ethyl acetate (2x). The combined organic extracts were washed with 1M HCl (2x), brine and dried over Na_2SO_4. The solvent was removed *in vacuo*. The residue was purified by flash chromatography on silica (methanol (5%)/dichloromethane (95%)) to yield **7** as a colourless oil (179 mg, 88%). R_f: 0.12, IR (neat): 3264 cm^{-1} (O-H/N-H), 2980 cm^{-1} (C-H), 1731,

1709 cm^{-1} (C=O).
^{1}H NMR (500 MHz, DMSO-d_6, δ) 8.15 (1H, d, J = 8.0 Hz, N**H**), 4.48-4.43 (1H, m, **H**αC), 4.19-4.11 (2H, m, OC**H**$_2$CCH), 3.74 (1H, dd, J = 9.6, 5.6 Hz, HαCC**H**HO), 3.64 (1H, dd, J = 9.6, 4.0 Hz, HαCCH**H**O), 3.43 (1H, t, J = 2.2 Hz, CH_2CC**H**), 2.81 (4H, s, OCC**H**$_2$C**H**$_2$CO), 2.67 (2H, t, J = 6.9 Hz, CNCOC**H**$_2$), 2.20 (2H, t, J = 6.8 Hz, C**H**$_2$CONH), 1.66-1.54 (4H, m, CH_2C**H**$_2$C**H**$_2$$CH_2$). ^{13}C NMR (125 MHz, DMSO-d_6, δ) 172.8, 172.0, 171.4, 170.3, 168.9, 79.8, 77.5, 69.0, 57.8, 52.0, 34.2, 29.9, 25.5, 25.3, 24.3, 23.7.
HRMS (ES+) calculated for $C_{16}H_{20}N_2O_8$ $[M+H]^+$ = 369.1299, found $[M+H]^+$: 369.1292 m/z.

2,5-dioxopyrrolidin-1-yl 2,2,2-trifluoroacetate



N-hydroxy succinimide (1.0 g, 8.7 mmol) was dissolved in a stirring solution of toluene (20 mL) and trifluoroacetic anhydride (1.4 mL, 10.0 mmol) and refluxed for 5 h. The solvent was removed *in vacuo* to yield TFA-NHS as a white crystalline solid (1.78 g, 98%).[155]
^{1}H NMR (500 MHz, $CDCl_3$, δ) 2.92 (4H, s, C**H**$_2$C**H**$_2$). ^{13}C NMR (125 MHz, $CDCl_3$, δ) 167.4, 153.8 (q, J_{CF} = 46.1 Hz), 114.1 (q, J_{CF} = 286.1 Hz), 25.7. ^{19}F NMR (470 MHz, $CDCl_3$, δ) -72.3.
All NMR data as per literature.[153]

Chapter 3

Synthesis of cleavable crosslinkers and crosslinking of *Staphylococcus aureus* biotin protein ligase

3.1 Introduction

The analytical challenges summarised in Section 1.3.2 have hindered the widespread use of the CXMS technique, resulting in the functionalisation of chemical crosslinkers with isotope labels, cleavable bonds and affinity tags (Section 1.3.3). Crosslinkers such as EGS-d_{12} and CBDPS (Figure 1.9) utilise a combination of these functionalities and are advantageous as the crosslinker is often able to address more than one analytical challenge at a time. The number of studies utilising crosslinkers with a combination of functionalities has been limited due to a lack of commercial availability and inflexibility of reported designs, which are generally not easily modified to suit different experiments. Hence, the modular synthetic protocol for crosslinker synthesis will be further exploited to synthesise crosslinkers containing an affinity tag and a cleavable bond in this chapter.

Two main goals of CXMS are the mapping of protein interaction networks and probing of macromolecular protein structure which are investigated through the formation of type 2 interprotein crosslinks. Hence, all crosslinkers synthesised using the modular synthetic protocol must be able to form covalent linkages between associating protein species to be effective within the CXMS workflow. The development of

a crosslinking assay using a protein which is known to form quaternary associations would therefore allow for the functional validation of novel crosslinkers synthesised. Hence this chapter will also explore the development of a crosslinking assay based upon the *Staphylococcus aureus* biotin protein ligase (*Sa*BPL) enzyme.

3.1.1 Cleavable bonds

Crosslinkers have been functionalised to include bonds which can be cleaved chemically, or *in vacuo* within the mass spectrometer using gas phase dissociation methods. Utilisation of bonds able to be cleaved in the gas phase is favourable, as extra chemical cleavage steps within the CXMS workflow are therefore not required. For cleavage in the gas phase to be achievable, it is often preferred for the bond within the crosslinker to dissociate at a lower activation energy than the amide bonds within the constituent peptides to reduce the number of competitive cleavage products (Section 1.4.2). The use of cleavable bonds allows for type 1 intrapeptide and type 2 interpeptide crosslinked species to be distinguished (Figure 1.4) and the site specific location of crosslinked residues made easier as peptide sequencing information is clearer.

Site specific location of crosslinked residues can be achieved for peptides containing more than one crosslink site, which eliminates some ambiguity currently possible when assigning crosslinks due to complex MS fragmentation behaviour, and allows for the determination of the most accurate distance constraints between protein residues. Assignment of crosslinked residues is especially complex in the case of type 2 crosslinked species as MS/MS sequencing attempts lead to co-fragmentation of each peptide, resulting in a complex MS/MS spectrum as discussed in Section (1.3.2). The use of a crosslinker with a cleavable bond, which dissociated prior to peptide backbone cleavage, would allow for the dissociation of the crosslinker, and hence the sequencing of each peptide individually in an MS3 approach, making it possible to determine the exact site of crosslink modification (Figure 3.1). MS3 approaches allow for an extra tandem MS step, breaking down the sample molecules into smaller constituents than is possible with traditional MS2 approaches.

Products formed through the dissociation of a cleavable bond have characteristic differences in mass which are dependent upon the crosslinker used, and the mechanism by which the bond cleaves. The fragmentation pattern of the disulfide containing crosslinker dithiobis(succinimidyl) propionate (DSP) is an example of this principle

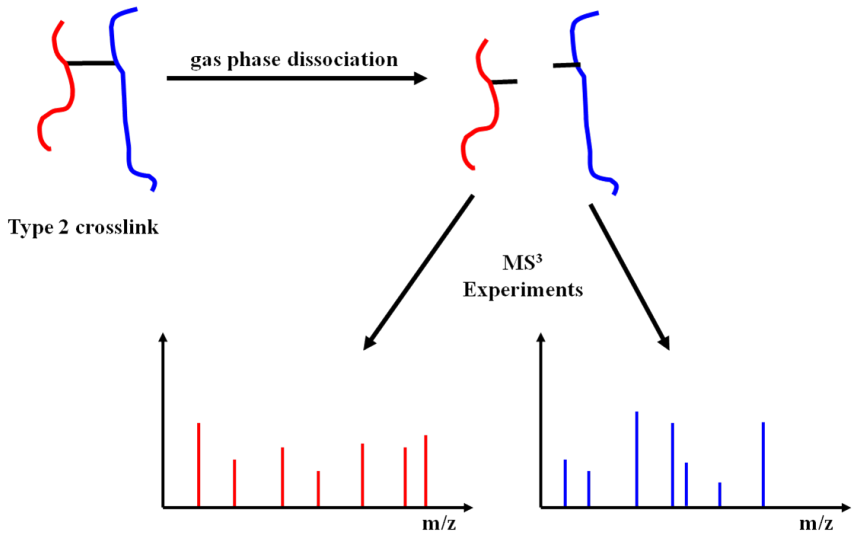


Figure 3.1: MS^3 sequencing of individual peptides (red and blue) after gas phase crosslinker dissociation within a type 2 crosslinked peptide.

(Figure 3.2). Low energy fragmentation of the disulfide bond in negative mode according to the mechanisms shown produces several characteristic signals with mass spacings of 32, 2, 32 Da.[156] Dedicated software such as XlinkX[82], and MeroX[81] have been developed which can identify the characteristic mass signature within a spectrum, thereby determining which signals correspond to intermolecular crosslinked species.

A crosslinker can also be designed to contain more than one cleavable bond within the spacer arm. Crosslinkers containing two cleavable bonds are classed as protein interaction reporters (PIR).[157] In the case of a type 2 crosslink, upon cleavage of the bonds using gas phase dissociation, the two constituent peptides are released, however, the center of the spacer arm between the two bonds is also freed, and acts as a reporter ion. The sum of the masses of the reporter ion and the two constituent modified peptides equals the mass of the parent ion, allowing for crosslinked species to be identified and peptides individually sequenced.

The combination of cleavable bonds and an affinity tag within the same crosslinker provides two methods to improve the efficiency of identification and analysis of crosslinked species. In addition to the advantages described for cleavable bonds in determining the site of crosslink formation, the use of affinity purification allows for the enrichment of crosslinked species from complex mixtures, simplifying the identification of crosslinked species by increasing their relative abundance.

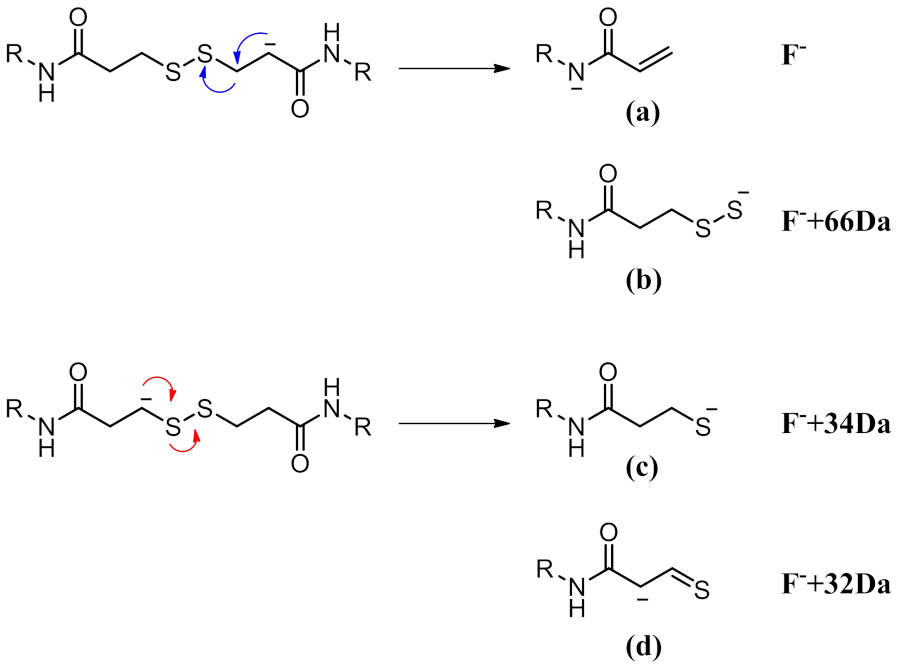


Figure 3.2: Gas phase dissociation mechanisms and products of DSP, yielding characteristic ions **(a)**, **(b)**, **(c)** and **(d)** according to nomenclature described by Bilusich *et al* and Zhang *et al.*[132,133] Ions **(a)** and **(d)** are formed through charge transfer via an ion-neutral complex. In the case of an asymmetrical crosslink, R chains will have different structures and masses, in which case 8 characteristic ions will be produced. $\mathbf{F}^-$ is the fragment with the smallest mass remaining from the crosslinker following CID. The mass differences of all of the ions are calculated from $\mathbf{F}^-$.

Incorporation of cleavable bonds into the crosslinker design proposed in Figure 2.2 (p. 28) can be achieved through customisation of the spacer arm module. The spacer arm module was designed to be incorporated using a monomethyl ester intermediate, and use of a monomethyl ester containing the chosen MS cleavable bonds would serve this purpose.

3.1.2 *Staphylococcus aureus* biotin protein ligase

The development of a protein based assay to probe crosslinker functionality is essential to prove the utility of the modular design. The assay requires a protein which is known to form interprotein associations *in vivo*. The biotin protein ligase enzyme from *Staphylococcus aureus* (*Sa*BPL) forms homodimers *in vivo*, contains a high natural abundance of lysine residues to be targeted by NHS reactive groups within the monomer structure and can be produced with relative ease through recombinant protein expression. Hence *Sa*BPL was chosen as the protein upon which the assay would be based.

Biotin protein ligases are ubiquitous enzymes directly responsible for the adenosine triphosphate (ATP) dependent attachment of biotin onto biotin-dependent enzymes and, in some organisms, the regulation of biotin biosynthesis. Through the natural process of biotinylation, the biotin and ATP substrates form biotinyl-5′-AMP within the active site.[158] Three classes of BPL enzymes have been classified based upon structure, of which bacterial BPLs fall into classes I and II, and mammalian and yeast BPLs into class III.[159,160] Class I BPL enzymes such as those from *Mycobacterium tuberculosis* and *Pyrococcus horikoshii* contain a catalytic domain only, to facilitate biotin attachment.[161,162] Class II BPL enzymes such as the the *Escherichia coli* BPL are bifunctional and contain the conserved catalytic domain, and also possess an N' terminal helix-turn-helix (HTH) deoxyribonucleic acid (DNA) binding domain.[163] The DNA binding domain facilitates the regulation of biotin biosynthesis, transport and fatty acid biosynthesis upon the formation of BPL homodimers, regulated by ligand binding. *Sa*BPL belongs to class II BPLs and acts as a transcriptional repressor when dimerised.[164,165] *Sa*BPL exists in two forms in solution, the apo form and ligand-bound holo form. *Sa*BPL dimerises and performs transcriptional repressor functions in the biotin, or biotinyl-5′-AMP bound holo form, though an equilibrium population of apo dimers can also exist in solution.[166]

*Sa*BPL presented a good crosslinking target due to its known propensity to form

dimers when ligand bound, and the abundance and position of lysine residues within the dimeric structure. Each *Sa*BPL monomer contains twenty two key lysine residues. The X-ray crystal structure of *Sa*BPL dimers bound to the natural substrate biotinyl-5′-AMP[164] revealed two lysine residues within the dimerisation interface which could be probed using the NHS activated crosslinkers synthesised from the modular protocol (Figure 3.3).

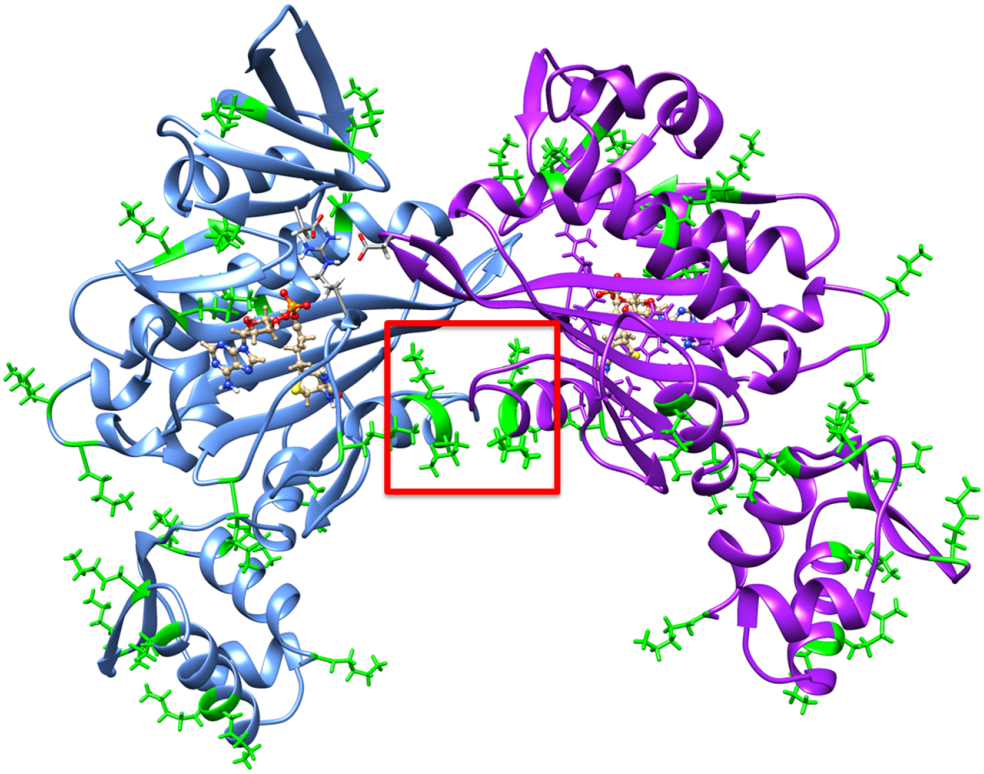


Figure 3.3: X-ray crystal structure by Pendini *et al*[164] of dimerised holo *Sa*BPL bound to biotinyl-5′-AMP. Lysines are displayed in green wire representation. The red box highlights the four lysine residues to be probed at the dimer interface.

The inter-lysine distances between each dimer are predicted to be lysine-99 (monomer 1) to lysine-99 (monomer 2), 8.01 Å and lysine-100 (monomer 1) to lysine-100 (monomer 2), 5.16Å. The distance between each residue is taken from the α carbon atom, as the lysine side chains are mobile within solution, and hence the X-ray crystal structure may not be representative of the dynamic side chain position. Successful crosslinking of *Sa*BPL with a crosslinker synthesised from the modular protocol would be indicated by the formation of covalently linked homodimers observed upon SDS-PAGE analysis.

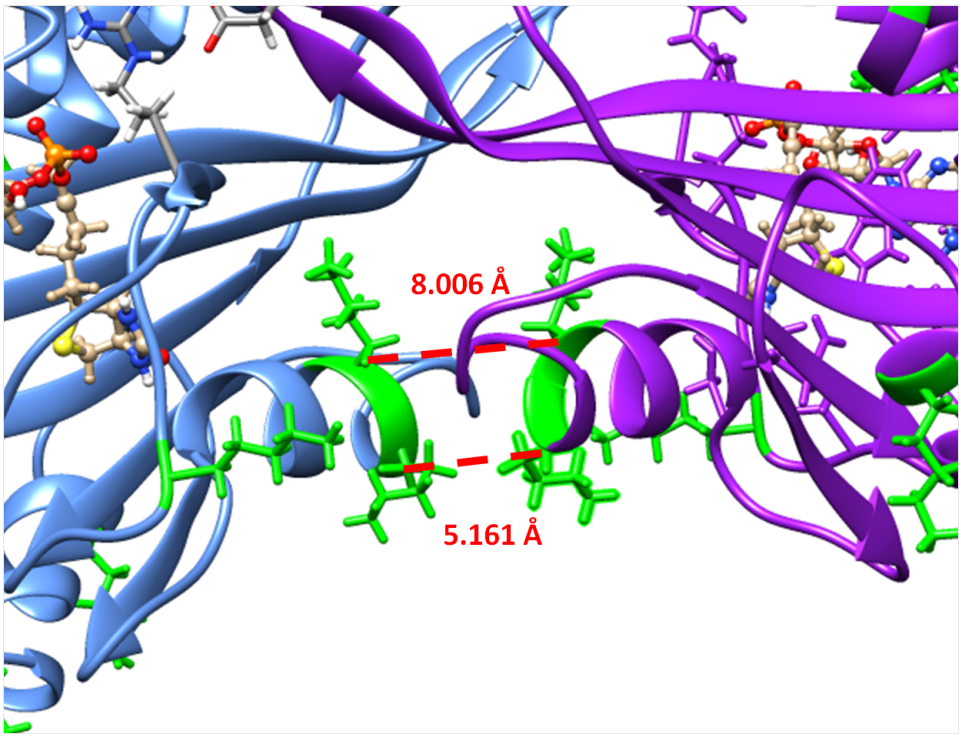


Figure 3.4: Excerpt from figure Figure 3.3 to highlight the distances between the two lysine 99 residues and two lysine 100 residues. The respective distances, Lys99-Lys99: 8.01 Å, Lys100-Lys100: 5.16 Å, are shown with a red lines from α-carbon to α-carbon.

3.1.3 Aims

In order to extend the modular synthesis discussed in Chapter 2, this work aims to synthesise crosslinkers **1b** and **1c** containing a cleavable bond as well as an affinity tag. The functionality of each crosslinker depends upon the propensity of the cleavable bond to dissociate at low collisional energies. Hence, the functionality of each cleavable bond will be established using CID to determine whether each will be viable for use in MS/MS experiments.

The formation of type 2 crosslinks between interacting protein species is pivotal for the characterisation of complex quaternary structure and the probing of protein interaction networks. Therefore, a crosslinking assay with *Sa*BPL which can be used to probe the ability of any synthesised crosslinkers to form type 2 crosslinks will be developed.

3.2 Crosslinker 1b

To address the first aim, the synthesis of crosslinker **1b** was attempted as the first design utilising the modular synthesis to combine a cleavable bond and an affinity tag.

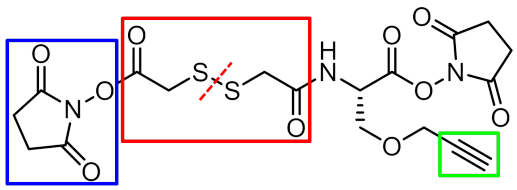


Figure 3.5: Crosslinker **1b**, cleavable bond denoted by red line.

Crosslinker **1b** was designed to follow the general crosslinker structure proposed in Section 2.2 (p. 28). The crosslinker contains NHS ester reactive group modules, an alkyne tag for Huisgen azide-alkyne cycloaddition and a spacer arm containing a disulfide bond. The disulfide bond has been shown to cleave using low energy CID in negative ion mode,[93,167,168] and hence was chosen as the labile bond to reduce competitive fragmentation with the peptide backbone. Based on the previous analysis of disulfide bonds, the crosslinker is expected to cleave to form a characteristic mass signature consisting of up to eight marker ions (Figure 3.6), with mass differences of 2, 378, 2 Da. Biotin is included on the crosslinked species as MS analysis will take place after Huisgen azide alkyne cycloaddition and avidin affinity chromatography enrichment.

The design of crosslinker **1b** addresses current challenges facing CXMS relating to the detection and identification of crosslinked peptides through the inclusion of an affinity tag, and the site specific location of crosslinked residues through the inclusion of a bond cleavable by gas phase dissociation.

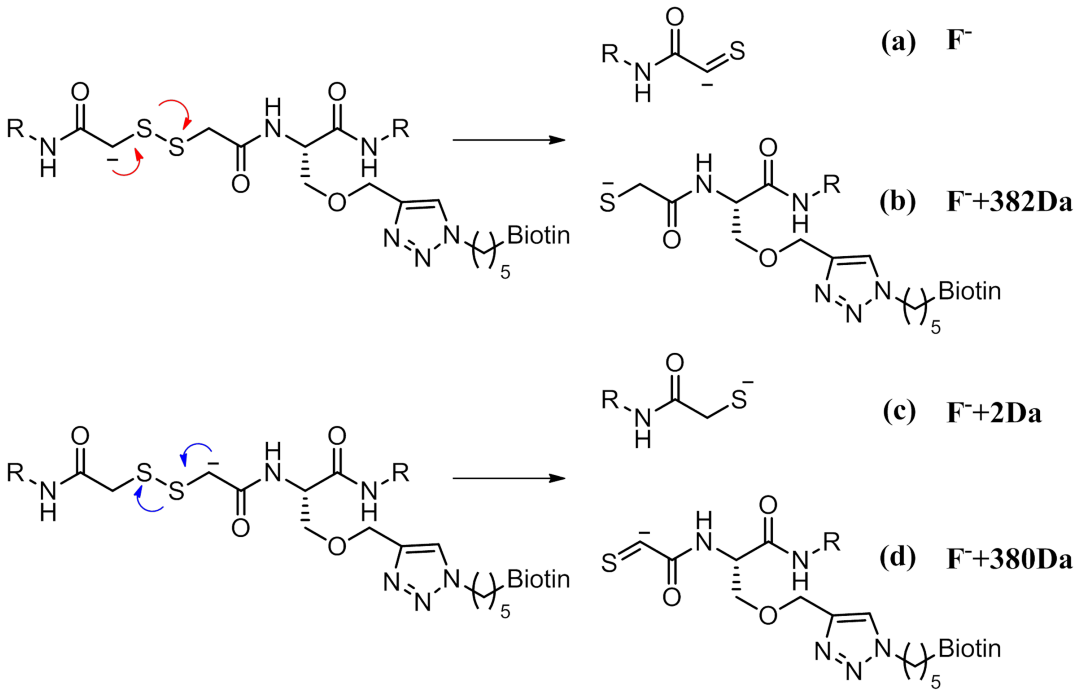
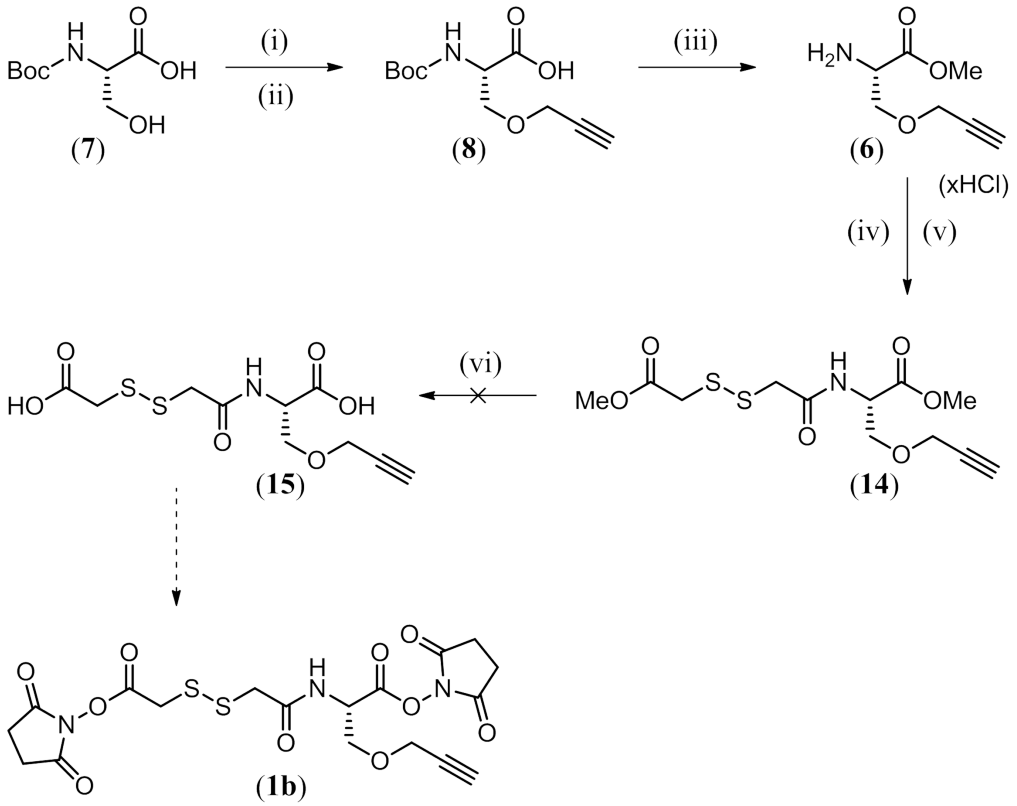


Figure 3.6: The expected dissociation mechanisms and CID products of crosslinker **1b**, **(a)**, **(b)**, **(c)** and **(d)** with masses relative to **(a)**. Ions **(a)** and **(d)** are formed through charge transfer via an ion-neutral complex. In the case of an asymmetrical crosslink, R chains will have different structures and masses, in which case 8 characteristic ions will be produced.

3.2.1 Synthesis of Crosslinker 1b

Synthesis of Crosslinker **1b** was attempted as shown in Scheme 3.1.



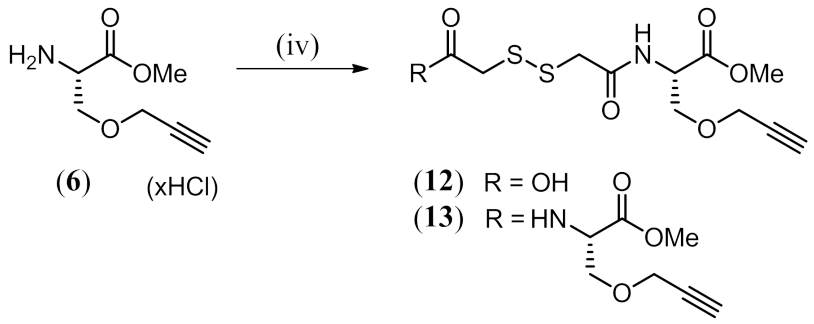
Scheme 3.1: Synthetic scheme for crosslinker**1b**. *Reagents and conditions*: (i) NaH, DMF, (ii) propargyl bromide (91%), (iii) MeOH, $SOCl_2$ (92%), (iv) Dithiodiglycolic acid, anh. DMF, DIPEA, HATU, (v) MeOH, $SOCl_2$ (27%).

Boc-L-Serine **2** was *O*-alkylated as described in 2.3.1 using sodium hydride (NaOH) and propargyl bromide (Scheme 2.2) to give alkyne **8** in 91% yield without further purification.

Alkyne **8** was treated with thionyl chloride in methanol to give hydrochloride salt **6** in 92% yield (Scheme 2.3, p. 34) without further purification, as described in 2.3.1, completing the alkyne module.

Coupling of the spacer arm and the alkyne module to form monoacid **12** was attempted (Scheme 3.2). The commercially available diacid, dithiodiglycolic acid (DDA), was utilised as the spacer arm, containing the disulfide group required for gas phase dissociation. The use of a diacid is not ideal at the peptide coupling step due to the possibility of formation of a 'diaddition' product **13**, in which the alkyne module would couple to both sides of the diacid. The diaddition product

would not be useful as a crosslinker due to its larger steric bulk in comparison with crosslinker **1b**. However, a diacid was used in place of a monomethyl ester as the corresponding monomethyl ester or anhydride were not commercially available, and reliable previous syntheses had not been reported.



Scheme 3.2: Coupling of hydrochloride salt **6** to the spacer arm using (iv) DDA, anh. DMF, HATU or EDC/HOBt, DIPEA to give a mixture of **12** and **13**.

In an attempt to reduce the amount of diaddition product formed, four different reaction conditions were tried utilising two types of coupling reagents, and different excesses of DDA (Table 3.1). A 4-molar excess of DDA was used in two of the attempts, to force the reaction of the alkyne module **8** with a 'new' molecule of DDA each time, thereby reducing the formation of **13**. The carbodiimide based coupling reagent EDC, and triazole based coupling reagent HATU were selected to determine whether different types of coupling reagents would favour the formation of **12** or **13**.

Table 3.1: Reaction conditions used and results obtained for the coupling of DDA to hydrochloride salt **6** to give monoacid **12**.

Eq. of DDA	**Reagents**	**Result**
1.0	EDC-HCl (1.3 eq), HOBt (1.3 eq), DIPEA (4.0 eq), DMF, 18h	Mixture
1.0	HATU (1.2 eq), DIPEA (4.0 eq), DMF, 18h	Mixture
4.0	EDC-HCl (1.3 eq), HOBt (1.3 eq), DIPEA (4.0 eq), DMF, 18h	Mixture
4.0	HATU (1.2 eq), DIPEA (4.0 eq), DMF, 18h	Mixture

In each case a mixture of both monoacid **12** and the diaddition product **13** were formed, confirmed through observation of the mass of each species by HRMS at 330.0098 m/z ($[M+Na]^+$, expected: 330.0081) and 459.1282 m/z ($[M+H]^+$, expected: 459.1260) respectively. Attempts were made to determine the ratio of product formation for each set of conditions using several techniques. Normal phase silica chromatography failed to yield the desired information as, whilst compound **13** could be isolated as a pure compound, monoacid **12** could not be isolated from

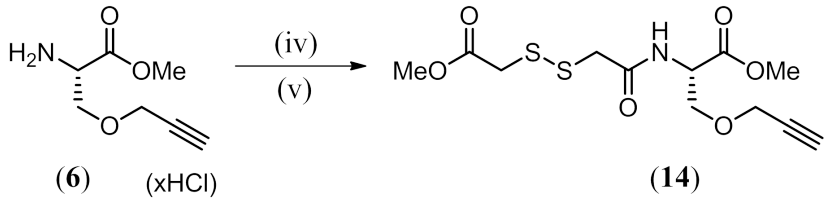
the DDA starting material for any of the reaction conditions in all of the solvent systems used, due to the large amount of streaking and co-elution observed. Acetic acid (1%) was added to the elution solvents to control the streaking, which made only a marginal difference to the outcome. The inability to isolate a pure yield of **12** prevented an accurate calculation of the product ratio using normal phase silica chromatography. The use of a combination of analytical and semi-prep high performance liquid chromatography (HPLC) was also attempted, as integration of the analytical HPLC trace peaks could give the desired ratio, and separation of the peaks using semi-prep HPLC could lead to peak identification. However, the lack of chromophores (conjugated pi-systems, and/or absorbing functional groups) within the structures of **12** and **13** resulted in an inability of the instruments to effectively detect each species for analysis. Definition of the product ratios was then attempted from the crude sample mixtures using NMR spectroscopy. Two signals which could be attributed to compounds **12** and **13** would be integrated and compared to obtain the product ratio. The crude NMR spectra were contaminated with DDA starting material and HOBt, as a base washing step could not be used in the work up due to the presence of a free carboxylic acid group in monoacid **12**. Signal similarity and overlap between products **12** and **13** also complicated the identification of separate signals, and common signals such as the CαH overlapped within the spectrum. Through the use of ^{1}H NMR, ^{13}C NMR and 2-D ^{1}H correlation spectroscopy (COSY) specific signals which could be used for comparison of integrals could not be attributed to a particular species, preventing the determination of the product ratio using NMR.

The product ratio of compounds **12** and **13** could not be determined, and monoacid **12** could not be isolated in pure yield using normal phase silica chromatography or HPLC. Hence to isolate compound **12**, an alternative process was developed. As the presence of a free carboxylic acid functional group was the cause of issues arising during normal phase silica chromatography purification, a method was devised utilising methanol and thionyl chloride in which monoacid **12** could be esterified to form diester **14** from the crude peptide coupling mixture, making the species more amenable to purification (Scheme 3.3). However, whilst the use of a large excess of DDA is more likely to suppress the formation of diaddition product **13** in the peptide coupling step, the presence of excess DDA during the esterification process would require large excesses of $SOCl_2$ to be used, and would form a large amount of side product, complicating the silica chromatography purification. Hence, the use of peptide coupling conditions with a 1 molar equivalent of DDA presented the most economical use of reagents over both steps (Table 3.2).

Table 3.2: Reaction conditions used and results obtained for the coupling of DDA to hydrochloride salt **6**, followed by esterification of crude mixture to give diester **14**

Eq. of DDA	**Reagents**	**Result**
1.0	(iv) EDC-HCl (1.3 eq), HOBt (1.3 eq), DIPEA (4.0 eq), DMF, 18h and (v) $SOCl_2$, MeOH, 6h	diester **14**, 22%
1.0	(iv) HATU (1.2 eq), DIPEA (4.0 eq), DMF, 18h and (v) $SOCl_2$, MeOH, 6h	diester **14**, 27%

The spacer arm, DDA, was coupled to hydrochloride salt **6** to yield a mixture of monoacid **12** and diaddition product **13**. After workup the mixture was esterified in the crude mixture, and purified by normal phase silica chromatography to give diester **14** (Scheme 3.3) isolated in pure yield using both EDC-HCl (22%) and HATU (27%). The triazole based coupling reagent HATU showed a slightly greater efficiency than EDC-HCl, and was hence used in all further syntheses of **14**. The ^{1}H NMR spectrum showed the appearance of two ester singlets, at 3.77 and 3.76 ppm, confirming the success of the esterification step, and hence the successful incorporation of the spacer arm module. Due to the higher degree of symmetry in compound **13** the ^{1}H NMR spectrum showed some marked differences from diester **14** which allowed for the identification of each product. Comparison between the ^{1}H NMR spectra of diaddition product **13** and diester **14** (Figure 3.7) reveals a difference in the methyl ester peaks of each compound, as **13** only has one methyl ester peak at 3.78 ppm due to the plane of symmetry, and **14** has two. Compound **13** also shows only one CH_2 signal from the spacer arm, a multiplet from 3.63 to 3.54, whereas diester **14** possesses signals for each CH_2 within the spacer arm and two multiplets from 3.64 to 3.58 and 3.58 to 3.51 ppm corresponding to a distinct separation between the CH_2 signals of the spacer arm.



Scheme 3.3: Coupling of alkyne module **6** to the spacer arm using (iv) DDA, anh. DMF, HATU, DIPEA to give a mixture of **12** and **13**, and subsequent esterification using (v) thionyl chloride, MeOH to give diester **14**.

Following the successful formation of diester **14**, hydrolysis to form diacid **15** was attempted (Scheme 3.4). Hydrolysis of methyl esters is commonly achieved through saponification with sodium or lithium hydroxide in a mixture of aqueous and organic solvents.[169,146] Hydrolysis of diester **14** was attempted using both sodium and

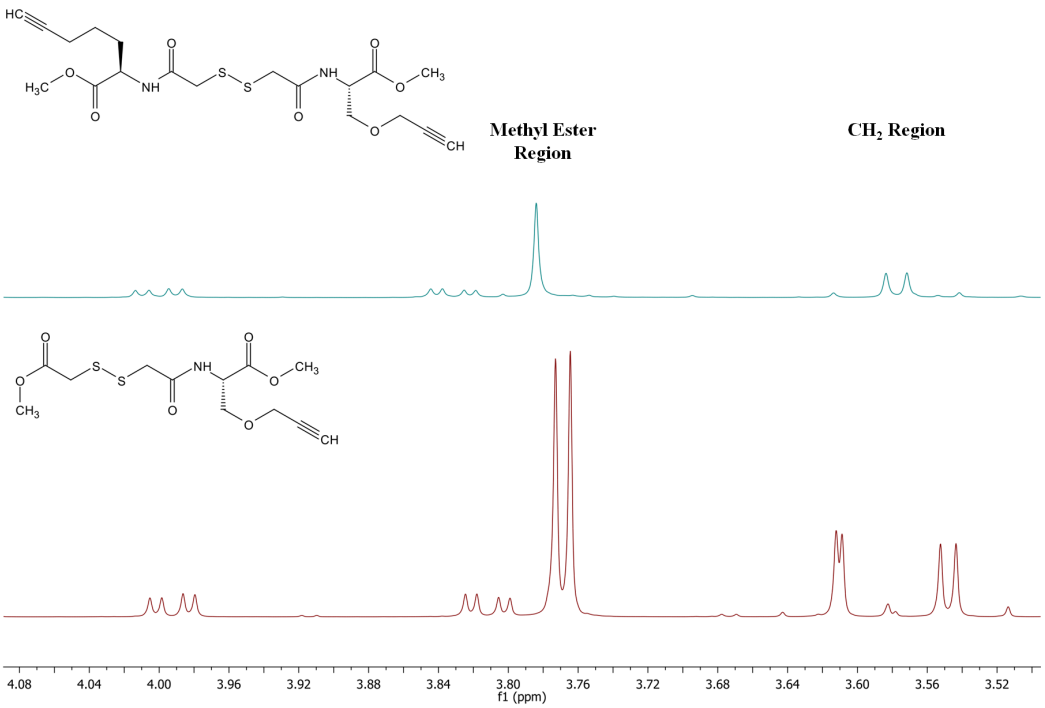
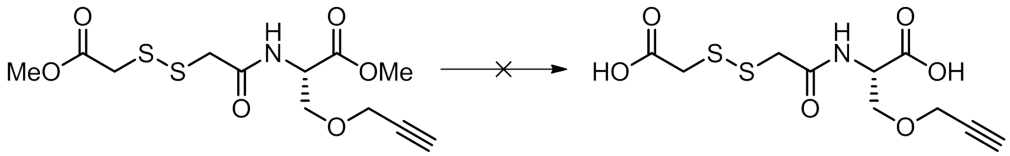


Figure 3.7: NMR comparison of **13** (top) and **14** (bottom), showing the differences between the methyl ester and spacer arm CH_2 peaks of the two compounds.

lithium hydroxide with conditions described in Table 3.3.



(14) **(15)**

Scheme 3.4: Hydrolysis of diester **14** to form diacid **15** was attempted using both NaOH and LiOH, however decomposition of starting material was observed.

Initial hydrolysis attempts using NaOH did not result in the isolation of diacid **15**. In the first instance, extraction of the diacid from the aqueous phase was attempted using ethyl acetate, however no desired product was isolated. A more polar solvent, butanol, was then used in several experiments to determine whether ethyl acetate was insufficiently polar to extract **15** from the aqueous phase. However, extraction with butanol also failed to isolate the diester starting material or diacid **15**. Starting material and diacid **15** could not be observed using NMR or HRMS in positive or negative mode from the crude reaction mixtures. Masses corresponding to decomposition products arising from reduction of the disulfide were observed by MS in negative mode, however conclusive NMR evidence could not be obtained for these structures. Hence, due to the inability to isolate starting material or product

Table 3.3: Reaction conditions used and results obtained for the hydrolysis of diester **14** to give diacid **15**.

Reagents	**Solvent, Rxn Time**	**Result**
1.6 M NaOH	THF, MeOH, 6h	Decomposition
1.6 M NaOH	THF, MeOH, 18h	Decomposition
1.0 M LiOH	THF, 6h	Decomposition
1.0 M LiOH	THF, 18h	Decomposition

and the lack of supporting NMR and MS data, it was surmised that saponification with NaOH resulted in decomposition of diester **14**.

The use of a weaker base, LiOH, was also attempted for saponification. The same issues pertaining to the inability to isolate starting material or product from crude and worked up mixtures remained prevalent. Amberlite IR-120(H^+) resin was used according to the procedure by Blanco *et al*[170] to cation exchange lithium ions attached to each carboxylic acid group for hydrogen ions, in favour of an acidic work up. However, diester **14** and diacid **15** still could not be detected using NMR or HRMS in positive or negative mode, again suggesting that diester **14** undergoes decomposition in basic conditions.

The decomposition of diester **14** could be occurring at the disulfide in dilute alkali, however, conclusive NMR evidence could not be obtained. However the degradation of disulfide bonds between cysteine residues within peptides and proteins using dilute alkali is documented,[171,172] and avoiding the use of NaOH and LiOH through the use of different protecting groups should be investigated in future studies.

3.2.2 Dissociation of the cleavable bond using CID

While it was not possible to form the target linker, it was still possible to investigate the MS fragmentation properties of the disulfide bond. The capacity of the disulfide bond to be cleaved using low energy CID was investigated through MS/MS analysis of **14**. Diester **14** was predicted to produce a slightly different marker ion pattern of 2, 125, 2 Da (Figure 3.8).

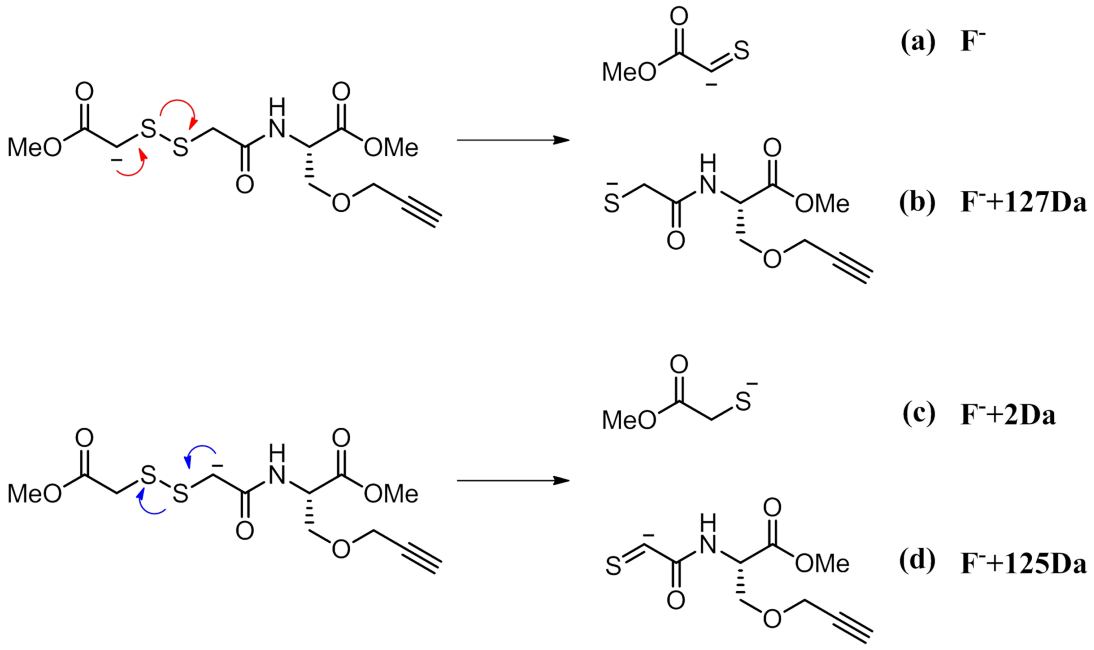


Figure 3.8: CID cleavage products of diester **14**, **(a)**, **(b)**, **(c)** and **(d)**. Ions **(a)** and **(d)** are formed through charge transfer via an ion-neutral complex.

During heterolytic bond cleavage within the mass spectrometer, one charged and one neutral species are formed. The species which is more substituted or has a greater mass is more likely to be able to stabilise the charge during bond dissociation. Hence it is more common to observe the larger mass product in the fragmentation of smaller molecules. As a result of this, ions **(b)** and **(d)** (Figure 3.8) are more likely to be observed from the low energy fragmentation of diester **14** than ions **(a)** and **(c)**.

The MS/MS spectrum of **14** (Figure 3.9) was obtained in negative mode using a 'low' collision energy of 15 V with Ar collision gas. Dissociation of the disulfide bond was successfully observed, yielding the expected **(b)** and **(d)** fragment ions at 230 and 228 m/z. As a result of this, the disulfide bond has been identified as a strong cleavable bond candidate for MS/MS experiments in the proposed modular design context.

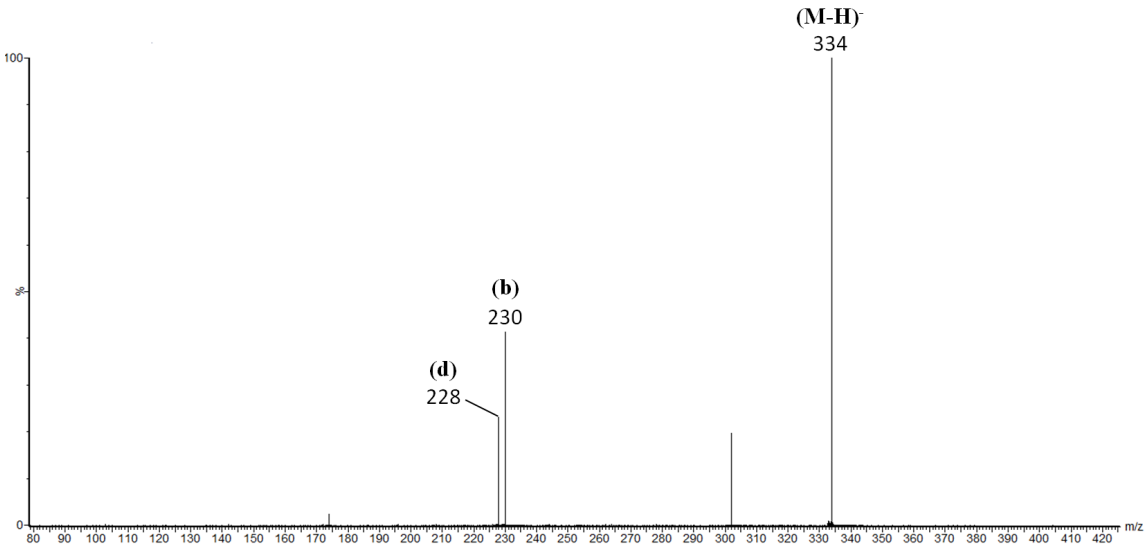


Figure 3.9: MS/MS spectrum of diester **14**

3.3 Crosslinker 1c

To further address the first aim of synthesising crosslinkers containing a cleavable bond and affinity tag, the synthesis of crosslinker **1c** was attempted, containing a cleavable bond in an elongated spacer arm, and an affinity tag.

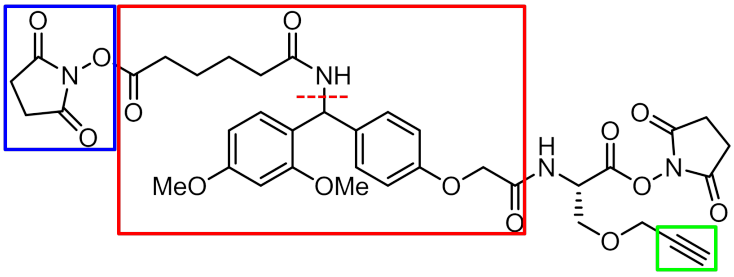


Figure 3.10: Crosslinker **1c**, cleavable bond denoted by red line.

Crosslinker **1c** was designed to comply with the modular synthesis detailed in Section 2.2, p. 6. Crosslinker **1c** contains a Rink structure reported by Henkel *et al* used in solid phase peptide synthesis, which has been shown to cleave at the carbon-nitrogen bond using low energy CID in positive mode by Tang *et al.*[173,174] The carbon-nitrogen bond produces a characteristic mass signature upon cleavage, yielding up to two ions (Figure 3.11) with a mass difference of 535 Da, which can be recognised using specialised software. Biotin is present within the crosslinked product as MS analysis takes place after Huisgen cycloaddition and avidin affinity purification.

The carbon-nitrogen bond is part of an elongated spacer arm module, which requires an extra peptide coupling step after coupling to the affinity tag module, to assemble the entire spacer arm. Whilst altering the spacer arm requires the addition of two more synthetic steps, the connection process, amide bond formation, remains the same. The first half of the spacer arm contains the carbon-nitrogen bond, and the second half consists of a simple alkyl chain. Due to the elongated spacer arm, crosslinker **1c** will have the capacity to probe protein residues which are further apart within a 3-D structure. The crosslinker utilises the modular structure proposed in Figure 2.2 (p. 28).

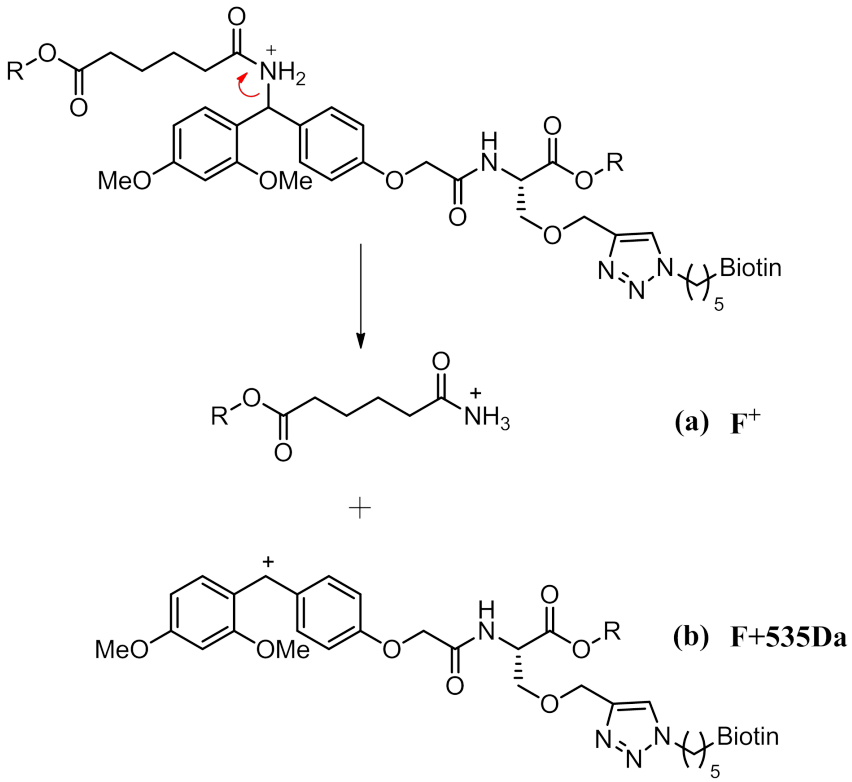
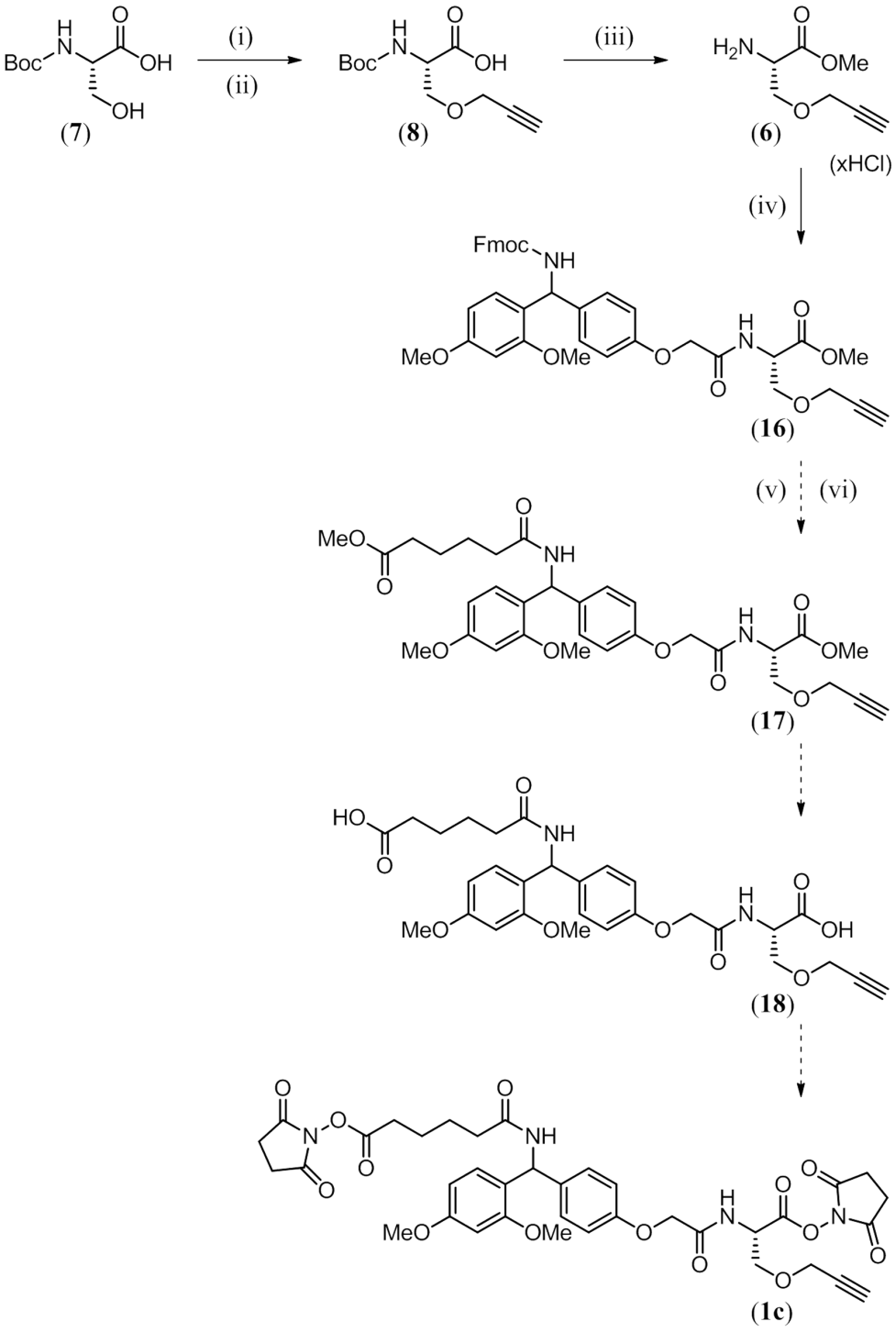


Figure 3.11: Possible CID mechanisms and products of Crosslinker **1c**, **(a)** and **(b)** with mass relative to **(a)**. Ion **(a)** is formed through charge transfer via an ion-neutral complex. In an asymmetrical crosslink, R chains will have different structures and masses, in which case 4 characteristic ions will be produced.

The design of crosslinker **1c** addresses the current analytical challenges facing the CXMS technique relating to the site specific location of crosslinked residues through the introduction of a bond cleavable by gas phase dissociation, and the detection and identification of crosslinked peptides through the inclusion of an affinity tag.

3.3.1 Synthesis of Crosslinker 1c

Synthesis of Crosslinker **1c** was attempted according to Scheme 3.5.



Scheme 3.5: Synthetic scheme for crosslinker **1c**. *Reagents and conditions*: (i) NaH, DMF, (ii) propargyl bromide, (iii) MeOH, $SOCl_2$, (iv) Rink amide linker, anh. DMF, DIPEA, HATU, (v) 20% piperidine, DCM, (vi) Monomethyl adipate, anh. DMF, DIPEA, HATU.

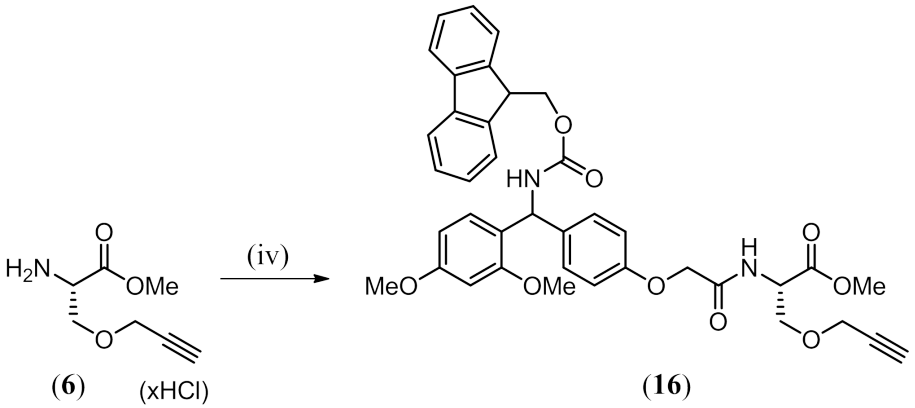
The alkyne module **8** was synthesised in 92% yield from **7** without further purification, as described in section 2.3.1, p. 32.

Assembly of the spacer arm in crosslinker **1c** required two peptide coupling steps and the removal of an Fmoc protecting group. The first half of the spacer arm contained the cleavable carbon-nitrogen bond within a commercially available monoacid, rink amide linker (RA). The RA also contained an Fmoc protected amine which could be utilised in the second peptide coupling step. Coupling of RA to hydrochloride salt **6** was attempted to yield monoester **16** using both carbodiimide and triazole based coupling reagents (Table 3.4).

Table 3.4: Reaction conditions used and results obtained for the coupling of RA to hydrochloride salt **6**.

Reagents	**Solvent, Rxn Time**	**Result**
RA (1.0 eq), EDC-HCl (1.3 eq), HOBt (1.3 eq), DIPEA (4.0 eq)	DMF, 18h	Starting material
RA (1.0 eq), HATU (1.2 eq), DIPEA (4.0 eq)	DMF, 18h	monoester **16**, 43%

Initial coupling efforts using the carbodiimide based coupling reagent EDC did not yield monoester **16**. However, coupling with the triazole based reagent HATU gave monoester **16** in 43% yield after purification (Scheme 3.6). HATU has been shown in literature to be more efficient than EDC and other carbodiimide based reagents for unhindered substrates, particularly for shorter reaction times,[144,143] consistent with the difference in reactivity observed in the formation of monoester **16**.



(**6**) (xHCl) (**16**)

Scheme 3.6: Coupling of alkyne module **6** to the spacer arm using (iv) RA, anh. DMF, HATU, DIPEA to give monoester **16**. FMOC structure shown to aid in clarity of NMR assignment.

COSY NMR spectroscopy was used to aid in the assignment of the HNMR spectrum of monoester **16** (Figure 3.12). Assignment of NH resonances was made by obser-

vation of correlations between the N**H** doublet signal at 8.39 and the multiplet at 4.65-4.60 ppm which corresponds to the **H**αC of the alkyne tag. Similarly, the N**H** doublet at 8.13 forming a correlation with the doublet at 6.06 ppm was assigned to the N**H** of the rink structure. The Fmoc aromatic signals could be identified through the correlation of 4 signals within the aromatic region, two doublets at 7.88 and 7.72, and two triplets at 7.41 and 7.32 ppm. No other aromatic structure within **12** could have 4 correlating hydrogen signals.

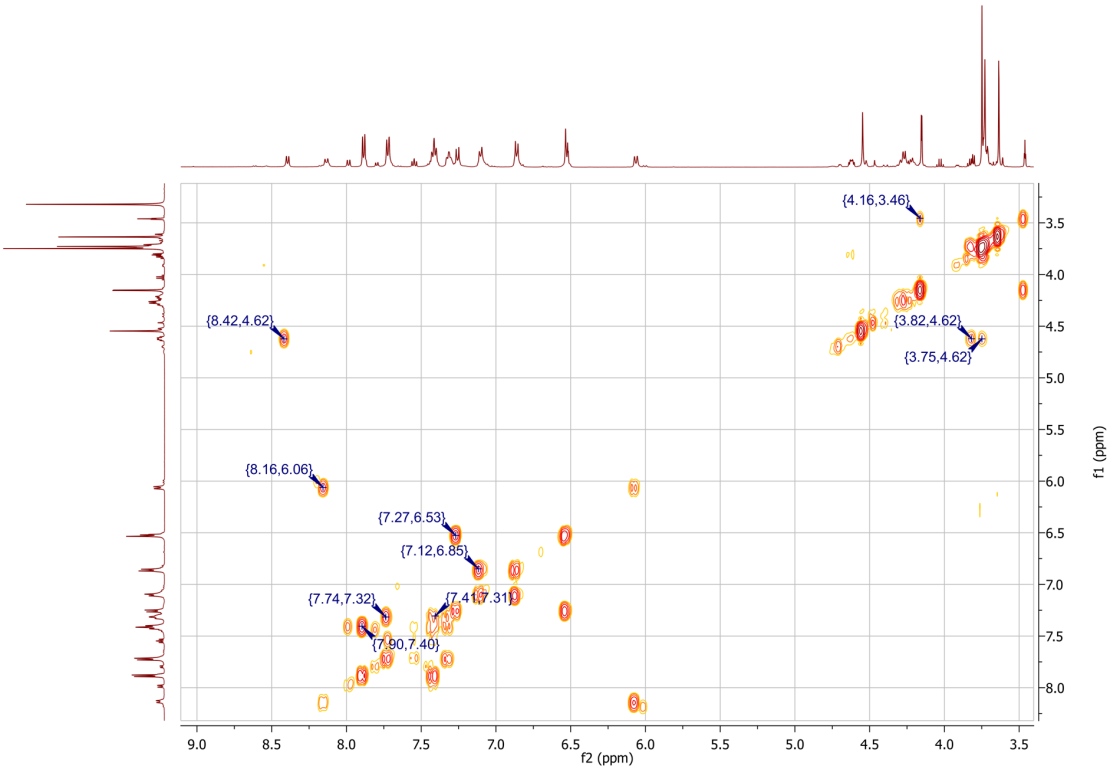


Figure 3.12: Hydrogen correlation spectrum of **16**.

Differentiation of the remaining aromatic signals between the rings containing the methyl esters and connected to the modified amino acid was more difficult. Correlation was observed between the 2H doublet at 7.10 ppm and the 2H doublet at 6.86 ppm, and between the 1H doublet at 7.26 ppm and 2H multiplet at 6.53-6.51 ppm. The correlation between 7.10 and 6.86 ppm was assigned to the aromatic ring connected to the modified amino acid as the signal integration indicated 4H, with each signal representing 2H as expected, as the ether ring contained a plane of symmetry. Hence the correlation between 7.26 and 6.53-6.51 was assigned to the methyl ester ring as the integration indicated 3H as expected.

Signals corresponding to the methyl ester peaks were observed in the ^{1}HNMR spec-

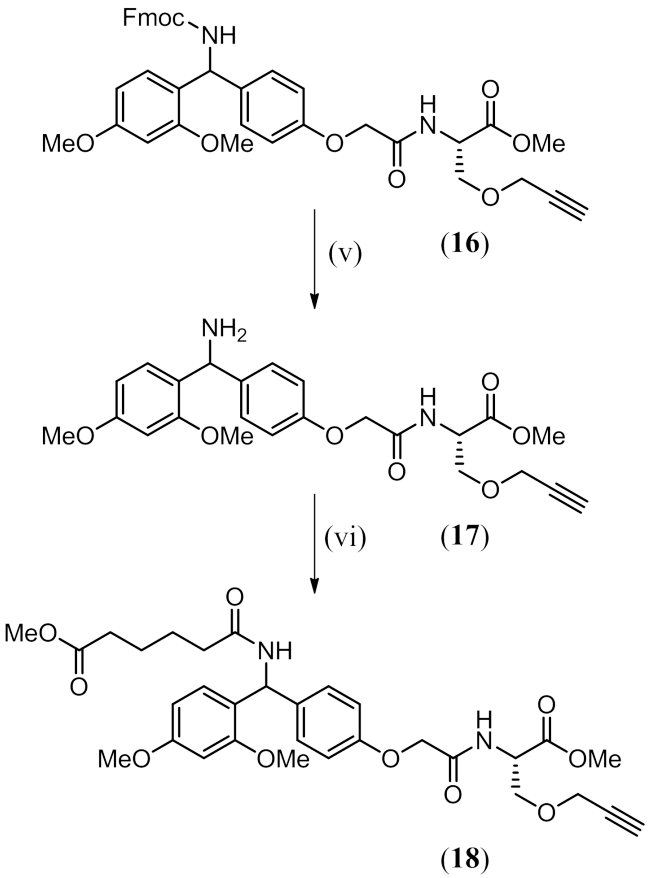
trum as expected overlapped with a multiplet at 3.76-3.72 and as a singlet at 3.64 ppm. The alkyne hydrogen triplet at 3.46 ppm and remaining signals appeared as expected, allowing for characterisation of monoester **16**.

Following the isolation of monoester **16**, sequential Fmoc deprotection and coupling steps were attempted to attach the second half of the spacer arm. The Fmoc group is labile in base such as diethylamine (DEA) and in particular secondary amines such as piperidine and piperazine capable of trapping the dibenzofulvene generated during removal.[146,175,176] The volatile base DEA was the first to be tried according to a procedure by Bodanszky *et al*,[177] as it was comparatively easier to remove than other amenable bases such as piperidine, and commonly utilised for solution phase peptide syntheses.[176]

Fmoc deprotection of **16** was attempted using DEA in DMF for 18 hours. DMF was removed through the formation of an azeotrope with toluene and NMR analysis of the crude amine **17** was attempted. However, no sample peaks were observed by NMR due to the presence of DMF. Starting material was observed by HRMS, with no evidence of **17**. As a result of this, Fmoc removal was then attempted using piperidine in DCM for 18 and 24 hours. Piperidine in DMF is commonly used in solid phase Fmoc removal protocols, taking between 10 and 40 minutes, as removal is better facilitated by a polar solvent. However, in solution Fmoc removal in DCM is documented on a scale of hours or days due to the less polar nature of the solvent.[146,178] The use of DCM was also desirable as it was comparatively easy to remove *in vacuo*, making it more amenable for subsequent analysis by NMR. The NMR spectrum of crude **17** yielded recognisable resonances. In methanol, one methyl ester and two methoxy peaks were observed at 3.76, 3.71 and 3.65 ppm and an alpha hydrogen signal at 4.68 ppm which corresponded to signals observed in the NMR of the starting material monoester **16**. The aromatic region could not be analysed to determine whether the Fmoc peaks had disappeared as the peaks did not resolve in any NMR solvent.

Despite inconclusive NMR analysis, attachment of the second half of the spacer arm to form diester **18** was then attempted through the coupling of the crude residue of **17** with monomethyl adipate (Scheme 3.7). Coupling was facilitated by HATU and DIPEA in DMF and purification attempted using normal phase silica chromatography in an ethyl acetate/petroleum ether solvent system. Diester **18** was detected using HRMS in both the 18 and 24 hour reactions, however a sufficiently pure sample for NMR analysis was not isolated. Further solvent systems must be trialled to obtain a pure sample of **18**. Yield of **18** may also be increased through

the use of piperidine in DMF for the removal of Fmoc, as the ability to form diester **18** through amine **17** has now been established.



Scheme 3.7: Fmoc deprotection of monoester **16** using (v) 20% piperidine/DCM to give amine **17**, and subsequent amide coupling (vi) monomethyl adipate, anh. DMF, HATU, DIPEA to give diester **18**.

3.3.2 Dissociation of the cleavable bond using CID

The ability of the carbon-nitrogen bond to be cleaved using low energy CID was investigated through the cleavage of **12**. Monoester **16** had not undergone Huisgen cycloaddition with biotin azide, hence cleavage products of compound **12** were predicted to produce ions at 440 and 240 m/z (Figure 3.13). The more stabilised ion is **(b)**, due to resonance stabilisation of the positive charge within the aromatic ring structures upon bond dissociation, hence it was more likely to observe ion **(b)** than **(a)**.

The cleavage spectrum of the sodiated ion of **16** was obtained in positive mode using a 'low' collision energy of 35 V with Ar collision gas. Dissociation of the carbon-

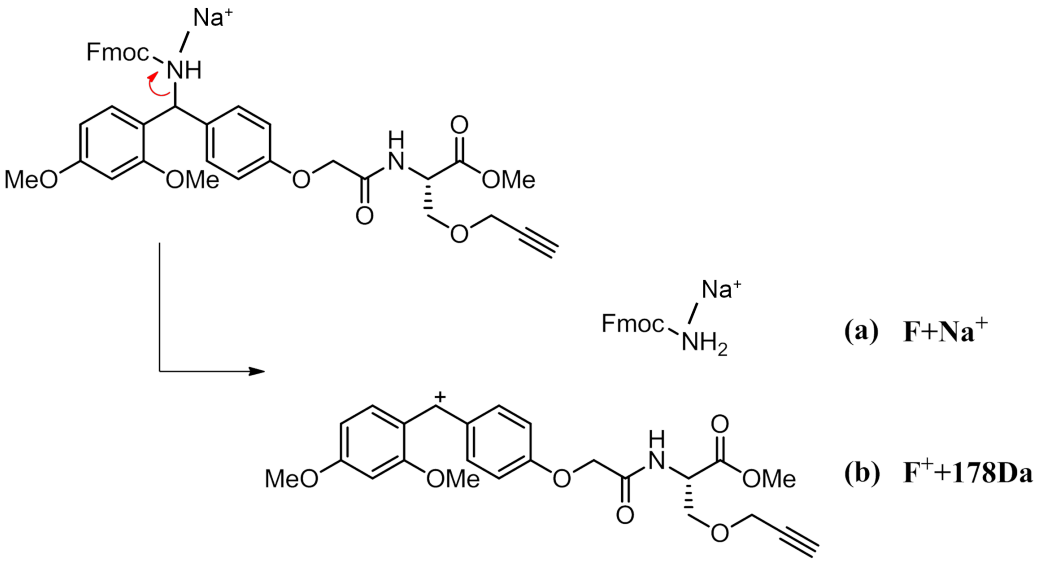


Figure 3.13: The CID cleavage products of monoester **16**, **(a)** and **(b)**, with mass relative to **(a)**. Ion **(a)** is formed through charge transfer via an ion-neutral complex.

nitrogen bond was successfully observed, yielding the predicted **(b)** fragment ion at 440 m/z.

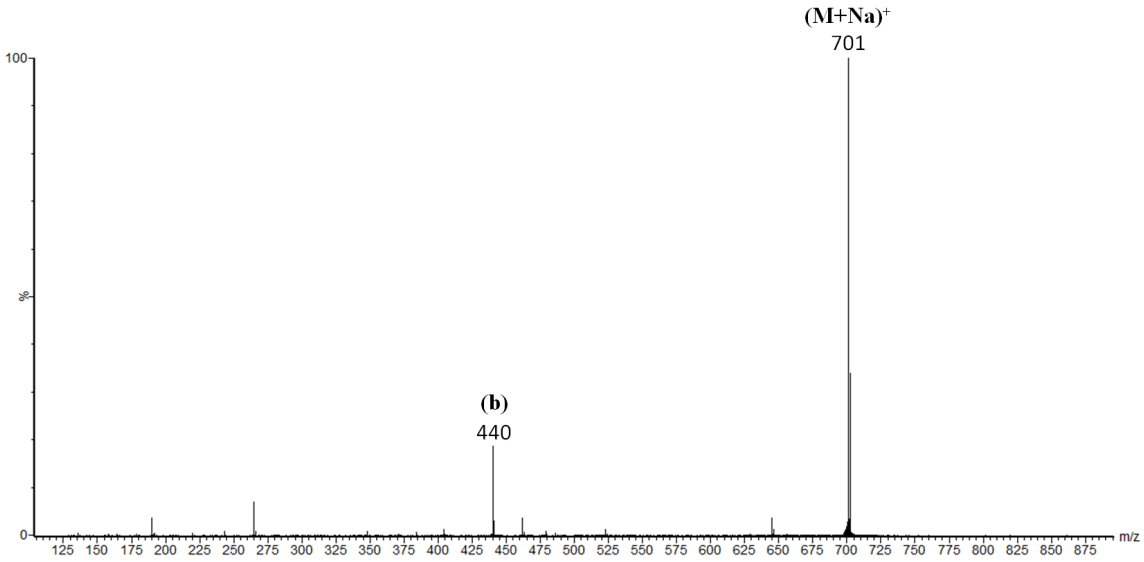


Figure 3.14: MS/MS spectrum of monoester **16**

3.4 Crosslinking of *Sa*BPL

For future assessment of the utility of crosslinkers synthesised using the modular synthetic protocol for the formation of interprotein crosslinks, a model crosslinking experiment was developed using the *Staphylococcus aureus* biotin protein ligase enzyme. An x-ray crystal structure reported by Pendini *et al*[164] showed the close proximity of two pairs of lysine residues within the dimerisation interface of *Sa*BPL which could be probed using NHS activated crosslinkers (Figure 3.3).

To develop the crosslinking experiment, two commercially available crosslinkers were used, dithiobis(succinimidyl) propionate (DSP), and disuccinimidyl suberate (DSS). Both crosslinkers are amine reactive, to target the abundant lysine residues, and consistent with the chemistry of the novel crosslinkers. DSP has a spacer arm length of 12.0 Å and DSS 11.4 Å, both of which were longer than the largest inter-lysine distance of 8.01 Å within the dimerisation interface to be probed (Figure 3.3). The natural ligand substrates biotin and ATP were used to induce dimerisation of *Sa*BPL. Different reaction times and molar excesses of DSS and DSP were trialled to optimise the crosslinking step, and the success of crosslinking was determined by SDS-PAGE analysis followed by Coomassie blue staining. The reducing agent dithiothreitol (DTT) commonly used in SDS-page denaturing loading buffers could not be used in experiments utilising DSP due to the possibility of reducing the disulfide bond within the crosslinker.

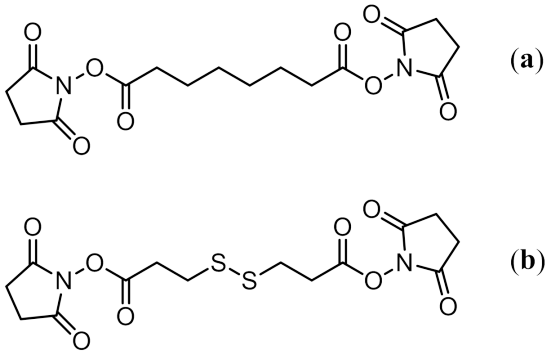


Figure 3.15: Commercially available crosslinkers **(a)** DSS and **(b)** DSP.

A key advantage of this system is that several controls could be used to ensure that the dimer observed was caused by the formation of physiologically relevant crosslinks. A monomeric F123G mutant of *Sa*BPL was utilised, in which a phenylalanine residue was mutated to a glycine. Mutation of this residue has been shown to limit *Sa*BPL homodimer formation in solution as it forms a critical interaction for the stabilisation of the homodimer.[166] The F123G control is important as the forma-

tion of false positive crosslinks, or artefactual associations between proteins which do not naturally associate in protein interaction studies, must be minimised. The association of proteins which do not naturally interact within a biological system can lead to erroneous pathway mapping.

A negative control experiment in which no crosslinker was added (minus crosslinker) whilst maintaining all other reaction conditions was also used. If formation of dimers was observed by SDS-PAGE analysis in the minus crosslinker experiment, dimerisation bands observed in experiments with DSP could not be attributed to crosslink formation. A negative control (no reaction) which was not incubated at 37 °C, contained no biotin and ATP to induce dimerisation and no crosslinking reagent was also included to ensure that the reaction conditions were not causing *Sa*BPL degradation. The molecular weights of the of the *Sa*BPL monomer and dimer are approximately 38 and 76 kDa respectively. For SDS-PAGE analysis following crosslinking, slightly larger masses were expected to be observed due to the addition of crosslinker to the monomer/dimer structure. Crosslink formation with DSS would be expected to contribute approximately 140 Da per crosslink and DSP 176 Da per crosslink, leading to an increase in overall monomer and dimer mass.

To optimise crosslinking conditions, protein:crosslinker ratios of 1:10, 1:20, 1:30, 1:40 and 1:50 for both DSS and DSP were used with reaction times of 5, 10, 15 and 30 minutes and 1, 2 and 3 hours. SDS-PAGE analysis of the selected experiments, namely 5 - 30 minute reaction times for DSS, the 1:30, 1:40 and 1:50 molar equivalents of DSP and the negative control samples are shown in Figure 3.16. Dimer formation was not observed in the F123G, minus crosslinker and no reaction controls, meaning that observation of dimer in experiments with the crosslinkers DSS and DSP could be attributed to biologically relevant crosslink formation. Dimer bands were observed at approximately 76 kDa in all WT *Sa*BPL crosslinking experiments with DSS and DSP, demonstrating the successful interprotein crosslinking of the *Sa*BPL enzyme. Dimer formation was observed in all reactions from 5-30 minutes using DSS. As a result of this, future crosslinking reactions should not require longer than 30 minutes to form interprotein crosslinks. Dimer formation was also observed for the 1:30, 1:40 and 1:50 molar excesses probed by DSP, hence future experiments with *Sa*BPL should only require a 1:30 molar excess. The crosslinking experiments using a molar excess of lower than 1:30 did not show dimer formation (data not shown). The requirement of a minimum of 1:30 molar excess of crosslinker per monomer was not unexpected as each *Sa*BPL monomer contains 22 lysine residues, and hydrolysis of the crosslinker and side reactions with other amino acid residues may also occur (Section 1.3). The formation of more dimer was not observed for

higher molar excesses of crosslinker or longer reaction times in comparison with a 1:30 protein:crosslinker ratio and reaction times greater than 30 minutes. Hence there is unlikely to be an advantage in using higher crosslinker ratios and longer reaction times in future experiments.

In summary, a model crosslinking experiment to determine crosslinker utility was established through the formation of *Sa*BPL homodimers using a 1:30 molar excess of DSS and DSP, at reaction times varying from 5 to 30 minutes, observed by SDS-PAGE analysis. The conditions specified should be suitable for future crosslinking experiments with crosslinkers synthesised using the modular synthetic protocol, such as **1a**, **1b**, and **1c**.

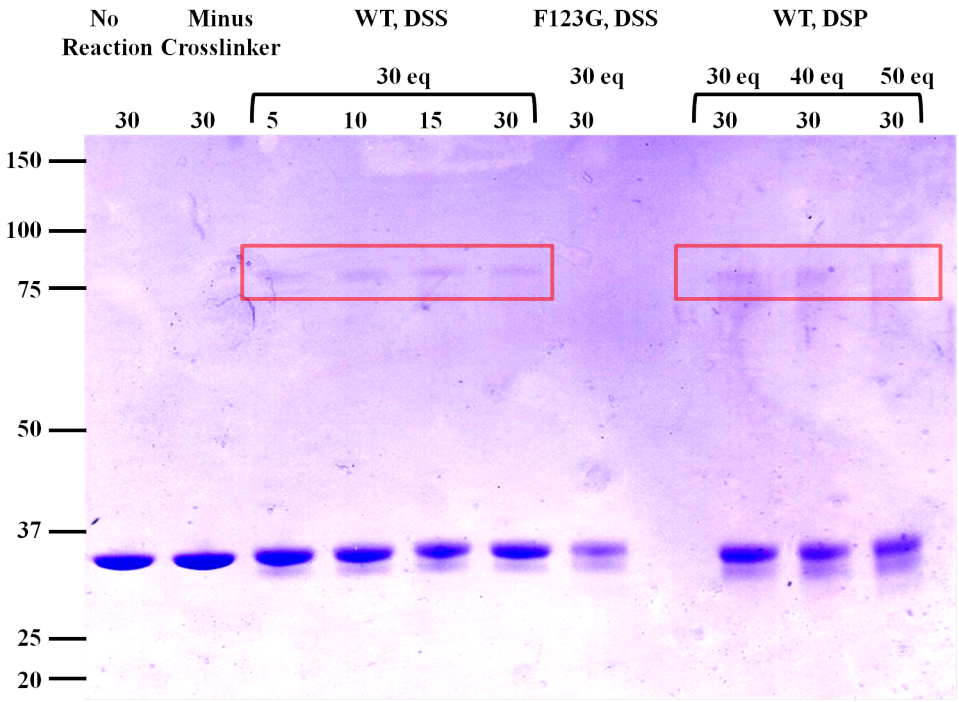


Figure 3.16: SDS-PAGE analysis of crosslinking reaction mixtures. Marker masses are shown on the left from 20 to 150 kDa, and reaction times in minutes above each lane. DSS displays a time course of crosslinking with different reaction times, and DSP shows crosslinking with different equivalents of crosslinker. Red boxes highlight *Sa*BPL dimer bands.

3.5 Conclusions and Future Work

This work aimed to further demonstrate the utility of a modular crosslinker synthesis by extending to more complex structures incorporating affinity tags as well as CID cleavable functional groups, as shown for crosslinkers **1b** and **1c**. The attachment of the spacer arm to the affinity tag module of crosslinker **1b** was successful, through peptide coupling with the triazole based coupling reagent HATU and subsequent esterification of the crude mixture of monoacid **12** and diaddition product **13** to yield diester **14**. Ensuing attempts to hydrolyse the methyl ester protecting groups did not yield diacid **15**, with decomposition observed. Consequently, further investigation is required to achieve the desired crosslinker **1b**.

The use of more labile or acid removable protecting groups in the synthesis of **1b** may allow for the isolation of diester **15**, and avoid the reduction of the disulfide bond. Formation of the disulfide bond *in situ* through the use of 2,2-bispyridyl disulfide as reported by Maruyama *et al*[179] may also provide an alternative approach to the synthesis of **1b**. Each half could be synthesised separately and assembled after methyl ester hydrolysis. However, the formation of the disulfide *in situ* could result in the formation of side products in which each half adds to form a symmetrical product, which would need to be separated from the desired asymmetric product.

Assembly of the spacer arm of crosslinker **1c** required two sequential deprotection and peptide coupling steps. The first half of the spacer arm containing the cleavable carbon-nitrogen bond was successfully attached to the affinity tag module using HATU, whilst coupling using EDC yielded starting material. Attempts to remove the Fmoc protecting group using DEA did not yield amine **17** as observed by NMR and HRMS. Fmoc removal in piperidine and DCM could not be confirmed by NMR. The crude sample of **17** was coupled to monomethyl adipate HATU and diester **18** was detected by HRMS, however a pure sample was not isolated in sufficient yield for NMR characterisation.

Diester **18** must be isolated and characterised by NMR. Yield of **18** may be improved by utilising piperidine in DMF rather than DCM, as polar solvents facilitate Fmoc removal more efficiently. Further investigation is required to define a solvent system to isolate **18** more effectively.

Low energy CID dissociation of the disulfide bond within diester **18** and the carbon-nitrogen bond of monoester **16** was attempted. Both the disulfide bond and the

carbon-nitrogen bond showed the desired dissociation at relatively low collisional energy values of 15 and 35 V respectively, validating the use of the cleavable bonds for CXMS studies. However, it would be necessary to confirm that the disulfide and carbon-nitrogen bonds cleave in preference to bonds within the peptide backbone, to ensure their utility within the MS3 crosslinking protocol.

A crosslinking assay was developed, based on the dimerisation of the *Sa*BPL enzyme, which could be utilised to determine the functionality of novel crosslinkers synthesised using the modular synthetic protocol. The commercially available crosslinkers DSS and DSP were used to develop this assay and successfully covalently crosslink *Sa*BPL dimers at a minimum molar excess of 1:30 with *Sa*BPL monomer and reaction time of 5-15 minutes at 37°C, utilising biotin and ATP as dimerisation inducers.

The utility of crosslinkers **1a**, **1b**, and **1c** should now be determined through the crosslinking assay devised with *Sa*BPL. Methods for Huisgen azide-alkyne cycloaddition between crosslinked peptides and biotin azide must be investigated, and a method for avidin affinity purification of crosslinked species defined in order to establish a complete method for sample preparation for CXMS using the modular synthetic method.

3.6 Acknowledgements

I would like to acknowledge and thank Louise Sternicki and Ashleigh Paparella for providing the *Sa*BPL utilised in the crosslinking assay, and Louise for assistance with the SDS-PAGE and setting up of crosslinking reactions. I acknowledge and thank Aimee Horsfall for assisting with several reactions during the optimisation of peptide coupling for crosslinker **1c**.

3.7 Experimental

3.7.1 Mass Spectrometry

MS/MS spectra were obtained using a Micromass Q-ToF 2 mass spectrometer (Waters, Manchester, UK). Samples were introduced into the mass spectrometer through

nano-electrospray ionisation using platinum coated borosilicate capillaries prepared in-house. Experimental conditions: capillary voltage 1.2 kV in negative mode and 1.4 kV in positive mode, source temperature 50 °C, and cone voltages between 30-40 V. Argon collision gas was used at 'low energy' values of 15-35 V.

3.7.2 *Sa*BPL

Materials and solutions

Reagents DSS and DSP, Biotin, MES SDS Running Buffer, Bradford Protein Reagent Concentrate and Novex Precast Gels were purchased from Thermo Fisher Scientific, Sigma Aldrich Bio-Rad and Invitrogen Life Technologies. Coomassie blue stain consisted of 0.2% (w/v) Coomassie brilliant blue, 10% ethanol (v/v) and 10% (v/v) acetic acid. Coomassie destain consisted of 10% (v/v) methanol and 10% acetic acid. 2x Gel loading buffer consisted of 0.1 M Tris pH 7, 8% (v/v) glycerol, 4% (w/v) SDS and 0.2% (w/v) bromophenol blue. WT and F123G *Sa*BPL were obtained courtesy of Louise Sternicki and Ashleigh Paparella at the School of Biological Sciences, University of Adelaide, Adelaide, South Australia. The WT and F123G monomeric mutant *Sa*BPL were produced by recombinant protein expression and stored in phosphate buffer comprised of 1x PBS, 100 mM KCl, 5% glycerol and 1 mM EDTA pH 8.0. Protein concentration was determined using a Bradford assay.[180]

Crosslinking and SDS-PAGE analysis

Crosslinking reactions were made up to a total volume of 10 µL and included WT and F123G monomeric mutant *Sa*BPL in the phosphate storage buffer at a concentration of 13 µM, saturating concentrations of biotin and ATP at 100 µM and 10 mM respectively and 10 mM $MgCl_2$. Concentrated stock solutions of 30 M DSS or DSP were made in DMSO and diluted into PBS to give the appropriate molar excesses of 1:20, 30, 40 and 50 of DSS or DSP while minimising the addition of DMSO. The reactions were incubated at 37 °C for 5 minutes to 3 hours. Samples were made up to 20 µL in 2x gel loading buffer and boiled at 100 °C for 5 minutes. Samples were separated on a 4-12 % Bis-Tris polyacrylamide precast gel and electrophoresed at 200 V in MES SDS running buffer. Staining of the resulting gel with Coomassie blue stain was undertaken for 1 hour before destaining with Coomassie destain at room temperature.

3.7.3 General Synthetic Methods

All starting materials were purchased from Sigma Aldrich, GL Biochem and Genetech unless otherwise stated, and were used as received. All NMR data was analysed using MestReNova 6.0 from Mestrelab Research S.L. All organic extracts were dried over anhydrous sodium sulfate. Thin layer chromatography was performed using Merck aluminium sheets with silica gel 60 F_{254}. Compounds were visualised using a UV lamp and/or with a potassium permanganate stain (3.0 g $KMnO_4$, 20.0 g K_2CO_3, 5 mL 10% NaOH, 300 mL H_2O) followed by heating. All Rf values were taken to the nearest 0.01. All reported yields were judged to be isolated yields, unless otherwise specified, by TLC and NMR. Flash chromatography was performed using Grace Davidson Discovery Sciences Davisil 40-63 grade silica gel, under positive nitrogen pressure. All solvents for chromatography were used as received. Infrared spectra were obtained using a Perkin Elmer BX FT-IR spectrometer. High field NMR spectra were recorded using an Agilent DD2 spectrometer operating at 500 MHz for ^1H and 125 MHz for ^{13}C. Solvents used to obtain spectra are specified. Chemical shifts are reported in ppm on the δ-scale relative to TMS (δ_H 0.00), $CDCl_3$ (δ_H 7.26, δ_C 77.16) or DMSO-d_6 (δ_H 2.50, δ_C 39.52). Spin multiplicities are reported as (br s) broad singlet, (s) singlet, (d) doublet, (t) triplet, (q) quartet, and (m) multiplet. All J values were rounded to the nearest 0.1 Hz and correspond to H-H coupling unless otherwise specified. All reactions done under an inert atmosphere (N_2) were carried out with oven dried glassware. Removal of solvent under reduced pressure refers to the removal of a large quantity of solvent by rotary evaporation followed by a high vacuum pump for a minimum of 1 hour. High resolution masses were measured using an Agilent Technologies 6230 Accurate-Mass TOF LC/MS to 0.0001 m/z.

General Procedure A: HATU Mediated Peptide Coupling
The respective amine was dissolved in dry N'N dimethylformamide (3 mL/100 mg of amine) and DIPEA (4.0 mol eq) added whilst stirring. The relevant carboxylic acid (1.0 mol eq) was added to the stirred solution followed by HATU (1.2 mol eq) and anhydrous HOBt (1.0 mol eq) and the solution stirred at room temperature under a nitrogen atmosphere overnight. The solution was partitioned between ethyl acetate and 1M HCl. The organic phase was then separated and washed with 1M HCl (2x), water (2x) and brine; dried over Na_2SO_4 and the solvent removed *in vacuo*. Where applicable the crude product was purified by flash chromatography on normal phase silica to yield the desired amides.

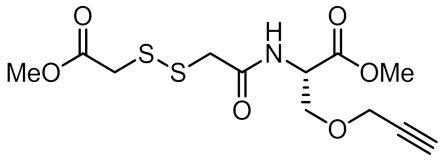
General Procedure B: EDC Mediated Peptide Coupling
The relevant carboxylic acid (1.0 mol eq) was dissolved in dry N'N dimethylformamide (3 mL/ 100 mg of amine) and EDC-HCl (1.3 mol eq) and anhydrous HOBt (1.3 mol eq) added whilst stirring. The relevant amine (1.0 mol eq) was added to the stirred solution, followed by DIPEA (4.0 mol eq) and the solution stirred at room temperature under a nitrogen atmosphere overnight. The solution was partitioned between ethyl acetate and 2M HCl. The organic phase was separated and washed with 2M HCl (2x), water (2x) and brine, dried over Na_2SO_4, and the solvent removed *in vacuo*. Where applicable the crude product was purified by flash chromatography on normal phase silica to yield the desired amides.

General Procedure C: Esterification of Carboxylic Acids
The relevant carboxylic acid was dissolved in methanol (100 mL/1 g of acid) and stirred on ice for 5 minutes. Thionyl chloride (2.0 mol eq) was added dropwise under a nitrogen atmosphere to the stirred solution. The mixture was stirred on ice for a further 15 min and then allowed to warm to room temperature and react for 6 h. The solvent was then removed *in vacuo*. The residue was re-dissolved in methanol and the solvent removed *in vacuo* three times to yield and oil. Where applicable the crude product was purified on normal phase silica to yield the desired esters.

3.7.4 Synthesis and Characterisation

Synthetic procedures and NMR data for alkyne **7** and hydrochloride salt **6** are reported in Chapter 2, Section 2.5.2, (p. 42).

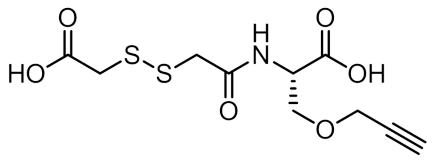
(*S*)-methyl 2-(2-((2-methoxy-2-oxoethyl)disulfanyl)acetamido)-3-(prop-2-yn-1-yloxy)propanoate (14)



Hydrochloride salt **6** (800 mg, 4.1 mmol) was coupled to dithiodiglycolic acid according to General Procedure A to yield a brown oil. The crude product was esterified according to General Procedure C to yield a brown oil. The residue was purified by flash chromatography on silica (ethyl acetate/petroleum ether, 1:1) to yield **14**

as an orange oil (378 mg, 27%). R_f: 0.26, IR (neat): 3280 cm^{-1} (N-H), 2953 cm^{-1} (C-H), 1734 cm^{-1} (C=O).
^{1}H NMR (500MHz, $CDCl_3$, δ) 7.20 (1H, d, J = 7.8 Hz, N**H**), 4.79 (1H, dt, J = 8.0, 3.3 Hz, **H**αC), 4.15 (2H, d, J = 2.4 Hz, OC**H**$_2$CCH), 3.99 (1H, dd, J = 9.5, 3.4 Hz, HαC**H**HO) 3.81 (1H, dd, J = 9.5, 3.3 Hz, HαCH**H**O), 3.77 (3H, s, OC**H**$_3$), 3.76 (3H, s, OC**H**$_3$), 3.64-3.58 (2H, m, C**H**$_2$SS), 3.58-3.51 (2H, m, SSC**H**$_2$), 2.45 (1H, t, J = 2.4 Hz, CH_2CC**H**). ^{13}C NMR (500 MHz, $CDCl_3$, δ) 170.3, 170.2, 168.0, 78.9, 75.3, 69.3, 58.6, 52.9, 52.8, 52.7, 42.5, 41.3.
HRMS (ES+) calculated for $C_{12}H_{17}NO_6S_2$ $[M+H]^+$ = 336.0576, found $[M+H]^+$: 336.0534 m/z.

(*S*)-2-(2-((carboxymethyl)disulfanyl)acetamido)-3-(prop-2-yn-1-yloxy)-propanoic acid (15)

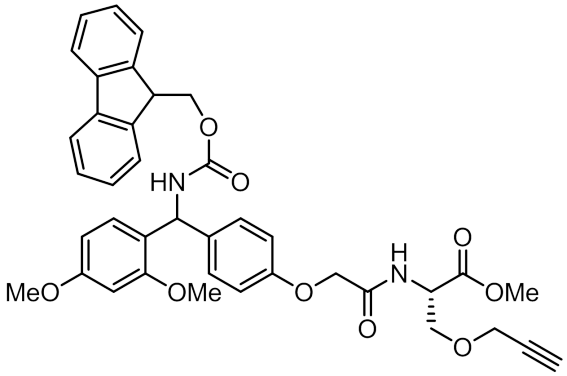


Synthesis of **15** was attempted using the reagents outlined in Table 3.3.

Using NaOH in THF and MeOH:
Diester **14** was dissolved in THF (2.0 mL/100 mg), 1.6 M sodium hydroxide (1.0 mL/100 mg) and methanol (1.0 mL/100 mg) and the mixture stirred at room temperature for 6 or 18h. The solvent was removed *in vacuo* and the residue partitioned between ethyl acetate and 1M HCl. The organic layer was isolated and the aqueous layer extracted with ethyl acetate (4x). The combined organic extracts were washed with brine and dried over Na_2SO_4. The solvent was removed *in vacuo* to yield a yellow oil. Decomposition was observed.

Using LiOH in THF:
Diester **14** was dissolved in THF (2.0 mL/100 mg) and 1.0 M LiOH (1.0 mL/100 mg) and the mixture stirred at room temperature for 6 or 18h. The THF was removed *in vacuo* and the resulting aqueous solution washed with diethyl ether (2x). The aqueous extract was treated with Amberlite IR-120 until reaching pH 6. The resin was filtered off and washed with deionised water. The filtrate was lyophilised to give a yellow oil. Decomposition was observed.

(2*S*)-methyl 2-(2-(4-(((((9H-fluoren-9-yl)methoxy)carbonyl)amino)(2,4-dimethoxyphenyl)methyl)phenoxy)acetamido)-3-(prop-2-yn-1-yloxy)pro-panoate (12)

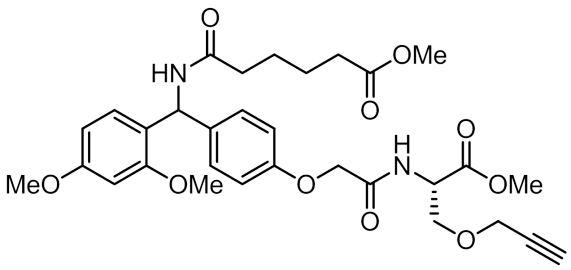


Hydrochloride salt **6** (500 mg, 2.6 mmol) was coupled to rink amide linker according to General Procedure A. The residue was purified by flash column chromatography on silica (ethyl acetate/petroleum ether, 1:1) to yield a tan solid (754 mg, 43%). Mp: 80.4 - 82.8 °C, R_f: 0.18, IR (neat): 3417 cm^{-1} (N-H), 3284 cm^{-1} (N-H), 2952 cm^{-1} (C-H), 1720 cm^{-1} (C=O).

^{1}H NMR (500MHz, $CDCl_3$, δ) 8.39 (1H, d, J = 8.0 Hz, N**H**CHCO), 8.13 (1H, d, J = 9.2 Hz, N**H**CHArom), 7.88 (2H, d, J = 7.5 Hz, FmocArom), 7.72 (2H, d, J = 8.2 Hz, FmocArom), 7.41 (2H, t, J = 7.5 Hz, FmocArom), 7.32 (2H, t, J = 6.1 Hz, FmocArom), 7.26 (1H, d, J = 8.9 Hz, MeOArom), 7.10 (2H, d, J = 8.4 Hz, Arom), 6.86 (2H, d, J = 8.7 Hz, Arom), 6.53-6.51 (2H, m, MeOArom), 6.06 (1H, d, J = 9.1 Hz, NHC**H**Arom), 4.65-4.60 (1H, m, **H**αC), 4.55 (2H, s, OC$\mathbf{H}_2$CONH), 4.31-4.20 (3H, m, C**H**C$\mathbf{H}_2$O), 4.15 (2H, d, J = 2.4 Hz, OC$\mathbf{H}_2$CCH), 3.82 (1H, dd, J = 9.8, 5.7 Hz, HαCC**H**HO), 3.76-3.72 (7H, m, OC$\mathbf{H}_3$, OC$\mathbf{H}_3$, HαCCH**H**O), 3.64 (3H, s, OC$\mathbf{H}_3$), 3.46 (1H, t, J = 2.3 Hz, CH_2CC**H**). ^{13}C NMR (125 MHz, DMSO-d_6, δ) 173.2, 171.1, 162.7, 160.0, 159.5, 158.6, 147.0, 146.9, 143.8, 138.8, 131.2, 130.9, 130.7, 130.4, 130.2, 128.4, 126.1, 123.2, 117.4, 107.8, 101.4, 82.8, 80.8, 71.5, 69.7, 68.7, 60.9, 58.7, 58.3, 55.3, 54.9, 54.2, 49.9.

HRMS (ES+) calculated for $C_{39}H_{38}N_2O_9$ $[M+Na]^+$ = 701.2474, found $[M+Na]^+$: 701.2499 m/z

methyl 6-(((2,4-dimethoxyphenyl)(4-(2-(((*S*)-1-methoxy-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-2-oxoethoxy)phenyl)methyl)amino)-6-oxo hexanoate (18)



Monoester **16** (900 mg, 1.3 mmol) was dissolved in anh. DCM (20 mL) and stirred for 5 min. To the solution was added piperidine (5 mL) under a nitrogen atmosphere and the mixture stirred at room temperature for 5h. The solvent was removed *in vacuo* to give an orange oil (763 mg). The residue was coupled to monomethyl adipate according to General Procedure A. Purification of the residue was attempted by flash chromatography on silica (ethyl acetate/petroleum ether, 6:4), however pure compound remains to be isolated.
HRMS (ES+) calculated for $C_{31}H_{38}N_2O_{10}$ $[M+Na]^+$ = 621.2423, found $[M+Na]^+$: 621.2424 m/z

Summary

The work presented in Chapter 2 details the design of novel modular crosslinker **1** consisting of three parts, namely reactive groups, a spacer arm and an affinity tag. Retrosynthetic analysis of **1** illustrates the ability to include a cleavable bond, isotopic label and affinity tag within a single design utilising only the three modules highlighted, and demonstrates that **1** can be synthesised in a minimum of five steps from commercially available Boc-Serine **7**. The utility of this design was demonstrated through the synthesis of monoester **11** and crosslinker **1a**. All three modules could be assembled effectively. Coupling of the spacer arm and affinity tag could be achieved using well characterised peptide coupling conditions mediated by HATU and DIPEA, and introduction of NHS reactive groups was achieved using N-trifluoroacetoxy succinimide with crosslinker **1a** synthesised in trace quantities. Challenges encountered in the synthetic process could be overcome by attempts to functionalise the crosslinker with different reactive groups such as maleimides or hydrazides.

Chapter 3 presented the synthesis of crosslinkers **1b** and **1c**, the first to be synthesised containing a combination of a cleavable bond and affinity tag using the synthetic method devised in Chapter 2. The spacer arm and affinity tag could be attached successfully to form **14**, however ester hydrolysis did not proceed to allow formation of **1b**. Attachment of the spacer arm and affinity tag was achieved to form diester **18** in trace quantities and further work is required to determine whether **1c** can be synthesised in improved yields.

A crosslinking assay utilising *Sa*BPL and the commercially available crosslinkers DSS and DSP was also described. The assay was developed to test all novel synthesised crosslinkers for the ability to form type 2 interprotein crosslinks which are pivotal in the study of interprotein associations and protein quaternary structure. *Sa*BPL was shown to form homodimers by SDS-PAGE analysis when crosslinked with both DSS and DSP, defining reaction conditions and providing a standard

reference to be used when testing any synthesised crosslinkers.

The work detailed in this thesis presents the first analytical ground work toward the development of a robust and adaptable method for the synthesis of modular chemical crosslinkers which have the potential to address the analytical challenges facing CXMS in the downstream analysis of crosslinked peptides. Low resolution techniques such as CXMS are becoming more widely used to aid in the structural characterisation of proteins and their role in complex biological processes. Continued innovation in the analysis of crosslinked species is crucial to the success of CXMS as a viable low resolution technique to determine protein structure, and to continue to enable the characterisation of the biological world where traditional techniques cannot.

Bibliography

[1] Branden, C. I.; Tooze, J. *Introduction to Protein Structure*, 2nd ed.; Garland Science, 1999.

[2] Sali, A.; Glaeser, R.; Earnest, T.; Baumeister, W. *Nature* **2003**, *422*, 216–225.

[3] Robinson, C. V.; Sali, A.; Baumeister, W. *Nature* **2007**, *450*, 973–982.

[4] Alber, F.; Dokudovskaya, S.; Veenhoff, L. M.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, A.; Karni-Schmidt, O.; Williams, R.; Chait, B. T.; Rout, M. P.; Sali, A. *Nature* **2007**, *450*, 683–694.

[5] Alberts, B. *Cell* **1998**, *92*, 291–294.

[6] Greer, J.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. *J. Med. Chem.* **1994**, *37*, 1035–1054.

[7] Wuthrich, K. *Science* **1989**, *243*, 45–50.

[8] Montelione, G. T.; Zheng, D.; Huang, Y. J.; Gunsalus, K. C.; Szyperski, T. *Nat. Stuct. Biol.* **2000**, *7*, 982–985.

[9] Varnay, I.; Truffault, V.; Djuranovic, S.; Ursinus, A.; Coles, M.; Kessler, H. *J. Am. Chem. Soc.* **2010**, *132*, 15698–15698.

[10] Pusey, M. L.; Liu, Z.-J.; Tempel, W.; Praissman, J.; Lin, D.; Wang, B.-C.; Gavira, J. A.; Ng, J. D. *Prog. Biophys. Mol. Bio.* **2005**,

[11] Bill, R. M.; Henderson, P. J. F.; Iwata, S.; Kunji, E. R. S.; Michel, H.; Neutze, R.; Newstead, S.; Poolman, B.; Tate, C. G.; Vogel, H. *Nat. Biotechnol.* **2011**, *29*, 335–340.

[12] Miyaguchi, K. *Biol. Cell.* **2014**, *106*, 323–345.

[13] Bai, X.-C.; Yan, C.; Yang, G.; Lu, P.; Ma, D.; Sun, L.; Zhou, R.; Scheres, S. H. W.; Shi, Y. *Nature* **2015**, *525*, 212–217.

[14] Bernado, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J. Am. Chem. Soc.* **2007**, *129*, 5656–5665.

[15] Mertens, H. D. T.; Svergun, D. I. *J. Stuct. Biol.* **2010**, *172*, 128–141.

[16] Kelly, S. M.; Jess, T. J.; Price, N. C. *Biochim. Biophys. Acta.* **2005**, *1751*, 119–139.

[17] Marion, J. D.; Van, D. N.; Bell, J. E.; Bell, J. K. *Anal. Biochem.* **2010**, *407*, 278–280.

[18] Sharon, M.; Robinson, C. V. *Annu. Rev. Biochem.* **2007**, *76*, 167–193.

[19] Heck, A. J. R. *Nat. Methods.* **2008**, *5*, 927–933.

[20] Pukala, T. L.; Ruotolo, B. T.; Zhou, M.; Politis, A.; Stefanescu, R.; Leary, J. A.; Robinson, C. V. *Structure* **2009**, *17*, 1235–1243.

[21] Sinz, A.; Arlt, C.; Chorev, D.; Sharon, M. *Prot. Sci.* **2015**, *24*, 1193–1209.

[22] Rossmann, M. G.; Morais, M. C.; Leiman, P. G.; Zhang, W. *Structure* **2005**, *13*, 355–362.

[23] Poliakov, A.; Duijn, E. V.; Lander, G.; Fu, C.-Y.; Johnson, J. E.; Prevelige, P. E.; Heck, A. J. R. *J. Struct. Biol.* **2007**, *157*, 371–383.

[24] Arlt, C.; Ihling, C. H.; Sinz, A. *Proteomics* **2015**, *15*, 2746–2755.

[25] Greber, B. J.; Boehringer, D.; Leitner, A.; Bieri, P.; Voigts-Hoffmann, F.; Erzberger, J. P.; Leibundgut, M.; Aebersold, R.; Ban, N. *Nature* **2013**, *505*, 515–519.

[26] Stengel, F.; Aebersold, R.; Robinson, C. V. *Mol. Cell. Proteomics.* **2012**, *11*, R111.014027.

[27] Herzog, F.; Kahraman, A.; Boehringer, D.; Mak, R.; Bracher, A.; Walzthoeni, T.; Leitner, A.; Beck, M.; Hartl, F. U.; Ban, N.; Malmstrom, L.; Aebersold, R. *Science* **2012**, *337*, 1348–1352.

[28] Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. *Nature* **2013**, *497*, 643–646.

[29] Sinz, A. *Mass. Spectrom. Rev.* **2006**, *25*, 663–682.

[30] Paramelle, D.; Miralles, G.; Subra, G.; Matrinez, J. *Proteomics* **2013**, *13*, 438–456.

[31] Maher, S.; M., J. F. P.; Taylor, S. *Rev. Mod. Phys.* **2015**, *87*, 113–130.

[32] Shukla, A. K.; Futrell, J. H. *J. Mass. Spectrom.* **2000**, *35*, 1069–1090.

[33] Kaltashov, I. A.; Eyles, S. J. *Mass Spectrometry in Structural Biology and Biophysics*, 2nd ed.; John Wiley & Sons Inc., 2012.

[34] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64–71.

[35] Fenn, J. B. *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 3871–3894.

[36] Karas, M.; Bachmann, D.; Hillenkamp, F. *Anal. Chem.* **1985**, *57*, 2935–2939.

[37] Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2301–2303.

[38] Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151–153.

[39] Cottrell, J. S. *J. Proteomics.* **2011**, *74*, 1842–1851.

[40] Boeri Erba, E.; Petosa, C. *Prot. Sci.* **2015**, *24*, 1176–1192.

[41] Verbeck, G. F.; Ruotolo, B. T.; Sawyer, H. A.; Gillig, K. J.; Russell, D. H. *J. B. T.* **2002**, *13*, 56–61.

[42] Lopez, A.; Tarrag, T.; Vilaseca, M.; Giralt, E. *New. J. Chem.* **2013**, *37*, 1283–1289.

[43] Pagel, K.; Natan, E.; Hall, Z.; Fersht, A. R.; Robinson, C. V. *Angew. Chem. Int. Edit.* **2013**, *52*, 361–365.

[44] Loo, J.; Berhane, B.; Kaddis, C.; Wooding, K.; Xie, Y.; Kaufman, S.; Chernushevich, I. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 998–1008.

[45] Smith, D. P.; Woods, L. A.; Radford, S. E.; Ashcroft, A. E. *Biophys. J.* **2011**, *101*, 1238–1247.

[46] Uetrecht, C.; Versluis, C.; Watts, N. R.; Wingfield, P. T.; Steven, A. C.; Heck, A. J. R. *Angew. Chem. Int. Edit.* **2008**, *47*, 6247–6251.

[47] Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. *Science* **2005**, *310*, 1658–1661.

[48] Konermann, L.; Pan, J.; Liu, Y.-H. *Chem. Soc. Rev.* **2011**, *40*, 1224–1234.

[49] Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R. *Prot. Sci.* **2013**, *22*, 530–544.

[50] Calabrese, A. N.; Pukala, T. L. *Aust. J. Chem.* **2013**, *66*, 749–759.

[51] Rappsilber, J. *J. Struct. Biol.* **2011**, *173*, 530–540.

[52] Clegg, C.; Hayes, D. *Eur. J. Biochem.* **1974**, *42*, 21–28.

[53] Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5802–5806.

[54] Bennett, K. L.; Kussmann, M.; Bjrk, P.; Godzwon, M.; Mikkelsen, M.; Srensen, P.; Roepstorff, P. *Prot. Sci.* **2000**, *9*, 1503–1518.

[55] Rappsilber, J.; Siniossoglou, S.; Hurt, E. C.; Mann, M. *Anal. Chem.* **2000**, *72*, 267–275.

[56] Leavell, M. D.; Novak, P.; Behrens, C. R.; Schoeniger, J. S.; Kruppa, G. H. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1604–1611.

[57] Kalkhof, S.; Sinz, A. *Ana. Bioanal. Chem.* **2008**, *392*, 305–312.

[58] Mdler, S.; Bich, C.; Touboul, D.; Zenobi, R. *J. Mass Spectrom.* **2009**, *44*, 694–706.

[59] Mclafferty, F. W.; Fridriksson, E. K.; Horn, D. M.; Lewis, M. A.; Zubarev, R. A. *Science* **1999**, *284*, 1289–1290.

[60] Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; Mclafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 806–812.

[61] Kruppa, G. H.; Schoeniger, J.; Young, M. M. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 155–162.

[62] Vandermarliere, E.; Mueller, M.; Martens, L. *Mass Spectrom. Rev.* **2013**, *32*, 453–465.

[63] Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 834–850.

[64] Chen, Z. A.; Jawhari, A.; Fischer, L.; Buchen, C.; Tahir, S.; Kamenski, T.; Rasmussen, M.; Lariviere, L.; BukowskiWills, J.; Nilges, M.; Cramer, P.; Rappsilber, J. *EMBO J.* **2010**, *29*, 717–726.

[65] Leitner, A.; Joachimiak, L. A.; Bracher, A.; Mnkemeyer, L.; Walzthoeni, T.; Chen, B.; Pechmann, S.; Holmes, S.; Cong, Y.; Ma, B.; Ludtke, S.; Chiu, W.; Hartl, F. U.; Aebersold, R.; Frydman, J. *Structure* **2012**, *20*, 814–825.

[66] Kalisman, N.; Adams, C. M.; Levitt, M. *Proc. Natl. Acad. Sci.* **2012**, *109*, 2884–2889.

[67] Liu, H.; Zhang, H.; Weisz, D. A.; Vidavsky, I.; Gross, M. L.; Pakrasi, H. B. *Proc. Natl. Acad. Sci. USA.* **2014**, *111*, 4638–4643.

[68] Bojja, R. S.; Andrake, M. D.; Weigand, S.; Merkel, G.; Yarychkivska, O.; Henderson, A.; Kummerling, M.; Skalka, A. M. *J. Biol. Chem.* **2011**, *286*, 17047–17059.

[69] Sinz, A. *Expert Rev. Proteomics.* **2014**, *11*, 733–743.

[70] Back, J. W.; de Jong, L.; Muijsers, A. O.; de Koster, C. G. *J. Mol. Biol.* **2003**, *331*, 303–313.

[71] Kluger, R.; Alagic, A. *Bioorg. Chem.* **2004**, *32*, 451–472.

[72] Nguyen-Huynh, N.-T.; Sharov, G.; Potel, C.; Fichter, P.; Trowitzsch, S.; Berger, I.; Lamour, V.; Schultz, P.; Potier, N.; Leize-Wagner, E. *Prot. Sci.* **2015**, *24*, 1232–1246.

[73] Guerrero, C.; Tagwerker, C.; Kaiser, P.; Huang, L. *Mol. Cell. Proteomics.* **2006**, *5*, 366–378.

[74] Navare, A. T.; Chavez, J. D.; Zheng, C.; Weisbrod, C. R.; Eng, J. K.; Siehnel, R.; Singh, P. K.; Manoil, C.; Bruce, J. E. *Structure* **2015**, *23*, 762–773.

[75] Smart, S. K.; Mackintosh, S. G.; Edmondson, R. D.; Taverna, S. D.; Tackett, A. J. *Protein. Sci.* **2009**, *18*, 1987–1997.

[76] Zorn, M.; Ihling, C. H.; Sinz, A.; Golbik, R.; Sawers, R. G. *J. Proteome Res.* **2014**, *13*, 5524–5535.

[77] Wells, M.; Tidow, H.; Rutherford, T. J.; Markwick, P.; Jensen, M. R.; Mylonas, E.; Svergun, D. I.; Blackledge, M.; Fersht, A. R. *Proc. Natl. Acad. Sci. USA.* **2008**, *105*, 5762–5767.

[78] Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R. *Mol. Cell. Proteomics.* **2010**, *9*, 1634–1649.

[79] Gao, Q.; Xue, S.; Doneanu, C. E.; Shaffer, S. A.; Goodlett, D. R.; Nelson, S. D. *Anal. Chem.* **2006**, *78*, 2145–2149.

[80] Gotze, M.; Pettelkau, J.; Schaks, S.; Bosse, K.; Ihling, C.; Krauth, F.; Fritzsche, R.; Kuhn, U.; Sinz, A. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 76–87.

[81] Gotze, M.; Pettelkau, J.; Fritzsche, R.; Ihling, C.; Schafer, M.; Sinz, A. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 83–97.

[82] Liu, F.; Rijkers, D. T.; Post, H.; Heck, A. J. *Nat. Methods.* **2015**, *12*, 1179–1184.

[83] Du, X.; Chowdhury, S. M.; Manes, N. P.; Wu, S.; Mayer, M. U.; Adkins, J. N.; Anderson, G. A.; Smith, R. D. *J Proteome Res.* **2011**, *10*, 923–931.

[84] Xu, H.; Zhang, L.; Freitas, M. A. *J. Proteome Res.* **2008**, *7*, 138–144.

[85] Taverner, T.; Hall, N. E.; O'Hair, R. A. J.; Simpson, R. J. *J. Biol. Chem.* **2002**, *277*, 46487–46492.

[86] de Koning, L. J.; Kasper, P. T.; W., B. J.; Nessen, M. A.; Vanrobaeys, F.; Van Beeumen, J.; Gherardi, E.; do Koster, C. G.; de Jong, L. *FEBS. J.* **2006**, *273*, 281–291.

[87] Fischer, L.; Chen, Z. A.; Rappsilber, J. *J. Proteomics* **2013**, *88*, 120–128.

[88] Walzthoeni, T.; Joachimiak, L. A.; Rosenberger, G.; Rost, H. L.; Malmstrom, L.; Leitner, A.; Frydman, J.; Aebersold, R. *Nat. Methods.* **2015**, *12*, 1185–1190.

[89] Muller, M. Q.; Dreiocker, F.; Ihling, C. H.; Schafer, M.; Sinz, A. *Anal Chem.* **2010**, *82*, 6958–6968.

[90] Muller, D. R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O. *Anal. Chem.* **2001**, *73*, 1927–1934.

[91] Kao, A.; Chiu, C.-L.; Vellucci, D.; Yang, Y.; Patel, V. R.; Guan, S.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L. *Mol. Cell. Proteomics.* **2011**, *10*, M110.002212.

[92] Lomant, A. J.; Fairbanks, G. *J. Chem. Biol.* **1976**, *104*, 243–261.

[93] Calabrese, A. N.; Good, N.; Wang, T.; He, J.; Bowie, J.; Pukala, T. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1364–1375.

[94] Liu, F.; Goshe, M. B. *Anal. Chem.* **2010**, *82*, 6215–6223.

[95] Soderblom, E. J.; Goshe, M. B. *Anal. Chem.* **2006**, *78*, 8059–8068.

[96] Soderblom, E. J.; Bobay, B. G.; Cavanagh, J.; B., G. M. *Rapid Common. Mass Spectrom.* **2007**, *21*, 3395–3408.

[97] Muller, M. Q.; Dreiocker, F.; Ihling, C. H.; Schfer, M.; Sinz, A. *J. Mass Spectrom.* **2010**, *45*, 880–891.

[98] Ilver, D.; Arnqvist, A.; gren, J.; Frick, I.-M.; Kersulyte, D.; Incecik, E. T.; Berg, D. E.; Covacci, A.; Engstrand, L.; Born, T. *Science* **1998**, *279*, 373–377.

[99] Kang, S.; Mou, L.; Lanman, J.; Velu, S.; Brouillette, W. J.; Prevelige, P. E. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 1719–1726.

[100] Sohn, C. H.; Agnew, H. D.; Lee, J. E.; Sweredoski, M. J.; Graham, R. L. J.; Smith, G. T.; Hess, S.; Czerwieniec, G.; Loo, J. A.; Heath, J. R.; Deshaies, R. J.; Beauchamp, J. L. *Anal. Chem.* **2012**, *84*, 2662–2669.

[101] Bobofchak, K. M.; Tarasov, E.; Olsen, K. W. *Biochim. Biophys. Acta.* **2008**, *1784*, 1410–1414.

[102] Paoli, B.; Pellarin, R.; Caflisch, A. *J. Phys. Chem. B.* **2010**, *114*, 2023–2027.

[103] Huisgen, R. *Angew. Chem. Int. Edit.* **1963**, *2*, 565–598.

[104] Vellucci, D.; Kao, A.; Kaake, R. M.; Rychnovsky, S. D.; Huang, L. *J. Am. Soc. Mass Spectrom* **2010**, *21*, 1432–1445.

[105] Nessen, M. A.; Kramer, G.; Back, J.; Baskin, J. M.; Smeenk, L. E.; de Koning, L. J.; van Maarseveen, J. H.; de Jong, L.; Bertozzi, C. R.; Hiemstra, H.; de Koster, C. G. *J. Proteome. Res.* **2009**, *8*, 3702–3711.

[106] Buncherd, H.; Nessen, M. A.; Nouse, N.; Stelder, S. K.; Roseboom, W.; Dekker, H. L.; Arents, J. C.; Smeenk, L. E.; Wanner, M. J.; van Maarseveen, J. H.; Yang, X.; Lewis, P. J.; de Koning, L. J.; de Koster, C. G.; de Jong, L. . *Proteomics* **2012**, *75*, 2205–2215.

[107] Chowdhury, S. M.; Du, X.; Tolic, N.; Wu, S.; Moore, R. J.; Mayer, M. U.; Smith, R. D.; Adkins, J. N. *Anal. Chem* **2009**, *81*, 5524–5532.

[108] Yan, F.; Che, F.-Y.; Rykunov, D.; Nieves, E.; Fiser, A.; Weiss, L. M.; Hogue Angeletti, R. *Anal. Chem* **2009**, *81*, 7149–7159.

[109] Wine, R.; Dial, J.; Tomer, K.; Borchers, C. *Anal. Chem.* **2002**, *74*, 1939–1945.

[110] Sinz, A.; Wang, K. *Anal. Bio. Chem.* **2004**, *331*, 27–32.

[111] Fujii, N.; Jacobsen, R. B.; Wood, N. L.; Schoeniger, J. S.; Guy, R. K. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 427–429.

[112] Hurst, G. B.; Lankford, T. K.; Kennel, S. J. *J. Am. Soc. Mass Spectrom* **2004**, *15*, 832–839.

[113] Trester-Zedlitz, M.; Kamada, K.; Burley, S. K.; Fenyo, D.; Chait, B. T.; Muir, T. W. *J. Am. Chem. Soc.* **2003**, *125*, 2416–2425.

[114] Petrochenko, E. V.; Serpa, J. J.; Borchers, C. H. *Mol. Cell. Proteomics* **2011**, *10*, M110 001420.

[115] Petrochenko, E. V.; Xiao, K.; Cable, J.; Chen, Y.; Dokholyan, N. V.; Borchers, C. H. *Mol. Cell Proteomics.* **2009**, *8*, 273–286.

[116] Petrochenko, E. V.; Olkhovik, V. K.; Borchers, C. H. *Mol. Cell. Proteomics.* **2005**, *4*, 1167–1179.

[117] Herbert, C. G.; Johnstone, R. A. W. *Mass Spectrometry Basics*, 1st ed.; CRC Press LLC, 2003.

[118] Lane, C. *Cell. Mol. Life Sci.* **2005**, *62*, 848–869.

[119] 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. *Clin. Biochem Rev.* **2003**, *24*, 3–12.

[120] Smith, R. D.; Light-wahl, K. J.; Winger, B. E.; Loo, J. A. *Org. Mass Spectrom.* **1992**, *27*, 811–821.

[121] Wilm, M.; Mann, M. *Int. J. Mass Spectrom. Ion. Proc.* **1994**, *136*, 167–180.

[122] Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1–8.

[123] Karas, M.; Bahr, U.; Dulcks, T. *Fresen. J. Anal. Chem.* **2000**, *366*, 669–676.

[124] Paul, W.; Steinwedel, H. *Z. Naturforsch* **1953**, *8*, 448–450.

[125] Stephens, W. E. *Phys. Rev.* **1946**, *69*, 691.

[126] Mamyrin, B. A. *Int. J. Mass Spectrom.* **2001**, *206*, 251–266.

[127] Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11*, 601.

[128] McCloskey, J. A. *Mass Spectrometry*, 1st ed.; Elsevier Science Publishing Co., 1990; pp 886–887.

[129] Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G. *Methods* **2005**, *35*, 211–222.

[130] Bowie, J. H.; Brinkworth, C.; Dua, S. *Mass Spectrom. Rev.* **2002**, *21*, 87–107.

[131] Boontheung, P.; Brinkworth, C. S.; Bowie, J. H.; Baudinette, R. V. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 287–292.

[132] Bilusich, D.; Bowie, J. H. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 619–628.

[133] Zhang, M.; Kaltashov, I. A. *Anal. Chem.* **2006**, *78*, 4820–4829.

[134] Hunter, M. J. M.; Ludwig, M. L. *J. Am. Chem. Soc.* **1962**, *84*, 3491–3504.

[135] Wallace, C. J.; Harris, D. E. *J. Biol. Chem.* **1984**, *217*, 589–594.

[136] Kovalenko, O. V.; Yang, X. H.; Hemler, M. E. *Mol. Cell Proteomics.* **2007**, *6*, 1855–1867.

[137] Kim, Y.; Ho, S. O.; Gassman, N. R.; Korlann, Y.; Landorf, E. V.; Collart, F. R.; Weiss, S. *Bioconjug. Chem.* **2008**, *19*, 786–791.

[138] Blencowe, A.; Hayes, W. *Soft Matter* **2005**, *1*, 178–205.

[139] Korshunova, G.; Sumbatyan, N.; Topin, A.; Mtchedlidze, M. *Mol. Biol.* **2000**, *34*, 823–839.

[140] Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Edit.* **2001**, *40*, 2004–2021.

[141] Prescher, J. A.; Bertozzi, C. R. *Nat. Chem. Biol.* **2005**, *1*, 13–21.

[142] Tieu, W. Design and synthesis of reaction intermediate derivatives as biotin protein ligase inhibitors. 2012.

[143] El-Faham, A.; Albericio, F. *Chem. Rev.* **2011**, *111*, 6557–6602.

[144] Valeur, E.; Bradley, M. *Chem. Soc. Rev.* **2009**, *38*, 606–631.

[145] Pehere, A. D.; Sumby, C. J.; Abell, A. D. *Org. Biomol. Chem.* **2013**, *11*, 425–429.

[146] Isidro-Llobet, A.; Alvarez, M.; Albericio, F. *Chem. Rev.* **2009**, *109*, 2455–2504.

[147] Gibson, S. E.; Lecci, C. *Angew. Chem. Int. Edit.* **2006**, *45*, 1364–1377.

[148] Barrett, G. C. *Chemistry and Biochemistry of the Amino Acids*, 1st ed.; Chapman and Hall, 1985.

[149] Barman, A. K.; Gour, N.; Verma, S. *ARKIVOC* **2013**, *2013*, 82–99.

[150] Han, S.-Y.; Kim, Y.-H. *Tetrahedron* **2004**, *60*, 2447–2467.

[151] Prasad, P.; Molla, M. R.; Cui, W.; Canakci, M.; Osborne, B.; Mager, J.; Thayumanavan, S. *Biomacromolecules* **2015**, *16*, 3491–3498.

[152] Heng, S.; Ngyuen, M.-C.; Kostecki, R.; Monro, T. M.; Abell, A. D. *RSC Advances* **2013**, *3*, 8308–8317.

[153] Adamczyk, M.; Chen, Y.-Y.; Fishpaugh, J. R.; Mattingly, P. G.; Pan, Y.; Shreder, K.; Yu, Z. *Bioconjugate Chem.* **2000**, *11*, 714–724.

[154] Khorana, H. G. *Chem. Rev.* **1953**, *53*, 145–166.

[155] Sakakibara, S.; Inukai, N. *B. Chem. Soc. Jpn.* **1965**, *38*, 1979–1984.

[156] Calabrese, A. N.; Wang, T.; Bowie, J. H.; Pukala, T. L. *Rapid Commun. Mass Spectrom.* **2013**, *27*, 238–248.

[157] Tang, X.; Bruce, J. E. *Mol. BioSyst.* **2010**, *6*, 939–947.

[158] Beckett, D. *J. Nutr.* **2009**, *139*, 167–170.

[159] Soares Da Costa, T. P.; Tieu, W.; Yap, P. N. R., M. Y.; Polyak, S. W.; Sejer-Pedersen, D.; Morona, R.; Turnidge, J. D.; Wallace, J. C.; Wilce, M. C. J.; Booker, G. W.; Abell, A. D. *J. Biol. Chem.* **2012**, *287*, 17823–17832.

[160] Paparella, A. S.; Soares da Costa, T. P.; Yap, M. Y.; Tieu, W.; Wilce, M. C. J.; Booker, G. W.; Abell, A. D.; Polyak, S. W. *Curr. Top. Med. Chem.* **2014**, *14*, 4–20.

[161] Gupta, V.; Gupta, R.; Khare, G.; Salunke, D.; Surolia, A.; Tyagi, A. *PLoS One* **2010**, *5*, e9222.

[162] Bagautdinov, B.; C., K.; Sugahara, M.; Kunishima, N. *J. Mol. Biol.* **2005**, *353*, 322–333.

[163] Eisenstein, E.; Beckett, D. *Biochemistry* **1999**, *38*, 13077–13084.

[164] Pendini, N. R.; Yap, M. Y.; Polyak, S. W.; Cowieson, N. P.; Abell, A.; Booker, G. W.; Wallace, J. C.; Wilce, J. A.; Wilce, M. C. J. *Prot. Sci.* **2013**, *22*, 762–773.

[165] Rodionov, D. A.; Mironov, A. A.; Gelfand, M. S. *Genome Res.* **2002**, *12*, 1507–1516.

[166] Soares da Costa, T. P.; Yap, M. Y.; Perugini, M. A.; Wallace, J. C.; Abell, A. D.; Wilce, M. C. J.; Polyak, S. W.; Booker, G. W. *Biochemistry* **2014**, *91*, 110–120.

[167] Makepeace, K. A. T.; Serpa, J. J.; Petrotchenko, E. V.; Borchers, C. H. *Methods* **2015**, *89*, 74–78.

[168] Bilusich, D.; Bowie, J. H. *Mass Spectrom. Rev.* **2009**, *28*, 20–34.

[169] Clayden, J.; Greeves, N.; Warren, S. *Organic Chemistry*, 2nd ed.; Oxford University Press, 2012.

[170] Blanco, B.; Sedes, A.; Pen, A.; Lamb, H.; Hawkins, A. R.; Castedo, L.; Gonzlez-Bello, C. *Org. Biomol. Chem.* **2012**, *10*, 3662–3676.

[171] Florence, T. M. *Biochem. J.* **1980**, *189*, 507–520.

[172] Galande, A. K.; Trent, J. O.; Spatola, A. F. *Biopolymers* **2003**, *71*, 534–551.

[173] Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E. *Anal. Chem.* **2005**, *77*, 311–318.

[174] Henkel, B.; Zeng, W.; Bayer, E. *Tetrahedron Lett.* **1997**, *38*, 3511–3512.

[175] Carpino, L. A.; Han, G. Y. *J. Org. Chem.* **1972**, *37*, 3404–3409.

[176] Carpino, L. A. *Accounts Chem. Res.* **1987**, *20*, 401–407.

[177] Bodanszky, M.; Bodanszky, A. *The practise of peptide synthesis*, 2nd ed.; Springer-Verlag, 1994.

[178] Pennington, M. W.; Dunn, B. M. *Peptide Synthesis Protocols*, 1st ed.; Humana Press, 1995; Vol. 35; Chapter 2.

[179] Maruyama, K.; Nagasawa, H.; Suzuki, A. *Peptides* **1999**, *20*, 881–884.

[180] Bradford, M. M. *Anal. Biochem.* **1976**, *72*, 248–254.

Appendix A

NMR Spectra

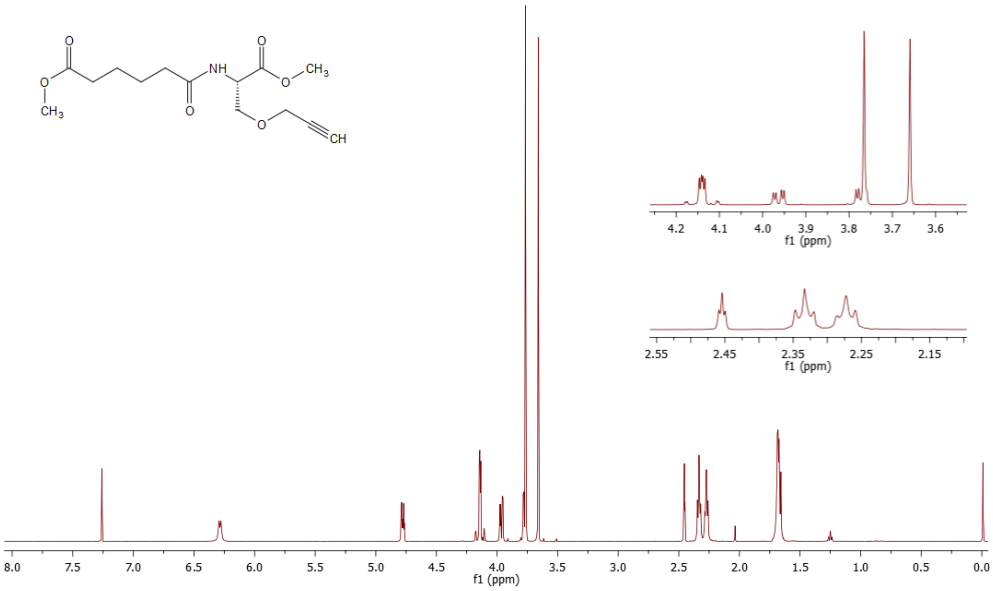


Figure A.1: Hydrogen NMR spectrum of diester **5** in $CDCl_3$. Expansion of methyl ester (top) and spacer arm triplet regions (bottom) shown in inset.

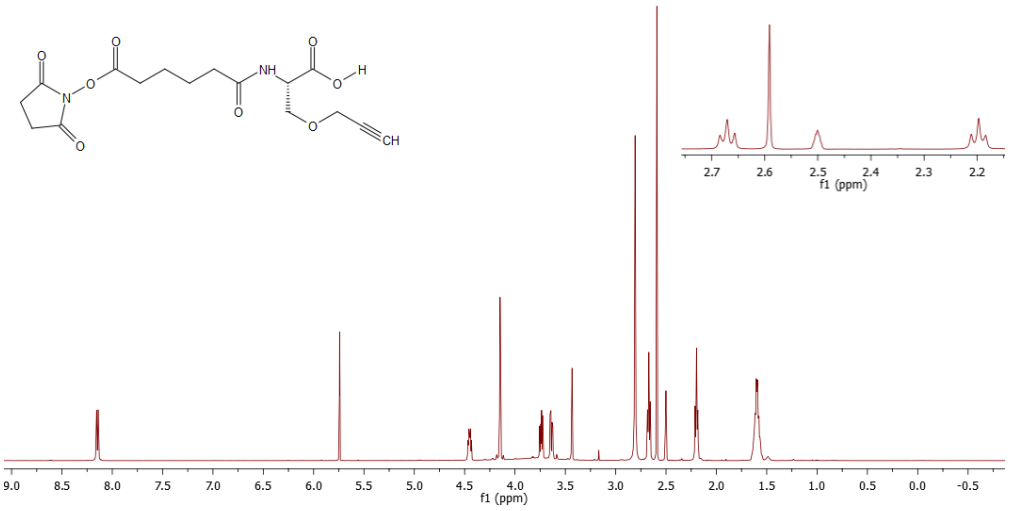


Figure A.2: Hydrogen NMR spectrum of monoester **7** in DMSO-d_6. Expansion of spacer arm triplet region shown in inset.

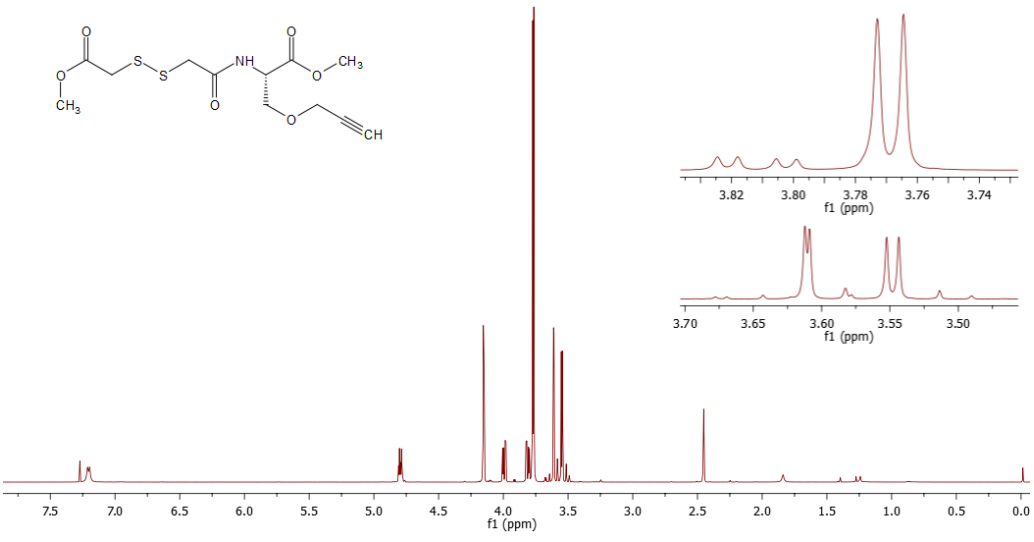


Figure A.3: Hydrogen NMR spectrum of diester **10**. Expansion of methyl ester (top) and spacer arm CH_2 regions (bottom) shown in inset.