MASS SPECTROMETRY-BASED STRUCTURAL INSIGHTS INTO PROTEIN ASSEMBLIES

A thesis submitted for the Degree of Doctor of Philosophy

By

Henry Sanders B. Sc. (Hons)

THE UNIVERSITY of ADELAIDE

December 2020

Table of Contents

i. Acknowledgements 1

ii. Declaration 2

iii. Abstract 3

Chapter 1: Introduction 6

1.1 Protein structure 6

1.2 Protein misfolding 8

1.3 Protein structure determination 11

1.3.1 Methods for protein structure determination 11

1.3.2 Mass spectrometry 11

1.3.3 Native mass spectrometry 15

1.3.4 Ion-mobility mass spectrometry 16

1.3.5 Cross-linking mass spectrometry 18

1.4 Aims 21

1.5 References 22

Chapter 2: The molecular chaperone β-casein prevents amorphous and fibrillar aggregation of α-lactalbumin by stabilisation of dynamic disorder 31

Chapter 3: Structural and mechanistic insights into the influence of phospholipids in amyloid fibril formation and inhibitor efficacy 56

Structural and mechanistic insights into the influence of phospholipids in amyloid fibril formation and inhibitor efficacy 58

3.1 Abstract 59

3.2 Introduction 61

3.3 Experimental Procedures 63

3.3.1 Preparation of Aβ40, αS and LUVs 64

3.3.2 Dynamic light scattering 64

3.3.3 Thioflavin T assays 64

3.3.4 Transmission electron microscopy 65
3.3.5 Immunoblotting 65
3.3.6 Ion-mobility mass spectrometry 66
3.4 Results 66
3.4.1 LUVs increase the rate of Aβ40 fibrillization and reduce EGCG-resveratrol efficacy 66
3.4.2 LUVs increase the rate of αS fibril elongation and reduce EGCG-resveratrol efficacy 69
3.4.3 LUVs and inhibitors differentially influence the distribution of protein aggregation states 71
3.4.4 IM-MS reveals structural changes induced by EGCG complexes with Aβ40 and αS monomers 72
3.5 Discussion 77
3.6 Acknowledgements 81
3.7 Competing Interests 81
3.8 Author Contributions 81
Chapter 4: Development of a modular synthetic protocol for protein crosslinkers 96
Development of a modular synthetic route for protein chemical cross-linking reagents 97
4.1 Abstract 99
4.2 Introduction 100
4.3 Results and Discussion 102
4.3.1 Development of a modular synthetic protocol 102
4.3.2 Cross-linking and fragmentation characterisation 105
4.3.3 Post cross-link derivatisation of the modular linker 106
4.4 Conclusion 109
4.5 Materials and methods 110
4.5.1 Materials and reagents 110

4.5.2 Synthesis of cross-linkers 110
4.5.3 Synthesis of compound **10** 110
4.5.4 Synthesis of compound **12** 111
4.5.5 General modular synthetic protocol 112
4.5.6 Synthesis of compound **4**, **6** and **7** 113
4.5.7 Cross-linking of AcAAKA 113
4.5.8 Cross-linking of hen egg lysozyme 113
4.5.9 Copper click reaction 113
4.5.10 SDS–polyacrylamide gel electrophoresis 113
4.5.11 Matrix-assisted laser desorption ionization mass spectrometry 114
4.6 Acknowledgements 114
4.7 Competing Interests 114
4.8 Author Contributions 114
4.9 References 115
4.10 Supplementary Information 119
4.10.1 Characterisation Data 122
Chapter 5: Towards the development of a modular synthetic route for protein chemical cross-linking reagents 129
5.1 Introduction 129
5.2 Development of a modular synthetic protocol 129
5.2.1 Linker design 129
5.3 Development of the synthetic route 131
5.3.1 Synthesis of succinimide linker 131
5.3.2 Synthesis of linker **1** 132
5.3.3 Synthesis of linker **2** 133
5.4 Incorporation of a CID-cleavable spacer arm 134

5.4.1 Linker **3** 135
5.4.2 Linker **4/4a** 137
5.5 Conclusions and future directions 141
5.6 Materials and methods 143
5.6.1 Materials and reagents 143
5.6.2 Cross-linking of AcAAKA 143
5.6.3 Tandem mass spectrometry 143
5.6.4 High-resolution mass spectrometry 143
5.7 Synthetic methods 144
5.7.1 Compound **12** 144
5.7.2 Succinimide linker 145
5.7.3 Linker **1** 147
5.7.4 Linker **2** 149
5.7.5 Linker **3** 151
5.7.6 Linker **4a** 153
5.7.7 Linker **4** 155
5.8 References 158
Chapter 6: Summary, Conclusion and Future Directions 160
6.1 β-CN stabilises both amorphous and fibrillar α-LA 160
6.2 Lipids reduce fibril inhibitor efficacy of Aβ40 and αS 161
6.3 Development of a modular protocol and linker library to expand XL-MS opportunities 162
6.4 References 164
Chapter 7: Statements of authorship 168

i. Acknowledgements

Firstly, I would like to thank my supervisor, Tara Pukala. She is one of the most patient people I know and this work would never have been possible without her unending effort. Her edits and expertise went far beyond the average supervisor and I thank her greatly for all her help.

Secondly, I want to thank the entire lab group for their support. In particular, BJ, who taught me so much towards the end of my PhD and gave me insight and experience into what the near future of research holds.

Without my friends and family this never would have been possible. My parents, Craig and Louise, whose unlimited support got me over the line. They are amazing people and their pride in what I do, despite never really understanding any of it, has been appreciated more than anything.

Finally, thanks to Sarah, who knew exactly how to manage my grumpiness and supported me no matter how much pressure I was under.

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

I acknowledge the support I have received for my research through the provision of an a Faculty of Sciences Divisional Scholarship from the University of Adelaide.

iii. Abstract

Complications with protein homeostasis, genetic mutations or post-translational modifications can lead to deficits in the correct folding, and therefore function, of proteins. These misfolding events are often associated with protein aggregation via amorphous and fibrillar pathways. Protein fibrils have long been associated with a range of neurodegenerative diseases, such as Alzheimer's and Parkinson's diseases, through the formation of fibrillar plaques that are deposited in diseased regions of the brain. Current understanding of misfolding diseases point to low molecular weight oligomers en route to fibril formation, instead of fibrils themselves, as the pathological species. More specifically, the relationship between lipid membranes and misfolded oligomers plays a role fibril kinetics and may be a vital component of disease aetiology via membrane disruption or leakage by oligomeric pore formation.

To date, there are no effective therapies associated with neurodegenerative misfolding diseases, despite great advances in our understandings. To treat diseases such as these we require a greater understand of the mechanisms at play. Traditional structure determination techniques such as X-Ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are often not amenable to the study of misfolding proteins given their transient nature, therefore new structure determination techniques must be exploited. This thesis utilised and developed a combined biophysical approach, with a focus on mass spectrometry (MS), to better understand protein misfolding and aggregation.

Protein quality control systems like chaperones combat misfolding but these can fail, leading to disease. The mechanisms of molecular chaperones depend in part on protein morphology and can interact differently with the same protein if it undergoes different structural stresses. In this work, ion-mobility mass spectrometry (IM-MS) was utilised to understand the protective mechanisms of a molecular chaperone, β-casein (β-CN), against the aggregation of α-lactalbumin (α-LA). α-LA is capable of both fibrillar and amorphous aggregation, both of which were inhibited by β-CN at substoichiometric concentrations. During amorphous aggregation, analytical size exclusion revealed α-LA formed stable, high molecular weight complex with β-CN. IM-MS coupled with collision-induced dissociation (CIU) described transient structural interactions wherein β-CN stabilised α-LA, as well as reduced conformational heterogeneity while in the gas-phase. Overall, this data demonstrates the

practicality of biophysical techniques, particularly IM-MS with CIU, to explore the interactions between misfolding proteins and chaperones.

α-LA is a powerful model for protein misfolding, but it is not associated with disease. Amyloid beta 40 (Aβ40) and α-synuclein (α-syn) proteins are implicated in Alzheimer's and Parkinsons's diseases, respectively. Here these intrinsically disordered proteins were examined to better understand their interaction with the cellular membrane and its effects on fibril inhibition by small molecules, EGCG and resveratrol. Large unilamellar vesicles (LUVs) were used to model the cellular membrane, and it was revealed that they both increased the rate of fibrillar aggregation and decreased the effectiveness of the known fibril inhibitors. Using oligomer specific immunoblotting and IM-MS it was shown that EGCG and resveratrol work through differing mechanisms, wherein resveratrol targeted the elongation phase of aggregation while EGCG targeted the nucleation phase. IM-MS showed EGCG to preference binding to more compact forms of monomeric Aβ40 and α-syn, while LUVs influenced conformational changes indicative of nucleation. These observations combined to form a more detailed mechanistic insight into the protein-lipid-inhibitor relationship and highlighted that current approaches toward drug design may be misguided if the effects of lipid membranes are not properly considered.

The transient nature of misfolding proteins can make structural determination complicated using traditional high-resolution structure determination techniques like X-ray crystallography and NMR spectroscopy. While useful, native- and IM-MS cannot offer specific detail in terms of tertiary structure and subunit architecture. Cross-linking MS (XL-MS) has emerged as a valuable complementary tool for uncovering structural information from proteins. While cross-linkers are available commercially, the rapid emergence of new analytical strategies means specific features or combinations of features are not readily available, and the highly variable nature of proteins means the optimum cross-linking regent may be system specific. To remedy this and enable to a wide range of cross-linker chemistries, a modular synthetic protocol was designed to allow for the incorporation custom reactive groups, spacer-arms, enrichment motifs and other features. The protocol was used to develop a small library containing 8 unique cross-linkers, each containing different features. The reagents were optimised using model peptides to monitor efficiency, labile spacer-arm fragmentation and derivatisation with enrichment and dye motifs. The wide applicability and

straightforward synthesis of the modular protocol aims to remedy the lacking diversity in available cross-linkers for broad uptake in XL-MS workflows.

Overall, the work in this thesis aims to highlight and expand MS-based techniques for protein structure determination. The presented examples showcase that investigating transient proteins require adaptable techniques like native MS, IM-MS and XL-MS. The combination of these and other biophysical techniques provide a toolset capable of handling a great range of protein structures deemed too complicated for traditional structure determination techniques.

Chapter 1: Introduction

1.1 Protein structure

Fundamental to life on Earth is the function of proteins which carry out essential biological tasks such as catalysis, transport and signal transmission. Proteins consist of amino acids transcribed from the genetic information encoded in DNA via messenger RNA. Most proteins, following transcription, begin the process of folding to form an intricately defined three-dimensional structure which allow them to perform a specific biological function. The sequence of amino acids in a chain give it specific molecular properties which dictate its defined structure, and therefore function, as a biomolecular machine. The importance of the protein structure-function relationship was realised in the 1950s when Anfinsen demonstrated the refolding of a denatured RNase enzyme, restoring its catalytic activity.[1]

The way in which proteins consistently fold into their native states is still not fully understood, although a number of models have been proposed to describe formation of protein structure. Even small polypeptides would take an astronomical amount of time to fold through the random sampling of conformations, according to Levinthal's paradox,[2] given the enormous range of potential structures. Despite this, the protein folding process is observed on a millisecond timescale.[2,3] The process of structure formation is thought to be accelerated by sampling a series of increasingly thermodynamically favourable intermediates throughout the process of folding, as described by the nucleation-condensation model.[4] The native state of a functional protein is usually thought to be the state in which its intramolecular contacts are at their lowest free energy. We can therefore think of protein folding as the polypeptide chain sampling structural conformations upon a three dimensional energy landscape in which the folding pathway is guided towards lower energy states (**Fig. 1.1**).[5] As a protein samples lower energy conformations it has fewer conformational options as it approaches its native state. Hence, although the initial structural freedom is immense, through the nucleation-condensation model a protein can quickly take on a lower energy state before more intricate folding events take place.[4]

The nucleation-condensation model proposes that, in the initial folding steps, a protein undergoes nucleation wherein large sections of the peptide chain form a native-like nucleus, which is followed by the condensation of the remaining unstructured protein. Smaller

proteins rely on a rate determining nucleation step that depends on key residues forming interactions that allow the remaining structure to form around it. Larger proteins can nucleate multiple hydrophobic residue-rich regions which form a hydrophobic core, avoiding solvent and allowing the remaining chain to condense and take on the native structure.[6,7]

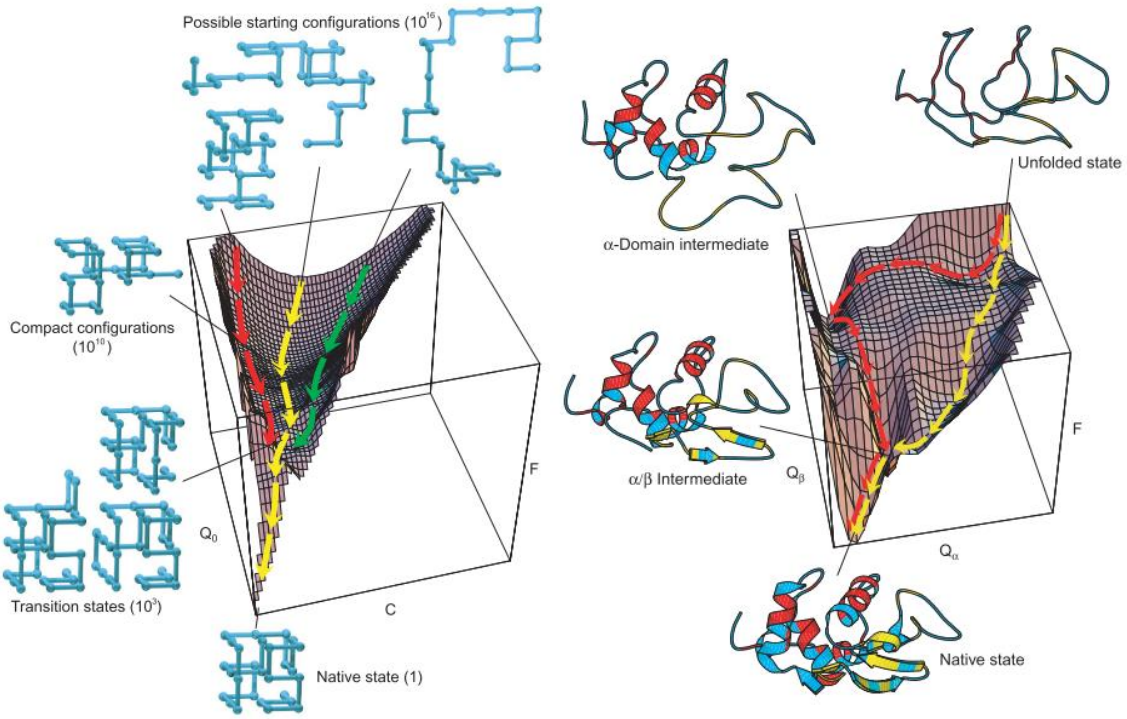


Figure 1.1. The free-energy landscape showing collapse of a polypeptide into its lowest energy state while conformational freedom is reduced (left). During folding, proteins can sit in intermediate states of low energy before reaching the native structure (right). Adapted with permission from Dinner, *et al.* (2000).[5]

Aside from this process of folding, proteins can undergo further modifications before they are utilised in the cell, for example, alternative RNA splicing, the formation of quaternary structure through complexation of multiple proteins and through post-translational modifcations (PTMs) such as acetylation or phosphorylation.[8] These processes help to account for the incredible versatility of the proteome attributed to such a comparitively small number protein-encoding genes within the human genome (20,000 – 25,000).[9]

1.2 Protein misfolding

Due to the complex nature of protein folding, the process is susceptible to malfunction, and proteins that are 'misfolded' will often be inactive or potentially toxic to the cell. Certain proteins known as 'misfolding proteins' are more prone to this than others. During the folding process proteins sample intermediate conformations within the energy landscape, sometimes straying from the correct folding process and going into an off-pathway state.[10] A disruption of cellular homeostasis,[11,12] posttranslational modifications (PTMs)[13] or genetic mutation[14] can lower the energy of these normally transient, unfolded intermediates, or in the case of proteolytic digestion, form entirely new intrinsically disordered proteins.[15]

At high enough concentrations, misfolded proteins can form aggregates. This can be due to changes in the cellular environment that cause the partial or full unfolding of proteins exposing their normally buried hydrophobic regions, their intrinsic propensity for alternative secondary structure formation or changes in the net protein charge, leading to intermolecular association.[16–19] These conditions can lead to two outcomes; the formation of amorphous aggregates or highly-ordered amyloid fibrils. Amorphous aggregates are largely considered non-toxic, with nuclear cataracts regarded as the best known disease association, and are formed by random hydrophobic associations of insoluble protein.[20] Fibrils are formed by the stacking of a conserved, repeating β-sheet structure.[21]

Amyloid fibrils consist of several 'protofibrils' or strand-like protein fibres consisting of repeating monomeric protein units stacked together perpendicularly against the fibril axis, held together through strong β-sheet interactions (**Fig. 1.2**). This process is typically labelled as nucleation-dependent, where misfolded proteins associate to form small oligomers beginning the 'lag phase'. The lag phase ends with the creation of a nucleus, following a shift toward an increase in β-sheet content,[22] which is capable of rapidly recruiting monomeric units to incorporate into the protofibril. The protofibril extends during an 'elongation phase' and multiple protofibrils combine to form a mature amyloid fibril, reaching equilibrium with the monomers in solution at the beginning of the 'plateau phase' (**Fig. 1.2**).

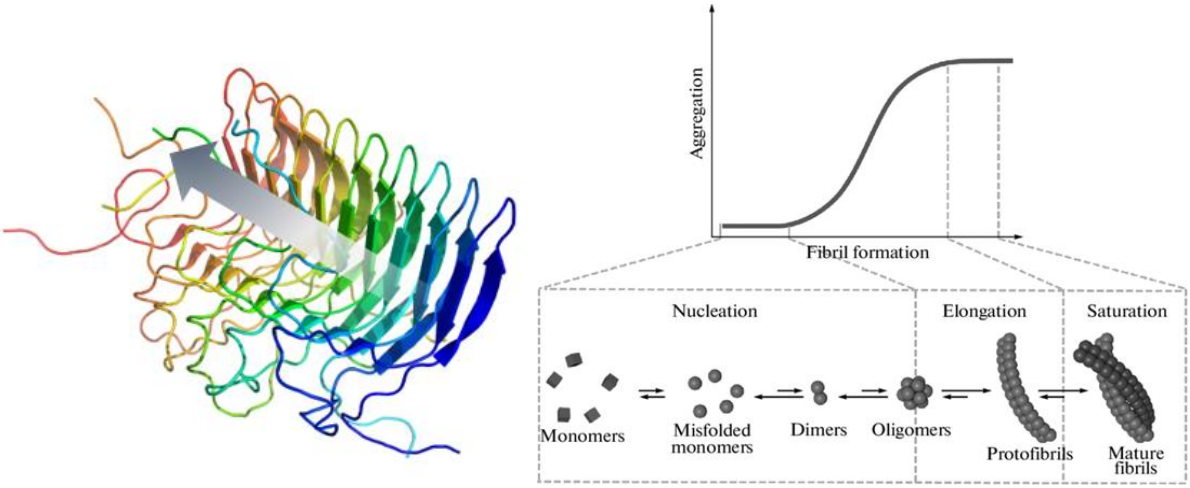


Figure 1.2. Pathway of amyloid aggregation. The stacking of amyloid proteins, held together by beta-sheet interactions perpendicular to the fibril axis (indicated by the arrow) (left). Adapted with permission from Wasmer, *et al.* (2008).[23] During the nucleation phase, misfolded proteins form low-molecular weight oligomers until a seed capable of fibrillation forms, beginning the elongation phase. Protofibrils then accumulate to form mature fibrils (right). Adapted with permission from Kulikova, *et al*. (2015)[24]

Amyloid fibril plaques are the pathological indicator for a wide variety of debilitating diseases known as misfolding diseases, including Alzheimer's, Parkinson's and Huntington's diseases.[25] While the exact mode of disease aetiology is not fully understood, the presence of amyloid fibril build up in the diseased cells of sufferers led to the formulation of the amyloid hypothesis,[26] implicating fibrils as the cause of disease. However, current research has shown that while plaques may be an indicator of misfolding disease, they may not actually be toxic,[27,28] whereas the causative agents instead are likely the low-molecular weight oligomeric species *en route* to fibril formation.[29–32]

The mechanism by which the low-molecular weight oligomers exert their toxic effects is still not well defined, although it has been suggested that the process may involve the interaction between cellular membranes and amino acid residues of the protein that are normally buried within the globular structure. Interactions between fibrillogenic proteins and lipid membranes, as well as other surfaces, have been shown to accelerate fibril formation rates and affect mature fibril morphology indicating some form of interaction is taking place[33–37] but the process, and whether it is varied or a unified mechanism is not well understood.[38–41]

Current opinion also suggests that oligomers may insert themselves within the bilayer or bind to its surface, leading to membrane thinning and membrane leakage, or form annular aggregates that allow for unchecked ion movement, potentially leading to apoptosis (**Fig. 1.3**).[36] This impairment of membrane function may potentially underlie the toxicity observed.

Due to the toxic effects of amyloid formation, cells have developed a variety of preventative measures to correct or destroy misfolded proteins. One component of the protein homeostasis network is formed by 'chaperones', proteins designed to increase the efficiency of successful protein folds or correct misfolded proteins. Chaperones belong to two broad mechanistic families, holding and folding chaperones.[42] Holding chaperones are energy independent and prevent further unfolding or aggregation of proteins, whereas folding chaperones are ATP-dependent and can 'refold' proteins to restore function. Unfortunately, the effectiveness of chaperones is not boundless, and notably many diseases due to protein misfolding begin with mutations to genes responsible for protein chaperones.[43–45] From an exogenous perspective, natural polyphenols have long been scrutinized for their ability to prevent the fibrillation of misfolding proteins.[46] For example, epigallocatechin-3-gallate (EGCG), the primary polyphenol in green tea, has shown to stall the progression of amyloid formation, as well as restructure fibrillated proteins.[47–49] While no effective therapeutic compounds currently exist to treat amyloid diseases at the target misfolding protein level, further attention on small molecule inhibitors such as natural polyphenols is required to exploit their effects more fully.

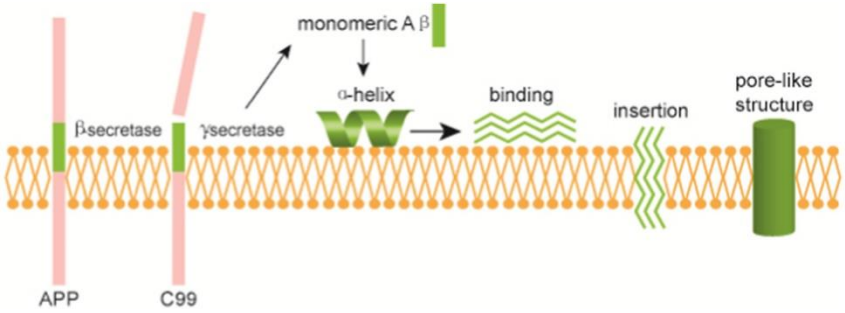


Figure 1.3. Endoproteolytic cleavage of amyloid precursor protein (APP) forms the misfolding protein, amyloid β (Aβ). Misfolding protein oligomers interact with the lipid bilayer, destabilising it and leading to apoptosis. Oligomers may interact on the surface, insert into the membrane or form annular pores, allowing for the unchecked mixing of the outer- and inner-cellular environment. Adapted with permission from Niu*, et al.* (2018)[50]

1.3 Protein structure determination

1.3.1 Methods for protein structure determination

To properly understand disease due to protein misfolding we must first understand the structure and molecular interactions of the proteins involved. One traditional method for protein structure determination is X-ray crystallography, a technique that is capable of atomic-level resolution but shows proteins as rigid structures locked within a crystal lattice.[51] This goes against the understanding that proteins sample a range of energetically favourable states governed by the thermodynamic landscape of the environment. On top of this, only between 2 – 10 % of proteins are amenable to crystallisation.[52] Nuclear magnetic resonance (NMR) spectroscopy is capable of revealing protein structure while in solution, taking into account their dynamic nature, but has a multitude of limitations such as the requirement for relatively large amounts of sample and typical application to proteins below 40 kDa (although NMR has been sporadically applied to significantly larger structures[53,54]). Cryo-electron microscopy (cryo-EM) has emerged as a powerful tool for protein structure determination in recent years, capable of 1 Å level resolution,[55] but is also not without limitations. Up until recently, cryo-EM was primarily used for the analysis of proteins at sizes exceeding 150 kDa, and although a small percentage (<1%) of structures below this have been determined, there are complications with image resolution that depend heavily on the heterogeneity of the protein of interest.[56,57]

These techniques have many strengths, in particular their resolution, but can require specific circumstances to be fully applicable. Misfolding proteins present intrinsic difficulties for structural determination due to their labile and transient nature, and as such, advances in the development of therapeutics and understanding disease aetiology have been limited.

1.3.2 Mass spectrometry

Mass spectrometry (MS) has emerged in recent decades as a powerful technique for uncovering structural features of proteins and is used as a complement or even alternative to more traditional structure determination techniques. While initially developed in the 1900s, technologies allowing the analysis of protein structure have only been available within the late 20th century.[58] MS has since proved itself an integral part of the protein structure

determination toolset due to its flexibility in handling both denatured and native protein samples, and ability to reveal peptide sequences, subunit stoichiometries, low-resolution 3D structure and more. The ability to handle a wide range of samples complimented by the hybrid approach of combining MS with other techniques, allows it to offer information on a wide range of proteins at different levels of structure.

The basis of MS relies on ionisation of a sample then separation of the analytes by their mass to charge ratio (*m/z*). Until the development of electrospray ionisation (ESI) and matrix-assisted laser desorption/ionization (MALDI), introduction of analytes to the gas phase would typically destroy macromolecules like proteins.[58] ESI as a 'soft ionisation' technique opened the door to the possibility of MS for small, non-volatile biomolecules all the way up to entire viral assemblies.[59]

ESI produces gas-phase ions by transferring them from the solution into the gas-phase in a 'gentle' manner that not only reduces the likelihood of molecular fragmentation but can also maintain non-covalent interactions.[60–63] The sample solution is ejected through a small opening, typically a capillary needle, between an electric potential. Due to the small size of solvent ejection tip the local electric field is very high, leading to the formation of a 'Taylor cone' as the sample aerosolises, producing droplets that rapidly desolvate (**Fig. 1.4**).[64] This process is often assisted by heat to aid in evaporation of clustered solvent molecules. Ionisation is possible in both positive and negative modes depending on the polarity of the electrical potential that is applied. In positive mode, protons, sodium, ammonium and other cations in the solvent can give droplets charge, as well as fixed charge motifs on the analyte. In negative mode this is achieved by the presence of hydroxide, acetate or deprotonation of acids in the sample.[65] As droplets travel toward the MS inlet they evaporate, increasing the relative charge density of the droplets until Coulombic repulsions trigger a 'droplet fission' which ejects a smaller, more highly charged droplet.[66] This process repeats until a solvent free, gas-phase ion is produced which can be separated and analysed,[66] although the hydrophilicity of many biomolecules can give rise to incomplete desolvation and result in less well resolved mass spectra.[67,68] Building upon this process has led to the development of nanoESI which utilises a small needle typically coated in conductive material to assist transfer of charge. This requires only nanolitre quantities of sample solution, typically at low micromolar protein concentrations, for analysis.[62] Aside from lower sample requirements,

nanoESI has also been shown to improve spectral resolution as smaller initial droplets effectively reduces the relative concentration of salts and potential contaminants in the droplet and aids desolvation.[69]

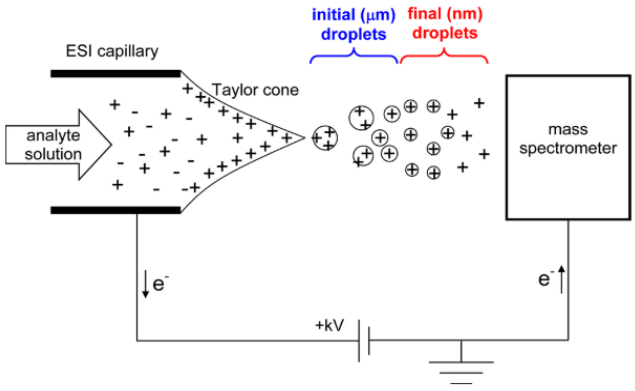


Figure 1.4. The nanoESI-MS process for ionisation of a protein. Solvent forms droplets which undergo desolvation followed by 'droplet fission' leading to solvent free gas-phase ions. Adapted from Konermann, *et al.* (2013).[66]

Following ionisation, analytes are separated by their mass to charge ratio (*m/z*). ESI is commonly paired with a time-of-flight (TOF) mass analyser for protein anlaysis. TOF separation works on the basis that low mass, highly charged species travelling over a fixed distance under the influence of an electric potential will reach the detector faster than high mass, lower charged species. Packets of ions are transmitted to the TOF analyser where they are pulsed for detection, and ion masses are resolved based on the time taken to pass the known distance of the analyser using a calibration approach (**Fig. 1.5**).

The quadrupole, another mass analyser, is often paired with ESI and TOF analysis as a hybrid MS instrument (Q-TOF). The quadrupole consists of four parallel rods to which RF and DC voltages are applied to produce oscillating electric fields. Ions are transmitted based on how they interact with the electric field generated, with ions of a particular *m/z* having a stable trajectory through the quadrupole.[60] This is advantageous because ions of a specific mass can be selected and subjected to a collision-induced dissociated (CID) cell before TOF separation. CID activates ions by propelling them into an inert gas, such as helium, nitrogen or argon. This

leads to an increase in the energy of the analyte causing it to fragment in a predictable manner. Calculating the expected mass losses due to amino acids detaching from the peptide allows for the structure of the protein can be determined.[70] The practicality of this technique paired with a QTOF MS lies in the ability to isolate individual masses, fragment them and then analyse the fragmentation products as this can give information such as subunit stoichiometries of proteins, small molecule identification and much more.[60,70–74]

Possibly the most important development in biological MS has been the advent of bottom-up proteomics, lending itself to the utilisation of CID as a protein sequencing technique. In this typical workflow, proteins are proteolytically digested and resulting peptides are separated by liquid chromatography (LC). Individual peptide ions are then selected based on *m/z* then fragmented by CID and the resulting products analysed. The sequential fragmentation pattern of activated peptides between the amide bond, known as *b/y*-ions, or the nitrogen and alpha-carbon, known as *c/z* ions, or between the carbonyl and the alpha-carbon, known as *x/y* ions, the peptide sequence can be reverse-engineered from the fragmentation spectra which is then cross-checked with an online database (**Fig. 1.5**).[75] The method of activation used will preference a particular output of fragmentation pattern, for example CID preferences *b/y* ion formation whereas another dissociation methods, electron transfer dissociation (ETD), preferences *c/z* ions.[76] Modern proteomics analysis of this nature typically utilises an orbitrap mass analyser due to superior mass resolution and accuracy when compared with other instruments.[77] Orbitraps, unlike TOF, do not measure mass by ion travel time along a known path; instead the ions oscillate around a central electrode, generating an image current which is detected before Fourier transformation to yield a mass spectrum.[78]

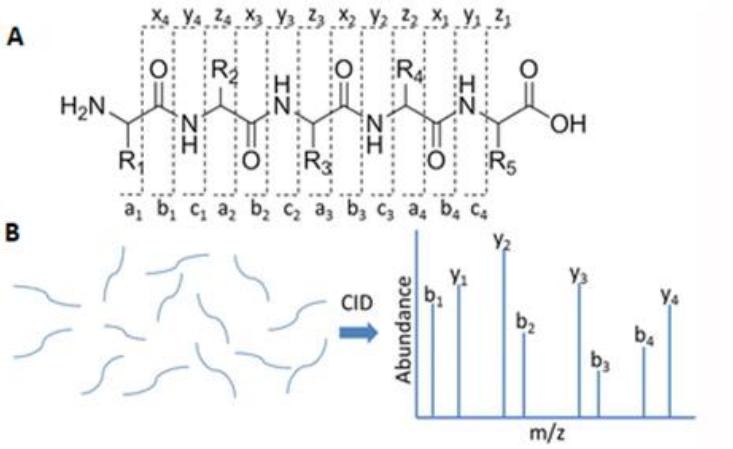


Figure 1.5. Bottom-up proteomics. (**A**) Peptides selected by MS/MS and subject to CID fragmentation yield *b/y* ions while analogous activation techniques can lead to *c/z* and *a/x* ions. (**B**) The fragmentation spectra are analysed by software, or by hand, calculating the mass losses due to amino acid dissociations giving a peptide sequence. These mass losses are crosschecked against online databases to identify protein with which they belong. Adapted from Zhang, *et al.* (2013).[79]

1.3.3 Native mass spectrometry

With the advent of ESI it became possible to not only ionise large biomolecules without fragmentation, but also maintain non-covalent interactions that determine protein higher order structure. Early work with 'native MS' revealed subunit stoichiometries of horse liver and yeast alcohol dehydrogenases as a dimer and tetramer, respectively, an experiment considered impossible years earlier.[80] Current applications of native MS reveal information on vastly more complicated protein complexes. For example Gordiyenko*, et al*.[81] revealed the GTP-bound translation initiation factor 2 to be a decamer rather than the previously thought pentamer, changing paradigms in structural biology. It is not only possible to observe protein-protein interactions by native MS, but a variety of studies apply CID and MS to understand protein-ligand binding, using CID activation to determine the dissociation constants of protein assemblies.[82] For example, Yen*, et al.*[83] investigated binding affinities between the human purinergic receptor $P2Y_1$ and adenosine 5'-diphosphate, -triphosphate and the drug MRS2500, highlighting the effectiveness of native MS as a tool for investigating non-covalent protein interactions. Reading*, et al.* used native MS and titration with detergents and lipids,

as well as temperature modulation, to establish equilibria between tetrameric and pentameric oligomers of the mechanosensitive channel of large conductance or MscL, demonstrating the ability to apply native MS to membrane proteins which are particularly challenging to structural biology.

While native MS is a powerful technique for investigating complex architectures, it can also be used to probe the relative foldedness of a protein through changes in the charge state distribution of the protein ion.[84] This is due to the fact that unfolded proteins expose more ionisable residues to the solvent, shifting the charge state distribution to more highly charged ions.[66] This observation has been exploited to gain insight into the stability of proteins,[85] for example, through monitoring the denaturation of ferricytochrome *c* by acid in differing methanol concentrations,[86] and in recent years has been further enhanced by coupling of ion mobility technologies with MS.

1.3.4 Ion-mobility mass spectrometry

This gentle conditions of native MS is particularly beneficial for the analysis proteins that are unable to be crystalised and are difficult to produce recombinantly, resulting in low protein yields like membrane proteins,[87] and misfolding protein aggregates.[88] Furthermore, native MS with ion-mobility (IM-MS) is capable of measuring higher-order structural features of proteins regardless of their dynamic nature, which is an intrinsic property of amyloid-forming proteins.

During IM-MS, ions are separated based on their three-dimensional, rotationally-averaged structure known as the collisional cross-section (CCS), as well as their *m/z*. In a traditional drift tube IM experiment, ions tumble through the IM cell under the influence of an applied voltage which moves them through a tube of inert gas passing in the opposite direction. Larger ions encounter more collisions with the gas, slowing their passage, whereas ions with a smaller CCS encounter fewer collisions and travel faster.[89] This yields an arrival time distribution (ATD) for ions which can be used to calculate their CCS, using the Mason-Schamp equation[90]:

$$\Omega = \frac{3ze}{16N}\left(\frac{2\pi}{\mu k_B T}\right)^{\frac{1}{2}}\frac{1}{K_0}$$

Where Ω is the CCS, K_0 is the reduced mobility (depending on temperature and pressure of the drift tube), z is the charge of the ion, e is the elementary charge, N is the number density of the drift gas, μ is the reduced mass of the ion-neutral drift pair has pair, k_B is the Boltzmann constant and T is the gas temperature.[91]

Since the concept of the original drift tube was described a century ago, many developments in IM-MS have been made. For example, travelling wave ion-mobility spectrometry (TWIMS), the mode of separation used in this thesis, was incorporated into the first commercial IM-MS instruments and offered enhanced detection of large analytes such as proteins. A TWIMS sector is comprised of a series of stacked ring ion guides capable of generating radiofrequencies. The radiofrequency voltages are set at alternating polarities which confine ions radially, while a current passing through the centre pushes ions axially. The axial voltage is pulsed, creating a 'travelling wave' that pushes the ions through the IM sector. Ions of a lower mobility experience a greater number of 'roll over' events, taking longer to reach the detector.[91] Larger ions experience more roll over events, interacting with the gas filled cell, giving separation and a means to measure CCS by a calibration approach from drift times (**Fig. 1.6**).[92] Height and velocity of the travelling wave can be tuned to optimise separation.

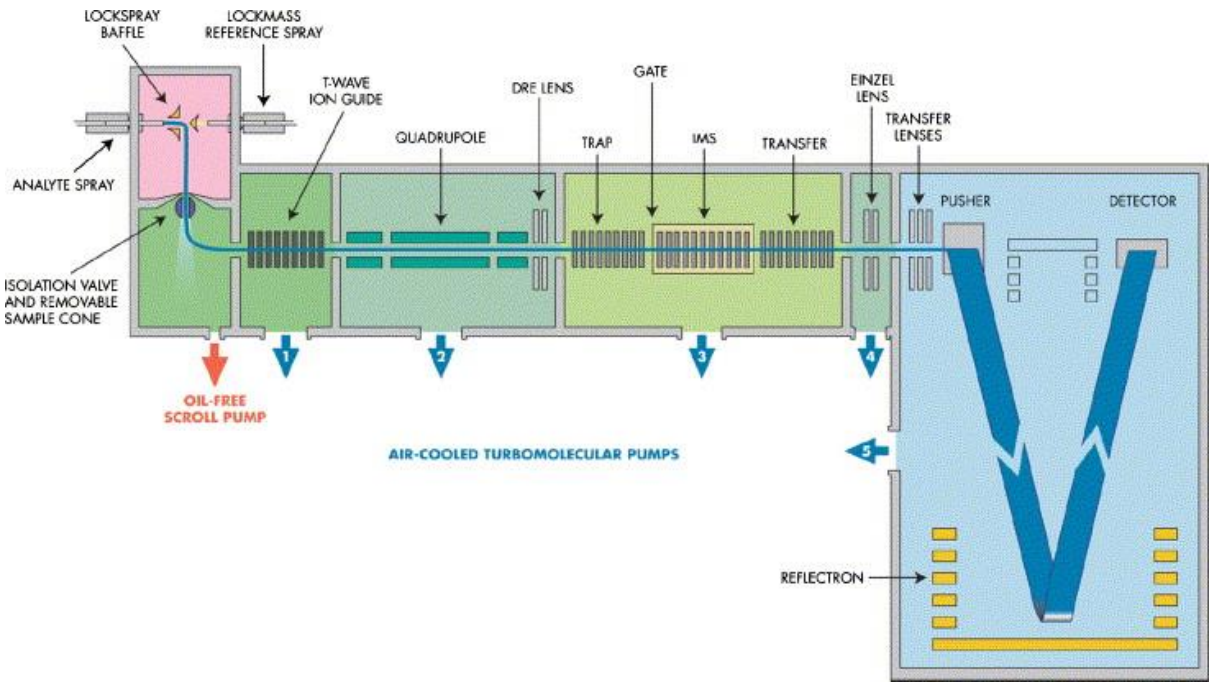


Figure 1.6. Schematic diagram of a Waters Synapt G1 IM-MS. The analyte is ionised by ESI and separated by its mass to charge ratio in the time-of-flight mass analyser. The individual mass can be selected by the quadrupole and be activated in either the trap or transfer CID cells. A TWIMS drift cell separates ions based on their mobility. Adapted with permission from Pringle*, et al* (2007).[93]

While originally applied to small molecules, the first studies using native IM-MS in a protein context confirmed the 11-member ring topology of *trp* RNA in the gas-phase in conformations consistent with high resolution structures, while binding of tryptophan and RNA were shown to stabilise the complex.[94] The ability to retain solution phase structural properties as demonstrated by IM highlighted the potential of the technique for protein structure determination. IM-MS is particularly well suited to proteins that are difficult to gain structural information on, like membrane and misfolding protein. This was showcased when Smith*, et al*. observed that β2-microglobulin formed elongated oligomeric structure before aggregation,[95] and more recently when Österlund*, et al.* demonstrated that IM-MS can be used to gain structural insights into highly transient amyloid β 42 (Aβ42) oligomers. Native IM-MS revealed Aβ42 formed hexameric micelle clusters which, when compared with computational models, showed pore like structures, shedding light on potential disease mechanisms.[96]

Further exploration of IM-MS as a protein structure determination technique has led to the development of collisional-induced unfolding (CIU). In this approach, the ATDs of native-like proteins in the gas-phase are monitored while CID voltages are increased. This allows for observation of protein unfolding as the ATD increases with increased collision energy. For example, utilisation of this technique has revealed that the allosteric binding of HIV protease inhibitors conferred stabilisation of human zinc metalloprotease (ZMPSTE24) by noting increased collision energies were required to unfold apo-ZMPSTE24.[97]

1.3.5 Cross-linking mass spectrometry

MS in conjunction with chemical crosslinking (XL-MS) is another emerging technique for understanding tertiary and quaternary protein structure. XL-MS works by chemically ligating

specific amino acids within a protein of interest, giving interspatial distance constraints that allow for the interrogation of protein binding interfaces, intramolecular contacts and stoichiometry.[98] A typical chemical crosslinker consists of two main components; the spacer arm and the reactive group (**Fig. 1.7**). The reactive group consists of a chemical moiety that irreversibly reacts with amino acids. Common reactive groups include n-hydroxy succinimide (NHS) for basic residues, hydrazide groups for acidic residues and maleimides for thiol containing residues.

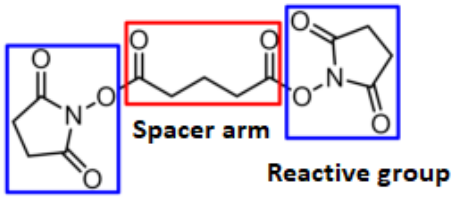


Figure 1.7. A simple chemical crosslinker, disuccinimidyl glutarate (DSG), with a carbon chain spacer arm and two n-hydroxy succinimide reactive groups which target basic protein sites like lysine.

Cross-linked peptides are typically sequenced by bottom-up proteomics (**Fig 1.8**).[98–101] Cross-linking has the potential to yield three kinds of modifications: interpeptide links, intrapeptide links and 'dead-end' links whereby a crosslinker has attached to the protein at one end but hydrolysed on the other. XL-MS software calculates the added masses from the three possible cross-link modifications and scans the MS/MS spectra and to identify matching sequence data, and therefore linkage sites.[100] The most commonly utilised examples of XL-MS software currently are pLink,[102] MeroX[103] and XlinkX.[104] Once identified, these linkages can give information on the interspatial arrangements of amino acid residues through the spacer-arm which, since it is a known length, acts as a 'molecular ruler'.[101] This data can be used to guide 3D computational structure modelling, build linkage maps and establish subunit topology. When combined with complimentary techniques like cryo-EM, XL-MS can reveal subunit stoichiometries, binding interfaces and 3D structures, perhaps best exemplified by Plaschka, *et al.* who determined the structure of co-activator complex Mediator (cMed) of *Saccharomyces cerevisiae*.[105] Low resolution EM data, likely due to heterogeneity of the cMed

structure, was aided by XL-MS to model the RNA polymerase II–Mediator complex and show how it was stabilised by several transcription factors.

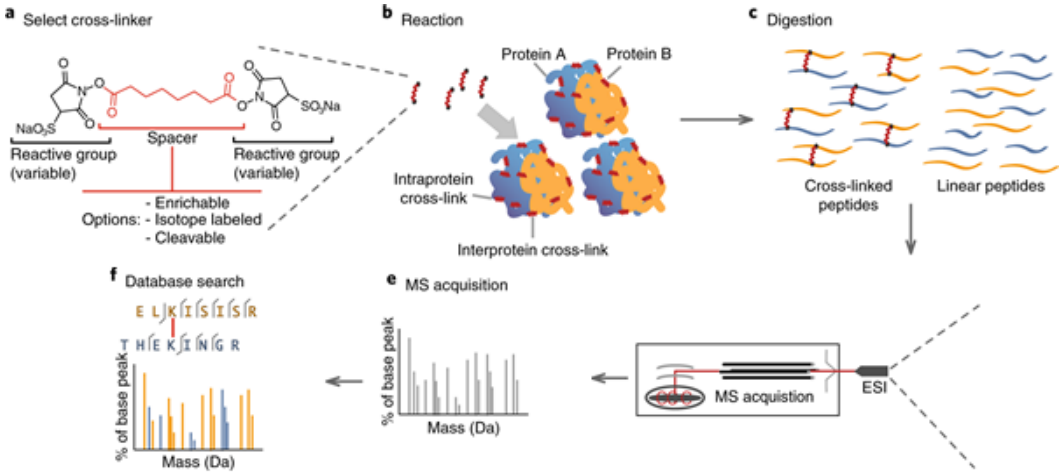


Figure 1.8. Once selected, a cross-linker is reacted with the protein of interest. The protein is digested, often from an excised gel electrophoresis band, then digested. The protein is then analysed by liquid chromatography MS and the cross-links are identified by spectral matching with database searches. Adapted with permission from O'Reilly, *et al*. (2018).[100]

XL-MS allows for protein structural information to be covalently 'locked' in place during analysis, yielding a 'snapshot' of a protein configuration. This presents advantages for capturing highly dynamic proteins in various states that are only accessible in solution. For example, this was demonstrated when Schmidt*, et al.* showed cross-linking could lock in place conformational changes of chloroplast ATPase.[106] Using XL-MS, different conformational states were covalently trapped, with and without phosphorylation of the ATPase, suggesting the post-translational modification was essential to the closed and opened form of the complex. The ability to study dynamic protein structures is not only relevant to folded native proteins but also of particular interest to elucidating structural information from misfolding proteins and oligomers. For example, McDonald*, et al.* utilised XL-MS to better understand the prion protein (PrP^{c}) which, due to regions of intrinsic disorder, is difficult to categorise with traditional protein structure determination techniques.[107] In this study they were able to define intramolecular contacts between the C- and N-terminal domain that are considered important to disease pathology.

The techniques listed describe the ability for MS interrogation of native structures and subunit architectures from proteins that are unamenable to traditional structure determination techniques. Protein structure determination is essential to understanding misfolding diseases for the development of therapeutics, however the analysis of amyloid proteins present special challenges.

1.4 Aims

This thesis aims to demonstrate the use of an MS-based approach to explore proteins that are less amenable to traditional structure determination techniques. i) The effects of chaperones on binding and stability of misfolding proteins will be explored. ii) The ability of lipids to impact protein aggregation and the efficacy of fibril inhibitors. iii) A modular synthetic protocol will be developed for synthesis of a cross-linker library to add diversity to a currently limited range of cross-linking reagents.

1.5 References

(1) Anfinsen, C. B. Principles That Govern the Folding of Protein Chains. *Science* **1973**, *181* (4096), 223–230. https://doi.org/10.1126/science.181.4096.223.

(2) Levinthal, C. Are There Pathways for Protein Folding? *J. Chim. Phys.* **1968**, *65* (1), 44–45. https://doi.org/10.1051/jcp/1968650044.

(3) Rojnuckarin, A.; Kim, S.; Subramaniam, S. Brownian Dynamics Simulations of Protein Folding: Access to Milliseconds Time Scale and Beyond. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (8), 4288–4292. https://doi.org/10.1073/pnas.95.8.4288.

(4) Fersht, A. R. Nucleation Mechanisms in Protein Folding. *Curr. Opin. Struct. Biol.* **1997**, *7* (1), 3–9. https://doi.org/10.1016/S0959-440X(97)80002-4.

(5) Dinner, A. R.; Šalib, A.; Smitha, L. J.; Dobsona, C. M.; Karplus, M. Understanding Protein Folding via Free-Energy Surfaces from Theory and Experiment. *Trends Biochem. Sci.* **2000**, *25* (7), 331–339. https://doi.org/10.1016/S0968-0004(00)01610-8.

(6) Khan TA, N. A. Protein Folding and Misfolding: A Perspective from Theory. *J. Glycomics Lipidomics* **2015**, *05* (01), 868–893. https://doi.org/10.4172/2153-0637.1000128.

(7) Fersht, A. R. Transition-State Structure as a Unifying Basis in Protein-Folding Mechanisms: Contact Order, Chain Topology, Stability, and the Extended Nucleus Mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (4), 1525–1529. https://doi.org/10.1073/pnas.97.4.1525.

(8) Gotts, S. *Mechanisms Underlying Enhanced Processing Efficiency in Neural Systems*, 4th editio.; Garland Science: New York, 2003; Vol. 32. https://doi.org/10.1024/0301-1526.32.1.54.

(9) Abdellah, Z.; Ahmadi, A.; Ahmed, S.; Aimable, M.; Ainscough, R.; Almeida, J.; Almond, C.; Ambler, A.; Ambrose, K.; Ambrose, K.; et al. Finishing the Euchromatic Sequence of the Human Genome. *Nature* **2004**, *431* (7011), 931–945. https://doi.org/10.1038/nature03001.

(10) Capaldi, A. P.; Kleanthous, C.; Radford, S. E. Im7 Folding Mechanism: Misfolding on a Path to the Native State. *Nat. Struct. Biol.* **2002**, *9* (3), 209–216. https://doi.org/10.1038/nsb757.

(11) Kelly, J. W. The Alternative Conformations of Amyloidogenic Proteins and Their Multi-Step Assembly Pathways. *Curr. Opin. Struct. Biol.* **1998**, *8* (1), 101–106. https://doi.org/10.1016/S0959-440X(98)80016-X.

(12) Barrow, C. J.; Yasuda, A.; Kenny, P. T. M.; Zagorski, M. G. Solution Conformations and Aggregational Properties of Synthetic Amyloid β-Peptides of Alzheimer's Disease. Analysis of Circular Dichroism Spectra. *J. Mol. Biol.* **1992**, *225* (4), 1075–1093. https://doi.org/10.1016/0022-2836(92)90106-T.

(13) Sambataro, F.; Pennuto, M. Post-Translational Modifications and Protein Quality Control in Motor Neuron and Polyglutamine Diseases. *Front. Mol. Neurosci.* **2017**, *10*. https://doi.org/10.3389/fnmol.2017.00082.

(14) Johnson, B. S.; Snead, D.; Lee, J. J.; McCaffery, J. M.; Shorter, J.; Gitler, A. D. TDP-43 Is Intrinsically Aggregation-Prone, and Amyotrophic Lateral Sclerosis-Linked Mutations Accelerate Aggregation and Increase Toxicity. *J. Biol. Chem.* **2009**, *284* (30), 20329–20339. https://doi.org/10.1074/jbc.M109.010264.

(15) Hardy, J. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science* **2002**, *297* (5580), 353–356. https://doi.org/10.1126/science.1072994.

(16) Dobson, C. M. Protein Folding and Misfolding. *Nature* **2003**, *426* (6968), 884–890. https://doi.org/10.1038/nature02261.

(17) Fink, A. L. Protein Aggregation: Folding Aggregates, Inclusion Bodies and Amyloid. *Fold. Des.* **1998**, *3* (1), R9–R23. https://doi.org/10.1016/S1359-0278(98)00002-9.

(18) Selkoe, D. J. Folding Proteins in Fatal Ways. *Nature* **2003**, *426* (6968), 900–904. https://doi.org/10.1038/nature02264.

(19) Chiti, F.; Stefani, M.; Taddei, N.; Ramponi, G.; Dobson, C. M. Rationalization of the Effects of Mutations on Peptide and Protein Aggregation Rates. *Nature* **2003**, *424* (6950), 805–808. https://doi.org/10.1038/nature01891.

(20) Stranks, S. D.; Ecroyd, H.; Van Sluyter, S.; Waters, E. J.; Carver, J. A.; Von Smekal, L. Model for Amorphous Aggregation Processes. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2009**, *80* (5), 051907. https://doi.org/10.1103/PhysRevE.80.051907.

(21) Dobson, C. M. Principles of Protein Folding, Misfolding and Aggregation. *Semin. Cell Dev. Biol.* **2004**, *15* (1), 3–16. https://doi.org/10.1016/j.semcdb.2003.12.008.

(22) Ekeowa, U. I.; Freeke, J.; Miranda, E.; Gooptu, B.; Bush, M. F.; Perez, J.; Teckman, J.; Robinson, C. V.; Lomas, D. A. Defining the Mechanism of Polymerization in the Serpinopathies. *Proc. Natl. Acad. Sci.* **2010**, *107* (40), 17146–17151. https://doi.org/10.1073/pnas.1004785107.

(23) Wasmer, C.; Lange, A.; Van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H. Amyloid Fibrils of the HET-s(218-289) Prion Form a β Solenoid with a Triangular Hydrophobic Core. *Science* **2008**, *319* (5869), 1523–1526. https://doi.org/10.1126/science.1151839.

(24) Kulikova, A. A.; Makarov, A. A.; Kozin, S. A. Roles of Zinc Ions and Structural Polymorphism of β-Amyloid in the Development of Alzheimer's Disease. *Mol. Biol.* **2015**, *49* (2), 217–230. https://doi.org/10.1134/S0026893315020065.

(25) Chiti, F.; Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **2006**, *75* (1), 333–366. https://doi.org/10.1146/annurev.biochem.75.101304.123901.

(26) Selkoe, D. J.; Hardy, J. The Amyloid Hypothesis of Alzheimer's Disease at 25 Years. *EMBO Mol. Med.* **2016**, *8* (6), 595–608. https://doi.org/10.15252/emmm.201606210.

(27) Dodart, J.-C.; Bales, K. R.; Gannon, K. S.; Greene, S. J.; DeMattos, R. B.; Mathis, C.; DeLong, C. A.; Wu, S.; Wu, X.; Holtzman, D. M.; et al. Immunization Reverses Memory Deficits without Reducing Brain Aβ Burden in Alzheimer's Disease Model. *Nat. Neurosci.* **2002**, *5* (5), 452–457. https://doi.org/10.1038/nn842.

(28) Nilsberth, C.; Westlind-Danielsson, A.; Eckman, C. B.; Condron, M. M.; Axelman, K.; Forsell, C.; Stenh, C.; Luthman, J.; Teplow, D. B.; Younkin, S. G.; et al. The "Arctic" APP Mutation (E693G) Causes Alzheimer's Disease by Enhanced Aβ Protofibril Formation. *Nat. Neurosci.* **2001**, *4* (9), 887–893. https://doi.org/10.1038/nn0901-887.

(29) Ryan, R. E.; Hill, S. J. Improving the Experiences and Health of People with Multimorbidity: Exploratory Research with Policymakers and Information Providers on Comorbid Arthritis. *Aust. J. Prim. Health* **2014**, *20* (2), 188–196. https://doi.org/10.1071/PY12083.

(30) Wang, J.; Dickson, D. W.; Trojanowski, J. Q.; Lee, V. M. Y. The Levels of Soluble versus Insoluble Brain Aβ Distinguish Alzheimer's Disease from Normal and Pathologic Aging. *Exp. Neurol.* **1999**, *158* (2), 328–337. https://doi.org/10.1006/exnr.1999.7085.

(31) Reixach, N.; Deechongkit, S.; Jiang, X.; Kelly, J. W.; Buxbaum, J. N. Tissue Damage in the Amyloidoses: Transthyretin Monomers and Nonnative Oligomers Are the Major Cytotoxic Species in Tissue Culture. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (9), 2817–2822. https://doi.org/10.1073/pnas.0400062101.

(32) Sangwan, S.; Zhao, A.; Adams, K. L.; Jayson, C. K.; Sawaya, M. R.; Guenther, E. L.; Pan, A. C.; Ngo, J.; Moore, D. M.; Soriaga, A. B.; et al. Atomic Structure of a Toxic, Oligomeric Segment of SOD1 Linked to Amyotrophic Lateral Sclerosis (ALS). *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (33), 8770–8775. https://doi.org/10.1073/pnas.1705091114.

(33) Sani, M.-A.; Gehman, J. D.; Separovic, F. Lipid Matrix Plays a Role in Abeta Fibril Kinetics and Morphology. *FEBS Lett.* **2011**, *585* (5), 749–754. https://doi.org/10.1016/j.febslet.2011.02.011.

(34) Hellstrand, E.; Nowacka, A.; Topgaard, D.; Linse, S.; Sparr, E. Membrane Lipid Co-Aggregation with α-Synuclein Fibrils. *PLoS One* **2013**, *8* (10), e77235. https://doi.org/10.1371/journal.pone.0077235.

(35) Terakawa, M. S.; Yagi, H.; Adachi, M.; Lee, Y. H.; Goto, Y. Small Liposomes Accelerate the Fibrillation of Amyloid β(1- 40). *J. Biol. Chem.* **2015**, *290* (2), 815–826. https://doi.org/10.1074/jbc.M114.592527.

(36) Burke, K. A.; Yates, E. A.; Legleiter, J. Biophysical Insights into How Surfaces, Including Lipid Membranes, Modulate Protein Aggregation Related to Neurodegeneration. *Front. Neurol.* **2013**, *4 MAR*, 17. https://doi.org/10.3389/fneur.2013.00017.

(37) Matsuzaki, K. How Do Membranes Initiate Alzheimers Disease? Formation of Toxic Amyloid Fibrils by the Amyloid β-Protein on Ganglioside Clusters. *Acc. Chem. Res.* **2014**, *47* (8), 2397–2404. https://doi.org/10.1021/ar500127z.

(38) Ho, C. S.; Khadka, N. K.; She, F.; Cai, J.; Pan, J. Polyglutamine Aggregates Impair Lipid Membrane Integrity and Enhance Lipid Membrane Rigidity. *Biochim. Biophys. Acta - Biomembr.* **2016**, *1858* (4), 661–670. https://doi.org/10.1016/j.bbamem.2016.01.016.

(39) Bode, D. C.; Freeley, M.; Nield, J.; Palma, M.; Viles, J. H. Amyloid-β Oligomers Have a Profound Detergent-like Effect on Lipid Membrane Bilayers, Imaged by Atomic Force and Electron Microscopy. *J. Biol. Chem.* **2019**, *294* (19), 7566–7572. https://doi.org/10.1074/jbc.AC118.007195.

(40) Serra-Batiste, M.; Ninot-Pedrosa, M.; Bayoumi, M.; Gairí, M.; Maglia, G.; Carulla, N. Aβ42 Assembles into Specific β-Barrel Pore-Forming Oligomers in Membrane-Mimicking Environments. *Proc. Natl. Acad. Sci.* **2016**, *113* (39), 10866–10871. https://doi.org/10.1073/pnas.1605104113.

(41) Perissinotto, F.; Rondelli, V.; Parisse, P.; Tormena, N.; Zunino, A.; Almásy, L.; Merkel, D. G.; Bottyán, L.; Sajti, S.; Casalis, L. GM1 Ganglioside Role in the Interaction of Alpha-Synuclein with Lipid Membranes: Morphology and Structure. *Biophys. Chem.* **2019**, *255*, 106272. https://doi.org/10.1016/j.bpc.2019.106272.

(42) Gregersen, N.; Bross, P.; Vang, S.; Christensen, J. H. Protein Misfolding and Human Disease. *Annu. Rev. Genomics Hum. Genet.* **2006**, *7* (1), 103–124. https://doi.org/10.1146/annurev.genom.7.080505.115737.

(43) Macario, A. J. L.; De Macario, E. C. Sick Chaperones and Ageing: A Perspective. *Ageing Res. Rev.* **2002**, *1* (2), 295–311. https://doi.org/10.1016/S1568-1637(01)00005-8.

(44) Uversky, V. N.; Fink, A. L. *Protein Misfolding, Aggregation, and Conformational Diseases*; 2006; Vol. 4. https://doi.org/10.1007/b136464.

(45) Soto, C. Unfolding the Role of Protein Misfolding in Neurodegenerative Diseases. *Nat. Rev. Neurosci.* **2003**, *4* (1), 49–60. https://doi.org/10.1038/nrn1007.

(46) Henríquez, G.; Gomez, A.; Guerrero, E.; Narayan, M. Potential Role of Natural Polyphenols against Protein Aggregation Toxicity: In Vitro, in Vivo, and Clinical Studies. *ACS Chem. Neurosci.* **2020**, *11* (19), 2915–2934. https://doi.org/10.1021/acschemneuro.0c00381.

(47) Hudson, S. A.; Ecroyd, H.; Dehle, F. C.; Musgrave, I. F.; Carver, J. A. (−)-Epigallocatechin-3-Gallate (EGCG) Maintains κ-Casein in Its Pre-Fibrillar State without Redirecting Its Aggregation Pathway. *J. Mol. Biol.* **2009**, *392* (3), 689–700. https://doi.org/10.1016/j.jmb.2009.07.031.

(48) Chandrashekaran, I. R.; Adda, C. G.; MacRaild, C. A.; Anders, R. F.; Norton, R. S. EGCG Disaggregates Amyloid-like Fibrils Formed by Plasmodium Falciparum Merozoite Surface Protein 2. *Arch. Biochem. Biophys.* **2011**, *513* (2), 153–157. https://doi.org/10.1016/j.abb.2011.07.008.

(49) Bieschke, J.; Russ, J.; Friedrich, R. P.; Ehrnhoefer, D. E.; Wobst, H.; Neugebauer, K.; Wanker, E. E. EGCG Remodels Mature α-Synuclein and Amyloid-β Fibrils and Reduces Cellular Toxicity. *Proc. Natl. Acad. Sci.* **2010**, *107* (17), 7710–7715. https://doi.org/10.1073/pnas.0910723107.

(50) Niu, Z.; Zhang, Z.; Zhao, W.; Yang, J. Interactions between Amyloid β Peptide and Lipid Membranes. *Biochim. Biophys. Acta - Biomembr.* **2018**, *1860* (9), 1663–1669. https://doi.org/10.1016/j.bbamem.2018.04.004.

(51) McPherson, A. *Introduction to Macromolecular Crystallography: Second Edition*; 2008. https://doi.org/10.1002/9780470391518.

(52) Wang, H.; Feng, L.; Webb, G. I.; Kurgan, L.; Song, J.; Lin, D. Critical Evaluation of Bioinformatics Tools for the Prediction of Protein Crystallization Propensity. *Brief. Bioinform.* **2018**. https://doi.org/10.1093/bib/bbx018.

(53) Vandermarliere, E.; Stes, E.; Gevaert, K.; Martens, L. Resolution of Protein Structure by Mass Spectrometry. *Mass Spectrom. Rev.* **2016**, *35* (6), 653–665. https://doi.org/10.1002/mas.21450.

(54) Takeuchi, K.; Baskaran, K.; Arthanari, H. Structure Determination Using Solution NMR: Is It Worth the Effort? *J. Magn. Reson.* **2019**, *306*, 195–201. https://doi.org/10.1016/j.jmr.2019.07.045.

(55) Yip, K. M.; Fischer, N.; Paknia, E.; Chari, A.; Stark, H. Atomic-Resolution Protein Structure Determination by Cryo-EM. *Nature* **2020**, *587* (7832), 157–161. https://doi.org/10.1038/s41586-020-2833-4.

(56) Renaud, J.-P.; Chari, A.; Ciferri, C.; Liu, W.; Rémigy, H.-W.; Stark, H.; Wiesmann, C. Cryo-EM in Drug Discovery: Achievements, Limitations and Prospects. *Nat. Rev. Drug Discov.* **2018**, *17* (7), 471–492. https://doi.org/10.1038/nrd.2018.77.

(57) Wu, M.; Lander, G. C. How Low Can We Go? Structure Determination of Small Biological Complexes Using Single-Particle Cryo-EM. *Curr. Opin. Struct. Biol.* **2020**, *64*, 9–16. https://doi.org/10.1016/j.sbi.2020.05.007.

(58) Griffiths, J. A Brief History of Mass Spectrometry. *Anal. Chem.* **2008**, *80* (15), 5678–5683. https://doi.org/10.1021/ac8013065.

(59) Uetrecht, C.; Barbu, I. M.; Shoemaker, G. K.; Van Duijn, E.; Heck, A. J. R. Interrogating Viral Capsid Assembly with Ion Mobility-Mass Spectrometry. *Nat. Chem.* **2011**, *3* (2), 126–132. https://doi.org/10.1038/nchem.947.

(60) Barth, M.; Schmidt, C. Native Mass Spectrometry-A Valuable Tool in Structural Biology. *J. Mass Spectrom.* **2020**, *55* (10), e4578. https://doi.org/10.1002/jms.4578.

(61) Allison, T. M.; Bechara, C. Structural Mass Spectrometry Comes of Age: New Insight into Protein Structure, Function and Interactions. *Biochem. Soc. Trans.* **2019**, *47* (1), 317–327. https://doi.org/10.1042/BST20180356.

(62) Heck, A. J. R. Native Mass Spectrometry: A Bridge between Interactomics and Structural Biology. *Nat. Methods* **2008**, *5* (11), 927–933. https://doi.org/10.1038/nmeth.1265.

(63) Loo, J. A. Studying Noncovalent Protein Complexes by Electrospray Ionization Mass Spectrometry. *Mass Spectrom. Rev.* **1997**, *16* (1), 1–23. https://doi.org/10.1002/(SICI)1098-2787(1997)16:1<1::AID-MAS1>3.0.CO;2-L.

(64) Kebarle, P.; Verkcerk, U. H. Electrospray: From Ions in Solution to Ions in the Gas Phase, What We Know Now. *Mass Spectrom. Rev.* **2009**, *28* (6), 898–917. https://doi.org/10.1002/mas.20247.

(65) Banerjee, S.; Mazumdar, S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int. J. Anal. Chem.* **2012**, *2012*, 1–40. https://doi.org/10.1155/2012/282574.

(66) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9. https://doi.org/10.1021/ac302789c.

(67) Cech, N. B.; Enke, C. G. Relating Electrospray Ionization Response to Nonpolar Character of Small Peptides. *Anal. Chem.* **2000**, *72* (13), 2717–2723. https://doi.org/10.1021/ac9914869.

(68) Schmidt, A.; Bahr, U.; Karas, M. Influence of Pressure in the First Pumping Stage on Analyte Desolvation and Fragmentation in Nano-ESI MS. *Anal. Chem.* **2001**, *73* (24), 6040–6046. https://doi.org/10.1021/ac010451h.

(69) Karas, M.; Bahr, U.; Dülcks, T. Nano-Electrospray Ionization Mass Spectrometry: Addressing Analytical Problems beyond Routine. *Fresenius. J. Anal. Chem.* **2000**, *366* (6–7), 669–676. https://doi.org/10.1007/s002160051561.

(70) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. *Methods Enzymol.* **2005**, *402* (1993), 148–185. https://doi.org/10.1016/S0076-6879(05)02005-7.

(71) Benesch, J. L. P.; Ruotolo, B. T. Mass Spectrometry: Come of Age for Structural and Dynamical Biology. *Curr. Opin. Struct. Biol.* **2011**, *21* (5), 641–649. https://doi.org/10.1016/j.sbi.2011.08.002.

(72) Sharon, M.; Robinson, C. V. The Role of Mass Spectrometry in Structure Elucidation of Dynamic Protein Complexes. *Annu. Rev. Biochem.* **2007**, *76* (1), 167–193. https://doi.org/10.1146/annurev.biochem.76.061005.090816.

(73) Taverner, T.; Hernández, H.; Sharon, M.; Ruotolo, B. T.; Matak-Vinković, D.; Devos, D.; Russell, R. B.; Robinson, C. V. Subunit Architecture of Intact Protein Complexes from Mass Spectrometry and Homology Modeling. *Acc. Chem. Res.* **2008**, *41* (5), 617–627. https://doi.org/10.1021/ar700218q.

(74) Laganowsky, A.; Reading, E.; Hopper, J. T. S.; Robinson, C. V. Mass Spectrometry of Intact Membrane Protein Complexes. *Nat. Protoc.* **2013**, *8* (4), 639–651. https://doi.org/10.1038/nprot.2013.024.

(75) Steen, H.; Mann, M. The Abc's (and Xyz's) of Peptide Sequencing. *Nat. Rev. Mol. Cell Biol.* **2004**, *5* (9), 699–711. https://doi.org/10.1038/nrm1468.

(76) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2016**, *88* (1), 30–51. https://doi.org/10.1021/acs.analchem.5b04563.

(77) Gillet, L. C.; Leitner, A.; Aebersold, R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu. Rev. Anal. Chem.* **2016**, *9*, 449–472. https://doi.org/10.1146/annurev-anchem-071015-041535.

(78) Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal. Chem.* **2015**, *8* (1), 61–80. https://doi.org/10.1146/annurev-anchem-071114-040325.

(79) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113* (4), 2343–2394. https://doi.org/10.1021/cr3003533.

(80) Loo, J. A. Observation of Large Subunit Protein Complexes by Electrospray Ionization Mass Spectrometry. *J. Mass Spectrom.* **1995**, *30* (1), 180–183.

https://doi.org/10.1002/jms.1190300127.

(81) Gordiyenko, Y.; Schmidt, C.; Jennings, M. D.; Matak-Vinkovic, D.; Pavitt, G. D.; Robinson, C. V. EIF2B Is a Decameric Guanine Nucleotide Exchange Factor with a Γ2ε2 Tetrameric Core. *Nat. Commun.* **2014**, *5* (1), 3902. https://doi.org/10.1038/ncomms4902.

(82) Chen, G.; Fan, M.; Liu, Y.; Sun, B.; Liu, M.; Wu, J.; Li, N.; Guo, M. Advances in MS Based Strategies for Probing Ligand-Target Interactions: Focus on Soft Ionization Mass Spectrometric Techniques. *Front. Chem.* **2019**, *7* (October), 1–17. https://doi.org/10.3389/fchem.2019.00703.

(83) Yen, H.-Y.; Hopper, J. T. S.; Liko, I.; Allison, T. M.; Zhu, Y.; Wang, D.; Stegmann, M.; Mohammed, S.; Wu, B.; Robinson, C. V. Ligand Binding to a G Protein–Coupled Receptor Captured in a Mass Spectrometer. *Sci. Adv.* **2017**, *3* (6), e1701016. https://doi.org/10.1126/sciadv.1701016.

(84) Testa, L.; Brocca, S.; Grandori, R. Charge-Surface Correlation in Electrospray Ionization of Folded and Unfolded Proteins. *Anal. Chem.* **2011**, *83* (17), 6459–6463. https://doi.org/10.1021/ac201740z.

(85) KALTASHOV, I.; ABZALIMOV, R. Do Ionic Charges in ESI MS Provide Useful Information on Macromolecular Structure? *J. Am. Soc. Mass Spectrom.* **2008**, *19* (9), 1239–1246. https://doi.org/10.1016/j.jasms.2008.05.018.

(86) Konermann, L.; Douglas, D. J. Acid-Induced Unfolding of Cytochrome c at Different Methanol Concentrations: Electrospray Ionization Mass Spectrometry Specifically Monitors Changes in the Tertiary Structure. *Biochemistry* **1997**, *36* (40), 12296–12302. https://doi.org/10.1021/bi971266u.

(87) Chorev, D. S.; Robinson, C. V. Protein Assemblies Ejected Directly from Native Membranes Yield Complexes for Mass Spectrometry. *Science* **2019**, *366* (6466), eaax3102. https://doi.org/10.1126/science.aax3102.

(88) Mitra, G. Application of Native Mass Spectrometry in Studying Intrinsically Disordered Proteins: A Special Focus on Neurodegenerative Diseases. *Biochim. Biophys. Acta - Proteins Proteomics* **2019**, *1867* (11), 140260. https://doi.org/10.1016/j.bbapap.2019.07.013.

(89) May, J. C.; McLean, J. A. Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Anal. Chem*. 2015. https://doi.org/10.1021/ac504720m.

(90) Mason, E. A.; Schamp, H. W. Mobility of Gaseous Lons in Weak Electric Fields. *Ann. Phys. (N. Y).* **1958**. https://doi.org/10.1016/0003-4916(58)90049-6.

(91) Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E. The Power of Ion Mobility-Mass Spectrometry for Structural Characterization and the Study of Conformational Dynamics. *Nat. Chem.* **2014**, *6* (4), 281–294. https://doi.org/10.1038/nchem.1889.

(92) Williams, D. M.; Pukala, T. L. Novel Insights into Protein Misfolding Diseases Revealed by Ion Mobility-Mass Spectrometry. *Mass Spectrom. Rev.* **2013**, *32* (3), 169–187. https://doi.org/10.1002/mas.21358.

(93) Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H. An Investigation of the Mobility Separation of Some Peptide and Protein Ions Using a New Hybrid Quadrupole/Travelling Wave IMS/Oa-ToF Instrument. *Int. J. Mass Spectrom.* **2007**, *261* (1), 1–12. https://doi.org/10.1016/j.ijms.2006.07.021.

(94) Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science* **2005**, *310* (5754), 1658–1661. https://doi.org/10.1126/science.1120177.

(95) Smith, D. P.; Radford, S. E.; Ashcroft, A. E. Elongated Oligomers in B2-Microglobulin Amyloid Assembly Revealed by Ion Mobility Spectrometry-Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (15), 6794–6798. https://doi.org/10.1073/pnas.0913046107.

(96) Österlund, N.; Moons, R.; Ilag, L. L.; Sobott, F.; Gräslund, A. Native Ion Mobility-Mass Spectrometry Reveals the Formation of β-Barrel Shaped Amyloid-β Hexamers in a Membrane-Mimicking Environment. *J. Am. Chem. Soc.* **2019**, *141* (26), 10440–10450. https://doi.org/10.1021/jacs.9b04596.

(97) Mehmood, S.; Marcoux, J.; Gault, J.; Quigley, A.; Michaelis, S.; Young, S. G.; Carpenter, E. P.; Robinson, C. V. Mass Spectrometry Captures Off-Target Drug Binding and Provides Mechanistic Insights into the Human Metalloprotease ZMPSTE24. *Nat. Chem.* **2016**, *8* (12), 1152–1158. https://doi.org/10.1038/nchem.2591.

(98) Back, J. W.; De Jong, L.; Muijsers, A. O.; De Koster, C. G. Chemical Cross-Linking and Mass Spectrometry for Protein Structural Modeling. *J. Mol. Biol.* **2003**, *331* (2), 303–313. https://doi.org/10.1016/S0022-2836(03)00721-6.

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

(100) O'Reilly, F. J.; Rappsilber, J. Cross-Linking Mass Spectrometry: Methods and Applications in Structural, Molecular and Systems Biology. *Nat. Struct. Mol. Biol.* **2018**, *25* (11), 1000–1008. https://doi.org/10.1038/s41594-018-0147-0.

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

(102) Yang, B.; Wu, Y.-J.; Zhu, M.; Fan, S.-B.; Lin, J.; Zhang, K.; Li, S.; Chi, H.; Li, Y.-X.; Chen, H.-F.; et al. Identification of Cross-Linked Peptides from Complex Samples. *Nat. Methods* **2012**, *9* (9), 904–906. https://doi.org/10.1038/nmeth.2099.

(103) Götze, M.; Pettelkau, J.; Fritzsche, R.; Ihling, C. H.; Schäfer, M.; Sinz, A. Automated Assignment of MS/MS Cleavable Cross-Links in Protein 3D-Structure Analysis. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (1), 83–97. https://doi.org/10.1007/s13361-014-1001-1.

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

(105) Plaschka, C.; Larivière, L.; Wenzeck, L.; Seizl, M.; Hemann, M.; Tegunov, D.; Petrotchenko, E. V.; Borchers, C. H.; Baumeister, W.; Herzog, F.; et al. Architecture of the RNA Polymerase II–Mediator Core Initiation Complex. *Nature* **2015**, *518* (7539), 376–380. https://doi.org/10.1038/nature14229.

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

(107) McDonald, A. J.; Leon, D. R.; Markham, K. A.; Wu, B.; Heckendorf, C. F.; Schilling, K.; Showalter, H. D.; Andrews, P. C.; McComb, M. E.; Pushie, M. J.; et al. Altered Domain Structure of the Prion Protein Caused by Cu2+ Binding and Functionally Relevant Mutations: Analysis by Cross-Linking, MS/MS, and NMR. *Structure* **2019**, *27* (6), 907-922.e5. https://doi.org/10.1016/j.str.2019.03.008.

Chapter 2: The molecular chaperone β-casein prevents amorphous and fibrillar aggregation of α-lactalbumin by stabilisation of dynamic disorder

Disclaimer and author contributions.

The data presented in this chapter has been peer reviewed and published in Biochemical Journal.

Sanders, H. M.; Jovcevski, B.; Carver, J. A.; Pukala, T. L. The Molecular Chaperone β-Casein Prevents Amorphous and Fibrillar Aggregation of α-Lactalbumin by Stabilisation of Dynamic Disorder. *Biochem. J.* 2020, *477* (3), 629–643. https://doi.org/10.1042/BCJ20190638.

The majority of the experiments, data analysis and manuscript writing were performed by Sanders, H. M.

Jovcevski, B., Carver, J. A. and Pukala, T. L. assisted with experimental design, data interpretation and editing of the manuscript.

Biochemical Journal (2020) **477** 629–643
https://doi.org/10.1042/BCJ20190638

PORTLAND PRESS

Research Article

The molecular chaperone β-casein prevents amorphous and fibrillar aggregation of α-lactalbumin by stabilisation of dynamic disorder

Henry M. Sanders[1], Blagojce Jovcevski[1], John A. Carver[2] and Tara L. Pukala[1]

[1]School of Physical Sciences, The University of Adelaide, Adelaide, SA 5005, Australia; [2]Research School of Chemistry, Australian National University, Acton, ACT 2601, Australia

Correspondence: Tara L. Pukala (tara.pukala@adelaide.edu.au)

Deficits in protein homeostasis (proteostasis) are typified by the partial unfolding or misfolding of native proteins leading to amorphous or fibrillar aggregation, events that have been closely associated with diseases including Alzheimer's and Parkinson's diseases. Molecular chaperones are intimately involved in maintaining proteostasis, and their mechanisms of action are in part dependent on the morphology of aggregation-prone proteins. This study utilised native ion mobility–mass spectrometry to provide molecular insights into the conformational properties and dynamics of a model protein, α-lactalbumin (α-LA), which aggregates in an amorphous or amyloid fibrillar manner controlled by appropriate selection of experimental conditions. The molecular chaperone β-casein (β-CN) is effective at inhibiting amorphous and fibrillar aggregation of α-LA at sub-stoichiometric ratios, with greater efficiency against fibril formation. Analytical size-exclusion chromatography demonstrates the interaction between β-CN and amorphously aggregating α-LA is stable, forming a soluble high molecular weight complex, whilst with fibril-forming α-LA the interaction is transient. Moreover, ion mobility–mass spectrometry (IM-MS) coupled with collision-induced unfolding (CIU) revealed that α-LA monomers undergo distinct conformational transitions during the initial stages of amorphous (order to disorder) and fibrillar (disorder to order) aggregation. The structural heterogeneity of monomeric α-LA during fibrillation is reduced in the presence of β-CN along with an enhancement in stability, which provides a potential means for preventing fibril formation. Together, this study demonstrates how IM-MS and CIU can investigate the unfolding of proteins as well as examine transient and dynamic protein–chaperone interactions, and thereby provides detailed insight into the mechanism of chaperone action and proteostasis mechanisms.

Introduction

Globular proteins rely on the formation of discrete, low-energy, three-dimensional native structures in order to carry out their biological function [1]. As newly synthesised proteins are produced, they undergo an 'on-pathway' folding process to the native state, involving condensation of hydrophobic residues into a non-solvent exposed core, while the remaining residues sample available structural conformations. The folding process is largely 'downhill', with successive structures being lower in energy with fewer degrees of freedom, thereby rapidly accelerating the process [1–3]. However, conditions of cellular stress can render proteins unable to adopt their native conformation, instead directing proteins towards 'off-pathway' folding mechanisms. Here, proteins populate partially unfolded, intermediate states which are prone to self-assembly through the formation of either amorphous aggregates or amyloid fibrils [4,5]. Amorphous aggregates form due to random hydrophobic associations between misfolded proteins that become insoluble after reaching a critical aggregate size [6]. In

Received: 2 September 2019
Revised: 10 January 2020
Accepted: 15 January 2020

Accepted Manuscript online: 15 January 2020
Version of Record published: 11 February 2020

Biochemical Journal (2020) 477 629–643
https://doi.org/10.1042/BCJ20190638

contrast, fibril formation results from a nucleated growth process whereby ordered monomeric units stack perpendicularly to the fibril axis, forming a cross-β-sheet structure. Multiple resultant 'protofibril' strands assemble to produce mature fibrils [7–9].

Molecular chaperones are a crucial constituent of the protein quality control network which regulates protein refolding, stabilisation, sequestration and degradation. Failure of this protein quality control network propagates the formation of misfolded and aggregation-prone proteins which are implicated in disease [10,11]. For example, Alzheimer's and Parkinson's diseases are two of the many disorders associated with the formation of amyloid fibrils [4]. Amorphous aggregates are considered non-toxic, with nuclear cataracts being the most notable amorphous aggregate-associated disease [12]. To date, there are no effective therapeutic approaches to treat neurodegenerative diseases in which misfolding proteins are implicated. Understanding distinctions between these off-folding pathways can allow for more effective therapeutic design, for example by redirecting misfolded proteins from a fibrillar to an amorphous pathway *in vivo*. Therefore, a robust model misfolding protein that is able to mimic physiologically relevant conditions and follow both amorphous and fibrillar aggregation pathways is a valuable tool for assessing potential inhibitors of protein aggregation.

Bovine α-lactalbumin (α-LA) is a 14.2 kDa acidic calcium-binding milk protein which plays an essential role in the biosynthesis of lactose within the mammary glands [13]. The protein consists of a large α-helical domain, a small β-sheet domain and a calcium-binding loop, a conformation which is stabilised by four disulfide bonds [14]. α-LA undergoes amorphous aggregation when reduced with dithiothreitol (DTT α-LA), but forms fibrillar aggregates when reduced and carboxymethylated (RCM α-LA) due to formation of a partially folded intermediate state [15–18]. As a result, α-LA is a useful system for investigating structural distinctions between amorphous and fibrillar aggregation, and identifying which conformations dictate these aggregation pathways. β-casein (β-CN) is an intrinsically disordered major milk protein [19] which has a monomer mass of ~24 kDa, and at temperatures above ~20°C, forms concentration-dependent oligomers known as micelles [20]. Like the other bovine caseins, β-CN exhibits chaperone activity to prevent the unfolding and amorphous aggregation of proteins under stress conditions such as elevated temperature [21]. β-CN also prevents amyloid fibril formation of κ-casein [22–24].

Difficulties associated with investigating the mechanisms of chaperone-mediated proteostasis and characterising the transient interactions between chaperones and aggregation-prone proteins mean that a combination of biophysical techniques is required to probe this process. Native mass spectrometry (MS) is well-suited for the analysis of protein–protein interactions, given the gentle nature and low sample requirements of nano-electrospray ionisation. MS enables the preservation and interrogation of low-abundant, heterogeneous assemblies formed during aggregation. When paired with ion mobility (IM), ion mobility–mass spectrometry (IM-MS) allows the size, stoichiometry and dynamics of protein assemblies to be studied. Moreover, the unfolding dynamics of proteins can be observed by gas-phase manipulation through collision-induced unfolding (CIU), enabling the measurement of protein stability and protein–protein interactions in the gas-phase, all of which are reflective of their solution-phase properties [25–29].

To understand the structural features that direct α-LA towards either the amorphous or fibrillar aggregation pathways, we utilised an integrated biophysical approach to observe the differences in monomeric α-LA structure and stability following chemical modification (i.e. reduction via DTT to give DTT α-LA compared with RCM α-LA). In addition, we have assessed the ability of β-CN to inhibit amorphous and fibrillar aggregation of DTT α-LA and RCM α-LA, respectively. In particular, IM-MS was implemented to probe changes in the conformation and unfolding dynamics of α-LA, and CIU was used to understand better the effects of β-CN in protecting against aggregation through stabilisation of α-LA monomers. Identifying distinct changes in conformation and dynamics that dictate which aggregation pathway proteins follow, as well as their interaction with molecular chaperones, provides a greater understanding of how chaperones arrest protein unfolding, misfolding and aggregation. Moreover, this study highlights how MS-based approaches can be utilised to probe transient and highly dynamic protein–protein interactions.

Experimental procedures

Bovine α-LA (type II, calcium-depleted, 85% pure, lyophilised), β-CN and all other reagents were purchased from Sigma–Aldrich (St. Louis, U.S.A.) unless stated otherwise. DTT was obtained from Astral Scientific (Gymea, Australia). Reduction and carboxymethylation (RCM) of α-LA was performed as described previously [30] and confirmed by observation of the associated mass increase using MS (Supplementary Figure S1). RCM

α-LA was buffer exchanged into ammonium acetate (100 mM) using an Amicon Ultra-4 centrifugal spin filter at 4°C. Samples were lyophilised and stored at −80°C until use.

Amorphous and fibrillar aggregation assays

The DTT-induced (2 mM final concentration) amorphous aggregation of α-LA (DTT α-LA; 100 μM) was monitored by the change in light scattering at 340 nm. Aggregation of α-LA was performed in 100 mM ammonium acetate (pH 7.0) in the absence and presence of β-CN at various molar ratios (1 : 1, 1 : 10 and 1 : 100, β-CN : DTT α-LA). Amorphous assays were plated (200 μl/well) into sealed 96-well microplates (Greiner Bio-One) and aggregation was monitored using a FLUOstar Optima microplate reader (BMG Lab Technologies, Melbourne, Australia) over a period of 6 h at 37°C, with an initial shaking period of 300 s. Fibrillar aggregation of RCM α-LA was monitored in real-time using a ThT fluorescence assay [15]. Fibrillation of RCM α-LA was performed in 100 mM ammonium acetate (pH 7.0) in the absence and presence of β-CN at various molar ratios (1 : 1, 1 : 2, 1 : 10, 1 : 20 and 1 : 100, β-CN : RCM α-LA). Fibrillar assays were plated (50 μl/well) into sealed 384-well microplates (Greiner Bio-One) and fibril formation was monitored using a FLUOstar Optima microplate reader (BMG Lab Technologies, Melbourne, Australia) over a period of 24 h at 37°C with shaking (double orbital, 50 rpm) for 30 s prior to fluorescence measurement. ThT fluorescence was measured using an excitation and emission wavelength of 440 and 490 nm, respectively. All assays were performed in triplicate and reported as mean ± SEM. The degree of inhibition (%) afforded by β-CN against amorphous and fibrillar aggregation was calculated as described previously [31].

Transmission electron microscopy

Three microliters of sample was taken immediately at the conclusion of the light scattering or ThT assay and placed directly onto a carbon-coated 400-mesh nickel TEM grid (ProSciTech, Thuringowa Central, Australia) [32]. Grids were washed with 0.22 μm filtered MilliQ water and stained with 2% (w/v) uranyl acetate solution. Samples were viewed using either a CM100 transmission electron microscope (Philips, Eindhoven, Netherlands) or a Tecnai G2 Spirit TEM (FEI, Oregon, U.S.A.) with a magnification of 45 000×.

Intrinsic tryptophan and bis-ANS fluorescence

Both intrinsic tryptophan and bis-ANS fluorescence studies were performed using a Cary Eclipse fluorescence spectrophotometer (Varian Inc., Mulgave, Australia) at room temperature. Excitation and emission slit widths were adjusted to 2.5 nm and scan speed was set at 60 nm/min. Native α-LA, DTT α-LA and RCM α-LA (100 μM) were prepared in 100 mM aqueous ammonium acetate (pH 7.0). Fluorescence was measured using a quartz cuvette of pathlength 1 cm. Intrinsic tryptophan fluorescence was measured using an excitation wavelength of 295 nm and emission was measured from 300 to 400 nm. To determine the exposed hydrophobicity of all α-LA forms, 4,4′-dianilino-1,1′-binaphthyl-5,5′-disulfonic acid (bis-ANS) (10 μM) was added and fluorescence emission was measured from 400 to 600 nm following excitation at 385 nm. The obtained spectra represent the average of at least 3 scans.

Analytical size-exclusion chromatography

The interaction between α-LA and β-CN was observed by SEC [33]. Samples (100 μM) following aggregation were centrifuged at 14 000×***g*** for 30 min at 4°C and subsequently loaded (500 μl) onto a Superdex 200 10/300 GL analytical-SEC (GE Healthcare, Illinois, U.S.A.) which was equilibrated with 100 mM ammonium acetate (pH 7.0) at a flow rate of 0.4 ml/min at room temperature. The SEC column was calibrated using standards (Sigma–Aldrich, Missouri, U.S.A.) containing bovine thyroglobulin (670 kDa), bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa) and horse myoglobin (17 kDa).

SDS–PAGE

Eluted fractions following analytical-SEC were further examined by SDS–PAGE (12% gels) (Bio-rad, California, U.S.A.) and bands were visualised using a Silver staining kit (Invitrogen, California, U.S.A.) for analytical-SEC fraction analysis and Coomassie Brilliant Blue stain (ThermoFisher, Massachusetts, U.S.A.) for band density analysis. Both techniques were performed according to the manufacturer's instructions. Band densities were calculated using GelAnalyzer 9.1 (http://www.gelanalyzer.com/).

Biochemical Journal (2020) 477 629–643
https://doi.org/10.1042/BCJ20190638

Ion mobility–mass spectrometry

The conformation of α-LA forms was investigated by IM-MS performed on a Synapt HDMS Q-TOF mass spectrometer (Waters Corporation, Manchester, U.K.) using a nano-electrospray ionisation source. Samples were prepared in 100 mM ammonium acetate (pH 7.0) to a final concentration of 25 μM. DTT α-LA was formed by the addition of DTT (2 mM) and RCM α-LA was incubated in the presence of β-CN (1 : 0.5 molar ratio; RCM α-LA : β-CN). Samples were loaded into platinum-coated borosilicate glass capillaries prepared in-house. Gentle source conditions were applied to minimise gas-phase structural changes prior to detection, with instrument parameters as follows: capillary voltage, 1.60 kV; sampling cone, 30 V; extraction cone, 1.5 V; trap/transfer collision energy, 10/15 V; trap gas, 5.5 l/h; backing gas, ∼4.5 mbar. The parameters for IM were as follows: IM cell wave height, 8 V; IM cell wave velocity, 350 m/s; transfer t-wave height, 8 V; transfer t-wave velocity, 250 m/s. Mass spectra and arrival time distributions (ATDs) were viewed using MassLynx (v4.1) and DriftScope (v2.1), respectively (Waters Corporation, Manchester, U.K.).

Collision-induced unfolding

The CIU dynamics of α-LA forms was investigated by observing the unfolding of the monomer^{7+} charge state (native α-LA at 2031 *m/z* and RCM α-LA at 2093 *m/z*) and IM spectra were collected under the conditions stated above. Unfolding of the monomer^{7+} ions was induced by increasing the trap collision energy (activation energy) by 2.5 V increments from 10 to 60 V. CIU heat maps were generated using CIUsuite with default parameters [34]. The proportion of unfolded α-LA with increasing activation energy was calculated by normalising the ATD to the highest intensity, taking the relative intensity of any unfolded species present as a function of folded α-LA [34,35]. A dose–response (sigmoidal) curve was applied to calculate the CIU_{50} (i.e. the voltage at which 50% of the monomer as unfolded) and the rate of unfolding (denoted by the Hill slope of the sigmoidal fit). The equation is shown below:

$$y = \text{bottom} + \frac{(\text{top} - \text{bottom})}{1 + 10^{(CIU_{50} - x) \times \text{rate of unfolding}}}$$

where bottom and top refer to the bottom and top of the plateau, i.e. the folded and unfolded states of α-LA. Curve fitting analyses were performed using Prism 7.0 (GraphPad Software, San Diego, U.S.A.) software.

Results

β-CN is more effective at inhibiting RCM α-LA fibril formation compared with DTT-induced amorphous aggregation

To evaluate the ability of β-CN to inhibit both amorphous and fibrillar aggregation of α-LA, chaperone assays were performed at various α-LA : β-CN molar ratios (Figure 1). In the absence of β-CN, insoluble amorphous aggregates of α-LA (100 μM) were induced upon addition of DTT (2 mM) in ammonium acetate (100 mM, pH 7.0), whereupon an increase in light scattering occurred at ∼0.75 h and reached a plateau indicative of amorphous aggregate formation after ∼2.5 h (Figure 1A, black), which was confirmed by TEM (Figure 1A, top inset). The lag-phase for DTT α-LA aggregation was unchanged in the presence of β-CN across all molar ratios (ranging from 1 : 1 to 100 : 1, DTT α-LA : β-CN), however, the subsequent rate and degree of light scattering decreased in a β-CN concentration-dependent manner (Figure 1C). β-CN exhibited intermediate ability at inhibiting the amorphous aggregation of α-LA with ∼60% inhibition at a 1 : 1 ratio (DTT α-LA : β-CN) and no inhibition at 100 : 1 (Figure 1C). In addition, the activity of β-CN was not affected by the presence of DTT as no increase in light scattering was observed for the β-CN only control, consistent with its absence of disulfide bonds (Supplementary Figure S2).

The ability of β-CN to inhibit α-LA fibril formation was assessed using a ThT aggregation assay (Figure 1B,D). In the absence of β-CN, RCM α-LA (200 μM) forms amyloid fibrils in ammonium acetate (100 mM, pH 7.0) as evidenced by an increase in ThT fluorescence from ∼7.5 h, indicative of fibril elongation and maturation without the need for $MgCl_2$, which previous studies (in phosphate buffer) have used to promote fibril formation [15]. In addition, the formation of mature RCM α-LA fibrils was confirmed by TEM (Figure 1C, top inset). The efficiency of β-CN as a fibril inhibitor was apparent from these assays, whereby 1 : 1 and 100 : 1 molar ratios (RCM α-LA : β-CN) exhibited near total inhibition and ∼33% inhibition, respectively (Figure 1D). In addition, near total inhibition was observed at a 2 : 1 molar ratio (RCM α-LA : β-CN) (Supplementary

Biochemical Journal (2020) **477** 629–643
https://doi.org/10.1042/BCJ20190638

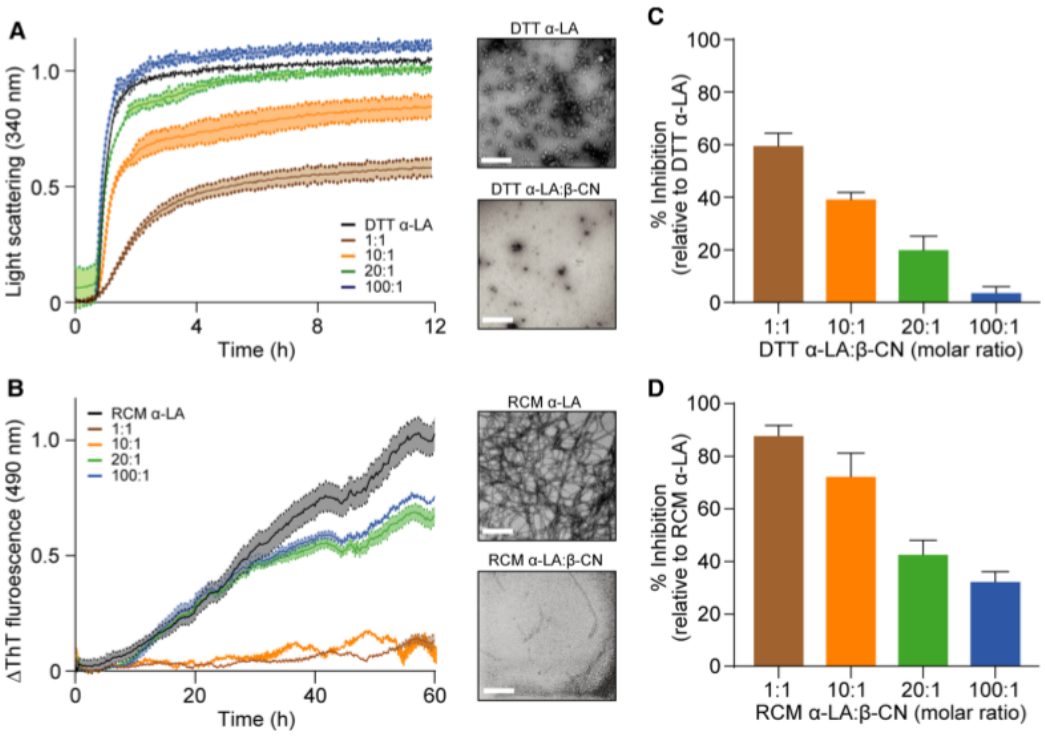


Figure 1. β-CN inhibits the amorphous and fibrillar aggregation of α-LA.

(**A**) DTT α-LA (100 μM) amorphous aggregation was monitored by the change in light scattering at 340 nm. Aggregation was induced with DTT (2 mM) in ammonium acetate (100 mM, pH 7.0) in the presence of various β-CN molar ratios. TEM image (**A**, inset) displays the morphology of amorphous aggregates in the presence (1 : 1 molar ratio) and the absence of β-CN. (**B**) Fibril formation of RCM α-LA was monitored by the change in ThT fluorescence at 490 nm (excitation at 440 nm) at various molar ratios of β-CN to determine chaperone activity. A corresponding TEM image shows the morphology of RCM α-LA fibrils in the presence (1 : 1 molar ratio) and the absence of β-CN (B, inset). (**C**) The concentration-dependent chaperone activity of β-CN against amorphous aggregation was calculated by comparing the degree of light scattering at the conclusion of the assay. (**D**) The concentration-dependent chaperone activity of β-CN against fibril formation was calculated by comparing the degree of ThT fluorescence at the conclusion of the assay. Data are reported as mean ± SEM ($n = 3$). Scale bars represent 400 μm.

Figure S2). To ensure the ThT data accurately reflects the fibril yields, the relative amount of soluble material was assessed following the 72-h aggregation period using SDS–PAGE (Supplementary Figure S3). A clear decrease in monomeric RCM α-LA is observed following fibril formation, which is restored to a significant degree in the presence of β-CN. Overall, the data demonstrate the potent ability of β-CN to prevent α-LA aggregation, with much greater efficiency towards inhibiting fibril formation.

Previous work has suggested the major difference between the two α-LA states is due to carboxymethylation of α-LA, which adds additional negative charge to the protein and, because of repulsive interactions, induces the protein to adopt a more extended and flexible state capable of forming structural motif(s) that are pre-requisite for fibril formation [16]. In contrast, reduced (DTT) α-LA adopts more rigid structures which lead to the formation of amorphous aggregates. Aggregation kinetics of the α-LA species is also a likely factor. Amyloid fibril formation is favoured over amorphous aggregation by slowly aggregating species, and Figures 1A,B, along with work by Kulig and Ecroyd [15], show that reduced (amorphously aggregating) α-LA aggregates at an earlier time point than (amyloid fibril-forming) RCM α-LA, i.e. the former has a shorter lag-phase.

Structural modifications of monomeric α-LA dictate its interaction with β-CN to prevent aggregation

To further probe the difference in the ability of β-CN to prevent amorphous and fibrillar aggregation of α-LA, a better understanding of the structural differences between monomeric α-LA forms is required. Intrinsic

Biochemical Journal (2020) 477 629–643
https://doi.org/10.1042/BCJ20190638

tryptophan and extrinsic bis-ANS fluorescence studies were employed to probe the tertiary structure and exposed hydrophobicity of aggregation-prone forms of α-LA. RCM α-LA demonstrated a 250% increase in tryptophan fluorescence compared with native α-LA, whilst DTT α-LA displayed only a 60% increase in fluorescence (Figure 2A). In addition, RCM α-LA displayed a large fluorescence red-shift (29 nm shift to 354 nm) compared with native α-LA (absorbance maxima at 325 nm) which was less prevalent for DTT α-LA with only a 17 nm red-shift to 342 nm (Figure 2A). In contrast, the dye bis-ANS, which monitors clustered exposed hydrophobicity of a protein, showed a fluorescence increase in 85% for DTT α-LA compared with RCM α-LA which only exhibited a 20% increase in fluorescence (Figure 2B). A control sample was also analysed to ensure that the excess DTT did not affect tryptophan fluorescence (Supplementary Figure S4). In contrast, the hydrophobic residues of DTT α-LA are more exposed to solvent compared with RCM α-LA and the tryptophan residues are more shielded, which is consistent with the protein adopting an intermediately folded, 'molten globule' state as previously reported [15,17,36].

From these fluorescence experiments, it is concluded that overall RCM α-LA exposes more of its four tryptophan residues to solvent due to unfolding than DTT α-LA whilst shielding hydrophobic residues. The observation in Figure 2B that DTT α-LA has greater ANS fluorescence than RCM α-LA is consistent with the general features adopted by both these forms of α-LA, i.e. RCM α-LA has a relatively extended (unfolded) conformation compared with DTT α-LA which adopts a more compact, molten globule structure that exposes the protein's hydrophobic core to solution, and therefore binds ANS readily. Amorphous aggregates are formed much faster than amyloid fibrils and DTT α-LA exposes more hydrophobicity. RCM α-LA exposes less hydrophobicity, despite being less ordered than native α-LA. This may be important for the prerequisites for following a fibril formation pathway rather than amorphous aggregation pathway.

Analytical-SEC was performed in order to determine the nature of the interaction between α-LA and β-CN under the same conditions used for both aggregation assays and MS (Figure 2C,D). In their native states, β-CN and α-LA each eluted as a single peak of ∼30 kDa in mass (13.5 ml) and 15 kDa in mass (17.2 ml), respectively (Figure 2C,D), consistent with their monomeric state. Prior to aggregation, DTT α-LA eluted as two, broad overlapping peaks (15.0 and 17.2 ml) indicative of monomeric α-LA and potentially low-molecular mass aggregates (e.g. dimer) (Figure 2C, purple). Importantly, the amount of soluble DTT α-LA decreased after aggregation with a small peak at ∼670 kDa indicative of aggregated α-LA (Figure 2C, red). In the presence of an equimolar ratio of β-CN, the intensity of the peak ∼8 ml (∼670 kDa) increased drastically corresponding to a high molecular weight (HMW) complex (Figure 2C, blue) containing both α-LA and β-CN, as determined by SDS–PAGE (Figure 2E). In the absence of β-CN, RCM α-LA eluted as a single peak at ∼15.8 ml before (Figure 2D, purple) and after (Figure 2D, red) incubation. In contrast with DTT α-LA, RCM α-LA in the presence of β-CN does not form a HMW complex as no peak was observed at earlier elution volumes (Figure 2D, blue). The data suggest that the interaction between DTT α-LA and β-CN leads to stable complex formation between the two proteins in order to prevent amorphous aggregation, whilst β-CN interacts with RCM α-LA transiently in order to inhibit fibril formation.

IM-MS reveals conformational transitions of monomeric DTT and RCM α-LA during early stages of aggregation

Native IM-MS allows for the gentle separation of proteins in the gas-phase based on both mass to charge ratio (m/z) and rotationally averaged size (or collisional cross section), the latter of which can be monitored as a function of changes in solution properties or gas-phase activation and unfolding, providing conformational insight (Supplementary Figure S5). IM-MS was utilised here to observe conformational transitions of DTT and RCM α-LA monomers during their initial stages of aggregation. The most prominent peak for both DTT and RCM α-LA was the monomer^{7+} charge state (M^{7+}) at m/z 2031 (Figure 3A, red box). From this charge state, we extracted the ATD, providing information on the mobility of the ion through a buffer gas, which is dependent upon the protein's three-dimensional structure. The ATD of the M^{7+} for DTT and RCM α-LA was analysed during the lag-phase of aggregation. Initially, the ATD of RCM α-LA M^{7+} exhibits two populations (at 5.8 and 8.1 ms) indicative of a predominantly compact (Figure 3B, blue box) and an unfolded state, respectively (Figure 3B, green box). The presence of the disordered state diminished during incubation, eventually disappearing after 2 h (Figure 3B). This trend was present for every major charge state, showing either an increase in compact structural populations, or a decrease in less compact populations (Supplementary Figure S6). In contrast, the presence of the unfolded state became more apparent in DTT α-LA during incubation after 1 h,

Biochemical Journal (2020) **477** 629–643
https://doi.org/10.1042/BCJ20190638

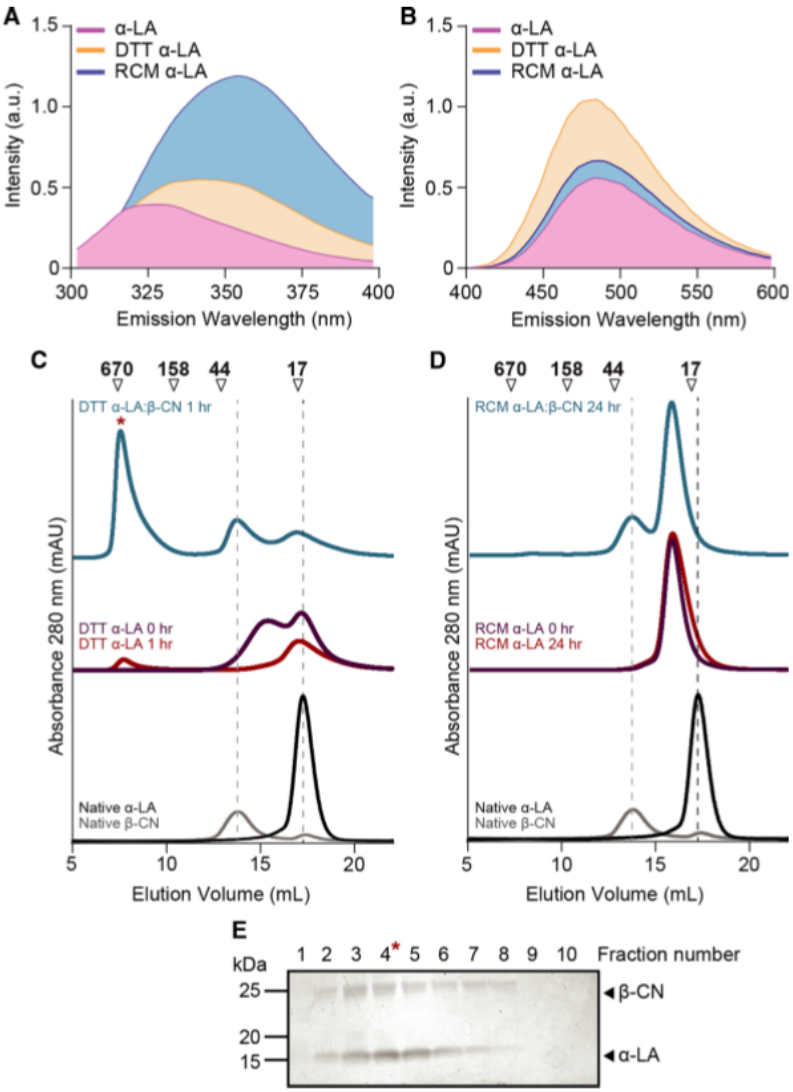


Figure 2. Structural modification of α-LA influences its interaction with β-CN.

(**A**) Intrinsic tryptophan fluorescence spectra of α-LA forms (100 μM) in 100 mM ammonium acetate (pH 7.0), exhibiting an increase in tryptophan exposure for DTT- and RCM α-LA relative to native α-LA. (**B**) bis-ANS (10 μM) fluorescence spectra measuring the clustered, exposed hydrophobicity of α-LA forms demonstrates an increase in exposed hydrophobicity for modified α-LA compared with native α-LA. (**C**) Analytical-SEC profiles of native α-LA (100 μM) and β-CN (100 μM). DTT α-LA (100 μM) at 0 h (purple) and 1 h (red) post-incubation at 37°C was compared with DTT α-LA in the presence of β-CN (100 μM, blue) forming a HMW complex (asterisk). Analytical-SEC experiments were performed in 100 mM ammonium acetate (pH 7.0). (**D**) Analytical-SEC profiles RCM α-LA (100 μM) at 0 h (purple), 24 h (red) post-incubation at 37°C was compared with RCM α-LA in the presence of β-CN (100 μM, blue). SEC experiments were performed in 100 mM ammonium acetate (pH 7.0) and elution volumes of molecular mass standards (kDa) are indicated above the chromatogram. (**E**) SDS–PAGE and silver staining of eluate fractions across the HMW peak from analytical-SEC of DTT α-LA in the presence of β-CN after 1 h of incubation. The asterisk indicates the fraction number (4) corresponding to the apex of the HMW peak.

despite hardly being present initially (Figure 3C). The ATDs of M^{7+} DTT and RCM α-LA were overlayed with the ATDs of their respective monomers at high and low activation energies (Supplementary Figure S7) to highlight that they are unfolded forms of the same species. RCM α-LA also showed a reduction in highly charged species indicating a compaction of structure, shielding residues from ionisation [37]. After incubation, DTT α-LA showed an increase in highly charged species indicative of unfolding (Supplementary Figure S8). Overall,

Biochemical Journal (2020) **477** 629–643
https://doi.org/10.1042/BCJ20190638

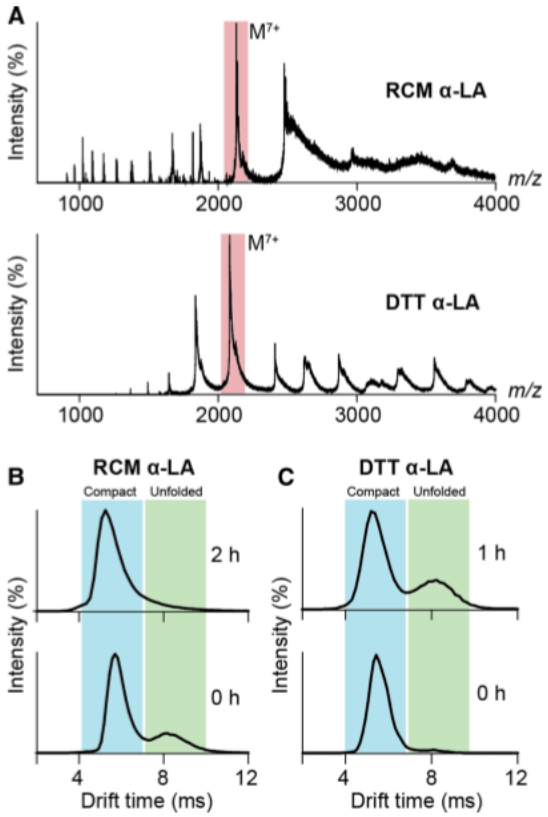


Figure 3. DTT and RCM α-LA undergo contrasting conformational transitions during the early stages of aggregation.
(**A**) Native MS of RCM (top panel) and DTT α-LA (bottom panel) (both 25 μM) at 0 h with the M^{7+} (*m/z* 2031) being the most abundant ion and selected for ATD analysis (red box). ATD of RCM (**B**) and DTT α-LA (**C**) M^{7+} during incubation at 37°C with shaking (300 rpm). The emergence of a disordered (unfolded) state (green box) during incubation of DTT α-LA contrasts to the disorder to order transition of RCM α-LA monomers during aggregation.

the data demonstrate that α-LA monomers follow disorder to order transitions (or vice versa) during initial stages of aggregation that is dependent on the aggregation pathway (i.e. amorphous or amyloid fibrillar).

IM-MS reveals differences in the conformation and unfolding propensity of RCM α-LA compared with native α-LA

The conformation and gas-phase unfolding dynamics of native and RCM α-LA in the absence and presence of β-CN were further examined by native IM-MS. The oligomeric state of native α-LA is predominantly monomeric with the M^{7+} charge state most prominent (Figure 4A, red box), as well as a population of dimer (D) (Figure 4A, top). RCM α-LA is also monomeric with a proportion of highly charged, unfolded monomer present (due to the increased exposure of ionisable residues as a result of its unfolded nature [37]) (Figure 4A, green box). A similar mass spectrum was also observed for RCM α-LA in the presence of β-CN, with both folded and unfolded monomeric RCM α-LA species present as well as monomeric β-CN (Figure 4A, bottom). The large proportion of highly charged monomeric signals indicative of unfolded RCM α-LA and lack of signals from oligomeric species evident in these spectra support the idea that the larger structures observed for RCM α-LA by SEC (Figure 2D) are extended, non-native states rather than higher-order aggregates.

To qualitatively probe the conformation and unfolding dynamics of α-LA, the ATD of the α-LA M^{7+} ions at 2031 *m/z* was monitored at low and high activation energy in the trap collision cell, allowing for resolution of folded (10–15 V) and unfolded (35–40 V) states (Figure 4B). Aside from being the most prominent charge state, the M^{7+} was chosen for IM-MS analyses in order to minimise charge state overlap from larger species (e.g. dimers) and was confirmed to be appropriate according to a power calculation which identifies suitable charge states for analysis [38]. The ATD for each folded and unfolded state also reflects the degree of heterogeneity within each

Biochemical Journal (2020) **477** 629–643
https://doi.org/10.1042/BCJ20190638

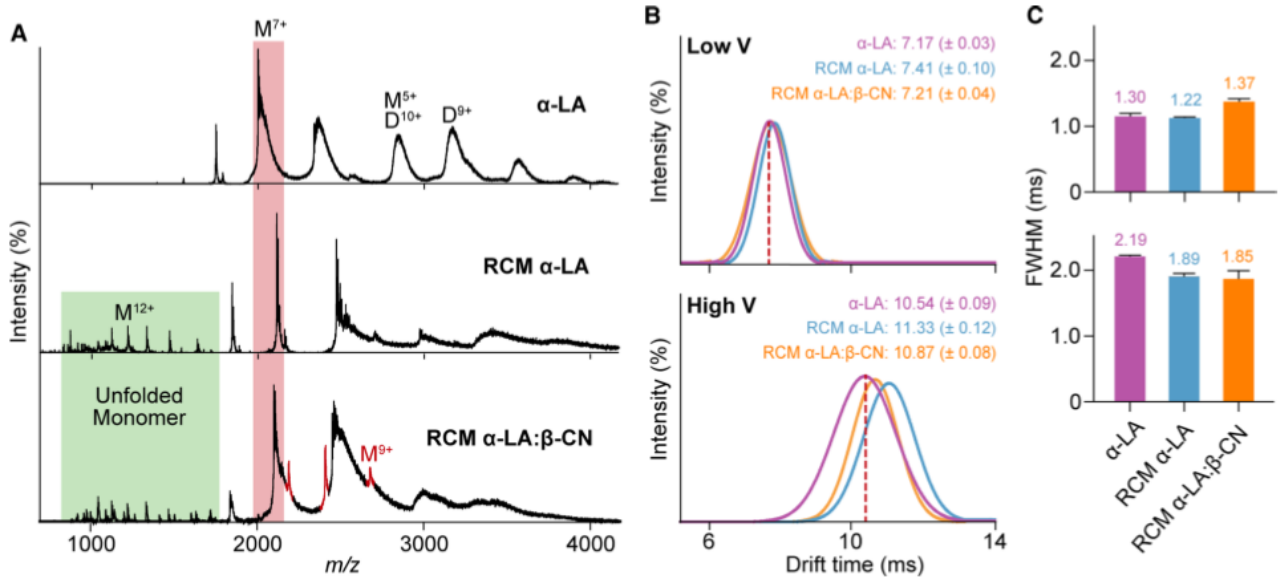


Figure 4. IM-MS analysis of native and RCM α-LA reveals differences in structure and dynamics with β-CN.

(**A**) Native MS of α-LA (top panel), RCM α-LA in the absence (middle panel) and presence (bottom panel) of β-CN (red) at a molar ratio of 2 : 1 (RCM α-LA : β-CN) after 15 min (final α-LA concentration, 25 μM). MS reveals a slight shift in mass due to the RCM of α-LA as well as the presence of unfolded monomer (green box). The M^{7+} (orange box) was selected for IM-MS and subsequent CIU analysis. (**B**) ATDs of the M^{7+} of α-LA forms under low (10–15 V, top panel) and high (35–40 V, bottom panel) activation energy. (**C**) FWHM time for α-LA M^{7+} ATDs at low (10 V, top panel) and high (40 V, bottom panel) activation energy. States (monomer: M, dimer: D) are indicated with charge state in superscript. Data in B and C are reported as mean ± SEM ($n = 3$).

population as represented by the full-width half maximum (FWHM) (Figure 4C). A larger FWHM indicates a more dynamic state that is sampling more conformations across a broader ATD, whereas a small FWHM indicates the ions adopt a relatively limited set of conformations. It should be noted that DTT α-LA was not similarly amenable to CIU studies due to the short time scale of amorphous aggregation preventing its analysis [17].

Under low activation energy, the ATD of all α-LA forms examined (native α-LA, RCM α-LA and RCM α-LA : β-CN) do not differ, with the exception that β-CN induced a slight broadening in the ATD for RCM α-LA compared with native and RCM α-LA alone (Figure 4B,C, top). Under high activation energy, however, changes in the ATD amongst α-LA forms were far more apparent, whereby RCM α-LA adopts a larger size compared with native α-LA as evidenced by a shift in the centroid of ATD to longer drift times (Figure 4B, bottom). Interestingly, this unfolding is somewhat attenuated in the presence of β-CN (Figure 4B, bottom). In addition, the conformational heterogeneity of RCM α-LA in the presence and absence of β-CN is reduced compared with native α-LA under high activation energy (Figure 4C, bottom). The data demonstrate that RCM α-LA unfolds more readily than native α-LA and that the presence of β-CN alters the structure of RCM α-LA, making it more native-like during activation.

β-CN enhances the stability of RCM α-LA monomers

We sought to quantitatively measure the stability RCM α-LA in the absence and presence of β-CN in the gas-phase by performing CIU IM-MS. Increasing activation (at increments of 2.5 V in the collision cell) of specific ions gradually shifted the ATD, allowing the stability and interconversion of individual species (i.e. folded to unfolded conformers) to be monitored (Figure 5A). Across all samples examined only two distinct populations were observed, even following extremely high activation up to 150 V, corresponding to folded and unfolded conformations, with a single unfolding transition and no distinct intermediates identified (Figure 5). Native α-LA required the most activation energy to unfold (Figure 5B, top panel). RCM α-LA required less activation energy to induce unfolding (Figure 5B, middle panel). The difference is presumably due to native α-LA having intact disulfide bonds unlike RCM α-LA and possibly the destabilising effect of eight additional negative charges as a result of carboxymethylation. Interestingly, RCM α-LA required a higher activation energy to

Biochemical Journal (2020) **477** 629–643
https://doi.org/10.1042/BCJ20190638

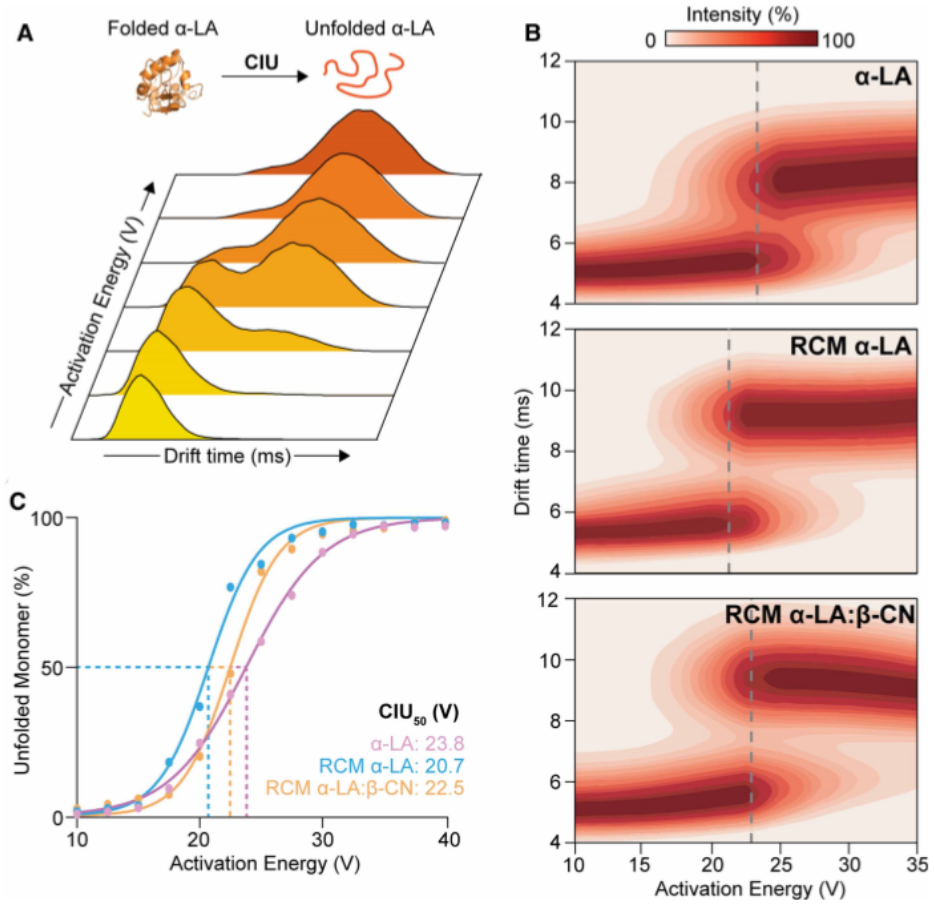


Figure 5. CIU analysis reveals β-CN enhances the stability of RCM α-LA.

(**A**) Schematic representation of the unfolding dynamics of native α-LA, demonstrating that increasing the gas-phase activation energy promotes the unfolding of α-LA as monitored for the α-LA M^{7+} ions. The gas-phase unfolding pathway comprised only two distinct populations (folded and unfolded) between 10 and 40 V. No additional transitions were observed with activation up to 150 V. (**B**) CIU heat maps of native α-LA M^{7+} (top panel), RCM α-LA in the absence (middle panel) and presence (bottom panel) of β-CN at a ratio of 2 : 1 (RCM α-LA : β-CN) with a 15 min incubation at 37°C. The single unfolding transition is indicated by grey dashed lines. (**C**) The abundance of unfolded α-LA M^{7+} was plotted as a function of activation energy for native α-LA (purple), RCM α-LA (blue) and RCM α-LA in the presence of β-CN (orange) at a ratio of 2 : 1 (RCM α-LA : β-CN) with a 15 min incubation in order to assess the stability. Data were fitted with a sigmoidal function and the stability was quantified by calculating the CIU_{50} (i.e. activation energy at which the abundance of unfolded monomer was 50%, dashed lines).

induce unfolding when in the presence of β-CN compared with RCM α-LA alone (Figure 5B, bottom panel). To quantitatively assess the stability of the various forms of α-LA, the relative populations of folded and unfolded α-LA were plotted as a function of activation energy and the energy required for 50% of α-LA monomer to be unfolded (CIU_{50}) was calculated as a measure of stability (Figure 5C). Native α-LA had a CIU_{50} of 23.8 V whereas RCM α-LA had a lower CIU_{50} of 20.7 V (Figure 5C). Moreover, native α-LA unfolded at a slower rate (0.14 ms/V) than RCM α-LA (0.21 ms/V) (Figure 5C). Interestingly, the presence of β-CN increased the CIU_{50} of RCM α-LA to 22.5 V and restored the rate of unfolding of RCM α-LA to that of RCM α-LA alone (Figure 5C). The data indicate that β-CN interacts with the monomer of RCM α-LA and enhances its stability as a means of preventing amyloid fibril formation.

Discussion

Improved understanding of the structural characteristics which direct proteins along unfolding pathways can inform approaches towards the treatment of protein misfolding diseases. Furthermore, understanding the

molecular basis that underlies molecular chaperone activity to prevent protein unfolding and aggregation can provide additional avenues for therapeutic intervention. Herein, we investigated the effects of the molecular chaperone β-CN on both amorphous and amyloid fibrillar aggregation of α-LA as representative examples of protein unfolding and its mitigation. In a broader context, the study also has relevance to the dairy industry as the chaperone action of β-CN is important in stabilising other milk proteins such as α-LA, for example under ultra-high temperature and pasteurisation processing [22,39].

β-CN inhibits the fibrillar aggregation of RCM α-LA more effectively than DTT α-LA amorphous aggregation (Figure 1B,D) although it has no observable effect on the lag-phase of α-LA aggregation in each case. While β-CN has been previously shown to inhibit both amorphous and fibrillar aggregation of a range of proteins, this is the first instance in which the chaperone activity of β-CN has been assessed against a single, aggregation-prone protein that can form either amorphous or fibrillar aggregates under physiologically relevant conditions (i.e. pH near 7.0 and 37°C) and without the need for harsh conditions or treatments that are not amenable to native MS (typically temperature >45°C, very acidic pH or using organic solvents). In general, molecular chaperones are classified into two classes, namely stabilising 'holdases' [40] and energy dependant 'foldases'. In agreement with previous studies [21–24,41,42], our data are consistent with β-CN functioning as a holdase to prevent protein aggregation without refolding aggregation-prone proteins to their native state.

Intrinsic tryptophan fluorescence and extrinsic bis-ANS fluorescence studies provided insights into the structural differences between α-LA forms (Figure 2A,B). Greater exposed hydrophobicity but near-native levels of tryptophan fluorescence exhibited by DTT α-LA implied a molten globule-like state in which secondary structure is mainly present while a dynamic overall structure exposes hydrophobic regions such as the protein's core. Conversely, increased tryptophan fluorescence for RCM α-LA indicated a lack of native structure with little exposed, clustered hydrophobicity, consistent with a non-native, relatively unfolded but stable structure as was previously observed by far-UV circular dichroism spectroscopy [15]. Therefore, a relative lack of exposed hydrophobicity for RCM α-LA may be important to promote fibril formation rather than amorphous aggregation, particularly as amorphous aggregates formed at a much faster rate than amyloid fibrils (compare Figure 1A,B).

IM-MS revealed distinct changes in the proportion of a disordered monomeric state of DTT α-LA and RCM α-LA with time (Figure 3), which may explain differences in the unfolding and aggregation pathways for each α-LA form. For RCM α-LA, the disappearance of a disordered species after 2 h of incubation (Figure 3B) leads to the adoption of a more ordered, compact intermediate conformation that is relatively long lived. It is likely this structure does not expose enough hydrophobicity to self-assemble amorphously but, over time, associates in an ordered, fibrillar form. The phenomenon of intrinsically disordered proteins reorganising and gaining structure prior to fibril formation has been observed previously and may be a key step in the formation of amyloid fibrils [43,44]. The opposite occurs for DTT α-LA. The Cys6–Cys120 disulfide bond of native α-LA is highly accessible to solvent and is reduced within 1 s upon the addition of DTT to give the triply disulfide-bonded form of α-LA which is relatively stable and ordered in conformation [17,45]. Complete reduction in the remaining disulfide bonds occurs after ∼5 min of incubation with DTT leading to a disordered molten globule conformation of α-LA that aggregates amorphously [17,45]. Analysis by NMR spectroscopy [17,45] reveals freshly incubated RCM α-LA adopts a similar disordered molten globule conformation to fully DTT-reduced α-LA (i.e. prior to its aggregation), consistent with the IM-MS results in Figure 3.

While β-CN forms complexes with other amorphously aggregating proteins to prevent their aggregation (41), β-CN more effectively protects against fibrillar RCM α-LA (Figure 1). Notably, this occurs without HMW complexation and despite the reduced hydrophobicity of RCM α-LA (Figure 2). *In vivo*, β-CN associated with the other caseins (of which there are three in bovine milk) to form the large amorphous aggregate known as the casein micelle [22–24]. Consistent with their intrinsically disordered nature, β-CN and the other caseins adopt a significant proportion of polyproline-II (PP-II) helical secondary structure, mostly within their proline, glutamine-rich regions [22–24]. These regions are involved in casein–casein interactions within the casein micelle and with other proteins during chaperone action [22–24]. Thus, hydrophobicity does not play a large role in protein interactions involving caseins [24], which may contribute to the greater efficiency of β-CN to inhibit α-LA fibril formation compared with α-LA amorphous aggregation, since the latter is dominated by hydrophobic association. From Raman spectroscopic studies, the molten globule state of α-LA at acidic pH adopts significant PP-II helical structure which most likely arises from unfolding of its β-sheet domain [46]. Furthermore, NMR spectroscopy shows that the molten globule state of α-LA present at acidic pH has a very similar conformation to that of RCM α-LA and fully DTT α-LA [45]. The presence of PP-II structure in RCM

Biochemical Journal (2020) 477 629–643
https://doi.org/10.1042/BCJ20190638

α-LA and β-CN, coupled with the prevalence of the PP-II conformation in protein–protein interactions [47], would predispose the PP-II structured regions in both proteins to co-associate, albeit transiently, and hence prevent the conversion of RCM α-LA to the amyloid fibrillar state.

β-CN is also known to undergo post-translational modifications that affect its chaperone activity. It is phosphorylated at five serine residues which form sites for binding calcium phosphate via nanocluster formation [22,41,48,49]. Dephosphorylation of β-CN reduces its chaperone activity against amorphous aggregation of ovotransferrin and apo α-LA, whilst remaining effective against fibril formation of RCM κ-casein [48], implying that electrostatic interactions play some part in β-CN chaperone action against amorphously aggregating proteins but are not involved in regulating chaperone action against fibril-forming proteins. β-CN exhibits very similar chaperone action to αs-casein [20] and sHsps. All are much better chaperones in preventing the aggregation of slowly aggregating proteins [20,50]. Thus, kinetics may also be a factor in the greater chaperone activity of β-CN in preventing RCM α-LA from aggregating, since DTT α-LA aggregates at a much faster rate than RCM α-LA (Figure 1).

We have investigated the heterogeneity, unfolding dynamics and stability of monomeric α-LA using IM-MS. Our data show that RCM α-LA is relatively more unfolded and structurally less stable than the native form (Figures 4 and 5). Interestingly, native and RCM α-LA showed similar ATD features (i.e. centroid and FWHM) (Figure 4B, top panel), suggesting that RCM α-LA exists in a relatively stable off-pathway structure with heterogeneity comparable to that of native α-LA. Upon unfolding (Figure 4B,C, bottom panel), native α-LA showed greater heterogeneity than RCM α-LA, likely due to RCM α-LA already existing in a partially disordered conformation and thereby having fewer conformational options available, whereas native α-LA can exist in all conformations between folded and disordered.

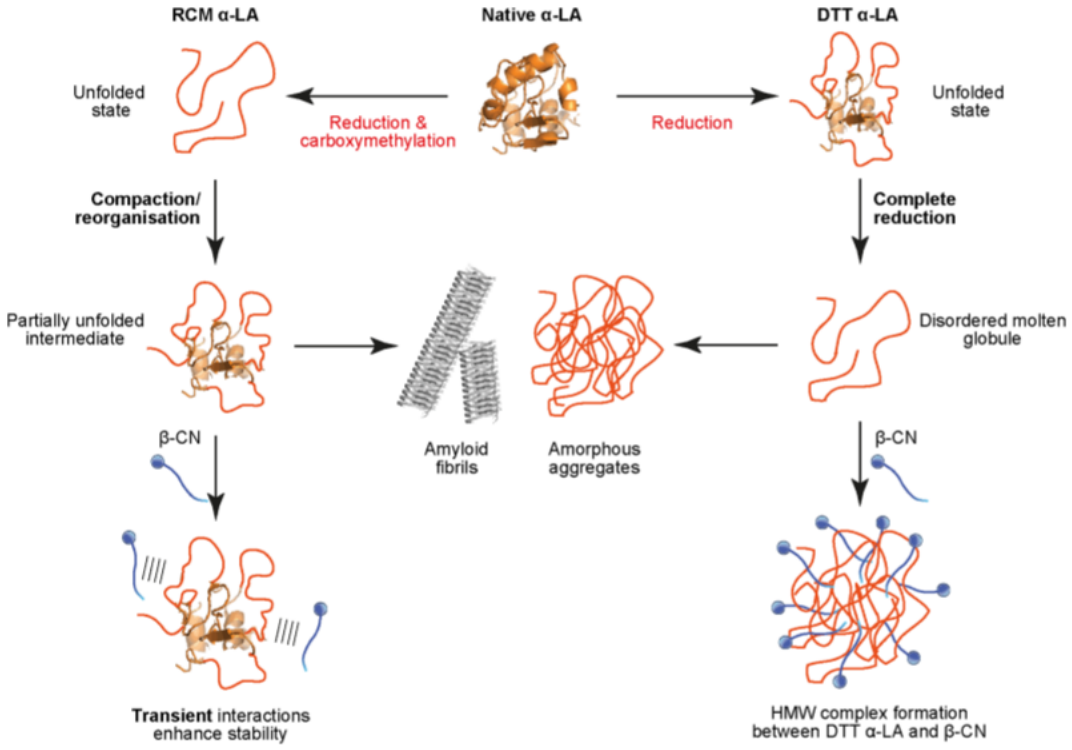


Figure 6. β-CN inhibits the aggregation of α-LA through two distinct mechanisms.

The sequential reduction in α-LA disulfide bonds initiates unfolding and induces an increase in exposed hydrophobicity and self-association to promote amorphous aggregation. The addition of β-CN sequesters disordered, hydrophobic α-LA into a HMW complex to prevent amorphous aggregation. Conversely, the complete reduction and carboxymethylation of α-LA results in the monomer condensing into a semi-stable intermediate with little exposed hydrophobicity which can form amyloid fibrils over time. In this scenario, β-CN interacts transiently with RCM α-LA intermediates to improve monomeric stability and inhibit amyloid fibril formation.

Structural changes from transition states correlate well with changes observed by CIU [38]. We showed that all α-LA forms unfold in a simple two-step mechanism. CIU analysis (Figure 5) also showed that RCM α-LA was more prone to unfolding as all its disulfide bonds are reduced, weakening its tertiary structure. However, the enhanced stability afforded by β-CN reduces the propensity of RCM α-LA to unfold. It appears, therefore, that β-CN inhibits fibril formation via a transient interaction (as confirmed by analytical-SEC, Figure 2) which induces conformational change in RCM α-LA promoting a more stable, native-like structure which is less prone to aggregation. Consistent with this, NMR studies also observed transient interactions between α-crystallin and RCM α-LA [45], whereas stabilisation and eventual complexation occurred between the disordered molten globule state of amorphously aggregating α-LA [17,45]. αs-Casein also prevents the amorphous aggregation of DTT α-LA by complexation, as monitored by NMR spectroscopy [21]. Figure 6 provides a schematic representation of the different mechanisms of β-CN chaperone action with amorphous and amyloid fibrillar aggregating α-LA as elucidated in this study. The mechanisms bear significant similarity to those for the inhibition of amorphous and amyloid fibrillar aggregation by sHsps [51–54].

In summary, DTT and RCM α-LA monomers have different structural properties apparent during the early stages of aggregation, with RCM α-LA displaying compaction to a less hydrophobic, stable intermediate whereas DTT α-LA rapidly unfolds to a molten globule state that exposes hydrophobicity and promotes the association with additional DTT α-LA monomers. These data are consistent with the notion that amorphous aggregation depends on exposing hydrophobicity whilst fibrillar aggregation relies on the formation of a more ordered monomeric intermediate with less exposed hydrophobicity that associates more slowly. While β-CN is able to protect against the amorphous aggregation of DTT α-LA, probably via hydrophobic and electrostatic interactions, a different and more efficient mechanism prevents fibril formation of RCM α-LA which may involve the mutual interaction of PP-II helical regions in both proteins. In doing so, β-CN interacts with RCM α-LA transiently to increase its stability, restricting its capacity to enter the amyloid fibril off-folding pathway (Figure 6). In a general sense, this study has provided insights into the conformational properties which induce amorphous or fibrillar aggregation and the mechanism of chaperone function in proteostasis. Finally, the use of IM-MS and CIU to investigate unfolding protein–chaperone interactions has been demonstrated, affording unique molecular insights compared with other biophysical techniques.

Abbreviations

β-CN, β-casein; ANS, 8-anilinonaphthalene-1-sulfonic acid; ATD, arrival time distribution; CIU, collision-induced dissociation; DTT, dithiothreitol; FWHM, full-width half maximum; HMW, high molecular weight; IM, ion mobility; MS, mass spectrometry; PP-II, polyproline-II; RCM, reduced and carboxymethylated; SEC, size-exclusion chromatography; sHsps, small heat-shock proteins; TEM, transmission electron microscopy; α-LA, α-lactalbumin.

Author Contributions

H.M.S., B.J., J.A.C. and T.L.P. designed the research; H.M.S. performed the experiments; H.M.S., B.J., J.A.C. and T.L.P. analysed experimental data; H.M.S., B.J., J.A.C. and T.L.P. wrote the manuscript.

Funding

This research was financially supported in part by an Australian Research Council Discovery Project Grant to T.L.P. (DP170102033) and a National Health and Medical Research Council Project Grant to J.A.C. (1068087).

Acknowledgements

H.M.S. is supported by a Faculty of Sciences Divisional Scholarship from the University of Adelaide. Many thanks to Michael Graetz for his assistance and expertise in regards to fluorescence and TEM data acquisition. We thank Adelaide Microscopy (University of Adelaide) for TEM technical assistance. We also thank Flinders Analytical (Flinders University, Australia) for access to IM-MS instrumentation. This research was financially supported in part by an Australian Research Council Discovery Project Grant to T.L.P. (DP170102033) and a National Health and Medical Research Council Project Grant to J.A.C. (1068087).

Open Access

Open access for this article was enabled by the participation of University of Adelaide in an all-inclusive *Read & Publish* pilot with Portland Press and the Biochemical Society under a transformative agreement with CAUL.

Biochemical Journal (2020) **477** 629–643
https://doi.org/10.1042/BCJ20190638

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

References

1 Dinner, A.R., Šalib, A., Smitha, L.J., Dobsona, C.M. and Karplus, M. (2000) Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci.* **25**, 331–339 https://doi.org/10.1016/S0968-0004(00)01610-8

2 Naeem, A., Khan, T.A. and Fazili, N.A. (2015) Protein folding and misfolding: a perspective from theory. *J. Glycom. Lipidom.* **5**, 868–893 https://doi.org/10.4172/2153-0637.1000128

3 Dobson, C.M. (2004) Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol.* **15**, 3–16 https://doi.org/10.1016/j.semcdb.2003.12.008

4 Knowles, T.P.J., Vendruscolo, M. and Dobson, C.M. (2014) The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15**, 384–396 https://doi.org/10.1038/nrm3810

5 Ecroyd, H. and Carver, J.A. (2008) Unraveling the mysteries of protein folding and misfolding. *IUBMB Life* **60**, 769–774 https://doi.org/10.1002/iub.117

6 Stranks, S.D., Ecroyd, H., Van Sluyter, S., Waters, E.J., Carver, J.A. and Von Smekal, L. (2009) Model for amorphous aggregation processes. *Phys. Rev. E* **80**, 051907 https://doi.org/10.1103/PhysRevE.80.051907

7 Kirschner, D.A., Abraham, C. and Selkoe, D.J. (1986) X-ray diffraction from intraneuronal paired helical filaments and extraneuronal amyloid fibers in Alzheimer disease indicates cross-beta conformation. *Proc. Natl Acad. Sci. U.S.A.* **83**, 503–507 https://doi.org/10.1073/pnas.83.2.503

8 Khurana, R., Ionescu-Zanetti, C., Pope, M., Li, J., Nielson, L., Ramirez-Alvarado, M. et al. (2003) A general model for amyloid fibril assembly based on morphological studies using atomic force microscopy. *Biophys. J.* **85**, 1135–1144 https://doi.org/10.1016/S0006-3495(03)74550-0

9 Rochet, J.C. and Lansbury, P.T. (2000) Amyloid fibrillogenesis: themes and variations. *Curr. Opin. Struct. Biol.* **10**, 60–68 https://doi.org/10.1016/S0959-440X(99)00049-4

10 Uversky, V.N. and Fink, A.L. (2006) Protein misfolding, aggregation, and conformational diseases. *Protein Rev.* **4**, 450 https://doi.org/10.1007/978-0-387-36534-3

11 Soto, C. (2003) Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat. Rev. Neurosci.* **4**, 49–60 https://doi.org/10.1038/nrn1007

12 Truscott, R.J.W. (2005) Age-related nuclear cataract - oxidation is the key. *Exp. Eye Res.* **80**, 709–725 https://doi.org/10.1016/j.exer.2004.12.007

13 Permyakov, E.A. and Berliner, L.J. (2000) α-Lactalbumin: structure and function. *FEBS Lett.* **473**, 269–274 https://doi.org/10.1016/S0014-5793(00)01546-5

14 Acharya, K.R., Ren, J.S., Stuart, D.I., Phillips, D.C. and Fenna, R.E. (1991) Crystal structure of human α-lactalbumin at 1.7 A resolution. *J. Mol. Biol.* **221**, 571–581 https://doi.org/10.1016/0022-2836(91)80073-4

15 Kulig, M. and Ecroyd, H. (2012) The small heat-shock protein αB-crystallin uses different mechanisms of chaperone action to prevent the amorphous versus fibrillar aggregation of α-lactalbumin. *Biochem. J.* **448**, 343–352 https://doi.org/10.1042/BJ20121187

16 Goers, J., Permyakov, S.E., Permyakov, E.A., Uversky, V.N. and Fink, A.L. (2002) Conformational prerequisites for α-lactalbumin fibrillation. *Biochemistry* **41**, 12546–12551 https://doi.org/10.1021/bi0262698

17 Carver, J.A., Lindner, R.A., Lyon, C., Canet, D., Hernandez, H., Dobson, C.M. et al. (2002) The interaction of the molecular chaperone α-crystallin with unfolding α-lactalbumin: a structural and kinetic spectroscopic study. *J. Mol. Biol.* **318**, 815–827 https://doi.org/10.1016/S0022-2836(02)00144-4

18 Bomhoff, G., Sloan, K., McLain, C., Gogol, E.P. and Fisher, M.T. (2006) The effects of the flavonoid baicalein and osmolytes on the Mg^{2+} accelerated aggregation/fibrillation of carboxymethylated bovine 1SS-α-lactalbumin. *Arch. Biochem. Biophys.* **453**, 75–86 https://doi.org/10.1016/j.abb.2006.02.001

19 Farrell, J., Wickham, E.D., Unruh, J.J., Qi, P.X. and Hoagland, P.D. (2001) Secondary structural studies of bovine caseins: temperature dependence of β-casein structure as analyzed by circular dichroism and FTIR spectroscopy and correlation with micellization. *Food Hydrocoll.* **15**, 341–354 https://doi.org/10.1016/S0268-005X(01)00080-7

20 Cragnell, C., Choi, J., Segad, M., Lee, S., Nilsson, L. and Skepö, M. (2017) Bovine β-casein has a polydisperse distribution of equilibrium micelles. *Food Hydrocoll.* **70**, 65–68 https://doi.org/10.1016/j.foodhyd.2017.03.021

21 Morgan, P.E., Treweek, T.M., Lindner, R.A., Price, W.E. and Carver, J.A. (2005) Casein proteins as molecular chaperones. *J. Agric. Food Chem.* **53**, 2670–2683 https://doi.org/10.1021/jf048329h

22 Holt, C., Carver, J.A., Ecroyd, H. and Thorn, D.C. (2013) Invited review: Caseins and the casein micelle: their biological functions, structures, and behavior in foods. *J. Dairy Sci.* **96**, 6127–6146 https://doi.org/10.3168/jds.2013-6831

23 Thorn, D.C., Ecroyd, H., Carver, J.A. and Holt, C. (2015) Casein structures in the context of unfolded proteins. *Int. Dairy J.* **46**, 2–11 https://doi.org/10.1016/j.idairyj.2014.07.008

24 Holt, C., Raynes, J.K. and Carver, J.A. (2019) Sequence characteristics responsible for protein–protein interactions in the intrinsically disordered regions of caseins, amelogenins, and small heat-shock proteins. *Biopolymers* **110**, e23319 https://doi.org/10.1002/bip.23319

25 Williams, D.M. and Pukala, T.L. (2013) Novel insights into protein misfolding diseases revealed by ion mobility-mass spectrometry. *Mass Spectrom. Rev.* **32**, 169–187 https://doi.org/10.1002/mas.21358

26 Politis, A., Stengel, F., Hall, Z., Hernández, H., Leitner, A., Walzthoeni, T. et al. (2014) A mass spectrometry-based hybrid method for structural modeling of protein complexes. *Nat. Methods* **11**, 403–406 https://doi.org/10.1038/nmeth.2841

27 Sharon, M. and Robinson, C.V. (2007) The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu. Rev. Biochem.* **76**, 167–193 https://doi.org/10.1146/annurev.biochem.76.061005.090816

28 Dixit, S.M., Polasky, D.A. and Ruotolo, B.T. (2018) Collision induced unfolding of isolated proteins in the gas phase: past, present, and future. *Curr. Opin. Chem. Biol.* **42**, 93–100 https://doi.org/10.1016/j.cbpa.2017.11.010

29 Rabuck-Gibbons, J.N., Lodge, J.M., Mapp, A.K. and Ruotolo, B.T. (2019) Collision-induced unfolding reveals unique fingerprints for remote protein interaction sites in the KIX regulation domain. *J. Am. Soc. Mass Spectrom.* **30**, 94–102 https://doi.org/10.1007/s13361-018-2043-6

30 Shechter, Y., Patchornik, A. and Burstein, Y. (1973) Selective reduction of cystine I–VIII in α-lactalbumin of bovine milk. *Biochemistry* **12**, 3407–3413 https://doi.org/10.1021/bi00742a007

31 Ecroyd, H. and Carver, J.A. (2008) The effect of small molecules in modulating the chaperone activity of αB-crystallin against ordered and disordered protein aggregation. *FEBS J.* **275**, 935–947 https://doi.org/10.1111/j.1742-4658.2008.06257.x

32 Carver, J.A., Duggan, P.J., Ecroyd, H., Liu, Y., Meyer, A.G. and Tranberg, C.E. (2010) Carboxymethylated-κ-casein: a convenient tool for the identification of polyphenolic inhibitors of amyloid fibril formation. *Bioorgan. Med. Chem.* **18**, 222–228 https://doi.org/10.1016/j.bmc.2009.10.063
33 Liu, Y., Jovcevski, B. and Pukala, T.L. (2019) C-phycocyanin from spirulina inhibits α-synuclein and amyloid-β fibril formation but not amorphous aggregation. *J. Nat. Prod.* **82**, 66–73 https://doi.org/10.1021/acs.jnatprod.8b00610
34 Eschweiler, J.D., Rabuck-Gibbons, J.N., Tian, Y. and Ruotolo, B.T. (2015) CIUSuite: a quantitative analysis package for collision induced unfolding measurements of gas-phase protein ions. *Anal. Chem.* **87**, 11516–11522 https://doi.org/10.1021/acs.analchem.5b03292
35 Jovcevski, B., Kelly, M.A., Aquilina, J.A., Benesch, J.L.P. and Ecroyd, H. (2017) Evaluating the effect of phosphorylation on the structure and dynamics of hsp27 dimers by means of ion mobility mass spectrometry. *Anal. Chem.* **89**, 13275–13282 https://doi.org/10.1021/acs.analchem.7b03328
36 Ewbank, J.J. and Creighton, T.E. (1993) Structural characterization of the disulfide folding intermediates of bovine α-lactalbumin. *Biochemistry* **32**, 3694–3707 https://doi.org/10.1021/bi00065a023
37 Konermann, L., Ahadi, E., Rodriguez, A.D. and Vahidi, S. (2013) Unraveling the mechanism of electrospray ionization. *Anal. Chem.* **85**, 2–9 https://doi.org/10.1021/ac302789c
38 Zhong, Y., Han, L. and Ruotolo, B.T. (2014) Collisional and coulombic unfolding of gas-phase proteins: high correlation to their domain structures in solution. *Angew. Chem. Int. Ed.* **53**, 9209–9212 https://doi.org/10.1002/anie.201403784
39 Raynes, J.K., Day, L., Augustin, M.A. and Carver, J.A. (2015) Structural differences between bovine A1 and A2 β-casein alter micelle self-assembly and influence molecular chaperone activity. *J. Dairy Sci.* **98**, 2172–2182 https://doi.org/10.3168/jds.2014-8800
40 Gregersen, N., Bross, P., Vang, S. and Christensen, J.H. (2006) Protein misfolding and human disease. *Annu. Rev. Genomics Hum. Genet.* **7**, 103–124 https://doi.org/10.1146/annurev.genom.7.080505.115737
41 Carver, J.A., Ecroyd, H., Truscott, R.J.W., Thorn, D.C. and Holt, C. (2018) Proteostasis and the regulation of intra- and extracellular protein aggregation by ATP-independent molecular chaperones: lens α-crystallins and milk caseins. *Acc. Chem. Res.* **51**, 745–752 https://doi.org/10.1021/acs.accounts.7b00250
42 Zhang, X., Fu, X., Zhang, H., Liu, C., Jiao, W. and Chang, Z. (2005) Chaperone-like activity of β-casein. *Int. J. Biochem. Cell Biol.* **37**, 1232–1240 https://doi.org/10.1016/j.biocel.2004.12.004
43 Das, S., Pukala, T.L. and Smid, S.D. (2018) Exploring the structural diversity in inhibitors of α-synuclein amyloidogenic folding, aggregation, and neurotoxicity. *Front. Chem.* **6**, 181 https://doi.org/10.3389/fchem.2018.00181
44 Pavlova, A., Cheng, C.-Y., Kinnebrew, M., Lew, J., Dahlquist, F.W. and Han, S. (2016) Protein structural and surface water rearrangement constitute major events in the earliest aggregation stages of tau. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E127–E136 https://doi.org/10.1073/pnas.1504415113
45 Lindner, R.A., Kapur, A. and Carver, J.A. (1997) The interaction of the molecular chaperone, α-crystallin, with molten globule states of bovine α-lactalbumin. *J. Biol. Chem.* **272**, 27722–27729 https://doi.org/10.1074/jbc.272.44.27722
46 Ashton, L. and Blanch, E.W. (2010) pH-induced conformational transitions in α-lactalbumin investigated with two-dimensional Raman correlation variance plots and moving windows. *J. Mol. Struct.* **974**, 132–138 https://doi.org/10.1016/j.molstruc.2010.03.005
47 Adzhubei, A.A., Sternberg, M.J.E. and Makarov, A.A. (2013) Polyproline-II helix in proteins: structure and function. *J. Mol. Biol.* **425**, 2100–2132 https://doi.org/10.1016/j.jmb.2013.03.018
48 Koudelka, T., Hoffmann, P. and Carver, J.A. (2009) Dephosphorylation of αS- and β-caseins and its effect on chaperone activity: a structural and functional investigation. *J. Agric. Food Chem.* **57**, 5956–5964 https://doi.org/10.1021/jf9008372
49 Holt, C. and Carver, J.A. (2012) Darwinian transformation of a "scarcely nutritious fluid" into milk. *J. Evol. Biol.* **25**, 1253–1263 https://doi.org/10.1111/j.1420-9101.2012.02509.x
50 Cox, D., Selig, E., Griffin, M.D.W., Carver, J.A. and Ecroyd, H. (2016) Small heat-shock proteins prevent α-synuclein aggregation via transient interactions and their efficacy is affected by the rate of aggregation. *J. Biol. Chem.* **291**, 22618–22629 https://doi.org/10.1074/jbc.M116.739250
51 Esposito, G., Garvey, M., Alverdi, V., Pettirossi, F., Corazza, A., Fogolari, F., et al. (2013) Monitoring the interaction between β2-microglobulin and the molecular chaperone αB-crystallin by NMR and mass spectrometry: αB-crystallin dissociates β2-microglobulin oligomers. *J. Biol. Chem.* **288**, 17844–17858 https://doi.org/10.1074/jbc.M112.448639
52 Cox, D., Carver, J.A. and Ecroyd, H. (2014) Preventing α-synuclein aggregation: the role of the small heat-shock molecular chaperone proteins. *Biochim. Biophys. Acta* **1842**, 1830–1843 https://doi.org/10.1016/j.bbadis.2014.06.024
53 Treweek, T.M., Meehan, S., Ecroyd, H. and Carver, J.A. (2015) Small heat-shock proteins: important players in regulating cellular proteostasis. *Cell. Mol. Life Sci.* **72**, 429–451 https://doi.org/10.1007/s00018-014-1754-5
54 Carver, J.A., Grosas, A.B., Ecroyd, H. and Quinlan, R.A. (2017) The functional roles of the unstructured N- and C-terminal regions in αB-crystallin and other mammalian small heat-shock proteins. *Cell Stress Chaperones* **22**, 627–638 https://doi.org/10.1007/s12192-017-0789-6

Supplementary information for:

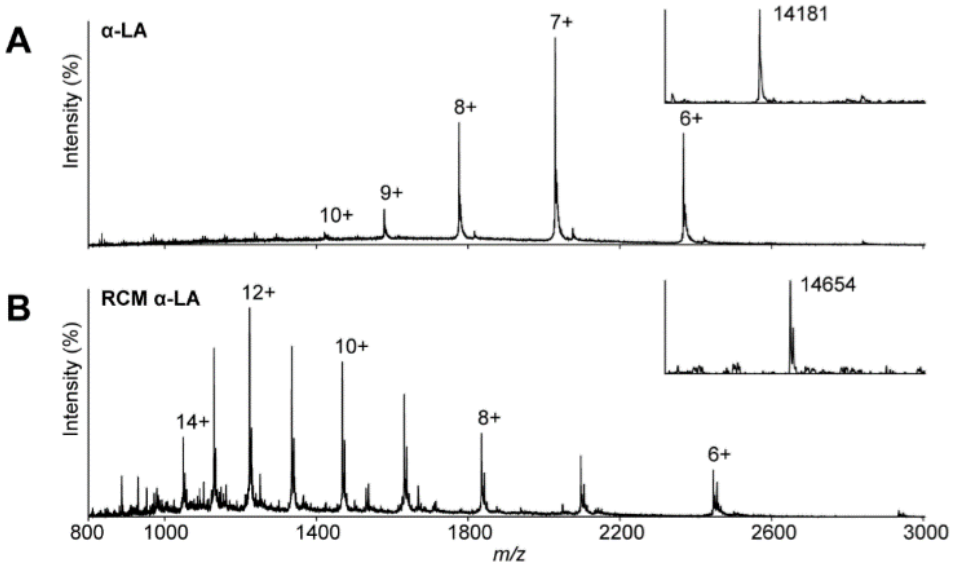


Figure S1. α-LA was fully reduced and carboxymethylated. Denatured mass spectrum of α-LA (top) and RCM α-LA (bottom) (both 100 µM) in 50% acetonitrile with 1% trifluoroacetic acid. Data was deconvoluted to give the molecular masses of α-LA at 14,181 Da (top, inset) and RCM α-LA at 14654 Da (bottom, inset). A mass difference of 473 Da corresponds to the reduction and carboxymethylation of all 8 cysteines.

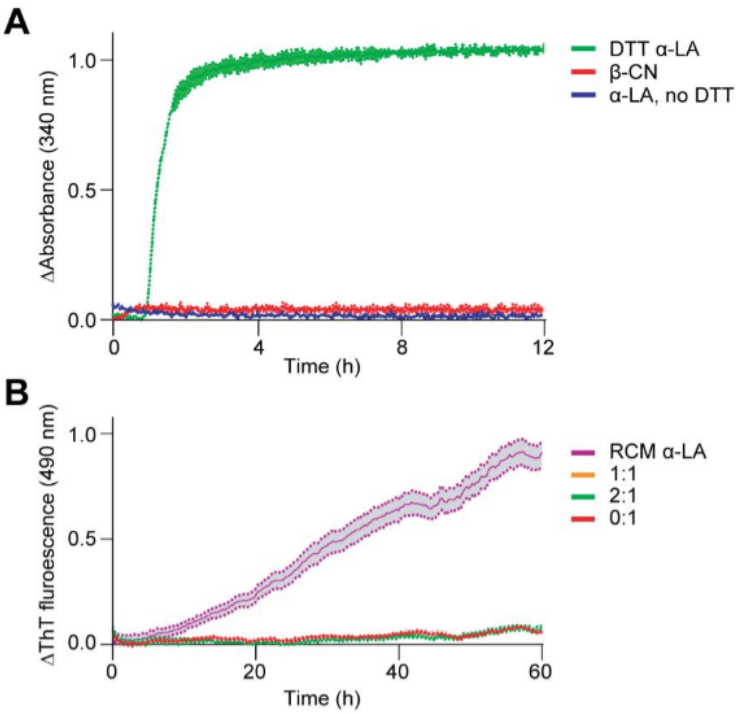


Figure S2. β-CN does not affect light scattering or ThT fluorescence assays. (A) Amorphous aggregation of α-LA (100 µM) was induced with DTT (2 mM) and monitored by the scattering of light at 340 nm. β-CN (100 µM) in the presence of DTT (2 mM) and α-LA (100 µM) in the absence of DTT displayed no effects on light scattering. (B) Fibril formation of RCM α-LA (200 µM) was monitored by the change in ThT fluorescence at 490 nm (excitation at 440nm) with varied molar ratios of β-CN (RCM α-LA : β-CN) to determine chaperone activity. β-CN (200 µM) in the absence of RCM α-LA showed no effects on ThT fluorescence. All samples were in ammonium acetate (100 mM, pH 7.0). Data in A and B are reported as mean ± SEM (n = 3).

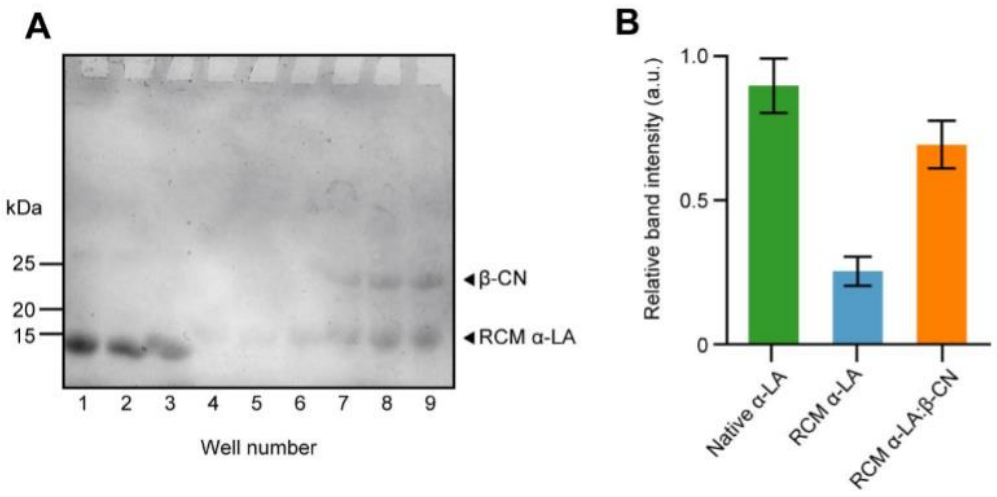


Figure S3. (A) SDS-PAGE and Coomassie staining of native α-LA (200 µM, well numbers 1-3), RCM α-LA (200 µM, well numbers 4-6) and RCM α-LA:β-CN (1:1, 200 µM, well numbers 7-9) analysed in triplicate following 72 hours of incubation. 10 µL of sample was added to each well. **(B)** Analysis of the band densities shows soluble levels of native α-LA and RCM α-LA in the presence of β-CN are not different within experimental error. RCM α-LA:β-CN shows a significantly higher band density than RCM α-LA alone, indicating β-CN protects RCM α-LA from fibrillar aggregation. Data are reported as mean ± SEM (n =3).

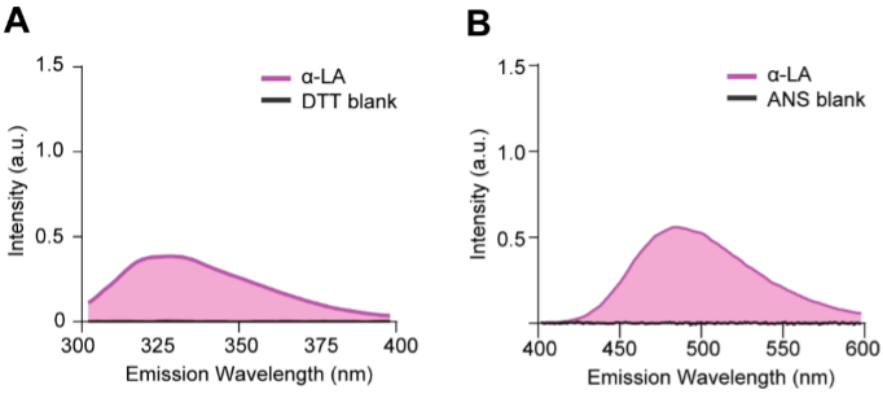


Figure S4. (A) Intrinsic tryptophan fluorescence spectra of α-LA (100 µM) and a DTT (2 mM) blank in 100 mM ammonium acetate (pH 7.0) showing DTT alone does not influence fluorescence. **(B)** Bis-ANS (10 µM) fluorescence spectra measuring the exposed hydrophobicity of α-LA (100 µM). In the absence of ANS (ANS blank) no extrinsic fluorescence is observed.

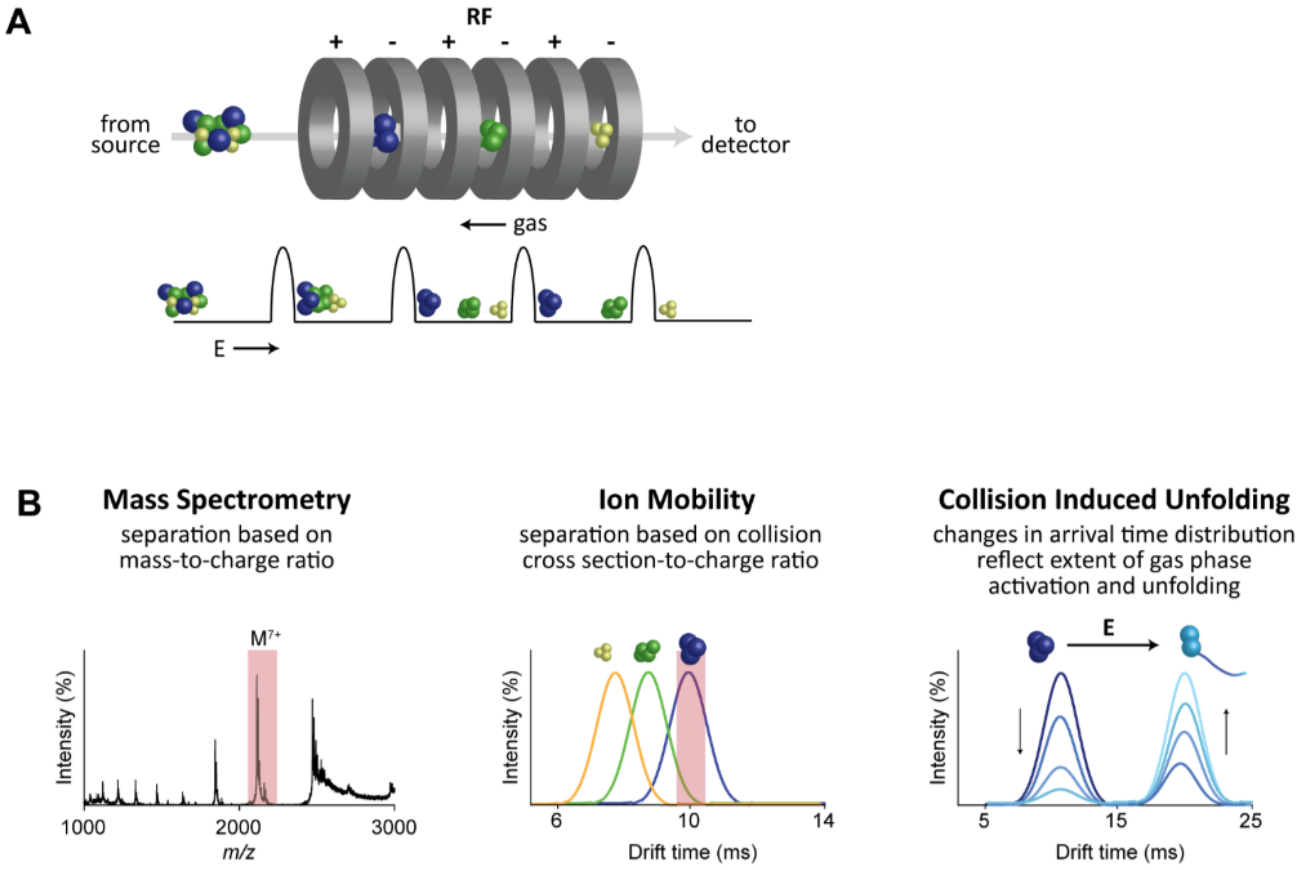


Figure S5. (A) Travelling wave ion mobility separation. Alternating phases of RF voltage are applied to a stacked ring ion guide (top), to which a sequence of symmetric potential waves is superimposed (bottom). The ions are propagated along the potential wave front against a reverse flow of buffer gas. Ions with low mobilities (large collision cross section) experience the most wave roll over events due to increased collisions with gas, and hence have the longest drift time. Figure reproduced with permission from Pukala, T. L. (2019) Importance of collision cross section measurements by ion mobility mass spectrometry in structural biology. Rapid Commun. Mass Spectrom. **33**, 72-82. **(B)** Mass spectrometry measures the mass-to-charge ratio of an ion to provide information about molecular composition (left). Importantly, the mobility of an ion (centre) is related to structural features such as charge and collision cross section, the latter of which can be used to infer information about three-dimensional conformations. In a collision induced unfolding experiment (left), ions of interest are subjected to collisional activation which induces unfolding of proteins or protein complexes in the gas phase. This is monitored by observing the corresponding shift in arrival time distribution to longer drift times with increasing populations of extended structures.

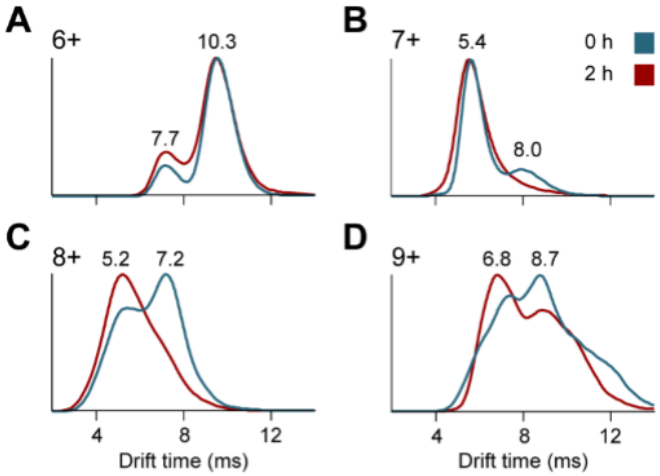


Figure S6. Arrival time distributions of M^{6+} **(A)**, M^{7+} **(B)**, M^{8+} **(C)** and M^{9+} **(D)** ions of RCM α-LA. Samples were incubated at 37 °C with shaking (300 rpm). Each charge state shows either a relative increase in the proportion of more compact species or a decrease in the proportion of less compact species following incubation for 2 hours.

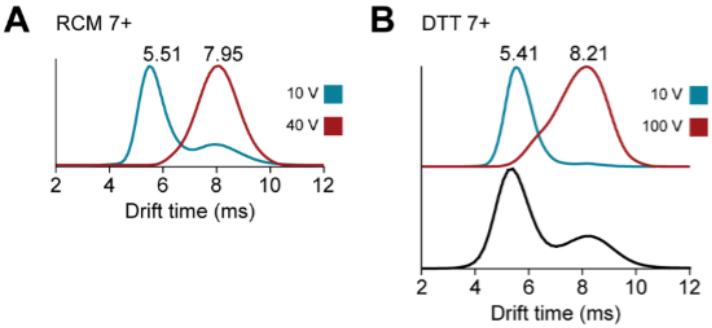


Figure S7. IM-MS analysis of DTT and RCM α-LA shows unfolded species align with their respective monomers when fully activated. **(A)** The ATDs of the extended M^{7+} RCM α-LA ions at 10 V (blue, drift time 7.95 ms) align with the fully activated species at 40 V (red). **(B)** The ATDs of the M^{7+} of native α-LA at 10 V (top panel, blue) and 100 V (top panel, red) align with the two species present in DTT α-LA after 2 hr incubation with DTT (bottom panel, black). Given the intact disulphide bonds, native α-LA required more gas phase activation than DTT α-LA to reach the fully unfolded state.

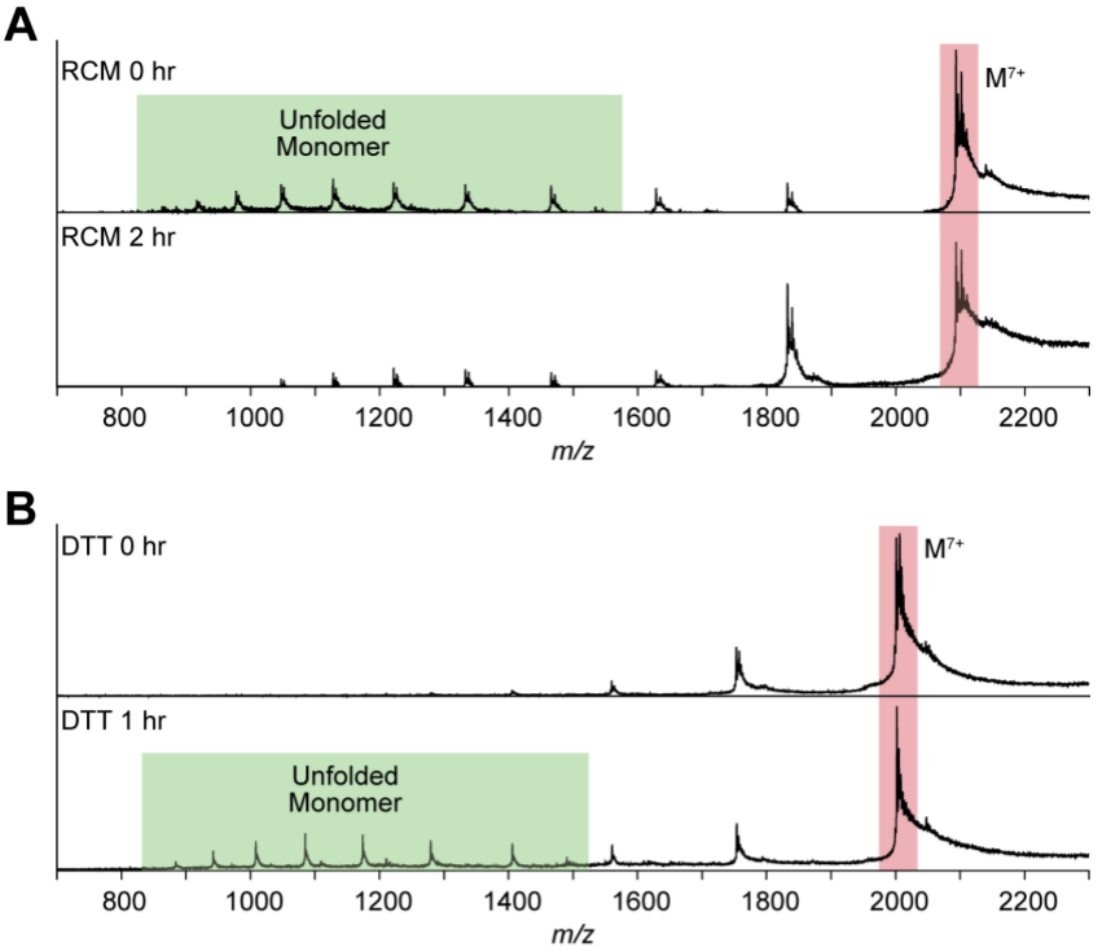


Figure S8. (A) RCM α-LA shows highly charged species (**A**, green box), following 2 hours of incubation (**A**, bottom) these highly charged species reduce in intensity indicating a reduction in unfolded, exposed ionisable residues. Initially, DTT α-LA shows few highly charged species (**B**, top). Following 1 hour of incubation highly charged species (**B**, green box) appear indicating α-LA unfolds to reveal more ionisable residues. All samples were incubated at 37 °C with shaking (300 rpm).

Chapter 3: Structural and mechanistic insights into the influence of phospholipids in amyloid fibril formation and inhibitor efficacy

Disclaimer and author contributions.

The data presented in this chapter has been peer reviewed and is under revision in FEBS Journal.

The majority of the experiments, data analysis and manuscript writing were performed by Sanders, H. M.

Jovcevski, B. and Pukala, T. L. assisted with experimental design, data interpretation and editing of the manuscript.

Structural and mechanistic insights into the influence of phospholipids in amyloid fibril formation and inhibitor efficacy

Henry M. Sanders[1], Blagojce Jovcevski[1], Tara L. Pukala[1*]

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

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

Keywords: protein unfolding, amyloid aggregation, amyloid inhibitors, lipid membrane, ion mobility-mass spectrometry

Running title: Lipids in amyloid fibril formation and inhibition

3.1 Abstract

Under certain cellular conditions functional proteins undergo misfolding, leading to a transition into oligomers which precede formation of amyloid fibrils. Misfolding proteins are associated with a variety of neurodegenerative diseases such as Alzheimer's disease and Parkinson's disease. While the importance of lipid membranes in misfolding and disease aetiology is broadly accepted, the influence of lipid membranes during therapeutic design has been largely overlooked. This study utilized an integrated biophysical approach to provide structural and mechanistic insights into the effects of two lipid membrane systems (anionic and zwitterionic) on the inhibition of Aβ40 and α-synuclein amyloid formation at the monomer, oligomer and fibril level. Large unilamellar vesicles (LUVs) were shown to substantially decrease the effectiveness of two well-known fibril inhibitors, (-)-epigallocatechin gallate (EGCG) and resveratrol, however, use of immunoblotting and native ion-mobility mass spectrometry revealed this occurs through varying mechanisms. Oligomeric populations in particular are differentially affected by LUVs in the presence of resveratrol, an elongation phase inhibitor, compared to EGCG, a nucleation targeted inhibitor. Native Ion-mobility mass spectrometry showed EGCG interacts with compact forms of monomeric Aβ40 and α-synuclein, while LUVs induce conformational transitions indicative of nucleation before elongation. The competing effects of the lipids and inhibitor, along with reduced inhibitor binding in the presence of LUVs, provides a mechanistic understanding for the decreased inhibitor efficacy in a lipid environment. Together, this study highlights that amyloid inhibitor design may be misguided if the effects of lipid membrane composition and architecture are not considered during development.

Abbreviations

Amyloid β (Aβ), α-synuclein (αS), Large unilamellar vesicles (LUVs), (-)-epigallocatechin gallate (EGCG), ion-mobility mass spectrometry (IMMS), arrival time distribution (ATD), collisional cross section (CCS), zwitterionic 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (PC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (PG), thioflavin T (ThT), transmission electron microscopy (TEM).

3.2 Introduction

Many intrinsically disordered proteins (IDPs) are prone to aggregation in the form of amyloid fibrils, and the accumulation of neuronal plaques containing these aggregates are hallmarks of a variety of neurodegenerative diseases. For example, extracellular plaques composed of amyloid β (Aβ) and intracellular Lewy bodies containing α-synuclein (αS) are associated with Alzheimer's disease and Parkinson's disease, respectively. However, despite strong correlative links, the molecular mechanisms by which these IDPs contribute to the disease aetiology is still poorly understood.

In periods of cellular stress, amyloidogenic proteins can enter an aggregation pathway commencing with a nucleation-phase in which misfolding proteins restructure to form aggregation-prone, prefibrillar species. This is followed by an elongation-phase in which monomeric proteins are recruited from solution to form mature fibrils [1,2]. Current research suggests the low molecular weight, soluble, oligomeric species formed early on the aggregation pathway are primarily responsible for neurotoxic effects rather than the mature amyloid fibrils [3–7]. Small, pre-fibrillar species have been shown to be more toxic than the plaques that they eventually form [6–8], and smaller oligomeric species have been shown to exert enhanced toxicity [9]. While focus has shifted in this regard from fibrils to monomers and oligomers, further research is required to understand the structure of toxic oligomeric species to better guide therapeutic development.

Importantly, many recent studies have highlighted the critical role that membrane composition and curvature plays in amyloid disease propagation through the interactions between misfolded oligomeric species and plasma/organelle membranes. Whether this is due to a unifying mechanism or a variety of molecular processes driving pathogenesis is yet to be fully understood [10–13]. For example, Aβ has been shown to interact with neuronal cell membranes whereby oligomeric forms of the protein cause membrane leakage and transmission of ions, disrupting ionic homeostasis particularly for Ca^{2+}, and resulting in neuronal cell death [11,14–16]. On top of this, studies show that lipid bilayers (and other lipid forms) may increase the aggregation propensity of misfolded proteins as well as alter monomeric and oligomeric structures [17–19]. Also of importance is membrane composition, as differing lipid mixtures induce a variety of different effects including increased rate of αS nucleation in the presence of anionic lipid membranes [20], while shorter chain phospholipids

have been shown to halt Aβ fibril formation at low concentrations and accelerate it at higher concentrations [21]. Anionic lipids have been shown to interact with many misfolding proteins such as, Aβ, αS and human islet amyloid polypeptide and increase rates of fibrillar elongation [22,23]. Other components of the cellular membrane such as cholesterol and gangliosides can also influence fibril kinetics and morphology [13,24–26]. The lipidome of neurons is complicated, diverse and is altered with age and disease states. After 20 years of age the total lipid content of the brain decreases, as does the concentration of sphingomyelins and cholesterol, although different neurons are affected more than others, further reinforcing the importance of understanding of amyloid formation in diverse lipid environments [27].

The economic and social burden caused by amyloid diseases makes the hunt for an effective therapeutic incredibly important [28]. Natural polyphenols are of great interest due to their antioxidant effects and have been shown extensively to directly inhibit the aggregation of a wide range of misfolding proteins, particularly Aβ and αS [29,30]. Two compounds that have been widely studied and demonstrated efficacious anti-amyloid activity, (-)-epigallocatechin gallate (EGCG) and resveratrol, are natural products isolated from green tea and red wine, respectively. EGCG has been shown to not only inhibit fibril formation but also remodel Aβ into smaller, nontoxic, amorphous aggregates [30–33]. Resveratrol, whilst a promising candidate in animal studies [34,35], has been shown to stop Aβ fibril formation through a capping mechanism, preventing large scale fibrilization while having no effect on oligomer formation [36–38]. Despite the promise of these studies, they and many others largely ignore the influence that a lipid environment plays on amyloid formation as well as the effectiveness of inhibitors in this highly dynamic environment [39]. To better develop effective small molecule inhibitors of misfolding and aggregating proteins we must understand their interactions in the physiological environments in which these inhibitors will be acting.

IDPs are, by definition, highly dynamic and heterogeneous, making them not amenable to study by traditional structural biology techniques such as x-ray crystallography due to the transient nature of the protein structure itself. Other methods like nuclear magnetic resonance (NMR) spectroscopy, circular dichroism (CD) spectroscopy and fluorescence spectroscopy present data on a global average rather than the collection of unique structural states and oligomers that constitute a transient mixture. The combination of nano-electrospray ionisation (nESI) with native ion-mobility mass spectrometry (IM-MS) allows for

gentle ionisation of proteins into the gas-phase, preserving important tertiary and quaternary structures, while separating based on mass to charge (*m/z*) as well as a rotationally averaged size (or collisional cross section, CCS) [40–42]. It therefore offers a useful alternative analytical method to probe the heterogeneous and dynamic states of IDPs during protein aggregation, and investigate the impact of small molecules on this process.

Here we have utilised IM-MS in an integrated biophysical approach to better understand the role of lipid membranes in protein misfolding diseases at distinct stages along the aggregation pathway, from monomeric structure to fibril morphology. We explore differences in the aggregation dynamics of Aβ40 and αS in the presence of two membrane models, namely unilamellar vesicles (LUVs) composed of zwitterionic 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (PC) or a mixture of anionic 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (PG) with PC (herein PCPG). These model systems were chosen as PC is a major phospholipid in human neuronal membranes and PG acts to create a negative bilayer surface charge, which are major components within grey and white matter [27]. The relative efficacy of two small molecule inhibitors, EGCG and resveratrol, was evaluated in these lipid environments and the effects of these variables on protein structure and aggregation propensity were interrogated. This study highlights the interplay between misfolding proteins, phospholipid membranes and fibril inhibitors, and may inform the design of more effective therapies to take into account the influence that lipid environments impose on disease development and inhibition.

3.3 Experimental Procedures

Human Aβ40 (lyophilised powder) was purchased from Anaspec (Fremont, USA). 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine and 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol), were purchased from Avanti Polar Lipids (Alabaster, USA). EGCG, resveratrol and all other reagents were purchased from Sigma-Aldrich (St. Louis, USA) unless otherwise specified.

3.3.1 Preparation of Aβ40, αS and LUVs

Expression and purification of human αS (UniProt accession number: P37840) was performed as described previously [43]. Pre-formed αS seed fibrils were produced from the monomeric protein preparation as described previously [44]. Aβ40 was pre-treated with NH_4OH to ensure monomerisation according to the method described by Ryan*, et al* [45]. Aliquots were snap frozen and stored at -80 °C and were only defrosted immediately before use. LUVs were extruded using an Avanti Mini-Extruder with 0.1 µm extruder filters (Millipore, Billerica, USA) according to the manufacturer's instructions. LUV diameters were confirmed by dynamic light scattering (DLS) and transmission electron microscopy (TEM).

3.3.2 Dynamic light scattering

LUV diameters were measured by dynamic light scattering using a Malvern Zetasizer nano ZS (Malvern Instruments, Malvern, UK). LUVs were characterised at a concentration of 1.5 mM with a 633 nm laser at a back scattering angle of 173°.

3.3.3 Thioflavin T assays

Fibrillar aggregation was monitored in real-time using a ThT fluorescence assay [46]. Primary nucleation-mediated fibrilization of Aβ40 (50 µM) and αS (150 µM) was performed in 100 mM ammonium acetate (pH 7.4) and thioflavin T (ThT) (40 µM) in the absence and presence of LUVs (1:30 molar ratio; protein:lipid) [24]. Inhibitor concentrations were based on previously established concentrations [33,47]. Fibrilization of αS (50 µM) in the presence of short seed fibrils (2.5 % w/w) was performed to examine the effects on fibril elongation in the absence and presence of LUVs and inhibitors. Samples were plated (50 µL/well) into sealed 384-well microplates (Greiner Bio-One, Kremsmünster, Austria) and fibril formation was monitored using a FLUOstar Optima microplate reader (BMG Lab Technologies, Melbourne, Australia) over a period of 20 h (Aβ40), 25 h (seeded αS) and 120 h (unseeded αS) at 37 °C with shaking (double orbital, 50 rpm) for 30 s prior to fluorescence measurement. ThT fluorescence was measured with an excitation and emission wavelength of 440 and 490 nm, respectively. All data, except for seeded αS assays, were fitted with non-linear sigmoidal dose response to calculate and compare the rate of fibril elongation as well as the extent of fibril formation (i.e.

inhibition). Seeded αS assays were fitted with a one-phase exponential association. All ThT kinetic data were analysed using Prism 8.0 (GraphPad Software, La Jolla, USA) and all assays were performed at least three times and are reported as mean ± SEM.

3.3.4 Transmission electron microscopy

Samples were taken immediately at the conclusion of the ThT assay and placed directly (3 µL) onto a carbon-coated 400-mesh nickel TEM grid (ProSciTech, Thuringowa Central, Australia) [32]. Grids were washed with 0.22 µm filtered MilliQ water and stained with 2% (w/v) uranyl acetate solution. Samples were viewed using either a CM100 transmission electron microscope (Philips, Eindhoven, Netherlands) or a Tecnai G2 Spirit TEM (FEI, Oregon, USA) with a magnification of 45 000×.

3.3.5 Immunoblotting

The relative abundance of various morphological states (i.e. monomer, oligomer or fibril) of Aβ40 and αS was assessed by immuno-dot blotting as described previously [33,48]. Samples were prepared in sealed 384-well microplates (Greiner Bio-One, Kremsmünster, Austria) and incubated using a FLUOstar Optima microplate reader (BMG Lab Technologies, Melbourne, Australia), replicating the conditions of the ThT kinetic assays in the absence of ThT. Briefly, 1.5 µL of sample at various time points during fibril formation were dotted onto nitrocellulose membranes and blocked with 5% skim milk in TBS-T at room temperature for 1 h. Membranes were washed four times prior to incubation with either an A11 anti-oligomer (1:5000 in 2% skim milk TBS-T) (Invitrogen, California, USA) or an anti-fibril (ab205339, 1:5000 in 2% skim milk TBS-T)) (Abcam, Cambridge, United Kingdom) at 4 °C overnight. An IgG-HRP secondary antibody (403001) and Clarity™ ECL substrate (both Bio-Rad, California, USA) were used for detection. The intensity of the blots was quantified using ImageJ [49], where experiments were performed in triplicate and data reported as mean ± SEM. Data was analysed by one-way ANOVA using Prism 8.0 (GraphPad Software, La Jolla, USA).

3.3.6 Ion-mobility mass spectrometry

IM-MS analysis was performed on a Synapt HDMS G1 Q-TOF mass spectrometer (Waters Corporation, Manchester, UK) using a nano-electrospray ionisation source. Samples were prepared in 100 mM ammonium acetate (pH 7.0) to a final protein concentration of 50 µM. Aβ40 and αS were incubated in the presence of LUVs (1:10 molar ratio; protein:LUV) and/or inhibitor (1:1 molar ratio; protein:inhibitor) for 10 min at room temperature before analysis. Samples were loaded into platinum-coated borosilicate glass capillaries prepared in-house. Gentle ionisation and gas-phase conditions were applied to minimise structural changes prior to detection [50], with instrument parameters as follows: capillary voltage, 1.60 kV; sampling cone, 30 V; extraction cone, 1.5 V; trap/transfer collision energy, 10-15 V; trap gas, 5.5 l/h; backing gas, ~4.5 mbar. The parameters for IM were as follows: IM cell wave height, 8 V; IM cell wave velocity, 350 m/s; transfer t-wave height, 8 V; transfer t-wave velocity, 250 m/s. Mass spectra and arrival time distributions (ATDs) were viewed using MassLynx (v4.1) and DriftScope (v2.1), respectively (Waters Corporation, Manchester, UK).

3.4 Results

3.4.1 LUVs increase the rate of Aβ40 fibrillization and reduce EGCG-resveratrol efficacy

The effects of LUVs consisting of either PC or PCPG on Aβ40 fibril formation were first assessed by performing ThT kinetic assays (**Fig. 1**). These aggregation assays were carried out in the presence of LUVs of certain size (i.e. 100 nm polycarbonate membranes were used to create homogenous LUVs) in order to prevent any influences on aggregation kinetics arising from differing vesicle sizes. The size distribution of the LUVs was confirmed by DLS (**Supp. Fig. 1**) which displayed a normal and narrow size distribution with diameters of approximately 100-300 nm. Furthermore, ThT experiments in the presence of LUVs alone were conducted to verify that the addition of lipids did not significantly affect ThT fluorescence properties (**Supp. Fig. 2**).

Aβ40 showed an increase in ThT fluorescence in the plateau phase as well as an increased rate of fibril elongation when incubated with both PC and PCPG LUVs (1:30 molar ratio, Aβ40:LUV), with PCPG LUVs elongating over six times faster compared to Aβ40 alone (**Supp. Table 1**). The formation of mature fibrils was confirmed by TEM (**Fig. 1B**) and interestingly,

Aβ40 fibrils incubated with LUVs appeared to be vesicle-bound (**Fig. 1C and D**), suggesting LUVs increase the fibrilization propensity of Aβ40 through interactions at the bilayer surface.

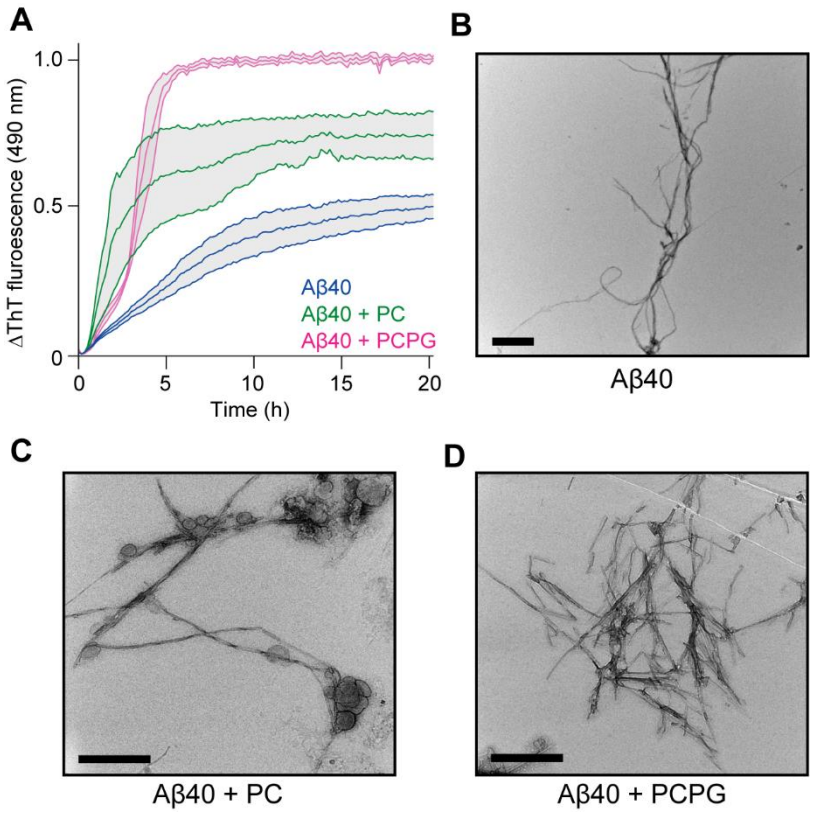


Figure 1. PC and PCPG LUVs alter the aggregation kinetics and fibril formation of Aβ40. (**A**) Aβ40 (50 μM) (blue) fibril formation was monitored by the change in ThT fluorescence in the presence of either PC (green) or PCPG (pink) LUVs (1:30, Aβ40:LUV). Data are reported as mean ± SEM (n = 3). TEM images show the morphology of Aβ40 alone (**B**) and in the presence of PC (**C**) and PCPG (**D**) LUVs at the same molar ratios. Scale bars represent 500 nm.

Additional ThT assays were conducted to assess the efficacy of small molecule amyloid inhibitors while in the presence of LUVs. Resveratrol and EGCG are known fibril inhibitors broadly effective against a variety of amyloidogenic proteins. However, they have differing modes of inhibition and differing levels of efficacy [51]. EGCG and resveratrol were added at a 1:1 molar ratio which has been used previously to assess inhibitory efficacy of these molecules [33,47]. Upon incubation with LUVs, both EGCG (**Fig. 2A**) and resveratrol (**Fig. 2B**) were found to be less effective as fibrilization inhibitors. Consistent with previous studies of Aβ40 [33,51], EGCG significantly limits ThT fluorescence at a 1:1 molar ratio, with a 95%

decrease in fluorescence at plateau compared to Aβ40 alone. However, when incubated with either LUV at a 1:30 ratio (Aβ40:LUV), the inhibitory efficacy of EGCG was reduced, with a plateau fluorescence decrease of 81% and 50% in the presence of PC and PCPG LUVs respectively (**Fig. 2A and Supp. Table 1**). Interestingly, Aβ40 in the presence of LUVs displayed a decrease in ThT fluorescence following approximately 7 and 15 hours of incubation for PCPG and PC respectively, indicative of decreased mature fibril formation. This observed loss in fluorescence may be attributed to LUV degradation during the assay [52] at which point EGCG is able to remodel mature fibrils into smaller, amorphous aggregates [32,53].

The efficacy of resveratrol as an inhibitor of Aβ40 aggregation was also examined in the presence of LUVs. Resveratrol is generally less effective at inhibiting Aβ40 fibril formation than EGCG, with greater ThT fluorescence observed at the same inhibitor concentration. Nevertheless, similar to EGCG, incubation with LUVs resulted in a further reduction in the inhibitory efficacy of resveratrol (**Fig. 2B and Supp. Table 1**). However, unlike EGCG, no difference was observed between the zwitterionic (PC) and anionic (PCPG) LUV systems, and no decline in fluorescence was observed over time. This latter point is consistent with the reported mechanism of action of resveratrol; despite being shown to remodel mature fibrils at high concentrations [53], resveratrol appears to cap fibril formation but is unable to disassemble fibrils [37]. Overall, the combined data demonstrate the ability of LUVs to reduce the efficacy of two well-known small molecule inhibitors of fibrilization. PCPG was found to affect Aβ40 aggregation and inhibitor efficacy more significantly, and anionic LUVs have also been shown in literature to have a greater affinity for Aβ40 [22,54]. For this reason, we chose to focus only on PCPG LUVs for subsequent experiments in this study.

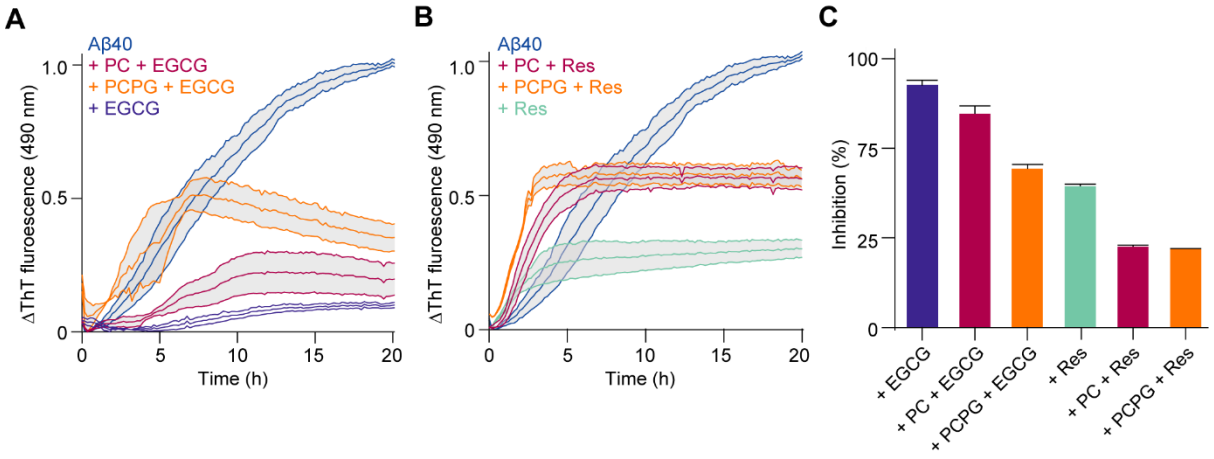


Figure 2. Inhibiting Aβ40 fibril formation by EGCG and resveratrol is compromised by PC and PCPG LUVs. The effect of LUVs on Aβ40 fibrillar inhibition by **(A)** EGCG (50 µM) and **(B)** resveratrol (50 µM) was monitored through ThT fluorescence assays. Aβ40 (50 µM) was incubated with either inhibitor in the presence and absence of LUVs (1:30, Aβ40:LUV). (**C**) The extent of fibril inhibition afforded by EGCG and resveratrol with and without LUVs was calculated by comparing the ThT fluorescence maximum across each treatment. Data are reported as mean ± SEM (n = 3).

3.4.2 LUVs increase the rate of αS fibril elongation and reduce EGCG-resveratrol efficacy

To investigate whether the effects of LUVs on fibrillar aggregation and inhibition by EGCG and resveratrol are consistent between protein systems, we also performed analogous ThT experiments using αS (**Fig. 3**). Compared to Aβ40, αS aggregates at a much slower rate and allows for *in vitro* kinetic analysis to be performed in the presence and absence of seed fibrils in a robust manner. This allows us to separately examine the inhibitory action of EGCG and resveratrol during primary nucleation and fibril elongation stages in the presence of LUVs. As observed for Aβ40, in the presence of PCPG LUVs, αS fibrils elongated faster (5 times greater rate) compared to αS alone during primary nucleation-mediated fibrilization (i.e. unseeded aggregation) (**Fig. 3A**). However, in the presence of seed fibrils (2.5% w/w), the rate of fibril elongation remained similar to αS alone (**Fig. 3B**). The data suggests that LUVs primarily modulate the rate of fibril elongation through alterations in monomeric or early oligomeric structures during the nucleation phase.

EGCG and resveratrol were subsequently introduced into these systems to ascertain whether the LUVs had similar effects on inhibitory efficacy against αS aggregation. In the unseeded assays, the inhibitory activity of resveratrol towards αS aggregation was completely nullified in the presence of PCPG LUVs (**Fig. 3A and C**), and in fact, αS fibrils with resveratrol and PCPG elongated faster than αS alone, despite the presence of inhibitor. The inhibitor efficacy of resveratrol was also greatly reduced in the seeded αS system (**Fig. 3B and D**), however the rate of elongation remained consistent across all inhibitor conditions. Contrastingly, EGCG was found to be far more effective as an inhibitor against both unseeded and seeded αS aggregation, displaying little ThT fluorescence in the presence of PCPG LUVs.

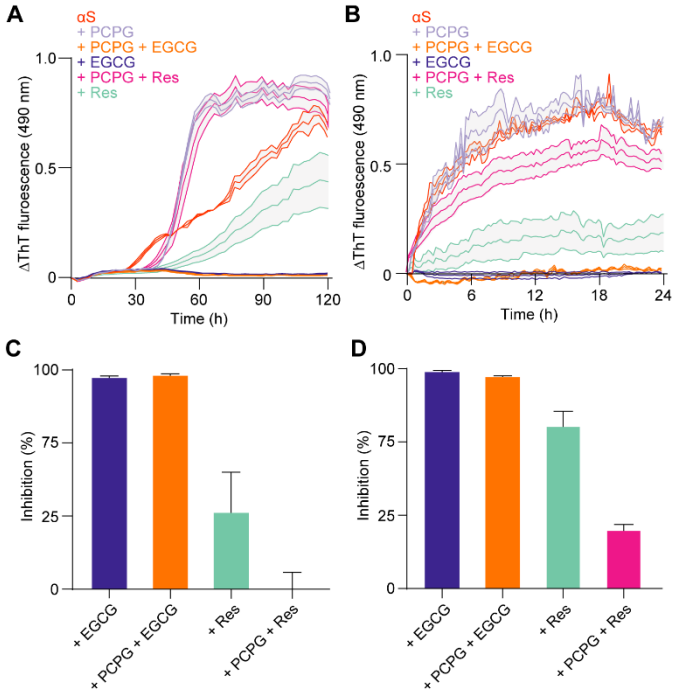


Figure 3. PCPG LUVs modulate the inhibitory efficacy of EGCG and resveratrol during seeded and unseeded αS aggregation. The effect of PCPG LUVs on unseeded and seeded αS fibril formation in the presence of EGCG (50 µM) and resveratrol (50 µM) was monitored through ThT fluorescence assays. **(A)** Unseeded αS (150 µM) and **(B)** seeded αS (50 µM with 2.5% w/w seed fibrils) was incubated with inhibitors in the presence and absence of LUVs (1:30 αS:LUV). The concentration-dependent inhibition activity of EGCG and resveratrol against unseeded **(C)** and seeded **(D)** αS fibril formation with and without LUVs was calculated by comparing the ThT fluorescence maximum across each treatment. Data are reported as mean ± SEM (n = 3).

3.4.3 LUVs and inhibitors differentially influence the distribution of protein aggregation states

ThT assays are only able to provide insight into the fibril content during the aggregation process, which ignores important details regarding monomeric structure and oligomeric populations. For these reasons, we also carried out immunoblotting experiments to ensure ThT-derived data mirrored relative fibril content across treatments, as well as to gain insight into pre-fibrillar oligomeric populations. Aβ40 fibril-specific antibodies were first used to validate ThT fluorescence data, with the relative intensity (i.e. abundance) of fibrillar species compared to Aβ40 alone as a function of time (**Supp. Fig. 3**). In general, the relative abundance of fibrils present in the presence of LUVs and/or EGCG-resveratrol corresponds to the ThT kinetic data.

We also performed immunoblotting to observe the abundance of pre-amyloid oligomers (using an A11 oligomer-specific antibody for both Aβ40 and αS) in the presence of LUVs and EGCG or resveratrol in order to gain additional mechanistic insight into their effects on early aggregation pathways (**Fig. 4**). The very short lag-phase of Aβ40 made discerning early oligomeric states difficult (**Fig. 4A and 4B**), however after 6 hours oligomeric abundances were observed to be lower in the presence of PCPG LUVs compared to Aβ40 alone. This may indicate that PCPG LUVs induce changes in oligomeric structure that are unrecognisable by anti-oligomeric antibodies, or alternatively, gives rise to faster aggregation leading to fewer oligomers present for antibody presentation (i.e. detection). Upon incubation with PCPG LUVs, αS showed a large increase in oligomer abundance in the early stages of fibril formation (after 24 hours, during the lag-phase). Then, similarly to Aβ40, a decrease in oligomer abundance was subsequently noted at 96 hours (**Fig. 4C and 4D**). Since the increase in oligomer abundance precedes the elongation-phase as shown by comparison with the unseeded αS ThT assay (**Supp. Fig. 4**), this suggests that PCPG LUVs increases oligomeric concentration and rapid fibril elongation. Interestingly, resveratrol increased the population of oligomers for both Aβ40 and αS when compared to the proteins in the presence of both resveratrol and PCPG. This is in contrast to Aβ40 and αS in the presence of EGCG, which showed lower oligomer formation compared to when incubated with both EGCG and PCPG, and further reinforces a different mechanism of action of the two inhibitors.

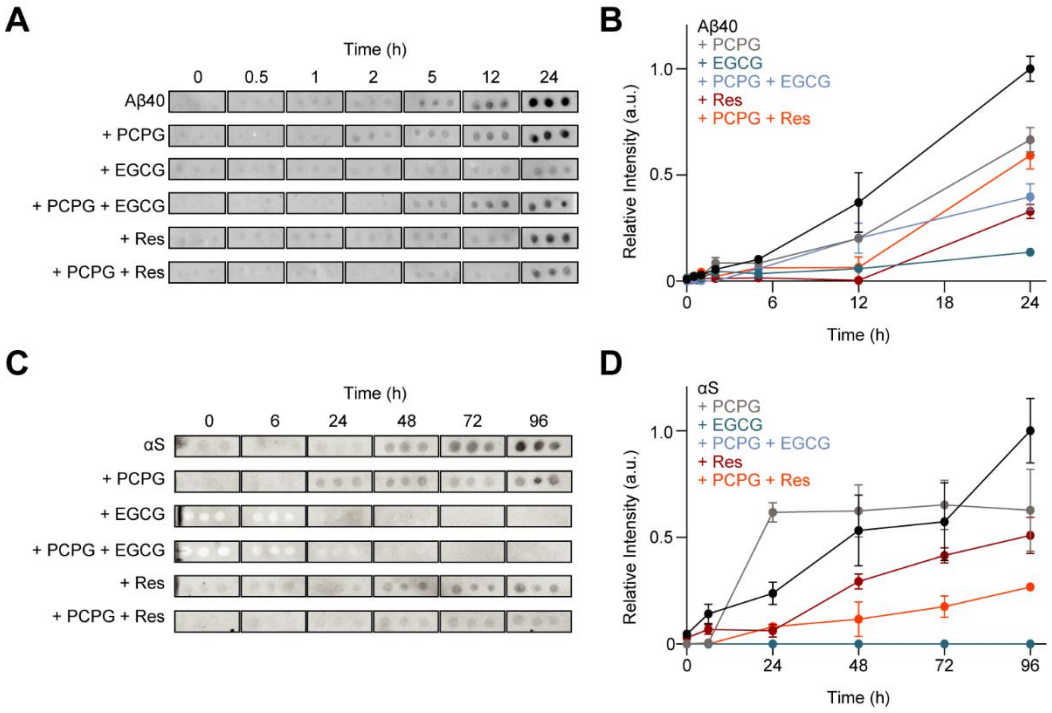


Figure 4. LUVs alter the abundance of αS and Aβ40 oligomers in the presence of EGCG and resveratrol during fibril formation. The effect of LUVs on oligomer formation during aggregation in the absence of presence of EGCG (50 µM) and resveratrol (50 µM) was observed by immunoblotting. **(A)** Aβ40 (50 µM) was incubated in the presence and absence of EGCG or resveratrol and LUVs (1:30, Aβ40:LUV) and the abundance of oligomers was quantified by densitometry **(B). (C)** αS (50 µM) was incubated in the presence and absence of EGCG or resveratrol and LUVs (1:30, αS:LUV) and the abundance of oligomers was quantified by densitometry **(D).** Data are reported as mean ± SEM (n = 3).

3.4.4 IM-MS reveals structural changes induced by EGCG complexes with Aβ40 and αS monomers

Thus far, our data suggests that LUVs interact with amyloidogenic proteins early in the fibrilization pathway, potentially through changes in the structure of monomeric or low molecular weight oligomeric species. For this reason, we applied native IM-MS to interrogate structural features of soluble protein forms to provide further mechanistic insight into the combined effects of PCPG LUVs on aggregation and inhibition by EGCG and resveratrol. The nESI process gives rise to multiply charged protein ions, with the degree of protonation

dependent upon solvent-exposed surface area and hence protein foldedness [55]. In native IM-MS experiments, ions are separated not only based on their *m/z* ratio, but also on rotationally averaged CCS values for a given charge state; the longer that ion takes to traverse the ion mobility sector, the larger its rotationally averaged CCS and *vice versa*. IM arrival time distributions can therefore be extracted from ion populations of a given *m/z* to provide structural information based on overall size.

We first analysed any changes in the monomeric structure of Aβ40 due to the addition of EGCG, PCPG LUVs or both (**Fig. 5A**). EGCG and PCPG were primarily investigated here as they had the greatest effect on both inhibition and aggregation respectively, and resveratrol was not found to form stable complexes with either of the proteins (**Supp. Fig 5**). While PCPG showed little effect on the native Aβ40 mass spectrum, the addition of EGCG produced peaks corresponding to Aβ40:EGCG complexes with stoichiometries up to a Aβ40 1:3 EGCG ratio (**Fig. 5A**). However, in the presence of both EGCG and PCPG LUVs, only Aβ40 1:1 EGCG complexes were observed. This reduction in EGCG binding to Aβ40 potentially indicates a mechanism in which PCPG LUVs reduces EGCG inhibitor efficacy by inducing structural changes that decrease EGCG binding affinity.

To better understand the overall structural transitions of Aβ40 under these varying conditions, the arrival time distribution (ATDs) for the major charge states were extracted and compared. In the presence of PCPG LUVs, Aβ40 monomers adopt a more compact structure (i.e. shorter drift times) when compared with Aβ40 alone (**Fig. 5B**), where significant ATD deviations were observed particularly at lower charge states. In contrast, Aβ40 in the presence of EGCG showed similar ATDs compared with Aβ40 alone, though with a notable reduction in ATD widths. This reduction in ATD width is indicative of a decrease in structural heterogeneity, as the protein samples fewer conformations during separation by IM. When Aβ40 was in the presence of both EGCG and PCPG we observed an apparent combination of effects, with an increase in more compact species as well as a decrease in ATD widths, most notable for the lower charge states (M^{3+}).

Following this general evaluation of the effects of PCPG and EGCG on Aβ40 structural heterogeneity, we analysed the ATDs of the M^{4+} (as the base peak in the spectra) more closely to interrogate changes to the monomeric structure of Aβ40 in detail (**Fig. 5C**). The M^{4+} ion of Aβ40 alone showed three main structural populations with drift times centred at

approximately 4.5 ms, 6 ms and 7 ms. No stable complexes were detected between Aβ40 and PC or PG lipids, and only small changes in monomeric structure were observed for the M^{4+} ion in the presence of PCPG LUVs. Following incubation with EGCG, only small structural changes in monomeric Aβ40 were observed, indicated by the disappearance of the most unfolded population at 7 ms and narrowing of the ATD, suggesting reduced heterogeneity.

Analysis of the EGCG:Aβ40 complexes revealed more significant changes in the drift time of the M^{4+} ions. For example, the 1:1 Aβ40:EGCG complex showed a large change from moderately compact structures with ATDs at approximately 6 ms towards more compact structures with an ATD centred at approximately 5 ms **(Fig. 5C)**, as well as the appearance of a more compact ion population at 4 ms. This shift in structural compaction was most clear for the 2:1 EGCG:Aβ40 complex showing a total transition to species with ATDs between 4-5.5 ms despite a large increase in mass (from a 4,330 Da monomer to a 5,247 Da complex). This implies that EGCG influences structural compaction upon binding, or is otherwise targeting more compact conformations for binding, which is consistent with previous observations by IM-MS and NMR [56,57]. Notably, an ATD feature centred at 7 ms was absent in the presence of EGCG but present when incubated with both EGCG and PCPG, along with a general broadening of the ATD. This suggests that PCPG restores the disordered state of Aβ40 (heterogeneity), while also simultaneously influencing transitions toward more compact structures, thereby increasing its propensity for aggregation. The competing mechanisms of EGCG and PCPG LUVs may in turn explain the decreased efficacy of fibril inhibitors in the lipid environment.

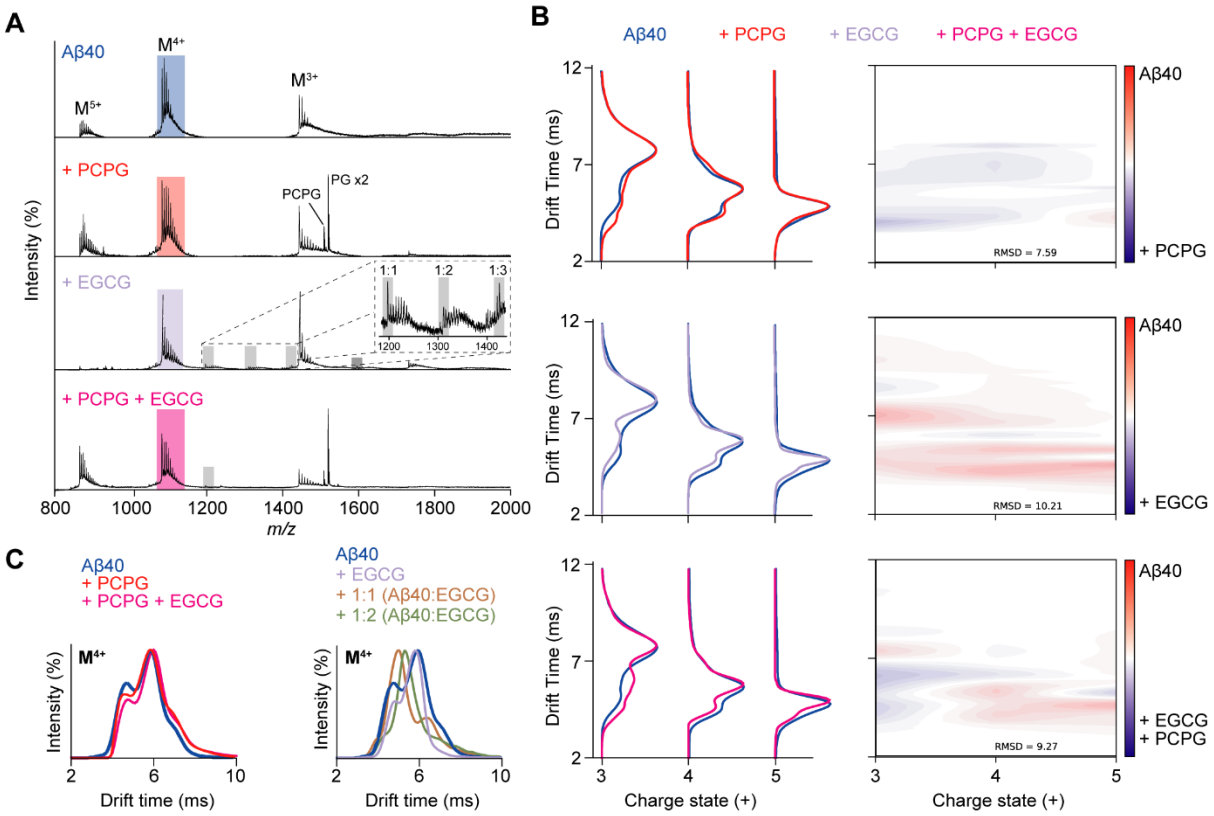


Figure 5. Native IM-MS unravels the structural heterogeneity of Aβ40 in the presence of EGCG and PCPG LUVs. (**A**) Native MS spectra of Aβ40 (50 μM) in the presence of PCPG (500 μM) and/or EGCG (50 μM). Complexes with EGCG (1:1, 2:1 and 3:1, EGCG:Aβ40) are primarily observed from the M^{4+} charge state (**A, inset**). (**B**) Heat map comparisons of Aβ40 ATDs for each charge state. Aβ40 compared to Aβ40 + PCPG (**C, left**), Aβ40 compared to Aβ40 + EGCG (**C, middle**) and Aβ40 compared to Aβ40 + PCPG + EGCG. (**C**) ATD analysis of Aβ40 M^{4+} in the absence of EGCG, M^{4+} in the presence of EGCG (50 μM), M^{4+} 1:1 Aβ40:EGCG complex and M^{4+} 1:2 Aβ40:EGCG complex.

Analogous native IM-MS experiments were performed to evaluate the effects of EGCG and PCPG LUVs on αS. Much like for Aβ40, αS was shown to form EGCG:αS complexes up to a 3:1. However, unlike for Aβ40, 1:1 lipid:protein complexes were also observed at lower charge states (M^{7+} and M^{8+}) under these conditions (**Fig. 6A**), although peak intensities were too small to obtain reliable ATDs or differentiate the identity of the lipid. Structural changes were then observed through shifts in ATDs when incubated with EGCG and PCPG, where extracted ATDs for each individual charge state were compared across the different conditions (**Fig. 6B**).

Overall, the ATD trends observed for αS were comparable with that of Aβ40. In the presence of PCPG LUVs, there was an increase in conformational heterogeneity and compaction as seen by the deviation in the heat maps towards lower drift times, which were most notable at lower charge states. The presence of EGCG induced αS to adopt little change in drift time except for a reduced peak intensity at 8 ms. Again, similarly to Aβ40, EGCG and PCPG LUVs induced ATD changes resembling a combination of effects from both EGCG and PCPG alone, although the EGCG effects seemed to slightly predominate which may hint at the greater inhibitory efficacy of EGCG toward αS fibrilization compared with Aβ40.

The 11+ charge state of αS was analysed by IM-MS (**Fig. 6C**) as it had the best compromise between IM resolution and low charge, which is typically indicative of native-like structure [55]. The M^{11+} ions of αS showed two distinct populations with ATDs at 8.6 ms and 9.9 ms, with the latter feature exhibiting greater intensity. Upon incubation with PCPG LUVs, the distribution of the αS structural populations shifted, preferencing the more compact structures with ATDs centred on 8.6 ms, and an increase in structural heterogeneity was observed. A more compact structure is often associated with the transition to an aggregation-prone state as the protein enters a lower energy, off-pathway structure prior to self-association [58–60], and hence these observations may rationalise the increased aggregation propensity (**Fig. 3A**). Similarly to Aβ40, the data suggested EGCG preferences binding to more compact structures as the 1:1 αS:EGCG ATD was observed at 8.6 ms, a phenomenon which has been previously observed [57]. Since compaction has been noted as a phenomenon occurring before the fibrillar aggregation of amyloid proteins, the high efficacy of EGCG may be due to targeting this process through complexation [60–62]. A reduction in the abundance of EGCG-protein complexes was again observed in the presence of PCPG LUVs, which, along with the competing structural effects of the inhibitor and lipids, provides further basis for decreased inhibitor efficacy in the presence of LUVs.

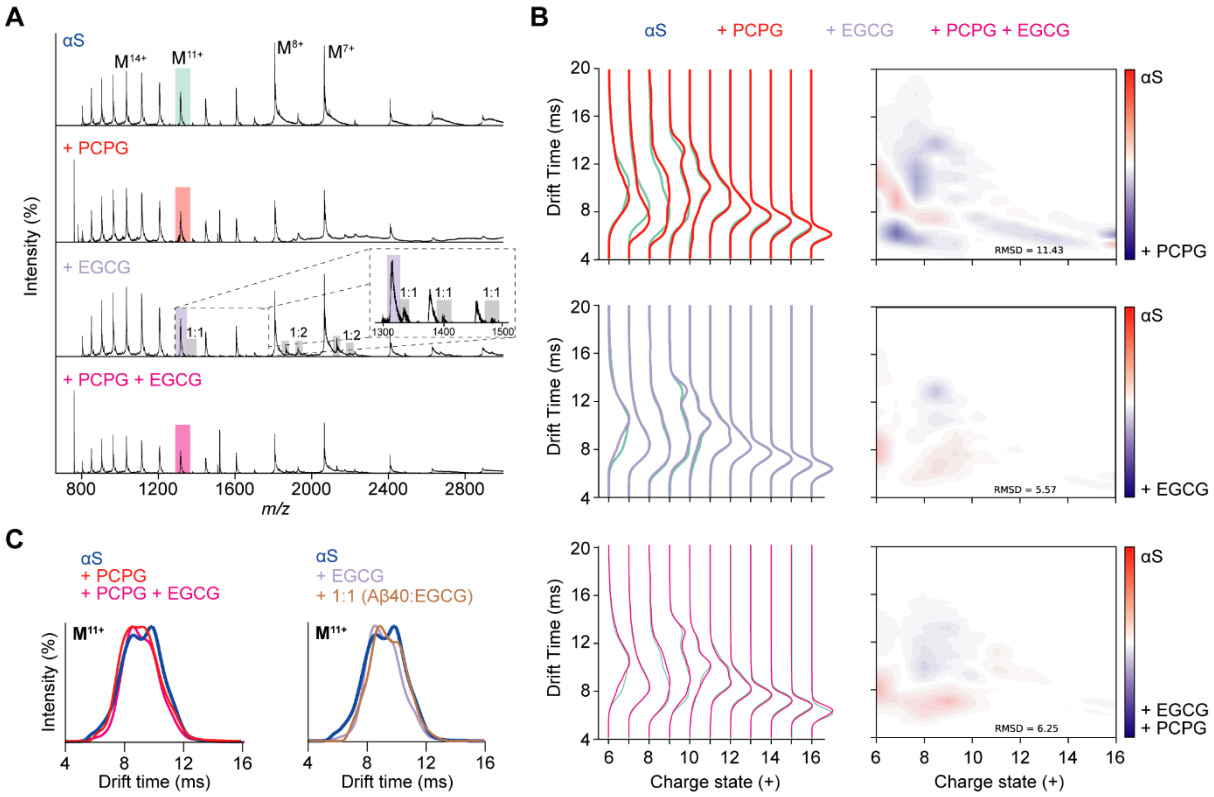


Figure 6. Native IM-MS examines the structural heterogeneity of αS in the presence of EGCG and PCPG LUVs (**A**) Native MS spectra of αS (50 µM) in the presence of PCPG (500 µM) and/or EGCG (50 µM). Complexes with EGCG (1:1, 2:1 EGCG:Aβ40) are primarily observed (**A, inset**). (**B**) Heatmap comparisons of αS ATDs for each charge state. αS vs αS + PCPG (**C, left**), αS vs αS + EGCG (**C, middle**) and αS vs. αS + PCPG + EGCG. (**C**) ATD analysis of 11+ αS monomer in the absence of EGCG, 11+ αS monomer in the presence of EGCG (50 µM) and 11+ 1:1 αS:EGCG complex.

3.5 Discussion

Effects of the lipid membrane are vitally important considerations in the fibrillar aggregation process which have only gained recent attention and focus. Lipid binding is important for both Aβ40 and αS in their functional states. For example, amyloid precursor protein, the protein from which Aβ40 is derived through proteolysis, is a membrane bound protein [25,63] and αS is thought to play a role as a lipid vesicle tethering protein via an N-terminal double-anchor mechanism [64–66]. Many studies have suggested that lipid membranes initiate aggregation

of Aβ40 and αS, acting as a catalytic surfaces for misfolding and fibrilization [20,67,68]. Since potential small molecule inhibitors may be competing against this process, it is important to consider their activity in the lipid environment. On top of this, the aetiology of misfolding diseases is also thought to be due to membrane interactions of misfolding monomer and/or oligomers through disruption, permeation and lysis of neuronal membranes [11,12,69,70]. Therefore, a better understand of this interplay is important to lead to successful therapy design. Given that recent work has shown the effectiveness of EGCG as a fibril inhibitor is environment dependent, including considerations such as oxidation status and pH [71], we investigated how lipid environments may affect the efficacy of two well-known inhibitors of fibrillization, EGCG and resveratrol, towards Aβ40 and αS fibril formation.

We have demonstrated that both PC and PCPG LUVs increase the rate of Aβ40 aggregation. However, this effect was more significant in the anionic lipid environment, supporting current opinions that suggest electrostatic effects play an important role in protein-lipid interactions and thereby have a more pronounced effect on aggregation propensity [18,22,72]. Similarly, LUVs increased the rate of αS fibril elongation by 8 times in the absence of preformed seed fibrils compared with a 1.1 times increase in elongation rate for seeded αS aggregation. This is consistent with a mechanism in which the lipids primarily influence aggregation in the nucleation phase through a pathway dependent on lipid composition and relative concentration of aggregating protein [20,26,73]. TEM data indicated LUVs were bound to fibrils, potentially due to elongating fibrils propagating from a localised protein concentration on the LUV surface. If electrostatic interactions increase the local concentration of protein on the LUV surface to a greater extent, this would explain the increased elongation rates in the presence of PCPG over PC LUVs, especially given that both Aβ40 and αS have been shown to interact with anionic LUVs preferentially [22,74,75].

Importantly, we have also demonstrated that the lipid environment drastically reduces the ability of EGCG and resveratrol to prevent fibril formation in both Aβ40 and αS systems. Anionic PCPG LUVs again had a more significant effect than zwitterionic PC LUVs in the case of EGCG and Aβ40 aggregation, although both LUVs reduced resveratrol inhibition equally. These differences highlight the different inhibition mechanisms of EGCG and resveratrol, where EGCG inhibits in the nucleation phase and can remodel fibrils, while resveratrol caps fibrils early in the elongation phase [30–33,36–38].

Immunoblotting revealed the abundance of oligomeric αS was much higher just prior to elongation when in the presence of PCPG LUVs. Recent research has shown the highest energy barrier in fibril elongation is the transition from oligomer to immature fibrils, and the majority of these oligomers do not form fibrils [76]. A higher oligomer concentration would therefore be expected to increase the probability of oligomer to fibril transitions, denoting a mechanistic explanation for the increased elongation rate. Interestingly, oligomer abundance was higher in the presence of resveratrol alone than both resveratrol and PCPG, which was in contradiction to the analogous conditions with EGCG and PCPG. This suggests EGCG decreases oligomeric concentrations more effectively than resveratrol. The increased oligomeric concentrations due to PCPG LUVs may outpace the ability for resveratrol to effectively cap fibrils, leading to more fibril formation than PCPG with EGCG.

Native MS revealed Aβ40 and αS formed complexes with EGCG while resveratrol did not, consistent with previous work [56,57]. However, the abundance of EGCG-protein complexes was notably reduced in the presence of lipid. In addition, native IM-MS data showed PCPG LUVs increased the extent of structural compaction across multiple charge states as well as increasing structural heterogeneity for both Aβ40 and αS. Compaction is a phenomenon previously suggested to initiate aggregation [60–62], while increased conformational heterogeneity is consistent with the formation of additional pre-amyloid states that provide greater probability for oligomer to fibril transitions [76]. In contrast, EGCG was shown to decrease the structural heterogeneity across charge states. If EGCG more specifically targets species early on the aggregation pathway (i.e. monomeric and smaller oligomeric species), it may intercept the mechanisms that increase nucleation as a result of the lipid environment. This may also explain why LUVs had a greater effect on resveratrol inhibition than EGCG, as resveratrol targets elongation more specifically and therefore cannot prevent the LUV effects [36,37].

In summary, we provide further evidence that lipid environments affect the structure and aggregation kinetics of amyloidogenic proteins, Aβ40 and αS, and add insight into inhibition mechanisms of EGCG and resveratrol in this environment. Importantly, LUVs appear to reduce the effects of these small molecule fibril inhibitors through competitive mechanisms affecting protein conformational states. In particular, the potent aggregation inhibitor EGCG complexes with monomers and reduces conformational heterogeneity while PCPG LUVs induce

structural compaction and increased heterogeneity of monomers for both Aβ40 and αS. Therefore, these opposing effects may explain a reduction in inhibition by LUVs. This study highlights the importance of the cellular membrane in the fibrillar aggregation of misfolded proteins, together with our relative lack of understanding on the structural effects that this environment induces on amyloid structure and dynamics. A lack of regard for the effects of the lipid environment during therapeutic design ultimately exemplifies how current developments have been ineffective in treating protein misfolding diseases.

3.6 Acknowledgements

H.M.S. was supported by a Faculty of Sciences Divisional Scholarship from the University of Adelaide. We thank Dr. Mandy Leung for her assistance with DLS data acquisition, and Adelaide Microscopy (University of Adelaide) for technical assistance with TEM. We also thank Flinders Analytical (Flinders University, Australia) for access to IM-MS instrumentation. This research was financially supported in part by an Australian Research Council Discovery Project Grant to T.L.P. (DP170102033).

3.7 Competing Interests

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

3.8 Author Contributions

H.M.S., B.J., and T.L.P. designed the research; H.M.S. and B.J. performed the experiments; H.M.S., B.J. and T.L.P. analysed experimental data; H.M.S., B.J. and T.L.P. wrote the manuscript.

1 Dobson CM (2004) Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol* **15**, 3–16.

2 Massi F & Straub JE (2001) Energy landscape theory for Alzheimer's amyloid β-peptide fibril elongation. *Proteins* **42**, 217–29.

3 Dodart J-C, Bales KR, Gannon KS, Greene SJ, DeMattos RB, Mathis C, DeLong CA, Wu S, Wu X, Holtzman DM & Paul SM (2002) Immunization reverses memory deficits without reducing brain Aβ burden in Alzheimer's disease model. *Nat Neurosci* **5**, 452–457.

4 Dahlgren KN, Manelli AM, Blaine Stine W, Baker LK, Krafft GA & Ladu MJ (2002) Oligomeric and fibrillar species of amyloid-β peptides differentially affect neuronal viability. *J Biol Chem* **277**, 32046–32053.

5 Lin WH, Ciccotosto GD, Giannakis E, Tew DJ, Perez K, Masters CL, Cappai R, Wade JD & Barnham KJ (2008) Amyloid-β peptide (Aβ) neurotoxicity is modulated by the rate of peptide aggregation: Aβ dimers and trimers correlate with neurotoxicity. *J Neurosci* **28**, 11950–11958.

6 Bieschke J, Herbst M, Wiglenda T, Friedrich RP, Boeddrich A, Schiele F, Kleckers D, Lopez del Amo JM, Grüning BA, Wang Q, Schmidt MR, Lurz R, Anwyl R, Schnoegl S, Fändrich M, Frank RF, Reif B, Günther S, Walsh DM & Wanker EE (2012) Small-molecule conversion of toxic oligomers to nontoxic β-sheet–rich amyloid fibrils. *Nat Chem Biol* **8**, 93–101.

7 Bucciantini M, Giannoni E, Chiti F, Baroni F, Taddei N, Ramponi G, Dobson CM & Stefani M (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**, 507–511.

8 Chen SW, Drakulic S, Deas E, Ouberai M, Aprile FA, Arranz R, Ness S, Roodveldt C, Guilliams T, De-Genst EJ, Klenerman D, Wood NW, Knowles TPJ, Alfonso C, Rivas G, Abramov AY, Valpuesta JM, Dobson CM & Cremades N (2015) Structural characterization of toxic oligomers that are kinetically trapped during α-synuclein fibril formation. *Proc Natl Acad Sci U S A* **112**, E1994–E2003.

9 Mannini B, Mulvihill E, Sgromo C, Cascella R, Khodarahmi R, Ramazzotti M, Dobson CM, Cecchi C & Chiti F (2014) Toxicity of protein oligomers is rationalized by a function combining size and surface hydrophobicity. *ACS Chem Biol* **9**, 2309–2317.

10 Ho CS, Khadka NK, She F, Cai J & Pan J (2016) Polyglutamine aggregates impair lipid membrane integrity and enhance lipid membrane rigidity. *Biochim Biophys Acta - Biomembr* **1858**, 661–670.

11 Bode DC, Freeley M, Nield J, Palma M & Viles JH (2019) Amyloid-β oligomers have a profound detergent-like effect on lipid membrane bilayers, imaged by atomic force and electron microscopy. *J Biol Chem* **294**, 7566–7572.

12 Serra-Batiste M, Ninot-Pedrosa M, Bayoumi M, Gairí M, Maglia G & Carulla N (2016) Aβ42 assembles into specific β-barrel pore-forming oligomers in membrane-mimicking environments. *Proc Natl Acad Sci* **113**, 10866–10871.

13 Perissinotto F, Rondelli V, Parisse P, Tormena N, Zunino A, Almásy L, Merkel DG, Bottyán L, Sajti S & Casalis L (2019) GM1 Ganglioside role in the interaction of Alpha-synuclein with lipid membranes: Morphology and structure. *Biophys Chem* **255**, 106272.

14 Kawahara M, Ohtsuka I, Yokoyama S, Kato-Negishi M & Sadakane Y (2011) Membrane incorporation, channel formation, and disruption of calcium homeostasis by Alzheimer's β-amyloid protein. *Int J Alzheimers Dis* **2011**, 1–17.

15 Flagmeier P, De S, Michaels TCT, Yang X, Dear AJ, Emanuelsson C, Vendruscolo M, Linse S, Klenerman D, Knowles TPJ & Dobson CM (2020) Direct measurement of lipid membrane disruption connects kinetics and toxicity of Aβ42 aggregation. *Nat Struct Mol Biol* **27**, 886–891.

16 Ugalde CL, Lawson VA, Finkelstein DI & Hill AF (2019) The role of lipids in -synuclein misfolding and neurotoxicity. *J Biol Chem* **294**, 9016–9028.

17 Jayasinghe SA & Langen R (2005) Lipid membranes modulate the structure of islet amyloid polypeptide. *Biochemistry* **44**, 12113–12119.

18 Terzi E, Hölzemann G & Seelig J (1995) Self-association of β-amyloid peptide (1-40) in solution and binding to lipid membranes. *J Mol Biol* **252**, 633–642.

19 Dikiy I & Eliezer D (2012) Folding and misfolding of alpha-synuclein on membranes. *Biochim Biophys Acta - Biomembr* **1818**, 1013–1018.

20 Galvagnion C, Buell AK, Meisl G, Michaels TCT, Vendruscolo M, Knowles TPJ & Dobson CM

(2015) Lipid vesicles trigger α-synuclein aggregation by stimulating primary nucleation. *Nat Chem Biol* **11**, 229–234.

21 Korshavn KJ, Satriano C, Lin Y, Zhang R, Dulchavsky M, Bhunia A, Ivanova MI, Lee Y-H, La Rosa C, Lim MH & Ramamoorthy A (2017) Reduced lipid bilayer thickness regulates the aggregation and cytotoxicity of amyloid-β. *J Biol Chem* **292**, 4638–4650.

22 Lin M-S, Chiu H-M, Fan F-J, Tsai H-T, Wang SSS, Chang Y & Chen W-Y (2007) Kinetics and enthalpy measurements of interaction between β-amyloid and liposomes by surface plasmon resonance and isothermal titration microcalorimetry. *Colloids Surf B* **58**, 231–236.

23 Seeligt J, Lehrmann R & Terzi E (1995) Domain formation induced by lipid-ion and lipid-peptide interactions. *Mol Membr Biol* **12**, 51–57.

24 Sani M-A, Gehman JD & Separovic F (2011) Lipid matrix plays a role in Abeta fibril kinetics and morphology. *FEBS Lett* **585**, 749–754.

25 Williams TL & Serpell LC (2011) Membrane and surface interactions of Alzheimer's Aβ peptide - insights into the mechanism of cytotoxicity. *FEBS J* **278**, 3905–3917.

26 Habchi J, Chia S, Galvagnion C, Michaels TCT, Bellaiche MMJ, Ruggeri FS, Sanguanini M, Idini I, Kumita JR, Sparr E, Linse S, Dobson CM, Knowles TPJ & Vendruscolo M (2018) Cholesterol catalyses Aβ42 aggregation through a heterogeneous nucleation pathway in the presence of lipid membranes. *Nat Chem* **10**, 673–683.

27 Naudí A, Cabré R, Jové M, Ayala V, Gonzalo H, Portero-Otín M, Ferrer I & Pamplona R (2015) Lipidomics of human brain aging and Alzheimer's disease pathology. *Int Rev Neurobiol* **122**, 133–89.

28 2019 Alzheimer's disease facts and figures (2019) *Alzheimer's Dement* **15**, 321–387.

29 Henríquez G, Gomez A, Guerrero E & Narayan M (2020) Potential role of natural polyphenols against protein aggregation toxicity: In vitro, in vivo, and clinical studies. *ACS Chem Neurosci* **11**, 2915–2934.

30 Freyssin A, Page G, Fauconneau B & Bilan A (2018) Natural polyphenols effects on protein aggregates in Alzheimer's and Parkinson's prion-like diseases. *Neural Regen Res* **13**, 955.

31 Huang Q, Zhao Q, Peng J, Yu Y, Wang C, Zou Y, Su Y, Zhu L, Wang C & Yang Y (2019) Peptide-Polyphenol (KLVFF/EGCG) binary modulators for inhibiting aggregation and neurotoxicity of amyloid-β peptide. *ACS Omega* **4**, 4233–4242.

32 Bieschke J, Russ J, Friedrich RP, Ehrnhoefer DE, Wobst H, Neugebauer K & Wanker EE (2010) EGCG remodels mature α-synuclein and amyloid-β fibrils and reduces cellular toxicity. *Proc Natl Acad Sci U S A* **107**, 7710–7715.

33 Ehrnhoefer DE, Bieschke J, Boeddrich A, Herbst M, Masino L, Lurz R, Engemann S, Pastore A & Wanker EE (2008) EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nat Struct Mol Biol* **15**, 558–566.

34 Corpas R, Griñán-Ferré C, Rodríguez-Farré E, Pallàs M & Sanfeliu C (2019) Resveratrol induces brain resilience against Alzheimer neurodegeneration through proteostasis enhancement. *Mol Neurobiol* **56**, 1502–1516.

35 Regitz C, Fitzenberger E, Mahn FL, Dußling LM & Wenzel U (2016) Resveratrol reduces amyloid-beta (Aβ1–42)-induced paralysis through targeting proteostasis in an Alzheimer model of Caenorhabditis elegans. *Eur J Nutr* **55**, 741–747.

36 Feng Y, Wang X ping, Yang S gao, Wang Y jiong, Zhang X, Du X ting, Sun X xia, Zhao M, Huang L & Liu R tian (2009) Resveratrol inhibits beta-amyloid oligomeric cytotoxicity but does not prevent oligomer formation. *Neurotoxicology* **30**, 986–995.

37 Fu Z, Aucoin D, Ahmed M, Ziliox M, Van Nostrand WE & Smith SO (2014) Capping of Aβ42 oligomers by small molecule inhibitors. *Biochemistry* **53**, 7893–7903.

38 Mishra R, Sellin D, Radovan D, Gohlke A & Winter R (2009) Inhibiting islet amyloid polypeptide fibril formation by the red wine compound resveratrol. *ChemBioChem* **10**, 445–449.

39 Muro E, Atilla-Gokcumen GE & Eggert US (2014) Lipids in cell biology: how can we understand them better? *Mol Biol Cell* **25**, 1819–1823.

40 Williams DM & Pukala TL (2013) Novel insights into protein misfolding diseases revealed by ion mobility-mass spectrometry. *Mass Spectrom Rev* **32**, 169–187.

41 Politis A, Stengel F, Hall Z, Hernández H, Leitner A, Walzthoeni T, Robinson C V. & Aebersold

R (2014) A mass spectrometry–based hybrid method for structural modeling of protein complexes. *Nat Methods* **11**, 403–406.

42 Sharon M & Robinson C V. (2007) The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu Rev Biochem* **76**, 167–193.

43 Volles MJ & Lansbury PT (2007) Relationships between the sequence of α-synuclein and its membrane affinity, fibrillization propensity, and yeast toxicity. *J Mol Biol* **366**, 1510–1522.

44 Grey M, Linse S, Nilsson H, Brundin P & Sparr E (2011) Membrane interaction of α-synuclein in different aggregation states. *J Parkinsons Dis* **1**, 359–371.

45 Ryan TM, Caine J, Mertens HDT, Kirby N, Nigro J, Breheney K, Waddington LJ, Streltsov VA, Curtain C, Masters CL & Roberts BR (2013) Ammonium hydroxide treatment of Aβ produces an aggregate free solution suitable for biophysical and cell culture characterization. *PeerJ* **1**, e73.

46 Kulig M & Ecroyd H (2012) The small heat-shock protein αB-crystallin uses different mechanisms of chaperone action to prevent the amorphous versus fibrillar aggregation of α-lactalbumin. *Biochem J* **448**, 343–352.

47 Hung VWS, Cheng XR, Li N, Veloso AJ & Kerman K (2013) Electrochemical detection of amyloid-beta aggregation in the presence of resveratrol. *J Electrochem Soc* **160**, G3097–G3101.

48 Kayed R, Head E, Thompson JL, McIntire TM, Milton SC, Cotman CW & Glabe CG (2003) Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **300**, 486–9.

49 Schneider CA, Rasband WS & Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675.

50 Liu J & Konermann L (2011) Protein–protein binding affinities in solution determined by electrospray mass spectrometry. *J Am Soc Mass Spectrom* **22**, 408–417.

51 Kantham S, Chan S, McColl G, Miles JA, Veliyath SK, Deora GS, Dighe SN, Khabbazi S, Parat MO & Ross BP (2017) Effect of the biphenyl neolignan honokiol on Aβ42-induced toxicity

in Caenorhabditis elegans, Aβ42 fibrillation, cholinesterase activity, DPPH radicals, and iron(II) chelation. *ACS Chem Neurosci* **8**, 1901–1912.

52 Lee JC-M, Bermudez H, Discher BM, Sheehan MA, Won Y-Y, Bates FS & Discher DE (2001) Preparation, stability, and in vitro performance of vesicles made with diblock copolymers. *Biotechnol Bioeng* **73**, 135–145.

53 Chandrashekaran IR, Adda CG, MacRaild CA, Anders RF & Norton RS (2011) EGCG disaggregates amyloid-like fibrils formed by Plasmodium falciparum merozoite surface protein 2. *Arch Biochem Biophys* **513**, 153–157.

54 McLaurin JA & Chakrabartty A (1997) Characterization of the interactions of Alzheimer β-amyloid peptides with phospholipid membranes. *Eur J Biochem* **245**, 355–363.

55 Konermann L, Ahadi E, Rodriguez AD & Vahidi S (2013) Unraveling the mechanism of electrospray ionization. *Anal Chem* **85**, 2–9.

56 Hyung SJ, Detoma AS, Brender JR, Lee S, Vivekanandan S, Kochi A, Choi JS, Ramamoorthy A, Ruotolo BT & Lim MH (2013) Insights into antiamyloidogenic properties of the green tea extract (-)-epigallocatechin-3-gallate toward metal-associated amyloid-β species. *Proc Natl Acad Sci U S A* **110**, 3743–3748.

57 Konijnenberg A, Ranica S, Narkiewicz J, Legname G, Grandori R, Sobott F & Natalello A (2016) Opposite structural effects of epigallocatechin-3-gallate and dopamine binding to α-synuclein. *Anal Chem* **88**, 8468–8475.

58 Smith DP, Radford SE & Ashcroft AE (2010) Elongated oligomers in β2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *Proc Natl Acad Sci U S A* **107**, 6794–6798.

59 Liu Y, Carver JA, Calabrese AN & Pukala TL (2014) Gallic acid interacts with α-synuclein to prevent the structural collapse necessary for its aggregation. *Biochim Biophys Acta - Proteins Proteomics* **1844**, 1481–1485.

60 Das S, Pukala TL & Smid SD (2018) Exploring the structural diversity in inhibitors of α-synuclein amyloidogenic folding, aggregation, and neurotoxicity. *Front Chem* **6**, 1–12.

61 Sanders HM, Jovcevski B, Carver JA & Pukala TL (2020) The molecular chaperone β-casein

prevents amorphous and fibrillar aggregation of α-lactalbumin by stabilisation of dynamic disorder. *Biochem J* **477**, 629–643.

62 Pavlova A, Cheng CY, Kinnebrew M, Lew J, Dahlquist FW & Han S (2016) Protein structural and surface water rearrangement constitute major events in the earliest aggregation stages of tau. *Proc Natl Acad Sci U S A* **113**, E127–E136.

63 Williams TL, Day IJ & Serpell LC (2010) The effect of Alzheimer's Aβ aggregation state on the permeation of biomimetic lipid vesicles. *Langmuir* **26**, 17260–17268.

64 Georgieva ER, Ramlall TF, Borbat PP, Freed JH & Eliezer D (2010) The lipid-binding domain of wild type and mutant α-synuclein: Compactness and interconversion between the broken and extended helix forms. *J Biol Chem* **285**, 28261–28274.

65 Fusco G, Pape T, Stephens AD, Mahou P, Costa AR, Kaminski CF, Kaminski Schierle GS, Vendruscolo M, Veglia G, Dobson CM & De Simone A (2016) Structural basis of synaptic vesicle assembly promoted by α-synuclein. *Nat Commun* **7**, 12563.

66 Killinger BA, Melki R, Brundin P & Kordower JH (2019) Endogenous alpha-synuclein monomers, oligomers and resulting pathology: let's talk about the lipids in the room. *npj Park Dis* **5**, 23.

67 Shrivastava AN, Aperia A, Melki R & Triller A (2017) Physico-pathologic mechanisms involved in neurodegeneration: Misfolded protein-plasma membrane interactions. *Neuron* **95**, 33–50.

68 Iyer A & Claessens MMAE (2019) Disruptive membrane interactions of alpha-synuclein aggregates. *Biochim Biophys Acta - Proteins Proteomics* **1867**, 468–482.

69 Sciacca MFM, Kotler SA, Brender JR, Chen J, Lee DK & Ramamoorthy A (2012) Two-step mechanism of membrane disruption by Aβ through membrane fragmentation and pore formation. *Biophys J* **103**, 702–710.

70 Fusco G, Chen SW, Williamson PTF, Cascella R, Perni M, Jarvis JA, Cecchi C, Vendruscolo M, Chiti F, Cremades N, Ying L, Dobson CM & De Simone A (2017) Structural basis of membrane disruption and cellular toxicity by α-synuclein oligomers. *Science* **358**, 1440–1443.

71 Sneideris T, Sakalauskas A, Sternke-Hoffmann R, Peduzzo A, Ziaunys M, Buell AK & Smirnovas V (2019) The environment is a key factor in determining the anti-amyloid efficacy of EGCG. *Biomolecules* **9**, 855.

72 Kremer JJ & Murphy RM (2003) Kinetics of adsorption of β-amyloid peptide Aβ(1-40) to lipid bilayers. *J Biochem Biophys Methods* **57**, 159–169.

73 Srivastava AK, Pittman JM, Zerweck J, Venkata BS, Moore PC, Sachleben JR & Meredith SC (2019) β-Amyloid aggregation and heterogeneous nucleation. *Protein Sci* **28**, 1567–1581.

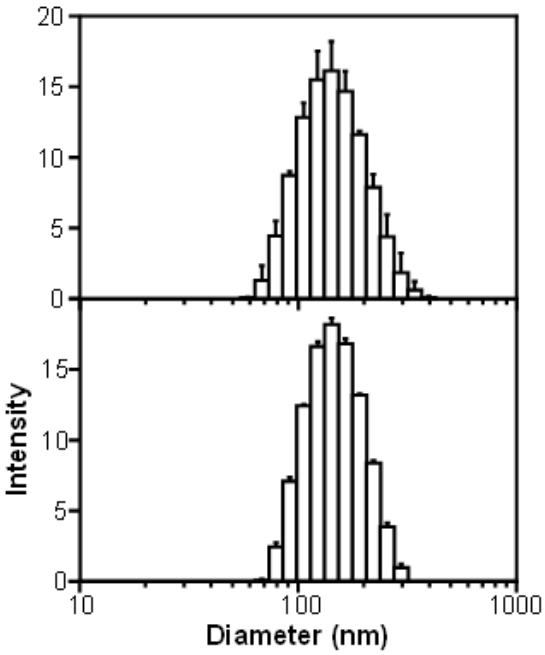
74 Bussell R & Eliezer D (2004) Effects of Parkinson's disease-linked mutations on the structure of lipid-associated α-synuclein. *Biochemistry* **43**, 4810–4818.

75 Rhoades E, Ramlall TF, Webb WW & Eliezer D (2006) Quantification of α-synuclein binding to lipid vesicles using fluorescence correlation spectroscopy. *Biophys J* **90**, 4692–4700.

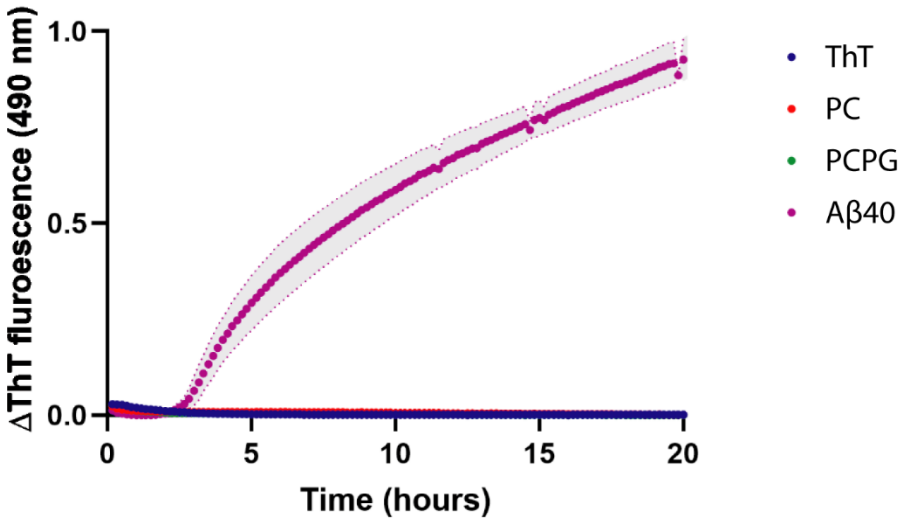
76 Michaels TCT, Šarić A, Curk S, Bernfur K, Arosio P, Meisl G, Dear AJ, Cohen SIA, Dobson CM, Vendruscolo M, Linse S & Knowles TPJ (2020) Dynamics of oligomer populations formed during the aggregation of Alzheimer's Aβ42 peptide. *Nat Chem* **12**, 445–451.

Supplementary Table 1. Rates and amyloid inhibition percentages derived from ThT assays (main text, Figures 1-3). All values except for seeded αS assays were derived from non-linear sigmoidal dose response models fit to relative maximum fluorescence. Seeded αS curves were fit with a one-phase exponential association model due to the lack of a lag phase. All assays were performed in triplicate and are reported as mean ± SEM.

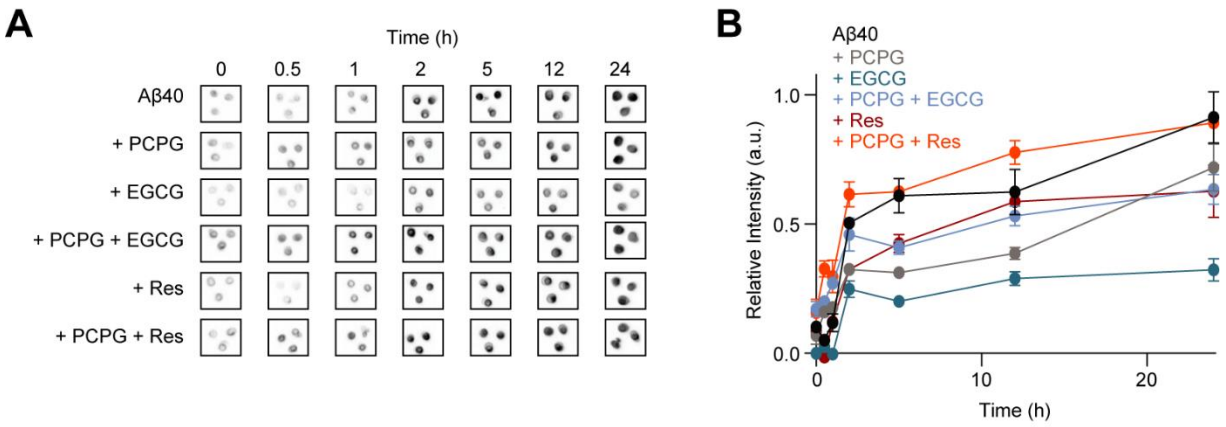
Condition	**Maximum fluorescence (Normalised ΔThT Fluorescence)**	**SEM**	**Rate of elongation (ΔThT Fluorescence per hour)**	**SEM**
Aβ40	0.994	0.0130	0.090	0.0167
+ PC	1.42	0.0121	0.161	0.0496
+ PCPG	1.93	0.00304	0.619	0.0275
Aβ40	1.01	0.0128	0.118	0.00795
Aβ40 + EGCG	0.0509	0.00294	0.354	0.0909
PC + EGCG	0.186	0.00606	0.328	0.0954
PCPG + EGCG	0.503	0.00755	0.546	0.172
Aβ40	1.03	0.0687	0.0722	0.0366
Aβ40 + resveratrol	0.290	0.00579	0.186	0.0669
Aβ40 + PC + resveratrol	0.576	0.00420	0.400	0.0408
Aβ40 + PCPG +	0.589	0.00310	0.803	0.082
Unseeded				
αS	0.982	0.0451	0.0153	0.00151
αS PCPG 1:30	1.06	0.00269	0.116	0.00634
αS EGCG 1:1	0.0312	0.0037	0.264	0.419
αS EGCG PCPG 1:30	0.0216	0.00444	0.387	0.253
αS Res 1:1	0.647	0.180	0.0265	0.00661
αS Res PCPG 1:30	1.00	0.0776	0.0912	0.00776
Seeded				
αS	0.684	0.00628	0.162	0.00666
αS PCPG 1:30	0.896	0.00988	0.198	0.0110
αS EGCG 1:1	0	0	0	0
αS EGCG PCPG 1:30	0	0	0	0
αS Res 1:1	0.308	0.0262	0.119	0.0367
αS EGCG PCPG 1:30	0.925	0.0219	0.126	0.0114



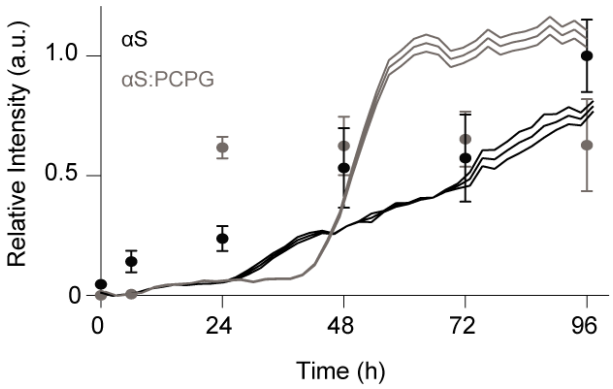
Supplementary Figure 1. Dynamic light scattering was performed to ensure PC (top) and PCPG (bottom) LUVs were of a homogenous size. LUVs were characterised at a concentration of 1.5 mM with a 633 nm laser at a back scattering angle of 173°.



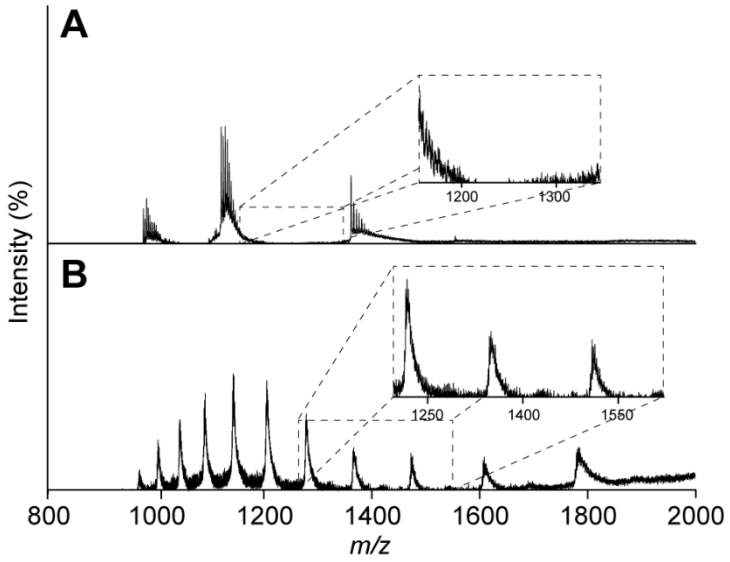
Supplementary Figure 2. A ThT fluorescence assay was performed to ensure the addition of lipids did not affect the ThT fluorescence under the conditions of the assays. Aβ40 (50 µM), PC and PCPG were separately incubated with ThT (50 µM) as well as ThT alone. Data are reported as mean ± SEM (n = 3).



Supplementary Figure 3. The effect of LUVs on fibril concentrations and fibril inhibition by EGCG (50 µM) and resveratrol (50 µM) was monitored through immuno-dot blot assays. **(A)** Aβ40 (50 µM) was incubated in the presence and absence of inhibitors and LUVs (1:30 Aβ40:LUV). **(B)** The relative concentrations of Aβ40 fibrils were quantified using densitometry. Data are reported as mean ± SEM (n = 3).



Supplementary Figure 4. LUVs alter the abundance of αS oligomers and increase the rate of fibril elongation. The effect of LUVs on oligomer formation during aggregation was observed by immunoblotting and ThT fluorescence. αS (50 μM) was incubated in the presence and absence of LUVs (1:30, αS:LUV) and the abundance of oligomers was quantified by densitometry **(circles)**. This data was compared with the effect of PCPG LUVs on unseeded αS fibril formation (150 μM) through ThT fluorescence assays in the presence and absence of LUVs (1:30 αS:LUV) **(lines)**. Data are reported as mean ± SEM (n = 3).



Supplementary Figure 5. Native MS shows resveratrol forms no stable complexes with Aβ40 or αS. (**A**) Native MS spectra of Aβ40 (50 μM) in the presence of resveratrol (50 μM). Aβ40 complexes with resveratrol were not observed (**A, inset**). (**B**) Native MS spectra of αS (50 μM) in the presence of resveratrol (50 μM). αS complexes with resveratrol were not observed (**B, inset**).

Chapter 4: Development of a modular synthetic protocol for protein crosslinkers

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

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

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

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

Disclaimer and author contributions.

The data presented in this chapter has been peer reviewed and is under revision in FEBS Journal.

The majority of the experiments, data analysis and manuscript writing were performed by Sanders, H. M.

Downey, K. assisted with baseline synthesis, data interpretation and experimental methods used from published thesis. Bubner, E. R. and Stephens, K. G. assisted in synthesis analysis and experimentation. Abell, A. D. and Pukala, T. L. assisted with experimental design, data interpretation and editing of the manuscript.

4.1 Abstract

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

4.2 Introduction

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

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

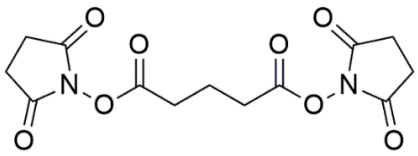


Figure 1. An example of a simple, commercially available cross-linker, disuccinimidyl glutarate (DSG), comprised of a non-cleavable 5 carbon chain between two lysine reactive NHS residues.

Following covalent cross-linking, as the length of the spacer arm is known, inter- and intra-protein upper distance constraints can be defined through site specific localisation of the cross-links.[3] These distance constraints allow for the interrogation of protein binding interfaces, intramolecular contacts and stoichiometry.[8] Furthermore, while techniques like crystallography require rigid structures for accurate models to be constructed, XL-MS can also capture regions that exhibit structural differences between conformational states, and can give information on the rearrangements and transient interactions of proteins which are often critical in controlling cellular processes.[2,3]

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

characteristic mass differences during MS analysis can improve linked peptide identification.[16]

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

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

4.3 Results and Discussion

4.3.1 Development of a modular synthetic protocol

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

motif to which enrichment or luminescent tags (or other functionalities) can be attached after cross-linking.

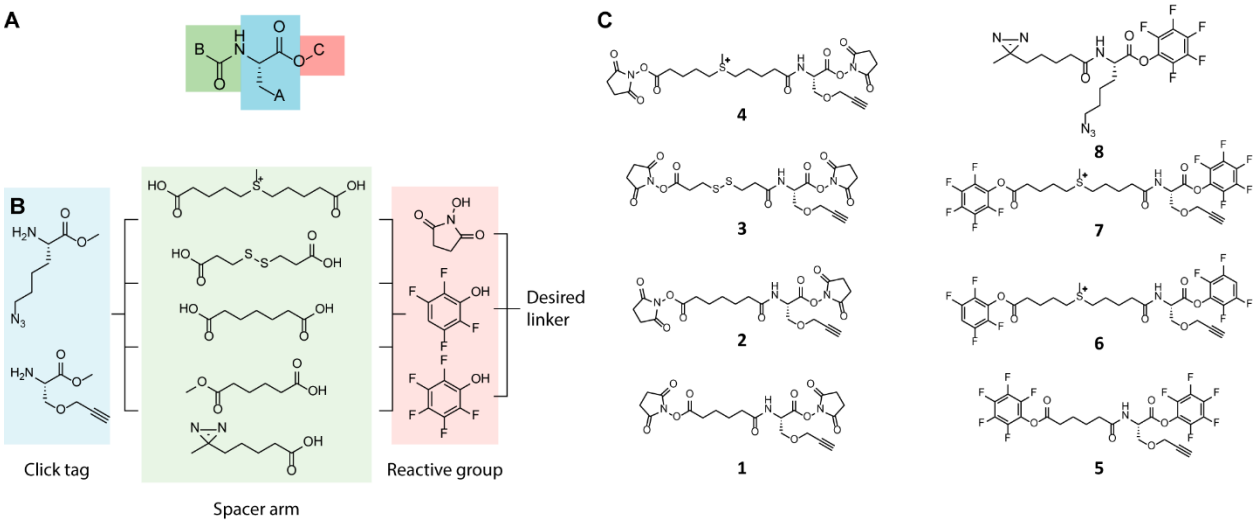


Figure 2. (**A**) The general structure of the modular linker includes 3 components; the 'click' modified amino acid core (**blue**), the spacer arm (**green**) and the reactive group (**red**), which can be synthesised in a combinatorial fashion (**B**). (**C**) Structures contained in the cross-linker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

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

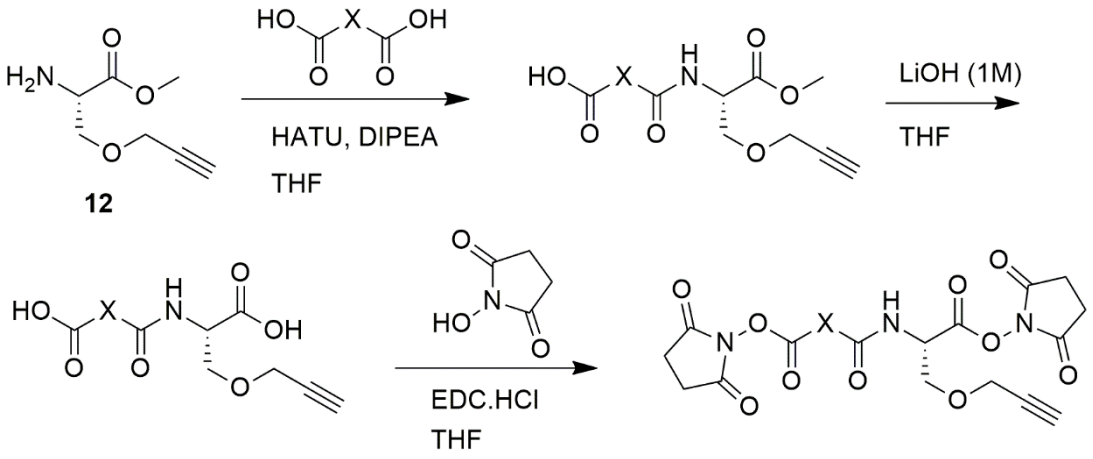
Following synthesis of the two base units, all subsequent reactions followed the same general reaction scheme wherein in the spacer-arm is coupled to the amino acid base unit by a peptide coupling, followed by a methyl ester hydrolysis and finally coupling of the reactive groups to the exposed acids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

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

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

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

Scheme 1. General synthetic protocol highlighting the modularity of the approach. Introducing a diacid (where X represents different spacer arm options) and reactive group of choice allows for synthesis of a custom linker.



4.3.2 Cross-linking and fragmentation characterisation

To demonstrate the reactivity of the synthesised library, a selection of amine reactive linkers was applied to a test peptide, acetylated AAKA (AcAAKA), which is a simple model system with a single primary amine available for cross-linking. Incubation of cross-linkers with AcAAKA at a high concentration allows for the simulation of the peptide linkages that could be expected in a proteolytic digest. Initially, compound **1** was used to form a cross-link between AcAAKA peptides and MS/MS analysis was performed on the linker peptide complex, the structure of which was confirmed by the expected fragmentation products (**Supp. Fig. 1**). Following this, the facile fragmentation of CID cleavable linkers was validated using a negative and positive ion mode cleavable linker (compounds **3** and **4** respectively). Both compounds were reacted with AcAAKA and MS/MS analysis revealed spectra consistent with expected dissociation patterns (**Fig. 2**). Compound **4** formed a 6-membered oxazoline ring under low energy CID, preferencing the non-serine side, yielding a 125 *m/z* mass difference between fragmentation products (**Fig. 3A**).[20] Capable instruments may then scan for this characteristic 125 *m/*z doublet during MS/MS under lower fragmentation energies and isolate each ion for higher energy MS sequencing experiments, allowing each linked peptide to be more easily identified and then sequenced, greatly improving site-specific localisation of the modification.[22] Similarly, the negative ion mode linker cleaved as expected forming a characteristic quartet fragmentation pattern, also suited to subsequent MS^3

sequencing experiments.[23] The facile fragmentation about the disufide bond in the linker chain is effected by either an enolate anion or by an anion situated directly adjacent to the disulfide.[24] There are therefore four potential products possible as fragmentation can take place on the serine or non-serine side of the disulphide bond, although the relative abundance of these signals in the MS/MS spectra of the linked AcAAKA peptides did not reveal significant preference for either side in this example (**Fig. 3B**).

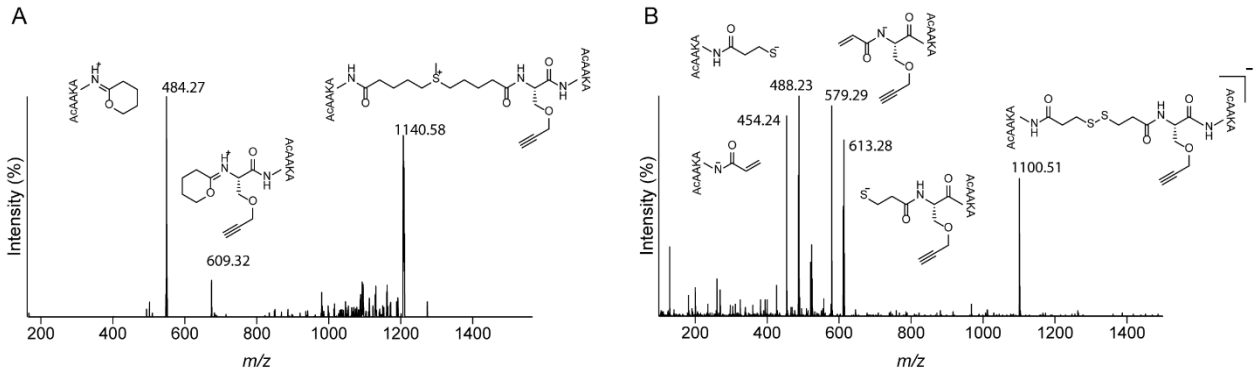


Figure 3. (**A**) Low CID energy of linker **3** cross-linked to AcAAKA yield two major daughter ions corresponding to 6-member oxazoline rings on the non-serine (484.27 *m/z*) or serine side (609.32 *m/z*) from the parent ion (1140.58 *m/z*). (**B**) The linker **4** parent ion (1100.51 *m/z*) yielded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

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

4.3.3 Post cross-link derivatisation of the modular linker

A feature of the modular linker design is the included alkyne/azide tag which can be derivatised subsequent to the reaction with proteins. This is advantageous since it avoids the inclusion of bulky groups such as biotin in the reactive linker, which can potentially restrict access to reactive sites on the protein during the cross-linking reaction. It also allows for flexibility in the choice of 'clickable' motif to include a range of functionalization. Alkyne and

azide groups on biotin, fluorescent tags, agarose UV-cleavable beads, peptides and oligos, amongst many other examples, are becoming increasingly available commercially, and the flexibility to apply several tags to a single cross-linked sample is attractive. After validation of covalent linkage, the ability for post-cross-linking derivatisation of the linkers through a copper catalysed Huisgen cycloaddition reaction was demonstrated with both biotin for enrichment and Cy3 for fluorescent visualisation.

Following covalent linkage between compound **1** and AcAAKA, a simple copper catalysed click reaction facilitated the addition of biotin azide. Evidence for the successful derivatisation was provided by MS, with MS/MS analysis of the expected $(M+H)^+$ ion of the product at 1294 *m/z* giving rise to expected fragmentation ions (**Fig. 4A**). Biotin affinity purification following this reaction allows for enrichment in a manner which is familiar and available to many biochemistry laboratories.

The flexibility offered by post-cross-linking derivatisation was further showcased using two different linkers, compound **5** and **6**. Both compounds and a commercially available control, disuccinimidyl sulfoxide (DSSO), were used to cross-link hen egg lysozyme and the resulting products were separated by SDS-PAGE and visualised using Coomassie Brilliant Blue stain (**Fig 4B**). Modification of the lysozyme monomer by cross-linkers was confirmed by MS (**Supp. Fig. 2**) with between 1 and 4 Linker 1 compounds covalently binding, whereas DSSO showed less dispersity with only 5 modifications. Despite differences in linker length that may affect lysine accessibilities, both custom linkers showed comparable cross-linking efficiency with the DSSO control, and dimer band densities were reflective of the monomer to dimer ratio in buffered solution.[25] Before PAGE, a sample of protein linked with compounds **5** and **6** was further covalently modified with cy3-azide by copper catalysed cycloaddition, and modifications were evidenced by an upward band shift in the gel due to the added mass of the cy3 tag. The gels were then visualised by cy3 fluorescence emission at 556 nm, following excitation at 553 nm. To emphasize how fluorescence tags could be used to increase visibility of hidden covalently modified protein complexes, smaller amounts of cy3-modified proteins were loaded on the gel to reduce the dimer intensity below limits of detection of the Coomassie stain. (**Fig 4B**). Notably, the band width of lysozyme:cy3-linker complexes increased, giving an indication of heterogenous mass changes due to attached cy3 tags. Upon irradiation and fluorescence

detection, the dimeric species that were undetectable using Coomassie staining alone were revealed by the cy3 dye.

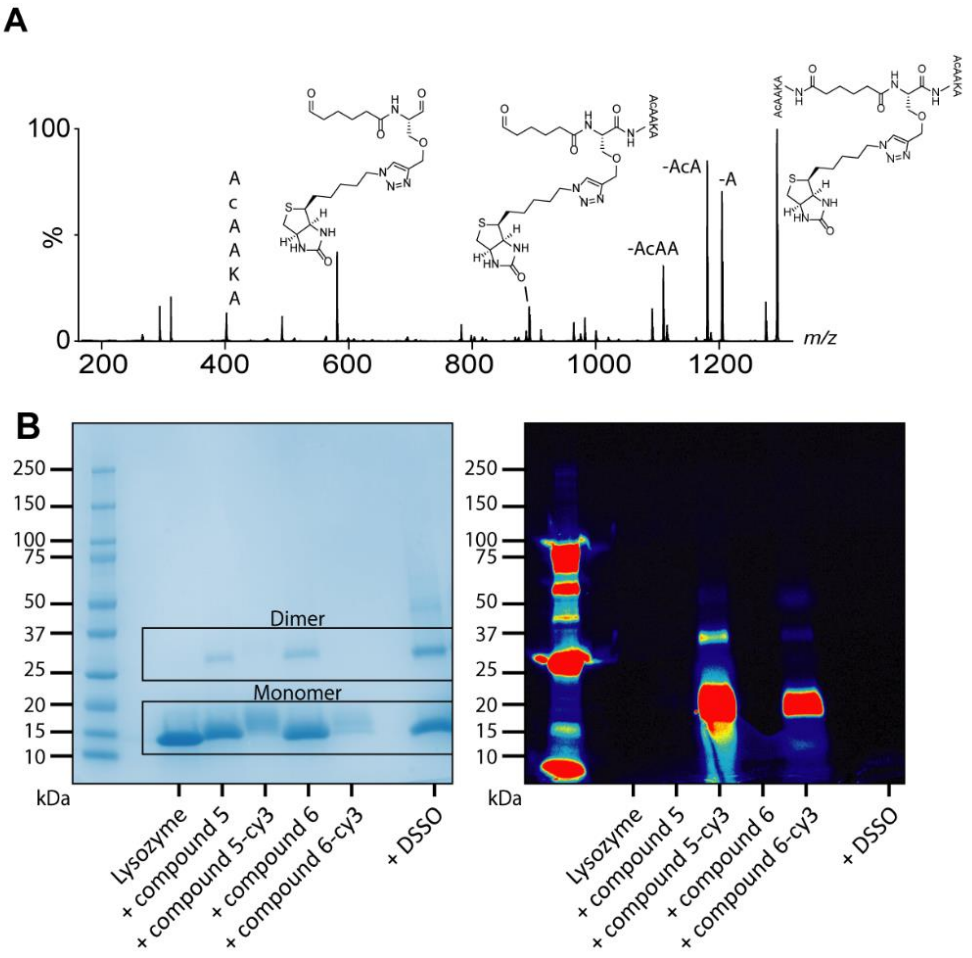


Figure 4. (**A**) CID was used to characterise cross-linked AcAAKA using linker **1** followed by post-linkage modification with biotin-azide using a Cu catalysed click cycloaddition reaction. The parent ion was identified (1293 *m/z*) and MS/MS analysis revealed characteristic fragment ions at 892 and 581 *m/z* that represent the parent ion minus one and two AcAAKA peptides, respectively. (**B, left**) SDS-PAGE revealed successful covalent linkage of lysozyme dimers using compounds **5** and **6** by comparison to a commercially available linker (DSSO), as well as post-linkage modification with cy3 (**B, right**). Fluorescence imaging highlights the ability to uncover hidden oligomers through dye derivatisation of linked samples.

The application of XL-MS has in recent years gone beyond *in vitro* samples and small numbers of interacting proteins, with cellular and tissue samples now subject to XL-MS for proteome

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

4.4 Conclusion

This study presents a modular synthetic protocol for the synthesis of protein cross-linking reagents capable of tackling the individual challenges associated with protein structure determination by XL-MS. The ability to exchange reactive groups and spacer arms as needed was showcased through the synthesis of a library containing 8 unique cross-linkers. The reagents analysed here displayed covalent linkages with model peptides and lysozyme with a comparable efficiency to that of commercially available cross-linkers. The introduction of cleavable spacer arms allows for improved identification strategies through diagnostic fragmentation peaks. In addition, centring the synthesis around an alkyne-/azide-modified amino acid allows for introduction of additional functionality through copper click chemistry, as we demonstrated through post-linkage modification of biotin for avidin-affinity purification or cy3 for improved visualisation through fluorescence. Overall, this synthetic protocol aims to reduce the complexity of XL-MS by offering a scaffold for greatly improving the diversity of

potential cross-linking reagents, intentionally designed for the highly variable requirements of the field.

4.5 Materials and methods

4.5.1 Materials and reagents

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

4.5.2 Synthesis of cross-linkers

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

4.5.3 Synthesis of compound 10

Triflic anhydride (25 mL, 146 mmol) was added to a solution of sodium azide (26.6 g, 410 mmol) in water (100 mL) while stirring at 0 °C. The reaction was allowed to warm to room temperature while stirring for 2 h, then extracted with dichloromethane (2 x 50 mL) and the combined organic extracts washed with saturated Na_2CO_3 (100 mL). The organic extract was added dropwise to a stirred solution of *N*-α-Boc-*L*-lysine (10.1 g, 41.1 mmol), $CuSO_4.5H_2O$ (1.11 g, 4.45 mmol) and K_2CO_3 (8.68 g, 62.8 mmol) in 2:1 (v/v) methanol/water (450 mL) at 0 °C. The solvent was removed under reduced pressure and the resulting residue diluted with

water (75 mL) and acidified by dropwise addition of 6 N HCl until the formation of a precipitate. The suspension was then diluted with potassium phosphate buffer (pH 6.2, 150 mL of 1:1 (v/v) 0.25 M K_2HPO_4/ 0.25 M KH_2PO_4) before further acidifying to pH 3 with 6 N HCl. The mixture was then extracted with ethyl acetate (2 x 150 mL). The combined organic extract was washed with water (100 mL) and brine (100 mL), then dried ($NaSO_4$) and the solvent removed under reduced pressure. The residue was dissolved chloroform (100 mL), filtered by suction filtration and the solvent removed under reduced pressure to obtain a yellow oil (10.4 g, 93%).

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

4.5.4 Synthesis of compound 12

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

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

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

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

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

4.5.5 General modular synthetic protocol

10 or **12** was dissolved in dry tetrahydrofuran (THF) (5 mL per 100 mg) with diisopropylethylamine (4.0 mol eq) while stirring. The desired spacer arm diacid (1.0 mol eq) was added to the solution followed by HATU (1.2 mol eq). The mixture was stirred overnight at room temperature under a nitrogen atmosphere. 1 M HCl (5 mL per 100 mg) was added and the solution was extracted with ethyl acetate (1.5 mL per 100 mg, 3×). The organic phase was washed with water (3×) and brine then dried over Na_2SO_4 and the solvent was removed *in vacuo*. The product was then purified by flash chromatography on normal phase silica to yield the desired amide.

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

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

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

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

4.5.7 Cross-linking of AcAAKA

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

4.5.8 Cross-linking of hen egg lysozyme

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

4.5.9 Copper click reaction

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

4.5.10 SDS–polyacrylamide gel electrophoresis

Lysozyme was cross-linked and examined by SDS–PAGE (12% gels) (Bio-rad, Hercules, CA, U.S.A.). Bands were visualised using Coomassie Brilliant Blue stain (ThermoFisher, Waltham,

MA, U.S.A.) according to the manufacturer's instructions. Fluorescence imaging was performed on a ChemiDoc gel imagine system (Bio-rad, Hercules, CA, U.S.A.) using an excitation wavelength of 553 nm and emission was measured at 556 nm.

4.5.11 Matrix-assisted laser desorption ionization mass spectrometry

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

4.6 Acknowledgements

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

4.7 Competing Interests

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

4.8 Author Contributions

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

4.9 References

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

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

(3) O'Reilly, F. J.; Rappsilber, J. Cross-Linking Mass Spectrometry: Methods and Applications in Structural, Molecular and Systems Biology. *Nat. Struct. Mol. Biol.* **2018**, *25* (11), 1000–1008. https://doi.org/10.1038/s41594-018-0147-0.

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

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

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

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

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

(9) Steigenberger, B.; Pieters, R. J.; Heck, A. J. R.; Scheltema, R. A. PhoX: An IMAC-Enrichable Cross-Linking Reagent. *ACS Cent. Sci.* **2019**, *5* (9), 1514–1522. https://doi.org/10.1021/acscentsci.9b00416.

(10) Petrotchenko, E. V.; Serpa, J. J.; Borchers, C. H. An Isotopically Coded CID-Cleavable

Biotinylated Cross-Linker for Structural Proteomics. *Mol. Cell. Proteomics* **2011**, *10* (2), 1–8. https://doi.org/10.1074/mcp.M110.001420.

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

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

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

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

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

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

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

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

https://doi.org/10.1021/acs.analchem.6b02240.

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

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

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

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

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

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

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

(26) Kaake, R. M.; Wang, X.; Burke, A.; Yu, C.; Kandur, W.; Yang, Y.; Novtisky, E. J.; Second, T.; Duan, J.; Kao, A.; et al. A New in Vivo Cross-Linking Mass Spectrometry Platform to Define Protein–Protein Interactions in Living Cells. *Mol. Cell. Proteomics* **2014**, *13* (12), 3533–3543. https://doi.org/10.1074/mcp.M114.042630.

(27) Matzinger, M.; Kandioller, W.; Doppler, P.; Heiss, E. H.; Mechtler, K. Fast and Highly Efficient

Affinity Enrichment of Azide-A-DSBSO Cross-Linked Peptides. *J. Proteome Res.* **2020**, *19* (5), 2071–2079. https://doi.org/10.1021/acs.jproteome.0c00003.

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

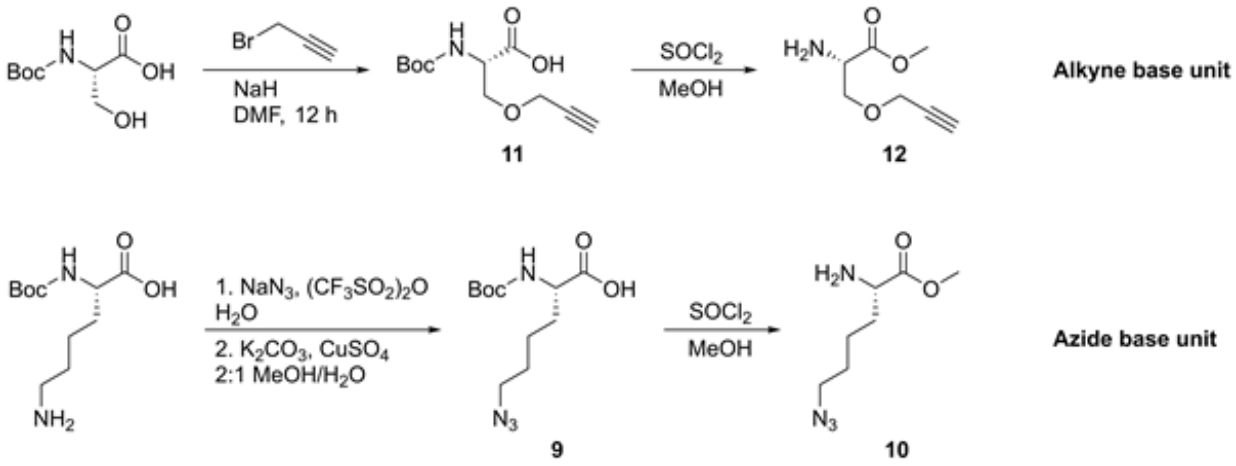
(29) Barman, A. K.; Gour, N.; Verma, S. Morphological Transition Triggered by Mannose Conjugation to a Cyclic Hexapeptide. *Arkivoc* **2012**, *2013* (2), 82–99. https://doi.org/10.3998/ark.5550190.0014.208.

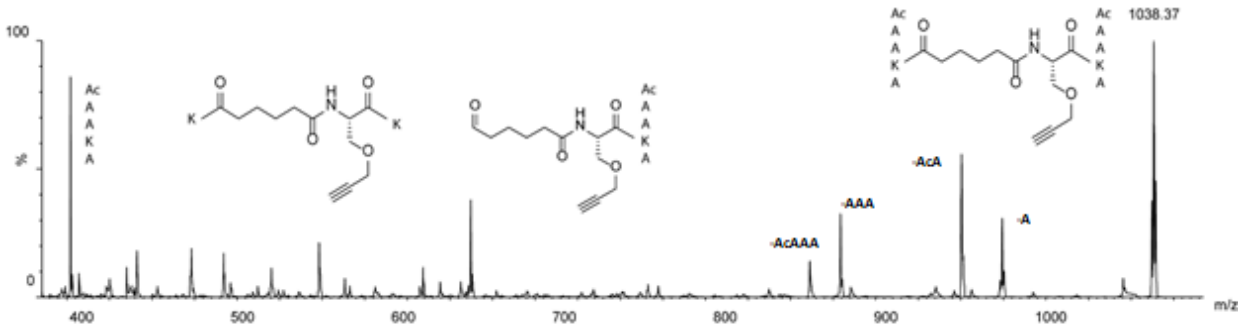
(30) Broadpharm. Click Chemsitry Protocols https://broadpharm.com/public/uploads/protocol_files/Click Chemistry Protocol.pdf (accessed Jun 25, 2020).

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

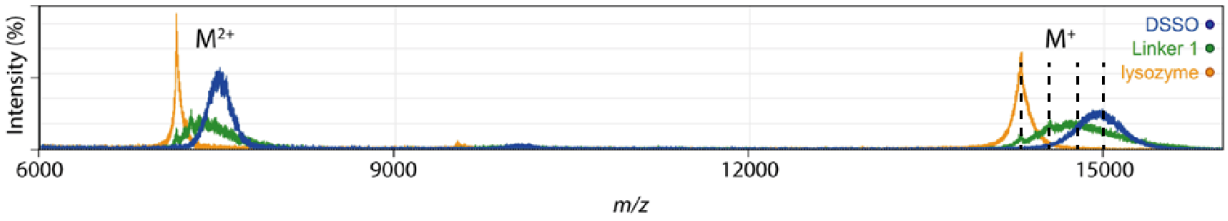
4.10 Supplementary Information

Supplementary Scheme 1.





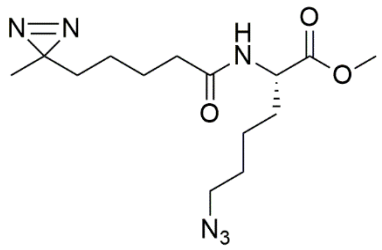
Supplemetary Figure 1. MS/MS analysis of AcAAKA peptide cross-linked by compound **1** (*m/z* $[M+H]^+$ = 1038). Expected fragmentation patterns were used to confirm the presence of the peptide-linker complex.



Supplementary Figure 2. MALDI spectrum shows doubly charged (approximately 7100 *m/z*) and singly charged (14,200 *m/z*) lysozyme. The protein is modified, increasing its mass, upon the covalent linkage with multiple cross-linkers. Dotted lines indicate additional modifications due to Linker 1.

4.10.1 Characterisation Data

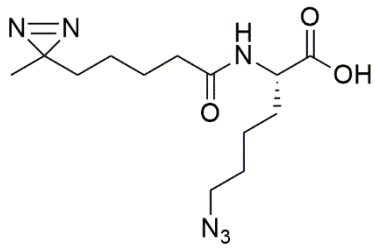
(S)-methyl 6-azido-2-(5-(3-methyl-3H-diazirin-3-yl)pentanamido)hexanoate



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

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

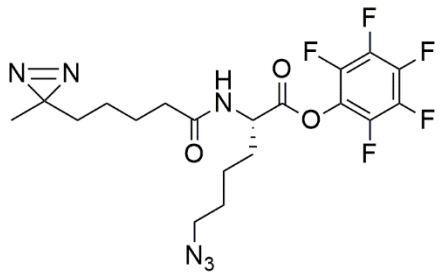
(S)-6-azido-2-(5-(3-methyl-3H-diazirin-3-yl)pentanamido)hexanoic acid



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

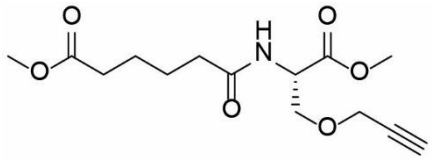
^{13}C NMR: (125 MHz, $CDCl_3$): δ 177.8, 176.5, 54.7, 53.7, 38.7, 36.5, 34.2, 31.0, 28.4, 27.7, 26.1, 25.0, 22.4 ppm.

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



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

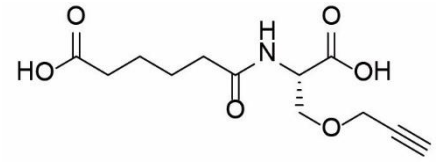
(S)-methyl 6-((1-methoxy-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-6-oxohexanoate



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

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

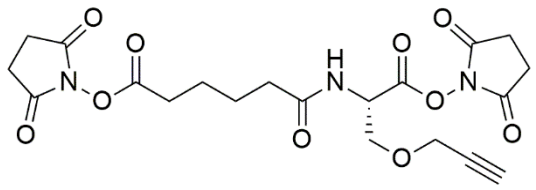
(S)-6-((1-carboxy-2-(prop-2-yn-1-yloxy)ethyl)amino)-6-oxohexanoic acid



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

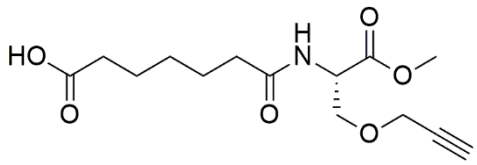
^{13}C NMR (125 MHz, $CDCl_3$, δ): 174.0, 172.5, 170.7, 79.0, 75.3, 69.6, 58.7, 52.8, 52.4, 51.7, 36.1, 33.9.

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



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

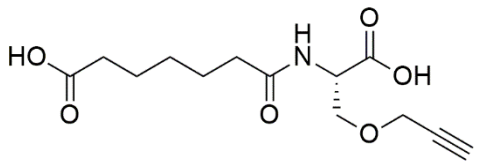
(S)-7-((1-methoxy-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-7-oxoheptanoic acid



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

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

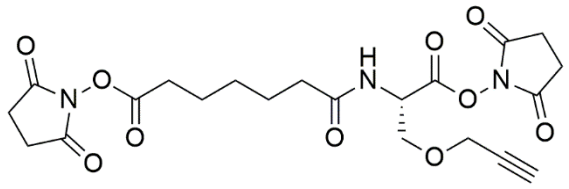
(S)-7-((1-carboxy-2-(prop-2-yn-1-yloxy)ethyl)amino)-7-oxoheptanoic acid



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

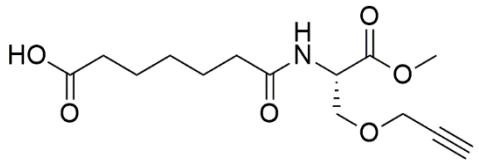
^{13}C NMR (125 MHz, CD_3OD, δ): 178.8, 177.4, 174.3, 81.4, 77.6, 71.5, 60.5, 55.1, 37.7, 36.0, 31.0, 27.8, 27.0.

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



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

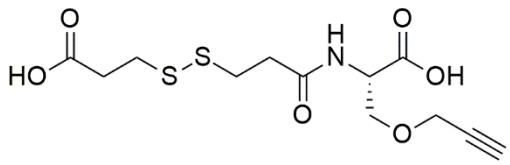
(S)-3-((3-((1-methoxy-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-3-oxopropyl)disulfanyl)propanoic acid



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

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

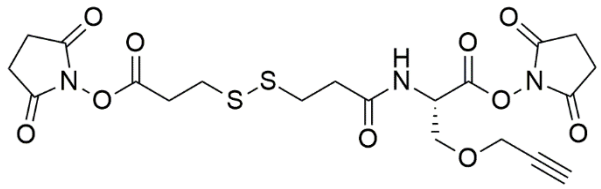
(S)-2-(3-((2-carboxyethyl)disulfanyl)propanamido)-3-(prop-2-yn-1-yloxy)propanoic acid



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

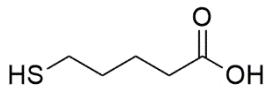
^{13}C NMR (125 MHz, CD_3OD, δ): 176.7, 176.7, 175.1, 81.4, 77.7, 71.5, 60.5, 55.2, 37.5, 36.2, 36.1, 35.6.

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



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

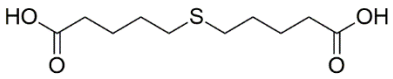
5-mercaptopentanoic acid



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

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

5,5'-thiodipentanoic acid

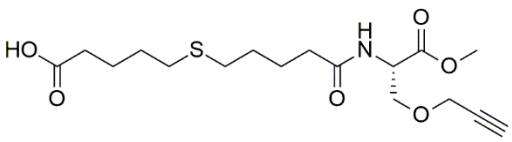


^{1}H NMR (500 MHz, $CDCl_3$, δ): 2.54 (t, *J* = 7.2 Hz, 4H), 2.40 (t, J = 7.2 Hz, 4H), 1.80 – 1.61 (m, 8H).

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

As NMR data as per literature[20]

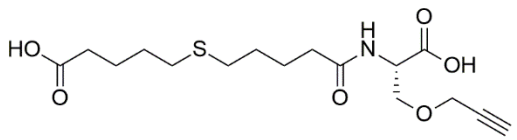
(S)-5-((5-((1-methoxy-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-5-oxopentyl)thio)pentanoic acid



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

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

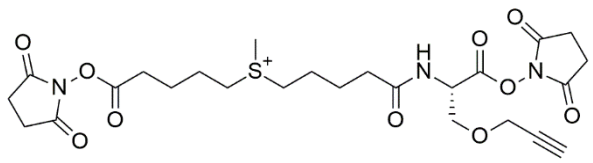
(S)-5-((5-((1-carboxy-2-(prop-2-yn-1-yloxy)ethyl)amino)-5-oxopentyl)thio)pentanoic acid



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

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

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



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

Chapter 5: Towards the development of a modular synthetic route for protein chemical cross-linking reagents

5.1 Introduction

During the conception and design of the synthetic strategy and production of resulting crosslinking reagents described in Chapter 4, a considerable amount of work was conducted that did not comprise the end product compiled for publication. This includes a range of additional reaction steps, alternative reactions and partially synthesised products, as well as additional linker reagents that were not included in the final manuscript. For this reason, this chapter serves as an supplement to discuss the broader synthetic approach and decisions made during the development of the synthetic procedure described in Chapter 4.

5.2 Development of a modular synthetic protocol

5.2.1 Linker design

To address the challenges associated with XL-MS, a synthetic route was designed that enabled the inclusion of different reactive groups, spacer-arms and affinity tags depending on the requirements of the experiment. Initially, the general linker design (**Fig. 1**) included NHS reactive groups, an exchangeable spacer-arm and an alkyne tag through which enrichment or post-linking functionalisation motifs could be introduced by 1,3 cycloaddition chemistry.

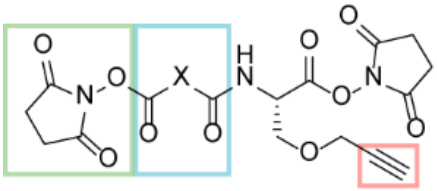


Figure 1: The general proposed cross-linker includes a reactive group (green), spacer arm (blue) and an alkyne unit with which affinity tags can be 'clicked' on through a Cu catalysed cycloaddition reaction. X represents different spacer arm options.

Retrosynthetic analysis of the general crosslinker devised in Chapter 4, **15**, was performed to explore synthetic routes toward synthesis (**Fig. 2**). The formation of **15** can be achieved by activating the carboxylic acid acids of a diacid, **14**, in the presence of NHS. The methyl diester,

13, can be formed through HATU-mediated peptide coupling between **12** and a methyl ester/acid, the length or composition of the carbon chain allows for tunability of the final spacer-arm component of the linker. Coupling **13** with a methyl ester/acid, rather than a diacid, ensures no di-addition side products are formed wherein **12** reacts on either side of the acid, as well as forming a diester **13** which reduces the compound polarity, making it easier to purify as the conjugate base can interact with the silica. N-Boc-protected serine is a cheap and commercially available amino acid, already protected for an O-alkylation with propargyl bromide to form **11**, that can in turn yield **12** with a one-pot thionyl chloride ($SOCl_2$) esterification, simultaneously deprotecting the amine and forming the methyl ester. **12** serves as the base unit for subsequent reactions with which different spacer arms can be attached.

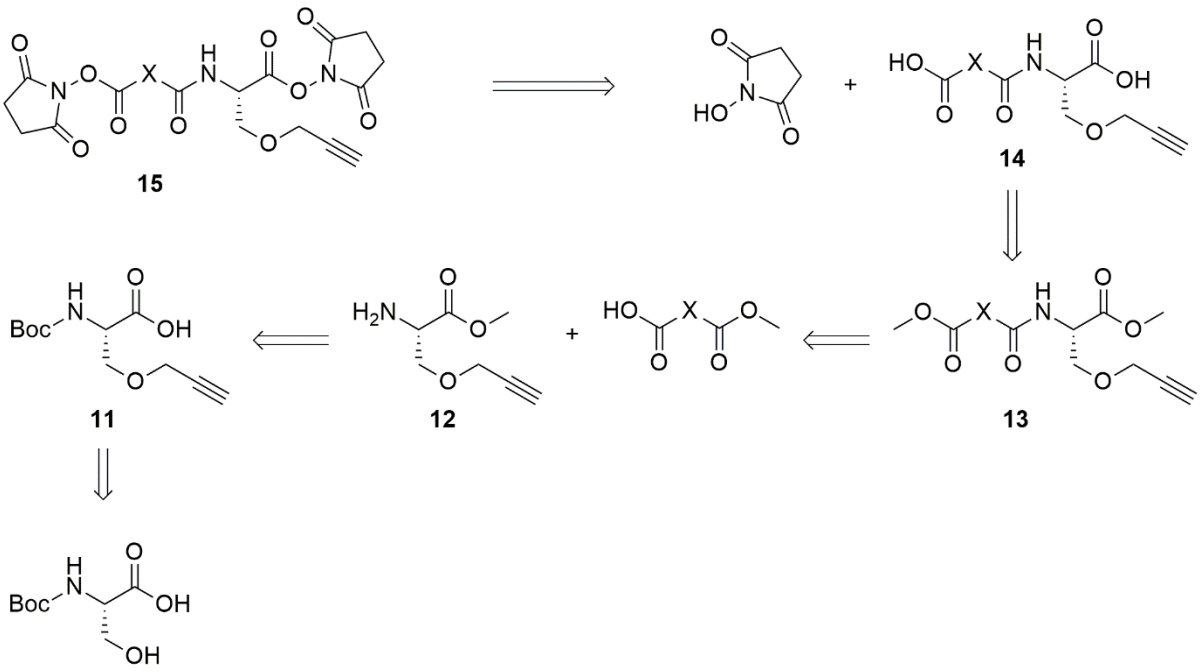


Figure 2: Retrosynthetic analysis of the proposed general linker.

Previous work centred around applications with a monomethyl adipate based spacer arm failed to yield the final product, linker **1** (described in Chapter 4), in a useful amount and only small quantities were detectable by MS.[1] To develop a modular synthetic protocol it was important to establish a synthetic route that was flexible enough to handle a variety of differing linker components with varied chemistries that may present challenges during synthesis. The work described here provides context for approaches that were trialled to

overcome the synthetic challenges of previous work and develop an adaptable synthetic route to create a cross-linker library that could be applied to biological samples.

5.3 Development of the synthetic route

To begin, a simpler spacer arm was applied to probe for issues in the synthetic procedure that were encountered in previous work.[1] Retrosynthetic analysis revealed that an amide coupling with succinic anhydride would yield a two-carbon length spacer arm with which NHS esterification conditions could be trialled (**Fig. 3**).

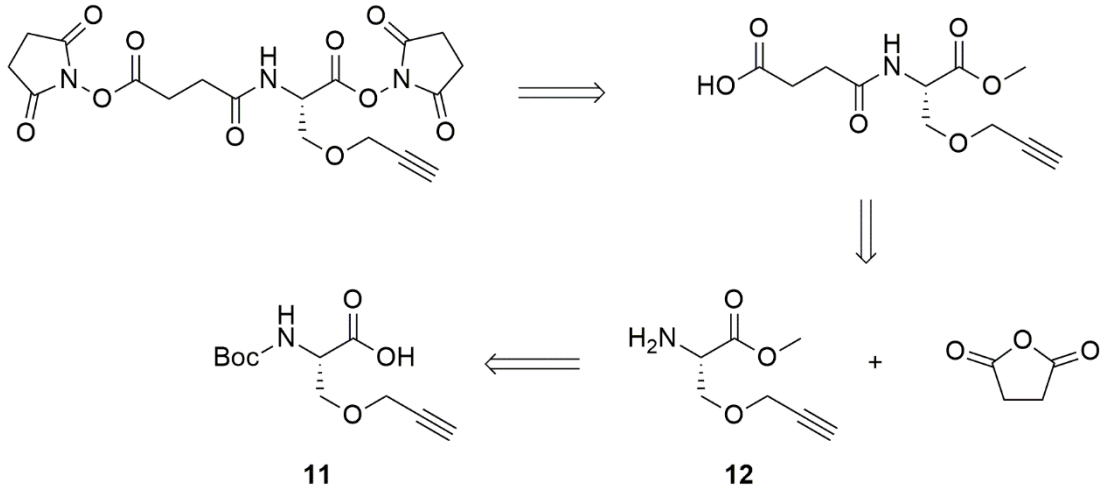


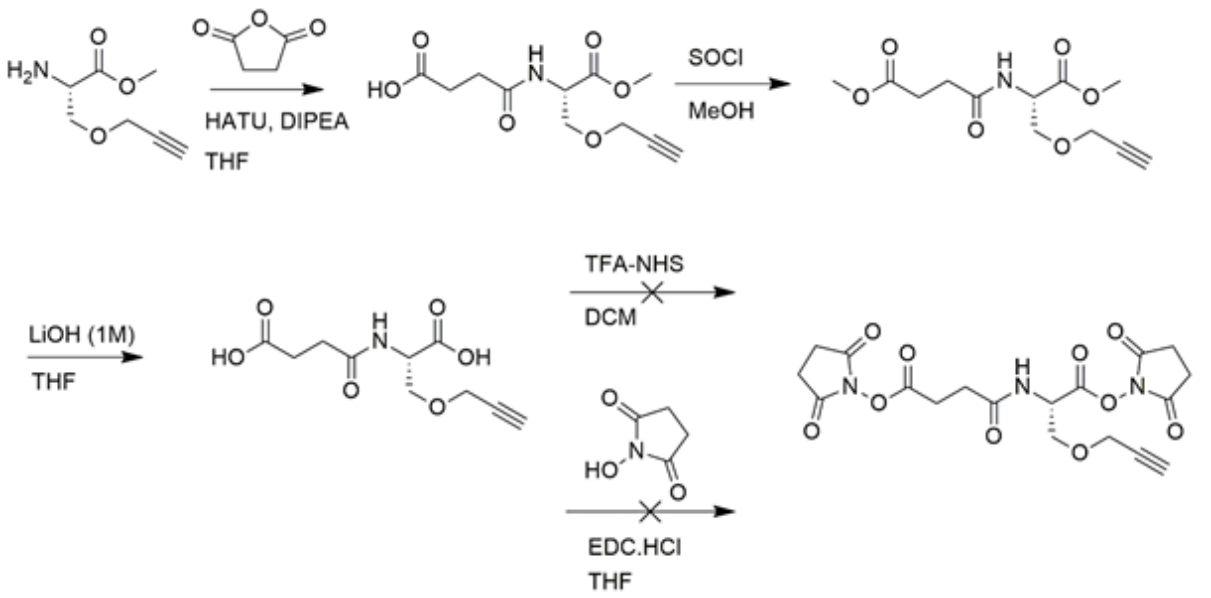
Figure 3: Retrosynthetic analysis of linker 1a.

5.3.1 Synthesis of succinimide linker

Initially, development of the synthetic protocol began by trialling synthesis of a simple linker which incorporated a succinimide spacer-arm. Synthesis of the succinimide linker (**Scheme 1**) successfully yielded the corresponding diacid but EDC coupling with NHS failed. Analysis by MS failed to produce the expect *m/z* and crude NMR showed no sign of the characteristic 8H singlet at approximately 2.7 ppm. This was consistent with previous work that described issues with the total synthesis of the cross-linker pertaining to the NHS esterification reaction.[1] To combat this, another reaction described by Rao, *et al.*[2] was trialled, using activated NHS in the form of trifluoroacetic acid-NHS (TFA-NHS) and pyridine to deprotonate

the acids for esterification. Unfortunately, this reaction too reaction yielded no product following MS and NMR analysis.

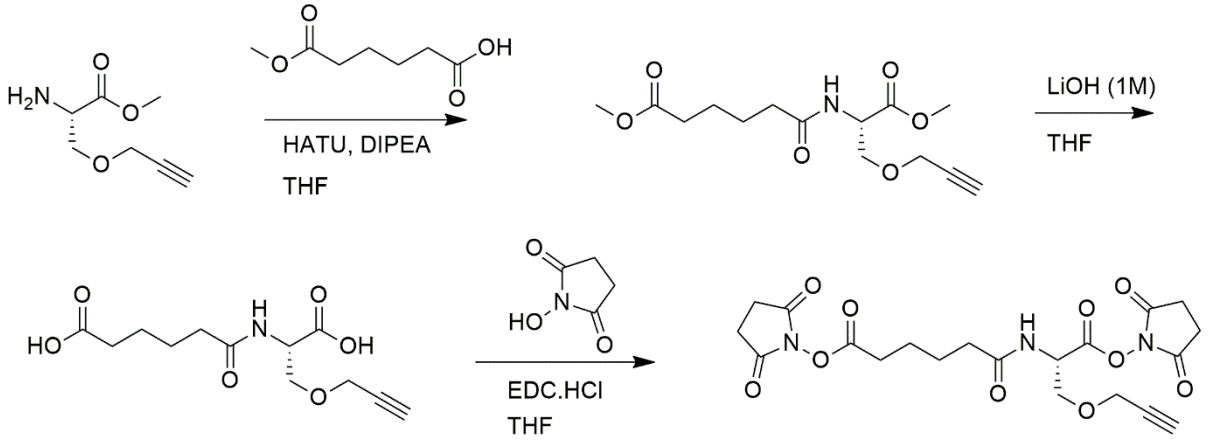
Scheme 1. Synthesis of succinimide linker



5.3.2 Synthesis of linker 1

Observing incomplete NHS esterification of the succinic anhydride-based linker, it was proposed that the length of the carbon chain may be too short, with the steric bulk of the serine-alkyne motif interfering with the NHS esterification reaction. For this reason, synthesis of a linker with a longer carbon chain was attempted. Replacing succinic anhydride with monomethyl adipate as the spacer arm would increase its length by two carbons, potentially creating space enough between the acids to allow for successful synthesis. Aside from being inexpensive and commercially available, monomethyl adipate also yields a di-methyl ester following HATU-mediated peptide coupling (**Scheme 2**), meaning a second esterification is not necessary in comparison with succinic anhydride. This synthesis successfully yielded linker **1** as confirmed by HRMS and was used immediately for cross-linking reaction, as described in Chapter 4.

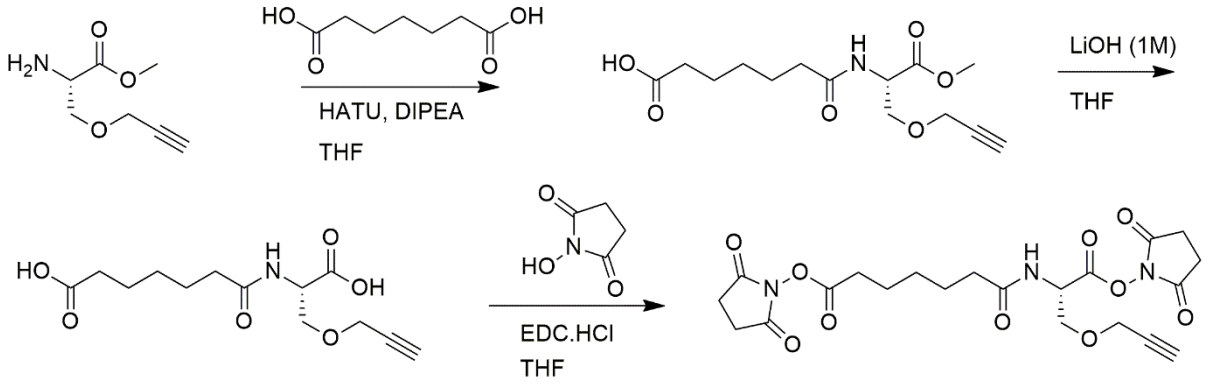
Scheme 2. Synthesis of linker **1**



5.3.3 Synthesis of linker **2**

Central to the proposed modular synthetic protocol is the ability to exchange spacer-arms to enable crosslinking reactions between protein sidechains that are further apart. Following the successful synthesis of linker **1,** a linker of a longer length was attempted using heptanedioic acid as a spacer-arm (**Scheme 3**). The synthesis followed the same procedure as described for linker **1** with the exception that, following the HATU coupling, the resultant amide was not protected by another methyl esterification and instead was purified with 50:50 EtOAc:petroleum ether with 0.1 % acetic acid to ensure protonation of the free acid. This allowed for successful separation of the product. Following ester hydrolysis, linker **2** was synthesised by coupling of NHS with EDC.

Scheme 3. Synthesis of linker **2**



5.4 Incorporation of a CID-cleavable spacer arm

Cross-linkers capable of CID-induced cleavage in both positive and negative ion mode were designed for synthesis. MS is capable of ionisation in both positive and negative modes. For this reason cross-linkers capable of labile fragmentation in both polarities were considered for design. Dithiobis(succinimidyl) propionate (DSP)[3] is a commercially available, negative-ion cleavable cross-linker that can be retrosynthetically broken down into NHS and 3,3'-dithiopropionic acid. The disulphide spacer arm is capable of fragmentation when negative ionisation forms either an enolate anion, or an anion situated directly adjacent the disulphide.[4] A characteristic, quartet fragmentation pattern is formed by each product, with spaces of 2 Da between the disulphide fragments and 66 Da between the enolate/disulphide fragments (**Fig. 4**).[4] Disuccinimidyl sulfoxide (DSSO)[5] and *S*-methyl 5,5'-thiodipentanoylhydroxysuccinimide[6] are positive-ion cleavable linkers that can be broken down to NHS and diacids by retrosynthetic analysis. DSSO rearranges upon protonation to form a characteristic doublet with a space of 50 Da, as the C-S bond is broken forming an enol and hydroxythio product. Upon protonation, *S*-methyl 5,5'-thiodipentanoylhydroxysuccinimide rearranges to form two potential products: a 6-membered oxazoline ring and a methyl ether product which forms a characteristic doublet with a 47 Da difference. Formation of the oxazoline ring requires an adjacent nitrogen which is derived from the amide bond formed during crosslinking with lysine, therefore linkage is necessary for fragmentation.

Important to the development of a general synthetic procedure is the easy incorporation of different spacer-arm moieties. The chosen linkers are derived from dicarboxylic acids and should therefore be available for coupling to the base unit, **12**, followed by NHS esterification.

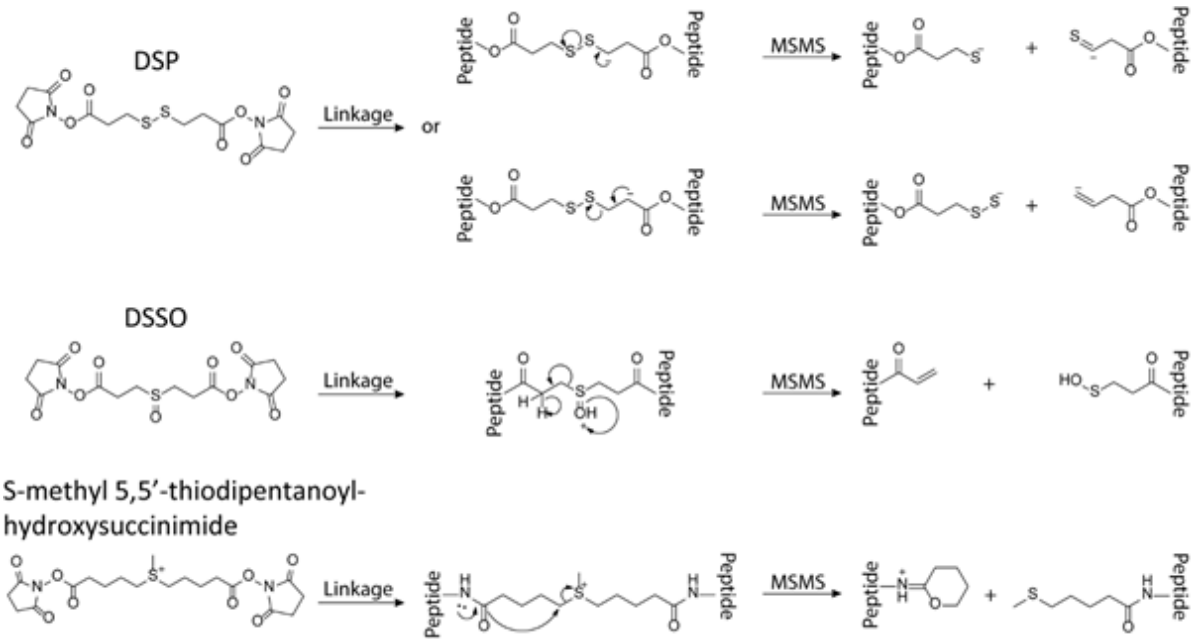


Figure 4: A schematic depicting the fragmentation pathways of three different MS labile linkers.

5.4.1 Linker **3**

Retrosynthetic analysis revealed that a disulphide spacer-arm, linker **3**, could be synthesised from a peptide coupling with the **12** but previous work highlighted that methyl ester hydrolysis with both LiOH and NaOH resulted in decomposition of the disulphide starting material,[1] a phenomenon that has been noted previously.[7,8] To avoid this, an alternative synthesis was first trialled whereby di-tert-butylazodicarboxylate (DBAD) protected 3-mercaptopropionic acid would be coupled with **12**. Following ester hydrolysis, introduction of free 3-mercaptoproionic acid would yield the disulphide before NHS esterification (**Fig. 5**).[9]

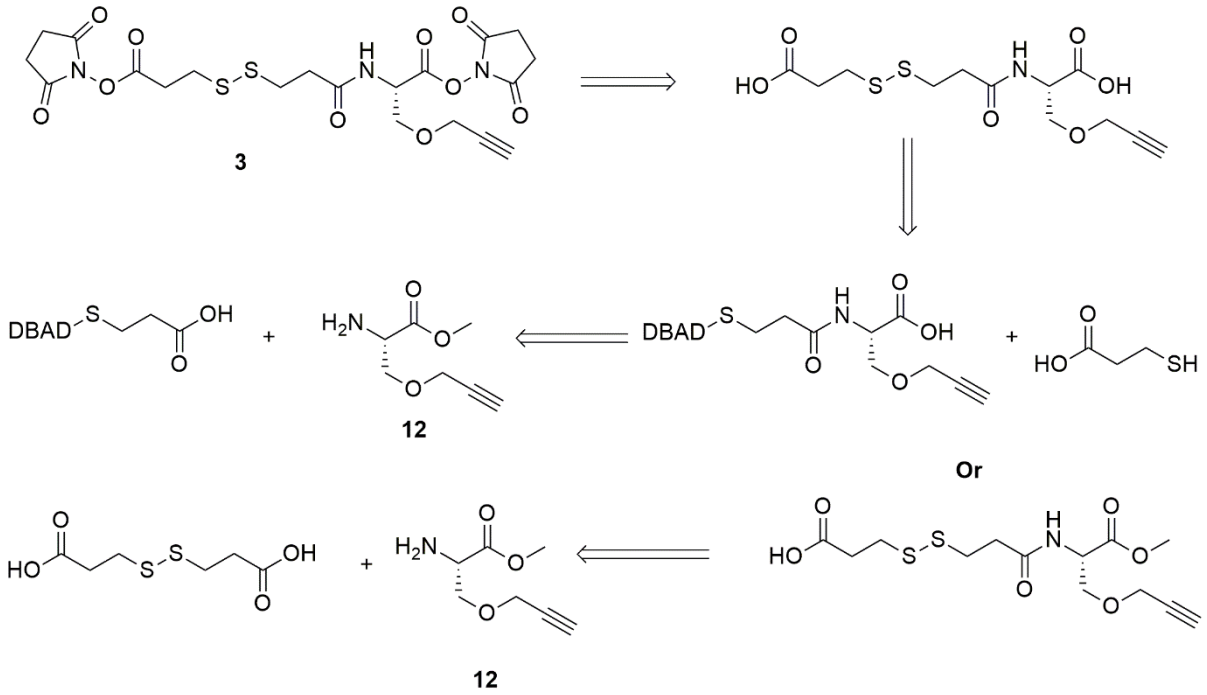
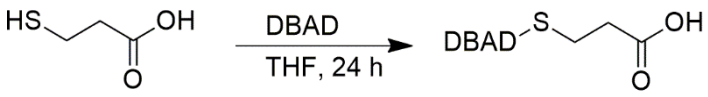


Figure 5. Retrosynthetic analysis showing synthesis of linker **3** is possible through two ways: using a DBAD protecting group and introducing a thioacid or via a coupling with a diacid.

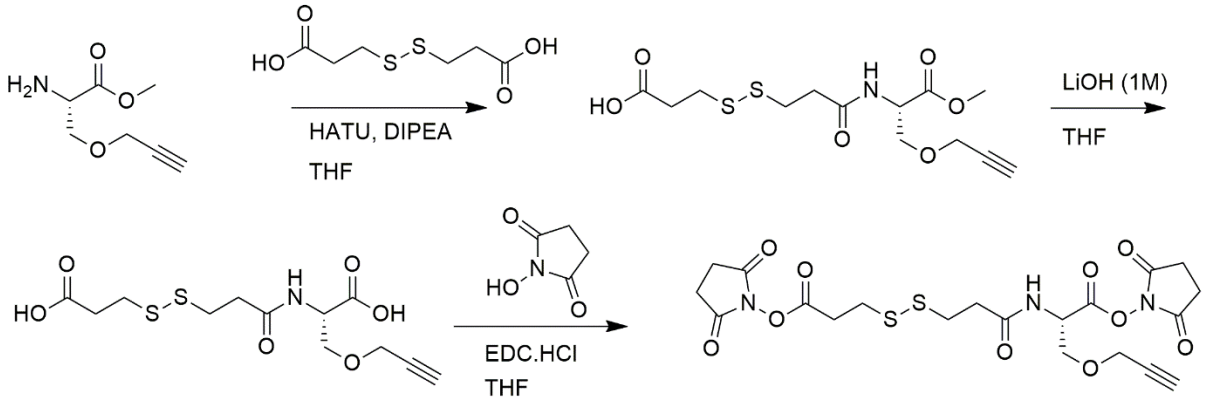
Initial synthesis of the DBAD protected thiol was attempted (**Scheme 4**) but reaction failed to yield any of the desired product. MS analysis did not show the expected ions and NMR revealed only a large singlet at 1.45 ppm corresponding to the di-tert butyl group, but showed no CH_2 triplets indicating the presence of 3-mercaptopropionic acid.

Scheme 4. Synthesis of DBAD protected 3-mercaptopropionic acid.



Following the failed attempt at forming the DBAD protected thiol, a peptide coupling with 3,3'-dithiodipropionic acid was instead attempted, guided by conditions established with linker **2** (**Scheme 5**). This synthesis successfully yielded the final product, linker **3**, which was immediately used for XL-MS, as described in Chapter 4.

Scheme 5. Synthesis of linker **3**.



5.4.2 Linker **4/4a**

A new reagent based on the commercially available positive cleavable cross-linker, DSSO was considered by retrosynthetic analysis (**Fig. 6**). The sulfoxide motif could be incorporated though the peptide coupling of 3'-thiodipropanoic acid, followed by NHS esterification and oxidation of the thioether by meta-chloroperoxybenzoic acid (mCPBA), as described by Kao, *et al*.[5] (**Scheme 6**).

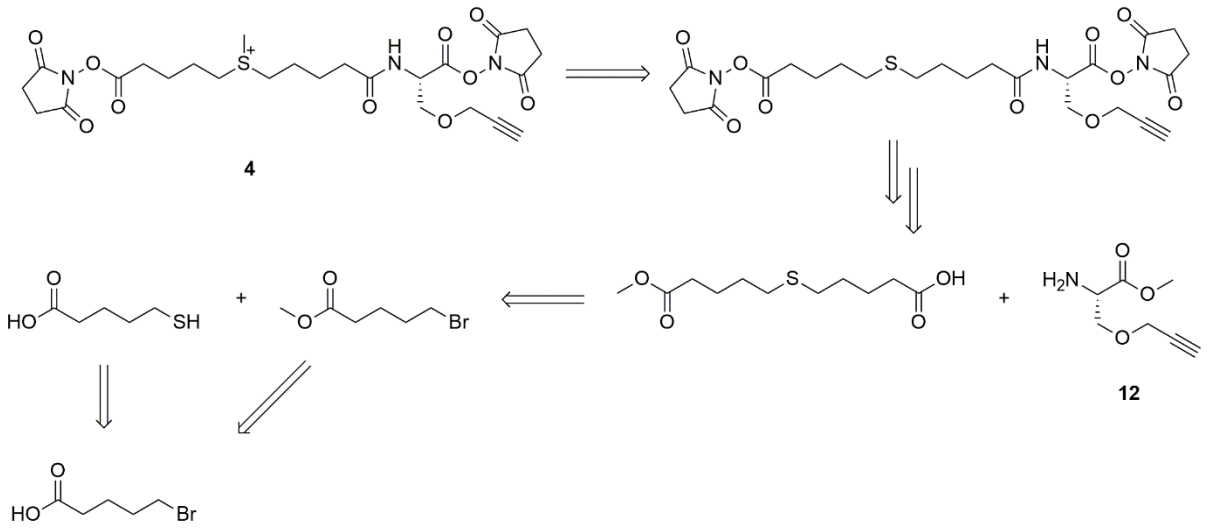
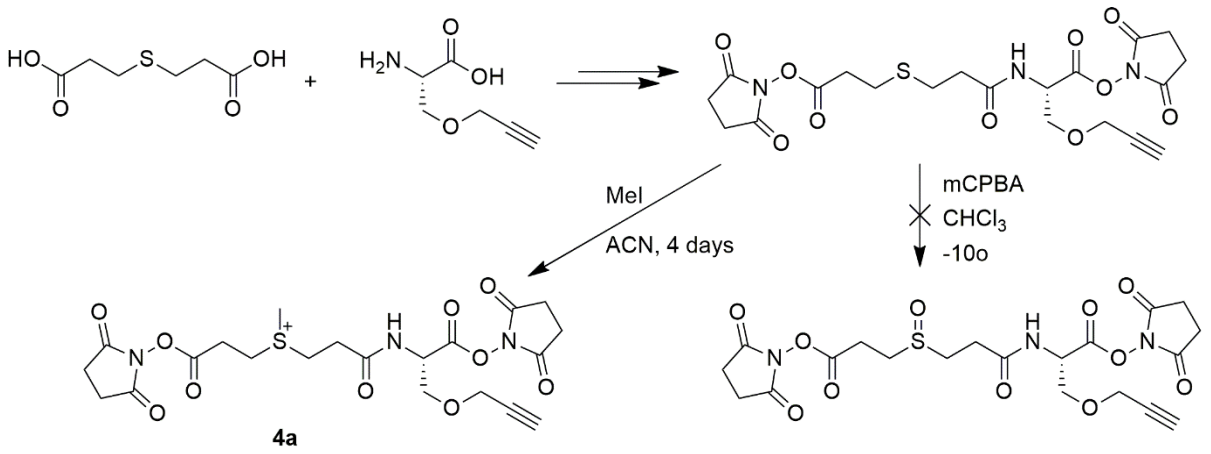


Figure 6. Retrosynthetic analysis of DSSO-based linker.

While formation of the sulphur containing cross-linker chain was a success following the same procedure as for linker **3**, oxidation with mCPBA was unsuccessful as the compound was found to preferentially form the sulfone, rather than the sulfoxide (**Scheme 6**). This was confirmed by HRMS with only the sulfone product, at 552.0887 Da, being detected. To determine whether the sulfone would still fragment under CID conditions similarly to the sulfoxide, and therefore retain functionality as an MS cleavable linker, the sulfone product was crosslinked with a test peptide, AcAAKA and subject to MS/MS analysis. The sulfone product was shown not to fragment at the spacer-arm as would be expected with the sulfoxide and instead fragment the linked test peptides preferentially. As can be seen in **Figure 7**, the MS/MS spectrum was dominated by b/y ion fragmentation along the peptide chain, while the spacer-arm remained intact. Potentially due to the added oxygen enabling resonance effects to stabilise the positive charge through CID.

As only the final oxidation step was unsuccessful, alternative reaction conditions were instead trialled to yield a methyl sulfonium product, similar in concept to *S*-methyl 5,5'-thiodipentanoylhydroxysuccinimide, but with a three carbon instead of 5 carbon chain. This would allow for testing the lability of the spacer-arm as well as extend the synthetic procedure. The methyl sulfonium linker was successfully synthesised in one step by methyl transfer using methyl iodide (2.2 mol eq.), as per Lu, *et al.*,[6] yielding linker **4a** which was used for analysis without further purification (**Scheme 6**).

Scheme 6. Synthesis of methyl sulfonium linker, **4a**, following the failed oxidation of the DSSO-based linker.



Following successful synthesis of this linker, **4a**, an MS/MS experiment was performed to investigate if the linker would fragment as intended. When subject to CID, the compound rearranged to form a 4-membered oxazoline ring, confirming the reaction as well as compatibility between the serine base unit and a methyl sulfonium containing spacer-arm. On top of this, the data showcases the first time a shorter carbon chain has been used as a methyl sulfonium based cleavable crosslinker.

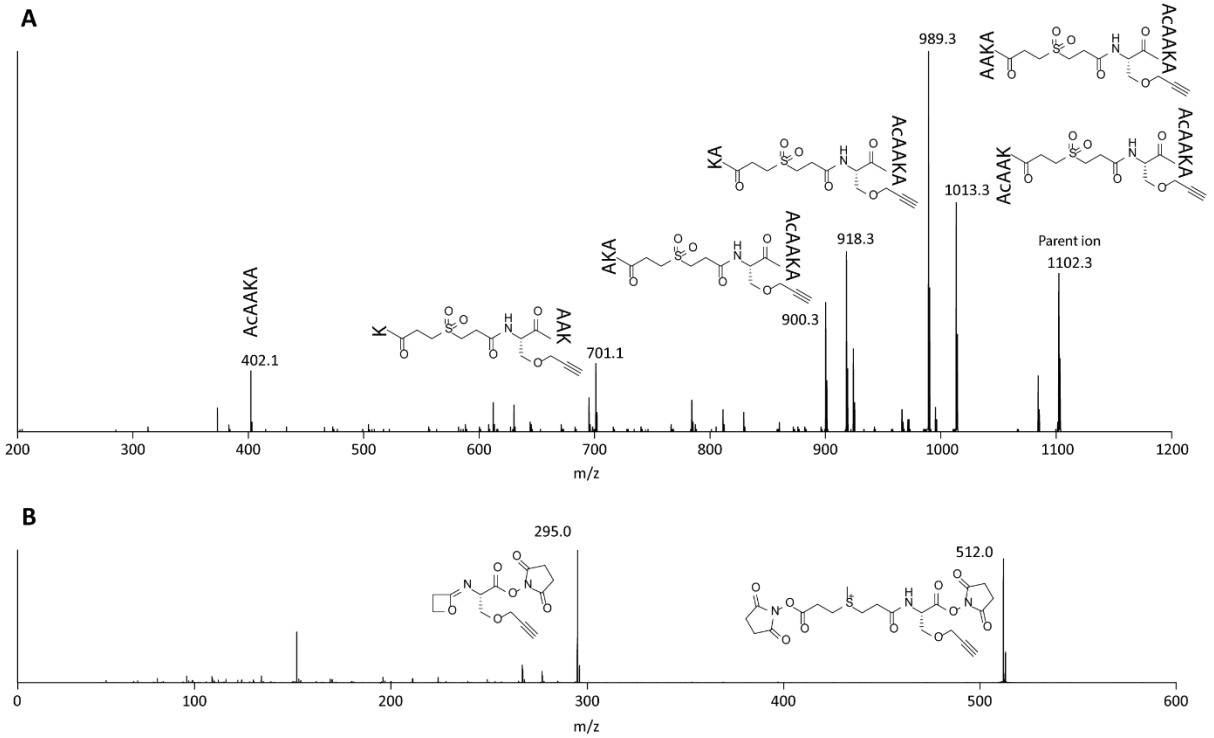


Figure 7. (**A**) Increasing CID energy instead lead to preferential fragmentation of the peptide backbone, rather than the spacer-arm. (**B**) When the sulphur was methylated, rather than oxidised, the spacer-arm fragmented preferentially upon increased CID energy, rearranging to form a 4-membered oxazoline ring.

Showing the linker design for a labile spacer-arm based on the fixed-charge methyl sulfonium ion was feasible, a longer version of the linker using 5,5'-thiodipentanoic acid, as per the design by Lu*, et al.*,[6] was synthesised. This design initially included peptide coupling with a mono-methyl ester derivative of 5,5'-thiodipentanoic acid, 5-((5-methoxy-5-oxopentyl)thio)pentanoic acid, as this would help reduce diaddition products arising if the serine unit could attack either side of the acid. The thioether synthesis was trialled (**Fig. 8**)

using Et_3N in DCM as well as K_2CO_3 and 12 crown-4-ether. Neither combination yielded the desired monoester/-acid product and therefore the diacid, 5,5'-thiodipentanoic acid, was coupled to the amine, **12**, by activation with HATU instead. This successfully yielded the desired amide which, following deprotection and methylation with methyl iodide, yielded Linker **4**, which was used for crosslinking reactions, as described in Chapter 4, without further purification (**Scheme 7**)**.**

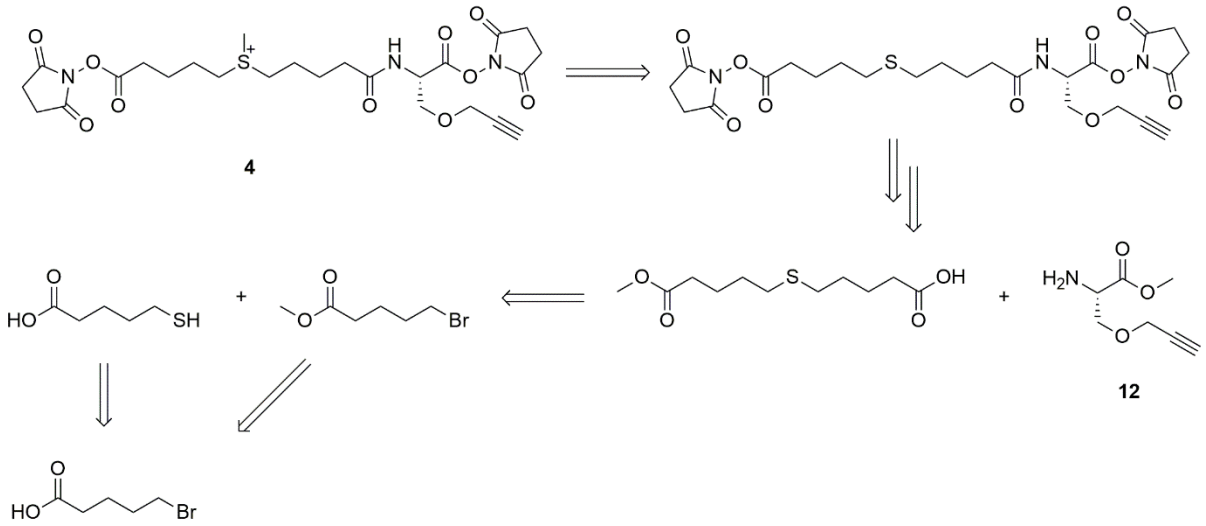
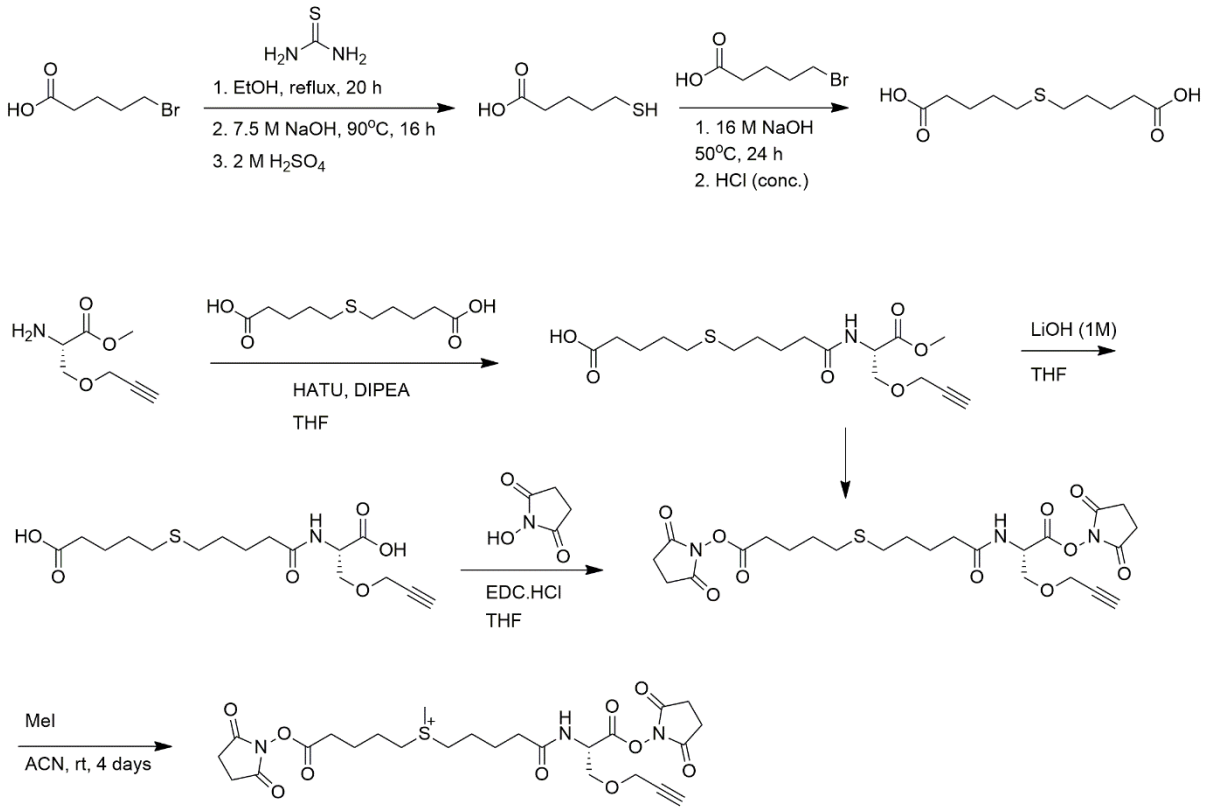


Figure 8. Retrosynthetic analysis of linker **4.**

Scheme 7. Synthesis of 5, 5'-thiodipentanoic acid which used as the spacer-arm coupling to form linker **4**

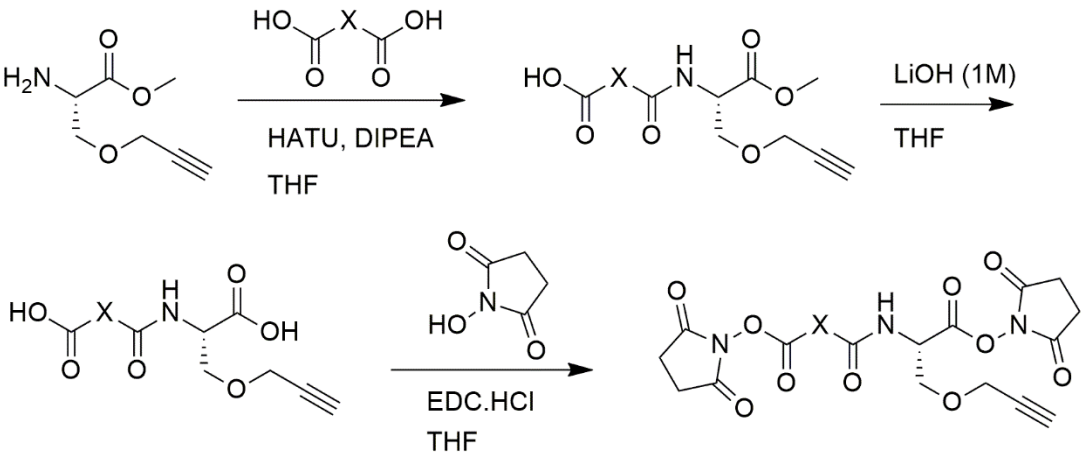


5.5 Conclusions and future directions

Combining the knowledge gathered through the successful synthesis of five cross-linking reagents, a general synthetic procedure was generated (**Scheme 8**). It was important to keep this synthetic procedure as consistent as possible, through the coupling of diacids (and monoacid/-esters), rather than unique syntheses for each spacer arm to ensure the protocol is simple to follow and therefore widely accessible. Based on the previous reactions, the general synthetic procedure developed is capable of handling the synthesis of all considered spacer-arm combinations, with an additional methylation step for linker **4** and **4a**. The flexibility of the peptide coupling allows for the incorporation of any acid based spacer-arm but could also allow for incorporation of a diazirine for non-specific heterobifunctional cross-linking. The esterification reaction allows for diversity wherein any ester-based reactive group can be incorporated. Finally, the alkyne tag allows for the ability to directly bind DBCO Sepharose beads directly for enrichment,[10] but could also be derivatised by biotin azide or azido-PEG3-phosphonic acid for avidin or IMAC affinity chromatography, respectively.

The work demonstrated in this chapter reflects on the considerations which developed the groundwork toward the synthesis of a more comprehensive range of cross-linker reagents, as shown in Chapter 4, that can be applied to biological samples. The advantages of diverse cross-linker choices may help to address some current analytical challenges with regards to XL-MS.

Scheme 8. General synthetic procedure. X represents different spacer arm options.



5.6 Materials and methods

5.6.1 Materials and reagents

Chemicals were purchased from Merck (Kenilworth, U.S.A.) or AK Scientific (Union City, U.S.A.). The peptide AcAAKA was synthesised in-house using standard Fmoc solid-phase methods on 2-chlorotrityl chloride resin (GL Biochem, Shanghai, China) and purified by high-performance liquid chromatography (HPLC) to greater than 95% purity as described previously.[4]

5.6.2 Cross-linking of AcAAKA

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

5.6.3 Tandem mass spectrometry

MS/MS spectra were collected using a Micromass Q-ToF 2 mass spectrometer (Waters, Manchester, UK) using a nano-electrospray ionisation source. Samples were prepared in 1:1 water:acetonitrile at a concentration of 10 µM. Samples were loaded into platinum-coated borosilicate glass capillaries prepared in-house. Instrument parameters as follows: capillary voltage, 1.60 kV; sampling cone, 30 V; extraction cone, 1.5 V; trap/transfer collision energy, 10/15 V; trap gas, 5.5 l/h; backing gas, ~4.5 mbar, collision energy 10 V. For CID, collision energy was increased to 50 V. Mass spectra were viewed using MassLynx (v4.1) (Waters Corporation, Manchester, U.K.).

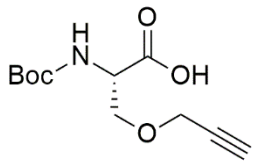
5.6.4 High-resolution mass spectrometry

High-resolution mass spectra were collected using an Agilent Technologies 6230 Accurate-Mass TOF spectrometer and were visualised using Qualitative Analysis (Version B.06.00) (MassHunter Workstation Software, Agilent Technologies, Santa Clara, U.S.A.)

5.7 Synthetic methods

5.7.1 Compound 12

(S)-2-((tert-butoxycarbonyl)amino)-3-(prop-2-yn-1-yloxy)propanoic acid (11)

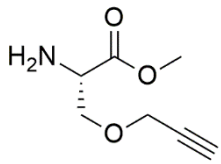


Compound **12** was synthesised by dissolving Boc-L-Serine) in DMF (50 mL per 1 g) and stirring on ice for 15 min. A 60 % wt/wt dispersion of sodium hydride (2.5 mol eq.) was added and the solution was stirred on ice for 45 min. The reaction mixture was allowed to warm to room temperature and then stir overnight under a nitrogen atmosphere. Water (15 mL) was added and stirred for 5 minutes. The solution was washed with diethyl ether, acidified to <3 pH with 1 M $KHSO_4$ and then extracted with ethyl acetate (3x 30 mL). The combined extracts were washed with water and dried over Na_2SO_4. The solvent was removed *in vacuo* to yield **11** as a yellow oil which was used without further purification with a yield of 91 %.

^{1}H NMR (500 MHz, DMSO-d_6, δ): 8.64 (br s, 3H), 4.37 (t, *J* = 3.3 Hz, 1H), 4,24 (dd, *J* = 16.1, 2.2 Hz ,1H), 4.20 (dd, *J* = 16.1, 2.2 Hz, 1H), 3.91 (dd, *J* = 10.5, 4.1 Hz, 1H), 3.85 (dd, *J* = 10.5, 3.0 Hz, 1H), 3.76 (s, 3H), 3.56 (t, *J* = 2.1 Hz, 1H)

All NMR data is consistent with literature.[11]

(S)-methyl 2-amino-3-(prop-2-yn-1-yloxy)propanoate (12)

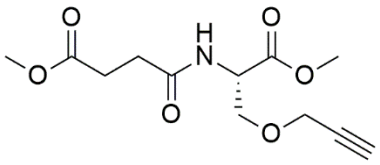


The acid (**11**) was dissolved in methanol (50 mL per 1 g) and stirred on ice for 15 min. Thionyl chloride was then added dropwise (2 mol eq.) under a nitrogen atmosphere. The solvent was removed *in vacuo* and the residue was co-evaporated with methanol three times to yield compound **12** as an orange oil, with a yield of 97 %.

^{1}H NMR (500 MHz, DMSO-d_6, δ): 8.64 (br s, 3H), 4.37 (t, J = 3.3 Hz, 1H), 4.24 (dd, J = 16.1, 2.2 Hz, 1H), 4.20 (dd, J = 16.1, 2.2 Hz, 1H), 3.91 (dd, J = 10.5, 4.1 Hz, 1H), 3.85 (dd, J = 10.6, 3.0 Hz, 1H), 3.76 (s, 3H) 3.56 (t, J = 2.1 Hz, 1H). ^{13}C NMR (125 MHz, CD_3OD, δ): 168.7, 79.4, 77.2, 67.6, 59.4, 54.3, 53.9. All NMR data is consistent with literature.[12]

5.7.2 Succinimide linker

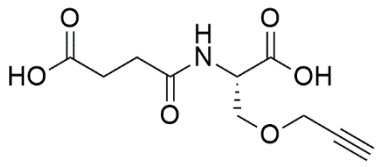
(S)-methyl 4-((1-methoxy-1-oxo-3-(prop-2-yn-1-yloxy)propan-2-yl)amino)-4-oxobutanoate



12 was dissolved in tetrahydrofuran (THF) (5 mL per 100 mg) with diisopropylethylamine (DIPEA) (4.0 mol eq) while stirring. Succinic anhydride (1.0 mol eq.) was added to the solution followed by HATU (1.2 mol eq.). The mixture was stirred overnight at room temperature under a nitrogen atmosphere. 1 M HCl (5 mL per 100 mg) was added and then extracted with ethyl acetate (1.5 mL per 100 mg, 3×). The organic phase was washed with water (3×) and brine then dried over N_2SO_4 and the solvent was removed *in vacuo*. As this would yield a monoacid/-ester the crude product was esterifying to aid in purification. The crude product was dissolved in methanol (50 mL per 1 g) and stirred on ice for 15 min. Thionyl chloride was then added dropwise (2.0 mol eq.) under a nitrogen atmosphere. The solvent was removed *in vacuo* and the residue was co-evaporated with methanol three times. The crude product was then purified by flash chromatography on normal phase silica to yield the desired amide **17** with a yield of 38%.

^{1}H NMR (500 MHz, $CDCl_3$, δ): 4.75 (dt, *J* = 8.2, 3.2 Hz, 1H), 4.12 (d, *J* = 2.4 Hz, 2H), 3.94 (dd, *J* = 9.5, 3.2 Hz, 1H), 3.74 (s, 3H), 3.78 – 3.71 (m, 1H), 3.67 (s, 3H), 2.68 – 2.53 (m, 4H), 2.45 (t, *J* = 2.4 Hz, 1H).

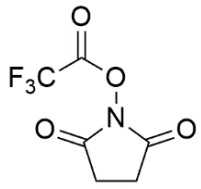
(S)-4-((1-carboxy-2-(prop-2-yn-1-yloxy)ethyl)amino)-4-oxobutanoic acid



The amide **16** was dissolved in THF (1 mL per 100 g) and stirred on ice for 15 minutes. 1 M LiOH (4 mol eq.) was added and the reaction was stirred until completion determined by thin layer chromatography (TLC). The solvent was removed *in vacuo* and the residue was partitioned between DCM and water. The organic layer was isolated and the aqueous layer was extracted with DCM (2×). The combined organic extracts were washed with brine and dried over N_2SO_4. The solvent was removed *in vacuo* to yield the desired diacid (**17**) as a yellow oil with a yield of 80%.

^{1}H NMR (500 MHz, CDO_3D, δ): δ 4.63 (t, *J* = 4.2 Hz, 1H), 4.19 (d, *J* = 2.4 Hz, 2H), 3.93 (dd, *J* = 9.6, 4.2 Hz, 1H), 3.78 (dd, *J* = 9.7, 3.6 Hz, 1H), 2.87 (t, *J* = 2.5 Hz, 1H), 2.68 – 2.55 (m, 4H).

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



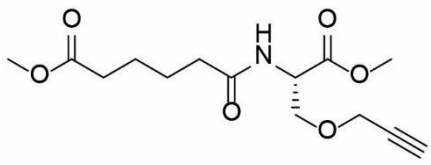
TFA-NHS was synthesised by dissolving NHS in toluene (20 mL per gram). TFA was added (1.2 eq.) and the mixture was refluxed for 5 h. The solvent was removed *in vacuo* to yield TFA-NHS as a white crystalline solid in near quantitative yield.

^{1}H NMR (500 MHz, CDCl3, δ) 2.92 (s, 4H).

^{19}F NMR (125 MHz, $CDCl_3$, δ) 167.4, 153.8 (q), 114.1, 25.7.

5.7.3 Linker 1

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



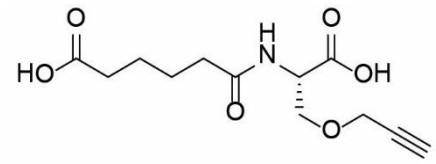
12 was dissolved in tetrahydrofuran (THF) (5 mL per 100 mg) with diisopropylethylamine (DIPEA) (4.0 mol eq) while stirring. Monomethyl adipate (1.0 mol eq.) was added to the solution followed by HATU (1.2 mol eq.). The mixture was stirred overnight at room temperature under a nitrogen atmosphere. 1 M HCl (5 mL per 100 mg) was added and then extracted with ethyl acetate (1.5 mL per 100 mg, 3×). The organic phase was washed with water (3×) and brine then dried over N_2SO_4 and the solvent was removed *in vacuo*.

The crude product was then purified by flash chromatography on normal phase silica to yield the desired amide with a yield of 19%.

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

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

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



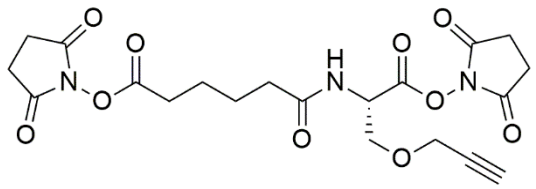
The amide was dissolved in THF (1 mL per 100 g) and stirred on ice for 15 minutes. 1 M LiOH (4 mol eq.) was added and the reaction was stirred until completion determined by thin layer chromatography (TLC). The solvent was removed *in vacuo* and the residue was partitioned between DCM and water. The organic layer was isolated and the aqueous layer was extracted with DCM (2×). The combined organic extracts were washed with brine and dried over N_2SO_4.

The solvent was removed *in vacuo* to yield the desired diacid as a yellow oil with a yield of 79%.

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

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

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

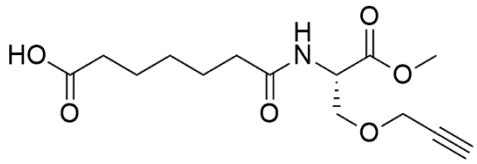


The diacid was dissolved in anhydrous THF (10 mL per 100 mg) with EDC-HCl (3.4 mol eq) and stirred at room temperature for 10 minutes. NHS was (6 eq.) added and the reaction was stirred at room temperature under a nitrogen atmosphere overnight. The solvent was removed *in vacuo* and the resulting residue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

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

5.7.4 Linker 2

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



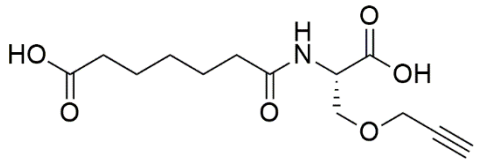
12 was dissolved in tetrahydrofuran (THF) (5 mL per 100 mg) with diisopropylethylamine (DIPEA) (4.0 mol eq) while stirring. Heptanedioic acid (1.0 mol eq.) was added to the solution followed by HATU (1.2 mol eq.). The mixture was stirred overnight at room temperature under a nitrogen atmosphere. 1 M HCl (5 mL per 100 mg) was added and then extracted with ethyl acetate (1.5 mL per 100 mg, 3×). The organic phase was washed with water (3×) and brine then dried over N_2SO_4 and the solvent was removed *in vacuo*.

The crude product was then purified by flash chromatography on normal phase silica to yield the desired amide with a yield of 16%.

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

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

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



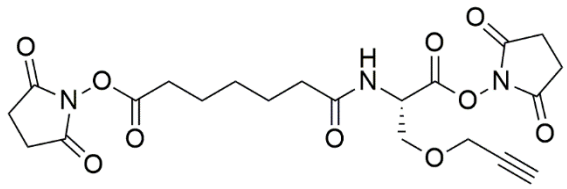
The amide was dissolved in THF (1 mL per 100 g) and stirred on ice for 15 minutes. 1 M LiOH (4 mol eq.) was added and the reaction was stirred until completion determined by thin layer chromatography (TLC). The solvent was removed *in vacuo* and the residue was partitioned between DCM and water. The organic layer was isolated and the aqueous layer was extracted with DCM (2×). The combined organic extracts were washed with brine and dried over N_2SO_4.

The solvent was removed *in vacuo* to yield the desired diacid as a yellow oil with a yield of 79%.

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

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

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

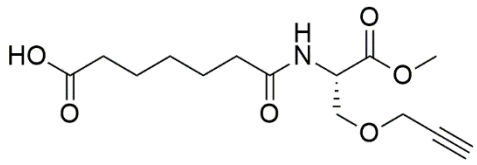


The diacid was dissolved in anhydrous THF (10 mL per 100 mg) with EDC-HCl (3.4 mol eq) and stirred at room temperature for 10 minutes. NHS was (6 eq.) added and the reaction was stirred at room temperature under a nitrogen atmosphere overnight. The solvent was removed *in vacuo* and the resulting residue was partitioned between DCM and water. The aqueous layer was extracted with DCM (3x) and the combined organic extracts were dried over $MgSO_4$ and filtered. The solvent was removed from the filtrate *in vacuo* to afford the NHS ester, Linker **2**, which was used for cross-linking reactions without further purification.

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

5.7.5 Linker **3**

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



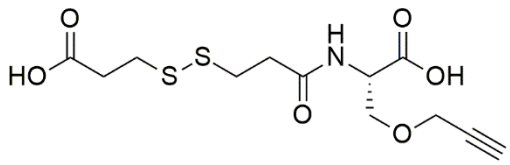
12 was dissolved in tetrahydrofuran (THF) (5 mL per 100 mg) with diisopropylethylamine (DIPEA) (4.0 mol eq) while stirring. 3,3'-dithiodipropanoic acid (1.0 mol eq.) was added to the solution followed by HATU (1.2 mol eq.). The mixture was stirred overnight at room temperature under a nitrogen atmosphere. 1 M HCl (5 mL per 100 mg) was added and then extracted with ethyl acetate (1.5 mL per 100 mg, 3×). The organic phase was washed with water (3×) and brine then dried over N_2SO_4 and the solvent was removed *in vacuo*.

The crude product was then purified by flash chromatography on normal phase silica to yield the desired amide with a yield of 16%.

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

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

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



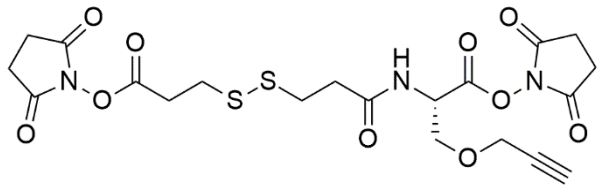
The amide was dissolved in THF (1 mL per 100 g) and stirred on ice for 15 minutes. 1 M LiOH (4 mol eq.) was added and the reaction was stirred until completion determined by thin layer chromatography (TLC). The solvent was removed *in vacuo* and the residue was partitioned between DCM and water. The organic layer was isolated and the aqueous layer was extracted

with DCM (2×). The combined organic extracts were washed with brine and dried over N_2SO_4. The solvent was removed *in vacuo* to yield the desired diacid as a yellow oil with a yield of 83%.

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

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

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

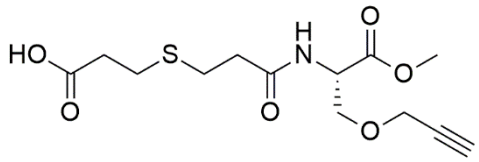


The diacid was dissolved in anhydrous THF (10 mL per 100 mg) with EDC-HCl (3.4 mol eq) and stirred at room temperature for 10 minutes. NHS was (6 eq.) added and the reaction was stirred at room temperature under a nitrogen atmosphere overnight. The solvent was removed *in vacuo* and the resulting residue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

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

5.7.6 Linker 4a

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



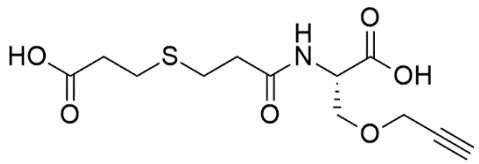
12 was dissolved in tetrahydrofuran (THF) (5 mL per 100 mg) with diisopropylethylamine (DIPEA) (4.0 mol eq) while stirring. 3,3'-thiodipropanoic acid (1.0 mol eq.) was added to the solution followed by HATU (1.2 mol eq.). The mixture was stirred overnight at room temperature under a nitrogen atmosphere. 1 M HCl (5 mL per 100 mg) was added and then extracted with ethyl acetate (1.5 mL per 100 mg, 3×). The organic phase was washed with water (3×) and brine then dried over N_2SO_4 and the solvent was removed *in vacuo*.

The crude product was then purified by flash chromatography on normal phase silica to yield the desired amide with a yield of 37%.

^{1}H NMR (500 MHz, CD_3OD, δ): 4.69 (dt, *J* = 3.81, 1.1 Hz, 1H), 4.18 (t, *J* = 2.3 Hz, 1H), 3.93 (dd, *J* = 9.6, 4.7 Hz, 1H), 3.77 (dd, *J* = 9.7, 3.7 Hz, 1H), 3.75 (s, 3H), 2.89 (t, *J* = 2.4 Hz, 1H), 2.80 (td, *J* = 7.3, 2.9 Hz, 4H), 2.64 (t, *J* = 7.1 Hz, 2H), 2.58 (t, *J* = 7.2 Hz, 2H).

^{13}C NMR (125 MHz, $CDCl_3$, δ): 179.10, 173.91, 173.25, 81.41, 77.91, 72.05, 61.28, 55.46, 55.14, 39.14, 37.21, 30.19, 29.47

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



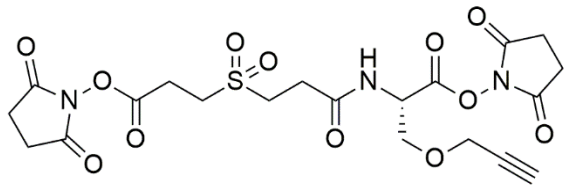
The amide was dissolved in THF (1 mL per 100 g) and stirred on ice for 15 minutes. 1 M LiOH (4 mol eq.) was added and the reaction was stirred until completion determined by thin layer chromatography (TLC). The solvent was removed *in vacuo* and the residue was partitioned between DCM and water. The organic layer was isolated and the aqueous layer was extracted

with DCM (2×). The combined organic extracts were washed with brine and dried over N_2SO_4. The solvent was removed *in vacuo* to yield the desired diacid as a yellow oil with near quantitative yield.

^{1}H NMR (500 MHz, CD_3OD, δ): 4.66 (dt, *J* = 3.6, 1.1 Hz, 1H), 4.19 (d, *J* = 2.4 Hz, 2H), 3.94 (dd, *J* = 9.6, 4.7 Hz, 1H), 3.79 (dd, *J* = 9.6, 3.6 Hz, 1H), 2.81 (dt, *J* = 10.1, 7.2 Hz, 4H), 2.60 (td, *J* = 7.2, 5.3 Hz, 4H).

^{13}C NMR (125 MHz, CD_3OD, δ): 176.9, 175.5, 174.2, 81.4, 77.6, 71.6, 60.5, 55.2, 38.4, 36.9, 29.7, 29.0.

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

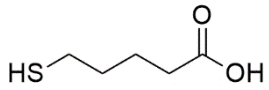


The diacid was dissolved in anhydrous THF (10 mL per 100 mg) with EDC-HCl (3.4 mol eq) and stirred at room temperature for 10 minutes. NHS was (6 eq.) added and the reaction was stirred at room temperature under a nitrogen atmosphere overnight. The solvent was removed *in vacuo* and the resulting residue was partitioned between DCM and water. The aqueous layer was extracted with DCM (3x) and the combined organic extracts were dried over $MgSO_4$ and filtered. The solvent was removed from the filtrate *in vacuo* to afford the NHS ester. The NHS ester was dissolved in chloroform (5 mL per 100 mg) at 0 °C and mixed with a solution of *m*-chloroperbenzoic acid (1 mol eq.) (3 mL per 100 mg). The solution was stirred for 5 minutes then solvent was removed *in vacuo* and the product was purified by flash chromatography on normal phase silica to yield the desired NHS that was used for cross-linking reactions immediately.

HRMS (ESI+) calculated for $C_{20}H_{23}N_4O_{12}S$ $[M+Na]^+$ = 552.0900, found $[M+Na]^+$: 552.0887

5.7.7 Linker **4**

5-mercaptopentanoic acid

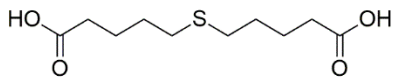


Based on synthetic procedures described by Lu, *et al.* (2008),[6] 5-bromovaleric acid and thiourea (1.5 mol eq.) were dissolved in ethanol (10 mL per 1 g) and refluxed for 20 h. The solvent was removed under reduced pressure and 7.5 M NaOH (aq) (15 mol eq.) was added. The mixture was stirred for 16 h at 90 °C. The mixture was cooled on ice and 2 M H_2SO_4 was added dropwise while stirring until pH 1. The product was then extracted with dichloromethane (2 × 100 mL). The combined organic extracts were dried over $MgSO_4$ and filtered. The solvent was removed from the filtrate *in vacuo* to afford 5-mercaptopentanoic acid as a colourless oil in near quantitative yield. The product was used without further purification.

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

As per NMR data is consistent with literature.[6]

5,5'-thiodipentanoic acid



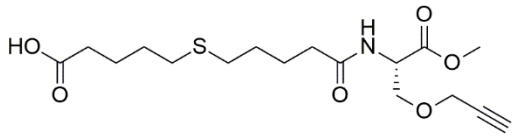
Using synthesis described in Lu, *et al.* (2008),[6] a solution of 5-bromovaleric acid in 16.7 M NaOH (2 mL per 1 g) was added dropwise to a mixture of 5-mercaptopentanoic acid in 16.7 M (2 mL per 1g). The mixture was stirred at 45 °C for 24 h. The mixture was cooled to room temperature and acidified to pH 1 with concentrated hydrochloric acid. The product was extracted with dichloromethane (5 × 50 mL) and the combined organic extracts were dried over $MgSO_4$ and filtered. The solvent was removed from the filtrate *in vacuo* to afford 5, 5'-thiodipentanoic acid as a white solid with a yield of 86%.

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

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

As NMR data as per literature[6]

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



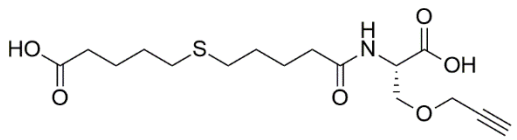
12 was dissolved in tetrahydrofuran (THF) (5 mL per 100 mg) with diisopropylethylamine (DIPEA) (4.0 mol eq) while stirring. 5, 5'-thiodipentanoic acid (1.0 mol eq.) was added to the solution followed by HATU (1.2 mol eq.). The mixture was stirred overnight at room temperature under a nitrogen atmosphere. 1 M HCl (5 mL per 100 mg) was added and then extracted with ethyl acetate (1.5 mL per 100 mg, 3×). The organic phase was washed with water (3×) and brine then dried over N_2SO_4 and the solvent was removed *in vacuo*.

The crude product was then purified by flash chromatography on normal phase silica to yield the desired amide with a yield of 20%.

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

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

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



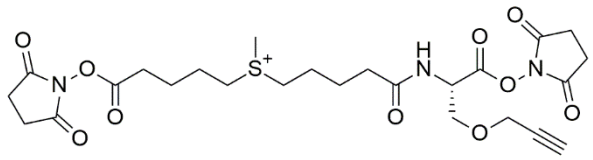
The amide was dissolved in THF (1 mL per 100 g) and stirred on ice for 15 minutes. 1 M LiOH (4 mol eq.) was added and the reaction was stirred until completion determined by thin layer

chromatography (TLC). The solvent was removed *in vacuo* and the residue was partitioned between DCM and water. The organic layer was isolated and the aqueous layer was extracted with DCM (2×). The combined organic extracts were washed with brine and dried over N_2SO_4. The solvent was removed *in vacuo* to yield the desired diacid as a yellow oil with 67% yield.

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

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

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



The diacid was dissolved in anhydrous THF (10 mL per 100 mg) with EDC-HCl (3.4 mol eq) and stirred at room temperature for 10 minutes. NHS was (6 eq.) added and the reaction was stirred at room temperature under a nitrogen atmosphere overnight. The solvent was removed *in vacuo* and the resulting residue was partitioned between DCM and water. The aqueous layer was extracted with DCM (3x) and the combined organic extracts were dried over $MgSO_4$ and filtered. The solvent was removed from the filtrate *in vacuo* to afford the NHS ester. The NHS ester was dissolved in acetonitrile (ACN) (1 mL per 100 mg) followed by methyl iodide (2.2 mol eq.) and left stirring under a N_2 atmosphere for 4 days successfully yielding linker **4a**, which was used for crosslinking reactions without further purification.

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

5.8 References

(1) Downey, K. M.; Prof, A.; Pukala, T.; Abell, P. A. Design and Synthesis of Protein Chemical Crosslinkers : A Modular Approach, University of Adelaide, **2017**.

(2) Rao, T. S.; Nampalli, S.; Sekher, P.; Kumar, S. TFA-NHS as Bifunctional Protecting Agent: Simultaneous Protection and Activation of Amino Carboxylic Acids. *Tetrahedron Lett.* **2002**, *43* (43), 7793–7795. https://doi.org/10.1016/S0040-4039(02)01759-8.

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

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

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

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

(7) Florence, T. M. Degradation of Protein Disulphide Bonds in Dilute Alkali. *Biochem. J.* **1980**, *189* (3), 507–520. https://doi.org/10.1042/bj1890507.

(8) Galande, A. K.; Trent, J. O.; Spatola, A. F. Understanding Base-Assisted Desulfurization Using a Variety of Disulfide-Bridged Peptides. *Biopolym. - Pept. Sci. Sect.* **2003**. https://doi.org/10.1002/bip.10532.

(9) He, R.; Pan, J.; Mayer, J. P.; Liu, F. Stepwise Construction of Disulfides in Peptides. *ChemBioChem* **2020**, *21* (8), 1101–1111. https://doi.org/10.1002/cbic.201900717.

(10) Matzinger, M.; Kandioller, W.; Doppler, P.; Heiss, E. H.; Mechtler, K. Fast and Highly Efficient Affinity Enrichment of Azide-A-DSBSO Cross-Linked Peptides. *J. Proteome Res.* **2020**, *19* (5), 2071–2079. https://doi.org/10.1021/acs.jproteome.0c00003.

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

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

Chapter 6: Summary, Conclusion and Future Directions

The focus of this thesis was to utilise an MS-based approached to better understand structures and interactions of proteins that are notoriously complicated to analyse with traditional structure determination techniques. Overall, the work here highlights the power of utilising an MS-based approach to tackle unique protein structure challenges, in particular for misfolding proteins whose intrinsic disorder demands an analytical technique capable of great adaptability.

6.1 β-CN stabilises both amorphous and fibrillar α-LA

The ability to monitor both amorphous and fibrillar aggregation from a single protein allows greater insight into the structural behaviour that designates which pathway it will take. This is of much interest as amorphous aggregates are widely regarded and non-toxic, while fibrillar aggregates are associated with a wide range of neurodegenerative diseases. Understanding the contrasting changes in structural effects that dictate a mode of aggregation may have implications in disease understanding or identification of therapeutic targets for drug design. The molecular chaperone β-CN was shown to inhibit the aggregation of both fibrillar RCM α-LA and amorphous DTT α-LA, with a greater efficacy observed toward RCM α-LA. Intrinsic and extrinsic fluorescence studies reveal differences in the two forms of α-LA, where DTT α-LA exposed far greater levels of hydrophobicity but exhibited native-like levels of tryptophan residues suggesting a molten-globule like state that undergoes rapid amorphous aggregation due to random interactions from exposed hydrophobicity. On the other hand, RCM α-LA indicated a lack of native structure, through a change in tryptophan fluorescence, but exposed little hydrophobic content. The unravelling in structure but lack of exposed hydrophobicity may be important to the fibrillation mechanism as RCM α-LA fibril formation is a slower process with a relatively lengthy nucleation phase, compared with amorphous aggregation. This may be important to fibrillar aggregation of RCM α-LA by striking a balance between avoiding the faster process of amorphous aggregation, due to fewer contacts between hydrophobic residues, while also remaining destabilised enough to sample off-pathway conformations during the lag phase before nucleation.

Although β-CN inhibited both forms of α-LA aggregation, it was more effective towards fibrillar RCM α-LA. Unlike DTT α-LA,[1–3] RCM α-LA was not shown to form a HMW complex by analysis SEC. IM-MS instead revealed that RCM α-LA was transiently stabilised by β-CN. CIU analysis revealed that RCM α-LA required lower CID voltages to induce unfolding than native α-LA but when RCM α-LA was subject to CIU in the presence of β-CN the unfolding voltage increased, suggesting β-CN afforded greater structural stability to RCM α-LA, despite no non-covalent interactions being observed by native MS. Overall, this data infers β-CN operates through two distinct mechanisms, depending on the mode of aggregation. β-CN locks DTT α-LA, in the form of a disordered molten globule, into a HMW complex while the partially unfolded intermediate of RCM α-LA is afforded structural stability through transient interactions with β-CN. Furthermore, this work highlights the power of CIU as a technique for investigating the effects on protein stability from molecular chaperones.

The application of CIU toward chaperone-protein interactions is underrepresented and could give a better understand to a whole host of protein-protein interactions. This technique is particularly suited for the examination of chaperones and inhibition relationships of misfolded proteins that still retain structure and therefore unfolding can be visualised, yet are still too dynamic for other techniques.

6.2 Lipids reduce fibril inhibitor efficacy of Aβ40 and αS

Current understanding implicates oligomeric, misfolding proteins interacting with lipid membranes in aetiology of misfolding diseases. Many studies have been conducted on Aβ40 and αS, the proteins associated with Alzheimer's and Parkinson's disease respectively, to better understand potentially deleterious effects from lipid membrane interactions.[4–7] Natural polyphenols are of interest in drug development due to their notable anti-amyloid effects[8,9] with EGCG[9–12] and resveratrol[13,14] being two prototypical examples. In this work, LUVs, one zwitterionic (PC) and one anionic (PCPG), were used as models for a lipid environment. Using ThT fluorescence assays, both LUV types were shown to increase fibrillar aggregation, with anionic lipids having a greater effect. LUVs were also shown to decrease the efficacy of both EGCG and resveratrol, with anionic LUVs again having a greater impact.

Oligomer specific immunoblotting revealed that EGCG reduced oligomeric content of Aβ40 and αS more effectively that resveratrol. This may be explained by EGCG having been

previously shown to affect oligomeric misfolding species,[9–12] while resveratrol has been described as inhibiting through a capping mechanism, preventing the elongation rather than nucleation of misfolding proteins.[15–17]

Finally, native MS and IM-MS was used to characterise non-covalent interactions and structural changes. Native MS showed that EGCG formed complexes with Aβ40 and αS, while resveratrol did not. IM-MS showed EGCG induced structural compaction across multiple charge states of Aβ40 and αS, a phenomenon witnessed before the initiation of fibrillar aggregation.[18–20] In contrast, EGCG interacted with compact forms of monomeric Aβ40 and αS, decreasing structural homogeneity. The direct competition between the structural effects of EGCG and LUVs may explain why inhibitory efficacy was reduced. Furthermore, as resveratrol inhibits more directly within the elongation phase, it cannot counter the effects of a lipid environment as effectively as EGCG. This could potentially explain its poorer performance as an inhibitor.

Overall, this work added insight into the poorly understood relationship between lipid membranes and misfolding proteins as well as highlighting the importance of considering lipids in aggregation inhibition studies. This study emphasizes that if the lipid environment is not taken into account during therapeutic design, the outcome may result in an ineffective treatment.

Further work is required to better understand the relationship between lipids, misfolding proteins and inhibitors. Utilising more protein models and inhibitors not based on natural polyphenols could yield more insight into these mechanisms. On top of this, different or more realistic lipid models, such as lipid nanodiscs and total brain lipid extract derived LUVs could reveal more about what different lipid characteristics can do to fibrillar inhibition.

6.3 Development of a modular protocol and linker library to expand XL-MS opportunities

XL-MS is fast emerging as one of the most important MS-based techniques, capable of handling proteins that are highly dynamic, like misfolding proteins.[21–23] In particular, the power of this technique lies in its ability to be combined with higher-resolution techniques, like cryo-EM.[24] To tackle gaps in the reagent toolset of XL-MS a modular synthetic protocol was designed to greatly expand the range of available cross-linker options that are currently

available. Centred around an amino acid base unit, derived from alkyne-modified serine or azide-modified lysine residues, spacer-arms can be peptide coupled followed by esterification with a reactive group. Furthermore, the azide/alkyne motifs allow for the possibility of post-linkage functionalisation opportunities. Using acid-based spacer-arms and hydroxyl-based reactive groups, a library of 8 linkers was synthesised. These linkers were tested on model peptides and proteins to showcase their linkage, labile spacer-arm fragmentation and post-linkage derivatisation with enrichment and dye tags. In summary, the described synthetic protocol aimed to increase the diversity of cross-linking reagents to better supplement the incredibly varied requirements of XL-MS experiments.

Development of the modular protocol for synthesis of XL-MS reagents leaves by design almost limitless opportunities for production of further derivatives, and hence the extended scope was too large to be explored entirely and much remains to be done. Potential routes to be explored include the expansion of reactive groups, particular those developed by Gutierrez, *et al.*[25,26] which require only a single reaction with the NHS esters already synthesised in Chapters 4 and 5. Following this, Linker **4a** showed, for the first time, the ability to form a labile 4-membered oxazoline fragmentation product. This would allow for a shorter chain length, fragmentable linker that is intrinsically charged, which may increase ionisability. Further optimisation of this linker could uncover its benefits as a useful reagent. Finally, due to linker base unit consisting of amino acids using isotopically labelled amino acids would allow for quantitation and characteristic identification.[27,28]

To conclude, many proteins are not amenable to traditional structure determination techniques but the need to understand their structure to aid in disease research and therapeutic development remains. MS is a powerful compliment, or alternative, to those techniques, and its full potential as a protein structure determination technique is still yet to be realised.

6.4 References

(1) Holt, C.; Carver, J. A.; Ecroyd, H.; Thorn, D. C. Invited Review: Caseins and the Casein Micelle: Their Biological Functions, Structures, and Behavior in Foods1. *J. Dairy Sci.* **2013**, *96* (10), 6127–6146. https://doi.org/10.3168/jds.2013-6831.

(2) Thorn, D. C.; Ecroyd, H.; Carver, J. A.; Holt, C. Casein Structures in the Context of Unfolded Proteins. *Int. Dairy J.* **2015**, *46*, 2–11. https://doi.org/10.1016/j.idairyj.2014.07.008.

(3) Holt, C.; Raynes, J. K.; Carver, J. A. Sequence Characteristics Responsible for Protein-protein Interactions in the Intrinsically Disordered Regions of Caseins, Amelogenins, and Small Heat-shock Proteins. *Biopolymers* **2019**, *110* (9), e23319. https://doi.org/10.1002/bip.23319.

(4) Kawahara, M.; Ohtsuka, I.; Yokoyama, S.; Kato-Negishi, M.; Sadakane, Y. Membrane Incorporation, Channel Formation, and Disruption of Calcium Homeostasis by Alzheimer's β-Amyloid Protein. *Int. J. Alzheimers. Dis.* **2011**, *2011*, 1–17. https://doi.org/10.4061/2011/304583.

(5) Bode, D. C.; Freeley, M.; Nield, J.; Palma, M.; Viles, J. H. Amyloid-β Oligomers Have a Profound Detergent-like Effect on Lipid Membrane Bilayers, Imaged by Atomic Force and Electron Microscopy. *J. Biol. Chem.* **2019**, *294* (19), 7566–7572. https://doi.org/10.1074/jbc.AC118.007195.

(6) Flagmeier, P.; De, S.; Michaels, T. C. T.; Yang, X.; Dear, A. J.; Emanuelsson, C.; Vendruscolo, M.; Linse, S.; Klenerman, D.; Knowles, T. P. J.; et al. Direct Measurement of Lipid Membrane Disruption Connects Kinetics and Toxicity of Aβ42 Aggregation. *Nat. Struct. Mol. Biol.* **2020**, *27* (10), 886–891. https://doi.org/10.1038/s41594-020-0471-z.

(7) Ugalde, C. L.; Lawson, V. A.; Finkelstein, D. I.; Hill, A. F. The Role of Lipids in -Synuclein Misfolding and Neurotoxicity. *J. Biol. Chem.* **2019**, *294* (23), 9016–9028. https://doi.org/10.1074/jbc.REV119.007500.

(8) Henríquez, G.; Gomez, A.; Guerrero, E.; Narayan, M. Potential Role of Natural Polyphenols against Protein Aggregation Toxicity: In Vitro, in Vivo, and Clinical Studies. *ACS Chem. Neurosci.* **2020**, *11* (19), 2915–2934.

https://doi.org/10.1021/acschemneuro.0c00381.

(9) Freyssin, A.; Page, G.; Fauconneau, B.; Bilan, A. Natural Polyphenols Effects on Protein Aggregates in Alzheimer's and Parkinson's Prion-like Diseases. *Neural Regen. Res.* **2018**, *13* (6), 955. https://doi.org/10.4103/1673-5374.233432.

(10) Huang, Q.; Zhao, Q.; Peng, J.; Yu, Y.; Wang, C.; Zou, Y.; Su, Y.; Zhu, L.; Wang, C.; Yang, Y. Peptide-Polyphenol (KLVFF/EGCG) Binary Modulators for Inhibiting Aggregation and Neurotoxicity of Amyloid-β Peptide. *ACS Omega* **2019**, *4* (2), 4233–4242. https://doi.org/10.1021/acsomega.8b02797.

(11) Bieschke, J.; Russ, J.; Friedrich, R. P.; Ehrnhoefer, D. E.; Wobst, H.; Neugebauer, K.; Wanker, E. E. EGCG Remodels Mature α-Synuclein and Amyloid-β Fibrils and Reduces Cellular Toxicity. *Proc. Natl. Acad. Sci.* **2010**, *107* (17), 7710–7715. https://doi.org/10.1073/pnas.0910723107.

(12) Ehrnhoefer, D. E.; Bieschke, J.; Boeddrich, A.; Herbst, M.; Masino, L.; Lurz, R.; Engemann, S.; Pastore, A.; Wanker, E. E. EGCG Redirects Amyloidogenic Polypeptides into Unstructured, off-Pathway Oligomers. *Nat. Struct. Mol. Biol.* **2008**, *15* (6), 558–566. https://doi.org/10.1038/nsmb.1437.

(13) Corpas, R.; Griñán-Ferré, C.; Rodríguez-Farré, E.; Pallàs, M.; Sanfeliu, C. Resveratrol Induces Brain Resilience against Alzheimer Neurodegeneration through Proteostasis Enhancement. *Mol. Neurobiol.* **2019**, *56* (2), 1502–1516. https://doi.org/10.1007/s12035-018-1157-y.

(14) Regitz, C.; Fitzenberger, E.; Mahn, F. L.; Dußling, L. M.; Wenzel, U. Resveratrol Reduces Amyloid-Beta (Aβ1–42)-Induced Paralysis through Targeting Proteostasis in an Alzheimer Model of Caenorhabditis Elegans. *Eur. J. Nutr.* **2016**, *55* (2), 741–747. https://doi.org/10.1007/s00394-015-0894-1.

(15) Feng, Y.; Wang, X. ping; Yang, S. gao; Wang, Y. jiong; Zhang, X.; Du, X. ting; Sun, X. xia; Zhao, M.; Huang, L.; Liu, R. tian. Resveratrol Inhibits Beta-Amyloid Oligomeric Cytotoxicity but Does Not Prevent Oligomer Formation. *Neurotoxicology* **2009**, *30* (6), 986–995. https://doi.org/10.1016/j.neuro.2009.08.013.

(16) Fu, Z.; Aucoin, D.; Ahmed, M.; Ziliox, M.; Van Nostrand, W. E.; Smith, S. O. Capping of

Aβ42 Oligomers by Small Molecule Inhibitors. *Biochemistry* **2014**, *53* (50), 7893–7903. https://doi.org/10.1021/bi500910b.

(17) Mishra, R.; Sellin, D.; Radovan, D.; Gohlke, A.; Winter, R. Inhibiting Islet Amyloid Polypeptide Fibril Formation by the Red Wine Compound Resveratrol. *ChemBioChem* **2009**, *10* (3), 445–449. https://doi.org/10.1002/cbic.200800762.

(18) Sanders, H. M.; Jovcevski, B.; Carver, J. A.; Pukala, T. L. The Molecular Chaperone β-Casein Prevents Amorphous and Fibrillar Aggregation of α-Lactalbumin by Stabilisation of Dynamic Disorder. *Biochem. J.* **2020**, *477* (3), 629–643. https://doi.org/10.1042/BCJ20190638.

(19) Das, S.; Pukala, T. L.; Smid, S. D. Exploring the Structural Diversity in Inhibitors of α-Synuclein Amyloidogenic Folding, Aggregation, and Neurotoxicity. *Front. Chem.* **2018**, *6* (MAY), 1–12. https://doi.org/10.3389/fchem.2018.00181.

(20) Pavlova, A.; Cheng, C. Y.; Kinnebrew, M.; Lew, J.; Dahlquist, F. W.; Han, S. Protein Structural and Surface Water Rearrangement Constitute Major Events in the Earliest Aggregation Stages of Tau. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (2), E127–E136. https://doi.org/10.1073/pnas.1504415113.

(21) McDonald, A. J.; Leon, D. R.; Markham, K. A.; Wu, B.; Heckendorf, C. F.; Schilling, K.; Showalter, H. D.; Andrews, P. C.; McComb, M. E.; Pushie, M. J.; et al. Altered Domain Structure of the Prion Protein Caused by Cu2+ Binding and Functionally Relevant Mutations: Analysis by Cross-Linking, MS/MS, and NMR. *Structure* **2019**, *27* (6), 907-922.e5. https://doi.org/10.1016/j.str.2019.03.008.

(22) Hall, Z.; Schmidt, C.; Politis, A. Uncovering the Early Assembly Mechanism for Amyloidogenic β 2 -Microglobulin Using Cross-Linking and Native Mass Spectrometry. *J. Biol. Chem.* **2016**, *291* (9), 4626–4637. https://doi.org/10.1074/jbc.M115.691063.

(23) Deroo, S.; Stengel, F.; Mohammadi, A.; Henry, N.; Hubin, E.; Krammer, E.-M.; Aebersold, R.; Raussens, V. Chemical Cross-Linking/Mass Spectrometry Maps the Amyloid β Peptide Binding Region on Both Apolipoprotein E Domains. *ACS Chem. Biol.* **2015**, *10* (4), 1010–1016. https://doi.org/10.1021/cb500994j.

(24) Schmidt, C.; Urlaub, H. Combining Cryo-Electron Microscopy (Cryo-EM) and Cross-

Linking Mass Spectrometry (XL-MS) for Structural Elucidation of Large Protein Assemblies. *Curr. Opin. Struct. Biol.* **2017**, *46*, 157–168. https://doi.org/10.1016/j.sbi.2017.10.005.

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

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

(27) Fischer, L.; Chen, Z. A.; Rappsilber, J. Quantitative Cross-Linking/Mass Spectrometry Using Isotope-Labelled Cross-Linkers. *J. Proteomics* **2013**, *88*, 120–128. https://doi.org/10.1016/j.jprot.2013.03.005.

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

Chapter 7: Statements of authorship

Statement of Authorship

Title of Paper	The molecular chaperone β-casein prevents amorphous and fibrillar aggregation of α-lactalbumin by stabilisation of dynamic disorder
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Sanders, H. M.; Jovcevski, B.; Carver, J. A.; Pukala, T. L. The Molecular Chaperone β-Casein Prevents Amorphous and Fibrillar Aggregation of α-Lactalbumin by Stabilisation of Dynamic Disorder. Biochem. J. 2020, 477 (3), 629–643. https://doi.org/10.1042/BCJ20190638.

Principal Author

Name of Principal Author (Candidate)	Henry Sanders		
Contribution to the Paper	Majority of experiments, data analysis and manuscript writing.		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	07/12/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Tara Puakla		
Contribution to the Paper	Oversight of project design, data interpretation, editing of manuscript		
Signature		Date	7/12/2020

Name of Co-Author	Blagojce Jovcevski		
Contribution to the Paper	Experimental design, perfoming experiments, data interpretation, editing of manuscript		
Signature		Date	7/12/2020

Please cut and paste additional co-author panels here as required.

Statement of Authorship

Title of Paper	The molecular chaperone β-casein prevents amorphous and fibrillar aggregation of α-lactalbumin by stabilisation of dynamic disorder
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	1. Sanders, H. M., Jovcevski, B., Carver, J. A., and Pukala, T. L. (2020) The molecular chaperone β-casein prevents amorphous and fibrillar aggregation of α-lactalbumin by stabilisation of dynamic disorder. Biochem. J. 477, 629–643

Principal Author

Name of Principal Author (Candidate)	Henry Sanders		
Contribution to the Paper	Majority of synthesis and experimental work, interpreted data and wrote manuscript.		
Overall percentage (%)	75%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	23-Nov-20

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	John A. Carver		
Contribution to the Paper	With Prof. Pukala, I developed the original idea for the project and contributed to editing and fine-tuning of the manuscript, for example in the Discussion section.		
Signature		Date	23 November 2020

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Please cut and paste additional co-author panels here as required.

Statement of Authorship

Title of Paper	Structural and mechanistic insights into the influence of phospholipids in amyloid fibril formation and inhibitor efficacy
Publication Status	☐ Published ☐ Accepted for Publication ☒ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Currently under review with FEBS J.

Principal Author

Name of Principal Author (Candidate)	Henry Sanders		
Contribution to the Paper	Majority of experiments, data analysis and manuscript writing.		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	07/12/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Tara Pukala		
Contribution to the Paper	Oversight of project design, data interpretation, editing of manuscript		
Signature		Date	7/12/2020

Name of Co-Author	Blagojce Jovcevski		
Contribution to the Paper	Experimental design, performing experiments, data interpretation, editing of manuscript		
Signature		Date	7/12/2020

Please cut and paste additional co-author panels here as required.

Statement of Authorship

Title of Paper	Development of a modular synthetic route for protein chemical cross-linking reagents
Publication Status	☐ Published ☐ Accepted for Publication ☒ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	

Principal Author

Name of Principal Author (Candidate)	Henry Sanders		
Contribution to the Paper	Majority of synthesis and experimental work, interpreted data and wrote manuscript.		
Overall percentage (%)	75%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	23-Nov-20

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Kayla Downey		
Contribution to the Paper	Baseline synthesis, data interpretation and experimental methods used from published thesis.		
Signature		Date	27-11-2020

Name of Co-Author			
Contribution to the Paper	Emily R Bubner Assisted in synthesis, analysis and experimentation.		
Signature		Date	07-12-2020

Please cut and paste additional co-author panels here as required.

Statement of Authorship

Name of Co-Author	Katherine G Stevens		
Contribution to the Paper	Contributed to design, synthesis and characterisation of photoreactive crosslinkers, click chemistry experiment design		
Signature		Date	6/12/20

Name of Co-Author	Andrew D Abell		
Contribution to the Paper	Experimental design, editing of manuscript		
Signature		Date	

Name of Co-Author	Tara L Pukala		
Contribution to the Paper	Provided advice related to experimental design and project aims, editing of manuscript		
Signature		Date	4/12/20

Please cut and paste additional co-author panels here as required.